



UNIVERSITÀ DI PISA

Facoltà di Ingegneria

TESI DI LAUREA SPECIALISTICA
IN INGEGNERIA ELETTRONICA

Design of a process monitor and
peripheral circuits enabling the
characterisation of CMOS 45 nm
Ultra Low Power and Litho Friendly
optimised standard cells

Candidato:

Claudio Tagliabue

Relatori:

Prof. Giuseppe Iannaccone

Prof. Stefano Di Pascoli

Dr. Ing. Agnese Bargagli Stoffi

Anno Accademico 2006/2007

Acknowledgments

Se questo lavoro è giunto infine a compimento, il merito non è soltanto mio.

Prima di tutto voglio ringraziare Agnese. Non solo per l'inesimabile supporto tecnico e l'inesauribile disponibilità di cui ho potuto beneficiare in tutte le fasi della tesi. E' grazie a lei, infatti, se ho fatto miei dei principi che trascendono la semplice dimensione lavorativa: mai dare nulla per scontato, mai credere a qualcosa solo perché altri lo hanno detto, e, soprattutto, mai piegare la testa davanti alle difficoltà.

Grazie anche ad Harold, l'altra persona su cui, durante i miei sei mesi a NXP, ho sempre potuto contare. La sua gioia e il suo entusiasmo si sono dimostrati un baluardo anche nei momenti più difficili.

Grazie a Fabio e Salvatore: le infinite discussioni sull'elettronica, sul futuro, e non solo, mi hanno aiutato in tante scelte, ma soprattutto hanno reso il lavoro un divertimento, sempre.

Grazie a Claudio: anche se giunto solo alla fine, i suoi preziosi consigli in un momento critico hanno fatto sì che potessi completare un'importante parte del mio progetto.

Grazie a Paola: alle sue molte doti ha aggiunto anche quelle di un'infinita pazienza e di un altruismo davvero non comune.

Infine grazie a Davide, Giovanni, Giuseppe ed Angelo: non c'è il minimo dubbio che senza di loro la mia permanenza in Olanda non sarebbe potuta essere l'esperienza straordinaria che questa tesi mi ha permesso di vivere.

Claudio

Pisa - 22 Febbraio 2008

Quid terrarum iuvare novitas potest? Quid cognitio
urbium aut locorum? In irritum cedit ista iactatio.
Quaeris quare te fuga ista non adiuvet? Tecum fugis.
Onus animi deponendum est: non ante tibi ullus
placebit locus.

LUCIUS ANNAEUS SENECA, *Epistulae morales ad
Lucilium, Liber III, 28*

Contents

1	Introduction	1
1.1	Design scaling	3
1.1.1	Design for Manufacturability	4
1.1.2	Design for Lithography	5
1.2	Ultra Low Power	7
1.2.1	Subthreshold regime	8
1.3	Testchip	12
1.3.1	Target Performance Measurements	13
1.3.2	Main Core modules	15
1.3.3	Description of a C^2 -Block	17
1.3.4	Digital Core	22
1.3.5	Process Monitors	23
1.3.6	Test setup considerations	24
2	Multiplexer	25
2.1	Project requirements	26
2.2	Switch element	27
2.2.1	Pass Gate	28
2.2.2	3-State	29
2.3	Multiplexer architecture	30
2.3.1	Single array	32
2.3.2	Multi stage	34
2.3.3	Standard library multiplexer	36

2.3.4	Frequency Divider mux	38
2.4	Performance comparison	40
2.4.1	I/O bound function	40
2.4.2	Variability	42
2.4.3	Area occupation	43
2.4.4	Power consumption	44
2.4.5	Switch element choice	45
2.5	One mux per <i>C</i> -Block	45
2.6	Technology change	48
2.6.1	Performance alterations	48
2.7	Extracted parameters	50
3	Selector	52
3.1	Project requirements	53
3.1.1	First RingO gate: NAND	54
3.1.2	First RingO gate: NOR	55
3.1.3	Special modes	55
3.2	Selector structure	56
3.2.1	Address decoding	57
3.2.2	BSC	59
3.2.3	BE array	60
3.3	Selex	62
4	Digital Core	65
4.1	Low power techniques for CMOS logic	65
4.1.1	Power Switching	65
4.1.2	Standby Voltage Scaling	66
4.1.3	Dynamic Voltage Scaling (DVS)	66
4.2	Project Requirements	67
4.3	Core design	68
4.3.1	Core structure	68
4.3.2	Core functionality	71

4.4	Simulation and implementation	71
5	Monitor	73
5.1	Lithography aberrations	73
5.2	Lithography monitor	75
5.3	Sensing circuit	78
5.3.1	Circuit topology	79
5.3.2	Working principle	81
5.3.3	Input voltage generation	82
5.3.4	Layout realisation	83
5.4	Monitor block implementation	86
A	Digital Core Truth Tables	89
B	Verilog-AMS and Verilog-A	95
B.1	Verilog	95
B.1.1	Verilog-AMS	95
B.1.2	Verilog-A	96
B.2	Project applications	96
B.2.1	Hierarchy	97
B.2.2	Verification	97
	Bibliography	98

List of Figures

1.1	Stochastic distributions	2
1.2	Reduced spread distributions	3
1.3	Modular Design	6
1.4	Transfer Function (TF)	10
1.5	Symmetrical TF	11
1.6	Top-level of the testchip and its modules	13
1.7	Block diagram of the <i>C</i> -Block: RingO arrays and control logic	17
1.8	Block diagram of the <i>C</i> ² -Block	21
2.1	2x1 multiplexer	25
2.2	Share of the address decoder	27
2.3	Pass Gate element	28
2.4	3-State Buffer element	29
2.5	Output signal dynamic amplitude vs. load capacitance	31
2.6	$\frac{\max[V_{out}]}{V_{DD}}$ and $\frac{\min[V_{out}]}{V_{DD}}$ vs. load capacitance	32
2.7	Array of <i>N</i> <i>switch elements</i> (Pass Gates)	33
2.8	Multi stage hierarchical mux (3 stages)	34
2.9	Standard cell: 4x1 mux	37
2.10	Comparison between a Mux and an FD mux	39
2.11	I/O bound functions for the 3 solutions	41
2.12	Block diagram of the <i>C</i> -block	46
2.13	Realisation of the 2x4x4x8 FD mux, modifying the 4x4x8	47
2.14	I/O bound function for the TSMC Hybrid FD mux	49

3.1	7 stages, inverter based RingO	54
3.2	Selector block diagram	56
3.3	Part of the NAND plane: most significant group	58
3.4	Part of the NOR plane	58
3.5	Block Enable realizing solutions	61
3.6	Selex and RingO arrays	64
4.1	Combinatorial net	69
4.2	FF-Comb block	69
4.3	FF-Comb chain	70
5.1	OPC and SRAFs applied in the mask definition process	74
5.2	Embodiment of the monitor	75
5.3	Layout of the monitor	76
5.4	Possible configurations	79
5.5	Proposed architecture	80
5.6	Proposed architecture with switches	80
5.7	Transient behaviour	83
5.8	V_{in} generator	83
5.9	Proposed architecture with switches	84
5.10	Layout realisations	85
5.11	Transient behaviour in post layout simulations	86
5.12	Basic element of the <i>Monitor Block</i>	87
5.13	<i>Monitor Block</i> with its 5 pins: <i>EN</i> , <i>SI</i> , <i>SE</i> , <i>clk</i> and <i>SO</i>	88

List of Tables

1.1	Optimised inverter performance	11
1.2	Special modes for the selectors	20
1.3	I/O and Pins of the RingOs	23
2.1	Pass Gate logic functionality	28
2.2	3-State Buffer logic functionality	30
2.3	Number of transistors for a 4x1 Mux	37
2.4	Monte Carlo simulation results	43
2.5	Number of transistors per multiplexer	44
2.6	Power consumption for a single multiplexer path	44
2.7	Number of transistors per multiplexer	46
2.8	Inverter RingOs operating frequencies	48
2.9	Monte Carlo simulation results	50
2.10	Monte Carlo simulation results at $V_{DD} = 1.2$ V	50
2.11	Inverter RingOs operating frequencies	51
3.1	NAND logic functionality	54
3.2	NOR logic functionality	55
3.3	Special modes for the selector	56
3.4	BSC outputs	60
3.5	Number of gates and of transistors per BE array	62
4.1	Performance specs for mobile applications	68
4.2	Truth Table of path 9: OUT<9>	71

5.1	Nominal characteristics and DC values	77
A.1	Truth Table of path 1: OUT<1>	89
A.2	Truth Table of path 2: OUT<2>	89
A.3	Truth Table of path 3: OUT<3>	90
A.4	Truth Table of path 4: OUT<4>	90
A.5	Truth Table of path 5: OUT<5>	90
A.6	Truth Table of path 6: OUT<6>	91
A.7	Truth Table of path 7: OUT<7>	91
A.8	Truth Table of path 8: OUT<8>	91
A.9	Truth Table of path 9: OUT<9>	92
A.10	Truth Table of path 10: OUT<10>	92
A.11	Truth Table of path 11: OUT<11>	92
A.12	Truth Table of path 12: OUT<12>	93
A.13	Truth Table of path 13: OUT<13>	93
A.14	Truth Table of path 14: OUT<14>	93
A.15	Truth Table of path 15: OUT<15>	94
A.16	Truth Table of path 16: OUT<16>	94

Chapter 1

Introduction

The evolution of the CMOS technology finds has been lately characterised by the scaling of transistor size and by the reduction of their power dissipation. Transistor scaling has always to be sought in regard of the integrated circuit robustness. The reduction of the supply voltage is a typical effect of this evolution: the smaller are the transistors, the lower is the supply voltage allowed across them. Obviously, a remarkable decrease of the power consumption can be achieved by lowering the supply voltage. In the last technology nodes the speed of the scaling process is decreasing, since the complexity of the technology increases with its size reduction, leading to two classes of difficulties:

- operational environment issues: decrease of the noise margins, and therefore robustness, due to the lowering of the supply voltages and signal ranges;
- technology related issues: reduction of the lithography accuracy, since the wavelength of the lasers used for the photo-lithographic process is no longer much smaller than the smallest device dimension.

One of the main aspect of these issues is the variability of the fabrication process. It is predictable that the value of all geometrical and electrical parameters will have a stochastic distribution: typically a Gaussian or a

logNorm distribution, depending from the feature taken into account (Figure 1.1).

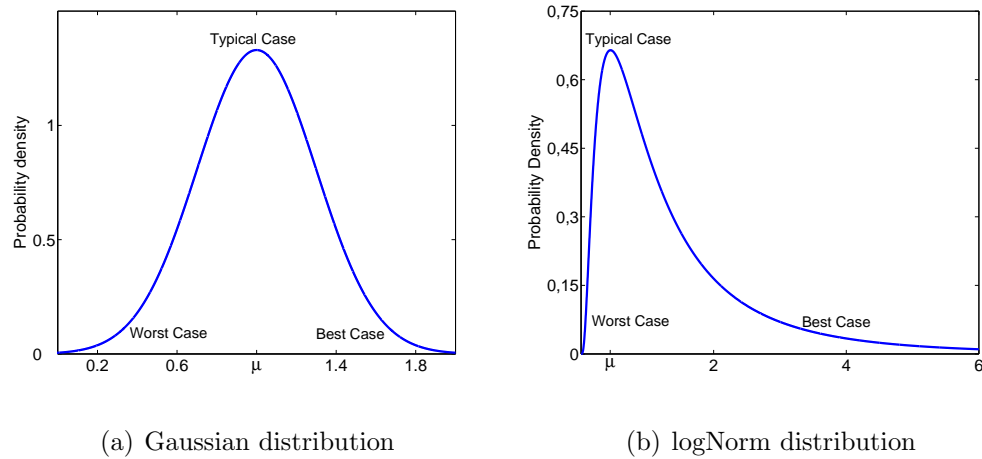


Figure 1.1: Stochastic distributions

Variability is measured as the difference between expected and actual performance. It can be attributed to design causes (model inaccuracy, design errors, parasitic elements), environmental causes (temperature variations, noise) or physical causes (variations in the manufacturing process). The effects of these factors are worse as the technological node gets smaller. The field of investigation of this work is focused on the physical causes leading to variability: it aims to solutions that could be implemented, alongside the standard libraries, to attain better device performance.

A common approach in digital circuit design consists in dimensioning for the expected worst case, so that, statistically speaking, the requested features are achieved in all cases. This method is called Corner Design or Worst Case Design. Its benefits are easily understood, but the over-design requested to follow it leads to a loss of performance and to a wider silicon area utilisation. Moreover it is unmistakable that further benefits may be obtained by decreasing this variability, thus reducing the spread of the distribution (Figure 1.2):

1. the probability to attain Typical Case features increases;

2. the Worst Case and Best Case features are closer to the Typical Case ones.

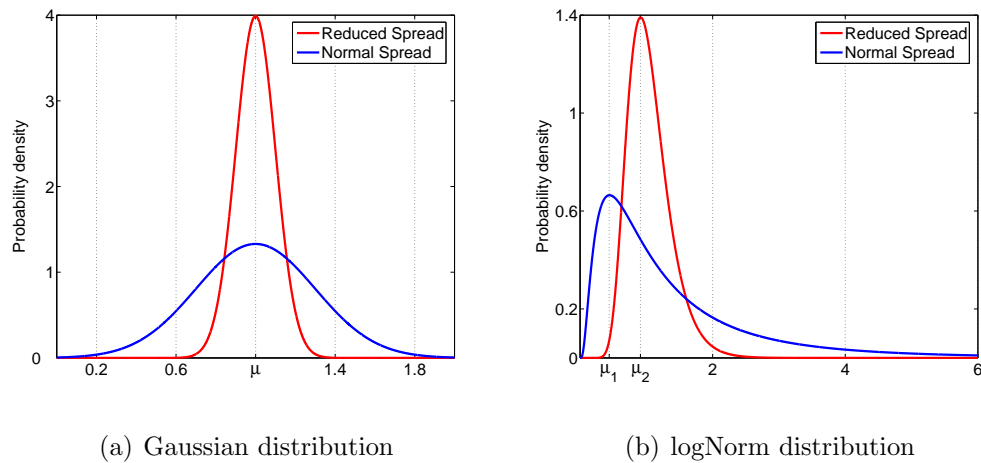


Figure 1.2: Reduced spread distributions

The target of this project is indeed to reduce the effects of the variability of the realisation process in a CMOS 45 nm technology node in digital circuits performances, using unconventional design methods.

1.1 Design scaling

The conventional approach for the re-design of a circuit in a new technology node consisted in its mere scaling from the previous node. Since the reduction of the lithography wavelength has lately not corresponded to the scaling of the device dimensions, the precise control of the dimensions of the litho process is no more achievable with the same techniques used in the past years.

Among the Resolution Enhancements Techniques (RET) investigated in the last years, the most common are:

- the OPC: Optical Proximity Correction;
- the SRAFs: the Sub Resolution Assist Features.

A brief description of OPC and SRAFs is given in Section 5.1. These techniques increase the complexity of mask design.

Dimensional control even at smaller dimensions could be reached using also alternative design approaches. The most commonly known are summarised under the acronyms DfM (*Design for Manufacturability*) and DfL (*Design for Lithography*).

1.1.1 Design for Manufacturability

For the time being, it is necessary to rely on innovations that extend the use of photolithography beyond the 45 nm node. Therefore, support from the design side might alleviate some of the expected problems when extending the use of 193 nm lithography into the sub 50 nm CMOS technologies. To improve the yield, thus, complex *Design for Manufacturability* design rules have already been used in most advanced technology nodes.

DfM includes a set of techniques to modify the design of ICs in order to make them more manufacturable, i.e. to improve their functional yield, parametric yield, and their reliability.

DfM consists also of a set of different methodologies trying to enforce some soft (recommended) design rules regarding the shapes and polygons of the physical layout of an integrated circuit. These DfM methodologies work primarily at full chip level. Additionally, worst-case simulations at different levels of abstraction are applied to minimize the impact of process variations on performance and other types of parametric yield loss.

To make the design as robust as possible to yield loss causes, some DfM techniques are:

- substitute higher yield cells where permitted by timing, power, and routability;
- increase the spacing and width of interconnect wires, where possible;
- optimize the amount of redundancy in internal memories;

- insert redundant vias in the design where possible.

These operations require a detailed understanding of yield loss mechanisms, since these changes trade off against one another. For example, introducing redundant vias reduces the chance of via problems, but increases the chance of unwanted shorts. The advantages and drawbacks, therefore, depend on the details of the yield loss models and the characteristics of the particular design.

More information about DfM methodologies can be found in [1], [2], [3], [4].

1.1.2 Design for Lithography

For the 45 nm node, however, DfM methodologies may be not enough to improve yield. DfL, also called lithofriendly design, litho-driven design or litho-centric DfM, is focused on more regular layout structures.

Lithofriendly layouts try to reduce variability by relaxing the minimum poly gate pitch, by minimizing the range of pitches present in the layout and by adding dummy poly lines. Even if the poly interconnect lines are allowed in the two orthogonal directions, often the horizontal lines are drawn wider, making them non critical for printing. This is actually possible because only the poly gate lines have significant influence for the variability of the circuit performances, while the poly interconnect may be neglected. The dummy poly lines are added to increase layout regularity: they have to be placed adjacent the poly gate lines, and their width should be the same.

These techniques give some main advantages that result in a better contrast (i.e. printability) of small features, and therefore improve production yield:

- illumination used in the lithographic process can be optimised for the chosen pitch and/or chosen orientation;
- optimum lens properties can be chosen (Numerical Aperture of the lens);

- it is easier to place assist features (SRAFs) and to apply post layout corrections (OPC).

It has been demonstrated [5] that these expedients allow to avoid dimensional variations due to decreased laser resolution and to phase conflicts (when phase shifting masks are used).

These litho-driven considerations lead to the conclusion that a modular and regular layout, with relaxed minimum distances, can considerably reduce the performance variability of an IC (Figure 1.3).

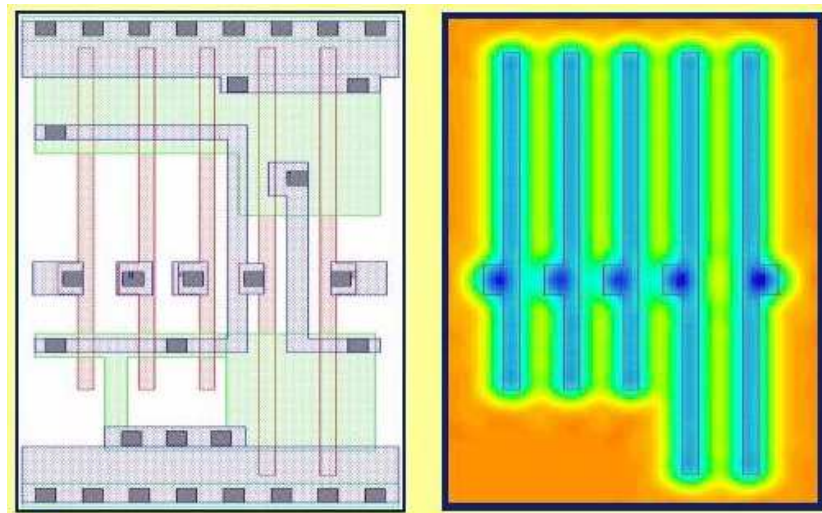


Figure 1.3: Modular Design

A possible drawback of this layout approach is that the circuit area could increase. For complex logic gates (Full adder and Flip-Flop) this increment is in a 5-11% order [5]. It may be proved that, however, in the realisation of a complete chip, there is no (or a small) area penalty paid for using lithofriendly design, since the most frequently used logic gates do not require a significant increase of area. In this project the litho-driven layout will aim to have approximately the same area of the conventional layout.

DfL, thus, simplifies the lithographic process, it supports SRAFs and OPC, and may reduce the mask costs. It may also lead to a more aggressive scaling and to yield improvement, due to a smaller set of patterns to be

printed. Moreover, more regularity in the standard cells may also lead to a better portability to the next technology nodes.

1.2 Ultra Low Power

Considering the rapid growth of the portable applications' market, less consuming circuits are a specific research target in the electronic design. Ultra Low Power (ULP) electronics is a new product area, rising alongside the low power, but characterised by even stronger power requirements.

Due to the great benefits deriving from a more parsimonious power consumption, ULP circuits are wide spread applications:

- handheld devices;
- medical applications (monitoring systems, medical instrumentation, implantable devices);
- wireless network systems;
- smart cards;
- RFIDs.

In Ultra Low Power circuits, where $V_{DD} < V_{Tn} + |V_{Tp}|$, all devices work necessarily in the subthreshold regime. A brief description of the subthreshold transistor models is given in the next paragraph.

As mentioned before, the easiest way to reduce energy consumption is the lowering of the supply voltages. The usually given expression of the average energy dissipated by a CMOS gate per clock period is:

$$E = E_{dynamic} + E_{leakage} = \alpha CV_{DD}^2 + I_{off}V_{DD}T$$

Where α is the *activity factor*, statistically determined; C is the load capacitance, I_{off} is the leakage current; $1/T$ is the operating frequency and V_{DD} is the supply voltage.

It is clear how the value of the supply voltage V_{DD} is directly responsible (both quadratically and linearly) for the energy consumption. Nonetheless, it must be taken into account that a design optimised only for energy dissipation would probably lose other fundamental features, such as speed and robustness. For this reason, ULP design should always take into account, in addition to the supply voltage, the circuit's architecture and the predicted activity factor and throughput.

Stricter constraints on the supply voltages leads to unusual trade-offs with other circuit's features, such as the working frequency, the sensitivity to environmental factors and even the circuit's area.

Usually, in ULP, Worst Case design leads to an unacceptable area overhead. Thus, to obtain more realistic results, Monte Carlo analysis are performed. Moreover, Monte Carlo analysis give information about the spread, thus consenting a design aware of the variability of the parameters.

Since the subthreshold logic is extremely sensitive to parameter variations, particular effort must be spent to obtain regular circuit design and layout. For these initial considerations, the link between the lithofriendly and ULP circuit optimisation is therefore evident.

Among the intents of this project, there is the creation of an Ultra Low Power design strategy, giving better results in performance variability than the conventional approach.

1.2.1 Subthreshold regime

Weak inversion region, also known as subthreshold regime, is defined as the saturation region of a transistor whose V_{GS} does not exceed the threshold voltage V_T [6].

