Analog Circuit Design

High-Speed A-D Converters, Automotive Electronics and Ultra-Low Power Wireless

Edited by
Arthur H.M. van Roermund
Herman Casier
Michiel Steyaert

Springer

ANALOG CIRCUIT DESIGN

Analog Circuit Design

High-Speed A-D Converters, Automotive Electronics and Ultra-Low Power Wireless

Edited by

ARTHUR H.M. VAN ROERMUND
Eindhoven University of Technology,
The Netherlands

HERMAN CASIER
AMI Semiconductor, Vilvoorde,
Belgium

and

MICHIEL STEYAERT
Katholieke Universiteit Leuven,
Belgium

Springer

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

ISBN-10 1-4020-5185-9 (HB)
ISBN-13 978-1-4020-5185-2 (HB)
ISBN-10 1-4020-5186-7 (e-book)
ISBN-13 978-1-4020-5186-9 (e-book)

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

www.springer.com

Printed on acid-free paper

Table of Contents

Preface

This book is part of the *Analog Circuit Design* series, and comprises three chapters, each discussing a very relevant topic that receives a lot of attention in analog circuit design at this moment:

1. High-Speed AD Converters
2. Automotive Electronics: EMC issues
3. Ultra-Low Power Wireless

The book contains eighteen tutorial papers, six per chapter. These papers are written by six experts in each field, reflecting their presentations held at the 15th workshop on Advances in Analog Circuit Design (AACD), Maastricht, The Netherlands, April 2006. The workshop was organised by Neil Bird of Philips Research The Netherlands, and the program committee consisted of Arthur van Roermund from Eindhoven University of Technology, The Netherlands; Herman Casier of AMI Semiconductor Belgium, and Michiel Steyaert from Katholieke Universiteit Leuven, Belgium; who are also the editors of this book.

This book is number 15 in the successful series of Analog Circuit Design, providing valuable information and excellent overviews of analog circuit design and related CAD, mainly in the fields of basic analog modules, mixed-signal electronics, AD and DA converters, RF systems, and automotive electronics. For the previous books in the series, see next page.

Analog Circuit Design is an essential reference source for analog circuit designers and researchers wishing to keep abreast with the latest developments in the field. The tutorial coverage also makes it suitable for use in an advanced design course.

We hope that also this book will provide a valuable contribution to our Analog Circuit Design community.

Arthur van Roermund

Previous Books in *Analog Circuit Design*

The series on Analog Circuit Design covers the topics of previous workshops on Advanced Analog Circuit Design (AACD), as listed in the following table:

2005	Limerick (Ireland)	RF Circuits: Wide Band, Front-Ends, DACs Design Methodology and Verif for RF and M/S Systems Low Power and Low Voltage
2004	Montreux (Swiss)	Sensor and Actuator Interface Electronics Integrated High-Voltage Electronics and Power Mgt Low-Power and High-Resolution ADCs
2003	Graz (Austria)	Fractional-N Synthesizers Design for Robustness Line and Bus drivers
2002	Spa (Belgium)	Structured Mixed-Mode Design Multi-Bit Sigma-Delta Converters Short-Range RF Circuits
2001	Noordwijk (The Netherlands)	Scalable Analog Circuits High-Speed D/A Converters RF Power Amplifiers
2000	Munich (Germany)	High-Speed A/D Converters Mixed-Signal Design PLLs and Synthesizers
1999	Nice (France)	XDSL and other Communication Systems RF-MOST Models and Behavioural Modelling Integrated Filters and Oscillators
1998	Copenhagen (Denmark)	1-Volt Electronics Mixed-Mode Systems LNAs and RF Power Amps for Telecom
1997	Como (Italy)	RF A/D Converters Sensor and Actuator Interfaces Low-Noise Oscillators, PLLs and Synthesizers

1996	Lausanne (Swiss)	RF CMOS Circuit Design Bandpass Sigma Delta and Other Data Converters Translineair Circuits
1995	Villach (Austria)	Low-Noise/Power/Voltage Mixed-Mode with CAD tools Voltage, Current and Time References
1994	Eindhoven (Netherlands)	Low-Power Low-Voltage Integrated Filters Smart Power
1993	Leuven (Belgium)	Mixed-Mode A/D Design Sensor Interfaces Communication Circuits
1992	Scheveningen (The Netherlands)	OpAmps ADC Analog CAD

PART I: HIGH-SPEED AD CONVERTERS

The first chapter of this book is on High-Speed AD Converters. It addresses both generic high-speed design issues and specific design examples. 'High-speed' can be interpreted in two ways. It can mean in absolute sense the most high-speed converters, like addressed in the first two papers. However, it can also have a meaning in relative sense: those converters where the speed is dictated by the application, but where the dynamic properties dominate the performance. Depending on the application specifications, several architectures come into the picture. The next four papers address design techniques and design examples, for different type of architectures, that shift the speed barriers towards higher values by alleviating the analog requirements, by correction and calibration in the digital domain, or by giving a better understanding and modeling, thus paving the way to more optimized designs.

The first two papers address design aspects of really high-speed ADCs, with 22GS/s and 20GS/s design examples; these papers include a more general and tutorial like review. The first one, of Peter Schvan, addresses a time-interleaved flash converter, operating at 22GS/s with 5 bit resolution, made in a bipolar technology. Sources of degradation at high speed are discussed before the design is elaborated. Bipolar has for several reasons been the technology of choice for very high-speed converters up till shortly. The next paper, of Ken Poulton, addresses the Bipolar/CMOS trends and the shift in the last five years to CMOS. The pipelined converter slices of the time-interleaved architecture are made in full-CMOS; a separate SiGe chip comprises the only function left for bipolar in this design: the buffering.

As suggested earlier, time-interleaved flash and time-interleaved pipelines are not the only architectures that are considered nowadays. Other architectures like successive approximation converters (SAR), subranging converters, time-inteleaved folding converters, and sub-harmonic limit-cycle sigma-delta (SLC-SDM) conversion will be addressed too, in the next papers.

Dieter Draxelmayr, in his paper, demonstrates similarities and differences between pipelined, algorithmic, and SAR converters. He addresses the various algorithmic and circuit level adaptations that can improve speed, and the calibration methods that can improve the accuracy while alleviating the requirements on the building blocks and technology.

The next paper, from Frank van der Goes, presents again a different architecture: the two-step subranging ADC. He addresses the design of a family of converters ranging from 8 bit to 10 bit at sampling rates from 50MS/s up to 200MS/s.

Robert Taft focuses on various ways of offset calibration and applies it to a 2x interleaving 8bit folding interpolating flash ADC operating at 1.6GS/s, where he solves the offset problem with a one-time foreground calibration.

Finally, we have a paper on the sigma-delta type of converters. These converters have several beneficial properties, but the main drawback is the limited bandwidth due to the required oversampling ratio. In this paper it is shown that also this speed boundary can be shifted, by a better understanding and modeling of the non-linear loop, especially by understanding the limit cycles. By using a new mode of operation, Subharmonic Limit-Cycle SDMs are obtained that achieve significantly improved performance.

Arthur van Roermund

22GS/s ADCs – IMPLEMENTATION CHOICES AND PERFORMACE TRADE-OFFS

Peter Schvan
Nortel, Ottawa, Canada

Abstract

Implementation options for very high speed ADCs are discussed by first reviewing performance limiting mechanisms then comparing the various architectures and building blocks typically used in these converters. Results for a 5b 22GS/s ADC are presented.

1. Introduction

Analog to digital converters are typically classified according to the input frequency range they are intended to operate in and the resolution they achieve over that frequency range. Since operating at increasing input frequencies require different design approaches it is possible to identify regions of operation that are addressed with a certain class of ADCs. For each class of ADC architectures and performance targets the designer faces a set of unique issues. Here I would like to review the challenges when designing for the highest operating frequency. I will review available architectures and circuit options and describe a particular implementation of a 22GS/s converter.

Applications for such high speed converters include data acquisition systems, semiconductor test equipments, radar and fiber optic transmission systems and other telecommunication applications. Many of these applications will pose additional requirements that will set limits to such secondary parameters as power dissipation size and BER. In particular economical consideration and the availability of limited board area make it desirable to integrate one or more ADCs with subsequent digital signal processing imposing limits on power dissipation and even on the type of technology available for the designer.

2. Sources of Performance Degradation

Before discussing suitable architectures for very high speed ADCs it is instructive to review the sources of fundamental limitations on ADC

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 3–15.

performance. Fig. 1 shows maximum achievable resolution when considering the impact of thermal noise, clock jitter and decision ambiguity. The effective number of bit, ENOB, was calculated using the well know relationship ENOB=(SNR-1.76)/6.02 where SNR is calculated for the different noise sources. The input-referred thermal noise voltage is:

$$V_{thermal}=\sqrt{4kT\,R_{eff}F_{in}} \tag{1}$$

where k is the Boltzmann's constant, T is temperature and R_{eff} is the effective thermal resistance, and F_{in} is the maximum input frequency. Here a pessimistically large R_{eff} of 1Kohm was used to represent all sources of thermal and shot noise at the input of the converter integrated up to the input frequency.

While thermal noise is unlikely to ever become a concern at frequencies above 1GHz it is inaccurate sampling that is typically considered as the most important cause of performance degradation. Clock jitter causes the ADC not to sample the input signal at precisely equal time intervals. Noise voltage as a result of clock jitter is:

$$V_{jitter}=\sqrt{2}\;\pi A F_{in}\tau_{jitter} \tag{2}$$

Here τ_{jitter} represents the rms clock jitter and A is the full scale input amplitude. The graph shows the resolution limits for 0.5 and 1ps rms values. If clock jitter is not reduced below 1ps resolution is limited to 6 bits for a 4GHz input signal. Another consequence of increasing sampling rate is that not enough time is available for the comparator to produce the necessary logic levels for the encoder circuit. When a comparator makes an ambiguous decision it can be treated as an additive noise. This noise can be approximated [1] as:

$$V_{ambiguity}=\sqrt{\frac{2AV_L}{3G}}\;exp\left(\frac{-1}{8F_{in}\tau_{latch}}\right) \tag{3}$$

Here V_L is the required logic level, G is the DC gain of the comparator and τ_{latch} is the regeneration time constant of the latch in the comparator. Assuming that the ADC is operating at Nyquist rate the time available for the latch to produce an output is $1/4F_{in}$. The time constant τ_{latch} can be approximated with the empirical formula $\tau_{latch}=2.5/\pi F_t$ where for this calculation F_t is assumed to be 150GHz. This time constant can be effectively increased by increasing the number of latches in the comparator. The graph shows the impact of the decision ambiguity related noise for single and double latch comparators. Although the model describing the latching process is highly simplified these results demonstrate that above 10 GHz decision ambiguity could become the resolution limiting mechanism.

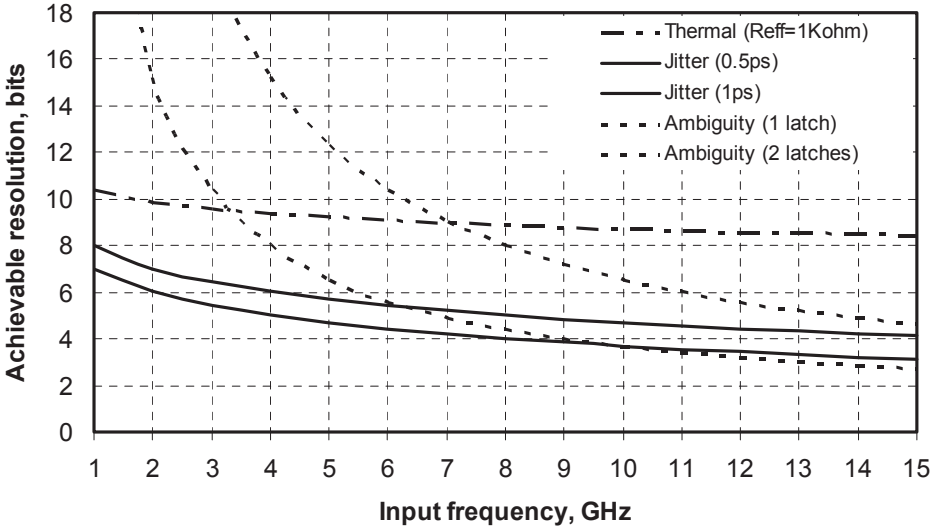


Fig.1. Performance limitation due to various sources of noise.

Apart from these sources of noise that add to quantization noise we also have to consider the impact of harmonic distortion when evaluating the performance of an ADC. The most important source of non-linear behavior prior to the decision process is signal dependent delay that leads to timing skew between the input signal and the sampling clock.

2. Very High Speed ADC Architectures

With currently available semiconductor technologies high speed device performance limits gate delay and settling time so that for sampling rates of several GS/s no multi step ADC architecture can be used restricting designers to the well known flash converter approach. While folding converters are similar to flash converters they place further limitation on the maximum input bandwidth. The only architecture that relaxes the demand on device speed is the time-interleaved converter. As shown in Fig. 2 the input data is sub-sampled with a series of N S/H circuits using phase shifted clocks, acting as an analog de-multiplexer, so that subsequent A to D conversion can take place in parallel using lower speed converters operating at 1/N of the target sampling rate.

In practical implementations it turns out that the open-loop generation of multi-phase clocks is not possible with the phase accuracy required for very high input frequencies. Therefore it requires a sophisticated phase adjustment feedback circuitry that involves processing the generated digital output to detect any phase mismatch so that it could be corrected continuously as long as it takes place at low speed. Similarly the large number of S/Hs and sub

ADCs in a practical implementation inevitable leads to gain and offset mismatch between the different parallel channels which, if uncorrected, lead to harmonic distortion. This again can be overcome with an appropriate feedback algorithm that compensates for these variations to achieve the desired performance. Time-interleaved ADC [2] achieved 20GS/s sampling rate producing 4.6 ENOB for 6GHz input at the expense of relatively high power dissipation required for the DSP. The following discussion will concentrate on flash type architectures.

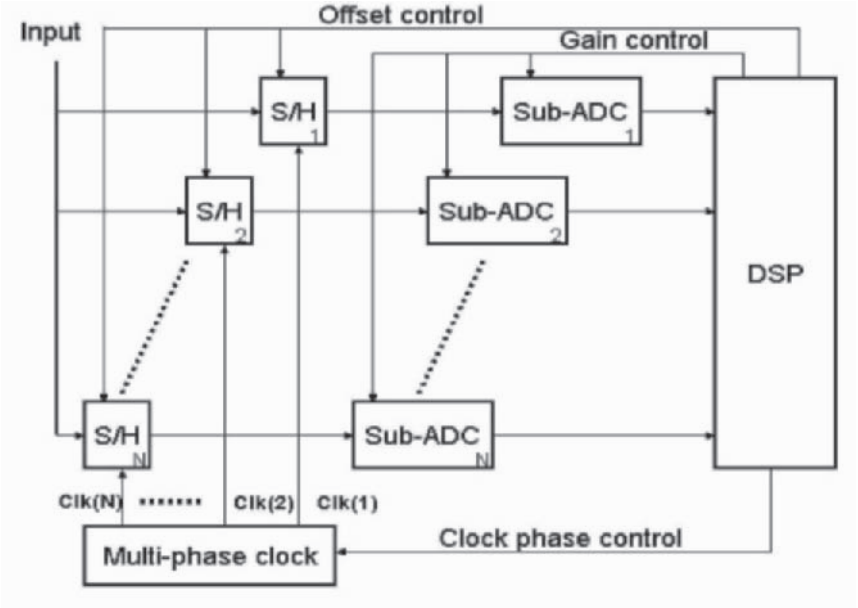


Fig.2. Time-interleave ADC architecture.

4. Critical Building Blocks

The circuit blocks that determine to a large extent the performance of an ADC include the S/H circuit, comparator and error correction scheme used. Next we will discuss some of the issues related to the selection of these blocks.

4.1 Sample and Hold Circuits

In a flash ADC the input signal is connected to a number of comparators which then make decisions on the signal level in parallel. It is usually beneficial to place a S/H circuit in front of these comparators to eliminate timing or rapidly changing signal related distortion. Examples for such S/H

circuits are shown in Fig. 3. S/H can be classified based on the type of device they employ to perform switching, accordingly we can talk about MOS, bipolar or diode based architectures.

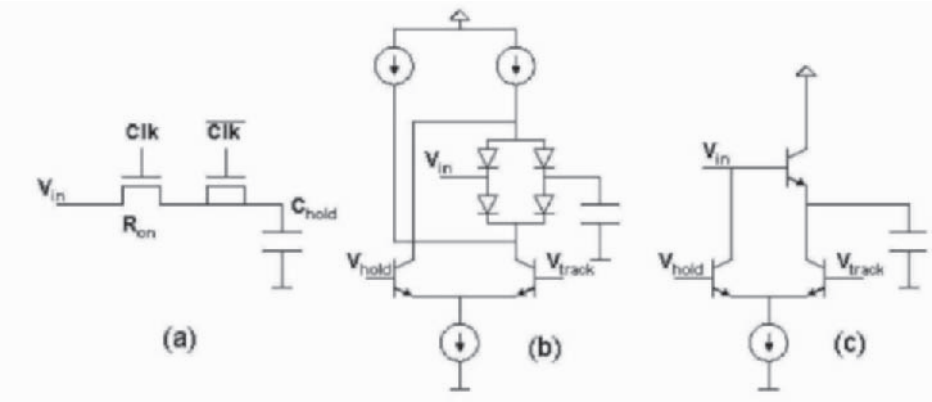


Fig.3. S/H circuit examples (a) MOS switch, (b) diode bridge, (c) switched emitter follower.

While MOS based S/H circuits require the lowest complexity they suffer from channel charge injection, clock feedthrough, input dependent sampling instance and high on resistance. While clock feedthrough can be mitigated with compensation techniques it limits the ratio between the size of the switching transistor and hold capacitor, C_{hold}. Voltage dependence can be reduced by bootstrapping, although that becomes difficult at high speed or by limiting the common mode and swing of the input signal. On resistance, however sets a maximum limit on the sampling rate of MOS switches. For practical applications on resistance, R_{on}, is limited to about 200ohm for a hold capacitance, of 100fF. That sets a minimum settling time of $5R_{on}C_{hold}$ to 100ps, assuming 5bit accuracy, which implies a maximum sampling rate of 5GS/s. Published CMOS S/H circuits have remained well below that level [3,4] around 1GS/s. Unfortunately shrinking CMOS technologies don't offer much relief since shorter gate length and thinner gate oxide with un-scaleable turn-on threshold voltage drastically limits the maximum voltage that can be applied to the switch and also the achievable input dynamic range. Importantly, however, in time-interleaved ADC where while sampling precision is important, clock cycle time is increased making MOS switches a viable option.

If a bipolar transistor is available switched emitter follower can achieve higher sampling rate. Here the main factors that affect the signal accuracy at high sampling rates are input buffer nonlinearity and signal dependent switch bias current. The performance limitation of the switched emitter follower results from the track-mode distortion when the bias current modulates its base emitter voltage by charging and discharging the hold capacitor. While an InP DHBT version of the T/H requiring 5.2V supply demonstrated 5b,

12GS/s performance [5] a SiGe bipolar implementation including the subsequent flash ADC achieved only 2.8b for 5GHz input and 10GS/s sampling rate [6].

The performance of S/H using a diode bridge has been analyzed in [7]. Sampling rate of 4GS/s with 6b accuracy was demonstrated [8] while using an experimental SiGe technology (Ft=375GHz) even a 40GS/s sampling rate for a 3b converter was reported [9]. Although a diode bridge has the potential to operate at very high sampling rates high speed Schottky diodes are rarely available making it an impractical choice.

4.2 Comparator

Since it appears that S/H operating around 20GS/s sampling rate is difficult to implement at best or may not even be possible with mainstream Si technologies currently available the remaining option is a flash converter without S/H.

Conceptually it should be easier to make a decision on the signal level at a given instance rather than to hold its precise value at the sampling instance. When no S/H is used the input and reference signals are directly connected to a set of parallel comparators which usually includes input amplifier and a decision circuit in the form of latch. In this type of flash ADC there is an increased demand placed on the input stage of the comparator. It should minimize loading on the input signal for maximum bandwidth, produce the difference between a set of reference values and the input signal, while preventing the input to couple to the reference network, it should also isolate the digital switching transients or “kickback” of the latches from the input and finally produce the zero transition to the decision block with a delay that is independent of the input signal.

Figure. 4 shows an input buffer where reference levels are provided in form of a differential current input [10]. The difference between input signal and reference signal is produced by emitter followers whose outputs are level shifted using the differential reference currents. The distortion that would result from varying emitter follower current and associated V_{be} drop is compensated by the voltage to current converter where the opposite V_{be} variation takes place. This design also provides a high level of isolation between the input and the resistive ladder compared to implementations where the input of the comparator is directly connected to the reference ladder. Bipolar design also guaranties very low offset in the order of 1-2mV that is negligible compared to typical LSB values.

The amplifier after the level shifter input stage is a modified Cherry-Hooper design [11]. The delay of the zero transition is the result of the limiting characteristics of the amplifier coupled with limited bandwidth. This delay dependence was investigated in [12]. It was demonstrated that third order distortion that results from the nonlinear characteristics of the amplifier

depends on the ratio between the full scale input of the ADC and the output linear range of the amplifier, and also on the ratio between the bandwidth and input frequency. This design can provide both high bandwidth and large output linear range.

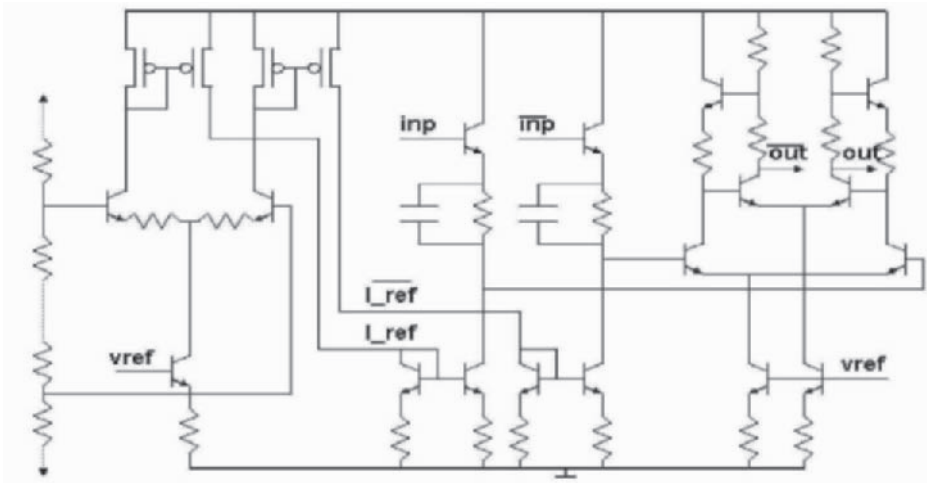


Fig.4. Pre-amplifier with differential current reference.

In high speed ADC the bit error rate (BER) is an important design criterion. As mentioned earlier errors occur as a result of metastability when in case of a small input signal at start of the latching process there is no sufficient time available for the flip-flop to produce the minimum required logic level for the encoding logic. While metastability depends primarily on the time constant, τ_{latch}, of the latch it can be mitigated by including more latches in the decision circuit which effectively increases the time available for producing a valid logic level at the expense of increased power dissipation. Since 2^N decision circuits are required, where N is the number of bits, this represents the largest contribution to the total power dissipation of the ADC.

4.3 Error Correction

Comparators in a flash converter generate what is commonly known as a "thermometer code". Due to extreme clock skew or other noise it is possible that these outputs produce an "illegal" combination of values or bubble. Standard method of guarding against it relies on a voting circuit that can verify the validity of the output and correct bubble errors [13]. Unfortunately these logic circuits require two consecutive logic operations that can not be performed at 22GHz clock rate without excessive power dissipation. Gray coding has the property to fail benignly in the presence of bubbles or even some of the invalid logic states that results from metastability. Gray

encoding ensures that a metastable output is passed on as a single unsettled bit so that only N latches are required (instead of 2^{N-1}) to further reduce error probability.

Since the usual ROM based encoding schemes is too slow for high conversion rates an analog summing approach [14] can be employed to perform the thermometer to Gray code conversion. The digital outputs of the comparators are converted to differential currents that are summed with alternating polarity. Gray code LSB is obtained by adding every second output 1, 3, 5 …31, in case of a 5 bit converter, the second LSB from outputs 2, 6 …30, and so on while the MSB is from output 16. The speed of this conversion is limited by the capacitance of the summing node associated with generating the LSB output due to the large number of interconnects. Finally, the binary output can be generated using exclusive OR gates after retiming the Gray code outputs.

5. A 5b 22GS/s ADC Example

The 22GS/s 5b ADC reported in [15] uses comparator and error correction scheme described above. The ADC has a fully differential architecture for noise rejection shown in Fig. 5.

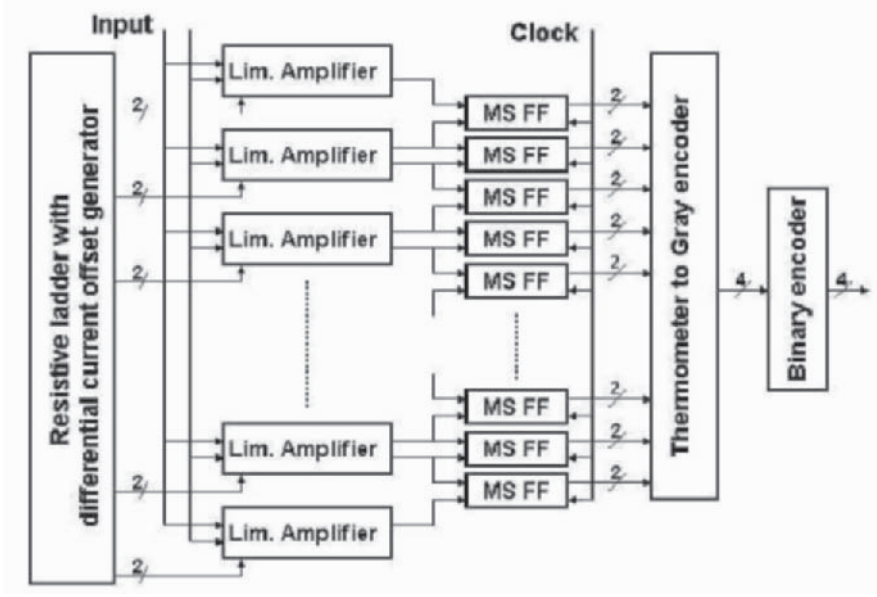


Fig.5. Architecture of the 22GS/s 5b ADC.

To achieve sufficiently high input bandwidth and minimize the length of input bus and therefore minimize clock skew it is advantageous reduce the number of amplifiers connected to that buss. Inputs to the comparators generating the LSB are produced through interpolation between the outputs of the amplifiers as shown in Fig. 5.
The ADC described here has been fabricated in a 0.13μm SiGe BiCMOS technology with 150GHz Ft bipolar transistors. MOS devices were only used in the bias circuits and for implementing the summing nodes for Grey code generation.

5.1 Experimental Results

To simplify testing a 6bit DAC [16] and linear output buffer with combined 35dB SFDR at 8GHz are integrated with the ADC allowing the reconstruction of the output signal. This test circuit is flip chip mounted directly onto a circuit board together with a linear AGC amplifier that generates the differential input for the ADC. The reconstructed outputs are fed into a sampling scope and a spectrum analyzer.

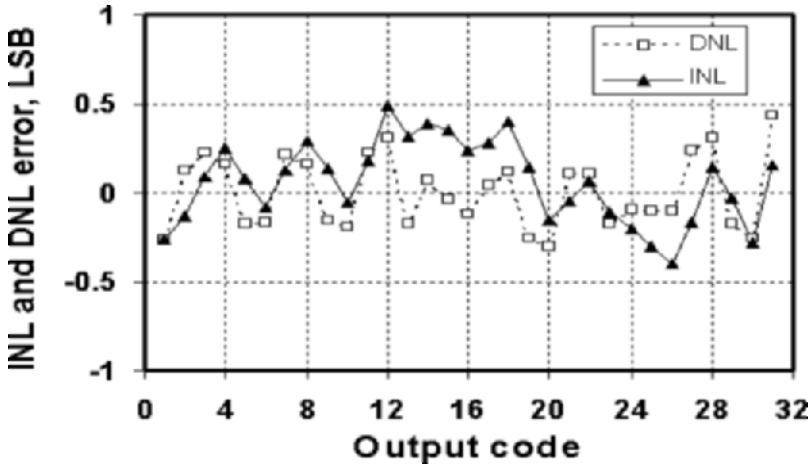


Fig.6. Measured INL and DNL at 22GS/s with 1GHz input signal.

INL and DNL were tested using standard code density measurement at 22GS/s sampling rate while driving the ADC with a full scale 1GHz sine wave input. Fig. 6 shows that both DNL and INL are below 0.5 LSB.

A typical output spectrum of the reconstructed signal is shown in Fig. 7. The spectrum is analyzed after correcting it for the sin(x)/x roll-off of the DAC and RF signal loss in the cables and connectors. The highest spur is the third harmonics of the input signal at -35 dB. For flash converters without S/H the main sources of SNRD degradation are clock jitter or skew and distortion caused by signal dependent delay and finite aperture time.

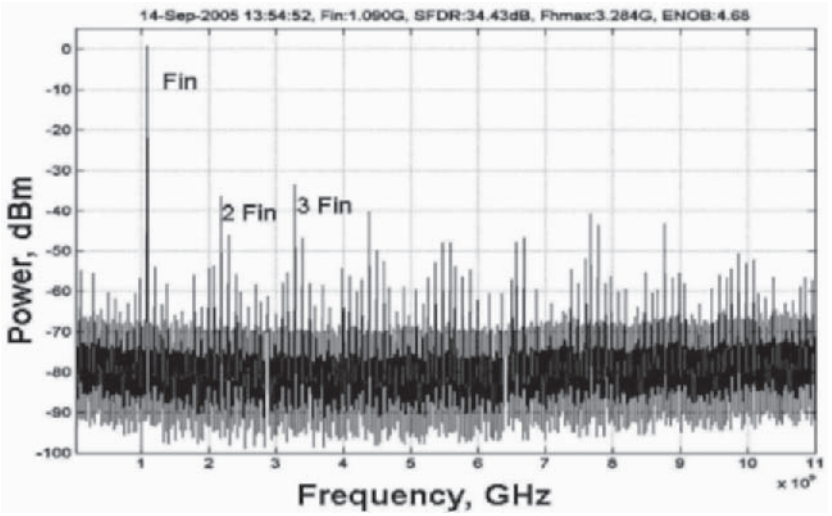


Fig.7. Frequency spectrum at 1GHz input and 22GS/s sampling rate.

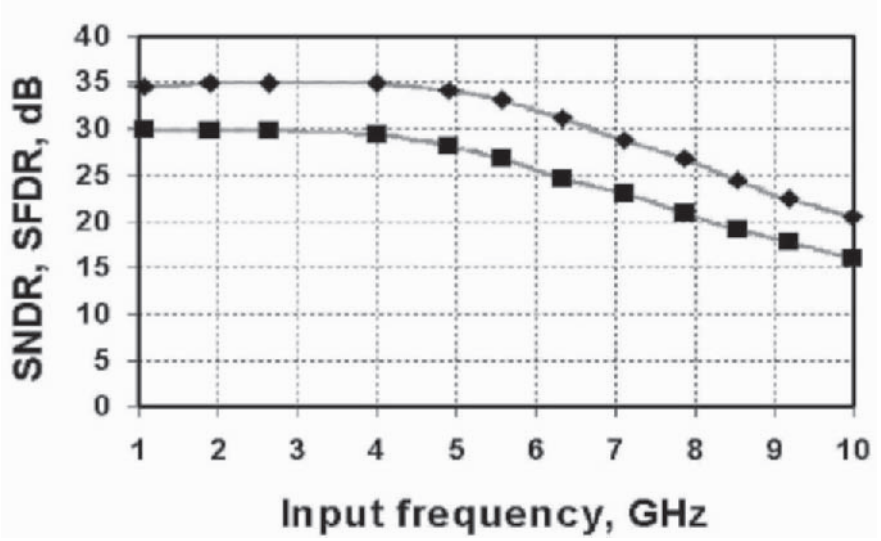


Fig.8. Measured SNDR and SFDR as a function of input frequency.

Figure 8 shows the measured SNRD and SFDR values for a series of input frequencies. These SNRD results indicate that ENOB stays above 4 up to 6GHz then drops to 3.5 at 7GHz input frequencies respectively. Clock jitter, that includes variation in propagation delay between data and clock signals, becomes noticeable between 4-5GHz. For input frequencies above 6GHz spurs related harmonics dominate SNRD as indicated by the parallel decline of SNRD with SFDR.

Bit error rate is measured by observing the reconstructed output on a sampling oscilloscope in accumulation mode that allows the counting of samples outside of the expected range. Fig. 9 shows the bit error rate for different sampling frequencies for a 1GHz input signal. Error rate increases with sampling frequency reaching 10^{-4} at 22GS/s which is within the acceptable range for the intended application. BER performance could be improved by reducing metastability of the comparators by adding latches to the decision circuit.

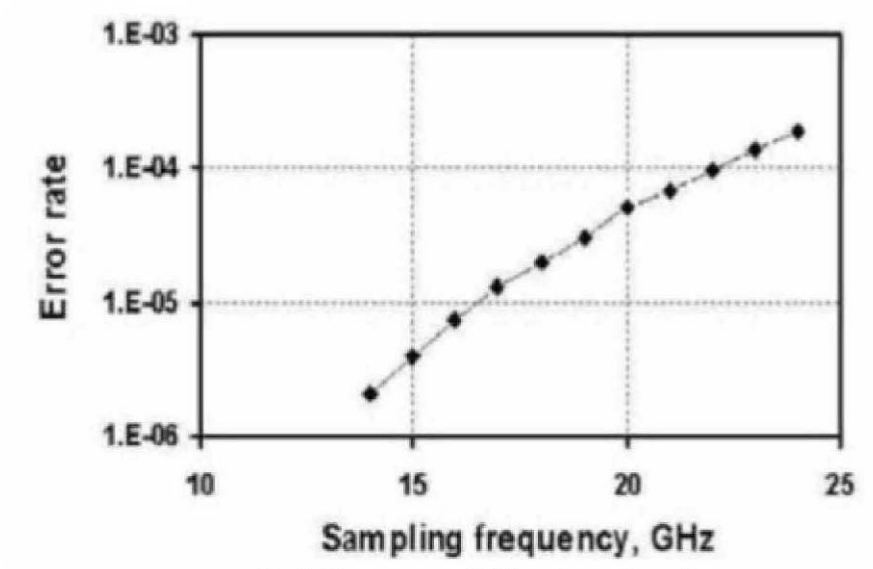


Fig.9. Bit error rate at different sampling rates.

Figure 10 shows the die photograph of the ADC. Power dissipation is 3W operating from a 3.3V supply. The layout is designed to ensure that both the data and clock reach the different comparators simultaneously. The comparators were arranged in a U shape around the input data distribution network to minimize clock and data propagation delay.

This is the first 5b 22GS/s ADC with sufficiently reduced power dissipation to be suitable for integration with a DSP for dispersion compensation in fiber optic transmission systems.

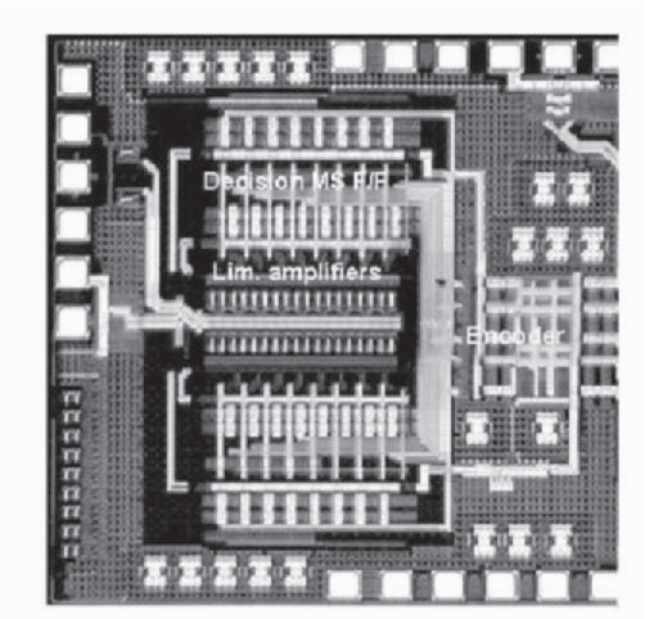


Fig.10. Die photograph of the 5b 22GS/s ADC.

6. Conclusions

Options for implementing ADCs with very high sampling rate have been reviewed. It was found that with current mainstream semiconductor technologies the two promising approaches are time-interleaved architecture or flash converter without S/H circuit. It was demonstrated that such a flash converter can achieve 22GS/s sampling rate and ENOB of 4 up to 6 GHz input frequency while consuming 3W.

References

[1] L.E. Larson, "High-speed analog-to-digital conversion with GaAs technology: Prospects, trends, and obstacles, 14th European Solid-State Circuits Conf. pp. 2871-2878, 1988.

[2] K. Poulton, et al., "A 20GS/s 8b ADC with a 1MB Memory in 0.18um CMOS," ISSCC Dig. Tech. Papers, pp. 318-319, February 2003.

[3] C. K. Ken, et al., "A Serial-Link Transceiver Based on 8-GSample/s A/D and D/A Converters in 0.25-u, CMOS," J. of Solid-State Circuits, pp. 1684-1692, vol. 36, No 11, 2001

[4] X. Jiang, et al., "A 2GS/s 6b ADC in 0.18um CMOS", ISSCC Dig. Tech. Papers, pp. 122-123, Feb. 2003.

[5] J. Lee, at al., "A 6b 12-GSample/s Track-and-Hold Amplifier in InP

DHBT Technology," J. of Solid-State Circuits, pp. 1533-1539, vol. 38, no. 9 2003

[6] J. Lee, et al., "A 5-b 10-GSample/s A/D Converter for 10-Gb/s Optical Receivers," GaAs IC Symposium, pp. 204-207, 2003.

[7] J. Jensen and L. E. Larson, "A Broadband 10-GHz Track-and-Hold in Si/SiGe HBT Technology," J. of Solid-State Circuits, pp. 325-330, vol. 36, no. 3, 2001

[8] K. Poulton, et al., "A 6-b, 4 GSa/s GaAs HBT ADC," J. of Solid-State Circuits, pp.1109-1118, vol. 30, No 10, 1995

[9] W. Cheng, et al., "A 3b 40GS/s ADC-DAC in 0.12mm SiGe," ISSCC Dig. Tech. Papers, pp. 262-263, 2004.

[10] Greshishchev, et al., "A Fully Integrated SiGe Receiver IC for 10-Gb/s Data Rate" J. of Solid-State Circuits, vol. 35, pg. 1949, 1998

[11] Y. Greshishchev and P. Schvan, "A 60-dB Gain, 55-dB Dynamic Range, 10-Gb/s Broad-Band SiGe HBT Limiting Amplifier," J. of Solid-State Circuits, pp. 1914-1920, vol. 34, no. 12, 1999

[12] R. J. van de Plassche and P. Baltus, "An 8-bit 100-MHz Full-Nyquist Analog-to-Digital Converter," J. Solid-State Circuits. pp. 1334-1344, vol. 23, no. 6, 1988.

[13] C. W. Mangelsdorf, "A 400-MHz Input Flash Converter with Error Correction," J. Solid-State Circuits. pp.184-191 vol. 25 no. 1, 1990

[14] K. Ono, et al., "Error Suppressing Encode Logic of FCDL in 6-bit Flash A/D Converter," BCTM Technical Digest, pp. 200-203, 1996.

[15] P. Schvan et al., "A 22GS/s 5b ADC in 0.13µm SiGe BiCMOS," ISSCC Dig. Tech. Papers, pp. 572-573, 2006.

[16] P. Schvan, D. Pollex, T. Bellingrath, "A 22GS/s 6b DAC with Integrated Digital Ramp Generator," ISSCC Dig. Tech. Papers, pp. 122-123, 2005.

ARCHITECTURES AND ISSUES FOR GIGASAMPLE/SECOND ADCs

Ken Poulton, Robert Neff, Brian Setterberg,
Bernd Wuppermann, Tom Kopley
Agilent Labs, Santa Clara, California

Abstract

Architectures for ADCs at 1 Gigasample/second (1 GSa/s) and beyond now include flash, folding and interpolating as well as the time interleaving of slower unit converters such as pipeline and even successive approximation ADCs. In addition, CMOS is taking over in this former bastion of bipolar technology. We describe the issues common to all architectures: bandwidth, power, I/O, data storage, and cost. We examine these issues in detail for the time-interleaved approach as exemplified by two 8-bit ADCs operating at 4 GSa/s and 20 GSa/s, implemented in CMOS.

1. Introduction

Before 2001, analog-to-digital converters (ADCs) at rates over 1 gigasample/second (1 GSa/s) were mainly used in digital oscilloscopes. The architectures used were flash and folding-and-interpolating, some of them time-interleaved up to 4 ways. Sample rates went up to 4 or 5 GSa/s; the technologies were bipolar and power and cost were high.

The last five years have seen a huge change in ADCs used at 1 GSa/s and beyond. On the application side, increasing disk drive data rates has created the initial need for cheap, relatively low-power 6-bit ADCs in the gigasample region. More recently, future high-bandwidth communications applications (both wireless and wireline) are driving interest in ADCs with sample rates near 1 GSa/s with power compatible with battery operation.

In technology, the last five years have seen a wave of mostly-CMOS converters. This is enabled by the increasing speed of CMOS processes, and required for integration of ADCs into larger CMOS chips for cost-reduction reasons. The surprising development is the use of CMOS and pipeline architectures even for non-embedded applications, up to the fastest 8-bit ADCs. This has been enabled by the development of massively time-interleaved architectures.

The issues that are specific to, or more critical in, gigasample ADCs include

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 17–32.

- Wide bandwidth with low distortion and noise. The analog performance of an ADC (e.g., BW, SNDR, SFDR) depends most strongly on the front-end sampling of the input signal. The dynamic range required depends on the application.
- High sample rate with low metastable error rate. Flash and folding ADCs can make their metastable rates arbitrarily low simply by adding more pipelining after their comparators (at the expense of power), but this is more difficult to resolve for multi-stage converters. Applications with a lot of memory or peak detection can drive the need for metastable error rates as low as 10^{-17}.
- Power dissipation. Gigasample ADCs are usually close to (or beyond) the knee in the power curve where power must be increased much more than linearly to increase the sample rate. Even converters with low absolute power may be near this point.
- Interleave overhead. While time interleaving allows one to increase sample rate, it brings along its own issues:
    - Generation of multiple clocks with sub-picosecond time alignment
    - Gain and offset alignment
    - Physical distribution of analog signals and clocks

  These add both power and complexity.
- I/O or sample memory. Driving the samples off-chip, or storing them on-chip can take as much power as the ADC itself.

These issues will be illustrated in the rest of the paper.

2. 20th-Century Technologies for Gigasample ADCs

Most gigasample ADCs before 2001 shared several characteristics:

- Bipolar IC technology. Briefly stated, the idea was that to go fast, you should use the fastest technology. In the late 90's, 25-GHz bipolar technology could hit 1.5 GSa/s [1] and 2 GSa/s [2] and 50-GHz GaAs HBT technology reached 4 GSa/s [3]. 8-bit CMOS ADCs were still in the 100-MSa/s range. 6-bit CMOS ADCs were technology-limited to 500 MSa/s.

  Secondly, bipolar transistor accuracy (e.g., offset in a diff pair) is about 10 times better than in CMOS technology. These two characteristics made bipolar the "obvious" choice for gigasample ADCs.
- Emphasis on low complexity. In most bipolar processes, the number of transistors that can be fabricated with high yield is very low compared to CMOS. In addition, a relatively large current (on the order of 1 mA) is required to achieve the full f_T of the transistor,

which leads to a practical limit of around 3000 full-speed transistors on a 10-W chip.

- Front-end track-and-hold (T/H). It is difficult to maintain low distortion at high input frequencies through the analog preprocessing in a folding ADC. The multiple parallel comparators inherent in both flash and folding architectures provide another avenue for high-frequency errors due to sample-time mismatches. The usual solution was to include a front-end T/H, but bipolar track and hold circuits such as diode bridge samplers or switched emitter followers are very power hungry.
- Parallel PECL outputs. Most designs limited data rates on their parallel output ports to 1-2 GHz by adding demultiplexing circuits and increasing the number of outputs by 2-4x. These are also power-hungry.
- High Power. The above-mentioned bipolar ADCs were all over 5 W well before microprocessors forced the development of many of today's heat-dissipation technologies.
- Custom packaging. While [1] is in a BGA package, earlier ADCs used custom ceramic packages to handle the bandwidth, I/O and power-dissipation requirements. Many used multi-chip assemblies [2] [4].

A package photo of such an ADC [2] is shown is.Figure 1. The core was a 7-bit bipolar folding and interpolating ADC, operating at 2 GSa/s. Two cores were on one chip, time-interleaved to get 4 GSa/s and voltage-interleaved to reach 8 bits. Two chips were time-interleaved to reach 8 GSa/s in an oscilloscope product. The chip included a preamp and a bipolar diode-bridge T/H for each ADC core, which dissipated nearly as much power as the ADC cores. The total power was 13 W.

To reduce the power required to communicate parallel data to the memory, the custom CMOS memory chip was placed next to the ADC and connected with chip-to-chip wirebonding.

3. Trends and Tools

3.1. Trends

Several trends have driven a different approach to gigasample ADCs in the last 5 years. The first is in IC technology. For over two decades, CMOS technology has been following Moore's Law and the ITRS roadmap with amazing success and predictability. Progress in CMOS has been steady because the demand and rewards for digital circuits with smaller features have been so clear and the market has been growing at a huge rate. CMOS

gate lengths have shrunk 33x in the last 25 years: from 3 um in 1980 to 90 nm, and gate delays have decreased by about the same amount.

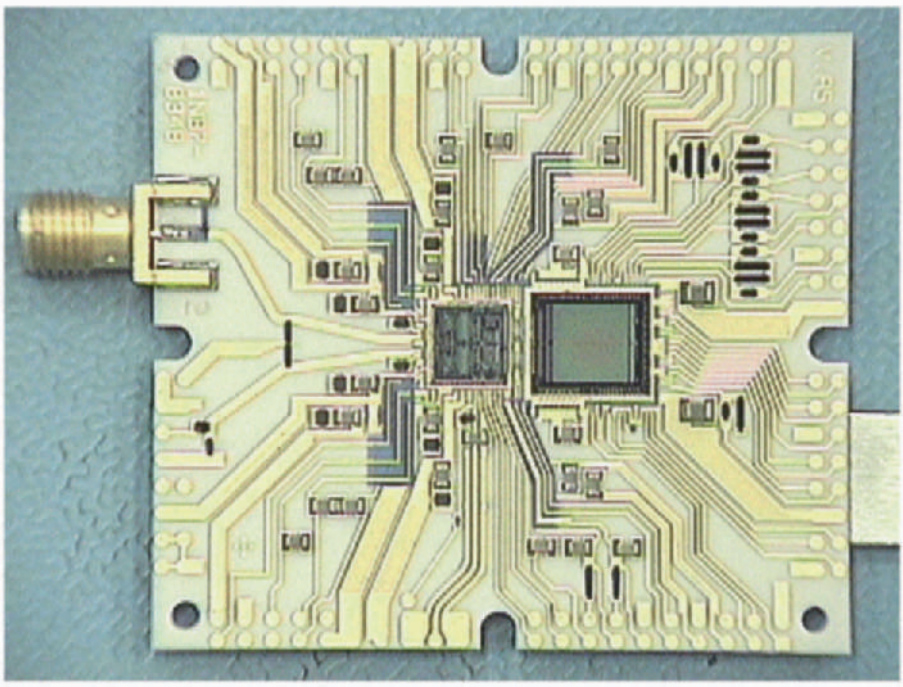


Fig. 1. Custom thick-film hybrid package with custom bipolar ADC and custom CMOS memory chip.

Bipolar technology, by contrast, has moved in bigger steps, but much less frequently, because its shrinking share of the semiconductor market makes it harder to finance process development. Bipolar and BiCMOS technology has also been driven more by niche markets, such as CPUs (a mirage of the 90's) and RF circuits in the last 5 years. Silicon and SiGe bipolar f_Ts have changed by about 40x, from 5 GHz in 1980, to 200 GHz now, but each period of stagnation has caused designers to consider other technologies. One of those periods was in the mid-90s.

A second aspect of technology scaling is the increasing level of integration required for lowest-cost solutions. This leads to the desire to integrate ADCs with other (mostly digital CMOS) circuits to reduce chip count and cost. Since CMOS wafers are much less expensive than BiCMOS, there is a strong motivation to move the ADCs into CMOS.

Even for chips that are standalone ADCs, the much lower cost of CMOS wafers and the dramatically lower power of CMOS logic makes CMOS ADCs attractive if the fundamental design problems around speed and accuracy can be overcome. The accuracy problems of CMOS were overcome very nicely in the late 80s by the use of switched-capacitor circuits and

sigma-delta ADC architectures, but neither of these lent themselves to gigasample speeds. But they at least made it clear that CMOS could be used for some ADCs and other analog circuits.

Another trend is in ADC architectures: adding more calibration and digital corrections to augment the analog circuit performance of known ADC architectures. This both drives the need for more logic on the ADC chip and alleviates the need for the accuracy in the transistors.

3.2. Power Efficiency Figure of Merit

In order to compare different ADC approaches, the ADC power figure of merit (FOM) was developed (and first reported at AACD in 1992 in [5]):

$$FOM = P / (2^{ENOB} \ F_{S\text{-}NYQ})$$

where $F_{S\text{-}NYQ}$ is the minimum of F_S and 2*RBW, the BW where the SNDR has dropped by 0.5 bit from the low-frequency value. Smaller values indicate more power-efficient and accurate ADCs. The values for Nyquist (non-oversampled) ADCs reported at ISSCC through 1998 were 5-1000 pJ/conversion-level; reported values have dropped steadily since then to a best value of 0.16 pJ/conv-level in 2006.

It's important to keep in mind the inconsistencies in the numbers used to compute FOM. The largest is whether digital power, especially I/O, is included. This may not affect low-speed ADCs very much, but can change the FOM by a factor of 2 for gigasample ADCs. Other circuits such as input buffering and T/H circuits may be key to performance but not included in FOM calculations.

But beyond those errors, there is typically a 10:1 range in FOM results reported in a given year, based on the goals of the IC. Some papers may be focused on demonstrating a new architectural twist or calibration method rather than power and accuracy. Many designs are aimed at hitting a specific sample rate and accuracy with the technology available, which may lead to "inefficient" implementations as measured by the FOM.

When designing a new converter, FOM results can be useful in trying to compare different architectures and extrapolating from published accuracy levels to the desired level.

One key pitfall in this process is that for ADCs limited by thermal noise, scaling for higher or lower SNR does not result in a constant FOM. In this case, power scales as the square of resolution 2^{ENOB}, not linearly as implied by the FOM. The reason to keep using this FOM is that when we look over the whole range of ADCs (from 3 bits to 20) this FOM provides a more nearly resolution-independent value than one that scales with a constant FOM. This is consistent with the observation that different architectures work better at different sample rate points.

4. Two 21st-Century Gigasample ADCs

Based on the 20th century ADC technologies, it seemed silly to consider using CMOS to build scope ADCs at 4 GSa/s and beyond. But the trends and cost differences were compelling enough to stimulate us to investigate.

Our first target for a CMOS ADC was for 8 bits at 4 GSa/s, aimed at replacing the above bipolar ADC with a lower-power and lower-cost CMOS ADC. We took the time-interleaved approach to an extreme: we would use many independent ADC "slices" operating in round-robin fashion to reach the desired sample rate. We settled on a number of new design principles:

- Choose the core ADC architecture primarily for power efficiency of the unit converter rather than raw speed
- Time-interleave as many slices as needed
- Let currents and device sizes be optimized for SNR rather than matching
- Use calibration to recover the necessary accuracy

These principles were chosen to take advantage to the strength of CMOS: lots of transistors, low costs for digital logic.

After the success of the 4-GSa/s ADC [6], we extended this approach to 20 GSa/s, creating what is still the fastest 8-bit ADC in any technology [7].

4.1. ADC Slice Architecture

The selection criteria for the unit ADC slice included:

- Power efficient (low FOM)
- Area efficient (low area/sample_rate)
- High sample rate (to minimize the overall complexity)
- Very low metastable error rate (conflicts with high sample rate)
- Scalability to future processes (especially for VDD scaling)

It quickly became apparent that pipeline ADCs were attractive at the 8-bit level. However, the typical switched-capacitor approach was both slow and the necessary capacitors were area-intensive.

Instead, we selected a current-mode pipeline. The simplified schematic is shown in Fig. 2. The basic circuit is a current mirror with a sampling switch between diode and mirror devices. The signal is carried in the current values rather than the voltages. This has several advantages: open-loop operation for high speed; current mirrors are more linear than other MOS open-loop amplifiers; no linear capacitors required; simplicity for small area and low power. Although the previous literature showed current-mode pipelines in the 10 MSa/s range, we determined that a simple design could operate at 125 MSa/s in 0.35-um CMOS.

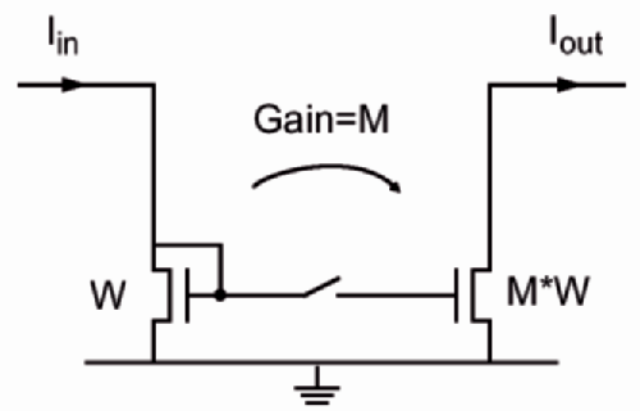


Fig. 2. Simplified Schematic of the Current-Mode Amplifier in the Pipeline.

When designed to meet the SNR needs of an 8-bit ADC, we had a small, low-power stage with very poor gain accuracy, far, far short of being able to achieve the required 8-bit gain accuracy in the first pipeline stage. We introduced redundancy with a reduced-radix architecture, resolving 1 bit per stage. (Note that many pipeline designs introduce redundancy with a 1.5-bit-per-stage design with stage gains of 2.00. This reduces the accuracy required of the comparators, but does not relax the gain accuracy required of the amplifier.) The nominal radix we chose was 1.6 to guarantee high yields with a large number of pipelines on a die. This in turn called for 12 pipeline stages to reach a resolution of 256 levels.

The resulting pipeline requires three additional circuits to make a complete slice: 1) an input T/H, 2) an input transconductor stage to convert the input voltage samples into current and 3) a digital radix converter to convert 12-bits of radix-1.6 data to 8 bits of binary.

During the calibration process, the effective gain of each stage is extracted, which leads to the coefficients to be loaded into the radix converter. In this approach, we extract the fabricated stage gains with (at least) 8-bit accuracy instead of having to control them to that accuracy.

4.2. Input T/H

Time interleaving can get us to an arbitrarily large sample rate, but tends to reduce analog bandwidth and accuracy somewhat. The input T/H is the key to analog performance, limiting the bandwidth, linearity and SNR of the whole system.

Fortunately, CMOS has one truly fast circuit: a series-FET sampling switch (Fig. 3). There are several conditions required to get the highest bandwidth from this sampler: there can be only one NMOS device in the signal path; the signal swing is small; the input common mode is near

ground; the clock has the fastest-possible falling edge; the circuit is differential; and C_{HOLD} is composed only of circuit parasitics.

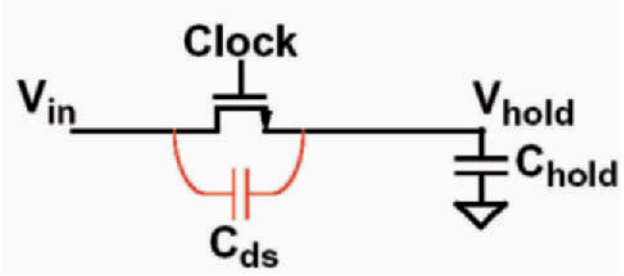


Fig. 3. Simplified Schematic of the Input T/H.

In 0.35 um technology, we got a 2 GHz bandwidth with –50 dBc HD3; in 0.18 um, we got a 7 GHz bandwidth with –50 dBc HD3.

4.3. T/H Architecture

There are several basic approaches to front-end T/Hs: single-T/H, tree structures, and time-interleaved T/Hs (one per slice, each directly connected to the input). The single-T/H and tree approaches have the key advantage that sample timing is set by a single T/H clock. However, they tend to be limited by the loading after the first T/H to small interleave factors.

Multiple interleaved T/Hs makes the T/H-to-ADC-slice connection very simple and relatively easy to lay out, but creates very demanding input distribution and timing alignment problems. The alignment accuracy required can be calculated by considering the apparent voltage error created by a small timing error in sampling the fastest-slewing input signal. As a rule of thumb, for a 1 GHz input signal, 1 ps rms of jitter and clock misalignment will limit performance to 7 effective bits.

For the 4-way interleaving in [4] we chose a tree structure and we implemented the T/H tree in a faster technology (GaAs MESFET) than the ADC. For the high interleave factors we chose for the CMOS ADCs, interleaved samplers was the only feasible approach. In the 4-GSa/s ADC, we were able to simply drive all the interleaved samplers directly from the 100-ohm differential input to the chip. For the 20-GSa/s ADC, driving 4 pF at 5 GHz was not feasible from 100 ohms, so we added a SiGe input buffer chip to drive the T/H capacitance with a low impedance.

For the 4-GSa/s ADC, we used 32 slices, for the 20-GSa/s ADC, 80 slices. For the 20-GSa/s ADC, our goal was 5 effective bits at 5 GHz input signal which required better than 600 fs rms timing error including both clock jitter and residual misalignment. The uncalibrated alignment we

achieved with careful design of parallel paths had around 10 ps rms misalignment, mostly due to random device fabrication mismatches in the clock paths. The calibration step adjusts digitally-controlled delay elements in each clock path.

4.4. T/H Clock Architecture

The chosen interleaved T/H architecture for the 4-GSa/s ADC requires the generation of 32 clocks at 125 MHz, each delayed by 250 ps from the previous. We targeted 1 ps rms for jitter and 1 ps rms for misalignment.

The simple way to generate these clocks would be a delay-locked-loop (DLL) with 16 stages of 250-ps delays (using both edges of the 16 stages to the 32 clocks). But the power required to reach 1 ps rms jitter with 16 stages would have been much larger than the total chip power budget. This is because power scales inversely with jitter squared as well as the accumulation of jitter down the delay chain.

The clocking architecture we use is shown in Figure 4. To generate the the 250-ps delays between adjacent clocks, we use a 4-stage DLL at 500 MHz, which reduces the power requirement by 16x. To get the 32 clocks at 125 MHz, we divide each 500 MHz phase by 4.

To get the required timing alignment, we introduce digitally-controlled delays in two places: 1) after the DLL and 2) after the divide-by-4 stages. The combination gives us a total delay-adjust range of around 100 ps and resolution of 250 fs.

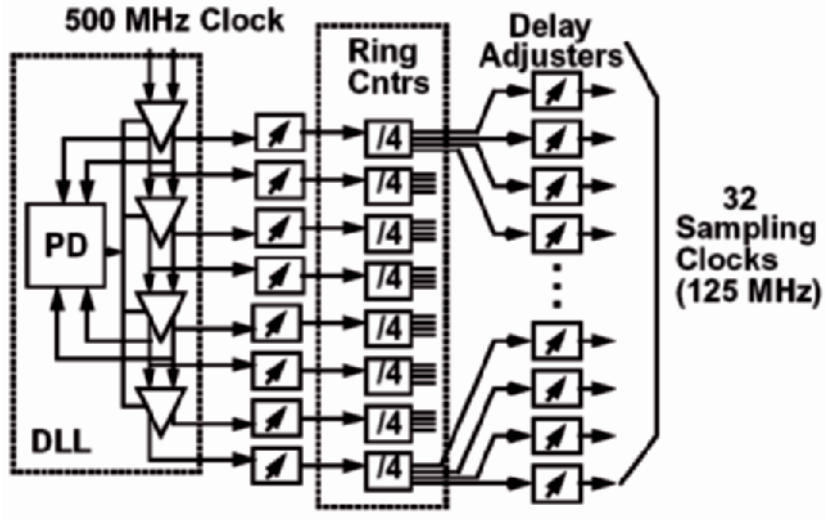


Fig. 4. Clock Generator.

4.5. Calibration

The overall block diagram of the 4-GSa/s ADC is shown in Fig. 5 with the calibrations highlighted.

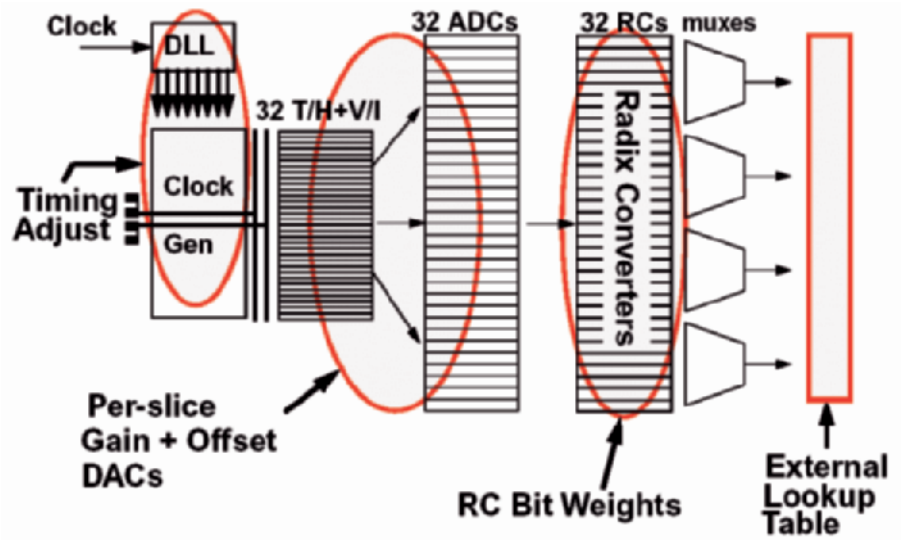


Fig. 5. Block Diagram Showing Calibrations.

All calibrations are done with special calibration signals switched into the input of the ADC. The calibrations are all digitally controlled, but we use a mix of analog and digital circuits to insert them into the signal path. The left-most calibration is the timing alignment, which uses digital values to adjust the timing of the clocks. The next one is the slice-to-slice alignment of gain and offset; this uses digital values to control per-slice gain and offset DACs. The third is the bit-weight values loaded into the Radix Converter. These may be thought of as the stage gains in each stage of each pipeline converter. The last is an external lookup table, which is used in some applications. This primarily corrects the minor third harmonic distortion created in the V/I stages in front of each current-mode ADC.

All of the calibration values must be calculated by software external to the ADC; we use the general-purpose CPU already present in an oscilloscope. However, once calibrated, the chip performs all the corrections except the optional lookup table in hardware in real time.

It has already been described how the use of calibration allows the use of less-accurate, and thus, smaller and lower-power circuits. In addition, the use of calibration relaxes the need to characterize and compensate for second- and third-order effects on gain accuracy, reducing design time. The use of iterative calibration allows the use of simple correction DACs with relatively poor gain control and linearity.

4.6. 20-GSa/s ADC

At 4 GSa/s, we sent the data out in LVDS in 4 parallel 8-bit words at 1 GByte/s for each word. For the next step to 20 GSa/s, we did not want to increase this to 20 words (320 pins), nor did we want to increase the word rate due to the difficulty of maintaining clock and data alignment. So we decided to put 1 MByte of sample memory on the ADC chip. The resulting chip is seen in Figure 6. It is evident that the memory dominates the chip. The SiGe chip to buffer the input capacitance is seen at the top; chip-to-chip wirebonding is used for the analog interface between the two chips.

4.7. Performance

The primary measure of performance for oscilloscope ADCs is signal to noise plus distortion ratio (SNDR), measured in dB or effective bits. In Fig. 7 we plot the performance of the previous bipolar 4-GSa/s ADC along with the 4- and 20-GSa/s CMOS ADCs. Despite operating at 1/3 the power, the CMOS 4-GSa/s ADC is more accurate than the bipolar one by about 0.5 effective bits.

The gain of the 20-GSa/s ADC is seen in the lower graph; it is flat out to 5 GHz. The accuracy decreases inversely to input frequency at the high end; this is due mainly to the jitter in this ADC's sampling clocks, totaling 600 fs rms.

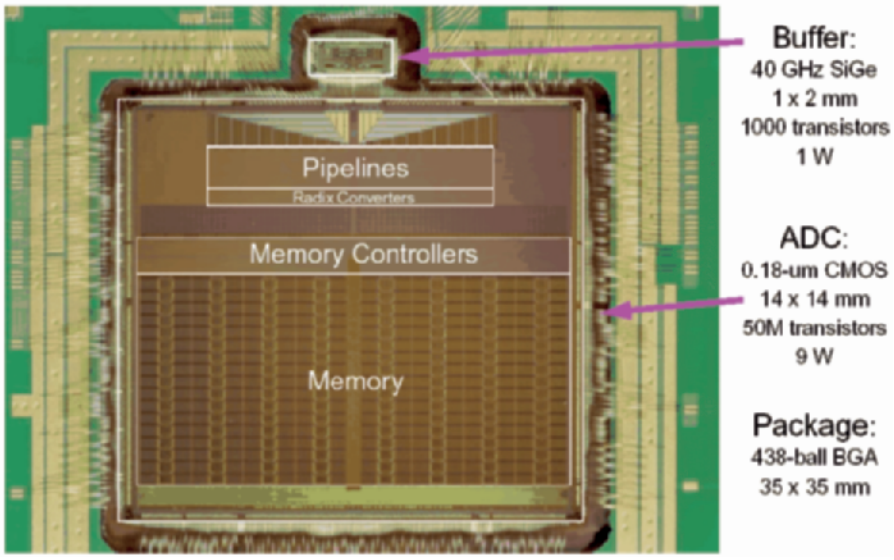


Fig. 6. 20-GSa/s ADC Multi-chip Module Photo.

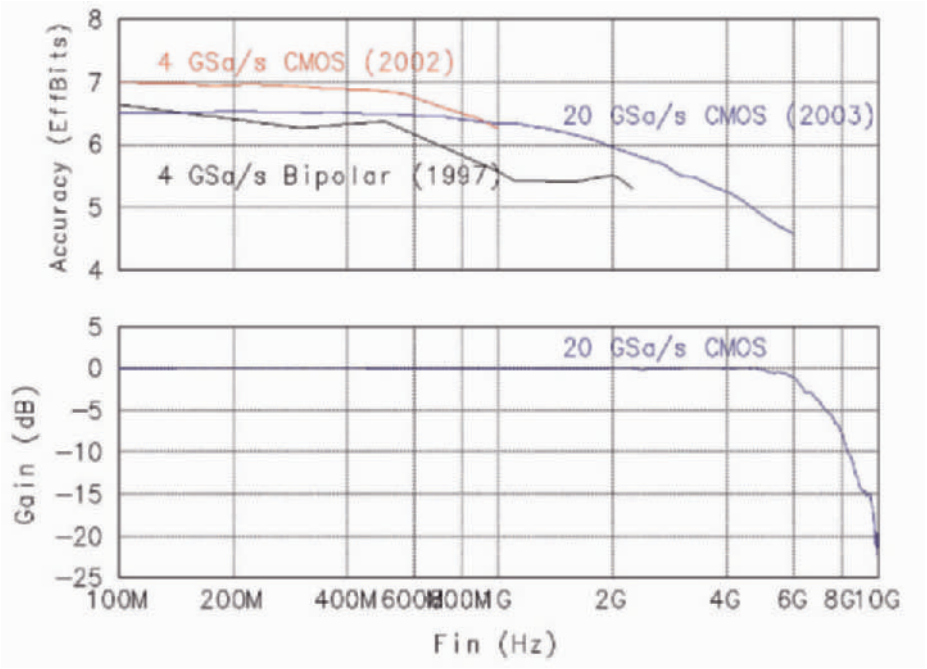


Fig. 7. ADC Effective Bits and Gain vs. Input Frequncy.

4.8. Power

The power breakdown of the two chips is shown in Table 1. The core ADC slices were chosen for power efficiency and provided good FOM values for their process generation.

However, the overall power FOM of the chips is around 10 times worse. This is due to several factors. On the analog side, the interleaved clock generator increases the power by about 150 mW. The power required for the full-speed outputs or on-chip memory nearly doubles the chip power. The choice of sample rate at 3-4x the input bandwidth provides high-fidelity waveforms, but further reduces the FOM compared to the simple Nyquist requirement of 2x the bandwidth. Finally, extending the bandwidth into the jitter-limited region (as seen in Fig. 7) increases the sample rate needed to 5-10 times the resolution bandwidth. These all contribute to better performance for oscilloscope products, but combine to make a very poor-looking whole-chip FOM.

Table 1. Power Breakdown and Figure of Merit of the CMOS ADCs.

	4 GSa/s CMOS	**20 GSa/s CMOS**
Input Buffer Chip	-	1.0 W
Pipeline ADC core	1.3 W	3.4
T/H + V/I	0.6	0.6
Radix Converters	0.45	0.6
DLL	0.15	0.13
Outputs	1.7	-
Memory	-	4.2
Total	**4.2 W**	**10.0 W**
Slice FOM	**4.5 pJ/conv-level**	**2.3 pJ/conv-level**
Chip FOM	**41 pJ/conv-level**	**27 pJ/conv-level**

5. Other 21st-Century ADCs

5.1. Architectures

The 2006 International Solid State Circuits Conference included many papers demonstrating the arrival of CMOS in the gigasample range. A 4-bit flash converter at 1.25 GSa/s [8] achieved a record FOM value because its low resolution can tolerate the bad threshold matching of CMOS comparators with no calibration.

The conference also showed several examples of new architectures arriving in the gigasample arena. One paper [9] showed a 6-bit subranging ADC, probably the first of that architecture to operate directly at 1 GSa/s. Another [10] described a 1-GSa/s ADC using 4-way time interleaving of pipeline ADCs to reach 11 bits. A third paper of note [11] described the interleaving of two 6-bit successive approximation converters to reach 600 MSa/s. Although not yet a gigasample ADC, it seems likely that with 4-way interleaving, we could see one of the slowest of all Nyquist ADC architectures used to go faster than 1 GSa/s.

And of course, all of these ADCs were implemented in 130-nm or 90-nm CMOS. The only BiCMOS ADC [12] was a 5-bit flash at 22 GSa/s.

5.2. Power Figure of Merit

A plot of power figure of merit vs. ADC input BW is shown in Fig. 8 for papers at ISSCC from 1998 through 2005 [13]. We observe a region in the lower right with no data points. This is not due to a fundamental limit, but rather many practical considerations that make it difficult to achieve a low FOM along with very high bandwidth: input buffering, clock generation,

signal distribution, jitter and I/O power. These are the same factors that cause our massively-interleaved ADCs to have reasonable FOMs for single ADC slices, but high FOM values for the whole chip.

The fact that this is not a fundamental limit is shown by the addition of points for ISSCC 2006 in Fig. 9. Many of these points are pushing into this formerly-empty region.

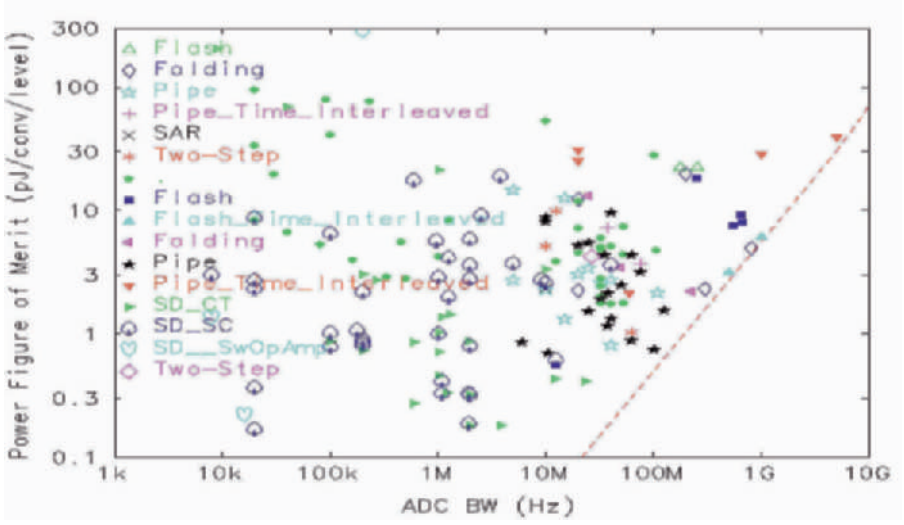


Fig. 8. FOM vs. ADC Input BW for ISSCC papers 1998 though 2005.

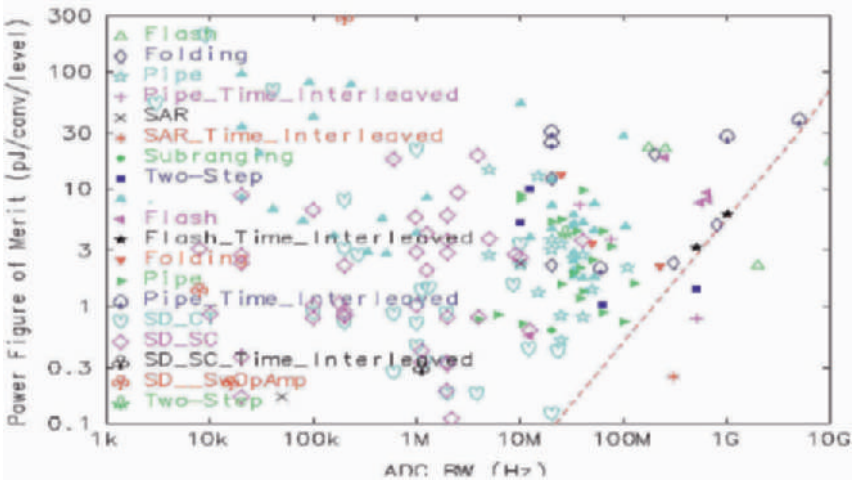


Fig. 9. Power Figure of Merit including 2006.

6. Summary

The turn of the millennium saw several major changes in gigasample ADCs. Most notable among these are: an increasing number of applications for such ADCs, an increased use of time interleaving enabling the use of slower ADC architectures as building blocks, and the march of CMOS into this formerly bipolar domain.

References

[1] MAX108 Datasheet, Maxim Integrated Products.

[2] Ken Poulton, Knud L. Knudsen, John Kerley, James Kang, Jon Tani, Eldon Cornish and Michael VanGrouw, "An 8-GSa/s 8-bit ADC System", IEEE VLSI Symposium on Circuits, 1997.

[3] Ken Poulton, Knud L. Knudsen, John J. Corcoran, Keh-Chung Wang, Randy Nubling, Richard Pierson, Mau-Chung F. Chang, Peter Asbeck and R.T. Huang, "A 6-bit, 4 GSa/s GaAs HBT ADC", IEEE Journal of Solid State Circuits, Oct 1995.

[4] Ken Poulton, John J. Corcoran, Thomas Hornak, "A 1 GHz 6-bit ADC System", IEEE Journal of Solid State Circuits, Dec 1987.

[5] J. Corcoran, "High Speed Sample and Hold and Analog-to-Digital Converter Circuits", AACD 1992, published in Johan H. Huijsing, Rudy J. van de Plassche, Willy M.C. Sansen "Analog Circuit Design", Kluwer Academic Publishers, December 1992, ISBN 0-7923-9288-4.

[6] Ken Poulton, Robert Neff, Art Muto, Wei Liu, Andy Burstein, Mehrdad Heshami "A 4-GSample/s 8b ADC in 0.35-um CMOS", IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 166-167, Feb 2002.

[7] Ken Poulton, Robert Neff, Brian Setterberg, Bernd Wuppermann, Tom Kopley, Robert Jewett, Jorge Pernillo, Charles Tan, Allen Montijo, "A 20-GSample/s 8b ADC with a 1-MByte Memory in 0.18-um CMOS", IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 318-319, Feb 2003.

[8] Van der Plas, Geert; Decoutere, Stefaan; Donnay, Stéphane, "A 0.16pJ/conversion-step 2.5mW 1.25GS/s 4b ADC in a 90nm Digital CMOS Process", IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 566-567, Feb 2006.

[9] Figueiredo, Pedro; Cardoso, Paulo; Lopes, Ana; Fachada, Carlos; Hamanishi, Naoyuki; Tanabe, Ken; Vital, João, "A 90nm CMOS 1.2V 6b 1GS/s Two-Step Subranging ADC", IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 568-569, Feb 2006.

[10] Gupta, Sandeep; Choi, Michael; Inerfield, Michael; Wang, Jingbo, "A 1GS/s 11b Time-Interleaved ADC in 0.13µm CMOS", IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 576-577, Feb 2006.
[11] Chen, Shuo-Wei; Brodersen, Robert, "A 6b 600MS/s 5.3mW Asynchronous ADC in 0.13um CMOS", IEEE International Solid State Circuits Conference Digest of Technical Papers, pp. 574-575, Feb 2006.
[12] P. Schvan, D. Pollex, S-C. Wang, C. Falt, N. Ben-Hamida, "A 22GS/s 5b ADC in 130nm SiGe BiCMOS", IEEE International Solid State Circuits Conference Digest of Technical Papers, Feb 2006.
[13] Data courtesy of Boris Murmann.

CONCEPTS AND IMPROVEMENTS IN PIPELINE AND SAR ADCS

Dieter Draxelmayr, Peter Bogner

Abstract

In this paper we want to review the development of pipelined and SAR ADCs from basic concepts to novel techniques. We will demonstrate that there are quite a lot of similarities and that there are a few major discrepancies. Based on this we draw comparisons and point out specific strengths and weaknesses of the individual architectures.

1. Introduction

Pipeline A/D-converters have received considerable attention over the last couple of years. At every ISSCC we were able to find several papers on this topic demonstrating a great deal of research interest and - of course - demonstrating step by step improvements of the underlying basic concepts. The main reason for this most probably is the capability of combining high performance in speed and resolution with low power consumption. Today already several authors have demonstrated a Figure of Merit (FoM) of less than 1pJ/conv. step. There exist several different FoM definitions. The FoM used for this is given by

$$FOM = \frac{P}{2^{enob} * 2 * BW} \tag{1}$$

(enob $\hat{=}$ effective number of bits,
BW $\hat{=}$ effective resolution bandwidth,
P $\hat{=}$ dissipated power)

One reason for the good performance is the use of binary search. Binary search is also known in software as a very efficient method to find a certain item. In electronics we use it to find a specific code corresponding to some input voltage. There are also other concepts which exploit binary search for A/D-conversion: Successive Approximation (mostly referred to as SAR, where the "R" stands for "Register"), and cyclic A/D-converters. SAR

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 33–52.

converters may also demonstrate very good performance, whereas cyclic converters fall a little bit behind, although the underlying principle seems to be very convincing and quite straightforward. Therefore in the remainder of this text we will concentrate on Pipeline and SAR, demonstrating the similarities and pointing out the differences.

2. Fundamentals: The Power of Binary Search

Figure 1 shows a basic diagram of a SAR converter. The input signal (if not static) is sampled and held by a Sample-and-Hold (SH) circuit. A Digital-to-Analog converter (DAC) produces an intermediate voltage that is compared to the sampled (and held) voltage. If these two voltages are equal, obviously the digital code at the DAC input corresponds to the input voltage. We can try several algorithms to find the appropriate input code and the binary SA-algorithm is a very popular one. Fig. 2 shows a widely used implementation. It is known as charge-redistribution ADC [1]. During the first phase the switch Sc is closed and the switches SMSB to SLSB are set to "input voltage". Next Sc is opened, which defines the sampling point of the input signal. Then the trapped charge is redistributed by setting the individual switches SMSB to SLSB either to the positive or to the negative reference voltage.

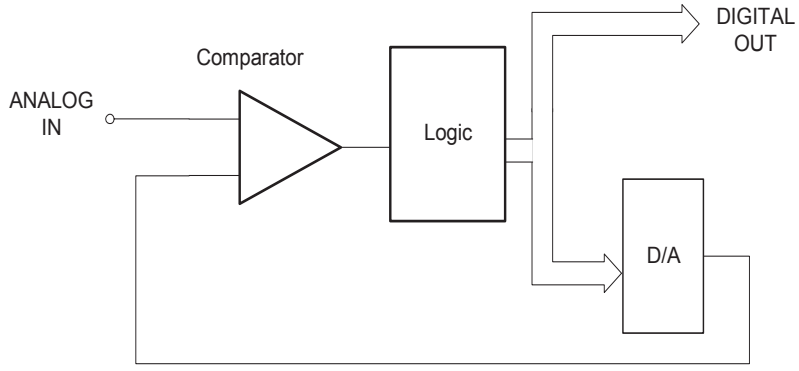
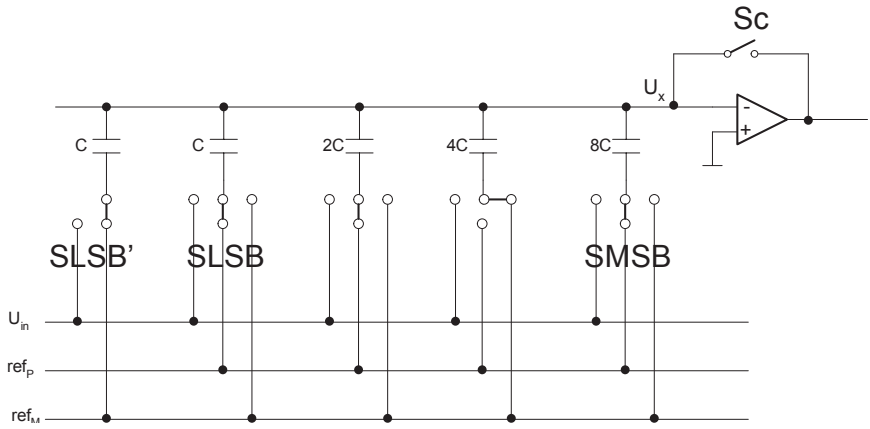


Fig. 1. Tracking Converter, SA converter.



Calculation Example: $U_{in} = ref_M + \frac{11}{16} * \left(ref_P - ref_M\right)$ (2)

$$\left(ref_P - U_X\right) * \left(C + 2C + 8C\right) + \left(ref_M - U_X\right) * \left(C + 4C\right) = Q = 16C * \left(U_{in} - U_{off}\right)$$

$$11C * ref_P - 11C * U_X + 5C * ref_M - 5 * CU_X = 16C * ref_M + 11C * ref_P - 11C * ref_M - 16C * U_{of}$$

$$U_X = U_{off}$$

Fig. 2. Charge-redistribution ADC: Redistribution.

One can show that we have found an equilibrium state (where the voltage at Vx is the same as at the time of sampling) if

$$Vin = \frac{(Vrefp - Vrefm) * \sum \left(di * Ci\right)}{Ctot} \quad (3)$$

Where Vin is the input voltage at the sampling time
Vrefp –Vrefm is the difference between the positive and the negative reference voltages
Di is the individual bit in the corresponding output code
Ci is the individual capacitor associated with the capacitive array
Ctot is the sum of all capacitors that have sampled the input voltage at the sampling time (the point in time where the switch Sc is opened).

This also explains why we have an extra C (operated by SLSB') in the array, because otherwise we would have got a gain error of 1 LSB. The calculations in the figure give an example for an input voltage that refers to code 11. It can be seen that the gain error in such a structure can be made quite small. Although the static offset is cancelled there still remains a dynamic part coming from charge injection and parasitic couplings.

Therefore the dominant error terms are given by offset and non-linearity (caused by element mismatch).

Fundamentally, there are three speed-limiting blocks in this circuit: the comparator, the logic (SAR) and the capacitive network.

The speed of the logic can be improved by optimizing the SAR towards minimum delay.

The speed of the capacitive network is given by the RC time constants. Fundamentally the capacitor size is limited by kT/C-noise, for practical purposes in many cases matching is the limiting factor. However, we can of course increase the size of the switches. Pushing this too hard tends to increase the power consumption. For the logic this is obvious (bigger load gives bigger dynamic power dissipation), but this is even more significant for the analog power consumption, because in order to ensure fast settling of the capacitive network we also have to dimension the reference drivers accordingly.

Finally we also have to work on the speed of the comparator. There are two well-known techniques to increase that: Reset switches (in order to insure a fast settling to the equilibrium state after a decision has been taken) and clamping. Fig. 3 shows an example of the latter.

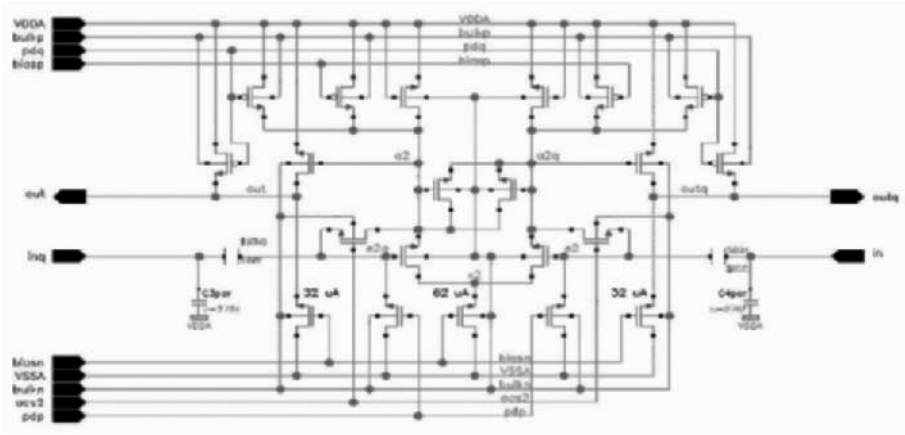


Fig. 3. Design Example: Comparator.

This is a fully differential structure where the input signal is capacitively coupled to a differential stage. In a basic implementation these capacitors are laid out in a capacitive array as shown in Fig. 2. We can see a current-starved diode as a load which at the same time prohibits the signals a2 and a2q from going too low. In addition these two signals are clamped against each other by the two anti-parallel MOS-diodes in the center of the drawing. The circuit is completed by source-follower output stages, sampling switches (operated by ocs2) and power down transistors.

To some extent we can consider a SAR converter as a basic version of a sub-ranging converter because we successively narrow down our search interval. This can be seen as a disadvantage because we have to keep high accuracy until the very last moment. Therefore it is reasonable to try to expand (amplify) the residual search interval to full scale again after each decision step. This is also known as algorithmic A/D-conversion. It operates according to the following steps:

1. Decide, whether your search value is in the upper or lower half of your search interval. This gives the setting of the corresponding bit (MSB for the first decision).
2. If upper half: subtract half of the interval size.
3. Amplify the remainder by 2.
4. Store the result as new input and restart with step 1.

Using this algorithm we can determine an arbitrary number of bits, just limited by the number of cycles we are going to run. A direct implementation of this can be found in a so-called cyclic ADC.

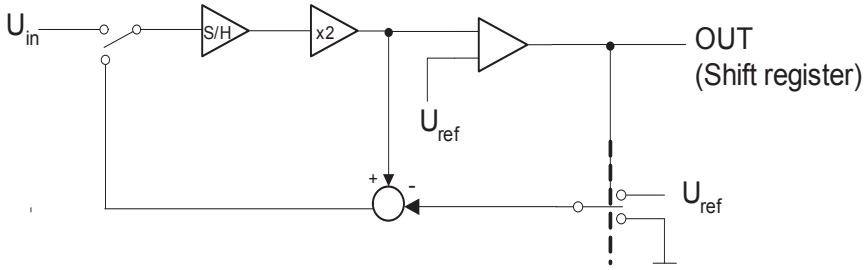


Fig. 4. Cyclic Converter.

At first sight this looks very attractive, because we can get an arbitrary number of bits using quite simple hardware: an OP-amp, a comparator (which could even make use of the existing OP-amp), a few capacitors and switches are sufficient. However, there are also stringent drawbacks: The capacitors again have to be sized at least according to kT/C-noise. This means that we must use quite large capacitors although for any decision following the MSB decision this would not be required. From a fundamental viewpoint this makes the design more power-hungry than it needs to be.

In order to solve this we should expand the algorithm into cycles that use dedicated hardware for each cycle. This can be dimensioned in a way that allows scaling in successive stages. At the same time all of the stages can operate in parallel thus producing a completely converted word on every clock cycle. This principle is known as Pipeline ADC. 5 shows a basic implementation:

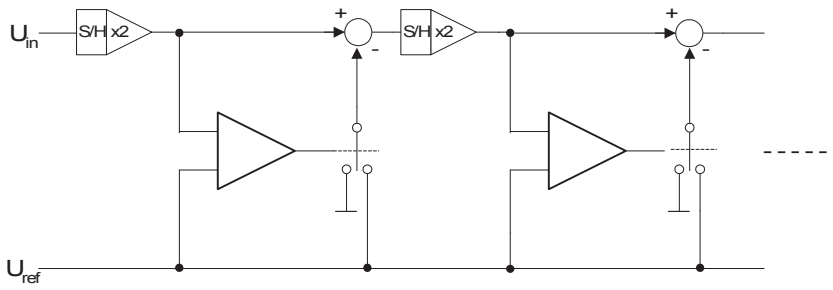


Fig. 5. Pipelined Converter.

Each stage (realizing a "cycle") is composed of a sampler, a comparator, a conditional subtractor and an amplifier with a gain of 2 (or 2^n in case of a multibit design).

The main difference to the SAR ADC is the presence of an operational amplifier (Op-amp) used in an accurate amplifier circuit. This amplifier is usually built in a feedback mode to achieve high accuracy. A feedback circuit is always slower and more problematic than an open loop structure (stability, settling behavior, …). Nevertheless, due to the pipelining all stages work in parallel and therefore we get a high conversion throughput. This speed advantage is the main reason for choosing a pipelined architecture for certain applications but it has the drawback of requiring an accurate amplification stage.

As already mentioned a pipelined ADC also contains a DAC as a building block. In the minimum implementation of 1 bit per stage the DAC is simply one capacitor switched to the negative or positive reference. For the pipelined ADC we also have the option to resolve more than one bit per stage. Fig. 6 shows one stage of this generalized pipelined ADC including an input sample and hold, a flash ADC (comparator), the DAC, the subtractor and the amplifier. In a switched capacitor implementation the circuit can be simplified by combining the SH and the DAC together with the subtractor and the amplifier to generate the so called Multiplying DAC (MDAC) shown in Fig. 7a and 7b. Figure 7a shows the MDAC in sampling mode, so the circuit acts like a standard sampling block. 7b shows the MDAC in multiplying mode whereby the capacitors are switched to the positive or negative reference depending on the decision of the flash ADC. It is interesting to see that this implementation is very similar to the charge-redistribution SAR. In both cases we have a binary network which acts at the same time as a sample and hold capacitor. After charge redistribution we get the new residue; in the case of the SAR we just determine the sign, in the case of a pipeline we amplify. The real implementation of the pipelined ADC is now reduced to two building blocks: the MDAC and the flash ADC (Fig. 8).

The heart of the pipeline ADC is the MDAC, which has the most significant influence on the performance of the ADC.

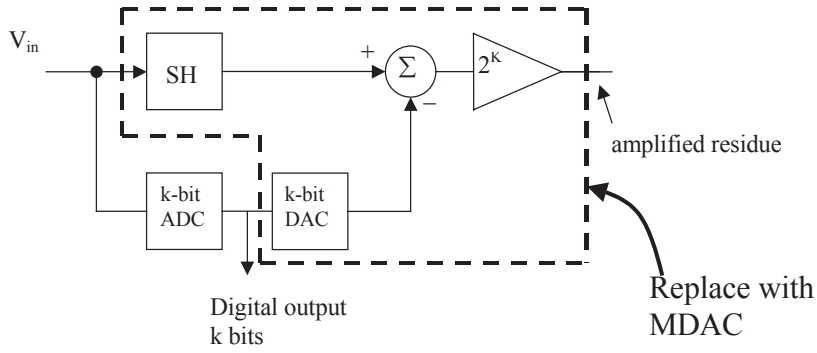


Fig. 6. Block diagram of a pipelined stage.

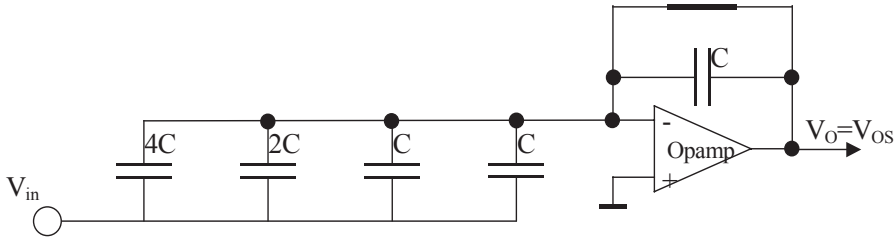


Fig. 7a. Switched capacitor implementation of the MDAC in sampling mode.

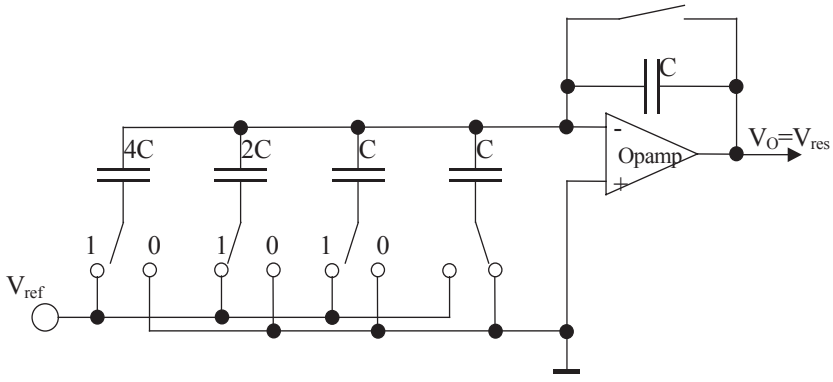


Fig. 7b. Switched capacitor implementation of the MDAC in amplifying mode.

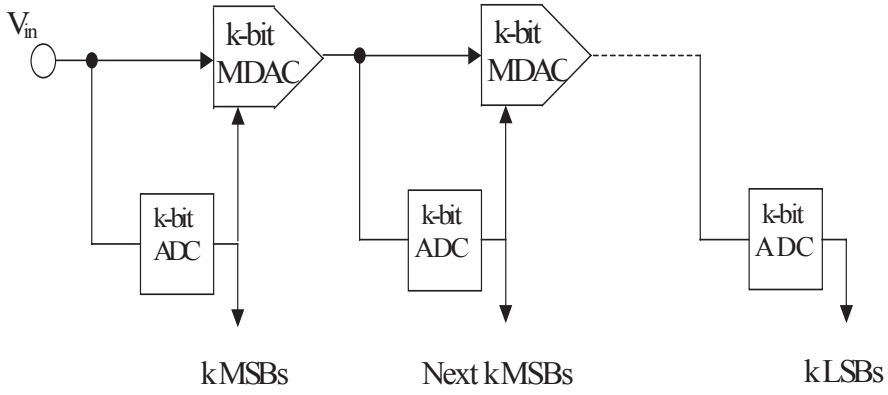


Fig. 8. Real implementation of a pipelined ADC.

3. Increasing the Speed

Although transistor level optimization is crucial in order to achieve high performance, it is foremost the choice of architecture which allows another step further. In case that the signal statistics are known there have been several proposals to speed up the process. If, for example, the input signal is very slow we can apply search algorithms which start in the vicinity of the previous sample. The simplest algorithm just looks at the difference between the input signal and the DAC signal and either increments or decrements a counter to adjust the DAC value by one LSB. This is known as “delta” or “tracking” conversion. Of course this only works well with slow input signals. There are also modifications of this like adaptive step-size control which try to overcome the most stringent restriction: the slow slew rate.

A different and more general approach is to examine the probability distribution of the signal. If you know, for example, that generally your input signal is small, then you can start with smaller search intervals at small codes. Of course the gain in conversion speed at small signals goes at the expense of an even bigger loss in speed if the assumption proves to be wrong. Nevertheless, statistically speaking this will speed up the total conversion process. A more detailed discussion can be found, e.g., in [2].

A more general approach to speed up the conversion process takes into account the properties of real circuitry: One can observe that at the speed limit the conversion of certain codes leads to greater error than the conversion of others. The reason for this is insufficient settling. During a binary search the DAC signal is stepped and a large step leading to a DAC signal in close vicinity to the input signal takes more time to do an accurate decision than a smaller step or a DAC signal that is quite far off from the input signal. There are two approaches to combat this: One can try to use

self-timed circuitry (an example can be found in [3]) or one can allow for some errors and that are corrected later in the conversion process. A quite straightforward implementation of the latter idea is to run several regular conversions and then to insert one or more "correction cycles" in order to eventually correct inaccurate decisions. Of course this only makes sense if the error was small, i.e. when only a few correction cycles are sufficient to derive a correct result. However, this is not a major restriction, as the primary source of error occurs at signals that have not finished settling.

Another method for correction of such errors is described in [4]. The paper presents the usage of a non-binary number base for coding. If the number base is smaller than 2 this leads to redundancy or code overlaps, since the sum of remaining smaller bits is always larger than the respective bit under test. This requires more conversion steps than in the binary case. However, it can run significantly faster yielding an overall gain in speed. For a hardware implementation 3 modifications of the basic SAR ADC are required: A non-binary network, a modified SAR and a code translator, since the output of the SAR will be non-binary. Fortunately these modifications can be carried out with only moderate effort. A non-binary network is relatively simple to implement. The major drawback is that it results in reduced matching. This is not a major issue if trimming is carried out anyway and there is also a self-calibration procedure available [5] (discussed in next chapter). It is also found that it is quite simple to modify the SAR if an appropriate number base is chosen. In a binary SAR a single "1" as "bit under test" is shifted through the SAR. If, for example, the coding "1.85" is taken, as in [4, 5], a value of "10001" must be shifted through, since this gives symmetrical code overlap in both directions (for example, in this case 10001000… is the midpoint between 10000000… and 01111111…). Therefore, the only modification required is to shift two "1s" instead of one. Finally, this non-binary code must be translated into a binary one. A straightforward approach for this translation could be via a simple ROM (or PROM, in the case of trimming). However, for longer code-words it saves effort if the binary representation of all the "bits" in the non-binary code-word resides in a memory. The conversion can be carried out by simply summing all the binary representations of the "bits" that have been set to 1 during the conversion process. Figure 9 shows a possible implementation of such a decoder. As can be seen the additions can commence during the conversion process, which requires a relatively simple arithmetic unit.

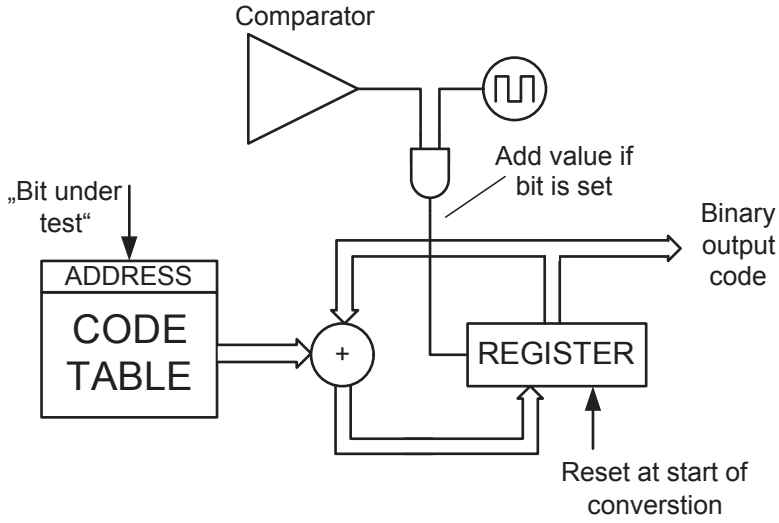


Fig. 9. Code translation logic.

The principle presented in [6] avoids some of the difficulties mentioned in the previous section. The paper describes the use of binary coding for a non-binary search. The only modification in this case is the usage of a somewhat more elaborate search machine, which still requires very little silicon area, especially in deep submicron processes. The general concept is that the next search interval is not defined by bisection, but by an increase in the search interval. This again allows for some settling errors and thus provides code-correction capabilities at the expense of more conversion cycles. Nonetheless, as demonstrated in [6] this can result in a significant increase in overall conversion speed.

In pipelined ADCs we similar principles are used. Redundancy is used in the comparators, which means that there are always more comparators in the flash ADC than necessary. This allows for errors in the flash ADC and relaxes heavily the design effort at the expense of increased circuit complexity. It also makes the design faster. As already mentioned, the simplest implementation of a pipeline stage is a single capacitor switched to the negative respective to the positive reference. For this decision one comparator is required. But this comparator must be very precise as a minor error could cause an overdrive in the amplifying stage. This can be avoided by using an additional comparator which leads to three comparison regions: switch the capacitor to the negative or the positive reference or to ground. This is the well known 1.5bit architecture, currently the most frequently used architecture [7].

The main speed limiting factor is the MDAC. In addition to the problems of the SA-converter which also exist in the pipelined ADC (settling, matching, …) an accurate settling of the amplifier is necessary. We find static and dynamic limitations. The main effect of the dynamic settling error

is a limitation of the sampling speed. This parameter can be influenced by the OP-amp current and the technology used.

The feedback circuitry in the MDAC mainly determines the conversion speed. To increase the speed the loop can be opened and an open loop amplifier can be used instead. This raises the problem of a poorly defined inter-stage gain. Such an architecture is used in [8] and is a very promising technique for nanometer technologies where very low open loop gains are observed. In this work the amplifier topology is a relatively simple differential amplifier. The open loop structure requires additional calibration techniques to overcome the inter-stage gain error.

4. Saving Power

Generally any technique that helps to make a design faster is also a potential candidate for saving power. For example, for a given speed a non-binary SA-converter places less stringent requirements on the comparator. This means, that the comparator can be run at lower current levels to achieve the same net speed. In pipelined converters there are even more options available.

The most power hungry part there is the MDAC. Here the most efficient and simplest way to save power is via stage scaling. That is that later stages are designed with smaller capacitors and less OP-amp current due to relaxed accuracy requirements.

Also the dedicated input sample and hold can be eliminated. This function can be taken over by the sample and hold present in the first stage. This saves much power at the cost of introducing several design problems. Firstly the sample and hold in the first stage causes inter-symbol interference because of the charging of the input sample capacitors to the reference voltage during the amplification phase of the first stage. This charging depends on the input signal (quantized input signal) and is visible at the input during the sampling phase. This can be solved by additional circuitry [12]. Additionally this save of the dedicated input sample and hold can cause mismatch between the input flash ADC and the first stage, which can lead to incorrect flash ADC decisions and to an overdrive of the first stage.

As discussed above, the 1.5bit architecture entails minimal hardware effort and is very robust and is therefore very often chosen. This architecture has an inter-stage gain of 2 (effective resolution of 1 bit per stage) and uses 2 comparators in the flash ADC (1.5bit). Many current implementations are made with a resolution of more than one bit per stage. These so called multi-bit architectures have the advantage of allowing the pipeline to be shortened, which means that fewer stages are necessary for a dedicated resolution of the ADC. Fewer stages need fewer OP-amps and at the same time allow a greater relaxation of accuracy requirements in the subsequent stages. The OP-amp is the most power hungry part of a pipelined ADC and therefore

power consumption can be reduced by using a multi-bit architecture at the expense of increased design complexity. The 14bit ADC in [13] shows the advantage of power saving due to a multi-bit input stage. This 14bit ADC is optimized by using an inter-stage gain of 8 (effective resolution of 3bit) in the first stage. For simple implementation the subsequent stages are implemented by a standard 1.5bit pipeline chain. The second stage is scaled down by a factor of 4, which gives a huge net power saving. This ADC reaches a very good FoM of 1.08pJ/conversion cycle at a supply voltage of 3V.

In a pipelined ADC the OP-amp is idle during the sampling phase (if correlated double sampling is not implemented). Theoretically, it could be switched off during this phase. However, a full power down of the OP-amp is not usually feasible as the power up phase would require too much time. However, it has been shown that the output stage of a 2-stage OP-amp can be switched off [14].

Another property of the pipeline ADC is the alternate processing of subsequent stages. This means that one stage carries out amplification while the other samples and vice versa. This can be taken advantage of to exchange one OP-amp between two subsequent stages. This design technique is also called OP-amp sharing and is a very effective method to save power [15]. The power saving can be a significant portion of the whole design. The drawback is a big increase in circuit complexity. Many switches must be added and the layout requires long interconnect lines. The added switches introduce a series resistance, which degrades the settling behavior of the pipelined stage. Additionally the OP-amp is not reset and the residual signal may hamper the accuracy of subsequent cycles.

Newer technologies with smaller device sizes generally have less parasitics and faster transistors. Literature shows that implementations in newer technologies usually possess a better FoM than implementations in older technologies. ADC's in 0.8um and older CMOS Technologies have a FoM much higher than 10pJ/conversion cycle ([9] from 1996 has 25pJ). Newer Designs in nanometer technologies achieve FoM's close to and below 1pJ/conversion cycle ([10] and [11]). The designs presented in these papers have a supply voltage of 1.8V at a technology size of 0.18um and 1.2V at 90nm respectively (results in 0.56 pJ/conversion cycle). This huge improvement in power saving is a combination of newer design techniques and the advantage of the new technology.

5. Matching Considerations

So far we have discussed several approaches to speed up the conversion process. Of course for high speed operation it is favorable to keep the capacitance values as small as possible (mainly dictated by kT/C-noise).

However, this approach usually yields insufficient matching. Basically there are different architectures yielding different linearity properties. A binary structured approach leads to the well-known saw-tooth characteristic in the INL plot. A unit-source approach gives very good DNL and has a smoother INL characteristic. Figure 10 shows a comparison of INL curves using the same unit elements. In both cases a linear error distribution in the unit elements has been assumed. In Fig. 10a they are grouped in a binary form (all the smaller elements form the MSB, the larger ones form the LSBs) whereas in Fig. 10b a thermometer coding (again starting with small elements going to large elements) is assumed. If one does not worry about this type of INL but good DNL is of greater importance, then the second approach is better. The only drawback of this is that it requires more hardware effort since individual steering for each individual element is required. The binary approach is simpler, but at the cost of problems in INL and DNL. Numerous ways to improve the linearity can be found in current literature. In case of a constant gradient (strictly linear error distribution) a symmetrical switching scheme helps. The idea is to bring together pairs of elements with the same error magnitude but opposite sign. So in a thermometer decoded array an element with an associated error is first activated and then the next activated element is placed on the opposite side of the array.

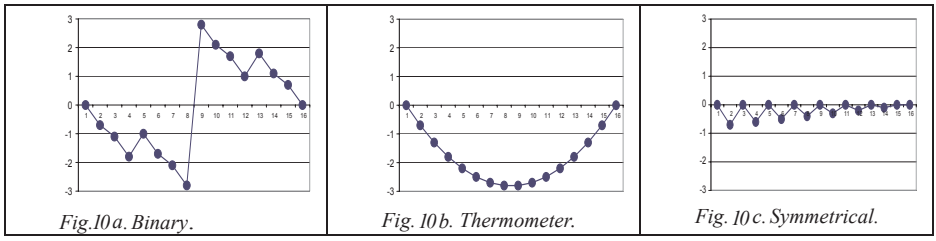


Fig.10a. Binary. *Fig. 10b. Thermometer.* *Fig. 10c. Symmetrical.*

Under the assumption of a linear error distribution this will then have the same error with the opposite sign. Fig. 10c shows the array used in Fig. 10a and Fig. 10b, but now the switch coding is implemented in a symmetrical way. This approach only works in the case of a linear error distribution. For different error distributions one has to go for different switching schemes. Since the error distribution in many cases is not very well known quite a number of switching schemes have been published in literature.

Since there is a tradeoff between INL and effort between a thermometer decoded array and a binary decoded array, a hybrid approach known as segmented decoding is also frequently used. MSBs are activated by a thermometer decoder, LSBs are done in a binary way.

6. Calibration Techniques

Finally a point is reached where all these measures are insufficient. In order to further improve linearity trimming could be carried out. A detailed investigation of this will not be presented in the context of this paper. Instead we want to have a closer look into several self-calibration techniques. In terms of self-calibration different approaches can be distinguished: e.g. foreground versus background, analog versus digital. In the following section several examples of these techniques will be presented.

Self-calibration of binary structures is based on the fact that each bit has to have the same size as the sum of all smaller bits. In other words: If a high-to-low transition is set on the MSB and a low-to-high transition on all the other bits ideally a zero reaction should be observed at the output. This can be easily tested and therefore individual bits can be adjusted until a perfectly binary weighted structure is obtained. This can be readily implemented, as done e.g. in the CS5014 (14 bit self-calibrating A/D-converter), but leads to the observation that each single bit has to be equipped with a number of auxiliary elements. A more compact self-calibration technique has been published in [16]. Here only a single array is required to correct for linearity errors at the expense of a somewhat more complex digital circuit. The correction methods shown so far all work in the analog domain, i.e. the analog values are made as close to ideal as possible.

[17] presents a different approach: Corrections are not carried in the analog domain, but in the digital domain. This means, that the equivalent code for each bit is stored in a memory and that the final output code is assembled as a sum of equivalent codes coming from memory. For binary structures this approach has the problem that missing codes cannot be corrected. For a non-binary (redundant) coding scheme however this works quite well. In [5] we have shown that it is even possible to do self-calibration for this type of non-binary structures. The approach used in [5] works as follows: Each bit can be represented as a fraction (smaller than 1) of the sum of all other bits. In other words: In order to get a zero reaction on the output a FS-step has to be applied at the bit under investigation and a somewhat smaller step, opposite direction, at all the smaller bits. Knowing the size of the smaller step means to know the size of the ratio. That smaller step can be digitized with the aid of the (still non-perfect) ADC. The interesting observation now is that this leads to a set of equations that can be solved yielding correct values for each bit.

The methods described so far are all foreground calibration methods. This means that the conversion process must be halted in order to do the calibration. [18] presents a method that tries to carry this out in the background. The idea is that, the test if two capacitors are equal can be done in a quite short amount of time (much shorter that a complete conversion cycle). So at the end of each conversion some time is reserved to do such a

measurement. This somewhat slows down the total conversion rate but allows for a continuous calibration, since the necessary digital calculations can be easily done in parallel to the normal conversion process.

It has taken some time until the idea of self-calibration has found its way also into pipeline converters but for linearity improvements calibration is a smart and beneficial method.

The static amplification in the MDAC is defined by

$$V = \frac{1}{\frac{Cf}{Ci} + \frac{1}{A0}} \tag{4}$$

To keep the gain well defined the open loop gain of the amplifier should be very high. This leads to very complex OP-amp architectures using cascodes with regulation amplifiers and other power hungry elements. The requirement of the high open loop gain is a big challenge, especially in new technologies. The modern CMOS processes are trimmed for high speed digital applications which also enables high speed (high sampling rate) analog designs, but do not offer good current sources. That means that the amplifier gain is very limited even with cascoding tricks. In a 65nm CMOS technology, a simple two stage miller amplifier without cascodes typically has a gain of less than 30dB. It is nearly impossible to shift the gain up to 60 to 70 dB which would be necessary to build an ADC with an accuracy greater than 10bit. This problem could be solved by trimming or by calibration means, as shown later.

Figure 11 shows the residue plot of a 1.5bit architecture in an ideal case (black) and in presence of an inter-stage gain error (red).

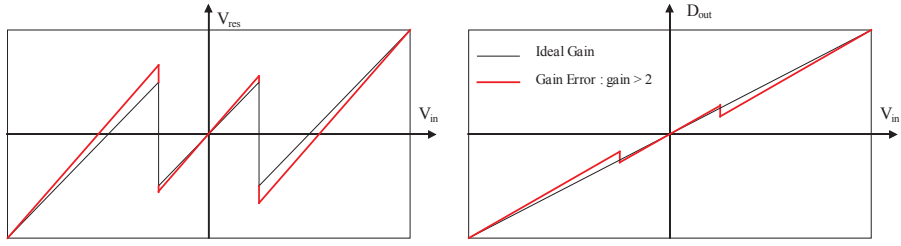


Fig. 11. Residuum and Digital output with interstage gain error.

The second major error is the DAC error, which is caused by mismatch of the DAC capacitors. The root and the remedy is the same as in the SAR as already discussed before. Figure 12 shows the effect to the residuum in the pipeline ADC in a 1.5bit architecture.

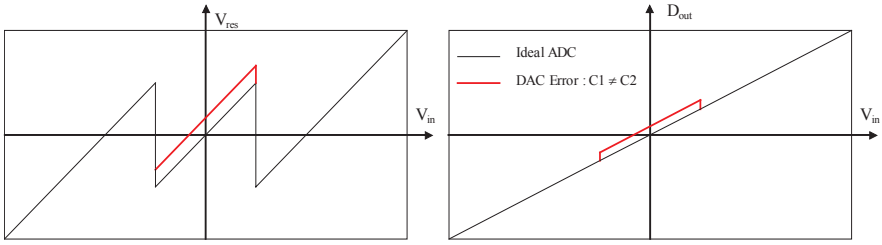


Fig. 12. Residuum and Digital output with DAC errors.

Both errors can be observed in the INL and DNL, as in the SAR. To enhance the linearity many calibration methods have been developed. Initial calibration methods in pipeline ADC's were very straightforward: compensate the error source in the analog domain. The design of [19], implemented in 1991, used a calibration in the analog domain with the aid of trimming capacitors for the DAC and the gain errors. For the calibration phase the normal operation of the ADC was interrupted, which is possible in certain applications (in this case a video application). The calibration sequence was implemented exactly as the method discussed for the SAR ADC. But it is obvious that an interruption of the normal operation is not possible in every application.

Nowadays, the trend for calibration is digital calibration. Such architectures exploit the possibilities in high speed CMOS technologies, which are perfectly suited for huge digital parts. The calibration is usually based on a correlation method and makes use of a calibration sequence that is fed into the ADC. A very straight forward method is described in [20]. It disconnects the input of the ADC and feeds in a calibration signal generated by an accurate sigma delta DAC. The ADC is fully calibrated for all non idealities in the digital domain.

A more enhanced method is background calibration. The ADC still works in a normal mode and performs the calibration transparently to the user. Several methods have been investigated in recent years since these are also most promising for future technologies. No interruption is necessary and therefore these ADCs can be used for any application. Most of the architectures have the same block diagram as shown in Fig. 13. A pseudo-random sequence is used as a calibration signal. In Fig. 13 only stage1 is calibrated. This can be extended to any stage, but of course the first stage has the most stringent accuracy requirements.

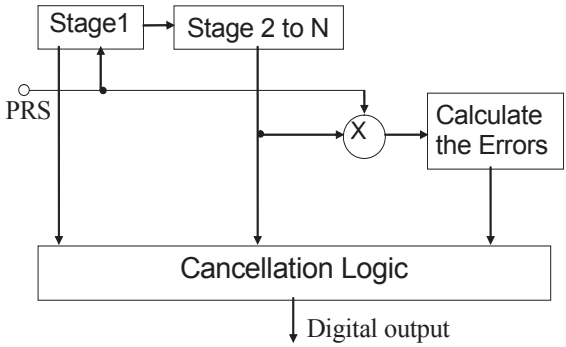


Fig. 13. Typical architecture for digital background calibration.

Figure 13 shows the architecture used in [12]. The signal is fed into the signal chain with the aid of an extended MDAC of stage 1. The inter-stage gain error and the DAC errors can be calculated with a correlation method and subtracted. [21] and [22] present a quite similar architecture. In these works dynamic element matching in the DAC is used to measure the errors and subtract them at the output. [23 to 25] also present methods to implement digital background calibrations. We expect further interesting developments to occur in this area in the near future.

7. Comparisons

In order to do a fair comparison, we should compare similar things. Although pipeline-, cyclic- and SA-converters use binary search as the fundamental operating principle, their circuit implementations are quite different. From the algorithmic point of view the pipelined and the cyclic converter are very similar: both of them use a compare / conditional subtract / amplify-by-two mechanism for each bit. Except for the fact that the pipelined converter operates at n bits (n … number of stages) in parallel, whereas the cyclic converter can handle only one bit per cycle, it also can be observed that the cyclic converter has to handle quite large capacitors in each cycle. The capacitors have to be scaled (at least) according to the kT/C-requirements for the full resolution. Since there is only a single stage, the LSBs also have to be handled using these big capacitors. This seems to be a major drawback with respect to the pipelined converter, since here usually a scaling of the individual stages can be carried out.

From the above, it can be seen that the pipelined converter can be expected to be not only faster, but also more power efficient than the cyclic converter. Therefore, the following section will compare the pipelined

converter with the SAR. Also the SAR converter has the drawback that scaling at subsequent stages can not be carried out, since there are no subsequent stages. However, comparing a charge-redistribution ADC with a pipelined converter reveals that the (analog-related) hardware of the SAR is much simpler. It consists of just a single capacitive network (sized the same way as the sampling capacitor in the pipelined converter) some switches and a comparator. The requirements for the comparator are more stringent in the SAR case, since the full resolution is needed. This is also true for the basic pipelined architecture, but by using some redundancy (e.g. using the 1.5b/stage architecture) these requirements can be relaxed significantly. Using redundancy in the SAR does not really help in this context, since there is no amplification between successive decisions.

A major difference between pipeline and SAR is the amplifier. Usually this is the most power-hungry building block in a pipelined converter, it is nonexistent in the SAR. So the main point of comparison is between the comparator within the SAR and the amplifier within the pipeline. Of course the power efficiency of a comparator can be much higher than the power efficiency of an amplifier, since in the latter case power is lost due to the stability requirements. On the other hand, full amplifier performance is only required for the first pipeline stage, whereas there is no scaling option for the comparator in a SAR. This could mean that for high resolutions the pipelined converter is more favorable than for low resolutions. In new technologies this picture may change because it is much easier to build high-resolution comparators than to build high-gain operational amplifiers, which is in favor of the SAR. There are some attempts to design simpler amplifiers. This involves calibrating away gain-related conversion errors. [8, 25] even take into account some nonlinear effects (more power efficient open loop amplifiers tend to introduce nonlinearity). Although this is very interesting to further increase power efficiency, in our opinion these techniques can not be called mature.

In the comparisons done so far the energy consumption had to be normalized towards an average bit cycle (or, to compare a pipelined cycle with a complete SAR conversion, since both of them deliver a full output word).

For making a more direct comparison a pipelined converter has to be compared to an array of time-interleaved SA-converters. The design described in [26] also delivers a full output word each clock cycle. The design in [3] goes even one step further: It combines redundancy and self-timing within a time-interleaved converter architecture. Both of the abovementioned papers show extremely good figures of merit far below 1pJ/conv. step. However, at the same time they are to some extent incomplete, since there are some remaining artifacts due to the time-interleaved nature and (in [3]) the code translation into binary is also missing. In addition, these are 6 bit converters whereas the resolution of

pipelined converters is usually higher. To our knowledge there are no recent demonstrations of higher resolution time-interleaved SAR-converters available. Since the demonstrated figures-of-merit of the abovementioned SAR converters are extremely attractive we look forward to see some more research in this area.

8. Conclusions

Comparisons between SAR and pipelined ADCs have been made. Many similarities can be found. The attractiveness of a SAR approach mainly lies in the very simple analog circuitry required, whereas the attractiveness of a pipelined ADC lies in its compactness at high sampling rates. Therefore, for higher sampling frequencies pipeline ADC's will still be used in future designs. However, for nanometer technologies ADCs with higher resolution (>8bit) should be implemented with some form of calibration. The alternative, a time-interleave SAR array, may be simple in terms of circuitry but suffers from mismatch issues and would benefit greatly from some form of calibration. We expect to see further research on this topic in the future.

References

[1] J. McCreary et al.: "A High-Speed, ALL-MOS Successive-Approximation Weighted Capacitor A/D Conversion Technique", ISSCC 1975, pp. 38-39.

[2] R. Peck, D. Schroeder: "A low-power entropy-coding analog/digital converter with integrated data compression", ESSCIRC 2003, pp.173-176.

[3] S.-W. Chen et al.: "A 6b 600MS/s 5.3mW Asynchronous ADC in 0.13µm CMOS", ISSCC 2006, session 31.

