PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADES DE ENGENHARIA E INFORMÁTICA
ENGENHARIA DE COMPUTAÇÃO

RICARDO AQUINO GUAZZELLI

VOLTAGE SCALING EFFECTS ON NCL CELLS:
Analysis and Characterization

Porto Alegre
2014

RICARDO AQUINO GUAZZELLI

VOLTAGE SCALING EFFECTS ON NCL CELLS:
Analysis and Characterization

End of Term work presented as part of the activities to obtain a degree of Computer Engineering at the Faculty of Engineering in the Pontifical Catholic University of Rio Grande do Sul.

Advisor: Fernando Gehm Moraes
Co-Advisor: Matheus Trevisan Moreira

Porto Alegre
2014
RICARDO AQUINO GUAZZELLI

ABSTRACT

Voltage scaling is an attractive way to achieve drasticly lower power consumption when performance is not a main constraint. This technique has been satisfying emerging applications such as sensor network and medical applications. However, current standard cell libraries in the literature disregard the possibility to combine voltage scaling with asynchronous circuits. Asynchronous design allows much more relaxed timing assumptions than synchronous designs and, consequently, may improve the operation of ultra-low-power devices. In this work, an experimental environment is developed to show the benefits and drawbacks of NCL gates operating in near/sub-threshold region. Moreover, a 8-bit Kogge-Stone full adder employing NCL gates is implemented in order to present a circuit level evaluation. The obtained results point out that for a 65n CMOS technology with $V_{th} = 0.48V$ the near-threshold region achieves higher delay/power efficiency, whereas the sub-threshold region impacts heavily on the performance of the cells.

LIST OF ABBREVIATIONS

ASIC	Application Specific Integrated Circuit
CAD	Computer Aided Design
CMOS	Complementary Metal-Oxide-Semiconductor
DI	Delay Insensitive
DIBL	Drain-Induced Barrier Lowering
DRC	Design Rule Checking
EDA	Electronic Design Automation
EDP	Energy-Delay Product
F04	Fan-out of Four
GAPH	Grupo de Apoio ao Projeto de Hardware
INWE	Inverse Narrow Width Effect
LDP	Leakage-Delay Product
LVS	Layout Versus Schematic
MIT	Massachusetts Institute of Technology
MOS	Metal-Oxide Semiconductor
NCL	Null Convention Logic
PCHB	Pre-Charged Half-Buffer
PDN	Pull-Down Network
PUN	Pull-Up Network
PVT	Process, Voltage and Temperature
QDI	Quasi Delay Insensitive
RSCE	Reverse Short Channel Effect
RTO	Return to One
RTZ	Return to Zero
SNM	Static Noise Margin
SoC	System on a Chip
STA	Static Timing Analysis

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

Advances in semiconductor processing allowed reducing the minimum feature size of transistors and wires in the last years. As a consequence, it is possible to significantly increase the number of transistors in a single chip. A good example of this increase is the System on a Chip (SoC) used in the Xbox One console, which contains five billion transistors [SELL14]. Moreover, higher performance and lower power features can be obtained, due to the lower resistances and capacitances from smaller transistors and wires [BEE10]. However, these advances bring challenges to circuit and computer aided design (CAD) tools designers. As the transistor feature size reaches fundamental atomic limits, they gradually behave less and less as an ideal switch and wires behave less and less as ideal electrical connections with negligible impedance [ITR11]. In addition, increased manufacturing parameter variation brings uncertainties to the processes of estimating/predicting the timing and power characteristics of circuits [ITR11].

Nowadays, the predominant digital circuit design style is synchronous. This style takes as a fundamental assumption that all components share a common and discrete notion of time, which is specified by a global clock signal distributed throughout the circuit. This signal controls every sequential element, typically flip-flops and latches. The value stored in these elements can only change when the clock signal switches its logic level in a given direction. This is what enables sequential blocks to perceive time as a discrete variable, allowing data flow from one register to the next as the clock signal switches. Such characteristic enables the designer to ignore wire and logic gates delays, as long as the worst case delay between two registers is never longer than the period of the clock signal controlling them [RAB03]. However, despite the fact that synchronous designs simplify the design process, they also bring challenges with regard to clock distribution, skew and power consumption. Accordingly, variations on manufacturing process operating conditions result in substantial variations on the electrical characteristics of devices, which result in delay and power variations [MAR11] [RAH12]. To cope with these problems, synchronous designs require margins in the period of the clock signal, which lead to costs in performance, power, area and design time [KIM13]

Asynchronous circuits are an alternative to overcome such design issues faced by contemporary synchronous designers. Unlike the synchronous paradigm's main assumption, these circuits do not rely on a discrete notion of time. In this way, the global clock signal is removed and local handshaking control blocks are added between adjacent storage elements, establishing the synchronization, communication and sequencing of operations [SPA01]. Figure 1 exemplifies the structural difference between the synchronous and asynchronous paradigms, using a linear pipeline case study. In the former, Figure 1(a), a clock signal *CLK*, common to all registers, controls the sequencing of events. At each pulse of this signal, the registers copy the values at their inputs to their outputs, *i.e.* data propagates one stage. For the asynchronous example in Figure 1(b), data propagates through local communication between each pair of registers and is controlled by the *CTRL* blocks. This fundamental assumption of local synchronization enables avoiding clock problems [SPA01].

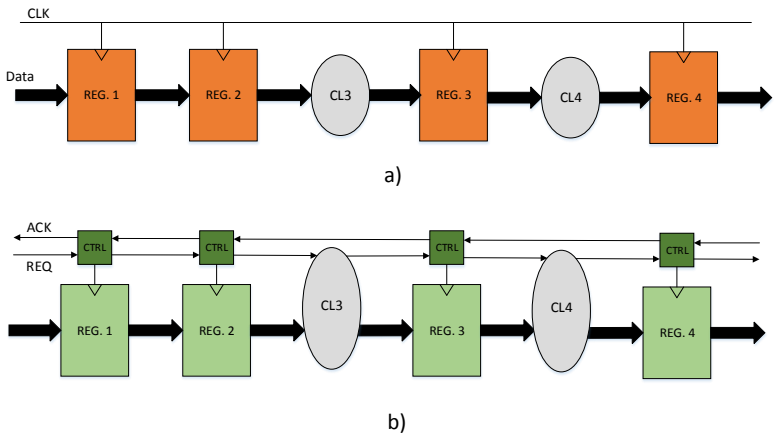


Figure 1 – Two basic alternative design styles for digital circuits: (a) a synchronous circuit; (b) an asynchronous circuit. Each CLi represents a combinational logic block, REG.i symbolize registers, and CTRL indicates control logic. Adapted from [SPA01].

Among the constraints imposed by contemporary designs, power consumption is an issue that has gained growing attention. Emerging applications, such as distributed sensor networks and wearable devices, define low energy consumption as a fundamental factor of the circuit functionality [CAL05]. In addition, with decreasing feature size, transistors have become increasingly leaky, augmenting static power dissipation, and have been challenging designers to meet power constraints [BEE10]. This has motivated the research on design techniques for minimizing energy or power for a given performance constraint. These efforts usually focus on high performance strong inversion operation (super-threshold region) and are implemented at the architectural level, where designers can reduce the computation workload or improve the architecture to achieve better power optimizations [IEO13]. At the circuit level, a compelling approach to lower power consumption is reducing the voltage supply, which is also called voltage scaling (VS). As the supply voltage is quadratic ally related to the dynamic power, VS is a promising low power technique [CHA10]. Taking this design option to the extreme, some low power systems have been created to operate in the sub-threshold region of transistors [LOT11]. This approach can allow the circuit to operate at very low voltage levels.

This work intends to contribute to the state of the art in asynchronous circuits design by exploring and characterizing a specific template for asynchronous design, namely Null Convention Logic (NCL) [FAN96], for different voltage levels. The work encompasses super-, near- and sub-threshold regions. NCL is a particularly interesting asynchronous template because it allows easier timing closure compared to other approaches and has been explored in many works available in the state of the art, such as [JEO08][FOL05][KOM02][LIG00][PAR14][REE12][MOR13d][MOR13b][MOR14a]. This analysis will provide the basis to further optimize NCL designs in sub-threshold operation.

1.1 Motivation

Applications focused on low energy consumption have motivated the research for energy-aware design techniques [CAL05] [LOT11]. Near- and sub-threshold operation can achieve drastic power savings, but this can only be used when speed is not a primary constraint. These approaches have shown good results to reduce power consumption in digital circuits [LOT11] [LIU12] [ZHO12]. Moreover, NCL asynchronous circuits are naturally suited to sub-threshold operation, given that they allow much more relaxed timing assumptions than synchronous designs [FAN96] [CHA10]. However, ultra-low power standard cell libraries proposed to date

are not capable to support asynchronous circuit projects [IEO13] [PNS13]. This End of Term work intends to contribute to fill this gap, and is circumscribed on the research developed by the Hardware Design Support Group (GAPH) at PUCRS, which has worked for a long time on several aspects of asynchronous circuit research. In fact, the group has a library called ASCEnD, which contains hundreds of asynchronous components [MOR11a][MOR11b][MOR14b], among which a large set of NCL gates is available. However, these gates were always used for designs that targeted nominal operating voltage. The first step to enable the usage of this library for voltage scaling applications is to understand and characterize its gates accordingly.

1.2 Contributions of this Work

This End of Term work provides a voltage scaling analysis using NCL gates. Device, gate and circuit evaluations are presented. For a better context introduction, the Author also presents the current state of near/sub-threshold and asynchronous standard cell libraries, as well as an initial voltage scaling analysis of basic CMOS gates.

1.3 Work Structure

This End of Term work is structured in the seven Sections. The introduction provides an initial context regarding the current state of CMOS technology, as well as benefits and challenges with emerging technologies and applications. The section entitled Concepts indicates basic points that support the presented work. The section called State of the Art gathers information about current near/sub-threshold and asynchronous standard cell libraries. Voltage Scaling on CMOS Cells is focused in the voltage scaling analysis of basic CMOS cells, while Voltage Scaling on NCL Cells Section aims to the experimental environment developed and obtained results. A Full-Adder Case Study shows a circuit level analysis of an 8 bits Kogge-Stone adder employing NCL gates. At last, Conclusions presents important results obtained in the work, as well as the projected future works are presented.

2. CONCEPTS

This Chapter presents a set of basic concepts about low-power design techniques and asynchronous circuits, which ease the discussion throughout this work.

2.1 Cell Characteristics

One of the most used approaches for designing integrated circuits (ICs) is the semi-custom standard-cell design style. This approach achieves significant reduction in fabrication costs and time due to the fact that designers rely on predesigned and pre-validated standard cells and electronic design automation (EDA) tools to construct complex circuits, avoiding the need of design each transistor from scratch. Standard cell libraries are usually provided by vendors and contain a wide number of logic gates with a range of fan-in and fan-out with different physical, electrical and logical characteristics. Each of the basic gates of a standard cell library is called a standard-cell, or simply cell. To allow EDA tools to use these cells, one must have their timing and power characteristics available. These characteristics enable the tool to perform logic optimizations and to meet timing and power constraints. This section details such characteristics and explores which are relevant for NCL cells.

2.1.1 Dynamic Switching Power

In CMOS circuits, the power dissipated while charging or discharging capacitances is denominated dynamic switching power [RAB03]. The dynamic switching power is dependent on the supply voltage V_{dd}, the switching frequency f_s, the initial and final voltages and the equivalent capacitance of a node [KUR04]. Also, a CMOS gate can be represented by a generic block diagram that is illustrated at Figure 2. Note that the CMOS gate is driving an output capacitor C_L. This output capacitor C_L represents the output load of the gate, which is created by different sources, such as drain-to-body junction, interconnection and gate oxide capacitances.

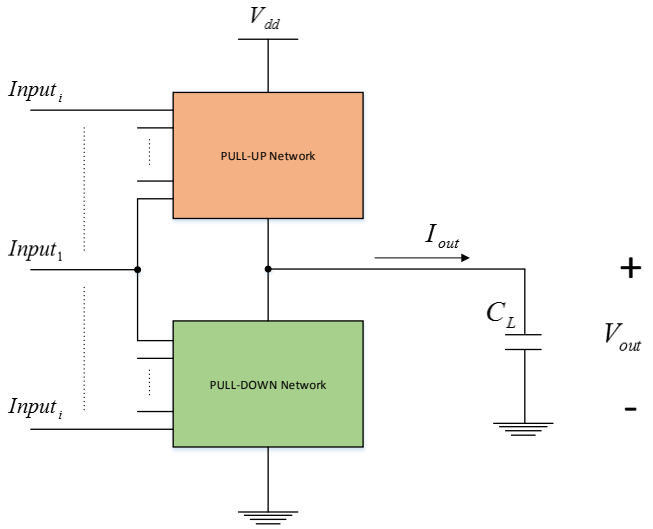


Figure 2 - Generic diagram of a CMOS gate. Adapted from [KUR04].

For instance, if a low-to-high transition occurs at the output node, the pull-up network is enabled, while the pull-down network is disabled. During this transition, the voltage supply V_{dd} provides a portion of current I_{out} that passes through the pull-up transistor and charges C_L. The power $P(t)$ provided from the power supply to charge C_L is

$$P(t) = I_{out}(t).V_{dd} \tag{1}$$

$$I_{out}(t) = C_L.\frac{dV_{out}(t)}{dt} \tag{2}$$

where V_{out} is the voltage at the output. The energy provided from the power supply for a $V_1 \rightarrow V_2$ transition $E_{V_1 \rightarrow V_2}$ at the output node is

$$E_{V_1 \rightarrow V_2} = \int_{t1}^{t2} P(t)dt = V_{DD}\int_{t1}^{t2} I_{out}(t)dt = C_L.V_{DD}\int_{V_1}^{V_2} dV_{out}(t) = C_L.V_{DD}.(V_2 - V_1) \tag{3}$$

$$V_{swing} = V_2 - V_1 \tag{4}$$

$$E_{V_1 \rightarrow V_2} = C_L.V_{DD}.V_{swing} \tag{5}$$

where $E_{V_1 \rightarrow V_2}$ is the energy provided from the power supply when charging C_L from an initial voltage V_1 to a final voltage V_2. t_1 and t_2 are the times for the output voltage V_{out} to reach V_1 and V_2, respectively. Note that in this case, V_1 is the upper limit voltage that represents a low logic level, whereas V_2 is the bottom limit voltage that represents a high logic level. Having the $V_1 \rightarrow V_2$ transition complete, the energy stored in the output capacitor C_L is

$$E_{C_L} = \int_{t1}^{t2} P_{C_L}(t)dt = \int_{t1}^{t2} V_{out}(t)I_{out}(t)dt = C_L \int_{V_1}^{V_2} V_{out}(t)dV_{out}(t) = \frac{C_L.({V_2}^2 - {V_1}^2)}{2} \quad (6)$$

where $P_{C_L}(t)$ is the instantaneous power stored in the output capacitor C_L. Note that remaining energy provided from the power supply is dissipated in the parasitic resistances of the pull-up transistors during the output transition [KUR04].

Taking now as example a high-to-low transition, the pull-up network is disabled and the pull-down network is now enabled. The current that passes through the pull-down network and discharge the output capacitor C_L still is represented by I_{out}, albeit with opposite polarity. The energy dissipated in this transition by the transistors in the pull-down network is

$$E_{V_2 \to V_1} = \int_{t1}^{t2} P_{pulldown}(t)dt = \int_{t1}^{t2} V_{out}(t)I_{out}(t)dt = -C_L \int_{V_2}^{V_1} V_{out}(t)dV_{out}(t) \quad (7)$$

$$E_{V_2 \to V_1} = -\frac{C_L\left({V_1}^2 - {V_2}^2\right)}{2} = \frac{C_L\left({V_2}^2 - {V_1}^2\right)}{2} = E_{C_L} \quad (8)$$

where $E_{V_2 \to V_1}$ is the energy dissipated by the pull-down network when the output capacitor C_L is discharged from V_2 toV_1. t_1 and t_2 are the times for the output voltage V_{out} to reach V_2 and V_1. Observing (6) and (8), it is possible to visualize that all the energy stored in the output capacitor C_L during a low-to-high transition is dissipated during the following high-to-low transition [KUR04].

With $E_{V_2 \to V_1}$, it is possible to determine the dissipated power while a node is charging or discharging its capacitances. Considering that a node periodically transitions between V_1 and V_2 with a period T_s and a frequency f_s, the average dynamic power consumed by the gate is [KUR04]

$$P = \frac{E_{V_2 \to V_1}}{T_s} = f_s.C_L.V_{dd}.V_{swing} \quad (9)$$

If statistical data is available in order to estimate the average number of transitions that occurs in this specific node while operating, the average dynamic power in this node i can be defined as

$$P_i = \alpha_i.f_s.C_L.V_{dd}.V_{swing} \quad (10)$$

where α_i is the probability that a state changing voltage transition will occur at the node i while operating at a certain clock cyle. As digital circuits usually employ V_{DD} and ground as high and low logical level respectively, the average dynamic power consumed by a gate is [KUR04] [IEO13] [PNS13]

$$P_i = \alpha_i.f_s.C_L.{V_{dd}}^2 \quad (11)$$

In this way, as (11) shows, the dynamic power of a gate is directly proportional to its output capacitance and its switching activity, represented by the product of α_i and f_s. On top of that, it is

proportional to voltage squared, which indicates that even small variations in the voltage can significantly affect the dynamic power of a given CMOS gate.

2.1.2 Leakage Power

Leakage power, or static power consumption, occurs when the circuit is quiescent and its inputs and outputs are static. Even though some transistors are operating in the cutoff region and ideally could cut the current flow from the power supply, still leakage current is drawn from the power supply. This happens due to the non-ideal off-state characteristics of a transistor. As described in [PNS13] [IEO13], The leakage power can be represented by:

$$P_{leakage} = I_{off}.V_{dd} \tag{12}$$

where V_{dd} is the supply voltage and I_{off} is the leakage current of the circuit. In long channel transistors, leakage currents are heavily influenced by weak inversion and reverse biased p-n junction diode currents [KUR04]. However, other leakage elements are relevant when short channel transistors are used, as it will be discussed in the next sections.