Below the threshold voltage, the current of the MOS transistors has an exponential dependence on V_{GS} .

Since in subthreshold region V_{GS} is less than V_T , the mobile charge Q_m is zero, while the depletion charge Q_D is larger than in strong inversion region. A small current still flows at the surface underneath the gate of the MOS

transistor, but flows even in the depletion layer [7].

The minimum Drain-Source voltage V_{DS} needed to operate in inversion is called V_{DSsat} . In strong inversion it is about $V_{DSsat} \approx V_{GS} - V_T$; while in the subthreshold regime it is about $3U_T$. Thus, to reach the saturation for a MOS transistor in weak inversion is enough to have a V_{DS} approximately three times the thermal voltage: $3U_T = 3\frac{kT}{q}$. The driving current in subthreshold regime is then given by

$$I_{DS} = \frac{W}{L} I_{D0} e^{\frac{V_{GS}-V_T}{\eta U_T}} \left(1 - e^{-\frac{V_{DS}}{U_T}} \right) \quad (1.1)$$

Where η is called *subthreshold slope factor*. And I_{D0} is:

$$\begin{aligned} I_{D0n} &= \mu_n C_{ox} U_T^2 & nMOSFETs \\ I_{D0p} &= -\mu_p C_{ox} U_T^2 & pMOSFETs \end{aligned} \quad (1.2)$$

A direct consequence of this exponential behaviour is the value of the transconductance g_m in weak inversion, obtained by taking the derivative of I_{DS} versus V_{GS} :

$$g_m^{wi} = \frac{W}{L} \frac{I_{D0}}{\eta U_T} e^{\frac{V_{GS}-V_T}{\eta U_T}} \left(1 - e^{-\frac{V_{DS}}{U_T}} \right) = \frac{I_{DS}}{\eta U_T} \quad (1.3)$$

It is illustrated how in this regime the transconductance is directly proportional to the current.

More significant to understand the MOS transistor transfer efficiency from input to output is the transconductance to current ratio $\frac{g_m}{I_{DS}}$, that in strong inversion is

$$\frac{g_m^{si}}{I_{DS}} = \frac{2}{V_{GS} - V_T} \quad (1.4)$$

whilst in weak inversion it assumes the value

$$\frac{g_m^{wi}}{I_{DS}} = \frac{1}{\eta U_T} \quad (1.5)$$

which is independent of the current. Moreover this is the highest value that can be achieved. Therefore, for circuits requiring high gain and that

may operate with small currents and low operating frequencies, this region is preferred.

The robustness of a digital gate can be pointed out by the slope of its transfer function, since it determines the noise margins and the capability to regenerate a noisy input signal into a full-dynamic output signal.

In addition, it must be considered that, due to the increased sensitivity to parameter variations in weak inversion, it is more difficult to have a symmetric transfer function in a CMOS inverter (Figure 1.4(a)).

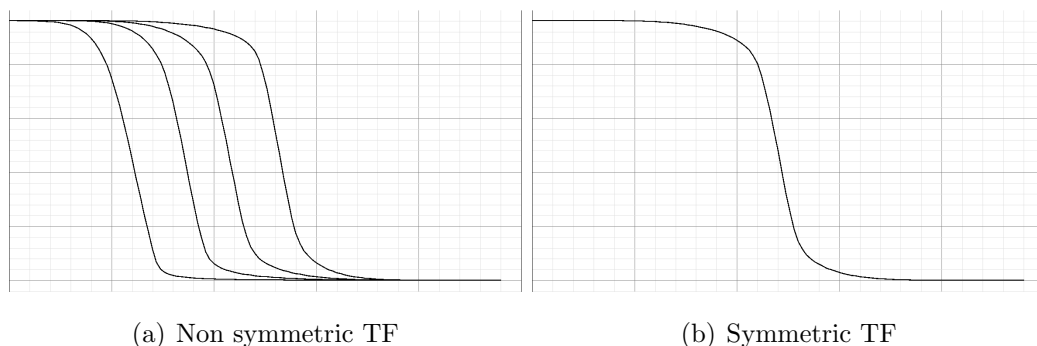


Figure 1.4: Transfer Function (TF)

It is however well known that good noise margin are obtained with a symmetrical transfer function (Figure 1.4(b)).

In a digital inverter the cross-over point is defined as the input voltage that should be applied to obtain an output voltage equal $\frac{V_{DD}}{2}$. At the cross-over point $I_{DS_n} = I_{DS_p}$, and $V_{DS_n} = |V_{DS_p}| = \frac{V_{DD}}{2}$. To obtain $I_{DS_n} = I_{DS_p}$, for an inverter operating in subthreshold regime, the $\frac{W_n}{W_p}$ ratio is typically different than in strong inversion.

Figure 1.5 shows the result of a simulation in CMOS 90nm technology node, displaying that in subthreshold region a symmetrical transfer function is given by $\frac{W_n}{W_p} \approx 1$.

For this reason, a definition of new libraries for the subthreshold region operation is needed. Table 1.1 shows a comparison between a standard inverter in strong inversion region, the same inverter in subthreshold regime

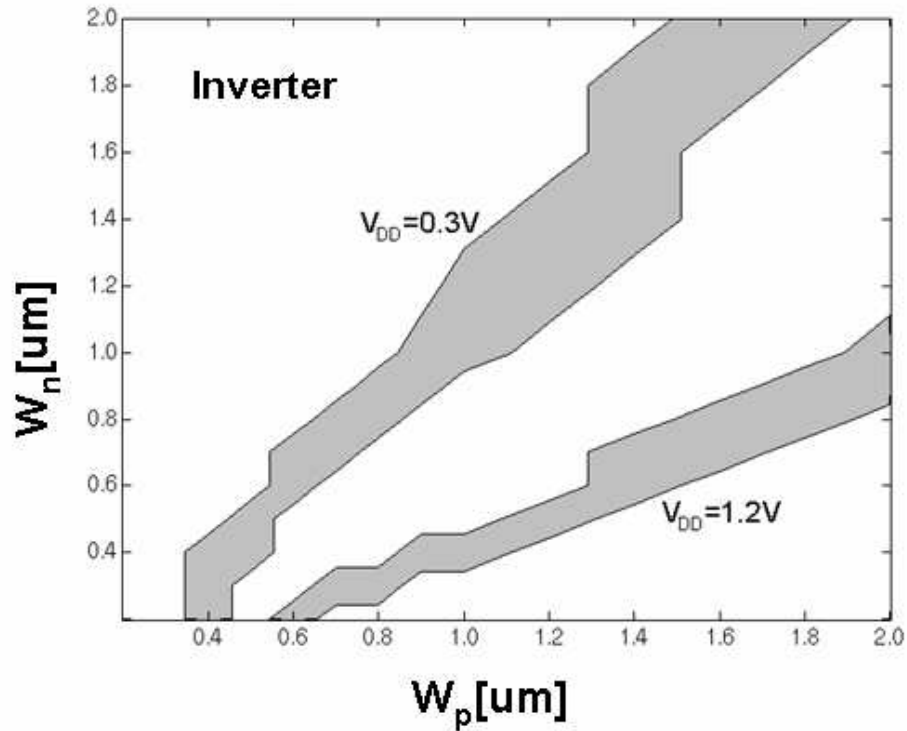


Figure 1.5: Symmetrical TF

and two different types of inverter optimised for the subthreshold operating region. The data refer to the CMOS 45 nm technology node.

	Std. CMOS		Optimised library	
	1.1 V		0.3 V	
	$W_p = 215nm$		$W_p = 120nm$	$W_p = 200nm$
			$W_n = 120nm$	$W_n = 120nm$
Max Freq (f_{max}) [MHz]	11.7e3	27.3	60.6 (+120 %)	58.9 (+115 %)
Power @ f_{max} [nW]	125e3	20.3	17.4 (-14 %)	21.57 (+6.2 %)
Switching Energy [$\frac{nW}{MHz}$]	10.6	0.74	0.28 (-61 %)	0.36 (-50 %)

Table 1.1: Optimised inverter performance

It is possible to observe that a standard cell used in subthreshold region

has a huge performance decrease. On the other side, better results are obtained, in the same operating region, with a standard-like inverter having $W_n = W_p$.

More information on ULP design strategies can be found in [8], [9], [10], [11].

1.3 Testchip

This project aims to design and to realize a testchip to investigate and to quantify the improvement of the circuit performances obtained through the design of dedicated litho-friendly (LF) and of the Ultra Low Power (ULP) standard-like libraries. The LF standard cell libraries are optimised for lithography using ultra regular layout styles. The ULP standard cell library is optimised to operate at extremely low supply voltage.

The main objective of the testchip is to get insight into the local and the global variability of relevant parameters for digital design, such as operating frequency and power consumption. In this testchip some structures are also included, to develop some innovative circuits that should help to monitor the quality of the technology process. The testchip is realised in a CMOS 45 nm process.

The planned testchip is made up the following blocks:

- Main Cores: one for each of the five designed libraries, plus a rotated version of a lithofriendly library. Each core contains a combinatorial logic block to measure the statistical parameters of the circuit's performances.
- Digital Core: a small digital core, where combinatorial and sequential logic are implemented together, to verify the circuit behaviour at extreme low voltages.
- Process Monitors: to verify the quality of the process and of the impact of a lithofriendly design approach on the fabrication process.

A representation of the entire testchip is reported in Figure 1.6.

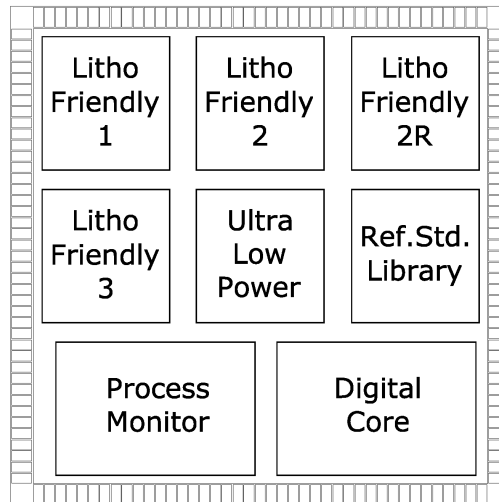


Figure 1.6: Top-level of the testchip and its modules

1.3.1 Target Performance Measurements

The aim of the circuit is the qualification of the realised standard cell libraries. The quantities that have to be measured in these structures are the active power P_{on} , the standby-power P_{off} of the circuits, the maximum operating frequency f , and the dependence of the active power on the operating frequency and the circuit activity. This testchip is designed to gain a strong insight about the robustness of digital circuits in nano metric devices. The mean (μ) and the standard deviation (σ) of these performance indicators are good measurements of circuit sensitivity to local and global variability.

To perform variability measurements, the basic structures realised on the testchip are ring oscillators (see Section 1.3.3). It is well known [5] that variability effects are mostly perceptible in the delay of the affected cell. Therefore, ring oscillators are used to attain statistical information on the delay of the cells composing them from frequency measurements. The operating frequency of a ring oscillator is given by:

$$f = \frac{1}{T} = \frac{1}{2 \sum_{i=1}^N \tau_{d_i}}$$

Where T is the ring oscillator period, N is the number of cells composing the ring oscillator, and τ_{d_i} is delay of the i -th cell, supposed to be approximately the same for the raise and fall commutation.

Given this relation, the mean value and the standard deviation of the ring oscillator period will be:

$$\langle T \rangle = \left\langle 2 \sum_{i=1}^N \tau_{d_i} \right\rangle = 2 \sum_{i=1}^N \langle \tau_{d_i} \rangle = 2N \langle \tau_d \rangle \quad (1.6)$$

$$\sigma_T^2 = 2N \cdot \sigma_{\tau_d}^2 \quad \text{and} \quad \sigma_T = \sqrt{2N} \cdot \sigma_{\tau_d} \quad (1.7)$$

For small relative variations, even if the relation between time and frequency is not linear, we have:

$$\frac{\Delta f}{f} \approx \frac{\Delta T}{T}$$

Therefore:

$$\frac{\sigma_f}{\langle f \rangle} \approx \frac{\sigma_T}{\langle T \rangle} \quad (1.8)$$

From equations 1.6, 1.7 and 1.8 is then possible to find the relation between the relative standard deviation of the measured frequency and of the delay time:

$$\frac{\sigma_{\tau_d}}{\langle \tau_d \rangle} = \frac{\sigma_T}{\sqrt{2N}} \cdot \frac{2N}{\langle T \rangle} = \sqrt{2N} \cdot \frac{\sigma_T}{\langle T \rangle} \approx \sqrt{2N} \cdot \frac{\sigma_f}{\langle f \rangle} \quad (1.9)$$

The active power P_{on} depends on f . Both active and standby power P_{on} and P_{off} are a function of supply voltage V_{DD} , back-bias voltages V_{bbp} and V_{bbn} , and temperature T . Therefore, the following measurement are required:

- P_{on} as a function of f , V_{DD} , V_{bbp} , V_{bbn} , and T ;

- P_{off} as a function of V_{DD} , V_{bbp} , V_{bbn} , and T .

In the design of the circuit, the following requirements must be fulfilled:

- Independent power supply connections for the different modules;
- independent back-bias connections (pwell and nwell);
- controllable activity α of the digital block.

Statistical information on local variability, measurements of many identical delay paths laid out at close distance are necessary. Statistical information on global variability is obtained by measuring different dies or samples.

To gain further insight into lithography properties, at least one of the test cores should be placed twice with different orientations; therefore one of the instances rotated of 90 degrees.

1.3.2 Main Core modules

In the testchip there are six digital modules which perform the same functionality. They are realised with different standard cell libraries, optimised for different goals.

The aims of these digital cores are first of all to prove that the design methods used to implement the digital libraries are efficient, and then to quantify the gain in performance obtained with these optimised digital libraries.

The modules are:

- **Reference Std. Library Module (REF)**, implemented with the cells from the reference standard library currently available in the digital flow. This module is implemented in the testchip to have a reference for the other modules in terms of power and frequency performance.
- **3 Lithofriendly Library Modules (LF<1:3>)**, developed with three different layout approaches. The main aim of these libraries is to improve the lithofriendliness of the design and therefore to reduce the

spread of the performance and to improve the yield of the digital circuits (see Section 1.1.2).

1. **LF1**: lithofriendliness is limited to the active areas: all the transistors are drawn with the same width and length, thus the standard-like cells are higher than the standard one. The area overhead is the most significant.
 2. **LF2**: lithofriendliness is extended to active area, poly and contacts. There is a small area overhead, and the metal layers are not designed for lithography.
 3. **LF3**: lithofriendliness is extended to all layers. The performances are in this way decreased, but the reduced spread obtained in this way is expected to compensate the performance loss.
- **Rotated Lithofriendly Library Module** is a rotated copy of **LF2**, to verify the effect of orientation on the performances of the circuit.
 - **Ultra Low Power Module (ULP)** makes use of a digital library optimised to operate at low voltage supplies (see Section 1.2).

The design of a large digital library is unfeasible with the limited available resources and it is not strictly necessary since the first goal of this project is the verification of the design concept. In this experiment the main design focus is limited to few combinatorial logic gates and to Flip-Flops. Although the limited number of cells, the measurement results can give a good insight of the trend of the performances of the digital logic that could be developed, if one of these libraries would be adopted in a real digital design. The key performances that are going to be observed in this experiment are maximum operating frequency, the power consumption and the robustness of these performances to process variations and device mismatch.

Each digital module is constituted by a C^2 -block: with this block the performances of the combinatorial logic are analysed. The C^2 -block is made

up by four C -Blocks. The basic structures of these combinatorial blocks are ring oscillators.

1.3.3 Description of a C^2 -Block

The C^2 Block is the part of the core module where the combinatorial logic is proven. It is made up by four C -blocks, each of them consisting in two arrays of 128 ring oscillators (RingOs). The C -Blocks differ for the logic gates with which the RingOs are realised, i.e. IVX, NAND, NOR or MXD.

Beside the arrays of ring oscillator, each C -Block has also a selector and a multiplexer to complete its functionality. The selector enables the oscillation of the required RingO(s), while the multiplexer selects which output node has to be available at the output pad. In Figure 1.7, the high level schematic of a C -Block is reported.

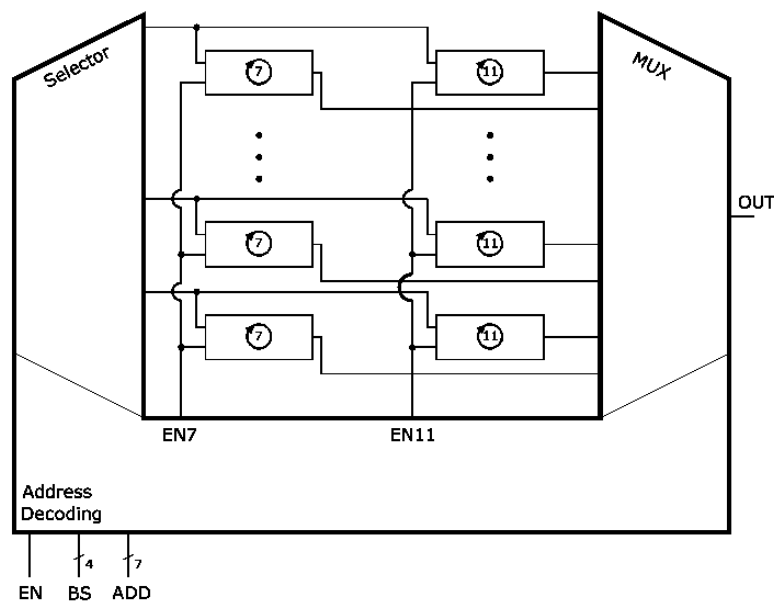


Figure 1.7: Block diagram of the C -Block: RingO arrays and control logic

In the following of this Section, a short introduction to the dimensioning of each part of this structure is given.

- **Ring Oscillators:** The basic structures of these combinatorial blocks are ring oscillators. They are developed with Inverter, with Nand, with Nor or with a mixture of these cells. To compare consistent data, in case of two inputs gates, the signal is sometime associated to the node closer to the output node. The logic depth of the ring oscillators, i.e. the number of combinatorial gates that are cascaded, has been determined as a trade off between two opposite conditions:
 - On one hand, the logic depth of the RingO should be minimised, since one of the main aims of this experiment is to verify if the optimised libraries reduce the performance (i.e. gate delay and thus RingO operation frequency) spread. As the number of the cascaded elements within the ring oscillator increases, the performances of the single gates are averaged out.
 - On the other hand, the oscillating frequency of a ring oscillator increases decreasing the number of cascaded gates. Since the maximum acceptable frequency is limited by the maximum speed of the frequency divider that is connected between the RingO and the measurement equipment, depending on the speed of the single gate a lower border for the logic depth of the RingOs is found.

For the given technology and the designed Flip-Flop, RingOs with a 7-logic depth satisfy the requirements above. Since this technology node is still not mature, RingOs with 11-logic depth are also realised to ensure the functionality of the circuit also if the frequency of the circuit is above the expected corner situations. To attain a relevant number of measurements, and therefore to be able to derive some statistical information about the spread, the number of RingOs must be sufficiently high. In this experiment 128 RingOs of each type are designed.

- **Control logic:** it is made up by a selector that decodes the addresses for the activation of the corresponding RingOs (see Chapter 3), and by a multiplexer that routes the output of the selected RingOs to the

output of the circuit (see Chapter 2). The inputs of the control logic are:

- 7-bit input Address ($ADD < 6 : 0 >$), which encode the selector line that has to be active.
- the enable signal (EN) that identifies which array has to be addressed;
- 4-bit input Block Select modes ($BS < 3 : 0 >$), active low, which functionality is reported in Table 1.2;
- Disable signal (DIS): that turns off the entire block.

Since the multiplexer must report to the output the signal of the ring oscillator enabled by the selector, the selector and the multiplexer may share the same address decoding logic. The output of the multiplexer presents a high frequency signal. The output signal is therefore sent to a frequency divider and afterwards to the output pad.

In the normal state, the selector enables only one of the 2x128 RingOs to oscillate, while the others are in the disable condition. Modifying the value of the address bits, all RingOs can be activated one after the other and the operating frequency of each of them can be measured at the output node of the multiplexer.

The selector presents also some special modes, which are coded through the Block Selector (BS) inputs. The aim of these special modes is to enable the simultaneous oscillation of more RingOs. When the circuit operates in these modes, the main goal is to measure the power consumption as a function of the circuit activity, while there is no interest to observe the output voltage of the multiplexer.

In Table 1.2 the special modes of the selector and the value of the control inputs are reported. Since the EN signals may never be active together, i.e. the 7-depth RingOs array can not be activated together

with the 11-depth RingOs array, the selector functionality is illustrated with just one EN signal.

Mode (Activity)	Control bits				
	DIS	BS3	BS2	BS1	BS0
Normal	0	1	1	1	1
25%	0	1	1	1	0
50%	0	1	1	0	0
75%	0	1	0	0	0
100%	0	0	0	0	0
0%	1	X	X	X	X

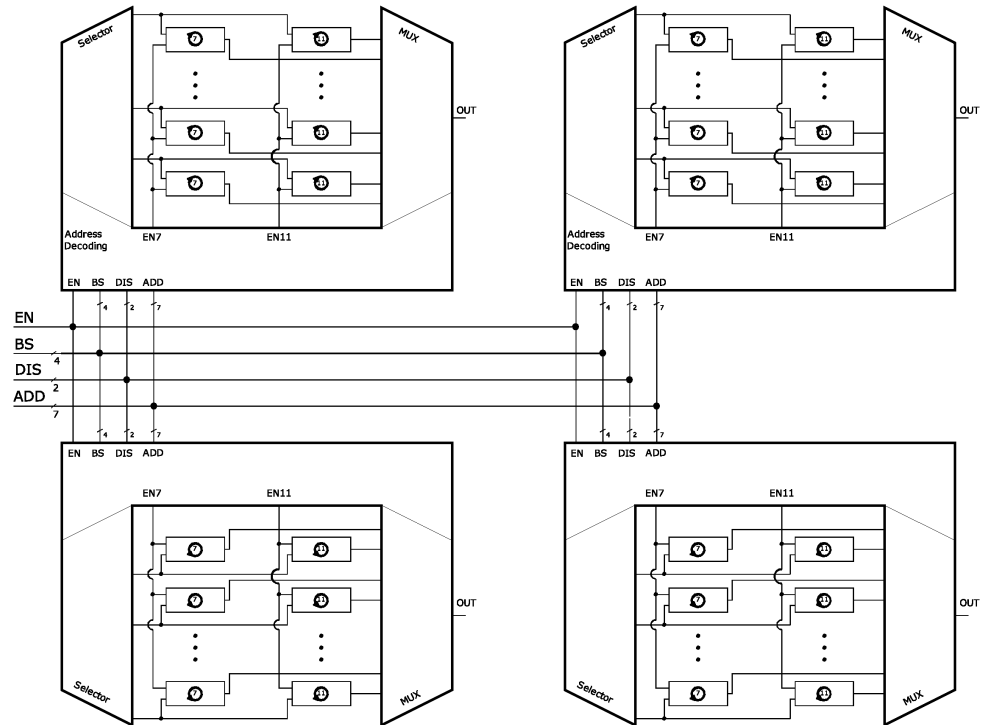
Table 1.2: Special modes for the selectors

- **Scan chain:** it is a structure of shift registers with serial input and serial output, whose information is loaded from an output pad during the initialization phase of the testing, and determines part of the test vectors. In this test structure, the information that can be stored in the scan chain is the value of the address, the value of the BS and the value of the DIS signal, i.e. 15 bits per module.