[4] Z.G. Boyacigiller et al.: "An Error-Correcting 14b /20µs CMOS A/D Converter", ISSCC 1981, pp. 62-63.

[5] D. Draxelmayr: "A Self-Calibration Technique for Redundant A/D Converters Providing 16b accuracy", ISSCC 1988, pp. 204-205.

[6] F. Kuttner: "A 1.2V 10b 20MSample/s Non-Binary Successive Approximation ADC in 0.13µm CMOS", ISSCC 2002, pp. 176-177.

[7] A. Abo, P. Gray: "A 1.5V, 10-bit, 14.3-MS/s CMOS Pipelined Analog-to-Digital Converter", IEEE JSSC, vol. 34, no. 5, pp. 599-606, May 1999.

[8] B. Murmann, B. E. Boser: "A 12b 75MS/s Pipeline Adc Using Open-Loop Residue Amplification", ISSCC 2003, pp. 328-329.

[9] G. Kang et al.: “A high performance CMOS A/D Converter for a digital camcorder system”, IEEE Transactions on consumer electronics, August 1996, Volume 42, Number 3, pp. 285-289.
[10] M. Yoshioka, M. Kudo, K. Gotoh, Y. Watanabe “A 10b 125MS/s 40mW Pipelined ADC in 0.18um CMOS” ISSCC Dig. Tech. Papers, Session 15, pp. 282-283, Feb. 2005
[11] D. Kurose, T. Ito, T. Ueno, T. Yamaji, T. Itakura “55-mW 200-MSPS 10-bit Pipeline ADCs for Wireless Receivers” ESSCIRC Proceedings, pp. 527-530., Sept. 2005.
[12] P. Bogner, F. Kuttner, C. Kropf, T. Hartig, M. Burian, H. Eul “A 14b 100MS/s digitally self calibrated Pipelined ADC in 0.13µm CMOS” ISSCC 2006, Session 12, Feb. 2006.
[13] D. Kelly et al.: “A 3V 340mW 14b 75MDPS CMOS ADC with 85dB SFDR at Nyquist” ISSCC 2001, pp. 134-135.
[14] H.C. Kim, et al.: “A 30mW 8b 200MS/s Pipelined CMOS ADC using a Switched-Opamp Technique” ISSCC 2005, Feb 2005, pp. 284-285.
[15] K. Nagaraj et al.: “A 250mW, 8b, 52Msample/s Parallel-Pielined A/D converter with Reduced Number of Amplifiers,” IEEE J. Solid-State Circuits, vol. 32, no. 3, pp. 312-320, March, 1997.
[16] H. Lee et al.: “A Self-Calibrating 15 Bit CMOS A/D-Converter”, IEEE Journ. of Solid-State Circuits, Dec. 1984, pp. 813-819.
[17] Y. Akazawa et al.: “New Linearity Error Correction Technology for A/D and D/A Converter LSI”, proceedings of the 14th Conference on Solid State Devices, Tokyo, 1982, pp. 115-119.
[18] F. Kuttner: US patent 5825316
[19] Y. Lim et al.: “A 13-b 2.5MHz Self-calibrated Pipelined A/D Converter in 3µm CMOS”, IEEE JSSC, vol. 26, no 4, April 1991, pp. 628-636.
[20] X. Li et al.: “A Fast and Accurate Calibration Method for High-speed High-resolution pipeline ADC”, IEEE JSSC II, 2002, pp. 800-803.
[21] E. Siragusa, I. Galton: “A Digitally Enhanced 1.8V 15b 40MS/s CMOS Pipelined ADC”, ISSCC 2004, pp. 452-453.
[22] P. Yu et al.: “A 14b 40MSample/s Pipelined ADC with DFCA”, ISSCC 2001, pp. 136-137.
[23] H. Liu et al.: “A 15b 20MS/s CMOS Pipelined ADC with Digital Background Calibration”, ISSCC 2004, pp. 454-455.
[24] K. Nair, R. Harjani: “A 96dB SFDR 50MS/s Digitally Enhanced CMOS Pipeline A/D Converter”, ISSCC 2004, pp. 456-457.
[25] C. Grace et al.: “A 12b 80MS/s Pipelined ADC with Bootstrapped Digital Calibration”, ISSCC 2004, pp. 460-461.
[26] D. Draxelmayr: “A 6b 600MHz 10mW ADC Array in Digital 90nm CMOS”, ISSCC 2004, pp. 264-265.

FAST AND POWER-EFFICIENT CMOS SUBRANGING ADCs

F.M.L. van der Goes, J. Mulder, C.M. Ward, C.-H. Lin,
D. Kruse, J.R. Westra, M. Lugthart, E. Arslan,
O. Bajdechi, R.J. van de Plassche and K. Bult
Broadcom Netherlands, The Netherlands

Abstract

This paper presents a two-step subranging ADC architecture based on interpolation, averaging, offset compensation and pipelining techniques. Application of these techniques results in fast and power-efficient converters with an accuracy between 8b and 12b.

1. Introduction

Area and power are important parameters for analog-to-digital converters (ADCs) integrated on large digital ICs. This is especially true if several ADCs have to be implemented on the same die. These embedded ADCs necessarily have to be designed in state-of-the-art digital CMOS processes, and therefore, have to work at low supply voltages, while special process options, such as low-threshold devices, are often not available.

The ADCs presented in this paper have an 8b to 12b resolution and clock speeds of 75MS/s and up. Several ADC architectures are capable of achieving these specifications, e.g. flash ADCs [1], pipeline ADCs [2–4], folding ADCs [5–9] and subranging ADCs [10–12]. However, when small die area, low-power and low-voltage operation are of primary importance, the two-step subranging architecture has proven to be a very suitable choice. A key advantage of this architecture is that the two-step approach allows for an area and power efficient design. Additionally, it can use simple differential-pair amplifiers, which are very well fit for low-voltage operation.

The basic architecture of the two-step subranging ADC, comprising a coarse ADC (CADC) and a fine ADC (FADC), is explained in Sec. 2. Sec. 3 discusses one of the key elements in this design, the voltage subtractor. The timing of this subranging ADC is explained in Sec. 4. Techniques that are used to obtain a low-power and low-area ADC are discussed in Sec. 5. Sec. 6 explains

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 53–71.

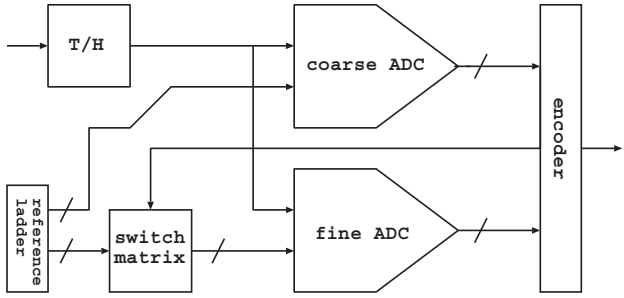


Fig. 1. Two-step subranging ADC architecture.

techniques to increase the maximum clock speed without increasing the power consumption. Finally, experimental results of various ADCs, implemented in 0.13μm CMOS technologies, are presented in Sec. 7. The design will be based on an 8b converter, but similar design strategies can be applied for 10b and 12b ADC's.

2. ADC Architecture

The ADCs presented in this paper use a two-step subranging architecture [13–16], shown in Fig. 1, comprising two flash ADCs, a reference ladder, a switch matrix and an encoder. The subranging ADC is preceded by an on-chip track-and-hold (T/H). Each conversion cycle, the CADC first performs a rough quantization of the ADC input voltage. The switch matrix then connects the FADC to the appropriate subrange of the reference ladder, determined by the CADC. The FADC compares the ADC input voltage against the selected reference voltages and quantizes the input voltage at full resolution. The digital encoder combines the CADC and FADC thermometer output code data into the final digital output code. Note that both the CADC and the FADC use the same reference ladder and the same T/H output signal. This guarantees matching of input and reference voltages between the two ADCs and saves chip area. A clock generator produces a non-overlapping two-phase clock, used to clock the switches inside the T/H and the ADC. The two clock phases are denoted by ϕ_1 and ϕ_2, respectively.

A minimum number of comparators is obtained when using both a $N/2$ bits CADC and a $N/2$ bits FADC, where N is the number of ADC bits. However, this requires the errors made by the CADC to be smaller than one least significant bit (LSB) of the FADC (LSB_{FADC}).[1] It is not possible to accomplish this without excessive power consumption inside the CADC. It is, therefore, com-

[1] Note that this does not apply to the CADC quantization errors.

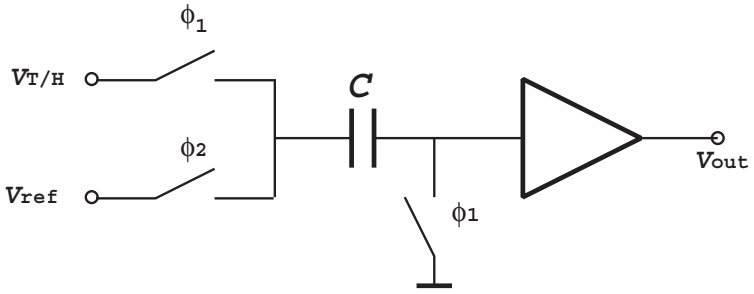


Fig. 2. Switched Capacitor Voltage Subtractor.

mon practice to use some amount of over-range [13–15, 17–19], which allows the FADC to correct for errors of the CADC. If the CADC errors are smaller than the over-range, the overall ADC performance is only determined by the accuracy of the FADC and not by the CADC performance.

Over-range can be obtained by either increasing the number of bits in the CADC or in the FADC. An important restriction of the CADC is that its latency has to be low, because the CADC decision has to be available before the FADC can start to convert the input signal. In the FADC, latency is less important and pipelining techniques can be used to decrease power consumption. Therefore, we preferred to increase the number of bits in the FADC, resulting in a 4b CADC and 5b FADC. For the 10b and 12b ADC, the allocation for the CADC/FADC is 5b/6b and 6b/7b respectively.

3. Voltage Subtraction

As discussed in Sec. 2, both the CADC and FADC compare the T/H output voltage $V_{\mathrm{T/H}}$ to a set of reference voltages. The CADC uses reference voltages that are covering the full input signal range and the FADC uses a set of reference voltages that are close to the input signal. A subtraction circuit is used to perform this comparison. The subtractor that is used in this subranging ADC is based on Switched Capacitor (SC) circuits. The basic principle of a SC circuit is shown in Fig. 2.

During the first phase (ϕ_1), the voltage $V_{\mathrm{T/H}}$ is sampled on capacitor C. During the next phase (ϕ_2), the amplifier amplifies the voltage difference $V_{\mathrm{ref}} - V_{\mathrm{T/H}}$. Bottom-plate sampling is used to improve the accuracy. An array of these SC circuits is applied in both the input of the FADC and CADC.

Besides the voltage subtraction, this circuit serves other purposes as well. First, the SC circuit allows an elegant timing between the T/H, CADC and

FADC. This will be discussed in Sec. 4. Secondly, the input capacitors allow the independent optimization of the T/H common-mode output voltage and the common-mode input voltages of the CADC/FADC.

4. Timing

A crucial point in the design of a subranging ADC is the timing between the CADC and the FADC, especially when a T/H drives the ADC. This is because the T/H output voltage is only available during Hold mode, which is half of the time. The FADC needs to wait for the digital output of the CADC to become available before it can start to convert the input signal. When the processing of both the CADC and FADC has to take place during Hold mode, they both are only active during part of the Hold mode. This is not very efficient and results in a low maximum clock frequency. A more efficient timing uses the fact that the SC circuit in Fig. 2, besides subtraction, also delays the T/H output voltage by half a clock cycle. This property can be used to delay the signal for only the FADC, allowing half a clock latency for the CADC.

Delaying only the input for the FADC and not for the CADC can easily be implemented by operating the SC circuit at the input of the CADC and FADC out of phase. In this way, both the CADC and FADC are connected to the T/H during Hold mode. The CADC is now in amplify mode while the FADC is in sampling mode. During Track mode, both ADC's operate in the opposite phase. This timing is shown in Fig. 3. Note that only the clock signal for the switch at the input of the amplifiers is different.

Since the FADC is a flash converter, the array of SC circuits acts as a distributed T/H. Usually, a distributed T/H have difficulties in accurately tracking of a signal. However, this T/H operates on a non-moving signal (the main T/H is in Hold mode) and high tracking accuracies can easily be obtained.

5. Low-Power Low-Area Techniques

This section describes techniques that are applied in the FADC and CADC in order to reduce power and area. Pipelining offset compensation, averaging and interpolation are being applied for this purpose. Before these techniques will be discussed, the unit amplifier cell will be considered. Both the CADC and FADC are flash converters and they consist of arrays of unit amplifier cells. This cell consists of a cascoded differential-pair with resistive loads, as shown in Fig. 4. A fully differential implementation is used to achieve low sensitivity to substrate noise and to achieve sufficient power supply rejection. The gain of the amplifier is set by noise and speed and is designed to be 4.

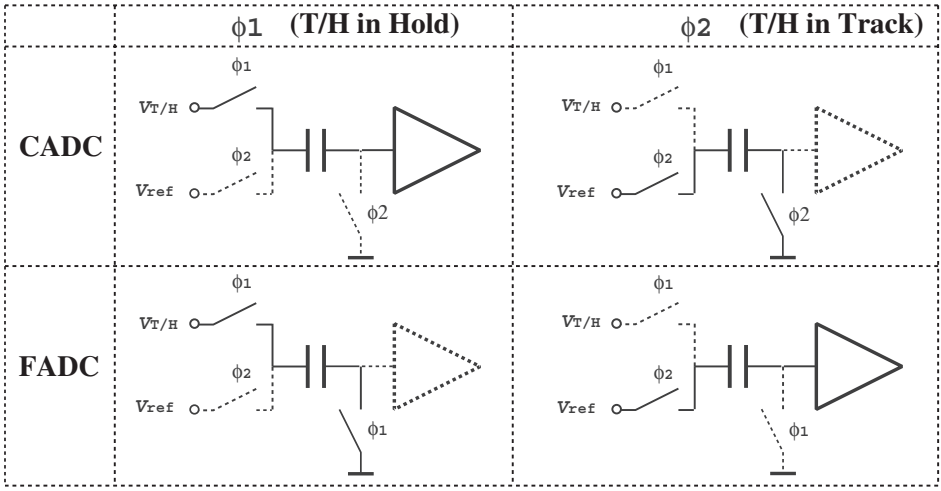


Fig. 3. CADC/FADC modes during both clock phases.

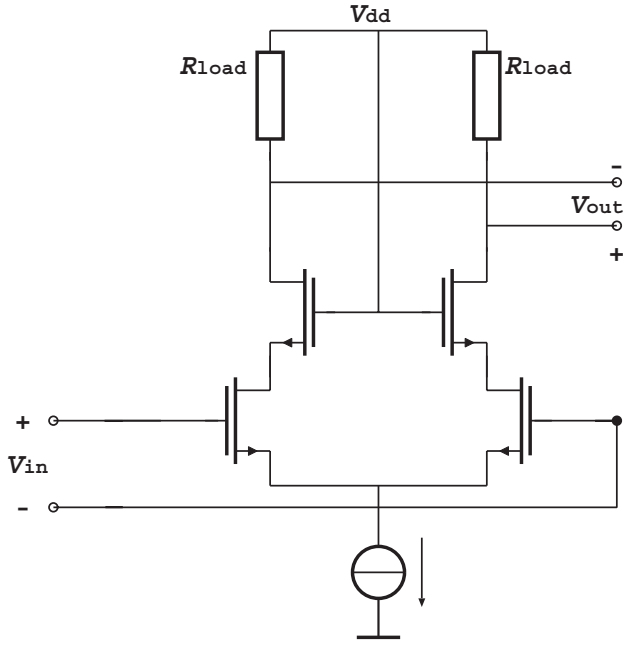


Fig. 4. Unit amplifier cell that is used in the CADC and FADC.

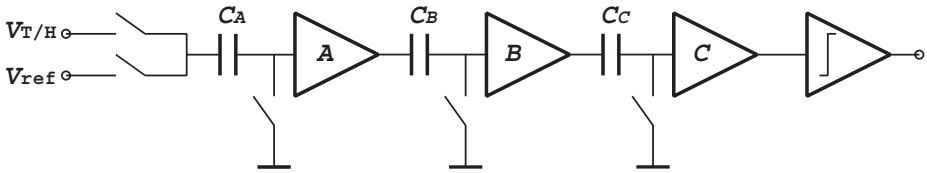


Fig. 5. Block diagram of the FADC.

5.1. Pipelining

In ADCs comprising cascaded arrays of amplifiers, pipelining is often used to improve conversion speed [7, 10, 14, 20]. In effect, conversion latency is traded against amplifier speed, and hence, power consumption.

The pipelining is enabled by inserting capacitors between the arrays of amplifiers. The 8b converter uses 3 arrays of amplifiers in cascade. This is mainly to overcome the offset of the comparators. The functional block diagram of the FADC is shown in Fig. 5. The amplifier arrays, denoted by $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$, are followed by an array of comparators. The 10b and 12b converter use 4 and 5 arrays of amplifiers in cascade.

The inter-stage capacitors that are used for pipelining are also used for auto-zeroing purposes, as will be discussed in Sec. 5.2.

5.2. Offset Compensation

Offset compensation to all diff-pairs is applied. As a consequence, small devices can be used in these amplifiers, allowing the use of lower bias currents, which helps to reduce the power consumption of the FADC. Two different implementations are applied. These are shown in Fig. 6.

The circuit in Fig. 6a uses a *closed-loop* technique. The offset is stored on the capacitors during the reset phase (switches S_1 are closed). In contrast, the circuit in Fig. 6b is an *open-loop* auto-zero technique [13, 21]. Here, the offset of the amplifier is stored on the output capacitors C_3 during the reset phase. The reset switches are implemented by the differential switch S_3 and the common-mode switches S_2. Although the common-mode switches also perform differential resetting, the differential switch is more effective.

Accurate offset compensation of the open-loop technique requires a good linearity of the amplifier and can therefore only be used in the front-end of the ADC where the signals are still small. The closed-loop technique doesn't have this linearity problem. Obviously, since the last amplifier array is not followed by capacitors, the closed-loop technique has to be applied here.

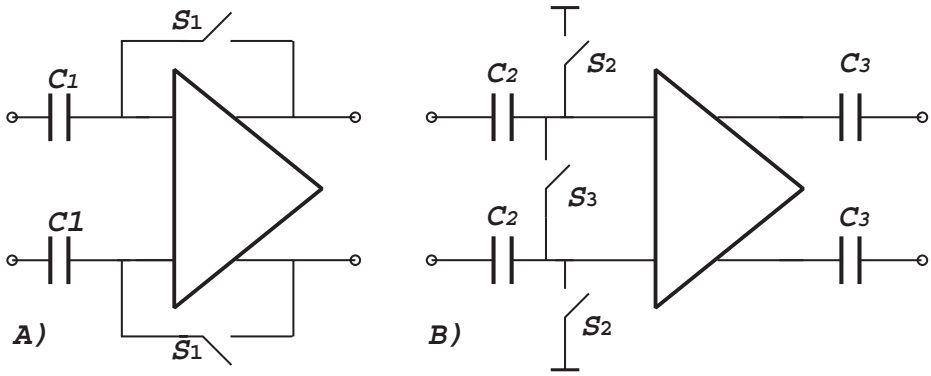


Fig. 6. Two implementations of offset compensation.

5.3. Averaging

Averaging is widely used in flash, folding and subranging ADCs [8, 9, 14, 22–25] to improve ADC performance. In an array of amplifiers, the contributions of random mismatch and noise to the output signals of neighboring amplifiers are mostly uncorrelated. Hence, when averaging the outputs of several amplifiers, the average output signal has a higher signal-to-noise ratio (SNR) than the individual output signals.

Although this subranging ADC design already implements offset compensation, averaging still has a beneficial influence on the performance. First of all, averaging helps to decrease the influence of noise generated in the amplifiers and switches. Secondly, the switches, required for auto-zeroing and pipelining, introduce random mismatches themselves, which cannot be auto-zeroed. Averaging decreases this charge-injection mismatch. As a result, amplifier bias levels and capacitor sizes can be reduced, which reduces both power consumption and chip area.

The capacitors already present between array $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ can be reused to implement $2\times$ capacitive averaging. As shown in Fig. 7, the input capacitors of row $\mathcal{B}$ and $\mathcal{C}$ are split into two equal halves, connecting to two different input voltages $V_{\mathrm{in},0}$ and $V_{\mathrm{in},1}$. When the amplifier is in reset-mode, the capacitors are charged to $V_{\mathrm{in},0}$ and $V_{\mathrm{in},1}$, respectively. During amplify-mode, the charges on the capacitors are redistributed and an output voltage V_{out} results that is proportional to the average value of $V_{\mathrm{in},0}$ and $V_{\mathrm{in},1}$, i.e., $V_{\mathrm{out}} \sim \frac{1}{2}(V_{\mathrm{in},0}+V_{\mathrm{in},1})$.

Of all three amplifier arrays, the amplifiers in the first array are most important with respect to the ADC performance. Therefore, it would be beneficial if, e.g., $4\times$ averaging of these amplifiers could be achieved, instead of the $2\times$ provided by the circuit shown in Fig. 7.

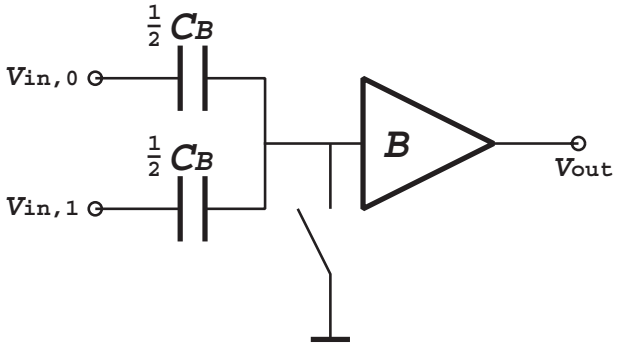


Fig. 7. Implementation of 2× capacitive averaging.

Normally, the optimal averaging window should be based on the shape of the g_{m} curve of the involved amplifiers. Since the input signals are still small, the best approximation of the g_{m} curve is a straight line and this results in a rectangular 'averaging window'.

The most efficient way to obtain this is by using the distributed averaging shown in Fig. 8. Although using only 2 capacitors at each differential-pair input, instead of 4, 4× averaging of the $\mathcal{A}$ amplifiers is achieved. This is accomplished by 'skipping' one $\mathcal{A}$ amplifier when connecting the two inputs of an amplifier in $\mathcal{B}$ to the outputs of two amplifiers in $\mathcal{A}$. The black-colored amplifiers demonstrate that this topology can indeed achieve 4× averaging of the $\mathcal{A}$ amplifiers. Suppose the FADC input signal, after quantization, is located halfway between the inputs of amplifiers $\mathcal{A}_1$ and $\mathcal{A}_2$, then the 'zero-crossing' at the output of $\mathcal{C}_1$ is determined by amplifiers, $\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ in array $\mathcal{A}$, amplifiers $\mathcal{B}_1$ and $\mathcal{B}_2$ in array $\mathcal{B}$ and amplifier $\mathcal{C}_1$ in array $\mathcal{C}$. It can be easily derived that all amplifiers within the averaging window of array $\mathcal{A}$ are equally weighted. The same applies to row $\mathcal{B}$.

5.4. Interpolation

Interpolation is generally used to reduce the number of amplifiers required in an ADC [5–11, 15, 16, 19, 24, 26, 27], thus decreasing layout complexity, which is imperative for achieving small die area and low-power operation. Moreover, interpolation reduces the number of reference voltages required, so less signals have to be routed in the layout. This helps to achieve a small area, and, because of less parasitic capacitance, it also reduces the power consumption of the ADC. Additionally, in subranging ADCs, interpolation reduces the number of switches required in the switch matrix [15]. Since the switch matrix

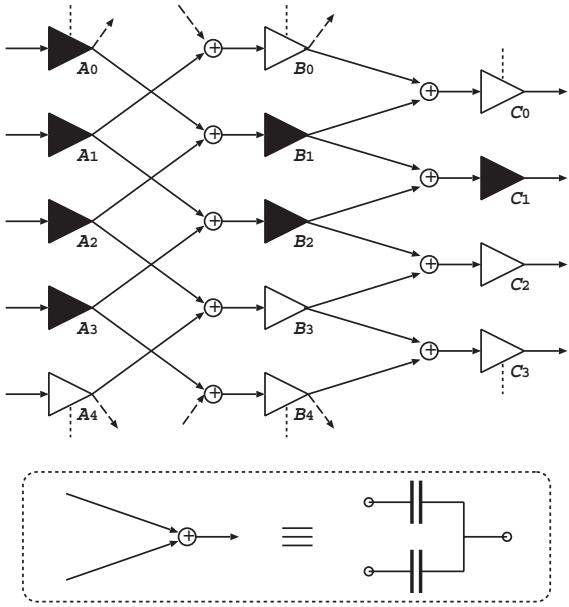


Fig. 8. Distributed averaging topology, providing 4× and 2× averaging of the amplifiers in array $\mathcal{A}$ and $\mathcal{B}$, respectively.

represents a significant capacitive load to the T/H and the reference ladder, interpolation indirectly helps to reduce the power consumption of the T/H and the reference ladder.

The ADC uses 2 different interpolation techniques, one based on splitting of capacitors and one based on splitting of differential-pairs. The last one is active interpolation.

A 2× capacitive interpolation technique is shown in Fig. 9, where the input capacitors in row $\mathcal{A}$ are reused to implement 2× capacitive interpolation. Each original capacitor $C_{\mathcal{A}}$ is replaced by two capacitors with half the size. Note the circuit similarity between 2× interpolation and the averaging shown in Figure 7.

The active interpolation is shown in Fig. 10. By combining the output currents of two differential-pairs, with input voltages $V_{\mathrm{in},0}$ and $V_{\mathrm{in},1}$, respectively, an output voltage V_{out} is obtained that is proportional to the average of $V_{\mathrm{in},0}$ and $V_{\mathrm{in},1}$, i.e., $V_{\mathrm{out}} \sim \frac{1}{2}(V_{\mathrm{in},0} + V_{\mathrm{in},1})$.

In this design, a total of 4× active interpolation is used. The first 2× active interpolation is implemented in the amplifiers of row $\mathcal{B}$. Since it is difficult to implement active interpolation in row $\mathcal{C}$, because of the closed-loop auto-zero method used in that array, the second 2× is implemented in the differential-

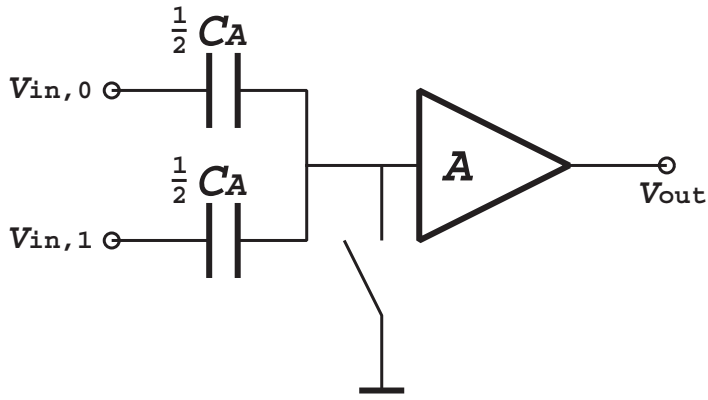


Fig. 9. Implementation of 2× capacitive interpolation.

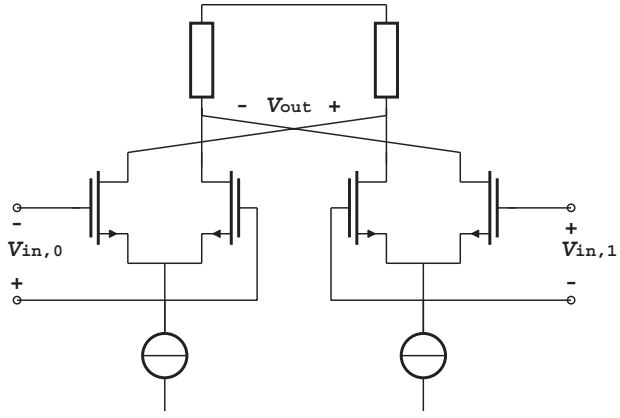


Fig. 10. Simplified schematic of a differential-pair amplifier providing 2× interpolation.

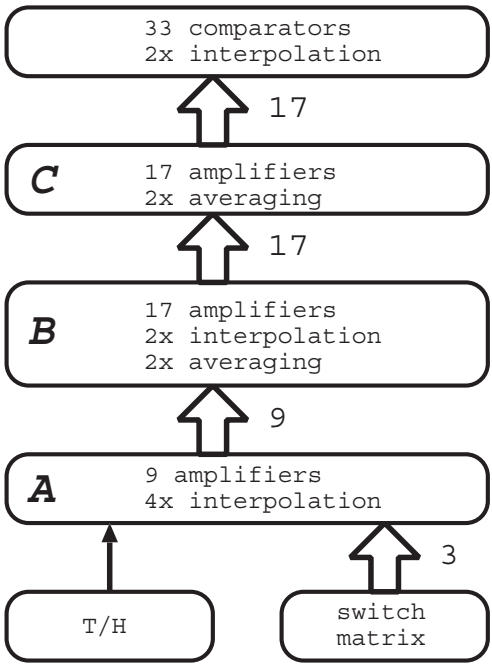


Fig. 11. Block diagram of the FADC.

pair input stage of the FADC comparators. This brings the total number of amplifiers in each array to 9, 17 and 17, respectively. This is shown in Fig. 11. A 4× capacitive interpolation at the input of array $\mathcal{A}$ results in a total of 16× interpolation and this results in only 3 output wires of the switch matrix. The amount of interpolation in the FADC for the 10b and 12b ADC is 32× and 64× respectively.

6. Speed Improvements

The finite on-resistance of switches S_2 and S_3, shown in Fig. 6b, is a major problem at high sampling rates. This is illustrated in Fig. 12. During clock phase ϕ_1, amplifier $\mathcal{A}$ is in reset-mode and samples $V_{\mathrm{T/H}}$. The current I_{charge} charging the input capacitor gives rise to a non-zero voltage V_ϵ at the input of $\mathcal{A}$ that settles exponentially down to zero. This error voltage is amplified by amplifier $\mathcal{A}$ and spread out in time, because of the finite speed of $\mathcal{A}$. Because the FADC uses pipelining, amplifier $\mathcal{B}$ is in amplify-mode and therefore further amplifies the error voltage. As a consequence, the error ripples through the complete FADC. Eventually, all error voltages settle to zero, but they considerably limit the maximum conversion speed of the FADC.

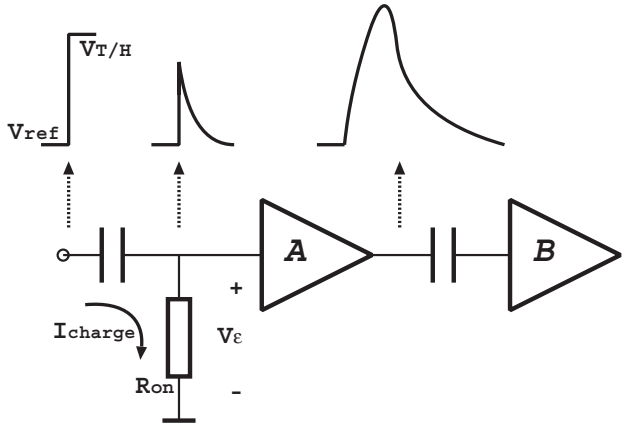


Fig. 12. Effect of finite on-resistance of the reset switches at high conversion rates.

Figure 13 shows an implementation of the reset switches in row $\mathcal{A}$ and $\mathcal{B}$ that significantly improves the pipelining operation in the FADC at high conversion rates. In reset-mode, when all switches are turned on, the on-resistance of switches M_4, M_5, M_6 and M_7 is equal. Switches M_4 and M_6 form a resistive divider, the output of which is a *common-mode* voltage. The same applies to M_5 and M_7. As a consequence, the *differential-mode* error voltage developing across the drain and source terminal of M_2 is not transferred to the gates of the differential-pair amplifier. In other words, the differential-mode transfer function of these cross-coupled switches equals zero in reset-mode. During amplify-mode, when only M_4 and M_7 are turned on, the differential-mode transfer function equals one. The fact that the error voltage no longer ripples through the complete FADC significantly improves the maximum conversion rate of the ADC.

The cross-coupled switches introduce some additional noise and charge-injection mismatch, which have only a small influence on the FADC performance in this design.

7. Experimental Results

This section gives a full measurement overview of the 8b converter and key numbers for 10b and 12b designs. The 8b ADC was fabricated in a 0.13μm CMOS process. A die micrograph is shown in Fig. 14. The ADC occupies only

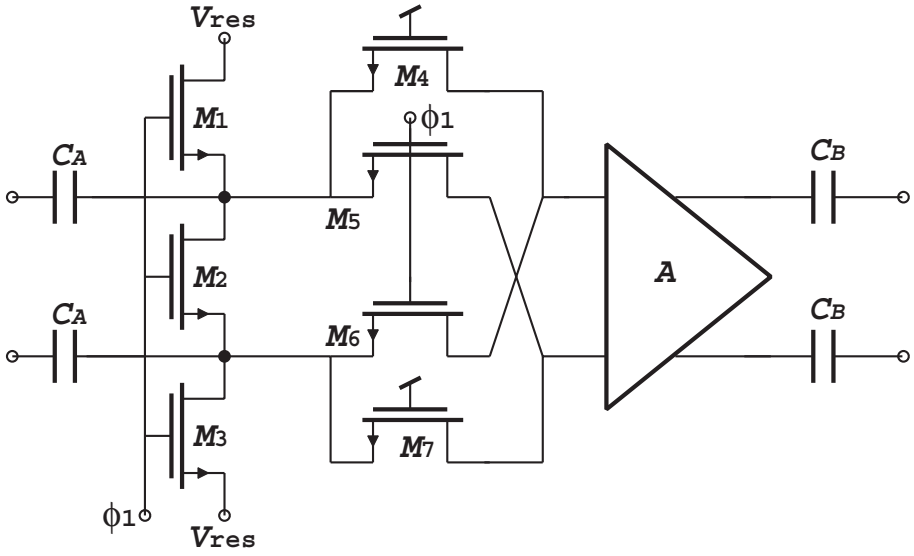


Fig. 13. Implementation of the reset switches in array $\mathcal{A}$ and $\mathcal{B}$ of the FADC.

290μm $\times$ 316μm = 0.09mm^2, including the on-chip T/H and clock generator circuit.

The reference ladder is implemented as a 390μm long continuous strip of poly silicon. In order to fit in the layout, the reference resistor had to be folded at least once. For symmetry with respect to the 17 reference taps, it was folded 8 times. An additional benefit is that the routing of the reference voltages to the CADC and switch matrix can be shorter, which reduces parasitic capacitance [28].

Figure 15 shows the measured differential nonlinearity (DNL) and integral nonlinearity (INL) curves of the ADC, measured at $f_{\mathrm{clk}} = 125$MS/s. The INL is smaller than ± 0.25b and the DNL smaller than ± 0.15b.

Figure 16 shows a clock sweep using an 8MHz input signal. The ENOB of the ADC equals 7.6b at 125MS/s, dropping half a bit to 7.1b at 220MS/s. At 125MS/s, the SNR equals 7.7b and the total harmonic distortion (THD) (including the 2^{nd} through 10^{th} harmonic) is 9.7b.

Figure 17 shows an input signal frequency sweep at 125MSample/s. The effective resolution bandwidth (ERBW) exceeds 100MHz.

Figure 18 shows the ADC output spectrum. The spurious-free dynamic range (SFDR) is 68dB.

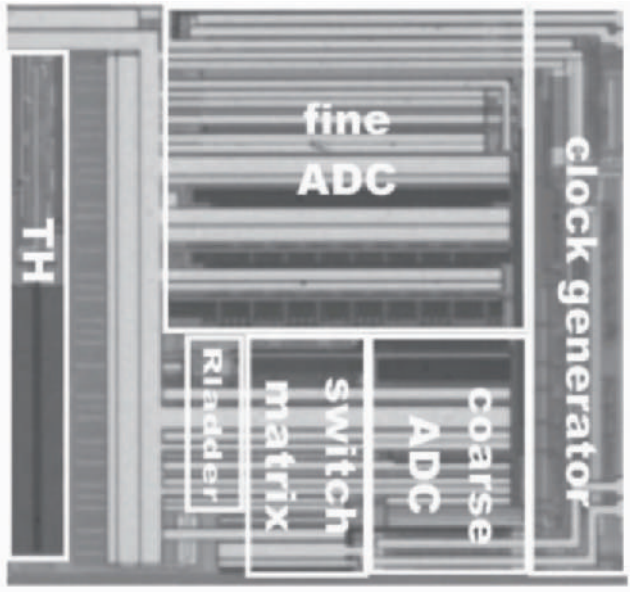


Fig. 14. Micrograph of the 8b ADC.

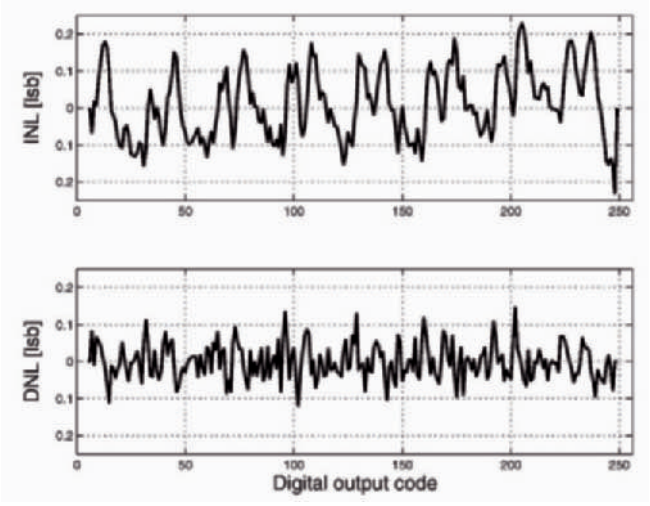


Fig. 15. Measured INL and DNL of the 8b ADC.

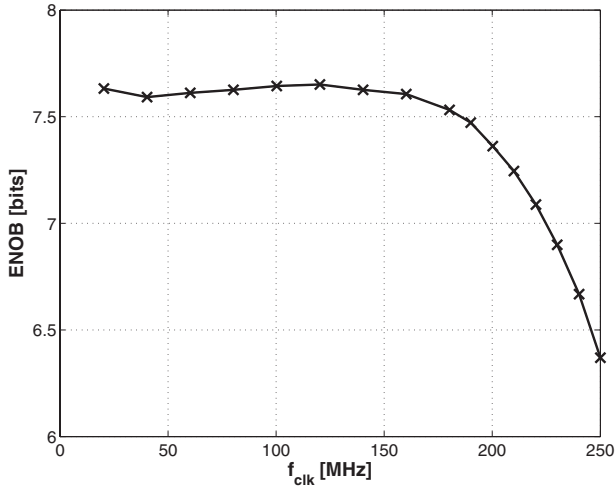


Fig. 16. Measured ADC performance vs f_{clk} at f_{in} = 8MHz.

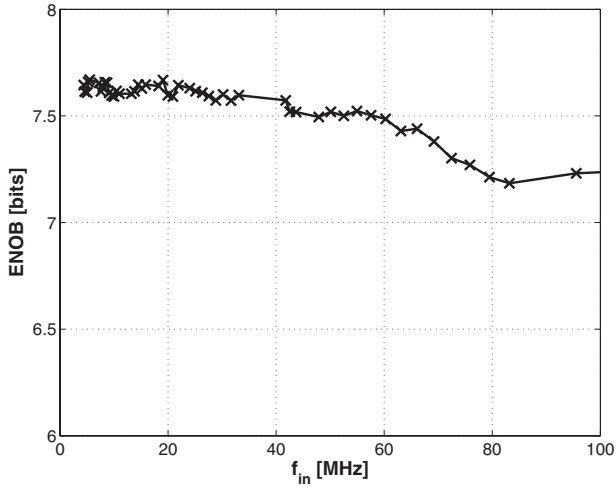


Fig. 17. Measured ADC performance vs f_{in} at f_{clk} = 125MS/s.

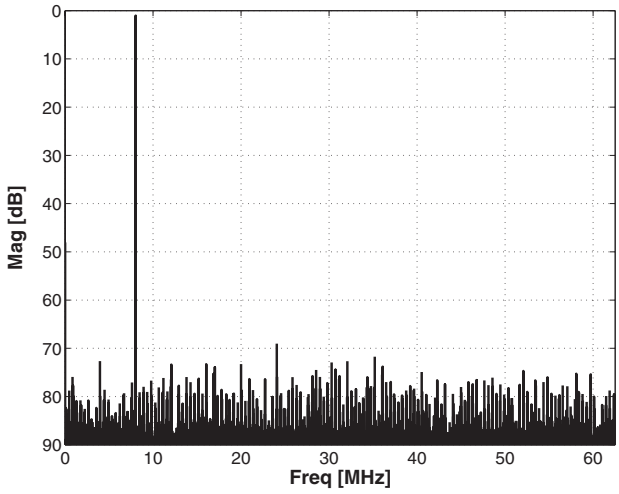


Fig. 18. Measured output spectrum of the 8b ADC.

Table 1. Performance comparison for 8b converters.

Ref.	Technology	area [mm^2]	F [pJ/conv]
[10]	0.6μm	1.8	88
[7]	0.5μm	1.7	794
[2]	0.5μm	10	7934
[28]	0.35μm	0.8	529
[11]	0.35μm	0.9	30
[4]	0.18μm	1.8	2.5
This work.	0.13μm	0.09	0.54

The total power consumption P equals 21mW, including T/H (6mW) and clock generator, but excluding the digital output buffers. The power consumption is approximately equally divided between the T/H, the clock generator and the ADC. The energy per conversion step is often used as a figure of merit (FOM) F. It is defined as $F = P/(2^{\mathrm{ENOB}} \times 2\mathrm{ERBW})$, and, for this design, equals 0.54pJ/conversion at 125MS/s. For comparison, Table 1 lists the area and FOM of this design together with a number of recently published 8b Nyquist CMOS ADCs running at comparable sampling frequencies. This shows that this design achieves best results for both area and FOM.

Key results of 8b, 10b and 12b ADC designs are listed in Table 2.

Table 2. Performance summary of 8b, 10b and 12b designs.

Resolution	8b	10b	12b
Conversion rate	150MSample/s	150MSample/s	75MSample/s
Process	0.13μm CMOS	0.13μm CMOS	0.13μm CMOS
Power supplies	1.2V / 2.5V	1.2V / 2.5V	1.2V / 2.5V
ADC Power	15mW	90mW	135mW
T/H Power	6mW	30mW	90mW
ENOB	7.6b	9.2b	11b
SNR	7.7b	9.4b	11.2b
THD	9.7b	10.5b	12.0b

8. Conclusion

The subranging ADC architecture uses simple differential-pair amplifiers, which makes it very suitable for low-voltage operation, required in state-of-the-art CMOS processes. Several techniques, such as active interpolation, pipelining, auto-zero offset compensation and distributed capacitive averaging, are applied to obtain 8b-12b subranging ADCs.

References

[1] Y. Yoshii, K. Asano, M. Nakamura, and C. Yamada. An 8 bit, 100 MS/s flash ADC. *IEEE Jour. of Solid-State Circ.*, 19(6):842–846, December 1984.

[2] J. Ming and S.H. Lewis. An 8-bit 80-Msample/s pipelined analog-to-digital converter with background calibration. *IEEE Jour. of Solid-State Circ.*, 36(10):1489–1497, October 2001.

[3] K. Poulton, R. Neff, N. Setterberg, B. Wuppermann, T. Kopley, R. Jewett, J. Pernillo, C. Tan, and A. Montijo. A 20GSample/s 8b ADC with 1MB memory in 0.18μm CMOS. In *ISSCC Dig. of Tech. Papers*, pages 318–319, 2003.

[4] S. Limotyrakis, S.D. Kulchycki, D. Su, and B.A. Wooley. A 150MS/s 8b 71mW time-interleaved ADC in 0.18μm CMOS. In *ISSCC Dig. of Tech. Papers*, pages 258–259, 2004.

[5] A.G.W. Venes and R.J. van de Plassche. An 80-MHz, 80-mW, 8-b CMOS folding A/D converter with distributed track-and-hold preprocessing. *IEEE Jour. of Solid-State Circ.*, 31(12):1846–1853, December 1996.

[6] M.P. Flynn and D.J. Allstot. CMOS folding A/D converters with current-mode interpolation. *IEEE Jour. of Solid-State Circ.*, 31(9):1248–1257, September 1996.

[7] M.-J. Choe, B.-S. Song, and K. Bacrania. An 8-b 100-MSample/s CMOS pipelined folding ADC. *IEEE Jour. of Solid-State Circ.*, 36(2):184–194, February 2001.

[8] G. Geelen and E. Paulus. An 8b 600MS/s 200mW CMOS folding A/D converter using an amplifier preset technique. In *ISSCC Dig. of Tech. Papers*, pages 254–255, 2004.

[9] R. Taft, C. Menkus, M.R. Tursi, O. Hidri, and V. Pons. A 1.8V 1.6GS/s 8b self-calibrating folding ADC with 7.26 ENOB at Nyquist frequency. In *ISSCC Dig. of Tech. Papers*, pages 252–253, 2004.

[10] Y.-T. Wang and B. Razavi. An 8-bit 150-MHz CMOS A/D converter. *IEEE Jour. of Solid-State Circ.*, 35(3):308–317, March 2000.

[11] R.C. Taft and M.R. Tursi. A 100-MS/s 8-b CMOS subranging ADC with sustained parametric performance from 3.8 V down to 2.2 V. *IEEE Jour. of Solid-State Circ.*, 36(3):331–338, March 2001.

[12] J. Mulder, C.M Ward, C.-H. Lin, D. Kruse, J.R. Westra, M.L. Lugthart, E. Arslan, R. van de Plassche, K. Bult, and F.M.L. van der Goes. A 21mW 8b 125MS/s ADC occupying 0.09mm^2 in 0.13μm CMOS. In *ISSCC Dig. of Tech. Papers*, pages 260–261, 2004.

[13] T. Matsuura, H. Kojima, E. Imaizumi, K. Usui, and S. Ueda. An 8-b 50-MHz 225-mW submicron CMOS ADC using saturation eliminated comparators. In *Proc. Custom Integrated Circ. Conf.*, pages 641–644, 1990.

[14] K. Kusumoto, A. Matsuzawa, and K. Murata. A 10-b 20-MHz 30-mW pipelined interpolating CMOS ADC. *IEEE Jour. of Solid-State Circ.*, 28(12):1200–1206, December 1993.

[15] B.P. Brandt and J. Lutsky. A 75-mW, 10-b, 20-MSPS CMOS subranging ADC with 9.5 effective bits at Nyquist. *IEEE Jour. of Solid-State Circ.*, 34(12):1788–1795, December 1999.

[16] H. van der Ploeg, G. Hoogzaad, H.A.H. Termeer, M. Vertregt, and R.L.J. Roovers. A 2.5-V 12-b 54-Msample/s 0.25-μm CMOS ADC in 1-mm^2 with mixed-signal chopping and calibration. *IEEE Jour. of Solid-State Circ.*, 36(12):1859–1867, December 2001.

[17] Y. Nishida, K. Sone, K. Amano, S. Matsuba, and A. Yukawa. An 8-bit 200MS/s 500mW BiCMOS ADC. In *Proc. Custom Integrated Circ. Conf.*, pages 207–210, 1995.

[18] M. Sugawara, Y. Yoshida, M. Mitsuishi, S. Nakamura, S. Nakaigawa, Y. Kunisaki, and H. Suzuki. A 2.5V 100MS/s 8bit ADC using pre-linearization input buffer and level up DAC/subtractor. In *VLSI Circ. Symp.*, pages 170–173, 1998.

[19] M. Clara, A. Wiesbauer, and F. Kuttner. A 1.8V fully embedded 10b 160MS/s two-step ADC in 0.18μm CMOS. In *Proc. Custom Integrated Circ. Conf.*, pages 437–440, 2002.

[20] T. Sigenobu, M. Ito, and T. Miki. An 8-bit 30MS/s 18mW ADC with 1.8V single power supply. In *VLSI Circ. Symp.*, pages 209–210, 2001.

[21] K. Nagaraj, F. Chen, T. Le, and T.R. Viswanathan. Efficient 6-bit A/D converter using a 1-bit folding front end. *IEEE Jour. of Solid-State Circ.*, 34(8):1056–1062, August 1999.

[22] K. Kattmann and J. Barrow. A technique for reducing differential non-linearity errors in flash A/D converters. In *ISSCC Dig. of Tech. Papers*, pages 170–171, 1991.

[23] K. Bult and A. Buchwald. An embedded 240-mW 10-b 50-MS/s CMOS ADC in 1-mm^2. *IEEE Jour. of Solid-State Circ.*, 32(12):1887–1895, December 1997.

[24] G. Hoogzaad and R. Roovers. A 65-mW, 10-bit, 40-Msample/s BiCMOS Nyquist ADC in 0.8 mm^2. *IEEE Jour. of Solid-State Circ.*, 34(12):1796–1802, December 1999.

[25] M. Choi and A. Abidi. A 6-b 1.3-Gsample/s A/D converter in 0.35-μm CMOS. *IEEE Jour. of Solid-State Circ.*, 36(12):1847–1858, December 2001.

[26] P. Vorenkamp and R. Roovers. A 12-b, 60-MSample/s cascaded folding and interpolating ADC. *IEEE Jour. of Solid-State Circ.*, 32(12):1876–1886, December 1997.

[27] H. van der Ploeg and R. Remmers. A 3.3-V, 10-b, 250MSample/s two-step ADC in 0.35-μm CMOS. *IEEE Jour. of Solid-State Circ.*, 34(12):1803–1811, December 1999.

[28] K. Yoon, J. Lee, D.-K. Jeong, and W. Kim. An 8-bit 125MS/s CMOS folding ADC for Gigabit Ethernet LSI. In *VLSI Circ. Symp.*, pages 212–213, 2000.

ADVANCES IN HIGH-SPEED ADC ARCHITECTURES USING OFFSET CALIBRATION

Robert Taft, Chris Menkus, Maria Rosaria Tursi, Ols Hidri,
Valerie Pons
National Semiconductor GmbH, Fürstenfeldbruck, Germany

Abstract

This paper describes how offset calibration can be used to enhance the speed-resolution performance of ADCs while maintaining low levels of power dissipation. After selecting a high-speed ADC topology, we will justify the need for offset calibration and then present a brief overview of different ADC calibration topologies, indicating their relative merits and drawbacks. We will then describe in more detail how one-time foreground calibration was used to achieve a 0.18 μm CMOS, 1.8V, 1.6GS/s, 8b folding-interpolating ADC with 7.26 ENOB at Nyquist, which only dissipates 774 mW.

1. Introduction

The continuous improvement in high-speed CMOS Analog-to-Digital Conver-ters (ADCs), as measured by their increased resolution and speed for a given power budget, has been driven by process improvements, primarily by geometry scaling. However, this trend of CMOS geometry scaling enabling the next level of ADC performance has greatly diminished, starting at about the 0.35 μm gate-length generation, primarily due to device mismatch improving only marginally coupled with the reduction of the power supply voltage with each successive deep-submicron technology. ADC performance advances will therefore come increasingly from design innovations, especially those which address the effects of offset stemming from device mismatch [1].

One of the most promising circuit techniques increasingly implemented in the past decade is that of offset calibration, which directly counteracts device mismatch. In this paper, after selecting a high-speed ADC topology, we will justify the need for offset calibration, and give a brief overview of different ADC calibration topologies. We will then describe in more detail why and how one-time foreground calibration was chosen to achieve a 0.18 μm CMOS 1.8V, 1.6GS/s, 8b folding ADC with 7.26 ENOB at Nyquist.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 73–96.

2. High-speed ADC Architectures

CMOS ADC converters resolve an input voltage by either parallel processing as in a flash-type converter, see Fig. 1, or sequential processing as in a successive-approximation-type converter. Parallel processing has the advantage of conversion speed, sequential processing the advantage of compactness and low power. Other types of converters, such as subranging, multi-step, or pipeline, are hybrids, performing limited parallel processing sequentially. For high-speed ADCs, we limited ourselves to those which have no decision feed-back loops.

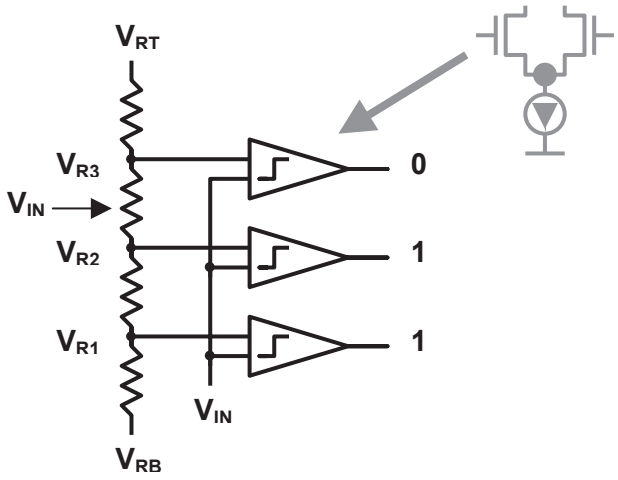


Fig. 1. Schematic for a 2-bit flash converter, which uses a resistor ladder to subdivide end-point reference voltages V_{RT} and V_{RB}. The comparators' output thermometer code "0, 1, 1," corresponds to a V_{IN} between voltages V_{R3} and V_{R2}. Note that there is no over-range indication; the binary encoded output is "10".

Figure 2 gives an example of a decision feed-back loop in a subranging flash converter [2]. By using the 2-bit flash ADC of Fig. 1 for the two coarse bits in combination with the two fine bits obtained from the subranged conversion of Fig. 2, we obtain a 4-bit converter with only 6 comparators, compared to 15 comparators required for a single-stage flash converter. In higher resolution ADCs, the number of comparators and amount of power saved due to subranging is substantially greater. However, this architecture is slower—not because it requires two steps—the coarse and fine conversions could be done in parallel in a staggered fashion. The problem is that the subranged conversion needs to decide to what part of the reference ladder to switch its comparators. This information is provided by the coarse comparison, and thus a decision feedback-loop exists between the coarse

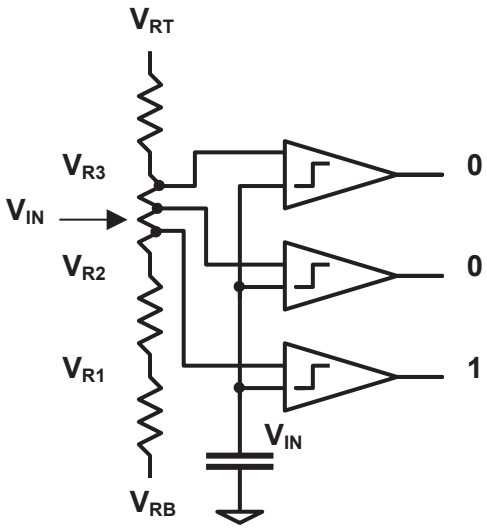


Fig. 2. Schematic for the second comparison of a 4-bit subranging converter, showing the evaluation of the two LSBs. The two MSBs "10" were obtained as shown in Fig. 1. The MSBs are not subtracted from V_{IN}. Rather, the input voltage V_{IN} is held, and the comparator reference inputs are switched onto the appropriate segment of the reference ladder. The binary representation of the LSBs is "01", which together with the MSBs "10" forms the 4-bit digital output code "1001".

and fine conversion steps of the converter. We believe that any high-speed ADC architecture which relies on decision feed-back loops to resolve an input signal is disadvantaged compared to those that can run open loop, especially when Code-Error-Rates and meta-stability are considered. Therefore, subranging, multi-step, pipeline and successive-approximation ADCs were excluded for our program.

Although the flash architecture has been used in the highest speed ADCs, it has the severe disadvantage that it requires 2^N-1 comparators for N bits of resolution, which results in a substantial area and power penalty starting at the 8-bit level. Furthermore, for 8 bits and higher resolutions, encoding the outputs of so many comparators presents an additional speed limitation. Fortunately, flash ADCs are not the only type without decision feed-back loops. An ingenious architecture proposed by van de Plassche and van der Grift [3] is able to subrange without prior knowledge of the coarse conversion. Thus the coarse and fine conversions are done in parallel, without decision feed-back loops, rather than sequentially. This type of ADC is now referred to as "folding," and is nearly identical in architecture to flash, but in the preamplifier stages between the input voltage and comparators, outputs are joined together in such a way that a single comparator can be used to match an input to a number (k) of distinct reference voltages. The number of comparators required is only

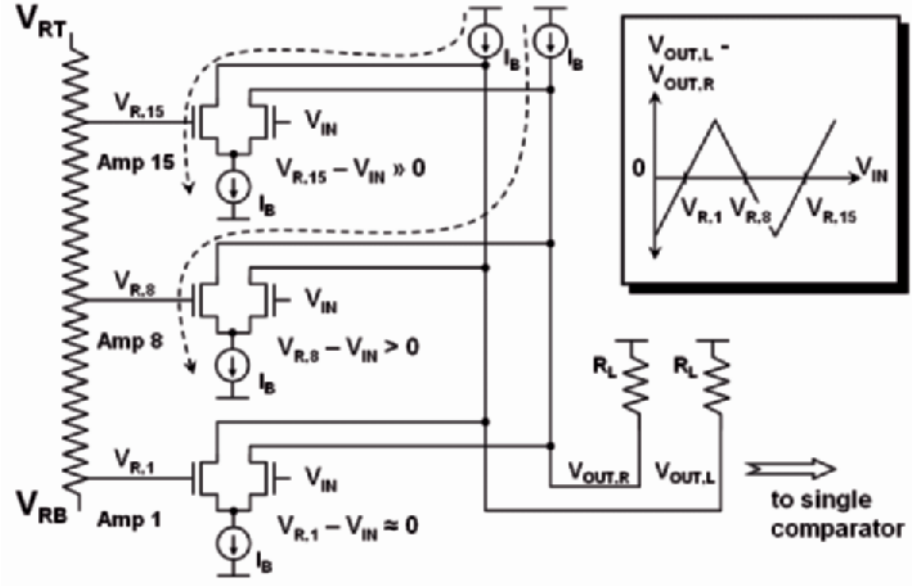


Fig. 3. Example of folding the outputs of an array of flash preamplifiers by 3. With V_{IN} near $V_{R.1}$, amps 8 and 15 are fully saturated, their tail currents supplied by the cancellation currents on the "folding bus," and they do not affect the output voltage.

2^N / k + k - 2, where k represents the order of folding. The additional k-2 included in this sum arises from the coarse comparators required to resolve which fold the input signal was in, as described later. Because of this large reduction in comparators, the folding architecture was chosen for this ADC. The principle of operation for "folding-by-3" is shown in Fig. 3. Here the number of preamplifiers of a flash converter is a multiple of three. These amplifiers are grouped such that the difference in the reference voltage of the top and middle amplifier and middle and bottom amplifier is equal, and the same for all of the groups of three. Rather than each preamplifier driving one comparator, each group of three preamplifiers only drives one comparator, by connecting their outputs in parallel, but with the middle amplifier in reverse polarity. This reverse connection of the middle amplifier results in each of the "zero-crossings" of the 3 preamplifiers forcing a zero-crossing at the comparator, and thus being detectable. Folding has one drawback, it is more sensitive to device offsets than flash. Thus, ever since the elegant solution of folding migrated from bipolar (where it was originally conceived [3]) to CMOS technologies, it has been plagued by the higher mismatches of CMOS pairs.

3. Justification of ADC Offset Calibration

Due to the simplicity of the flash converter, it will be used to now justify offset calibration. The comparators of Fig. 1 are usually formed by a differential pair as part of a regenerative amplifier, whose inherent offset voltage, δV_T, with standard deviation $\sigma_{\delta VT}$ introduces an error in quantizing the input voltage. If preceding preamplifiers exist, then this same argument applies to the differential pair of the preamplifiers. The offset voltage δV_T is usually dominated by the MOS threshold variation of each input device, which is determined by the statistical fluctuation of dopant atoms in the depletion region of the transistor. The random position of the dopant atoms results from ion implantation and thermal annealing, and is given by a Poisson distribution. I.e., it is not a manufacturing issue that can be solved by process evolution, short of placing and keeping the dopant atoms deterministically in the presence of thermal cycles, a prohibitive thought. Keyes [4] first quantified the effect this variation had on the standard deviation of the threshold voltage, σ_{VT}, of square geometry transistors:

$$\sigma_{VT} = \frac{q}{2C_{OX}}\sqrt{\frac{\pi N_A}{2}}(W_{eff}L_{eff})^{-3/8}(x_D)^{1/4} \tag{1}$$

where q is the fundamental charge, C_{OX} the gate oxide capacitance per unit area, N_A the well dopant concentration in the depletion region, L_{eff} the effective channel length, W_{eff} the effective channel width, and x_D the MOSFET depletion depth. This equation correlated very successfully with the measured V_T fluctuations for near minimum NMOS devices in a 0.5 μm SRAM process [5].

More general models have been developed for arbitrary ratios of W_{eff} / L_{eff}, see [6,7]. These have a 1/2 power dependence on L_{eff} and W_{eff}. An especially useful form is given by Mizuno which replaces the depletion depth x_D with the surface potential ϕ_B [7]:

$$\sigma_{VT} = \frac{\sqrt{2}\, q^{3/4}\epsilon_S^{1/4}\phi_B^{1/4}}{\epsilon_{OX}} t_{OX} N_A^{1/4}(W_{eff}L_{eff})^{-1/2} \tag{2}$$

where ε_S is the permittivity of silicon, ε_{OX} the permittivity of the gate oxide, t_{OX} the thickness of the gate oxide, and all other variables as defined for the previous equation. Please note that the derivation given in the *IEEE Trans. On Elec. Dev.* by Mizuno is in error by a factor of 2, as is his final equation 11. A similar but simpler derivation given by Watt in [6] yields the same (corrected) result.

The surface potential is a very weak function of doping, with $\phi_B \sim 0.40$ V at 300K for well dopant concentrations of 10^{17} cm^{-3}. Since ϕ_B varies less than 4% for an order of magnitude change in N_A, we can treat it as a constant.

Note that equation 2 describes the 1-sigma spread of V_T. When V_T mismatch is measured, it is the difference between two identical distributions, resulting in a standard deviation $\sigma_{\delta VT}$ which is root-2 times as large, assuming no correlation. I.e.:

$$\sigma_{\delta VT} = \sqrt{2}\,\sigma_{VT} \tag{3}$$

Evaluating all constants while combining equations 2 and 3 results in the expression:

$$\sigma_{\delta VT} = 3.72 \times 10^{-9}\, t_{OX} N_A^{1/4} (W_{eff} L_{eff})^{-1/2} \ \left(\mathrm{V}\ \frac{\mathrm{cm}^{3/4}\,\mu\mathrm{m}}{\text{Å}}\right) \tag{4}$$

Equation 4 shows that for same area transistors, process scaling will reduce $\sigma_{\delta VT}$ only if the oxide thickness reduces faster than the fourth-root of the well doping. For constant-supply scaling, this was not the case, and $\sigma_{\delta VT}$ remained relatively flat. However, the product $\sigma_{\delta VT} C_{OX}$ grew rapidly, so that for the same offset, a larger gate capacitance and therefore power dissipation resulted. For process scaling including supply scaling which started below the 0.5 μm generation, $\sigma_{\delta VT}$ improves marginally with process technology, since the well doping does not need to increase as dramatically with both supply and channel length reduced. However, this small improvement is offset by the reduced signal swing resulting from the reduced headroom the scaled supply voltage creates. Again, the $\sigma_{\delta VT} C_{OX}$ product still worsened.

One way to take advantage of small-geometry processes to obtain both high resolution and high speed ADC converters is to increase the devices size for the offset-sensitive circuits, since this reduces the $\sigma_{\delta VT}$ inversely with square-root of area, see equation 4. However, the increased gate capacitance results in an unacceptable power penalty for high-speed ADCs. Another approach is auto-zeroing, which although successful on intermediate speed converters, is not well suited for high-speed ADCs, due to the introduction of a phase for storing the offset, not to mention the need for auto-offset-storing elements such as capacitors, which bring extra devices and parasitics into the circuit.

Another approach which indirectly addresses offsets is that of interleaving. If N ADCs sample the input in turn, then each ADC has N times longer to resolve the input. The increased effective conversion speed is countered by an N-fold increase in layout area, since each of the interleaved channels needs sufficiently large devices to have low offsets. However, this approach does allow auto-zeroing ADCs, which overcome

offsets, to be used. In the case of interleaving auto-zeroing ADCs, though, each channel must have sufficiently large capacitors to maintain the commensurate device noise for the target SNR. A further complication is driving multiple ADCs from a single source. Not only does the loading of the source need to be carefully controlled, but so does the timing. We do use interleaving-by-2 (N=2) for the ADC presented here, since this gives a speed benefit without too large an area penalty or too much complexity in timing. However, since we don't use auto-zeroing to obtain the highest converter speed, we still needed to address device mismatch. The most direct way to tackle the problem is to calibrate out the offsets, in a way that does not impede the analog circuits in any way, and does not add significantly to the power or area budget.

4. ADC Calibration Topologies

Various offset calibration topologies have been used, primarily on slower, higher resolution pipeline ADCs. Both foreground calibration during which the converter cannot be used [8] and background calibration during which the ADC can continue resolving the input [9] have been demonstrated. Other background techniques are applicable beyond their pipeline ADC implementation, and include introducing a small uncorrelated signal on the input, and driving the error term to zero [10]; having one redundant interleaved ADC, which can be swapped into the interleaved array, to allow each ADC channel to be calibrated [11]; and using a queue-based cascaded Track&Hold, which runs more slowly than the ADC, to allow periodic introduction of calibration vectors [12]. These latter two techniques are not suitable for this high-speed ADC.

One other technique which has gained some acceptance is post-ADC linearity and timing improvements using DSP filtering/algorithms on the digital data. The coefficients of the digital filter are typically set for a specific Nyquist zone during calibration at final test. This approach is outside the scope of this work, but clearly has the draw-back that some types of ADC errors can not be corrected in the digital domain, and that the corrections require a continuously running DSP with its associated significant power dissipation.

We use a one-time foreground offset cancellation technique, similar to the one demonstrated on a high-speed flash ADC converter [13]. However, in this work, the calibration is performed fully on chip, without the need for externally applied calibration vectors or even for the removal of the input signal. We believe this work documents the first time calibration has been combined with folding, and further demonstrates how complementary these two approaches are.

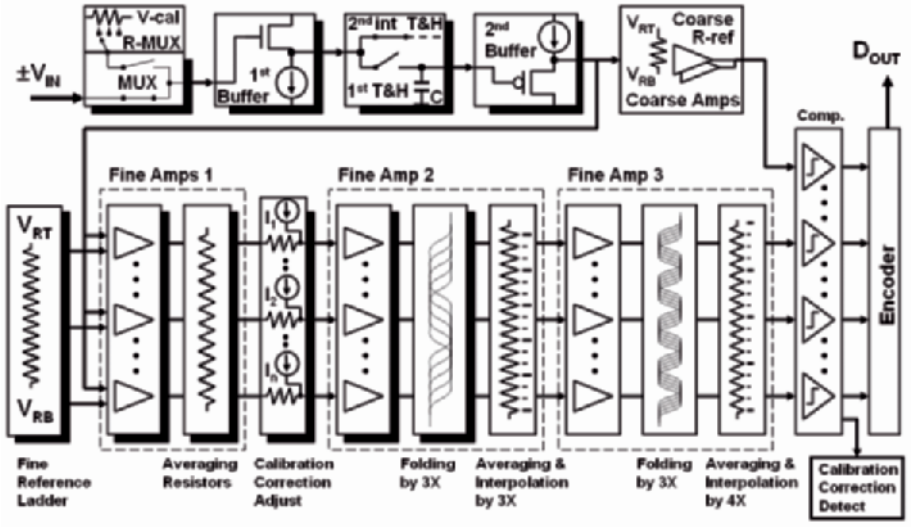


Fig. 4. Block diagram of this calibrated folding-interpolating ADC, showing only the 1st interleaved channel. Offsets and nonlinearities arising from elements inside the shadowed blocks are fully compensated for by the calibration correction applied between fine amps 1 and 2. Fine Amps 1, 2, and 3 are each made of an array of 27 amplifiers.

5. Self-Calibrating Folding-Interpolating ADC Architecture

A complete overview of the ADC from MUX'ed calibration vectors at the input buffer to the encoder outputs is given in Fig. 4. After the 1st buffer which follows the input MUX, the Track&Hold, distributed fine Amps 1, 2, and 3, comparators, coarse comparators, encoder, and LVDS outputs are all interleaved by two, running at 800 MHz for this nominally 1.6 GS/s differential ADC. In the absence of interleaving, this is the highest speed CMOS 8-bit single-channel ADC published to date.

Note the large number of front-end blocks whose linearity or offsets are calibrated by a single correction point between preamplifiers 1 and 2. They are indicated by shadowed boxes in Fig. 4. We now describe in detail each of the major circuit blocks sequentially.

5.1. Input MUX

By applying the calibration vectors at the very beginning of the signal chain, the number of functional blocks whose offsets can be corrected are maximized. However, this requires an on-chip analog input MUX with very

low distortion at GHz input frequencies, to maintain good Nyquist performance. This was achieved using a constant V_{GST} NMOS pass-gate, shown in Fig. 5. The gate-to-source voltage of pass-gate M0 is set by the combined on-voltage of M1 and M2. Since the body effects of transistors M0 and M1 nearly cancel, and M2 with its well tied to its source has no body effect, this NMOS pass-gate posses a constant on-resistance and therefore very low distortion. Furthermore, the constant on-resistance is achieved with very low power, since only a small bias current is required to maintain the bias voltage. The transient response is assured by capacitive coupling between the gate and source of M0. Unlike the more conventional constant-V_{GS} sampling switch used in the Track&Hold, this pass-gate only switches at the start and end of calibration.

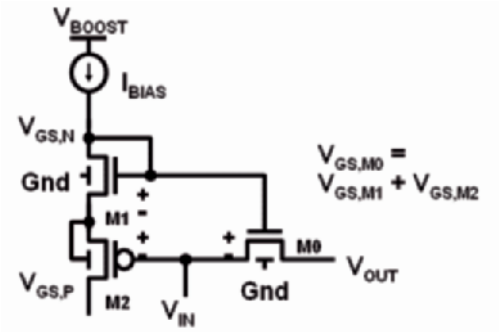


Fig. 5. The high-bandwidth, low-distortion input-MUX gate is low power and has nearly constant-V_{GST} (body effects of M1 and M0 nearly cancel, M2 has no body effect).

5.2. Interleaved Track&Hold

Because of the very high conversion speed, an open-loop topology was chosen for the Track&Hold. The pseudo-differential architecture uses two separate but identical circuits for the positive and negative part of the differential input signal, one of which is shown in Fig. 6. As depicted, the Track&Hold employs interleaving. After a shared first buffer, the signal splits into two interleaved paths running at half the converter speed. Each interleaved path is made up of a sampling switch, S_S, the sampling cap, C_S, and a second buffer. The first buffer reduces input signal loading and kickback noise, while the second buffer drives the first array of preamplifiers, which follows.

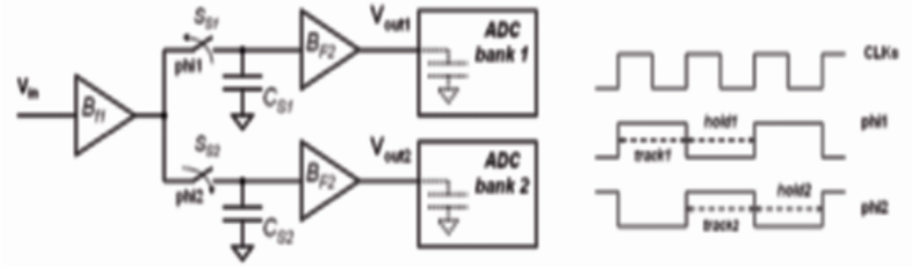


Fig. 6. Interleaved Track&Hold and timing diagram.

The most important advantage of interleaving-by-2 is that both the track time and the hold time of the Track&Hold are more than doubled (these times are *more than doubled* because the effective period doubles, but the switching time between track and hold does not—an appreciable effect at these conversion speeds!). This greatly simplifies the design of the sampling switch and second buffer. However, interleaving can greatly degrade the SNR because of mismatches between the two interleaved paths. Timing mismatches, which become increasingly important at higher input frequencies, add to the SNR degradation due to gain and offset mismatches. Each of these sources of potential error was successfully addressed in this design. Gain and offset for both interleaved channels are matched by sharing the same input buffer and including the Track&Hold in the calibration path. The sampling aperture offset between the two interleaved channels, t_A, was minimized using a master sampling clock, *samp*, as shown in Fig. 7, and previously used in [14]. As seen, the sampling instant is determined by the falling edge of the full speed clock signal *samp*, which is common to the two interleaved sampling switches. Local half speed signals, $track_1$ and $track_2$, provide the interleaving function between the two channels, but are not timing critical. The resulting aperture offset t_A was extracted as < 0.35 ps (peak-peak) at F_S = 1 GS/s and F_{IN} = 1.5 GHz for all nine parts examined, using the F_S / 2 - F_{IN} spur (P_{tA}) with equation:

$$P_{tA}\,(\mathrm{dB}) = 20 \log10\,(2\pi\, F_{IN}\, t_{A,RMS}) \tag{5}$$

where $t_A = t_{A,RMS}$ x 2, see e.g. lower spectrum in Fig. 13. Exacting layout using both step-and-repeat methodology, to reduce sensitivity to mask misalignment, and cell mirroring, to allow symmetric clock distributions, was made possible by creating mirror-invariant layout for the Track&Hold. As a result, the t_A dependence on operating conditions showed that t_A was not limited by layout, but by F_S / 2 noise (e.g , the LVDS drivers run at F_S / 2), likely fed into the clocking through the power bussing.

Also shown, as part of Fig. 7, is a simplified schematic of the constant-V_{GS} switch used for sampling. The track-phase boost capacitor, previously charged to V_{ON}, is connected between gate and source of the sampling

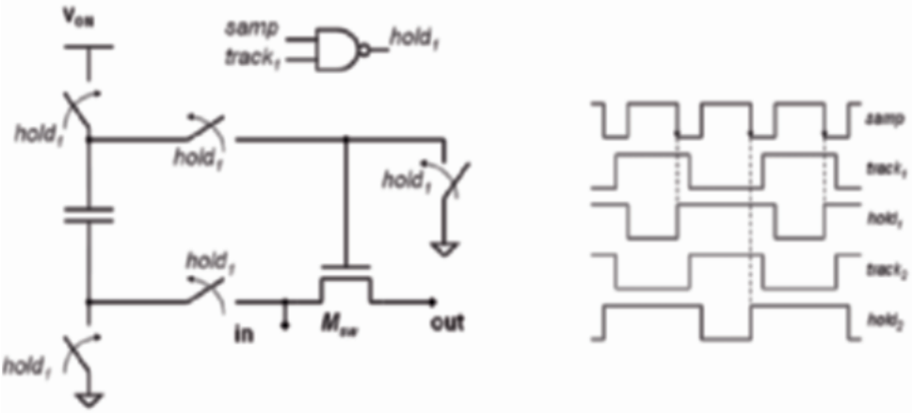


Fig. 7. Simplified constant-V_{GS} switch for the 1st interleaved channel. Note that the sampling switch for the 2nd interleaved channel is identical but driven by $hold_2$, generated by $track_2$ and samp. The timing diagram shows the $hold_1$ and $hold_2$ rising edges determined by the samp falling edge.

switch, thus acting as a battery and providing an almost constant gate-to-source voltage during the track phase. This pseudo-differential architecture would achieve the required 8-bit linearity despite the switch not being compensated for the body effect (as the MUX pass-gate is), even without being placed inside the calibration path.

5.3. Preamplifiers

A combination of best design practices was used to reduce power as much as possible. For example, to amplify a signal by a factor of 27, it is more efficient to cascade three amplifiers, each with a gain of 3, than to use just one single, high-gain amplifier.

The preamplifiers following the Track&Hold serve three purposes: 1) To amplify the small voltage difference between the signal and reference to be much larger than the input-referred comparator offset, 2) To generate new crossing points through "interpolation," which reduces the number of reference voltages the input signal needs to be compared to, and 3) To "fold" the crossing points onto a reduced number of shared wires, which reduces the number of required comparators.

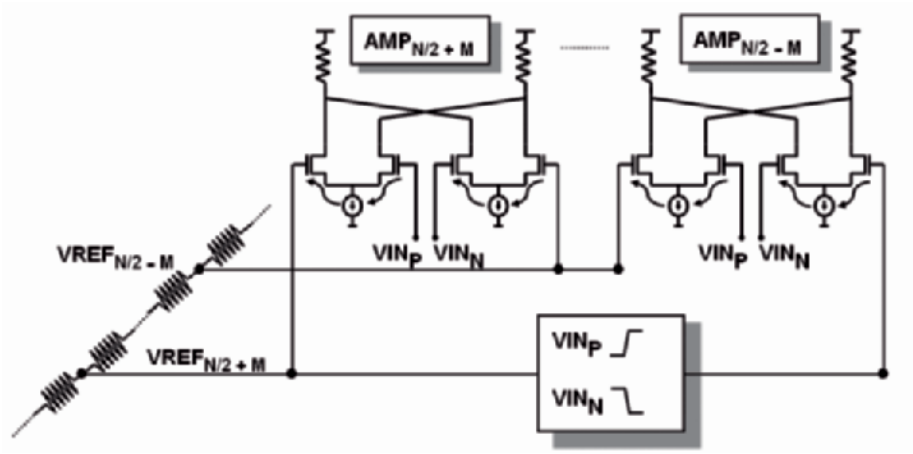


Fig. 8. The first preamplifiers are differential differencing amplifiers. Their connection to the flash-like reference ladder is indicated.

The first stage preamplifiers convert the pseudo-differential (ground-referenced) Track&Hold output into a true differential signal for better supply rejection and processing ease. This is accomplished with differential differencing amplifiers, as shown in Fig. 8. The differential reference voltages are generated from a scaled version of a precise bandgap reference, mirrored across a single resistor ladder, and tapped at evenly spaced intervals. Important is that the reference signals are low enough impedance to settle transient spikes within one comparison period. Using a single reference ladder has the advantage that kickback from the differential pairs is cancelled to the first order, since each tap point is coupled to both polarities of the input signal through the gate-to-source capacitances of two amplifiers.

5.4. Interpolation, Folding, and Averaging

We briefly describe the well-known circuit techniques of interpolation, folding, and averaging, which are used extensively in the preamplification stages of the fine channel [3,15,16].

Interpolation allows the generation of intermediate voltages, so that the input voltage does not need to be compared to 255 distinct reference voltages at the front-end of a flash-like converter. Each input to reference voltage comparison at the front end would require its own amplifier, consuming an unacceptable amount of power and area. “Virtual” crossing points, which aren’t the result of the input crossing a physical reference voltage, are created through resistor interpolation at subsequent stages of amplification. In this

design, the second and third preamplifier stages generate three and four crossing points, respectively, for every one "powered" crossing point.

Interpolation only reduces the required number of preamplifiers, but the comparators still consume significant power and area. Likewise, a large number of comparator outputs leads to increasingly complex digital encoding logic. In this design, a total folding order of nine is achieved by cascading two stages of folding-by-three, as in [16]. The existing second and third preamplifier arrays are available for this purpose. In each of these arrays, 9 X 3 differential pairs are joined at the load into nine amplifier triplets, each sharing a common load, whose outputs are interpolated and made available to the following stage. The number of comparators is thus reduced by nine.

Minimizing the folding order of one stage to 3 and using cascade folding has the advantage of reducing parasitic capacitance at the folding nodes, since fewer amplifier outputs are tied to the same bus. Nonetheless, the large physical separation of the amplifiers tied to one folding bus results in a significant parasitic capacitance. To mitigate the effects of this capacitance, we use a folded cascode output stage as shown in Fig. 9 (folded cascode is used in contrast to telescopic cascode, and is unrelated to the term folding used to describe this ADC architecture). Although a folded cascode consumes extra power, in this case, the extra power is negligible, since it is shared between three amplifier differential pairs. The low source impedance shields the capacitance of the folding bus from the high-impedance output, and reduces the voltage swing at the differential pair drains, thus reducing kickback to the previous stage.

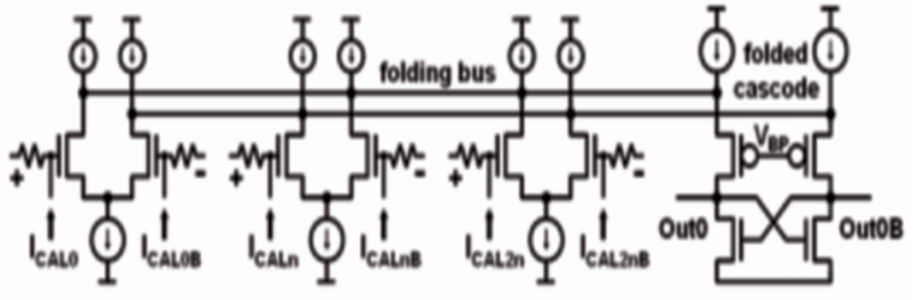


Fig. 9. Preamp2 showing three amplifiers grouped onto a folding bus. The folded-cascode load mitigates the effect of the large load at the amplifiers' outputs.

Some information is, of course, lost in the 9-to-1 mapping of folding, and so seven "coarse" comparators are used to locate which of the nine folds the input signal was in. Since the coarse comparators work in parallel, there are no decision feedback loops, and high-speed throughput is possible.

Folding adds extra offset because the two saturated pairs not involved in a decision need to have their tail currents exactly cancelled or matched. Therefore, averaging is used to reduce the effects of these offsets, as described in [15]. The principle is to statistically reduce the influence of one amplifier's offset by making each amplifier output the superposition of its own and its neighbor's outputs, assuming that the neighbor's offsets are uncorrelated. Effective device noise, while not dominant in this design, is also reduced in the same way. The more amplifiers that contribute to a single output, the greater the improvement. Relatively low-valued resistors join neighboring amplifier outputs and, for preamplifiers 2 and 3, make use of the same resistors already used for interpolation.

Each of the three preamplifier arrays incorporates averaging. However, its benefit is strongest in the first amplifier stage, where more neighboring amplifiers remain unsaturated, thus increasing the number of amplifiers, which pull the "active" amplifier's output towards the correct value. Later amplifier stages, although benefiting less from averaging, benefit more from the input-referred offset being divided down by the gain of the earlier amplifier stages.

5.5. Calibration

After a start-up delay following power-on, twenty-seven calibration vectors from an on-chip linear reference resistor are MUXed to the input buffer during a user-transparent 29 μs calibration cycle, during which time the analog input signal is automatically disconnected from the converter input. As in the first preamplifier reference ladder, the calibration vectors are generated from a scaled version of the bandgap reference, mirrored across a large-area resistor, and tapped at precise intervals. Offsets at each of the distributed and cascaded amplifier 1 and 2 pairs are corrected using a simple current DAC which adjusts the voltage across resistors placed in series between the output of each Amp 1 and the input of the following Amp 2 diff pair. The series resistors are necessary to localize the offset correction going into the less-averaged Amp 2, while still correcting for offsets in the more-averaged Amp 1.

The DAC is shown in block form, "Calibration Correction Adjust" in Fig. 4. The corrective current injection is shown as a single-ended signal in this block diagram, and correctly as a differential signal in Fig. 9. Although the original intent was to calibrate the offsets of "fine" preamps 1 and 2, this architecture calibrates most of the critical analog blocks as shown in Fig. 4, where they are indicated with shadow boxes. The linearity of the input MUX and 1st buffer are calibrated, as are the linearity, offset, and gain of the Track&Hold and 2nd buffer. This latter calibration correction was crucial in enabling an interleaved architecture, as previously described. Offsets in the fine reference ladder linearity and gain are calibrated, as are those in preamp

1, including its load and averaging resistors, and the series resistors introduced for applying the calibration correction itself. Finally, offsets in Amp 2, its load, its averaging, and its folding are calibrated. The first offsets

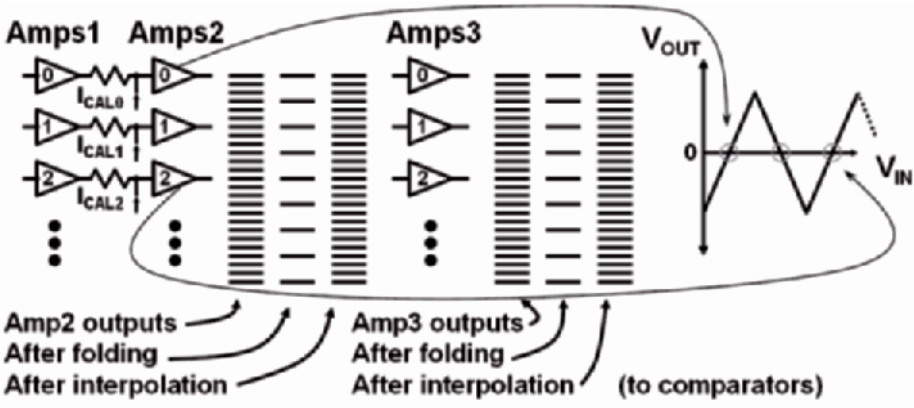


Fig. 10. During calibration, N_C comparator outputs determine the $I_{CAL1,2,}$.. adjustments for N_A amp1's, while V_{IN} is stepped with a linear on-chip resistor. The amplifier zero-crossings are folded onto comparator zero-crossings, resulting in a minimum $N_C = N_A / k$. Fine Amps 1,2, and 3 are each made of an array of 27 amplifiers.

not calibrated are in the interpolation at the output of Amp 2, since they don't map back uniquely to the calibration adjustments. The offsets of Amp 3 and later blocks, although not calibrated, have a small contribution to the input-referred offsets, due to the gain of the previous two stages of amplification.

Many of the blocks in the calibration path already have at least 8 bits of linearity. Their pre-calibration accuracy reduces the required dynamic calibration range, simplifying the digital calibrator circuitry and correction DACs. Furthermore, to ensure robustness, neither timing nor settling offsets are calibrated. This leads to a very stable calibration over clock frequency and time. I.e., a calibration performed with F_S = 100 MS/s is fully valid even if the clock frequency is changed to F_S = 1.8 GS/s. Further, the key parameter calibrated, INL, measured repeatedly after a single power-up calibration, showed no significant drift over 50 hours of operation.

During calibration, rather than interpreting the encoded output, a limited number of comparator outputs, N_C, are examined directly, as indicated by the block "Calibration Correction Detect" of Fig. 4, which contains latches and combinational logic to adjust the correction DACs. To simplify calibration of each of the distributed Amp 1 and 2 pairs, the number of these input amplifiers, N_A, must be an integer multiple of the order of folding, k (k = 3 x 3 = 9 in this ADC). This maps the folded zero-crossings of the ADC transfer function back onto the Amp 1 and 2 zero-crossings, see Fig. 10. Without

interpolation, this relation is mathematically necessarily true. However, by maintaining it even in the presence of interpolation, the number of comparators examined, N_C, are minimized. The transfer-function

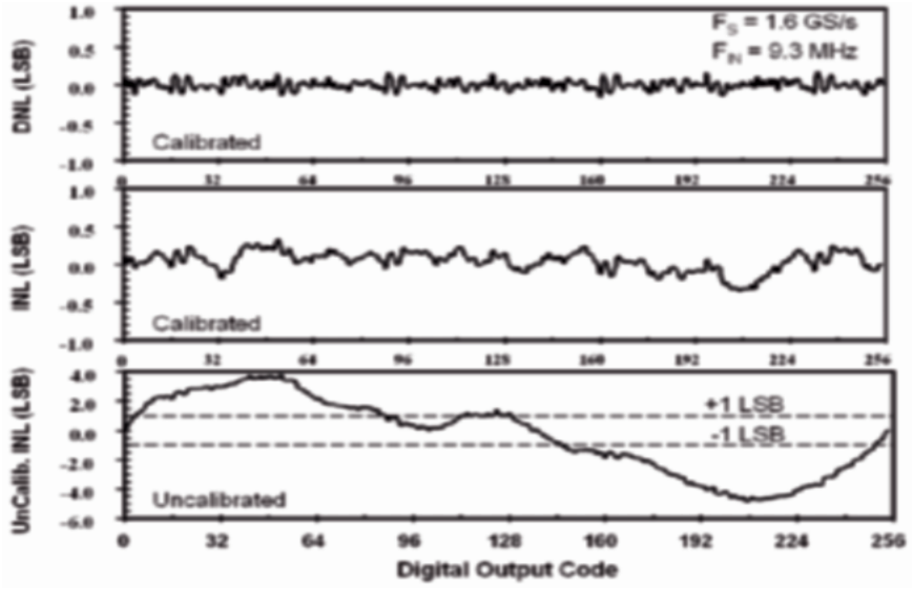


Fig. 11. Measured DNL and INL at 1.6 GS/s.

zero-crossings are detected using comparator outputs already present for normal operation, avoiding additional analog circuitry. The number of comparator outputs thus examined is $N_C = N_A / k$, assuming $N_A \geq k$. If $N_A = k$, then only a single comparator output is used, eliminating the inter-comparator offset error during calibration. In this design, $N_A = 27$, and $N_C = 3$.

6. Experimental Results

This calibrated-folding ADC is able to operate substantially above 1.6 GS/s. The test system limits the maximum conversion rate to 1.8 GS/s, where 7.5 ENOB are achieved for a 50 MHz input. However, since this is a Nyquist converter, we report all results here at 1.6 GS/s, at which conversion speed the Nyquist performance is not compromised.

Figure 11 shows the Differential NonLinearity (DNL < ±0.15 LSB) and Integral NonLinearity (INL < ±0.35 LSB) at 1.6 GS/s, extracted with a 9.3175 MHz sinewave input. In addition, INL is plotted for one interleaved

channel with calibration disabled. The uncalibrated INL curve is typically ≤ ±4 LSB, but varies widely in shape between parts. Fig. 12 plots ENOBs versus input frequency, maintaining nearly 7.0 ENOBs up to 1.1 GHz input. The near Nyquist performance for CMOS ADCs recently published at

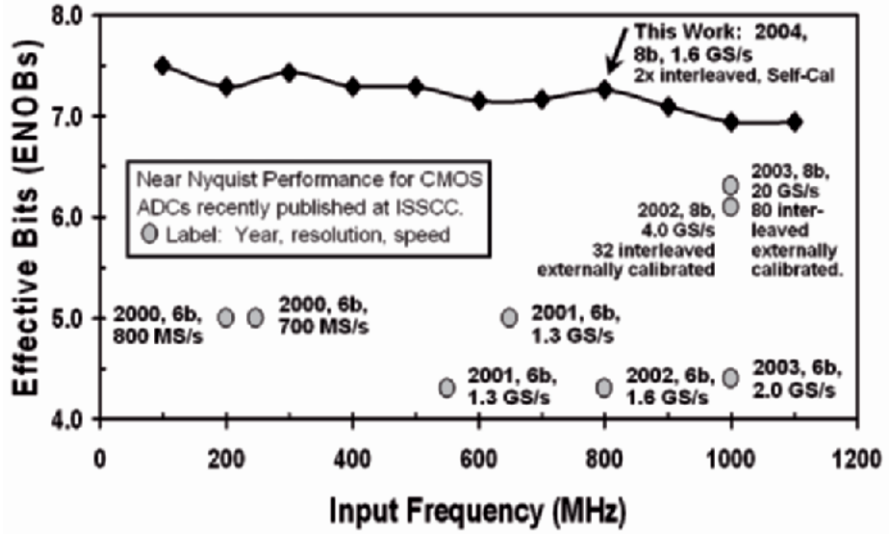


Fig. 12. Measured ENOBs versus F_{IN} at 1.6 GS/s. The near Nyquist performance for CMOS ADCs recently published at ISSCC are also indicated [14,17-23].

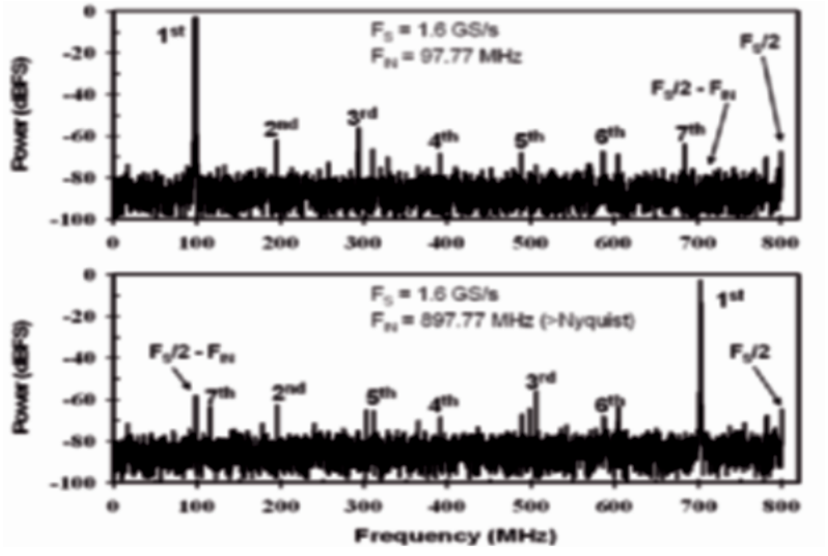


Fig. 13. Measured Spectrums at 1.6 GS/s for F_{IN} = 97.7 MHz and 897.77 MHz.

ISSCC are also indicated [14,17-23]. Note that performance approaching this work has only been achieved in CMOS technologies using heavily interleaved pipeline ADCs, which then results in extensive timing calibration between the individual banks. Furthermore, this previous work applied the calibration at the system level, with external calibration vectors required at the input.

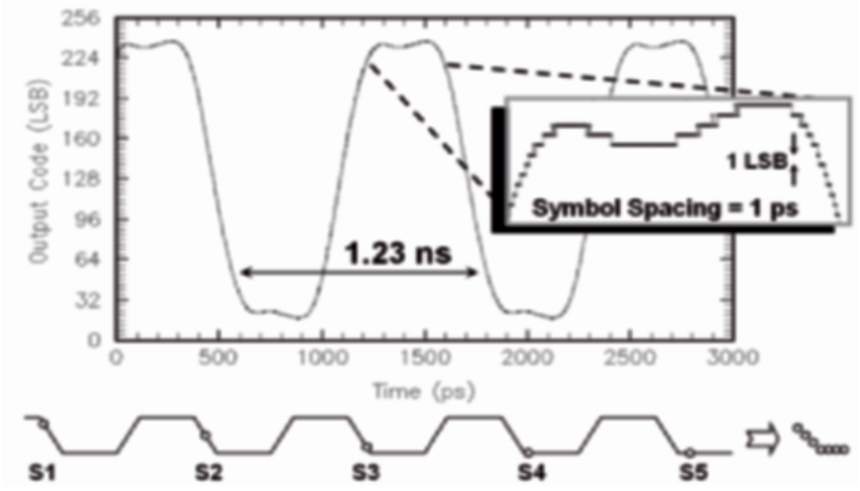


Fig. 14. Measured 800 MHz Beat-Frequency test at 1.8V. The converter samples a 810.7 MHz, 0.70 VDIFF Squarewave at 810 MS/s. The principle of the test is indicated below the plot.

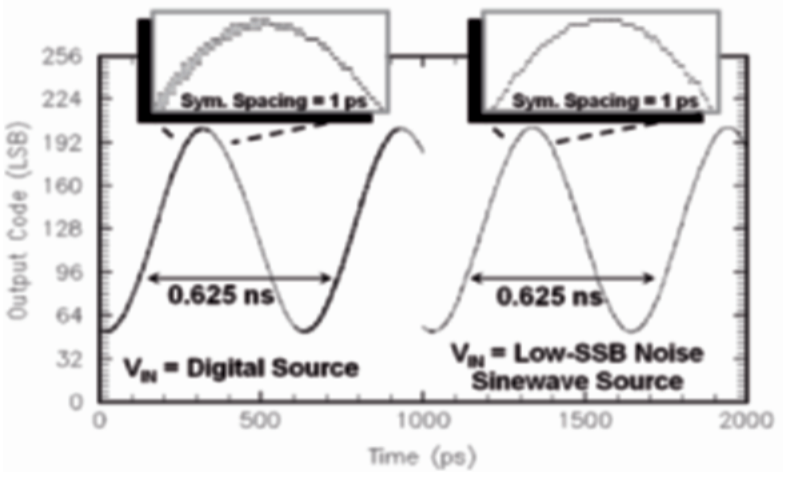


Fig. 15. Measured 1.6 GS/s Beat-Frequency test at 1.8V. The converter samples a 1.6026 GHz signal at 1.6 GS/s. The digital source showed deterministic jitter above 1.5 GHz.

INL and DNL can be misleading in determining an ADC's performance, because they are based on an averaging or histogram measurement, and because they are usually performed at low input frequencies. A much more stringent test, besides the classical spectrum plot of Fig. 13, is the beat frequency test, in which the input and sampling frequency are almost identical, resulting in a low frequency reconstructed signal corresponding to the difference in or "beat between" F_S and F_{IN}. Fig. 14 shows the measured 800 MHz beat-frequency test at 1.8V. The converter samples a

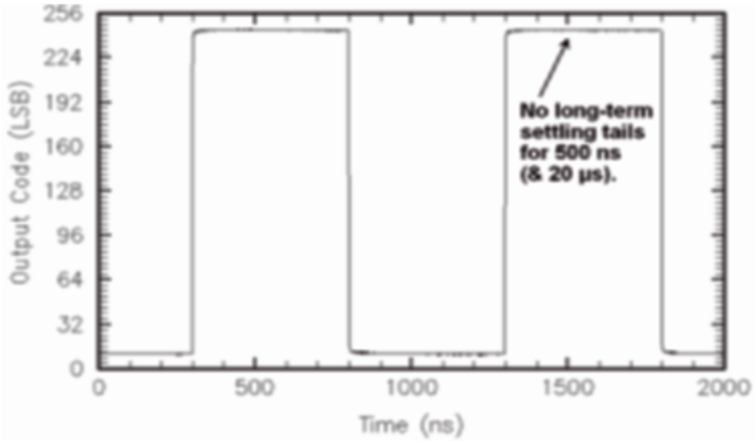


Fig. 16. Measured Pulse-Response at 1.8V and 1.6 GS/s. V_{IN} = 1 MHz, 0.70 Vp-p differential squarewave.

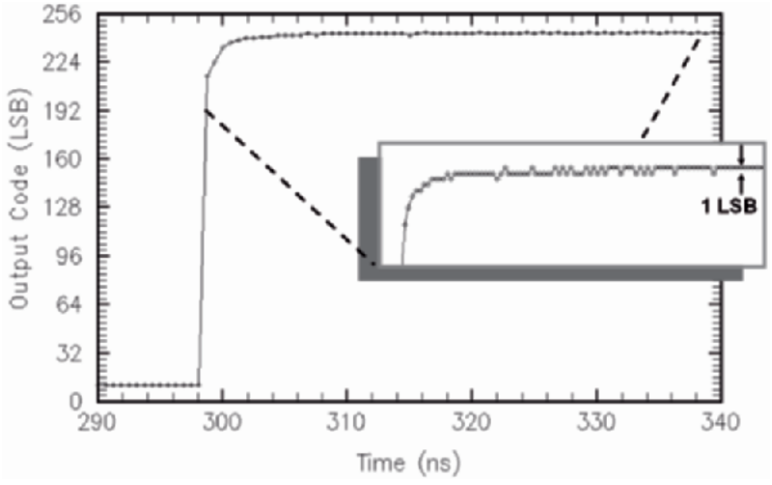


Fig. 17. Zoom-in of previous Pulse-Response at 1.8V and 1.6 GS/s.

810.7 MHz, 0.70 V_{DIFF} Squarewave at 810 MS/s. A graphical representation of the test principle is indicated in the lower part of the figure. Fig. 15 repeats the same test at 1.6 GS/s, to be consistent in reporting the ADCs performance at this higher conversion rate. Since the third harmonic is substantially above the 1.75 GHz input bandwidth of the ADC, the resulting waveform looks nearly like a sinewave. However, this stringent test showed less the performance limitations of the ADC, than the test equipment, which has an even-odd deterministic jitter above 1.5 GHz. We documented that this performance degradation was indeed the digital pulse generator by replacing this source with a high-purity sinewave at the same 1.6 GHz input frequency, with substantially better results, as seen.

Table I. Performance Summary at 1.6 GS/s for single ADC.

	F_{IN} = 97.77 MHz	F_{IN} = 797.77 MHz (Nyquist)
Sample Rate, F_S	1.6 GS/s	
Resolution	8 bits	
Max DNL	±0.15 LSB	
Max INL	±0.35 LSB	
SNR	48 dB	46 dB
SFDR	61 dB	56 dB
THD	- 57 dB	- 57 dB
ENOBs	7.60	7.26
Interleave aperture offset	< 0.35 ps @ F_S = 1 GS/s & F_{IN} = 1.5 GHz	
Input (–3 dB) Bandwidth	> 1.75 GHz	
Resolution (-0.5 ENOB) Bandwidth	1.0 GHz	
Input Range	±400 mV differential	
Input Capacitance	1.8 pF (to gnd, w/o package)	
Input Termination	50 Ω (100 Ω differential)	
Single Supply	1.8 V	
Analog (DC) Current	245 mA	
Switching (AC) Current	185 mA	
LVDS Output Drivers	90 mA	
ADC Core Power (w/o outputs)	774 mW	
ADC core area	3.6 mm^2	
ADC die area	16 mm^2 (for dual ADC, pad limited)	
Package	128-pin EPQFP	
Technology	0.18 µm CMOS (1-poly, 5-metal) No capacitor module nor dual-gate process	

One final set of measurements, pulse responses, is presented in Figs. 16 and 17. These are extremely important for test and measurement applications, where a settling tail would create an unacceptable baseline or gain wander. No settling tails within the 8-bit resolution of the ADC were attributable to the converter, and we once again were limited by the accuracy of our test equipment. The immediate transient response, however, is limited by the input bandwidth of the ADC.

Table I shows a performance summary at 1.6 GS/s and 1.8 V. All measurements presented in this paper are for a single ADC, although the 128-pin Exposed Paddle QFP package houses a dual ADC, as indicated in the total die area. Unlike most publications on GHz ADCs, these results are

with the outputs delivering the full undecimated data, using sixteen 800 MS/s LVDS output drivers. The total power with an active input, including the LVDS drivers, is 936 mW, which has been substantially reduced from the 1.4 W reported in our initial conference presentation [24]. The improvement was achieved by reducing the device sizes in analog blocks. Compared to the earlier publication, the 100 MHz distortion and ENOB values were also improved. Additional power savings by scaling the device sizes even further are possible due to the use of calibration.

The -3dB input bandwidth of 1.75 GHz includes the socket, board, and balun, and is assumed to be higher for the ADC alone. A zoom-in die photo of one ADC is shown in Fig. 18. The large and accurate calibration reference generator is shown, as is the calibration control logic, and the individual calibration memory with adjustment current DACs. The input MUX and Track&Hold are quite large, in part because metal-metal caps

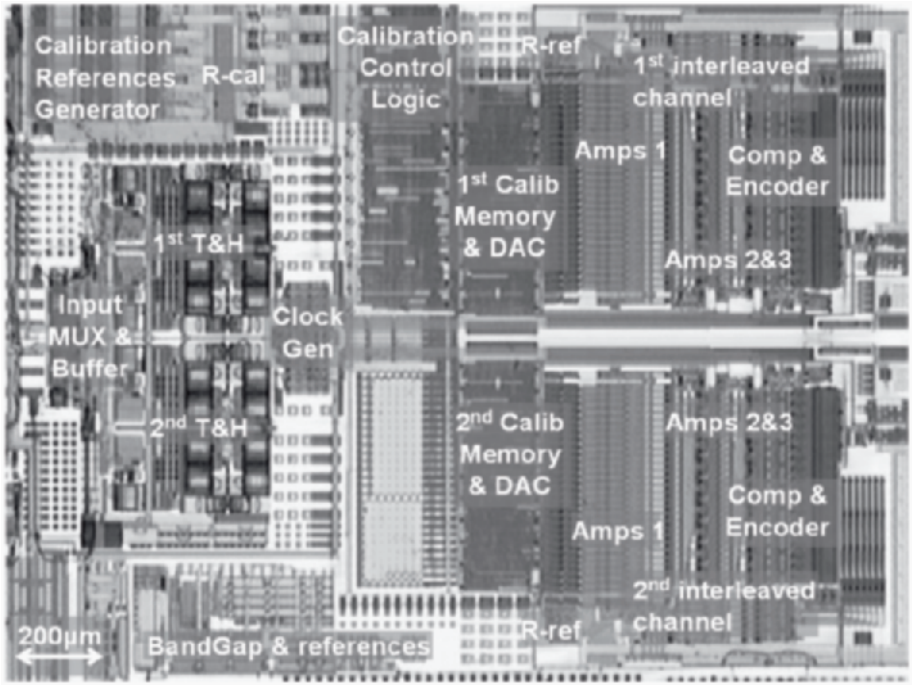


Fig. 18. Die photo of single ADC showing on-chip calibrator and 2X interleaving.

were used for the sampling capacitors and boost capacitors. The amplifiers shown are those for the initial conference presentation, consistent with the higher power consumption. The power and performance reported here corresponds to a smaller layout, where device and layout parasitics were substantially reduced. For this die photo, metals 4 and 5 have been removed to show the underlying structure.

7. Conclusion

We have demonstrated for the first time the significant performance enhancement calibration brings to folding-interpolating Analog-to-Digital Converters (ADCs). Offset calibration and folding have proven very complementary techniques to enhance the performance of converters. This resulted in a 1.8V ADC in 0.18um CMOS achieving a conversion rate exceeding 1.6 GS/s, since the amplifier device sizes were minimized to maximize speed, without the restriction of device matching. At an F_{IN} of 50 MHz, 7.5 ENOBs were maintained up to 1.8 GS/s, the conversion speed limited only by the test system. At 1.6 GS/s, the ADC achieved 0.15 LSB DNL, 0.35 LSB INL, 7.6 ENOB at 100 MHz input, and 7.26 ENOB at Nyquist. The ADC uses 2X interleaving, with each bank running at 800 MHz for this nominally 1.6 GS/s differential ADC. In the absence of interleaving, this is the highest speed CMOS 8-bit single-channel ADC published to date. At this speed, current consumption from a single 1.8V supply was 245 mA analog, 185 mA digital, and 90 mA for the LVDS drivers.

The ac performance is approximately 1.5 ENOBs higher compared to the same circuit with calibration disabled. The use of best design practices to optimize the ADC linearity prior to introducing calibration resulted in this small required dynamic calibration range, simplifying the calibrator circuitry and resulting in stable continuous performance over time without recalibration. Therefore, the fully on-chip calibration is performed automatically, just one time at power-up.

Acknowledgements

The authors wish to thank Andrew Glenny, Don Archer, George Cleveland, Susan Luschas, Rick Cullen, Servando Aguilar, Wen-Jung Hsu, Ben Watts, Bob Hartford, Fred Bowman, Jacque Margolycz, Neeraj Pendse, Jaime Bayan, Li Zhang, Sean Duggan, Andreas Tüchler, Eric Blom, Satoshi Sakurai, Helmut Strauch, Norbert Drescher for contributions; Frank Trautmann, Christian Glassner, Martin Embacher, Shoba Rao, Ian King, Chung-Yu Chen, Leo Azevedo, Paul McCormack, Alex Cherkassky, Mark May, Pier Andrea Francese for discussions; Gabriele Manganaro, David Boisvert, Peter Holloway and Pat Tucci for strong management support.

References

[1] M. Steyaert and K. Uyttenhove, "Speed-power-accuracy trade-off in high-speed analog-to-digital converters: now and in the future…", *Proceedings of the Ninth Workshop on Advances in Analog Circuit Design (AACD '00),* April 2000, pp. 1-22.

[2] R. C. Taft and M. R. Tursi, "A 100-MSPS 8-b CMOS subranging ADC with parametric operation from 3.8V down to 2.2 V", *IEEE J. Solid-State Circuits*, vol. 36, no. 3, pp. 331-338, March, 2001.

[3] R. J. van de Plassche and R. E. J. van der Grift, "A high-speed 7 bit A/D converter", *IEEE J. Solid-State Circuits*, Vol. SC-14, No. 6, December 1979, pp. 938-943.

[4] R. Keyes, "The effect of randomness in the distribution of impurity atoms on FET thresholds", *Applied Physics*, Vol. 8, p. 251, 1975.

[5] J. D. Burnett, K. Erington, C. Subramanian, and K. Baker, "Implications of fundamental threshold voltage variations for high-density SRAM and logic circuits," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 15-16, 1994.

[6] J. T. Watt, "Modeling the performance of liquid-nitrogen cooled CMOS VLSI," *Tech. Rep. Tech Rep. #G725-3, Stanford University*, CA, 1989.

[7] T. Mizuno, J. Okamura, and A. Toriumi, "Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFETs," *IEEE Transactions on Electron Devices*, vol. ED-41, no. 11, pp. 2216-21, 1994.

[8] I. Opris, L. Lewicki, and B. Wong, "A Singled-Ended 12b 20MSample/s Self-Calibrating Pipeline A/D Converter", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 138-139, Feb., 1998.

[9] J. Ming and S. H. Lewis "An 8b 80MSample/s Pipelined ADC with Background Calibration", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 42-43, Feb., 2000.

[10] D. Fu, K. Dyer, S. Lewis, P. Hurst, "Digital Background Calibration of a 10b 40Msample/s Parallel Pipelined ADC", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 140-141, Feb., 1998.

[11] K. Dyer, D. Fu, S. Lewis, P. Hurst, "Analog Background Calibration of a 10b 40Msample/s Parallel Pipelined ADC", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 142-143, Feb., 1998.

[12] O. E. Erdoğan, P. J. Hurse, and S. H. Lewis, "A 12b Digital-Background-Calibrated Algorithmic ADC with –90dB THD", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 316-317, Feb., 1999.

[13] W. Ellersick, C.-K. K. Yang, M. Horowitz, and W. Dally, "GAD: A 12-GS/s CMOS 4-bit A/D Converter for an Equalized Multi-Level Link", in *IEEE Symp. on VLSI Circuits Digest of Tech. Papers*, pp. 49-52, June, 1999.

[14] K. Nagaraj, D. A. Martin, M. Wolfe, R. Chattopadhyay, S. Pavan, J. Cancio, and T. R. Viswanathan, "A Dual-Mode 700-MSamples/s 6-bit 200-MSamples/s 7-bit A/D Converter in a 0.25-µm Digital CMOS Process", *IEEE J. Solid-State Circuits*, vol. 35, no. 12, pp. 1760-1768, Dec., 2000.

[15] K. Kattmann and J. Barrow, "A Technique for Reducing Differential Non-Linearity Errors in Flash A/D Converters," in *IEEE ISSCC Dig. of Tech. Papers*, pp. 170-171, Feb., 1991.

[16] K. Bult and A. Buchwald, "An Embedded 240-mW 10-b 50-MS/s CMOS ADC in 1mm^2", *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1887-1895, Dec., 1997.

[17] K. Sushihara, H. Kimura, Y. Okamoto, K. Nishimura, and A. Matsuzawa, "A 6b 800Msamples/s CMOS A/D Converter", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 428-429, Feb., 2000.

[18] G. Geelen, "A 6b 1.1Gsample/s CMOS A/D Converter", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 128-129, Feb., 2001.

[19] M. Choi and A. A. Abidi, "A 6b 1.3GSample/s A/D Converter in 0.35 µm CMOS", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 126-127, Feb., 2001.

[20] P. Scholtens and M. Vertregt, "A 6 bit 1.6GS/s Flash ADC in 0.18µm CMOS using Averaging Termination", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 168-169, Feb., 2002.

[21] K. Poulton, R. Neff, A. Muto, W. Liu, A. Burstein, M. Heshami, "A 4Gsample/s 8b ADC in 0.35µm CMOS", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 166-167, Feb., 2002.

[22] X. Jiang, Z. Wang, M. F. Chang, "A 2GS/s 6b ADC in 0.18µm CMOS", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 322-323, Feb., 2003.

[23] K. Poulton, R. Neff, B. Setterberg, B. Wuppermann, T. Kopley, R. Jewett, J. Pernillo, C. Tan, A. Montijo, " A 20GS/s 8b ADC with a 1MB Memory in 0.18µm CMOS", in *IEEE ISSCC Dig. of Tech. Papers*, pp. 318-319, Feb., 2003.

[24] R. C. Taft, C. Menkus, M. R. Tursi, O. Hidri and V. Pons, "A 1.8V 1.6GS/s 8b self-calibrating folding ADC with 7.26 ENOB at Nyquist frequency," in *ISSCC Digest of Technical Papers*, pp. 252-526, Feb., 2004.

SUB-HARMONIC LIMIT-CYCLE SIGMA-DELTA MODULATION, APPLIED TO AD CONVERSION

Sotir Ouzounov[1,2], Hans Hegt[2] and Arthur van Roermund[2]

[1]Philips Semiconductors, Prof. Holstlaan 4, 5656AA, Eindhoven, The Netherlands
[2]Eindhoven University of Technology, Den Dolech 2, 5600 MB, Eindhoven, The Netherlands

Abstract

In this paper a new mode of operation for Sigma-Delta Modulation (SDM) is proposed and applied to AD conversion. The proposed operation is identified as a sub-harmonic limit-cycle mode. It is shown, that in such sub-harmonic mode a more aggressive loop filter function can be applied that results in a better modulator performance. Moreover, most of the building blocks in the SDM loop operate at a frequency that is an integer fraction of the applied sampling frequency. That brings several very attractive advantages: a reduction of the required power consumption per converted bandwidth, immunity to excessive loop delays and to DAC waveform asymmetry, and a higher tolerance to clock imperfections. The proposed SDM ADC offers an alternative to low-speed low-performance low-cost and high-speed high-performance and high-cost implementations by introducing a new trade-off between limit cycle frequency and clock frequency.

The limit cycle modes are studied via a graphical application of the Describing Function (DF) theory. A second order continuous time SDM with 5MHz conversion bandwidth and 1GHz sampling frequency is used as a test case for the evaluation of the performance of the proposed type of modulators. The modulator is implemented in a 1.8V, 0.18 micron digital CMOS process. The theoretical discussion is validated with high level and transistor simulations, and measurements.

1. Introduction

Currently, multiple communication applications begin to utilize SDMs with large bandwidths (above 10MHz) and with high accuracy (better than 14bit). Despite the technology advances, those requirements pose a significant challenge for SDM. At the same time, the room for architectural and conceptual improvements of SDMs seams to be exhausted and the implementations are striving to achieve high performance mainly through an increase of cost and complexity. One of the main cost factors is the clock

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 97–125.

speed. An increase in the sampling frequency requires an increase of the gain-bandwidth product of the building blocks, faster settling times and power hungry decimation filters. Alternatively, the performance can be improved via introduction of multi-bit quantizers or multi-stage (MASH) implementations. Those alternatives increase quite significantly the complexity of the design and again lead to an increase of area and power consumption.

Here, we discuss in particular, high-order single-bit SDM designs. In Table I, several state of the art implementations are selected that realize single-loop single-bit SDMs. The designs are ordered in an increasing order L of the loop filter and an increasing signal bandwidth. The table illustrates ADCs that are designed for different frequency ranges, application areas and in different CMOS processes.

Table 1. Comparison: evaluation vs. measurements.

Ref.	L	Q bits	Signal BW, Hz	f_S, Hz	OSR	DR, dB WNM	**Diff. I, dB**	DR, dB Sim.*	**Diff. II, dB**	DR, dB, **Meas.**	SNDR, dB, **Meas.**	Process	Applica-tion
[1]	3	1	16k	1.538M	100	126.4	**49.4**	106.7	**28.7**	77	62	0.5µm CMOS	audio
[2]	3	1	25k	2M	100	126.4	**38.4**	106.7	**18.7**	88	85	0.35 µ m CMOS	audio
[3]	4	1	400	256k	320	>205	**>100**	173	**68**	105		0.6 µm CMOS	Measure-ment
[4]	4	1	600k	256M	210	186	**100**	153.7	**67.7**	86		90 nm CMOS	
[5]	5	1	300k	21.07M	32	134	**37**	85	**5**	80	77	0.35 µm CMOS	AM/FM radio
[6]	5	1	1M	64M	32	134	**58**	85	**9**	76	75.5	0.18 µm CMOS	Bluetooth receiver
[7]	5	1	200k	26M	48	153	**63**	105	**15**	90		0.18 µm CMOS	GSM
			1.228M	76.8M	64	167	**84**	120	**47**	83			CDMA
			3.84M	153.6M	40	144.7	**70.7**	96	**22**	74			UMTS
[8]	5	1	3.1M	400M	64	167	**105**	120	**58**	62	56	0.6 µm CMOS	

One of the goals of the table is to illustrate the difference in the ADC performance as evaluated analytically or simulated and the actual measurement results from some of the best-reported implementations. The analytical prediction for the achievable dynamic range (DR) is made with the help of the white noise model (WNM) of the quantizer in the SDM loop [9], [10]. The WNM treats the quantizer as a noise source with a white spectral distribution independent from the input signal. The WNM takes into account the loop filter characteristics, the applied OSR and the resolution of the

quantizer. For higher order loop filters the zeros in the loop filter characteristics are often optimized in order to assure maximum quantization noise suppression in a certain bandwidth.

With the help of the WNM, the achievable dynamic range is evaluated with respect to the oversampling ratio (OSR) and the filter order. For full scale input signal the dynamic range [9] is given as:

$$DR_{WNM} = \frac{6(2L+1)OSR^{2L+1}}{\pi^{2L}} \tag{1}$$

The difference between the WNM prediction and the measurements is given in column 8 of Table 1. The SDM implementations (with respect to *L* and *OSR*) from Table 1 are also evaluated with high-level simulations. A generally available tool [11] is used for the evaluation. The results are noted in column 9 and the difference with the measurements is given in column 10. Several observations from the data in Table I have to be pointed out:

- The reported measurement results show a significant deviation from the estimation that can be made with the WNM. Moreover, the deviation increases with the increase of the order of the loop filter and the applied OSR.
- In some cases, the ideal high-level simulations predict performance much closer to that of the implemented ADCs. However, the difference with the measurements varies a lot between the designs and is worse for high-speed applications.
- The measured signal to noise and distortion ratio (SNDR), when given in the reference, is also included in the table in column 12. It shows that in practice there are implementation issues that further decrease the performance of the ADC with respect to the simulations.

Based on the above observations, the following conclusions can be drawn:

- The comparison: analytical model – simulation – measurement brings questions about the applicability of the WNM in modern SD design and respectively, about the level of understanding of the SDM operation.
- The SDM design is put on rather heuristic background that relies on simulations. Despite their evident importance, the simulations may not include all important design parameters and may not always bring new insights in the SDM operation and behaviour.

A step further in the conceptual understanding and theoretical description of the SDMs can be done if the modulators are treated as non-linear, closed-loop, sampled systems. Undoubtfully, such an approach is supported by a known phenomenon in the SDM operation, defined generally as limit cycle behaviour. Here, we define as limit cycles the periodic modes that appear at the output bit-stream of single-bit SDMs. It is well known that those modes

can cause appearance of spurious tones in the signal band and can lead to a significant deterioration of the performance of the modulator in terms of SNDR and DR. However, the WNM of the quantizer in the SDM loop, cannot not predict those modes as they originate from the non-linear nature of the SDMs. Respectively, an analytical model on the SDM operation is required that is capable of describing the limit cycle behaviour.

Though some recent studies [12] undertake a deep look in the limit cycle behaviour in SDM, the phenomenon is mainly studied heuristically via extensive simulations. This has resulted in SDMs parameterisation that follows simple "rules of the thumb" but without solid analytical background. The non-linear properties of the modulators and the phase-gain relation in the closed-loop are rarely studied analytically. That is why the WNM is unable to predict the appearance of limit cycles and their impact on the performance of the system.

For the understanding and the description of the limit cycles, a 'non-linear look' at SDM operation is required. In the 50ties and 60ties major developments in the treatment of closed loop non-linear systems were made in the context of automated control [13], [14] and [15]. Those works led to a generalized approach for the treatment of non-linear systems: the Describing Function (DF) representation of the non-linear element [16]. So far, in the context of SDM, those theories have received little attention. In this work, the DF theory is used for the modelling of the quantizer function and for the investigation of limit cycles modes in SDMs.

