

Analog Circuit Design

RF Circuits: Wide band, Front-Ends, DAC's
Design Methodology and Verification for RF
and Mixed-Signal Systems
Low Power and Low Voltage

Edited by

Michiel Steyaert

Arthur H.M. van Roermund

Johan H. Huijsing



Springer

ANALOG CIRCUIT DESIGN

Analog Circuit Design

RF Circuits: Wide band, Front-Ends, DAC's,
Design Methodology and Verification for RF
and Mixed-Signal Systems, Low Power
and Low Voltage

Edited by

Michiel Steyaert

*Katholieke Universiteit Leuven,
Belgium*

Arthur H.M. van Roermund

*Eindhoven University of Technology,
The Netherlands*

Johan H. Huijsing

*Delft University of Technology,
The Netherlands*

 Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 10 1-4020-3884-4 (HB)
ISBN 13 978-1-4020-3884-6 (HB)
ISBN 10 1-4020-3885-2 (e-book)
ISBN 13 978-1-4020-3885-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Printed on acid-free paper

All Rights Reserved

© 2006 Springer

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 in the Netherlands.

Table of Contents

Preface	vii
Part I: RF Circuitis: wide band, Front-Ends, DAC's	
Introduction	1
Ultrawideband Transceivers	
John R. Long	3
High Data Rate Transmission over Wireless Local Area Networks	
Katelijjn Vleugels.....	15
Low Power Bluetooth Single-Chip Design	
Marc Borremans, Paul Goetschalckx	25
RF DAC's: output impedance and distortion	
Jurgen Deveugele, Michiel Steyaert.....	45
High-Speed Bandpass ADCs	
R. Schreier	65
High-Speed Digital to Analog Converters	
Konstantinos Doris, Arthur van Roermund.....	91
Part II: Design Methodology and Verification for RF and Mixed-Signal Systems	
Introduction	111
Design Methodology and Model Generation for Complex Analog Blocks	
Georges Gielen.....	113
Automated Macromodelling for Simulation of Signals and Noise in Mixed-Signal/RF Systems	
Jaijeet Roychowdhury	143

A New Methodology for System Verification of RFIC Circuit Blocks Dave Morris.....	169
Platform-Based RF-System Design Peter Baltus	195
Practical Test and BIST Solutions for High Performance Data Converters Degang Chen	215
Simulation of Functional Mixed Signal Test Damien Walsh, Aine Joyce, Dave Patrick	243
Part III: Low Power and Low Voltage Introduction	249
The Effect of Technology Scaling on Power Dissipation in Analog Circuits Klaas Bult.....	251
Low-Voltage, Low-Power Basic Circuits Andrea Baschiroto, Stefano D'Amico, Piero Malcovati.....	291
0.5 V Analog Integrated Circuits Peter Kinget, Shouri Chatterjee, and Yannis Tsividis.....	329
Limits on ADC Power Dissipation Boris Murmann	351
Ultra Low-Power Low-Voltage Analog Integrated Filter Design Wouter A. Serdijn, Sandro A. P. Haddad, Jader A. De Lima	369
Wireless Inductive Transfer of Power and Data Robert Puers, Koenraad Van Schuylenbergh, Michael Catrysse, Bart Hermans	395

Preface

The book contains the contribution of 18 tutorials of the 14th workshop on Advances in Analog Circuit Design. Each part discusses a specific to-date topic on new and valuable design ideas in the area of analog circuit design. Each part is presented by six experts in that field and state of the art information is shared and overviewed. This book is number 14 in this successful series of Analog Circuit Design, providing valuable information and excellent overviews of analog circuit design, CAD and RF systems. These books can be seen as a reference to those people involved in analog and mixed signal design.

This years' workshop was held in Limerick, Ireland and organized by B. Hunt from Analog Devices, Ireland.

The topics of 2005 are:

RF Circuits: wide band, front-ends, DAC's
Design Methodology and Verification of RF and Mixed-
Signal Systems
Low Power and Low Voltage

The other topics covered before in this series:

1992 Scheveningen (NL):
Opamps, ADC, Analog CAD

1993 Leuven (B):
Mixed-mode A/D design, Sensor interfaces, Communication
circuits

1994 Eindhoven (NL)
Low-power low-voltage, Integrated filters, Smart power

1995 Villach (A)

Low-noise/power/voltage, Mixed-mode with CAD tools,
Volt., curr. & time references

1996 Lausanne (CH)

RF CMOS circuit design, Bandpass SD & other data conv.,
Translinear circuits

1997 Como (I)

RF A/D Converters, Sensor & Actuator interfaces, Low-noise
osc., PLLs & synth.

1998 Copenhagen (DK)

1-volt electronics, Design mixed-mode systems, LNAs & RF
poweramps telecom

1999 Nice (F)

XDSL and other comm. Systems, RF-MOST models and
behav. m., Integrated filters and oscillators

2000 Munich (D)

High-speed A/D converters, Mixed signal design, PLLs and
Synthesizers

2001 Noordwijk (NL)

Scalable analog circuits, High-speed D/A converters, RF
power amplifiers

2002 Spa (B)

Structured Mixed-Mode Design, Multi-bit Sigma-Delta
Converters, Short Range RF Circuits

2003 Graz (A)

Fractional-N Synthesis, Design for Robustness, Line and Bus
Drivers

2004 Montreux (Sw)

Sensor and Actuator Interface Electronics, Integrated High-Voltage Electronics and Power Management, Low-Power and High-Resolution ADC's

I sincerely hope that this series provide valuable contributions to our Analog Circuit Design community.

Michiel. Steyaert

Part I: RF Circuits: wide band, Front-Ends, DAC's

The trends in RF circuits is since several years towards the fully integration. Secondly we see the further requirements to the needs of higher data-rates and as such higher bandwidths. The discussion of shifting the digital boundary closer and closer to the antenna, results in the ever increasing requirements for AD/DA converters. For that in this part the different trends are discussed through different systems over different building blocks.

The first topic addresses a total different communication systems: UWB (Ultra Wide Band). There are two approaches towards ultra wide band systems. The first is an 'extension' of the WLAN OFDM system, the other one is the use of impulse radio systems. For the later one, new architecture, new circuit structures and new topologies are required. As such still a long way has to be performed. The first approach is basically an extension of the WLAN. For that the second paper handles in detail WLAN systems. The many years of research in those systems has resulted nowadays in the extremely high integration in RF CMOS. The third paper addresses even more standard products, Bluetooth devices. Low power, fully integration deals nowadays with digital interference effects and the requirements towards deep submicron.

The next three papers deal with the AD/DA converters. The first and the last one deal with design issues for DAC topologies with clock rates reaching 1GHz. It is clear that at that moment dynamic performances are becoming the dominant issue. Especially the finite output impedance of the topologies, in combination with the signal dependency results in important distortion components. The last paper deals in detail about the timing issues and design trade offs for the master-slave latch and driver circuits. The last but one paper discusses high speed band pass ADC's. To perform AD conversion at a high IF frequency, continuous time ADC's are proposed. Implementation issues are discussed and by using active RC integrators, high performance and high integration can be obtained, however at the cost of power drain.

Michiel Steyaert

ULTRAWIDEBAND TRANSCEIVERS

John R. Long
Electronics Research Laboratory/DIMES
Delft University of Technology
Mekelweg 4, 2628CD Delft, The Netherlands

Abstract

An overview of existing ultrawideband (UWB) technologies is presented in this paper, including multi-band OFDM (MB-OFDM, scalable for data rates from 55-480Mb/s). Time-domain impulse radio and wideband FM approaches to UWB for low (<100 kb/s) and medium data rates (100 kb/s-10 Mb/s) are also described.

1. Introduction

Ultrawideband (UWB) communication technology is defined as any scheme that occupies more than 500MHz bandwidth, or where the ratio of channel bandwidth to centre frequency is larger than 20%. Early UWB system development concentrated on imaging radar, which is used for precise location finding and imaging. The recent interest in UWB communication systems arises from the desire for high-speed, short-range networking (e.g., to support multimedia applications), although UWB technology can also be used in low power, low bit-rate applications. UWB has the potential to support a number of applications more effectively than other short-range wireless alternatives, such as the 802.11 or Bluetooth systems, as illustrated by the data throughput versus distance curves of Fig. 1. The IEEE 802.15.3a group has proposed a physical layer standard for IC development that has led to the development of commercial UWB chipsets by a number of vendors.

The motivation for wideband transmission can be seen from Shannon's theorem, which relates the signal-to-noise ratio (S/N) and bandwidth (W) of a system to the channel capacity (C). For low S/N ratios,

$$C = W \log_2(1 + S/N) \approx W(S/N) \quad \text{Eq. 1.}$$

Eq. 1 predicts that capacity can be improved by either increasing the effective signal-to-noise ratio or by increasing the system bandwidth. For conventional narrowband systems, bandwidth improvements have been realized by decreasing

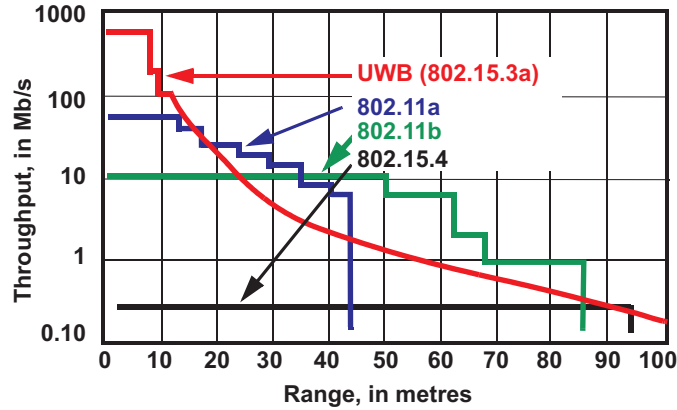


Fig. 1: Comparison of data throughput and range for IEEE 802 standards.

the range (thereby decreasing the S/N ratio) or through the use of error correcting coding. The GHz bandwidths available in an ultrawideband system allows large increases in capacity without compromising range or adding overhead by coding. The recent ruling by the Federal Communications Commission in the United States permits use of the 3.1-10.6GHz band for communications with a average power spectral density (PSD) to less than -41dBm (measured in a 1MHz bandwidth using an isotropic antenna) as shown in Fig. 2. By restricting the PSD, the received power is constrained at a given distance. The typical S/N ratio will be low (approx. 0dB) for these systems. Therefore, using as much of the allocated bandwidth as possible is the most effective way of achieving higher data rates, although advanced forward error correcting codes may be used (at the cost of complexity) to realize further gains. A few of the commercial narrowband systems shown in Fig. 2, such as DCS-1800 and 802.11 LAN (not to scale) are strong sources of potential interference, and so co-existence of UWB with other systems must be addressed in any practical system implementation.

2. Multiband OFDM (MB-OFDM)

The proposed standard for high data-rate applications using UWB technology (IEEE 802.15.3a) is multiband OFDM [1], which offers bit rates ranging from 55 to 480 Mbit/s. In the proposed standard, the 3-10GHz spectrum approved for indoor use is divided into 14 bands that are 528MHz wide. For the first generation of MB-OFDM systems, potential interference from WLAN and other commercial sources are limited, as only bands 1-3 are used (see Fig. 3). These bands lie between the 2.4GHz ISM and 5-6GHz bands used by 802.11 WLAN. MB-

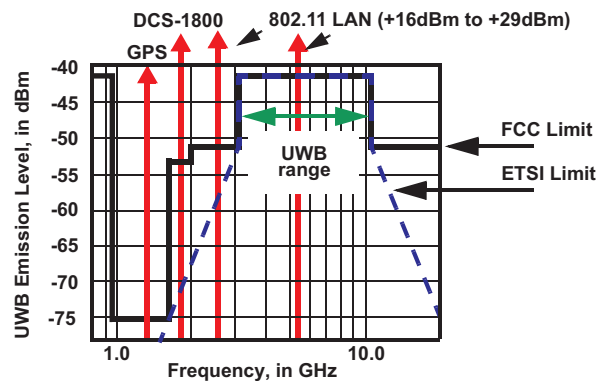


Fig. 2: UWB indoor spectral mask.

OFDM is therefore scalable, and channel capacity can be added as technology improves or capacity requirements increase by adding more 528MHz wide bands to the system.

The OFDM symbols are interleaved across all transmit bands to add frequency diversity into the system and provide robustness against multi-path and other types of interference. One advantage of using OFDM, is that tones can be switched off near frequencies (or in bands) which must be protected from interference. Since each MB-OFDM band is only 528MHz wide, this reduces the demands on the bandwidth of the signals which the transmitter and receiver must process. A guard interval is inserted between OFDM symbols in order to allow sufficient time to with between channels, however, switching must be achieved within 9ns.

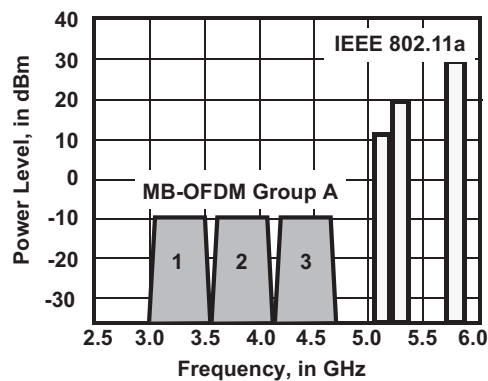


Fig. 3: Frequency bands proposed for the first generation of MB-OFDM.

The architecture of the MB-OFDM transmitter and receiver is similar to other OFDM systems. This allows manufacturers to leverage existing OFDM designs

for the development of MB-OFDM ICs. Restrictions on the transmit constellation size and signal processing overhead allow simplified implementations. For data rates below 80Mb/s a full I/Q transmitter is not required. This reduces the size of the analog portion of the transmit chain on an IC by about one-half.

One method of implementing a fast switching source for bands 1-3 is shown in Fig. 4. the centre frequencies for the sub-bands are 3432, 3960 and 4488MHz, respectively. Frequency division from a master PLL source produces a number of sub-frequencies, and single-sideband mixers are then used to combine the desired tones to create local oscillators centred in each sub-band under digital control (the select function in Fig. 4).

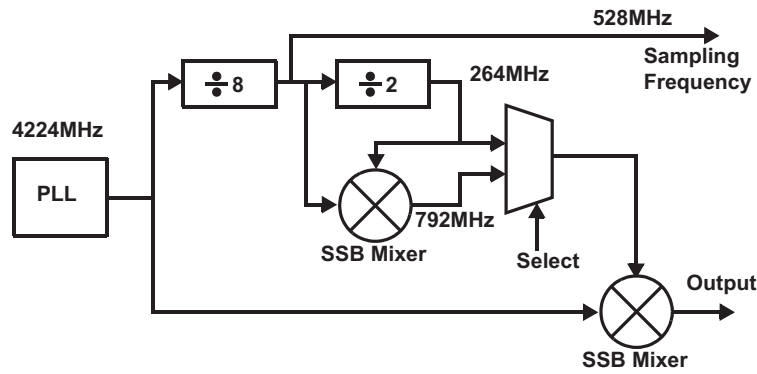


Fig. 4: Fast-switching frequency synthesis.

On the transmit side, OFDM produces a peak-to-average ratio of 21dB for the transmit signal. The required RF power output is

$$-41.25\text{dBm} \cdot \text{MHz} + 10\log(528) = -14\text{dBm} \quad \text{Eq. 2.}$$

Adding 10dB margin to ensure linearity and assuming a Class A (linear) power amplifier with 10-20% efficiency, the dc power consumption required is

$$P_{\text{DC}} = P_{\text{ac}}/\eta = -4\text{dBm}/0.1 = 4\text{mW} \quad \text{Eq. 3.}$$

Other circuitry will swamp out power consumption of the power amplifier, unlike other wireless systems where the power consumed by the power amp dominates.

A simulated link budget [1] for the 110Mbps/s data rate predicts a 6.6dB noise figure receiver is required with a sensitivity of -80.5dBm (assuming a 3-band transceiver, -10.3dBm transmit power, 6dB link margin and 0dBi gain antennas). Power consumption of a 130nm CMOS implementation operating at 110Mb/s is

projected to consume 156mW in transmit and 205mW in receive modes, and require 7.1mm² die area. Power consumption falls to 128mW in transmit and 155mW in receive modes when scaled to the 90nm technology node in CMOS, and would require 5.2mm² die area.

3. Time Domain Impulse Radio

The first UWB radio systems used a sequence of short-duration pulses to convey information in a time-domain radio. Each pulse, or wavelet, consists of a number of cycles of a sinusoid with a Gaussian-shaped amplitude envelope. In the frequency domain, each pulse has a broad spectral shape with a slow roll-off. For example, a 5 cycle Gaussian wavelet satisfies the FCC mask of Fig. 2 after out-of-band filtering by the transmit antenna. In a low-cost communication system, a simple wavelet and modulation scheme are chosen to simplify the implementation and minimize power consumption. Data is typically encoded as a sequence of pulses with time-varying position (PPM) with a peak amplitude on the order of 100mV. Data scrambling or coding is used to ensure sufficient timing information for extraction at the receiver. Low gain wideband antennas provide only a modest gain, so pulse amplification is required before detection of the pulses using a time correlator and timing extraction, as illustrated in Fig. 5.

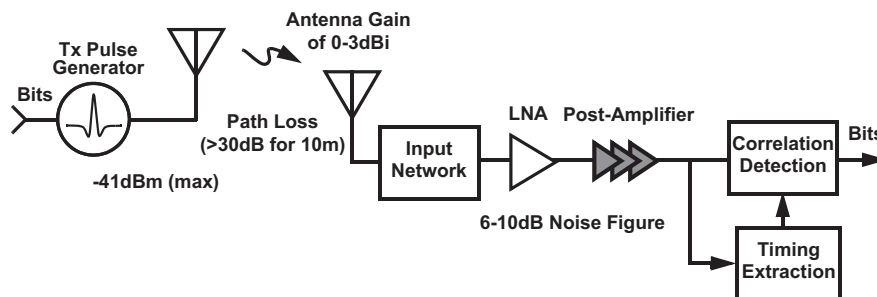


Fig. 5: Impulse radio transceiver.

UWB radar pulse generators use step recovery diodes, which require individual circuit trimming and a number of discrete components. Digital pulse generation circuits based on simple building blocks such as counters, flip-flops and logic gates have the advantage that they are scalable, reproducible, and compatible with other digital VLSI circuits in a system-on-a-chip (SoC) implementation. It should be noted that only a small amount of analog filtering is required to make the output signal compliant with the FCC or ETSI spectral masks.

In the correlation receiver of Fig. 5, a locally-generated wavelet is compared and

matched with the received pulse. Any distortion of the received pulse due to the time-varying channel characteristics, or bandlimiting by the antennas and receiver preamplifiers) makes correlation difficult.

Using a transmitted reference scheme, as proposed by Hoctor and Tomlinson [2], alleviates these problems and is less complex than the rake receivers used in other time-domain UWB systems. An “autocorrelation receiver” [3] based on this concept is shown in Fig. 6. Two pulses per symbols are transmitted, separated by delay, τ_d . The first pulse is the reference for the second pulse, and the relative phase between the pulse doublet is modulated in time by the transmit data. At the receiver, the first pulse is delayed by τ_d , and then correlated with the second pulse (i.e., the received reference pulse becomes the template used for correlation). Aside from providing a more accurate reference for correlation, this receiver does not require pulse synchronization. The delay between the two pulses is used for synchronization.

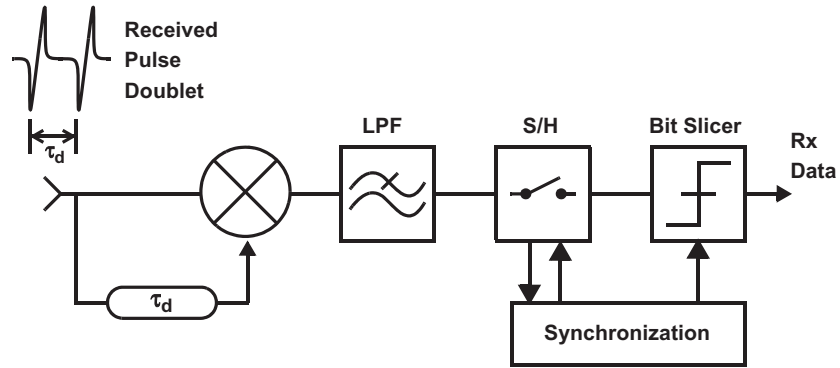


Fig. 6: Autocorrelation receiver.

4. Ultrawideband FM

Frequency modulation has the unique property that the RF bandwidth B_{RF} is not only related to the bandwidth f_m of the modulating signal, but also to the modulation index β , which can be chosen freely. The approximate bandwidth of a UWB-FM signal is given by Carson’s rule

$$B_{RF} = 2(\beta + 1)f_m = 2(\Delta f + f_m) \quad \text{Eq. 4.}$$

Choosing ($\beta \gg 1$) yields an ultrawideband signal that can occupy a bandwidth within the RF oscillator’s tuning range. The signal bandwidth can be easily

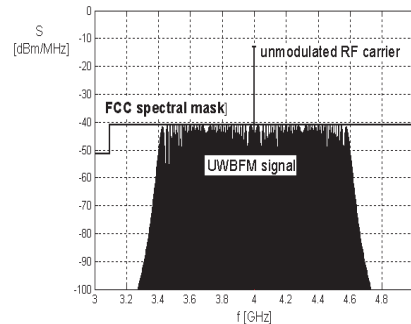


Fig. 7: 4GHz carrier and UWB-FM signal ($f_{SUB} = 1$ MHz and $\beta = 600$) [5].

adjusted by modifying the deviation (Δf) of the wideband FM signal.

The power spectral density of the wideband FM signal has the shape of the probability density function of the modulating signal, so a triangular sub-carrier with a uniform probability density function gives a flat RF spectrum. A triangular sub-carrier is relatively straightforward to generate using integrated circuit techniques.

Fig. 7 shows an example of the spectral density of a UWB-FM signal obtained using a -13dBm triangular sub-carrier. The sub-carrier frequency is 1 MHz and the deviation Δf is 600 MHz (modulation index $\beta=600$). The spectral density is lowered by a factor of $10 \log_{10}(\beta) = 28$ dB. This UWB signal is FCC compliant.

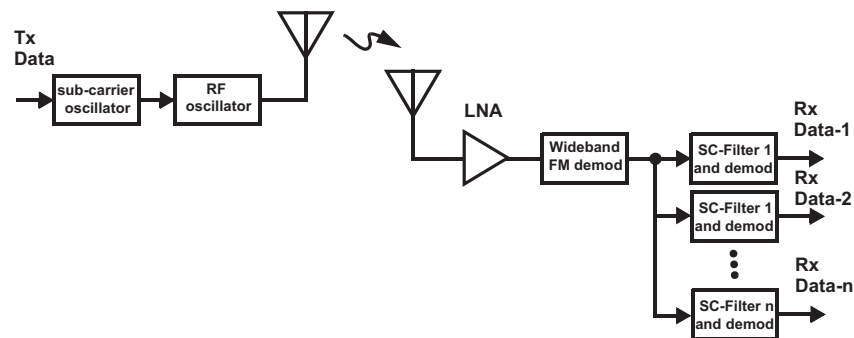


Fig. 8: UWB-FM transceiver block diagram.

The block diagram of a UWB-FM transceiver is shown in Fig. 8 [4]. Digital data is modulated on a low-frequency sub-carrier (typically 1 MHz for 100 kbit/s data) using FSK techniques (e.g., modulation index $\beta_{SUB} = 2$). The modulated sub-carrier modulates the RF oscillator, yielding the constant envelope UWB sig-

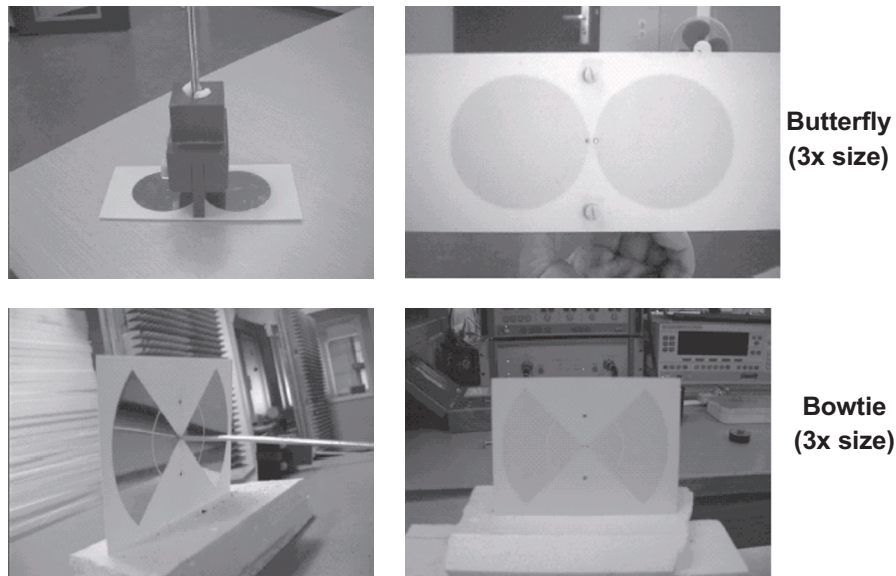


Fig. 9: Examples of ultrawideband antennas [6].

nal.

The receiver demodulates the UWB-FM signal without frequency translation. It is relatively simple, and no local oscillator or carrier synchronization are required. Fig. 16 shows a block diagram of a UWB-FM receiver. The receiver consists of a wideband FM demodulator, and one (or several) low-frequency sub-carrier filtering, amplification stages, and sub-carrier filters and demodulators. Because of its simplicity, extensive hardware or power-hungry digital signal processing are not needed to implement a UWB-FM transceiver. This scheme is ideally suited to low and medium data-rate applications, where battery lifetime, weight and form factor are the most important design considerations.

4. Ultrawideband Hardware

Ultrawideband systems also have special hardware considerations. Antennas must be capable of transmitting or receiving GHz bandwidth signals with minimal dispersion, distortion and attenuation. Their response cannot be peaked using LC resonant circuits, because this narrows the bandwidth and causes phase distortion. Two examples of planar ultrawideband antennas are shown in Fig. 9. When scaled for use in the 3-10GHz frequency range, these 50Ohm antennas are

approximately $4 \times 3 \text{ cm}^2$ in area, and have a gain of approximately 2dBi. Antenna gain is low for omnidirectional designs, and received power is proportional to $1/f^2$ (2 antennas), so attenuation over a 10m link can be greater than 30dB. Path loss is much greater when obstructions are present.

Broadband amplifiers are needed to compensate for the path loss between transmitter and receiver. A low-noise preamplifier topology suitable for CMOS integration is shown in Fig. 10. The input and output matching networks must cover a much wider range in frequency compared to a narrowband preamplifier. At the output, a source follower stage provides a simple broadband interface between the low Q-factor resonant load of cascode amplifier M1/M2 and subsequent stages.

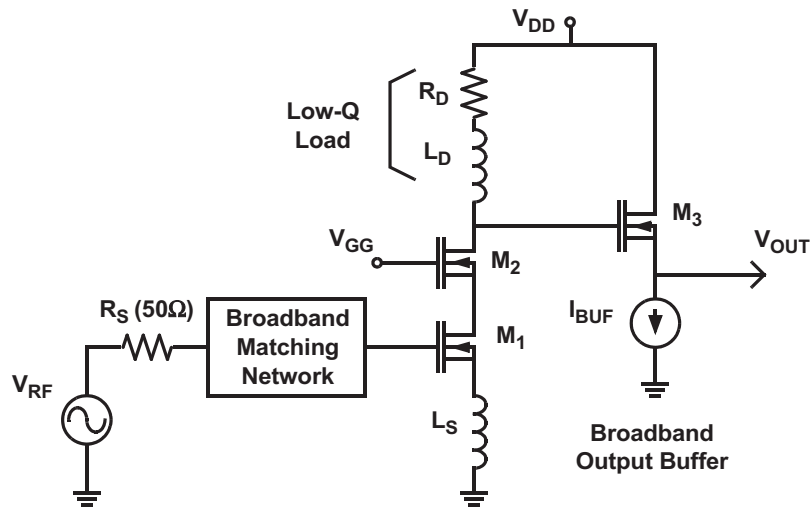


Fig. 10: Multi-octave low-noise amplifier topology.

On the input side, a multi-LC section matching network is used. The reflection coefficient of a passive LC ladder matching network with Chebyshev coefficients is illustrated in Fig. 11. A 2 or 3-stage matching network is needed to realize the desired reflection coefficient for a 3-10GHz application. Gyration of the source inductance together with the input capacitance is used to set the input termination resistance as part of a series resonant circuit. Implementing the matching network using on-chip components introduces losses that compromise the noise figure of the amplifier and consume valuable chip area. On the other hand, an off-chip matching network requires trimming in manufacture and increases the number of components and form factor of the radio.

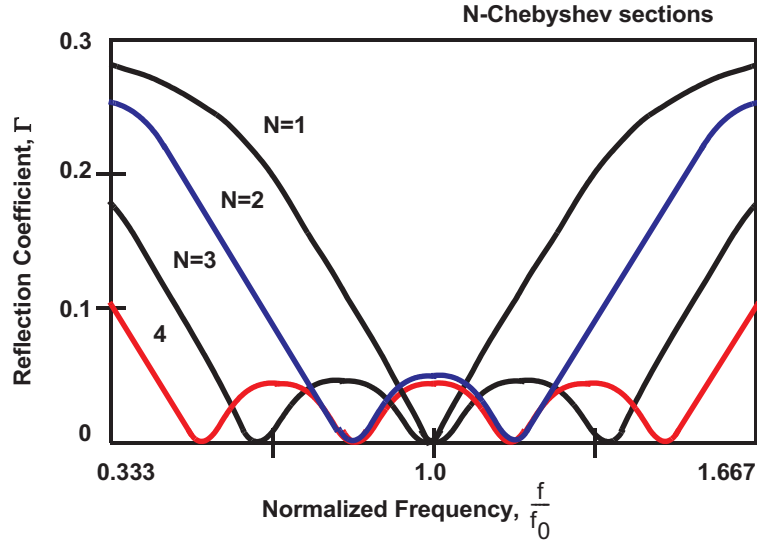


Fig. 11: Reflection coefficient of multi-section matching networks.

Recently reported results from 2 ultrawideband LNAs are listed below in Table 1. The CMOS preamplifier uses a 3 stage LC ladder network for input matching, while the SiGe bipolar amplifier uses a simpler 2 stage design. The power consumption of the bipolar amplifier is 3 times higher than the CMOS LNA, however, it has over 12dB more gain and 4.5dB lower noise figure.

Table 1: Performance comparison of recently reported UWB LNAs.

Technology	Gain	S11 (min.)	IIP3	P _D	NF
0.18 μ m SiGe [8]	22dB	-10dB	-5.5dBm (3.5GHz)	27mW (2.7V)	< 4.5dB (3-10GHz)
0.18 μ m CMOS [7]	9.3dB	-10dB	-6.7dBm (6GHz)	9mW (1.8V)	< 9dB (3-10GHz)

5. Conclusions

Ultrawideband technology in its various forms offers high data throughput for a given level of power consumption when compared with conventional radios.

Multiband OFDM can coexist with other systems and conform to world-wide standards and regulations. It offers an efficient trade-off between bit-rate, complexity and power consumption for short range use (< 20m), and can be implemented efficiently in CMOS by leveraging existing designs developed for

802.11 WLAN systems. It is already very close to a commercial reality in applications such as wireless USB interfaces.

Other UWB schemes, such as time-domain and UWB-FM radio, are better suited to low data rate applications, but they are further from commercialization. However, these simpler modulation schemes and hardware will consume less power and can reduce radio complexity and form factor. A time-domain solution has the additional advantage of compatibility and scalability with VLSI design methodologies and circuits.

References

- [1] Anuj Batra, et al, "Multi-band OFDM Physical Layer Proposal", document IEEE 802.15-03/267r6, September 2003.
- [2] R.T. Hocht and H.W. Tomlinson, "Delay-Hopped Transmitted-Reference RF Communications", Proceedings of the IEEE Conference on Ultra Wideband Systems and Technologies, pp. 265-270, May 2002.
- [3] S. Lee, S. Bagga and W.A. Serdijn, "A Quadrature Downconversion Auto-correlation Receiver Architecture for UWB", Proc. UWBST & IWUWBS2004, Kyoto, Japan, May 19-21, 2004.
- [4] J. Gerrits and J. Farserotu, "Ultra Wideband FM: A Straightforward Frequency Domain Approach", Proc. of the European Microwave Conference, Munich, Germany, October 2003, pp. 853 – 856.
- [5] J.F.M. Gerrits, J. Farserotu and J.R. Long, "UWB Considerations for MAGNET Systems", Proceedings of the ESSCIRC, Leuven BE, 2004, pp. 45-56.
- [6] Photos courtesy of A. Yarovoy, IRCTR-TU Delft.
- [7] A. Bevilacqua and A. Niknejad, "An Ultra-Wideband CMOS LNA for 3.1 to 10.6 GHz Wireless Receivers," ISSCC 2004, pp. 382 – 383.
- [8] A. Ismail and A. Abidi, "A 3 to 10 GHz LNA Using Wideband LC-ladder Matching Network," ISSCC 2004, pp. 384 – 385.
- [9] "First report and order, revision of part 15 of the commission's rules regarding ultra-wideband transmission systems", FCC, Washington DC, ET Docket 98-153, 2002.
- [10] J. Foerster, E. Green, S. Somayazulu, and D. Leeper, "Ultrawideband

technology for short- or medium-range wireless communications,” *Intel Technology Journal*, 2001.

- [11] IEEE 802.15-03/157r1: Understanding UWB – Principles & Implications for Low-Power Communications.
- [12] M. Z. Win and R. A. Scholtz, “Impulse radio: How it works,” *IEEE Comm. Letters*, vol. 2, pp. 36–38, February 1998.
- [13] A/D precision requirements for an ultra-wideband radio receiver: Newaskar, P.P.; Blazquez, R.; Chandrakasan, A.R.; *Signal Processing Systems*, 2002. [SIPS '02]. IEEE Workshop on, 16-18 Oct. 2002; Pages:270 – 275.

HIGH DATA RATE TRANSMISSION OVER WIRELESS LOCAL AREA NETWORKS

Katelijjn Vleugels
H-Stream Wireless Inc., Los Altos, California 94022, USA

Abstract

This paper discusses new trends in emerging WLAN systems and applications, and their implications on the architecture and circuit implementation of next-generation 802.11 communication ICs. Higher data rates, longer transmission ranges, lower cost and higher system capacity are putting new constraints on the baseband and RF circuits constituting future 802.11 transceivers. Several new circuit techniques drawn from recent publications [2]-[6] that address such constraints are presented.

1. Introduction

Recent years have seen a tremendous growth in the deployment of Wireless Local Area Networks (WLANs) in the home, the enterprise as well as public hot spots. The proliferation of WLANs can be attributed to a number of factors including the adoption of the 802.11 industry standard and rigorous interoperability testing, the development of higher performance wireless LAN equipment, rapid reductions in product pricing and the increased importance of user convenience and mobility.

The three WLAN standards in existence today are based on the IEEE 802.11b, the IEEE 802.11g and the IEEE 802.11a specifications respectively. The 802.11b- and 802.11g-based products both operate in the 2.4-GHz unlicensed ISM band with an aggregate bandwidth of 83.5 MHz and a total of 11 channels, only three of which are non-overlapping. The 802.11a-based products operate in one of three subbands of the 5-GHz unlicensed national information infrastructure (UNII) band with an aggregate bandwidth of 580 MHz, supporting 24 independent 802.11 channels.

The higher achievable data rates and better spectral efficiency make 802.11g and 802.11a the preferred standards in high data rate applications. Furthermore,

when overall system capacity is a concern, 802.11a is the standard of choice because of the significantly larger number of non-overlapping channels available in the 5-GHz band.

The 802.11g and 802.11a standards are both based on an orthogonal frequency-division multiplexing (OFDM) modulation technique, supporting data rates as high as 54 Mbps, or even 108 Mbps when special channel bonding techniques are used. The OFDM modulated signal consists of 52 carriers, occupying a 20 MHz channel. Each carrier is 312.5 kHz wide, giving raw data rates from 125 kbps to 1.125 Mbps per carrier depending on the modulation type (BPSK, QPSK, 16-QAM or 64-QAM) and the error-correction rate code (1/2 or 3/4).

This paper first identifies the trends evolving from new WLAN applications like home entertainment, and its adoption in PDAs or cell phones. Next, we investigate the implications of such trends on the architecture of next-generation 802.11 WLAN RFICs. The paper concludes with a detailed discussion of circuit implementation techniques to improve the performance of high data rate 802.11 WLAN RF transceivers.

2. Trends in 802.11 wireless LAN

Driven by its successful widespread deployment and its adoption in new and emerging applications, the development of 802.11 wireless LAN systems is experiencing a trend towards higher data rates, longer transmission ranges, lower cost and higher system capacity.

As described in [1], different approaches are being pursued to obtain even higher data rates beyond 54 Mbps. In one approach, the signaling rate is doubled resulting in a wider bandwidth transmission. This technique, also referred to as the channel bonding mode increases the achievable raw data rate from 54 Mbps to 108 Mbps, yet increases the channel bandwidth from about 17 MHz to 34 MHz. While the channel bonding mode has little effect on the implementation of the RF circuitry, it imposes more stringent requirements on the baseband circuitry inside the RFIC. It requires the cutoff frequency of the baseband filters in both the transmitter and receiver paths to be programmable by a factor of two. Furthermore, the higher signal bandwidth means that the analog-to-digital converters in the receiver path, and the digital-to-analog converters in the transmit path must be clocked at twice the sampling frequency, resulting in a higher power dissipation.

Alternatively, higher data rates can be achieved by using multiple transmit and receive chains in parallel. This technique is also referred to as Multiple Input

Multiple Output (MIMO) systems. A first MIMO method, MIMO A/G, uses beamforming and Maximum Ratio Combining (MRC) to extend the range at which a given data rate will succeed. It is compatible with existing 802.11 a/g equipment, and although the best performance is achieved when implemented at both ends of the communication link, more than half of the potential benefit can be realized when only one end implements the technique [1]. A second MIMO method, MIMO S/M, used more sophisticated encoding techniques to fully leverage the availability of multiple transmit and receive chains. The advantage of MIMO S/M over MIMO A/G is that, in addition to extending the range at which a given data rate succeeds, it also allows to increase the data rates over short distances. The main disadvantages of MIMO S/M are the increased digital implementation complexity, the incompatibility with existing equipment, and the requirement to have MIMO S/M implemented at both ends of the communication link before any performance improvement is achieved.

Where the trends for higher data rates and longer ranges are leading to more complex, and potentially more power hungry systems, wide adoption sets strict requirements on production yield, cost and the number of external components used in the system. Highly integrated System-on-Chips (SoCs) are presented in [2] and [3]. By integrating all the radio building blocks, as well as the physical layer and MAC sections into a single chip, a cost-effective solution can be obtained. Furthermore, to minimize the number of external components, a radio architecture amenable to a high level of integration is critical. Two radio architectures, the direct conversion architecture and the two-step sliding IF architecture, as proposed in [4] can meet the integration requirements of future 802.11 RFICs. The overall component count can furthermore be decreased by integrating external components such as the antenna T/R switch, the power amplifier (PA), the low-noise amplifier (LNA) and RF baluns [3].

Driven by new emerging applications like wireless Voice-over-IP, the limited system capacity available in the 2.4-GHz band is gradually becoming more of a concern. With only 3 non-overlapping channels and interference from other wireless technologies like cordless phones, microwaves and Bluetooth, the 2.4-GHz band is approaching saturation. While 802.11a, operating in the 5-GHz band holds great promise because of its much larger capacity, dual-band operation RF transceivers will be important to ensure compatibility with the existing infrastructure of 802.11 b and g equipment, while at the same time leveraging the much larger system capacity available in the 5-GHz band. It is common for dual-band transceivers to share IF and baseband circuits, while each frequency band typically has its dedicated RF circuits. Dual-band operation also impacts the implementation of the synthesizer, in that it has to provide the Local Oscillator (LO) signals for both the 2.4-GHz and 5-GHz transmit and receive paths.

3. Recent developments in 802.11 wireless LAN RFICs

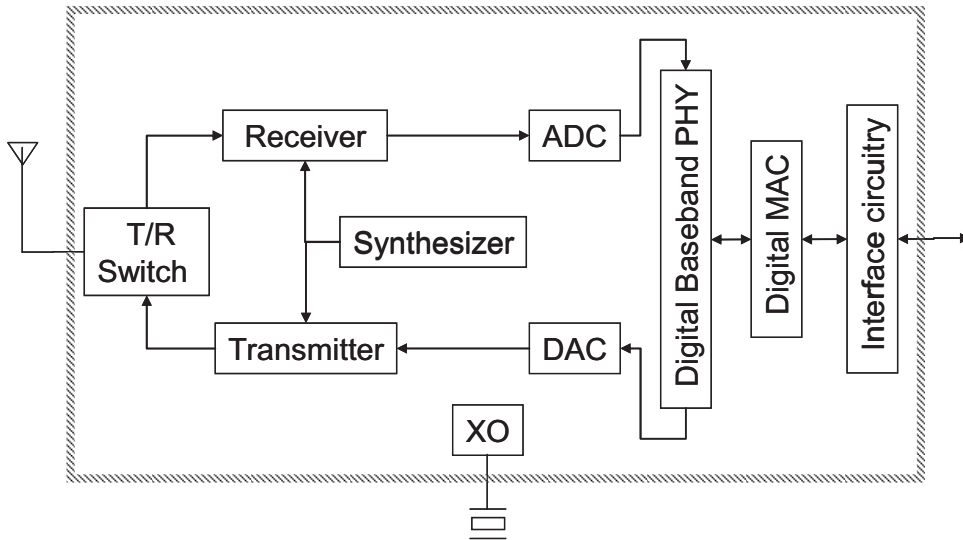


Fig.1. Block diagram of an integrated IEEE 802.11 WLAN SoC

Fig.1 shows the block diagram of a highly integrated 802.11 WLAN SoC, integrating all of the radio components as well as the physical layer and MAC sections. The IC essentially connects the RF antenna to the digital host computer. One big advantage of integrating the analog and digital components on a single SoC is that closed-loop RF calibration techniques can be used to correct for analog circuit non-idealities [2].

However, single-chip integration can affect the system performance in several ways. The coupling of switching noise on the supply and bias voltages can reduce the receive sensitivity, elevate the phase noise and degrade the Error Vector Magnitude (EVM) of the transmitted signal. Several techniques to prevent the corruption of sensitive analog and RF signals are proposed in [2]. They include the use of fully differential circuits to obtain first-order rejection of common-mode digital switching noise, the use of separate or star-connected power supplies to reduce supply crosstalk, the use of on-chip voltage regulators for sensitive circuits, and the use of a deep N-well trench to reduce substrate coupling.

As mentioned earlier, two transceiver architectures amenable to high levels of integration, are the direct conversion architecture and the two-step sliding IF architecture. Two well-known problems of the direct conversion architecture is DC offset and pulling issues between the power amplifier and the synthesizer.

To overcome VCO pulling, different frequency planning schemes have been proposed. In [5], the voltage-controlled oscillator (VCO) operates at two or four times the LO frequency, depending on the frequency band. To accommodate this, the VCO's effective operation range must be from 9600-11800 MHz, increasing the synthesizer's power dissipation. Furthermore, pulling can still occur in the presence of a strong second harmonic in the transmitter. Alternatively, an offset VCO architecture as in [6] can be used. This is illustrated in Fig.2. Since the VCO is running at 2/3rds of the transmitter carrier frequency, no pulling will occur. Furthermore, the excessive power dissipation with running the VCO at 10 GHz is avoided, at the cost of an extra mixing operation in the LO path.

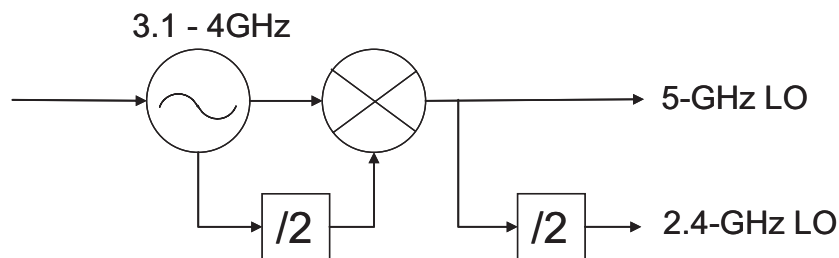


Fig.2. Offset VCO architecture

In [4], the pulling problem is overcome by using a two-step conversion, rather than a direct conversion architecture. As illustrated in [1], the two-step conversion approach can be very similar to the approach of Fig.2, with the main difference that the additional mixing occurs in the signal path, rather than the LO path. This is illustrated in Fig.3. The two approaches become even more similar if the first and second mixing stages in the two step approach are merged into a single stage with a single set of RF inductive loads as shown in Fig.4.

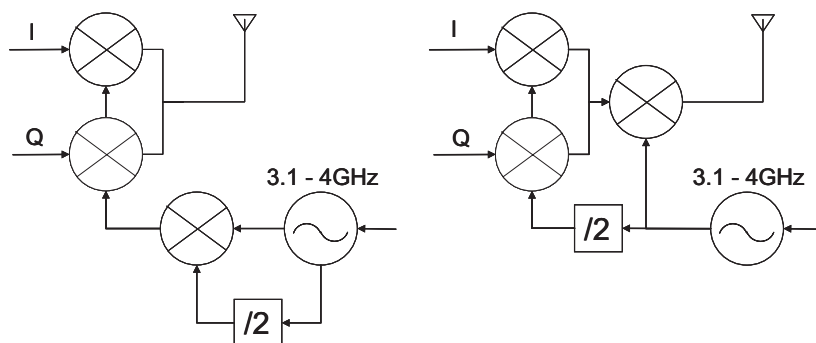


Fig.3. Direct conversion versus two-step conversion radio architecture

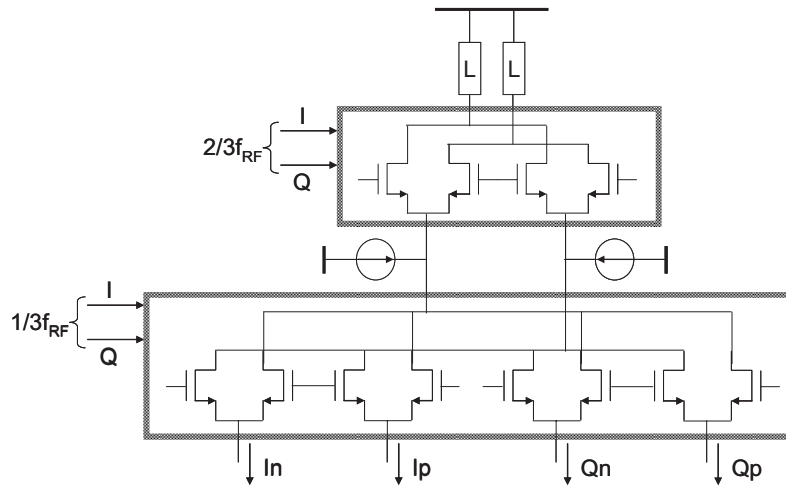


Fig.4. Stacked mixer topology with single inductive load

4. Circuit implementation techniques

This section presents a select set of circuit techniques that can be implemented to meet the requirements of future 802.11 RFICs. Circuits discussed include a high-frequency divide-by-two circuit with switchable inductors [4], an integrated dynamically biased power amplifier [4], and on-chip T/R switch [3].

4.1. High-frequency divide-by-two circuit with switchable inductors

The high-frequency divide-by-two circuit, discussed in [4] and shown in Fig.5, consists of two inductively loaded current-mode flip-flops in a feedback loop. The inductive loads tune out the capacitive load associated with the feedback divider, the I and Q buffers, as well as parasitic wiring capacitance. The locking range of the divider is increased using switchable inductors. The use of switchable inductors, rather than switched capacitors has the advantage that a relatively constant output load impedance is achieved across the wide frequency range [4].

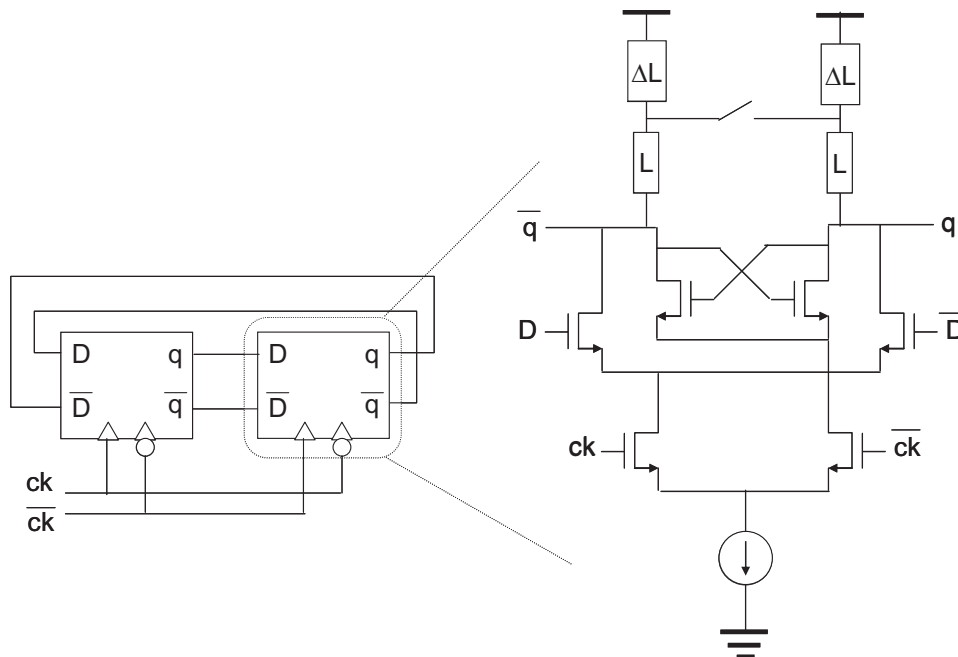


Fig.5. High-frequency divide-by-two circuit with switchable inductors

4.2. Integrated dynamically biased power amplifier

To reduce the external component count, it is desirable to integrate external components like the external power amplifier. However, the high linearity requirements of accommodating data rates up to 54 Mbps tends to result in excessive power consumption when the external PA is integrated with the 802.11 RFIC. Typically, a class-A power amplifier is used to meet the stringent linearity requirements, where a fixed dc bias current is chosen to accommodate the peak signals, resulting in a poor PA power efficiency if the signal amplitude is below the maximum level. Especially for modulation schemes like OFDM, with large a peak-to-average ratios (PAR), the power dissipation can be reduced substantially with a dynamically biased output stage whose DC current is proportional to the envelope of the output signal [6]. A dynamically biased power amplifier was proposed in [4] and simplified schematic is shown in Fig.6. A dynamically biased amplifier dissipates very low power at small-signal levels, and the power dissipation increases only during the signal, making it very attractive for high data rate OFDM signaling where the peak-to-average ratio can be as large as 17 dB, but signal peaks are very infrequent [4].

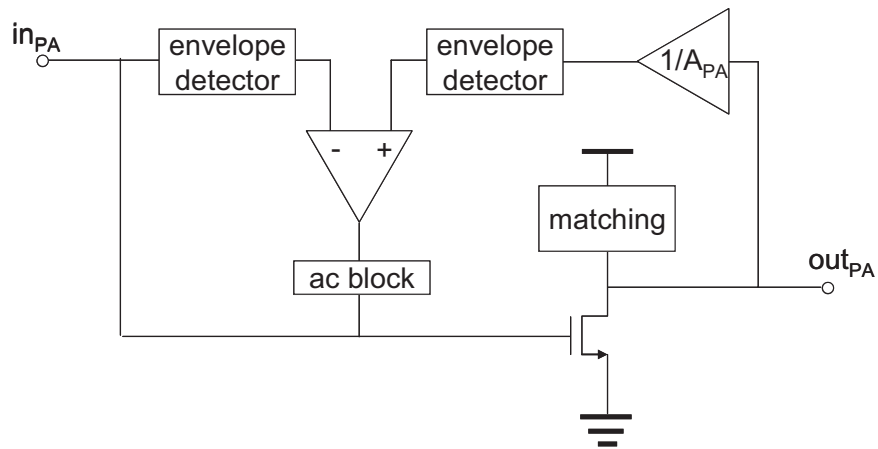


Fig.6. Dynamically biased power amplifier

4.3. Integrated T/R switch

T/R switch circuits typically serve a dual purpose: performing antenna selection as well as time division duplexing (TDD) between the receiver (Rx) and transmit (Tx) paths. To accommodate both functions, two series transistors are typically needed, introducing significant insertion loss.

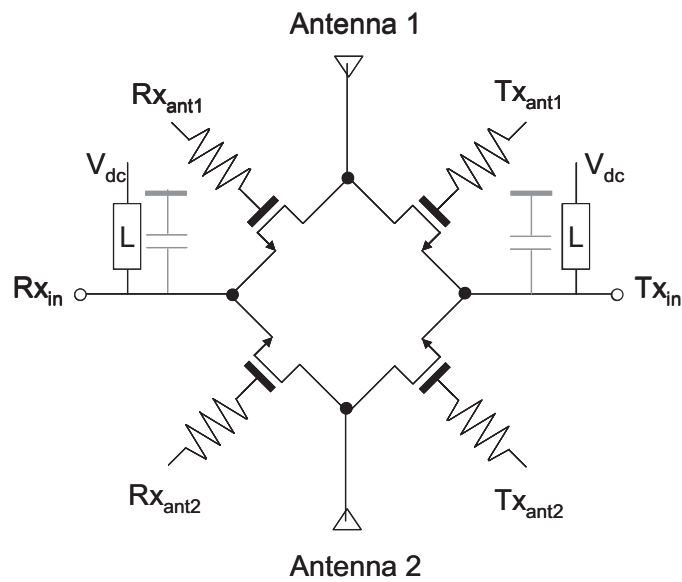


Fig.7. On-chip T/R switch circuit

The ring structure-based on-chip T/R switch proposed in [3] and shown in Fig.7 requires only a single switch in the signal path, resulting in a measured insertion loss of about 1.8 dB while at the same time achieving better than 15dB of isolation in each node.

5. Conclusions

In this paper we discuss new trends evolving from emerging new WLAN applications, and their impact on the architecture and circuit implementation requirements for next-generation 802.11 WLAN RFICs. Higher data rates and dual band operation are putting more stringent requirements on the linearity and tuning range requirements of the RF circuits. A dynamic biasing technique is proposed as a way of achieving high linearity without excessive power dissipation. Finally, low cost is the driving force behind highly integrated single-chip communication SoCs with no or minimal number of external components. Integration of the power amplifier and T/R switch circuits are important steps in that direction.

References

- [1] W. J. McFarland, "WLAN System Trends and the Implications for WLAN RFICs," IEEE Radio Frequency Integrated Circuits Symposium, 2004.
- [2] S. Mehta, et al., "An 802.11g WLAN SoC," ISSCC Dig. Tech. Papers, pp. 94-95, Feb. 2005.
- [3] H. Darabi et al., "A Fully Integrated SoC for 802.11b in 0.18 μ m CMOS," ISSCC Dig. Tech. Papers, pp. 96-97, Feb. 2005.
- [4] M. Zargari, et al., "A Single-Chip Dual-Band Tri-Mode CMOS Transceiver for IEEE 802.11a/b/g Wireless LAN," IEEE J. Solid-State Circuits, Vol. 39, pp. 2239-2249, Dec. 2004.
- [5] L. Perraud, et al., "A Direct-Conversion CMOS Transceiver for the 802.11a/b/g WLAN Standard Utilizing a Cartesian Feedback Transmitter," IEEE J. Solid-State Circuits, Vol. 39, pp. 2226-2238, Dec. 2004.
- [6] P. Zhang, et. Al., "A Direct Conversion CMOS Transceiver for IEEE 802.11a WLANs," ISSCC Technical Digest, pp. 354-355, Feb. 2003.

Low Power Bluetooth Single-Chip Design

Marc Borremans, Paul Goetschalckx

Marc.Borremans@st.com || Paul.Goetschalckx@st.com
STMicroelectronics Belgium nv,
Excelsiorlaan 44-46, B-1930 Zaventem, Belgium.

Abstract

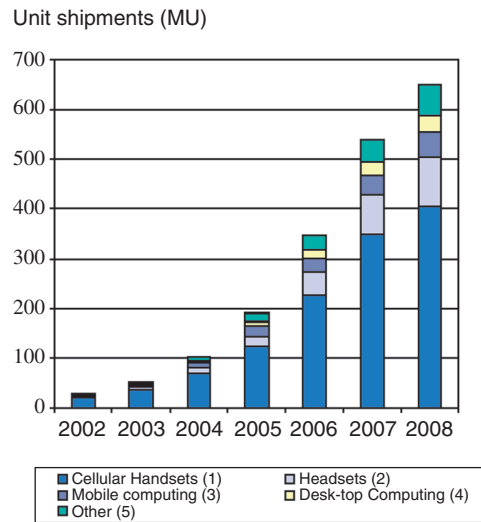
This paper describes the implementation of a second generation Bluetooth single chip in 0.13 μ m technology. The considerations in the concept phase on technology and topology level are highlighted. The main targeted market segment is the cellular applications, urging for very low power consumption in all operation modes. The presented chip presents a very competitive consumption in active operation and also an excellent power-down current consumption. The physical layer implementation of the transmitter part of the chip is presented as a case study of the active power reduction.

1. Introduction

The first generation of Bluetooth solutions brought the wireless RF connectivity function of the Bluetooth standard v1.1. The digital base-band processor and the RF transceiver were often split in a two-chip solution.

The second generation of Bluetooth systems focuses on low cost, single-chip solutions. On top, emphasis came on improvements in RF performance, intelligent use of the spectrum, and user experience improvements brought by the Bluetooth v1.2 standard.

Figure 1 gives an overview of the market expectations for Bluetooth, by application segment, until 2008. Cellular phones embedding the Bluetooth functionality represent by far the largest market segment for Bluetooth devices. Bluetooth connectivity is indeed available in nearly all new cellular phones. Concurrently, wireless headsets are today a very important application. The presented Bluetooth component is specifically designed for cellular handsets .



*Fig.1: Total and application specific Bluetooth market expectations.
(Source IMS).*

The cellular phone application brings a number of specific requirements related to the hardware :

- The low power consumption, to save battery life-time, is ultra important.
- A low cost Bill Of Material (BOM) associated to the component
- A small size “footprint” and “height” of the complete Bluetooth solution on a Printed Circuit Board (PCB).
- A very low Bit Error Rate for maintaining good voice quality.
- The co-existence with the cellular radio and other RF applications embedded in the cellular phone.

The design of the presented Bluetooth component, differentiates by its very low power consumption in all operating modes, offering good RF performance and respecting the other requirements listed.

Low power consumption is important, in the operational modes and in the low power modes as defined by the Bluetooth standard. Also in complete power down mode, when the component is not functioning at all, low leakage current (or low permanent bias current) is essential for the battery life time.

This first section briefly described the target application and market for the Bluetooth component and the circuits developed. Consequently the main specific requirements are set.

The second section covers the technology choice and the system partitioning of the full Bluetooth component during the concept phase. The current consumption in low power modes and in power-down mode will be heavily impacted by these decisions. In operational mode, the current consumption is mainly determined by the radio part of the circuit.

It will be shown in the third section how the implementation of this -mainly analog- part of the chip has been optimized towards low power consumption in functional mode. The transmitter case study is presented in detail, illustrating some important considerations targeting low cost and low current consumption. Finally, a plot of the chip and some measurement results are shown.

2. The concept phase

The Bluetooth component presented integrates the digital base-band processor and memories, plus the RF transceiver in a 0.13 μm CMOS process.

A number of technical choices have to be made in the definition phase of the component in order to match the cost, the technical and the time-to-market requirements.

The choice for silicon technology interacts with the choice for architecture and circuit implementation.

2.1. IP and technology selection

The transceiver architecture was inherited from a stand-alone Bluetooth radio product in a CMOS 0.25 μm technology [1].

The diagram below (Fig. 2) depicts the criteria for the silicon technology choice and the technology options. The Bluetooth component is implemented in ST's CMOS 0.13 μm low power process with a dual gate oxide for 2V5 transistors.

Circuits running at RF frequency that get real benefit from low geometry transistor parameters and lower parasitics are designed with 0.13 μm transistors. This concerns mainly the circuits connected to the antenna like LNA and PPA and high frequency circuits in the PLL, like I-Q-generation, the LO buffers and the prescaler. Of course, also the digital parts like the base-band processor, the demodulation and modulation logic and the delta-sigma modulator in the PLL are implemented in 0.13 μm transistors.

Circuits that benefit from a higher supply voltage, mainly the IF part of the transceiver because of the dynamic range, are designed in the 2V5 dual gate oxide transistors.

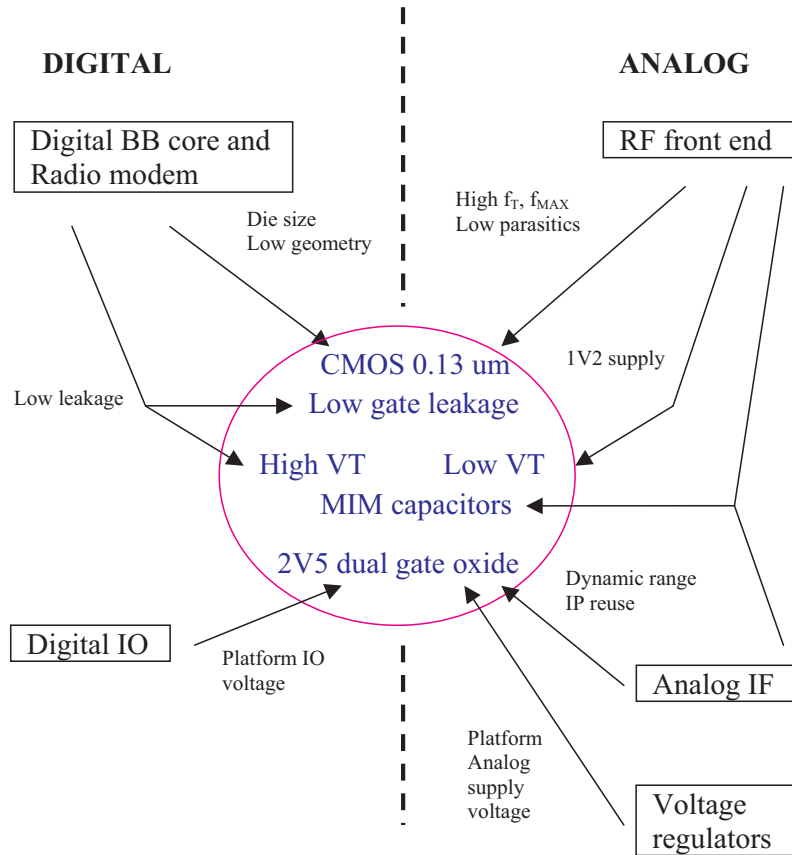


Fig.2 : Criteria impacting the technology and technology options selection.

The 2V5 dual gate transistor option is also used for a number of LDO on-chip supply regulators that serve two goals :

- 1) provide a clean supply to the load by establishing a power supply rejection towards the external supply and separate the supply of different loads to avoid cross-talk.
- 2) provide the means to completely switch off the supply of a circuit not in use, to eliminate the leakage current.

All voltage regulators, including the one that powers the base-band logic, are controlled by a power management circuit. The latter is implemented in 2V5 dual gate oxide logic, which exhibits a minimum remaining leakage current.

2.2. Top level implementation highlights: crosstalk

Managing cross-talk between different circuit building blocks is one of the main challenges in developing a fully integrated transceiver and even more in developing a single-chip base-band plus transceiver component.

As discussed above, crosstalk via the power supplies is counter-acted by a deliberated grouping of circuits to specific voltage regulators. On chip power supply and ground routing has been studied in detail and star connections have been applied where appropriate. MIM capacitances running along the power supply tracks offer extra supply decoupling at high frequency.

A lot of attention has been paid to avoid cross-talk via the ESD protection structures and the tracks routed for ESD protection. As ESD standards require pin to pin protection, even digital IO's have connection paths to the analog IO's via the ESD protection implementation. The voltage regulators and analog circuits are carefully grouped and separated from each other on distinct IO supply rails. Back-to-back diodes are put in series between the rails of critical analog parts and a common ESD protection track.

Of course, switching noise originated from digital parts needs a lot of care and attention. Possible sources are mainly the large digital base-band processor circuit and the digital circuits in the fractional-N PLL. The logic has been implemented by "splitted supply" standard cells and IO's, allowing separate power connections for the switching currents and for the substrate and well connections. This is particularly important in case of wire bonding, the most recent version of the chip presented uses flip-chip bumping. In order to maintain latch-up immunity at the package pin level, grounds and supplies are connected in the substrate of the BGA substrate in case of wire bonding.

The main digital base-band processor part is separated from the transceiver part by an on-chip isolation wall in the silicon substrate. Also "aggressive" and "sensitive" transceiver parts are separated from each other by these isolation structures.

3. The Radio Architecture

Fig.3 shows the top-level topology of the RF transceiver integrated in the STLC2500 single chip Bluetooth.

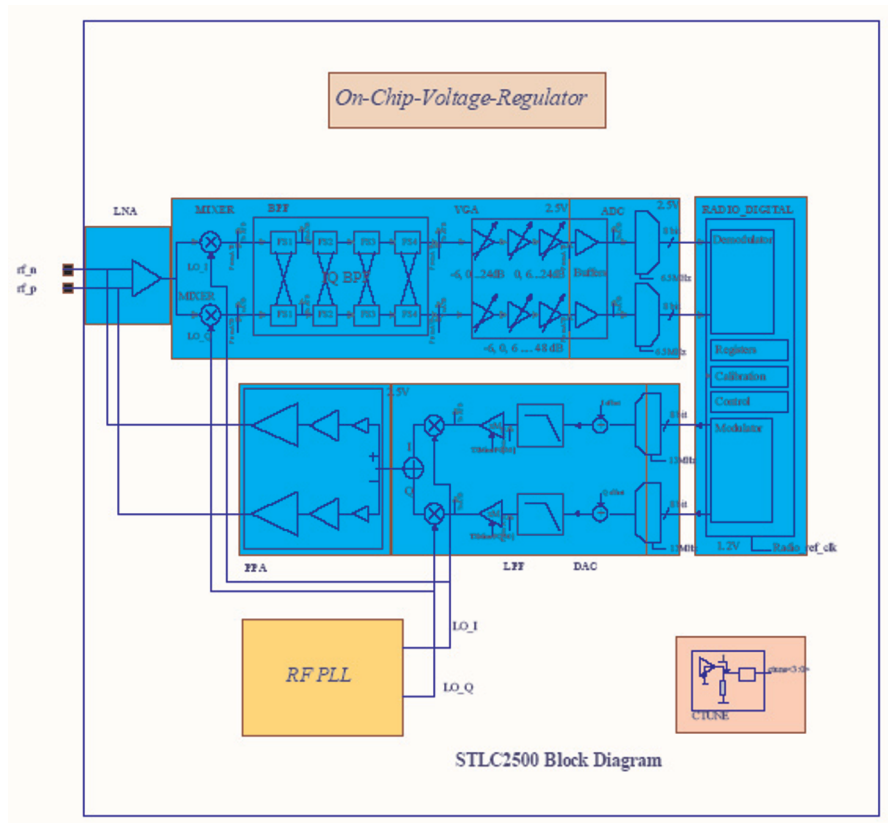


Fig.3 : Topology of the Bluetooth transceiver.

The receiver implements a low-IF architecture for Bluetooth modulated input signals. The mixers are driven by two quadrature signals which are locally generated from a VCO signal running at twice the channel frequency. The output signals in the I signal path and Q signal path are bandpass filtered by an active poly-phase bandpass filter for channel filtering and image rejection. This filter is automatically calibrated to compensate for process variations.

The output of the bandpass filter is amplified by a VGA to the optimal input range for the A/D converters. Further filtering is done in the digital domain.

The digital part demodulates the GFSK coded bit stream by evaluating the phase information in the I and Q signals. The digital part recovers the receive bit clock. It also extracts RSSI data by calculating the signal strength and uses this information to control the overall gain amplification in the receive path.

The transmitter takes the serial input transmit data from the base-band processor. This data is GFSK modulated to I and Q signals. It is converted to analog signals with 8-bit D/A converters. These analog signals are low-pass filtered. Here again, an automatic calibration is integrated. The signal is then applied to the direct up-conversion mixers, which uses the same VCO as the receiver. At the end of the transmit chain, a multi-stage class AB output amplifier provides the final amplification to the RF signal.

The on-chip VCO is fully integrated, including the tank resonator circuitry. It is the heart of a completely integrated fractional-N PLL. The oscillator frequency for the various Bluetooth channels is programmed by the digital radio control section. An auto-calibration algorithm centres the PLL frequency despite all possible process variations.

The testability of the chip is enhanced by an analog testbus that allows to observe and drive the main points in the signal paths, for characterisation purposes and for final ATE test in the production line.

4. Case Study : The Transmitter

The next sections provide more detail on the design trade-offs used in the transmitter part of the chip.

4.1. Targets

The constant envelope frequency modulation of the Bluetooth system allows several topologies to be used for the transmitter part of the chip.

The most widely spread topologies are the IQ up-conversion of a low frequency signal to the channel frequency and also the direct modulation of the oscillator. The IQ up-conversion topologies can be segmented based on IF frequency and on the PLL topology.

The directly frequency modulated oscillator control can be analog or digital [5]. In order to take optimal benefit from the experience gained by a first generation Bluetooth radio, predecessor of the presented chip, and in order to have a fast time to market, the existing direct IQ up-conversion transmitter topology has been used as a starting point [1].

An additional major advantage of this topology is the easy upgrading to the next generation Bluetooth v2.0 standard, called “Enhanced Data Rate”, which uses non-constant envelope modulation. In the direct oscillator modulation topology

this requires amplitude modulation in the power amplifier, increasing the complexity and the risk significantly.

The new TX design focussed on an aggressive reduction of the current consumption in the chip, while assuring an excellent behaviour on the signal quality and the RF performance.

The Bluetooth specification defines TX output power within -6 to +4 dBm as class II, and TX output power up to 20 dBm maximum as class I application. The terminology “class 1.5” is used below for a TX output power which is above class II ratings but considerably below the maximum of class I.

	Starting point[1]	Result classII	Result class1.5
Technology	0.25 um	0.13/0.28 um	0.13/0.28 um
Pout [dBm]	0dBm	3dBm	7dBm
Pout [mW]	1mW	2mW	5mW
Current PA	34-44 mA	8.5 mA	11.7 mA
Current IQ-Mixer	7.2 mA	1.2 mA	2 mA
Current LPF	0.7 mA	0.24 mA	0.24 mA
Current DAC	0.6 mA	0.37 mA	0.37 mA
TOTAL Current	42.5-52.5 mA	10.3 mA	14.3 mA

Table 1 : Transmitter design results compared to the starting point.

Table 1 gives a quantitative representation of the achieved power and current consumption targets for the different sub-circuits in the transmitter part compared to the starting point.

The table illustrates the main challenge in the transmitter design : to reduce extremely the power consumption in the integrated power amplifier. At the same time the delivered output power even had to be doubled (+3dB) compared to our original solution. The target output power was set close to the maximum limit of the Bluetooth class II specifications in order to guarantee a maximal range.

4.2. A Class-1.5 prototype

In order to increase the transmission range, the power can be increased even above the +4dBm class II upper limit. In this case the device becomes a device in the Bluetooth class I category which ranges up to 20dBm. As a power control mechanism is mandatory for this operating range, a flexible and very fine resolution power control mechanism has been implemented.

For cellular handset applications, it is interesting to extend the transmission range by increasing the power with a few dB, without the need to add an external PA consuming a lot of current. To accommodate the class 1.5, a boost of the output power has been implemented. The increase in output power is

partly done by increasing the gain in the low frequency part of the TX signal path and partly by dynamically adapting the gain of the power amplifier. The combination of these two mechanisms results in 4 dB more (=250%) power compared to the class II mode.

The designed 7dBm (=5mW) output power is confirmed by the first measurement results. The additional current consumption is approximately 1 mA per dB. So, for achieving the 7dBm power level only 4mA additional current is required. This is also indicated in Table 1. This excess current can be divided into 0.8mA in the mixer and 3.2mA for the dynamic current increase in the PA. It is important to note that this additional power capability has been implemented without any additional cost (neither area nor current) at nominal class-II operating mode, as no extra stages have been used to increase the power.

4.3. Power Supply Constraints

The increased speed of the 0.13 μ m CMOS transistors has a beneficial impact on the power consumption of the RF circuits. On the other hand, the low supply voltage limits the choice of circuit topologies. The tolerance of the on-chip regulators further limits the supply voltage range. As a result, circuits implemented with 0.13 μ m transistors need to perform down to a 1.2 V minimum supply voltage.

4.4. The power amplifier

The GFSK modulation of the Bluetooth data allows the use of more power efficient topologies compared to the linear class A topology. A multi-stage class AB topology was selected in order to increase the efficiency of the amplifier. Theoretically more power efficient topologies requiring coils were not selected in order to avoid the large area consumption of integrated inductors and to avoid the possible EM coupling of the RF signal and its harmonics from the amplifier's coils into the integrated inductor of the VCO.

The first stage of the power amplifier delivers the voltage amplification of the signal, while the final stage drives the required load impedance.

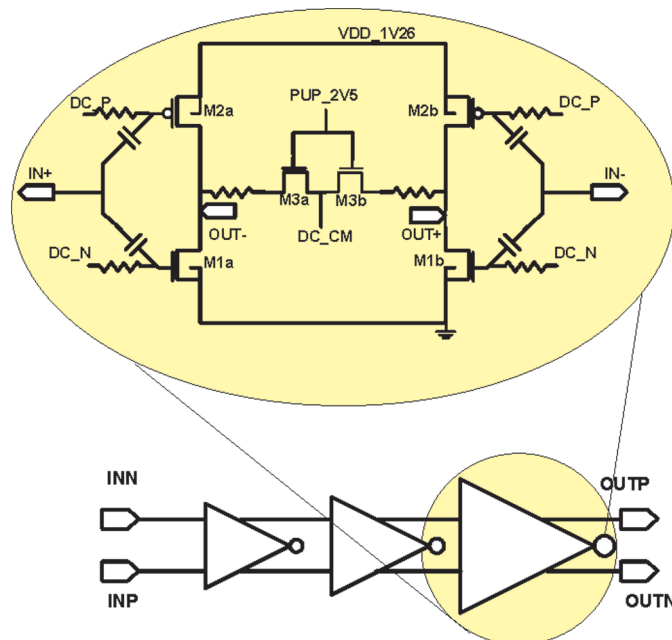


Fig.4.: Simplified schematic of the power amplifier

The best power efficient combination of I-Q up-conversion mixer and PA resulted in the final topology by which additional gain is provided by a third stage in the PA. This allows for a low current consumption I-Q mixer. The resulting total current consumption of the 3-stage PA including all bias circuitry is only 8.5 mA at +3 dBm output power. This number contains a significant part of dynamic current. The dynamic current and the mixer current decrease with a few mA for low output power settings via the power control mechanism.

A simplified schematic of the PA is given in Fig.4. It shows the three-stage amplifier. The transistor detail of the last stage is included. The RF input signal is applied at the IN+ and IN- terminals. It is ac-coupled using MIM capacitors to the amplifying transistors M1 and M2. Those transistors are operating in class-AB mode. They are respectively biased by nodes DC_N and DC_B, that themselves are determined by a bias current in a diode connected transistor. The bias currents provide on-chip automatic compensation for temperature and process variations. In this way, the variation of the gain over corners and temperature is very limited.

The DC operating point is determined by the node DC_CM which is resistively connected to the drains of the amplifying transistors. The transistors M3a and M3b guarantee the isolation between the positive and negative output during a receive slot. This is necessary as the RF output terminals are shared between the

LNA and the PA. The RX-TX switch is integrated on-chip by carefully powering up and down the respective sub-circuits.

4.5. The DAC, the LPF, the programmable gain amplifier and the mixer

The 8-bit signal DACs, the low pass filter and the programmable gain amplifier have been implemented in 2.5V dual gate oxide transistors. The high voltage headroom for these devices allowed to stack the different devices between the power rails. In this way, the current consumption is minimised.

The IQ mixer, being the interface element between the low frequency and the RF part of the signal path, was also implemented in the 2V5 dual gate oxide transistors optimally using the higher voltage range.

The I-Q up-conversion mixer

The mixer topology is based on a Gilbert cell mixer topology with some modifications resulting in superior in-band linearity [2]. Third order distortion components at LO3BB are smaller than -50dBc . The required LO signal amplitude is less than 300mV amplitude for any technology corner or temperature. This small LO amplitude allows for lower power consumption in the LO-buffers of the PLL.

A current-mode topology

The current consumption has been optimized by using a flattened approach for the tx-signal path. The building blocks are not longer considered as individual blocks with buffering interfaces. Traditionally, this interfacing constitutes a significant amount in the total current consumption in order to provide sufficient linearity.

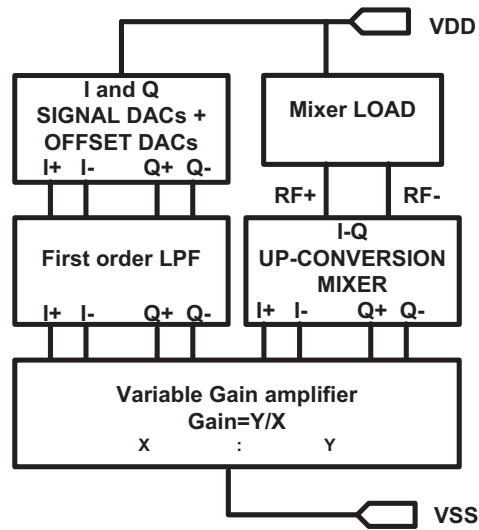


Fig.5: Schematic overview of the transmitter analog section without PA.

The picture (Fig.5) shows the implementation of the DAC, the low-pass filter, the variable gain amplifier and the mixer. It clearly shows how the current consumption for interfacing has been avoided: the current from the DAC flows through the filter into the variable gain amplifier. The amplified current directly drives the switches of the mixer. There are only two DC-current branches. One for the DAC and LPF and one for the mixer. The relation between these two is given by the programmable current mirror. A very small DC/ac current ratio can be used because the signal stays in current-mode. The whole system functions well over all corners and ranges with a DC/ac ratio of only 1.15.

The D/A converters

The I and Q current-steering D/A converters have a 4-4 segmented architecture. An intrinsic 8 bit accuracy is guaranteed by sizing the pMOS reference transistors using the following formulas for the current accuracy.

$$\frac{\sigma(I)}{I} \leq \frac{1}{2.C.\sqrt{2^N}} \quad \text{with} \quad C = \text{norm}^{-1}\left(0.5 + \frac{\text{yield}}{2}\right)$$

with :

- $\frac{\sigma(I)}{I}$: the relative standard deviation of a unit current source
- N :the resolution of the converter
- Norm⁻¹ : the inverse cumulative normal distribution, integrated between -x and x.
- Yield: the relative number of converters with an INL < 0.5LSB.

Based on this formula and the size versus matching relation [4], the dimensions of the current source transistors is determined:

$$W = \sqrt{\frac{1}{KP\left(\frac{\sigma(I)}{I}\right)^2} \cdot \left[\frac{A_\beta^2}{(V_{GS} - V_T)^2} + \frac{4A_{VT}^2}{(V_{GS} - V_T)^4} \right]}$$

$$L = \sqrt{\frac{KP}{4I\left(\frac{\sigma(I)}{I}\right)^2} \cdot [A_\beta^2 (V_{GS} - V_T)^2 + 4A_{VT}^2]}$$

A lot of care has been spent during design and layout in order to avoid distance and side effects by using common centroid structures for the I and Q DAC. Dummy rows and columns provide full symmetry. All interconnections have also been placed systematically and symmetrically over the current source matrix.

This approach results in the targeted intrinsic accuracy. The INL and DNL measurement results are shown in Figure 6 to illustrate this statement. This figure shows the data collection on 646 devices for INL and DNL results on the signal DACs. The parameter m represents the mean of the distribution, s is the sigma of the normal distribution. The pictures shows that the targeted 8-bit accuracy is even achieved for more than 6 sigma.

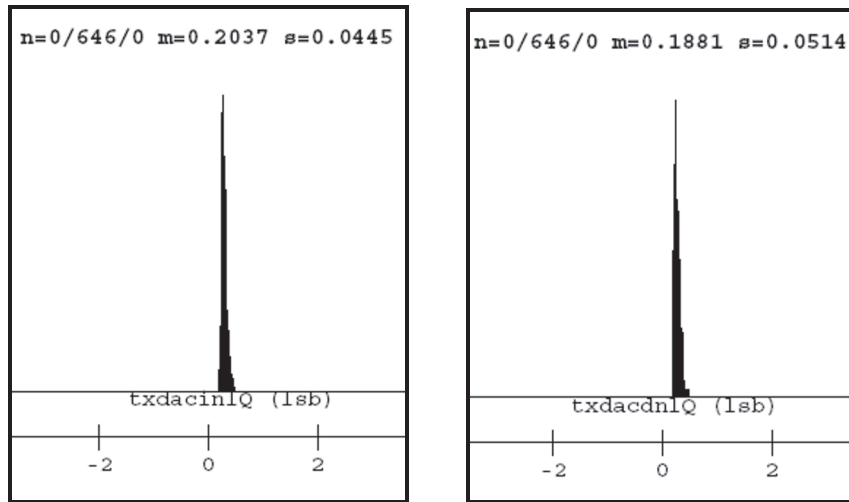


Fig.6.: INL and DNL measurement results.

The dynamic behaviour has been guaranteed using extreme care in the synchronisation of the control bits and by using limited swing for the control of the switches. Also the I and Q offset DACs (see section 4.7) have been integrated in the signal DAC structure.

The Low Pass Filter

Thanks to the very good Spurious Free Dynamic range at the output of the DAC, the filter mainly needs to suppress the clock-alias components on the I and Q analog signal currents. A first order filter is sufficient to obtain more than 10 dB margin on the adjacent channel power specification. This first order filter is implemented using the $1/g_m$ impedance of a cascode transistor and an additional capacitance.

The variable gain amplifier

The variable gain amplifier is implemented as a programmable current mirror. The design of this building block is mainly constrained by the area consumption, the offset between the I and Q part of the signal and the harmonic distortion introduced by this circuit.

The next sections go into details on some aspects of the design optimisation of the circuits.

4.6. Linearity

The current generated by the DAC is highly linear, with a 3rd harmonic suppression over 60dBc.

The LPF causes some linearity degradation due to the time variant operating point of the cascode transistor. Extra current sources were introduced to minimize this effect and to obtain a linearity around 50dBc. An additional current of 80 μ A per signal branch is injected so that the total DC current is 128 μ A, leading to a DC/AC ratio of 3.2 in the cascode. In the consumption table this 320 μ A surplus current is noted as the LPF current. The pMOS added current is eliminated by nMOS DC current sources avoiding the current to be amplified to the mixer. A lot of attention has been paid on the matching of these nMOS and pMOS currents.

The programmable current amplifier causes some degradation of the linearity, mainly due to the parasitic capacitance in the current mirror which are a consequence of the large size of the transistors. Note that the Early effect is non-dominant thanks to the long device lengths used.

The third order baseband harmonic is directly up-converted in the mixer to LO3BB. An other important contributor to this distortion component is the intrinsic non-linearity of the up-conversion mixer. The mixer was designed to reach at least 50dBc worst case suppression of this 3rd harmonic. Thanks to some modifications [2] on the classical mixer, this very good linearity can be achieved. This is confirmed by measurements.

Taking everything together, the LO3BB linearity is still better than 40dBc worst case over all ranges and all technology corners.

The mixer is fully functional at a DC/AC ratio of only 1.15 for the input current. In order to guarantee this ratio in all conditions, a typical value for this ratio of 1.20 is fixed at the DAC output. At the LPF output, 1.18 is guaranteed. The delta is caused by the difference between the nMOS and pMOS added currents through the cascode.

4.7. Carrier suppression : calibration versus intrinsic offset accuracy.

The spurious component at carrier frequency is mainly determined by the DC offset current at the I and at the Q low frequency inputs of the mixer.

In order to cancel this offset, I and Q differential offset DACs with a resolution equal to $\frac{1}{2}$ LSB of the signal DAC have been implemented. These offset DACs

are integrated in the same layout structure as the signal DACs. In this way, good accuracy and correlation with the signal currents is assured as well as a very small additional area consumption.

However, sample by sample calibration is expensive. Especially for products intended for mass production, the elimination of a calibration cost has an important impact on the cost of the component. Each building block has been designed in order to avoid this calibration. In the mixer, the main contributor to DC offset and therefore to carrier feed-through is the current amplifier. The main contributor to the offset in this current mirror is the VT matching of the transistors.

Calculations show that the carrier suppression is only dependent on technology parameters, on the square root of the current and on the length of the transistors:

$$\frac{I_{ac}}{\sigma_{IDC}} \sim \sqrt{I_{DC}} * L$$

It is interesting to note that the current offset and therefore the carrier suppression are only determined by the L of the MOS transistor once the current has been fixed. The length is a tradeoff between linearity, transistor size, power consumption and carrier rejection.

L=3 μ m brings to a carrier rejection of 38dBc (4 σ). These values would not require any trimming.

Mismatch in the switch transistors, the PLL phase and amplitude, the resistor and capacitance inaccuracy and the Early effect are second order contributors.

$$\sigma_{IDC} = \sqrt{\sigma^2_{I_MIRROR} + \sigma^2_{I_FILTER_N} + \sigma^2_{I_FILTER_P} + \sigma^2_{I_DAC}}$$

The total offset determining the carrier component is also determined by the inaccuracy on the additional current sources for the filter linearity and a non-dominant contribution by the DAC. In the design, a similar part of the carrier budget was attributed to the LPF+DAC as to the mirror in the variable gain amplifier. In this way, a carrier rejection of at least 35dBc is achieved.

5. Chip photograph

Figure 7 shows a picture of the STLC2500 chip. The transceiver part and the digital part can clearly be distinguished. Also the isolation structure between the digital and analog part is clearly visible.

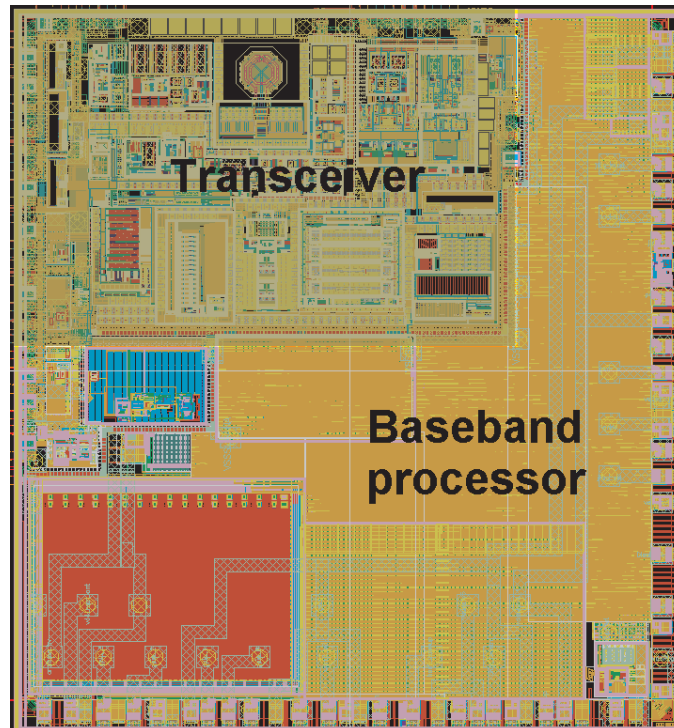


Fig.7.: Plot of the presented STLC2500 BT single chip

6. Measurements Results

The presented chip has been extensively characterized over process corners and temperature on bench set-ups and by test equipment in the production line. The STLC 2500 has been fully Bluetooth qualified over the -40 to $+85$ degC temperature range.

As an illustration, a few selected measurements of the chip when operating in class 1.5 mode are shown.

Figure 8 shows the measured output power at the device's pin when operating in class-1.5 mode. $+7$ dBm output power is reached at room temperature. The variation over the extended temperature range -40 to $+100$ degC is less than ± 1 dB.

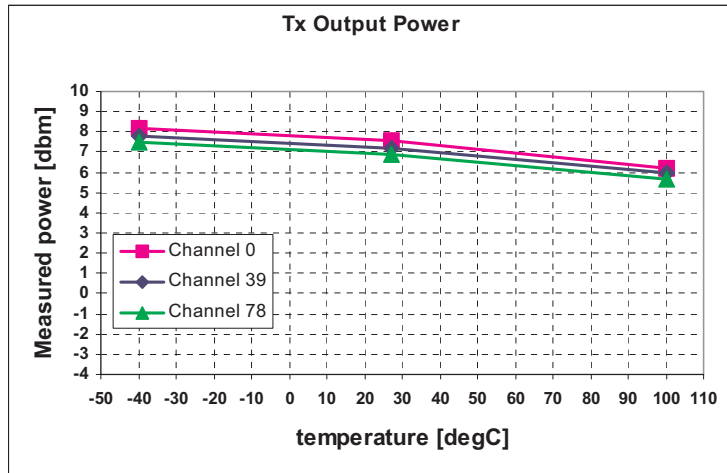


Fig.8: Measured output power for the “class-1.5” mode of the presented STLC2500 BT single chip

Figure 9 shows the adjacent channel power in the same operating mode. The graphs show the clean spectrum with a large margin to the Bluetooth specification (indicated in red). The margin is maintained over all temperature extremes.

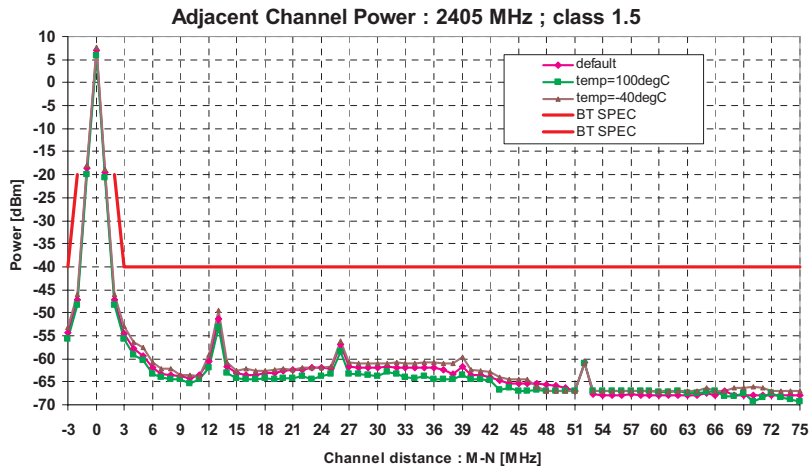


Fig.9: Measured Adjacent channel power for the “class-1.5” mode of the presented STLC2500 BT single chip

7. Conclusions

The implementation of a second generation Bluetooth single chip in 0.13 μ m technology has been presented. The presented chip has an excellent and very competitive power consumption making it the preferred solution for cellular applications. This low power consumption is achieved by specific technology selection, by including an integrated power management control unit and by analog circuit topologies and implementation techniques focused on low power consumption of the analog transceiver. The implementation of the transmitter part of the chip is used as a case study. It shows how a significant power consumption reduction has been reached. Design trade-offs on gain, linearity, area consumption, carrier feedthrough and accuracy have been covered. The presentation is illustrated with some measurement data on the presented chip and first measurement results on the “class1.5” high power version are shown.

References

- [1] J. Craninckx, “A Fully Integrated Single-chip Bluetooth Transceiver”, Proceedings of the AACD 2002 Workshop, April 2002.
- [2] Marc Borremans, “UP CONVERTER MIXER LINEARIZATION IMPROVEMENT”, European patent application 03076146.3 // US patent application 10/826,874
- [3] Anne Van den Bosch et al., “A 10-bit 1GS/s Nyquist Current steering D/A Converter”, IEEE JSSC, Vol.36, no. 3, pp. 315-324, March 2001.
- [4] M. Pelgrom et al., “Matching Properties of MOS transistors”, IEEE JSSC, vol.24, no.5, pp. 1433-1439, Oct. 1989
- [5] R. Staszewsky et al, “All-Digital TX Frequency Synthesizer and Discrete-Time Receiver for Bluetooth Radio in 130-nm CMOS”, IEEE JSSC, VOL. 39, NO.12, Dec. 2004.

RF DAC's: output impedance and distortion

Jurgen Deveugele, Michiel Steyaert
Kasteelpark Arenberg 10
3001 Leuven
Katholieke Universiteit Leuven: ESAT-MICAS

Abstract

The output impedance of a current-steering DAC is setting a lower limit for the second-order distortion [1]. At low frequencies it is not much of a factor. The output resistance can be quite high. At higher frequencies the capacitances gravely reduce the output impedance. As the distortion is mainly second order, some propose to use differential outputs for high frequency signals [2]. Unfortunately, this analysis is not complete. Likewise effects cause severe third-order distortion. Before envisaging RF DAC's this problem must be identified. In this paper we study the problem in depth and give design guidelines.

1. Introduction

Current-steering DAC's are the high-speed DAC's of choice for the moment [3, 4, 5, 6]. They have two main advantages over other structures. First, they do not require high-speed opamps with good linearity at those speeds. Second, they do not require any nodes with large capacitors to be charged or discharged at high speeds. The current-steering architecture thus seems to be a good candidate for a RF DAC.

As discussed in [1] the reduction of the output impedance at high frequencies does introduce distortion. It was believed that this problem could be bypassed by using differential structures [2], but our detailed analysis shows that a large portion of the distortion is third order. For single-tone generation, the third-order distortion can be outside the band of interest. But the output signal of a generic RF DAC has a more complex spectrum. The third-order distortion is then in-band, making filtering the distortion out of the signal not an option.

2. Calculating the distortion

In order to understand the gravity of the problem we do calculate the distortion. We however take a different approach than the one used in [1]. We do not

calculate how big the output impedance must be. We do stress the importance of the troublesome part of the impedance: the capacitive part. At low frequencies the output resistance can be designed to be high enough. The problem starts if the non-linear current that charges and discharges the parasitic capacitances starts to grow relatively big compared to the load current. We identified two mechanisms that generate these non-linear currents. They are described in the two sections below. For ease of calculations we assumed that the DAC's are unary-decoded.

2.1. The maximum output impedance

The output resistance can be made sufficiently high. The output impedance however has a capacitive part. We consider a typical current source with cascode and switches sub-circuit as it is drawn in figure 1. The two parasitic capacitances of importance, C_0 and C_1 are added to the circuit. We desire to calculate the output impedance of the black part in figure 1. The small-signal schematic to calculate the output impedance is shown in figure 2. We are however interested in the upper limit of this impedance rather than in the exact impedance of this circuit. One upper limit is set by capacitance C_1 and the gain of the switch transistor M_{sw} in the on region. Let us consider the circuit drawn in figure 3(a). The switch transistor cascodes the impedance Z_{eq} . Z_{eq} is chosen to be the small-signal impedance formed by the current-source transistor M_{cs} and the cascode transistor M_{cas} . By doing so the small-signal circuit shown in figure 3(b) is equivalent to the one in figure 2. The output impedance Z_{out} is calculated as

$$Z_{out} = r_{osw} + A_{sw}Z_{eq} \quad (1)$$

An upper-limit for the output impedance can be found assuming that the output resistance of the current source and the cascode are infinite. The output resistance can not be infinite, but it can be very high by using gain boosting [7] cascode transistor M_{cas} for example. The output impedance then is

$$Z_{out} = r_{osw} + A_{sw} \frac{1}{j2\pi f C_1} \quad (2)$$

An upper limit for the output impedance is thus set by no more than the gain A_{sw} of the switch transistor and the capacitance C_1 as seen in figure 1. This upper limit is independent of any gain boosting of the current source or cascode transistor. Gain boosting [7] of the switch itself can raise this maximum, but gain boosting the switch is not an easy task as the input signal is digital.

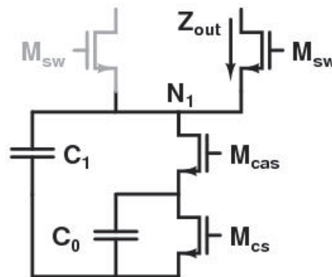


Fig. 1. Typical current source, cascode and switches sub-circuit.

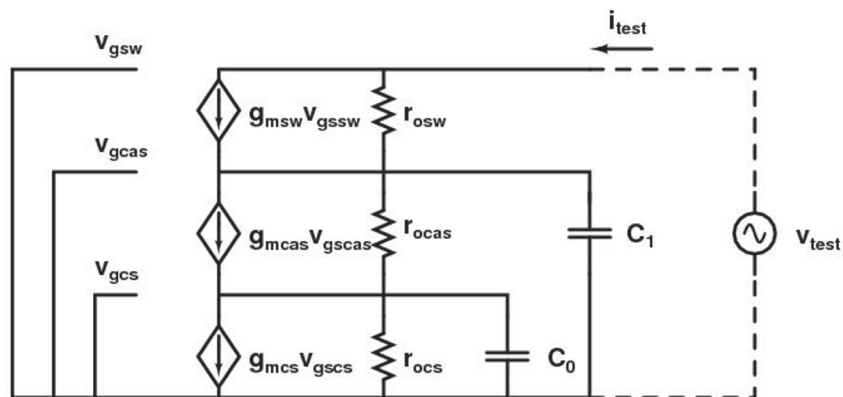


Fig. 2. Small-signal circuit of the black part of the schematic seen in figure 1.

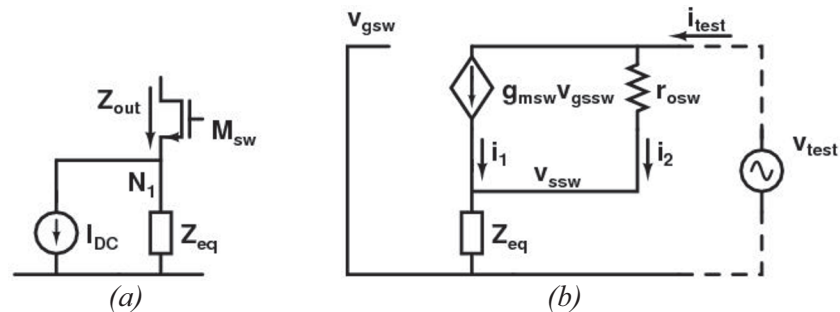


Fig. 3. (a) Schematic for determining the upper-limit of the impedance. (b) Small-signal circuit for determining this upper-limit.

2.2. The non-linear capacitive load

We have seen in section 2.1 that the gain of the switch transistor and the capacitance on the node at the source of the switch are setting an upper-limit for the output impedance. We decided to calculate the distortion caused by the non-linear capacitive load rather than to use the output impedance as in [1].

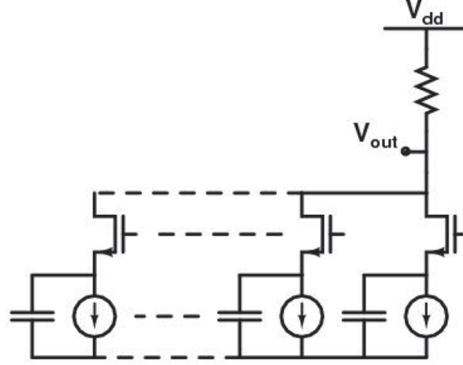


Fig. 4. Single-ended branch of the simplified converter.

In figure 4 we have a single-ended branch of our simplified converter. We see that the number of capacitors switched to the output node depends on the output code. For input code x , x capacitors are coupled to the output. The other $2^N - 1 - x$ capacitors are then connected to the differential output. N is the number of bits. In order to ease the analysis we assume that the converter has an infinite resolution and sampling rate. This allows us to use differential calculus.

The additional charge δQ that flows to the connected capacitors during the time interval δt is proportional to the connected capacitance and to the voltage change of the capacitive nodes:

$$\frac{\delta Q}{\delta t} = C_{connected} \frac{1}{A_{sw}} \frac{\delta V_{out}}{\delta t} \quad (3)$$

The change of charge over time equals to the instantaneous (non-linear) current $I_{nl,1}$:

$$I_{nl,1} = -C_{total} 2\pi f \left(\frac{\cos(2\pi ft)}{2} + \frac{\sin(4\pi ft)}{4} \right) \frac{1}{A_{sw}} V_p \quad (4)$$

C_{total} is the sum of the capacitances C_1 connected to node N_1 over all the unary-weighted blocks. V_p equals to the output voltage amplitude. Equation (4) has two frequency components. The first one is the cosine part and has the same fundamental frequency as the signal, but a different phase. This part slightly modulates the amplitude of the signal. The second part describes the second harmonic. We can calculate the SFDR using the sine term component of equation (4):

$$SFDR = 20 \log \left(\frac{\overline{I_p}}{I_{nl,1}} \right) = 20 \log \left(\frac{2A_{sw}}{\pi f R_L C_{total}} \right) \quad (5)$$

expressed in dB. R_L is the load resistor. This agrees with the findings in [1]. The form of equation (5) once again emphasizes the importance of the gain of the switch. Compared to [1] we do not find higher-order non-linear terms in our results. This difference is caused by the fact that in [1] the feedback takes the distortion caused by the distortion spurious into account. Reviewing the findings in [1] one can see that, for well designed converters, the third or higher-order distortion is negligible. This lead in [2] to the conclusion that in a differential converter the cascode can be omitted as the third-order distortion caused by the output impedance is very low.

2.3. Switching of the capacitive load

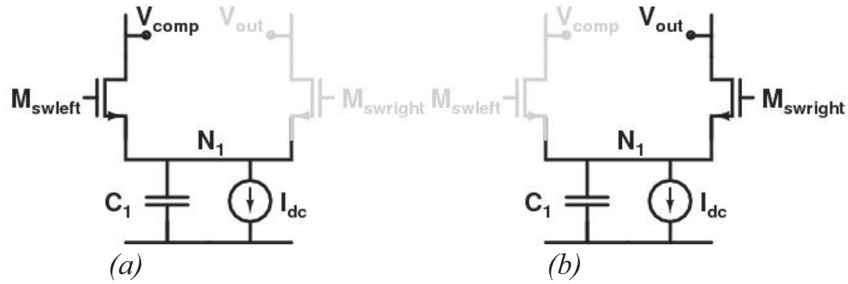


Fig. 5. (a) Switch switched to left output. M_{swleft} cascodes node N_1 from V_{comp} . (b) Switch switched to right output. $M_{swright}$ now cascodes node N_1 from V_{out} .

In this subsection we model the influence of the switching of the capacitance from one node to another node. Let us consider a very small increment of the input code. For a converter with infinite precision this causes some capacitances to be switched from the complementary output to the output. Before switching, as in figure 5(a), the voltage on node N_1 was equal to the voltage on the complementary output V_{comp} divided by the gain of the switch A_{sw} . After switching, as in figure 5(b), the voltage on node N_1 will evolve to the voltage on the output V_{out} divided by the gain of the switch A_{sw} . The charging of this capacitive node causes a current flow trough the right-hand switch. This current is superimposed on the desired current. It is calculated as:

$$\frac{\delta Q}{\delta t} = \frac{V_{out} - V_{comp}}{A_{sw}} \frac{\delta C_{out}}{\delta t} \quad (6)$$

C_{out} is the sum of the capacitances C_1 that are connected to the output node. This charging is proportional to the instantaneous difference between the output voltage and the complementary output voltage, inversely proportional to the gain of the switches and proportional to the amount of extra capacitance switched to the output. The non-linear current $I_{nl,2}$ can be calculated as:

$$I_{nl,2} = -C_{total} 2\pi f \frac{\sin(4\pi ft)}{2} \frac{1}{A_{sw}} V_p \quad (7)$$

The results in equation (7) do resemble a lot to equation (4). There is however one important aspect we did not yet take into account. The calculated current in equation (7) is only correct *if* the input code is rising. Only additional capacitors that are switched to the output do cause current to flow trough the output. Capacitors that are switched from the output to the complementary output do not cause any current flow trough the output. They only cause current to flow trough the complementary output. This is the case if the input code decreases. The current in equation (7) still has to be multiplied by a modified signum function:

$$\text{sgn}(\cos(2\pi ft)) = \begin{cases} 1, & \cos(2\pi ft) > 0 \\ 0, & \cos(2\pi ft) < 0 \end{cases} \quad (8)$$

The correct equation for the non-linear output current is thus:

$$I_{nl,2} = -C_{total} 2\pi f \frac{\sin(4\pi ft)}{2} \text{sgn}(\cos(2\pi ft)) \frac{1}{A_{sw}} V_p \quad (9)$$

We are mainly interested in the frequency domain. Equation (9) has a second-order term with amplitude

$$I_{nl_second-order,2} = -C_{total} 2\pi f \frac{1}{4A_{sw}} V_p \quad (10)$$

and a third-order component with amplitude:

$$I_{nl_third-order,2} = -C_{total} 2\pi f \frac{1}{8A_{sw}} V_p \quad (11)$$

The total second-order distortion is given by summing the results given in equations (7) and (10). Their amplitudes are in phase. The amplitude of the second-order distortion for both effects combined is thus twice as large as the amplitude calculated in [1]. What is a lot worse is that the amplitude of the third-order distortion is not negligible at all. It has half the amplitude of what was originally calculated for the second-order distortion in [1]. The maximum attainable SFDR for a differential DAC is then given by:

$$SFDR = 20 \log \left(\frac{\overline{I_p}}{I_{nl_third-order}} \right) = 20 \log \left(\frac{4A_{sw}}{\pi f R_L C_{total}} \right) \quad (12)$$

Figure 6 shows the plot of the non-linear currents of equations (4) and (9) and their sum below on the picture. The corresponding spectra are depicted in figure 7.

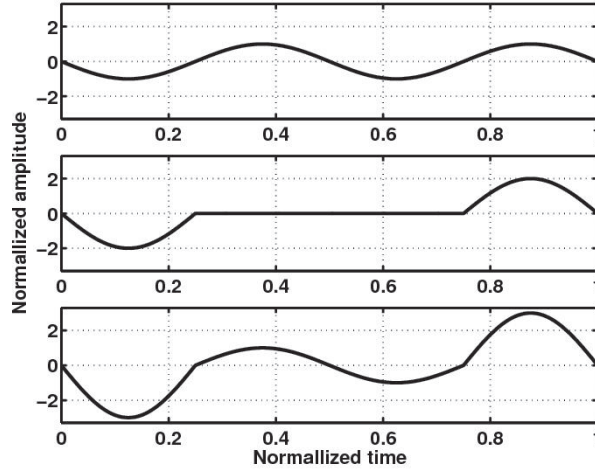


Fig. 6. Plot of the two non-linear currents (above plots) and their sum (bottom plot) in the time domain. The current in the top-most plot depicts the second harmonic component of the non-linear current.

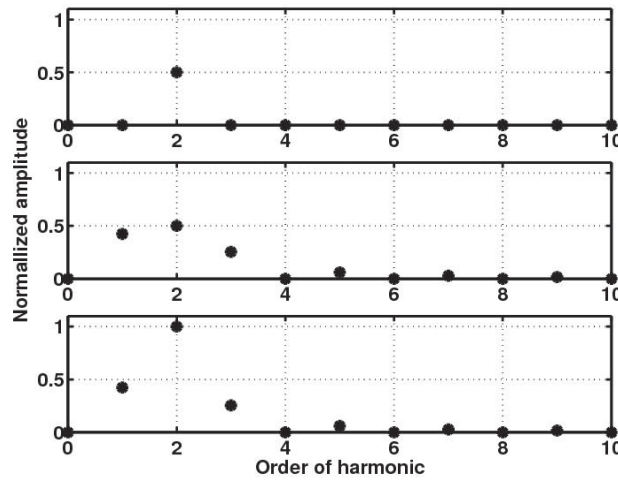


Fig. 7. Plot of the two non-linear currents (above plots) and their sum (bottom plot) in the frequency domain.

3. Design measures to decrease the distortion

In this section we take a look at different options to decrease the distortion caused by capacitance N_1 seen in figure 1. As seen in equation (12) the SFDR in a differential DAC is limited by the third-order distortion and this limit is given by:

$$\frac{4A_{sw}}{\pi f R_L C_{total}}$$

If we want to maximize the SFDR then we have the following options:

- decrease the signal frequency
- decrease the load resistor R_L
- decrease the capacitance C_{total}
- increase the gain of the switch transistor

The first one, decreasing the signal frequency, is not an option. A RF DAC is a DAC with a very high output frequency.

3.1. Decreasing the load resistor

Decreasing the load resistor on the other hand is a viable option. Unfortunately, for an unchanged load current this means that the output voltage and the output power decreases. If we want to keep the output voltage swing then we do need to increase the load current. Doubling the load current for example however doubles the sizes of the switches and the cascades, if the operation points remain unchanged. Transistors with double the size have almost double the parasitic capacitances. Reducing the load resistor hence does only help if it equals to reducing the output voltage.

Although that reducing the output amplitude seems to be a very undesirable thing to do, it is not all bad. RF DAC's can be used in single-hop transmitters or even in software radios. Let's first take a look at a conventional super-heterodyne transmitter [8], depicted in figure 8. The DAC is followed by a low-pass filter to remove the Nyquist images. The Nyquist images are the images of the signal that are generated at

$$f = i f_{clock} + f_{signal} \quad (13)$$

and at

$$f = i f_{clock} - f_{signal} \quad (14)$$

for every integer value of i equal to or greater than 1. This is inherent to all clocked DAC's due to the sampled nature of the converters. The filtered signal is then mixed with an Intermediate Frequency (IF) local oscillator signal. The mixer often has a low conversion gain and a intermediate amplifier must be used. The images, generated during the mixing of the signal, are removed by the

first band-pass filter. The filtered signal is then fed into the second mixer stage, where it is mixed with the RF local oscillator signal. This signal is filtered by the second band-pass filter to remove the images generated by the mixing. Finally, the signal is amplified by the power amplifier. The first stages in the power amplifier restore the signal strength that has been lost in the second mixer stage.

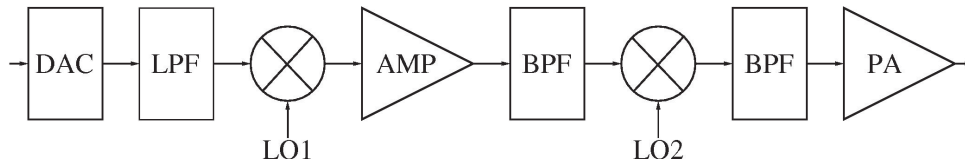


Fig. 8. Block schematics of a super-heterodyne transmitter.

In figure 9 a single-hop transmitter is depicted. Compared to figure 8 the IF mixing stage, the first band-pass filter and the amplifier are removed out of the block diagram of figure 9. This can be done if the DAC generates a high frequency signal. The signal must be modulated in the digital domain. It must be centered around the IF frequency that was used in the super-heterodyne transmitter. The single-hop transmitter thus requires a high-speed DAC.

If we would only remove the first mixing stage and the first band-pass filter in figure 8 then we could allow the DAC to have a smaller output amplitude. If we can remove one mixing stage by using a higher frequency DAC then we do not necessarily need to have a high output amplitude. Of course part of the linearity problems of the DAC are then solved by the amplifier. But its specifications are not harder than they are in the original super-heterodyne transmitter. Still, one mixing stage can be removed. This reduces the distortion and the noise figure. Also, one band-pass filter can be removed. If it was an active filter then the power consumption and the distortion drops. If it was a passive filter, like for example an external SAW filter, then no signal power is lost in the filter. In an optimized design the signal level of the DAC and the amplification can be further tuned.