A C^2 -Block is obtained joining four C -Blocks (Figure 1.8). During the measurement at most one of the four C -Blocks must be active. To save some silicon area, the C -Blocks of each digital module may share the address lines, the BS and the EN signals, while the DIS signals differ for each C -Block.

The routed signals in the C^2 -block are therefore:

- **ADD<9-0>:** 7 Address + 3 coded DIS = 10 bits Address;
- **BS<3-0>:** Block select modes (active low);
- **EN:** Enable signal.

Figure 1.8: Block diagram of the C^2 -Block

I/O and Pins of the ring oscillators module

The I/O signals of each module are:

- V_{DD} : supply voltage for RingOs
- V_{DDS} : pMOSFET bulk connection for RingOs
- V_{SS} : ground for RingOs
- V_{SSS} : nMOSFET bulk connection for RingOs
- V_{DDH} : supply voltage for control logic
- V_{SSH} : ground for control logic
- OUT : the output of the multiplexer

- *SI*: Input of the scan chain
- *SO*: Output of the scan chain
- *CLKs*: Clock signal of the scan chain

Each main core needs separate supplies voltage pins to ensure a complete independence of the blocks during the power consumption measurements. Each module must present also a separate OUT pin.

There are three choices for the scan chain:

1. a safe choice is to reduce the length of the scan chain, so that each module has its own, that would require 3 pins, for a total of 18 pins;
2. 5 pins could be saved sharing the same clock signal for all the scan chains (*CLKs*);
3. if the number of available pins is limited the scan chain of all modules is just one, and the only 3 pins are necessary for the six modules. This solution has as drawback the fact that the scan chain must connect block that may be placed at a relative quite large distance in the chip.

In Table 1.3 the number of I/Os and of PINs for each module and the total for all six modules is reported. Between bracket are the value needed to realise alternative solutions that ensure the same logical functionality and the same measurement capability, but which are more complex in the layout phase, and therefore may introduce more risks in the realisation.

1.3.4 Digital Core

In Section 1.2.1 the basic principles of the subthreshold regime are given. Moreover, it is discussed how an optimised design strategy is necessary to improve the performance of circuits operating in subthreshold region.

Table 1.1 points out how, for the inverter performance, standard cells can be optimised to give better results in subthreshold regime.

	I/Os	PINs
V_{DD}	1	6
V_{DDS}	1	6
V_{SS}	1	6
V_{SSS}	1	6
V_{DDH}	1	6
V_{SSH}	1	6
SI	1	6(1)
SO	1	6(1)
CLKs	1	6(1)
OUT	1	6
TOTAL	10	54(49-39)

Table 1.3: I/O and Pins of the RingOs

However, since power consumption in handheld systems is one of the most significant constraints, all strategies leading to energy saving are explored. It is an emerging idea to make these systems work at low frequencies in subthreshold regime when they are in idle, lowering their V_{DD} .

In Chapter 4 the digital core design is discussed. In the digital core realised on the testchip, standard combinatorial and sequential logic are implemented together. The aim of this core is to verify how a circuit designed to work in strong inversion region behaves at extreme low voltages.

1.3.5 Process Monitors

In Section 1.1 the main problems of the design scaling are discussed. Moreover, in Section 1.1.2 the lithofriendly approach is introduced. For the new technology nodes, such as the CMOS 45nm used in this project, the actual advantages and drawbacks of a lithofriendly design approach are still to be quantified.

Therefore, process monitors are realised on the testchip in order to find

out the actual reliability and robustness of the process itself, and to point out the necessity to adopt litho-driven design methodologies. The task of the monitor is to verify the presence of a systematical error introduced by pattern aberrations.

In Chapter 5 the design of analog process monitor is discussed.

1.3.6 Test setup considerations

Considerations about the available test facility:

- Maximum 8 power supplies available at the same time. Two are required for the pad ring, so 6 are left for the cores on the test chip.
- Maximum realistic input signal frequency = 500 MHz.
- Maximum realistic output signal frequency = 100 MHz.
- PGA package with up to 256 pins is preferred.

Chapter 2

Multiplexer

A multiplexer, or mux, is a circuit used to select one out of many analog or digital data sources and to output that source into a single channel. This process is called multiplexing.

A multiplexer is an ideal multi-input, single-output switch. A signal called *selector* specifies which one of the multiple inputs has to be forwarded to the output.

A multiplexer with N inputs needs M *selector* bits, where $2^M \geq N$.

In Figure 2.1 a 2 inputs, 1 output multiplexer (2x1 mux) is shown. Thus, in this case, $N = 2$ and $M = \log_2(N) = 1$.

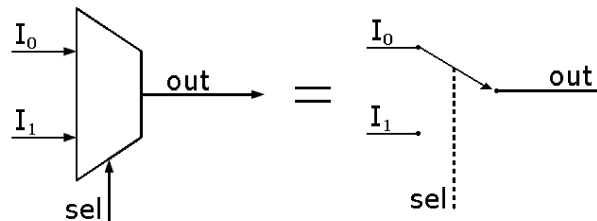


Figure 2.1: 2x1 multiplexer

In this Chapter the design of a multiplexer is discussed. In Section 2.1 the project requirements for this work are analysed. In Section 2.2 an overview of the basic mux core cells is given. In Section 2.3 different multiplexer architectures are analysed, and the choice of an architecture fulfilling the project requirements is discussed. Then, a performance comparison between multi-

plexers based on different core cells is given in Section 2.4, using simulations results. Section 2.5 deals with a modification of the project specs, adopted to reduce the area occupation of a mux, hence not decreasing its performances. In Section 2.6 the performance modifications given by a technology change are presented. Section 2.7 shows the effects of the parasitic parameters on the multiplexer performance.

2.1 Project requirements

In each C -block two arrays of 128 ring oscillators are present: one made up by 7 logic depth RingOs and the other by 11-logic-depth RingOs. The oscillation frequency for a 7-logic-depth RingO is about 5 GHz in the typical case, and can range from about 1.4 GHz to 10 GHz in the corner cases. The frequency of the 11-logic-depth RingOs is lower and thus non critical.

In this project, a 128x1 multiplexer is required to select one out of 128 ring oscillator outputs for each array. The main aim of the multiplexer is to forward the selected signal without frequency distortions. The capability to achieve this target in different conditions is called robustness. The realisation of a robust multiplexer in the mentioned frequency range and for the adopted supply voltage (1.1 V) is vital for this project. It has to be noticed that the selection signals are at low frequency, therefore the critical signals for the robustness are the mux inputs only. Alongside the robustness, area occupancy and power consumption are considered among the multiplexer performances, and their reduction has to be achieved.

Since $N = 128$, $M = \log_2(128) = 7$ selection bits are needed. These bits are decoded in order to obtain 128 mutually exclusive signals: when one of them is active the remaining 127 must be inactive. Since the 7 and the 11 logic depth RingOs are never active at the same time, the multiplexers for the 2 arrays have the same structure, and share the same selection bits. For this reason they will be hence no more considered separately.

Moreover, the 7 selection bits are the same arriving to the selector block. Therefore the decoding logic between the two blocks may be shared, routing

the 128 mutually exclusive signals from the selection block to the multiplexer (Figure 2.2). This solution is justified by the proximity of the selector to the multiplexer.

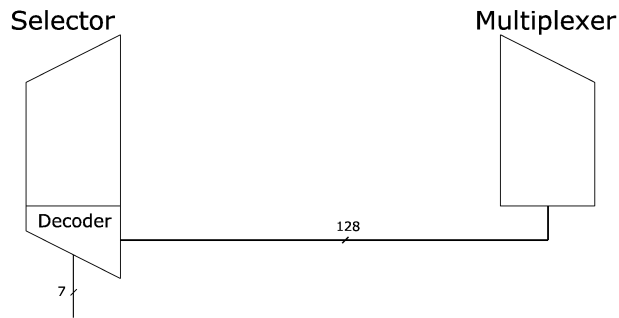


Figure 2.2: Share of the address decoder

After the multiplexer stage, a frequency divider is realised to attain a factor thousand division of the signal frequency. In this way, the frequency of a signal forwarded to a pad fulfills the requirement of the measurement equipment in any operating condition.

2.2 Switch element

As mentioned above, an ideal multiplexer is nothing more than a multi-input switch. Different circuit realisations of multiplexers exist, depending on the way the switch function is implemented.

The core element of the multiplexer is the *switch element*: depending on the value of its selection signal, its input signal can be either forwarded toward the output or cut off. All outputs of the *switch elements* are connected together at the multiplexer output. For this reason, when a *switch element* is inactive, it must not drive the output. The switch cell must thus have the possibility to set its output to a floating mode.

In this project two different types of multiplexer are taken into account. They have the same structure, but they differ in the *switch element*. The two used *switch elements* are the Pass Gate and the 3-State Buffer.

2.2.1 Pass Gate

A Pass Gate is made up by a pMOS and an nMOS transistor whose drains and sources are connected together (Figure 2.3).

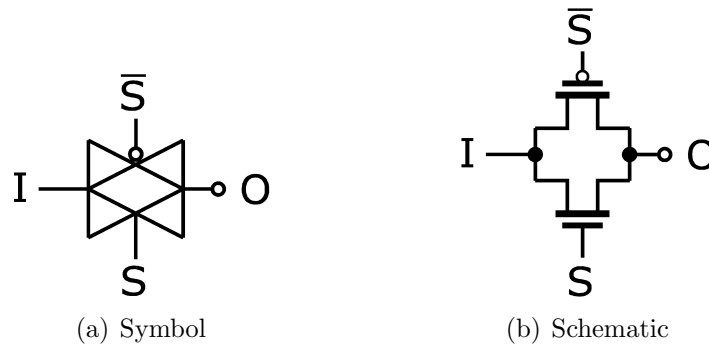


Figure 2.3: Pass Gate element

When the selection bit S is high, the channel of both pMOS and nMOS transistor is formed, thus the input I and the output O can be roughly considered shorted. When the selection bit is low, none of the MOS transistors has a $|V_{GS}| > |V_T|$, therefore the input I and the output O are open-circuited, and the O node can be considered floating.

The logic functionality of a Pass Gate is displayed in Table 2.1.

I	S	O
0	0	Z
1	0	Z
0	1	0
1	1	1

Table 2.1: Pass Gate logic functionality

A Pass Gate 8x1 multiplexer, simulated in a 45 nm CMOS technology, operates correctly up to 5 GHz with nominal supply voltage of 1.1 V and minimum size transistors ($W = 120$ nm and $L = 40$ nm).

Since the Pass Gate gain is always equal or smaller than 1, its input signal cannot be enhanced. For this reason, using Pass Gates as *switch elements*

for the multiplexer does not guarantee high robustness.

On the other side, this is the solution assuring the lower area occupancy, since each *switch element* is constituted by two minimum dimension transistors only.

2.2.2 3-State

A simple way to realise a 3-State Buffer is shown in Figure 2.4: the two MOSFETs closer to the output node (nMS and pMS) are driven by the selection bit S (nMOS) and the negated selection bit \bar{S} (pMOS); the two MOSFETs next to the power rails (nMI and pMI) are driven by the input signal I .

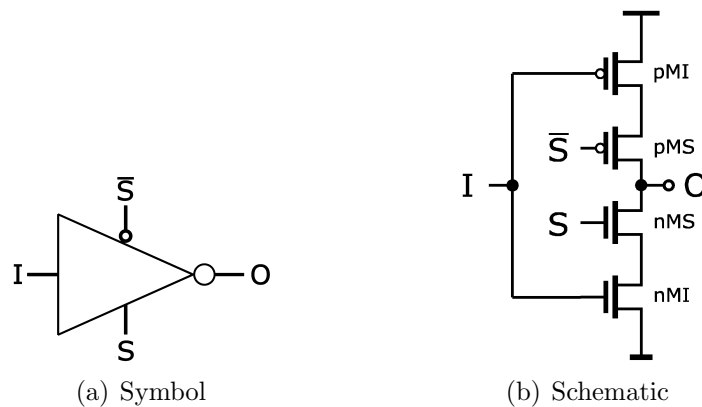


Figure 2.4: 3-State Buffer element

When the selection bit S is high, nMS and pMS can be considered as closed switches, and the circuit behaves like a simple CMOS inverter driven by the input I , whose output is O . When the selection bit S is low, nMS and pMS are approximately open switches, and the output O is floating.

The logic functionality of a 3-State Buffer is displayed in Table 2.2.

The 3-State Buffer described above is just an inverter (given by nMI and pMI) driven by a high frequency signal, and two cascaded switches (nMS and pMS) driven by a low frequency signal. The critical part for robustness is thus the inverter: a correct dimensioning for nMI and pMI is vital. Through

I	S	O
0	0	Z
1	0	Z
0	1	1
1	1	0

Table 2.2: 3-State Buffer logic functionality

simulations it has been noticed that good results were obtained by having $W_{pMI} \simeq 1,414 \cdot W_{nMI}$, where W_{nMI} and W_{pMI} are respectively nMI and pMI widths.

The 3-State Buffer gain is $\frac{g_{mMI}}{g_{mout}}$, with g_{mout} directly dependent on W_{MS} . With these transistor dimensions, the 3-State Buffer gain is greater than $10dB$. Therefore, when a signal is selected through a 3-State Buffer multiplexer, its logic levels are restored.

Unfortunately, the higher robustness of this solution is paid in area: this *switch element* requires 4 transistors instead of the 2 needed by the Pass Gate, and they may not be at minimum size. To achieve the required speed the total area is greater than two times the Pass Gate area.

This solution requires also supply power, while the Pass Gates are passive circuits, disconnected by the power rails.

2.3 Multiplexer architecture

In this Section different multiplexer architectures are analysed. For each one of these architectures advantages and drawbacks are pointed out. In particular, the architecture analysis focuses on the capability to fulfill the requirements of this project.

In Section 2.1 it has been mentioned that the robustness of a mux can be evaluated as its capability to forward the selected signal without frequency distortions. A limitation to the multiplexer robustness is given by the maximum capacitance drivable by a single *switch element*.

Since the RingOs operating frequency is known, the maximum drivable load capacitance of a single *switch element* is determined through simulations at that very frequency. The results below showed are obtained using Pass Gates as *switch element*.

Figure 2.5 shows that for a $C_L \geq 3.5$ fF the Pass Gate multiplexer output is less than 90% of the amplitude of its input. Since the nMOS and the pMOS transistors have the same dimensions, this loss is asymmetrical for the high and the low logic level of the signal, as shown in Figure 2.6. Thus the value of 3.5 fF is way too high.

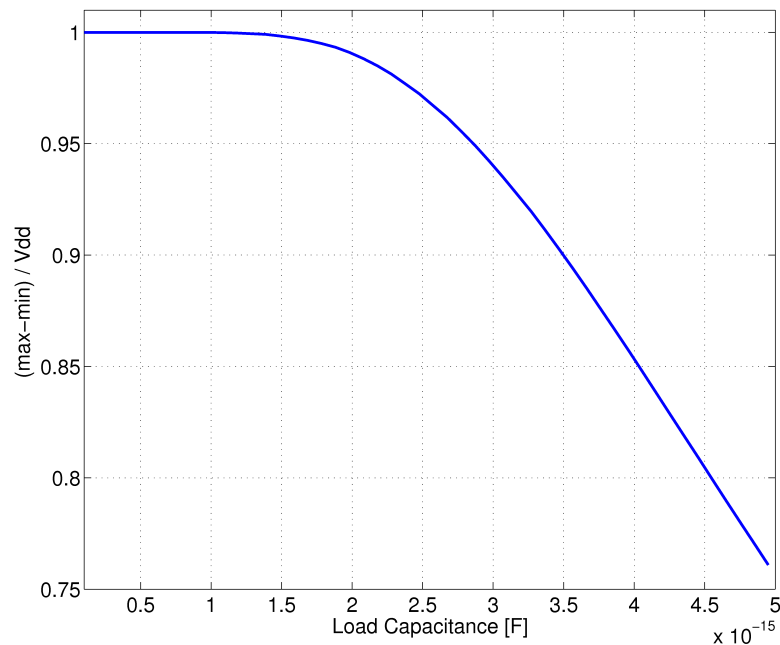


Figure 2.5: Output signal dynamic amplitude vs. load capacitance

Figure 2.6 shows that, at the operating input frequency, the output waveforms have acceptable values of both maximum and minimum amplitudes for a load capacitance lower than 2.5 fF. For the 3-State Buffers very similar results are obtained.

In the following, for the discussed architectures the value of the output capacitance is derived, in order to evaluate their reliability.

The two main design approaches to realise a multiplexer are a non hier-

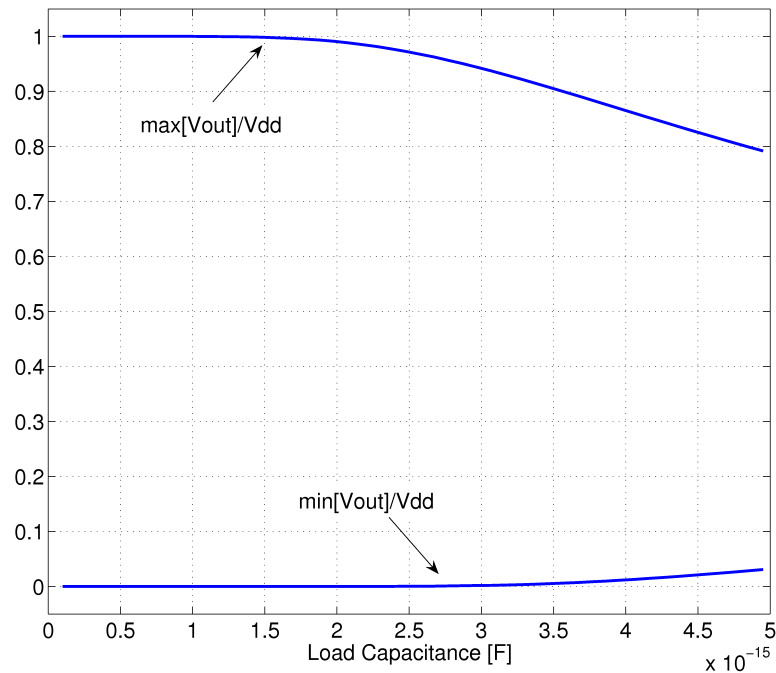


Figure 2.6: $\frac{\max[V_{out}]}{V_{DD}}$ and $\frac{\min[V_{out}]}{V_{DD}}$ vs. load capacitance

archical architecture (Section 2.3.1) and a multi stage or hierarchical architecture (Section 2.3.2).

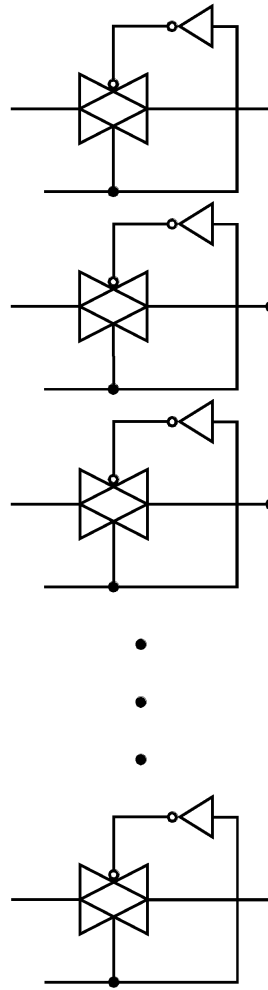
In Section 2.3.4 and 2.3.3 possible modifications to further improve the hierarchical architecture functionalities are discussed.

2.3.1 Single array

The simplest method to realise a multiplexer is to connect the output of all the *switch elements* at the output of the multiplexer. This solution is known as non hierarchical mux or single array mux.

If the multiplexer is organised as an array of N elements (Figure 2.7), the load capacitance for a single *switch element* is given by the parallel of the output capacitances of the N *switch elements* plus the load of the following stage. Therefore:

$$C_L = C'_L + C_{EXT} = N \cdot C_{OUT}^{se} + C_{EXT}$$

Figure 2.7: Array of N *switch elements* (Pass Gates)

If a Pass Gate *switch element* is chosen, the capacitance at the output of a single element (C_{OUT}^{PG}) is mainly given by the parallel of the two Drain-to-Body capacitances of its nMOS and pMOS transistor. Thus:

$$C_{OUT}^{PG} \approx C_{BDn} + C_{BDp} \simeq C_J L_c (W_n + W_p) = 2 \cdot C_J L_c W_{min}$$

Where C_J (the junction capacitance) and L_c (the contact length) can be considered, in first approximation, technology defined parameters.

The output capacitance C_{OUT}^{PG} of the designed Pass Gate, determined

through a Spectre simulation, is ca. 200 aF; while the total output capacitance C'_L of the unloaded mux results ca. 24.5 fF, thus close to the theoretical value $C'_L = 128 \cdot C_{OUT}^{PG} \simeq 128 \cdot 0.2 \text{ fF} = 25.6 \text{ fF}$.

Since the capacitance at the output of a 3-State Buffer is comparable to the Pass Gate one, and since it is no critical to have the inner transistors (nMS and pMS) at minimum width (W_{min}), $C_{OUT}^{PG} \approx C_{OUT}^{3SB}$. For this reason, there is no evident advantage in the load capacitance value using the Pass Gate solution rather than the 3-State Buffer one.