The paper is structured as follows: Firstly, a limit cycle model of the SDM operation is introduced. The possibility for operation in a sub-harmonic limit cycle mode is explained conceptually. Secondly, for the self-consistency of the paper, the relevant DF fundamentals are summarized and put in the context of SDMs evaluation. Then the limit cycle behaviour is explained with the help of Asynchronous Sigma-Delta Modulators (ASDM). A graphical determination of the limit cycles is demonstrated in the gain-phase plane with the help of the DF theory. Section 5 elaborates further on the limit cycle modes in synchronous SDMs. The impact of the extra loop delay and the hysteresis in the quantizer is incorporated in the analytical evaluation. Again, the limit cycles' behaviour is established graphically. Section 6 discusses the advantages and disadvantages of the sub-harmonic limit cycle operation. A comparison with the standard SDM design is made and the benefits of the new mode of operation are discussed. Finally, in Section 7, the hardware implementations of an ASDM and a SDM operating at a sub-harmonic limit cycle are presented.

2. SDM: Limit Cycle Model

A limit cycle model of SDM operation can be developed with the help of the system shown in Fig. 1. It is a closed-loop system built with a continuous time linear filtering block $L(j\omega)$ and a binary quantizer modelled with its transfer function $N(A)$ defined with respect to the amplitude A at its input. The sampling operation is embedded within the loop. The loop filter is of a second order and has a low-pass character. Here, the study is limited to a second order system because it is relatively simple and its properties can be easily visualized and interpreted.

In the general SDM treatment, the sampling operation is performed before or within the quantizer. Moreover, in practically all reported SDM implementations the quantizer is an internally sampled circuit. The ideal quantizer is memoryless and its position with respect to the sampling operation is of no consequence for the loop operation. Here, to decouple the amplitude quantization from the sampling, the sampling is performed after the quantizer.

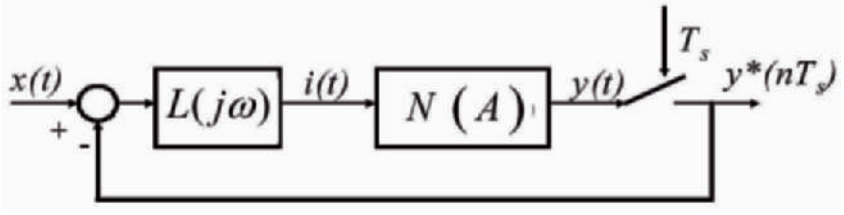


Fig. 1. Block diagram of SDM.

For the construction of the limit cycle model for SDM, firstly, the properties of the output bitstream in idle operation (*x(t)=0*) are considered. For an unbiased system the average value of the output signal should also be zero. Thus, under idle condition, due to the full symmetry of the loop and the presence of an integrator, the bitstream can consists only of N ones followed by N zeros. This requirement presupposes that the only oscillations that can exist in the loop are even integer multiples N of the sampling period T_s ($N = 2k$, and k=1, 2...) [17].

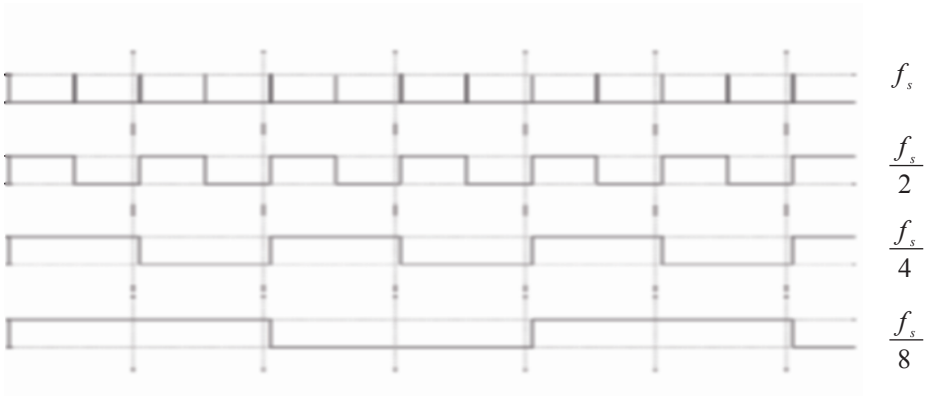


Fig. 2. Idle limit cycles: possible output waveforms.

In Fig. 2, three possible idle limit cycle modes are illustrated, respectively with frequencies $f_s/2$, $f_s/4$ and $f_s/8$. All of them can be defined as sub-harmonic modes because they have a frequency that is an even integer fraction of the applied sampling frequency. Further in the paper, the conditions for operation in one limit cycle mode or another are elaborated in detail. Here, we state that, in contradiction with the general perception of the SDM operation, the limit cycle with frequency $f_s/2$ is not always the only possible one. Moreover, it may not result in the best performance for the system. In fact the designer can choose the sub-harmonic mode and the frequency of operation of the SDM loop such that the best trade-off for the system performance can be made. Several notions are stated here and are going to be elaborated on further in the paper:

- Every SDM operates at one or more limit cycles. In fact, the situation when only one limit cycle is possible in the system can be considered as a rare case. In most practical situations, at least two limit cycles are possible and the SDM state depends entirely on the instantaneous amplitude of the input signal.
- The SDM will be considered stable when it is able to produce a bitstream representation of an arbitrary (for example, zero, DC, sinusoidal or Gaussian) input signal with magnitude within its dynamic range. This somewhat broad definition of stability is required in order to distinguish from instable modes that lead to a permanent clipping of the output to either +1 or –1, such that the SDM is not able to process the input signal. The above definition of stability identifies as stable also the SDMs that operate at a very low limit cycle frequency. This definition is needed

because quite often those modes are described as instable, only because their behaviour drastically deviates from the predictions of the WNM.

In order to facilitate the interpretation of the applied approach, next, the basic concepts of the DF theory are presented.

3. DF: Basic Definitions

The describing functions build quasi-linear approximations of the non-linear element according to a certain linearization rule and with respect to preliminary defined input signals, for example sinusoids. The linearization rule implements a criterion for the evaluation of the approximation. The most common criterion used in practice is the minimum mean square difference between the approximated output and the actual output of the non-linear element. Different describing functions can be derived for the same non-linear element with respect to different input signals. The basic and most often used DF is the sinusoidal input describing function (SIDF). It is derived under the assumption that there is a single harmonic component (pure sine) in front of the non-linear element. When applied to SDM in idle mode, that means that the signal in front of the quantizer is approximated with a single harmonic component with frequency of half the frequency of the applied clock signal. An assumption is made that in idle mode only the fundamental component of the limit cycle is propagated though the loop while its harmonics are very much suppressed. The validity of this assumption when the system uses higher order filters and oversampling ratios is in most cases evident. Here we aim qualitative evaluations and a possible small inaccuracy of the DF result is of a little consequence.

The SIDF for a binary quantizer and a binary quantizer with hysteresis h for a sine input signal with amplitude A are given respectively as:

$$N(A) = \frac{4}{\pi A} \tag{2}$$

$$N(A,\varphi) = \frac{4}{\pi A}\sqrt{1-\left(\frac{h}{A}\right)^2} - j\frac{4h}{\pi A^2} = \frac{4}{\pi A}e^{-j\sin^{-1}(h/A)} \qquad A>h \tag{3}$$

The quantizer output levels are normalized to +/-1 and φ accounts for the phase rotation of the complex function (3). In fact the ideal quantizer can be seen as a special case of the quantizer with hysteresis, when h=0.

4. Evaluation of Limit Cycle Oscillations in ASDM

The interpretation of limit cycle behaviour in SDM is significantly facilitated by a study of a system for which the limit cycle is an evident and inherent property. Such a system is the Asynchronous Sigma-Delta Modulator (ASDM).

4.1. ASDM Operation and Properties

There are several reasons to start the study of limit cycle behaviour with a closer look at ASDM:

- ASDM [18], [19] can have a very similar structure to that of a typical continuous time SDM. It is a closed-loop system (Fig. 3) built with a continuous time linear filtering block $L(j\omega)$ and a binary quantizer modelled with its describing function $N(A)$.

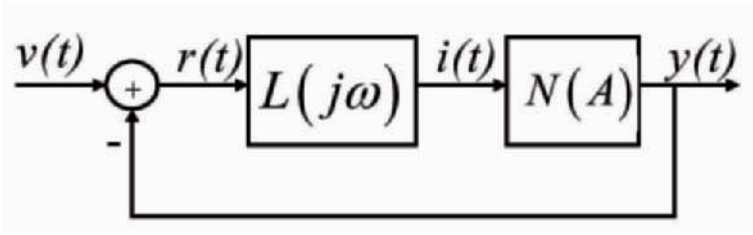


Fig. 3. ASDM Block diagram.

- The ASDM transforms the information in the amplitude of the input signal into time information in a binary output signal, without an external sampling signal. The amplitude-time transformation in the ASDM is done via an inherent periodic self-oscillation: a limit cycle. The limit cycle frequency ω_c can be regarded as a counterpart of the clock frequency in SDM and is defined [20] as the main ASDM design parameter that determines the spectral properties of the output signal and the quality of the amplitude-time transformation.

A detailed description of the properties of the modulator can be found elsewhere [21]. Here we pay special attention to the determination of the limit cycle behaviour of the modulator.

The possibility for self-oscillation in closed-loop linear systems is determined after an investigation of the phase-gain relation in the loop. A basic criterion can be used (most known as Barkhausen criterion) that states, that a closed-loop system is prone to self-oscillation at a frequency for which an integer multiple of 360 degrees of phase shift occurs and the gain is 1. In

the ASDM context such a criterion can be established with the help of the DF representation of the quantizer, because it allows application of linear theory. Undamped oscillations that are found in the linearized systems are then interpreted as limit cycle oscillations in the non-linear system. For consistency, the amplitude *A* of the limit cycle is defined and observed at a fixed position in the system, for example, for the signal *i(t)* in front of the non-linear element. The observation point is concluded from the fact that this is the point from the signal propagation path that exhibits continuous variations in amplitude with respect to the system parameters and the input signal *v(t)*. The existence and the parameters of the limit cycle oscillation are established for zero input signal, *v(t)=0*. The system in Fig. 3 is described with the following set of equations:

$$\begin{cases} I(j\omega) = -L(j\omega)Y(j\omega) \\ Y(j\omega) = N(A)I(j\omega) \end{cases} \tag{4}$$

The solution of (4) is given by the following equation:

$$1 + N(A_c)L(j\omega_c) = 0 \tag{5}$$

Equation (5) can be used in several ways in the study of limit cycles' properties. Firstly, it can be solved analytically for the determination of the limit cycle amplitude A_c and frequency ω_c. Alternatively, a graphical solution of the equation can be sought.

4.2. Graphical Prediction of Limit Cycle Oscillations (ASDM)

A solution of (5) can be found if the transfer characteristics of $L(j\omega)$, parameterised with respect to frequency and of $-1/N(A,\varphi)$, parameterised with respect to amplitude A are plotted in the same co-ordinate system. Then the crossing points of the two characteristics determine the solutions of equation (5) and consequently the limit cycle points. For the purpose, magnitude-phase plots, polar or root-locus representations can be employed. Here a magnitude-phase plot is used. An example for a second order system with quantizer defined as in (3) and loop filter defined as in (6) is given in Fig. 4. After deriving the magnitude and the phase of (6), the solution of (5) is found graphically.

$$L(j\omega) = \frac{(j\omega + \omega_z)\omega_p^2}{(j\omega + \omega_p)^2 \omega_z} \tag{6}$$

The DC gain in (6) is normalized to 1 and for the second order system two coinciding poles and a zero are introduced. The solution of (5) is given by the crossing point of the linear and the non-linear transfer characteristics, as indicated on the picture. For the chosen test case only one limit cycle is possible. The limit cycle frequency can be altered via modifications of $L(j\omega)$ and/or the quantizer. For example, the crossing point can be changed via a modification of the zero position in $L(j\omega)$ or via modification of the hysteresis value.

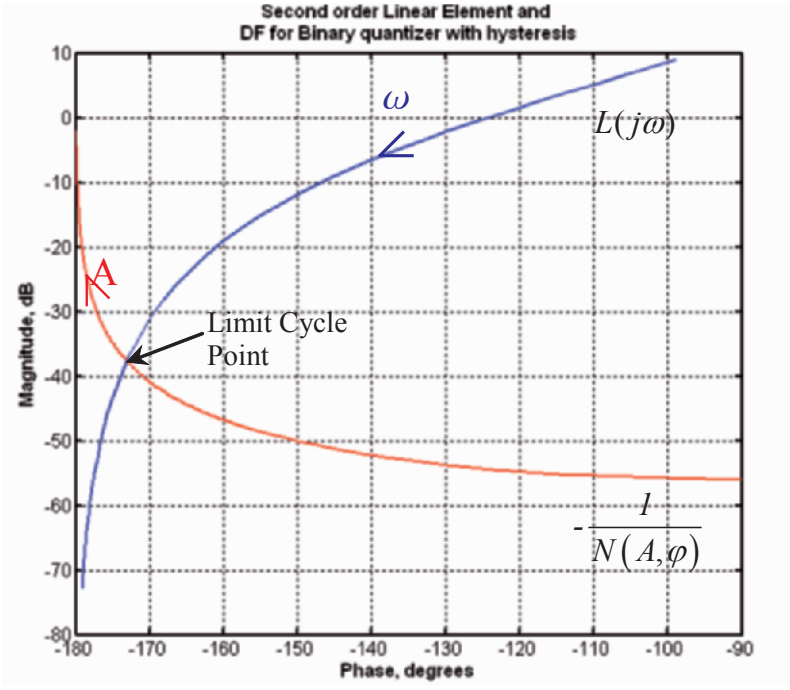


Fig. 4. Graphical determination of limit cycles in second order ASDM.

This approach allows a simple and quick estimate of the limit cycle oscillations. Its main advantage can be seen not only in the establishment of the limit cycle parameters but also in the visualization of the trends and dependences in the system. The evaluation of the limit cycles is confined to

plotting the two complex functions $L(j\omega)$ and $-1/N(A,\varphi)$ in a way that can give most insight in the system behaviour and trends.

In practice, if the initial values of A_c ω_c are different from those determined from the crossing point, the loop mechanism would adjust them so that the limit cycle point is reached. In the described case, the DF is giving a very good prediction for the existence, frequency and amplitude of the limit cycle.

5. Limit Cycles in SDM

The SDM can be approached with the limit cycle model created for its asynchronous counterpart. The sampling operation is embedded within the loop, as shown in Fig. 1.

5.1. Graphical Evaluation of the Limit Cycle Modes in SDM

In the context of SDM, we need to investigate the interaction between the sampling clock with frequency ω_s and the conditions for asynchronous limit cycle as established in the previous section for the ASDM. The limit cycle modes in a SDM can be established if the sampling operation is treated as a process that introduces phase delay in the loop. A phase shift is introduced that is dependant on the clock frequency ω_s. Again, firstly, the idle limit cycle modes are studied as those determine stable points in the system operation. The situation when the input signal could induce a limit cycle that does not exist in idle mode can be considered as a rare special case.

The phase delay that is introduced in the loop depends on the particular sub-harmonic mode, such that per mode *N*, the delay, normalized to the period of the sampling clock, can have any value in the range $(0:2\pi/N)$. The limit cycle with frequency $\omega_s/2$ will be called first mode. All other limit cycles with lower frequencies are called sub-harmonic limit cycle modes and are referred to with their respective order *N*. For the analytical evaluation the sampling delay in the loop can be taken into account via a modification of the describing function of the quantizer. The new sampled describing function is given as $N^*(A,\varphi,\tau_{s,\mathrm{N}})$, where the sampling delay

for limit cycle mode N is $\tau_{s,N}$. Limit cycle oscillations occur for sets of A, φ, $\tau_{s,i}$ and ω for which:

$$L(j\omega) = -1/N^*\left(A, \varphi, \tau_{s,i}\right) \tag{7}$$

The DF recognizes the signal in front of the quantizer only at the sampling instants. The frequencies, for which the sampling operation can add enough delay into the loop such that the locus $-1/N(A)$ is crossed, correspond to limit cycle modes. The idle limit cycle model is illustrated in Fig. 5. An ideal binary quantizer is used in the evaluation. The DF of such a quantizer (2) is dependent only on the sign of the driving signal and is a real function of the amplitude A, thus it is a straight line on the gain-phase plot. Furthermore, the DF assumes that the input to the quantizer is sinusoidal and only the fundamental component of the output signal is taken into account in the analysis. Thus for the idle case when no other signal components are propagated in the loop but the clock signal or its integer multiples, the DF is defined with respect to the idle oscillations in the loop.

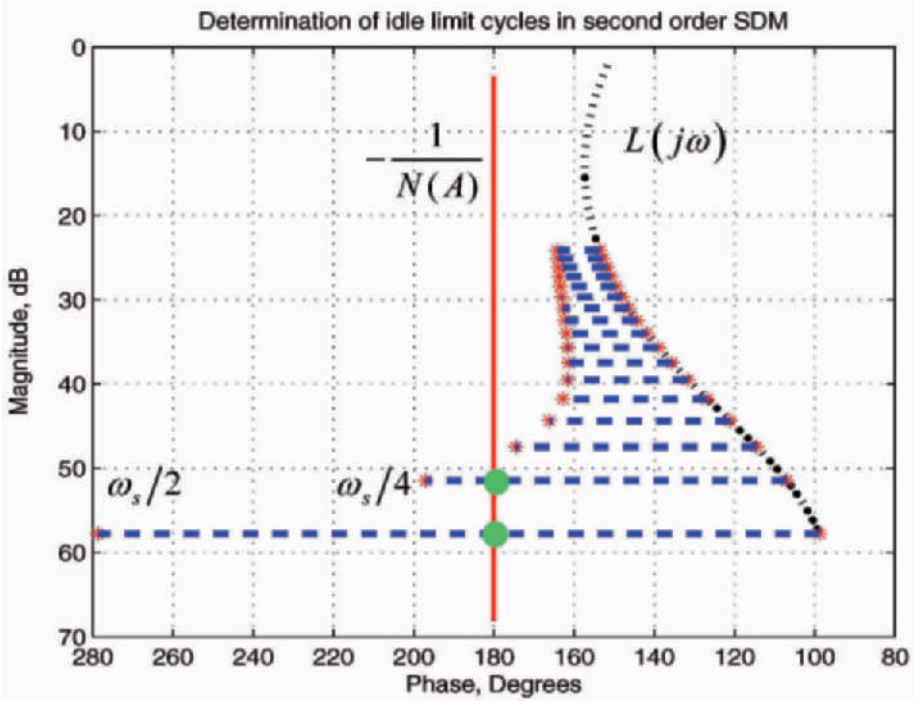


Fig. 5. Idle limit cycles in second order SDM.

The delay for each ω_s/N frequency is shown as a line originating from the discrete points $L(j\omega_s/N)$ and with length equal to $2\pi/N$. For our test case, two limit cycle modes are identified with frequencies $\omega_s/2$ and $\omega_s/4$. Limit cycles with lower frequency are not possible because the total phase delay in the loop does not provide 360 degrees phase shift. The first limit cycle with frequency $\omega_s/2$ is the only mode that is considered in practice for the design of SDMs. The assumption that the loop always operates with $\omega_s/2$ is the base for the white noise model. However, the second limit cycle mode that corresponds to $\omega_s/4$ should also be taken into account. An increase of the loop gain shifts the $L(j\omega)$ locus downwards and results in higher amplitude of the limit cycle; however, its frequency remains the same, because the phase characteristic of the filter remains the same. The loop can be forced to operate at one limit cycle or another with a proper phase compensation of the linear part of the loop. Fig. 5 shows that even simple second order sigma delta modulators can enter multiple sub-harmonic limit cycles. The number of possible sub-harmonic modes depends on the filter transfer characteristics and the applied sampling frequency. For the example in Fig. 5, the filter coinciding poles are positioned at low frequency and for that frequency range the filter causes maximally 180 degrees phase shift. The zero is positioned at high frequencies and reduces the phase shift back to 90 degrees.

An important remark is that in SDM all limit cycles, if simultaneously possible, have harmonically related frequencies. An inspection of Fig. 5 shows that for the studied filter example, loop operation with frequency $\omega_s/2$, can be achieved only if the sampling is applied in the region where the double pole introduces phase rotation less than 90 degree (if it is above 90 degrees, the first sub-harmonic mode 2, will also be possible). This however, would result in a filter with very high frequency poles and respectively with bad suppression of the quantization noise in the baseband. A practical second order SDM implementation has the poles of the loop filter as close to the origin as possible. From Fig. 5 it can be seen that such filter configuration results in the simultaneous possibility for at least two idle limit cycle modes. The active mode for zero input can be established with perturbation analysis. It shows that for zero input the limit cycle with $\omega_s/4$ is stable and active, while the limit cycle with frequency $\omega_s/2$ is entered for busy operation and for certain amplitudes of the input signal.

The amplitude and the frequency of the limit cycle are modulated by the input signal. That leads to the appearance of spectral components positioned around the limit cycle frequency and with energy that depends on the amplitude of the input signal. In order to prevent the appearance of these components in the baseband, the limit cycle frequency has to be much higher than the desired conversion bandwidth. In the SDM case, the sampling operation de-correlates the limit cycle tones from the input signal and spreads them in frequency. The tones are not distinguishable from the noise floor even for high input levels under the condition that the limit cycle frequency is high enough. When more than one limit cycle is possible, with the increase of the input signal, the system jumps between the limit cycles depending on the instantaneous amplitude of the signal in front the quantizer.

5.2. Impact of the Extra Loop Delay and the Hysteresis

From the limit cycle point of view, the most important practical constraints are introduced by the extra loop delay and the hysteresis in the quantizer. Those two are underlined among the numerous other practical problems that may appear during an implementation, because they have a dominant impact on the performance and are present in any implementation. For example, in any physically realizable system the propagation delay is not zero and some hysteresis is present in the quantizer. The limit cycle model offers a better understanding of the mechanism through which those effects deteriorate the SDM performance. The impact of the total delay τ in the loop can be incorporated in the filter transfer function:

$$L(j\omega) = \frac{(j\omega + \omega_z)\omega_p^2}{(j\omega + \omega_p)^2 \omega_z} e^{-j\omega\tau} \tag{8}$$

In Fig. 6, the limit cycle behaviour of an SDM with a binary quantizer is evaluated for two values of the loop delay.

The delay decreases the phase margin to the low frequency limit cycles. For the studied filter configuration and clock frequency, a delay of 20% of the clock frequency already led to the appearance of $\omega_s/8$ limit cycle. The impact of the delay for higher frequencies is higher. A delay that approaches the period of the sampling clock leads to operation at a very low frequency limit cycle and has a severe deteriorating effect on the performance of the modulator.

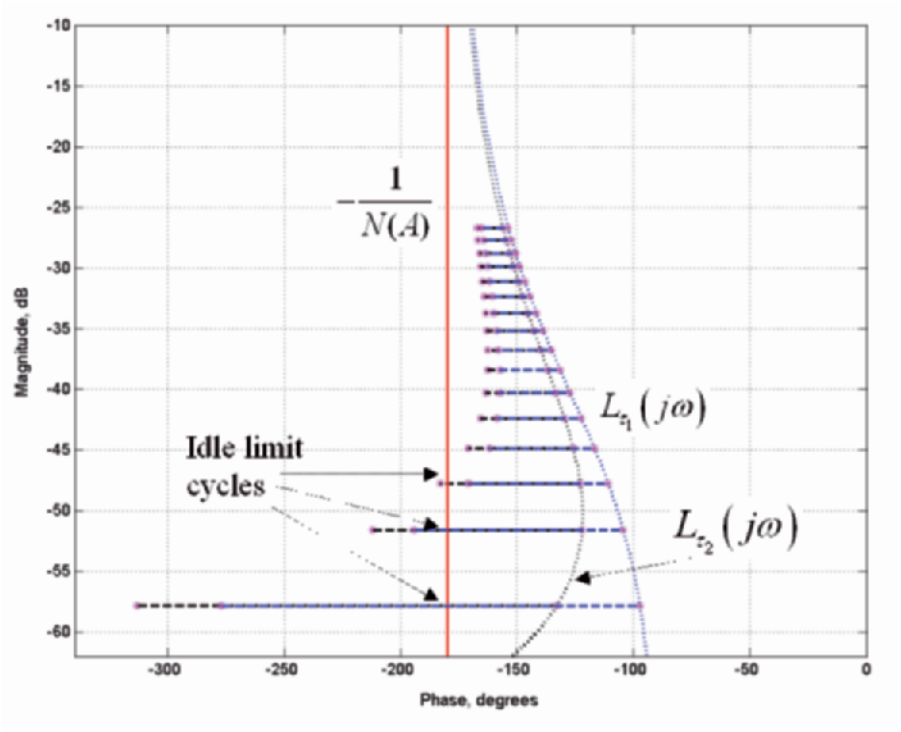


Fig. 6. Evaluation of the impact of the delay into a second order SDM, a) for delay $\tau_1 = 0$*; b) for delay* $\tau_2 = 0.2T_{CLK}$.

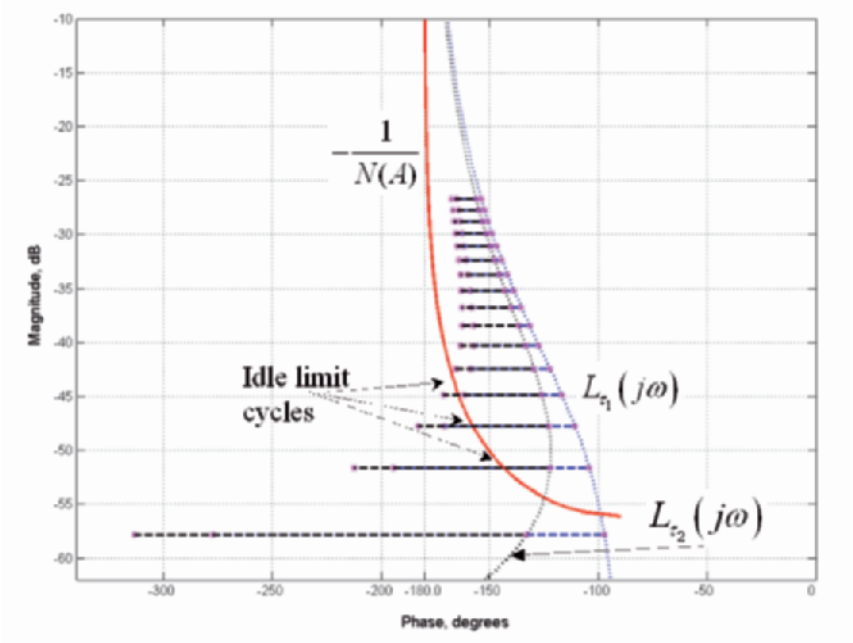


Fig. 7. Evaluation of the impact of the quantizer hysteresis, a) for delay $\tau_1 = 0$*; b) for delay* $\tau_2 = 0.2T_{CLK}$.

The impact of a hysteresis in the quantizer is illustrated in Fig. 7. The hysteresis also limits the phase margin in the loop. For the example, as shown in Fig. 7, the hysteresis renders limit cycle $\omega_s/2$ impossible.

The graphical application of the limit cycle model of the SDM operation can serve as a tool for a quick evaluation of the available phase margin in the loop and the possible limit cycle modes.

6. SDMs Operating at Sub-Harmonic Limit Cycles

From the discussion so far, it is clear that every SDM is prone to entering one or more idle limit cycle modes. The possible modes need to be carefully investigated and taken into account in the transistor level design of the building blocks. The design of $L(j\omega)$ (especially for higher order transfer functions) should be done with considerations for the possible limit cycles for the chosen sampling frequency.

The most common approach used in practice is to aim always for maximum limit cycle frequency equal to $\omega_s/2$. That requires a significant GBW from the linear part. The requirement is becoming much more stringent when clock speeds, at the edge of the technology, are desired. Achieving a GBW product that is high enough for those clock speeds can be very expensive, if possible at all. Even when the technology allows operation at $\omega_s/2$ that would leave very little room for extra processing inside the loop because of the very high sensitivity to extra loop delays [22]. An alternative approach is possible: The SDM loop can be deliberately designed to operate at one or more sub-harmonic limit cycles i.e. with an even integer fraction of the clock frequency. For example the sub-harmonic modes shown in Fig. 6 can be used. In this case, only the sampling switches operate at the full clock speed. The rest of the loop operates at the limit cycle frequency, thus with a significantly lower average speed.

6.1. Calculation of the Baseband Quantization Noise Power

More insight in the consequences of working at a limit cycle can be acquired after the evaluation of the quantization noise power P_N within bandwidth B:

$$P_N = \int_0^B E(f)\frac{2}{\left|1+L(f)\right|^2}df \tag{9}$$

where E(f), is the quantization noise spectrum. Firstly, the white noise model for a binary quantizer with a quantization step of 2 is used: $E(f)=1/(3f_s)$. For a second order system, (9) is evaluated as:

$$P_N \approx \frac{\pi^4}{15}\left(\frac{2B}{f_s}\right)^5 \tag{10}$$

From (10) follows that P_N decreases with 15dB for each doubling of f_s. Secondly, the quantization noise spectrum $E_{LC}(f)$ that arises from the sampling of a square wave with frequency f_c (the limit cycle frequency) with clock frequency f_s, is evaluated [23], [24] as:

$$E_{LC}(f) = \frac{8}{3} f_c T_s^2 \tag{11}$$

Calculation (11) has to be done with respect to the lowest possible limit cycle as predicted by the DF evaluation. When we take into account that the loop operates at an even fraction $N= f_s /f_c$ of the clock frequency, from (11), P_{NLC} is:

$$P_{NLC} = \frac{8}{3} f_c T_s^2 \frac{\pi^4}{5 f_c^4}(2B)^5 = \frac{8\pi^4}{15} N^3 \left(\frac{2B}{f_s}\right)^5 \tag{12}$$

From (12) it can be concluded that when the SDM operates at a limit cycle, a penalty of 9dB is introduced for each reduction by two of the loop frequency. In Table 2, several SDMs operating with different sampling frequencies are shown. Standard SDM 1 operates with the highest f_s and is used as a reference. The reduction of the performance due to a decrease of the sampling speed is illustrated. A comparison between SDM 2 and Limit Cycle SDM 1 (LCSDM 1) shows that when the two loops operate at the same frequency ($f_s/4$ in this case), LCSDM shows 6dB better performance. The improvement is 12dB for LCSDM 2 with respect to SDM 3. The results show that it is far more advantageous to force the SDM to function at a low limit cycle frequency than just to decrease the clock f_s.

Table 2. Comparison: Standard SDM and Limit Cycle SDM.

	Clock Frequency	Limit Cycle Frequency	P_N, dB penalty	Simulated $SNDR_{50\%,FS}$
Standard SDM 1	f_s	$f_s/2$	reference	81.7dB
Standard SDM 2	$f_s/2$	$f_c=f_s/4$	+15dB	66.5dB
Standard SDM 3	$f_s/4$	$f_c=f_s/8$	+30dB	51.7dB
LCSDM 1	f_s	$f_c=f_s/4$	+9dB	72.5dB
LCSDM 2	f_s	$f_c=f_s/8$	+18dB	63.7dB

In the last column of Table 2, Matlab simulation results are given for the SNDR in each case. A second order system is evaluated with a baseband of 5MHz and clock frequency of 1GHz. In Fig. 8 the output spectra of LCSDM 2 operating at idle mode and very small input are compared. In an idle mode the modulator operates with limit cycle frequency of $f_s/4$. When a very small signal is applied the second limit cycle mode $f_s/8$, as predicted by the

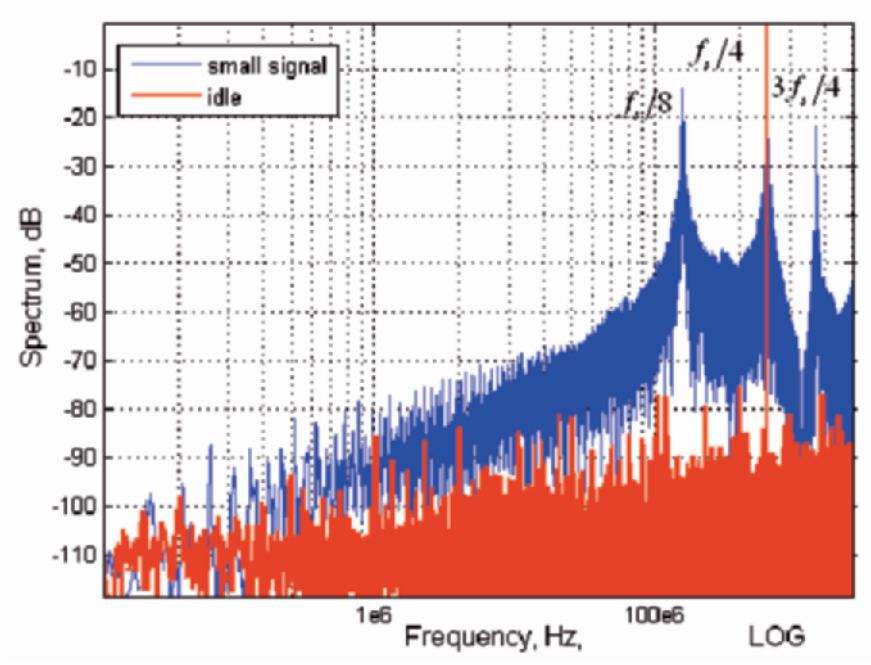


Fig. 8. Idle limit cycle operation with $f_c = f_s/4$ and small signal limit cycle operation with lowest frequency $f_c = f_s/8$.

DF evaluation, is also active for certain instantaneous amplitudes of the input signal. The limit cycle modes of operation can be recognized from the spectral peaks at the limit cycle frequencies. The simulation results confirm the theoretical expectations.

In Fig. 9, the output spectrum is shown for a large input signal. A 1MHz sinusoidal input signal with amplitude of 50% (-6dB) of Full Scale (FS) is used. The spectral peaks due the limit cycles are spread into the noise floor and are not distinguishable any more.

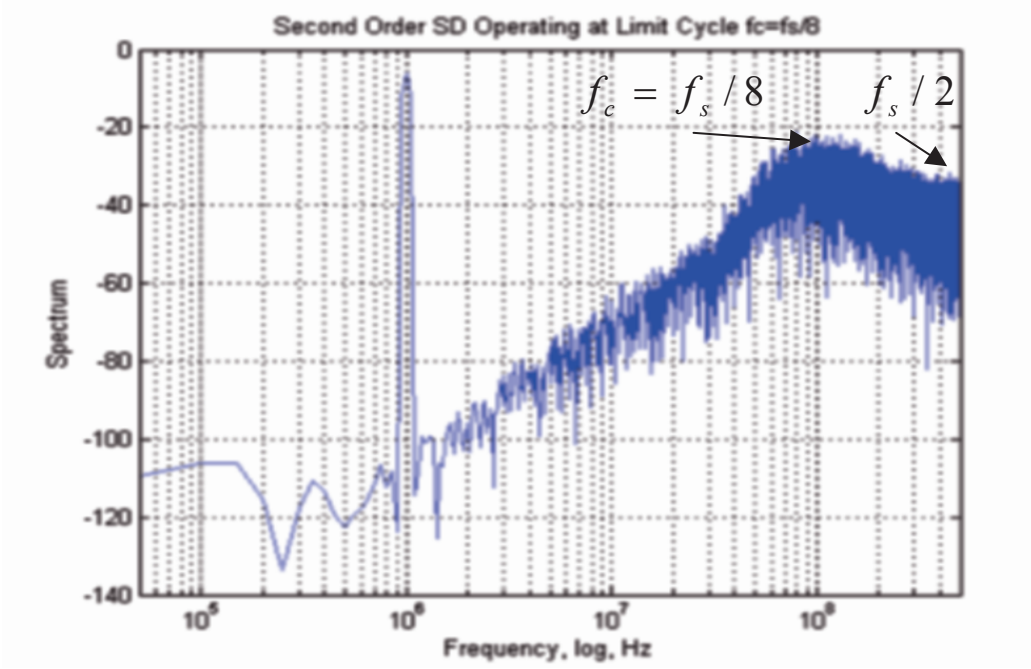


Fig. 9. Simulated output spectrum for 50% modulation depth.

6.2. Advantages of Sub-Harmonic Operation

In section 6.1 the performance of the SDMs that are designed to operate at a sub-harmonic limit cycle was compared with respect to the standard design approach. The same filter function was used in the evaluation. However, the limit cycle model allows an optimisation of the loop filter transfer characteristics with respect to desired limit cycle behaviour and performance. In the second order system the filter characteristics are manipulated via the frequency position of the zero. When the zero is shifted to a higher

frequency, the system displays the highest SNR and DR. That is due to the fact that the loop filter provides most of the phase rotation needed for entering limit cycle oscillation. Equivalently, that means that the quantization noise in the signal band of interest is more suppressed by the loop filter. A further increase of the zero frequency makes that effect stronger, but, however, also decreases the phase margin for the next (with lower frequency) sub-harmonic limit cycle.

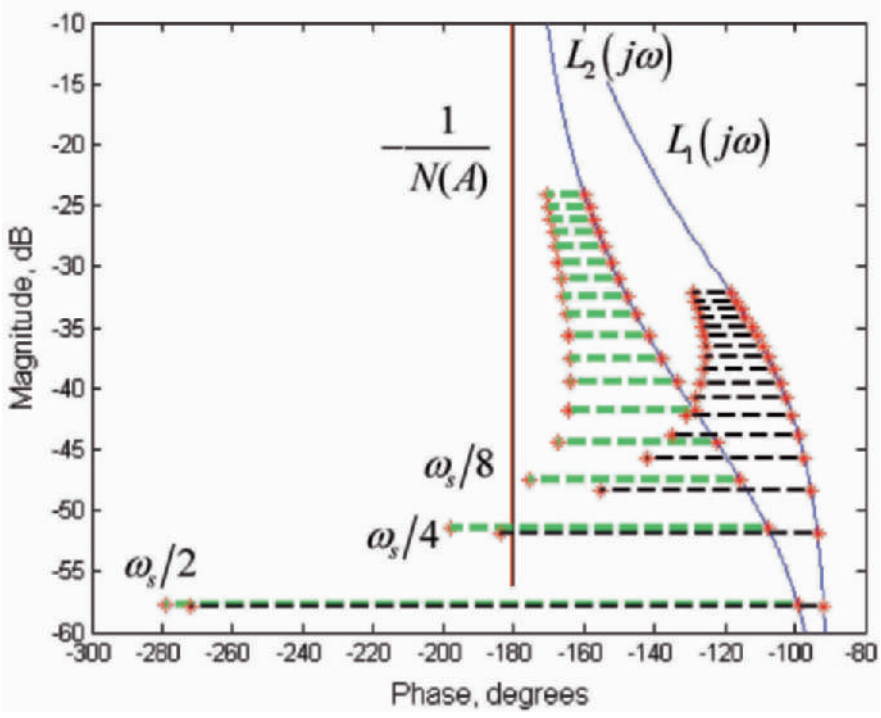


Fig. 10. Filter optimisation for sub-harmonic limit cycle operation.

The possibility to apply a stronger filter function is illustrated in Fig. 10. Two second order filters: $L_1(j\omega)$ and $L_2(j\omega)$ of type (6) are compared that differ only with respect to the frequency position of the zero, such that in $L_2(j\omega)$ the zero is $\omega_{z,2} = 10\omega_{z,1}$. This choice improves the noise shaping with 4dB, however, it also decreases the phase margin to the next sub-harmonic limit cycle with frequency $f_s/8$ from 20 to 4 degrees. The trade-off: performance-phase margin can be optimally resolved for a particular technology spread.

One of the reasons for the deviation of the measurement results from the available methods for prediction is the susceptibility of higher order SDMs to enter unpredicted limit cycles. Moreover, those may have a strong deteriorating effect only for certain input amplitudes or clock speeds. On top of that, when the SDM operates at full clock speed it is very sensitive to excessive loop delays and clock imperfections. In that respect, the proposed sub-harmonic limit cycle operation has several important advantages:

- The GBW that is required for the linear part of the loop corresponds to that required for a SDM operating at a fraction of the clock speed. This leads to a simpler implementation and a significant reduction of the power consumption.
- The loop delay is taken into account in the determination of the limit cycle modes and is made a part of the loop mechanism. Higher extra delay can be tolerated.
- The proposed mode of operation introduces a trade-off between conversion bandwidth, limit cycle frequency and clock frequency and thus gives an additional degree of freedom in the design of SDM.

7. Hardware Implementations

Prototypes of a second-order ASDM and a second-order LCSDM are implemented in a digital 0.18μm 1.8V CMOS process. A differential implementation of the second-order ASDM is described with the block diagram shown in Fig. 11. The stages g_{m1} and g_{m2} are transconductors (voltage-to-current converters) that, together with the capacitances C_{int1} and C_{int2}, implement continuous-time integrators. The blocks FB_1 and FB_2 represent the feedback transfer. In practice they are implemented as switched-current sources that are controlled by the output signal. The blocks A_1 and A_2 are linear gain stages. Their purpose is to decrease the effective hysteresis value and reduce the design requirements for the quantizer with respect to speed and power consumption. A binary quantizer with hysteresis completes the ASDM loop. A detailed description of the measured ASDM performance is described in [21]. Here an example for the measured output spectrum is given in Fig. 14.

The hardware implementation of the LCSDM is based on the same building blocks. A sampling latch is placed after the quantizer Fig. 12, that digitises the two-level signal provided by the asynchronous quantizer. The position of the sampler is attractive from an implementation point of view because the sampler acts on a well-defined two level signal. In this way, the probability for metastability in the sampling can be significantly decreased because the time duration of the weak signal provided to the sampling switch

is minimized. The sampling delay, the switching and the propagation times in the quantizer and the sampling switch are incorporated in the limit cycle mechanism.

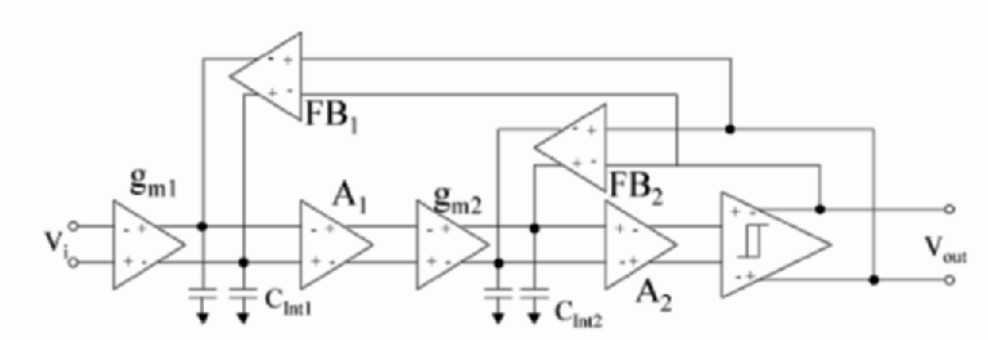


Fig. 11. Block diagram of the transistor implementation of a second order ASDM.

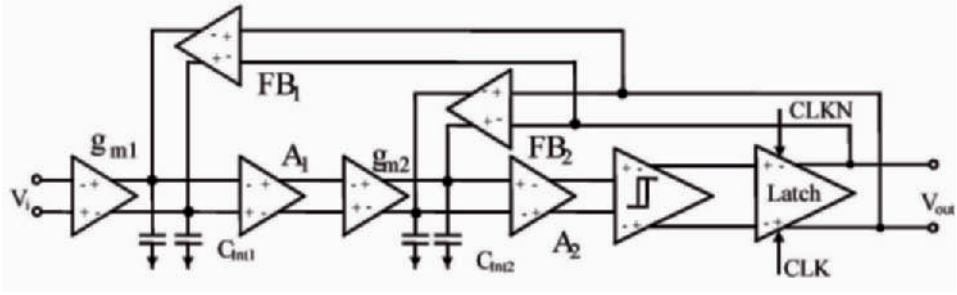


Fig. 12. Block diagram of the transistor implementation of a second order LCSDM.

A critical block in the implementations of both asynchronous and synchronous SDMs is the input voltage-to-current converter gm_1 as it determines the linearity and the noise performance of the modulators and the achievable DR. A detailed description of the used implementation can be found in [25]. Here, in Fig. 13, an example of its SFDR is given, measured for the ASDM implementation.

The second-order LCSDM was designed to operate with a lowest sub-harmonic limit cycle of 125MHz when sampled with 1GHz clock.

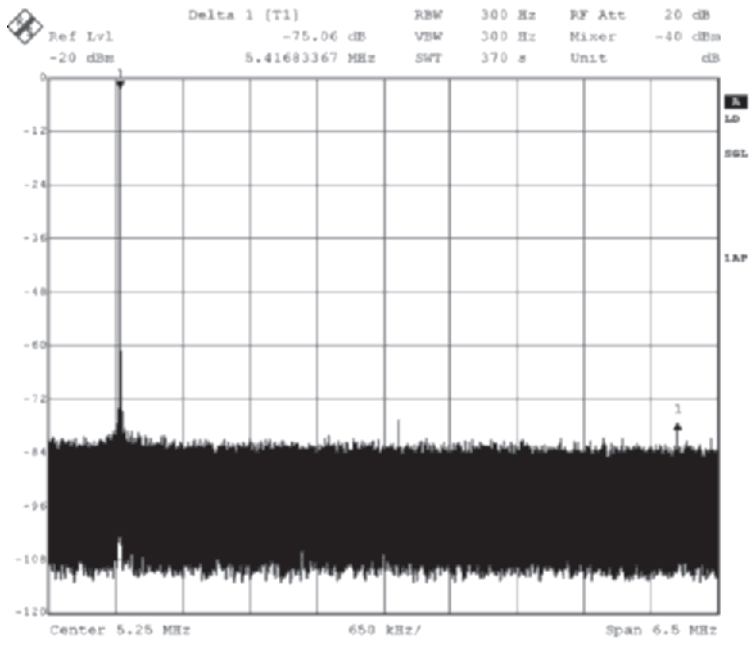


Fig. 13. SFDR for ASDM measured at 80% Modulation Depth, for an input frequency of

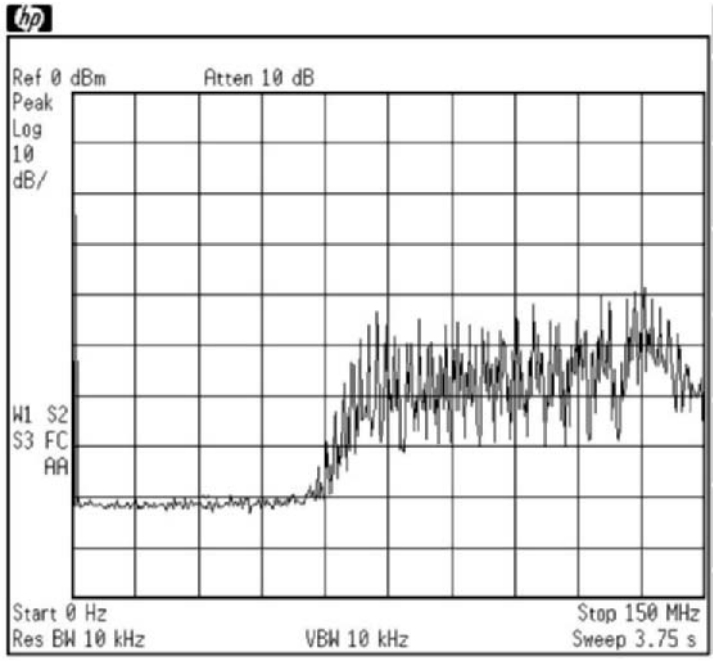


Fig. 14. Measured output spectrum of the second-order ASMD for a 1MHz test signal and a modulation depth of 50%.

In Fig. 15, a transistor simulation of the implemented LCSDM is shown. The output spectrum for an idle operation with limit cycle $f_c = f_s/4$ is compared with the spectrum with an input signal. The simulated behaviour matches the prediction of the model and the Matlab simulations.

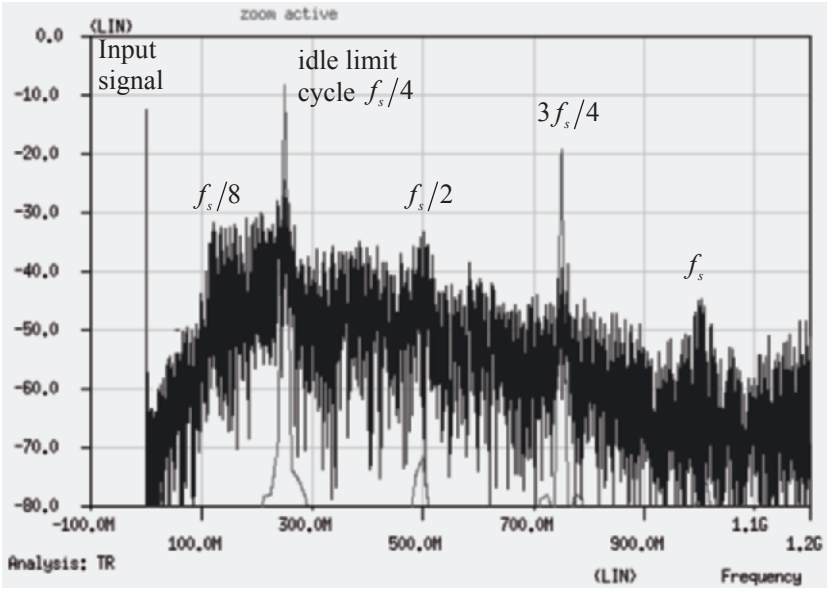


Fig. 15. Idle limit cycle with $f_c = f_s/4$ *and busy operation with lowest limit cycle frequency of* $f_c = f_s/8$.

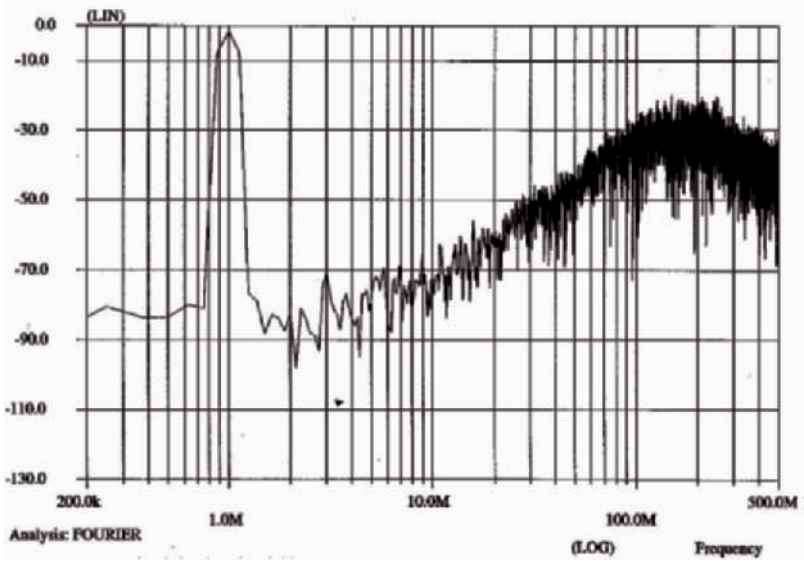


Fig. 16. Transistor simulation of the output spectrum.

In Fig. 16 the simulated output spectrum of the transistor implementation for 1MHz sine input signal is illustrated. A SNDR of 64dB is achieved in a bandwidth of 5MHz. This result matches exactly the expectation given in the last column of Table 2.

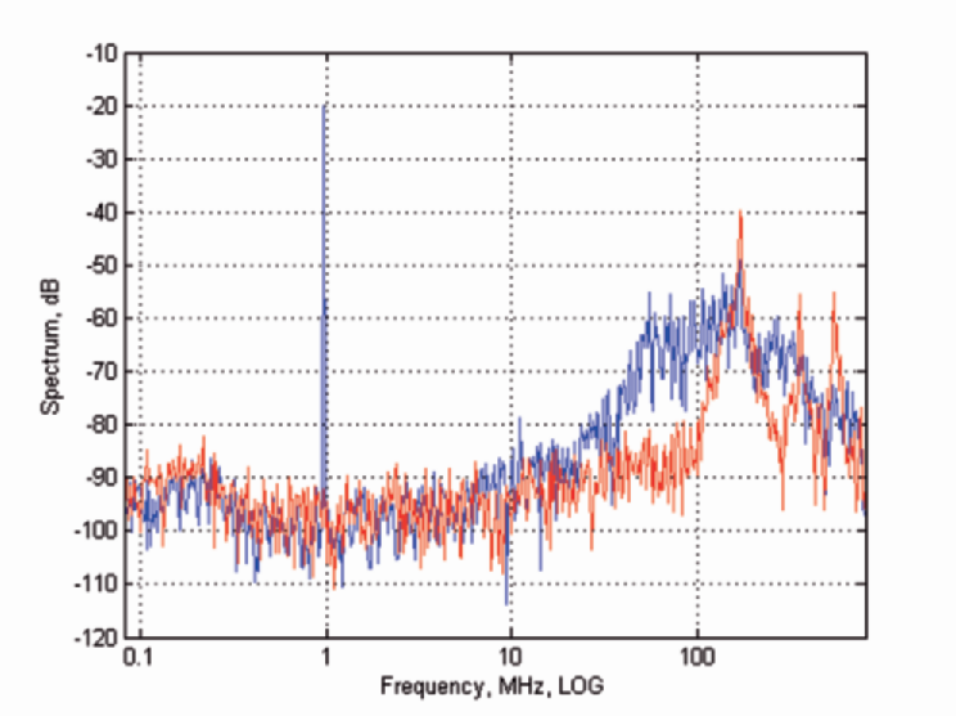


Fig. 17. Measured output spectra of the second- order LCSMD with 1GHz clock and an idle limit cycle of 125MHz and for -6dB FS input signal.

In Fig. 17, the idle output spectrum is compared with the output spectrum for busy operation. The measured IC exhibits an idle limit cycle of 125MHz or $f_c = f_s/8$. This idle limit cycle frequency is twice lower than the intended one: $f_c = f_s/4$. Respectively, in active operation the system enters even lower limit cycles. The lowest measured limit cycle has frequency of 83MHz that corresponds to $f_c = f_s/12$. The measured properties are summarized in Table 3. The measured SNDR of 55dB matches the performance expectations for $f_c = f_s/12$. This result can be compared to the designs in Table 1 with the help of the following figure of merit (FOM):

$$FOM = \frac{P}{2^{ENOB} 2BW} \tag{13}$$

Lower FOM is an indication of a better power efficiency.

Table 3. Summary of the measured performance.

0.18µm, 1.8V CMOS	LCSDM
Area total (core), mm^2	0.068(0.05)
Limit cycle frequency, MHz	83
Clock frequency, MHz	1000
Signal Bandwidth, MHz	5
SNDR [10k-5MHz], dB	55
Current consumption (core)	2.1mA

Table 4. Performance comparison.

FOM, *pJ/conv.*	[1]	[2]	[3]	[4]	[5]	[6]	This work
	7.4	0.8	470	6	1	0.4	0.8

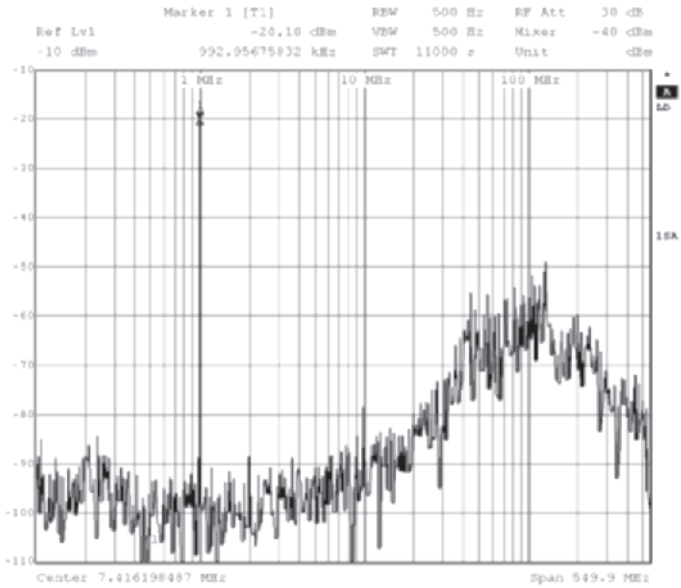


Fig. 18. Measured output spectrum of a LCSDM for –6dB FS input signal.

In Fig. 18 the measured output spectrum of a LCSDM driven with 1MHz sinusoidal input signal is shown. The spectral bump that corresponds to idle limit cycle of $f_c = f_s/8$ is clearly visible.

8. Conclusions

A new limit cycle model of SDM operation was proposed. It identifies the limit cycles in SDM as a dominant phenomenon that determines the properties and performance of the ADC. The existence of limit cycles in second order asynchronous and synchronous sigma-delta modulators was investigated via a graphical application of the Describing Function method. The determination of the frequency and amplitude of the possible limit cycles was shown. A new, sub-harmonic limit cycle mode for the SDM was investigated. It was shown that this mode of operation leads to a reduction of the total power consumption and significantly reduces the design complexity. The expected behaviour and performance were confirmed with simulations and measurements.

References

[1] V. Peluso, P. Vancorenland, A.M. Marques, M.S.J. Steyaert, and W. Sansen, "A 900-mV low-power A/D converter with 77-dB dynamic range", IEEE JSSC, vol. 33, pp. 1887–1897, Dec. 1998.

[2] Libin Yao, M. Steyaert, and W. Sansen, "A 1-V 140-W 88-dB Audio Sigma-Delta Modulator in 90-nm CMOS", IEEE JSSC, vol. 39, No. 11, pp. 1809–1818, Nov. 2004.

[3] Simona Brigati, "A Fourth-Order Single-Bit Switched-Capacitor ΔΣ Modulator for Distributed Sensor Applications", IEEE Transactions on Instrumentation and Measurement, vol. 53, No. 2, pp. 266–270, April 2004.

[4] A. Das, R. Hezar, R. Byrd, G. Gomez, B. Haroun, "A 4th-order 86dB CT ΔΣ ADC with Two Amplifiers in 90nm CMOS", ISSCC, 2005, pp. 496–612.

[5] E.J. van der Zwan, K. Philips, and C.A.A. Bastiaansen, "A 10.7-MHz IF-to-baseband A/D conversion system for AM/FM radio receivers," IEEE JSSC, vol. 35, pp. 1810–1819, Dec. 2000.

[6] K. Philips, "A 4.4 mW 76 dB SD ADC for Bluetooth receivers", ISSCC, Feb. 2003, pp. 64–65.

[7] R.H.M. van Veldhoven, "A triple-mode continuous-time ΔΣ modulator with switched-capacitor feedback DAC for a GSM-EDGE/-CDMA200/UMTS receiver", IEEE JSSC, vol. 38, Dec. 2003, pp. 2069–2076.

[8] L. Luh, J. Choma Jr., J. Drapper, “A 400 MHz 5th Order Continuous-Time Switched-Current Sigma Delta Modulator”, ESSCIRC 2000.
[9] M.W. Hauser, “Principles of oversampling A/D conversion,” J. Audio Eng. Soc., vol. 39, pp. 3–26, Jan./Feb. 1991.
[10] S. Norsworthy, R. Schreier, and G. Temes, Eds., Delta-Sigma Data Converters: Theory, Design and Simulation. New York: IEEE, 1997.
[11] R. Schreier. “The Delta-Sigma Toolbox version 7.1”, Dec. 11, 2004, Available at: http://www.mathworks.com/matlabcentral/fileexchange/-index.jsp
[12] D. Reefman, J. Reiss, E. Janssen, M. Sandler, “Description of Limit Cycles in Sigma-Delta Modulators”, IEEE TCAS I, Vol. 52, June 2005, pp. 1211–1223.
[13] B.C. Kuo, “Analysis and Synthesis of Sampled-Data Control Systems”, Prentice Hall, 1963.
[14] J.C. Gille, M.J. Pelegrin, P. Decaulne, “Feedback Control Systems, Analysis, Synthesis and Design”, McGraw-Hill, 1959.
[15] D.P. Atherthon, “Nonlinear Control Engineering”, VNR, 1982.
[16] A. Gelb, W.V. Velde, “Multiple-Input Describing Functions and Non-linear system design”, McGraw-Hill, 1968.
[17] H.C. Torng, W.E. Meserve, “Determination of Periodic Modes in Relay Servomechanisms Employing Sampled Data”, IRE Transation on Automatic Control, pp. 425–440, September 1961.
[18] E. Roza, “Analog-to-Digital Conversion via Duty-Cycle Modulation,” IEEE TCAS-II: Analog and Digital Signal Processing, vol. 44, No. 11, pp. 907–914, 1997.
[19] S. Ouzounov, E. Roza, H. Hegt, G. van der Weide, A. van Roermund, “An 8MHz, 72 dB SFDR Asynchronous Sigma-Delta Modulator with 1.5mW Power Dissipation”, Proc. Symp. on VLSI Circuits 2004, pp. 88–91.
[20] S. Ouzounov, E. Roza, H. Hegt, G. van der Weide, A. van Roermund, “Design of High-Performance Asynchronous Sigma Delta Modulators with a Binary Quantizer with Hysteresis”, Proc. CICC 2004, pp. 181–184.
[21] S. Ouzounov, E. Roza, H. Hegt, G. van der Weide, A. van Roermund, “Analysis and Design of High-Performance Asynchronous Sigma-Delta Modulators with a Binary Quantizer”, IEEE JSSC, vol. 41, No. 3, pp. 588–596, March 2006.
[22] J.A. Cherry, W.M. Snelgrove, “Excess Loop Delay in Continuous-Time Delta-Sigma Modulators”, IEEE TCAS-II: Analog and Digital Signal Processing, vol. 46, No. 4, pp. 376–389, 1999.
[23] E. Roza, “Poly-Phase Sigma-Delta Modulation,” IEEE TCAS-II:Analog and Digital Signal Processing, vol. 44, No. 11, pp. 915–923, 1997.

[24] J.E. Iwersen, “Calculated quantizing noise of single-integration delta modulation coders,” Bell Syst. Tech. J., vol. 48, p. 2359, Sept. 1969.
[25] S. Ouzounov, E. Roza, H. Hegt, G. van der Weide, A. van Roermund, “A CMOS V-I converter with 75dB SFDR and 360mW power consumption”, IEEE JSSC, vol. 40, No. 7, pp. 1527–1532, July 2005.

Part II: Automotive Electronics: EMC issues

Automotive electronics are very demanding due to their very high safety and reliability requirements and the harsh environment in which they operate. The main design challenges posed by this harsh environment translate in a high voltage supply with high power interferences, a high ambient temperature and high electromagnetic (EMC) and electrostatic (ESD) interferences. EMC in particular is a difficult topic to cope with during IC design. Where design for EMC has long been regarded as an art rather than a science, this situation has changed in recent years and this is discussed in the next chapter.

The radiated and conducted EMC of an automotive module is a function of the behavior of the chip, the package and the PCB subsystem and on their interactions. Analog simulation of such complete system, including all large-signal RF noise sources and coupling paths is clearly not feasible. The first paper by Thomas Steinecke discusses a design and verification flow for EM emission. The complexity is reduced by splitting the problem in three software tools: EXPO targets the IC-feasibility phase, NEMO generates dynamic current profiles of digital sub-modules and XcitePI is a sign-off simulation of the complete module.

Injection of large power EMC signals on a chip input and high voltage electrical transients, conducted on the supply lines or coupled to an input generate large substrate currents in junction-isolated technologies. The resulting design challenge is similar to controlling the substrate currents generated by switching inductive loads. Michael Schenkel describes in his paper the two types of substrate current and their effects in relation to technology variations and technological, layout and circuit protection strategies.

The next three papers address various aspects of EM immunity. The first paper by Franco Fiori analyzes the distortion generated by EMC signals on the input differential pair of an operational amplifier and on MOS switches in Switched Capacitor circuits. A double differential input pair is presented, which largely improves the immunity of the operational amplifier while a non-linear model is derived for the MOS switch to predict the demodulation of the EMC input signal. The combination of the opamp model with the MOS switch model allows an accurate prediction of the EMC behavior of a complete SC circuit.

Derek Bernardon, in his paper, discusses the DPI test on various building blocks. First, the current flow in the package and the coupling effects to the substrate of the RF power signals, generated in the DPI test, are evaluated. Based on the understanding of these effects, circuit improvements are devised for a comparator on a smart power chip. In a second example,

different circuit concepts for an internal voltage regulator are compared and validated. The last example shows how this understanding is used to further optimize the electromagnetic immunity of an existing bandgap circuit.

In the last of the three papers on EM immunity, Aarnout Wieers proposes a structured methodology to evaluate the immunity of analog continuous time circuits. Instead of very time consuming transient analyses for many frequencies in the wide EMC frequency range, a methodology, based on AC analysis and large-signal node impedance is proposed. This methodology is first explained on a simple emitter follower and then applied on a more complex CAN receiver with a resonance in the cabling and the PCB attached to it. The last example is a current reference, based on an, EMC sensitive, external resistor.

Finally, the last paper of this chapter proposes a new approach to extend IC emission modeling into the EMC domain. Mart Coenen first discusses the three on-chip noise sources and their different modeling methods. The application with its impedances and transfer function is the next model component and the measurement setup and signal interpretation is the last component of the model. The proposed modeling concept resolves most of the present EMC emission modeling issues and is currently under development and discussion in international standardization bodies.

MODELING AND VERIFICATION TECHNIQUES TO ENSURE SYSTEM-WIDE ELECTROMAGNETIC RELIABILITY

Thomas Steinecke, Infineon Technologies, Munich, Germany
thomas.steinecke@infineon.com

Abstract

Electromagnetic emission needs to be considered in the design flow of automotive ICs. Complex ICs require on one hand early forecasts and accurate sign-offs for EMI, on the other hand their EMI modeling and simulation is very complex. Approaches and solutions are presented in this paper.

1. Introduction

High frequency electromagnetic emission (EMI) observed in automotive electronic control units (ECU) is often originated by dynamic switching currents of switching logic on a silicon chip, e.g. microcontroller. To help designing low-EMI ICs and low-EMI printed circuit boards, the RF noise sources and coupling paths need to be described by electrical models and simulated by analog engines like Spice. However, complex circuits consisting of millions of transistors set natural limits for detailed models due to data space and simulation time. The need for simplified and nevertheless accurate dynamic current profiles becomes evident. A typical EMC design and verification flow for ICs must start at an early design phase of chip and package to run feasibility studies finding most appropriate placement and routing configurations on chip and in the package.

As soon as detailed layouts of chip and package are available, an "EMC sign-off procedure" has to be performed, based on accurate layout and netlist simulations. This takes place before the complete IC design is released for fabrication.

This paper describes the techniques and tools related to the EMC design and verification flow, discussing the expected accuracies.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 129–150.

2. EMC Design and Verification Flow for Integrated Circuits

To understand the purpose of the involved software tools it is helpful to divide the ECU into chip, IC-package and PCB subsystems. Talking about electromagnetic emission, noise sources are located on the chip and the RF noise is propagated through on-chip, package and PCB traces. From there, the RF is either conducted into connected cables of a car supply and communication net or radiated by means of PCB traces acting as antennas into the environment. Talking about electromagnetic immunity, the noise sources are located in the environment of the ECU and propagated through PCB, package and on-chip traces, where the noise finally penetrates transistors, possibly causing malfunction. Figure 1 illustrates the two possible noise propagation directions. This paper will discuss only the emission direction. Immunity modelling is a future activity.

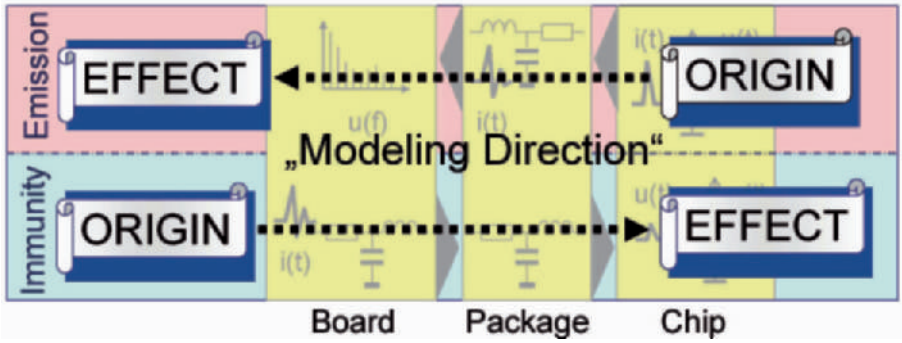


Fig. 1. EMC Modelling Directions.

Figure 2 shows the EMC design and verification flow as it is installed at the automotive microcontroller development group of Infineon Technologies. It consists basically of 3 software tools:

- “EXPO” (Expert system of Power supply) targets to support EMC design feasibility studies of ICs before starting the design phase.
- “XcitePI” (Power integrity simulation) is used for netlist/layout-based EMC sign-off simulations prior to production start.
- “NEMO” (Netlist-based Emission Modeller) creates the dynamic current profiles of digital modules. These models are required by EXPO and XcitePI for proper noise simulation.

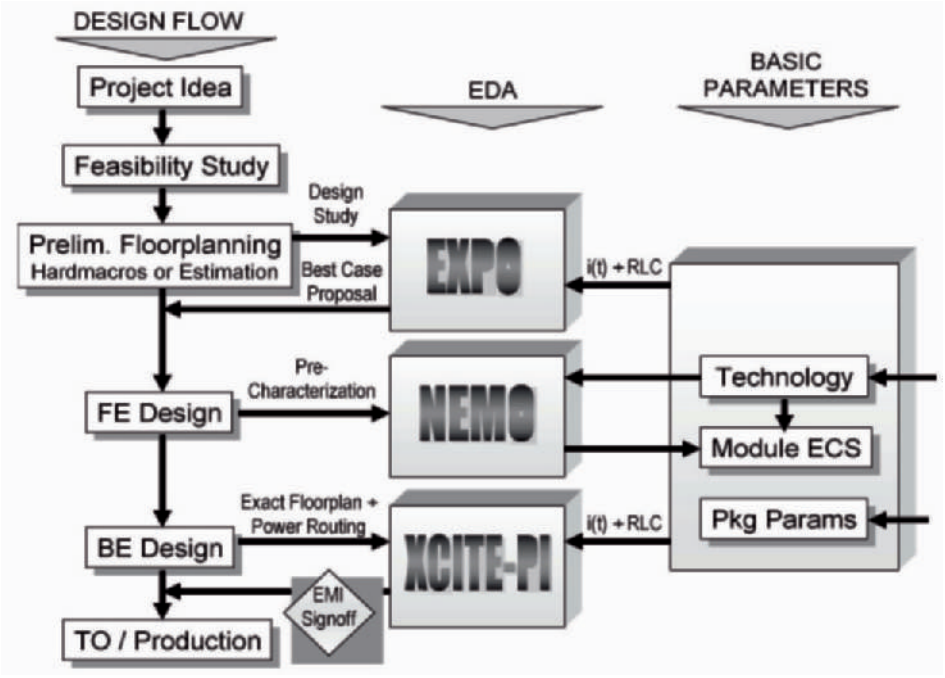


Fig. 2. EMC Simulation Flow for Ics.

3. EXPO: Feasibility Studies for EMC Design Optimization

EXPO allows fast changes in module placement, module size and power connections. It does not rely on pattern simulations, but uses pre-characterized typical or worst-case dynamic current models of all types of modules (e.g. digital, analog, flash memory, SRAM, I/O) instead. The software pilot design and first test case was supported by the Medea+/BMBF funding project MESDIE, Fkz 01 M 3061 H [1].

The user is guided by a graphical Windows interface through the entry procedures of on-chip layout and connectivity data and also package layout and connectivity data.

Figure 3a shows the floorplan configuration window for on-chip design data. The whole chip area is divided into a selectable number of equally distributed tiles. Every tile is assigned either a pre-characterized module type as described above or a pre-characterized power supply segment like capacitive buffered VDD-VSS traces. In addition, any tile can be assigned a power domain connection, thus representing e.g. a power grid.

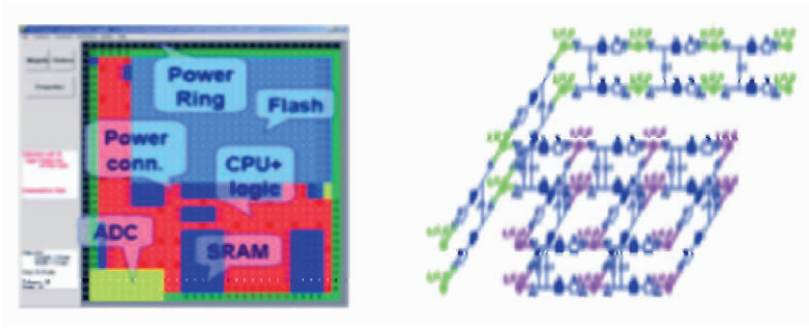


Fig. 3. (a) EXPO on-chip design data entry; (b) resulting netlist.

Figure 3b shows the structure of the resulting on-chip power supply netlist. The netlist consists of specific dynamic current profiles for all types of functional blocks like digital logic, flash memory, SRAM, ADC, etc. There is exactly one mesh segment for every tile of the chip in the netlist. According to the tile size, the module type specific current profile is scaled down, resulting in an equal current distribution over the whole module. Additionally, the on-chip substrate coupling is also reflected by the netlist.

These current sources are connected through RLC networks to the on-chip power supply network. Shape and amplitude of the peak is ideally derived from analog simulation of the chip. For very complex chips, analog transistor simulations cannot be handled anymore. In this case, analytic or behavioural models are required to handle simulation time and data size. This analytic modelling is still under development; the current status of automatic current profiling from digital netlists is described in paragraph 5.2. Meanwhile an alternate current profiling approach from measurement is applied. Providing a pre-characterized dynamic current pulse of a "reference module" in a certain technology, this current pulse can be scaled by measuring the average DC current which equals the integral of the current peak. Current shape remains unchanged, and peak amplitude is adapted according the measured DC current. A sample current peak is shown in Fig. 4; I_{noise} is the measured average current. It is either derived from simulations or from measurements. In the case of measurements, the amplitude is calculated from the measured DC current, and the timing is calculated from technology-specific gate propagation delays.

After input of the chip layout data, the package design data must be added. Figure 5a shows the configuration window for package design data and Fig. 5b shows the structure of the resulting package power supply netlist. Through naming conventions, all on-chip power pads are connected to their corresponding bond wires which are part of the package model.

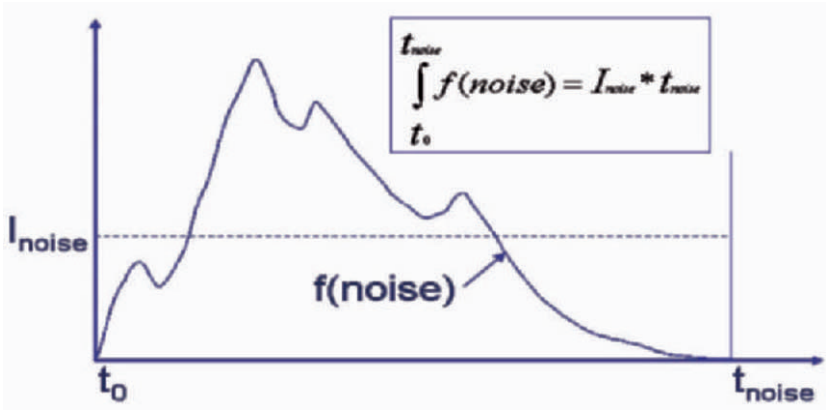


Fig. 4. Current peak used for module profiling.

EXPO provides the user with a selection table for the parasitic effects to be simulated. These are: SSN, EMI (emission), crosstalk between power domains, ESD (discharge paths). The complete netlist including the simulation directives is then executed by HSpice. The results can be observed in EXPO and overlaid with measurement curves to prove their correlation and thus the quality of the EXPO chip/package models.

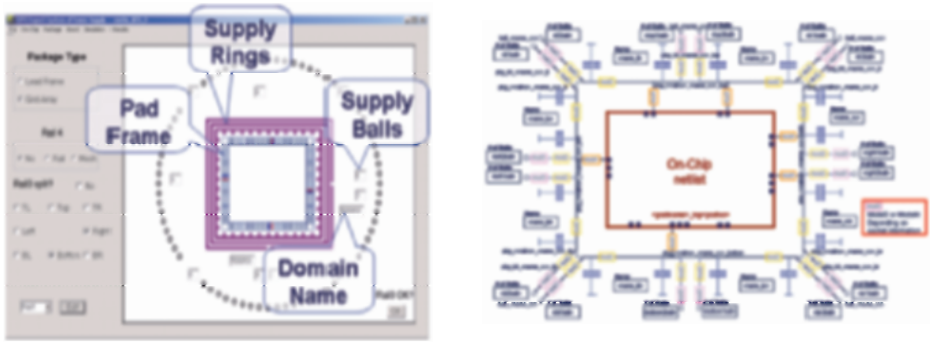


Fig. 5. (a) EXPO package design data entry; (b) resulting netlist.

Complete modelling has been done for the 32-bit microcontrollers TC1796 and TC1766. The following case studies introduce the modelling and simulation capabilities supported by EXPO.

3.1. Board Impedance

Figure 6 shows the comparison of impedance for a PCB power domain between measurement and simulation. The simulation result matches quite well to the measurement. There is a small difference on the resonance

frequency at about 400MHz and a further mismatch above 800MHz. The comparisons of impedance for all other powers as well as grounds show similar results.

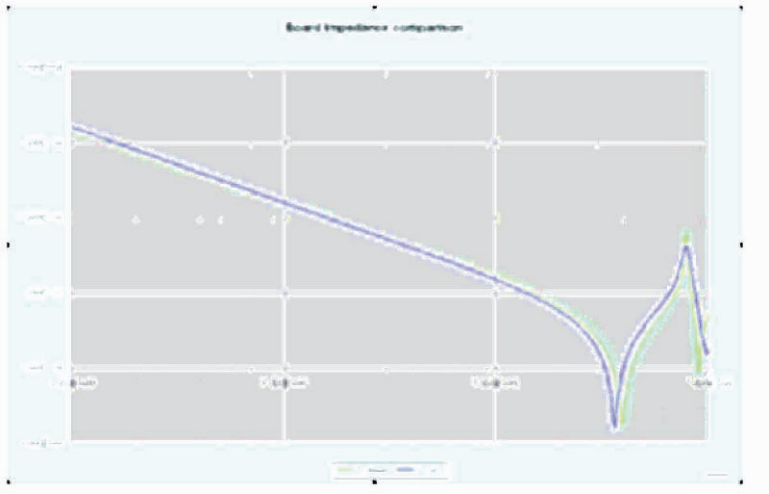


Fig. 6. Simulated and measured test board impedance.

3.2. VDE Standard Simulation

Figure 8 shows a simulation result of conducted emission based on the VDE 150Ohm (IEC 61967-4) standard. In this case the power supply was simulated and measured where the noise sources are connected. The noise sources simulate the active level of the core which runs with a frequency of 150MHz. Figure 7 indicates that EXPO simulation and conducted emission measurement are in good agreement. The overall difference stays below 5dBµV which is in the range of measurement inaccuracy.

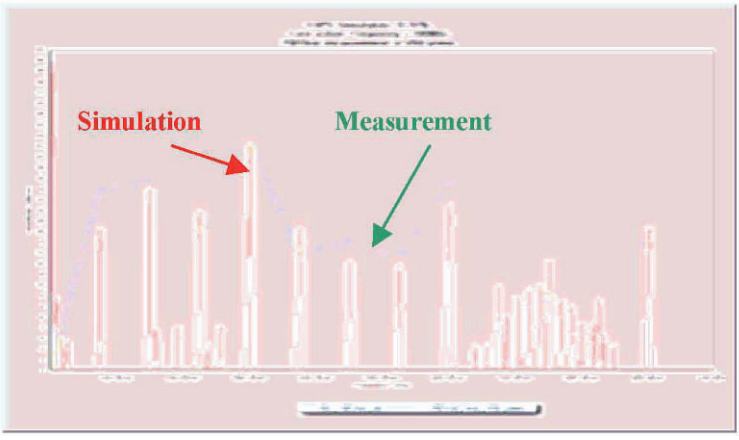


Fig. 7. Simulated and measured conducted emission.

A good agreement is seen up to 600MHz. Depending on the inaccurate modelling of the board impedance above 800MHz an increased drift between simulation and measurement can be observed in this range.

3.3. Power Crosstalk Simulation

Power crosstalk means that a noisy power domain can disturb another power domain. If several different power domains are used on a design, power noise crosstalk has to be considered for board as well as package design. There are two coupling paths between the power domains. On the one hand side, the noise is transmitted via electromagnetic fields. And on the other hand, the noise is coupled via the impedance of the package and board. Figure 8 shows a simulation example from EXPO, indicating the crosstalk from the TC1796 digital core supply to the external memory bus supply (VDDE). Microcontroller measurements focus on power noise crosstalk according the "Generic IC Test Specification for EMC" ("BISS Paper") [2].

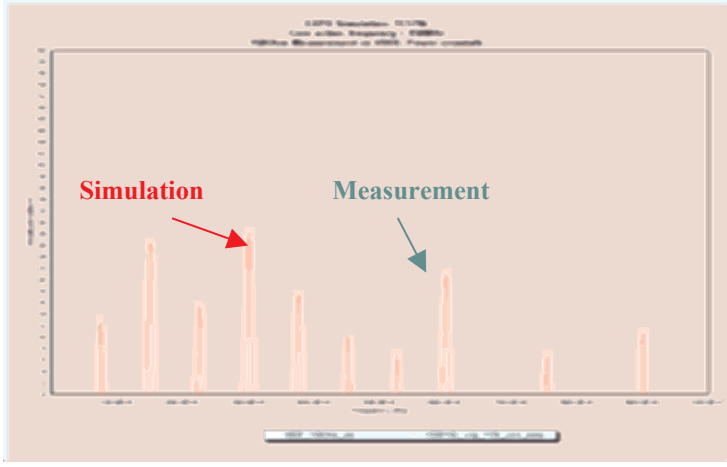


Fig. 8. Simulated and measured power noise crosstalk.

3.4. SAE Standard Simulation

Additionally, the tool EXPO allows the user to analyze the chip with package and board design for radiated emission. Thus the electromagnetic emission according the SAE standard J1752/3 (IEC 61967-2) can be also simulated.

The SAE measurement is performed by 2 measurements, one measurement with 0 degree and one with 90 degrees test board rotation. The resulted mathematical product has to be calculated from these two measurement vectors.

The difficulty on the simulation is to combine both radiated emission measurements into one simulation. For this reason, the physical design of the chip as well as package must be known. Figure 9 indicates a nice correlation between simulation and measurement of radiated emissions generated by the TC1796.

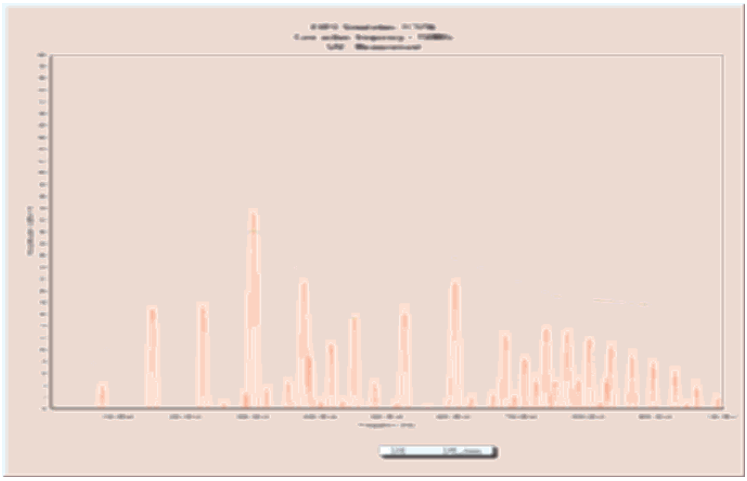


Fig. 9. Simulated and measured radiated emission

4. XcitePI: Accurate Simulation for EMC Design Sign-Off

Electromagnetic emission caused by switching activity of large clocked ICs like microcontrollers is an important quality criterion which may lead to customer rejection of an IC and missed design-wins. Thus it is essential to provide modeling and simulation tools which allow to predict the EMI behaviour of complex ICs before costly masks are generated and silicon is fabricated. In contrast to the tool EXPO, which is intended to be used for design case studies and early power integrity forecasts, the modeling and simulation approach with XcitePI is going to be established as an EMI sign-off process for automotive microcontrollers. In cooperation with the tool vendor Sigrity, the 32-bit microcontroller TC1796 was modelled, and noise distribution on chip and through the package has been simulated. This IC is complex enough to really investigate the noise propagation paths and even self-disturbance of the chip.

Once switching noise is generated at certain locations (circuit domains) on an IC, the noise often propagates more easily to other locations (circuit domains) of the chip through the package power and ground structures rather than directly through the chip power grid itself. This is particularly true for flip-chip packages, but also for wire-bonded chips.

In the case of the Infineon 32-bit microcontroller TC1796, the behaviour of the IC plus its chip package is analyzed to understand the different ways of noise distribution. Figure 10 shows the complexity of this product. The TC1796 consists of approx. 30 million transistors and is mounted in a P-BGA 416 package. The µC contains ca. 10 power supply domains, of which the main supplies are connected to rings on the package substrate. The whole chip/package system is therefore very complex.

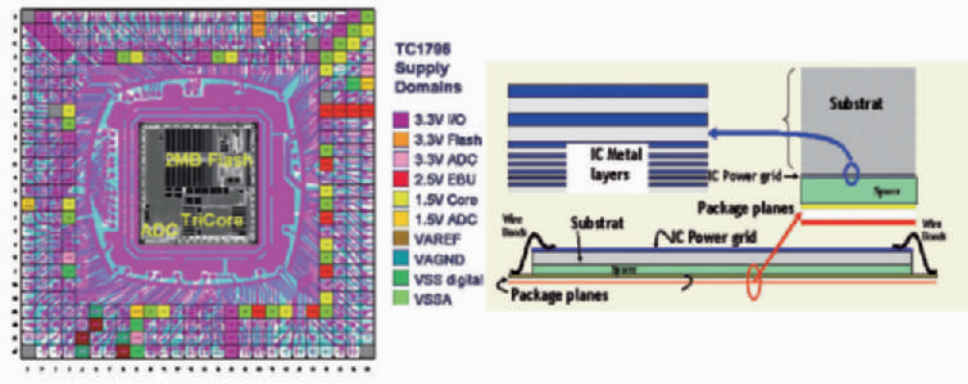


Fig. 10. Silicon chip and BGA package of the TC1796.

The goal is to gain a reliable model for dynamic current flow through the microcontroller's wire bonds/balls. The results have to reflect chip design changes (module placement, power routing) and package design changes (BGA substrate, integrated passives).

An alternate approach for the RLC parasitic extraction of the on-chip power supply system [3] was developed in the Medea+/BMBF funding project MESDIE, Fkz 01 M 3061 H [1].

4.1. Chip and Package Modeling

The switching activities on the microcontroller need to be modelled with the help of high-quality on-chip current sources, which are distributed all over the silicon die as a function of the individual circuit domains. These current sources are connected by the on-chip power supply network, realized by a combination of grid and ring structure. The power network also has to contain the distributed on-chip capacitors and leakage currents. This system is modelled by co-design [4], [7] in a retroactive way, such that current paths through the package are re-entered into the silicon.

Results of this co-simulation are dynamic currents or voltages at various points of interest, e.g. power supply package balls. The dynamic currents and

voltages can afterwards be re-used by PCB simulations considering noise decoupling circuits.

Lumped RLC or simplified S-parameter models would fail to consider off-chip package effects, as they are unable to take into account the distributed interaction between IC power grid and chip package, and are thereby unable to determine the true transient noise distribution and behavior in the power delivery system.

Consequently accurate modeling of the noise behaviour of high-speed ICs therefore necessitates the modeling of distributed electromagnetic wave propagation effects in the package planes and the distributed interactions between IC power grid and package structures, as done in Speed2000 [7],[5] plus Co-Simulator [7] plus XcitePI [6].

This way it is possible to evaluate different decoupling schemes, to simulate the impacts of various distributions and characteristics of the given source excitations. It is also possible to assess the effects of chip packages and one can perform what-if comparisons to optimize electrical performance, power grid physical parameters etc.

4.2. Electrical Description of Microcontroller and BGA

Major goal is a high-quality correlation between simulation results and EMI measurements in the time domain. Once the good quality of the simulation model has been demonstrated, models for new chips and their packages are trusted to deliver realistic simulation results already during the IC design phase without having physical measurement values.

The TC1796 is wire-bonded to a two-layer BGA as shown in Fig. 11.

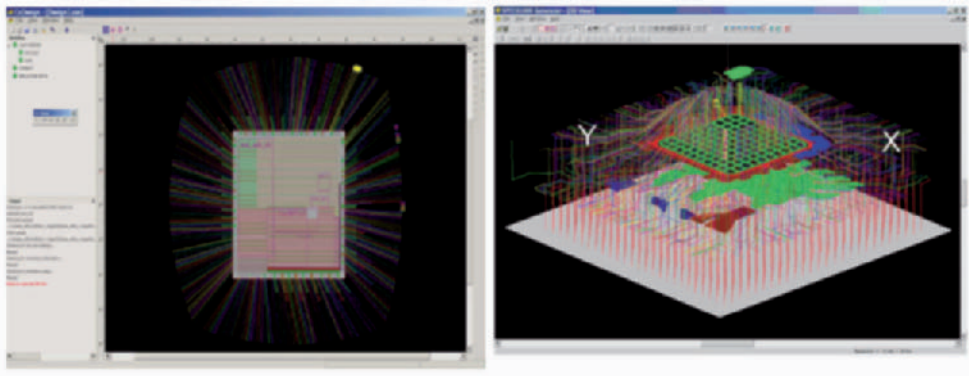


Fig. 11. 416 wirebonds connect the 32 bit microcontroller with its 2-layer BGA.

The power distribution system of the microcontroller is distributed over the 7 metal layers of the CMOS chip plus the power distribution system of the BGA. The on-chip power distribution consists of ring buses, regular

power grids of core logic and on-chip memory and the more or less regular power nets grids in the analog circuit part of the μC.

The power net distribution system of the two-layer BGA connects the μC power net distribution system with corresponding PCB power nets, which have to supply the chip with power within the corresponding total design.

All static and dynamic electrical characteristics of the materials, needed to form the IC and the BGA, are considered in the XcitePI and Speed2000 simulations. The full wave solver algorithms of XcitePI and Speed2000 cover automatically the coupling of all vias and the inter-plane coupling of all involved power patches and power planes. The power trace coupling on chip is covered automatically; the power trace coupling on the BGA with Speed2000 can also be automated on request. The power net grid geometry of the μC is described by tables. The table data can be manually generated in XcitePI or in Microsoft Excel tables or can be read in automatically through LEF/DEF (Cadence Library and Design Exchange Data), defining the microcontroller design. The EXCEL tables can be converted into .xml data. XML (Extensible Markup Language) is the XcitePI input format.

The power integrity (PI) analysis for the microcontroller is performed within XcitePI [6]. The TC1796 has 9 VDD and 4 GND power net systems; the BGA contains 4 VDD and a common GND power net system.

The PI-analysis of the BGA, done with Speed2000, covers the BGA plus the wire-bonds between the BGA and the microcontroller. The BGA geometry can be read into Speed2000 through the APD (Cadence Advanced Package Designer) format *.mcm.

As only the microcontroller and BGA without corresponding PCB were simulated, it is assumed that all BGA power solder balls see the idealized supply voltages of the corresponding voltage regulator modules. Their internal resistance is idealized with 1mOhm.

4.3. Power Integrity (PI) and On-Chip Circuit Switching

As explained above, the switching activities on the microcontroller need to be modelled with the help of high-quality on-chip current sources, which are distributed over the silicon die.

Several current sources are connected to the on-chip power supply network, realized by a combination of grid and ring structure. The power network also contains distributed on-chip capacitors.

The current sources model the switching activities of the individual logic circuit domains on the microcontroller. Thus the overall switching current is determined by the architecture of the microcontroller. The format is SPICE piece-wise linear. The implemented example of such a modelling is shown in

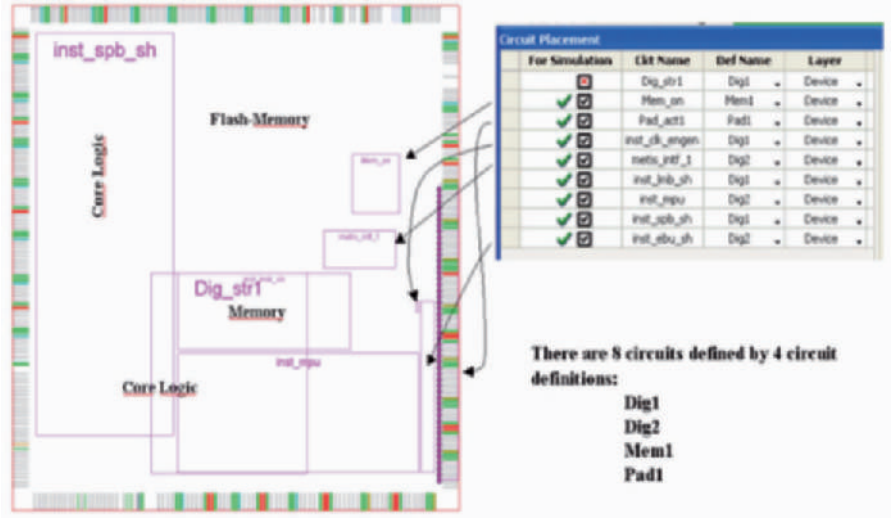


Fig. 12. On chip circuits are characterized by their "current profiles".

4.4. Co-Simulation of Chip and Package

The chip and package current distribution has to be simulated as a function of given circuit domain switching activities. During the input procedure of the co-design analysis, the input to the chip data and the chip-package data can still be modified. The simulation and the run-time parameters are defined here exclusively. Those are e.g. the locations at which the noise voltages or the dynamic current behaviour on the local power distribution is of interest. The noise sources need to be provided by the chip designers. A way of automatic generation of dynamic current profiles for large digital function modules is described in [8].

Figure 13a is an example of a chosen set of view points for a given switching activity on the µC with and without BGA. The circuit signature repetition rate is about 6 ns. From these individual analysis results it can be concluded that the power distribution shows considerable variations over time and that these variations are larger with BGA than without BGA. As minimum it can be concluded that PI simulations of an IC alone do not deliver a complete picture of the power distribution variations on the IC under real world application conditions. Combined with measurements, the shown simulation techniques allow a better insight into the existing power distribution implementation and its characteristics.

As can be seen from Fig. 13b, the power potential variations are considerably higher for microcontroller with BGA, caused by the noise coupling through the package.

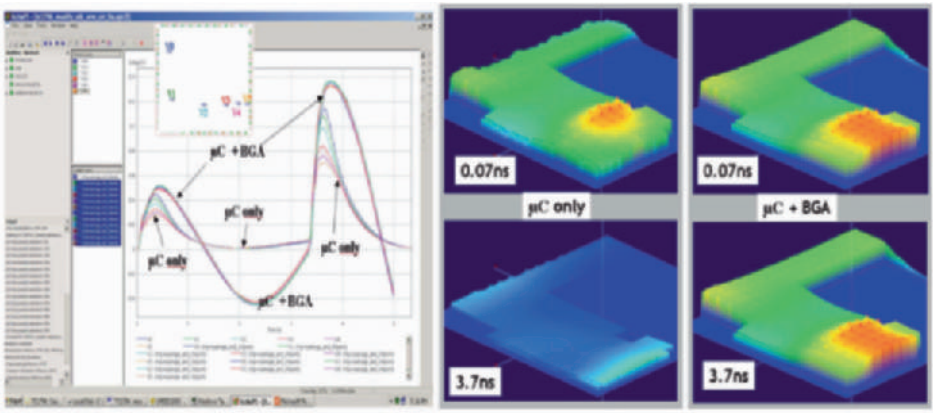


Fig. 13. Power potential distribution as function of the location with and without package.

5. NEMO: Generation of Dynamic Current Profiles

Complex ICs with millions of transistors generating switching current – thus causing electromagnetic emission – are impossible to simulate using analog circuit models and real test patterns. Time and data space limitations demand for alternate modeling strategies. The approach followed by the NEMO tool for modelling digital functional blocks avoids time-consuming pattern simulations. It uses pre-characterized switching models of single gates. Variations of input slope and output load of gates are considered by analytical online calculations. For single gates, mixed gates and flip-flops of a 130nm CMOS standard cell library, these calculations have been approved and applied to regular and irregular functional blocks. The simulation results of these analytical models show good correlation with SPICE simulations using BSIM4 (Berkeley Short-channel IGFET Model) transistor models. This work is being funded by the "Bayerische Forschungsstiftung" project MISEA [10].

The modelling software "NEMO" (Netlist-based Emission MOdels) imports a Verilog gate netlist of a digital function block. The gate interconnections are extracted and used as control parameters for the analytical scaling of the pre-characterized gate current profiles. Gate propagation delays are calculated and used for the superposition of all single-gate switching currents. The result is an equivalent current source (ECS) of the complete digital module. The ECS can be used to create CEM models (Chip Emission Models) to be used for chip floorplan optimization [11] or ICEM models for PCB (Printed Circuit Board) layout optimization as proposed in [9]; see Fig. 14.

Once this software has successfully applied to big circuits and the activity level can be scaled, these ECS are used as noise excitation sources by EXPO and XcitePI.

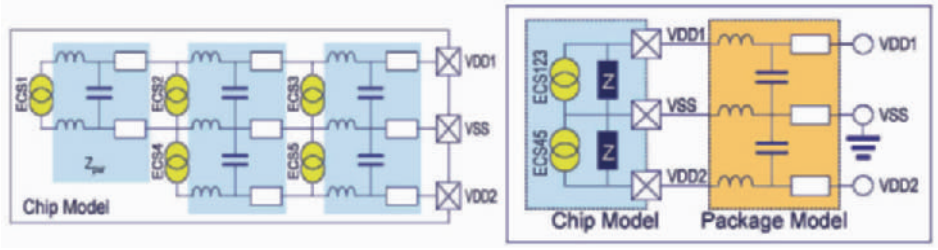


Fig. 14. (a) Chip Emission Model; (b) IC Emission Model.

Since NEMO implementation is still ongoing, this paper can describe only the recent results. Next steps will be (1) the application on huge real standard cell modules as being used in 32-bit microcontrollers, (2) scaling of switching activity, and (3) the automatic generation of CEM models which consider the spatial distribution of large modules.

5.1. Standard Cell Characterization

The dynamic switching current of a cell depends on various parameters. Supposing a consistent power supply and a continuous manufacturing process, it depends finally on the input slope and the output load of the gate. Figure 15 shows as an example the analyzed behaviour of an inverter during characterization. The difference to other types of gates is the number of inputs/outputs and the corresponding input/output slope. Prior to generating module-specific current profiles, a full standard cell library characterization procedure has to be performed. This characterization procedure has to be repeated for every single standard cell which is part of the selected library. We will see in the subsequent explanations that special care has to be taken especially for mixed gates and flip-flops as well as register and RAM macros. However, the deducted set of characteristic parameters applies to all cell types, thus it can be automated. The time spent once on this characterization pays very well for the fast subsequent dynamic current simulations.

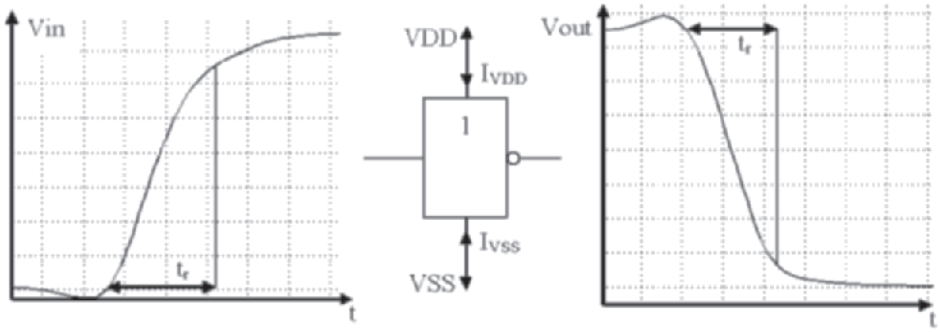


Fig. 15. For characterization, the cell is connected directly to the constant power supply.

The slope of the input voltage is measured on rising edges between zero and 90%, on falling edges between VDD and 10% of VDD. At the beginning of a transition, the voltage has a short overshoot; finally, it converges at the end similar to an exponential function. Therefore the curve progression at the end is unusable, but the first part of the transition is highly important. Especially flip-flops cause in this phase a comparatively slow rise of the output voltage. The positive current direction is appointed from the power supply into the cells for all simulations and graphics.

For determining all effects influencing the input slope exactly, the entire circuit needs to be considered. The output load is also not defined by the subsequent cell only. A compromise solution for reducing the complexity while preserving a good accuracy is to simulate the standard cell with just a minimum of surrounding gates representing the environmental circuit. Another tolerable simplification is to replace the connected gates with so called equivalent inverters. The influence to the input slope and the current profile depending on the output load is almost the same when replacing complex gates with these inverter cells, where the transistor types and sizes have the appropriate specification. In some cases, a serial resistor between the equivalent inverter and the power supply should be added. Such cases apply if the gate connected to the output consists of two or more transistors in series.

Figure 16 shows the finally used circuit including all the above-mentioned simplifications to emulate the surrounding circuit for simulations of single gates in various conditions.

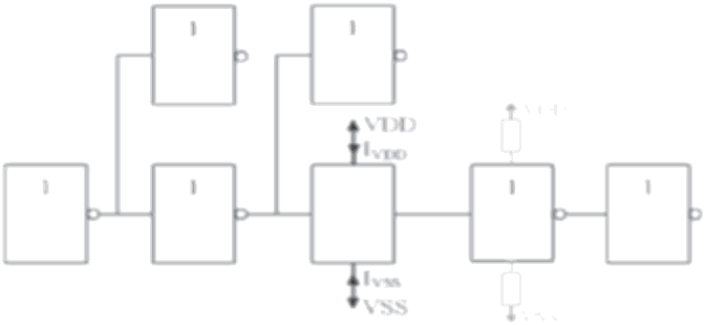


Fig. 16. Schematic of simulation circuit during the characterization process.

The purpose of the two subsequent gates is the simulation of different output loads. Hence the first one is of course the most important gate, it must be modeled preferably well and if necessary with additional resistors. The second equivalent inverter is less significant, so it is sufficient to model it with an adequate approximation. Different input slopes can be adjusted with the preliminary gates. Most important are the transistor sizes of the two gates directly connected to the input of the cell that is to be characterized. The input slope of the previous stage is also important, therefore additional gates are added. By varying the transistor sizes of the gates within this circuit, the results in most cases are covered very precisely, and the accuracy of the characterization in exceptional cases is still a good approximation.

Hence the previously mentioned simulated conditions can be an assortment only; it must be possible to interpolate them. Therefore the current profile characteristics are extracted in the form of significant points. Here, these reference points are maxima, minima and zero points. Because the profile characteristics deviate from each kind of cell, the considered number of maxima and minima between two zero points has to be variable. Figure 4 shows a comparison of three different profiles in several ambient conditions.

Figure 17 (a) shows a simple buffer switching from '0' to '1'. Different conditions cause different amplitudes and delays, but there is no need of more than one maximum between the zero points. Figure 17 (b) shows the current profile caused by a flip-flop with a rising edge of the clock including an output change from '0' to '1'. Comparing the amplitude of the two coexisting maxima between the zero points shows the need of characteristic points associated to each of them, because it is not guaranteed that one of them is the global maximum in all conditions. Fig. 17 (c) shows the profile of a flip-flop with a falling edge on the clock input. Hence it is triggered by rising edges, no output changes may occur and the output load is irrelevant. The shape of the profiles show that in this case three reference points are obvious to characterize the profile most beneficial.

These steps for characterizing a cell are independent from the application, so it can be done prior to the circuit simulation. When saving the results in a library, the characterization process has to be done in ideal case only once per technology. Due to the employment of such a library of pre-characterized cells, most simulation effort is avoided within the real chip simulation. This means, that only a fraction of computation is necessary, leading to an enormous simulation speed-up.

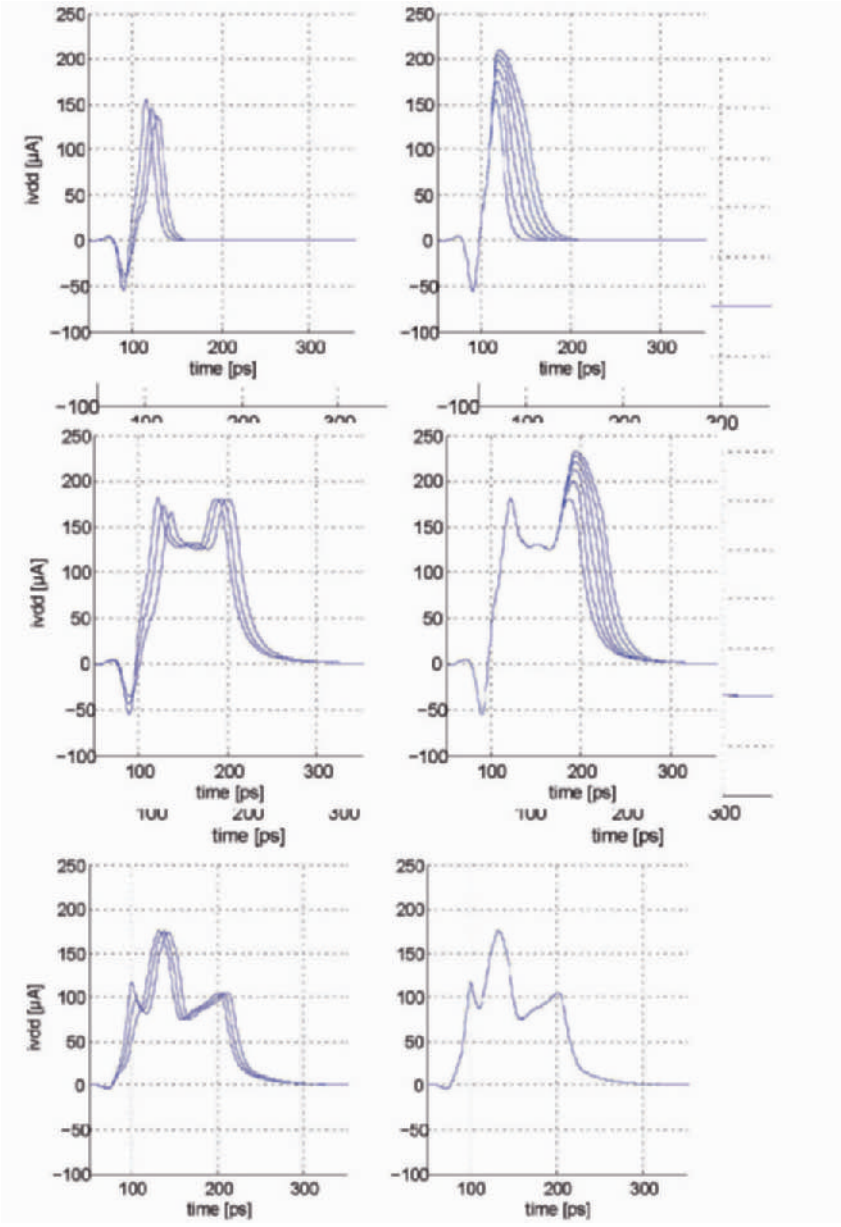


Fig. 17. Comparison of different profile types under several conditions (a)(b)(c).

5.2. Module Equivalent Current Source

An IC module can be modelled as an equivalent current source that represents the dynamic switching currents within one clock cycle. Calculating the current profile of the entire circuit means basically to calculate the profiles for each switching event of the single gates and to sum them up with the corresponding delays.

Figure 18 illustrates the discussed steps for modelling the dynamic switching current of a complete module within this section. It shows the schematic and the extracted paths of the single switching events. The resulting current profiles are shown in Fig. 19.

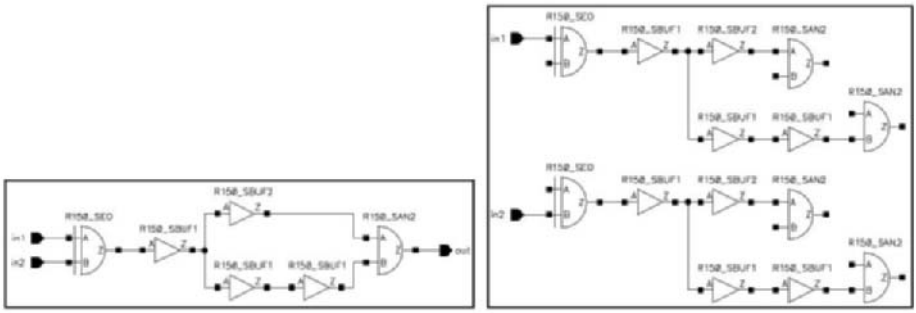


Fig. 18. (a) shows the schematic and (b) the extracted paths of a circuit example.

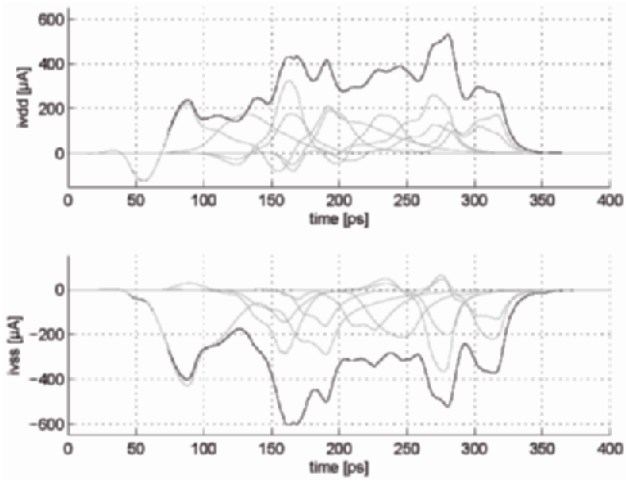


Fig. 19. Switching currents of single gates (gray) and overall current of the module (black).

Hence the module circuit is based on its netlist, it can be divided into paths of consecutively switching gates.

Each path begins with a module input port or a flip-flop followed by the subsequent logic gates. Every single path consists of a chain of consecutively switching gates and ends with either a module output port or an input pin of a flip-flop. Remarkable is, that the figure of extracted paths consists of more gates than the original schematic, because some gates may change their output value more than once. For example in an XOR gate, where both inputs change with a certain delay from low to high, the output state changes twice. The first rising edge causes the output to change from low to high, the second edge back to low. As shown in the example, the subsequent gates must be considered twice within one single clock cycle. Contrariwise, when replacing the XOR by an AND gate, the first rising edge on an input causes no transition of the output and the following gates have to be considered only when both inputs are high. These different conditions have to be taken into account when gathering the switching activities.

The first step for determining the current profile for each switching event is to figure out the input slope as well as the output load of each single gate. These values are essentially required for using the characterized cells discussed in the previous section. The output load as well as the parallel gates and the transistor sizes of the previous gate are given by the circuit specification in terms of its netlist, gate types and extracted paths. The input slope is nothing else than the output slope of the previous gate, but the slope is no parameter within the characterization simulation circuit. Therefore a kind of successive approximation by adapting the transistor sizes of the very first parallel gate (Fig. 3 upper left) is necessary. If all parameters of a switching gate are known, the configuration of the characterization circuit is possible and the calculation of the current profile can be done. This calculation consists actually of two main algorithms. Due to the fact that the cell characterization can be done with a finite number of simulations only, an interpolation of existing configurations is necessary for getting the actual characteristic values like input slope, output slope, propagation delay as well as the minima, maxima and zero points of the current profile. Since there are altogether six parameters for configuring the characterization circuit, the algorithm must be a six-dimensional interpolation as well. A good explanation of the interpolation method including all utilized formulas can be found in [12]. The final tool to obtain the current profile is a scaling algorithm which stretches a base profile of the particular cell to the recently calculated properties.

Taking all these considerations and calculations into account and summing up the scaled profiles to one vector with the appropriate time offset (delay), the result is the current profile of the complete module. This finally obtained profile represents the desired equivalent current source for the module.

5.3. Simulation Results

NEMO provides all the previously mentioned considerations. The results respecting the accuracy are shown with the following example. Figure 20 shows the test circuit schematic that is similar to the previous example in the version with an AND gate.

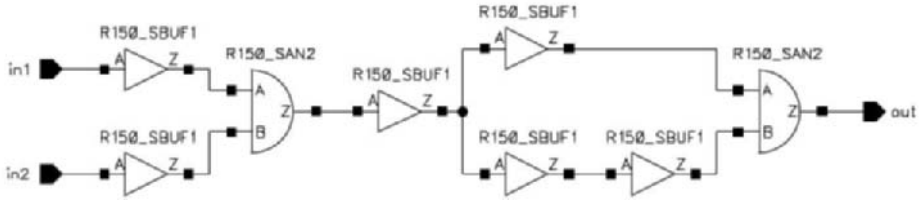


Fig. 20. Schematic of a small test circuit.

Here also both inputs are switched from low to high. “in2” follows “in1” with a delay of 100ps. In Fig. 21 can be recognized, that after the transition of “in1” only the corresponding buffer switches, whereas the AND gate contributes only a small current. The reason of the small current profile is that only two transistors of the AND gate become active, but the and-condition is not fulfilled. Because the output stays low, no subsequent gates become active at this time.

After the mentioned delay of 100ps, the rising edge on “in2” occurs and the condition for an output change of the AND gate is fulfilled. This causes all subsequent gates to change their states, resulting in the shown current profiles.

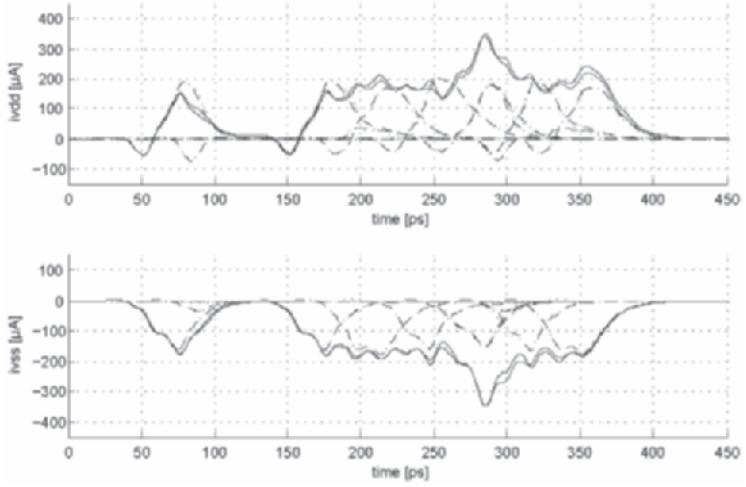


Fig. 21. Comparison of Spice simulation result (gray) and NEMO simulation result (black). Dashed lines represent single gate currents and solid lines represent the overall current.

The plots in Fig. 21 show the comparison of the simulation result of SPICE and the result achieved with the tool NEMO. It can be recognized, that the calculated curves cover the simulation results of SPICE very well. The deviations of the curves are generally in the absolute amplitude of the current. For EMC simulations di/dt is more important than the amplitude. Electromagnetic emission is more related to the shape than the amplitude, and the shape of the current profile is met very well by the NEMO simulations.

6. Conclusions

For EMI modeling and simulation, things are not settled yet. The more complex integrated circuits are, the more complicated the generation of EMI models becomes. Due to the importance of very early prediction of expected electromagnetic emission, EMI modeling and simulation must become part of the IC design flow. Basis for reliable simulation results is the proper description of the dynamic switching currents of digital and analog IC function modules. This task is fulfilled by the proprietary tool NEMO. These "current profiles" are taken by the proprietary tool EXPO for early EMI design case studies, and by the commercial tool XcitePI for EMI sign-off before IC tapeout. The model quality needs to be continuously improved. This work is currently continued at the companies Infineon and Sigrity, supported by the University of Erlangen-Nuremberg.

References

[1] MESDIE project internet pages: www.mesdie.org
[2] Generic IC EMC Test Specification Version 1.0 © 2004 Bosch, Infineon, SiemensVDO (BISS)
[3] Modeling, Simulation and Measurement of Conducted Emissions on Chip-Level; T. Steinecke, H. Koehne, M. Schmidt; EMC Compo 2004, Angers, France
[4] On-Chip Power Integrity Including Package Effects; John Kane, Jiayuan Fang; SOCcentral Publications, March 11, 2005 www.soccentral.com/results.asp?EntryID=12237
[5] Signal and Power Integrity Analysis of large complex PCBs : Theory, Implementation, Simulation and Result Verification by Measurements; Ekkehard Miersch; PIERS 2004, p. 164
[6] Sigrity's XcitePITM performs dynamic noise simulation of the full-chip power grid structure. See www.sigrity.com

[7] Sigrity's CoDesign, combines XcitePI with SPEED2000, Sigrity's time domain PCB and Chip-Package PI- and SI- analysis tool, to get a combined PI-analysis of IC plus Chip-Package
[8] High Level Modeling of Dynamic Switching Currents in VLSI IC Modules; A. Gstoettner, T. Steinecke, M. Huemer; EMC Compo 2005, Munich
[9] "Cookbook for Integrated Circuit model ICEM", IEC standard proposal, project number 62014-3, Oct. 2001.
[10] "MISEA – Modellierung Integrierter Schaltungen für die EMV-Simulation in der Automobilindustrie (Modeling of Integrated Circuit Devices for Automotive EMC Simulation)." MISEA project internet pages: http://www.misea.de
[11] D. Hesidenz, T. Steinecke, "Chip-Package EMI Modeling and Simulation Tool EXPO", EMC Compo 2005, Munich.
[12] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B.P. Flannery, "Numerical Recipes in C", Cambridge University Press, 1992, ch. 3.

SUBSTRATE CURRENT FORMATION, EFFECTS, AND PROTECTION STRATEGIES

Michael Schenkel
Synopsys Switzerland LLC, Affolternstrasse 52, CH-8050 Zurich
michael.schenkel@synopsys.com

Abstract

Semiconductor companies face the demand for 'first silicon success' to be competitive. Substrate current effects are one major cause for redesigns in junction-isolated Smart Power technologies and they put 'first silicon success' at risk. Substrate currents are difficult to control because they are three-dimensional in nature and are strongly layout dependent. Moreover, protection measures show counteracting effects, and minority and majority carrier effects may occur at the same time.

The influence of different protection measures and technology variations on two substrate current effects (substrate potential shift and minority carrier collection) has been investigated by measurements and calibrated full-chip 3D device simulations, including transient effects. The different protection measures show counteracting effects and can therefore only be optimized for specific layouts and application conditions. However, some general recommendations for layout partitioning and the placement of substrate contacts can be derived.

1. Introduction

Smart Power ICs, the integration of high-voltage, high-power transistors together with their low-voltage, and low-power analog and digital control circuits on the same chip reduce cost and electromagnetic interference (EMI) while increasing performance and reliability. One application is motor management ICs that combine four power transistors in an H-bridge configuration and their driving and protection circuitry [1]. In the case of junction isolation, which is still today the favored solution for Smart Power ICs, reverse biased junctions ensure the electrical isolation between components in normal operation mode. However, the typically reverse-biased junctions can be driven in forward conduction by the external circuit. This occurs if the power transistors switch off an inductive load, voltage spikes emerge on supply lines, or simply by accidentally exchanging the supply and ground pin. As a result of the forward-biased isolation junction, parasitic bipolar transistors are activated and a large substrate current flows, which may lead to latch-up in the control circuits and the loss of IC

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 151–182.

functionality. This often leads to costly redesigns with the risk of losing the order because of today's very short time-to-market demand. Moreover, for safety critical applications (e.g. automotive electronics), substrate current effects lead to serious quality and reliability problems if they are not detected before delivery of the systems.

In contrast to substrate noise in mixed-signal ICs, substrate currents injected by parasitic bipolar transistors are currents of low frequency (switching times in the order of microseconds) and are of considerable magnitude (hundreds of milliamperes), leading to a potential shift in the whole substrate of up to several volts. Due to the distributed 3D behavior of substrate currents and their effects protection strategies are strongly layout dependent. Furthermore, protection strategies must cope with minority and majority carriers (counteracting effects) and should consume little space.

Protection structures used against parasitic substrate currents consist of substrate contacts, sometimes together with 'collector-tubs' for the minority carriers [3, 4, 5]. Substrate contacts and collector-tubs can be placed and connected in such a way that a retarding field in the substrate arises and hinders the injected minority carriers from diffusing in the direction of the control circuitry [6, 7, 8].

