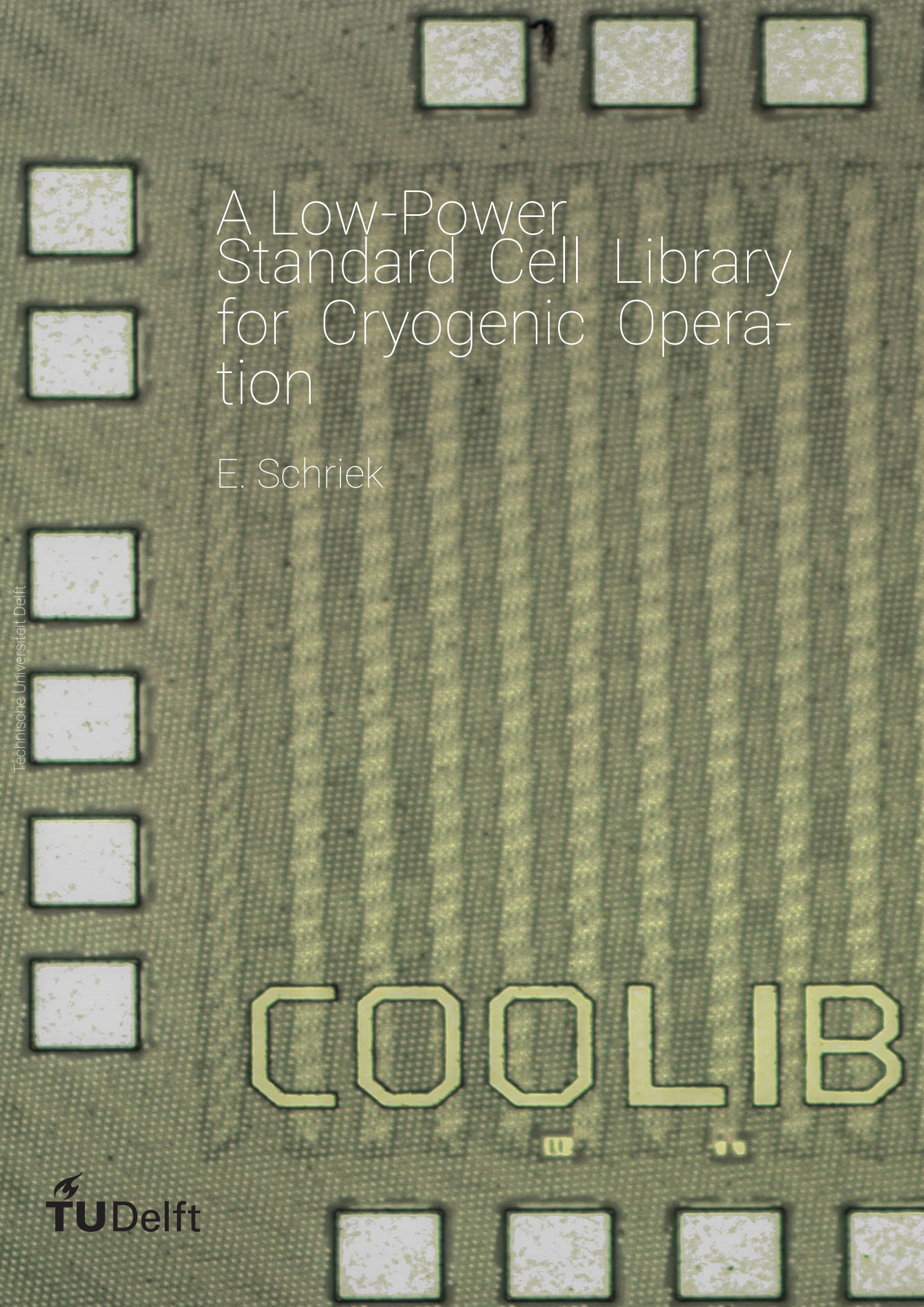
# A Low-Power Standard Cell Library for Cryogenic Operation

E. Schriek

Technische Universiteit Delft




TU Delft

# A Low-Power Standard Cell Library for Cryogenic Operation

by

E. Schriek

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday October 25, 2018 at 04:00 PM.

| | | |
|---|---|---|
| Student number: | 4386965 | |
| Project duration: | September 4, 2017 – October 25, 2018 | |
| Thesis committee: | Prof. dr. E. Charbon, | TU Delft, Supervisor |
| | Dr. F. Sebastiano | TU Delft, Daily Supervisor |
| | Dr. ir. T.G.R.M. van Leuken, | TU Delft |

*This thesis is confidential and cannot be made public until October 31, 2019.*



TU Delft Delft University of Technology

# Abstract

With the advances towards a quantum computer incorporating many qubits, the demand for a scalable read-out and control platform rises. Recent proposals for such a platform exploit the maturity of CMOS technology, to implement the various analog, RF and digital circuits required. To allow for scaling, the electronics that constitute this platform are translated close to the qubits in a cryogenic environment. Limited power dissipation can be absorbed in this environment, stressing the need for low-power circuits.
Presently, digital control is carried out by FPGAs due to the rapidly changing control parameters. FPGAs are expected to be replaced in the future by ASICs, as they are inferior in terms of performance, area and most critically power dissipation. Conventional standard cells used for digital ASIC implementation are expected to be sub-optimal with respect to low-power and cryogenic operation, due to a shift in device characteristics in this environment. This thesis proposes the first minimal standard cell library designed for this environment, with low-power operation in mind, that enables future digital and mixed-signal designs to address the requirements demanded by the cryogenic environment.

# Acknowledgements

I would like to thank my thesis supervisor, Prof. Edoardo Charbon, for offering and trusting me with this challenging project. Edoardo enabled me to work independently, while maintaining an advisory role to the project and myself. Even though Edoardo was abroad most of the time, he kept a close relation to the project by weekly meetings, regardless of his location.
Secondly I would like to thank Dr. Fabio Sebastiano, for being my daily supervisor. Many technical details were easily fleshed out by him due to his expertise in the analog domain.
I would like to thank the members of the 'CoolGroup' for the many educational sessions outside of my field of expertise. Specifically, I would like to thank the other master students that are part of the group: Job, Jaco and Sriram, for cooperative thinking about general issues and helping me with issues I encountered specific to the analog domain.
Finally, I would like to thank my parents for mental support, as well for providing a relaxed working environment at home and means of daily commute to Delft.

*E. Schriek*
*Hendrik-Ido-Ambacht, September 2018*

# Contents


List of Figures 1

1 Introduction 3
1.1 Quantum Computing . . . 3
1.2 Classical Electronics and Quantum Computing . . . 3
1.3 Motivation . . . 4
1.4 Thesis Outline . . . 4

2 (Ultra) Low Power CMOS 5
2.1 The Sub/Near-Threshold Region . . . 5
2.1.1 Theoretical Limits on Supply Voltage and Noise Margins . . . 6
2.1.2 Minimum-Power/Energy Point . . . 7
2.1.3 Secondary Effects . . . 8
2.1.4 Variation . . . 9
2.2 The Sub/Near-Threshold Region at LHe . . . 10
2.2.1 Theoretical Limits on Supply Voltage and Noise Margins . . . 10
2.2.2 Minimum-Power/Energy Point . . . 10
2.2.3 Secondary Effects . . . 11
2.2.4 Variation . . . 11
2.2.5 Latch-up . . . 12
2.3 Logic Families/Styles . . . 13
2.3.1 Static CMOS . . . 13
2.3.2 Dynamic CMOS . . . 14
2.3.3 Dynamic Threshold CMOS (DTMOS) . . . 15
2.3.4 Transmission Gate Logic . . . 16
2.3.5 Current Mode Logic . . . 17
2.3.6 Other Families . . . 17
2.4 Low-Power Standard Cell Libraries . . . 17
2.4.1 Process . . . 18

3 The Standard Cell Library, 'CooLib' 19
3.1 Proposed Standard Cell Layout Properties . . . 19
3.2 Proposed Standard Cell Library Content . . . 23
3.2.1 Low-Power Implementation Cells . . . 24
3.2.2 Physical-Only cells . . . 28
3.3 Timing Information Extraction . . . 29
3.3.1 The Liberty Format . . . 30
3.3.2 Simulation Based Standard Cell Characterisation . . . 31
3.4 Physical Information Extraction . . . 31
3.5 Standard Cell Verification . . . 32

4 Test Chip Design and Implementation 35
4.1 Design & Architecture . . . 35
4.1.1 On-Chip Delay Testing . . . 35
4.1.2 Test Circuits . . . 36
4.2 Implementation . . . 37
4.2.1 Static Logic Synthesis . . . 37
4.2.2 Domino Logic Synthesis . . . 37
4.2.3 Hierarchical, Partition Based Implementation with Multiple Library Domains . . . 39
4.2.4 Power Domains . . . 40

4.3 Verification . . . 41
4.3.1 Functionality & Timing . . . 41
4.3.2 Power Grid . . . 42
4.4 Finalised Layout and Implementation Issues . . . 43

**5 Cryogenic RISC-V** **45**
5.1 RISC-V . . . 45
5.2 SRAM Programming . . . 45
5.3 Implementation and Verification . . . 47
5.4 Finalised Layout and Implementation Issues . . . 47

**6 Testing & Results** **49**
6.1 Test Setup . . . 49
6.2 Automated Test Pattern Generation . . . 49
6.3 Field-Programmable Gate Array Programming . . . 50
6.3.1 Dynamic Clock Generation . . . 50
6.4 Results . . . 50
6.4.1 300 K Delay Measurements . . . 52
6.4.2 300 K Leakage Power Measurements . . . 53
6.4.3 300 K Dynamic Power Measurements . . . 55

**7 Conclusions and Future Work** **57**
7.1 Conclusions . . . 57
7.2 Future Recommendations . . . 57

**Chip Gallery** **59**

**Bibliography** **61**

**A Standard Cell Layouts** **65**

**B RISC-V Programming & Firmware** **77**

# List of Figures


1.1 Proposed cryo-CMOS controller for the control and readout of qubits. [1] . . . . 4

2.1 Noise margins as a function of imbalance factor $IF$, for various values of $V_{DD}$ in the subthreshold regime. . . . . 7
2.2 Minimum Energy Point considering TSMC40LP parameters and $C_L = 1fF$, $\alpha = 0.1$ and LD=10 . 8
2.3 Inverter FO4 falltime simulation (Monte-Carlo) displaying the presence of the RSCE effect. Optimal length is indicated. . . . . 9
2.4 Noise margins for room temperature (a), versus 4.2 K prediction for $n = 34.9$ (b), and $n = 1$ (c). . 11
2.5 Minimum-energy point comparison between 300 K and 4.2 K. . . . . 12
2.6 Sensitivity to $V_{TH}$ mismatch and/or difference. . . . . 12
2.7 Basic implementation of a dynamic logic gate. . . . . 14
2.8 Basic implementation of a domino logic gate. . . . . 15
2.9 NOR gate build from transmission gates. . . . . 16
2.10 Basic CML structure. . . . . 17
2.11 Stacking factor $X_{STACK}$ as a function of $V_{DD}$ . . . . . 18

3.1 Average improvement for $\tau_{HL}$ (left) and $\tau_{LH}$ (right) over 1000 simulations (Monte Carlo). Results below the 100% yield line should not be considered. . . . . 20
3.2 Improvement in for $\sigma_{HL}$ (left) and $\sigma_{LH}$ (right) over 1000 simulations (Monte Carlo). Results below the 100% yield line should not be considered. . . . . 20
3.3 Additional well taps on cell boundary. . . . . 21
3.4 IR estimation under 4.2 K operating conditions. . . . . 22
3.5 Performance and functional yield ($3\sigma$) for popular DFF architectures (logic styles) in the subthreshold region. . . . . 24
3.6 Schematic (a) and layout (b) of the proposed transmission gate D-type flip-flop (without reset). 24
3.7 Schematic (a) and layout (b) of the proposed domino OR gate. . . . . 25
3.8 Proposed two-stage low-to-high level shifter. $V_{DDL}$ represents the low voltage level, $V_{DDI}$ and $V_{DDH}$ represent the intermediate and high (nominal) supply levels. . . . . 26
3.9 Regions of successful LH conversion in terms of $V_{DDL}$ and $V_{DDI}$. The green region represents SS/FF/TT global process corners, while the red and blue represent the SF and FS corners respectively. . . . . 27
3.10 LH shifter double height ($2 \times 2.10\mu m$) standard cell layout. . . . . 27
3.11 Level down conversion, (a) shows conventional inverter based down shifting and its problem, (b) shows the proposed solution. . . . . 28
3.12 (a) shows the level-down shifter double-height ($2 \times 2.10\mu m$) layout, (b) shows the schematic. . . 28
3.13 Physical-only cells, (a) shows a 32-track filler cell, (b) shows the 16-track decap cell. . . . . 29
3.14 Comparison between full layout and abstract view of the same cell. . . . . 32

4.1 Overview of the BIST circuit. . . . . 36
4.2 . . . . . 38
4.3 Logic duplication to remove a trapped inverter at the AND gate's output. . . . . 39
4.4 Logic rearrangements to avoid unnecessary logic duplication. . . . . 40
4.5 . . . . . 41
4.6 Power domain 'PDLow' graphical representation. . . . . 42
4.7 . . . . . 42
4.8 Isolation provided by DNW and short-circuit (dotted line) due to dynamic back-biasing when placed in a common DNW. RW stands for p-well in DNW. . . . . 43
4.9 Full layout of the implemented test chip. . . . . 44

5.1 Logic surrounding the SRAM for programming and readout. . . . . 46
5.2 IR drop across the RISC-V core. . . . . 47
5.3 Finalised PicoRV32 implementation layout with CooLib-Static (and TSMC SRAM blackbox). . . 48

6.1 Complete test set-up. Supplies labelled with the VCC prefix are used for the FPGA. . . . . 50
6.2 Simplified state diagram for the automated testing of the test chip. . . . . 51
6.3 The range of synthesizable frequencies on the Xilinx Artix-7, with $F_{CRYSTAL} = 100MHz$. Points where the relative error between two neighbouring frequencies is above 0.5% are indicated by a red dot. . . . . 52
6.4 300 K delay measurements for the test chip. . . . . 53
6.5 300 K leakage power measurements for the test chip. . . . . 54
6.6 Low-to-high level shifter static power consumption at the $V_{DDH}$ supply, over the full range of $V_{DDL}$. . . . . 54
6.7 300 K dynamic power measurements for the test chip. . . . . 55

# 1 Introduction

## 1.1. Quantum Computing

The use of quantum computers promises advantages over 'classical' computers in solving certain intractable problems [2]. For instance, optimisation and factorisation problems are anticipated to experience an exponential speedup [3]. Quantum computing aims to exploit the phenomena inherited by quantum mechanics to provide speedup, specifically superposition and entanglement.
Contrary to classic computers, where a binary system of bits is used, quantum computers use so-called qubits. In detail, qubits are capable of being in a superposition of two states, whereas a bit can only have one state, logic one or zero. The qubit's state is described as $\alpha|0\rangle + \beta|1\rangle$, where $\alpha$ and $\beta$ are probabilities of finding the qubit in state $|0\rangle$ or $|1\rangle$ respectively. Two numbers are hence necessary to describe the state of a single qubit, contrary to a classical bit where a single number suffices. Moreover, an N-qubit systems requires $2^N$ numbers to describe its state, and cannot be described by classical computers due to memory limitations for reasonable numbers of qubits.

## 1.2. Classical Electronics and Quantum Computing

The development of quantum computers is currently at the state where a small number of qubits is implemented, which is far from approaching the number of qubits required for the problems just described. Even with this small number of qubits, classical computers and conventional electronics are currently used for the readout and control of the qubits or quantum computer. The small number of qubits allow for these electronic devices to operate at room temperature, with long cables connecting to the qubits, that reside at a stable (20-100) milli-Kelvin operating temperature provided by dilution refrigerators. Towards implementation of a larger number of qubits this type of control is expected to become challenging. To cope with the issue, it is proposed to move the readout and control electronics closer to the qubits, which currently lies at the 4.2 K operating temperature. One such architecture for readout and control is displayed in figure 1.1. Operating at this temperature creates new challenges as circuits implemented in CMOS behave differently at this temperature. Additionally, the cooling power of dilution fridges is restricted to several Watts [4], which has to be shared amongst several circuits.
Many of the analog circuits which are part of the readout and control chain have already been implemented with 4.2 K operation in mind [4], while other blocks are under active development. However, the digital control has received less attention, which could be perceived as curious since the digital control is responsible for a significant chunk of power dissipation, with respect to the power budget of a few watt. The state-of-the-art in digital control at 4.2 K exploit of-the-shelf FPGAs to provide a digital control platform which is capable of dealing with the rapid developments in the number of qubits due to its reconfigurability. Certain FPGAs have been characterised at 4.2 K [5], showing attractive changes in device characteristics (i.e. performance) while confirming their functionality at 4.2 K.
Future development stages where the digital control becomes fixed do not require the FPGA's ability to reconfigure. Moreover, at this stage the performance, power and area (PPA) efficiency of the FPGA falls short compared to that of an ASIC [6]. As a side note, developments much further into the possible future might include integration of qubits onto the same die where the ASIC resides, to further stress the need for ASICs, but for now this remains speculation. Low-power techniques for digital implementation, for instance power

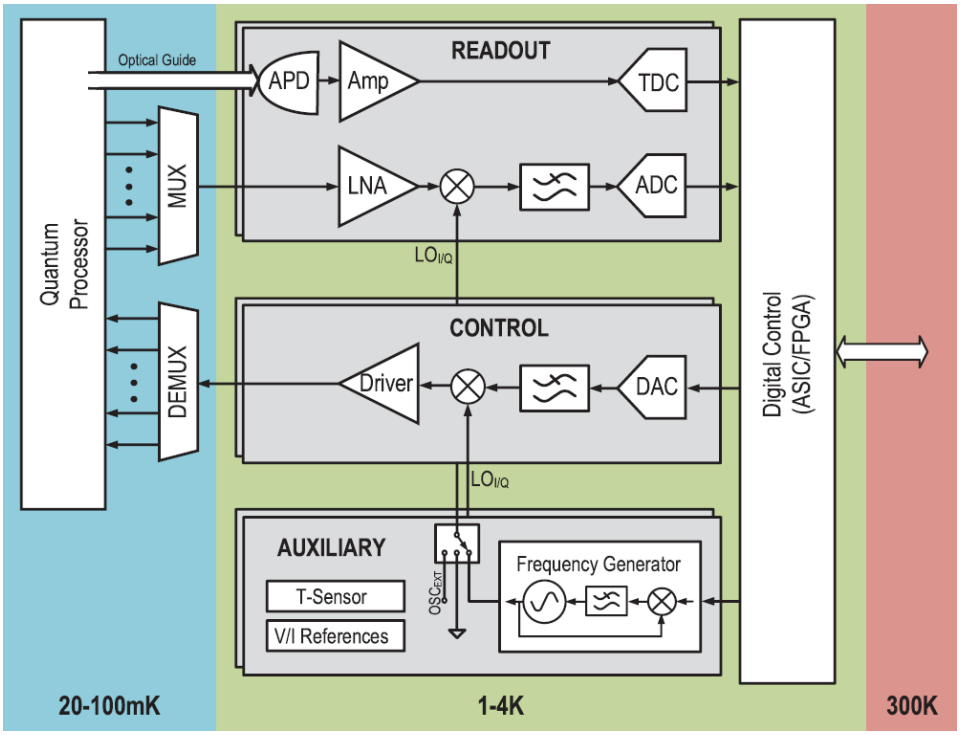



Figure 1.1: Proposed cryo-CMOS controller for the control and readout of qubits. [1]

gating, allow ASICs to increase the power-efficiency gap to FPGAs even further. This stresses the need for low-power digital control in the (near-) future, as any reduction in power dissipated by the digital control opens up headroom for the qubits and analog circuits.

## 1.3. Motivation

While low-power digital implementation techniques are widely available, an opportunity is found in the core of digital circuits. Any semi-custom digital circuit is implemented with the use of standard cells, which are part of a standard cell library, comprising large varieties of boolean functions as well sequential logic such as flip-flops. This library is typically supplied by the foundry and designed for nominal operation, in terms of operating temperature and voltage.
The change in device (MOSFET) characteristics at 4.2 K compared to 300 K render these conventional standard cells potentially sub-optimal in terms of performance and power. Hence, this work aims at exploiting the change in device characteristics at the standard cell level of hierarchy in digital implementation, to achieve more efficient low-power operation.

## 1.4. Thesis Outline

Chapter 2 opens with an introduction into low-power CMOS, followed by an analysis of its properties and how they are predicted to translate by operating at 4.2 K. Chapter 3 discusses the proposed standard cell library 'CooLib' in terms of design choices, additionally it discusses the process of building the auxiliary components needed for a usable standard cell library. Chapter 4 presents one of two chips implemented with the proposed standard cells, alongside typical standard cells, which aims at quantification of the library's performance. Chapter 5 presents the second chip implemented solely with the proposed standard cell library, which implements a simple RISC-V processor. Chapter 6 concerns the test setup and presents the test results. Finally, a conclusion is drawn in chapter 7.