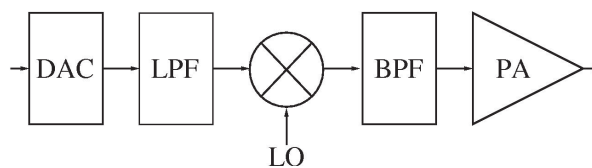


Fig. 9 Block schematics of a single-hop transmitter.

RF DAC's could be used in direct transmitters or in software radios. Compared to the block given in figure 9 the direct transmitter would not need the mixing stage nor the band-pass filter. This increase the signal integrity and the DAC can have a lower output amplitude while still providing the same signal amplitude to

the power amplifier. Again, optimum signal levels can be chosen to increase the overall performance. An additional advantage of lowering the load resistor is that the pole at the output is pushed to higher frequencies.

3.2. Decreasing the capacitance

Another option is decreasing the capacitance. The main contribution comes from the source capacitances of the two switches. A small fraction is contributed by the drain capacitance of the cascode. Optimizing the capacitance guides us to the following design rule: the switches are given a larger $V_{gs}-V_T$ than the cascode as this minimizes the source capacitances of the switches. An additional benefit is that the capacitances C_{gd} of the switches are reduced, reducing feed-through of the switch input signal to the output.

In the layout minimum source structures are chosen for the switches and minimum drain structures are chosen for the cascode, as seen in figure 10(a). Figure 10(b) shows another option to minimize the source capacitance of the switches. The source is now common to the two switches. This layout results in slightly smaller capacitance compared to the layout in figure 10(a) but it is to be avoided. The asymmetry in the switches potentially results in different switching times for the right and left switch, leading to spurious frequencies. Sometimes it is useful to choose non minimum L for the cascode in order to increase the output resistance. Remember that the limit stated in equation (2) is an upper-limit. It might thus be needed to raise the cascode output resistance. Increasing the length of this transistor does increase the drain capacitance, but it does not increase it too much.

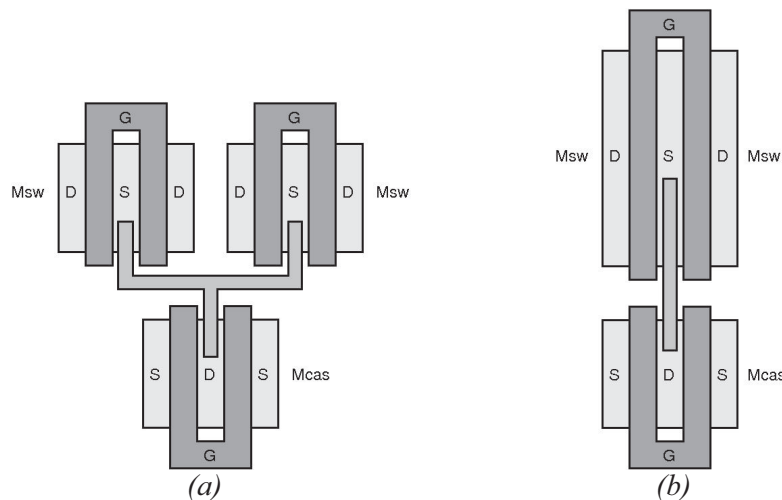


Fig. 10. (a) Layout of the switches and cascodes that optimizes the dynamic impedance by sharing drains and sources. (b) Asymmetric option to minimize the source capacitance of the switches.

$$A = 1 + g_m r_{out} \quad (15)$$

In older technologies, with line-widths well above those used in sub-micron technologies, g_m is given by [9]:

$$g_m = \frac{2I_{DS}}{V_{GS} - V_T} \quad (16)$$

and r_{out} is given by:

$$r_{out} = \frac{1}{g_{ds}} = \frac{V_E L}{I_{DSsat}} \quad (17)$$

with V_E the Early voltage. I_{DSsat} equals to I_{DS} if the cascode transistor is operated in the saturation region. Under those conditions equation (15) can be rewritten using equations (16) and (17):

$$A = 1 + \frac{2V_E L}{V_{GS} - V_T} \quad (18)$$

Equation (18) states that the gain of the switches can be increased by reducing the overdrive voltage $V_{GS} - V_T$. It is maximized if the transistor is operated in weak inversion, where equation (16) no longer holds. The drawback is that the parasitic capacitance increases if the overdrive voltage $V_{GS} - V_T$ is decreased: larger widths are needed if the device must switch the same current. For the strong inversion region the width is proportional to:

$$I_{DS} \sim \frac{W}{L} (V_{GS} - V_T)^2 \quad (19)$$

Combining equation (18) and (19) we see that the ratio of gain over capacitance is optimized by using large overdrive voltages $V_{GS} - V_T$. In deep sub-micron the equations become more complex, but this trade-off is still to be made.

The gain of the switch can also be boosted by gain-boosting the switch itself. Gain boosting however is developed for transistors that remain in the on-region. Adapting the technique so that it does work on switched transistors is very challenging. It has not yet been done to the authors knowledge.

3.4. Sizing strategy for the switches and the cascodes

In the previous sections we stated that it is important to maximize the ratio between the gain and the capacitance at node N_1 in figure 1. In the design of a

current-steering DAC it is also important to keep the structure of each block as much as possible identical to the other blocks. This leads to a first sizing strategy for the switches and the cascodes: the switches and cascodes are designed for a LSB segment. This means that we determine the sizes for one LSB segment and then use only multiples of this segment for the bigger currents.

This is a good strategy for low accuracy converters, but not for converters with high precision. The LSB current gets very small. Minimum size switches and cascodes are then sufficient to conduct the current, even with low overdrive voltages. The capacitance C_{total} , as defined in section 3, then equals to 2^N-1 times the parasitic capacitance of two minimum sized switches and one minimum sized cascode. As the number of bits N increases, this capacitance grows to large. Therefore we first size the MSB segments and then scale down to the LSB segments. Minimum size transistors are then used for the switches and cascodes of a few LSB segments. The LSB segments are then no longer scaled versions of the MSB segments. This slightly deteriorates the dynamic performance of the converter, but this is more than compensated by having smaller parasitic capacitances.

3.5. Using cascodes on top of the switches

Another option to increase the ratio between the gain and the capacitance is to use the structure shown in figure 12. The capacitance on node N_2 includes the source capacitance of only one transistor and the smaller drain capacitance of another transistor. For the effects described in section 2.2 the ratio between the gain of the cascode on top of the switches and the capacitor on node N_2 now sets the upper-limit. As the usable overdrive voltages are smaller in this structure the upper-limit for the single-ended SFDR is only slightly higher with our new structure. The even-order distortion fortunately can be reduced by using differential structures.

The main advantage of this structure is that now node N_3 is buffered with both a switch, operating in the saturation region, and a cascode above the switch. The upper-limit for the differential-ended SFDR is now determined by the ratio between the combined gain of the switch and the cascode above the switch and the capacitance at node N_3 .

The maximum achievable SFDR for the differential DAC shown in figure 12 is now given by:

$$SFDR = 20 \log \left(\frac{\bar{I}_p}{I_{nl_third-order}} \right) = 20 \log \left(\frac{4A_{sw} A_{cas_top}}{\pi f R_L C_{total,2}} \right) \quad (20)$$

where $C_{total,2}$ is the sum of the capacitances connected on node N3. It is only slightly larger than the sum of the capacitances C_{total} connected on node N1. The upper-limit for the differential-ended SFDR is now roughly increased by the gain factor of the cascode on top of the switches.

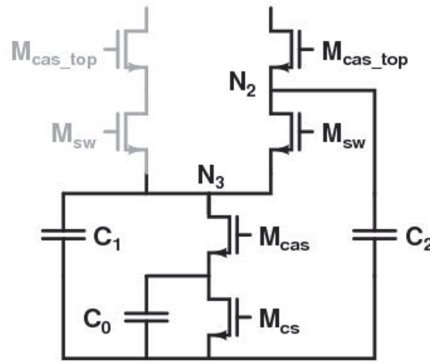


Fig. 12. Cascodes on top of the switches.

The structure has been used in [11]. The dynamic performance of this converter was however limited by the driver structure. The structure shown in figure 12 has potential drawbacks. The cascodes on top of the switches are continuously switched on and off. It is important to note that the gate voltage of these cascodes remains constant. The switching action is caused by the change in the source voltage of these transistors. This is depicted in figure 13. The voltage on node N2 is well defined if the switch is on. It is then determined by the bias voltage of the cascode on top of the switch, and the overdrive voltage required to conduct the current through this cascode. In our case the gate of this cascode is connected to the power supply V_{dd} . All voltages are plotted relative to V_{dd} . The digital blocks and the switches use a lower power supply, indicated by V_{sw} on the plot. If the switch is turned off, then node N2 is charged fast as long as the cascode on top of it remains in the strong inversion region. After that the charging slows down as the cascode enters the sub-threshold region. If the switch is switched on again then the initial conditions on node N2 are dependent on how long the switch has been turned off. This gives rise to code dependencies. It is therefore imperative that the bulk of the charging is over before the end of one clock period.

This switched operation clearly does not correspond to the normal operation of a cascode. Therefore it is important to not only have simulation results, which have often poor accuracy in strongly non-linear devices as DAC's, but measurement results as well.

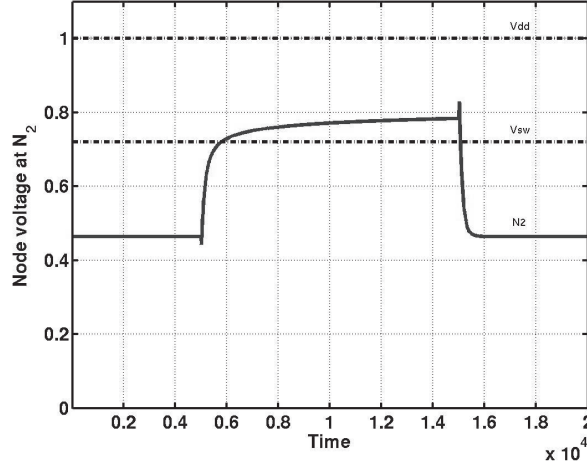


Fig. 13. Discharging of node N2.

3.6. Implementation and measurement results

In [12] we have used the proposed structure with cascodes on top of the switches in a 10-bit binary-weighted DAC. Although the main purpose was to demonstrate that a binary-weighted DAC can have good dynamic linearity, it also demonstrated that our switching structure can achieve good linearity. The basic structure is shown in figure 14. The actual DAC contains many of these basic structures in parallel.

Transistor M12 is used as the current source. This transistor has a large width and length and hence a lot of parasitic capacitance. Each current-source transistor is embedded in the large current-source-array. The current-source transistors are connected through long wires. These wires add a lot of interconnect capacitance between the drain node of the current-source transistor and the ground node. The large capacitance on this node is shielded from node N9 by cascode transistor M11. The use of this cascode is vital as the capacitance on node N9 must be low in order to increase the SFDR. This can be seen in equation 20.

Transistors M9 and M10 are switches. Transistors M7 and M8 are the cascodes above the switches. Transistors M1-M6 and inverters I1-I2 compose the drivers of the switches.

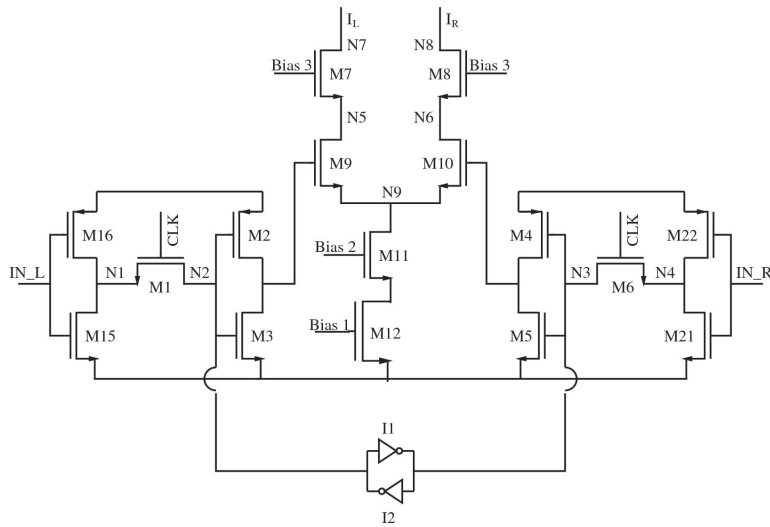


Fig. 14. Basic scheme of our current-steering DAC.

The structure is very simple and uses no bootstrapping, no gain-boosting and no feedback. Therefore it is easier to analyze and design and it is very low power. Figure 15 shows an overview of the SFDR performance at 250 MS/s.

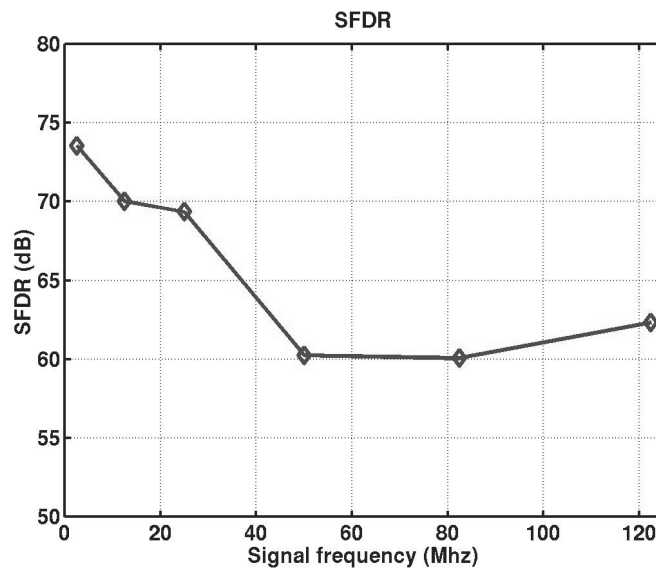


Fig. 15. SFDR performance of the converter at 250MS/s.

Figure 16 shows a dual-tone test. The IMD is at -67dBc in the shown frequency band.

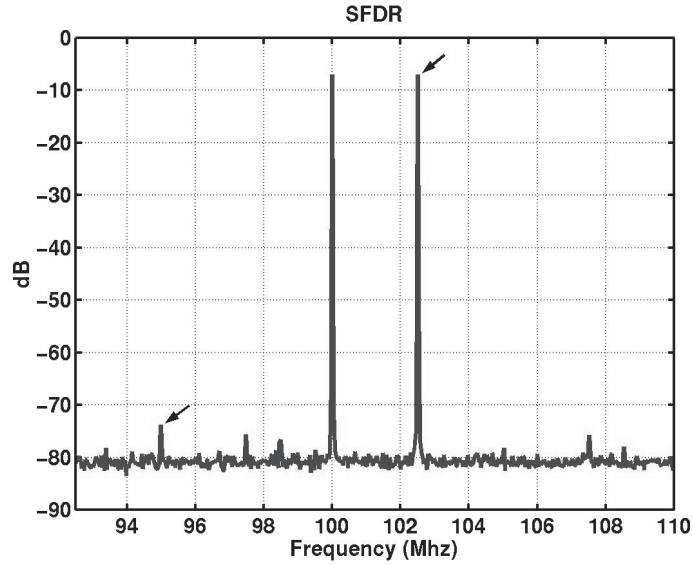


Fig. 16. Dual-tone test at 250MS/s.

Table 1 summarizes the performance of the converter. We decided to measure the dynamic performance for two different load currents. The dynamic performance is depending on the combination of the latches and the switches. It is optimized for a certain operating point. The chip is designed to have a 10 mA load current. By measuring the performance with a load current that is only 5 mA we operate the device far outside its operation region. The minimum SFDR over the Nyquist band drops with as few as 2.2 dB while reducing the load current with a factor of 2. This demonstrates that our structure is quite robust to variations in the operation point and therefore expected to be quite tolerant against process variations. Also important to notice is that the power consumption is rather low. The converter only consumes 4 mW plus the power consumed by the load current.

Table 1.

INL, DNL	< 0.1 LSB
Active area	< 0.35 mm ²
SFDR ($I_{Load} = 10$ mA)	> 60 dB
SFDR ($I_{Load} = 5$ mA)	> 57.8 dB
Glitch energy	2.64 pV.s
Power consumption at Nyquist	4 mW + $I_{Load} \times 1.8$ V

3.7. Conclusions

In this paper we discussed the importance of the output impedance for current-steering DAC's. Compared to previous publications we argued that the third-order distortion can not be ignored. This is important as it states for the first time that using differential structures does not overcome the impedance problem. The factors of importance are identified so that we have guidelines for better design. Several design options were discussed and the measurement results of an actual chip are discussed.

References

- [1] A. van den Bosch, M.S.J. Steyaert, W. Sansen, "SFDR-bandwidth limitations for high speed high resolution current steering CMOS D/A converters", Electronics, Circuits and Systems, 1999. Proceedings of ICECS '99. The 6th IEEE International Conference on, Volume 3, 5-8 Sept. 1999, pp. 1193-1196
- [2] Luschas S. and Lee H.-S., "Output impedance requirements for DACs", Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on, Volume 1, 25-28 May 2003, pp. 861-864
- [3] C-H. Lin and K. Bult, "A 10-b, 500-MSample/s CMOS DAC in 0.6 mm²", Solid-State Circuits, IEEE Journal of, Volume 33, Issue 12, Dec. 1998, pp. 1948-1958
- [4] A. van den Bosch, M.A.F. Borremans, M.S.J. Steyaert, W. Sansen, "10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter", Solid-State Circuits, IEEE Journal of, Volume 36, Issue 3, March 2001, pp. 315-324
- [5] A. Van den Bosch, M.A.F. Borremans, M. Steyaert and W. Sansen, "A 10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter", Solid-State Circuits, IEEE Journal of, Volume 36, Issue 3, March 2001, pp. 315-324
- [6] K. Doris, J. Briaire, D. Leenaerts, M. Vertregt and A. van Roermund, "A 12b 500 MS/s DAC with >70dB SFDR up to 120 Mhz in 0.18µm CMOS", ISSCC 2005, Feb. 7, pp.116-117
- [7] K. Bult and G.J.G.M. Geelen, "A fast-settling CMOS op amp for SC circuits with 90-dB DC gain", Solid-State Circuits, IEEE Journal of, Volume 25, Issue 6, Dec. 1990, pp. 1379-1384
- [8] Leon W. Couch, "Digital and Analog Communication", Systems Prentice Hall International, 1997
- [9] Kenneth R. Laker and Willy M. C. Sansen, "Design of analog integrated circuits and systems", McGraw-Hill, 1994
- [10] W. Schofield, D. Mercer, L.S. Onge, "A 16b 400MS/s DAC with <-80dBc IMD to 300MHz and <-60dBm/Hz noise power spectral density", Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC. 2003 IEEE International, Pages:126 - 482 vol.1

- [11] J. Bastos, A.M. Marques, M.S.J. Steyaert and W. Sansen, "A 12-bit intrinsic accuracy high-speed CMOS DAC", *Solid-State Circuits, IEEE Journal of*, Volume 33, Issue 12, Dec. 1998, pp. 1959-1969
- [12] J. Deveugele And M. Steyaert, "A 10b 250MS/s binary-weighted current-steering DAC", *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*, 15-19 Feb. 2004, pp. 362-532

HIGH-SPEED BANDPASS ADCs

R. Schreier
Analog Devices, Inc., Wilmington MA

Abstract

A bandpass ADC digitizes a bandpass signal directly, without prior conversion to baseband. Bandpass ADCs are well-suited to wired and wireless receivers, and can reduce system complexity, increase integration and improve performance. This paper describes architectures for bandpass and quadrature bandpass ADCs and examines several circuit considerations associated with operation at sampling rates in the 100-MHz range.

1. Introduction

Narrowband analog bandpass signals are found in a wide variety of telecommunications systems, including cellular telephony, radio, and television. In these applications, the desired radio frequency signal is usually extremely narrowband, having a bandwidth which is often less than 1% of the carrier frequency. In addition, most of the intermediate frequencies in a superheterodyne (superhet) receiver designed for such applications are also narrowband. In such systems, the ability to digitize a narrowband signal with a bandpass ADC offers advantages in terms of such critical performance metrics as dynamic range, power and cost.

Fig. 1 shows the block diagram of a typical multi-step superhet receiver with a digital back-end. As the figure shows, a sequence of filter/amplify/mix operations is used to convert the desired signal from radio frequency (RF), typically in the VHF (30 MHz to 300 MHz) or UHF (300 MHz to 3 GHz) range, down to one or more intermediate frequencies (IFs) and finally down to baseband, where

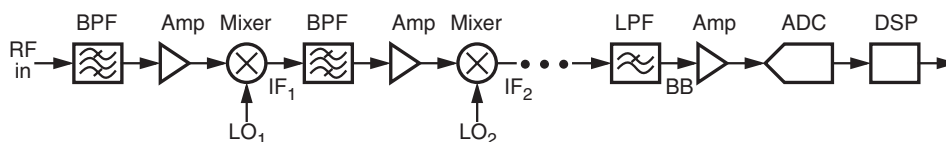


Fig. 1. A typical superheterodyne receiver with a digital back-end.

the signal is digitized by an ADC and processed by a digital signal processor (DSP). The venerable superhet architecture is preferred in applications that require high dynamic range and a high degree of immunity to interferers. A superhet radio achieves its immunity to interferers by progressively refining the signal with a sequence of increasingly narrow filters until only the desired signal remains. These filters operate on successively lower IFs in order to keep the complexity of the filters within reasonable limits.

In the early stages of the receiver, an important function of these filters is to attenuate undesired signals and noise which would mix down to the next IF along with the desired signal. Since the complexity of such *image filters* increases as the relative separation of the desired signal and the image signal decreases, and since the separation is equal to twice the IF, each downconversion stage typically reduces the carrier frequency of the desired signal by no more than about a factor of 10. For example, the standard first IF of an FM radio is 10.7 MHz, which is about $1/10^{\text{th}}$ of the 88-108-MHz RF. Since the bandwidth of this FM signal is less than 200 kHz, further downconversion and filtering could be applied so that the ADC would only have to deal with the desired 200-kHz-wide baseband signal. These extra conversion and filtering operations could be eliminated if the ADC were able to digitize the 10.7-MHz IF signal directly, and if it could do so with adequate dynamic range. Receiver systems for other applications experience similar simplifications when the ADC is shifted closer to the front end.

An ADC which supports this early conversion to digital form is the *bandpass ADC* [1]. With a bandpass ADC, the architecture of Fig. 1 can be reduced to that shown in Fig. 2. A bandpass ADC concentrates its conversion effort on the band of interest only, and can therefore be more efficient than an ADC which digitizes the entire band from dc to the IF.

Replacing the mixer in Fig. 2 with a quadrature mixer, and replacing the bandpass ADC with a quadrature bandpass ADC [2] allows the system to dispense with the image filter, yielding the simplified system shown in Fig. 3. For the ul-

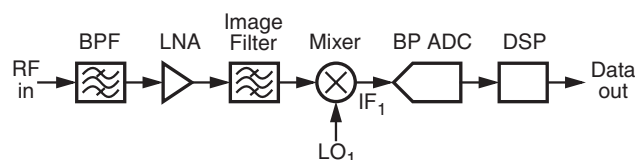


Fig. 2. A receiver employing a bandpass ADC at IF_1 , the first IF.

mate in diagrammatic simplicity, the software radio receiver architecture shown in Fig. 4 absorbs all analog functions into the ADC [3]. In this receiver, all standard-specific parameters (center frequency, bandwidth, modulation format and access protocol) are defined through digital signal processing. This ambitious architecture puts an enormous burden on the ADC, and thus systems which employ this architecture tend to consume much more power than systems which are optimized for a small number of communications standards.

The above considerations, namely system simplification and improved power efficiency, provide some justification for pursuing the design of bandpass ADCs in general. A final reason to consider the bandpass approach is that, by keeping the signal band away from dc, a bandpass ADC preserves the spectral separation between the signal of interest and various low-frequency noise sources and distortion components such as $1/f$ noise and even-order intermodulation products.

Since bandpass ADCs with center frequencies in the range of a few MHz and bandwidths in the range of a few hundred kHz already exist in commercial form, the focus of this paper is on high-speed bandpass ADCs. The impetus for bandpass ADCs having center frequencies in the tens of MHz and bandwidths in the MHz range comes from the desire to place the ADC as close to the antenna as possible, and also to broaden the application area of bandpass ADCs.

Pushing the ADC toward the front-end of the radio increases the ADC's dynamic range requirements since the ADC must now cope with the potentially-much-larger interfering signals that otherwise would have been removed by analog filters. The dynamic range requirements are usually so high that a delta-sigma ($\Delta\Sigma$) architecture is the best, or even the only, choice.

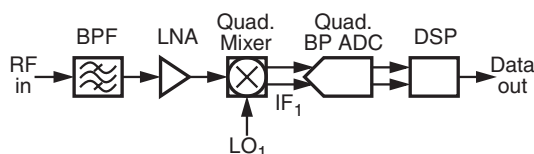


Fig. 3. A receiver employing a quadrature bandpass ADC.

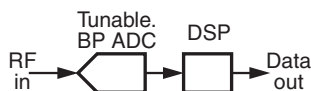


Fig. 4. A software radio receiver.

However, $\Delta\Sigma$ converters rely on the use of oversampling. A $\Delta\Sigma$ converter typically requires an oversampling ratio (OSR) of at least 8, with 30 being a fairly representative number. Returning to the example of the 10.7-MHz IF of an FM radio, direct application of standard (lowpass) $\Delta\Sigma$ techniques would require a sampling rate in the 600-MHz range—a rather daunting number. However, since OSR is the ratio of the sampling rate to twice the signal *bandwidth*, achieving an OSR of 30 for a signal with a 200-kHz bandwidth actually only requires a sampling rate of 12 MHz, which is a much more practical number.¹ As described in the literature [1-5], the key to achieving this reduced sampling rate is to concentrate the zeros of the $\Delta\Sigma$ modulator's noise transfer function (NTF) in the band of interest, specifically in the vicinity of 10.7 MHz for the case of an FM radio, rather than have the NTF zeros centered on dc.

The shift in the NTF zeros away from dc requires the use of resonators instead of integrators in the modulator's loop filter. Suitable resonators can be realized in either discrete-time form (using switched-capacitor, switched-current or switched-op-amp techniques) or in continuous-time form (using gm-C, active-RC, LC or transmission-line techniques). Since high-frequency resonators are most conveniently implemented with continuous-time circuits, and since continuous-time circuits endow the ADC with inherent anti-aliasing, this paper will only consider continuous-time resonators.

The other critical circuit block in a high-speed BP $\Delta\Sigma$ ADC is the first feedback DAC. Important circuit considerations for this block are also described in this paper, but before delving into circuits, this paper examines some of the architectural alternatives available to the bandpass ADC designer.

2. Modulator Architecture

The following subsections discuss two architecture-level decisions that need to be made before detailed circuit design can begin. The first decision is whether to make a quadrature bandpass ADC or a regular bandpass ADC. Quadrature ADCs are more *complex* (pun intended), but offer performance and system-level advantages. The second decision involves the selection of modulator topology (feedback, feedforward or hybrid). Other high-level decisions include the selec-

1. For convenience in the digital post-processing, the IF is typically located at a simple rational fraction of the sampling rate. For our FM example, setting $f_0 = 3f_s/4$ results in $f_s = 14.2666$ MHz and an OSR of about 36, while setting $f_0 = f_s/4$ results in $f_s = 42.8$ MHz and $OSR \approx 100$. Either choice is a reasonable one.

tion of sampling rate and center frequency, the choice between discrete-time and continuous-time signal processing, the choice of a single-loop or a multi-loop (cascade) topology, and the selection of the number of quantization levels.

Since frequency-planning considerations and OSR requirements usually constrain the sampling rate and center frequency to the point where very little design freedom is available, this paper will not explore the trade-offs associated with these design parameters. The speed and inherent anti-aliasing advantages of continuous-time circuitry were already mentioned, and these provide the justification for not considering discrete-time circuits further. Similarly, since this paper is concerned with systems which are wideband, the use of multi-bit quantization is virtually a necessity.

Lastly, this paper assumes single-loop topology will be used. Multi-loop topologies have significant advantages in the context of wideband systems, but since a multi-loop system requires its noise-cancelling digital filter to match the NTF of the $\Delta\Sigma$ modulator, and since the NTF a high-speed continuous-time $\Delta\Sigma$ ADC tends to be ill-controlled, this matching can be difficult to achieve. Calibration techniques have been successfully used to achieve the required matching [6], but this paper will focus on single-loop systems as these are free of such concerns and since a single-loop system is often the starting point for a multi-loop system.

2.1. Bandpass vs. Quadrature Bandpass

Fig. 5 illustrates how a bandpass $\Delta\Sigma$ ADC system appears to the system designer. The input to the ADC is usually an IF signal, but in some cases may be at RF. The high-speed output of the $\Delta\Sigma$ modulator contains the desired signal surrounded by shaped quantization noise, plus interfering signals. The digital output of the modulator is mixed to dc by a digital quadrature mixer, and then

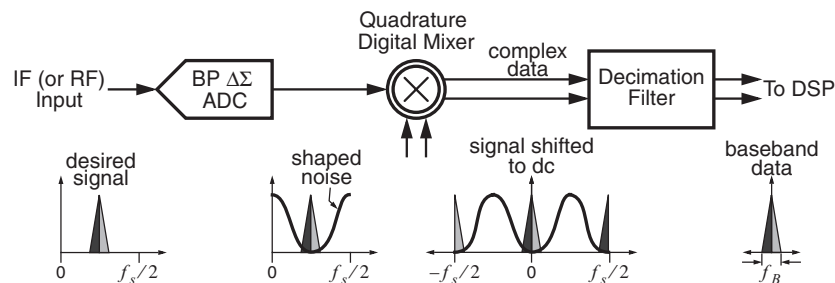


Fig. 5. A bandpass $\Delta\Sigma$ ADC system.

lowpass-filtered and decimated by a quadrature lowpass digital decimation filter, so that the final output is reduced-rate baseband digital data containing only the desired signal.

The oversampling ratio of a bandpass system is defined in the same manner as in a lowpass system, namely $OSR = f_s/(2f_B)$, so that $OSR = 1$ corresponds to Nyquist-rate sampling. Note that for a bandpass signal, f_B is the *two-sided* bandwidth.

In a lowpass system, the modulator output can be decimated by a factor of OSR without loss of information, since the minimum sample rate at the output of the decimation filter is $2f_B$. In a bandpass system, however, the minimum sample rate at the output of the decimation filter is only f_B because the output of the decimation filter is complex data. Thus, the data from a bandpass ADC can be decimated by a factor as high as $2 \times OSR$ without loss of information.

Just as a bandpass modulator can exploit the narrowband character of its input, a *quadrature $\Delta\Sigma$ modulator* can exploit the additional information available in a quadrature signal¹. Fig. 6 illustrates the main signal-processing operations that occur within a quadrature $\Delta\Sigma$ ADC system. A quadrature signal, such as that produced by a quadrature mixer, is applied to a quadrature $\Delta\Sigma$ modulator which outputs a digital quadrature signal containing the desired signal and the shaped noise

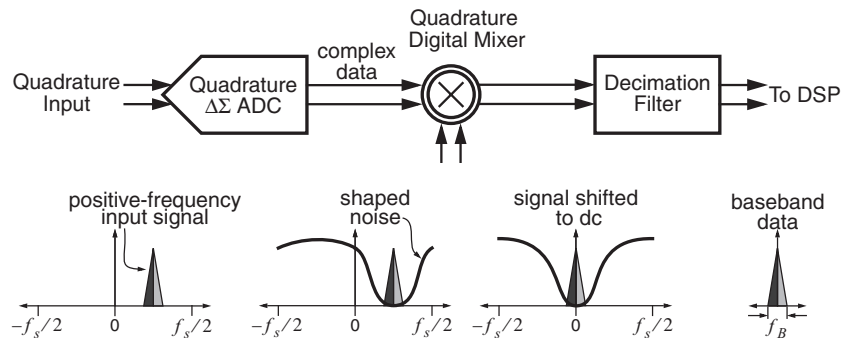


Fig. 6. A quadrature $\Delta\Sigma$ ADC system.

1. A quadrature, or complex, signal consists of two real signals, commonly denoted either by I (for in-phase) and Q (for quadrature phase), or by re (for real) and im (for imaginary). The key difference between a real signal and quadrature signal is that the spectrum of a quadrature signal need not be symmetric about zero frequency. For a quadrature signal, positive frequencies are truly distinct from negative frequencies.

quantization noise. The distinguishing feature of a quadrature modulator is that its NTF need only attenuate positive-frequency (or negative-frequency) quantization noise. In a sense, a quadrature converter is more efficient than a bandpass converter because no power is wasted digitizing the negative-frequency content of the input. As in a real (non-quadrature) system, the modulator output is mixed to baseband by a digital quadrature mixer and filtered by a quadrature decimation filter to produce Nyquist-rate baseband data.

For a real system, signals beyond $f_s/2$ suffer from aliasing, whereas for a quadrature system the corresponding limits are $\pm f_s/2$. The total alias-free bandwidth is thus f_s . In order for $OSR = 1$ to correspond to no oversampling, the OSR of a quadrature system is defined as $OSR = f_s/f_B$. In other words, for a given signal bandwidth and sampling rate, a quadrature modulator has an OSR that is twice that of a real modulator. Lastly, since the minimum output data rate is f_B , decimation by a factor of OSR is appropriate for a quadrature system.

A quadrature $\Delta\Sigma$ modulator can be either lowpass or bandpass. However, since a quadrature lowpass modulator is equivalent to a pair of regular lowpass modulators operating independently on the components of the quadrature signal, the advantages of a quadrature lowpass modulator over competing architectures are not as pronounced as they are for a quadrature bandpass modulator. In the bandpass case, a quadrature modulator is useful because it effectively doubles OSR , and it does so without doubling the hardware. Specifically, a bandpass modulator having n in-band zeros requires a loop filter of order $2n$, containing $2n$ op-amps, whereas a quadrature modulator having n in-band zeros requires a complex loop filter of order n , which also contains $2n$ op-amps.

As an example, Fig. 7 depicts a 6th-order NTF for a bandpass modulator, while Fig. 8 does likewise for a 3rd-order quadrature bandpass modulator. Both modulators require 6 op amps in their implementation. Also, both modulators employ an oversampling ratio of 32 and theoretically achieve 16-bit SQNR performance with 4-bit quantizers. However, since $f_B = f_s/OSR$ in a quadrature system but only $f_B = f_s/(2 \cdot OSR)$ in a real modulator, the quadrature modulator has a bandwidth which is twice that of the real modulator, assuming a common sampling rate.

The above example illustrates the primary advantage of quadrature modulation, namely a doubling of the signal bandwidth, for a given OSR and sampling rate. Also, as mentioned in the introduction, quadrature modulation facilitates the elimination of an image filter. The primary disadvantage of a quadrature system is increased complexity, in particular a doubling in the number of quantizers,

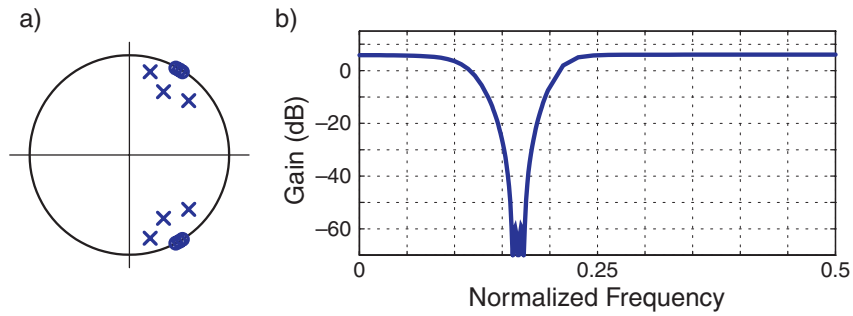


Fig. 7. a) Pole-zero and b) magnitude plots for a bandpass NTF with $f_0 = f_s/6$ and $OSR = 32$.

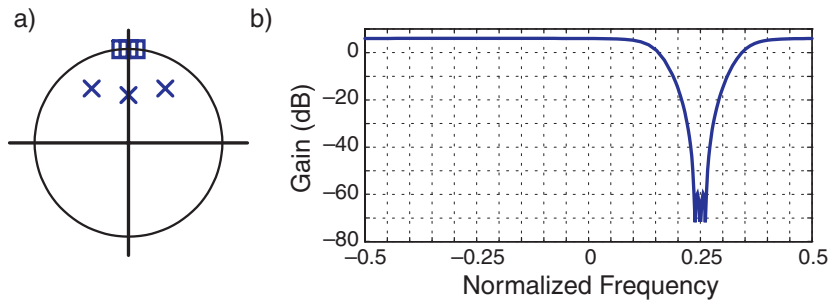


Fig. 8. a) Pole-zero and b) magnitude plots for a quadrature bandpass NTF with $f_0 = +f_s/4$ and $OSR = 32$.

feedback DACs and output bits. (Bear in mind that the complexity of a quadrature bandpass modulator's loop filter is essentially the same as that of a real bandpass modulator.) Although space limitations preclude a detailed discussion of the topic, it is important for the reader to be aware of the fact that quadrature systems are also sensitive to path mismatch. The degree of sensitivity is usually severe enough to require the addition of one or more image zeros to the NTF. More details can be found in Ch. 5 of [5].

2.2. Loop Filter Architectures

Bandpass modulators possess the same architectural variety as lowpass modulators, and the trade-offs between the different structures are also essentially the same. Bandpass modulators can be implemented in single-loop or cascade form, with a similar trade-off between improved stability and increased sensitivity to

analog non-idealities such as parameter errors and finite op-amp gain. Likewise, the loop filter of a bandpass modulator can be constructed using any of the conventional forms found in lowpass modulators, including feedback, feedforward and hybrid topologies, with similar trade-offs between internal dynamic range and STF quality.

For example, Fig. 9 shows the structure of the loop filter of a 4th-order bandpass modulator which employs a feedback topology, while Fig. 10 does likewise for a feedforward topology. Fig. 11 plots representative signal transfer functions (STFs) for these two topologies. As the figure shows, the STF associated with the feedback topology has an attractive bandpass shape, whereas the STF associated with the feedforward topology has out-of-band peaks. These peaks make a feedforward modulator vulnerable to large-amplitude interfering signals in the vicinity of the STF peaks. For this reason, a feedforward topology should only be used when the incoming signal has been adequately filtered. As described in the literature, the main motivation for adopting a feedforward architecture is that it reduces the dynamic range requirements in the all-important first resonator.

When f_0 is a substantial fraction of the sampling rate, there is strong coupling between the two integrators that comprise a resonator, and thus the resonator

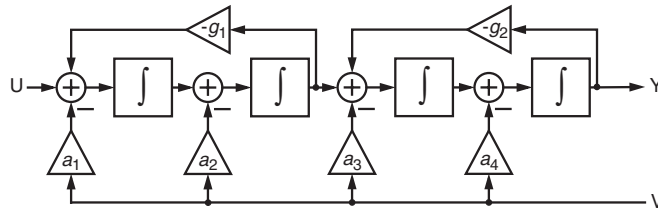


Fig. 9. Loop filter of a 4th-order bandpass modulator employing the standard feedback topology.

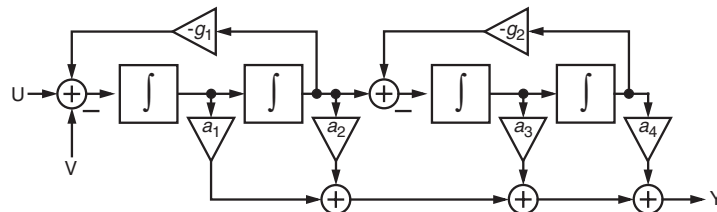


Fig. 10. Loop filter of a 4th-order bandpass modulator employing the standard feedforward topology.

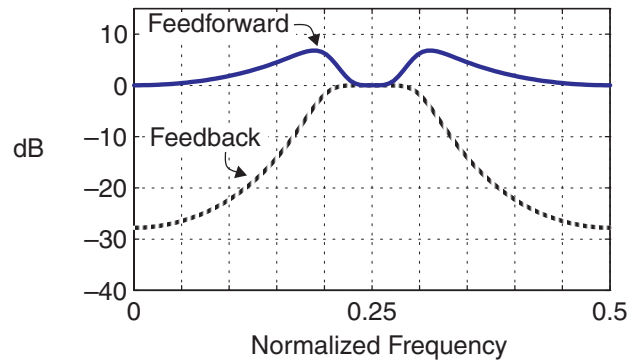


Fig. 11. Representative STFs for modulators employing feedforward and feedback topologies.

output may be taken from the first integrator, as shown in Fig. 12. Taking the resonator output from the first, rather than the second, integrator's output changes the transfer function of the resonator from $\omega_0^2/(s^2 + \omega_0^2)$, which is a lowpass response, to $s\omega_0/(s^2 + \omega_0^2)$, which is a bandpass one. Since the bandpass response has a null at dc, it is clear that a lowpass modulator cannot make use of these bandpass resonators. However, a bandpass modulator can. Since the $n/2$ resonators in a bandpass modulator may either be of the lowpass or bandpass variety, there are $2^{n/2}$ possible lowpass/bandpass resonator combinations for a given loop-filter category such as the feedback, feedforward or any of the hybrid categories.

Fig. 13 illustrates how adding a feedforward path and thus connecting the output of one resonator to both of the integrators in the next resonator can eliminate one of the feedback coefficients (i.e., one of the feedback DACs) in a bandpass modulator. Since the transfer function from V to Y is the same in Fig. 13 as that of

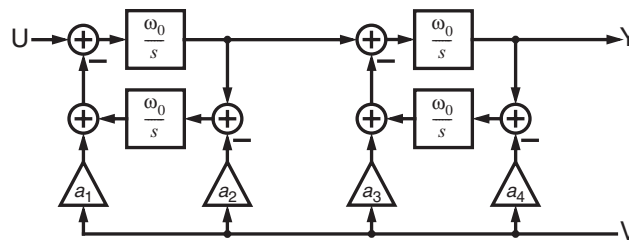


Fig. 12. Loop filter of a 4th-order bandpass modulator employing a feedback topology with bandpass resonators.

Fig. 12, the noise transfer function of a modulator employing the loop filter of Fig. 13 will be the same as that of a modulator employing the loop filter of Fig. 12. (Of course, the signal transfer functions may not be the same.) This transformation may be applied to each resonator section except the last one, thereby cutting the required number of DACs by nearly 50%. This transformation is helpful in the construction of a bandpass modulators which employs LC tanks as the resonance elements.

Fig. 14 shows a portion of a loop filter which encompasses all of the above variants. Each resonator section is coupled to the next through 4 arbitrary gains, so the choice of a lowpass vs. a bandpass section is simply a special case in which all coefficients are zero except for one. The feedback DACs are not shown, and could be added to any or all of the integrator summing junctions, according to whether a feedback, feedforward or hybrid modulator topology is used.

Fig. 15 shows the structure of a quadrature modulator employing a feedback topology. As the figure shows, the resonators which make up the modulator's loop filter are again special cases of Fig. 14, in which $c_1 = c_4 = c_x$ and

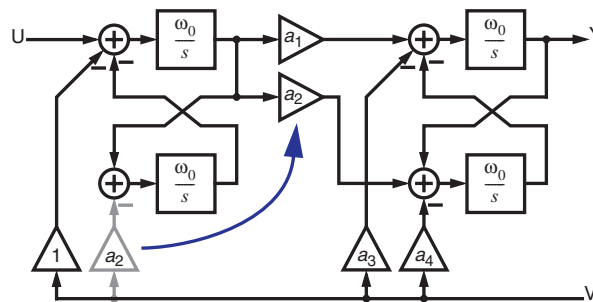


Fig. 13. Eliminating a feedback DAC by adding a feedforward path.

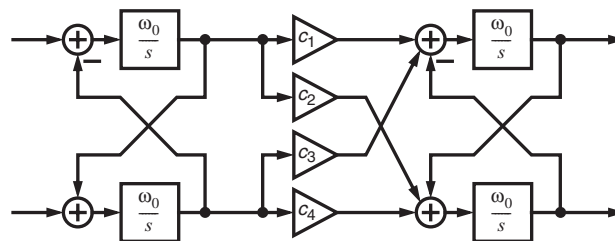


Fig. 14. Internal structure of a more general loop filter for a bandpass modulator.

$c_2 = -c_3 = c_y$, where $c = c_x + jc_y$ is the complex coefficient linking the first and second resonators.

The above diagrams contain a common element, namely a pair of integrators that have been cross-coupled to form a resonator. The next section describes circuits which can implement these elements.

3. Resonator Design

A lowpass modulator needs good *integrators*, whereas a bandpass modulator needs good *resonators*. The degradation to modulator performance caused by a finite quality factor (Q) in the resonators of a bandpass modulator is analogous to the degradation caused by finite dc gain in the integrators of a lowpass modulator: both cause reduced SQNR and increased susceptibility to tonal behavior. The SQNR degradation is significant when Q falls below f_0/f_B . Thus, in order to take full advantage of a high value of *OSR*, the Q of each resonator should be high. Conversely, when the signal is not especially narrowband, i.e. when f_0/f_B is not very high, the Q requirements for nearly ideal operation are relaxed. The resonant frequency of the resonator must be accurate for similar reasons. A frequency error that is an appreciable fraction of f_B , say 20%, is usually close to the level of significance. Once again, a high value for *OSR* dictates more stringent accuracy requirements, unless the NTF has been designed to have

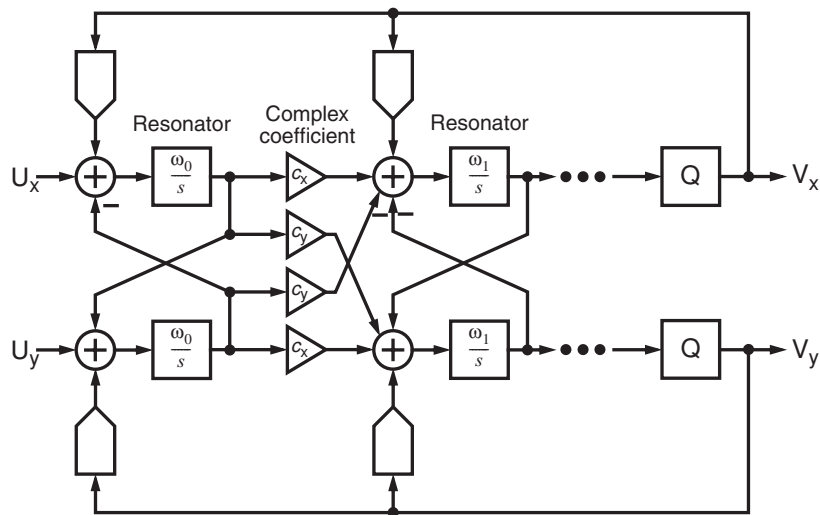


Fig. 15. A quadrature modulator employing the feedback topology.

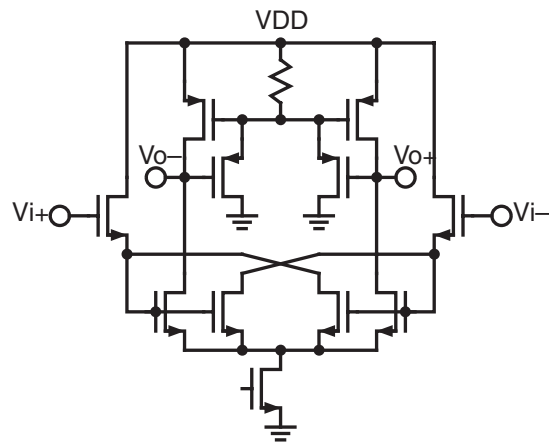


Fig. 17. A low-phase-shift transconductor (Fig. 7 of [7].)

3.2. Active-RC Resonator

Fig. 18 shows the structure of an active-RC resonator. Here the center frequency is given by $\omega_0 = 1/(RC)$, and once again the highly variable RC product necessitates the use of tuning. Tuning may be accomplished by adjusting R (continuously via MOS devices, or in discrete steps using a resistor array), by adjusting C (here an array is most practical), or by a combination of the two approaches. Once again, configuring the resonator as an oscillator is straightforward and eliminates the need for a replica block, but can only be done while the converter is off-line.

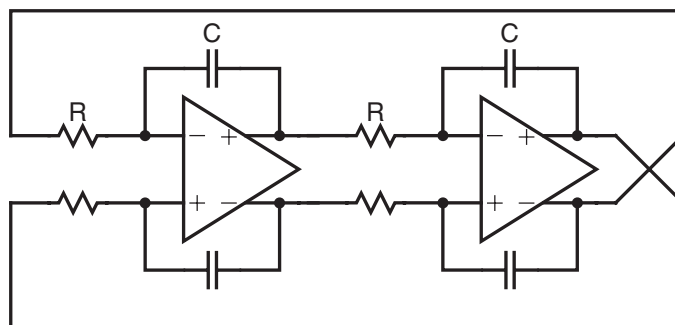


Fig. 18. An active-RC resonator.

Fig. 19 contains a derivation of the pole location for a two-integrator loop. In order to guarantee a high Q , the gain of the op amp must also be fairly high. Specifically, $Q = 25$ requires a gain of 100 at f_0 . Assuming $f_0 = 25$ MHz and a first-order roll-off leads to a gain-bandwidth product of $GBW = 2.5$ GHz, which is rather high.

However, the pole Q depends on the phase of the op amp gain as well as on its magnitude. If the phase shift of the op amp is 45 degrees at f_0 , then the pole of the loop slides along the imaginary axis and the Q of the system remains high. This shift in resonant frequency is not problematic, since f_0 has to be tuned anyway. Fig. 20a shows a circuit which has the required phase shift. If the two ground symbols are replaced by the virtual grounds produced at the inputs of the op amps as indicated in the two-integrator cascade, then the resulting resonator (shown in Fig. 20b) has a Q which is insensitive to the gain (g_m) of the op amp.

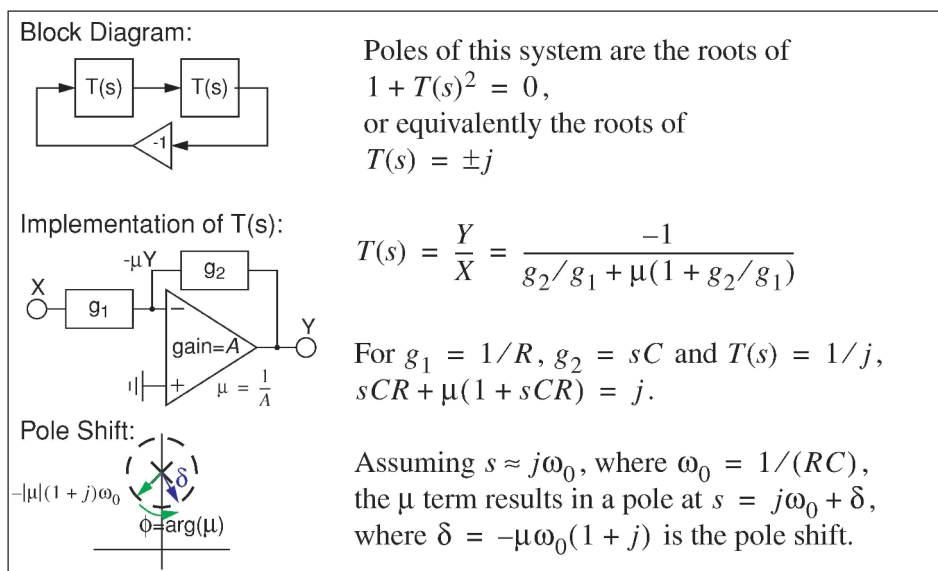


Fig. 19. Derivation of the pole shift in an active-RC resonator.

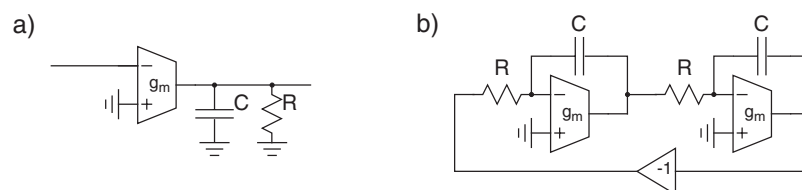


Fig. 20. A circuit with 45° phase shift at $\omega_0 = 1/(RC)$ and the associated resonator structure.

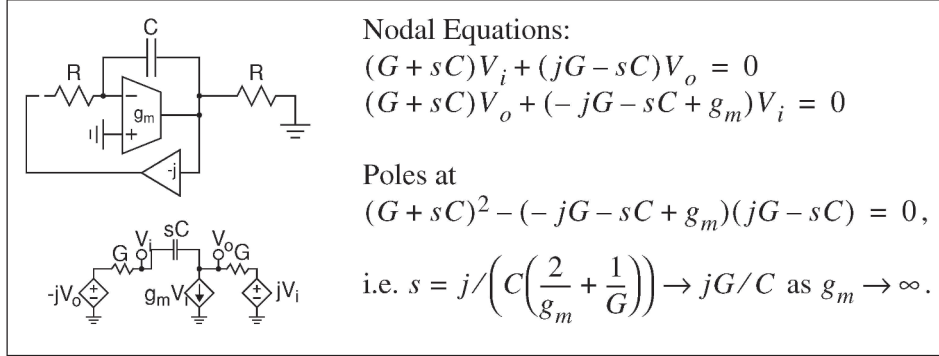


Fig. 21. Quadrature representation of Fig. 20b, and associated analysis.

This technique reduces the required GBW by an order of magnitude. All that is needed is a transconductor with low internal phase shift at f_0 , such as that of Fig. 17.

The configuration in Fig. 20b produces a resonance that is nominally at $s = 1/(R_1 C_1)$. Fig. 21, confirms the analysis of Fig. 19 using a quadrature representation of the resonator, and once again establishes that using a pure transconductance for the amplifiers lowers the resonant frequency, but does not degrade resonator Q . Fig. 22 repeats this analysis with non-zero switch resistance in the capacitor array and finite bandwidth in the transconductor. This analysis indicates that finite bandwidth pushes the pole to the right, while non-zero switch resistance pushes the pole to the left. With an infinite-bandwidth g_m , the Q of the resonator is $Q = R_1/(2R_{sw})$. (Thus, for $Q > 25$, we need $R_{sw} < R_1/50$.) With a zero-resistance switch, the Q of the resonator is $Q = -f_u/(4f_0)$ where f_u is the unity-gain bandwidth of the transconductor when loaded by resistance R_1 . (Once again, $Q > 25$ for $f_0 = 25$ MHz requires $f_u > 2.5$ GHz.) However, as shown in Fig. 22, a fortuitous cancellation happens if $f_u = \pi/(R_{sw}C)$. This cancellation is somewhat process-sensitive, but can typically be relied upon to reduce the transconductance bandwidth requirement by a factor of 2.

3.3. LC Resonator

The last resonator to be considered in this paper is the LC tank driven by a current source, shown in Fig. 23. From the viewpoint of complete integration, this topology represents a backwards step. On-chip inductors possess only a few nanohenries of inductance, and so would only be useful if the center frequency is

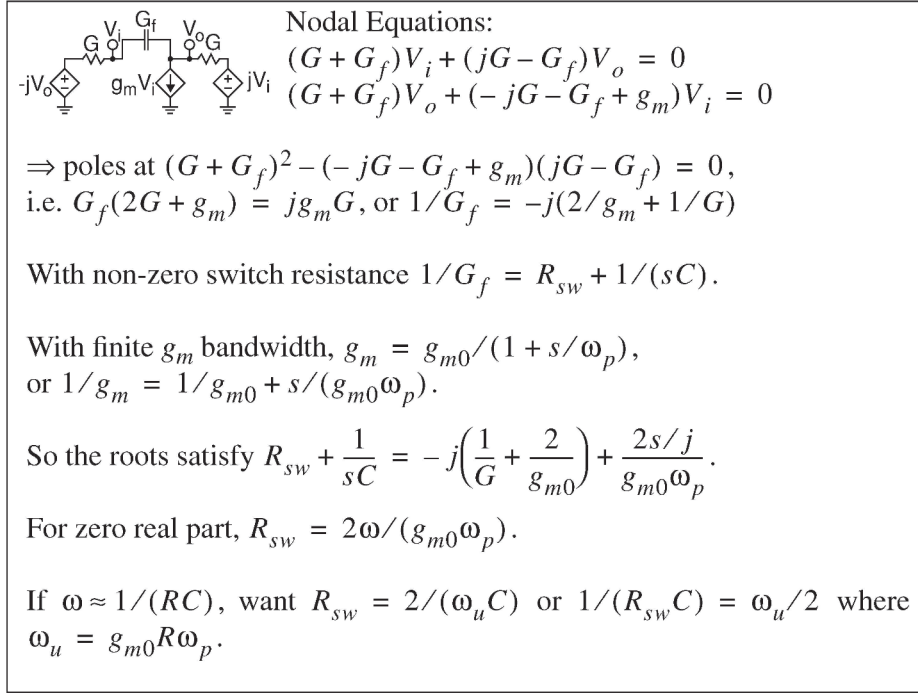


Fig. 22. Analysis of Fig. 21 with finite switch resistance and finite transconductor bandwidth.

above 1 GHz or so. Since such high frequencies are currently beyond the reach of existing mainstream technologies, most bandpass modulators which exploit inductors have relied on external components. As with the G_m -C and active-RC resonators, the accuracy of the LC tank's center frequency is determined by the accuracy of its components. Since discrete inductors with tolerances on the order 2% and $Q > 50$ are available, as are capacitors with even tighter tolerances and higher Q , it is possible to implement a high- Q LC resonator without incorporating means for tuning. Furthermore, since inductors and capacitors are ideally noiseless, a resonator based on an LC tank enjoys an enormous noise advantage over the preceding resonator circuits. The distortion of an LC tank is also quite small compared to what can be achieved with active circuitry. Lastly, since an

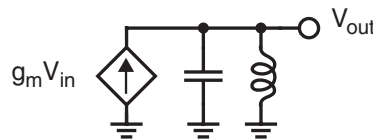


Fig. 23. A resonator based on an LC tank.

LC tank implements a *physical resonance* (as opposed to the *synthesized resonance* of an active-RC resonator), an LC resonator needs no bias power. Despite its resistance to integration, an LC tank possesses a number of important attributes (namely low noise, distortion and power!) that make its use in a band-pass converter highly advantageous [8].

The main drawbacks associated with the use of an LC tank are lower integration, lower center-frequency programmability, and the fact that a purely passive quadrature LC resonator does not exist.

4. Feedback DAC

Since a $\Delta\Sigma$ modulator is a feedback system, the performance of the system can be no better than its feedback element, namely the multi-bit DAC. The key performance specifications of the DAC are noise and linearity, and in the case of a high-speed bandpass converter, the ability to produce a clean high-frequency spectrum. Since CMOS current-mode DACs have been demonstrated to operate well at sampling rates in the 300 MHz range [9], these DACs are excellent candidates for this critical feedback function. This section quantifies two important design considerations for such DACs, namely noise and the nonlinearity due to non-ideal element dynamics. Other concerns in the DAC include $1/f$ noise and matching, but since these can be addressed by allocating sufficient area to the current sources and, in the case of matching, by using mismatch-shaping, these challenges can be overcome.

4.1. Thermal Noise

For a MOS device in saturation, the 1-sided spectral density of the output current noise is

$$S_{I_d}(f) = \frac{8kTg_m}{3}. \quad (1)$$

If the full-scale current of the DAC is I_{FS} , then the peak differential output current $(I_A - I_B)/2$ is $I_{FS}/2$ and the signal power for a -3 -dBFS output signal is

$$S^2 = \frac{0.5(I_{FS}/2)^2}{2} = \frac{(I_{FS})^2}{16}. \quad (2)$$

Assuming the square-law for a MOS device holds, $I_{FS} = K(\Delta V)^2$ and $g_m = 2K\Delta V$, where $\Delta V = V_{gs} - V_t$, so that

$$S^2 = \frac{K^2(\Delta V)^4}{16} \quad (3)$$

and the noise power in the differential output for a bandwidth B is¹

$$N^2 = \frac{S_{I_d}(f)}{4}B = \frac{4kTBK\Delta V}{3} \quad (4)$$

so that the signal-to-noise ratio is

$$\left(\frac{S}{N}\right)^2 = \frac{3K(\Delta V)^3}{64kTB} = \frac{3I_{FS}\Delta V}{64kTB}. \quad (5)$$

As a numerical example, consider $B = 5$ MHz and $\text{SNR} = 100$ dB. According to Eq. (5), we need $(I_{FS})(\Delta V) = 4.5$ mW. So assuming $\Delta V = 0.3$ V, then the full-scale current must be $I_{FS} = 15$ mA. Clearly achieving a high SNR in the DAC is purely a matter of allocating sufficient power to the DAC.

4.2. Element Dynamics

The switching behavior of the current sources is a well-known but less well-understood source of error. Appendix A shows that if an element's dynamics are dependent only on the previous state, then the time-domain response can be broken down into linear and nonlinear (error) components as shown in Fig. 24. The signals labelled w_1 and w_2 are waveforms which represent the linear portion of the element's response, while w_0 represents clock feed-through and e represents the nonlinear error. According to this model, e gets added to or subtracted from the output in each clock period, depending on whether the data changed state or stayed the same.

The w_0 , w_1 , w_2 and e waveforms can be computed from simulation of the element's response to 00, 01, 10 and 11 data patterns using the formula shown in Fig. 24 and derived in Appendix A. Specifically, the error waveform is given by

$$e = \frac{(w_{01} + w_{10}) - (w_{00} + w_{11})}{4}, \quad (6)$$

which shows that zero nonlinearity results when the sum of the 01 and 10 responses match the sum of the 00 and 11 responses. If the 00 and 11 responses are assumed to be flat lines, then the condition for zero nonlinearity is that the 01

1. $I_d = (I_A - I_B)/2 \Rightarrow N_{I_d}^2 = (N_{I_A}^2 + N_{I_B}^2)/4$. Now, $I_{FS} = I_A + I_B$, so $N_{I_{FS}}^2 = N_{I_A}^2 + N_{I_B}^2$ and thus $N_{I_d}^2 = N_{I_{FS}}^2/4$.

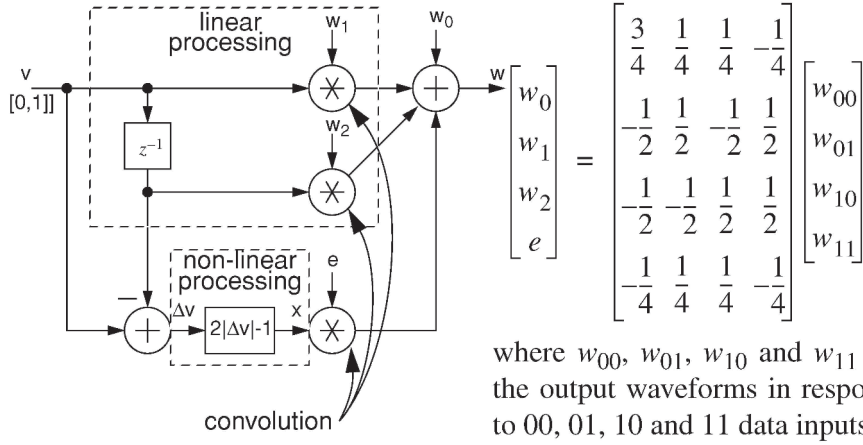


Fig. 24. Model for nonlinear element dynamics.

and 10 responses must be complementary. For a differential system with perfect symmetry, Eq. 6 is satisfied automatically, but since imbalance can allow the single-ended errors to leak through, it is important for the single-ended performance to be sufficiently high.

Once the error waveform has been computed, its impact on the spectrum of the DAC's output can be determined by convolving e with the switching sequence

$$x = \sum_{i=1}^M (2|\Delta v_i| - 1), \quad (7)$$

where v_i is the 1-bit (0/1) control signal for the i^{th} element. (If the data is thermometer-coded then x can be computed using $x = 2|\Delta v| - M$, whereas if the ADC data is mismatch-shaped then Eq. 7 must be used.)

This convolution can be done by stitching together appropriately scaled copies of e . Taking the Fourier transform of the waveform so constructed gives the spectrum of the error caused by DAC dynamics. A more efficient process is to multiply the Fourier transform of e by the spectrum of the x sequence. Both procedures assume that the error waveform is the same for every element.

Appendix A examines the DAC element shown in Fig. 25 using the model shown in Fig. 24. This circuit uses a 3V PMOS device as the current source in order to support a large ΔV , namely 1.5 V, so that sufficiently low noise can be achieved with a small DAC current. Appendix A's analysis of the single-ended

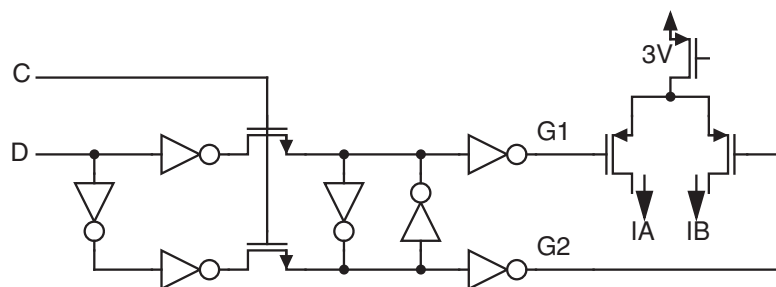


Fig. 25. Example DAC element and associated driver.

dynamic performance indicates that this circuit will behave adequately at a clock frequency of 100 MHz, with only a modest amount of cancellation (8 dB) from differential operation.

5. Conclusions

A bandpass $\Delta\Sigma$ ADC's ability to digitize an IF signal efficiently and with high linearity allows this ADC architecture to simplify the design of a high-performance receiver. This paper discussed two architecture-level choices (bandpass vs. quadrature bandpass modulation and selection of the loop topology), and also discussed a few of the important considerations in the design the two most critical blocks in a bandpass converter (the first feedback DAC and the first resonator).

It was shown that a quadrature bandpass converter has performance and system advantages, but is more complex than a regular bandpass converter. Several topologies for bandpass modulators were presented, and the important features of each were identified.

For the resonator block, three approaches were considered: two active (gm-C and active-RC) and one passive (LC). The LC resonator has unbeatable noise, power and distortion performance, but is resistant to integration and is not easily adapted to the construction of quadrature modulators. It was shown that using a transconductance element for the amplifier and adding series resistance to the integrating capacitor allows an RC resonator to reliably achieve high Q values for center frequencies in the tens of MHz, thereby making an active-RC resonator an excellent choice for both bandpass and quadrature bandpass modulators.

Lastly, this paper examined the issues of thermal noise and element dynamics in a current-mode CMOS DAC. Achieving low noise was demonstrated to be simply a matter of burning enough power. A signal-processing model of the nonlinearity caused by element dynamics was derived and then used to quantify the performance of a simple current-mode DAC implemented in 0.18- μm technology. The model indicates that element dynamics should not be a limiting factor for a 25-MHz center frequency.

Numerous other challenges exist in the construction of a bandpass ADC with a center frequency in the tens of MHz, a bandwidth of a few MHz and a dynamic range of 90 dB, but this paper shows that the first-level design challenges are manageable with existing techniques and technologies. The reader should expect to see several such converters reported in the literature in the next few years.

Appendix A: Modeling Element Dynamics

If the response of an element is dependent only on its previous state, then it is possible to construct a complete output waveform by concatenating waveforms according to the following table, where w_{ij} represents the output waveform over one clock period in response to a transition from state i to state j .

$v(n-1)$	$v(n)$	$w(t), t \in [nT, (n+1)T]$	w_{lin}
0	0	w_{00}	w_0
0	1	w_{01}	$w_0 + w_1$
1	0	w_{10}	$w_0 + w_2$
1	1	w_{11}	$w_0 + w_1 + w_2$

Table 1: Waveform look-up table and waveforms from linear model.

We want to model this behavior with a linear system plus offset:

$$w_{lin} = w_0 + w_1 v(n) + w_2 v(n-1) \quad (8)$$

while minimizing the error

$$e = w - w_{lin} \quad (9)$$

in the mean-square sense. Table 1 lists the output from the linear model alongside the actual output. Our goal is to match the two end columns as best we can.

Assuming that all transitions are equi-probable, this modeling problem can be solved by computing the w_0 , w_1 and w_2 waveforms which solve the least-squares problem

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} w_{00} \\ w_{01} \\ w_{10} \\ w_{11} \end{bmatrix}, \text{ or } Ax = b. \quad (10)$$

The least-squares solution is

$$x = (A^T A)^{-1} A^T b, \text{ or } \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} w_{00} \\ w_{01} \\ w_{10} \\ w_{11} \end{bmatrix} \quad (11)$$

so that the error $b - Ax$ is

$$\begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} w_{00} \\ w_{01} \\ w_{10} \\ w_{11} \end{bmatrix} = \begin{bmatrix} -e \\ e \\ e \\ -e \end{bmatrix}, \quad (12)$$

where

$$e = \frac{(w_{01} + w_{10}) - (w_{00} + w_{11})}{4}. \quad (13)$$

As Eq. 12 shows, the error waveform e needs to be added to the output whenever the input changes, and subtracted from the output when the input stays constant. This requirement is implemented in the portion of Fig. 24 marked “nonlinear processing.”

Fig. 26 shows the simulated single-ended output waveforms over a 3-ns interval, as well as the waveforms computed using Eq. 11 and Eq. 13, for the circuit given in Fig. 25.

As described in the body of this paper, nonlinear switching dynamics causes the spectrum of the error waveform to be multiplied by the spectrum of the x sequence. Fig. 27 shows the Fourier transform of e (scaled by $f_{\text{CLK}} = 100$ MHz to account for the fact that this waveform is produced each clock period). The magnitude of the error waveform in the vicinity of 25 MHz is -101 dB.

Fig. 28 shows the spectrum of a 256-point, 33-level v sequence as well as the spectrum of the corresponding x sequence (assuming no mismatch-shaping is used). The signal spectrum has a -9 dBFS tone just to the right of band-center, whereas the switching noise spectrum contains a pair of in-band spurs, the larger of which has an amplitude of -6 dBFS. (Fortunately, when mismatch-shaped data is used the spectrum of x looks quite white. Unfortunately, when mismatch-shaping is used the power in the x signal is larger and does not decrease when signals are small. Imposing constraints on the mismatch-shaping logic which make the x signal more benign was explored in [10].) The signal tone will be attenuated by $\text{sinc}(1/3) = -1.7$ dB and so should have an amplitude of

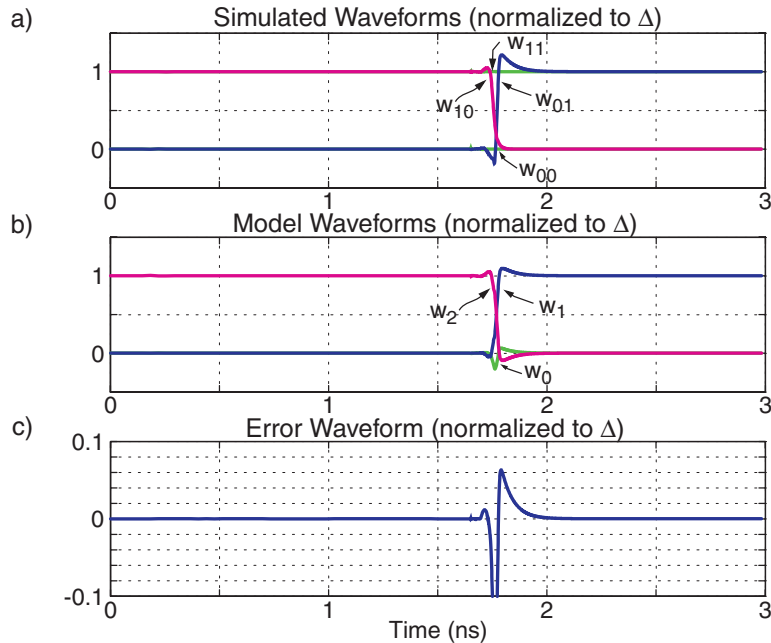


Fig. 26. Single-ended waveforms from the example circuit.

-11 dBFS at the DAC output, while the amplitude of the error spur should be $-6 - 101 = -107$ dBFS. Thus, in the absence of any differential cancellation, the DAC limits the SFDR to 96 dB. In order to reach 100 dB, only 4 dB of differential cancellation is needed.

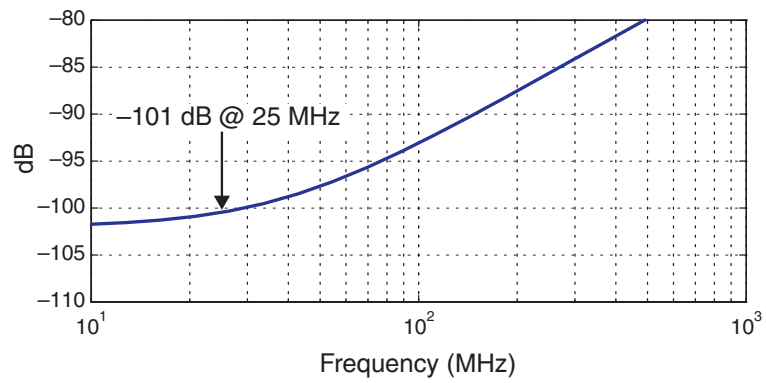


Fig. 27. Fourier transform of e .

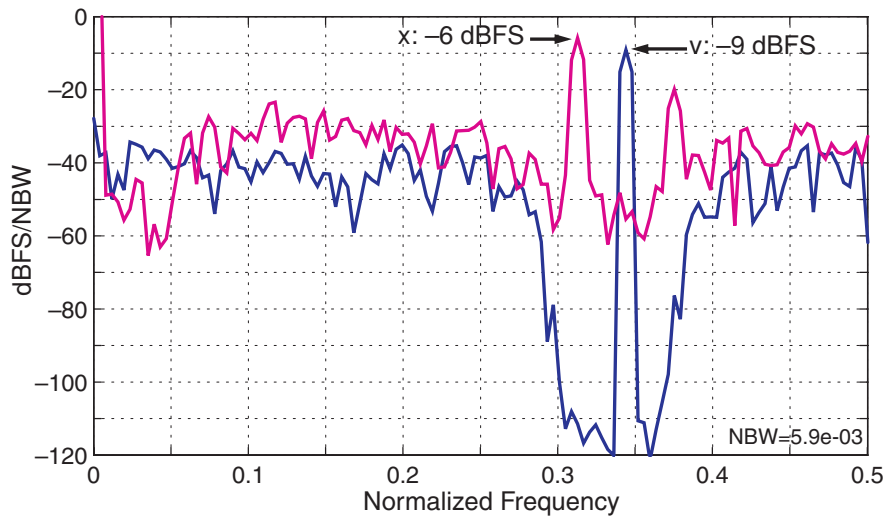


Fig. 28. v and x spectra for a 256-point 33-level data set, no mismatch-shaping.

References

- [1] T. H. Pearce and A. C. Baker, "Analogue to digital conversion requirements for HF radio receivers," *Proceedings of the IEE Colloquium on system aspects and applications of ADCs for radar, sonar and communications*, London, Nov. 1987, Digest No. 1987/92.
- [2] S. A. Jantzi, K. W. Martin and A. S. Sedra, "Quadrature bandpass $\Delta\Sigma$ modulation for digital radio," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 12, pp. 1935-1950, Dec. 1997.
- [3] G. Raghavan, J. F. Jensen, J. Laskowski, M. Kardos, M. G. Case, M. Sokolich and S. Thomas III, "Architecture, design, and test of continuous-time tunable intermediate-frequency bandpass delta-sigma modulators," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 1, pp. 5-13, Jan. 2001.
- [4] J. Van Engelen and R. Van De Plassche, *Bandpass sigma delta modulators— stability analysis, performance and design aspects*, Norwell, MA: Kluwer Academic Publishers 1999.
- [5] R. Schreier and G. C. Temes, *Understanding Delta-Sigma Data Converters*, John Wiley & Sons Inc., Hoboken, New Jersey, 2005.
- [6] L. J. Breems et al., "A cascaded continuous-time $\Sigma\Delta$ modulator with 67 dB dynamic range in 10 MHz bandwidth," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 72-73, Feb. 2004.
- [7] J. A. E. P. Van Engelen, R.J. Van De Plassche, E. Stikvoort and A. G. Venes, "A sixth-order continuous-time bandpass sigma-delta modulator for digital radio IF," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 12, pp. 1753 -1764, Dec. 1999.
- [8] R. Schreier, J. Lloyd, L. Singer, D. Paterson, M. Timko, M. Hensley, G. Patterson, K. Behel, J. Zhou and W. J. Martin, "A 10-300 MHz IF-digitizing IC with 90-105 dB dynamic range and 15-333 kHz bandwidth," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 12, pp. 1636-1644, December 2002.
- [9] W. Schofield, D. Mercer, and L.S. Onge, "A 16b 400MS/s DAC with <-80 dBc IMD to 300 MHz and <-160 dBm/Hz noise power spectral density," *IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pp. 126-127, Feb. 2003.
- [10] T. Shui and R. Schreier, "Mismatch shaping for a current-mode multibit delta-sigma DAC," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 331-338, March 1999.

HIGH-SPEED DIGITAL TO ANALOG CONVERTERS

Konstantinos Doris¹, Arthur van Roermund²

¹Philips Research Laboratories,
Eindhoven, The Netherlands

Technische Universiteit Eindhoven
Eindhoven, The Netherlands

Abstract

High-speed Digital-to-Analog Converters (DAC) are used in single- and multi-carrier communication applications because they simplify the number of mixing and filtering operations in the analog domain. In these applications, CMOS realizations that offer high-frequency linearity over broad bandwidths are required. The Current Steering architecture is the most suitable candidate, however, many nonlinear mechanisms limit its linearity: high sampling rates are possible, but good linearity is achieved only at small fractions of the Nyquist band, or at a large power and area penalty. Here, a rational design process will be described which demonstrates that high frequency linearity can be achieved at a low cost in power consumption and silicon area.

1. Introduction

High-speed Digital-to-Analog Converters (DAC), and especially CMOS implementations, are used in multi-carrier communication applications because they reduce signal processing operations in the analog domain. The ultimately flexible transmitter architecture is depicted in figure 1. In these applications, the DAC is required to process broadband signals with power spectral densities that span over several hundreds of MHz dependent on the application. To further simplify the subsequent low-pass filtering and to allow efficient implementation of pre-distortion techniques for high-data rate communications sampling rates multiple times higher than the actual transmitted signal bandwidth are required.

To make such type of transmitters possible, the DAC should maintain high linearity and low noise levels over this frequency range. Therefore, frequency dependent linearity specifications such as Spurious-Free-Dynamic-Range (SFDR) or Intermodulation-Distortion (IMD) are of primary importance. Typical values are more than 60dB SFDR over the complete bandwidth range.

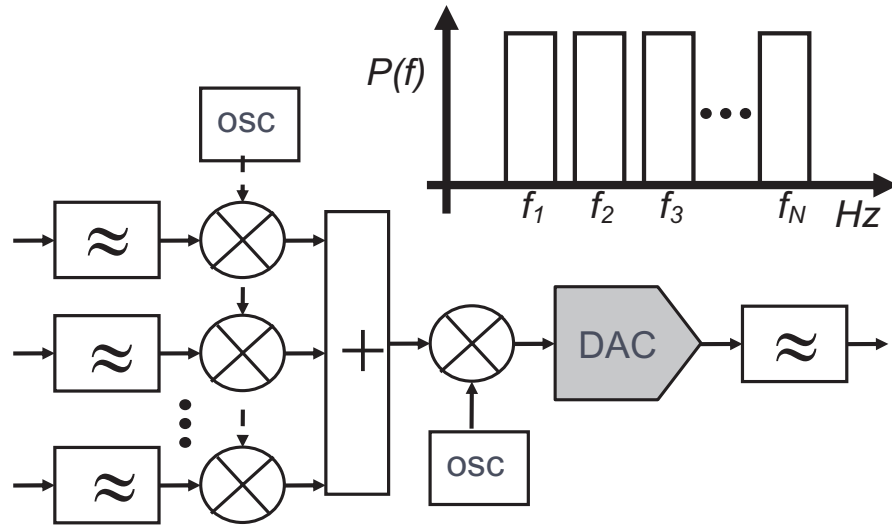


Figure 1: Multi-carrier transmitter with the DAC placed very close to the antenna.

The segmented Current Steering DAC (CS DAC) is the best architecture nowadays to deliver the combination of dynamic range and speeds at high frequencies. While these DAC's offer already hundreds of MHz of sampling rates in modern CMOS processes [1-5], it is often the case that they lack high frequency linearity, and especially without the use of output re-sampling circuits [6,7-8]: for these DACs, usually the SFDR starts at a very high value (e.g. 80dB) at kHz signal frequencies and then drops abruptly (e.g. 20-40dB/dec) as frequency exceeds a few MHz. Recently, the performance obtained in the CMOS DAC's [9-11] showed that good high-frequency linearity at sampling rates close or more than 1GHz can be achieved without the large costs in power consumption and silicon area associated with non-CMOS DAC's [8], often more than an order of magnitude larger.

In this work, the current status of CS DAC's will be given with respect to where, why, and how they fail in the context of high-speed operation. Designing a wide dynamic range high-speed DAC requires a thorough understanding and proper addressing of the error mechanisms that limit their performance. An example will be given on how this can be done at a low cost in power consumption and area describing main design aspects of a CMOS 12bit 500MS/s DAC [11-12].

Current Steering DAC's: where do they fail?

The basic architecture of a CS-DAC is shown in figure 2. It consists of:

- a current source network where the currents are generated in accordance to binary and thermometer coding;
- current switches that select which currents are to be added to form the analog representation of the input code;
- circuits that synchronize data with the clock before driving the switches (clock generator, clock distribution network, and clock elements);
- a digital binary to thermometer decoder for the decoding operations.

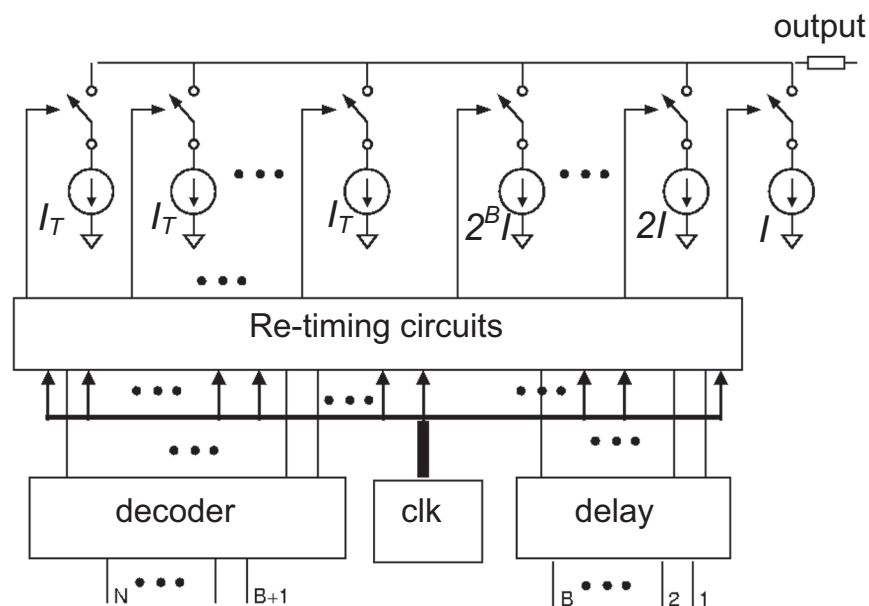


Figure 2: A thermometer-binary segmented CS DAC architecture.

Technological options for DAC realization include GaAs CMOS, Si Bipolar, SiGe Bipolar, and BiCMOS. CMOS is today's mainstream option to integrate the DAC as part of a larger VLSI system, therefore it is preferred in many cases. With respect to the design techniques being applied, the main characteristics of non-CMOS DACs are:

- Full differential current steering topology for every circuit in the signal flow. ECL levels for input and clock, small swing for the rest of the circuits.

- Partitioning in a few thermometric bits (3-5)
- Master-slave latches before the switches, latch buffers to filter switching noise of the latches and condition the data properly. Low swing differential signals everywhere, and especially at the switches.
- Speed optimized switched current cells.
- BJT cascoded resistors as current sources of the thermometric part for Bipolar DACs, transistors for GaAs, and R-2R ladders for the binary part.
- No output buffer, and direct connection of the current switches to the output.
- Re-sampling at the output in many occasions.
- Multiple supply networks (analog, digital) to separate interference of digital switching noise in critical analog circuits.
- DC accuracy achieved with inherent matching or post fabrication methods (e.g. laser trimming).

Such type of DAC's offer in most cases significant advantages with respect to high speed operation. Applications such as arbitrary waveform generators for testing equipment are the main drive to build Gsample/s DACs. Earlier examples include GaAs DAC's, e.g. for 14 and 16 bits [14,16], respectively reaching rates up to 2 Gsample/s. The most recent examples include a 10b 1.6Gsample/s GaAs DAC showing feasibility of conversion in the second Nyquist frequency range, a 15 bit 1.2 GSample/s [8] and a 6 bit 22 GSample/s [15], both implemented with SiGe BiCMOS processes.

A typical CMOS implementation is shown in figure 3. The main characteristics of CMOS DAC's are:

- Single-ended CMOS-logic signals for all circuits in the signal flow and the clock, and differential signals for the current cell.
- Partitioning between a medium to large thermometer part (5-8) and a relatively small binary part.
- Single latches, which are based on cross-coupled CMOS inverters and reduced-swing drivers with modified complementary data crossing point.
- Differential current switches, and cascoded current sources, which are usually constructed with transistors.
- Re-sampling circuits (e.g. Track-and-Hold) at the outputs in some occasions.
- Calibration and switching sequences to deal with DC matching errors.

Sampling rates of CMOS DAC's are lower than their non-CMOS counterparts. Recently a couple of DAC's with more than 1Gsample/s were presented in [3] and [10] for 10 and 14 bits, respectively while there exist plenty examples with sampling rates of several hundreds MHz for 12-16 bits [1,2,4,5,7,9-11].

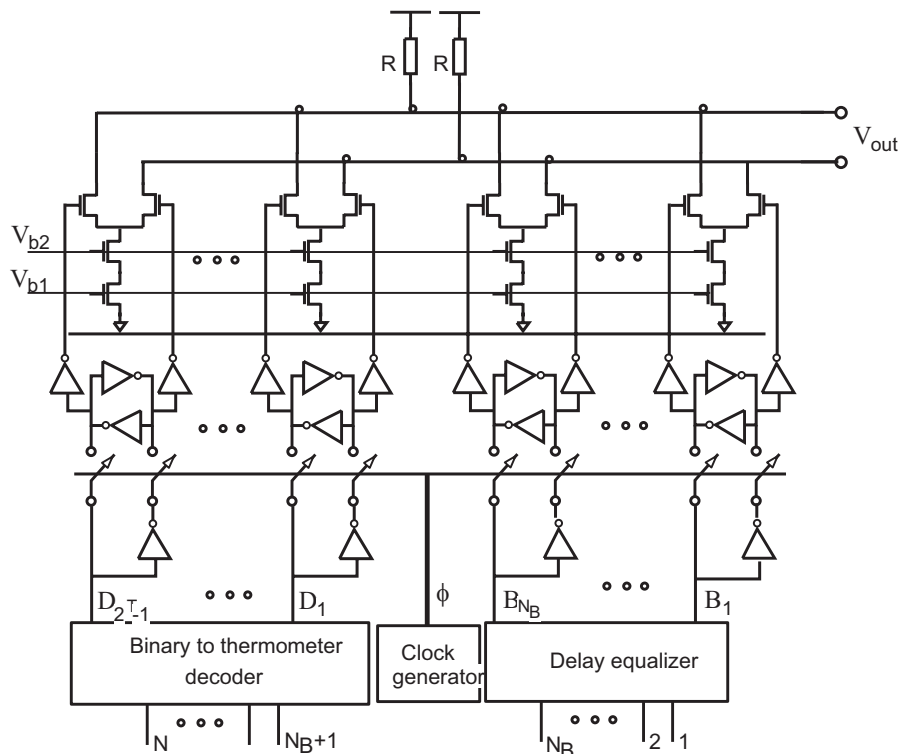


Figure 3: A thermometer-binary segmented architecture.

Achieving high sampling rates and many bits is not identical with good high frequency linearity. The maximum sampling rate of a DAC in practice indicates the maximum sampling rate at which digital logic functions still operate properly. It says nothing of the quality of the analog signals being converted. Similarly, the number of bits is merely an indication of static linearity, which is less of interest at high frequencies. In [16] a 12 bit DAC (with 14 bit static accuracy) is reported at 1 Gsample/s which delivers a mere 52 dB SFDR at just 100 MHz, and 62 dB using an output sampler. In [14] despite the 2 Gsample/s rates offered by a 14 bit GaAs DAC, only 58 dB are obtained at 62 MHz signal frequencies with 0.75 Gsample/s rate. However, this situation seems to be changing; the DAC presented in [8] showed that despite a linearity far less than 15 bit, SFDR values of 65 dB can be reached up to 600MHz of signal frequencies. Notice however, the cost of 6 Watts and roughly 30 mm² that were used for the cause of obtaining such performance.

Such levels are prohibitive for most applications other than measurement equipment. Obtaining high frequency linearity was shown to open paths in exploiting the second Nyquist frequency band for even less signal processing operations at the analog domain [17].

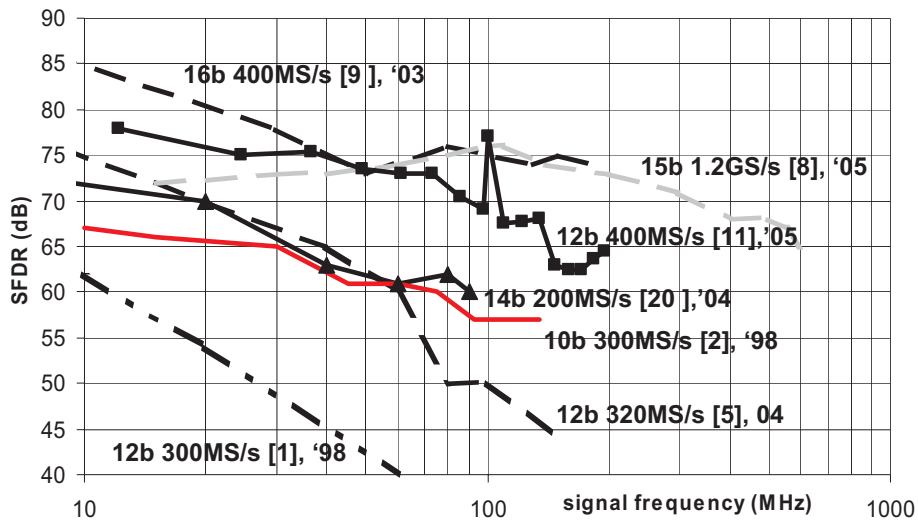


Figure 4 : Representative SFDR vs frequency plots from recent literature.

A similar situation existed until very recently for CMOS DAC's. Figure 4 shows representative plots of the SFDR situation that applies nowadays. The SiGe DAC [8] was added for comparison (the grey line). Notice the large contrast between the number of bits and the SFDR at high frequencies for all DAC's. With the exception of a few, it can be stated that nowadays most DAC's suffer from a rapid linearity degradation as the frequency increases for more than 1/10 of the sampling rate. Clearly, this is related with the dynamic behavior of these circuits. Yet, this situation is also improving as it can be seen from the figure, e.g. [9,11] and without the huge penalties in power consumption and area compared to non CMOS DAC's: approximately 400 and 200mW where spent in [9] and [11], respectively. In [10], 65 dB SFDR where reported for up to 260MHz from a 1.4 Gsample/s CMOS DAC at a total power of 400mW. Therefore, basic questions appear with respect to which are the limitations at high frequencies, what exactly is their impact, and how can they be prevented or addressed such that high frequency linearity can be obtained with CMOS at a reasonable power and area budget.

High frequency linearity limitations

Although simple at first sight, the CS-DAC exhibits many nonlinear error generation mechanisms: that is, this architecture is not characterized by one error mechanism that can be clearly identified as a limiting factor at higher frequencies but by many mechanisms, which are coupled with each other at signal and circuit level. Thus, when one is trying to adopt a circuit solution and optimize the design to solve one error mechanism, he, or she, often influences negatively other mechanisms. The most important error generation mechanisms are:

1. Process mismatch in the current sources.
2. Output resistance and capacitance modulation by the input signal.
3. Nonlinear switch operation that creates spikes in their source node.
4. Charge feedthrough from the switch control nodes to the output and from the common source switch node to the biasing nodes.
5. Local variations of the individual current pulses due:
 - a. Mismatch in current and clock switches, latches and their drivers.
 - b. Mismatch between the decoder gates that drive the latches, and the associated interconnect RC's.
 - c. Clock skew between clocking locations due to interconnect length differences, transmission line effects, etc.
 - d. Unequal interconnect length at the output summing node network.
6. Interference due to feed-through of switching signals on the biasing nodes.
7. Power supply and substrate related effects.
8. Clock (timing) jitter.

To design CS-DAC's with good high frequency linearity, these error mechanisms must be addressed properly, and preferably independently of each other. This requires their dependencies with signal, architecture, circuit and layout parameters to be well understood. To improve their understanding, an error classification has been proposed in [12-13] in accordance to a set of basic error properties. Special cases of error mechanisms were analyzed hierarchically. In this way, errors are not examined only as a separate case independently of each other but also in view of their common and differences.

A major distinction is between amplitude and timing errors; the impact of timing errors scales up with frequency, but for each class member this is determined by its other properties. Most problems relevant for high frequencies are timing errors.

Another important distinction is between local and global timing errors. Local timing errors appear as waveform differences between the unit current pulses, e.g.

in a sample-to-sample transition each unit current pulse has a different delay from the rest. The signal quality is determined by the average of all local timing errors involved in the sample-to-sample transition. This gives a smooth linearity degradation of the signal quality vs. its frequency of roughly 10dB/dec. Local errors can be further categorized to spatially random (mostly related to mismatch) and deterministic (related to interconnect length different) dependent on their spatial characteristics. In figure 4, the DAC's [2,8,9,11] seem to be characterized by such errors, at least, for a significant part of the frequency points. Furthermore, these DAC's limited by such behavior seem to have the best overall performance in their year context.

Global timing errors appear in the same way for all current pulses generated for a sample-to-sample transition. All current pulses have the same shape with respect to each other during the code transition, but this shape depends on the sample-to-sample transition. Global errors are mostly associated with global nodes: the clock node, the biasing and supply nodes (e.g. power supply and substrate bounce, switching interference at the biasing nodes), the output node (e.g. nonlinear output capacitance), etc. The error usually depends on the signal values, signal derivative, etc., thus they cause linearity degradation that scales quite often with 20-40dB/dec. Global timing errors can be further divided in random and deterministic. Evaluation of a large set of data from measurement plots published in open literature indicates that performance in literature is usually dominated by global timing errors. Therefore, it is of primary importance to reduce these types of errors to the minimum in order to obtain good levels of high frequency linearity. The DAC's in [2,8,9,11] have succeeded at minimizing global errors. Elimination or reduction of global errors needs to be done in ways that do not increase local timing errors. In the following sections, such a design approach will be described.

3. Dealing with high frequency linearity: a design example

In this section some important aspects of a design example will be given to demonstrate a structured way of dealing with error mechanisms that limit high frequency linearity. More details of the described IC are given in [11-12]. It is a 12bit 500MS/sec Current Steering DAC realized in a CMOS 0.18 μm .

The aim in this design is to avoid the generation of errors that lead to non-linearity and therefore attempts to use techniques that suppress their effect in the output signal are avoided. The following principles are followed: (a) prevention of error mechanisms is preferred than suppression, or compensation; (b) global error

mechanisms are eliminated, or translated to local ones, which are easier to cope with; (c) simple techniques are applied, e.g. no calibration or error control loops.

3.2 Architecture

The architecture employed is fully differential and consists of 6 thermometer and 6 binary code bits (fig. 5). Current Mode Logic (CML) is used for all logic circuits. Low swing differential input signals control low swing buffers of the DAC and feed the 6b CML thermometer decoder. A delay equalizer for the least significant bits ensures that all the latches capture the data synchronously. The master-slave latches provide synchronized data to the switched current cells (SI). The differential output is directed to an off-chip resistive load. Two-level local biasing is applied. The background of these choices is explained in this section.

3.2.1 Signaling and Logic

For low mismatch based local timing errors, and low global timing errors due to supply and substrate noise there are opposing demands. The former calls for fast switching signals and many thermometer bits (many elements), the latter for slow switching signals and a few thermometer bits, primarily due to strong disturbances generated by CMOS logic circuits. At the same time, for low deterministic local timing errors due to interconnect length differences, a few thermometer bits are also required. Therefore, a tradeoff appears on the choice of proper segmentation and on the allowed steepness on the signals during transitions.

The high common mode noise rejection achieved with differential signals [19] in combination with the low supply disturbance generation offered by low swing Common Mode Logic (CML) decouples local and global timing problems and facilitates better focus on each error class separately. In other words, the steepness of the switching signals can be increased to deal with timing errors caused by mismatch, and the number of thermometer bits to help in averaging errors better, without the subsequent penalties of switching disturbances as it happens with CMOS logic. The constant power consumption of CML compared to CMOS logic's dependency with frequency is another advantage as well.

At the analog output side, differential signaling reduces substantially the errors due to nonlinear settling and DAC output impedance because the distortion generated by these problems is mainly of second order.

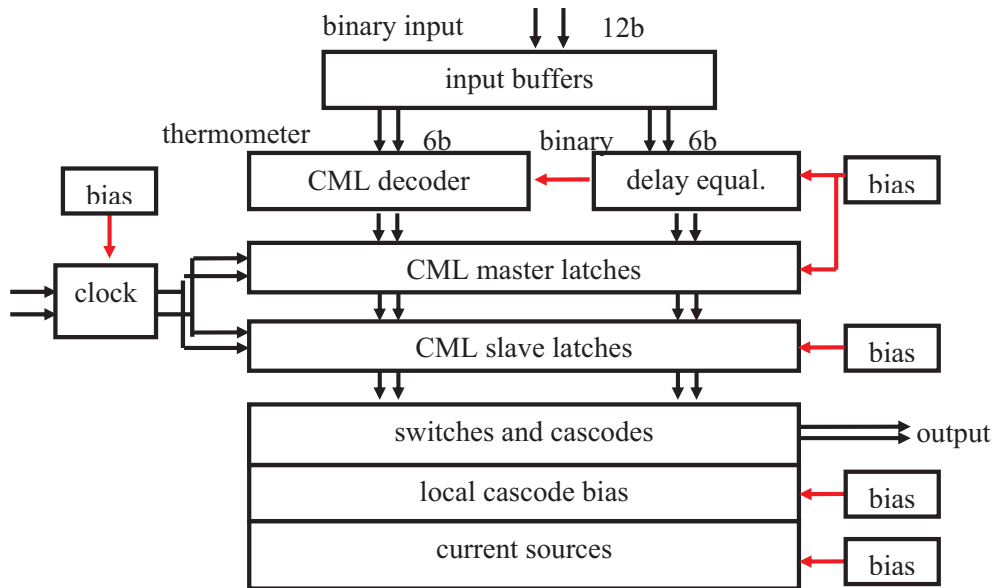


Figure 5: The DAC architecture.

3.2.2 Power supplies and biasing

CML prevents the dominant portion of data-dependent supply disturbances, however, remaining disturbances need attention as well. In addition, the biasing voltages for CML need to be shielded properly. To prevent error generation from these disturbances, separate power supplies and biasing nodes are used for the clock, decoder and master-latches, slave-latches and drivers, and current sources. This localizes disturbances in each type of circuits (latches separately from decoder, etc.) instead of distributing it globally.

With each circuit type biased separately, there remains still the issue of interference within the same circuit type (e.g. from latch to latch). To reduce further these error mechanisms, additional local prevention techniques are used. Local decoupling capacitance per latch was added. For the current cells, where a similar problem appears from the coupling of the switch tail node spikes to the global biasing nodes, each individual cell has a local bias at the source cascode transistor. Finally, multiple of pins per supplies are used in the package to reduce the inductance of the bonding wires interfacing the on-chip to the off-chip supplies.

In summary, CML logic, differential signals, and several localization techniques were applied to translate global error mechanisms to local. In this way, strong global error related nonlinear distortion is avoided, while segmentation and signal steepness is to be used as a degree of freedom for low local timing errors.

3.2.3 Thermometer/binary partitioning

The 12 binary input bits were partitioned in 6 thermometer (MSB's) and 6 binary bits (LSB's), represented by N_T and N_B , respectively. This choice was based on local timing errors, speed, area, and power consumption. Both random errors, e.g. matching at the current cell, driver and clock switches, and deterministic ones, e.g. clock and output interconnect differences, were taken into account.

For local random errors, according to the calculations made in [12-13] for 74 dB of signal to total distortion power ratio at the Nyquist rate for a DAC operating at 200 and 400MS/sec, a one sigma spread of 2.8 and 1.4 psec is required for 6 thermometer bits. For 8 bits, the calculations indicate 5.6 and 2.8 psec, respectively. Large N_T provides better averaging of local errors, thus better performance, or more relaxed timing specifications.

Transistor level analysis in a circuit chain consisting of a latch, driver and current cell indicates some additional effects. For a fixed power, the impact of the errors contributed by latches, drivers and switches scales differently with N_T . The impact of mismatch errors at the drivers and current cell switches are reduced when N_T increases but the impact of the latches increase, because the additional number of latches and associated extra interconnect reduces the clock slope, thus increases the errors faster than what averaging improves. Deterministic local errors influence specifications as well. Averaging still applies here, but the local error magnitude increases faster with increased N_T than the averaging benefits do. Consideration of all the above factors with a design target for random errors around 3 psec spread resulted in a decision to use 6 thermometer bits.

3.3. Building block design

The implementation of the basic building blocks will be described in this section.

3.3.1 Switched current cell

The schematic of the switched current (SI) cell is shown in fig. 6. The current sources were sized to the 12b level for random mismatch. Each current source is

partitioned into four sub-arrays and each sub-source is biased locally to reduce gradient effects. Layout techniques, developed in [12], were applied to reduce systematic mismatch effects to the 12b level.

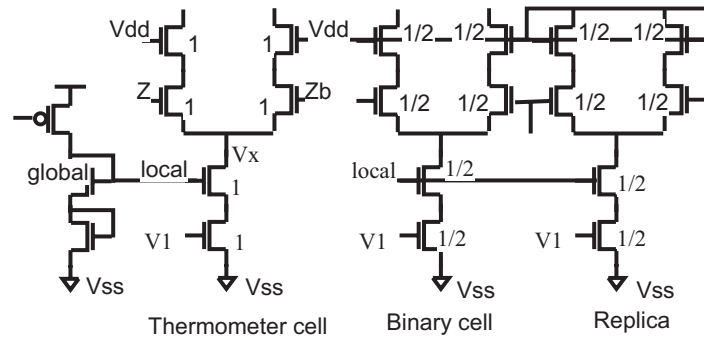


Figure 6: Thermometer and binary SI cells

The conventional addition of inactive capacitance to the binary switches to match their delay to the thermometer ones leads to timing differences between them: inactive capacitances are linear whereas active capacitances are not.

The replica cells shown in fig. 6 ensure that all transients have an identical shape. The use of differential output signals reduces the output impedance requirements for each SI cell significantly compared to single ended output signals. Therefore, there is more design freedom for the switches, cascodes and current source. A single-cascode boosts the output impedance, it shields the current source from spikes and reduces spike interference to the bias lines. Local source cascode biasing [14] was applied to isolate the global bias line from the collective interference due to spikes of the common switch nodes (V_x in fig. 4) of every switching current cell, transforming once more the global error mechanism to local.

Charge feed-through is reduced using local switch cascodes and low swing switch driver signals. Feed-through compensation [10,18] is avoided because it doubles the switching logic, the supply disturbances, the clock load and the common switch node capacitance, thus increasing errors and power consumption. Data-dependent variations on the common switch node [9] are reduced to acceptable levels by switch cascodes. The low data swing inherently rises the switching cross-point and therefore reduces the spike on the common switching node.

Significant attention was paid to design switches that contribute less than 1 psec spread of random local timing errors. The dimensions of the switches were selected based on circuit optimization on the basis of the combined effect of the switch mismatch, its gate capacitance, and the self-capacitance of the driver and the corresponding interconnect [12].

3.3.2 Master-Slave Latch

The data signals generated at the output of the decoder are subject to the effects of different logic depths, device process mismatch, interconnect length differences, cross-coupling. These effects cause large waveform shape variations and delays. Latches and drivers receive these widely different waveforms and generate clear, identical and very accurately synchronized ones to drive the SI cells.

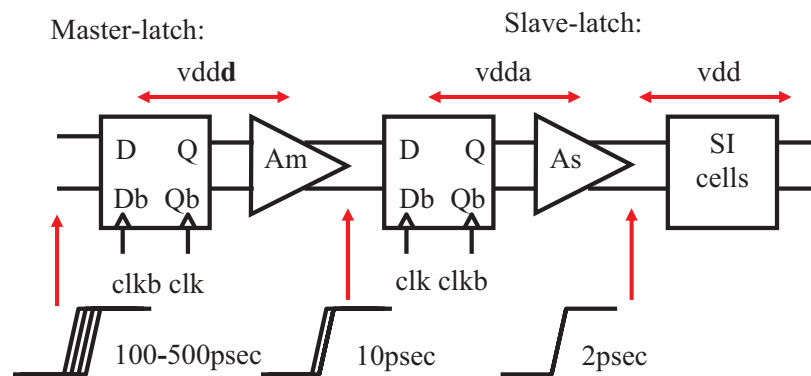


Figure 7: Block level schematic of the Master-slave latch.

High-speed CMOS DACs use single latches based on variations of the cross-coupled CMOS inverter. CMOS inverters are used as drivers for the current switches. As sampling rates increase, the many tasks assigned to one single latch become more difficult to accomplish. Moreover, this topology suffers from several local timing errors origins related that appear during transitions [12].

In this design, a CML master-slave (MS) latch is used (fig. 7 and 8) for its low swing differential operation, low power supply disturbance, and low power consumption at high speed. This topology proves capable of low local timing errors as well. The tasks assigned to the latch are divided in two latches to deal with them more efficiently. The master latch receives decoded data and removes delays,

spikes, etc. Clean and well-synchronized data (<10 psec) are passed differentially to the slave latch at the next clock phase. The slave latch and driver refer now the data on the local cleaner power supply, attenuate any remaining data-dependent effects and provides precise timing (1-2 psec spread), steep edges, and low swing.

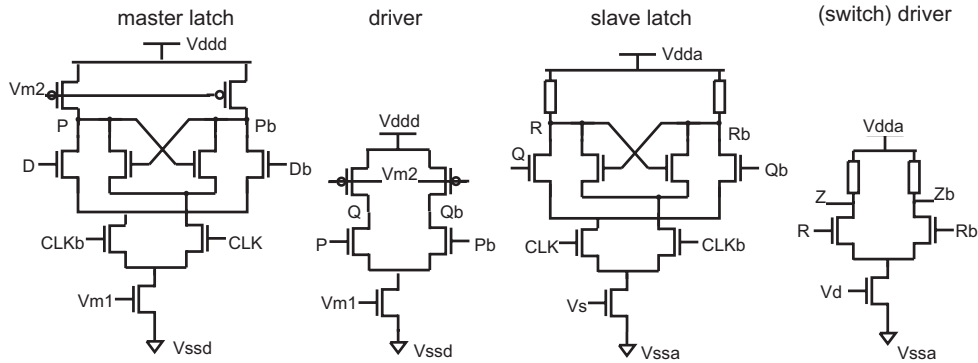


Figure 8: Circuit schematic of the master-slave latch.

Significant attention is paid to all sources of local timing errors such as interconnect geometry, slopes, ratios of driving and load capacitances of latches, mismatch parameters, etc.

3.4 Layout

Layout design plays a crucial role in the performance of the converter. A lot of attention has been paid in realizing a well-structured layout. All circuits layouts have been made manually, and many circuits have been extracted and back-annotated for simulations. The main aspects of the layout will be described here.

The layout can be seen with the aid of the die photo in fig. 9. On the right side of the figure the arrays of the input buffers, the decoder, the MSB/LSB delay equalizer and the master latches are located (region A). The slave latches, drivers, cascoded switches, and the current source cascodes are located in region B. Left of region B are the Vdda/Vssa rails and their decoupling, the local cascode biasing circuits, the output interconnects, and other biasing wires (region C). The foremost left part of the figure shows the current source array and its biasing circuit (region D). The clock buffer is located at the top of region B

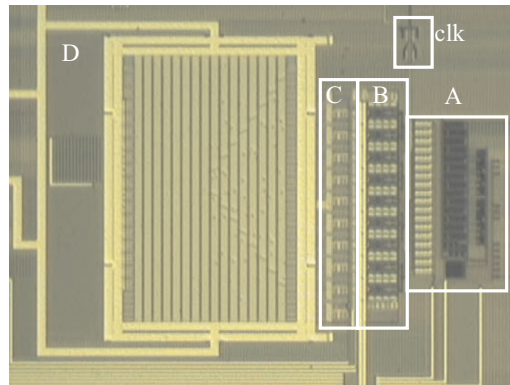


Figure 9: Die photo.

Data flow from the left of the picture to the right. The differential clock network splits in two parts for slave and master latches, respectively. A combination of a primary and secondary binary trees connected with a rail to average errors is used for the slave clock. The output currents are summed with binary trees and a rail.

3.5 Measurements

The DAC is realized in a CMOS 0.18 μm single poly 5 metal-layer process and placed in a LQFP 80-pin package. Measurements of a 15mA full scale current are shown in the pictures. Fig. 10 indicates an INL of 1LSB and a DNL less than 0.6LSB's. The SFDR vs. (normalized) signal frequency is shown in fig. 11. Starting close to 80dB, for sample frequencies (f_s) up to 350MS/s, the SFDR is larger than 70dB up to 123MHz and 66dB close to Nyquist. At 400MS/s the SFDR stays higher than 70dB up to 100MHz and 65dB at Nyquist. At 500MS/s the SFDR drops with a 20dB/dec up to 120MHz down to 60dB and stays noticeably constant up to Nyquist. Between 300-400MS/s and for f/f_s between 0.1-0.35 local deterministic timing errors limit the performance: the smooth degradation of 10dB/dec of signal frequency and the linear drop with f_s are characteristic of this error class. Beyond 400MS/s decoding errors change the DAC behavior. The maximum f_s with roughly 60dB at full Nyquist is 500MS/s. Even at 600MSample/s (not shown), the performance is still characterized by a lower, yet constant SFDR.

The DAC consumes 216mW from a 1.8V supply (160mW without the clock buffer) independent on frequency and occupies 1.13mm². Table 1 shows the

performance summary. An SFDR comparison with other recent DAC's was already in Figure 4.