2.1.3 Transition Delay

When characterizing the delay aspects of a CMOS gate, the output transition delay must be considered. The transition delay is basically the time that the output takes to switch its binary logical level, *i.e.* to transition from a valid 0/1 to a valid 1/0 [RAB03]. Classically, designers assume that the transition times occur between 10% and 90% of the output voltage swing [RAB03], albeit in some cases other values such as 20% and 80% can be used. The definition of these values is typically technology and library specific and must ensure tolerable noise margins for all cells. Figure 3 illustrates the rise and fall transition delays out an output signal. As the figure shows, a low-to-high (rising) transition t_r is the time it takes for the signal to switch from 10% of V_{dd} to 90% of V_{dd}. A high-to-low (falling) transition t_f, on the other hand, is the time it takes for the signal to switch from 90% off V_{dd} to 10% of V_{dd}. Transition delay is also known as the slew or slope of a transition [RAB03].

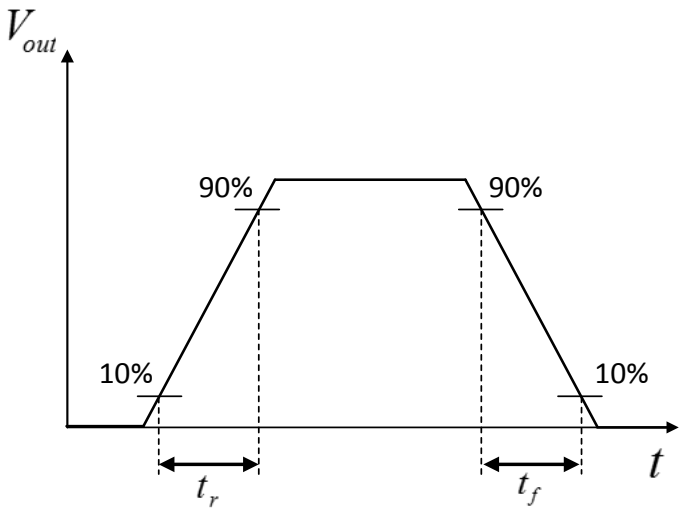


Figure 3 - Rise and fall transition delays.

2.1.4 Propagation Delay

Propagation delay is another important metric for cells characterization. This metric represents the time it takes for a transition in a given input to generate a transition in an output and is typically used by synthesis tools during static timing analysis (STA) for calculating the delay of logic paths. It is measured from the point when the transition in the input signal crosses the switching threshold V_1 to the point when the transition in the output signal crosses the switching threshold V_2. The propagation delay is expressed according to [RAB03]

$$t_p = \int_{V_1}^{V_2} \frac{C_L(v)}{I_{out}(v)} dv \tag{13}$$

Classically, designers assume a switching threshold (V_1 and V_2) of 50% of V_{dd} [RAB03], however in some case other values could be adopted. For better illustration, Figure 4 shows the input and output signal of an inverter and points the switch threshold values of each propagation delay. As the figure shows, a high-to-low (falling) propagation t_f is the time it takes for the input signal to switch from 50% of V_{dd} and propagate to the output, making it gfalls to 50% of V_{dd}. Meanwhile, low-to-high (rising) propagation is the time it takes for the input signal to switch from 50% off V_{dd} and propagate to the output, making it rises to 50% of V_{dd}.In some cases, it is possible to use 40% as the switching threshold for rise transition and 60% for fall transitions [RAB03].

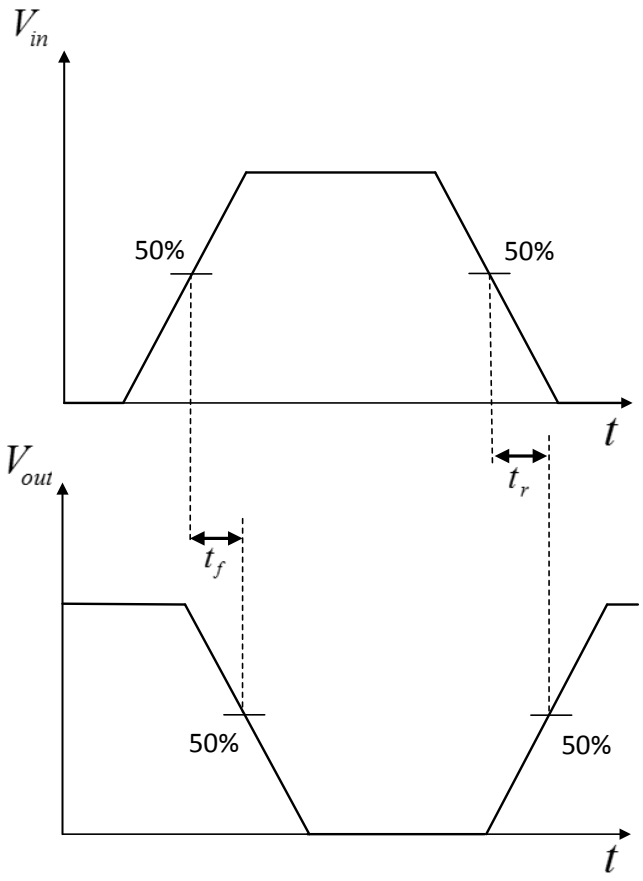


Figure 4 - Example of a fall and rise propagation delay for an inverter gate.

2.2 Super-threshold Operation

The majority of contemporary ICs is designed to operate at super-threshold voltages, *i.e.* with a V_{dd} bigger than V_{th}. In this case, the gate-to-source V_{GS} voltage in the metal-oxide-semiconductor (MOS) transistor is usually higher than the threshold voltage V_{th} ($V_{GS} > V_{th}$). According to the unified MOS model [RAB03], the drain current I_D of a transistor in super-threshold operation can be expressed as

$$I_D = \mu_o C_{ox} \frac{W}{L} \left(V_{GT} V_{min} - \frac{{V_{min}}^2}{2} \right) (1 + \lambda V_{DS}) \;\; for\, V_{GT} \geq 0 \tag{14}$$

$$with\, V_{min} = \min(V_{GT}, V_{DS}, V_{Dsat})$$
$$V_{GT} = V_{DS} - V_{th}$$

where μ_o is the charge-carrier effective mobility, W is the gate width, L is the gate length, C_{ox} is the gate oxide capacitance per unit area, V_{Dsat} is the drain-to-source voltage when critical electrical field is reached (velocity saturation) and λ is the channel-length modulation parameter. λ is responsible to model the current dependence on drain voltage due to the early effect, which is intensified in short-channel transistors [RAB03]. Note that V_{min} only keeps the lower value among V_{GT}, V_{DS} and V_{Dsat}. Despite the fact the Equation (14) only holds when $V_{GT} \geq 0$, the drain current I_D is also heavily dependent of V_{DS}, value that depends on the operating voltage. In this way, recalling Equation (13), the propagation delay of a CMOS cell is inversely proportional to the operating voltage.

2.3 Sub-threshold Operation

2.3.1 Sub-threshold Conduction

The sub-threshold or weak-inversion conduction is an effect present in MOS transistors when the gate-to-source V_{GS} voltage is equal to or lower than the threshold voltage V_{th} ($V_{GS} \leq V_{th}$). Operating in this region, the transistor is neither completely turned on nor turned off [RAB03]. In most digital applications, the sub-threshold current is caused by parasitic leakage currents and is accordingly undesirable, because it is seen as a deviation from the ideal switch-like behavior of the MOS transistor [LIU12] [RAB03]. As described by Ieong *et al.* in [IEO13], the current I_{Dsub} while operating in the sub-threshold region is approximated by:

$$I_{Dsub} = I_o e^{\left(\frac{V_{GS} - V_{th} + \eta V_{DS}}{n V_t}\right)} \left(1 - e^{\frac{-V_{DS}}{V_t}}\right) \tag{16}$$

$$I_o = \mu_o C_{ox} \frac{W}{L} (n-1) {V_t}^2 \tag{17}$$

where I_o is the nominal current, n is the sub-threshold slope factor, V_t is the thermal voltage, and η is the Drain-Induced Barrier Lowering (DIBL) coefficient. Besides, μ_o is the surface mobility, C_{ox} is the unit area oxide capacitance. The value $\mu_o . C_{ox}$ stands for the intrinsic transconductance. From Equation (16), it is possible to see that the current has an exponential relationship with V_{GS} and V_{th}.

Operating in the sub-threshold region can reduce both static and dynamic power consumptions. This is possible due to the relationship between the supply voltage and dissipated power, expressed in Equations (11)

and (12). According to (11), the reduction of the supply voltage leads to quadratic savings in dynamic power consumption, while (12) shows that static power is reduced linearly. Furthermore, the use of sub-threshold design may enable the use of energy harvesting[1]. As a large share of low power applications use batteries as an external power supply, power savings enabled by sub-threshold design can increase the time between recharges [PNS13]. Nonetheless, circuit operation in the sub-threshold region suffers from problems, most coming in the form of performance degradation and higher sensibility to process variations.

In a CMOS standard cell, the output fall and rise transitions rely on the I_{on}/I_{off} ratio, where I_{on} is the active current and I_{off} is the leakage current [PNS13]. If the I_{on}/I_{off} ratio is too small, the active pull-down (PDN) or pull-up transistor networks (PUN) may not have strength enough to change the logic level of the output, making the circuit fail. Due to the reduced supply voltage and active current, sub-threshold operation implies a smaller I_{on}/I_{off} ratio [PNS13], which also interferes in the performance of sub-threshold circuits, increasing delays by several orders of magnitude. Moreover, PVT variability, transistor sizing or channel doping are responsible for variations of the threshold voltage, changing the transistor currents and, consequently, the I_{on}/I_{off} ratio [PNS13].

2.4 Planar Bulk CMOS in Sub-threshold Operation

As planar bulk CMOS processes are most often optimized for super-threshold operation, the use of devices for sub-threshold operation may lead to unexpected effects. This Section explores the most relevant of these effects to help the understanding of how each one interferes in the characteristics of MOS transistors. This enables assessing the challenges of sub-threshold design.

2.4.1 Drain-Induced Barrier Lowering (DIBL)

Drain-induced barrier lowering or DIBL is a short channel effect in MOSFETs referring originally to a reduction of a transistor threshold voltage at higher drain voltages. In a classic planar field effect transistor with a long channel, the bottleneck in channel formation occurs far enough from the drain contact in order to be electrostatically shielded from the drain by the combination of the substrate and gate. Hence, classically, the threshold voltage was independent of drain voltage. In short channel devices this is no longer true. The drain is close enough to source, and so a high drain voltage can *open the bottleneck* and turn the transistor on prematurely. Figure 5 illustrates the DIBL effect in a transistor. As part of the depletion in very short channel transistors is affected by the drain and source bias and lower gate voltage is necessary to deplete the transistor, the barrier for electron injection between drain and source decreases. Consequently, DIBL causes I_{DS} and sub-threshold leakage to increase with higher V_{DS} [WES10]. In order to see the impact difference of DIBL in long and short channel transistors, Figure 6 shows the potential variation along the channel in both transistor types.

[1] Energy harvesting (also known as power harvesting or energy scavenging) is the process in which energy is captured from a system's environment and converted into usable electric power. Energy harvesting allows electronics to operate where there's no conventional power source, eliminating the need to run wires or make frequent visits to replace batteries. An energy harvesting system generally includes circuitry to charge an energy storage cell, and manage the power, providing regulation and protection.

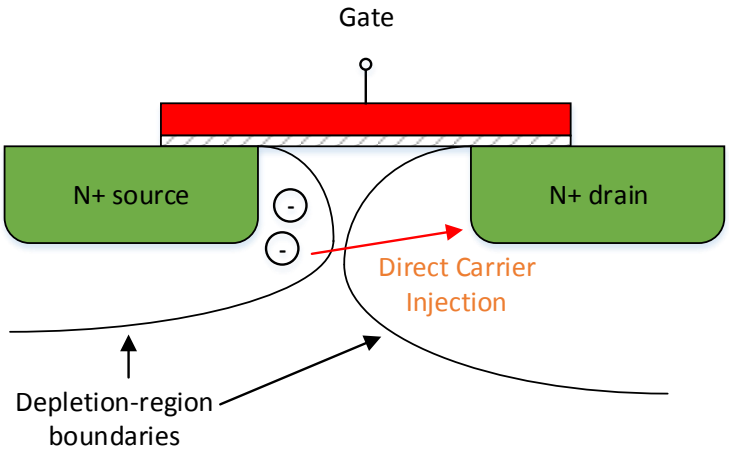


Figure 5 - DIBL effect in a NMOS transistor. Adapted from [SARAS].

This effect is controlled in super-threshold design because *halo implants* (an additional technology step in transistor fabrication) are placed in the channel next to drain and source [PNS13]. However, if V_{dd} is reduced, the drain voltage is reduced as well. Hence, sub-threshold operation minimizes the impact of DIBL on the transistor.

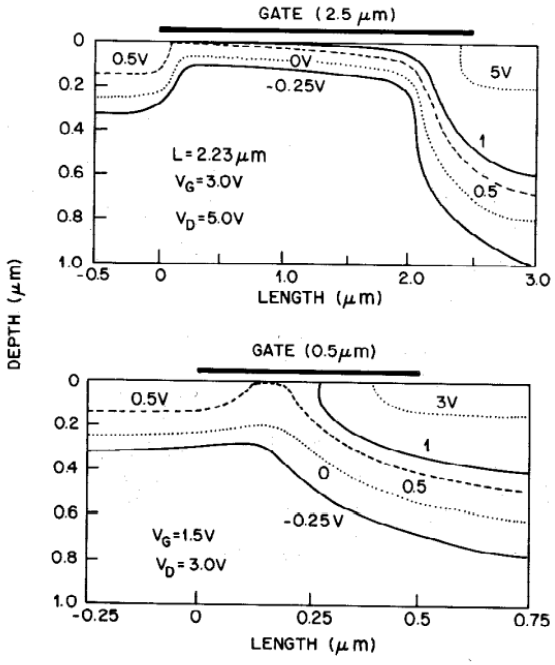


Figure 6 - Potential variation along the channel illustrating DIBL impact. Extracted from [SARAS].

2.4.2 Reverse Short Channel Effect (RSCE)

Reverse Short Channel Effect or RSCE is a result of halo implants, which are used in current MOSFET technologies to improve scaling and control of short channel effects [RIO02]. However, these implants generate a non-uniform channel doping profile in the device. For better illustration, Figure 7 shows the net doping levels along the device's length.

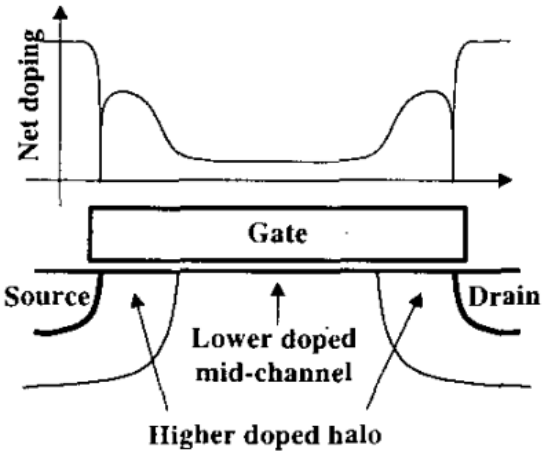


Figure 7 - Non-uniform channel doping resulting from typical halo process. Extracted from [RIO02].

RSCE implies an increase of V_{th} when the channel length of the transistor decreases. In super-threshold operation, RSCE affects the functionality of the transistor less than DIBL. Thus, short channel transistors are faster than long channel transistors [PNS13]. However, DIBL is minimized in sub-threshold operation and RSCE can be a dominant effect that leads to increased delay for minimum channel lengths [LIU12] [PNS13].

2.4.3 Inverse Narrow Width Effect (INWE)

The Inverse Narrow Width Effect or INWE implies in the decrease of the threshold voltage when the transistor's width is decreased. It is generated by the Shallow Trench Isolations (STI) or fully recessed isolation oxide, which are applied at the edges of the transistors to improve the electrical field in these zones [PNS13]. Figure 8 shows the surface potential of a MOSFET with fully recessed isolation oxide.

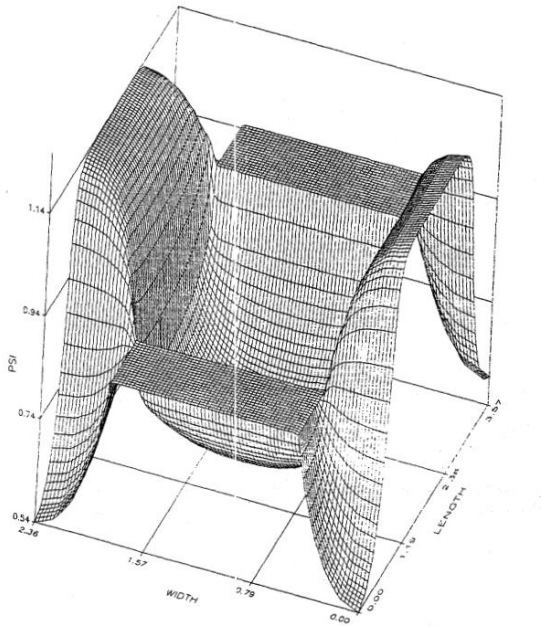


Figure 8 - Surface potential of a short-geometry MOSFET. Extracted from [AKE87].

It is possible to see the enhanced potentials at the edges of the channel that are generated by gate fields terminating on the sidewalls of the channel [AKE87]. With the present fields, the decreasing of the transistor's width implies in a significant contribution from the sidewall gate field capacitances to the gate capacitance C_g, which inflicts variations to the threshold voltage. An alternative to minimize INWE is using smaller transistor widths and apply transistor fingering when higher currents are needed [PNS13].

