Thesis for the Degree of Licentiate of Engineering

CMOS Data Converters for Closed-Loop mmWave Transmitters

Victor Åberg

CHALMERS

Department of Computer Science and Engineering
Chalmers University of Technology
Göteborg, Sweden, 2020

CMOS Data Converters for Closed-Loop mmWave Transmitters
Victor Åberg

Technical report 212L
ISSN 1652-876X

Department of Computer Science and Engineering
Chalmers University of Technology
SE–412 96 Göteborg
Sweden
Telephone: +46–(0)31–772 10 00

Cover: Chip photo of the 2×6 bit RF IQ modulator presented in chapter 3.

Printed by Chalmers digitaltryck
Chalmers Tekniska Högskola
Göteborg, Sweden, 2020

CMOS Data Converters for Closed-Loop mmWave Transmitters

Victor Åberg
Department of Computer Science and Engineering
Chalmers University of Technology

Abstract

With the increased amount of data consumed in mobile communication systems, new solutions for the infrastructure are needed. Massive multiple input multiple output (MIMO) is seen as a key enabler for providing this increased capacity. With the use of a large number of transmitters, the cost of each transmitter must be low. Closed-loop transmitters, featuring high-speed data converters is a promising option for achieving this reduced unit cost.

In this thesis, both digital-to-analog (D/A) and analog-to-digital (A/D) converters suitable for wideband operation in millimeter wave (mmWave) massive MIMO transmitters are demonstrated. A 2×6 bit radio frequency digital-to-analog converter (RF-DAC)-based in-phase quadrature (IQ) modulator is demonstrated as a compact building block, that to a large extent realizes the transmit path in a closed-loop mmWave transmitter. The evaluation of an successive-approximation register (SAR) analog-to-digital converter (ADC) is also presented in this thesis. Methods for connecting simulated and measured performance has been studied in order to achieve a better understanding about the alternating comparator topology.

These contributions show great potential for enabling closed-loop mmWave transmitters for massive MIMO transmitter realizations.

Keywords: CMOS, Data converter, IQ-modulator, mmWave, RF-DAC, SAR ADC

Publications

This thesis is based on the work contained in the following papers:

[A] V. Åberg, C. Fager, and L. Svensson, "Design Considerations and Evaluation of a High-Speed SAR ADC", in *IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC)*, 2018, pp. 1-6.

[B] V. Åberg, C. Fager, and L. Svensson, "A 2×6b 8GS/s 17-24GHz I/Q RF-DAC Transmitter in 22nm FDSOI CMOS", submitted to *IEEE Solid-State Circuits Letters*, 2020, pp. 1-4.

Contents

Acknowledgement

First and foremost, I want to express my deepest gratitude to my main supervisor, Dr. Lars Svensson, for his immense support, deep understanding, and for the perspectives to all the unexpected events that has taken place during this journey. Also, thanks for always being there, guiding me around this uncharted territory.

I want to thank my co-supervisor, prof. Christian Fager, for his immense support, for sharing his deep knowledge about modelling and measurements, and most importantly, for always being there to help.

I want to thank my examiner, Prof. Per Larsson-Edefors, for his support, input, and ideas.

I want to thank my office mates: Dr. Christoffer Fougstedt, Dr. Kevin Cushon, Erik Börjeson, Dr. Erik Ryman, Dr. Mohammed Abdulaziz, and Han Zhou for all the interesting discussion, fun times, and for the help with mastering the CAD tools.

I want to thank Dr. Lena Perterson, Prof. Jan Jonsson, and Alexandra Angerd for your support and understanding when I, at times, was tied up with completing my designs for tape-out, thus not being able to fully help out with teaching.

I also want to thank MEL for their support in the measurement lab. I want to thank Globalfoundries for sponsoring the manufacturing of the 22 nm chips. I want to thank Johan, Thomas, and Jean-Pierre at Keysight for rapidly sourcing measurement equipment and software at times when my measurements were in limbo, caused by unforeseen events. I want to thank my co-workers at the Computer Engineering division for all interesting discussions. I want thank the GHz center at Chalmers, Vinnova, and the industry partners in the ELEMENT project for their financial support and for interesting discussions and input during the project meetings.

In addition, I also want to thank my fellow guiders and board members in Varbergs NSF Scoutkår for their support and understanding

when I, at times, have been too busy to help out with the activities.

Finally, I would like to express my deepest gratitude to my parents Eva and Jan and my sister Anna for their fantastic support throughout this journey.

Victor Åberg
Göteborg, 2020

Acronyms

2D	2 dimensional
5G	fifth generation
A/D	analog-to-digital
ADC	analog-to-digital converter
AM	amplitude modulation
AWG	arbitrary waveform generator
BB	baseband
BER	bit-error rate
CDAC	capacitive digital-to-analog converter
CMOS	complementary metal oxide semiconductor
CW	continuous wave
D/A	digital-to-analog
DAC	digital-to-analog converter
DC	direct current
DE	drain efficiency
DNL	differential non-linearity
DPD	digital pre-distortion
DR	dynamic range
DSP	digital signal processing
ENOB	effective number of bits
EVM	error vector magnitude
FDSOI	fully-depleted silicon-on-insulator
FF	flip-flop
FFT	Fast Fourier transform
FoM	figure of merit
FoMS	Schreier figure of merit
FoMW	Walden figure of merit
FPGA	field-programmable gate array
GaAs	gallium arsenide

HD	harmonic distortion
I	in-phase
IF	intermediate frequency
INL	integral non-linearity
IQ	in-phase quadrature
ISI	intersymbol interference
LO	local oscillator
LSB	least significant bit
MIMO	multiple input multiple output
mmWave	millimeter wave
MSB	most significant bit
NR	new radio
OSR	oversampling ratio
PA	power amplifier
PAE	power added efficiency
PAPR	peak-to-average power ratio
PCB	printed circuit board
PPF	polyphase filter
PPG	pulse pattern generator
PRBS	pseudorandom binary sequence
PSD	power spectral density
Q	quadrature-phase
QAM	quadrature amplitude modulation
QFN	quad flat no-lead
RF	radio frequency
RF-DAC	radio frequency digital-to-analog converter
RMS	root mean square
RRC	root raised cosine
SAR	successive-approximation register
SFDR	spurious free dynamic range
SiGe	silicon germanium
SNDR	spurious free dynamic range
SNR	signal-to-noise ratio
SRAM	static random-access memory
TI	time interleave
VHDL	very high speed integrated circuit hardware description language
ZOH	zero-order hold

Chapter 1

Introduction

We continuously consume more and more data as we spend more time online: working, communicating with friends and family, and entertaining ourselves. For the last couple of years, a yearly 60 % increase in global mobile traffic has been observed [1]. This trend is expected to continue for the coming years. To support this increased demand, new technologies are needed.

Traditionally, frequencies in the sub-6 GHz band has been used for mobile communication [2]. However, these frequency bands are highly crowded, supporting not only mobile communication but also satellite communication, defence systems, and aeronautical systems just to mention a few [2]. Allocating additional frequency resources for mobile communication in order to support the increased demands is therefore hard. In the newly-introduced fifth generation (5G) new radio (NR) mobile communication standard, millimeter wave (mmWave) bands in frequency range 2 (24.5–40 GHz) are allocated for multi-Gb/s communication [3].

An important component in the realisation of 5G NR mmWave transmitters is the use of beamforming and massive multiple input multiple output (MIMO) [4]. Arranging tens or hundreds of antennas in arrays enables multi-beam operation, allowing for multiple parallel channels using the same spectral resources [5]. The high power consumption resulting from assigning a transceiver to each antenna element in the array is seen one of the major challenges that needs to be solved before the deployment of mmWave massive MIMO systems [6].

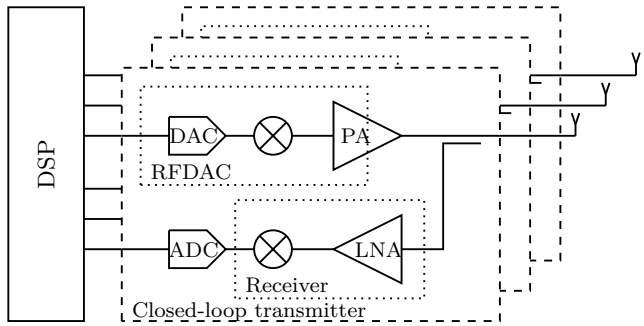


Figure 1.1: Illustration of generic closed-loop transmitter in a MIMO setting.

1.1 Closed-loop mmWave transmitter

Realizing a mmWave massive MIMO transmitter where each antenna element is individually driven, is a major challenge. This not only due to the power consumption, but also due to the small wavelength of mmWave signals, resulting in small antenna separation, thus limiting the allowable size of each individual transceiver.

Using higher-complexity digital techniques to compensate for analog impairments is seen as a key enabler for relaxed hardware requirements, allowing for reduced size and increased efficiency [7]. To train the digital pre-distortion (DPD) algorithms, they need to know the behaviour of the transmitted signal, thus requiring a feedback path [8]. A generic closed-loop array transmitter is shown in figure 1.1. It shows how the transmitted signal is tapped and feed back into the digital signal processing (DSP), adapting the transmitted signal to reduce the effect of the analog impairments. This feedback path can either be in constant operation, or be used during shorter training sequences. In any case the feedback path must handle the entire transmitted bandwidth, thus requiring analog-to-digital converters (ADCs) with both high sample rate and high bandwidth.

To further reduce the transmitter cost, alternative architectures are needed, combining multiple analog functionalities into a single building block [6]. Radio frequency digital-to-analog converters (RF-DACs) are seen as a promising option, combining the digital-to-analog converter (DAC), mixer, and to some extent the power amplifier (PA) into a single building block, bringing a significantly reduced transmitter cost [9, 10].

1.2 CMOS integration

Integration of mmWave circuits have for long been done in compound semiconductors, such as silicon germanium (SiGe) and gallium arsenide (GaAs), as these provide excellent analog performance [11, Ch. 11]. GaAs is often chosen for its higher output power [12] while SiGe enables tighter integration [11, Ch. 11].

Thanks to its aggressive scaling, tight integration, and power efficient operation, there is no technology that can compete with complementary metal oxide semiconductor (CMOS) when it comes to the realization of digital circuits.

In massive MIMO, a tight integration between analog and digital components is needed to achieve a low transmitter cost, making CMOS the only viable option. This forces analog building blocks, operating at mmWave frequencies to also be implemented in CMOS [13].

1.3 Goal

The work reported in this thesis addresses fully-depleted silicon-on-insulator (FDSOI) CMOS implementations of both the data converters shown in figure 1.1. The goal is to explore the design space for circuits at the required performance level, using state-of-the-art silicon processes; and also to provide contributions in the testing and characterization of the experimental circuits.

1.4 Thesis outline

The continuation of the thesis is outlined as follows: Theoretical framework and fundamental concepts for designing and evaluating RF-DACs are presented in chapter 2. The design and evaluation of an radio frequency (RF) in-phase quadrature (IQ) modulator is then presented in chapter 3. These two chapters focuses on the design and evaluation of RF-DAC based IQ modulators. The two following chapters will focus on the design and evaluation of an ADC, starting with the theoretical framework, fundamental concepts and topological alternatives in chapter 4. The design and evaluation of a successive-approximation register (SAR) ADC will then follow in chapter 5. The thesis will then be concluded in chapter 6, summarizing the contributions and looking at future work.

Chapter 2

Background on RF digital-to-analog conversion

This chapter will present the theoretical framework needed to understand the principles behind a modulator. The chapter will also cover different topological options for RF-DAC based IQ modulators and metrics for evaluating them.

2.1 Modulation theory

The modulator is an essential part of a transmitter; it is responsible for placing the information that is to be transmitted onto the carrier. A continuous wave (CW) signal cannot transmit any information since its behaviour is static over time, although it is alternating within a period. The amount of information that can be transmitted on a white Gaussian noise channel is given by the Shannon theorem, equation (2.1) [14]. This equation gives the total channel capacity C, given the channel bandwidth B and signal-to-noise ratio (SNR).

$$C = B \log_{10}(1 + SNR) \quad [\text{b/s}] \tag{2.1}$$

Below, the theory behind modulation will be studied in more detail, starting with the mathematical definition. The principle behind modulation will then be described, followed by the principles behind pulse shaping.

2.1.1 Mathematical definition of modulation

The mathematical definition of modulation is a multiplication between a time-varying signal $v_m(t)$ and the CW carrier $v_c(t)$, given by equation (2.2). Let us first study amplitude modulation (AM), using a sine wave as the modulating signal v_m, equation (2.3). Multiplying the CW carrier, equation (2.2), with the modulation, equation (2.3), gives the modulated signal in equation (2.4).

$$V_c = v_c \sin(2\pi f_c t) \tag{2.2}$$

$$v_m(t) = v_a \sin(2\pi f_s t) \tag{2.3}$$

$$v(t) = v_a \sin(2\pi f_s t) \cdot v_c \sin(2\pi f_c t) = v_a v_c \sin(2\pi (f_c \pm f_s) t) \tag{2.4}$$

Finally, let us now consider a time-varying amplitude and phase signal v_m, given by equation (2.5). Multiplying this signal to the carrier, equation (2.2), then gives equation (2.6).

$$v_m(t) = v_a(t) \sin(\varphi(t)) \tag{2.5}$$

$$v(t) = v_a(t) \sin(\varphi(t)) \cdot v_c \sin(2\pi f_c t) = v_c v_a(t) \sin(2\pi f_c t \pm \varphi(t)) \tag{2.6}$$

In a communication system, the time-varying signals, $v_a(t)$ and $\varphi(t)$, represents the movement between discrete points, so called constellation points, located in the complex plane. All the constellation points for a given modulation format forms a constellation diagram, representing all the values that $v_a(t)$ and $\varphi(t)$ can take.

2.1.2 Modulation principle in time and frequency

Based on the mathematical definition above, we can now study the waveforms used to form modulated signals, both in time and frequency domain.

In the time domain, we have the CW signal shown in figure 2.1a. The data to be transmitted is represented by a sequence of symbols, each having a unique combination of amplitude and phase defined by its

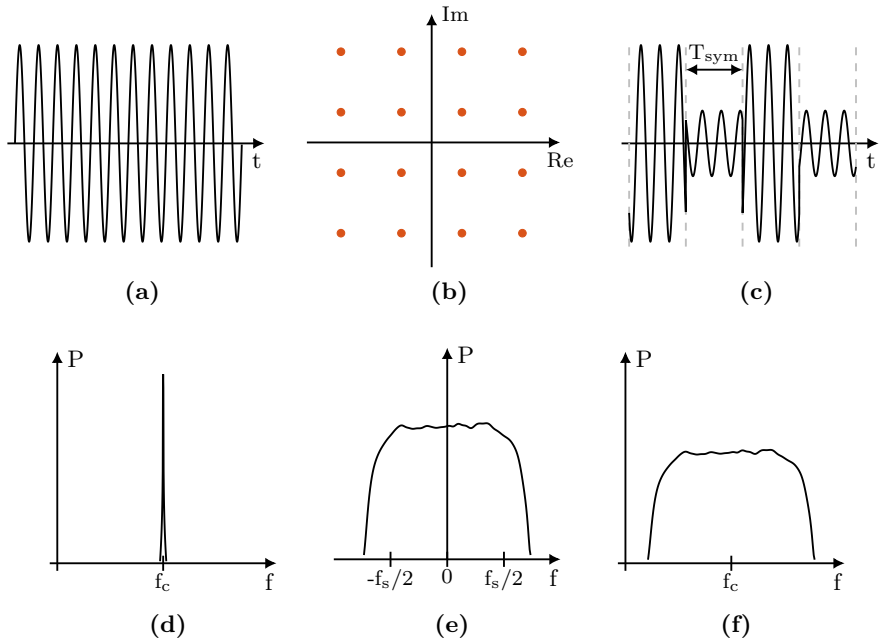


Figure 2.1: Illustration of the modulation principle in both time (a)-(c) and frequency (d)-(f) domain.

location in the constellation diagram, as shown in figure 2.1b. This time-varying amplitude and phase signal is multiplied with the CW signal, forming the modulated signal shown in figure 2.1c. In this figure we can observe how the amplitude and phase change at uniformly separated instances in time, with a spacing equal to the symbol duration T_{sym}.

Now, let us look at the principle behind modulation in the frequency domain. The CW signal is represented by a narrow pulse in figure 2.1d, ideally consisting of a single frequency, although noise causes it to spread out. From equation (2.6), we know that the frequency domain baseband (BB) data will appear at both sides of the carrier. The BB frequency domain spectrum presented in figure 2.1e, uses a oversampled pulse-shaping filter to provide a band-limited signal (the signal in figure 2.1c does not use such a filter). Here, two samples per symbol are used, thus relating the sample rate to the symbol duration as follows: $T_{sym} = 2/f_s$. The spectral behaviour is the same, both for fully random signals and random sequences with finite length, given that the latter is long enough

to appear random. In the time domain, the modulation was applied through a multiplication. The corresponding operation in the frequency domain is an addition between the BB signal and the carrier, thus, only shifting the center frequency of the BB signal, locating it around the carrier f_c as shown in figure 2.1f.

From figure 2.1e, it appears like the BB signal bandwidth is doubled, as the signal covers the band $[-f_s/2, f_s/2]$. For real signals, such as those used in AM, the spectrum is symmetrical around the carrier, thus, only a single side-band needs to be transmitted. For complex signals, such as those used in quadrature amplitude modulation (QAM), the spectrum is asymmetrical around the carrier, making both side-bands important. The asymmetrical spectrum is a result of combining the two orthogonal signals which make up the complex QAM signals. As one DAC is needed on the transmit-side and one ADC is needed on the receive-side, for each orthogonal component, the effective sample rate is doubled as the orthogonal signals are processed in parallel, thus resulting in the increased bandwidth.

2.1.3 Modulation formats

Above we presented how data is modulated on to the carrier. The chosen modulation format will not only affect the complexity of the modulator implementation. It will also set the amount of data that can be transmitted for a given bandwidth. The spectral efficiency S is a metric of the number of bits that can be transmitted per spectral resource. Equation (2.7) gives the spectral efficiency for a bit-rate R_b and a bandwidth BW.

$$S = \frac{R_b}{\mathrm{BW}} \quad [\mathrm{b}/(\mathrm{Hz\,s})] \tag{2.7}$$

With limited spectral resources, a high spectral efficiency is important in order to achieve a high bandwidth. The spectral efficiency is listed in table 2.1 for a few different modulation formats with varying complexity.

Each symbol in a modulated signal is represented by a unique amplitude, or by a unique phase, or by a unique combination of these two. A constellation diagram shows all the symbols that can be transmitted, for a given modulation format, arranging them after their amplitude and phase. The constellation diagrams for a few different modulation

Table 2.1: Spectral efficiency for different modulation formats.

Modulation format	Spectral efficiency [b/(Hz s)]
OOK	1
PAM-4	2
QPSK	2
16-QAM	4
32-QAM	5
64-QAM	6
256-QAM	8
1024-QAM	10

formats of varying complexity are shown in figure 2.2. The principle behind a few of them will be described below.

The simplest AM format is OOK, shown in figure 2.2a, which builds on toggling the carrier on and off according to the bit-stream. This is however not a very effective modulation format as can be seen in table 2.1. The principle behind OOK can be extended further with the introduction of additional amplitude levels and sign. A constellation diagram for a PAM with four levels is shown in figure 2.2b. The spectral efficiency is given as $\log_2$ of the number of levels used. Clearly, it can be seen that the number of levels needed rapidly increase with demands for higher spectral efficiency.

PAM4 introduces the phase as a means to transfer information. By further using the phase it becomes possible to construct 2 dimensional (2D) constellations, giving higher spectral efficiency while requiring fewer levels along each direction. The simplest 2D modulation format is QPSK, shown in figure 2.2c, which modulates only the phase of the carrier (the amplitude is constant regardless of the transmitted data). By introducing additional amplitude levels to the 2D modulation formats, we introduce QAM. This set of modulation formats was first introduced in 1960 [15], and few years later, the square QAM typically seen today was introduced [16]. A few different QAM formats are show in figures 2.2d to 2.2f, containing 16, 32, and 64 constellation points respectively. For these modulation formats, the number of levels needed along each orthogonal direction is the square-root (rounded up) of the

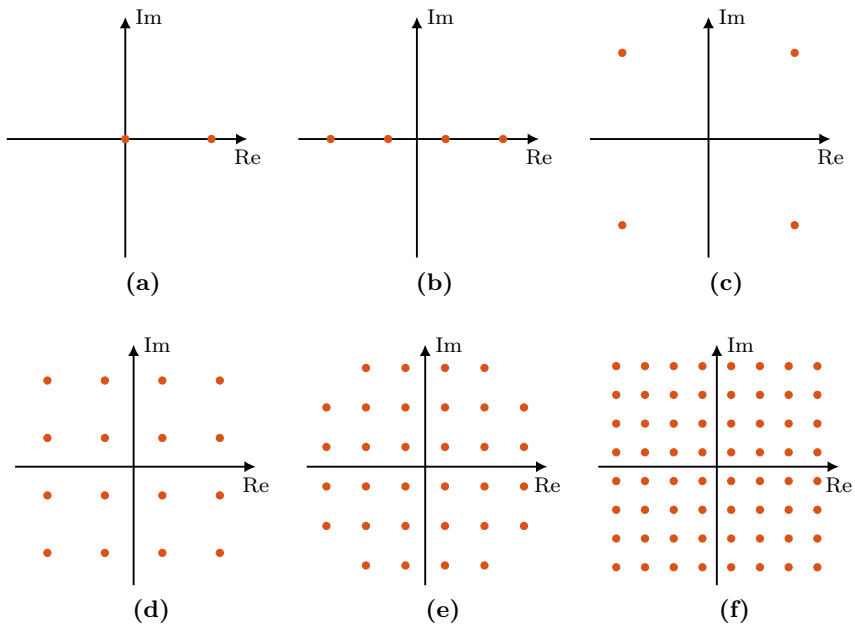


Figure 2.2: Illustration of a few commonly used modulation formats, ranging from simple formats, such as OOK (a), PAM (b), and QPSK (c), to more complex QAM formats (d)-(f).

number of constellation points for the given modulation format, clearly showing a large gain compared to using PAM. If the grid contains more locations than the number of constellation points provided by the modulation format, the locations in the corners are not used, as can be seen in figure 2.2e. The reason is that these locations suffers more from non-linearities in the transmitter as these locations have the highest signal power, thus making them more suitable to exclude.

The mapping of digital data onto the constellation points is typically done using Gray encoding [17, Ch. 3]. The benefit with this encoding is that the bit sequence mapped to each constellation point only differs by one bit for the closest located constellation points. Thereby the bit-error rate (BER) is minimized when detecting the received data as the constellation points most likely to be erroneously detected are the ones located closest to the intended value. Mapping principles can be used to reduce the peak-to-average power ratio (PAPR), where scrambling methods, such as block level codes are popular [18].

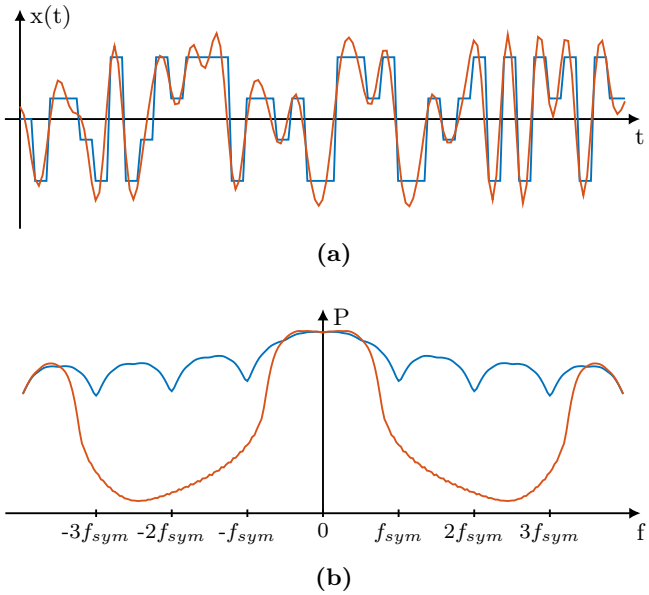


Figure 2.3: Time domain behaviour is shown in (a) for an AM signal outputted by a DAC modelled by a ZOH function (blue) and the corresponding pulse-shaped signal (orange). The frequency domain behaviour is shown in figure 2.3b.

2.1.4 Pulse-shaping

The DAC output can be modelled by a block that connects a digital code to an analog value and a block that keeps this value constant for the entire sample duration. The time domain output signal for a DAC modelled by a zero-order hold (ZOH) function is shown in figure 2.3a together with the corresponding signal fed through a pulse shaping filter. In the figure it can been seen how the ZOH gives a box-shaped signal. The signal passing through the pulse-shaping filter is smother, at the cost of overshooting the range. The frequency domain behaviour is shown in figure 2.3b for the two signals. The ZOH function here translates into a sinc function with zeros at every multiple of the symbol frequency f_{sym}. The oversampled and pulse-shaped signal suppresses the signal further outside the transmitted band but the images around f_s cannot be compensated for.

In modern communication systems, requirements are not only tough on the signal quality within the transmitted band, but also in the neigh-

bouring bands, making it important to suppress the out-of-band emissions, such as the ones caused by the ZOH. In addition, non-linear distortion in the channel introduces intersymbol interference (ISI), making symbols dependent on each other. By properly selecting a filter, the signal can both be band-limited and ISI can be prevented at the sampling instances [19, Ch. 8]. The root raised cosine (RRC) filter is popular in communication systems as the same filter can implemented both in the transmitter and in the receiver for an overall ISI-free channel.

Combining the pulse-shaping filter with oversampling introduces additional points along the path from one constellation point to the next, bringing a smoother signal [17, Ch. 4]. The oversampling ratio (OSR) is defined as the ratio between the sample rate and the symbol rate in the system. Oversampling also simplifies the requirements on the analog filter needed to suppress images, as they are further separated by the increased sample rate [17, Ch. 4]. While oversampling mainly affects the out-of-band properties, it can also improve the in-band performance. As the transition rates when moving from one constellation point to another is reduced, smother transitions are allowed, stressing the circuit less. To fully benefit from oversampling, an increased DAC resolution is needed in order to provide additional levels in between the constellations points, used to smoothen the transition from one symbol to the next.