The value found for C'_L is more than 10 times higher than the 2.5 fF limit. Therefore, despite its simplicity, a non hierarchical solution is unacceptable in this technology node for a 128x1 multiplexer.

2.3.2 Multi stage

Due the load capacitance limit, a hierarchical architecture for the multiplexer has been adopted. The mux is redesigned with more than one stage. Figure 2.8 shows the case of 3 cascaded stages.

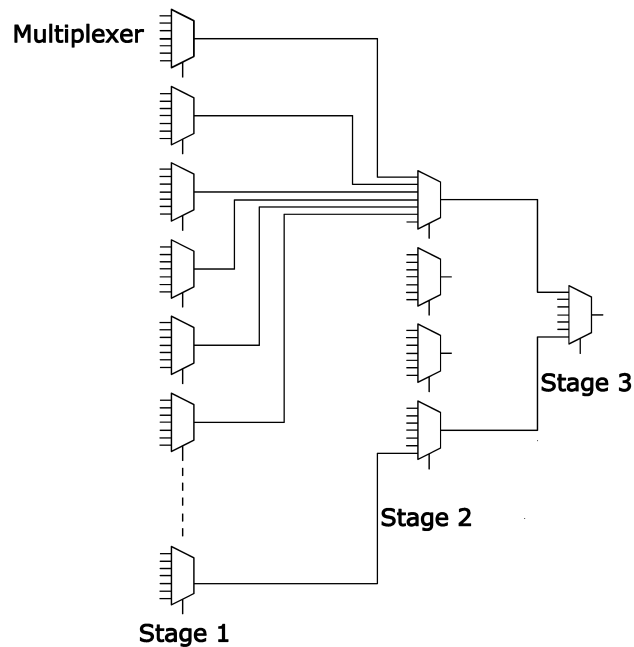


Figure 2.8: Multi stage hierarchical mux (3 stages)

Adopting a hierarchical architecture increases the multiplexer area, and makes its structure more complicated. Both these effects get worst as the number of stages increases. On the other hand, having more than one stage reduces the dimensions of the single stage components, and thus their capacitances. Moreover, a hierarchical multiplexer is controlled through a hierarchical addressing. Therefore it is no longer necessary to route the $N = 128$ selection signals from the decoder realised in the selector to the mux (Figure 2.2). The number of signals to route depends on the chosen architecture.

Two stages mux

First, let us consider a 2 stages 128x1 multiplexer, made up by 16 8x1 mux (first stage) whose outputs are connected to a 16x1 mux (second stage).

In this case, to perform the selection, 8 mutually exclusive signals (or 3 coded bits) are needed for the first stage and 16 for the second stage. The number of signals to be routed from the selector is thus 24.

The load capacitance for the first stage results:

$$\begin{aligned} C_L &= C'_L + C_{EXT} = C_L^{1^{st}stage} + C_{in}^{2^{nd}stage} = 16 \cdot C_{OUT}^{se} + C_{in}^{2^{nd}stage} \simeq \\ &\simeq 16 \cdot 0.2 \text{ fF} + 0.5 \text{ fF} = 3.7 \text{ fF} \end{aligned}$$

Even if the value of C_L is lower than for the single array multiplexer, it is still too high to be driven by a signal at the required operating frequency.

Therefore, a two stages multiplexer does not fit the requirements.

Three stages mux

A three stages hierarchy is then considered. The chosen hierarchical architecture is 4x4x8:

1. The first stage has 128 inputs and 32 outputs. It is made up by 32 4x1 muxs.

2. The second stage has as inputs the 32 outputs of the previous stage, and it has 8 outputs. It is made up by 8 4x1 muxs.
3. The third stage has as inputs the 8 outputs of the previous stage, and it generates the output of the entire 128x1 multiplexer. This stage is just one 8x1 mux.

In this case, to perform the selection, 4 mutually exclusive signals (or 2 coded bits) are needed for the first stage, 4 for the second stage, and 8 for the third stage. The number of signals to be routed from the selector is thus 16.

The critical stage for the load capacitance is the last one, since is there that the highest value of C'_L is found:

$$C'_L = 8 \cdot C_{OUT}^{se} \simeq 8 \cdot 0.2 \text{ fF} = 1.6 \text{ fF}$$

that is less than the 2.5 fF limit. A three stage multiplexer thus fits the requirements.

2.3.3 Standard library multiplexer

Alongside the two main solutions (Pass Gate mux and 3-State Buffer mux), another multiplexer type has been taken into account in the performance evaluation. This multiplexer type is a standard library cell, based essentially on 3-State Buffers. The standard cell is a 4x1 mux, counting 20 transistors (Figure 2.9). It is important to notice that the choice of a 4x1 multiplexer fits the requirements of the 4x4x8 hierarchical architecture discussed above.

From Figure 2.9 can be seen that the operative principle of the standard cell mux is the same of a 3-State Buffer mux. The selection signals are $S0 : S3$; the inputs are $D0 : D3$. The main advantage of this solution is given by the area reduction obtained by a clever logic minimization. For this cell, there is no need to have both the select and the negated select signals to activate a path, as it happens in the Pass Gate and the 3-State Buffer cells shown in Figures 2.3 and 2.4.

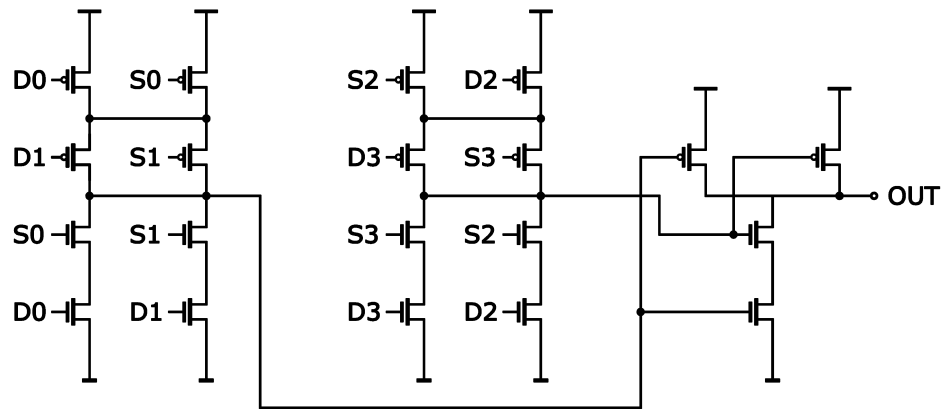


Figure 2.9: Standard cell: 4x1 mux

Area occupancy

A rough evaluation of the area occupancy of a 4x1 mux can be given counting the number of its transistors. Table 2.3 shows the total number of transistors for the three solutions, considering the inverters needed to negate the selection signals (4 for the 3-State Buffers and the Pass Gates, none for the standard cell) and the ones to regenerate the output signal (1 for the 3-State Buffers and the Pass Gates, 2 to take in account the NAND of the standard cell).

	<i>Switch elements</i>	FETs per <i>se</i>	Inverters	Total
Pass Gate	4	2	5	18
3-State Buffer	4	4	5	26
Std Cell	4	4	2	20

Table 2.3: Number of transistors for a 4x1 Mux

It is then possible to realise a hybrid multi stage multiplexer, using the 3-State based standard cells for the first and the second stage to ensure robustness, and a 8x1 Pass Gate mux for the last stage to minimize size.

2.3.4 Frequency Divider mux

Several simulations have been run to compare the multi stage multiplexers based on Pass Gates, on 3-State Buffers and the hybrid multiplexer.

In the typical case simulations (TT), all 3 solutions give comparable results. In the corner simulations (FF and SS), the Pass Gate evidences its lower robustness.

Although all 3 architectures present an output waveform with acceptable amplitude, duty cycle and correct frequency, further architectures have been investigated to obtain a multiplexer closer to the ideal functionality.

In the three stage mux, the last stage is the critical one for the output capacitance value, and thus is the one that can more distort the signal.

It has been mentioned in Section 1.3.3 that the frequency of the mux output must be divided by a factor thousand in order to fulfill the requirements of the measurement equipment. For this reason a frequency divider is present. It can be implemented by the cascade of 10 Flip-Flops.

Since the output of a Flip-Flop is a square wave with duty cycle $\delta = 50\%$ and half the frequency of its input, a FF could be used to regenerate the signal before the third stage of the mux. Therefore, a stage of Flip-Flops may be inserted before the last stage of the multiplexer.

The circuit thus obtained is called Frequency Divider multiplexer (FD mux), and performs the multiplexing and a factor 2 division of the input signal frequency.

With a FD mux, the input signals of the third stage, the critical one, are completely regenerated and their frequency is halved.

In Figure 2.10 a comparison between the last stage waveforms of a three stage multiplexer and a three stage FD mux is given. Both multiplexers use 3-State Buffers as *switch elements*. The data shown below are obtained in a typical case simulation (TT), thus the driving signal frequency is about 5 GHz.

Even in the more robust solution, the 3-State Buffers multiplexer, the advantages in robustness given by the FD mux are perceptible. Using a FD

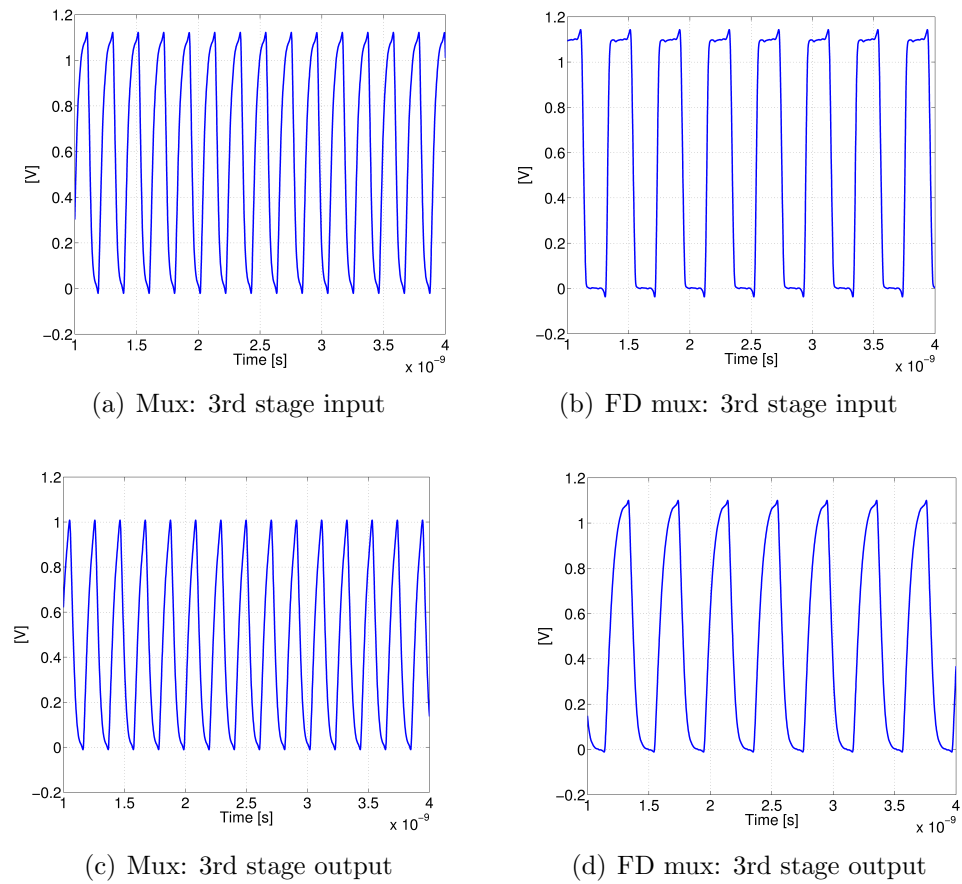


Figure 2.10: Comparison between a Mux and an FD mux

mux the multiplexer output is a rail-to-rail signal, with $\delta \approx 50\%$, and smaller raise and fall times. None of these characteristics is reached in a simple three stage mux.

The advantage of this architecture is that it makes the last stage not critical for the mux functionality.

The FD mux architecture has been therefore adopted.

The main drawback of this architecture is that, since a Flip-Flop is needed before each of the 8 inputs of the last stage, it occupies a slightly larger area.

2.4 Performance comparison

In Section 2.3.4 the choice of a 4x4x8 multiplexer with a factor 2 frequency divider before the last stage was discussed. Defined the architecture, 3 paths can be followed in the realisation of the multiplexer, depending on the used *switch element*:

1. Pass Gate multiplexer;
2. 3-State Buffer multiplexer;
3. Hybrid multiplexer, using the 3-State based standard cells for the first and the second stage, and a 8x1 Pass Gate mux for the last stage.

Among the evaluated performances, in this Section a comparison between the three solutions is carried out taking into account the following features:

- input/output functional bound;
- variability introduced by the multiplexer itself;
- area occupation;
- power consumption.

The choice of the *switch element*, based on this comparison, is discussed in Section 2.4.5.

2.4.1 I/O bound function

Because of the insertion of a factor 2 frequency divider before the last stage, the three stage FD mux is no longer a linear system. For this reason, its behaviour can no more be described by a transfer function. Nonetheless, it is possible to derive a functional relation between the input and the output signals of the multiplexer. In particular, the relation between input and output frequencies has been hence defined bound function.

For the 3 solutions the bound functions are derived in the typical case (TT), to evaluate their reliability. It is useful to remind that the nominal frequency of a 7 stages ring oscillator is about 5 GHz.

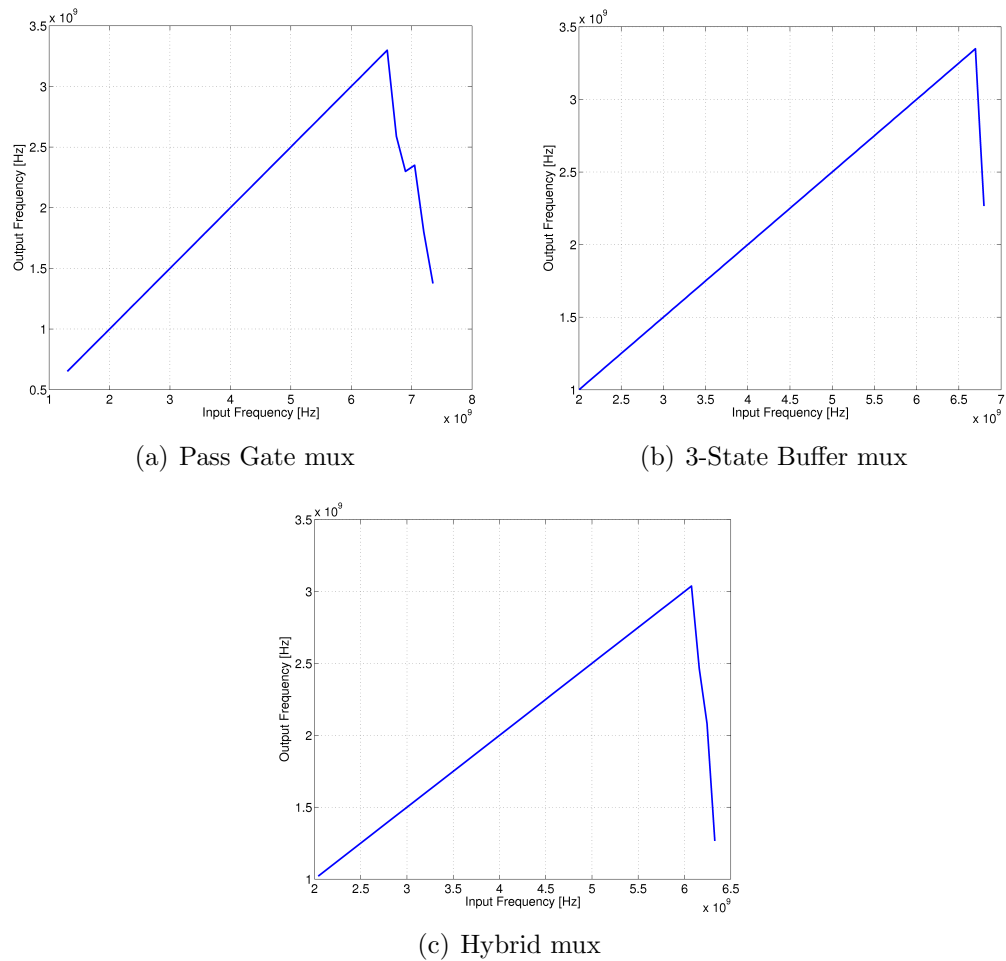


Figure 2.11: I/O bound functions for the 3 solutions

From Figure 2.11, it can be seen that between the input and the output frequencies a linear relation exists. This relation is maintained in a bandwidth that is approximately the same for the 3 solutions: up to about 6.7 GHz for the Pass Gate and the 3-State Buffer multiplexers, and up to about 6.1 GHz for the hybrid mux. The slope factor of the bound functions is $1/2$, due to the frequency divider.

2.4.2 Variability

As mentioned in Chapter 1, among the aims of this project there is the measurement of the process variability, to investigate solutions that may reduce the spread of technology parameters. This analysis is carried out through the design of ring oscillators and Flip-Flops. For the RingOs, the statistical distribution of the parameters spread may be estimated through the analysis of their oscillating frequency.

The multiplexer must only select the signal generated by one of the ring oscillators, and forward it to a pad, in order to have it available for the frequency measurement. From the signal frequency is then possible to quantify the average variations of the delay of the gates making up the ring oscillators.

Therefore the mux must not introduce unwanted variations to the output frequency of the oscillation that has to be measured.

To quantify the variations introduced by the different multiplexer types, Monte Carlo simulations are used. Based on the bound functions shown in Figure 2.11, the behaviour of each multiplexer is simulated in 5 points, given by the following input frequencies: 2 GHz, 3.5 GHz, 5 GHz (the expected operating frequency), 5.3 GHz and 6 GHz.

For each point, 500 iterations are carried out to obtain an acceptable statistical significance.

The result of each of the 15 Monte Carlo simulations (5 points for each of the 3 solutions) is a stochastic distribution. The data obtained are shown in Table 2.4.

From the distribution mean value μ and standard variation σ , the relative variability $\frac{\sigma}{\mu}$ is derived.

At the operating frequency of 5 GHz all 3 solutions present a very low variability: $\frac{\sigma}{\mu} < 77ppm$.

At lower frequencies the results are even better.

At 5.3 GHz variability is proved to be still very low: $\frac{\sigma}{\mu} < 94ppm$.

At 6 GHz the introduced variations are unbearable, since this frequency is the closer to the upper limit of the FD mux, especially in the hybrid case

		2 GHz	3.5 GHz	5 GHz	5.3 GHz	6 GHz
PG	μ [GHz]	0.9999999	1.75	2.4999999	2.6500055	2.9244825
	σ [Hz]	3788.5	3463.9	1611.1	126.68e3	275.6e6
	variation	3.8 ppm	2 ppm	0.6 ppm	47.8 ppm	9.4%
3SB	μ [GHz]	0.9999969	1.7499956	2.499993	2.6499925	2.9315076
	σ [Hz]	4992.7	5048.7	5904.2	6646.6	260.102e6
	variation	5 ppm	2.9 ppm	2.4 ppm	2.5 ppm	8.9%
Hybrid	μ [GHz]	1.0000002	1.7500001	2.4999909	2.6500006	2.7805446
	σ [Hz]	3522.7	2653.9	191.74e3	248.24e3	450.464e6
	variation	3.5 ppm	1.5 ppm	77 ppm	93.7 ppm	16.2%

Table 2.4: Monte Carlo simulation results

(Figure 2.11). Nonetheless, since simulations proved that the frequency of the ring oscillators varies in a very small range around its operating point (about $5\text{ GHz} \pm 3\%$) in the typical case, the results at 6 GHz are not of major concern.

Even if at the nominal operating frequency the variations introduced by the hybrid FD mux, the worst performing one, are 130 times larger than the results obtained with the Pass Gate FD mux, their effect on the output signal is much lower than the variations introduced by the RingOs. Therefore, at this stage, all 3 solutions are still available.

2.4.3 Area occupation

A common way to estimate area occupation for digital circuits is to count the number of transistors.

In Table 2.3 the number of transistors needed to realise a 4x1 multiplexer in the 3 different solutions is quantified.

In Table 2.5 the same count is carried out for a 128x1 FD mux. The amount of *switch elements* (3SB: 3-State Buffer; PG: Pass Gate; Std: standard 4x1 mux), Flip-Flops (FF) and inverters (IVX) is quantified. The number of transistors for each element is reported between brackets. In the last column the total number of transistors for each mux solution is calculated.

	<i>Switch elements</i>			Logic		Tot
	PG (2)	3SB (4)	Std (20)	IVX (2)	FF (26)	
Pass Gate	168	0	0	330	8	1204
3-State Buffer	0	168	0	490	8	1860
Hybrid	8	0	40	168	8	1360

Table 2.5: Number of transistors per multiplexer

To count the number of inverters, both the ones needed to regenerate the signals between the stages and the ones to negate the selection bits are taken into account. As expected, the Pass Gate solution guarantees the lower area occupation. For the hybrid solution, since the negated selection bits are not needed, the area is slightly higher than for the Pass Gate solution, but still much lower than for the 3-State Buffer mux.

2.4.4 Power consumption

The power consumption for a single path through the multiplexer is simulated. The power from the supply voltage is evaluated apart from the power absorbed from the ring oscillator (Table 2.6)

	Supply [μ W]	RingO [μ W]	Total consumption [μ W]
Pass Gate	161.47	5.389	166.859
3-State Buffer	189.68	0.165	189.845
Hybrid	175.12	0.805	175.925

Table 2.6: Power consumption for a single multiplexer path

The trend results to be very similar to the one found for the area, since the total power consumption raises as the number of transistor increases.

It can be observed that the Pass Gate solution, although is the least consuming, absorbs more power from the ring oscillators than the other solutions. This is due to the intrinsic nature of the Pass Gate, that does not regenerate the input signal, but brings it directly to the output. Subtracting

current from the ring oscillators could affect their frequency and even prevent them from oscillating. Thus, the robustness of the Pass Gate mux must be verified in the worst case corner.

On the other side the 3-State Buffer solution needs almost no power from the ring oscillators, but absorbs more power from the supply pin than the other solutions.

As for the area occupation, the hybrid solution is a trade off between the other ones.