2.5 Asynchronous Circuits

Most of today's digital systems are synchronous, at least locally synchronous. They employ a global (or locally global) clock signal to synchronize the operation of all storage elements, creating the abstraction of a discrete notion of time. In an asynchronous circuit, on the other hand, sequential components communicate without a clock signal [MYE01]. Here, handshake protocols are used to perform synchronization, communication and sequencing of operations [SPA01]. The discrete-time abstraction of synchronous circuits helps simplifying the design, but removing it can grant several benefits, as lower power consumption, higher operating speed, lower electromagnetic noise emission, and avoiding clock distribution problems [BEE10] [MYE01] [HAU95] [SPA01]. However, differently from synchronous designs, asynchronous circuits can rely on different templates, each with its own benefits and drawbacks [BEE10].

2.5.1 Quasi-Delay-Insensitive Design

Among the different asynchronous design templates available in the literature, bundled-data and quasi-delay-insensitive (QDI) are the main template families. Despite the fact that bundled-data design can benefit to a some extent of the use of conventional design tools due to its similarity to synchronous circuits, bundled-data templates still requires extra care with the definition and verification of timing constraints between data and control signals. An alternative to avoid these issues is to encode the control signals within the data channel, which is the main strategy adopted by QDI design. In fact, QDI design is reported by Martin and Nyström as the most practical template, due to its relaxed timing constraints [MAR06]. Its structure requires a choice of handshaking protocol and a delay-insensitive (DI) code to represent data.

One of the most used DI codes is called dual-rail (1-of-2 channel) [MAR06]. Dual-rail channels embed the request signal within data signals, by representing each bit with two wires. This of course brings the need of extra hardware, but relaxes timing matching. In dual-rail channels each bit of data is encoded by two wires: *d.t* and *d.f* and rely on the classic return-to-zero 4-phase (RTZ) handshake protocol [SPA01]. Table 1 presents the basic codification for these channels. The request signal is regarded as asserted when *d.t* and *d.f* assume different logical levels. In order to represent a '1' logic level, it is necessary to set *d.t* high and *d.f* low. The representation of a '0' logic level follows an opposite convention: *d.t* is set low and *d.f* high. Between each pair of valid data a spacer must always be signaled. In the case of this RTZ example, all wires must return to zero. Note that both signals set to logic 1 is defined as an invalid state. For instance, assuming a 2 bits channel, if it is desired to represent the value "11", *d0.t* and *d1.t* must be set high and *d0.f* and *d1.f* set low. When both signals are set to logic 0, this represents a spacer or empty state. Figure 9 (a) illustrates two examples of the RTZ 4-phase handshake protocol. As an initial state, all data signals are reset in the beginning of the communication cycle, indicating a spacer. Then, the data channel presents a valid data codification (marked as 1 in Figure 9 (a)). As a consequence, the ack signal is asserted, signalizing that the data was

computed (2). Next, the data channel shows a spacer and exhibits that the data is no longer valid (3). At last, the ack signal is reset, finalizing the communication cycle (4). The first communication cycle indicates the transmission of "11", whereas the second cycle indicates the transmission of "01" through the data channel.

Table 1 –Codification for a 1 bit dual-rail channel using RTZ and RTO handshake protocols.

Signals		Value	
d.t	***d.f***	**Using RTZ**	**Using RTO**
0	0	**Spacer**	**Invalid**
0	1	**0**	**1**
1	0	**1**	**0**
1	1	**Invalid**	**Spacer**

It is also possible to implement dual-rail channels using a variation of the RTZ convention, called Return-to-One (RTO) [MOR12]. This variation assumes that data is represented as the negation of data in RTZ. In this case, spacers are represented by all wires at 1 and all wires at 0 represent an invalid state. Also, data is signaled by a wire at 0. As Table 1 shows, for a 1 bit channel, a logic '1' is represented by *d.t* at 0 and *d.f.* at 1 and a logic '0' by *d.t.* at 1 and *d.f.* at 0. For this case, Figure 9 exhibits two examples of the RTO 4-phase handshake protocol. Note that the Figure 9 (b) shows the same transmission as Figure 9 (a), albeit all data signals *d.f* and *d.f* are inverted.

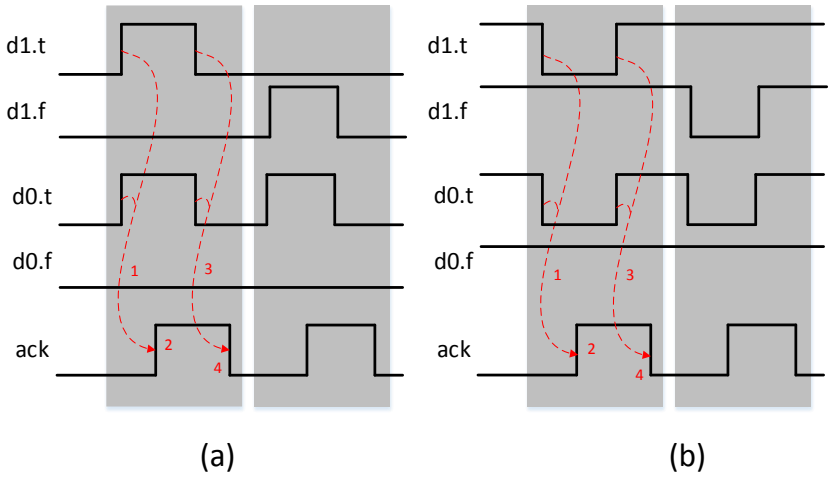


Figure 9 - Example of data transmission through a 2 bits dual rail channel based on (a) RTZ and (b) RTO. Adapted from [SPA01].

2.5.2 Null Convention Logic

Theseus Logic, Inc. proposed the NCL logic family [FAN96] to implement 1-of-n QDI asynchronous circuits. Since then, NCL has been applied to deal with power problems [ZHO10] [GUA10], to design high speed circuits [WUJ10] [YAN11] and fault tolerant applications [LOD12], among other uses. Employing NCL gates permits power-, area- and speed-efficient QDI design with a standard-cell-based approach, as opposed to other asynchronous templates that require full-custom approaches. NCL gates couple a threshold function with positive integer weights assigned to inputs to the use of a hysteresis mechanism. Figure 10 shows a generic symbol for a NCL gate with a threshold function M and N inputs, each one with a weight w_i. According to the NCL gate function, each input can have different weight values. However, if the weights are not informed, weight 1 is assumed for all.

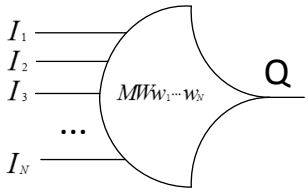


Figure 10 - Basic NCL gate symbol.

The output of the NCL gate switches according the following premises: (1) a high-to-low transition only occurs when all inputs are at logic level 0; (2) a low-to-high transition occurs when the sum of weights for inputs at logic level 1 is bigger or equal to the threshold M. In case the inputs do not satisfy the threshold function, the output holds its previous state. These characteristics demonstrate how NCL gates are similar to a classic asynchronous component, the C-element. In fact, a C-element is a special case of NCL-gate where $M = N$.

3. STATE OF THE ART

This Chapter presents an overview of recent propositions of sub-threshold and asynchronous standard cell libraries in the literature. It covers mostly the aspects and design flow characteristics of each library.

3.1 Pons et al. Low-Power Standard Cell Design [PNS13]

In this work, the authors present a sub-threshold standard cell library based on an 180nm CMOS technology that was optimized to be supplied at 400mV. The library contains 33 combinational cells, such as inverters, NANDs, NORs, XNORs and some complex gates. In addition, there are 12 sequential cells, including latches and flip-flops. The authors restricted the number of the cells' inputs to three or less, to restrict I_{on}/I_{off} degradation. Cells with more than three inputs present higher I_{off}, which contribute to a lower I_{on}/I_{off} ratio.

Regarding the sizing methodology, transistors are sized in two steps. First, the channel length L of all transistors is increased, to minimize RSCE, consequently increasing I_{on}. This step also decreases the sensibility to process variations. Next, transistors' widths are upsized, to increase I_{on} and the drive strength even more. Concerning INWE minimization, the sizing in this second step is done using transistor multi-fingering. Besides these steps, there is also a specific step that modifies sequential cells. A particular technique, adopts clocked feedback in latches and flip-flops (which means that feedback loops are only enabled in the memorization mode, not during the writing of new data). This modification helps sequential cells, by disconnecting the feedback structure during write operations, thus avoiding output degradation. The design flow is similar to previous low-power libraries, but it adds changes to support sub-threshold design. The library characterization, logic synthesis, place and route are performed using Cadence tools. However, this library has no specific support to the design of asynchronous circuits.

3.2 Ieong et al. Standard Cell Library Design for Ultra-Low-Power Biomedical Applications [IEO13]

This work also presents an 180nm CMOS sub-threshold standard cell library. According to results reported in the article, this library achieved the lowest power consumption at 600mV. Unlike the approach of Pons et al. [PNS13], this work pledges that transistors are to be kept as small as possible, to minimize parasitic capacitances. Hence, the channel length of transistors was kept in its smallest possible value. Through simulations using inverter chains, the authors indicate that maintaining a 1:1 P/N ratio provides the lowest power consumption. This work employs the logical effort methodology [SUT99], based on conventional sizing techniques. Initially, the characteristic inverter is sized and then other gates with stacked transistors are sized according to the topology. The next design flow steps (cell characterization, logic synthesis, place and route) are achieved using existing CAD tools. Again, the library itself and the tools employed in the design flow are not fully thought to support asynchronous circuit design.

3.3 Liu et al. Standard Cell Sizing Methodology for Sub-threshold Operation [LIU12]

The authors propose a new transistor sizing methodology for standard cells in sub-threshold operation. Their approach is based on balancing the N and P networks using statistical formulations to minimize short-channel effects and process variations. These formulations come from the conclusion that the threshold voltage and current of an NMOS transistor operating in sub-threshold regime obey to a Normal and Log-Normal distributions, respectively. The resulting equation allows to choose between maximizing cell current with area constraints or minimizing cell area with current constraints. This work also introduces transistor sizing formulations for combinational logic, as well as optimization techniques for flip-flop design. With this methodology, the authors created a standard cell library using a 90nm CMOS technology. The library contains 144 cells, including inverters, logic gates and flip-flops. The cell characterization process covers slow process (SS) and typical process (TT) corners at 25°C and 300mV.

Despite the fact that this library presents a sound methodology, still the authors do not include any asynchronous components considerations in their work.

3.4 Lotze and Manoli Standard-Cell-Based Design using Schmitt Trigger Logic [LOT11]

Differently from the previously discussed works, the authors show here a new topology approach using a Schmitt Trigger logic. With the hysteresis characteristic of Schmitt Trigger gates, the authors aim to reduce leakage current from the critical output node, improving the I_{on}/I_{off} ratio. To do so, all logic cells were redesigned and the Schmitt Trigger principle was applied to build the new transistor topology of the cells. This is done by replicating both the PUN and the PDN, and adding a hysteresis transistor between the original PUN/PDN and its respective replicated network. The main drawback of this technique is a significant area overhead.

The authors present sizing information regarding a NAND2 gate with Schmitt Trigger logic. The methodology is based on the Voltage Transfer Curve (VTC) of the cell and transistors are sized to achieve good margin noises. Consequently, the VTC must be balanced in the center, where $V_m = V_{dd}/2$. However, the fact that the Schmitt Trigger logic is mixed with the gate logic complicates the sizing process. Almost every transistor must have its width and length dimensions sized differently.

Using the described design process, the authors created a standard cell library in a 130nm CMOS technology. This library only contains inverters, two-input NANDs, two-input NORs and flip-flops. Although the library has a small number of gates, the authors highlight that this limitation helps achieving a significant reduction on the voltage supply. They demonstrate that their cells achieve extremely low voltage supplies, in the range of 62mV to 90mV.

3.5 Zhou et al. Dual-Width Standard Cell Library for Near/Sub-threshold Operation [ZHO12]

This work proposes an INWE-aware sub-threshold sizing methodology and a dual-width standard cell library, which contains inverters, buffers, NANDs, NORs, MUXes, latches and flip-flops. Dual-with stands for the use of two set of gates. The first set is sized following a sizing methodology proposed by the authors. The latter set is composed by minimum-sized gates. A CMOS 40nm technology was used to implement the library. The fan-in of all cells was limited to two or less.

Trying to minimize INWE, the authors applied the same technique proposed by Pons et al. [PNS13]. The idea is to layout transistors with minimum-width multiple fingers. Therefore, the transistor width and the threshold voltage remain unchanged, making the drain current proportional to the transistor width increase. Compared to a conventional sizing methodology, the new one improves performance and power consumption of gates while operating at lower voltages, but increases gate area. However, this area overhead was minimized when the new sizing methodology was applied to the design of large circuits.

The proposed library contains two types of gates. The first one consists in gates designed with the proposed sizing methodology. The second type is based on minimum-sized gates. Mixing these types, the authors achieve further power optimization, while keeping performance improvements. The approach was to dedicate the INWE-aware sized gates to critical paths of the circuits, while minimum-sized gates are used in non-critical paths.

Unfortunately, asynchronous circuit components are not included in the scope of this work.

3.6 Liao and Hutchens Robust Ultra-Low-Power Sizing for Cell Libraries [LIA12]

Having a similar approach proposed by [PNS13] and [ZHO12], the authors adopt the multiple fingers technique to minimize INWE. RSCE is covered as well by increasing the transistor's channel length. Only three cells were considered (INV, NAND and NOR) with one driving strength. All evaluations were performed using a 180nm CMOS technology with $V_{dd} = 400mV$.

Initially, the authors present a threshold voltage and drain current evaluation of PMOS and NMOS transistors according their dimensions. Thus, threshold variations can be related to the transistor's dimensions and, consequently, to INWE and RSCE. At the end of evaluation, the authors indicate how intense INWE and RSCE affect PMOS and NMOS transistors and show similar results obtained by other authors [PNS13] [ZHO12], indicating that applying transistors with minimum width and multiple fingers mitigate INWE and achieve higher drain currents.

Having the results of this evaluation, the authors propose a cell library sizing methodology and design it employing the multiple fingers technique. The sizing methodology also considers the Energy-Delay Product (EDP) and Static Noise Margin (SNM) to define the best transistor dimensions. When cells were sized focusing only EDP, the authors achieve delay reduction of 72%~94% and EDP reduction of 76%~90%. However, the authors achieve delay reduction of 44%~80% and EDP reduction of 22%~70% when considering EDP and SNM optimization.

3.7 Kwong Sub-threshold Cell Library and Methodology [KWO06]

This work is a product of a Master Thesis published at the Massachusetts Institute of Technology (MIT). The author proposes a sub-threshold methodology and a 65nm CMOS technology cell library containing 56 gates. The library also was integrated in a CAD design flow. The optimum voltage supply for this library was set as 250mV and the number of inputs of all cells was limited to three or less. All basic logic functions with two and three inputs were provided, and the library also includes several variations of flip-flops and latches.

The adopted sizing methodology is based on the gate topology, minimum energy and process variation. Sizing single-stage gates, the author indicates the use of the basic sizing method. Both NMOS and PMOS width sizes are increased to achieve higher drive strengths. For multiple-stage gates, the main sizing strategy is to make the output stage identical to the single-stage gate with the same drive strength. Moreover, the logic effort technique [SUT99] is employed to minimize delay through all stages of the day. Transmission gates were designed with minimum width since no delay or energy benefits are achieved increasing them.

The next steps of the method were performed by conventional CAD tools. However, the author warns that during cell characterization, all gates should be characterized at several supply voltages. Hence, the designer can predict how the circuit would behave when it is necessary to change the voltage supply due to external variations, such as temperature and workload.

Again, the design of components for asynchronous circuits were not covered in this work.

3.8 The ASCEnD Standard Cell Library [MOR13a]

The ASCEnD library was proposed by the GAPH group at the Pontifical Catholic University of Rio Grande do Sul. This library was devised to support asynchronous circuits design and works in tandem with the basic standard cell library of the underlying technology. ASCEnD employs the STMicroelectronics 65nm bulk technology and currently contains over six hundred components, such as C-elements and NCL gates. Currently, the library counts with a fully automated design flow, from the transistor sizing step to the cell layout generation.

The sizing methodology is divided in two steps. Initially, the cell schematic is designed with the help of an in-house tool called ROGen (the name is an acronym of Ring Oscillator Generator). This tools uses the Cadence Spectre electric simulator to generate a large number of simulation scenarios for the cell. Its output provides power and delay information for the simulated schematic. After that, another in-house tool, called CeS (from Cell Specifier), processes the information obtained from ROGen. The designer informs a cost function to trade-off power and performance and CeS then selects the most appropriate transistor dimensions.

Regarding the next steps, cells are generated using Astran [ZIE14], a dedicated tool to produce a cell layout compatible with the Cadence Virtuoso. The cell layout is validated using Mentor Calibre DRC and LVS. Mentor Calibre PEX is also used to extract parasitics from the verified layout. Finally, another in-house tool is used for the cell electrical characterization. This tool, called LiChEn [MOR13c] (from Library Characterization Environment), processes the extracted circuit, computing all static states and transition arcs and characterizes the cell, quantifying its transition and propagation delays, input capacitance, internal, switching and leakage power.

Despite the fact that the ASCEnD library fully supports asynchronous circuits, ultra-low power design is out of the scope of the library, at least in its current state.

3.9 Maurine et al. Standard Cell Library [MAU03]

Proposed by the French labs TIMA and LETI, this library was designed to support QDI asynchronous circuits. Its first version used a 130nm gate length technology, called TAL-130. However, it later evolved to a 65nm gate length version, which is called TAL-65. These libraries contain several variations of C-elements and latches. The sizing strategy adopted comprises five design rules: (1) balance the currents flowing through the PDN and PUN to balance active and RTZ phases; (2) design at least the drives X0, X1, X2 and X4 for each cell function. Hence, the library should cover a span of loads; (3) design components with a given drive strength to provide the same current capability as the inverter with the same drive strength; (4) accommodate weak and critical loads in two functional stages. This rule focuses on minimizing area while preserving speed performance; (5) Avoid logic decompositions where the state holding element drives the output node.

Unfortunately, there is not much further information about the design flow used, which restricts a deeper the analysis of this library.