A drawback with using oversampling is that it increases the required sample rate for the DAC for a given symbol rate, or it reduces the symbol rate for a given sample rate. Achieving both high bandwidth and high OSR is hard as circuit technology limits the achievable sample rate for DACs.

2.2 Figures of merit

To fairly compare different designs, fair comparison measures that does not favour a single topology are needed. While there exist several generally accepted figures of merit (FoMs) for ADCs, as described in section 4.3, this is not the case for RF-DACs. One reason is that there is no fixed output interface for RF-DACs, as output power and load impedance are part of the specifications. Another reason is that RF-DACs can be used to generate modulated signals, making it highly beneficial to co-integrate large portions of a modulator into a single block, further complicating the comparison of different designs. In addition, they can be designed for different output power levels, operate

at different frequency bands and be optimized for different modulation formats, making a fair comparison even tougher.

There are several traditional DAC measures which in theory could be of interest also for RF-DACs, such as linearity and spurious free dynamic range (SNDR). However, the linearity is a static measure, not giving any insight about the performance for wideband signals. It might also be complex to accurately measure the steady-state linearity for a RF-DAC due to a potentially high carrier frequency. The narrow-band signals needed for SNDR measurements would also not mimic the actual wideband performance. Below, a few different FoMs commonly used for evaluating RF-DACs will be presented.

2.2.1 Error vector magnitude

The error vector magnitude (EVM) is a measure of how accurate the transmitted constellation is compared to the intended one. EVM is typically measured as the root mean square (RMS) error normalized to either the average constellation power $\mathrm{EVM}_{\mathrm{rms}}$, or to the highest power for any constellation point $\mathrm{EVM}_{\mathrm{rms,peak}}$.

An illustration of the error vector for a single point is shown in figure 2.4. From this figure, the EVM for a single point can be calculated as in equation (2.9), where equation (2.8) shows the definition of the error vector. This equation can then be extended into the average metrics commonly used for performance comparison. Equation (2.10) shows the expression for $\mathrm{EVM}_{\mathrm{rms}}$ while equation (2.11) shows the expression for the $\mathrm{EVM}_{\mathrm{rms,peak}}$. The metric normalized to the RMS power makes it possible to compare the performance for different modulation formats [20, Ch. 5] while the one normalized to the peak constellation point is commonly adopted in instruments [21, 22].

$$\mathrm{IQ}_{err} = \mathrm{IQ}_{meas} - \mathrm{IQ}_{ref} \tag{2.8}$$

$$\mathrm{EVM}\,[n] = \frac{|\mathrm{IQ}_{err}\,[n]|}{\left|\mathrm{IQ}_{ref}\,[n]\right|} \tag{2.9}$$

$$\mathrm{EVM}_{rms} = \frac{\mathrm{rms}\left(|\mathrm{IQ}_{err}|\right)}{\mathrm{rms}\left(\left|\mathrm{IQ}_{ref}\right|\right)} \tag{2.10}$$

$$\mathrm{EVM}_{rms,peak} = \frac{\mathrm{rms}\left(|\mathrm{IQ}_{err}|\right)}{\max\left(\left|\mathrm{IQ}_{ref}\right|\right)} \tag{2.11}$$

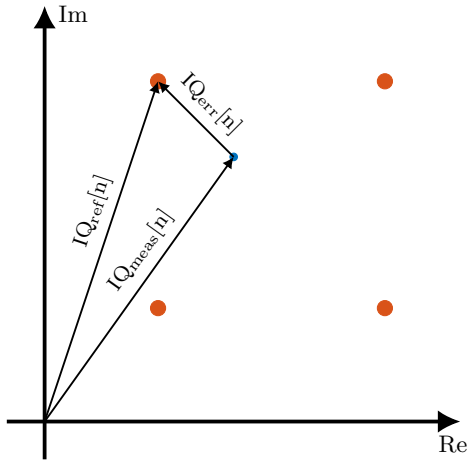


Figure 2.4: Illustration of EVM calculation for a single point.

In a real communication system, the received constellation diagram will be affected by noise coming from three sources: the transmitter, the channel, and the receiver. This noise will introduce errors in the transmitted data stream. The BER is a measure of how many erroneous bits that are received in relation to the total number of received bits. The BER is also tightly connected to the error in the received constellation diagram. By connecting the BER to the EVM through the SNR, it becomes possible to study the relationship between BER and EVM in a Gaussian channel. For QAM, the bit error probability can be calculated using equation (2.12), by inserting equation (2.13), assuming Gaussian noise [20, Ch. 2]. In the following equations, M is the number of constellation points in the modulation format, BW is the bandwidth of the system, and BR is the bit rate. $Q(x)$ is the Q-function, often used in error-probability expressions [20, Ch. 2].

$$P_b(M) = \frac{2P_{e\sqrt{M}}}{\log_2 M}\left(1 - \frac{1}{2}P_{e\sqrt{M}}\right) \tag{2.12}$$

$$P_{e\sqrt{M}} \cong \frac{2\left(\sqrt{M}-1\right)}{\sqrt{M}}Q\left(\sqrt{3\frac{\log_2 M}{M-1}\frac{\mathrm{BW}}{\mathrm{BR}}\mathrm{SNR}}\right) \tag{2.13}$$

For a channel where bit errors are caused by additive white Gaussian noise [23], the SNR will correspond to the average signal power over the

average error. The relationship between $\mathrm{EVM_{rms}}$ and SNR is given by equation (2.14).

$$\mathrm{EVM}_{rms} = \frac{1}{\sqrt{\mathrm{SNR}}} \tag{2.14}$$

In order to relate the average power to the peak constellation point a scale factor is needed. The crest factor c, relates the magnitude of the maximum constellation point to the average constellation magnitude, while PAPR relates the corresponding power ratio. Equation (2.15) shows the expression commonly used to describe this modulation format dependent ratio [24–26]. Here, M is the number constellation points for a given modulation format. As EVM is inversely proportional to the signal power, $\mathrm{EVM_{rms,peak}}$ will therefore be related to $\mathrm{EVM_{rms}}$ as in equation (2.16).

$$c = \sqrt{\mathrm{PAPR}} = \sqrt{3\frac{\sqrt{M}-1}{\sqrt{M}+1}} \tag{2.15}$$

$$\mathrm{EVM_{rms,peak}} = \frac{1}{c}\mathrm{EVM_{rms}} \tag{2.16}$$

The relationship between EVM and BER is shown in figure 2.5 for several different modulation formats of different complexity. The EVM is normalized to the average constellation for these curves. In high-capacity communication systems, an input BER of 10^{-3} is typically considered as the DSP typically includes error correcting capabilities, thereby further reducing the BER [23].

From the diagram in figure 2.5 we can observe several things. We can see that as the number of constellation points is quadrupled, the required EVM must be halved for a constant BER. It can also be observed that the slope is larger for a BER of 1×10^{-6} than for a BER of 1×10^{-3}, making the system more sensitive to noise when BER requirements are tougher.

Although EVM gives a good measure on the accuracy of the transmitted constellation, the measure still comes with its limitations when it comes to comparing different designs. As it does not account for the DAC resolution, it is important to keep this in mind when comparing figures. As the modulator resolution is connected to the complexity of the intended modulation format, the ratio between the number of required levels and the number of levels available must also be considered. In theory, a high resolution modulator should be able to provide

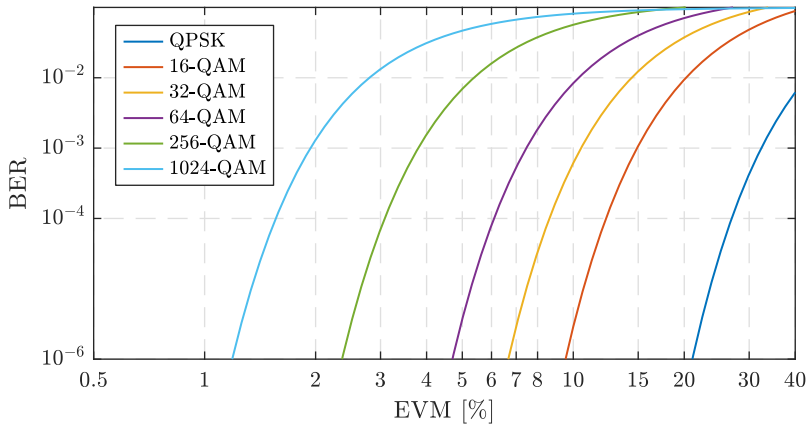


Figure 2.5: BER plotted versus EVM for different modulation formats with EVM normalized to average power.

a better EVM than a modulator with lower resolution given the same sample rate and modulation format. It also is important to account for the transmitted data rate as the non-linear behaviour of the modulator might be further stressed by wideband signals.

2.2.2 Efficiency metrics

Efficiency measures are often of interest when comparing highly integrated components. For PAs, both drain efficiency (DE) and power added efficiency (PAE) are common measures used to compare the efficiency of the implementation.

$$\mathrm{DE} = \frac{P_{RF,out}}{P_{DC}} \tag{2.17}$$

$$\mathrm{PAE} = \frac{P_{RF,out} - P_{RF,in}}{P_{DC}} \tag{2.18}$$

While DE only accounts for the efficiency at the output the PA equation (2.17), PAE also accounts for the input power and the power consumed in potential pre-amplifier stages as in equation (2.18). This makes PAE a much more usable metric. However, for RF-DACs, all the power consumption will not be located in the analog path as some functionality needs to be implemented as digital circuits. It is not clear

if this consumption should be included in the PAE or not, making the measure hard to use.

Another complication is the vast span of topologies used for implementing RF-DACs. To some extent, the principles for achieving a high output power are the same both for RF-DACs and for PAs. Again, comparing with PAs, a large span of different PA topologies exist. When comparing PA implementations, one tries to keep the topology as similar as possible, as the topology might have significant impact on the efficiency. For RF-DACs on the other hand, the relatively low number of publications makes this kind of categorization impossible as other things not related to the output power also need to be considered, such as the data rate.

2.2.3 Energy consumption per bit

The energy consumed per transmitted bit is a comparison metric used when comparing the performance both in entire communication systems and in individual subsystems [9]. This is a good method as it makes it possible to study effects on a system level even though focus might be on a significantly smaller block. There are however subsystems which are less suitable for this kind of comparison, such as RF-DACs, as the output power level is not compensated for. Subtracting the output power will not be sufficient as the power consumption is highly dependent on the intended output power. Other parts of the RF-DAC must also be designed for the intended output power, thus connecting the power consumption for the rest of the RF-DAC to the output power level.

2.3 Topological alternatives

Transmitters can be highly customized depending on their intended operation. Two transmitter topologies, the direct conversion and the superheterodyne, have traditionally been adopted in communication systems [27, Ch. 14]. This however changed with the introduction of wideband RF-DAC-based transmitters [28–30]. Since their introduction in 2004, there has been an increased interest for RF-DAC-based transmitters, driven by increased demands of tighter system integration. Both Cartesian [9, 10, 31–37] and polar [38–42] transmitters has been demonstrated, operating over a large range of frequencies ranging from a few GHz [28–30, 40–42] to 60 GHz and above [32–39].

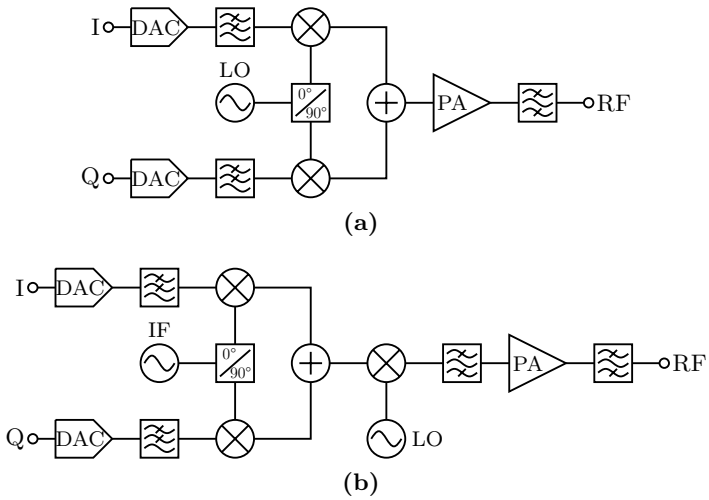


Figure 2.6: Illustrations of the direct converting transmitter (a) and the superheterodyne transmitter (b).

2.3.1 The classical transmitter approach

The two transmitter topologies traditionally used are both based on similar principles, up-converting the BB data using mixers [27, Ch. 14]. Generic block diagrams for the direct-converting and the superheterodyne transmitter are shown in figures 2.6a and 2.6b respectively. The direct-converting transmitter is compact as it only requires a single up-conversion. However, as only a single up-conversion is used, potential local oscillator (LO) leakage will end up in the center of the transmitted band, making it impossible to filter out. The superheterodyne transmitter solves this problem by introducing an intermediate frequency (IF). Carefully selecting the IF and LO frequencies makes it possible to filter out any LO leakage as the IF provides an offset between the transmitted band and the LO. This however comes at the cost of additional hardware.

Although both topologies in principle are fairly simple, many different versions are presented in literature, addressing different weaknesses with these topologies. These will however not be studied here.

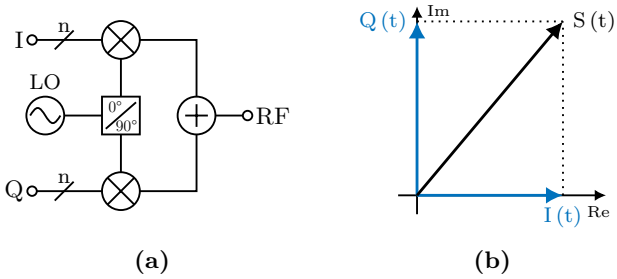


Figure 2.7: Illustration of the RF-DAC based Cartesian modulator (a) and the vector combining (b).

2.3.2 Cartesian RF-DAC

The Cartesian RF-DAC based modulator is to a large extent based on the direct-converting transmitter topology presented above. However, there is one significant difference; rather than providing analog BB data to the mixer using DACs, digital control signals are used to control the operation of the mixer. To allow this control of the mixer, it is divided into several smaller cells that are individually controlled. Through this individual control, it becomes possible to combine the up-conversion with the digital-to-analog (D/A) conversion. The similarities between the two topologies can clearly be seen when comparing the RF-DAC-based topology in figure 2.7a with the direct-converting topology in figure 2.6a. In figure 2.7b the vector combining of the in-phase (I) and quadrature-phase (Q) signals is shown.

2.3.3 Polar RF-DAC

The idea behind the polar RF-DAC-based topology is very similar to the Cartesian topology. However, rather than using two RF-DACs to realize the orthogonal signals in the Cartesian coordinate system, the polar topology operates based on a polar coordinate system, thus only requiring a single RF-DAC for the amplitude modulation [39]. The phase modulation is applied to the LO signal at an earlier stage. A block diagram for a generic RF-DAC-based polar transmitter is shown in figure 2.8a. The vector summation of the phase φ and amplitude A is shown in figure 2.8b.

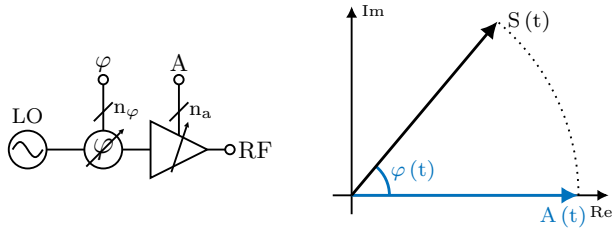


Figure 2.8: Illustration of the RF-DAC based polar modulator (a) and the vector summation (b).

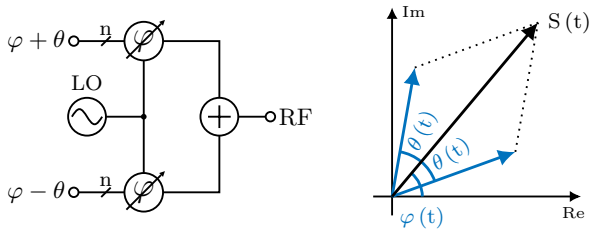


Figure 2.9: Illustration of the RF-DAC based outphasing modulator (a) and the vector summation of the outphasing signals (b).

2.3.4 Outphasing transmitter

The principle behind outphasing differs a bit from the two approaches presented above. It focuses on keeping a constant amplitude level, eliminating the need for an amplitude modulator. Rather the phase modulation is used to generate the amplitude levels needed, making it possible to use highly efficient amplifiers rather than RF-DACs [43].

The modulation is applied using two parallel phase modulators, one having a positive phase offset and the other a negative one. Each modulator has a fixed magnitude of half the desired maximum output level. The amplitude modulation is achieved by adjusting the outphasing angle θ and then combine the two channels. A block diagram for the outphasing transmitter is shown in figure 2.9a. The vector summation of the outphasing signals is shown in figure 2.9b.

2.3.5 Topology comparison

The Cartesian topology has the advantage that it operates in the same coordinate system as the complex modulation formats commonly used. The square grid making up a M-QAM constellation, see figure 2.2, closely matches the grid formed by the Cartesian modulator. However, a weakness with the topology is when it is desired to adjust the phase of the transmitted signal, such as in beam-forming MIMO systems [44–46]. Rotating the constellation would result in a mismatch between the grid generated by the modulator and the constellation to be transmitted, reducing the modulation accuracy. However, as phase adjustments for beam movements are occurring at a less frequent rate than the transmitted data is updated, it becomes possible to use an external phase modulator located either before or after the Cartesian modulator.

The polar topology on the other hand with its circular grid, is more suitable for systems requiring phase rotation. However, when a static phase is needed, the polar topology becomes less suitable for square-shaped constellations as a large set of the available points will not be used. Fitting a square into a circle will always result in areas of the circle not being covered by the square. Another drawback with the polar topology is the complexity of implementing phase modulators. RF-DACs have been demonstrated to operate at high speed with moderate resolution [9, 39]. The same cannot be said about phase modulators. In some cases, Cartesian based modulators have been used as phase modulators [39, 40], although this in not their primary function as they simultaneously provide both amplitude and phase modulation. Transmission-line-based phase modulators have been demonstrated [47–49], but these were intended for beam-forming applications and cannot provide the sample rate required for polar modulation. The phase modulation waveform also needs to be synced to the amplitude signal with sub-sample precision for good performance [39], which further complicates the implementation of a high-bandwidth polar modulator.

The outphasing modulator is very similar to using two parallel polar transmitters, although no amplitude modulation is needed. As no amplitude modulation is needed, highly efficient amplifiers could be used, thereby in principle result in a high efficiency [43, 50]. However, high-speed phase modulation is a problem that the outphasing topology shares with the polar topology. Another problem with the phase modulation is the non-linear signal processing [51], resulting in a large bandwidth expansion, thus requiring a large OSR which is problematic

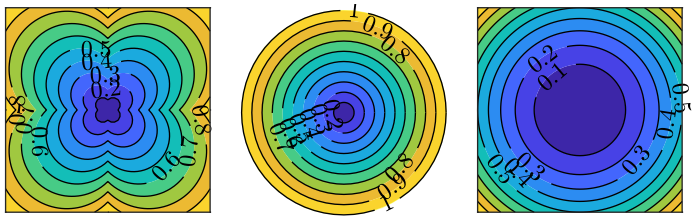


Figure 2.10: Contour plots of the normalized DE for the different modulator topologies, comparing the code dependency for Cartesian (a), polar (b), and outphasing (c).

for wideband modulators. A rule of thumb is that polar transmitters require roughly an oversampling of ten times the signal bandwidth to account for the bandwidth expansion [52]. However, polar transmitter realizations requiring a lower OSR have been demonstrated [39].

2.4 Code dependent efficiency

For RF-DAC-based modulators, not only the peak efficiency is of interest; also the code-dependent efficiency is important, especially for modulated signals with a large PAPR. On average, the output power for modulated signals is significantly lower than the peak output power, thus resulting in a significant efficiency reduction [10]. Figure 2.10 shows contour plots of the code dependent DE for, Cartesian, polar and outphasing modulators normalized to the peak DE achieved by each topology.

From figure 2.10 it might be hard to see how the PAPR caused by the modulation will affect the efficiency. The relative modulated DE is plotted versus PAPR in figure 2.11 for the three transmitter topologies. It can here be seen how the outphasing transmitter achieves the lowest efficiency while the Cartesian and polar are fairly similar.

2.5 Non-overlapping LO

In most analog systems, the duty cycle is not considered as pure sine wave signals are more desired than pulse-shaped signals. For pulse-

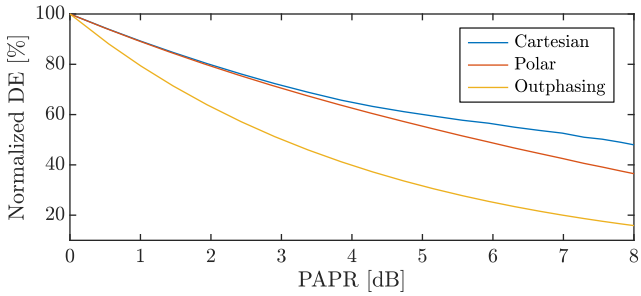


Figure 2.11: Modulation dependent DE as a function of PAPR, ploted for Cartesian, polar and outphasing transmitters. The DE is normalized to the peak for each individual transmitter topology.

shaped signals, the duty cycle is defined as the ratio between the duration of the pulse and the signal period at half the pulse magnitude. As pulse-based signals have a high harmonic content, they might cause undesired interference. Mixers are however the exception, where strong LO signals are desired for more idealized operation [11, Ch. 13]. Here it is typically not a problem to use 50 % duty cycle signals, making it possible to use clipped sine wave signals to drive the LO port on the mixer. Since RF-DACs can be seen as a group of parallel mixers, one might think that 50 % duty cycle would be fine for them as well. This is however only true for polar transmitters, which only use a single RF-DAC, thus not having any summation on the output. For other RF-DAC-based topologies, the outputs from several parallel RF-DACs are summed to form the desired signal.

The different waveforms shown in figure 2.12 are calculated based on the four LO signals, to generate the top right point in figure 2.4. The two differential signal are defined as in equation (2.19) where RF_{diff+} is generated by LO_0 and LO_{90}. The same also goes for RF_{diff-}, which is generated by LO_{180} and LO_{270}. The resulting single-ended signal is then given by equation (2.20). The 50 % and 25 % duty cycle cases will be presented in more detail below.

$$\mathrm{RF}_{diff+} = \mathrm{LO}_0 + \mathrm{LO}_{90} \tag{2.19}$$

$$\mathrm{RF}_{sing} = \mathrm{RF}_{diff+} - \mathrm{RF}_{diff-} \tag{2.20}$$

The Cartesian topology relies on the summation of the orthogonal

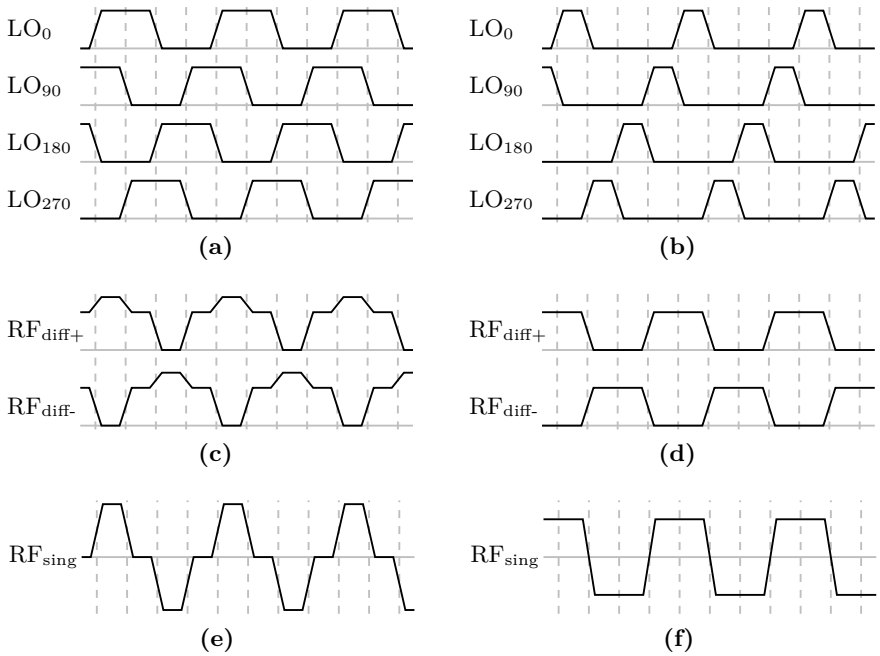
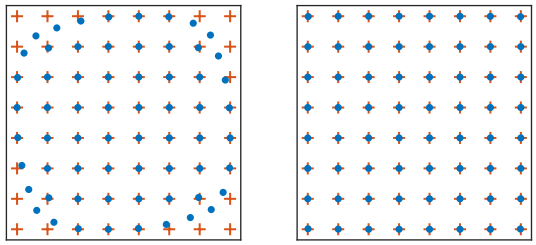


Figure 2.12: Quadrature LO signals with 50 % duty cycle (a) and 25 % duty cycle (b). LO_0 and LO_{90} are then combined to form RF_{diff+} (c) and (d). The resulting RF signal is shown for the 50 % and 25 % cases in (e) and (f) respectively.

signals generated by the parallel RF-DACs. However, complete orthogonality can only be achieved if there is no interaction between the RF-DACs [33]. Interaction-free operation will however not be fulfilled for 50 % duty cycle LO signals [33]. As can be seen in figure 2.12a, the four LO signals needed in a fully differential quadrature modulator overlap with each other, thus causing intermodulation [33]. The extreme cases when both RF-DACs are fully on become most troublesome as the largest interference occurs here, although the problem is present for all output combinations.

In figure 2.12c, RF_{diff+} is generated from LO_0 and LO_{90} to form the top right most point in figure 2.4. Combining the 50 % duty cycle LO signals to form the differential components results in a large duty cycle expansion as illustrated in figure 2.12c. To mimic the effects of saturation in the transistors when both LO signals are combined, a smaller



(a) (b)

Figure 2.13: Effects on output constellation when directly combining the RF signals at the RF-DAC outputs, using 50 % duty cycle LO signals (a) and 25 % duty cycle (b) LOs signals. The reference grid is shown in orange.

than ideal increase is shown. Combining the differential signals in figure 2.12c into a single-ended signal, shown in figure 2.12e, results in large portions of the differential signals cancel each other with a reduced efficiency as the result. The resulting waveform, with a small duty cycle, will have a high harmonic content that needs to be filtered out.