The focus of this paper is on the placement of substrate contacts and layout-dependent effects in motor control H-bridges. Static as well as transient effects are investigated. Substrate current 'safe' areas for sensitive circuits are identified dependently on relevant parameters such as operating point, doping, substrate contact placement, and so on. All this is performed by using Technology Computer-Aided Design (TCAD) validated by measurements. Due to the presence of minority carriers, numerical simulations must be carried out with a device simulator because a circuit simulator neglects minority carriers.

2. The Origin of Substrate Currents

About 90% of substrate currents originate from recirculating currents due to the switching of inductive loads. Only a small number is due to disturbances on supply lines during operation.

2.1. Inductive Load Switching

H-bridges, which are considered in this paper, consist of four lateral power transistors (LDMOSFETs), which are interconnected with the load (motor) as shown in Fig. 1. The most common way to drive an inductive load (motor) is with pulse width modulation (PWM). Usually, the voltage of the output terminal of one leg (e.g. OUT1) is held stationary while the output terminal of the opposite leg (e.g. OUT2) is switched between ground and V_{supply}. A

masking time is always introduced to avoid a short circuit between the top and bottom transistors of the same leg.

The inductance of the motor windings attempts to keep its current constant during switching. This leads to the need for free-wheeling diodes in circuits where inductive loads are switched. The intrinsic body-to-drain diode of LDMOSFETs is often used as a free-wheeling diode.

During motor operation, the current flows from V_{supply} through the high-side power transistor H1, the motor, and the low-side power transistor L2 to ground (solid green line in Fig. 1). During the masking time, all four power transistors are switched off. However, the inductance of the motor drives the current (up to 5 A) through the inherent body-to-drain free-wheeling diodes of the other two power transistors L1 and H2 (dashed red line in Fig. 1), and generates a potential at the source of H2 (=OUT2) above V_{supply} and at the drain of L1 (=OUT1) below ground due to the voltage drop V_f across the free-wheeling diodes. The above-supply and below-ground voltages determine the forward bias of the emitter-to-base junctions of the parasitic bipolar transistors and lead to carrier injection into the substrate.

Below-ground and above-supply voltages in the range of -1.3 V and 1.5 V, respectively, have been measured during typical PWM mode of operation.

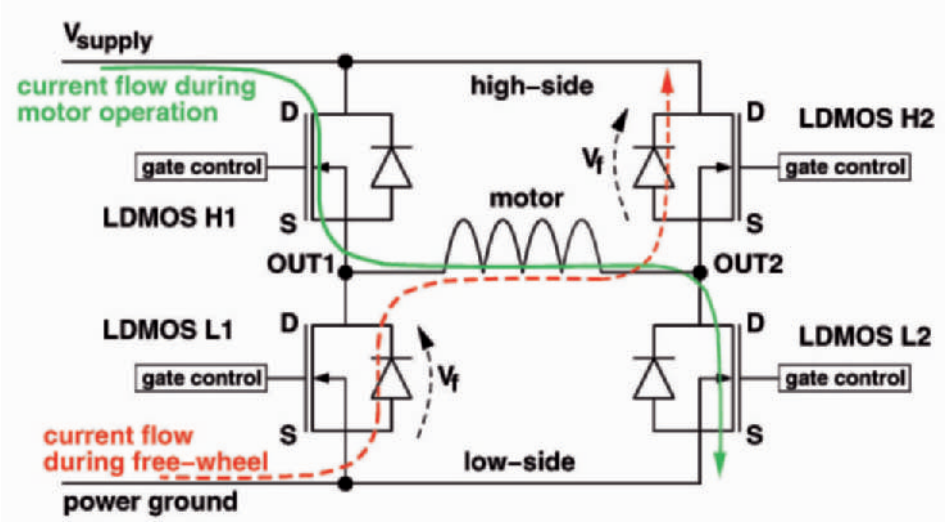


Fig. 1. Motor control H-bridge with two high-side transistors (H1 and H2), two low-side transistors (L1 and L2), and their inherent free-wheeling diodes.

If the drain contact of a low-side power transistor is driven below ground, the parasitic NPN transistor forms as follows [1, 2]. All n-doped regions of the power transistor (n-epi, n-well, n-sinker, n–buried layer) are at a negative potential compared to the grounded p-substrate. The substrate is usually grounded at the back of the chip and/or at substrate contacts at the

top. The junction between the p-substrate and n-regions becomes forward biased and emits minority carriers into the substrate. The n-regions of the high-side power transistors at V_{supply} act as collectors for the lateral parasitic NPN transistor, as shown in Fig. 2. Other n-regions (n-tubs) on a positive potential, e.g. those of sensitive devices in the control circuitry, also collect minority carriers. The collected carriers can disturb the functionality of components in these n-tubs or induce CMOS latch-up [1, 2, 6, 3, 4].

If a current sense resistor is placed between the low-side LDMOS source and ground and the substrate is connected to ground, the forward-biasing voltage across the drain-to-substrate diode is the sum of the voltage drop across the free-wheeling diode and the sense resistor. In the extreme case, this can cause all the recirculation current to flow through the substrate.

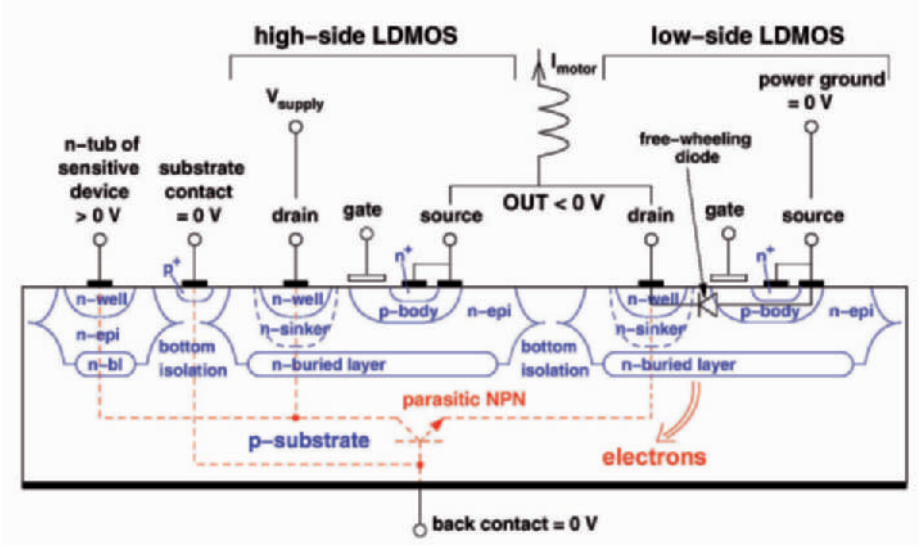


Fig. 2. Simplified cross section of an H-bridge in a Smart Power technology showing the parasitic NPN transistor and free-wheeling diode.

If the source of the high-side transistor is driven above V_{supply} and the junction between p-body and n-epi (free-wheeling diode) becomes forward biased, the source emits holes into the n-epi region. Since all n-doped regions (n-well, n-epi, n–buried layer, and n-sinker) are on a positive potential (V_{supply}) compared to the grounded p-substrate, the junction between the p-substrate and n-regions is reverse biased and acts as a collector for the parasitic vertical PNP transistor, as shown in Fig. 3. The parasitic vertical PNP transistor injects majority carriers (holes) into the substrate, leading to a positive potential shift in the substrate, which may bias isolation junctions forward [1, 2].

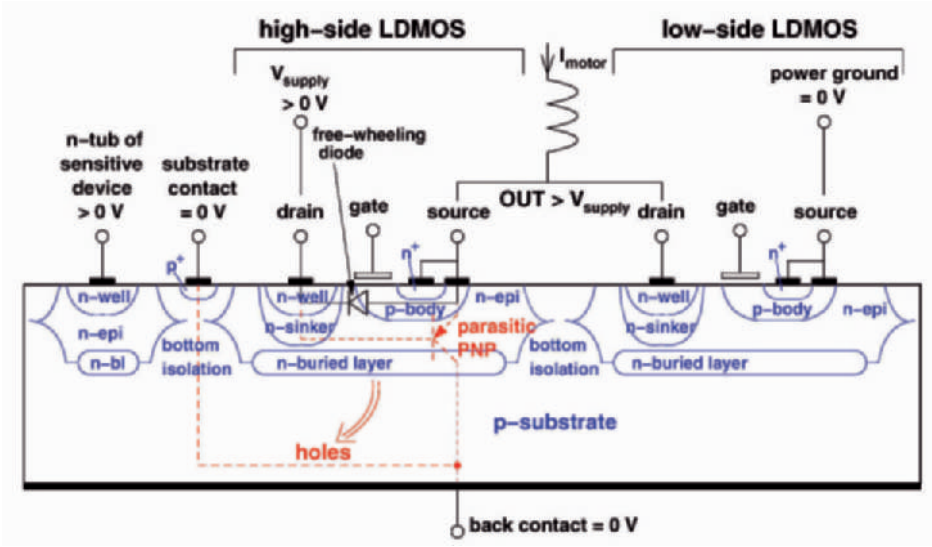


Fig. 3. Simplified cross section of an H-bridge in a Smart Power technology showing the parasitic PNP transistor and free-wheeling diode which, is identical to the emitter-to-base junction of the parasitic PNP transistor.

2.2. Disturbances on Supply Line

Integrated circuits in automotive applications are subject to significant electrical disturbances. Transients originating from various sources such as the ignition system are coupled into electronic devices during standard operation. Immunity against such transients is mandatory for ICs to ensure safe operation of the vehicle. Immunity against transients on supply lines is tested according to the ISO 7637 test standard.

Figure 4 shows a single test pulse of type 3a and type 3b according to the nomenclature of ISO 7637. Test pulse 3a is a pulse superimposed on the battery voltage with a negative amplitude of up to 150 V, and a rise time of 5 ns, and a fall time of 100 ns, depicted in solid black. Test pulse 3b has a positive amplitude of up to 100 V superimposed on the battery voltage of the car with the same rise and fall times, depicted as red dashes. These pulses are applied to the device under test (DUT) in bursts of 100 pulses.

Due to the short rise time and the large amplitude of the pulses, capacitive coupling is severe when applying these pulses.

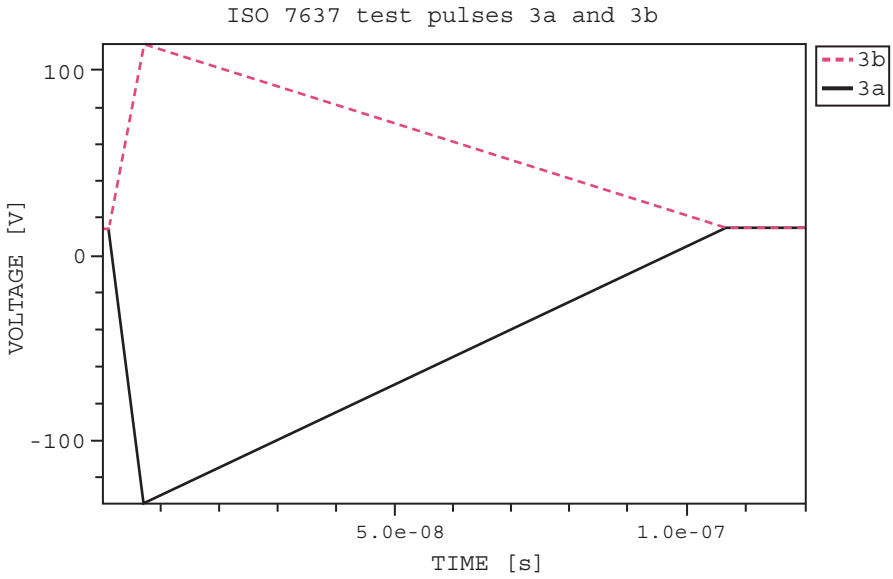


Fig. 4. Single ISO 7637 test pulses.

3. Calibration and Verification of Device Simulation

A test chip has been designed, which is dedicated to measure fundamental dependencies such as potential and minority carrier collection versus distance, direction, and size of the collecting structure or injection conditions and the influence of the protection measures. Measurements also had to serve for the calibration and validation of device simulations.

The layout of the test chips has been designed as closely as possible to H-bridge layouts of existing products but with reduced geometrical complexity to obtain simple structures for simulation. Four rectangular, individually controllable, LDMOS power transistors (L1, L2, H1, and H2), long substrate contacts (LSCs), small substrate contacts (SSCs) at different positions (e.g. s19), a large n-tub as minority carrier collector, and small n-tubs of various sizes at different positions (e.g. n20) have been implemented (cf. Fig. 5). In addition, n-tubs with adjacent substrate contacts (SCs) have been implemented on the test chip to investigate the influence of the substrate contact bias on the collection of minority carriers. Through the SSCs, consisting of a p^+-diffusion connected to a pad, the substrate potential could be measured.

The calibration of device simulation has been performed on full-chip 3D finite element (FE) structures of the test chip (similar to the structure show in Fig. 20). The challenge of full-chip 3D device simulation lies in a strategy to simplify simplification, the complicated chip topology and to omit irrelevant

structures (i.e. topology reduction) [9]. Without the complexity of the device simulation structure would go far beyond today's computing power.

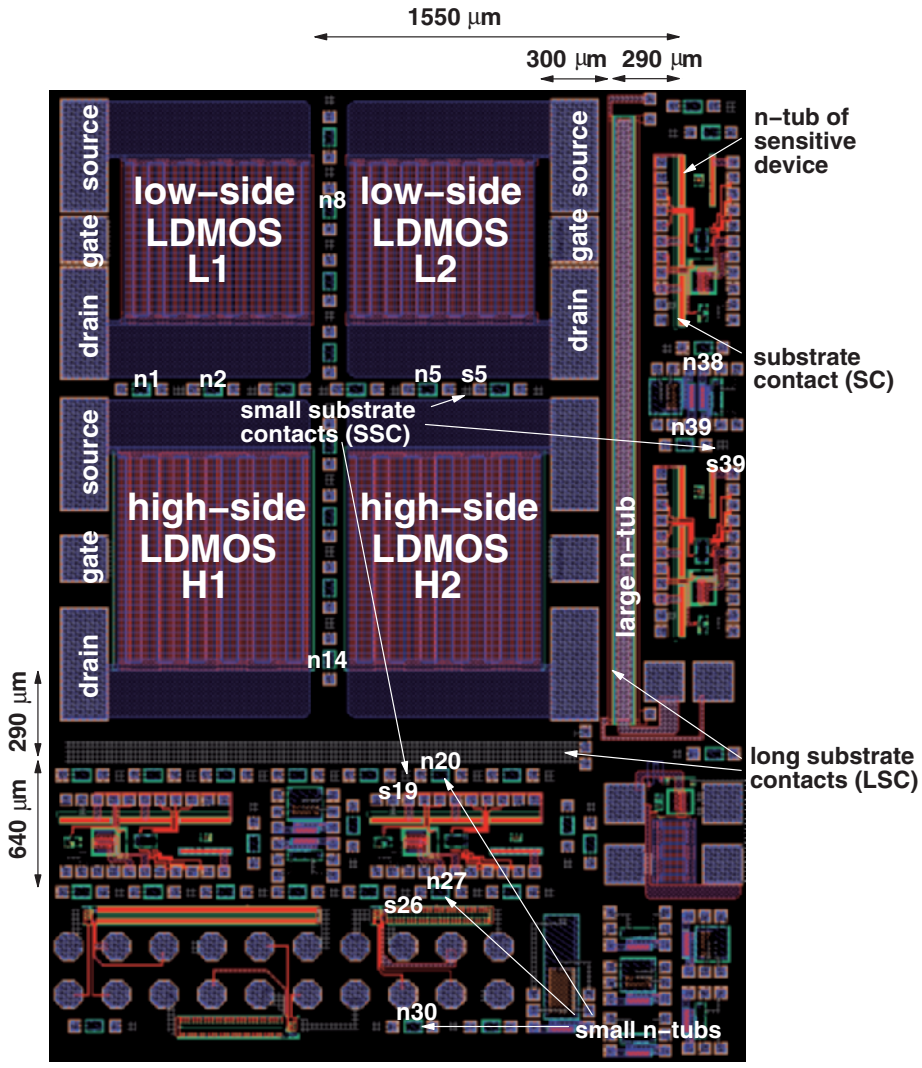


Fig. 5. Test chip in a 0.6 µm Smart Power technology with four LDMOS transistors, and substrate contacts and n-tubs of different size and at different locations.

As important parameters for calibration, the substrate doping concentration N_A, the minority carrier (electron) lifetime τ_n in the substrate, the barrier height Φ_B of the Schottky-type contact at the back of the chip, and the nonnegligible external (contact) resistors have been identified. All these parameters have been calibrated by matching measurements and simulation results of structures whose characteristics depend mainly (or solely) on one

of these unknown parameters [9]. Thereby, calibration intends to extract the unknown parameters within a physically plausible range.

The calibration procedure, as well as the application of static and transient full-chip 3D device simulations, has been validated by comparing measurements and simulation results. Fig. 6 shows the good agreement between the simulated and measured negative shift of the substrate potential due to minority carrier injection and validates, therefore, the calibration and the applied simplifications.

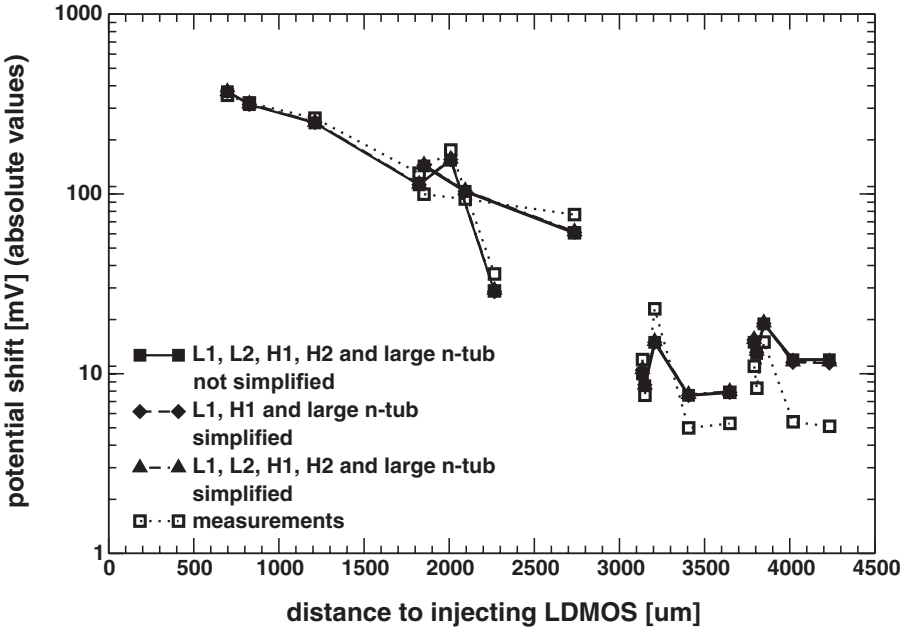


Fig. 6. Negative potential shift as a function of the distance to the center of the injecting LDMOS L2. Simulation and measurements agree well, and simulations with different simplification steps differ only by 6%, i.e. calibration and simplifications are validated.

Figure 7 shows the LDMOS L2 drain voltage, the resulting substrate current, and the collected n-tub current for two different grounding configurations (substrate contact (SC) adjacent to n-tub grounded or not). The applied voltage and timing conditions are close to recirculating application conditions. The switching off of the drain-to-substrate diode of

the LDMOS L2, responsible for substrate current injection, shows reverse recovery with a typical reverse current peak. In the case of a floating SC, the n-tub current reaches its maximum after switching off the substrate current injection, i.e. a considerable amount of current is collected after injection has already stopped. The good agreement between measured and simulated minority carrier collection validates the calibration and the applied simplification steps.

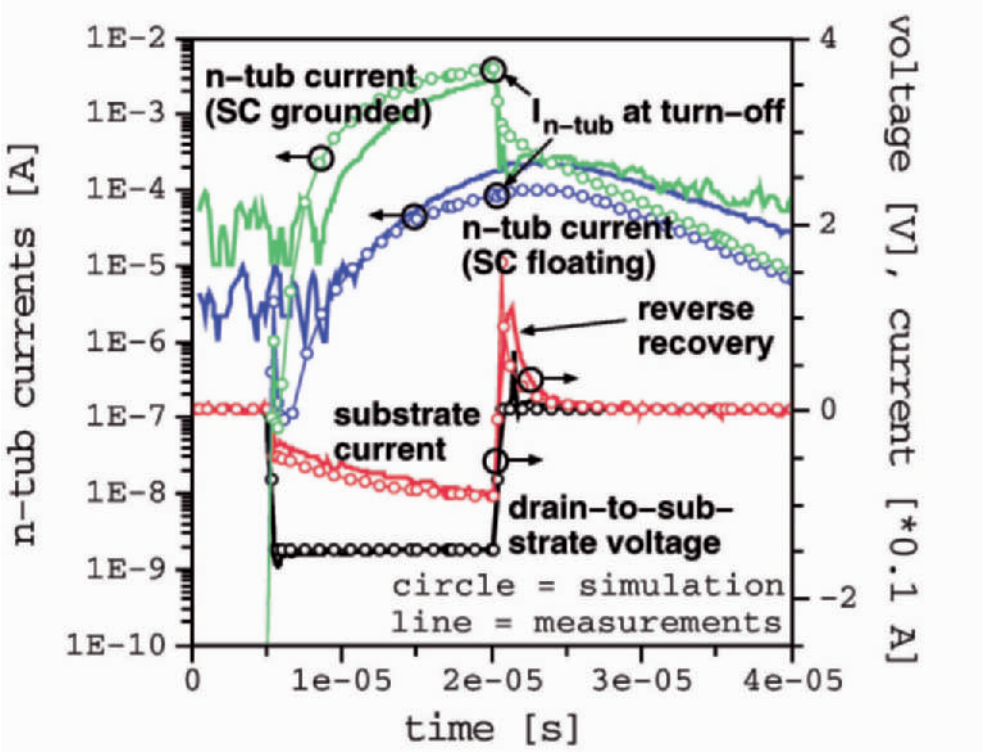


Fig. 7. Collection of transient injected minority carriers. A below-ground voltage of -1.5 V leads to 80 mA substrate current. The collected current is 2.6 mA and 170 μA, respectively, depending on SC bias (LSC and back contact grounded).

4. Substrate Current Effects

After the calibration of the relevant technology parameters and the validation of a topology reduction methodology for power stages and collecting structures [9], investigations of substrate current effects by using full-chip 3D device simulations have been performed.

4.1. Positive Potential Shift Due to Majority Carrier Injection

Holes, which are injected into the substrate by the parasitic PNP transistor at above-supply condition, flow to the substrate contacts and cause a positive potential shift in the substrate. This positive potential shift can lead to failures of junction isolations and high destructive currents due to latch-up, especially at n-tubs on low potential. A positive potential shift between 200 mV and 400 mV may already bias isolation junctions forward, especially at higher temperatures.

Two sets of simulations have been performed out on a full-chip 3D structure as validated in the preceding section. In the first set, different grounding configurations have been investigated, in the second, technology variations were investigated.

Figure 8 shows the comparison of different grounding configurations and the following conclusions can be drawn:

- A grounded back contact reduces the potential shifts by a factor of 10–30. (Most of the substrate current flows to the back contact. Without a grounded back contact, all the substrate current flows to the LSC, which leads to a significantly higher potential shift at s19 and s26.)
- An Ohmic back contact shows no advantage over the present Schottky-type back contact.
- A lower series resistor of the LSC helps to reduce the positive potential shift.

Figure 9 shows the comparison of different technology variations. The following points can be stated:

- A reduction of the substrate doping by a factor of 6 leads to slightly less substrate current due to the higher resistivity of the substrate but a considerably higher potential shift because of the higher LSC current.

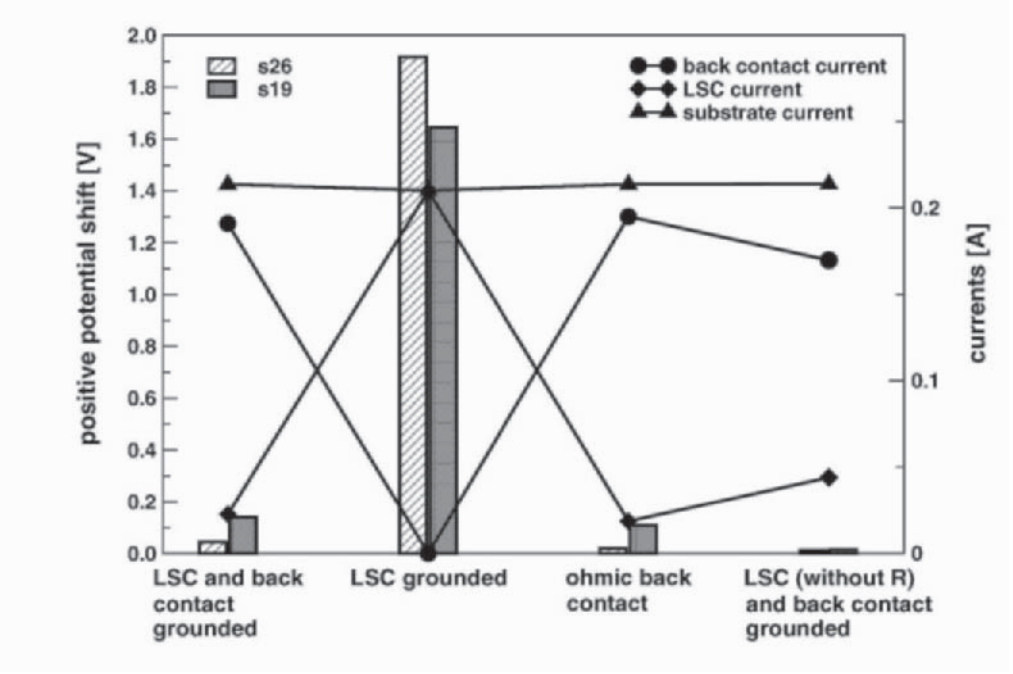


Fig. 8. Simulated positive potential shift at two different locations, substrate current, LSC current, and back-contact current for different grounding configurations.

- An increase of the substrate doping by a factor of 1.6 leads to a slightly lower potential shift because of a lower LSC current.
- A ten-times higher isolation doping leads to increased currents in the substrate, the back contact, and the LSC and, therefore, leads to a higher potential shift.
- A ten-times lower isolation doping reduces the substrate current and, consequently, the back-contact and LSC current, and the potential shift.

4.2. N-Tub Currents Due to Minority Carrier Injection

The parasitic NPN transistor injects minority carriers (electrons) into the p-substrate at a below-ground condition. This leads to a negative potential shift

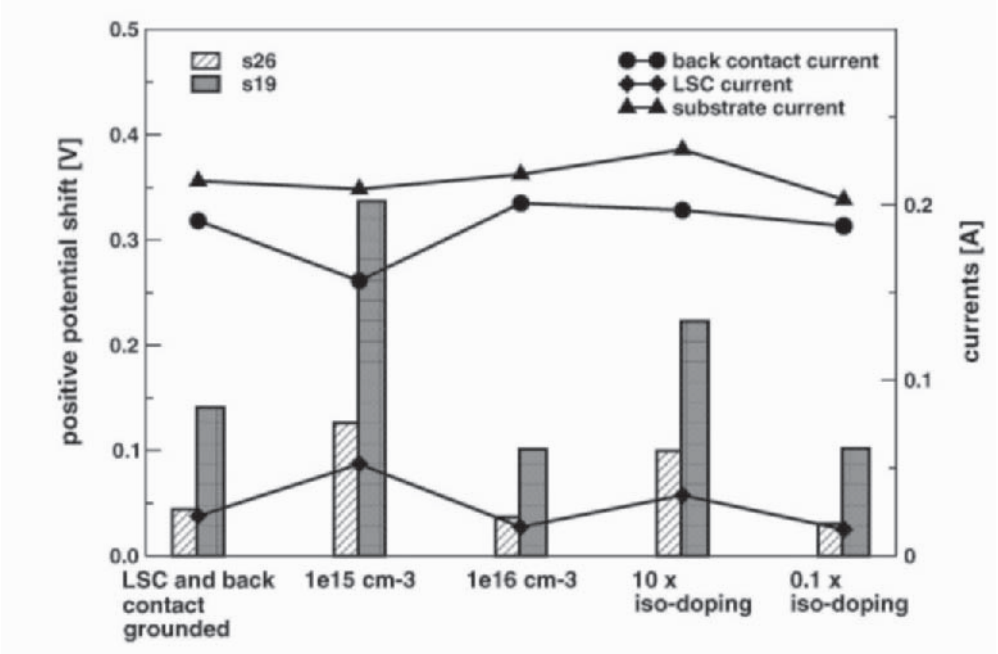


Fig. 9. Simulated positive potential shift at two different locations, substrate current, LSC current, and back-contact current for different technology variations.

in the substrate, which is not harmful in itself. However, the injected electrons are collected by any n-tub in the p-substrate. Because of the long minority carrier lifetime in the substrate, the electrons can reach n-tubs that are several hundred micrometers away from the injecting LDMOS. Electrons, which are collected by an n-tub, can disturb the devices or circuits placed in that n-tub. Therefore, e.g. bipolar PNP transistors, which have the n-tub as base diffusion, are dangerous in the present technology because the collected electron current is added to the base current of the transistor, which of course changes the collector current significantly. Bipolar NPN transistors are less vulnerable because the parasitic current is only added to the collector current. However, the linearity of current mirror circuits depends directly on the ratio of the collector currents of the two transistors used. Therefore,

current mirror circuits can easily be disturbed by substrate currents. For example, the immunity of bandgap reference circuits to substrate currents is below 1 μA collected current [10]. A PMOS transistor in a minority carrier–collecting n-tub is also disturbed by the collected electrons through the body effect, which leads to a change in threshold voltage.

Two sets of simulations have been performed on the same full-chip 3D structure as previously described. In the first set, different grounding configurations and bias conditions have been investigated, in the second, technology variations were investigated. Compared are the quasi-static n-tub current at t = 20 μs (cf. Fig. 7), the capacitive n-tub current due to capacitive coupling of the substrate potential change (dv/dt) to the space charge capacitance of the n-tub (cf. Fig. 12), the back-contact current, and the LSC current.

Figure 10 shows the comparison of different grounding configurations and bias conditions. Most of the substrate current flows into the LSC if it is grounded, because of the Schottky-type back contact (cf. [9]). The most important findings are:

- An open back contact leads to less n-tub current (due to less substrate current) but a higher capacitive n-tub current.
- An Ohmic back contact minimizes the capacitive n-tub current but leads to more quasi-static n-tub current because of a higher substrate current.
- A grounded SC close to an n-tub suppresses capacitive n-tub currents but leads to an n-tub current, which is approximately one order of magnitude higher (due to an altered electric field in the substrate between the SC and LSC [11]).
- If a grounded LSC is placed behind the injecting LDMOS L2 (in the opposite direction to the n-tub), the quasi-static n-tub current is more than 20 times lower and the capacitive n-tub current is 2.7 times lower. This is because the electric field, which builds up in the substrate, accelerates electrons away from the n-tub.
- An n-tub on low potential, e.g. +0.7 V, is very vulnerable to capacitive n-tub currents because of the high space-charge capacitance due to the small space-charge width.

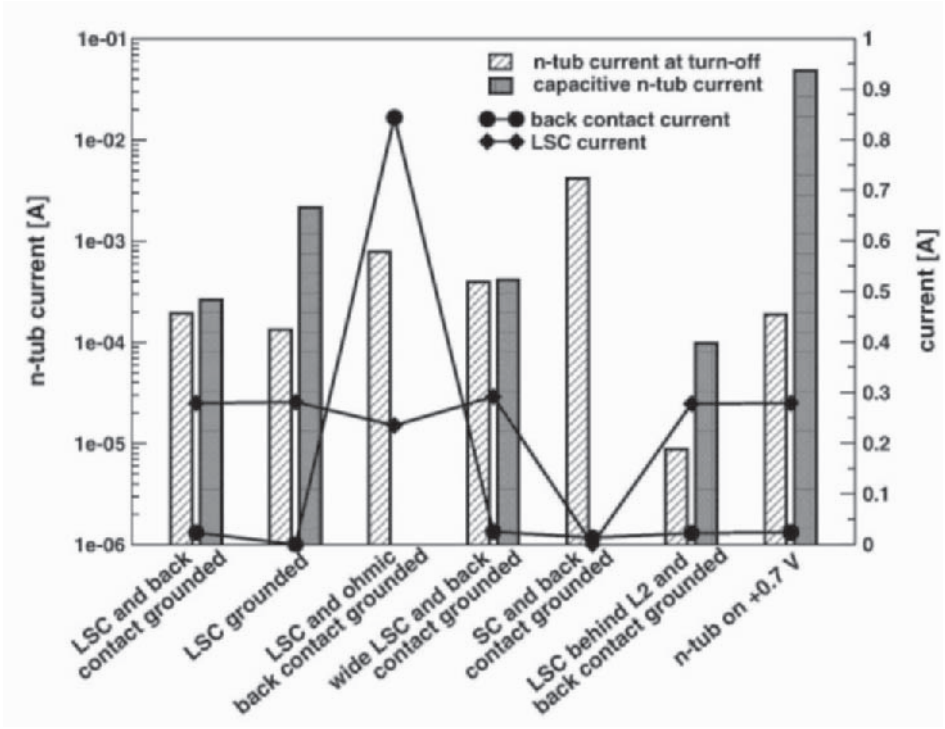


Fig. 10. Collected n-tub current at turn-off, capacitive n-tub current, back-contact current, and LSC current for different grounding configurations and bias conditions.

Figure 11 shows the comparison of different technology variations. The following conclusions can be drawn:

- Ten-times shorter lifetimes lead to less substrate current and a four orders of magnitude lower quasi-static n-tub current, whereas the capacitive n-tub current decrease is much smaller.
- Ten-times longer lifetimes lead to almost ten-times higher quasi-static n-tub current, whereas no capacitive n-tub current appears.
- An increase of the substrate doping by a factor of 1.6 leads to slightly lower substrate current and, therefore, a lower n-tub current.
- A six-times lower substrate doping leads to an increased substrate current and an increased quasi-static n-tub current. The capacitive n-tub current is also higher.
- The isolation doping as well as surface recombination of 1e4 cm/s at the chip sidewalls have almost no influence on the n-tub currents.

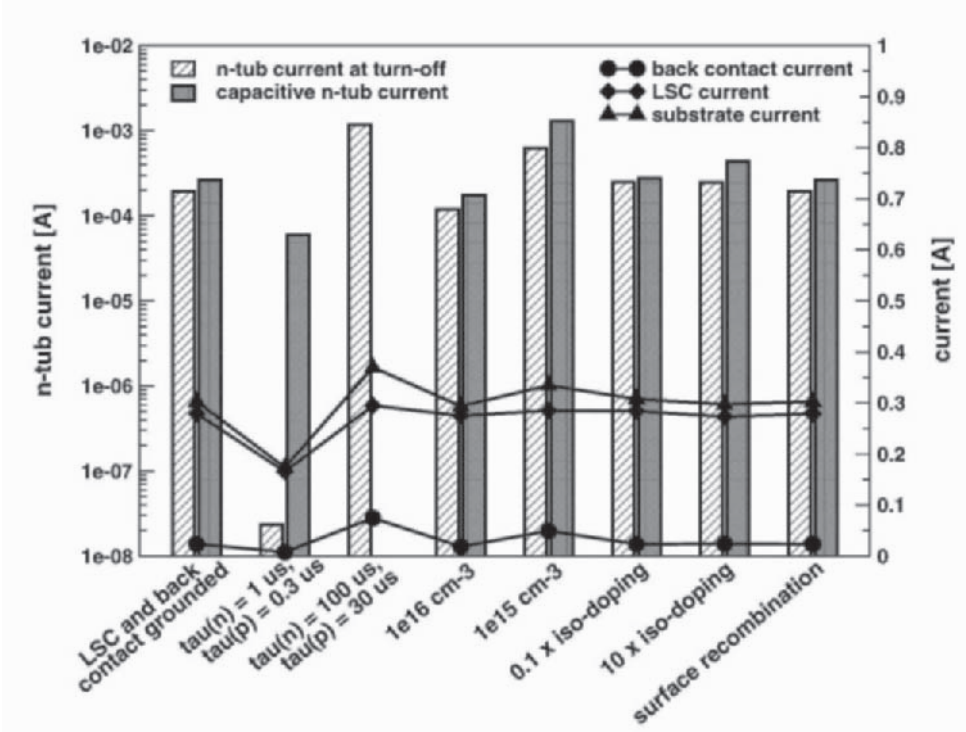


Fig. 11. Collected n-tub current at turn-off, capacitive n-tub current, back-contact current, LSC current, and substrate current for different technology variations.

Reverse recovery, the extraction of excess carriers during diode turn-off, leads to a high opposite (positive) current in the substrate and, therefore, to a positive potential shift. The reverse recovery occurs if the high-side LDMOS is switched on and the potential at the drain of the low-side LDMOS changes quickly from below-ground to V_{supply}. The influence of this transient positive shift of the substrate potential has been investigated further by device simulations. Fig. 12 shows the simulated waveforms at 30 V/µs.

Reverse-recovery current peaks reach values of several hundreds of milliamperes. For dv/dt > 20 V/µs, the substrate potential shift V_{sub} is above 650 mV and consequential n-tub currents of several milliamperes occur due to the short forward-bias of the isolation junction if the n-tub is on low potential (e.g. +0.7 V). Such high n-tub currents may trigger latch-up and are a real threat to correct IC functionality. The n-tub was 600 µm from the injecting LDMOS in this investigation.

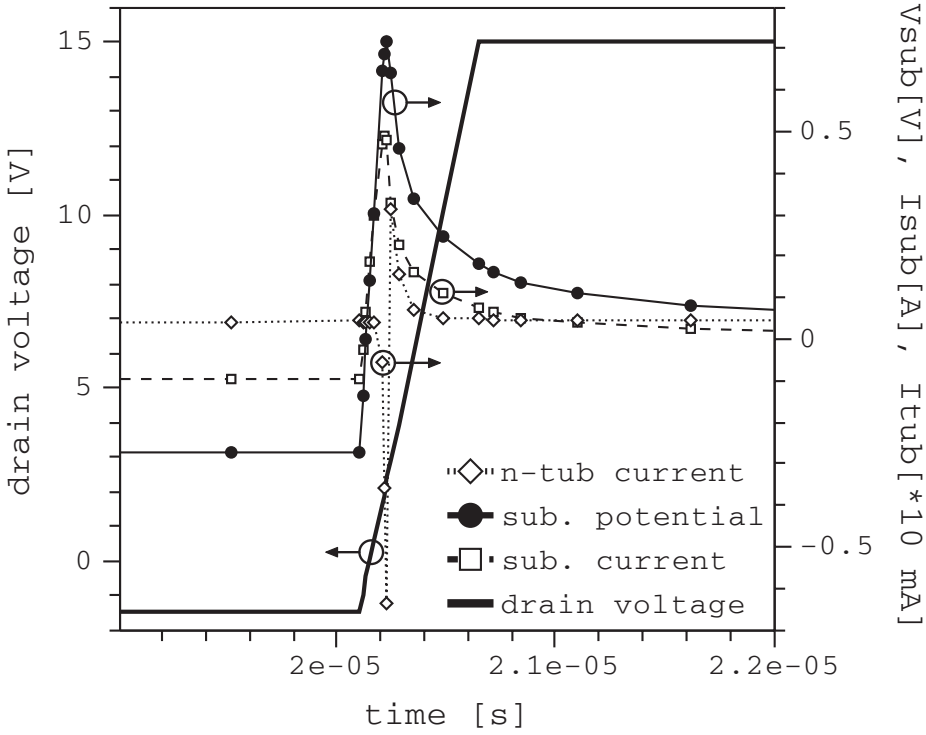


Fig. 12. Reverse recovery of the junction between n-region and p-substrate during high-side LDMOS turn-on. The reverse recovery current leads to a high positive potential shift in the substrate.

4.3. Case Study: Destruction of ESD Protection Diode

Electrostatic discharge (ESD) protection of pads of different power domains is usually performed with a circuit as shown in Fig. 13. Whereas ESD testing of the protection structure did not show weaknesses, diode D1 was destroyed during switching of an inductive load connected to the V_{out} pad.

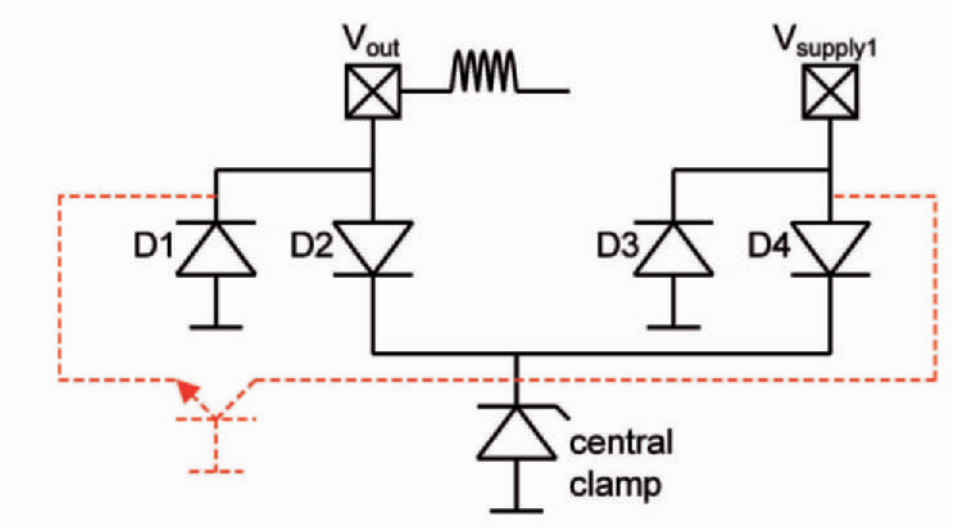


Fig. 13. Test structure of antiparallel coupling diodes to protect different power domains.

Different tests of single devices did not show a destruction only test of the complete structure. Therefore, the opening of a parasitic path had to be assumed. The switching of an inductive load leads to a voltage undershoot at V_{out}. A look at the cross section in Fig. 14 reveals the inherently present parasitic NPN transistor. The below-ground voltage leads to minority (electron) injection and activates the parasitic NPN transistor. Due to the direct connection of the n-epi of diode D4, which is the collector of the parasitic NPN transistor, to $V_{supply1}$ a high current started to flow and burned out the n-body to p-substrate junction of diode D1.

The activation of the parasitic NPN transistor could be confirmed by device simulations and transient latch-up (TLU) measurements.

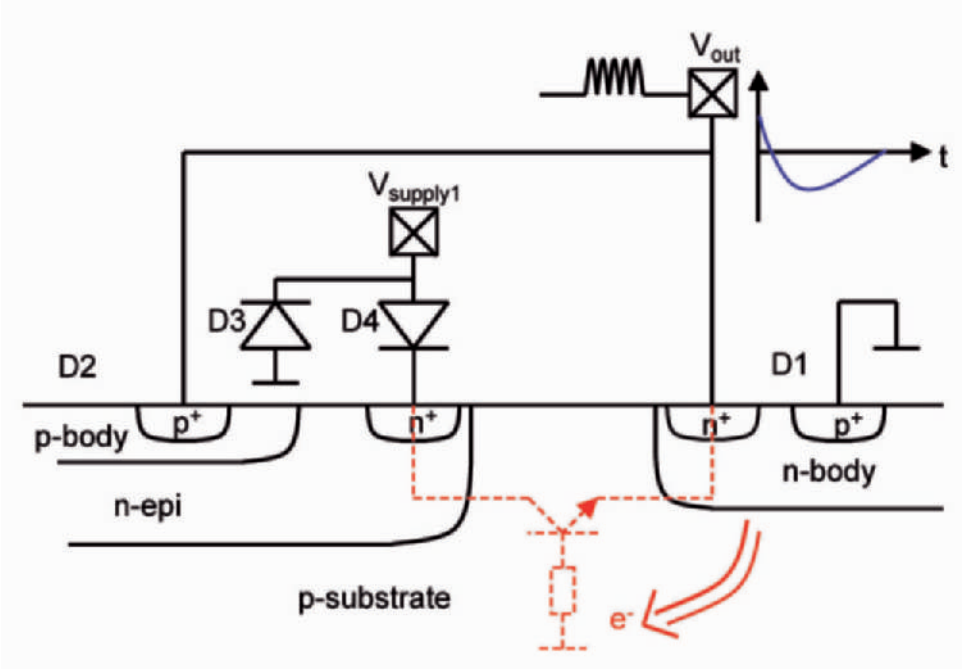


Fig. 14. Cross section of the relevant part of the ESD protection circuit. The below-ground voltage at V_{out} leads to minority (electron) injection into the substrate and activates the parasitic NPN transistor (red dashes).

4.4. Case Study: Notching Up of DMOS Due to Voltage Pulse on Supply Line

ESD protection circuits must be not only tested against ESD pulses but also against disturbances during operation. ISO 7637 pulses as shown in Section 2.2. are used for qualification. Fig. 16 shows an ESD protection circuit that has been tested with ISO 7637 type 3a and type 3b pulses, and has shown destruction of the ESD protection DMOS in the case of the type 3b pulse.

Measurements of the gate voltage of the DMOS showed an increase of the gate voltage which was much larger than the increase expected due to capacitive coupling via the gate-to-drain capacitance. Thus, another coupling

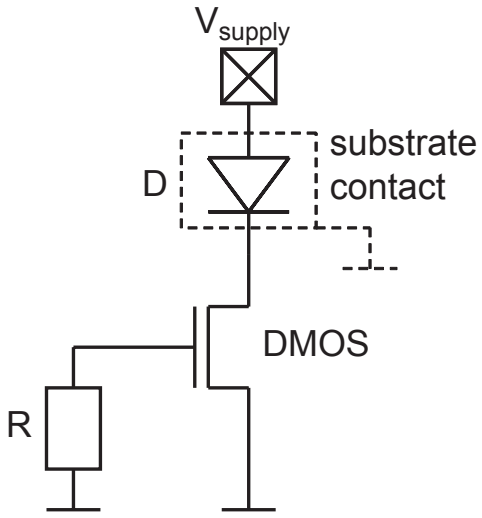


Fig. 15. Schematic of the ESD protection circuit.

mechanism had to play a role. The induced gate voltage opened the DMOS transistor, and a destructive current flowed from V_{supply} through the DMOS transistor. A grounded substrate contact ring around diode D reduced the coupling to the gate, and the gate voltage stayed below a critical value, not opening the DMOS transistor.

Investigations using device simulations revealed a strong coupling through the substrate between the cathode of diode D and the gate resistor R. Furthermore, the influence of a grounded substrate contact ring around diode D could be confirmed. Fig. 16 shows a cross section of the relevant part of the circuit.

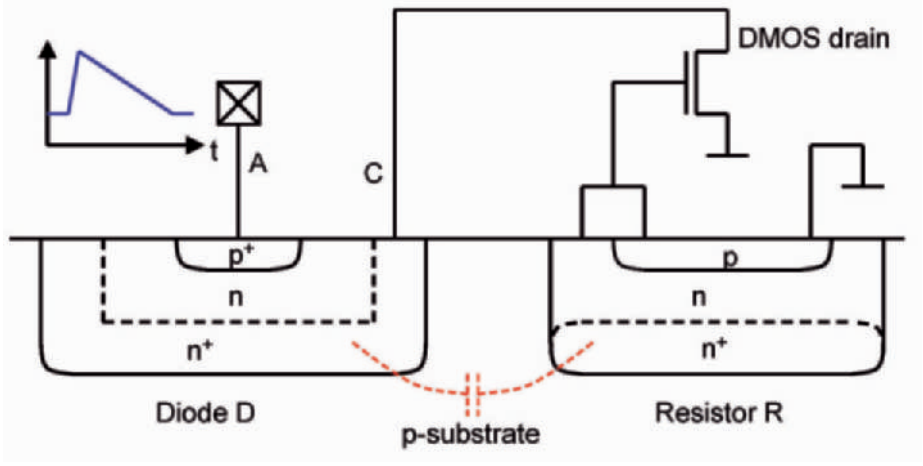


Fig. 16. Cross section of the relevant part of the ESD protection circuit. The fast-rising voltage pulse at the anode of diode D coupled through the substrate to the gate resistor R and opened the DMOS transistor.

An analysis of the temperature dependence showed that the coupling via the substrate increased at higher temperatures. This leads to the situation that, even with a grounded substrate contact ring, the coupling is strong enough to induce a harmful voltage at the gate of the DMOS transistor, as seen in Fig. 17.

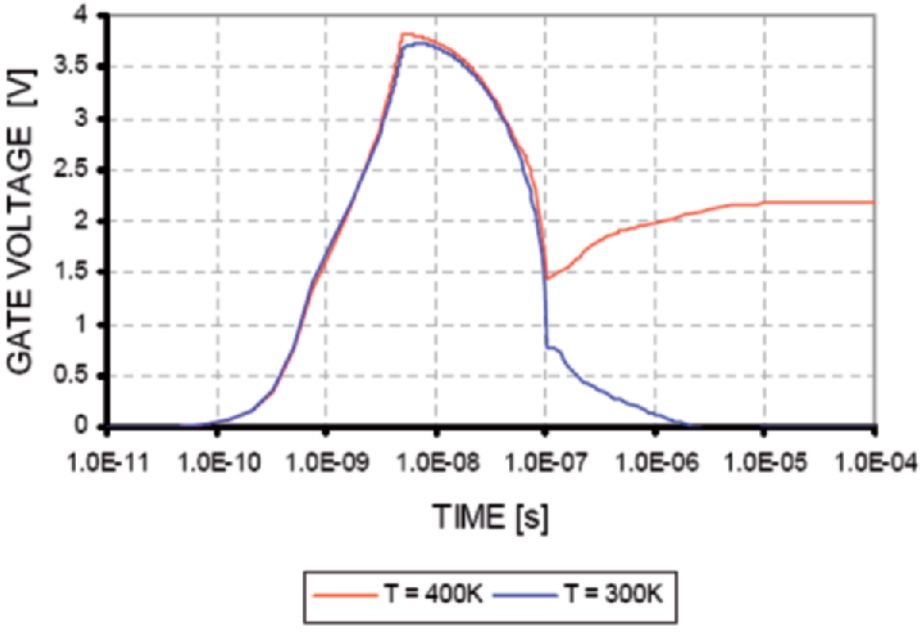


Fig. 17. Simulated gate voltage at two different temperatures. A substrate contact ring around diode D has been grounded in both cases.

5. Protection Strategies

The most efficient way to reduce substrate current effects would be to reduce substrate current injection. Lower substrate currents will also increase the efficiency of Smart Power ICs because less power will be dissipated in the substrate during substrate current injection. The reduction of efficiency due to this additional power dissipation is considerable.

5.1. Technological Measures

The gain of the parasitic PNP transistor depends very strongly on the buried layer and sinker doping. Thus, a high buried layer and sinker doping around high voltage devices is desirable and reduces the gain of the parasitic PNP transistor, but at higher costs.

A high substrate doping and a low isolation doping lead to less substrate contact current and, therefore, a lower positive potential shift (cf. Fig. 9).

In the case of the parasitic NPN transistor, the free-wheeling diode IV-characteristic determines the forward bias of the emitter-to-base junction of the parasitic NPN transistor. Thus, the lower the voltage drop across the free-wheeling diode during recirculation, the lower the forward bias of the minority carrier–injecting junction between the n-regions and p-substrate.

A higher substrate doping (and consequently shorter carrier lifetimes) leads not only to lower substrate current injection but also to lower base resistance and, therefore, higher base-to-emitter voltage. A trade-off between high-side capability and the suppression of minority carrier injection must be found since lower doped substrates allow higher breakdown voltages but minority carriers reach farther into the substrate. On the other hand, highly doped substrates lead to the early recombination of minority carriers, but high breakdown voltages are harder to achieve and it is difficult to create a retarding field in the substrate as used in active barrier concepts (see Section 5.2). This lead to the development of p^- epi on top of p^{++} substrates. As shown in [12], the use of active barriers in p^- epi on p^{++} substrates is feasible whereas the use of deep trenches decreases the minority carrier injection. In addition, the influence of the p^- epi thickness on the active barrier effectiveness has been investigated.

Another interesting approach is to introduce highly doped buried layers below the CMOS part of a chip, which help as recombination barriers [13].

For high-voltage Smart Power technologies, the trend is towards silicon-on-insulator (SOI) [14]. SOI-based isolation techniques can suppress substrate current injection completely. However, transient coupling and the danger of latch-up triggering increases due to the higher parasitic capacitances of SOI structures and the costs of SOI processing are also high [15].

5.2. Substrate Contacts and Their Placement

In an H-bridge during recirculation, both minority and majority carriers are injected into the substrate. Protection measures for motor control H-bridges can, therefore, only be derived by simulations of concurrent minority and majority carrier injection as shown in this section.

Figure 18 shows an application-near simulation structure used for the investigation of substrate contact placement. Substrate contact LSC1 has

been located directly in front of the high-side transistors, LSC2 is behind the low-side transistors, and SC is adjacent to the minority carrier–collecting n-tub. The n-tub has been located 300 µm from the LSC1 and has an area of 90 x 290 µm^2. In addition, a small, floating, substrate potential–sensing SSC has been placed beside the n-tub. Application-near below-ground and above-supply voltages and voltage slopes have been applied. The n-tub has been set to +5 V. Fig. 18 shows the contour plot of the hole quasi-Fermi potential at the chip surface during concurrent electron and hole injection; back contact as well as LSC1 are grounded.

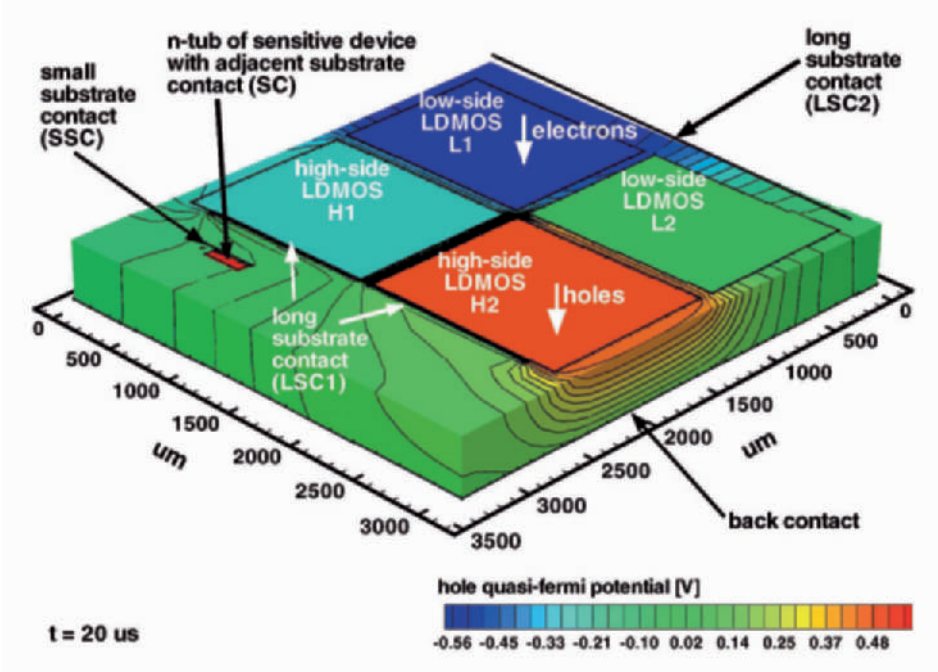


Fig. 18. Application-near full-chip 3D device simulation structure and contour plot of the simulated hole quasi-Fermi potential at the chip surface during concurrent electron and hole injection. LSC1 and back contact are grounded.

In Table 1 and Table 2 different substrate grounding configurations are compared. The quasi-static values of the n-tub current and the potential shift have been taken at 20 µs (cf. Fig. 7), immediately before turning off of the

below-ground and above-supply conditions. The peak of the n-tub current waveform has been taken as capacitive n-tub current value. The peak of the substrate potential waveform has been taken as the transient potential shift value, which is due to the reverse recovery of the junction between the buried layer and the substrate.

Table 1 shows that a grounded LSC2 behind the low-side power transistors leads to a more than three times smaller quasi-static n-tub current than with a grounded LSC1. However, the displacement current is highest. An Ohmic back contact leads to the lowest quasi-static n-tub current as well as the lowest capacitive n-tub current. A grounded SC near the collecting n-tub results in a 2.5 times higher quasi-static current but in a more than 2.5 times lower displacement current. Without a grounded back contact, large quasi-static as well as displacement currents occur. Quasi-static n-tub currents do not exceed 0.4 μA in this configuration. This is mainly due to the shielding effect of the majority carrier–injecting high-sides. On the other hand, during switching, capacitive n-tub currents can reach 60 μA.

Table 1. Collected n-tub currents.

Grounded substrate contacts	Quasi-static n-tub current [μA] at 20 μs	Capacitive n-tub current [μA] at turn-off
LSC1, back contact	0.12	50
LSC2, back contact	0.037	61
LSC1, Ohmic back contact	0.01	7.8
LSC1, SC, back contact	0.29	19
LSC2, SC, back contact	0.25	26
LCS1, (w/o back contact)	0.39	58

From Table 2, the following conclusions can be drawn with respect to potential shift: A grounded LSC2 behind the low-side LDMOS leads to the highest quasi-static potential shift, but the potential is negative. An Ohmic back contact leads to the lowest quasi-static potential shift. A grounded SC results in the lowest transient potential shift and also a very low quasi-static potential shift. If the back contact is not grounded, the transient shift of the substrate potential reaches several volts due to the high reverse-recovery current peak. The substrate potential shift at quasi-static conditions is slightly negative and does not exceed 220 mV. During switching, however, transient potential shifts reach 0.3 V, or even several volts if the back contact is not grounded, and can lead to junction isolation failure if the n-tub lies on a low potential.

Table 2. Substrate potential at SSC near the n-tub.

Grounded substrate contacts	Quasi-static substrate potential [V] at 20 μs	Transient substrate potential [V] at turn-off
LSC1, back contact	-0.083	0.317
LSC2, back contact	-0.220	0.346
LSC1, Ohmic back contact	-0.006	0.286
LSC1, SC, back contact	-0.040	0.137
LSC2, SC, back contact	-0.091	0.156
LCS1, (w/o back contact)	-0.041	2.274

Regarding protection measures against potential shift and n-tub current, the following conclusions can be drawn:

- The lowest n-tub current as well as the substrate potential shift can be achieved by an Ohmic back contact.
- A floating back contact leads to the highest n-tub current and, especially, to the highest substrate potential shift and should, therefore, be avoided whenever possible.
- All other grounding configurations with substrate contacts on top of the chip show counteracting trends and must be chosen according to the specific requirements. For example, if a certain quasi-static n-tub current is acceptable, a grounded substrate contact near the n-tub reduces the capacitive n-tub current as well as the quasi-static and transient substrate potential shift.
- A high-side power LDMOS, which injects majority carriers, should be placed between a minority–carrier injecting low-side power LDMOS and sensitive devices or circuits in the control part. In such a configuration, the injected minority carriers (electrons) recombine with the injected holes before they reach the n-tubs. Therefore, a lower quasi-static n-tub current results, whereby care must be taken with respect to the positive potential shift due to the injected majority carriers.

If only low-side power transistors are on a chip (multiple low-side drivers), the shielding effect of the majority carrier–injecting high-side is absent. If only minority carriers are involved, so-called active barriers are a good measure as shown in the next section.

5.3. Active Barrier Concepts

An active barrier concept was first published in [6, 1]. Active barriers are most useful in (multiple) low-side ICs where no high-side stages are present and can, therefore, be placed between the low-side and the control part. A detailed investigation of this active barrier concept can be found in [7], whereas in [8] an improvement at higher currents is presented.

Two active barrier concepts have been simulated and compared. Concept v2 intentionally introduces a parasitic NPN transistor T_b between the low-side stage and the control circuitry. The collector of T_b is short circuited by a substrate contact to its base, which drives T_b into saturation due to the voltage drop across the substrate resistance between base and collector. In [6], a slightly different structure (see concept v3 in Table 3) is presented, which leads to an even larger attenuation of the electron diffusion towards the control part. In addition, a simulation with only a grounded substrate contact SC3 has been made, which is the grounding configuration with the lowest n-tub current if no active barrier is used. Table 3 lists the three configurations and Fig. 20 shows schematically the 2D simulation structure used.

Table 3. Investigated active barrier configurations.

Name	Description	Reference
v1	SC3 grounded	[9]
v2	SC1 grounded, large n-tub1 shorted to SC2	[1, 7, 8]
v3	SC3 grounded, large n-tub2 shorted to SC1	[6]

Figure 19 compares the n-tub currents of the three different concepts. The substrate current is the same for all three configurations. The reference structure v1 exhibits the Gummel plot of a parasitic transistor with a gain much lower than 1. Active barrier concept v2 starts to suppress the minority carrier injection after a certain threshold voltage is reached, but it is very inefficient at higher current levels, whereas v3 reduces the n-tub current over five orders of magnitude even at higher substrate currents.

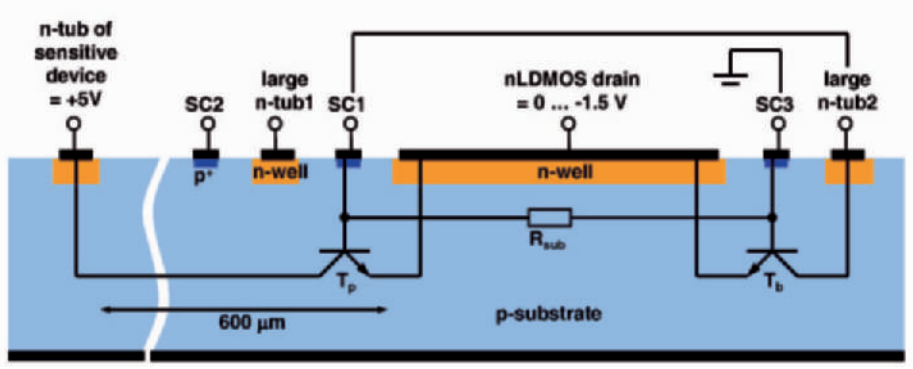


Fig. 19. Two-dimensional simulation structure used to investigate active barrier concepts. The lumped element representation of concept v3 is shown.

The functionality of v3 can be explained as follows. Three parts of the IV-characteristics must be distinguished.

Below a drain voltage of about –0.6 V, the protection does not work and the n-tub current is the collector current of the parasitic NPN transistor T_p formed by the NLDMOS drain, the p-substrate, and the n-tub. Most of the current flows between the NLDMOS drain and SC3, whereas the large n-tub2 already collects minority carriers, which flow from SC1 and R_{sub} to SC3.

Between –0.6 V and –0.8 V, the voltage drop across the substrate resistance R_{sub}, which is due to the current collected by the large n-tub2, is large enough to shut off the parasitic transistor T_p. The junction between the large n-tub2 and the substrate becomes forward biased (T_b in saturation).

Above –0.8 V, the potential at SC1 is almost as negative as at the NLDMOS drain and prevents the parasitic transistor T_p from switching on again. The protection works fully even at higher substrate currents. A retarding field builds up which accelerates the injected minority carriers towards SC3 and away from the n-tub. In configuration v2, however, the voltage drop between the emitter and base of the parasitic NPN transistor at higher currents causes the transistor to leave saturation and the protection efficiency decreases.

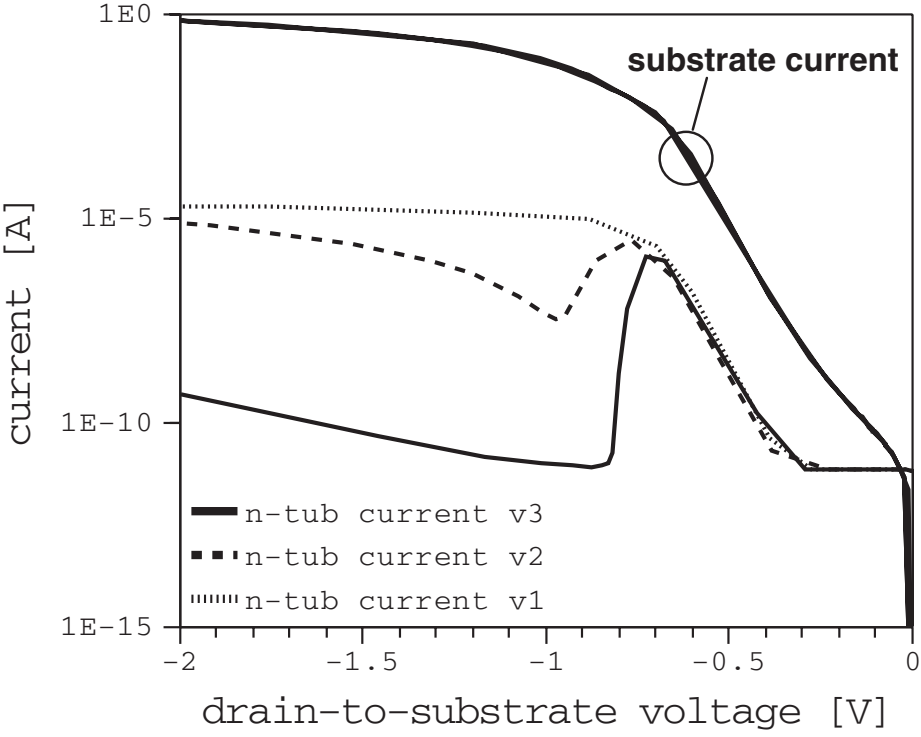


Fig. 20. Comparison of active barrier concepts v1, v2, and v3.

As also presented in the literature [7, 8], the resistor of the floating short, between the large n-tub and the substrate contact, strongly influences the efficiency of active barrier concepts, especially at higher currents. The lower the resistor the higher the efficiency. This may be a disadvantage for the practical implementation of configuration v3.

Furthermore, in [13] it is shown that active barriers only protect circuitries close to the barrier even if four active barrier configuration v2's are connected in series, because minority carriers flow below the sphere of influence of the barrier and reach CMOS parts far away from the barrier. In [16], a modified configuration v4 is presented, which shows higher effectiveness at high injection currents. The configuration v4 short-circuits large n-tub1 to SC1 and grounds SC2. A comparison of configuration v4 and configuration v2 as well as a single large n-tub on a positive potential between injector and sensitive circuitry are shown. The paper also discusses double v4 and double v2 configurations and a combined v4–v2

configuration. It is also shown that an additional ground contact disturbs the efficiency of the active barrier and that the efficiency strongly depends on the width of the large n-tub.

Recently, active bipolar/MOS pull-down protection structures have been published [17]. Their principle consists in locally lowering the substrate potential in order to revert the injecting isolation diode to reverse bias mode. Whereas the bipolar pull-down protection is self-triggered, the MOS gate is biased by a control circuit that detects the negative voltage and then switches on the MOS with a high positive voltage. Due to the highly positive gate voltage, the MOS operates in the linear region with a low on-state resistance and the below-ground potential is tied to a substrate contact ring around the injector. The efficiency of the protection decreases at high current levels because the voltage drop reduces the negative bias of the substrate contact ring.

6. Substrate Current Investigation by Means of Circuit Simulation

In [18], compact modeling of the parasitic NPN transistor is presented. Existing BJT models for circuit simulations are not suitable due to the complex structure of the parasitic NPN transistor (strong inhomogeneous current flow, a base width of up to a few hundred micrometers, and multiple base contacts and collectors). Therefore, a new compact model has been implemented in Verlog-A.

The inputs are the voltages at the contacts as well as technology and layout parameters, and the outputs are the currents. Some functions have been reformulated to avoid singularities that worsen convergence. Comparison of 3D device simulations and circuit simulations is show in Fig. 21, and the agreement of the collected n-tub current is within a factor of 2.5. The accuracy decreases if too few substrate contacts are present because the influence of the electric field on the minority (electron) density is neglected in the model. The simulation time is about 30 seconds compared to hours or even days for 3D device simulations.

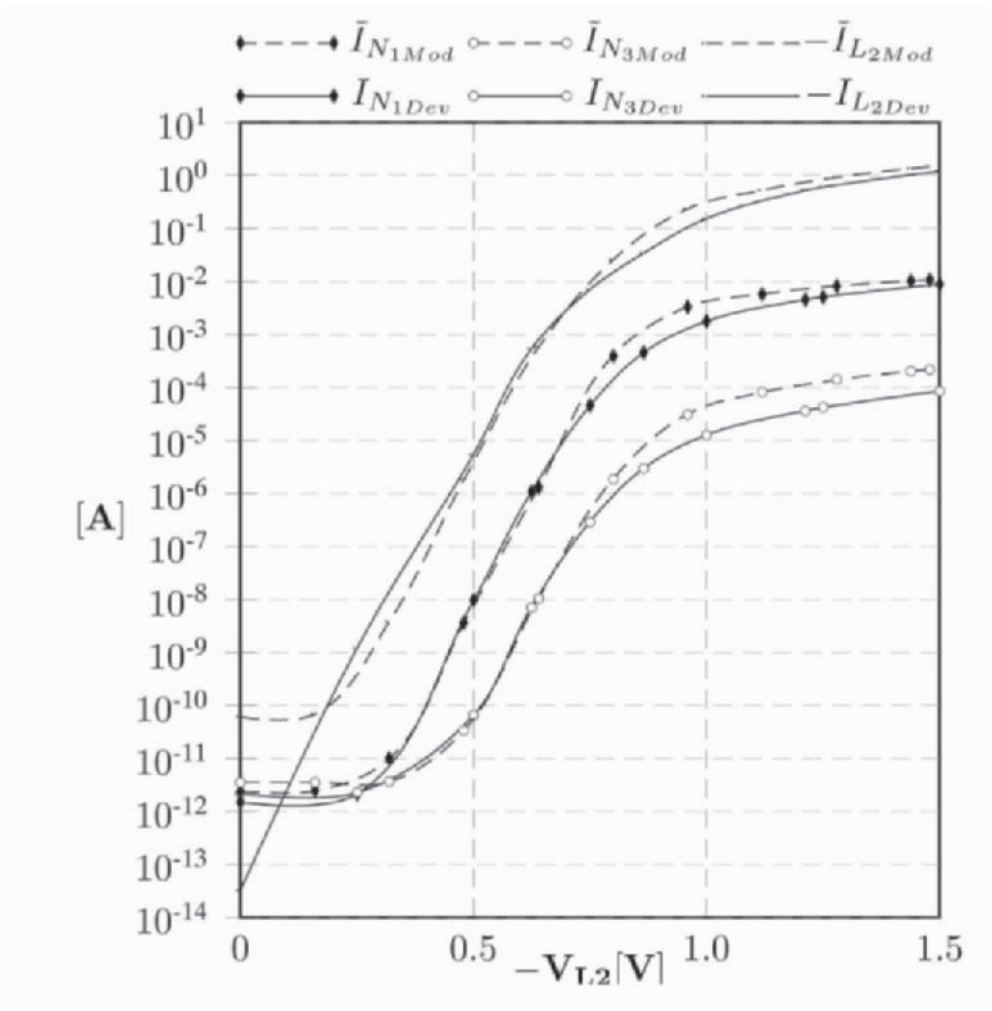


Fig. 21. Simulation results from device simulation (index $_{Dev}$) and circuit simulation employing the new model (index $_{Mod}$) from [18].

7. Conclusions

Substrate current effects are one major risk for Smart Power IC functionality. Most substrate currents originate from recirculating currents during inductive load switching. Therefore, circuits driving inductive loads should always be checked carefully.

Substrate currents are difficult to control because they are three-dimensional in nature, strongly layout dependent, and protection measures, namely grounded substrate contacts, show counteracting effects on them. A good grounding is desirable to reduce the substrate potential shift and

transient effects, but increases the substrate current and collected current. Therefore, optimized measures against substrate current effects are only possible for specific layouts and application conditions.

Despite the counteracting tendency of substrate contact placements, three recommendations can be stated:

- To prevent substrate potential shift (from high-side drivers) and transient coupling, place grounded substrate contacts close to the drivers and/or sensitive devices or circuits in the control part.
- To prevent minority carrier effects (from low-side drivers), place sensitive devices or circuits of the control part as far away as possible from these drivers (the relevant parameter is diffusion length), and avoid grounded substrate contacts in the control circuitry and close to low-side drivers in the direction to the control part.
- The best measure against substrate current effects is a grounded Ohmic back contact. The worst configuration is a floating back contact.

From these recommendations a good partitioning places low-side drivers along the edges of a chip. If high-side drivers are required in the design, they should be placed between low-side drivers and the control circuitry of the chip. If no high-side stages are required (multiple low-side ICs), active barrier concepts or active pull-down protection structures minimize the diffusion of minority carriers towards the protected area. However, the effectiveness of different active barrier concepts strongly depends on the specific layout, grounding configuration, and substrate doping.

A highly doped substrate is favorable to suppress substrate currents but it leads to lower high side blocking capability, and it is more difficult to generate the retarding field used in active barrier concepts. Highly doped buried layers below the injecting device and/or below sensitive circuits are another technological way to reduce substrate current effects.

No results at higher temperature are shown in this paper due to limited space. Of course, most effects and parameters are temperature dependent. However, the large temperature dependence of the free-wheeling diode governs the behavior of the substrate currents at higher temperatures. Moreover, the temperature dependence of the Schottky-type back contact strongly influences the minority carrier injection [9].

8. Acknowledgements

The author would like to thank W. Wilkening (Robert Bosch GmbH) for valuable discussions, M. Schaldach (formerly ETH Zurich) for providing his investigation into disturbances on the supply line, and J. Oehmen (University Hanover) for providing Fig. 21.

References

[1] B. Murari, F. Bertotti, G.A. Vignola, “Smart power ICs: technologies and applications”, Springer, Berlin, 1996.

[2] R. J. Widlar, “Controlling Substrate Currents in Junction-Isolated IC’s”, IEEE Journal of Solid-State Circuits, Vol. 26, No. 8, August 91, pp. 1090-1097.

[3] G. de Cremoux, E. Dubois, S. Bardy, J. Lebailly, “Simulations and Measurements of Cross-Talk Phenomena in BiCMOS Technology for Hard Disk Drives”, International Electron Devices Meeting. Technical Digest, IEEE, New York, 1996, pp. 481-484.

[4] W. W. T. Chan, J. K. O. Sin, S. S. Wong, “A Novel Crosstalk Isolation Structure for Bulk CMOS Power IC’s”, IEEE Transactions on Electron Devices, Vol. 45, No. 7, July 1998, pp. 1580-1586.

[5] V. Venkatesan, Q. Nguyen, A. Bose, P. Parris, “DC substrate coupling between LDMOS and CMOS devices in Hyperintegration I technology”, Proceedings of the 1993 Bipolar/BiCOMS Circuits and Technology Meeting, IEEE, New York, 1998, pp. 57-60.

[6] R. Peppiette, “A New Protection Technique for Ground Recirculation Parasitics in Monolithic Power I.C.,s”, Sanken Technical Report, Vol. 26, No. 1, 1994, pp. 91-97.

[7] O. Gonnard, G. Charitat, Ph. Lance, M. Suquet, M. Bafleur, J. P. Laine, A. Peyre-Lavigne, “Multi-ring Active Analogic Protection for Minority Carrier Injection Suppression in Smart Power Technology”, Proceedings of the International Symposium on Power Semiconductor Devices & ICs, ISPSD 2001, IEEE, Osaka, 2001, pp. 351-354.

[8] S. Gupta, J. C. Beckman, S. L. Kosier, “Improved Latch-up Immunity in Junction-Isolated Smart Power ICs with Unbiased Guard Ring”, IEEE Electron Device Letters, Vol. 22, No. 12, December 2001, pp. 600-602.

[9] M. Schenkel, “Substrate Current Effects in Smart Power ICs”, PhD Thesis ETH-No. 14925, Hartung-Gorre Printing House, Konstanz, Germany, 2003.

[10] W. Horn, H. Zitta, “A Robust Smart Power Bandgap Reference Circuit for Use in an Automotive Environment”, IEEE Journal of Solid-State Circuits, Vol. 37, No. 7, July 2002, pp. 949-952.

[11] M. Schenkel, P. Pfäffli, W. Wilkening, D. Aemmer, and W. Fichtner, “Transient Minority Carrier Collection from the Substrate in Smart Power Design”, Proceedings of European Solid-State Device Research Conference, Nuremberg (Germany), 2001, pp. 411-414.

[12] V. Khemka, V. Parthasarathy, R. Zhu, B. Bose, T. Roggenbauer, “Trade-Off Between High-Side Capability and Substrate Minority Carrier Injection in Deep Sub-Micron Smart Power Technologies”,

Proceedings of the International Symposium on Power Semiconductor Devices & ICs, ISPSD 2003, IEEE, Cambridge, 2003.

[13] M. Zitouni, E. de Fresart, R. De Souza, X. Lin, J. Morrison, P. Parris, "New protection structure against minority carrier injection", Proceedings of the 2003 Bipolar/BiCOMS Circuits and Technology Meeting, IEEE, Toulouse, 2003.

[14] B. Murari, C. Contiero, R. Gariboldi, S. Sueri, and A. Russo, "Smart Power Technologies Evolution", Thirty-Fifth IAS Annual Meeting and World Conference on Industrial Applications of Electrical Energy, Vol. 5, 2000, pp. P-10 - P-19.

[15] E. N. Stefanov, G. Charitat, N. Nolhier, P. Rossel, "Transient Behaviour of Isolation Architectures in Smart Power Integrated Circuits", EPE '97, Trondheim, 1997, pp. 3.036-3.041.

[16] T. K. H. Starke, P. M. Holland, S. Hussain, W. M. Jamal, P. A. Mawby, P. M. Igic, "Highly Effective Junction Isolation Structures for PICs Based on Standard CMOS Process", IEEE Transactions on Electron Devices, Vol. 51, No. 7, July 2004, pp. 1178-1184.

[17] J. P. Laine, O. Gonnard, G. Charitat, M. Bafleur, L. Bertolini, "Substrate current control in Smart Power IC's with a Flexible Protection Structure", Proceedings of the 2002 Bipolar/BiCOMS Circuits and Technology Meeting, IEEE, Monterey, 2002, pp. 36-40.

[18] J. Oehmen, L. Hedrich, M. Olbrich, E. Barke, "A Methodology for Modeling Lateral Parasitic Transistors in Smart Power ICs", Behavioral Modeling and Simulation Workshop 2005, IEEE, 2005.

ON THE SUSCEPTIBILITY OF ANALOG CIRCUIT TO EMI

Franco Fiori
Politecnico di Torino, Electronics Dept.
franco.fiori@polito.it

Abstract

This chapter deals with the effect of electromagnetic interference on the operation of analog front-end like operational amplifier and switched capacitor circuits. In particular, the offset, which is generated in the input differential stage of an operational amplifier is evaluated by an analytical model and a new circuit topology immune to RFI is proposed. Furthermore, the distortion of RFI in a basic switched capacitor circuit is analyzed and a new simple model for MOS switches is derived. Such a model is employed to predict RFI-induced upset in complex SC circuits.

1. Introduction

The electromagnetic emission delivered by modern electronic equipments has significantly increased the level of electromagnetic pollution and nowadays the amplitude of radio frequency interference (RFI), which are collected by electronic system harnesses, can be significantly higher than that of nominal signals. Currently, the amplitude of RFI reaching integrated circuit is kept under control by RFI filters and shields, which are usually inserted at printed circuit board (PCB) level. However, in several automotive applications these solutions are not applicable because most of control, communication and power circuits are fully integrated on silicon as it is in smart power ICs.

In these cases, integrated circuits must be designed to be intrinsically immune to electromagnetic interference (EMI) without the support of any off-chip filter. To this purpose, this chapter presents the last research achievements on the susceptibility of integrated analog front-end to EMI, focusing on the effect of RFI on operational amplifiers and switched-capacitor circuits.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 183–202.

2. CMOS Opamp Susceptibility to EMI

When RFI is superimposed to the input nominal signals of a feedback operational amplifier (opamp), its output nominal signal is definitively corrupted by demodulated RFI and in the specific case of continuous wave (CW) interference, the opamp output signal is affected by a DC output offset voltage, whose value depends on the RFI amplitude and frequency [1] [2].

To this purpose, in previous researches this phenomenon has been ascribed to the distortion of RFI in the opamp input stage [3] and, for this reason, in the following of this chapter, the distortion of RFI in a CMOS differential pair is discussed and a new circuit topology immune to RFI is presented.

2.1 EMI Distortion in CMOS Differential Pair

Common CMOS opamps are composed of cascaded amplifying stages the first of which is usually a differential pair like that shown in Fig. 1, because it amplifies the difference of the input voltages v_{ninv} and v_{inv} i.e.

$$v_d = v_{ninv} - v_{inv},$$

while it rejects the common mode voltage

$$v_{cm} = \frac{v_{ninv} + v_{inv}}{2}.$$

In theory, the output differential current

$$i_d = i_{D1} - i_{D2},$$

does not depend on the common mode voltage if M1-M2 and their loads L1-L2 are perfectly matched. With the assumption of perfect matching and considering the transistors M1-M2 biased in the saturation region, the drain current of each transistor can be written as

$$i_{Di} = \beta \left(v_{GSi} - V_{TH} \right)^2 \quad i = 1,2$$

in which $\beta = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)$ where μ_n is the mobility of electrons in nMOS devices, C_{ox} is the capacitance of the gate oxide per unit of area and W and L

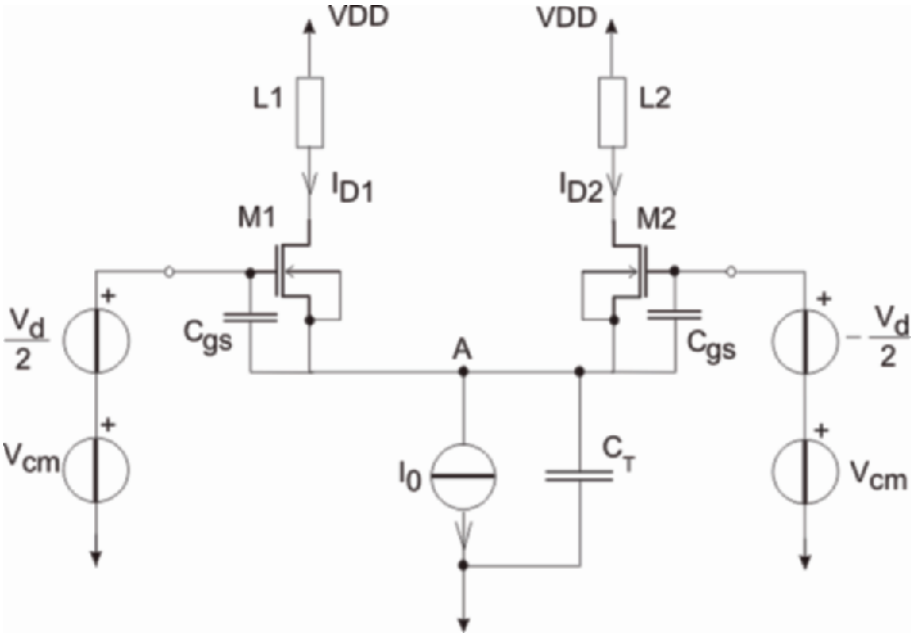


Fig. 1. CMOS differential pair.

are respectively the width and the length of the gate area of the MOS devices of the pair, V_{TH} is the threshold voltage and v_{GSi} is the gate-source voltage.

On the basis of the above mentioned transistor model, the mean value of the differential output current is derived in order to evaluate the RFI-induced output offset current. To this purpose, the gate-source voltage of each transistor is expressed as

$$v_{GSi} = \overline{v_{GSi}} + v_{gsi}(t) \quad , \quad i = 1,2$$

where $\overline{v_{GSi}}$ $v_{gsi}(t)$ are the mean value and the time variant component of the gate-source voltage, and the mean value of the differential output current can be written as

$$\overline{i_d} = \beta\left(\left(\overline{v_{GS1}} - V_{TH}\right)^2 - \left(\overline{v_{GS2}} - V_{TH}\right)^2 + \left(\overline{v_{gs1}^2} - \overline{v_{gs2}^2}\right)\right).$$

Now, assuming that the mean-value of the input differential voltage is null ($\overline{v_d} = 0$) the following equation holds

$$\overline{v_{GS1}} = \overline{v_{GS2}},$$

and the mean-value of the output differential current can be written as

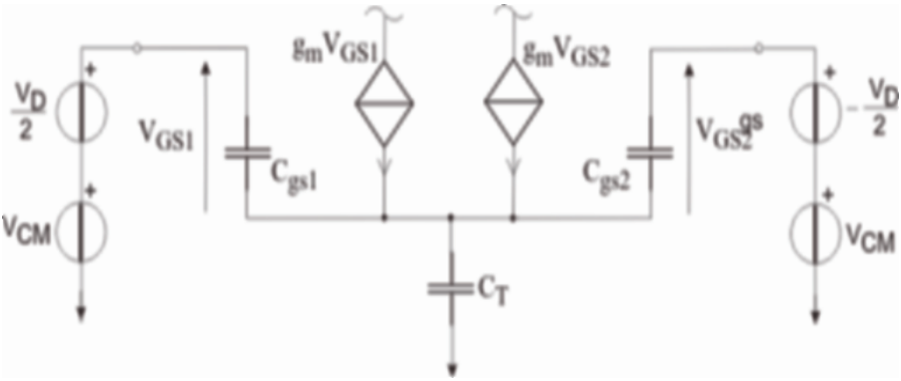


Fig. 2. Small signal equivalent circuit.

$$\overline{i_d} = \beta\left(\overline{v_{gs1}^2} - \overline{v_{gs2}^2}\right) \tag{1}$$

in which

$$\overline{v_{gsi}^2} = \int_{-\infty}^{+\infty} S_{V_{GSi}}(\omega)d\omega = \int_{-\infty}^{+\infty} V_{GSi}(\omega)V_{GSi}^*(\omega)d\omega \tag{2}$$

where $S_{V_{GSi}}(\omega)$ is the power spectral density of the gate-source voltage, which can be expressed in terms of its Fourier transform.

With reference to the small signal equivalent circuit of the differential pair as it is shown in Fig. 2, the gate-source voltage of the transistors M1 and M2 can be written as

$$V_{GS1}(\omega) = H(\omega)V_{CM}(\omega) + \frac{V_D}{2} \tag{3}$$

$$V_{GS2}(\omega) = H(\omega)V_{CM}(\omega) - \frac{V_D}{2} \tag{4}$$

where

$$H(\omega) = \frac{j\omega C_T}{2g_m + j\omega\left(C_T + 2C_{gs}\right)}$$

in which g_m is the transconductance of M1 and M2, C_T is the parasitic capacitance between the node A and the ground (see Fig. 2), while C_{gs} is the gate-source capacitance.
Substituting (3) and (4) in (2) then in (1), the mean value of the differential output current can be expressed as

$$\overline{i_d} = \beta \int_{-\infty}^{+\infty} \mathrm{Re}\{H(\omega)V_D(\omega)V_{CM}(\omega)\}d\omega\,. \tag{5}$$

In the specific case of $v_d = V_{d\_pk}\cos(\omega_0 t)$ and $v_{cm} = V_{cm\_pk}\cos(\omega_0 t + \theta)$, eqn. (5) becomes

$$\overline{i_d} = \frac{\beta}{2} V_{d\_pk} V_{cm\_pk} |H(\omega)| \cos(\Phi_H + \theta), \tag{6}$$

in which $\Phi_H = \arctan\left(\dfrac{\mathrm{Im}\{H\}}{\mathrm{Re}\{H\}}\right)$.

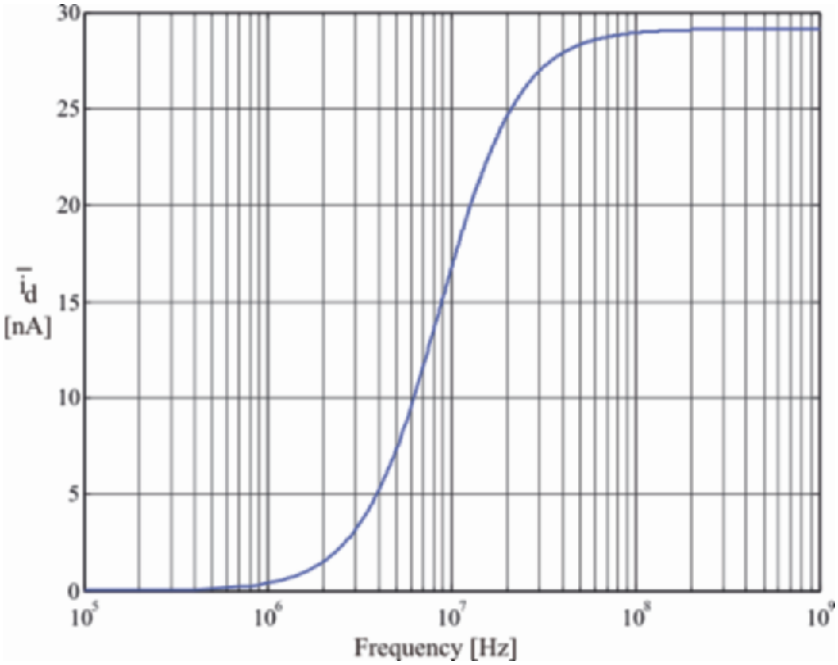


Fig. 3. Output offset current induced by RFI in a differential pair. (W/L)=100/10, I0=20uA, CT=2.1pF.

Figure 3 shows the output offset current versus frequency of an nMOS differential pair with (W/L)=100/10, I0=20uA and C_T = 2.1pF with $V_{d\_pk} = 20mV$ and $V_{cm\_pk} = 10mV$, which has been obtained by (6). Once evaluated $\overline{i_d}$, the input offset voltage of an operational amplifier is computed as

$$v_{off} = \frac{\overline{i_d}}{g_m}$$

where g_m is the transconductance of the input differential stage.

In the next section, the above derived model is employed to design an opamp input stage immune to RFI.

2.2 The Double Differential Pair

In the previous section it has been shown that the simultaneous presence of common mode and differential mode voltages generates in a differential pair an output offset current (see eqn. (5)).

Lets now consider two distinct differential pairs (M1-M2, M3-M4 in Fig. 4), which are driven by the same input signals and whose output ports are cross connected as it is shown in Fig. 4. In this circuit, the overall output offset current can be evaluated as the difference of the offset current, which is generated in the pair M1-M2 (#1) and in the pair M3-M4 (#2), that is

$$\overline{i_d} = \overline{i_{d1}} - \overline{i_{d2}} . \tag{7}$$

Substituting eqn. (5) in (7), the output offset current can be written as

$$\overline{i_d} = \int_{-\infty}^{+\infty} \mathrm{Re}\{[\beta_1 H_1(\omega) - \beta_2 H_2(\omega)] V_D(\omega) V_{CM}(\omega)\} d\omega$$

where

$$\beta_i H_i(\omega) = \frac{j\omega C_{Ti}\beta_i}{2g_{mi} + j\omega(C_{Ti} + 2C_{gsi})}. \qquad i = 1,2$$

In order to achieve the immunity to EMI, it must be verified that $\overline{i_d} = 0$ i.e.

$$\beta_1 H_1(\omega) = \beta_2 H_2(\omega), \tag{8}$$

which is verified if two identical differential pairs are employed. However, this choice gives the trivial solution of zero output current for any inputs, including the nominal input signals. For this reason, eqn. (8) must be satisfied under the assumption $g_{m1} \neq g_{m2}$.

From eqn. (8) follows

$$\frac{\beta_1 C_{T1}}{C_{T1} + 2C_{gs1}} = \frac{\beta_2 C_{T2}}{C_{T2} + 2C_{gs2}}, \tag{9}$$

$$\frac{g_{m1}}{C_{T1} + 2C_{gs1}} = \frac{g_{m2}}{C_{T2} + 2C_{gs2}}, \tag{10}$$

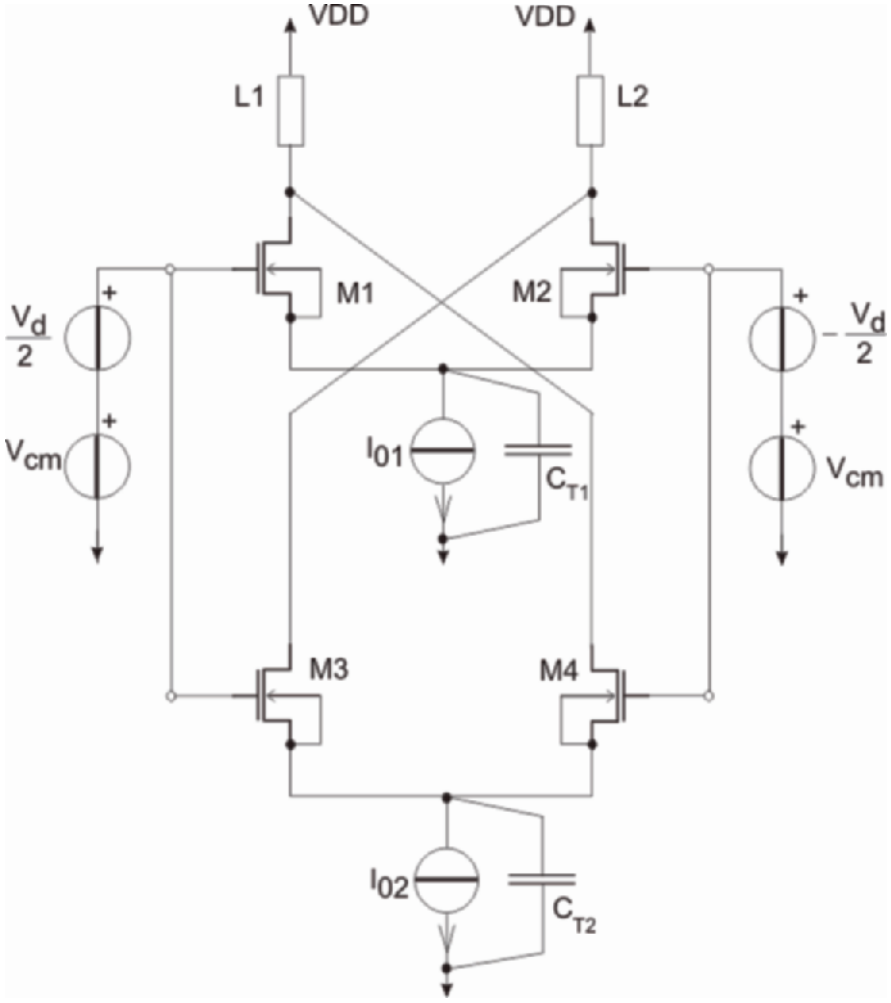


Fig. 4. The double differential pair.

and

$$\left(\frac{W}{L}\right)_2 = \frac{\left(1+\frac{C_{gs2}}{C_{T2}}\right)}{\left(1+\frac{C_{gs1}}{C_{T1}}\right)}\left(\frac{W}{L}\right)_1,$$

$$\left(\frac{C_{T2}}{C_{T1}}\right)^2 = \frac{\left(\frac{W}{L}\right)_2}{\left(\frac{W}{L}\right)_1}\frac{I_{02}}{I_{01}},$$

which can be employed to design the auxiliary differential pair M3-M4 in order to compensate the offset current, which is generated in the main differential pair M1-M2 by RFI.

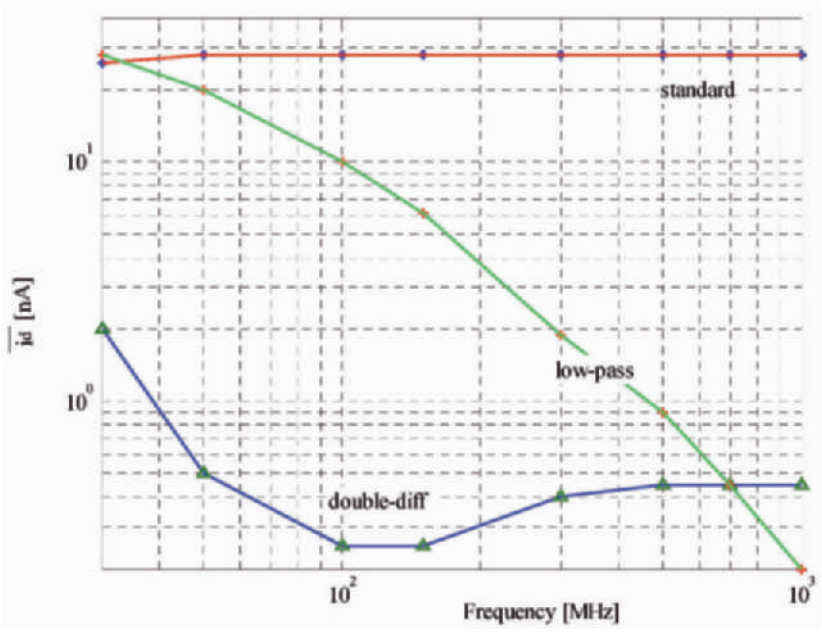


Fig. 5. Output offset current versus frequency. (*) standard differential pair, (+) standard differential pair with input low-pass filter, (Δ) Double differential pair M1-M2, M3-M4 in Fig. 4.

In particular, it can be observed that C_{T2} is minimized if $\frac{C_{gs2}}{C_{T2}} \approx 0$, hence, under this assumption, if the parameters of the pair M1-M2 are known and once defined the ratio $\frac{I_{02}}{I_{01}}$, the aspect ratio of M3, M4 and the capacitance C_{T2} can be derived. Furthermore, I_{01} is known because its value derives from opamp base-band design specifications, while I_{02} is defined on the basis of the maximum RFI signal amplitude reaching the M3-M4 input terminals.

In order to make evident the above described design method, a differential pair M1-M2 with $(W/L)_1 = 100/10$, $C_{T1} = 2.1$ pF, $I_{01} = 20\mu A$ and $C_{gs1} = 850$ fF is considered. On the basis of these data and assuming $\frac{C_{gs2}}{C_{T2}} \approx 0$ and $I_{02} = 5\mu A$, the aspect ratio of M3-M4 is equal to $\left(\frac{W}{L}\right)_2 = 5.5/1$ and the capacitance $C_{T2} = 778$ fF.

With reference to above mentioned parameters, the output offset current, which is induced in the double differential pair by common and differential mode interference has been evaluated by time domain simulation. To this purpose, Fig. 5 shows the output offset current versus frequency for the differential pair M1-M2 (standard), the double differential pair (double-diff) and a differential pair (low-pass) whose input signals have been filtered by two passive low-pass filters (cut-off frequency is 60 MHz (low-pass)). A lower value of the cut-off frequency requires a proper design of the opamp frequency compensation network.

These input stages have been excited by the same radio frequency differential mode $V_{d\_pk} = 20mV$ and common mode voltage $V_{cm\_pk} = 10mV$.

3. EMI Susceptibility of Switched Capacitor Circuit

The failures, which are induced by RFI in SC circuits, i.e. in circuits that only includes MOS switches, capacitors and opamps connected in a non-degenerate way, are related, in general, both to the opamp(s) RFI-induced distortion and to the high-frequency behavior of MOS switches.

While the nonlinear effects of RFI in opamp circuits has been discussed in the previous chapter, in the following the high-frequency distortion of MOS switches is discussed.

3.1 EMI Distortion in Elementary MOS Switch

To this purpose, the basic SC circuit in Fig. 6a is considered. In such circuit, the input signal

$$v_{IN} = V_{IN} + v_{in}(t) \tag{11}$$

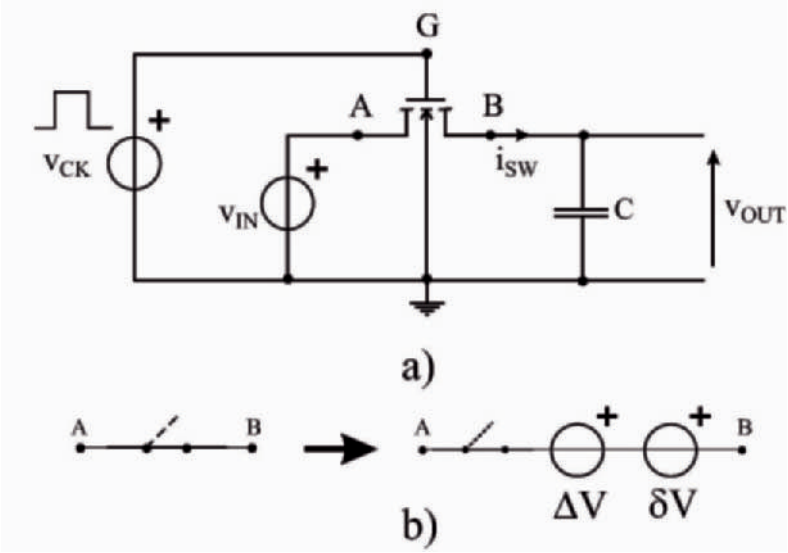


Fig. 6. MOS Switch Distortion: a) Basic SC circuit. b) RFI-induced errors in closed-MOS switches.

includes a nominal DC signal V_{IN} and a zero-mean RFI term $v_{in}(t)$, as shown in Fig. 7. When the switch is on and RFI is added to the DC input signal, the output voltage

$$v_{OUT} = V_{OUT} + v_{out}(t), \tag{12}$$

includes a DC component $V_{OUT} \neq V_{IN}$ and a zero-mean component $v_{out}(t) \neq v_{in}(t)$. As a consequence, with reference to Fig. 7, the capacitor voltage V_S at time T_s, when the switch is turned off, can be expressed as

$$\begin{aligned} V_S &= V_{IN} + (V_{OUT} - V_{IN}) + v_{out}(T_s) + V_{CH} \\ &= V_{IN} + \Delta V + \delta V + V_{CH} \end{aligned}, \tag{13}$$

where V_{IN} is the nominal DC input voltage, ΔV is the RFI-induced offset error and $\delta V = v_{out}(T_s)$ is the error component which depends on T_s and

which can be described in terms of its root mean square (rms) value over the possible sampling instants. Finally, V_{CH} is an error term due to charge injection that is not affected by RFI and which will not be considered hereafter.

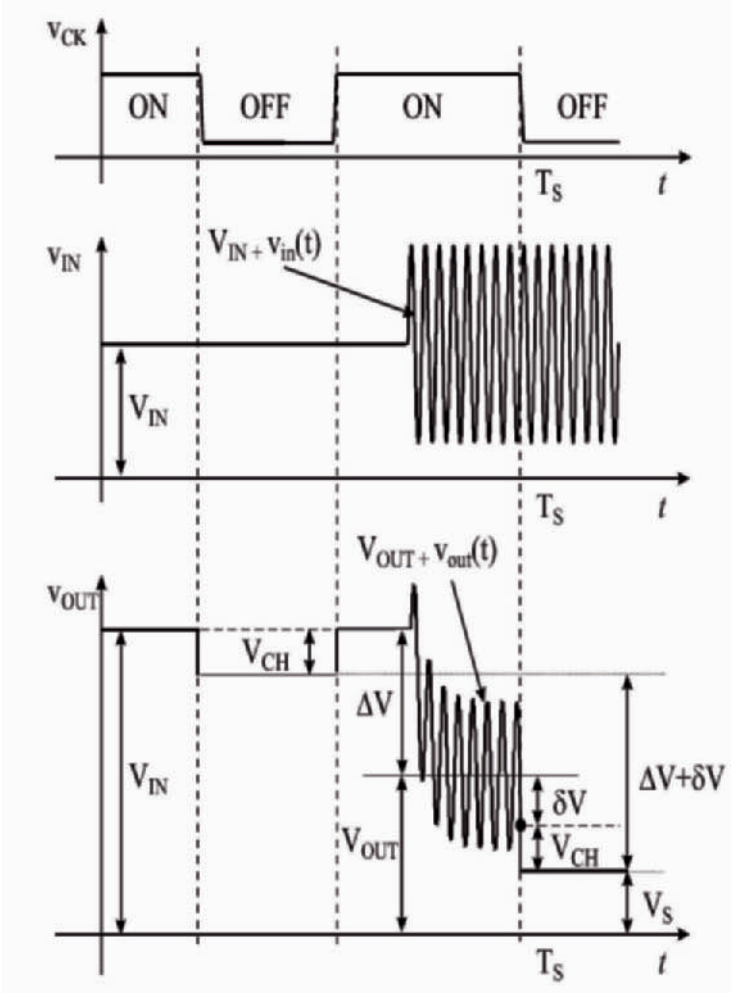


Fig. 7. Typical waveforms of the circuit in Fig. 6a in the presence of RFI.

From eqn. (13), the RFI-induced errors due to on-switch distortion can be predicted by standard baseband SC circuit analysis if the MOS switches are replaced as shown in Fig. 6b. The RFI-induced errors ΔV and the rms value of δV, which depend on switch parameters and on the SC circuit configuration, are predicted in the following by a simple analytical model.

3.2 Prediction of Errors due to MOS Switch Distortion

The RFI-induced errors ΔV and δV are predicted employing the bulk-referred EKV MOS transistor model [4]. According to this model, with reference to Fig. 6a, the current which flows through a closed MOS switch is expressed as

$$i_{SW} = n\beta\left(v_P - \frac{v_A + v_B}{2}\right)(v_A - v_B), \tag{14}$$

where v_A and v_B are the bulk-referred terminal voltages and $\beta = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)$ in which μ_n is the mobility of electrons in nMOS devices, C_{ox} is the capacitance of the gate oxide per unit of area and W and L are respectively the width and the length of the gate area of the MOS device.
Furthermore,

$$v_P = v_G - V_{T0} - \gamma\left\lfloor\sqrt{v_G - V_{T0} + \xi^2} - \xi\right\rfloor \tag{15}$$

in which v_G is the bulk-referred gate voltage, V_{T0} is the threshold voltage and ξ, γ and n are technology parameters which are described in [4].

Equation (14) is valid for MOS transistors in the linear region, i.e. under the assumption that both $v_A < v_P$ and $v_B < v_P$, and it is employed in the following in order to predict the RFI-induced errors ΔV and δV. To this purpose, with reference to Fig. 6a, from eqns. (11), (12) and (14) the current i_{SW} can be expressed as

$$i_{SW}=n\beta\left[V_P\left(V_{IN}-V_{OUT}+v_{in}-v_{out}\right)-\frac{1}{2}\left(V_{IN}+v_{in}\right)^2+\frac{1}{2}\left(V_{OUT}+v_{out}\right)^2\right], \tag{16}$$

where V_P is the value of v_P given by eqn.(15) when the switch is on.

The current i_{SW} also flows through the capacitor C and, consequently, in steady state condition, its mean value $\overline{i_{SW}}$ is zero, i.e., from (16),

$$V_P\left(V_{IN}-V_{OUT}\right)-\frac{1}{2}\left[V_{IN}^2+\overline{v_{in}^2}-V_{OUT}^2-\overline{v_{out}^2}\right]=0 \tag{17}$$

Where $\overline{v_{in}^2}$ and $\overline{v_{out}^2}$ are the mean-square values of the RFI voltages v_{in} and v_{out} respectively. Eqn.(17) is solved in order to calculate the DC output voltage, V_{OUT}, and the RFI-induced offset $\Delta V=V_{OUT}-V_{IN}$ is consequently expressed as

$$\Delta V=\left(1-\sqrt{1+\frac{\overline{v_{in}^2}-\overline{v_{out}^2}}{\left(V_P-V_{IN}\right)^2}}\right)\left(V_P-V_{IN}\right) \tag{18}$$

The mean-square values of the RFI voltages $\overline{v_{in}^2}$ and $\overline{v_{out}^2}$ in (18) can be expressed in terms of the RFI voltage which is superimposed onto the external input by frequency domain linear analysis. In particular, in the case of continuous wave (CW) RFI,

$$\overline{v_{in}^2}=\frac{V_{in}^2}{2} \quad \text{and} \quad \overline{v_{out}^2}=\frac{1}{1+\omega^2 r_{ON}^2 C^2}\frac{V_{in}^2}{2}, \tag{19}$$

where V_{in} is the peak amplitude of CW RFI and, from (14),

$$r_{ON}=\left.\frac{\partial v_{AB}}{\partial i_{SW}}\right|_{v_A,v_B=V_{IN}}=\frac{1}{n\beta\left(V_{IN}-V_P\right)} \tag{20}$$

Furthermore, from linear analysis, the root-mean-square (rms) value of the error term δV in Eqn. (13) can be expressed as

$$\delta V_{rms} = \sqrt{\overline{v_{out}^2}}, \tag{21}$$

It can be observed that Eqn.(18) and (21), which have been derived with reference to the basic SC circuit in Fig. 6a, are valid for any MOS closed switch under the hypothesis that $\overline{i_{SW}} = 0$, which is true for any switch in practical non-degenerate SC networks. On the basis of ΔV and δV, on-switches can be replaced as shown in Fig. 6b and the RFI-induced errors in complex SC circuits can be predicted by baseband SC circuit analysis.

3.3 Prediction of RFI-Induced Errors in Complex SC-Circuits

The errors which are induced by RFI in a generic, non-degenerate, SC circuit can be predicted on the basis of the results about MOS switch distortion, that have been derived in the previous Section, and with the opamp model, which is described in [3].

To this purpose, during each clock phase, the RFI-induced errors ΔV and δV in closed switches should be calculated and they should be replaced with the equivalent circuit shown in Fig. 6b. The effect of RFI on the operation of any opamp included in the circuit should be taken into account adding an input offset voltage source to each opamp, which is affected by RFI.

The circuit, which has been obtained, is equivalent in terms of RFI-induced errors to the original SC circuit and it can be analyzed by baseband SC circuit analysis techniques [5]. On the basis of such an analysis, the nominal SC circuit output voltage(s), the RFI-induced offset in the output voltage(s), and the rms value of the RFI-induced fluctuations in the samples of the output voltage(s), which depend on the phase of RFI waveform with respect to the edges of the clock waveform, can be predicted.

The approach, which has been outlined above, is applied now to the unit-gain fully differential SC voltage amplifier in Fig. 8 and the predictions which have been obtained are compared with the results of time-domain computer simulations.

In particular, the unit gain SC amplifier of Fig. 8 has been designed referring to a 0.8 μm CMOS technology process with a power supply voltage of 5V. It includes a fully differential two-stage Miller opamp circuit,

poly capacitors and nMOS switches that are driven by non-overlapping square wave phase signals φ_1 and φ_2 with a frequency of 100 kHz and with rising and falling edges of 1ns. A DC common mode input voltage $V_{CM} = 2.5\,\mathrm{V}$ and a nominal DC differential voltage $V_{CM} = 200$ mV have been applied to the amplifier inputs and CW RFI with a frequency in the range 10MHz-1GHz has been superimposed onto the non-inverting input voltage as shown in Fig. 8. With reference to this particular configuration, it can be observed that RFI is never superimposed onto the opamp input terminals and the opamp circuit does not contribute to RFI-induced distortion.

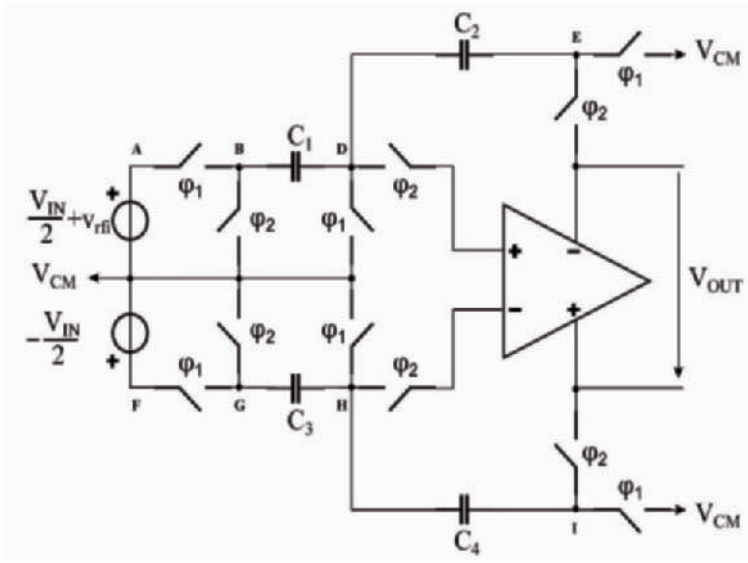


Fig. 8. Fully Differential SC Amplifier.

The circuit in Fig. 8 has been simulated by Spectre, a Spice-like circuit simulator, for different amplitudes and frequencies of CW RFI. For each value of RFI amplitude and frequency, eight different phases of the CW RFI waveform have been considered, eight time-domain simulations have been performed and the errors in the amplifier output voltage due to RFI have been evaluated.

In Fig. 9, such errors are plotted with reference to a 500MHz CW RFI signal with a peak amplitude of 1V and the contribution of the RFI-induced offset ΔV_{OUT} and of the phase dependent error δV_{OUT} are evidenced.

The value of ΔV_{OUT} and the rms value of δV_{OUT} have been estimated by calculating respectively the average and the standard deviation of the output voltages which have been obtained from the simulations for different phases of RFI. Such analysis has been repeated for different amplitudes and frequencies of RFI performing extremely time-consuming transient computer simulations and the value of the RFI-induced offset error ΔV_{OUT} and the rms value of δV_{OUT} have been compared with model predictions.

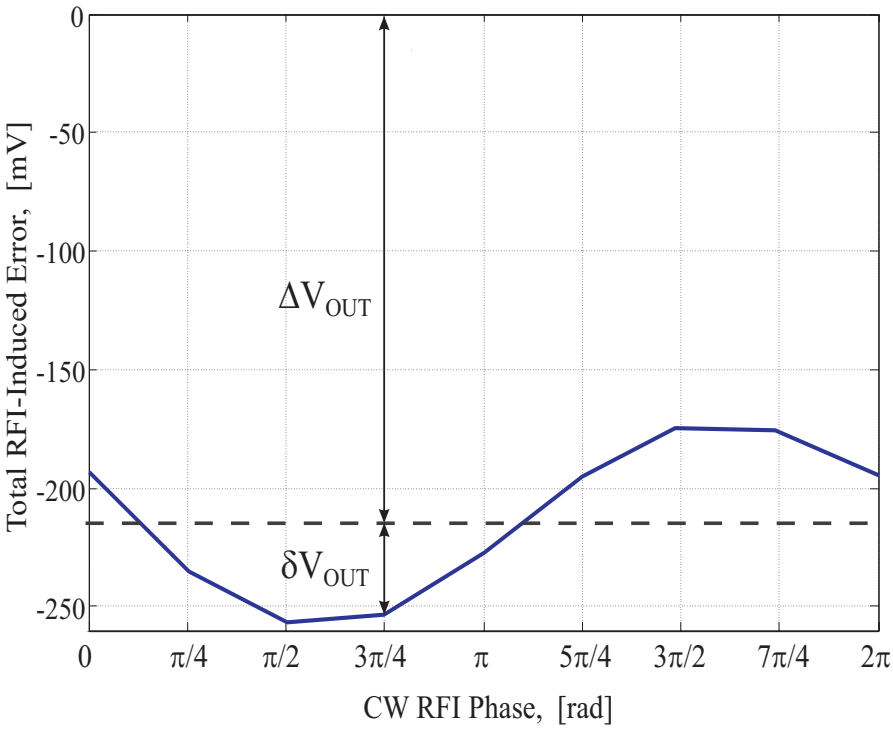


Fig. 9. RFI-Induced Errors Vs. CW RFI Phase (computer simulation).

In Fig. 10, in particular, the RFI-induced offset ΔV_{OUT} has been plotted versus the peak amplitude of a 100MHz CW RFI signal, in Fig. 11 the RFI-induced offset ΔV_{OUT} has been plotted versus frequency for a CW RFI signal with a peak amplitude of 1V and finally, in Fig. 12 the rms value of the error δV_{OUT} has been plotted versus frequency for a CW RFI signal with a peak amplitude of 1V. In all these plots, model predictions are shown in solid line while crosses represent computer simulation results. It can be observed that model predictions are in good agreement with time-domain

computer simulation results, as a consequence, the model which is presented in this paper is suitable to describe the effects of RFI in the operation of SC circuits.

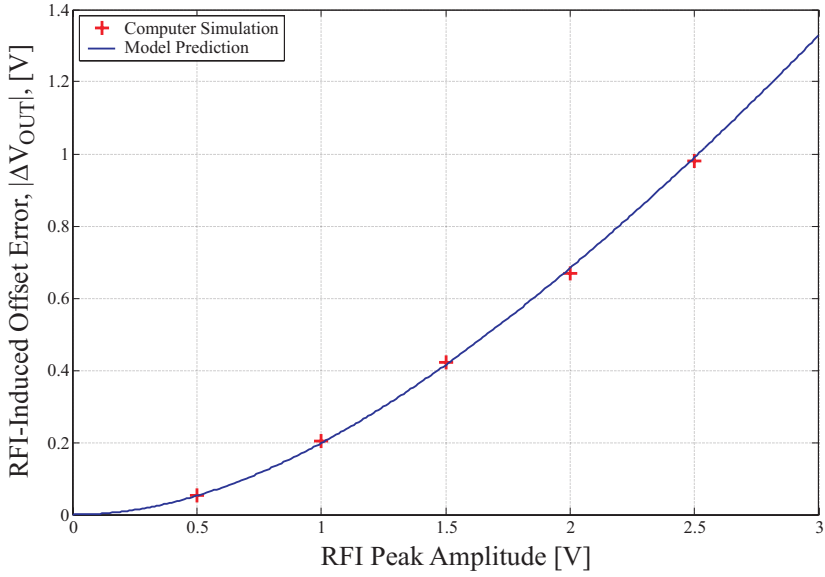


Fig. 10. RFI-Induced Offset ΔV_{OUT} *Vs. CW RFI Peak Amplitude* V_{in}.

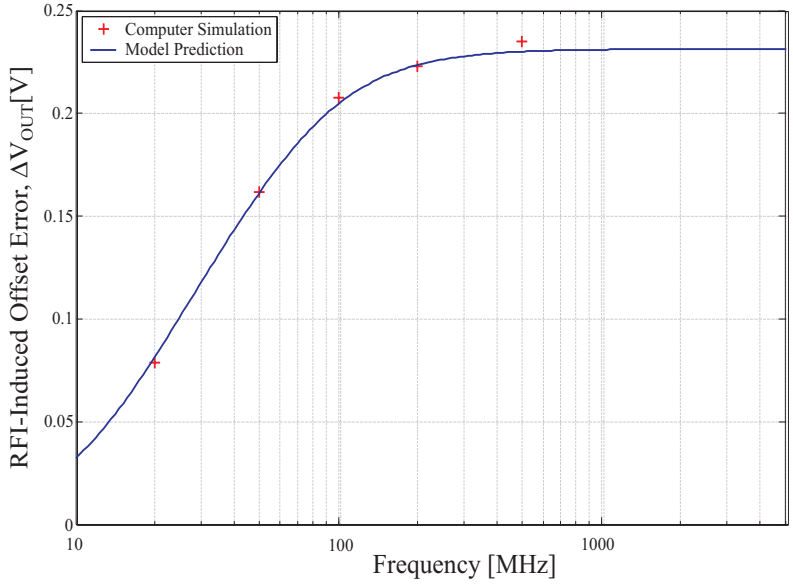


Fig. 11. RFI-Induced Offset ΔV_{OUT} *Vs. CW RFI Frequency.*

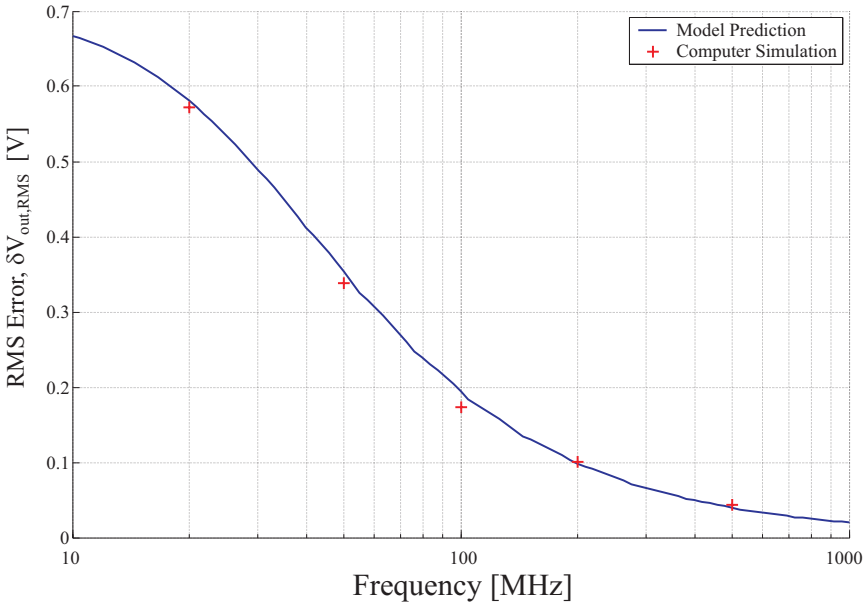


Fig. 12. Root-mean-square value of ΔV_{OUT} *Vs. CW RFI Frequency.*

10. Conclusions

In this chapter the last research achievements on the susceptibility to EMI of integrated analog front-end have been presented. In particular, the effect of RFI on the nominal performance of feedback CMOS opamps has been presented and a new input stage immune to RFI has been described. Furthermore, the susceptibility to RFI of switched capacitor front-ends has been discussed and a new nonlinear model of MOS switches to predict the demodulation of RFI has been presented and it has been employed in a SC amplifier.