# 2

# (Ultra) Low Power CMOS

Digital Integrated-Circuits in the present time are predominantly implemented using Complementary Metal-Oxide-Semiconductors, or CMOS, where both N- and P-type transistors are used. The choice for this particular style of implementation over other styles, such as transistor-transistor logic (TTL), arises from its low static power consumption and large noise margins. Furthermore, its implementation (layout) allows for very high gate densities, and is thus the standard for very-large-scale integration (VLSI).
Power dissipated by a single logic gate can be described by the following equation, which consists of three contributors, dynamic power, leakage power and short-circuit power [7, 8]

$$P_{total} = P_{dynamic} + P_{sc} + P_{leak}, \tag{2.1}$$

with

$$P_{dynamic} = \frac{1}{2} C_L V_{DD}^2 f \alpha; \qquad P_{sc} = I_{peak} V_{DD} t_{sc} f; \qquad P_{leak} = I_{leak} V_{DD}, \tag{2.2}$$

where $C_L$ is the capacitive load driven by the gate, $V_{DD}$ is the supply voltage, $f$ is the switching frequency, $\alpha$ represents the activity factor, $I_{peak}$ is the peak current drawn when both transistors are on, with $t_{sc}$ being the duration of this period, and finally $I_{leak}$ representing the current which flows when the transistor is off ($V_{GS} = 0$). Clearly, the most effective way to reduce power is to lower the supply voltage due to its quadratic impact. Lowering the supply voltage is not trivial however, as it imposes negative implications for, most importantly, propagation delay and noise immunity. These implications rapidly worsen when the supply voltage drops below the threshold voltage of the transistors, where the transistor is said to operate in the subthreshold region. Even though the implications of operating in the subthreshold region are detrimental to performance, the context of this project makes it necessary to operate in this particular region. Thankfully, the subthreshold region is often exploited due to its properties, and is thus well-understood and modelled. This chapter will start with a more in-depth look into the subthreshold region, along with related topics appropriate to the region. Secondly, the subthreshold region in context of cryogenic operation will be discussed. Finally, a qualitative comparison between logic families and/or styles will be made, again in the context of cryogenic, subthreshold operation.

## 2.1. The Sub/Near-Threshold Region

The sub-threshold region's main characteristic is the exponential $V_{GS} - I_{DS}$ relationship. As stated before, this relationship renders the sub-threshold region unattractive for high performance circuits and/or applications. However, applications which do not yield the high performance constraint might benefit from operating in the sub-threshold region. For example, battery-powered applications are typically not under this constraint, and lowering the supply voltage affects energy consumption in a similar manner to power consumption, increasing application operating time on a single charge of battery as a consequence. The subthreshold current $I_{DS}$, meaning the current which flows from the drain to source terminal, is typically modelled by [9]

$$I_{DS} = I_0 \frac{W}{L} e^{\frac{V_{GS}-V_{TH}}{n v_t}} (1 - e^{-\frac{V_{DS}}{v_t}}); \qquad I_0 = \mu_0 C_{ox} \frac{W}{L} (n-1) v_t^2, \tag{2.3}$$

where $W/L$ is the aspect ratio between transistor width and length, $V_{TH}$ is the threshold voltage, $V_{GS}$ is the gate-source voltage, $V_{DS}$ is the drain-source voltage, $v_t = kT/q$ is the thermal voltage, and finally $I_0$ and $n$ are technology-dependent parameters. It is important to note that $V_{TH}$ depends on $V_{DS}$, through the drain-induced barrier lowering (DIBL) effect, and $V_{BS}$ through the body effect. This can be analytically expressed as [9]

$$V_{TH} = V_{TH0} - \lambda_{DS} V_{DS} - \lambda_{BS} V_{BS}, \tag{2.4}$$

with $V_{TH0}$ representing the zero-bias threshold voltage (when $V_{DS} = V_{GS} = 0V$), whereas $\lambda_{BS}$ and $\lambda_{DS}$ are technology-dependent coefficients which represent the 'strength' of the DIBL and body effect.
The expression for $I_{DS}$ confirms the exponential $V_{GS} - I_{DS}$ relationship and hence the quick performance degradation. As $V_{DD}$ lowers, performance degradation is further increased by DIBL, as the DIBL effect implies that $V_{TH}$ increases as $V_{DS}$ lowers. DIBL was found to be negligible above 130-nm nodes, but increases severely with the scaling of $L_{eff}$ [10].
As stated earlier, operating in the subthreshold region has implications with respect to noise immunity. Noise immunity in CMOS is typically expressed in terms of noise margins. Existing work has shown that strength imbalance between N- and P-MOS transistors in a single gate (i.e. inverter) affects these noise margins in a negative manner. Furthermore, it was found that eventually, as $V_{DD}$ decreases, noise margins become negative which imposes a restriction on the minimum supply voltage. This theoretical restriction of supply voltage, as well as noise margins, will be discussed in more detail in section 2.1.1. Other characteristics of the subthreshold region will be discussed as follows. The so-called minimum-energy point, which is a selling point of subthreshold region, will be discussed in 2.1.2. Short-channel effects with affect transistor characteristic in the subthreshold region will be discussed in 2.1.3. Finally, a more demanding issue of the subthreshold region appearing in smaller process nodes, variation, will be discussed in 2.1.4.

### 2.1.1. Theoretical Limits on Supply Voltage and Noise Margins

In the previous section it was briefly stated that N- and P-MOS strength imbalance imposes restrictions on $V_{DD,min}$. Before discussing this in more detail, theoretical limits considering perfect balance between N- and P-MOS should be considered, where perfect balance implies equal current driving capabilities disregarding transistor dimension. Lower bounds on $V_{DD,min}$ have been extensively explored in literature, however agreement on a single expression for $V_{DD,min}$ is difficult to find. The most commonly found expression, which also has a clear derivation [11], is found to be [12, 13]

$$V_{DD,min} = 2\frac{kT}{q}\left(1 + \frac{C_{fs}}{C_{ox} + C_{dep}}\right)\ln\left(2 + \frac{C_{dep}}{C_{ox}}\right), \tag{2.5}$$

which reduces to $V_{DD,min} \approx 2\frac{kT}{q}\ln(2) \approx 36mV$ at 300 K. It should be indicated that the last term of this expression contains the subthreshold slope factor $n$, as $n = 1 + \frac{C_{dep}}{C_{ox}}$. Different boundaries for $V_{DD,min}$ found in literature are integer multiples of the thermal voltage, $V_{DD,min} > 2...4\frac{kT}{q}$ [7, 14]. This boundary appears to be incorrect, as the lowest $V_{DD}$ found in literature at room temperature, at 62 mV, is well below $4\frac{kT}{q}$ [15]. Succesful operation at 62 mV was achieved using a schmitt-trigger logic style which also claims a theoretical limit lower than 36 mV, namely 31.5 mV [16]. Although this decrease is practically negligible, this indicates other parameters play a role in achieving a low $V_{DD,min}$, for this schmitt-trigger logic is claimed to result from a larger $I_{ON}/I_{OFF}$ ratio. The larger $I_{ON}/I_{OFF}$ leads to better signal swing, naturally leading to better noise margins. The fact that the 36 mV, or close to, has never been achieved can be explained by the subthreshold slope factor which is always greater than unity in practice.
Further explanation can be found in N- and P-MOS strength imbalance. $V_{DD,min}$ lies deep in the subthreshold region, meaning that any imbalance in threshold voltage between N- and P-MOS transistors is amplified compared to superthreshold operation. Designing for perfect balance is thus an impossible task under the effects of inter- and intra-die process variations. Since it is desirable to operate at a supply voltage where positive noise margins exist, an expression was derived for $V_{DD,min}$ in [17], which takes imbalance as well as the 'natural' limits into account

$$V_{DD,min} \approx 2v_t + nv_t \ln(IF), \tag{2.6}$$

where IF, the imbalance factor, is the strength ratio between the stronger and the weaker transistor (effectively $I_{DS,P}/I_{DS,N}$ or vice-versa). This work claims $V_{DD,min} \approx 116mV$ in a 65-nm process, which is well above

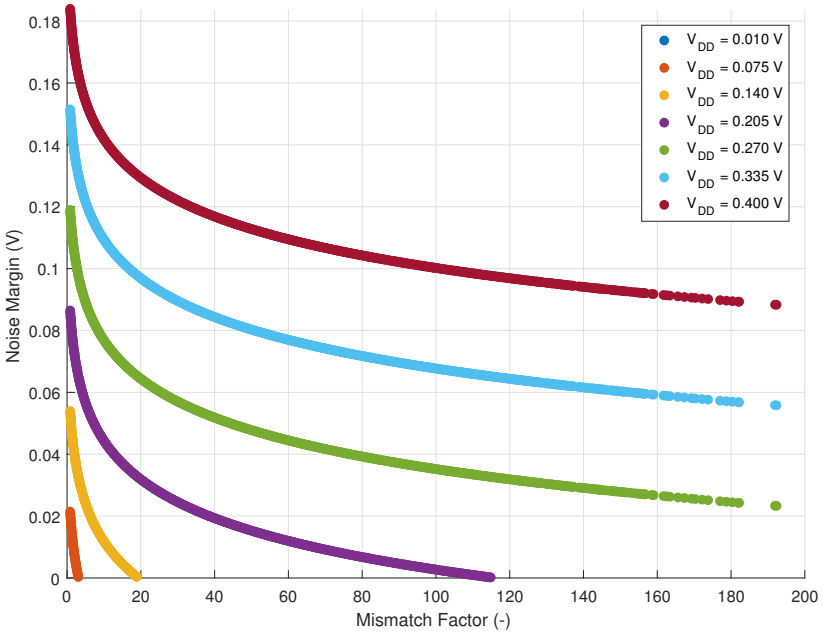



Figure 2.1: Noise margins as a function of imbalance factor $IF$, for various values of $V_{DD}$ in the subthreshold regime.

the theoretical boundaries that do not take imbalance into account. Various other ultra-low voltage applications have been demonstrated to operate around this voltage, possibly indicating that this boundary is more practical to work at.
In the previous discussion, noise margins are often mentioned. To understand static noise margins in low-voltage CMOS, one must understand that low-voltage acts as ratioed logic [9]. In detail, low-voltage inherits an $I_{ON}/I_{OFF}$ ratio much smaller compared to nominal $V_{DD}$ operation, in such low quantities that it starts acting as ratioed logic. In this case, the output high (low) fails to reach $V_{DD}$ ($V_{SS}$). This is known as a form of signal degradation, and worsens as $V_{DD}$ decreases. For a detailed analysis the reader is referred to [9]. Another important factor is the strength ratio, or the imbalance factor. A strong NMOS affects (decreases) the output high level $V_{OH}$, while a strong PMOS affects (increases) the output low level $V_{OL}$. The author of this work performed an in-depth analysis, deriving expressions for $\Delta V_{OH}$ and $\Delta V_{OL}$, which were found to match well

$$\Delta V_{OH} = \upsilon_t \frac{\beta_n}{\beta_p} e^{-\frac{V_{DD}}{n_p \upsilon_t}}; \qquad \Delta V_{OL} = \upsilon_t \frac{\beta_p}{\beta_n} e^{-\frac{V_{DD}}{n_n \upsilon_t}}, \tag{2.7}$$

where $\beta_p$ and $\beta_n$ represent device strength, similar to $I_{DS}$. Subsequently, in the same work, an expression for static noise margin is found to be

$$NM = min(NM_L, NM_H) \tag{2.8}$$

$$= \frac{V_{DD}}{2} - \upsilon_t - \upsilon t \frac{n}{2} \ln(IF). \tag{2.9}$$

Equations 2.7 and 2.8 are powerful tools to analyse noise immunity in the subthreshold regime. Figure 2.1 shows the noise margin as a function of imbalance factor $IF$, where $IF$ was taken as Gaussian distribution to replicate $\sigma_{V_{TH}}$. Clearly, the noise margin degrades with $V_{DD}$, although a logic gate can sustain considerable imbalance at $V_{DD} = 200mV$.

### 2.1.2. Minimum-Power/Energy Point

One of the features of the subthreshold region is the so-called minimum-energy point. At this particular point with respect to supply voltage, the energy consumed of some circuit is at a global minimum. Naturally, this

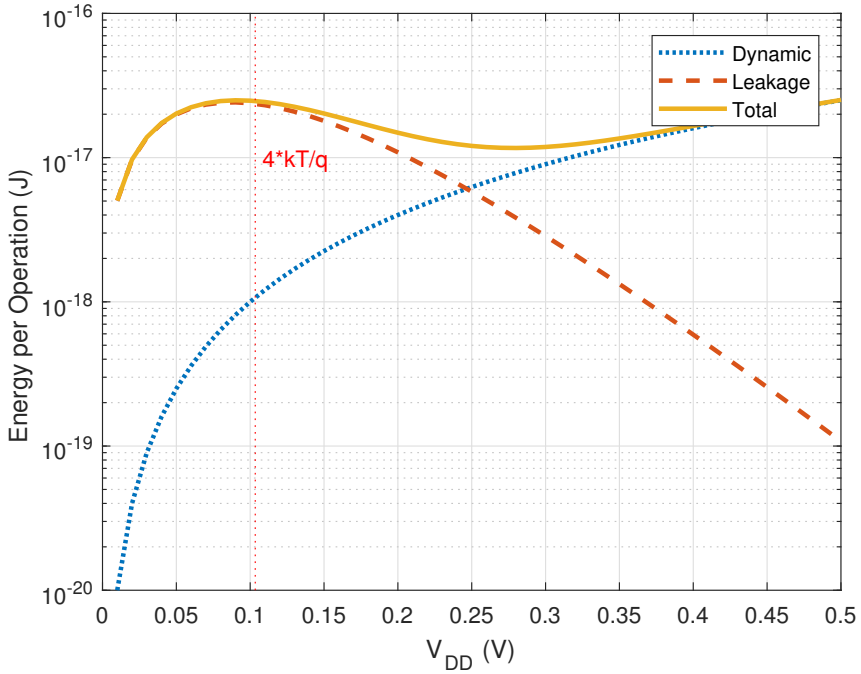



Figure 2.2: Minimum Energy Point considering TSMC40LP parameters and $C_L = 1fF$, $\alpha = 0.1$ and LD=10

point is of interest to (mobile) battery powered application, as stated earlier. The minimum-energy point typically lies deep in the subthreshold region [8]. Similar to power, energy is composed of a dynamic, leakage and short-circuit portion

$$E_{total} = E_{dynamic} + E_{leak} + E_{sc} \tag{2.10}$$
$$= C_L V_{DD}^2 \alpha + I_{leak} V_{DD} t_p LD + Q_{sc} V_{DD}, \tag{2.11}$$

where $LD$ is the logic depth, or the number of gates considered, and the charge $Q_{sc}$ represents the current drawn for the short-circuit duration.
Figure 2.2 shows the minimum energy point (MEP) for $C_L = 1fF$, $\alpha = 0.1$ and $LD = 10$, with process parameters extracted from SPICE simulation. The critical difference between energy and power is that energy, specifically leakage, is dependent on propagation delay $t_p$. This implies that, while $I_{leak}$ decreases with $V_{DD}$, $t_p$ will increase exponentially and eventually leakage energy will dominate dynamic energy. At that point the minimum energy consumption is achieved, and decreasing $V_{DD}$ will only incur unnecessary penalties in energy consumption. Finally, for higher activity factors $\alpha$ the MEP will shift up and to a lower $V_{DD}$. If more than one gate is to be considered, meaning $LD$ increases, the MEP will shift up and to the right.
Contrary to energy, power consumption decreases monotonically with $V_{DD}$, since there is no dependence on propagation delay.

### 2.1.3. Secondary Effects

Two secondary effects should be considered when operating in the subthreshold region, the reverse short-channel effect (RSCE) and the inverse narrow-width effect (INWE), as both effects substantially modulate the threshold voltage. Short channel devices display a phenomenon where $V_{TH}$ decreases with device length [7, 18], which is known as SCE. As device dimensions scale, the effect worsens with increasing DIBL. To mitigate the problem, HALO doping (pocket implants) is used [18, 19], which consists of highly doped regions at the drain and source, increasing the overall doping concentration in the channel. Increasing the length of a transistor with HALO doping separates the HALO implants, lowering the dopant concentration in the centre of the channel. In turn, this increases the depletion layer under the gate, lowering $V_{TH}$ [19]. This effect is known as the reverse short-channel effect, and as the name implies is the opposite of the short-channel effect. In the subthreshold region, SCE is attenuated by a reduction of DIBL, meaning the RSCE effect is

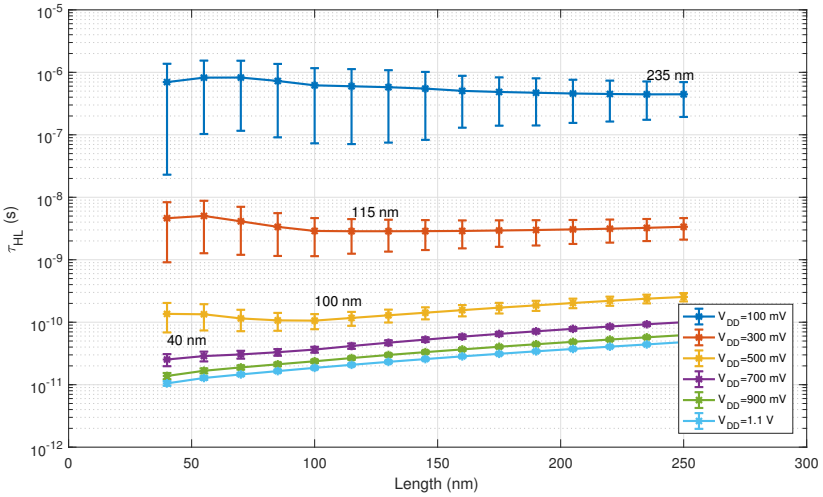



Figure 2.3: Inverter FO4 falltime simulation (Monte-Carlo) displaying the presence of the RSCE effect. Optimal length is indicated.

very well present [18]. An attractive property of exploiting the RSCE effect is a reduction in propagation delay, due to the exponential $V_{GS} - I_{DS}$ relationship. Secondly, RSCE increases the depletion width under the gate, lowering the depletion capacitance $C_{dep}$. This implies that the subthreshold slope factor $n$ lowers, since $n = 1 + \frac{C_{dep}}{C_{ox}}$. Finally, the gate capacitance of a transistor in the subthreshold region is given by the oxide capacitance $C_{ox}$ in series with the depletion capacitance $C_{dep}$, which implies that RSCE lowers the gate capacitance. A SPICE simulation, shown in figure 2.3, of an inverter driving four equally sized inverters shows that the optimal length in the subthreshold region is indeed larger than minimum, confirming the theory as defined in literature.
The narrow-width effect is a result of local oxidation of silicon (LOCOS), which is a process step developed to insulate transistors from each other, by providing a gradual reduction in channel depth [19]. The use of this insulating technique allows the depletion region to extend outside of the channel width, under the isolating field oxide [7]. Consequently, more gate voltage is needed to create the channel, which is more significant for small channel widths. Hence, the narrow-width effect (NWE) increases $V_{TH}$ with decreasing $W$. LOCOS has in time been replaced by shallow trench isolation (STI) [19], creating an abrupt end to the channel instead of the gradual reduction of LOCOS. This abrupt ending, and the defects introduced create fringing fields at the channel corners, reducing the threshold voltage when the width of the channel is smaller. The threshold voltage under the effect of the fringe fields is expressed as [19]

$$V_{TH0} = V_{FB} + \psi_s + E_{ox}\frac{t_{ox}}{1 + F/W}. \tag{2.12}$$

For details, the reader should refer to [20]. This effect is coined as the inverse narrow-width effect, as it reduces $V_{TH}$ as $W$ decreases.

### 2.1.4. Variation

Variation between transistor, or more commonly known as intra-die variation (mismatch) is becoming an increasingly bigger problem as device dimensions scale. Random spatial reliability effects, which are decided during manufacturing, are the main source of mismatch in the form of threshold voltage variation [21, 22]. Random spatial reliability effects include random dopant fluctuations and line edge/width roughness. Mismatch has larger implications for the subthreshold compared to the superthreshold. The $V_{TH}$ distribution is typically Gaussian, resulting in a Gaussian distribution for $I_{ON}$ and thus delay in the superthreshold. In the subthreshold, the resulting distribution for $I_{ON}$ is log-normal due to the exponential dependence of $I_{ON}$ on $V_{TH}$. As a result, a larger design margin is needed to achieve a certain yield compared to the superthreshold. To combat the issue, increasing logic depth is effective as it quickly moves the log-normal to a Gaussian distribution, as dictated by the Central Limit Theorem [23]. The second solution which acts on a per transistor basis is to increase device dimensions, as dictated by Pelgrom's law [24]. As mismatch only worsens with newer process nodes, reaching the theoretical lower limits on supply voltage becomes harder as well. Circuits

trying to explore this theoretical regime require perfect N- and P-MOS balance, as discussed in section 2.1.1. Surely, minimum sized transistor, and standard cells provided by the foundry, would be ineffective for this purpose.