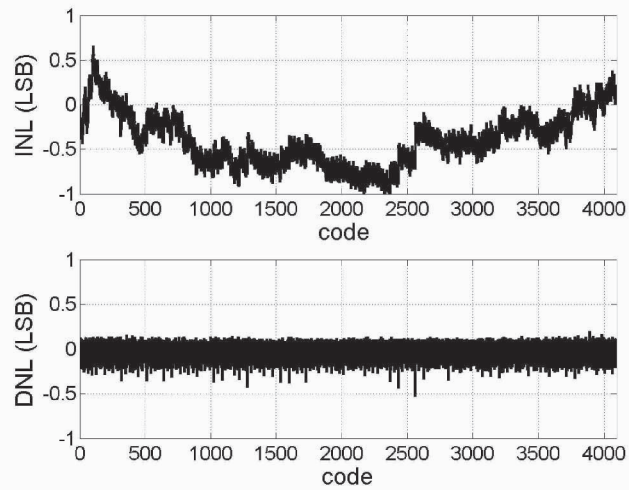


Figure 10: Measured INL and DNL.

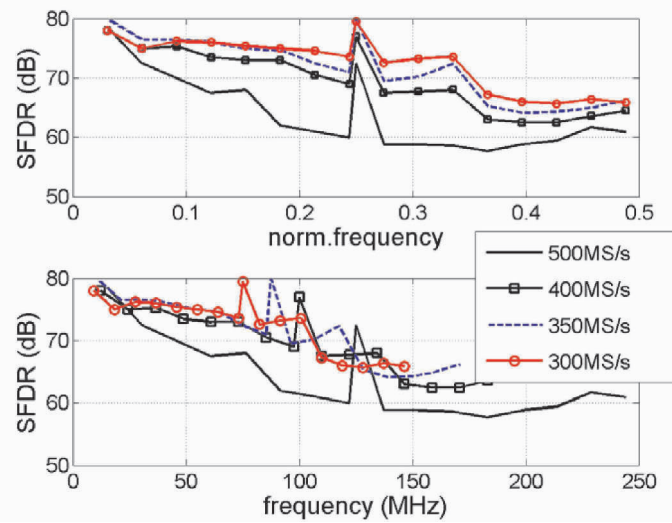


Figure 11: Spurious Free Dynamic Range vs. (normalized) signal frequency.

4. Conclusions

High speed Digital to Analog Converters are realized with the Current Steering technique. Such DAC's can deliver very high sampling rates and many bits, but it suffers from many nonlinear error mechanisms at high frequencies that limit linearity. Significant progress in increasing high frequency linearity has been made the last few years both for CMOS and non CMOS processes.

The main difficulty with Current Steering DAC's is that the nonlinear error generation mechanisms are coupled with each other at circuit and architectural level. Very often one error mechanism is reduced with circuit design practices at the expense of increasing another.

An example of a rational design approach was presented to deal with the complexity of dealing with each error mechanism independently from the others. The approach is based on classification of error mechanisms and simplicity of design solutions. Its efficiency was demonstrated with a wide-bandwidth, high dynamic range 12bit 500MS/s CMOS Current Steering DAC, which was realized in a 0.18 μ m CMOS process. Main design aspects of this IC were explained.

Such an approach facilitates the shift of DAC operations further towards the antenna side of a transmitter allowing larger degrees of digital transmitter architectures and more versatile digital frequency synthesis.

Table 1: Performance summary

Process info	CMOS 0.18 μ m, 1.8V, 5M1P
Sample rate	350-600 MSample/sec
Resolution	12bits
INL/DNL	1LSB/0.6LSB
SFDR @350MS/s	80-65dB from 10 to 175MHz
SFDR @400MS/s	78-64dB from 10 to 200MHz
SFDR @500MS/s	78-58dB from 10 to 250MHz
Power at 15mA	216 mW
Area	1.13mm ²

References

- [1] J.Bastos et.al., "A 12-Bit Intrinsic Accuracy High-Speed CMOS DAC," IEEE Journal of Solid State Circuits, vol. 29, no. 10, pp. 1180-1185, Oct. 1994.
- [2] C. Lin and K. Bult, "A 10-b, 500-Msample/s CMOS DAC in 0.6mm²," IEEE Journal of Solid State Circuits, vol. 33, no. 12, pp. 1948-1958, Dec. 1998.
- [3] A.van den Bosch et.al., "A 10-bit 1-Gsample/s Nyquist Current-Steering CMOS D/A Converter," IEEE Custom Integrated Circuit Conf. (CICC), 2000, pp. 265-268.
- [4] A.van den Bosch et. al., "A 12b 500Msample/s Current Steering CMOS D/A Converter," IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, 2001, pp. 366-367.
- [5] K. O'Sullivan, et al., "A 12-bit 320-Msample/s Current Steering CMOS D/A Converter in 0.44 mm²," IEEE Journal of Solid State Circuits, vol. 39, no. 7, pp. 1064-1072, Jul. 2004.
- [6] K. Khanoyan et.al., "A 400-MHz, 10bit charge domain CMOS D/A converter low-spurious signal frequency synthesis", Analog Circuit Design, Kluwer Academic Publishers, 2002.
- [7] A. Bugeja and B.S. Song, "A Self-Trimming 14-b 100-MS/s CMOS DAC", IEEE Journal of Solid State Circuits, vol. 35, num. 12. pp. 1841-1852, Dec. 2000.
- [8] B. Jewett et. al., "A 1.2 GS/s 15b DAC for Precision Signal Generation," IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 110-112, Feb. 2005.
- [9] W. Schofield et al., "A 16b 400MS/s DAC with <-80dBc IMD to 300MHz and <-160dBm/Hz Noise Power Spectral Density", IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, Feb. 2003.
- [10] B. Schafferer et al., "A 3V CMOS 400mW 14b 1.4GS/s DAC for Multi-Carrier Applications", IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 360-361, Feb. 2004.
- [11] K. Doris et. al., "A 12b 500MS/s DAC with >70dB SFDR up to 120MHz in 0.18um CMOS", IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 116-118, Feb. 2005.
- [12] K. Doris, "High-speed D/A Converters: from Analysis and Synthesis Concepts to IC Implementation", Ph.D. Thesis, Technical University Eindhoven, Sept. 2004.
- [13] K. Doris et.al., "High-speed Digital to Analog Converter issues with applications to Sigma-Delta Modulators," Analog Circuit Design, Kluwer Academic Publishers, 2002, ISBN 1-4020-7216-3.

- [14] F.G. Weiss and T.G. Bowman, "A 14-Bit, 1 Gs/s DAC for Direct Digital Synthesis Applications", Gallium Arsenide Integrated Circuit Symposium, 1991, pp. 361-364.
- [15] P. Schvan, et.al., "A 22GS/s 6b DAC with Integrated Digital Ramp Generator", IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 122-123, Feb. 2005.
- [16] K-C. Hsieh, et.al., "A 12-bit 1-Gword/s GaAs Digital-to-Analog Converter System", IEEE Journal of Solid State Circuits, vol 22, no. 6, pp. 1048-1054, Dec. 1987.
- [17] M-J. Choe, et. al., "A 1.6GS/s 12b Return-to-Zero GaAs RF DAC for Multiple Nyquist Operation", IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 112-114, Feb. 2005.
- [18] S. Park et.al., "A Digital-to-Analog Converter Based on Differential-Quad Switching," IEEE Journal of Solid State Circuits, vol. 37. no 10, pp. 1335-1338, Oct. 2002.
- [19] M. Ingels and M. Steyaert, "Design strategies and decoupling techniques for reducing the effects of electrical interference in mixed-mode IC's", IEEE Journal of Solid State Circuits, vol 32, no. 7, pp. 1136-1141, July 1997.
- [20] Q. Huang, et.al., "A 200MS/s 14b 97mW DAC in 0.18/spl mu/m CMOS", IEEE International Solid State Circuits Conf. (ISSCC) Dig. Tech. Papers, pp. 364-366, Feb. 2005.

Part II: Design Methodology and Verification for RF and Mixed-Signal Systems

Complexity of analog design increases. The required accuracy goes up, technology offers new challenges, but also new threats, and system complexity rises. The translation from system specs to subsystem specs becomes cumbersome, with the non-linear relations involved. The reverse is also true: the number of relevant parasitic and impairments from technology increases, and these in turn should be translated to their impact at the higher levels of the system. How can we efficiently simulate circuits embedded in larger and increasingly non-linear systems; or systems with very large dynamic ranges in time constants, in signal domain, etc.? And, having designed, how can we test tomorrow's ICs with embedded circuits that cannot easily be accessed, and with such a complexity that complexity of testing becomes a real problem? How is that exploding complexity controlled?

At the same time, the pressure from business on shortening design time increases, facing us with a real conflict to be solved; but how?

Abstraction and reuse are key words here. They can be translated to issues like design methodology, reuse, and automation; platform-based design; high-level modelling and model generators; new generations of simulators; simulation of test environments; etc. One thing is clear: the disciplines involved grows.

This, in short, is the topic of this part. People from different disciplines, from businesses and universities from a variety of countries, all involved in design methodology and verification for RF and mixed-signal systems, have been brought together, to highlight the problem from different angles of view.

The question the analog circuit designer faces is: become a dinosaur (and thus die) by staying away from developments around him, or getting strongly involved in all those developments, being completely occupied with the problems arising there, leaving no time for circuit design? The truth will be somewhere in between. Getting informed of the various developments, however, is a first prerequisite to make the right choices.

Arthur van Roermund

DESIGN METHODOLOGY AND MODEL GENERATION FOR COMPLEX ANALOG BLOCKS

Georges Gielen
ESAT-MICAS, Katholieke Universiteit Leuven
Kasteelpark Arenberg 10
3001 Leuven, Belgium
gielen@esat.kuleuven.ac.be

Abstract

An overview is presented of the design methodology and related modeling needs for complex analog and RF blocks in mixed-signal integrated systems (ASICs, SoCs, SiPs). The design of these integrated systems is characterized by growing design complexities and shortening time to market constraints. Handling these requires mixed-signal design methodologies and flows that include system-level architectural explorations and hierarchical design refinements with behavioral models in the top-down design path, and detailed behavioral model extraction and efficient mixed-signal behavioral simulation in the bottom-up verification path. Techniques to generate analog behavioral models, including regression-based methods as well as model-order reduction techniques, are described in detail. Also the generation of performance models for analog circuit synthesis and of symbolic models that provide designers with insight in the relationships governing the performance behavior of a circuit are described

1. Introduction

With the evolution towards ultra-deep-submicron and nanometer CMOS technologies [1], the design of complex integrated systems, be it ASICs, SoCs or SiPs, is emerging in consumer-market applications such as telecom and multimedia, but also in more traditional application domains like automotive or instrumentation. Driven by cost reduction, these markets demand for low-cost optimized and highly integrated solutions with very demanding performance specifications. These integrated systems are increasingly mixed-signal designs, embedding on a single die both high-performance analog or mixed-signal blocks and possibly sensitive RF frontends together with the complex digital circuitry

(multiple processors, a couple of logic blocks, and several large memory blocks) that forms the core of most electronic systems today. In addition to the technical challenges related to the increasing design complexity and the problems posed by analog-digital integration, shortening time-to-market constraints put pressure on the design methodology and tools used to design these systems.

Hence the design of today's integrated systems calls for mixed-signal design methodologies and flows that include system-level architectural explorations and hierarchical design refinements with behavioral models in the top-down design path to reduce the chance of design iterations and to improve the overall optimality of the design solution [2]. In addition, to avoid design errors before tape-out, detailed behavioral model extraction and efficient mixed-signal behavioral simulation are needed in the bottom-up verification path. This chapter presents an overview of the model generation methods used in this context.

The chapter is organized as follows. Section 2 addresses mixed-signal design methodologies and describes techniques and examples for architectural exploration and top-down hierarchical design refinement. Section 3 describes analog and mixed-signal behavioral simulation and gives an overview of techniques to automatically generate analog behavioral models, including regression-based methods as well as model-order reduction techniques. Section 4 then describes the generation of performance models for analog circuit synthesis. Finally, section 5 presents methods to generate symbolic models that provide designers with insight in the relationships governing the performance behavior of a circuit. Conclusions are drawn in section 6.

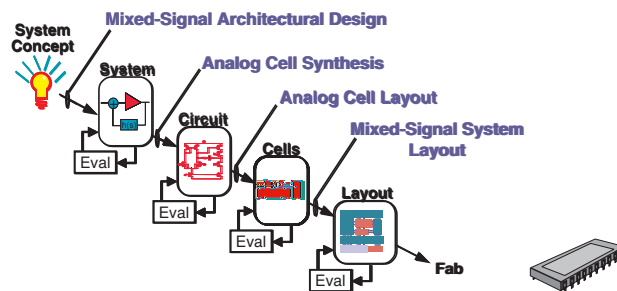


Fig. 1. Top-down view of the mixed-signal IC design process.

2. Top-down mixed-signal design methodology

The growing complexity of the systems that can be integrated on a single die today, in combination with the tightening time-to-market constraints, results in a growing design productivity gap. That is why new design methodologies are being developed that allow designers to shift to a higher level of design

abstraction, such as the use of platform-based design, object-oriented system-level hierarchical design refinement flows, hardware-software co-design, and IP reuse, on top of the already established use of CAD tools for logic synthesis and digital place & route. However, these flows have to be extended to incorporate the embedded analog/RF blocks.

A typical top-down design flow for mixed-signal integrated systems may look as shown in Fig. 1, where the following distinct phases can be identified: system specification, architectural design, cell design, cell layout and system layout assembly [2,3]. The advantages of adopting a top-down design methodology are:

- the possibility to perform system architectural exploration and a better overall system optimization (e.g. finding an architecture that consumes less power) at a high level before starting detailed circuit implementations;
- the elimination of problems that often cause overall design iterations, like the anticipation of problems related to interfacing different blocks;
- the possibility to do early test development in parallel to the actual block design; etc.

The ultimate advantage of top-down design therefore is to catch problems early in the design flow and as a result have a higher chance of first-time success with fewer or no overall design iterations, hence shortening design time, while at the same time obtaining a better overall system design. The methodology however does not come for free and requires some investment from the design team, especially in terms of high-level modeling and setting up a sufficient model library for the targeted application domain. Even then there remains the risk that also at higher levels in the design hierarchy low-level details (e.g. matching limitations, circuit nonidealities, layout effects) may be important to determine the feasibility or optimality of an analog solution. The high-level models used therefore must include such effects to the extent possible, but it remains difficult in practice to anticipate or model everything accurately at higher levels. Besides the models, efficient simulation methods are also needed at the architectural level in order to allow efficient interactive explorations. The subjects of system exploration and simulation as well as behavioral modeling will now be discussed in more detail.

2.1. System-level architectural exploration

The general objective of analog architectural system exploration is twofold [4,5]. First of all, a proper (and preferably optimal) architecture for the system has to be decided upon. Secondly, the required specifications for each of the blocks in the chosen architecture must be determined, so that the overall system meets its requirements at minimum implementation cost (power, chip area, etc.).

The aim of a system exploration environment is to provide the system designer with the platform and the supporting tool set to explore in a short time different architectural alternatives and to take the above decisions based on quantified rather than heuristic information.

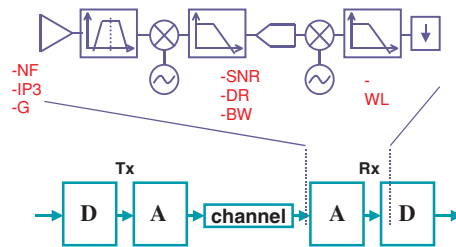


Fig. 2. Digital telecommunication link, indicating a possible receiver front-end architecture with some building block specifications to be determined during frontend architectural exploration.

Consider for instance the digital telecommunication link of Fig. 2. It is clear that digital bits are going into the link to be transmitted over the channel, and that the received signals are being converted again in digital bits. One of the major considerations in digital telecom system design is the bit error rate, which characterizes the reliability of the link. This bit error rate is impacted by the characteristics of the transmission channel itself, but also by the architecture chosen for the transmitter and receiver frontend and by the performances achieved and the nonidealities exhibited by the analog/RF blocks in this frontend. For example, the noise figure and nonlinear distortion of the input low-noise amplifier (LNA) are key parameters. Similarly, the resolution and sampling speed of the analog-to-digital converter (ADC) used may have a large influence on the bit error rate, but it also determines the requirements for the other analog subblocks: a higher ADC resolution may relax the filtering requirements in the transceiver, resulting in simpler filter structures, though it will also consume more power and chip area than a lower-resolution converter. At the same time, the best trade-off solution, i.e. the minimum required ADC resolution and therefore also the minimum power and area, depends on the architecture chosen for the transceiver frontend.

Clearly, there is a large interaction between system-level architectural decisions and the performance requirements for the different subblocks, which on their turn are bounded by technological limits that shift with every new technology process being employed. Hence it is important to offer designers an exploration environment where they can define different frontend architectures and analyse and compare their performance quantitatively and derive the necessary building block specifications. Today the alternative architectures that are explored are still to be provided by the system designer, but future tools might also derive or

synthesize these architectures automatically from a high-level language description [6].

The important ingredients that are needed to set up such an architectural exploration environment are [4,5] :

- a fast high-level simulation method that allows to evaluate the performance (e.g. SNR or BER) of the frontend;
- a library of high-level (behavioral) models for the building blocks used in the targeted application domain, including a correct modeling of the important building block nonidealities (offset, noise, distortion, mirror signals, phase noise, etc.);
- power and area estimation models that, starting from the block specifications, allow estimation of the power consumption and chip area that would be consumed by a real implementation of the block, without really designing the block.

The above ingredients allow a system designer to interactively explore frontend architectures. Combining this with an optimization engine would additionally allow optimization of the selected frontend architecture in determining the optimal building block requirements as to meet the system requirements at minimum implementation cost (power/area). Repeating this optimization for different architectures then makes a quantitative comparison between these architectures possible before they are implemented down to the transistor level. In addition, the high-level exploration environment would also help in deciding on other important system-level decisions, such as determining the optimal partitioning between analog and digital implementations in a mixed-signal system [7], or deciding on the frequency planning of the system, all based on quantitative data rather than ad-hoc heuristics or past experiences.

As the above aspects are not sufficiently available in present commercial system-level simulators like SPW, COSSAP, ADS or Matlab/Simulink, more effective and more efficient solutions are being developed. To make system-level exploration really fast and interactive, dedicated algorithms can be developed that speed up the calculations by maximally exploiting the properties of the system under investigation and using proper approximations where possible. ORCA for instance is targeted towards telecom applications and uses dedicated signal spectral manipulations to gain efficiency [8]. A more recent development is the FAST tool which performs a time-domain dataflow type of simulation without iterations [9] and which easily allows dataflow co-simulation with digital blocks. Compared to current commercial simulators, this simulator is more efficient by using block processing instead of point-by-point calculations for the different time points in circuits without feedback. In addition, the signals are represented as complex equivalent baseband signals with multiple carriers. The signal representation is local and fully optimized as

the signal at each node in the circuit can have a set of multiple carriers and each corresponding equivalent baseband component can be sampled with a different time step depending on its bandwidth. Large feedback loops, especially when they contain nonlinearities, are however more difficult to handle with this approach. A method to efficiently simulate bit error rates with this simulator has been presented in [10].

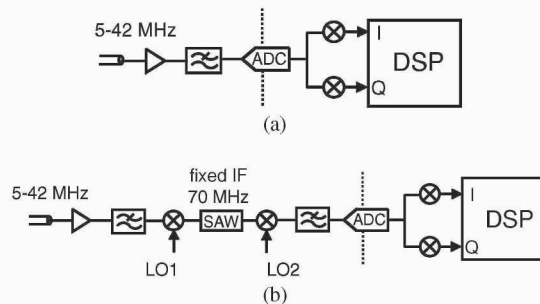


Fig. 3. Two possible architectures for a cable TV application : (a) all-digital architecture, (b) classical architecture.

Example

As an example [4,5], consider a frontend for a cable TV modem receiver, based on the MCNS standard. The MCNS frequency band for upstream communication on the CATV network is from 5 to 42 MHz (extended subsplit band). Two architectures are shown in Fig. 3 : (a) an all-digital architecture where both the channel selection and the downconversion are done in the digital domain, and (b) the classical architecture where the channel selection is performed in the analog domain.

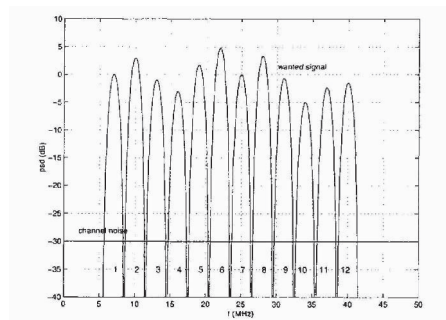


Fig. 4. Typical input spectrum for a CATV frontend architecture using 12 QAM-16 channels.

A typical input spectrum is shown in Fig. 4. For this example we have used 12 QAM-16 channels with a 3 MHz bandwidth. We assume a signal variation of

the different channels of maximally ± 5 dB around the average level. The average channel noise is 30 dB below this level. Fig. 5 shows the spectrum of the selected channel as simulated by ORCA [8] for the all-digital architecture of Fig. 3a at the receiver output after digital channel selection and quadrature downconversion. The wanted channel signal and the effects of the channel noise, the ADC quantization noise, and the second- and third-order distortion are generated separately, providing useful feedback to the system designer. The resulting SNDR is equal to 22.7 dB in this case, which corresponds to a symbol error rate of less than 10^{-10} for QAM-16.

By performing the same analysis for different architectures and by linking the required subblock specifications to the power and/or chip area required to implement the subblocks, a quantitative comparison of different alternative architectures becomes possible with respect to 1) their suitability to implement the system specifications, and 2) the corresponding implementation cost in power consumption and/or silicon real estate. To assess the latter, high-level power and/or area estimators must be used to quantify the implementation cost. In this way the system designer can choose the most promising architecture for the application at hand.

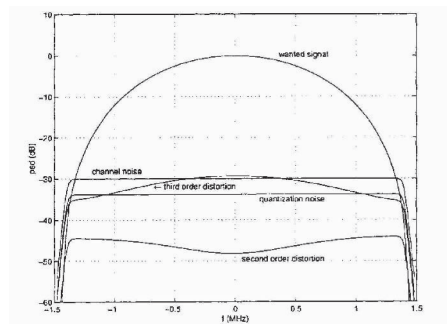


Fig. 5. Simulated spectrum of the selected channel for the all-digital CATV architecture at the receiver output.

Fig. 6 shows a comparison between the estimated total power consumption required by the all-digital and by the classical CATV receiver architectures of Fig. 3 as a function of the required SNR [11]. These results were obtained with the simulator FAST [9]. Clearly, for the technology used in the experiment, the classical architecture still required much less power than the all-digital solution.

Finally, Fig. 7 shows the result of a BER simulation with the FAST tool for a 5-GHz 802.11 WLAN architecture [9]. The straight curve shows the result without taking into account nonlinear distortion caused by the building blocks; the dashed curve takes this distortion into account. Clearly, the BER considerably worsens in the presence of nonlinear distortion. Note that the whole BER

analysis was performed in a simulation time which is two orders of magnitude faster than traditional Monte-Carlo analysis performed on a large number of OFDM symbols.

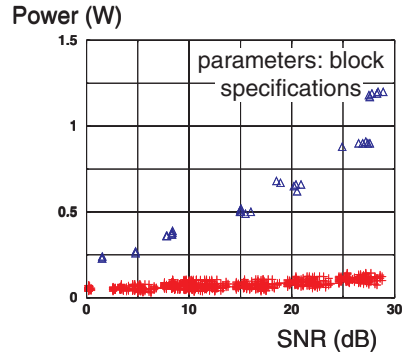


Fig. 6. Power consumption comparison between the all-digital CATV architecture (triangles) and the classical architecture (crosses) as a function of the required SNR [11].

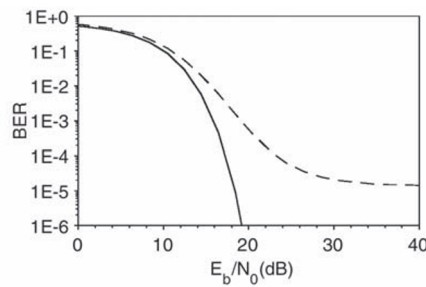


Fig. 7. Simulated BER analysis result for a 5-GHz 802.11 WLAN architecture with (dashed) and without (straight) nonlinear distortion of the building blocks included [10].

2.2. Top-down analog block design

Top-down design is already heavily used in industry today for the design of complex analog blocks like Delta-Sigma converters or phase-locked loops (PLL). In these cases first a high-level design of the block is done with the block represented as an architecture of subblocks, each modeled with a behavioral model that includes the major nonidealities as parameters, rather than a transistor schematic. This step is often done using Matlab/Simulink and it allows to determine the optimal architecture of the block at this level, together with the minimum requirements for the subblocks (e.g. integrators, quantizers, VCO,

etc.), so that the entire block meets its requirements in some optimal sense. This is then followed by a detailed device-level (SPICE) design step for each of the chosen architecture's subblocks, targeted to the derived subblock specifications. This is now illustrated for a phase-locked loop (PLL).

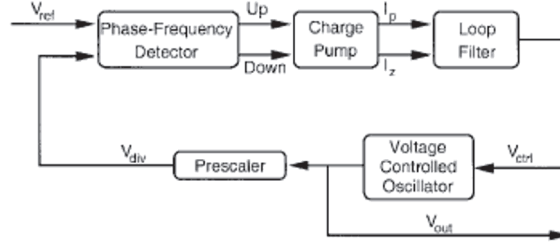


Fig. 8. Basic block diagram of a phase-locked loop analog block.

Example

The basic block diagram of a PLL is shown in Fig. 8. If all subblocks like the phase-frequency detector or the voltage-controlled oscillator (VCO) are represented by behavioral models instead of device-level circuits, then enormous time savings in simulation time can be obtained during the design and verification phase of the PLL. For example, for requirements arising from a GSM-1800 design example (frequency range around 1.8 GHz, phase noise -121 dBc/Hz @ 600 kHz frequency offset, settling time of the loop for channel frequency changes below 1 ms within 1e-6 accuracy), the following characteristics can be derived for the PLL subblocks using behavioral simulations with generic behavioral models for the subblocks [12] : $A_{L_{PF}} = 1$, $K_{VCO} = 1e6$ Hz/V, $N_{div} = 64$, $f_{L_{PF}} = 100$ kHz. These specifications are then the starting point for the device-level design of each of the subblocks.

For the bottom-up system verification phase of a system, more detailed behavioral models have to be generated that are tuned towards the actual circuit design. For example, an accurate behavioral model for a designed VCO is given by the following equation set :

$$v_{out}(t) = A_0(v_{in}(t)) + \sum_{k=1}^{k=N} A_k(v_{in}(t)) \cdot \sin(\Phi_k(t)) \quad (1)$$

$$\Phi_k(t) = \varphi_k(v_{in}(t)) + 2\pi \int_{t_0}^t k \cdot [h_{stat2dyn}(\tau) \otimes f_{stat}(v_{in}(\tau))] d\tau$$

where Φ_k is the phase of each harmonic k in the VCO output, A_k and φ_k characterize the (nonlinear) static characteristic of a VCO, and $h_{stat2dyn}$ characterizes the dynamic voltage-phase behavior of a VCO, both as extracted from circuit-level simulations of the real circuit. For example, Fig. 9 shows the

frequency response of both the original device-level circuit (red) and the extracted behavioral model (blue) for a low-frequency sinusoidal input signal. You can see that this input signal creates a side lobe near the carrier that is represented by the model within 0.25 dB accuracy compared to the original transistor-level circuit, while the gain in simulation time is more than 30x [12].

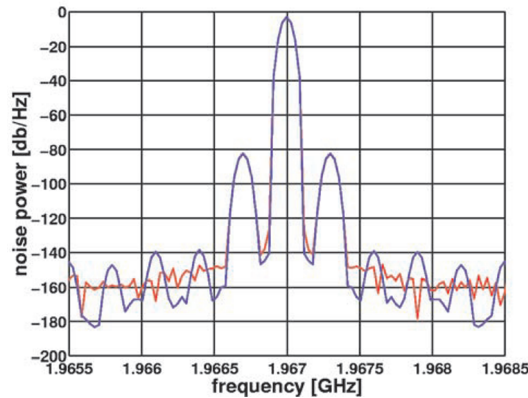


Fig. 9. Frequency response of an extracted behavioral VCO model (blue) compared to the underlying device-level circuit response (red) [12].

3. Behavioral modeling and model generation

There are (at least) four reasons for using higher-level analog modeling (functional, behavioral or macro modeling) for describing and simulating mixed-signal systems [2] :

- The simulation time of circuits with widely spaced time constants (e.g. oversampling converters, phase-locked loops, etc.) is quite large since the time-step control mechanism of the analog solver follows the fastest signals in the circuit. Use of higher-level modeling for the blocks will accelerate the simulation of these systems, particularly if the “fast” time-scale behavior can be “abstracted away”, e.g. by replacing transistor-level descriptions of RF blocks by baseband-equivalent behavioral models.
- In a top-down design methodology based on hierarchical design refinement (like Fig. 1) at higher levels of the design hierarchy, there is a need for higher-level models describing the pin-to-pin behavior of the circuits in a mathematical format rather than representing it as an internal structural netlist of components. This is unavoidable during top-down design since at higher levels in the design hierarchy the details of the underlying circuit implementation are simply not yet known and hence only generic mathematical models can be used.

- A third use of behavioral models is during bottom-up system verification when these models are needed to reduce the CPU time required to simulate the block as part of a larger system. The difference is that in this case the underlying implementation is known in detail, and that peculiarities of the block's implementation can be incorporated as much as possible in the extracted model without slowing down the simulation too much.
- Fourthly, when providing or using analog IP macrocells in a system-on-a-chip context, the virtual component (ViC) has to be accompanied by an executable model that efficiently models the pin-to-pin behavior of the virtual component. This model can then be used in system-level design and verification, by the SoC integrating company, even without knowing the detailed circuit implementation of the macrocell [13].

For all these reasons analog/mixed-signal behavioral simulation models are needed that describe analog circuits at a higher level than the circuit level, i.e. that describe the input-output behavior of the circuit in a mathematical model rather than as a structural network of basic devices. These higher-level models must describe the desired behavior of the block (like amplification, filtering, mixing or quantization) and simulate efficiently, while still including the major nonidealities of real implementations with sufficient accuracy.

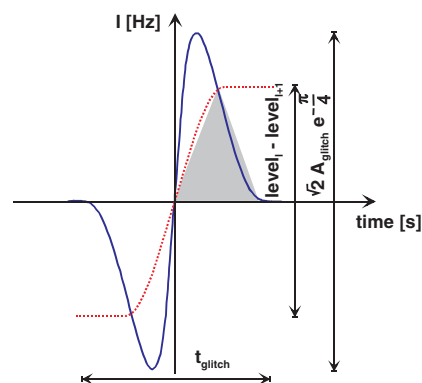


Fig. 10. Typical dynamic behavior of a current-steering digital-to-analog converter output when switching the digital input code.

Example

For example, the realistic dynamic behavior (including settling time and glitch behavior) of a current-steering DAC as shown in Fig. 10 can easily be described by superposition of an exponentially damped sine (modeling the glitch behavior) and a shifted hyperbolic tangent (modeling the settling behavior) [14]:

$$i_{out} = A_{gl} \sin\left(\frac{2\pi}{t_{gl}}(t-t_0)\right) \exp\left(-\text{sign}(t-t_0)\frac{2\pi}{t_{gl}}(t-t_0)\right) + \frac{level_{i+1} - level_i}{2} \tanh\left(\frac{2\pi}{t_{gl}}(t-t_0)\right) + \frac{level_{i+1} + level_i}{2} \quad (2)$$

where $level_i$ and $level_{i+1}$ are the DAC output levels before and after the considered transition, and where A_{gl} , t_0 and t_{gl} are parameters that need to be determined, e.g. by regression fitting to simulation results of a real circuit. Fig. 11 compares the response of the behavioral model (with parameter values extracted from SPICE simulations of the original circuit) with SPICE simulation results of the original circuit. The speed-up in CPU time is a factor 874 (!!)

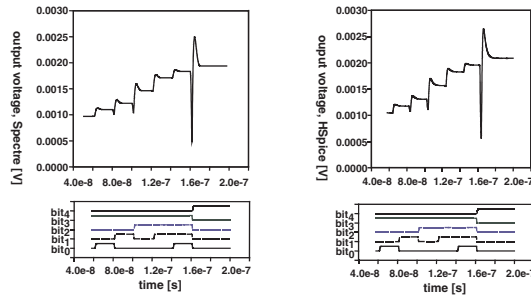


Fig. 11. Comparison between the device-level simulation results (on the right) and the response of the behavioral model (on the left) [14].

The different analog hardware description levels considered in design practice today are [15] :

- the *circuit level* is the traditional level where a circuit is simulated as an network of physical devices;
- in a *macromodel* an equivalent but computationally cheaper circuit representation is used that has approximately the same behavior as the original circuit. Equivalent sources combine the effect of several other elements that are eliminated from the netlist. The simulation speed-up is roughly proportional to the number of nonlinear devices that can be eliminated;
- in a *behavioral model* a purely mathematical description of the input-output behavior of the block is used. This typically will be in the form of a set of differential-algebraic equations (DAE) and/or transfer functions. Conservation laws still have to be satisfied;
- in a *functional model* also a purely mathematical description of the input-output behavior of the block is used, but conservation laws are not enforced and the simulated system turns into a kind of signal-flow diagram.

The industrial use of analog higher-level (functional, behavioral, macro) modeling is today enabled by the availability of standardized mixed-signal hardware description languages such as VHDL-AMS [16,17] and Verilog-AMS [18,19], both of which are extensions of the corresponding digital hardware description languages, and both of which are supported by commercial simulators today. These languages allow description and simulation of separate analog circuits, separate digital circuits and mixed analog-digital circuits, at the above different abstraction levels. In general they also allow description and simulation of both electrical and non-electrical systems, as long as they can be modeled by a set of (nonlinear) differential-algebraic equations. Note that while originally restricted to low-to-medium frequencies with lumped elements only, the standardization of the extension of these languages, e.g. VHDL-AMS, towards the RF/microwave domain with distributed elements has been started in recent years.

3.1. Analog behavioral model generation techniques

One of the largest problems today is the lack of systematic methods to create good analog behavioral or performance models – a skill not yet mastered by the majority of analog designers – as well as the lack of any tools to automate this process. Fortunately, in recent years research has started to develop methods that can automatically create models for analog circuits, both behavioral models for behavioral simulation and performance models for circuit sizing. Techniques used here can roughly be divided into fitting or regression approaches, constructive approaches and model-order reduction methods.

3.1.1. Fitting or regression methods

In the *fitting or regression approaches* a parameterized mathematical model is proposed by the model developer and the values of the parameters \mathbf{p} are selected as to best approximate the known circuit behavior. A systematic approach to regression-based model construction consists of several steps:

1. Selection of an appropriate model structure or template. The possible choices of model are vast. Some of the more common include polynomials, rational functions [20], and neural networks. Recently EDA researchers have begun to utilize results from statistical inference [21] and data mining [22], and we expect to see regression tree, k-nearest neighbor, and kernel forms such as support vector machines [23,24,25] to become more prominent in the future. Posynomial forms have attracted particular interest for optimization applications [26,27], as optimization problems involving models of this form can be recast as convex programs, leading to very efficient sizing of analog circuits.
2. Creation and/or selection of the simulation data to which to fit the model

via an appropriate design-of-experiments scheme [27].

3. Selection of a model fidelity criterion. For example, the model can be fit by a least-square error optimization where the model response matches the simulated (or measured) time-domain response of the real circuit as closely as possible in an average sense [28]. This is schematically depicted in Fig. 12. The error could for instance be calculated as :

$$error = \int_0^T \|v_{out,real}(t) - v_{out,model}(t)\|^2 dt \quad (3)$$

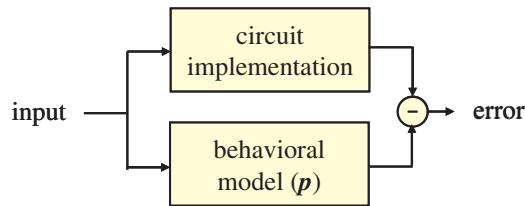


Figure 12. Basic flow of fitting or regression approach for analog behavioral model generation.

4. Selection of the optimization procedure to select the parameters (in some cases this step and the previous are combined), such as by gradient descent or other gradient-based optimization, “boosting” [29], or stochastic optimization such as Markov-chain Monte Carlo or simulated annealing.
5. Validation of the final model. Without specific attention paid to model validation, it is quite common to find “overfit” models. Such models may have small error for the simulation data on which they were “trained”, but very poor accuracy when slightly different circuit excitations are introduced when the model is put into use. Regularization may be introduced in step 3 to attempt to suppress such behavior, e.g. by modifying the model fidelity criterion to penalize large coefficients in a least-squares fit.

It should be clear that these fitting approaches can in principle be very generic as they consider the block as a black box and only look at the input-output behavior of the block which can easily be simulated (or measured). Once the model is generated, it becomes an implicit model of the circuit. However, hiding in each of the steps outlined above are daunting practical challenges. Chief among these comes the first step: for any hope of success, first a good model template must be proposed, which is not always trivial to do in an accurate way without knowing the details of the circuit. Even when good choices are possible, it may happen that the resulting model is specific for one particular implementation of

the circuit. Likewise, the training set must exercise all possible operating modes of the circuit, but these can be hard to predict in advance.

To address these challenges, progress made in other areas such as in research on time series prediction (e.g. support vector machines [24,25]) and data mining techniques [22] are being pursued.

3.1.2. Symbolic constructive model generation methods

The second class of methods, the *constructive approaches*, try to generate or build a model from the underlying circuit description. Inherently these are therefore white-box methods as the resulting model is specific for the particular circuit, but there is a higher guarantee than with the fitting methods that it tracks the real circuit behavior well in a wider range. One approach for instance uses symbolic analysis techniques to first generate the exact set of describing algebraic/differential equations of the circuit, which are then simplified within a given error bound of the exact response using both global and local simplifications [30]. The resulting simplified set of equations then constitutes the behavioral model of the circuit and tracks nicely the behavior of the circuit. The biggest drawback however is that the error estimation is difficult and for nonlinear circuits heavily depends on the targeted response. Up till now, the gains in CPU time obtained in this way are not high enough for practical circuits. More research in this area is definitely needed.

3.1.3. Model-order reduction methods

The third group of methods, the *model-order reduction methods*, are mathematical techniques that generate a model for a given circuit by direct analysis and manipulation of its detailed, low-level description, for example the nonlinear differential equations in a SPICE simulator, or the resistor-capacitor model describing extracted interconnect. Classical model-order reduction algorithms take as input a linear, time-invariant set of differential equations describing a state-space model of the circuit, for example

$$\frac{dx}{dt} = Ax + Bu ; y = Cx + Du \quad (4)$$

where x represents the circuit state, u the circuit inputs, y the circuit outputs, and the matrices A , B , C and D determine the circuit properties. As output model-order reduction methods produce a similar state-space model $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}$, but with a state vector \tilde{x} (thus matrix description) of lower dimensionality, i.e. of lower order :

$$\frac{d\tilde{x}}{dt} = \tilde{A} \tilde{x} + \tilde{B} u ; \tilde{y} = \tilde{C} \tilde{x} + \tilde{D} u \quad (5)$$

These reduced-order models simulate much more efficiently, while approximating the exact response, for example matching the original model closely up to some specified frequency.

Originally developed to reduce the complexity of interconnect networks for timing analysis, techniques such as asymptotic waveform evaluation (AWE) [31] used Padé approximation to generate a lower-order model for the response of the linear interconnect network. The early AWE efforts used explicit moment matching techniques which were not numerically stable, and thus could not produce higher-order models that were needed to model circuits more complicated than resistor-capacitor networks, and Padé approximations often generate unstable and/or non-passive reduced-order models. Subsequent developments using Krylov-subspace-based methods [32,33] resulted in methods like PVL (Padé via Lanczos) that overcome many of the deficiencies of the earlier AWE efforts, and passive model construction is now guaranteed via projection-via-congruence such as used in PRIMA [34].

In recent years, similar techniques have also been extended in an effort to create reduced-order macromodels for analog/RF circuits. Techniques have been developed for time-varying models, particularly periodically time-varying circuits [35,36], and for weakly nonlinear circuits via polynomial-type methods that have a strong relation to Volterra series [37,38,36]. Current research focuses on methods to model more strongly nonlinear circuits (e.g. using trajectory piecewise-linear [39] or piecewise-polynomial approximations [40]) and is starting to overlap with the construction of performance models, through the mutual connection to the regression and data mining ideas [22,24,25].

Despite the progress made so far, still more research in the area of automatic or systematic behavioral model generation or model-order reduction is certainly needed.

3.2. Power and area model generation techniques

Besides behavioral models, the other crucial element to compare different architectural alternatives and to explore trade-offs during system-level exploration and optimization are accurate and efficient power and area estimators [41]. They allow to assess and compare the optimality of different design alternatives. Such estimators are functions that predict the power or area that is going to be consumed by a circuit implementation of an analog block (e.g. an analog-to-digital converter) with given specification values (e.g. resolution and speed). Since the implementation of the block is not yet known during high-level system design and considering the large number of different possible implementations for a block, it is very difficult to generate these estimators with high absolute accuracy. However, for the purpose of comparing different design alternatives during architectural exploration (as discussed in

section 2.1), the tracking accuracy of estimators with varying block specifications is of much more importance.

Such functions can be obtained in two ways. A first possibility is the derivation of analytic functions or procedures that return the power or area estimate given the block's specifications. An example of a general yet relatively accurate power estimator that was derived based on the underlying operating principles for the whole class of CMOS high-speed Nyquist-rate analog-to-digital converters (such as flash, two-step, pipelined, etc. architectures) is given by [41] :

$$power = \frac{V_{dd}^2 \cdot L_{min} \cdot (F_{sample} + F_{signal})}{10^{(-0.1525 \cdot ENOB + 4.8381)}} \quad (6)$$

The estimator is technology scalable, has been fitted with published data of real converters, and for more than 85% of the designs checked, the estimator has an accuracy better than 2.2x. Similar functions are developed for other blocks, but of course often a more elaborate procedure is needed than a simple formula. For example, for the case of high-speed continuous-time filters [41], a crude filter synthesis procedure in combination with operational transconductor amplifier behavioral models had to be developed to generate accurate results, because the implementation details and hence the power and area vary quite largely with the specifications.

A second possibility to develop power/area estimators is to extract them from a whole set of data samples from available or generated designs through interpolation or fitting of a predefined function or an implicit function like e.g. a neural network. As these methods do not rely on underlying operating principles, extrapolations of the models have no guaranteed accuracy.

In addition to power and area estimators also **feasibility functions** are needed that limit the high-level optimization to realizable values of the building block specifications. These can be implemented under the form of functions (e.g. a trained neural network or a support vector machine [42]) that return whether a block is feasible or not, or of the geometrically calculated feasible performance space of a circuit (e.g. using polytopes [43] or using radial base functions [44]).

4. Performance modeling in analog synthesis

While the basic level of design abstraction for analog circuits is mainly still the transistor level, commercial CAD tool support for analog cell-level circuit and layout synthesis is currently emerging. There has been remarkable progress at research level over the past decade, and in recent years several commercial offerings have appeared on the market. Gielen and Rutenbar [2] offer a fairly complete survey of the area. Analog synthesis consists of two major steps: (1) circuit synthesis followed by (2) layout synthesis. Most of the basic techniques in both circuit and layout synthesis rely on powerful numerical optimization

engines coupled to “evaluation engines” that qualify the merit of some evolving analog circuit or layout candidate. The basic scheme of optimization-based analog circuit sizing is shown in Fig. 13. High-level models also form a key part of several analog circuit synthesis and optimization systems, both at the circuit level as well as for the hierarchical synthesis of more complex blocks as will be discussed next.

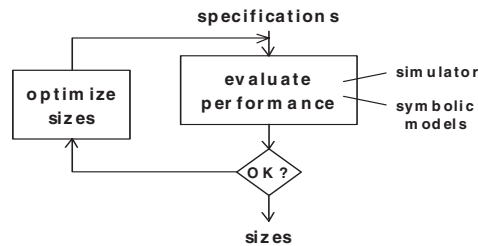


Figure 13. Basic flow of optimization-based analog circuit sizing.

The most general but also by far the slowest automated circuit sizing solution is to call a transistor-level simulator (SPICE) as evaluation engine during each iteration of the optimization of the circuit. These methods therefore couple robust numerical optimization with full SPICE simulation, making it possible to synthesize designs using the same modeling and verification tool infrastructure and accuracy levels that human experts use for manual design, be it at the expense of large CPU times (hours or days of optimization time). Example tools include the FRIDGE tool [45] tool and the ANACONDA tool [46]. The latter tool cuts down on the CPU time by using a global optimization algorithm based on stochastic pattern search that inherently contains parallelism and therefore can easily be distributed over a pool of workstations, to try out and simulate 50,000 to 100,000 circuit candidates in a few hours. These brute-force approaches require very little advance modeling work to prepare for any new circuit topology and have the same accuracy as SPICE. The major drawback is the large CPU times for all optimizations. In [47] ANACONDA/MAELSTROM in combination with macromodeling techniques to bridge the hierarchical levels, was applied to an industrial-scale analog system (the equalizer/filter frontend for an ADSL CODEC). Again, the experiments demonstrated that the synthesis results are comparable to or sometimes better than manual design !!

The huge CPU time consumption of the straightforward simulation-based optimization approaches (sometimes also called “simulator in the loop”) can be reduced significantly by replacing the simulations by model evaluations. These models can be behavioral simulation models as described above, effectively calling behavioral simulation during every optimization iteration, or they can be what are termed performance models [27]. An example of the behavioral approach is the DAISY tool which provides efficient high-level synthesis of

discrete-time $\Delta\Sigma$ modulators [48] based on a behavioral-simulation-based optimization strategy. The high-level optimization approach determines both the optimum modulator topology and the required building block specifications, such that the system specifications — mainly accuracy (dynamic range) and signal bandwidth — are satisfied at the lowest possible power consumption. A genetic-based differential evolution algorithm is used in combination with a fast dedicated $\Delta\Sigma$ behavioral simulator to realistically analyze and optimize the modulator performance. Recently the DAISY tool was also extended to continuous-time $\Delta\Sigma$ modulators [49]. For the synthesis of more complex analog blocks, an hierarchical approach is needed, in which higher-level models are indispensable to bridge between the different levels.

The other alternative to speed up circuit synthesis is to use performance models [27] to evaluate rather than simulate the performance of the candidate circuit solution at each iteration of the optimization. Rather than traditional behavioral models, which model the input-output behavior of a circuit, performance models directly relate the achievable performances of a circuit (e.g. gain, bandwidth, or slew rate) to the design variables (e.g. device sizes and biasing). Fig. 14 for example shows part of such a performance model, displaying the phase margin as a function of two design variables for an operational amplifier [25]. In such *model-based synthesis* procedure, calls to the transistor-level simulation are then replaced by calls to evaluate the performance model, resulting in substantial speedups of the overall synthesis, once the performance models have been created and calibrated. The latter is a one-time up-front investment that has to be done only once for each circuit in each technology.

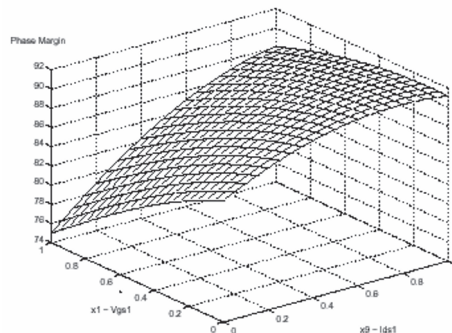


Figure 14. Performance model of the phase margin as a function of two design variables for an opamp (subset of the actual multi-dimensional performance model).

The question remains how such performance models can be generated accurately. Most approaches for performance model generation are based on fitting or regression methods where the parameters of a template model are fitted

to have the model match as closely as possible a sample set of simulated data points. The use of simulated data points guarantees SPICE-level accuracies. A recent example of such fitting approach is the automatic generation of posynomial performance models for analog circuits, that are created by fitting a pre-assumed posynomial equation template to simulation data created according to some design of experiments scheme [27]. Such a posynomial model could then for instance be used in the very efficient sizing of analog circuits through convex circuit optimization. To improve these methods, all progress made in other research areas such as in time series prediction (e.g. support vector machines [25]) or data mining techniques [22] could be applied here as well. For example, Fig. 15 shows results of two different performance models for the same characteristic of an opamp. The graphs plot predicted versus actual values for the gain-bandwidth (GBW) of an opamp. On the left, traditional design of experiments (DOE) techniques have been applied across the performance space of the circuit, resulting in large spread in prediction errors, whereas on the right novel model generation technology [50] is used that results in much better prediction accuracies across the entire performance space, as needed for reliable model-based circuit optimization with guaranteed SPICE-level accuracies.

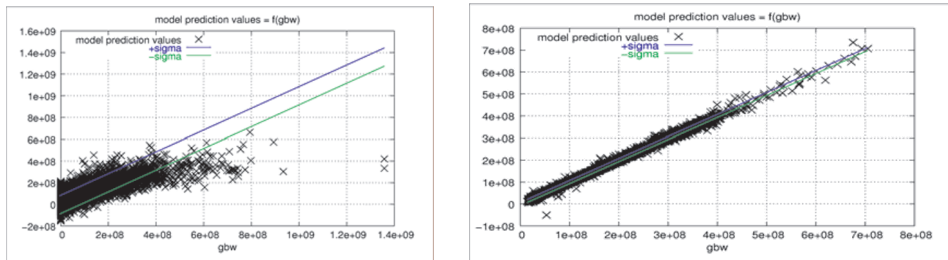


Figure 15. Predicted versus actual values for the gain-bandwidth (GBW) of an opamp: on the left as generated with traditional design of experiments techniques, on the right as generated with novel model generation technology [50].

Despite the progress made so far, still more research in the area of automatic performance model generation is needed to reduce analog synthesis times, especially for hierarchical synthesis of complex analog blocks. This field is a hot research area at the moment.

5. Symbolic modeling of analog and RF circuits

Analog design is a very complex and knowledge-intensive process, which heavily relies on circuit understanding and related design heuristics. Symbolic circuit analysis techniques have been developed to help designers gain a better understanding of a circuit's behavior. A symbolic simulator is a computer tool

that takes as input an ordinary (SPICE-type) netlist and returns as output (simplified) analytic expressions for the requested circuit network functions in terms of the symbolic representations of the frequency variable and (some or all of) the circuit elements [51,52]. They perform the same function that designers traditionally do by hand analysis (even the simplification). The difference is that the analysis is now done by the computer, which is much faster, can handle more complex circuits and does not make as many errors. An example of a complicated BiCMOS opamp is shown in Fig. 16. The (simplified) analytic expression for the differential small-signal gain of this opamp has been analyzed with the SYMBA tool [53], and is shown below in terms of the small-signal parameters of the opamp's devices :

$$A_{V0} = \frac{g_{m,M2}}{g_{m,M1}} \frac{g_{m,M4}}{\left(\frac{g_{o,M4}g_{o,M5}}{g_{m,M5} + g_{mb,M5}} + \frac{G_a + g_{o,M9} + g_{o,Q2}}{\beta_{Q2}} \right)} \quad (7)$$

The symbolic expression gives a better insight into which small-signal circuit parameters predominantly determine the gain in this opamp and how the user has to design the circuit to meet a certain gain constraint. In this way, symbolic circuit analysis is complementary to numerical (SPICE) circuit simulation, which was described in the previous section. Symbolic analysis provides a different perspective that is more suited for obtaining insight in a circuit's behavior and for circuit explorations, whereas numerical simulation is more appropriate for detailed design validation once a design point has been decided upon.

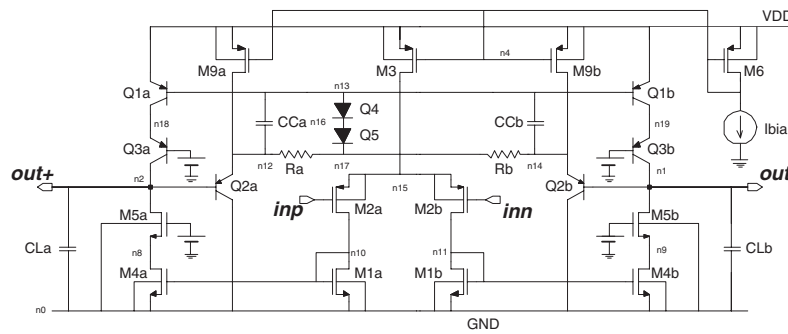


Figure 16. BiCMOS operational amplifier to illustrate symbolic analysis.

At this moment, only symbolic analysis of linear or small-signal linearized circuits in the frequency domain is possible, both for continuous-time and discrete-time (switched) analog circuits [51,52,54]. In this way, symbolic expressions can be generated for transfer functions, impedances, noise functions,

etc. In addition to understanding the first-order functional behavior of an analog circuit, a good understanding of the second-order effects in a circuit is equally important for the correct functioning of the design in its system application later on. Typical examples are the PSRR and the CMRR of a circuit, which are limited by the mismatches between circuit elements. These mismatches are represented symbolically in the formulas. Another example is the distortion or intermodulation behavior, which is critical in telecom applications. To this end, the technique of symbolic simulation has been extended to the symbolic analysis of distortion and intermodulation in weakly nonlinear analog circuits where the nonlinearity coefficients of the device small-signal elements appear in the expressions [37]. For example, the (simplified) symbolic expression for the second-order output intercept point for the feedback circuit of Fig. 17 for frequencies up to the gain-bandwidth can be generated by symbolic analysis as :

$$OIP2_{LF} = \frac{2g_{m_{M1}}g_{m_{M3}}}{-g_{m_{M1}}K_{2g_{m_{M3}}} + g_{m_{M3}}(K_{2g_{m_{M1B}}} - K_{2g_{m_{M1A}}})} \quad (8)$$

where the K_{2x} coefficient represents the second-order nonlinearity coefficient of the small-signal element x . Note that the mismatch between transistors M1A and M1B is crucial for the distortion at lower frequencies in this circuit.

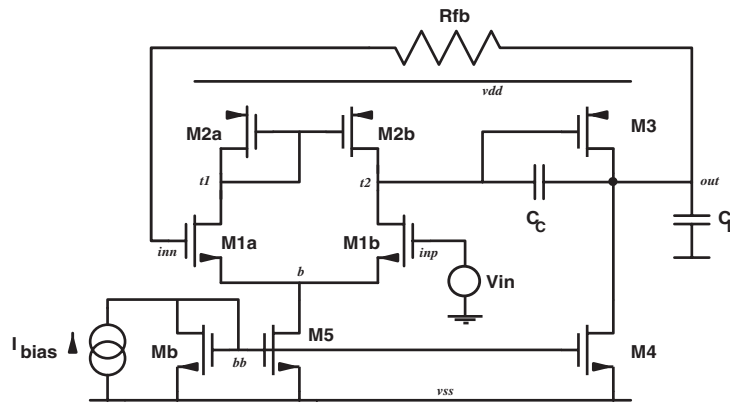


Figure 17. CMOS opamp with feedback to illustrate symbolic distortion analysis.

Exact symbolic solutions for network functions, however, are too complex for linear(ized) circuits of practical size, and even impossible to calculate for many nonlinear effects. Even rather small circuits lead to an astronomically high number of terms in the expressions, that can neither be handled by the computer nor interpreted by the circuit designer. Therefore, since the late eighties, and in principle similar to what designers do during hand calculations, dedicated

symbolic analysis tools have been developed that use heuristic simplification and pruning algorithms based on the relative importance of the different circuit elements to reduce the complexity of the resulting expressions and retain only the dominant contributions within user-controlled error tolerances. Examples of such tools are ISAAC [54], SYNAP [55] and ASAP [56] among many others. Although successful for relatively small circuits, the fast increase of the CPU time with the circuit size restricted their applicability to circuits between 10 and 15 transistors only, which was too small for many practical applications.

In the past years, however, an algorithmic breakthrough in the field of symbolic circuit analysis has been realized. The techniques of simplification before and during the symbolic expression generation, as implemented in tools like SYMBA [53] and RAINIER [57], highly reduce the computation time and therefore enable the symbolic analysis of large analog circuits of practical size (like the entire 741 opamp or the example of Fig. 16). In simplification before generation (SBG), the circuit schematic, or some associated matrix or graph(s), are simplified before the symbolic analysis starts [58,59]. In simplification during generation (SDG), instead of generating the exact symbolic expression followed by pruning the unimportant contributions, the desired simplified expression is built up directly by generating the contributing dominant terms one by one in decreasing order of magnitude, until the expression has been generated with the desired accuracy [53,57]. In addition, the technique of determinant decision diagrams (DDD) has been developed as a very efficient canonical representation of symbolic determinants in a compact nested format [60]. The advantage is that all operations on these DDD's are linear with the size of the DDD, but the DDD itself is not always linear with the size of the circuit. Very efficient methods have been developed using these DDD's [60,61].

All these techniques, however, still result in large, expanded expressions, which restricts their usefulness for larger circuits. Therefore, for really large circuits, the technique of hierarchical decomposition has been developed [62,63]. The circuit is recursively decomposed into loosely connected subcircuits. The lowest-level subcircuits are analyzed separately and the resulting symbolic expressions are combined according to the decomposition hierarchy. This results in the global nested expression for the complete circuit, which is much more compact than the expanded expression. The CPU time increases about linearly with the circuit size, provided that the coupling between the different subcircuits is not too strong. Also the DDD technique has been combined with hierarchical analysis in [64].

Another recent extension is towards the symbolic analysis of linear periodically time-varying circuits, such as mixers [65]. The approach generalises the concept of transfer functions to harmonic transfer matrices, generating symbolic expressions for the transfer function from any frequency band from the circuit's input signal to any frequency band from the circuit's output signal.

In addition, recent approaches have also started to appear that generate symbolic expressions for large-signal and transient characteristics, for instance using piecewise-linear approximations or using regression methods that fit simulation data to predefined symbolic templates. A recent example of such fitting approach is the automatic generation of symbolic posynomial performance models for analog circuits, that are created by fitting a pre-assumed posynomial equation template to simulation data created according to some design-of-experiments scheme [27]. Very recently even a template-free approach has been presented where no predefined fitting template is used, but where the “template” is evolved dynamically using genetic optimization with a canonical-form grammar that adds extra terms or functions to the evolving symbolic expression until sufficient accuracy is obtained for the symbolic results with respect to the reference set of simulation data [66]. This kind of methods are very promising, since they are no longer limited to simple device models nor to small-signal characteristics only – they basically work for whatever characteristic can be simulated – but they still need further research.

Based on the many research results in this area over the last decade, it can be expected that symbolic analysis techniques will soon be part of the standard tool suite of every analog designer, as an add-on to numerical simulation.

6. Conclusions

The last few years have seen significant advances in both design methodology and CAD tool support for analog, mixed-signal and RF designs. The emergence of commercial analog/mixed-signal (AMS) simulators supporting multiple analog abstraction levels (functional, behavioral, macromodel and circuit level) enables top-down design flows in many industrial scenarios. In addition, there is increasing progress in system-level modeling and analysis allowing architectural exploration of entire systems, as well as in mixed-signal verification for both functional verification and to anticipate problems related to embedding the analog blocks in a digital environment. A crucial element to enable this is the development of techniques to generate efficient behavioral models. An overview of model generation techniques has been given, including regression-based methods as well as model-order reduction techniques. Also the generation of performance models for analog circuit synthesis and of symbolic models that provide designers with insight in the relationships governing the performance behavior of a circuit has been described. Despite the enormous progress, model generation remains a difficult art that needs more research work towards automatic and reliable model generation techniques.

References

- [1] "International Technology Roadmap for Semiconductors 2003," <http://public.itrs.net>.
- [2] G. Gielen, R. Rutenbar, "Computer-aided design of analog and mixed-signal integrated circuits," Proceedings of the IEEE, Vol. 88, No. 12, pp. 1825-1854, December 2000.
- [3] H. Chang, E. Charbon, U. Choudhury, A. Demir, E. Felt, E. Liu, E. Malavasi, A. Sangiovanni-Vincentelli, I. Vassiliou, "A top-down constraint-driven design methodology for analog integrated circuits," Kluwer Academic Publishers, 1997.
- [4] G. Gielen, "Modeling and analysis techniques for system-level architectural design of telecom frontends," IEEE Transactions on Microwave Theory and Techniques, Vol. 50, No. 1, pp. 360-368, January 2002.
- [5] G. Gielen, "Top-down design of mixed-mode systems: challenges and solutions," chapter 18 in Advances in Analog Circuit Design (edited by Huijsing, Sansen, Van der Plassche), Kluwer Academic Publishers, 1999.
- [6] "The MEDEA+ Design Automation Roadmap 2003", <http://www.medeas.org>.
- [7] S. Donnay, G. Gielen, W. Sansen, "High-level analog/digital partitioning in low-power signal processing applications," proceedings 7th international workshop on Power and Timing Modeling, Optimization and Simulation (PATMOS), pp. 47-56, 1997.
- [8] J. Crols, S. Donnay, M. Steyaert, G. Gielen, "A high-level design and optimization tool for analog RF receiver front-ends," proceedings International Conference on Computer-Aided Design (ICCAD), pp. 550-553, 1995.
- [9] P. Wambacq, G. Vandersteen, Y. Rolain, P. Dobrovolny, M. Goffioul, S. Donnay, "Dataflow simulation of mixed-signal communication circuits using a local multirate, multicarrier signal representation," IEEE Transactions on Circuits and Systems, Part I, Vol. 49, No. 11, pp. 1554-1562, November 2002.
- [10] G. Vandersteen et al., "Efficient bit-error-rate estimation of multicarrier transceivers," proceedings Design, Automation and Test in Europe (DATE) conference, pp. 164-168, 2001.
- [11] P. Wambacq, G. Vandersteen, S. Donnay, M. Engels, I. Bolsens, E. Lauwers, P. Vanassche, G. Gielen, "High-level simulation and power modeling of mixed-signal front-ends for digital telecommunications," proceedings International Conference on Electronics, Circuits and Systems (ICECS), pp. 525-528, September 1999.
- [12] Bart De Smedt, Georges Gielen, "Models for systematic design and verification of frequency synthesizers," IEEE Transactions on Circuits and Systems, Part II: Analog and Digital Signal Processing, Vol. 46, No. 10, pp. 1301-1308, October 1999.
- [13] Virtual Socket Interface Alliance, several documents including VSIA Architecture Document and Analog/Mixed-Signal Extension Document, <http://www.vsia.org>.
- [14] J. Vandenbussche et al., "Systematic design of high-accuracy current-

- steering D/A converter macrocells for integrated VLSI systems,” *IEEE Transactions on Circuits and Systems, Part II*, Vol. 48, No. 3, pp. 300-309, March 2001.
- [15] A. Vachoux, J.-M. Bergé, O. Levia, J. Rouillard (editors), “Analog and mixed-signal hardware description languages,” Kluwer Academic Publishers, 1997.
- [16] E. Christen, K. Bakalar, “VHDL-AMS - a hardware description language for analog and mixed-signal applications,” *IEEE Transactions on Circuits and Systems, part II: Analog and Digital Signal Processing*, Vol. 46, No. 10, October 1999.
- [17] IEEE 1076.1 Working Group, “Analog and mixed-signal extensions to VHDL,” <http://www.eda.org/vhdl-ams/>.
- [18] K. Kundert, O. Zinke, “The designer’s guide to Verilog-AMS,” Kluwer Academic Publishers, 2004.
- [19] “Verilog-AMS Language Reference Manual” version 2.1, <http://www.eda.org/verilog-ams/>.
- [20] B. Antao, F. El-Turky, “Automatic analog model generation for behavioral simulation,” proceedings IEEE Custom Integrated Circuits Conference (CICC), pp 12.2.1-12.2.4, May 1992.
- [21] T. Hastie, R. Tibshirani, J. Friedman, “The elements of statistical learning,” Springer Verlag, 2001.
- [22] H. Liu, A. Singhee, R. Rutenbar, L.R. Carley, “Remembrance of circuits past: macromodeling by data mining in large analog design spaces,” proceedings Design Automation Conference (DAC), pp. 437-442, June 2002.
- [23] B. Scholkopf, A. Smola, “Learning with Kernels,” MIT Press, 2001.
- [24] J. Phillips, J. Afonso, A. Oliveira, L.M. Silveira, “Analog macromodeling using kernel methods,” proceedings IEEE/ACM International Conference on Computer-Aided Design (ICCAD), November 2003.
- [25] Th. Kiely, G. Gielen, “Performance modeling of analog integrated circuits using least-squares support vector machines,” proceedings Design, Automation and Test in Europe (DATE) conference, pp. 448-453, February 2004.
- [26] M. Hershenson, S. Boyd, T. Lee, “Optimal design of a CMOS op-amp via geometric programming,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 20, pp. 1-21, January 2001.
- [27] W. Daems, G. Gielen, W. Sansen, “Simulation-based generation of posynomial performance models for the sizing of analog integrated circuits,” *IEEE Transactions on Computer-Aided Design*, Vol. 22, No. 5, pp. 517-534, May 2003.
- [28] R. Saleh, B. Antao, J. Singh, “Multi-level and mixed-domain simulation of analog circuits and systems,” *IEEE Transaction on Computer-Aided Design*, Vol. 15, pp 68-82, January 1996.
- [29] Y. Freund, R. Schapire, “A short introduction to boosting,” *Journal of Japanese Society for Artificial Intelligence*, Vol. 14, pp. 771-780, 1999.
- [30] C. Borchers, L. Hedrich, E. Barke, “Equation-based behavioral model generation for nonlinear analog circuits,” proceedings IEEE/ACM Design Automation Conference (DAC), pp. 236-239, 1996.

- [31] L. Pillage, R. Rohrer, "Asymptotic waveform evaluation for timing analysis," *IEEE Transactions on Computer-Aided Design*, Vol. 9, No. 4, pp. 352-366, April 1990.
- [32] L.M. Silveira et al., "A coordinate-transformed Arnoldi algorithm for generating guaranteed stable reduced-order models of RLC circuits," proceedings *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 288-294, 1996.
- [33] P. Feldmann, R. Freund, "Efficient linear circuit analysis by Padé approximation via the Lanczos process," *IEEE Transactions on Computer-Aided Design*, Vol. 14, pp 639-649, 1995.
- [34] A. Odabasioglu, M. Celik, L. Pileggi, "PRIMA: passive reduced-order interconnect macromodeling algorithm," *IEEE Transactions on Computer-Aided Design*, Vol. 17, No. 8, pp. 645-654, August 1998.
- [35] J. Roychowdhury, "Reduced-order modeling of time-varying systems," *IEEE Transactions on Circuits and Systems, Part II*, Vol. 46, No. 10, pp. 1273-1288, October 1999.
- [36] J. Phillips, "Projection-based approaches for model reduction of weakly nonlinear, time-varying systems," *IEEE Transactions on Computer-Aided Design*, Vol. 22, pp. 171-187, 2003.
- [37] P. Wambacq, G. Gielen, P. Kinget, W. Sansen, "High-frequency distortion analysis of analog integrated circuits," *IEEE Transactions on Circuits and Systems, Part II*, Vol. 46, No. 3, pp. 335-345, March 1999.
- [38] L. Peng, L. Pileggi, "NORM: compact model order reduction of weakly nonlinear systems," proceedings *Design Automation Conference*, pp. 472-477, June 2003.
- [39] M. Rewienski, J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices," *IEEE Transactions on Computer-Aided Design*, Vol. 22, No. 2, pp. 155-170, February 2003.
- [40] Ning Dong, J. Roychowdhury, "Piecewise polynomial nonlinear model reduction," proceedings *Design Automation Conference*, pp. 484-489, June 2003.
- [41] E. Lauwers, G. Gielen, "Power estimation methods for analog circuits for architectural exploration of integrated systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Vol. 10, No. 2, pp. 155 - 162, April 2002.
- [42] F. De Bernardinis, M. Jordan, A. Sangiovanni-Vincentelli, "Support vector machines for analog circuit performance representation," proceedings *Design Automation Conference (DAC)*, pp. 964-969, June 2003.
- [43] P. Veselinovic et al., "A flexible topology selection program as part of an analog synthesis system," proceedings *IEEE European Design & Test Conference (ED&TC)*, pp. 119-123, 1995.
- [44] R. Harjani, J. Shao, "Feasibility and performance region modeling of analog and digital circuits," *Kluwer International Journal on Analog Integrated Circuits and Signal Processing*, Vol. 10, No. 1, pp. 23-43, January 1996.
- [45] F. Medeiro et al., "A statistical optimization-based approach for automated sizing of analog cells," proceedings *ACM/IEEE International Conference on Computer-Aided Design (ICCAD)*, pp. 594-597, 1994.

- [46] R. Phelps, M. Krasnicki, R. Rutenbar, L.R. Carley, J. Hellums, "Anaconda: simulation-based synthesis of analog circuits via stochastic pattern search," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 19, No. 6, pp. 703–717, June 2000.
- [47] R. Phelps, M. Krasnicki, R. Rutenbar, L.R. Carley, J. Hellums, "A case study of synthesis for industrial-scale analog IP: redesign of the equalizer/filter frontend for an ADSL CODEC," *proceedings ACM/IEEE Design Automation Conference (DAC)*, pp. 1-6, 2000.
- [48] K. Francken, G. Gielen, "A high-level simulation and synthesis environment for Delta-Sigma modulators," *IEEE Transactions on Computer-Aided Design*, Vol. 22, No. 8, pp. 1049-1061, August 2003.
- [49] G. Gielen, K. Francken, E. Martens, M. Vogels, "An analytical integration method for the simulation of continuous-time $\Delta\Sigma$ modulators," *IEEE Transactions on Computer-Aided Design*, Vol. 23, No. 3, pp. 389-399, March 2004.
- [50] Courtesy of Kimotion Technologies (<http://www.kimotion.com>).
- [51] G. Gielen, P. Wambacq, W. Sansen, "Symbolic analysis methods and applications for analog circuits: a tutorial overview," *The Proceedings of the IEEE*, Vol. 82, No. 2, pp. 287-304, February 1994.
- [52] F. Fernández, A. Rodríguez-Vázquez, J. Huertas, G. Gielen, "Symbolic analysis techniques - Applications to analog design automation," *IEEE Press*, 1998.
- [53] P. Wambacq, F. Fernández, G. Gielen, W. Sansen, A. Rodríguez-Vázquez, "Efficient symbolic computation of approximated small-signal characteristics," *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 3, pp. 327-330, March 1995.
- [54] G. Gielen, H. Walscharts, W. Sansen, "ISAAC: a symbolic simulator for analog integrated circuits," *IEEE Journal of Solid-State Circuits*, Vol. 24, No. 6, pp. 1587-1596, December 1989.
- [55] S. Seda, M. Degrauwe, W. Fichtner, "A symbolic analysis tool for analog circuit design automation," *proceedings IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 488-491, 1988.
- [56] F. Fernández, A. Rodríguez-Vázquez, J. Huertas, "Interactive AC modeling and characterization of analog circuits via symbolic analysis," *Kluwer International Journal on Analog Integrated Circuits and Signal Processing*, Vol. 1, pp. 183-208, November 1991.
- [57] Q. Yu, C. Sechen, "A unified approach to the approximate symbolic analysis of large analog integrated circuits," *IEEE Transactions on Circuits and Systems, Part I*, Vol. 43, No. 8, pp. 656-669, August 1996.
- [58] W. Daems, G. Gielen, W. Sansen, "Circuit simplification for the symbolic analysis of analog integrated circuits," *IEEE Transactions on Computer-Aided Design*, Vol. 21, No. 4, pp. 395-407, April 2002.
- [59] J. Hsu, C. Sechen, "DC small signal symbolic analysis of large analog integrated circuits," *IEEE Transactions on Circuits and Systems, Part I*, Vol. 41, No. 12, pp. 817-828, December 1994.
- [60] C.-J. Shi, X.-D. Tan, "Canonical symbolic analysis of large analog circuits with determinant decision diagrams," *IEEE Transactions on Computer-Aided Design*, Vol. 19, No. 1, pp. 1-18, January 2000.
- [61] W. Verhaegen, G. Gielen, "Efficient DDD-based symbolic analysis of

- linear analog circuits,” IEEE Transactions on Circuits and Systems, Part II: Analog and Digital Signal Processing, Vol. 49, No. 7, pp. 474-487, July 2002.
- [62] J. Starzyk, A. Konczykowska, “Flowgraph analysis of large electronic networks,” IEEE Transactions on Circuits and Systems, Vol. 33, No. 3, pp. 302-315, March 1986.
- [63] O. Guerra, E. Roca, F. Fernández, A. Rodríguez-Vázquez, “A hierarchical approach for the symbolic analysis of large analog integrated circuits,” proceedings IEEE Design Automation and Test in Europe conference (DATE), pp. 48-52, 2000.
- [64] X.-D. Tan, C.-J. Shi, “Hierarchical symbolic analysis of analog integrated circuits via determinant decision diagrams,” IEEE Transactions on Computer-Aided Design, Vol. 19, No. 4, pp. 401-412, April 2000.
- [65] P. Vanassche, G. Gielen, W. Sansen, “Symbolic modeling of periodically time-varying systems using harmonic transfer matrices,” IEEE Transactions on Computer-Aided Design, Vol. 21, No. 9, pp. 1011-1024, September 2002.
- [66] T. McConaghy, T. Eeckelaert, G. Gielen, “CAFFEINE: template-free symbolic model generation of analog circuits via canonical form functions and genetic programming,” proceedings IEEE Design Automation and Test in Europe (DATE), pp. 1082-1087, 2005.

Automated Macromodelling for Simulation of Signals and Noise in Mixed-Signal/RF Systems

Jaijeet Roychowdhury*

Abstract

During the design of electronic circuits and systems, particularly those for RF communications, the need to abstract a subsystem from a greater level of detail to one at a lower level of detail arises frequently. One important application is to generate simple, yet accurate, system-level macromodels that capture circuit-level non-idealities such as distortion. In recent years, computational (“algorithmic”) techniques have been developed that are capable of automating this abstraction process for broad classes of differential-equation-based systems (including nonlinear ones). In this paper, we review the main ideas and techniques behind such algorithmic macromodelling methods.

1 Introduction

Electronic systems today, especially those for communications and sensing, are typically composed of a complex mix of digital, analog and RF circuit blocks. Simulating or verifying such systems is critical for discovering and correcting problems prior to fabrication, in order to avoid re-fabrication which is typically very expensive. Simulating entire systems to the extent needed to generate confidence in the correctness of the to-be-fabricated product is, however, also usually very challenging in terms of computation time.

A common and useful approach towards verification in such situations, both during early system design and after detailed block design, is to replace large and/or complex blocks by small *macromodels* that replicate their input-output functionality well, and verify the macromodelled system. The macromodelled system can be simulated rapidly in order to evaluate different choices of design-space parameters. Such a macromodel-based

*Jaijeet Roychowdhury is with the University of Minnesota, Minneapolis, USA.

verification process affords circuit, system and architecture designers considerable flexibility and convenience through the design process, especially if performed hierarchically using macromodels of differing sizes and fidelity.

The key issue in the above methodology is, of course, the creation of macromodels that represent the blocks of the system well. This is a challenging task for the wide variety of communication and other circuit blocks in use today. The most prevalent approach towards creating macromodels is *manual abstraction*. Macromodels are usually created by the same person who designs the original block, often aided by simulations. While this is the only feasible approach today for many complex blocks, it does have a number of disadvantages compared to the *automated alternatives* that are the subject of this paper. Simulation often does not provide abstracted parameters of interest directly (such as poles, residues, modulation factors, *etc.*); obtaining them by manual postprocessing of simulation results is inconvenient, computationally expensive and error-prone. Manual structural abstraction of a block can easily miss the very nonidealities or interactions that detailed verification is meant to discover. With semiconductor device dimensions shrinking below 100nm and non-idealities (such as substrate/interconnect coupling, degraded device characteristics, *etc.*) becoming increasingly critical, the fidelity of manually-generated macromodels to the real subsystems to be fabricated eventually is becoming increasingly suspect. Adequate incorporation of non-idealities into behavioral models, if at all possible by hand, is typically complex and laborious. Generally speaking, manual macromodelling is heuristic, time-consuming and highly reliant on detailed internal knowledge of the block under consideration, which is often unavailable when subsystems that are not designed in-house are utilized. As a result, the potential time-to-market improvement via macromodel-based verification can be substantially negated by the time and resources needed to first generate the macromodels.

It is in this context that there has been considerable interest in *automated techniques* for the creation of macromodels. Such techniques take a detailed description of a block – for example, a SPICE-level circuit netlist – and generate, via an automated computational procedure, a much smaller macromodel. The macromodel, fundamentally a small system of equations, is usually translated into Matlab/Simulink form for use at the system level. Such an automated approach, *i.e.*, one that remains sustainable as devices shrink from deep submicron to nano-scale, is essential for realistic exploration of the design space in current and future communication circuits and systems.

Several broad methodologies for automated macromodelling have been proposed. One is to generalize, abstract and automate the manual macromodelling process. For example, common topological elements in a circuit are recognized, approximated and conglomerated (*e.g.*, [16, 61]) to create a macromodel. Another class of approaches attempts to

capture *symbolic* macromodels that capture the system's input-output relationship, *e.g.*, [35, 54–56, 59, 65]. Yet another class (*e.g.*, [4, 15, 21]) employs a *black-box* methodology. Data is collected via many simulations or measurements of the full system and a regression-based model created that can predict outputs from inputs. Various methods are available for the regression, including data mining, multi-dimensional tables, neural networks, genetic algorithms, *etc.*...

In this paper, we focus on another methodology for macromodelling, often termed *algorithmic*. Algorithmic macromodelling methods approach the problem as the transformation of a large set of mathematical equations to a much smaller one. The principal advantage of these methods is generality - so long as the equations of the original system are available numerically (*e.g.*, from within SPICE), knowledge of circuit structure, operating principles, *etc.*, is not critical. A single algorithmic method may therefore apply to *entire classes* of physical systems, encompassing circuits and functionalities that may be very disparate. Four such classes, namely linear time invariant (LTI), linear time varying (LTV), nonlinear (non-oscillatory) and oscillatory are discussed in Sections 2, 3, 4 and 5 of this paper. Algorithmic methods also tend to be more rigorous about important issues such as fidelity and stability, and often provide better guarantees of such characteristics than other methods.

2 Macromodelling Linear Time Invariant (LTI) Systems

Often referred to as reduced-order modelling (ROM) or model-order reduction (MOR), automated model generation methods for Linear Time-Invariant (LTI) systems are the most mature amongst algorithmic macromodelling methods. Any block composed of resistors, capacitors, inductors, linear controlled sources and distributed interconnect models is LTI (often referred to simply as “linear”). The development of LTI MOR methods has been driven largely by the need to “compress” the huge interconnect networks, such as clock distribution nets, that arise in large digital circuits and systems. Replacing these networks by small macromodels makes it feasible to complete accurate timing simulations of digital systems at reasonable computational expense. Although interconnect-centric applications have been the main domain for LTI reduction, it is appropriate for any system that is linear and time-invariant. For example, “linear amplifiers”, *i.e.*, linearizations of mixed-signal amplifier blocks, are good candidates for LTI MOR methods.

Figure 1 depicts the basic structure of an LTI block. $u(t)$ represents the inputs to the system, and $y(t)$ the outputs, in the time domain; in the Laplace (or frequency) domain, their transforms are $U(s)$ and $Y(s)$ respectively. The definitive property of any LTI system [67] is that the input and output are related by convolution with an *impulse response* $h(t)$

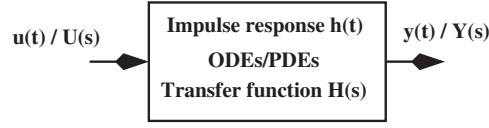


Figure 1: Linear Time Invariant block

in the time domain, *i.e.*, $y(t) = x(t) * h(t)$). Equivalently, their transforms are related by multiplication with a *system transfer function* $H(s)$, *i.e.*, $Y(s) = H(s)X(s)$. Note that there may be many internal nodes or variables within the block. The goal of LTI MOR methods is to replace the block by one with far fewer internal variables, yet with an acceptably similar impulse response or transfer function.

In the majority of circuit applications, the LTI block is described to the MOR method as a set of differential equations, *i.e.*,

$$\begin{aligned} E\dot{x} &= Ax(t) + Bu(t) \\ y(t) &= C^T x(t) + Du(t) \end{aligned} \quad (1)$$

In (1), $u(t)$ represents the input waveforms to the block and $y(t)$ the outputs. Both are relatively few in number compared to the size of $x(t)$, the state of the internal variables of the block. A , B , C , D and E are constant matrices. Such differential equations can be easily formed from SPICE netlists or AHDL descriptions; especially for interconnect applications, the dimension n of $x(t)$ can be very large.

The first issue in LTI ROM is to determine what aspect of the transfer function of the original system should be retained by the reduced system; in other words, what metric of fidelity is appropriate. In their seminal 1990 paper [39], Pileggi and Rohrer used *moments* of the transfer function as fidelity metrics, to be preserved by the model reduction process. The moments m_i of an LTI transfer function $H(s)$ are related to its derivatives, *i.e.*,

$$m_1 = \left. \frac{dH(s)}{ds} \right|_{s=s_0}, \quad m_2 = \left. \frac{d^2H(s)}{ds^2} \right|_{s=s_0}, \dots, \quad (2)$$

where s_0 is a frequency point of interest. Moments can be shown to be related to practically useful metrics, such as delay in interconnects.

In [39], Pileggi and Rohrer proposed a technique, Asymptotic Waveform Evaluation (AWE), for constructing a reduced model for the system (1). AWE first computes a number of moments of the full system (1), then uses these in another set of linear equations, the solution of which results in the reduced model. Such a procedure is termed *explicit moment matching*. The key property of AWE was that it could be shown to produce reduced models whose first several moments (at a given frequency point s_0) were identical to those of the

full system. The computation involved in forming the reduced model was roughly linear in the size of the (large) original system.

While explicit moment matching via AWE proved valuable and was quickly applied to interconnect reduction, it was also observed to become numerically inaccurate as the size of the reduced model increased beyond about 10. To alleviate these, variations based on matching moments at *multiple frequency points* were proposed [2] that improved numerical accuracy. Nevertheless, the fundamental issue of numerical inaccuracy as reduced model sizes grew remained.

In 1994, Gallivan et al [9] and Feldmann/Freund [7, 8] identified the reason for this numerical inaccuracy. Computing the k^{th} moment explicitly involves evaluating terms of the form $A^{-k}r$, *i.e.*, the k^{th} member of the *Krylov subspace* of A and r . If A has well separated eigenvalues (as it typically does for circuit matrices), then for $k \sim 10$ and above, only the dominant eigenvalue contributes to these terms, with non-dominant ones receding into numerical insignificance. Furthermore, even with the moments available accurately, the procedure of finding the reduced model is also poorly conditioned.

Recognizing that these are not limitations fundamental to the goal of model reduction, [7, 9] proposed alternatives. They showed that numerically robust procedures for computing Krylov subspaces, such as the Lanczos and Arnoldi (*e.g.*, [50]) methods, could be used to produce reduced models that match any given number of moments of the full system. These approaches, called *Krylov-subspace MOR techniques*, do not compute the moments of the full system explicitly at any point, *i.e.*, they perform *implicit moment matching*. In addition to matching moments in the spirit of AWE, Krylov-subspace methods were also shown to capture well the dominant poles and residues of the system. The Padé-via-Lanczos (PVL) technique [7] gained rapid acceptance within the MOR community by demonstrating its numerical robustness in reducing the DEC Alpha chip's clock distribution network.

Krylov-subspace methods are best viewed as reducing the system (1) via *projection* [11]. They produce two projection matrices, $V \in \mathcal{R}^{n \times q}$ and $W^T \in \mathcal{R}^{q \times n}$, such that the reduced system is obtained as

$$\begin{aligned} \underbrace{W^T E}_{\hat{E}} \dot{x} &= \underbrace{W^T A V}_{\hat{A}} x(t) + \underbrace{W^T B}_{\hat{B}} u(t) \\ y(t) &= \underbrace{C^T V}_{\hat{C}^T} x(t) + Du(t). \end{aligned} \quad (3)$$

For the reduction to be practically meaningful, q , the size of the reduced system, must be much smaller than n , the size of the original. If the Lanczos process is used, then $W^T V \approx I$ (*i.e.*, the two projection bases are bi-orthogonal). If the Arnoldi process is applied, then $W = V$ and $W^T V = I$.

The development of Krylov-subspace projection methods marked an important milestone in LTI macromodelling. However, reduced models produced by both AWE and Krylov methods retained the possibility of *violating passivity*, or even being *unstable*. A system is passive if it cannot generate energy under any circumstances; it is stable if for any bounded inputs, its response remains bounded. In LTI circuit applications, passivity guarantees stability. Passivity is a natural characteristic of many LTI networks, especially interconnect networks. It is essential that reduced models of these networks also be passive, since the converse implies that under some situation of connectivity, the reduced system will become unstable and diverge unboundedly from the response of the original system.

The issue of stability of reduced models was recognized early in [9], and the superiority of Krylov-subspace methods over AWE in this regard also noted. Silveira et al [22] proposed a co-ordinate transformed Arnoldi method that guaranteed stability, but not passivity. Kerns et al [18] proposed reduction of admittance-matrix-based systems by applying a series of non-square congruence transformations. Such transformations preserve passivity properties while also retaining important poles of the system. However, this approach does not guarantee matching of system moments. A symmetric version of PVL with improved passivity and stability properties was proposed by Freund and Feldmann in 1996 [42].

The passivity-retaining properties of congruence transformations were incorporated within Arnoldi-based reduction methods for RLC networks by Odabasioglu et al [31, 32] in 1997, resulting in an algorithm dubbed PRIMA (Passive Reduced-Order Interconnect Macromodelling Algorithm). By exploiting the structure of RLC network matrices, PRIMA was able to preserve passivity *and* match moments. Methods for Lanczos-based passivity preservation [41, 66] followed.

All the above LTI MOR methods, based on Krylov-subspace computations, are efficient (*i.e.*, approximately linear-time) for reducing large systems. The reduced models produced by Krylov-subspace reduction methods are not, however, optimal, *i.e.*, they do not necessarily minimize the error for a macromodel of given size. The theory of balanced realizations, well known in the areas of linear systems and control, provides a framework in which this optimality can be evaluated. LTI reduced-order modelling methods based on *truncated* balanced realizations (TBR) (*e.g.*, [13, 14]) have been proposed. Balanced realizations are a canonical form for linear differential equation systems that “balance” controllability and observability properties. While balanced realizations are attractive in that they produce more compact macromodels for a given accuracy, the process of generating the macromodels is computationally very expensive, *i.e.*, cubic in the size of the original system. However, recent methods [23] that combine Krylov-subspace techniques with TBR methods have been successful in approaching the improved compactness of

TBR, while substantially retaining the attractive computational cost of Krylov methods.

3 Macromodelling Linear Time Varying (LTV) Systems

3.1 Linear Time Varying (LTV) Macromodelling

LTI macromodelling methods, while valuable tools in their domain, are inapplicable to many functional blocks in mixed-signal systems, which are usually nonlinear in nature. For example, distortion or clipping in amplifiers, switching and sampling behaviour, *etc.*, cannot be captured by LTI models. In general, generating macromodels for nonlinear systems (see Section 4) is a difficult task.

However, a class of nonlinear circuits (including RF mixing, switched-capacitor and sampling circuits) can be usefully modelled as *linear time-varying* (LTV) systems. The key difference between LTV systems and LTI ones is that if the input to an LTV system is time-shifted, it does not necessarily result in the same time shift of the output. The system remains linear, in the sense that if the input is scaled, the output scales similarly. This latter property holds, at least ideally, for the input-to-output relationship of circuits such as mixers or samplers. It is the effect of a separate local oscillator or clock signal in the circuit, *independent of the signal input*, that confers the time-varying property. This is intuitive for sampling circuits, where a time-shift of the input, relative to the clock, can be easily seen not to result in the same time-shift of the original output – simply because the clock edge samples a different time-sample of the input signal. In the frequency domain, more appropriate for mixers, it is the time-varying nature that confers the key property of frequency shifting of input signals. The time-varying nature of the system can be “strongly nonlinear”, with devices switching on and off – this does not impact the linearity of the signal input-to-output path.

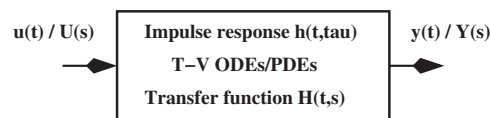


Figure 2: Linear Time Varying block

Figure 2 depicts the basic structure of an LTV system block. Similar to LTI systems, LTV systems can also be completely characterized by impulse responses or transfer functions; however, these are now functions of two variables, the first capturing the time-variation of the system, the second the changes of the input [67]. The detailed behaviour

of the system is described using time-varying differential equations, *e.g.*,

$$\begin{aligned} E(t)\dot{x} &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)^T x(t) + D(t)u(t). \end{aligned} \quad (4)$$

Time variation in the system is captured by the dependence of A , B , C , D and E on t . In many case of practical interest, this time-variation is periodic. For example, in mixers, the local oscillator input is often a sine or a square wave; switched or clocked systems are driven by periodic clocks.

The goal of macromodelling LTV systems is similar to that for LTI ones: to replace (4) by a system identical in form, but with the state vector $x(t)$ much smaller in dimension than the original. Again, the key requirement is to retain meaningful correspondence between the transfer functions of the original and reduced systems.

Because of the time-variation of the impulse response and transfer function, LTI MOR methods cannot directly be applied to LTV systems. However, Roychowdhury [45–47] showed that LTI model reduction techniques can be applied to LTV systems, by first reformulating (4) as an LTI system similar to (1), but with extra *artificial inputs* that capture the time-variation. The reformulation first separates the input and system time variations explicitly using multiple time scales [48] in order to obtain an operator expression for $H(t, s)$. This expression is then evaluated using periodic steady-state methods [20, 44, 57] to obtain an LTI system with extra artificial inputs. Once this LTI system is reduced to a smaller one using any LTI MOR technique, the reduced LTI system is reformulated back into the LTV system form (4). The use of different LTI MOR methods within this framework has been demonstrated, including explicit moment matching [45] and Krylov-subspace methods [36, 46, 47]. Moreover, Phillips [36] showed that the LTV-to-LTI reformulation could be performed using standard linear system theory concepts [67], without the use of multiple time scales.

4 Macromodelling Non-oscillatory Nonlinear Systems

While wires, interconnect, and passive lumped elements are purely linear, any mixed-signal circuit block containing semiconductor devices is nonlinear. Nonlinearity is, in fact, a fundamental feature of any block that provides signal gain, or performs any function more complex than linear filtering. Even though linear approximations of many nonlinear blocks are central to their design and intended operation, it is usually important to consider the impact of nonlinearities with a view to limiting their impact. For example, in “linear” amplifiers and mixers, distortion and intermodulation, caused solely by nonlinearities, must typically be guaranteed not to exceed a very small fraction of the output of

the linearized system. This is especially true for traditional RF and microwave designs. Such *weakly nonlinear systems* comprise an important class of blocks that can benefit from macromodelling.

Additionally, many nonlinear blocks of interest are not designed to be approximately linear in operation. Examples include digital gates, switches, comparators, *etc.*, which are intended to switch abruptly between two states. While such operation is obviously natural for purely digital systems, strongly nonlinear behaviour is also exploited in analog blocks such as sampling circuits, switching mixers, analog-to-digital converters *etc.*. Furthermore, oscillators and PLLs, which are common and basic components in mixed-signal systems, exhibit complex dynamics which are fundamentally strongly nonlinear.

Unlike for the classes of linear systems considered in the previous sections, no technique currently exists that is capable, even in principle, of producing a macromodel that conforms to any reasonable fidelity metric for *completely general* nonlinear systems. The difficulty stems from the fact that nonlinear systems are richly varied, with extremely complex dynamical behaviour possible that is very far from being exhaustively investigated or understood. This is in contrast to linear dynamical systems, for which comprehensive mathematical theories exist (see, *e.g.*, [67]) that are universally applicable. In view of the diversity and complexity of nonlinear systems in general, it is difficult to conceive of a single overarching theory or method that can be employed for effective macromodelling of an arbitrary nonlinear block. It is not surprising, therefore, that macromodelling of nonlinear systems has tended to be manual, relying heavily on domain-specific knowledge for specialized circuit classes, such as ADCs, phase detectors, *etc.*

In recent years, however, linear macromodelling methods have been extended to handle weakly nonlinear systems. Other techniques based on piecewise approximations have also been devised that are applicable some strongly nonlinear systems. As described below in more detail, these approaches start from a general nonlinear differential equation description of the full system, but first approximate it to a more restrictive form, which is then reduced to yield a macromodel of the same form. The starting point is a set of nonlinear differential-algebraic equations (DAEs) of the form

$$\begin{aligned}\dot{q}(x(t)) &= f(x(t)) + bu(t) \\ y(t) &= c^T x(t),\end{aligned}\tag{5}$$

where $f(\cdot)$ and $q(\cdot)$ are nonlinear vector functions.

4.1 Polynomial-based weakly nonlinear methods

To appreciate the basic principles behind weakly nonlinear macromodelling, it is first necessary to understand how the full system can be treated if the nonlinearities in (5)

are approximated by low-order polynomials. The polynomial approximation concept is simply an extension of linearization, with $f(x)$ and $q(x)$ replaced by the first few terms of a Taylor series about an expansion point x_0 (typically the DC solution); for example,

$$f(x) = f(x_0) + A_1(x - x_0) + A_2(x - x_0)^{\textcircled{2}} + \dots, \quad (6)$$

where $a^{\textcircled{i}}$ represents the Kronecker product of a with itself i times. When (6) and its $q(\cdot)$ counterpart are used in (5), a system of polynomial differential equations results. If $q(x) = x$ (assumed for simplicity), these equations are of the form

$$\begin{aligned} \dot{x}(t) &= f(x_0) + A_1(x - x_0) + A_2(x - x_0)^{\textcircled{2}} + \dots + bu(t) \\ y(t) &= c^T x(t). \end{aligned} \quad (7)$$

The utility of this polynomial system is that it becomes possible to leverage an existing body of knowledge on weakly polynomial differential equation systems, *i.e.*, systems where the higher-order nonlinear terms in (6) are small compared to the linear term. In particular, *Volterra series theory* [51] and weakly-nonlinear perturbation techniques [29] justify a relaxation-like approach for such systems, which proceeds as follows. First, the response of the linear system, ignoring higher-order polynomial terms, is computed – denote this response by $x_1(t)$. Next, $x_1(t)$ is inserted into the quadratic term $A_2(x - x_0)^{\textcircled{2}}$ (denoted a *distortion input*), the original input is *substituted* by this waveform, and the *linear* system solved again to obtain a *perturbation due to the quadratic term* – denote this by $x_2(t)$. The sum of x_1 and x_2 is then substituted into the cubic term to obtain another weak perturbation, the linear system solved again for $x_3(t)$, and so on. The final solution is the sum of x_1 , x_2 , x_3 and so on. An attractive feature of this approach is that the perturbations x_2 , x_3 , *etc.*, which are available *separately* in this approach, correspond to quantities like distortion and intermodulation which are of interest in design. Note that at every stage, to compute the perturbation response, a *linear* system is solved – nonlinearities are captured via the distortion inputs to these systems.

The basic idea behind macromodelling weakly nonlinear systems is to exploit this fact; in other words, to apply linear macromodelling techniques, appropriately modified to account for distortion inputs, to each stage of the relaxation process above. In the first such approach, proposed in 1999 by Roychowdhury [47], the linear system is first reduced by LTI MOR methods to a system of size q_1 , as shown in Figure 3, via a projection basis obtained using Krylov-subspace methods. The distortion inputs for the quadratic perturbation system are then expressed in terms of the *reduced* state vector of the linear term, to obtain an input vector of size q_1^2 . The quadratic perturbation system (which has the same linear system matrix, but a different input vector) is then again reduced via another projection basis, to size q_2 . This process is continued for higher order terms. The overall

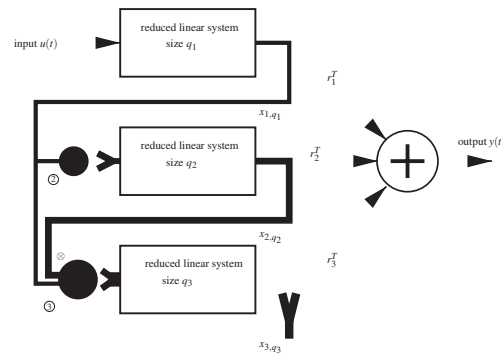


Figure 3: Block structure of reduced polynomial system

reduced model is the union of the separate reduced models with outputs summed together, as depicted in Figure 3.

By tailoring projection bases for each nonlinearly-perturbed linear system, this approach focusses on accuracy; however, this is achieved at the cost of increased macro-model size $q_1 + q_2 + \dots$. Recognizing the size issue, Phillips in 2000 [37, 38] proposed that a *single* projection basis be applied to the system (7) (analogous to LTI MOR systems), and also observed that Carlemann bilinearization [49] could be employed to obtain a canonical equation form. Intuitively, the use of a single projection basis consolidates the commonality in the three reduced models shown in Figure 3, leading to smaller overall models.

In 2003, Li and Pileggi proposed the NORM method [34], which combines and extends the above two approaches. Similar to [47], NORM generates tailored projection bases for each perturbed linear system, but instead of retaining separate macromodels as in Figure 3, it compresses these projection bases into a single projection basis. NORM then employs this single projection basis to reduce the system (7) as proposed in [38]. A particularly attractive property of NORM is that it produces a macromodel that matches a number of *multidimensional moments* of the Volterra series kernels [51] of the system – indeed, the distortion terms for each perturbed system are pruned to ensure matching of a specified number of moments. The authors of NORM also include a variant that matches moments at multiple frequency points.

4.2 Piecewise approximation methods

The polynomial approximations discussed above are excellent when the intended operation of the system exercises only weak nonlinearities, as in power amplifiers, “linear”

mixers, *etc.*. Outside a relatively small range of validity, however, polynomials are well known to be extremely poor global approximators. This limitation is illustrated in Figure 4, where it can be seen that, outside a local region where there is a good match, even a sixth-degree Taylor-series approximation diverges dramatically from the function it is meant to represent.

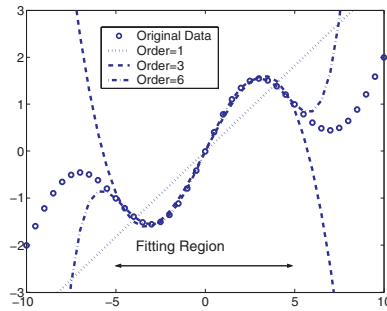


Figure 4: Limitations of global polynomial approximations

It is for this reason that other ways of approximating (5) that have better global approximation properties than polynomials have been sought. One approach is to represent the nonlinear functions $f(\cdot)$ and $q(\cdot)$ in (5) by *piecewise linear (PWL)* segments. The state space is split into a number of disjoint regions, and within each region, a linear approximation is used that matches the nonlinear function approximately within the region. By using a sufficiently large number of regions, the nonlinear function can be represented accurately over the entire domain of interest. From a macromodelling perspective, the motivation for PWL approximations is that since the system is linear within each region, linear macromodelling methods can be leveraged.

Piecewise linear approximations are not new in circuit simulation, having been employed in the past most notably in attempts to solve the DC operating point problem [17, 33]. One concern with these methods is a potential exponential explosion in the number of regions as the dimension of the state space grows. This is especially the case when each elemental device within the circuit is first represented in piecewise form, and the system of circuit equations constructed from these piecewise elements. A combinatorial growth of polytope regions results, via cross-products of the hyperplanes that demarcate piecewise regions within individual devices.

To circumvent the explosion of regions, which would severely limit the simplicity of a small macromodel, Rewiński and White proposed the Trajectory PWL method (TPWL) [43] in 2001. In TPWL, a reasonable number of “center points” is first selected

along a simulation trajectory in the the state space, generated by exciting the circuit with a representative training input. Around each center point, system nonlinearities are approximated by linearization, with the region of validity of the linearization defined *implicitly*, as consisting of all points that are closer to the given center point than to any other. Thus there are only as many piecewise regions as center points, and combinatorial explosion resulting from intersections of hyperplanes is avoided. The implicit piecewise regions in TPWL are in fact identical to the Voronoi regions defined by the collection of center points chosen.

Within each piecewise region, the TPWL approach simply reduces the linear system using existing LTI MOR methods to obtain a reduced linear model. The reduced linear models of all the piecewise regions are finally stitched together using a *scalar weight function* to form a single-piece reduced model. The weight function identifies, using a closest-distance metric, whether a test point in the state space is within a particular piecewise region, and weights the corresponding reduced linear system appropriately.

The TPWL method, by virtue of its use of inherently better PWL global approximation, avoids the blow-up that occurs when polynomial-based methods are used with large inputs. It is thus better suited for circuits with strong nonlinearities, such as comparators, digital gates, *etc.* However, because PWL approximations do not capture higher-order derivative information, TPWL's ability to reproduce small-signal distortion or intermodulation is limited.

To address this limitation, Dong and Roychowdhury proposed a piecewise polynomial (PWP) extension [25] of TPWL in 2003. PWP combines weakly nonlinear MOR techniques with the piecewise idea, by approximating the nonlinear function in each piecewise region by a polynomial, rather than a purely linear, Taylor expansion. Each piecewise polynomial region is reduced using one of the polynomial MOR methods outlined above, and the resulting polynomial reduced stitched together with a scalar weight function, similar to TPWL. Thanks to its piecewise nature, PWP is able to handle strong nonlinearities globally; because of its use of local Taylor expansions in each region, it is also able to capture small-signal distortion and intermodulation well. Thus PWP expands the scope of applicability of nonlinear macromodelling to encompass blocks in which strong and weak nonlinearities both play an important functional rôle.

We illustrate PWP using the fully differential op-amp shown in Figure 5. The circuit comprises 50 MOSFETs and 39 nodes. It was designed to provide about 70dB of DC gain, with a slew rate of $20V/\mu s$ and an open-loop 3dB-bandwidth of $f_0 \approx 10kHz$. The PWP-generated macromodel was of size 19. We compare the macromodel against the full SPICE-level op-amp using a number of analyses and performance metrics, representative of actual use in a real industrial design flow.

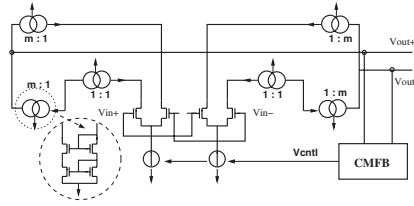


Figure 5: Current-mirror op-amp with 50 MOSFETs and 39 nodes

Figure 6 shows the results of performing DC sweep analyses of both the original circuit and the PWP-generated macromodel. Note the excellent match. Figure 7 compares Bode

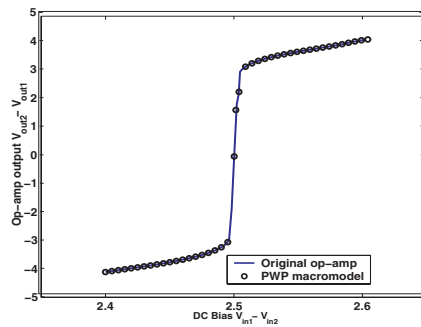


Figure 6: DC sweep of op-amp

plots obtained by AC analysis; two AC sweeps, obtained at different DC bias points, are shown. Note that PWP provides excellent matches around *each* bias point.

If the op-amp is used as a linear amplifier with small inputs, distortion and intermodulation are important performance metrics. As mentioned earlier, one of the strengths of PWP-generated macromodels is that weak nonlinearities, responsible for distortion and intermodulation, are captured well. Such weakly nonlinear effects are best simulated using frequency-domain harmonic balance (HB) analysis, for which we choose the one-tone sinusoidal input $V_{in1} - V_{in2} = A \sin(2\pi \times 100t)$ and $C_{load} = 10pF$. The input magnitude A is swept over several decades, and the first two harmonics plotted in Figure 8. It can be seen that for the entire input range, there is an excellent match of the distortion component from the macromodel vs that of the full circuit. Note that the *same* macromodel is used for this harmonic balance simulation as for all the other analyses presented. Speedups of about $8.1\times$ were obtained for the harmonic balance simulations.

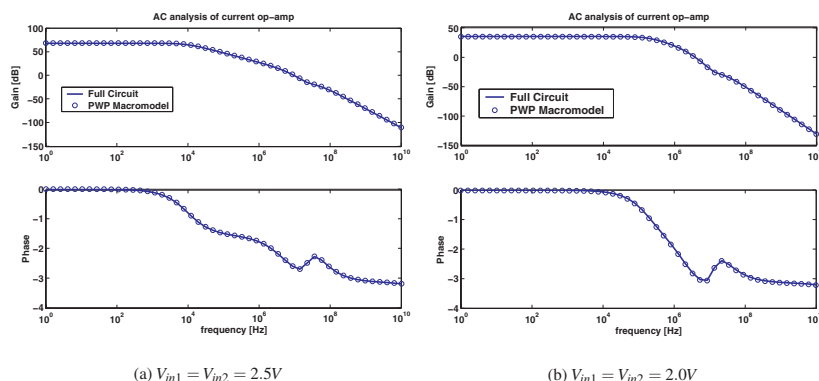


Figure 7: AC analysis with different DC bias

Another strength of PWP is that it can capture the effects of strong nonlinearities excited by large signal swings. To demonstrate this, a transient analysis was run with a large, rapidly-rising input; the resulting waveforms are shown in Figure 9. The slope of the input was chosen to excite slew-rate limiting, a dynamical phenomenon caused by strong nonlinearities (saturation of differential amplifier structures). Note the excellent match between the original circuit and the macromodel. The macromodel-based simulation ran about $8\times$ faster.

5 Macromodelling oscillatory systems

Oscillators are ubiquitous in electronic systems. They generate periodic signals, typically sinusoidal or square-like waves, that are used for a variety of purposes. From the standpoint of both simulation and macromodelling, oscillators present special challenges. Traditional circuit simulators such as SPICE [27, 40] consume significant computer time to simulate the transient behavior of oscillators. As a result, specialized techniques based on using *phase macromodels* (e.g., [1, 3, 10, 12, 24, 28, 30, 52, 58, 60]) have been developed for the simulation of oscillator-based systems. The most basic class of phase macromodels assumes a *linear relationship* between input perturbations and the output phase of an oscillator. A general, time-varying expression for the phase $\phi(t)$ can be given by

$$\phi(t) = \sum_{k=1}^n \int_{-\infty}^{\infty} h_{\phi}^k(t, \tau) i_k(\tau) d\tau. \quad (8)$$

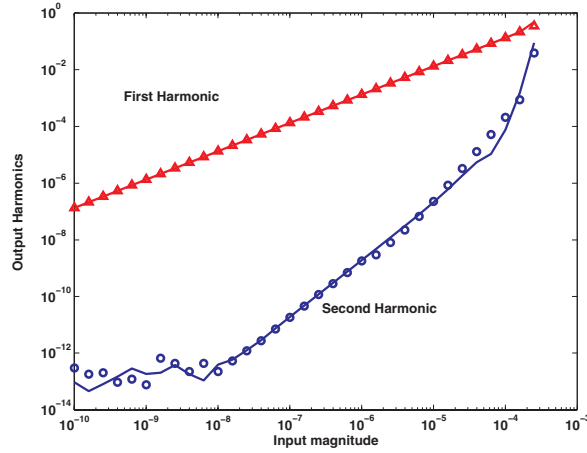


Figure 8: Harmonic analysis of current-mirror op-amp: solid line – full op-amp; discrete point – PWP model

The summation is over all perturbations i_k to the circuit; $h_\phi^k(t, \tau)$ denotes a time-varying impulse response to the k^{th} noise source. Very frequently, time-invariant simplifications of (8) are employed [19].

Linear models suffer, however, from a number of important deficiencies. In particular, they have been shown to be inadequate for capturing fundamentally nonlinear effects such as injection locking [62]. As a result, automatically-extracted nonlinear phase models have recently been proposed [5, 6, 62–64] that are considerably more accurate than linear ones. The nonlinear phase macromodel has the form

$$\dot{\alpha}(t) = v_1^T(t + \alpha(t)) \cdot b(t). \quad (9)$$

In the above equation, $v_1(t)$ is called the *perturbation projection vector (PPV)* [6]; it is a periodic vector function of time, with the same period as that of the unperturbed oscillator. A key difference between the nonlinear phase model (9) and traditional linear phase models is the inclusion of the phase shift $\alpha(t)$ inside the perturbation projection vector $v_1(t)$. $\alpha(t)$ in the nonlinear phase model has units of time; the equivalent phase shift, in radians, can be obtained by multiplying $\alpha(t)$ by the free-running oscillation frequency ω_0 .

We illustrate the uses of (9) by applying it to model the VCO inside a simple PLL [53], shown in block form in Figure 10. Using the nonlinear macromodel (9), we simulate the transient behavior of the PLL and compare the results with full simulation and lin-

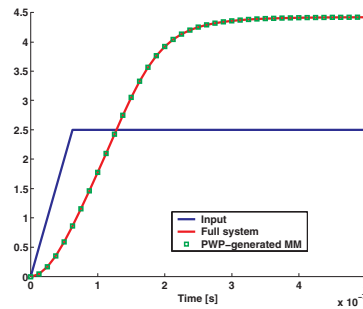


Figure 9: Transient analysis of current-mirror op-amp with fast step input

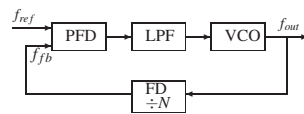


Figure 10: Functional block diagram of a PLL.

ear models. The simulations encompass several important effects, including static phase offset, step response and cycle slipping.

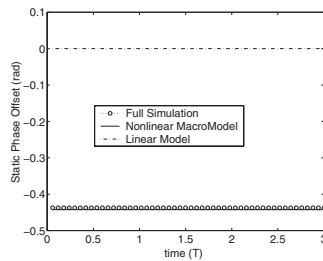


Figure 11: The static phase offset of PLL when $f_{ref} = f_0$.

Figure 11 depicts the static phase offset of the PLL when a reference signal of the same frequency as the VCO's free-running frequency is applied. The PLL is simulated to locked steady state and the phase difference between the reference and the VCO output is shown. The fact that the LPF is not a perfect one results in high-frequency AC components being fed to the VCO, affecting its static phase offset. Observe that both full simulation and the nonlinear macromodel (9) predict identical static phase offsets of about 0.43 radians.

Note also that the linear phase macromodel fails to capture this correctly, reporting a static phase offset of 0.

Figure 12 depicts the step response of the PLL at different reference frequencies. Figure 12(a) shows the step responses using full simulation, the linear phase model and the nonlinear macromodel when the reference frequency is $1.07f_0$. With this reference signal, both linear and nonlinear macromodels track the reference frequency well, although, as expected, the nonlinear model provides a more accurate simulation than the linear one. When the reference frequency is increased to $1.074f_0$, however, the linear phase macromodel is unable to track the reference correctly, as shown in Figure 12(b). However, the nonlinear macromodel remains accurate. The breakdown of the linear model is even more apparent when the reference frequency is increased to $1.083f_0$, at which the PLL is unable to achieve stable lock. Note that the nonlinear macromodel remains accurate.