Acknowledgement

This work was partially supported by MIUR - PRIN 2005.

References

[1] ——"Integrated Circuit Electromagnetic Susceptibility Handbook", Tech. Rep. MDC E1929, McDonnell Douglas Astronautic Company, St. Louis, MO, 1978.

[2] J. G. Tront, J. J. Whalen, C. E. Larson, J. M. Roe, “Computer-aided analysis of RFI effects in operational amplifiers”, IEEE Trans. On EMC, vol. 21, n.2, Nov. 1979, pp. 297-306.

[3] F. Fiori, P.S. Crovetti, “Nonlinear Effects of Radio-Frequency Interference in MOS Operational Amplifiers”, IEEE Trans. on Circ. and Syst. I, vol. 49, n.3, Mar. 2002, pp. 367-372.

[4] C.C. Enz, F. Krummenacher, E.A. Vittoz, “An Analytical MOS Transistor Model Valid in All Region of Operation and Dedicated to Low-Voltage and Low-Current Application”, in Analog Int. Circ. and Sign. Proc. Journ. vol. 8, n.7, July 1995, pp. 83-114.

[5] P.S. Crovetti, F. Fiori, “Nonlinear Effects of RFI in Switched Capacitor Circuits”, accepted for publication as regular paper on TCAS I (preview available on IEEEXplore)

IMPROVED ELECTROMAGNETIC IMMUNITY CIRCUIT DESIGN

Derek Bernardon
Infineon Technologies
Villach, Austria

Abstract

The design of integrated circuits for automotive applications has to meet challenging EMI requirements. In order to determine the cause of most common failures it is necessary to perform research in order to improve concept, design and layout of new or present products. In this article the concept/design analysis and results of fundamental blocks such as comparator/opamp, internal voltage supply and bandgap are presented.

1. Introduction

The objectives of EMI research for improved circuit design are:

- Reduce R&D costs by reducing number of redesigns due to Electro-Magnetic Compatibility
- Implementation of simulation techniques that give insight concerning potential weaknesses in circuit design and layout
- Determination of root causes in concept, circuit design, and layout for most common failures
- Determination of "best practice" to be used during concept, design and layout to avoid EMC issues, especially due to immunity

To achieve difficult EMI specifications for a system IC, it is crucial for the fundamental biasing blocks to be robust, otherwise they will limit the performance of the whole IC. This paper presents concept/design analysis and EMI results for a comparator/opamp, internal voltage supply, and a bandgap.

2. Direct Power Injection

In order to gain a deeper understanding of immunity issues, it is necessary to perform DPI (Direct Power Injection) tests, which permit to identify failures in an easier manner, because it permits to inject RF power into a single pin at a time as illustrated in Fig. 1. Standard: ISO 11452-7(12-95); IEC 62132-3(01-00) [1].

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 203–217.

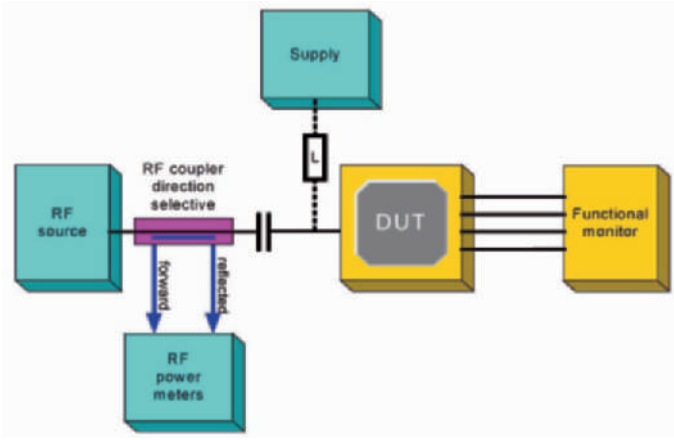


Fig. 1. DPI Measurement Setup.

The DPI test consists of injecting RF power, through a coupling capacitor, at a regulated constant forward power. This power is then increased until the monitoring unit does not report a failure, once the first failure is detected then the program automatically steps up the frequency and searches again at which RF power level the DUT (Device Under Test) fails. The DUT consists of the application module with the IC under examination and all its' external components.

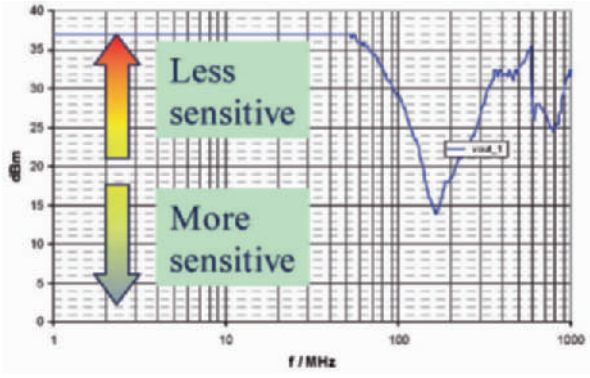


Fig. 2. Results of a DPI Test.

Figure 2 illustrates a typical result of a DPI measurement, where 37dBm are equivalent to 5W forward power injected into a pin of the module.

3. Coupling Mechanism of RF Power Inside the IC for SPT Technology [2]

Depending on the product there are various types of packages, which may have a significant impact on EMI results. As illustrated in Fig. 3 the RF substrate currents have different paths for the different type of packages and ground connections.

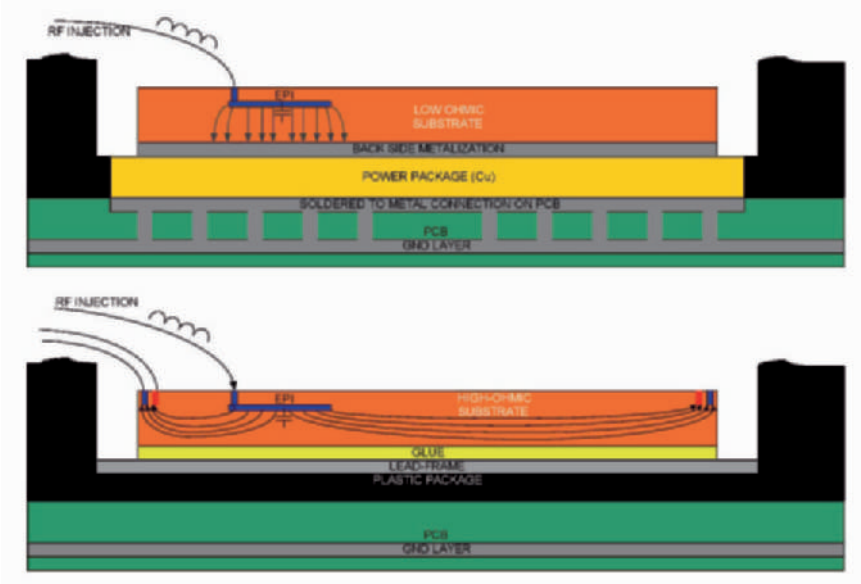


Fig. 3. RF Current Flow Depending on Package.

RF substrate currents can differ significantly from the DC ones, because of the fact they see different impedances than exclusively the resistive ones. A bonding wire has significant impedance at high frequencies, and in many products the ground/substrate connection is made via a bonding wire, causing the ground connection to have higher impedance, which causes the RF substrate currents to flow more through the substrate looking for other lower ohmic connections to ground. In such a case placing devices at farther distances does not have a significant impact on EMI performance. Although there is an electrical connection to the lead frame, which is externally connected to the PCB ground, the glue, and the fact that the backside of the die has no metallization layer, seems not to permit significant conduction of RF substrate currents, thus the majority will find its way back through substrate connections or through the circuits, which is what should be avoided. However when a power package is used with backside metallization then better EMI performance can be achieved due to the fact that RF

substrate currents have low inductive and resistive paths to ground. This does not mean that they will not affect neighbouring devices or blocks, however distance does in this case prove to be effective in improving EMI performance.

3.1. EMI Coupling for Comparator/Opamp

On a testchip (Fig. 4) DPI tests were performed, and the effect of the substrate currents was observed on a small comparator located within the red circle. The comparator is not electrically connected to the power DMOSs, except for channel 4, which input is directly connected to the drain Q4 (Fig. 4). Fig. 5 illustrates the immunity performance of the testchip where the output of the comparator alone was monitored, and RF power was injected into the drains of the four different channels (Q1 to Q4), one at a time. It is possible to observe in Fig. 5 just how the substrate currents affected this block. This IC had neither backside metallization nor a power package. And it can be observed how the RF currents injected at channel 3, which is located completely on the other side of the IC, affect the comparators' performance in Fig. 5.

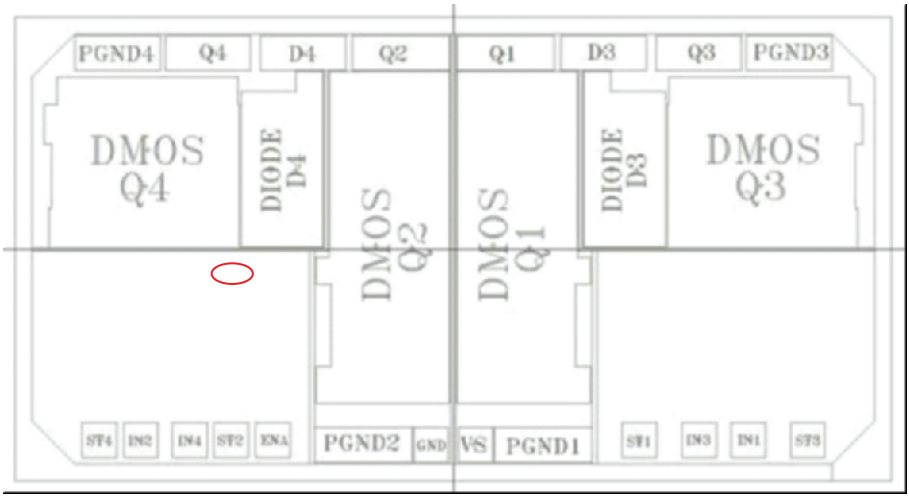


Fig. 4. Layout of Testchip.

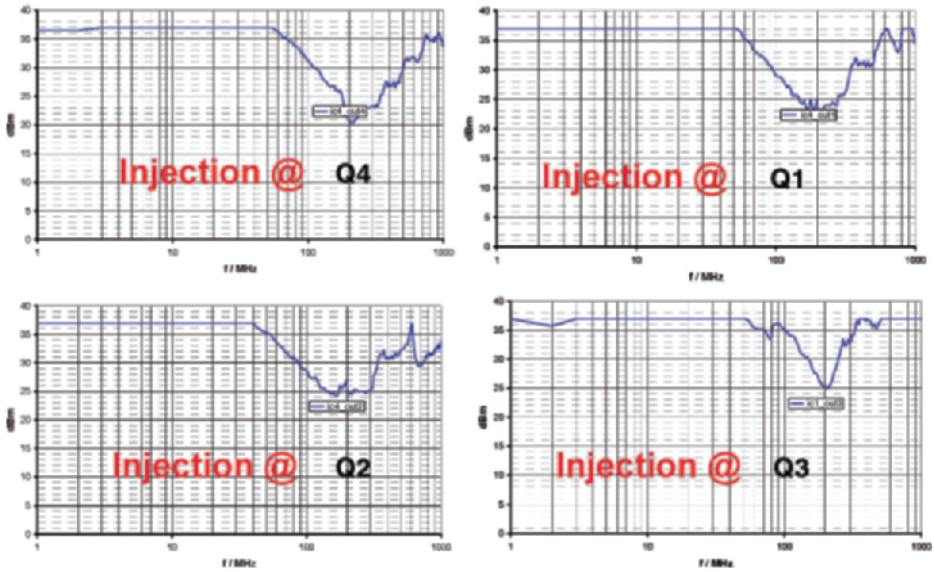


Fig. 5. DPI Results for Comparator at Different Injection Points.

The frequency range where substrate currents are sufficiently large to disturb neighbouring blocks is from 40MHz to 1GHz. It is interesting to observe how channel 3 although being adjacent to the border of the die, where the P+ ring and N-sinker ring are located, nevertheless injects current throughout the whole substrate, when one would expect that the majority of the current would have the tendency to flow into these rings that are right beside the device having a low resistive path to ground. Results however show that a significant amount of current still flows throughout the whole substrate, although some part of the current does flow into the rings. It would almost appear as if the current almost flows radially outwards from the power DMOS into substrate.

The dominant coupling mechanism is evident in this case; Fig. 6 illustrates the cross-section of the devices that are involved.

Since the drain of the DMOS is the EPI, it has the best parasitic capacitance to substrate, which makes it the perfect coupling mechanism for RF currents. The comparator uses PNPs as its' input stage, where the base is the EPI, which couples in RF substrate currents and amplifies them since its'

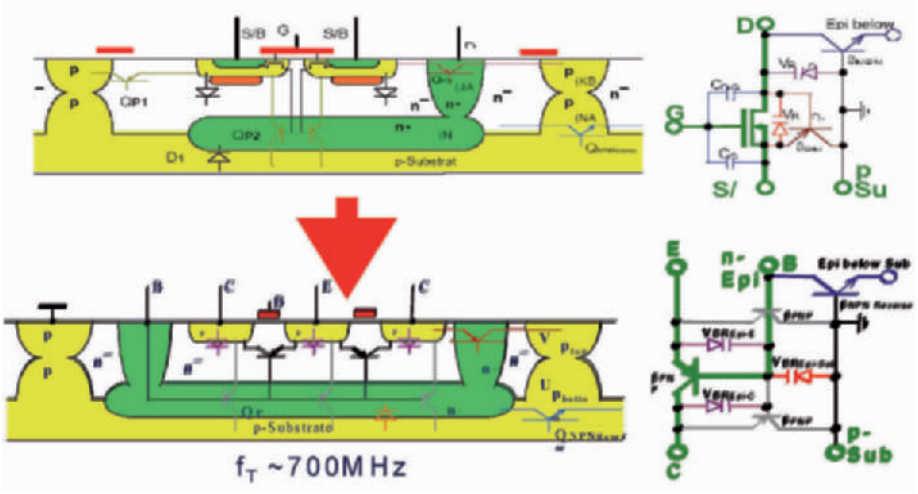


Fig. 6. RF Power Coupling Mechanism through Substrate.

transient frequency is 700MHz. However it has been observed that when modifications are made to improve reverse current effects, these will also improve EMI performance, it is just that for EMI more factors need to be taken into consideration. To avoid that RF substrate currents dangerously affect your IC it is necessary for all EPIs to be connected to low ohmic nodes. To improve EMI performance on this testchip a P+ substrate ring was introduced around the PNP input stage.

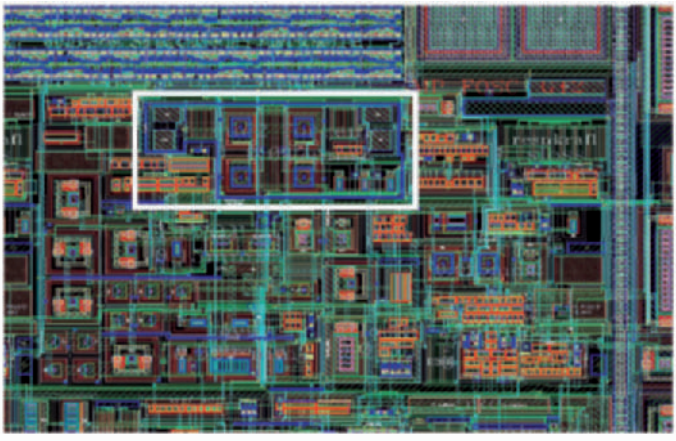


Fig. 7. P+ Ring around input PNP structure.

And a low pass filter was introduced (Fig. 8) to improve the performance at low frequencies, since the performance at high frequencies (greater than 100MHz) is determined by the RF substrate currents in this specific case. The filters are necessary since the comparator is directly connected to the pin where RF power is injected to. In Fig. 8 it is possible to observe two filters, the first filter is needed to reduce the RF signal that appears at the drain of the power DMOS when it is in high impedance mode. The second filter serves two purposes, the first is to additionally filter the RF signal, and the second is to transform the RF signal into a common mode signal so that the comparator may reject it.

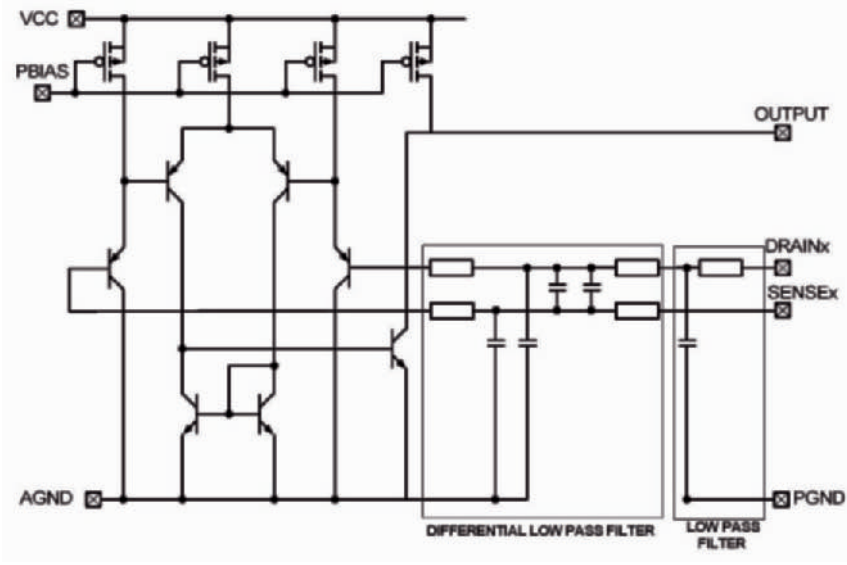


Fig. 8. Filtering for Comparator/Opamp directly connected to injection point.

Figure 9 illustrates the results where it is possible to observe the improvement due to the low pass filter and P+ ring. The low pass filter improved the low frequency direct coupling, and the P+ ring reduced the RF substrate current injection into the input stage at high frequencies.

4. Internal Voltage Supply

The internal voltage supply is a critical block, because it supplies the most sensitive blocks of a system IC, therefore if it does not have a good

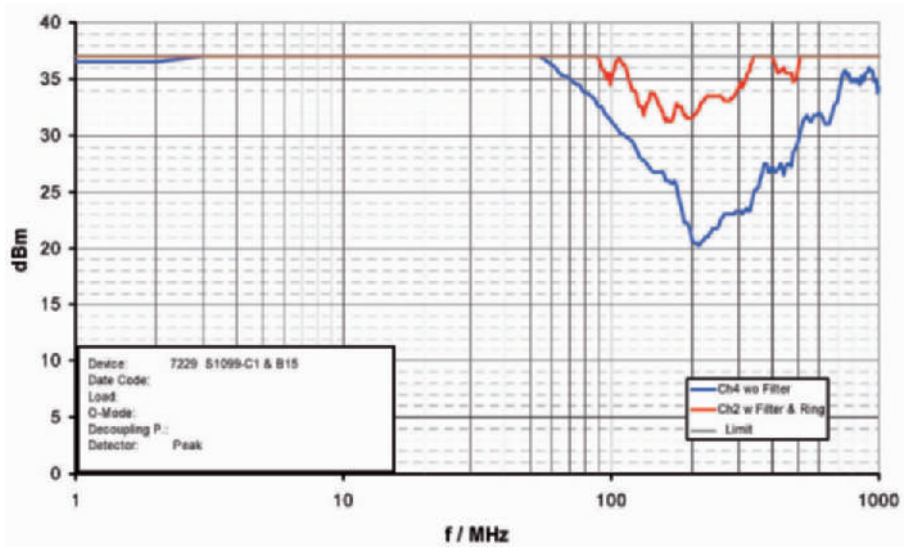


Fig. 9. DPI results.

immunity, this single block will limit the performance of the entire IC. The question is what is the best concept to use in order to achieve maximum EMI performance?

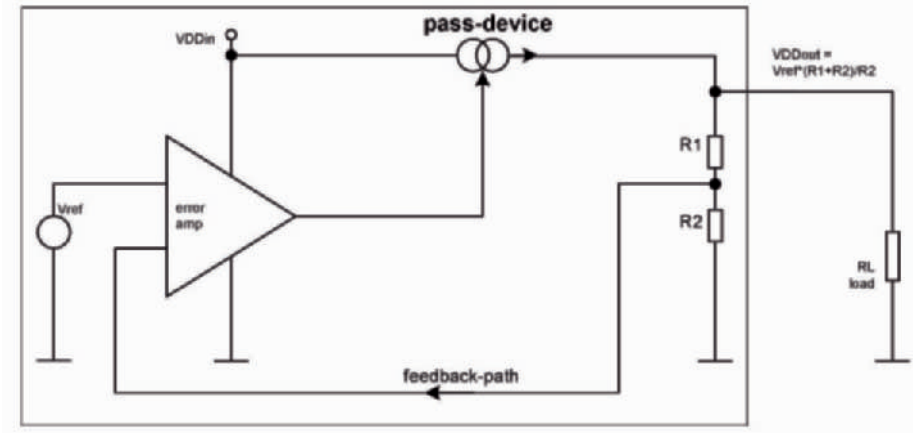


Fig. 10.

First it is necessary to determine which pass device is the best for the internal voltage supply in order to obtain best EMI performance, assuming it will be directly connected to battery, thus tested with 5 Watts of injected RF power.

Regulation Device	Advantage	Disadvantage
PMOS & DMOS in parallel	■ Low Vds drop due to > Vgs ■ No charge pump required ■Good SOA	■ Larger than DMOS w/ Charge Pump ■ Stability when PMOS is active ■ Only limited reverse polarity possible ■Bad EMC performance
DMOS	■ Low output impedance ■ Low Vds drop possible ■ Good SOA	■ Charge pump required→RF Emissions ■ No reverse polarity operation due to body diode unless back-back concept used → Area
NPN	■ Good stability ■ Good load response ■ Lowest output impedance	■ does not operate at very low Vce ■ No reverse polarity operation due to low Veb breakdown voltage ■ Low Vceo limits max. operating range ■ SOA weak
PNP	■ Very low Vec drop ■ Reverse polarity protection possible	■ Very slow and large device ■stability

Fig. 11. Pros and Cons of various output stages.

The most promising devices are the DMOS with charge pump and DMOS in parallel with a PMOS, so a testchip was designed with the following concepts in order to verify which one had the best EMI performance. Fig. 12 is a concept that was previously used with DMOS and charge-pump, Fig. 13 is a concept that was also previously used with an output DMOS in parallel with a PMOS, and Fig. 14 illustrates a “new” concept that is designed specifically to improve EMI performance. Concepts 1 and 3 have large capacitors at both gate and source of the DMOS to improve immunity, but concept 3 has an additional start-up circuitry, which is connected directly to the battery, and in steady state the internal voltage regulator then supplies all biasing structures. Small transfer gates are used to isolate biasing and control blocks from the start-up circuitry, in order to avoid direct coupling from the battery.

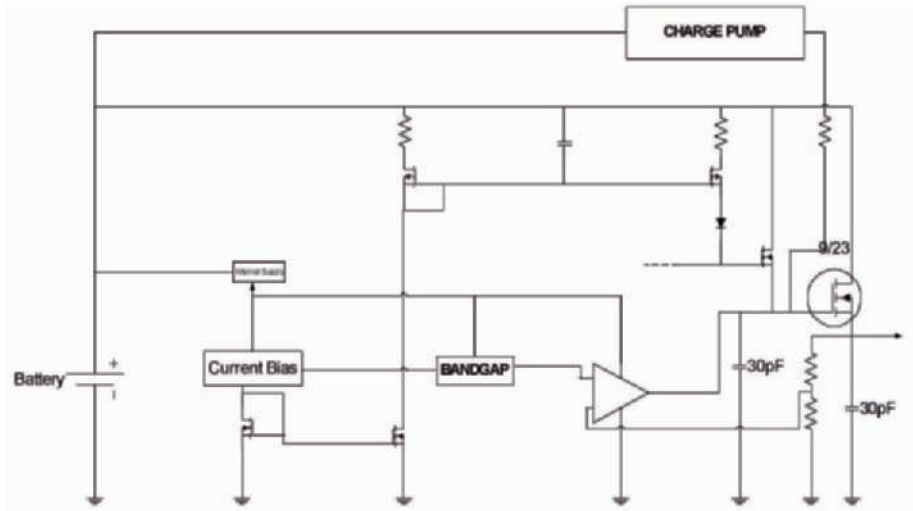


Fig. 12. Concept 1.

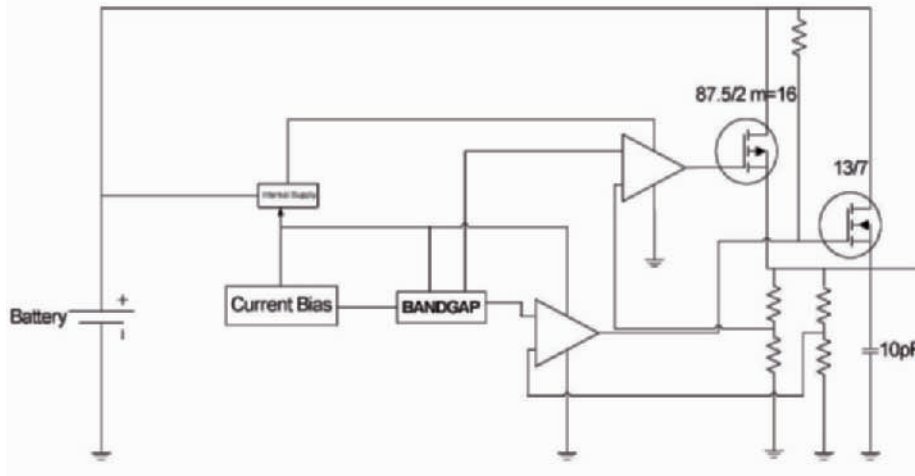


Fig. 13. Concept 2.

This way it provides “protection” for the current biasing and bandgap, by using the linear regulator as a RF filter, and has significantly less capacitance than concept 1.

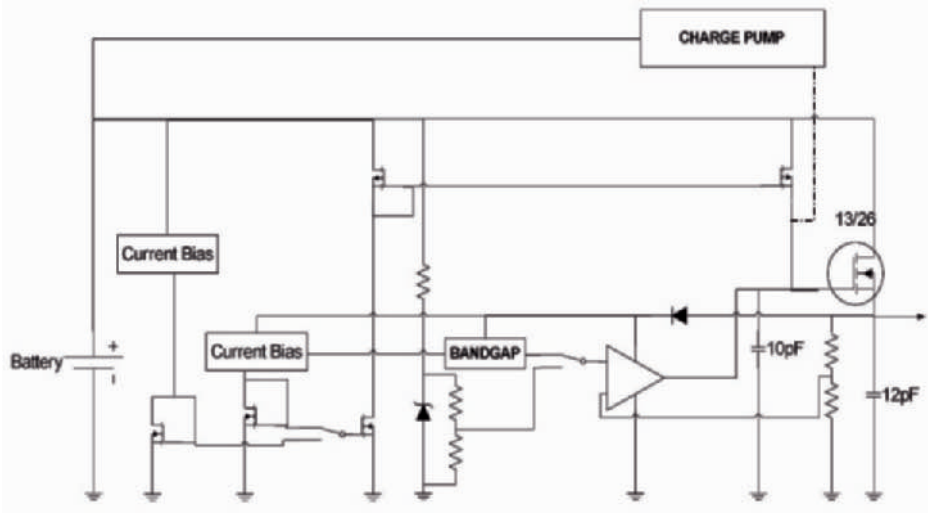


Fig. 14. Concept 3.

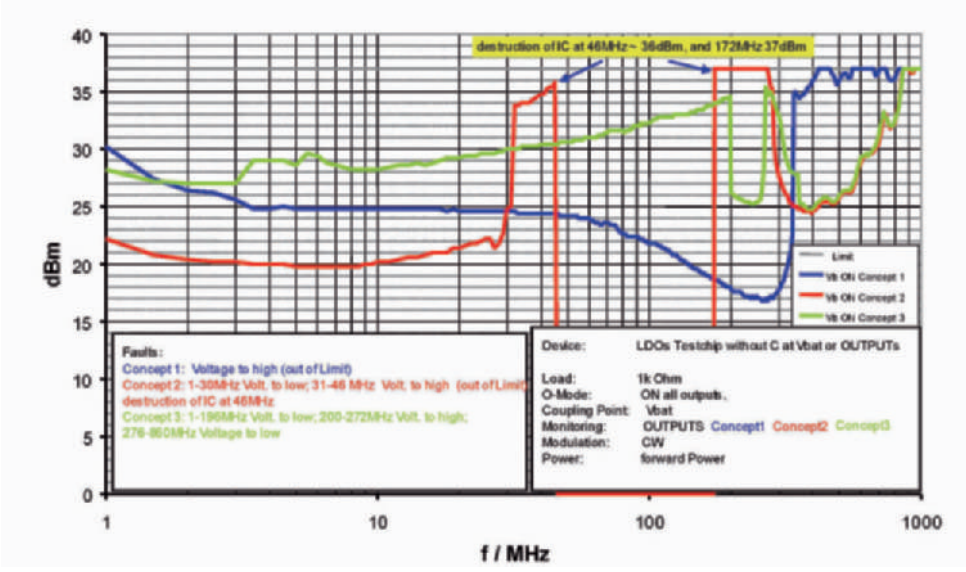


Fig. 15. DPI Results.

In Fig. 15 the EMI results show that concept 3 performs better than the other two, although it is smaller in area, with a minimum of 25dBm. These tests were performed without any capacitance at the battery pin. Concept 2 with a PMOS in parallel with a DMOS, had a failure which led to the destruction of the IC. Concept 1 has a strong weakness in the medium frequency range, with a minimum of 17dBm at 200-300MHz. Concept 3 has

the best performance in the low-medium frequency range because it has the least amount of RF power that reaches the gate of the power DMOS. Once this was understood a small modification was made to concept 3 to improve its' performance as illustrated in Fig. 16.

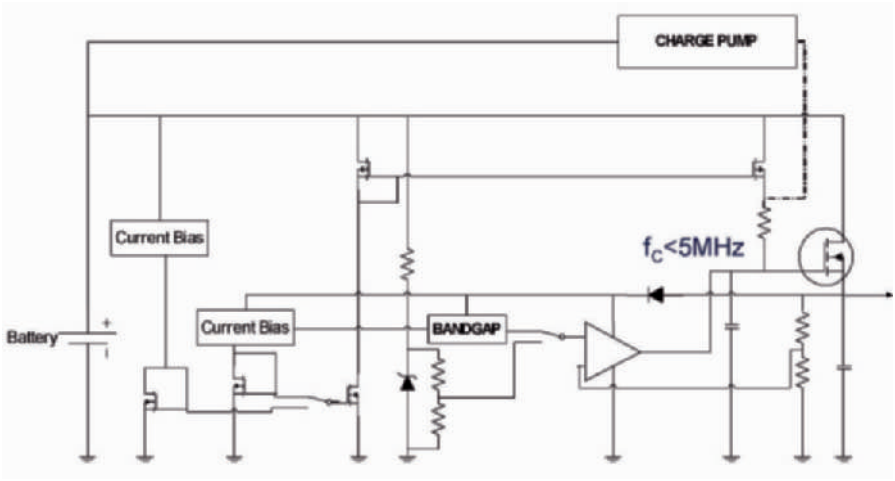


Fig. 16. Improved Concept.

By inserting a simple resistor in series to the PMOS a passive filter is formed which attenuates further the introduction of RF power into the gate of the power DMOS. It is also important to place a P+ ring around the DMOS to reduce the emission of RF substrate currents.

5. Bandgap

In Fig. 17 the bandgap [3] shows that all the EPIs are connected to the internal power supply. This reduced the effect of reverse DC currents, which also is the best approach for improving EMI performance. However the internal voltage supply does not have low impedance over the whole frequency range up to 1GHz, and therefore causes the local supply of the bandgap to have a superposition of a significant ac component due to the RF currents which are flowing through the EPI connections of the transistors (shown with green dots in Fig. 17). Due to the fact the line may have significant parasitics, it is important to place a capacitor near the bandgap (C4 Fig. 18), or else the series resistance and inductance of the metal and/or poly lines will reduce the filtering effect of the capacitance at high frequencies, making the insertion of the capacitor useless.

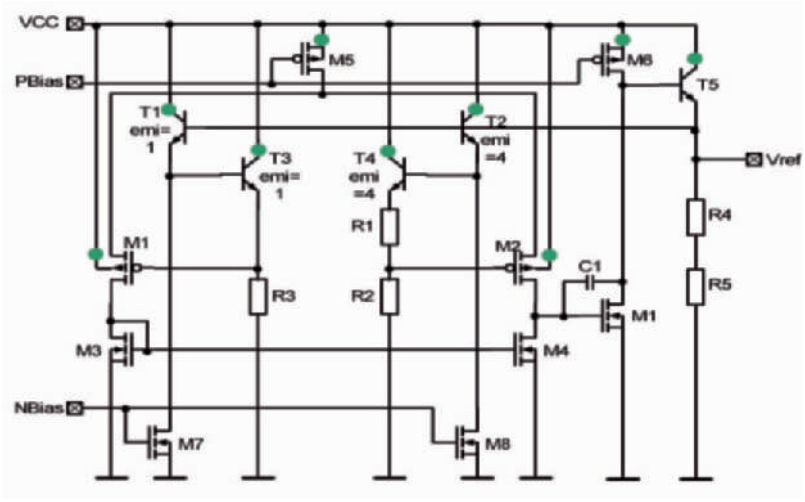


Fig. 17. Bandgap.

After analysing the circuit in simulation under RF power injection at the supply of the bandgap it is possible to observe just how bad the attenuation is from power supply to output voltage for frequencies above the bandwidth of the opamp (Fig. 19). In order to improve the attenuation for higher frequencies it is necessary to introduce some capacitances as illustrated in Fig. 18 (C2, C3, and C4).

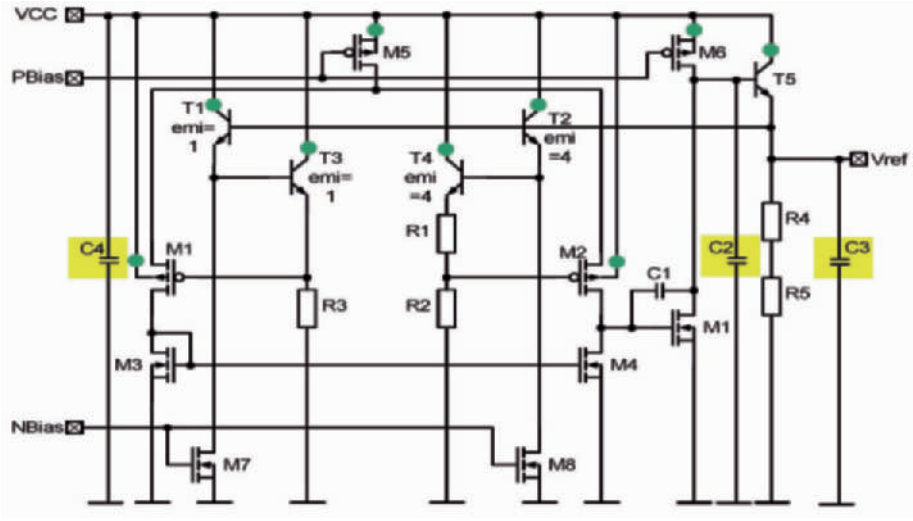


Fig. 18. Bandgap Optimized for EMI.

In order to properly simulate the effect of the RF currents introduced via the EPI wells into the supply line, it is necessary to include the supply line inductance, resistance, and approximate impedance of the output resistance of the internal voltage regulator, and emulate this RF substrate current with an ac current source from the local VCC to ground. Fig. 19 illustrates the difference of the two different concepts.

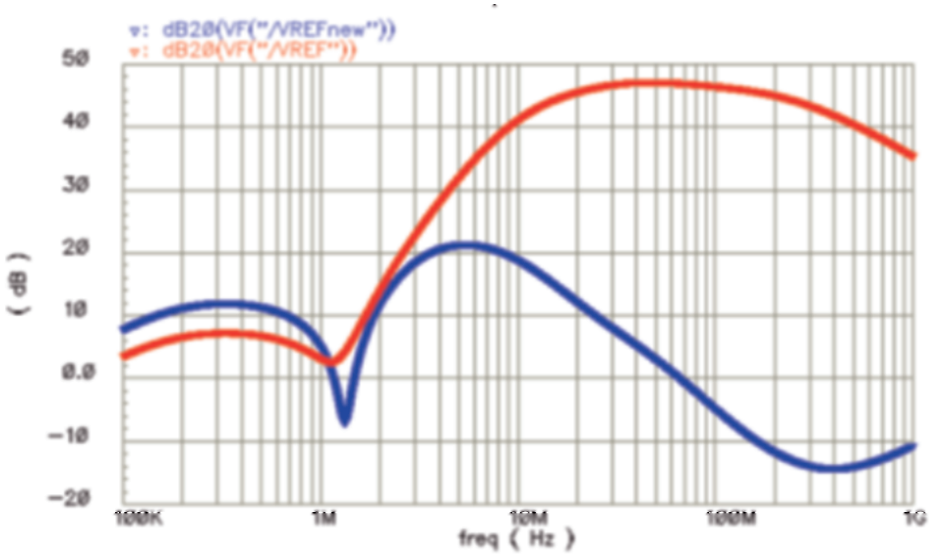


Fig. 19. Bandgap RF current to Output Voltage transfer Function.

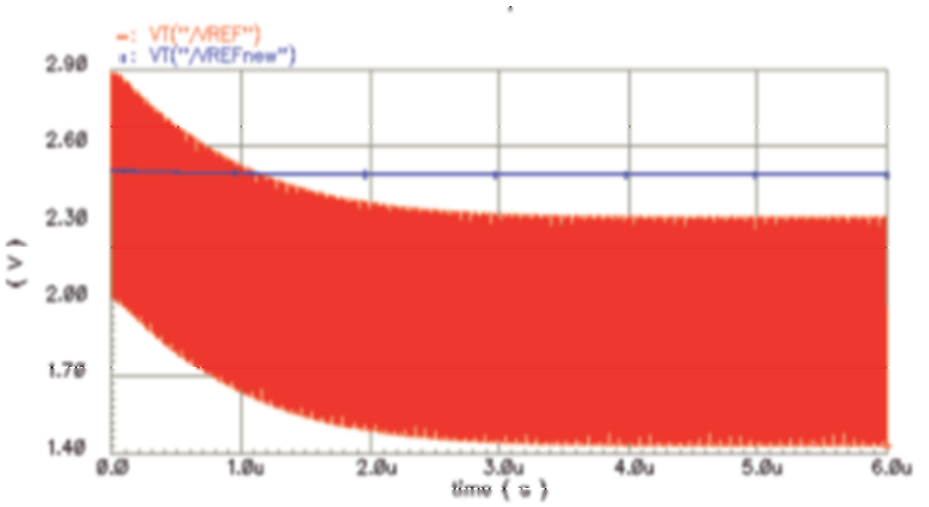


Fig. 20. Bandgap Transient Results.

An additional reason why the capacitance at the supply is so important is that the input structure of the bandgap is unsymmetrical, which causes a phase shift among the signals that reach the opamp, causing a DC shift as shown in Fig. 20.

Both the DC level variation and the amount of coupled RF power to the output voltage are improved in the new concept.

10. Conclusions

This paper presented results of concept/design analysis for a comparator/ opamp, internal voltage supply, and bandgap for obtaining better EMI robustness. It also highlights the root causes of failures and specific coupling mechanisms for SPT technology, and explains how simulations can be performed for EPI coupled RF substrate current injection, and then proposes solutions to improve their performance. The purpose of this paper was to highlight just how important it is to design all fundamental blocks in a robust manner against RF substrate currents and conducted emissions so that these would not limit the system performance.

References

[1] http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CS-NUMBER=38302&ICS1=43&ICS2=40&ICS3=10

[2] M. Stecher et al., “Key Features of a Smart Power Technology for Automotive Applications”, Conference on Integrated Power Systems CIPS 2002, Bremen, Germany

[3] H. Zitta, W. Horn, C. Lenzhofer, “Automotive IC-Design”, Analog Circuit Design, Kluwer Academic Publisher 2004 pp.137-151

METHODOLOGY AND CASE STUDY FOR HIGH IMMUNITY AUTOMOTIVE DESIGN

Aarnout Wieers, Herman Casier

AMI Semiconductor, Belgium
Aarnout_Wieers@amis.com

Abstract

This paper presents a structured methodology that, despite it is not conclusive for all cases, can support a design engineer to develop and/or debug electronic circuits for electromagnetic susceptibility (EMS). This methodology is well suited to investigate continuous time analog electronic systems, modules or circuits and does not require any special tools. The method is flexible, can be applied in many cases by simple reasoning and is customizable towards specific needs.

1. Introduction

In modern automotive and industrial environments, EMC (Electro Magnetic Coupling) requirements rapidly increase. Realizing high EMC immunity (high EMI or low EMS), combined with realizing low EMC emissions (low EME) are crucial differentiators for OEMs to select electronic components and systems. Therefore component suppliers make large investments in design, measuring and debugging techniques to optimise the EMC performance of their electronic systems and components and to maintain their competitive advantage. Several design guidelines exist to support low EMS design but due to its complex nature, debugging EMS on system or component level stays a difficult problem. For automotive applications the EMS spectrum starts at 150KHz and extends up to 1GHz with tendency to further increase. Parasitic components are often poorly known and can show large variability. Correct modelling is a real challenge and more pragmatic methods are used. This paper introduces a structured and effective methodology that uses standard AC analysis and a set of intuitive rules and criteria to help a design engineer in designing and/or debug an electronic circuit for low EMS. Despite that this method is not conclusive in all cases, we believe that it offers an effective instrument to detect and cure EMS weaknesses in many cases.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 219–237.

2. EMC Properties

EMC is dominated by four characteristic properties: 1) the level of injected power, 2) the range of EMC frequencies injected, 3) the amplitude modulation characteristics used to modulate the EMC power and finally 4) the pass/fail criterion used to evaluate the circuit immunity for EMC. Pass/fail criteria are typically defined by two properties: 1) the maximum allowed deviation of a given parameter. If a parameter exceeds this value, this is considered as a failure. And 2) an EMC power level for which no failure may occur.

To save high measurement and rework costs on the finished module, component suppliers use different types of pre-compliance tests. The most common used pre-compliance tests are direct power injection (DPI)(6) and bulk current injection (BCI). In the examples presented in this paper, DPI will be assumed but the same methods can be applied using BCI.

3. Origins of EM Susceptibility

EMC disturbances can enter into electronic circuits via different electrical paths. The majority of these paths are defined by parasitics components and couplings. Parasitics introduced by PCB tracks, cable harness, bonding wires, leads, substrate coupling etc… are not well known but can play an important role in injecting EMC into a circuit. Especially these couplings and parasitics make EMS a problem that is difficult to grasp and complex to analyze.

Linear circuits can attenuate EMC signals but cannot introduce signal asymmetry, which contains an additional DC component. Or in other words, a linear attenuated signal stays linear and the average of the resulting signal stays equal to zero. On the contrary, active components in general behave non-linear. E.g. a MOS is characterized by its quadratic law in strong inversion and a bipolar or diode by its exponential law. Applying EMC on such a non-linear component, translates the symmetrical input signal into an asymmetrical output signal. This effect is called rectification. If such not-symmetrical signal is filtered, by means of a capacitor, charge accumulation or pumping will occur and a DC shift is introduced, which can de-bias the circuit out of its operational region.

Classifying EMS problems into failure groups, the most common reasons that lead to low EMS are charge-pumping effects, coupling effects, circuit mismatches and resonance. In many cases, coupling, matching and resonance effect are just initial reasons that finally lead to rectification and pumping. Two types of rectification can be distinguished: soft and hard rectification. With soft rectification we indicate that the device under EMC stress shows a modulated DC operating point, which is not large enough to cut off the device. Hard rectification on the other hand will periodically or continuously cut off a device when EMC stress is applied. To illustrate rectification in more detail, a

source follower input buffer as depicted in Figure 1 is studied. Current source Ibias together with R1 and C1 represent a non-ideal current source biasing of the follower Q1.

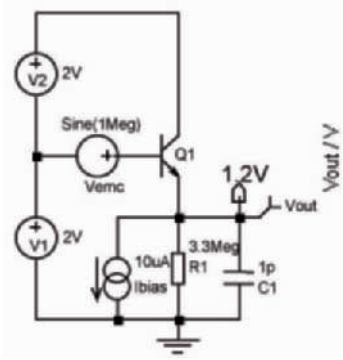


Fig. 1. Source follower circuit.

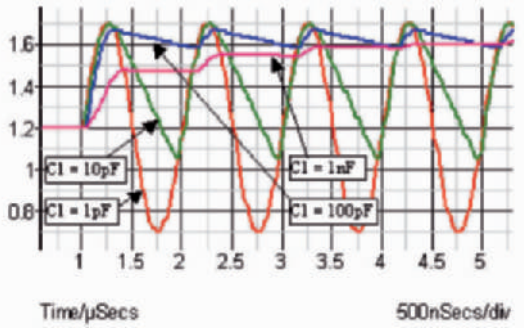


Fig. 2. Source follower charge pump effect.

3.1. Large-Signal Analysis

A large signal approach is used to analyze the source follower circuit in figure 1.

Case 1: C1 close to zero: In order to evaluate the EMS behavior, a sine wave voltage source is added in series with the base of the source follower Q1. Using large signal considerations, the emitter of Q1 will fast charge the circuit node OUT but will not discharge the same node. Discharging takes mainly place over Ibias and a neglectable part over R1. Since C1 is close to zero, the current source Ibias is sufficiently strong to discharge the node OUT completely within every sine wave period. Even with an EMC signal of 1 volt, the DC operating point of Q1 stays quasi constant. Despite this large signal amplitude and the exponential relationship between the collector current and the base emitter voltage of Q1, the follower output remains to behave linear. The lower waveform depicted in figure 2 shows the follower output with C1 = 1pF.

Case 2: C1 is replaced by a dominant capacitor >> 1pF: Due to the exponential behaviour of Q1, charging of C1 does not fundamentally change. This can be seen in figure 2 just after the delay and when the first sine wave period starts. However as soon as the input signal goes down and C1 starts to discharge, the emitter of Q1 slowly discharges while the base follows the fast EMC signal. Or in other words, the bandwidth of the emitter node is too low to follow the base signal. Dependent on the applied EMC amplitude, Vbe will vary, strongly vary or even switch off Q1 while the emitter node slowly discharges. The DC operating point of Q1 is no longer maintained and the non-linear transfer curve

of Q1 becomes visible. The capacitor C1 accumulates this non-linear voltage and the output OUT shows a positive DC shift. This effect is called charge pumping. The three upper curves in figure 2 show this effect in different gradation for the capacitor values C1=10pF, C1=100pF and C1=1nF. Besides a DC shift, the sine wave amplitude is reduced (limited bandwidth) and the signal shape is distorted. Note that charge pumping always moves into the direction to cut off the non-linear device that creates it. The peak value of the sine wave tends to remain the same as the peak value of the sine wave without charge pumping.

3.2. Limitations of Large-Signal Analysis

The approach used above to analyze the follower circuit is based on large signal considerations but lacks to be conclusive when large frequency bands need to be considered. Standard SPICE transient simulations offer a good method to analyze circuits for large signal behaviour. For simulation purposes, EMS is often modelled by means of direct power injection (DPI), typically realized by an AC coupled HF sine wave voltage source connected to the pin under test. Transient simulations are performed in the time domain. This has two consequences: 1) In order to recognize charge pumping, multiple periods of accumulation of charge are usually required. And 2) transient simulations are limited to one frequency at a time. Consequently, we risk leaving certain frequency dependencies such as narrow band feed forward couplings and resonance chains with high Q undiscovered. To avoid this, the complete EMC spectrum (typically 150KHz to 1GHz) should be simulated and many transient simulations would be required. Limited design time and quite long simulation times do not allow performing simulations in an exhaustive manner. Furthermore, it is not straightforward to take into account the large spreads that several parasitics tend to have. Monte Carlo analysis can offer a solution here but again at the cost of simulation time. Another theoretical method that drastically could improve simulation times while being able to recognize non-linearity is a sensitivity analysis. However two major drawbacks make this type of simulation unattractive: a) Sensitivity analysis is by nature a DC type of analysis that excludes all frequency dependencies including capacitors and inductors. This loss of frequency information makes its usage questionable. b) Sensitivity is only checked for one DC operating point. As EMC tends to be large signal, in general multiple sensitivity simulations will be required, which is again not attractive. We therefore continue our investigation and explore alternative ways.

3.3. Small-Signal Analysis

EMS is by nature defined in the frequency domain. AC analysis is well suited to evaluate circuit behavior in the frequency domain and in many cases, AC analysis can be performed intuitively; but can such an intuitive approach be used to predict EMS? The following paragraph will answer this question.
The components found in an analog circuit can be classified in 4 groups:
1) The EMC input source: EMS signals are induced by electro magnetic fields picked up by some kind of antenna; for example a cable harness or a metal track on a PCB. The currents and voltages induced in this way can be very large and are mainly determined by the physical characteristics of the antenna and by frequency dependent impedance of the circuit node. Pre-compliance tests such as DPI typically use a HF sine wave voltage source to model the induced EMC signal. This source is AC coupled to the circuit node under test and has a source impedance of 50 ohm. An AC analysis is well suited to analyze both the HF voltage source and the impedance over frequency of a circuit node. There however is still one problem. Impedance alone is not enough to define the ability of a circuit to handle current. To illustrate this, we use again a follower circuit but now inject the EMC signal on the emitter follower output as depicted in figure 3. For low frequencies the follower Q1 can source high currents but sinking current is strongly limited by the current source Ibias and its output impedance R1. Despite this limited sink capability, the AC output impedance, with no EMC signal applied, is found to be low. This is due to the linearization of the follower characteristics in the bias point made in an AC analysis. In this case, this largely overestimates the AC output impedance for currents sourced to the output OUT and underestimates the AC output impedance for currents sinked from the output OUT. The reason is found in the asymmetrical current capability of the circuit. Assume we apply DPI to the follower output. During the positive half-period of the DPI sine wave, the DC bias current in Q1 will diminish or even switch off Q1 and the impedance on the pin OUT will drastically increase. This effect can not be recognized from a small signal AC analysis. Still it is this limited current capability, rather than the small signal AC impedance, that will determine the final capability for the circuit to take up EMC power. Fortunately, this limitation is in general only valid for the node on which we directly inject EMC. Once the EMC signal has entered a branch deeper inside the circuit, in general, no direct path exists towards the external EMC source. This implies that the maximum voltages and currents are no longer defined by the EMC source itself but are now defined by the circuit topology and the circuit supply. Or in other words, starting from this branch EMC can be considered as a superposed HF sine wave signal in series with the branch input rather than a power source that injects from external. This important observation allows using AC analysis inside circuits in order to predict EMS problems.

2) Linear components: Linear components such as resistors, capacitors and inductors attenuate a signal but the resulting signal remains linear. No rectification occurs and no DC shifts are introduced. Therefore an AC analysis is well suited to describe the behavior of these elements even when large signals are applied.
3) Linear equivalents: Elements like a PCB track, a cable harness, a bond wire etc… can be approximated with simplified HF models. In general these networks are linear and can be modeled with resistors, capacitors and inductors. AC analysis is again well suited to describe these circuits, also for large signals.
4) Non-linear components: Non-linear components such as diodes, bipolars and MOS transistors show a non-linear transfer characteristic. Most of these devices have a voltage input and a current output. The output current flows into the load impedance, where it is again translated into a voltage. This voltage drives the next stage or branch in a circuit and the process repeats. Non-linear components not necessarily create non-linear signals. As an example take a classical current mirror circuit consisting of 2 transistors, a bipolar input diode and a second, identical bipolar, connected parallel with base and emitter to the first bipolar. When a linear sine wave voltage is superimposed on the base-emitter input of the bipolar diode, an exponential collector current will be generated in both transistors. If this exponential current of the output transistor is injected into a linear load-impedance, e.g. in a resistor. The exponential current is converted into an exponential voltage and rectification will occur with a risk for charge pumping. If instead the exponential collector current of the output transistor is injected into a diode (e.g. the incoming branch of another bipolar current mirror) then the exponential current is first converted by the logarithmic law of the diode again into a linear voltage and no rectification nor a DC shift will occur at this point. This basic but essential observation leads to the conclusion that as well the transfer characteristic as also the load impedance are crucial elements to determine charge-pumping effects. In most cases, simple reasoning is enough to recognize the linear and non-linear transfer characteristics in a circuit. By means of AC analysis the behavior of a node-impedance over frequency is determined. When combining both techniques together, this can lead to an efficient method to predict charge-pumping effects taking all frequencies into account.

3.4. Defining Pass and Fail Criteria

Up to now, rectification has been studied without defining a pass/fail criterion. Soft and hard rectification has been discussed and it has been shown that the amount of rectification depends as well on the amplitude of the input signal as on the non-linearity of the chosen topology. But how can this all be

linked to the EMS pass/fail criteria defined for a circuit? One way to approach this problem is by using initial and maximum impedances. The technique is demonstrated by means of the emitter follower as depicted in figure 3. The output impedance on the node OUT is given by:

$$Zout(f) = \frac{1}{Gm + \frac{1}{R1} + j\omega C1} \tag{1}$$

To evaluate the circuit during EMC conditions, an additional sine wave current, Iemc, is injected into the node OUT. The frequency of the sine wave is chosen to be well below the bandwidth of the circuit. In the positive alternance of the sine wave, the DC current in the follower will reduce proportionally to the injected EMC current. Neglecting R1 and C1, the follower will remain to have a bias current as long as IemcMax < Ibias and equation (1) will hold, where:

$$Gm = Ic / \frac{kT}{q} \approx 40 * Ic \tag{2}$$

When IemcMax >= Ibias, Q1 will have no bias current and the transistor will cut off. Recalculating the impedance on the node, now with Q1 cut off becomes:

$$ZoutMax(f) = \frac{1}{\frac{1}{R1} + j\omega C1} = \frac{R1}{1 + j\omega R1C1} \tag{3}$$

In figure 4, the upper curve shows the impedance on node Out with Q1 cut off and the lower curve the impedance with Q1 biased. Or in other words, we have found the initial output impedance, Zout(f), when no EMC is applied (lower curve) and the maximum output impedance, ZoutMax(f), when a large EMC signal is applied (upper curve).

From figure 4, three information's can be deducted: a) The output impedance changes by up to three orders of magnitude. b) For each frequency the initial and maximum impedance can be retrieved and the circuit susceptibility can be found. c) For high frequencies, C1 becomes dominant in defining the output impedance and both curves converge.

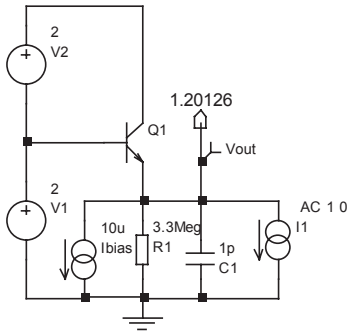


Fig. 3. Emitter follower circuit.

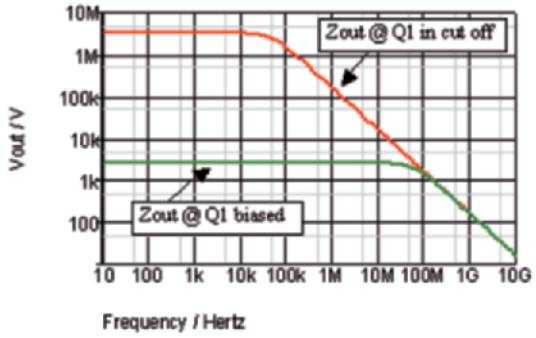


Fig. 4: Follower output impedance.

For low frequencies, at the fail criterion, when Q1 just cuts off, IemcMax = Ibias. From this and the maximum impedance, given by equation (3), the maximum EMC power the circuit can handle is calculated as:

$$P = I^2 R \longrightarrow PemcMax(f) = Ibias^2 * ZoutMax(f) \tag{4}$$

For higher frequencies, the impedance of C1 comes in parallel with the current Ibias and hence PemcMax(f) increases accordingly.

Using a fail criterion, together with the maximum impedance, we where able to deduct the circuit EMC susceptibility over frequency. But how to determine the EMC susceptibility if only a parameter shift of a few percents is allowed? Applying a more strict fail criterion is a solution. Instead of allowing Q1 to cut off, we require Q1 to maintain a minimal bias current larger then a certain percentage of the DC bias current. This will limit the non-linearity and thus the rectification effect. For Zout >> 50Ω (the DPI source impedance) Vemc is equal to the applied DPI signal and:

$$Itot = Ibias + Iemc \quad \text{and} \quad Vbe(tot) = Vbe(Ibias) - Vemc \tag{5}$$

With Itot, the total bias current in Q1, Ibias, the DC bias current and Iemc, the additional EMC current injected. Vbe(tot), the total base-emitter voltage of Q1, Vbe(Ibias), the DC bias current in Q1 and Vemc the applied DPI voltage. As Q1 is a bipolar, Ic = Is*exp(Vbe/Ut), with Ut = kT/q. Substitution in equation (6) results in:

$$\frac{Itot}{Ibias} = \frac{Ibias + Iemc}{Ibias} = e^{\left(\frac{Vbe(Ibias) - Vemc}{Ut} - \frac{Vbe(Ibias)}{Ut}\right)} = e^{\left(\frac{-Vemc}{Ut}\right)} \tag{6}$$

Rewriting and taking the average by integrating over one sine wave period, the DC shift introduced by the non-linear transfer curve of Q1 is given by:

$$1+\frac{Iemc}{Ibias}=e^{\left(\frac{-Vemc}{Ut}\right)}\longrightarrow 1+\frac{I\_mean}{Ibias}=\frac{\omega}{2\pi}\int_0^{\frac{2\pi}{\omega}}\left(e^{\left(\frac{-Vemc*Sin(\omega t)}{Ut}\right)}\right)dt \qquad (7)$$

With I_mean the DC current shift as a result of charge-pumping. In figure 5, 1+I_mean and I_mean are depicted as a function of Vemc/Ut. Note that for soft rectification, the current DC shift, I_mean, created by charge-pumping can be approximated as:

$$I\_mean\approx\left(\frac{Vemc}{2*Ut}\right)^2 \qquad (8)$$

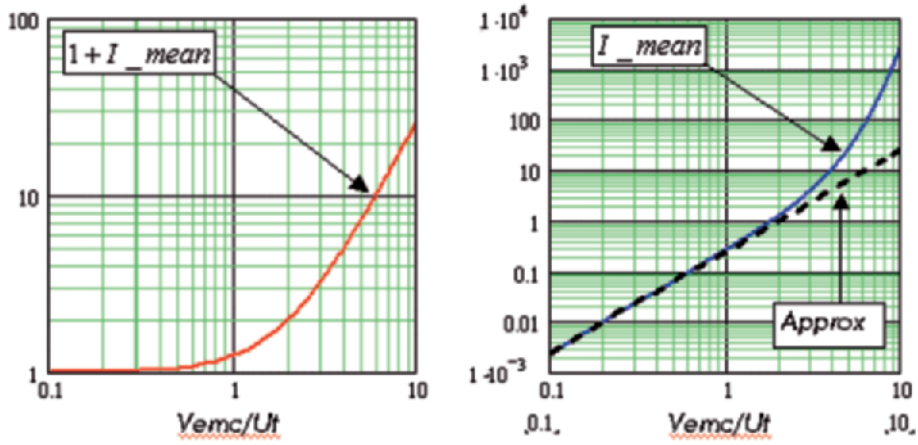


Fig. 5. The DC current shift due to charge-pumping, 1+I_mean and I_mean.

The method and equations above demonstrate that as well for hard as for soft failure criteria the circuit susceptibility over frequency can be deducted and it was sufficient to combine AC analysis with a set of well-known DC circuit properties. Generalizing this concept, it seems plausible that for any continuous time system we can use this technique based on AC analysis to predict EMC sensitivity over frequency. Compared to modern electronic systems, the complexity of a follower circuit is of course small and can be not considered as representative. In the following section the method is therefore applied on a more complex circuit. In order to demonstrate the flexibility of

the method, a CAN receiver input is not only analyzed, but also design and optimized for EMS.

4. Design of a Low EMS CAN Receiver

As 2nd example a CAN receiver is designed. The CAN receiver requires to operate up to 1Mbit/sec and with a propagation delay below 200nsec. A high speed, low gain comparator input stage as depicted in Figure 6 is selected as starting point.

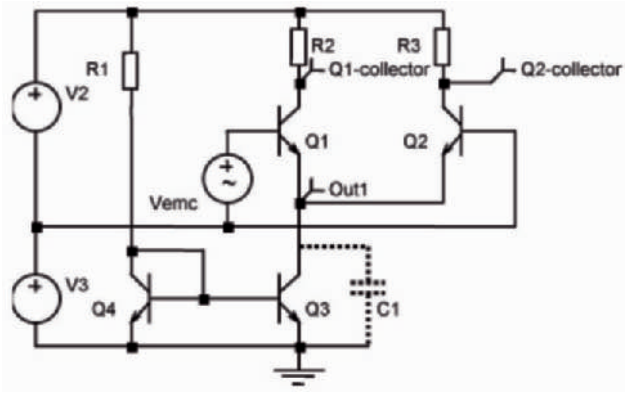


Fig. 6. Schematic of the CAN receiver input stage.

Step 1: Definition of the EMC requirements: EMS robust design means no rectification and/or no accumulation. The elements that can introduce non-linear behaviour are the two bipolar devices, Q1 and Q2, forming a differential pair. (Q1 and Q2 could also be replaced by MOS devices). As EMC failing criterion, we choose that one or both of the input bipolar devices show a clear shift in DC operating point. This can happen under three conditions: 1) when the input common mode signal exceeds the input compliance voltage range and 2) when the bandwidth of the common emitter node (virtual ground) of the differential pair is far below the EMC frequency applied. 3) When the input differential signal exceeds the linear input compliance voltage range.
Step 2: Impedance and AC analysis: The pole on the virtual ground (common emitters) of the differential pair is expressed by:

$$F_{-3db} = \frac{Gm}{2 * \pi * C1} \tag{9}$$

where C1 stands for the collector capacitance of the bias current mirror. Assuming C1 = 1pF and Ic = 10uA, the pole is calculated at ~63MHz. For frequencies above this pole, the virtual ground of the comparator input stage

becomes too slow to follow the input signal. The common emitter voltage rectifies and charge pumping occurs. Figure 7 shows the comparator outputs with C1 = 1pF (no charge pumping) and C1 = 100pF (charge pumping).

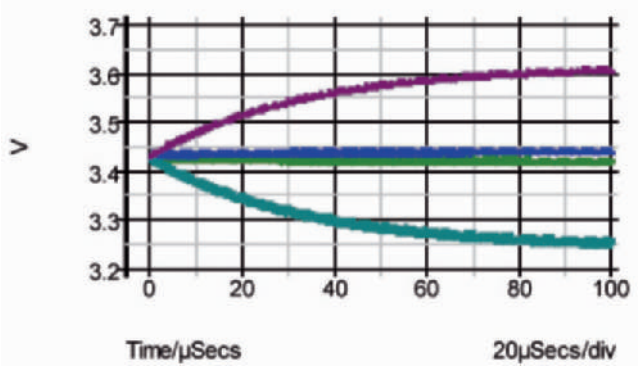


Fig. 7. Rectification for C1 = 1pF and C1 = 100pF @ F=10MHz.

To avoid charge pumping, two options exist: 1) the incoming signal is attenuated by means of a resistive divider or by means of a low pass filter with cut-off frequency well below the comparator pole or 2) the bandwidth of the comparator stage is increased beyond the bandwidth of the incoming EMC signal. Increasing the bandwidth has three major disadvantages: a) it results in a higher current consumption. b) It makes the circuit more sensitive to high-frequency noise and disturbances. c) The EMC requirements for following circuit branches of the comparator stage increase. As a general rule, it is therefore preferred to limit the incoming signal bandwidth rather than to increase the circuit bandwidth, whenever possible. The CAN receiver in this example requires to operate at a speed of 1Mbit/sec with a settling time of less than 200nsec. These requirements directly define the lower limit for the comparator bandwidth. To keep the circuit complexity low, a passive first order RC filter is added in front of the comparator inputs. The settling time of 200nsec requires a minimum bandwidth of about 15MHz (3 tau). When integrated on silicon, the component spread of an RC filter is in the order of 50%. This moves the –3dB point to a typical value of 30MHz. The attenuation of a 1st order LPF filter is limited to 20dB/decade. 20dB attenuation is only reached for a the filter bandwidth above 300MHz, while the comparator bandwidth is limited to 63MHz. It's clear that all these constraints together push to move to high circuit bandwidths, which for obvious reasons are undesired.

A better solution can be found by introducing attenuation by means of a resistive divider. The attenuation of 20dB is valid for all frequencies and therefore relaxes the input range and linearity requirements of all subsequent stages. The LPF, with a typical –3dB point at 30MHz, is also added to further

reduce the high-frequency requirements of the subsequent stages. Designing for typically 30MHz with 50% component variations gives a worst case –3dB point at 45MHz. Since this is very close to the pole of the comparator, we here choose to also boost the bandwidth of the virtual ground of the comparator differential stage. This is done by simply replacing the bias current source with a bias resistor. Compared to the current bias, the resistor has a much lower impedance and a significant smaller parasitic capacitance. Both properties help to move the comparator -3dB point easily by a factor 3. Drawback of the resistor bias is the soft rectification introduced by the common mode dependency of the comparator bias current and its non-linear effect on the comparator gain.

Step 3: Input impedance and dominant external components: The receiver inputs are connected to the CAN bus. The CAN bus is terminated at each end with a resistor of 120ohm and two capacitors of 220pF to ground. The value of these capacitors is restricted by the CAN settling time requirement of 200nsec. The filter input impedance is assumed to be always much higher than the 120 ohm termination resistor. This observation leads to two important conclusions: a) For frequencies below the first pole nearly all EMC power is injected into the two 120 ohm bus terminators + the 50ohm EMC source impedance. With 4 Watt EMC power injected in 2 times 120 ohm in parallel, the expected RMS signal amplitude is:

$$Vrms(f) = \sqrt{Zout(f) * Pemc(f)} = \sqrt{4W * 120\Omega // 120\Omega} = 15.491V \qquad (10)$$

Looking to the components connected to the bus, immediately two reasons are recognized that can lead to rectification: a) the ESD protection diodes will clamp and b) the compliance of the input common mode is exceeded. This compliance is limited to only 1.25V due to the required gain of the first stage and the input signal dependent bias variations. Adding large external capacitors is not feasible, as this would limit the transmission speed and settling times. Adding a lower external shunt resistor will increase the current consumption and limit the number of CAN nodes that can be connected on a bus. Adding series inductance is not effective for low frequencies and besides that also expensive. One possible solution left is to use a resistive divider that attenuates the input signal for all frequencies. The minimum attenuation required is given by:

$$Attenuation = \frac{VemcMax}{VInputCM} = \frac{15.491V * \sqrt{2}}{1.25V} \sim 17.53 \qquad (11)$$

And second the LV ESD protection diodes need to be replaced by symmetrical HV ESD protection devices that do not conduct when +22Vp or –22Vp is applied on the CAN lines. These ESD protections are represented by the double zener diodes in figure 8.

Attenuating the input signal by a factor 10 has clear technical disadvantages, e.g. increased offsets (the input offset is multiplied by the attenuation), high sensitivity to the divider mismatches (divider mismatches translate a common mode signal into a differential mode signal), a lower signal to noise ratio etc… Nevertheless, in this specific case, the possible solutions to deal with EMS are very limited. Without looking for alternative solutions, which is beyond the cope of this paper, it is here accepted to make this tough compromise on all parameters in favour of avoiding charge pumping. The final CAN receiver circuit solution is depicted in figure 8. The biasing current source is replaced by a resistor R1 and an input divider R7-10 and an input LPF R9-10/C7-8 are added.

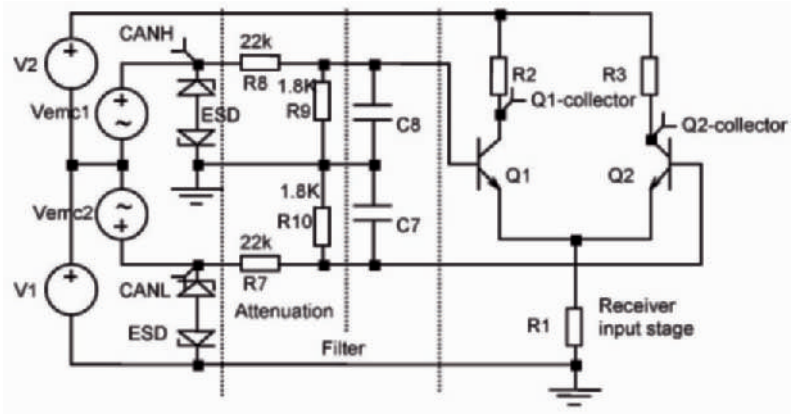


Fig. 8. EMS optimised CAN receiver stage.

Step 4: Checks for resonance: Resonance circuits exist everywhere. In many cases resonance chains are present in the form of a combination of PCB tracks, cabling, parasitics, couplings etc. This observation makes it complex to come to an exact analysis and engineers tend to use a heuristic approach. Figure 9 depicts a basic model of the CAN bus including cabling and the PCB. On both sides of the CAN bus, line termination resistors of 120ohm and two capacitors are present. In the middle of the CAN bus, we inject common mode the AC signal and inside the ASIC we enter on the resistive divider and LP filter. For simplicity, the input impedance of the comparator input stage was neglected.

The AC simulation result is depicted in figure 10. The curve labeled "original" shows a resonance at about 80MHz with an amplitude of +2dB, even after the resistive divider and Low Pass Filter. This resonance presents a risk and the circuit is again reviewed. Lowering the –3dB point of the LPF is not an option. Making the LPF a second order filter will only bring a small improvement at the cost of additional area. Re-evaluating the CAN bus model, the 80MHz is mainly defined by C1, C2 and L3. Lowering C1 and C2 or lowering L3 will move the resonance to higher frequencies but the resonance amplitude will increase.

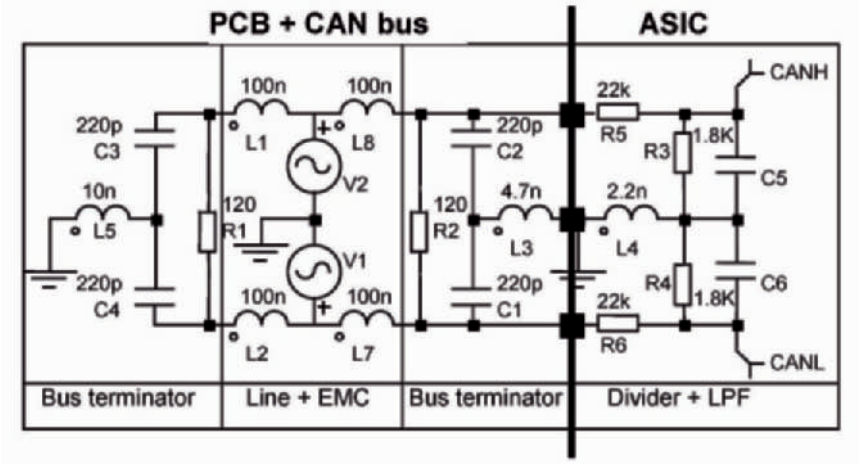


Fig. 9. Basic model of the CAN bus.

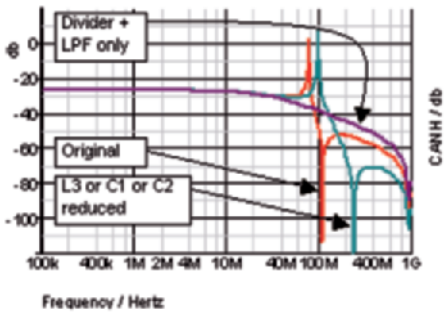


Fig. 10. CAN bus: resonance.

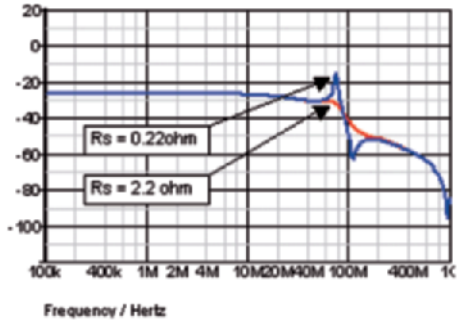


Fig. 11. CAN bus: damped resonance.

Therefore moving the frequency with one decade, which due to physical PCB limitations is difficult to achieve, gives only a small improvement. A much better technique is to introduce damping to drastically lower the resonance amplitude.

We can realize this by adding a small series resistor in the LC loop, for example in series with L3 to ground or in series with both capacitors C1 and C2. Figure 11 shows the impact of damping for 2 resistor values, Rs=0.22ohm and Rs=2.2ohm. The small resistor values make it easy to realize this without affecting the overall circuit performance. In many cases the parasitic resistance of the capacitor will be sufficient.

Step 5: Verification by transient analysis: To confirm the final solution, a large signal transient simulation using DPI is performed. The pole and impedance information deducted from small signal analysis is used to define the

frequencies to run transient simulations for. In this example, 5 frequencies are selected:
1) The minimum and maximum required EMC frequencies. These frequencies define the boundaries of the specification.
2) The resonance frequency, the most vulnerable frequency in this circuit, justifies an explicit check. Some soft rectification is expected.
3) The typical comparator pole frequency. At this frequency the comparator has to operate without rectification.
4) The typical LP filter frequency. At this frequency the additional attenuation of the LPF starts to act.

Transient simulations have been used to proof that no rectification occurs. To keep the picture readable, Figure 12 only demonstrates 3 of the 5 frequencies.

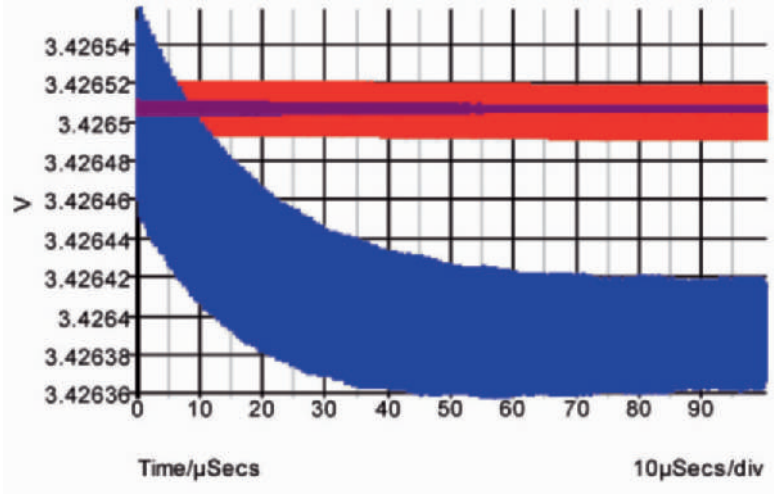


Fig. 12. CAN receiver: Transient simulation results of V(Q1_collector) with Femc = 30MHz (LPF), 100MHz (cmp pole) and 77.288MHz (resonance).

The two upper curves (overlapping) show the comparator output Q1_collector at 30MHz and 100MHz, while the lower curve was simulated at the resonance frequency of 77.286MHz. As can be seen from this graph only at 77.286MHz soft rectification appears that introduces a minor DC shift << 1%. This result is in line with the performed AC analysis.

Example 3: Debugging a current reference

As a 3rd example, the current reference as depicted in figure 13 is debugged for EMS. In order to do so, a DPI source is connected to the input node and used to emulate EMC.

Step 1: Identify rectifiers: Primary devices that can rectify are: The follower NMOS device, the diode in the current mirror and the ESD protection diode. Indirect rectifications could appear due to limited input and output range of the operational amplifier and from the output PMOS device in the current mirror.
Step 2: Circuit analysis: A small signal analysis is made for the nodes that can lead to rectification. It's not the goal to make exact AC calculations here but to fast explore trends and values applicable for the circuit by means of your intuition. Detailed analysis is used only at the moment vulnerability is found.
Small signal: For frequencies below the Opamp GBW, the NMOS follower has a low output impedance, mainly defined by 1/Gm in parallel with Rext and Cpin. Due to the low output impedance, the bandwidth will be in the order of several hundreds of MHz. A linear approximation for the follower impedance is depicted in Figure 13. For high Due to the negative feedback introduced by the operational amplifier, the impedance of the follower is divided by the gain. A typical Opamp has a gain of 100dB, a dominant pole at about 1Hz and a gain bandwidth product (GBW) of about 1MHz. Also this curve is added to figure 13.

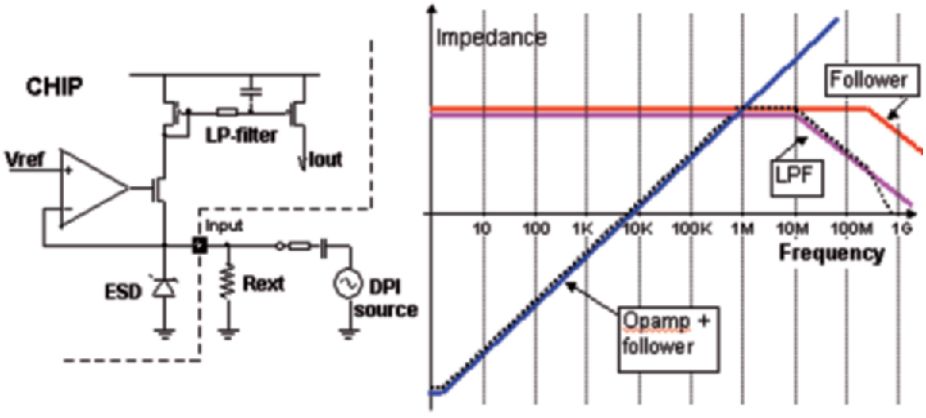


Fig. 13. Schematic and AC properties of the current reference circuit.

Large signal: For frequencies below the GBW of the Opamp, the gate will follow the source and the DC bias current will vary linearly proportional to the EMC amplitude. For moderate EMC levels, there is no rectification and no charge pumping since the capacitance and the impedance on the input node are very low. The linear drain current however is injected into the non-linear mirror diode. This diode voltage is hence non-linear but the exact inverse non-linearity of the output transistor restores this and the output current is linear. Note that we silently have assumed that the bandwidth of the mirror is well above the GBW of the Opamp, which in general will be true. The EMC signal is superposed on the output current but there is no DC shift on the mirror output current.

Step 3: Fail criterion: The Opamp will loose to stay in feedback as soon as the DC current in the MOS follower device becomes zero. Applying this fail criterion, the Opamp plays no role any more (feedback loop is broken) and the maximum impedance on the output node is approximated by Rext//Cpin. This shows a direct limitation for the EMC susceptibility. Note that compared to the follower circuit, the initial impedance is divided by the gain (opamp in feedback) but the maximum impedance stays the same (opamp not in feedback).

Small signal: For frequencies above the Opamp GBW, the Opamp becomes too slow to follow the input signal and acts as a low pass filter. The gate of the follower MOS no longer follows the source. Using some simplifications, we consider the gate to become a DC point. Depending on the opamp topology, also a DC shift on the opamp output will occur due to rectification effects inside the opamp, thus moving also the DC operating point of the follower. For simplicity, we here assume the opamp itself not to rectify, even for large signal amplitudes.

Large signal: When the EMC signal modulates the input voltage, an exponential or quadratic drain current is created, dependent on the operation region of the MOSFET. This non-linear drain current is fed into the mirror diode. We consider here only the case where there is no LPF or the pole of the LPF is well beyond the GBW of the Opamp:

a) For frequencies between the GBW and below the LPF –3dB point, the LPF has no impact and the non-linear drain current is directly mirrored to the output. So also the output current is exponential or quadratic and when filtered on a capacitor, a DC shift will appear.

b) For frequencies above the –3dB point of the LPF, the situation changes. The exponential or quadratic drain current of the NMOS is converted into a linear voltage over the mirror diode if this PMOS device is in the same operation

Table 1. Overview of the circuit behaviour.

Node	**Freq. below GBW Opamp**	**Freq. below LPF + above GBW**	**Freq. above LPF**
Id NMOS	Linear output current	Exponential or Quadratic output current	Exponential or Quadratic output current
Gate NMOS	Follows Vemc	DC	DC
Vmir_diode	Logarithmic or square-root voltage	Linear voltage	Linear voltage
Vout LPF	No attenuation of Vmir diode	No attenuation of Vmir diode	Attenuated Vmir diode
Iout	Linear output current, no attenuation of EMC input signal	Exponential or Quadratic output current, no attenuation of EMC input signal: hard rectification & pumping	Attenuated exponential or quadratic output current: soft rectification and pumping

region of the NMOS input transistor. This linear or softly rectified diode-voltage is attenuated by the LPF and only a fraction of the AC component is passed to the VGS of the mirror output MOS. Consequently, the output current has a much smaller EMC content and no significant rectification occurs on the current output. The table below summarizes the different circuit operations:

We now review a few solutions that will reduce the rectification for frequencies between the GBW and the –3dB point of the LPF. The fundamental problem causing the rectification is the exponential or quadratic drain current generated because the Opamp cannot follow the input signal and the gate of the NMOS becomes a DC signal. Adding a capacitor from the input pin to ground is not a good solution. In order to attenuate the EMC signal, the pole on the input pin needs to falls well below the Opamp GBW, which is impractical and poses a difficult stability problem. A better solution is to decouple the gate from the Opamp output by means of a series resistor and an extra capacitor placed over the VGS of the NMOS. In this way, for high frequencies, VGS is shorted and basically disconnected from the Opamp output. The incoming EMC signal is now converted in an exponential or quadratic output current but the high-frequency signal is largely attenuated before it reaches the PMOS diode. This will strongly reduce the rectification effect. Or in other words, we introduce a feed forward path that bypasses the Opamp and shorts the VGS of the NMOS.

Summary

In this article, a method was described that uses the small signal approach combined with well-known DC characteristics of a circuit to determine and quantify EMC susceptibility. This method was applied on three practical examples and shown to be effective as well to design a circuit as to debug a circuit. Despite the method has its limitations and was not exhaustively evaluated, it has proven to be of valuable help to a designer to analyse circuits for EMS. The method does not require any special knowledge or tools and can be applied on analog continuous time circuits. More work is ongoing to further optimise and extend this method and also to integrate it into the design flow.

References

(1) H. Casier, *"Electronic circuits in an automotive environment"*, Tutorial presentation T3, ISSCC 2004, Feb. 15, 2004, San Francisco.

(2) M. De Mey, *"Robustness in Analog Design"*, Proceedings of the 12th Workshop on the Advances in Analogue Circuit Design, AACD 2003,

(3) B. Deutschmann, *"Improvement of System Robustness through EMC Optimization"*, Proceedings of the 12th Workshop on the Advances in Analogue Circuit Design, AACD 2003, 15-17 April 2003, Graz Austria.

(4) IEC 62132-4, "Integrated Circuits – Measurement of Electromagnetic Immunity – 150 kHz to 1 GHz, Part 4: Direct RF Power Injection Method".

IC MODELLING FOR EMC
New Developments

Mart Coenen
Philips Applied Technologies, EM&C
High Tech Campus 26
5656AE Eindhoven, the Netherlands
mart.coenen@philips.com

Abstract

Modelling of ICs is mostly "just" restricted to the functional behavior of the design. Extending the modelling requirements into the EMC domain, including signal and power integrity and incorporating the application with the measurement environment requires a different approach. Fortunately, the IC in its package has only a limited number of pins/balls to the PCB connected thereto. A new approach is given which is presently under discussion and development in international standardization (IEC62433-x) by IEC SC47A/WG2, but also by other bodies like: EIA,IEEE [10, 12, 15] and/or proprietly organizations.

1. Introduction

Over the last two decades, the EMC modelling of ICs has become a major topic to enable a forecast of IC designs in their intended application conditions. Doing so has indicated the necessity to take measures at the application level, like: shielding and filtering at the application level, to fulfill the mandatory EMC requirements at the product level. An EMC margin i.e. lack of margin will indicate the necessity of measures required which will affect the overall bill-of-materials and as such will forecast the likelihood of an economic success. In addition, areas for design improvements can be identified, before a tape-out or a design re-spin.

At first the RF emission of relative small and simple digital circuits in their application were considered. Thereafter, various attempts were explored to model RF immunity of circuits, at first for operational amplifiers, followed by analysis of digital applications (standard logic families). The remaining field for exploration will be to analyze and model the impulse immunity of circuits to electric and cosmic phenomena.

The first step in signal integrity (SI) modelling was to use IBIS models [2], which forecasted the behavior of I/O signals off-chip. Under the assumption e.g. with simple logic functions, that I/O signals are dominant,

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 239–251.

this information can also be used to forecast RF emission. But for more complex IC designs, the contribution of the IC's core will dominate the RF emission. To extend the IBIS models, new concepts were created like ICEM, LECCS and IMIC [3, 4, 5] to represent the "noise" created by the digital core with a limited number of equivalent sources and lumped elements. Driven by the capabilities of these new models, also IBIS was extended to enable the forecast of power integrity (PI) by the release of version 4.1.

From the above it can be noted that most EMC analysis is carried out by using analogue circuit simulators and that the measurant therefore needs to be conductively. When simulation results need to be compared to measurements either the requirements at the product level (IEC CISPR-x) [9] or the EMC IC measurement techniques as defined in IEC 61967-x have to be used [6, 7, 8]. In either case, the levels are given as conductive requirements and/or by radiated limits.

To enable EMC modelling to an extend, an open interface format, agreed upon by all IC suppliers, is necessary to enable system integrators to carry out PI/SI and EMC system analysis. This open interface format will at first probably be irrelevant for the IC manufacturer to support his "internal" design and simulation environment for his internal development process.

However, even for IC suppliers who have to deliver the proof on concept for their newly designed systems, supplementary data on accompanying devices is inevitable to forecast the performance of the whole system.

The IC related information has to be known at the boundary of the die, package level or even at the module level (Black-Box), determined by the requirements of the system integrators. When System-in-Package (SiP) designs are to be created, bare die boundary signal information is necessary. At the next level of system integration, the information needs to be known at the boundary of the IC package irrelevant on its content; single or multiple dies, with or without integrated passives.

2. RF Emission Root Causes

Considering the RF emission from an IC (die), the root causes will be threefold:

- Noise from the I/Os, including their supply paths; V_{ssE}/V_{ddE}
- Noise from the core; V_{ssD}, V_{ddD} or
- Noise from the substrate $V_{ss(IS)}$-network

These three noise sources can be considered independent from each other and ultimately at the boundary of the IC, a combination of the three will appear; one dominant with the other two superimposed (crosstalk) with lower amplitudes.

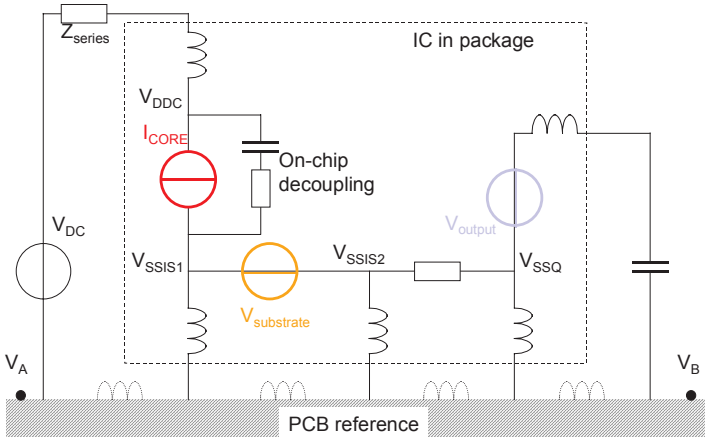


Fig. 1. Examples of core, V_{ss}-net/substrate and I/O noise sources.

2.1. Noise from I/Os

When the noise of the I/Os is considered, the full transistor based model can propriety be used. A high level of model-order-reduction (MoR) is necessary to bring the complexity down to whatever is processable by analogue circuit simulators. Furthermore, propriety information about the circuit topology and the process used needs to be hidden for the end-user. Examples of I/O modelling and their consequences are given in figure 2: HSpice, 3-Wire and IBIS vers. 2.1.

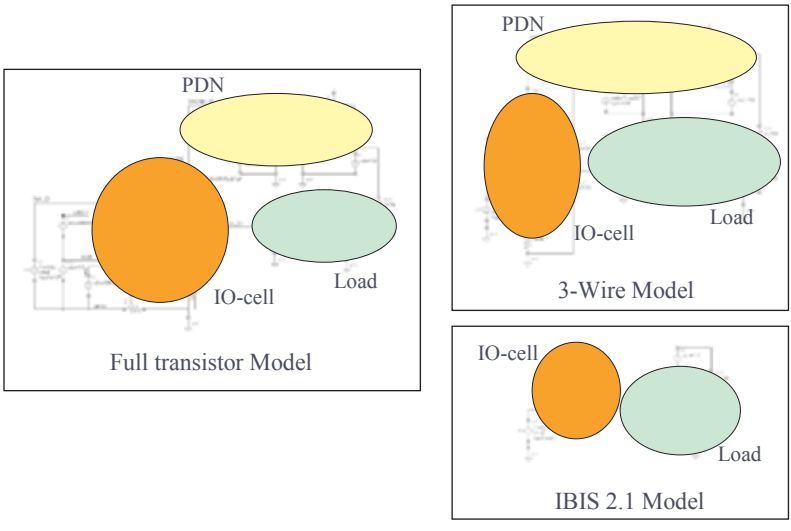


Fig. 2. Examples of I/O modelling and its consequences for PI and SI.

The simulation results in figures 3 and 4 give the worst-case PI and SI results of 61 simultaneous switching outputs and a full transistor based model is compared to a 3-Wire model. This 3-Wire model has been developed in close collaboration with Fraunhofer Institute IZM during the MEDEA^{+} Mesdie program, and an IBIS Version 2.1 model, up till today the only version supported in the Cadence design flow. The 3-Wire model has nearly the same amount of active nodes as the IBIS model but is capable to forecast PI as well as SI but only requires 6 parameters. In this simulation example, the 3-Wire model has not been corrected for the delay that is incorporated in the full transistor and IBIS model [13, 14, 16, 17].

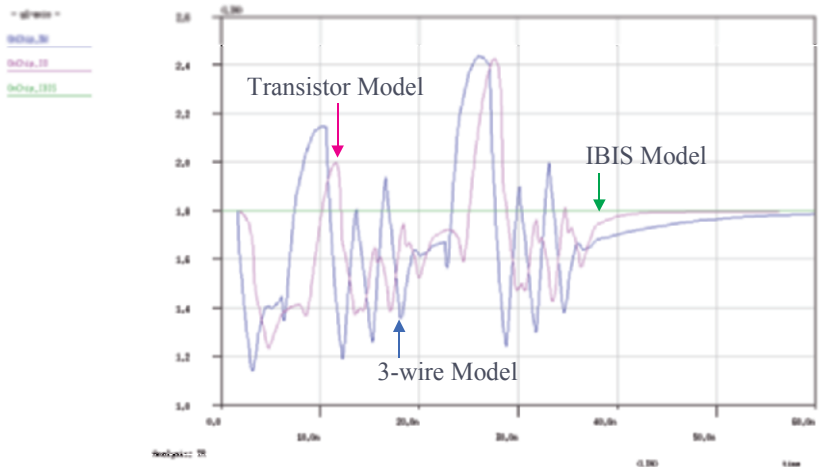


Fig. 3. Examples of I/O modelling and its PI results.

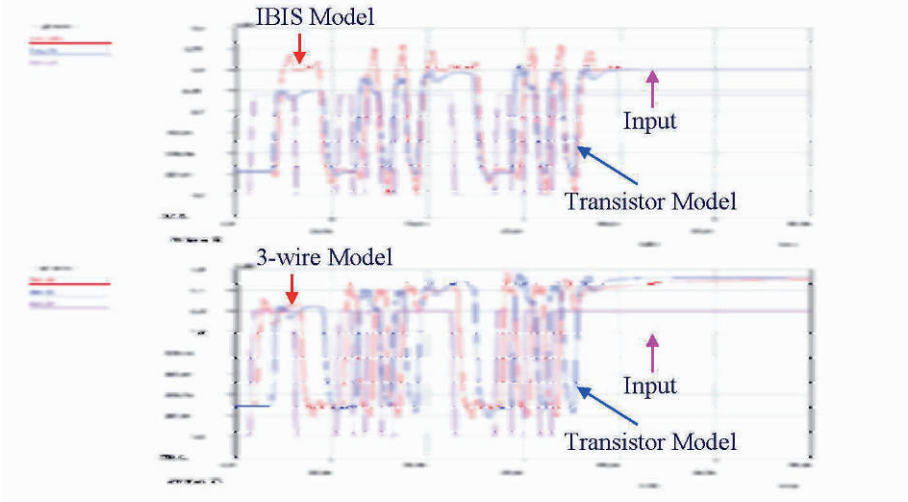


Fig. 4. Examples of I/O modelling and its SI results.

In figure 3, it can be seen that the internal supply voltage variation (PI) of the full-transistor model and the 3-Wire model are close as where the IBIS model only indicates an ideal unaffected supply voltage as a consequence of the model used. In figure 4 the SI responses for the 3 models used are given. Here too, the 3-Wire model follows the full-transistor model when the lack of compensation for the delay is taken into account. The IBIS model overestimates the signal swing but still provides a good forecast of the functional signal in the logic threshold region.

With RF design, the use of all information; PCB, package and IC information is quite common but for these kind of applications the number of signals to be considered is quite limited, typically 4; two for the differential signalling and its supply and ground. With these designs, all signals are typically referred to a 50 Ω impedance environment and they are well described using S-parameter representation for the passive networks involved [1, 10, 11].

With digital applications using 32, 64 or 128-bit wide busses, the PI and SI modelling complexity grows accordingly and other descriptive methods and/or model order reduction need to be applied. Digital source impedances are state dependent and the package and PCB topologies are less RF defined. Furthermore, signal transitions are “random”, skew is NOT predetermined and the slew-rate of the signals is data rate and PI dependent. Last but not least, the application (off-chip decoupling) affects the power integrity and as such signaling will be history related (decoupling capacitance charge recovery time).

2.2. Noise from the Core

Considering that digital cores contain several million transistors, used in a limited number of standard cell configurations, the local cell supply voltage, the excitation (activity) but also the location of that standard cell within the core area will determine its contribution in the noise of the core, called “internal activity” (IA). The contribution to the whole core supply current will be determined by: the activity of other cells nearby which may act as local decoupling, the on-chip decoupling present and the topology of the supply network. The whole supply network, called “power distribution network” (PDN) with ICEM can be used to describe the bounce in the supply i.e. ground network of the core.

2.3. Noise from the Substrate

The supply currents flowing in the core and I/O area will determine the static voltage IR-drop in the substrate i.e. V_{ss}-network which is in parallel to that substrate. The IR-drop can be calculated by the end of the design when

layout and routing is finished e.g. by using Voltage Storm from Cadence. Alternatively, proprietly scripts have been written which forecast the IR-drop when a rough floor-planning, power routing structure and power pin allocation is known, see figure 5. In the bottom plane the sources are indicated as well as the ground pin allocations and in an X/Y/Z plane, the IR-drop voltage as function of location is given (referred to "null" volt at the ground pins).

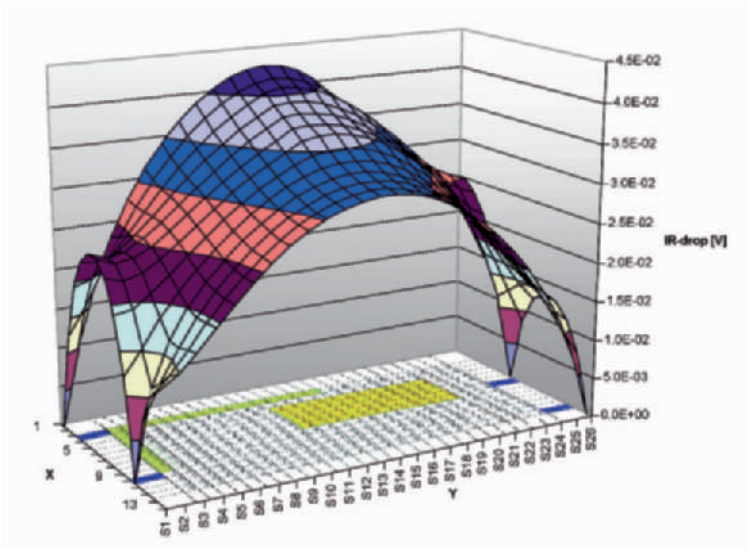


Fig. 5. IR-drop estimation of non-homogeneous designs.

A distinct has to be made between the peak noise and the average "noise" voltage that is just a result of the average or DC currents and the equivalent resistive V_{ss}-network (in parallel to the substrate resistance via contact holes). The averaging effect is again a function of the amount of on-chip decoupling capacitance and the supply network resistance but also of the off-chip supply application used.

When homogeneous supply currents through the IC; V_{ddD} to V_{ssD} net are considered through a homogeneous resistive network, the static delta voltage at the boundaries (along the iso-bares) will be zero. However, when high-speed busses with several agents communicate between left to right or vise versa, the supply current drawn from the core supply network will remain the same though the noise induced in the substrate/ V_{ss}-network will reverse its sign.

The way in which the V_{ss}-net/substrate noise source propagates further will be determined by the application in the IC-package/SiP and/or at the PCB. When all multiple V_{ss}-pads are short-circuited to a common ground

e.g. die pad, the remaining voltage towards the PCB will be substantially less due to the low impedance in parallel to the substrate/ V_{ss}-network impedance, figure 6.

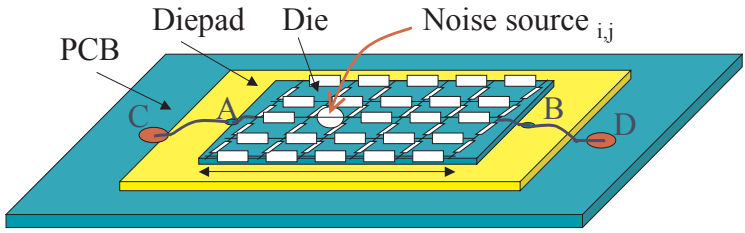


Fig. 6. Propagation of substrate/V_{ss}-net noise.

3. Application

The type and placement of the off-chip decoupling elements will determine the RF supply current flowing off-chip onto the PCB, both for core and I/O. These off-chip RF currents will induce noise on the PCB. When no RF or AC current is allowed off-chip e.g. by using an off-chip DC current source to supply the IC core or by using Kelvin contacts [16], the ground-bounce between the V_{ss} on the IC and the V_{ss} plane of the PCB will diminish. In nearly every case, the on-chip supply bounce is (pre-) determined by the ratio of active versus static cells and the decoupling capacitance on-chip. By the above measures, the equivalent core impedance has become irrelevant when high off-chip series impedances are being used.

4. Signal Interpretation

Different from using a broadband oscilloscope, the RF emission is typically measured by using an EMI receiver or RF spectrum analyzer (SA) with a prescribed bandwidth. For frequencies between 30 and 1000 MHz a typical bandwidth of 120 kHz is used, see IEC CISPR-16-x [9]. For automotive applications only a bandwidth of 10 kHz is used in this frequency range. This means that the measured signal is weighted over a longer period (» 1/BW, to allow the band-pass filter to respond). When according IEC CISPR the quasi-peak detector is used, the measurement time is extended to several ms per frequency interval.

With steady-state repetitive signals the RF (power) spectrum or spectral density is fixed. When using functional, more random like signals and/or different modes of operation, this is not the case and the measured result will

be data and mode of operation dependent. A specific example is the use of data bursts on busses. Here convolution applies between the steady-state random data and the burst length versus repetition rate (on/off keying). Last but not least, using spread-spectrum modulation on the clocks used may affect the integrated measurement results determined by the ratio of the frequency spreading versus the bandwidth that with the signal is measured, but this will hardly affect the time related instantaneous responses.

5. New Approach for EMC Simulations

ICs can be black box represented as active N-ports. The N-port can be made passive by using equivalent impedances that behave according a certain averaging over time as indicated above. An impedance [Z], admittance [Y] or S-parameter matrix can represent these passive networks [11]. Then in addition, equivalent voltage sources have to be added in series with each pole of the N-port, see figure 7. These voltage sources may represent arbitrary wave shapes of which their phase relation (or delay) is defined towards a unique point of reference.

The filtering as provided by the measurement system (SA) can be transferred through the passive network without affecting the overall response; response weighting is both distributive as well as associative. This representation allows determining the influence of off-chip loading to the currents resulting (sensitivity analysis). The model as presented can be dealt with according the analogue simulator format requirements of the ALF [10].

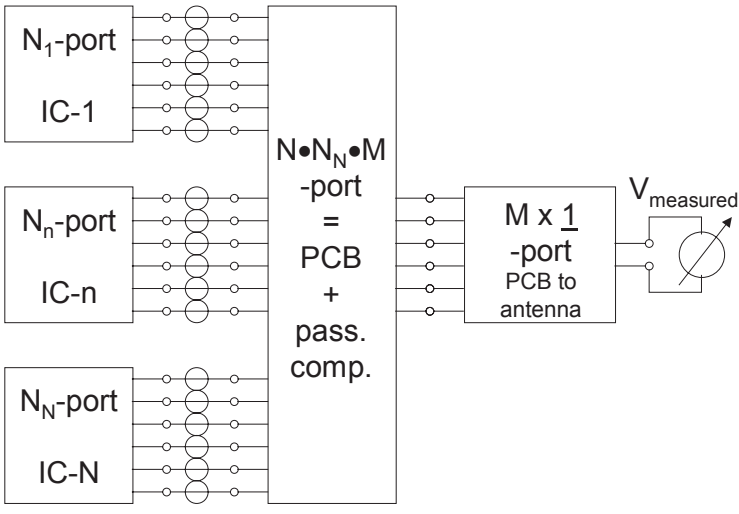


Fig. 7. ICs as N-port models.

The loading of the N-port will only be slightly affected by the load, considering the typical condition as defined in the specified in the datasheet. When either a capacitive or a resistive load is expected (within ranges) at each driver, the forecasted signal source with that node will only slightly divert with the real application.

When load impedances are known for each pole of the N-port together with the IC's excitation, then an equivalent circuit with (N-1) current sources or (N-1) voltage sources without any impedance can represent the whole IC.

All the ICs used shall be represented by N_n-ports that are all connected to passive networks being the SiP or the PCB incorporating discrete lumped elements and/or the lumped element representation of the IC-package or laminate. The PCB shall be represented as an N_n-port to each IC and to its external connectors by a M-port, figure 7. All the voltage sources in series with all poles shall represent the signals, as they would appear in the real application. Macro modelling applies for each level, dependent on the physical size of the application and the upper frequencies used.

Ultimately, the measurement port is typically a single port, connected to the external measurement system either conductive or by radiation. For each of the measurement methodology chosen, a set of frequency dependent transfer functions from each IC node towards the measurement system will result.

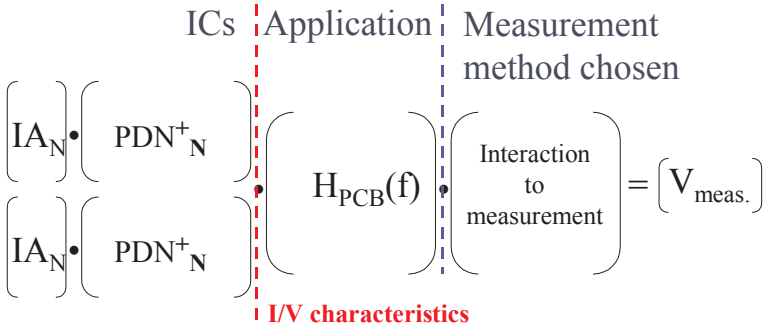


Fig. 8. Total EMC IC modeling framework.

For the most simplified version of the impedance (admittance) matrix, only (part of) the diagonal impedances (admittances) can be provided. Doing so ignores all internal IC crosstalk to the other nodes in the N-port as each node has become independent.

Other reduction algorithms to reduce complexity can be achieved by setting “zero-s” when values below (above) certain levels. Either from the boundary of the IC (SiP) or from the boundary of the PCB, this equivalent source representation can now be used with 3-D EM solvers: CST MWS, Ansoft HFSS, etc.

At least four options for IC data representation are available:

- All source impedances, amplitudes and phases; Load i.e. application insensitive
- All source amplitudes and relative phases known; Load sensitive
- Only diagonal source impedances, amplitudes and phases; Load sensitive and restricted
- Only partly diagonal source impedances, amplitudes and phases (e.g. IBIS & ICEM)

The info required can be derived by simulations or from near-field scanning measurements.

This approach allows for two kinds of optimization:

- Either the transfer function from a certain IC node needs to be minimized towards the measurement port; IC-package, SiP, PCB level or
- The equivalent excitation of the IC needs to be minimized or impedance to be optimized to those critical transfer functions (with were notified to give minimum insertion losses towards the measurant).

5.1. Worst-Case

While considering the real application, the ICs have to demonstrate reliable operation under the worst-case mode of operation. The worst-case with power integrity and signal integrity may not coincide as switching speeds degrade when the supply voltages reduce. Other influences will result from Process spread, off-chip supply Voltage and Temperature (PVT).

5.2. Test and Measurement

For all of the leaded IC packages, the currents through each individual pin can be measured as function of time (with a delay towards a given reference) or as function of frequency (with a certain phase relation towards a given reference). An EM field scanning measurement technique suited to derive these parameters is described in TR IEC 61967-3 [6]. Taking both the local E-field, to represent the voltage on the pin, and the H-filed through that pin, representing the current, the node can be characterized.

5.3. Relation to Existing Proposals

The mathematical modeling proposed allows an open interface format for the definition of parameters of ICs used to enable RF emission modelling. Most analogue circuit simulators; (H)Spice, Spectre, Sabre, APLAC are incapable of incorporating complex 3-D structures in their analysis. 3-D EM solvers are typically incapable of dealing with complex source and load structures. The presently available formats; IBIS, ICEM, LECCS, IMIC, 3-Wire model or encrypted full transistor models are limited in their applicability and/or will only represent part of the real circuits.

6. Acknowledgement

The modelling and standardization work is carried out in close cooperation with the MEDEA$^+$ Mesdie and Parachute partners [18, 19] and the members collaborating in the international standardization committees from IEC SC47A/WG2 and 9.

7. Conclusions

- Although the modelling concept is still under development, this concept resolves most of the present EMC modelling issues and indicates the possibilities for future enhancements.
- The framework as described allows the inclusion of most of the presently used modelling formats (upwards compatible). Impedance and source representation format conversion routines can to be used from the MatLab libraries.
- The open interface format allows the use of all information; full matrix, just the main diagonal or even less. Less detail in source definition will give less reliable simulation results but the format still can be used.
- The data required can be derived from full transistor models (by the vendor) or obtained from measurements (EM field surface scan) when leaded packages are used.
- Existing EMC modelling development yields to company propriety formats that block the EM modelling of a full application incorporating ICs from various suppliers.
- The interaction-modelling format as proposed is used in a new framework of international standards (IEC 62433-x) as indicated in figure 9.

Part 1: General → **NP**
Part 2: RF emission conducted → **CD = ICEM revised**
Part 3: RF emission radiated → **NP + partly ICEM**
Part 4: RF immunity conducted → **NP + partly ICEM**
Part 5: RF immunity radiated → **NP**
Part 6: Impulse immunity → **NP**
Part 7: Intra-IC compatibility → **NP + partly ICEM → TR**

NP = New work item proposal
CD = Committee draft
CDV = Committee draft for voting
TR = Technical report
TS = Technical specification
FDIS = Final draft International standard
IS = International standard

Fig. 9. EMC IC modelling framework proposal.

References

[1] R. Mavaddat, Network Scattering Parameters, World Science, 1996
[2] IBIS Modelling Cookbook For IBIS Version 4.0, Government Electronics and Information Technology Association and The IBIS Open Forum, 2005
[3] Models of integrated circuits for EMI behavioral simulation ICEM, IEC 62014-3
[4] Linear Equivalent Circuit and Current Source for I/O (LECCS-I/O) Modeling of IC Power Current for EMI Simulation, http://www.sdl.hitachi.co.jp/english/summary/2004/520161/ or 510028/
[5] Models of integrated circuits for EMI behavioral simulation IMIC, IEC 62404
[6] Measurement of Electromagnetic Emission up to 1 GHz, IEC 61967-x
[7] Measurement of Electromagnetic Immunity up to 1 GHz, IEC 62132-x
[8] Measurement of Impulse Immunity, IEC 62215-x
[9] IEC CISPR-xx, (see www.iec.ch)
[10] Advanced Library Format, IEEE 1603, 2003
[11] Touchstone File Format Specification, Rev 1.1, EIA/IBIS open Forum, 2003
[12] A.L. Drozd, Selected Methods for Validating Computational Electromagnetic Modelling Techniques, IEEE, 2005
[13] Z.Y. Yang, S. Huq, V. Arumugham, I. Park, Enhancement of IBIS Modelling Capability in Simultaneous Switching Noise (SSN) and

Other Power Integrity Related Simulations- Proposal, Implementation and Validation, IEEE 2005
[14] R. Kaw, I. Novak, M. Swaminathan, Towards Developing a Standard Input/Output Format for PDN Modelling and Simulation Tools, IEEE 2005
[15] M. Coenen, R. de Jager, Standardization for EMC IC Modelling, EMC COMPO 2005
[16] M. Coenen, D. de Greef, Optimization techniques for minimizing IR-drop and supply bounce, EMC COMPO 2005
[17] E. Engin, Modeling of Lossy Interconnects and Packages with Non-Ideal Power/Ground Planes, PhD Thesis, University of Hannover, 2004
[18] MEDEA$^+$ Mesdie, project A509, www.medeaplus.org
[19] MEDEA$^+$ Parachute, project 2A701, www.medeaplus.org
[20] Electromagnetic Compatibility of Integrated Circuits; http://www.springer.com/sgw/cda/frontpage/0,11855,7-40109-22-86706278-0,00.html

Part III: Ultra Low Power Wireless

To achieve information transfer in flexible portable systems wireless communication is the best option. As such the power drain in the wireless circuits are becoming extremely important. In the area of low date-rate systems (sensor area networks) ultra low power wireless circuits are required. For that in this chapter the advances in ultra low power wireless systems is discussed.

Since a few years the Ultra Wide Band impulse radio is under study. In the first paper receiver architectures and their circuits for high data-rate impulse UWB receivers are described. A good optimization between wide band antenna specifications and circuits can improve the efficiency in the allowed spectrum. However for receiving high quality signals in a correlation receiver the pulse waveform is very important. For that in this approach consecutive pulses are transmitted with a predefined delay between them. The first act as reference and the second is modulated. However to realize in the receiver accurate delay functions requires specific circuits, which are described and analyzed.

In the second paper attention is focused on wireless systems for sensing networks. Of course the required data-rate is very small, and as such ultra low power receivers are required. This has been received by exploring special techniques such as bulk acoustic wave components in combination with super-regenerative transceiver topologies. Circuits such as oscillators with power drains of 100 μW and passive receivers with power drains of 200 nW have been demonstrated.

The next paper discusses the optimization on system level for low-date rate impulse UWB systems. This is in the 0-960 MHz band. Since ultra low power is up most important a careful selection of implementation in the analog or digital area of the different building blocks has been analyzed. It is clear that with the existing technologies the front-ends are better off in the analog area. As such receivers with energy consumptions of less than 20 nJ/bit should be feasible.

Also in medical implantable system more and more wireless communication is present. Either the power is provided by means of an inductive coupling or either by an internal non-rechargeable battery. Since the operation requires typically 5 to 7 years ultra low power is mandatory. As such power down topologies are commonly used. However fast power up circuits, especially in the clock reference circuits, can reduce the power drain considerable. In this fourth paper quick start circuits are as such discussed in detail.

The next paper discusses transceiver circuits for wireless sensor networks. Besides the circuits, the optimization of the communication

protocol can have serious impacts on the total power drain. In this work also transceiver circuits are analyzed and the main power drain, besides the power amplifier, is in the PLL circuitry. PLL circuits with 1 mA have been achieved. By a careful combination of protocol and circuits an average power drain of 27 μW for 32 bytes packet every 30s has been demonstrated.

In the last paper a frequency hopping spread spectrum synthesizer is discussed. Commonly fractional N synthesizers are selected, but they have to deal with increased phase noise coming from the sigma-delta modulator. By using Direct Digital Frequency Synthesis this problem can be avoided. However high speed DAC's are required in this approach. The advantage is that predistortion of the varactors can easily be implemented.

Michiel Steyaert

ULTRA-WIDEBAND RADIO: UNCONVENTIONAL CIRCUIT SOLUTIONS FOR UNCONVENTIONAL COMMUNICATION

S. Bagga, S. A. P. Haddad, W. A. Serdijn, and J. R. Long

Electronics Research Laboratory, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Delft University of Technology, Mekelweg 4, 2628CD, Delft, The Netherlands
E-mail: {s.bagga, s.haddad, w.a.serdijn, j.r.long}@ewi.tudelft.nl)

Abstract

The governing body in the United States, the Federal Communications Commission (FCC), has issued ultra-wideband regulations, under Part15 of the Commission's rules on April 22, 2002. Therefore, ultra-wideband transmissions (intentional emissions) under certain frequency and power limitations have been permitted. The US has openly endorsed UWB based consumer products. Ultra-wideband offers significant contributions and advantages but simultaneously a number of challenges also need to be addressed. One of the key challenges is the co-design of an impulse generator and miniaturized antennas for ultra-wideband impulse radio. This has been designed and fabricated in 0.18μm CMOS technology. Measurements show the correct operation of the circuit for supply voltages of 1.8V and a power consumption of 45mW. The output pulse approximates a Gaussian monocycle having a pulse duration of about 375ps. Proper modulation of the pulse in time is confirmed [1]. In addition, to meet the stringent FCC stipulated frequency spectrum, an orthonormal [2] ladder filter with a Daubechies' impulse response is employed. The filter is implemented using novel 2-stage gm-C cells employing negative feedback. Simulation results in CMOS 0.13μm technology show that this pulse generator requires a total current of 30mA at a 1.2V power supply. The frequency coverage of the simulated waveform is about 85% of the FCC mask.

In receivers, narrowband interference has been acknowledged as a serious impairment for UWB performance. Only limited research has been addressed to evaluate the effects on transmitted reference (TR) schemes. By employing "frequency wrapping", a new architecture, the Quadrature Downconversion Autocorrelation Receiver (QDAR) has been developed [3]. It does not suffer from timing and template matching problems, and it also circumvents processing at high frequencies, thereby reducing the on-chip circuit complexity and power consumption. Moreover, it not only offers narrowband interference rejection but also relaxes the timing accuracy of the delay circuitry in the auto-correlation function. For the QDAR, a quantized

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 255–283.

analog delay [4] and a filter based delay line [5] have been designed. For a time delay of 0.55ns, simulation results in CMOS 0.13μm technology show that the quantized analog delay requires a total current of 36.7mA at a 1.6V power supply. For the filter based delay, a total current of 30mA at a 1.2V power supply is required to achieve a time delay of 0.72ns. Finally, this paper also addresses double-loop negative feedback amplifier techniques to realize true wideband LNA's [6].

Index Terms–broadband, delay lines, quadrature downconversion, FCC, impulse radio, integrated circuits, low-noise amplifiers, negative feedback, transceivers, transmitted reference, ultra-wideband

1. Introduction

In today's marketplace for emerging communication technologies, the focal point of attention is ultra-wideband radio as it not only promises enhanced data throughput with low-power consumption, but also provides high immunity against electromagnetic interference (EMI) and robustness to fading. It is expected that future short-range indoor ultra-wideband (UWB) telecommunication systems will operate in the frequency band from 3.1-10.6GHz according to the Federal Communications Commission (FCC) mask [7-11]. Europe and Asia are expected to follow shortly, albeit with minor differences in the allocated energy frequency spectra and emission levels.

When implemented as *impulse radio* (IR-UWB) (i.e., where the information is transmitted by very short EM pulses [7], [12]) or *Orthogonal Frequency Division Multiplexing* (OFDM-UWB), this new communication technology may revolutionize the way we think in wireless technology by modulating data in time rather than in frequency with low-power consumption. Depending on which form of UWB (i.e. impulse radio, OFDM or DS-UWB) is utilized, applications for this technology range from home-RF to detection and localization.

As associated with any emerging technology, there are several challenging design issues in UWB that need to be addressed. From the perspective of traditional narrow-band systems, the wideband nature of the front-end architecture employed in UWB systems requires a totally different design methodology of both the UWB front-end architecture and its constituting UWB circuit building blocks. It is seen that when designing low-power, single-chip, integrated UWB radio solutions, the analog circuits consume the majority of the total power. While low-power digital solutions are readily used for large-scale designs, no comparable techniques have yet emerged for the analog design components. In the following sections, we investigate unorthodox design methodologies for ir-UWB transceiver architectures.

2. Co-design of an Impulse Generator and Miniaturized Antennas for IR-UWB

The co-design of the impulse generator and miniaturized antennas is described in this section. Pulse position modulation is used to encode the binary transmitted data [13], [14]. The waveform to be transmitted is the Gaussian monocycle due to its intrinsic time-frequency resolution product [11], [15], which is important for applications such as positioning and imaging. The impulse generator consists of a cascade of a fast triangular pulse generator and a Gaussian filter (i.e., a filter with a Gaussian impulse response) [16], [17]. The filter is implemented as a cascade of three complex first-order systems (CFOS), which, in turn, consist of gm-C sections that employ differential pairs with partial positive feedback. The entire transmitter is the combination of the modulator with the impulse generator and the antenna (Figure 1).

The modulator is placed in front of the impulse generator as it requires less hardware complexity to delay binary signals as compared to continuous-time signals. In order to obtain pulse position modulation, a ramp is generated, whose slope depends on the information signal [20] [21]. The ramp is then fed to the input of the comparator that compares the momentary value of the ramp with a fixed threshold and generates a trigger [22]. This edge is then used to drive the triangular pulse generator, which consecutively triggers the impulse response of the succeeding pulse shaping network, being a Gaussian filter.

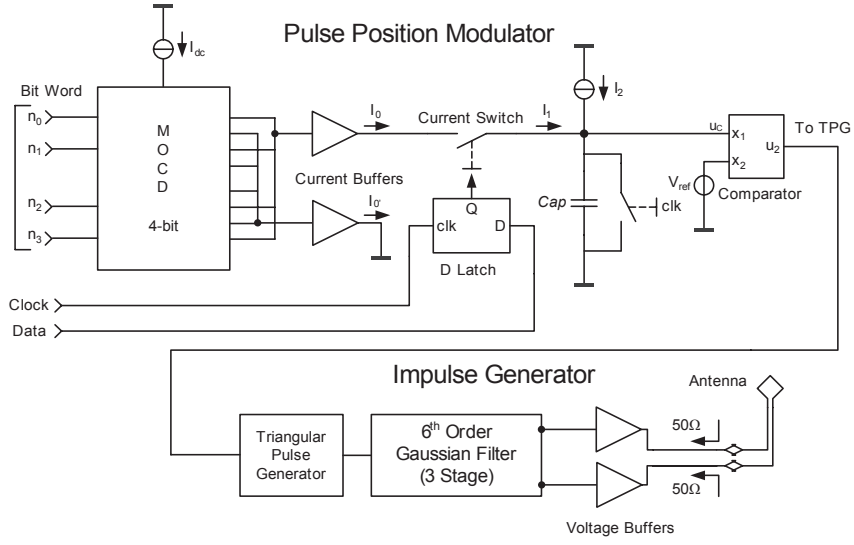


Fig. 1. Block diagram of impulse generator and modulator.

The short transient pulses fired by the impulse generator are transmitted by an antenna. The antenna is closely integrated with the transmitter and it should be very well matched (i.e. achieve a VSWR<2 within the entire operational frequency band). Integration with RF circuits gives additional freedom in antenna design as the antenna input impedance is not limited to 50Ω. Balanced feeding can be realized without a balun by using a differential amplifier in the receiver and an impulse generator with a differential output in the transmitter. The desired radiation pattern of the antenna should be omni-directional.