With the 25 % duty cycle LO signals shown in figure 2.12b, it can be seen that the LO signals do not overlap, thereby ideally combine into the differential signals shown in figure 2.12d. This results in fully differential signals, not overlapping each other at any time, bringing a nice 50 % duty cycle RF signal when combined into the single-ended signal shown in figure 2.12f.

Constellation diagrams for modulated signals using 50 % and 25 % duty cycle LO signal are shown in figure 2.13. With 50 % duty cycle LO signals, a significant compression can be seen for the corner points, significantly reducing the EVM. This effect cannot be seen for 25 % duty cycle LO signals as they are non-overlapping.

The principle of 25 % duty cycle LO signals has been demonstrated both in mixer applications [53–55] and in RF-DAC based transmitters [33], although at low frequency. Wideband LO generators for 25 % duty cycle operating at DC–62 GHz [56] have also been demonstrated, albeit in SiGe process.

The RF-DAC efficiency is tightly connected to the duty cycle, although this might not be fully intuitive. Consider the differential outputs for the RF-DAC in the 50 % duty cycle case shown in figure 2.12c.

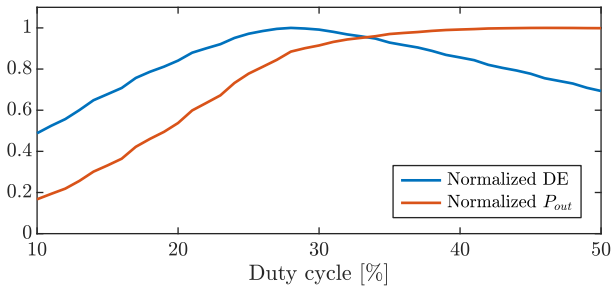


Figure 2.14: Normalized DE and output power as a function of the duty cycle.

Here it can be seen that the RF-DACs are active 75 % of the time, resulting in a conduction angle of 75 %, which translates into an amplifier operating in class-AB. Ideally, it would be desired to have class-B operation, which corresponds to a conduction angle of 50 %. As 25 % duty cycle LO signals are fully isolated, a conduction angle of 50 % is achieved, as can be seen in figure 2.12d.

In figure 2.14, the normalized DE and output power is plotted as a function of duty cycle. Depending on for which duty cycle a saturated output power is desired, the peak seen for the DE will move. It will however not be possible to get a maximum efficiency at a duty cycle of 50 %.

Chapter 3

RF IQ modulator design and evaluation

RF IQ modulators have been presented to operate at a large set of frequency bands, ranging from a few GHz up to 60 GHz and beyond [9, 10, 23, 31–34, 36, 37, 39, 57, 58]. In [9], a RF-DAC based IQ modulator is described, having a resolution of 2×6 bit, operating at 20–32 GHz with a sample rate of 5 GS/s, although simulations indicate a sample rate approaching 16 GS/s. A 2×10 bit RF-DAC based IQ modulator operating at 20–32 GHz with a sample rate of 2 GS/s is demonstrated in [10]. A polar modulator having a 4 bit resolution, operating at 60 GHz is demonstrated in [39]. In this design, the amplitude modulator is clocked at 10 GHz. The phase modulation is implemented using a Cartesian up-converting mixer where analog BB data is fed onto the chip, generating the phase modulated LO signal. A 2×9 bit RF-DAC based IQ modulator is demonstrated with an operating range of 85–95 GHz and a clock rate of 15 GHz [34]. This IQ modulator has however only been verified using OOK, by toggling the most significant bit (MSB). A RF-DAC based IQ modulator with a 2×2 bit resolution, operating at 94 GHz with a sample rate of 20 GHz is demonstrated in [23].

I will below describe the realization of a 17–24 GHz, 2×6 bit RF IQ modulator, operating at 10 GS/s. A block diagram of the modulator is shown in figure 3.1. A LO at two times the desired carrier frequency is fed into the divider which generates the quadrature LO signals. These are then fed through inverter buffers to boost the drive-strength. Two RF-DACs are used to generate the I- and Q-components before being combined at the output. An on-chip memory is used to provide the data

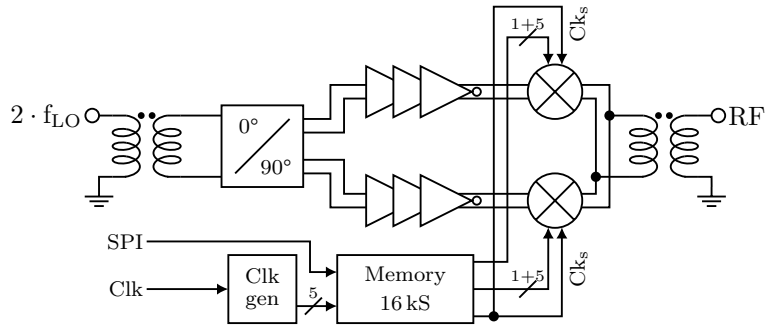


Figure 3.1: Block diagram showing the implementation of the RF IQ modulator.

streams that are fed into the RF-DACs.

3.1 Proof of concept

Before implementing the entire RF IQ modulator, a simpler RF-DAC was implemented as a proof of concept, testing the principle. This RF-DAC used a 5 bit resolution and only supported magnitude modulation. In this proof of concept, data was externally fed onto the chip together with a sample clock using an field-programmable gate array (FPGA), providing a quick evaluation of the principles.

This evaluation highlighted the complexity of having multiple differential high-speed signals connected to the chip, not only significantly impacting the layout of the chip but also the testability. Providing six parallel data streams while simultaneously probing the RF signals proved too complex, highlighting the limitations with this test-setup.

3.2 Unit cell

The unit cell is the lowest level building block in RF-DACs. It can be used to form thermometer based implementations where each cell has its own control signal. Also, it can be used to form hierarchical blocks, where several unit cells are grouped together, for example into bit-cells. This flexibility makes the unit cell usable in many types of implementations. The basic principle behind a mixing unit cell is a set

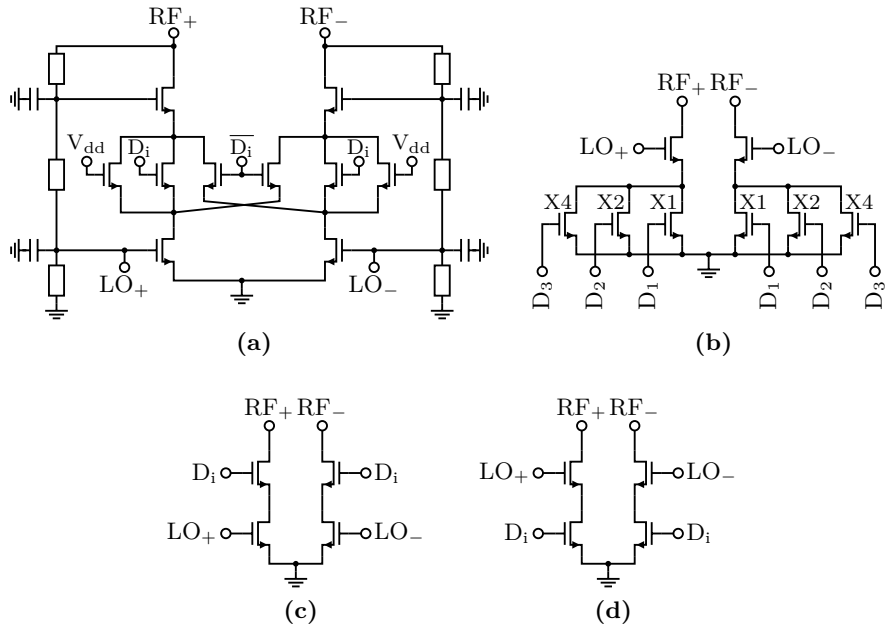


Figure 3.2: Examples of different unit cell arrangements seen in literature. In (a) an arrangement for constant bias current is shown. The shared LO principle is shown in (b). The individually paired configurations with the data switch on top (c) and below (d) the LO-switch.

of stacked transistors, one pair for LO signals and another pair for data signals.

Implementations described in literature sometimes directly go for binary weighted structures and do not rely on unit cells [9]. There are many different considerations behind the realization of a RF-DAC; the topology is one of them. The important point is however that all implementations require some sort of scaled structure for realizing the output levels (with thermometer based implementation being the exception). Many different topologies for arranging the LO- and data-switches are presented in literature; three alternatives will be presented below.

In stacked topologies where cascodes are used to increase the output power, a cancelling unit cell topology is commonly used. In this topology, the data-switches are placed on top of the LO-switches, and three parallel data-switches are used in each differential leg. Two of these data-switches are connected to one of the LO signals and the third one

is connected to the LO signal having opposite polarity [9,37], as shown in figure 3.2a. One transistor controls the bias level, while the others control the output level through RF summation or cancellation. As two transistors constantly are active, a constant bias current is achieved for all digital control signals. This constant bias condition is important in cascodes as the supply voltage must be evenly distributed over all the stacked transistors making up the cascode.

Another approach, that shares some principles with the stacked topology, uses multiple parallel data switches, all sharing a common LO-switch. The data-switches are placed in the bottom, controlling the bias current through the LO switch [23,31], as shown in figure 3.2b. While only a single set of LO-switches are needed, the topology suffers from a large parasitic capacitance in the node connecting the data-switches to the LO-switches.

The third principle builds on combining a data-switch and a LO-switch, either placing the data-switch on the top as in figure 3.2c [10,36] or by placing the data-switch in the bottom as in figure 3.2d [39]. These arrangements brings a reduced parasitic capacitance between the LO- and data-switch compared to the case when the LO-switch was shared. The topologies presented in figures 3.2b to 3.2d all have in common that the total bias current for the RF-DAC is dependent on the number of active unit cells, thus improving the efficiency when bot all unit cells are active.

Although all the topologies presented above are very different, they all only realize the magnitude part of the modulation. The sign, that is the ability to turn the phase 180°, is typically implemented in between the quadrature LO generation and the RF-DACs [9,10,36,37,39]. This sign generation is based on swapping the polarity of the LO signals connected to the RF-DACs This method however requires the sign generation to be synced to the amplitude generation. An alternative approach for the sign generation is to implement it into the unit cells. This can be done by duplicating the magnitude generation while swapping the LO connections, thus doubling the footprint of the unit cell.

The LO leakage cancellation brought by arrangement of the LO- and data-switches greatly impacts the overall RF-DAC performance. Since none of the topologies presented above have sign generation, they all suffer form potentially large LO leakage. Although the cancellation approach builds on suppressing the signal when no output is desired, leakage will still occur due to transistor mismatch. For the case when

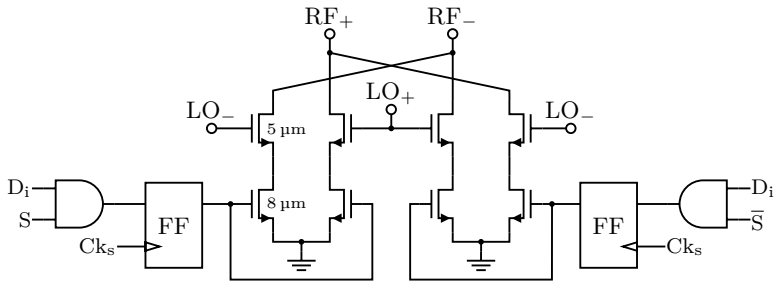


Figure 3.3: A schematic of the unit cell implementation showing both RF path with LO and data switches as well as the FFs that are used to align the data.

the data-switch is placed on top of the LO-switch, cancellation is brought by the blocking provided by the deactivated data-switch [10,36]. Further LO leakage cancellation is claimed to be achieved by also gating the LO signal reaching the LO-switch [10]. This would however add additional delays for activation of the cell, making it less suitable for wideband operation.

The unit-cell topology presented in figure 3.3 seeks to mitigate these drawbacks. The unit cell uses top placed LO transistors and slightly larger data switches placed below. The size of the LO-switches are set by the desired output power, while slightly larger data-switches are used to improve the efficiency. For the RF-DAC to be able to handle wideband signals and high OSR, it is important that the data-switches can operate at high speed. To reduce effects of jitter in the data switching, flip-flops (FFs) have been placed in the unit cells in order to align all the data switching events to a single signal, the sample clock. The sign generation has also been moved into the unit cell, using two parallel mixers with the LO signals swapped, as seen in figure 3.3. Placing the sign generation within the unit cell also brings inherent LO leakage cancellation at unit-cell level.

The largest contribution to LO leakage is the capacitive coupling between the gate and drain terminals of the LO-switches. Through the introduction of an additional mixer, a leakage-path with opposite polarity is introduced, thus significantly reducing the LO leakage.

The LO leakage has been simulated for the two topologies presented in figures 3.2c and 3.2d, for the case when additional disabling of the LO switches is used [10], and for the unit cell topology presented in

figure 3.3. Schematic level simulations has been performed on the four topologies mentioned above to estimate their LO leakage, using the sizes for the LO- and data-switch in figure 3.3. For an implemented circuit, additional leakage caused by the routing is expected, further degrading the LO leakage performance. As the topology in figure 3.2d suffers from the highest LO leakage, this topology will be used as the reference. The topology presented in figure 3.2c brings a LO leakage reduction of 25 dB. Deactivating the LO signal when the cell is not used brings an additional 20 dB LO leakage reduction compared to the topology in figure 3.2c which this principle is used for. The unit cell presented in figure 3.3 shows excellent LO leakage cancellation, showing a 90 dB leakage reduction compared to figure 3.2d. However, as this principle builds on cancellation of the leakage paths, mismatch is the transistor sizes will significantly increase the leakage. In the case of a $\pm 1\,\%$ transistor size mismatch, the LO leakage performance closely matches the performance achieved by disabling the LO signal driving the switch when it is not in use.

In principle, it is possible to use a single data switch shared by the two LO switches in the mixer, as long as the total size is maintained. However, with the introduction of the sign generation, one signal needs to be crossed in the unit cell layout, as can be understood from figure 3.3, further increasing the complexity. The dual data switches give more freedom for the implementation, allowing the LO signals to be crossed rather than the RF signals.

In modern CMOS processes, the metal routing in the bottom metal layers is very narrow and thin, severely limiting ability to transfer current, especially direct current (DC). With the total power consumed by RF-DACs mainly being consumed in the unit cells, a good ground is required to handle this total current. This is currently the limiting factor for the unit cell size. The intense routing of LO and RF signals above the mixers limits the available area for ground routing.

3.3 RF IQ modulator core topology

The RF IQ modulator core arranges the unit cells in a hierarchy to form the different output levels. Two different hierarchies are often used: the binary topology, where unit cells are combined into binary weighted groups, or the thermometer topology, where the unit cells are individually activated depending on the input code. These topologies are most

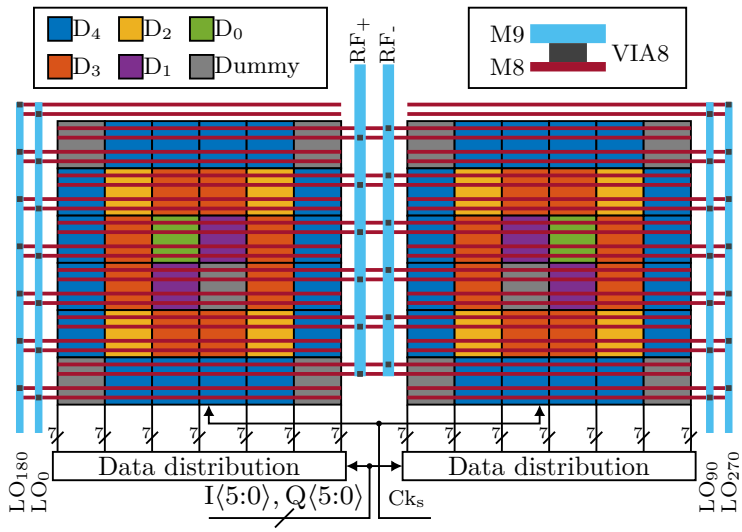


Figure 3.4: Illustration of the unit cell placement for the binary scheme together with the routing of the LO and RF signals. The data distribution blocks contains the binary to thermometer encoder.

commonly used for low and medium resolution RF-DACs. At higher resolutions, a combination of these topologies can often be found, where the MSBs are thermometer encoded and the least significant bits (LSBs) are binary encoded. The RF IQ modulator presented here features the possibility to choose between binary and thermometer encoding.

For the binary encoding scheme, the unit cells are distributed in a common-centroid fashion, as can be seen in figure 3.4. For the thermometer-encoded mode, the cells are activated in a chain starting at the corner where the LO signals are fed in, moving up and down column-wise until reaching the corner where the RF signal is outputted. The Q-block is a mirrored replica of the I-block to further improve matching.

The FFs located in the unit cells, see figure 3.3, reduce skew caused by the data routing by isolating the data switches from the data routing. Instead, the control signals for the data switches are derived by the sample clock. This makes the clock signal Ck_s the only signal that needs to be carefully aligned at the input of all unit cells. The clock signal is routed to the center of the RF-DAC before being distributed through a H-tree in order to reduce the switching skew at the unit cells. Dummy loads are placed at the locations in the H-tree where no cells

are connected, further balancing the clock distribution.

The unit cells are distributed over a fairly large area, which makes it important to consider delays in the routing of the LO and RF signals for optimum combining of the unit cell outputs. The routing is implemented as mirrored F-shaped structures in the top metal layers, as shown in figure 3.4. This routing is designed to give equal delay from the LO input to the RF output through any unit cell.

In addition to the LO leakage, phase inversion is something that can be done either inside the unit cell or before it. Generating the sign before the RF-DAC reduces the complexity of the unit cell it self. However, for good performance using wideband signals, the sign generation needs to be precisely synced to the amplitude switching in the unit cells. The exact propagation delay from the phase generation to the unit cell will not be a multiple of the clock period, which makes this alignment complex. An alternative is to move the sign generation into the unit cell, eliminating the need for critical alignment of the sign and data. The sign generation is implemented by duplicating parts of the unit cell or the entire structure, at an obvious cost in cell area.

3.4 LO generation

Circuits for delivering quadrature LO signals can be divided into three categories: 90° hybrids [10, 23, 31, 32], polyphase filters (PPFs) [59], and divider-based topologies [9]. The hybrid-based quadrature LO generators are often implemented using transmission-lines, although a lumped realization is an option. A quarter wavelength transmission-line will realize a 90° phase shift. With increased frequency, the absolute length decreases, making the hybrid approach most suitable for high frequency implementations. In air, the wavelength for a 30 GHz signal is 1 cm and on chip, it is roughly a quarter of this, making a $\lambda/4$ transmission-line roughly 700 µm long. Although there are methods for reducing the size of 90° hybrids [60], their size will still be substantial below 30 GHz. Another limitation with 90° hybrids is that they can only generate 50 % duty cycle LO signals.

High frequency PPFs are typically implemented as passive structures, using resistors and capacitors. This topology features significant loss, requiring amplifiers to boost the signal level. Another limitation with the PPF topology is that it can, just like the 90° hybrid, only generate 50 % duty cycle LO signals.

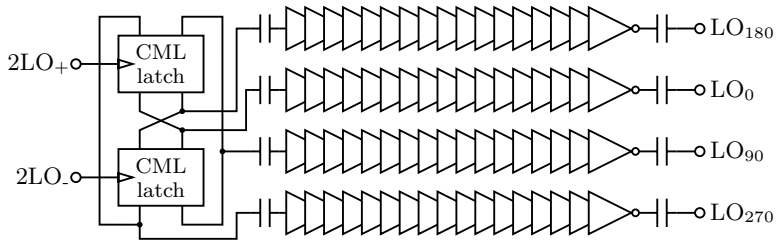


Figure 3.5: Schematic of the quadrature LO generation including the divider and DC-block capacitors.

The divider-based topology can in theory operate over the largest relative frequency band. It builds on using two cross-coupled latches, clocked with opposite sign. The two latches will output signals at half the clock frequency having 90° phase difference. In principle, divider-based quadrature LO generators have no lower frequency limit for their operation. However, due to the double input frequency required, there is an upper frequency limit for which the topology will work. Divider based quadrature LO generators have been demonstrated to generate upto 32 GHz LO signals in CMOS [9] and 62 GHz in SiGe [56]. With the limited driving capability of the divider, significant buffering is needed. However, thanks to the switching behaviour of the divider it is possible to use inverter-based buffers, thus making it suitable for CMOS integration.

The divider-based topology allows for a tunable duty cycle as the bias level can be adjusted at the buffer input. This is good as transistors often have a threshold voltage smaller than half the supply voltage, effectively increasing the duty cycle. The 50 % duty cycle signals generated by the hybrid and PPF can also be used to generate non-overlapping LO signals using AND-gates [56], however, this results in a fixed duty cycle.

A schematic of the implemented LO generator is shown in figure 3.5. Due to its potential for high operational bandwidth and ability to provide non-overlapping LO signals, the divider based quadrature LO generation was chosen. Two cross-coupled dividers are used to generate the non-overlapping LO signals. These signals are then fed through four parallel 17 inverter buffers to increase the drive strength. The final two inverters in the buffer chain have the same size in order to sharpen the output transitions.

3.5 Testability

The testing and evaluating of a design is just as important as the design it-self. Already when designing the circuit, it is important to make sure that it can be evaluated in a good way. The test infrastructure must be flexible in order to support several different test cases. The data streams used to control the RF-DACs can be generated using several different sources, each featuring its own set of challenges. The data streams can be externally provided to the chip, using either a FPGA or a waveform generator. An alternative is to use generators implemented on the chip, either pseudorandom binary sequence (PRBS) generators or memories.

Most RF-DACs presented in literature use externally fed data for controlling the functionality [10, 31, 34–37, 39]. This method simplifies the on-chip design, as the data stream can be directly fed to the RF-DAC. By serializing the data stream, fewer parallel interfaces are needed. This is at the cost of demultiplexers needed to parallelize the data. The problem here is the complexity of providing and aligning a large number of high-speed control signals without electrically and practically affecting the RF measurements. There are cases where limitations in the external test equipment have limited the possibilities to fully test the chip [34].

Externally-fed data streams can either be provided by an arbitrary waveform generator (AWG) or a pulse pattern generator (PPG), however, these typically come with a limited number of ports, limiting the number of parallel channels that can be used. An alternative here would be to use an FPGA, as the number of parallel channels here can be increased. It will however be more complex to fine-tune the delays on a FPGA and a significant design effort is needed for the implementing the test system on the FPGA.

An alternative to externally feeding the data onto the chip is to use an on-chip memory that stores a record of data points [9,33]. This record is then used to control the RF-DACs during the measurements, giving almost the same flexibility as the external option. There are however two limitations with the on-chip memory solution; the record length is limited by the amount of memory that can fit onto the chip, and the transmitter cannot be used in a real link, as the record length would be too short. While the on-chip solution brings a simpler measurement set-up, it comes at the cost of a larger design effort for the chip as the memory also needs to be integrated together with the rest of the design. An important factor with the memory is that it should not limit

the testability of the circuit, either in terms of data rate or in terms of record length. A too-short memory will significantly limit the testability of the chip [9]. It should be noticed that when an oversampled pulse-shaping filter is used for spectral shaping, the number of transmitted symbols becomes the record length divided by the OSR. To achieve a good statistics for the EVM calculation, the record length will depend on the modulation format used. High complexity QAM formats have a larger number of unique symbols resulting in the need of a longer record and thereby a larger memory.

The method chosen for providing the digital data to the RF-DACs is an on-chip static random-access memory (SRAM) memory. With the 2×6 bit configuration, a total of 12 bits will be needed for each symbol, resulting in a digital bandwidth of 120 Gb/s between the memory and the RF-DACs at a sample rate of 10 GS/s. A 16 kS memory has been chosen, as it provides a good balance between footprint and record length. It consists of two blocks, a memory array and a serializer. The memory array is configured as 1024 words each containing 16 samples. These 16 samples are then serialized through a pipelined multiplexer tree, allowing the memory array to operate at 1/16 of the output clock rate. Programming is performed through a simple serial interface implemented in the memory block. The block has been designed in VHDL, using a SRAM memory array generated by a memory compiler. The design has then been synthesized and placed-and-routed with a target output clock rate of 10 GHz, using a standard digital design flow, providing a standard-cell based implementation of the memory block.

3.6 Manufactured chip

In order to measure the implemented RF IQ modulator it needs to be accessed externally. During the design, all signals that needs to be accessed externally must be connected to a pad. Although the RF-DAC is a differential circuit, measuring differential signals is complex; thus a circuit element, a so called balun, is needed for converting the input signal from single-ended to differential. On the output a balun will also be needed for converting the differential signal into a single-ended one.

The RF IQ modulator has be fabricated in Globalfoundries 22 nm FDSOI CMOS process. The entire design measures 1 mm by 1.18 mm. The modulator core including baluns occupy 225 µm by 650 µm out of which the core itself occupy 225 µm by 190 µm. A chip photo is shown

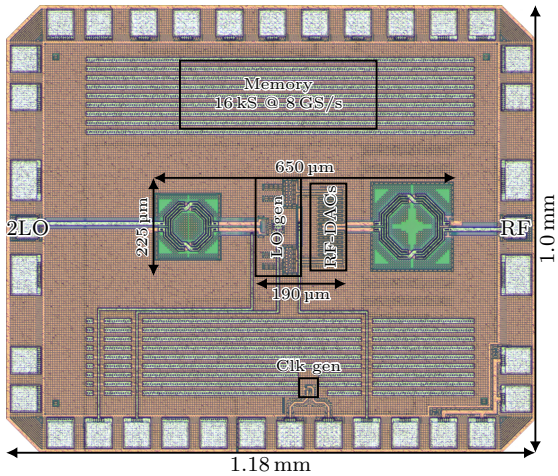


Figure 3.6: Chip photo highlighting the placement of the different building blocks and some key dimensions.

in figure 3.6, highlighting all the essential blocks in the IQ modulator.

3.7 Measurements

The measurements used to characterize the RF IQ modulator can be divided into two parts: static and modulated measurements. These will be presented in detail below.

All measurements have been performed using on-chip probing of the RF signals, while all the DC and control signals are bonded out onto a printed circuit board (PCB) onto which the chip in mounted. The LO is in all measurements provided by a vector network analyzer, while sample clock is provided either by the vector network analyzer or by an external signal generator.

3.7.1 Static measurements

There are two types of static measurements that are of interest: the output power as a function of frequency, and the static RF output and drain efficiency for all the different output codes.