3.10 USC Asynchronous Standard Cell Libraries [BEE11] [FER04]

The Asynchronous CAD and VLSI group from the University of Southern California (USC) has successfully implemented two asynchronous libraries. Designed in the context of a Ph.D. thesis [FER04], the first library contains a set of basic cells that are specifically used to implement a QDI template based on single-track full-buffer (STFB). The technology used in this case was Taiwan Semiconductor Manufacturing Company (TSMC) 250nm bulk CMOS technology.

According to Ferretti [FER04], the sizing methodology follows single-size strategy for each STFB function. The process is manual and employs extensive analysis via SPICE simulation. Although the used technology allows the implementation of smaller transistors, the adopted minimum NMOS and PMOS transistor width was 0.6µm and 1.4 µm, respectively. In addition, the author determined that the width of each NMOS transistor in an N-stack should be $k * 1.2$ µm, where k is the number of stacked NMOS transistors in the PDN. The sizing is also based on the fan-out of 4 (FO4) rule. This means that one inverter should efficiently drive four times its own input load.

The second library was proposed by Beerel et al. [BEE11], which also present an automated flow for Pre-Charged Half-Buffer (PCHB) asynchronous template design. The library contains control circuit cells and C-elements with 4 inputs or less. It also implements all 2-input and 3-input combinational logic functions, as well as various 4-, 5- and 6-input combinational functions. As the library is proprietary, it is not possible to retrieve much detailed information about the sizing methodology, or discovering whether the transistor sizing step is automated.

3.11 Comparison Table

Table 2 shows the main aspects of each of the explored works on asynchronous and sub-threshold libraries design. Minimum voltage, sizing methodology and the used CMOS technology of each library are indicated. In addition, the table indicates whether the library support asynchronous design and/or sub-threshold operation. By analyzing the table, it is evident the lack of a cell library in the literature that support both sub-threshold operation and asynchronous circuits. This makes this work even more attractive. Exploring and evaluating sub-threshold NCL gates contributes to fill this gap. In addition, it allows a better view of the benefits and drawback that asynchronous circuits and sub-threshold operation may provide when applied together.

Table 2 - Comparison table showing main characteristics of each standard cell library.

Cell Library	Nominal Vdd	Minimum Vdd	Sizing Methodology	CMOS Technology	Asynchronous Support	Low voltage Support
Pons et al. [PNS13]	1.8V	400mV	RSCE/INWE-aware	180nm	No	Yes
Ieong et al. [IEO13]	-	600mV	Logical effort methodology	180nm	No	Yes
Liu et al. [LIU12]	-	300mV	Minimum short-channel effects and PVT	90nm	No	Yes
Lotze and Manoli [LOT11]	-	62mV	Minimum Supply Voltage and Schmitt Trigger usage	130nm	No	Yes
Zhou et al. [ZHO12]	1.2V	300mV	INWE-aware and minimum energy	40nm	No	Yes
Kwong [KWO06]	1.2V	250mV	Minimum energy and PVT	65nm	No	Yes
Liao and Hutchens [LIA12]	-	400mV	RSCE/INWE-aware	180nm	No	Yes
ASCEnD [MOR13a]	1V	1V	Cost function power/performance	65nm	Yes	No
Maurine [MAU03]	1.2V	1.2V	Balanced active and RTZ phases	130nm	Yes	No
USC [FER04]	2.5V	2.5V	Single-size strategy	250nm	Yes	No
USC [BEE11]	-	-	-	-	Yes	No

4. VOLTAGE SCALING ON CMOS CELLS

The first practical aspect of this work is to understand gathering all the information provided in the previous Chapters, this Section presents experiments conducted to understand on the impact of voltage scaling in CMOS gates. Timing and power impact are the main covered characteristics. The target technology used is a 65nm bulk CMOS technology from STMicroelectronics. All results reported here were obtained using the Cadence analog simulator Spectre.

4.1 Transistor

As a first analysis, lets consider an NMOS transistor with minimum dimensions ($W = 0.135\mu$ and $L = 0.06\mu$). Figure 11 shows the I_D versus V_{DS} characteristic of the NMOS transistor for different V_{GS} values. The curve present a quasi-linear region at low V_{DS}, when $V_{DS} < V_{GS}$, and a quasi-saturation region when V_{DS} reaches higher values, when $V_{DS} \geq V_{GS}$. Recalling Equation (13), the quasi-linear region is modeled by the center term $\left(V_{GT}V_{min} - \frac{{V_{min}}^2}{2}\right)$ of the equation, whereas the quasi-saturation region is modeled by the rightmost term $(1 + \lambda V_{DS})$. Note that I_D has a similar curve for sub-threshold operation, albeit with significant lower values, which can reach the scale of nA. Accordingly, Figure 12 plots I_D versus V_{DS} for the same NMOS transistor for voltages below its threshold voltage, 0.48V. Note that the charts are presenting in a logarithmic scale to highlight the low currents that take place in this region of operation. Albeit these currents are orders of magnitude smaller than those in the super-threshold domain, they present the same behavior, quasi-linear when $V_{DS} < V_{GS}$ and quasi-saturation region when $V_{DS} \geq V_{GS}$. Recalling Equation (16), its rightmost term $(1 - e^{\frac{-V_{DS}}{V_t}})$ models the roll-off current at small V_{DS}, which represents the quasi-linear region. In the meantime, the quasi-saturation slope is a consequence from DIBL and is modeled by η [KWO06].

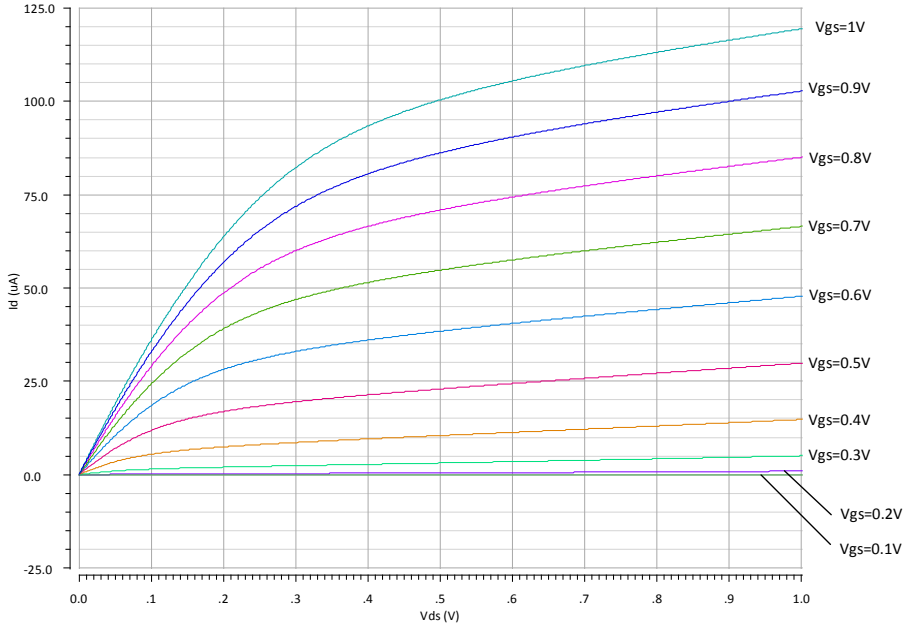


Figure 11 - I_{DS} versus V_{DS} characteristic of an NMOS transistor with $V_{th} = 0.48V$.

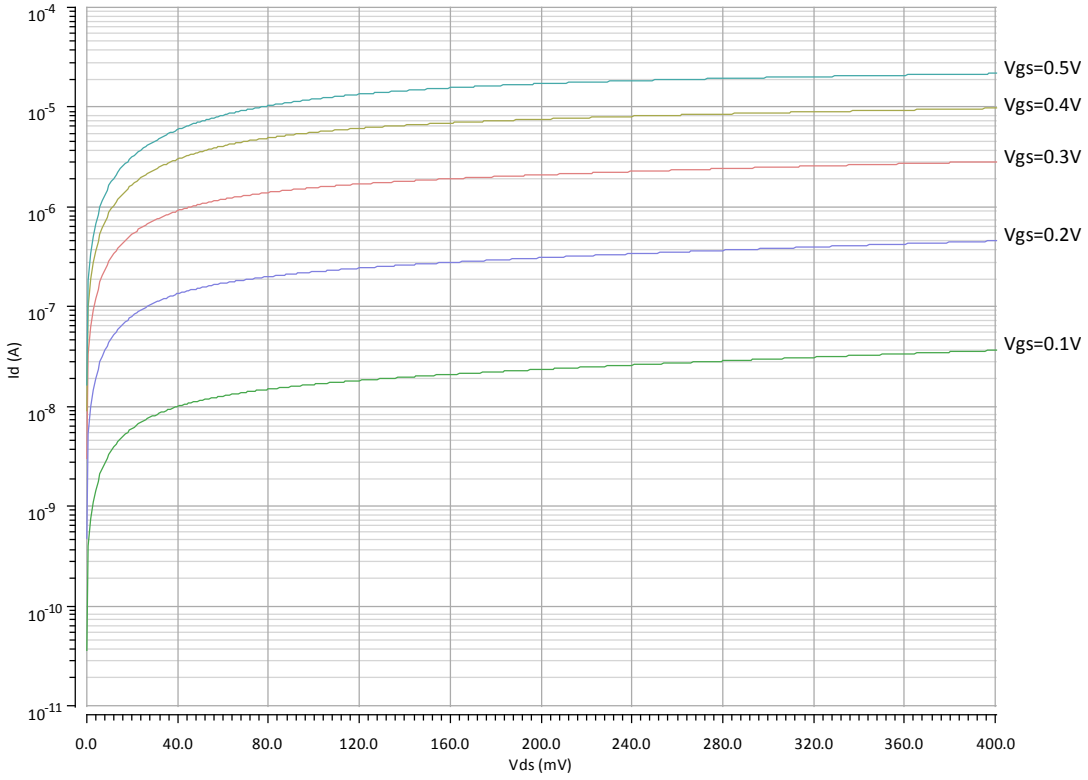


Figure 12 - I_{DS} versus V_{DS} characteristic of an NMOS transistor for subthreshold operation.

A closer analysis of Figure 12 shows a challenge in sub-threshold operation. For example, consider $V_{GS} = 0.1V$. In this case, the maximum current it can delivers is in the order of tens of nA. This is equivalent

to leakage current of the MOS transistors of this technology in the cut off region, roughly 40 nA. In this way, ensuring that cells operate under such conditions can be a challenging task, because leakage and active currents must be well balanced.

4.2 Inverter Gate

Initially, let us consider the inverter gate from Figure 13, which was selected from one of the libraries provided by the target technology vendor. Both NMOS and PMOS are Standard Threshold Voltage (SVT) transistors. It is possible to see in Voltage Transfer Curve (VTC) in Figure 13 the output degradation while decreasing V_{dd}. Such degradation presents itself in terms of a poor voltage transfer curve, a reduced slope in the charts of Figure 13, which leads to performance and power degradation, due to effects like longer short circuit periods of time while switching. With moderate values of V_{dd}, the VTCs are similar to traditional super-threshold curves. Albeit, reaching extremely low values of V_{dd}, below 0.3V, the output shows a strong degradation. This is due to the exponential decreasing of I_{on}/I_{off}. Such analysis corroborates that it is not possible to neglect the leakage currents in the gate while operating in sub-threshold region.

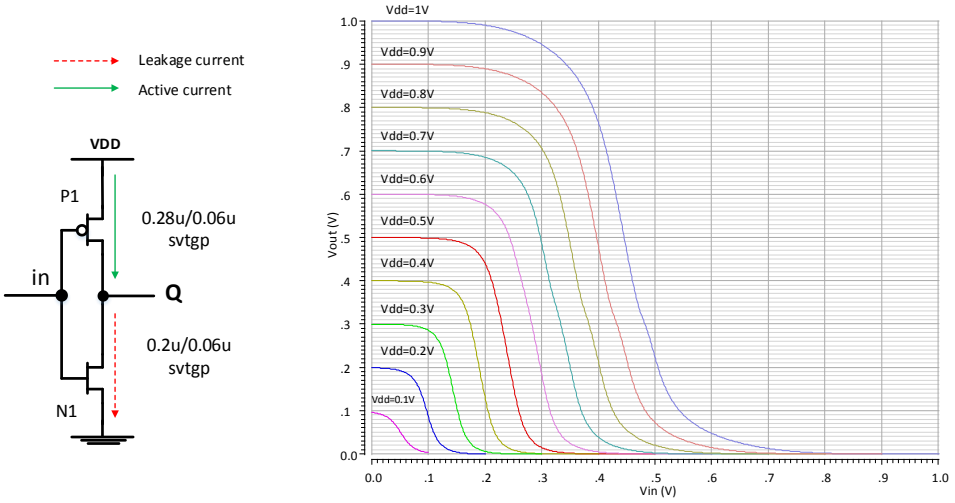


Figure 13 – Evaluated inverter gate and multiple VTC curves with $V_{DD} = \{0.3V, 0.25V, 0.2V, 0.15V, 0.1V\}$.

A characteristic that suffers huge impact in sub-threshold operation is performance. As I_{on} also decreases exponentially in sub-threshold region, which is expressed in Equation (16), the gate's performance decreases significantly and becomes very sensitive to any voltage variation. This relation between I_{on} and performance can be seen in Equation (13) as I_{on} is inversely proportional to the propagation delay. Figure 14 shows the delay degradation of the inverter with an F04 equivalent output load while scaling V_{DD}. As V_{dd} reaches values below the threshold voltage, the gate's transition delay increases about 100x-1,000x. In contrast, the gate's energy while transitioning also shows a more significant decreasing when V_{DD} hits the sub-threshold region, which can be seen at Figure 14 as well. Unfortunately, the delay impact is much higher than the energy saving in the gate at sub-threshold voltages. In fact, another metric that correlates energy and performance, Energy-Delay Product (EDP), shows that the its minimum point is $V_{DD} \sim 0.55V$ (see Figure 15). This means that the voltage range in which the gate is most energy efficient, is in the near-threshold region,

indicating that sub-threshold operation is more challenging and its advantages rely on specific design techniques, such as the one discussed in [LOT11].

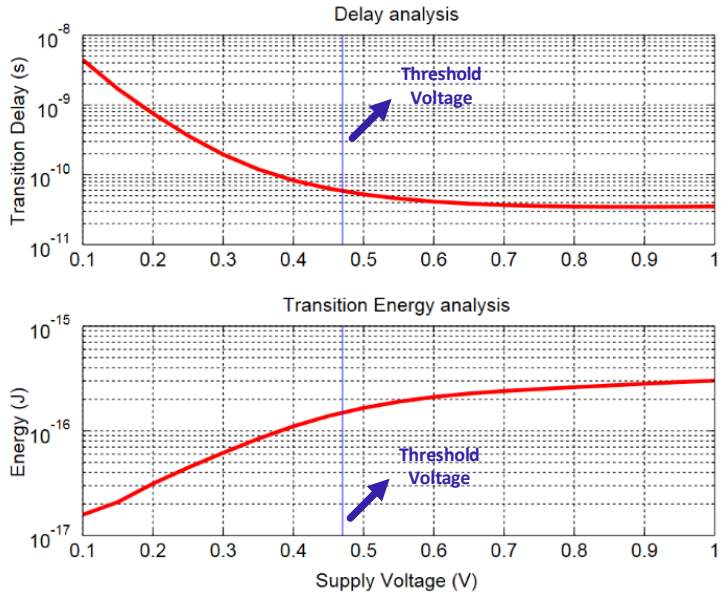


Figure 14 - Transition delay and transition energy of the evaluated inverter. FO4 load used.

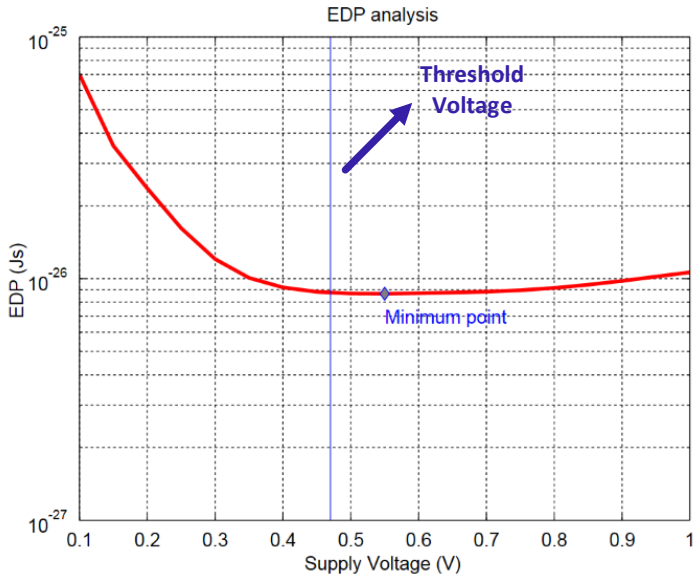


Figure 15 - Energy Delay Product (EDP) of the evaluated inverter.

It is also important to consider the leakage power of gates in sub-threshold operation. Figure 16 shows the leakage power of the evaluated inverter gate when the output is at high and low logical levels. With V_{DD} below threshold voltage, the leakage power is reduced by 10x-100x. Despite the fact that this is an expressive decrease, the delay increase in sub-threshold operation still is much more significant. Hence, a similar trade-off to the one observed for energy can be seen with the Leakage-Delay Product (LDP), check Figure 17. When V_{dd} hits the sub-threshold region, the delay increases exponentially and influences heavily in the Leakage-Delay Product. However, the leakage reduction is significant enough to achieve the LDP minimum

point in sub-threshold region. In fact, Figure 17 indicates that the LDP minimum point is also in the near-threshold region, $V_{DD} \sim 0.45V$.