2.4.5 Switch element choice

As mentioned before the 3-State Buffer FD multiplexer is more robust than the Pass Gate FD mux, especially in the corner cases (FF and SS). On the other side, Pass Gate FD mux proved to be better in area occupation and supply power consumption.

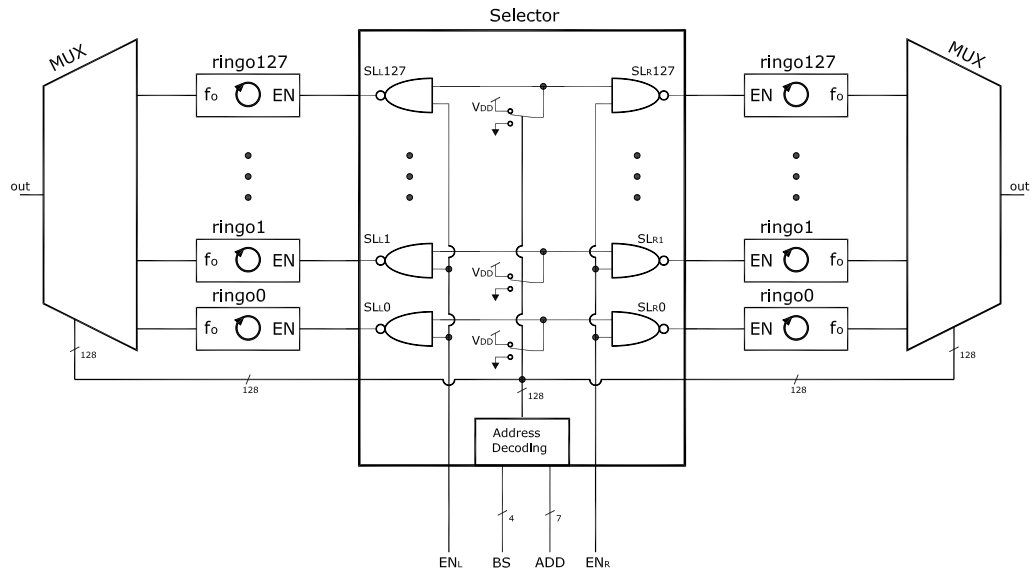
Between these two solutions, the hybrid FD multiplexer resulted to be almost as robust as the 3-State Buffer one, since is also based on 3-State elements, but less area and power consuming. Therefore, for the multiplexer realisation, a hybrid solution is adopted, using the 3-State based standard cells for the first and the second stage, and a 8x1 Pass Gate mux for the last stage.

2.5 One mux per C -Block

In this Section, a possible modification of the project specifications is discussed, in order to reduce the area occupation of the multiplexer, without decreasing its performances.

For each C -Block two arrays of 128 ring oscillators are present: one composed by 7 logic depth RingOs and the other one by 11 logic depth RingOs. To select one out of the 128 oscillators a FD multiplexer per array is implemented. The block diagram of the circuit is shown in Figure 2.12.

An area improvement can be obtained using a single multiplexer to select

Figure 2.12: Block diagram of the *C*-block

one oscillation out of the 256 coming from the 2 arrays. A 256x1 FD mux can be designed combining the two 128x1 FD mux described above. This solution would save a large part of the logic block and one output pad.

The two possible architectures to implement the 256x1 FD multiplexer are:

1. 2x4x4x8 FD mux: modify the hybrid 4x4x8 FD multiplexer adding 256 Pass Gate as first stage;
2. 4x4x4x4 FD mux: completely based on the standard cell (4x1 mux).

The area estimation for the two solutions clearly point out the advantage of the first one (Table 2.7).

	PG (2)	Std (20)	Tot
2x4x4x8 mux	256+8	40	1328
4x4x4x4 mux	0	85	1700

Table 2.7: Number of transistors per multiplexer

Moreover, the adoption of a 4x4x4x4 solution unable the sharing of the selector decoder, since it is designed for a 4x4x8 multi stage architecture. On the contrary, switching from the 4x4x8 to the 2x4x4x8 architecture is very simple, since the first stage may be driven by the EN_L and EN_R bits, with no need of additional logic.

Thus, only one 2x4x4x8 hybrid FD multiplexer per C -Block is realised. It is implemented with 128 2x1 Pass Gate mux as first stage, the 3-State based standard cells for the second and the third stages, and a 8x1 Pass Gate mux for the last stage (Figure 2.13).

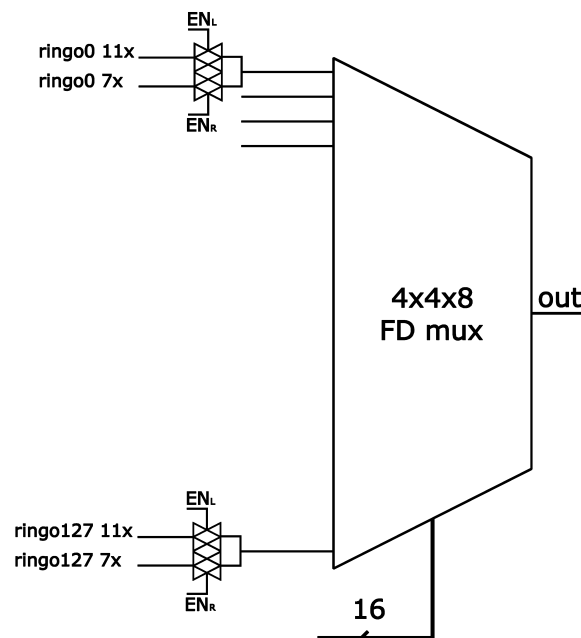


Figure 2.13: Realisation of the 2x4x4x8 FD mux, modifying the 4x4x8

As discussed in Section 2.4.4, realising the first stage with Pass Gates as *switch elements* could decrease the mux robustness, since the power absorbed from the RingOs could prevent them from oscillating. Nonetheless, simulations proved that the 256x1 FD mux has no decreased functionality, and the overall robustness is guaranteed by the intrinsic robustness of the standard cells constituting the second and the third stage.

2.6 Technology change

During the development of this project, a technology change occurred. All the data and the result discussed until this point refer to the Crolles CMOS 45 nm technology flow. However, due to business choices, the TSMC CMOS 45 nm has been adopted.

Even if the fundamental design choices made for the multiplexer are still valid, the technology change necessarily affects the results given in this Chapter. In this Section, a brief overview of the most significant differences is presented.

The considerations presented in Section 2.3 maintain their reliability in the TSMC flow, therefore the architectural choice of a hybrid FD multiplexer (see Section 2.4.5) is kept in the new technology.

2.6.1 Performance alterations

The main difference encountered in the technology change is in the operating frequency of the ring oscillators. In Table 2.8 the operating frequency modification for 7 stages and 11 stages ring oscillators is reported (inverter based).

		Crolles	TSMC
Operating frequency	RingO 7x	5 GHz	12 GHz
	RingO 11x	3.2 GHz	7.6 GHz

Table 2.8: Inverter RingOs operating frequencies

A significant increase in the operating frequency is noticeable. Moreover, the I/O bound function of the Crolles hybrid mux reported in Figure 2.11(c) testifies its inadequacy for the new frequency specifications. In Figure 2.14, the I/O bound function for a FD hybrid mux realised with *Low* V_T TSMC transistors is presented.

Between the input and the output frequencies a linear relation is attained for a bandwidth up to about 13.6 GHz, thus greater than the expected

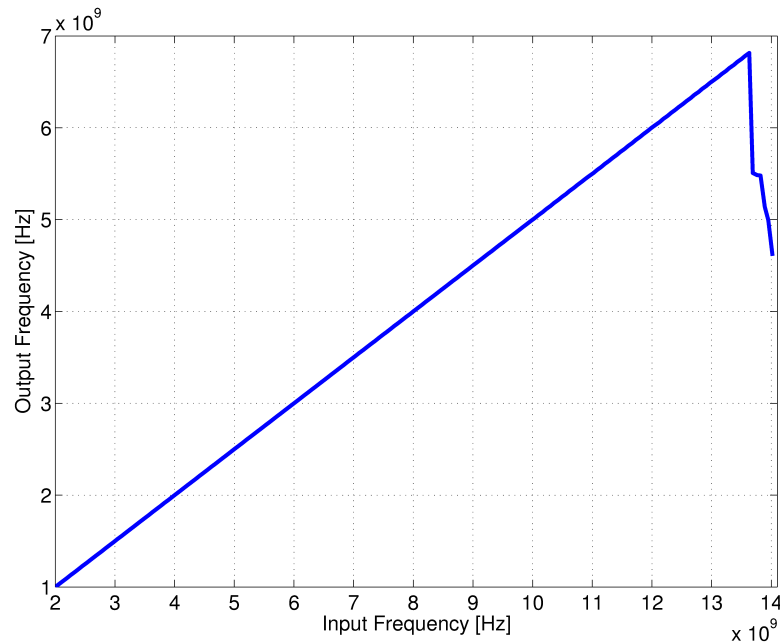


Figure 2.14: I/O bound function for the TSMC Hybrid FD mux

oscillating frequency of the RingO 7x. The slope factor of the bound function is $1/2$, due to the frequency divider.

In Section 2.4.2, the need of Monte Carlo simulations to quantify the variations introduced by the multiplexer itself is discussed. Even for the new multiplexer, Monte Carlo simulations are used.

Based on the bound function shown in Figure 2.14, the multiplexer behaviour is simulated in 7 points, given by the following input frequencies: 2 GHz, 3.5 GHz, 5 GHz, 6 GHz, 10 GHz, 11 GHz and 12 GHz (the expected operating frequency). For each point, 400 iterations are carried out to attain an acceptable statistical significance. The result of each Monte Carlo simulation is a stochastic distribution, and the obtained data are shown in Table 2.9.

In all simulations variability is proven to be very low: $\frac{\sigma}{\mu} < 12ppm$.

However, in the simulation run at the expected operating frequency (12 GHz), out of 400 iterations, 90 produced a non acceptable output signal, due to the decreased voltage levels throughout the mux. Even if rising the supply volt-

input	typ	Monte Carlo	
f_{in} [GHz]	f_{out} [GHz]	μ [GHz]	σ [Hz]
2.00	1.00	1.00	105.76
3.50	1.75	1.75	146.07
5.00	2.50	2.50	171.83
6.00	3.00	3.00	180.37
10.00	5.00	5.00	2006.5
11.00	5.50	5.50	2213.1
12.0	6.00	6.00*	67447*

Table 2.9: Monte Carlo simulation results

age V_{DD} from 1.1 V to 1.2 V solves the problem (see Table 2.10), this result does not represent a concern, since a discrepancy is expected between the ideal and the real RingOs operating frequency (see Section 2.7).

input	typ	Monte Carlo	
f_{in} [GHz]	f_{out} [GHz]	μ [GHz]	σ [Hz]
12.0	6.00	6.00	0.181

Table 2.10: Monte Carlo simulation results at $V_{DD} = 1.2$ V

2.7 Extracted parameters

To quantify the discrepancy between the expected and the actual operating frequency of a ring oscillator, parasitic parameters are extracted from the layout realisation of the Main Core.

It is known that the operating frequency of a ring oscillator is given by the delay of the cells constituting it:

$$f = \frac{1}{2 \cdot (t_{d_1} + t_{d_2} + \dots + t_{d_n})}$$

The delays t_d are given by

$$t_{d_i} \propto (R_{DS_i} \cdot C_{L_i})$$

Where C_{L_i} , the load capacitance of the i -th stage, is given by $C_{in_{i+1}}$, the input capacitance of the $(i + 1)$ -th stage. However, if parasitic capacitances are considered, for each stage C_{L_i} becomes

$$C_{L_i} = C_{in_{i+1}} + C_{par_i}$$

Therefore, the operating frequency is expected to be lower in actual circuit realisations than in schematics.

Simulations from the Main Core extracted view prove the operating frequencies to be much lower than expected. In Table 2.11 the lowering due to the parasitic elements is reported. The data refer to inverter based ring oscillators.

		Views	
		Schematic	Extracted
Operating frequency	RingO 7x	12 GHz	4.4 GHz
	RingO 11x	7.6 GHz	2.8 GHz

Table 2.11: Inverter RingOs operating frequencies

Table 2.9 proves that for these values of operating frequency the variability introduced by the multiplexer is so low that can not affect the variability measurement for the ring oscillators.

Chapter 3

Selector

The main core of the testchip is based on ring oscillators, organised in structures called *C*-Blocks, as described in Section 1.3.3.

To singly activate the ring oscillators (RingOs) present in each *C*-Block, mutual exclusive selection signals are needed. In order to generate these selection bits from coded addresses, a specific circuit, hence named selector, is designed.

An additional logic, a multiplexer, is then necessary to route the output of the active oscillator to the output pad, where it may be observed.

In this Chapter the design of the selector is discussed.

In Section 3.1 the project specifications are given. Furthermore, two alternative selection circuits are analysed and compared in Sections 3.1.1 and 3.1.2; special operating modes for the selector are described in Section 3.1.3.

A structural description of the selector architecture is given in Section 3.2, where the circuit blocks are discussed.

Section 3.3 deals with a modification of the architecture, that allows to share the decoding logic between the selector and the multiplexer.

3.1 Project requirements

In each C -Block two arrays of 128 ring oscillators are present; one is constituted by 11-depth RingOs, the other one by 7-depth RingOs. Since only one of the 256 oscillator outputs can be forwarded to the output pad at a given time, 256 mutual exclusive selection signals are needed.

Thus, for each C -Block a 256 lines selector is realised.

The line that has to be selected is coded by 1 enable bit, that identifies the array, and by 7 address bits, that identify the ring oscillator ($2^7 = 128$).

A C^2 -Block is made up by 4 C -Blocks, whose RingOs are realised with different logic gates: inverters (INV), NANDs, NORs and RingOs realised with a mixture of those gates (MXD).

The C^2 -Block has a single output pad, where the output of the selected RingOs of the selected C -Block is multiplexed. Therefore, at most one of the 4 C -Blocks present in each C^2 -Block need to be active at the time. For this reason a disable signal for each C -Block is present.

Furthermore, beside the selection of single RingOs, the selector has to enable special operating modes, described in Section 3.1.3.

The selector is thus characterised by the following input/output signals:

- 7 bit input Address $ADD < 6 : 0 >$, coding the selector line that has to be active;
- 1 bit input $EN11$, specifying the selected array ($EN11=1$ for the 11-depth RingOs, $EN11=0$ for the 7-depth RingOs);
- 1 bit input DIS , to disable the selection of the complete C -Block;
- 4 bit input Block Select bits $BSB < 3 : 0 >$, selecting the special modes;
- 256 bit output Select $SEL < 255 : 0 >$, enabling the corresponding ring oscillator ($SEL < 127 : 0 >$ for the 7-depth RingOs, $SEL < 255 : 128 >$ for the 11-depth RingOs).

As mentioned above, each C -Block is characterised by a different ring oscillator architecture. As an example, in Figure 3.1 is shown a 7-depth INV ring oscillator.

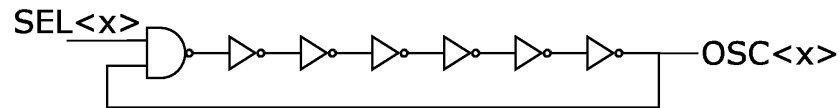


Figure 3.1: 7 stages, inverter based RingO

The first gate of a ring oscillator must have the capability to activate and to stop the oscillation. In this project, the first RingO gate is either a NAND or a NOR gate. Depending on the first gate, the selection signal must be active with different logic values.

In Sections 3.1.1 and 3.1.2 the different selector specifications deriving from the used first gate are given.

In Section 3.1.3 the special operating modes for the selector are described, and the specs for their implementation are given.

3.1.1 First RingO gate: NAND

In the NAND and in the INV based ring oscillators, the first gate is a NAND (Figure 3.1).

Table 3.1 shows the NAND gate functionality.

In 1	In 2	Out
0	0	1
0	1	1
1	0	1
1	1	0

Table 3.1: NAND logic functionality

If the $SEL < x >$ signal is low, the NAND output is forced to a high logic value. In this case the RingO does not oscillate. On the other hand, if

the $SEL < x >$ signal is high, the NAND output is determined by the other input. In this case, with an odd number of gates, the chain oscillates.

For this type of ring oscillator, the idle logic value is thus low. And to activate it the selection signal must be high.

3.1.2 First RingO gate: NOR

In the NOR and in the MXD based ring oscillators, the first gate is a NOR.

Table 3.2 shows the NOR gate functionality.

In 1	In 2	Out
0	0	1
0	1	0
1	0	0
1	1	0

Table 3.2: NOR logic functionality

If the $SEL < x >$ signal is high, the NOR output is forced to a low logic value. In this case the RingO does not oscillate. On the other hand, if the $SEL < x >$ signal is low, the NOR output is determined by the other input. In this case, with an odd number of gates, the RingO oscillates.

Thus, the behaviour of a RingO using a NOR as first gate is opposite to the one of a RingO with a NAND.

For this type of ring oscillator, the idle logic value is therefore high. And to activate it the selection signal is active low.

3.1.3 Special modes

The selector presents also four special modes, coded through the Block Select bits (BSB). These special modes enable the simultaneous oscillation of 1/4, 1/2, 3/4 or all the RingOs of the enabled C -Block. When the circuit operates in these modes, the main goal is to measure the power consumption as a function of the circuit activity. Thus there is no interest, in the special

modes, to observe the output voltage of the C -Block. In Table 3.3 the special modes and the corresponding value of the BSB are reported.

Mode (Activity)	Control bits			
	BSB3	BSB2	BSB1	BSB0
Normal	1	1	1	1
25%	1	1	1	0
50%	1	1	0	0
75%	1	0	0	0
100%	0	0	0	0

Table 3.3: Special modes for the selector

3.2 Selector structure

The structure of the designed selector of a C -Block is shown in Figure 3.2. The decoding of the 7 address bits is carried out by the two decoding (DEC) blocks: DEC_{NAND} and DEC_{NOR} represent respectively the NAND plane and the NOR plane of a common decoder.

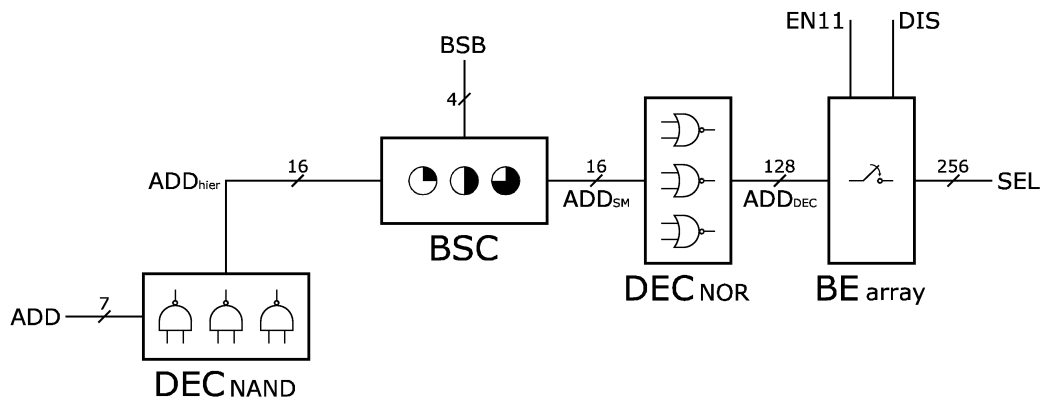


Figure 3.2: Selector block diagram

The Bit Selection Control (BSC) block implements the special modes activation.

The Block Enable (BE) array provides the enabling and disabling signals for the single ring oscillators, or it may disable the entire C -Block.

3.2.1 Address decoding

The address decoding takes place in two phases. The first phase is performed by the DEC_{NAND} block, the second phase by the DEC_{NOR} block.

Between these two phases the BSC block is present, in order to enable the special modes. Despite that, the address decoding analysis is hence carried out not considering the BSC block, that will be further discussed in Section 3.2.2. Therefore, some of the following considerations are only true figuring the Bit Selection Control block as transparent.

DEC_{NAND}

The inputs of the DEC_{NAND} block are the 7 address bits, ADD .

The outputs of the DEC_{NAND} block are 16 partially decoded address bits, hierarchically organised, ADD_{hier} .

In the DEC_{NAND} block the 7 ADD bits are negated one by one. Hence these bits are divided in 3 groups:

- Most Significant group (MSG): $ADD < 6 : 4 >$ and $\overline{ADD} < 6 : 4 >$.
- Middle group (MDG): $ADD < 3 : 2 >$ and $\overline{ADD} < 3 : 2 >$;
- Less Significant group (LSG): $ADD < 0 : 1 >$ and $\overline{ADD} < 0 : 1 >$;

For each group, every combination is used as input of a NAND plane. The outputs generated by the NAND plane are thus:

$$2^{LSG} + 2^{MDG} + 2^{MSG} = 2^2 + 2^2 + 2^3 = 4 + 4 + 8 = 16$$

In this way, 4 2-inputs NANDs are needed for the LSG, 4 for the MDG, and 8 3-inputs NANDs for the MSG (Figure 3.3).

The NAND plane architecture described above is a hierarchical 4x4x8 architecture. Due to the nature of a NAND gate, the hierarchical addresses thus obtained are active low signals.

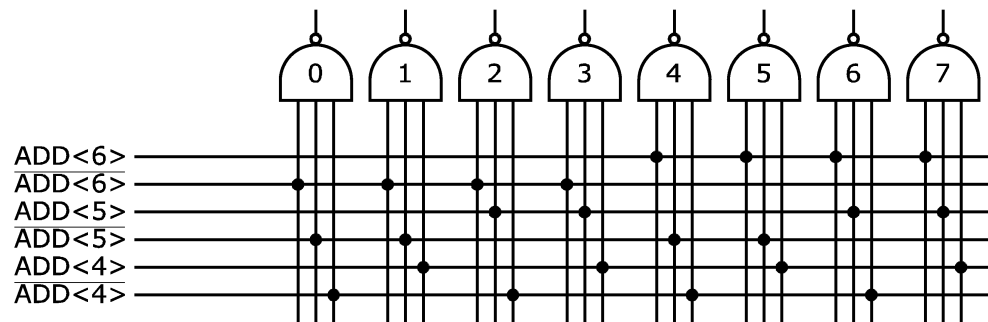


Figure 3.3: Part of the NAND plane: most significant group

DEC_{NOR}

The inputs of the DEC_{NOR} block are ADD_{SM} , obtained by the 16 hierarchical address bits generated by DEC_{NAND} through the BSC block, to activate the special modes (see Section 3.2.2).