In the power-vs-frequency measurements, it is important to also verify the functional behaviour for modulated signals. Static CW signals

are not as sensitive to imbalances between the I- and Q-channels as modulated signals, even though both channels are fully activated. It could therefore be possible to measure a large CW bandwidth, especially for divider based topologies which in principle do not have a lower frequency limit for proper operation, and still have a functional modulated-signal bandwidth which is only a small fraction of this CW bandwidth. In this sense, the 3 dB bandwidth is misleading since it only indicates the behaviour for CW signals, which cannot carry any information.

During the static measurements described here, the output power has been measured using a Keysight vector network analyzer, with on-chip power calibration. In parallel, a Keysight oscilloscope has been used to measure a narrow-band highly oversampled modulated signal, in order to verify proper functionality at each frequency point.

The static RF behaviour gives a good prediction about what artifacts that will be seen in the modulated measurements. The results cannot only be used for estimating IQ gain and phase imbalances, but also to study potential discontinuities in magnitude or phase. The results collected here can also be combined with the DC power consumption, to calculate the DE.

Measuring the output power and phase is straightforward, given that the input and output frequency is the same. Here, since a divider-based LO generation is used, the definition of phase is unclear. However, since only the relative phase change is of interest, it is possible to use a reference tone at the output frequency for phase comparison. This approach however requires that the LO signal and reference signal sources have a stable long-term phase lock to each other. For these measurements, the vector network analyzer was configured in frequency offset mode, providing the option to have separated transmit and receive frequency ranges [61]. Simultaneously measuring the RF-DAC supply current makes it possible to calculate the efficiency for each output code.

3.7.2 Modulated measurements

To study the behaviour with modulated signals, the signal needs to be analysed both in time and frequency domain, checking both the constellation diagram and spectrum. It is also important to evaluate the performance for various modulation formants, OSRs, and sample rates in order to understand the limitations of the circuit.

For these measurements, the on-chip memory was programmed with the filtered and up-sampled BB signal (these calculations were done off-

chip). The spectrum was measured using a Keysight vector network analyzer in spectrum analyzer mode, while the time domain behaviour is captured using a Keysight oscilloscope, directly capturing the RF signal. The down-conversion and post-processing is then performed offline. This approach makes it possible to get calibrated power spectral density (PSD).

3.8 Performance

The chip has been evaluated both for its static and dynamic properties. Measurements have been performed between 17–24 GHz, with a maximum sample rate of 8 GS/s, limited by the speed of the memory. We are fairly confident about the memory being the limiting factor, as its supply current dropped unexpectedly when the clock frequency was further increased. An increased supply voltage extended the operation range slightly, indicating that this is the limiting factor. The results will be divided into static and dynamic performance according to the description of the measurement process listed in sections 3.7.1 and 3.7.2.

During the measurements, we observed a limited tuning range for the divider and LO buffers. This has limited the possibility to study the benefits of using non-overlapping LO signals. In the measurements, we have observed a behaviour indicating a sub-50 % duty-cycle, but we have not been able to verify the value exactly. Trying to adjust for larger duty-cycle resulted in a large IQ-gain and -phase imbalance.

3.8.1 Static performance

As stated above, the static performance of the RF IQ modulator consists of two parts; the operational bandwidth of the modulator, and the static behaviour for each combination of output codes.

The output power for the CW range is presented in figure 3.7. This is the band in which a narrowband modulated signal can be properly transmitted with a IQ phase error less than $\pm 10°$. The bias voltage at the input of the buffer chain has been adjusted at every measured frequency point to minimize the phase error. Although the RF IQ modulator is functional over the range 17–25 GHz, the 3 dB bandwidth is 17–24 GHz.

While the power consumption for the quadrature LO generation is independent on the modulator's output signal, it is highly dependent on the frequency of the LO. Figure 3.7 also shows the power consumption for the LO generation at different frequencies. This power consumption

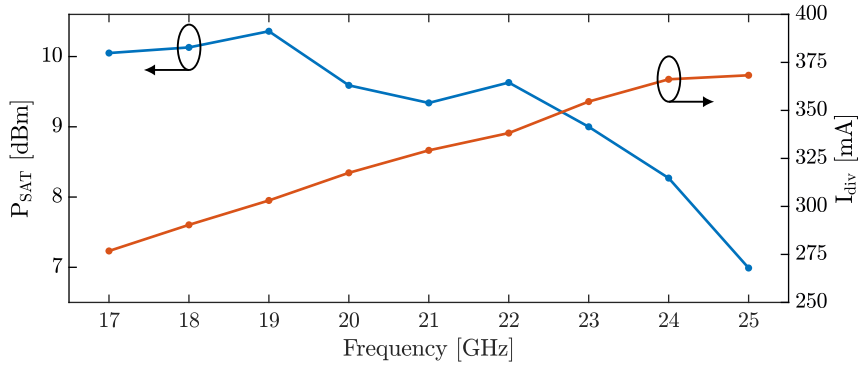


Figure 3.7: Static output power and quadrature LO generator current consumption for some carrier frequencies.

is measured for a supply voltage of 1 V. The increased supply voltage, in contrast to the 0.8 V the buffer was designed for, has been needed for keeping the functionality at the higher carrier frequencies. The peak output power achieved when all cells are active is 10.4 dBm, while consuming 78 mA from a 0.9 V supply. This gives a peak drain efficiency of 15.6 % at 19 GHz.

Figure 3.8 shows the measured output power and phase for all combinations of I and Q at a carrier frequency of 21 GHz[1]. The figure does not use any scales as it is the power and phase that it measured and then converted into a complex magnitude representation. For the conversion to result in a voltage, something needs to be assumed about the load. In the figure, a small gain error can be observed both between the two channels and between the two phases of a single RF-DAC. From the figure it can be observed that positive Q-codes results in a bit larger output power, roughly 1 dB higher than the maximum codes in the other directions achieve. For all the other directions the variation lies within 0.5 dB. It can also be observed that there is a small phase drift between the points belonging to each I code. This behaviour matches the order in which the points are swept. A rotated sweep order have also been evaluated with the same behaviour as described above, but rotated in

[1] As the circuit includes a divider for the quadrature LO generation, there is a risk of having a 180° phase rotation of the measured phase. These events are considered to be caused by the measurement equipment and they are therefore rotated back to the original phase.

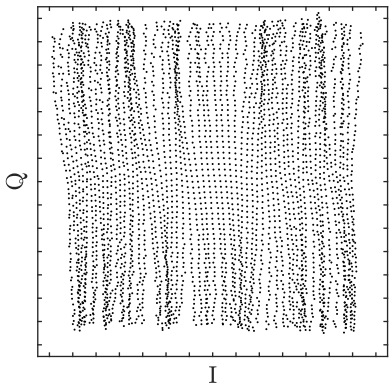


Figure 3.8: Static RF behaviour for all combinations of digital input codes.

the way as the swept order. As the effect follows the swept order, it indicates some type of memory effect or drift in the measurement setup. The source of this effect is not fully understood.

In figure 3.9 a contour plot shows the drain efficiency when the output code is changed. It can here be observed that the efficiency is highest in the corners where both the I- and Q-channel are active. This is the case when non-overlapping LO signals are used. When 50 % duty cycle signal are used, the maximum efficiency is achieved when only one channel is active.

As all the different combinations of inputs are studied, it becomes possible to also check the LO leakage, as the RF-DAC includes the off-state. An LO leakage cancellation of −29 dBc is achieved when all unit cells are off. However, the true LO leakage performance for the unit cells cannot be measured due to additional leakage caused by DAC-level routing. Parasitic extraction based simulations of the DAC-level routing confirms the measured leakage level. The lowest output level is achieved when one unit cell is active in each RF-DAC, giving a LO level of 40 dB below the highest output power. As the LO leakage is only in the range of a LSB it is possible to compensate for on the digital side.

3.8.2 Modulated performance

The modulated measurements have been performed at a few different modulation formats and sample rates using an OSR of 7 and a carrier frequency of 21 GHz. A few different lower OSRs were also tested,

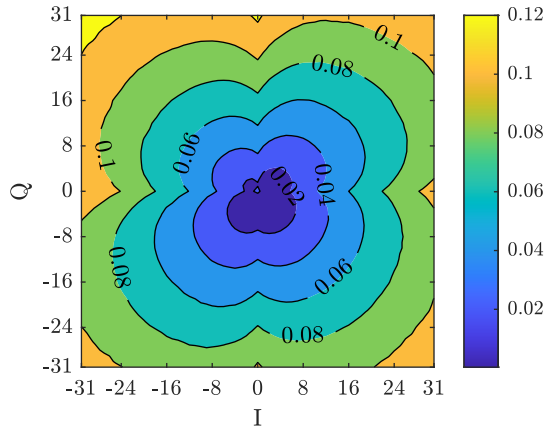


Figure 3.9: Contour plot of the DE for all combinations of input codes.

but this was the one resulting in the highest data rate without having a severe degradation in EVM. Figure 3.10 shows the EVM for three modulation formats, 16-QAM, 32-QAM, and 64-QAM, at different data rates, corresponding to sample rates in the range 1–8 GS/s. In the figure it can be observed that the EVM not only rises for high data rates which is expected, but also for low data rates. This increased EVM at low data rates is thought to be caused by a smaller instability in the circuit that increases the noise level. The exact source has however not been found. For all different combinations of modulation format and sample rate, a simple linear algorithm has been used to compensate for IQ-gain and -phase imbalance and for the DC offset caused by the LO leakage.

With modulated measurements, the divider and buffer chains consume a total of 330 mA from a 1 V supply. The logic and clock distribution in the RF-DACs consumed 20 mA from a 0.925 V supply. The average output power, for a 64-QAM signal is 2.5 dBm, giving a drain current of 28 mA from a 0.9 V supply. This gives a drain efficiency of 7.1 % for the modulated signals. The difference between the static and average output power closely matches the PAPR for the modulated signal.

Figure 3.11 shows constellation diagrams and PSDs for 16-QAM at two different data rates. For figures 3.11a and 3.11b a data rate of 3.66 Gb/s is achieved at a sample rate of 6.4 GS/s. The EVM is in this case 8.0 %. With a sample rate of 8 GS/s, a data rate of 4.57 Gb/s is achieved at an EVM of 12.6 %. The constellation diagram and PSD

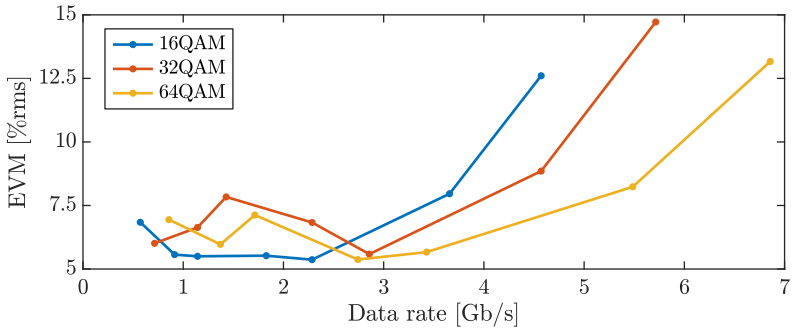


Figure 3.10: EVM for 16-, 32-, and 64-QAM signals, using an OSR of 7 for various different data rates

are shown in figures 3.11c and 3.11d respectively. For a BER lower than 1×10^{-3}, an EVM lower than 15 % will be needed, as shown in figure 2.5.

The constellation diagrams and PSDs for a 32-QAM modulation are shown at two different data rates in figure 3.12. Figures 3.12a and 3.12b shows the constellation diagram and the PSD for a sample rate of 6.4 GS/s giving a data rate of 4.57 Gb/s. This is for an EVM of 8.9 %. With a sample rate of 8 GS/s, an EVM of 14.7 % is achieved with a data rate of 5.71 Gb/s. The constellation diagram and PSD are shown in figures 3.12c and 3.12d. An EVM of less than 10.6 % is needed to achieve a BER lower than 1×10^{-3}, as can be seen in figure 2.5.

In figures 3.13a and 3.13b, the constellation diagram and the PSD are shown for a data rate of 3.43 Gb/s using 64-QAM modulation. This is achieved at a sample rate of 4 GS/s, resulting in an EVM of 5.7 %. An EVM of 8.2 % and a data rate of 5.49 Gb/s is achieved at 6.4 GS/s. The constellation diagram and PSD are shown in figures 3.13c and 3.13d. The EVM requirement for a BER lower than 1×10^{-3} is 7.5 % for 64-QAM, as can be seen in figure 2.5.

In the constellation diagrams at the lower data rates, figures 3.11a, 3.12a and 3.13a, the constellation points closest to the center has tighter grouping compared to the ones located further away from the center. This is especially clear for the 64-QAM modulation shown in figure 3.13a. This is a fully expected behaviour when lying close to the maximum output power level, as the outer points experience a larger non-linearity than the inner ones, thereby making the circuit generate more noise. For the higher data rates this behaviour changes, as can be seen in

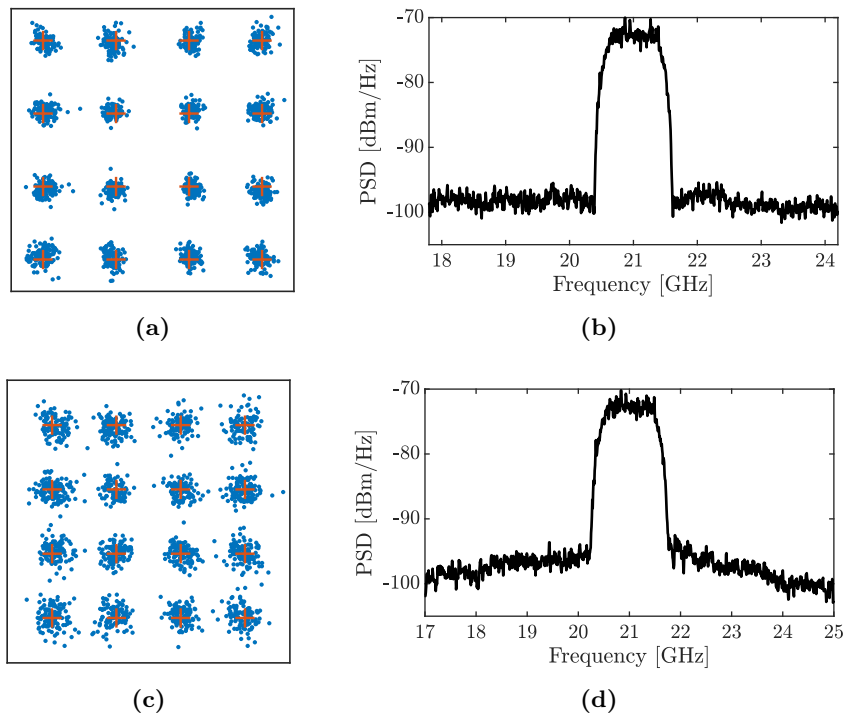


Figure 3.11: Constellation diagram and spectrum for 16-QAM at 3.66 Gb/s, (a)-(b), and at 4.57 Gb/s, (c)-(d)

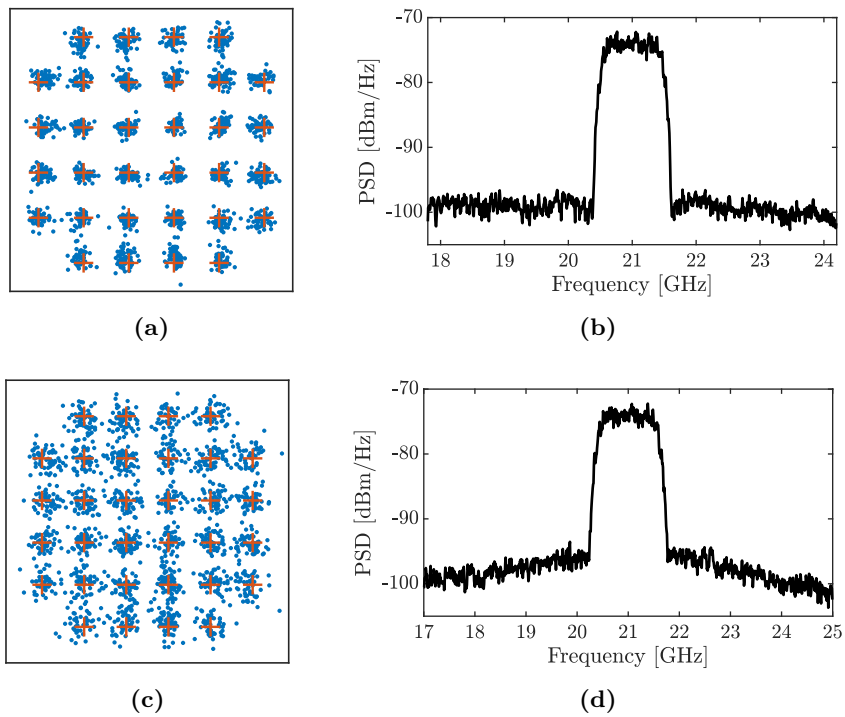


Figure 3.12: Constellation diagram and spectrum for 32-QAM at 4.57 Gb/s, (a)-(b), and at 5.71 Gb/s, (c)-(d)

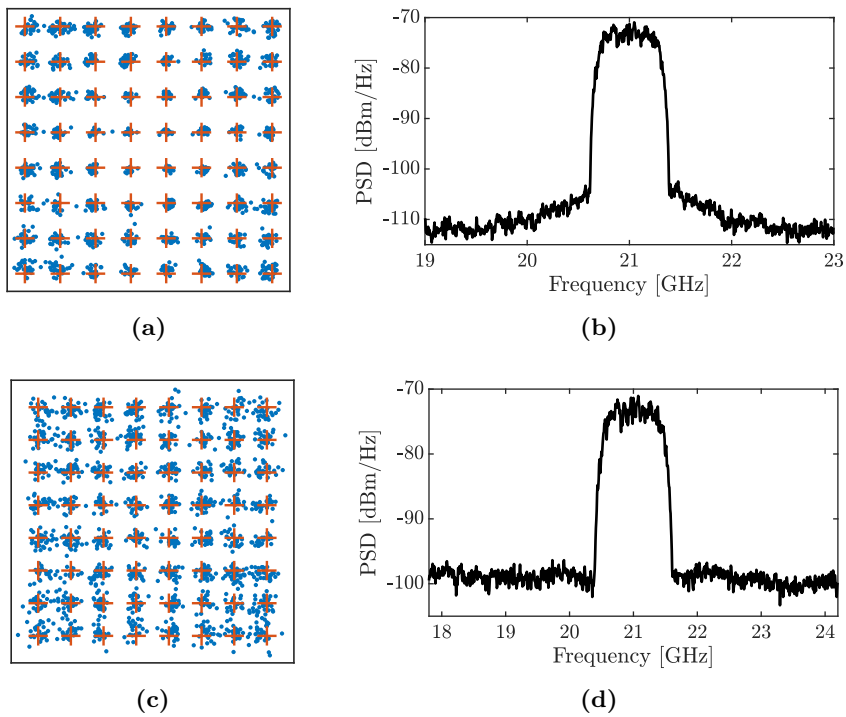


Figure 3.13: Constellation diagram and spectrum for 64-QAM at 3.43 Gb/s, (a)-(b), and at 5.49 Gb/s, (c)-(d)

Table 3.1: Performance comparison with state-of-the-art.

	This work	[9]	[10]	[31]
Technology	22 nm SOI	45 nm SOI	28 nm	45 nm SOI
Resolution [b]	2×6	2×6	2×10	2×4
Sample Rate [GS/s]	8	6.5	2	–
Freq., (3 dB BW) [GHz]	17–24* (17–25)*	15–32 (18–32)	20–32 (–)	42–47 (38–49)
Peak Psat [dBm]	10.4	19.9	19.0	21.3
Drain eff. [%]	15.6	15.6	34.4	16**
Modulation format, Data rate [Gb/s], EVM [%rms]	16QAM, 4.6, 12.6; 32QAM, 4.6, 8.9; 64QAM, 3.4, 5.7	16QAM, 26, 7.9; 32QAM, 30, 6.9; 64QAM, 12, 4	64QAM, 3, 3.6; 256QAM, 1, 2.8	QPSK, 1.25, 4; 16QAM, 0.04, 8
Area (core) [mm^2]	1.18 (0.15)	2.42 (–)	1.6 (0.2)	1.15 (–)

* Working range for modulated signals, CW range is larger. ** PAE.

figures 3.11c, 3.12c and 3.13c. The groups are here closer to be equally sized at all the constellation points, not as clear in figure 3.13c, indicating that the EVM is limited by the internal noise in the circuit, rather than the non-linearity as in the cases before.

Chapter 4

Background on analog-to-digital conversion

This chapter will provide some theoretical background knowledge needed for understanding of the design and evaluation of the converter presented in chapter 5. This includes both fundamental concepts, topological options, and evaluation metrics.

4.1 Conversion theory

The principle of analog-to-digital (A/D) conversion is based on the concept of taking a continuous signal in both time and amplitude and represent it using discrete points in time and a limited set of amplitude levels. These concepts are known as sampling and quantization and will be introduced one by one before being combined.

4.1.1 Sampling

Sampling is the process of converting a continuous-time signal into a discrete-time representation of the signal. Mathematically, this process is often represented by a multiplication of a signal with a sequence of uniformly distributed Dirac pulses [62, Ch. 2]. This case is often referred to as uniform sampling and the sampled signal x_s is then given by equation (4.1), where $T_s = 1/f_s$ is the inter-sample spacing [20, Ch. 2]. Although uniform sampling is most commonly adopted, non-uniform sampling can be useful in some cases. Non-uniform sampling however relies on a significantly more complex mathematical framework [62, Ch. 2].

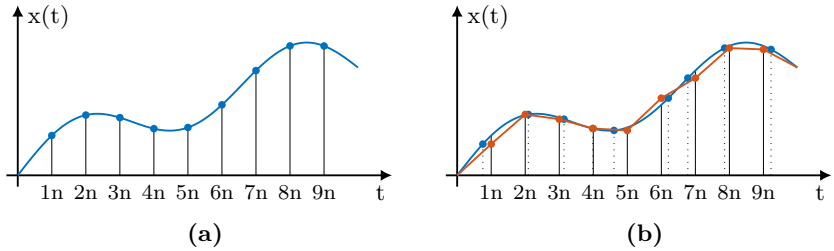


Figure 4.1: Illustration a signal sampled (a) and the effect of jitter in the sampling instance (b). Here the signal is sampled as the blue points while the orange points shows the resulting signal, assuming no jitter.

$$x_s(t) = \sum_{n=-\infty}^{\infty} x(nT_s)\,\delta(t - nT_s) \tag{4.1}$$

A uniformly sampled signal is shown in figure 4.1a. As all samples are assumed to be equally distributed when uniform sampling is used, any variations in the sampling instance are undesired. In a real circuit, it is however impossible to achieve fully uniform sampling [63, Ch. 1]. Small variations in the sampling instance are caused both by systematic errors, known as skew and by random variations, known as jitter [62, Ch. 2]. Differences in path propagation time is an example of a static error while noise coupled through buffer chains is an example for a random variation. With the assumption of equally distributed samples, any variation in the sampling instance will introduce an error in the sampled signal. Although jitter is small relative to the sampling time [62, Ch. 2], the illustration in figure 4.1b shows the effects of an exaggerated jitter.

The non-linear nature of sampling is shown when transforming the sequence of pulses into the frequency domain [63, Ch. 1]. Images of the signal will appear along the frequency axis at all multiples of the sampling frequency f_s. To avoid overlapping images, a band limited signal is needed [62, Ch. 2]. Equation (4.2) presents the relationship between the sample rate f_s and the signal bandwidth BW for no overlap. This relationship is known as the Nyquist criterion was first indicated by Harry Nyquist [64], before Claude Shannon formulated into the expression, equation (4.2), although he formulated it in words [14].

$$BW \leq \frac{f_s}{2} \tag{4.2}$$

In order to be able to reconstruct a signal based on its sampled representation, the Nyquist criterion states that the signal must be band-limited. This also goes for noise, as any out of band noise will be folded into the into the range 0–$f_s/2$, also known as the Nyquist band. To fully band-limit the signal, an anti-aliasing filter is used to, remove undesired signals outside the band of interest.

Although the Nyquist criterion states the ratio between the signal bandwidth and the sample rate, it is not limited to the frequency range 0–$f_s/2$. Higher order Nyquist bands can also be used by extending equation (4.2) to equation (4.3), where the entire band must be located within the same Nyquist band [63, Ch. 1]. Thus, reducing the freedom in choosing sample rate, while avoiding undesired folding.

$$f_H - f_L \leq \frac{f_s}{2} \tag{4.3}$$

In practical implementations of sampling circuits, the sample and hold representation is a more accurate representation [20, Ch. 2]. In a ADC, the signal must be stored while the signal is quantized, making this representation more accurate.

4.1.2 Quantization

Quantization is in principle very similar to sampling, but rather than representing a continuous-time signal at discrete points in time, a continuous-amplitude signal is represented by discrete levels. There is however one significant difference, a sample represents a unique instance in time while a quantization level represents a range of continuous-amplitude values.

The quantization level is selected by rounding the continuous signal to the closest level, thereby introducing an error, know as the quantization error [62, Ch. 4]. The number of quantization levels in an ADC is often defined by its resolution N in bits, giving the quantization step size Δ in equation (4.4), where X_{FS} is the full scale range of the quantizer [63, Ch. 1]. A signal, its quantized representation and the resulting quantization error are shown for two different quantization levels in figure 4.2. In figure 4.2a, the output codes and boarders for the quantization bins are also shown.

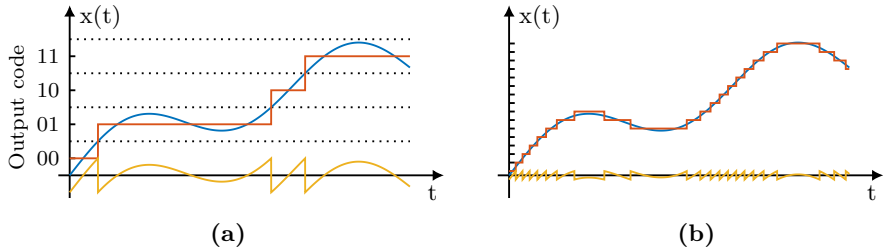


Figure 4.2: Signal, quantization representation and quantization error for a 2 bit resolution (a), and a 4 bit resolution (b). In (a), the output codes and boarder are also shown.

$$\Delta = \frac{X_{FS}}{2^N} \tag{4.4}$$

A mid-point mapping (used in figure 4.2) is often used to represent the location of the quantization level with the range of values it covers, giving a symmetrical quantization error [63, Ch. 1]. However, both the endpoints can also be used to represent the quantization level, while keeping the dynamic range for the quantization error. In an ADC the quantization level is represented by a digital code, while analog quantization levels often exist within the converter itself.