## 2.2. The Sub/Near-Threshold Region at LHe

Cryogenic CMOS has perceived quite some attention in recent times, in context of quantum computation. Subsequently, it is known how important transistor parameters alter. If equation 2.3 is considered, three important parameters change significantly at 4.2 K. First, the threshold voltage $V_{TH}$ is generally found to increase by approximately 100 mV [25]. Secondly, the mobility experiences an increase of approximately twofold. Finally, the subthreshold slope $SS$ is well below the $60mV/decade$ limit imposed by room temperature operation, even though the subthreshold slope factor $n$ increases by an order of magnitude ($SS = \ln(10)\frac{kT}{q}n$). It should be noted that other parameters are affected, such as transistor transconductance, however these parameters are of little interest to digital circuits. The net effect of the altered parameters is a slightly lower $I_{DS}$ in the subthreshold regime, which is due to the increase in threshold voltage $V_{TH}$. Even though mobility increases, $V_{TH}$ dominates due to exponential $I_{DS} - V_{GS}$ relation. Naturally, the net effect in the superthreshold regime is a slight increase of $I_{DS}$, as this exponential relation is not present. The decrease of the subthreshold slope $SS$ implies a significant increase of the $I_{ON}/I_{OFF}$ ratio. Furthermore, $I_{OFF}$ (leakage) decreases by orders of magnitude [25].

The purpose of this section is to compare, how the different characteristics, effects and issues translate when moving to liquid helium temperature, or 4.2 K. Most parameters of significance are known, which allows for a prediction using existing models and expressions. It is important to treat the following subsections as such.

### 2.2.1. Theoretical Limits on Supply Voltage and Noise Margins

The absolute theoretical limit on supply voltage at room temperature (300 K) was found to be approximately 36 mV, as discussed in section 2.1.1. If the assumption is made that the subthreshold slope is ideal, implying $n = 1$, then the theoretical limit at 4.2 K evaluates to $V_{DD,min} \approx 2\frac{kT}{q}\ln(2) \approx 0.48mV$. This reduction, compared to room temperature, follows only from a reduction in the thermal voltage, which is based on constants. This limit might seem promising from a power perspective, however finding an application where delays caused by operating at this supply voltage are allowed might prove difficult. The subthreshold slope factor $n$ at room temperature is typically close to one, experimental results at 4.2 K turn out significantly higher [25, 26]. With a value for $n = 34.9$ taken from these results (40-nm process), $V_{DD,min} \approx 2.47mV$.

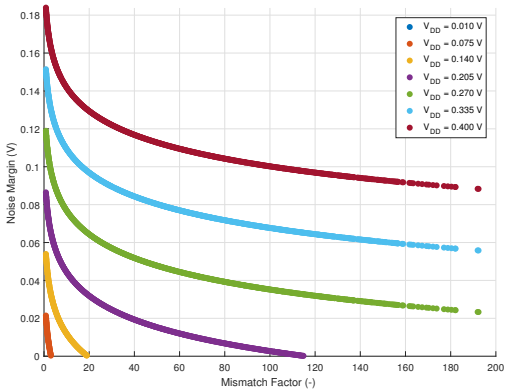
One particular paper [27] already confirmed that the theoretical limit is indeed lower in cryogenic environment, as 27 mV was achieved at 77 K (compared to 70 mV at 300 K). Unfortunately, the author did not attempt 4.2 K operation, neither is there literature attempting limits at 4.2 K available.

Previously, the analysis was made that (deep) subthreshold acts as ratioed logic. The significance of this analysis was that (subthreshold) noise margins are a strong function of the transistor $I_{ON}/I_{OFF}$ ratio. Given the decline in subthreshold slope $SS$, and thus increase in $I_{ON}/I_{OFF}$, noise margins are expected to become better at 4.2 K. Figure 2.4 shows a comparison between noise margins at 300 K and 4.2 K (estimated). At 4.2 K, noise margins are estimated to remain close to $\frac{V_{DD}}{2} - v_t$ (the ideal value), due to the decrease in $SS$. Figure 2.4c shows an estimation of the noise margins for the ideal $SS$. In this scenario, the subthreshold slope steepness approximates a vertical line, i.e. the transistor shows near ideal on/off switching behaviour. Under this scenario, noise margins are estimated to be close to the ideal value $\frac{V_{DD}}{2} - v_t$ under arbitrary imbalance conditions.

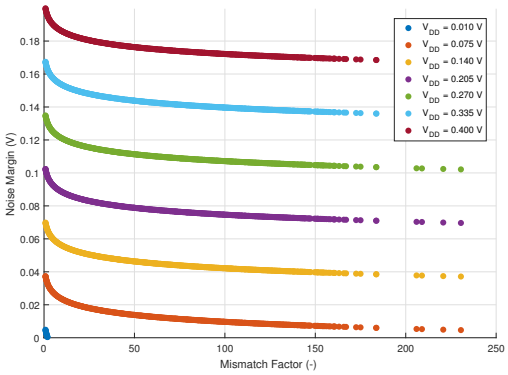
### 2.2.2. Minimum-Power/Energy Point

The minimum-energy point present in the subthreshold region has been discussed in section 2.1.2. By using parameters for mobility, threshold voltage and subthreshold slope factor from existing experiments at 4.2 K, it is possible to predict the minimum-energy point at 4.2 K. Figure 2.5 shows a comparison using to the plot shown in 2.1.2.

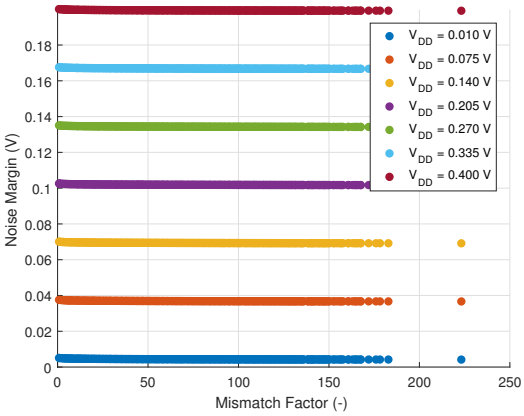
Clearly, the reduction in leakage current offered by operating at 4.2 K moves the MEP to the left. The reduction in leakage energy provides an overall reduction in the total energy, translating in the MEP moving downwards in comparison to room temperature. The reduction in leakage energy is confirmed by measurement results in [25], where a significant reduction in leakage current is shown for a 40-nm process. This reduction in leakage current is mainly attributed by a steeper subthreshold slope.




(a) Noise margin prediction for 300 K with typical subthreshold slope $SS$ ($n = 1.4$).




(b) Noise margin prediction for 4.2 K with experimental subthreshold slope $SS$ ($n = 34.9$).




(c) Noise margin prediction for 4.2 K with ideal subthreshold slope $SS$ ($n = 1$).

Figure 2.4: Noise margins for room temperature (a), versus 4.2 K prediction for $n = 34.9$ (b), and $n = 1$ (c).

### 2.2.3. Secondary Effects

In section 2.1.3 secondary effects on transistor behaviour were discussed. Additionally, it was shown how effective it is to exploit these effects in the subthreshold regime. For RSCE, it has been established that the effect is significantly reduced [28–30] down to 30 K. Explanation is found in the temperature dependence of the bulk Fermi potential, which is found to be less sensitive to doping concentration at low temperatures. One should remember that RSCE is effectively a modulation of the doping concentration. It is presumed that this reduction continues down to 4.2 K.

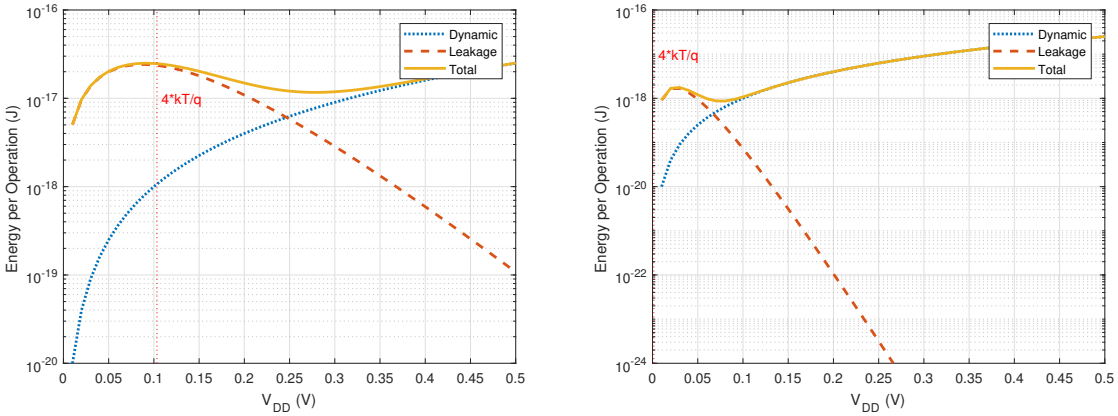
INWE is an effect which has gotten no attention below 77 K. A single research [30] presents experimental results for the INWE effect down to 77 K. In this work, it was found that the INWE effect is significantly larger for the PMOS. For the NMOS, no clear conclusion was adopted.

### 2.2.4. Variation

The main source of variation, or mismatch, at room temperature is due to random dopant fluctuations (RDF). To predict the significance of variation at 4.2 K, a commonly used metric named the coefficient of variation (CV) is used [23] [31]

$$\frac{\sigma_{\tau_{PD}}}{\mu_{\tau_{PD}}} \approx \frac{\sigma_{I_{ON}}}{\mu_{I_{OFF}}} = \sqrt{e^{\frac{\sigma_{V_{TH}}}{nv_t}} - 1}. \tag{2.13}$$

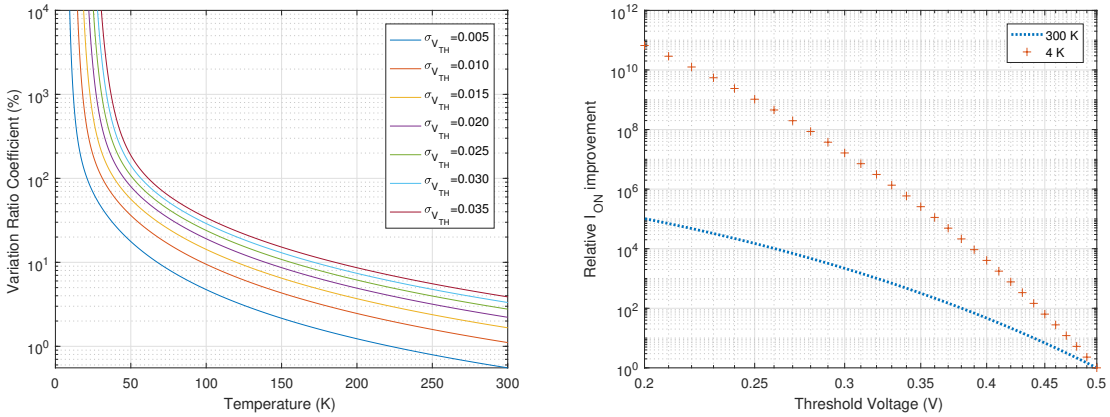
If $n$ is interpolated from 300 K values to the experimental values found at 4.2 K over temperature, an indication of the CV over temperature is given in figure 2.6a. Following this metric, a significant increase in mismatch of




(a) Minimum-energy point at 300 K.

(b) Minimum-energy point at 4.2 K.

Figure 2.5: Minimum-energy point comparison between 300 K and 4.2 K.




(a) Interpolation of the coefficient of variation for mismatch on $I_{ON}$.

(b) Relative $I_{ON}$ improvement of 300 K versus 4 K ($n = 1$ versus $n = 34.9$)

Figure 2.6: Sensitivity to $V_{TH}$ mismatch and/or difference.

$I_{ON}$, or propagation delay, is to be expected. An possible explanation could lie in the subthreshold slope $SS$; an equal deviation in threshold voltage $V_{TH}$ results in a greater deviation of $I_{ON}$ at 4 K, as compared to 300 K. This is visualised in figure 2.6b.

Recent work shows experimental results for a 40-nm process [32], confirming the expected increase in mismatch. At moderate inversion, or near-threshold, mismatch is found to degrade up to 10x. The experienced increase in mismatch is likely to be explained by carrier freeze-out modulating the carrier concentration, amplifying the effects of RDF.

### 2.2.5. Latch-up

The combination of well and substrates in CMOS processes results in parasitic $n-p-n-p$ structures, creating parasitic bipolar transistors [7]. Special care should be taken to not forward bias one of these bipolar transistors, doing so will cause it to feed current into the base of the other bipolar transistor. This creates a positive feedback loop between the two bipolar transistor, which increases the current until the circuit breaks down. Avoiding latch-up is simple, $R_{nwell}$ and $R_{psub}$ should be small. This is normally achieved by placing well and substrate contacts in the neighbourhood of the transistors. While $R_{nwell}$ and $R_{psub}$ should be minimised, excessive currents through the substrate could still lead to large IR values, potentially forward biasing the parasitic bipolar transistor.

Substrate current is a function of $I_{DS}$ [33], implying the subthreshold regime is 'safer' to operate in. Substrate/n-well resistance has been found to increase by several orders of magnitude, due to freeze-out of the substrate

at 4.2 K [1]. Considering the requirements to avoid latch-up, this substantial increase is alarming as it could lead to the destruction of the circuit. Some literature exists on latch-up at cryogenic temperatures, older research claimed that latch-up at cryogenic temperatures was impossible. To sustain the latch-up effect at room temperature, the multiplicative current gain of the bipolar transistors should be larger than unity. However, bipolar transistor current gain, $\beta_{no}$ and $\beta_{po}$, drop far below unity at 4.2 K [26], rendering latch-up close to impossible. This claim was found to be true down to liquid nitrogen temperatures (77 K) [34]. It was found that below approximately 50 K, other mechanisms for carrier generation start to play a role, specifically shallow level impact ionization (SLII) [34, 35]. For this region, the condition for latch-up was found to be

$$\beta_{no}\beta_{po} > (M_{sn}M_{sp})^{-1}; \qquad M_s = \frac{N}{n} I_{ii}(F), \tag{2.14}$$

where $N$ is the density of doping atoms, $n$ the number of free carriers, and $I_{ii}(F)$ the ionization ratio, which is a function of the electric field and is zero below the threshold for impact ionization $F_{ii}$. This implies that while $\beta_{no}\beta_{po} < 1$, $M_s$ could reach sufficiently high values to sustain the latch-up condition. Unfortunately, no models and/or experimental results can be found with respect to SLII.
Apart from loop gain, power supply and associated circuits must be able to provide a minimum holding current $I_H$ and holding voltage $V_H$ to maintain latched state [34–36]. In this research it was found that $V_H$ is considerably higher compared to 300 K, while $I_H$ is only slightly higher (for different process nodes).
To conclude, while information about latch-up at 4.2 K is scarce, it appears latch-up immunity is better compared to 300 K. Still, due to the destructive nature of the effect special care should be taken into account.

## 2.3. Logic Families/Styles

In section 2.1, the subthreshold and its characteristics have been discussed. Since it is clear that operating in the subthreshold region is not without consequences, many logic families, or styles, have been analysed with respect to the regime. Furthermore, based on predictions for 4.2 K operation, certain logic families might show obvious advantages to others. Hence, the purpose of this section is to qualitatively (and quantitatively to some extend) compare logic families, based on literature, with an eye on subthreshold, cryogenic operation.

### 2.3.1. Static CMOS

Static (complementary) MOS is the industry standard for implementing digital circuits. All of the analysis regarding noise margins and the minimum-energy in section 2.1 is based on this logic family. Any commercial design automation software, i.e. synthesis, will selectively support static CMOS. Hence, static CMOS is the recommended logic family [7], and is in practice impossible to avoid.

### 2.3.2. Dynamic CMOS

Dynamic CMOS relies on temporary storage of charge to provide the logic high $V_{OH}$ and low $V_{OL}$ levels. Similar to static CMOS, dynamic logic consumes no static power. Dynamic logic works in two phases, precharge and evaluation, where the phase is determined by the clock signal $CLK$ [7]. The basic schematic of a dynamic gate is shown in figure 2.7. When $CLK = 0$ (precharge), $C_L$ is charged to $V_{DD}$, while the path to ground is closed by the bottom NMOS transistor. Subsequently, when $CLK = 1$ (evaluation), the top transistor is disabled and $C_L$ is either discharged to ground through the pull-down network (PDN) or remains untouched. $V_Y$ is thus a function of the NMOS transistors in the pull down network, and is implemented in a similar manner to static CMOS. Advantages of the dynamic structure is a lower number of transistors ($N+2$ versus $2N$ of static logic), and thus a lower total input capacitance. Secondly, speed (performance) is higher due to a decrease in load capacitance and to an important characteristic of dynamic logic.

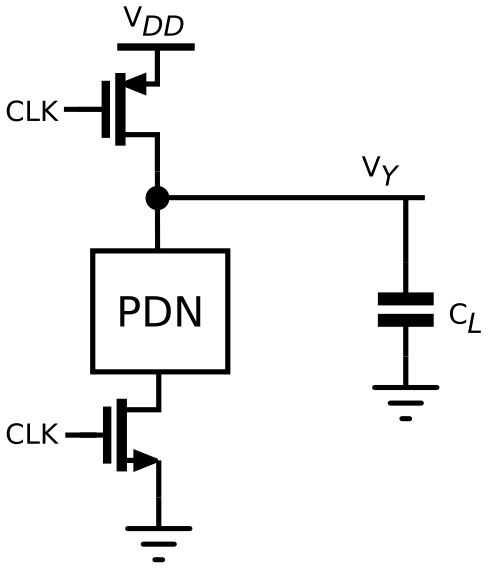



Figure 2.7: Basic implementation of a dynamic logic gate.

Consider the scenario where $C_L$ is precharged and waiting for evaluation. If no switching (evaluation) occurs, $C_L$ remains high, resulting in $\tau_{pLH} = 0$. If large logic depths are considered, the total propagation delay of that path can be reduced significantly if some of the dynamic gates in the path do not evaluate. Downsides of the dynamic logic style lie in dynamic power consumption and signal integrity issues. Since dynamic logic acts on $CLK$, it might in the worse case scenario consume power while charging and discharging $C_L$. Hence, dynamic logic typically consumes more dynamic power than static CMOS [7]. Signal integrity issues arise from, most importantly, charge leakage of $C_L$ and charge sharing. Charge leakage appears naturally due to the leakage current of the transistors in the pull-down network. Charge sharing takes place whenever the charge in $C_L$ is redistributed over the parasitic capacitances of the transistors in the pull-down network due to partial evaluation. Both issues can be solved by adding additional transistors, to prevent leakage in the case of charge leakage, and to precharge parasitic capacitances in the case of charge sharing. Both solutions induce area and capacitance penalties.

In the end there is a more destructive issue with pure dynamic logic that prevents it from being used in reality. As a matter of fact, cascading of dynamic logic is impossible. The output of every dynamic gate $V_Y$, meaning the inputs to gates in the next stage, are precharged to $V_{DD}$ (one). This means that $C_L$ might be unintentionally discharged at the beginning of the evaluation phase [7]. This implies that correct operation is ensured if the inputs can only make a single transition from zero to one during the evaluation phase. This countermeasure is effectively implemented with an 'extension' on the dynamic logic family, which is named domino logic.

#### Domino Logic

Domino logic solves the issue of cascading of dynamic gates by placing a static inverter at the original output $V_Y$. The basic domino gate structure is displayed in figure 2.8. The result of this static inverter is a precharged value of $V_Y = 0$, and selectively transitions during the evaluation phase from zero to one. This solves the critical cascading issue of dynamic logic. A number of side effects are introduced with the addition of the static inverter; an increase in noise immunity, a reduction in $C_L$ (when cascading) and the inverter can be used to prevent charge leakage and charge sharing.

To prevent charge leakage, an additional PMOS transistor is introduced which has its gate connected to $V_Y$. This means that when $C_L$ is precharged, $V_Y$ is zero, turning on the transistor and maintaining the voltage at the intermediate node. Subsequently when $V_Y = 1$ during evaluation, the transistor is off such that it does not impact the discharging of $C_L$. This transistor is called a weak-keeper and is usually employed in domino logic [37]. Similar to combating charge leakage, the output of the static inverter can be used to precharge the parasitic capacitances residing in pull-down network of the following domino logic stage.

One issue that emerges from using a static inverter at the gate's output is that only non-inverting functions can be implemented. In most cases, inverting functions can be replaced by non-inverting functions as dictated by De Morgans's Law. This requires an additional logic transformation step during synthesis. Finally, one can use dual-rail (differential) domino logic to provide both inverting and non-inverting functions at the same time.

Domino logic is claimed to show a significant increase (10x) in power-delay product (PDP) over static logic in the subthreshold regime [38]. The same work presents an analysis of noise immunity of subthreshold domino logic, which is considerably different from static logic. Domino logic fails to provide a correct output when the intermediate voltage drops to the extent where it crosses the static inverter switching threshold. It was found that $V_{noise}$, i.e. the noise voltage present at one of the domino gate's input, needs to be very close to $V_{DD}$, thus implying excellent noise immunity [38]. Consider the reduction in transistor leakage current and increase in static noise margins at 4.2 K. Potentially, a 'true' domino logic implementation without weak-keeper might function correctly under these improvements.