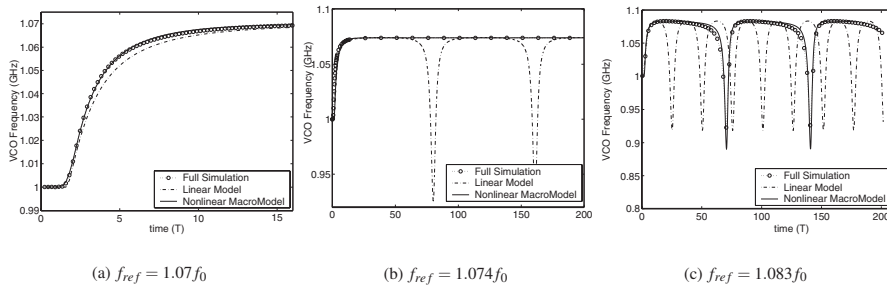


Figure 12: The step response of the PLL under different reference frequencies.

Finally, Figure 13 illustrates the prediction of cycle slipping. A reference frequency $f_{ref} = 1.07f_0$ is provided and the PLL is brought to locked steady state. When a sinusoidal perturbation of amplitude $5mA$ and duration 10 VCO periods is injected, the PLL loses lock. As shown in Figure 13(a), the phase difference between the reference signal and the VCO output slips over many VCO cycles, until finally, lock is re-achieved with a phase-shift of -2π . Both nonlinear and linear macromodels predict the qualitative phenomenon correctly in this case, with the nonlinear macromodel matching the full simulation better than the linear one. When the injection amplitude is reduced to $3mA$, however, Figure 13(b), the linear macromodel fails, still predicting a cycle slip. In reality, the PLL is able to recover without slipping a cycle, as predicted by both the nonlinear macromodel and full simulation.

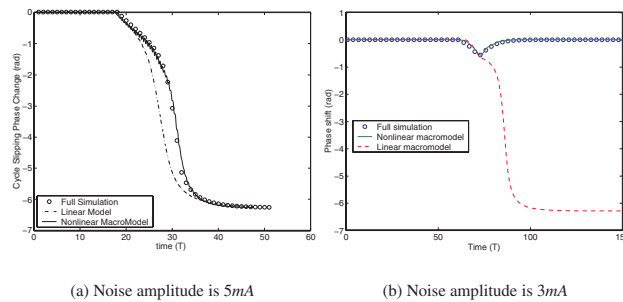


Figure 13: Cycle slipping in the PLL under different noise amplitudes.

6 Conclusion

Automated bottom-up macromodelling is rapidly becoming critical for the effective hierarchical verification of large mixed-signal systems. We have provided an overview of several of the main algorithmic macromodelling approaches available today. Linear time-invariant methods, the subject of research for more than a decade, have already proven their usefulness for interconnect analysis. Issues such as the fidelity, compactness, dynamical stability and passivity of generated macromodels have been identified and addressed. Extensions to linear time-varying systems, useful for mixers and sampling circuits, have also been demonstrated to produce useful, compact models. Interest in macromodelling nonlinear systems has grown rapidly over the last few years and a number of promising approaches have emerged. The important special case of oscillatory nonlinear circuits has, in particular, seen significant advances which are already close to being deployed in commercial CAD tools. It is very likely that further research in automated nonlinear macromodelling will translate into useful tools that are of broad practical use in the future.

Acknowledgments

This overview has relied extensively on the work of several graduate students in the Analog System Verification Group of the University of Minnesota. In particular, the sections on PWP and on oscillator/PLL macromodelling have reproduced work of Ning Dong, Xiaolue Lai and Yayun Wan [25, 26, 62–64]. I would also like to thank Joel Phillips for many in-depth discussions related to macromodelling.

References

- [1] A. Demir, E. Liu, A.L. Sangiovanni-Vincentelli and I. Vassiliou. Behavioral simulation techniques for phase/delay-locked systems. In Proceedings of the Custom Integrated Circuits Conference 1994, pages 453–456, May 1994.
- [2] E. Chiprout and M.S. Nakhla. Asymptotic Waveform Evaluation. Kluwer, Norwell, MA, 1994.
- [3] A. Costantini, C. Florian, and G. Vannini. Vco behavioral modeling based on the non-linear integral approach. IEEE International Symposium on Circuits and Systems, 2:137–140, May 2002.
- [4] D. Schreurs, J. Wood, N. Tuffillaro, D. Usikov, L. Barford, and D.E. Root. The construction and evaluation of behavioral models for microwave devices based on time-domain large-signal measurements. In Proc. IEEE IEDM, pages 819–822, December 2000.
- [5] A. Demir, A. Mehrotra, and J. Roychowdhury. Phase noise in oscillators: a unifying theory and numerical methods for characterization. IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl., 47:655–674, May 2000.
- [6] A. Demir and J. Roychowdhury. A Reliable and Efficient Procedure for Oscillator PPV Computation, with Phase Noise Macromodelling Applications. IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl., pages 188–197, February 2003.
- [7] P. Feldmann and R.W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. IEEE Trans. CAD, 14(5):639–649, May 1995.
- [8] R.W. Freund and P. Feldmann. Efficient Small-signal Circuit Analysis And Sensitivity Computations With The Pvl Algorithm. Proc. ICCAD, pages 404–411, November 1995.
- [9] K. Gallivan, E. Grimme, and P. Van Dooren. Asymptotic waveform evaluation via a lanczos method. Appl. Math. Lett., 7:75–80, 1994.
- [10] W. Gardner. Introduction to Random Processes. McGraw-Hill, New York, 1986.
- [11] E.J. Grimme. Krylov Projection Methods for Model Reduction. PhD thesis, University of Illinois, EE Dept, Urbana-Champaign, 1997.
- [12] A. Hajimiri and T.H. Lee. A general theory of phase noise in electrical oscillators. IEEE J. Solid-State Ckts., 33:179–194, February 1998.

- [13] J. Phillips, L. Daniel, and L.M. Silveira. Guaranteed passive balancing transformations for model order reduction. In Proc. IEEE DAC, pages 52–57, June 2002.
- [14] J-R. Li and J. White. Efficient model reduction of interconnect via approximate system gramians. In Proc. ICCAD, pages 380–383, November 1999.
- [15] J. Wood and D.E. Root. The behavioral modeling of microwave/RF ICs using nonlinear time series analysis. In IEEE MTT-S Digest, pages 791–794, June 2003.
- [16] K. Francken, M. Vogels, E. Martens, and G. Gielen. A behavioral simulation tool for continuous-time Δ/Σ modulators. In Proc. ICCAD, pages 229–233, November 2002.
- [17] J. Katzenelson and L.H. Seitelman. An iterative method for solution of nonlinear resistive networks. Technical Report TM65-1375-3, AT&T Bell Laboratories.
- [18] K.J. Kerns, I.L. Wemple, and A.T. Yang. Stable and efficient reduction of substrate model networks using congruence transforms. In Proc. ICCAD, pages 207–214, November 1995.
- [19] K. Kundert. Predicting the Phase Noise and Jitter of PLL-Based Frequency Synthesizers. www.designers-guide.com, 2002.
- [20] K.S. Kundert, J.K. White, and A. Sangiovanni-Vincentelli. Steady-state methods for simulating analog and microwave circuits. Kluwer Academic Publishers, 1990.
- [21] L. Hongzhou, A. Singhee, R. Rutenbar, and L.R. Carley. Remembrance of circuits past: macromodeling by data mining in large analog design spaces. In Proc. IEEE DAC, pages 437–442, June 2002.
- [22] L.M. Silveira, M. Kamon, I. Elfadel, and J. White. A coordinate-transformed Arnoldi algorithm for generating guaranteed stable reduced-order models of RLC circuits. In Proc. ICCAD, pages 288–294, November 1996.
- [23] M. Kamon, F. Wang, and J. White. Generating nearly optimally compact models from Krylov-subspace based reduced-order models. IEEE Trans. Ckts. Syst. – II: Sig. Proc., pages 239–248, April 2000.
- [24] M. F. Mar. An event-driven pll behavioral model with applications to design driven noise modeling. In Proc. Behav. Model and Simul.(BMAS), 1999.
- [25] N. Dong and J. Roychowdhury. Piecewise Polynomial Model Order Reduction. In Proc. IEEE DAC, pages 484–489, June 2003.

- [26] N. Dong and J. Roychowdhury. Automated Extraction of Broadly-Applicable Non-linear Analog Macromodels from SPICE-level Descriptions. In Proc. IEEE CICC, October 2004.
- [27] L.W. Nagel. SPICE2: a computer program to simulate semiconductor circuits. PhD thesis, EECS Dept., Univ. Calif. Berkeley, Elec. Res. Lab., 1975. Memorandum no. ERL-M520.
- [28] O. Narayan and J. Roychowdhury. Analysing Oscillators using Multitime PDEs. IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl., 50(7):894–903, July 2003.
- [29] A. Nayfeh and B. Balachandran. Applied Nonlinear Dynamics. Wiley, 1995.
- [30] E. Ngoya and R. Larchevèque. Envelop transient analysis: a new method for the transient and steady state analysis of microwave communication circuits and systems. In Proc. IEEE MTT Symp., 1996.
- [31] A. Odabasioglu, M. Celik, and L.T. Pileggi. PRIMA: passive reduced-order interconnect macromodelling algorithm. In Proc. ICCAD, pages 58–65, November 1997.
- [32] A. Odabasioglu, M. Celik, and L.T. Pileggi. PRIMA: passive reduced-order interconnect macromodelling algorithm. IEEE Trans. CAD, pages 645–654, August 1998.
- [33] Fujisawa Ohtsuki and Kumagai. Existence theorems and a solution algorithm for piecewise linear resistor networks. SIAM J. Math. Anal., 8, February 1977.
- [34] P. Li and L. Pileggi. NORM: Compact Model Order Reduction of Weakly Nonlinear Systems. In Proc. IEEE DAC, pages 472–477, June 2003.
- [35] P. Vanassche, G. Gielen, and W. Sansen. Constructing symbolic models for the input/output behavior of periodically time-varying systems using harmonic transfer matrices. In Proc. IEEE DATE Conference, pages 279–284, March 2002.
- [36] J. Phillips. Model Reduction of Time-Varying Linear Systems Using Approximate Multipoint Krylov-Subspace Projectors. In Proc. ICCAD, November 1998.
- [37] J. Phillips. Projection-based approaches for model reduction of weakly nonlinear, time-varying systems. IEEE Trans. CAD, 22(2):171–187, February 2000.
- [38] J. Phillips. Projection frameworks for model reduction of weakly nonlinear systems. In Proc. IEEE DAC, June 2000.
- [39] L.T. Pillage and R.A. Rohrer. Asymptotic waveform evaluation for timing analysis. IEEE Trans. CAD, 9:352–366, April 1990.

- [40] T.L. Quarles. Analysis of Performance and Convergence Issues for Circuit Simulation. PhD thesis, EECS Dept., Univ. Calif. Berkeley, Elec. Res. Lab., April 1989. Memorandum no. UCB/ERL M89/42.
- [41] R. Freund. Passive reduced-order models for interconnect simulation and their computation via Krylov-subspace algorithms. In Proc. IEEE DAC, pages 195–200, June 1999.
- [42] R. Freund and P. Feldmann. Reduced-order modeling of large passive linear circuits by means of the SyPVL algorithm. In Proc. ICCAD, pages 280–287, November 1996.
- [43] M. Rewienski and J. White. A Trajectory Piecewise-Linear Approach to Model Order Reduction and Fast Simulation of Nonlinear Circuits and Micromachined Devices. In Proc. ICCAD, November 2001.
- [44] M. Rösch and K.J. Antreich. Schnell stationäre simulation nichtlinearer schaltungen im frequenzbereich. AEÜ, 46(3):168–176, 1992.
- [45] J. Roychowdhury. MPDE methods for efficient analysis of wireless systems. In Proc. IEEE CICC, May 1998.
- [46] J. Roychowdhury. Reduced-order modelling of linear time-varying systems. In Proc. ICCAD, November 1998.
- [47] J. Roychowdhury. Reduced-order modelling of time-varying systems. IEEE Trans. Ckts. Syst. – II: Sig. Proc., 46(10), November 1999.
- [48] J. Roychowdhury. Analysing circuits with widely-separated time scales using numerical PDE methods. IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl., May 2001.
- [49] W. Rugh. Nonlinear System Theory - The Volterra-Wiener Approach. Johns Hopkins Univ Press, 1981.
- [50] Y. Saad. Iterative methods for sparse linear systems. PWS, Boston, 1996.
- [51] M. Schetzen. The Volterra and Wiener Theories of Nonlinear Systems. John Wiley, 1980.
- [52] M. Schwab. Determination of the steady state of an oscillator by a combined time-frequency method. IEEE Trans. Microwave Theory Tech., 39:1391–1402, August 1991.

- [53] J. L. Stensby. Phase-locked loops: Theory and applications. CRC Press, New York, 1997.
- [54] S.X.-D Tan and C.J.-R Shi. Efficient DDD-based term generation algorithm for analog circuit behavioral modeling. In Proc. IEEE ASP-DAC, pages 789–794, January 2003.
- [55] S.X.-D Tan and C.J.-R Shi. Efficient DDD-based term generation algorithm for analog circuit behavioral modeling. In Proc. IEEE DATE Conference, pages 1108–1009, March 2003.
- [56] S.X.-D Tan and C.J.-R Shi. Efficient very large scale integration power/ground network sizing based on equivalent circuit modeling. IEEE Trans. CAD, 22(3):277–284, March 2003.
- [57] R. Telichevesky, K. Kundert, and J. White. Efficient steady-state analysis based on matrix-free krylov subspace methods. In Proc. IEEE DAC, pages 480–484, 1995.
- [58] P. Vanassche, G.G.E. Gielen, and W. Sansen. Behavioral modeling of coupled harmonic oscillators. IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, 22(8):1017–1026, August 2003.
- [59] W. Daems, G. Gielen, and W. Sansen. A fitting approach to generate symbolic expressions for linear and nonlinear analog circuit performance characteristics. In Proc. IEEE DATE Conference, pages 268–273, March 2002.
- [60] L. Wu, H.W. Jin, and W.C. Black. Nonlinear behavioral modeling and simulation of phase-locked and delay-locked systems. In Proceedings of IEEE CICC, 2000, pages 447–450, May 2000.
- [61] X. Huang, C.S. Gathercole, and H.A. Mantooth. Modeling nonlinear dynamics in analog circuits via root localization. IEEE Trans. CAD, 22(7):895–907, July 2003.
- [62] X. Lai and J. Roychowdhury. Capturing injection locking via nonlinear phase domain macromodels. IEEE Trans. MTT, 52(9):2251–2261, September 2004.
- [63] X. Lai and J. Roychowdhury. Fast, accurate prediction of PLL jitter induced by power grid noise. In Proc. IEEE CICC, May 2004.
- [64] X. Lai, Y. Wan and J. Roychowdhury. Fast PLL Simulation Using Nonlinear VCO Macromodels for Accurate Prediction of Jitter and Cycle-Slipping due to Loop Non-idealities and Supply Noise. In Proc. IEEE ASP-DAC, January 2005.

- [65] Y. Qicheng and C. Sechen. A unified approach to the approximate symbolic analysis of large analog integrated circuits. IEEE Trans. Ckts. Syst. – I: Fund. Th. Appl., pages 656–669, August 1996.
- [66] Z. Bai, R. Freund, and P. Feldmann. How to make theoretically passive reduced-order models passive in practice. In Proc. IEEE CICC, pages 207–210, May 1998.
- [67] L. A. Zadeh and C. A. Desoer. Linear System Theory: The State-Space Approach. McGraw-Hill Series in System Science. McGraw-Hill, New York, 1963.

A NEW METHODOLOGY FOR SYSTEM VERIFICATION OF RFIC CIRCUIT BLOCKS

Dave Morris, Agilent EESof EDA
Cheadle Royal Business Park, Cheshire, SK8 3GR England

Abstract

This paper describes a new RFIC design flow methodology and is based upon the RFIC Reference Flow recently jointly developed by Cadence, Agilent and Helic. The flow described addresses many of the problems associated with the performance verification of RFIC circuit blocks developed for wireless systems applications. The flow is based upon the Cadence[®] Virtuoso[®] custom design platform and utilises simulation engines and additional capability provided in Agilent Technologies RF Design Environment (RFDE).

1. Introduction

A typical RFIC design project requires a diverse range of engineering skills and knowledge to bring a design to fruition. The design team will usually include engineers with system design responsibilities and others with circuit design responsibilities. It is the system designers who are responsible for interpreting the wireless system specifications, formulating the RF system architecture and defining the requirements of individual RF circuit blocks (amplifiers, mixers etc...). In the past it has been particularly difficult for the RF circuit designer to verify the performance of his transistor level designs against the wireless system specification. The methodology outlined in this paper fills this important gap in the RFIC design flow and addresses the problems of circuit and system designers having to try to correlate circuit specifications with the wireless system specification.

The methodology outlined provides RF circuit designers with the ability to perform verification simulations on their designs against both traditional circuit specifications (Gain, Noise Figure etc) and the wireless system specifications (EVM, ACPR etc).

To demonstrate the new methodology, the RF portion of an 802.11b transceiver IC has been selected as a test case. The following sections provide details of the simulation techniques employed, using time, frequency and mixed

time-frequency domain methods, to verify the performance of individual circuit blocks against both circuit specifications and system specifications.

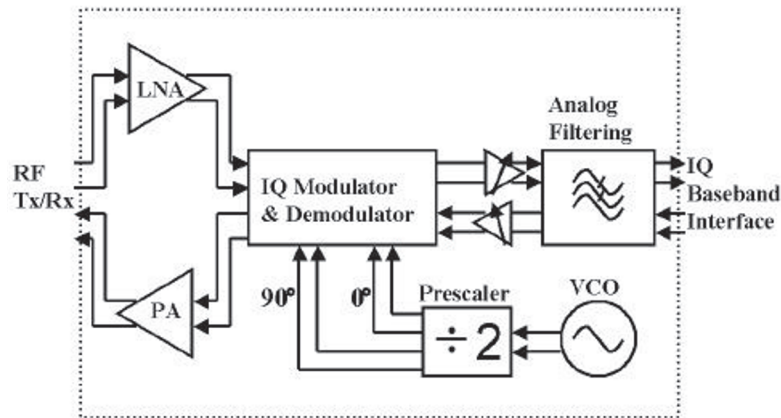


Fig.1. Overall 802.11b Transceiver Block Diagram.

2 Summary of a Typical RFIC Design Flow

The design of an RFIC presents a number of both technological and scheduling challenges. Large scale RF ICs will typically consist of a wide range of on-chip functionality including digital, analog and RF blocks. The ability to design and accurately simulate these various functional blocks efficiently is clearly important, both to reduce the time to tapeout and to increase the probability of a successful first-spin design. This paper will focus on a new methodology, which enables designers to undertake a comprehensive verification of the RF functional blocks in a seamless manner.

2.1 Top Down System Design

The wireless RFIC design process will usually start with the system designer modeling the entire RFIC. Typically behavioural modeling will be used initially to model all of the various functional blocks. In addition, the system designer must also create a testbench in which the RF stimulus is generated and meaningful measurements can be extracted. The testbench is used to simulate the entire RFIC and verify the performance of the RFIC against the key wireless specifications such as EVM (Error Vector Magnitude) or ACPR (Adjacent Channel Power Ratio). This initial simulation or series of simulations provides information about the expected ideal performance characteristics of the design, together with details of the design partitioning and requirements of the individual functional blocks. The resulting specifications for the functional RF

blocks then form the basis for the RF circuit designer to develop the transistor level implementation.

The RF circuit designer will be provided with a requirements specification, written in terms of circuit performance parameters such as gain, gain compression, noise figure etc. and will simulate a preliminary transistor level description of the circuit. The simulation method selected may be a time domain, frequency domain or mixed time-frequency domain, depending on the type of circuit (amplifier, mixer, oscillator...) and the performance parameter under investigation. Through an iterative process, and perhaps making use of optimisation techniques, the design will be refined until each of the RF functional blocks performance specifications has been satisfied. At this stage in the flow, the circuit designer has verified the performance of the RF circuit block against the functional block specifications, but has not yet verified the performance of his circuit in the wireless system. This is clearly an important gap in the verification process. In the past designers may have relied upon experience or 'rules of thumb' to correlate the circuit performance parameters with the wireless system performance parameters. For example, if the noise figure is 'X' dB, then the expected associated EVM may be 'Y' %.

The ability to seamlessly perform mixed-level simulations in which a behavioural model can be replaced by the full transistor level model and simulated in the system test bench would be highly desirable.

2.2 Bottom Up RF Circuit Verification

It is the responsibility of the RF circuit designer to develop a transistor level circuit, which provides adequate performance and satisfies the functional block requirements defined by the system engineer. Several factors are key if the RF circuit design is going to be successful. Perhaps the most fundamental requirement for success is a set of device models which are silicon accurate. Whilst foundries will typically provide simulation models as part of a process design kit (PDK), occasionally, the circuit designer may wish to supplement the PDK by creating his own models. For example an electro-magnetic (EM) simulator may be used to create a rich s-parameter model for a spiral inductor, or test chips and physical measurements may be used for the same purpose.

Assuming that the RF circuit designer has available a set of silicon accurate simulation models, traditional circuit simulation techniques may be employed to accurately predict the circuit performance. Simulations may be performed in the time domain, frequency domain, or mixed time-frequency domain. Typically, for RF design, the ability to perform frequency domain simulation using engines like Harmonic Balance can offer significant benefits.

Being a frequency domain simulator, Harmonic Balance allows frequency dependant models such as dispersive elements or s-parameter models to be easily included. Similarly there is no simulation time penalty associated with selecting widely spaced analysis frequencies. Typical practical examples of simulations which benefit from frequency domain simulation are low IF mixer simulations, intermodulation distortion simulations etc... However, even using these advanced circuit simulators, the circuit designer is still only able to verify that his circuit satisfies the circuit specifications. They do not provide the ability to verify the performance of the circuit under representative drive conditions in the wireless system application.

The wireless standards generate waveforms with unique modulation and framing structures. The same standards also require designs to be tested based on burst structure with pilot, idle, and active portions and with measurements specific to a portion of or on the composite waveform. Often these measurements require meeting specifications for different data rates and sometimes they need resolution at the bit level, requiring fully compliant parameterized sources and measurements. The possibility to run this type of simulation at the system testbench level but include a transistor level representation of the circuit / device under test (DUT) in a mixed-level system simulation is clearly highly desirable. The recent introduction by Agilent Technologies of the RFDE Wireless Test Benches (WTB) into the Cadence® Virtuoso® custom design platform now makes this possible. The WTB allows both system and circuit designers to assess the performance of a functional RF circuit block against the modulated performance figures of merit such as EVM, ACPR and BER for a particular wireless standard.

Before describing the WTB concept in further detail, it is necessary to outline the framework into which the WTB fits.

3. RF Design Environment (RFDE)

The RF Design Environment (RFDE) is intended to enable large-scale RF/mixed-signal IC design development in the Cadence® Virtuoso® custom design platform. It integrates Agilent Technologies best-in-class frequency and mixed-domain RF simulation technologies into the Cadence analog and mixed signal design flow framework.

RFDE comprises three fundamental elements, the simulation engines (ADSSim), a specialized library (adsLib) and a results viewer/post-processing facility (Data Display). The ADSSim simulators are accessed from the Cadence Analog Design Environment (ADE) in much the same way that Spectre is accessed from ADE. This arrangement is illustrated pictorially in Fig. 2

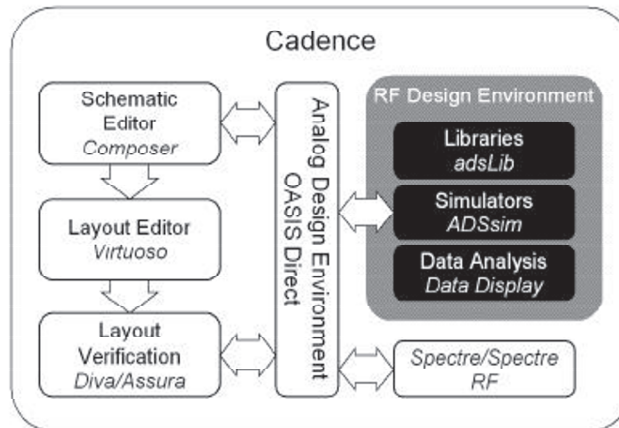


Fig.2. RFDE Integration into Cadence® Virtuoso® custom design platform

3.1. RFDE Simulation Engines

At the heart of RFDE lie the simulation engines, referred to as ADSsim. These engines provide designers with access to a range of simulation technology, which has been developed specifically for the simulation of RF circuits. The following sections provide a brief overview of each of the simulation types available under RFDE.

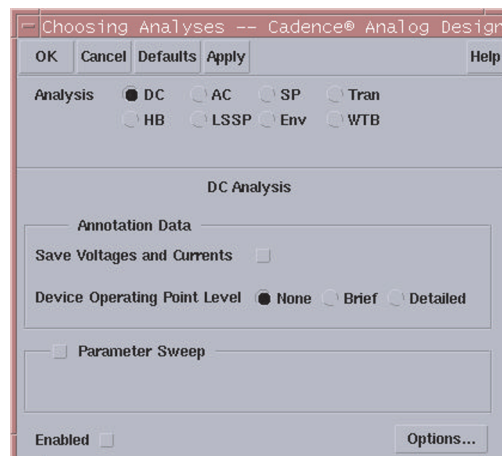


Fig.3. ADSsim simulators available within RFDE

3.1.1 AC & S-parameter Simulation

The AC/S-parameter simulators are linear frequency domain circuit simulators. These are used to analyze RF circuits operating under linear conditions. This simulation technology can be applied very effectively to the design of passive RF and small-signal active RF circuits found in many wireless applications.

3.1.2 Harmonic Balance Simulation

Harmonic Balance is a frequency domain simulator that efficiently produces steady-state results for circuits excited by periodic signals. The simulation produces spectral data directly, which does not incur a simulation time penalty based on the frequency spacing of multiple large signals, nor from having low and high frequency signals present in the circuit. Also, since the signal frequency is available at simulation time, it is possible to include model effects that are best described as a function of frequency (e.g. frequency response, dispersion). A practical example of such a model could be a spiral inductor, which could be represented as an s-parameter matrix (using either measured data or data simulated using an EM simulator)

The Harmonic Balance method assumes the input stimulus consists of a relatively small number of steady-state sinusoids. Therefore the solution can be expressed as a sum of steady-state sinusoids, which includes the input frequencies together with any significant harmonics or mixing terms.

$$v(t) = \text{real} \left(\sum_{k=0}^N V_k e^{j\omega_k t} \right)$$

Where V_k is the complex amplitude and phase at each of a limited set of N frequencies ω . The simulator converts nonlinear differential equations into the frequency domain, where it becomes a system of nonlinear algebraic equations:

$$j\omega_k * F_k(q(v(t))) + F_k(f(v(t))) + V_k * H(j\omega_k) = I_m(\omega_k)$$

Where $F_k()$ signifies the k^{th} spectral component of a Fourier transformation. The harmonic balance simulator must then simultaneously solve this system of N nonlinear algebraic equations for all the V_k values. The nonlinear devices are still evaluated in the time domain by using an inverse Fourier transformation to convert the V_k values into the $v(t)$ waveform prior to evaluating the nonlinear $q()$ and $f()$ functions. This means that standard SPICE-type, nonlinear current and charge waveforms are transformed into the

frequency domain at each iteration so their spectral values can be used in the frequency domain equations. Since most harmonic balance simulators use Newton-Raphson techniques, the derivatives (nonlinear resistance and capacitance) must also be computed in the time domain and transformed into the frequency domain.

The circuit node voltages take on a set of amplitudes and phases for all frequency components. The currents flowing from nodes into linear elements, including all distributed elements are calculated by means of straightforward frequency-domain linear analysis. Currents from nodes into nonlinear elements are calculated in the time-domain. A frequency-domain representation of all currents flowing away from all nodes is available. According to Kirchoff's Current Law (KCL), the currents should sum to zero at all nodes. The probability of obtaining this result on the first iteration is extremely small.

Therefore, an error function is formulated by calculating the sum of currents at all nodes. This error function is a measure of the amount by which KCL is violated and is used to adjust the voltage amplitudes and phases. If the method converges (that is, if the error function is driven to a given small value), then the resulting voltage amplitudes and phases approximate the steady-state solution.

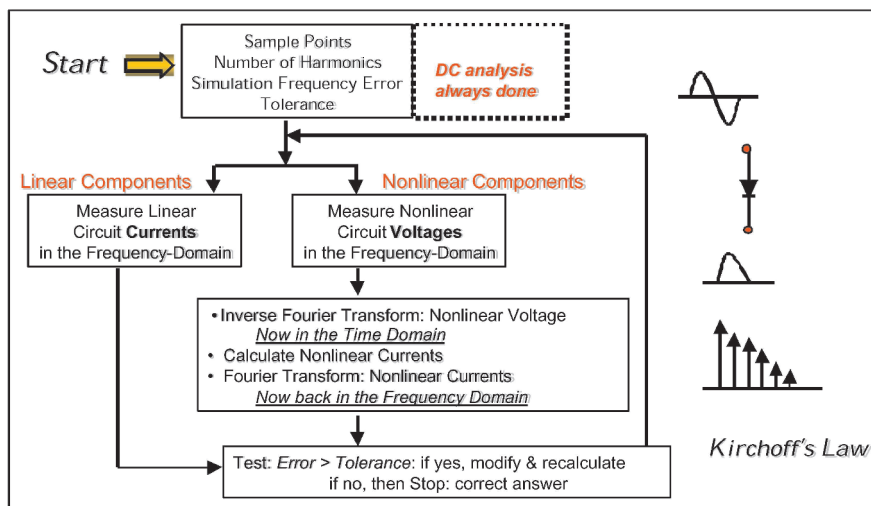


Fig.4. Harmonic Balance Simulation Flowchart

In the context of high frequency circuit and system simulation, harmonic balance has a number of advantages over conventional time-domain analysis:

- Designers are usually most interested in a system's steady-state behavior. Many high frequency circuits contain long time constants that require conventional transient methods to integrate over many periods of the lowest-frequency sinusoid to reach steady state. Harmonic balance on the other hand, captures the steady-state spectral response directly.
- The applied voltage sources are typically multi-tone sinusoids that may have very narrow or very widely spaced frequencies. It is not uncommon for the highest frequency present in the response to be many orders of magnitude greater than the lowest frequency. Transient analysis would require integration over an enormous number of periods of the highest-frequency sinusoid. The time involved in carrying out the integration is prohibitive in many practical cases.
- At high frequencies, many linear models are best represented in the frequency domain. Simulating such elements in the time domain by means of convolution can result in problems related to accuracy, causality, or stability

Typical practical applications of Harmonic Balance include the simulation of non-linear noise, gain compression, harmonic distortion, and intermodulation distortion in circuits such as power amplifiers and mixers. In addition Harmonic Balance lends itself well to oscillator analysis.

3.1.3 Circuit Envelope Simulation

Circuit Envelope's technology permits the analysis of complex RF signals by employing a hybrid time and frequency domain approach. It samples the modulation envelope (amplitude and phase, or I and Q) of the carrier in the time domain and then calculates the discrete spectrum of the carrier and its harmonics for each envelope time samples. Thus, the output from the simulator is a time-varying spectrum, which may be used to extract useful information. When compared to solutions using time domain, circuit envelope is most efficient when there is a large difference between the high-frequency carrier and the low-frequency time variation.

Circuit envelope simulation combines elements of harmonic balance and time-domain simulation techniques. Like harmonic balance, circuit envelope simulation describes the nonlinear behavior of circuits and the harmonic content of the signals. Unlike harmonic balance, however, circuit envelope simulation extends over time. It is not constrained to describe steady-state behavior only. In

effect, circuit envelope simulation depicts a time-varying series of harmonic balance results.

$$v(t) = \text{real} \left(\sum_{k=0}^N V_k(t) e^{j\omega_k t} \right)$$

In circuit envelope simulation, input waveforms are represented as RF carriers with modulation envelopes that are described in the time domain. The input waveform consists of a carrier term and a time-varying term that describes the modulation that is applied to the carrier.

Amplitude, phase, and frequency modulation, or combination of these can be applied, and there is no requirement that the signal be described as a summation of sinusoids or steady state. This makes it possible to represent digitally modulated (pseudo random) input waveforms realistically.

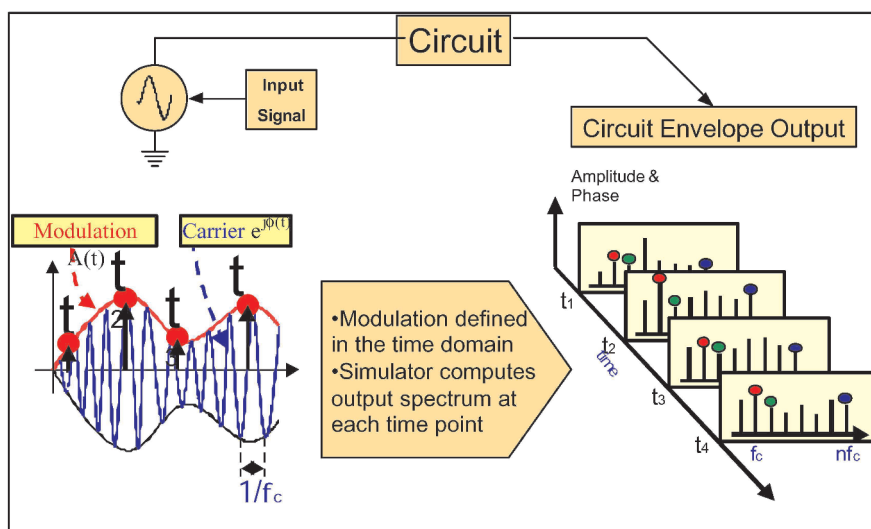


Fig.5. Circuit Envelope Simulation

In circuit envelope simulation, the discrete time-varying spectrum of the carrier and its harmonics is simulated at each time point over the time interval. If the circuit includes frequency mixing intermodulation terms are also computed at each time point. Amplitude and phase data at each time point in the simulation is then saved into the simulation results file. These results, in the time domain, show amplitude, phase, I/Q, frequency, and harmonics as a function of time for the output and any other node of the circuit if desired.

By taking the Fourier transform of the amplitude and phase data from the simulation of any spectral component (for example the fundamental), frequency domain results around that spectral component can be presented. The Fourier transform is used (in effect) to convert the amplitude and phase data from the simulation back into the frequency domain. This makes it possible to examine results such as spectral regrowth, adjacent-channel power, intermodulation distortion, and phase noise.

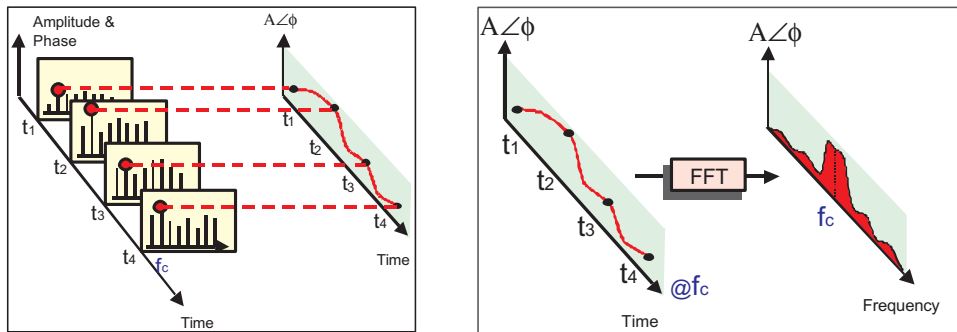


Fig.6. Circuit Envelope Time Domain & Frequency Domain Results

Typical practical applications of Circuit Envelope include the analysis of amplifier and mixer circuits operating in pulsed RF environments or digitally modulated RF environments. Measurements of spectral regrowth and adjacent channel power ratio (ACPR) may be studied. Other circuit types that are analysed efficiently in circuit envelope include phase locked loops (PLL's), Gain Control Loops (AGC's), Voltage Controlled Oscillators (VCO's), Voltage Controlled Amplifiers (VCA's), and Modulators.

3.1.4 Transient/Convolution Simulation

The Convolution Simulator is an advanced time-domain simulator that extends the capability of High Frequency SPICE by accurately simulating frequency-dependent components (distributed elements, S-parameter data files, transmission lines, etc.) in a time-domain simulation. The Convolution Simulator evaluates high-frequency effects such as skin effect, dispersion, and higher frequency loss.

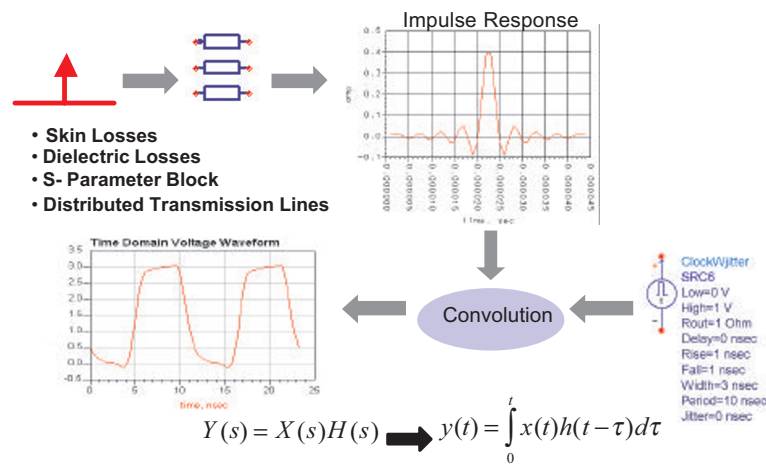


Fig.7. Convolution Simulation

Convolution converts the frequency-domain information from all the distributed elements to the time-domain, effectively resulting in the impulse response of those elements. The time-domain-input signals at the element's terminals are convolved with the impulse-response of the element to yield the output signals. Elements that have exact lumped equivalent models –including nonlinear elements – are characterized entirely in the time domain without using the impulse responses.

Typical application examples include analyzing transient conditions where the effects of dispersion and discontinuities are significant, and observing the effects of off-chip elements, and chip-to-board interactions in IC simulations.

3.1.5 Momentum Simulation

Momentum is a planar electromagnetic (EM) simulator that enables RF and microwave designers to significantly expand the range and accuracy of their passive circuits and circuit models. The ability to analyze arbitrary shapes, on multiple layers and to consider real-world design geometries when simulating coupling and parasitic effects, makes Momentum an indispensable tool for customized passive circuit design. The simulator is based on the Method of Moments (MoM) technology, which is particularly efficient for analyzing planar conductor and resistor geometries. The method of moments (MoM) technique is

based upon the work of R.F. Harrington. It is based on older theory that uses weighted residuals and variational calculus. Detailed information on the method of moments and Green's theorem can be found in Field Computation by Moment Methods [1].

Momentum may be accessed from the Cadence Virtuoso layout environment and may be used to compute S-, Y-, and Z-parameters of general planar circuits. These EM accurate models can then be used directly in RFDE circuit simulators including Harmonic Balance, Convolution, Circuit Envelope and Wireless Test Bench.

3.1.6 Wireless Test Bench

A Wireless Test Bench (WTB) is a collection of pre-configured parameterised sources, measurements, and post-processing setups based on published specifications of a wireless standard, packaged in an executable simulation flow.

Wireless Test Benches offer system-level wireless signal sources and standards measurements from within the Cadence® Virtuoso® custom design platform. Several pre-configured wireless test benches are available as an RFDE option, and provide fully parameterized sources and measurements to help meet today's complex wireless standards. Currently these include WLAN, 3GPP and TD-SCDMA.

Additional flexibility allows System architects to develop their own customized test benches early in the development cycle and export them from Agilent's Advanced Design System (ADS) into RFDE. RFIC circuit designers can then access the test benches from within the Cadence® Virtuoso® custom design platform to verify their circuit designs against the wireless system specifications.

3.2 The adsLib library

RFDE provides an additional library of components for application with ADSsim. These components include a number of simulation related components such as signal sources including controlled, frequency domain, time domain, noise and modulated sources. These sources are typically used in harmonic balance or circuit envelope analysis. In addition, the library also contains a number of simulation models, which help extend the scope of simulation to include some off-chip circuitry. For example an extensive library of models is included for standard transmission line constructs such as microstrip, stripline etc. The key benefit of having these models available is that it becomes

relatively easy to include some off-chip design, distributed matching for example and simulate this together with the on-chip design.

3.3 The Data Display

Following a simulation using one of the RFDE simulation engines, the simulation results are written to a file called a dataset file. The Data Display is the environment used to access the simulation results and display the data in an appropriate way, for example data may be plotted on a scalar plot, a smith chart, a text listing etc. In addition the data display environment also provides the capability to post process simulation data by allowing the user to create mathematical expressions to extract useful performance measurements.

4 WLAN 802.11b Transceiver Simulations

The transmit section of an 802.11b transceiver IC will be used as a vehicle to demonstrate the RF system verification flow described in the previous sections. The circuits & simulations described in the following sections, form part of the RFIC Reference Flow recently jointly developed by Cadence, Agilent and Helic.

Some of the key steps associated with the simulation and verification of a number of individual RF functional blocks will now be described. Initially verification will be performed against the circuit specification parameters. An important extension to the verification process will also be described which allows the circuit designer to verify the circuit performance against the RF wireless system (802.11b) specification parameters.

The RF functional blocks utilised in this 802.11b transmit chain include a Voltage Controlled Oscillator (VCO), frequency divider, transmit mixer, baseband chain, and power amplifier. Details of the RF design partitioning can be seen in Fig 8.

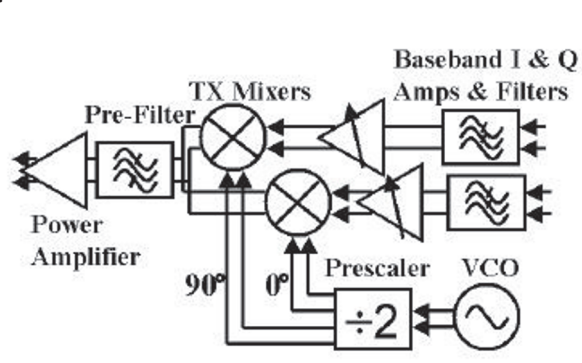


Fig.8. WLAN 802.11b Transmit Chain Block Diagram.

4.1 VCO Simulations

Oscillator circuits are well suited to being simulated in the frequency domain using the harmonic balance simulator available with RFDE. Oscillators are a unique class of circuits - they have no inputs (other than DC sources) and generate an RF signal at some frequency. The goal of the simulation is to determine both the spectrum of the output signal as well as its frequency.

Two alternative methods may be employed in the course of setting-up an oscillator simulation using RFDE harmonic balance. Either the designer may place a special simulation component into the oscillator feedback loop, or the designer must identify a nodal connection in the simulation setup. In the former case, the simulator uses the formal theory of loop gain to determine the point of oscillation where the loop gain = $1+j0$. Whilst this technique is robust, it can cause problems downstream in the RFIC design flow, specifically during layout versus schematic (LVS) checking. In the later case, a special voltage source is connected to the specified node(s) by the simulator. This source generates a voltage at only the fundamental frequency of oscillation; it does not load the circuit at any other frequency. The oscillator analysis then adjusts the frequency and injected voltage until the current supplied by this source is zero. Theoretically this is solving for the point where the admittance (current injected divided by voltage value) of the source is zero. When the source supplies no current but a non-zero voltage is present, this is the point at which the oscillation is self-sustaining.

Harmonic balance is also able to handle parameter sweeps (such as tuning voltage) very efficiently because the harmonic balance engine utilises the simulated results from parameter n as a starting point for subsequent simulation using parameter $(n+1)$.

Being a frequency-domain simulator, harmonic balance is also able to utilise frequency-domain models, which can be highly desirable. For example, in this case an on-chip spiral inductor is used in the resonator structure of the VCO and the enhanced model accuracy of the s -parameter model over an equivalent circuit model may prove vital for a silicon accurate simulation. The s -parameter model could originate from various sources. For example it may come from network analyzer measurements on a test chip, or it may come from an electromagnetic simulation. In this case the RFDE EM simulator (Momentum) was utilised to create an s -parameter model for the inductor. The ability to set-up and run Momentum simulations directly from the Cadence Virtuoso Layout environment simplifies the process of model creation considerably by avoiding the need to transfer layout information between disparate tools.

In setting up the VCO simulation, the designer must specify an initial guess for the frequency of oscillation. This initial guess does not need to be specified particularly accurately, because a harmonic-balance search algorithm will vary the operating frequency to find the oscillation frequency.

In this simulation, the tuning voltage, V_t , is being swept to assess the tuning characteristic of the oscillator. The simulator performs a separate oscillator analysis for each swept variable value, then attempts to find the actual oscillation condition for each value. Following the analysis, it is possible to display the frequency of oscillation and the output power of that frequency as a function of the tuning voltage V_t .

The results of the VCO analysis are provided in Fig.9, this data display shows the waveforms, frequency versus tuning voltage, output spectrum and tuning sensitivity. Typically a phase noise simulation would follow this type of basic oscillator analysis.

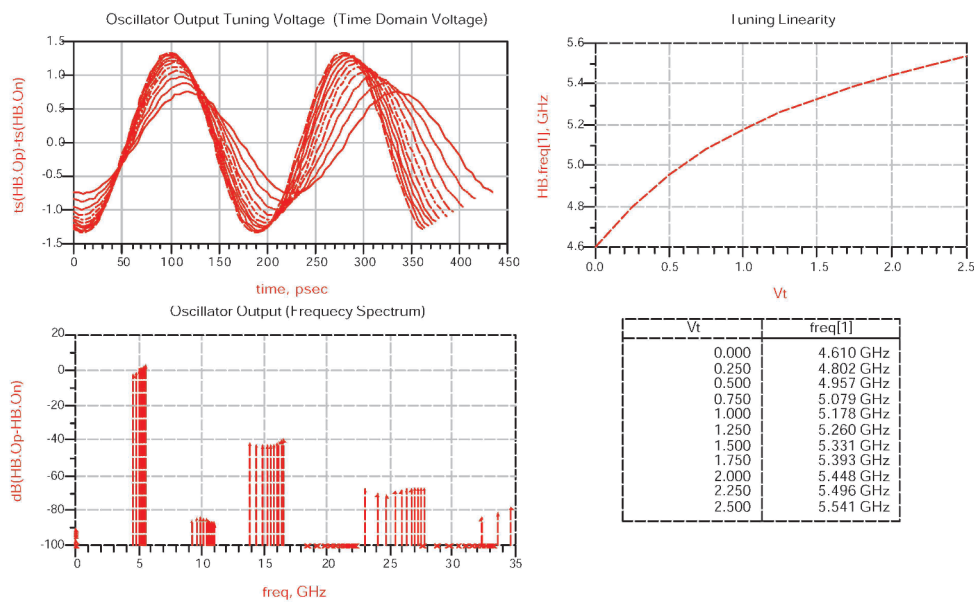


Fig.9. VCO Simulation Results

4.2 Upconverter Mixer Simulations

There are a number of simulations that a designer may wish to run on the upconverting mixer. These include voltage conversion gain versus input (IF) amplitude, voltage conversion gain versus LO amplitude, as well as input and output-referred third-order intercept point simulations. Harmonic balance is particularly well suited for mixer simulations, which often require closely spaced input tones for IP3 simulation. Also, the IF or baseband frequency may be orders of magnitude lower than the RF frequency, but harmonic balance simulation time is independent of the spacing between analysis frequencies.

4.3 Power Amplifier Simulations

RFDE provides the RF circuit designer with a range of capabilities that are useful through the whole design and verification phase of the project. This section will focus on the development and verification of a power amplifier (PA) block to illustrate some of the key steps in the design and verification simulations.

During the early stages of the design, it is likely that the designer will be interested in simulating fundamental characteristics of the selected transistor such as I-V curves and Gm-versus-bias and possibly performing load pull, and stability analysis simulations. After the preliminary design has been developed, the PA performance will be verified against traditional circuit parameters such as gain, gain compression, TOI, noise figure etc. Importantly, RFDE also provides a means for the designer to extend the level of verification by allowing simulation of the PA transistor level circuit using a realistic representation of the complex modulated RF signal encountered in the wireless system. System level measurements such as EVM, BER or ACPR will typically be made following such a simulation.

Sections 4.3.1 & 4.3.2 provide details of how RFDE simulation may be used in the early stages of the PA design, to assess biasing, stability and impedance matching requirements. An iterative process of refinement would then be used to create the finished design. This process is not described in this paper. Sections 4.3.3 and 4.3.4 provide details of how the finished PA design might be simulated to verify compliance against the circuit level specifications. Section 4.3.5 provides details of how RFDE may be used to verify the performance of the finished PA against the WLAN (802.11b) system specifications.

4.3.1 I-V Curve Simulations

This section illustrates an I-V curve simulation of the PA device. The results shown in Fig.10 were achieved using the DC and s-parameter simulation capability provided in RFDE. In this case, parametric sweeps have been used to sweep both drain and gate voltages in order to generate the IV curves. Note that the RFDE data display environment is used to display simulated results. Post processing of simulated results is also possible by writing mathematical expressions into the data display environment.

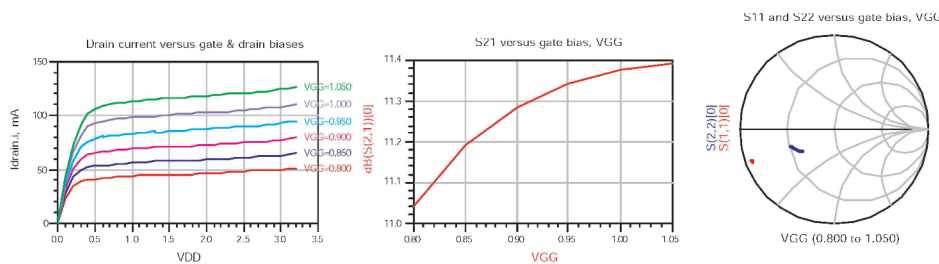


Fig.10. PA Device: Simulated I-V Curve Data

4.3.2 Load-Pull Contour Simulations

Power amplifier designers often perform load pull simulations on their devices to determine the complex load impedance required to maximise power delivered, maximize power-added efficiency, or minimize intermodulation distortion, etc [2].

The simulation setup illustrated in Fig.11 shows the PA output FET with a capacitor and resistor connected between the gate and drain to improve stability, ideal bias networks, and a source and load. The bias networks have current probes and wire names, for calculating the DC power consumption, which is required for computing the power-added efficiency (PAE). There are also current probes and wire labels at the RF input, for computing the input impedance, and at the RF output, for computing the power delivered to the complex load impedance.

Harmonic balance simulation will be used for this analysis. During the simulation both the phase and magnitude of the load reflection coefficient are swept. Note that only the load reflection coefficient at the fundamental frequency is being swept. The reflection coefficients at harmonic frequencies are set to

arbitrary user defined values (usually impedances close to an open or a short give best power-added efficiency).

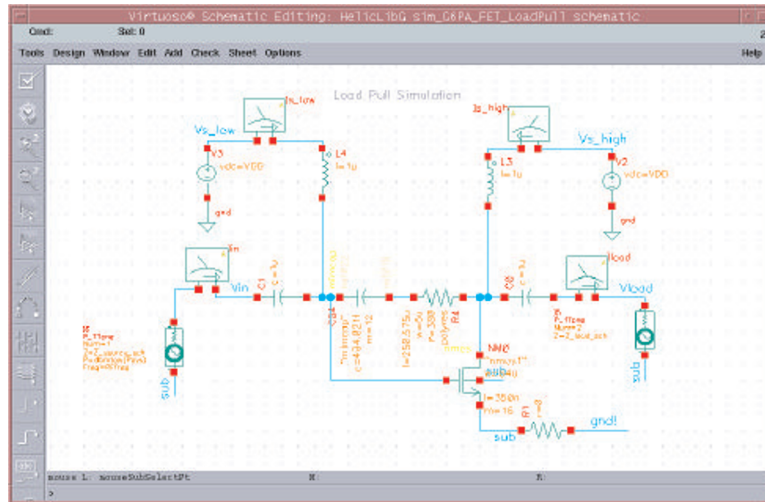


Fig.11. Load Pull Simulation Schematic.

Typically when running a load pull simulation, a designer might start out with a coarse sampling of the entire Smith Chart, and then limit the ranges of the phase and amplitude sweeps. A 1-tone harmonic balance simulation is run for each combination of reflection coefficient phase and magnitude.

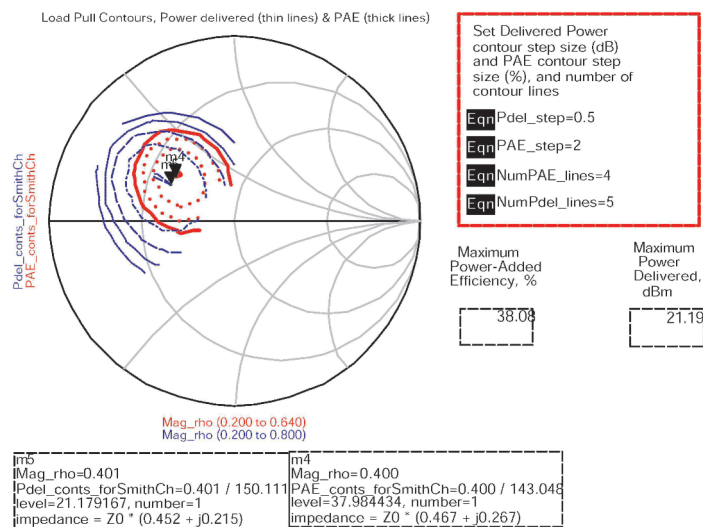


Fig.12. Load Pull Simulation Results –Power Delivered & PAE Contours

4.3.3 Power Amplifier Verification : Gain, 1-dB Compression, PAE

In this section we will see how RFDE simulation engines may be used effectively to simulate the finished power amplifier circuit. We are interested here in verifying that the PA performance satisfies a number of key circuit performance specifications such as 1-dB gain compression point, third order intercept point etc.

Firstly, a simulation will be undertaken to examine the gain compression of the PA. Harmonic Balance is used to sweep the available RF source power and measurements are made on the corresponding RF output power. This simulation provides the designer with details of the small-signal gain, the 1-dB compression point and the shape of the compression characteristics (i.e. Is it a slow roll-over into compression or a hard compression). In this case the RF source power is swept from -30 to -12.5 dBm using 25dB steps, then from -11 to -4 dBm using 1dB steps.

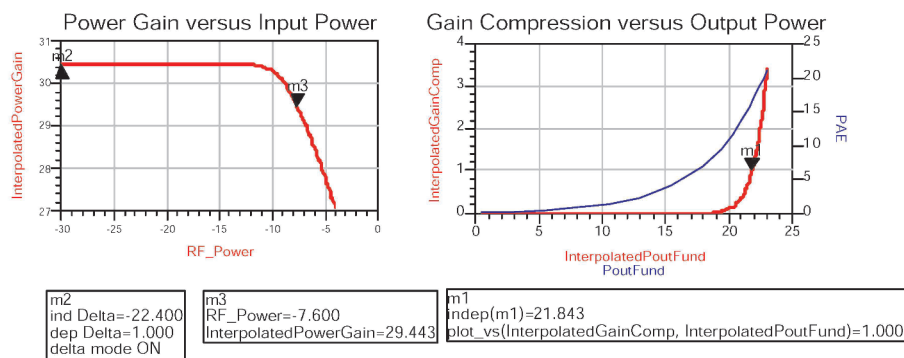


Fig.13. PA Gain Compression & PAE Simulation

Following the simulation, the gain compression results may be plotted in different ways. Fig.13 illustrates how some simple equations can be used to calculate and plot the gain compression as a function of either RF input or RF output power level. Similarly, equations have been used to calculate the power added efficiency (PAE) of the amplifier at any given RF input/output power level. From these results we are able to extract the following circuit specification measurements.

- Small Signal Power Gain = 30.4dB
- 1-dB Gain Compression Point = 21.8dBm
- PAE (@ 1-dB GCP) = 16%

4.3.4 Power Amplifier Verification : Two Tone TOI

In this section we will see how RFDE simulation engines may be used effectively to extend the verification of the power amplifier design by performing a two tone intermodulation simulation to assess the third order intercept point (TOI) performance. Typically intermodulation simulation/testing is performed with two or more tones, closely spaced in frequency. In such situations, the harmonic balance simulator provides a particularly efficient simulation method since the harmonic balance simulation time is independent of both the absolute frequencies specified for the test tones and the test tone spacing.

To simulate the TOI (third-order intercept) point of the PA, a similar arrangement will be used to that described in section 4.3.3. However for this simulation, two large-signal tones will be used as the RF input test signal. The two input tones are applied at frequencies 2.450GHz and 2.451GHz. The harmonic balance analysis is setup to sweep the available composite source power from -30 to -15 dBm in 2.5 dB steps, and then from -14 to -8 dBm in 1 dB steps.

The harmonic balance simulation setup provides the user with the ability to specify how many harmonics of the fundamental test tones to include in the analysis. In addition, the user may specify the highest order intermodulation (mixing) product to include in the simulation. Increasing the order will lead to a more accurate simulation, but will also require more time and memory. For this simulation, we are interested in the third order products and so the highest order intermodulation product was set to 5.

Fig.14 shows the simulated input and output-referred TOI points, in addition to the small-signal power gain, PAE, DC power consumption, and gain compression. The output power at the 1-dB gain compression point occurs at a lower level with two input tones, than with one input tone. Note from the plots of the fundamental and 3rd-order terms versus input power that slope of the 3rd-order term varies dramatically from the classically predicted 3:1 gradient line, especially as the amplifier is driven into compression. It is important to ensure that the IP3 point is computed from values of the fundamental and 3rd-order terms that are well below the compression point. Otherwise the TOI point, which is computed from an extrapolation, will be incorrect. For this amplifier, the output-referred TOI point is about 36 dBm, and the input-referred TOI point is about 5.8 dBm.

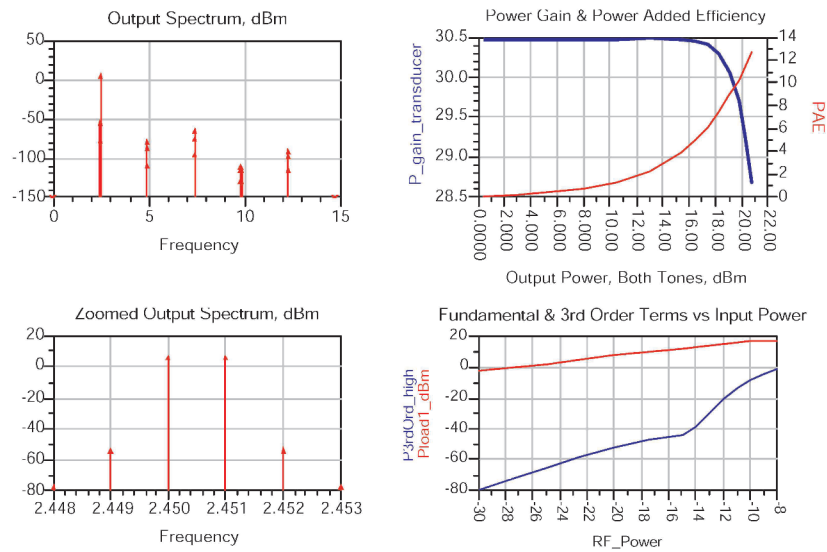


Fig.14. PA Gain Compression, PAE and TOI Simulation

As the RF input signals become large enough to drive the amplifier into compression, the third-order intermodulation distortion becomes higher than would be predicted by extrapolation from the TOI point. This implies that the distortion generated by the transistor level amplifier may be worse than predicted by a behavioural model using just a TOI parameter to model the intermodulation distortion. In other words, although we appear to be satisfying the circuit specifications for the power amplifier, it is possible that the behavioural modeling used in the top-down system level simulations was not sufficiently rich to capture some potential problems and there is a possibility that the power amplifier will not function correctly in the WLAN 802.11b system application.

4.3.5 Power Amplifier Verification in WLAN 802.11b System

In the previous sections we verified the performance of the power amplifier design against circuit level performance parameters. We have also identified the possibility that behavioural modeling used by the system designer during the top-down design phase may not be adequate to accurately model the distortion generated by the actual transistor level power amplifier. This in turn leads to uncertainty about whether the power amplifier distortion will lead to non-compliance against the wireless system level specifications.

This section describes an additional verification step, which can now be taken to confirm whether or not that the power amplifier will function correctly in the wireless system application. This step will utilise the Wireless test Bench (WTB) simulation capability in RFDE. The WLAN WTB utilises test signals and measurements defined in the standards [3-7]. Fig.15 illustrates the top-level schematic created for the WTB simulation. Notice that the schematic contains neither signal sources nor measurement terminations. In fact this schematic represents only the device under test (DUT), which in this case is simply the power amplifier circuit. Notice also that the schematic uses ideal baluns on the RF input & output to interface the differential inputs and outputs of the power amplifier to the WTB sources and measurements.

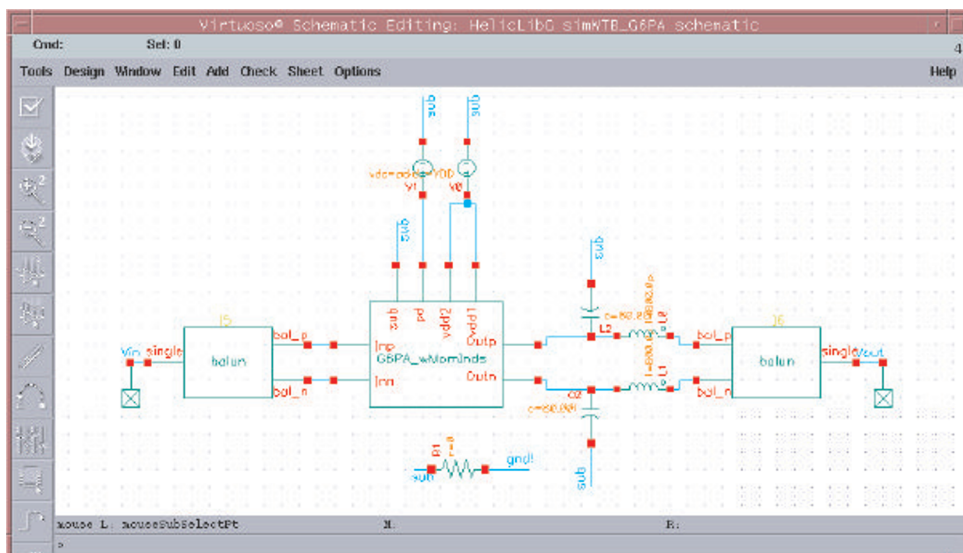


Fig.15. Power Amplifier Top Level Schematic

The WTB simulation capability is accessed from the Analog Design Environment (ADE) in the same way as any of the other RFDE simulators. Having selected the WTB simulator, the user is presented with the simulation setup dialogue box detailed in Fig.16



Fig.16. WTB Simulation Setup

Currently the WTB provides verification capability for WLAN, 3GPP and TD-SCDMA standards. In addition, custom WTBs can be created in Agilent Technologies Advanced Design System (ADS) and exported from ADS for use in the Cadence ADE.

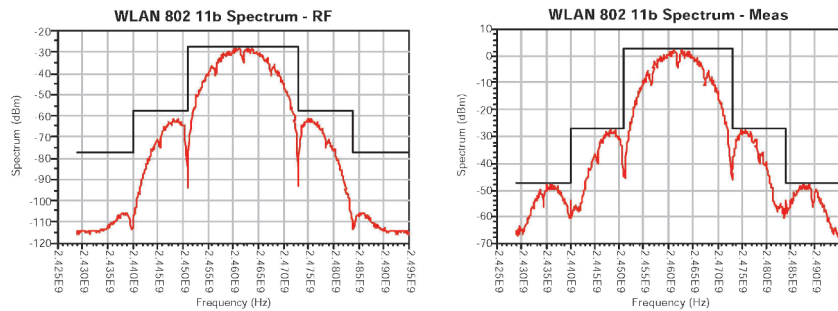
The setup dialogue prompts the user to define the node on the DUT that should be connected to the modulated RF signal source, and to select the node on the DUT that should be connected to the termination & measurement portion of the WTB. A further series of setup options allows the designer to decide precisely which measurements he would like to perform.

In this example the RF source and measurement frequency was set to 2.462GHz, channel 11 of the WLAN 802.11b standard. Initially the power amplifier was simulated with an input RF power level of -10dBm.

Following the WTB simulation a pre-configured data display window opens. Fig.17 illustrates the spectrum measured on the output of the power amplifier. With a -10dBm RF input level, the amplifier is operating just below the 1-dB compression point and the spectrum is within the specification limits indicated by the mask. The EVM measurement was also available following the simulation and in this case was 2.2%.

WLAN_802_11b_TX Test Bench - Spectrum Measurement

real(RF_Source) / (1 MHz)	real(RF_Power_dBm)	real(RF_R)	...FMeasurement) / (1 MHz)	real(Meas_R)
2462.000	-10.000	50.000	2462.000	50.000



WLAN Specification: Transmitted Spectrum Request on IEEE Std 802.11b-1999, section 18.4.7.3

Power levels in dB_r (relative to the maximum spectral density of the signal)

Frequency offset (MHz)	<=-22	-11	0	11	>=22
Power level (dB _r) <=	-50	-30	0	-30	-50

Fig.17. WTB Simulation Results : Spectrum Measurement@-10dBmInput

The simulation was then repeated using an input power level of -9 dBm. This corresponds to operating the power amplifier at its 1-dB compression point. In this case the spectrum fails to stay within the mask limits and the EVM has increased to 2.7%. One simple conclusion from this analysis is that the PA should not be operated at or beyond the 1-dB compression point in this WLAN 802.11b application.

5. Conclusions

This paper has outlined some of the key features and benefits available to RFIC circuit designers utilising Agilent Technologies RFDE simulators within the Cadence[®] Virtuoso[®] custom design platform. In particular, a new methodology has been described which allows RFIC circuit designers to close the verification gap by simulating transistor level circuit designs using realistic representations of the complex modulated RF signals encountered in wireless system applications. Thus allowing the circuit performance to be evaluated in terms of system level parameters such as EVM, BER or ACPR.

Acknowledgments

Many thanks to Andy Howard, Application Engineer, Agilent EEsof EDA, 1400 Fountain Grove Parkway, Santa Rosa, CA 95403-1738, USA for providing much of the material used in section 4

References

- [1] R. F. Harrington, Field Computation by Moment Methods. Maxillan, New York (1968).
- [2] S.C Cripps, RF Power Amplifiers for Wireless Communications, Artech House (1999)
- [3] IEEE Std 802.11a-1999, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz Band," 1999.
- [4] ETSI TS 101 475 v1.2.1, "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer," November, 2000.
- [5] IEEE P802.11g/D8.2, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Further Higher Data Rate Extension in the 2.4 GHz Band," April, 2003.
- [6] CCITT, Recommendation O.151(10/92).
- [7] CCITT, Recommendation O.153(10/92)

PLATFORM-BASED RF-SYSTEM DESIGN

Peter Baltus

Philips Semiconductors Advanced Systems Labs
and Eindhoven University of Technology
Eindhoven, The Netherlands

Abstract

This paper describes a platform-based design approach for RF systems. The design approach is based on a common, modular architecture, a collection of reusable, configurable RF building blocks, and a method for implementing transceivers using these building blocks.

1. Introduction

The number of systems that use radio links is increasing quickly, and, in parallel to that, the number of standards for such radio links is increasing quickly as well. Recently, we have seen for example WLAN systems based on IEEE 802.11b, 802.11a, 802.11g, 802.11n standards, WMAN systems based on IEEE 802.16a, IEEE802.16d, and IEEE 802.16e standards, WPAN based on the Bluetooth standard and its extensions to medium and high data rates, ultra-wideband (UWB) in two fundamentally different types (multi-band OFDM and time domain) as proposed for standardization in IEEE 802.15.3a, Zigbee, cordless standards such as DECT, PHS, CT0, CT1, CT2, cellular standards such as GSM in 450MHz, 480MHz, 850MHz, 900MHz (in standard, extended, and railway variants) bands, DCS 1800, PCS 1900, with or without GPRS and EDGE extensions, AMPS, IS-95, IS-98, IS-136, UMTS, PDC in high and low bands, CDMA2000 and TD-SCDMA. This list is just a subset of relatively recent and popular communication standards. In the area of broadcast standards, a similar growth in radio link standards is ongoing.

This creates new challenges for system and semiconductor designers. The most obvious one is the challenge to develop systems and components for each of these different standards in time and with limited resources. To complicate matters, many applications require multiple radio links, either for different purposes (e.g. Bluetooth in a cellular phone) or for compatibility with different systems at various locations (e.g. different types of cellular phone networks in different countries). The number of radio links that a typical consumer expects

to find in especially handheld equipment is increasing as well. Modern high-end cellular phones typically include multi-band cellular radio links in addition to WLAN and WPAN radio links, as well as FM radio broadcast receivers. In the near future, this will probably be extended with television reception (DVB-H and/or DVB-T) and GPS radio links. Also, near-field communication links, ultra-wideband links and WMAN radio links are expected to appear in handheld terminals in the not-too-distant future.

The introduction of multiple radio links in a single device, such as a handheld terminal, is often not as simple as putting multiple radio components next to each other. There are at least four issues to consider:

1. Co-existence of multiple radio links.
2. Interference between transceivers.
3. Antenna placement, interaction, and interface with multiple transceivers.
4. Optimization/sharing of resources.

These issues can be solved more easily by a combined and integrated design of the multiple radio links. With N radio link types, however, this could ultimately require the design of 2^N combinations of radio transceivers. Since the number N is already becoming so big that developing the individual radio transceivers in time and with limited resources is becoming non-obvious, the development of 2^N multi-mode multi-band radio transceivers using traditional custom design approaches for each design is impossible.

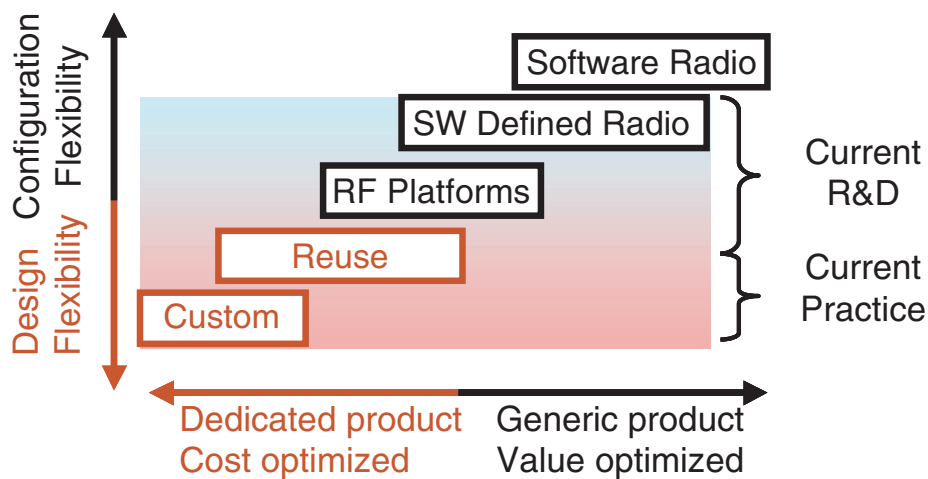


Figure 1: Solutions for multi-mode multi-band transceivers

Several approaches are being considered to solve this problem (Figure 1), ranging from the current approaches such as full-custom design and various forms of reuse, to a full software radio consisting of an antenna, data converters,

and all remaining signal processing in software. These approaches, and the transceivers developed using them, have different trade-offs with respect to cost and flexibility. Custom design tends to result in transceivers with the ultimate cost optimisations, since there is almost unlimited flexibility in the design phase. Even the dimensions and layout of individual devices can be optimised using this approach. However, the investment in effort and time needed to develop a single transceiver in this way is larger than with any other method.

Reuse is a way to reduce the development effort and time of transceivers by reusing parts of previous transceiver developments. In the strict sense, this would be “black-box” IP reuse, where circuits are reused without changing (or even knowing) their internal implementation. This results in reduced flexibility during design, and can also result in a transceiver with a higher cost than a fully custom-designed transceiver, in return for a faster and cheaper development effort. Because the transceiver can now enter the market earlier, it can have a higher value, depending on the price erosion of the specific market. In practice, RF reuse is often implemented in a less strict manner, where the circuits to be reused are modified and optimised to get a better trade-off between time & effort invested and cost & performance of the result. It is expected that this practice will shift over time towards more strict and formal reuse methods, enabled by a growing experience with reuse, better reuse methods and processes, and a higher level of acceptance by individual designers.

On the other extreme are software defined radio and software radio, that aim to provide a single product solution to multiple (and ultimately all) radio requirements. In software radios, all signal processing is done completely in software, whereas in software defined radio the signal path is still partially in analog and/or dedicated digital hardware. In this case, the parameters of the signal path can be adjusted through software. Both software radio and software defined radio offer a different type of flexibility than custom designed transceivers and transceivers based on reuse: the properties of the radio can be changed after it has been fabricated, allowing it to be adapted to new requirements without the need to modify the hardware. Especially in cases where fabrication is slow and/or expensive, this flexibility adds a lot of value in allowing the quick and cheap development of multi-mode multi-band transceivers.

It is even possible to change the properties of a transceiver during operation, to deal with changes in environment, properties of the signal, or even to communicate with systems based on a different standard. This flexibility does come at a price, however, since the extra complexity of software defined radios and software radios results in extra cost, higher power dissipation and reduced performance. These disadvantages are likely to decrease over time because of

improvements in technology. Also, when the number of modes and bands that needs to be supported increases, the ease of resource sharing in software defined radios and software radios will probably compensate for the extra complexity at some point in the future. For now, software defined radios are subject of research projects, and software radios are still further into the future.

Platform-based transceivers are an intermediate solution in terms of flexibility and cost/value optimisation. Platform-based RF design will be discussed in more detail in the section 2. The system aspects of platform-based RF design will be discussed in section 3.

2. Platform-based RF design

Platform-based design has become an accepted approach for *digital* circuits. It is a logical next step when reuse of sub circuits by itself does not provide a sufficiently fast time-to-market. A direct translation of the platform-based design method for digital circuits to RF design is not likely to be successful, however, because of a number of basic differences between digital and RF design:

1. Digital design is based on the robustness to noise, interference and non-linearity that is inherent to the processing of quantized, typically binary, signals. Distortion is completely irrelevant digital circuits, the desired signal level is typically equal to the maximum signal level, and noise levels and crosstalk can be as high as 20dBs below the maximum signal. In RF circuits, the noise level is much further below the maximum signal, in many cases 80dB or more, and the desired signal can be around the noise level while interferer levels exceed the desired signal by several orders of magnitude. This puts very high requirements on the linearity and noise performance of RF circuits, often close to the limits that can be achieved in the IC technology used.
2. Digital design is based on the robustness to delay inherent to the processing of time-discrete signals. This is achieved at the expense of a much large margin between clock frequencies and the unity-gain bandwidth of the individual devices. Typically, the clock frequency is less than 1% of the unity-gain bandwidth. RF circuits often operate with signals around 20% of the unity-gain bandwidth of individual devices, and therefore have a much smaller margin between desired and achievable performance.
3. The complexity in terms of circuit elements in RF circuits tends to be much less than in digital circuits. Whereas a typical RF transceiver circuit has in the order of 1000 devices, modern microprocessors use around 6

orders of magnitude more devices to implement their highly complex required functionality.

Generally, the complexity of digital circuits is in the management of the desired functionality, whereas the complexity of RF circuits is in the management of undesired behavior. The platform-based approaches to digital design target an improvement in the management of desired behavior. Since the desired behavior of typical RF circuits is very simple, these approaches add little value to existing RF development methods. Obviously, a different approach to platform-based RF design is needed, that focuses on the management of undesired behavior.

The sensitivity for this undesired behavior in RF circuits is caused by the small margins in time and amplitude between desired and achievable performance. This makes very accurate modeling and simulation of all (parasitic) properties of RF circuits crucial. Since there are so many possible parasitic effects that could potentially affect RF performance, it is not (yet) feasible to model all of them accurately. As a result, there is often a significant discrepancy between the performance of a circuit as predicted by a circuit simulator and the measured performance of the actual IC. “First time right” IC design is much more usual for digital IC’s than for RF IC’s. The iterations through fabrication usually dominate the overall development time. Figure 2 illustrates this, using typical numbers for RF IC design projects.

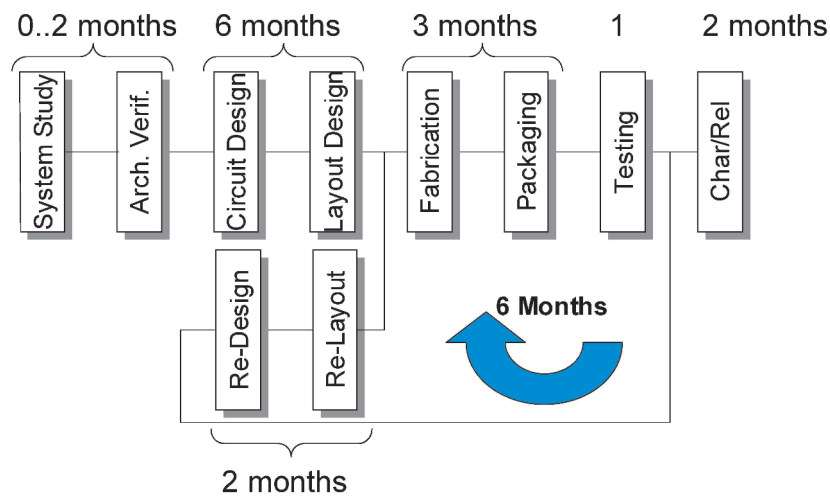


Figure 2 Typical RF transceiver design timeline

The deficiencies in modeling can be divided into two categories:

1. Deficiencies in the modeled behavior of an individual sub circuit.
2. Deficiencies in the modeled interaction between multiple sub circuits.

The first category can be eliminated by a strict reuse approach at the sub circuit level. If only proven sub circuits are used in a design, of which all relevant parameters are accurately known, then there should be no surprises of the first category when measuring the fabricated IC. In practice, such strict reuse seldom occurs, since circuits need to be adapted to a new technology, or to different requirements of a new transceiver¹.

The second category includes effects such as substrate crosstalk, temperature effects and gradients, power supply noise generation and sensitivity, interaction through package parasitic elements such as bondwires and pins, impact of load and source impedances (especially when these are non-linear), parasitic effects of interconnect between sub circuits, etc.

Both categories can be addressed with a platform-based design method. A platform-based RF design method consists of:

- A common, modular architecture.
- A library of reusable, configurable building blocks that are optimized for use in this architecture.
- A method for developing multi-mode multi-band transceivers using these IP blocks and architecture.

The common architecture ensures that the building blocks fit together without modifications, enabling strict reuse of these blocks. Also, the number of blocks required for covering a significant application area is reduced, since only a single architecture needs to be supported. Finally, the building blocks can be transferred to a new technology and verified with limited effort once such a technology becomes available.

The impact on time-to-market of the interaction between sub circuit blocks is addressed by assembling the sub circuits into transceiver through system-in-package (SiP) integration rather than monolithic integration (Figure 3). This has two effects:

1. The interaction between the blocks is limited because of the larger physical separation between the blocks, as well as the division of power supply nets and lack of a common (monolithic) substrate.
2. Any remaining interactions can be addressed at the SiP integration level, without the need to iterate over fabrication loops through the IC fab. Since SiP fabrication tends to be much faster than monolithic integration, this further reduces the time-to-market. In addition, with the increasing cost of IC mask sets, the development cost will be reduced as well.

¹ An additional reason is the “not-invented-here syndrome” that some organizations and individuals suffer from. Although this is a very serious consideration, it is outside the scope of this paper.

The cost of a SiP transceiver, using appropriate technologies and design approaches, tends to be similar to a monolithic transceiver. Even if the cost is slightly higher, the difference is usually a lot less than the price erosion in the time gained by the improved time-to-market. When the market matures, next generations of the product can be more integrated towards a monolithic transceiver.

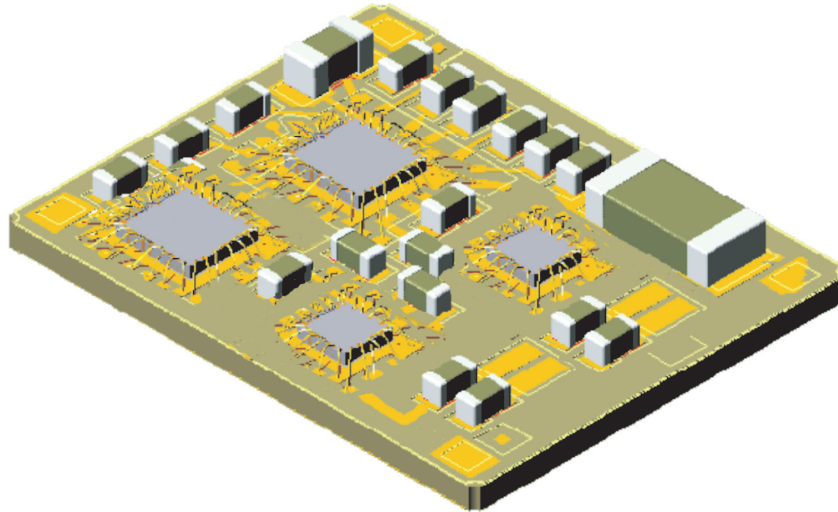


Figure 3 Drawing of a typical SiP with individual building blocks assembled on a substrate and surrounded by SMD components.

When defining a platform, a very important choice is the scope. The scope can be defined as the combination of all specifications for which a transceiver implementation based on the platform can be made. In design space, the specification of such a transceiver is represented by a single point (Figure 4). The scope of a platform therefore is a collection of such points, referred to as a “scope specification cloud” in the context of this paper. The specification of a transceiver depends on the specifications of the building blocks. For example, the noise figure of the transceiver depends on the gains and noise figures of the individual building blocks. For a specific architecture, there is a unique translation from building block specifications to transceiver specification.

The inverse is not true, however: the same transceiver specification can be achieved by an infinite number of combinations of specifications for the building blocks. Therefore, there are degrees of freedom in deriving the specification clouds of the building blocks from the platform scope specification cloud. Since it is important to keep the number of building blocks low to reduce

initial investment and maintenance costs of the building block library, these degrees of freedom can be used to reduce the number of building blocks. This can be accomplished in two steps:

1. Using the degrees of freedom in the decomposition of the specification cloud, the points in the design space for the individual building blocks are clustered together in a small number of groups.
2. The building blocks are configurable, so that a single block can achieve the specifications represented by the multiple points in such a group, if the points are not too far apart.

The specification parameters of a configurable block are represented by a cloud in design space (Figure 4). Using these two steps, the total number of blocks can be drastically reduced.

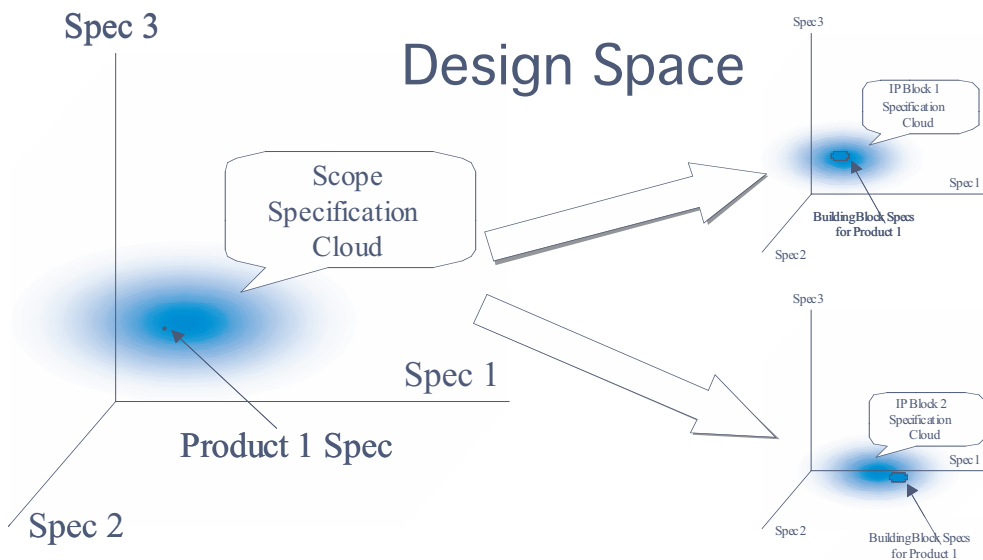


Figure 4 Design space for the platform and for the individual building blocks.

One common aspect of design methods that exceed the full custom level is that these methods can be split into two parallel processes and corresponding design flows (Figure 5):

1. The generation of IP blocks.
2. The design of a product based on these IP blocks.

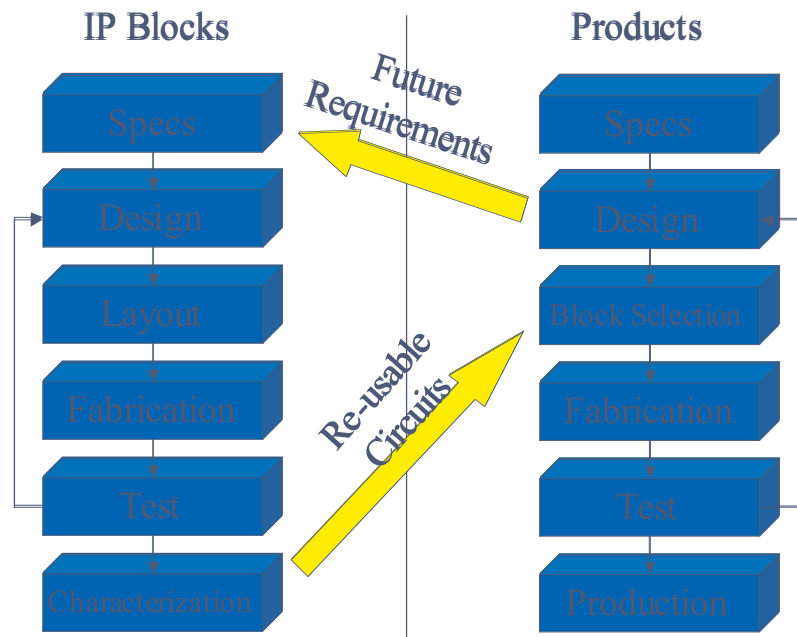


Figure 5 Platform-based design flows for developing building blocks (IP blocks) and for developing transceivers (products) based on these building blocks

For platform-based RF design, the first process is very similar to the IC design of transceiver sub circuits. The main differences are:

- The circuits need to be configurable.
- The sub circuits need to be designed as stand-alone building blocks on individual IC's.
- The specifications for the circuits are derived through the platform scope decomposition as described above.

Design of configurable RF circuits in the context of platform-based RF circuit design is described in [1] and [2]. The configuration of these circuits is achieved through configuration pins that change the parameters of the circuit depending on the value of these pins. Platform-based RF system design for the development of transceiver products using these configurable RF circuits will be discussed in the next section.

3. Platform-based RF system design

A number of new problems at RF system design level need to be solved when using a platform-based approach. These include:

- Defining a common, modular architecture.

- Creating specifications for a minimum set of building blocks to cover the platform scope.
- Transceiver system simulations using behavioral models based on measured performance of the building blocks.
- Characterization of the building blocks.
- Validation of the block specifications.
- Selection of the best blocks and configuration settings to implement transceiver specifications

These problems and their solutions will be discussed in individual sub-sections in the remainder of this section. As an example, a platform scope is used in these subsections consisting of transceivers for the following standards:

- GSM-450, GSM-480, GSM-850, GSM-900, GSM-900 extended, GSM-900 railways, DCS-1800, PCS-1900, with and without GPRS and EDGE extensions
- IS-91 (NAMPS, CAPS), IS-95/98 (digital mode), IS-136 (digital mode)
- DECT
- Bluetooth
- WLAN IEEE 802.11b, 802.11g, 802.11a (USA/EU/Japan modes)
- GPS (receive-only)
- All multi-mode multi-band combinations of these standards

By grouping related standards together, this list can be reduced to 15 single standards and their multi-mode multi-band combinations.

3.1. Defining a common, modular architecture

A common architecture for platform-based RF systems design needs to support all transceivers within the scope of the platform with acceptable cost, size and performance. This can be achieved by checking each potential architecture exhaustively against all points in the platform scope specification cloud. Obviously, existing transceivers that meet the requirements for a point in the platform scope can be used to skip the related checks. In this way it was determined that a zero-IF/low-IF architecture can be used as a common architecture for the scope defined at the beginning of this section. A super-heterodyne architecture would also meet the performance requirements of this scope, but doubts with respect to cost and size of the total transceiver (including external filters) made this a less desirable alternative.

The granularity of the building blocks in this common architecture is a trade-off between flexibility and cost efficiency. Small building blocks enable more flexibility (with in the extreme case individual transistors and passive components), whereas larger building blocks can be more cost-efficient, especially if the building blocks become so small, and the number of connections so large, that the building blocks become bondpad limited. This

might also increase the assembly cost due to the many interconnections required. For the platform scope described at the beginning of this section, the optimum trade-off was found to be at around 5 building blocks per transceiver SiP for the assembly (die- and wirebonding) and (laminated) substrate technologies currently available.

Given this number of building blocks, the partitioning boundaries need to be defined in such way that the number of connections between the modules is minimal, since this will reduce assembly cost, bondpad overhead of the building blocks, and power dissipation of the interfaces between the building blocks. The potential advantage of using different technologies for different building blocks was also taken into account. The result is shown in Figure 6. The transceiver is partitioned into:

- A down-converter (LNA/mixer)
- An up-converter (upmixer/driver)
- An LO generator (VCO/synthesizer)
- A power amplifier (PA)
- An IF block (filter/VGA)

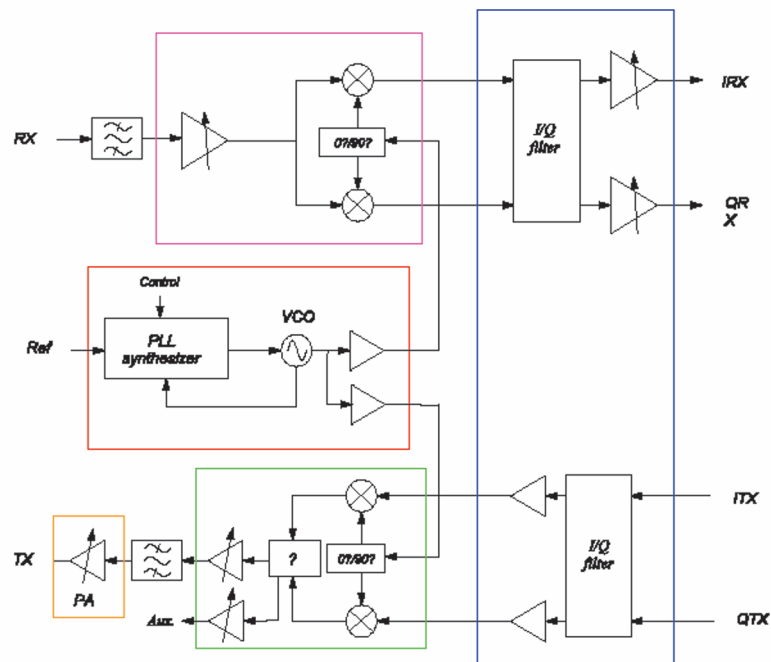


Figure 6 Partitioning of the common, modular architecture

The interfaces between the blocks need to be standardized in order to allow flexible combinations of different blocks. For practical reasons, all interfaces have been defined as $100\times$ balanced signals ($2 \times 50\times$). The balancing reduces emission of signals and susceptibility to interference from other signals. The $100\times$ impedance level allows for easy measurements and characterization. It also ensures that transmission lines for this impedance can be realized on almost any substrate technology.

The interfaces in this architecture are:

- Antenna to down-converter
- Up-converter to PA
- PA to antenna
- IF block to and from baseband
- Down-converter to IF block
- IF block to up-converter
- LO generator to down-converter
- LO generator to up-converter

From these interfaces, the first four also exist in traditional single-chip transceivers with an external PA. The bottom four are extra, but two of them are low-frequency and can easily be implemented in a power-efficient way. This leaves only two extra RF interfaces compared to a traditional transceiver partitioning, and therefore it can be expected that there will be only a very limited increase in power dissipation because of these interfaces. For multi-mode multi-band transceivers, the number of building blocks can quickly increase, depending on the amount of sharing required. An optimized granularity, e.g. by combing up- and downconverters with the LO generator in a single block, provides better flexibility versus efficiency trade-offs for such transceivers.

3.2. Creating specifications for a minimum set of building blocks

There is not yet an algorithm for finding the minimum number of configurable building blocks that, together, cover a scope specification cloud. Such an algorithm would require a formal description of the performance trade-offs that can be achieved through configurability. This could be based on [3]. It would also require an estimate of the cost in terms of performance, power dissipation, and chip area as a function of the configuration range. If not for this cost, a single building block with an infinite configuration range would be the obvious solution. A universal configuration versus cost trade-off function has not been found yet either, although work in this area is being carried out for specific circuits [2].

Since an algorithm for finding the minimum number of blocks is not yet available, a set of building blocks was defined by a small team of very experienced RF designers. They studied the specification clouds of the for many individual parameter combinations. Examples of specification clouds for such parameter combinations are shown in Figure 7 and Figure 8. They then developed scenarios for the decompositions of the scope cloud into specification clouds for the building blocks, using their experience with previous transceiver circuits and sub circuits to estimate realistic configuration ranges.

Without any grouping, the 15 standards would have required 75 building blocks. The team of RF designers ended up with 19 blocks, using conservative estimates of achievable configuration ranges. It is expected that work as described in [1] and [2] will result in further reduction of this number. Even so, the result is already quite useful since it results in a reduction of the development effort by a factor of about 4 for single-mode single-band transceivers, and a factor of about 59 when including all multi-mode multi-band transceivers in this scope specification cloud. This shows the reuse potential of platform-based RF system design.

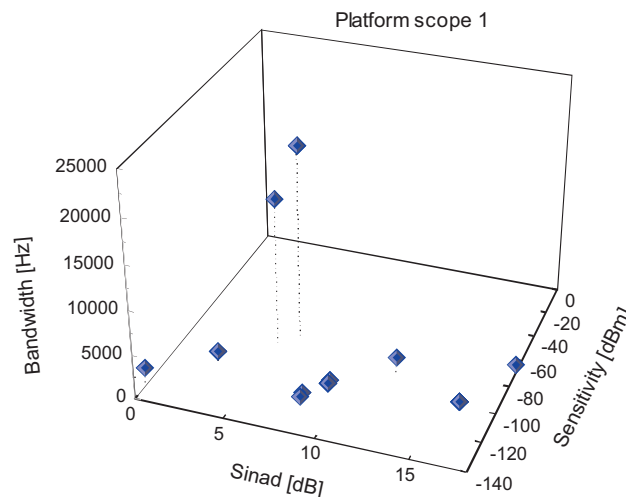


Figure 7 Scope specification cloud for receiver Sinad, sensitivity, and bandwidth

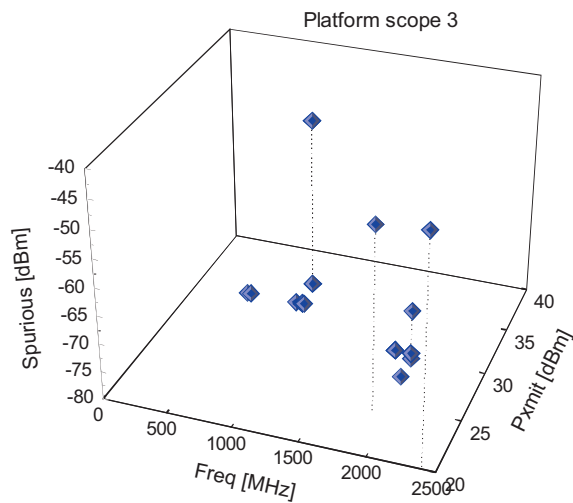


Figure 8 Scope specification cloud for transmitter frequency, output power and spurious

3.3. Transceiver system simulations using behavioral models

When designing a transceiver at the system level using a platform-based approach, simulation of the transceiver at the transistor level is quite unattractive for several reasons:

- It is less efficient than simulation using behavioral models at the building block level, especially when running complex simulations such as bit-error rate simulations including the baseband.
- It is less accurate because the transistor level models are based on measurements of individual devices and do not model all parasitic effects such as substrate, temperature, interconnect etc. (see section 2). Behavioral models can use parameters based on measurements of the actual building blocks, and include the effects of all parasitic elements in the circuit.
- In the future, building blocks might be shared between companies, and it might become important to prevent disclosure of the internal structure of such blocks, while still sharing the resulting behavior.

Fortunately, a lot of work has already been carried out on RF behavioral models [4][5][6], and many EDA suppliers provide RF behavioral model support in their system simulators. These models are typically of a somewhat smaller granularity than the building blocks described in this paper, but this can be worked around by combining several of these models into a sub circuit which then serves as the behavioral model for the building block. This approach is used

in ADS as shown in Figure 11. The parameters for these models are obtained through characterization based on measurements of the actual building blocks, and stored in tables that contain the relevant parameter values for each combination of configuration parameters.

3.4. Characterization of the building blocks

The characterization of the building blocks requires a lot of measurements, since each of the performance parameters of the building block needs to be measured at each combination of configuration parameter settings. Even with only two configuration pins, and only ten values of each configuration parameter, all performance parameters need to be measured 100 times. With 19 building blocks, this is only practical using an automated measurement set-up. The power supply and configuration pin voltages are set through programmable power supplies. The input signals are generated by a pair of signal generators that can be modulated through an AWG, in order to cover the different standards. Two generators are used to allow for two-tone IP3 measurements. A third generator provides the LO signal for the up-converter and down-converter building blocks. Alternatively, the input signal can be generated by a noise source for noise figure measurements through an automatic RF switching network. The output signal can be analyzed, again through an automatic RF switching network, by a spectrum analyzer. The measurement results are automatically saved in a file that can be used as input for the table-driven behavioral models in the system simulator. A diagram of the automated measurement set-up is shown in Figure 9.

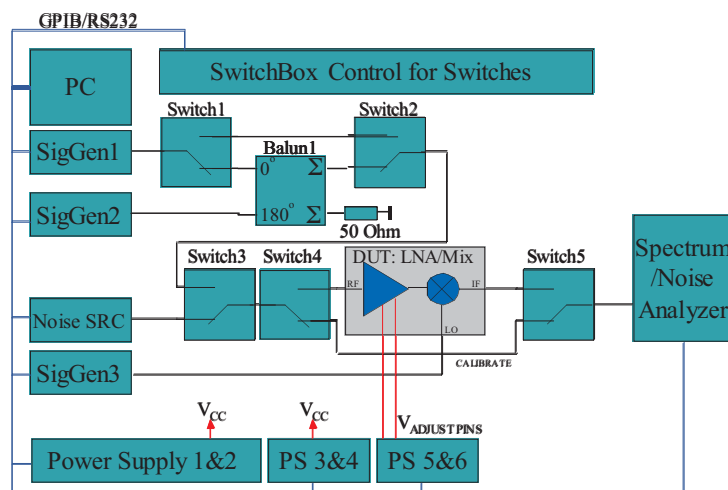


Figure 9 Schematic diagram of the automated characterization set-up

The measurement set-up as it is currently being used is shown in Figure 10.



Figure 10 Automated measurement set-up for characterization of configurable RF building blocks

3.5. Validation of the block specifications

After the definition of the building blocks, it is necessary to validate that indeed all standards, as represented in the scope specification cloud, can be covered. This has been accomplished by using the behavioral models and a system simulator. The model parameters are derived from the specifications of the building blocks, and the simulator is used to check the system specifications for these blocks connected according to the common, modular architecture (Figure 11).

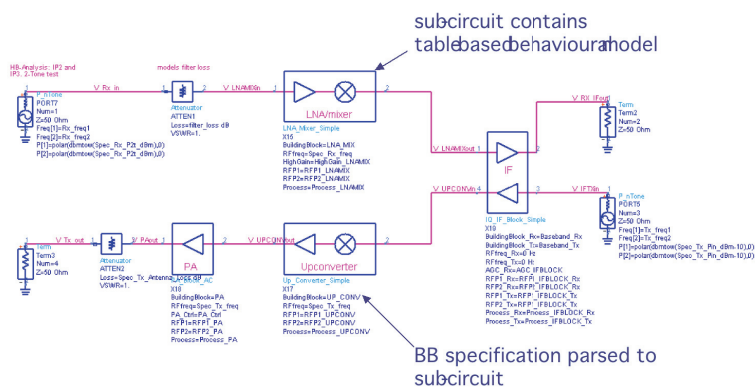


Figure 11 Validation test bench

The flow used to validate the coverage of the platform scope is shown in Figure 12. For each standard, a number of tests are run through the system simulator. If the specifications for the point in the platform scope specification cloud are not met, the configuration parameters of one or more building blocks are adjusted and the simulation is run again. If in any test any configuration parameter of any building block is changed, all previous tests are invalidated and the validation for this specific point in the platform scope is restarted. Only when all tests are passed with a single combination of configuration parameters, the specific point in the platform scope is validated. The building block specifications are only validated when all points in the platform scope are validated.

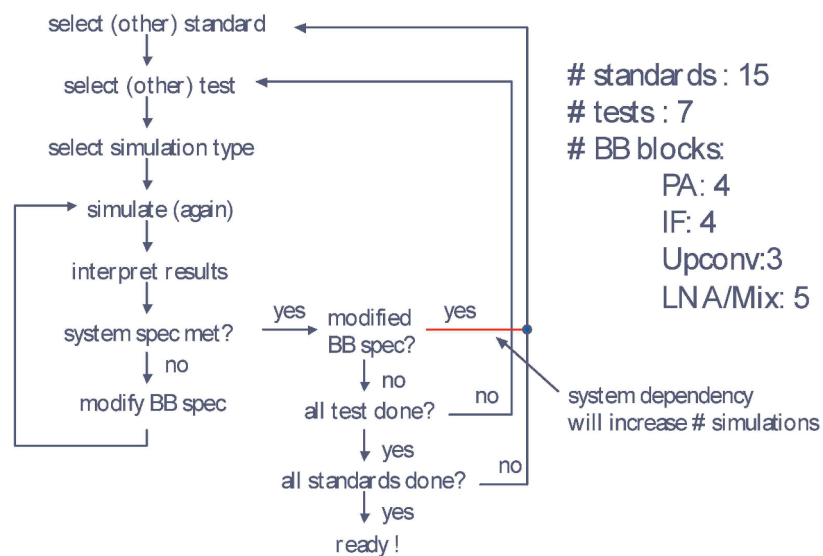


Figure 12 Validation flow

3.6. Selection of blocks and configuration settings

When using a platform-based system design approach, one of the first design choices is the selection of the most appropriate building blocks and the settings of the configuration parameters for these blocks. Even for a single-mode single-band transceiver, 5 blocks need to be selected from the library of 19 blocks, giving 720 possible combinations. Each of these building blocks has 2 configuration parameters, and at least 10 values for each parameter would have to be considered, giving 10^{10} possible parameter combinations for each combination of building blocks. Exhaustively trying all possible combinations of building blocks and parameter settings is therefore not yet practical.

Genetic algorithms were found to be a suitable approach to solving this complexity problem [7], and more specifically, a differential evolution algorithm [8] was used. This algorithm finds the solution of the problem described above in about 20 seconds on a modern PC, well within acceptable time limits.

4. Conclusions

A platform-based design approach is feasible for RF transceiver design. It covers an application area, as defined by the platform scope specification cloud, with a small number of reusable, configurable building blocks, and is capable of much higher reuse counts than is typical for non-platform-based reuse approaches. Moreover, it provides an evolutionary path towards software-defined radio in the future.

Acknowledgements

The work described in this paper was carried out by the RF platform project team at the Philips semiconductors Advanced Systems Labs, especially by Stefan Crijns, Oswald Moonen, Arno Neelen, Marc Vlemmings, and Martin Barnasconi. The block and configuration setting algorithm was developed by Haris Duric as part of his masters thesis. Investigations into the limits of configurable and reusable RF circuits are carried out by Maja Vidojkovic as part of her PhD research.

References

- [1] Maja Vidojkovic, Johan van der Tang, Peter Baltus and Arthur van Roermund, "RF Platform: A Case Study on Mixers", ProRisc 2003, Eindhoven, The Netherlands
- [2] Maja Vidojkovic, Johan van der Tang, Peter Baltus* and Arthur van Roermund, "A Gilbert cell mixer with a digitally controlled performance space", ProRisc 2004, Eindhoven, The Netherlands
- [3] Peter Baltus, "Minimum Power Design of RF Front Ends", thesis, Eindhoven University of Technology, The Netherlands, 2004
- [4] Hyunchul Ku; Kenney, J.S, "Behavioral modeling of nonlinear RF power amplifiers considering memory effects", Microwave Theory and Techniques, IEEE Transactions on , Volume: 51 , Issue: 12 , Dec. 2003, Pages:2495 – 2504

- [5] Gharaibeh, K.M.; Steer, M.B, “Characterization of cross modulation in multichannel amplifiers using a statistically based behavioral modeling technique”, Microwave Theory and Techniques, IEEE Transactions on, Volume: 51 , Issue: 12 , Dec. 2003, Pages:2434 – 2444
- [6] Wood, J.; Root, D.E.; Tuffillaro, N.B., “A behavioral modeling approach to nonlinear model-order reduction for RF/microwave ICs and systems”, Microwave Theory and Techniques, IEEE Transactions on , Volume: 52, Issue: 9, Sept. 2004, Pages:2274 – 2284
- [7] Haris Duric, “Method for Optimisation of Mobile and Wireless Transceiver Design”, Masters Thesis, Eindhoven University of Technology, 2002
- [8] Storn, R. and K. Price, “Differential evolution – A simple and efficient adaptive scheme for global optimization over continuous spaces”, International Computer Science Institute, 1995, TR-95-012

PRACTICAL TEST AND BIST SOLUTIONS FOR HIGH PERFORMANCE DATA CONVERTERS

Degang Chen

Dept. of Electrical & Computer Engineering
Iowa State University, Ames, IA 50011, USA

Abstract

Data converters (ADCs and DACs) are the most widely used mixed-signal integrated circuits. The relentless push towards higher performance and lower costs has made data converter testing increasingly challenging, especially for those deeply embedded in a large system on chip (SoC) or system in package (SiP). With no solutions visible on the horizon, the ITRS has named BIST of analog and mixed-signal circuits as part of some identified “most daunting challenges” facing the semiconductor industry. This paper introduces some practical solutions for high performance ADC and DAC testing that can be used in both production test and built-in-self-test environments. In particular, four data converter testing strategies will be reviewed: 1) a stimulus error identification and removal algorithm enabling all transition points testing of 16 bit ADCs using 7 bit linear signal sources; 2) a cyclically switched DDEM DAC implemented as a on-chip stimulus source for ADC code density test, achieving better than 16 bit linearity; 3) a high-speed high-resolution DAC testing strategy using very low resolution digitizers; and 4) a BIST strategy using low resolution DDEM ADC for high performance DAC testing.

1. Introduction

Data converters (ADCs and DACs) are identified as the most prominent and widely used mixed-signal circuit in today’s integrated mixed-signal circuit design. With the increasing resolution and/or increasing conversion speed, testing of data converters is becoming increasingly more challenging [1, 2]. As more analog and RF circuitry is integrated in high-volume systems-on-a-chip (SoCs) applications [1] most of today’s data converters are deeply embedded in large system design, which adds significant extra difficulty in testing. Built-in-self-test is the most promising solution to test those deeply-embedded data

converters [2-4]. More significantly self calibration based on BIST can lead to post-fabrication performance improvement.

The ubiquitous belief in the IC test community is that in order to accurately test a data converter, instruments for generating stimulus input and/or taking output measurements must have accuracy that is at least 10 times or 3 bits better than the device under test (DUT). This widely accepted rule has been written into several IEEE standards. Guided by these and other literature, the data converter design and test community has invested numerous efforts attempting to integrate high accuracy testing circuitry on-chip for the purpose of achieving ADC/DAC BIST [5-8]. However, in BIST environments, sufficiently high accuracy and high speed instrumentation circuitry is extremely difficult to obtain due to the requirement of implementing it on a small sacrificial die area. The reported on-chip stimulus sources or measurement devices either cost too much design effort and die area, or lack the capability to test data converters with modestly high resolution. For example, the best on-chip linear ramp generator [5] is only 11-bit linear, which is sufficient for testing only up to 9-bit ADCs if the test error is expected to be at the $\frac{1}{4}$ LSB level. To overcome this bottleneck, test methods that can use low-cost stimulus sources and low-cost measurement devices but still achieve accurate testing results for high performance data converters must be developed before practical BIST of data converter can become reality. [9-10].

In this paper, we review four recently developed methodologies that offer great potential for becoming practical solutions to high performance data converter test and BIST. The first method is an ADC test algorithm called stimulus error identification and removal (SEIR) [11, 12]. The algorithm use two easy-to-generate nonlinear signals offset by a constant voltage as stimulus input to the ADC under test and use the redundant information in a system identification algorithm to identify the signal generator. Once the signal generator is identified, its error component can be compensated for in the ADC's input output data. This in turn allows for the accurate testing of the ADC. Furthermore, a simple test setup strategy can make the algorithm insensitive to test environment nonstationarities. Simulation and experimental results show that the proposed methodology can accurately test 16-bit ADCs using 7-bit linear signals in an environment with more than 100 ppm nonstationarity in the test window.

While the first method can provide accurate test of all transition points of the ADC, it does require floating point DSP capability that may not be readily available in stand alone ADC products. The second method provides a very-low-cost on-chip stimulus source for use in the traditional code density test of ADCs. The stimulus source is implemented as a minimum sized all-digital DAC, controlled by a very simple cyclic switching logic using the Deterministic Dynamic Element Matching (DDEM) technique [13, 14]. The DAC has 12 bit apparent resolution and the fabricated DACs have 9 to 10 bit INL linearity

performance. Experimental results show that the DDEM DAC stimulus source can be used to test 14-bit ADCs with test error bounded by ± 0.3 LSB. It outperformed any previously reported on-chip stimulus source by at least 4 bits in terms of ADC BIST performance. The robust performance, low cost and fast design make the DDEM DAC a qualified on-chip stimulus source for high performance ADC BIST.

The third method is concerned with at-speed full-code transition level measurement for high-speed high-resolution DACs [15]. Since DAC speed/resolution performance is typically higher than corresponding ADC performance, at-speed testing of high performance DACs is very challenging. The proposed method uses very low resolution ADCs as the high speed digitizer. Appropriate dithering is incorporated in the test algorithm to effectively increase the ADC quantization resolution and to prevent information loss due to large quantization errors. Simulation results show that the static linearity of 14 bit DACs can be tested to better than 1 LSB accuracy, and dynamic performance of more than 85 dB SFDR can be tested with 1 dB accuracy, using 6-bit ADCs. Experimental results demonstrate accurate testing of 12 bit DACs using 6-bit and 7-bit ADCs.

Like the second method, the fourth method is intended for a code-density based BIST solution well suited for on-chip DAC linearity testing instead of ADC testing [16]. Also similar to the second method, the fourth method employs the DDEM technique but uses it with a simple low-resolution flash ADC which will be used as the measurement device for digitizing the DAC output. A simple second step fine quantization stage and a simple input dithering DAC are also incorporated. Numerical simulation shows that the proposed flash DDEM ADC, which has a 6-bit coarse DDEM stage, an 8-bit fine stage and a 5-bit dithering DAC, with linearity of all the blocks less than 6 bits, is capable of testing 14-bit DACs accurately.