4.1.3 Combining sampling and quantization

Above, sampling and quantization have been treated individually with a continuous behaviour along the axis not discretized. For a sampled signal, its magnitude is still continuous, while the quantized signal is still continuous in time. However, an ADC performs both these operations within each conversion cycle, outputting a signal that is both sampled and quantized. An illustration of a signal simultaneously sampled and quantized is shown in figure 4.3, where uniform stepping along both axes can be observed. In the figure, a ZOH is assumed for representation of the signal between the samples.

An effect of combining both sampling and quantization is that we now not only have a sampled signal but also a unique quantization error for each sample. As can be seen in this figure, the size of the error varies at each sample, giving a noise like behaviour.

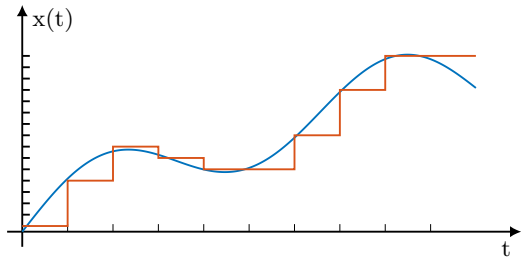


Figure 4.3: Sampling and quantization of a continuous time signal.

4.2 Theoretical and practical limits

There is a set of both fundamental and practical limits defining the limits for the achievable performance for an ADC. While the fundamental limits brought by the theoretical limits defines the boundaries for a given configuration, the practical limits are more connected to effects of variations and noise that in theory are considered ideal. These variations further limit the performance, but the effects can be minimized at design time, however, typically at a cost. It is still important to consider these practical limits already at an early stage as they will affect the converter performance.

4.2.1 Quantization noise

The rounding operation taking place in the quantization process inevitably leads to the introduction of an error. This fundamental error can only become zero when an infinite number of quantization levels are used, which is not realizable in practice [63, Ch. 1]. For a mid point representation, the maximum error is $\Delta/2$, but its average value is highly dependent on the signal [63, Ch. 1]. Under specific circumstances, it is possible to treat the quantization error as noise. First, the probability of reaching any quantization level should be equal, and the quantization errors should be uncorrelated with the input signal [63, Ch. 1]. This forces the signal to be changing over time, and the amplitude must be large enough to occupy a large portion of the dynamic range. Second, a large set of quantization levels should also be used and these must have a uniform spacing.

The average quantization error is calculated based on the assumption of uniform distribution of the errors. With a uniformly distributed

quantization error $p(\epsilon_Q) = 1/\Delta$ within the range $[(-\Delta)/2, \Delta/2]$, the average quantization noise is given by equation (4.5).

$$P_Q = \int_{-\infty}^{\infty} \epsilon_Q^2 p(\epsilon_Q)\, \mathrm{d}\epsilon_Q = \int_{(-\Delta)/2}^{\Delta/2} \frac{\epsilon_Q^2}{\Delta} \mathrm{d}\epsilon_Q = \frac{\Delta^2}{12} \tag{4.5}$$

With the average signal power for a full-scale sine wave given by equation (4.6), the theoretical SNR limit can be defined. Equation (4.7) gives the SNR in dB as a function of the resolution. This expression is also useful when evaluating the performance of converters as will be discovered later.

$$P_{sig} = \frac{1}{T} \int_0^T \left(\frac{X_{FS}}{2} \sin(\omega t) \right)^2 \mathrm{d}t = \frac{X_{FS}^2}{8} = \frac{\left(2^N \Delta\right)^2}{8} \tag{4.6}$$

$$\mathrm{SNR}_Q = 10 \log_{10} \left(\frac{2^{2N} \Delta^2}{8} \frac{12}{\Delta^2} \right) = 6.02N + 1.76 \quad [\mathrm{dB}] \tag{4.7}$$

4.2.2 Thermal noise

With the quantization noise bringing the fundamental noise limit, it is important that other noise sources are minimized in order to keep the overall noise level low. Thermal noise is another fundamental noise source limiting the performance of ADCs [63, Ch. 1]. The sample circuit consists of a sampling switch and a sampling capacitor. Ideally, the sample switch has no on-resistance, however, in practice a non-zero on-resistance unavoidable. This equivalent resistor introduces noise due to the random movement of electrons within it [11, Ch. 10]. The movement is directly connected to the temperature of the device. This noise is white in nature, spanning the entire frequency spectrum, with a noise power equation (4.8) related to the temperature and resistance [62, Ch. 2].

$$\overline{v_{n,R_s}^2} = \overline{4kTR_s} \tag{4.8}$$

The combination of the sample capacitor and the non-ideal sample switch brings a low-pass response to the sampling circuit, thereby limiting the total sampled noise power [62, Ch. 2]. Equation (4.9) gives the transfer function for the band-limited thermal noise power. The total thermal noise power sampled is then given by equation (4.10).

$$v_{n,C_s}^2(\omega) = \frac{\overline{4kTR_s}}{1+(\omega R_s C_s)^2} \tag{4.9}$$

$$P_{n,C_s} = \int_0^\infty \frac{\overline{4kTR_s}}{1+(2\pi f R_s C_s)^2}\mathrm{d}f = \frac{kT}{C_s} \quad [\mathrm{W}] \tag{4.10}$$

In order to specify the size of the sampling capacitor we need to relate thermal noise to the quantization noise. By setting them equal, the size of the sampling capacitor C_s is given by equation (4.11). A larger sampling capacitor will however in most cases be beneficial as the SNR is degraded by 3 dB when both these noise sources are equal.

$$C_s = \frac{kT}{v_{n,C_s}^2} = \frac{12\cdot 2^{2N}kT}{V_{FS}^2} \quad [\mathrm{F}] \tag{4.11}$$

For differential sample circuits, the sampled noise power will increase by a factor of two, given that each channel is equal to the corresponding single-ended channel [65, Ch. 6]. At the same time, the sampled signal power increases by a factor of four compared to the single-ended case. This, given constant magnitude on the channels both in the differential and single-ended case. This would bring an SNR improvement of 3 dB, allowing for the sample capacitors in each differential channel to be reduced by 50 %. This results in the same requirement for the minimum sampling capacitor both in the single-ended case and in the differential case.

4.2.3 Jitter

Variations in the actual sampling instance, jitter, introduces additional errors as uniform sampling is typically assumed in the DSP. Jitter will cause the signal to be sampled at inaccurate points in time, thereby bringing an incorrect signal level to the assumed sampling point [62, Ch. 2]. The magnitude of the error is dependent both on the inaccuracy in the sampling instance $\delta(t)$ and on the slope of the signal at this point [63, Ch. 1]. For a single sine wave signal, the noise power caused by jitter is given by equation (4.12) which together with the signal power gives the SNR in equation (4.13). Expressions for more complex signals can be calculated through their derivatives.

$$x_j^2(t) = A^2\omega_{in}^2\cos^2(\omega_{in}t)\,\delta(t) \tag{4.12}$$

$$\mathrm{SNR}_{ji} = 20\log_{10}\left(\omega_{in}\delta_{rms}\right) \quad [\mathrm{dB}] \tag{4.13}$$

Just like thermal noise, jitter can also be related to the quantization noise, assuming a full-scale input signal. Equation (4.14) shows the average tolerable jitter relative signal frequency and quantizer resolution for a resulting noise contribution equal to the quantization noise. For high-speed, high-resolution converters jitter can become a significant problem that is hard to overcome. In normal digital circuits, a jitter in the orders of tens of $\mathrm{ps_{rms}}$ is often seen [62, Ch. 2]. Reaching below $100\,\mathrm{fs_{rms}}$ is however significantly more complex with only a few cases being demonstrated below this level [66], assuming jitter to be the limiting factor in these cases.

$$\delta_{rms} = \sqrt{\frac{2}{3}}\frac{1}{2^N\omega_{in}} \quad [\mathrm{s}] \tag{4.14}$$

4.2.4 Matching

No device is completely accurate. All devices, capacitors, resistors, and transistors implemented on chip experience variations [62, Ch. 5]. In some parts of the converter, the relative accuracy between components, known as matching, is important for proper operation. In an ADC, good matching is most important in the analog building-blocks such as the capacitive digital-to-analog converter (CDAC) and comparators. However, in some cases, such as for comparators, the effects of mismatch can be compensated for by adopting calibration [67]. Other mismatch effects, such as the linearity of the reference voltage scaling, can to some extent be compensated for in the DSP following the ADC.

The implementation of the decision levels is highly dependent of the converter topology, resulting in different requirements for accurate matching. For SAR converters, the sampling capacitor often doubles as the reference generator, making matching of the capacitors making up the CDAC critical. For low-resolution converters, the minimum capacitor size needed for good matching will result in a significantly larger total capacitance than needed for the thermal noise suppression. For high resolution converters, the required total capacitance needed to account for the thermal noise might result in additional mismatch effects caused by the distribution of the capacitors.

4.3 Figures of merit

Evaluating the performance of an ADC includes measurements of both dynamic and static properties. The sampled energy gives a theoretical connection between the converter resolution and energy consumption per conversion, although the ratio between sampled energy and total energy consumed is highly dependent on design. Based on demonstrated implementations, it can be observed that the power consumption is dependent on both bandwidth and resolution [62, Ch. 4]. Comparison metrics based on the bandwidth, resolution and power can be used to compare different implementations [68].

The somewhat standardized input and output interfaces in ADCs, allows for the creation of fair comparison metrics. This in contrast to RF-DACs where no good comparison metrics exist, as mentioned in section 2.2.

4.3.1 Dynamic metrics

The dynamic behaviour of a converter is often defined by two metrics: the SNDR and the spurious free dynamic range (SFDR) [65, Ch. 1]. In some cases, effective number of bits (ENOB) is used as a substitute for SNDR [63, Ch. 2].

The SNDR is calculated as the ratio between signal and noise power including distortion, as in equation (4.15), for a single sine wave signal [65, Ch. 1]. SFDR on the other hand is the ratio between the signal and the largest spurious tone [63, Ch. 2]. Both these metrics involve performing a Fast Fourier transform (FFT) on the converter output data, transforming it into the frequency domain. These metrics are most commonly presented in dB since the ratio between signal and noise can be several orders of magnitude.

$$\text{SNDR} = 10\log_{10}\left(\frac{P_{signal}}{P_{noise+distortion}}\right) \quad \text{[dB]} \tag{4.15}$$

With the quantization noise setting the fundamental limit for the SNDR, a good indication of the converter performance is its effective resolution [62, Ch. 4]. Reversing equation (4.7) gives the effective resolution, equation (4.16), based on the measured SNDR. The effective resolution makes it possible to quickly assess the performance of a converter by comparing its resolution to the measured ENOB.

$$\text{ENOB} = \frac{\text{SNDR} - 1.76}{6.02} \quad \text{[b]} \tag{4.16}$$

SFDR gives the usable dynamic range (DR) without risking the signal being masked by a spurious tone [63, Ch. 2]. In many cases, SFDR is limited by the third order harmonic, then equalling the harmonic distortion (HD). However, any spurious tone can limit the SFDR, while only harmonics are considered in HD.

4.3.2 Static metrics

Ideally, an ADC will only show errors related to quantizing the signal. However, in practice additional errors depend on the circuit topology, the process used for implementation, the resolution of the converter, and on the chosen sample rate, will further reduce the performance [62, Ch. 4]. All these effects will affect the linearity of the converter, causing unequally sized quantization steps.

The linearity of the converter is mostly considered to be a static measure [63, Ch. 2; 65, Ch. 1], although at circuit level, it might be affected by dynamic properties in the circuit. Linearity is a measure of how non-linear effects caused by non-uniform quantization affects the ideal transfer curve. Two different metrics, the differential non-linearity (DNL) and the integral non-linearity (INL) are commonly used together when evaluating the linearity [62, Ch. 4]. Although both numbers have a value for each output code, the maximum value is often reported as a single value. This can however hide important information only given by the entire data-set. Both INL and DNL are typically presented in LSBs, giving a tight connection to the quantization step size, although the absolute value in volt, or a fraction of the DR in % also can be used [63, Ch. 2].

DNL is the step size deviation from the ideal step size Δ. Ideally, the DNL should be 0 LSB, but more importantly, it should be larger than -1 LSB as a DNL below indicates missing codes, thus resulting in digital output codes not reachable by any analog input signal [62, Ch. 4]. The DNL is given by equation (4.17), where $\Delta_r(k)$ is the actual step size given by $X_{k+1} - X_k$ [63, Ch. 2].

$$DNL(k) = \frac{\Delta_r(k) - \Delta}{\Delta} \tag{4.17}$$

INL is the total deviation from the ideal linear transfer function. It is tightly connected to the DNL through equation (4.18), giving the true

interpolated INL. This definition of INL can however in some cases be a bit pessimistic as in also includes potential linear gain errors. Two alternative definitions that account for this is the best linear fit and the end-point fit [63, Ch. 2]. The end-point fit is the most commonly used, giving an INL of 0 LSB at the two end-points by compensating for offset and gain error [62, Ch. 4; 63, Ch. 2]. Assuming gain and offset compensation is typically not a problem as these are linear effects that easily can be compensated for using DSP after the converter.

$$INL\,(k) = \sum_{i=1}^{k} DNL(i) \tag{4.18}$$

$$INL\,(k) = (1+G)\sum_{i=1}^{k} DNL(i) \tag{4.19}$$

Although both DNL and INL are metrics of linearity they have different effects on the performance. A large uncorrelated DNL will behave like white noise adding on top of the quantization noise [63, Ch. 2]. A large INL on the other hand indicates a deviation from the linear transfer function, thus causing HD [63, Ch. 2].

4.3.3 Comparison metrics

Comparing different ADCs based on the static and dynamic metrics presented above in combination with the power consumption will not result in a fair comparison as the converter complexity is not considered. The energy consumption per conversion is a good metric to use for estimations of block-level power consumption in a system. However, it does not handle the increased complexity coming with increased resolution, thereby making it an unsuitable option for comparing converters with different resolution.

Although there is a large set of comparison metrics, two are commonly used for comparing different ADC implementations while accounting for the complexity related to the resolution [62, Ch. 4; 63, Ch. 2]. These two metrics are Walden figure of merit (FoMW) and Schreier figure of merit (FoMS). Both these metrics are a result of empirically fitting ADC data over two different performance regions [69].

FoMW normalizes the consumed energy by the effective number of quantization levels, equation (4.20) [62, Ch. 4; 63, Ch. 2; 70]. This definition is however the inverse of the originally defined one [68, 71].

This metric is most suitable for comparisons of converters with moderate resolution as it does not fully account for the increased complexity in noise limited designs [66, 69].

$$\text{FoMW} = \frac{P_{DC}}{2^{ENOB} f_s} \quad \text{[fJ/conv.-step]} \tag{4.20}$$

The other popular comparison metric is the FoMS, equation (4.21) [62, Ch. 4; 63, Ch. 2]. However, originally this metric used the DR rather than the SNDR thereby not accounting for the distortion [72, Ch. 9; 73, Ch. 1]. This metric better suits high resolution converters as the energy per conversion is not scaling sufficiently for lower resolutions [66].

$$\text{FoMS} = \text{SNDR} + 10\log_{10}\left(\frac{f_s}{2P_{DC}}\right) \quad \text{[dB]} \tag{4.21}$$

Thanks to the long history of designing ADCs, a large volume of designs have been published. Several surveys have studied trends in ADC performance, using these comparison metrics as part of the comparison [70]. One such survey is performed by Boris Murmann who collects the latest presented ADCs from ISSCC and VLSI Symposium each year, making comparisons possible [66]. When performing these comparisons it is important to use the FoMs with care as the trade-off between power and resolution is significantly more complex than implied by these metrics [69].

4.4 High-speed converter topologies

High-speed ADCs can be implemented using several different converter topologies. Among them, we find the flash converter, the pipeline topology and the SAR topology [62, Ch. 8]. In addition to these, multiple ADCs can be time interleaved in order to further increase the sample rate [63, Ch. 4]. In recent time, the SAR topology has been the most popular topology for high-speed moderate resolution ADCs [66]. Although these three generic topologies often are used for categorization, implementations presented in literature often combine features from different topologies [66], thus making the categories a bit more diffuse.

All the three topologies presented below implements single channel converters. Time interleave (TI) multiple converters makes it possible to increase the sample rate beyond what a single converter can achieve [62, Ch. 9]. This is done by sampling each converter with a phase-shifted

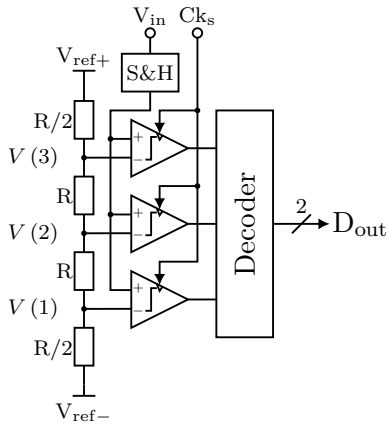


Figure 4.4: Illustration of a 2 bit flash converter.

clock signal, thereby making the converters sample at different points in time. It is important to remember that the Nyquist criterion is only fulfilled when all channels are combined and not for each channel by itself [62, Ch. 9].

Single channel converters have been demonstrated to achieve sample rates of a few GS/s [74–76]. TI converters on the other hand has been demonstrated to reach 100 GS/s [77].

4.4.1 Flash

The flash converter is the fastest single-channel converter topology, performing an entire conversion within the duration of a single comparison [62, Ch. 8] by simultaneously performing all comparisons in parallel [63, Ch. 4]. This is great for high-speed operation as flash converters achieves sample rates that otherwise would have required TI converters [66]. However, it comes at the cost of a high power consumption and a large footprint as the number of comparators scale with $2^N - 1$, only making this topology suitable for low resolution converters [63, Ch. 4]. As all levels are compared simultaneously, the result will be thermometer coded, thus a thermometer to binary conversion is needed in order to retrieve a binary result [62, Ch. 8].

A 2 bit flash converter is illustrated in figure 4.4, showing the parallel operation of all the comparators. The reference level generation is implemented using a resistive divider, providing one reference level for each comparator. The decoder is used to convert the thermometer data

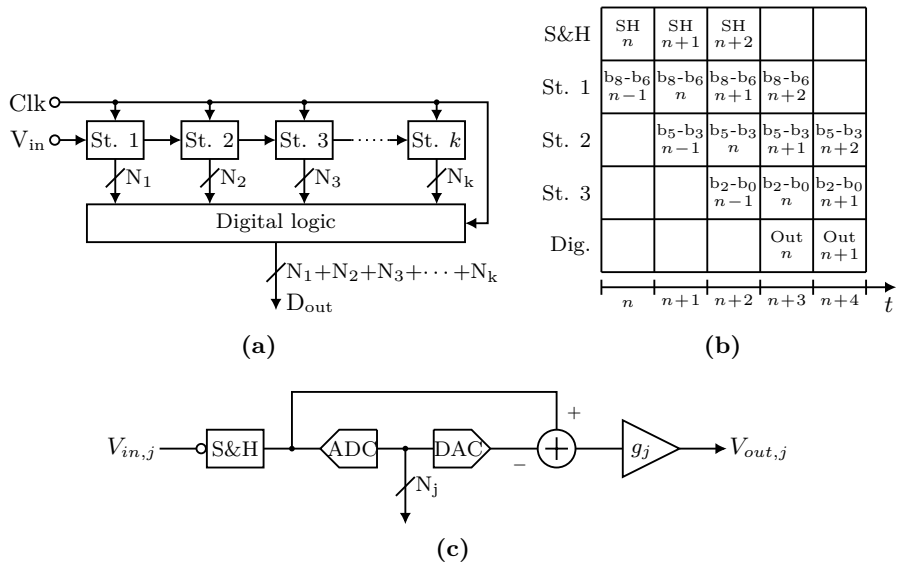


Figure 4.5: Illustration of the pipeline converter topology, (a), the timing at which the different bits are resolved, (b), and a generic pipeline stage (c).

into binary data.

4.4.2 Pipeline

Just like flash converters, pipeline converters build on parallelism in space. The difference is that pipeline converters cascade multiple low-resolution converters, thus distributing the conversion over time [63, Ch. 4]. Pipeline converters mostly resolves a single bit in each stage [62, Ch. 8]. However, in principle a more complex ADC and DAC can be placed in each stage, increasing the number of resolved bits in each stage. The gain accuracy and comparator offset limits the achievable performance [62, Ch. 8]. The re-sampling of the signal in between each stage also introduces latency from when the signal enters the converter to the time when the digital result can be collected.

A generic pipeline converter is illustrated in figure 4.5a, using the generic stage topology shown in figure 4.5c. A timing diagram showing when the different bits are outputted from the different stages is shown in figure 4.5b. The generic stage topology in figure 4.5c shows how the

signal first is sampled in the ADC, resolving the N_j bits. The DAC then subtracts this result from the signal, producing a residue signal that is then amplified before reaching the next stage. The residual signal is less than an LSB in each stage. The amplifier gain is less than or equal to the resolution of the previous converter, thereby filling the range of the next converter.

The amplifiers needed for amplifying the residual signal between each signal is in most cases the limiting component in pipeline converters [62, Ch. 8]. While a gain of 2, can be achieved with high accuracy, higher gain becomes more complex to get accurate due to matching, thus limiting the usable resolution in each stage. Thanks to the stage gain that resets the DR for the next stage, making the pipeline converter suitable for high resolution implementations. Traditionally, the high power consumption of the amplifiers has limited the use of pipeline converters with many stages. However, with the ring-amp topology, highly accuracy and power efficient amplifiers make the pipeline converters a competitive option to SAR converters [78, 79].

4.4.3 Successive-approximation register

In contrast to the pipeline topology, SAR converters build on a iterations in time, successively getting closer to the result for each iteration while only using a single comparator [62, Ch. 8]. A simplified converter is illustrated in figure 4.6a, illustrating the components needed. A timing diagram showing when the different bits are generated is shown in figure 4.6b. The input signal is first sampled and then compared against a reference value. The result of the comparison is then used to update the reference level, reducing the range in which the signal is located. For binary scaled reference levels, the number of iterations equals the resolution in bits.

Thanks to the limited set of analog components and lack of amplifiers, the SAR topology shows great potential for low power consumption [62, Ch. 8]. These properties are further enhanced by the use of charge-redistribution based reference level generation as power consumption in the reference generation is significantly reduced [62, Ch. 8]. Thanks to the switching operation, SAR converters are highly suitable for CMOS integration, being the topology that has benefited most from the transistor scaling [62, Ch. 8].

Most high-speed SAR converters are found in the 6–10 bit range, locating them between flash and pipeline converters [66]. The limiting

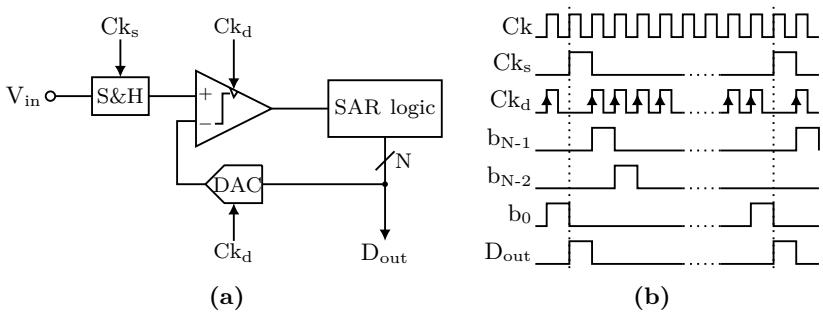


Figure 4.6: Illustration of the SAR converter topology (a) and the timing for when the bits are resolved (b).

factor is a combination of the accuracy of the reference level generation, limited by the capacitive matching, and comparator offset voltage [62, Ch. 8].

4.4.4 Topological trends

With an increased performance for smaller feature size technologies and a changing resolution requirement on high-speed converters, the topology of choice has also been changing [66]. Pipelined converters where for a long time the choice in high-speed applications, but thanks to the increased process performance, the SAR topology became more popular bringing reduced power consumption, however at lower resolution [66]. The pipeline topology might thanks to the development of the ring-Amp again be an option to consider for high-speed moderate converters. They also show potential for significantly higher resolution than SAR converters, thereby making them a very interesting option [78, 79].

4.5 Charge-redistribution

In principle, any DAC can be used to generate the reference levels in a SAR converter. Traditionally, resistive structures have been the dominant choice, suffering from a large power consumption [62, Ch. 8]. Capacitor-based, so-called charge-redistribution networks have become a more popular choice in low-power converters, taking advantage of the good switching properties provided by the CMOS technology [62, Ch. 8].

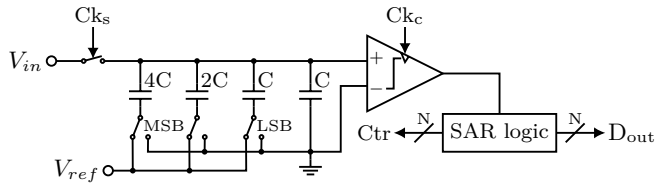


Figure 4.7: Illustration of a 3 bit charge redistribution network connected to a comparator.

Charge-redistribution builds on using the sampling capacitor to generate decision levels. An illustration of a 3 bit charge-redistribution network together with a comparator is shown in figure 4.7. The sampling capacitor is arranged as several parallel capacitors, binary scaled in this case. The back-plates of the capacitors are pre-set to the reference voltage before sampling and the signal is then sampled onto the capacitors top-plate. First the MSB is toggled to ground, changing the voltage on the top-plate according to equation (4.22). The comparator is then used to decide if the switch setting should be kept or not. The process in then repeated until all bits are resolved. When all stages have been switched, the residual voltage on the top-plate should be less than 1 LSB.

$$V_{top} = V_{in} - \frac{4C}{4C + 2C + C + C} V_{ref} \quad [\text{V}] \tag{4.22}$$

Differential charge-redistribution networks brings the advantage of not needing any toggling before resolving the MSB as this level is inherent from the topology [62, Ch. 8].