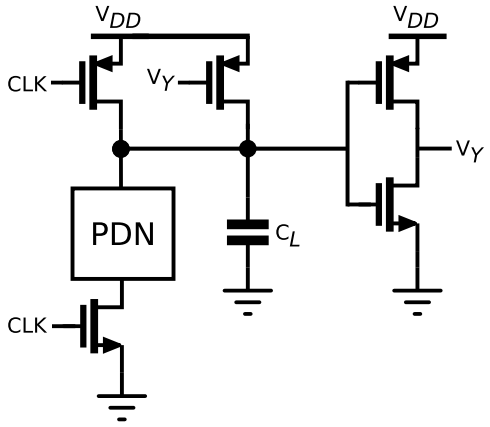



Figure 2.8: Basic implementation of a domino logic gate.

Due to the high performance characteristic, domino logic has found its place in high performance ASIC design [37]. In the state-of-the-art, it appears as domino logic is avoided due to its power consumption and general increased complexity in implementation [39].

### 2.3.3. Dynamic Threshold CMOS (DTMOS)

Dynamic Threshold CMOS is a family similar to static CMOS. In fact, any static CMOS gate can be implemented in DTMOS by a near exact copy. This arises from the fact that DTMOS *is* static CMOS, under different body bias conditions. In more detail, the gate of each transistor is connected to the body (or bulk) of that transistor. Turning a transistor on will forward-bias ($V_{BS} > 0$ in case of NMOS) the device, lowering the threshold voltage. When the device is off, the body bias conditions are identical to static CMOS (reverse-biased). This creates a family compatible with static CMOS, but superior to it in terms of performance $I_{ON}$. Furthermore, DTMOS is known to show less susceptibility to mismatch (variation) [40] and to have a superior threshold slope factor $n$ compared to static CMOS [26]. Implications of the latter is an increased $I_{ON}/I_{OFF}$ ratio, and thus larger static noise margins.

Operating DTMOS at nominal voltages is dangerous in context of latch-up, applying $V_{BS} = 1.1$ to an NMOS will certainly trigger the parasitic bipolar transistors formed by the various wells. Hence, subthreshold operation is perfect from this point of view, as $V_{BS}$ will not reach critical levels. On the subject of 4.2 K operation, a decrease in subthreshold slope compared to static CMOS seems to be confirmed by the work in [26]. This decrease does not appear to increase towards 4.2 K, hence the reduction in subthreshold slope factor $n$ introduced by DTMOS can be considered negligible in light of the already significant attenuation of $n$ by lowering the temperature. A major downside of DTMOS is the need of a triple-well process, to isolate each individual NMOS transistor in its own p-well. Similarly, each PMOS should reside in its own n-well, however this does not require a special process. Even if a triple-well process is used, significant area overhead is attributed by hot-well spacing, which is the minimum distance between wells with different voltages. If one were to implement a DTMOS inverter under these constraints, an area increase in area of 8x is to be expected for a 40-nm process. Evidently this is unacceptable for very-large-scale integration.

It is important to stress the fact that DTMOS is a concept that can be applied to basically any logic family. For example, B-DTNMOS [41] exploits the DTMOS concept in domino logic by using the clock signal inherent to domino logic as back bias, to experience the advantages of DTMOS such as increased performance and resistance to variation.

### 2.3.4. Transmission Gate Logic

Transmission gate (TG) logic is a popular choice in literature for subthreshold operation [42]. Transmission gate logic is essentially a form of pass-transistor logic, without the downsides of pass-transistor logic. A figure of a NOR implemented with transmission gates is shown in 2.9.

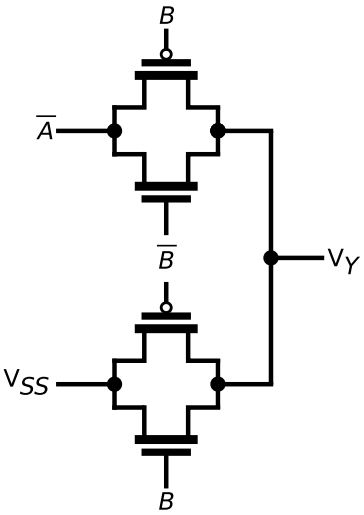



Figure 2.9: NOR gate build from transmission gates.

In pass-transistor logic as well transmission-gate, primary inputs are allowed to connect to both transistor gate and transistor source-drain terminals, contrary to selectively transistor gates in static logic. Allowing this attempts to lower the number of transistors required to implement some logic gate, reducing the total input capacitance of that gate. For example, an AND logic gate can be implemented with only two (pass) transistors [7]. This aggressive reduction in the number of transistors is not without consequences, gates implemented with pass-transistors exhibit voltage swing degradation at the output(s). For instance, an NMOS transistor can only pull some node to $V_{DD} - V_{TH}$, instead of $V_{DD}$ for the output high level. A PMOS transistor exhibits a similar characteristic for the output low level. Transmission gate logic is a direct solution for this problem, and consists of a PMOS transistor in parallel with an NMOS transistor. This means that although the NMOS is poor at passing an logic one, the PMOS will compensate for it as it is in the same path (and vice-versa). While the transmission gate logic style is effective in solving the main issue of pass transistor logic, it introduces a new issue by requiring differential inputs. As the gate typically does not provide differential outputs, additional gates in to form of static inverters is needed.

The need for static inverters when using TG logic is further exacerbated by the fact that TG logic is non-regenerative [7]. This means that TG logic can never 'restore' degraded inputs signals to the ideal values of $V_{DD}$ and $V_{SS}$ at the output, contrary to regenerative families such as static and dynamic CMOS. Degradation of signals in TG logic is a cause of the series resistance associated with the transmission gate, and other possible effects such as noise and crosstalk. To cope with the problem, static inverters are inserted at regular intervals. Another result of the series resistance of transmission gates is that delay in a network of $n$ transmission gates is proportional to $n^2$ [7] (when modelled as RC network). This scaling requires buffer inserted at regular intervals, where the optimum number of transmission gates per segment is typically three or four [7]. Luckily, as buffers consist typically back-to-back static inverters, the regeneration problem is solved concurrently by applying this optimisation.

Transmission gate logic imposes attractive properties for subthreshold operation. Primarily, TG logic is inherently more resistant to mismatch than static CMOS [42]; the two types of transistors are always together in the same path, compensating for each others weakness. Secondly, TG logic shows a significant reduction in leakage power since there is no direct path from supply to ground. For the NOR topology shown in 2.9, a reduction in leakage power of 19x was observed compared to a static NOR implementation [42]. In the context of 4.2 K operation, TG logic might endure larger improvements over static CMOS with respect to mismatch, as mismatch increases. On the other hand, the low leakage power might become a redundant feature, as leakage is already significantly decreasing.

### 2.3.5. Current Mode Logic

Current mode logic (CML) achieves high-performance, low-power operation by using low voltage signal swings. Low voltage differential signalling (LVDS) is a commonly used example of the principle. CML is a differential logic family, which imposes benefits with respect to noise immunity. Furthermore, since the current that flows through the branches of a CML gate can be adjusted, the family provides a very wide operating range in terms of performance and power. Figure 2.10 shows the basic structure, typically the load resistors $R_L$ are implemented as PMOS transistors connected to ground. While its main properties are attractive for low-power, high-performance applications, the family suffers from several complications. First of all, static power consumption exists as one of the NMOS transistors from the gate's input pair is always (partially) on. Secondly, interfacing with conventional full swing logic is not possible without additional effort, for instance the swing is too small to exceed the logic threshold of a static inverter with the same supply voltage. Finally, the bias circuitry for the current source demands additional area overhead and headroom for the current source is required. Similar to the application of DTMOS to dynamic logic, dynamic operation has been applied to CML to remove the static power consumption [43] at the expense of increased design complexity.

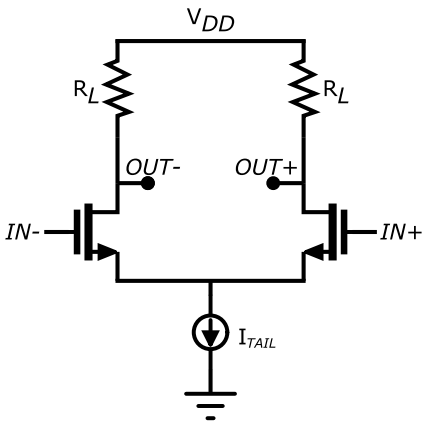


Figure 2.10: Basic CML structure.

### 2.3.6. Other Families

It is important to stress the fact that there exists a vast amount of literature on different (sub-)families. These families have been considered, but removed from further discussion as they are typically less proven or poorly (not) compared to well known families. Some logic families such as pass-gate logic are simply disregarded as they cope with issues that cannot be dealt with.

## 2.4. Low-Power Standard Cell Libraries

Various sizing strategies to exploit the subthreshold regime have been proposed to be employed in standard cells. In particular, sizing is driven by secondary transistor effects RSCE and INWE, but also by variation. The work in [44] and [19] shows considerable reduction in power and area, while improving performance by exploiting the inverse-narrow-width-effect. The sizing strategy proposed to exploit this secondary effect is to use minimum-sized transistors with more than one finger, making sure INWE is as strong as possible for each transistor present. The reverse short channel effect can be exploited concurrently to INWE, further impacting power and performance in a attractable manner [18, 19].

To account for the increase in variation in the subthreshold regime, stacking of devices is found to be effective [19, 45]. It should be noted that exploiting RSCE also combats mismatch, as $\sigma_{V_{TH}}$ is proportional to $1/\sqrt{WL}$ (Pelgrom's Law[24]). Stacking of devices (transistors) is done when exploiting INWE, stressing the importance of INWE based sizing in the subthreshold regime. Finally, body biasing (i.e. DTMOS) can help in mitigating variation [45, 46].

Considerations regarding cell strength, logic style, fan-in and transistor type can be made in the subthreshold regime. Large cell strengths are of little importance in subthreshold standard cell libraries, as performance is typically not a concern in this regime. The choice in logic style can affect many parameters in significant ways, section 2.3.4 is a clear example of this. High fan-in ($> 2$) gates should specifically be avoided in a subthreshold standard cell library [22]. High fan-in gates typically have multiple transistors stacked in series, where the height of this stack is equal to the fan-in. Each of this transistors induces a drain-source voltage $V_{DS}$ drop, which translates into an increase of the threshold voltage $V_{TH}$ as dictated by drain-induced barrier-lowering (DIBL). Hence, increasing the number of series transistors translates into an exponential reduction in current in the subthreshold regime. Figure 2.11 shows this behaviour in SPICE simulation, where $X_{STACK}$ represents the total reduction in current through the stack. Ideally, for three series transistors one would expect a reduction of 3x in total current, but this is clearly not the case in the subthreshold region. Two fan-in gates, which is the minimum for most functions, shows the least amount of susceptibility to the effect. This confirms the observations made in literature.

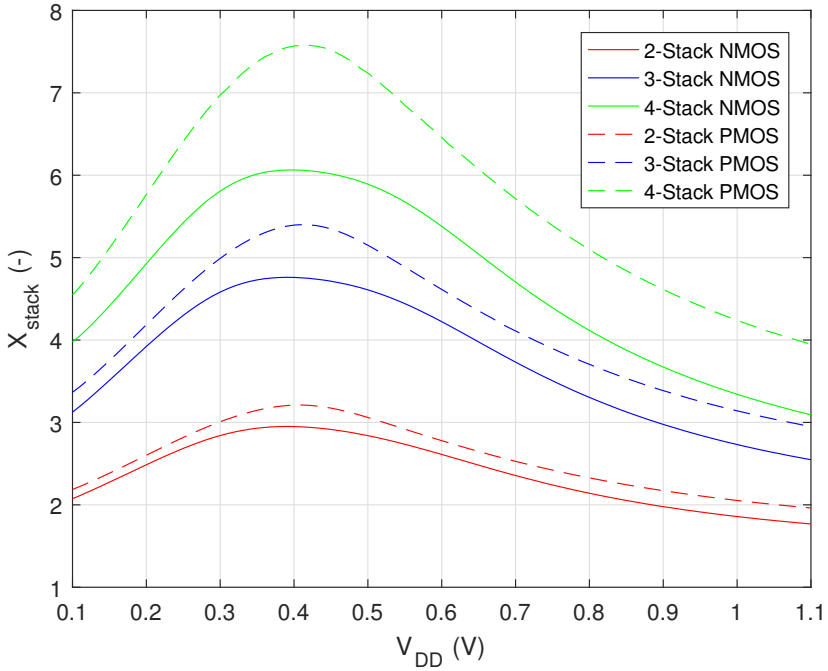



Figure 2.11: Stacking factor $X_{STACK}$ as a function of $V_{DD}$

### 2.4.1. Process

Typically fully-depleted silicon-on-insulator (FD-SOI) is a better choice over a bulk (standard) process for several reasons. The following enumeration summarises the enhancements of FD-SOI over a bulk process.

1. Lower parasitic capacitances, resulting in general faster and lower power operation [47, 48].
2. Superior subthreshold slope $SS$ [48, 49].
3. Lower DIBL coefficient. Implies a smaller raise in threshold voltage when moving to low $V_{DD}$. More importantly, the performance gap between bulk and FD-SOI becomes larger as $V_{DD}$ becomes lower. For instance, the work in [49] found a 32% increase in performance at $V_{DD} = 1.0V$, while a 84% increase was observed at $V_{DD} = 0.6V$.
4. Absent random dopant fluctuations, as there are no dopants in the channel. A reduction in $\Delta_{V_{TH}}$ of 40% has been reported [49].
5. Immune to latch-up, as the parasitic bipolar transistor structure is not present [47].
6. Very effective forward/reverse body biasing, in principle any bulk voltage can be applied due to the lack of parasitic bipolar transistors [49].

FD-SOI offers attractive properties for subthreshold operation and is consequently considered the standard for low-voltage high-performance operation.

# 3

# The Standard Cell Library, 'CooLib'

This chapter focuses on standard cell libraries, in particular on details regarding implementation. First, a literature review on low-power standard cell libraries will be given, to get an understanding of the state-of-the-art at room temperature. Secondly, a proposal is made for a cryogenic, lower-power standard cell library, based on the analysis from chapter 2. Finally, the process of building a standard cell library compatible with commercial design automation tools is explained.

## 3.1. Proposed Standard Cell Layout Properties

Based on the analysis made in chapter 2, the following standard cell characteristics are proposed (table 3.1). A justification of the proposed characteristics is provided subsequently.

Table 3.1: Comparison between state-of-the-art low-power, subthreshold libraries and the proposed cryo-subthreshold library.

| 300 K Low-Power | 4.2 K Proposed Low-Power |
|---|---|
| Small cells | Large cells |
| Possibly Multi-$V_{TH}$ (MTCMOS) | Exclusively low-$V_{TH}$ |
| $L > L_{min}$ (RSCE aware) | $L = L_{min}$ (RSCE reduced at 4.2 K) |
| $W = W_{min}$ (INWE aware) | $W = W_{min}$ (INWE aware) |
| Low drive strengths | Low drive strengths |
| Restricted to low fan-in | Restricted to low fan-in |
| Possibly forward/reverse biased | Possibly forward/reverse biased |
| Separate well-tap cells | Integrated well-taps |
| Ideally implemented in FD-SOI | - |

Large cells are proposed to oppose the increase in mismatch at cryogenic temperatures. Standard cell dimensions are expressed as multiples of the lowest layer metal (and possible higher layers) pitch, which is more commonly known as a routing track. This property makes it trivial to align standard cells on the routing grid. Specifically, a cell height of fifteen tracks ($0.14\mu m$) is adopted, which leads to a cell height of $2.10\mu m$ in the given process (TSMC 40-nm CMOS). As a side note, the cell width is determined by cell function and is again expressed as multiples of the track width.
The tiny amount of information available with respect to the INWE effect implies that the effect worsens at 4.2 K, implying larger performance gains by maintaining the minimum device width. Hence, the INWE sizing strategy is adopted similar to 300 K. The minimum width constraint for transistors in combinations with a 15-track height allows for four transistors per side (N & P-type). The expected improvements incorporating the INWE sizing strategy (stacked minimum sized transistors) are depicted in terms of transition time mean and standard deviation in figure 3.1 and figure 3.2 respectively.
What is interesting to note from figure 3.1 is that the relative performance increase is inversely proportional to $V_{DD}$. A possible explanation is similar to the behaviour found in 2.11, since DIBL diminishes as $V_{DD}$ approaches zero, the overall effect of the INWE effect on $I_{ON}$ is amplified. Secondly, the observation can be made that the sizing strategy shows similar improvements for both N- and P-MOS transistors. Taking a closer

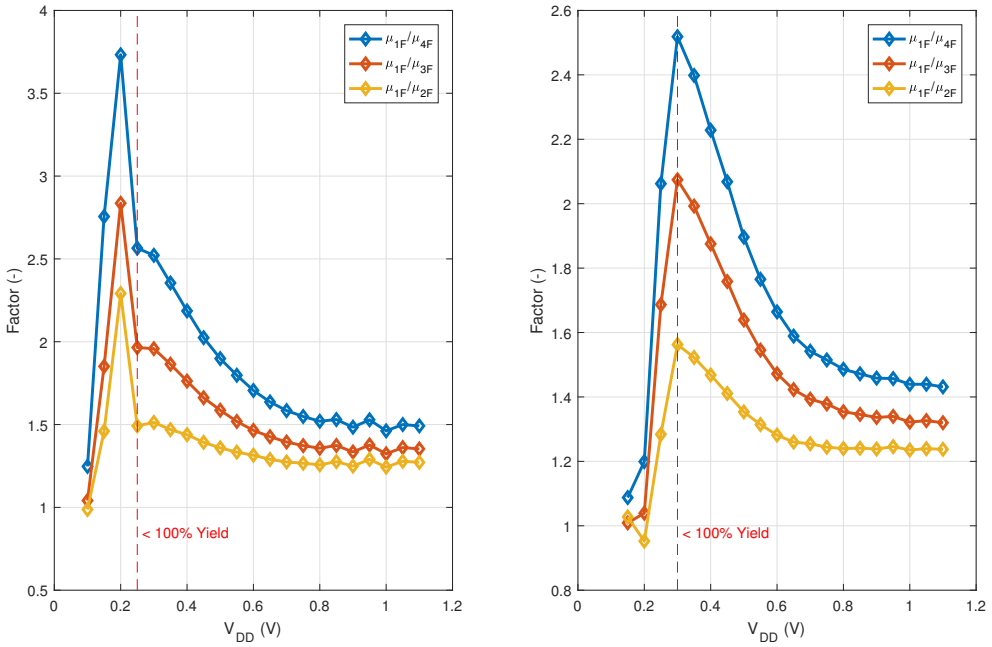



Figure 3.1: Average improvement for $\tau_{HL}$ (left) and $\tau_{LH}$ (right) over 1000 simulations (Monte Carlo). Results below the 100% yield line should not be considered.

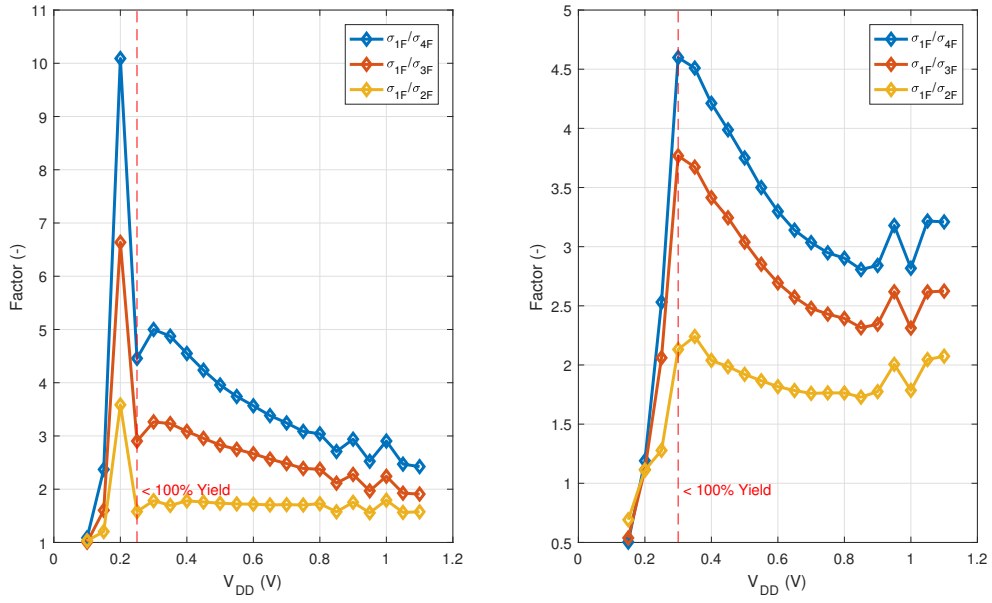



Figure 3.2: Improvement in for $\sigma_{HL}$ (left) and $\sigma_{LH}$ (right) over 1000 simulations (Monte Carlo). Results below the 100% yield line should not be considered.

look to figure 3.2, a similar trend is followed for the standard deviation of the transition times. Evidently, a reduction in standard deviation is again found to be inversely proportional with $V_{DD}$. This shows that apart from device up-sizing (Pelgrom's law [24]), the subthreshold region offers a secondary countermeasure for mismatch, which scales with $V_{DD}$. The aforementioned properties thus make an INWE aware strategy the preferred choice.

Considering standard cell drive strengths, the restriction of selectively low drive strengths is adopted for 'CooLib'. Motivation lies in the fact that power is the component of interest, contrary to performance. Secondly, for practical and logistical reasons in context of the project duration, it is infeasible to implement a large set of distinct drive strengths.