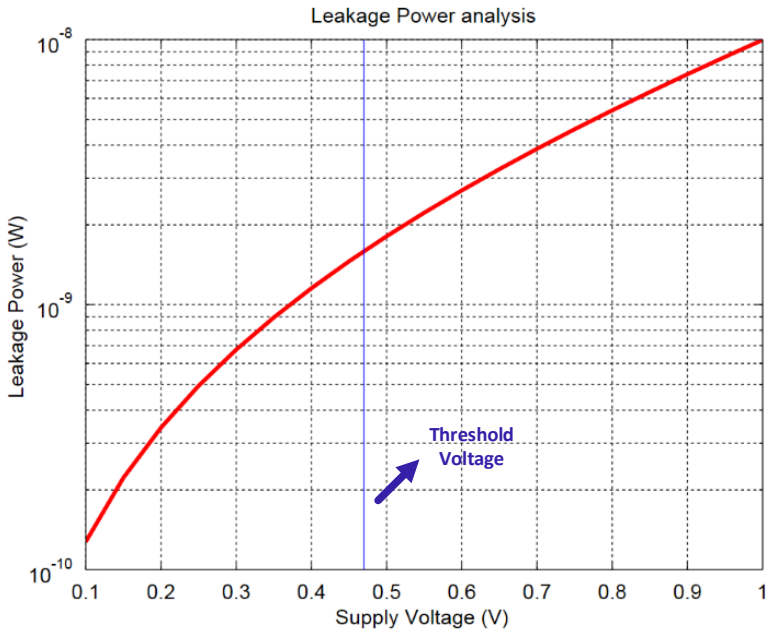


Figure 16 – Leakage power of the evaluated inverter gate.

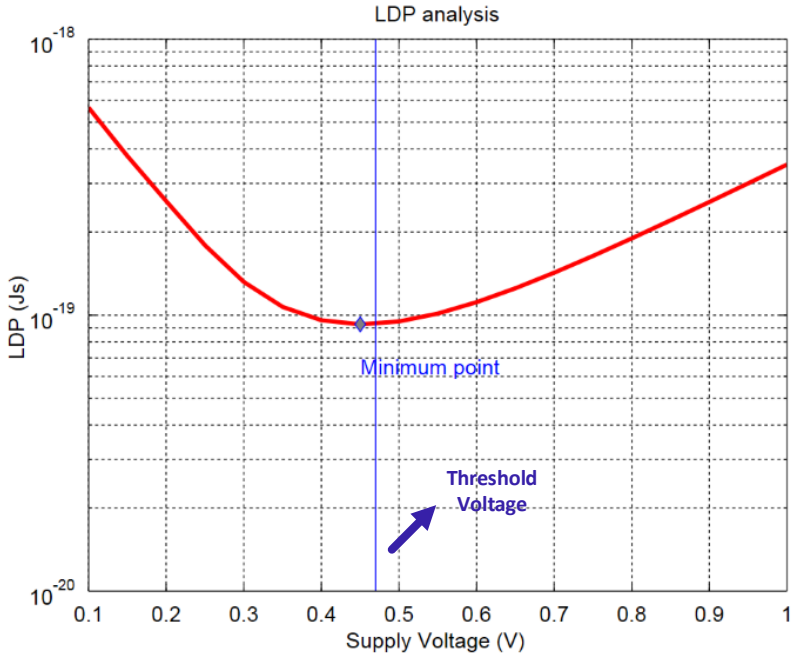


Figure 17 - Leakage-Delay Product of the evaluated inverter.

4.3 NAND and NOR Gates

With the initial results from the inverter analysis, it is important to take a step further and observe the impacts of voltage scaling in more complex CMOS gates. Thus, NAND and NOR gates are considered, as

they allow to evaluate the effect of stacking PMOS and NMOS transistors separately, providing a comprehensive analysis that can be extended to other logic gates. To follow the same criteria as the inverter scenario, all gates were selected from one of the libraries provided by the target technology vendor. With NAND gates, it is possible to see the impacts of stacking multiple NMOS transistors in series, whereas NOR gates shows the impacts of stacking multiple PMOS transistors in series. Table 3 shows the NMOS's width W_{NMOS} and the PMOS's width W_{PMOS} of each evaluated NAND/NOR gate. Note that all gates contains SVT transistors with $L_{min} = 0.06\mu$. Timing and power characteristics are covered as the previous evaluation.

Table 3 - NAND and NOR gates information.

Gate Name	n° inputs	Drive Strength	W_{NMOS}	W_{PMOS}
NAND2X2	**2**	X2	$\mathbf{0.2\mu}$	$\mathbf{0.28\mu}$
NAND3X2	**3**	X2	$\mathbf{0.2\mu}$	$\mathbf{0.28\mu}$
NAND4X2	**4**	X2	$\mathbf{0.22\mu}$	$\mathbf{0.28\mu}$
NOR2X2	**2**	X2	$\mathbf{0.2\mu}$	$\mathbf{0.35\mu}$
NOR3X2	**3**	X2	$\mathbf{0.25\mu}$	$\mathbf{0.72\mu}$
NOR4X2	**4**	X2	$\mathbf{0.25\mu}$	$\mathbf{0.94\mu}$

Figure 18 shows the delay and energy of the NAND2X2 and NOR2X2 gates while reducing V_{dd}. As expected, similarly to the inverter, both gates demonstrate an exponential delay degradation when V_{dd} hits below the threshold voltage, showing a degradation of 10-100x. Meanwhile, the transition energy of the NAND2X2 and NOR2X2 gates also shows a significant reduction when the gate is operating below de threshold voltage, albeit this reduction is not higher than the delay increase. This trade-off can be seen in Figure 19, which indicates the EDP of the NAND2X2 and NOR2X2 gates while reducing V_{dd}. When V_{DD} hits the near-threshold voltage region, the EDP of both gates reaches the minimum point (at V_{dd}~$0.6V$). However, when V_{dd} is below the threshold voltage, the EDP curve starts to rise again, a consequence of the huge delay degradation. Observing Figure 19, it is also possible to note that the NOR gate achieves higher EDP values when compared to the NAND gate. This limitation to the NOR2X2 gate is due to the fact that the NOR gates has PMOS transistors in series. As PMOS transistors has lower carrier mobility than the NMOS in the target technology, NOR gates achieves lower I_{on}/I_{off} ratios. Consequently, avoiding structure with stacking PMOS is advised.

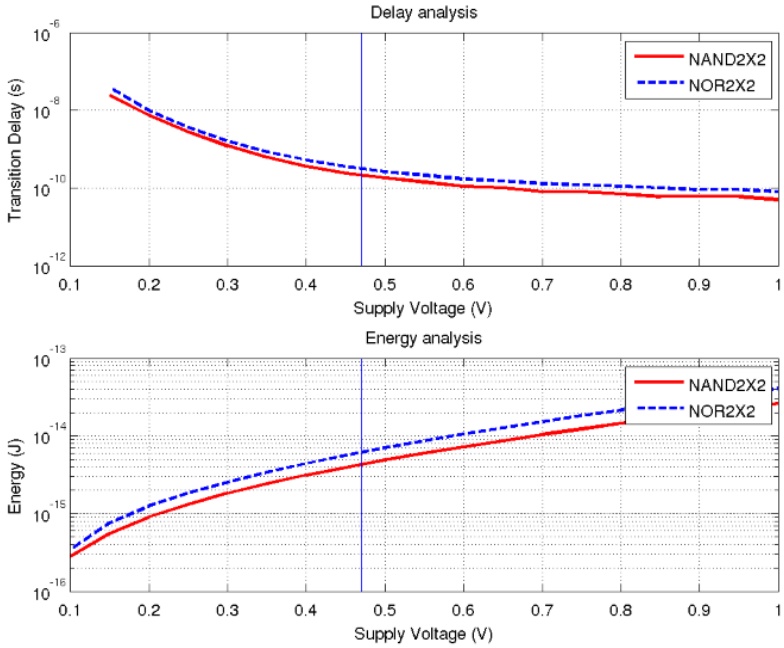


Figure 18 -Delay and Energy of the NAND2X2 and NOR2X2 gates while reducing V_{dd}.

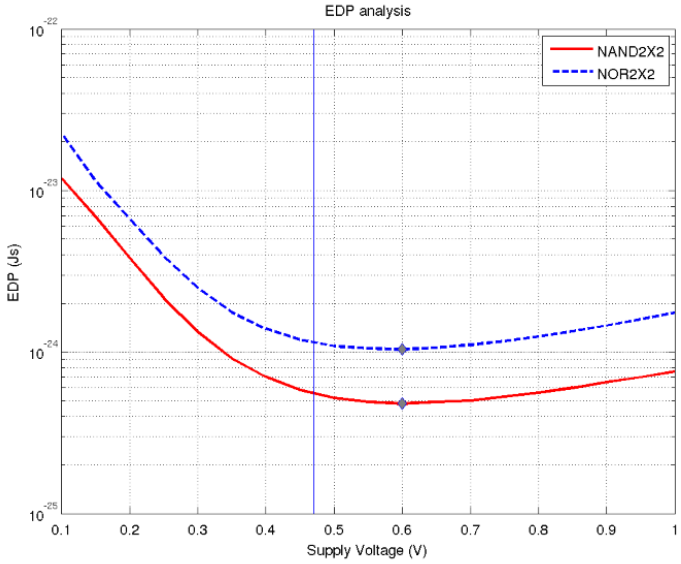


Figure 19 - EDP of the NAND2X2 and NOR2X2 gates while reducing V_{dd}.

Regarding leakage power, Figure 20 shows the leakage power reduction when V_{DD} is scaled. When V_{dd} reaches the threshold voltage, the leakage power is reduced almost 10x. In order to see the trade-off between leakage power reduction and delay degradation, Figure 21 indicates the LDP of the NAND2X2 and NOR2X2 gates while reducing V_{dd}. Both gates have their LDP minimum point at V_{dd}~$0.6V$, with the NAND2X2 gate reducing LDP by 1x and the NOR2X2 reducing by 1.5x. Recalling Equation (12), leakage power is directly proportional to V_{dd} and I_{off}. As variations in the latter are negligible, leakage reduces proportionally to the former. However, delay, as explained before, suffers higher degradation at subthreshold voltages. This explains why the sweet spot for LDP is in the near threshold region, where delay overheads are still not as large.

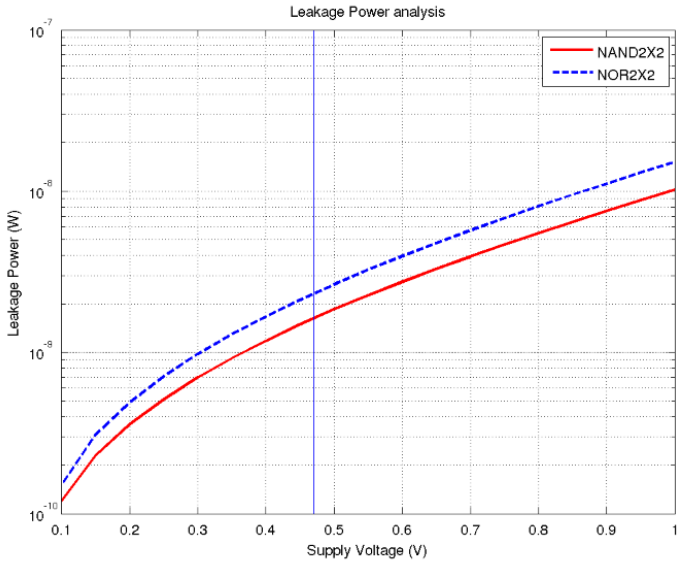


Figure 20 - Leakage power of NAND2X2 and NOR2X2 gates while reducing V_{DD}.

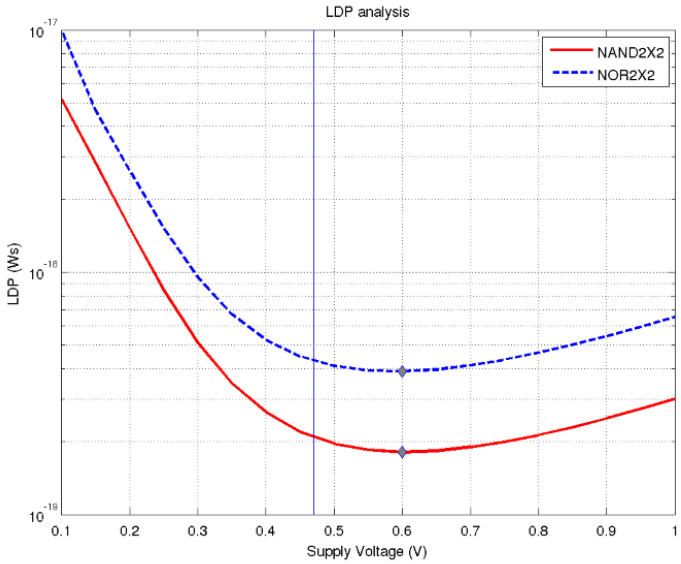


Figure 21 - LDP of NAND2X2 and NOR2X2 gates while reducing V_{DD}.

At last, gates with more than two inputs are considered. As the I_{on}/I_{off} ratio suffers great impact when V_{dd} is reduced, stacking more transistor compromises even further the I_{on}/I_{off} ratio of the gate. In a NAND gate, for example, if multiple NMOS transistor are stacked in series, the PDN may not able to discharge the output capacitance due to two main aspects: (1) the lower current I_{on} provided by the PDN network; (2) the PUN has multiple PMOS in parallel, increasing leakage current I_{off}. Figure 22 shows the EDP of a NAND gate with 2, 3 and 4 inputs. As the figure indicates, if more inputs are employed in the gate, higher EDP values are achieved and, consequently, higher delay and energy degradation. Note that even with 4 inputs, the NAND gate was able to operate with $V_{dd} = 0.1V$ at the cost of high delay penalty. In this way, it is advised to keep the number of inputs of cells reduced when designing circuits for voltage scale applications. Moreover, LDP is affected with different number of inputs. Observing Figure 23, the LDP is very similar for different number of inputs, indicating that it does not depend on that characteristic of the gate.

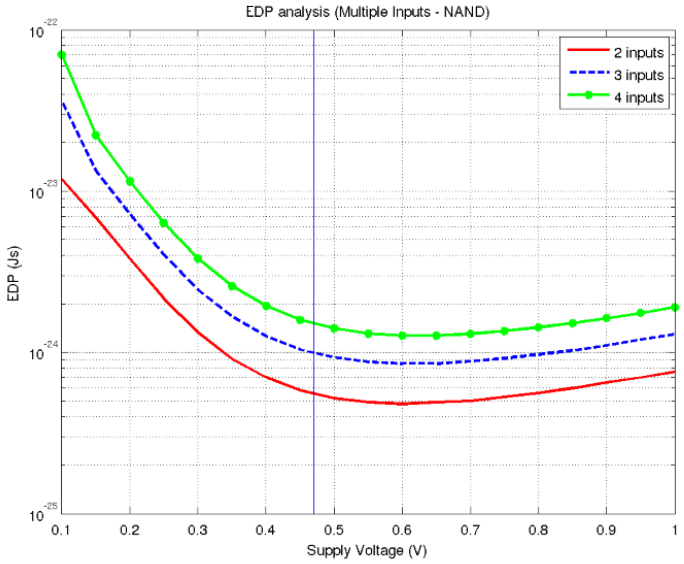


Figure 22 - EDP comparison: NAND with 2, 3 and 4 inputs.

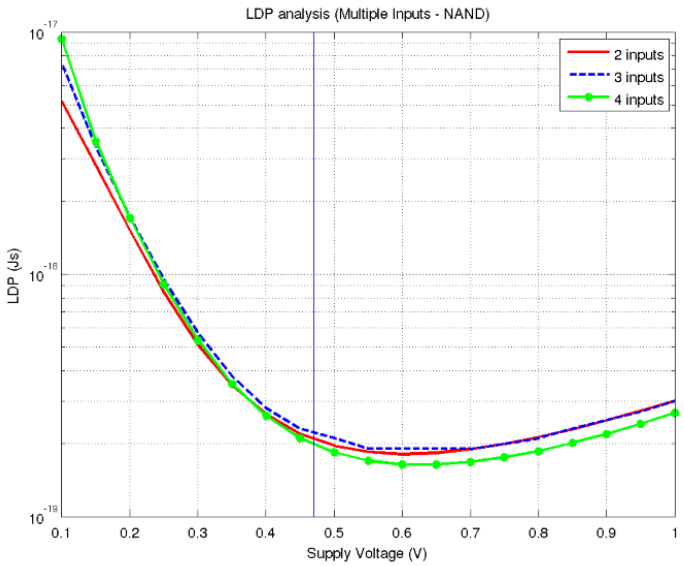


Figure 23 - LDP comparison: NAND with 2, 3 and 4 inputs.

5. VOLTAGE SCALING ON NCL CELLS

This section focuses in the description of the experimental environment developed in this work and the analysis of the obtained results for single NCL.

5.1 Experimental Environment

As mentioned in Section 3, current standard cell libraries are focused in either asynchronous support or in near/sub-threshold operation, but not both. This implies that components used in asynchronous circuits are not characterized and analyzed for lower supply voltages. Hence, an experimental environment was developed in order to enable the characterization of asynchronous in multiple supply voltages. Albeit the analysis in this work focused on NCL cells, other asynchronous cells can also be characterized with this environment. Currently, the experimental environment uses three main tools for cell characterization, simulation and plot generation. Figure 24 shows the flow adopted of the experimental environment, indicating the main tools and their respective position in the flow. The Voltage Scaling and MATLAB scripts shown in Figure 24 are implemented in Schell Script and are responsible to parse the tools' output files.