The outputs of the DEC_{NOR} block are the decoded addresses called ADD_{DEC} . During the normal operation, they are 128 mutual exclusive signals selecting one out of the 128 ring oscillator composing a RingO array.

In the DEC_{NOR} block, the 16 input bits are divided in 3 hierarchical groups, as in the NAND plane: 4x4x8. Every combination of 3 bits, each one of them coming from a different hierarchical group, is used as input of a 3-inputs NOR gate. Therefore, this block is a NOR plane made up by $4 \cdot 4 \cdot 8 = 128$ 3-inputs NOR gates (Figure 3.4).

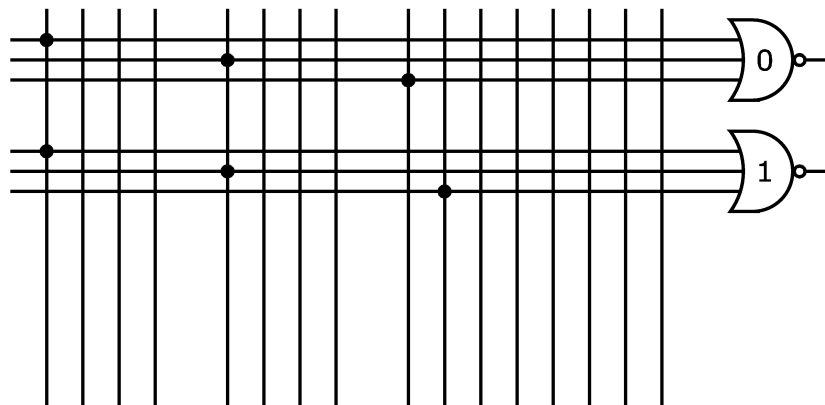


Figure 3.4: Part of the NOR plane

Due to the nature of a NOR port, the mutual exclusive selection signals obtained in this way are active high.

3.2.2 BSC

As shown in Figure 3.2, between the two address decoding phases the Bit Select Control block is inserted. This block is responsible for the activation of the special modes.

Each special mode is designed to give, during the measuring, information on the power consumption. In the special modes more than one ring oscillator is active at the same instant.

The inputs of the *BSC* block are:

- $ADD_{hier} < 15 : 0 >$: the 16 partially decoded, hierarchically organised bits generated by the DEC_{NAND} block;
- $BSB < 3 : 0 >$: selecting the special modes (see Section 3.1.3).

The outputs of the *BSC* block are 16 partially decoded signals, hierarchically organised: ADD_{SM} .

The special modes selection is given by the Block Select bits, coded in Table 3.3. Depending on the working mode, 25%, 50%, 75% or 100% of the 128 RingOs of the array is active.

The *BSC* modifies the input signals according to the information delivered by the Block Select bits. In particular, it modifies the addresses selecting more than one of them active at the same time. Therefore, when the new addresses are decoded by the DEC_{NOR} block, the selection signals are no more mutual exclusive.

In Table 3.4 the outputs of the *BSC* block in the 25%, the 50%, the 75% and the 100% mode are shown. It has to be remembered that, at this stage, the signals are active low.

Mode (Activity)	Hierarchic address (4x4x8)		
	$ADD_{SM} < 15 : 12 >$	$ADD_{SM} < 11 : 8 >$	$ADD_{SM} < 7 : 0 >$
25%	1110	0000	0000 0000
50%	1100	0000	0000 0000
75%	1000	0000	0000 0000
100%	0000	0000	0000 0000

Table 3.4: BSC outputs

3.2.3 BE array

The Block Enable array is composed by 256 enablers, one for each ring oscillator contained in a C -Block.

The inputs of the BE array are:

- $ADD_{DEC} < 127 : 0 >$: the 128 address bits produced by the DEC_{NOR} block;
- DIS : the C -Block disable signal;
- $EN11$: the array selector.

The outputs of the BE array are the 256 selection signals constituting the final outputs of the selector: $SEL < 255 : 0 >$.

In order to enable a specific ring oscillator, 3 conditions must be verified at the same time:

1. The C -Block must be selected, thus its DIS signal must be low;
2. the array containing the RingO must be selected, therefore $EN11$ must be high if the RingO is a 11-depth ring oscillator, or must be low if it is a 7-depth one;
3. the address must match the index of the RingO to be activated.

As discussed in Section 3.1.1 and 3.1.2, depending on the first gate of the ring oscillator the selection is given either by a high (NAND case) or a low (NOR case) logic value.

Thus, the SEL signals for a RingO array using NAND as first gates are obtained as

$$SEL < i > = \begin{cases} \overline{EN11} \& \overline{DIS} \& ADD_{DEC} < i > & \text{for } i \in [0, 127] \\ EN11 \& \overline{DIS} \& ADD_{DEC} < i - 128 > & \text{for } i \in [128, 255] \end{cases}$$

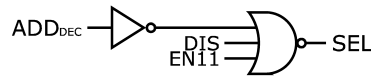
On the other hand, for the NOR case

$$SEL < i > = \begin{cases} \overline{EN11} \& \overline{DIS} \& \overline{ADD_{DEC} < i >} & \text{for } i \in [0, 127] \\ EN11 \& \overline{DIS} \& \overline{ADD_{DEC} < i - 128 >} & \text{for } i \in [128, 255] \end{cases}$$

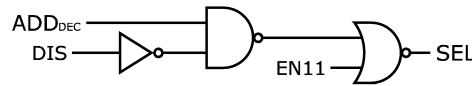
In the following, the NAND case only is discussed. All the given considerations are easily derivable in the NOR case. For a matter of simplicity, only one selection signal, $SEL < i >$ with $i \in [0, 127]$, is analysed.

In order to realise these logic functions, 3 different solutions were explored (Figure 3.5):

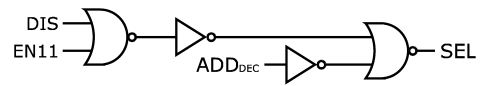
1. NOR solution: $SEL < i > = \overline{DIS \parallel EN11 \parallel \overline{ADD_{DEC} < i >}}$;
2. NAND and NOR solution: $SEL < i > = \overline{(\overline{DIS} \& \overline{ADD_{DEC} < i >}) \parallel EN11}$;
3. OR and NOR solution: $SEL < i > = \overline{(\overline{DIS} \parallel \overline{EN11}) \parallel \overline{ADD_{DEC} < i >}}$;



(a) NOR solution



(b) NAND and NOR solution



(c) OR and NOR solution

Figure 3.5: Block Enable realizing solutions

Since no solution presents clear advantages or disadvantages in the functionality, a choice has been made to reduce the area occupancy.

The NOR solution needs a 3-inputs NOR gate for each signal to be generated, thus 256. In addition, all the 128 ADD_{DEC} signals are negated, therefore 128 inverters are present. A further inverter is then needed to negate the $EN11$ signal when $i \in [128, 255]$. For the transistors account it has to be considered that 3-inputs NOR gates are made by 6 MOSFETs instead of 4.

In the NAND and NOR solution 128 NAND gates and 1 inverter are needed to generate the input signals for the 256 NOR gates giving the outputs. A further inverter is then needed to negate the $EN11$ signal when $i \in [128, 255]$.

The OR and NOR solution needs 1 NOR gate and 1 inverter to generate a global disable signal. Another NOR gate and 2 further inverters are needed to generate the global disable signal when $i \in [128, 255]$. 256 NOR gates generate the outputs from this global disable and the ADD_{DEC} signals, negated by 128 inverters.

In Table 3.5 the number of transistors needed to realise a BE array in the 3 solutions is reported.

Solution	Gates			MOS account
	NOR	NAND	INV	
NOR	256	0	129	1794
NAND and NOR	256	128	2	1540
OR and NOR	258	0	131	1294

Table 3.5: Number of gates and of transistors per BE array

Since the third solution is the one that guarantees the least area occupancy, it is the chosen one.

3.3 Selex

In Chapter 2, the design of the multiplexer is discussed. For each C -Block a multiplexer is present, whose function is to forward one out of the 256 RingO

outputs to a pad.

This multiplexer must be active during the ‘Normal Mode’ only, i.e. $BSB = 1111$.

Since the signal to be forwarded is the one produced by the selected ring oscillator, the multiplexer and the selector may share the same decoding logic. This solution is justified by the proximity of the selector to the multiplexer.

Moreover, the discussed multiplexer has a hierarchical structure, chosen to be a 4x4x8.

For this reason, it does not require to route the 256 completely decoded addresses by the selector, but only the 16 partially decoded addresses, hierarchically organised, coming from the DEC_{NAND} block. It has been mentioned in Section 3.2 that the ADD_{hier} bits are low active, therefore, a bench of 16 inverters is present in order to restore the correct logic values before delivering the signals to the mux.

A block including the selector and the multiplex is thus realised. Figure 3.6 shows this new block, hence called Selex.

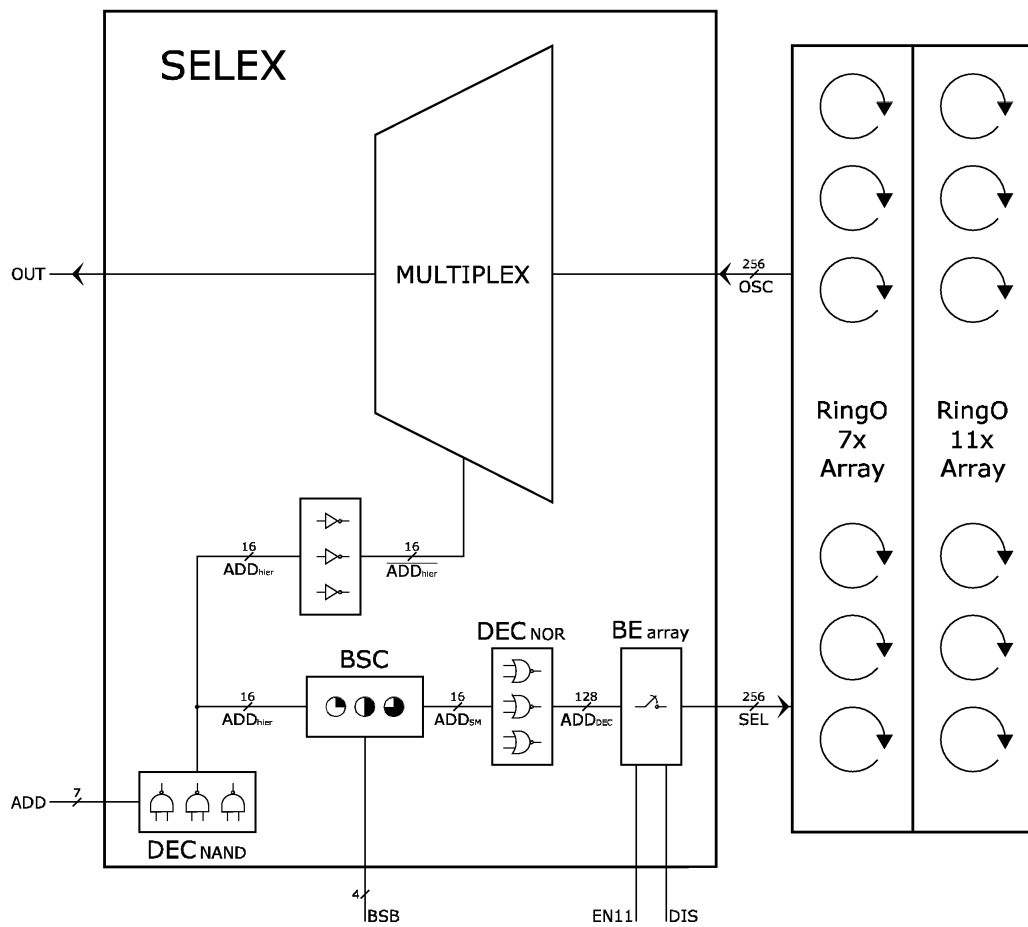


Figure 3.6: Selex and RingO arrays

Chapter 4

Digital Core

In Section 1.2 the benefits of a parsimonious power consumption are discussed. Although power consumption is a major concern in all electronics, less consuming circuits are becoming a specific research target for some electronic markets. Power is a design constraint especially for portable computers and mobile communication devices, but in many cases the design process should not subordinate it to performance.

In Section 4.1 some common low power methodologies are introduced. Among them, the *Standby voltage scaling* has been afterwards used in this project. In Section 4.2 the basic idea of the digital core and the project specifications are given. A further analysis of the core structure, functionality and performance is presented in Section 4.3. The behavioural simulations and the layout realisation of the core itself are described in Section 4.4.

4.1 Low power techniques for CMOS logic

4.1.1 Power Switching

This technique consists in disconnecting the entire circuit or some of its blocks from the power rails during their idle phases, in order to avoid leakage power. Several aspects must be taken into consideration to evaluate the benefits from using this technique in a specific design [12]:

- a beforehand division of the circuit in logical domains that may be

separately turned off must be defined;

- a control logic that determines the strategy according which the circuit should go in (or exit from) the idle state must be designed and implemented;
- the time and the power overhead necessary to recover a stage from its idle state must be taken into account;
- the power gating noise introduced by the current peaks given by the power switching must be taken into account;
- sleep transistors must be dimensioned as a trade off between resistivity and leakage [13] [14];
- multi- V_T is almost a must for this technique.

4.1.2 Standby Voltage Scaling

Standby Voltage Scaling is the reduction of the supply voltage applied to the circuit during its idle condition to a value at which the state of the circuit can be retained in the FF. It is an alternative to the power switching, adopted to preserve the information about the logical state of the nodes and to reduce the power-on time [15].

Moreover, by cleverly reducing the supply voltage, the idle circuits may continue to work, with reduced performance, in subthreshold regime (see Section 1.2.1).

This technique requires an efficient variable supply voltage generator.

4.1.3 Dynamic Voltage Scaling (DVS)

With this technique the value of the supply voltage and/or the operating frequency is scaled down every time the performance of the circuit is not critical, attaining a lower power consumption. The value of the supply voltage is determined by a circuit that periodically evaluates the time behaviour

of the different tasks carried on by the system; this circuit is called voltage scheduler.

A classification of DVS systems, that differ one from the other in the voltage scheduler realisation, has been proposed in [16].

4.2 Project Requirements

One of the energy saving techniques discussed above is the *Standby Voltage Scaling*. As the system during its idle phase works in subthreshold regime (see Section 1.2.1) the Ultra Low Power operating mode ensures some power saving, whilst the circuits may continue to work at reduced operating frequencies. The benefits of this approach are clearly visible: even in idle phase it is possible to have the system performing non critical operations.

As displayed in Table 1.1, the performance of common standard cells drops dramatically by scaling the supply voltage V_{DD} from 1.1 V to 0.3 V. This performance drop is significantly perceptible in the operating frequency, that, for a standard inverter, goes from 11.7 GHz to 27 MHz. However, since in idle phase the performances are not a major concern, it makes sense to explore this solution.

Among the structures realised on the testchip, a digital core is designed. In the digital core, standard combinatorial and sequential logic are implemented together. The aim of the digital core is to test the flexibility of a simple combinatorial and sequential block to work in strong inversion region and in subthreshold regime, at different times.

The core is constituted by several combinatorial paths, up to a logic depth of 20, whose inputs and outputs come from and end in Flip-Flop registers. Among these paths the most and the least frequently switching are the most time critical.

The standard cells constituting the combinatorial paths are Inverters, Nand (2 inputs, 3 inputs and 4 inputs), Nor (2 inputs, 3 inputs and 4 inputs), Xor (2 inputs, 3 inputs and 4 inputs) and Full Adders of the *TSMC CMOS_{45 nm}, Standard V_T* digital standard library.

The main application of the test that will be carried out with the digital core is the mobile communication market. Therefore, the performance requirements are the ones reported on Table 4.1.

State	Normal	Idle
Min operating Frequency	52 MHz	32 KHz
Supply voltage	1.1 V	0.4 V

Table 4.1: Performance specs for mobile applications

4.3 Core design

4.3.1 Core structure

The digital core has a hierarchical structure. For simplicity reasons, in the following a Bottom-Up description of the core design is given.

- The basic cell of this core is a combinatorial net. In Figure 4.1, the generic combinatorial net used in the design is displayed. It has 4 inputs (A , B , C and D) and one output.

The combinatorial net has a 20-logic depth for the longest path, and a 6-logic depth for the shortest path. The last four stages are inverters used as buffers, to restore the signal levels. Moreover, the longest path is designed to be also the slowest, since it flows through the critical path of each standard cell (i.e. the one constituted by the transistors that are the furthest from the output node).

- Four combinatorial nets are inserted after a 4 bit Flip-Flop register, connected in order to have a different functionality for each one of them (Figure 4.2). The used Flip-Flops are scannable FF with asynchronous reset. This block is called FF-Comb block.

A FF-Comb block is a sequential circuit with 4 data inputs and 4 data outputs. Each data output is a logic function of the 4 data inputs. The FF-Comb block has a latency of 1 clock cycle.

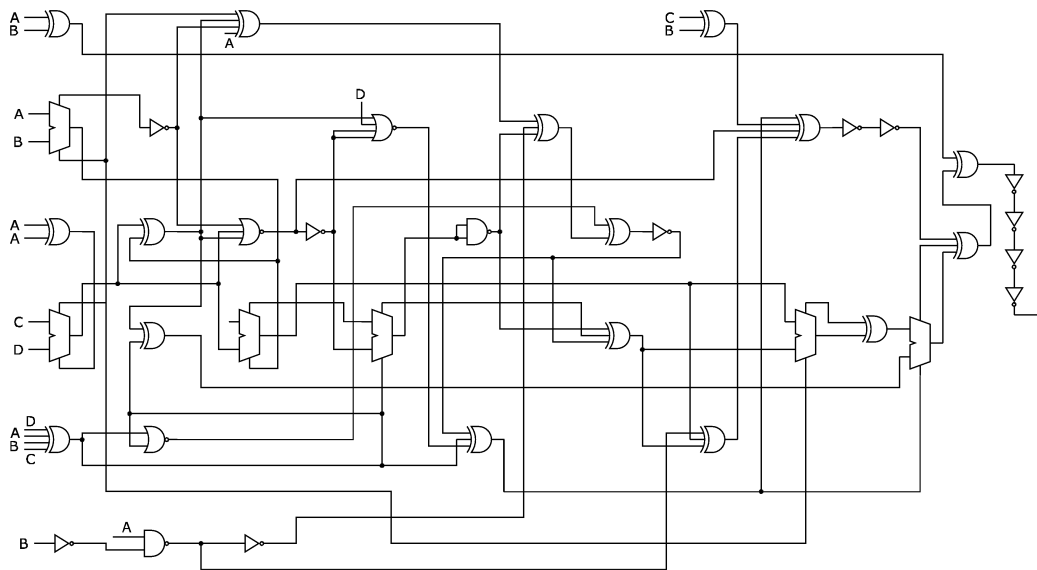


Figure 4.1: Combinatorial net

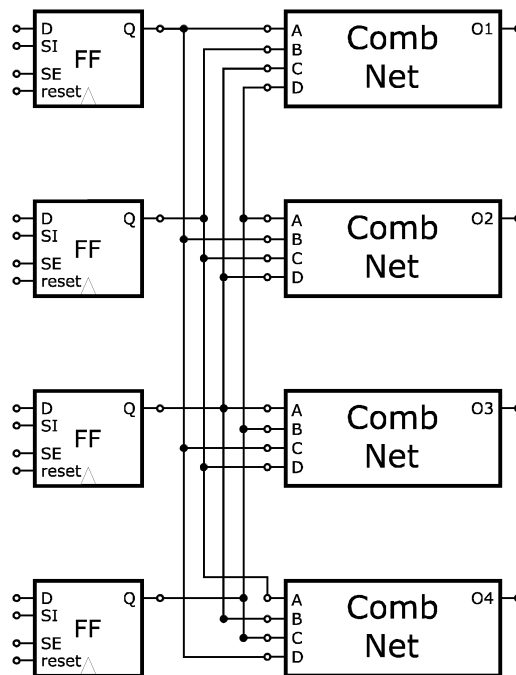


Figure 4.2: FF-Comb block

- A sequential net composed by 20 FF-Comb blocks is designed. After the last FF-Comb block a Flip-Flop register is inserted (Figure 4.3).

This net takes the name of FF-Comb chain.

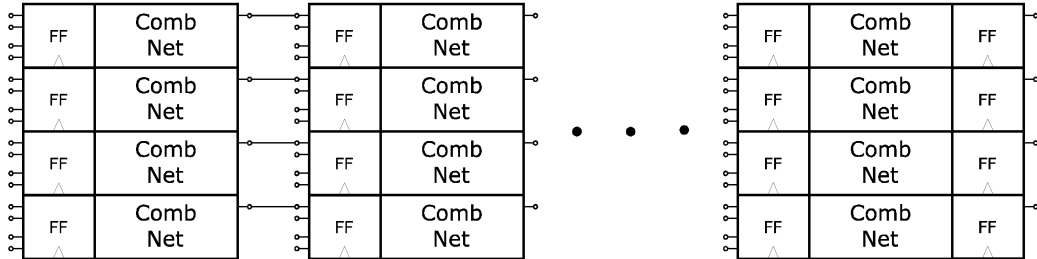


Figure 4.3: FF-Comb chain

A FF-Comb chain is once again a sequential circuit with 4 data inputs and 4 data outputs. Each data output is a logic function of the 4 data inputs. The FF-Comb chain has a latency of 21 clock cycles.

- Different FF-Comb chains may be obtained re-arranging the connections between FF-Comb blocks. In the uppermost level of the digital core, four of these different FF-Comb chains are placed. The digital core is thus a sequential circuit with 16 data inputs and 16 data outputs. Each data output is a logic function of just 4 out of the 16 data inputs. The digital core has a latency of 21 clock cycles.

The Flip-Flop constituting the registers are connected together in order to implement a scan chain, i.e. the Q pin of each Flip-Flop is connected with the *ScanIn* (SI) of the following Flip-Flop.

The scan chain gives two main advantages:

1. since the functionality of this circuit will be tested through a *load and clock* procedure, the input data can be loaded serially reducing the number of input pins from 16 to 2;
2. all the 20 stages of each path can be analysed and/or forced during measurement.