2. Precision ADC Test with SEIR

This section will briefly review the SEIR algorithm. More details can be found in [11, 12]. An n -bit ADC has $N=2^n$ distinct output codes. The static input-output characteristic of the ADC can be modeled as

$$D(x) = \begin{cases} 0, & x \leq T_0, \\ k, & T_{k-1} < x \leq T_k, \\ N-1, & T_{N-2} < x, \end{cases} \quad (1)$$

where D is the output code, x is the input voltage, and T_k is the k -th transition voltage. The ADC integral non-linearity (INL) at code k , INL_k , is defined as

$$INL_k = \frac{T_k - T_0}{T_{N-2} - T_0} (N - 2) - k. \quad (2)$$

The overall INL is the maximum magnitude of INL_k 's,

$$INL = \max_k \{|INL_k|\}. \quad (3)$$

A real ramp signal can be modeled as

$$x(t) = x_{os} + \eta t + F(t), \quad (4)$$

where x_{os} is a DC offset, ηt is a linear component, and $F(t)$ is a nonlinear component. Without affecting the linearity test results, (4) can be normalized and written as

$$x(t) = t + F(t). \quad (5)$$

Transition time t_k is defined to be the time instance at which the ramp signal is equal to the k^{th} transition level,

$$T_k = x(t_k). \quad (6)$$

If $F(t)$ is known and t_k 's are measured, T_k , INL_k and ADC linearity can be calculated by using equations above. However, the input nonlinearity $F(t)$ is usually unknown. To identify this nonlinearity, it is expanded over a set of M basis function $F_j(t)$'s with unknown coefficient a_j 's as

$$F(t) = \sum_{j=1}^M a_j F_j(t). \quad (7)$$

The SEIR algorithm uses two stimulus ramp signals with a constant offset α :

$$x_1(t) = t + F(t) \quad (8)$$

$$x_2(t) = x_1(t) - \alpha \quad (9)$$

Correspondingly, two sets of histogram data $H_{k,1}$'s and $H_{k,2}$'s and two estimates for a transition level T_k , can be obtained:

$$\hat{T}_{k,1} = \hat{t}_{k,1} + \sum_{j=1}^M a_j F_j(\hat{t}_{k,1}) = T_k + e_{k,1} \quad (10)$$

$$\hat{T}_{k,2} = \hat{t}_{k,2} + \sum_{j=1}^M a_j F_j(\hat{t}_{k,2}) - \alpha = T_k + e_{k,2}, \quad (11)$$

where $e_{k,1}$ and $e_{k,2}$ are estimation errors and the transition times are estimated from the histogram data $H_{k,1}$'s and $H_{k,2}$'s as

$$\hat{t}_{k,1} = \sum_{i=0}^k H_{i,1} / \sum_{i=0}^{N-1} H_{i,1} \quad (12)$$

$$\hat{t}_{k,2} = \sum_{i=0}^k H_{i,2} / \sum_{i=0}^{N-1} H_{i,2}. \quad (13)$$

By subtracting (10) from (11), we can eliminate the involvement of the unknown transition level T_k , and obtain an equation involving only the M a_j 's and α . As k takes different values, we obtain $N-1$ such equations. They can be used to robustly estimate the unknowns by using the standard least squares (LS) method to minimize the error energy:

$$\{\hat{a}_j's, \hat{\alpha}\} = \arg \min \left\{ \sum_{k=0}^{N-2} \left[\hat{t}_{k,1} - \hat{t}_{k,2} + \sum_{j=1}^M a_j (F_j(\hat{t}_{k,1}) - F_j(\hat{t}_{k,2})) + \alpha \right]^2 \right\}. \quad (14)$$

With the knowledge of ramp nonlinearity a_j 's, we can remove their effects on the histogram data and accurately identify the transition level as

$$\hat{T}_k = \hat{t}_{k,1} + \sum_{j=1}^M \hat{a}_j F_j(\hat{t}_{k,1}). \quad (15)$$

Thus ADC's linearity performance can be estimated by applying (2) and (3). Figure 1 gives an example where a 14-bit ADC is under test by using 16 parameters for input nonlinearity identification.

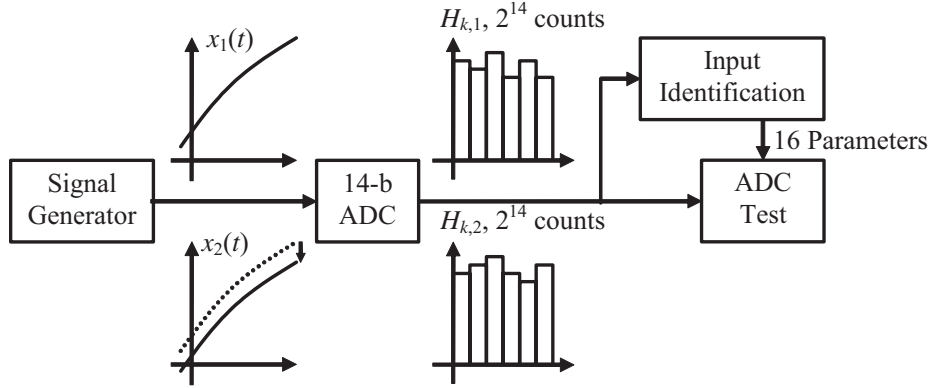


Figure 1 Test of a 14-bit ADC using SEIR algorithm.

Extensive simulation has been conducted in Matlab using behavioral level models of the ADC and the signal generator. Different ADC architectures (including flash as well as pipeline) have been simulated with various resolution levels (from 10 bits to 16 bits). Signal generators with different levels of nonlinearity ranging from 2 bit linear to 7 bit linear have been used. The constant shift levels have also been varied from 0.1% to 2% of the ADC input full range. Measurement noise in the test environment is assumed to be Gaussian and independent with zero mean and variance comparable to one LSB. Thermal nonstationarity of the test environment is also included. Figure 2 illustrate one representative example of testing a 14 bit ADC. The upper graph contains two curves, the actual INL_k plot and the estimated INL_k plot of the ADC. The difference between the two is shown in the lower graph, indicating testing accuracy to the 14 bit level. Figure 3 shows the testing results for 64 randomly generated 16-bit ADCs using 7 bit linear ramp signal and 32 samples per ADC code bin. The results show that test errors are consistently with 0.5 LSB.

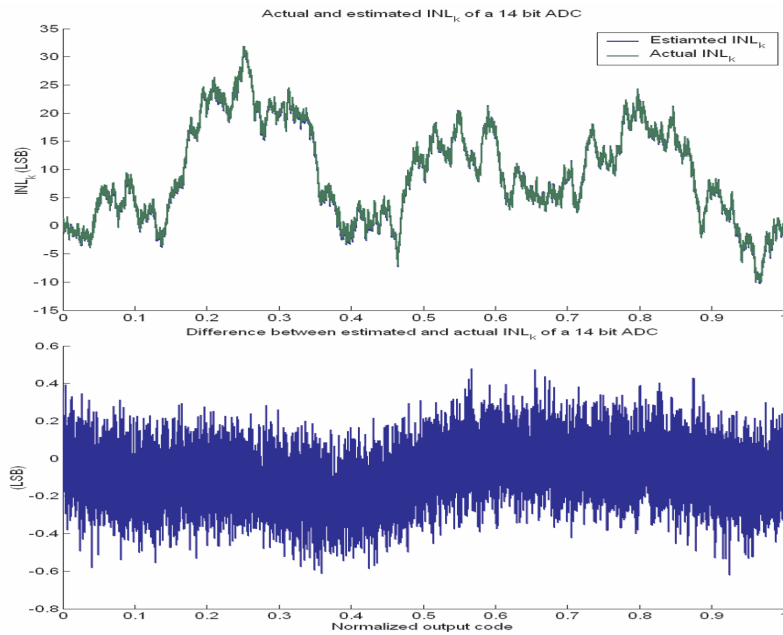


Figure 2 A representative example of a 14-b ADC

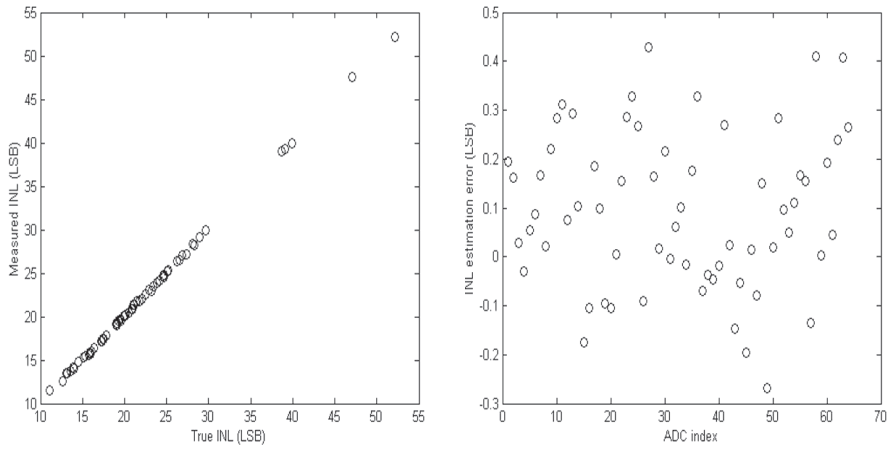


Figure 3 Linearity testing of 64 different 16-bit ADCs with 7-bit linear ramps, 32 samples per ADC code, maximum error within 0.5 LSB

Commercially available 16-bit ADCs were also tested to verify the performance of the SEIR algorithm. The sample used as the device under test was a laser trimmed 16-bit successive-approximation register (SAR) ADC with excellent linearity performance of about 1.5 LSB typical INL . Description of the test hardware will be omitted here. It suffices to say that experienced engineers at Texas Instruments tested the INL_k of the ADC using both the traditional

method and the SEIR method. 32 samples per code were used to keep the test time reasonable. The two nonlinear signals were synthetically generated with a nonlinear waveform having about 7-bit linearity. The DC offset between the two signals was set to about 0.05% of full range. Proper time-interleaving between the two signals was used to cancel up to the fifth order gradient errors due to environmental nonstationarity. Test results are shown in Figure 4. Results from the traditional method using 20 bit linear source were plotted as the top curve. The corresponding measured INL is 1.66 LSB. Results from the SEIR algorithm using 7-bit linear signals with 10 basis functions were plotted as the lower curve. The estimated INL with nonlinear signals is 1.77 LSB. The difference in INL estimation is only 0.11 LSB.

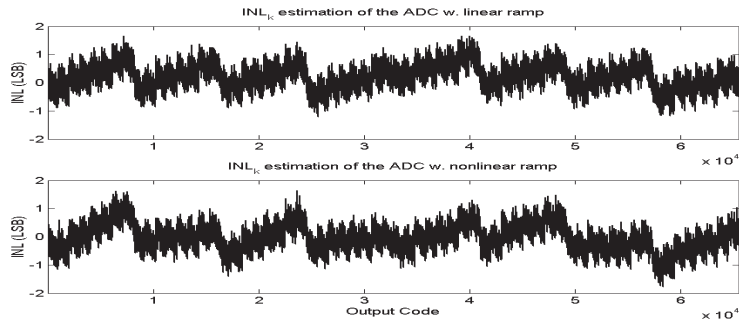


Figure 4 INL_k measurement of a 16 bit high performance SAR ADC. Top: test results with 20-bit linear signal. Bottom: test results with 7-bit linear signals

3. DDEM DAC for ADC Testing

In this section we will briefly review the Deterministic Dynamic Element Matching approach as applied to a current steering thermometer-coded (TC) DAC to produce a signal source for ADC BIST [13, 14]. Normally an n -bit current steering TC DAC has $2^n - 1$ current elements. For any input code k in the range of $[0, 2^n - 1]$, the first k current elements out of the total $2^n - 1$ are selected and turned on. The output current from these k elements can be forced to flow through a resistor R_C , and the voltage across R_C can serve as the output voltage for code k . For a normal DAC, the element selection for each DAC code k is fixed. However, in the DDEM DAC case, for each input code k , the current element selection has different combinations. The pattern of the switching sequence for all the DAC code k will determine the DAC output properties.

In an n -bit DDEM DAC, there are $N = 2^n$ current elements (i_1, i_2, \dots, i_N) with one extra element added to a normal DAC. For each input code k , the DDEM DAC produces p samples, each with a different current element combination. Here p is super-even and termed the DDEM iteration number. The DDEM DAC switching pattern is called the Cyclic DDEM Switching Sequence.

To show the switching sequence, the current sources are arranged conceptually and sequentially around a circle, as seen in Figure 5, to visualize a wrapping effect whereby the N^{th} current source is adjacent to the first current source. The physical layout of the current sources need not have any geometric association with this cyclic visualization. p index current elements are selected from all the N elements denoted by the sequence $i_1, i_{1+q}, i_{1+2q}, \dots, i_{1+(p-1)q}$, where q is defined as $q=N/p$. These index current sources are uniformly spaced around the circle. For each input code k , $1 \leq k \leq N$, the DAC generates p output voltages. Each output voltage is obtained by switching on k current sources consecutively starting with one of the p index current sources. Thus, the d^{th} sample ($1 \leq d \leq p$) is obtained by switching on k current sources consecutively starting with $i_{1+(d-1)q}$ and continuing around the circle in the clock-wise direction. For example, if $n=5$, $N=32$, $p=8$ ($q=N/p=4$) and $k=10$, the 1st voltage sample would be generated by switching on I_1 to I_{10} ; the 2nd voltage sample would be generated by switching on I_5 to I_{14} ; and so on.

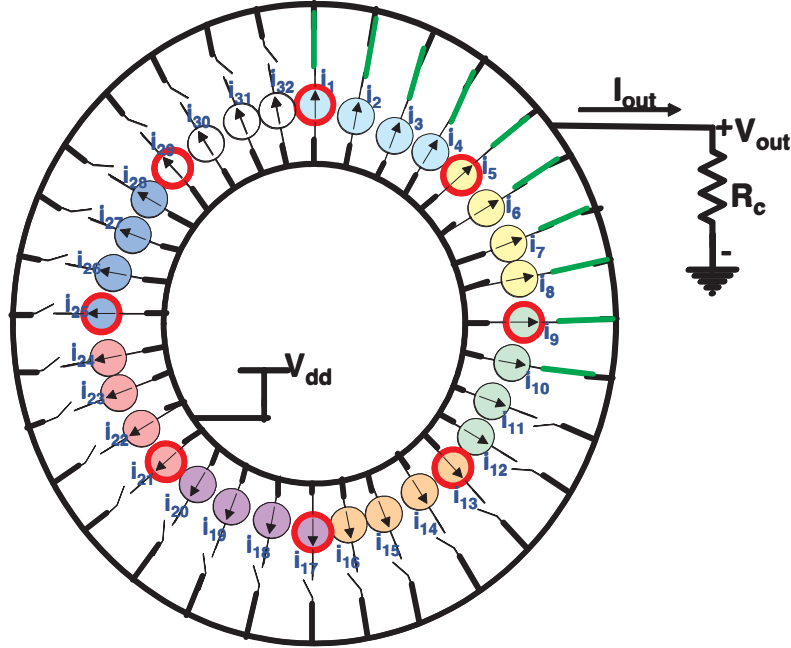


Figure 5 Cyclic DDEM switching: $n=5$, $N=32$, $p=8$, $q=4$, $k=10$; 1st output sample

It was shown in [14] that the DDEM DAC can achieve an equivalent linearity n_{eq} given by,

$$n_{eq} = \log_2 p + ENOB_{DAC} \quad (16)$$

where $ENOB_{DAC}$ is the effective number of bits of the un-switched DAC in terms of INL performance. Hence, each time we double the iteration number p ,

we can increase the DDEM DAC's linearity by one bit. This is shown to be true for p up to $2^{(ENOB_{DAC})}$. For example, if the un-switched DAC has 12 bit apparent resolution with about 9-bit linearity, then we can expect p to increase the DDEM DAC's linearity with p up to about 512.

To verify our theoretical results, a new DDEM DAC with 12 bit un-switched resolution and maximum $p=512$ was designed, fabricated, and tested, in 0.5 μ m CMOS technology. The DDEM DAC is composed of current source elements and the DDEM control logic. The 12-bit current steering DAC has 4096 current source elements, which occupies most part of the chip area. As DDEM can handle the random mismatching errors, minimum sized devices can be used to save the die area. A resulted benefit from using minimum sized devices is that parasitic capacitance is also minimized; hence the DDEM DAC operation speed can be very high. The adopted current source structure is the single-supply positive-output 3 PMOS transistor current source as depicted in Figure 6. The three PMOS transistors T_1 - T_3 form the basic current source structure. T_1 's gate is connected to V_r which provides a biasing voltage, and therefore the current flowing through T_1 is controlled by V_r . The 4 reference switching transistors T_4 - T_7 are also shown in Figure 6. When the control bit is high, T_2 's gate is connected to V_h and T_3 's gate is connected to V_l . Since V_h is set to be higher than V_l , when the control bit is high, the current from the drain of T_1 will go through T_3 to the I_{op} node. Similarly, when the control bit is low, T_1 's output current will go through T_2 to the I_{on} node. By this way, the current source functions symmetrically when the control bit signal is set between high and low. Two resistors R_p and R_n are connected to I_{op} and I_{on} respectively. The current from I_{op}/I_{on} will be collected on R_p/R_n , and the voltage difference across R_p and R_n serves as the output voltage.

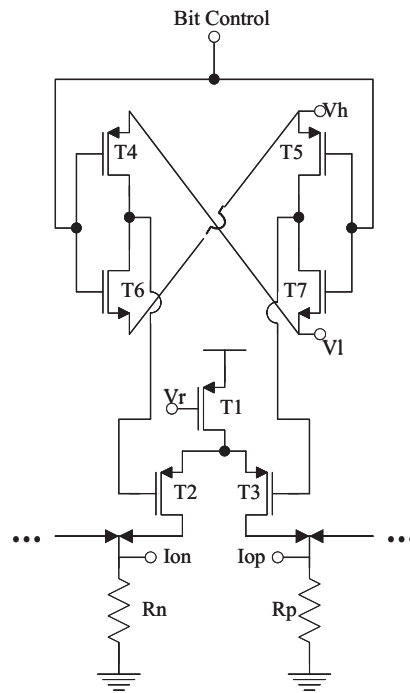


Figure 6 Current Steering Element Structure

Simulation shows that the settling error of the current source in Figure 4 is less than 0.02% within 5ns. Actually, the settling errors can also be handled by the DDEM algorithm. Hence the DDEM DAC can be operated at a very high speed up to hundreds of MHz, benefited from the simple structure and small device size. Furthermore, to maximize the speed, the two reference voltages V_h and V_l should be set properly. The difference between V_h and V_l should be chosen carefully such that when the control bit signal changes, the current from T1 can be almost totally switched to either T₂ or T₃ while during switching none of the 3 transistors T₁-T₃ will go into the deep triode region. The appropriate values for V_r , V_h and V_l can be found through transistor-level simulation.

Since a 12-bit DAC is estimated to have about 9 bit linearity, we limit the DDEM iteration number p to be 512. Thus, we group the current source elements into 512 groups. Each group contains 8 current elements that share the same DDEM control unit. The 8 elements inside each group will be switched on sequentially when the DAC's base clock signal advances, and the DDEM control logic clock frequency is 1/8 of the DAC clock frequency.

The Cyclic DDEM Switching Sequence and the DDEM control logic implementation and operation are very simple. The control logic circuit is a 512-bit shift register ring with each unit controlling one current source element group. The simple 6-transistor shift register unit with 2 CMOS inverter and 2

NMOS transistor switches in series as shown in Figure 7 was adopted to achieve high speed with small die area. Two-phase non-overlapping clock signals whose frequency is only 1/8 of the DAC clock frequency are generated on-chip to drive the control logic. Starting from the all-zero state, one of the register units is selected as the index point and a logical '1' is continuously pumped into this unit. Then each time the DDEM control clock signal advances, one more register unit is set to '1'. Thus the DAC output a monotonic ramp voltage by clustering/releasing current on R_p/R_n . More ramps are obtained by changing the index point position.

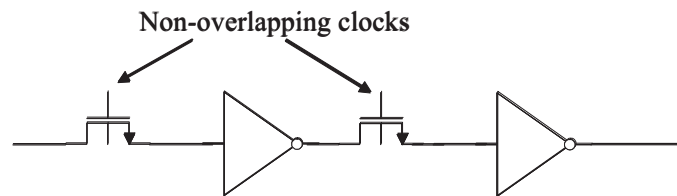


Figure 7 DDEM control logic unit

The 12-bit DDEM DAC was fabricated in $0.50\mu\text{m}$ standard CMOS process. The core die size is $1.5\text{mm}\times 1.4\text{mm}=2.1\text{mm}^2$. The die photo is shown in Figure 8. The power supply voltages are 5V for both digital and analog parts. When driving the 22 ohm resistance loads, the differential output range is -1.1~1.1 volts. The actual output range can be tuned by changing the resistance loads or the biasing voltage. The DDEM DAC was tested on a Credence Electra IMS (Integrated Measurement System) tester. The tester provides the power supply, biasing and reference voltages, RESET signal, two-phase non-overlapping clocks and 9-bit DDEM iteration control signals. The output voltage across R_p and R_n is sampled using an 18-bit digitizer. When the DDEM DAC was clocked at 100MHz, neat ramps can be observed from the oscilloscope. However, since the 18-bit digitizer can not operate at 100MHz, the DDEM DAC's output voltages were only measured using 1MHz clocks.

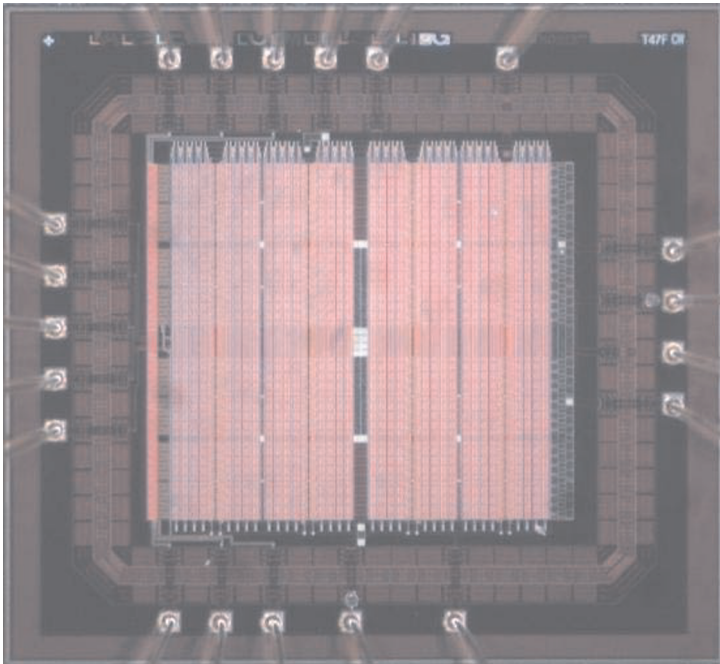


Figure 8 Photo of the DDEM DAC die

Figure 9 shows that without DDEM, the original 12-bit DAC has an INL error of 2.3 LSBs, which means the original DAC is about 9 to 10 bit linear. By Equation (16), the DDEM DAC is expected to have an equivalent linearity of 18 to 19 bits when $p=512$.

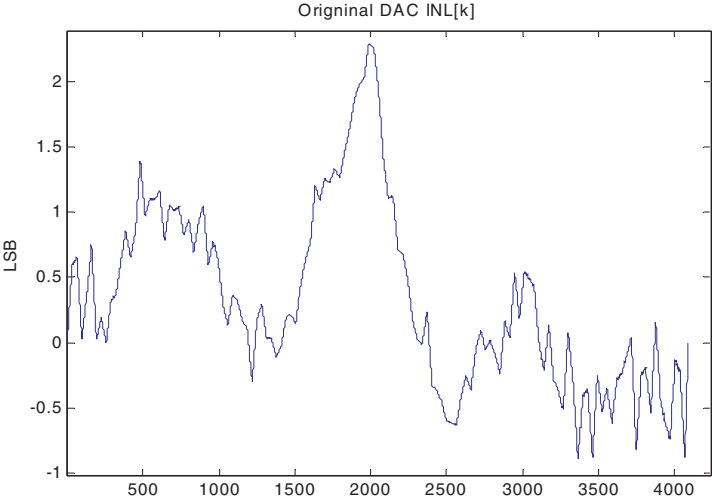


Figure 9 INL_k of the Original 12-bit DAC (Experimental)

To evaluate the DDEM DAC performance, the measured DDEM DAC output samples were used as the stimulus source to test simulated 14-bit ADCs. In Each test, the estimated INL_k curve using the DDEM DAC was compared to the ADC true INL_k curve, and the difference is recorded as the ADC test error. The result for a typical ADC test is shown in Figure 10. The INL_k estimation error is bounded by ± 0.3 ADC LSB, which means the DDEM DAC has achieved an equivalent linearity of at least 16 bits ($14 - \log_2 0.3 = 15.7$). The actual test performance is 2 to 3-bit lower than what is predicted by theoretical analysis. However, the measured performance may be limited by the digitizer used which has 18-bit resolution.

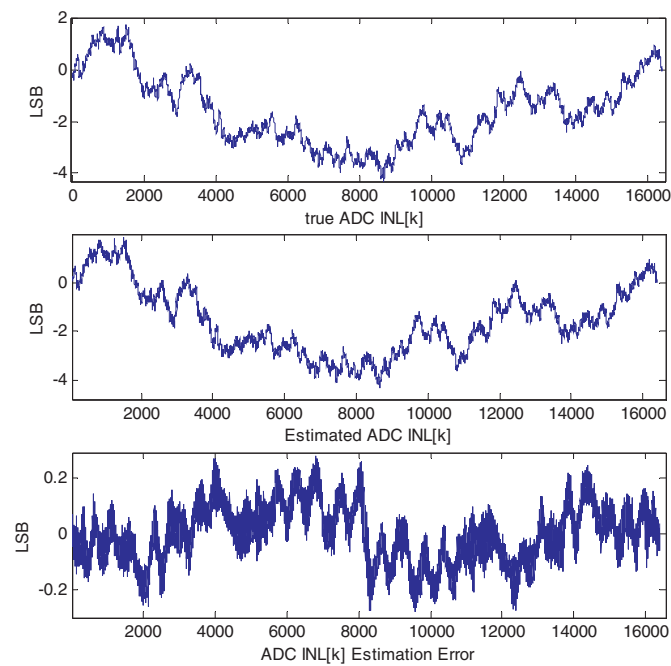


Figure 10 ADC INL_k test curves using 12-bit DDEM DAC

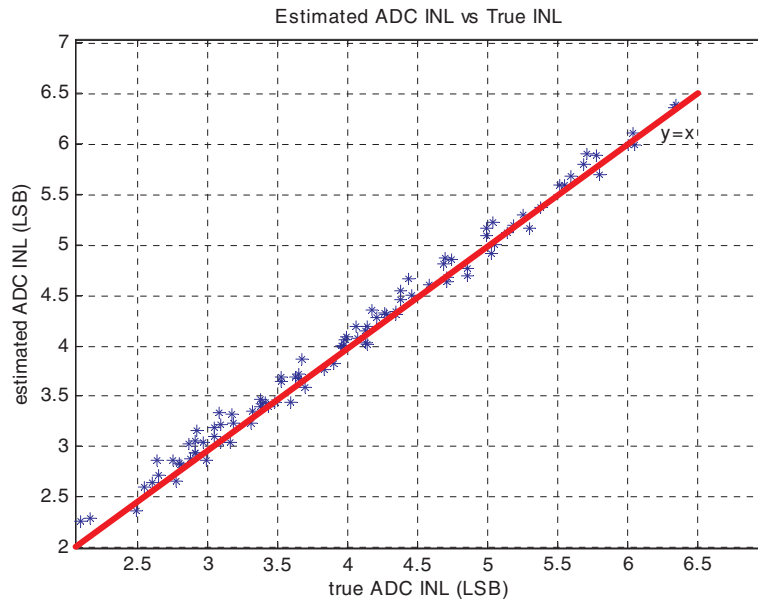


Figure 11 ADCs' True INL vs Estimated INL

To verify the robustness of the DDEM DAC performance, the measured DDEM DAC was used to test 100 simulated 14-bit ADCs with different amount of INL errors. The estimated ADC INL's using the DDEM DAC as test stimulus source versus true ADC INL's is shown in Figure 11. The ADC INL estimation errors (defined as the estimated INL minus the true INL) range from -0.16 to 0.24 ADC LSB. Table 1 provides a comparison of between this work and other on-chip stimulus sources for ADC test in literature. Only those having experimental results are listed in this table. As can be seen this new DDEM DAC outperforms any previously reported on-chip stimulus source by at least 4 bits. Since this DDEM DAC design is a "digital" DAC, it can be easily scaled down to newer technologies. Compared to other source generators, this new DDEM DAC is very die area efficient.

Table 1 Performance Comparison

Source Generator	Year	Source Type	Die Area & Technology	Performance
This work	2005	DDEM DAC	2.1mm ² @ 0.5μm CMOS	16 bits
1 st DDEM DAC [13]	2004	DDEM DAC	0.4mm ² @ 0.5μm CMOS	12 bits
B. Provost et. al. [5]	2003	Linear Ramp	0.18mm ² @ 0.18μm CMOS	11 bits
C. Jansson et. al. [6]	1994	Linear Ramp	N/A @ 2μm CMOS	8 bits

4. High performance DAC Testing with Low Resolution ADCs

A DAC's static linearity is characterized by its integral nonlinearity (*INL*) and differential nonlinearity (*DNL*). The fit-line *INL* of an n-bit DAC at code *k* is defined as

$$INL_k = (N-1) \frac{v_k - v_0}{v_{N-1} - v_0} - k \text{ (LSB)}, k = 0, 1 \dots N-1, \quad (17)$$

where $N = 2^n$ and v_k is the output voltage associated with k . The unit LSB, standing for the least significant bit, is the averaged voltage increment,

$$1 \text{ LSB} = \frac{v_{N-1} - v_0}{N-1}. \quad (18)$$

INL_0 and INL_{N-1} are equal to 0 under this definition, which is a straightforward result of the fit line definition. The expression of *INL* is

$$INL = \max_k \{ | INL_k | \}. \quad (19)$$

Definitions of code-wise and overall *DNL* are

$$DNL_k = (N-1) \frac{v_k - v_{k-1}}{v_{N-1} - v_0} - 1 \text{ (LSB)}, k = 1 \dots N-1, \quad (20)$$

$$\text{and } DNL = \max_k \{ | DNL_k | \}. \quad (21)$$

Dynamic performance of a DAC is usually characterized by its frequency response. One of the commonly used spectral specification is the spurious free dynamic range (SFDR), defined as the difference between the amplitude of the fundamental component and that of the maximum spurious component in the output spectrum of the DAC sinusoidal response,

$$SFDR = A_1 - \max \{ A_j, j = 2, 3, \dots \} \text{ (dB)}. \quad (22)$$

In the standard approaches to DAC testing, a high resolution digitizer is used to measure the output voltage of the DAC for each DAC input code. For static linearity testing, the DAC input code is sequentially increased from minimum to maximum at a slow rate to ensure accurate measurement of the output. For spectral testing, sinusoidal input is used, and a high speed measurement device is required. If the measurement needs to be done on-chip, the availability of such measurement devices becomes a great challenge. In particular, if one has a high resolution (14+ bits) DAC with update rate in the GHz range, there is no available method for at speed test with sufficient resolution.

The proposed strategy uses a low-resolution measurement ADC (m-ADC) and a dithering DAC (d-DAC) to test a high-performance DAC, the device under test (DUT), as shown in Figure 12. The m-ADC can take the form of a flash structure for high-speed sampling. Since its resolution is much lower than the DUT, information about the fine details of the DAC output variations will be lost due to coarse quantization. To prevent information loss, the d-DAC output is scaled by a small factor α and added to the output of the DUT before presented to the input of the m-ADC. The d-DAC can be simply a known-to-be-operational device from the same product family of the DUT.

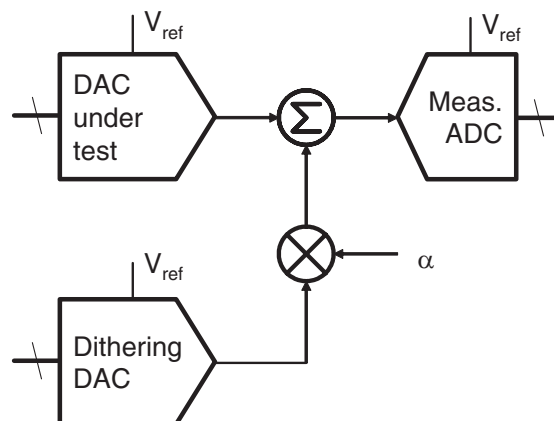


Figure 12 DAC test with a low-resolution ADC and dithering.

For quasi-static linearity testing, the DAC (DUT) is commanded to generate one output ramp for each dithering level. To ensure continuation of information, the full dithering range is selected to be 3 LSB of the m-ADC. Hence for each DAC input code, 3 or 4 different m-ADC output code may be obtained due to different dithering levels. For each DAC input code, the m-ADC output is sorted into a histogram of 3 or 4 bins.

After all the measurement data is collected, a joint identification algorithm can be used to identify the m-ADC's transition levels as well as the main DAC's

output levels. Instead of going through the algorithm itself, we point out that the whole setup can be viewed from the perspective of the ADC testing algorithm in section 2. The main DAC (DUT) can be viewed as the unknown nonlinear signal generator. The d-DAC is providing many different levels of constant offset. The difference now is that the m-ADC has fewer parameters to be identified (63 for 6 bit ADC). Hence, our algorithm first identifies the transition levels of the ADC to the 16-bit accuracy level and uses this information together with the dithering information to identify the output levels of the main DAC.

To test DAC spectral performance, the DUT will repeatedly generate a periodic waveform and the d-DAC will dither each complete period of the signal by a different voltage level. The low-resolution m-ADC will sample the dithered waveform. The m-ADC output code together with the dithering information will be used to estimate the DAC output voltage level for a given input code. With the estimated voltage samples, the SFDR of the DAC output waveform can be calculated using FFT. In addition to ramp dithering, sine waves can also be used for dithering in spectral testing. Both the DUT and the d-DAC are generating sinusoidal signals at different frequencies, while the m-ADC is digitizing the dithered waveform.

Large numbers of 14-bit DACs were tested in simulations using 6-bit m-ADC with 5-6 bit linearity and 12-bit dithering DAC with 9-10 bit linearity. Both static and dynamic testing situations are simulated. The proposed algorithm can consistently test both the static linearity and spectral performance accurately to the 14-bit level. Figure 13 shows a representative case of simulation results on testing a 14-bit DAC using 6-bit ADC as measurement device with 12 bit dithering. The INL_k testing errors indicate the 14 bit testing accuracy was achieved.

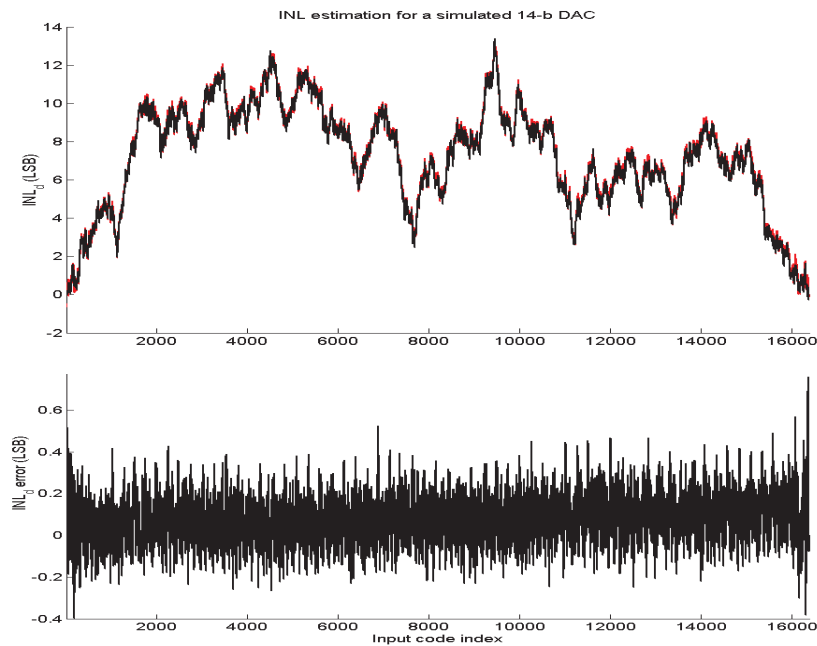


Figure 13 Representative simulation results on INL testing of a 14-bit DAC using 6-bit ADC as measurement device. Top: actual and estimated INLk plots; bottom: difference between the two

Figure 14 summarizes the simulation results for SFDR testing of 64 different 14-bit DACs using 6-bit measurement ADCs with dithering. The DACs' true SFDR as measured using an ideal infinite resolution ADC is on the horizontal axis. The SFDR testing errors for each DAC goes to the vertical axis. Results in Figure 14 indicate that all SFDR test errors are with 1.5 dB using 4098 point FFT. Notice that these 64 DACs have true SFDR ranging from less than 75 dB to more than 90 dB.

Preliminary experiments were carried out to validate the performance of the proposed DAC testing algorithm using low-resolution ADCs. We used a Conejo baseboard by Innovative Integration in our experiments. This board has four 16-bit DACs, four 14-bit ADCs, and a TI DSP on board. As a comparison reference, a sine wave signal with a synthesized distortion was first measured

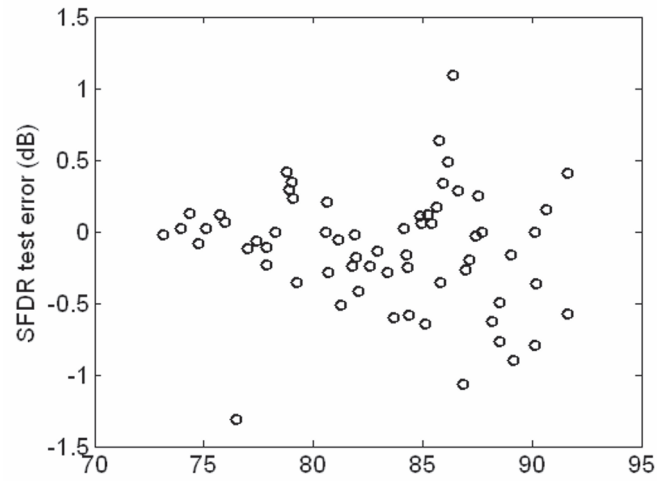


Figure 14 14-bit DAC SFDR test error using 6-bit m-ADC.

using a 14-bit ADC. The data length was chosen as 2048 in experiments, containing 11 periods. The signal's FFT spectrum is plotted in Figure 15. The measured SFDR was 59.91 dB.

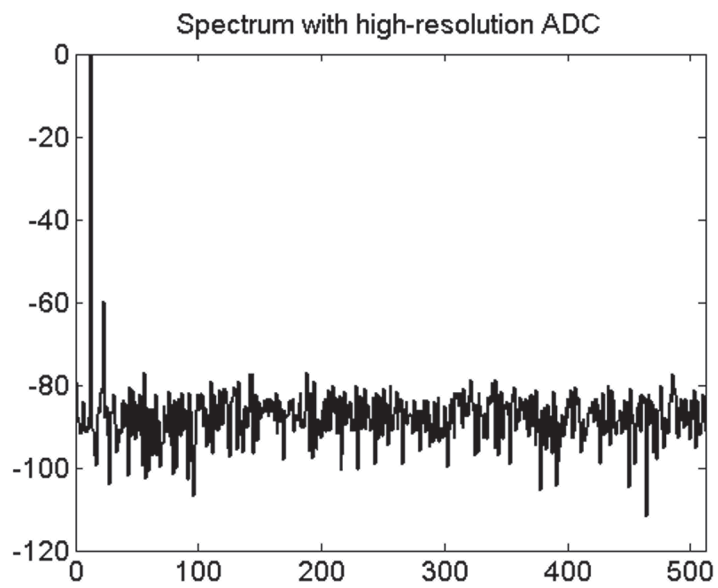


Figure 15 Estimated spectrum of a synthesized 12-bit DAC using a 14-bit ADC as the measurement device

The same DAC generating the same signal was then tested by using a 6-bit ADC with 8-bit dithering. The range of the dithering signal was 5% of the output of the DUT. The dithered output was quantized and processed by the proposed algorithm. FFT was used to generate the spectrum plotted in Figure 16. The estimated SFDR was accurately tested to be 59.23 dB. Since the proposed method uses more data points to avoid information loss, this also led to a reduced noise floor in the spectrum.

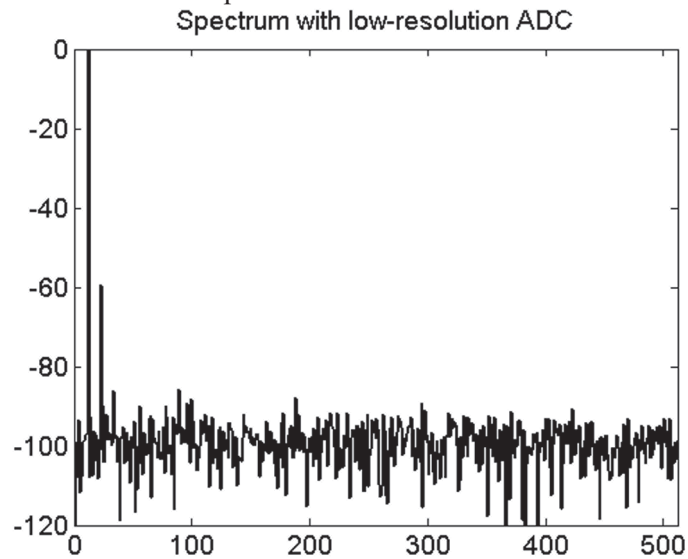


Figure 16 Estimated spectrum of the same 12-bit DAC but using 6-bit ADC as measurement device together with dithering

For comparison, the same DAC generating the same signal was also tested using the same 6-bit ADC as measurement device but without using the proposed algorithm. Figure 17 shows the testing results, which are clearly wrong since they are very different from those in Figure 15.

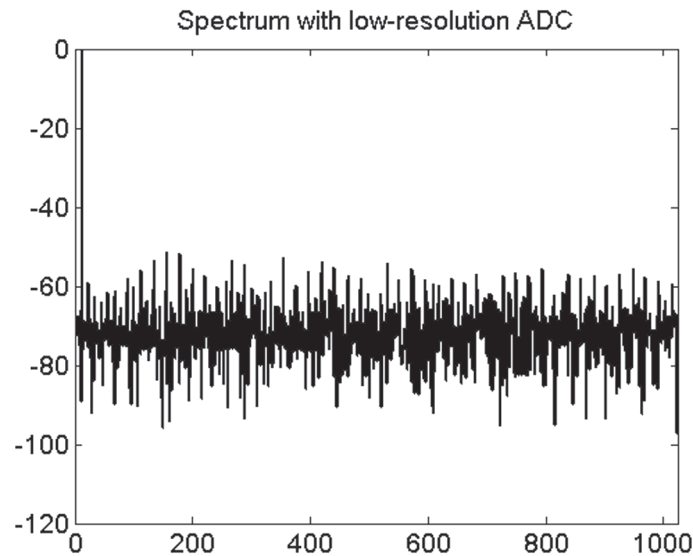


Figure 17 Estimated spectrum of the same 12-bit DAC but using 6-bit ADC as measurement device without the proposed algorithm

5. DDEM Flash ADC for DAC BIST

The method in last section can be used to accurately measurement each and every DAC output voltage levels. In many applications, the DAC output can be guaranteed by design to have a monotonic relationship with the DAC input code. In this section, we will review a DDEM ADC based method that is suitable for BIST of DAC nonlinearity. Unlike the method in section 3 which uses DDEM with a DAC for ADC testing, here we apply DDEM to the resistors in a R-string of a low resolution flash ADC for testing DAC nonlinearity. The DDEM switching rearranges resistors to form different R-strings, which leads to different sets of ADC transition points. Assume mismatch errors in resistors are generated from a normal distribution with a zero mean and a standard deviation $\sigma \cdot R_0$, where R_0 is the desired resistance value. The overall distribution of all the possible transition points is nearly uniform, which is a desired distribution of ADC transition points to be used in DAC testing.

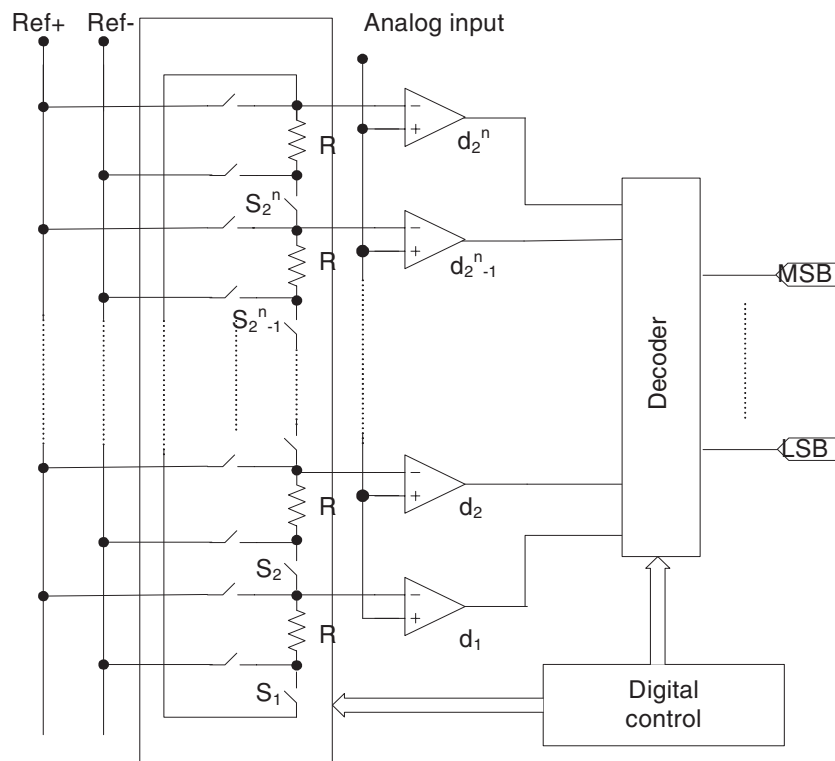


Figure 18 Structure of an n-bit DDEM flash ADC

Figure 18 shows the structure of an n-bit DDEM flash ADC. Similar to a typical flash ADC, an R-string with N resistors forms a voltage divider that provides reference voltages, where $N=2^n$. The decoder converts thermometer codes generated by the comparators into binary codes. Different from the conventional flash structure, resistors are physically connected as a loop via switches in the DDEM ADC. The loop can be broken at different positions by opening specific switches to build different R-strings, consequently different ADCs. Each time, one of P switches (uniformly spaced along the loop), S_i for $i=(j-1)*q+1, j=1, 2, \dots, P$, is open, where P is selected so that $q=N/P$ is an integer. Connecting the two nodes of the open switch to external reference voltages, a set of internal reference voltages is generated. Therefore, P digital outputs are available for one analog input, quantized by the DDEM ADC with different sets of reference voltages. In this DDEM structure the maximum value of P is N .

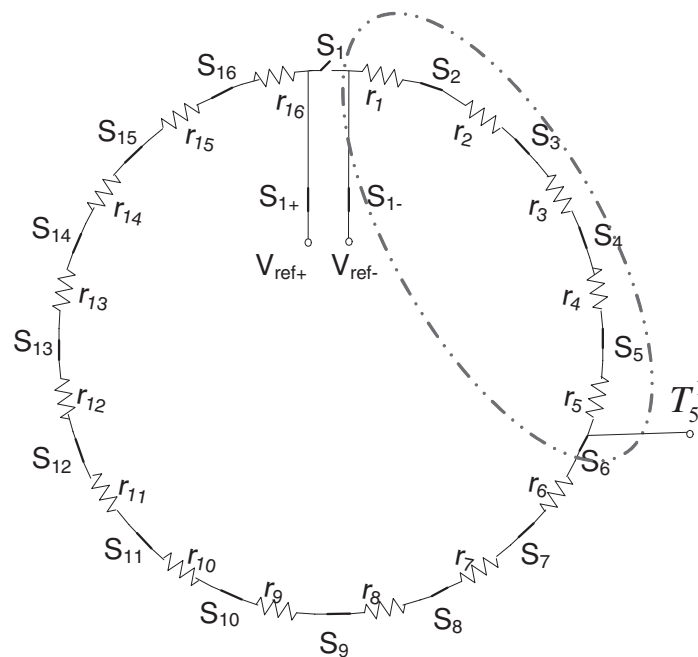


Figure 19 Switching of a 4-bit DDEM flash ADC with $P=4$

Figure 19 illustrates an R-string with 2^4 resistors in a 4-bit DDEM ADC. When $P=4$, one (S_1 shown) of the 4 switches (S_1, S_5, S_9, S_{13}) can be open at different times to obtain 4 different R-strings, which lead to 4 sets of transition points. For each output code of the DDEM flash ADC we have P transition points. It can be shown that the distribution of all the transition points ($P \cdot 2^n$) are nearly uniform in the ADC input range.

Flash ADCs provide the fastest conversion from an analog signal to a digital code and is ideal for applications requiring a large bandwidth. However, the resolution of flash ADCs is restricted to 8 bits by the large amount of power consumption, area, and input capacitance introduced by the 2^n comparators. To make the scheme suitable to high-resolution test, a fine flash ADC quantization and an input dithering DAC is incorporated with the DDEM stage. Figure 20 illustrates the structure of the proposed BIST approach and Figure 21 shows the block diagram of the two-step DDEM flash ADC.

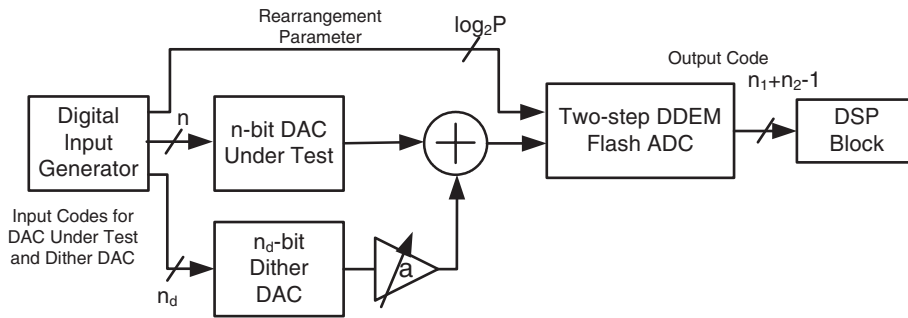


Figure 20 Block diagram of the proposed BIST scheme

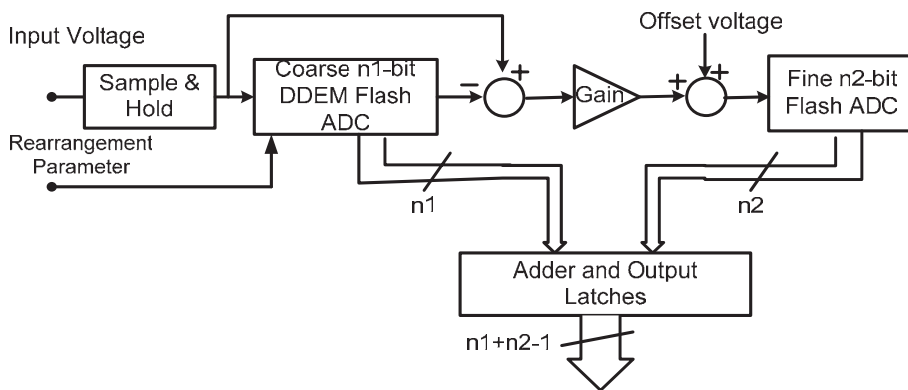


Figure 21 Block diagram of the two-step DDEM flash ADC

The two-step structure comprises a sample-and-hold stage, an n_1 -bit coarse DDEM flash ADC and an n_2 -bit fine flash ADC, a residual voltage generator, a gain stage, and a digital adder and output latches. The sample-and-hold stage is needed to compensate for the time delay in the coarse quantization and reconstruction steps. The coarse ADC does the conversion for the first n_1 bits. A residual voltage is generated by subtracting from the analog input the reference voltage right smaller than it, determined by the coarse ADC output, and the difference is amplified by the gain stage. In order to avoid missing codes, the full-scale range of the fine flash ADC is set to be equivalent to 2 LSBs of the coarse system. A constant offset voltage is added to the residual voltages to move them up to a desired input level for the fine ADC, where the middle part of the fine ADC's input range is. This shift operation can compensate for the errors in residual voltages introduced by comparator offset voltages. The final output code is a summation of shifted coarse and fine codes. In this DDEM structure, mismatches in the coarse resistor strings are desired to spread out distributions of transition points after DDEM. This low matching requirement dramatically reduces the area consumption of the R-string. Because the full scale range of the fine stage is only equivalent to 2 LSBs of the coarse

stage, the fine stage can greatly increase the test ability of the whole ADC, and accuracy and linearity of the fine stage are not critical to the test performance.

The output of the dithering DAC is added to the output of the DAC under test, and the sum is taken as the input to the DDEM ADC. The full scale output range of the dithering DAC is adjustable and very small relative to the ADC input range, e.g. several LSBs of the original first stage flash ADC. That ensures the shifted DAC output signal is still covered by the middle linear part of DDEM ADC transition points. For each output of the dithering DAC, output voltages of the DAC under test are shifted up by a small offset. It is equivalent to shifting all the transition points of the ADC to the opposite direction by an equal amount. Assume the resolution of the dithering DAC is n_d . The DDEM ADC's transition points are shifted 2^{n_d} times. The nonlinearity error in the dithering DAC introduced by component mismatches can be neglected because of its small output range.

To verify the proposed structure, simulations in MATLAB are carried out. In simulation, a 14-bit DAC is modeled as a device under test. The two-step DDEM flash ADC has a 6-bit coarse stage, and an 8-bit fine stage. The resistor strings in these two ADCs are generated from a Gaussian distribution with a nominal value of 1 and $\sigma = 0.05$ to match practical situations. The linearity of the original coarse stage and fine stage are nearly 6 bits and 7 bits respectively.

Figure 22 shows the INL_k estimation error when only the two-step DDEM ADC with $P=64$ is used to test the 14 bit DAC. The max INL_k estimation error is about 2LSB in 14-bit level, that means the tester has about 12-bit test performance. It agrees with our theoretical analysis.

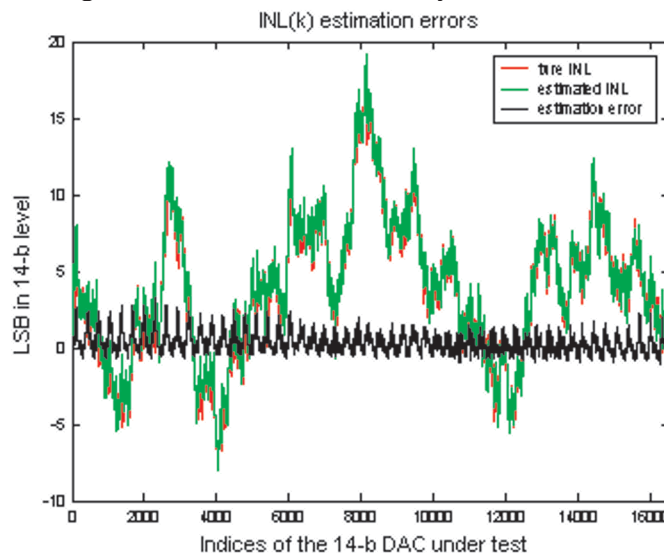


Figure 22 INL_k estimation error with $P=64$ and no dithering DAC, 14 bits DAC under test

By adjusting the value of P and adding a dithering DAC, we can reduce the estimation error. Figure 23 illustrates the estimation results when $P=16$ and a 5-bit dithering DAC is used. The result shows that with the above configuration the INL_k error is under 0.4LSB in 14-bit level and the INL error is 0.13LSB .

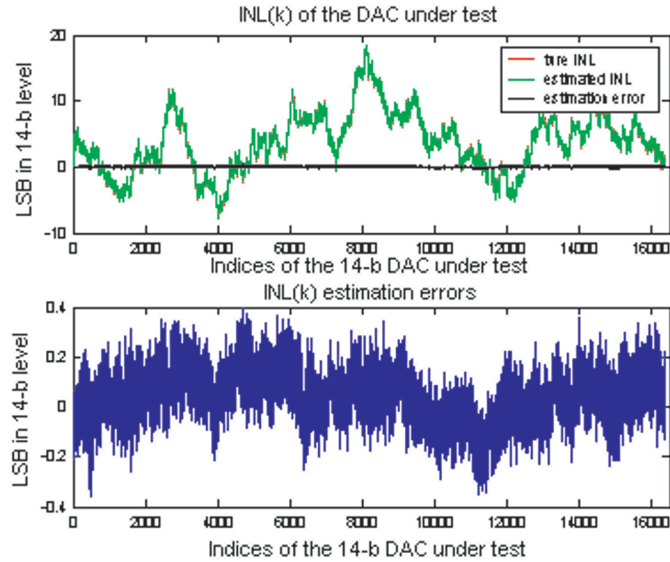


Figure 23 INL_k estimation error with $P=16$ and 5-bit dithering DAC, 14 bits DAC under test

In the analysis, we have shown that the test ability of DDEM depends on the distribution of the mismatch errors. In order to validate the robustness of the algorithm, different DDEM ADCs are implemented. In this simulation, we use 100 different DDEM ADCs, with the coarse stage nearly 6-bit linearity, to test 100 different 14-bit DACs. Figure 24 shows the relationship between the estimated INL values of different DACs and the true values, where the estimation errors are less than 0.586LSB and the INLs of the DACs are in the range from 5LSB to 25LSB . The results show that with P equal to 16 and a 5-bit dithering DAC, the proposed two-step DDEM ADC is capable of testing 14-bit DACs.

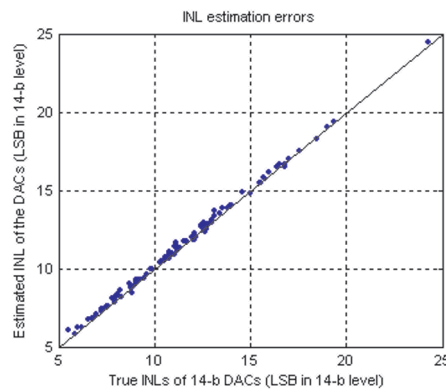


Figure 20 Estimated INL vs true INL for 100 14-bit DACs tested by 100 DDEM ADC

6. Conclusion

We have reviewed a family of practical methodologies that are suitable for testing or built-in-self-test of high performance data converters. Every method uses practical, easy-to-implement stimulus signal generators for ADC testing or low-resolution digitizers for DAC testing, but still achieves high accuracy testing results. Such methods offer great potential for being incorporated as on-chip test solutions. They can serve as enabling technology for general AMS test and BIST in deeply embedded SoC. Finally, these methods can be incorporated into on-chip self calibration for performance and yield enhancement.

References

- [1] "International Technology Roadmap for Semiconductors," 2003 edition, <http://public.itrs.net/Files/2003ITRS/Home2003.htm>
- [2] M. Burns and G.W.Roberts, "An Introduction to Mixed-Signal IC Test and Measurement," Oxford University Press, New York, USA 2000
- [3] R. de Vries, T. Zwemstra, E.M.J.G. Bruls, P.P.L. Regtien, "Built-in self-test methodology for A/D converters"; Proceedings 1997 European Design and Test Conference, 17-20 March 1997 pp. 353 -358
- [4] F. Azaiz, S. Bernard, Y. Bertrand and M. Renovell, "Towards an ADC BIST scheme using the histogram test technique"; Proceedings 2000 IEEE European Test Workshop, 23-26 May 2000, pp. 53 -58
- [5] B. Provost and E. Sanchez-Sinencio, "On-chip ramp generators for mixed-signal BIST and ADC self-test", IEEE Journal of Solid-State Circuits, Volume: 38 Issue: 2 , Feb. 2003, pp. 263 -273.
- [6] C. Jansson, K. Chen and C. Svensson, "Linear, polynomial and exponential ramp generators with automatic slope adjustment", IEEE

- Transactions on Circuits and Systems I: Fundamental Theory and Applications, Vol. 41, No. 2, Feb. 1994, pp. 181-185.
- [7] J. Wang; E. Sanchez-Sinencio and F. Maloberti, "Very linear ramp-generators for high resolution ADC BIST and calibration", Proceedings of the 43rd IEEE MWSCAS, pp: 908 -911, Aug. 2000
 - [8] S. Bernard, F. Azais, Y. Bertrand and M. Renovell, "A high accuracy triangle-wave signal generator for on-chip ADC testing", The Seventh IEEE European Test Workshop Proceedings., pp89 -94 May 2002
 - [9] K. L. Parthasarathy, Le Jin, D. Chen and R. L. Geiger, "A Modified Histogram Approach for Accurate Self-Characterization of Analog-to-Digital Converters", Proceedings of 2002 IEEE ISCAS, Az, May 2002
 - [10] K. Parthasarathy, T. Kuyel, D. Price, Le Jin, D. Chen and R. L. Geiger, "BIST and Production Testing of ADCs Using Imprecise Stimulus", ACM Trans. on Design Automation of Electronic Systems, Oct. 2003
 - [11] L. Jin, K. Parthasarathy, T. Kuyel, D. Chen and R. Geiger, "Linearity Testing of Precision Analog-to-Digital Converters Using Stationary Nonlinear Inputs," Int. Test Conf., pp. 218-227, Charlotte, NC, Oct 2003.
 - [12] L. Jin, K. Parthasarathy, T. Kuyel, D. Chen and R. Geiger, "High-Performance ADC Linearity Test Using Low-Precision Signals in Non-Stationary Environments," submitted to 2005 International Test Conference, Austin, TX, Oct 2005.
 - [13] H. Jiang, B. Olleta, D. Chen and R. L. Geiger, "Testing High Resolution ADCs with Low Resolution/ Accuracy Deterministic Dynamic Element Matched DACs", 2004 Int. Test Conference, Charlotte, NC, Oct 2004
 - [14] H. Jiang, D. Chen, and R. L. Geiger, "A Low-Cost On-Chip Stimulus Source with 16-Bit Equivalent Linearity for ADC Built-In-Self-Test", submitted to 2005 International Test Conference, Austin, TX, Oct 2005.
 - [15] L. Jin, T. Kuyel, D. Chen and R. Geiger, "Testing of Precision DACs Using Low-Resolution ADCs with Dithering," submitted to 2005 International Test Conference, Austin, TX, Oct 2005.
 - [16] Hanqing Xing, Degang Chen, and Randall Geiger, "A New Accurate and Efficient Linearity BIST Approach for High Resolution DACs Using DDEM Flash ADCs," submitted to 2005 International Test Conference, Austin, TX, Oct 2005.

Simulation of Functional Mixed Signal Test

Damien Walsh, Aine Joyce, Dave Patrick
Analog Devices Limerick

damien.walsh@analog.com, aine.joyce@analog.com, dave.patrick@analog.com

Abstract

The late 1990's saw an increasing use of virtual test environments in the test development community, to allow quicker development of mixed signal test programs. Following the industry downturn in 2001, automatic test equipment (ATE) vendors have had to reassess their support for virtual test. This paper will detail an alternative approach developed to address the simulation of mixed signal test programs.

1. Introduction

Virtual test allows a test development engineer use their test program to drive a simulation of both the ATE and the device under test (DUT). Figure 1 shows the typical setup of the virtual test environment.

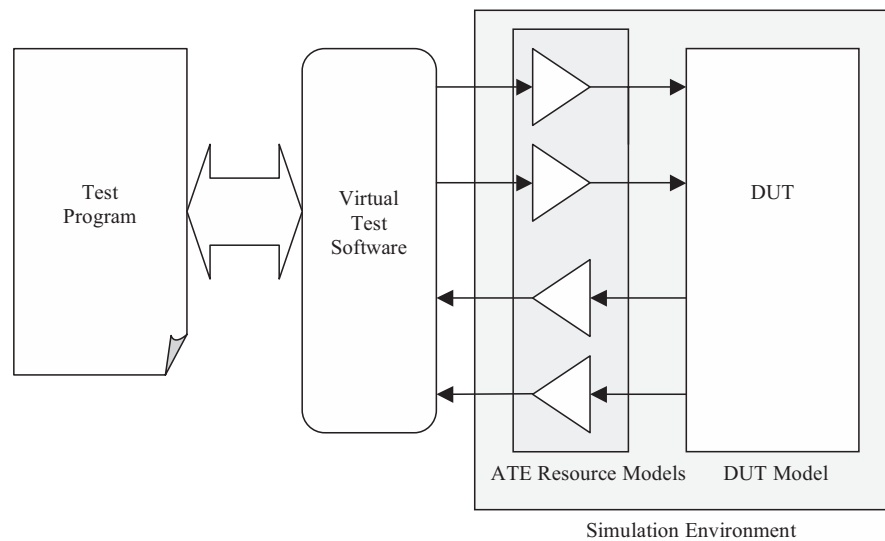


Fig.1. Typical Virtual Test Simulation Architecture.

The test program runs outside of the simulation environment and the virtual test software transfers stimulus and response information between the test program

and the ATE models running in simulation. The debug of the test program and DUT is accomplished using the standard ATE development tools. Typically the resources modeled would include digital pin drivers, voltage / current sources, arbitrary waveform generators and digitizers.

Virtual test was used on several products in Analog Devices Limerick and demonstrated that significant savings in terms of both ATE debug time and time to market could be made. However, following the announcement that support from some of our ATE vendors for virtual test would be phased out due to financial pressures resulting from the downturn in 2001, a team was established to investigate an alternative approach to simulation of mixed signal test. As part of this process the following weaknesses of virtual test were identified.

- 1) High Maintenance. Virtual test was seen as a high maintenance activity by the CAD group. This was partly due to complexity of the environment as well as the need to maintain capability with ATE system software revisions.
- 2) Portability. Simulations generated by the test development engineer could not be easily run by the rest of the development team, as an understanding of the ATE system software was required. This made it difficult to share simulation results especially if test development and design were not geographically co-located.
- 3) Requirements. As virtual test requires a test program, simulations could not typically be run until the test development engineer had designed hardware and commenced their program development. This typically delayed virtual test being run until just before design completion. Therefore virtual test was limited in helping define design-for-test requirements.

2. Tester Independent Simulation

As well as virtual test several other simulation tools were being used by test development engineers at the time. These tools provided a very basic digital and analogue capability when compared to virtual test. However it was found that as much benefit was being gained by the users of these tools when compared with virtual test users. Measures such as program development time, and pre-tape out bugs found, were all comparable. Therefore it was decided that any new simulation tool should try to merge the benefits of these tools in terms of ease of use, low maintenance, etc. with the benefits of virtual test e.g. rapid ATE program debug.

Ivy was chosen as the name for the simulation tool.

An outline of the simulation environment is shown in figure 2 below. In simulation, the ATE resource models have been replaced with Ivy models of typical ATE resources. Rather than using the test program to control the simulation, it is controlled via a simple text file. In taking this approach we have made the environment test system independent.

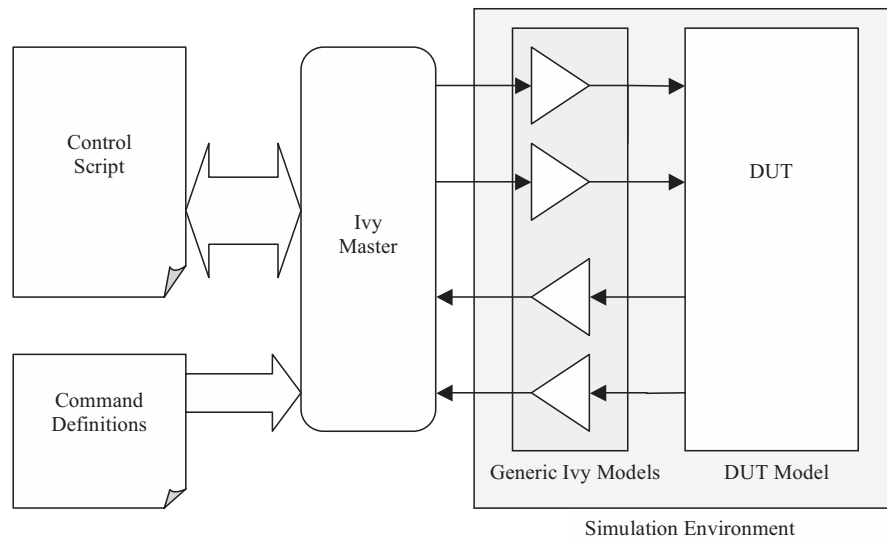


Fig2. Ivy Simulation Architecture.

This has major benefits in terms of CAD maintenance because we are no longer coupled to ATE system revisions. It also makes the tool available to the widest possible test development audience. Indeed many of the users are not test development engineers but design engineers, hence addressing the issue of simulation portability. Since no ATE knowledge is required to use the environment, simulations may be carried out by any of the development team. The major concern with adopting this approach was losing the ATE program debug capability that is a key benefit of virtual test. It will be discussed later how this has been addressed.

The Ivy simulation environment is controlled via an extendable language. The command definition file allows the user to change how the stimulus is applied to the DUT. This allows the same basic Ivy models to supply stimulus to, and capture response from, a wide variety of devices.

3. Transaction Level Simulation

Ivy uses the concept of a transaction level simulation. Here the stimulus and response to the mixed signal device is driven by high level commands, called from the control script. The command definition file provides Ivy with the necessary information to translate these high level commands and their arguments into the required stimulus signals for the DUT. Similarly the DUT response is captured and processed using information in the command definition file to provide high level response data back to the user. Figure 3 below illustrates the information flow during a typical simulation transaction.

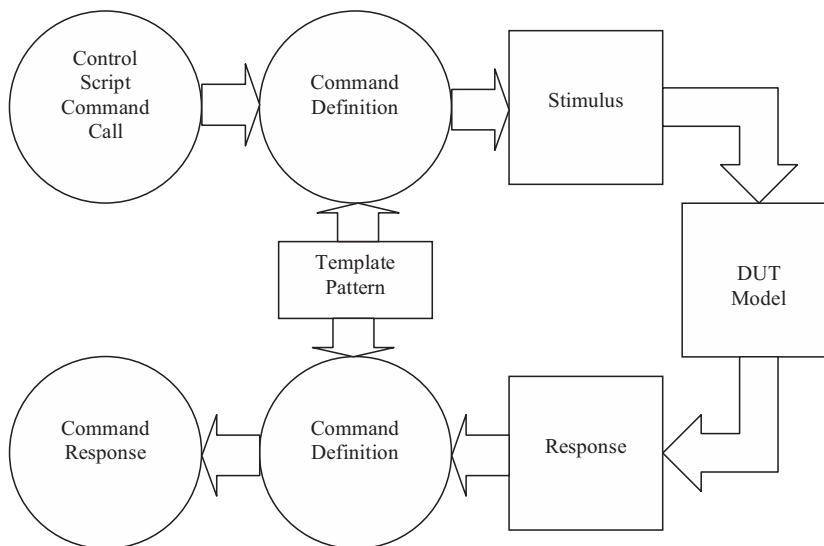


Fig3. Transaction Level Simulation Information Flow.

Commands would typically be defined that describe how information is written to and read from the device. The command definition describes how the data in the template pattern should be either modified or captured. Hence command arguments can be used to modify the stimulus pattern, and command return values can be generated from the DUT responses. For example, a read command would typically specify the register to be read, and the value being read would be generated from the DUT response. This resultant control script command and response for reading 0x55 from register location 0x7d from the DUT would be as follows:

```

Input:    read 0x7d
Response: read 0x7d -> 0x55
  
```

4. ATE Linkage

In order for the test development engineer to gain the most from simulation it is imperative that the work undertaken in simulation is easily ported to the ATE system. Ivy addresses this by defining a new vector file format that is supported by the majority of ATE systems used in Analog Devices Limerick. This file format, called a “.sim” file, is a cycle based format that includes timing and format information, dc levels information and vector information. It completely describes all of the relevant pattern information. Tools have been developed on our ATE systems that both directly read this format and save existing ATE patterns in the .sim file format. Tools have also been developed that will automatically generate ATE test program code from the command definition file so that the same transaction level environment is now available on the ATE as in simulation.

Ivy allows the test development engineer to capture an entire simulation run, composed of multiple transactions into a single “.sim” file that can be loaded directly onto the test system. By implementing the above ways of transferring simulation runs onto the ATE we can drastically reduce the amount of debug required on the ATE system. By allowing a route from ATE back into simulation we can also utilize our simulation tools to aid in the debugging of silicon issues discovered on the ATE system during product evaluations, thus further reducing ATE system time requirements for the test development engineer once actual silicon is available.

5. Ivy Capabilities

The simulation environments used in conjunction with Ivy allow a wide range of mixed signal simulations to be carried out. Ivy is typically used in full chip simulations using either Adice or an Adice / Verilog co-simulation.

Adice is a circuit simulator developed for use within Analog Devices in the design of analog, mixed signal, and RF integrated circuits. Adice provides a flexible, reliable environment for analog designers to explore the many degrees of freedom in analog circuit design. Additionally, Adice was specifically designed to perform simulations of large mixed signal chips using arbitrary mixtures of transistor-level and behavioral-level models of both the analog and digital circuitry. High level models may be constructed using the generic model library supplied with Adice. Custom models may be written in the Adice Modeling Language or subsets of the Verilog and Verilog-A languages for simulation directly in Adice. Alternatively, co-simulations may be performed with Adice and a full-featured Verilog simulator.

In addition to the traditional dc, ac, and transient analyses common among SPICE-type simulators, Adice can also perform periodic steady state analysis and linear small signal analysis of periodically time varying circuits. These types of analysis are useful for RF circuits, switched capacitor circuits, and other circuits that are driven by a large periodic clock or carrier frequency. An interface from Adice to SpectreRF (a product of Cadence Design Systems) provides access to the additional RF simulation features of SpectreRF from within the Adice environment.

Regression Testing

Ivy allows simulation responses to be compared against known good responses, thus allowing the development of regression test simulation or self checking test benches to be developed. When developing regression test simulations using analogue responses, it is possible to define limits against which the simulated response may be checked against. The known good response for digital response may be either user defined, describing expect operation of the DUT, or the response generated by a known good simulation.

Analogue Performance Simulations

As Adice allows arbitrary mixtures of transistor-level and behavioral-level models of both analogue and digital circuitry, it is possible to verify performance of key analogue blocks even when simulating at full chip level. A full range of analogue source and capture instruments are supported in the Ivy environment that enable this.

6. Conclusions

Ivy provides a different approach to mixed signal test simulations compared to virtual test. Because we do not require an ATE test program and hardware to be developed before running simulations, we allow the test development engineer to become involved in simulations much earlier in the process. This enables the test development engineer to become fully involved in the design for testability and design for manufacturability phases of the design process. Ivy aids in the design of the test solution, whereas virtual test allows the validation of a developed test solution. Also as Ivy requires no ATE knowledge in order to be used it has been taken up by IC design engineers as their mixed signal test bench for full chip simulations also. This means that it is now very easy for the test development engineer to take a much fuller role in the simulation verification phase.

Part III: Low Power and Low Voltage

Introduction

Scaling and wireless force the circuit designer to Low Power and Low Voltage designs. The six tutorial papers of this section cover these fields.

The first introductory paper is from Klaas Bult: “Scaling and the consequences for Voltage and Power”.

The second paper by Andrea Baschirottio from the University of Lecce, Italy, and Picro Malcovati from University of Pavia, Italy, comes up with basic blocks in “Low-Voltage Low-Power Bias Circuits”.

The third paper by Peter Kinget from Columbia University, USA, discusses a “0.5 V Opamp Circuit Design”.

The fourth paper by Boris Murmann from Stanford University, USA, describes “Limits on ADC Power Dissipation”.

Wouter Serdijn from Delft University of Technology, Netherlands, treats “Ultra Low-Power Low-Voltage Analog Integrated Filter Design”.

Finally Bob Puers from Catholic University of Leuven, Belgium, shows applications of “Wireless Inductive Transfer of Power and Data”.

Johan Huijsing

The Effect of Technology Scaling on Power Dissipation in Analog Circuits

Klaas Bult

Broadcom Netherlands - Bunnik - The Netherlands

Abstract

A general approach for Power Dissipation estimates in Analog circuits as a function of Technology scaling is introduced. It is shown that as technology progresses to smaller dimensions and lower supply voltages, matching dominated circuits are expected to see a reduction in power dissipation whereas noise dominated circuits will see an increase. These finds are applied to ADC architectures like Flash and Pipeline ADC's and it is shown why Pipeline ADC's survive better on a high, thick-oxide supply voltage whereas Flash ADC's benefit from the technology's thinner oxides. As a result of these calculations an adaptation to the most popular Figure-of-Merit (FOM) for ADC's is proposed.

1. Introduction

Cost reduction and miniaturization have driven CMOS down the path of ever shrinking feature sizes and growing die sizes. Since we entered the sub-micron technologies more than ten years ago, scaling of the supply voltage has become a technological necessity. In the same period, integration of analog together with digital circuits on the same die has become an economic necessity. Systems-on-a-chip (SoC) are now State-of-the-Art.

Many papers have been published since to discuss the many challenges especially analog designers face to maintain good performance from their circuits [1-22], of which the supply voltage scaling is the most dominant one. We now have more than a decade of experience in this area and porting designs from one technology to the next has become the daily life of many analog designers.

On the premises that power can buy any performance, power dissipation has become the currency to measure the merits of a circuit. Most Figures-of-Merit (FOM) are based on power dissipation. The question what the effect of voltage scaling is on power dissipation is a very basic one. Many authors have predicted an increase in power dissipation as a result of voltage scaling [6, 14, 19, 22], some have predicted equal power dissipation [20] and some have predicted a decrease in power dissipation [15, 17, 18].

The goal of this paper is to present a unified approach to predict power dissipation under the pressure of voltage scaling and answer questions like why pipeline converters survive better on thick-oxide high- V_{dd} approaches whereas flash-based converters really benefit from technology scaling. A proper choice of architecture may be strongest trump-card in analog arsenal and selection of the most appropriate Figure-of-Merit is an important guideline in that choice.

Section 2 presents the general approach for power estimates and ends with some preliminary conclusions. Section 3 discusses the design implications of the results found in section 2. Section 4 considers the caveats of the theory used and discusses the implications of the ITRS 2003 predictions [47]. In section 5 an adaptation to a well known Figure-of-Merit is proposed and the final conclusions are discussed in section 6.

2. A General Approach for Power-Estimates in Analog Circuits as a function of Technology Scaling

In this section a more general approach to estimating the effect of Technology Scaling on Power Dissipation will be proposed. As a vehicle to drive this discussion, we will focus on a simple gain-stage as shown in fig. 1: a MOS transistor, biased by a current source, driving a load-capacitor. Many other circuits could have been chosen for this purpose (i.e. differential pair, source follower, 2-

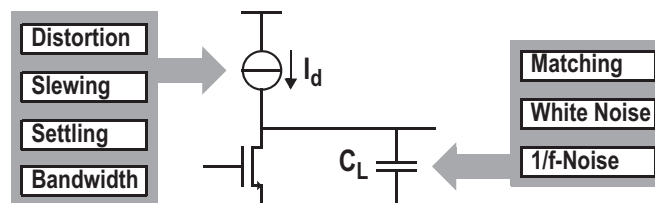


Fig.1. A single-transistor gain-stage.

stage OpAmp, etc.), but this circuit combines general purpose and simplicity and results from the following discussion can in most cases easily be translated to other examples. For simplicity reasons, we will also assume the current source and the load capacitance to be ideal, i.e. without parasitic capacitance, resistance or noise contributions.

We will start by assuming certain design specifications like Bandwidth (BW), Signal Frequency (F_{sig}), Distortion ($HD2$ or $HD3$) and Dynamic Range (DR) have to be met. Through the process of scaling those specifications will be kept constant. The calculation procedure is outlined in the next section.

2.1 Calculation Procedure and Assumptions

In the following we will use a calculation procedure similar to what has been used in [15]. In order to be able to calculate the minimal necessary power dissipation we need information from the *Process Technology*, from *System-Level* requirements and from *Circuit-Level* implementation aspects. Below, the calculation procedure will be outlined along with the basic assumptions we have been using to simplify this discussion.

From the *Process Technology* we need to know the Oxide Thickness T_{ox} and the supply voltage V_{dd} . We will assume that the Oxide Thickness scales proportional to the technology's minimum feature size, expressed as L_{min} [15]:

$$T_{ox} = \lambda L_{min} \quad (1)$$

with $\lambda \approx 0.03$ (see Fig.2). For the supply voltage we assume that, below a minimum feature size of 0.7[um], it scales proportional to the technology, i.e. to L_{min} and above 0.7[um], the supply voltage is constant, i.e. 5[V], as shown in Fig. 3. This has been the case for all technologies up to now [15] and is also predicted to be the case in the near future [47].

On order to keep the results of the calculations below as broadly applicable as possible, we will assume as little as possible on *Circuit-Level*. However, we will assume a constant Voltage Efficiency η_{vol} , defined as:

$$\eta_{vol} = \frac{V_{sig, p-p}}{V_{dd}} \quad (2)$$

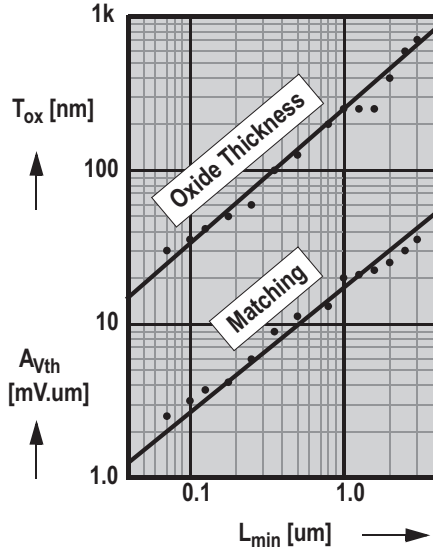


Fig.2. Oxide-Thickness T_{ox} and Matching parameter A_{Vth} versus minimum channel-length L_{min} [15].

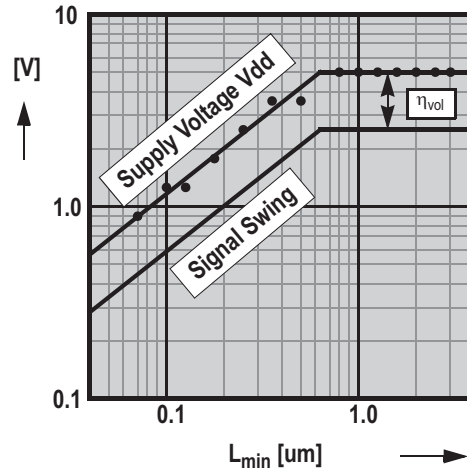


Fig.3. Supply-Voltage V_{dd} and Signal-Swing V_{sig} versus minimum channel-length L_{min} [15].

which means that we will assume that the signal swing in our circuits is a fixed percentage of the supply voltage, say for instance 80%. We will also assume a constant Current Efficiency (η_{cur}), defined as:

$$\eta_{cur} = \frac{I_d}{I_{dd}} \quad (3)$$

which means that we will assume that the total current drawn from the supply is a fixed multiple of the minimal required bias current, or in other words, we will assume that the bias current needed in a certain branch is a fixed percentage, say 25% of the total supply current. This overhead is generally used for biasing, using differential circuits, reserving some headroom for spreads, etc. The validity and reasoning behind the above assumptions are discussed in section 4.

From the *System-Level* we need to know the design goals such as the required circuit Bandwidth (BW) or maximum signal Frequencies (F_{sig}), the required Slew-Rate (SR), the required distortion level (THD) and the required Dynamic Range (DR), defined as:

$$DR = \frac{V_{sig,rms}}{\sqrt{\sum V_{unwanted}^2}}, \quad (4)$$

where $V_{unwanted}$ can consist of white noise, $V_{n,white}$, 1/f-noise, $V_{n,1/f}$ or offset voltages, V_{offset} , and $V_{sig,rms} = V_{sig,p-p}/(2\sqrt{2})$. Although in actual designs the powers of the individual unwanted signals have to be added to obtain the true Dynamic Range, in the calculations below we will assume, for simplicity, one source of unwanted signals to be dominant in each situation.

The *Calculation Procedure* is then as follows (see Fig.4). Starting from the supply voltage V_{dd} , we calculate the signal swing V_{sig} by using the voltage efficiency η_{vol} . Combining the signal swing V_{sig} and using the desired Dynamic Range (DR) allows us to calculate the maximum level of unwanted signals ($V_{unwanted}$) that can be tolerated in this design. The level of unwanted signals in turn determines the size of the Load Capacitor (C_L). Once we have the value of the Load Capacitance, we use Bandwidth (BW), Settling (SET), Slew-Rate (SR) or Distortion (THD) specifications to determine the necessary bias current (I_d) for this design. Using the current efficiency η_{cur} , we are able to calculate the required Supply Current (I_{dd}), which multiplied by the supply voltage V_{dd} yields the Power Dissipation (P). The entire procedure is depicted in Fig. 4

2.2 Dynamic Range

The Dynamic Range plays a very dominant role in determining the Power Dissipation. Using equation (4), we can determine the maximum level of the power of the unwanted signals:

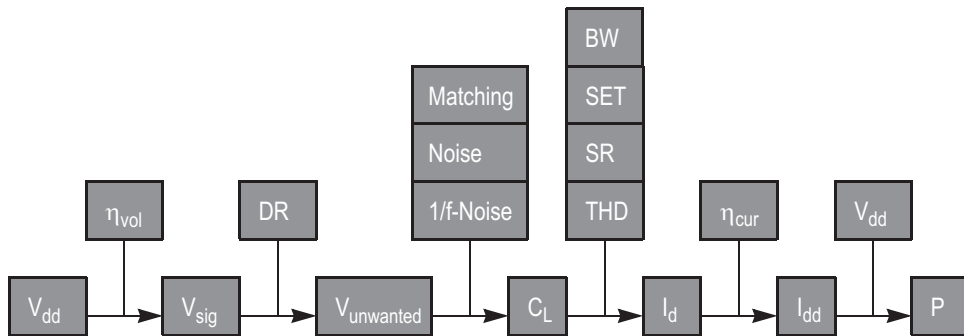


Fig. 4. The calculation procedure used in this paper.

$$V_{unwanted}^2 = \frac{V_{sig,rms}^2}{DR^2} \quad (5)$$

In the following sections, we will use either matching, thermal noise or 1/f-noise to determine the size of the capacitor needed to obtain the required Dynamic Range.

2.3 Capacitor Size

2.3.1 Matching Dominated Capacitance

Offset voltages can in certain cases, like the input stages of a Flash ADC, set the low-end limit for the Dynamic Range. Low offset voltages require large gate areas and hence large input capacitance. In this section we will assume the offset voltages to be the dominant unwanted signal:

$$V_{unwanted} = V_{offset} \quad (6)$$

As described in [23,24], the offset voltage is dependent on a Technology dependent parameter A_{Vth} and the square-root out of the gate-area:

$$V_{offset} = \frac{n_\sigma A_{Vth}}{\sqrt{WL}}, \quad (7)$$

in which n_σ equals the number of sigma's required for sufficient yield. In [25] it was shown that parameter A_{Vth} is proportional to the Oxide-Thickness:

$$A_{Vth} = \gamma T_{ox}, \quad (8)$$

in which γ is a Technology independent constant. The gate-capacitance of the transistors that have to fulfill the matching requirements can be calculated using:

$$C_{gate} = \varepsilon_0 \varepsilon_r WL / T_{ox}. \quad (9)$$

Using (5) together with (6), (7), (8) and (9) yields:

$$C_{gate} = \varepsilon_0 \varepsilon_r \gamma^2 n_{\sigma}^2 T_{ox} \frac{DR^2}{V_{sig,rms}^2}. \quad (10)$$

This is the minimum gate-capacitance of the transistors satisfying the matching requirements. Since T_{ox} is linearly dependent on *Technology* (see (1) and Fig. 2) and since V_{sig} depends on the *Technology* as shown in Fig. 3, the minimum gate-capacitance needed for sufficient matching is proportional to L_{min} above the $0.7\mu\text{m}$ *Technology* and proportional to $1/L_{min}$ below (after) that *Technology*, as shown in Fig. 5. From this figure it is clear that for matching requirements, there is a minimum gate-capacitance occurring at the $0.7\mu\text{m}$ *Technology*.

2.3.2 White-Noise Dominated Capacitance

In many cases White Noise (Thermal Noise) is setting the lower end of the Dynamic Range:

$$V_{unwanted} = V_{noise}. \quad (11)$$

Total integrated white noise power is given by:

$$V_{noise}^2 = \frac{kT}{C_{noise}}. \quad (12)$$

Using (5), (11) and (12) yields:

$$C_{noise} = kT \frac{DR^2}{V_{sig,rms}^2}. \quad (13)$$

This is the minimum value for the noise limiting capacitance. Fig. 5 depicts the dependence of C_{noise} on the *Technology* parameter L_{min} . As shown, below $0.7\mu\text{m}$, C_{noise} is inversely proportional to the square of the minimum channel-length L_{min} .

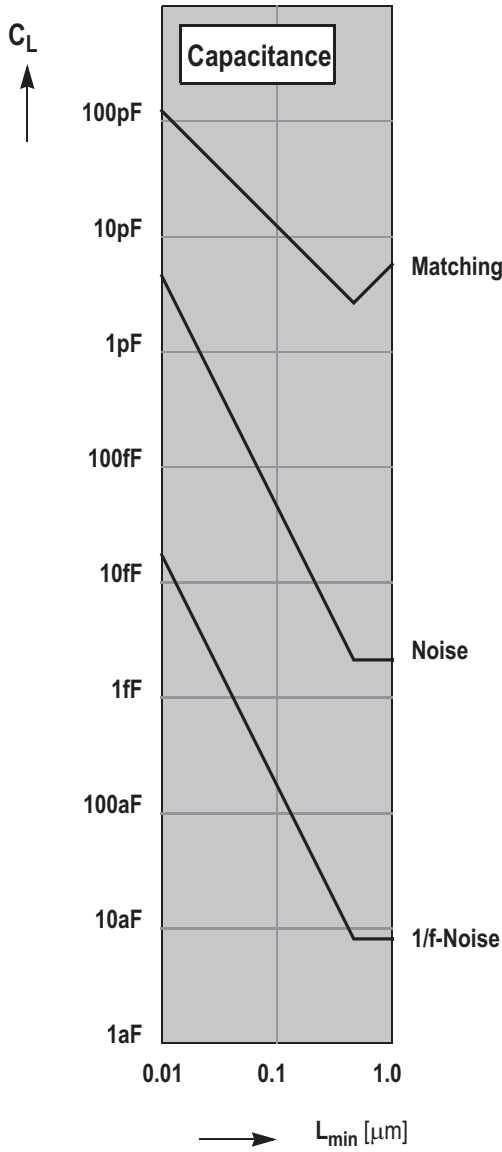


Table 1: Constants and Design Parameters

The numbers in this table have been used to generate Figs. 5, 10 and 11.

ϵ_0	8.85e-12 [C ² /J m]
ϵ_r	4
$K_{1/f}$	1e-24 [J]
k	1.38e-23 [J/K]
T	300 [K]
γ	0.9[v]
λ	0.03
n	3
V_{gt}	100 [mV]
F_{sig}	10 [MHz]
η_{cur}	25 [%]
η_{vol}	80 [%]
DR	60 [dB]
HD ₂	-60 [dB]
HD ₃	-60 [dB]
f_1	1 [Hz]
f_2	100 [MHz]

Fig. 5. Matching, 1/f-Noise and Thermal-Noise dominated Minimum Capacitance as a function of the Technologies minimum Channel-Length L_{min} . The numbers in Table 1 have been used for absolute positioning of the curves.

2.3.3 1/f-Noise Dominated Capacitance

In a certain number of situations, especially in low frequency applications, 1/f-noise is determining the lower-end of the Dynamic Range. Examples of such situations may be found in input amplifiers, in for instance Audio applications or Sensor electronics. In these situations 1/f-noise may be the dominant unwanted signal:

$$V_{unwanted} = V_{n, 1/f}. \quad (14)$$

The 1/f-noise power is given by [48, 49, 50]:

$$V_{n, 1/f}^2 = \frac{K_f}{C_{ox}WL} \frac{1}{f} \quad (15)$$

Parameter K_f is a technology determined parameter with a value of approximately 10^{-24} . In mature technologies K_f is fairly constant and in first order approximation does not seem to scale with technology [46]. In expression (15) C_{ox} is the oxide-capacitance per unit area. The gate-capacitance may be written as:

$$C_{gate} = C_{ox}WL. \quad (16)$$

Using (16) in (15) and integrating the total 1/f-noise power between frequency f_1 and f_2 yields:

$$\hat{V}_{n, 1/f}^2 = \int_{f_1}^{f_2} \frac{K_f}{C_{gate}} \frac{1}{f} df, \quad (17)$$

which in turn becomes:

$$\hat{V}_{n, 1/f}^2 = \frac{K_f}{C_{gate}} \ln(f_2/f_1). \quad (18)$$

From (18) and (5) we may calculate the 1/f-noise dominated gate-capacitance:

$$C_{1/f} = C_{gate} = \frac{\ln(f_2/f_1)K_f DR^2}{V_{sig,rms}^2}. \quad (19)$$

This capacitance is the minimum gate-capacitance that results in the desired Dynamic Range in situations where 1/f-noise dominates the unwanted signals. Fig. 5 depicts the dependence of this capacitance as a function of *Technology*. As is shown, below 0.7 μm , $C_{1/f}$ is inversely proportional to the square of the minimum channel-length L_{min} .

2.3.4 Relative Capacitor Sizes versus Technology

Fig. 5. depicts the dependency of the various capacitances on *Technology* and also their relative sizes. The numbers used for this figure are in Table 1. The results apply to 1 single device. It can be concluded from this picture, that when *Matching* comes in to play, it will be dominant over Thermal or 1/f-Noise demands with respect to capacitor size. At the 90nm *Technology* node and with 3-sigma designs for *Matching*, the ratio between the capacitor size required for *Matching* and for Thermal-Noise is about 200 and the ratio between the capacitor size for Thermal Noise and 1/f-Noise is also about 200. Of course, the 1/f-Noise dominated capacitor size depends on the low and high frequency limits f_1 and f_2 in (19). But since this dependence is only logarithmic, the value of f_2/f_1 has very little bearing on the final value of the capacitance.

In the case of *Matching*, more than 1 device is needed, like for instance in a Flash-ADC, 2^N devices would be needed. This emphasizes the need for matching improve techniques like averaging [42, 43] and offset cancellation [54] in those designs.

2.4 Current Value

In the previous section we have determined the size of the capacitor needed to obtain the desired Dynamic Range. In this section we will calculate the amount of current needed to drive that capacitance with a certain performance. We will do so under the dynamic performance conditions of Bandwidth (BW), Slew-rate (SR), Settling (SET) and Distortion. In the case of distortion we will distinguish between openloop conditions (HD2, HD3), closed loop (feedback) conditions (HD2-CL, HD3-CL) and Discrete-Time conditions (HD2-DT).

In Discrete-Time situations we will find the Clock-Frequency (F_{clk}) in many expressions. In order to simplify comparisons with the results from the Continuous-Time domain, we will assume operation at Nyquist Frequencies: $F_{sig} = F_{clk}/2$.

2.4.1 General Expression for the Transconductance Gm

To determine the required current value in various design situations, we need to have an expression for the transconductance g_m . To make these calculations as widely usable as possible, we will use an expression valid in both *Strong-* as well as *Weak Inversion*. In *Strong Inversion* the trans conductance g_m may be written as:

$$g_m = \frac{2I_d}{V_{gs} - V_t}, \quad (20)$$

whereas in *Weak Inversion* we have:

$$g_m = \frac{qI_d}{nkT} = \frac{I_d}{40mV}, \quad (21)$$

with $n \approx 1.5$ being the body-effect factor. As it appears, in both situations the transconductance can be written as the current divided by a voltage:

$$g_m = \frac{2I_d}{V_{gt}}. \quad (22)$$

In *Strong-Inversion* this voltage becomes:

$$V_{gt} = V_{gs} - V_t, \quad (23)$$

whereas in *Weak-Inversion*:

$$V_{gt} = (2nkT)/q. \quad (24)$$

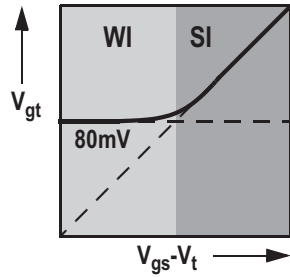
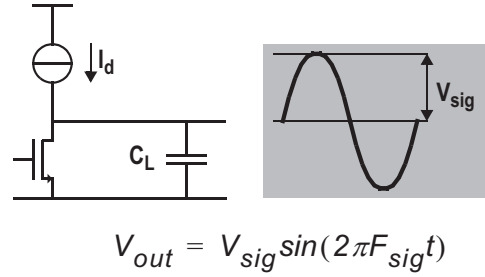


Fig. 6. V_{gt} as a function of $(V_{gs} - V_t)$ in Strong-Inversion (SI) and in Weak-Inversion (WI).



$$V_{out} = V_{sig} \sin(2\pi F_{sig} t)$$

Fig. 7. Preventing Slewng requires a charging current larger than:

$$I_{SR} = 2\pi F_{sig} V_{sig} C_L$$

Fig. 6 shows the voltage V_{gt} as a function of $(V_{gs} - V_t)$. As is shown, in Strong Inversion V_{gt} is identical to $(V_{gs} - V_t)$, whereas in Weak Inversion V_{gt} limits to a value of about 80mV at room temperature. Equation (22) is now valid at all points of operation and is a very useful expression. We will use this expression in the following sections on determining the required amount of bias current I_d under different conditions.

2.4.2 Bandwidth Dominated Current

In certain situations, especially when signal-levels are low, only a certain Bandwidth (BW) is required. The maximum achievable bandwidth is the Unity-Gain Frequency (F_u). Any DC-gain A_0 that may be required lowers the bandwidth by the same amount. This might be achieved by applying a load resistor in parallel to C_L in the circuit of Fig. 1, or by using feedback. In either case, the bandwidth is set by:

$$BW = \frac{F_u}{A_0} = \frac{g_m}{2\pi A_0 C_{gate}}. \quad (25)$$

To maintain uniformity with other results to come in this paper, we will use F_{sig} for the maximum signal frequency in our expressions: $F_{sig} = BW$. Substituting (22) in (25) and solving for the required current results in:

$$I_{BW} = \pi V_{gt} F_{sig} A_0 C_{gate}. \quad (26)$$

Combining equation (26) with (23) and (24) (see Fig. 6) brings us to the conclusion that Bandwidth is achieved for the lowest current in *Weak-Inversion*, as that achieves the lowest effective V_{gt} .

2.4.3 Slew-Rate Dominated Current

In many situations however, signal swings are larger and only achieving a certain Bandwidth is not sufficient. In order to prevent Slewing (see Fig. 7), the required current is:

$$I_{SR} = 2\pi F_{sig} V_{sig} C_L. \quad (27)$$

Comparing to (26) shows that, usually, the minimum current to prevent Slewing is larger than the current required for a certain Bandwidth. However, the two requirements become equal for:

$$V_{sig} = V_{gt}/2. \quad (28)$$

This means that signal-swings have to be small enough so that the active device is never overwhelmed by the input signal. As V_{gt} according to (24) and Fig. 6 never goes below $80mV$ (at room temperature), this means that signal-levels below $40mV$ will never cause Slewing.

2.4.4. Distortion Dominated Current in Openloop Condition

In many continuous-time cases however, preventing Slewing is not sufficient either and a certain Distortion level has to be achieved. This situation is more complex than the previous situations and depends on whether Feedback is used or not and whether we operate under Continuous-Time or Discrete-Time conditions. In this section we will assume no Feedback (Open-Loop) and a Continuous-Time situation. Considering the circuit and its Frequency Domain transfer-function in Fig. 8, it has been shown [51] that:

$$HD_2 = \frac{1V_{in}}{4V_{gt}}. \quad (29)$$

However, (29) is a function of the input amplitude V_{in} . To find the HD_2 as a function of the output signal amplitude V_{sig} , we have to know the gain between input and output at the frequency of interest. We will assume a transfer function as shown in Fig. 8 and furthermore we will also assume that the input signal frequency is in that part of the transfer function where the slope is -1 and below the Unity-Gain Frequency (F_u). This is the most interesting situation since going beyond F_u would mean an attenuation instead of a gain and going to frequencies below the -3dB-Bandwidth of the circuit makes achieving good THD only easier. Under these conditions (29) may be re-written as:

$$HD_2 = \frac{1V_{sig}}{4AV_{gt}} = \frac{1V_{sig}F_{sig}}{4V_{gt}F_u}, \quad (30)$$

where A equals the gain at F_{sig} and A is substituted by F_u/F_{sig} (see Fig. 8). The Unity-Gain Frequency (F_u) in turn is a function of g_m and C_L :

$$F_u = g_m/(2\pi C_L). \quad (31)$$

Substituting (22) and (31) in (30) and solving for I_d yields:

$$I_{HD2} = I_d = \frac{\pi V_{sig} F_{sig} C_L}{4 HD_2}. \quad (32)$$

As is shown in Appendix A, a similar calculation for differential pairs yields a third-harmonic distortion:

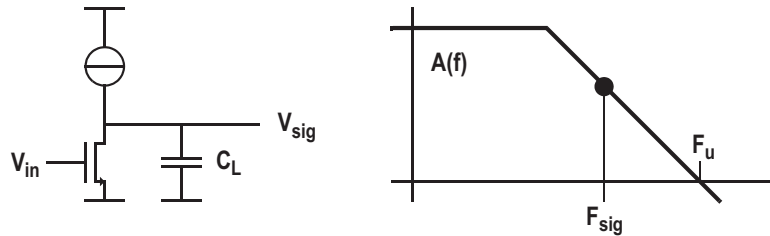


Fig. 8. Single-Transistor Gain-Stage and its Frequency-Domain transfer-function

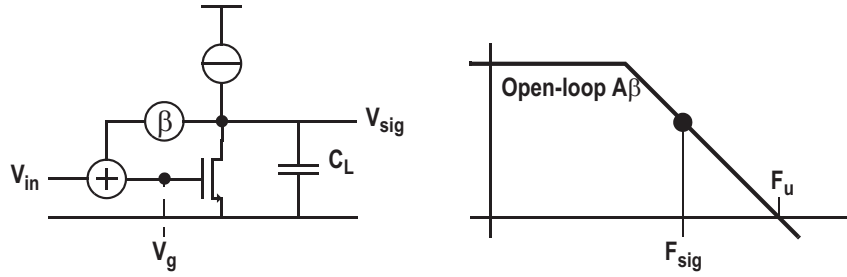


Fig. 9. Single-Transistor Amplifier in Feedback situation. On the right is the Bode-diagram of the Loop-Gain $A\beta$.

$$I_{HD3} = \frac{\pi}{2\sqrt{2}} \frac{V_{sig} F_{sig} C_L}{\sqrt{HD_3}}. \quad (33)$$

Equation (32) shows that achieving better distortion means increasing power dissipation. Comparing (32) to (27) reveals that, in general, achieving good distortion takes more current than just preventing Slewing. The ratio between the required currents is:

$$\frac{I_{HD2}}{I_{SR}} = \frac{8}{HD_2}. \quad (34)$$

Equation (34) shows that on the edge of Slewing the circuit of Fig. 8 achieves $HD_2 = 1/8$, in general a very poor distortion number. Requiring better Distortion numbers increases the bias current inversely proportional to the decrease of the Distortion.

2.4.5. Distortion Dominated Current in Feedback Condition

Using Feedback improves the Distortion and therefore the required amount of current for a certain distortion level will go down. Fig. 9 shows the circuit of Fig. 8 with a negative feedback factor β applied to it. An Open-loop transfer-function $A\beta$ is assumed as depicted in Fig. 9 on the right hand side. As many textbooks [48] show, the Distortion in a feedback situation improves with the amount of loop-gain $A\beta$. Assuming operation in the point indicated in the graph of fig. 9 and using (29) divided by $A\beta$ we obtain:

$$HD_2 = \frac{1V_{in}}{4V_{gt}} \frac{1}{A\beta} = \frac{1}{4} \frac{V_{sig}}{A^2 \beta V_{gt}} = \frac{1}{4\beta V_{gt}} \left(\frac{F_{sig}}{F_u} \right)^2 \quad (35)$$

in which the gain A was substituted by: $A = F_u / F_{sig}$ (see Fig. 9). Substituting (31) and (22) in to (35) leads to:

$$I_d^2 = \frac{\pi^2 V_{sig}^2 F_{sig}^2 C_L^2}{4 \beta HD_2}, \quad (36)$$

which may be re-written as:

$$I_{HD2-CL} = I_d = \frac{\pi V_{sig}^{\frac{1}{2}} V_{gt}^{\frac{1}{2}} F_{sig} C_L}{2 \sqrt{\beta HD_2}}. \quad (37)$$

Comparison with (32) reveals that the dominant effect of feedback on the current required for a certain amount of distortion is that, instead of dividing by HD_2 (equation (32)), (37) shows a division by $\sqrt{\beta HD_2}$, where especially the square-root helps dramatically in reducing the current demands.

In Appendix B a derivation is outlined similar to the one in this paragraph where it is shown that in differential pairs the required amount of current for the 3rd Harmonic Distortion (HD_3) is:

$$I_{HD3-CL} = \frac{\pi V_{sig}^{\frac{2}{3}} V_{gt}^{\frac{1}{3}} F_{sig} C_L}{4 \sqrt[3]{\beta HD_3}}, \quad (38)$$

and in general, from a Taylor series expansion:

$$I_{HDn-CL} \propto \frac{\pi V_{sig}^{\frac{n-1}{n}} V_{gt}^{\frac{1}{n}} F_{sig} C_L}{2^{n-1} \sqrt[n]{\beta HD_n}}. \quad (39)$$

2.4.6. Settling Dominated Current

Preventing Slewing (section 2.4.3) may not be sufficient in many cases and settling to an absolute accuracy is required. This is the case in Pipeline ADC's for instance. In this case we are operating in a discrete time-domain at a clock frequency of F_{clk} . Usually, half a clock period is available for settling. We will assume that settling to an accuracy of $1/DR$ is required:

$$\frac{1}{DR} = \exp\left(-\frac{T_{clk}}{2\tau}\right) = \exp\left(\frac{-g_m}{4F_{clk}C_L}\right). \quad (40)$$

Taking the logarithm of both sides of (40) and using (22) for g_m yields:

$$\ln(DR) = \frac{g_m}{4F_{clk}C_L} = \frac{I_{Set}}{2F_{clk}C_L V_{gt}}, \quad (41)$$

which leads to:

$$I_{Set} = \ln(DR)4F_{sig}C_L V_{gt}, \quad (42)$$

where operation at Nyquist-frequency ($F_{sig} = F_{clk}/2$) is assumed. The above equation shows a very weak dependence on distortion and signal swing and strong dependence on signal frequency, load capacitance and transistor biasing (V_{gt}). Comparing this result to Bandwidth dominated situations (equation (26)) reveals that settling to an absolute accuracy is a more stringent requirement than achieving a certain Bandwidth. The ratio between the two requirements is:

$$\frac{I_{Set}}{I_{BW}} = \frac{4}{\pi} \ln(DR). \quad (43)$$

For a 10-bit system ($DR = 60\text{dB}$) that ratio is approximately 8.8, indicating that settling to an accuracy of 60dB requires approximately 8.8 times more current than just achieving sufficient Bandwidth for the same signal frequency. Apart from this Dynamic Range dependent ratio, the behavior of this current requirement as a function of Technology is the same as that of Bandwidth dominated situations.

2.4.7. Settling to a certain Distortion-Level

Not in all Discrete-Time applications settling to an absolute accuracy (1/DR) is required. Sometimes settling until a certain Distortion-Level is achieved is sufficient, like for instance in many Switched-Capacitor circuits. In this section we calculate how much current it takes to settle to a certain pre-described Distortion-Level.

As shown in section 2.4.5 equation (35), the Closed-Loop Continuous-Time Distortion is:

$$HD_2 = \frac{1V_{in}}{4V_{gt}} \frac{1}{A\beta} = \frac{1}{4} \frac{V_{sig}}{A^2\beta V_{gt}}. \quad (44)$$

From this we see that to obtain a level of Distortion HD_2 , a gain would be required of:

$$A = \sqrt{\frac{V_{sig}}{4HD_2\beta V_{gt}}}. \quad (45)$$

In a Discrete-Time situation, this means that we must have settled sufficiently to obtain the equivalent of this gain. That means:

$$A = \exp\left(\frac{T_{clk}}{2\tau}\right) = \exp\left(\frac{g_m}{2F_{clk}C_L}\right). \quad (46)$$

Equating (45) to (46), substituting (22) and solving for I_d results in:

$$I_d = \ln\left(\sqrt{\frac{V_{sig}}{4\beta HD_2 V_{gt}}}\right) F_{clk} C_L V_{gt}. \quad (47)$$

Assuming Nyquist-Frequency operation, the required current for obtaining a certain Distortion-level in a Discrete-Time situation is:

$$I_{HD2-DT} = \ln\left(\frac{V_{sig}}{4\beta HD_2 V_{gt}}\right) F_{sig} C_L V_{gt} . \quad (48)$$

This expression has a similar shape as the result found in the previous section for settling. It is very weakly dependent on the distortion and signal swing and strongly dependent on signal frequency, load capacitance and transistor biasing (V_{gt}). Taking the ratio between this result (48) and the result obtained in the previous section on the required current for settling to an absolute accuracy (42) results in:

$$\frac{I_{HD2-DT}}{I_{Set}} = \frac{\ln\left(\frac{V_{sig}}{4\beta HD_2 V_{gt}}\right)}{4\ln(DR)} , \quad (49)$$

which shows that under most circumstances settling to achieve a certain Distortion-Level requires only 25% of the current required for settling to absolute accuracy.

The requirements for a certain distortion-level in Continuous-Time (37) versus Discrete-Time (47) conditions are equal if:

$$\ln\left(\frac{V_{sig}}{4\beta HD_2 V_{gt}}\right) = \frac{\pi}{4} \sqrt{\frac{V_{sig}}{\beta V_{gt} HD_2}} . \quad (50)$$

From this equation it is easy to see that the requirement of a certain continuous-time distortion-level is always more difficult to meet than the requirement of settling to the required gain for that distortion-level.

2.4.8. Relative Current Values versus Technology

Fig. 10 shows all previously discussed bias-currents as a function of the Technology (L_{min}). The equations used to generate these curves are (26), (27), (32), (33), (37), (38), (42) and (48). To determine the absolute positions of the curves, the design parameters listed in Table 1 have been used. As can be seen

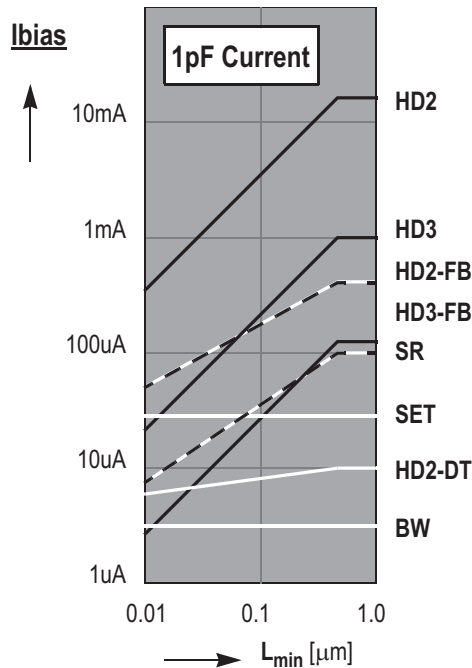


Fig. 10.