In a practical implementation, not all the capacitance in the charge-redistribution network will be switchable due to parasitic capacitance in the routing. This reduces the dynamic range of the charge-redistribution network. The reference voltage used in the charge-redistribution can however be used to compensate for this effect [80]. Equation (4.23) gives the reference voltage needed for a desired dynamic range when non-switchable capacitance is present in the charge-redistribution network.

$$V_{ref} = V_{FS} \frac{C_{sw} + C_p}{C_{sw}} \quad [\text{V}] \tag{4.23}$$

4.6 Redundant scaling

The most compact digital representation of an analog signal is when it is binary encoded. Most ADCs presented in literature are also binary scaled as this results in the smallest number of comparisons [66]. Although binary encoding gives the smallest number of comparisons in a SAR converter, it will not provide any redundancy against errors [81]. In principle, it is not a problem to assume correct decisions, however; in high-speed converters it becomes costly to guarantee fully settled decision levels. Redundancy could then be used to reduce the settling time requirement, increasing the speed of the converters [62, Ch. 8].

For a charge-redistribution based DAC, two factors sets the total settling time: the RC time-constant τ of the charge-redistribution cell, and the number of time-constants required for sufficient settling. The RC time-constant is set by the switch resistance R in combination with the capacitance C loading it. Reducing C is typically not possible due to matching and thermal noise concerns. Reducing R through a large switch would on the other hand load the charge-redistribution network with additional parasitic capacitance. The number of time-constants needed is on the other hand dependent on the scaling used [62, Ch. 8]. With binary scaling, reducing the number of time constants τ is not possible as this would introduce errors. Redundant scaling can on the other hand be used to reduce the number of time constants as the redundancy correct for incorrect decisions caused by incomplete settling [82].

In order to simplify the description of the different scaling principles a few definitions will be used. A *stage* corresponds to one iteration in the SAR algorithm. A *level* is one of the outputs from the reference generator against which the signal is compared. With redundant scaling, several control codes might result in the same output level. A *path* is a sequence of decisions and corresponding levels that have been visited during the previous stages. Although several paths might result in the same output, they are all unique. A *step* is the separation between decision levels at two successive stages. In addition to this, it will also be assumed that all control bits have a binary representation, that is, they are either 0 or 1.

Three different scaling principles will be described in more detail below: the binary, the generalized non-binary, and the compensating scaling.

Starting with the binary scaling, it builds on halving the step at each stage. Under error-free operation, the difference between the signal and

its quantized representation will always be smaller than 0.5 LSB [80]. However, an erroneous decision along the way will result in a significantly larger error. An example of a binary scaled 5 bit decision tree is shown in figure 4.8a, showing both correct operation and the effect of an erroneous decision. The reference level is calculated according to equation (4.24) and the output code is provided by equation (4.25) giving the stage results $d\,(i)$.

$$V_{ref}\,(k) = 2^N \left(2^{-1} + \sum_{i=1}^{k} -1^{d(i-1)+1} 2^{-1} \right), \quad (k = 1, 2, ..., N) \tag{4.24}$$

$$D_{out} = d\,(1)\,2^{N-1} + d\,(2)\,2^{N-2} + \cdots + d\,(N) = \sum_{i=1}^{N} d\,(i)\,2^{N-i} \tag{4.25}$$

The simplest redundant scaling algorithm is non-binary scaling which uses a fixed radix $\gamma = 2^{N/M}$ with $N \leq M$, where N is the resolution and M the number of comparisons [81]. The resulting radix, in the range $1 < \gamma \leq 2$, will most likely not result in integer sized steps. A lower value on γ will result in increased redundancy at the cost of a larger number of comparisons.

The generalized non-binary scaling algorithm is a generalization of the principles behind the non-binary scaling algorithm, allowing more freedom in the placement of the redundancy. The generalization extends the definition of the radix, allowing individual radix in different stages, thus providing integer sized steps and more freedom in the redundancy placement.

Two propositions define the requirements for the generalized non-binary scaling. If the first statement, equation (4.26), is fulfilled, a correct result will be obtained even if an erroneous decision have been made at stage k [81]. The second statement, equation (4.27), defines the total number of redundant paths, given the redundancy $q\,(i)$ at stage i and the over-range r.

$$|V_{in} - V_{ref}\,(k)| < q\,(k) \tag{4.26}$$

$$2^M - 2^N = \sum_{i=1}^{M-1} 2^i q\,(i) + 2r \tag{4.27}$$

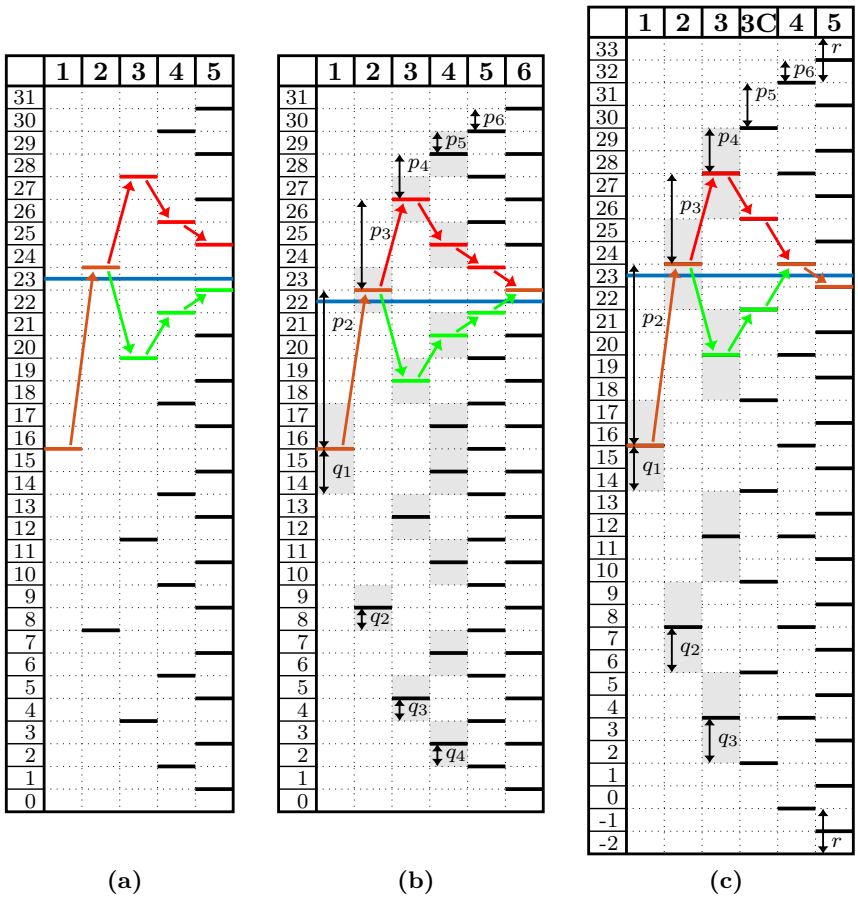


Figure 4.8: Illustration of binary scaled reference levels (a), scaling according to the generalized redundant scaling algorithm (b), and the compensating scaling (c). The generalized scaling uses $p = [16, 7, 4, 2, 1, 1]$ and $q = [2, 1, 1, 1, 0, 0]$. For all three scaling principles, the resulting path is shown for both fully correct decisions and in the case of an incorrect decision at stage 2. The input level is represented by the blue line, gray filled boxes shows the redundancy, orange lines indicates shared decision levels. The green and red paths represent the paths taken by correct decision and in the case of an erroneous decision at stage 2.

With an increased number of decision levels for each successive stage, thus increasing the cost of adding redundancy, making redundancy more costly at later stages in the conversion cycle. Two conditions must be fulfilled when calculating the steps [81], giving the output range $[-r, 2^N - 1 + r]$. The first, equation (4.28), gives the size of the first stage, being located in the middle of the range. The second, equation (4.29), gives the relationship for the total number of decision levels based on the size of each step. The step $p\,(k+1)$ is then given by equation (4.30) for a chosen redundancy $q\,(k)$ at stage k and the redundancy at the following stages.

$$p\,(1) = 2^{N-1} \tag{4.28}$$

$$\sum_{i=1}^{M} p\,(i) = 2^N - 1 + 2r \tag{4.29}$$

$$p\,(k+1) = 2^{M-k-1} - q\,(k) - \sum_{i=k+1}^{M-1} 2^{i-k-1} q\,(i) \tag{4.30}$$

The decision levels at stage k is then calculated using equation (4.31) and the digital output code is provided by equation (4.32) [81]. It should be noted that while binary scaling can be referenced both from 0 and the mid-point, the mid-point is the only point at which a redundant scaling is symmetric. The last part of equation (4.32) is needed to align the output code and reference level while keeping the binary range starting at 0. Excluding this part would result in the output code pointing at the next level above which becomes problematic for the top-most data value.

$$V_{ref}\,(k) = p\,(1) + \sum_{i=2}^{k} -1^{d(i-1)+1} p\,(i)\,, \quad (k = 1, 2, ..., M) \tag{4.31}$$

$$D_{out} = p\,(1) + \sum_{i=2}^{M} -1^{d(i-1)+1} p\,(i) + 0.5 \cdot -1^{d(M)+1} - 0.5 \tag{4.32}$$

An example of a 5 bit generalized scaling using $M = 6$, $p = [16, 7, 4, 2, 1, 1]$, $q = [2, 1, 1, 1, 0, 0]$, and no over-range is shown in figure 4.8b. Here, both the path for fully correct decisions and for the path when an

erroneous decision is made are shown. As can be seen, the redundancy covers the erroneous decision, although an alternative path has been taken to the result.

The final redundant scaling principle presented in the compensating principle [80]. It builds on using binary scaled stages and then introduces compensation stages with the same step as the stage before its placement, thereby bringing redundancy to all the earlier stages [80]. This principle is a special case covered by the generic scaling algorithm. The level shift introduced by the compensation stage corresponds to half a step, thereby introducing over-range. As a result, the larger redundancy used, the more over-range becomes introduced, thus reducing the usable DR of the converter [80]. An example of a 5 bit compensating scaling is presented in figure 4.8c, introducing a compensation stage after the third stage.

Where in the conversion cycle redundancy becomes most useful depends on what type of event it should cover. In the case of random decision errors in the comparator, redundancy should be located in the last stage as an error in any previous stage can be covered for as long as the error is small. In the case when redundancy is used for settling-time reductions by allowing decisions at incomplete settling of the reference level, it becomes more useful as a large settling-time reduction is achieved. The stage-dependent settling-time is caused by the requirement of a fixed allowable settling error, typically 0.5 LSB [81, 82], while the range is reduced in each stage. With redundancy, the tolerable settling error is increased, thus further reducing the settling time. However, in order to achieve reduced conversion time with redundant scaling, the cost of adding additional decisions must be gained by the reduced settling-time provided by the redundant scaling.

Chapter 5

ADC design and evaluation

Parts of the work with the SAR ADC presented below has already been presented in [83]. This includes the schematic level design and all the related topological design choices. Some of these choices and their background will be repeated below when the ADC design is presented. This is in order to bring a base for the evaluation of the converter which is the main contribution presented in this work.

SAR converters have for some time been seen as the most suitable option for implementing high-speed converters at a resolution of 6–10 bits [66]. They are very energy efficient and highly suitable for integration in modern small feature-size CMOS processes, thanks to the small number of analog components [62, Ch. 8]. As SAR converters iteratively converts the analog signal into a digital one, the iteration time limits the achievable sample rate. In order to reach the very highest sample rates for a single-channel converter, a reduction of the loop-time is needed. Commonly, this reduction is achieved by employing multiple comparators rather than a single one which the generic SAR topology uses [82,84,85]. The comparators can be arranged so that the number of iterations are reduced by resolving multiple bits in each cycle [85–87], or by reducing the time for each iteration by using alternating comparators, thereby eliminating the reset time from the loop-time [82, 84, 88–90].

Multi-bit per cycle converters often demonstrate a higher SNDR, however, at slightly lower sample rate than achieved using alternating comparators [82, 84, 85, 87]. While alternating comparators often target high sample rate [82, 88–90], the combination of high sample rate and high SNDR can still be achieved [84].

The alternating comparator topologies presented in [82, 84] are both

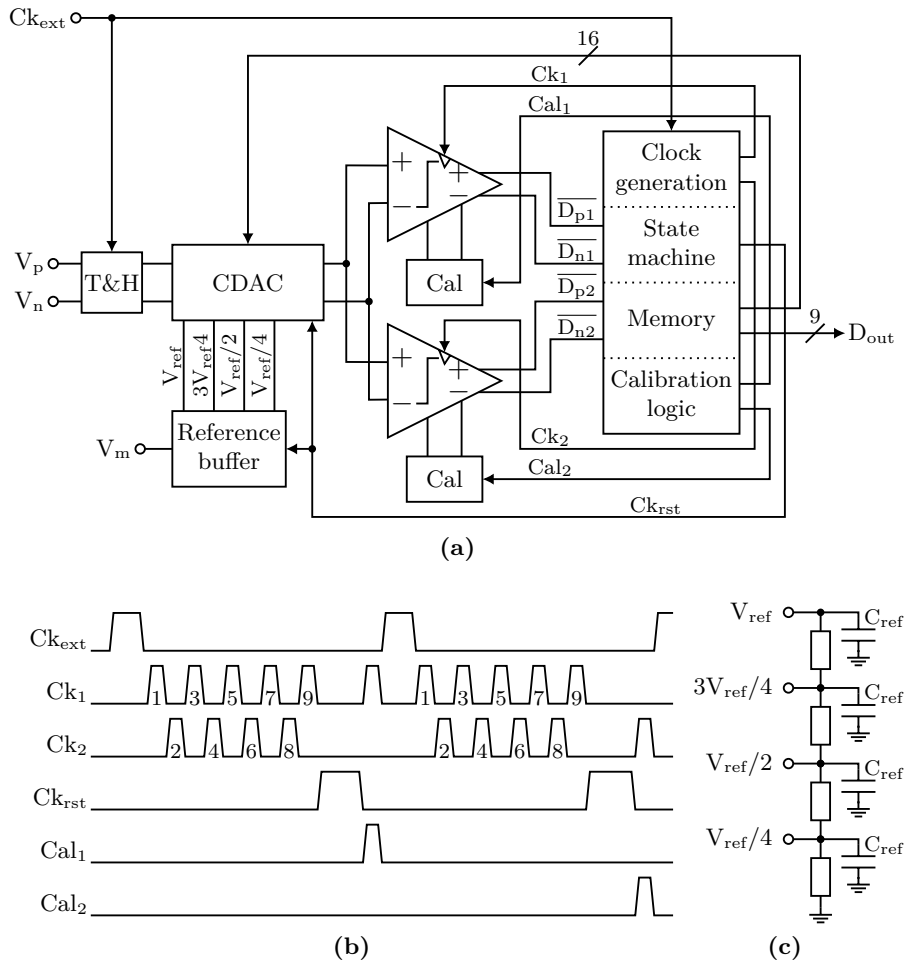


Figure 5.1: Illustration of the ADC topology, in (a), featuring alternating comparators, CDAC, and asynchronous logic. The timing for the alternating comparators is shown in (b). A schematic for the fractional reference voltage generation is shown in (c).

very similar, relying on redundantly scaled reference generators. The main difference is that [82] relies on a pair of comparators while [84] uses both coarse and fine comparator pairs. The SAR topology presented here, largely matches the topology presented in [82]. The cost of additional comparators as used in [84] was considered to be too high. The converter is shown in figure 5.1, featuring redundantly scaled reference levels, offset voltage calibration, asynchronous control logic, and an internal reference voltage buffer. The internal reference voltage is controlled by an external reference, allowing free tuning. A high-bandwidth track and hold circuit is used for sampling the analog signal onto the CDAC.

A few of the core building blocks will be presented in more detail below followed by measurements and performance evaluation. The focus will be on connecting the measured performance to the performance indicated by simulations, in an attempt to understand the factors limiting the performance.

5.1 Alternating comparator topology

The benefit of using alternating comparators comes from eliminating the comparator reset time from the critical timing path [82]. This, as only one comparator is needed for the comparison, making it possible for the other one to reset itself. The timing for the alternating comparators is shown in figure 5.1b.

Using multiple comparators, their relative matching to each other becomes important. Comparator mismatch appears as an input referred offset voltage, thus moving the tipping point for the comparator. When a single comparator is used, offset voltage is not a large problem as it is constant for all comparisons, making it possible to compensate for using DSP. In alternating comparators, offset voltage is more troublesome as the comparator in use changes for every comparison, thereby making it more complex to identify and compensate for using DSP. Rather, the comparators needs to be calibrated in order for these effects to be reduced.

Although two comparators is the minimum number needed for alternating operation, nothing precludes the used of additional comparators. Both coarse and fine alternating pairs are used in [84]. However, one could also see the potential of alternating among more than two identical comparators in order to give more time for reset. The cost, only

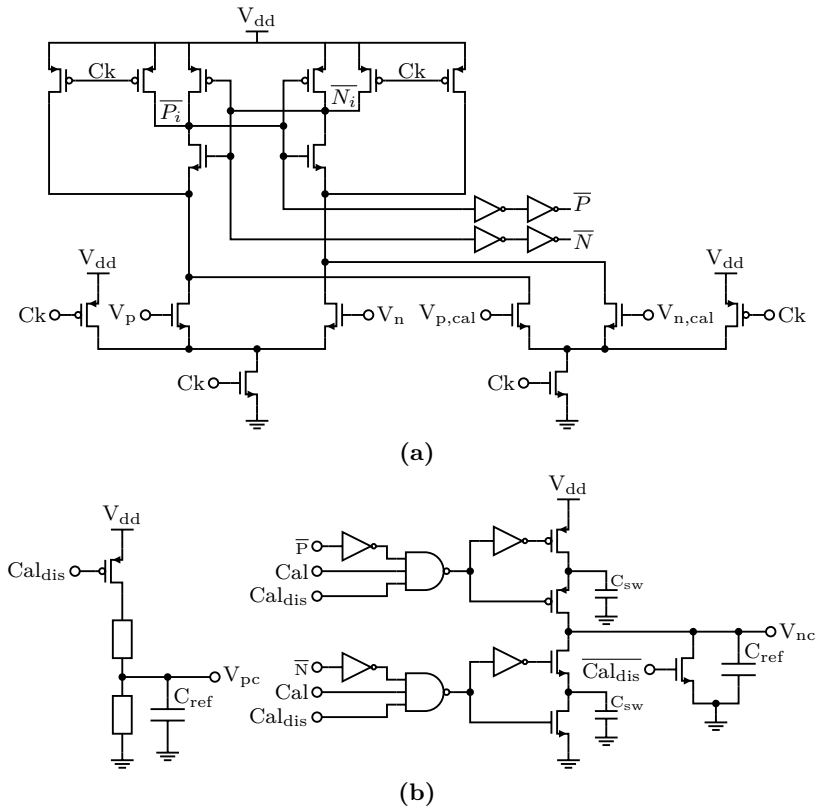


Figure 5.2: Schematic of the strong-ARM comparator used (a) and the calibration circuitry used for reducing effects of comparator offset voltage (b).

looking at the comparators is not that large [84]. The cost rather lies in additional load on the CDAC and more significantly in the footprint of the calibration circuit needed for each converter, often significantly larger than the comparator itself [84].

With its core functionality, the comparator topology greatly affects the achievable performance. It must provide fast decision time and small offset voltage, and not rely on internal bias currents as these increase the power consumption. Both the dynamic latch [84, 91–93] and the strong-ARM comparator [82, 94–96] are popular topologies in SAR converters, fulfilling the above requirements. The dynamic latch is based on a distributed topology that gives a higher internal gain, thereby increasing the decision time for small input differences. The strong-ARM topology on the other hand uses a single stack of transistors, thereby increasing the decision speed for large input signals. As the voltage difference compared for most comparisons is fairly large in a SAR ADC, the increased decision speed becomes beneficial. A schematic of the comparator together with the calibration logic is shown in figure 5.2a.

The device mismatch and layout asymmetries are both sources to offset voltage in the comparator. As the latching part of a comparator often relies on cross-coupled inverters [82, 84]. When cross-coupled inverters are used, a fully symmetric comparator cannot be achieved. As long as this asymmetry is small, it can however be compensated for through comparator calibration. Comparator calibration can however not compensate for offset voltage caused by large rapid common mode variations, as the comparators are calibrated only at the operating point at which the calibration is performed. Large offset voltage variations can be expected when the common-mode voltage is changed, making it important to keep it constant.

The calibration scheme chosen for the comparator builds on switched charge control [67]. The two calibration transistors are placed in parallel with the input transistors. These are then driven by a charge switching network enabled during the calibration phase. Two large capacitors stores the calibration voltage. The voltage on one side is generated through a resistive divider while the other is generated by adding or subtracting a small charge from the storage capacitor, thus changing the voltage [67]. This is a flexible, precise and low-power calibration approach, although it constantly need to be in operation as the storage capacitors experience leakage. This leakage causes the calibration voltage to slowly drift towards 0, thereby making the calibration inaccurate.

5.2 Capacitive DAC

The CDAC performs two functions, it both acts as a sampling capacitor and as a reference level generator. The reference levels are generated through charge-redistribution; by switching the back-plates of the capacitors, it is possible to change the voltage on the top-plate. Most CDACs are organized in a single-balanced topology, that is, switching only occurs in one of the legs at a time [80, 93]. This however would cause the common-mode voltage to change for each decision. In a double-balanced CDAC, switching takes place after every comparator decision, thus keeping the common-mode voltage constant [82]. This however comes at the cost of twice as many capacitive elements, although the total capacitance can be kept constant. With a constant common-mode voltage, a smaller comparator offset voltage would result.

With the double-balanced topology it becomes possible to both add and subtract charge from the CDAC. This as one of the parallel capacitors in each stage is connected to V_{ref} while the other is connected to ground while sampling. Switching the grounded signal to V_{ref} adds charge to the leg, thus increasing the voltage on the top-plate while the opposite happens if a signal is switched from V_{ref} to ground.

In this implementation a double-balanced redundantly scaled CDAC, shown in figure 5.3, is used. It is based on the scaling [128, 64, 28, 16, 8, 4, 4, 2, 1], where fractional reference voltages are used at the final two stages in order to reduce the number of unit cells needed. The number of unit cells used in each stage is shown in figure 5.3. This scaling does not result in the largest settling time reduction, as that would result in significantly increased usage of fractional voltages, reducing the linearity. The fractional reference voltage will cause increased settling time as they are generated through resistive dividers, thus having a larger series resistance to the reference buffer.

Matching does not depend only in the unit cell. Common centroid placement schemes are commonly used reduce effects of mismatch [62, Ch. 5]. However, as the reference voltage is stored in capacitive buffers mainly located outside the CDAC, and routed to each unit cell, the path resistance will affect the settling time. Thereby it also becomes important to limit the number of unit cells simultaneously switching on each row as these cells share the same reference voltage routing. This is illustrated in figure 5.4. The chosen unit cell placement for the CDAC together with the top-plate routing and reset-switch placement is shown in figure 5.5.

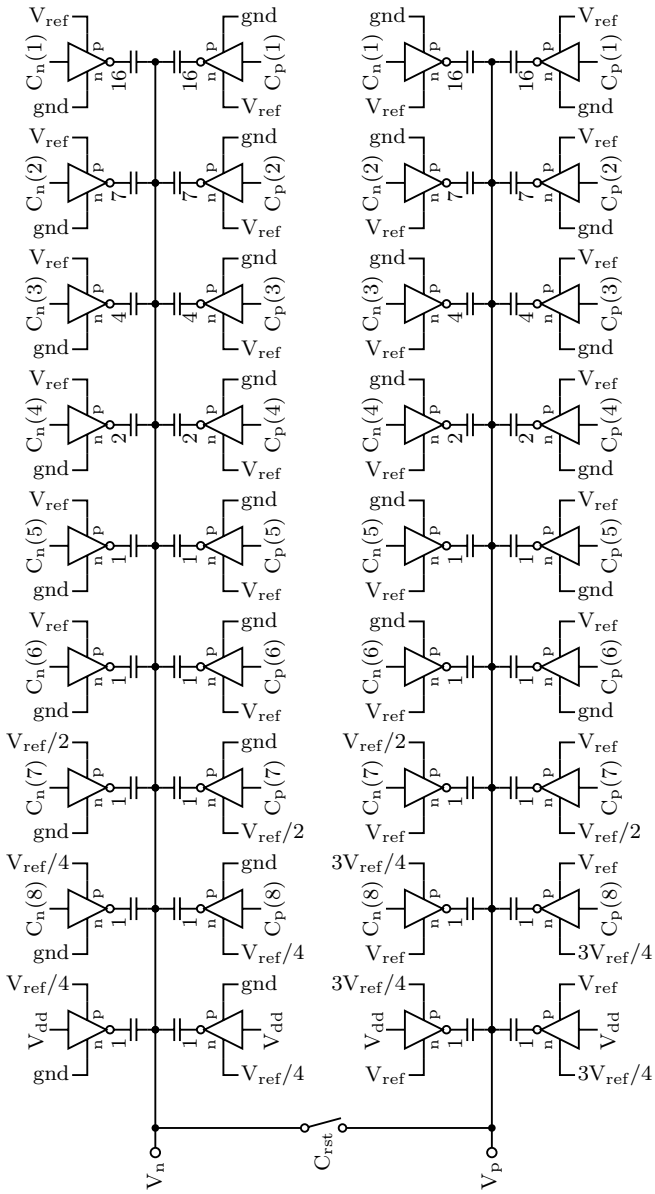


Figure 5.3: Schematic representation of the double balanced CDAC. The number of parallel unit cells in each stage is highlighted by the number next to the cell.

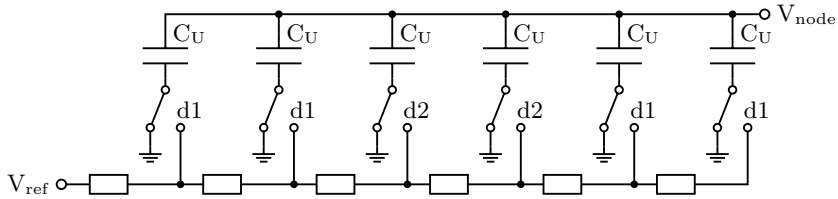


Figure 5.4: Illustration of switching-dependent unit cell placement used to minimize the effects of the trace resistance.