The signals *Clock* (CLK), *ScanEnable* (SE) and *Reset* are common to the Flip-Flops, and are distributed with a balanced tree. The external pins of the digital core are thus only 5: SI , CLK , SE , *Reset* and SO .

4.3.2 Core functionality

The digital core described above has 16 combinatorial and sequential paths. Each path is designed to have a different functionality, in order to test as much random functions as possible. As an example, in Table A.16 the Truth Table of the ninth path is reported.

IN<11:12>	IN<9:10>	00	01	11	10
00		1	0	1	1
01		0	1	0	1
11		0	0	0	0
10		0	1	0	1

Table 4.2: Truth Table of path 9: OUT<9>

In appendix A, the logic functionality of all the 16 paths is given.

4.4 Simulation and implementation

The circuit has to be simulated to verify both its behavioural function and its time enclosure. To fulfill the project requirements given in Section 4.2, the simulations of the digital core have been run in the two conditions described in Table 4.1.

Since the dimensions of the complete digital core are considerable, to test the digital functionality of all paths Verilog models are used. In particular, Verilog-AMS and Verilog-A models describing the standard cells are written. It is in this way faster to attain a behavioural verification of the complete core. In appendix B more details are given about VerilogA and VerilogAMS Hardware Description Languages.

Simulations proved the expected functionality of all the 16 paths to be the same for both testing states: Normal and Idle (see Table 4.1).

Since this experiment want to be a proving concept for a typical design, no modification of the design flow or semi-custom design are allowed. The design has to be realised in a full automated procedure, using standard synthesis tools. The adoption of Verilog models consented to use an automatic *Place & Route* utility for the layout implementation of the digital core.

Therefore, the layout of the core has been automatically generated using the Cadence software Encounter, starting from:

- the layout of the instantiated standard cells;
- the Verilog netlist of the core;
- the definition of the design constraints.

After the automatic *Place & Route*, some handcrafted layout modifications have been done to complete the back-end design of the chip, since the available design flow did not yet support all features needed for this latest technology node.

Chapter 5

Monitor

Among the blocks on the testchip, a lithography performance monitor is present. Section 5.1 illustrates the need of such kind of monitors to find out the impact of patterning aberrations. A possible monitor architecture is then proposed in Section 5.2; while two circuits to evaluate the monitor results are designed in Section 5.3. In Section 5.4 the description of the chosen implementation is then given.

The task of the monitor is to verify the presence of a systematical error introduced by pattern aberrations in an assigned structure. Therefore, the results are not evaluated quantifying the entity of the error, but rather measuring its occurrences in more than one circuit realisation on silicon.

5.1 Lithography aberrations

Photolithography is the common technique used to manufacture CMOS devices. The substrate is covered with a layer of photoresist. Then, using a patterned photo-mask, the wafer is exposed to light, so that the pattern of the photo-mask is transferred to the photoresist. Since the printing patterns have sub-wavelength dimensions, the photoresist pattern may differ from the photo-mask one, leading to a deterioration of the quality and of the yield of the fabricated ICs.

Thus, corrections, also called compensations, for the shortcomings in the masks are needed.

One of these methods is the optical proximity correction (OPC). It consists in distorting the pattern printed on the photo-mask so that the final result in the photoresist is closer to the desired design. Figure 5.1(a) shows how OPC is applied in the mask-definition process. The right mask pattern is used during lithography, to avoid roundings in the diffusion region, that may cause poor control on the gate width. Better imaging results can be achieved by using Sub Resolution Assist Features (SRAFs), such as scattering bars and hammerheads, which are not printed onto the wafer, but help to reduce resolution enhancement variations across the mask (Figure 5.1(b)).

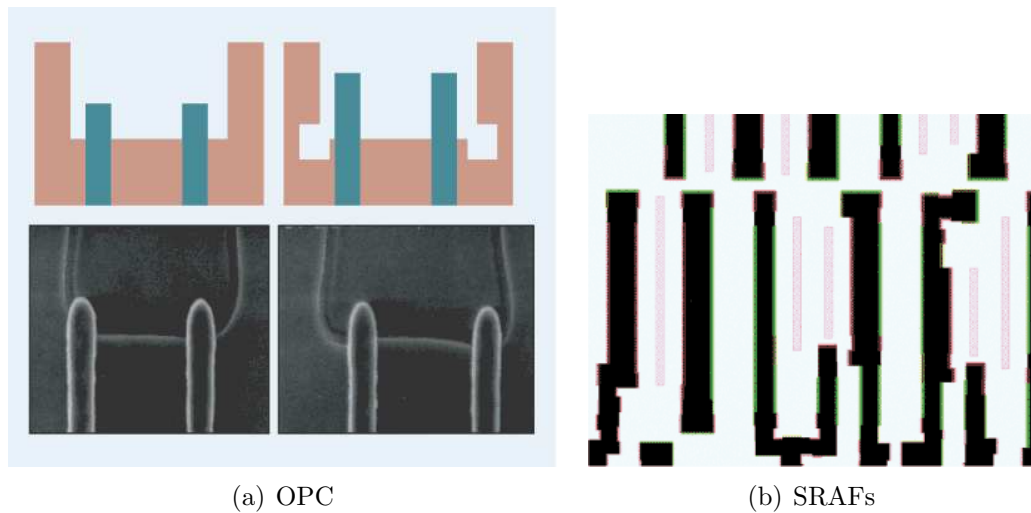


Figure 5.1: OPC and SRAFs applied in the mask definition process

In Section 1.1 alternative design approaches, such as DfM (*Design for Manufacturability*) and DfL (*Design for Lithography*), are discussed. Moreover, the definition of lithofriendliness is given.

With the shrinking of the node size, lithography is an increasing source of variability. The performance of both analog and digital designs may be affected by this kind of variations.

Analog designs make large use of transistor pairs (e.g. in differential amplifiers), whose functionality may be corrupted by lithography variations. However, the large dimensions of the transistors commonly used in these designs make analog circuits less sensitive to variations.

On the other hand, common digital designs use minimum dimension devices; memories present also positive feedback paths. Therefore the design functionality may result to be much more sensitive to slightly variations of the transistor dimensions.

Nonetheless, the influence of lithography on the operation and performance of digital circuits might be less clear and less visible. In order to decide how much lithofriendliness should be included in our libraries and designs, it would be valuable to be able to measure the influence of the lithography on the robustness at transistor level and to extrapolate this knowledge to a digital circuit.

5.2 Lithography monitor

A lithofriendly monitor is designed to study the variations of the channel length of MOS devices due to proximity effects. The information gained with this monitor may also give an indication of the lithographic robustness of the digital cell library and design.

In Figure 5.2 the embodiment of the monitor core is shown. It shows the parallel connection of three transistors, that ideally should be identical, controlled by the same gate voltage.

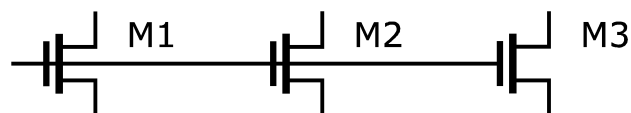


Figure 5.2: Embodiment of the monitor

Figure 5.3(a) shows how the layout should be if an ideal lithography would be available. Figure 5.3(b) displays instead what may be realised on silicon, due to non ideal imaging or printing during the lithographic process.

The layout of the three transistors explains how process variability could affect, in a different way, each device.

For our experiment, the transistors are dimensioned so that they become as sensitive as possible to these lithographic imperfections: i.e. the three

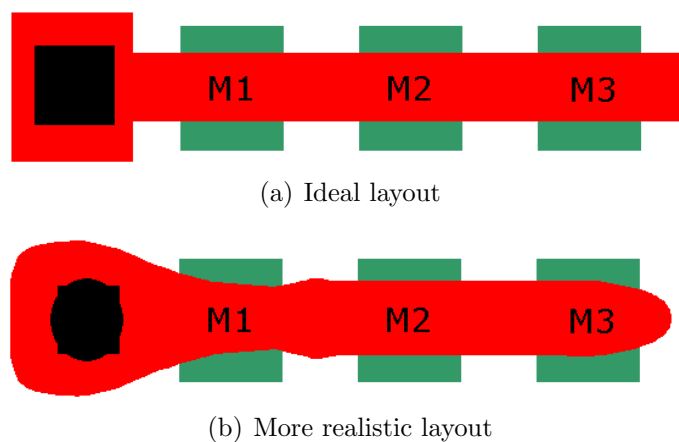


Figure 5.3: Layout of the monitor

transistors have minimum widths and lengths, and they are positioned at minimum spacings. Even if respecting the fab layout rules the realised transistors should be exactly alike, differences in the effective channel lengths of the three identically drawn transistors are expected.

The intent of this experiment is, in fact, to evaluate the robustness of the realisation process. If for minimum dimensioned transistors the layout rules do not guarantee equal performance of the three transistors above described, lithofriendly design is needed.

Due to the layout topology, transistor $M2$ should show the least lithographic imperfections and show the most homogeneous channel length. It is therefore used as a reference. Transistor $M1$ is expected to have an average channel length ($L1$) larger than transistor $M2$ ($L2$), due to its proximity to the contact. On the other side, transistor $M3$ is expected to have a reduced length due to the rounding of the poly line toward its end.

It is well known ([7], [17], [18], [19]) that the Drain-Source current of a nMOS in saturation region is given, in first approximation, by:

$$I_{DS} = \frac{\mu_n C_{ox} W}{2L} (V_{GS} - V_T)^2$$

Assuming that current dependence on the length can be expressed as the sum of two independent components, i.e. that the effect of the variation

of the dimension ratio can be observed separately from the variation of the threshold voltage of the transistors due to the length variation, then the variation of the current can be written as:

$$\begin{aligned}
\Delta I_{DS} &= \left. \frac{\partial I_{DS}}{\partial L} \right|_{V_{GS}-V_T=const} \cdot \partial L + \left. \frac{\partial I_{DS}}{\partial V_T} \right|_{L=const} \cdot \frac{\partial V_T}{\partial L} \cdot \partial L = \\
&= -\frac{I_{DS}}{L} \cdot \partial L + \left(-\frac{2I_{DS}}{V_{GS} - V_T} \right) \cdot \frac{\partial V_T}{\partial L} \cdot \partial L = \\
&= -\left(\frac{I_{DS}}{L} + \frac{2I_{DS}}{V_{GS} - V_T} \cdot \frac{\partial V_T}{\partial L} \right) \cdot \partial L = -\gamma \cdot \partial L \quad (5.1)
\end{aligned}$$

To derive the variation of the current as a function of the transistor length or the given technology, a nMOS is simulated. The nominal characteristics of a nMOS, as well as its DC operating values, are reported in Table 5.1.

W	150nm	L	40nm
V_{DS}	1.1 V	V_{GS}	0.6V
V_T	0.34V	I_{DS}	16.6 μ A

Table 5.1: Nominal characteristics and DC values

It results that for this transistor:

$$\left. \frac{\partial I_{DS}}{\partial L} \right|_{V_{GS}-V_T=const} = -\frac{I_{DS}}{L} = -0.42 \frac{\mu\text{A}}{\text{nm}} \quad (5.2)$$

$$\left. \frac{\partial I_{DS}}{\partial V_T} \right|_{L=const} = -\frac{2I_{DS}}{V_{GS} - V_T} = -127.7 \frac{\mu\text{A}}{\text{V}} \quad (5.3)$$

Assuming that the variation of the length due to the imperfection of the masks is limited, the threshold voltage of the transistor can be described as a linear function of the length, thus its derivative can be considered in first approximation a given value:

$$\frac{\partial V_T}{\partial L} = 1.92 \frac{\text{mV}}{\text{nm}} \quad (5.4)$$

Therefore it can be derived that:

$$\begin{aligned} \left. \frac{\Delta I_{DS}}{\Delta L} \right|_{nominal} &= -\gamma_{nom} = - \left(0.42 \frac{\mu A}{nm} + 127.7 \frac{\mu A}{V} \cdot 1.92 \frac{mV}{nm} \right) \\ &= -(0.42 + 0.25) \frac{\mu A}{nm} = -0.67 \frac{\mu A}{nm} \end{aligned} \quad (5.5)$$

The value of γ_{nom} , that is defined as positive, is very close to the simulated value:

$$\left. \frac{\Delta I_{DS}}{\Delta L} \right|_{simulated} = -0.63 \frac{\mu A}{nm}$$

Therefore, the variation of the drive current is proportional to the variation of the average channel length.

To evaluate the actual variations among length of the transistor lengths, a circuit capable to sense the difference of drive currents is needed.

5.3 Sensing circuit

Since mask imperfections are no deterministic effects, to gain knowledge on the statistical distribution of the variation, rather than a measurement on a single circuit, measurements on several theoretically equal circuits should be performed. Therefore, a large number of monitors is realised, and the result of each of them is evaluated with a Pass or Fail logic.

As a matter of fact, a statistical measurement gives a more reliable outcome than single entity measurements for the investigation of systematic errors. On the other side, the Pass or Fail approach obviously leads to a lower accuracy in the effective entity of the discrepancy, but it was chosen since it dramatically decreases the required number of pads of the testchip.

The main task of the monitor is to reveal a variation of the average channel length of a transistor with respect to a reference one. In Section 5.2 it has been discussed how a variation in the channel length affects the drive current

of a transistor. Therefore, a circuit capable to sense this current variation is realised.

Referring to Figure 5.2, to evaluate the presence of layout variations of the channel lengths between $M1$ and $M2$ (used as reference), and between $M2$ (used as reference) and $M3$, two sensing circuits are designed.

In Section 5.3.1, only the circuit for the comparison between $M1$ and $M2$ is discussed. At schematic level the two circuits differ only for the transistor under test, however in Section 5.3.3 a schematic description of the circuit for the comparison between $M3$ and $M2$ is given.

5.3.1 Circuit topology

To measure the difference of the driving currents between two non identical MOS transistors with shorted gates, the devices may be used as a differential pair or in a current mirror configuration (Figure 5.4).

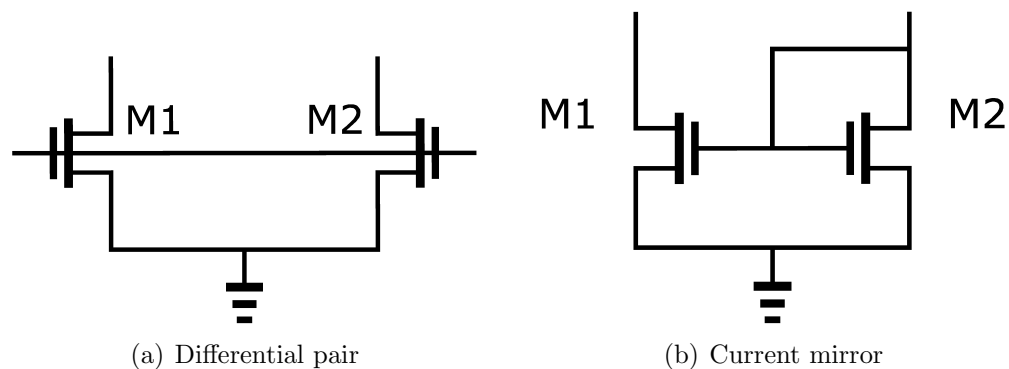


Figure 5.4: Possible configurations

These configurations could be used in the first stage of a differential amplifier. However, simulations proved that an amplifier with more than 2 stages is necessary to generate a sufficient gain to produce an acceptable output voltage level. For area and circuit simplicity, this solution is therefore not acceptable.

Since the gain is then crucial, an architecture giving an infinite gain is adopted (Figure 5.5). The two inverters constitute a latch. A latch is a

metastable system: in its instability condition the gain is infinite.

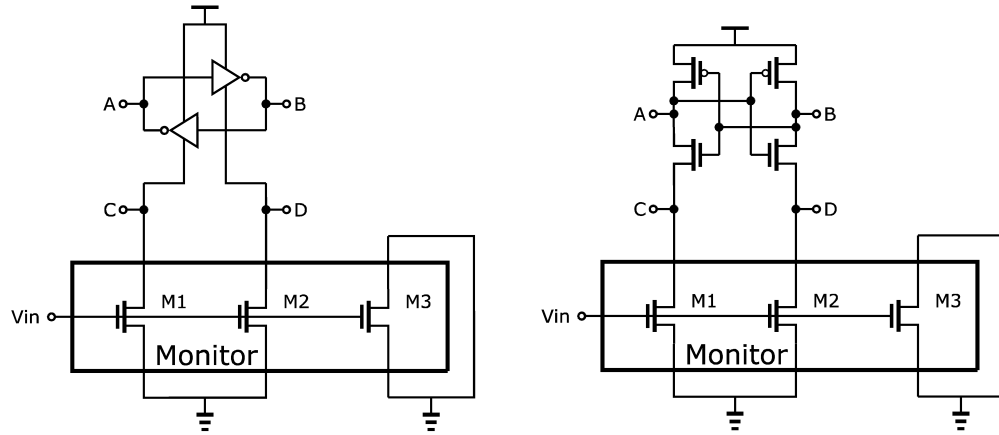


Figure 5.5: Proposed architecture

As described in Figure 5.5, the circuit reaches its stable condition as soon as the supply voltage is given. Therefore, an activation circuit is needed. In Figure 5.6 one nMOS and four pMOS with switch functionality are added to the above circuit.

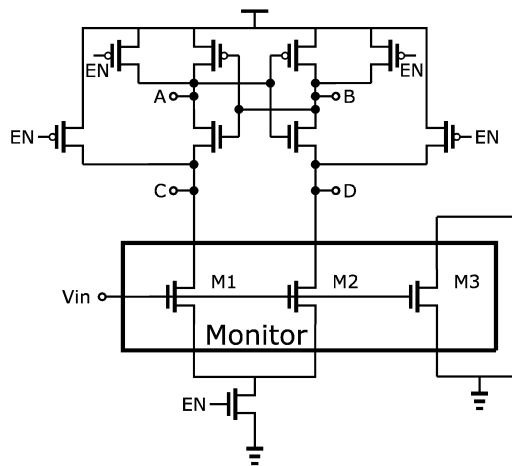


Figure 5.6: Proposed architecture with switches

5.3.2 Working principle

The working principle of the architecture displayed in Figure 5.6 is quite simple:

- When the EN signal is low, the nMOS switch is open, so that the sources of the monitor transistors are floating. In the meantime, the pMOS switches are closed, and the nodes A , B , C and D are approximately at V_{DD} . The system is off.
- When the EN signal goes high, the nMOS switch is closed, connecting the sources of the monitor transistor to ground. Moreover, the pMOS switches are open, and the system enters its instability condition. If the two inverters constituting the latch are identical, the stability condition is determined by the monitor transistors, $M1$ and $M2$.
- Once the stability condition is reached, the latch is in idle, and the total power consumption is given by the leakage power only. At this stage, only a commutation of the EN signal can restore the instability.

To evaluate how the mismatch between the monitor transistors under test ($M1$ and $M2$) affects the output nodes voltages, a transient analysis of the instability condition is needed. Two fundamental hypothesis in this analysis are assumed:

1. the 2 inverters constituting the latch are identical;
2. the 2 transistors under test are identical, unless for their average channel length L .

When the nMOS used as a footer closes, due to a low→high transition of the EN signal, both $M1$ and $M2$ start to drive current. In the meantime, the pMOS switches become open, and the A , B , C and D nodes voltages are no more fixed to V_{DD} , but they are free to vary.

Assuming $L2 = L_{nom}$ and $L1 = L_{nom} + \Delta L$,

- if $\Delta L > 0$, then (see equation 5.5):

$$I_{DS1} = I_{DSnom} + \Delta I_{DS1} = I_{DSnom} \cdot [1 - \gamma_{nom} \cdot \Delta L]$$

Therefore $M2$ drives more current than $M1$, and the voltage of the node D drops more rapidly than the node C voltage. The same thing happens respectively for nodes B and A .

Since node B reaches the logic value '0' before node A , the latch falls in the stability condition: with $B = 0$ and consequently $A = 1$ (Figure 5.7(a)).

- If $\Delta L < 0$, and thus $L1$ is smaller than $L2$:

$$I_{DS1} = I_{DSnom} + \Delta I_{DS1} = I_{DSnom} \cdot [1 + \gamma_{nom} \cdot |\Delta L|]$$

Therefore, the circuit has the opposite behaviour, and reaches the opposite ability condition: $B = 1$ and consequently $A = 0$ (Figure 5.7(b)).

- If $\Delta L = 0$, $I_{DS1} = I_{DS2} = I_{DSnom}$. Thus, the voltage of the nodes C and D drops simultaneously, the system enters in metastability and the reached final stability condition is stochastic.

Figure 5.7 shows the transient simulations of the circuit with $V_{DD} = 1.1$ V, $L_{nom} = 40$ nm, $W_{nom} = 150$ nm and $\Delta L = \pm 0.05$ nm.

Schematic simulations proved the monitor to be sensitive to very small variations of the average channel length. On the other hand, to make the two discussed hypothesis true, a specific layout strategy is followed (see Section 5.3.4).

5.3.3 Input voltage generation

To generate the V_{in} voltage, chosen to be $\frac{V_{DD}}{2}$, 2 diode-connected identical pMOS transistors are added to the circuit (Figure 5.8).

The two pMOS are in series, and thus they drive the same current. Therefore, the V_{in} generated in this way reaches the $\frac{V_{DD}}{2}$ value.