The effect of Technology Scaling on the current necessary to drive a $1[\text{pF}]$ capacitance. The numbers listed in Table 1 have been used to obtain the absolute position of the curves. For bandwidth (BW) a gain of $A_0 = 1$ is used.

from the figure, the curves are grouped in 3 different groups: the bandwidth and settling related group (solid white lines), the distortion in feedback group (grey lines) and the slew-rate and open-loop distortion group.

The solid white lines include settling (SET), discrete-time distortion (HD2-DT) and bandwidth (BW). These situations show either no, or very small dependence on Technology. The current ratio between settling (SET) and bandwidth (BW) dominated situations is given by (43) and is approximately 8.8 at this level of dynamic range (DR). The discrete-time distortion curve (HD2-DT) is right in between these two situations for the somewhat older technologies and goes slightly in the direction of bandwidth for the newer technologies. All 3 situations are independent (or almost independent) of Technology.

The grey lines indicates the distortion in feedback circuits (HD2-FB and HD3-FB). These curves show a decrease in power as a result of Technology scaling, although the effect is less than proportional. It is obvious from Fig. 10 that feedback helps to achieve a certain distortion-level at a much lower current. However, the feedback requires a certain loop-gain from the amplifier which in turn requires a certain amount of current. It is the dependency on the loop-gain of the current that causes the reduction in current consumption to be less than proportional to the Technology (see (35) and (37)).

The group with the solid black lines include HD2, HD3 and SR and shows a reduction in current proportional to Technology. However, especially in the older technologies, they require significantly more bias-current than the bandwidth or settling situations.

It is important to note from Fig. 10 that there are huge differences in the levels of current required in the various situations, ranging from 3uA (Bandwidth) to 20mA (HD_2 , towards the larger L_{min}). The bandwidth (BW) dominated situation requires the least amount of current, as was expected. Again, in this graph a constant V_{gt} of 100mV is assumed for this situation. In reality, V_{gt} could be chosen differently, mostly higher, especially in older technologies ($> 0.13\mu\text{m}$). This would cause the curve (actually more a band of curves for this situation) to bend up towards the older technologies. For newer technologies V_{gt} is (under most circumstances) going to get lower but will be limited to 80mV once weak-inversion operation is reached (see also Fig. 6).

At 100nm Technology (where an assumption of $V_{gt} = 100\text{mV}$ is not far from reality), the ratio between Slew-Rate and Bandwidth dominated bias-current is about 8. The ratio between the Distortion (HD2) and Slew-Rate (SR) dominated bias-current (given by (34)) equals 125 at $HD_2 = 60\text{dB}$, for all Technologies. It can be concluded that, if (continuous-time) distortion is important, it will be dominating the required amount of bias-current, rather than Slew-Rate, Settling or Bandwidth. Discrete-Time distortion (HD2-DT) behaves more like the bandwidth or settling situation. It only shows a very weak dependence on Technology. Below the pivotal Technology of $0.7\mu\text{m}$ the current required for achieving that distortion-level decreases slightly, but not dramatically.

Please note that this graph assumes a fixed load-capacitance and scaling of the load-capacitance is not included in these results. The next section will include the effect of Technology scaling on both the load-capacitance as well as the current.

2.5 Power Dissipation

In the previous paragraphs we have derived expressions for the capacitor and the current values under different circumstances. Each of the expressions for the capacitor sizes (section 2.3) can now be combined with each of the expressions for the current value (section 2.4) to obtain a value for the device bias current I_d . To calculate the current drawn from the supply we will use the inverse of (3):

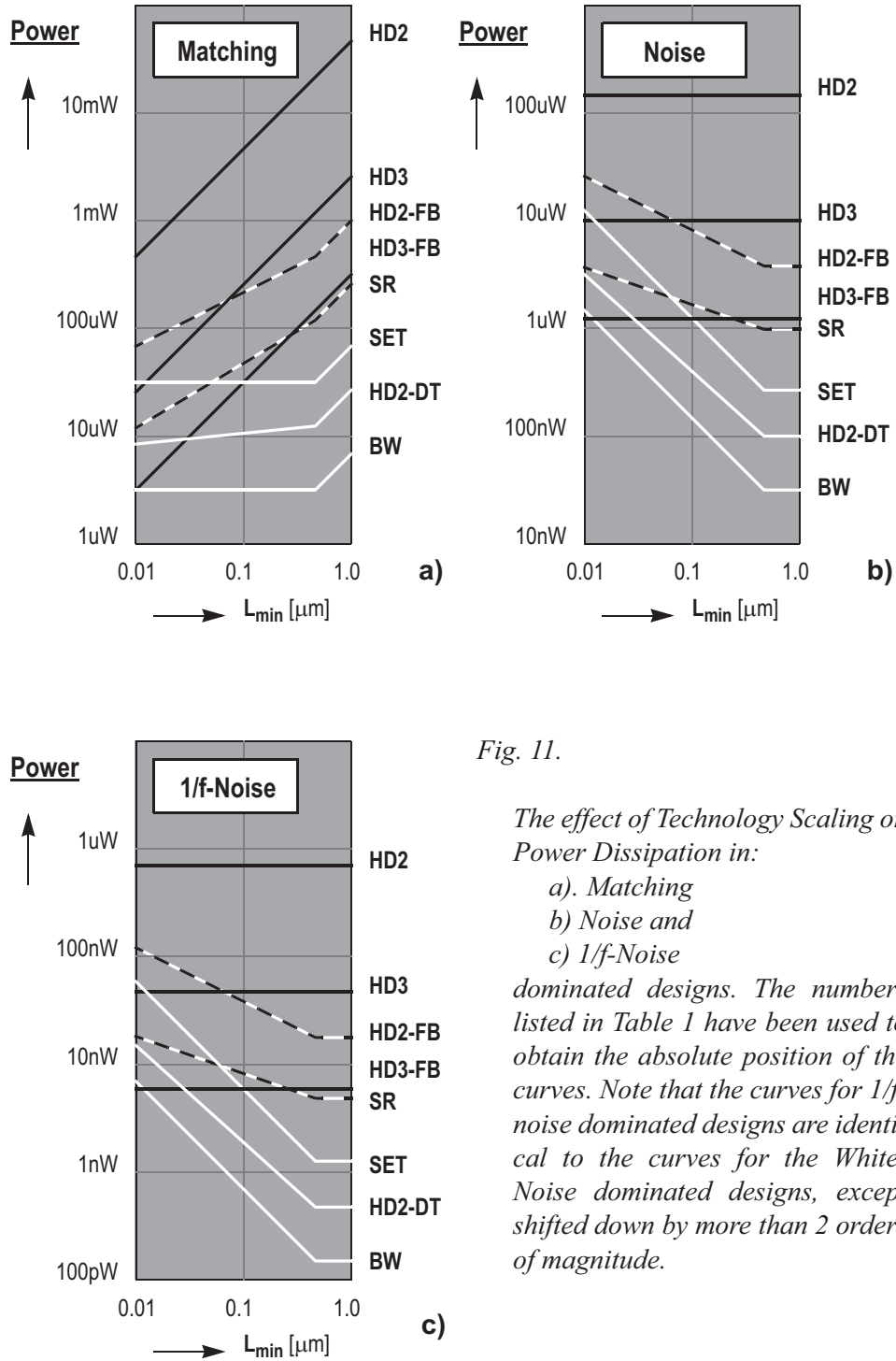


Fig. 11.

The effect of Technology Scaling on Power Dissipation in:

- a). Matching
- b) Noise and
- c) 1/f-Noise

dominated designs. The numbers listed in Table 1 have been used to obtain the absolute position of the curves. Note that the curves for 1/f-noise dominated designs are identical to the curves for the White-Noise dominated designs, except shifted down by more than 2 orders of magnitude.

Table 2. Expressions for Power Dissipation as a function of Technology and Design parameters.

Note that all expressions need to be multiplied by the factor in the upper-left corner and by the appropriate factor from the 2nd column.

Power Estimate $\frac{2\pi F_{sig} DR^2}{\eta_{vol} \eta_{cur}}$	Matching	Noise	1/f-Noise
Multiplying Factor	$\varepsilon_0 \varepsilon_r \gamma^2 n_{\sigma}^2$	kT	$\ln(f_2/f_1) K_f$
BW	$\frac{T_{ox} V_{gt}}{2A_0 \eta_{vol} V_{dd}}$	$\frac{V_{gt}}{2A_0 \eta_{vol} V_{dd}}$	
HD2-DT	$\frac{T_{ox} \ln\left(\frac{\eta_{vol} V_{dd}}{4\beta HD_2 V_{gt}}\right) V_{gt}}{2\pi \eta_{vol} V_{dd}}$	$\frac{\ln\left(\frac{\eta_{vol} V_{dd}}{4\beta HD_2 V_{gt}}\right) V_{gt}}{2\pi \eta_{vol} V_{dd}}$	
SET	$\frac{T_{ox} 2\ln(DR) V_{gt}}{\pi \eta_{vol} V_{dd}}$	$\frac{2\ln(DR) V_{gt}}{\pi \eta_{vol} V_{dd}}$	
SR	T_{ox}	1	
HD2-FB	$\frac{T_{ox} \sqrt{V_{gt}}}{\sqrt{\eta_{vol} V_{dd} \beta HD_2}}$	$\frac{\sqrt{V_{gt}}}{\sqrt{\eta_{vol} V_{dd} \beta HD_2}}$	
HD3	$\frac{T_{ox}}{4\sqrt{HD_3}}$	$\frac{1}{4\sqrt{HD_3}}$	
HD2	$\frac{T_{ox}}{8HD_2}$	$\frac{1}{8HD_2}$	

$$I_{dd} = \frac{I_d}{\eta_{cur}}. \quad (51)$$

Multiplying by V_{dd} yields the Power Dissipation:

$$P = I_{dd}V_{dd}. \quad (52)$$

Table 2 shows an overview of all expressions obtained in the above described manner. Equation (2) was used to substitute $\eta_{vol}V_{dd}$ for V_{sig} , in order to show the dependence on the supply voltage. Note the multiplication factor in the upper left corner which needs to be applied to all results in the table. That factor, which contains F_{sig} , DR^2 , η_{vol} and η_{cur} , is common to all the results. For each column there is also a separate multiplication factor in the 2nd row of the table. Please note that although Noise and 1/f-Noise are combined in to 1 single column, they do have separate multiplication factors in the 2nd row.

To get a feeling for the relative values of these expressions Fig. 11 was produced. In order to get a consistent comparison, the numbers from Table 1 have been used again. Please note that the different y-axis in Fig. 11a, 11b and 11c are shifted from each other. Although the numbers chosen as design goals are arbitrary, it is clear from this figure that if Matching is important (Fig. 11a), it will dominate the power requirements. The worst-case corner occurs when the capacitance is determined by matching and the current by (openloop) distortion requirements. It is also clear that the effect of 1/f-noise on power dissipation is indeed very small.

2.5.1 Power Dissipation in Matching dominated designs

The most striking aspect of the matching dominated power dissipation depicted in Fig. 11a, is that non of the curves scale up with technology. This means that all matching dominate designs are either indifferent or benefit from technology scaling. In the cases of the bandwidth-limited and discrete-time designs (the white lines), the power dissipation goes down until the pivotal technology of 0.7 μ m after which it remains flat and independent of technology. These situations however do require the least amount of power.

2.5.2 Power Dissipation in White and 1/f-Noise dominated designs

Fig. 11b shows the noise dominated design situations. Note that all the curves are relatively low compared to the curves from Fig. 11a. It is noteworthy that, in contrast to the matching dominated design situations, non of the curves in Fig. 11b show a decrease in power dissipation when technology progresses to smaller dimensions. The openloop distortion and slew-rate curves put the highest demands on power dissipation, but remain flat and independent of technology, whereas all other curves show an increasing power as a result of technology scaling. Note that the curves in Fig. 11c (1/f-Noise) are a scaled copy of the curves in Fig. 11b (White Noise). This can also be seen in Table 2. From that table it is clear that the ratio in power dissipation between these two situations is $kT/(\ln(f_2/f_1)K_f)$, which is approximately 200 at room temperature.

2.6 Preliminary Conclusions

Assuming our starting points as stated in section 2.1 are still valid (to which we will come back in paragraph 4), we may come to the following conclusions.

Matching dominated design requires the highest amount of power dissipation for a given set of design specifications. The combination with openloop distortion (solid black lines in Fig. 11a) is the most severe with respect to power, but is also more rare in real designs. Slew-Rate and matching is a combination that behaves in the same way as matching and distortion. An example of such a combination can be found in the pre-amps of Flash ADC's. Matching together with distortion specifications in a feedback situation (grey lines in Fig. 11a) is a very common combination. An example of such a situation is for instance a Track & Hold amplifier during the track-phase, while driving a Flash-based ADC. But although these situations require the highest amount of power dissipation, the scaling with technology is beneficial, as the power decreases with the technology's minimum feature-size.

The opposite can be said about the combination of noise and distortion in feedback (the grey lines in Fig.11b). This is a very common situation and can be found for instance in a Track & Hold amplifier in track-mode (driving a Pipeline ADC), in Fixed-Gain amplifiers or in some continuous-time filters (like Opamp-RC filters). The required amount of power is significantly lower than in matching dominated situations and also significantly lower than in openloop situations. However, this combination does not benefit from technology scaling, but shows

an increase in power dissipation for shrinking technologies. For HD2-FB we find an increase of about 3 times per decade of technology scaling, whereas for the more common HD3-FB an increase of about 2 times is found.

The combination of noise and openloop distortion, as may be found in LNA's and gm-C filters, shows no dependence on technology in Fig. 11b (solid black lines). This means that these types of circuit could scale well, as long as the signal-swing scales with the supply voltage (equation (2)) and does not get dictated by external factors.

Table 3: *Designs examples categorized according to the dominating performance parameters.*

Design Situation	Matching	Noise & 1/f-Noise
BW		- Sensor Pre-Amp - LNA
HD2-DT HD3-DT	- T&H in Hold-mode driving a Flash-ADC	- Switched-Capacitor Circuits - T&H in Hold-mode while driving a Pipeline ADC
SET	- Folding ADC's	- Pipeline ADC - Cyclic ADC
SR	- Flash-ADC - Sub-Ranging ADC	
HD2-FB HD3-FB	- T&H in Track-mode driving a Flash-ADC	- OpAmp-RC Filters
HD2 HD3		- LNA - gm-C Filters

As already concluded in section 2.4.6, direct comparison between the required power-levels for achieving a certain distortion-level in Continuous-Time versus Discrete-Time situations (HD2 versus HD2-DT) shows that it is easier to achieve a certain distortion-level in Discrete-Time situations than in Continuous-Time situations. However, Technology scaling has an adverse effect on the power dissipation in Discrete-Time situations if combined with Noise requirements (solid white lines in Fig. 11b). As can be clearly seen in Fig. 11b, for bandwidth limited (BW), discrete-time distortion limited (HD2-DT) or settling limited designs (SET), the power dissipation will go up for shrinking tech-

nologies. This is a point of concern since this could apply to many design situations. As will be discussed in section 3, these conclusions ask for a judicious choice of architecture to take advantage as much as possible of the benefits technology scaling is offering.

Table 3 shows an overview of design examples classified according to the parameters that dominate the capacitance and the current

2.7 Data from Literature

To show that the estimates for power dissipation derived in this section are close to reality, information on 6-bit ADC's was gathered out of the open literature [59-73] to enable a comparison between measured data on power dissipation and power estimates based on the above theory. Six bit ADC's were chosen for this comparison as most 6-bit design share the same architecture, that of a Flash-ADC as depicted in Fig. 13.

We assume a Track & Hold Amplifier, 2 arrays of pre-amplifiers, an array of comparators and decoding logic. Furthermore we assume that there are 64 pre-amplifiers or comparators in each array. It is reasonable to assume that the power dissipation of the Track & Hold Amplifier equals the total power dissipation of the pre-amps of the first array and that the sum of the power dissipation of the 2nd array and the comparator array also equals the power consumed by the first array. In that case we have 3 sections with equal power dissipation: the T&H amplifier, the 1st array of pre-amps and the combination of the 2nd array of pre-amps and the comparators.

We use equations (10) and (27) to obtain a Matching and Slew-Rate dominated power estimate of 1 single pre-amp of the first array and multiply that times 64 for the number of stages and times 3 to include the T&H amplifier, the 2nd array of pre-amps and the comparators. Each paper [59-73] quotes a number for the Effective Number of Bits (ENOB) and Resolution Bandwidth. In our estimates we use 2^{ENOB} as an estimate for the Dynamic range (DR) and the Resolution BW as an estimate for the maximum signal frequency F_{sig} . Table 4 gives an overview of the numbers used in the estimates.

Using the above approach we made estimates for the power dissipation of each design based on the ENOB and Resolution BandWidth numbers. Fig. 12 shows a plot of the Power Dissipation quoted in the paper versus our estimate for that design. The thick line indicates the position at which the estimated power

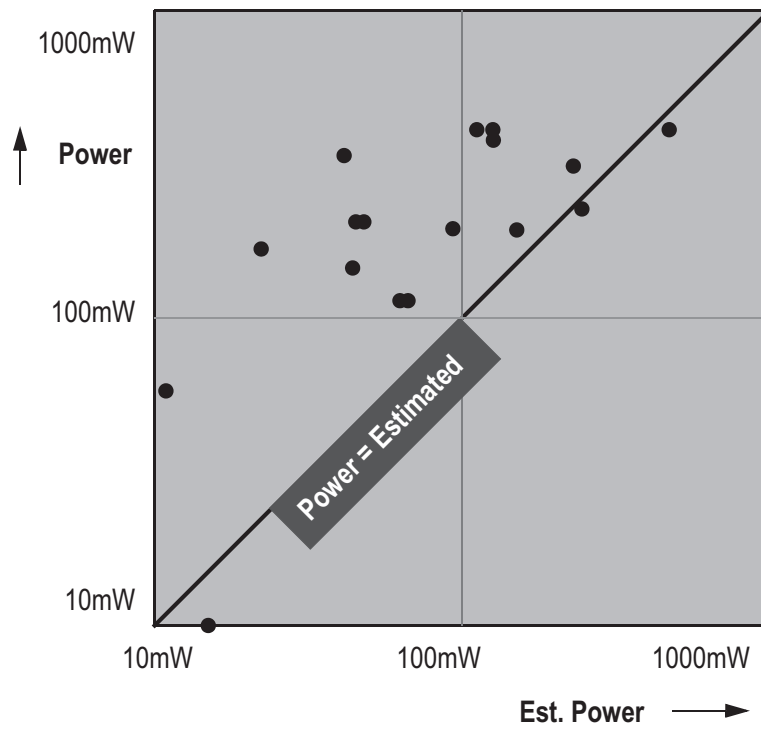


Fig. 12. Power Dissipation of 6-bit Flash ADC's from literature [15] versus a Power Estimate using the method described in this section.

Table 4: Constants and Design Parameters used in generating Fig. 12.

N	$3 \times 64 = 192$
Fsig	Resolution BW from paper
DR	2^{ENOB} from paper
η_{vol}	60 [%]
η_{cur}	25 [%]
γ	1.0 [V]
n	3 sigma

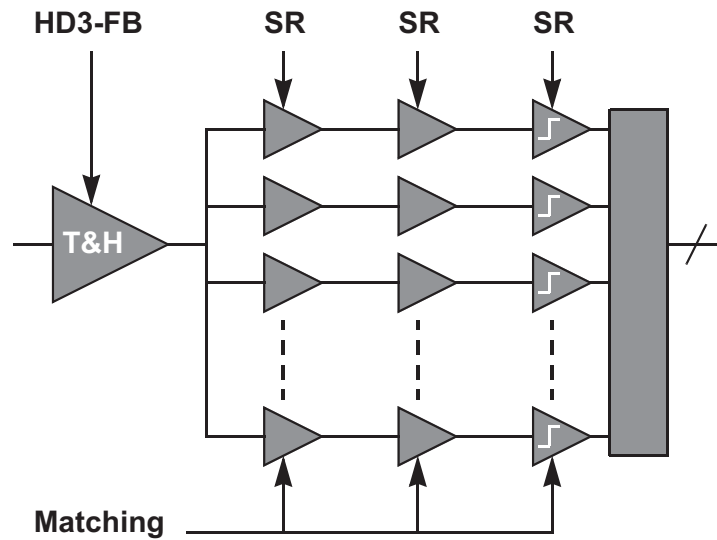


Fig. 13. A Flash-based ADC. The dominant capacitance and current determining factors are indicated in the figure.

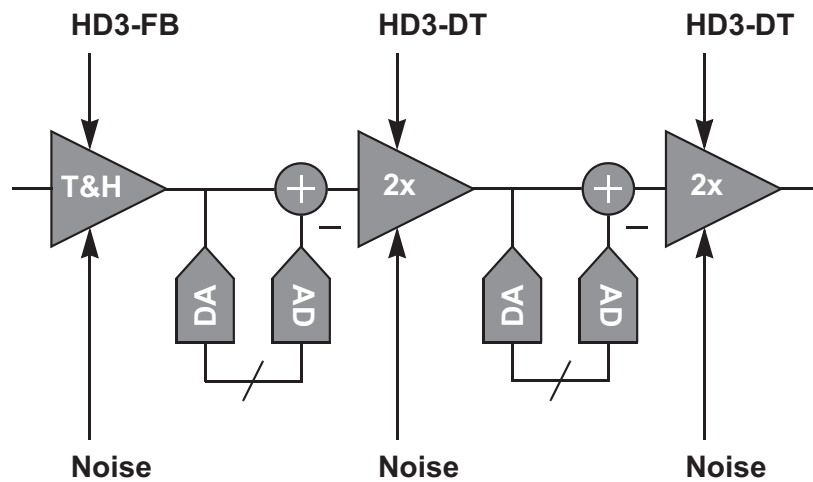


Fig. 14. A Pipeline ADC. The dominant capacitance and current determining factors are indicated in the figure.

exactly equals the measured power dissipation. As can be seen from Fig. 12, the estimate is very close to the real power dissipation, with in most cases the measured power dissipation some what higher than the estimate. There are 3 designs that quote a slightly lower power dissipation than what we estimated. Keep in mind though that none of the papers quoted a yield number ($n\sigma$) or mentioned anything about voltage (η_{vol}) or current efficiency (η_{cur}). For those parameters the estimates of table 4 were used.

With most of the designs very close to the power estimate and at least within half an order of magnitude, Fig. 12 shows that the power estimation procedure described in this section are not merely of theoretical value, but are very close to reality.

3. Design Implications

The results from the previous paragraph indicate that there are huge differences in both power-level as well as the way the power dissipation scales with technology, dependent on the architecture chosen. In order to achieve minimum power-level in general and benefit as much as possible from the technology scaling, important choices have to be made on architectural level. In this paragraph we will indicate some of the consequences of the finds of the previous paragraphs.

3.1 Necessity of Matching Improvement Techniques

As was concluded in the previous paragraphs, and is also clearly shown by Fig. 11, matching demands increase power dissipation enormously. On top of that, one has to consider that, when matching is concerned, usually multiple elements (N) come in to play, as was clearly shown in the example of section 2.7 ($N=192$). This stresses the need for techniques that improve matching, like for instance Averaging [42, 43], Calibration [56, 57], Offset-cancellation [54], Chopping and Dynamic Element Matching [58]. Orders of magnitude improvement in power dissipation can be obtained and sometimes combinations of different techniques can be used.

Using these techniques brings great benefits but usually does not change the behavior of the architecture with respect to the influence of technology scaling on power dissipation. However, if pushed to the extreme, an initially matching dominated architecture may end up to become noise dominated instead.

3.2 Pipeline versus Flash-based ADC's

Figs. 14 and 15 show generic circuit diagrams of a Flash and a Pipeline ADC. Indicated in the figures are the dominant capacitance and current determining factors.

As is shown, the Flash-ADC is limited by matching and Slew-rate and scales with technology as indicated by the grey line labeled SR in Fig. 11a. The worst-case point in the diagram of Fig. 13 however, is the Track & Hold amplifier driving the matching dominated input capacitance of the Flash-ADC, which scales less than proportional to technology, as shown by the curve labeled HD3-FB in Fig.11a. This would lead to the conclusion that as technology shrinks, the power dissipation in the ADC goes down faster than the power dissipation in the T&H amplifier. This coincides with what designers are observing these days, that the T&H amplifier becomes more and more dominant in the power dissipation of a Flash-based ADC.

The Pipeline-ADC (Fig. 14) is dominated by noise and distortion in a feedback situation (HD3-FB). The dependence of power dissipation of this ADC on technology is shown in Fig. 11b, as indicated by the grey line labeled HD3-FB. As the figure shows, this combination of requirements leads to a power dissipation which increases as technology progresses to smaller dimensions. Fortunately, the increase is not very strong and equals to approximately a doubling of the power dissipation for every decade in technology scaling. Nevertheless, it can be concluded here that Pipeline-ADC's do not benefit from technology scaling. This also agrees with what designers are experiencing nowadays and explains why Pipeline-ADC's are usually designed using Thick-Oxide transistors running from a higher supply voltage (i.e. 3.3[V] or 2.5[V] in a 1.8 [V] or 1.2 [V] process).

Comparing the power-scaling of Flash-based ADC's to Pipeline ADC's could lead to a preliminary conclusion that Flash-based ADC benefit from technology scaling while Pipeline-ADC's have to resort to the thicker oxides usually available in deep sub-micron processes. This raises the question whether over time more ADC design will be Flash-based.

4. Caveats and ITRS predictions

The results calculated in paragraph 2 and the preliminary conclusions from paragraph 3 are all based on the assumptions stated in paragraph 2.1. Linear scaling of the technology was assumed, together with continuing improvement of matching as a result of technology scaling. Moreover, a constant current and voltage efficiency was assumed, of which at least the latter may be questionable. In this paragraph we will discuss some of these starting points and evaluate to which extend the conclusions from the previous paragraph remain valid.

4.1 Scaling of Matching

In the early phases of a technology node matching is usually considerably worse than when the process is well matured. This inevitably leads to discussions among designers about the scaling of matching. Theory [25] predicts that

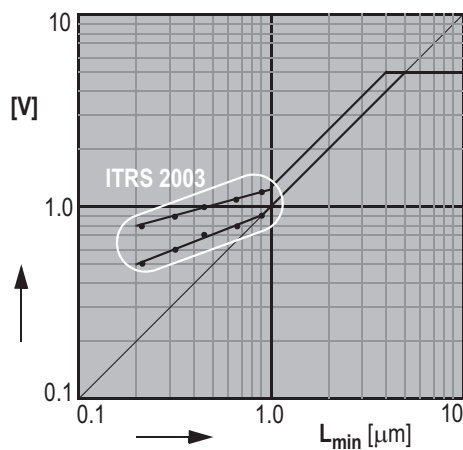


Fig. 15. The ITRS 2003 predictions of the high-performance and low-power digital supply voltage for 90nm, 65nm, 45nm, 32nm and 22nm technologies.

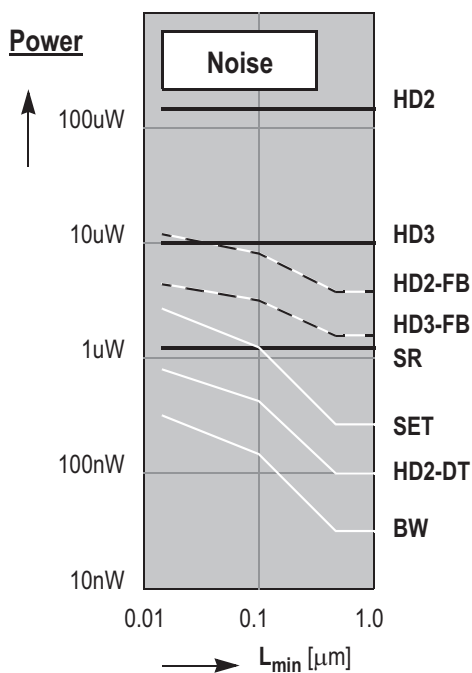


Fig. 16. The result of the less aggressive supply voltage scaling predicted by the ITRS 2003 on Noise dominated designs. All the curves (except for the SR, HD2 and HD3) bend down to a reduced power dissipation.

matching keeps improving with thinner oxides. This theory is based on the notion that threshold voltage mismatch between matched pairs of transistors is predominantly based on fluctuations of dopants in the channel. As the charge fluctuations are calculated back to a mismatch voltage the gate, by dividing through the oxide-capacitance, thinner gate-oxides reduce the input-referred V_{th} -mismatch. History (Fig. 2) shows that, until now, that has indeed been the case and predictions of the Mixed-Signal Design Roadmap of the ITRS2001 [46] also align with equation (8).

More recent work however, showed that fluctuations of the poly-silicon gate depletion charge [55] and fluctuations of fractions of the source and drain implants penetrating through the gate material [45] start contributing considerably to V_{th} -mismatch. With the supply voltage shrinking and the number of integrated transistors on a typical large digital system getting in to the hundreds of millions (requiring 6σ designs), threshold-voltage mismatch starts to effect the noise-margins in digital design [44]. This shifts the importance of mismatch from a niche high precision analog corner to the mainstream of digital design and as such will get much more attention. New solutions will be invented to combat these issues such that digital design will stay on Moore's track and analog design will only benefit from that.

4.2 Moving from Strong- to Weak-Inversion increases Distortion

Lowering of the supply voltage will not only lower the available headroom for signal-swing, but also reduced the bias-voltages (V_{gt}) thereby pushing the devices from strong-inversion in to weak-inversion. While weak-inversion operation has great benefits and entire industries are built on it, the inherent distortion of the voltage to current transfer is higher in weak-inversion than in strong-inversion. This is due to the more non-linear nature of the exponential relation in weak-inversion as compared to the square-law behavior in strong-inversion. However, as indicated by (35), output referred distortion benefits from a large gain A from input to output. A large gain requires a large g_m and the ratio between g_m and I_d is largest in weak-inversion. So, where weak-inversion exhibits the larger intrinsic distortion, it does benefit from the highest g_m/I_d ratio. It can be shown that the increase in distortion in an openloop situation for a given amount of current, by going from strong to weak-inversion, does not exceed 2x.

4.3 Intrinsic Gain ($g_m r_{out}$) reduces

The intrinsic MOS-transistor gain, $g_m r_{out}$, has been decreasing as technology-scaling allows for smaller channel-length. Even the mechanism that predominantly determines the output-impedance has changed. Where it used to be channel-length shortening, it is now Drain-Induced Barrier-Lowering (DIBL) that governs the output impedance. The predictions of the Mixed-Signal Design Roadmap of the ITRS2001 [46] on this is that at minimum channel-length a minimum value of 20 would be maintained over all technologies, at least until the 22nm node is reached in 2016.

4.4 Voltage Efficiency

The calculations performed in paragraph 2 and especially the preliminary conclusions of paragraph 2.6 are based on the assumption of a constant voltage efficiency as defined by (2). The constant reduction of the supply voltage, the less than proportional reduction in threshold voltage and the 80 [mV] lower limit of V_{gt} (Fig. 6) may endanger that assumption. Several papers [26-40, 52] have been published addressing this problem. Two of the most severe problems are Switches and OpAmps.

Switches at a low supply voltage may pose a significant problem. The use of transmission-type switches (NMOS and PMOS in parallel) may not be an option as the on-resistance may vary too much over the range from V_{ss} to V_{dd} . There may be even situation where the sum of the NMOS and PMOS V_{th} 's are larger than the supply voltage V_{dd} and a gap appears somewhere in the middle of the range where there is no conduction. Use of an NMOS-switch only can sometimes be a solution, but does shift the usable signal range close to V_{ss} and often limits the signal swing too much.

Many papers and several techniques have been published to combat this situation. Use of low- V_{th} devices may be a solution [26, 27], but at higher cost. Several different clock-boosting techniques [28-31] have been proposed, though the necessary circuitry is fairly complicated and use of those techniques is usually limited to just a few critical switches. Reliability is also a problem, since it is hard not to stress the devices.

The Switched OpAmp technique [32-36] is a true circuit solution that does not stress the process, but due to the necessity of powering down amplifier stages, the maximum clock-rate is reduced to allow for sufficient recovery-time.

OpAmps at a very low supply voltages have problems with the available signal swing at the output as well as a very limited common-mode range at the input. The limited input common-mode range often does not allow non-inverting amplifier architectures. Even for inverting architectures CM level-shift techniques may be necessary [34, 41]. The limited supply voltage may not allow for cascoding and certainly prohibits the use of telescopic architectures. This pushes the designs in the direction of two-stage designs, in order to obtain sufficient gain. The first stage only needs to handle a very limited signal swing at its output and can be cascoded, whereas the second stage does not require a large gain and can be implemented as a regular common-source amplifier. However, this does reduce the maximum achievable unity-gain frequency.

Maybe the most positive news in terms of voltage efficiency is the ITRS 2003 [47] prediction of the supply voltage V_{dd} as shown in Fig. 14. The same figure also shows the progression of the supply voltage up until now. Apparently a shift in the predictions has taken place where the new view on supply voltage scaling is that it will be substantially less aggressive. The predictions are that even the low-power digital V_{dd} will not go below 0.5V, at least until 2018. Good circuit performance at 0.5 [V] supply, even in today's processes have already been shown [52].

A less aggressive supply voltage scaling also would have a significant effect on the power dissipation estimates from Fig. 11. While it does not have any effect on the HD2, HD3 and SR curves (they are V_{dd} independent), all the other curves (HD2-FB, HD3-FB, SET, HD2-DT and BW) will bend down to a lower power dissipation than what is predicted in Fig. 11. This is an important result, as these situations, if combined with Noise (Fig. 11b), all have a tendency of showing an increasing power dissipation for shrinking technologies. Fig. 16 shows the results of the less aggressive supply voltage scaling as predicted by the ITRS 2003 (Fig. 15), on the power dissipation in Noise dominated circuits (similar to Fig. 11b). As can be seen from the figure, this has a significant positive effect on the prediction of power dissipation and almost brings the increase of power dissipation to a halt.

4.5 Gate-Leakage

Gate-leakage is a phenomenon relatively new to CMOS design. It is caused by tunneling directly through the thin oxide of the MOS device and shows an exponential dependence on the voltage across the oxide [53]. In [22] it is shown that gate-leakage could effect circuit design in a number of aspects. The

intrinsic current gain ($g_m r_{out}$) is rapidly deteriorating as a function of both technology as well as channel-length. The same thing is true for device matching. The dependence on channel-length is less of a problem as the channel-length itself is under the control of the designer, but the degradation as a function of technology is more serious and could become a significant problem at 45nm and beyond [22]. Gate-leakage also affects the maximum hold-time of a capacitor in a feedback-loop around an OpAmp and as such puts a lower limit on the clock-frequency of Switched-Capacitor and Track & Hold circuits. A positive aspects of this phenomenon however, is that the problem becomes more prone when the transistor size becomes large relative to the capacitor size, which is when the circuit is pushed to its high-frequency limit. At that moment however, a lower limit on the clock-frequency may be less of a concern.

5. Figure of Merit

To be able to compare designs that are comparable in their goals but slightly different in various performance parameters, Figures-of-Merit (FOM) are often introduced. A very common FOM for ADC's is:

$$FOM_1 = \frac{P}{2^{ENOB} ResBW}, \quad (53)$$

which is often expressed in pJ per conversion. Note that strictly speaking this is actually a 'Figure of de-Merits' a lower result is better. From the results of Table 2 however, we see that all expressions for Power Dissipation have the term $F_{sig} DR^2$ in common. Equating $DR = 2^{ENOB}$ and $F_{sig} = ResBW$ leads to the conclusion that:

$$FOM_2 = \frac{P}{2^{2ENOB} ResBW} \quad (54)$$

would make a better FOM. For matching dominated designs normalizing on T_{ox} or L_{min} is even more appropriate:

$$FOM_3 = \frac{P}{2^{2ENOB} L_{min} ResBW}. \quad (55)$$

Both (54) and (55) do give a different perspective on what is the 'best' design.

6. Conclusions

A general approach to power dissipation estimation has been proposed. It has been shown that in general the capacitor size is set by either matching or noise requirements together with the dynamic range and the maximum available signal swing, which in turn is mainly determined by the supply voltage. The current necessary to drive this capacitance is determined by dynamic performance parameters like slew-rate, settling or distortion. The final power dissipation is found by selecting the appropriate combination of design parameters and expressions have been presented for the power dissipation in all these cases.

It has been shown that matching dominated designs exhibit a decreasing or equal power dissipation for shrinking technologies whereas noise dominated designs show an increasing or equal power dissipation. It has also been shown that matching dominated designs require the highest amount of power dissipation, which stresses the need for matching improvement techniques like averaging, calibration, offset cancellation, dynamic element matching or chopping techniques.

As flash-ADC's are matching based, they benefit from technology scaling. Hence they use regular thin-oxide devices and operate from the regular supply voltage. Pipeline converters however, are noise limited and are adversely affected by technology scaling. As a result they benefit from the use of thick-oxide devices and operate usually from a higher supply voltage.

Several caveats to these power dissipation predictions are mentioned, in which gate-leakage and reduced scaling of matching may be most threatening. Finally, an adapted Figure-of-Merit for ADC's is proposed, to better reflect the merits of the circuit as opposed to the merits of the technology.

Appendix A

In this appendix the 3rd-order distortion in openloop situations is calculated for a differential pair in a similar way as was done for the 2nd-order harmonic distortion in section 2.4.4.

Differential circuitry is often used to suppress influence from common-mode signals and also to cancel out any even order distortion. If a differential pair is used for that purpose, the dominant harmonic will be the 3rd-order distortion HD3. In general, the shape of the 3rd-order distortion will be of the form [51]:

$$HD_3 = \frac{1}{m} \left(\frac{V_{in}}{V_{gt}} \right)^2, \quad (A.1)$$

where $m=8$ is assumed here for simplicity. Using the same reasoning as in section 2.4.4, we may re-write this expression in terms of the voltage swing at the output (V_{sig}) in the following way:

$$HD_3 = \frac{1}{8} \left(\frac{V_{sig}}{V_{gt}} \right)^2 \frac{1}{A^2} = \frac{1}{8} \left(\frac{V_{sig}}{V_{gt}} \right)^2 \frac{F_{sig}^2}{F_u^2}. \quad (A.2)$$

Substituting (31) and (22) in (A.2) and solving for I_d yields:

$$I_d = \frac{\pi}{2\sqrt{2}} \frac{V_{sig} F_{sig} C_L}{\sqrt{HD_3}}. \quad (A.3)$$

When comparing (A.3) to equation (33) shows that it is usually a lot easier to achieve good distortion in differential circuits than in single-ended circuits. The square-root term reduced the required amount of current dramatically.

Appendix B

In this appendix the 3rd-order distortion in closed-loop situations is calculated for a differential pair in a similar way as was done for the 2nd-order harmonic distortion in section 2.4.5.

Here we start with equation (A.1) and apply feedback with a loopgain of $A\beta$ to it:

$$HD_3 = \frac{1}{8} \left(\frac{V_{in}}{V_{gt}} \right)^2 \frac{1}{A\beta}. \quad (\text{B.1})$$

Using the same reasoning as in section 2.4.5, we may re-write this expression in terms of the voltage swing at the output (V_{sig}) in the following way:

$$HD_3 = \frac{1}{8} \left(\frac{V_{sig}}{V_{gt}} \right)^2 \frac{1}{A^3 \beta}. \quad (\text{B.2})$$

Substituting A by F_u/F_{sig} yields:

$$HD_3 = \frac{1}{8} \left(\frac{V_{sig}}{V_{gt}} \right)^2 \frac{F_{sig}^3}{\beta F_u^3}. \quad (\text{B.3})$$

Substituting (31) and (22) in (B.3) and solving for I_d yields:

$$I_d = \frac{\pi V_{sig}^{\frac{2}{3}} V_{gt}^{\frac{1}{3}} F_{sig} C_L}{4 \sqrt[3]{\beta HD_3}}. \quad (\text{B.4})$$

In general higher harmonics may be calculated in a similar fashion:

$$I_d \propto \frac{\pi V_{sig}^{\frac{n-1}{n}} V_{gt}^{\frac{1}{n}} F_{sig} C_L}{2^{n-1} \sqrt[n]{\beta HD_n}}. \quad (\text{B.5})$$

References

- [1] B. Hosticka et al., "Low-Voltage CMOS Analog Circuits", IEEE Trans. on Circ. and Syst., vol. 42, no. 11, pp. 864-872, Nov. 1995.
- [2] W. Sansen, "Challenges in Analog IC Design in Submicron CMOS Technologies", Analog and Mixed IC Design, IEEE-CAS Region 8 Workshop, pp. 72-78, Sept. 1996.
- [3] P. Kinget and M. Steyaert, "Impact of transistor mismatch on the speed-accuracy-power trade-off of analog CMOS circuits", Proc. IEEE Custom Integrated Circuit Conference, CICC96, pp.333-336, 1996.
- [4] M. Steyaert et al., "Custom Analog Low Power Design: The problem of low-voltage and mismatch", Proc. IEEE Custom Int. Circ. Conf., CICC97, pp.285-292, 1997.
- [5] V. Prodanov and M. Green, "Design Techniques and Paradigms Toward Design of Low-Voltage CMOS Analog Circuits", Proc. 1997 IEEE International Symposium on Circuits and Systems, pp. 129-132, June 1997.
- [6] W. Sansen et al., "Towards Sub 1V Analog Integrated Circuits in Submicron Standard CMOS Technologies", IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers, pp. 186-187, Feb. 1998.
- [7] Q. Huang et al., "The Impact of Scaling Down to Deep Submicron on CMOS RF Circuits", IEEE J. Solid-State Circuits, vol. 33, no. 7, pp. 1023-1036, July 1998.
- [8] W. Sansen, "Mixed Analog-Digital Design Challenges", IEEE Colloq. System on a Chip, pp. 1/1 - 1/6, Sept. 1998.
- [9] R. Castello et al. "High-Frequency Analog Filters in Deep-Submicron CMOS Technologies", IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers, pp.74-75, Feb. 1999.
- [10] K. Bult, "Analog Broadband Communication Circuits in Deep Sub-Micron CMOS", IEEE Int. Solid-State Circ. Conf. Dig. Tech. Papers, pp.76-77, Feb. 1999.
- [11] J. Fattaruso, "Low-Voltage Analog CMOS Circuit Techniques", Proc. Int. Symp. on VLSI Tech., Syst. and Appl., pp. 286-289, 1999.
- [12] D. Foty, "Taking a Deep Look at Analog CMOS", IEEE Circuits & Devices, pp. 23-28, March 1999.
- [13] D. Buss, "Device Issues in the Integration of Analog/RF Functions in Deep Submicron Digital CMOS", IEDM Techn. Dig., pp. 423-426, 1999.
- [14] A.J. Annema, "Analog Circuit Performance and Process Scaling", IEEE Trans. on Circ. and Syst., vol. 46, no. 6, pp. 711-725, June 1999.
- [15] K. Bult, "Analog Design in Deep Sub-Micron CMOS", Proc. ESSCIRC, pp. 11-17, Sept. 2000.
- [16] M. Steyaert et al., "Speed-Power-Accuracy Trade-off in high-speed Analog-to-Digital Converters: Now and in the future...", Proc. AACD, Tegernsee, April 2000.

- [17] J. Lin and B. Haroun, "An Embedded 0.8B/480uW 6b/22MHz Flash ADC in 0.13um Digital CMOS Process using Nonlinear Double-Interpolation Technique", IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers, pp.308-309, Feb. 2002
- [18] D. Buss, "Technology in the Internet Age", IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers, pp.118-21, Feb. 2002
- [19] K. Uyttenhove and M. Steyaert, "Speed-Power-Accuracy Trade-off in High-Speed CMOS ADC's", IEEE Trans. on Circ. and Syst. II: Anal. and Dig. Sig. Proc., pp. 280-287, April 2002.
- [20] Q. Huang, "Low Voltage and Low Power Aspects of Data Converter Design", Proc. ESSCIRC, pp. 29 - 35, Sept. 2004.
- [21] M. Vertregt and P. Scholtens, "Assessment of the merits of CMOS technology scaling for analog circuit design", Proc. ESSCIRC, pp. 57 - 63, Sept. 2004.
- [22] A. Annema et al, "Analog Circuits in Ultra-Deep-Submicron CMOS", IEEE J. of Solid-State Circ., vol 40, no. 1, pp. 132-143, Jan. 2005.
- [23] K. Lakshmikumar et al., "Characterization and Modelling of Mismatch in MOS Transistor for Precision Analog Design", IEEE J. of Solid-State Circ., vol SC-21, no. 6, pp. 1057-11066, Dec. 1986
- [24] M. Pelgrom et al., "Matching Properties of MOS Transistors", IEEE J. of Solid-State Circ., vol 24, no. 5, pp. 1433-1439, Oct. 1989.
- [25] T. Mizuno et al., "Experimental Study of Threshold Voltage Fluctuation Due to Statistical Variation of Channel Dopant Number in MOSFET's", IEEE Trans. on Elec. Dev. vol. 41, no.11, pp. 2216-2221, Nov. 1994.
- [26] T. Ohguro et al., "0.18 μm low voltage / low power RF CMOS with zero V_{th} analog MOSFET's made by undoped epitaxial channel technique", IEEE IEDM Tech. Dig., pp.837-840, 1997.
- [27] S. Bazarjani and W.M. Snelgrove, "Low Voltage SC Circuit Design with Low- V_t MOSFET's", IEEE Int. Symp. on Circ. and Syst., ISCAS95, pp. 1021-1024, 1995.
- [28] T. Cho and P. Gray, "A 10 b, 20 MSample/s, 35 mW Pipeline A/D Converter",
- [29] T. Brooks et al., "A Cascaded Sigma-Delta Pipeline A/D Converter with 1.25 MHz Signal Bandwidth and 89 dB SNR",
- [30] A. Abo and P. Gray, "A 1.5 V, 10-bit, 14.3-MS/s CMOS Pipeline Analog-to-Digital Converter", IEEE J. of Solid-State Circ., vol 34, no. 5, pp. 599-606, May 1999.
- [31] U. Moon et al., "Switched-Capacitor Circuit Techniques in Submicron Low-Voltage CMOS". IEEE Int. Conf. on VLSI and CAD, ICVC99, pp. 349-358, 1999.
- [32] J. Crols and M. Steyaert, "Switched-OpAmp: An Approach to Realize Full CMOS Switched-Capacitor Circuits at Very Low Power Supply Voltages", IEEE J. of Solid-State Circ., vol 29, no. 12, pp. 936-942, Aug. 1994.

- [33] V. Peluso et al., "A 1.5-V-100- μ W SD Modulator with 12-b Dynamic Range Using the Switched-OpAmp Technique", *IEEE J. of Solid-State Circ.*, vol 32, no. 7, pp. 943-952, July 1997.
- [34] A. Bashiroto and R. Castello, "A 1-V 1.8-MHz CMOS Switched-OpAmp SC Filter with Rail-to-Rail Output Swing", *IEEE J. of Solid-State Circ.*, vol 32, no. 12, pp. 1979-1986, Dec. 1997.
- [35] V. Peluso et al., "A 900-mV Low-Power SD A/D Converter with 77-dB Dynamic Range", *IEEE J. of Solid-State Circ.*, vol 33, no. 12, pp. 1887-1897, Dec. 1998.
- [36] L. Dai and R. Harjani, "CMOS Switched-Op-Amp-Based Sample-and-Hold Circuit", *IEEE J. of Solid-State Circ.*, vol 35, no. 1, pp. 109-113, Jan. 2000.
- [37] R. Hogervorst et al., "A Compact Power-Efficient 3V CMOS Rail-to-Rail Input/Output Operational Amplifier for VLSI Cell Libraries", *IEEE J. of Solid-State Circ.*, vol 29, no. 12, pp. 1505-1513, Dec. 1994.
- [38] W. Redman-White, "A High Bandwidth Constant g_m and Slew-Rate Rail-to-Rail CMOS Input Circuit and its Application to Analog Cells for Low Voltage VLSI Systems", *IEEE J. of Solid-State Circ.*, vol 32, no. 5, pp. 701-711, May 1997.
- [39] B. Blalock et al., "Designing 1-V Op Amps Using Standard Digital CMOS Technology", *IEEE Trans. on Circ. and Syst.*, vol. 45, no. 7, pp. 769-780, July 1998.
- [40] T. Duisters and E.C. Dijkmans, "A -90-dB THD Rail-to-Rail Input Opamp Using a New Local Charge Pump in CMOS", *IEEE J. of Solid-State Circ.*, vol 33, no. 7, pp. 947-955, July 1998.
- [41] S. Karthikeyan et al., "Low-Voltage Analog Circuit Design Based on Biased Inverting Opamp Configuration", *IEEE Trans. on Circ. and Syst.*, vol. 47, no. 3, pp. 176-184, March 2000.
- [42] K. Kattmann, J. Barrow, "A technique for reducing differential non-linearity errors in flash A/D converters", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp. 170 - 171, Feb. 1991.
- [43] K. Bult and A. Buchwald, "An embedded 240-mW 10-b 50-MS/s CMOS ADC in 1-mm²", *IEEE J. of Solid-State Circ.*, vol 32, no. 12, pp. 1887-1895, Dec. 1990.
- [44] H. Tuinhout, "Impact of parametric mismatch and fluctuations on performance and yield of deep-submicron CMOS technologies", in *Proc. ESSDERC*, pp. 95-101, 2002.
- [45] J. Dubois et al, "Impact of source/drain implants on threshold voltage matching in deep sub-micron CMOS technologies", in *Proc. ESSDERC*, pp. 115-118, 2002.
- [46] R. Brederlow et al., "A mixed-Signal Design Roadmap", *ITRS 2001 in IEEE Design & Test of Computers*, pp. 34-46, 2001.
- [47] ITRS 2003, <http://public.itrs.net/Files/2003ITRS/Home2003.htm>
- [48] P. Gray and R. Meyer, "Analysis and Design of Analog Integrated Circuits", John Wiley & Sons, Inc., New York, 1992.

- [49] K. Laker and W. Sansen, "Design of Analog Integrated Circuits and Systems", McGraw-Hill, Inc., New York, 1994.
- [50] Y. Tsividis, "Operation and Modelling of The MOS Transistor", McGraw-Hill Book Company, New York, 1987.
- [51] K. Bult, "CMOS Analog Square-Law Circuits", Ph.D. Thesis, Twente University, 1988.
- [52] S. Chatterjee et al., "A 0.5V Filter with PLL-Based Tuning in 0.18 μ m CMOS", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.506-507, Feb. 2005
- [53] R. van Langevelde et al., "Gate current: Modelling, DL extraction and impact on RF performance", in *IEDM Tech. Dig.*, pp. 289-292, 2001.
- [54] J. Mulder et al., "A 21-mW 8-b 125-MSample/s ADC in 0.09-mm² 0.13- μ m CMOS", *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2116 - 2125, Dec. 2004.
- [55] H.P. Tuinhout et al., "Effects of Gate Depletion and Boron Penetration on Matching of Deep Submicron CMOS Transistors", *IEDM 97, Tech. Dig.* pp. 631-635, 1997.
- [56] K. Dyer et al., "Analog background calibration of a 10b 40Msample/s parallel pipelined ADC", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp. 142 - 143, Feb. 1998.
- [57] E. Siragusa and I.Galton; "A digitally enhanced 1.8-V 15-bit 40-MSample/s CMOS pipelined ADC", *IEEE J. of Solid-State Circ.*, pp. 2126 - 2138, Dec. 2004.
- [58] R. v.d.Plassche, "Integrated Analog-to-Digital and Digital-to-Analog Converters", *Kluwer Academic Publishers*, Dordrecht, The Netherlands, 1994.
- [59] K.McCall et al. "A 6-bit 125 MHz CMOS A/D Converter", *Proc. IEEE Custom Int. Circ. Conf.*, CICC, 1992.
- [60] M.Flynn and D.Allstot, "CMOS Folding ADCs with Current-Mode Interpolation", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.274-275, Feb. 1995.
- [61] F.Paillardet and P.Robert, "A 3.3 V 6 bits 60 MHz CMOS Dual ADC", *IEEE Trans. on Cons. Elec.*, vol. 41, no. 3, pp. 880-883, Aug. 1995.
- [62] J.Spalding and D.Dalton,"A 200MSample/s 6b Flash ADC in 0.6 μ m CMOS", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.320-321, Feb. 1996.
- [63] R.Roovers and M.Steyaert, "A 175 Ms/s, 6b, 160 mW, 3.3 V CMOS A/D Converter", *IEEE J. of Solid-State Circ.*, vol 31, no. 7, pp. 938-944, July 1996.
- [64] S.Tsukamoto et al., "A CMOS 6-b, 200 MSample/s, 3 V-Supply A/D Converter for a PRML Read Channel LSI", *IEEE J. of Solid-State Circ.*, vol 31, no. 11, pp. 1831-1836, Nov. 1996.
- [65] D.Dalton et al., "A 200-MSPS 6-Bit Flash ADC in 0.6- μ m CMOS", *IEEE Trans. on Circ. and Syst.*, vol. 45, no. 11, pp. 1433-1444, Nov. 1998.
- [66] M.Flynn and B.Sheahan, "A 400-MSample/s 6-b CMOS Folding and Interpolating ADC", *IEEE J. of Solid-State Circ.*, vol 33, no. 12, pp. 1932-1938, Dec. 1998.

- [67] S.Tsukamoto et al., "A CMOS 6-b, 400-MSample/s ADC with Error Correction", *IEEE J. of Solid-State Circ.*, vol 33, no. 12, pp. 1939-1947, Dec. 1998.
- [68] Y.Tamba and K.Yamakido, "A CMOS 6b 500MSample/s ADC for a Hard Disk Drive Read Channel", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.324-325, Feb. 1999.
- [69] K.Yoon et al., "A 6b 500MSample/s CMOS Flash ADC with a Background Interpolated Auto-Zero Technique", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.326-327, Feb. 1999.
- [70] I.Mehr and D.Dalton, "A 500-MSample/s, 6-Bit Nyquist-Rate ADC for Disk-Drive Read-Channel Applications", *IEEE J. of Solid-State Circ.*, vol 34, no. 7, pp. 912-920, July 1999.
- [71] K.Nagaraj et al., "Efficient 6-Bit A/D Converter Using a 1-Bit Folding Front End", *IEEE J. of Solid-State Circ.*, vol 34, no. 8, pp. 1056-1062, Aug. 1999.
- [72] K.Nagaraj et al., "A 700MSample/s 6b Read Channel A/D Converter with 7b Servo Mode", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.426-427, Feb. 2000.
- [73] K.Sushihara et al., "A 6b 800MSample/s CMOS A/D Converter", *IEEE Int. Solid-State Circ. Conf., Dig. Tech. Papers*, pp.428-429, Feb. 2000.

Low-Voltage, Low-Power Basic Circuits

Andrea Baschiroto¹, Stefano D'Amico¹, Piero Malcovati²

¹Department of Innovation Engineering - University of Lecce – Lecce – Italy

²Department of Electrical Engineering – University of Pavia – Pavia – Italy

Abstract

In this presentation several solutions for operating analog circuits at low power and/or low voltage will be discussed. Different approaches will be presented at transistor level, at circuit level and at system level.

1. Introduction

Technological evolution and market requirements are pushing towards low-voltage and low-power integrated circuits. This need comes from technology shrink (which lowers the breakdown voltage of the devices) and from the increasing demand for portable (battery operated) systems, as illustrated by the power supply voltage forecast of the ITRS Roadmap reported in Fig. 1.

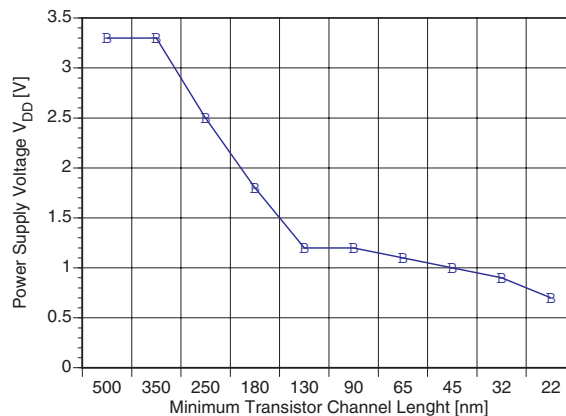


Fig. 1 – Evolution of the power supply voltage as a function of the technology node according to the ITRS Roadmap

Before entering into the discussion on low-voltage and low-power basic circuits it is worth to clarify a couple of fundamental questions:

- does low-voltage imply low-power for analog circuits?
- what does low-voltage really mean?

For the digital systems the power consumption is proportional to $C \cdot V_{DD}^2$. Therefore, the reduction of the supply voltage together with the reduction of the parasitic capacitances C in scaled technologies definitely corresponds to a reduction of the power consumption. On the other hand, for analog systems the situation is often the opposite, since the reduction of the supply voltage, if no specific countermeasures are taken, results in a reduction of the signal amplitude that can be processed, and, as a consequence, of the dynamic range that can be obtained with the same power consumption. In fact, for a given supply V_{DD} , the maximum voltage swing possible in an analog system is about

$$\text{Eq.(1)} \quad SW = [V_{DD} - 2V_{OV}],$$

where V_{OV} denotes the upper and lower saturation voltages of the output stage of an analog circuit (typically V_{OV} is the overdrive voltage of a MOS transistor). The power consumption (P) can be obtained multiplying V_{DD} by the total current I :

$$\text{Eq.(2)} \quad P = V_{DD} I.$$

On the other hand, if the noise N^2 , as in most analog systems, is limited by the thermal component, it is inversely proportional to a fraction of the total current I :

$$\text{Eq.(3)} \quad N^2 \propto \frac{1}{\alpha \cdot I} = \frac{V_{DD}}{\alpha \cdot P}.$$

The analog system dynamic range (DR) can then be written as

$$\text{Eq.(4)} \quad DR = \frac{SW^2}{N^2} \propto \frac{(V_{DD} - 2V_{OV})^2}{\frac{1}{\alpha \cdot I}} = \alpha \frac{P}{V_{DD}} (V_{DD} - 2V_{OV})^2.$$

A given DR therefore requires a power consumption proportional to

$$\text{Eq.(5)} \quad P \propto \frac{DR \cdot V_{DD}}{\alpha (V_{DD} - 2V_{OV})^2}$$

or a current consumption given by

$$\text{Eq.(6)} \quad I \propto \frac{DR}{\alpha (V_{DD} - 2V_{OV})^2}$$

From Eq.(5) and Eq.(6), it appears that both the power and the current consumption of analog circuits increase when V_{DD} decreases while maintaining constant the dynamic range, as qualitatively shown in Fig. 2.

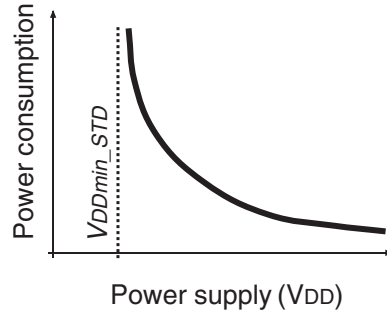


Fig. 2 – Power consumption vs. power supply for a given dynamic range

This can be also seen from Table I, where the performance of some significant analog systems is compared. They are all $\Sigma\Delta$ modulators, but their performance is limited by thermal noise. They are compared using two possible figures of merit (F_P and F_I)

$$\text{Eq.(7)} \quad F_P = \frac{4 \cdot k \cdot T \cdot DR^2 \cdot BW}{P}$$

and

$$\text{Eq.(8)} \quad F_I = \frac{4 \cdot k \cdot T \cdot DR^2 \cdot BW}{I} = \frac{4 \cdot k \cdot T \cdot DR^2 \cdot BW \cdot V_{DD}}{P}$$

The figure of merit F_P takes into account the power dissipation, while F_I the current consumption. Therefore, F_I does not consider the obvious power consumption reduction due to the scaling of V_{DD} , i.e. it considers only the increase in power consumption required to maintain the dynamic range.

TABLE I – PERFORMANCE COMPARISON

<i>Reference</i>	<i>Year</i>	V_{DD} [V]	<i>DR</i> [dB]	<i>BW</i> [kHz]	<i>P</i> [mW]	F_P [$\times 10^6$]	F_I [$\times 10^6$]
Dessouky [1]	2001	1	88	25	1	261	261
Peluso [2]	1998	0.9	77	16	0.04	332	299
Libin [3]	2004	1	88	20	0.14	1493	1493
Gaggl [4]	2005	1.5	82	1100	15	193	289
Rabii [5]	1997	1.8	99	25	2.5	1316	2369
Williams [6]	1994	5	104	50	47	443	2214
Nys [7]	1997	5	112	0.4	2.175	483	2415
Wang [8]	2003	5	113	20	115	575	2875
YuQing [9]	2003	5	114	20	34	2448	12240

It appears that the value of F_I for systems operating at 5 V (even if realized with technologies not at the state-of-the-art) is better than that for more recent implementations at low-voltage. The same applies for F_P with some exceptions (such as [5] and [3]) where specific low-voltage techniques have been used to limit the power consumption penalty. In spite of this general power consumption increase, low-voltage analog circuits are needed for mixed-signal systems-on-a-chip, where the supply voltage is determined by the digital section requirements, and therefore proper solutions for the power-efficient implementation of analog building blocks at low voltage must be developed.

From Fig. 2 it also appears that there is a certain value of the supply voltage (V_{DDmin_STD}) below which an analog circuit designed with standard techniques cannot any longer operate. To allow an analog circuit to operate with supply voltages below V_{DDmin_STD} again specific design techniques are required. This consideration is useful to answer the second fundamental question. We can define “low-voltage system” a system that operates with a power supply voltage lower than V_{DDmin_STD} (i.e. a system for which specific design techniques for low voltage operation are required).

The value of V_{DDmin_STD} depends on the kind of circuit considered. Therefore, it is not possible to give a universal definition of “low voltage”, but this definition depends on the function and the topology of the considered circuit or system. For example in a bandgap reference voltage generator it is quite easy to identify V_{DDmin_STD} with the value of the bandgap voltage in silicon (1.2 V) plus V_{OV} . When $V_{DD} < 1.2 \text{ V} + V_{OV}$, indeed, the traditional reference voltage $V_{BG} = 1.2 \text{ V}$ obviously cannot be any longer generated, thus requiring specific circuit techniques to produce a lower output voltage with the same characteristics of stability and accuracy. In sampled-data systems it can be demonstrated that the value of V_{DDmin_STD} is limited by the proper operation of the switches. This is because the availability of good switches allows the operational amplifiers to be properly operated and biased.

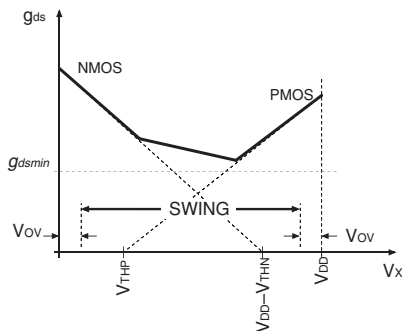


Fig. 3 – Switch conductance with large V_{DD}

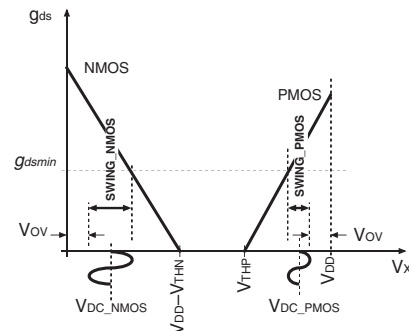


Fig. 4 – Switch conductance with low V_{DD}

For a given technology, the switch conductance is passing from the situation of Fig. 3 with a large value of V_{DD} to the situation of Fig. 4 with a low value of V_{DD} .

A critical voltage region centered around $V_{DD}/2$ appears where no switch is conducting and then rail-to-rail output swing cannot be achieved. The value of V_{DDmin_STD} can then be extracted analyzing the operation of the transmission gate, as in Fig. 5.

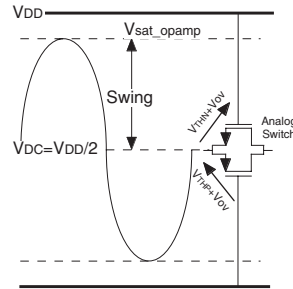


Fig. 5 – Transmission gate with minimum supply voltage

The most critical situation is when the signal to be sampled is equal to $V_{DD}/2$. In this case, the use of transmission gates is efficient for:

$$\text{Eq.(9)} \quad V_{DDmin_STD} > V_{THN} + V_{THP} + 2 \cdot V_{OV}$$

The use of transmission gates allows us to exploit standard techniques (for both operational amplifiers and switches). Therefore novel solution for the design of low-voltage sampled data systems are needed either at circuit or at system level only when the power supply voltage is:

$$\text{Eq.(10)} \quad V_{DD} > V_{THN} + V_{THP} + 2 \cdot V_{OV}$$

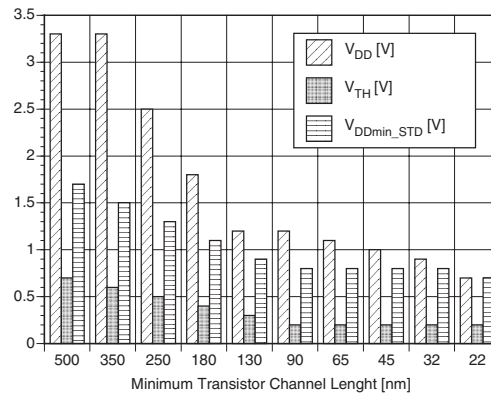


Fig. 6 – The foreseen maximum supply voltage for the next years

Notice that V_{DDmin_STD} is technology dependent and for the forecast technologies in the future years (see Fig. 6), this condition is going to appear for technologies be-

yond 30 nm (even beyond 65 nm when considering worst cases and design margins).

For other kind of circuits, such as continuous-time filters or operational amplifiers the value of V_{DDmin_STD} is again different and can be typically determined with simple considerations based on the circuit topology.

In the following sections we will consider the most important techniques for implementing circuits (Section 2) and systems (Section 3) operated with power supply voltages approaching or lower than V_{DDmin_STD} and for minimizing the power consumption penalty associated with low-voltage operation.

2. Low-Voltage Circuit Techniques

In this section we will review different circuit techniques for achieving low-voltage operation with high-efficiency (i.e. minimizing the power consumption penalty). In particular, we will consider low-voltage design issues in the MOS transistors and in some of the most important analog building blocks, namely current mirrors, bandgap references, operational amplifiers and common-mode feedback circuits.

2.1 MOS Transistors

The minimum voltage required to operate a MOS transistor is typically determined by two parameters, namely the threshold voltage V_{TH} and the overdrive voltage $V_{OV} = V_{GS} - V_{TH}$, as for example in Eq.(9). Therefore, in order to achieve low-voltage operation the designer has to minimize these two parameters. The threshold voltage can be lowered by modifying the process (technologies with low-threshold transistors are sometimes available at the expense of higher production cost) or by using special circuit techniques. However, in most cases, V_{TH} is a fixed parameter, on which the designer has no control. On the other hand, the overdrive voltage is under the designer control. By minimizing V_{OV} and/or the bias current, however, the MOS transistors end up operating in the weak inversion region, which is therefore the most common operating condition of the transistors in low-voltage and low-power circuits [10]. The drain current of a MOS transistor in the weak inversion region is given by:

$$\text{Eq.(11)} \quad I_D = I_S \cdot e^{\frac{V_G - V_{TH}}{nV_T}} \cdot \left(e^{\frac{V_S}{V_T}} - e^{\frac{V_D}{V_T}} \right),$$

where $V_T = k \cdot T / q$ denotes the thermal voltage and n the slope parameter. Transistors operated in this region feature advantages and disadvantages, summarized in Table II.

TABLE II – ADVANTAGES AND DISADVANTAGES OF MOS TRANSISTORS OPERATED IN THE WEAK INVERSION REGION

<i>Advantages</i>	<i>Disadvantages</i>
Minimum V_{OV}	Maximum drain current mismatch
Minimum gate capacitance	Maximum output noise current for a given I_D
Maximum ratio g_m/I_D and voltage gain	Low speed, $f_T \cong \frac{\mu \cdot V_T}{2 \cdot \pi \cdot L^2}$

In particular, it is worth to consider the output current mismatch, which has impact on the achievable performance of several circuits. The output current mismatch in a MOS transistor is given by

$$\text{Eq.(12)} \quad \sigma\left(\frac{\delta I_D}{I_D}\right) = \sqrt{\sigma_\beta^2 + \left(\frac{g_m}{I_D} \cdot \sigma_T\right)^2},$$

where σ_B and σ_T are the mismatches of $\beta = \mu \cdot C_{ox}$ and V_{TH} , respectively. In the weak inversion region g_m/I_D is maximum and hence the second term of Eq.(12) increases, leading to a large mismatch. With typical values of σ_B and σ_T the output current mismatch of MOS transistors in the weak inversion region can be as large as 10%. Nonetheless, in bandpass applications, the offset may be not critical and then MOS device may be efficiently operated in weak inversion [11].

An interesting idea for reducing the value of the MOS transistor threshold voltage without modifying the process is to bias the device with a negative bulk-source voltage ($V_{BS} < 0$) [12]. Indeed, the threshold voltage of a MOS transistor is given by

$$\text{Eq.(13)} \quad V_{TH} = V_{TH0} + \gamma \left(\sqrt{|2 \cdot \phi_F - V_{BS}|} - \sqrt{|2 \cdot \phi_F|} \right),$$

where V_{TH0} is zero bias threshold voltage, γ the bulk effect factor and ϕ_F the Fermi potential. The bulk bias V_{BS} is normally positive, which leads to an increase of the threshold voltage with respect to V_{TH0} . However, by biasing the transistor with $V_{BS} < 0$ V, we can actually decrease the threshold voltage (see Eq.(13)). To reduce the threshold voltage as much as possible, the device has to be bulk biased as high as possible. However, this will forward bias the bulk-source diode, which is also the base-emitter diode of the associated parasitic bipolar transistor, thereby turning on this BJT. The absolute value of V_{BS} is limited by how much current this BJT can tolerate. Moreover, the parasitic BJT introduces additional noise in the MOS transistor and might lead to latch-up.

2.2 Current Mirrors

Current mirrors are among the most important building blocks for realizing analog integrated circuits. The conventional current mirror, shown in Fig. 7.a, requires at

least a power supply voltage $V_{DDmin} = V_{TH} + 2 \cdot V_{OV}$ to properly operate (V_{OV} is required by the input current source) and therefore it is not a particularly critical block for low-voltage operation. The output impedance of this current mirror, however, is relatively low and is worsening in scaled down technologies, making the use of this circuit unpractical in many cases. The traditional way of increasing the output impedance of a current mirror is the use of cascode structures (Fig. 7.b). A cascode current mirror, however, requires a minimum supply voltage $V_{DDmin} = 2 \cdot V_{TH} + 3 \cdot V_{OV}$, which enables the use of this structure only for $V_{DD} > 1.8$ V.

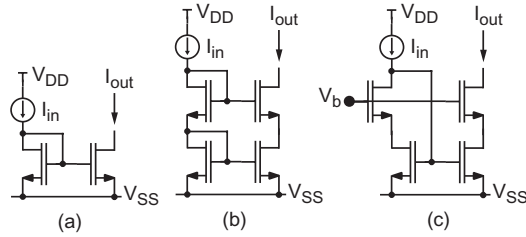


Fig. 7 – Conventional current mirror (a), conventional cascode current mirror (b), high swing cascode current mirror (c)

The most common solution for achieving a sufficiently large output impedance in a current mirror without increasing V_{DDmin} is the so called “high-swing” current mirror, whose schematic is shown in Fig. 7.c. This circuit requires a minimum power supply voltage $V_{DDmin} = V_{TH} + 2 \cdot V_{OV}$ as the conventional current mirror, but achieves an output impedance of the same order of magnitude as the cascode current mirror, as illustrated in Fig. 8.

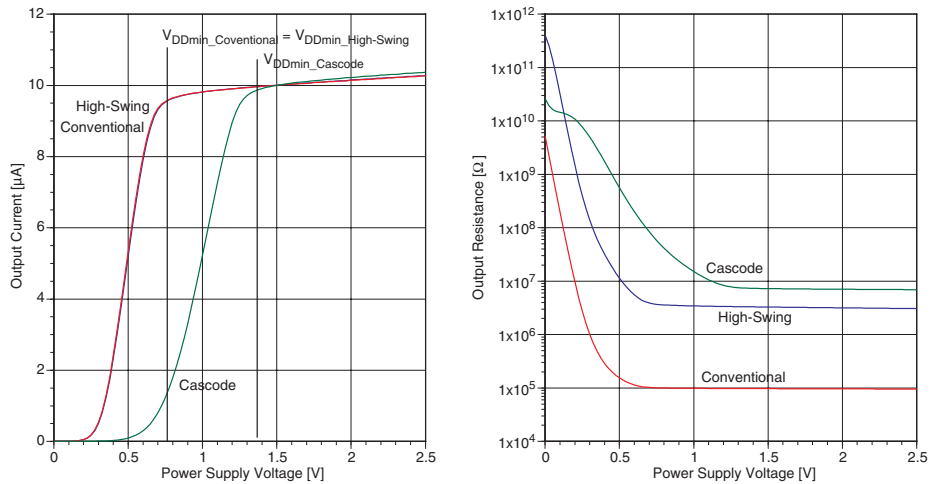


Fig. 8 – Output current and output resistance as a function of the power supply voltage for the conventional, cascode and high-swing current mirrors

2.3 Bandgap Reference Voltage Generators

Conventional bandgap reference structures produce a reference voltage of about 1.2 V with minimum sensitivity to temperature variations. Of course, as already mentioned, when the supply voltage goes down below $1.2 \text{ V} + V_{OV}$ (V_{DDmin_STD}), it is no longer possible to use the conventional structures [13].

Two components build up the output voltage of a bandgap reference circuit. One is the voltage across a directly biased diode (V_{BE}) and the other is a term proportional to the absolute temperature (PTAT). The negative temperature coefficient of the former term compensates for the positive temperature coefficient of the latter. If $V_T = k \cdot T/q$ is used to obtain a PTAT voltage, it is well known that (at ambient temperature) it has to be multiplied by approximately 22 to compensate for the temperature dependence of the diode voltage. If this condition is satisfied, the generated bandgap voltage becomes approximately $V_{BG} = 1.2 \text{ V}$. Using a supply voltage (V_{DD}) lower than $V_{DDmin_STD} = 1.2 \text{ V} + V_{OV}$, a fraction of V_{BG} with appropriate temperature features has to be generated. Since the bandgap voltage is given by

$$\text{Eq.(14)} \quad V_{BG} = V_{BE} + n \frac{k \cdot T}{q},$$

a fraction of the conventional bandgap voltage is achieved by scaling both terms of Eq.(14), using currents terms proportional to V_{BE} and to V_T , respectively. These currents are suitably added and transformed into a voltage with a resistor. The temperature dependence of the resistors used is compensated by fabricating them with the same kind of material. Fig. 9 shows the schematic diagram of a circuit, which implements the described operation [14].

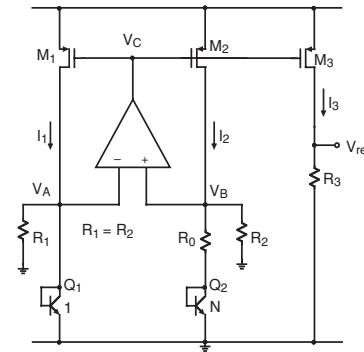


Fig. 9 – Schematic of the low-voltage bandgap circuit.

Two diode connected bipolar transistors with emitter area ratio N drain the same current, leading to a ΔV_{BE} equal to $V_T \ln(N)$. Therefore, the current in R_0 is PTAT. The operational amplifier forces the two voltages V_A and V_B to be equal, thus producing a current in the nominally equal resistors R_1 and R_2 proportional to V_{BE} . As a result, the current in M_1 , M_2 and M_3 ($I_1 = I_2 = I_3$) is given by:

$$\text{Eq.(15)} \quad I_1 = \frac{V_T \cdot \ln(N)}{R_0} + \frac{V_{BE}}{R_1}.$$

The output voltage is then given by

$$\text{Eq.(16)} \quad V_{out} = I_1 \cdot R_3 = V_T \left[\frac{R_3 \cdot \ln(N)}{R_0} \right] + V_{BE} \frac{R_3}{R_1} = \frac{R_3}{R_1} \left[\frac{R_1 \cdot \ln(N)}{R_0} V_T + V_{BE} \right]$$

The compensation of the temperature coefficients of V_T and V_{BE} is ensured by choosing values of N and of the R_1/R_0 ratio which satisfy

$$\text{Eq.(17)} \quad \frac{R_1 \cdot \ln(N)}{R_0} = 22.$$

Moreover, since transistors M_1 , M_2 and M_3 maintain almost the same drain-source voltage V_{DS} , independently of the actual supply voltage, the power supply rejection ratio of the circuit is only determined by the operational amplifier. By inspection of the circuit, it can be observed that the minimum supply voltage is determined by $V_{BE} + V_{OV}$. However, the supply voltage used must also ensure proper operation of the operational amplifier (at least $V_{TH} + 2 \cdot V_{OV} + V_{OV}$ for a CMOS implementation) and, indeed, this is the true limit of the circuit.

2.4 Operational Amplifiers

The MOS transistor output impedance of scaled-down technologies decreases and, as a consequence, also the achievable gain-per-stage decreases. In addition, at low supply voltage, stacked configurations (for example cascode) are not possible.

In this situation a sufficiently large operational amplifier (opamp) gain can be achieved by adopting multistage structures which, however, for stability reason, tend to have a relatively small bandwidth as compared with single stage structures. On the other hand they present the advantage of allowing us, at the first order, to separately optimize the different stages. In this way it is possible to operate with the input common-mode voltage (V_{in_DC}) and the output common-mode voltage (V_{out_DC}).

Regarding the input stage, the most feasible solution in order to reduce noise coupling and offset appears to be the differential structure. The simplest differential input stage is shown in Fig. 10. In this case, PMOS input transistors with NMOS load devices have been used as an example. The minimum supply voltage for proper operation (V_{DDmin}) is given by

$$\text{Eq.(18)} \quad V_{DDmin} = \max \left\{ 3 \cdot V_{OV} + V_{outpp}, V_{in_DC} + V_{TH_P} + 2 \cdot V_{OV} \right\}$$

The first condition is forced by the three stacked devices M_1 , M_3 , M_5 that have to operate in the saturation region (i.e. $V_{DS} > V_{OV}$), assuming an output swing V_{outpp} .

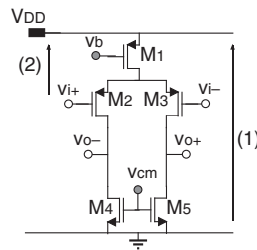


Fig. 10 – Differential input stage

The second condition is forced by the sum of the V_{GS} drop across the input device and the V_{DS} of the PMOS current source. In the case of V_{TH} larger than $V_{OV} + V_{outpp}$ and $V_{in_DC} = 0$ (which appears to be the optimal bias voltage for the input stage), the theoretical minimum supply voltage V_{DDmin} is obtained and it is given by

$$\text{Eq. (19)} \quad V_{DDmin} = V_{TH} + 2 \cdot V_{OV}.$$

Regarding the output stage, the key issue is to maximize the output swing, thus leading to $V_{out_DC} = V_{DD}/2$.

The complete scheme of a differential amplifier is shown in Fig. 11 [15]. It allows to satisfy both the above conditions.

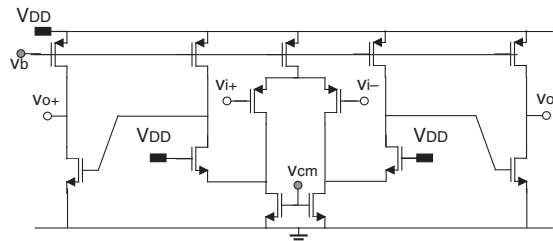


Fig. 11 – Complete two-stage amplifier

In this circuit the folding structure has been used to ensure that the V_D of M_4 , M_5 is biased slightly higher than one V_{OV} above ground. This allows us to maintain all the devices (for the possible exception of the cascode device itself) in saturation region at all times. This is because either at the source of the cascode device and at the gate of the NMOS device in the output stage the voltage swing is quite small. The minimum supply voltage V_{DDmin} for this circuit is still given by $V_{TH} + 2 \cdot V_{OV}$, assuming that V_{TH} is the largest between the NMOS and the PMOS threshold voltages. The structure of Fig. 11 corresponds to a fully differential amplifier. In the case of a single-ended structure, a current mirror must be implemented to realize the differential-to-single ended transformation. The use a diode connected MOS in the signal path, typical of classical current mirror topologies, is not possible since it would increase the minimum supply voltage. Fig. 12 shows a possible circuit ca-

pable of operating with $V_{DDmin} = V_{TH} + 2 \cdot V_{OV}$: a low-voltage current mirror is used (indicated with dashed line).

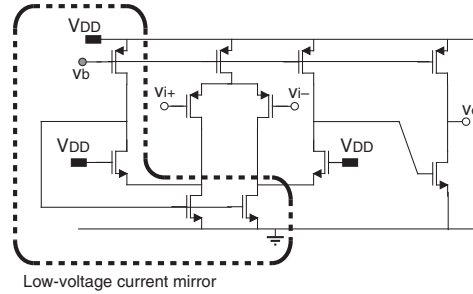


Fig. 12 – Single-ended opamp with low-voltage current mirror

In alternative to the above class-A input stage, a class-AB input stage is also available [2]. It is based on the input differential pair of Fig. 13. Applying a differential signal to this circuit, two equal output currents (I_{out}) are generated. The complete class-AB opamp scheme, which uses also the low-voltage current mirror, is shown in Fig. 14. It is able to operate with $V_{DDmin} = V_{TH} + 2 \cdot V_{OV}$. This stage can be used when a large capacitive load is present and the power consumption must be reduced. However, as it is usually the case in a class AB stage, its structure is relatively complex, thus increasing noise and offset.

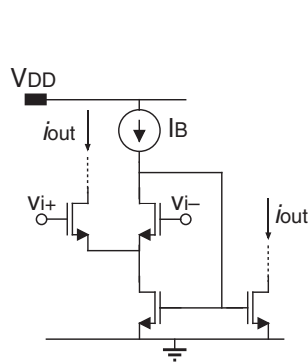


Fig. 13 – Class-AB input pair

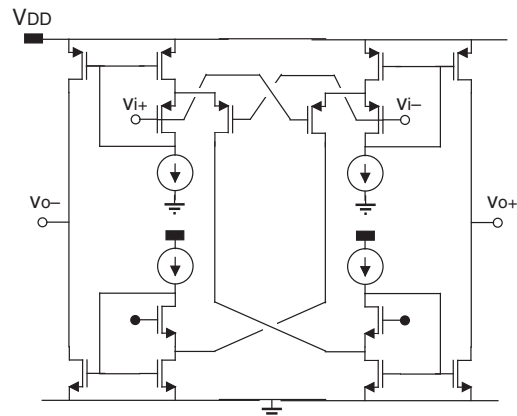


Fig. 14 – Class-AB opamp

An interesting circuit for low-voltage applications is the bulk driven opamp. In this circuit, whose simplified schematic is shown in Fig. 15 [16], the input signal is applied to the bulk of two MOS transistors. This technique allows us to achieve rail-to-rail input common-mode swing requiring a supply voltage as low as $V_{DDmin} = V_{TH} + 2 \cdot V_{OV}$. The drawback of this circuit is that the transconductance value changes dramatically (about 2 times) with the common-mode input voltage. Moreover, the equivalent input referred noise of a bulk-driven MOS amplifier is

larger than the conventional gate-driven MOS amplifier because of the small trans-conductance provided by the input transistors.

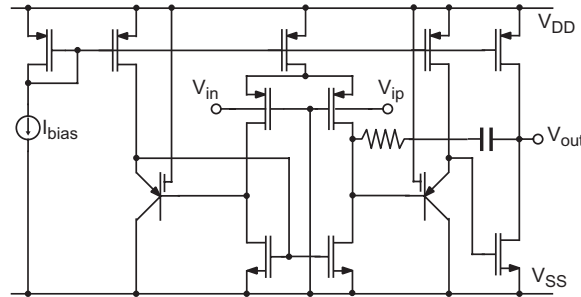


Fig. 15 – Bulk driven opamp

2.5 Common-Mode Feedback

For a fully differential opamp an additional common-mode feedback (CMFB) circuit is required to control the output common-mode voltage. This can be done using a continuous-time or a sample-data (dynamic) approach.

For a continuous-time solution, the key problem is that the inputs of the CMFB circuit must be dc-connected to the opamp output nodes, which are located around $V_{DD}/2$. For low supply voltage $V_{DD}/2$ is lower than V_{TH} . No MOS gate can therefore be directly connected to the opamp output node. Fig. 16 shows a circuit that uses a passive level shifter to circumvent this limitation. In this case a trade-off exists between the amplitude of the signal present at the CMFB circuit input and the amount of level shifting. In addition the resistive level shifter decreases the output stage gain. Finally, this scheme operates with a V_{DDmin} which is $V_{TH} + 3 \cdot V_{OV}$, i.e. larger than the value by the rest of the opamp. This means that in some cases the CMFB becomes the limiting factor for the V_{DD} minimization.

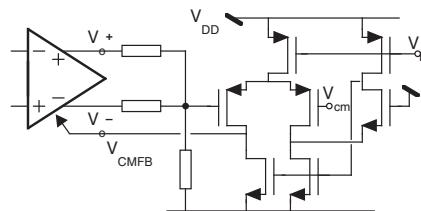


Fig. 16 – Low-voltage continuous-time CMFB circuit

On the other hand, for a dynamic CMFB circuit the key problem is to properly turn on and off the switches. This could be easily done using a voltage multiplier and for this case no further discussion is necessary. Two further solutions (active and passive) for a low-voltage dynamic CMFB circuit are shown in Fig. 17 ([17], [18], [15]).

3.1 Low-Power Solutions

The possible approaches for reducing the power consumption at system level are mainly related to operation timing in the device under development and to the accurate control of the signal swing. This is of course directly applied for sampled-data systems (mainly realized with Switched-Capacitor or Switched-Current techniques), while few examples of continuous-time systems will be given.

The main power reduction techniques based on operation timing are:

- The duty-cycle technique: this corresponds to turn-on active circuits only when they are needed, while the circuits are turned-off for the rest of the time. These circuits consume power only when they are turned-on and the overall average power consumption is reduced. This is only possible when the system allows that the output signal of the circuit to be turned-off is not needed during the off-state. This means that the signal processing is required only in certain periods. On the other hand, the circuits have to be able to quickly recover the state acquired before the turn-off, and this recovering has to be done within the available time slot.
- The time-sharing technique: once that it is possible to disable the operation of some parts of the device, these parts, instead of being turned-off (with the duty-cycle technique), may be connected to different parts of the device, avoiding the duplication of active device and, as a consequence, of their power consumption.

3.1.1 The duty-cycle technique

The duty-cycle technique corresponds to activate the consuming circuit (or a part of it) only for the necessary time for signal processing, while for the rest of the time it is turned-off. A descriptive timing diagram is shown in Fig. 18. During the turn-on time the circuit is turned-on; in this time slot the circuit is recovering its nominal operation condition from turn-off condition, and then it is not processing the signal. During the signal processing time, the circuit is effectively processing the signal. After this the circuit is turned-off and remain idle for the turn-off time.

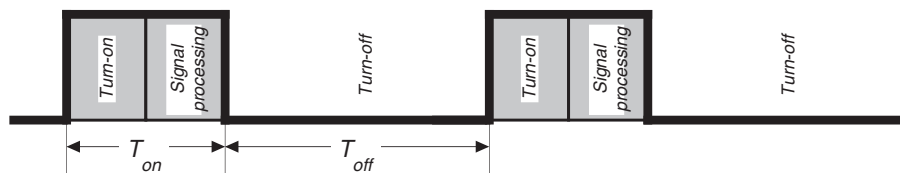


Fig. 18 – The duty-cycle technique

In some cases, the duty-cycle is not fixed and the circuit is turned on only when it is needed by the application, by an ‘activation’ signal. In some other cases, the system is not requiring this feature, and the turn-on/off timing is regulated by a fixed clock. The average power consumption can now be expressed as:

$$\text{Eq. (20)} \quad P_{av} = \frac{P_{on} \cdot T_{on}}{T_{on} + T_{off}}$$

The relative weight of T_{on} with respect to the total period ($T_{on} + T_{off}$) gives the net power saving advantage of the technique. Thus this technique would give a significant advantage when $T_{on} \ll T_{off}$. Notice that T_{on} includes also the turn-on time, which is dependent on time constants of the full system, which can be electrical or of different nature (mechanical, thermal, etc...). These time constants should be minimum, but in some cases they are out of the control of the circuit designer and so they may severely impact on the application of this technique. In the following a couple of cases, which are representative of the concept, are given. Both of them refers to sensor interface, which are power demanding during sensing, but the system requires a measurement with a very low data-rate (1/min or less).

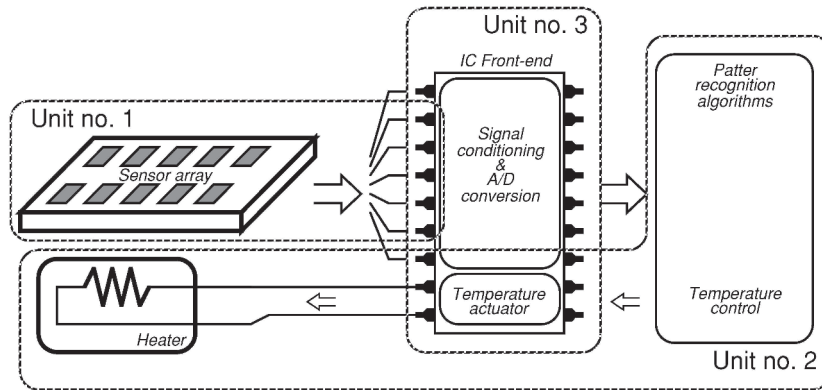


Fig. 19 – The gas-sensor structure

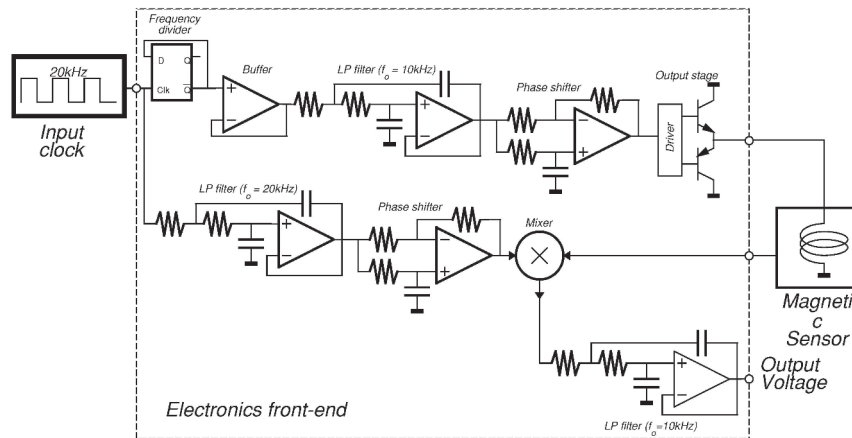


Fig. 20 – The electronic compass structure

Fig. 19 shows a gas-sensor system under development [19]. In this system a measurement is required only every two minutes. In the structure the most power hungry block is the heater, which has to supply a larger current to heat the sensor to the target temperature. Thus the adopted solution is to turn on/ff the full structure every two-minutes and leaving the system active only for the time needed for the start-up (which includes and is dominated by the thermal time constant of the sensor substrate to be heated to 200° C) and for the measurement, which are in total about 500 μ s. The duty-cycle is then of $5 \cdot 10^{-4}$ %. As a consequence the overall average power consumption is 1/240.000 time lower than the full operation power consumption.

A similar concept will be used in a fully-integrated dual-axis electronic compass under development, shown in Fig. 20 [20]. In this case the magnetic sensor is a flux-gate, which has to be excited with a large current (to saturate the ferromagnetic material). On the other hand the required measurement rate is about 1Hz, while the measurement process (including also the start-up, which in this case is negligible) requires only 0.2 ms. A 0.2% duty-cycle can then be adopted, with the consequent power consumption reduction.

Similar concepts are also available in many other situations, like, for instance, in some telecommunication systems, in which the receiver is completely turned-on only when a coming signal to be processed has been recognized. This reduces the receiver power consumption when no signal is coming. In same way, also the transmitter is turned-off when no transmission is required, which may occur when the terminal is inactive or in a TDD communication scheme during the time frame not allowed for the transceiver. In this case the power saving is considerable since the Power Amplifier is one of the most power hungry block in the transceiver.

3.1.2 *The time-sharing technique*

The time-sharing technique is an improvement of the duty-cycle technique and corresponds to the use of the same circuit (which would be disabled in a duty-cycle scheme) in different positions of the systems. This means that some parts of the circuit may be in idle state and only one part of the circuit is active. As a consequence the application of this technique is possible only in those cases in which the output signals in some nodes are not read, and the relative driving force can be nulled. Many examples of this technique are given in SC circuits. In the basic SC biquadratic cell of Fig. 21.a, the opamp embedded in the 1st integrator is active only during phase 1, while the opamp of the 2nd integrator is active only during phase 2. As a consequence, the same opamp can be shared between the two integrators., as shown in Fig. 21.b.

For a power budget, in the standard scheme of Fig. 21.a, the two opamp has to settle exactly in the same time slot of the single opamp of Fig. 21.b. As a consequence each of the two opamp of Fig. 21.a consumes the same power of the single opamp of Fig. 21.b. This means a net power saving of about 50%.

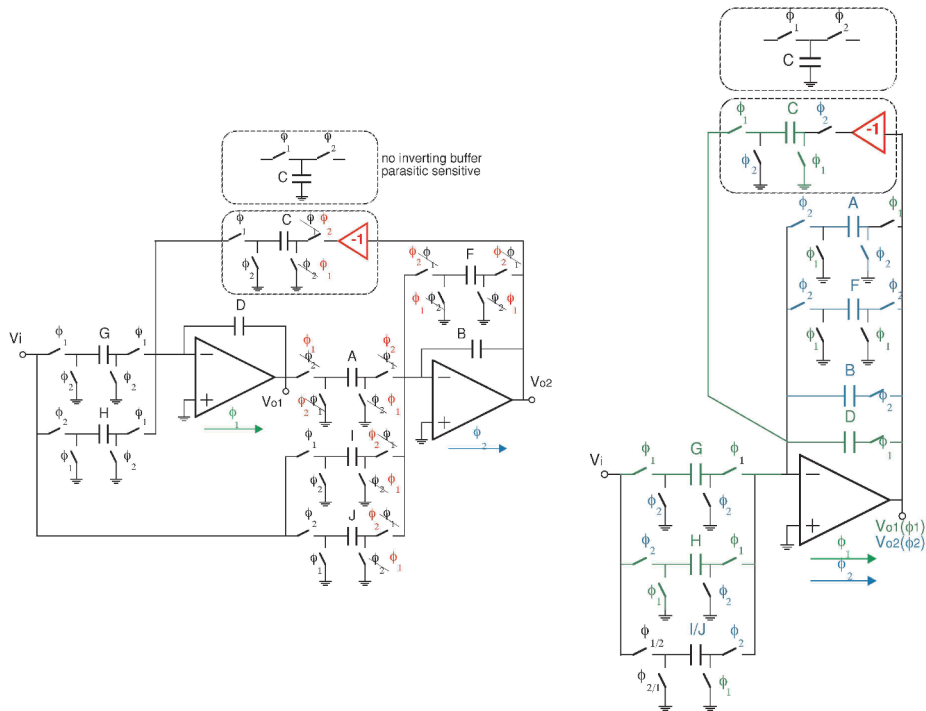


Fig. 21 – Time-sharing operations for a SC biquadratic cell

The main limitations of this solution are the following:

- the voltages of two integrator output nodes may be largely different, and so the opamp (whose output node is switched between the two output nodes) is forced to quickly update its output voltage and this is achieved with a slew-rate, which may be larger than that required for the opamp of Fig. 21.a;
- the opamp used in different integrators requires to connect/disconnect the integrating capacitor, and this may corrupt the information stored as a charge on the integrating capacitor due to charge injection, clock feedthrough, etc.

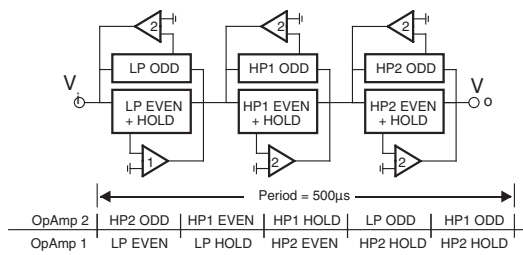


Fig. 22 – Time-sharing operations for a SC cascade structure

The concept above described for the basic biquad cell, can be extended to more complex SC structures, like cascade of biquads as shown in Fig. 22 [21], ladder structures [22], and $\Sigma\Delta$ modulators [11]. In these cases some opamps are used in different positions and at an operating rate (f_{op}) higher than the external rate (f_{ext}) of the system.

Examples of the time-sharing are also given by several SC compensation scheme. In these cases, some features of the active devices (typically the opamp) should require a large power to fit the specification. The alternative approach is then to design a low-performance opamp and use one time slot to calibrate it. This concept can be done for the improvement of the overall SC system performance w.r.t. to opamp limitations [23], [24], [25], [26], [27], in:

- $1/f$ & low-frequency noise & offset (CDS technique)
- finite opamp gain
- finite opamp bandwidth (double-sampling technique)

Fig. 23 shows a very popular SC structure implementing the Correlated Double Sampling (CDS) technique. During phase ϕ_1 , the structure is self-calibrating to reduce the offset and $1/f$ noise at the output node and to compensate for the effects of the opamp finite gain (at least for low-frequency input signal). As a consequence, the opamp could be designed to exhibit lower $1/f$ noise and offset performance, but with lower power consumption.

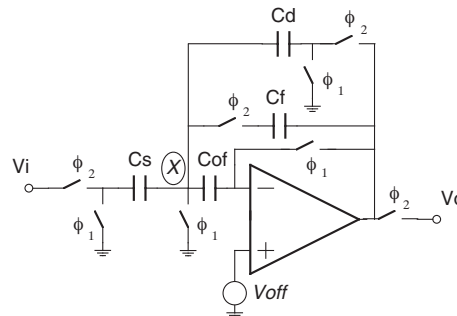


Fig. 23 – A ‘time-sharing’ structure with CDS for offset, $1/f$ and finite gain compensation

Another application of the time-sharing operation is given by the double-sampling technique, which gives an immediate advantage of two to the opamp speed requirements, which reflects on lower power consumption. The output value of the standard SC integrator of Fig. 24.a is read only at the end of ϕ_2 , and the time available for the opamp to settle is $T_s/2$. The equivalent Double-Sampled structure is shown in Fig. 24.b. The capacitor values for the two structures are the same, and thus they implement exactly the same transfer function. The time evolutions for the two structures are compared in Fig. 25.

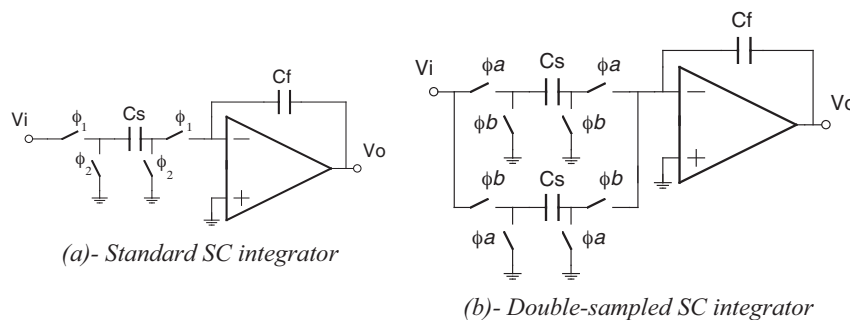


Fig. 24 – Standard and Double-Sampled SC integrator structures

In the Double-Sampled integrator the opamp has to settle within phase ϕ_A , which is as long as the sampling period T_s . Therefore for the Double-Sampled SC integrator, the time available for the opamp to settle is doubled w.r.t. the standard solution. This advantage can be used to reduce at low sampling frequency: a smaller bandwidth to be guaranteed by the opamp reduces its power consumption.

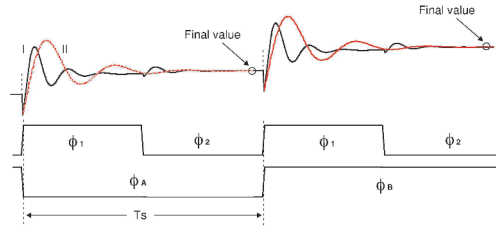


Fig. 25 – Standard (Line I) and Double-sampled (Line II) SC integrator operation

The cost of the double-sampled structure is the doubling of all the switched-capacitors. In addition, in the case of a small mismatch between the two parallel paths, mismatch energy could be present around $F_s/2$ [28].

3.1.3 Dynamic reduction in $\Sigma\Delta$ modulator ([29], [4])

Another possible approach to reduce power consumption has been applied to $\Sigma\Delta$ modulator. It based on the concept that the power consumption depends also on the signal amplitude (slew-rate, etc...). For this reason, in the scheme of Fig. 26 the input signal is forwarded to the quantizer input. This means that the loop filter is processing only the quantization noise, whose amplitude can be strongly reduced if a multibit quantizer is adopted. In this way a small output swing is required to the loop filter, which may exhibit better linearity at lower power consumption.

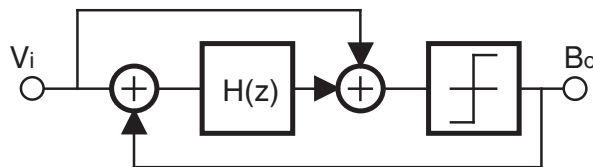


Fig. 26 - LP $\Sigma\Delta$ Modulator

3.1.4 Dynamically power-optimized circuits ([30])

Another approach to reduce/optimize the overall power consumption consists in adapt the current level w.r.t. the signal level. An example of this concept is shown in Fig. 27.

For low signal level, lower noise is required to achieve a given DR and so large power consumption is needed. This corresponds to operate with phase ϕ_{DR} active. On the other hand, for large signal level, the power consumption can be reduced and so the opamp on the bottom can be turned-off. In both phase the transfer function is the same. The critical issue of this technique is the time needed for the bottom part when it is turned-off to be updated before being connected to the top part.

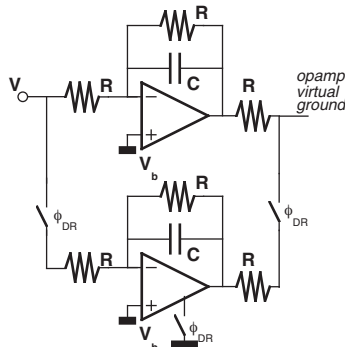


Fig. 27 - Dynamically optimize active-RC building block

3.2 Low-Voltage Solutions

3.2.1 Low-Voltage Continuous-Time Systems

LV continuous-time systems may be divided in open-loop (like gm-C filters) or closed-loop (like active-RC filters) structures.

Any discussion about open-loop structure deals with the differential stage voltage limitations, as already discussed in the opamp section. The minimum voltage supply (V_{DDmin}) is limited by the output-to-input connection of two similar stages, as shown in Fig. 28. This means that the V_{DDmin} depends both on input and output stage limitation, and it is given by:

$$\text{Eq. (21)} \quad V_{DDmin} = (V_{GSdiff} + V_{sat_top}) + 2 \cdot V_{sw} + V_{sat_bottom} = V_{TH} + 3 \cdot V_{ov} + 2 \cdot V_{sw}$$

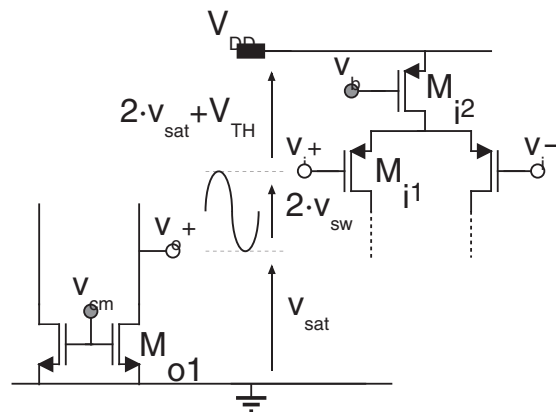


Fig. 28 - LV Gm-C filter connection

As a further definition of the above expression, the V_{GSdiff} is correlated with the maximum signal amplitude i.e. with V_{sw} . A slightly reduction of the V_{DDmin} is achievable by using a pseudo-differential structure [31], which avoids the use of

the tail current generator M_{i2} and, as a consequence, the contribution of V_{sat_top} is cancelled. Of course a pseudo-differential structure exhibits a worse CM-signal rejection and a worse CM-signal control. To guarantee these features at the same level of fully-differential stages, additional circuitry is necessary, which increases the overall power consumption.

On the other hand, in the design of a closed-loop structure the use of the virtual ground separates the dependence of V_{DDmin} on both input and output stage. To maximize output swing, the output stage is designed to fix $V_{out}=V_{DD}/2$. The supply minimization deals with the virtual ground, i.e. the input common-mode voltage of the embedded opamps. Fig. 29.a shows a possible solution for the basic active-RC integrator embedded in a closed loop system. Without any level shift (i.e. no I_o and no R_B), for cell coupling it has to verify $V_i=V_o=V_{DD}/2$. This would require a significantly large V_{DDmin} (given approximately by $V_{DDmin}=V_{DD}/2+V_{TH}+2\cdot V_{ov}$. Assuming a rail-to-rail output stage i.e. $V_{DD}=2\cdot V_{sw}+2\cdot V_{ov}$, it results:

$$\text{Eq.(22)} \quad V_{DDmin}=V_{sw}+V_{TH}+3\cdot V_{ov}$$

A V_{DDmin} reduction is obtained with a level shift from the output to the input, decoupling the input stage requirements from the output stage requirements. The level shift can be realized as shown in Fig. 29. In Fig. 29.a the level shift is realized with a current source I_o . The input stage is optimizing by biasing V_b just one V_{ov} (as required by the I_o current source). The optimum value for the current I_o to set $V_{out_DC}=V_{DD}/2$ is given by:

$$\text{Eq.(23)} \quad I_o = \frac{V_{DD}/2 - V_b}{R_i}$$

Using this scheme the minimum supply voltage is given by:

$$\text{Eq.(24)} \quad V_{DDmin} = (V_{TH}+2\cdot V_{ov}) + V_b = (V_{TH}+2\cdot V_{ov}) + V_{ov} = V_{TH} + 3\cdot V_{ov}$$

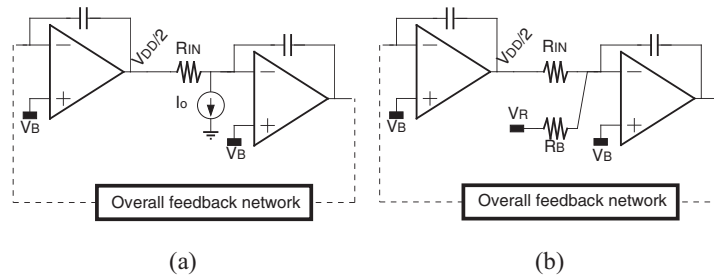


Fig. 29 – LV Active-RC dumped integrator

The above V_{DDmin} is just one V_{ov} of the absolute minimum supply voltage for analog circuits. The above scheme could be conceptually implemented by replacing the current generator I_o with a resistor R_B connected to V_R that sinks the current $(V_B-V_R)/R_B$, as shown Fig. 29.b. In this scheme, the opamp output dc-voltage is:

$$\text{Eq. (25)} \quad V_{out_dc} = V_B \cdot (1 + R_{IN}/R_B) - V_R \cdot R_{IN}/R_B$$

which is not limited to V_{sat} of the current source, but can be extended to be very close to the rail (but cannot reach it).

The above simple schemes are the high-level generalization of a number of biasing scheme operating at low-voltage (like the switched-opamp technique described in the following paragraphs). An example of its application in a mixed continuous-time sampled-data system is the sampler of SAR ADC that is shown in Fig. 30 [32].

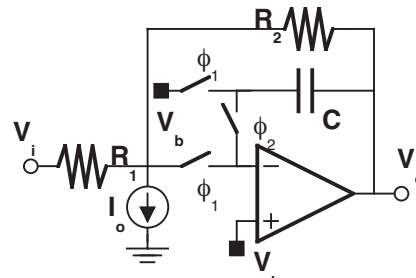


Fig. 30 – LV Sampler

3.2.2 Low-voltage Sampled-data Systems (Switched-capacitor - SC)

In the following the case of SC circuits will be addressed, since they appear a good compromise between circuit complexity and achievable performance. The typical SC integrator is shown in Fig. 31, where it is indicated that different supply voltages can be used to bias the opamps and the switches (and their relative driving circuits).

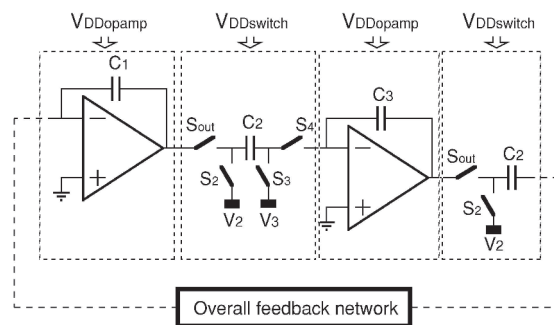


Fig. 31 – Typical SC integrator

3.2.2.1 Processing a reduced signal amplitude with standard SC solutions

The fundamental limitation to the operation of SC circuits at low-voltage are due to switches operations. From Fig. 4, it appears that at low-voltage using standard complementary switches results in a reduced (or zero) possible signal swing. In addition, this signal swing is strictly possible only in two regions: one is closed to ground and the other is closed to V_{DD} . In these region using complementary switches is no more advantageous, and so single MOS switches (NMOS-only or PMOS only) can be used. An NMOS-only switch may then be adopted when the signal swing is closed to ground, while a PMOS-only switch would be used when the signal swing is closed to V_{DD} . This however makes the possible signal swing to depend on the power supply. In fact, let's consider the case of a NMOS-only switch connected to a signal biased at V_{DC} and with a signal swing V_{SW} . The resulting switch conductance is shown in Fig. 32.

In this case the minimum supply voltage V_{DDmin} results to be given by:

$$\text{Eq.}(26) \quad V_{DDmin} = V_{DC} + V_{SW} + V_{THN}(V_{DC} + V_{SW}) + V_{ov}$$

Notice that the NMOS threshold voltage V_{THN} depends on the voltage, whom the switch is connected to, i.e. $(V_{DC} + V_{ov})$ due to the body effect. As a consequence, V_{DDmin} depends directly and indirectly on V_{SW} . This low-voltage SC design approach has been adopted in the design of a Sample&Hold whose scheme is shown in Fig. 33 [33]. It presents a pseudo-differential (PD, i.e. two single-ended structures driven with opposite signals) double-sample (DS, i.e. the input is sampled during both clock phases) structure. It operates as follows. During phase 1, C_{IP} and C_{IM} sample the input signal, referred to V_{DD} . During phase 2, C_{IP} and C_{IM} are connected in the opamp feedback, producing the output sample¹.