The adoption of low fan-in gates is motivated by the explanation given in section 2.4, the reader is encouraged to refer to that section for a more detailed analysis.

Finally, due to the expected reduction in leakage current $I_{OFF}$ and increase threshold voltage $V_{TH}$, only low threshold voltage (LVT) devices are used. To allow for even more aggressive reductions in threshold voltage, the 'CooLib' standard cells are proposed to incorporate possible forward-biasing by means of a multi-rail format. More specific, apart from the two 'default' rails $V_{DD}$ and $V_{SS}$, two bias rails $V_{NW}$ and $V_{PW}$ are provided which are required to connect to the n-well and p-substrate, respectively. An issue closely related to forward biasing is latch-up, as voltages applied to the substrate might forward bias the parasitic BJTs present in the substrate, which is one of the conditions to trigger and maintain latch-up (as discussed in section 2.2.5). In the context of 4.2 K, it was discussed that latch-up is highly unpredictable for reasons already stated in section 2.2.5. The significant increase in substrate resistance [1] combined with an increased substrate current might lead to significant IR values.

To estimate limits on forward biasing at 4.2 K, substrate resistance measurements from [1] are combined with room temperature simulations of substrate current. Note that this a very rough estimate, but it is the best estimate currently possible. Figure 3.4 shows the results. Based on this estimation, operating at the nominal biasing conditions is safe even at 4.2 K where the substrate resistance becomes significant. Operating at forward biasing conditions displays significant IR at $V_{DD} \approx 700mV$, indicating that caution is advisable. It should be noted that for these simulations, the forward biasing voltage is equal to the supply voltage, $V_{DD} = V_{BS}$. Ultimately, the back-bias voltage can be chosen arbitrarily, although limited to the boundaries defined by junction breakdown boundaries, regardless of $V_{DD}$. This implies that larger back-bias voltages could be possible (without generating a large IR) at lower supply voltages, since $I_{BS}$ is a strong function of the drain current $I_{DS}$.

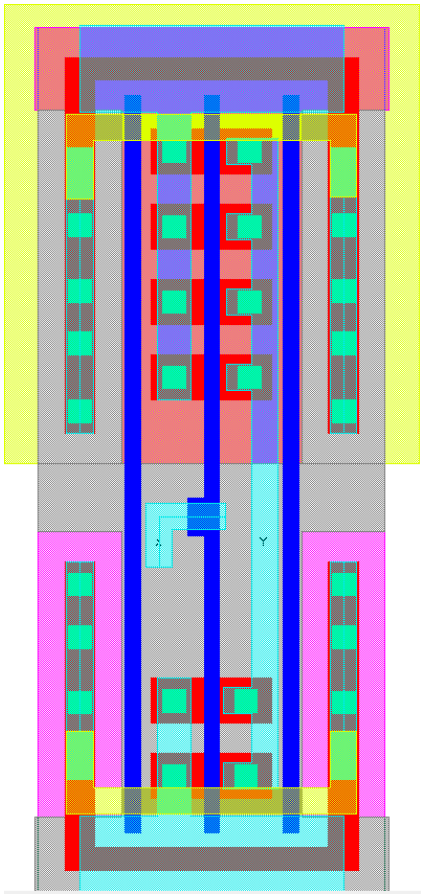
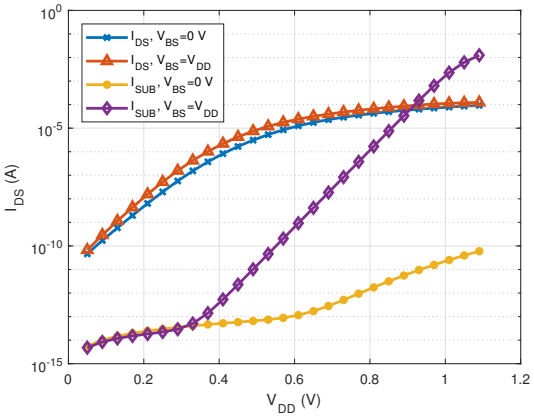
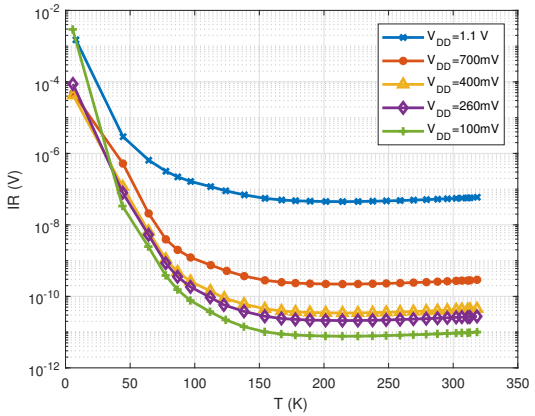


Figure 3.3: Additional well taps on cell boundary.

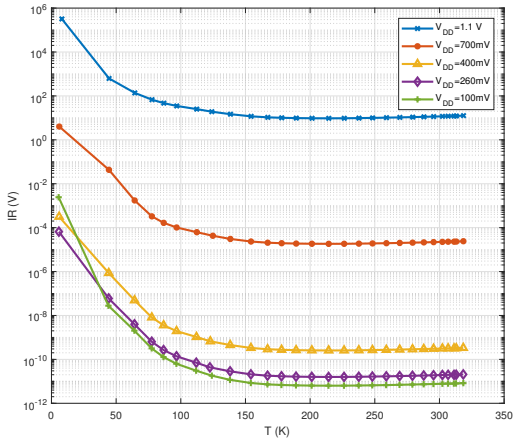
Additionally, to compensate for the increase in substrate and well resistance, well taps should be placed more frequently. To keep the area overhead induced by the additional well taps small, well tap standard cells are avoided. A better option was found to integrate vertical stripes of well taps on standard cell boundaries, as the distance (proportional to the resistance) from the transistors to the taps (pick-up) is minimised this way. Figure 3.3 shows the standard cell layout template with the well taps in place applied to a basic inverter. The pull-up section of the cell contains a total of ten well taps to the n-well, contrary to the eight well taps of the pull-down section, as the pull-up section typically is larger in (vertical) size. This approach to the placement of well taps minimises the distance as much as possible, however the distance from transistors present in the horizontal centre of large cells is larger than that of others. Still, any cell dimension is (far) below the distance for which the well resistance was measured in figure 3.4, which means the IR estimation is too pessimistic for most standard cells proposed.




(a) Simulation of $I_{SUB}$ ($I_{DS}$) for a minimum-sized LVT NMOS (TSMC 40-nm) with $V_{BS}$=$V_{DD}$.




(b) Estimation of voltage ($I_{SUB} \times R_{PSUB}$) across substrate under nominal biasing conditions $V_{BS} = 0$ (NMOS).




(c) Estimation of voltage ($I_{SUB} \times R_{PSUB}$) across substrate under forward biasing conditions $V_{BS} = V_{DD}$ (NMOS).

Figure 3.4: IR estimation under 4.2 K operating conditions.

## 3.2. Proposed Standard Cell Library Content

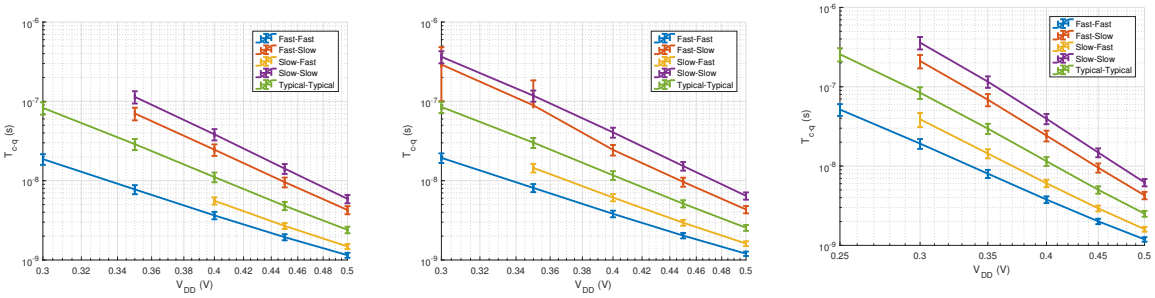
Given the duration of the project, only a minimal set of cells (with respect to function and strength) can be implemented. The main requirement of this set is that it should allow commercial synthesis tools to implement most functions described in RTL. Furthermore, specific cells are needed to convert the nominal voltage-levels provided by I/O cells into voltage-levels lying in the subthreshold region. Table 3.2 lists the proposed cells with their function and additional information. The layout of all cells proposed is found in appendix A.

Table 3.2: Proposed standard cell library content.

| Name | Boolean Function | Logic Family (Style) | Drive Strengths (Sizes) | Fan-in |
|---|---|---|---|---|
| Inverter | $Y = \neg X$ | Static | 1/2/4X | 1 |
| (Clock) buffer | $Y = X$ | Static | 1/2/4X | 1 |
| Nand-2 | $Y = A \wedge B$ | Static | 1/2X | 2 |
| And-Or-Invert-2 | $Y = (A \wedge B) \vee C$ | Static | 1/2X | 3 |
| Positive Edge D-type Flip-Flop (No Reset) | N/A | Transmission Gate | 1/2X | 1 |
| Multiplexer-2 | $Y = (A \wedge \neg S) \vee (B \wedge S)$ | Transmission Gate | 1X | 2 |
| And-2 | $Y = (A \wedge B) \wedge CLK$ | Domino | 1/2X | 2 (3) |
| Or-2 | $Y = (A \vee B) \wedge CLK$ | Domino | 1/2X | 2 (3) |
| And-Or-2 | $Y = (A \wedge B \vee C) \wedge CLK$ | Domino | 1/2X | 2 (3) |
| Level-Shifter (Up) | $Y = X$ | Static | 1X | 1 |
| Level-Shifter (Down) | $Y = X$ | Static | 1X | 1 |
| Power Gate (Header) | N/A | Static | 1X | N/A |
| Decap | N/A | Cross-Coupled | 8/16 Track | N/A |
| Tie-Hi | N/A | Static | 1X | N/A |
| Tie-Lo | N/A | Static | 1X | N/A |
| Filler | N/A | N/A | 1/2/4/8/16/32/64 Track | N/A |

The core set of combinational cells is implemented with the static logic family. This decision is forced by commercial tools, as they primarily focus on (require) static logic. In principle, only the NAND is required to implement any combinational circuit due to its functional completeness, however most synthesis tools require the presence of inverters as well. Inverters and buffers are implemented in three distinct drive strengths, contrary to all other cells, as they are used in e.g. clock tree synthesis and dealing with hold violations. Also note that the basic OR gate is missing. This is due to the fact that a static OR gate implementation consists of two PMOS transistors in series as pull-up, and two parallel NMOS transistors as pull-down. Based on the analysis in section 2.1, it is likely that the pull-down network in the OR gate configuration will be the dominant factor, limiting $V_{DD,min}$ at the subthreshold region due to swing degradation. As an alternative, a and-or-invert (AOI) cell is provided. The AOI gate suffers from the same imbalance issue as the OR gate, but only under specific input conditions. The AOI gate is not low fan-in (low fan-in is considered two), but its structure consists of two series transistors at maximum, therefore not suffering of the issues presented in figure 2.11. Additionally, a multiplexer is added to the combinational set as they are commonly used, specifically in the chip referred to in chapter 4. The static implementation of a multiplexer is rather bulky (inefficient), contrary to the transmission gate style which presents the most simple implementation of a multiplexer. A static inverter is integrated into the cell to provide the complementary inputs.

The set of sequential cells consists of only one cell, which is the simplest possible D-type flip-flop (DFF), without reset. In the context of this project, no other types of flip-flops, or latches, are needed. Furthermore, this single cell allows implementation of any register for finite-state machine purposes. The logic style for this cell was determined to be based on transmission gate, due to its excellent performance in the subthreshold region in terms of $V_{DD,min}$ [50]. A SPICE simulation over process corners with mismatch confirms this as shown in figure 3.5. This is likely to be explained by the resistivity to mismatch (imbalance) inherited by the logic style. Other popular choices for subthreshold DFFs include a pass gate based architecture, and a architecture commonly known as PowerPC 603 (also partially transmission gate based) [51], of which simulation results are shown in figures 3.5a and 3.5b respectively. While these two are slightly faster than the pure transmission gate based implementation, they are more prone to low-voltage operation (and process corners), exacerbating the need for the pure transmission gate based implementation. Note that cell area for

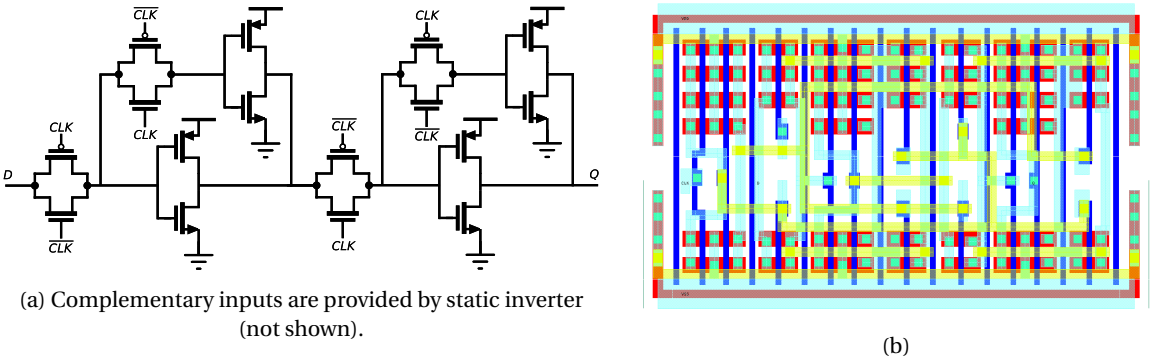



(a) Pass gate based DFF architecture.

(b) PowerPC 603 DFF architecture.

(c) Transmission gate based DFF architecture.

Figure 3.5: Performance and functional yield ($3\sigma$) for popular DFF architectures (logic styles) in the subthreshold region.




(a) Complementary inputs are provided by static inverter (not shown).

(b)

Figure 3.6: Schematic (a) and layout (b) of the proposed transmission gate D-type flip-flop (without reset).

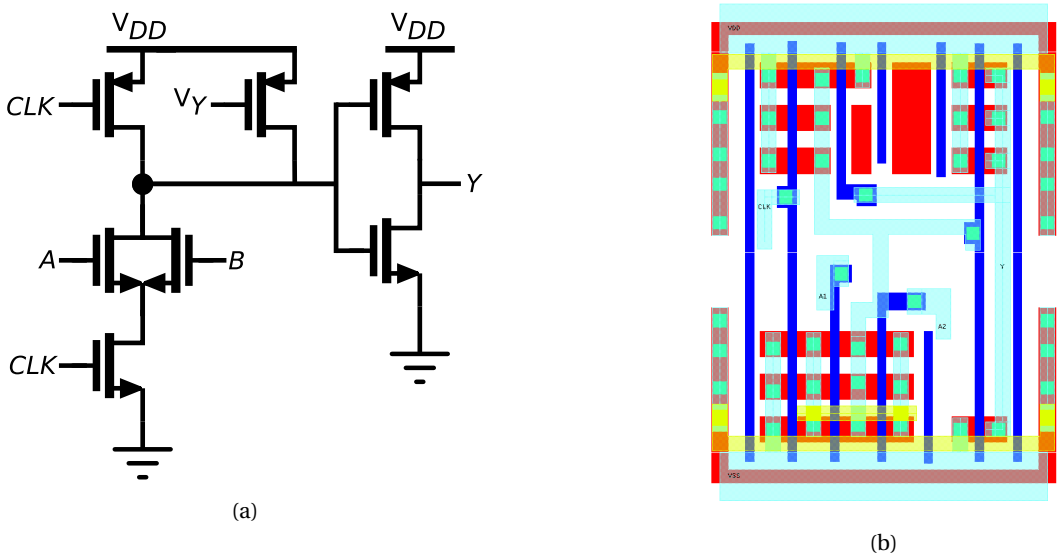
the different architectures is not of interest for this standard cell library. Figure 3.6 shows the schematic and implemented layout following the properties proposed in section 3.1.
Finally, a set of combinational domino logic cells is implemented, given this particular logic family might benefit the most from operating at 4.2 K, as discussed in section 2.3.2. To test the behaviour of 'true' domino logic at 4.2 K, a version of every domino cell is provided without weak-keeper. Contrary to static logic, an OR gate should be implemented without pull-up/down imbalance in domino logic. The pull-down network of a domino gate houses at least two series NMOS transistors (three in AND gate scenario), which is more than desirable as per figure 2.11. This could hamper the domino logic's performance in subthreshold, even though NMOS transistors are less susceptible to series stacking compared to PMOS transistors. Figure 3.7 shows a typical domino gate schematic and proposed layout.
In the end, while current mode logic is a viable choice due to its characteristics (section 2.3.5), it is avoided for several reasons. Mainly, the differential structure imposes incompatibility with commercial synthesis tools. Secondly, having a bias current per cell adds significant design and implementation complexity. Finally, its incompatibility with static CMOS makes it hard to justify CML standard cells. However, CML is considered to be perfectly viable for (full-) custom blocks.

### 3.2.1. Low-Power Implementation Cells

Standard cells specific to low-power implementation include level-shifters, power gates, retention flip-flops and isolation cells amongst others. Low-voltage operation without proper level conversion is impossible, as low-voltage signal swing typically lies below the logic threshold $V_M$ (ideally $V_M = V_{DD}/2$ [7]) of gates operating at the nominal voltage. Hence, level-shifters take care of converting signals between power domains, where there is typically a low-voltage power domain and a nominal (high) voltage-domain. Additionally a difference is found in converting from low to high (LH) and vice-versa (HL) in terms of cell architecture.

#### Low-to-high shifting

LH conversion is typically done using a differential half-latch based structure [52], but has significant disadvantages in terms of area (in the subthreshold) as the circuit is ratioed. More importantly, in most LH




(a) (b)

Figure 3.7: Schematic (a) and layout (b) of the proposed domino OR gate.

architectures sizing is not trivial, in context of 4.2 K this would impose very unpredictable behaviour. The requirements for LH conversion architecture are thus simplicity (or predictability towards sizing) and very low-voltage up-conversion ($< 200mV$). Conversion delay, static power consumption and energy are typical metrics which are critical for LH conversion as well. Given the project context , energy can be neglected and delay has a lower priority compared to the aforementioned two requirements. While static power consumption is important in the context of the project, preference is given towards lower $V_{DD}$ conversion, i.e. an architecture which consumes considerable static power but achieves very low $V_{DD,min}$ will be the preferred choice. Motivation for this trade-off is twofold, primarily it is desirable to test over a very large range of $V_{DD}$ (LH conversion should not limit $V_{DD,min}$) and the number of LH cells in typical circuits is small, meaning the overall contribution in static power consumption can be considered negligible. An architecture which was found to the fulfil requirements is based on a (PMOS) differential pair with active current mirror [53]. Initial simulations showed that even without sizing this structure was able to convert signals as low as 300 mV to the nominal core voltage of 1.1 V. However, conversion below 200 mV was found to be troublesome. To deal with this problem an additional stage was introduced to convert the low voltage signal into an intermediate voltage signal, which is subsequently converted into the required output voltage signal. This approach was also employed in [15], where a three-stage LH shifter was proposed to achieve 62 mV operation. While adding an intermediate voltage level produces additional difficulties in i.e. power routing, it was found to be unavoidable in the search of very low $V_{DD}$. Figure 3.8 shows the final schematic of the LH shifter. After the two differential stages a buffer connected to the high voltage level is introduced to fully restore the signal swing, as the logic high output of the second stage was found to be about 80-90% of $V_{DDH}$.
The LH shifter was found to be functional (SPICE simulation, TT corner) down to 20 mV, which remarkably is below the theoretical limit at room temperature. This simulation was performed with ideal inputs however, with realistic inputs provided by inverters the lower limit was found to be approximately 50 mV. While the range provided by this LH shifter is excellent, it was found to be very sensitive to mismatch and global process corners. Over a Monte Carlo simulations of 1000 points, the functional yield at $V_{DD,low} = 30mV$ was found to be about 10%. With respect to process corners, the SF corner was found to be functional down to $V_{DD,low} \approx 10mV$, while the FS corner showed failure at $V_{DD,low} \approx 140mV$. Additionally, function yield was identical over the TT, SS and FF corners, clearly indicating that imbalance is the limiting factor. To balance FS and SF corners, the PMOS input pair is implemented with high-threshold voltage devices, furthermore the length is slightly increased. To deal with mismatch the amount of fingers was increased to 24, at which a functional yield of $3\sigma$ was achieved for $V_{DD,low} = 30mV$. It should be noted that tuning the intermediate voltage $V_{DDI}$ can be used to combat the effects of global process corners. Figure 3.9 illustrates this, a dot represents a successful conversion from the low-voltage to the high-voltage level, given the intermediate voltage level on the Y-axis. The green area represents Monte Carlo results over the TT, SS and FF corners and shows that $V_{DDI}$ can almost be picked arbitrarily. The red area represents the SF corner, while the blue area shows the FS

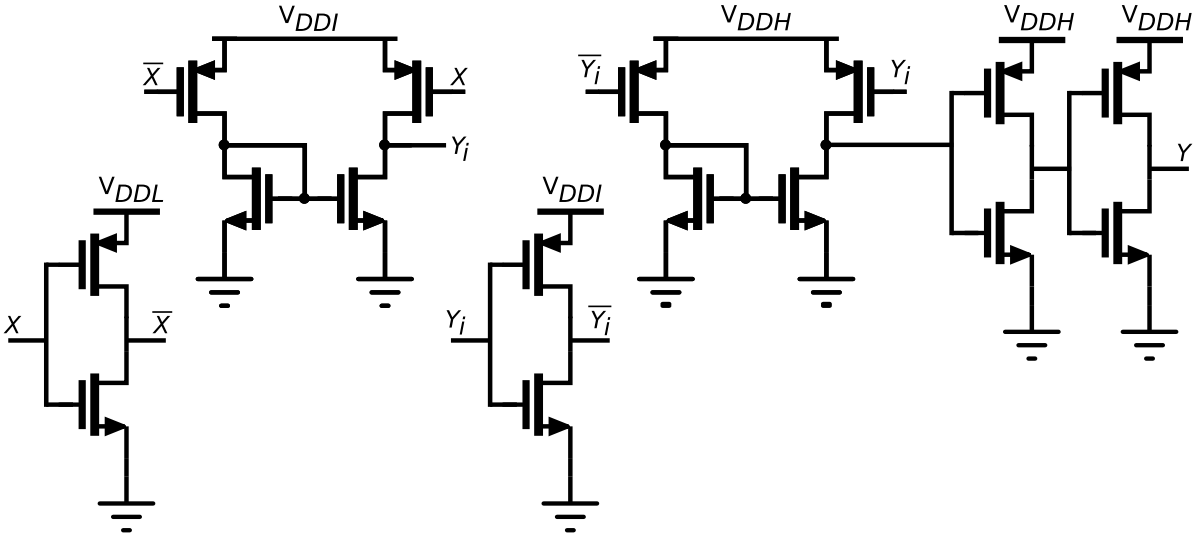



Figure 3.8: Proposed two-stage low-to-high level shifter. $V_{DDL}$ represents the low voltage level, $V_{DDI}$ and $V_{DDH}$ represent the intermediate and high (nominal) supply levels.

corner. Intuitively, in the FS corner successful conversion is achieved for higher $V_{DDI}$, since the PMOS input pairs are weaker. Finally the purple area indicates the region which is common to all corners, which appears as low as $V_{DDL} = 110mV$.
To conclude, the layout of the LH shifter is shown in figure 3.10. The NMOS pair (per stage) forming the current mirror is placed in a column to avoid long interconnect between the two. This leads to a double height standard cell layout. The secondary power supplies $V_{DDI}$ and $V_{DDH}$ are implemented as secondary metal stripes across the PMOS transistors. Note that there are two rails per secondary supply which are not connected internally (to the cell). This implies that both rails should be explicitly connected during the power routing stage of a typical digital implementation flow. The area of the standard cell is not fully occupied due to a size mismatch of the P- and N-type transistors. More exotic LH shifter cells, for instance with cell enable, could make use this available area.