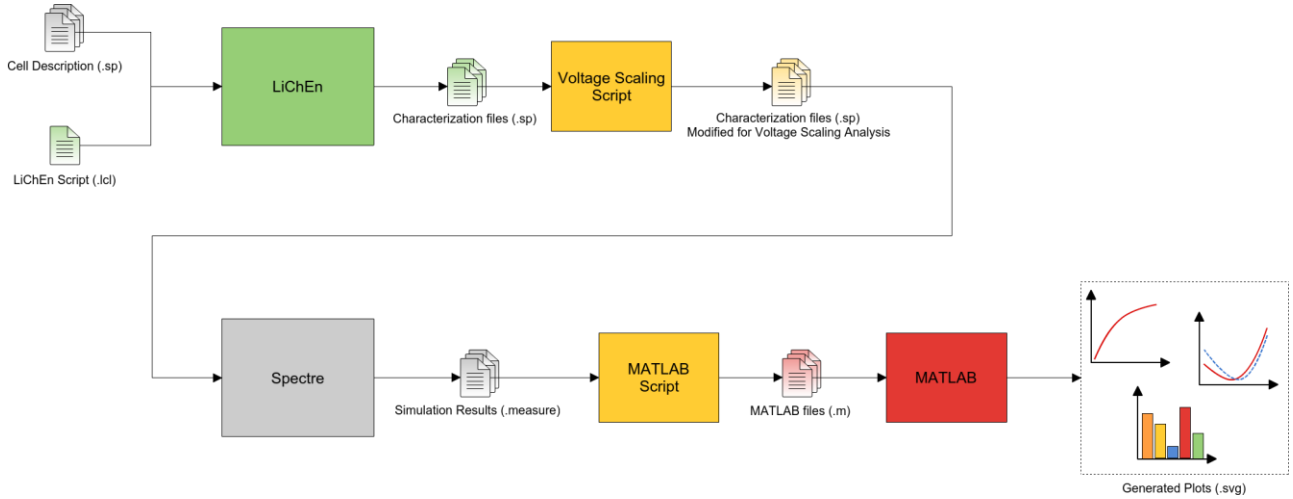


Figure 24 - Experimental environment flow.

For cell characterization, the LiChEn [MOR13c] tool from the ASCEnD Design Flow [MOR13a] is employed to generate simulation files with all the input-to-output transition arcs and static states of the target cells. This automation in the characterization step reduces the time and error of a manual analysis of each cell arc. Note that all characterization files from LiChEn uses the SPICE language. As multiple V_{DD} are needed for a voltage scaling analysis and LiChEn only supports one V_{dd} per characterization, it is necessary to invoke LiChEn several times, each time using a different V_{dd}. To do so, the flow represented in Figure 25 was adopted. Basically, the flowchart indicates that LiChEn is invoked multiple times and V_{dd} is decreased until it reaches a pre-defined minimum V_{dd} – This work uses the minimum $V_{dd} = 0.1V$ as it is the minimum supply voltage that achieves operation correctness of several NCL gates (considering a typical process corner and environment temperature as 25℃). Note that the flowchart indicates the file modifications performed by the

Voltage Scaling script. Executing the adopted flowchart creates a directory hierarchy, which can be seen in Figure 26. This directory hierarchy arranges all characterization files in three levels: libraries, cells and arcs/states. At the first level (library), the directories arrange the files according to the adopted V_{dd}. As the name suggests, the second level (cell) separates the files according the cell. The former level (arc) separates individually each arc, displaying whether the arc is dynamic, internal or static state.

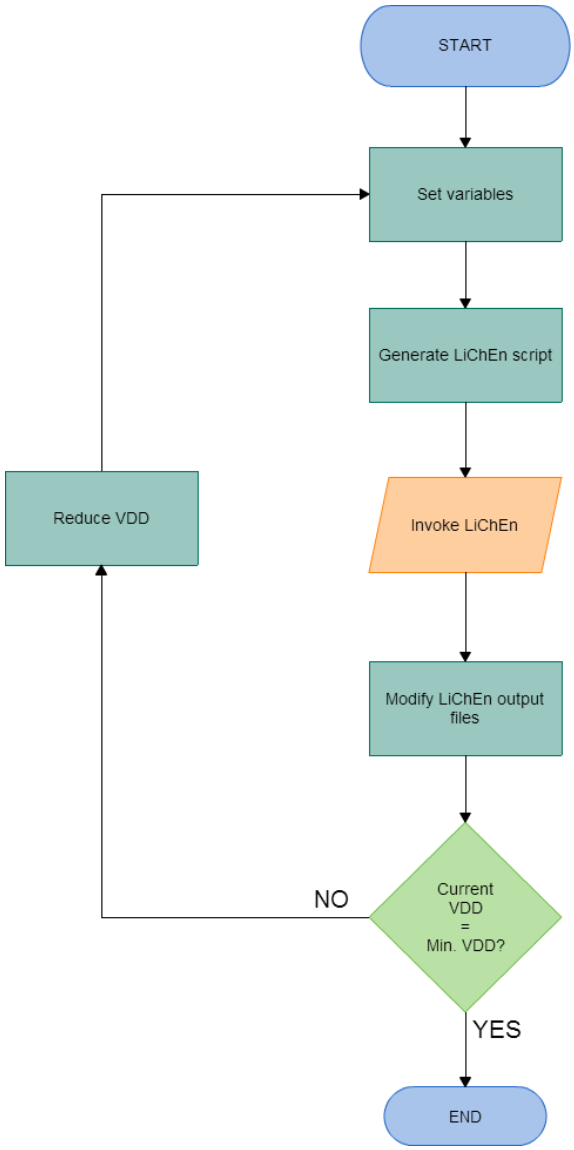


Figure 25 - Flowchart for voltage scaling characterization. LiChEn is invoked several times, each time generating all characterization files for the selected V_{dd}.

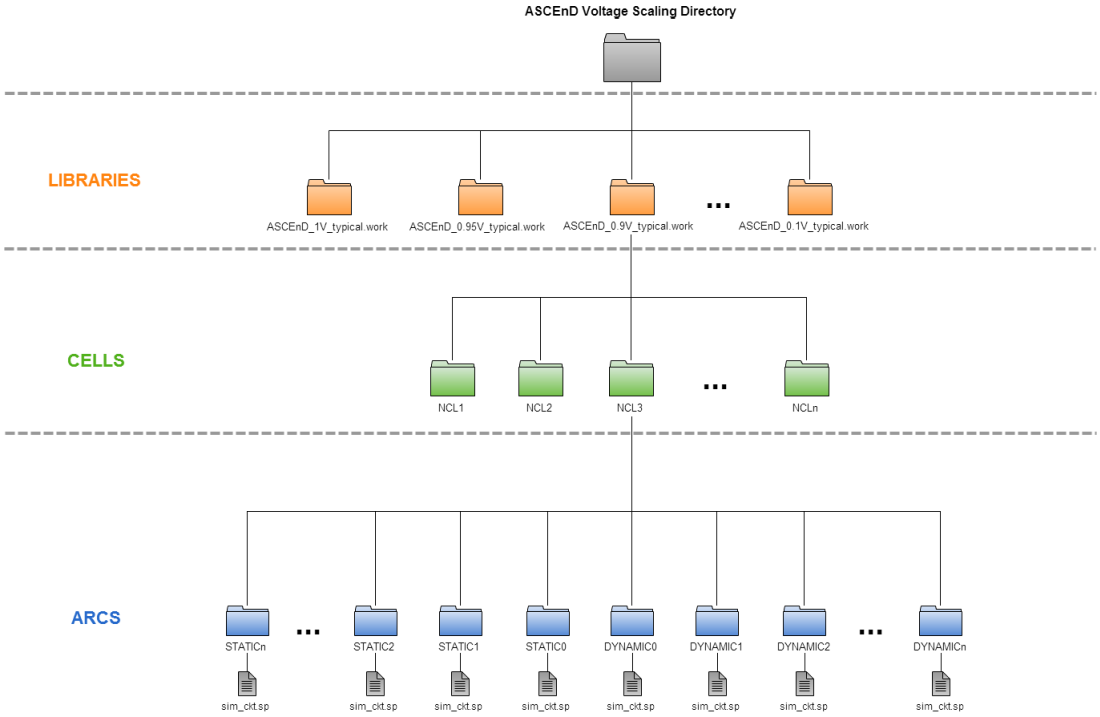


Figure 26 - Directory hierarchy of the characterization files.

Although LiChEn automates significantly the characterization process, there are some issues in near/sub-threshold operation that LiChEn disregard. Initially, LiChEn uses a ramp slope as an input in dynamic and internal arcs. This type of slope is acceptable in super-threshold operation. However in near/sub-threshold operation, the use of a ramp slope as an input is not realistic due to the low current at the input and output of the cells, which generate long slopes. In addition, LiChEn dismiss the presence of noise in the input. In near/sub-threshold operation, again, input noise must be considered since certain cells may not be able to represent a high logic level as 100% of V_{dd} or a low logic level as 0% of V_{dd}. Thus, the characterization files provided by LiChEn are modified by the Voltage Scaling script, indicated in Figure 24.

The Voltage Scaling script accesses all characterization files provided by LiChEn and adjusts them to support a realistic input slope and input noise. For better illustration, Figure 27 (a) shows an example of an original characterization file provided by LiChEn, while Figure 27 (b) shows the same characterization file but with the modification from the Voltage Scaling script. For a realistic input slope, a 2-stage inverter chain was added in each input. All added inverters have the same drive strength as the NCL gate. To introduce input noise, the waveform of each input source were modified such as the high logic level is represented by 90% of V_{dd} and low logic level by 10% of V_{dd}. After modifying all characterization files, each arc can be simulated for power and delay extraction.

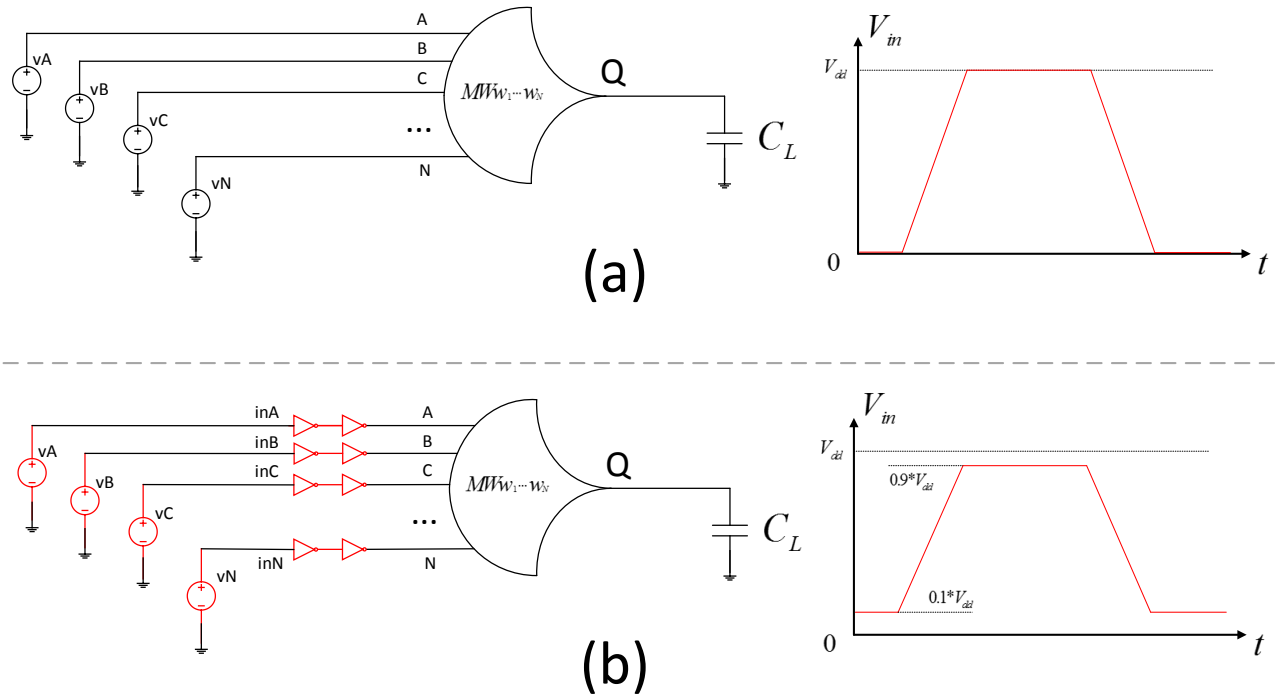


Figure 27 - Example of simulation setup of (a) original file and (b) modified file.

Regarding the simulation and plot generation step, the Spectre simulator and MATLAB are used, respectively. All results from the simulation are filtered by MATLAB script (Figure 24), which also generates MATLAB files from a template. These MATLAB files can be executed in MATLAB and generates plots regarding transition and propagation delay, transition energy, EDP, leakage power and LDP.

5.2 Experiments and Discussion

With the experimental environment presented in subsection 5.1, it is possible to select a basic set of NCL gates for characterization and evaluation. In total, nine NCL gates were selected and each NCL gate employed different characteristcs. Table 4 shows all selected NCL gates and their respective topology, family, function, number of inputs and drive strength. Regarding their transistor implementation, referred here as topology, NCL gates may be realized with several distinct topologies and electrical characteristics. Propositions exist that use differential logic [YAS10], multi-threshold CMOS technologies [AKE87] [BAI08], static, semi-static or dynamic topologies [SOB98] to design NCL gates. For simplicity, this work restrains attention to static and semi-static implementations only. Figure 28 (a) and (b) illustrate the semi-static and static topologies.

In the semi-static topology, the PDN and PUN are respectively represented by the blocks SET and RESET. The hysteresis function is implemented by the output inverter *P1/N1* – responsible for displaying the output *Q* – and the feedback inverter *P0/N0*. This structure ensures that the output keeps its value in case neither SET nor RESET functions are asserted. Usually, the feedback inverter has minimum size, due to the fact that it is only used for maintaining the output stable. Note that in some cases, this topology requires careful transistors sizing due to the resistance imposed by the feedback inverter. The static topology employs a structure similar to that of the semi-static topology. However, it adds two structures that control the feedback inverter: HOLD0 and HOLD1. The HOLD0 block is the complement of RESET and the HOLD1

block is the complement of SET. These additions control the feedback inverter, turning it off when the output is switching. Hence, the static topology reduces interferences while switching the output and enables a less constrained design and transistors sizing. This makes the static topology more suited for voltage scaling applications and throughout this work all NCL cells are assumed to employ a static topology.

The NCL gates also can be classified in two families: NCL and NCL+ [MOR13d]. The traditional NCL is suited for the RTZ protocol, whereas NCL+ were implemented to support the RTO protocol. Moreover, NCL and NCL+ gates may employ a positive unate or negative unate function. The only difference between these functions in NCL gates is the polarity of the output node. Recalling Figure 28, take notice that both NCL topologies (a) and (b) have the signal *Q* and *Qn*, where *Qn* is the negative value of *Q*. If a NCL gate has a positive unate function, then the output node is *Q*. However, if a NCL gate has a negative function, then *Qn* is used as output node. Evaluating both scenarios allows a more comprehensive analysis of the effects of voltage scaling on NCL cells, as they can implement either positive or negative unate functions, as described in [MOR14a]. Also, the choice for different functions allows assessing how NCL cells with different number of inputs behave as voltage is scaled. The naming style of the gates is defined as follows:

- The prefix *ST_* indicates a static topology;
- *NCL* indicates that the gate belongs to the NCL family and *NCLP* that it belongs to the NCL+ family. A preceeding *I* identifies that the function is negative unate, otherwise it is positive unate;
- After the family comes the threshold function: first a number to indicate the threshold, followed by the weights identifier *W*, which is followed by the weights of all inputs;
- Next, the *OF* identifier is followed by the number of inputs in the gate;
- Finally the driving strength is represented as a number preceded by the identifier *X*.

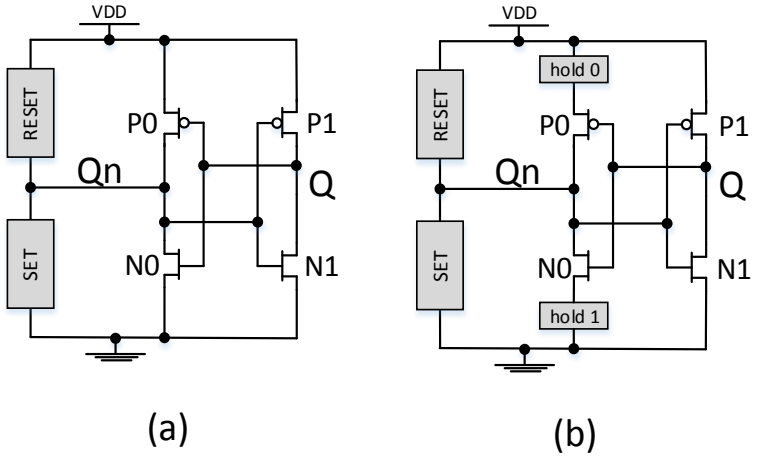
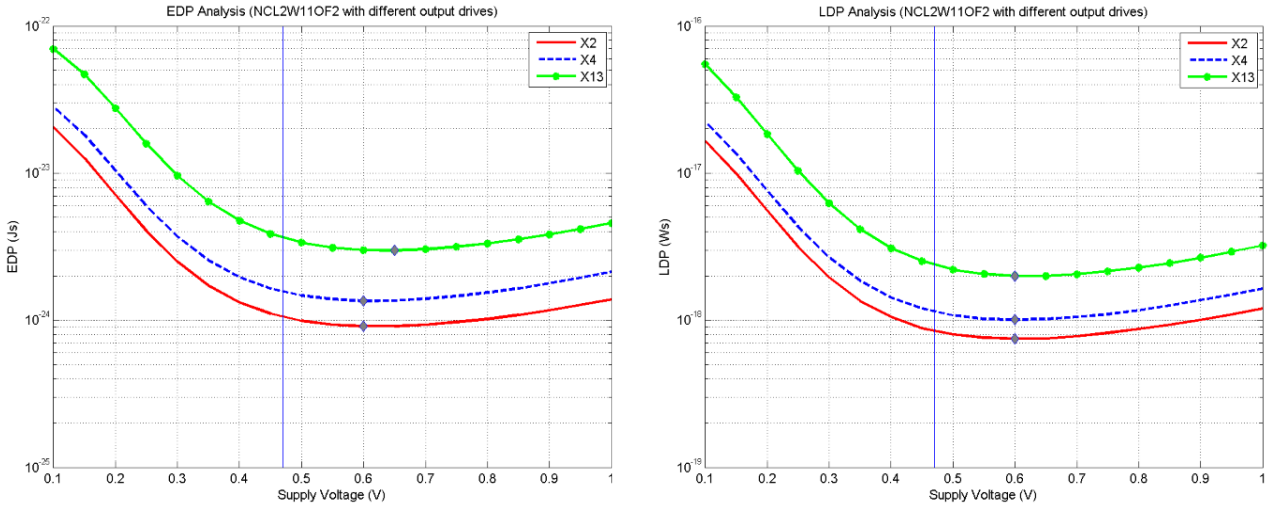


Figure 28 - Semi-static (a) and static (b) topologies.