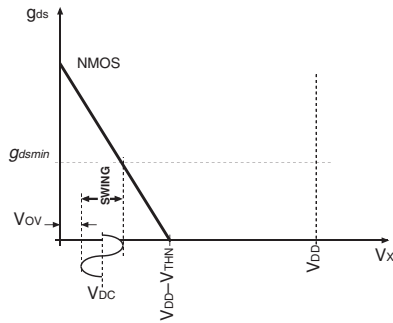


Fig. 32 – Possible swing for an NMOS-only switch

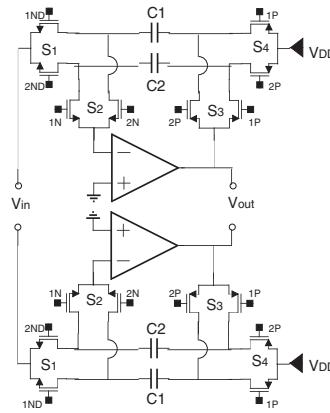


Fig. 33 – PD-DS S&H circuit

Switch operations are guaranteed by proper control of the voltage at the node where the switches are connected to, as shown in Fig. 34. The opamp input dc-voltage is set to ground by the feedback action. Switch S_2 is then realized with a single NMOS device. Switch S_4 is connected to V_{DD} , and then it is realized using a PMOS device. The input signal dc-voltage V_{in_dc} is set close to ground: this allows to realize S_1 with a single NMOS device. The opamp output dc-voltage (V_{out_dc}) is then fixed at the value:

$$\text{Eq.}(27) \quad V_{out_dc} = V_{DD} - V_{in_dc}$$

and results to be close to V_{DD} . S_3 is then realized with a PMOS device. The minimum supply required by the structure is then fixed by proper operation of switches S_1 and S_3 , and it is given by:

¹ This structure offers the following advantages. The PD structure avoids the implementation of a Common-Mode Feedback circuit, a critical block for low-voltage circuits. The DS structure avoids any opamp reset phase which causes in single-sampled structures large output steps. Slew-rate requirements are then relaxed, also because the opamp has to charge only the output load, and not the feedback capacitor. Furthermore the opamp always operates with feedback factor equal to one, achieving maximum speed of response. Finally a negligible droop-rate is expected assuming an opamp MOS input device without input current.

$$\text{Eq.}(28) \quad V_{DDn} > V_{in_dc} + V_{sw} + V_{THn}(V_{in_dc} + V_{sw}) + V_{ov}$$

$$\text{Eq.}(29) \quad V_{DDp} > V_{sat} + 2 \cdot V_{sw} + V_{ov} + V_{THp}(V_{out_dc} - V_{sw})$$

where V_{sw} is the peak of the single-ended signal amplitude, V_{THn} and V_{THp} are the maximum values of the NMOS and PMOS threshold voltages obtained for the body effect evaluated at the maximum level of the signal swing. V_{sat} is the minimum distance from V_{DD} for the opamp output node before it enters in the saturation region. As previously anticipated, V_{DD} depends from the signal swing directly and indirectly, through the dependence of the V_{TH} from the signal value. This aspects can be studied plotting the available output swing vs. the power supply V_{DD} . In the proposed design the following values have been used: $V_{ov}=50\text{mV}$, $V_{sat}=80\text{mV}$. The V_{DDmin} vs. differential output swing is given in Fig. 35. The line with the stars indicates the technological typical case for both NMOS and PMOS, while all the other lines indicate all the possible combinations of NMOS and PMOS worst cases for the used $0.5\mu\text{m}$ CMOS technology. For the typical case 600mV_{pp} output swing are possible with 1.2V power supply.

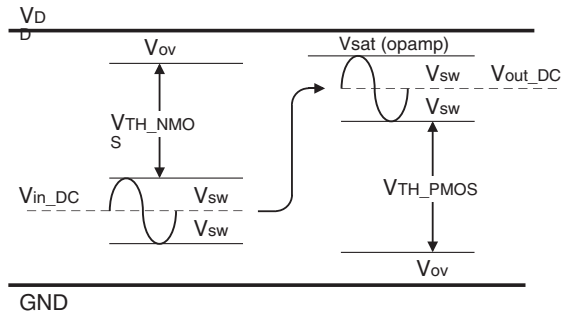


Fig. 34 – Voltage drop across the switches

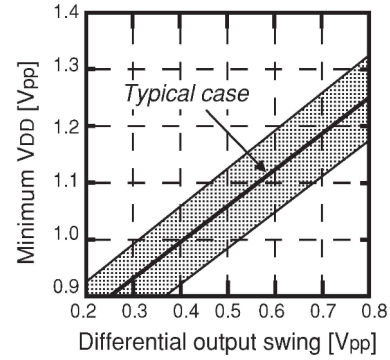


Fig. 35 – V_{DDmin} vs. output swing

The main advantage of this approach is the fact that it uses only standard block design (at the cost of a reduced and technology-dependent signal swing) and so there are no critical limitations to operate with a high sampling frequency.

3.2.2.2 Processing a rail-to-rail signal amplitude with novel solutions

As the supply voltage reduces, accordingly the available output swing strongly reduces, and, as a consequence, also the DR. It is therefore mandatory to maximize the output swing, which has to be rail-to-rail (using $V_{out_dc}=V_{DD}/2$). This can be done with the following approaches: the on-chip supply voltage multiplier, the on-chip clock multiplier and the switched-opamp technique. These design approaches are different with respect to the supply used for the switch section and for the opamp section (as shown in Fig. 31). Depending on this choice of opamps and

switches supply, switches and opamp operation are guaranteed in different ways and, as a consequence, particular problems, which limit its applications arise.

3.2.2.2.1 On-chip supply voltage multiplier ([34])

If the SC designer wants to re-use all his know-how, the only possible design approach is to generate on-chip an auxiliary supply $V_{DDmulti}$ to be used to power the complete SC filter. In this way the SC filter is designed using the available analog cells for opamps and switches, operating from the multiplied supplied voltage (i.e. with $V_{DDswitch}=V_{DDopamp}=V_{DDmulti}$).

The on-chip supply voltage multiplier suffers from the following limitations:

- the technology robustness: the scaled-down technology presents the maximum acceptable electric field between gate and channel (for gate oxide break-down) and between drain and source (for hot electrons damage) must be reduced and this results in an absolute limit to the value of the multiplied supply voltage;
- the need to supply a dc-current from the multiplied supply forces to use an external capacitor: an additional cost, not feasible for other system considerations;
- the conversion efficiency of the charge-pump cannot be 100% and this could limit the application of this approach in battery operated portable systems;

For these arguments this approach appears the least feasible for future applications and it will not be discussed any further.

3.2.2.2.2 On-chip clock voltage multiplier ([35], [36])

A second and more feasible alternative to operate low-voltage SC filters is the use of on-chip clock multiplier to drive only the switches, while the opamps operate from the low-supply voltage. Thus the voltage multiplier has only to drive the capacitive load due to the switch gates, while it is not required to supply any dc-current to the opamp. No external capacitor is then required and the SC filter is fully integrated. Using this design approach the switches can operate as in standard SC circuit working with higher power supply.

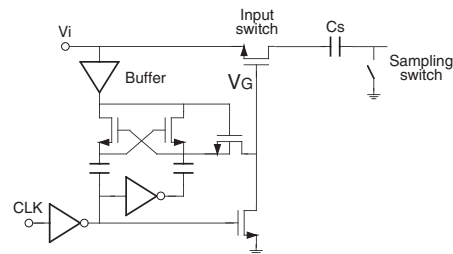


Fig. 36 – A possible clock multiplier

On the other hand, the opamp has to be properly designed in order to operate from the reduced supply voltage. In particular the opamp input dc-voltage is necessary to be set $V_{in_DC}=0$ (this will be explained later), while the opamp output dc voltage is set to $V_{out_DC}=V_{DD}/2$ to achieve rail-to-rail output swing. These dc levels are not equal and so a voltage level shift must be implemented. Such a level shift can be efficiently implemented using SC technique. In this way the operation is possible due to the full functionality of the switches at any input voltage using the multiplied clock supply. In the scheme of Fig. 31, assuming $V_2=V_{out_DC}$ and $V_3=V_{in_DC}$ gives the proper dc-voltage at the opamp input and at the output nodes. This design

the switches are driven with the maximum overdrive, i.e. $V_{DD}-V_{TH}$. The minimum supply voltage required for proper operation of the switches is then given by:

$$\text{Eq.(30)} \quad V_{DDmin} = V_{TH} + V_{ov}$$

where V_{TH} is the larger of the two threshold voltages (N-type and P-type). The V_{DDmin} value is of the same order as the minimum supply voltage for the digital CMOS circuits operation.

As previously described, also for the SOA technique is necessary to implement a level shift due to the difference between the opamp input and output dc-voltages. This is efficiently implemented in the scheme of Fig. 37 with the switched-capacitor C_{DC} , which gives a fixed charge injection into virtual ground. The charge balance at the opamp input node allows to evaluate the amount of the level shift as:

$$\text{Eq.(31)} \quad -C_{IN} \cdot V_{out_DC} - C_{DC} \cdot V_{DD} = C_{IN} \cdot (V_{in_DC} - V_{DD}) + C_{DC} \cdot V_{in_DC} = 0$$

Since V_{in_DC} is set to ground, the opamp output dc-voltage V_{out_DC} is fixed at:

$$\text{Eq.(32)} \quad V_{out_DC} = V_{DD} \cdot \frac{C_{IN}}{C_{DC}}$$

To set V_{out_dc} to be equal to $V_{DD}/2$ it is necessary to design $C_{DC} = C_{IN}/2$. This allows to satisfy all the points previously stated with the scheme of Fig. 37. This concept can be also shown considering the scheme of Fig. 29.b as the equivalent continuous-time scheme of Fig. 37. The key advantage of the SC technique is that the proper phasing of C_{DC} realizes a negative impedance, and this allows to fix $V_{in_dc} = 0$. The main problems presented by the SOA can be summarize as follows:

- only the non-inverting and delayed SC integrator has been up to now proposed in literature. Thus, a sign change must be properly implemented to close the basic two-integrators loop and to build high-order filters. This problem is still open for the single ended structure and the only proposed solution is using an extra inverting stage;
- any unaccuracy in the C_{DC} size gives an extra offset at the output node which limits the output swing;
- any noise and disturb present on V_{DD} is injected into the signal path.

A fully differential structure can alleviate all of the above problems. In fact:

- a fully-differential architecture provides both signal polarities at each node, which allows to build high order structures without any extra elements (e.g. inverting stage);
- any disturbance (offset or noise) injected by C_{DC} results in a common mode signal which is largely rejected by the fully-differential operation.

Nonetheless the above advantages, fully-differential structures present a drawback, which is their need for a Common Mode FeedBack (CMFB) circuit, which becomes critical at low supply voltage as discussed in the previous sections. In addition to this, the SOA design approach still suffers for the following open problems:

- a SOA structure uses an opamp which is turned on and off. The opamp turn on time results to be the main limitation in the increasing the sampling frequency;
- the output signal of a SOA structure is available only during one clock phase, because during the other clock phase the output is set to zero. If the output signal is read as a continuous-time waveform the zero-output phase has two effects: a gain loss of a factor of 2, and an increased distortion. This second drawback is due to the large output steps resulting in slew-rate-limited signal transient and glitches. However when the SOA integrator precedes a sampled-data system (like an ADC) the SOA output signal is sampled only when it is valid and both the above problems are cancelled;
- the input coupling switch sees the entire voltage swing and so is still critical: only ac-coupling through a capacitor appears a good solution, up to now.

3.2.2.2.4 Turn-on time reduction: the Unity-Gain Feedback technique ([38])

The Unity-Gain Feedback technique allows to reduce the turn-on time of the opamp. This technique does not to completely turn-off the opamp during the off phase, but it biases it in a stand-by condition.

The relative scheme for the basic SC integrator is shown in Fig. 38. In the off phase, the output nodes are driven by a battery in the feedback to $V_{DD} - V_{sat}$, without turning-off the opamp. This dramatically reduces the turn-on time, allowing the use of higher sampling frequency. A possible drawback of this approach is the residual signal at the output during the off-phase.

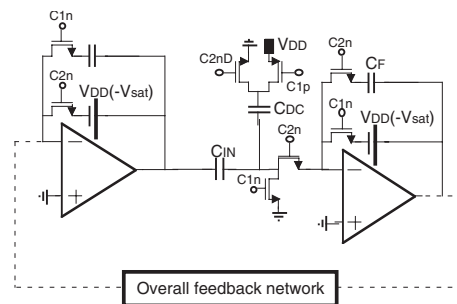


Fig. 38 – The Unity-Gain-Feedback Technique SC integrator

3.2.2.2.5 The input series switch

One of the main lacks of the SOA technique is the implementation of the series switch to be connected at the input signal, which can exhibit rail-to-rail signal swing. The input signal cannot be directly connected to a gate since its dc-voltage is set to $V_{DD}/2$, which is higher than V_{TH} . Thus a possible solution consists in connecting the input signal to a passive impedance to be connected to a some kind of virtual ground.

3.2.2.2.5.1 The Active Switched-Opamp series switch ([39])

Fig. 39 shows the conceptual scheme of a possible solution: it consists in a switched-buffer, implemented with a switched-opamp in inverting configuration. In the case of $V_{batt}=0V$ and V_{in_DC} set to ground (as previously described), and V_{out_DC} fixed to $V_{DD}/2$, V_{s_DC} results to be set to $-V_{DD}/2$, not a feasible value for the

previous stage operation. On the other hand, if $V_{batt}=V_{DD}/2$, node X acts like a virtual ground set to $V_{DD}/2$, and the bias condition becomes: $V_{out_DC}=V_{DD}/2-V_{s_DC}$. Using $V_{s_DC}=V_{DD}/2$ for rail-to-rail input swing of the preceding stage, V_{o_DC} is set to $V_{DD}/2$. Notice that in R_1 and R_2 only signal current flows and, with $R_1=R_2$, V_o follows V_s with negative unitary gain.

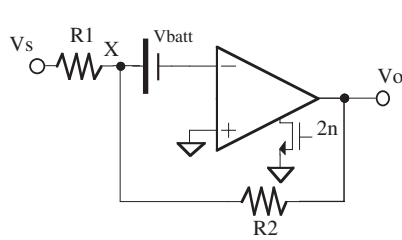


Fig. 39 – The series switch scheme

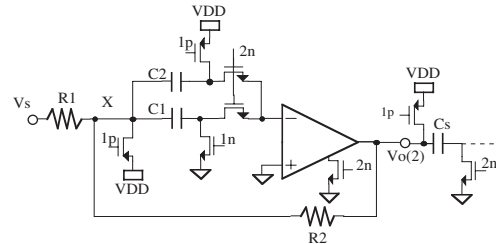


Fig. 40 – The switched-opamp buffer

Fig. 40 shows the complete circuit. In this circuit the battery V_{batt} is implemented with SC technique and operates as follows. During phase 1 capacitor C_1 is charged to V_{DD} , while capacitor C_2 has no charge on its nodes since they are both connected to V_{DD} . During phase 2 capacitors C_1 and C_2 are connected in parallel. Using $C_1=C_2$, across both capacitors a voltage equal to $V_{DD}/2$ results. In this phase no charge is added to C_1 and C_2 , which then act like a battery from opamp input node (set to ground) to node X which results to be set to $V_{DD}/2$, as required. During phase 2 the feedback network (R_1 - R_2) is active since the opamp inverting input node is set to ground. This forces V_X to be set to $V_{DD}/2$, as required, and $V_o(2)$ follows $V_s(2)$. The value of $V_o(2)$ is sampled on C_s which is the input capacitor for the following stage in which it injects its charge during the following phase 1.

3.2.2.2.5.2 The switched-RC technique ([40])

An alternative solution for the input series switch is given by the switched-RC technique, as shown in Fig. 41.

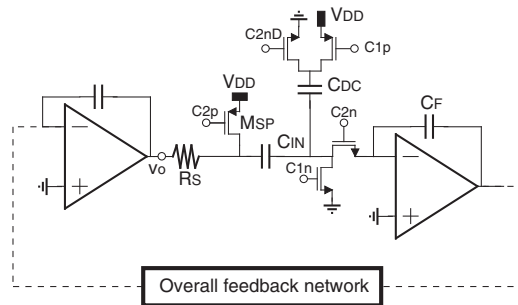


Fig. 41 – The switched-RC integrator technique

In this case, the opamp driving force is never turned-off. However, canceling the series switch would results is a large output current when the output node is con-

nected to the reference voltage (ground or V_{DD}). Thus, at the driving opamp output node a series resistance is connected in order to limit the output current.

The main draw-back of this technique is that, when the series switch would be turned-off, there is a residual signal (given by the resistive partition of R_I with the on-resistance of M_{SP}) which is continuously injected in the integrator. And this signal partition is also non-linear due to the switch on-resistance non-linearity.

4. References

- [1] M. Dessouky and A. Kaiser, "Very low-voltage digital-audio delta-sigma modulator with 88-dB range using local switch bootstrapping", *IEEE J. of Solid-State Circuits*, pp. 349-355, 2001.
- [2] V. Peluso, P. Vancorenland, A. M. Marques, M. S. J. Steyaert and W. Sansen, "A 900-mV low-power delta-sigma A/D converter with 77-dB dynamic range", *IEEE J. of Solid-State Circuits*, pp. 1887-1897, December 1998.
- [3] Yao Libin, M. S. J. Steyaert and W. Sansen, "A 1-V 140- μ W 88-dB audio sigma-delta modulator in 90-nm CMOS", *IEEE J. of Solid-State Circuits*, pp. 1809-1818, December 2004.
- [4] R. Gaggl, M. Inversi and A. Wiesbauer, "A Power Optimized 14-Bit SC Delta-Sigma Modulator for ADSL CO Applications", *ISSCC 2004 Digest of Technical Papers*, San Francisco, USA, pp. 82-83, February 2004.
- [5] S. Rabbii and B. A. Wooley, "A 1.8-V digital-audio sigma-delta modulator in 0.8 μ m CMOS", *IEEE J. of Solid-State Circuits*, pp. 783-796, December 1997.
- [6] L. A. Williams III and B. A. Wooley, "A third-order sigma-delta modulator with extended dynamic range", *IEEE J. of Solid-State Circuits*, pp. 193-202, 1994.
- [7] O. Nys and R. K. Henderson, "A 19-bit low-power multibit sigma-delta ADC based on data weighted averaging", *IEEE J. of Solid-State Circuits*, pp. 933-942, 1997.
- [8] C. B. Wang, S. Ishizuka and B. Y. Liu, "A 113-dB DSD audio ADC using a density-modulated dithering scheme", *IEEE J. of Solid-State Circuits*, pp. 114-119, 2003.
- [9] Yang YuQing, A. Chokhawala, M. Alexander, J. Melanson and D. Hester, "A 114-dB 68-mW Chopper-stabilized stereo multibit audio ADC in 5.62 μ m²", *IEEE J. of Solid-State Circuits*, pp. 2061-2068, 2003.
- [10] L. Lentola, A. Mozzi, A. Neviani, and A. Baschiroto, "A 1 μ A Front-End for Pacemaker Atrial Sensing Channels with Early Sensing Capability", *IEEE Transactions on Circuits and Systems - Part II - August 2003 - pp. 397-403*
- [11] V. Colonna, G. Gandolfi, F. Stefani, A. Baschiroto, "A 10.7MHz, Self-Calibrated, Switched-Capacitor Based, Multibit 2nd-Order Bandpass $\Sigma\Delta$ Modulator with On-Chip Switched-Buffer ", *IEEE J. of Solid-State Circuits*, pp. 1341-1346, Aug. 2004

- [12] S. Chatterjee, Y. Tsvividis, P. Kinget, "0.5V Filter with PLL-Based Tuning in 0.18 μ m CMOS", ISSCC 2005 Digest of Technical Papers, San Francisco, USA, pp. 506-507, February 2005
- [13] H. Banba, H. Shiga, A. Umezawa, T. Miyabata, T. Tanzawa, S. Atsumi and K. Sakuii, "A CMOS bandgap reference circuit with sub-1-V operation", IEEE Journal of Solid-State Circuits, vol. 34, pp. 670-674, May 1999.
- [14] P. Malcovati, F. Maloberti, C. Fiocchi and M. Pruzzi, "Curvature Compensated BiCMOS Bandgap with 1 V Supply Voltage", IEEE J. of Solid-State Circuits, vol. 36, pp. 1076-1081, July 2001.
- [15] A. Baschiroto, R. Castello, "A 1V 1.8MHz CMOS Switched-opamp SC filter with rail-to-rail output swing", IEEE J. of Solid State Circuits, pp. 1979-1986, December 1997.
- [16] B. J. Blalock, P.E. Allen and G.A. Rincon-Mora, "Designing 1-V opamp using standard digital CMOS technology." IEEE Transactions on Circuits and Systems II, vol. 45, pp. 930-936, July 1998.
- [17] M. Waltari, and K.A.I. Halonen, "1-V 9-Bit Pipelined Switched-Opamp ADC", IEEE J. of Solid State Circuits, January 2001, pp.129-134
- [18] A. Baschiroto, P. Cusinato, G. Montagna, R. Castello, "Fully differential, switched capacitor, operational amplifier circuit with common-mode controlled output", USA patent (issued) - 6,411,166 B2 -25/6/2002
- [19] <http://ims.unipv.it/prin03/>
- [20] A. Baschiroto, E. Dallago, P. Malcovati, M. Marchesi, G. Venchi, "Precise vector-2D magnetic field sensor system for electronic compass", IEEE Sensors 2004 - Vienna (Austria) - Oct. 2004 - pp. 1028-1031
- [21] R. Castello, A. G. Grassi and S. Donati, "A 500-nA sixth-order bandpass SC filter", IEEE J. of Solid-State Circuits, vol. 25, pp. 669-676, June 1990.
- [22] F. Montecchi, "Time-shared switched-capacitor ladder filters insensitive to parasitic effects", IEEE Transactions on Circuits and Systems, vol. 31, pp. 349-353, April 1984.
- [23] C. Enz, and G.C. Temes, "Circuit techniques for reducing the effects of opamp imperfections: autozeroing, correlated double sampling, and chopper stabilization", Proc. IEEE, Nov. 1996, pp. 1584-1614
- [24] K. Haug, F. Maloberti, G.C. Temes, "Switched-capacitor integrators with low finite-gain sensitivity" Electronics Letters, vol. 21, pp.1156-1157, Nov. 1985
- [25] K. Nagaraj, J. Vlach, T.R. Viswanathan, and K. Singhal, "Switched-Capacitor Integrator with Reduced Sensitivity to Amplifier Gain," Electronics Letters, vol. 22, pp. 1103-1105, Oct. 1986
- [26] K. Nagaraj, T.R. Viswanathan, K. Singhal, and J. Vlach, "Switched-capacitor circuits with reduced sensitivity to amplifier gain," IEEE Trans. on Circuits and Systems, vol. CAS-34, pp. 571-574, May 1987.
- [27] L.E. Larson, and G.C. Temes, "Switched-capacitor building-blocks with reduced sensitivity to finite amplifier gain, bandwidth, and offset voltage",

- IEEE International Symposium on Circuits and Systems (ISCAS '87) , pp. 334-338.
- [28] J. J. F. Rijns and H. Wallinga, "Spectral analysis of double-sampling switched-capacitor filters", IEEE Transactions on Circuits and Systems, pp. 1269-1279, Nov. 1991.
- [29] J. Silva, U. Moon, J. Steensgaard, and G.C. Temes, "Wideband low-distortion delta-sigma ADC topology", IEE Electronics Letters, 7th June 2001,
- [30] M. Ozgun, Y. Tsvividis, and G. Burra, "Dynamically Power-Optimized Channel-Select Filter for Zero-IF GSM", ISSCC 2005 Digest of Technical Papers, San Francisco, USA, pp. 504-505, February 2005
- [31] F. Rezzi, A. Baschiroto, and R. Castello, "3V 12-55MHz BiCMOS pseudo-differential continuous-time filter", IEEE Transaction on Circuits and Systems - II - Nov. 1995 - pag. 896-903
- [32] S. Morteza pour and E. K. F. Lee, "A 1-V, 8-bit successive approximation ADC in standard CMOS process", IEEE J. of Solid State Circuits, Apr. 2000
- [33] A. Baschiroto, "A low-voltage Sample&Hold circuit in standard CMOS technology operating at 40Ms/s", IEEE Transactions on Circ. and Syst. - Part II - vol. 48, no. 4, April 2001, pp. 394-399
- [34] G. Nicollini, A. Nagari, P. Confalonieri, C. Crippa, "A -80 dB THD, 4Vpp switched capacitor filter for 1.5 V battery-operated systems", IEEE J. of Solid-State Circuits, vol. 31, pp. 1214 - 1219, August 1996
- [35] T.B. Cho, P.R. Gray, "A 10b, 20Ms/s, 35mW pipeline A/D converter", IEEE J. of Solid-State Circuits, vol. 30, pp. 166 - 172, March 1996
- [36] A.M. Abo, P.R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter", IEEE J. of Solid-State Circuits, pp. 599 - 606, May 1999
- [37] J. Crols, and M. Steyaert, "Switched-Opamp: an approach to realize full CMOS switched-capacitor circuits at very low power supply voltages", IEEE J. of Solid-State Circuits, Vol. SC-29, no.8, pp.936-942, August 1994
- [38] M. Keskin, U.-K. Moon, , and G. C. Temes, "A 1-V 10-MHz Clock-Rate 13-Bit CMOS $\Sigma\Delta$ Modulator Using Unity-Gain-Reset Opamps", IEEE JSSC, July 2002
- [39] A. Baschiroto, R. Castello, G.P. Montagna, "An active series switch for switched-opamp circuits", IEE Electronics Letters - 9th July 1998, Vol. 34, no. 14 - pag. 1365-1366
- [40] G.-C. Ahn, D.-Y. Chang, M. Brown, N. Ozaki, H. Youra, K. Yamamura, K. Hamashita, K. Takasuka, G. Temes, Un-Ku Moon, "A 0.6V 82dB Delta-Sigma Audio ADC Using Switched-RC Integrators", ISSCC 2005 Digest of Technical Papers, San Francisco, USA, pp. 166-167, February 2005

0.5 V ANALOG INTEGRATED CIRCUITS

Peter Kinget, Shouri Chatterjee, and Yannis Tsividis
Columbia University
New York, NY, USA

Abstract

Semiconductor technology scaling has enabled function density increases and cost reductions by orders of magnitudes, but for shrinking device sizes the operating voltages have to be reduced. As we move into the nanoscale semiconductor technologies, power supply voltages well below 1 V are projected. The design of MOS analog circuits operating from a power supply voltage of 0.5 V is discussed in this paper. The scaling of traditional circuit topologies is not possible anymore and new circuit topologies and biasing strategies have to be developed. Several design examples are presented. The circuit implementations of gate and body-input 0.5 V operational transconductance amplifiers and their robust biasing are discussed. These building blocks are combined for the realization of active varactor-tuned RC filters operating from 0.5 V using standard devices with a $|V_T|$ of 0.5V in a standard 0.18 μm CMOS technology.

1 Introduction

Analog circuits provide the connection of digital computing signal processing systems to the physical world. As such the true power of digital signal and information processing can only be exploited if analog interfaces with corresponding performance are available. Cost and size considerations push towards a co-integration of the analog interfaces and the digital computing/signal processing on a single die, thus in the same technology.

The International Technology Roadmap for Semiconductors (ITRS) [1] gives us a unique opportunity to look into the projected future of semiconductor technology and identify design challenges early (Fig. 1). The linewidth of CMOS

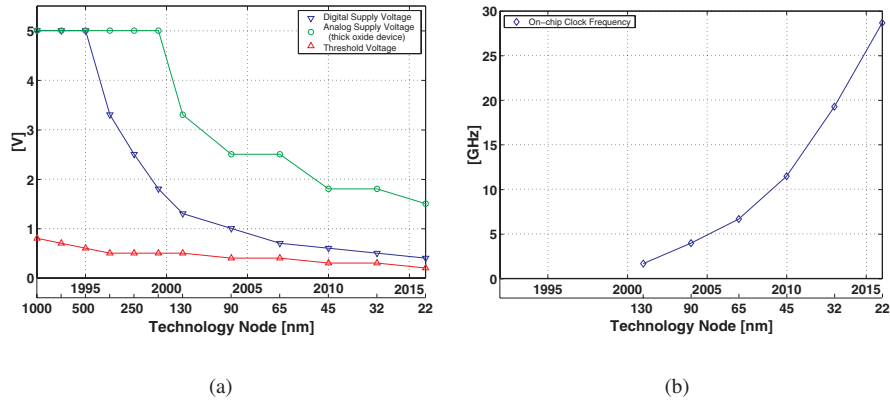


Figure 1: (a) Supply voltage and threshold voltage scaling and (b) on-chip clock frequency scaling according to [1]

technologies is projected to keep scaling deeper into nanoscale dimensions for the next two decades so the functionality density, the intrinsic speed of the devices and thus the signal processing capability will keep increasing. However, in order to maintain reliability, to reduce power density and to avoid thermal problems, the maximum supply voltage has to be scaled down appropriately. Fig. 1 shows the projections for the supply voltage and on-chip clock frequencies. The supply voltage scaling is beneficial for digital circuits since it reduces the power consumption quadratically. To maintain good ON/OFF characteristics of the MOS transistors for digital logic the transistor's threshold voltage V_T cannot be reduced as aggressively because static leakage levels would become too large. A minimum standard V_T of about 0.2 to 0.3 V is foreseen. By about the year 2013 at the 32 nm node a power supply voltage of 0.5 V is projected for high performance digital circuits. Also important to note is the fact that the scaling of the supply voltage for nanometer technologies is mainly driven by reliability and breakdown concerns. Consequently any internal voltage boosting of the external low voltage may not be possible.

These low power supply voltages and the relatively high device threshold voltages are a major obstacle for the realization and performance of analog circuits. Smaller supply voltages result in smaller available signal swings. The reduction of circuit errors due to thermal noise or offset voltages often leads to higher power consumption [2–6]. In addition, devices used in high speed linear circuits need to be biased in moderate or strong inversion with a minimum voltage overdrive ($V_{GS} - V_T$) (approx. 0.15 to 0.2 V) resulting in a $V_{DS,sat}$ requirement of about 0.15 V. Typical analog building blocks require a supply voltage which is several $V_{DS,sat}$ plus the signal swing, or a V_T plus several $V_{DS,sat}$

plus the signal swing. At supply voltages below 1 V the design of analog circuits becomes very challenging since the traditional circuit techniques run out of voltage headroom (see e.g., [5–10]).

These challenges can be addressed with technology modifications or with circuit design solutions. A straightforward technology solution is to add thick oxide devices that are less aggressively scaled; these are slower but can operate with larger supply voltages (see Fig. 1(a)). They allow a resizing or sometimes even a reuse of I/O and analog building blocks. Another technology option is to include low V_T [11] or native devices (zero V_T). These offer some extra headroom in circuits [12], but low V_T devices typically require an extra mask; native devices are typically less well characterized or modeled and sometimes have less reproducible characteristics. Extra semiconductor processing steps and masks result in extra cost and turn-around time. Since the analog interfaces typically occupy only 5 to 30% of the die area on large system-on-a-chip (SoC) circuits, the increased cost is hard to justify economically in large volume applications.

In the past decade we have witnessed significant design innovations to reduce the supply voltage of analog circuits from 5 V to 3.3 V, to 2.5 V, and recently to 1.8 V and even 1.3 V. Clock voltage boosting [13, 14], the switched-opamp circuit technique [15], back-gate driven circuits [16–18], rail-to-rail input stages [19, 20], multi-stage amplifiers with nested-Miller compensation [19, 21, 22], and level-shift techniques [23] are a few examples. Several amplifiers operating at 1 V [18], [24], [25] and down to 0.9 V [26] have been demonstrated. Sub-1V analog-to-digital converters [27–30] have also been recently demonstrated.

2 Low voltage analog circuit design challenges and opportunities

Operating a MOS device at low voltages For applications requiring high bandwidths or high clock and sampling rates, MOS devices are biased in the strong inversion region, i.e., $(V_{GS} - V_T) \geq 0.2$ V [31]. The device acts as a voltage controlled current source or transconductor as long as $V_{DS} \geq V_{DS,sat}$ with $V_{DS,sat} = (V_{GS} - V_T)/\alpha$ so that it operates in saturation. Typical values for α are between 1 and 1.5 and a good estimate for $V_{DS,sat}$ at the edge of strong inversion is about 0.15 V [31]. A MOS transistor can also be operated in its weak inversion region for $(V_{GS} - V_T) \leq -0.05$ V or -0.1 V. This offers very high transconductance/current efficiencies and low power operation but the bandwidth is limited. The minimal V_{DS} to maintain the device in saturation is now about $4kT/q$ to $5kT/q$, or 0.1 to 0.125 V [31]. So, in any region of operation, we need to maintain a drain-source bias of about 0.15 V. It is im-

portant to remark that this requirement is *independent of the threshold voltage V_T of the device*. On the other hand, the gate-source bias for the device V_{GS} is $V_T + (V_{GS} - V_T)$ and is thus strongly dependent on the V_T of the devices as well as the region of operation.

Challenges at 0.5 V The most basic way to achieve amplification with a MOS transistor is the common source configuration with an active load¹ as shown in Fig. 2. The required input (gate) bias is $V_T + (V_{GS} - V_T)$ and the optimal output (drain) bias is $V_{DD}/2$ for an output swing of $V_{DD} - 2V_{DS,sat}$. At 0.5 V V_{DD} two limitations can occur: the output bias is typically smaller than the input bias²; and, the input swing is very limited. Clearly, it becomes very difficult to design circuits with large input and output swings. However, as long as a sufficiently large gain exists between input and output, this is not a strong limitation.

With a 0.5 V supply, it is very difficult to use a common drain configuration (Fig. 2). The output can swing sufficiently but since there is no gain between the input and the output, the input bias and signal swing would require voltage levels above the supply.

In a common gate configuration the input signal, output signal and $3V_{DS,sat}$ are stacked; even if we assume a large voltage gain for the stage, the available output swing is too small for most applications. A common gate stage (or folded cascode) can be embedded in an amplifier if followed by sufficient gain so that no significant swings are needed at the common gate output. Similarly, cascode topologies with all devices in saturation³ are excluded at 0.5 V since they require a stack of the output swing and $4V_{DS,sat}$ (about 0.6 V).

Of the basic transistor configurations only the common source configuration has the potential to operate at supply voltages of 0.5 V. It is again important to remark that this limitation stems from the required $V_{DS,sat}$ of about 0.15 V and is independent of the value of V_T .

Feedback and virtual grounds The input and output bias level differences can be accommodated by using feedback topologies which keep the amplifiers inputs at virtual ground and allow for a level-shift between the output and the virtual ground notes as illustrated in Fig. 3 [5, 23]. Similar level shifts can be accommodated in switched capacitor circuits.

¹We assume the active load is implemented with a single transistor biased as a current source.

²Only if V_T is very small or negative, or if the $(V_{GS} - V_T)$ is kept very small – which implies the device goes into weak inversion, – equal input and output bias can be achieved.

³To get the full benefit of a cascode topology we need to use cascode devices in the signal device and active load device, resulting in a stack of 4 devices.

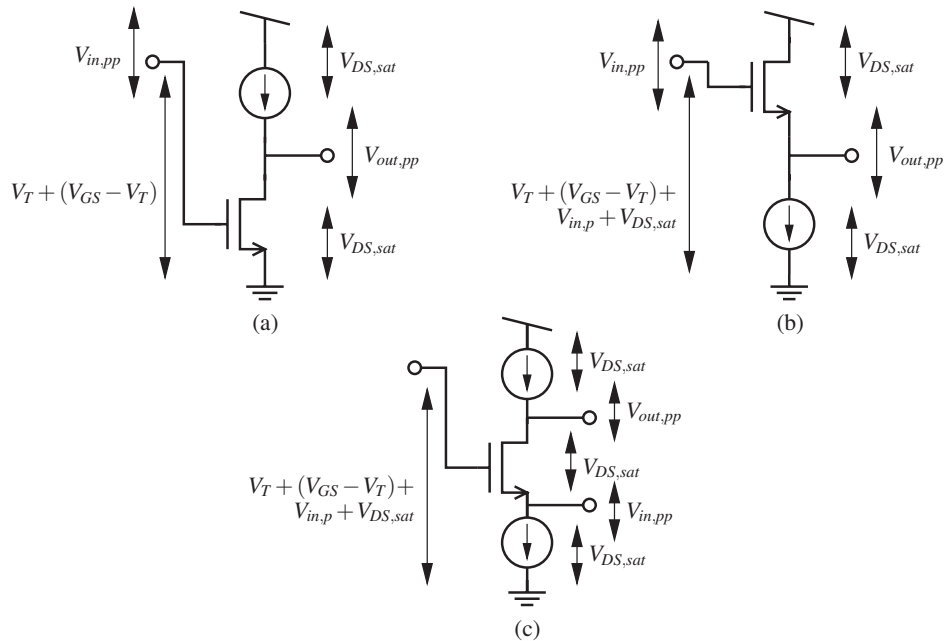


Figure 2: Voltage ranges in (a) common source, (b) common drain, and (c) common gate configurations.

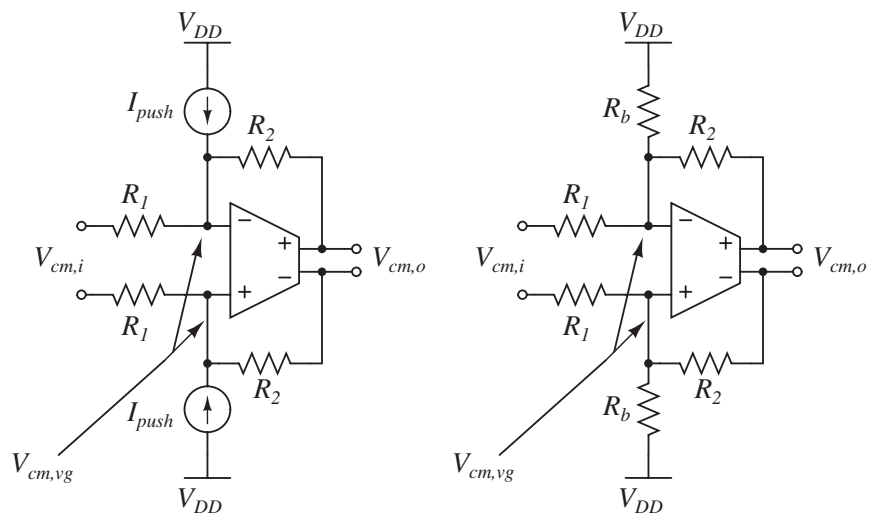


Figure 3: The injection of DC currents into the virtual ground nodes allows a level shift between the virtual ground and amplifier outputs. As long as the loop gain is large, the signal swing at the virtual ground remains very small and large output swings can be achieved [5, 23].

Opportunities at 0.5 V: the body terminal Operation from a small supply voltage (≤ 0.5 V) offers the advantage that the risk of turning on any of the parasitic bipolar devices in the circuit is largely eliminated, provided that supply transient overvoltages are adequately kept under control. This enables the use of *forward biasing* for the body-source junction which results in a reduction of the threshold voltage V_T [32–35]. Traditionally only the body terminal of pMOS devices could be accessed in n-well processes, but modern MOS processes offer the availability of nMOS devices in a separate well so that their body terminal can be biased independently.

Forward body bias has been used in digital applications to tune the V_T so that a more consistent circuit performance over process and temperature and thus a higher yield is obtained [32–34, 36]. Interestingly, a Low-Voltage-Swapped-Body-Bias (LVSB) design style has been proposed [37] where the body of the nMOS is tied to the positive supply and the body of the pMOS is tied to the negative supply. High speed or low power consumption is obtained and correct functionality for an operating temperature up to 75°C has been demonstrated [37]. As mentioned, forward body bias also allows to adjust and reduce the threshold voltage of the device. In [38, 39] e.g., we typically use a forward bias of 0.25 V which results in a reduction of the V_T by 50 mV for a standard device in a $0.18\ \mu\text{m}$ CMOS technology.

The availability of the body terminal thus offers two opportunities. The signal can be applied to the body (back-gate) of the device [16–18, 38] whereas the gate is used to bias the device; or, when we apply the signal to the gate, we can use the body (back-gate) to control the bias of the device [39]. Both techniques will be illustrated in the subsequent sections.

Bipolar devices The built-in potential of silicon PN junction is about 0.7 V which excludes bipolar devices for true low voltage circuits at 0.5 V.

3 Fully Differential Operational Transconductance Amplifiers

Fully differential circuits [40, 41] are standard in contemporary analog integrated circuits due to their large signal swing and better supply and substrate interference robustness. At 0.5 V, we have to rely on those properties even more and fully differential topologies are a must. It is important to point out that the correct operation of differential topologies relies on the availability of good common-mode rejection; not only needs the differential-mode gain to be significantly larger than the common-mode gain, but also the common-mode

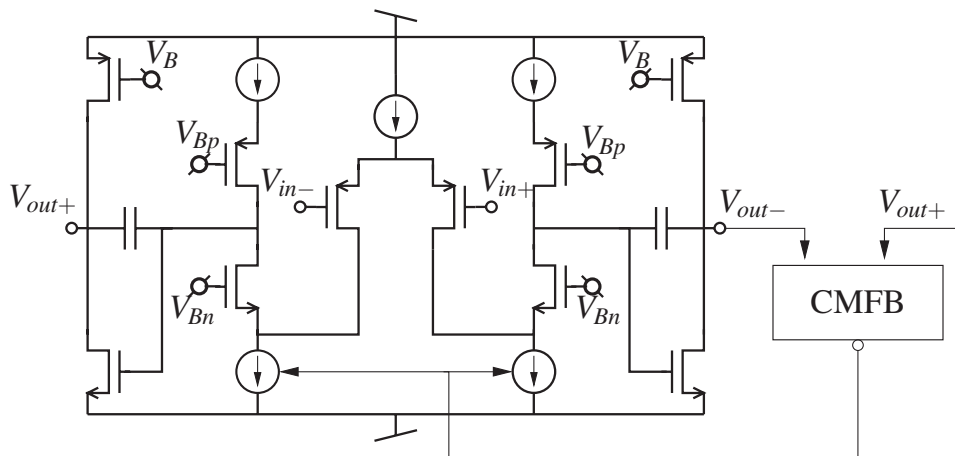


Figure 4: *Folded cascode operational transconductance amplifier*

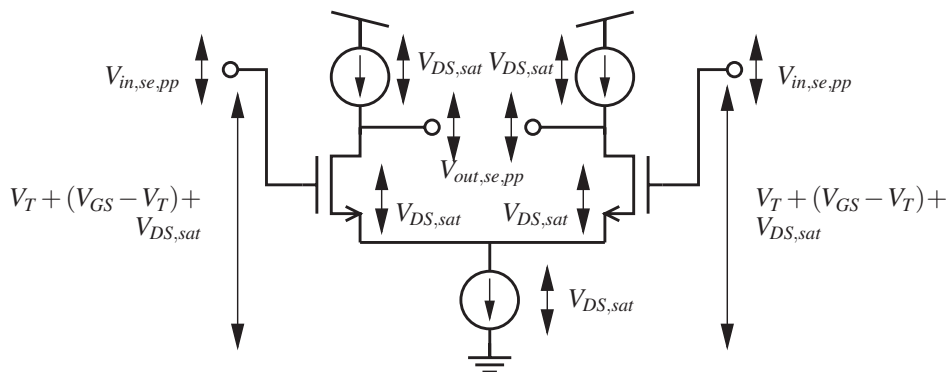


Figure 5: *Bias and signal ranges in a differential pair.*

gain needs to be sufficiently smaller than 1 in the presence of positive feedback loops in the common-mode signal path.

Two stage, folded cascode transconductance amplifiers, shown in Fig. 4, are often used for low voltage applications (see e.g., [6, 23, 42]). The differential pair is the standard input structure for an operational (transconductance) amplifier. For a differential input signal, the differential signal current is proportional to the g_m of the input pair; for a common-mode input signal, the common-mode output current is determined by the conductance of the tail current source and is thus very small; the small response to common-mode signals in combination with a wide-band common-mode feedback (CMFB) provide a small common-mode gain and strong common-mode rejection [40, 41].

The signal ranges and biasing in a differential pair are illustrated in Fig. 5.

Due to the stacking of three devices, its output swing is very limited; however, by adding a second gain stage after the input stage this limitation can be overcome as shown in Fig. 4. The main challenge is the required input bias of $V_T + (V_{GS} - V_T) + V_{DS,sat}$; supposing the inputs can be biased at 0.5 V, the resulting maximum allowed V_T is only 0.15 V for strong inversion operation. Even when such a low V_T is available, in practice the inputs of the OTA will need to be about 0.15 V below the supply – see, e.g., Fig. 3 – so that strong inversion operation of this stage becomes impractical with a 0.5 V supply. So, for any technology where V_T is larger than 0.15 V there is a need to develop differential input structures with good common-mode rejection.

The design of wide-band common-mode feedback loops is also very challenging at 0.5 V. The output common-mode is set at 0.25 V ($V_{DD}/2$) for maximum swing so that it is very difficult to develop a wide-band error amplifier. We will discuss local common-mode feedback as an alternative solution in subsequent sections.

Telescopic amplifiers [40,41] are also widely used for analog integrated circuits thanks to their relative simplicity and intrinsic high operation speed. High gain is achieved in these configurations by using (folded) cascode topologies and is further enhanced with gain boosting. None of these topologies can be easily used at 0.5 V due to required device stacks and limited output swing. At 0.5 V we have to rely on multi-stage topologies to achieve sufficient gain. Due to the unavailability of the common drain stage the realization of a low output impedance required for the implementation of an operational amplifier also becomes very difficult and we are limited to operational transconductance amplifiers. For most on-chip applications the loads are capacitive or the load impedances can be kept sufficiently large that this is not a very significant limitation, especially in feedback circuits where the loop gain reduces the effective output impedance.

In the subsequent paragraphs two OTA designs will be introduced that can operate from a 0.5 V supply. We will also briefly discuss the application of such OTAs in a larger analog signal processing function.

3.1 Body-input OTA

Single-ended body-input operational amplifiers have been investigated for low voltage applications down to 0.7 V [16], [18], [24], [25], [26]. At supply voltages of 0.5 V or below, there is low risk for latchup in the circuit (assuming that supply transient overvoltages are limited) and the signals can be connected to the body node of the MOS devices without restrictions. For an input common mode at $V_{DD}/2$ (0.25 V), a small forward bias for the body-source junction is also introduced; this lowers the V_T and further increases the inversion level. Operation near the weak-moderate inversion boundary is preferred, in order to

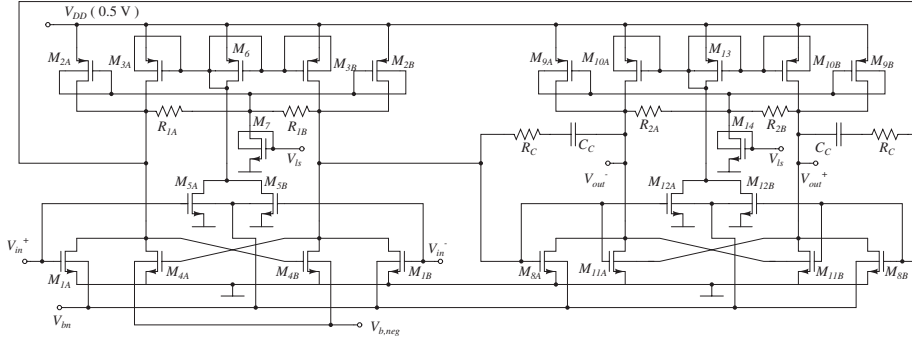


Figure 9: Two-stage fully differential operational amplifier with Miller compensation

DC small signal differential gain is:

$$A_{dd} = \frac{V_{out}^+ - V_{out}^-}{V_{in}^+ - V_{in}^-} = \frac{g_{m1}}{g_{ds1} + g_{ds2} + g_{ds3} + g_{ds4} + 1/R - g_{m4}} \quad (3)$$

The common-mode small signal gain is given by:

$$A_{cc} = \frac{V_{out}^+ + V_{out}^-}{V_{in}^+ + V_{in}^-} = -\frac{g_{m1} - g_{m5} \frac{g_{m3} + g_{mb3}}{g_{m6} + g_{mb6}}}{g_{m4} + g_{m2} + g_{mb2}} \quad (4)$$

If the W/L ratios of M_1, M_3, M_5, M_6 are such that $(\frac{W}{L})_1 / (\frac{W}{L})_5 = (\frac{W}{L})_3 / (\frac{W}{L})_6$, then the common-mode gain will be zero. In this design, we made $(\frac{W}{L})_3 / (\frac{W}{L})_6 = 0.5 \cdot (\frac{W}{L})_1 / (\frac{W}{L})_5$ so that we get 6 dB of rejection through the common-mode feed-forward path. Overall each stage has a differential gain A_{dd} of 25 dB and a common-mode gain A_{cc} of -10 dB.

To obtain a DC gain of about 50 dB two gain stages are cascaded to form a two stage operational transconductance amplifier as shown in Fig. 9. The input stage's output common-mode bias of 0.4 V guarantees that the input devices of the second stage are correctly biased. The output common-mode voltage of the second stage is set to 0.25 V by decreasing the DC drop across R_{2A} and R_{2B} to 0.15 V. Similarly to what is done in the input stage, a negative resistance implemented by pair M_{11A} and M_{11B} is used in the second stage, only its body terminal can operate from the low common voltage at the output and its body transconductance is used to provide a negative conductance; its gate transconductance is put in parallel with the input transconductance.

The operational amplifier is stabilized through the Miller capacitors C_C across the second stage. The gain-bandwidth product is approximately $g_{m1} / (2\pi C_C)$

where g_{m_1} is the gate transconductance of M_{IA} and M_{IB} and the second pole frequency of the amplifier is at $g_{m_8}/(2\pi C_L)$ where g_{m_8} is the gate transconductance of M_{8A} and M_{8B} and C_L is the output load capacitance. The series resistor R_C moves the zero introduced by C_C from the right-half-plane to the left-half-plane.

A prototype gate-input OTA was designed and fabricated in a 0.18 μm triple-well CMOS process [39]. The OTA has a DC small signal gain of 55 dB, a nominal unity gain bandwidth of 15 MHz and a phase margin of 60° for a load of 10 pF; it consumes 80 μW from a 0.5 V supply.

4 Bias circuits

Robust biasing strategies are crucial to obtain circuits with a consistent performance over process, temperature and supply voltage variations. At 0.5 V, the need for robust biasing only becomes more pressing, since any shifts in the bias points can significantly affect the signal swings (and thus the dynamic range) or the device operation points. To illustrate some biasing techniques for ultra-low voltages we will review the automatic on-chip generation of the bias voltages and current for the gate-input OTA in Fig. 9 [39].

Error amplifier Typical bias loops make extensive use of active feedback loops in combination with replica circuits. These loops require an error amplifier to servo the bias voltages or currents to their desired value. Most of the 0.5 V amplifier design techniques presented so far rely on the bias circuits, and thus the error amplifier design can not rely on them.

A carefully sized inverter can be configured as an ultra-low voltage inverting error amplifier that compares the input voltage to its input transition voltage and amplifies the difference. With a 0.5 V supply and an input transition of 0.25 V the amplifier's devices operate in weak inversion. The transition voltage of the error amplifiers can be adjusted by controlling the bodies of the nMOS devices. This is done automatically by using a reference error amplifier in a negative feedback arrangement as shown in Fig. 10. The amplifier consists of three identical stages as shown. The input transition voltage is set to $V_{DD}/2$ independent of variations in process and temperature as follows: if the input transition voltage of "ErrorAmpA" is smaller than V_{test} , the input voltage of "ErrorAmpB" decreases, the output voltage V_{inv} of "ErrorAmpC" decreases, the biasing of the nMOS devices is reduced, and the input transition voltage increases. Similarly, when the input transition voltage is larger than V_{test} , the feedback will react and decrease the input transition voltage. The feedback loop accurately sets the input transition voltage to the $V_{test} = V_{DD}/2$, or 0.25 V in this case. The feedback loop is stabilized with C_C and a zero-canceling series resistor R_C .

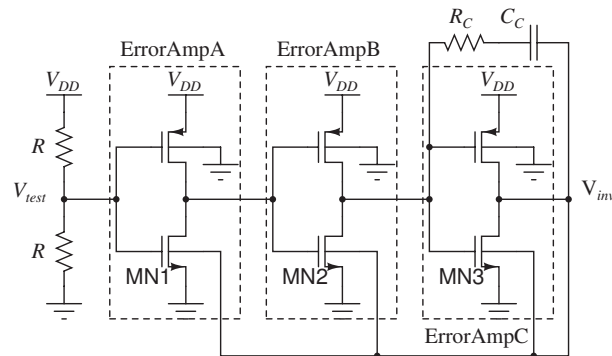


Figure 10: Reference error amplifier biasing loop to fix the input transition voltage to $V_{DD}/2$.

An error amplifier with its body biased from V_{inv} in Fig. 10 is now functionally equivalent to an inverting differential amplifier with its non-inverting terminal tied to a 0.25 V reference. The use of such error amplifiers is described in the following subsections.

Generation of a level shift voltage The gate-input OTA in Fig. 9 requires an accurate level shift voltage to bias the pMOS load transistors compared to the output common-mode level. This level shift is set with the bias current I_{ls} through a resistor (see Fig. 9). A current source is designed using a single nMOS device. To increase the inversion level of this device, its bias voltage is applied both through the gate and the body. A replica of this current source is used in the biasing circuit as shown in Fig. 11. The error amplifier servoes the current in the current source so that a voltage drop of $V_{DD}/2$ is established across a replica resistor. The voltage developed by this circuit, V_{ls} , is connected to M7 and M12 in Fig. 9, as shown there. By appropriate sizing the resistors and current sources in the amplifiers the desired level-shifting voltage drops can now be generated in the OTA.

DC output common mode control The bias voltage V_{bn} in Fig. 9 adjusts the biasing level of the nMOS input devices in relation to the pMOS load devices and allows to control the output common-mode voltage of the OTA. The biasing loop shown in Fig. 12 adjusts the V_{bn} of a replica of one amplifier stage with an input common-mode voltage of 0.4 V so that its the DC output common-mode is 0.25 V. The generated V_{bn} is then used in all gate-input OTAs on the same chip. Note that the error amplifier based loop is only used to set the DC value of the output common-mode and is not part of the common-mode feedback which

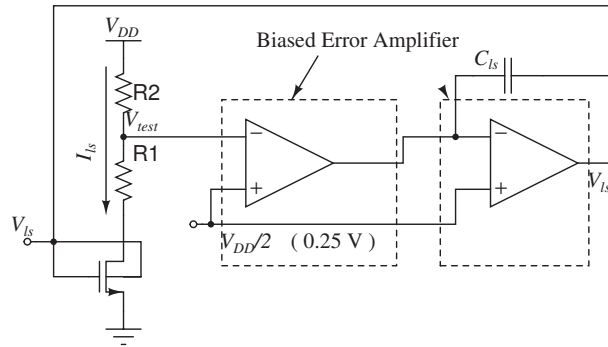


Figure 11: *Generation of level-shift voltages*

is provided locally in each amplifier.

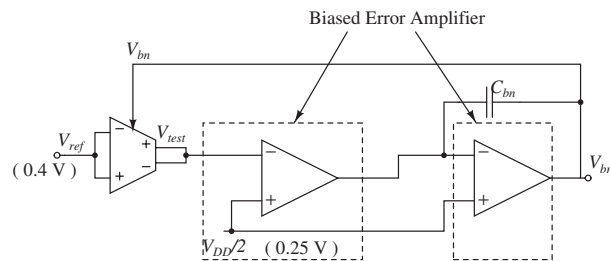


Figure 12: *Biasing bodies of input nMOS devices and setting the DC output common-mode.*

5 Application example: 0.5 V active filter

We now discuss an example application of 0.5 V gate-input OTAs in the design of a fully integrated active R-C filter [39]. In order to operate an active filter from a 0.5 V power supply we have to address the additional challenge of filter tuning.

The available tuning options are limited at 0.5 V. Standard devices cannot be used as switches in the signal path so that switching capacitors and resistors in or out the signal path (see e.g., [44]) is not possible. MOSFET-C techniques [45,46] using standard devices would require tuning voltages above the supply voltage to accommodate sufficient signal swing and linearity. The design of linear, tunable transconductors at 0.5 V is also very challenging.

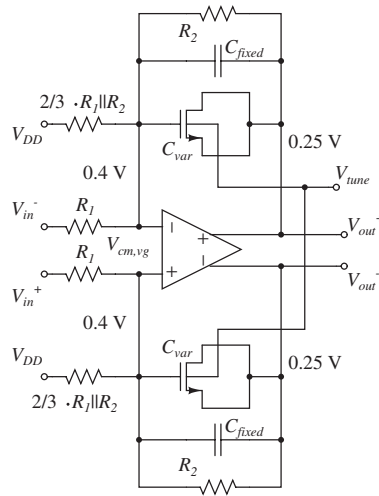


Figure 13: *Tunable 0.5 V integrator*

In [39] a varactor-R filter structure is presented where the filter capacitors are partly replaced with variable MOS capacitors (varactors) as shown in the lossy integrator schematic in Fig. 13. The gate of the varactors is at 0.4 V, the source and drain are connected to each other and are at 0.25 V. The body voltage is now biased to control the inversion in the device. The gate-source capacitance can be varied widely by moving the device from weak inversion towards strong inversion by changing the tuning voltage applied to the body. The varactor capacitance per unit area is low, and this capacitance is voltage dependent, but these drawbacks are mitigated by the fact that only a small varactor is used in parallel with a larger fixed capacitor.

Using this tunable integrator, a 5th-order low-pass elliptic filter with a 135 kHz cut-off frequency operating from 0.5 V has been demonstrated [39] based on a frequency-scaled version of the design in [45]. The complete filter schematic is shown in Fig. 14. Note that level shifting resistors are used to set the correct common-mode level at the inputs of the OTAs; since the OTA input nodes are kept at virtual short due to the loop gain in the feedback, the level shift resistors do not influence the filter characteristics.

A prototype was designed using standard devices with a $|V_T|$ of 0.5 V in a 0.18 μm CMOS technology and also included an on-chip PLL operating from 0.5 V to generate the filter tuning voltages, and on-chip biasing circuits to generate the bias voltages and currents for the gate-input OTAs.

The filter was extensively characterized (see [39]). Fig. 15 shows the measured frequency response for the filter operating using power supply voltages from 0.45 V to 0.60 V which clearly demonstrates the robust operation of the

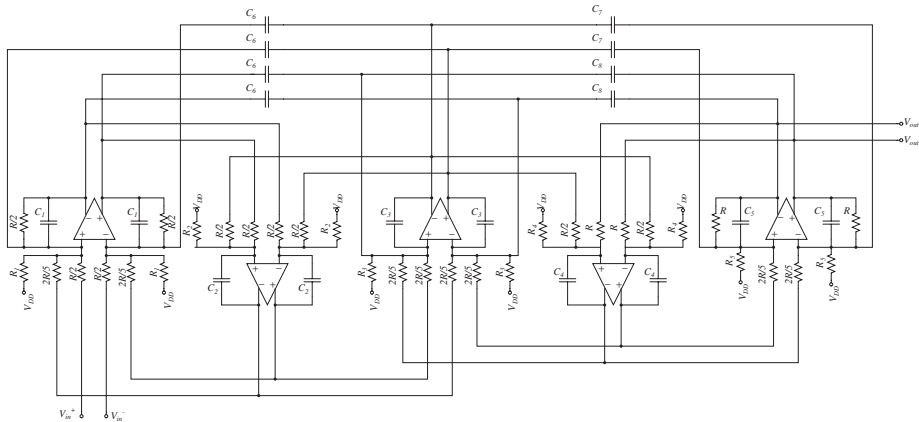


Figure 14: *Low-voltage 5th order low-pass elliptic filter*

filter and its biasing and tuning loops. The simulated and measured overall dynamic range – ratio of input amplitude at which there is 1% THD to the input referred noise – is about 57 dB when operating from 0.5 V supply with a current consumption of 2.16 mA. Correct functionality was verified for ambient temperatures from 5°C to 75°C .

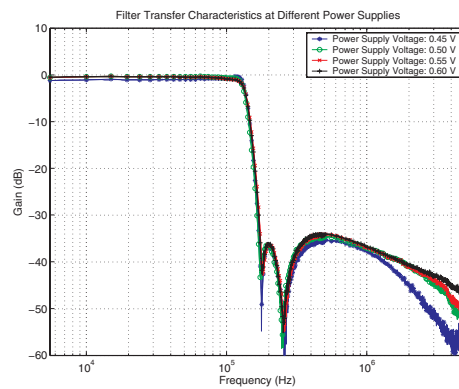


Figure 15: *Measured filter transfer characteristics at different power supply voltages with the on-chip tuning PLL active.*

6 Conclusions

In this paper we reviewed the challenges of designing analog circuits operat-

ing with a power supply of 0.5 V. We introduced two different fully differential OTA topologies that maintain good characteristics while being able to operate at 0.5 V using standard MOS devices with a $|V_T|$ of 0.5 V in a 0.18 μm CMOS technology. The use of such OTAs to build analog signal processing circuits was illustrated with the design of a 0.5 V 5th order fully integrated active filter. This work demonstrates that analog circuits can be designed to operate from a 0.5 V supply, even when the device nominal threshold is as high as the supply voltage itself. It can thus be expected that, with lower-threshold devices, 0.5 V analog circuits can attain even better performance.

Acknowledgment

The authors would like to thank Analog Devices, Inc. for funding part of this work.

References

- [1] "The International Technology Roadmap for Semiconductors (2001 edition)," see <http://public.itrs.net>.
- [2] E. Vittoz, "Future of analog in the VLSI environment," in *Proceedings ISCAS*, May 1990, pp. 1372–1375.
- [3] P. Kinget, "Implications of mismatch on analog VLSI," in *Analog VLSI Integration of Parallel Signal Processing Systems*. Ph.D. Thesis, K.U.Leuven (Belgium), May 1996, ch. 3, pp. 19–74.
- [4] P. Kinget and M. Steyaert, "Impact of transistor mismatch on the speed-accuracy-power trade-off of analog CMOS circuits," in *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, May 1996, pp. 333–336.
- [5] K. Bult, "Analog design in deep sub-micron CMOS," in *Proceedings European Solid-State Circuits Conference (ESSCIRC)*, September 2000, pp. 11–17.
- [6] Q. Huang, "Low voltage and low power aspects of data converter design," in *Proceedings European Solid-State Circuits Conference (ESSCIRC)*, September 2004, pp. 29–35.
- [7] B. Hosticka, W. Brockherde, D. Hammerschmidt, and R. Kokozinski, "Low-voltage CMOS analog circuits," *IEEE Trans. Circuits Syst. I*, vol. 42, no. 11, pp. 864–872, November 1995.

- [8] M. Steyaert, V. Peluso, J. Bastos, P. Kinget, and W. Sansen, "Custom analog low power design: the problem of low voltage and mismatch," in *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, May 1997, pp. 285–292.
- [9] W. Sansen, M. Steyaert, V. Peluso, and E. Peeters, "Toward sub 1 V analog integrated circuits in submicron standard CMOS technologies," in *Digest of Technical Papers IEEE International Solid-State Circuits Conference (ISSCC)*, February 1998, pp. 186–187.
- [10] J. Fattaruso, "Low-voltage analog CMOS circuit techniques," in *International Symposium on VLSI Technology, Systems, and Applications*, June 1999, pp. 286–289.
- [11] T. Ohguro, H. Naruse, H. Sugaya, E. Morifuji, S. Nakamura, T. Yoshitomi, T. Morimoto, H. Kimijima, S. Momose, Y. Katsumata, and H. Iwai, "An 0.18- μm CMOS for mixed digital and analog applications with zero-volt- V_{th} epitaxial-channel MOSFETs," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1378–1383, 1999.
- [12] S. Bazarjani and W. Snelgrove, "Low voltage SC circuit design with low- V_{t} MOSFETs," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, 1995, pp. 1021–1024.
- [13] T. Cho and P. Gray, "A 10b, 20Msample/s, 35mW pipeline A/D converter," *IEEE J. Solid-State Circuits*, vol. 30, no. 4, pp. 166–172, March 1995.
- [14] U. Moon, G. Temes, E. Bidari, M. Keskin, L. Wu, J. Steensgaard, and F. Maloberti, "Switched-capacitor circuit techniques in submicron low-voltage CMOS," in *IEEE International Conference on VLSI and CAD (ICVC '99)*, 1999, pp. 349–358.
- [15] J. Crols and M. Steyaert, "Switched-opamp: an approach to realize full CMOS switched-capacitor circuits at very low power supply voltages," *IEEE J. Solid-State Circuits*, vol. 29, no. 8, pp. 936–942, August 1994.
- [16] A. Guzinski, M. Bialko, and J. Matheau, "Body driven differential amplifier for application in continuous-time active-C filter," *Proceedings of ECCD*, pp. 315–319, 1987.
- [17] F. Dielacher, J. Houtmann, and J. Resinger, "A software programmable CMOS telephone circuit," *IEEE J. Solid-State Circuits*, vol. 26, no. 7, pp. 1015–1026, July 1991.

- [18] B. Blalock, P. Allen, and G. Rincon-Mora, "Designing 1-V op amps using standard digital CMOS technology," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 7, pp. 769–780, July 1998.
- [19] J. Huijsing and D. Linebarfer, "Low-voltage operational amplifier with rail-to-rail input and output ranges," *IEEE J. Solid-State Circuits*, vol. SC-20, no. 6, Dec. 1985.
- [20] R. Hogervorst, J. Tero, R. Eschauzier, and J. Huijsing, "A compact power-efficient 3V CMOS rail-to-rail input/output operational amplifier for VLSI cell libraries," *IEEE J. Solid-State Circuits*, vol. 29, no. 12, pp. 1505–1513, December 1994.
- [21] R. Eschauzier, L. Kerklaan, and J. Huijsing, "A 100-MHz 100-dB operational amplifier with multipath nested Miller compensation structure," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1709–1717, December 1992.
- [22] S. Pernici, G. Nicollini, and R. Castello, "A CMOS low-distortion fully differential power amplifier with double nested Miller compensation," *IEEE J. Solid-State Circuits*, vol. 28, no. 7, pp. 758–763, July 1993.
- [23] S. Karthikeyan, S. Mortezapour, A. Tammineedi, and E. Lee, "Low-voltage analog circuit design based on biased inverting opamp configuration," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 3, pp. 176–184, March 2000.
- [24] K. Lasanen, E. Raisanen-Ruotsalainen, and J. Kostamovaara, "A 1-V 5 μ W CMOS-opamp with bulk-driven input transistors," in *Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems*, 2000, pp. 1038–1041.
- [25] T. Lehmann and M. Cassia, "1-V power supply CMOS cascode amplifier," *IEEE J. Solid-State Circuits*, vol. 36, no. 7, pp. 1082–1086, July 2001.
- [26] T. Stockstad and H. Yoshizawa, "A 0.9-V 0.5- μ A rail-to-rail CMOS operational amplifier," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 286–292, 2002.
- [27] V. Peluso, P. Vancorenland, A. M. Marques, M. Steyaert, and W. Sansen, "A 900-mV low-power $\Delta\Sigma$ A/D converter with 77-dB dynamic range," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 1887–1897, Dec. 1998.

- [28] D. Chang, G. Ahn, and U. Moon, "Sub-1-V design techniques for high-linearity multistage/pipelined analog-to-digital converters," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 1, pp. 1–12, January 2005.
- [29] G. Ahn, D. Chang, M. Brown, N. Ozaki, H. Youra, K. Yamamura, K. Hamashita, K. Takasuka, G. Temes, and U. Moon, "A 0.6 V 82 dB $\Delta\Sigma$ audio ADC using switched-RC integrators," in *Digest of Technical Papers IEEE International Solid-State Circuits Conference (ISSCC)*, February 2005, pp. 166–167.
- [30] J. Sauerbrey, D. Schmitt-Landsiedel, and R. Thewes, "A 0.5-V 1- μ W successive approximation ADC," *IEEE J. Solid-State Circuits*, vol. 38, pp. 1261–1265, July 2003.
- [31] Y. Tsvividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. Oxford University Press, 2003.
- [32] T. Kobayashi and T. Sakurai, "Self-adjusting threshold-voltage scheme (SATS) for low-voltage high-speed operation," in *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, May 1994, pp. 271–274.
- [33] V. R. Kaenel, M. D. Pardoen, E. Dijkstra, and E. A. Vittoz, "Automatic adjustment of threshold and supply voltages for minimum power consumption in CMOS digital circuits," *IEEE Symposium on Low Power Electronics*, pp. 78–79, 1994.
- [34] M.-J. Chen, J.-S. Ho, T.-H. Huang, C.-H. Yang, Y.-N. Jou, and T. Wu, "Back-gate forward bias method for low-voltage CMOS digital circuits," *IEEE Trans. Electron Devices*, vol. 43, no. 6, pp. 904–910, June 1996.
- [35] A. L. Coban, P. E. Allen, and X. Shi, "Low-voltage analog IC design in CMOS technology," *IEEE Trans. Circuits Syst. I*, vol. 42, no. 11, pp. 955–958, November 1995.
- [36] J. W. Tschanz, J. T. Kao, S. G. Narendra, R. Nair, D. A. Antoniadis, A. P. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, pp. 1396–1402, November 2002.
- [37] S. Narendra, J. Tschanz, J. Hofsheier, B. Bloechel, S. Vangal, Y. Hoskote, S. Tang, D. Somasekhar, A. Keshavarzi, V. Erraguntla, G. Dermer, N. Borkar, S. Borkar, and V. De, "Ultra-low voltage circuits and processor

- in 180nm to 90nm technologies with a swapped-body biasing technique,” in *IEEE J. Solid-State Circuits*, February 2004, pp. 156–157.
- [38] S. Chatterjee, Y. Tsvividis, and P. Kinget, “A 0.5-V bulk-input fully differential operational transconductance amplifier,” in *Proceedings European Solid-State Circuits Conference (ESSCIRC)*, September 2004, pp. 147–150.
- [39] ———, “A 0.5-V filter with PLL-based tuning in 0.18 μ m CMOS,” in *Digest of Technical Papers IEEE International Solid-State Circuits Conference (ISSCC)*, February 2005, pp. 506–507.
- [40] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. New York, NY (USA): John Wiley & Sons, 2001.
- [41] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York, NY (USA): Mc Graw Hill, 2001.
- [42] M. Banu, J. Khoury, and Y. Tsvividis, “Fully balanced operational amplifiers with accurate output balancing,” *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1410–1414, December 1988.
- [43] F. Rezzi, A. Baschiroto, and R. Castello, “A 3 V 12-55 MHz BiCMOS pseudo-differential continuous-time filter,” *IEEE Trans. Circuits Syst. I*, vol. 42, no. 11, pp. 896–903, November 1995.
- [44] H. Huang and E. Lee, “Design of low-voltage CMOS continuous-time filter with on-chip automatic tuning,” *IEEE J. Solid-State Circuits*, vol. 36, no. 8, pp. 1168–1177, August 2001.
- [45] M. Banu and Y. Tsvividis, “An elliptic continuous-time CMOS filter with on-chip automatic tuning,” *IEEE J. Solid-State Circuits*, vol. SC-20, no. 6, pp. 1114–1121, Dec. 1985.
- [46] Y. Tsvividis, M. Banu, and J. Khoury, “Continuous-time MOSFET-C filters in VLSI,” *IEEE Trans. Circuits Syst.*, vol. 33, no. 2, pp. 125–139, February 1986.

LIMITS ON ADC POWER DISSIPATION

Boris Murmann
Center for Integrated Systems
Stanford University
Stanford, CA 94305-4070
murmam@stanford.edu

Abstract

The need for low power A/D conversion in a large number of applications has fueled a trend towards ever-improving ADC power efficiency. This article investigates practical limits to this development by analyzing the minimum power needed in the constituent building blocks of today's ADCs. A comparison with state-of-the-art experimental data shows that future improvements in power efficiency may be limited to less than one order of magnitude, unless future work strives to depart from traditional paradigms on a circuit and system level.

1. Introduction

With the recent shift towards digital information processing, low power A/D conversion has evolved as a key requirement in many electronic systems. Especially in portable applications, restrictions on the available power or energy tend to dictate a stringent upper bound for the maximum affordable energy per A/D conversion.

While feature size scaling has enabled the possibility of implementing extremely fast ADCs in standard CMOS technology [1], the resulting power dissipation at these technology limits is often prohibitively high. In modern applications, where the power budget is typically only a fraction of a Watt, power efficiency rather than technology speed upper bounds ADC throughput.

Over the past decade, tremendous progress has been made in reducing ADC power dissipation. Hence, one is tempted to ask: Are we approaching "fundamental" limits? How much more improvement can we hope for? These questions are difficult to answer with great precision, but this article attempts to provide a feel for what is possible in today's technology.

In the following discussion, we will limit ourselves to the analysis of popular ADC topologies in terms of their constituent building blocks; namely gain stages, integrators, and preamplifier-latch circuits in CMOS technology. As a

result, the derived results will allow us to speculate only about the limits of evolutionary progress, and must neglect the impact of fundamentally different and potential disruptive approaches, as for instance photonic conversion [2]. Following this introduction, Section 2 will review some of the previously stated fundamental limits pertaining to the most primitive analog signal processing elements. After a comparison with state-of-the art data, Section 3 adds practical and technological aspects that lead to an improved basis for reasoning about realistic limits. Sections 4 and 5 comment the observed results, while Section 6 outlines promising future directions.

2. Fundamental Limits

Fundamental limits for power dissipation in basic analog circuit blocks have been stated in a number of publications [3-9]. In the following discussion, we invoke some of these previous results, while limiting ourselves to voltage mode processing. Furthermore, we assume that only capacitors are available for energy storage, neglecting the fact that inductors may in some cases help lower the power dissipation of a particular function [10, 11].

First, consider an amplifier that drives a single capacitor with a rail-to-rail continuous time sinusoid of frequency f_{sig} . Assuming that the relevant thermal noise contribution is given by the total integrated noise

$$P_n = \frac{k_B T}{C}, \quad (1)$$

it follows that the minimum power dissipation of the driving amplifier is

$$P = M \cdot k_B T \cdot SNR \cdot f_{sig}. \quad (2)$$

In the above expressions, k_B is the Boltzmann constant, T is the absolute temperature, SNR is signal-to-noise ratio, and M is a constant that depends on amplifier topology. A minimum value of $M=8$ corresponds to ideal class-B operation [3]. For a class-A amplifier, it follows that $M=8\pi$ [6]. The additional factor of π stems from the class-B/class-A peak efficiency ratio, given by $(\pi/4)/(1/4)$.

While continuous time signal processing has recently gained popularity in sigma-delta ADCs, the majority of converters are still based on switched capacitor circuits. Hence, we now derive similar expressions for the case in which the amplifier processes a sampled data signal, rather than a pure sinusoid. For simplicity, we focus on the popular case of class-A amplification. Fig. 1 illustrates the general setup for this derivation. The amplifier input is driven by a sampled data representation of a sine wave sampled at Nyquist frequency ($f_s=2 \cdot f_{sig}$) and peak amplitude. In the following paragraphs, we consider two different settling characteristics.

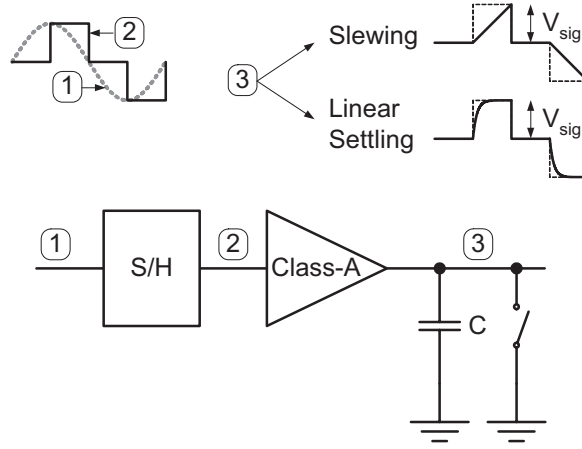


Fig. 1. Class-A sampled data processor.

First, assume that the amplifier output slews from its reset state to the signal value (V_{sig}) within $1/2$ sampling period. The minimum quiescent current in the driving amplifier is then given by

$$I_{bias} = C \cdot \frac{dV}{dt} = 4 \cdot C \cdot V_{sig} \cdot f_{sig}. \quad (3)$$

Assuming that the entire supply voltage is used for signal swing, the power dissipation becomes

$$P = 2 \cdot V_{sig} \cdot I_{bias}. \quad (4)$$

With

$$SNR = \frac{0.5 \cdot V_{sig}^2}{k_B T / C}, \quad (5)$$

we can now combine the above equations to yield

$$P = 16 \cdot k_B T \cdot SNR \cdot f_{sig}. \quad (6)$$

Consider now a second case in which the amplifier does not slew, and settles with a purely linear response of the form $(1 - e^{-t/\tau})$. In this case, the minimum amplifier bias current is set by the value needed in the initial transient (largest dV/dt). By differentiating, we find

$$I_{bias} = C \cdot V_{sig} \cdot \frac{1}{\tau}, \quad (7)$$

where τ is the settling time constant of the amplifier. Assuming that N time constant are needed to achieve the desired settling precision, it follows that

$$\tau = \frac{1}{4 \cdot f_{sig}} \cdot \frac{1}{N}, \quad (8)$$

Since the amplifier output is typically required to settle within a very small fractional error of V_{sig} , (5) remains approximately valid and we can combine the above equations to obtain

$$P \cong 16 \cdot N \cdot k_B T \cdot SNR \cdot f_{sig} \quad (9)$$

An estimate for the minimum value of N can be found by considering that the circuit must usually settle to within at least 1/2 LSB precision at a certain bit resolution B . Neglecting all other error contributors, and with the crude assumption that $B \cong [SNR(dB) - 1.76] / 6.02$, it follows that

$$N \cong -\ln\left(\frac{1}{2} \frac{2}{2^B}\right) \cong \ln\left(2^{\lceil SNR(dB) - 1.76 \rceil / 6.02}\right) \cong \frac{SNR(dB)}{9} \quad (10)$$

For instance, at $SNR=60$ dB, $N \cong 6.7$; and at $SNR=90$ dB, $N \cong 10$.

The above result indicates that purely linear settling is significantly less power efficient than the (fictitious) case of complete slewing. Many practical sampled data circuits slew initially, and then transition into linear settling. In this case, the power dissipation lies somewhere between the values given by (6) and (9).

It is interesting to compare the above power limits to actual ADC data. Fig. 2 shows the derived asymptotes together with experimental data reported at the IEEE International Solid-State Circuits Conference (ISSCC). In this data set, bandpass-type ADCs were included using the available conversion bandwidth in place of f_{sig} . Whenever SNR was not specified in the experimental results, $SNDR$ (signal-to-noise and distortion ratio) was used as a crude replacement.

As a first observation from Fig. 1, one should note that the power efficiency of the experimental parts varies by more than two orders of magnitude at almost any given SNR level. This spread can be attributed to several factors. First, the power efficiency of ADCs has been improving over time, which can be seen by the gap between data fits based on consecutive four-year periods (1998-2001 and 2002-2005). Secondly, it is clear that not all reported ADCs were optimized for low power, and more importantly, they all differ in robustness and functionality. For instance, oversampled bandpass ADCs tend to expend additional power at the benefit of reduced anti-aliasing filter requirements and integrated downconversion.

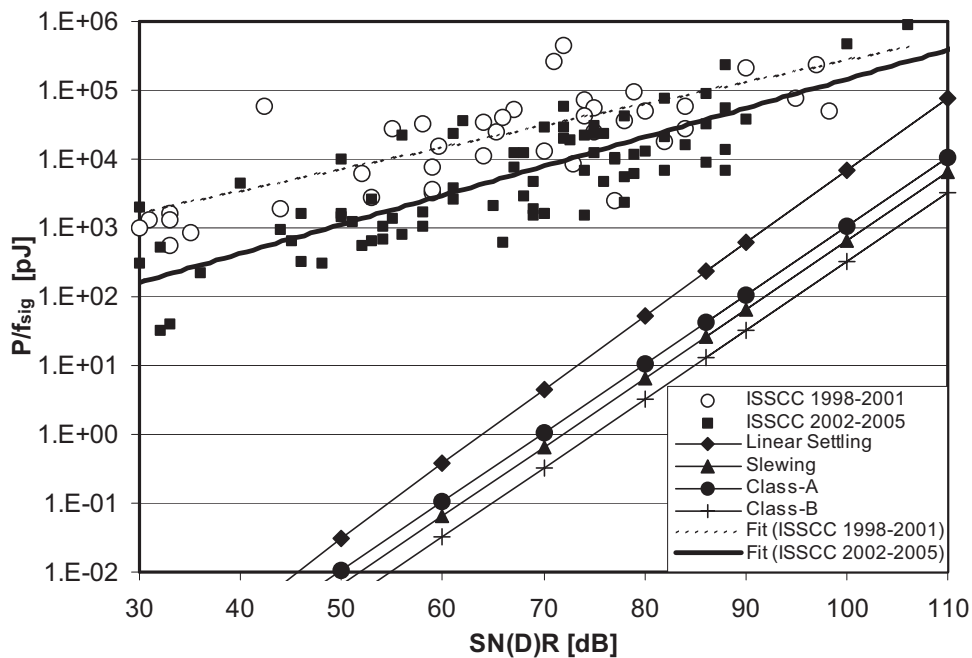


Fig. 2. Experimental ADC data and fundamental asymptotes for analog circuits containing a single amplifier and capacitor.

Despite the shortcomings of the above single-metric survey, it is clear and not very surprising that today's ADCs exceed the fundamental power limits by orders of magnitude. After all, these asymptotes hold only for single capacitor circuits and do not include unavoidable architectural and circuit overhead that is needed to realize a functional ADC in today's technology.

3. Architectural Considerations

In this section, we now augment the above results in order to derive an improved basis for reasoning about more realistic and ADC specific power limits. To accomplish this, we include several additional factors such as architectural complexity, excess noise, and efficiency in voltage and current usage. Clearly, such an analysis can only be carried out with specific ADC realizations in mind. Therefore, the final results should not be viewed as globally fundamental, but rather as an answer to the question: "How much further can we improve ADC power based on evolutionary progress, without changing the underlying circuit principles?" To begin, we loosely partition the *SNR* axis into three regions: High, medium and low *SNR*.

$$\alpha = \frac{2 \cdot V_{sig}}{V_{DD}}. \quad (12)$$

The excess noise factor n_f can be identified by computing the total output referred noise in the redistribution phase (ϕ_2)

$$P_n = P_{n,\phi_1} + P_{n,\phi_2} = 2 \frac{kT}{C} + \frac{kT}{C} \gamma \cdot \frac{1}{\beta} \cdot \left(1 + \frac{g_{m2}}{g_{m1}} \right) = n_f \cdot \frac{kT}{C}, \quad (13)$$

where γ is the MOS devices' white noise factor [12]. The last term in (11) arises from biasing considerations, which we will consider next.

Using (7) and the fact that this circuit's time constant is

$$\tau = \frac{1}{\beta} \frac{C}{g_{m1}}, \quad (14)$$

it follows that settling with a purely linear response requires

$$\frac{g_{m1}}{I_{bias}} \leq \frac{1}{\beta \cdot V_{sig}}. \quad (15)$$

If g_{m1}/I_{bias} is chosen smaller than the right hand value of (15), the circuit still settles linearly, but the power dissipation increases with the last term in (11), which is then larger than one. It is interesting to note that in this case, the power dissipation is no longer independent of the available signal swing. If g_{m1}/I_{bias} is increased beyond a value that satisfies (15), the circuit will begin to slew, and the power dissipation is ultimately bounded by a certain slewing limit, similar to that discussed in Section 2.

In practice, the choice of g_{m1}/I_{bias} varies between designs, and is a strong function of speed objectives. To proceed, we simply assume that (15) holds with equality. Depending on the particular design, this may be either optimistic or pessimistic with respect to the power dissipation predicted by (11).

Fig. 6 shows the resulting plot of (11) (labeled as "SC Stage"), with the following parameter choices: $\gamma=1$, $\beta=0.5$ ($C_g=0$, i.e. infinite device f_T), $g_{m1}=g_{m2}$ ($n_f=6$) and $\alpha=2/3$. The plot also contains limit lines for other SNR regions and topologies, which we will discuss next.

3.2. Medium SNR

ADCs in the medium SNR range have been implemented using a wide variety of architectures and circuit topologies. For simplicity, we focus here on two popular, and in some sense complementary approaches: Switched-capacitor pipeline ADCs and continuous time sigma-delta modulators.

Fig. 4 shows a simplified, conceptual block diagram of a tapered pipeline ADC, emphasizing only the required signal path amplifier stages and the position of

the comparators. For simplicity, we assume that each stage resolves one bit, and thus requires an amplifier gain of two. In practice, the amplification is accomplished using a switched capacitor stage similar to that of Fig. 3.

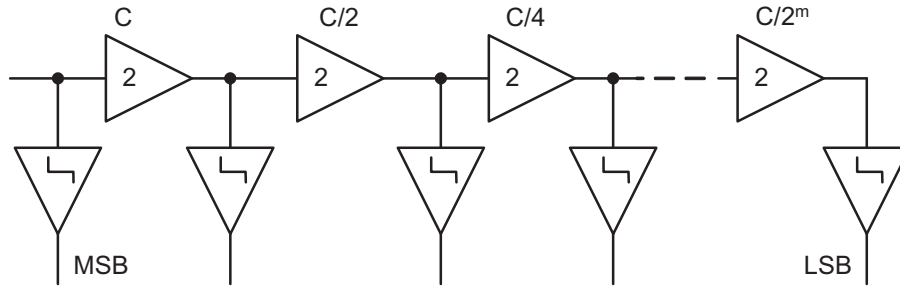


Fig. 4. Ideally tapered pipeline ADC signal path.

As shown in [13], the power-optimum tapering factor for the sampling capacitances in each stage is approximately equal to the stage gain. Hence, in an ideally tapered pipeline that corresponds to our example, the first stage's sampling capacitance C is reduced by a factor of two in each stage, down to $C/2^m$, where m is the total number of amplifiers used. In this ideal case, the first stage consumes about half of the total power and contributes half of the total input referred noise. Therefore, an ideally tapered pipeline ADC will consume approximately four times the power of a single switched capacitor gain stage designed for the same SNR .

Unfortunately, ideal stage tapering is hard to achieve in practice. Consider for instance a 14-bit design (13 amplifier stages) with a typical value of $C=5\text{pF}$. In this case, $C/128$ corresponds to only 40fF. At such levels, the physical dimensions of the capacitors become very small, and the actual stage input capacitance tends to be dominated by unavoidable parasitics from switches, comparators and wires. As a result, stage scaling terminates at a certain capacitance level in most practical designs.

Table 1 below constructs a numerical example on how this effect may impact the power dissipation of a pipeline ADC. As a starting point, consider a 14-bit design (13 amplifiers). If capacitance scaling discontinues at $C/128$, this means that the last two stages will carry additional, unwanted capacitance. Assuming that power dissipation is proportional to capacitance, this translates into a small power penalty (relative power is 4.06 compared to 4 in ideal case). The penalty worsens in designs with lower resolution, since a larger fraction of backend stages carry unwanted capacitance.

Fig. 6 shows a limit line for pipeline ADC power dissipation using the relative power numbers at the bottom of Table 1 as a multiplier for the single stage switched capacitor power dissipation (Section 3.1).

Table. 1. Pipeline ADC capacitor scaling example.

Number of Amplifiers	13	12	11	10
Stage Capacitances	1	1/4	1/16	1/64
	1/2	1/8	1/32	1/128
	1/4	1/16	1/64	1/128
	1/8	1/32	1/128	1/128
	1/16	1/64	1/128	1/128
	1/32	1/128	1/128	1/128
	1/64	1/128	1/128	1/128
	1/128	1/128	1/128	
	1/128	1/128		
	1/128			
ΣC	2.03	0.54	0.17	0.086
C_{single}	1/2	1/8	1/32	1/128
Relative Power Pipeline/Single SC Stage ($\Sigma C/C_{\text{single}}$)	4.06	4.32	5.44	11.01

Note again that this end result is fairly crude since it avoids a large number of technicalities. For instance, we neglected to account for the decrease in settling requirements in LSB stages and also the relative cost for a gain of two (instead of one, as assumed in section 3.1). Nevertheless, the key point here is to observe that the steep slope of 4x increase in power per bit ($\sim 6\text{dB}$ increase in SNR) does not hold at moderate resolution.

Alternative to pipeline ADCs, the implementation of medium SNR ADCs using continuous time sigma-delta architectures is gaining popularity. In these converters, power dissipation is dominated by the continuous time integrators, which are often implemented using active-RC or Gm-C stages.

In the following analysis, we will look at the power dissipation of a Gm-C integrator as an example. Fig. 5 shows a conceptual implementation using a simple MOS differential pair. Assuming that the input of this circuit is not driven beyond the weakly nonlinear range of its transfer function, we can express the differential output current available to the capacitive load as

$$\frac{i_{od}}{I_{bias}} \cong \frac{v_{id}}{V_{ov}} - \frac{1}{8} \left(\frac{v_{id}}{V_{ov}} \right)^3, \quad (16)$$

where $V_{ov} = V_{GS} - V_t$ is the quiescent point gate overdrive of the differential pair transistors, which are assumed to be ideal square law devices.

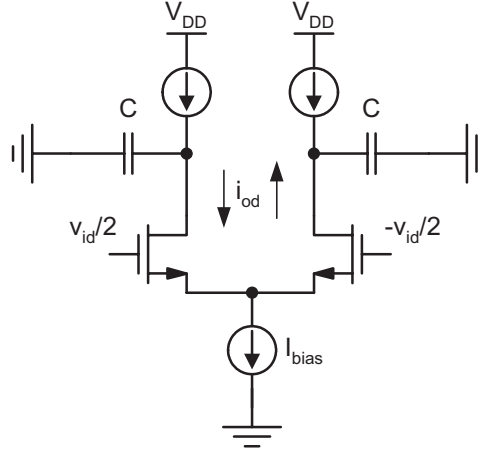


Fig. 5. Conceptual schematic of a Gm-C integrator.

In order to keep nonlinear distortion below a certain limit, one must restrict the differential input below a small fraction of the quiescent point gate overdrive. For instance, considering third order intermodulation (IM_3), it can be shown that

$$IM_3 \cong \frac{3}{32} \left(\frac{v_{id,max}}{V_{ov}} \right)^2. \quad (17)$$

Assuming that the implementation is constrained by IM_3 , it is now interesting to investigate how much of the bias current (I_{bias}) can be steered into the capacitive load. For this purpose, we introduce the current steering efficiency

$$\eta_{cur} = \frac{i_{od,max}}{I_{bias}}. \quad (18)$$

By approximating (16) with its linear term, and using (17) and (18), it then follows that.

$$\eta_{cur} \cong \frac{v_{id,max}}{V_{ov}} = \sqrt{\frac{32}{3} IM_3}. \quad (19)$$

For instance, in order to achieve $IM_3=60\text{dB}$, $\eta_{cur} \cong 0.1$. This means that a maximum of only 10% of the invested bias current can be steered into the load. It is straightforward to show that this argument still holds in a resistively degenerated differential pair. Degeneration increases the useable input voltage range, but also reduces transconductance, and therefore does not alter the current steering efficiency.

Factoring this penalty into the fundamental class-A limit of (2) (with $M=8\pi$), we can now find the minimum power dissipation of our primitive Gm-C integrator:

$$P = 8\pi \cdot n_f \cdot \frac{1}{\alpha} \cdot k_B T \cdot SNR \cdot f_{sig} \cdot \frac{I}{\sqrt{\frac{32}{3} IM_3}}, \quad (20)$$

where α is defined by (12). In the above result, it is assumed that the relevant noise contribution is the total integrated noise given by (1). It can be argued that this result is pessimistic, because a certain amount of noise is typically filtered out by subsequent circuitry [7]. Since this effect is hard to quantify, mostly because we would also need to include the additional power spent for filtering, we conclude this discussion by regarding (20) as a reasonable bound for approximate arguments. In Fig. 6, (20) is plotted for $n_f=2$, $\alpha=2/3$ and assuming $IM_3=SNR$ (Line labeled “Gm-C Stage”).

3.3. Low SNR

As discussed in [5], the power dissipation imposed by thermal noise is usually orders of magnitude below that dictated by matching and/or minimum realizable capacitance. To see this, consider e.g. $SNR=30\text{dB}$ (~ 5 -bit resolution). Using (5), with $V_{sig}=0.5\text{V}$ as an example, yields $C=16\text{aF}$, which is an unrealizable capacitance from a practical perspective.

To derive an applicable power bound for the low SNR region, we consider the example of a matching-limited, flash-type ADC. As shown in [5], the minimum power dissipation of a matching limited circuit is given by

$$P = 24 \cdot C_{ox} \cdot A_{VT}^2 \cdot f_{sig} \cdot \left(\frac{V_{sig,rms}}{3 \cdot \sigma_{Vos}} \right)^2, \quad (21)$$

where C_{ox} and A_{VT} are technology constants, and σ_{Vos} is the standard deviation of the accuracy limiting offset voltage. In this result, it is assumed that all the power is spent driving matching limited gate capacitance using a class-B amplifier. We now expand this result as follows:

- Assume a B -bit, flash-like structure, which necessitates 2^B matching limited components
- Set $3 \cdot \sigma_{Vos}$ equal to 1/2 LSB of the converter
- Assume class-A operation (additional factor of π)
- Include dynamic power dissipation (E_{dyn} =dynamic energy per clock cycle (one half of signal cycle) and component)

With these modifications, (21) becomes

$$P = \left(12\pi \cdot \frac{1}{\alpha} \cdot C_{ox} \cdot A_{VT}^2 \cdot 2^{3B} + 2 \cdot E_{dyn} \cdot 2^B \right) \cdot f_{sig}. \quad (22)$$

The above expression is plotted in Fig. 6 with $B=[SNR(\text{dB})-1.76]/6.02$ and $\alpha=2/3$ (Curve labeled “Flash”). All other parameters are chosen based on typical 0.13 μm technology data: $C_{ox}=15\text{fF}/\mu\text{m}^2$ and $A_{VT}=3\text{mV}\mu\text{m}$ [14]. A typical logic gate in 0.13 μm technology consumes about 6fJ per cycle [15]. With the assumption of a minimum digital complexity on the order of 10 gates needed for latching, buffering and decoding the final flash output, a value of $E_{dyn}=60\text{fJ}$ is used.

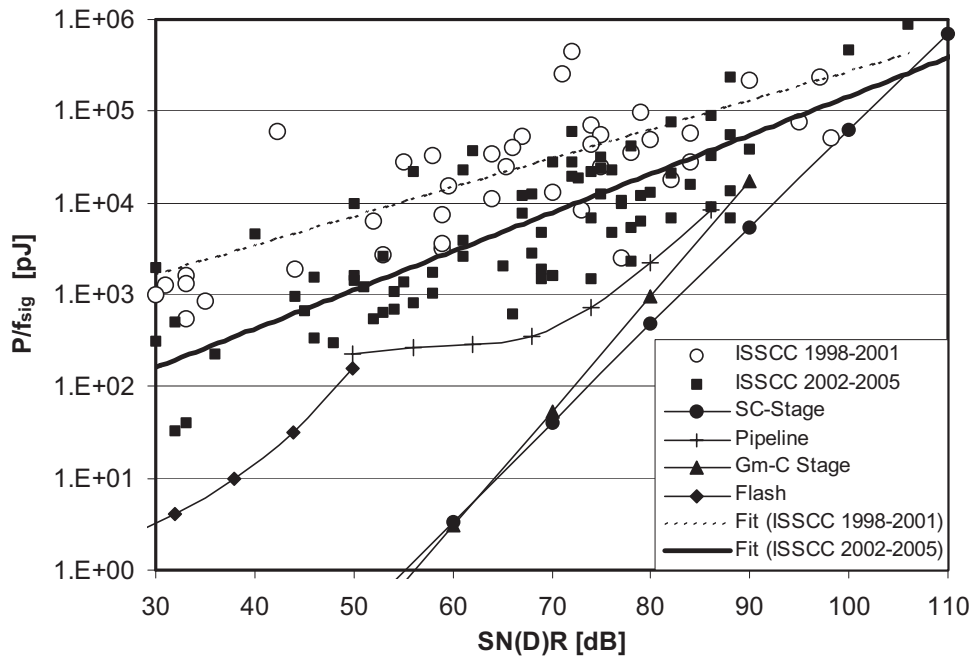


Fig. 6. Experimental ADC data and practical asymptotes for ADC power dissipation.

4. Comments and Interpretation

The above derivations are based on a number of crude assumptions, some of which are pessimistic, but not unreasonable for comparisons on a log-scale. For simplicity, a large number of additional practical factors were not considered. Most designs consume extra power e.g. for biasing, reference generation, clocking and front-end S/H circuitry. Furthermore, their SNR performance can be impaired by quantization noise and differential nonlinearity (low and medium SNR). In addition, a significant amount of power overhead is usually needed to construct high-gain multistage amplifiers that help ensure good precision, linearity and low drift.

Despite these missing factors and approximations used, the main observation from Fig. 6 is obvious: Today's ADCs, optimized for low power, have come very close to practical power limits imposed by their architectures and underlying circuit topologies.

In terms of technology, the result of Fig. 6 is based on only a few assumptions and typical parameter values for 0.13 μm CMOS designs (fractional swing, minimum capacitance, matching coefficient, dynamic energy). Further scaling in feature size will hold mixed blessings for different *SNR* objectives and topologies. As discussed in [16], it is likely that power will decrease only in circuits that are limited by matching (low *SNR*). This prediction is based on the assumption that matching coefficients (e.g. A_{VT}) will continue to scale down proportional to oxide thickness.

At moderate and high *SNR*, the implications of reduced supply voltages are often viewed as a detrimental factor, leading to the common prediction that power will increase with feature size reduction in noise limited designs (e.g. [17]). In reality, we haven't seen such an increase for several reasons. First, the skills and creativity of designers usually lead to inherently more power efficient, optimized circuits with each process generation. Secondly, it is often the case that the increased device speed of new technologies can be traded in for power savings. Higher transit frequency allows biasing MOS devices in moderate or even weak inversion, where the transconductor efficiency g_m/I_D is significantly larger than in strong inversion. This tradeoff is implicitly factored into equation (11). As long as g_m/I_D can increase proportional to the reduction in supply voltage, the power dissipation of noise limited circuits is unaffected by technology scaling. This situation applies to ADCs that are scaled in feature size while keeping their throughput constant, e.g. at video rate.

5. Figure of Merit Considerations

Several Figures of Merit (FOM) have been introduced to compare the performance of ADC designs using a single number. The most popular figure of merit is given by [18, 19]

$$FOM_1 = \frac{P}{2^{ENOB} \cdot f_{sig}}, \quad (23)$$

where $ENOB = (SNDR[\text{dB}] - 1.76) / 6.02$ is the converter's effective number of bits. This metric is based on the purely empirical observation that a fit to ADC data across all performance regions tends to show an approximate 2x power increase per bit. For instance, the bold trend line in Fig. 2 (ISSCC 2002-2005) has a slope of 1.8x per 6-dB increase in *SNR*.

An alternative FOM that is tied to fundamental tradeoffs is given by [16]

$$FOM_2 = \frac{P}{2^{2ENOB} \cdot f_{sig}}, \quad (24)$$

which suggest a power increase of 4x per bit added. This slope corresponds to that of the fundamental limit lines in Fig. 2, and is also in line with several local slopes of the projected technology limits (Fig. 6). For instance, the slope of the “Pipeline” curve is approximately 4x/6dB at the high *SNR* end.

From the discussion presented in this paper, it is clear that figures of merit must be used with care. For instance, it is unreasonable to compare particular designs with very different *SNR* objectives using any single number figure of merit. This is simply because in practice, the tradeoffs between resolution and power are far more complex than those implied by (23) or (24).

One way to address this problem is to introduce more complex figures of merit that contain architecture and technology parameters [16, 20]. Such FOMs, however, tend to be subjective, and would also need to change with the introduction of new technologies. As a simple alternative, it is often suitable to compare P/f_{sig} of particular designs with approximately the same *SNR* or *SNDR*. In cases where the tradeoff slopes are obvious (e.g. fully noise limited design), a FOM in the form of (24) may be suitable to account for the merit of *SNR* within a small region of comparison (e.g. *SNR*=80...90dB).

As seen from Fig. 2, the metric P/f_{sig} is also well suited for general surveys on state-of-the art (as opposed to comparing a small set of particular designs), since it allows visualizing the rates of progress across the various *SNR* regions.

6. Future Opportunities and Conclusion

Unless we manage to depart from the current route of evolutionary progress, mostly driven by refinements in device technology and pure circuit optimization, it is unlikely that ADC power can improve by more than an additional order of magnitude in the future. While this outlook seems bleak, it is also clear that there exists a wealth of alternative directions and unrealized opportunities on all levels of ADC design.

Today, the majority of applications view ADCs as static “black boxes” that deliver precisely linear transfer functions with fixed speed and *SNR*, determined by peak performance requirements. Especially in radio receivers, large *average* power savings are possible when ADC resolution and speed are dynamically adjusted to satisfy the minimum instantaneous performance needs. Furthermore, it is conceivable to embed ADCs within a system as weakly non-linear, but digitally correctable “one-to-one mappers” rather than perfect quantizers. For instance, within a communications system, it is possible to “equalize” the ADC together with the communication channel itself [21, 22].

On a circuit level, the potential benefit of a “digitally assisted” approach for A/D conversion has long been recognized. While a single A/D conversion with

$SNR=60\text{dB}$ consumes about 1.5nJ (Typical $P/f_{sig} = 2 \cdot P/f_s \cong 3\text{nJ}$, see Fig. 2), the same energy can be used to toggle roughly 250,000 logic gates in $0.13\mu\text{m}$ CMOS technology (6fJ/gate). Fig. 7 shows a graph where this calculation has been generalized for various SNR levels and digital feature sizes (L). This chart is based on typical ADC energy per conversion numbers that follow from the bold fit line of Fig. 2 (ISSCC 2002-2005), and assuming that digital switching energy scales approximately with $L \cdot V_{DD}^2$ (V_{DD} values taken from CMOS scaling roadmap). Especially at high SNR , several tens of thousand gates can be considered as “free” in terms of energy overhead, and can be used for digital calibration and postprocessing.

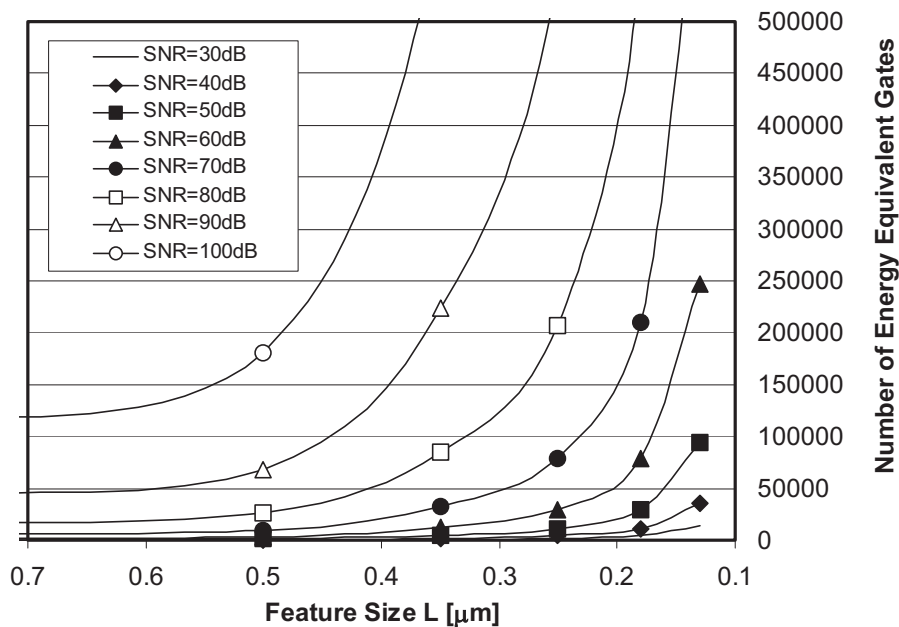


Fig. 7. Number of energy equivalent logic gates as a function of digital feature size and ADC SNR. Example: A single conversion using an ADC with 60dB SNR consumes as much energy as switching $\sim 250,000$ gates in $0.13\mu\text{m}$ technology.

As shown in [23], digital post-processing capabilities can be leveraged to replace precision amplifiers in a pipeline ADC with simple, low power open-loop stages. It is foreseeable that more such advanced digital compensation techniques will be developed in the future. By digitally eliminating all forms of static and dynamic nonlinearity errors in ADCs, it may ultimately be possible to approach class-B efficiency limits.

Overall, future improvements in ADC power dissipation are likely to come from a combination of the above aspects: Improved system embedding and reducing analog circuit complexity and precision to the bare minimum.

References

- [1] K. Poulton *et al.*, "A 20 GS/s 8b ADC with a 1MB Memory in 0.18 μ m CMOS," *ISSCC Dig. Techn. Papers*, pp. 318-319, Feb. 2003.
- [2] G. C. Valley *et al.*, "Photonic Analog-to-Digital Converters: Fundamental and Practical Limits," *SPIE: Integrated Optical Devices, Nanostructures, and Displays*, pp. 96-106, Jan. 2004.
- [3] E. A. Vittoz, "Future of Analog in the VLSI Environment," *Proc. ISCAS*, pp. 1372-1375, May 1990.
- [4] R. Castello and P. Gray, "Performance Limitations in Switched-Capacitor Filters," *IEEE Trans. Ckt. Syst.*, pp. 865-876, Sept. 1985.
- [5] P. Kinget and M. Steyaert, "Impact of transistor mismatch on the speed-accuracy-power trade-off of analog CMOS circuits," *Proc. CICC*, pp. 333-336, 1996.
- [6] A.-J. Annema, "Analog circuit performance and process scaling," *IEEE Trans. Ckts. Syst. II*, pp. 711-725, June 1999.
- [7] A.-J. Annema *et al.*, "Analog Circuits in Ultra-Deep-Submicron CMOS," *IEEE J. of Solid-State Circuits*, pp. 132-143, Jan. 2005.
- [8] S. Rabbii and B. A. Wooley, *The Design of Low-Voltage, Low-Power Sigma-Delta Modulators*: Kluwer Academic Publishers, 1998.
- [9] K. Uyttenhove and M. S. J. Steyaert, "Speed-Power-Accuracy Tradeoff in High-Speed CMOS ADCs," *IEEE Trans. Ckts. Syst. II*, pp. 280-287, Apr. 2002.
- [10] R. Schreier *et al.*, "A 10-300-MHz IF-digitizing IC with 90-105-dB dynamic range and 15-333-kHz bandwidth," *IEEE J. of Solid-State Circuits*, pp. 1636-44, Dec. 2002.
- [11] W. B. Kuhn *et al.*, "Dynamic range performance of on-chip RF bandpass filters," *IEEE Trans. Ckt. Syst. II*, pp. 685-694, Oct. 2003.
- [12] A. J. Scholten *et al.*, "Noise modeling for RF CMOS circuit simulation," *IEEE Trans. Electron Devices*, pp. 618-632, March 2003.
- [13] D. W. Cline and P. R. Gray, "A power optimized 13-b 5 MSamples/s pipelined analog-to-digital converter in 1.2 μ m CMOS," *IEEE J. Solid-State Circuits*, pp. 294-303, March 1996.
- [14] C. H. Diaz *et al.*, "CMOS technology for MS/RF SoC," *IEEE Trans. Electron Devices*, pp. 557-566, March 2003.
- [15] UMC, "Virtual Silicon 0.13 μ m Library," www.umc.com.
- [16] K. Bult, "The Effect of Technology Scaling on Power Dissipation in Analog Circuits," *Proc. 14th Workshop on Advances in Analog Circuit Design (AACD)*, pp. 211-250, April 2005.
- [17] K. Bult, "Broadband communication circuits in pure digital deep sub-micron CMOS," *ISSCC Dig. Techn. Papers*, pp. 76-77, Feb. 1999.

- [18] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, pp. 539-50, April 1999.
- [19] F. Goodenough, "Analog technology of all varieties dominate ISSCC," *Electronic Design*, pp. 96, Feb. 19, 1996.
- [20] M. Vogels and G. Gielen, "Architectural selection of A/D converters," *Proc. Design Automation Conference*, pp. 974-977, 2003.
- [21] W. Namgoong, "A channelized digital ultrawideband receiver," *IEEE Trans. Wireless Communications*, pp. 502-510, March 2003.
- [22] Y. Oh and B. Murmann, "System Embedded ADC Calibration for OFDM Receivers," to be published.
- [23] B. Murmann and B. E. Boser, *Digitally Assisted Pipeline ADCs*: Kluwer Academic Publishers, 2004.

Ultra Low-Power Low-Voltage Analog Integrated Filter Design

Wouter A. Serdijn Sandro A.P. Haddad and Jader A. De Lima

Abstract

Filtering is an indispensable elementary signal processing function in many electronic systems. In many critical applications, e.g., in portable, wearable, implantable and injectable devices, one should maximize the dynamic range and, at the same time, minimize the power consumption of the filter. This joint optimization can take place in different phases, the filter transfer function design phase, the filter topology design phase, and the filter circuit design phase.

In the filter transfer function design phase, the filter's functional input-output relation is mapped on a suitable filter transfer function. Two approximation techniques are introduced: the Padé approximation and the L_2 approximation. The Padé approximation is employed to approximate the Laplace transform of the desired filter transfer function by a suitable rational function around a selected point. The L_2 approximation offers a more global approximation, i.e., not concentrating on one particular point, and has the advantage that it can be applied in the time domain as well as in the Laplace domain.

In the filter topology design phase, the filter transfer function is mapped on a suitable filter topology. For this, the filter transfer function is written in the form of a state-space description, which subsequently is optimized for dynamic range, sparsity and sensitivity. In the determination and optimization of the dynamic range the filter's controllability and observability gramians play an important role. Dynamic range optimization boils down to transforming the controllability gramian such that it becomes a diagonal matrix with equal diagonal entries, transforming the observability gramian such that it also becomes a diagonal matrix, and capacitance distribution. To improve the state-space matrices' sparsity the dynamic-range optimized matrices can be transformed into a form that describes an orthonormal ladder filter. After applying capacitance distribution, a filter topology is found that is not too complex and has a dynamic range that is close (i.e., within a few dBs) to optimal.

Finally, in the filter circuit design phase, the filter topology is mapped on a circuit. A classification of integrators is presented. Falling in the category of transconductance-capacitance (gm-C) integrators, a novel nA/V CMOS transconductor for ultra-low power low-frequency gm-C filters is introduced. Its input transistors are kept in the triode-region to benefit from the lowest g_m/I_D ratio. The g_m is adjusted by a well defined (W/L) and V_{DS} , the latter a replica of the tuning voltage V_{TUNE} . The resulting design complies with $V_{DD}=1.5V$ and a $0.35\mu m$ CMOS process. Its transconductance ranges from $1.1nA/V$ to $5.5nA/V$ for $10mV \leq V_{TUNE} \leq 50mV$.