### High-to-low shifting

High-to-low (HL) shifting is typically considered trivial in the sense that simple inverters are capable of doing so. Normally a signal coming from a high-voltage (nominal) domain can be applied to an inverter which is connected to a low-voltage supply, providing the low-voltage inverter signal at its output. Considering the nominal voltage of the process in context, this method works down to approximately 200 mV (from 1.1 V). After this point significant under- and over-shoot starts appearing, which is a result of the steepness of the input signals in conjunction with the parasitic gate-drain capacitance of the inverter [7]. Furthermore, at these very low supply voltages the charge stored in this parasitic node cannot be shunted quickly enough, causing over- and under-shoot. Under-shoot is particularly destructive, as it limits the output in reaching the logic high level, or $V_{DDL}$. This is displayed in figure 3.11a, where a 400 mV input signal is applied, and $V_{DDL} = 50mV$. The 400 mV output swing in this simulation is the result of a prior down conversion to get a realistic transition time, note there is no over- and under-shoot issues in this first stage. The conversion from 400 mV into 50 mV shows significant undershoot, which indeed prevents the signal from reaching the logic high level in time. Note that it will eventually reach the logic high level, but it induces a significant penalty in terms of delay. The obvious solution is to limit (decrease) the slew rate, giving the transistors effectively more time to shunt the parasitic charge. However this also implies penalties in terms of delay, as aggressive up-sizing of device length is one way to accomplish this.
The proposed solution is as follows: since an intermediate voltage level is already required for low-to-high shifting, it can be used in high-to-low shifting without inducing any additional penalties. Hence, the high (nominal) 1.1 V signal is converted into the intermediate supply voltage which ranges from 200 mV to 400 mV as seen before. If this intermediate voltage signal swing $V_{OUT,S1}$ is inverted and applied to the second stage inverter PMOS transistor as dynamic (reverse) back bias, $V_{B,PMOS} = \neg V_{OUT,S1}$, the dynamic back bias couples to the inverter output through the parasitic capacitance between the bulk and drain ($C_{DB}$). The result of this capacitive coupling is that the signal overshoots when it would undershoot in the normal case. This

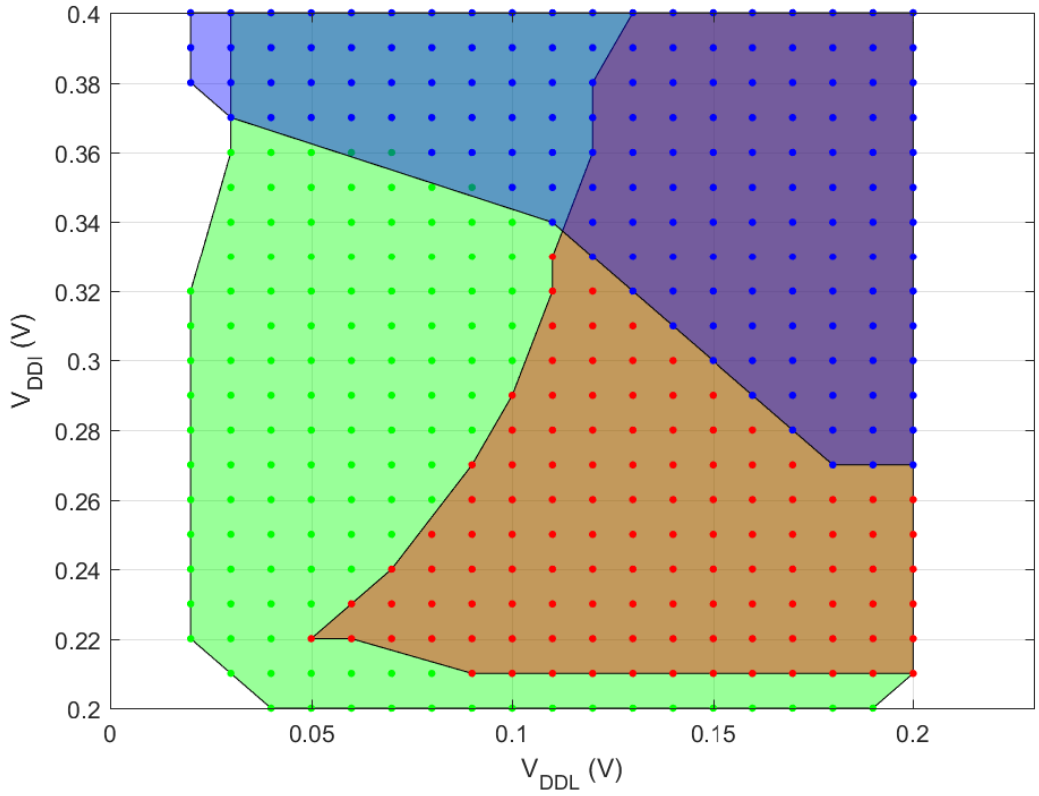



Figure 3.9: Regions of successful LH conversion in terms of $V_{DDL}$ and $V_{DDI}$. The green region represents SS/FF/TT global process corners, while the red and blue represent the SF and FS corners respectively.

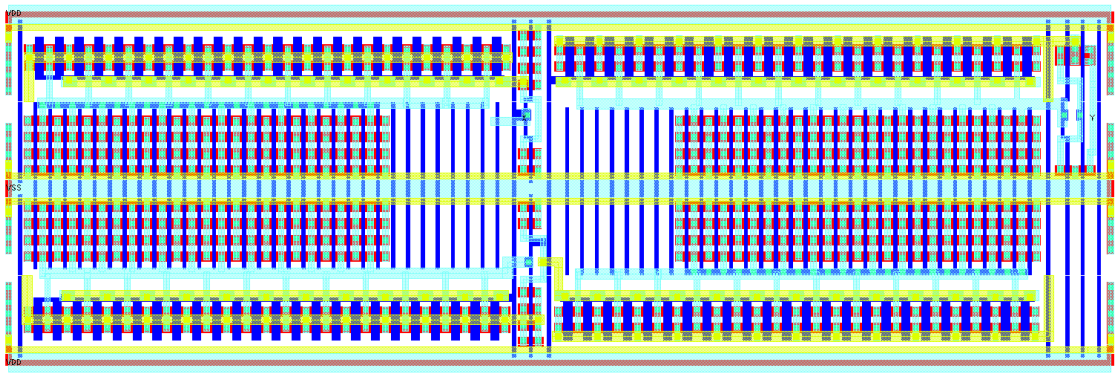
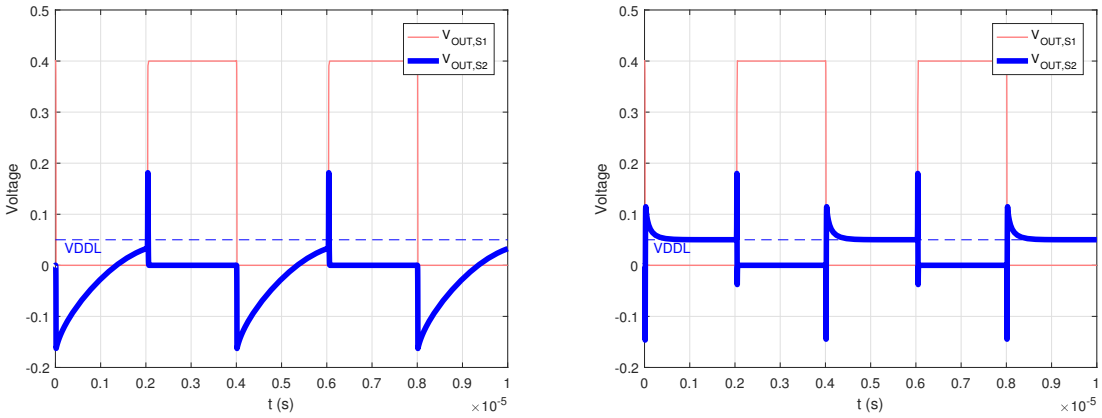
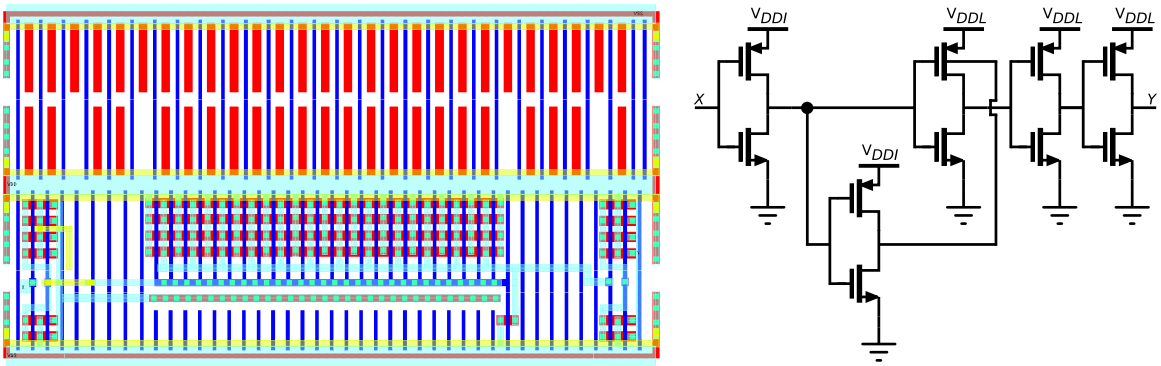


Figure 3.10: LH shifter double height ($2 \times 2.10 \mu m$) standard cell layout.




(a) (b)

Figure 3.11: Level down conversion, (a) shows conventional inverter based down shifting and its problem, (b) shows the proposed solution.




(a) (b)

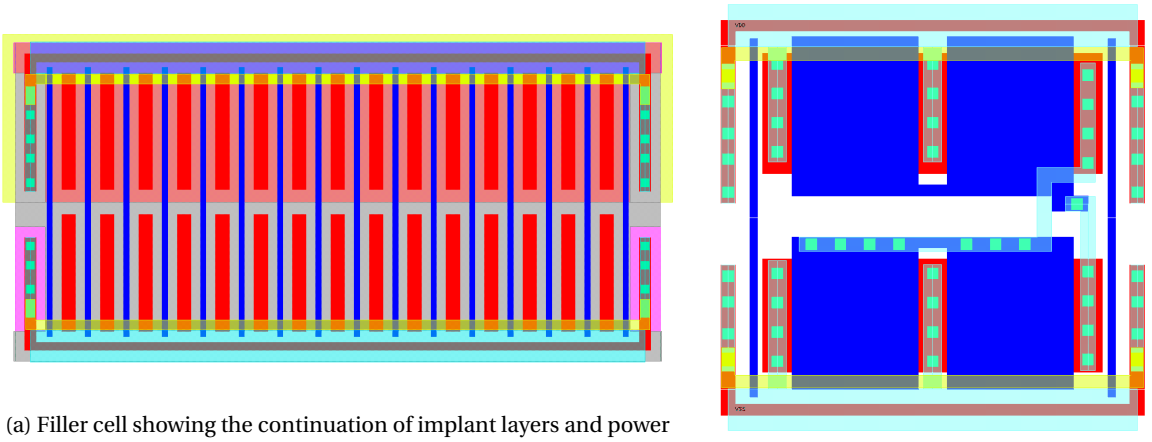
Figure 3.12: (a) shows the level-down shifter double-height ($2 \times 2.10 \mu m$) layout, (b) shows the schematic.

is displayed in figure 3.11a, the schematic is display in figure 3.12b. This solution allows down conversion to any $V_{DDL}$, at the expensive of additional power supply routing and slightly less trivial layout. In terms of layout, isolation is needed for the PMOS transistor with dynamic back bias, which is achieved by placing it in a separate n-well. Unfortunately the hot-well spacing (the distance required between wells with a different potential) requires a double-height standard cell layout, whereas the topside is effectively a filler. This layout is shown in figure 3.12a. Because of the intermediate voltage swing at the input of second stage inverter and the dynamic reverse back bias applied to the pull-up, it is advantageous in terms of output swing to increase the size of the pull-up network. Specifically, a ratio of 96-to-1 was found to work well.

### 3.2.2. Physical-Only cells

A set of physical-only cells (without boolean function) in the form of fillers, tie cells and decap cells are provided as they are (typically) required during digital implementation. Fillers are important as they connect the active implants (n+ and p+), as well as n-wells and power rails throughout an entire row. Fillers should come in various distinct widths, where the width is an integer multiple of the metal one routing pitch (track). Therefore a selection of seven widths from one track to 64 tracks is made. Fillers consist of dummy polysilicon and diffusion area, as per recommendation of the foundry (also improves density). Additionally the 'CooLib' fillers include the well-taps which are found in conventional (with function) cells such that they aid in lowering the substrate resistance. To illustrate, figure 3.13a shows the layout of the 32-track filler.

Tie-low and tie-high cells are provided to avoid direct connections (ESD concerns) to power and ground rails when there is need for a constant input. The tie cells are simply put static inverters, but with gate(s) connected



(a) Filler cell showing the continuation of implant layers and power rails.

(b)

Figure 3.13: Physical-only cells, (a) shows a 32-track filler cell, (b) shows the 16-track decap cell.

in a diode like fashion.

Finally some decap cells are provided to help mitigate IR drop issues during digital implementation. A trade-off in decap design can be made between effective capacitance $C_{eff}$ offered by the cell, cell leakage and ESD rating (effective series resistance $R_{eff}$). In the context of 4.2 K operation, focus should lie on optimising $C_{eff}$ and $R_{eff}$ due to expected reduction in leakage current. Typically two types of decaps are considered, one based on moscaps and one based on a cross-coupled structure, where the PMOS (NMOS) gate is connected to the drain of the NMOS (PMOS). The latter structure offers better ESD protection (higher $R_{eff}$), but may compromise the frequency response as a result. High frequency operation is not intended with the proposed standard cells, hence this cross-coupled structure is the preferred one. Optimal sizes in terms of transistor sizes were found based on the equations derived in [54]. A 16-track configuration with $W_n = 0.62\mu m$, $W_p = 0.66\mu m$ and two fingers of both transistor types is found to perform well (compared to an existing cross-coupled decap cell supplied by the foundry), which can be viewed in figure 3.13b. Similar, an 8-track decap cell is offered to provide some level of choice.

## 3.3. Timing Information Extraction

An implemented digital circuit should not impose timing violations, as these might disrupt the functionality of the chip. Most commonly known timing violations are setup and hold time violations with respect to sequential elements (i.e. flip-flop). Considering the case of a simple D-type positive-edge triggered flip-flop, setup time represents the time margin where the data signal should be stable before the positive clock edge appears. Similarly, hold time represents the time margin where the data signals should remain stable after the positive clock edge. Clearly, both setup and hold constrains are a function of the flip-flop's internal structure. Setup and hold analysis is typically performed in separate (extreme) process corners, setup is performed in the slow-slow corner, while hold analysis is performed in the fast-fast corner. This analysis is performed by design automation tools, such as synthesis and place-and-route tools, and can subsequently act on this result of this analysis.

Surely a SPICE simulation running on typical VLSI digital chips would be unsuitable for this analysis, as execution time increases quickly with the number of gates (transistors) in the simulation. Instead, standard cells are once characterised with the result of this characterisation stored in a database in the form of look-up tables. Consequently, design automation tools can interpret the information from this database and perform quick analysis by table look-up.

During the characterisation process additional information is captured in the form of power/energy consumption and signal integrity (noise propagation). All of this information is stored along timing data on a per cell basis. The most commonly used (standard) format is the liberty format[55], which will receive a more in depth required for proper characterisation of standard cells.

### 3.3.1. The Liberty Format

As stated before, the liberty format mainly consists of look-up tables for delay, power and noise propagation. For delay and power, these tables are two-dimensional containing slew rate (transition time) on one axis, and output load on the other axis. The dimensions of these matrices can be defined, i.e. larger tables would require more simulations. In the end, larger tables might be slightly more accurate when the design-automation tool performs interpolation. The table is bound by maximum values for slew rate and output load, where the maximum output load is often determined by constraining the maximum slew rate. Slew rate is measured between a lower and upper threshold, which might be different for rising and falling transitions. For older processes, these thresholds were typically between 10 and 90% of the full voltage swing [7], as this was the region where the slew rate was most linear. For process nodes of 40-nm and below, this band is reduced to 30% and 70% respectively, as this is the most linear region for these processes [56]. This transition band is only 40% of $V_{DD}$, which results in a misrepresenting slew rate. To translate 30-70% band to 10-90%, a slew derate factor is introduced and is determined as follows

$$\text{slew\_derate} = (70-30)/(90-10) = 0.5 \tag{3.1}$$

The slew derate factor can be specified in the liberty format, such that design-automation tools correctly determine the slew rate. Further important information stored in the liberty file include input/output voltage levels, input pin capacitances, output pin (boolean) functions and the units of the values present in the look-up tables. To illustrate, listing 3.1 shows the falling edge timing section of a simple inverter.

```
1  pin(Y) {
2      direction : output ;
3      function : "(!X)" ;
4      max_capacitance : 0.01607 ;
5      min_capacitance : 0.0004 ;
6      output_voltage : default ;
7      related_ground_pin : VSS ;
8      related_power_pin : VDD ;
9
10     timing() {
11             related_pin : "X" ;
12             timing_sense : negative_unate ;
13             timing_type : combinational ;
14
15             cell_fall(tmg_ntin_oload_7x7) {
16               index_1("0.02, 0.06281, 0.2131, 0.5001, 0.9472, 1.575, 2.4");
17               index_2("0.0004, 0.0006918, 0.001716, 0.003673, 0.00672, 0.011,
                     ↪ 0.01662");
18               values("0.008743, 0.009859, 0.01326, 0.01864, 0.02667, 0.038,
                     ↪ 0.05279",\
19                      "0.01211, 0.01388, 0.01923, 0.02762, 0.03816, 0.05034,
                           ↪ 0.06505",\
20                      "0.01596, 0.01864, 0.02699, 0.04018, 0.05709, 0.07698,
                           ↪ 0.09927",\
21                      "0.01763, 0.02108, 0.03218, 0.04966, 0.07228, 0.09915, 0.13",\
22                      "0.01761, 0.02218, 0.03492, 0.05633, 0.08531, 0.1192, 0.1581",\
23                      "0.01679, 0.02086, 0.03493, 0.06, 0.09339, 0.1338, 0.1793",\
24                      "0.01552, 0.01968, 0.03441, 0.06053, 0.0975, 0.1435, 0.1972");
25             }
26         }
27 }
```

Listing 3.1: A segment of liberty timing information.