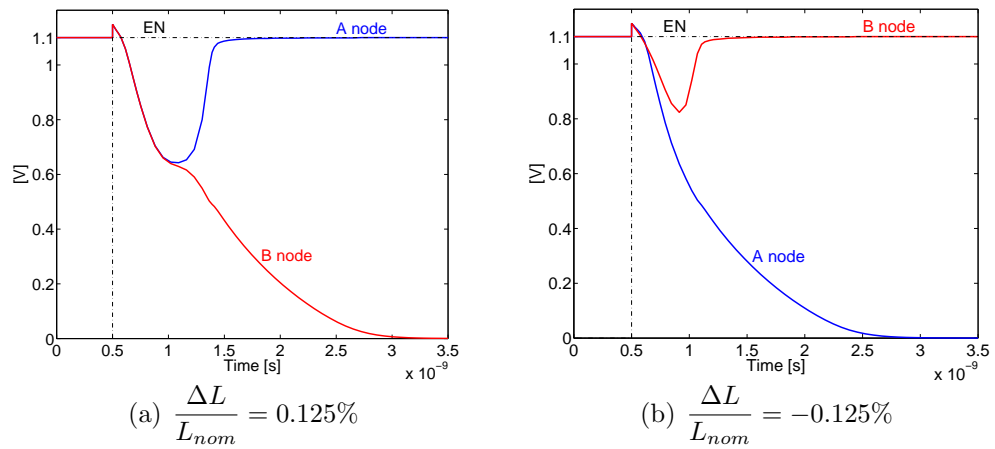
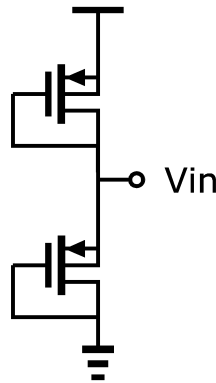


Figure 5.7: Transient behaviour

Figure 5.8: V_{in} generator

To ensure the same V_T , the two pMOSFETs are realised in two different *n-wells*, so that their $V_{BS} = 0$.

For the comparison between the monitor transistors $M2$ and $M3$, a sensing circuit similar to the one described in Figure 5.6 is designed (Figure 5.9).

Also this circuit makes use of an input voltage generator designed as the one in Figure 5.8.

5.3.4 Layout realisation

In Section 5.3.2 two fundamental hypothesis for the functioning of the circuit are given:

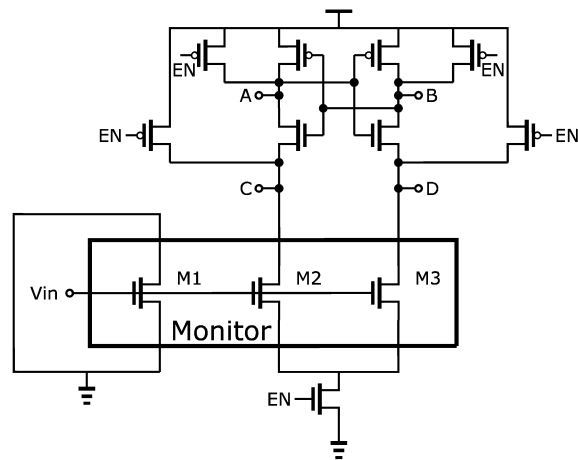


Figure 5.9: Proposed architecture with switches

1. the 2 inverters constituting the latch are identical;
2. the 2 transistors under test are identical, unless for their average channel length L .

To make them true, common matching strategies [20] are used:

- the transistors are drawn in order to match if translated;
- all transistors have the same orientation;
- the transistors constituting the latch are not at minimum dimensions ($W_n = 600$ nm, $L_n = 100$ nm, $W_p = 880$ nm, $L_p = 100$ nm);
- the matching transistors are drawn as close as allowed by the layout rules;
- the transistors constituting the latch have a *common centroid* disposition.

In Figure 5.10(a) the layout realisation of the latch transistors is displayed: the *common centroid* disposition is clearly visible. In Figure 5.10(b) the complete layout of the monitor cell is shown: the monitor cell includes the monitor itself, the sensing circuit, and the V_{in} generator.

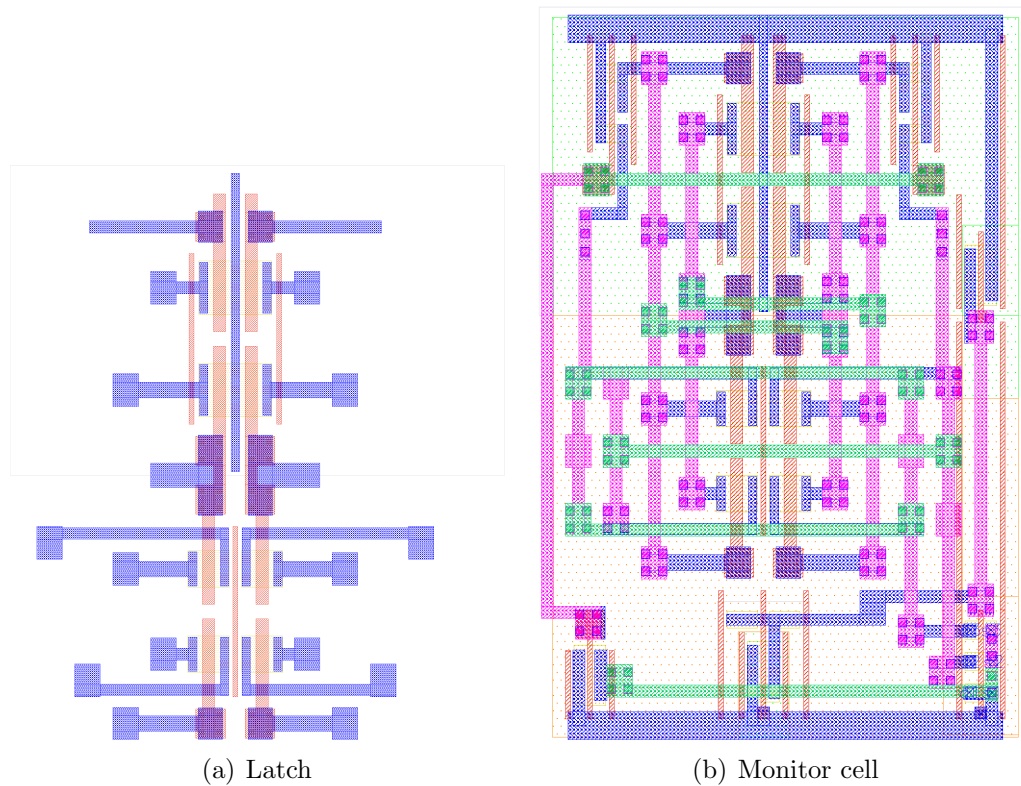


Figure 5.10: Layout realisations

These matching strategies consented to fulfill the above discussed hypothesis. Moreover, post layout and Monte Carlo simulations were run to prove the functionality of the sensing circuit taking into account the mismatch and the process variability. The following results are obtained assigning the channel length of transistor $M1$ as a parameter, and not actually making it dependent on the layout.

Comparing the transient behaviour of the schematic extracted from the layout (Figure 5.11) with the original one (Figure 5.7), it is proved that the implemented layout does not affect the circuit functionality. For the same variation of the channel length $L1$, indeed, the A and B node voltages have similar trend.

On the other hand, the response time is longer in post layout simulations, due to the parasitic elements now taken into account. In schematic simula-

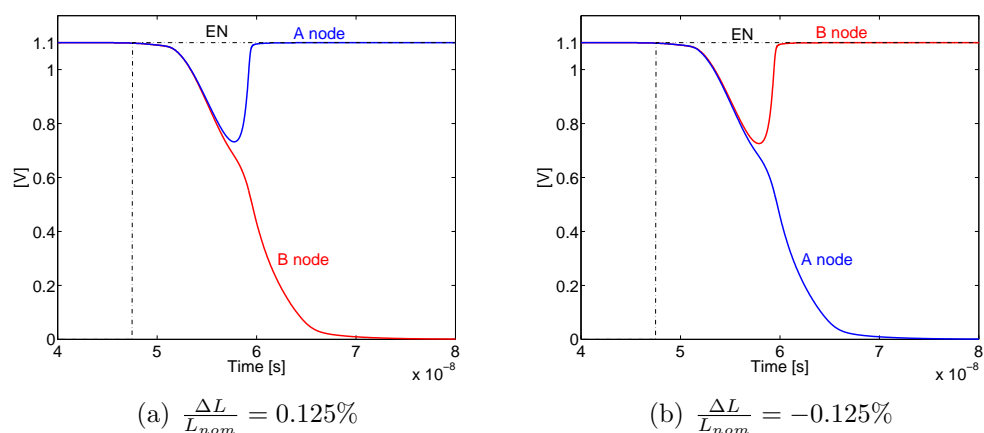


Figure 5.11: Transient behaviour in post layout simulations

tions the time needed to reach the stability condition is about 2.5 ns, while in post layout simulations it is about 22.5 ns.

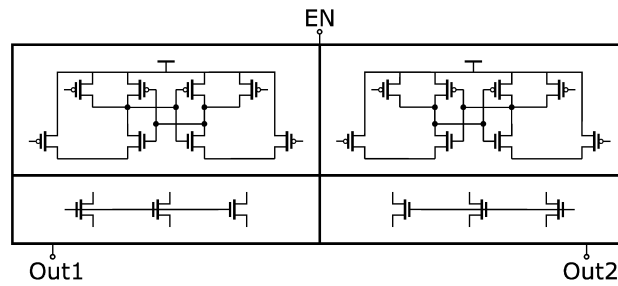
Nonetheless, the set up time is not a concern for the measurement, since once the stability condition is reached only a commutation of the EN signal can modify the output state.

Post layout Monte Carlo simulations proved that in approximately 98% of the trials the mismatch and process variations do not affect the monitor functionality, that results therefore almost only sensitive to the transistor length.

5.4 Monitor block implementation

In Section 5.3 the choice of a statistical measurement is discussed. In order to achieve a statistical significance, a large number of monitor cells is realised on silicon, constituting the *Monitor Block*.

The basic element of the *Monitor Block* is given by two monitors, one for the comparison between $M1$ and $M2$ and the other one for the comparison between $M2$ and $M3$, and their sensing circuit. The two monitors are realised as close as possible (approximately at a distance of 100 nm), so that they may be considered almost identical. A block description of the basic *Monitor*

Figure 5.12: Basic element of the *Monitor Block*

Block element is displayed in Figure 5.12.

The *Monitor Block* itself is then given up by 400 of these basic elements. For for the result caption a scan chain is implemented. Therefore, only 5 pads are needed in the test chip to give the stimuli and to probe the outputs of the *Monitor Block*: namely the *EN* signal, and the *ScanIn*, *ScanEnable*, *ScanOut* and *clk* signal for the scan chain. A block description of the *Monitor Block* is given in Figure 5.13.

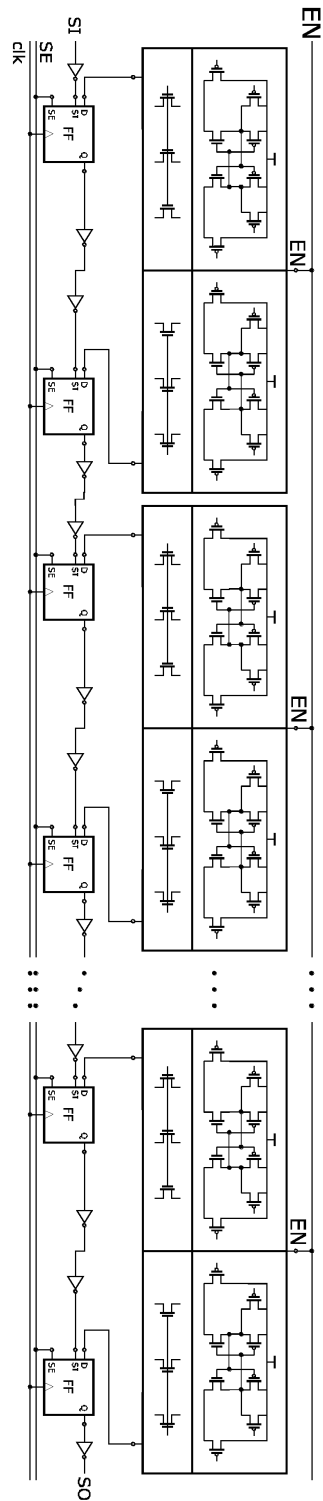


Figure 5.13: *Monitor Block* with its 5 pins: *EN*, *SI*, *SE*, *clk* and *SO*

Appendix A

Digital Core Truth Tables

In the following, the truth tables of the 16 combinatorial and sequential paths constituting the Digital Core (see Chapter 4) are displayed.

IN<3:4> \ IN<1:2>	00	01	11	10
00	0	1	0	0
01	1	0	1	0
11	1	0	0	0
01	0	1	0	0

Table A.1: Truth Table of path 1: OUT<1>

IN<3:4> \ IN<1:2>	00	01	11	10
00	0	1	0	0
01	1	0	1	0
11	0	1	0	0
01	0	0	0	0

Table A.2: Truth Table of path 2: OUT<2>

IN<3:4> \ IN<1:2>	00	01	11	10
00	1	0	1	0
01	0	1	0	0
11	0	0	0	0
01	0	0	0	0

Table A.3: Truth Table of path 3: OUT<3>

IN<3:4> \ IN<1:2>	00	01	11	10
00	0	1	0	0
01	0	0	1	0
11	0	1	0	0
01	0	0	0	0

Table A.4: Truth Table of path 4: OUT<4>

IN<7:8> \ IN<5:6>	00	01	11	10
00	1	0	1	1
01	0	1	1	1
11	0	0	0	0
01	0	1	0	0

Table A.5: Truth Table of path 5: OUT<5>

IN<7:8> \ IN<5:6>	00	01	11	10
00	0	1	0	0
01	1	0	1	0
11	0	1	0	0
01	0	0	0	1

Table A.6: Truth Table of path 6: OUT<6>

IN<7:8> \ IN<5:6>	00	01	11	10
00	1	0	1	0
01	0	1	0	0
11	0	0	0	0
01	0	0	0	1

Table A.7: Truth Table of path 7: OUT<7>

IN<7:8> \ IN<5:6>	00	01	11	10
00	0	1	0	0
01	0	0	1	0
11	0	1	0	0
01	0	0	0	1

Table A.8: Truth Table of path 8: OUT<8>

IN<11:12> \ IN<9:10>	00	01	11	10
00	1	0	1	1
01	0	1	0	1
11	0	0	0	0
10	0	1	0	1

Table A.9: Truth Table of path 9: OUT<9>

IN<11:12> \ IN<9:10>	00	01	11	10
00	0	1	0	0
01	1	1	1	0
11	1	1	0	0
01	0	0	0	0

Table A.10: Truth Table of path 10: OUT<10>

IN<11:12> \ IN<9:10>	00	01	11	10
00	1	0	1	0
01	0	1	0	0
11	1	0	0	0
01	0	0	0	0

Table A.11: Truth Table of path 11: OUT<11>

IN<11:12> \ IN<9:10>	00	01	11	10
00	0	1	0	0
01	0	0	1	0
11	1	1	0	0
01	0	0	0	1

Table A.12: Truth Table of path 12: OUT<12>

IN<15:16> \ IN<13:14>	00	01	11	10
00	0	1	0	1
01	1	1	1	1
11	0	1	0	0
01	0	1	0	1

Table A.13: Truth Table of path 13: OUT<13>

IN<15:16> \ IN<13:14>	00	01	11	10
00	0	0	0	1
01	1	1	0	1
11	1	0	0	0
01	0	1	0	0

Table A.14: Truth Table of path 14: OUT<14>

IN<15:16> \ IN<13:14>	00	01	11	10
00	0	1	0	0
01	1	1	0	1
11	1	0	0	0
01	0	1	0	0

Table A.15: Truth Table of path 15: OUT<15>

IN<15:16> \ IN<13:14>	00	01	11	10
00	0	0	0	1
01	0	0	0	0
11	1	0	0	0
01	0	0	0	1

Table A.16: Truth Table of path 16: OUT<16>

Appendix B

Verilog-AMS and Verilog-A

B.1 Verilog

Verilog (acronym for Verify Logic) is a Hardware Description Language (HDL) used to model electronic systems. The language supports the design, verification, and implementation of analog, digital, and mixed-signal circuits at various levels of abstraction [21].

The main difference between hardware description languages and conventional programming language in the execution of statements, that in HDLs is not strictly sequential. A Verilog design consists of a hierarchy of modules. Modules are defined with a set of input, output, and bidirectional ports. Internally, a module contains a list of wires and registers. Concurrent and sequential statements define the behaviour of the module by defining the relationships between the ports, wires, and registers. Sequential statements are placed inside a special sections (begin/end) and executed in sequential order within the block. All concurrent statements and all begin/end blocks in the design are executed in parallel.

B.1.1 Verilog-AMS

Verilog-AMS is a derivative of Verilog. It includes Analog and Mixed-Signal extensions (AMS) in order to define the behaviour of analog and mixed-signal systems [22].

The aim of Verilog-AMS standard is to create and use modules that encapsulate high-level behavioural descriptions as well as structural descriptions of systems and components. Verilog-AMS is an industry standard modeling language for mixed signal circuits. It provides both continuous-time and event-driven modeling semantics, and so is suitable for analog, digital, and mixed analog/digital circuits. It is particularly well suited for verification of analog, mixed-signal and RF integrated circuits.

B.1.2 Verilog-A

Verilog-A is an industry standard modeling language for analog circuits. It is the continuous-time subset of Verilog-AMS.

Verilog-A was created out of a need to standardise the Spectre behavioural language in face of competition from VHDL, which was absorbing analog capability from other languages (e.g. MAST). Verilog-A is an all-analog subset of Verilog-AMS.

Unfortunately, the original goal of a single language standard is still to be achieved.

B.2 Project applications

In the project described in this work, a large use of HDLs has been done. Given the large dimensions of the circuits implemented in the testchip (see Section 1.3), Verilog-AMS and Verilog-A are used for the verification of several digital and mixed-signal circuits. The testchip blocks described and verified with the HDLs are:

- Main Cores:
 1. Ring oscillators;
 2. Multiplexer (Chapter 2);
 3. Selector (Chapter 3).

- Digital Core (Chapter 4).

The main benefit given by this choice is the possibility to verify the functionality of simple circuits as long as more complex ones.

B.2.1 Hierarchy

At testchip level design, a hierarchical description of the above mentioned cores is given. For several of the hierarchical levels, a Verilog description is present. The further the hierarchy is descended, the more detailed the Verilog descriptions become. This strategy is adopted in order to have, alongside the hierarchical design, a hierarchical verification both at schematic and at layout level.

B.2.2 Verification

Moreover, using Analog and Analog Mixed-Signal Verilog extensions, it is possible to simulate a circuit whose blocks are defined partly with behavioural modules or described at MOSFET level. It is therefore possible to properly evaluate the functionality of sub-circuits described at netlist level while the other blocks of the cores are defined in a high-level language. This opportunity is beneficial for two aspects:

- during design, blocks may be verified even before the entire system is completed;
- during final verification, the complexity of the simulation can be reduced, excluding the netlist description of non critical blocks.

The simulator used both for the Verilog entities and for the schematic circuits is Spectre.

Bibliography

- [1] K. Preston White, Jr., R.N. Athay, and W.J. Trybula. Applying dfm in the semiconductor industry. In *Electronics Manufacturing Technology Symposium. 'Manufacturing Technologies - Present and Future', Seventeenth IEEE/CPMT International*, pages 438–441, 2-4 Oct 1995.
- [2] T. C. Kuo and Hong-Chao Zhang. Design for manufacturability and design for x: Concepts, applications, and perspectives. In *IEEE/CPMT International Electronic Manufacturing Technology (IEMT) Symposium*, 1995.
- [3] R. Wilcox, T. Forhan, G. Starkey, and D. Turner. Design for manufacturability: a key to semiconductor manufacturing excellence. In *Advanced Semiconductor Manufacturing Conference and Workshop (IEEE/SEMI)*, pages 308–313, 23-25 Sep 1998.
- [4] L. Riviere-Cazaux, K. Lucas, and J. Fitch. Integration of design for manufacturability (dfm) practices in design flows. In *Sixth International Symposium on Quality of Electronic Design (ISQED)*, pages 102–106, 21-23 March 2005.
- [5] M. Garg, A. Kumar, J. van Wingerden, and L. Le Cam. Litho-driven layouts for reducing performance variability. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 4, pages 3551– 3554, 2005.
- [6] Y. Tsividis. *Operation and Modeling of Mos Transistor*. McGraw-Hill, 1988.

-
- [7] K.M. Laker and W.M.C. Sansen. *Design of analog integrated circuits and systems*. McGraw-Hill, 1994.
- [8] H. Soeleman and K. Roy. Ultra-low power digital subthreshold logic circuits. In *Proceedings of the International Symposium on Low Power Electronics and Design*, pages 94–96, 1999.
- [9] H. Soeleman, K. Roy, and B.C. Paul. Robust subthreshold logic for ultra-low power operation. In *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, volume 9, pages 90–99, Feb 2001.
- [10] B.C. Paul, A. Raychowdhury, and K. Roy. Device optimization for digital subthreshold logic operation. In *IEEE Transactions on Electron devices*, volume 52, pages 237–247, Feb 2005.
- [11] Jinhui Chen, L.T. Clark, and Yu Cao. Robust design of high fan-in/out subthreshold circuits. In *Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers and Processors (ICCD)*, pages 405–410, 2-5 Oct. 2005.
- [12] J. Hailin and S.R. Marek-Sadowska, M.; Nassif. Benefits and costs of power-gating technique. In *Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers and Processors (ICCD)*, pages 559–566, 2-5 Oct. 2005.
- [13] J.T. Kao, A.P. Chandrakasan, and D. Antoniadis. Transistor sizing issues and tool for multi-threshold cmos technology. In *Proceedings of the 34th Design Automation Conference (DAC)*, pages 409–414, 9-13 Jun 1997.
- [14] J.T. Kao and A.P. Chandrakasan. Dual-threshold voltage techniques for low-power digital circuits. In *IEEE Journal of Solid-State Circuits*, volume 35, pages 1009–1018, Jul 2000.

-
- [15] B. Calhoun and A.P. Chandrakasan. Standby voltage scaling for reduced power. In *Proceedings of the IEEE Custom Integrated Circuits Conference*, pages 639–642, 21-24 Sept. 2003.
- [16] Gang Qu. What is the limit of energy saving by dynamic voltage scaling? In *IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, pages 560–563, 2001.
- [17] P.R. Gray and R.G. Meyer. *Analysis and design of analog integrated circuits*. John Wiley & Sons, Inc., fourth edition, 2001.
- [18] J. Millman and A. Grabel. *Microelectronics*. McGraw-Hill, second edition, 1994.
- [19] B. Razavi. *Design of analog CMOS integrated circuits*. McGraw-Hill, second edition, 2001.
- [20] A. Hastings. *The art of analog layout*. Prentice Hall, second edition, 2001.
- [21] D.E. Thomas and P.R. Moorby. *The Verilog Hardware Description Language*. Springer, fifth edition, 2002.
- [22] K.S. Kundert and O. Zinke. *The Designer's Guide to Verilog-AMS*. Springer, first edition, 2004.