Relying on our past experience with elliptically shaped dipoles [18], we have developed a so-called butterfly antenna to be used in UWB communications [19]. The optimized length of the butterfly antenna is about 2.2cm and the optimal flair ellipticity is 0.9. The antenna flairs are separated by 3.2mm. The experimental antenna is fed by a double semi-rigid cable. This antenna has been considered as a prototype for an integrated "*generator-antenna*" system.

2.1. Complex First-Order Filters

The Gabor transform [16] can be used to implement a Gaussian filter as its impulse responses are approximated Gaussian window functions, being the first and the second derivatives of Gaussian, respectively. The Gaussian filter can be implemented by a cascade of complex first-order systems. Gaussian monocycles have an excellent time-frequency resolution product making them ideal for applications such as imaging and positioning. A complex filter can be realized by means of cross-coupled real filters. Its structure exhibits similar characteristics as an ordinary second order system. We can cascade CFOS's as shown in Figure 2 in order to make a reasonable approximation to a Gaussian function.

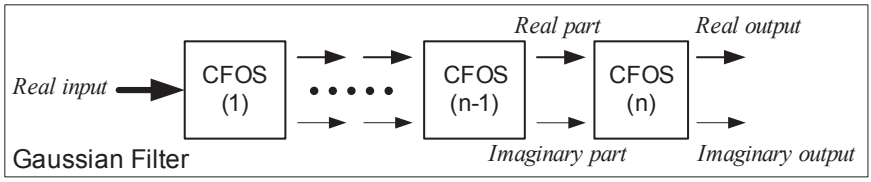


Fig. 2. Cascaded CFOS stages.

The envelope of the impulse response of these (*n+1*) CFOS stages connected in cascade is given by [16],

$$h(t) = \left(c_{re} + jc_{imag}\right)^{n+1} \frac{t^n}{n!} e^{\sigma t} U(t) \tag{1}$$

where U(t) denotes a step unit function. Moreover, through statistical analysis, it is well-known that when n!→∞, the Poisson function approaches a Gaussian function. As one can see in Figure 3, an improvement in the approximation to a Gaussian impulse response is obtained for a larger number of stages. In the next section, a cascade of three CFOS stages will be employed to approximate a Gaussian monocycle waveform.

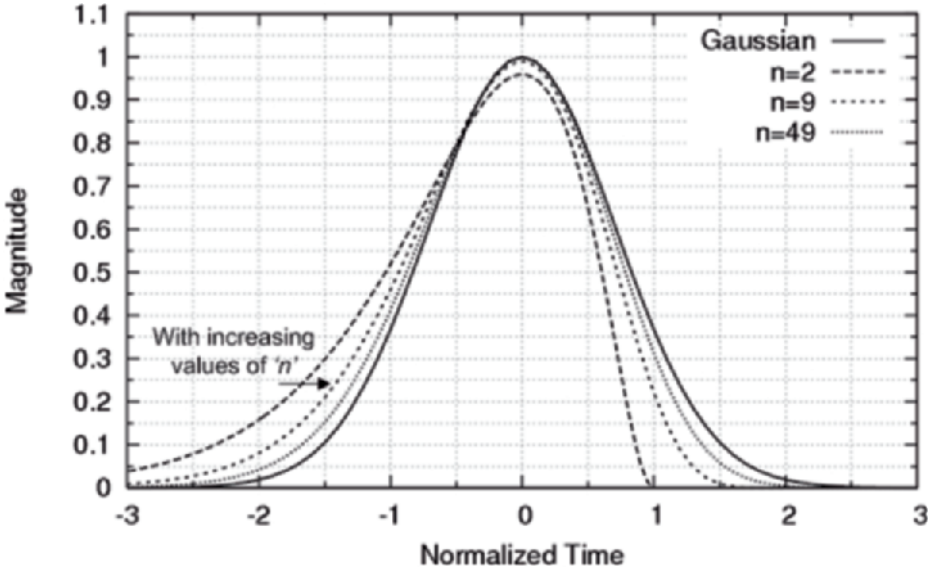


Fig. 3. Impulse response of a cascade of first-order systems with an increasing the number of stages.

2.2. Circuit Synthesis

In this section, the circuit design of the impulse generator will be discussed. For the modulator, please refer to [1].

2.2.1. 6th Order Gaussian Filter

A single CFOS stage using a differential pair arrangement with *partial positive feedback* (PPF) [23] is shown in Figure 4. As expected, the inclusion of the PPF stage as active load enhancement not only increases the DC gain but also the unity gain frequency. The significant increase in gain and bandwidth is contributed to the increase in the effective transconductance of

the stage. If L is the loop gain, then the gain of the amplifier is enhanced by a factor of $1/(1- L)$. L should be bounded, such that it has a lower and upper bound of 0 and 1, respectively.

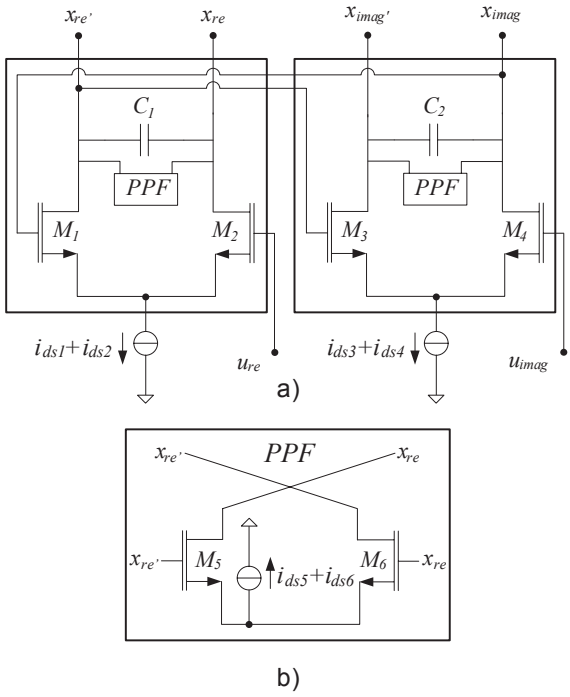


Fig. 4. a) CFOS employing two differential pairs with gain enhancement by partial positive feedback (PPF); b) implementation of PPF.

2.2.2. Triangular Pulse Generator

The triangular pulse generator is made up of a cascade of inverter stages, followed by an NAND gate function [1]. The key purpose of this block is to generate an impulse-like function that is able to evoke the impulse response of the succeeding pulse shaping network (see Figure 5).

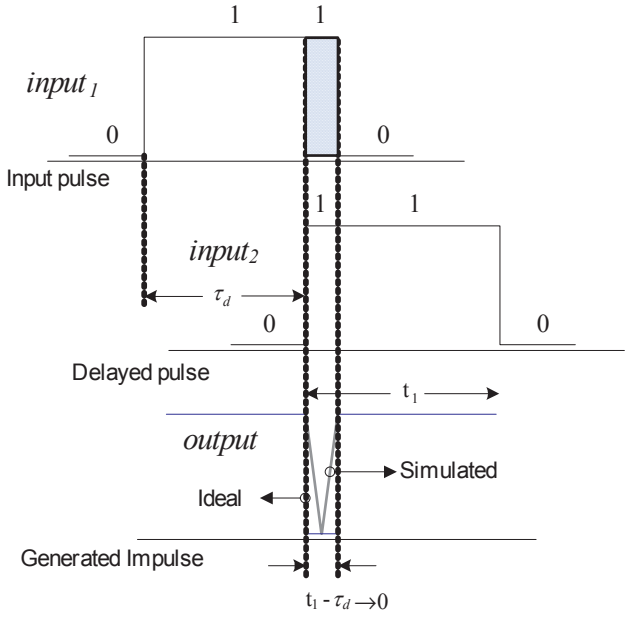


Fig. 5. Input and output waveforms of triangular pulse generator.

A Monte Carlo analysis is run in order to estimate the circuit's sensitivity. In [1] it is seen that the Gaussian monocycle is relatively unlikely to show a substantial discrepancy as a result of process and mismatch variations.

2.3. Antenna Design Issues

Large common-mode currents were present on the feeding cables during the experimental verification of the prototype design. The common-mode current and its excitation are explained by Figure 6 (left). For any symmetrical antenna fed via a conventional asymmetrical feeding line (e.g., a semi-rigid cable), the current received from the generator is divided in two components: the differential-mode current and the common mode current. The former excites the antenna while the latter propagates along the outer surface of the feeding line and causes parasitic radiation. The relative magnitude of the current components is determined by the ratio of the common-mode and differential-mode antenna impedances.

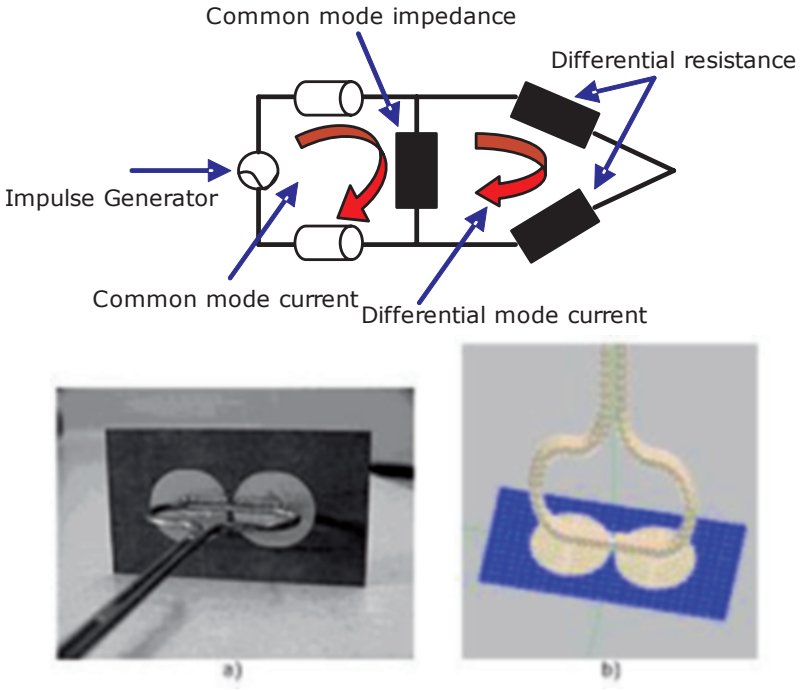


Fig. 6. Top) Excitation of the common mode and differential mode currents Bottom) a) Butterfly antenna with loop feed circuit photo; b) Computational model.

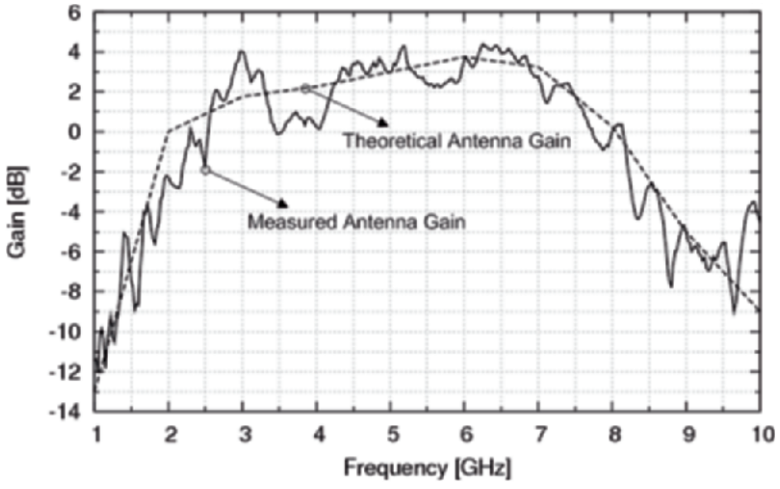


Fig. 7. The antenna gain versus frequency: theory and experimental.

As a result of the common-mode current, the antenna gain in the boresight direction oscillates with frequency (Figure 7). The antenna radiation patterns also show an oscillatory behavior.

In order to avoid excitation of common mode current, we developed an UWB balun [1]. Comparison of the antenna gain for two antenna designs (with and without a balun) is shown in Figure 8. It can be seen that the balun enlarges the antenna bandwidth by approximately 1GHz due to better radiation at frequencies above 6.5GHz.

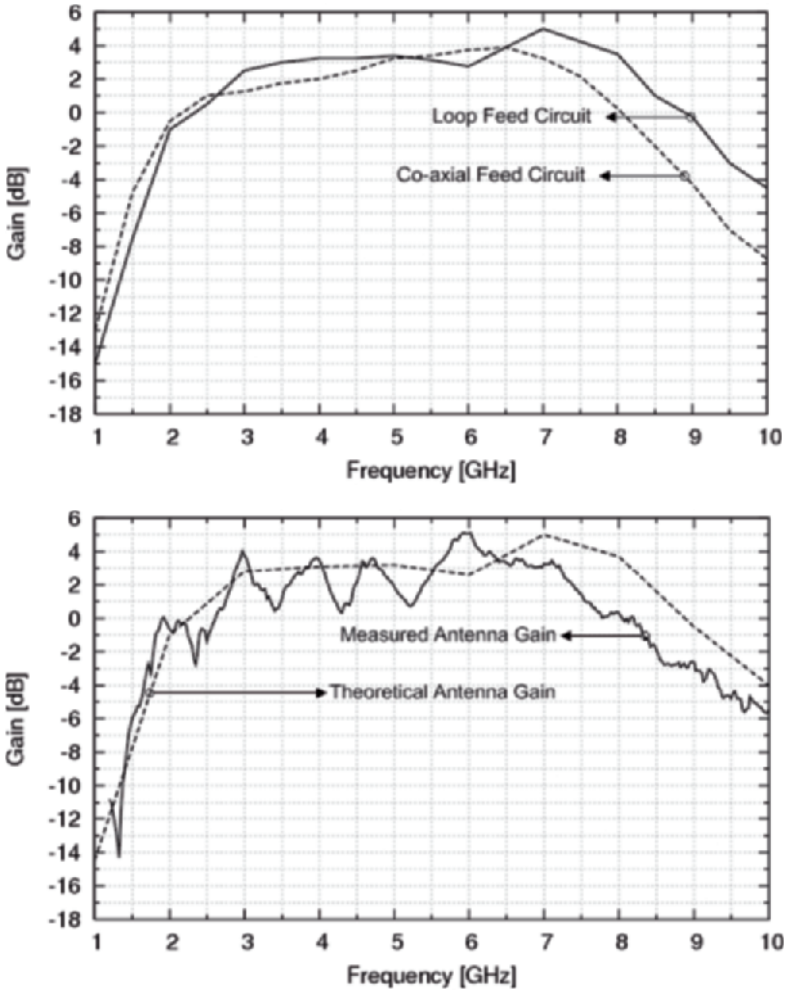


Fig. 8. Top) The antenna gain for two feeding circuits: with (drawn) and without (dashed) a balun; Bottom) Antenna gain for butterfly antenna with loop feeding line: theoretical (dashed) and measured (drawn).

2.4. Measurements

In this section, we investigate the integration of the antenna with the impulse generator. First, the chip is mounted on a PCB to be integrated with

different antennas. The output waveforms from the impulse generator are measured using the setup shown in [1]. The reference current (I_{dc}) fed to the MOCD is set to 1mA at a power supply of 1.8V. Figure 9 (left) shows the measured Gaussian monocycle waveforms. Furthermore, to verify pulse position modulation, a clock signal (1-5MHz), which acts as the binary input signal, is streamed into the D-latch. A bit code (0001) is chosen to verify that the least significant bit would vary the position of the pulse by approximately 315ps (i.e., for I_1=0.625mA).

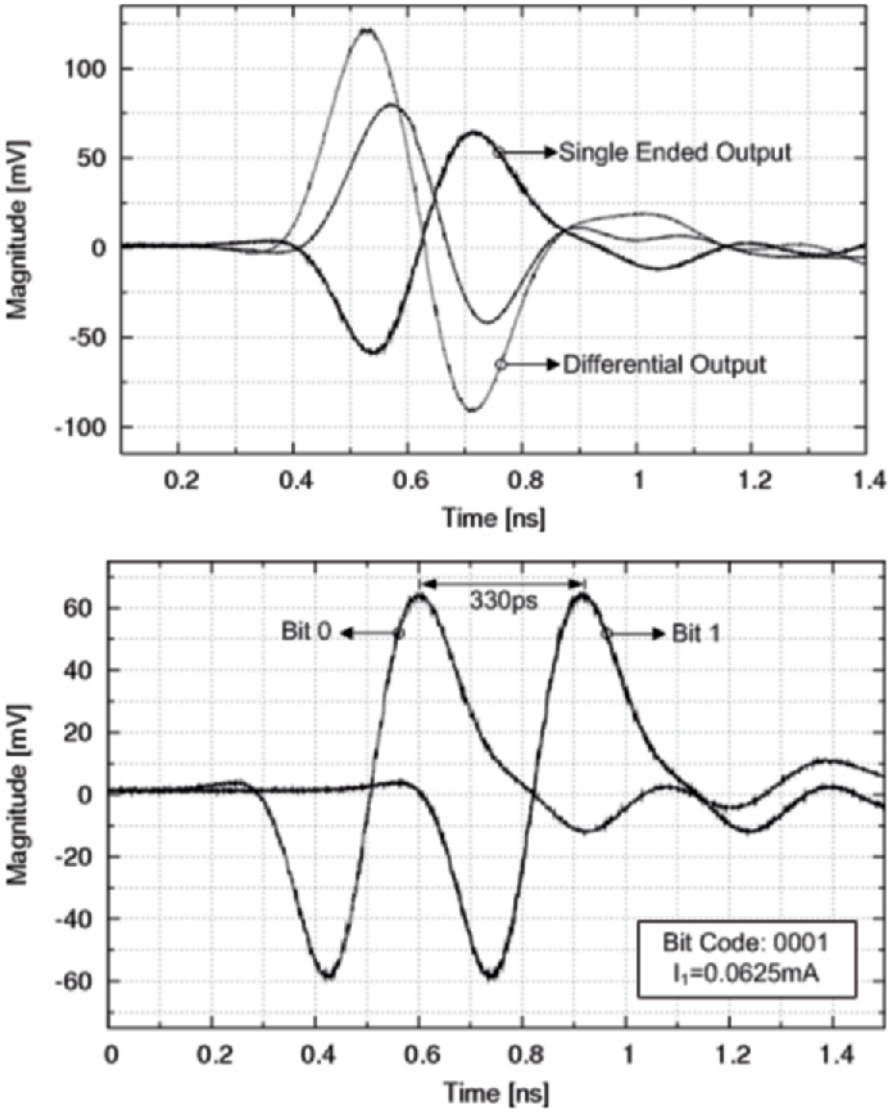


Fig. 9. Top) Gaussian monocycle and differential output; Bottom) Pulse position modulation for bit code 0001.

All measurements were carried out using a TEM-horn antenna as the receiving antenna placed approximately 35cm from the transmitting antenna. Properly designed TEM-horn antennas faithfully reproduce waveforms of the received pulse as their transfer function is flat in the operational frequency band [24]. Figure 10 compares the radiated waveform from the differentially fed butterfly antenna with the measured differential waveform from the generator.

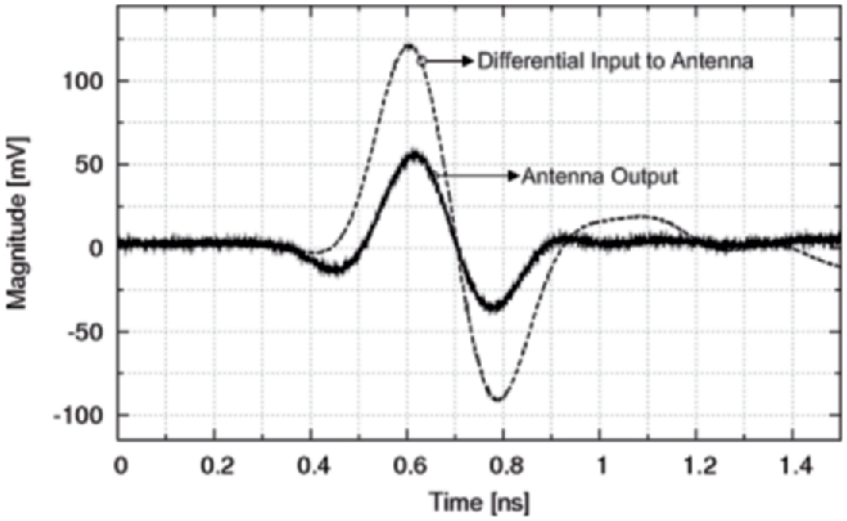


Fig. 10. Waveform (i.e. with 100Ω load) versus waveform transmitted by butterfly antenna.

3. An FCC Compliant Pulse Generator for IR-UWB Communications

As ultra-wideband systems transmit at very low spectral densities and occupy a large amount of bandwidth, it thus is unequivocal that the pulse generator's performance be optimized for maximum energy efficiency. In literature, it is seen that one of the most attractive qualities of Daubechies' wavelets [25] is that they are window-like functions in the frequency domain. This single characteristic accounts for their unique localization property and therefore the Daubechies' wavelet can be used as the transmitted waveform in impulse radio UWB. In Fig. 11, to the right is the mother wavelet and to the left is the scaling function, which is often used to generate the mother wavelet.

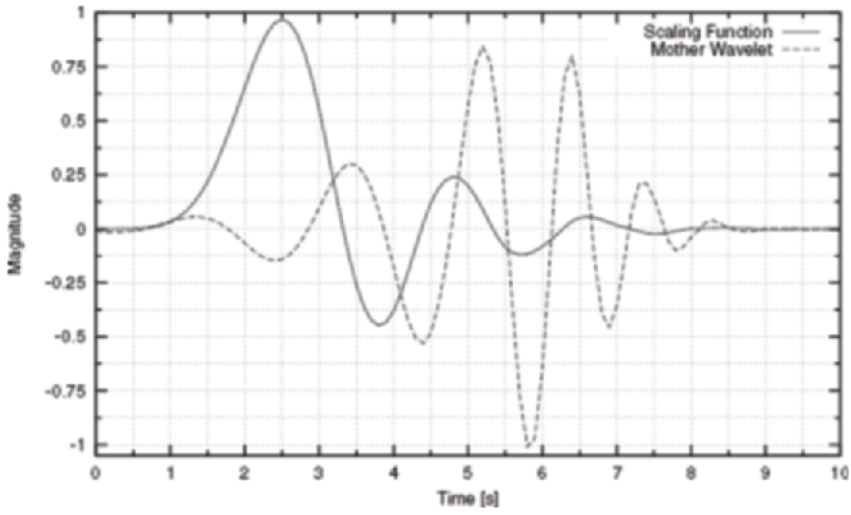


Fig. 11. Scaling Function and Mother Wavelet of Order 8.

3.1. Pulse Generator Model

As seen in Figure 12, by generating a window-like response in the baseband (i.e. with the Daubechies' filter) and then through upconversion, the energy efficiency of the pulse generator can be matched to that of the FCC frequency mask. The focus here is to implement the Daubechies' filter in CMOS technology. For detection in the receiver, the absolute shape of the transmitted waveform is not relevant [3]. Biphase modulation of the transmitted waveform can be achieved by alternating the polarities of the "impulses" that are used to drive the Daubechies' scaling function filter. In the next section, the transfer function of the Daubechies' filter is derived.

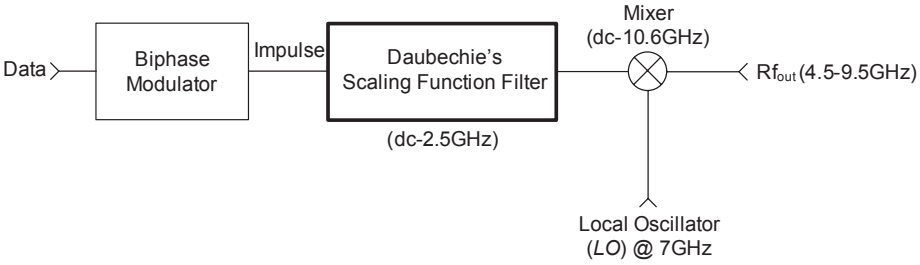


Fig. 12. Proposed pulse generator block diagram.

3.2. Transfer Function and State-Space Synthesis

Trade-offs between roll-off, attenuation and circuit complexity are taken into consideration prior to choosing the filter order of the Daubechies' filter. An eighth order filter is chosen with a bandwidth from dc-2.5GHz. The transfer function of the filter is generated using a Padé approximation [26]. The magnitude transfer as well as the phase response corresponds to that of a Daubechies' scaling function. Once the desired transfer function is formulated, its state-space description is then determined.

Among known standard state-space descriptions, such as the canonical, the diagonal and the modal, the orthonormal ladder form is notable since it is by definition semi-optimized for dynamic range due to the specific structure of the matrices. With a state-space approach, the filter can be optimized for dynamic range, sensitivity, sparsity and coefficient values.

3.3. Scaling: Capacitance and Coefficient Values

Transconductance amplifiers will form the basic building blocks to implement the state-space description coefficients of the filer. The integrators are implemented as capacitors with a normalized value of 1F. The corresponding matrices A, B, and C have extremely large coefficients corresponding to large gm values, which are not physically feasible at circuit level. By scaling the capacitors (cap=0.1pF) and α_1, one consequently scales matrices A and B. Coefficients of matrix C can too be down scaled by α_2, without affecting the response of the filter.

$$\begin{aligned} A^* &= cap \cdot A \\ B^* &= \alpha_1 \cdot cap \cdot B \\ C^* &= \alpha_2 \cdot C \end{aligned} \tag{2}$$

The block diagram of the state-space filter is shown in Figure 13 and has 22 non-zero coefficients.

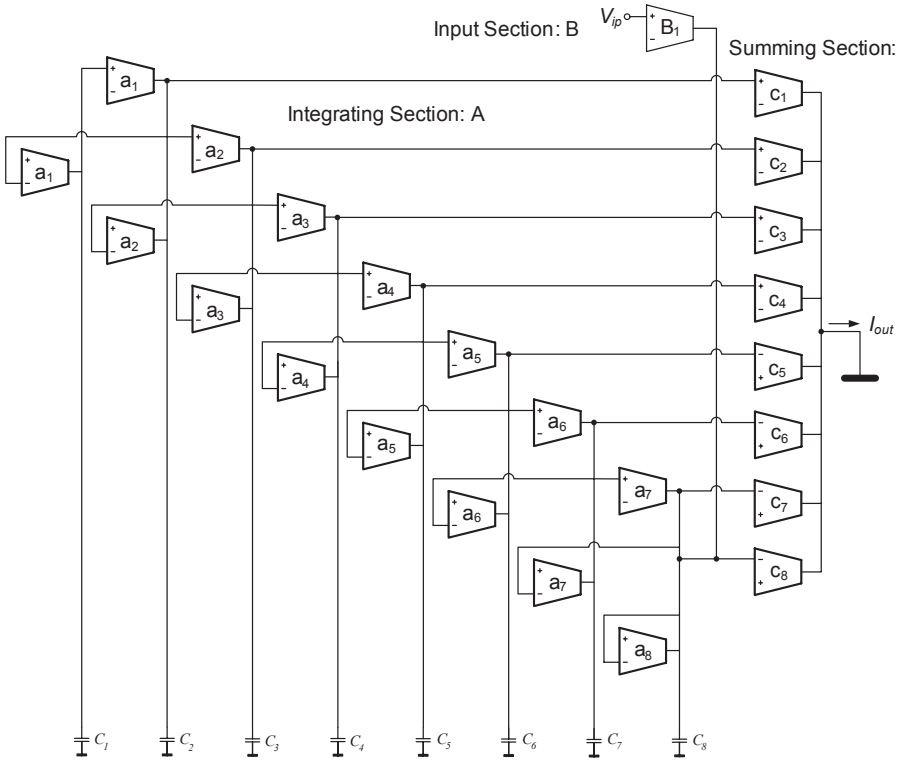


Fig. 13. Complete state-space filter structure.

3.4. Transconductance Amplifier

The transconductance amplifier is implemented using a negative feedback structure consisting of an active circuit, which implements a nullor and a feedback network (see Figure 14).

The nullor is a high-gain block with all the transfer parameters infinite. It is realized using a common source (*CS*) stage formed by transistor (M_1) at the input, and a non-inverting differential pair (M_2-M_3) at the output. The feedback network is made up of a resistor ***R***.

As compared to a single-stage implementation, a 2-stage nullor improves the loop gain, which yields higher linearity as well as bandwidth at the expense of power consumption. In reference to stability, frequency compensation in the form of pole-zero cancellation may be applied to this transconductance amplifier by means of a resistor-capacitor network connected between the gate of M_2 and the source of M_1. For biasing of the differential structure, the common-mode voltage (v_{cm}) is sensed at the outputs by R_{bias} and is compared to the desired reference voltage using a

voltage-controlled current source (*VCCS*). Its implementation is shown on the bottom of Figure 14. The output current delivered by the VCCS is then applied to a virtual ground node, V_X.

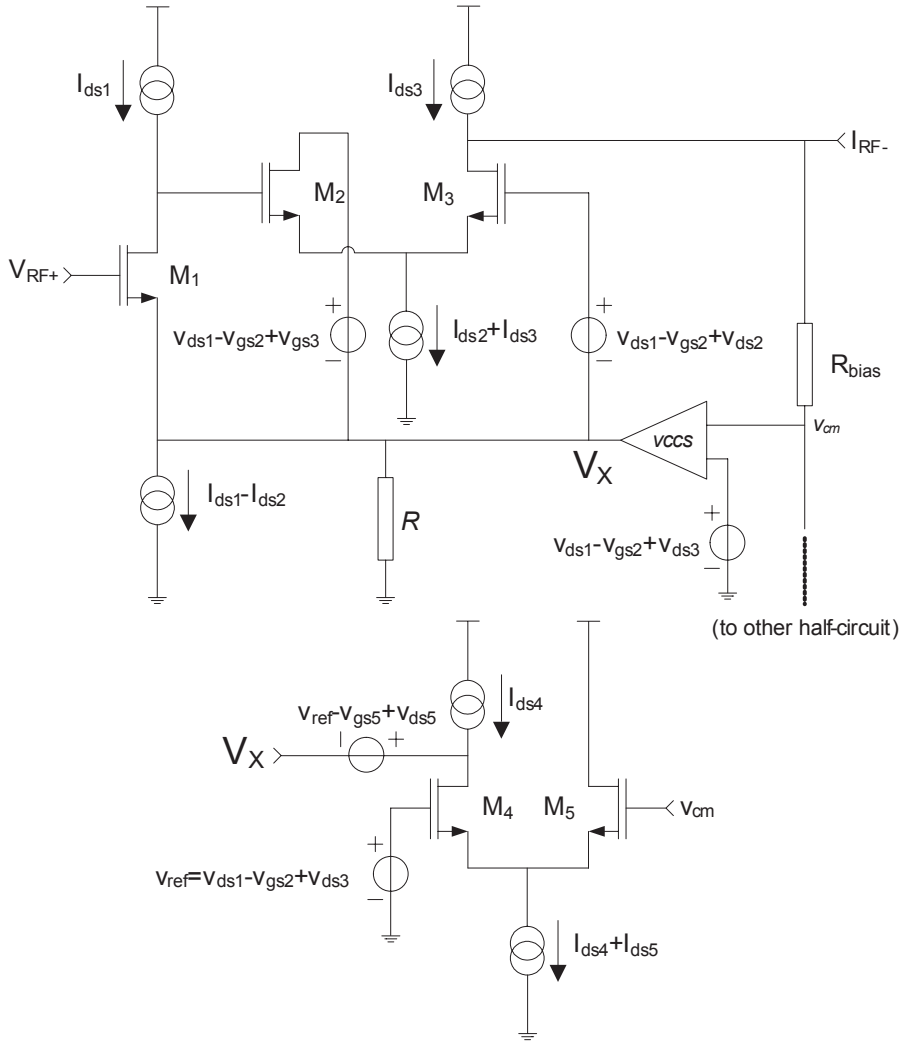


Fig. 14. Negative feedback gm amplifier (half-circuit).

3.5. Simulations

The simulated impulse response of the Daubechies' scaling function filter and the upconverted waveform are seen in Figure 15. By scaling down the capacitance even lower than 0.1pF as well as the coefficients in the matrices *A* and *B*, smaller pulse widths are attainable because of the trade-off between

bandwidth and gain. Figure 16 shows the frequency spectrum of the Daubechies' scaling function upconverted with a 7GHz carrier. As seen in the figure, the pulse generator's performance can now be optimized for maximum energy efficiency.

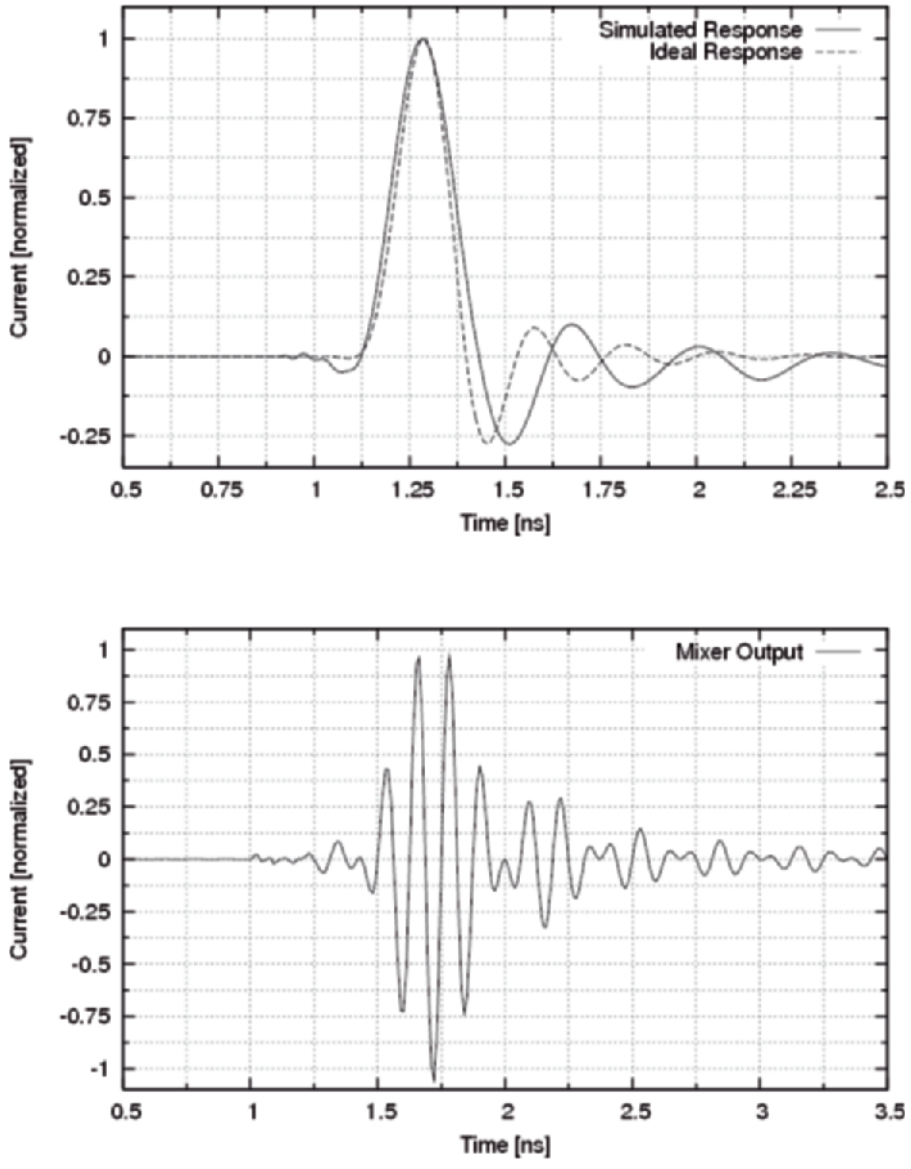


Fig. 15. Top) Impulse response of the 8th order Daubechies' function filter Bottom) Upconverted waveform.

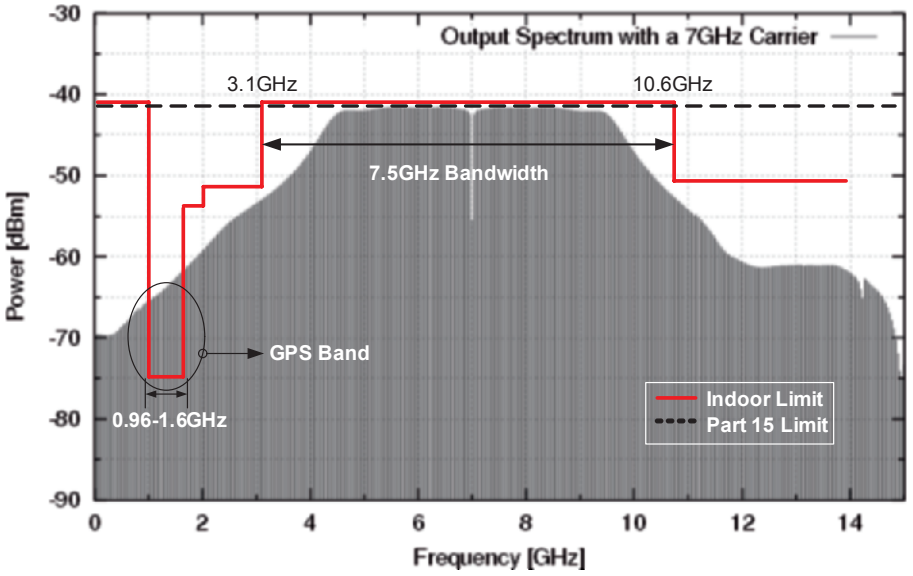


Fig. 16. Frequency spectrum of the upconverted waveform.

It must be noted that there is still some residual signal in the 0.96-1.6GHz frequency band. However, this will be filtered out by the UWB antenna [1]. The simulation parameters of the delay filter are given in Table 1.

Table 1. Simulation Parameters.

Specifications	Simulated
Waveform	Scaling Function of Daubechies' Mother Wavelet
Bandwidth	4.2-9.7GHz
Frequency coverage	@ 85%
Current consumption	25mA @ 1.2V
Size	1.25mm^2
Process	IBM CMOS 0.13μm

4. Quadrature Downconversion Autocorrelation Receiver

In the transmit reference scheme proposed by Hoctor and Tomlinson [27], consecutive pulses are transmitted with a predefined delay τ_d between them. The first pulse acts as a reference, whereas the second pulse is modulated. The autocorrelation receiver correlates the incoming signal with a delayed version of the previous signal. The absolute value of the output after integration is in fact the energy of the pulse while the polarity of the output contains the data.

A *Quadrature Downconversion Autocorrelation Receiver* (see Figure 17) [3] is designed to operate in the presence of strong narrowband interference, while still being able to detect the incoming UWB signal.

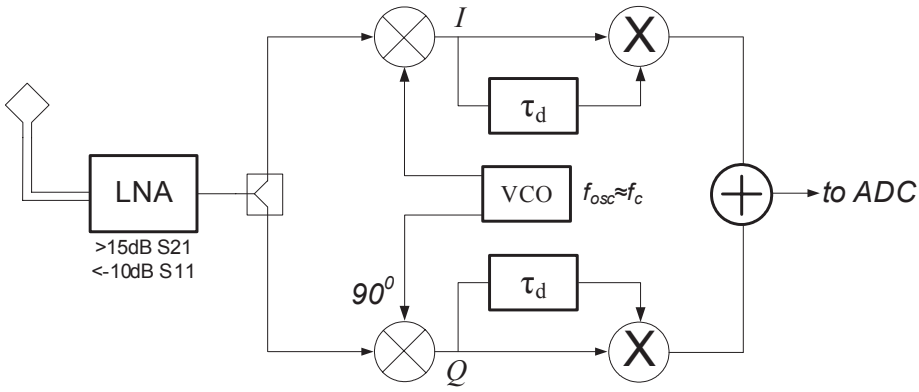


Fig. 17. Quadrature downconversion autocorrelation receiver.

The QDAR employs the principle of frequency wrapping; it folds the ultra-wideband frequency spectrum around the origin. At the same time, the narrowband interferers at 5.5 and 2.4 GHz are positioned outside the band of interest and are simply removed by the means of a band-pass filter (see Fig. 3). Besides resolving the issues of synchronization and capturing multipath energy, the QDAR exploits the fact that detection with an autocorrelation receiver is feasible as long as the relative polarity and shape of consecutive pulses is preserved.

In regards to the transmit reference scheme and from an implementation point of view, one is drawn to the conclusion that the bottleneck to this concept is the physical realization of an accurate continuous-time delay required to execute the autocorrelation function at high frequencies. Therefore, we propose two types of time delay circuits, as seen in the following sections.

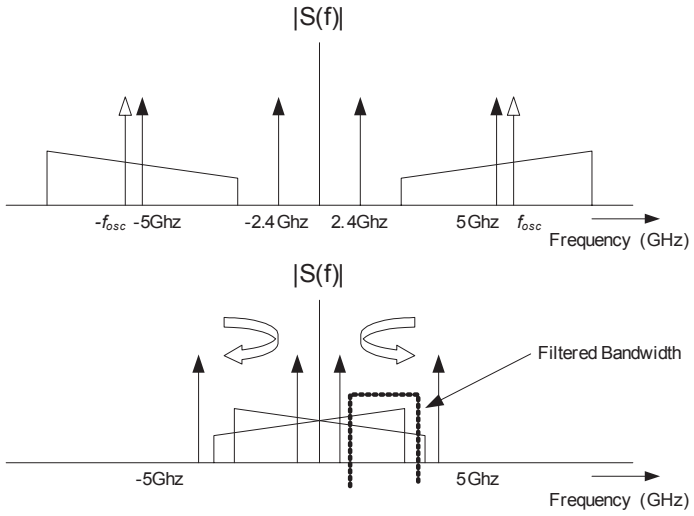


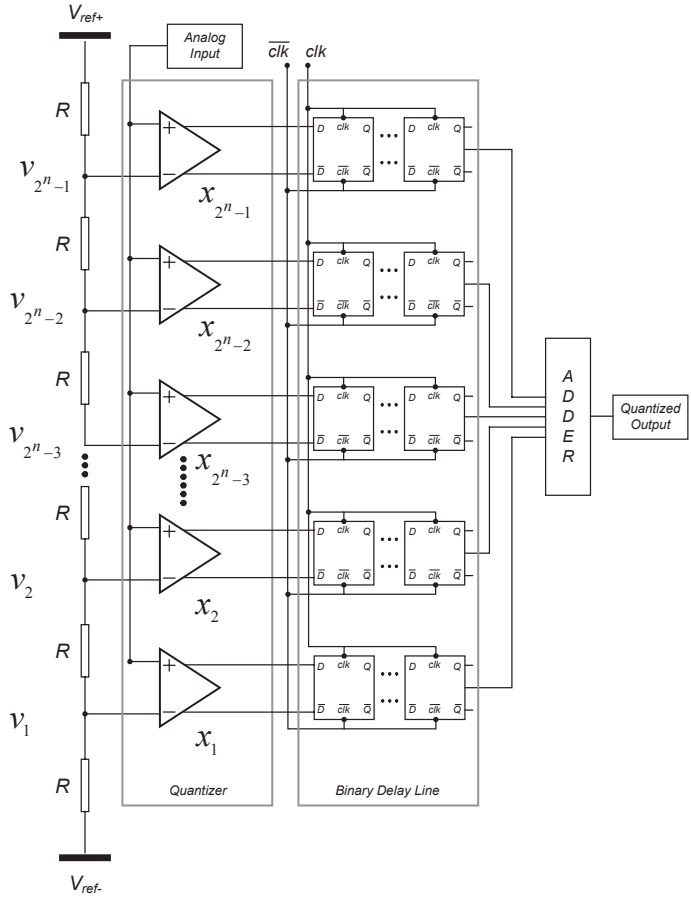
Fig. 18. Top) Frequency spectrum before downconversion Bottom) after downconversion.

5. A Quantized Analog Delay (QAD)

As delaying a binary signal requires significantly less hardware complexity as compared to a continuous-time signal, the continuous-time incoming signal should be *"quantized"* before it is delayed. Consequently, this section presents a quantized analog delay [4] to be used in the QDAR's autocorrelation function.

5.1. System Architecture

As illustrated in Figure 19 (left), the proposed quantized delay is comprised of a quantizer or *"n"* bits, multiple binary delay lines and an adder circuit. Being the foremost element, the quantizer consists of a series of comparators (2^n-*1*), each one comparing the input signal to a unique reference voltage. The comparator outputs connect to binary delay lines, which constitute a cascade of synchronized D-latches. In autocorrelation receivers the accuracy of the delay element is paramount for proper detection, hence a synchronous delay line is employed. Finally, by using an adder circuit, the outputs available at each delay line are linked together to reconstruct the quantized signal.



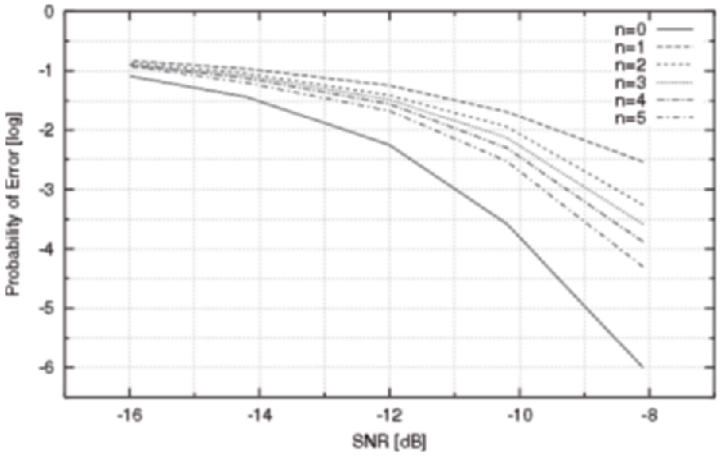


Fig. 19. Top) Quantized Analog Delay *derived from a conventional Flash ADC. Bottom) Probability of error for an* AWGN limited system *for different quantization levels.*

5.2. System Analysis

The quality of the approximation of the incoming signal is directly related to the number of bits of the quantizer (n) and thus influences the power consumption. For varying quantization levels as well as signal-to-noise ratios (SNR), the probability of error is plotted in Figure 19 for the QDAR with two QAD's. This quantitative analysis undoubtedly shows that by increasing the amount of quantization levels, one is capable of reducing the probability of error during autocorrelation. In addition, the value of n is determined based upon the trade-off between power consumption and bit precision, as well as taking into account the probability of error at the desired SNR. $n=0$ refers to the state of no quantization.

5.3. Simulations

In UWB systems, it is generally seen that either the 1st or the 2nd derivative of the Gaussian monocycle is used as the transmitted pulse. Thus, a Morlet waveform (i.e., derived from a Gaussian waveform) is used as the input signal to the QAD with a center frequency at 4.6GHz.

Figure 20 illustrates the performance of the quantized analog delay synchronized to a clock frequency (f_{clk}). The time delay (τ_d) between the analog and the quantized signal is calculated using the following equation,

$$\tau_d = \frac{f_{clk}^{-1}}{2} \cdot (n_o) \tag{3}$$

where n_o denotes the number of D-latches per level. Just to re-iterate, the delay time is paramount for autocorrelation and not the exact preservation of the waveform. A detailed list of specifications is given in Table 2. The time delay of 549ps seen here differs from the actual value by only 0.02%, which is acceptable based on the plots in Figure 6 in [4]. A realistic 10 GHz clock signal was used. In addition, Figure 21 illustrates the simulated and extracted waveforms.

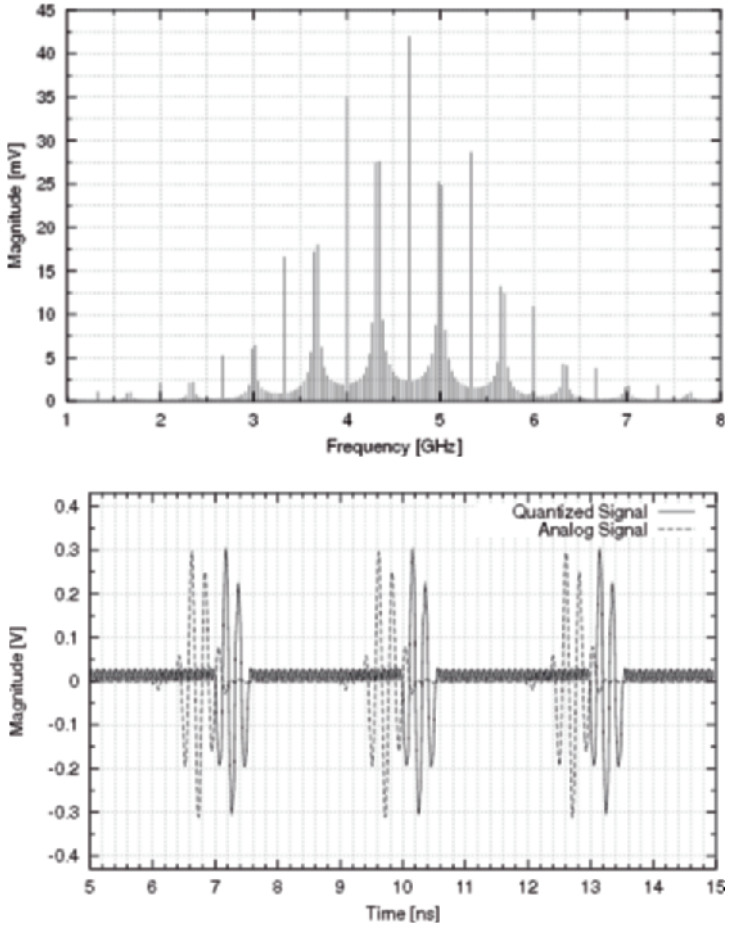


Fig. 20. Top) Frequency spectrum of the input Morlet waveform Bottom) Quantized signal vs. analog signal.

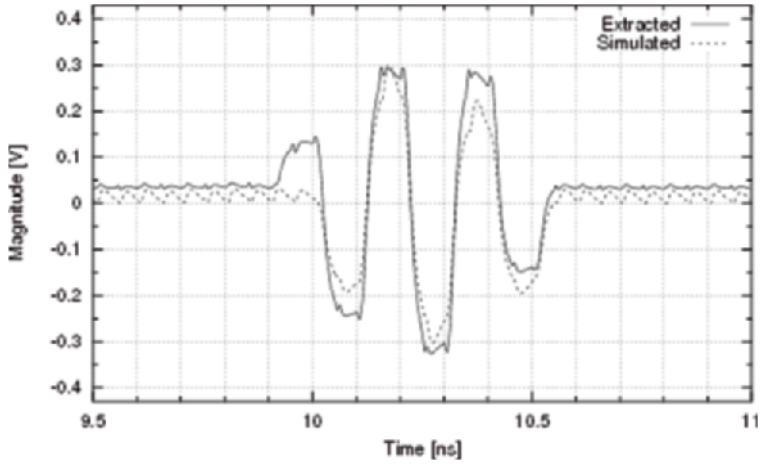


Fig. 21. Quantized signal – simulated vs. extracted.

Table 2. Simulation parameters.

Specifications	Simulated (w.r.t Morlet Waveform)	
Quantizer Precision (n)	3	
Time delay (τd) n_o=10	Actual 550ps	Simulated 549ps
Current consumption	36.7mA @ 1.6V	
Max. operating freq.	4.6GHz	
Clock frequency (fclk)	10GHz	
Process	IBM CMOS 0.13μm	
Size	1.25mm^2	

6. A Filter Based Delay

As the interferers may be stronger than the UWB signal their presence after downconversion and filtering may still deteriorate the outcome of the autocorrelation function. Hence, to further reduce the influence of the narrowband interferers, frequency selectivity is introduced in the delay element [5]. Assuming that the incoming waveform at the receiver, *x(t)*, is a Morlet given by,

$$x(t) = \cos(\alpha \cdot (t-\tau)) \cdot e^{-(t-\tau)^2} \tag{4}$$

where α is the center frequency and τ is the time instant at which the Morlet waveform is centered, the impulse response, *h(t)*, of the delay filter is chosen to ***match*** the incoming signal (i.e. *h(t)=x(t)*) and therefore the output of the time delay, *y(t)*, is the convolution of the incoming signal and the impulse response of the filter, given by,

$$y(t) = \int_0^t \cos(\alpha \cdot (\tau_1 - \tau)) \cdot e^{-(\tau_1-\tau)^2} \cdot \cos(\alpha \cdot (t-\tau_1-\tau)) \cdot e^{-(t-\tau_1-\tau)^2} d\tau_1 \tag{5}$$

From comparing (4) with (5), one sees the distinct resemblance between the convolved waveform and the incoming Morlet signal. The design procedure of the filter based time delay is the same as described in III.

The simulated impulse response of the Morlet filter is seen in Figure 22. By scaling down the capacitance even lower than 0.2pF as well as the coefficients in the matrices *A* and *B*, smaller pulse widths are attainable because of the trade-off between bandwidth and gain.

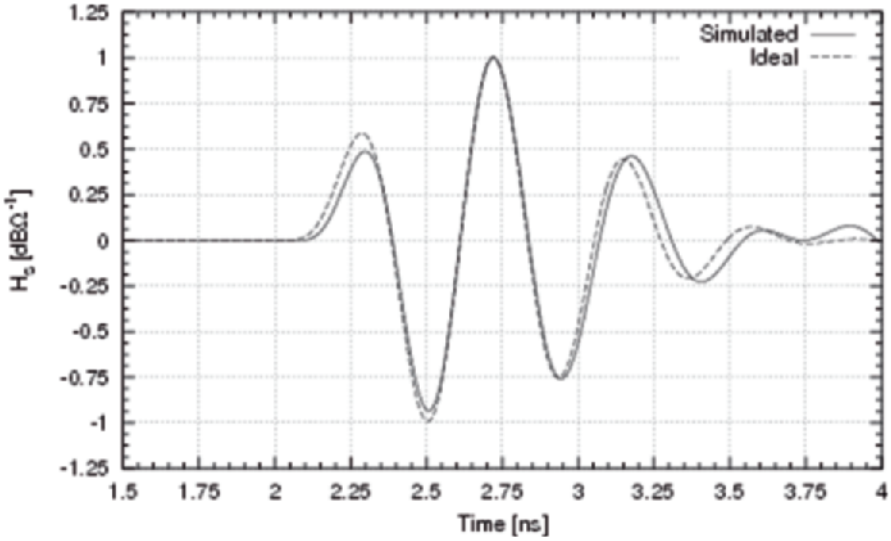


Fig. 22. Impulse response of an 8th order Morlet filter centered at 2.2 GHz.

Figure 23 shows the time delay (τ) between the Morlet waveform and the convolved signal.

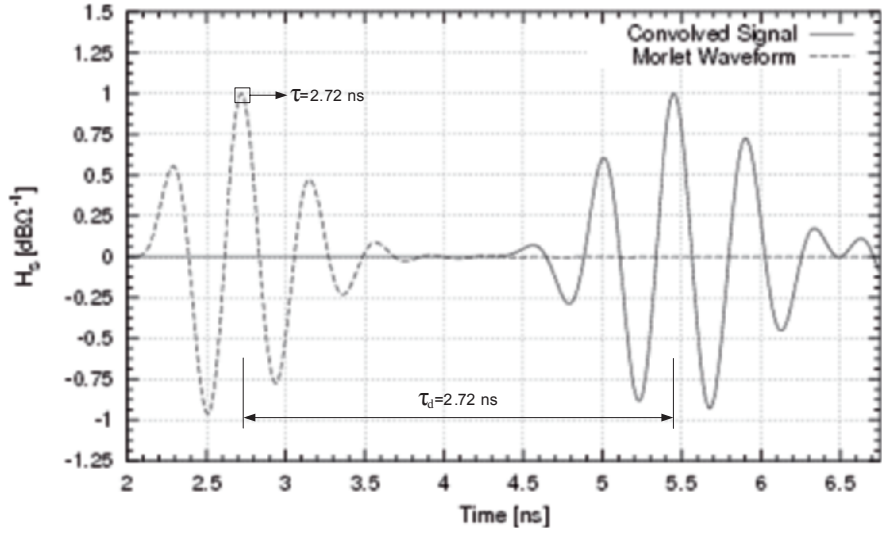


Fig. 23. Time delay between the incoming Morlet waveform and convolved signal.

7. Dual-Loop Negative-Feedback Broadband Low-Noise Amplifier Design

In order to achieve a broad bandwidth, a dual-loop negative feedback topology is favorable because of the possibility to accurately define the input impedance [28]. This section presents a power-to-power (i.e. having accurate input and output impedances) dual-loop negative feedback LNA with a transformer in one of the loops. The topology allows for separate (orthogonal) noise matching and impedance matching.

The design procedure takes into account the influence of transformer non-idealities on the noise and impedance performance of the LNA.

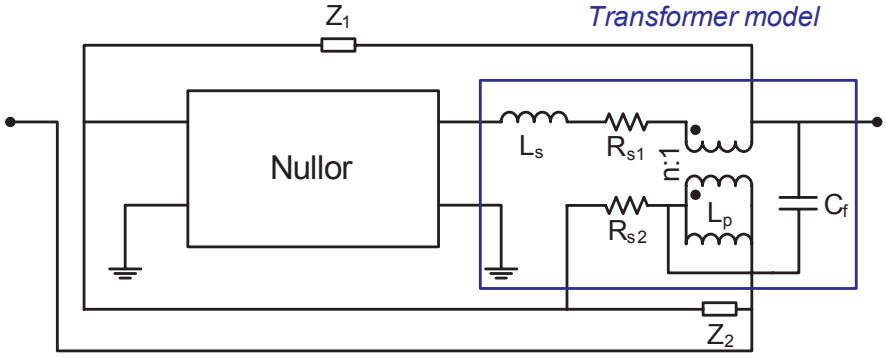


Fig. 24. Power-to-Power amplifier topology with transformer model.

Figure 24 shows the power-to-power dual-loop negative feedback topology with the transformer model and a nullor as the active part. L_s is the series inductance of the primary winding of the transformer. L_p is the parallel inductance of the secondary winding. R_{s1} and R_{s2} are parasitic series resistances. C_f models the parasitic capacitance between the windings. The substrate loss can be considered negligible due to the high resistivity of the GaAs substrate. All the impedances in the figure, Z_1, Z_2, R_{s1} and R_{s2} have series noise sources, which can be transformed to the input of the LNA to analyze the noise performance. Ignoring parasitic capacitance C_f for the moment, the input equivalent noise power spectral density (in V_2/rad/s) equals:

$$S_{Vn,eq,in}(\omega) = 4kT \frac{\omega^2 Lp^2 \frac{(2R_2 - Rs)^2}{R_1} + \left(R_2^2 + \omega^2 Lp^2\right)R_2 + R_2^2 R_{S2}}{\omega^2 Lp^2} \quad (6)$$

When designing the input impedance, the parasitic capacitance of the transformer needs to be considered. C_f shown in Figure 24 models the effective parasitic capacitance in the feedback loop originating mainly from the port-to-port capacitance and which introduces a pole and a zero in the frequency response of the input impedance (Z_{in}). This pole narrows the band of Z_{in} to several hundreds of MHz. To broaden the bandwidth, Z_2 is made inductive. As a result, the input impedance becomes,

$$Z_{in} = \frac{(R_2 + S \cdot L_2) \cdot (1+n) \cdot R_L \cdot Lp \cdot Cf \cdot R_1 \cdot (S + \omega_{Zero1}) \cdot (S + \omega_{zero2})}{n \cdot R_L (S \cdot Lp + R_2 + S \cdot L_2 + R_{s2}) \cdot (1 + S \cdot R_1 \cdot Cf)} \quad (7)$$

where R_L is the load impedance.

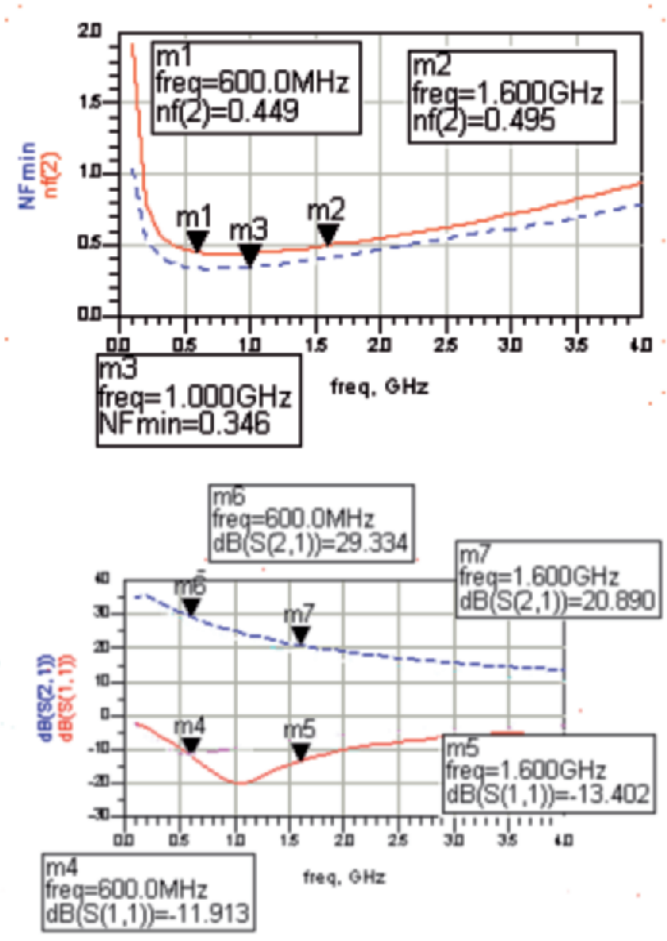


Fig. 25. Top) Noise Figure; Bottom) S11 and S21.

The LNA is to be fabricated in the 0.2μm GaAs process of OMMIC, Philips. The process provides P-HEMTs with cut-off frequencies around 60GHz. The die area amounts to $1.5\times1mm^2$. Due to the fact that the manufacturer provides all the models of real elements including the transmission lines and bends for connection, the simulation results provide a good prediction of the actual performance. Figure 25 shows the simulation results of the LNA. The noise figure is smaller than 0.5dB between 0.6-1.6GHz. The simulation of *stability* was made from 10 Hz to 100 GHz. The stability circles, together with the input and output reflection coefficient, show that the LNA is stable.

8. Conclusions

Impulse radio based UWB has its advantages but at the same time, the impression of straightforwardness in implementation is misleading. UWB is not free from frequency translation and neither from filtering. Most UWB front-end will require some sort of frequency translation (i.e. downconversion to baseband) in order to reduce the constraints on the RF analog blocks (e.g., the delay lines) as well as the mixed signal circuitry. It would be fair to consider that the radio requirements for radio based UWB are demanding (e.g. such as meeting the spectral mask definition in the transmitter). Trade-offs in the areas of antenna design, IC technology should be taken into account in order to achieve a robust design for a radio based UWB system. Only through unconventional methodologies, such as the co-design approach described in this paper, will one deliver a UWB radio with the optimum cost, performance and power consumption mix.

In this paper we introduced design techniques for an ir-UWB transmitter as well as the receiver. We have shown the co-design of an impulse generator and antennas, a Daubechies based impulse generator for the frequency optimization of the FCC mask, a receiver topology (i.e. the QDAR), which relaxes the design constraints on the delay lines in the auto-correlation function and double-loop negative feedback amplifier techniques to realize true wideband LNA's for the receiver front-end.

References

[1] S. Bagga, A.V. Vorobyov, S.A.P. Haddad, A.G. Yarovoy, W.A. Serdijn and J.R. Long, "Co-design of an Impulse Generator and Miniaturized Antennas for IR-UWB," in S*pecial Issue on UWB of the IEEE Transactions on MTT*, April 2006 (to appear).

[2] D.A. Johns, W.M. Snelgrove and A.S. Sedra, "Orthonormal Ladder Filters," *IEEE Transactions on Circuits and Systems*, vol. 36, pp. 337-343, March 1989.

[3] S. Lee, S. Bagga and W.A. Serdijn, "A Quadrature Downconversion Autocorrelation Receiver Architecture for UWB," in *IEEE Joint UWBST & IWUWBS 2004*, pp. 19-21, May 2004.

[4] S. Bagga, L. Zhang, W.A. Serdijn, J.R. Long and E.B. Busking, "A Quantized Analog Delay for an ir-UWB Quadrature Downconversion Autocorrelation Receiver (QDAR)," in *IEEE International Conference on Ultrawideband*, pp. 328-332, September 2005.

[5] S. Bagga, S.A.P. Haddad, W.A. Serdijn, J.R. Long and E.B. Busking, "A Delay Filter for an ir-UWB Front-End," in *IEEE International Conference on Ultrawideband*, pp. 323-327, September 2005.

[6] J. Xu, B. Woestenburg, J.G. Bij de Vaate and W.A. Serdijn, "GaAs 0.5dB NF Dual-Loop Negative-Feedback Broadband Low-Noise Amplifier IC," *IEEE Electronics Letters*, vol. 41, n0. 14, pp.780-782, July 2005.

[7] FCC 02-48 – Federal Communication Commission, "First Report and Order regarding Ultra Wideband Transmission System", *http://www.fcc.gov, adopted*: February 14, 2002, released: April 22, 2002.

[8] D. Porcino and W. Hirt, "Ultra-wideband radio technology: potential and challenges ahead," *IEEE Communications Magazine*, vol. 41, no. 7, pp. 66-74, July 2003.

[9] L. Yang and G. B. Giannakis, "Ultra-wideband communications: an idea whoes time has come," *IEEE Signal Processing Magazine*, vol. 21, no. 76, pp. 26-54, November 2004.

[10] M. Z. Win and R. A. Scholtz, "Impulse radio: how it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36-38, February 1998.

[11] S. Roy, J. R. Foerster, V. S. Somayazulu, and D. G. Leeper, "Ultrawideband radio design: the promise of high-speed, short-range wireless connectivity", *Proc. of the IEEE*, vol. 92, no. 2, pp. 295-311, Feb. 2004.

[12] A. Yarovoy, "Ultra-Wideband Systems", *Proc. 33rd European Microwave Conference*, pp. 597-600, October 2003.

[13] H. G. Schantz, "A brief history of ultra wideband antennas", *IEEE Aerospace and Electromagnetic System magazine,* vol. 19, issue 4, pp. 22-26, April 2004.

[14] R. A. Scholtz, "Multiple Access with Time-Hopping Impulse Modulation," *Proc. MILCOM*, October 1993.

[15] X. Chen, S. Kiaei, "Monocycle shapes for ultra wideband system", *IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 597-600, May 2002.

[16] H. Kanada, N. Aoshima, "Analog Gabor Transform Filter with Complex First Order System," in *Proc. SICE*, pp. 925-930, 1997
[17] J. M. Wilson, "Ultra-Wideband/a Disruptive RF Technology?" *Version 1.3, Intel Research and Development*, September 10, 2002.
[18] R.V. de Jongh, A.G. Yarovoy, L.P. Ligthart, I.V. Kaploun, A.D. Schukin, "Design and analysis of new GPR antenna concepts", *Proceedings, Seventh International Conference on Ground-Penetrating Radar*, Lawrence, vol. 1, pp.81-86, 1998.
[19] A.G. Yarovoy, R. Pugliese, J.R. Zijderveld, L.P. Ligthart, "Antenna Development for UWB Impulse Radio," *34th European Microwave Conference*, pp. 1257-1260, 2004.
[20] H. C. Morgan, W. H. Boyd, "Transmission of Electronic Information by Pulse Position Modulation Utilizing Low Average Power," *United States Patent*, No. 5,586,145, December, 1996.
[21] M. Saint-Laurent, M. Swaminathan, "A Digitally Adjustable Resistor for Path Delay Characterization in High-Frequency Microprocessors," *IEEE Journal of Solid State Circuits*, 2002.
[22] E.W. Justh, F.J. Kub, "Analogue CMOS continuous-time tapped delay line circuit," in *IEEE Electronic Letters*, vol. 31, no. 21, October 1995.
[23] R. Wang, R. Harjani, "Partial Positive Feedback for gain-Enhancement of Low-Power CMOS OTAs*," Analog Integrated Circuits and Signal Processing*, no. 8, pp. 21-35, 1995.
[24] A. G. Yarovoy, J. H. Zijderveld, A. D. Schukin, I. V. Kaploun, and L. P. Ligthart, "Dielectric Wedge Antenna for UWB Applications," *2005 IEEE International Conference on Ultra-Wideband*, Zurich, Switzerland, pp.186-189, September 2005.
[25] I. Daubechies's, "Ten Lectures on Wavelets," *Society for Industrial and Applied Mathematics*, Philadelphia, vol. 61, pp. 194-202, 1994.
[26] S.A.P. Haddad, S. Bagga and W.A. Serdijn, "Log-Domain Wavelet Bases," in *Proceedings IEEE International Symposium of Circuits and Systems*, May 2004.
[27] R.T. Hoctor 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.
[28] E.H. Nordholt, "Classes and Properties of Multi-loop Negative-Feedback Amplifiers," *IEEE Transactions on Circuits and Systems*, vol. cas-28, no. 3, pp. 203-211, March 1981.

CIRCUITS AND TECHNOLOGIES FOR WIRELESS SENSING

Brian Otis[*], Nathan Pletcher[†], Shailesh Rai[*], Fred Burghardt[†], Jan Rabaey†
[*]University of Washington, Seattle, WA, USA
[†]University of California, Berkeley, CA, USA

Abstract

This work describes ongoing research into system architectures, circuit design techniques, and new technologies applicable to low power wireless sensing. We will present completed proof-of-concept research as well as propose ideas for future architectures. It is shown that MEMS-based transceiver blocks in combination with a dedicated carrier sense receiver can substantially reduce the communications energy of a sensor network. Additionally, novel methods for power regulation and modular packaging will be introduced.

1. Introduction

The focus of this work is on low power communications circuits and supporting technologies for wireless sensor networks. Among the many design considerations for sensor networks transceivers, three stand out as the most crucial: power consumption, implementation volume, and cost. This paper describes research in three areas aimed at enabling new applications in wireless sensing. The first research topic involves the utilization of RF MEMS components in the design of transceiver blocks. The goal of this effort is to shift the burden of filtering RF signals from the active circuitry to passive micromechanical components, relaxing the specifications on the active circuitry. The second topic involves the design of a carrier sense receiver aimed at reducing the overall energy and latency of the network. While operating at very low quiescent power levels, this receiver will sense a transmit beacon and enable the main data radio when a packet is available on the network. The third topic concerns miniature assembly and packaging techniques to allow a modular-yet-small form-factor. Section 2 will describe our system requirements and a prototype proof-of-concept integrated chip. Section 3 proposes new building blocks of future low power transceivers. Section 4 describes a carrier sense architecture to increase link efficiency. Section 5 presents an overview of power management and enclosure design for low power wireless sensors.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 285-301.

2. System Requirements

Of the many challenges involved in the implementation of ultra-small wireless sensors, one of the largest is the design of the communications link. Typically, the network must support peer-to-peer connections between nodes, so the radio link must be symmetric. This eliminates the possibility of using RFID-type passive radios for receiving data. Since both sides of the data link are energy constrained, the receiver sensitivity must be high. Thus, for a given link range, a tradeoff exists between design for low receiver noise and high transmit power level. Size and cost considerations are also paramount in such ubiquitous systems. Unfortunately, a traditional implementation with multiple off-chip components adds prohibitively to the cost and form-factor, so an entirely thin-film, batch fabricated solution is desirable.

There has been great progress in the MEMS community in recent years in designing MEMS filters and switches for use in communications systems at very high frequencies. From an RF circuit architect point of view, these emerging technologies hold the potential to dramatically reduce the power consumption and complexity of the RF active circuitry. For example, a steep bandpass channel select filter can ease the linearity requirements of the LNA or even allow a sub-sampled receive architecture. Also, since MEMS devices are batch-fabricated on semiconductor substrates, they are physically small and ultimately integratable with the electronics.

As a specific example, recent advances in BAW resonator technology have allowed very high quality factors (>1000) at GHz frequencies [1]. Measured data from a 2.4GHz BAW resonator is shown in Fig. 1.

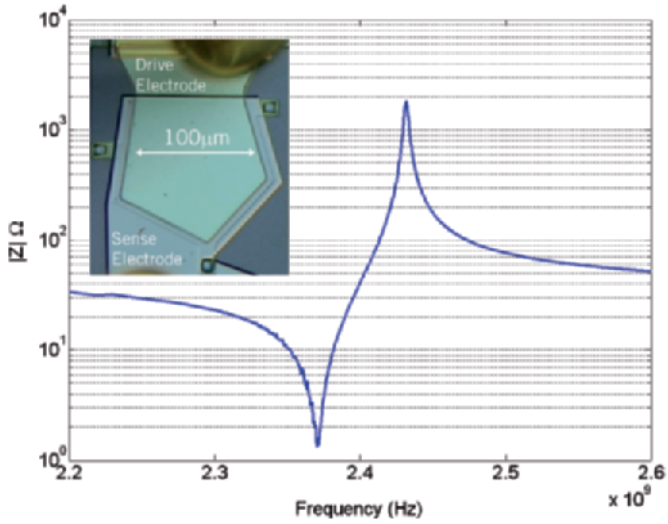


Fig. 1. Measured impedance of single 2.4GHz BAW Resonator. Inset photograph shows the 100um × 100um active resonator area.

The high Q nature of the structure is clearly visible in the impedance plot. In this proof-of-concept work we utilize the high Q and stable oscillation for creating RF oscillator circuits. In past implementations, we have leveraged this unique transfer function to construct an extremely selective RF amplifier [2]. The form factor of the resulting system can be further reduced by wafer thinning and flip-chip assembly.

These techniques were used to demonstrate a second-generation low power super-regenerative transceiver. This design used two high-frequency 1mm^2 BAW resonators alongside a 2mm^2 0.13μm CMOS chip [3].

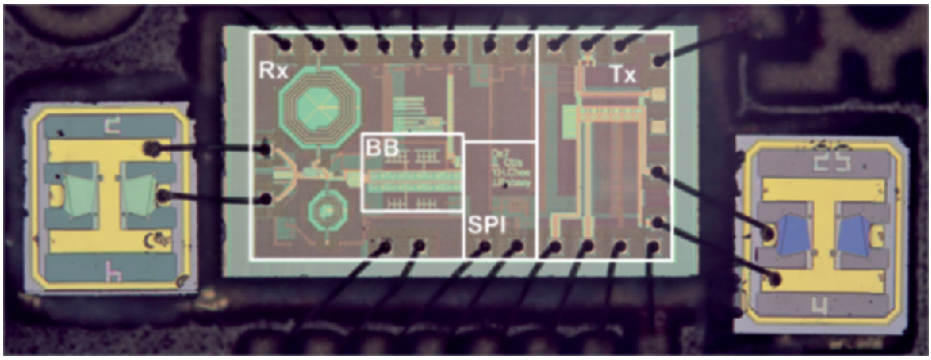


Fig. 2. CMOS/BAW Resonator Transceiver Implementation.

The receiver consumes less than 400μW when active, and supports datarates up to 50kbps. A three-wire SPI interface bus controls all receiver functionality, including variable datarate, RF chain bias current, baseband gain, and baseband bandwidth. A sensitivity of approximately -100dBm permits a very low transmitted power level. A small capacitor bank allows frequency compensation of the relatively large BAW temperature coefficient. The remainder of this brief builds on these concepts and introduces research leading to next generation wireless devices.

3. BAW-Referenced Tunable Oscillators

In previous work, BAW-referenced CMOS RF oscillators have been demonstrated [4]. However, the usefulness of fixed-frequency oscillators is limited and tunability is desired in many radio architectures. This section discusses preliminary work in the development of another key building block of low power transceivers. Quadrature voltage controlled oscillators (QVCO) are a crucial component in highly-integrated image-rejection transceivers. In these architectures, quadrature sinusoids are needed for I/Q modulation/demodulation. In low-IF, zero-IF or wide-band multi-standard

transceivers [5], quadrature signal generation by the VCO requires a minimization of the power consumption, quadrature phase error, and phase noise. These conflicting requirements become even more severe in the implementation of highly integrated, low power transceivers for sensor networks. Several creative implementation techniques exist for quadrature signal generation [6][7][8]. The goal of this work is to overcome the tuning limitations of a BAW referenced oscillator to construct a low phase noise QVCO.

3.1. VCO Design Using MEMS Resonators

High Q resonators are appealing for oscillator design since they promise low phase noise for a given signal power. There are, however, two main challenges that limit their applicability to use in a VCO. First, high Q structures are inherently limited in their tuning range. The insensitivity of the oscillation frequency to external impedances is one of the great advantages of the ubiquitous quartz crystal oscillator. This is in direct conflict with the need for a QVCO with a wide tuning range. Additionally, unlike LC resonators, MEMS resonators do not provide a stable DC oscillator operating point, introducing additional circuit design challenges.

Fundamentally, the structure of BAW can be visualized as a capacitor whose dielectric material is piezoelectric. Excitation of the acoustic wave occurs upon the application of an alternating voltage across the capacitor, which causes the piezoelectric material to cyclically expand and contract, thus eliciting energy oscillation between mechanical and electrical domains. Fig. 4 shows an equivalent electrical circuit of a BAW, where L_x, R_x and C_x represent motional inductance, resistance and capacitance, respectively. The impedances C_o and R_o represent shunt capacitance and resistance and Z_s and Z_p represent the series and parallel loading on the resonator due to the CMOS circuitry, respectively.

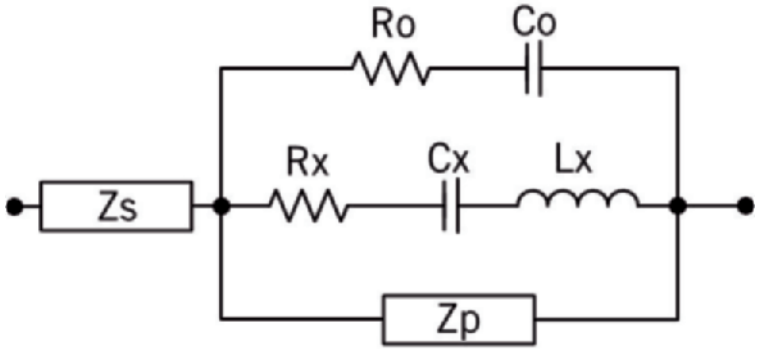


Fig. 4. Electrical equivalent of a BAW resonator.

The resonance frequency can be tuned by modifying Z_s and/or Z_p from the CMOS circuit. For example, as shown in Fig. 5, the frequency tuning can be achieved either by digitally tuning a capacitor bank in parallel with resonator to change its resonance frequency or by electrically switching different resonators. Alternatively, if electrostatic MEMS resonators are employed, the resonance frequency can be modified by varying the resonator bias voltage [9].

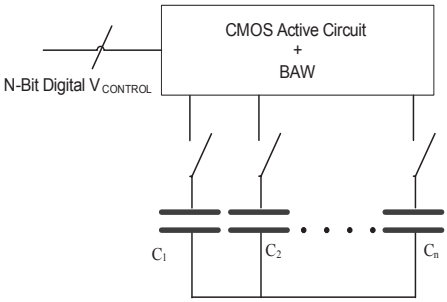


Fig. 5 (a). Tuning using capacitor bank.

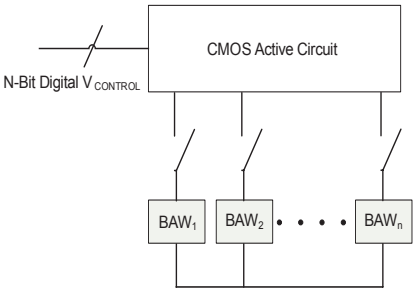


Fig. 5 (b). Tuning using multiple BAWs.

3.2. Tuning Issues and PLL Architectures Using QVCOs

The electrical tuning of a VCO using a MEMS resonator becomes a critical design issue because the inherent mechanical properties of the resonator itself cannot be dynamically changed. One viable option is to digitally switch between multiple resonators to obtain course variable frequencies. The digital switching of high Q MEMS resonators operated in their parallel resonance can be realized since the switch resistance will not dramatically degrade the resonator quality factor. The sensitivity of oscillation frequency to capacitive variation using MEMS devices is much less than the case of LC resonator design with lower Q, and thus it requires very fine frequency control with relatively large switched capacitors.

A low power, low phase-noise QVCO can increase the efficiency of sensor node transceivers and allow more efficient modulation schemes. This ongoing research involves the design and fabrication of a MEMS-referenced QVCO in the GHz frequency range.

4. Carrier Sense Radios

In a sensor network environment, where communication is relatively infrequent and packet lengths are short, the power consumed by a node while monitoring the channel for activity can become significant. The most

common solution is heavy duty cycling of the transceiver physical layer, which greatly reduces the average power consumption. However, in order for communication to be successful, there must be a rendezvous method to ensure that both the transmitting and receiving nodes are active simultaneously. In this section, the circuit techniques for a hardware implementation of a truly asynchronous rendezvous scheme are discussed.

4.1. Motivation

Several different classes of rendezvous methods have been proposed [10]. In a fully synchronous approach, all nodes are globally synchronized and wake up simultaneously. While attractive for very dense networks, it is be difficult to maintain global synchroniszation in an ad-hoc network. So-called pseudo-asynchronous rendezvous techniques combine periodic duty-cycling with a beaconing process, where one node transmits its desire to communicate repeatedly until the receiving node enters the active state. Pseudo-asynchronous rendezvous obviates the need for global synchronization but increases the network latency. In addition, the power consumed by the beaconing process can become significant.

An alternative approach is a truly asynchronous rendezvous, where nodes continuously monitor the channel for beacons using an auxiliary carrier sense receiver (CSR), and wake up only when data is being transmitted. The system is illustrated in Fig. 6, where the carrier sense receiver detects a beacon signal and activates the main receiver for data transfer. Because the CSR is continuously monitoring the channel, its power consumption must be a fraction of that of the main receiver to have a net energy savings. However, the power consumed by transmitter beaconing in the pseudo-asynchronous approach is greatly reduced. In addition, the network latency is diminished because the carrier sense receiver typically responds to the first beacon signal sent by the transmitter.

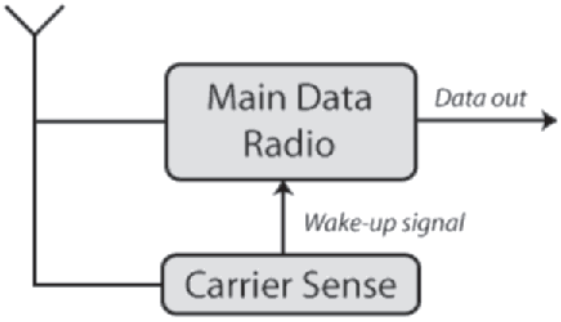


Fig. 6. Block diagram of the asynchronous carrier sense concept.

Clearly, a truly asynchronous rendezvous scheme using an auxiliary CSR may provide power savings and improved network latency. In the next section, the design considerations for such a receiver are outlined in more detail.

4.2. Design Considerations

The design specifications of the CSR are closely tied to the network characteristics of the target application. For example, if packet traffic in the network is low, there will be long periods of monitoring the channel without communication and the CSR power consumption must be very small. On the other hand, under high traffic conditions the time spent monitoring compared to communicating will be smaller, relaxing the power specification of the CSR. Additionally, the properties of the main data radio strongly influence the design of the CSR. In terms of power, the *ratio* of power consumption of the main radio to the CSR is the important metric. Consequently, a low power main radio requires a correspondingly low power CSR. It is also critical that the two receivers have equal communication range. Clearly, if the CSR has shorter range than the main radio, then it will be impossible to awaken some nodes that are within the communication range of the main radio. For similar reasons, it is desirable for the two radios to use approximately equal carrier frequencies. Varying channel conditions, such as deep fading, can occur in one band but not in another, again with the result that either the CSR or main receiver is unreachable. Furthermore, similar carrier frequencies allow the two receivers to share the same antenna, which is necessary from an integration perspective.

The CSR system is also unique because, compared to a traditional receiver, a higher bit error rate (BER) can be tolerated. For the CSR, bit errors translate into either missed wakeups or false wakeups. For the former case, the transmitter must re-send the wakeup beacon; in the latter case the main receiver is awoken erroneously by the CSR and no data is received. In both cases, the overall power increases due to unnecessary operation of the main data radio. In addition, missed wakeup errors negatively impact the network latency. However, preliminary simulations indicate that the overall system power will not increase appreciably until the rate of CSR errors approaches 10^{-1}. Fig. 7. shows a system simulation of the global average power consumption as the amount of CSR errors are increased.

Compared to conventional wireless receivers, the acceptable error is about 2 orders of magnitude higher, allowing a relaxed SNR specification at the output of the CSR.

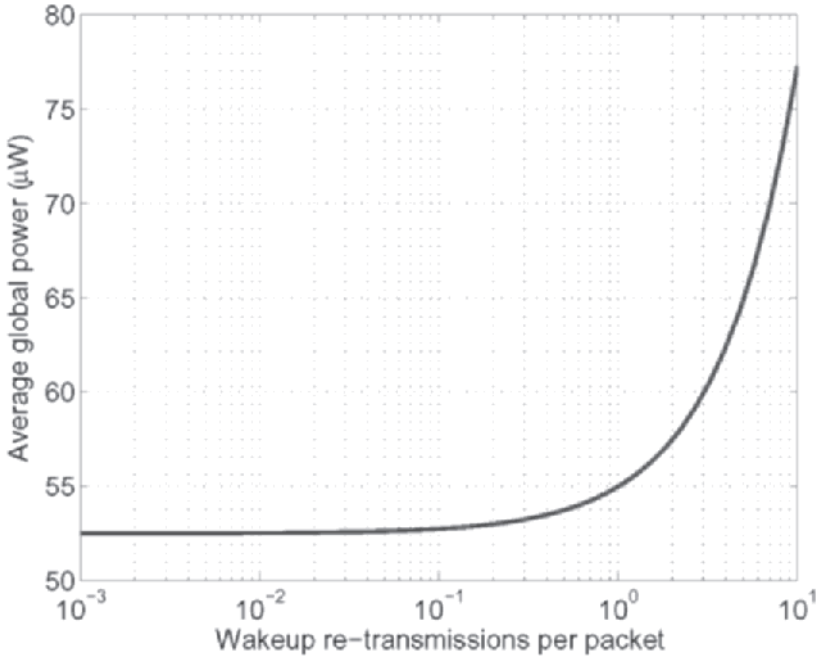


Fig. 7. Effect of carrier sense errors on average power consumption.

Examining the considerations above, the CSR specifications for a typical sensor network application can be determined. State-of-the-art transceivers for sensor nodes consume 1mW or less for the receiver and several milliwatts for the transmitter [3]. High levels of integration are required to reduce node size and cost, leading to the choice of relatively high carrier frequency. A summary of the resulting CSR specifications, designed to operate with the main data radio from [3], is shown in Table 1.

Table1. CSR specifications for a sensor node application.

Carrier frequency	2 GHz
Range (0-3dBm Tx power)	10 m
Power consumption	< 50μW

The challenging power consumption requirement (50μW) dictates a greatly simplified RF frontend and aggressive use of power-saving circuit techniques. Next, two example circuit design blocks for the CSR are discussed.

4.3. A Passive 1.9GHz Receiver

In order to explore the limits of minimum power receiver design, the first example illustrates a "passive" tuned RF receiver operating at 1.9GHz. The frontend can be considered passive because there are no active gain blocks before the frequency downconversion. The RF input signal passes through a high-Q channel select filter and is fed directly to an envelope detector comprised of an NMOS transistor operating in the subthreshold region. The envelope detector acts as a self-mixer, converting the RF-modulated signal down to baseband. The frontend schematic and die photo are shown in Fig. 8.

In this design, a BAW resonator is employed to achieve high-Q filtering in a small form factor. The BAW can be seen in the photo, wire-bonded to the CMOS die. The measured performance of the passive receiver is summarized in Fig. 9. Although the power consumption is very low, the sensitivity is inadequate for a peer-to-peer sensor network environment. Nevertheless, the passive receiver implementation establishes a benchmark

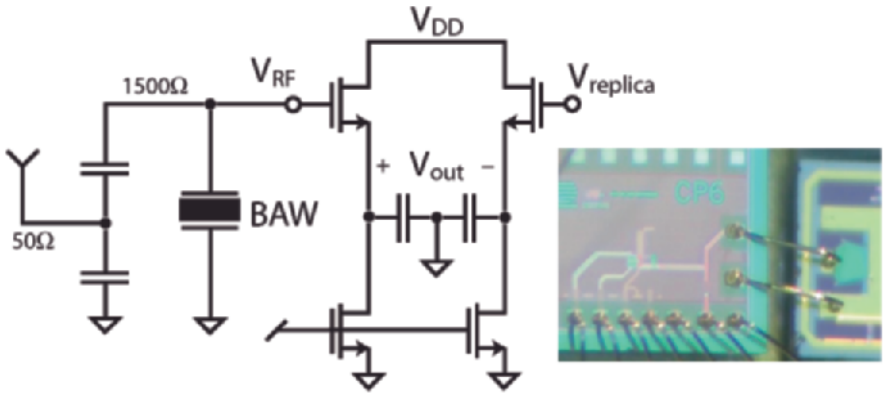


Fig. 8. Passive receiver schematic and implementation die photo.

Measured Performance

Technology (CMOS)	0.13µm
Rx power	200 nW
RF bandwidth (-3dB)	4 MHz
Sensitivity (12dB SNR)	-38 dBm

Measured $|S_{11}|$: -9.3dB

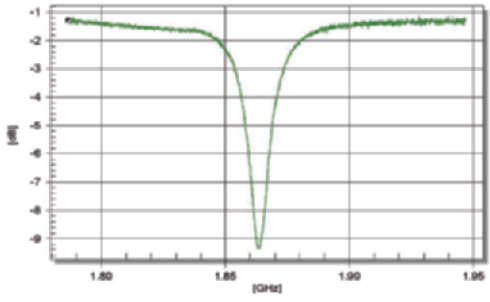


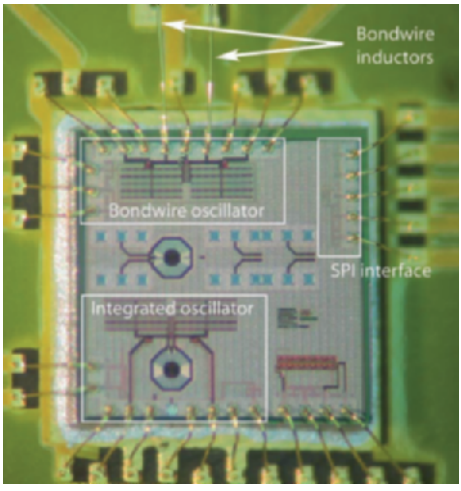
Fig. 9. Passive receiver input match and performance summary.

for a minimal power, high frequency receiver design. In the next section, the power performance of an active circuit, the RF oscillator, is investigated.

4.4. A Low Power Tunable 1.9GHz Oscillator

The RF oscillator is a component employed in most receiver architectures, and its performance has a significant influence on overall receiver performance. However, the oscillator also frequently consumes a significant portion of the overall power budget. The oscillator described in [11] aims to ascertain the feasibility of an ultra-low power RF oscillator for use in the implementation of the CSR.

Two principal techniques are utilized in the design to lower the power consumption. First, the main transconductor devices are designed to run in the weak inversion regime, where the available transconductance is higher for a given bias current. In modern scaled CMOS technologies, sufficient device f_t is available to allow transistors to be biased in weak inversion, even for operation at RF. Secondly, the architecture and biasing are chosen to enable the use of a 0.5V supply, which is less than half the nominal supply for the 0.13µm CMOS technology. In addition, bondwire inductors (3nH) are used to take advantage of their high quality factor. Fully digital frequency tuning is accomplished through a bank for switched capacitors, providing 150MHz of tuning range around 1.9GHz with minimum frequency steps of less than 200kHz.



Measured Performance

Supply voltage	0.5 V
Power	100 µW
Center frequency	1.87 GHz
Tuning range	155 MHz
Tuning resolution	200 kHz
Differential output swing	250 mV
Phase noise @ 1 MHz offset	-114dBc/Hz

Fig. 10. Low power oscillator die photo and performance summary.

Two oscillators were fabricated in 0.13µm CMOS, with one oscillator using an integrated inductor to allow performance comparison with the bondwire version. A photo of the bonded die is shown in Fig. 10, along with a summary of the measured performance.

The bondwire oscillator was also tested at supply voltages below 0.5V, and the circuit operated without performance degradation down to supply of 0.3V. At this bias point, the test oscillator demonstrated at least 100mV output swing while consuming around 60µW. The implementation validates the subthreshold design techniques and demonstrates the feasibility of an RF oscillator operating under 100µW. Together with the passive receiver proof-of-concept, the low power oscillator establishes a baseline for the design of the 50µW CSR.

5. System Design Issues

The circuits discussed so far were designed to consume very little power and realize high levels of integration. However, in order to properly utilize these advances, the system built around the radios must also be efficient and small. For example, supply voltage generation and packaging cannot be neglected and can add significant power and size overhead to a miniature sensor node. In this section, ongoing work in the areas of power management and enclosure desigis discussed.

5.1. Power Management

The effect of power management on overall system power efficiency is often overlooked. Power consumption of commercial off-the-shelf (COTS) devices has historically been orders of magnitude below the total system power budget. However, as the total system power enters the microwatt range, the proportion of power lost in management becomes significant. For example, Fig. 11 shows the efficiency graph of a typical step-down switching regulator. Below 1mA the efficiency drops sharply to the point where the regulator consumes nearly 70% of the source power when delivering 100 µA, which is still greater than the average requirement expected from advanced sensor electronics.

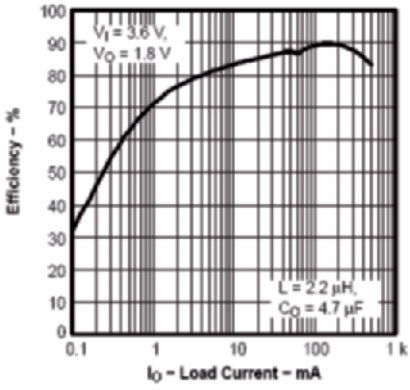


Fig. 11. Efficency vs. Load Current for typical switching regulator.

If voltage regulation requirements can be relaxed, the switched-capacitor charge pump offers improved efficiency at low loads, since the internal losses are mostly due to the internal oscillator and equivalent resistance of the flying and storage capacitors. If demands are light, the oscillator switching frequency can be low and I*R losses in the capacitors are small. One potential drawback to the prototypical switched-capacitor circuit is an unregulated output. Regulated charge pumps are available but incur a power penalty for an oscillator modulation circuit. In general, however, switched capacitor circuits tend to be more suitable for low power electronics than inductor-based solutions.

Two important points derive from the above observations: 1) any time a voltage conversion is required from source to supply rail, significant efficiency losses can be expected, and 2) circuits that can tolerate an unregulated supply reduce losses due to power management. Ideally, of course, all electronics in a sensor node would utilize the voltage directly from a battery to avoid power management losses. With typical circuit design techniques, this is not feasible, especially since the power supply on a sensor node will be powered from a volatile energy harvesting source.

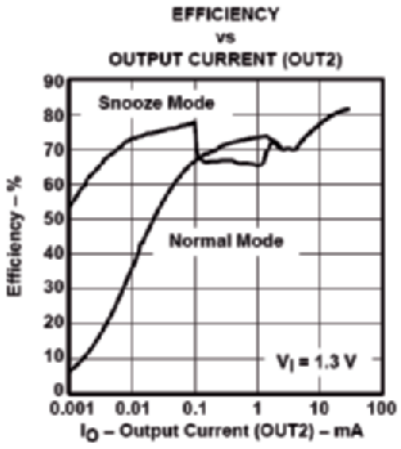


Fig. 12. Efficiency vs. Load Current for TPS60313 charge-pump

Off-the-shelf devices are typically designed to provide tens or hundreds of milliamps of current. Some, like the TI TPS60313, have operating modes where sections of the regulator are unused but the outputs are still operational at low load currents. Fig. 12 shows the efficiency graph of the TPS60313; snooze mode offers greatly improved efficiency at load currents under 100 μA.

Even so, with the best available COTS devices losses due to power management are high when operating in the average power range enabled by research outlined in this paper. In fact, in low duty-cycle circuits where rails are gated in quiescent mode, power consumed by the management circuitry approaches 100% of average.

If voltage level conversion and regulation is required, the best way to limit losses is to customize the management circuit for the application. Based on a standard switched-capacitor ladder architecture, the device pictured in Fig. 13 will consume less than 100 nanowatts while providing all required supply voltages for the sensor node. Fig. 14 is a schematic diagram of the converter block. Control logic runs on a current-starved ring oscillator

while an ADC measures the output, controlling an up/down frequency register. A current DAC controls a second ring oscillator for gate drive. The on-chip low dropout linear regulators (LDO) provide regulation, consuming on the order of 10nW of power each.

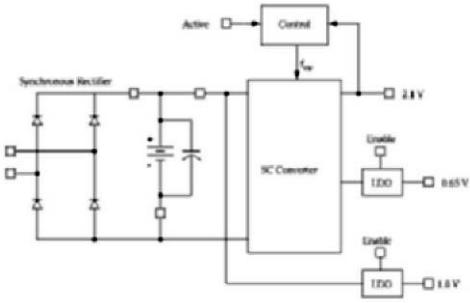


Fig. 13. Custom regulator block diagram. Boxes represent package terminals.

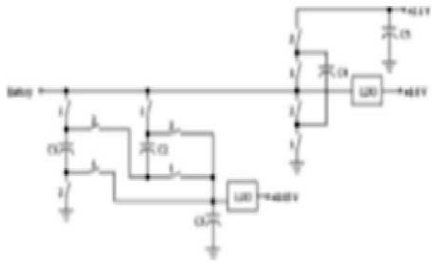


Fig. 14. Switched capacitor converter detail.

5.1. Packaging Considerations

To become truly ubiquitous, the entire sensor node platform must be extremely small. There are many parameters that affect size, among them component count, degree of integration, printed circuit board (PCB) form factor, and modularity requirements. Although high levels of integration will reduce component count, component size, and PCB size, it adversely affects *modularity*. The flexibility enabled by modularity generally means an increased need for interconnect. Interconnect by itself requires physical space but does not add capability to the sensor. In that sense it is the physical analog of power management: usually essential but mostly undesirable. Form factor is deceptively important for a sensor node, particularly if it is to be self powered, because the power scavenging technology will impose dimensional constraints. For example, solar panels are always thin and flat while an electromagnetic motion transducer has significant size in three dimensions. The solar panel solution matches nicely to a traditional, flat, rectangular, monolithic PCB. On the other hand, the electronics for a motion transducer would demand a non-conventional "3-d" assembly.

One of our low power sensor applications is a self-powered tire pressure monitoring system. The energy scavenger in this prototype is a motion transducer called a "shaker". The shaker is composed of a tube with embedded wire windings and a magnet inside the tube sliding from one end

to the other as the tire rotates. When the magnet passes the windings, an AC current is generated in the windings. A single traditional rectangular PCB does not accommodate this form factor efficiently, so the electronics for the tire pressure sensor are assembled onto a stack of PCBs roughly forming a cube.

Another advantage of this arrangement is the ability to change parts of the circuit, such as a radio, without rebuilding the powertrain or digital components of the node. As noted previously, this modularity comes at the cost of increased interconnect. The target dimension of the cube is $1cm^3$ so space for interconnect is highly constrained.

One of the smallest available multi-pin connectors is an elastomeric core material wrapped by a foil with conductors on the order of 0.05mm wide on 0.1mm centers. The connectors can be less than 1mm in height and width. Contact is made by pressure against an immersion silver or gold PCB pad, so they do not require soldering. Fig. 15 is a photograph of the top of an elastomeric connector.

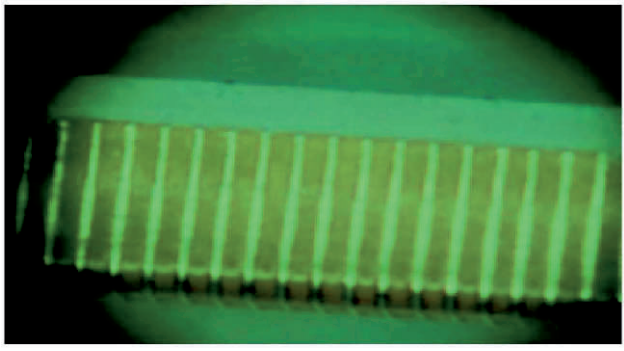


Fig. 15. Top view of a 1mm wide elastomeric connector showing the conductors.

Four PCBs comprise the electronics stack: a battery/rectifier board, a digital board, a sensor board, and a radio board. Power management is distributed. Fig. 16 is a layout diagram of the sensor board. The main component is a tire pressure sensor die set mounted via chip-on-board (COB). The pads surrounding the core are contact points for four elastomeric connectors, one on each side. Precise vertical board separation and elastomer compression is controlled by a small plastic ring shown as the inner square on the sensor board layout. This ring also acts as an inner retainer for the

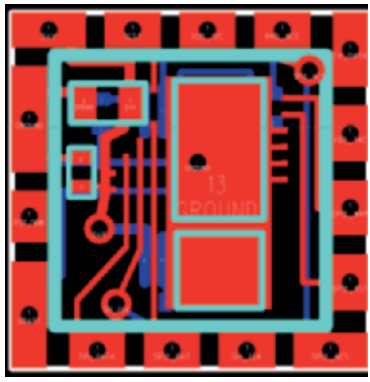


Fig. 16. The sensor board.

connectors. The entire assembly, boards, rings, and connectors, is inserted into a thin-walled plastic cube. Fig. 17 shows a physical mockup with the internal spacing rings and a test PCB. This prototype uses sample connectors (the gold color inside the cube) that are one-third taller than required by the electronic components. The dimensions of the mockup are 1cm in x and y, and the operational node will be 1cm or less in height.

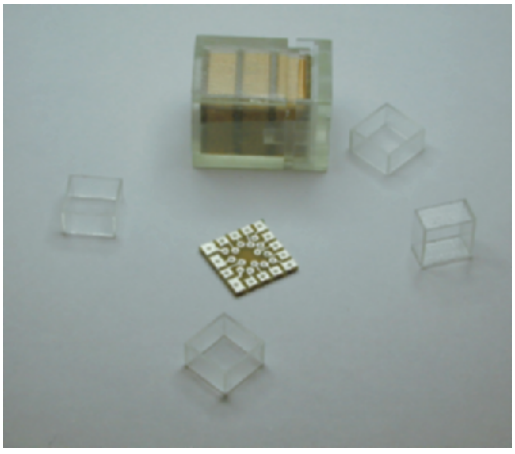


Fig. 17. Components of a physical mockup.

The prototype was used to determine reliability, contact resistance, and current capacity of the connectors. Performance in all three areas was exceptionally good: Reliable connection was made with the specified 5% connector deflection, top to bottom resistance (four PCBs and three connectors) is 0.06 ohm, and there was no noticeable degradation in connectivity when conducting 200 mA over more than 5 minutes, pad to pad. Since peak current through any given interconnect path will be less than 2 mA these connectors are more than adequate.

An operational "PicoCube" is expected in early June 2006. In the first version, a TI MSP430 microcontroller will be used for gathering sensor samples and controlling the radio for packet transmission. Both the microcontroller and sensor are in deep sleep mode most of the time; the sensor contains a timer that interrupts the microcontroller every six seconds. Power management for the "always on" sensor and μC supplies will be performed by the TI TPS60313 switched capacitor charge pump described above. Since the sensor and microcontroller are in deep sleep mode between samples, only the charge pump is fully active. Due to the TPS60313 snooze mode, average power is about 6μW, dominated by the charge pump. With the integrated regulator, average power will be reduced to the sub-microwatt range, dominated by the microcontroller. Future efforts to reduce system power may include a customized digital section, which is expected to push average power into the low hundreds of nanowatts.

6. Conclusions

This work has described three proposed techniques for implementing low power wireless links. First, the use of RF MEMS components to explore new transceiver blocks and architectures was described. A proof of concept system was presented, and future proposed work was outlined. Secondly, an architecture featuring a carrier sense receiver was presented. This would allow a large reduction in average link power consumption and latency. The carrier sense radio has a very aggressive power specification, and multiple techniques to achieve this goal were proposed. Finally, a custom, distributed power train is proposed to perform efficient supply regulation at very low current levels. This work also presented recent efforts in developing robust, miniature 3d packaging techniques that allow the modular integration of the electronics, sensor, and energy harvesting devices.

Acknowledgement

The authors would like to acknowledge the work by Mike Seeman (U.C. Berkeley Electrical Engineering), and Mike Koplow (U.C. Berkeley Mechanical Engineering).

References

[1] R. Ruby et. al., “Ultra-miniature high-Q filters and duplexers using FBAR technology”, IEEE ISSCC, San Francisco, CA, Feb 2001.

[2] B. Otis et al., “An Ultra-Low Power MEMS-Based Two-Channel Transceiver for Wireless Sensor Networks,” IEEE Symposium on VLSI Circuits, June, 2004.

[3] B. Otis, Y.H. Chee, J. Rabaey, “A 400uW Rx, 1.6mW Tx Super-regenerative transceiver for Wireless Sensor Networks”, IEEE ISSCC, San Francisco, CA, Feb 2005.

[4] B. Otis, J. Rabaey, “A 300uW, 1.9-GHz CMOS Oscillator Utilizing Micromachined Resonators,” IEEE Journal of Solid-State Circuits, vol. 38, July 2003, pp. 1271–1274.

[5] J. Crols and M. S. J. Steyaert, “Low-IF Topologies for High-Performance Analog Front Ends of Fully Integrated Receivers,” IEEE Trans. Circuits Syst. II, vol. 45, March 1998, pp. 269–282.

[6] M. Tiebout, “Low-Power Low-Phase-Noise Differentially Tuned Quadrature VCO Design in Standard CMOS,” IEEE J. Solid-State Circuits, vol. 36, July 2001, pp. 1018–1024.

[7] A. Fard, P. Andreani, "A Low-Phase-Noise Wide-Band CMOS Quadrature VCO for Multi-Standard RF Front-Ends," in Proc. 2005 RFIC Symp, February 2005, pp. 539–543.

[8] A. W. L. Ng, H. C. Loung "A 1 V 17 GHz 5 mW Quadrature VCO based on Transformer Coupling," IEEE ISSCC, San Francisco, CA, Feb 2006.

[9] W. Hsu, A. Brown, and K. Cioffi, "A Programmable MEMS FSK Transmitter," IEEE ISSCC, San Francisco, CA, Feb 2006.

[11] E. Lin, J Rabaey, A. Wolisz, "Power-effficient rendez-vous schemes for dense wireless sensor networks", 2004 IEEE International Conference on Communications, June 2004, pp. 3769–3776.

[12] N. Pletcher and J. Rabaey, "A 100uW, 1.9GHz Oscillator with Fully Digital Frequency Tuning," Proceedings of the 31st European Solid State Circuits Conference (ESSCIRC 2005), Grenoble, France, Sept. 12-16, pp. 387–390.

DESIGN OF AN ENERGY-EFFICIENT PULSED UWB RECEIVER

M. Verhelst, W. Vereecken, N. Van Helleputte,
G. Gielen, M. Steyaert and W. Dehaene

MICAS – ESAT – K.U. Leuven
Kasteelpark Arenberg 10, 3001 Heverlee, Belgium
mverhels@esat.kuleuven.be

Abstract

This paper studies the different power-performance trade-offs at architectural and block level to come to the most energy-efficient UWB system for operation in the 0-960MHz frequency band. This is achieved by designing for *the lowest energy per useful received bit.* Different receiver architectures are explored and compared against each other. After the selection of the most optimal architecture, the trade-offs inside the different analog building blocks of this receiver are studied. Our results show that the most energy-efficient solution makes use of a complex analog correlation UWB receiver, with an LNA with a $P_{in,1dBc}$ back-off of -5dB, 3-bit ADC's and a 500MHz QVCO. The requirements for the ADC offset, QVCO phase noise and mixer linearity are rather relaxed, which enables a low-power implementation.

1. Introduction

During the last decade, ultra-wideband (UWB) communication has attracted a lot of attention from the research community. One of its flavors, pulsed UWB communication ([1]), is based on the transmission of ultra-short pulses in time. As a result, the energy of these pulses is spread over a wide frequency range, with a very low power spectral density. In 2002, the Federal Communications Commission (FCC) defined UWB as any wireless transmission scheme that occupies a fractional bandwidth $\geq$ 20%, or has more than 500MHz of absolute bandwidth [2]. The fractional bandwidth is defined as W/f_c, with W the transmission bandwidth and f_c the band center frequency. Furthermore it must meet the spectrum mask specified by the FCC, which allows UWB in the 0-960MHz (for imaging, ranging) and 3.1-10.6GHz bands.

Not only does this wide bandwidth allow communication with very high data rates over a short distance. It also enables very accurate localization, robustness against fading and multipath channels and simple, low-power

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 303–329.

transmitter and receiver designs. As a result, ultra-wideband becomes also increasingly popular for power-efficient ranging, imaging and distance measurement with communication at low data rates. In this application domain, the main implementation issues are the energy consumption and robustness against noise and interference rather than achieving high data rates.

Due to the low transmit power levels, the loose linearity requirement and its duty cycled nature the transmitter will not contribute significantly to the energy consumption of an UWB transceiver. The focus of the design of an energy-efficient UWB communication system should hence be on the receiver, which has the difficult task to synchronize to the very short and broadband pulses.

In this paper, the different power-performance trade-offs at architectural and block level will be studied aiming for the most energy-efficient UWB receiver for operation in the 0-960MHz frequency band. The total receiver energy (including header reception, synchronization energy,...) to receive a single bit with a certain bit error rate is a very good metric to compare different designs against each other [3]. By designing for *the lowest energy per useful received bit*, both communication theory and implementation theory are considered and weighted carefully. Since the possible energy savings are the largest at the highest levels of abstraction, it is very important to start this exploration at architectural level. Therefore, this paper will, after a short description of the UWB system in section 2, determine the optimal receiver architecture in section 3. In section 4 the different building blocks of the analog front-end are studied to find the optimal joint power-performance trade-off for each of them. During this optimization, the implications of every block level decision on the total system's performance are taken into account. This system-block level co-design results in an energy-efficient UWB receiver design for low-power, low data rate applications. Finally section 5 studies the migration from the 0-960MHz band to the more common 3-5GHz band. The implications on the performance and power consumption of the receiver alternatives are examined to investigate whether the same conclusions still hold. Section 6 concludes the paper.

2. UWB System

In UWB communications, data is modulated on a stream of short-duration pulses. Figure 1 (solid line) shows such a 0-960MHz pulse in the time and frequency domain. Every bit is replaced by a set of N_s pulses, with N_s the length of the spreading code used. This spreading code is employed to increase the processing gain of the receiver, enable multi-user access and reduce the spectral spikes of the transmitted waveform. Assuming BPSK

data modulation, the polarity of every pulse is determined by the transmitted data bit and the corresponding bit of the used spreading code. As a result, the transmitted waveform can be described as:

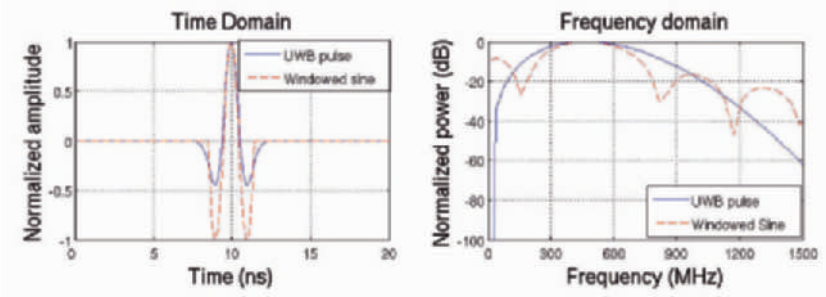


Fig. 1. Time and frequency domain plots of an UWB pulse and a windowed sine.

$$s_{tx}(t) = \sum_{n=-\infty}^{\infty} b_n \sum_{k=0}^{N_s-1} a_k \sqrt{E_p}\, w_{tx}(t - kT_p - nT_s) \tag{1}$$

with pulse repetition period T_p, symbol period $T_s = N_s T_p$, total energy per pulse E_p. $b_n \in \{-1, 1\}$ and $a_k \in \{-1, 1\}$ are respectively the n^{th} transmitted information bit and the k^{th} bit of the used spreading code of length N_s. $w_{tx}(t)$ is the transmitted pulse form, with unit energy.

The received signal at the receiver antenna can be denoted as:

$$s_{rx}(t) = \sum_{n=-\infty}^{\infty} b_n \sum_{k=0}^{N_s-1} a_k \sqrt{E_p}\, w_{rx}(t - kT_p - nT_s) + n(t) + i(t) \tag{2}$$

with n(t) additive white Gaussian noise (AWGN) with zero mean and power spectral density $N_0/2$ over the bandwidth W of the receiver and i(t) interference caused by other users of the spectrum. $w_{rx}(t)$ is the received pulse waveform, distorted by the multipath channel and antenna.

To detect the transmitted data bits b_n in the receiver, the incoming signal $s_{rx}(t)$ has to be decorrelated with both the received pulse form $w_{rx}(t)$ and the used spreading code. This is depicted in figure 2, which shows a general UWB receiver. After amplification, the incoming pulses are correlated with a locally stored pulse template $w_{templ}(t)$. This consists of a multiplication, followed by an integration (or accumulation). The result of this correlation is multiplied with the spreading code in a second correlator, which outputs the decorrelated data bits.

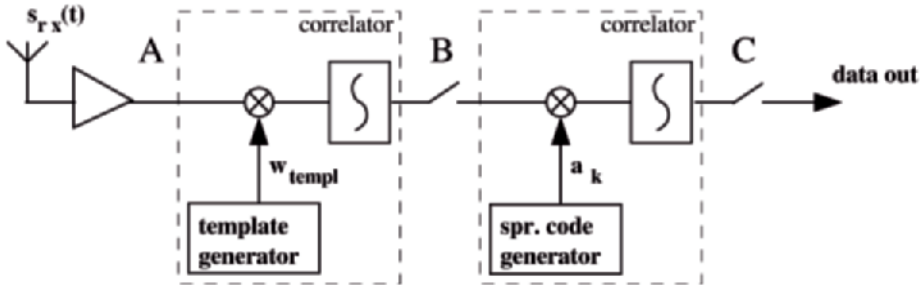


Fig. 2. General system view of a pulsed UWB receiver.

3. Architectural Level Exploration

In this stage of the design process, different system architectures will be compared against each other. Since it is too time consuming to come up with a detailed implementation of all architectures under study, it is not possible to do a perfect power estimation of every proposal, nor to take all underlying circuit non-idealities into account. Using as good as possible power approximations and performance simulations, the decision can be made which architectures should be studied into more detail. As the receiver is dominant in the power consumption of the UWB system, only the non-idealities that have a large effect on the receiver power consumption (acquisition power) or its performance (channel transfer function, interference, imperfect pulse matching) will be taken into account at this stage of the design process.

3.1. Position of the ADC

As can be seen in figure 2, the ADC can be placed at position A, B or C. This section will analyze and compare the different alternatives.

Placing the ADC at position A results in a fully digital architecture [4]. This has the advantage of being very flexible, scalable and easy to design. Moreover, a very fast acquisition is possible, due to the ability to do many parallel correlations in the digital domain. In this architecture, the ADC however has to work at Nyquist rate, i.e. at approximately 2GHz. This high sampling speed, together with a high resolution will cause the ADC to consume most of the power. [5] proposes to reduce the resolution to 1 bit to save power. Both alternatives (called ‘Inf-bit FD’, resp. ‘1-bit FD’) will be explored in this section.

The next alternative (analog correlation receiver, called 'AC', [6]) places the boundary between the analog and digital domain after the pulse matched filter at point B. Only the correlation with the spreading code is executed in the digital domain. This seriously relaxes the ADC sampling speed, but increases the complexity of the other analog blocks. It also becomes harder to do Rake reception, since the pulse template will now have to be constructed in the analog domain. Possible implementations for this template generator will be discussed in section 3.2.

Finally, placing the ADC at position C (called fully analog receiver, 'FA') reduces the ADC sampling speed even more. The difficulty of generating the pulse template in the analog domain remains, and the acquisition speed will also be a problem, since it is impossible to do many correlations in parallel in the analog domain.

The performance and power consumption of the alternatives discussed above will now be investigated:

A) PERFORMANCE

Assuming Gaussian noise, the decision variable at the output of the receiver will be Gaussian with conditional mean:

$$E[r \mid b_n] = \sum_{k=0}^{N_s-1} \int_0^{T_p} b_n \sqrt{E_p}\, w_{rx}(t) w_{templ}(t)\ \delta t = b_n N_s \sqrt{E_p \eta_{cap}} \tag{3}$$

η_{cap} is the energy capture efficiency of the receiver, which is a measure for how well the matching template correlates to the incoming (distorted) signal:

$$\eta_{cap} = \frac{\int_{-\infty}^{\infty} w_{rx}(t) w_{templ}(t)\ \delta t}{\int_{-\infty}^{\infty} w_{rx}^2(t)\ \delta t \int_{-\infty}^{\infty} w_{templ}^2(t)\ \delta t} \tag{4}$$

Assuming a noiseless template, the variance of the decision vector will only be caused by the noise present at the input of the receiver:

$$\sigma^2 = \sum_{k=0}^{N_s-1} E[\int_0^{T_p} n(t) w_{templ}(t)\ \delta t \int_0^{T_p} n(u) w_{templ}(u)\ \delta u] = N_s \frac{N_0}{2} \tag{5}$$

As a result the bit error probability (Pe) of the receiver can be written as:

$$Pe = Q\left(\sqrt{\frac{2N_s E_p \eta_{cap}}{N_0}}\right), \quad Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(\frac{-t^2}{2})\ \delta t, \quad for\ x > 0 \qquad (6)$$

This formula gives the theoretically ideal performance of the UWB receiver. This performance will degrade due to the finite-resolution ADC's and due to an imperfect matching of the template with the incoming signal (η_{cap}<1). Based on a Matlab model, these effects have been simulated for the proposed UWB receivers, operating in various environments: perfect (AWGN) channels, line of sight (LOS) channels, non-line of sight (NLOS) channels and interference dominated (with interference data from real measurements) NLOS channels. The LOS and NLOS simulations are averaged over 100 802.15.4a channels [17]. The high-resolution fully digital (Inf-bit FD), low-resolution fully digital (1-bit FD), analog correlating (AC) and fully analog (FA) receiver are simulated. In the fully digital receivers, the pulse template is constructed by averaging the incoming signal over a training sequence of 50 bits (=N_A), at the beginning of every data packet of 500 bits (=N_D). The acquisition of the analog correlating receiver will take longer (100 bits), because less can be parallelized. The acquisition method used here, is explained in [15]. The length of the acquisition in the fully analog receiver finally goes up with an additional factor N_s, due to the inability to correlate in parallel with different shifted versions of the spreading code. Since an infinite-bit ADC is assumed for the AC and FA receiver, their performance curves fall on top of each other. The simulations are done for N_s = 10, Tp = 40ns. Figure 3 shows the results.

As can be seen from the plots, the analog correlating receiver performs particularly bad in NLOS channels. This is due to the fact that this receiver is unable to do perfect channel compensation. The 1-bit fully digital receiver performance degrades significantly in the presence of interference. The average performance (=required input SNR for a Pe = 1e-3) of every receiver over the 3 realistic environments (LOS, NLOS and interference dominated NLOS) will be used to compute the consumed energy per useful received bit. This is possible since (under AWGN) the difference in performance of all receivers, translates to a difference in required spreading code length to obtain a predefined Pe and hence results in a different power consumption.