### 3.3.2. Simulation Based Standard Cell Characterisation

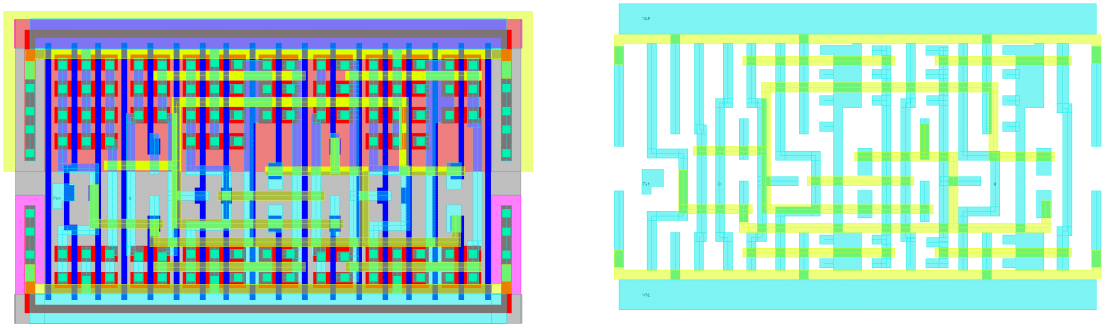
As the implementation of the 'CooLib' standard cells targets an existing technology (TSMC 40-nm), accurate foundry supplied models for SPICE simulation are available. One can in theory perform manual SPICE simulations on standard cells and populate the liberty database manually, but this would be a tedious task. As a result, commercial solutions exist which are specifically aimed to automate standard cell characterisation. Most commonly known tools are Liberate from Cadence, and SiliconSmart from Synopsys. Both solutions offer all of the state-of-the-art models, support for standard cells, I/O, complex (low-power) cells and memories. Furthermore, they support massively parallel operation, function recognition and features such as data-sheet generation. Both solutions can be considered equal, all features required for standard cell characterisation for the given process node are available. Synopsys SiliconSmart was ultimately chosen for its clear documentation and well-defined characterisation flow.

SiliconSmart [57] takes SPICE models of the standard cells as input, as well as configuration files. The SPICE models should contain parasitic R and C values to replicated the manufactured cell as much as possible. For this project, parasitic extraction was performed with Calibre xRC in batch mode to perform automatic extraction on the large amount of cells. Perhaps the most challenging aspect of characterisation is the configuration of the software. This includes setup of process corners, voltage levels and temperature (PVT), setup of liberty specific parameters and setup of cell configuration files. For PVT corners, best-case (BC) and worst-case (WC) are important as they are used in setup and hold analysis. The PVT values for these corners were set to match the PVT values available in the liberty supplied with the standard cells for the target process. In addition, for 'CooLib' additional back bias voltages need to be defined per corner, which were set to the default reverse bias voltages. Finally, secondary supplies for level-shifter cells are set to match $V_{DD}$ as they are not meant to be fixed voltages. For timing closure in the presence of secondary supplies and back bias supplies, additional timing corners are typically created such that setup and hold can be analysed under these secondary conditions. An important detail to note is that having supplies for back biasing and/or secondary supplies is not compatible with the 'vanilla' liberty format. For this purpose, the liberty multi-rail format exists which comes in two versions, V1 and V2. After trial and error, V2 was the only format found to be compatible with the place-and-route tool. The actual process of characterisation with a complete configuration consists of several steps. Primarily, cell netlists with parasitics are imported and functional recognition is applied. The next step consists of setting additional cell parameters, for instance the area attribute, and configuration based on the desired source model to be used for delay, power and noise calculation. Without going into detail of the available souce models, the Effective Current Source Model (ECSM, Cadence) is chosen as it is recommended by the place and route tool used in chapter 4. As a side note, qualification of the ECSM timing output consists of a consistency check between the different available models, implying that a choice for a model follows not from the results, but rather from the output format.

The subsequent step is the actual characterisation. Given the relative small number of cells to characterise, Monte Carlo simulations (1000) were used to gather variation data. For larger libraries this method is infeasible due to the run-time associated with Monte Carlo, and thus other methods are preferred. The final step consists of exporting the data to useful formats. Timing, power and noise data is aggregated in distinct liberty files based on the corner. Variation data is typically exported as a side file consisting of coefficients (effectively sigmas), which is modelled based on the Advanced/Parametric On-Chip Variation (AOCV/POCV) format. Even thought these side formats are deemed sufficient for a 40-nm process [55], a newer format is introduced by the name of Liberty Variance Format (LVF) which aims to further reduce timing pessimism induced by variation than AOCV/POCV. This format contains variation data of every timing arc in the cell, contrary to AOCV/POCV, and is integrated as an additional table in the liberty file. As mismatch increases at the target 4.2 K operating point [32], any reduction in timing pessimism is desired, therefore the LVF format is preferred.

## 3.4. Physical Information Extraction

The previous section explained how software tools obtain timing closure using a standardised format. Software tools that specifically target digital implementation require additional physical information to correctly place standard cells, and create interconnect between them. Furthermore, the metal routing which exists inside a single cell should be avoided when creating interconnect. In principle, software tools could work with the full layout of a cell, however this layout contains much more detail than needed, possibly hampering performance of software tools in the context of VLSI. As a result, a minimal set of information is once extracted, most importantly defining cell boundaries, metal polygons and pin locations. This required physical



(a) Full layout of the positive edge-trigger D-type flip-flop.

(b) Abstract view of the positive edge-trigger D-type flip-flop.

Figure 3.14: Comparison between full layout and abstract view of the same cell.

information is captured in the Library Exchange Format (LEF) [58], for each standard cell in a library.
The result of an extraction is an 'abstract' view, figure 3.14 highlights the difference between a full layout and an abstract layout. Clearly, all information regarding the transistors (including polysilicon gates) and implant layers is discarded, leaving selectively metal. Metal, or interconnect, receives additional annotation in the abstract view; every metal polygon is per definition labelled as routing blockage (obstruction), furthermore pin and power rails receive additional labelling. Information which is not present in the graphical representation (figure 3.14b) includes the cell bounding box, allowed cell orientation, transistor drain/source diffusion area and local metal densities. The cell bounding box and orientation determines how cells abut when they are placed in rows. Transistor drain/source diffusion area can be used by software tools to perform early antenna violation checks, as they are typically fixed by changing routing ordering. Finally, metal density look-up tables per cell allow software tools to quickly estimate the global metal density over large areas. Note that antenna information and metal density information are optional, but most definitely useful.
Generation of abstract views is automated, similar to characterisation of standard cells. Cadence provides a clear abstract generation tool, 'Virtuoso Abstract Generator', which requires one time configuration for an entire library. Initially standard cells are binned, i.e. core and I/O cells, subsequently each bin allows for a different extraction approach since they are different in structure. For this particular project only cells in the core bin exist, simplifying the process. The tool loads technology information, i.e. layer names and mappings, and provides an initial configuration to generate simple abstract views. Additional configuration is needed for the cell boundary (box), as the default cell boundary is determined by the dimensions of common drawing layers in the layout. As the proposed standard cells are built to intentionally overlap, the cell boundary should be such that digital implementation tool performs this perfectly. A clear identification point for the cell boundary in this context are the two bias voltage rails implemented in metal two. Hence, the tool is instructed to use metal two to determine the boundary, and a static correction is applied to move the cell boundary on the vertical centre of the metal two track. Similarly, the layers for diffusion area recognition (antenna) require manual input.
If generation of abstract views is successful the information can be exported in the LEF standard. Apart from the aforementioned cell information, the LEF file can be exported with process information such as metal pitch and direction. This process information is mandatory as it provides digital implementation tools the fundamentals required for routing. Furthermore, for each bin available in the library a site is defined in the LEF file, which contains cell dimensions and allowed orientations. This information is for instance required in the floorplanning step of a typical digital implementation flow. For 'CooLib', two core sites are defined: a site which defines the dimension of typical cells, $2.1 \times 0.14$, and a site which is twice this height, $4.2 \times 0.14$. The latter is required for both level-shifting cells.

## 3.5. Standard Cell Verification

As discussed in previous sections, standard cells undergo a process where physical and timing related information is extracted as required by digital implementation software. Naturally, the extracted information should be verified. For instance, errors in the timing information might give false timing closure, while errors in the physical information might lead to large numbers of design rule violations and connectivity issues.
Verification the physical LEF information is by far the most time-consuming, compared to the others. The

software tools which generates abstract views has a simple verification step which connects to digital implementation software and performs simple placements. Unfortunately the software tools supported for this step are old, and unusable during this project due to licensing issues. As a result, LEF verification was performed during actual chip implementation (chapter 4), which at least covers more than just simple scenarios. Parts of the chip were implemented with the 'CooLib' LEF in Cadence Innovus, subsequently the layout was exported to the analog environment and checked with DRC. Initial numbers of design rule violations by incorrect LEF (abstract) specification were excessive, due to poor abutment of cells and their implant layers. Consequently, cell layout issues were identified and resolved, followed by a new iteration of physical information extraction. This time-consuming process was repeated several times to get initial layouts DRC clean. Unfortunately, as implementation continued and the complexity of these implementations continued to increase, new issues were identified in rare scenarios, requiring additional iterations. To conclude, even with very careful layout planning, and abstract view creation, mistakes will pass unnoticed.
Verification of liberty outputs, i.e. cell timing, cannot be done in software as it is based on SPICE simulations. The SPICE simulations are based on foundry-supplied models, which are assumed to be verified. True verification of cell timing information extracted by simulations would thus require measurement of the cell produced in silicon. For this purpose, the cells will be placed on a different chip aimed at characterising individual structures. It should be noted that liberty files can be qualified for consistency, accuracy and completeness by the characterisation software (Synopsys SiliconSmart). Consistency checks look for relative errors between timing measurements resulting from different source models, which effectively acts as a software verification of simulations. Since the liberty format is quite complex, a different software tool by Synopsys, 'Library Checker', checks the syntax and semantics of all the attributes present in the liberty file.
For actual cell layout verification, in the context of cell function, the characterisation tool offers a function based recognition. In this flow, the tool performs quick SPICE simulations to determine the function of the netlist (layout). Since the characterisation tool applies stimuli appropriate to cell function, and thus expects certain responses, the cell layout is verified if characterisation succeeds.

# 4

# Test Chip Design and Implementation

## 4.1. Design & Architecture

To quantify the performance of the standard cell library, it is proposed to implemented commonly found combinational circuits using 'CooLib-Static' and 'CooLib-Domino' logic cells. For these circuits the critical metrics to measure are power and performance. To a lesser extent, energy is interesting because of the minimum-energy point discussion in chapter 2. Finally, it is desirable to test the combinational blocks for faults, for instance stuck-at faults.

For delay testing of the combinational blocks, at-speed testing is required. The simplest solution would be to expose the in- and outputs of some chip to the outside world, such that control and measurement can be performed with high-performance instruments. However, as this particular chip is to be tested in a dewar of liquid-helium this solution is not practical due to the amount of required cables. Furthermore, the additional delay imposed by these cables would obscure the delay measurements. Therefore, on-chip delay testing is desired.

Finally, a comparison between typical standard cells supplied by the foundry is appropriate. For instance, CooLib standard cells might only outperform typical standard cells if the set of typical cells is restricted to the ones implemented in chapter 3. An unrestricted (foundry supplied) implementation may outperform the CooLib simply because there is a larger array of cells and functions available. The result of this comparison is particularly important for digital chips to be implemented in the future, but it might also indicate the need for continued development of the CooLib.

To get accurate current measurement readings for power and energy calculations, each test circuit should ideally have its own supply pad. To elaborate, each test circuit contributes a certain leakage current when it is not switching, which is relative to the size of the circuit. For one, it is expected that the leakage current of CooLib test circuit might be significant at room temperature, which might imply that the total leakage current dominates the switching current of the circuit to be measured. However, too much area is required to fit all of the pads, furthermore it would complicate the measurement setup. A common solution to eliminate leakage current is to completely disconnect some module from the supply, also known as power gating. Surely these power gates inherit leakage current themselves, but they are typically implemented with high-threshold voltage devices to keep this at a minimum. Power gating per circuit to be tested is therefore desired for this chip.

### 4.1.1. On-Chip Delay Testing

When it comes to measuring delays (on-chip), a typical circuit which is employed is a time-to-digital (TDC) converter. A TDC measures the time between a start and stop signal, and is typically able to perform this with a high resolution (picoseconds). However, if a large combinational block is considered with many inputs, this would require either many TDCs or multiplexing of some sort. More importantly, design of a TDC in the context of 4.2 K operation is a problem of its own, due to the lack of simulation models.

A solution which deals with the problem of having many in- and outputs is proposed in [59]. This proposed Built-In Self-Test (BIST) method determines circuit delay by timing failure. To summarise, a clock edge is used to 'launch' some input pattern into the combinational block, while a subsequent clock edge is used to 'capture' the output pattern of the block. The time between the two clock edges is reduced until the wrong

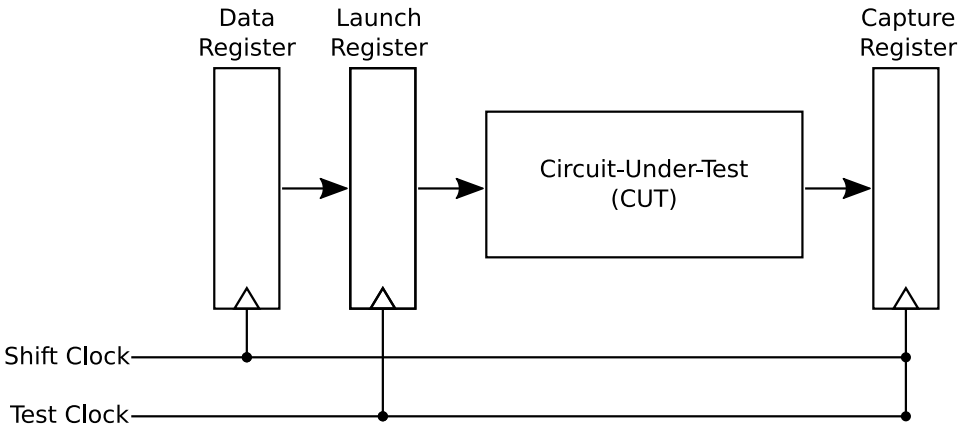



Figure 4.1: Overview of the BIST circuit.

output is produced, at this point the propagation delay of the combinational block is larger than the time between the two clock edges. The adapted BIST for this particular chip is displayed in figure 4.1.
The data register is a serial-in parallel-out register intended to receive input patterns via a simple serial interface. The launch register sits at the output of the data register, such that the input pattern is properly launched on the test clock edge. Finally, the output pattern is captured on the test clock edge in the capture register. The capture register is a parallel-in serial-out register, as it is also capable of shifting the captured data out via the simple serial interface. It should be noted that the measured delay will include timing components other that the propagation delay. Specifically, the clock-to-Q delay $\tau_{clk-q}$ of the launch register and the clock skew $\tau_{skew}$ between the launch and capture register, which gives [59]

$$\tau_{delay} = \tau_{cut} + \tau_{clk-q} - \tau_{skew} \tag{4.1}$$

where $\tau_{cut}$ represents the propagation delay of the circuit under test. Additionally, clock jitter might affect the effective period of the clock signal. This can be accounted for by repeated testing at the same clock frequency, which can be used to build a histogram of test outputs. The test clock can in principle be generated on-chip, possibly with ring-oscillators or a configurable phase-locked loop. However, doing so imposes similar issues as to implementing a TDC at 4.2 K. Since low-voltage, and thus low(er)-performance is of interest, fast clock frequencies (> 1 GHz) can be disregarded. Clock speeds which are reasonable for low-voltage operation, up-to several hundreds of MHz, can easily be brought in from outside of the chip in a single-ended fashion contrary to high-speed clock signals that require LVDS. Therefore, this particular BIST method relies on external test clock generation.

### 4.1.2. Test Circuits

As previously mentioned, different circuits should be implemented in different standard cell libraries for appropriate comparisons. A total of four versions can be considered for this chip, CooLib-Static, CooLib-Domino, TSMC40LP-Restricted and TSMC40LP-Unrestricted. The difference in the restricted and unrestricted version of TSMC40LP lies in the set of cells which is allowed for use during synthesis. Specifically, the restricted set contains the same cell functions and drive strengths as the ones found in CooLib-Static.
Due to the lack of a (N)OR gate available in CooLib-Static, the commonly used ISCAS-85 hardware benchmark cannot be used due to the format of the netlists. A benchmark suite which was found to be perfectly suitable is the EPFL combinational benchmark suite [60], as the netlists provided are implemented with just inverters and AND-2 gates. Three benchmarks were taken from this benchmark suite, with differences in size and number of in- and outputs. Additionally, a generic unsigned multiplier was chosen as test circuit, as the multiplier from the EPFL benchmark suite was found to be too big in size.
Initial choice for a majority voter with 1001 inputs was found to be very difficult to implement with respect to clock skew between the shift register elements (because of the low clock strength buffers available). Therefore this benchmark was swapped for a round-robin arbiter with a more 'gentle' 256 inputs. Table 4.1 summarises the test circuits chosen. In total, this results to sixteen test circuits. To provide a 'back-up' in case the CooLib implementation of the BIST fails, the CooLib-Static flavours of the test circuits are copied and implemented with TSMC's cells. Finally, a copy of the CooLib-Domino int-to-float converter is implemented using cells

Table 4.1: Test circuits (functions) implemented.

| Name | Inputs | Outputs | AND-nodes |
|---|---|---|---|
| Int-to-float converter | 10 | 7 | 260 |
| Sine | 24 | 25 | 1458 |
| Unsigned Multiplier | 16 | 16 | 2700 |
| Round-Robin Arbiter | 256 | 129 | 11839 |

without weak-keeper to observe the 'true' domino operation at 4.2 K.

## 4.2. Implementation

The implementation flow used for this chip follows the typical digital implementation depicted in figure 4.2a. Software tools from the Cadence Digital Design suite are chosen for their superior PPA (performance/power/area) [61], integration between different tools of the suite, and prior experience with the tools.

### 4.2.1. Static Logic Synthesis

Cadence Genus is used for the synthesis of CooLib-Static and TSMC40LP-(Un)restricted. Genus is the only commercial synthesis tool which was found to be compatible with CooLib-Static as the minimal set of cells required by Genus consists of inverters and (N)AND gates. Synthesis tools such as Synopsys DC also require the presence of an OR gate, which is not available for reasons stated in chapter 3. A feature offered by Genus which is particularly useful for the synthesis of different standard cell types is the ability to create library domains. A library domain is linked to a Liberty database, and the library domain can be specified per module (block) as displayed in listing 4.1. This allows the use of the same HDL description for all four flavours of the module.

```
1 create_library_domain {CoolLib_Static CoolLib_Dynamic TSMC40LP TSMC40LP_Full}
2 ...
3 set_db [vfind /libraries -library_domain CoolLib_Static] .library \
4     ${SILICONSMART_MODEL_DIR}/synopsys/CoolLibrary_TSMC40_ecsm_lvf_${CoolLib_CORNER
      ↪ }.lib
5 ...
6 set_db module:$DESIGN_TOP/epfl_sin/ .library_domain CoolLib_Static
```

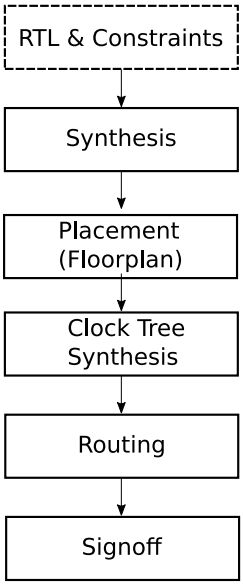
Listing 4.1: Library domains in Cadence Genus.

Another non-standard synthesis flag which was used for this chip is the preservation of the design hierarchy during synthesis. This particular option is very useful for the type of implementation used in section 4.2.3. Domino logic is not synthesised in Genus, which is explained in 4.2.2, but the domino netlists are still read by Genus for merging purposes. It is therefore important to set preservation (don't touch) flags on the domino modules as well. Finally, clocks are specified as per SDC (Synopsys Design Constraint) standard. Three clock domains exist consisting of a slow clock for serial in- and output, a non-fixed test clock and a non-fixed domino logic clock. Even though the test- and domino-clocks are not fixed to a single frequency, they are still specified as it is required for clock tree synthesis. Additionally, timing paths between the three clock domains are specified as false paths, as they are asynchronous clock-domain crossings (CDC). This is not an issue, as no inter clock-domain communication exists.