Table 4 - Selected NCL gates with their respective topology type, number of inputs and drive strength.

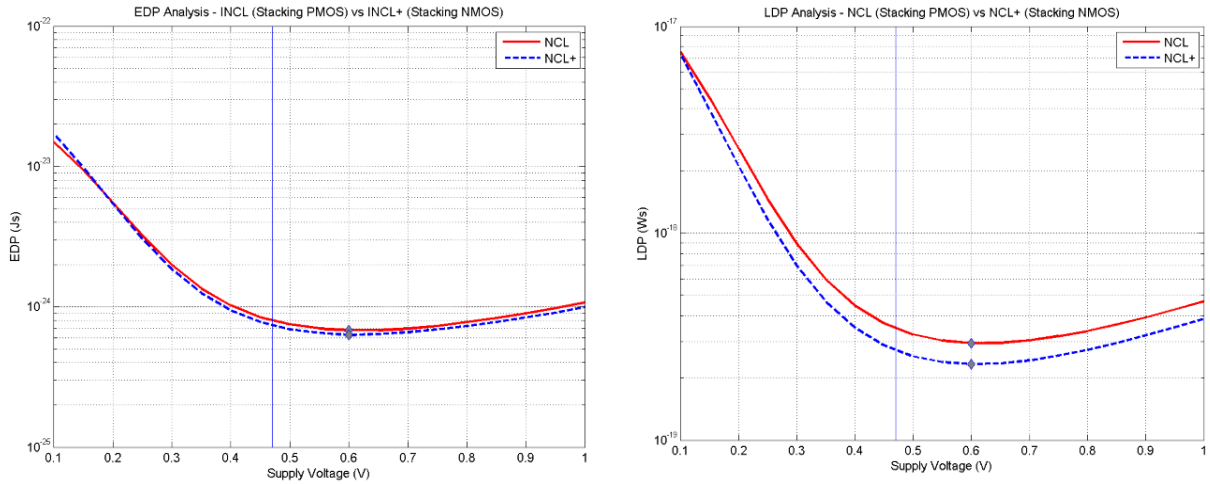
Gate Name	Unateness	Topology	Family	# inputs	Drive Strength
ST_INCL1W11OF2X4	Negative	Static	NCL	2	X4
ST_INCL2W11OF2X4	Negative	Static	NCL	2	X4
ST_NCL2W11OF2X2	Positive	Static	NCL	2	X2
ST_NCL2W11OF2X4	Positive	Static	NCL	2	X4
ST_NCL2W11OF2X13	Positive	Static	NCL	2	X13
ST_NCL3W111OF3X4	Positive	Static	NCL	3	X4
ST_NCL5W2211OF4X4	Positive	Static	NCL	4	X4
ST_INCLP1W11OF2X4	Negative	Static	NCL+	2	X4
ST_NCLP5W2211OF4X4	Positive	Static	NCL+	4	X4

After selecting the basic set of NCL gates, the developed voltage scaling environment characterized the NCL gates. The collected results were used to perform three analyses: driving strength, family and function dependence. Figure 29 (a) and (b) show the EDP and LDP of the ST_NCL2W11OF2 gate with three driving strengths: X2, X4 and X13 respectively. This allows understanding how different driving strengths are affected as the voltage is scaled. As expected, the use of higher drive strengths leads to larger energy and leakage power, increasing the EDP and LDP. However, the trend of the charts is similar for the different driving strengths, displaying a low dependency of driving strength in voltage scale trade-offs. Regarding NCL families, Figure 30 indicates the comparison between NCL and NCL+ families. Similarly to the previous analysis, EDP (c) and LDP (d) are considered. Note that the NCL achieves both higher EDP and LDP, principally for LDP results. This can be explained due to the fact that NCL gates usually stack PMOS transistors in series, while NCL+ gates stakes NMOS transistor. As a consequence, NCL+ should achieve higher I_{on}/I_{off} ratios, similarly to the NANDs and NORs analysis. Finally, Figure 31 (e) and (f) shows the comparison between positive and negative unate NCL gates. As the figure suggests, NCL gates employing negative unate functions achieve lower EDP and LDP. However, different from the previous cases, the lower EDP and LDP achieved by the negative unate function is not a consequence of a lower delay. In fact, observing Figure 32, it is possible to see that the negative unate function has higher delay and lower transition energy, both consequences of its structure, which employs only one level of logic and does not need a big output inverter.



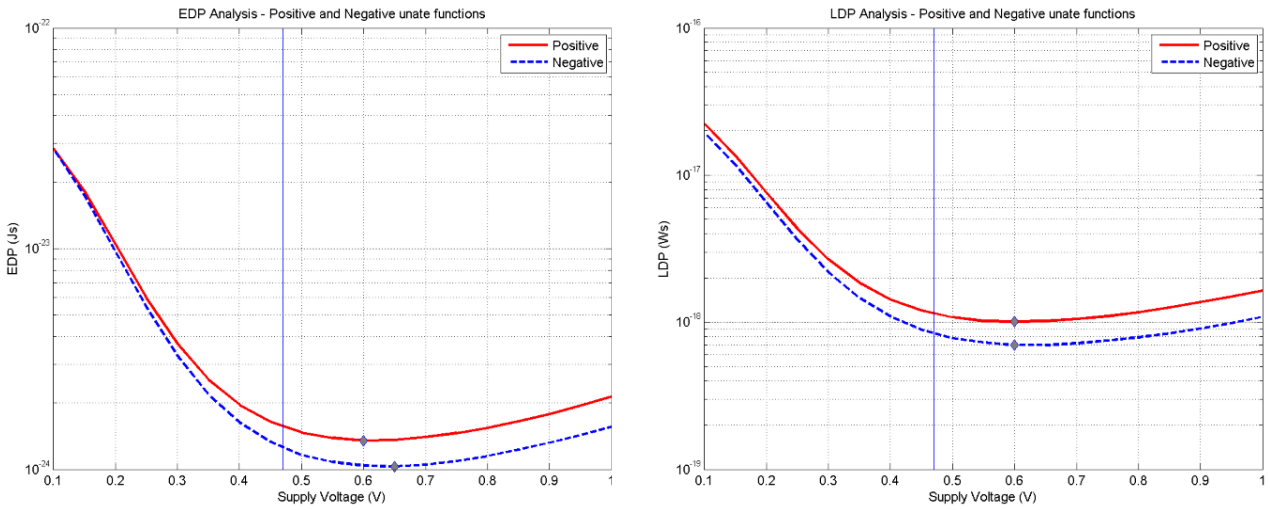
(a) (b)

Figure 29 – EDP (a) and LDP (b) of ST_NCL2W11OF2 gate with multiple drive strengths (X2, X4 and x13).



(c) (d)

Figure 30 – NCL and NCL+ comparison. EDP (c) and LDP (d) of ST_NCL1W11OF2 and ST_NCLP1W11OF2 gates with drive strength X4.



(e) (f)

Figure 31 - Function comparison. EDP (e) and LDP (f) of ST_NCL2W11OF2 and ST_INCL2W11OF2 with drive strength X4.

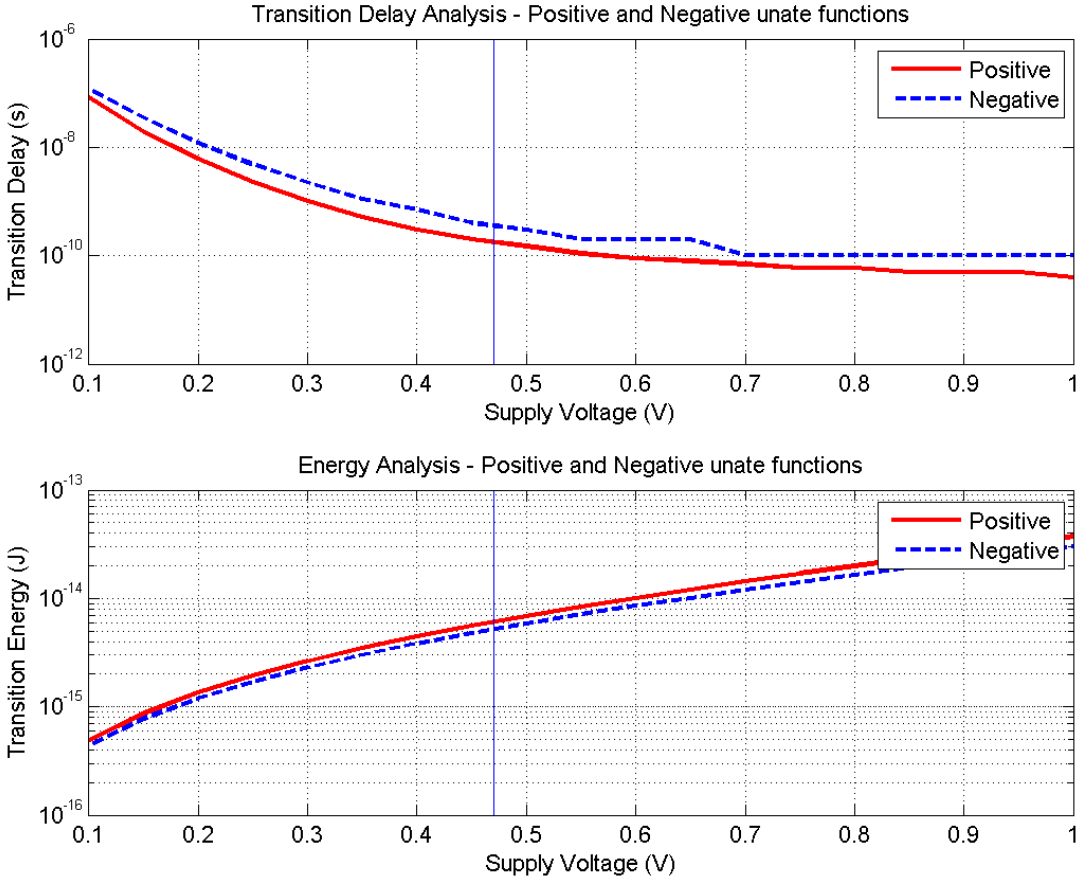


Figure 32 - Transition delay and energy of of ST_NCL2W11OF2 and ST_INCL2W11OF2 with drive strength X4.

6. A FULL-ADDER CASE STUDY

Another experiment allowed assessing the effects of voltage scaling on NCL cells on more complex structures. Accordingly, an 8 bits Kogge-Stone adder was designed and mapped to the cells of the ASCEnD library. The adder is similar to the one described in [MOR14c] and its basic the block diagram is showed in Figure 33. Note that this block diagram represents its classic single-rail implementation, for the sake of simplicity. For the synthesis process it was translated to dual-rail and mapped using the approach proposed in [MOR14a].

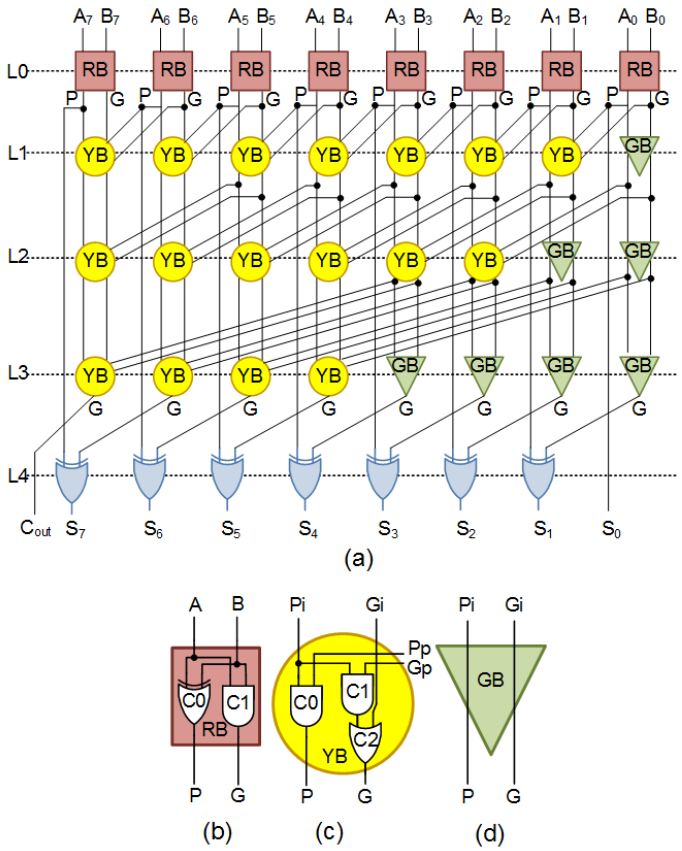


Figure 33 – 8 bits Kogge Stone (a) block diagram and basic blocks that compose it: (b) red box, (c) yellow box and (d) green box. Adapted from [MOR14c].

The choice for this adder as a case study was due to its availability in the GAPH research group and has already been used and validated for NCL synthesis. The choice for a small version (8 bits) has been taken in order to reduce synthesis and simulation time. Nevertheless, this does not compromise the generality of the results presented here because the main objective of evaluating such a circuit is to provide a more realistic environment for evaluating the NCL cells. Accordingly, in such an environment, all cells (other than those fed by the primary inputs) are driven by other NCL cells and scaling the voltage of the whole circuit impacts on all cells at the same time. This is different from the controlled environment presented in Section 5, where each cell was evaluated separately.

As it is discussed in [MOR14a], the circuit was synthesized using both NCL and NCL+ gates. Timing constraints were iteratively tightened until there was 0 slack. This enabled a scenario with no underutilization of the driving strength of the NCL cells. Synthesis relied on the Cadence Framework using the RTL Compiler. After synthesis, the tool automatically exported the mapped netlist was to a Verilog description. This netlist was then exported to a Spice description using Mentor Calibre. This description was employed in an analog mixed signal environment for validation as well as the circuit's performance and power analysis. As showed in Figure 34, this environment had a random data producer designed in SystemC in order to provide stimuli to the case study adder. The generated outputs were then computed by a data consumer and analyzer, also designed in SystemC. The interface between the Spice netlist of the case-study and the digital testbench was realized using analog to digital and digital to analog converters (ADC and DAC) described in VHDL-AMS, available in the Cadence Framework. The environment relied on Cadence analog and digital simulators, Spectre and NCSim, respectively, both from the Incisive suite.

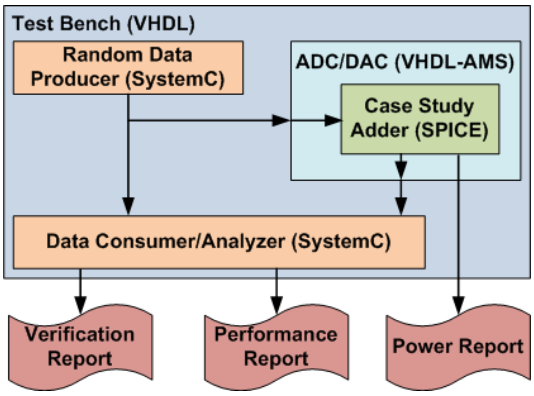


Figure 34 – Environment for validation and performance and power analysis of the case study circuit.

During simulation, the digital testbench generated a verification report to ensure the operation correctness after the synthesis process together with a performance report for obtaining latency metrics of the case study. Also, power reports were automatically generated by the analog simulator through .measure constructions in Spice language. This allowed for good precision of the collected data. Using this environment, a voltage scaling setup was realized, varying the voltage in the analog simulator, with the required configurations in the ADC and DAC modules, from 0.1V to 1V in steps of 50mV. For each run, all data was collected and stored for comparison. Note that the same random generated stimuli set was used for all runs. The obtained results are summarized in the charts of Figure 35, displaying three characteristics of the circuit as a function of its voltage supply level: propagation delay, energy per operation (EPO) and energy delay product (EDP).

The minimum operating voltage for the case study was 0.15V. Below this point, the circuit did not operate correctly, due to delay variations on the cells, as discussed in Section 4. As the charts show, degradation in the propagation delay of the circuit is more accentuated at sub-threshold voltage levels, as previously indicated by the NCL evaluations. Also, energy savings are very substancial until 0.2V. However, at the minimum operating voltage (0.15V) there is a modification in the trend of the chart and it consumers more EPO than at 0.2V. This is explained by the fact that at this point leakage power is highly significant and the cells have longer short circuit periods while switching, until the active PUN/PDN network over powers the cut off PDN/PUN network. Taking for example 0.2V, it presents a reduction of roughly 14x in EPO for an increase of approximately 42x in the propagation delay is presented.