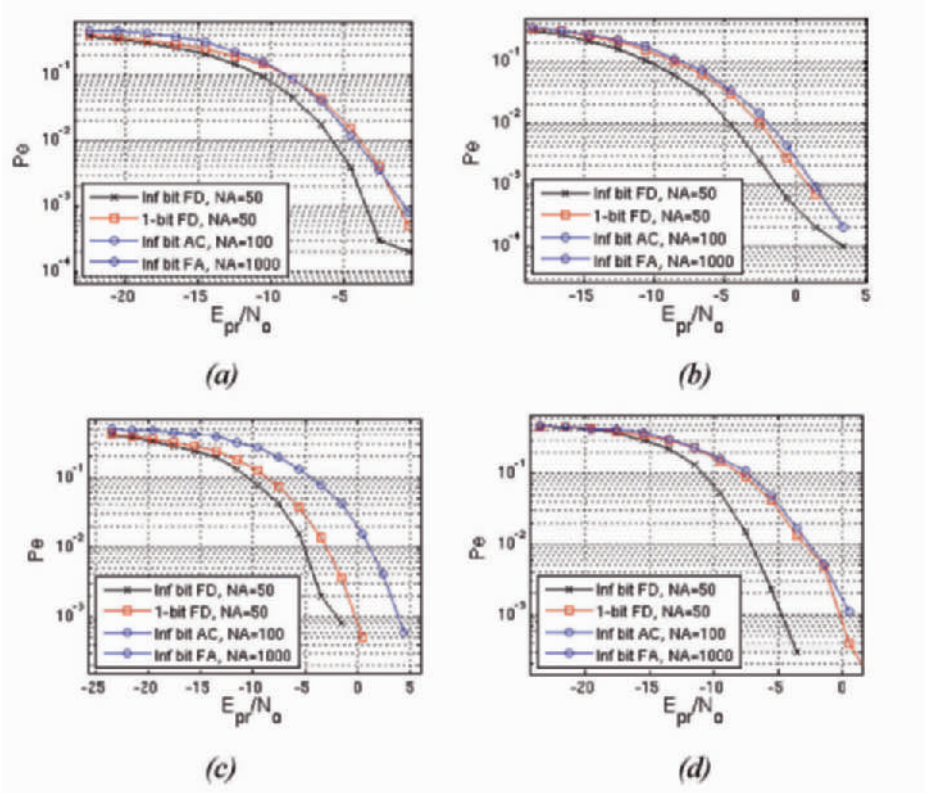


Fig. 3. Bit error probability (Pe) of the FD, AC and FA receiver in a gaussian channel (a), LOS channel (b), NLOS channel (c), interference-dominated NLOS channel (d). N_s=10, T_p = 40ns, E_{pr} = pulse energy at the receiver entrance.

B) POWER

The power consumption of every alternative has to be estimated. Various figures of merit (FOM) and circuits described in literature will be used to do this first estimation. The infinite-bit fully digital (Inf-bit FD) alternative will be replaced by a more realistic 4-bit version.

All architectures can use the same LNA. An LNA for UWB communication in the 0-960MHz band is described in [7] and consumes 0.72mW. The fully digital solution does not have other analog blocks operating on the incoming signal. The AC and the FA alternatives need an analog template generator and an integrating mixer. The mixer has to be wideband (500MHz) and its power consumption can hence be extracted from [8]. The template generation can be implemented in many ways. Section 3.2 will explore this block into more detail. Finally note that two mixers and ADC's are used in these CA and FA alternative, since two orthogonal analog paths are needed to take care of synchronization and clock offset tracking.

The power consumption of the ADC's can be estimated based on a commonly used FOM [9]:

$$FOM_{ADC} = \frac{2^N * f_{sampl}}{P_{ADC}} \tag{7}$$

with N the number of effective bits of the ADC, f_{sampl} the sample frequency and P_{ADC} the ADC power consumption. In a survey of over 100 ADC's published in literature between 1978 and 1999, the FOMs range from 1e10 to 1.2e12 [9]. In our computations a FOM of 5e11 is used. The VCO and PLL of the fully digital alternatives have to work at 2GS/s and with low phase noise [10]. Most of the hybrid and fully analog solutions on the other hand, only need a less accurate 500MHz signal, which allows them to position the pulse template with a resolution of 2nsec [11]. The power consumption of the digital part finally depends on the clock frequency of the digital part and on the effective capacity switched every cycle (physical capacity*switching activity). The latter will be different during acquisition (A) and during data reception (D). During acquisition a lot of correlations will have to be done in parallel, to come to packet synchronization and find the start of the spreading code. The digital part of the AC and FA alternatives consume much less, since they only have to work at pulse rate (speed = /80) resp. bit rate (speed = /800).

Table 1. Computation effective energy/bit for the different alternatives.

	4-bit FD	1-bit FD	AC	FA	Remarks
LNA	0.72mW	0.72mW	0.72mW	0.72mW	[7]
an. templ. generator	/	/	≥ 0.3mW	≥ 0.3mW	see section 3.2
mixer	/	/	2*3.7mW	2*3.7mW	[8]
ADC	64mW (2GSs, 4b)	4mW (2Gss, 1b)	2*0.8mW (25MHz, 4b)	2*0.08mW (2.5MHz, 4b)	[9]
VCO+PLL	7.6mW (2Gss)	7.6mW (2GHz)	0.3mW (500MHz)	0.3mW (500MHz)	[10], [11]
digital	(A) 75mW (D) 35mW	(A) 30mW (D) 14mW	(A) 5mW (D) 2.5mW	(A) 1mW (D) 1mW	[4]
Total power	(A)147mW (D)108 mW	(A)42.3mW (D)26.3mW	(A) ≥15.4mW (D) ≥12.9mW	(A) ≥10.0mW (D) ≥10.0mW	
L_{perf}	0dB	3.0dB	5.0dB	5.0dB	(Fig. 3)
N_A/N_D	50/500	50/500	100/500	1000/500	
Energy/bit @ Pe=1e-3	49.1nJ/bit	24.4nJ/bit	≥20.1nJ/bit	≥37.9nJ/bit	

After adding the contributions of the different blocks to the power consumption, a total power consumption of every alternative during acquisition (P_A) and data reception (P_D) can be computed (see table 1). To compute the consumed energy per effectively received bit, also the performance and the length of the preamble have to be taken into account. The performance loss (L_{perf}) can be derived from figure 3 and is normalized to 0dB for the Inf-bit FD solution. The average consumed energy per effectively received bit (under AWGN conditions) can be computed as [3]:

$$energy/bit = \frac{P_A * N_A * T_p + P_D * N_D * T_p}{N_D} * L_{perf} \quad (8)$$

From this computation it is clear that it isn't smart to go after the least power consuming alternative, nor the alternative with the best performance. The overall optimal architecture is the architecture with the best trade-off between these two parameters, or in this case alternative AC. When it is possible to use a reconfigurable topology, alternative AC is suited best during acquisition and alternative FA for data reception.

3.2. Front-End Architecture

Correlating with the pulse template in the analog domain will however only be optimal if the pulse template can be generated efficiently in the analog domain. This section takes the exploration for the most energy-efficient architecture one step further by exploring various alternatives for the analog front-end. They all differ in the way they generate the pulse template for the first correlation.

The first alternative (figure 4(a), called 'DAC') makes use of a perfect pulse template. Since it is impossible to generate the perfect pulse template in the analog domain, it will have to be stored in the digital domain and converted to the analog domain with a fast DAC. Three similar analog branches have to be used here in parallel to do early-late tracking. One very fast (4-bit) ADC is still needed during acquisition to learn the pulse template. Alternative b (called CAC) is described in [14]. It will not try to reconstruct the ideal pulse template. Instead it will use a windowed sine as pulse template. This results in a performance loss, but reduces the power consumption drastically. To speed up acquisition and enable clock offset tracking, the incoming signal will at the same time be mixed with the orthogonal signal: a windowed cosine. The last alternative (c, called 'TR') will not create a local version of the pulse template. Instead every pulse is replaced by two pulses: a reference pulse and an information bearing pulse. This transmitted reference ([12]) system will delay the reference pulse and use it as a template for the second, information bearing pulse. This alternative has the advantage of being very simple and straightforward.

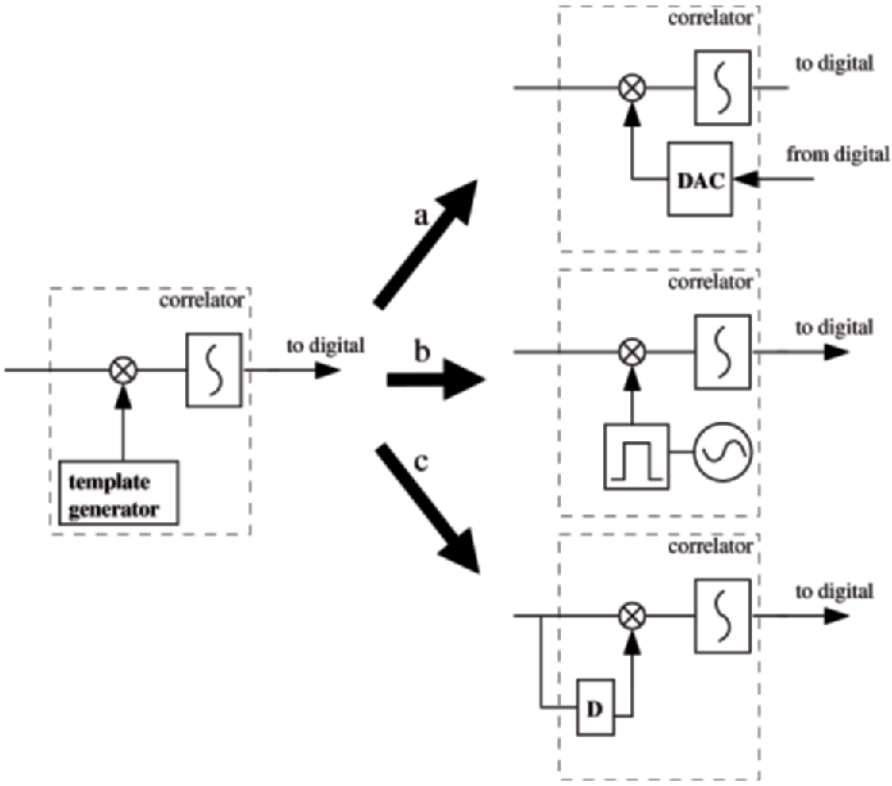


Fig. 4. The three alternatives for the pulse template generation.

Transmitted reference receivers need hardly any synchronization and automatically capture all multipath components of the received signal. On the other hand, they suffer a large performance degradation due to the very noisy template [12]. The performance and power consumption of every alternative will be investigated below.

A) PERFORMANCE

The performance of alternative DAC and CAC can also be computed with formula (6). η_{cap} will be 1 for ideal template matching (alternative DAC). Simulations show an η_{cap} of approximately 0.85 for alternative CAC under AWGN. The performance of alternative TR can however not be computed with this formula, since the variance of the decision vector (formula (5)) will contain an additional noise term due to the noisy template. The performance of this alternative can be expressed as [13]:

$$Pe = Q\left(\left[\frac{2}{N_s}\left(\frac{N_0}{\eta_{cap}E_p}\right) + \frac{WT_{corr}}{N_s}\left(\frac{N_0}{\eta_{cap}E_p}\right)^2\right]^{-1/2}\right) \tag{9}$$

with T_{corr} the duration of the correlation operation. A large T_{corr} will result in a large η_{cap}, but also in a large noise-cross-noise term and vice versa. This effect can also be seen in the simulation results of figure 5. As can be seen from formula (9) the difference between the performance of the transmitted reference system and the other alternatives is proportional to N_s.

B) POWER

The power consumption of every alternative can be computed in the same way as in table 1. The first alternative (DAC) learns the pulse template in the digital domain during the acquisition phase. Its power consumption is hence equal to alternative 'Inf-bit FD' in section 3.1. During data reception, the fast ADC is disabled and a fast (2Gsps) 5-bit DAC is used to generate the pulse

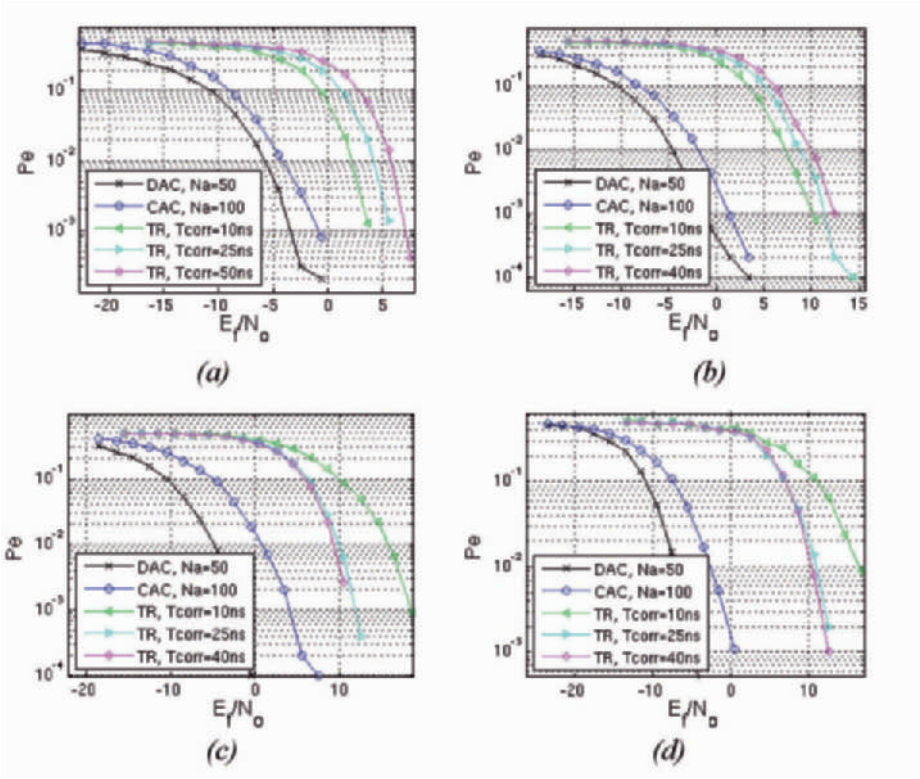


Fig. 5. Performance of the DAC, CA and TR receiver in a gaussian channel (a), LOS channel (b), NLOS channel (c), interference-dominated NLOS channel (d). N_s=10, T_p = 40ns, $E_f = 2*E_{pr}$ (TR); $E_f = E_p$ (others).

template in the analog domain. Three analog paths (mixer+DAC) are needed in the DAC receiver to enable early-late detection. A significant part of the power consumption also goes to the fast shift registers and register file necessary to steer these DAC's. The second alternative is the complex analog correlating receiver. The template generation hardly consumes additional

power here, since only a sine and cosine wave have to be generated. This is easily done by replacing the (already accounted for) VCO by a QVCO. The power consumption of the QVCO is assumed to be twice the one of the corresponding VCO, which is a slight overestimation. Remark that the power consumption of the ADC and mixers are doubled in this architecture due to the two orthogonal paths in the front-end. Finally, in the transmitted reference system (TR) the template generator only consists of a delay line. Up to the authors' knowledge, the only published analog delay line for UWB communications can be found in [16]. Although this design is made for UWB communications in the 3-5GHz band, the results can be migrated to the 0-960MHz band. This yields a power consumption of 13.1mW for a maximal delay of 2.5nsec. This small delay is however not enough for UWB communication in the 0-960MHz band, but can be increased at the cost of additional power consumption.

In the same way as in table 1, the total power consumption, average performance loss (based on figure 5) and the average consumed energy per effectively received bit are computed (see table 2). The results show that the TR system is not favorable due to its bad performance. This effect reduces/increases for smaller/larger N_s. Both the DAC and the CAC receiver show the best performance and have a roughly equal estimated energy per useful received bit. The DAC receiver however needs a high speed ADC, together with 3 parallel analog paths, each containing a mixer, an integrator and a 5-bit 2Gsps DAC. This of course would result in a huge chip area, which makes the CAC receiver a much more elegant solution.

Table 2. Computation effective energy/bit for alternatives a, b, c.

	DAC (a)	CAC (b)	TR (c)	Remarks
LNA	0.72mW	0.72mW	0.72mW	[7]
mixer(s)	(A)0mW (D)3*3.7mW	2*3.7mW	3.7mW	[8]
template generation	(A)0mW (D)2+3*3mW	see (Q)VCO	>13.1mW [16]	
ADC	(A)64mW (D)3*0.8mW	2*0.8mW (25MHz, 4b)	0.8mW (25MHz, 4b)	[9]
VCO+PLL	7.6mW (2Gss)	2*0.3mW (500MHz)	0.3mW (500MHz)	[10], [11]
digital (w.o. shiftreg)	(A)75mW (D)2.5mW	(A)5mW (D)2.5mW	2.5mW	[4]
Total power	(A)147.3mW (D)35.3mW	(A)15.4mW (D)12.9mW	(A)0mW (D)>21.2mW	
L_{perf}	0dB	5.0dB	15.2dB	(fig. 5)
N_A/N_D	50/500	100/500	0/500	
Energy/bit @ Pe=1e-3	20.0nJ/bit	20.1nJ/bit	>281.2nJ/bit	

3.3. Conclusion

This section explored different architectures for a pulsed UWB receiver based on the energy consumption per effectively received data bit [3]. It became clear that it isn't smart to go after the least power consuming alternative, nor the alternative with the best performance. The optimal architecture is the architecture with the best trade-off between these two parameters. The next section will further explore the most promising architecture: the complex analog correlating receiver, described in [14]. The block diagram of this receiver is plotted in figure 6.

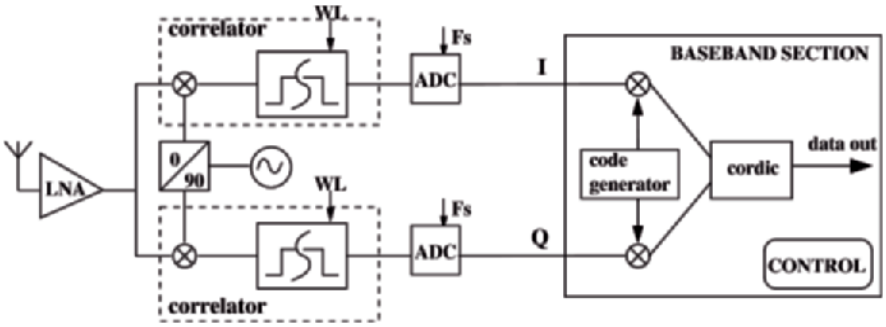


Fig. 6. Block diagram of the complex analog correlating receiver.

4. Block Level Exploration

This section will explore the different building blocks of the complex analog correlating receiver of figure 6. For every block (LNA, correlator, QVCO and ADC), the power-performance trade-off is studied at block level to find the energy-optimal values for the degrees of freedom of every block.

4.1. LNA

Due to its wide bandwidth, in-band interferers are one of the largest problems of UWB communications. This has a severe impact on the linearity requirement and power consumption of the LNA, which are studied in this paragraph.

A) SIGNAL/INTERFERENCE STRENGTH

Our LNA exploration starts with the computation of the expected signal and interference power at the input of the receiver. The FCC imposes for the pulse spectrum a -49.2 dBm/MHz EIRP limit (216-960MHz) [2,18]. Assuming a fictive isotropic reference antenna at the lower end of the

spectrum (240 MHz) and applying the Friis equation results in a maximum available received power P_R at the terminals of the receiving antenna of:

$$P_R = P_T G_T G_R \frac{\lambda^2}{(4\pi d)^2} = -22.4dBm - 20 - 20\log_{10}(d), \qquad (10)$$

where P_T is the average transmitted power, G_T and G_R are the transmitting resp. receiving antenna gain which is chosen to be 0dBi at the lowest notch in the pulse spectrum, d is the distance between transmitter and receiver and λ the wavelength at the lower end of the spectrum. As a result, the available pulse power at the input of the receiver will vary from -43dBm to -63dBm at 1 resp. 10 meter link distance in line-of-sight (LOS) conditions. From the average received power, the peak pulse power values can now be determined:

$$P_{R,peak} = \frac{P_R * T_p}{PW} \qquad (11)$$

For a pulse period T_p of 40ns and a pulse width PW of 4ns, the peak power ranges from -33dBm to -53dBm in the previously stated conditions. As a comparison: the integrated thermal noise power in the band of interest (240 – 720MHz) at 298K is less than -87dBm.

Our simulations are based on a power class 5 GSM interferer under worst-case conditions. The maximum output power averaged over a transmitting timeslot is 29dBm [24]. The interferer power available at the UWB antenna will then vary from 9dBm to -11dBm at 1 resp. 10 meter distance in LOS conditions. This leads to a signal to interferer power ratio (SIR) of -22dB to -62dB at the input of the LNA. In contrast to carrier-based applications, distortion emerges at the lower end of the SIR range. The BER performance of this case study cannot be predicted using the well-known Gaussian distribution based CDF. The modulation technique of the GSM interferer is Gaussian filtered minimum-shift keying (GMSK), which is a constant-envelope modulation scheme. The bandwidth of this interferer is small enough to consider it as a fixed sinusoidal carrier, which has a sinusoidal distribution [19]. In paragraph C the performance degradation due to this interferer will be simulated.

B) LNA PERFORMANCE: THEORY

Although interferers are suppressed by the transfer function of the correlator, this does not alter the LNA's linearity requirements, since it is still exposed to the full incoming RF power, filtered only by the antenna characteristics. A 50Ω resistive terminated differential LNA is employed for this case study. The available interferer power at the output of the antenna terminals spans from -11dBm to 9dBm, which leads to a peak voltage amplitude of -21dBV to -1dBV between the gates of the input MOS transistors.

The input-referred 1dB compression voltage ($V_{in,1dBc}$) is a useful parameter in the UWB situation. The simplified transfer characteristic of a MOST differential pair is given by [20]:

$$\frac{I_{od,peak}}{I_B} = \begin{cases} \dfrac{V_{id,peak}}{V_{GS}-V_T} - \dfrac{1}{8}\left(\dfrac{V_{id,peak}}{V_{GS}-V_T}\right)^3 - \cdots & \text{for } \dfrac{V_{id,peak}}{V_{GS}-V_T} < 1.63, \\ 1.09 & \text{else} \end{cases} \tag{12}$$

with V_{id} the differential input voltage, i_{od} the differential output current and $I_B/2$ the biasing current of each transistor. The 1dB compression voltage of this amplifier is [20]:

$$V_{in,1dBc} = 1.141 \cdot (V_{GS} - V_T). \tag{13}$$

Input signals exceeding this point will cause additional signal distortion, which results in bit error rate performance degradation.

C) LNA PERFORMANCE: SIMULATION

Figure 7 shows simulation results of the receiver performance (bit error rate) as a function of the back-off of the LNA (the ratio of the input amplitude of the LNA to $V_{in,1dBc}$). The product of the signal to interference ratio (SIR) and N_s is kept constant during the simulations, equal to 2.55. This results in a bit error rate of 1% in a distortionless front end. As expected, no noticeable performance degradation will occur as long as the back-off factor is small. Figure 7 shows that starting from -5dB back-off, the bit error rate (P_e) will increase about 10dB/decade, which is proportional to the interferer power. Finally, notice that once full clipping occurs, the performance degradation becomes again independent of the back-off factor. It follows that a minimum data throughput can be guaranteed, even for extreme operating conditions.

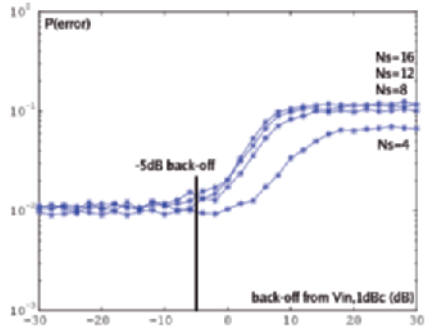


Fig. 7. Bit error rate versus back-off from Vin,1dBc, while SIR·N_s is 2.55.

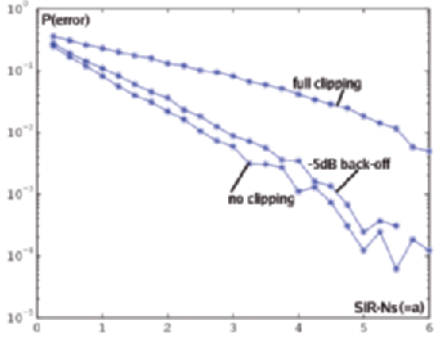


Fig. 8. Bit error rate versus SIR·PBB (only valid for N_s>4 and SIR<-3dB).

Figure 8 shows the same receiver performance as a function of $SIR \cdot N_s$ for a sinusoidal interferer. Three operation regions are shown: distortionless (no clipping, $V_{in,1dBc}$ large), full clipping of the pre-amplifier ($V_{in,1dBc}$ small) and operation at -5dB back-off. By increasing the back-off factor, the performance curve will shift more and more towards the curve for full clipping.

D) TRADING POWER FOR PERFORMANCE

The simulations show that a back-off factor of less than -5dB is needed to avoid a performance reduction due to the LNA. For a realistic peak voltage amplitude of -10dBV, this results in a required compression point of -5dBV ($V_{in,1dBc}$ = 0.56V) at the input of the differential LNA. In case of the simplified MOST amplifier (equation (13)), this compression voltage of -5dBV will require a minimum overdrive voltage of (V_{GS}-V_T) = 0.49V. Since this number is difficult to achieve in today's low-voltage technologies and will result in an increased power consumption, it should be explored whether it isn't better to allow some distortion in this case:

If the back-off factor can be decreased, the power efficiency of the LNA can be increased. Distortion however will now decrease the performance. To compensate for this loss, the number of pulses per bit will have to go up, which will again increase the power consumption per received bit.

In Figure 7, it can be seen that reducing the back-off by 3dB (doubling the power efficiency) will increase the P_e by a factor 2. From equations (12) and (13), one can derive that the current efficiency I_{od}/I_B is inversely proportional to the input-referred 1dB compression voltage $V_{in,1dBc}$. As a result, the power consumption is proportional to the power back-off. Since the contribution of the LNA to the total power consumption was estimated around 10%, increasing P_e by a factor of 2 reduces the overall power consumption by 5%.

The performance loss due to the reduction in LNA power has to be compensated by an increase of the number of pulses per bit (N_s). In the vicinity of the 1dB compression point and for a P_e=1e-2, a decrease of the performance by a factor of 2 can be compensated by increasing N_s by 18% (see figure 8). This however also increases the energy consumption per bit by 18%, which is more than the 5% we initially gained by reducing the back-off factor. It can hence be concluded that in the most energy-efficient solution no clipping is allowed and the LNA needs a back-off around -5dB.

4.2. Correlator

A proper choice of the mixer architecture can solve a lot of distortion problems. The most straightforward circuit implementation of the correlator is the four-quadrant mixer connected to a capacitor which performs the integrating part of the correlation (figure 9). The differential RF input signal

is injected in parallel with the mixer lines. The cascode transistors offer good impedance matching at the output of the LNA.

The parallel injection makes the mixer less sensitive to distortion since the complete correlation operation is current driven. Ideally the mixing transistors act as low-impedance switches and the RF input current is injected in the sampling capacitor which converts the injected pulse energy into the demodulated output voltage. As long as the bias current of the circuit is larger than the injected current, the effect of distortion is negligible.

The following measurement results were obtained from a UWB test radio front end (0.18μm CMOS) for the 3.1-10GHz band, but can also applied to the lower UWB frequency spectrum. The bias current through each branch of the mixer is 327μA. The mixer frequency of the receiver is 6GHz, while an interferer signal at 6.0008GHz is supplied at the mixer inputs. At the output of the receiver, the interferer power in the spectrum of the demodulated samples (93.75Mpulses/sec) was determined.

Figure 10 shows the relative fundamental and IM_3 components versus the absolute input power as they were read from the spectrum analyzer. Distortion levels remain very low until the input power exceeds 0dBm. The main reason is that the pulse charge-to-voltage conversion by the sampling capacitor is a linear operation. The voltage followers buffering the sampling capacitor will finally start to clip due to supply voltage constraints. Mixer power can thus be neglected compared to the LNA power consumption.

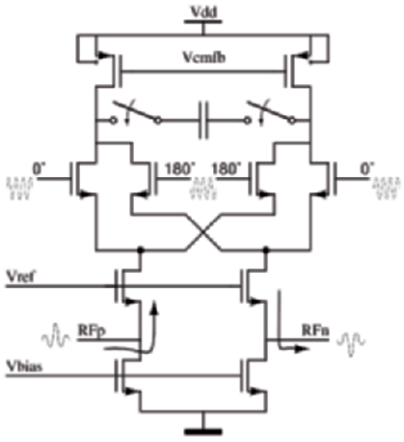


Fig. 9. Four-quadrant mixer with cascode inputs.

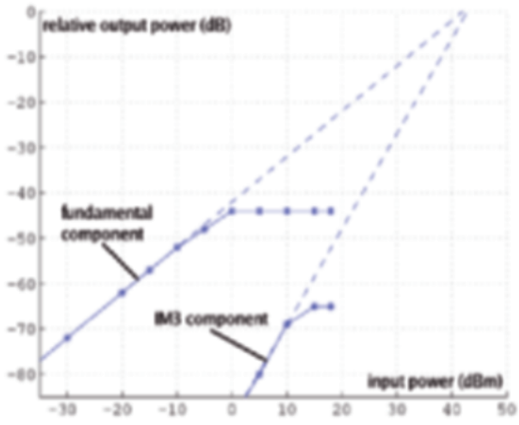


Fig. 10. Fundamental and IM3 components versus input power.

4.3. Template Waveform & QVCO

After the LNA, the received signal is correlated with windowed sine and cosine waves. The window is implemented using the switches seen in figure 9. When the switches are open, the window is closed. During the acquisition phase, the digital back-end will search for the optimal window position, which is when the signal peak falls in the middle of the window. When the window gets misplaced, less signal energy will be captured but the noise will remain the same, so the received SNR will degrade. Figure 11 shows the SNR-degradation for several window lengths as a function of the window misplacement. It is clear that a longer window is less sensitive to severe window misplacement, but on the other hand, also introduces more noise. The figure shows that when no misplacement is present, the longest window (6ns) performs worse than the smallest window (3ns). If the window placement is controlled with a clock frequency f_c, the maximum timing error is half the clock period. Figure 12 shows the control clock frequency versus the SNR degradation. Because the dynamic power consumption scales linearly with f_c, one can conclude that it is not power-efficient to use a clock frequency higher than 400MHz.

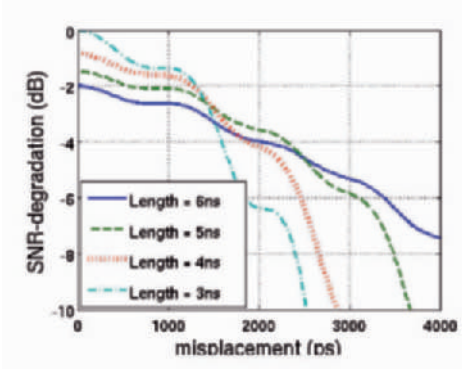


Fig. 11. SNR degradation due to window misplacement.

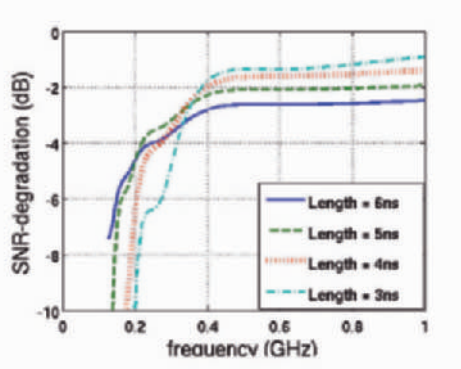


Fig. 12. SNR-degradation versus control clock frequency.

Assuming BPSK, the received (I,Q) data points at the input of the digital back-end (see figure 6) will be located along the I-axis of the constellation diagram, 180° apart from each other. The phase $\theta = arctan(Q/I)$ will be used to determine the received data point and yields information regarding clock offset and synchronization. Due to QVCO non-idealities, the points will have the tendency to shift from the ideal position. In the next paragraph the influence of typical non-idealities of the QVCO on the receiver will be investigated.

A) QUADRATURE OUTPUT IMBALANCE

The first non-ideality studied is the phase error ϕ between the two QVCO outputs (see figure 6). Let $s_{rx}(t) = s_{tx}(t) \otimes h(t) + n(t)$ denote the received signal. $s_{tx}(t)$ represents the transmitted signal, $h(t)$ is the channel response and $n(t)$ is a noise term. The I/Q outputs (before the ADC) and the corresponding phase θ are then given by:

$$\begin{aligned} I &= \int_{WL} \cos(\omega t + \phi).s_{rx}(t).dt \\ &= \cos(\phi).\int_{WL} \cos(\omega t).s_{rx}(t).dt - \\ &\quad \sin(\phi).\int_{WL} \sin(\omega t).s_{rx}(t).dt \\ &= \cos(\phi).I_{Ideal} - \sin(\phi).Q_{Ideal} \end{aligned} \qquad \begin{aligned} Q &= \int_{WL} \sin(\omega t).s_{rx}(t).dt \\ &= Q_{Ideal} \\ \theta &= \arctan\left(\frac{Q}{I}\right) \\ &= \arctan\left(\frac{\tan(\theta_{Ideal})}{\cos(\phi) - \sin(\phi).\tan(\theta_{Ideal})}\right) \end{aligned} \tag{14}$$

where I_{Ideal} and Q_{Ideal} are the I/Q-outputs that would have been received with an ideal quadrature VCO ($\phi = 0$) and θ_{Ideal} is the corresponding phase. Any phase error ϕ will introduce an error in the received phase θ. A good measure for the efficiency of a modulation scheme is the Euclidean distance between the received points in the constellation diagram. Due to the erroneous received phase, the Euclidean distance can degrade. Phase errors of QVCO's are typically less then a few degrees. A value of 10 degrees will yield a maximum loss of Euclidean distance of less than 9% which, related to SNR, equals a performance loss of 0.4dB. So for small values of ϕ and keeping in mind the BPSK modulation, the error on the received phase is negligible.

Due to clock offset, the received points in the constellation diagram will have the tendency to slowly rotate on a circle with a constant speed of rotation. The digital back-end can estimate this rotation based on previously received signals and compensate for it by inversely rotating the next received point with the proper amount. The result is that all received points are rotated back to the I-axis. Due to the phase error ϕ however, the points will rotate along an ellipse in stead of along a circle and the speed of rotation is no longer constant. No perfect correction is possible and the effect of this is that less energy is present in the corrected I-output as can be seen in figure 13. Figure 14 shows the worst-case SNR-degradation due to this effect. It is noteworthy that even for quite large values of ϕ only a very small SNR-degradation is noticeable when BPSK is used.

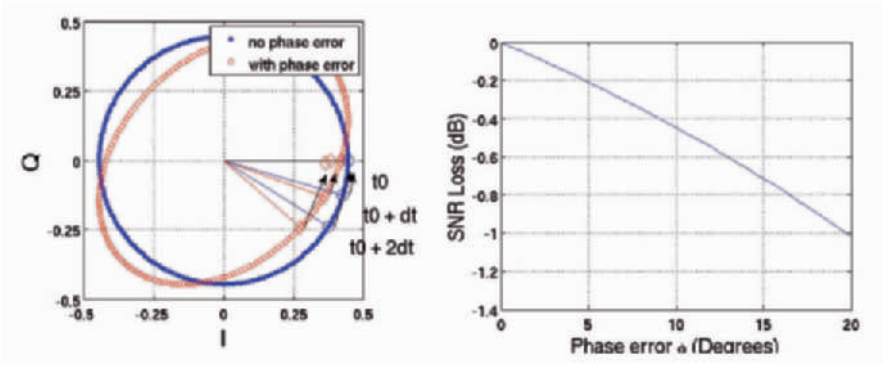


Fig. 13. Principle of correction for clock-offset and effect of phase error.

Fig. 14. SNR Degradation due to ϕ.

In order to verify the above statements, simulations were carried out with Simulink. Channel effects were included making use of the IEEE 802.15.4a channels specific for baseband UWB. Figure 15 shows BER plots for different values of phase error ϕ, assuming a reasonable 20ppm clock offset. One can clearly see only a fairly small influence which is in agreement with the above.

B) PHASE NOISE

Phase noise is another typical non-ideal behavior any VCO will exhibit. In the proposed receiver architecture, the effect of phase noise will be that the points in the constellation diagram will move around the ideal point. Figure 16 shows simulations with a phase noise of -100dBc/Hz at 1MHz offset. It is clear that phase noise has only a small effect even for fairly modest values when BPSK is used, since there is almost no degradation in the Euclidean distance between two opposite points in the constellation diagram. This is not surprising. Figure 1 showed the spectra of the UWB-pulse and the windowed sine. Due to phase noise, the spectrum of the windowed sine will slightly shift randomly around the desired center frequency. Given the wideband nature of the spectra, most of the signal energy is still captured.

When the VCO output is used as a reference for the digital clocks, phase noise relates to jitter which has an effect on the window placement. Jitter can be seen as an added random small timing error (order of magnitude 10ps – 100ps) on top of the already present fixed window misplacement. In figure 11 it is shown that jitter up to 100ps has almost no effect when the initial window misplacement is less than 1ns.

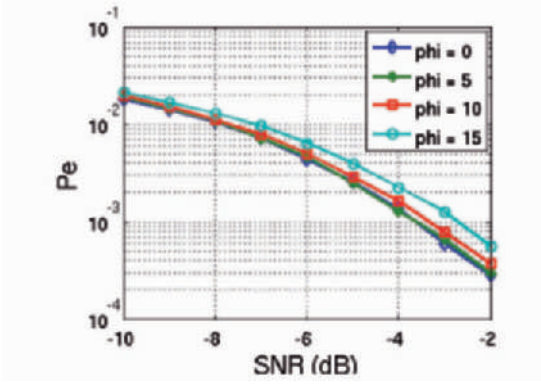


Fig. 15. BER-plots for different values of phase error ϕ.

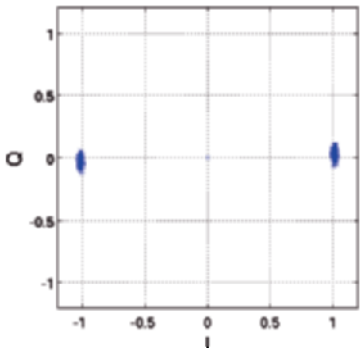


Fig. 16. Effect of phase noise in the IQ-diagram(-100dBc/Hz @ 1MHz).

We can conclude that the specifications for the QVCO in the proposed receiver architecture are quite relaxed. The oscillation frequency is only 500MHz. Phase noise of -100dBc/Hz at an offset of 1MHz and a quadrature phase error of up to 10 degrees can be tolerated. Literature research of low-power quadrature VCO's shows that these requirements can be met with for example coupled differential LC VCO's using power saving techniques as current reuse and back-gate biasing [21,22]. Since the specification for the phase noise is quite low, even ring oscillators [23] can be used in order to lower the overall power consumption.

4.4. ADC's and Digital Backend

Not only the power consumption of the ADC, but also that of the digital backend is highly dependent on the number of bits of the ADC. It is hence very important to keep this number as low as possible, without degrading the performance too much. Figure 17 (a) shows the bit error probability as a function of the input SNR for various ADC resolutions. In this simulation, Gaussian noise is assumed at the entrance of the receiver A perfect AGC makes sure the input signal spans the complete ADC input range. Due to the decorrelation with the spreading code in the digital domain (and the resulting processing gain), the SNR at the input of the ADC will be low. Therefore, the quantization noise can be fairly large, without degrading the receiver performance too much. Our simulations show for a bit error probability of 1e-4 a degradation of respectively 1.2dB, 0.6dB and 0.15dB for a 1-bit, 2-bit and 3-bit ADC respectively in comparison with an infinite-bit ADC. Figure 17 (b) shows the same simulation, but with interference dominated input noise. Due to the non-Gaussian distribution of the noise, the ADC's will suffer from more clipping. As a result, the performance loss for low resolution ADC's increases in relation to the ideal AWGN simulations.

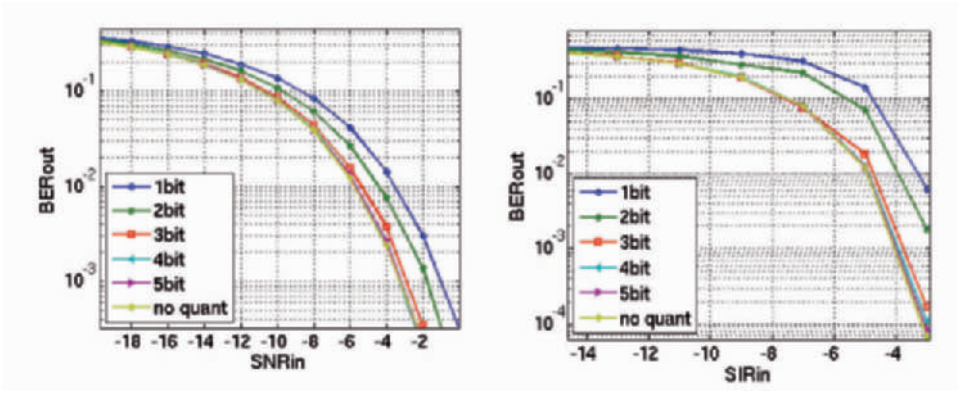


Fig. 17. Bit error probability for different values of the ADC resolution in the presence of AGWN noise (a) and in the presence of interferers (b) as a function of the SNR at the input of the ADC's, N_s=10.

Assuming a flash ADC, the power consumption of the ADC increases exponentially with the number of bits of the ADC (equation (7)). The power of the digital domain can be split in a fixed part and a part that increases linearly with the number of ADC bits. Figure 18 shows the estimated power consumption of the CAC receiver with 1-bit to 5-bit ADC's, based on previous assumptions and the power numbers from table 2.

The second plot of figure 18 combines this power consumption and the performance degradation due to quantization noise (based on figure 17 (b)) to come to the power per useful bit. Reducing the number of ADC bits decreases the power consumption at the cost of a decrease in performance. Figure 18 shows an optimum for a receiver with 3-bit ADC's. Note that this optimum is computed for N_s=10. Increasing this code length, would result in a decrease in the SNR at the input of the ADC's. This causes the curves of figure 17 to come closer to the ideal curve and as a result the optimum will shift to a smaller number of bits for the ADC's.

Figure 19 shows the performance degradation due to ADC offset for a 3-bit ADC CAC receiver. As can be seen, this degradation is very limited, which proves that no power should be wasted in optimizing this parameter. The sensitivity to this offset increases when moving to lower-resolution ADC's.

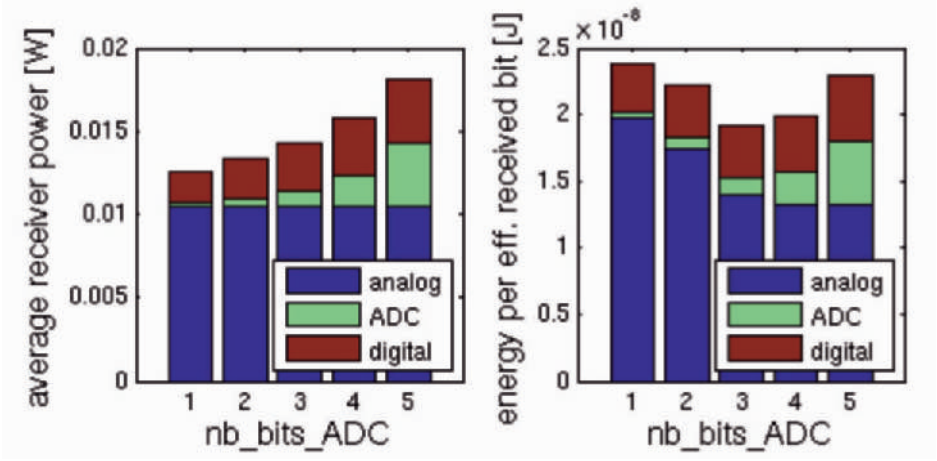


Fig. 18. Total power and energy per effectively received bit for receivers with 1-bit to 5-bit ADC's.

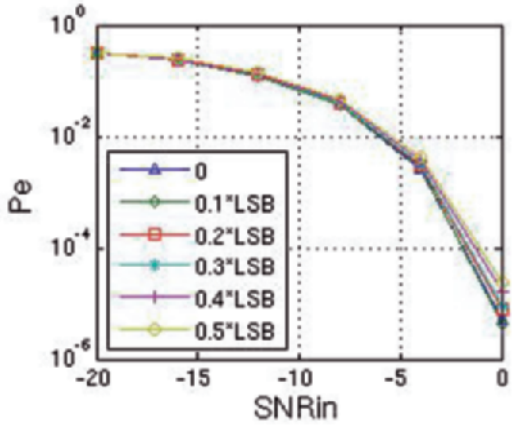


Fig. 19. Bit error probability for different values of the ADC offset (for a 3-bit ADC) as a function of the SNR at the input of the ADC's, N_s=10.

5. Migration to the 3-5GHz Band

Two frequency bands are available for UWB, 0Hz-960MHz and 3.1GHz-10.6GHz. The previous sections all dealt with signaling in the lower band. This section will investigate the trends in performance and power consumption after migration to a higher frequency band for the different receiver architectures described in section 3. Since 802.11a is a strong interferer, present at 5GHz, a frequency band of 3.1GHz-4.9GHz will be used.

A) PERFORMANCE

Almost two times as much bandwidth is available here and fewer interferers are present resulting in at least 3dB better performance. Furthermore the antenna gain will increase significantly unless an unrealistic large antenna for the baseband UWB would be used. On the other hand path loss will increase. Since the center frequency is about eight times higher, path loss will introduce about 18dB additional performance loss. Some additional performance loss is expected due to more scattering at higher frequencies. It can be expected that the overall picture will provide a slightly better performance in the 3.1GHz-4.9GHz band compared to the 0Hz-960MHz.

Formulas 6 and 9 can be used to evaluate the performance of the different architectures relative to each other. The wider bandwidth allows a larger E_p and hence a smaller N_s. Since scattering and path loss increase, η_{cap} will decrease in multipath environments. This effect will be the largest in the CAC architecture, which can not do perfect Rake reception. Looking at formulas 6 and 9 and taking into account the aforementioned effects, one can conclude that the different architectures will perform the same relative to each other, except the CAC receiver, which will have a performance decrease in multipath NLOS environments.

B) POWER

Again all architectures can share the same LNA. It is obvious that the LNA has to operate at higher frequencies over a wider range, which contributes to power consumption. After the LNA, several options are available to deal with the incoming signal, centered around 4GHz. The signal can be downconverted to baseband and the same architectures, as discussed in section 3, can be used. The same conclusions apply, keeping in mind that a 2GHz wide band is used in stead of 1GHz. Another solution is to directly feed the passband signal to the receiver without downconverting. Recall that the FD architectures don't have analog blocks other than the LNA, whereas the AC and FA correlate in the analog domain. The fully digital architectures employed ADC's sampling at the Nyquist rate and digital circuitry clocked at this frequency. Employing techniques as subsampling, still at least 4Gss are needed to sample the 2GHz wide signals. Compared to the 2Gss, one can state that the overall power consumption will be roughly twice as high for the fully digital receivers. The AC and FA architectures require a template generator for which three possible alternatives were discussed, the DAC, CAC and TR. The DAC alternative needs to generate a signal with two times as much bandwidth, meaning that power consumption of this template generator will roughly double. The CAC uses a QVCO, which has to be clocked at an eight times higher frequency. Since the signal bandwidth is only twice as high, this means that the phasenoise specification will roughly become four times more stringent. An oscillator with a higher Q is required resulting in a higher power consumption. The delay line described in [16],

was designed for this higher band and provides a delay of 550ps consuming 58mW. After this first correlation, signals at pulse rate are present. Since the overall performance is slightly better in the higher band, fewer pulses per bit can be used for the same bit-error-rate. The ADC's and the digital back-end need to handle signals only at pulse rate resulting in slightly lower power consumption for these parts.

Looking at table 1 and 2 and taking into account these trends, one can conclude that the CAC receiver still outperform the other template generator in terms of power consumption. However, the specifications for the analog blocks are not as relaxed as in the lower band and the overall power is higher.

To finally make a fair comparison between these receiver alternatives and between the high and the low frequency band, the energy per useful transmitted bit should again be compared. This requires a more detailed investigation of the performance and the power changes which is beyond the scope of this paper. These first trends however seem to indicate that the CAC receiver stays the most promising architecture in LOS and light multipath environments. In dense multipath environments, the DAC and 1-bit FD receiver becomes better.

6. Conclusions

This paper has studied the power-performance trade-offs in an ultra-low-energy pulsed UWB receiver for the 0-960MHz band to obtain the design consuming the least energy per effectively received bit. After an exploration at architectural level, the complex analog correlating receiver has been selected. For this receiver, the trade-offs inside the different building blocks are investigated. The most energy-efficient solution makes uses of an LNA with a $P_{in,1dBc}$ back-off of -5dB, 3-bit ADC's and a 500MHz QVCO. The requirements for the ADC offset, QVCO phase noise and mixer linearity are rather relaxed, which enables a truly low-power implementation.

Acknowledgments

The authors would like to thank I. O'Donnell from the Berkeley Wireless Research Center (BWRC) for the interference measurements.

References

[1] M. Win and R. Scholtz, "Impulse radio: How it works," IEEE Communications Letters, Vol. 2, No. 2, Feb. 1998, pp. 36–38.

[2] FCC, "First report and order," FCC 02-48, Feb. 2002.

[3] M.J. Ammer, “Low power Synchronization for Wireless Communication,” 2004 Ph.D.
[4] R. Blazquez, P. Newaskar, F. Lee, A. Chandrakasan, “A Baseband Processor for Impulse Ultra-wideband Communications,” JSSC, Sept. 2005, vol. 40, nr 9, pp. 1821–1828.
[5] I. O'Donnell, M. Chen, S. Wang and R. Brodersen, “An Integrated, Low-Power, Ultra-Wideband Transceiver Architecture for Low-Rate, Indoor Wireless Systems,” IEEE CAS Workshop on Wireless Communications and Networking, Sept. 2002.
[6] M. Verhelst and W. Dehaene, “System design of an ultra-low power, low data rate, pulsed UWB receiver in the 0-960MHz band,” IEEE ICC, May 2005, pp. 2812–2817.
[7] S. Wang, A. Niknejad, and R. Brodersen, “A sub-mW 960-MHz ultra-wideband CMOS LNA,” RFIC Digest of Papers, 2005, pp. 35–38.
[8] B. Razavi et al., “A UWB CMOS Transceiver,” JSSC, Dec. 2005 vol. 40, nr. 12, pp 2555–2562.
[9] R. Walden, “Analog to Digital Converter Survey and Analysis,” IEEE Journal on Selected Areas in Communications, Vol. 17, No. 4, April 1999.
[10] M. Toyama, S. Dosho and N. Yanagisawa, “A Design of a Compact 2GHz -PLL with a New Adaptive Active Loop Filter Circuit,” Symposium on VLSI Circuits, 2003, June 2003, pp. 185–188.
[11] P. Raha, “A 0.6-1.2V Low-Power Configurable PLL Architecture for 6GHz-300MHz Applications in a 90nm CMOS Process,” Symposium on VLSI circuits, pp. 232-235, June 2004.
[12] R. Hoctor and H. Tomlinson, “Delay-Hopped Transmitted-Reference RF Communications,” Proceedings of the IEEE Conference on Ultra Wideband Systems and Technologies, pp. 265–270, May 2002.
[13] R. Scholtz, D. Pozar and W. Namgoong, “Ultra-Wideband Radio,” EURASIP Journal on Applied Signal Processing, 2005, Vol. 3, pp. 252–272.
[14] M. Verhelst and W. Dehaene, “System design of an ultra-low power, low data rate, pulsed UWB receiver in the 0-960MHz band,” IEEE ICC, May 2005, Vol. 4, pp. 2812–2817.
[15] M. Verhelst and W. Dehaene, “Complex analog correlating pulsed UWB-receiver in realistic 0-1GHz channels,” AusWireless06, accepted for publication.
[16] S. Bagga, L. Zhang, W. Serdijn, J. Long and E. Busking, “A Quantized Analog Delay for an ir-UWB Quadrature Downconversion Receiver,” IEEE International Conference on Ultra-Wideband, Sept. 2005.
[17] IEEE802.15.4a, “15-04-0505-04-004a-UWB-Channel-Model-for-under-1-GHz,” 2004, www.ieee802.org/15/pub/TG4a.html.

[18] FCC, "Radiated emission limits, general requirements," Federal Communications Commission, Code of Federal Regulations, Title 47, Part 15 "Radio Frequency Devices," Section 15.209, Sep. 2005.
[19] L. W. Couch II, Digital and Analog Communication Systems, Prentice Hall, 1997.
[20] W. Sansen, "Distortion in elementary transistor circuits," IEEE Trans. Circuits Syst. II, Vol. 46, pp. 315–324, Mar. 1999.
[21] Nam-Jin Oh and Sang-Gug Lee, "Current Reused LC VCOs," IEEE Microwave and Wireless Components Letters, Vol. 15, No. 11, Nov. 2005.
[22] Hye-Ryoung Kim, Choong-Yul Cha, Seung-Min Oh, Moon-Su Yang, and Sang-Gug Lee, "A Very Low-Power Quadrature VCO With Back-Gate Coupling," JSSC, June 2004, vol. 39, nr 6.
[23] M. Grözing, B. Phillip, and M. Berroth, "CMOS ring oscillator with quadrature outputs and 100 MHz to 3.5 GHz tuning range," Proceedings of ESSCIRC03, pp. 679–682, September 2003.
[24] GSM 05.05 Version 5.2.0, July 1996; Digital cellular telecommunications system (Phase 2+); Radio transmission and reception.

DESIGN CONCEPTS FOR WIRELESS COMMUNICATION IN IMPLANTABLE MEDICAL APPLICATIONS

Michel De Mey, Craig Christensen, Shane Blanchard
AMI Semiconductor Belgium BVBA
Senneberg, J. Monnetlaan, Vilvoorde, B-1804 Belgium
AMI Semiconductor
2003 Buckskin Road, Pocatello, ID83201 USA

Abstract

The number of available electronic implantable devices is increasing every year. The complexity and functionality of these devices is also increasing at a significant rate. Communications with these complex devices for physician adjustment and to collect data is becoming increasingly important to provide the maximum patient comfort and the most effective treatment for medical conditions. A Communications band has been allocated in both North American and Europe specifically for communications with implanted devices. This paper reviews some of the unique requirements for an implanted wireless transceiver, regulatory requirements of the Medical Implantable Communication Service (MICS) band, provides a description of a product competing in this arena and a unique crystal startup circuit that can significantly reduce power consumption in power critical implant applications.

1. Introduction

In 1960 the first totally implantable pacemaker was implanted in a human. Since that time the number of applications of implantable electronics has been increasing at a steady pace. In 2006 roughly $10 billion will be spent on cardiac rhythm management devices alone. Implantable medical devices (IMDs) are successful in the treatment of many diseases and neurological disorders. These devices can interact directly with the brain and nervous system and target the treatment of a range of problems such as heart regulation/stimulating, hearing loss, epilepsy, depression, and even obesity. Given the aging population, there is an increasing need for more advanced healthcare treatments, including wireless implant devices that are able to deliver ongoing and cost effective monitoring of a patients condition.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 331–343.

2. Design Constraints of Implantable Devices

Implantable medical devices, communicating with the outside world have design constraints, which create unique challenges for system designers as well as the IC designers. Careful analysis of these impacts is required to find the best compromise amongst them.

2.1. Ultra Low Power Consumption

There are 2 strategies currently utilized to provide power to medical implantable devices. One is to supply the system by means of an internal non-rechargeable battery. The other is to provide power by means of inductive coupling. Experimentation is on going with a number of energy scavenging techniques, however if they are to be used, they must have little impact on the patient. These techniques, if successful will not fundamentally alter the ultra low power design constraint.

Implanted battery powered devices have a typical requirement of operation for 5 to 7 years. This means that the system designer as well as the IC designer need to analyze and reduce the power consumption as much as possible.

One of the best ways to conserve power is the most evident one. Turn as much of the circuitry off when its operation is not needed.

However in the sensor/transceiver application it is unknown by the receiver when data is being transmitted. Systems have evolved in a way that the transmission message is generally preceded with a header of some specified length. It is then possible to periodically turn on the receiver to look for the header transmission. This type of operation is known as Sniff Mode™. Every 0.5 to 16 seconds the receiver wakes up and listens to find out if it is being interrogated. If it detects no signal the device goes back to sleep. If a transmission is detected the receiver remains on to receive the transmitted data.

In sleep mode the digital domain will typically still have the power supply connected in order to keep the states of the device. Static leakage and characterization thereof is crucial for the IC design. At feature sizes below 350nm the off state leakage rises rapidly due to lower V_T scaling. At feature sizes below 130nm the gate leakage start to play a significant role. Process technology choice is of vital importance in implantable medical devices.

In order to quantify the threshold voltage effect, the standby power consumption P_{stat} can be modeled by the following formula:

$$P_{stat}=I_0*10^{-(V_T/S)}*V_{DD} \quad (1)$$

Where V_T is the device threshold voltage, I_0 is a constant, proportional to the total transistor width of the chip. S is the sub threshold slope. (for a .35

um technology this value is typically 80mV/decade). Vdd is the supply voltage in sleep mode. As a reference the impact of V_T on the static power consumption at a typical Vdd of 1V is plotted in figure 1.

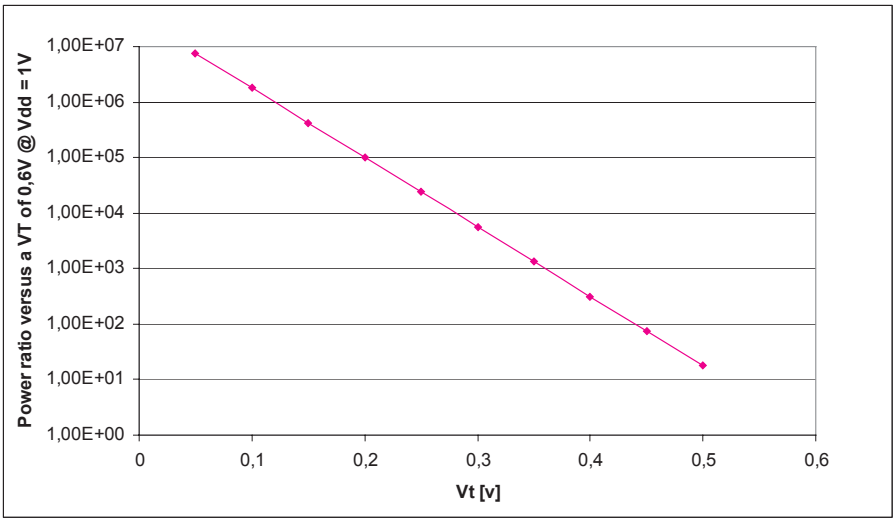


Fig. 1. Vt influence on Standby Power (reference Vt = 0,6V).

Hence techniques are introduced to reduce the leakage power such as multi Vt CMOS technologies and back biasing to increase the Vt during idle mode.

It is critical that the analog blocks have a power down state that will be controlled by the system's power management controller. The faster the system can be powered up, perform it's function and power down, the less power will be needed. An analog function that typically requires a long power up time is a crystal oscillator. In section 4 the detailed and patented implementation of a quick start crystal oscillator will be described.

2.2. The Human Body as Medium

Transmission in the human body has it's own challenges. The human body is composed of materials that are partially conductive and that have different dielectric constants. Furthermore when analyzing the implemented antenna the antenna can not be viewed in isolation of the body, the human body should be viewed as part of the antenna. This is because the radiation pattern and efficiency are influenced by the size, shape, location and composition of the body. The antenna, hence the implanted medical device will behave differently when implanted in the arm of a person, than when implanted in the abdomen.

The effective permittivity ε_{er} and conductivity σ_e of different human tissues, relevant for IMDs are given in table 1.

Table 1. Dielectric parameters human tissue at 403.5MHz.

Tissue	ε_{er}	σ_e (S/m)
Muscle	57.1	0.797
Fat	5.6	0.041
Lung	23.8	0.375
Skin (dry)	46.7	0.690
Skin (wet)	49.8	0.670
Bone Cancellous	22.4	0.235
Brain gray matter	57.4	0.739
Brain white matter	42.0	0.445

Testing of antenna and system performance of an implanted antenna in the lab is done by means of tissue simulating fluids. These fluids are the same ones that are used to measure specific absorption rates (SAR) in the evaluation of mobile handsets. These fluids are combinations of water, sugar, salt (NaCl) and Hex (Hydroxyethylcellulose).

The layered structure of the human body, like the interface between muscle and fat, gives rise to reflections that need to be taken into account.

Additionally the antennas are small in size, which makes the antennas for implantable medical devices very inefficient. The receivers will have requirements for high sensitivity and low noise figures to overcome the path loss, poor antenna performance and low transmit power levels.

2.3. Regulatory Requirements

In the early 1970s low frequency inductive links were the most frequent communication methods. They were characterized by low data rates (in the order of 1-30Kb/s) and operated in a frequency range of tens to hundreds of kHz.

The inductive links also had the limitation of being very short range operation.

Newer systems are designed to overcome the limitations of low data rate and short range. These systems are ultra low power RF, operating in the 433MHz to 915MHz Industrial, Scientific and Medical (ISM) bands or in the recent established 402-405MHz Medical Implantable Communication Service band (MICS band).

The Medical Implant Communications Service (MICS) is an ultra-low power mobile radio service for transmitting data in support of diagnostic or therapeutic functions associated with implanted medical devices. It serves implanted applications such as cardiac pacemakers and defibrillators. The

typical MICS system block diagram is represented in figure 2. The allocation of this band supports the use of longer range (typ. 2m), high speed wireless links.

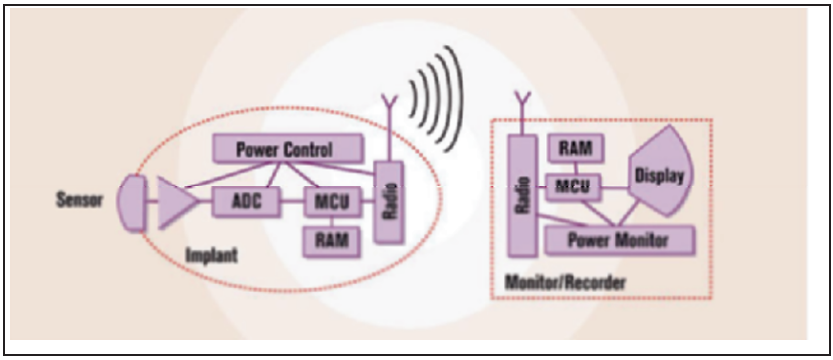


Fig. 2. The typical MICS system.

The key Medical Implantable communication services regulatory requirements are listed below:

- The controller must initiate all non-medical event communications
- The controller must perform a Clear Channel Assessment (CCA) prior to any MICS Transmissions. The controller scans all 10 of the 300kHz channels and is allowed to transmit on the channel with the lowest ambient signal level.
- The implanted Medical Device may transmit a medical event without a CCA
- The implant scans channels for the controller TX signals
- 300KHz Maximum Emission Bandwidth
- +/- 100 ppm Frequency Stability
- 25 microwatts Equivalent Isotropic Radiated Power (EIRP) [-16 dBm]
- Data Communication Only (No Voice Communication is allowed)
- Radio Must Be Certified

2.4. Other Requirements

Besides low power requirements there is the mission critical requirement for the use of technologies, processes and design techniques with a high reliability. Providing components and services for implantable medical devices requires extensive certification. The Federal Drug Administration (FDA) regularly inspects device manufacturers as well as their suppliers. Techniques applied in the frame of QS-9000 for automotive equipment manufacturers are applied to medical designs like Design Failure Mode and

Effect Analysis. Design techniques in order to reduce soft error rate (SER) sensitivity are also applied.

Furthermore it is desired to minimize the usage of external components. Medical grade components are costly and every connection or solder joint adds additional reliability risk.

The trauma of implantation needs to be minimized, which brings requirements of miniaturization and size and the implantable medical device needs to be built out of materials that ensure biocompatibility.

3. A Frequency Agile Transceiver for Low Data Rate RF in Medical Applications

In the section we will describe the latest product, the AMIS-53000, out of the AMI Semiconductor ASTRIC™ family. The product is providing a good overview of the functional analog and digital design blocks required for transceivers in medical applications. The product is suited for low to moderate data rate (up to 128kbps), low power, sub 1GHz, narrow band, FSK/GFSK/OOK multiple channel wireless applications. The integrated circuit is designed in a 350nm technology that includes EEPROM.

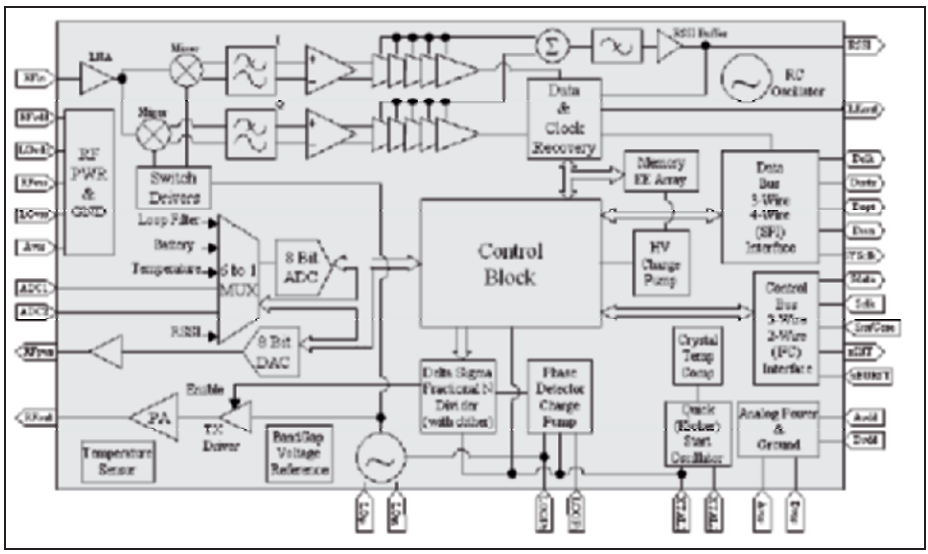


Fig. 3. Circuit diagram.

A low IF conversion receiver architecture was selected for the implementation. The mixer operates at 500kHz offset from the received carrier and achieves 1dB NF, and more than +80dBm IIP2. [1] The receiver 3dB blocking wideband selectivity is better than 60dB for frequency offset +- 1MHz and higher.

The transmit current is typical 50mA (15dBm), the sleep mode current of the device is less then 1uA. Features like Sniff mode™ and multi channel frequency setting simplify the design of a MICS compliant protocol.

The device requires a single external crystal working with an internal VCO and PLL to generate frequencies from 300 to 928MHz. The device supports On/Off Keying (OOK) at data rates from 1 to 19,2 kbps and (Gaussian) Frequency Shift keying (FSK/GFSK) at data rates from 1 to 128kbps. Internal capacitors eliminate the need for external load capacitors. Fast start up is achieved with the patented quick start crystal oscillator.

4. Quick Start Crystal Oscillator Design Example

As mentioned in the design constraints, implantable medical devices require reliable fast power up and power down. One of the most difficult circuits to turn on quickly is the crystal oscillator. The design example given here will show how the start-up time can be reduced from 5 to 10 milliseconds [2] to 5 to 10 microseconds. The fundamental design principle is to get the crystal to oscillate quickly by energizing the crystal tank with a controlled stimulus until there is enough energy to sustain oscillation [3]. The circuit is used in several AMIS transceivers to turn on the receiver in Sniff Mode ™ and is implemented in AMI Semiconductor's 0.35um CMOS process.

.

4.1. Top-Level Architecture

Figure 1 shows a top-level block diagram of the crystal oscillator and quick start circuitry.

The circuit has four major blocks. The crystal oscillator block, a temperature compensated binary trimmed ring oscillator, control logic and calibration circuitry. The calibration circuitry is used to trim the ring oscillator frequency to match the resonant frequency of the crystal.

The objective of quickly starting the crystal oscillator circuit is to get enough energy into the quartz crystal so that the differential voltage across its terminal will be large enough to have the output buffer to produce a square wave output signal. To achieve this, it is necessary to integrate as much current into the inductor, as per the crystal equivalent electrical model as fast as possible (4).

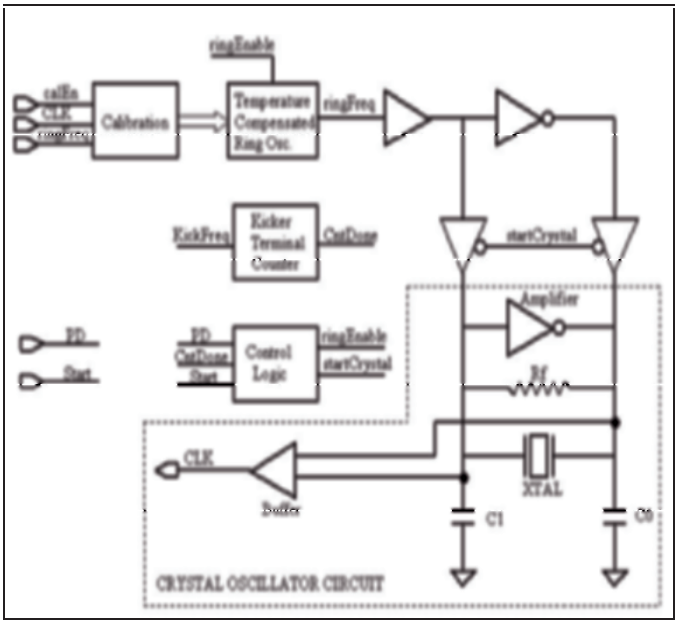


Fig. 4. Top level architecture.

$$i_L = \frac{1}{Ls} \cdot \int_0^t v \cdot dt + i_L(t_0) \tag{4}$$

This is accomplished by applying an alternating differential voltage signal across the terminal of the crystal at a rate that is as close to the crystal's resonant frequency as possible.

By applying the AC signal at the resonant frequency, the impedance across the terminals is at a minimum, thus maximizing current flow through the crystal. This signal is applied until there is enough energy to provide a stable clock signal. If the frequency of the applied signal is different than the resonant frequency of the crystal, a beat note occurs at a frequency equaling the difference between the applied signal and the resonant frequency. When this occurs the build up of energy into to crystal will have a cyclic characteristic. This means that the energy will build up and decay at the rate equal to the frequency of the beat note. To maximize the amount of time available for energizing the crystal (one half of a beat note cycle), an accurate alternating differential stimulus voltage with a frequency as close to the crystal as possible needs to be designed.

4.2. Quick Start Circuitry

A binary weighted current controlled ring oscillator is used to provide the alternating differential energizing stimulus to the crystal. A series of current controlled delay cells are cascaded with the inverting output fed back to the input to form a ring oscillator. Each delay cell contains a capacitor that is charged and discharged at a rate dependent upon a binary weighted controlled current source.

To ensure that the ring oscillator frequency would match the resonant frequency of the crystal, a few design requirements had to be made. This design required that the frequency of the ring oscillator had to be voltage, temperature and process independent.

To meet the voltage independency, the voltage supply was regulated to a voltage of 100mV below the minimum specified battery operating range. The regulated circuit used a constant voltage reference provided by a band-gap circuit, which varied only a few millivolts over its voltage and temperature operating range. Process variation was removed by calibrating the ring oscillator once the system is powered. This will be discussed in the following section. The temperature independency required a tightly controlled temperature compensation circuit. The temperature compensated circuit provides a current to the ring oscillator that has a DC component that sets the frequency of the ring oscillator and a linear positive sloped temperature dependent component (5).

$$I_{Bias} = I_{PTAT} \cdot Temperature + I_{DC} \tag{5}$$

The temperature dependent current will provide more current to the ring oscillator when temperature increases and CMOS transistors slow down and inversely provide less current when the circuit speeds up at colder temperatures. This current compensation has a temperature coefficient (TC) that inversely matches the TC of the ring oscillator.

However, the ring oscillator requires a different compensation slope when the trim word varied from 00h to FFh. To adjust for a varying compensation slope, a current mirroring structure was designed to provide a separate compensation slope for both trim boundaries. The inputs to the temperature compensation circuit consisted of a DC current source provided from the band-gap circuit and a PTAT (proportional to absolute temperature) current source. The DC current supply is used to set the low and high frequency limits of the ring oscillator. PTAT current is needed for temperature compensation.

As seen in figure 4, these two currents are mirrored to four independent transistors. Each transistor is sized to adjust for the lower and upper frequency range as well as adjust the compensation slope for the entire trim range. The currents are summed together and connected to each of the ring oscillator's delay cells.

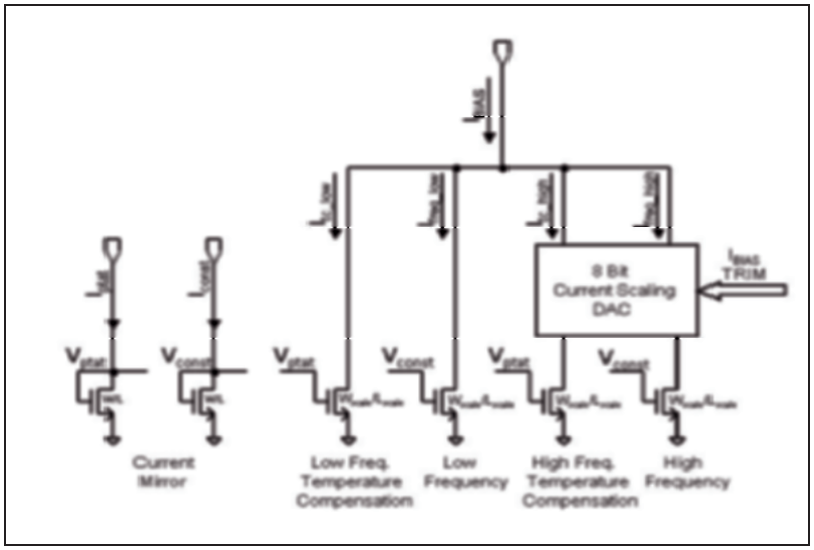


Fig. 5. Temperature compensation.

Knowing the frequency deviation for the ring oscillator through simulation and the resonant frequency of the crystal, control circuitry can be used to quickly start the crystal oscillator.

By calculating a worst-case difference between the ring oscillator and the crystal, a terminal counter can be programmed to optimize the duration of the excitation cycles of the crystal. Optimal duration is defined when the energy in the crystal has peaked in the first beat note. When start is asserted high, a state machine resets the terminal counter, enables the tri-state buffers and waits for the terminal counter to terminate. After the crystal has been started, the state machine disables the tri-state buffer and power downs the ring oscillator.

Once the state machine has completed the quick start cycle, the crystal has enough energy to sustain a stable oscillation and provide enough voltage swing to the buffer to produce a square clock signal. When start is de-asserted, the state machine will reset and wait for another start to go high. In between quick start cycles, the crystal oscillator is powered down and the energy inside the crystal will fully dissipate, otherwise the quick start circuit wouldn't be necessary.

4.3. Self Calibration Circuit

In order to alleviate frequency variation due to process, a digital circuit is used to trim the ring oscillator to be within an acceptable frequency tolerance of the crystal's resonant frequency. When the battery voltage is first applied to the system, a signal is asserted (calEn) that initiates a calibration routine. The calibration first waits 15 to 20 milliseconds for the crystal to self-start. Once the crystal has self-started, the ring oscillator is also started. The calibration circuit in figure 5 consists of two terminal counters - one being clocked by the buffered clock output of the crystal oscillator circuit and the other by the ring oscillator.

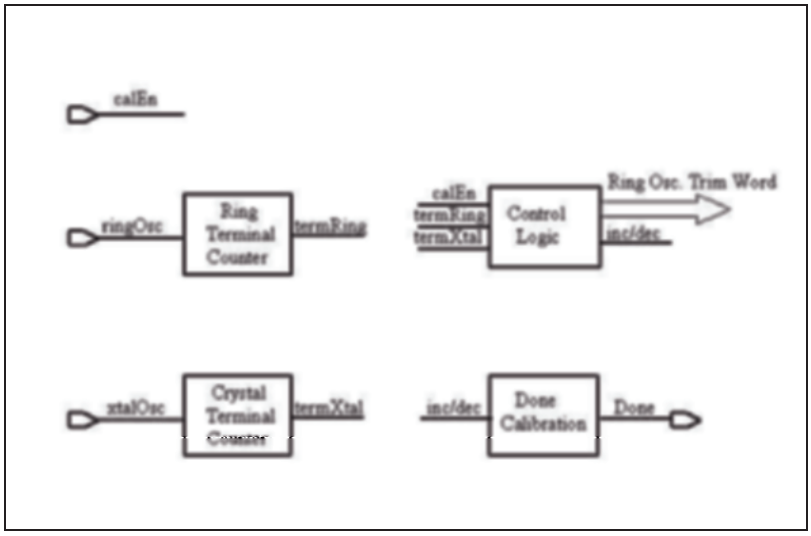


Fig. 6. Calibration Circuit.

A state machine is used to reset the counters and waits to determine which counter terminates first. Depending on which counter terminated first, the state machine will either increment or decrement the binary weighted current trim value controlling the frequency of the ring oscillator. Time is allowed for the ring oscillator to settle to its new frequency and the terminal counters are reset and the process happens again. Once the ring oscillator's frequency is adjusted to within an LSB of the crystals frequency, the increment and decrementing of the trim word will begin to toggle back and forth. After a series of back and forth toggles between incrementing and decrementing an output will be asserted high to stop the calibration. This process will not occur again until the battery is replaced in the system or software initiates a re-calibration. By calibrating, differences in crystal frequency due to process, component and board layout variations can be eliminated.

4.4. Start Up Sequence

With the implementation of the quick start oscillator circuit, the turn on time is approximately 10us. By characterizing the ring oscillator, a maximum deviation between the crystal resonant frequency and the ring oscillator allowed excitation of the crystal to be between 5us to 8us. After the crystal oscillator is started and the amplifier returns to its steady state, the total turn on time is 10us. In figure 6, a plot of the quick start is shown. The XTAL1 and XTAL2 signals are the voltages at the crystal pins. The signal REF_CLK plots the reference clock.

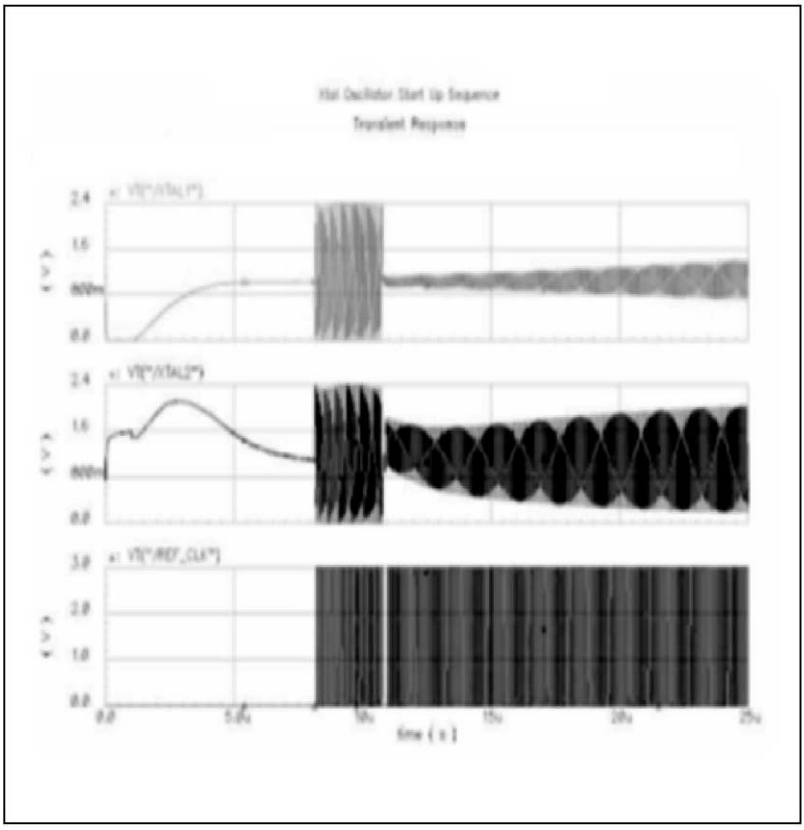


Fig. 7. Start Up Sequence.

Even though the crystal energy hasn't reached its steady state after excitation, the clock is close enough to its steady state value to allow the system to function correctly. In the Sniff Mode^{TM} application, the PLL can begin locking during the excitation of the crystal and is locked within 40us after the system has started. After another 50us of operation, the wireless receiver, which uses the crystal for a frequency reference, is able to

determine whether or not signal is present. If not, the receiver will turn to the off state.

5. Conclusions

Communications with increasing complex implanted medical devices is becoming a requirement to allow these devices to be used in their most effective manor, providing the greatest benefits for the patient. The severe conditions placed upon the wireless components due to the in body environment and limited power availability require special design techniques to be used to maximize performance and minimize power consumption. This paper reviews some of the limitations imposed by the implant environment and the wireless regulations. A product operating in this environment is described and an innovative circuit that can quickly start a crystal oscillator is also presented. This circuit can significantly reduce power consumption on startup, a critical parameter of implanted devices.

References

[1] C. Christensen, "A Low Power, Low Data Rate Integrated 433 MHz RF Transceiver in CMOS," in 2003 IEEE UGIM Symposium Digest, Boise, July 2003.

[2] A. Rusznyak, "Start-Up Time of CMOS Oscillators," IEEE Transactions on Circuits and Systems, vol. CAS-34, No. 3, March 1987, pp. 259–268.

[3] E. A. Vittoz, "High-Performance Crystal Oscillator Circuits: Theory and Application," IEEE Journal of Solid-State Circuits, vol. 23, No. 3, June 1988 pp. 774–783.

WiseNET, an Ultra Low-Power RF Transceiver SoC and Communication Protocol Solution for Wireless Sensor Networks

Vincent Peiris, Amre El Hoiydi, Antoine Ribordy, Erwan Le Roux, Thierry Melly, David Ruffieux, Franz Pengg, Frédéric Giroud, Nicolas Raemy, Martin Kucera, Lauri Sumanen, Patrick Volet, S. Cserveny, Claude Arm, Pierre-David Pfister, Ricardo Caseiro
CSEM SA, Swiss Center for Microelectronics and Microtechnology

Abstract

Autonomy and size are the most important challenges faced when designing radios for distributed wireless sensor networks (WSN). Reducing power consumption requires optimization across all the layers of the communication stack. The WSN platform developed at CSEM therefore uses a co-design approach that combines WiseMAC, a low-power media access control protocol, with the WiseNET SoC, a complex system-on-chip sensor node to exploit the intimate relationship between the MAC-layer and the radio transceiver parameters. This paper reviews the design and realization of the WiseNET SoC featuring a low-power 1V short-range UHF radio transceiver implemented on a 0.18μm standard digital CMOS process. The WiseNET radio consumes only 2.3mW in receive mode with a sensitivity smaller than −108-dBm at a BER of 10^{-3} for 25kb/s FSK in the 433MHz ISM band. The design, simulation and validation of the WiseMAC protocol are also detailed in the context of the deployment of a small ad-hoc network experiment at CSEM, demonstrating that the consumption of the WiseNET solution is more than an order of magnitude lower than a comparable Zigbee-based solution.

1. Introduction

A wireless sensor network (WSN) designates a system composed of numerous nodes distributed over an area in order to collect information. Applications of sensor networks are numerous, ranging from environmental monitoring, home and building automation to industrial control [1].

Most applications of wireless sensor networks require both a low sensor node cost and a lifetime of multiple years. These two application requirements translate in the ultra low-power system requirement. Low power design must be addressed both at the level of hardware and software.

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 345–375.

On the hardware side, low power design can be addressed at the technological, logical and system levels [2]. The technological level refers to the integrated circuit (IC), for example with the proper choice of circuit voltage and current levels, their frequency of operation or the minimization of parasitic capacitances on high-frequency nodes. The logical level refers to power aware circuit design with the implementation of power-efficient clocking schemes, and designing the blocks for fast turn-on, turn-off and turn-around times. The system level concerns the minimization of the energy consumed between the transceiver IC, the CPU or the other peripheral circuits that may be embedded on a specific application PCB.

On the software side, power aware design consists on the one hand to adapt the CPU frequency to the processing needs, to go into sleep mode whenever possible, and to turn on hardware peripherals only when needed. One of the most power-consuming hardware peripheral in a wireless sensor is the radio interface, whose power management needs to be addressed by the application and by all layers of the communication protocol stack. The application may decide to turn off the transceiver during specific periods, but when data must be transmitted, the communication protocol stack must also rely on energy efficient mechanisms.

A clear objective for the WiseNET project running at CSEM was to ultimately reach the lowest possible power consumption at both hardware and software levels. The overall design therefore required an intimate co-design approach by building up progressively the pieces of the physical layer in parallel with those of the communication protocol.

This co-design approach resulted in the design of both a low-power sensor node system-on-chip hardware (WiseNET SoC) and a low-power communication protocol stack (WiseNET Stack). A crucial component of the WiseNET stack is the WiseMAC protocol. The WiseMAC protocol is a medium access control protocol specifically designed to meet the requirements of sensor networks. It trades latency and maximum throughput against low power consumption under low traffic conditions and high energy efficiency under high traffic conditions.

Some important constraints underlie the design of the WiseNET SoC and protocol:

1) In order to achieve the targets of low cost and of multiple years lifetime, it was deliberately chosen to use a single AA alkaline battery as the energy source. This constraints the supply voltage of the WiseNET SoC from 1.6V to as low as 1V. Furthermore, with a capacity of 2.6Ah and a leakage of 10% of the initial capacity (amounting to roughly 30μA leakage current), it can be observed [3] that the average power consumption must be below 100μW to reach a lifetime of more than two years.

2) Ultra low power consumption requires shutting down the radio transceiver whenever possible. In order to achieve low latency communication, any protocol requires to frequently switch the transceiver on and off. To minimize the energy spent in transitions, it is important to minimize the energy required to turn the transceiver on. In the WiseNET project, a wake-up mechanism (called *preamble sampling*) based on a periodic measurement of the energy present on the radio channel was selected. As a result, special care has been given to the minimization of the energy spent for sampling.

3) A transmission power was chosen to be set at to 10dBm which is the maximum allowed in the targeted 433MHz frequency band. The motivations for this choice are on one hand the need of many applications to have a long transmission range and the fact that having a variable transmission power leads to inefficiencies. On the other hand, because the WiseMAC protocol uses a wake-up scheme based on periodically listening, on the contrary to other protocols such as 802.15.4 that require the periodic transmission of beacons, it was possible to afford a large power consumption in transmit state without significantly penalizing the overall average power consumption.

This paper presents the WiseNET SoC in Section 2 and the WiseMAC protocol in Section 3. Section 4 presents multi-hop sensor networking experimental results from the WiseNET project, and Section 5 concludes.

2. The WiseNET Transceiver SoC

2.1 Overview of the System-On-Chip

The radio transceiver peripheral is one of the most critical blocks in terms of power consumption at the hardware level for a wireless sensor node. The deliberate choice of a 1.6V alkaline cell as the power source implies that the hardware must be compliant with a supply voltage as low as 1V which is the end-of-life voltage of the battery.

This low-voltage requirement brings in an interesting design constraint for the radio. Previously published radio circuits clearly confirm the interest and feasibility of low voltage 6V alkaline cell as the power source implies that the hardware must be compliant with a supply voltage as low as 1V which is the end-of-life voltage of the battery radio blocks such as a 1V CMOS RF receiver-only in [4], a 1V RF transmitter-only in [5] and a 1V transceiver in [6] and [7]. The WiseNET SoC presented here is a fully integrated 1V wireless system-on-a-chip, whose block diagram is illustrated in Figure 1. It is clearly a complete system in the sense that it embeds not only a low-datarate short-distance low-power UHF radio transceiver, but also

a sensor interface with signal conditioning and with analog-to-digital conversion, a control unit based on a low-power RISC micro-controller with a on-chip SRAM, and a power management unit.

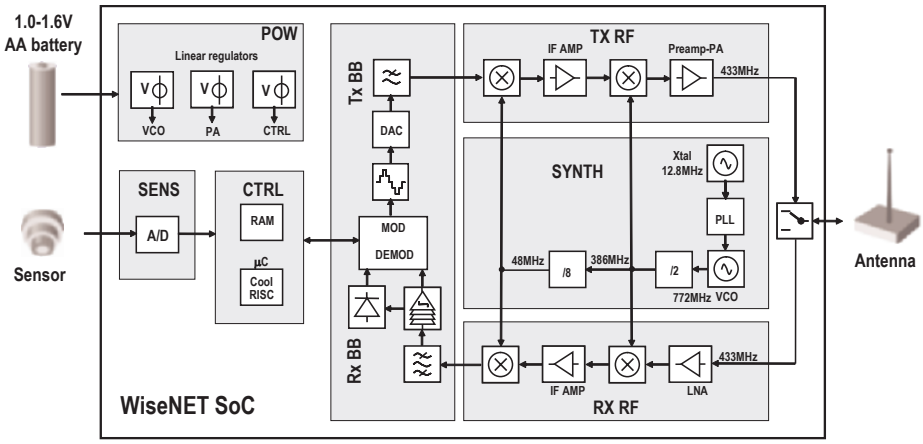


Fig. 1. Block diagram of the WiseNET SoC.

An important design constraint was to achieve a high degree of integration of all the system blocks on a single die. For this purpose, the chosen technology was a "standard digital" 0.18µm CMOS process. As a matter of a fact, the low V_T and high f_T of the 0.18µm transistors allow 1V operation at sub-GHz frequency by operating the devices in weak and moderate inversion regions where the g_m/I ratio remains near its optimum value [8], thus even for the circuits of the RF front-end of the WiseNET SoC.

Furthermore, the choice of such a deep-submicron CMOS process enables the realization of highly integrated systems especially in the case where the SoC includes large and complex digital blocks jointly with analog or RF circuitry. The low cost issue for the wireless sensor node dictated the deliberate choice of a "standard digital" CMOS process, which means without precision components such as capacitors and resistors, nor dedicated RF technology options such as substrate isolation and thick top-level metallization. From a commercial point of view, the high degree of integration achieved on the standard digital 0.18µm CMOS process enables both the miniaturization of the WiseNET sensor nodes, because most system parts happen to be integrated on the SoC, as well as the reduction of the overall cost thanks to the reduced bill of materials.

The WiseNET system-on-a-chip in Figure 1 is organized around a control unit (CTRL) based on a CoolRISC C816L core and peripherals [9], with embedded low-leakage SRAM [10]. This is a low-power oriented RISC micro-controller that is used for the configuration of all the system blocks,

especially the sensor's signal acquisition, conditioning and monitoring throughout the sensor interface, and the operation of the radio link with the operation of the WiseNET stack. The sensor interface (SENS) is built in order to allow typical low frequency wireless sensors to be handled, and comprises an analog-to-digital conversion stage, whose output data can wake up and then be processed by the control unit. The power management unit (POW) starts from a 1.0V-1.6V AA battery voltage and generates the supply voltages for supplying noise-sensitive blocks in the transceiver as well as the control unit by means of dedicated linear regulators.

The UHF radio transceiver consists in a receive (RXRF) and a transmit (TXRF) front-end that targets the 433MHz ISM band. It uses a common back-end for the receiver's and the transmitter's baseband channels (RXBB and TXBB). The RF front-ends also include a common frequency synthesizer block (SYNTH). The following sections describe the main radio characteristics and building blocks.

2.2. Main Radio Characteristics

The transceiver is a short-range low-datarate radio, whose main communication characteristics are summarized in *Table 1*.

Table 1. Main WiseNET radio communication specifications.

Parameter	Condition	Value	Unit
Supply voltage	AA alkaline battery	1.0-1.6	V
Operating frequency	ISM band	433.050 – 434.790	MHz
Primary frequencies		433.3, 433.9, 433.5	MHz
Channel spacing		600	kHz
Propagation range	Target	~2k outdoor - ~10 indoor	m
Data rate	FSK, min / typ / max	12.5 / 25 / 100	kb/s
Frequency deviation	FSK, min / typ / max	12.5 / 50 / 100	kHz

The receiver and transmitter channels are built according to a two-step super-heterodyne architecture such as in [11]. The first and second local oscillator (LO) frequencies are set to respectively 8/9 and 1/9 of the operating RF value. Both the receiver and the transmitter share a common frequency synthesizer PLL and division chain for the generation of the LO signals, and thus the same frequency plan.

For a 433MHz RF, this yields an intermediate frequency f_{IF}=48MHz with a LO frequency f_{LO}=386MHz. The frequency synthesizer uses a VCO running at f_{VCO}=772MHz and a dedicated /2, /8 division chain for generating the required LO signals at, respectively, 386MHz and 48MHz.

2.3. The Receiver Channel

The architecture of the receiver is illustrated in Figure 2.

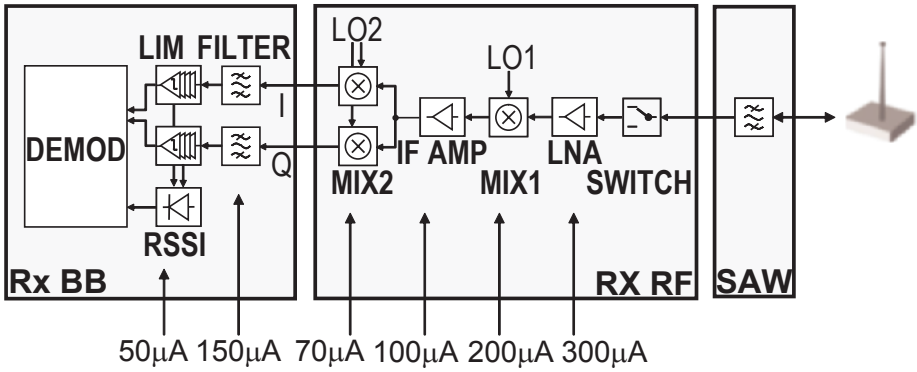


Fig. 2. Overview of the receiver architecture with the current consumption breakdown for the main sub-blocks.

The receiver path consists of an LNA, a first down-conversion mixer (MIX1) to the intermediate frequency, an IF amplifier and a second direct conversion mixer (MIX2). The backend includes a baseband filter, limiters, RSSI and an FSK demodulator.

With the 1V supply constraint, designing for a reasonable rejection of strong out-of-band interferers would require a very high dynamic range, which would translate into prohibitive current consumption for a wireless sensor node. Therefore an external SAW RF filter is used.

Although the super-heterodyne architecture means additional blocks at the intermediate frequency when compared to direct conversion architectures, it proves to be an interesting choice in term of power consumption. Operation at a rather high IF of 48MHz happens not penalize their current consumption, because even with a low bias current, the required 60dB of SFDR can be achieved easily while presenting the required frequency response for the IF blocks thanks to the high-f_T MOS transistors of the 0.18μm CMOS process. In addition, gain can be added at IF in order to reduce the contributions of the following stages to the overall noise, at much lower current expense than if the same gain had to be applied on the RF stage. Yet another advantage of the chosen super-heterodyne architecture is to allow the introduction of AC-coupling blocks between each stage that enable the suppression of DC components along the signal path.

With respect to mixer noise, and particularly 1/f noise, the super-heterodyne architecture presents the advantage of relaxing the design constraints on the first and second down-conversion mixers [12]. For the first

down-conversion mixer from RF to IF (MIX1), the flicker noise contribution is not an issue as the output frequency (f_{IF}=48MHz) has been set by construction well above the 1/f noise corner frequency for the 0.18μm process, which lies around a few hundreds of kHz with reasonable device sizes and bias conditions. If a direct-conversion, or even a low-IF, architecture had been chosen, the design constraints on the direct-conversion mixer would have been much tougher as the output frequency, in this case in baseband, would be lying below the 1/f noise corner frequency, making the 1/f noise contribution of the mixer significantly higher. The 1/f noise contribution would then need to be reduced as low as possible by increasing the bandwidth of the mixer's internal node up to at least several times the RF bandwidth [13], at the expense of high current biasing. Hence, a direct conversion approach was not chosen here, with respect to the affordable current budget for a wireless sensor node.

These 1/f noise constraints nevertheless apply to the second down-conversion stage (MIX2) which is clearly a direct-conversion mixer. Fortunately, the impact appears not so stringent when compared to a RF direct-conversion mixer as the input frequency is here the IF and not RF. Taking advantage of the high f_T capability provided by the 0.18μm process, the bandwidth of the internal node of MIX2 can be set in the order of a GHz, therefore easily achieving a ratio between internal node bandwidth and input frequency (f_{IF}=48MHz) well above ten, hence with sufficient margin for reducing significantly the 1/f noise contribution of the down-conversion mixer.

Considering the high-frequency section of the receiver channel, the RF input drives a single-ended LNA built around an MOS transistor with integrated inductive load and source degeneration. With the use of an external high-Q input inductor for impedance matching, sufficient gain is achieved for a low bias current since the MOS is operated in moderate inversion (I_b=300μA, A_v=24dB). A single-balanced first mixer is used with the input device sized in strong inversion for linearity purposes whereas the switching devices remain in weak inversion to comply with the 1V supply. At low bias current, noise and linearity considerations limit the mixer output load and therefore the mixer gain (I_b=200μA, A_v=6.5dB). Altogether, the LNA and mixer present a 60dB SFDR. The image frequency is rejected by the SAW filter. IF amplification is achieved with a g_m-pair whose gain is digitally adjustable with three 4.5dB steps by switching different degeneration resistor values, and biased in moderate inversion with 100μA. Down-conversion to DC is done using a I-Q double-balanced mixer whose input devices are biased at the highest possible inversion mode at 1V, yielding a 6dB gain and 76dB of SFDR for a 100μA bias current. The mixer load includes a programmable 40-470kHz second order passive filter adding rejection of out-off band signals without degrading linearity. Its cut-off frequency is placed sufficiently high taking into account the +/-25% RC

dispersion due to process variations in order not to disturb the subsequent baseband channel filter characteristics.

On the baseband side, the channel filter consists of a 12.5kHz first order high-pass pole to reject DC and 1/f contributions, and a Butterworth 3rd order active low-pass continuous-time g_m-C filter whose cut-off frequency (50kHz nominal) is automatically adjusted to the crystal oscillator reference of the radio PLL and can be programmed between 25kHz and 150kHz. A cascade of eleven 6dB gain stages with a limiter at the output provides saturation before feeding into the digital FSK demodulator. The nominal FSK operation is 25kb/s with a 50kHz frequency deviation, but the channel can be programmed for 12.5kb/s to 100kb/s data rates and frequency deviations within 12.5kHz and 100kHz. An RSSI function with a range of 66dB and 3dB resolution is provided by collecting and rectifying the baseband signal along the 6dB gain stages. The 3dB resolution is achieved through interpolation between the taps of the cascade of 6dB stages.

The receiver current is 2.3mA for a -111dBm sensitivity in FSK (at 25kb/s, 10^{-3} BER, without SAW and Rx/Tx switch). BER measurements are given vs. input power in Figure 3. A plot of the spurious immunity versus frequency offset is given in Figure 4.

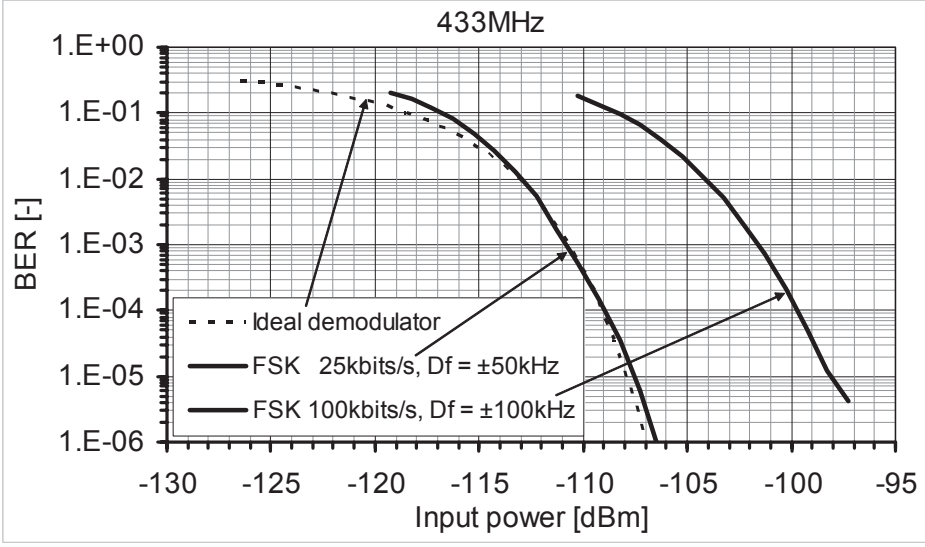


Fig. 3. BER measurement for the receiver.

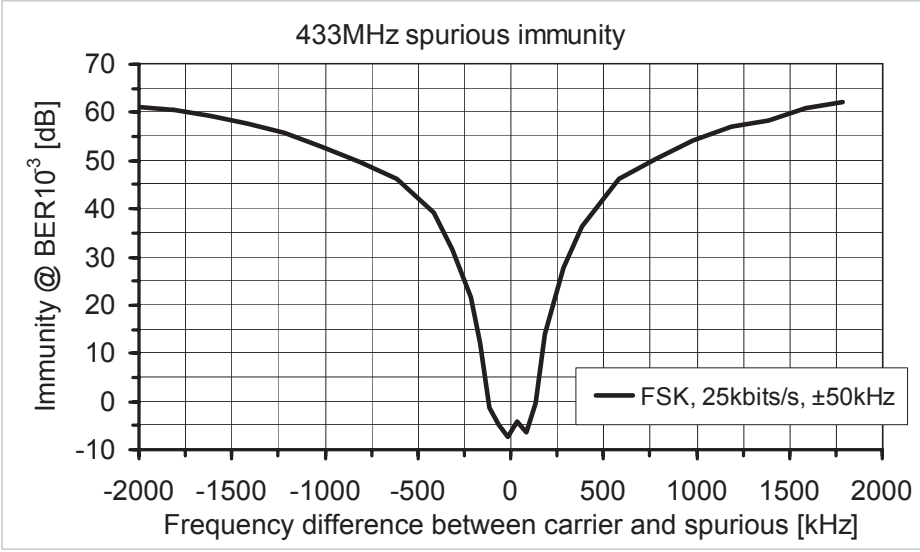


Fig. 4. Plot of the receiver spurious immunity.

2.4. The Transmitter Channel

The transmitter architecture is illustrated in Figure 5.

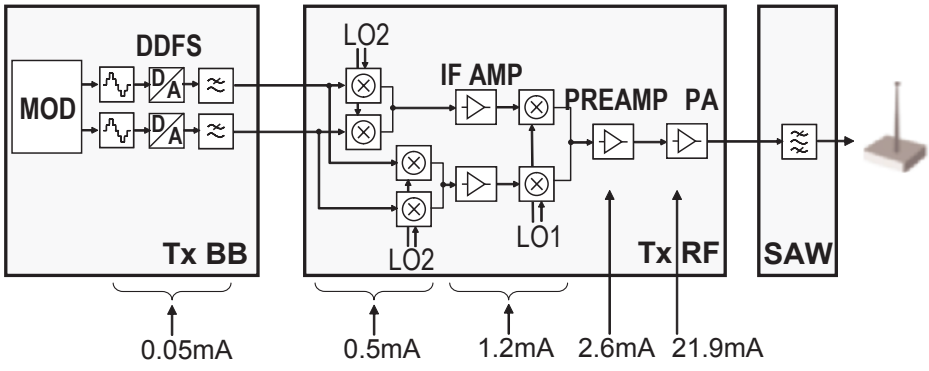


Fig. 5. Overview of the transmitter architecture with the current consumption breakdown.

The transmitter uses the same PLL and LO generation block as the receiver. This has the advantage of allowing for fast Rx/Tx turn-around times, which is an important issue for communication protocols.

The 433MHz transmitter back-end starts with the baseband modulator using a digital sine wave generator. I and Q signals are converted from a digital table containing one period of a sine- and cosine-wave. The transmitter front-end uses a two-step up-conversion to RF, with the same IF value as in the receiver. The first up-converter relies on an I-Q double-balanced mixer whose switching devices are biased in weak inversion to comply with the 1V operation. The input device is an MOS transistor biased in the triode region for maximal linearity, in order to reduce as much as possible the distortion on the baseband signal.

The same I-Q up-conversion principles are used to build the second mixer stage. The baseband signals present the largest possible amplitude for linearity reasons, which means a low overall conversion gain of 0dB with a current consumption of 1.7mA for the entire conversion chain. The up-converted signal feeds into a two-stage pre-amplifier that consists of a first differential stage and a second pseudo-differential stage which drives a differential on-chip resonant network. The pre-amplifier is followed by a class C differential PA. Due to the constant envelope signal of the FSK modulation, a highly nonlinear, high-efficiency transmitter architecture can be used. The PA's transistors present open-drain outputs which are connected to external inductors. The PA yields a maximum efficiency on a load of about 120Ω. An external differential-to-single-ended SAW filter is used to further suppress the out-of-band spurious contents.

The total transmitter current is 27.6mA for a 10dBm output at 1V supply for the 433MHz output. Measurements of the output power and global efficiency (PA and preamplifier) vs. supply voltage are given in Figure 6, and a plot of the output spectrum is given in Figure 7.

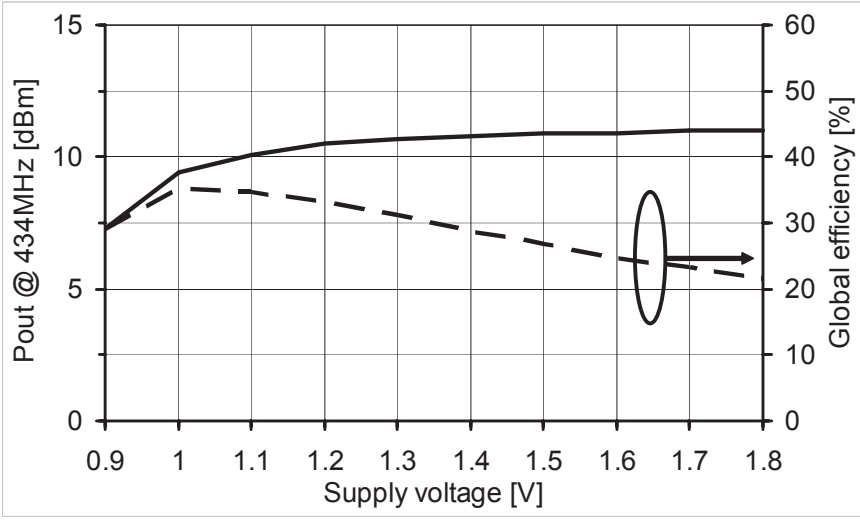


Fig. 6. Transmitter output power and efficiency.

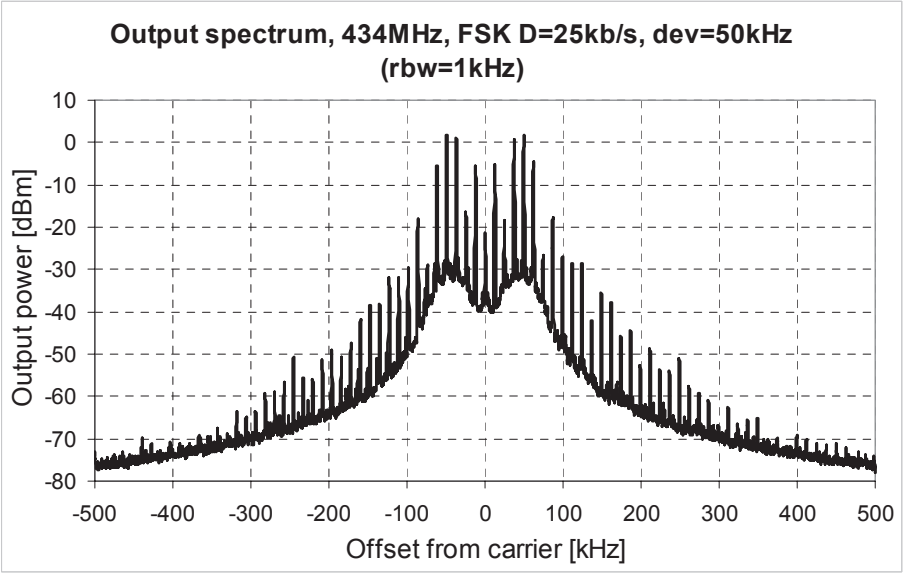


Fig. 7. Transmitter output spectrum.

2.5. The Rx/Tx/Antenna Interface

As illustrated in Figure 8, the RF front-end uses an integrated Rx/Tx switch and drives an external differential SAW filter. The latter reduces the spurious contents in Tx and provides stronger immunity against out-of band interferers in Rx, and rejection of the image frequency. The use of an on-chip Rx/Tx switch structure presents the advantages of saving an external Rx/Tx component in the bill of materials, but also the need of a high-voltage control signal for commuting such external parts.

In receive mode, the RF switches are turned on, whereas the PA output devices are cut off. One output node of the SAW is grounded by the switch on the left, and the signal from the antenna goes through the SAW in single-ended mode. The signal then passes through the switch on the right and the matching network before feeding to the LNA input.

In transmit mode, the RF switches are turned off. The LNA input is thus isolated from the PA output that directly drives the external inductors in differential mode. Because the RF switch on the SAW output is cut off, the SAW device does also the differential to single-ended conversion of the signal before feeding to the antenna.

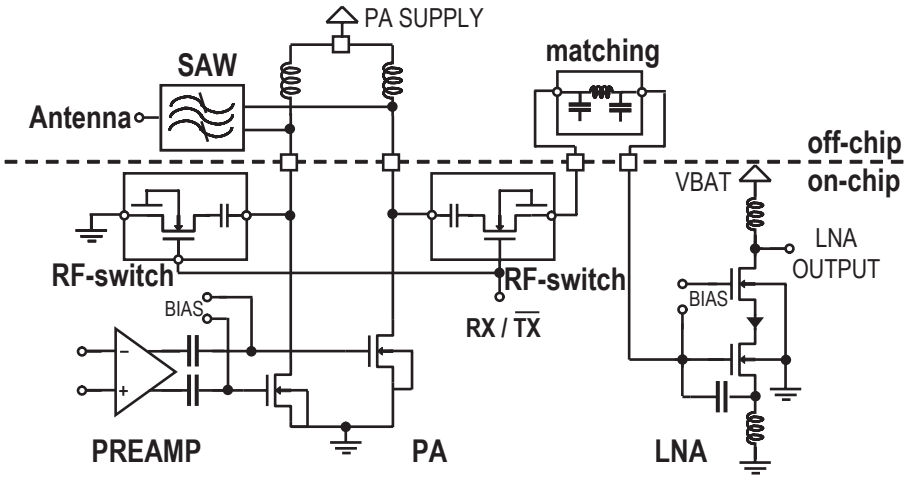


Fig. 8. Overview of Rx, Tx and antenna interface.

The Rx/Tx switch structure presents an insertion loss of 2dB and an isolation of 38dB.

2.6. The Frequency Synthesizer

The architecture of the frequency synthesizer is sketched in Figure 9.

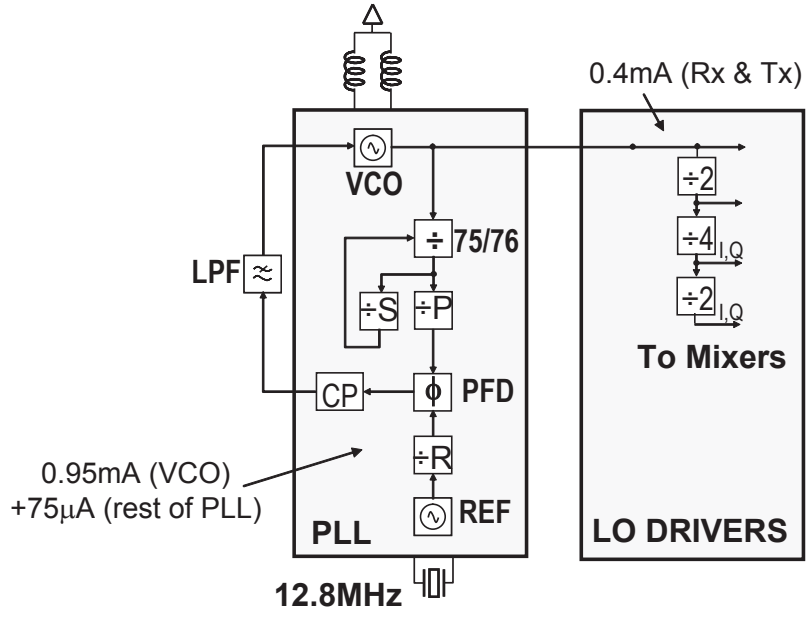


Fig. 9. Overview of frequency synthesizer with the current breakdown for the major blocks.

An integer-N synthesizer architecture has been chosen to generate the different LO signals because it presents a good trade-off between power consumption, simplicity and required settling times for the WiseNET application. The integer-N divider loop is built around a VCO running at f_{VCO}=772MHz. At the output of the VCO, a dual-modulus divider implemented with dynamic latches is used to divide by 75/76 and is combined with a programmable swallow counter that counts up to P, but detects an intermediate state S, where it switches the modulus of the front divider, on the overall providing f_{VCO}=(75(P)+S)fref. The other PLL blocks such as the charge pump and the loop filter are all implemented in a differential manner to achieve a good immunity against supply noise. The loop filter is external so that large capacitors can be used to reduce the PLL in-band noise. The reference signal is provided by a 12.8MHz crystal oscillator followed by a divide-by-R counter. By slightly changing the division ratio (different R, P, S triplet), narrower but unequally spaced steps can be achieved. Within the bands of interest, the maximum step between any two channels over the 433MHz ISM band is narrower than 6.25kHz, although the reference frequency is close to 100kHz.

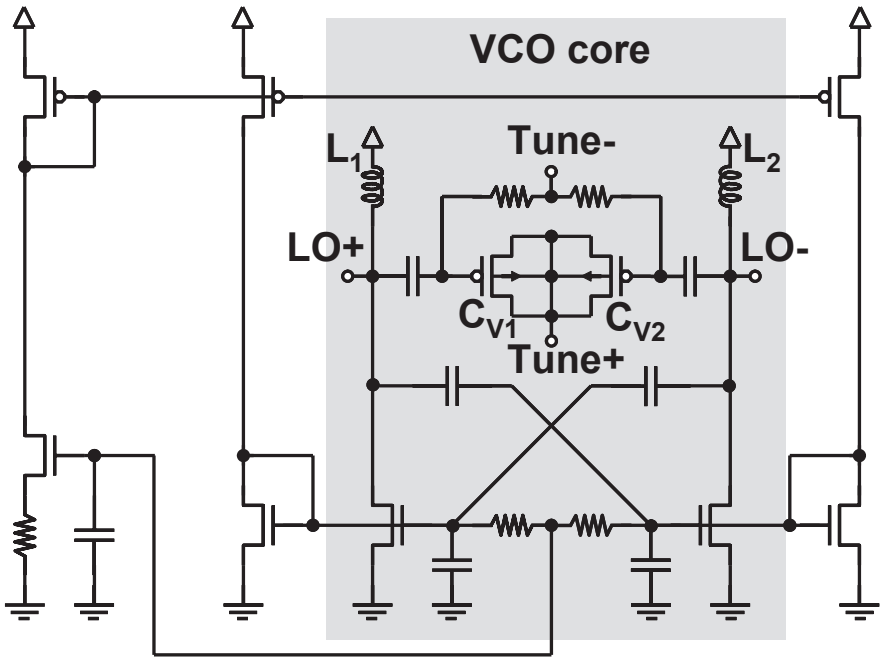


Fig. 10. Schematic of the VCO.

In order to keep the VCO current consumption as low as possible and provide low phase noise, the tank is built with two external high-Q inductors. The variable part of the tank is integrated on-chip with a differential

accumulation MOS varactor yielding a global loaded Q-factor close to 20. To limit further the VCO power dissipation, an amplitude regulation loop is used. A schematic of the VCO is illustrated in Figure 10.

The power consumption of the PLL is 1mA with 950μA for the VCO, the latter generating 1.8V peak-to-peak differential signals. The PLL division chain consumes only 50μA for dividing the f_{VCO}=772MHz LO signal down to the 100kHz reference frequency. A few other microamperes are sufficient for the charge pump and loop filter in lock mode.

The plot on Figure 11 shows a measured spectrum of the PLL. The spur at a reference frequency of 100kHz is lower than -55dBc and the phase noise at 600kHz offset is below -110dBc/Hz.

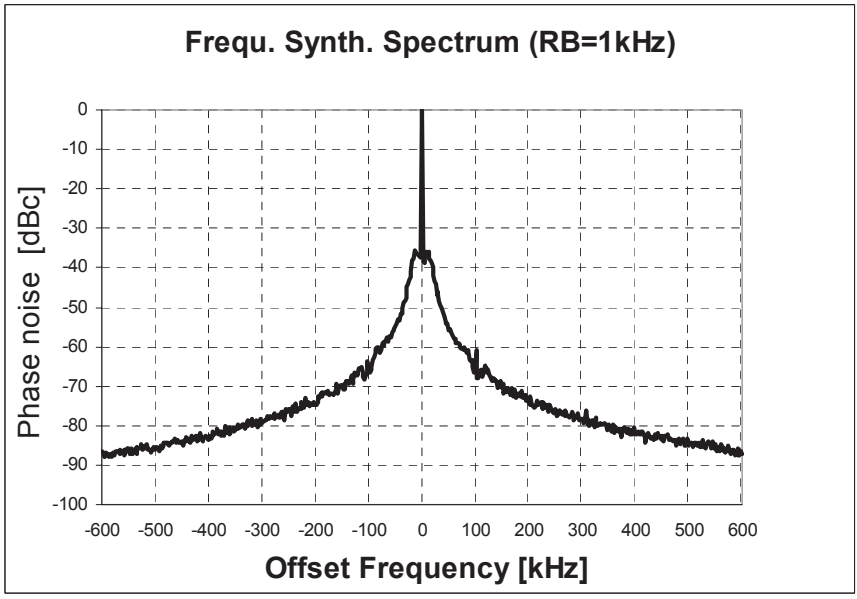


Fig. 11. Plot of the VCO phase noise.

2.7. WiseNET Transceiver Implementation

The main measured characteristics of the transceiver section of the WiseNET SoC are listed in Table 2.

Table 2. Main WiseNET radio measurements.

Parameter	Condition	Value	Unit
General	Typical conditions: Vbat=1.2V, T=25°C (-25°C/+75°C range), FSK D=25kb/s Δf=50kHz		
Supply voltage		1.0-1.6	V
Supply current	Rx: VCO + PLL + Rx	2.3	mA
	Tx: VCO + PLL + Tx, Pout=10dBm	27.6	mA
	Sleep mode	3.5	µA
	Micro-controller & digital blocks	50	µA/MHz
Data rate	Min / typ / max	12.5 / 25 / 100	Kb/s
Frequency deviation	Min / typ / max	12.5 / 50 / 100	kHz
Receiver	(All measurements without SAW & switch)		
Max input signal		10	dBm
RSSI dynamic range	Incl. 14dB AGC & Rx/Tx switch atten.	101	dB
Noise figure		4	dB
IIP3	spurs at 600kHz & 1250kHz	-29	dBm
FSK sensitivity	25kb/s, BER=10^{-3}, BW=100kHz	-111	dBm
Transmitter	(All measurements without SAW, but with switch)		
FSK output power	433MHz ISM band	10.5	dBm
Tx/Rx switch	Insertion loss / Isolation	-2 / -38	dB
PLL			
Phase noise	At 600kHz offset	<-110	dBc/Hz
Synthesis steps	For 100kHz min reference frequency	6.25	kHz
Timing			
VCO startup time	X-tal on, locked at +/-50kHz	350	µs
Channel switching time	VCO locked at +/-50kHz	<200	µs
Rx/Tx turnaround time	Onto same RF channel	<200	µs

The current breakdown in Rx and Tx modes is illustrated in Figure 12.

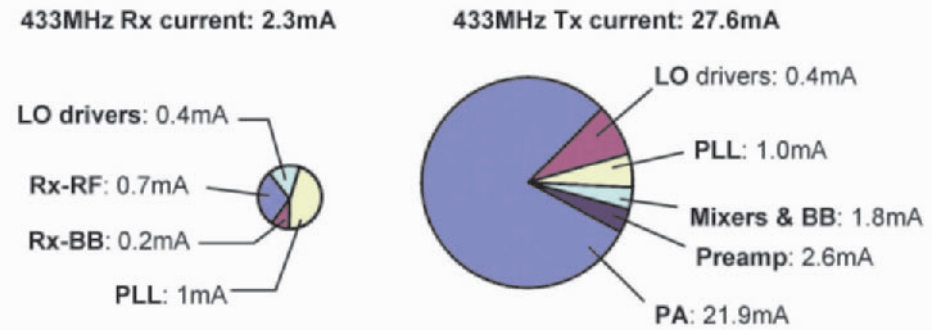


Fig. 12. Current breakdown in receive and transmit modes.

As will be detailed in section 3, the WiseMAC protocol implies that the sensor nodes are expected to be more in receive mode than in transmit mode. Therefore, from a global power efficiency point of view, special care was given to the receiver current consumption rather than to the transmit current. Roughly two-thirds of the current is used by the VCO and the LO drivers,

the remaining third by the RF front end with a fraction for the backend. This was the motivation for choosing an external high-Q inductor, rather than integrating the tank's inductor on-chip, because the degraded quality factor of on-chip inductors would have translated into a much larger PLL current [4], with a direct impact on the Rx current.