Wouter A. Serdijn and Sandro A.P. Haddad are with the Electronics Research Laboratory, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, the Netherlands, +31 (0)15 27-81715, serdijn@ieee.org, <http://elca.et.tudelft.nl/~wout>

Jader A. De Lima is with Freescale Semiconductor, Inc., Brazil

To illustrate the entire filter design procedure, a dynamic translinear Morlet filter is designed. Simulations and measurements demonstrate an excellent approximation of the Morlet wavelet base. The circuit operates from a 1.2-V supply and a bias current of $1.2\mu\text{A}$.

Index Terms

Filters, Integrators, Analog Integrated Circuits, Low Voltage, Low Power, Dynamic Translinear, Log-Domain, Gm-C, State Space Optimization, Dynamic Range, Sensitivity, Sparsity

I. INTRODUCTION

FILTERING is an indispensable elementary signal processing function in many electronic systems. Filters are either used for *selection*, i.e., to separate desired signals from other signals and noise by making use of their differences in energy-frequency spectra, or for *shaping*, i.e., to change the energy-frequency spectrum of a single, desired signal. In practice, a piece of electronic apparatus that does not contain at least one rudimentary filter can hardly be found.

Traditionally, filters operated in the continuous-time domain and have been designed as resistively terminated lossless discrete inductor-capacitor (LC) filters. When we wish to realize the filter on chip, however, often, at least for most sub-gigahertz applications, this implies giving up the use of inductors. The Laplace transform of filter transfer functions that can be realized with capacitive and resistive elements only have real poles in the left half of the complex Laplace plane, while often transfer functions with complex poles are called for. These are only realizable if active circuits are added.

With the introduction of active circuits in filters, resulting in active filters, two fundamental problems are introduced. First, unlike passive reactances, active elements produce noise and distortion. For this reason, active filters are bound to exhibit a limited dynamic range, defined as the ratio of the largest and the smallest signal level that the filter can handle. Second, unlike passive reactances, active elements dissipate energy. Thus power has to be supplied. In many critical applications, e.g., in portable, wearable, implantable and injectable devices, one should maximize the dynamic range and, at the same time, minimize the power consumption of the filter. This joint optimization can take place in different phases:

1. the filter transfer function design phase,
2. the filter topology design phase, and
3. the filter circuit design phase.

In the first phase, the *filter transfer function design phase*, the filter's functional input-output relation is mapped on a suitable filter transfer function, whose Laplace transform can be described by a strictly proper rational function of low order. For

obvious reasons only the implementation of a causal stable filter is feasible, meaning that it will have a proper rational transfer function that has all its poles in the complex left half plane and the degree of the numerator polynomial does not exceed the denominator degree.

In Section II, two approximation techniques will be introduced: the *Padé approximation* and the *L_2 approximation*. The Padé approximation is employed to approximate the Laplace transform of the desired filter transfer function by a suitable rational function around a selected point. The L_2 approximation offers a more global approximation, i.e., not concentrating on one particular point, and has the advantage that it can be applied in the time domain as well as in the Laplace domain.

In the second phase, the *filter topology design phase*, the filter transfer function is mapped on a suitable filter topology. In such a topology, the input node, the output node and the filter's main building blocks, the integrators¹, are interconnected. An equivalent method for describing the topology of the filter is the state space description, in which matrices are used to describe the connectivity of the integrators and the coupling of the input and the output. An n -th order filter can always be constructed by means of n integrators.

In Section III, the filter's state-space description will be optimized for dynamic range, sparsity and sensitivity. It will be shown that dynamic range optimization boils down to transforming the controllability gramian such that it becomes a diagonal matrix with equal diagonal entries, transforming the observability gramian such that it also becomes a diagonal matrix, and capacitance distribution. To improve the state-space matrices' sparsity the dynamic-range optimized matrices can be transformed into a form that describes an orthonormal ladder filter. After applying capacitance distribution, a filter topology is found that is not too complex and has a dynamic range that is close (i.e., within a few dBs) to optimal.

Finally, in the *filter circuit design phase*, the filter topology is mapped on a circuit. This includes the implementation of the integrators, the interconnection circuitry and their biasing subcircuits in a suitable IC technology. In Section IV, a novel nA/V CMOS transconductor for ultra-low power low-frequency gm-C filters will be introduced, employing transistors operating in strong inversion and in the triode region. Contrary to previous designs, its transconductance depends on the size of the input transistors and a control voltage only.

To illustrate the entire filter design procedure, in Section V a 10th-order dynamic translinear *Morlet* filter will be presented. Simulations and measurements will demonstrate an excellent approximation of the Morlet wavelet base. The circuit operates from a 1.2-V supply and a bias current of $1.2\mu\text{A}$.

¹ Although there is no preference for either a differentiator or integrator from a transfer-function or topological point of view, at circuit level, the use of differentiators often gives rise to high-frequency problems or instability. Therefore, in a filter, almost always integrators are employed.

II. DESIGNING THE FILTER TRANSFER FUNCTION

In the filter transfer function design phase, the aim is to generate a transfer function that satisfies the desired specifications, which may concern, in the frequency domain: the amplitude (or magnitude) response, the phase response — together with the amplitude response grouped in the two so-called Bode plots —, the group delay, the cutoff frequency, the passband/stopband loss, the passband/stopband edges, the amplitude/phase/delay distortion; and in the time domain, the impulse/step responses (including the overshoot, delay time and rise time).

The available methods for generating the filter transfer function can be classified as *closed-form* or *iterative*. In closed-form methods, the transfer function is derived from a set of closed-form formulas or transformations. Some classical closed-form solutions are the so-called Butterworth, Chebyshev, Bessel-Thompson and elliptic approximations. Iterative methods entail a considerable amount of computation but can be used to design filters with arbitrary responses.

If the desired filter transfer function does not have an explicit expression, then the splines interpolation method [1] can be used to generate the desired (idealized) filter transfer function that can be used as a starting point for the filter design process.

Taking into account that in active filters the power consumption and the dynamic range are proportional and inversely proportional to the order of the filter, respectively, in this phase, the joint optimization of power consumption and dynamic range means finding a low-order approximation of the Laplace transform of the desired filter transfer function. In the sequel we will deal with two, relatively unknown, techniques to come to such an approximation: the Padé approximation and the L_2 approximation.

A. Padé approximation

The Padé approximation [2] is employed to approximate the Laplace transform of the desired filter transfer function $G(s)$ by a suitable rational function $H(s)$ and is characterized by the property that the coefficients of the Taylor series expansion of $H(s)$ around a selected point $s = s_0$ coincide with the corresponding Taylor series coefficients of $G(s)$ up to the highest possible order, given the pre-specified degrees of the numerator and denominator polynomials of $H(s)$. If we denote the Padé approximation $H(s)$ at $s = s_0$ and of order (m, n) , with $m \leq n$, by

$$H(s) = \frac{p_0(s - s_0)^m + p_1(s - s_0)^{m-1} + \dots + p_m}{(s - s_0)^n + q_1(s - s_0)^{n-1} + \dots + q_n}, \quad (1)$$

then there are $n + m + 1$ degrees of freedom, which generically makes it possible to match exactly the first $n + m + 1$ coefficients of the Taylor series expansion of $G(s)$ around $s = s_0$. As this matching problem can easily be rewritten as a system of

$n + m + 1$ linear equations in the $n + m + 1$ variables $p_0, p_1, \dots, p_m, q_1, \dots, q_n$, a unique solution is obtained that is easy to compute. Moreover, a good match is guaranteed between the given function $G(s)$ and its approximation $H(s)$ in a neighborhood of the selected point s_0 .

However, there are also some disadvantages which limit the practical applicability of this technique [3]. One important issue concerns the selection of the point s_0 . Note that a good approximation of $G(s)$ around one point in the (complex) Laplace domain is not a requirement *per se*. A second important issue concerns stability, which does not automatically result from the Padé approximation technique. For example, if emphasis is put on obtaining a good fit for a particular s_0 , it may easily happen that the resulting approximation becomes unstable. The trade-off between a good fit near a certain point $s = s_0$ and stability is a non-trivial problem. A third issue concerns the choice of the degrees m and n of the numerator and denominator polynomials of the rational approximation $H(s)$. An unfortunate choice may yield an inconsistent system of equations or an unstable approximation.

B. L_2 approximation

An alternative to the Padé approximation is the so-called L_2 approximation, which offers a number of advantages [3]. First, on the conceptual level, it is quite appropriate to use the L_2 norm to measure the quality of an approximation $H(s)$ to the function $G(s)$. Another advantage of L_2 approximation is that it can be applied in the time domain as well as in the Laplace domain. According to Parseval's equality, minimization of the squared L_2 norm of the difference between $G(s)$ and $H(s)$ over the imaginary axis $s = j\omega$ is equivalent to minimization of the squared L_2 norm of the difference between $g(t)$ and $h(t)$.

Particularly in the case of low order approximation, the L_2 approximation problem can be approached in a simple and straightforward way using standard numerical optimization techniques and software.

III. DESIGNING THE FILTER TOPOLOGY

After we have completed the design of the filter transfer function, it is time to design the filter topology. As there are many possible state-space descriptions for a certain transfer function, there are many possible filter topologies. We will concentrate on finding a filter topology that is optimized for both dynamic range and power consumption.

As is well known from linear systems theory (see, e.g., [4]) any causal linear filter of finite order n can be represented in the Laplace domain as a state-space system (A, B, C, D) described by a set of associated polynomial equations of the form:

$$sX(s) = AX(s) + BU(s), \quad (2)$$

$$Y(s) = CX(s) + DU(s), \quad (3)$$

where $U(s)$ denotes the scalar input to the filter, $Y(s)$ the scalar filter output and $X(s)$ the state vector. The transfer function of the filter is given by:

$$H(s) = C(sI - A)^{-1}B + D. \quad (4)$$

A system's dynamic range is essentially determined by the maximum processable signal magnitude and the internally generated noise. It is well known that the system's controllability and observability gramians play a key role in the determination and optimization of the dynamic range [5], [6]. The controllability (K) and observability (W) gramians are derived from the state space description and are computed by solving the equivalent Lyapunov equations

$$AK + KA^T + 2\pi BB^T = 0, \quad (5)$$

$$A^TW + WA + 2\pi C^TC = 0. \quad (6)$$

As the dynamic range of a circuit is defined as the ratio of the maximum and the minimum signal level that it can process, optimization of the dynamic range is equivalent to the simultaneous maximization of the (distortionless) output swing and the minimization of the overall noise contribution. In [7], Rocha gives a geometric interpretation of the optimization of the dynamic range. A visualization of the optimization procedure can be seen in Fig. 1, for a system with three state variables. The output swing is related via the controllability gramian to the space of 'occurring' state-space vectors. Under the assumption of a random input signal, the shape of this space is generally a multidimensional ellipsoid. The constraint that each integrator has a maximum representation capacity (M) defines a multidimensional cuboid, which, for a distortionless transfer, should contain the former mentioned ellipsoid completely. As the mean square radius of the ellipsoid is equivalent to the maximum output swing, the output swing is maximal when the mean square radius is. This can occur if and only if the ellipsoid becomes a spheroid. In that case the controllability gramian is a diagonal matrix with equal diagonal entries, which means that all axes of the ellipsoid have equal length. Thus, the first optimization step boils down to a similarity transform, such that the controllability gramian of the new system becomes a diagonal matrix with equal diagonal entries. In the second step of the optimization procedure, the system is optimized with respect to its noise contribution. Rocha defines another ellipsoid, which describes the noise that is added to the state vector in each direction. While preserving the result of the first optimization step, it is possible to rotate the state space, such that the observability gramian becomes a diagonal matrix as well. In that case, the axes of the noise ellipsoid are aligned with the 'system axes'.

In [7] it is shown that, in order to maximize the dynamic range of the system, one should minimize the objective functional, which represents the relative improvement of the dynamic range and contains all parameters which are subject to manipulation by the designer. The objective functional is given by

$$F_{DR} = \frac{\max_i k_{ii}}{(2\pi)^2} \sum_i \frac{\alpha_i}{C_i} w_{ii}, \quad (7)$$

where k_{ii} and w_{ii} are the main diagonal elements of K and W , respectively, $\alpha_i = \sum_j |A_{ij}|$ is the absolute sum of the elements on the i -th row of A , and C_i is the capacitance in integrator i .

Finally, profiting from the well-known fact that the relative noise contribution of an integrator decreases when the capacitance and bias current increase, we apply noise scaling, i.e., we match an optimal capacitance distribution to the noise contributions of each individual integrator, viz. the diagonal entries of W combined with the coefficients in matrix A , resulting in [7]

$$C_i = \frac{\sqrt{\alpha_i w_{ii} k_{ii}}}{\sum_j \sqrt{\alpha_j w_{jj} k_{jj}}}. \quad (8)$$

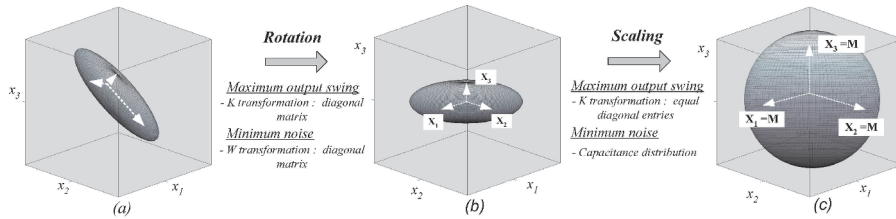


Fig. 1. Dynamic range optimization based on the similarity transformation of K and W and capacitance distribution. The coordinate axes represent the state variables and the cuboid represents the maximum signal amplitude (M) that the integrators are able to handle. (a) The initial state space representation (ellipsoid) is usually not well adapted to the integrator's representations capacity bounds (cuboid). (b) The (rotated) ellipsoid's principal axes are now aligned to the coordinate axes, as a result of the diagonalization procedure to the matrices K and W . (c) Finally, the optimized state space representation is obtained by scaling the state variables and the noise. Note that the sphere represents the maximum possible mean square radius which can be fitted into the integrator's capacity cuboid.

The drawback of a dynamic-range optimal system is that its state-space matrices are generally fully dense, i.e., all the entries of the A , B , C matrices are filled with nonzero elements. These coefficients will have to be mapped on circuit components and will result in a complex circuit with a large number of interconnections. For high-order filters it is therefore necessary to investigate how a realization of the desired

transfer function having sparser state-space matrices would compare to the one having maximal dynamic range. Also, when designing high-order filters, it is very desirable to concentrate on circuits that are less sensitive to component variations. It is known that an optimal dynamic range system will also have optimal, i.e., minimal, sensitivity [8]. For a less complex circuit, it is possible, for instance, to reduce A to upper triangular by a Schur decomposition and by this reducing the number of non-zero coefficients in A . However, this transformation leads to an increase in the system noise and consequently to an increase in the objective functional (7). Another possibility is the *orthonormal ladder structure* [9], which is significantly sparser than the fully dense A matrix of the dynamic-range optimal system and the Schur decomposition and still presents a good behavior with respect to sensitivity. Fig. 2 shows a block diagram of a general orthonormal ladder filter [9]. As shown in the block diagram, the filter output is obtained from a linear combination of the outputs of all integrators.

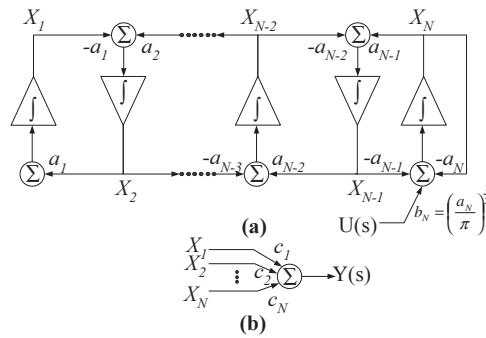


Fig. 2. Block diagram of an orthonormal ladder filter, (a) Leapfrog structure; (b) Output summing stage

The A matrix of an orthonormal ladder filter is tridiagonal and is very nearly skew-symmetric except for a single nonzero diagonal element. The B vector consists of all zeros except for the N th element. Another property of orthonormal ladder filters is the fact that the resulting circuits are inherently state scaled, i.e., the controllability gramian is already a identity matrix. The drawback of this structure is that the system is not optimized with respect to its noise contribution. However, if an optimal capacitance distribution is applied to this suboptimal system, it can still yield some extra gain compared to the case of equal capacitances. Often this leads to a filter topology that is not too complex and has a dynamic range that is close (i.e., within a few dBs) to optimal.

IV. DESIGNING THE FILTER CIRCUIT

After an optimal filter topology has been selected and the appropriate coefficients have been chosen, it's time to design the filter circuit, or more specifically, design the filter's main building block, viz. the integrator.

A. Four integrator classes

In order to be able to construct the filter topology, the transfer of the integrators should be dimensionless. On a chip, the integrating element is a capacitor, which can be employed as a (passive) capacitance or as part of an active transcapacitance (amplifier) and whose transfer has a dimension equal to $[\Omega]$. To realize a dimensionless integrator transfer function, we thus need an additional (trans)conductance. Hence, four types of integrators can be distinguished:

- a conductance-capacitance integrators,
- b conductance-transcapacitance integrators,
- c transconductance-capacitance (gm-C) integrators, and
- d transconductance-transcapacitance integrators.

Fig. 3 depicts the four integrator types that implement a voltage-to-voltage integration.

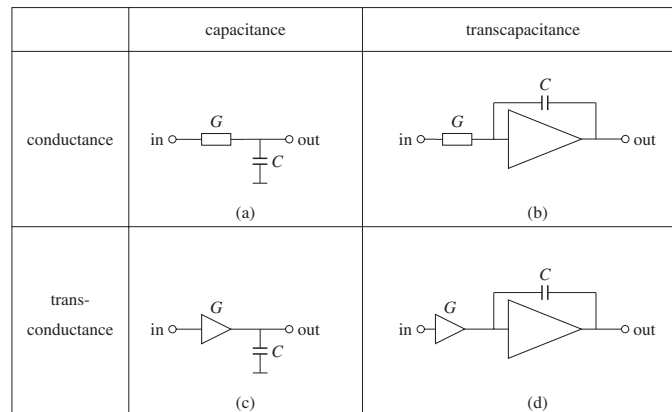


Fig. 3. Four classes of integrators

The *conductance-capacitance* integrator does not use active components. Both the required conductance and integration are implemented passively. As a result, using this type of integrator, it is not possible to implement filter transfer functions with complex poles.

The second type of integrator, the *conductance-transcapacitance* integrator, does not have this drawback and is thus used more often. In this type of integrator, the realization of the actual integration function is an active transcapacitance, often comprising an operational amplifier (op amp) having a capacitor in its (shunt) feedback path. The opamp can be designed to operate rail-to-rail at the output terminals, so full advantage is taken of the supply voltage, which entails an optimal dynamic range. The conductance can be integrated as a diffused resistor, but it could also be implemented as an MOS transistor in the triode region thus yielding a *MOSFET-C* integrator [10].

The third integrator type, the *transconductance-capacitance* (gm-C) integrator, makes use of active conductances, i.e., transconductances. The advantage of transconductors is that they are able to operate at relatively high frequencies, because their parasitic capacitances are in parallel with the integrator capacitors. Thus, they can be accounted for easily in the dimensioning of the required capacitor [11]. A major drawback, however, is that it is very difficult to implement transconductors with rail-to-rail input capability.

The fourth type of integrator is the *transconductance-transcapacitance* integrator. This integrator has no advantages over the second and third integrators mentioned. An important disadvantage is the use of two active parts, both adding to the distortion, the power consumption and the noise production.

In conclusion, the second and third type of integrators are preferred when designing filters. For both types of integrators an active part is required.

B. *ELIL and ELIN*

As integrators consist of two parts, a (trans)conductance and a (trans)capacitance, based on the relation of the intermediate quantity to the input and/or output quantity, linear integrators, our main filter building blocks, can be further classified into two categories [12]:

- externally linear, internally linear (ELIL), and
- externally linear, internally non-linear (ELIN).

Most of the known integrator types fall into the first category, being ELIL. In ELIL integrators the intermediate quantity is linearly related to the input and output quantities. Among them are the integrator topologies that are commonly referred to as *gm-C*, *MOSFET-C*, *opamp-RC*, *RC* and even (albeit discrete time rather than continuous time) *switched-capacitor (SC) integrators*. As in ultra low-power (i.e., nano- and micro-power) applications, resistors would become too large for integration on chip, occupying a large chip area, having a small bandwidth or have large absolute tolerances, and MOSFET conductances are bound to a limited dynamic range, we will not deal with these any further in the sequel. Instead, we will introduce a novel type

of transconductor, employing MOSFETs operating in the triode region as (active) transconductors.

For the second category, that of ELIN integrators, it holds that their external behavior is precisely linear, yet the intermediate quantity is non-linearly related to its input and output quantities. In here we find the subcategory of instantaneous companding² integrators, i.e., the degree of compression/expansion at a given instant depends only on the value of signals at that instant [12], [13]. Belonging to this subcategory, the class of *dynamic translinear* [13] (also known as *log-domain* [14], [15], [16] or *exponential state-space* [17]) is probably the most well known. To the subcategory of companding integrators, albeit discrete-time rather than continuous-time, also belong *switched current* [18] and *switched MOSFET* [19], [20] integrators. We will give an example of a dynamic translinear wavelet filter for biomedical applications in the next section.

But first, as promised, we will introduce a transconductor employing MOSFETs operating in the triode region.

C. A compact CMOS triode transconductor

On-chip realizations of large time constants are often required to design low cutoff-frequency (in the Hz and sub-Hz range) continuous-time filters in applications such as integrated sensors, biomedical signal processing and neural networks. To limit capacitors to practical values, a transconductor with an extremely small transconductance g_m (typically a few nA/V) is needed.

Previous works on low-voltage low-power CMOS techniques for obtaining very-low transconductances essentially concentrated on the combination of voltage attenuation at the input, source degeneration in the transconductor core and current splitting at the output [21], [22], [23], [24], keeping the transconductor input transistor(s) in saturation; whereas the lowest g_m/I_D ratio is obtained in strong-inversion triode-region (SI-TR).

In [25], a low- g_m pseudo-differential transconductor based on a four-quadrant multiplication scheme is presented, in which the drain voltage of a triode-operating transistor follows the incoming signal. Nevertheless, because triode operation needs to be sustained, the input-signal swing is rather limited. Moreover, this solution only applies to balanced structures. Although triode-transconductors, in which the signal is directly connected to the input-transistor gate, have been successfully employed in high-frequency gm-C filters [26], [27], their potential for very-low frequency filter design has not been addressed as yet.

Here we present a novel SI-TR transconductor for application in ultra low-power low-frequency gm-C filters, in which, contrary to previous approaches, the transcon-

² Companding is a combination of *compressing* and *expanding*

ductance, g_m , is being controlled by a voltage rather than by a current. In a SI-TR MOSFET, by connecting the source terminal to one of the supply rails, a control voltage applied to the drain linearly adjusts g_m , as the latter scales with the drain-source voltage V_{DS} . Since (W/L) offers a degree of freedom in the design of a particular transconductance, V_{DS} values well above the equivalent noise and offset of the bias circuit can be set, while still obtaining a very-low g_m . Consequently, filters with more predictable transfer functions can be implemented. Owing to its extended linearity, the SI-TR transconductor also handles larger signals, with no need for linearization techniques.

The proposed transconductor is depicted in Fig. 4 [28]. Input transistors M_{1A} - M_{1B}

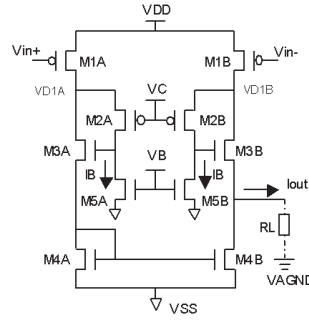


Fig. 4. Proposed triode-transconductor

have their drain voltages regulated by an auxiliary amplifier that comprises M_{2A} - M_{2B} , M_{3A} - M_{3B} and bias current sources M_{5A} - M_{5B} . A simple current mirror M_{4A} - M_{4B} provides a single-ended output. All transistors are assumed to be pair-wise matched. Although the gate-source voltages of M_{3A} and M_{4A} are stacked, their values are below the threshold voltages, so that the circuit still complies with low-voltage requirements. The gate-voltage of M_{2A} - M_{2B} is set to $V_C = V_{TUNE} - V_{GS2}$, whereas V_B imposes a bias current I_B through M_{5A} - M_{5B} . Both voltages V_B and V_C are generated on chip. Referring V_{TUNE} to V_{DD} , the transconductance of the entire circuit becomes:

$$g_m = g_{m_1} = \beta_1 V_{TUNE}, \quad (9)$$

with $\beta_1 = (W/L)_1 \mu_p C_{ox}$.

P-type input transistors were chosen because of their lower mobility and 1/f-noise coefficients as compared to similar parameters of n-MOSFETs. Except for M_{1A} - M_{1B} that stay in SI-TR, all remaining devices work in weak inversion and saturation. Assuming M_{5A} and M_{5B} to be ideal current sources, the transconductor output resistance

r_{out} is given by

$$r_{\text{out}} \approx r_{ds1}(1 + g_{m2}r_{ds2}) \quad (10)$$

Even though a common-drain configuration (M_{3B}) is seen from the output node, the transconductor still exhibits a relatively high output resistance, as the loop gain around M_{2B} and M_{3B} is relatively large.

Internal voltages V_B and V_C are derived from the circuit shown in Fig. 5. The generator is structurally alike the transconductor, with M_{1G} , M_{2G} and M_{3G} matched to their counterparts. An opamp equates the drain voltage to external voltage V_{TUNE} , so that $V_C \approx V_{\text{TUNE}} |V_{GS2G}|$. Since $V_{GS2G} = V_{GS2A} = V_{GS2B}$, the expected value of V_C is achieved. A low-voltage OTA, with a topology similar to the one in [27], is employed as opamp. A proper setting of the current gain B ($B > 1$) in current mirror M_{4G} - M_{5G} guarantees an optimal signal swing at both input and output, ensuring class-A operation of the transconductor.

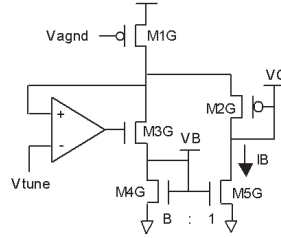


Fig. 5. Bias generator

Analysis of the noise performance of the proposed transconductor reveals that, as $g_{m1}r_{ds1} \ll 1$, the noise is dominated by the noise contributions of M_{2A} and M_{2B} . Their equivalent input noise voltage power spectral densities $S_{v_n,2A/B,eq}$, in $[V^2/\text{Hz}]$ equal

$$S_{v_n,eq} = \frac{2kT/g_{m2}}{(g_{m1}r_{ds1})^2} \quad (11)$$

which is the minimum one can achieve from an SI-TR transconductor.

As the gate length of M_1 is chosen considerably long to obtain a very-low g_{m1} , its $1/f$ noise is naturally minimized.

To back up the theoretical analysis, a SI-TR transconductor with g_m in the order of nA/V was designed. The design complies with $V_{DD} = 1.5\text{V}$ and a standard $0.35\mu\text{m}$ n-well CMOS process, with typical parameters $V_{Tn} = 0.50\text{V}$, $V_{Tp} = -0.60\text{V}$, $g_n = 0.58\text{V}^{1/2}$, $g_p = 0.45\text{V}^{1/2}$, $\mu_n = 403\text{cm}^2/\text{Vs}$, $\mu_p = 129\text{cm}^2/\text{Vs}$ and $C_{ox} = 446\text{nF}/\text{cm}^2$. Flicker-noise coefficients are $\text{KF}_n = 2.81\text{e-}27\text{A}^2\text{s}/\text{V}$, $\text{KF}_p = 1.09\text{e-}27\text{A}^2\text{s}/\text{V}$, $\text{AF}_n = 1.40$, $\text{AF}_p = 1.29$ and $\text{EF}_n = \text{EF}_p = 1$.

The tuning interval ranges from 10mV to 50mV, which implies $1.1\text{nA/V} \leq g_{m_1} \leq 5.5\text{nA/V}$. The optimal V_{AGND} is 0.6V, theoretically limiting the signal amplitude to 185mV. Transistor sizes (in $\mu\text{m}/\mu\text{m}$) are $(W/L)_1 = (1.2/600)$, $(W/L)_2 = (10/100)$, $(W/L)_3 = (12/2.4)$ and $(W/L)_4 = (W/L)_5 = (40/40)$. These dimensions maximize the signal swing at both input and output and trade off $1/f$ -noise and layout area. At nominal $V_{\text{TUNE}} = 20\text{mV}$, the calculated g_{m_1} and common-mode current $I_{D_{1,CM}}$ are 2.2nA/V and 0.63nA , respectively. Setting $B=1.5$ results in $I_B \approx 0.25\text{nA}$, a good compromise between signal swing, $1/f$ -noise, thermal noise and auxiliary-amplifier power consumption.

Simulations were carried out using PSPICE 9.2 with Bsim3v3 models. For a $1\text{k}\Omega$ load, fixing $V_{\text{in-}}$ at V_{AGND} and sweeping $V_{\text{in+}}$, the g_{m_1} dependence on the tuning voltage ($10\text{mV} \leq V_{\text{TUNE}} \leq 50\text{mV}$) is plotted in Fig. 6. The transconductance remains almost constant in the linear region, scaling linearly with V_{DS1} .

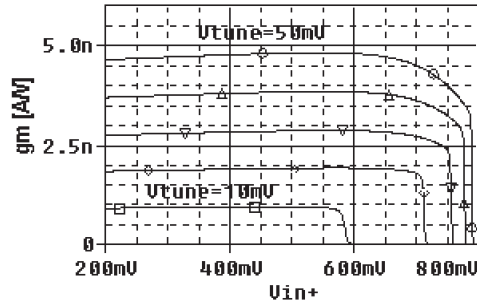


Fig. 6. Dependence of g_m on signal level and tuning

Transconductor noise figures from PSPICE are in excellent agreement with the performed noise calculations. The transconductor equivalent noise voltage for a 100mHz – 10Hz bandwidth is $260\mu\text{V}_{\text{RMS}}$. Similarly, the input-referred noise of the V_C generator is $42\mu\text{V}_{\text{RMS}}$, so that for the lowest V_{TUNE} of 10mV , a tuning-to-noise ratio (TNR) of 47dB is obtained. Given that transistor geometries are well defined in modern fabrication processes, g_m can be controlled to a good extent, as it relies on $(W/L)_1$ and V_{TUNE} only.

V. A 10TH-ORDER ULTRA LOW-POWER LOW-VOLTAGE DYNAMIC TRANSLINEAR WAVELET FILTER

This last section illustrates the design procedure outlined in the previous sections for implementing a filter whose impulse response is a Morlet [29]. The real part of this particular wavelet is of special interest for the local analysis of non-stationary signals as can be found in electrocardiograms. Its application in pacemaker frontends makes

an ultra low-power implementation mandatory. In the coming subsections, we first derive a suitable Morlet filter transfer function. Subsequently, we optimize the Morlet filter state-space description. Finally, we implement the optimized (orthonormal) ladder structure with log-domain integrators as main building blocks. Simulations and measurements that prove the correctness and robustness of the proposed design methodology will be provided as well.

A. Designing the Morlet filter transfer function

The design of the Morlet filter transfer function takes off with the (real part of the) desired impulse response $g(t)$ of the Morlet filter, i.e., a Gaussian-windowed sinusoid:

$$g(t) = \cos(5\sqrt{2}t)e^{-(t-3)^2}, \quad t \geq 0. \quad (12)$$

Since only causal filters can be implemented, this function is truncated at $t = 0$ and a time shift $t_0 = 3$ is introduced. The choice of this time shift involves an important trade-off that has to be made with care. If t_0 is chosen too small, the truncation error becomes too large. On the other hand, if t_0 is chosen too large, the function to be approximated will become very flat near $t = 0$. This effectively introduces a time-delay, which implies that a good fit can only be achieved with a filter of high order and thus compromises the power consumption.

The Laplace transform of (12) is not yet a suitable rational function and thus a low-order approximation has to be made. A [8/10] Padé approximation yields [30]:

$$H(s) = \frac{0.9s^8 - 13s^7 + 177s^6 - 618s^5 + 345s^4 + 7 \cdot 10^4 s^3 - 4 \cdot 10^5 s^2 + 2 \cdot 10^6 s - 3 \cdot 10^6}{s^{10} + 13s^9 + 336s^8 + 3 \cdot 10^3 s^7 + 4 \cdot 10^4 s^6 + 2 \cdot 10^5 s^5 + 2 \cdot 10^6 s^4 + 8 \cdot 10^6 s^3 + 4 \cdot 10^7 s^2 + 9 \cdot 10^7 s + 3 \cdot 10^8}. \quad (13)$$

Fig. 7 depicts the ideal ($g(t)$) and the approximated ($h(t)$) Morlet filter impulse responses, respectively. A good fit can be observed.

B. Designing the Morlet filter topology

Applying the state-space optimization method described in Section III, we find that the objective functional F_{DR} becomes equal to 96.98. This is the absolute minimum value of the objective functional associated with this transfer function.

To improve the state-space matrices' sparsity without compromising the dynamic range and sensitivity to parameter variations too much, an orthonormal ladder structure is implemented. The A , B , C and D matrices of this structure for the defined transfer function are given by:

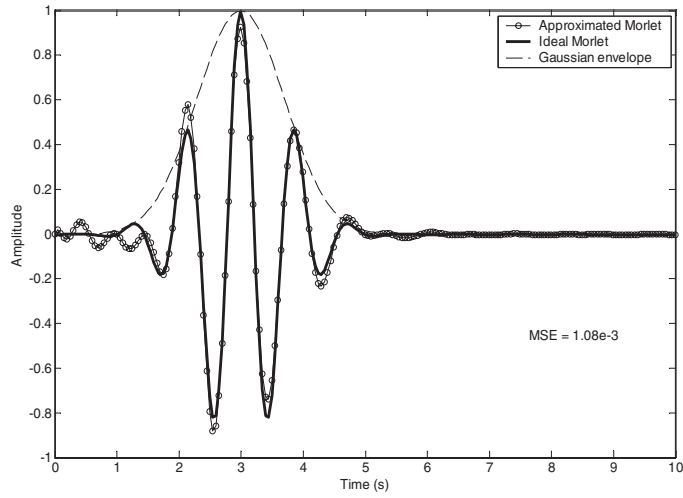


Fig. 7. Impulse response of the Morlet filter: the ideal impulse (dashed line) and the approximated impulse (solid line)

$$\begin{aligned}
 A &= \begin{bmatrix} 0 & 6.54 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -6.54 & 0 & 1.83 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1.83 & 0 & 6.59 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -6.59 & 0 & 2.72 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2.72 & 0 & 6.37 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6.37 & 0 & 3.89 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -3.89 & 0 & 6.27 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -6.27 & 0 & 5.88 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5.88 & 0 & 10.47 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -10.47 & -13.31 \end{bmatrix}, \\
 B &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 2.05 \end{bmatrix}, \\
 C &= [0.75 \quad -1.34 \quad 0.75 \quad 0.68 \quad -0.57 \quad 0.44 \quad -0.002 \quad -0.10 \quad 0.04 \quad 0], \\
 D &= [0]
 \end{aligned} \tag{14}$$

In order to minimize the noise contribution, an optimal capacitance distribution is applied, resulting in a normalized capacitance distribution $(C_1, \dots, C_{10}) = C'(0.142, 0.162, 0.110, 0.117, 0.086, 0.091, 0.073, 0.080, 0.073, 0.061)$, where C' represents the unit-less value of the total capacitance expressed in F. This leads to an objective functional $F_{DR} = 147.90$, which is not so far from the optimum case. The dynamic range has decreased by only 1.83dB.

C. Designing the Morlet filter circuit

A simple bipolar multiple-input low-power log-domain integrator [31] will be used as the basic building block for the implementation of the above state space description. This log-domain integrator is shown in Fig. 8 [31]. A pair of log-domain cells with opposite polarities and an integrating capacitor form the core of the integrator. V_{ip} and V_{in} are the noninverting and inverting input voltages, respectively, and the input currents are I_{ip} and I_{in} , which are superimposed on the dc bias currents. The output voltage V_o is given by the voltage across the capacitor. The circuit is composed of two identical log-domain cells, a voltage buffer and a current mirror. The log-domain cells Q_1 - Q_2 and Q_3 - Q_4 generate the log-domain currents I_{c2} and I_{c4} , respectively. A voltage buffer realized by Q_5 - Q_6 is inserted between them. Therefore, the output log-domain voltage V_o at the emitter of Q_2 also appears at the emitter of Q_4 . Finally, to obtain a log-domain integrator equation, we use a current mirror Q_7 - Q_8 to realize the difference between the two log-domain currents on the capacitor node. The connection from the bases of transistors Q_7 and Q_8 to the collector of Q_6 closes the feedback loop around Q_6 and Q_7 . This connection is convenient because it ensures that the overall voltage headroom is minimized. The equation that relates the input and output voltages to the current flowing in the integrating capacitor becomes

$$C_i \frac{dV_o}{dt} = (I_o + I_{ip}) e^{\frac{V_{ip} - V_o}{V_T}} - (I_o + I_{in}) e^{\frac{V_{in} - V_o}{V_T}}. \quad (15)$$

Notice that the input and output voltages of the integrator are at the same dc level. Therefore log-domain filter synthesis can easily be achieved by direct coupling of these integrators.

D. Synthesis of the log-domain state-space filter

By applying a simple mapping to the linear state-space equations (14), we can obtain the corresponding log-domain circuit realization which employs the above log-domain integrator.

The block diagram of the log-domain implementation of (14) is illustrated in Fig. 9, using the universal log-domain cell symbol described in [32] and shown in Fig. 8b.

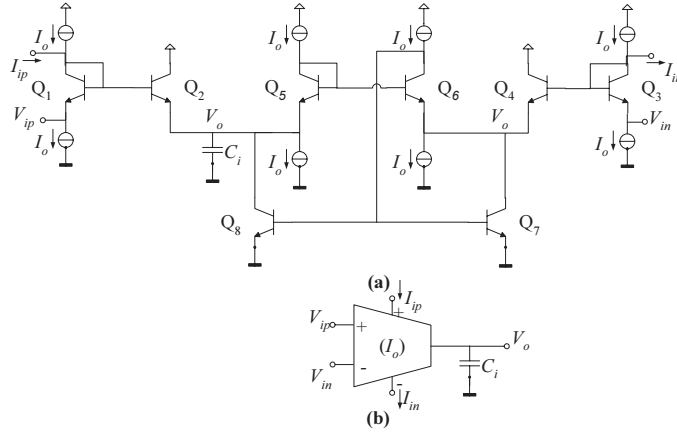


Fig. 8. a) The multiple-input low power log-domain integrator, and (b) its symbol [30]

Note that each column of the filter structure corresponds to a row in the state-space formulation. The parameter A_{ij} is implemented by the corresponding log-domain integrator with bias current $I_{A_{ij}}$, defined by a current matrix A_I

$$A_I = V_T C_i \cdot A \quad (16)$$

The input section, as governed by the state-space vector B , is realized by the first row from the top of Fig. 9. The parameter B is related to the current by

$$B = \frac{I_o}{V_T C_i} \quad (17)$$

Consequently, the B coefficients are not individually controllable by bias currents, and they have to be set equal to each other or to zero. Fortunately, this is the case in (14), where only one non-zero parameter of the B vector is present, as then it is not necessary to transpose the state-space system. The bias current vector C_I , which controls the vector C , is defined as

$$C_I = I_o \cdot C \quad (18)$$

E. Simulation and measurement results

To validate the circuit principle, we have simulated the log-domain state-space filter using models of IBM's $0.18\mu\text{m}$ BiCMOS IC technology. The circuit has been designed to operate from a 1.2V supply. Fig. 10 shows the impulse response of the wavelet

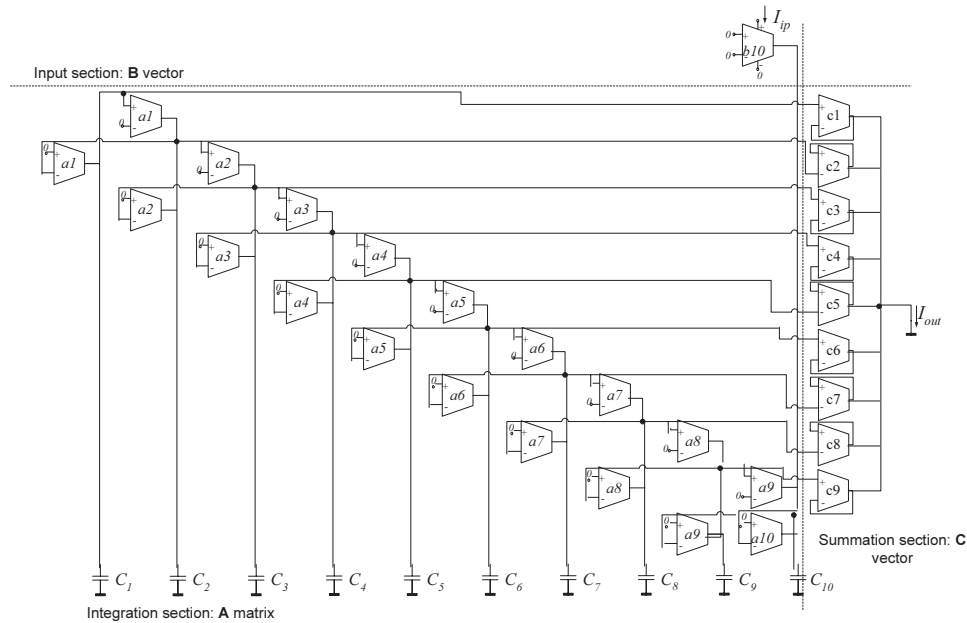


Fig. 9. Complete state-space filter structure

filter. The excellent approximation of the Morlet wavelet can be compared with the ideal Morlet function to confirm the performance of the log-domain filter. Fig. 11 shows the Monte Carlo analysis for process and mismatch variation of the technology in use. As evident from the Monte Carlo simulation (i.e. after 100 runs), the system characteristics show insensitivity towards both absolute and relative variations in the process parameters. Even though the impulse response may be slightly affected, the targeted wavelet analysis will be preserved.

Subsequently, the Morlet filter was implemented in the same IC technology. Fig. 12 shows a photomicrograph of the chip. The 10 integrator capacitors are clearly visible. Fig. 13 shows the measured impulse response. An excellent agreement with both the simulated impulse response and the ideal Morlet function (Fig. 10) can be observed.

The total filter's current consumption is $1.5\mu\text{A}$ with a 100pF total capacitance. The output current presents an offset of approximately 46.61pA . The rms output current noise is 66.97pA , resulting in a DR at the 1-dB compression point of approximately 30dB. The power efficiency of any bandpass continuous-time filter is a figure of merit to be able to compare various filter topologies and can be estimated by means of the power dissipation per pole, center frequency (f_c), and quality factor (Q) defined as

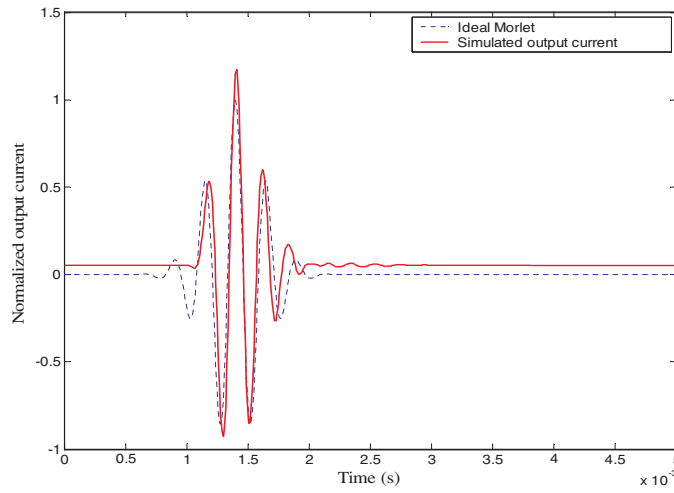


Fig. 10. Simulated impulse response

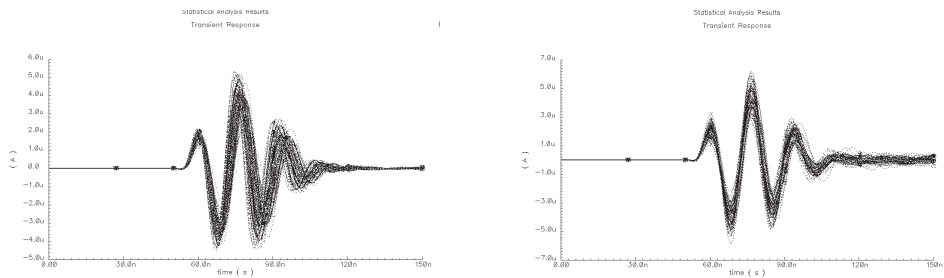


Fig. 11. Monte Carlo analysis (a) process variation, (b) mismatch variation

[33]

$$\text{Power per pole \& bandwidth} = \frac{P_{diss}}{n \cdot f_c \cdot Q}, \tag{19}$$

where P_{diss} is the total power dissipation and n is the order of the filter. The power efficiency of this filter equals 11.83pJ.

By changing the values of the bias currents along a dyadic sequence, one can obtain the impulse responses of a dyadic scale system, as illustrated in Fig. 14. Alternatively, one also may change the capacitance values, C_i . To implement a wavelet system, which usually consists of 5 dyadic scales, one needs to implement a filter bank (a parallel

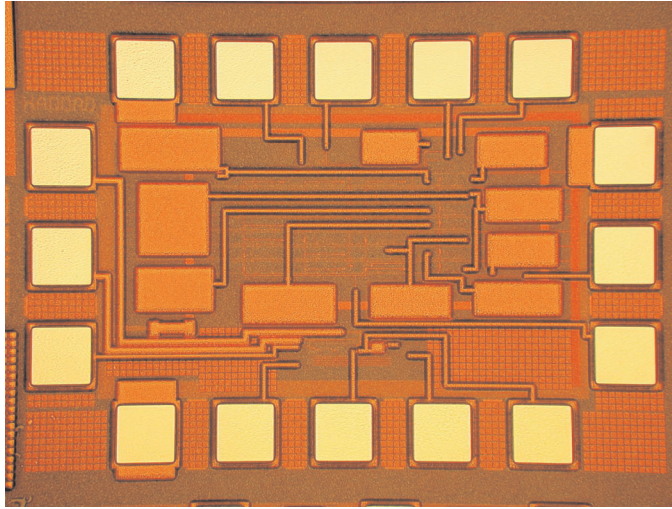


Fig. 12. Photomicrograph of the implemented Morlet filter

structure) with a total capacitance of 193.75pF, preserving the same bias current. This result indicates that a wavelet system is feasible.

Finally, in order to show that the same procedure can be applied for high frequency applications, we tuned the frequency response of the filter by varying the bias current over about four decades with center frequencies ranging from 5.8kHz to 58MHz, while preserving the impulse response waveform. Again, one can obtain the wavelet scales around this frequency (i.e. 58 MHz) by either scaling the current or the capacitance value accordingly. The performance of the filter is summarized in Table I.

VI. CONCLUSIONS

Filtering is an indispensable elementary signal processing function in many electronic systems. In many critical applications, e.g., in portable, wearable, implantable and injectable devices, one should maximize the dynamic range and, at the same time, minimize the power consumption of the filter. This joint optimization can take place in different phases, the filter transfer function design phase, the filter topology design phase, and the filter circuit design phase.

In the filter transfer function design phase, the filter functional input-output relation is mapped on a suitable filter transfer function. Two approximation techniques were introduced: the Padé approximation and the L_2 approximation. The Padé approximation is employed to approximate the Laplace transform of the desired filter

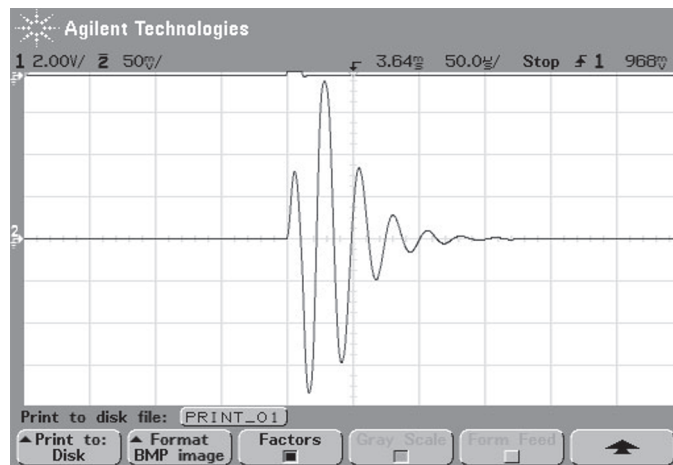


Fig. 13. Measured impulse response

transfer function $G(s)$ by a suitable rational function around a selected point. The L_2 approximation offers a more global approximation, i.e., not concentrating on one particular point, and has the advantage that it can be applied in the time domain as well as in the Laplace domain. It is based on the minimization of the squared L_2 norm of the difference between the desired transfer function and the approximation $H(s)$ over the imaginary axis $s = j\omega$, which is equivalent to minimization of the squared L_2 norm of the difference between $g(t)$ and $h(t)$.

In the filter topology design phase, the filter transfer function is mapped on a suitable filter topology. For this, the filter transfer function is written in the form of a state-space description, which subsequently is optimized for dynamic range, sparsity and sensitivity. In the determination and optimization of the dynamic range the filter's controllability and observability gramians play an important role. Dynamic range optimization boils down to transforming the controllability gramian such that it becomes a diagonal matrix with equal diagonal entries, transforming the observability gramian such that it also becomes a diagonal matrix, and capacitance distribution. To improve the state-space matrices' sparsity the dynamic-range optimized matrices can be transformed into a form that describes an orthonormal ladder filter. After applying capacitance distribution, a filter topology is found that is not too complex and has a dynamic range that is close (i.e., within a few dBs) to optimal.

Finally, in the filter circuit design phase, the filter topology is mapped on a circuit. A classification of integrators was presented. Falling in the category of transconductance-capacitance (gm-C) integrators, a novel nA/V CMOS transconductor for

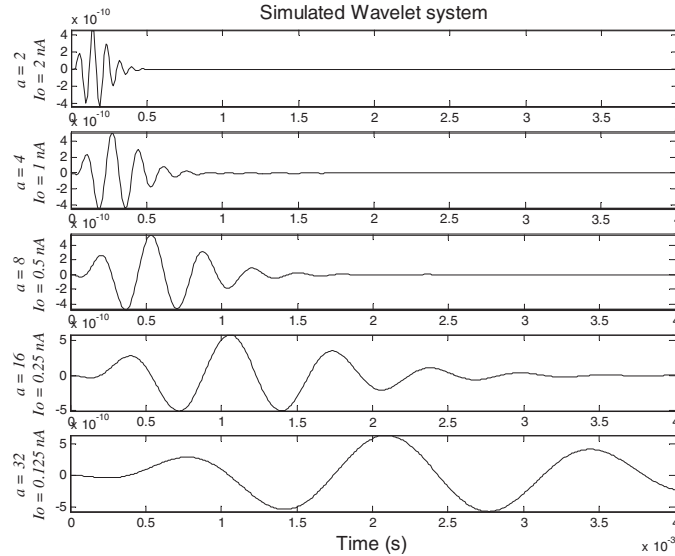


Fig. 14. Simulated impulse responses of a Morlet-based wavelet system with 5 scales. The scales are obtained by varying the current (from 0.125nA to 2nA) or the capacitance (from 100pF to 6.25pF).

ultra-low power low-frequency gm-C filters was introduced. Its input transistors are kept in the triode-region to benefit from the lowest g_m/I_D ratio. The g_m is adjusted by a well defined (W/L) and V_{DS} , the latter a replica of the tuning voltage V_{TUNE} . The resulting design complies with $V_{DD}=1.5V$ and a $0.35\mu m$ CMOS process. Its transconductance ranges from 1.1nA/V to 5.5nA/V for $10mV \leq V_{TUNE} \leq 50mV$.

To illustrate the entire filter design procedure, a dynamic translinear Morlet filter has been designed. Simulations and measurements demonstrate an excellent approximation of the Morlet wavelet base. The circuit operates from a 1.2-V supply and a bias current of $1.2\mu A$.

ACKNOWLEDGEMENT

The author would like to acknowledge the many contributions to this work made by Sumit Bagga, Igor Filanovsky, Gert Groenewold, Joel Karel, Cleber Marques, Bert Monna, Jan Mulder, Ralf Peeters, Daniel Rocha, Pietro Salvo, Arie van Staveren, Nanko Verwaal and Ronald Westra, D. Harame and IBM Microelectronics for fabrication access and fabrication of the testchip, and the financial support of part of this work by FAPESP, The State of São Paulo Research Foundation, CNPq, the Brazilian National Counsel of Technological and Scientific Development, and STW, the Dutch

Technology	0.18 μ m BiCMOS	
Bias current	$I_o = 1\text{nA}$	$I_o = 10\mu\text{A}$
Total capacitance	100pF	100pF
Supply voltage	1.2V	1.8V
Center frequency (f_c)	5.8kHz	58MHz
Power dissipation	1.5 μ W	24.3mW
Dynamic Range (1-dB)	30 dB	30 dB
Noise current (rms)	66.97pA	481.3nA
Supply voltage range	1V - 1.6V	1.7V - 2.1V
Power dissipation per pole f_c and Q	11.834pJ	13.96pJ

TABLE I
PERFORMANCE PER SCALE FOR TWO DIFFERENT OPERATING FREQUENCIES

Technology Foundation.

REFERENCES

- [1] D. Zwillinger, *Standard Mathematical Tables and Formulae*, CRC Press, 30st edition, 1996.
- [2] G.A. Baker Jr., *Essentials of Padé Approximants*, Academic Press, 1975.
- [3] J.M.H. Karel, R.L.M. Peeters, R.L. Westra, S.A.P. Haddad and W.A. Serdijn, *Wavelet approximation for implementation in dynamic translinear circuits*, proc. IFAC World Congress, Prague, Czech Republic, July 4-8, 2005.
- [4] T. Kailath, *Linear Systems*, Prentice Hall, 1980.
- [5] L. Thiele, *On the sensitivity of linear state-space systems*, IEEE Transactions on Circuits and Systems, Vol. 33, No. 5, pp. 502-510, May 1986.
- [6] M. Snelgrove and A.S. Sedra, *Synthesis and analysis of state-space active filters using intermediate transfer function*, IEEE Transactions on Circuits and Systems, Vol. 33, No. 3, pp. 287-301, March 1986.
- [7] D.P.W.M. Rocha, *Optimal Design of Analogue Low-power Systems, A strongly directional hearing-aid adapter*, PhD thesis, Delft University of Technology, April 2003.
- [8] G. Groenewold, *Optimal dynamic range integrators*, IEEE Transactions on Circuits and Systems I, Vol. 39, No. 8, pp. 614-627, August 1992.
- [9] D.A. Johns, W.M. Snelgrove, and A.S. Sedra, *Orthonormal ladder filters*, IEEE Transactions on Circuits and Systems, Vol. 36, No. 3, pp. 337-343, March 1989.
- [10] K.S. Tan and P.R. Gray, *Fully-integrated analog filters using bipolar-JFET technology*, IEEE Journal on Solid-State Circuits, Vol. 13, No. 6, pp. 814-821, December 1978.
- [11] K.W. Moulding and G.A. Wilson, *A fully-integrated five-gyrator filter at video frequencies*, IEEE Journal on Solid-State Circuits, Vol. 13, No. 3, pp. 303-307, June 1978.
- [12] Y. Tsvividis, *Externally linear, time-invariant systems and their application to companding signal processing*, IEEE Transactions on Circuits and Systems II, Vol. 44, No. 2, pp. 65-85, Feb. 1997.
- [13] J. Mulder, W.A. Serdijn, A.C. van der Woerd, and A.H.M. van Roermond, *Dynamic Translinear and Log Domain Circuits: Analysis and Synthesis*, Kluwer Academic Publishers, Boston, 1999
- [14] R. W. Adams, *Filtering in the log domain*, Proc. 63rd Convention A.E.S., May 1979, pp. 1470-1476.
- [15] E. Seevinck, *Companding current-mode integrator: A new circuit principle for continuous-time monolithic filters*, Electronics Letters, Vol. 26, No. 24, pp. 2046-2047, Nov. 1990.
- [16] D. R. Frey, *Log-domain filtering: An approach to current-mode filtering*, Proc. Inst. Elect. Eng., Pt. G, Vol. 140, No. 6, pp. 406-416, Dec. 1993.

- [17] D.R. Frey, *Exponential state space filters: A generic current mode design strategy*, IEEE Transactions on Circuits and Systems I, Vol. 43, No. 1, pp. 34-42, Jan. 1996.
- [18] C. Toumazou, J.B. Hughes and N.C. Battersby (editors), *Switched currents: an analogue technique for digital technology*, Peter Peregrinus, London, 1993.
- [19] F.A. Farag, C. Galup-Montoro and M.C. Schneider, *Digitally programmable switched-current FIR filter for low-voltage applications*, IEEE Journal on Solid-State Circuits, Vol. 35, No. 4, April 2000, pp. 637-641.
- [20] L.C.C. Marques, W.A. Serdijn, C. Galup-Montoro and M.C. Schneider, *A switched-MOSFET programmable low-voltage filter*, proc. SBMicro/SBCCI 2002, Brazil, September 9-14, 2002.
- [21] A. Veeravalli, E. Sanchez-Sinencio and J. Silva-Martinez, *Transconductance amplifier structures with very small transconductances: A comparative design approach*, IEEE Journal of Solid-State Circuits, Vol. 37, No. 6, pp. 770-775, June 2002.
- [22] J. Silva-Martinez and J. Salcedo-Suner *IC voltage to current transducers with very small transconductances*, Analog Integrated Circuits and Signal Processing, Vol. 13, pp. 285-293, 1997.
- [23] M. Steyaert, P. Kinget, W.M.C. Sansen and J. van der Spiegel, *Full integration of extremely large time constants in CMOS*, Electronics Letters, Vol. 27, No. 10, pp. 790-791, 1991.
- [24] A. Arnaud and C. Galup-Montoro, *A fully integrated 0.5-7 Hz CMOS bandpass amplifier*, Proc. of IEEE ISCAS, Vol. 1, pp. 445-448, Vancouver, Canada, 2004.
- [25] A. Veeravalli, E. Sanchez-Sinencio and J. Silva-Martinez, *A CMOS transconductance amplifier architecture with wide tuning range for very low frequency applications*, IEEE Journal of Solid-State Circuits, Vol. 37, No. 6, pp. 776-781, June 2002.
- [26] J. Penneck, *CMOS triode transconductor for continuous-time active integrated filters*, Electronics Letters, Vol. 21, No. 18, August 1985.
- [27] J.A. de Lima and C. Dualibe, *A linearly-tunable CMOS transconductor with improved common-mode stability and its application to gm-C filters*, IEEE Transactions on Circuits and Systems II, Vol. 48, No. 7, pp. 649-660, July 2001.
- [28] J.A. de Lima and W.A. Serdijn, *A compact nA/V CMOS triode transconductor and its application to very-low frequency filters*, Proc. IEEE International Symposium on Circuits and Systems, Kobe, Japan, May 23-26, 2005.
- [29] P. Goupillaud, A. Grossmann, and J. Morlet, *Cycle-octave and related transforms in seismic signal analysis*, Geoprospection, Vol. 23, pp. 85-102, 1984-1985.
- [30] S.A.P. Haddad, S. Bagga and W.A. Serdijn, *Log-domain wavelet bases*, Proc. IEEE International Symposium on Circuits and Systems, Vancouver, Canada, May 23-26, 2004.
- [31] M.N. El-Gamal and G.W. Roberts, *A 1.2V npn-only integrator for log-domain filtering*, IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Vol. 49, No. 4, pp. 257-265, April 2002.
- [32] G.W. Roberts and V.W. Leung, *Design and Analysis of Integrator-Based Log-Domain Filter Circuits*, Kluwer Academic Publishers, the Netherlands, 2000.
- [33] C. Toumazou, G. Moschytz and B. Gilbert, *Trade-Offs in Analog Circuit Design*, Kluwer Academic Publishers, the Netherlands, 2002.

WIRELESS INDUCTIVE TRANSFER OF POWER AND DATA

Robert Puers, Koenraad Van Schuylenbergh, Michael Catrysse, Bart Hermans
Katholieke Universiteit Leuven, ESAT-MICAS
Kasteelpark Arenberg 10
B-3001 Leuven, Belgium
puers@esat.kuleuven.ac.be

Abstract

This text discusses the possibilities when designing a wireless inductive link that works both as an energy link, to power up a remote device, as well as a communication link to retrieve data and to write data to the same remote device, using the same set of inductive coils. Datatransmission from the measurement system to a base unit is achieved by applying absorption modulation, datatransmission to the measurement system by applying amplitude modulation. Some basic formulae and design considerations are given, and a full example applicable to an implantable device is given.

1. Introduction

Inductive links find a widespread use in modern medicine, and are more precisely in use in implantable electronic devices. Pacemakers, defibrillators and cochlear implants are already well established examples, while retinal implants [1–4], neuro-muscular stimulation and recording devices [5–9] and instrumented orthopaedic implants [10,11] are still under development or used on a laboratory scale. For most of these long-term implantable devices, inductive powering is preferred to batteries because of reliability reasons. Datatransmission is sometimes integrated in these systems (uplink in case of the recording devices, downlink in case of stimulation devices), or a second wireless link is used. The present paper discusses a system that contains all in one.

A schematic overview of the system is given in Fig. 1. It will be optimised towards maximal power transfer efficiency and misalignment tolerance: a minimal amount of power transfer should be guaranteed within certain limits of coil separation and lateral or angular misalignment. The system that will be demonstrated is developed for a “smart” orthopaedic implant [11]: the

algorithms for the processing, storage and transmission of the measured data can be reprogrammed in situ, using the downlink data transmission, while the measured data can be transmitted to a PC base station, using the uplink data transmission. In general, the proposed system, that consists of two coils only, can be used for the powering, control and data retrieval of any isolated device, that is within reach of the near field of such a coil, but is not connected to it. In Section 2, the inductive link principles are introduced and the calculation and optimisation of the link is discussed. Section 3 describes the design of the complete inductive powering system and Section 4 the integration of bi-directional data transmission. In Section 5, misalignment analysis of the link is discussed. Finally, Section 6 shows an example, and some conclusions are drawn.

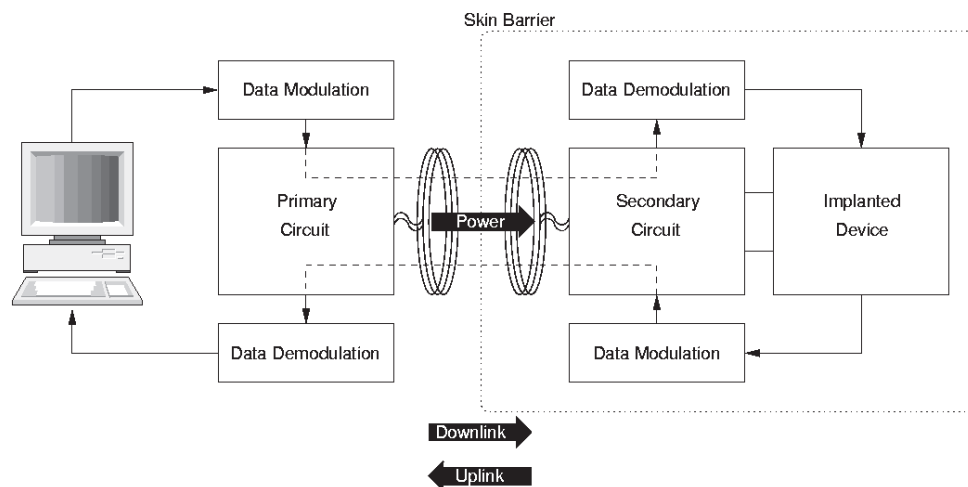


Fig.1. Basic block diagram of an inductive link for power and data transfer.

2. Powering the secondary coil

Coupled coils are used to transfer energy to remote electronics. The a.c. voltage induced in a secondary coil is then rectified to supply some remote circuitry (Figure 2, top). A voltage regulator is also added to smooth out the variation on the induced voltage caused by variations in the coupling and loading conditions. It is a common practise in inductive link design, to represent the power consumption of the remote electronics, the rectifier and the regulator by an equivalent a.c. resistor $R_{load 2}$. Note that the value of this resistor is function of the amplitude of the received voltage: the larger this voltage, the more the regulator has to cut power and the more it dissipates.

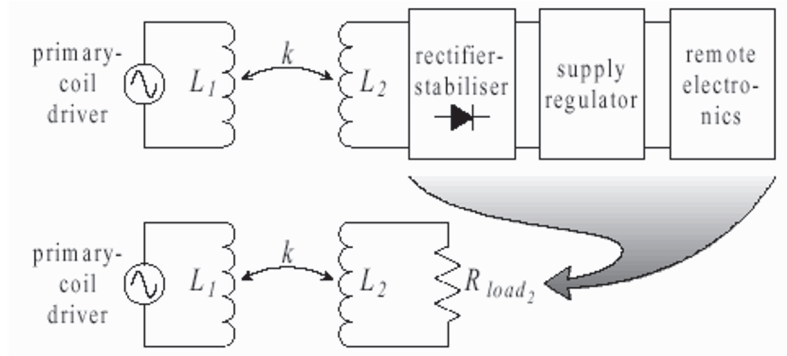


Fig.2. Representation of the load circuit by an equivalent resistor.

Figure 3 demonstrates how the transfer efficiency from input source V_{prim} to the equivalent a.c. load is calculated. The coil-loss resistors R_{S1} and R_{S2} have been added to model the link losses. The link efficiency is computed from the resistive dividers $R_{S1} - R_{\text{eq}}$ in the primary and $R_{\text{load}2} - R_{S2}$ in the secondary:

$$\eta_{\text{link}} = \left(\frac{R_{\text{eq}}}{R_{\text{eq}} + R_{S1}} \right) \left(\frac{R_{\text{load}2}}{R_{\text{load}2} + R_{S2}} \right) = \left(\frac{k}{n} \right)^2 \frac{R_{\text{load}2} \omega^2 L_{S1}^2 k^4}{(R + R_{S1}) \omega^2 L_{S1}^2 k^4 + R^2 R_{S1}}$$

This expression becomes easier to interpret by using the coil quality factors Q_{LS1} and Q_{LS2} , where

$$Q_{LS1} \equiv \frac{\omega L_{S1}}{R_{S1}}, \quad Q_{LS2} \equiv \frac{\omega L_{S2}}{R_{S2}}$$

The maximal link efficiency is only function of two parameters : $k^2 Q_{LS1}$ and Q_{LS2}

$$\eta_{\text{link}_{\text{max}}} = \frac{k^2 Q_{LS1} Q_{LS2}}{2 + k^2 Q_{LS1} Q_{LS2} + 2 \sqrt{1 + Q_{LS2}^2 + k^2 Q_{LS1} Q_{LS2}}}$$

This maximal efficiency increases with the coil coupling and quality factors. It remains, though, impractically low for realistic quality factors (in the range of 100) and weak to moderate coupling (k below 5 %). One of the reasons of the poor link efficiency at low coupling comes from the secondary leak inductance $L_{S2} (1-k^2)$. This inductance is much larger than the useful load R_{S2} at weak coupling and demands for a high induced voltage $k.n.v_1$. A higher v_1 involves higher primary coil current and hence higher (resistive) losses. It is therefore common in inductive link design to cancel the secondary leak inductance with a capacitor C_2 (Figure 4).

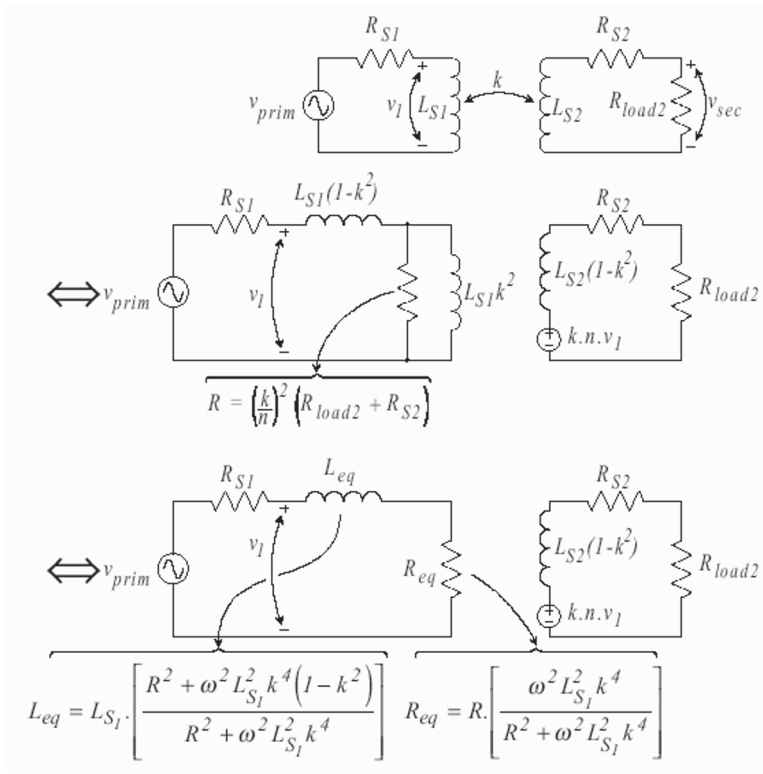


Fig.3: The equivalent link circuits for the calculation of the link efficiency

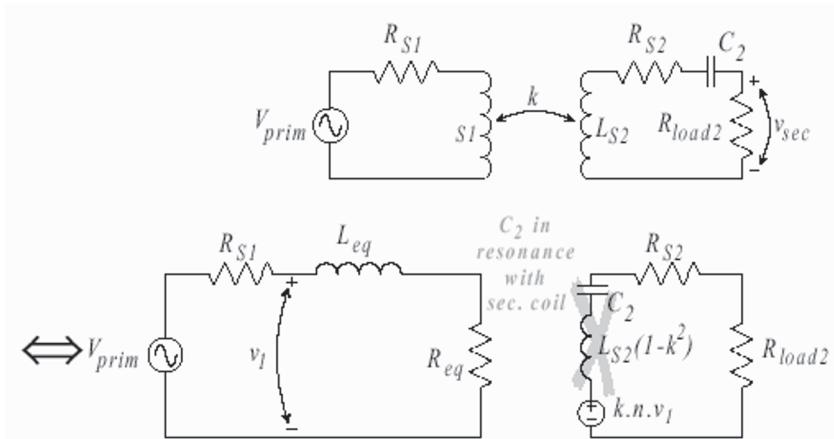


Fig.4. Cancelling of the secondary leak inductance $L_{S2} (1-k^2)$ with a series capacitor C_2 .

The maximal link efficiency for such a series-tuned link is calculate and equals:

$$\eta_{link_{max}} = \frac{k^2 Q_{LS1} Q_{LS2}}{\left(1 + \sqrt{1 + k^2 Q_{LS1} Q_{LS2}}\right)^2}$$

This maximal link efficiency is only function of one single parameter $k^2 Q_{LS1} Q_{LS2}$. Figure 5 clearly shows that secondary resonance indeed improves the link efficiency. This diagram, though, may be a bit misleading, because it erroneously suggests that a lower secondary coil quality factor corresponds to a higher link efficiency in case of a non-resonant secondary. The reason is that $X = k^2 Q_{LS1} Q_{LS2}$ is taken as horizontal coordinate. A lower secondary coil quality factor for the same X , automatically implies a higher primary coil quality factor. The corresponding reduction in the primary dissipation overcompensates the increased secondary loss. Some lines of constant $k^2 Q_{LS1}$ have therefore been added to indicate that increasing Q_{LS2} for the same $k^2 Q_{LS1}$ really increases the efficiency. Note that high-efficiency series-resonant links can only be realized for small load, as a high load dampens the secondary tank resonance too much. A series-resonant link has thus a current-source output characteristic. A voltage-source type output is achieved by parallel resonance of the secondary. The idea of canceling the leak inductance remains, but the tank capacitor is now placed in parallel to the secondary coil.

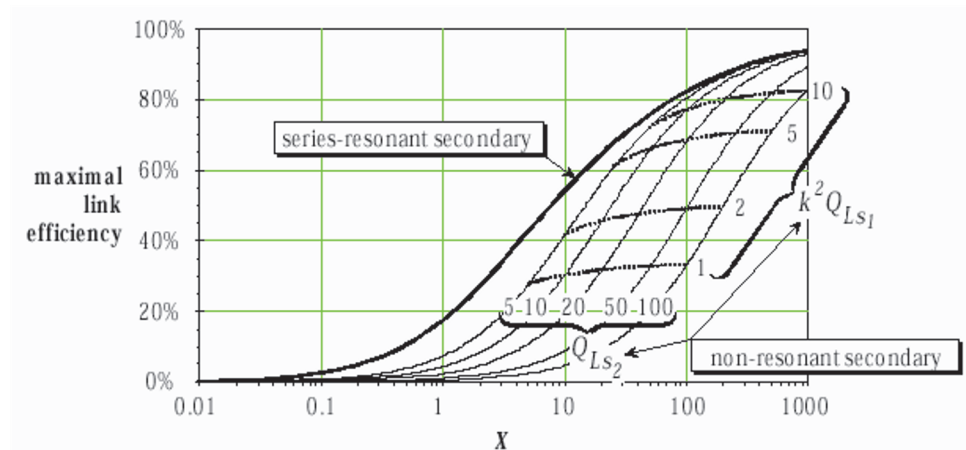


Fig.4. The maximal link efficiency for links with series resonant secondary compared to the link efficiency of non-resonant links as a function of X

2.1. Primary coil driver

A dedicated amplifier is needed to drive high currents into the primary coil in order to generate the magnetic fields required for the power transfer. The driver requirements were summarized by Gutmann in [12]

- It is preferable to use a switch-mode amplifier where the active elements operate as a switch so they only draw currents without carrying a voltage. This will drastically minimize the dissipation in the active elements and avoid its breakdown.
- The driver output should be a pure sinusoid because only the fundamental component is received at the secondary. The harmonic components do not contribute to the power transfer with a tuned secondary coil, but do cause losses in the primary.
- The primary inductance is tuned with a resonant capacitor. The latter is to cancel out the large primary leakage inductance that typically occurs with small coupling factors. This leakage inductance causes a large primary coil voltage, given the large coil currents that are required for the inductive powering.

Basically, the primary leakage inductance can be compensated by either a series or a parallel capacitor. Both ways have their benefits and drawbacks. The use of a series-resonant capacitor lowers the amplifier output voltage but matches the demand for a high output current. The amplifier's output stage needs high-current transistors that require a large base or gate current that also contributes to the driver losses (power MOSFETs have a large gate capacitor due to their large die sizes and take thus large a.c. gate currents). Inductance canceling with a parallel-resonant capacitor lowers the amplifier output current but maintains the need for a high output voltage. The output stage then contains high-voltage transistors that have large parasitic capacitors. The conclusion is that neither series-tuning nor parallel-tuning enables low-voltage and low-current operation. Luckily, the class-E amplifiers feature a double-tuned circuit. They feature a series-tuned coil with a second capacitor in parallel. This offers an elegant solution as it combines the benefits of low-current operation of a parallel resonance with the low-voltage operation of a series resonance.

There are primarily two approaches for driving a link primary:

- Most links use a small-signal master oscillator (MO) followed by a power amplifier (PA) connected to the primary coil. This MOPA set-up works fine, but at higher coupling, an effect known as pole splitting occurs (Figure 6). If the coil coupling is raised, the apparent primary inductance lowers and the primary resonance frequency augments. This effect is negligible at low coupling, but coupling factors (above about 10 %), do get primary tanks out of resonance. A feedback loop in their primary coil driver can be introduced to automatically

adapt the driver frequency to the changing tank tuning [13]. Pole splitting also occurs for the secondary tank. The equivalent secondary impedance of the driven primary tank dampens the secondary tank and lowers its resonance frequency with increasing coupling. This effect is, however, small.

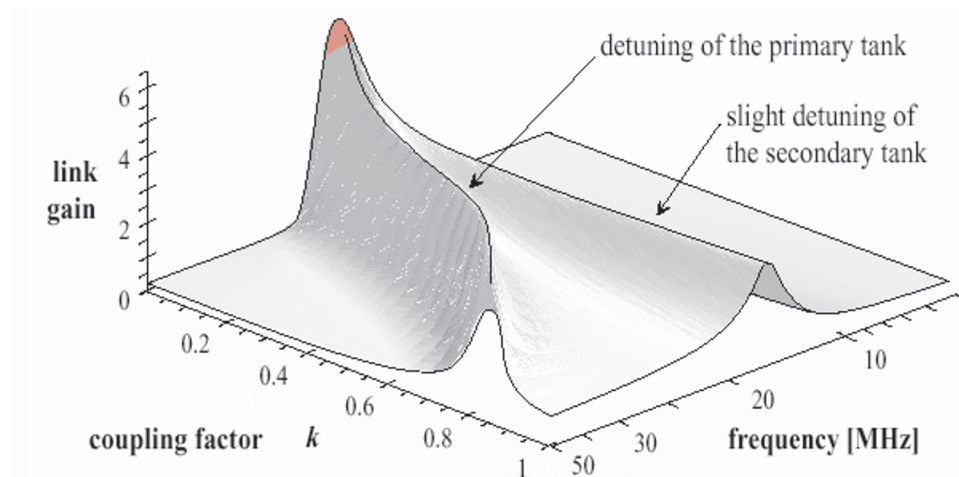


Fig.6 : Pole-splitting effect simulated on a link with a series-resonant primary and a parallel-resonant secondary, synchronous tuned at 20 MHz. $L_{S1} = L_{S2} = 1.96 \mu\text{H}$, $R_{S1} = R_{S2} = 5.18 \Omega$, $R_{load2} = 1 \text{ k}\Omega$

2.2. Link optimisation

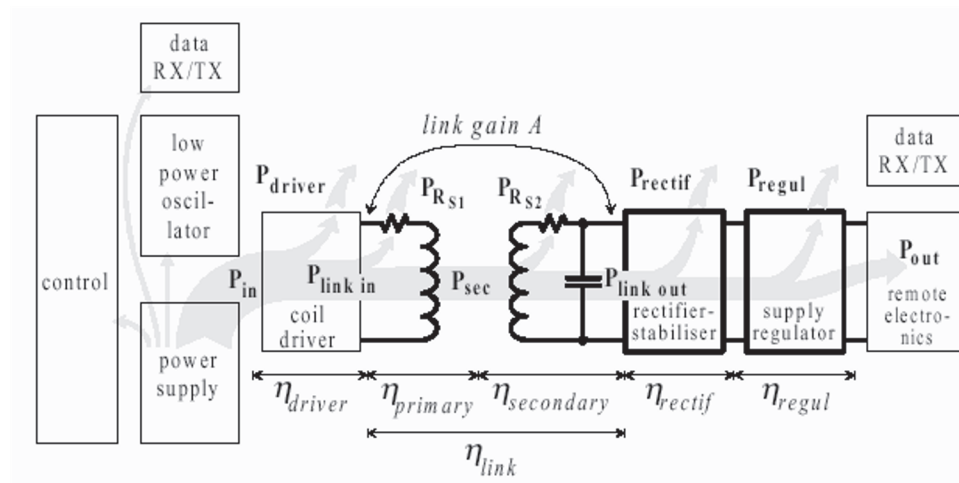


Fig.7 : The power flow in the inductive powering system

Numerous publications appeared on link design and optimisation since the

original paper of Shuder et al. [14]. Most authors use their own notations and give their own design procedure that fitted their specific application. This makes it for the interested reader difficult to obtain an overview of the field of available circuits and techniques, and to compare the optimization methods. But, a closer look reveals that most designs can be reduced to two basic philosophies:

- optimization of the link efficiency η_{link} , assuming secondary tank resonance, and
- de-sensitizing the link gain to coupling variations by critical coupling, assuming a resonant primary and a resonant secondary tank.

The differences between the design procedures are mostly found in the assumptions and simplifications that were (often implicitly) taken, with a strong dependency on the envisaged application. It is, for instance, common to assume that $k^2 Q_{LS1} Q_{LS2}$ is much larger than one. However, we found that this assumption is hard to maintain at low coupling factors.

Note that these methods only focus on finding the set of coils and capacitors that delivers the most optimal link efficiency η_{link} (Figure 7). The inductive powering system is never considered as a whole, assuming that the driver, rectifier and regulator losses are small compared to the link loss. This assumption is again hard to hold at small coupling. The weak coupling leaves the primary coil driver more or less freewheeling and makes the driver's losses dominant in the overall system's energy budget. Another important and popular assumption is that the coupling does not influence the tank resonance frequencies. The tanks are calculated as separated circuits with a resonance determined by their own L and C. This becomes a problem at the higher end of the coupling scale, where pole splitting occurs.

2.3. Link efficiency

The primary link efficiency is defined as the ratio of the power P_{sec} that reaches the secondary circuit to the power $P_{link\ in}$ put into the inductive link. It is calculated as the ratio of the power dissipated in R_{eq} to the total power dissipated in both R_{eq} and R_{S1} (Figure 8).

Hence,

$$\eta_{primary} \equiv \frac{P_{sec}}{P_{link\ in}} = \frac{R_{eq}}{R_{eq} + R_{S1}}$$

or

$$\eta_{primary} = \frac{k^2 Q_{LS1} Q_{LS2}}{1 + \frac{Q_{LS2}}{\alpha} + k^2 Q_{LS1} Q_{LS2}}$$

where

$$\alpha \equiv \omega_{res} P C_2 R_{load_2}$$

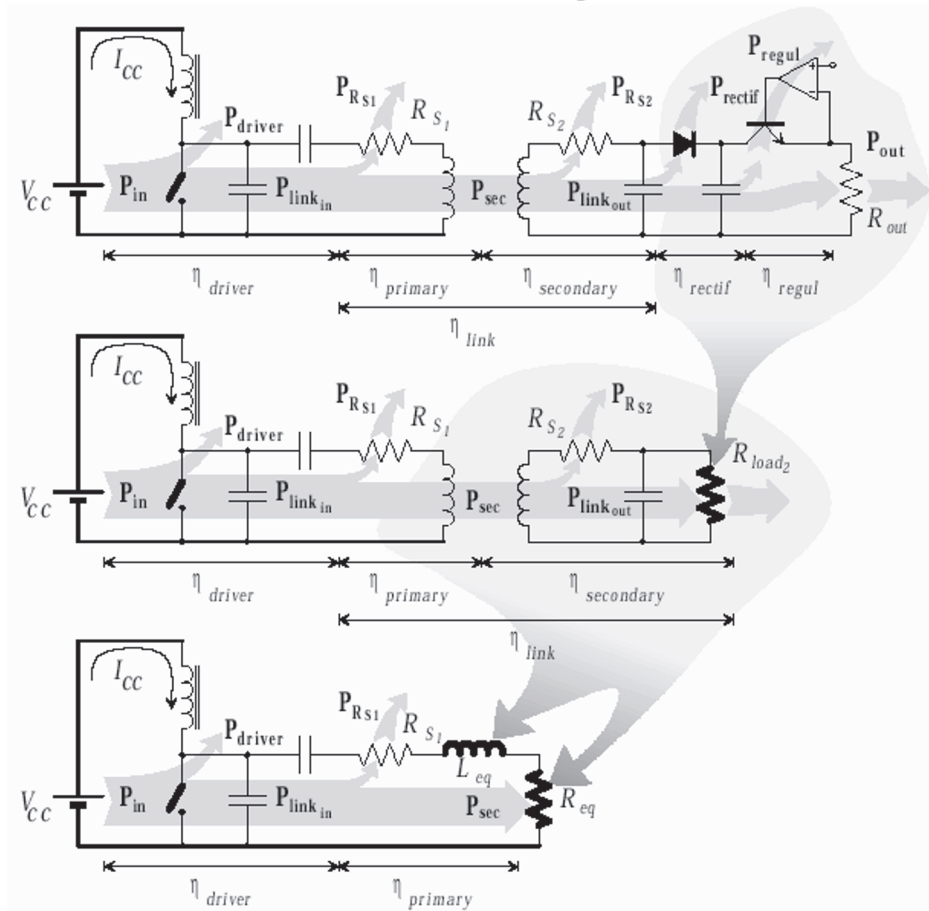


Fig.8: The power distribution in an inductive link. The concept is illustrated on a class-E driven link.

The total link efficiency is the product of the primary and the secondary link efficiencies. The expression is valid for all links with a parallel-resonant secondary, regardless whether the primary coil is tuned or not.

$$\eta_{link} = \frac{k^2 Q_{LS1} Q_{LS2}^2}{\left(1 + \frac{Q_{LS2}}{\alpha} + k^2 Q_{LS1} Q_{LS2}\right) (\alpha + Q_{LS2})}$$

3. Link design

Table 1
Geometric specifications of the inductive link

	Primary coil	Secondary coil
Shape	15 turns pancake/disk	12 turns solenoid
Diameter (mm)	60	20
Nominal distance (mm)	30	30
Operating frequency (kHz)	700	700

A design tool for the calculation of the electrical and magnetic properties of inductive links [15] based on formulae proposed by [16–19] can be implemented in MATLAB. The tool includes the influence of parasitic effects such as skin and proximity effect [20–21] and parasitic capacitance [22]. Together with the link formulae, the design tool can yield the optimization of the inductive link towards maximal efficiency and misalignment tolerance. Table 1 gives an overview of the obtained geometrical specifications. Maximization of the misalignment tolerance results in a pancake/disk shaped primary coil, with a larger diameter than the secondary coil. A rule of thumb that can be obtained from the optimization, is that maximal efficiency is obtained for a primary coil diameter that equals twice the distance between the two coils. The operating frequency is set at 700 kHz in order to avoid biological tissue damage. For the 500 kHz to 4 MHz band, no biological effects have been reported, in contrast to the extreme low frequency (ELF) band and the microwave (MW) band [23,24]. The only possible health risks for the proposed frequency are burns or temperature raise in tissue [25] and electroshocks due to unwanted contacts. As the used power level is lower than the maximal allowed level [26], the risk for burns or temperature raise is minimal and by using appropriate insulation of the coils, electroshocks are avoided.

The electrical and magnetic properties of the link can also be simulated using Finite Element software, such as FASTHENRY [27]. In Table 2, FASTHENRY-simulations, calculations using the MATLAB-scripts and measurements are compared. Q represents the quality factor of a coil, k is the coupling factor of the inductive link.

It can be seen that both simulations and calculations give a good approximation of the measurements. Only for the quality factor of the primary coil Q1, both methods give an overestimation. This is probably due to the fact that the used

formulae and the FASTHENRY-tool are developed for microcoils and not very well suited for larger coils. Although both methods give a comparable result, the MATLAB-scripts are preferred to the FASTHENRY-simulations, as they offer the possibility of being integrated in an optimization loop and both set-up and calculation times are shorter.

Table 2

Simulated, calculated and measured properties for the inductive link

	Simulated	Calculated	Measured
L1 (μH)	9.87	9.39	10.3
Q1	220.0	215.0	137.0
L2 (μH)	3.80	4.15	4.03
Q2	52.0	62.5	45.0
M (μH)	0.31	0.31	0.29
k	0.051	0.050	0.045

The complete inductive powering system is depicted in Fig. 8. The primary coil is driven by a Class E amplifier [28]. The Class E topology was chosen for its high efficiency, which is theoretically 100%. Taking into account the parasitic losses, such as the resistance of the RF Choke L_{RFC} and the on resistance of the switch S, realistic efficiencies of about 80% are obtained.

The secondary circuit consists of a simple rectifier and regulator circuit. The capacitor is added to the secondary coil to form a resonant receiver circuit at the operating frequency. In this way, the power transfer efficiency is increased. The chosen topology consists of a minimal amount of components in order to minimize the dimensions of the implantable circuit.

The input impedance of the implantable monitoring device R_{load} and the rectifier and regulator circuit can easily be transformed into the impedance $R_{\text{load}2}$, as shown in Fig. 8 middle. It can be transformed into the bottom Fig. 8. In this way, the Class E amplifier can be easily designed. Both the primary and the secondary circuit can be built using low cost, commercial off the shelf (COTS) components.

The inductive powering system was designed to deliver 50 mW to the implantable device, at a supply voltage of 5 V. The total power transfer efficiency, which is given by the multiplication of the primary circuit efficiency, the link efficiency and the secondary circuit efficiency, is 36%, while the calculated power transfer efficiency is 44%, mainly due to the parasitic losses of the inductive link. The supply voltage of the Class E amplifier is 2 V.

4. Integration of the bidirectional data communication

Up-link : Absorption modulation. Figure 9 shows the inductive power link, adapted for absorption modulation [29]. A MOS transistor S , used as a switch is introduced in the secondary circuit. By turning the switch S on and off, the input impedance of the secondary circuit is varied. This impedance variation is transmitted in a load variation of the primary circuit. This load variation can be ‘sensed’ by measuring the current through the primary coil $L1$. Therefore, a transformer $L3-L4$ is introduced in the primary circuit. The current variation through the coils $L1-L3$ is now transformed to a voltage variation over $L4$. This signal can then be decoded, using an envelope detector, a high pass filter, a low pass filter and a comparator are used to restore the signal.

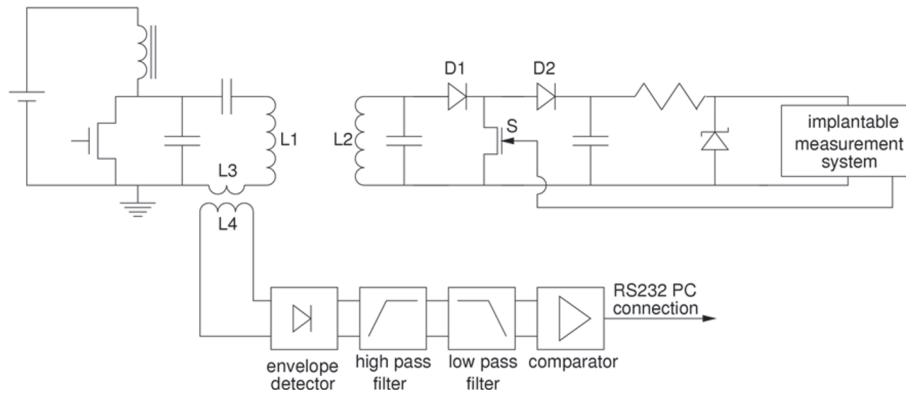


Figure 9. Inductive power link adapted for datatransmission from the implantable measurement system to the base unit, using absorption modulation

A second diode $D2$ is also added in the secondary circuit, to avoid leakage currents to the capacitors. Care has been taken that the absorption modulation does not jeopardize the desired power delivery to the measurement circuit.

The major advantage of using absorption modulation for uplink datacommunication is the low power consumption: the secondary circuit is not acting as a transmitter and power transfer is still possible during data-transmission. The major drawback is a limited transmission range. The gate voltage of S is chosen in such a way that the modulation depth is 5%. Due to the load variation, the Class E amplifier detunes from its ideal operation into a Classes C–E regime [30]: the amplifier only runs in a Class E regime for a single load condition, for which it is originally designed. At other load conditions, the amplifier runs in the lower efficient Classes C–E regime. This reduces the efficiency of the amplifier, but an overall power transfer efficiency of 23.4% can still be maintained.

Down-link : Amplitude modulation. Figure 9 shows the inductive power link, adapted for datatransmission from the base unit to the measurement system. Amplitude modulation is applied to the Class E driver [31], using a MOS transistor S in the driver circuit (acting as a switch) and a resistor R . Amplitude modulation has the advantage of enabling simple encoding and decoding circuits. The major drawback is the decrease in efficiency of the Class E driver. However, this modulation is preferred to frequency or to phase modulation for its basic encoding and decoding circuits, as in this way, the dimensions of the implantable circuit can be kept small. The decoding circuit consists of an envelope detector, a bandpass filter and a comparator. All blocks are built using commercially off-the-shelf components operating at 5 V, and can be powered by the inductive powering system. To make sure that the decoding circuit can be powered from the inductively delivered voltage level, the receiver capacitor is split into a voltage divider $C1$ - $C2$ and the high impedance input of the decoding circuit is connected to it.

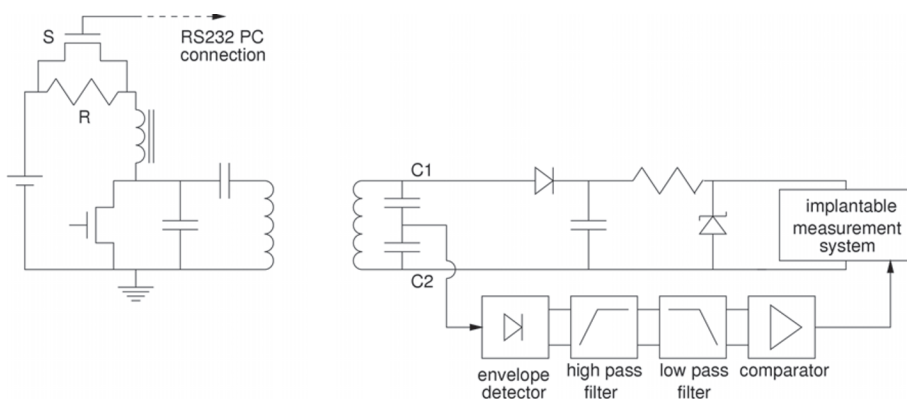


Figure 10. Inductive power link adapted for datatransmission from the base unit to the implantable measurement system, using amplitude modulation

Bi-directional datatransmission. The complete system, capable of powering and bi-directional datatransmission consists of the combination of the circuits shown in Fig. 9 and Fig. 10. In this way, a half-duplex communication link is achieved. The maximal bit rate achieved under test was 60,000 bits/s, with a carrier frequency of 700 kHz. For practical applications, however, the bit rate is set to 19,200 bits/s to match an RS232 link to a PC.

In Fig. 11 the different types of coil misalignment are defined. The decrease in power transfer efficiency of the inductive powering system is caused by the lower mutual inductance M due to the coil misalignment. It can be that this will

5. Misalignment analysis

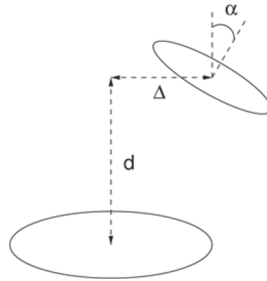


Figure 11. Definition of the misalignment parameters: d = distance, Δ = lateral misalignment, α = angular misalignment.

result in a lower link efficiency η_{link} . The link input impedance Z_{link} will decrease as well, causing the Class E amplifier to run in a Classes C–E regime, resulting in a lower primary circuit efficiency.

Fig. 12 shows the power transfer efficiency of the inductive powering system as a function of the distance between the two coils. It can be seen that the predicted values, obtained from calculations with the above mentioned design tool, give a good approximation of the effective efficiency. The power transfer efficiency during downlink transmission equals the efficiency without data-transmission, while a maximal difference of 50% in efficiency was measured during uplink transmission (absorption modulation). A power transfer efficiency of 30% (15% during uplink data-transmission) is guaranteed for a coil separation of 4 cm.

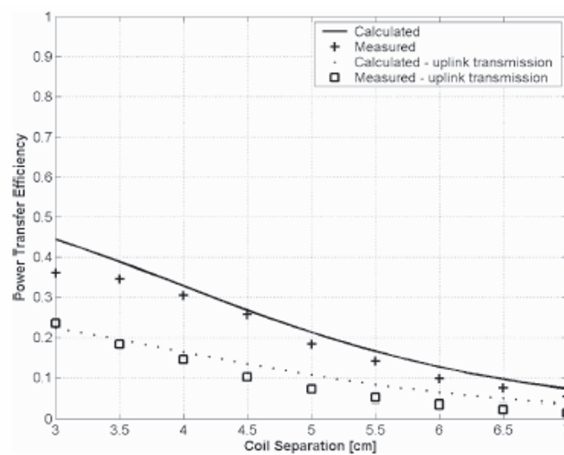


Fig. 12. Power transfer efficiency as a function of separation

In Fig. 13, the power transfer efficiency is plotted versus the lateral misalignment of the two coils. Again, the calculated values give a good prediction of the measured efficiency. For a lateral misalignment of 3 cm (50% of the primary coil radius), a power transfer efficiency of 29% is guaranteed. During uplink transmission, this efficiency falls to 13%.

Fig. 14 gives the measured power transfer efficiency as a function of the angular misalignment, compared to the calculated efficiency. For an angular misalignment of 45° , an efficiency of 29% is guaranteed during powering without data-transmission and with downlink data-transmission and an efficiency of 13% during powering uplink data-transmission.

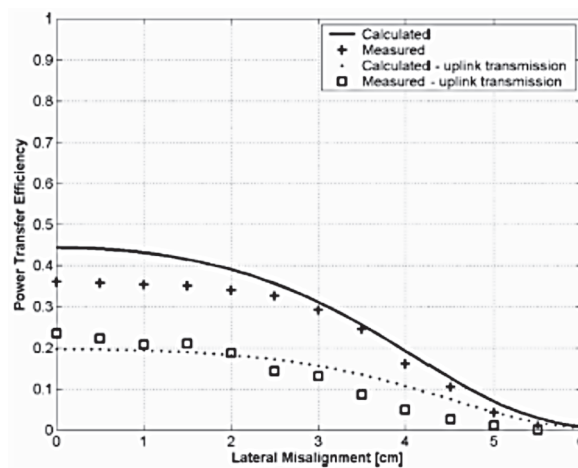


Fig. 13. Power transfer efficiency as a function of lateral misalignment

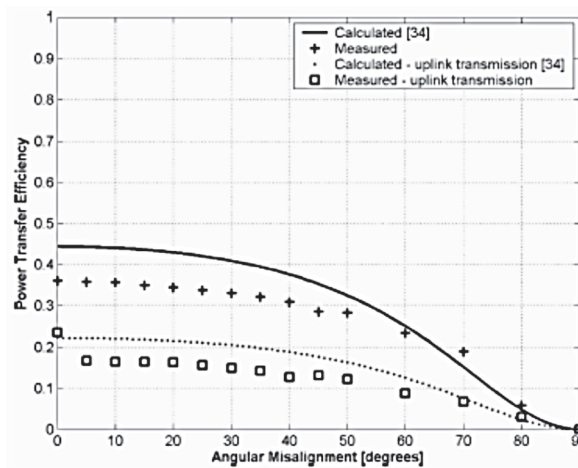


Fig. 14. Power transfer efficiency as a function of angular misalignment

6. Example of such a link

Orthopedic implants like nailplates, hip prostheses, ... can be instrumented with sensors to measure e.g. temperature, overload or fatigue stress on the implant or (unwanted) movement of the implant caused by loosening of the implant.

If these measurements are carried out for a longer period of time, percutaneous links or batteries are avoided to power the implanted measurement system. Instead, inductive links are used to deliver power to the implant. Figure 15 shows a schematic drawing of an inductive powering system that is designed to monitor the fracture healing of a femur bone. By monitoring the stress in the implant, important information on the fracture healing can be obtained. Moreover, the technique allows for a faster and optimal therapy.

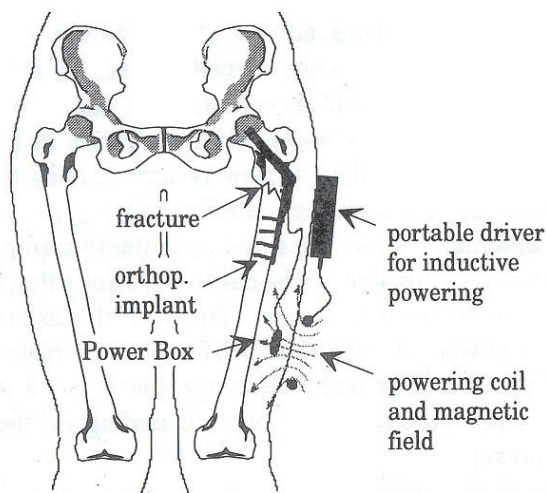


Fig. 15: Inductive powering used for fracture repair monitoring and improved therapy [11]

The coupling consists of a driver (in this case a Class E amplifier), two coils, a rectifier and a regulator. The two coils, of which one is implanted, form a loosely coupled transformer. The externally placed, primary coil is driven by a power amplifier. Figure 16 illustrates the external driver with the (large) flexible outer coil. The coil generates an alternating magnetic field of 700 kHz, which is partly picked up by the secondary, implanted coil. The sinusoidal signal,

received in this way, is then rectified and stabilized at the desired internal supply voltage for the internal load. This load is the electronic read-out and conversion circuitry of the sensors. These consist in strain monitoring devices, the signal of which is then transmitted to the outside world. In Figure 17 the total implant is shown, where the internal coil is housed in a ceramic box, fabricated out of a biocompatible, machinable ceramic (MacorTM). This box contains the secondary coil, rectification and stabilization circuitry. The stress monitor sensor and its interface electronics are housed in a cavity inside the metal implant. Connection between both parts is performed by hermetically sealed feedthrough connections.

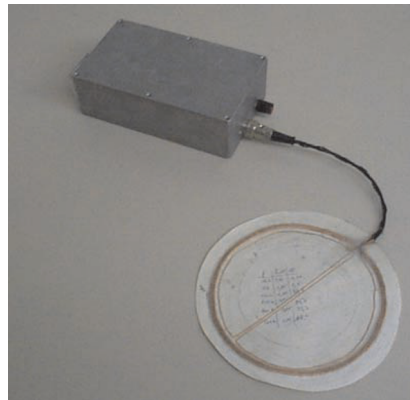


Fig. 16: External power driver (Class E-type) with external coil

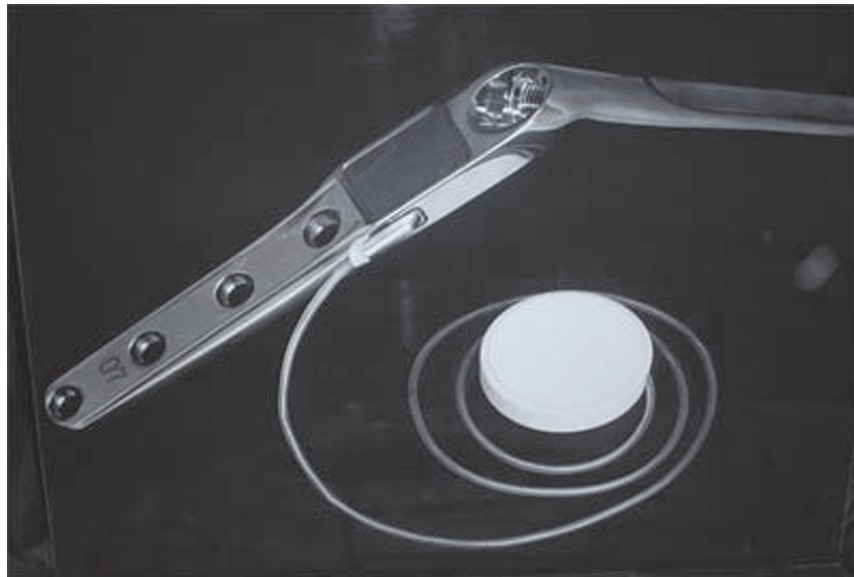


Fig. 17 : The total implant for fracture repair therapy monitoring

This paper presents a unique inductive powering system, that combines power transfer with bi-directional data-transmission. The design of the system has been highly automated, using a self-developed design tool. This design tool gives a good prediction of the effective properties of the system. Although low cost commercial off the shelf components were used, an overall efficiency of 36% was obtained for the delivery of 50 mW over a distance of 30 mm. Moreover, bi-directional data-transmission has been added to the inductive powering system, which makes the system suitable for implantable monitoring and stimulating devices. Future work will focus on the miniaturization of the secondary circuit, resulting in an implantable ASIC.

7. Conclusions

Power delivery to and bi-directional datatransmission with an implantable system were proven to be successful, using the proposed circuit. The use of one single link makes the system easily applicable. Datatransmission from the implantable system to the base unit can be used for the retrieval of measurements and to correct for misalignment, datatransmission to the implantable system can be used to instruct the implantable system (e.g. measurement results on-demand, perform a calibration, stimulation, ...). In a next step, the secondary circuit will be miniaturised, allowing the system to be applied for many biomedical applications such as auditory and visionary aid prostheses, neural prostheses, instrumented orthopedic implants, The system can also be used for industrial applications, such as measurements on rotating parts.

References

- [1] W. Liu, K. Vichienchom, M. Clements, S.C. DeMarco, C. Hughes, E. McGucken, M.S. Humayun, E. de Juan, J.D. Weiland, R. Greenberg, "A neuro-stimulus chip with telemetry unit for retinal prosthetic device", *IEEE J. Solid-State Circuits* 35 (10) (2000) 1487–1497.
- [2] G.J. Suaning, N.H. Lovell, "CMOS neurostimulation ASIC with 100 channels, scaleable output, and bi-directional radio-frequency telemetry", *IEEE Trans. Biomed. Eng.* 48 (2) (2001) 248–260.
- [3] M. Schwarz, R. Hauschild, B.J. Hosticka, J. Huppertz, T. Kneip, S. Kolnsberg, H.K. Trieu, "Single-chip CMOS image sensors for a retina implant system", *IEEE Trans. Circuits Syst. II* 46 (7) (1999) 870–877.
- [4] K. Stangel, S. Kolnsberg, D. Hammerschmidt, B.J. Hosticka, H.K. Trieu,

- W. Mokwa, "A programmable intraocular CMOS pressure sensor system implant", *IEEE J. Solid-State Circuits* 36 (7) (2001) 1094–1100.
- [5] J.A. Von Arx, K. Najafi, "A wireless single-chip telemetry-powered neural stimulation system", *Proc. IEEE Int. Solid-State Circuits Conf.* (1999) 214–215.
- [6] B. Ziaie, M.D. Nardin, A.R. Coghlan, K. Najafi, "A single-channel implantable microstimulator for functional neuromuscular stimulation", *IEEE Trans. Biomed. Eng.* 44 (10) (1997) 909–920.
- [7] T. Akin, K. Najafi, R.M. Bradley, "A wireless implantable multi-channel digital neural recording system for a micromachined sieve electrode", *IEEE J. Solid-State Circuits* 33 (1) (1998) 109–118.
- [8] B. Smith, Z. Tang, M.W. Johnson, S. Pourmehdi, M.M. Gazdik, J.R. Buckett, P.H. Peckham, "An externally powered, multichannel, implantable stimulator-telemeter for control of paralyzed muscle", *IEEE Trans. Biomed. Eng.* 45 (4) (1998) 463–475.
- [9] T. Cameron, G.E. Loeb, R.A. Peck, J.H. Schulman, P. Strojnik, P.R. Troyk, "Micromodular implants to provide electrical stimulation of paralyzed muscles and limbs", *IEEE Trans. Biomed. Eng.* 44 (9) (1997) 781–790.
- [10] F. Graichen, G. Bergmann, A. Rohlmann, "Hip endoprosthesis for in vivo measurement of joint force and temperature", *J. Biomechanics* 32 (1999) 1113–1117.
- [11] F. Burny, M. Donkerwolcke, F. Moulart, R. Bourgois, R. Puers, K. Van Schuylenbergh, M. Barbosa, O. Paiva, F. Rodes, J.B. Bégueret, P. Lawes, "Concept, design and fabrication of smart orthopaedic implants", *Med. Eng. Phys.* 22 (2000) 469–479.
- [12] R.J. Gutmann, "Application of RF circuit design principles to distributed power converters," *IEEE Trans. Ind. Electron. Contr. Instrum.*, vol. IECI-27, pp. 156-164, 1980.
- [13] Miller J.A., G. Bélanger and T. Mussivand, "Development of an autotuned transcutaneous energy transfer system," *ASAIO Journal*, vol. 39, pp. M706-M710, 1993.
- [14] Schuder J.C., H.E. Stephenson and J.F. Townsend, "High-level electromagnetic energy transfer through a closed chest wall," *Inst. Radio Engrs. Int. Conv. Record*, vol. 9, pp. 119-126, 1961.
- [15] K. Van Schuylenbergh, R. Puers, "A computer assisted methodology for inductive link design for implant applications", in: P. Mancini, S. Fioretti, C. Cristalli, R. Bedini (Eds.), *Biotelemetry XII*, Editrice Universitaria Litografia Felici, Pisa, 1993, pp. 392–400.
- [16] M. Soma, D.C. Galbraith, R.L. White, "Radio-frequency coils in implantable devices: misalignment analysis and design procedure", *IEEE Trans. Biomed. Eng.* 34 (4) (1987) 276–282.
- [17] E.S. Hochmair, "System optimization for improved accuracy in tran-

- scutaneous signal and power transmission”, *IEEE Trans. Biomed. Eng.* 31 (2) (1984) 177–187.
- [18] C.M. Zierhofer, E.S. Hochmair, “Geometric approach for coupling enhancement of magnetically coupled coils”, *IEEE Trans. Biomed. Eng.* 43 (7) (1996) 708–714.
- [19] S. Babic, C. Akyel, “Improvement in calculation of the self- and mutual inductance of thin-wall solenoids and disk coils”, *IEEE Trans. Magn.* 36 (4) (2000) 1970–1975.
- [20] P. Ravazzani, J. Ruohonen, G. Tognola, F. Anfosso, M. Ollikainen, R.J. Ilmoniemi, F. Grandori, “Frequency-related effects in the optimization of coils for the magnetic stimulation of the nervous system”, *IEEE Trans. Biomed. Eng.* 49 (5) (2002) 463–471.
- [21] B.L. Ooi, D.X. Xu, P.S. Kooi, F.J. Lin, “An improved prediction of series resistance in spiral inductor modeling with Eddy-current effect”, *IEEE Trans. Microwave Theory Techn.* 50 (9) (2002) 2202–2206.
- [22] A. Massarini, M.K. Kazimierczuk, “Self-capacitance of inductors”, *IEEE Trans. Power Electron.* 12 (4) (1997) 671–676. [23] S.F. Cleary, Cellular effects of electromagnetic radiation, *IEEE Eng. Med. Biol.* 6 (1) (1987) 26–30.
- [24] R.J. Smialowicz, “Immunologic effects of nonionizing electromagnetic radiation”, *IEEE Eng. Med. Biol.* 6 (1) (1987) 47–51.
- [25] E.R. Adair, “Thermophysiological effects of electromagnetic radiation”, *IEEE Eng. Med. Biol.* 6 (1) (1987) 37–41.
- [26] M.H. Repacholi, “Radiofrequency electromagnetic field exposure standards”, *IEEE Eng. Med. Biol.* 6 (1) (1987) 18–21.
- [27] M. Kamon, M.J. Tsuk, J. White, FASTHENRY: “A multipole-accelerated 3-D inductance extraction program”, *IEEE Trans. Microwave Theory Techn.* 42 (9) (1994) 1750–1758.
- [28] M.K. Kazimierczuk, D. Czarkowski, “Resonant Power Converters”, John Wiley & Sons, New York, 1995, pp. 347–365.
- [29] P.A. Neukomm, H. Kündig, “Passive wireless actuator control and sensor signal transmission”, *Sens. Actuators A* 21–23 (1990) 258–262.
- [30] M.K. Kazimierczuk, W.A. Tabisz, “Classes C–E high-efficiency tuned power amplifier”, *IEEE Trans. Circuits Syst.* 36 (3) (1989) 421–428.
- [31] M. Kazimierczuk, “Collector amplitude modulation of the Class E tuned power amplifier”, *IEEE Trans. Circuits Syst.* 31 (6) (1984) 543–549.