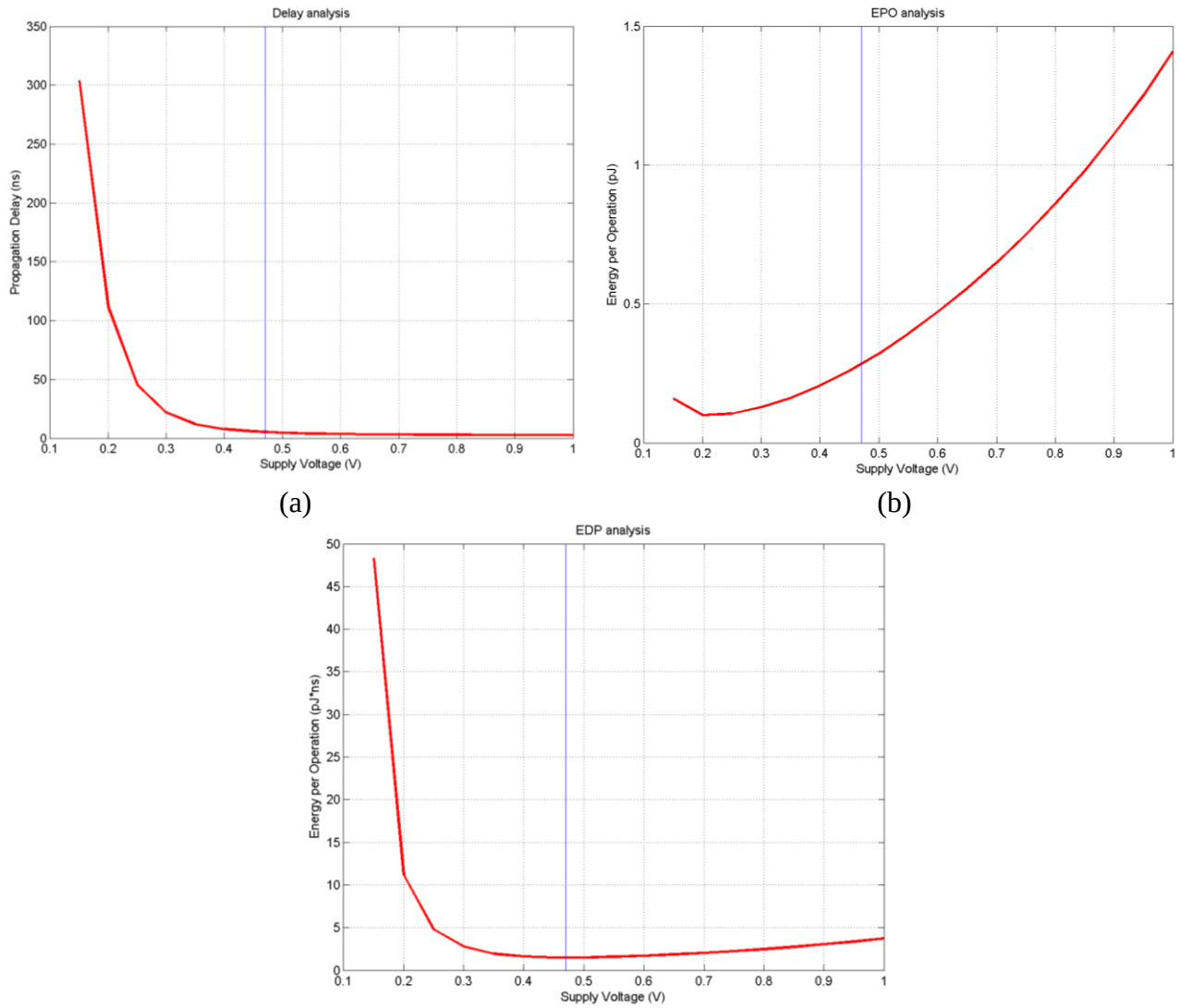


Figure 35 – Results for the 8 bit Kogge-Stone case study: (a) propagation delay, (b) energy per operation and (c) energy delay product.

For ultra-low-power applications the 14x reduction on EPO can look very attractive. However the high price of 42x the delay can be prohibitively constraining. The EDP metric allows an analysis of the sweet spot for operating voltage taking into account both propagation delay and EPO. Accordingly, as Figure 35 (c) shows, this point is between 0.45V and 0.5V, in the near-threshold region. Taking for example 0.45V, its EDP is roughly 2.5x smaller than the EDP at 1V. This translates to a 5.4x reduction on EPO for a cost of 2.2x in propagation delay. Albeit the reduction in energy is not as large as that at minimum operating voltage, the costs in propagation delay are much smaller. These results corroborate the analyses presented in previous sections and indicate that the most efficient delay and energy voltage for NCL circuits is in the near-threshold region.

7. CONCLUSIONS

During the development of this End of Term work, two main topics were covered: voltage scaling and asynchronous circuits. Studying these topics gave to the Author not only the opportunity to explore new topics that are not covered during the undergraduate course, but also made him review basic microelectronic aspects, as well as digital circuit principles. Furthermore, in order to perform the tasks during this End of Term work, it was necessary to explore available EDA tools from Cadence, as well as the ASCEnD Design Flow tools. This enabled a better understanding regarding the use and structure of a VLSI design flow. With the developed tasks, the Author demonstrates the effects of voltage scaling in a device level and the huge delay/power trade-offs that occur when voltage scaling is applied in basic CMOS gates. Besides the basic CMOS gates, NCL gates are evaluated as well. To automate this process, an experimental environment is presented, incorporating the LiChEn tool in order to characterize and evaluate the trade-offs of each NCL gate. Among these trade-offs, the obtained results indicate four main observations: (1) better delay/power trade-offs are achieved in near-threshold operation; (2) sub-threshold operation impose a heavy delay increase; (3) structures with PMOS transistors in series should be avoided when voltage scaling is considered and (4) restricting gates with more than 3 inputs is suggested. As a last evaluation, an 8 bits Kogge-stone adder is generated using NCL gates from the ASCEnD library in a VHDL-AMS environment. The results obtained from this last evaluation corroborate to the previous analyses.

7.1 Future work

The developed work represents an initial step of the integration between voltage scaling and asynchronous circuits. In order to introduce the voltage scaling support in the ASCEnD Design Flow, some aspects developed into the experimental environment may be added into the LiChEn tool. Nevertheless, a deeper analysis regarding near/sub-threshold operation may indicate optimizations that could be incorporated into the design flow or even into the gates' structure.

REFERENCES

[AKE87] Akers, L. A.; Sugino, M.; Ford, J. M. “Characterization of the Inverse-Narrow-Width Effect”. IEEE Transactions on Electron Devices, Vol. 34, 1987, pp: 2476-2484.

[BAI08] Bailey, A. D.; Di, J.; Smith, S.C.; Mantooth, H. A. “Ultra-low power delay-insensitive circuit design”. In: International Midwest Symposium on Circuits and Systems (MWSCAS), 2008, pp. 503-506.

[BEE10] Beerel, P. A.; Ozdag, R. O.; Ferretti, M. “A Designer’s Guide to Asynchronous VLSI”. Cambridge University Press, 2010, 339p.

[BEE11] Beerel, P.; Dimou, G.; Lines, A. “Proteus: An ASIC Flow for GHz Asynchronous Designs,” IEEE Design & Test of Computers, 28(5), Sep. 2011, pp. 36–51.

[CAL05] Calhoun, B. H.; Wang, A.; Chandrakasan, A. “Modeling and sizing for minimum energy operation in subthresold circuits”. IEEE Journal of Solid-State Circuits, 40(9), Sep. 2005, pp. 1778-1786.

[CHA10] Chang, I.; Park, S.; Roy, K. “Exploring asynchronous design techniques for process-tolerant and energy-efficient sub-threshold operation”. IEEE Journal of Solid-State Circuits, 45(2), Feb. 2010, pp. 401–410.

[FAN96] Fant, K. M.; Brandt, S. A. “NULL convention logic: a complete and consistent logic for asynchronous digital circuit synthesis”. In: International Conference on Application Specific Systems, Architectures and Processors (ASAP), 1996, pp. 261-273.

[FER04] Ferretti, M. “Single-Track Asynchronous Pipeline Template”. PhD Thesis, Faculty of Electrical Engineering, University of Southern California, 2004, 170p.

[FOL05] Folco, B., Brégier, V., Fesquet, L. and Renaudin , M., “Technology mapping for area optimized quasi delay insensitive circuits,” In: IFIP Conference on Very Large Scale Integration Systems, 2005, pp. 55–69

[GUA10] Guan X.; Liu Y.; Yang Y. “Performance Analysis of Low Power Null Convention Logic Units with Power Cutoff”. In: Asia-Pacific Conference on Wearable Computing Systems (APWCS), 2010, pp. 55-58.

[JEO08] Jeong, C.; Nowick, S.M., "Technology Mapping and Cell Merger for Asynchronous Threshold Networks," In: IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol.27, no.4, 2008, pp. 659-672

[HAU95] Hauck, S. “Asynchronous Design Methodologies: An Overview”. Proceedings of the IEEE, 83(1), Jan. 1995, pp. 69-93.

[IEO13] Ieong, C.-I.; Li, M.; Law, M.-K.; Mak, P.-I.; Vai, M.-I.; Mak, P.-U.; Wan, F.; Martins, R. P. “Standard cell library design with voltage scaling and transistor sizing for ultra-low-power biomedical applications”. In: International Conference of Electron Devices and Solid-State Circuits (EDSSC), 2013, pp. 1-2

[ITR11] ITRS. “International Technology Roadmap for Semiconductors 2011 Edition”. Design Chapter. Available at http://www.itrs.net/Links/2011ITRS/2011Chapters/2011Design.pdf.

[KIM13] S. Kim et al. “Razor-lite: A side-channel error-detection register fortiming-margin recovery in 45nm SOI CMOS,” In: IEEE International Solid-State Circuits Conference, 2013, pp. 264–266.

[KOM02] Kondratyev, A.; Lwin, K., "Design of asynchronous circuits using synchronous CAD tools," In: IEEE Design & Test of Computers, vol.19, no.4, 2002, pp. 107-117

[KUR04] Kursun, V. “Supply and Threshold Voltage Scaling Techniques in CMOS Circuits”. PhD Thesis, Department of Electrical and Computer Engineering, University of Rochester, 2004, 330p.

[KWO06] Kwong, J. Y. S. “A Sub-threshold Cell Library and Methodology”. Master Thesis, Faculty of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2006, 102p.

[LIA12] Ran Liao; Hutchens, C., "Digital Circuit Design for Robust Ultra-Low-Power Cell Library Using Optimum Fingers," IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), pp.446-449, 2012

[LIG00] Ligthart, M.; Fant, K.; Smith, R.; Taubin, A.; Kondratyev, A., "Asynchronous design using commercial HDL synthesis tools," In: Sixth International Symposium on Advanced Research in Asynchronous Circuits and Systems (ASYNC), 2000, pp. 114-125.

[LIU12] Liu, B.; Ashouei, M.; Huisken, J.; de Gyvez, J. P. “Standard cell sizing for sub-threshold operation”. In: Design Automation Conference (DAC), 2012, pp. 962-967.

[LOD12] Lodhi, F. K.; Hasan, O.; Hasan, S. R.; Awwad, F. “Modified null convention logic pipeline to detect soft errors in both null and data phases”. In: International Midwest Symposium on Circuits and Systems (MWCAS), 2012, pp. 402-405.

[LOT11] Lotze, N.; Manoli, Y. “A 62mV 0.13µm CMOS standard-cell-based design technique using schmitt-trigger logic”. In: International Solid State Circuits Conference (ISSCC), 2011, pp. 340-342.

[MAR90] Martin, A. J. “The limitations to delay-insensitivity in asynchronous circuits”. In: 6th Conference on Advanced Research in VLSI (AUSCRYPT), 1990, pp. 263-278.

[MAR06] Martin, AJ.; Nystrom, M., "Asynchronous Techniques for System-on-Chip Design," Proceedings of the IEEE , vol.94, no.6, pp. 1089-1120, 2006

[MAR11] Maricau, E.; Gielen, G., "Transistor aging-induced degradation of analog circuits: Impact analysis and design guidelines." In: Proceedings of the ESSCIRC, 2011 , pp. 243-246 [MAU03]Maurine, P.; Rigaud, J.; Bouesse, F.; Sicard, G.; Renaudin, M. “Static Implementation of QDI Asynchronous Primitives”. In: International Workshop on Power and Timing Modeling, Optimization and Simulation (PATMOS), 2003, pp. 181-191.

[MOR11a] Moreira, M.; Oliveira, B.; Pontes, J.; Calazans, N., "A 65nm standard cell set and flow dedicated to automated asynchronous circuits design," In: IEEE International SOC Conference (SOCC), 2011, pp. 99-104.

[MOR11b] Moreira, M.; Oliveira, B.; Pontes, J.; Moraes, F.; Calazans, N., "Adapting a C-element design flow for low power," In: 18th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2011, pp. 45-48.

[MOR12] Moreira, M. T.; Guazzelli, R. A.; Calazans, N. L. V. “Return-to-one protocol for reducing static power in C-elements of QDI circuits employing m-of-n codes”. In: Symposium on Integrated Circuits and Systems Design (SBCCI), 2012, pp. 1-6.

[MOR13a] Moreira, M. T.; Calazans, N. L. V. “Design of standard-cell libraries for asynchronous circuits with the ASCEnD flow”. In: IEEE Annual Computer Society Symposium on VLSI (ISVLSI), 2013, pp; 217-218.

[MOR13b] Moreira, M. T.; Oliveira, C. H. M.; Porto, R. C.; Calazans, N. L. V. “Design of NCL gates with the ASCEnD flow”. In: Latin American Symposium on Circuits and Systems (LASCAS), 2013, pp. 1-4.

[MOR13c] Moreira, M. T.; Oliveira, C. H. M.; Calazans, N. L.V.; Ost, L. C. “LiChEn: Automed Electrical Characterization of Asynchronous Standard Cell Libraries”. In: Euromicro Conference on Digital System design (DSD), 2013, pp. 933-940.

[MOR13d] Moreira, M.T.; Oliveira, C.H.M.; Porto, R.C.; Calazans, N.L.V., "NCL+: Return-to-one Null Convention Logic," IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS), 2013, pp. 836-839.

[MOR14a] Moreira, M.; Neutzling, A.; Martins, M.; Reis, A.; Ribas, R.; Calazans, N., "Semi-custom NCL Design with Commercial EDA Frameworks: Is it Possible?," In: 20th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC), 2014, pp. 53-60.

[MOR14b] Moreira, M. T.; Arendt, M. E.; Ziesemer JR., A.; Reis, R. A. da L., Calazans, N. L. V. ASCEnD2: Automated Asynchronous Cells Design. In: 27th Symposium on Integrated Circuits and Systems Design (SBCCI), 2014. pp. 183-188.

[MOR14c] Moreira, M. T., Trojan, G., Moraes, F. G., Calazans, N. L. V. “Spatially Distributed Dual-Spacer Null Convention Logic Design.” In: Journal of Low Power Electronics (Print), 2014. pp. 313-320.

[MYE01] Myers, C. “Asynchronous Circuit Design”. New York: John Wiley & Sons, Inc. 2001, 422p.

[PAR14] Parsan, F.A.; Al-Assadi, W.K.; Smith, S.C., "Gate Mapping Automation for Asynchronous NULL Convention Logic Circuits," In: IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol.22, no.1, 2014, pp. 99-112.

[PEE96] Peeters, A. “Single-Rail Handshake Circuits”. Phd Thesis, Eindhoven University of Technology, June 1996, 186p.

[PNS13] Pons, M.; Nagel, J.-L. ; Severac, D.; Morgan, M.; Sigg, D.; Ruedi, P.-F.; Piguet, C. “Ultra low-power standard cell design using planar bulk CMOS in sub-threshold operation”. In: International Workshop on Power and Timing Modeling, Optimization and Simulation (PATMOS), 2013, pp. 9-15.

[RAB03] Rabaey, J. M.; Chandrakasan A.; Nikolic, B. “Digital Integrated Circuits: A Design Perspective”. Upper Saddle River: Pearson Education, 2003, 761p.

[RAH12] Rahimi, A.; Benini, L.; Gupta, R.K., "Analysis of instruction-level vulnerability to dynamic voltage and temperature variations." Design, Automation & Test in Europe Conference & Exhibition (DATE), 2012 , pp. 1102-1105.

[REE12] Reese, R.B.; Smith, S.C.; Thornton, M.A., "Uncle - An RTL Approach to Asynchronous Design," In: 18th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC), 2012, pp. 65-72.

[RIO02] Rios, R.; Shih, W.; Shah, A.; Mudanai, S.; Packan, P. ; Sandford, T. ; Mistry, K. “A three-transistor threshold voltage model for halo processes”. In: International Electron Devices Meeting, 2002, pp. 113-116.

[SARAS] Stanford University, Prof Saraswat. “MOS Transistor Review”. Available at http://www.stanford.edu/class/ee316/MOSFET_Handout5.pdf , accessed in 2014.

[SELL14] Sell, J.; O'Connor, P., "The Xbox One System on a Chip and Kinect Sensor." In: IEEE Micro, 2014, Vol.34, no 2, pp. 44-53.

[SOB98] Sobelman, G. E.; Fant, K. “CMOS circuit design of threshold gates with hysteresis”. In: International Symposium on Circuits and Systems (ISCAS), Vol. 2, 1998, pp. 61-64.

[SPA01] Sparso, J.; Furber, S. “Principles of Asynchronous Circuit Design – A Systems Perspective”. London: Kluwer Academic Publishers, 2001, 337p.

[SUT99] Sutherland, I. e.; Sproull, R. F.; Harris, D. F. “Logical Effort: designing Fast CMOS Circuits”. The Morgan Kaufmann Series in Computer Architecture and Design, 199, 296p.

[WES10] Weste, N. H. E.; Harris, D. M. “CMOS VLSI Design: A Circuit and System Perspective”. Addison-Wesley; 4^{th} edition, 2010, 864p.

[WUJ10] Wu, J.; Choi, M. “Latency & area measurement and optimization of asynchronous nanowire crossbar system”. In: Instrumentation and Measurement Technology Conference (I2MTC), 2010, pp. 1596-1601.

[YAN11] Yang, Y.; Yang Y.; Zhu Z.; Zhou D. “A high-speed asynchronous array multiplier based on multi-threshold semi-static NULL convention logic pipeline”. In: International Conference on ASIC (ASICON), 2011, pp. 633-636.

[YAS10] Yancey, S.; Smith, S. C. “A differential design for C-elements and NCL gates”. In: International Midwest Symposium on Circuits and Systems (MWSCAS), 2010, pp. 632-635.

[ZHO10] Zhou, L.; Smith, S. C.; Di, J. “Bit-Wise MTNCL: An ultra-low power bit-wise pipelined asynchronous circuit design methodology”. In: International Midwest Symposium on Circuits and Systems (MWSCAS), 2010, pp. 217-220.

[ZHO12] Zhou, J.; Jayapal, S.; Büsze, B.; Huang, L.; Stuyt, J. “A 40 nm Dual-Width Standard Cell Library for Near/Sub-Threshold Operation”. IEEE Transactions on Circuits and Systems I: Regular Papers, 59(11), Nov. 2012, pp. 2569-2577.

[ZIE14] Ziesemer Jr., A.; Reis, R. A. L.; Moreira, M. T.; Arendt, M. E.; Calzans, N. L. V. “Automatic Layout synthesis with ASTRAN Applied to Asynchronous Cells”. In: IEEE Latin American Symposium on Circuits and Systems (LASCAS), 2014, Accepted for publication.