For the transmitter, an output power of 10dBm was chosen. This was to achieve the largest possible range knowing that with the WiseMAC protocol and assuming a low average traffic, it would not penalize significantly the average power consumption. This translates in a total Tx current of 25.8mA under 1V. Unlike the receiver, where all blocks, especially the VCO and LO drivers, need to be optimized carefully, the transmitter requires essentially the pre-amplifier and the PA to be designed with care, as they are by far the largest contributors.

A micrograph of the WiseNET SoC is given in Figure 13. The integrated circuit is implemented in TSMC's standard digital 0.18µm CMOS and has a size of 3.3mm by 3.6mm.

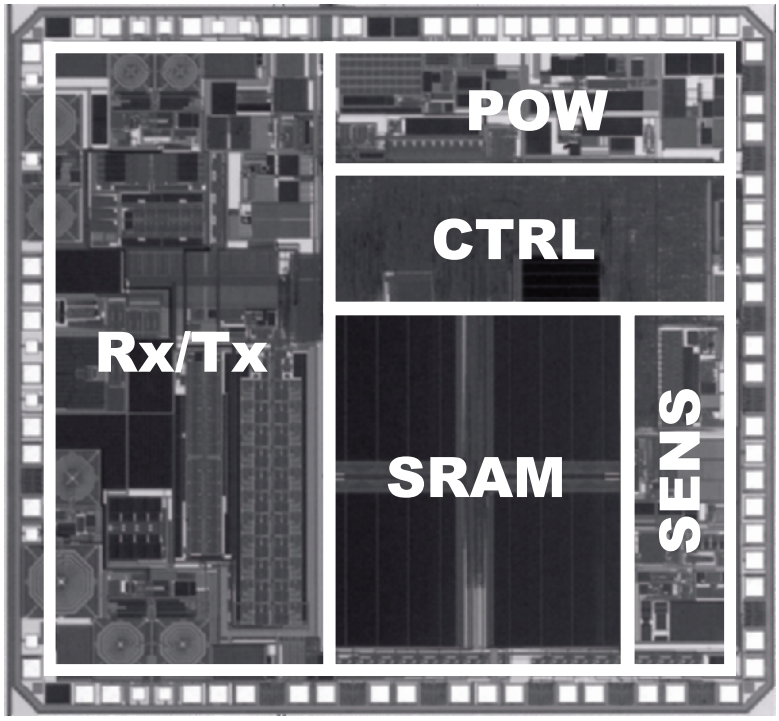


Fig. 13. Die photography.

3. WiseMAC for Low Power Multi-Hop Communication

3.1. Energy Efficient Communication in Sensor Networks

With today's technology, regardless of potential optimizations at the physical layer, the power consumption of a RF transceiver suitable for WSNs is above 1mW in receive mode, and much more in transmit mode. In order to achieve an average power consumption enabling years of lifetime on a low cost battery, the average power consumption should be kept below 100μW, as mentioned in Section 1. It is hence mandatory to turn the radio transceiver off most of the time. A transceiver may not listen to the channel all the time. A duty cycle of a few percent at maximum can be tolerated.

Because the transceiver of the sensor nodes may only be turned on during a small fraction of the time, there is a clear need for algorithms that organize the sensor nodes such that the source and the destination of a communication are turned-on at the same time. The WiseMAC protocol was thus designed in this specific context, and the following sections will highlight its main features and describe how power saving can be done efficiently at the MAC layer.

3.2. Medium Access Control Protocols for Sensor Networks

Because the requirements of WSNs are different from those of wireless LANs and mobile computer networks (MANETs), a specific MAC protocol is required. Such a protocol must minimize idle listening (the time spent listening to an empty channel), overhearing (the time spent listening to traffic addressed to other nodes), collisions and its own overhead (headers and MAC level signaling).

A variety of WSN MAC protocols have been proposed recently, the major proposals can be classified into *scheduled* and *unscheduled* protocols [3] as illustrated in Figure 14.

Scheduled protocols [16][17][18][24][25][26] rely on a network wide synchronization and the use of time division multiple access (TDMA). TDMA appears at first glance as a very appealing protocol for WSNs. It causes neither overhearing nor collisions and sensor nodes may sleep in between assigned communication slots. On the other hand, it is only energy efficient when transporting periodic traffic, while many sensor network applications can be expected to generate sporadic or bursty traffic. Secondly, the setup of the TDMA schedule between nodes can be a complex issue requiring usually a complex protocol implementing some distributed consensus.

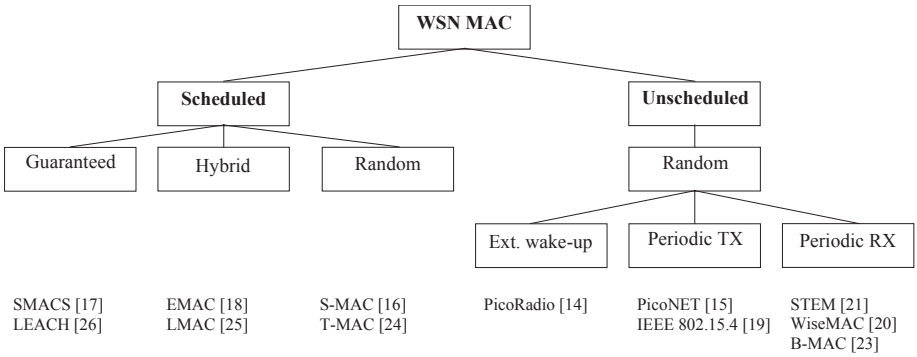


Fig. 14. A classification of WSN MAC protocols.

Unscheduled protocols [14][21][23][15][19] or WiseMAC [20] do not need a network wide synchronization. They use a mechanism to wake-up the destination of a message when needed. Such wake-up mechanisms include the use of an external ultra low-power wake-up hardware, the detection of a message transmission using a form of periodic listening or the announcement of one's availability for packet reception through the periodic broadcast of a beacon.

In the WiseNET project, it was deliberately chosen to design a MAC protocol that is unscheduled, contention based with a wake-up scheme using periodic listening. The choice of a contention protocol was motivated in order to avoid the complex tasks of synchronizing a multi-hop network and of allocating time slots. Many examples such as Ethernet [27] and IEEE 802.11 [28] have shown that they can provide very good performance in most cases. Because the power consumption in transmit mode is most of the time larger than the one in receive mode, the periodic listening approach was preferred for the wake-up scheme to the periodic transmissions in order to minimize the basic power consumption of the protocol [3].

The next section will introduce the WiseMAC protocol, an unscheduled contention protocol using a wake-up scheme based on periodic listening.

3.3. WiseMAC Algorithm

WiseMAC [20][3] is based on the preamble sampling technique. This technique consists in regularly sampling the medium to check for activity. By *sampling the medium*, it is meant listening to the radio channel for the duration required to measure the received power (i.e. a few symbols). All sensor nodes in a network sample the medium with the same constant period. Their relative sampling schedule offsets are independent. If the medium is

found busy, a sensor node continues to listen until a data frame is received or until the medium becomes idle again.

At the transmitter, a wake-up preamble of size equal to the sampling period is added in front of every data frame to ensure that the receiver will be awake when the data portion of the packet arrives. This technique provides a very low power consumption when the channel is idle. The disadvantages of this protocol are that the (long) wake-up preambles cause a throughput limitation and a large power consumption overhead in transmission and reception. The overhead in reception is not only bared by the intended destination, but also by all other nodes overhearing the transmission.

The WiseMAC protocol aims at reducing the length of this costly wake-up preamble. The solution used by WiseMAC consists in learning the sampling schedule of the direct neighbors to use a wake-up preamble of minimized size. This simple scheme provides a significant improvement compared to the basic preamble sampling protocol, as well as to S-MAC [16] and T-MAC [24].

Because the wireless medium is error prone, a link level acknowledgement scheme is required to recover from packet losses. The WiseMAC ACK packets are not only used to carry the acknowledgement for a received data packet, but also to inform the other party of the remaining time until the next sampling time. In this way, a node can keep a table of sampling time offsets of all its usual destinations up-to-date. Using this information, a node transmits a packet just at the right time, with a wake-up preamble of minimized size, as illustrated in Figure 15.

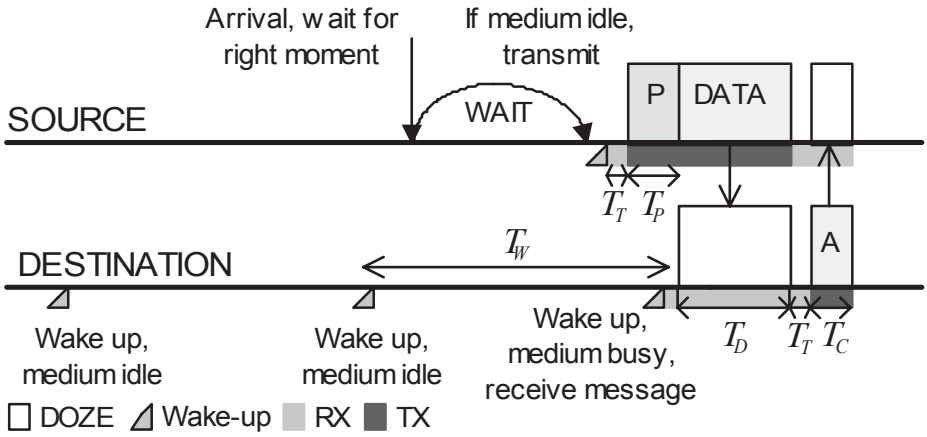


Fig. 15. Principle of operation of WiseMAC.

The duration of the wake-up preamble must cover the potential clock drift between the clock at the source and at the destination. This drift is proportional to the time since the last re-synchronization (i.e. the last time an

acknowledgement was received). The first communication between two nodes will always be done using a preamble as long as the wake-up period. Once some timing information is acquired, a wake-up preamble of reduced size can be used. The length of the wake-up preamble being proportional to the interval between communications, it will be small when the traffic is high. As a result, the per-packet overhead decreases with increasing traffic. In low traffic conditions, the per-packet overhead is high, but the average power consumption caused by this overhead is low. One of the major WiseMAC features is thus to provide a low average power consumption in low traffic conditions and a high energy efficiency in high traffic conditions.

Overhearing is naturally mitigated when the traffic is high, thanks to the combined use of the preamble sampling technique and the minimization of the wake-up preamble duration. As already mentioned, sensor nodes are not synchronized among themselves, their relative sampling schedule offsets are independent. In high traffic conditions, the duration of the wake-up preamble being smaller than the sampling period, short transmission are likely to fall in between sampling instants of potential overhearers.

The synchronization mechanism of WiseMAC can introduce a risk of systematic collision. Indeed, in a sensor network, a tree network topology with a number of sensors sending data through a multi-hop network to a sink often occurs. In this situation, many nodes are operating as relays along the path towards the sink. If a number of sensor nodes try to send a data packet to the same relay, at the same scheduled sampling time and with wake-up preambles of approximately identical size, there are high probabilities to obtain a collision. To mitigate such collisions, it is necessary to add a medium reservation preamble of randomized length in front of the wake-up preamble. The sensor node that has picked the longest medium reservation preamble will start its transmission sooner, and thereby reserve the medium.

Additional schemes, include the repetition of data frames in long preambles to minimize overhearing, the transmission of data packets in bursts thank to “more to come” indication in the packet header, the avoidance of interrupting data-ack transactions through a mandatory inter-frame space larger than the data-ack interval, the selection of a receive threshold well above the noise floor to minimize useless wake-ups and the selection of a carrier sensing threshold below the receive threshold to mitigate the hidden node effect.

3.4. Performance

The performance of the WiseMAC protocol has been studied through simulations, theory and real world implementation.

In order to simulate a low power protocol such as WiseMAC, it is required to implement in the simulation tool a model of the radio transceiver that closely matches the behavior of the real radio transceiver with respect to

the different states that the transceiver can be in and the time and energy spent in the transitions between states. Indeed, with a low power protocol, the energy spent in transitions can be dominating. Figure 16 shows the finite state machine of the implemented radio model.

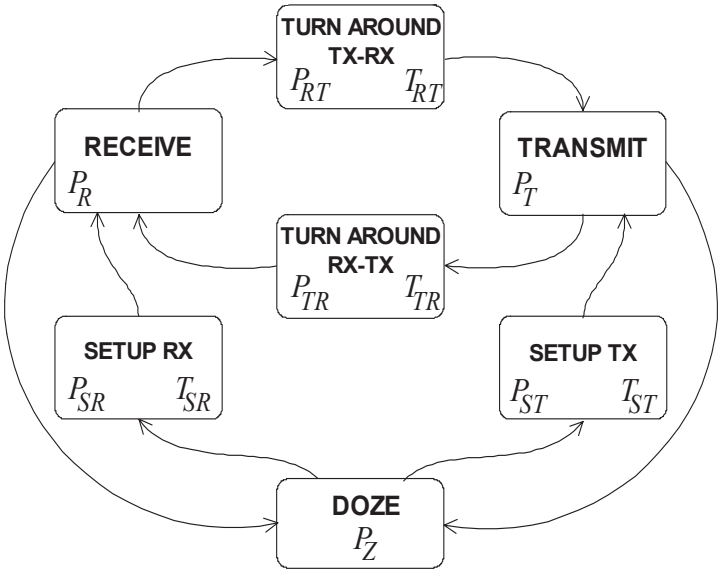


Fig. 16. Radio Layer Model.

In the DOZE state, the radio consumes a current of a few μA only. The radio quartz is turned off. In the SETUP_RX and the SETUP_TX states, the radio quartz and the radio are powering on. When in RECEIVE mode, the radio is able to measure the received signal strength and demodulate incoming data. The radio is in the TRANSMIT state when transmitting data.

The average power consumed in the receive, turn around and transmit states have been defined as $P_R = P_{TR} = P_{RT} = 2.3\text{mW}$ and $P_T = 27.6\text{mW}$ and the turn around times of $T_{RT} = T_{TR} = 200\mu\text{s}$ have been defined according to the data in Table 2. The setup time and energy of a wake-up sequence have been measured, as illustrated in Fig. 17, through observation of the consumed current until a valid RSSI measurement can be started. The power on phase into RX needs $T_{SR} = 1.8\text{ms}$ and the required power integrates to approximately 1.7μJ, resulting in an average power of $P_{SR} = 0.9\text{mW}$. Because the final stage power amplifier in transmit mode is only powered on when effectively transmitting data bytes, the same values apply for T_{ST} and P_{ST}.

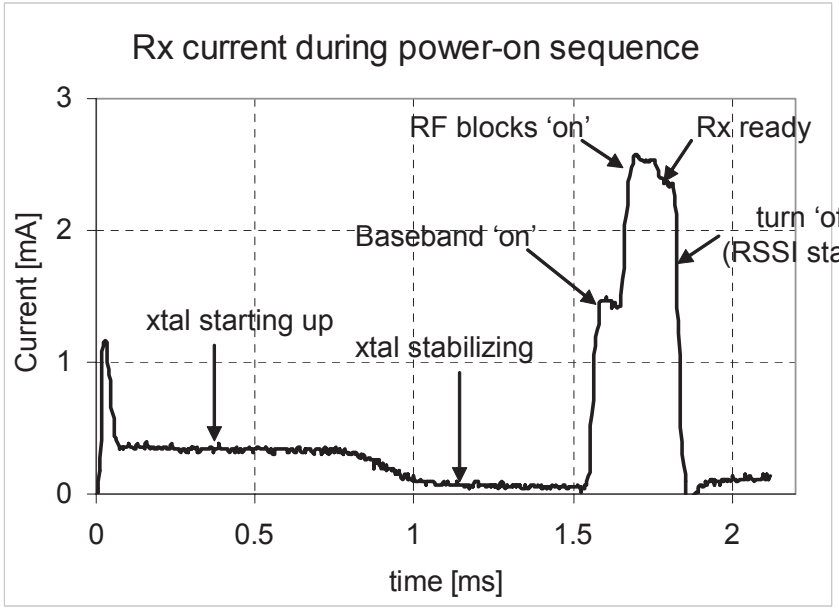


Fig. 17. Current consumption of the WiseNET SoC when powering on into RX, until RSSI is stable, then turn off.

As a comparison point, some measured parameters of the WiseNET solution are compared in Table 3 with those of an off-the-shelf state-of-the-art XE1203 transceiver combined with the XE88LC06A microcontroller from Semtech [29][30][31], which was also used within the WiseNET project for implementing the WiseMAC protocol. In the case of the wake-up sequence illustrated in Figure 17, the energy gain reaches roughly a factor of 30 with the WiseNET SoC.

Table 3. Comparison of some parameters of the WiseNET solution vs a state-of-the-art off-the-shelf XE1203 solution.

Parameter	WiseNET SoC	XE1203F & XE88LC06A
Supply voltage	1V	3V
Rx Power	2.3mW	45mW
Tx Power	27.6mW	32.3mW
Rx wake-up sequence	1.7μJ	50μJ

Based on the radio layer model of Figure 16, the WiseMAC protocol has been modeled using the Glomosim simulation platform. To permit comparison, the S-MAC [16] and T-MAC [24] protocols were modeled as well. S-MAC (Sensor-MAC) is a well-known MAC protocol specifically designed for sensor networks. This protocol can be seen as a lower power and lower traffic version of the power save mode defined by the IEEE 802.11 standard [28] for ad-hoc networks. T-MAC is an improvement of the S-MAC protocol providing support for a higher throughput.

Figure 18 shows the average power consumption of a node forwarding packets in a multi-hop network as a function of traffic for different protocols. A lattice topology has been used with Poisson traffic flowing in parallel. The average power consumption is computed for a central node. Traffic flowing through neighbors is creating interference and overhearing overhead to this central node (see [20][3] for more details). The circles show the average power consumption that is consumed when using the WiseMAC protocol with the WiseNET SoC. Each circle is the result of a simulation; at the end of which the average power consumption is computed through recording of the time spent in each state of the radio transceiver model. The dotted line along these circles represents the theoretical power consumption of WiseMAC (see [3]). A good match can be observed. With very low traffic (1 packet forwarded every 1000s), the power consumption goes below 20μW. In this area, the power consumption of the wake-up scheme, i.e. preamble sampling, is dominating. When the traffic increases, the power consumption with WiseMAC increases as well. A power consumption of about 100μW is observed for a traffic of 1 packet forwarded every 10s. It is interesting to observe that the power consumption of WiseMAC approaches, in the high traffic region, the one of the ideal protocol. With the ideal protocol, a node wakes up "by magic" whenever a packet needs to be received. The power consumption of the ideal protocol has been computed mathematically disregarding any collisions.

A comparison with S-MAC and T-MAC shows that these two protocols can either provide a low power consumption (with a duty cycle of 1%) or provide the support of a high traffic (with a duty cycle of 10%) while WiseMAC supports both.

An upper bound on the power consumption is given by the CSMA/CA protocol with which a transceiver is either in receive or transmit mode and therefore consume at least the power $P_R = 2.3\text{mW}$.

As a conclusion, it can be observed that WiseMAC can provide an ultra-low power consumption in low traffic conditions, can support a relatively high traffic, and that it becomes energy efficient when traffic increases.

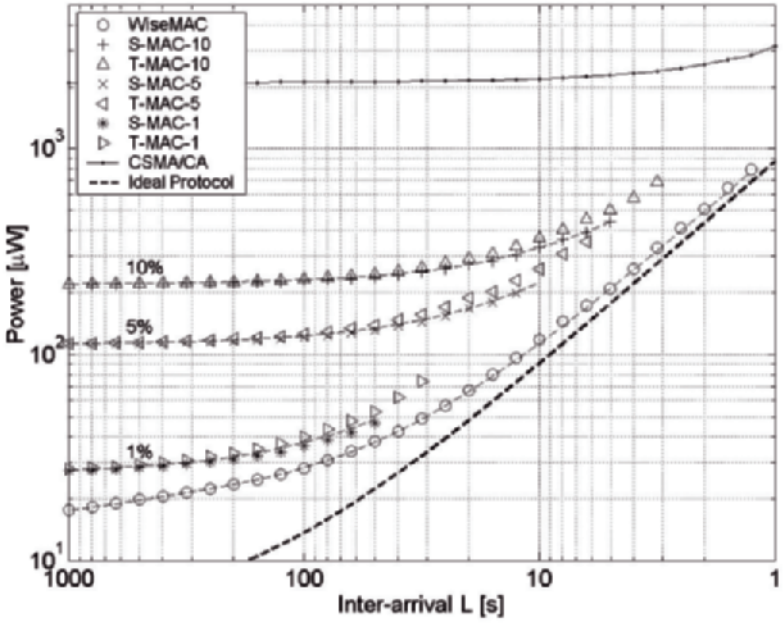


Fig. 18. Average power consumed as a function of traffic.

The IEEE 802.15.4 standard has recently been defined for personal area networks, targeting low cost and lower throughput than Bluetooth. The ZigBee alliance has specified routing protocols enabling the formation of multi-hop networks based on IEEE 802.15.4 devices, making IEEE 802.15.4/ZigBee a potential basis for the implementation of multi-hop sensor networks. In the proposed routing protocol, every router (i.e. relay) must be IEEE 802.15.4 coordinator. Power saving in IEEE 802.15.4 is made possible through the periodic transmission of a beacon. Downlink traffic in announced in the beacons and uplink traffic can be transmitted right after the beacon. A coordinator can sleep after having sent a 19 bytes long beacon and waited in receive state during 55 bytes for potential uplink traffic. The devices of a coordinator can sleep in between receptions of the beacons.

Figure 19 shows the average power consumption of a relay node receiving a packet of 32 bytes every 30s and forwarding it further in a multi-hop network. The power consumption is computed theoretically for WiseMAC, S-MAC and IEEE 802.15.4, assuming the use of the same underlying hardware (the WiseNET SoC), and the use of the same wake-up period of 250ms for all protocols. The wake-up period is the sampling period with WiseMAC, the frame duration with S-MAC and the beacon period with IEEE 802.15.4. It is chosen equal for all three protocol to provide comparable hop delays.

It can be observed that WiseMAC provides a power consumption of only 27µW, enabling multiyear lifetime on a single alkaline battery. With S-MAC, the power consumption if 5 times higher, due to the higher overhead caused by periodically listening during the duration of a few bytes, instead of a few symbols with WiseMAC. With IEEE 802.15.4 ZigBee, the power consumption is 36 times higher because a node must periodically receive a beacon from its parent, periodically transmit a beacon to its children and wait for potential uplink traffic from its children.

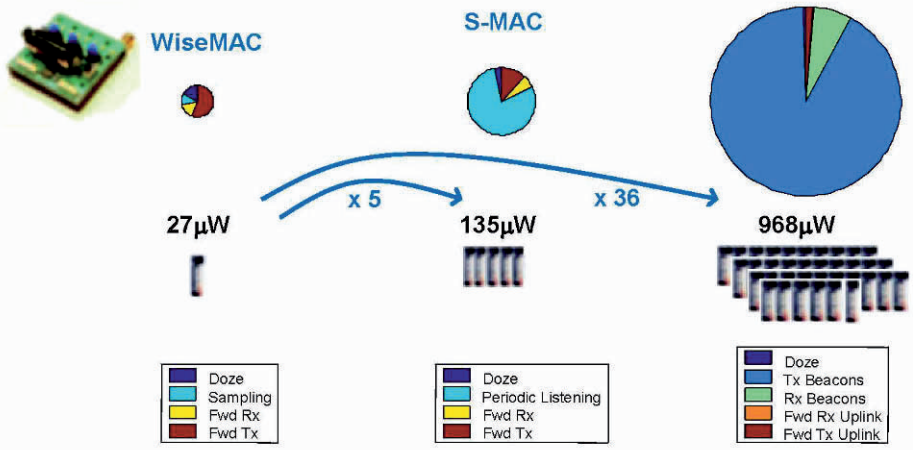


Fig. 19. Average power consumption when forwarding a 32 bytes packet every 30s with WiseMAC, S-MAC and 802.15.4 ZigBee, assuming the use of the WiseNET SoC.

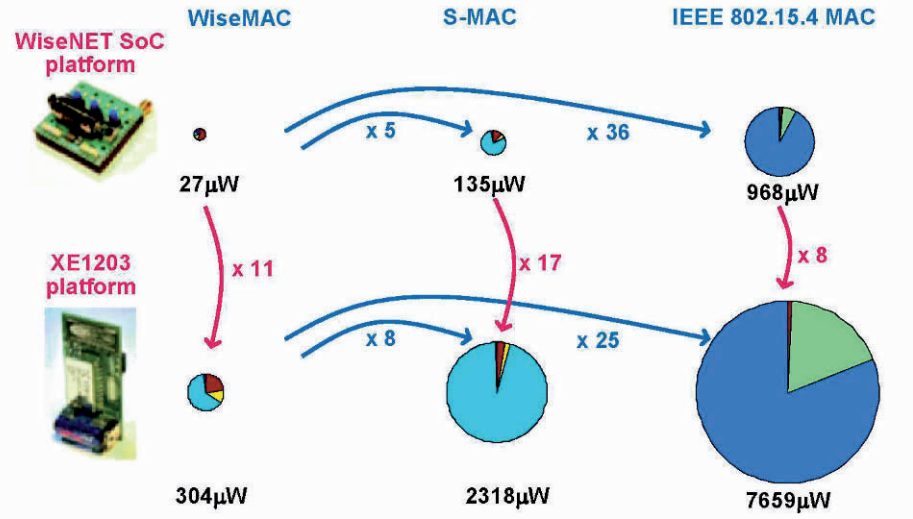


Fig. 20. Average power consumption when forwarding a 32 bytes packet every 30s with WiseMAC, S-MAC and 802.15.4 ZigBee, for the WiseNET SoC platform compared with the XE1203 platform.

When including the parameters of an off-the-shelf XE1203 transceiver [29][30][31] in the comparison, as illustrated in Figure 20, the power consumption calculated with IEEE 802.15.4 ZigBee is approximately 250 times larger than with the WiseNET SoC running WiseMAC.

4. Experimentation

The WiseNET protocol stack has been implemented on the WiseNET SoC following the structure shown in Figure 21.

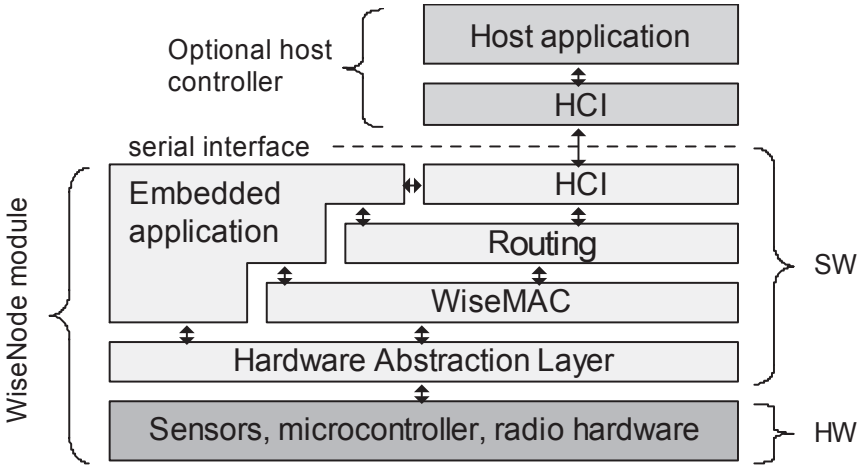


Fig. 21. The WiseNET protocol stack.

The hardware abstraction layer is a set of drivers that can be used by the upper layers to access radio, timer subsystem, serial and sensor interfaces. The communication stack is composed of the WiseMAC protocol and of the routing layer. Depending on the computing resources requirements of the application, it is possible to have either the application embedded in the SoC, or running on a second microcontroller. In the latter case, communication with the host microcontroller is done through a serial interface (SPI or UART) using a packet based HCI (Host Controller Interface) protocol. In the second case, the WiseNET SoC becomes a simple-to-use low power multi-hop communication modem.

In its current state, the WiseNET protocol stack uses about 19kB, leaving 3.6kB for the application, as illustrated in Figure 22. From these 19kB, the MAC layer accounts for 5kB and the routing layer for 1.3kB. The rest is needed for drivers and low-level “gcc” libraries.

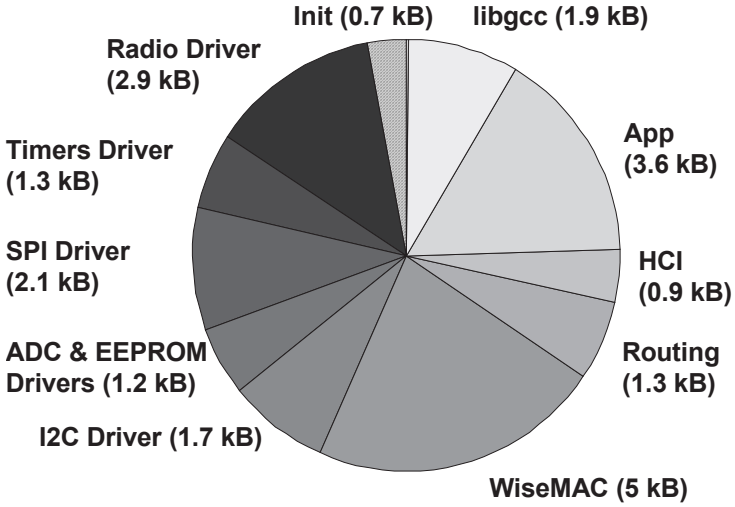


Fig. 22. The WiseNET protocol stack's module sizes.

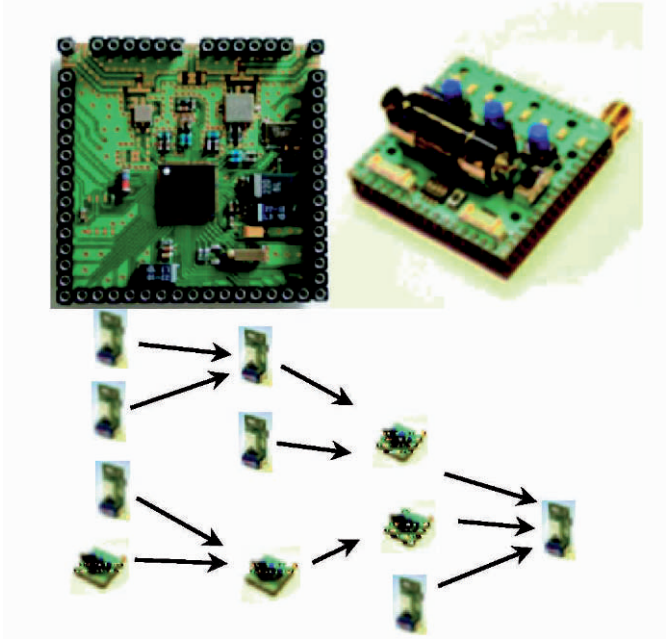


Fig. 23. Wireless Sensor Node using the WiseNET SoC (left, middle) and example test network topology.

A test network has been deployed using wireless sensor nodes based on the WiseNET SoC as illustrated in Figure 23 (left). The lower part includes the SoC, crystals, filters and decoupling capacitors. The upper part holds the battery and the user interface (LEDs, buttons, switches). In order to have a sufficient number of nodes from the very start of the project, a second type of nodes [29][30] with over-the-air compatibility has been designed using the off-the-shelf XE88LC06A controller and the XE1203 radio from Semtech [31]. Source code compatibility between the two platforms has been obtained thank to the definition of the hardware abstraction layer. Test network with multi-hop topology have been deployed at CSEM using a mix of both platforms. The average power consumption of the radio transceiver of a node could be computed through the measurement by every node of the amount of time spent with the radio in the receive and in the transmit states. It was possible to confirm experimentally the power consumption predicted by theory (see [3] for more details).

5. Conclusions

This paper addresses the different issues faced when designing ultra low-power WSN with a particular emphasis on the radio hardware that was co-designed with a dedicated MAC protocol.

The WiseNET transceiver SoC was designed and specifically optimized for the new WiseMAC protocol specially developed for WSN. It has been integrated on a standard digital 0.18μm deep-submicron CMOS process. It runs from a single 1.5V battery and operates down to 1.0V while consuming only 2.3mW in receive mode. It achieves a −108dBm sensitivity for 25kb/s FSK datarate with a 10^{-3} BER. The transmit power is 10dBm for a current consumption of 27.6mA under 1.0V. In addition to this low-power radio, the WiseNET SoC also includes all the functions required for data acquisition, processing and storage of the information provided by the sensor.

The WiseNET solution comprising the WiseNET SoC together with the WiseMAC protocol consumes more than an order of magnitude less power than comparable solutions available today, using for example the IEEE 802.15.4 standard.

Acknowlegments

This work was partly supported by industrial partners, both on the technical and financial sides. It was also partly financed by the NCCR-MICS program of the Swiss National Fund. The authors also wish to acknowledge Gaston Laurens, Thierry Porcheron, Thierry Scordilis and Gilles Durant from Hager Research for leveraging their expertise on real-world low-power radio

solutions and contributing to the WiseNET SoC specification, design, and characterization. The Semtech design team, with Stéphane Bories, Jean Golta, Philippe Graber, Clovis Lapaire, Philippe Mosch, Michel Moser, Olivier Nys, and Sylvio Todeschini, is also aknowledged for the implementation and test of the CoolRisc system, the sensor interface, the power management unit, and the assembly of the WiseNET SoC.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: A survey. IEEE Communications Magazine, 40(8):102-114, Aug. 2002.

[2] Paul J.M. Havinga. Mobile Multimedia Systems. PhD thesis, University of Twente, Feb. 2000.

[3] Amre El-Hoiydi, Energy efficient medium access control for wireless sensor networks, PhD Thesis, EPFL, no 3285 (2005)

[4] D. Ruffieux et al., "A 1.2mW RDS Reciever for Portable Applications," IEEE J. Solid State Circuits, vol 39(7), July 2004, pp 995-1005.

[5] P. Jacobs et al., "A 0.35µm CMOS fractional-N Transmitter for 315/433/868/915 MHz ISM Applications", Proc. IEEE European Solid State Circuits Conf., Sept 2003, pp. 425-428

[6] A.-S. Porret et al., "An Ultralow-Power UHF Transceiver Integrated in a Standard Digital CMOS Process : Architecture and Receiver," IEEE J. Solid State Circuits, vol. 36, no. 3, 2001, pp. 452-466

[7] T. Melly et al., "An Ultralow-Power UHF Transceiver Integrated in a Standard Digital CMOS Process : Transmitter," IEEE J. Solid State Circuits, vol. 36, no. 3, 2001, pp. 467-472

[8] C. Enz et al., "An MOS Transistor Model for RF IC Design Valid in All Regions of Operation," IEEE Trans. Microwave Theory and Techniques, Jan. 2002, pp. 342-359

[9] "CoolRISC816 8-bit Microprocessor Core Databook", Semtech Inc, see http://www.semtech.com

[10] S. Cserveny et al., "Noise Margin in Low Power SRAM Cells," in Proc. PATMOS Conf. Sept 2004

[11] H. Darabi et al., "A 4.5-mW 900MHz CMOS Receiver for Wireless Paging," IEEE J. Solid State Circuits, vol. 35, no.3, pp. 1085-1096

[12] T. Melly et al., "An Analysis of Flicker Noise Rejection in Low-Power and Low-Voltage CMOS Mixers," IEEE J. Solid State Circuits, vol 36, No 1, pp. 102-109, January 2001

[13] T. Melly, "Conception d'un Emetteur-Récepteur à Faible Consommation Integré en Technologie CMOS," Ph.D. dissertation, Swiss Federal Intitute of Technology, Lausanne (EPFL), 2000

[14] J.M. Rabaey, M.J. Ammer, J.L. da Silva Jr., D. Patel, and S. Roundy. PicoRadio supports ad hoc ultra-low power wireless networking. IEEE Computer, 33(7):42–48, July 2000.
[15] F. Bennett et al. Piconet: embedded mobile networking. IEEE Journal on Personal Communications, 4(5):8-15, Oct 1997.
[16] W. Ye, J. Heidemann, and D. Estrin. An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In Proc. IEEE INFOCOM Conf., 2002.
[17] K. Sohrabi, J. Gao, V. Ailawadhi, and G.J. Pottie. Protocols for self-organization of a wireless sensor network. IEEE Personal Communications, 7(5):16-27, Oct. 2000.
[18] L.F.W. van Hoesel, T. Nieberg, H.J. Kip, and P.J.M. Havinga. Advantages of a TDMA based, energy-efficient, self-organizing MAC protocol for WSNs. In Proc. IEEE 59th Vehicular Technology Conference, VTC 2004-Spring., volume 3, pages 1598-1602, May 2004.
[19] 802.15.4 LAN MAN Standards Committee of the IEEE Computer Society. IEEE 802.15.4, Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for Low Rate Wireless Personal Area Networks (LR-WPANS). IEEE, 2003.
[20] A. El-Hoiydi and J.D. Decotignie. WiseMAC: An Ultra Low Power MAC Protocol for Multi-hop Wireless Sensor Networks. In Proceedings of the First International Workshop on Algorithmic Aspects of Wireless Sensor Networks (ALGOSENSORS 2004), Lecture Notes in Computer Science, LNCS 3121, pages 18-31. Springer-Verlag, July 2004.
[21] C. Schurgers, V. Tsiatsis, and M.B. Srivastava. STEM: Topology Management for Energy Efficient Sensor Networks. In Proc. IEEE Aerospace Conf., volume 3, pages 1099-1108, Mar 2002.
[22] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: A survey. IEEE Communications Magazine, 40(8):102-114, Aug. 2002.
[23] Joseph Polastre, Jason Hill, and David Culler. Versatile Low Power Media Access for Wireless Sensor Networks. In Proc. Second ACM International Conference on Embedded Networked Sensor Systems (SenSys'04), Nov. 2004.
[24] Tijs van Dam and Koen Langendoen. An adaptive energy-efficient MAC protocol for wireless sensor networks. In Proc. First ACM International Conference on Embedded Networked Sensor Systems (SenSys'03), pages 171-180, Nov. 2003.
[25] L.F.W. van Hoesel and P.J.M. Havinga. A Lightweight Medium Access Protocol (LMAC) for Wireless Sensor Networks: Reducing Preamble Transmissions and Transceiver State Switches. In Proc. INSS, Japan, June 2004.

[26] Wendi Rabiner Heinzelmann, Anantha Chandrakasan, and Hari Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In Proc. Hawaii International Conference on System Sciences, Jan. 2000.
[27] LAN MAN Standards Committee of the IEEE Computer Society. 802.3 - Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specications. IEEE, 2002.
[28] LAN MAN Standards Committee of the IEEE Computer Society. 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Speci¯cations. IEEE, 1999.
[29] A. El-Hoiydi, V. Moser, J.-D. Decotignie, P. Volet, “Implementation of an Ultra Low-Power Self-Organizing Multi-Hop Wireless Sensor Network”, 2004 CSEM Scientific Report, p. 74.
[30] A. El-Hoiydi, J.-D. Decotignie, Poster Abstract: Low Power Multi-Hop Sensor Networking using WiseMAC, EWSN 2006.
[31] http://www.semtech.com

ULTRA-LOW POWER FREQUENCY-HOPPING SPREAD SPECTRUM TRANSMITTERS AND RECEIVERS

Emanuele Lopelli*, Johan van der Tang†,
Arthur H.M. van Roermund*
*Mixed-signal Microelectronics Group, Eindhoven University of Technology, Eindhoven, The Netherlands
†Semiconductor Ideas to the Market (ItoM), Eindhoven, The Netherlands

Abstract

This paper examines system and circuit design techniques for a "micro-Watt node" operating at a power level low enough to enable the use of an energy scavenging source. Despite several architectures have been investigated in order to reduce the overall system power consumption, none of them is able to guarantee robustness of the link and ultra-low power consumption at the same time. A survey of the most advanced architectures meant for ultra-low power transceivers is described. Advantages and drawbacks of all these systems are discussed and the reasons for an architecture based on Frequency-Hopping (FH) Spread-Spectrum (SS) are discussed. Finally a novel FH synthesizer based on a digital pre-distortion architecture is proposed in order to reduce the power consumption of the hopping synthesizer. The FH architecture together with a frequency offset robust demodulation technique allows a reduction by a factor 8 of the power consumption compared to the state-of-the-art synthesizers. Furthermore, a single RF block front-end is obtained combining the VCO and the PA. The novel RF front-end can be directly coupled to the antenna through a balun and the system is able to deliver -18 dBm output power on a 50 Ω load at 1 mA current consumption (2 V power supply). To prove the new synthesizer principle a communication link in the 902-928 MHz ISM band has been set-up. The receiver, mainly software with a flexible RF front-end, adopted a ST-DFT demodulation algorithm and achieved a BER smaller than 1.1% at -25 dBm output power, with TX and RX antennas placed at 8 meters distance in a NLOS condition and in a common office environment.

1. Introduction

In recent years, a trend toward a world in which people will be surrounded by networked devices that are sensitive and adaptive to their needs can be foreseen. This trend has been expressed in a vision called

A.H.M. van Roermund et al. (eds.), Analog Circuit Design, 377–411.

Ambient Intelligence (AmI). It is possible to partition this world into three different classes of devices called "Watt nodes","milli-Watt nodes" and "micro-Watt nodes" [1].

The "Watt nodes" and the "milli-Watt nodes" demand a further improvement in technology scaling to meet the low-power target. In contrast, the design of a "micro-Watt" node requires to meet the limit of miniaturization, cost reduction and power consumption. Therefore, the complexity of this task is not in the number of transistors but in the capability to optimally combine technologies, circuit and protocol innovation to obtain the utmost simplicity of the wireless node.

At the system level there are several challenges engineers have to face to achieve such a reduction in terms of power consumption. Like in a paging channel, a low-power wake-up circuitry will be required together with RSSI measurement circuits, so that energy expenditure can be optimally adjusted. Finally, advances in energy sources and scavenging techniques are mandatory to avoid battery replacement.

2. Energy Scavenging Techniques

Several scavenging techniques have been studied in the recent years. However it is unlikely that a single solution will satisfy the total application space. For example a solar cell requires minimum lighting conditions, a piezoelectric generator sufficient vibration, and a Carnot-based generator sufficient temperature gradient.

One of the most common scavenging techniques is to harvest energy from an RF signal. An electric field of 1 $\mathrm{V/m}$ yields only 0.26 $\mu\mathrm{W/cm^2}$, but such field strengths are quite rare [2]. This technique is generally used for RFID tags, which have a power consumption between 1 and 100 μW. Energy can be harvested by using solar cells. While 1 $\mathrm{cm^2}$ of standard solar cells produces around 100 mW under bright sun, it only generates no more than 100 μW in a typically illuminated office [2]. Also thermoelectric conversion can be used as an energy scavenging technique. Unfortunately the Carnot cycle limits the use of this technique for small temperature gradients by squeezing the efficiency below 5% for about 15 degrees temperature difference [2]. Another possible solution to the scavenging problem can be found using vibrational energy. If 1 $\mathrm{cm^3}$ volume is considered, then up to 4 μW power can be generated from a typical human motion, whereas 800 μW can be harvested from machine-induced stimuli [2].

Considering the aforementioned state-of-the-art in energy scavenging techniques, researchers have plenty of room for improvement both in increasing energy harvesting efficiency and in reducing the overall power consumption of wireless nodes.

3. Low-power Wireless Systems Trends

Increasing demands in wireless systems for sensing and monitoring applications have culminated in a high demand for low power autonomous devices. While AmI and the autonomous node concepts still are far from an everyday reality, a lot of effort has been put in the elaboration of novel standards, which can better match the low power trend. Following this novel trend, standards like Bluetooth and Zigbee have been proposed in the recent years.

Whereas these standards have reduced power consumption at expenses of a lower data-rate and Quality-of-Services (QoS), their protocol and their physical layer are still too complex to be implemented in an autonomous wireless device. Indeed, even though recently a simplified Bluetooth Low End Extension (LEE) [3] has been released, the overall transceiver power consumption remains too high for an autonomous node. While the data-rate has been scaled down to less than 0.5 Mbps and other constraints have been simplified both at the physical and MAC layer, a state-of-the-art current consumption of 9 mA (at -2 dBm transmitted power) in Tx mode and 8 mA in Rx mode have been reported for a Bluetooth transceiver [3].

In the effort to reduce the transceiver power consumption, the Zigbee standard has been recently released (Preliminary standard by Philips in 2001). Whereas the data-rate is reduced with respect to the Bluetooth standard, the overall transceiver power consumption remains high due to complex MAC and physical layers. A recent work [4] showed a Zigbee compatible radio transceiver, with 21 mW power consumption in Rx mode and 30 mW in Tx mode (at 0 dBm output power). It follows that these standards cannot be used in the design of a "micro-Watt node" where an average power consumption of less than 100 μW and therefore, a peak power consumption lower than 10 mW, is required for duty-cycles lower than 1%.

4. Recent Advances in Ultra-Low Power Transceivers

Starting with university research, the interest in ultra-low power wireless devices has increasingly spread among companies as well. In the wide scenario of ultra low-power devices, various pioneering investigations have been conducted to prove the feasibility of a "micro-Watt node" in terms of power consumption and robustness of the communication link.

4.1. Research in Industries

Several wireless products, which claim to be ultra-low power, are present on the market. Rarely these products can be used as core block for a "micro-Watt node". Pioneering researches toward the development of this kind of wireless nodes can be found in [5][6].

The Eco node [5] has been designed to monitor the spontaneous motion of preterm infants using the 2.4 GHz ISM band at 1 Mbps data-rate. While showing a good form factor (648 mm^3 by 1.6 grams), its power consumption is still far away from the minimum target required by an autonomous node. Indeed, even at 10 kbps and -5 dBm output power, it consumes 20.4 mW in Tx mode and 57 mW in Rx mode (at 1 Mbps) considering only the radio device. Whereas these nodes are designed for duty-cycled operation, as stated in Section 3, peak power consumption should not exceed 10 mW. Robustness of the link by frequency diversity is achieved by using a Frequency-Hopping Spread Spectrum (FHSS) technique.

The Telos node [6] complies with the ZigBee standard. As a result, while having a reduced data-rate (250 kbps), it has an overall power consumption of around 73 mW from 1.8 V power supply at 0 dBm transmitted power.

4.2. Research in Universities

Different universities are involved in pioneering research on ultra-low power devices and networks. At Berkeley university an ultra-low power MEMS-based transceiver has been developed [7]. Whereas using a 1.9 GHz carrier frequency and only two channels, the receiver power consumption is 3 mA from a 1.2 V power supply. The data rate is 40 kbps at 1.6 dBm output power. The low receiver power consumption is mainly

obtained by using a high-Q MEMS resonator implemented as a Thin-Film Bulk Acoustic Resonator (FBAR). If more channels are needed, like in the case of an FHSS transceiver, the hardware requirement will increase linearly with the number of channels, making this choice impractical from a low-power point of view. The transmitter part adopts direct modulation of the oscillator and MEMS technology, eliminating power hungry blocks like PLL and mixers, therefore reducing the overall power consumption.

Two major drawbacks can be foreseen in the proposed architecture. While reducing the circuit and technological gap toward a "micro-Watt node", it relies on non-standard components (MEMS), which will increase the cost and will require higher driving voltage. Furthermore it lacks on robustness due to the use of only two channels, while requiring a linear increase of the power consumption with the channels number, if a more robust frequency diversity scheme has to be implemented.

At the CSEM institute the WiseNet [8] project aims to optimize both the MAC and the physical layer to obtain a robust, low-power solution for sensor networks. Whereas not using the worldwide available 2.4 GHz ISM band but the lower 433 MHz ISM band, it achieves a power consumption of only 1.8 mW from a minimum supply voltage of 0.9 V in Rx mode. This result was achieved by a combination of circuit and system innovative techniques and the use of the low-frequency 433 MHz band, which reduces the power consumption of the most power hungry blocks like the frequency synthesizer. In Tx mode a high power consumption of 31.5 mW was reported mainly due to the choice of a high output power of 10 dBm. Data-rate is 25 kbps with FSK modulation.

The proposed solution, while relying partially on the frequency band choice to reduce the power consumption, still requires external components like high-Q inductors for the LC-tank circuit and external RF filters, which will deteriorate form-factor and power consumption at higher frequencies.

5. System Level Aspects for Wireless Microsensor Nodes

At system level, several trade-offs are present in order to optimize the overall power consumption of wireless nodes. Several power models

have been proposed in order to predict the optimum data-rate, which can minimize the overall power consumption. In [9] a simple energy model has been developed showing that, while increasing the data-rate and the efficiency of the power amplifier can reduce the energy per bit required to send the information, the inefficiency introduced by the short packet size (common in low data-rate applications) demands for a reduction in the start-up time.

Another simple model has been developed in [10]. The model considers both transmission and reception of data-packets, while considering PA efficiency and power consumption during transient times. From this model a fixed energy cost due to the oscillator power consumption and the receiver circuitry dissipation can be reduced by increasing the data-rate. Considering the energy per bit, results show that the optimum data-rate should range between 10 and 100 Mbps. In [10] the wireless node has a duty-cycle of around 2%, which is still too high for very simple applications like a normal indoor temperature sensor or a light switch. For these kinds of applications it is possible to imagine that a single message, no longer than 1000 bits, is sent every 15 minutes.

We can approximate the average power required to transmit a data packet $\mathrm{L_{packet}}$ with the following relation:

$$\mathrm{P_d} = \mathrm{P_{tx}}\frac{\mathrm{T_{tx}}}{\mathrm{T}} + \mathrm{P_{diss}}\frac{(\mathrm{T_{tx}} + \mathrm{T_{wu}})}{\mathrm{T}} + \mathrm{P_{idle}}\frac{(\mathrm{T} - \mathrm{T_{tx}} - \mathrm{T_{wu}})}{\mathrm{T}} \qquad (1)$$

where $\mathrm{P_{tx}}$ is the power radiated from the antenna, $\mathrm{T_{tx}}$ is the time required for each transmission, $\mathrm{T_{wu}}$ is the wake-up time of the transmitter (i.e. the time required to start up the circuitry), $\mathrm{P_{idle}}$ is the power dissipated in the idle mode when most of the transmitter functions are off, $\mathrm{P_{diss}}$ is the non-radiated power dissipated by the transmitter and T is the time interval between two consecutive transmissions.

Given a required $\mathrm{E_b/N_0}$ for a given BER, a certain receiver noise factor NF, a certain noise bandwidth $\mathrm{B_{noise}}$, a certain attenuation due to the channel Loss[1], the data bandwidth $\mathrm{B_{data}}$ and the one sided noise spectral density $\mathrm{N_0}$, the required transmitted power can be derived:

$$\mathrm{P_{tx}} = \mathrm{N_0} \times \frac{\mathrm{B_{noise}}}{\mathrm{B_{data}}} \times \frac{\mathrm{E_b}}{\mathrm{N_0}} \times \frac{\mathrm{NF} \cdot \mathrm{D}}{\mathrm{Loss}} \qquad (2)$$

[1]To calculate the path loss the following expression has been used: $\mathrm{Loss}=27.56 - 20\mathrm{Log_{10}}(\mathrm{f_c}) - 20\mathrm{Log_{10}}(\mathrm{r_0}) - 10 \cdot \mathrm{n} \cdot \mathrm{Log_{10}}(\frac{\mathrm{r}}{\mathrm{r_0}})$ where $\mathrm{f_c}$ is the carrier frequency expressed in MHz, r is the communication distance, $\mathrm{r_0}$ the reference distance for free-space propagation (unobstructed transmission distance which is less than 2 or 3 meters in an indoor environment) and n is the path-loss exponent.

The transmission time depends on the data-rate (D) and on the packet size:

$$T_{tx} = \frac{L_{packet}}{D} \tag{3}$$

Substituting (2) and (3) in (1), an expression for the average transmitter power consumption can be derived as a function of the data-rate.

$$P_d = K\frac{L_{packet}}{T} + P_{diss}\left(\frac{L_{packet}}{D \cdot T} + \frac{T_{wu}}{T}\right) + P_{idle} - P_{idle}\frac{L_{packet}}{D \cdot T} - P_{idle}\frac{T_{wu}}{T} \tag{4}$$

where

$$K = N_0 \times \frac{B_{noise}}{B_{data}} \times \frac{E_b}{N_0} \times \frac{NF}{Loss} \tag{5}$$

As can be seen from (4) and (5), the power required to transmit the data packet can effectively be reduced by increasing the data-rate. On the other hand, the improvement achievable by increasing the data-rate is substantially reduced when the total average power dissipation is dominated by the idle power. In Fig. 1(a) the minimum data-rate required to have an average power consumption equal to P_{idle} + 10% is shown. The power consumption can be considered to be dominated by the idle current in very-low duty cycle systems. Consequently, if the idle current decreases, the minimum required data-rate has to increase accordingly to have an average power consumption that does not exceed the power consumed in the idle mode by more than 10%. In Fig. 1(b) the simulation has been performed at a transmission rate of one data packet every 15 minutes. The data-rate strongly depends on P_{idle} but it is anyhow lower than 35 kbps.

Now looking at commercial low-power transmitter products, it can be seen that idle power consumptions larger than 5 μW are very common. Therefore, from Fig. 1(b), it can be deducted that increasing the data-rate will have a very small influence in decreasing the average transmitter power consumption. Furthermore, this will lead to a much more complex hardware (analog and digital) and larger leakage currents. The final effect is an increase in the power consumption in the idle mode and therefore, an increase in the average transmitter power consumption.

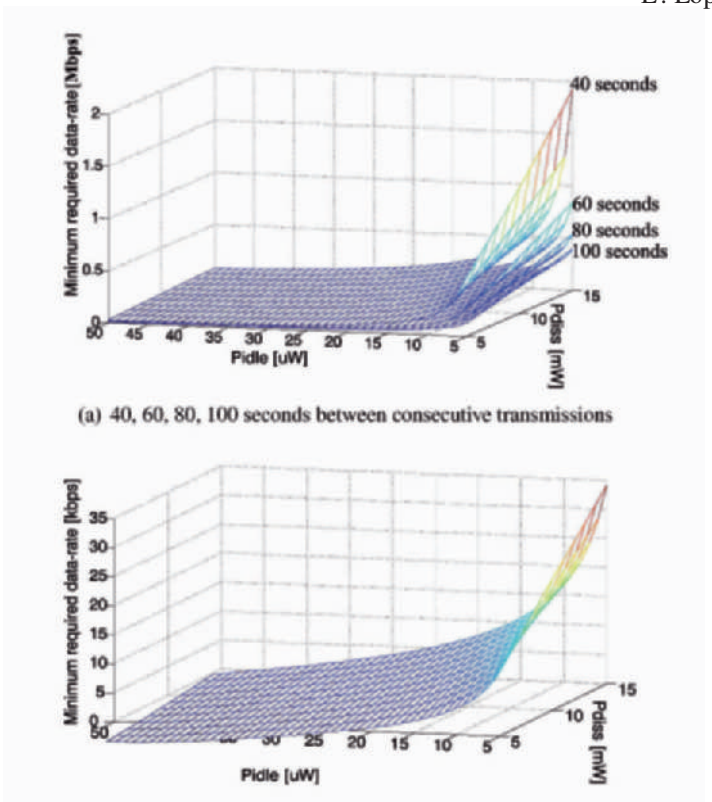


Fig. 1. Minimum data-rates to have an average power consumption equal to P_{idle}*+10% versus* P_{idle} *and* P_{diss}*;* L_{packet}*=1000 bits,* T_{wu}*=100μs, NF=20 dB,* $\frac{E_b}{N_0=15dB}$.

5.1. System Architecture Trends

Different radio architectures have been recently studied in order to reduce the power consumption. Some of these architectures comprise Ultra-Wideband (UWB) transceivers, Super-Regenerative transceivers, Subsampling transceivers as well as Spread-Spectrum based transceivers (both Frequency Hopping and Direct Sequence).

5.1.1. Ultra Wideband Transceivers

Among different architectures suitable for an ultra-low power implementation, UWB based systems are gaining more and more attention. FCC rules specify UWB technology as any wireless transmission scheme that occupies more than 500 MHz of absolute bandwidth.

The most important characteristic of UWB systems is the capability to operate in the power-limited regime. In this regime, the channel capacity increases almost linearly with power, whereas at high SNR it

increases only as the logarithm of the signal power as shown by the Shannon theorem

$$\mathrm{C} = \mathrm{BW} \times \log_2(1 + \frac{\mathrm{P_S}}{\mathrm{P_N}}) \tag{6}$$

where $\mathrm{P_S}$ is the average signal power at the receiver, $\mathrm{P_N}$ is the average noise power at the receiver and BW is the channel bandwidth. For low data-rate applications (small C), it can be seen from (6) that the required SNR can be very small given an available bandwidth in excess of several hundreds MHz. A small SNR translates in a small transmitted power and as a result in a reduction of the overall transmitter power consumption.

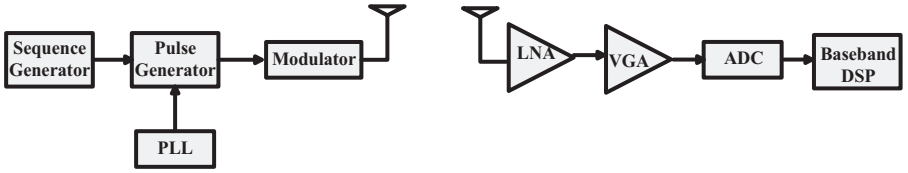


Fig. 2. The building blocks of an impulse based UWB transceiver.

Although UWB transceivers can have reduced hardware complexity, they pose several challenges in terms of power consumption. In Fig. 2 a schematic block diagram of an UWB transceiver is shown. The biggest challenge in terms of power consumption is the ADC. If all the available bandwidth is used, the sampling rate has to be in the order of several Gsamples per second. Furthermore, the ADC should have a very wide dynamic range to resolve the wanted signal from the strong interferers. This implies the use of low-resolution full-flash converters. It can be proven [11] that a 4-bit, 15 GHz flash ADC can easily consume hundreds of milliwatts of power. Even if a 1-bit ADC at 2 Gsample/s is used, the predicted power consumption of the ADC remains around 5 mW [12]. Furthermore, the requirement on the clock generation circuitry can be very demanding in terms of jitter.

Besides these drawbacks, wideband LNA and antenna design are challenging when the used bandwidth is in excess of some Gigahertz. The antenna gain, for example, should be proportional to the frequency [13], but most conventional antennas do not satisfy this requirement. LNA design appears quite challenging when looking at power consumption of state-of-the-art wideband LNAs [14]. A wideband LNA consumes between 9 and 30 mW making it very difficult to fulfill a

constraint of maximum 10 mW peak power consumption for the overall transceiver. Neverthless several successful designs are recently published showing the potential of UWB systems, their power consumption remains too high to be implemented in a "micro-Watt node". In [14] the total power consumption is around 136 mW at 100% duty cycle. In [15] a power consumption of 2 mW has been reported for the pulse generator only.

5.1.2. RF-ID, Subsampling and Super-Regenerative Architectures

In the wide arena of low-power architectures, RF-IDs represent a good solution when the applications scenario requires an asymmetric network. In this case the "micro-Watt node" needs to transmit data and to receive only a wake-up signal. The required energy is harvested from the RF signal coming from the interrogator. In [16] the interrogator operates at the maximum output power of 4 W, while generating by inductive coupling 2.7 μW. This power allows a backscattering-based transponder to send OOK-modulated data back to the interrogator in a 12 meters range using the 2.4 GHz ISM band. Unfortunately the limited amount of intelligence at the transmitter side makes this architecture not flexible and only suitable in a highly asymmetric wireless scenario.

The Nyquist theorem has been explored in subsampling based receiver in order to reduce the overall power consumption. The power consumption of analog blocks mainly depends on the operating frequency. Applying the theory of bandpass sampling [17], it can be proven that the analog front-end can be considerably simplified reducing the operating frequency. This has the potential to lead to a very low power receiver implementation. Unfortunately due to the noise aliasing, it can be proven that the noise degradation in decibels is:

$$\mathrm{D} = 10\mathrm{Log}_{10}(1 + \frac{2\mathrm{MN_p}}{\mathrm{N_0}}) \tag{7}$$

where M is the ratio between the carrier frequency and the sampling frequency, $\mathrm{N_0}$ is the white noise spectral density and $\mathrm{N_p}$ is the BPF filtered version of $\mathrm{N_0}$. In this sense, the choice of the BPF filter as well as the choice of the sampling frequency become quite critical. Beside this, the phase noise specification of the sampling oscillator becomes quite demanding. Indeed, the phase noise is amplified by M^2 requiring a careful design of the VCO. Consequently, when interferers are present, a poor phase noise characteristic can degrade the BER through reciprocal

mixing considerably. Consequently, up to now, this architecture has been used mainly in interferer-free scenarios (space applications) [18].

Super-regenerative architectures date back to Armstrong, who invented the principle. Despite many years of development, they still suffer from poor selectivity and lack of stability, while having the potential to be low power. Furthermore it is restricted to OOK modulation techniques only.

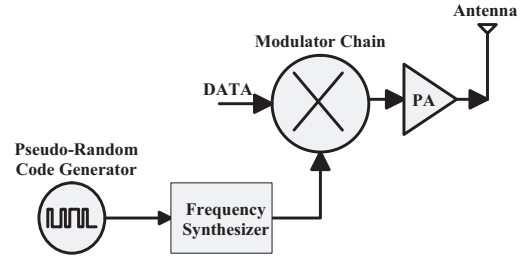
In [19] bulk acoustic wave (BAW) resonators are used to reduce the power consumption and to provide selectivity. In spite of achieving an overall power consumption of 450μW, it relies on non-standard technologies (BAW resonators), which will increase cost and form factor of the "micro-Watt node".

In [20] a 1.2 mW receiver has been designed and fabricated in 0.35-μm CMOS technology. Even though the power consumption is very close to the requirements of a "micro-Watt node", selectivity is quite poor. Indeed, to demodulate the wanted signal in the presence of a jamming tone placed 4 MHz far from the wanted channel with a BER of 0.1%, the jamming tone has to be no more than 12 dB higher than the desired signal. Generally, to achieve a reliable communication, the receiver should be able to handle interferers which have a power level 40 dB higher than the wanted signal with a BER smaller than 0.1%. This specification is very demanding for a super-regenerative architecture and it requires the use of non-standard components like BAW resonator to achieve a better selectivity.

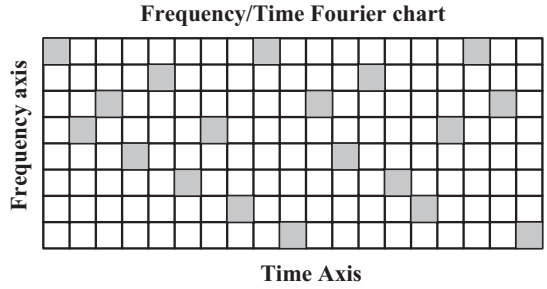
5.2. Spread Spectrum Systems

Among the various competing SS techniques, the FH system, in which the transmitting carrier frequency is changing according to a prescribed pseudorandom sequence, offers an attractive solution in terms of hardware complexity, system performance and power consumption.

FH also enables the simplicity of using FSK modulation, which allows the possibility of employing a direct-conversion receiver architecture. In this way, a greater integration level and a lower power consumption can be achieved. Furthermore, FSK modulation can be easily superimposed on the hopping carrier by simple digital techniques. Finally, the modulated output has a constant envelope, and is amenable to non-linear power amplification. A schematic block diagram of an FH transmitter is depicted in Fig. 3(a), while the distribution of the signal in the time-frequency plane is depicted in Fig. 3(b).



(a) Block diagram of a frequency hopping transmitter



(b) Frequency hopping signal

Fig. 3. A simplified example of an FHSS transmitter.

5.2.1. DSSS vs. FHSS

Differently from a DSSS system, the spreading code is applied on the carrier frequency rather than on the modulated data. This choice has some advantages as well as some drawbacks. In terms of power consumption, the wake-up time required by the system will be considerably lower. The wake-up time takes into account also the synchronization time. It can be proven that the average acquisition time for an SS system is

$$\overline{\mathrm{T_s}} = (\mathrm{C}-1)\mathrm{T_{da}}\left(\frac{2-\mathrm{P_d}}{2\mathrm{P_d}}\right) + \frac{\mathrm{T_i}}{\mathrm{P_d}} \tag{8}$$

where $\mathrm{T_i}$ is the integration time for the evaluation of each cell in the time-frequency plane, $\mathrm{P_d}$ is the probability of detection when the correct cell is being evaluated, $\mathrm{T_{da}}$ is the average dwell time at an incorrect phase cell, C is the total number of cells.

Now, assuming that no frequency uncertainty is present, there will be a time misalignment between the two PN sequences at the transmitter

and receiver side equal to ΔT_i. Therefore, while for a DSSS the system has to be synchronized within $\pm T_c/2$ where T_c is the chip duration, an FHSS system needs to be synchronized within $\pm T_{symb}/2$, where T_{symb} is the symbol period. Due to the fact that in a DSSS system the processing gain is related to the ratio between the chip rate and the symbol rate, the chip period is at least an order of magnitude smaller than the symbol period. As a result, the number of cells that have to be evaluated in a DSSS system is considerably bigger than in an FHSS system. From (8) the mean DSSS synchronization time is bigger than in the case of an FHSS system. This will increase the wake-up time and therefore the overall system power consumption.

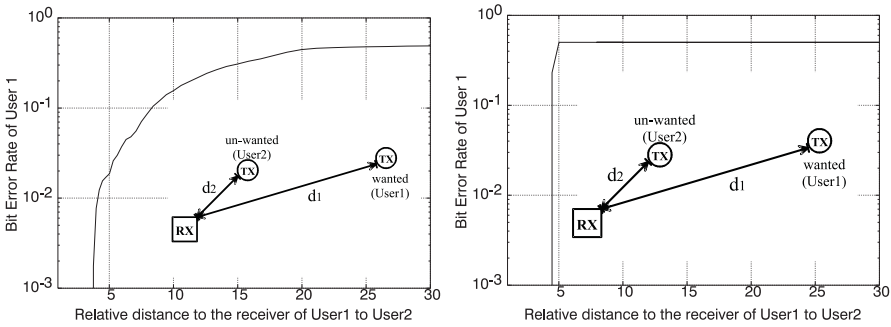


Fig. 4. Near-far sensitivity comparison between FHSS and DSSS.

Spread-Spectrum systems are generally affected by the so called near-far problem. The near-far problem is the major limitation in DSSS systems and increases the complexity of the transceiver due to the need for power control circuitry. Without any power control, the performances of a DSSS system in an environment in which other DSSS systems are present can be heavily spoiled.

As shown in Fig. 4, in an FHSS system, when User 1 is 5 times further away from the receiver compared to User 2 , the BER is still below 5%. At the same relative distance, the DSSS system has already a BER of about 50%. The results shown in these plots have been obtained using Simulink^{TM} models, in which the users were interfering continuously with each other and only an Additive White Gaussian Noise (AWGN) channel has been considered. In reality, thanks to the low data rates, the probability that 2 or more users will communicate simultaneously, is in the order of a few percent. Consequently, the average bit error rate should be scaled down accordingly. For example, a BER of 10% will

translate to a BER in the order of 0.1%. In conclusion, there is a high probability that in a DSSS system an Automatic Gain Control (AGC) should be used while, in an FHSS system, it can be easily avoided, reducing the complexity and the power consumption of the system.

Unfortunately, the requirements on the hopping synthesizer in terms of settling time and accuracy in the frequency synthesis are quite stringent. Therefore, the implementation of FHSS functionalities in a "micro-Watt node", requires a novel simplified architecture with a power consumption an order of magnitude smaller than the state-of-the-art.

6. Frequency Hopping Spread Spectrum Synthesizers

The main challenges in the implementation of an FHSS system result from the requirements for agile and accurate frequency hopping at very low power levels. Several ways to generate the hopping bins have been proposed in the recent years. Whereas widely used also for high data-rate links, fractional-N [21] and Direct-Digital Frequency Synthesizers (DDFS) [22] based synthesizers are relatively power hungry. In [21] the synthesizer power consumption is 55 mW from a 2.5 V power supply while in [22] the synthesizer dissipates 40 mW from a 3 V power supply.

Fractional-N based synthesizers have to deal with increased phase noise coming from the $\Sigma - \Delta$ modulator. Indeed, the quantization noise is filtered by the PLL phase transfer function and converted into phase noise [23]. Concluding, a trade-off between phase noise and settling time exists. Phase noise can increase the system BER by reciprocal mixing. Therefore, it should be kept low by increasing the reference frequency or the order of the loop filter [24]. This will increase in both cases the overall power consumption.

Recently, a new system concept has been proposed based on Analog-Double Quadrature Sampling (A-DQS) to relax the specifications on the synthesizer [25]. The channel selection originally performed by the synthesizer can be partitioned to the A-DQS. In this way, the step size of the PLL synthesizer can be doubled and the locking position of the LO in the entire frequency band is halved. Nevertheless, the power consumption remains too high, mainly due to the ADC requirements [26].

In a DDFS, a sinewave is synthesized in the digital domain through the use of a simple accumulator, which produces, as an output, a ramp proportional to the desired frequency, and a phase-to-sine amplitude

converter. In the simplest case this converter is a Read-Only Memory (ROM). A DAC and a Low-Pass Filter (LPF) are used to convert the sinusoid samples into an analog waveform.

The main operation in a phase accumulator of N-bit length is the N-bit addition. From [27] the energy required for an addition by an Arithmetic and Logical Unit (ALU) can be considered in the range of 300 pJ per addition (16-bit addition). If a bandwidth of 9.6 MHz (64 channels spaced by 150 kHz in the 915 MHz ISM band) should be synthesized using the architecture proposed in [22], then a 25 MHz reference clock is needed and the power consumption of the phase accumulator can be predicted by the following equation:

$$\mathrm{P_{phase-acc}} = \mathrm{f_{ref-clk}} \times \mathrm{E_{add}} \tag{9}$$

where $\mathrm{f_{ref-clk}}$ is the reference clock frequency and $\mathrm{E_{add}}$ is the energy per N-bit addition (300 pJ). From (9) the predicted power consumption for the phase accumulator is 7.5 mW. The second block in a DDFS, which to a great extent has significant power consumption, is the ROM.

Generally, even if a resolution of few Hertz is needed to keep the spurious level low, practical words longer than 14 bits will lead to a very large ROM even if compression techniques are employed. Considering a truncated 14-bit phase word and a 12-bit wordlength for the amplitude mapping, then the size of the ROM will be approximately 192 kbit. Splitting, for convenience, the ROM in three banks of approximately 64 kbit, then it can be implemented by three $2^8 \times 2^8$ matrices.

Defining the storage array as a $2^{\mathrm{n}} = 2^{\mathrm{n-k}} \times 2^{\mathrm{k}}$ matrix with 2^{n} memory cells, $2^{\mathrm{n-k}}$ rows and 2^{k} columns, then from [28] most of the power consumption in a ROM comes from the pre-charge or the evaluation of the 2^{k} memory cells. Therefore, we can approximate the total power consumption by the following:

$$\mathrm{P_{memcell}} = 3\frac{2^{\mathrm{k}}}{2}(\mathrm{c_{int}}\mathrm{l_{column}} + 2^{\mathrm{n-k}}\mathrm{C_{tr}})\mathrm{V_{dd}}\mathrm{V_{swing}} \tag{10}$$

where $\mathrm{P_{memcell}}$ is the approximated power consumption of the ROM, the factor 3 takes into account that the total memory required has been split into three ROMs of smaller size, $\mathrm{c_{int}}$ is the capacitance of a unit wire length with minimum width, $\mathrm{C_{tr}}$ is the minimum size gate capacitance, $\mathrm{V_{dd}}$ is the power supply voltage and $\mathrm{V_{swing}}$ is the voltage swing of each memory cell. Defining the memory cell as $\mathrm{d_m} \times \mathrm{d_m}$ square, then the column interconnection length of the memory matrix is $\mathrm{l_{column}} = 2^{\mathrm{n-k}}\mathrm{d_m}$.

Now, considering as an example the CMOS 0.18 μm technology, a 5.4 mW predicted power consumption for the ROM is obtained. As a result, the peak power consumption of the DDFS excluding DACs is higher than the 10 mW limit of a "micro-Watt node".

6.1. Pre-Distortion-Based Hopping Frequency Synthesizers

To reduce the overall power consumption to a level that allows the implementation of "micro-Watt node", a new architecture is proposed based on the direct synthesis of frequency bins. The schematic block diagram of the proposed architecture is shown in Fig. 5.

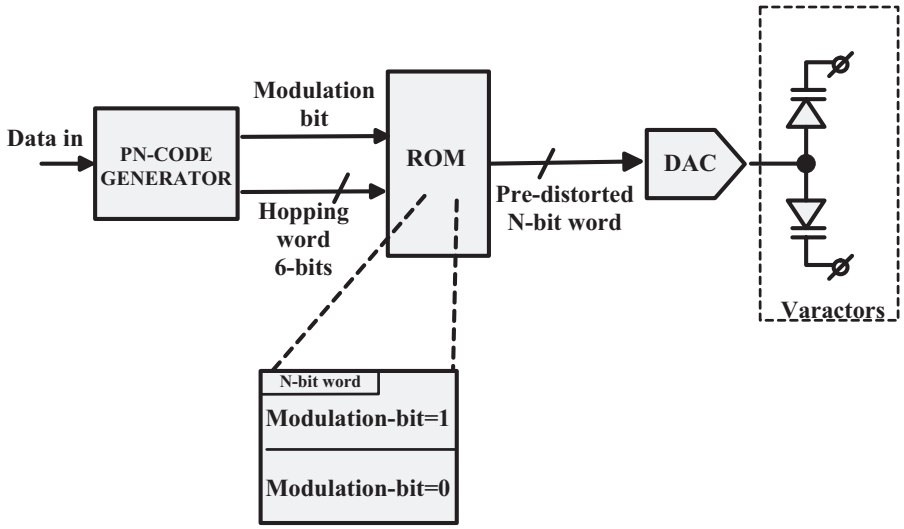


Fig. 5. Proposed hopping frequency synthesizer architecture.

The incoming data, together with the wanted hopping code, addresses a particular word cell in the ROM. The ROM has been split into two blocks depending on the modulation bit. In each memory cell the pre-distorted word that will drive the DAC with a defined data bit applied is stored. The DAC then drives directly the varactor array changing the capacitance and therefore the oscillation frequency according to the desired frequency bin and data.

The direct frequency synthesis is based on the well known relation between frequency and tank capacitance in a resonator based VCO (LC-VCO for example):

$$f_{osc} = \frac{1}{2\pi\sqrt{LC}} \tag{11}$$

where L and C are respectively the total capacitance and inductance of the tank and f_{osc} is the oscillation frequency. It is possible to change the oscillation frequency of the VCO by varying the capacitance of the tank. This, in practice, is realized by using a varactor diode, which has a capacitance that varies non-linearly with its reverse voltage across it. Therefore, applying the correct voltages to the varactor diodes, it is possible to synthesize all the required frequency bins with minimum hardware complexity (virtually only a VCO).

In an FHSS system the various frequency bins are addressed in a pseudo-random fashion. Pseudo-random codes are generated in the digital domain, while the varactor diodes require an analog control voltage. Consequently, a DAC is required as interface between the digital world and the analog world. It is important to notice that the DAC employed in a DDFS based frequency synthesizer works at a much higher frequency. Indeed, it should be able to synthesize frequencies in the order of MHz with certain spectral characteristics.

In this new approach the DAC has to be able to give a steady-state voltage at a rate comparable to the hopping rate (kHz range) without any constraint on its dynamic linearity (as long as the DAC settling time is much smaller than the dwell time). This will considerably simplify the design of the DAC allowing a current consumption as low as 100 μA [29].

6.1.1. PN-Codes Orthogonality

The architecture shown in Fig. 5 is an entirely feed-forward system. Due to the fact that no feedback loop is present, errors cannot be corrected. As a result, the system should be accurate in the absolute sense. In CDMA applications, like the one envisioned for a micro-Watt network, orthogonality of PN codes is of primary importance. Indeed, a loss of orthogonality will produce a huge degradation of the BER when two nodes with non-orthogonal codes communicate at the same time. Any non-linear behavior in the transmitter chain depicted in Fig. 5 can produce a loss of orthogonality between two orthogonal-generated PN codes.

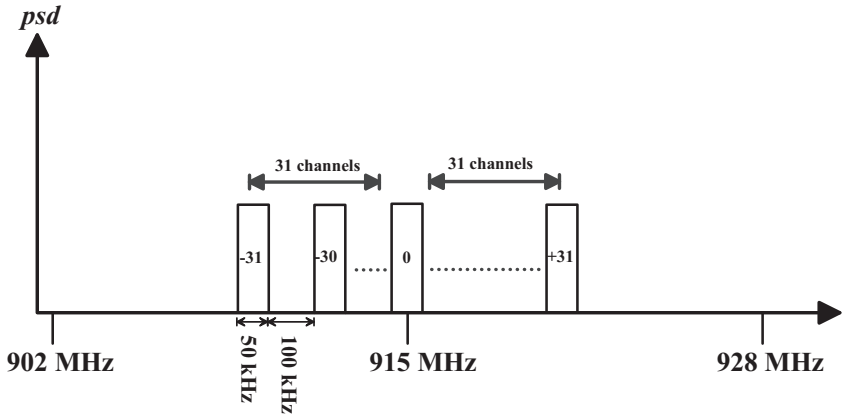


Fig. 6. Channel allocation scheme.

Given a 150 kHz frequency separation between the center frequencies of two adjacent channels, then the error in the synthesized frequency should not exceed 100 kHz (50 kHz channel bandwidth) if code orthogonality has to be preserved (see Fig. 6).

6.1.2. Frequency Offset Effect on Phase Noise Specifications

Unfortunately the allowed maximal frequency error has to be smaller than the inter-channel spacing. When the inter-channel distance reduces, while the orthogonality of the PN codes is preserved, phase noise of the unwanted channel will reduce the SNR in the wanted channel. Moreover, through reciprocal mixing the SNR can be reduced well below the minimum requirement to demodulate the incoming data with a certain BER. While increasing the phase noise performance of the synthesizer will allow to take into account this drawback, it requires an increase of the overall power consumption. The only solution, when the system is power constrained, resides in the reduction of the nonlinearities.

Therefore, the upper bound on the maximum allowed frequency error after pre-distortion is set by the phase noise specification. The most critical specification for the phase noise is set by the third and beyond channel [30], where the phase noise requirement is -100 dBc/Hz. The influence of the frequency offset is different depending on the relative position between the wanted and unwanted channels. It has the greatest effect when the wanted and unwanted channel are adjacent and a

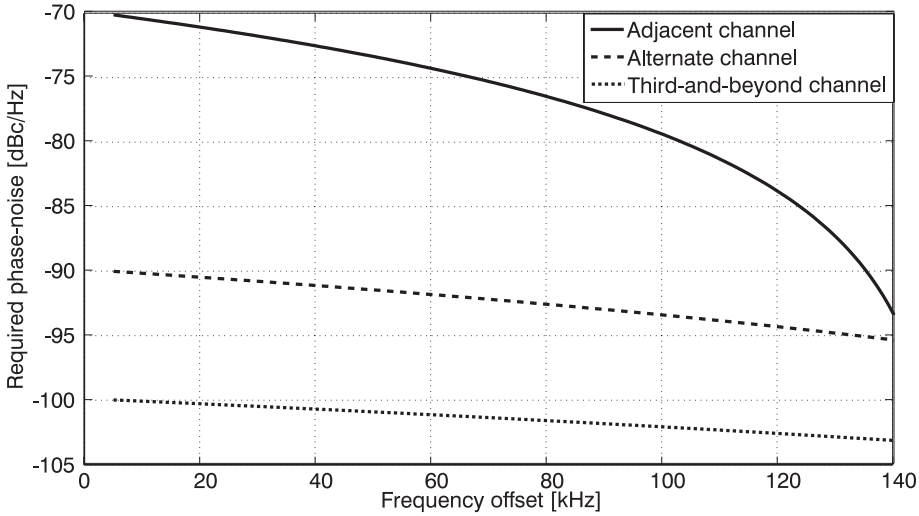


Fig. 7. Phase noise requirements vs. frequency offset (Phase noise requirements are: -70, -90, -100 dBc/Hz for the adjacent, alternate and third-and-beyond channel respectively, when no residual frequency offset is present).

smaller effect when they are two or more channels apart. This is clearly shown in Fig. 7. Anyhow, from Fig. 7, the most critical specification is still on the third-and-beyond channel regardless of the frequency offset. In order not to degrade the BER through reciprocal mixing, a 0.5 dB maximum degradation on the phase noise has been considered[2]. Under this condition, a maximum frequency error of 25 kHz can be tolerated.

6.1.3. Effect of Quantization Errors and Non-Linearities on the Frequency Synthesis Accuracy

The chain composed by the DAC, the varactor and the oscillator, introduces the following non-linearities:

- Square-root relation between frequency and capacitance
- C-V characteristic of the varactor

[2]The error probability of a non-coherent BFSK modulated signal is given by $\frac{1}{2}e^{-\frac{E_b}{2N_0}}$. Considering a 0.1% initial BER, a 0.5 dB degradation in the phase noise translates in a 0.5 dB degradation in the SNR at the demodulator input and therefore in a BER close to 0.2%.

- DAC Integral Non-linearity (INL)

A real DAC has a finite number of bits and therefore, the quantization error will play a role in setting the residual frequency error. While passing through the previous non-linearities, the quantization error translates in a non-linear frequency error. Besides this effect, also the INL of the DAC plays a role in the frequency synthesis accuracy. The aim of the pre-distortion algorithm is to compensate the effects caused by these non-linearities in the digital domain by applying pre-distorted digital codes to the DAC.

Looking at the varactor non-linearity and the square root relation between frequency and capacitance in an LC resonator, the following considerations apply. When the overall capacitance is the smallest (at higher frequencies), an error on the capacitance due to uncompensated non-linearity, will produce a larger error in the synthesized frequency. In this situation, the reverse voltage applied to the varactor is at its maximum value (for example -1.6 V) and the quantization error will have the greatest effect. In this region, however, the varactor exhibits a highly linear behavior, while close to the minimum reverse voltage (for example -0.2 V), it is highly non-linear. In this case the capacitance has its maximum value and the synthesized frequency is the lowest possible.

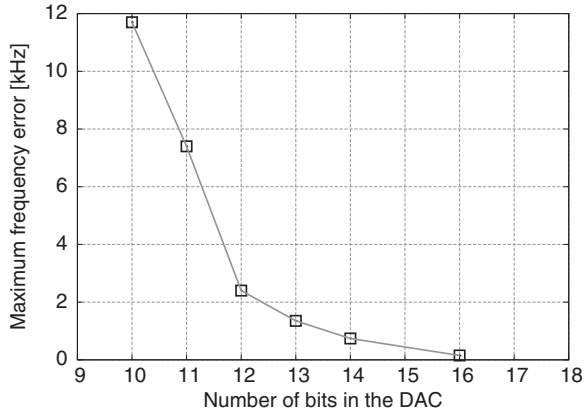
It can be proven that the maximum allowed capacitance error[3] in the two cases is given by the following equations:

$$\Delta C_{\sqrt{LC}} = \frac{1}{[(2\pi(f_{max} + \Delta f)]^2 L} - C_{min} \tag{12}$$

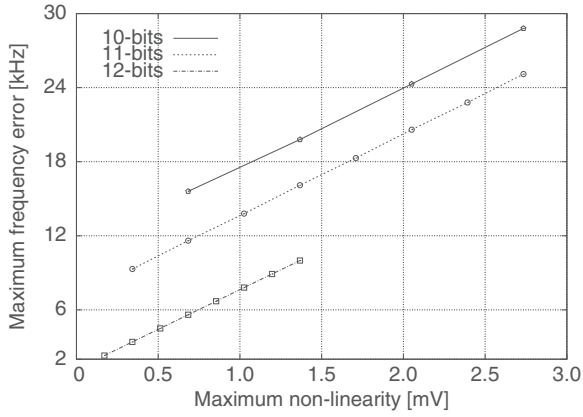
$$\Delta C_{var} = \frac{1}{[2\pi(f_{min} - \Delta f)]^2 L} - C_{max} \tag{13}$$

where f_{max} is the highest channel center frequency, f_{min} is the lowest channel center frequency, Δf is the maximum allowed frequency offset, C_{min} is the capacitance at the highest channel frequency, C_{max} is the capacitance at the lowest channel frequency and L is the LC-tank inductance value. Considering a 1.4 V swing on the varactor control voltage, an inductance value of 4.1 nH, and 64 channels placed around 915 MHz (ISM band), it can be found that the varactor non-linearity produces the largest frequency error, due to the quantization error in the DAC, requiring a minimum voltage step of about 2.2 mV, which

[3]The capacitance value is calculated so that the maximum residual frequency error after pre-correction is below the required specification of 25 kHz.



(a) Required DAC resolution



(b) Effect of DAC INL on the frequency error

Fig. 8. Simulated DAC resolution and INL requirements.

translates in at least 10 bits resolution. Using a 10-bit DAC translates in a frequency offset smaller than 25 kHz and roughly equal to 15 kHz according to theory.

The system has been simulated in Simulink$^{\mathrm{TM}}$ and the results are shown in Fig. 8. At 10-bits DAC resolution the residual frequency error is around 12 kHz (Fig. 8(a)), a value very close to the predicted one of 15 kHz.

The varactor has been modeled with a 6^{th} order polynomial. The choice of the polynomial order depends on the maximum allowed capacitance error resulting from the approximation. It should be no more

than 10% of the other error sources in such a way to consider it negligible. The maximum frequency error, using a 6^{th} order polynomial with respect to the interpolated points, is around 3.5 kHz as shown in Fig. 9. The maximum errors are also placed at the border of the varactor control voltage (-1.6 and -0.2 V) and are minimum in the middle range. Increasing the order of the interpolating polynomial above the 6^{th} order does not produce remarkable improvements. As a result, a 6^{th} order polynomial has been chosen for the interpolation.

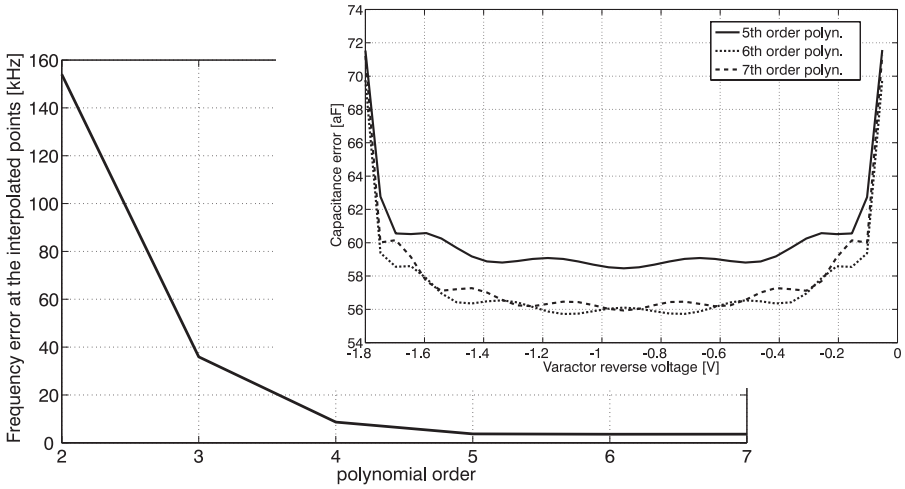


Fig. 9. Dependence of frequency error from the order of the interpolating polynomial.

The last non-linearity in the transmitter chain is the one introduced by the DAC INL. The DAC INL has been modeled with a second-order non-linearity and the maximum INL (INL_{max}) placed at the middle code. Therefore, the output of the DAC can be modeled as a function of the number of bits and the INL_{max} using the following expression

$$V_{\text{out}}^{\text{DAC}} = \frac{-\text{INL}_{\text{max}}}{2^{\text{N}-1}(2^{\text{N}}-1-2^{\text{N}-1})} X_{\text{in}}^2 + [\text{INL}_{\text{max}} \times \frac{2^{\text{N}}-1}{2^{\text{N}-1}(2^{\text{N}}-1-2^{\text{N}-1})} + \frac{\text{FSR}}{2^{\text{N}}-1}] X_{\text{in}} + 0.2 \quad (14)$$

where FSR is the full-scale range of the DAC[4], N is the number of bits and X_{in} is the digital N-bits input code. As can be seen in Fig. 8(b), a

[4]The voltage range for the varactor ϵ [-1.6;-0.2] V.

10-bits DAC with an $\mathrm{INL}_{\max}$ smaller than 1.5 LSB fulfills the required specifications. On the other hand, INL requirements can be traded with DAC resolution giving a degree of freedom in the design (see Fig. 8(b)).

6.2. Demodulation and Robustness

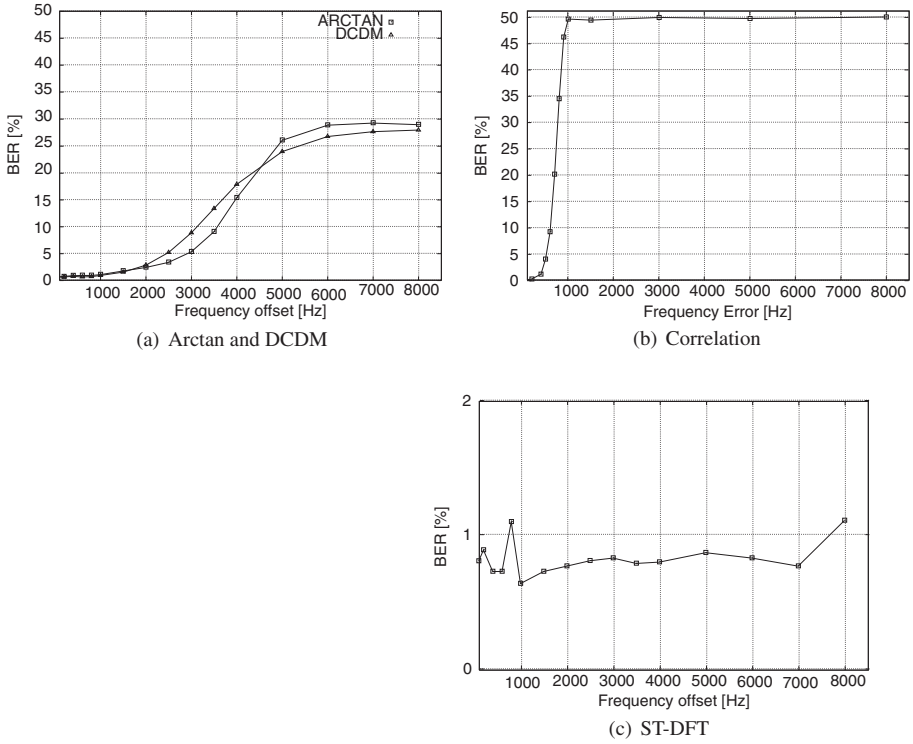
Though the 25 kHz upper-bound frequency error fulfills the requirements in terms of phase noise degradation and orthogonality of the codes, it poses problems at the receiver side. Indeed FSK modulation is very sensitive to frequency offsets. If a correlation-based demodulator is considered, then the frequency error is not allowed to exceed the data rate [31].

In the case of data rates between 1 and 10 kbps, this will pose strict requirements on the DAC linearity and resolution as well as on the maximum process spread of the varactor C-V curves.

Concluding, it is important to find an algorithm and a demodulator topology, which can successfully demodulate the incoming data under offsets as big as 25 kHz. In [31] several demodulator topologies have been studied, with respect to their performances, in the presence of frequency errors. The results are shown in Fig. 10. As can be seen, the Short-Time DFT (ST-DFT) algorithm, through differential encoding, shows a remarkable immunity against static frequency offsets. Furthermore it can be applied in the digital domain which has the potential to be low power. Indeed, if a zero-IF architecture is employed, due to the small signal bandwidth, the operating frequency of the ADC will be around 100 ksample/s at the Nyquist rate (signal bandwidth equal to 50 kHz).

If a correlation based demodulator is used, the residual frequency offset has to be much smaller than the data rate in order to not degrade the receiver BER performance. This means that either the data-rate has to be increased or the residual frequency error has to be reduced considerably. Both the Arctangent and Digital Cross-Differentiate-Multiply (DCDM) demodulators use the phase information embedded in the signal to recover the transmitted data. Due to their similarities, both the DCDM and the Arctangent demodulators have the same performances. They exhibit a higher immunity to the frequency offset but it is still not sufficient for the required application.

As shown in [32], the capability of the ST-DFT algorithm to reject the frequency offset depends upon the condition that the offset is slowly



(a) Arctan and DCDM

(b) Correlation

(c) ST-DFT

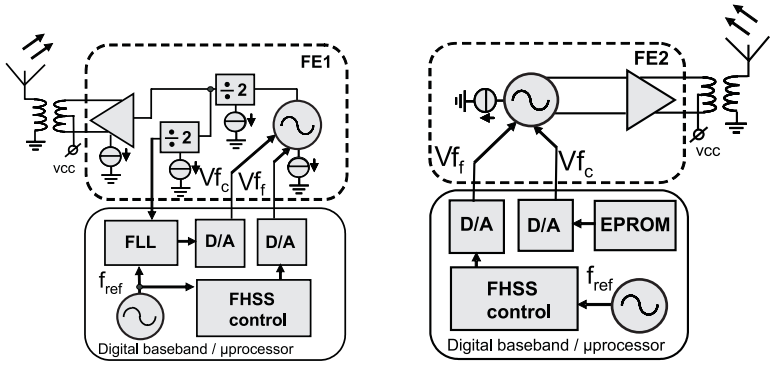
Fig. 10. BER vs frequency offset with $\mathrm{E_b/N_0} = 12\mathrm{dB}$.

varying. In other words, the frequency offset should vary at a rate smaller than the data-rate. Therefore, the offset between two consecutive bits can be considered the same and it will be canceled out when differential encoding is applied. In this way also the frequency error due to temperature and power supply variations can be tracked and actively canceled out without requiring any additional circuitry.

In the particular case of an FHSS system, each bit is sent on a different channel, which is affected by a different offset due to the INL distribution of the DAC as well as the varactor non-linearity. This means that two different bits will have two different offsets and, therefore, the simple differential encoding cannot cancel it out. This can be solved in the transmitter by sending information about the frequency offset every bit and with a receiver that tracks this information such that it cancels out the residual offset for each bit. The description of the aforementioned algorithm is beyond the scope of this paper and it will not be discussed further.

7. FHSS Pre-Distortion Based Transmitter Design

Two different transmitter RF front-ends have been realized in bipolar Silicon-on-Anything (SOA) technology. The front-end has to be as simple as possible while pushing the complexity in the digital domain. This is achieved by using the pre-distortion technique previously described. All the digital words stored in the ROM (see Fig. 5) are calculated so that they take into account the DAC quantization error and its effect via a non-linear transfer function (square root and varactor) on the frequency error. The INL of the DAC has to be compensated separately for each IC or can be left uncompensated if it is smaller then 1.5 LSB for a 10-bit DAC(see Fig. 8(b)). The two architectures are depicted in Fig. 11.



(a) Oscillator-divider based front-end (b) Single building block front-end

Fig. 11. Transmitter RF front-end.

In Fig. 11(a) a common oscillator-divider front-end is shown. The system front-end consists of a resonator based LC-VCO, a divider and an output stage able to deliver -25 dBm power on a 50 Ω load. To minimize oscillator pulling the VCO operates at 1.8 GHz and is divided down to the TX frequency. V_{f_c} is the coarse control voltage and it is used to calibrate the hopping channels inside the ISM band. The second divider, connected to a second buffer, is used inside an FLL loop for the initial center frequency calibration. Once this calibration is performed, the FLL and related dividers are powered down, thus not contributing to the total power dissipation. V_{f_f} is the fine control voltage and it is used

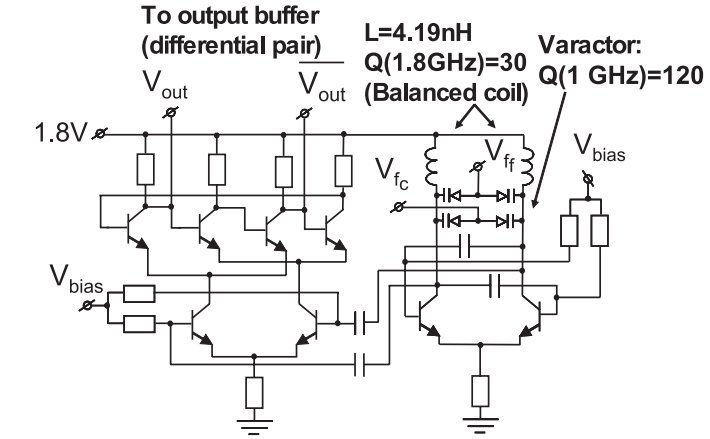
to synthesize the 64 channel frequencies. The FHSS control, which is realized in baseband, generates the PN hopping code, and it outputs the N-bits pre-distorted digital words. These words are applied to the D/A converter to synthesize the required control voltage for the fine varactor bank.

Figure 11(b) shows a single building block RF front-end. It is a combination of VCO and PA, of which the coarse frequency calibration code can be factory set and stored in an EPROM. This code will be converted into an analog voltage by a DAC. The output of the DAC will directly drive the coarse varactor on the power-VCO. The synthesis of the channel frequencies is obtained in the same way as for the architecture shown in Fig. 11(a). In this case, the oscillator center frequency is at 915 MHz instead of 1.83 GHz, which can reduce the system power consumption. Furthermore, no PA is required, but the VCO is directly coupled to the antenna through a balun.

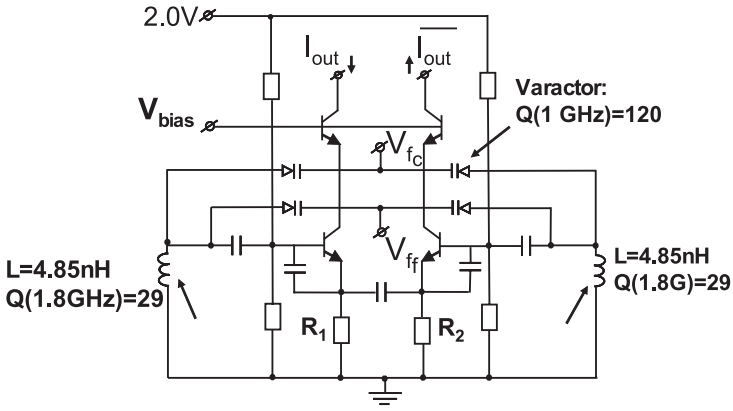
The baseband part has been implemented using discrete components for both architectures. The microprocessor is a PIC18F627A, which consumes 12 μA at 32 kHz, while the DAC is an AD7392, which draws 100 μA. The transistor level schematics of the two front-ends are depicted in Fig. 12.

The divider in Fig. 12(a) is a traveling wave divider. In this design the often present external base-resistors of the upper stage are not used and the design is optimized by proper transistor dimensioning to have maximum divider sensitivity at 1.8 GHz and a low power dissipation (200 μA). The output buffer (not shown) is a differential pair (2 mA). The oscillator core consumes 550 μA while achieving a -109 dBc/Hz phase noise at 450 kHz from the carrier.

The second architecture (Fig. 12(b)) consists only of a directly modulated RF cascoded Colpitts power VCO. Cross-coupled oscillators (Fig. 12(a)) have been preferred over other topologies for monolithic integrated circuit implementation because they are easily realized using CMOS technology and differential circuitry. Due to the use of the tail current source, cross-coupled oscillator topologies present phase noise performances worse than classical types of oscillator with one of the active device ports grounded. In addition, classical type of oscillator topologies provide larger oscillation amplitudes for a given bias current because there is no voltage drop of the DC current across the current source element enabling, in this way, optimization of the power consumption.



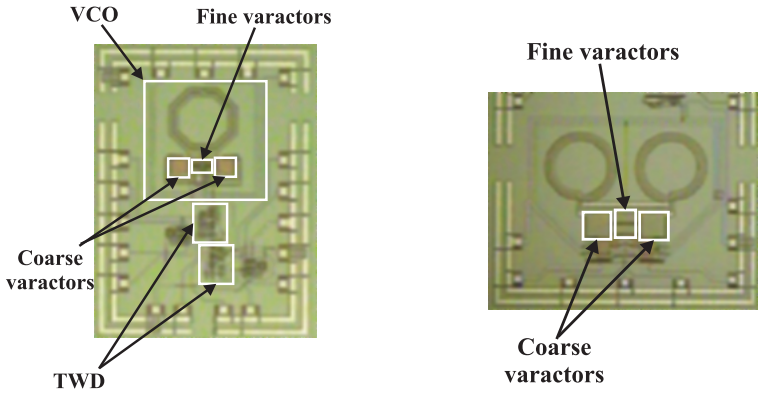
(a) FE1 transistor-level schematic



(b) FE2 transistor-level schematic

Fig. 12. Transistor-level schematic of the realized front-ends.

One of these configurations is based on the Colpitts topology. Isolation problems can degrade VCO performances in terms of phase noise and frequency stability through phenomena like VCO pulling. This problem arises from changes in the load conditions, and therefore requires a buffer stage between the VCO and the output stage, which will increase the overall power consumption. To minimize the power consumption, the VCO and its buffer are connected in series to reuse the bias current between the two stages. For these reasons, a common collector Colpitts oscillator has been chosen together with a common base buffer stage, arranged in a cascode configuration.



(a) FE1 die photo (b) FE2 die photo

Fig. 13. FE1 and FE2 die photos.

In this way, the current consumption is minimized, pulling is reduced due to better isolation between the tank and the load and no PA is required, instead the cascode stage can directly drive the antenna through a balun. As a result, a single block RF front-end is obtained. A differential configuration has been chosen to reduce the second order harmonic distortion term, which can degrade performances at the receiver side in a zero-IF architecture.

The bias current level largely depends on the required output power levels. It is desirable that phase noise requirements are met over a wide output power range. The output power can then be controlled by changing the bias current while meeting the BER specifications over the entire control range.

7.1. Experimental Results

The two RF frond-end ICs are shown in Fig. 13(a) and Fig. 13(b). The FE1 front-end occupies 2.8 mm^2 while the FE2 front-end 3.6 mm^2. Fig. 14 shows the required bias current for the power VCO versus the output power and phase noise performance at 450 kHz far from the carrier. The required bias current ranges approximately between 1 and 2 mA to obtain an output power between -18 dBm and -5 dBm. Simulated and measured results show a good agreement allowing minimization of the current consumption for a given set of output power and phase noise specifications. Indeed in Fig. 14 the phase noise varies between -102 dBc/Hz and -115 dBc/Hz when the bias current changes between 1 and

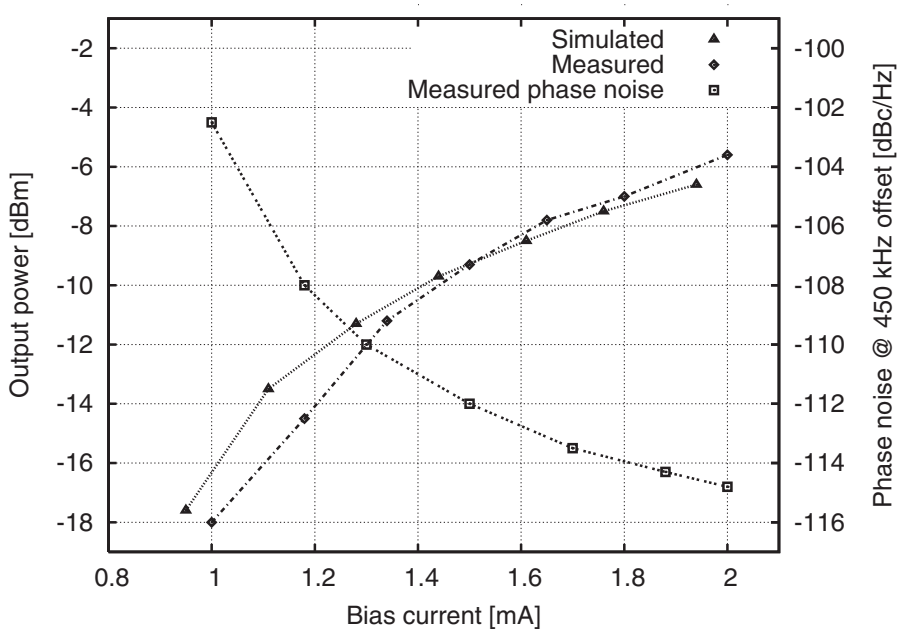


Fig. 14. Output power and phase noise versus bias current at 450 kHz offset.

2 mA, which is better than the required specification of -100 dBc/Hz at 450 kHz from the carrier.

In Fig. 15 the whole chain, from the look-up table in the microprocessor up to the evenly spaced FH spectrum is shown. In the measurement results shown in Fig. 15, the 64 channels are addressed sequentially rather than in a pseudorandom fashion. As can be seen, the output of the DAC has a non-linear shape in time due to the predistortion algorithm. Moreover, the channels are equally spaced with a maximum frequency error smaller than 5 kHz. This frequency error is smaller than theoretically predicted, due to the higher DAC resolution. Indeed, a 12-bit DAC has been used even if a 10-bit DAC was sufficient as explained in Section 6.1.

To demonstrate a reliable wireless link at the specified output power levels and phase noise specifications and to prove the pre-distortion concept, a FH transmitter has been realized. Two isotropic antennas placed 8 meters apart were used in a Non-Line-of-Sight (NLOS) configuration in a normal office environment.

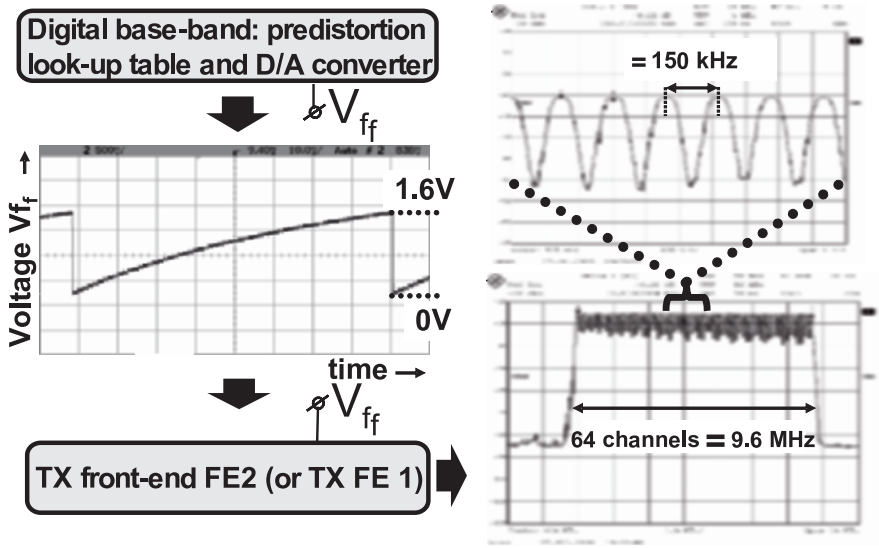


Fig. 15. Pre-distortion chain with measured DAC pre-distorted output voltage and output spectrum of TX FE2 (FE1 has an identical spectrum).

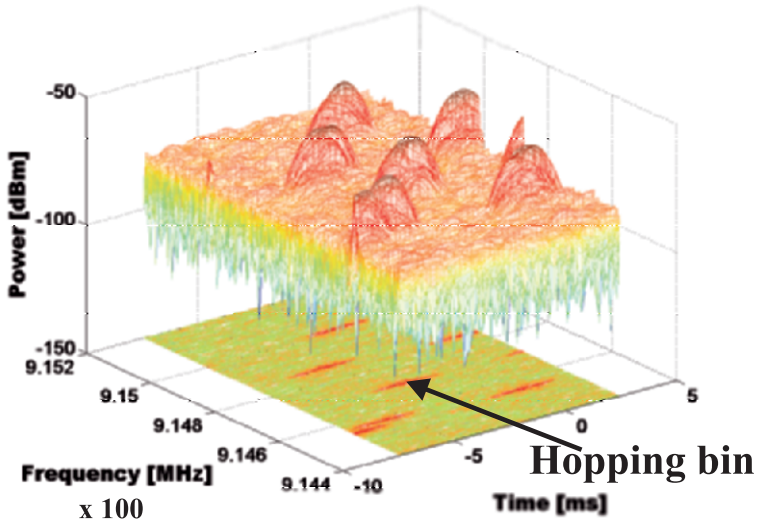
The transmitted power has been set to -25 dBm while the receiver employs a ST-DFT demodulation algorithm and a super-heterodyne architecture. The measured received power is on average -75 dBm. In this condition, the measured raw BER is lower than 1.1% at 1 kbps data rate and 1 khop/s hopping rate. The overall measured transmitter power consumption is 2.4 mW (for FE2) from a 2 V power supply and 5.4 mW (including the buffer) for FE1 from 1.8 V power supply. The baseband and mixed signal circuitry including the DSP dissipate 0.4 mW mostly consumed in the DACs[5].

In Fig. 16 the received spectrum before demodulation is shown together with a summary of the performance of the realized front-ends.

8. Conclusions

The implementation of a "micro-Watt network", envisioned in the AmI concept, has a unique set of design constraints that focuses attention on small required bandwidth, low duty cycle operation, short communication range, and the most challenging requirements in average and peak power demand. Nevertheless, a robust wireless link is mandatory,

[5]The baseband power consumption can be reduced to 0.2 mW if the coarse calibration DAC is switched off after coarse calibration is achieved.



(a) Received power spectrum (8 meters distance, NLOS condition and -25 dBm transmitted power)

Parameter	FE1	FE2(1mA)	FE2(2mA)	Unit
Technology	SOA	SOA		
Active chip area	2.8	3.6		mm^2
V_{cc}	1.8	2		V
Max. Front-end current (20 samples)	2.8	1	2	mA
Max. Front-end dissipation (20 samples)	5	2	4	mW
Baseband dissipation	0.4			mW
Total active FHSS TX dissipation	5.4	2.4	4.4	mW
Min. coarse tuning range (20 samples)	94.4	50		MHz
Min. fine tuning range (20 samples)	8.9	10		MHz
Worst case phase noise (20 samples)	-109	-102	-115	dBc/Hz @ 450 kHz
Output power	-25	-18	-5	dBm
Raw BER (8 meter distance, P_{out}=-25 dBm)	$< 1.1E-3$			

(b) RF front-ends performance summary

Fig. 16. Received spectrum and RF front-ends performance summary.

requiring spread spectrum techniques implemented in a low-cost technology.

Whereas increasing the data-rate will help reducing the fixed power cost coming mainly from the synthesizer, it has been shown that when the idle current is dominant (at duty cycle lower than 0.1%), low data-rate is a better choice.

Between the two most common SS techniques (DSSS and FHSS), an FHSS system has been found as the most robust and simple architecture to be implemented in a "micro-Watt node". The possibility to use the FSK modulation and a switching type PA, together with smaller wake-up time and no need for an AGC system, offers an unique possibility to lower the overall power consumption by an order of magnitude.

Problems arise due to the complexity of the hopping synthesizer in terms of accuracy and settling time in the frequency synthesis. A new architecture, based on the direct synthesis of frequency bins, is proposed in this article. The proposed architecture reduces the complexity of the hopping synthesizer, reducing by a factor 8 the power consumption with respect to state-of-the-art hopping synthesizers. The required accuracy in the synthesis of the frequency bins is obtained by a combination of digital pre-distortion and offset robust demodulation techniques. In this way, an absolute-accurate synthesizer is obtained and no feedback loop is used.

To demonstrate the feasibility of the digital pre-distortion concept, an FHSS transmitter has been realized in SOA technology. Two front-ends have been realized, one based on the common combination of VCO and divider and a second one, that reduces the RF front-end to a single block by combining the VCO and the PA.

Measurement results showed the possibility to achieve a BER smaller than 1.1% at -25 dBm transmitted power, NLOS condition in a common office environment with a distance between TX and RX antennas of around 8 meters. An overall power consumption for the complete FHSS transmitter as low as 2.4 mW at -18 dBm transmitted power has been reported. This is 6 times less than the state-of-the-art FHSS transmitters (see Table 1).

Table 1. FHSS products comparison

Company/Type	Operating Voltage	Frequency band	Current consumption
Chipcon CC1050	2.1-3.6 V	300-1000 MHz	10 mA @ 868 MHz, -5 dBm
Analog Devices ADF7010	2.3-3.6 V	902-928 MHz	20 mA @ 0 dBm
Chipcon CC2550	1.8-3.6 V	2.4 GHz	12.8 mA @ -12 dBm
TI Dolphin	2.2-3.6 V	902-928 MHz	35 mA @ +7 dBm
Nordic nRF24E2	1.9-3.6 V	2.4 GHz	10.5 mA @ -5 dBm
Nordic nRF2402	1.9-3.6 V	2.4 GHz	10 mA @ -5 dBm
Nordic nRF24L01	1.9-3.6 V	2.4 GHz	7 mA @ -18 dBm
THIS WORK [a]	**1.8-2.0 V**	**902-928 MHz**	**1.1 mA @ -18 dBm, 2.1 mA @ -5 dBm**

[a]Coarse calibration DAC switched off

Acknowledgment

This work is supported by Senter-Novem (IOP), The Netherlands, in the framework of IOP projects. The authors want to thank Pieter Harpe and Hans Hegt for the useful discussions on the topic and the help with the review of the article.

References

[1] H. D. Man, "Ambient intelligence: Gigascale dreams and nanoscale realities," in *International Solid-State Circuit Conference*, Feb. 2005, pp. 29–35.

[2] J. A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," in *IEEE Pervasive Computing*, vol. 34, Jan. 2005.

[3] M. Honkanen, A.Lappetelinen, and K. Kiveks, "Low end extension for Bluetooth," in *IEEE Radio and Wireless Conf.*, Sept. 2004, pp. 199–202.

[4] P. Choi and *et al.*, "An experimental coin-sized radio for extremely low-power WPAN (IEEE 802.15.4) application at 2.4 GHz," *IEEE J. Solid-State Circuits*, vol. 38, pp. 2258–2268, Dec. 2003.

[5] C. Park, J. Liu, and P. H. Chou, "Eco: an ultra-compact low-power wireless sensor node for real-time motion monitoring," in 4^{th} *Int. Symposium on Information Processing in Sensor Networks*, Apr. 2005, pp. 398–403.

[6] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in 4^{th} *Int. Symposium on Information Processing in Sensor Networks*, Apr. 2005, pp. 364–369.

[7] B. P. Otis, Y. H. Chee, R. Lu, N. M. Pletcher, and J. M. Rabaey, "An ultra-low power MEMS-based two-channel transceiver for wireless sensor networks," in *Symposium on* $VLSI$ *circuits*, June 2004, pp. 20–23.

[8] C. Enz, A. El-Hoiydi, J.-D. Decotignie, and V. Peiris, "WiseNet: An ultralow-power wireless sensor network solution," in *Computer*, vol. 37, Aug. 2004.

[9] B. H. Calhoun and *et al.*, "Design considerations for ultra-low energy wireless microsensor nodes," *IEEE Trans. Comput.*, vol. 454, pp. 727–740, June 2005.

[10] A. Y. Wang and C. G. Sodini, "A simple model for wireless microsensor transceivers," in *Global Telecommunications Conference*, Nov. 2004, pp. 3205–3209.

[11] P. Heydari, "A study of low-power ultra wideband radio transceiver architectures," in *Wireless Communications and Networking Conference*, Mar. 2005, pp. 758–763.

[12] I. D. O'Donnell and R. W. Brodersen, "An ultra-wideband transceiver architecture for low power, low rate, wireless systems," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 1623–1631, Sept. 2005.

[13] R. H. T. Bates and G. A. Burrell, "Towards faithful radio transmission of very wideband signals," *IEEE Trans. Antennas Propagat.*, vol. AP-20, pp. 684–690, Nov. 1972.

[14] J. Ryckaert and *et al.*, "Ultra-wide-band transmitter for low-power wireless body area networks: design and evaluation," *IEEE Trans. Circuits Syst. I*, vol. 52, pp. 2515–2525, Dec. 2005.

[15] L. Stoica and *et al.*, "An ultrawideband system architecture for tag based wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 1632–1645, Sept. 2005.

[16] J.-P. Curty and *et al.*, "Remotely powered addressable UHF RFID integrated system," *IEEE J. Solid-State Circuits*, vol. 40, pp. 2193–2202, Nov. 2005.

[17] R. G. Vaughan, N. L. Scott, and D. R. White, "The theory of bandpass sampling," *IEEE Trans. Signal Processing*, vol. 39, pp. 1973–1984, Sept. 1991.

[18] M. R. Yuce and W. Liu, "A low-power multirate differential PSK receiver for space applications," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 2074–2084, Nov. 2005.

[19] B. Otis, Y. H. Chee, and J. Rabaey, "A 400 μW-RX, 1.6 mW-TX super-regenerative transceiver for wireless sensor networks," in *International Solid-State Circuit Conference*, Feb. 2005, pp. 396–397.

[20] A. Vouilloz, M. Declercq, and C. Dehollain, "A low-power CMOS super-regenerative receiver at 1 GHz," *IEEE J. Solid-State Circuits*, vol. 36, pp. 440–451, Mar. 2001.

[21] D. Theil and *et al.*, "A fully integrated CMOS frequency synthesizer for Bluetooth," in *Radio Frequency Integrated Circuits Symposium*, May 2001, pp. 103–106.

[22] G. Chang, A. Rofougaran, M.-K. Kuand, A. A. Abidi, and H. Samueli, "A low-power CMOS digitally synthesized 0-13 MHz agile sinewave generator," in *International Solid-State Circuit Conference*, Feb. 1994, pp. 32–33.

[23] T. A. D. Riley, M. A. Copeland, and T. A. Kwasniewski, "Delta-sigma modulation in fractional-N frequency synthesis," *IEEE J. Solid-State Circuits*, vol. 28, pp. 553–559, May 1993.

[24] N. Christoffers and *et al.*, "High loop-filter-order $\Sigma-\Delta$-fractional-N frequency synthesizers for use in frequency-hopping-spread-spectrum communication-systems," in *International Symposium on Circuits and Systems*, vol. 2, May 2003, pp. 216–219.

[25] P.-I. Mak and *et al.*, "Frequency-downconversion and IF channel selection A-DQS sample-and-hold pair for two-step-channel-select low-IF receiver," in 10^{th} IEEE International conference on Electronics, Circuits and Systems, vol. 2, Dec. 2003, pp. 479–482.

[26] ——, "Frequency-downconversion and IF channel selection A-DQS sample-and-hold pair for two-step-channel-select low-IF receiver," in *International Symposium on Circuits and Systems*, vol. 1, May 2004, pp. 1068–1071.

[27] V. Kalyanaraman and *et al.*, "A power dissipation comparison of ALU-architectures for ASIPs," in *ECCTD 2005*, vol. 2, Sept. 2005, pp. 217–220.

[28] D. Liu and C. Svensson, "Power consumption estimation in CMOS VLSI chips," *IEEE J. Solid-State Circuits*, vol. 29, pp. 663–670, June 1994.

[29] Analog Devices, "AD7392 data sheet." [Online]. Available: www.analog.com

[30] E. Lopelli, J. van der Tang, and A. van Roermound, "System-level analysis of an ultra-low power, low data-rate FHSS transceiver," in *ProRisc*, Nov. 2004. [Online]. Available: http://www2.stw.nl/programmas/prorisc/proc2000/

[31] E. Lopelli and *et al.*, "A FSK demodulator comparison for ultra-low power, low data-rate wireless links in ISM bands," in *ECCTD 2005*, vol. 2, Sept. 2005, pp. 259–262.

[32] S. Hara and *et al.*, "A novel FSK demodulation method using short-time DFT analysis for LEO satellite communication systems," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 625–633, 1997.

Analog Circuit Design

High-Speed A-D Converters, Automotive Electronics and Ultra-Low Power Wireless

Edited by Arthur H.M. van Roermund, Herman Casier, Michiel Steyaert

Analog Circuit Design contains in total 18 tutorials. They reflect the contributions of 6 experts in each of the three fields covered by the three chapters mentioned in the subtitle, as presented at the 15th workshop on Advances in Analog Circuit Design (AACD) held in Maastricht, April 2006.

This book is number 15 in this successful series of Analog Circuit Design, providing valuable information and excellent overviews of analog circuit design and related CAD, mainly in the fields of basic analog modules, mixed-signal electronics, AD and DA converters, RF systems, and automotive electronics.

Analog Circuit Design is an essential reference source for analog circuit designers and researchers wishing to keep abreast with the latest developments in the field. The tutorial coverage also makes it suitable for use in an advanced design course.

ISBN 978-1-4020-5185-2
9 781402 051852

springer.com