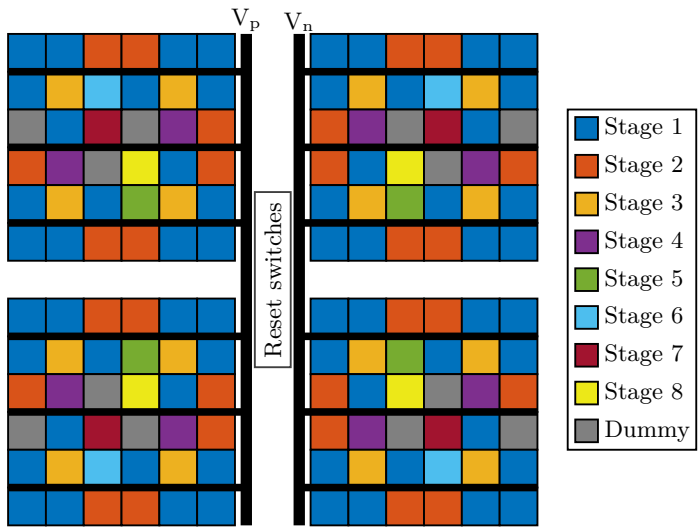


Figure 5.5: Illustration of the common centroid unit cell placement, accounting for switching-dependent settling time, used in the CDAC.

5.3 Asynchronous logic

With a comparator decision time logarithmically dependent on the voltage difference applied on the comparator input, a large variation in decision time can be expected [82]. In a binary scaled converter only a single decision is smaller than 1 LSB, while in a redundantly scaled converter this might increase to two comparisons. The stage at which this long decision will happen will not be known as it is signal dependent.

The majority of all digital systems uses synchronous logic, where each cycle is defined by a synchronization signal such as a clock. However, in a synchronous SAR converter, the large variation in decision time would result in the need of margins in each iteration, accounting for a long decision. With asynchronous logic, it becomes possible to share the margin as comparisons start directly after the previous finishes, thereby improving speed [82, 84]. A drawback is however that the timing cannot be tuned, a potential problem when it comes to testing the circuit.

The SAR logic consists of two parallel asynchronous blocks: a clock generator controlling the comparators, and a state machine for keeping track of the progress and storing the comparison results properly. Although both blocks share the comparator outputs as common inputs, it still becomes important to keep the delay fairly equal in the two loops in order to prevent unstable operation where blocks get out of sync. The asynchronous paths have therefore been designed for equal number of transistor stages, making the logic more tolerable to variations in transistor performance.

Although a conversion cycle is performed completely asynchronously, the synchronous sampling is used to initialize the asynchronous logic, thereby preventing it from propagating the effect of an incomplete conversion to the next stage. There is no additional cost as the main purpose of the sampling clock is to provide uniformly distributed samples.

5.4 Testability

In order to facilitate a thorough evaluation for the ADC, it is important that the design support various different tests. The largest complexity lies in storing the data generated in the converter. As the results are asynchronously generated, one bit at a time, and deleted at the end of the conversion in order to prepare for the next conversion, the data must be re-captured in order to be synced to the sampling clock. A

similar state machine as the one used in the converter is also used for re-capturing the data and sync it to the clock.

Just as for evaluating the RF-DAC, there exist two options for storing the data, either internally using an on-chip memory or externally by feeding the data off-chip at conversion rate. Internal memories are most commonly adopted for high-speed converters [82, 84]. It however comes with an increased design effort. Directly feeding the data off-chip comes with less design effort but might cause increased noise on the chip due to the output switching activity. Directly outputted data also allows for more freedom in detecting the source of potential incorrect behaviour when syncing the data to the clock.

With limited access to memory compilers in the 28 nm process used, designing a memory interface became too complex. Thus arose the need to feed data off-chip at the sample rate. The output drives were designed to both function as differential drivers as well as single-ended drivers, allowing for analysing the actual comparator outputs. The comparator outputs are differential when operating properly, making differential signalling possible. Incomplete conversion could be detected by monitoring the single-ended signals by checking if they both where 0. Clocking the outputs on both edges where chosen to reduce the output switching activity.

Externally controlling the master reference allows for adjustments of the dynamic range. By also having external access to the internally generate reference voltage, it not only becomes possible to monitor the function of the reference voltage generation. By providing a larger voltage on the internal reference voltage pin than on the master reference voltage pin, the internal reference voltage generation can be overridden, giving full control of the reference voltage used.

5.5 Manufactured chip

The ADC has been implemented and fabricated in STs 28 nm FDSOI CMOS process. The entire design measures 1.12 mm by 1.12 mm. The ADC core measures 87.5 µm by 78.5 µm, giving a total size of 6869 μm^2. A chip photo is shown in figure 5.6 together with a zoomed-in illustration of the ADC core, taken from the layout view in the design tool. This illustration highlights the placement of all the essential components in the converter.

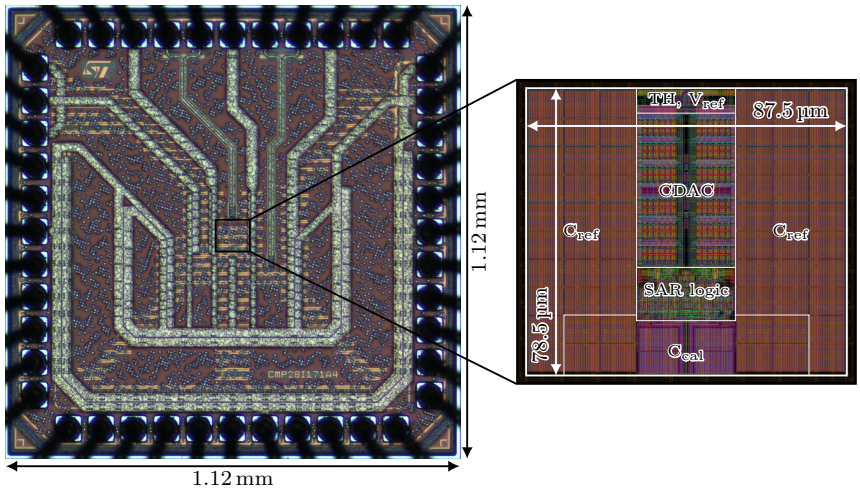


Figure 5.6: Chip photo showing the fabricated chip together with an illustration of the ADC captured from the design environment.

5.6 Measurements

To evaluate the performance for an ADC, both the static and the dynamic properties needs to be evaluated. The dynamic properties are evaluated by measuring a single tone at various frequencies. The SNDR and SFDR can then be extracted from the output spectrum. When measuring the SNDR, typically a 1 dB back-off from full-scale is used to avoid clipping effects while utilizing the full range.

Measuring the linearity is often done by feeding a saw-tooth signal, however, in high-speed converters this is not doable due to the complexity of generating a precise saw-tooth waveform [62, Ch. 11]. In converters where comparators rely on background calibration of the offset voltage, a slowly changing input signal might cause the calibration to drift. Rather statistical metrics are used to evaluate the converter linearity [62, Ch. 11]. The statistical distribution of a sine wave signal result is more samples located at the outer borders of the conversion range while a smaller number will fall within the range.

The chip has been mounted in a QFN package which has been soldered onto a PCB. All measurement equipment has then been connected to the PCB. The measurements have been performed using two signal generators, one for generating the clock at 8 times the desired sample

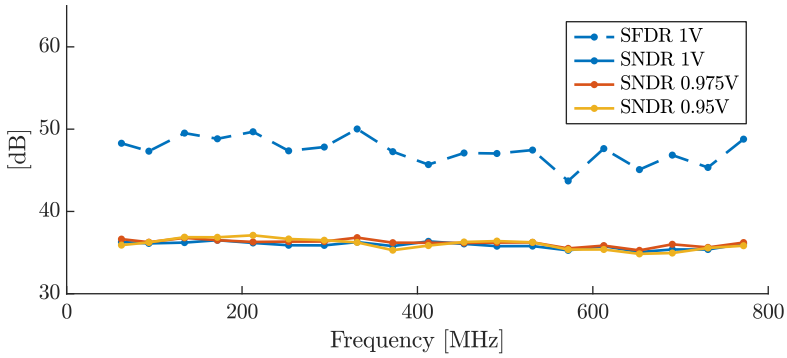


Figure 5.7: Measured SFDR and SNDR at various signal frequencies using $P_{in} = -1\,\mathrm{dB_{FS}}$ and $f_s = 800\,\mathrm{MS/s}$.

rate, and the other to be used for generating the signal to be measured. A logic analyzer was used for capturing the output signals, allowing for capturing of different record lengths, all being a power of two. The signal frequencies measured where chosen to result in an integer number of periods, given the record length and the sample rate chosen to eliminate FFT artefacts appearing from aperiodic signals.

5.6.1 Measured performance

All measurements have been performed at a nominal sample rate of 800 MS/s, using a full-scale range of $500\,\mathrm{mV_{pp,diff}}$. The measured SFDR and SNDR are plotted for various input signal frequencies in figure 5.7. From the figure it can be seen that the SNDR is tightly grouped for supply voltages in the range 0.95–1 V. With a nominal supply voltage of 1 V, this indicates that the chosen sample rate is not too high to result in incomplete comparisons. An output spectrum for a 371.9 MHz sine input is shown in figure 5.8. Here a large number of spurs can be observed, limiting both the SNDR and SFDR. The DNL and INL are shown in figures 5.9a and 5.9b.

The converter achieves a measured SNDR of 35.8 dB, giving an ENOB of 5.65, and an SFDR of 46.8 dB. The measured power consumption is 2.16 mW at a 1 V supply. A total of 2.7 pJ is consumed for each conversion, giving a FoMW of 53.6 fJ/conv.-step. An DNL between −1 LSB and 2 LSB and an INL smaller than ±2.5 LSB has also been measured. The performance is summarized and compared with

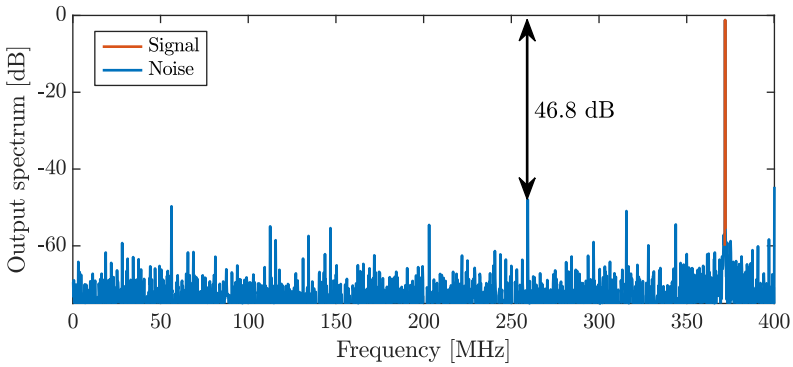
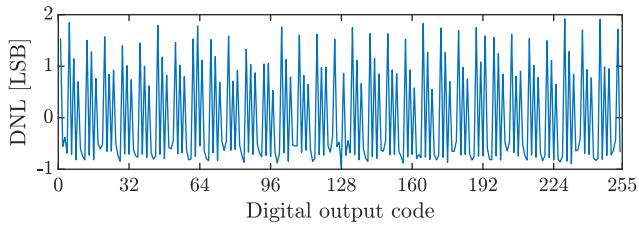
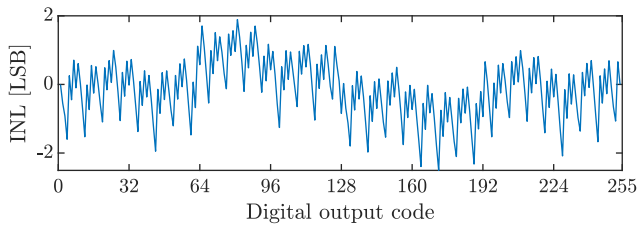


Figure 5.8: ADC output spectrum for $f_{sig} = 371.9\,\text{MHz}$, $P_{in} = -1\,\text{dB}_{\text{FS}}$, and $f_s = 800\,\text{MS/s}$.



(a)



(b)

Figure 5.9: Measured DNL (a), and INL (b).

Table 5.1: Performance comparison.

	[82]	[84]	[85]	This work
Technology	32 nm SOI	28 nm	28 nm	28 nm SOI
Sample rate [GS/s]	1.2	1	0.75	0.8
Resolution [b]	8	8	8	8
SNDR [dB]	39.3	43.6	45.2	35.8
SFDR [dB]	49.9	58.6	57.5	46.8
Power [mW]	3.06	3.2	4.5	2.16
FOMW [fJ/conv.-step]	34	25.9	41	53.6
Area [mm^2]	0.0015	0.006 75	0.004	0.0069

other state-of-the-art converters in table 5.1.

5.7 Evaluation and performance analysis

The measured power consumption closely matches the simulated values. A constant dynamic performance over a 50 mV supply voltage range indicates that the converter is not speed limited with a 1 V supply. The dynamic performance on the other hand is significantly lower than indicated by simulations.

Evaluating the performance of an asynchronous circuit is significantly more tricky than for a synchronous circuit. As the timing of the circuit is independent of the operation frequency, a key parameter is removed when trying to understand the reason behind the degraded performance. For example, some effects, such as incomplete comparator reset, may in synchronous logic be reduced by reducing the clock rate, thereby allowing more time for them. However, when the reset time is derived asynchronously, this will not be possible without changing other operating conditions, such as the supply voltage.

The measured data has been studied in search for patterns that may point towards a problem. We have then tried to replicate these patters using simulations. Three different error types are studied below: leakage from the output drivers, noise expansion around the signal, and comparator offset voltage.

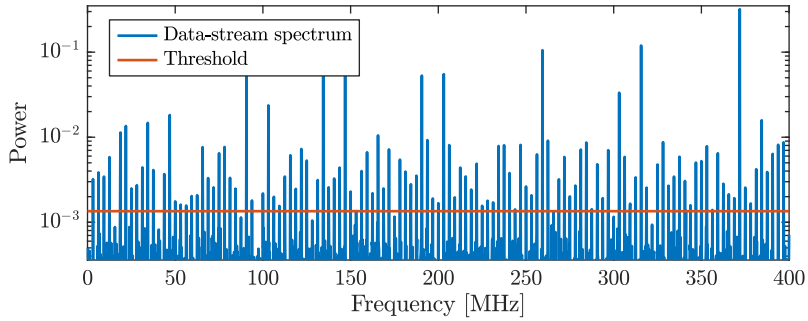


Figure 5.10: Summed power spectrum for the 9 data streams.

5.7.1 Output clock and data feedthrough

As mentioned before, the spectrum presented in figure 5.8 shows a large number of significant spurs which are not predicted by simulations. The large spur located at 400 MHz corresponds to the frequency of the half-rate clock used for syncing the output data. This indicates that this clock is leaking into the input signal. Furthermore, this and other spurs do not follow the input signal power, thereby pointing towards a digital origin.

Based on the assumption that the digital noise is dominated by the switching of the output pins, including the clock, an individual spectrum for each of the outputs has been calculated. The spectral power is then summed per frequency bin. The result is presented in figure 5.10. The bins with large values, indicating potential leakage, are sparsely distributed at frequencies depending on the signal frequency and the sample rate.

Comparing the frequencies at which spurs occur in figure 5.8 with the bins showing high total spectral power in figure 5.10, an overlap can be observed. However, a large value in the combined spectrum might not be connected to a large spur, and a large spur might be represented by a smaller leakage values. This is because in reality, the coupling of the different data streams is significantly more complex than a simple addition. Rather than just picking bins experiencing obvious leakage, a more methodical approach was chosen. A threshold was introduced and for all the bins in figure 5.10 exceeding the threshold, the value in the corresponding frequency bin in figure 5.8 was replaced by the average noise power. This operation will suppress all the spurs likely

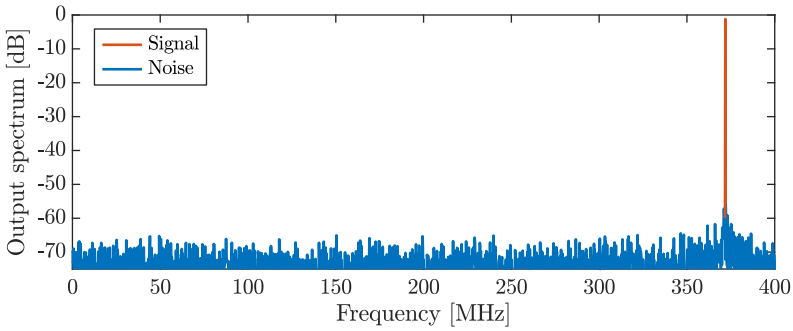


Figure 5.11: Output spectrum cleaned from data-dependent spurs.

caused by feedthrough from the digital pins. A lower threshold will result in suppression of more spurs and a larger improvement in SNDR. While sweeping the threshold, we observed no improvement in the SNDR after adjusting 6 % of the frequency bins. A spectrum with the data-dependent spurs removed is shown in figure 5.11. The estimated noise power caused by the data-dependent spurs is −38.4 dBc giving an SNDR of 37.6 dB for the cleaned spectrum. For the input frequency chosen, the suppressed spurs also include harmonics of the input frequency. Circuit simulations predict these to be significantly smaller than shown in figure 5.8, but these predictions cannot be verified.

5.7.2 Excess noise around the signal

Zooming in on figure 5.11, a small increase in the noise level around the signal can be observed, as shown in figure 5.12a. This increased noise is most likely caused by a small misalignment between the signal frequency and the sample rate. The accumulative noise power starting at the signal frequency and moving away from it are shown in figure 5.12b. With evenly-distributed noise, the accumulated noise should grow linearly, and the extrapolated intersection should be at the value 0 at the signal frequency. In the case of excess noise, the intersection point of the linear extrapolations will move upwards, representing the one-sided excess noise contribution. The intersection might here not be located at the signal frequency as the noise on each side can be different.

Looking at figure 5.12a, it would be expected that this contribution is small. This noise contribution is estimated to −45.8 dBc, thus causing a significantly smaller SNDR degradation than the data-dependent spurs.

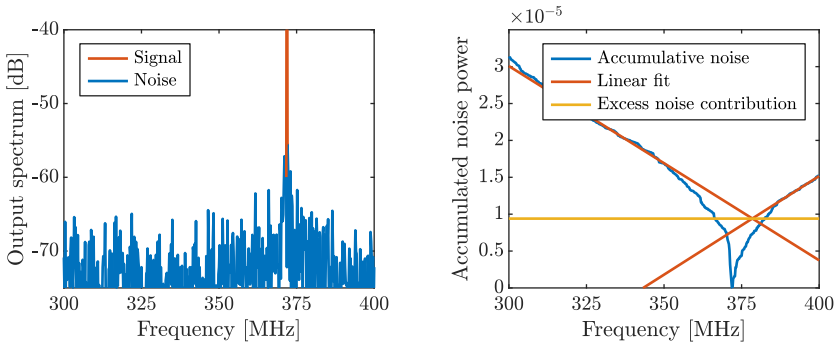


Figure 5.12: A zoomed-in portion of the spectrum in figure 5.11 is shown in (a). The accumulated noise power is shown in (b) together with linear fitted noise slopes. The intersection level gives the excess noise contribution per side.

5.7.3 Comparator offset voltage

The alternating-comparator SAR topology is sensitive to offset voltages, as pointed out above. The comparator calibration scheme chosen in this implementation compensates for static and slowly varying offsets. However, the offset voltage changing from decision to decision, that may be caused by the asynchronous timing cannot be calibrated for.

For comparators relying of positive feedback, such as the strong-ARM used here, the decision time depends on the input-voltage difference: a small difference needs longer time to cause the imbalance which is then resolved by the feedback. The mechanism is the same as for FF metastability resolution. The simulated decision time as a function of the input voltage is shown in figure 5.13b.

In addition, the comparators rely on complete equalization of its internal nodes during reset, for a low input-referred offset voltage. Disregarding any circuit asymmetries, any voltage difference remaining between the two decision nodes will translate into an offset voltage. The comparator reset phase is intended to reduce this difference enough, so that the offset voltage contribution becomes insignificant. As shown in figure 5.2, all decision nodes are pulled high by the reset switches, and will exponentially approach the supply voltage as set by the decision node capacitance and the resistance of the reset-switch. While the sign

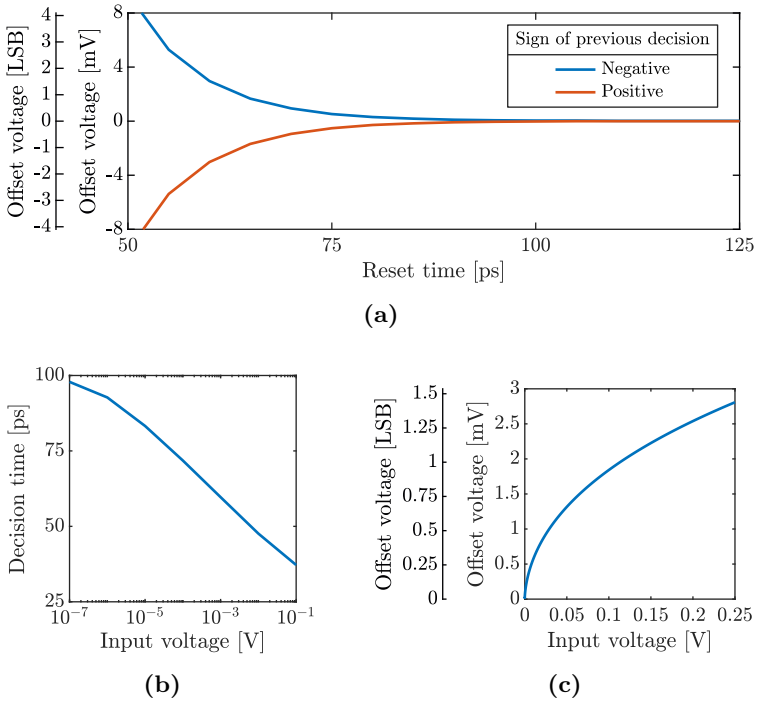


Figure 5.13: Simulated reset-induced offset voltage. The input-referred offset voltage as a function of reset duration for a previous positive or negative decision (a). The comparator decision time as a function of input voltage (b). The resulting offset voltage as a function of the offset voltage for the previous decision (c).

of the voltage difference will be kept during the entire reset phase, the magnitude (and therefore the offset voltage in the up-coming decision) will depend on the time spent in reset. The simulated offset voltage as a function of reset time is shown in figure 5.13a.

Combining figures 5.13a and 5.13b gives the comparator offset voltage as a function of input voltage for the previous decision, shown in figure 5.13c. Most decisions are performed at low input voltage differences: roughly 50 % of the comparisons have a voltage difference of less than 10 mV.

While effects of comparator offset voltage will appear as white noise in the spectrum, thereby increasing the noise floor, it offers few clues about the offset values. Rather, we have used histograms of the output

values to estimate the offset voltages. Just as for the linearity measurements, the distribution of samples should follow the bathtub shape given a single sine wave input signal [97]. Most of the samples will fall into the bins closest to the edges while the bins in the middle should have similar number of hits. If an offset voltage difference between the two comparators is present, the histogram will no longer be smooth for the middle bins. Instead, the number of samples per bin will fall into a repetitive pattern, depending on the offset voltage in the two comparators and at the reference scaling used.

A behavioural model was constructed to analyse what influence on the performance the offset voltage have. In addition to using fixed offset voltages, the model also incorporates the reset-induced offset voltage presented in figure 5.13c. The redundant scaling is also modelled in order to account for the decision level symmetries. For the first two decisions in each conversion, complete reset is assumed as both comparators are reset during the sampling. The behavioural model includes three tunable parameters, the fixed comparator offset voltage unique to each comparator and a scale factor for the relationship presented in figure 5.13c. We tuned these parameters to recreate the behaviour observed in the ADC. Without calibration, the scale factor was set to 1.05 and the fixed offset voltages to 1.9 LSB for comparator one and 0 LSB for comparator two. The central part of the histogram for the behavioural model is shown in figure 5.14a together with the histogram for the measured data to the right in figure 5.14b. The estimated noise contribution due to offset voltage is -40 dBc.

With calibration enabled, we arrived at the same scale factor value. The fixed offset voltage for comparator one was 0 LSB and for comparator two 0.95 LSB. The resulting noise contribution then becomes -44 dBc. In a similar manner as above, the histograms for the behavioural simulation and the measured data are shown in figures 5.14c and 5.14d.

The noise contribution for the data-dependent spurs calculated above is for the 371.9 MHz tone presented in figure 5.8. The requirements for the offset voltage estimation is however a bit different, thereby resulting in different results for the data-dependent spurs. With calibration disabled the measured SNDR was 36.2 dB. After removing the data-dependent spurs, an SNDR of 37.8 dB was achieved, giving a noise contribution of -39.8 dBc. With calibration activated, an SNDR of 37.6 dB was measured. This gives an SNDR for the cleaned spectrum of 38.8 dB,

giving a noise contribution of 44 dBc.

Adding the noise contributions from the excess noise around the signal and the comparator offset voltages gives a total noise contribution of −36.9 dBc with calibration disabled and −41.8 dBc with calibration enabled. The resulting SNDR after removing these noise sources is 44.5 dB with calibration disables and 41.7 dB when calibration is enabled. Circuit level simulations indicated an SNDR of 43 dB. This close agreement between the estimated SNDR and the simulated indicates that the estimated contributions of the studied noise sources are realistic. The estimated SNDR with calibration disabled appears to be a bit optimistic. However, the simple behavioural model used to estimate the offset voltage simplifies the complex relationship in the real circuit, thereby explaining the difference observed here.

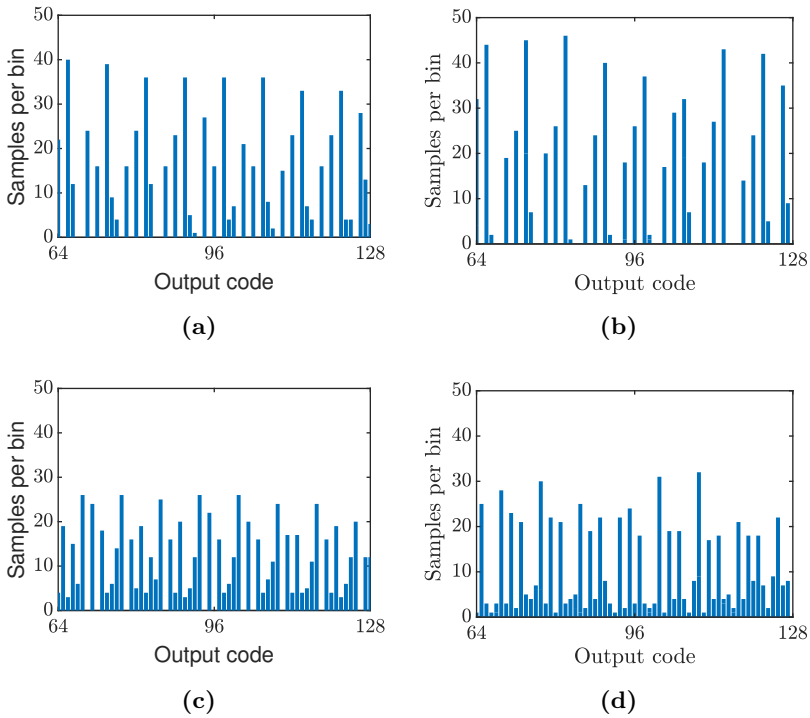


Figure 5.14: Histogram of measured data over 4096 samples without calibration (b), and with calibration (d). The corresponding histograms for the behavioural model used to estimate offset voltage without calibration (a), and with calibration (c). The estimated offset voltages with out calibration are 1.9 LSB and 0 LSB for comparator 1 and 2 respectively. With calibration enabled, the offset voltages are 0 LSB and 0.95 LSB for the two comparators. Only the central part of the histograms are included here.

Chapter 6

Conclusions

With the increased demands on higher data rates, new transmitters are needed. With the introduction of massive MIMO transceivers, operating at mmWave frequencies, compact and low cost transmitter realizations are needed. The closed-loop transmitter opens the path towards lower complexity analog implementation through an extensive use of DSP algorithms. For this, low cost and highly integratable data converters are needed.

In this thesis, I have presented data converters both for D/A and A/D conversion. These data converters have been designed to enable low-cost closed-loop mmWave transmitters needed for realizing massive MIMO transmitters having hundreds of antennas.

A 2×6 bit RF-DAC-based IQ modulator operating at mmWave frequencies have been demonstrated. It combines multiple functions into a single unit, bringing a compact and effective building block. The fabricated RF IQ modulator has been shown to operate upto 8 GHz, reaching data rates of 4.6 Gb/s while being able to handle various different modulation formats.

The evaluation of an 800 MS/s SAR ADC has also been presented in this thesis. For this alternating comparator based converter, the difference between simulated and measured performance was studied in order to understand the limitations in our fabricated circuit. Methods for estimating data-dependent switching noise and offset voltage between the alternating comparators have been presented.

6.1 Future work

A number of different areas are of interest to explore in future research. These include both the data converters them self and co-integration of them into larger systems.

6.1.1 Increased RF-DAC resolution

In order to fulfil the EVM requirements for the high complexity modulation formats, such as 256-QAM, used in modern communication systems [3], the RF-DAC resolution needs to be increased. With an increased resolution the number of unit cells rapidly increase, doubling for each additional bit of resolution, thus rapidly making it complex to support all connections needed in a flat hierarchy. This makes it beneficial to explore what benefits a hierarchical unit cell arrangement has both on the design-flow and on the RF-DAC performance.

6.1.2 Quadrature LO generation

The quadrature LO generation is the block within the current RF-DAC design with the highest power consumption. It therefore becomes important to explore different topologies, not only focusing on their power consumption but also consider their footprint and bandwidth.

Something else that here would be beneficial to explore further is generation of non-overlapping 25 % duty cycle LOs. Non-overlapping LOs show great potential but are highly complex to generate at mmWave frequencies, making it important to further explore their generation.

6.1.3 Co-integration with PA

Thanks to their small footprint, RF-DAC based IQ modulators can be tightly integrated with other components, such as PAs, enabling exotic topological options to explore. One such topology would be a Doherty PA, where the input path is split and individual input signals are provided to the different branches.

6.1.4 Closed-loop transmitter demonstration

To better understand the benefits with the feedback path, it would be beneficial to study it in an actual circuit. Although measurements

to some extent can mimic the closed-loop behaviour, a tightly integrated observation receiver will experience analog impairments just as the transmitter, thus reducing its accuracy. Implementing a closed-loop transmitter is a large project requiring a large joint effort from several persons. At the same time, it enables studies on how the observation path improves the performance, hopefully bringing understanding on what performance the different building blocks need to achieve.

To enable the implementation of a close-loop transmitter, a new ADC implementation will be needed in the same process as all the other building blocks. Implementing a new ADC allows for further exploration of the alternating comparator topology, using the understanding that we have gained through the evaluation presented in this thesis.

References

[1] P. Jonsson, S. Carson, G. Blennerud, J. K. Shim, B. Arendse, A. Husseini, P. Lindberg, and K. Öhman, "Ericsson mobility report," November 2019. [Online]. Available: https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf

[2] "The european table of frequency allocations and applications in the frequency range 8.3 kHz to 3000 GHz (ECA table)," March 2019. [Online]. Available: https://www.ecodocdb.dk/download/2ca5fcbd-4090/ERCREP025.pdf

[3] *5G; NR; Base Station (BS) radio transmission and reception*, ETSI 3GPP Std. 138.104 (V 15.8.0), Jan. 2020.

[4] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The new 5G radio access technology," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 24–30, 2017.

[5] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[6] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[7] X. Yang, M. Matthaiou, J. Yang, C. Wen, F. Gao, and S. Jin, "Hardware-constrained millimeter-wave systems for 5G: Challenges, opportunities, and solutions," *IEEE Communications Magazine*, vol. 57, no. 1, pp. 44–50, 2019.

[8] P. Suryasarman and A. Springer, "A comparative analysis of adaptive digital predistortion algorithms for multiple antenna transmitters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 5, pp. 1412–1420, May 2015.

[9] S. Shopov, N. Cahoon, and S. P. Voinigescu, "Ultra-broadband I/Q RF-DAC transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 12, pp. 5411–5421, Dec 2017.

[10] H. J. Qian, Y. Shu, J. Zhou, and X. Luo, "A 20-32-GHz quadrature digital transmitter using synthesized impedance variation compensation," *IEEE Journal of Solid-State Circuits*, pp. 1–13, 2020.

[11] D. M. Pozar, *Microwave Engineering.* Wiley, 2012.

[12] H. Wang, F. Wang, S. Li, T.-Y. Huang, A. S. Ahmed, N. S. Mannem, J. Lee, E. Garay, D. Munzer, C. Snyder, S. Lee, H. T. Nguyen, and M. E. D. Smith. (2020, feb) Power amplifiers performance survey 2000-present. [Online]. Available: https://gems.ece.gatech.edu/PA_survey.html

[13] J. Zhang, X. Ge, Q. Li, M. Guizani, and Y. Zhang, "5G millimeter-wave antenna array: Design and challenges," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 106–112, 2017.

[14] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[15] C. Cahn, "Combined digital phase and amplitude modulation communication systems," *IRE Transactions on Communications Systems*, vol. 8, no. 3, pp. 150–155, Sep. 1960.

[16] C. Campopiano and B. Glazer, "A coherent digital amplitude and phase modulation scheme," *IRE Transactions on Communications Systems*, vol. 10, no. 1, pp. 90–95, March 1962.

[17] L. L. Hanzo, S. X. Ng, T. Keller, and W. Webb, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems.* IEEE, 2004.

[18] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Transactions on Broadcasting*, vol. 54, no. 2, pp. 257–268, 2008.

[19] U. Dalal, *Wireless Communication and Networks*. Oxford University Press, 2015.

[20] Q. Gu, *RF System Design of Transceivers for Wireless Communications*. Springer, 2005.

[21] "Using error vector magnitude measurements to analyze and troubleshoot vector-modulated signals," Agilent Technologies, techreport 89400-14. [Online]. Available: https://literature.cdn.keysight.com/litweb/pdf/5965-2898E.pdf

[22] "Critical RF measurements in cable, satellite and terrestrial DTV systems," Tektronix, techreport 2TW-17370-2. [Online]. Available: https://download.tek.com/document/2TW_17370_2_HR.pdf

[23] H. Al-Rubaye and G. M. Rebeiz, "W -band direct-modulation >20-Gb/s transmit and receive building blocks in 32-nm SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 9, pp. 2277–2291, Sep. 2017.

[24] M. Talonen and S. Lindfors, "Power consumption model for linear RF power amplifiers with rectangular M-QAM modulation," in *4th International Symposium on Wireless Communication Systems*, Oct 2007, pp. 682–685.

[25] Shuguang Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, Sep. 2005.

[26] K. Kouassi, G. Andrieux, and J.-F. Diouris, "PAPR distribution for single carrier M-QAM modulations," *Wireless Personal Communications*, vol. 104, no. 2, pp. 727–738, 2019.

[27] A. Grebennikov, *RF and Microwave Transmitter Design*, ser. Wiley Series in Microwave and Optical Engineering. Wiley, 2011.

[28] S. Luschas, R. Schreier, and Hae-Seung Lee, "Radio frequency digital-to-analog converter," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1462–1467, Sep. 2004.

[29] P. Eloranta and P. Seppinen, "Direct-digital RF modulator IC in 0.13 µm CMOS for wide-band multi-radio applications," in *IEEE International Solid-State Circuits Conference, (ISSCC)*, Feb 2005, pp. 532–615 Vol. 1.

[30] P. Eloranta, P. Seppinen, S. Kallioinen, T. Saarela, and A. Parssinen, "A multimode transmitter in 0.13 µm CMOS using direct-digital RF modulator," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 12, pp. 2774–2784, Dec 2007.

[31] A. Agah, W. Wang, P. Asbeck, L. Larson, and J. Buckwalter, "A 42 to 47-GHz, 8-bit I/Q digital-to-RF converter with 21-dBm P_{sat} and 16% PAE in 45-nm SOI CMOS," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2013, pp. 249–252.

[32] A. Balteanu, I. Sarkas, E. Dacquay, A. Tomkins, G. M. Rebeiz, P. M. Asbeck, and S. P. Voinigescu, "A 2-Bit, 24 dBm, millimeter-wave SOI CMOS power-DAC Cell for watt-level high-efficiency, fully digital m-ary QAM transmitters," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 5, pp. 1126–1137, May 2013.

[33] S. M. Alavi, R. B. Staszewski, L. C. N. de Vreede, and J. R. Long, "A wideband 2×13-bit all-digital I/Q RF-DAC," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 732–752, April 2014.

[34] S. Shopov, A. Balteanu, and S. P. Voinigescu, "A 19 dBm, 15 Gbaud, 9 bit SOI CMOS power-DAC cell for high-order QAM W-band transmitters," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 7, pp. 1653–1664, July 2014.

[35] H. Al-Rubaye and G. M. Rebeiz, "A 20 Gbit/s RFDAC-based direct-modulation W-band transmitter in 32nm SOI CMOS," in *IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct 2016, pp. 1–4.

[36] K. Dasgupta, S. Daneshgar, C. Thakkar, K. Datta, J. Jaussi, and B. Casper, "A 25 Gb/s 60 GHz digital power amplifier in 28nm CMOS," in *43rd IEEE European Solid State Circuits Conference, (ESSCIRC)*, Sept 2017, pp. 207–210.

[37] S. Shopov, O. D. Gurbuz, G. M. Rebeiz, and S. P. Voinigescu, "A D-band digital transmitter with 64-QAM and OFDM free-space constellation formation," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 7, pp. 2012–2022, July 2018.

[38] K. Khalaf, V. Vidojkovic, K. Vaesen, J. R. Long, W. Van Thillo, and P. Wambacq, "A digitally modulated 60GHz polar transmitter in 40nm CMOS," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2014, pp. 159–162.

[39] K. Khalaf, V. Vidojkovic, K. Vaesen, M. Libois, G. Mangraviti, V. Szortyka, C. Li, B. Verbruggen, M. Ingels, A. Bourdoux, C. Soens, W. V. Thillo, J. R. Long, and P. Wambacq, "Digitally modulated CMOS polar transmitters for highly-efficient mm-wave wireless communication," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 7, pp. 1579–1592, July 2016.

[40] H. J. Qian, J. O. Liang, N. Zhu, P. Gao, and X. Luo, "A 3.1-7 GHz 40-nm CMOS digital polar transmitter with high data-rate and feed-forward operation," in *IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, Aug 2016, pp. 1–3.

[41] M. Hashemi, Y. Shen, M. Mehrpoo, M. Acar, R. van Leuken, M. S. Alavi, and L. de Vreede, "An intrinsically linear wideband digital polar PA featuring AM-AM and AM-PM corrections through nonlinear sizing, overdrive-voltage control, and multiphase RF clocking," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2017, pp. 300–301.

[42] T. Buckel, P. Preyler, A. Klinkan, D. Hamidovic, C. Preissl, T. Mayer, S. Tertinek, S. Brandstaetter, C. Wicpalek, A. Springer, and R. Weigel, "A novel digital-intensive hybrid polar-I/Q RF transmitter architecture," *IEEE Transactions on Circuits and Systems I: Regular Papers*, pp. 1–14, 2018.

[43] D. Zhao, S. Kulkarni, and P. Reynaert, "A 60-GHz outphasing transmitter in 40-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 12, pp. 3172–3183, Dec 2012.

[44] B. Sadhu, Y. Tousi, J. Hallin, S. Sahl, S. Reynolds, . Renström, K. Sjögren, O. Haapalahti, N. Mazor, B. Bokinge, G. Weibull, H. Bengtsson, A. Carlinger, E. Westesson, J. E. Thillberg, L. Rexberg, M. Yeck, X. Gu, D. Friedman, and A. Valdes-Garcia, "A 28GHz 32-element phased-array transceiver IC with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5G communication," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2017, pp. 128–129.

[45] T. Maruyama, K. Tsutsumi, E. Taniguchi, and M. Shimozawa, "1.4 deg.-rms 6-bit vector-sum phase shifter calibrating I-Q generator error by VGA for high SHF wide-band massive MIMO in 5G," in *Asia-Pacific Microwave Conference (APMC)*, Dec 2016, pp. 1–4.

[46] W. Yamamoto, K. Tsutsumi, T. Maruyama, T. Fujiwara, T. Hagiwara, A. Osawa, and M. Shimozawa, "A 28GHz 4-channel transmit/receive RF core-chip with highly-accurate phase shifter for high SHF wide-band massive MIMO in 5G," in *Asia-Pacific Microwave Conference (APMC)*, Nov 2018, pp. 753–755.

[47] Y. Tousi and A. Valdes-Garcia, "A Ka-band digitally-controlled phase shifter with sub-degree phase precision," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, May 2016, pp. 356–359.

[48] D. Huang, L. Zhang, D. Li, L. Zhang, Y. Wang, and Z. Yu, "A 60 GHz 360° 5-bit phase shifter with constant IL compensation followed by a normal amplifier with ±1 dB gain variation and 0.6 dBm $OP_{\text{-1dB}}$," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. PP, no. 99, pp. 1–1, 2016.

[49] F. Meng, K. Ma, K. S. Yeo, and S. Xu, "A 57-to-64-GHz 0.094-mm^2 5-bit passive phase shifter in 65-nm CMOS," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 5, pp. 1917–1925, May 2016.

[50] J. Lemberg, M. Kosunen, E. Roverato, M. Martelius, K. Stadius, L. Anttila, M. Valkama, and J. Ryynänen, "Digital interpolating phase modulator for wideband outphasing transmitters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 5, pp. 705–715, May 2016.

[51] A. K. Mustafa, S. Ahmed, and M. Faulkner, "Bandwidth limitation for the constant envelope components of an OFDM signal in a LINC architecture," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 9, pp. 2502–2510, Sep. 2013.

[52] J. Zhuang, K. Waheed, and R. B. Staszewski, "A technique to reduce phase/frequency modulation bandwidth in a polar RF transmitter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 8, pp. 2196–2207, Aug 2010.

[53] A. Mirzaei, H. Darabi, J. C. Leete, and Y. Chang, "Analysis and optimization of direct-conversion receivers with 25% duty-cycle current-driven passive mixers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 9, pp. 2353–2366, Sep. 2010.

[54] J. Han and K. Kwon, "RF receiver front-end employing IIP2-enhanced 25% duty-cycle quadrature passive mixer for advanced cellular applications," *IEEE Access*, vol. 8, pp. 8166–8177, 2020.

[55] A. Mirzaei, D. Murphy, and H. Darabi, "Analysis of direct-conversion IQ transmitters with 25% duty-cycle passive mixers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 10, pp. 2318–2331, Oct 2011.

[56] N. Weiss, S. Shopov, P. Schvan, P. Chevalier, A. Cathelin, and S. P. Voinigescu, "DC-62 GHz 4-phase 25% duty cycle quadrature clock generator," in *IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, Oct 2017, pp. 1–4.

[57] H. Jin, D. Kim, and B. Kim, "Efficient digital quadrature transmitter based on IQ cell sharing," *IEEE Journal of Solid-State Circuits*, vol. PP, no. 99, pp. 1–13, 2017.

[58] M. Ingels, D. Dermit, Y. Liu, H. Cappelle, and J. Craninckx, "A 2x14bit digital transmitter with memoryless current unit cells and integrated AM/PM calibration," in *43rd IEEE European Solid State Circuits Conference (ESSCIRC)*, Sept 2017, pp. 324–327.

[59] D. Zhao and P. Reynaert, "A 40 nm CMOS E-band transmitter with compact and symmetrical layout floor-plans," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2560–2571, Nov 2015.

[60] C. Tseng and C. Chang, "A rigorous design methodology for compact planar branch-line and rat-race couplers with asymmetrical t-structures," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 7, pp. 2085–2092, 2012.

[61] Keysight Technologies, *Keysight PNA Help*. [Online]. Available: http://na.support.keysight.com/pna/help/index.html

[62] M. Pelgrom, *Analog-to-Digital Conversion*. Springer, 2017.

[63] F. Maloberti, *Data Converters*. Springer, 2007.

[64] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[65] M. Gustavsson, J. J. Wikner, and N. N. Tan, *CMOS Data Converters for Communications*. Springer, 2002.

[66] B. Murmann. (2020) ADC performance survey 1997-2020. [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[67] M. Miyahara, Y. Asada, D. Paik, and A. Matsuzawa, "A low-noise self-calibrating dynamic comparator for high-speed ADCs," in *IEEE Asian Solid-State Circuits Conference (ASSCC)*, Nov 2008, pp. 269–272.

[68] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, Apr 1999.

[69] B. Murmann, "Limits on ADC power dissipation," in *Analog Circuit Design*, M. Steyaert, J. H. Huijsing, and A. H. van Roermund, Eds. Springer, 2006, ch. 16, pp. 351–367.

[70] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," in *IEEE Custom Integrated Circuits Conference (CICC)*, Sept 2008, pp. 105–112.

[71] R. Walden, "Analog-to-digital converter technology comparison," in *16th Annual Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, Oct 1994, pp. 217–219.

[72] R. Schreier and G. C. Temes, *Understanding Delta-Sigma Data Converters*, 1st ed. Wiley-IEEE Press, 2005.

[73] S. Pavan, R. Schreier, and G. C. Temes, *Understanding Delta-Sigma Data Converters*, 2nd ed. Wiley-IEEE Press, 2017.

[74] B. Raghavan, A. Varzaghani, L. Rao, H. Park, Xiaochen Yang, Zhi Huang, Yu Chen, R. Kattamuri, Chunhui Wu, B. Zhang, Jun Cao, A. Momtaz, and N. Kocaman, "A 125 mW 8.5–11.5 Gb/s serial link transceiver with a dual path 6-bit ADC/5-tap DFE receiver and a 4-tap FFE transmitter in 28 nm CMOS," in *IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, 2016, pp. 1–2.

[75] L. Kull, T. Toifl, M. Schmatz, P. Francese, C. Menolfi, M. Braendli, M. Kossel, T. Morf, T. Andersen, and Y. Leblebici, "A 3.1mW 8b 1.2GS/s single-channel asynchronous SAR ADC with alternate comparators for enhanced speed in 32nm digital SOI CMOS," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2013, pp. 468–469.

[76] Z. Zheng, L. Wei, J. Lagos, E. Martens, Y. Zhu, C. Chan, J. Craninckx, and R. P. Martins, "A single-channel 5.5mW 3.3GS/s 6b fully dynamic pipelined ADC with post-amplification residue generation," in *IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 254–256.

[77] L. Kull, J. Pliva, T. Toifl, M. Schmatz, P. A. Francese, C. Menolfi, M. Brändli, M. Kossel, T. Morf, T. M. Andersen, and Y. Leblebici, "Implementation of low-power 6-8 b 30-90 GS/s time-interleaved ADCs with optimized input bandwidth in 32 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 3, pp. 636–648, March 2016.

[78] B. Hershberg, B. v. Liempd, N. Markulic, J. Lagos, E. Martens, D. Dermit, and J. Craninckx, "A 6-to-600MS/s fully dynamic ringamp pipelined ADC with asynchronous event-driven clocking in 16nm," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2019, pp. 68–70.

[79] B. Hershberg, D. Dermit, B. v. Liempd, E. Martens, N. Markulic, J. Lagos, and J. Craninckx, "A 3.2GS/s 10 ENOB 61mW ringamp ADC in 16nm with background monitoring of distortion," in *IEEE International Solid- State Circuits Conference - (ISSCC)*, Feb 2019, pp. 58–60.

[80] C.-C. Liu, S.-J. Chang, G.-Y. Huang, Y.-Z. Lin, C.-M. Huang, C.-H. Huang, L. Bu, and C.-C. Tsai, "A 10b 100MS/s 1.13mW SAR ADC with binary-scaled error compensation," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2010, pp. 386–387.

[81] T. Ogawa, H. Kobayashi, M. Hotta, Y. Takahashi, H. San, and N. Takai, "SAR ADC algorithm with redundancy," in *IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, Nov 2008, pp. 268–271.

[82] L. Kull, T. Toifl, M. Schmatz, P. Francese, C. Menolfi, M. Brandli, M. Kossel, T. Morf, T. Andersen, and Y. Leblebici, "A 3.1 mW 8b 1.2 GS/s single-channel asynchronous SAR ADC with alternate comparators for enhanced speed in 32 nm digital SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3049–3058, Dec 2013.

[83] V. Åberg, "Desing of 28 nm FD-SOI CMOS 800 MS/s SAR ADC for wireless applications," Master's thesis, 2016.

[84] G. Wang, K. Sun, Q. Zhang, S. Elahmadi, and P. Gui, "A 43.6-dB SNDR 1-GS/s single-channel SAR ADC using coarse and fine comparators with background comparator offset calibration," in *43rd IEEE European Solid State Circuits Conference (ESSCIRC)*, Sept 2017, pp. 175–178.

[85] Y. C. Lien, "A 4.5-mW 8-b 750-MS/s 2-b/step asynchronous subranged SAR ADC in 28-nm CMOS technology," in *Symposium on VLSI Circuits (VLSIC)*, June 2012, pp. 88–89.

[86] H. Wei, C. H. Chan, U. F. Chio, S. W. Sin, U. Seng-Pan, R. Martins, and F. Maloberti, "A 0.024mm^2 8b 400MS/s SAR ADC with 2b/cycle and resistive DAC in 65nm CMOS," in *IEEE International Solid-State Circuits Conference*, Feb 2011, pp. 188–190.

[87] H. K. Hong, W. Kim, H. W. Kang, S. J. Park, M. Choi, H. J. Park, and S. T. Ryu, "A decision-error-tolerant 45 nm CMOS 7b 1 GS/s nonbinary 2b/cycle SAR ADC," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 2, pp. 543–555, Feb 2015.

[88] L. Kull, T. Toifl, M. Schmatz, P. Francese, C. Menolfi, M. Braendli, M. Kossel, T. Morf, T. Andersen, and Y. Leblebici, "A 35mW 8b 8.8 GS/s SAR ADC with low-power capacitive reference buffers in 32nm digital SOI CMOS," in *Symposium on VLSI Circuits (VLSIC)*, June 2013, pp. C260–C261.

[89] L. Kull, T. Toifl, M. Schmatz, P. Francese, C. Menolfi, M. Braendli, M. Kossel, T. Morf, T. Andersen, and Y. Leblebici, "A 90GS/s 8b 667mW 64x interleaved SAR ADC in 32nm digital SOI CMOS," in *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2014, pp. 378–379.

[90] L. Kull, D. Luu, C. Menolfi, M. Braendli, P. A. Francese, T. Morf, M. Kossel, A. Cevrero, I. Ozkaya, and T. Toifl, "A 24-to-72 GS/s 8b time-interleaved SAR ADC with 2.0-to-3.3pJ/conversion and >30dB SNDR at Nyquist in 14nm CMOS FinFET," in *IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb 2018, pp. 358–360.

[91] A. Bekal, R. Joshi, M. Goswami, B. Singh, and A. Srivatsava, "An Improved Dynamic Latch Based Comparator for 8-Bit Asynchronous SAR ADC," in *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2015, pp. 178–182.

[92] H. Jeon, Y.-B. Kim, and M. Choi, "Offset voltage analysis of dynamic latched comparator," in *IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug 2011, pp. 1–4.

[93] J.-H. Tsai, H.-H. Wang, Y.-C. Yen, C.-M. Lai, Y.-J. Chen, P.-C. Huang, P.-H. Hsieh, H. Chen, and C.-C. Lee, "A 0.003 mm^2 10 b 240 MS/s 0.7 mW SAR ADC in 28 nm CMOS with digital error correction and correlated-reversed switching," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 6, pp. 1382–1398, June 2015.

[94] Y. Huang, H. Schleifer, and D. Killat, "Design and analysis of novel dynamic latched comparator with reduced kickback noise for high-speed ADCs," in *European Conference on Circuit Theory and Design (ECCTD)*, Sept 2013, pp. 1–4.

[95] R. Palani and R. Harjani, "A 220-MS/s 9-Bit 2X time-interleaved SAR ADC with a 133-fF input capacitance and a FOM of 37 fJ/conv in 65-nm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 11, pp. 1053–1057, Nov 2015.

[96] Y. Sinangil and A. Chandrakasan, "An embedded energy monitoring circuit for a 128kbit SRAM with body-biased sense-amplifiers," in *IEEE Asian Solid State Circuits Conference (A-SSCC)*, Nov 2012, pp. 69–72.

[97] J. Doernberg, H. . Lee, and D. A. Hodges, "Full-speed testing of A/D converters," *IEEE Journal of Solid-State Circuits*, vol. 19, no. 6, pp. 820–827, Dec 1984.