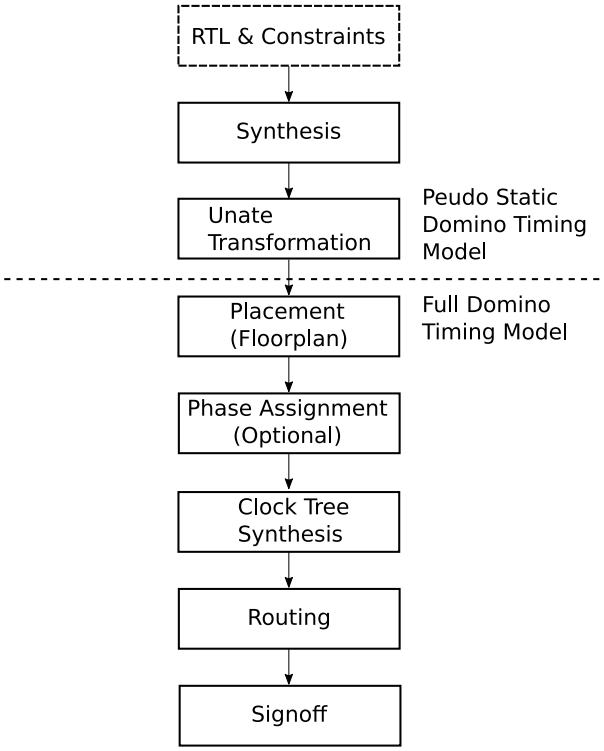
### 4.2.2. Domino Logic Synthesis

In section 2.3.2 it was briefly stated that digital implementation with domino logic is considered more demanding. The non-inverting nature of domino logic dictates that it will never fail because of the main issue found in conventional inverting dynamic logic, unwanted leakage of dynamic node, as domino gate outputs are restricted to only transition from zero to one (positive unate). If an inverter was to placed in between domino logic gates, this restriction would be violated.

Problems in implementation, specifically synthesis, arise from the fact that commercial tools do not support domino logic. For instance, Cadence Genus and Synopsys DC will report that the basic set of cells (inverter and (N)AND/(N)OR) is not present when targeting a pure domino logic standard cell library, even though they are available. To still use commercial tools, alternative design flows are presented which rely on a so-called




(a) Typical digital implementation flow based on static logic.




(b) The implementation flow used for the domino logic.

Figure 4.2

pseudo static domino timing model [37, 62]. This pseudo static domino timing model is simply put the full domino timing model (Liberty), with clock inputs manually removed to selectively leave the boolean function of the cell (full details are discussed in [37]. The result of using this pseudo static domino timing model is that the commercial (synthesis) tool thinks it is doing static synthesis, mapping an RTL description into domino logic gates. However, the resulting netlist will likely contain many inverters, which violates the restriction previously stated. Therefore, it is required to transform the netlist with inverters (negative unate) into a netlist without inverters in between domino logic gates (positive unate). This logic transformation is performed with an algorithm named bubble pushing, where De Morgan's laws are recursively applied until all inverters are either in the front or back of the domino logic gates (block), preserving the (positive) unate requirement of domino logic. Commercial synthesis tools do not implement this logic transformation, therefore it was implemented manually. Figure 4.2 highlights the difference between the typical static digital implementation flow and the altered domino flow based on the flow presented in [37]. Note that phase assignment is out of scope for this context, as it is a (pipelining) technique, involving multiple clock domains, to increase domino logic performance even further.

### Unate Transformation

The bubble pushing algorithm is implemented in C++, in about 3000 lines of code, and only supports the domino logic gates present in the CooLib. The input to the algorithm should be a mapped (Verilog) netlist, which contains domino logic gates and inverters. For this project, Synopsys DC was used in conjuction with the pseudo static timing model to achieve this.
The C++ implementation parses the mapped netlist using simple regular expressions, which is not the fastest solution, but deemed sufficient for this project. For each gate a data structure (vertex) is populated which contains general gate information, i.e. function. Afterwards, all of the vertices are connected by edges to create a directed acyclic graph which corresponds to the netlist. To start the recursive bubble pushing algorithm, this particular implementation starts at the output vertexes meaning it will push the inverters to the primary inputs of the netlist.

Table 4.2: The amount of vertices (gates) and edges (interconnect) before and after unate transformation.

| Module | Edges Original | Vertices Original | Edges Unate | Vertices Unate | Edge Overhead | Vertex Overhead |
|---|---|---|---|---|---|---|
| Int2float | 245 | 270 | 214 | 250 | -12.7% | -7.4% |
| 16x16 Mult | 2334 | 2430 | 3611 | 3739 | +54.7% | +53.9% |
| Sin | 6121 | 6195 | 9799 | 9901 | +60.1 % | +59.8 % |
| Majority Voter | 10483 | 11486 | 15886 | 17890 | +51.5% | +55.8% |

While traversing the DAG it might happen that an inverter cannot be removed by applying De Morgan's law, as both the inverting and non-inverting output of the prior gate are needed simultaneously. In this scenario the inverter is considered trapped. Trapped inverters are removed by logic duplication, where the entire fan-in logic cone is duplicated to provide both inverting and non-inverting outputs at the cost of area. Under special timing constraints trapped inverters can be allowed [63], however this would require timing annotation of the parsed netlist. To keep things manageable, and under the uncertainty of timing at the target operating temperature, selectively logic duplication is applied. Figure 4.3 illustrates the trapped inverter and logic duplication principle.

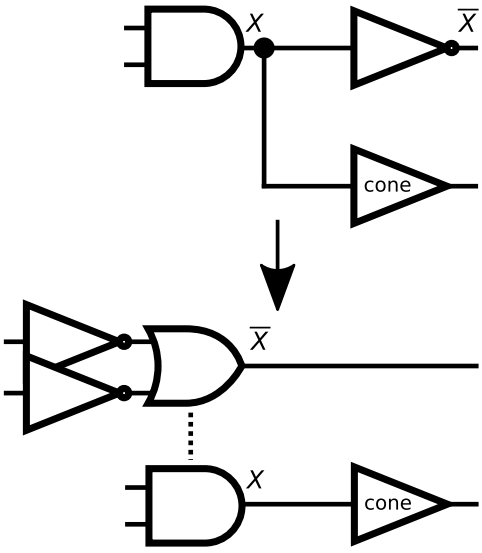



Figure 4.3: Logic duplication to remove a trapped inverter at the AND gate's output.

A naive implementation of logic duplication leads to excessive area overhead and an exponential increase in run-time due to the recursive nature of the algorithm. However, the maximum area overhead induced by cone duplication should be 100%, as only the inverted and non-inverted output of a gate are ever needed. Hence, redundant duplication of a gate can be avoided if a gate is associated (linked) with its duplicated gate. Two scenarios can be encountered as the linked gate can be inverting or non-inverting with respect to the original gate. Figure 4.4a shows the first scenario and how the trapped inverter is removed by logic rearrangement. In a similar fashion, while duplicating a logic cone it might be beneficial to check the associated gates even when no (trapped) inverter is encountered. This is illustrated in figure 4.4b where insertion of a temporary dummy inverter, which is trapped by definition, allows for removal of logic gates by rearrangement of subsequent logic. These transformations inherit a significant reduction in logic duplication.

The result of the unate transformation is displayed in table 4.2. As expected, most circuits (combinational blocks) suffer from an increase in area due to logic duplication. However, a curious case is found for the Int2float circuit where a reduction is observed. Verification of the algorithm is achieved by an exhaustive input pattern sweep, where the outputs of the unate circuit are compared to the original static circuit. For circuits with a large number of input bits, exhaustive sweeps are infeasible and replaced by applying a large number of randomised input stimuli.

### 4.2.3. Hierarchical, Partition Based Implementation with Multiple Library Domains

The layout part, i.e. floorplanning and routing, of the typical digital implementation flow is performed with Cadence Innovus. For simple chips, the entire mapped netlist might be implemented at once. The existence of library domains renders this impossible, as specific combinations of library domains are incompatible with each other, meaning they cannot be placed in the same row. The only way of dealing with this issue in Innovus is by partitioning. Partitioning allows modules to be implemented individually, from placement all the way to GDS export. If all non-compatible modules are implemented in this manner, they can be assembled into the top-level when they are finished. This is also implies that the top-level is implemented separately. This hierarchical partition based approach comes with additional complexity in implementation. For instance, timing budgets for partitions need to be derived to allow individual implementation of the top- and block-level(s). Furthermore, a pin assignment per partition is required to ensure correct routing when the top-level

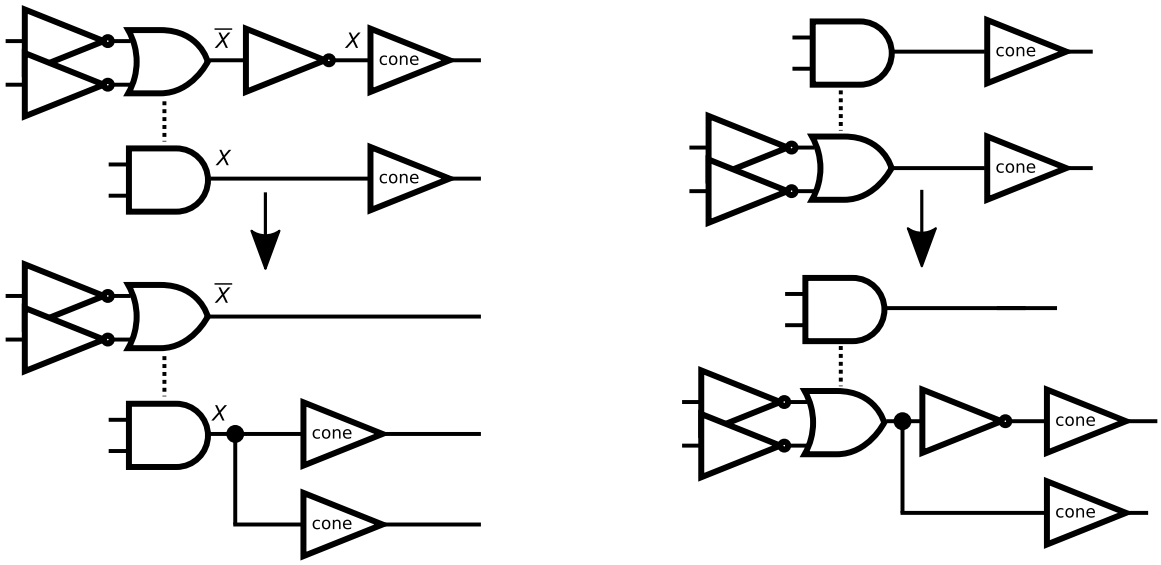
and the partitions are assembled. Figure 4.5a shows the layout (floorplan) after partitioning with some power routing in place. For each module and its respective library domains a partition is created.
For each type of library domain a script is used to implement the block. Each version of this script is similar in terms of flow, but contains subtle differences. For instance, the domino implementation script has different clock tree synthesis parameters than static implementation. Some blocks require nested partitioning as they mix library domains on the module hierarchy level. For example, the TSMC40LP combinational blocks contain CooLib-Static flip-flops for the BIST implementation, as they are expected to achieve lower $V_{DD}$. Similarly redundant copies of the combinational CooLib-Static blocks with TSMC40LP BIST implementation are provided as back-up solution. These blocks require a second-level of partitioning, which is automated by two additional flavours of scripts. The layout of such a nested block is shown in figure 4.5b, where the transition between library domains is indicated by the power rings of the inner block.

### 4.2.4. Power Domains

Each of the partitions previously discussed is defined as power domain, as they can be powered-down by power gates. The layout of the power gate is found in appendix A, and is a header type power gate that can be placed in both ring and column style. As a side note, the power gates are placed in the top-level partition and the connection to the partition is made after assembly of the partitions. The supplies and secondary supplies are equal for all power domains, the only difference lies in the power domain shut-off condition and back-bias supplies. Specification of the power domains is done in the Common Power Format (CPF), which allows specification of rules such as operating voltage. The low-power specific cells should be defined in this format, as Innovus is able to automatically infer where these cells should be placed. For instance, listing 4.2 shows the CPF 2.0 specification for both level-shifter cells.

```
1 define_level_shifter_cell -cells { LEVELDOWN }  \
2 -input_voltage_range {0.8:1.1} -output_voltage_range {0.1:1.0} \
3 -direction down -output_power_pin VDD -input_power_pin VDDH -ground VSS \
4 -valid_location to
5
6 define_level_shifter_cell -cells { LEVELUP }  \
7 -input_voltage_range {0.1:1.0} -output_voltage_range {0.8:1.1} \
8 -direction up -output_power_pin VDDH -input_power_pin VDD -ground VSS \
9 -valid_location from
```

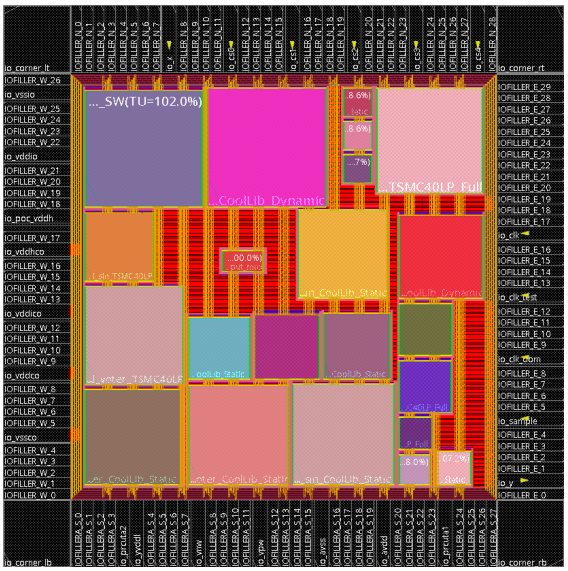
Listing 4.2: Level-shifter specification in the CPF 2.0 standard.



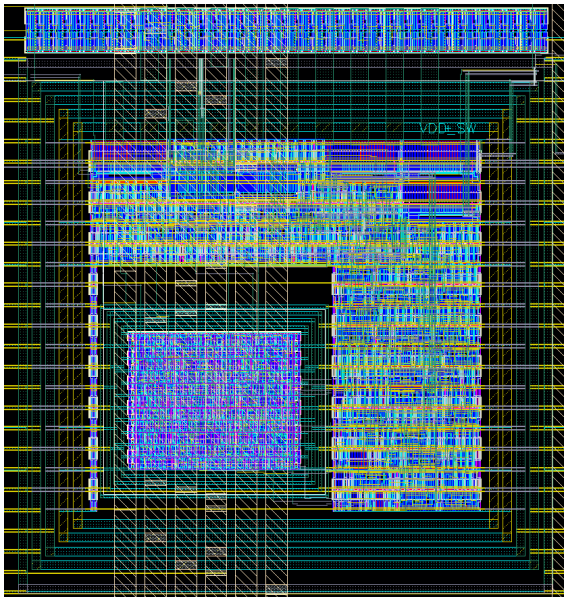

(a) Merging of a trapped inverter if the associated (linked) duplicated gate is inverting.

(b) Dummy inverter insertion allows removal of duplicated logic cone.

Figure 4.4: Logic rearrangements to avoid unnecessary logic duplication.




(a) Top-level layout with partitions and early power routing in place.




(b) Nested partition together with power gates in place.

Figure 4.5

Since hierarchical implementation is used for the partitions, a similar hierarchical approach in applying CPF is adopted. A single block-level CPF specification is created of a module which is repeatedly used for all 21 partitions. To allow for small options in this specification, parameters can be supplied. One such parameter is the power domain shut-off condition. Figure 4.6 shows a graphical representation of the power domain configuration which is nearly common to all partitions.
As shown, each power domain only contains a single output. However since the power domain can be powered-down, it is required to insert isolation at this single output as its value will be undefined in case of shut-down. For this purpose, the NAND gate available in CooLib-Static is sufficient as no state retention is needed, although care should be taken as the output is inverted by this gate.

## 4.3. Verification

Verification of the layout after place and route is performed by exchanging the GDS between the digital and analog Cadence environment, and subsequently running layout-versus-schematic (LVS) using Mentor Graphics Calibre. Additionally, design rule checking (DRC) is done in a similar manner.

### 4.3.1. Functionality & Timing

Functionality is verified at various stages during the implementation flow. Undoubtedly, functional verification is important during and after writing the HDL such that correct operation without timing information is ensured. Simulation during this step also provides a 'golden' reference for simulations later in the implementation flow. Simulations are also performed post-synthesis and post-implementation (or routing). Post-implementation simulation can be considered the most critical, as this simulation should replicate the yet to be fabricated chip as close as possible in terms of timing. Timing at the target operating conditions of 4.2 K is unknown due to the lack of models, therefore sign-off timing analysis and verification is not performed. To emulate a potentially more extreme environment at 4.2 K, additional positive slack is specified for setup and hold constraints, on top of the conventional best- and worst-corner analysis. Functional and post-synthesis simulations are performed with Mentor Graphics QuestaSim, while post-implementation simulations are performed with Cadence Incisive. Incisive offers support for simulations with CPF annotation (verification), and is therefore a more suitable way of verifying the power domain semantics (i.e. shutdown).

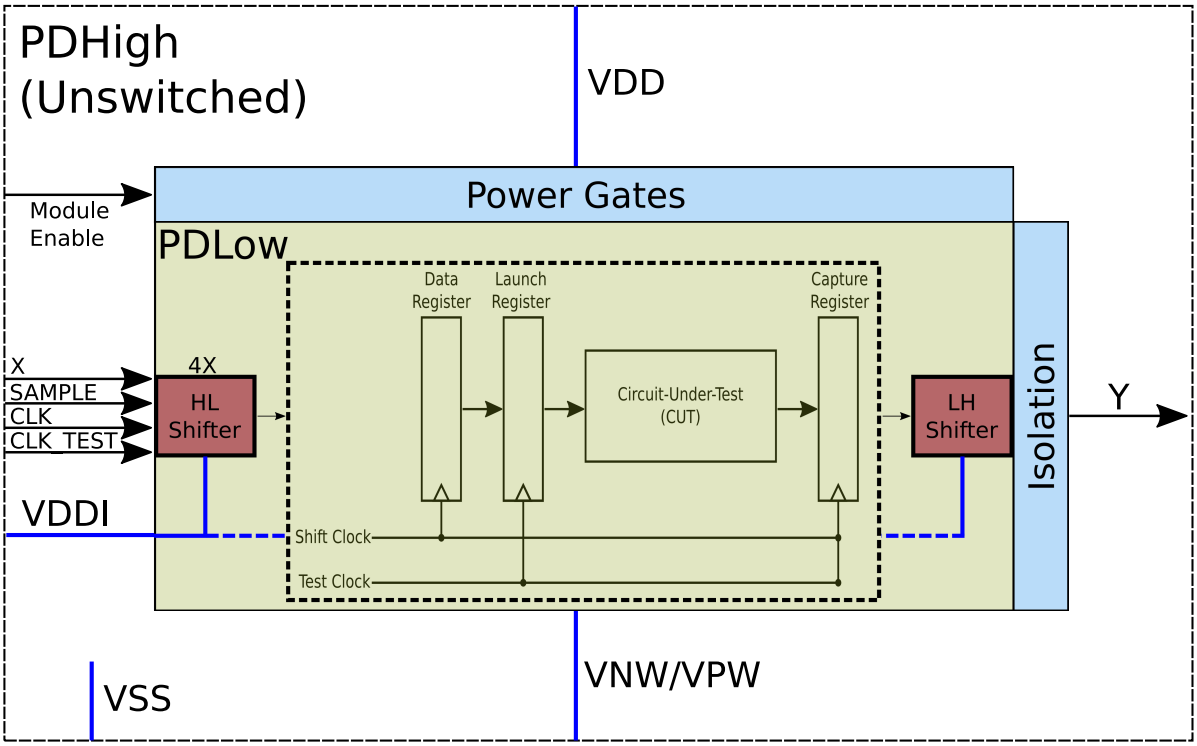



Figure 4.6: Power domain 'PDLow' graphical representation.

### 4.3.2. Power Grid

Verification of the power grid in terms of IR drop is performed using Cadence Voltus. Voltus uses a power-grid library which contains electrical characteristics per cell.

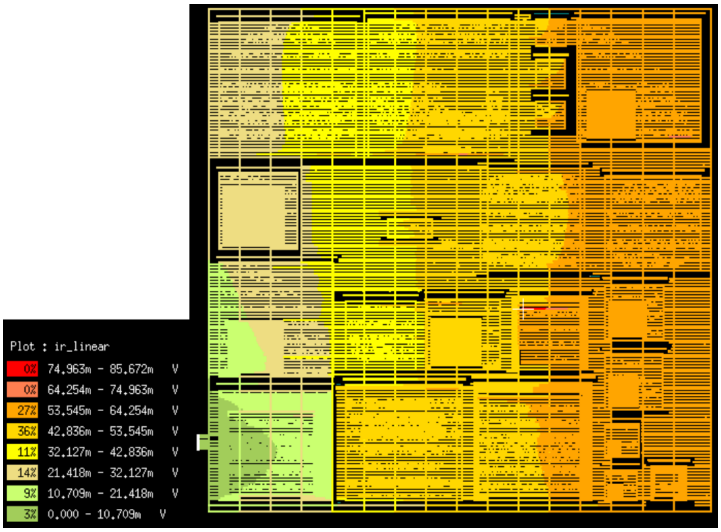



Figure 4.7

For instance, it contains the capacitance as seen from $V_{DD}$ to the cell, which in turn can be used during IR drop analysis by using the fact that cells which are off act as decap. Generation of the power-grid library is a one-time effort, Voltus performs SPICE simulations on the cell netlists (with parasitics) to create a database. Additionally, the power-gate available in the library receives additional characterisation on important parameters such as $I_{ON}$ and $R_{DS}$, which might be used for optimisation.

The total capacitance offered by the power grid (approximately 600 pF) was found to be sufficiently large to not use decap. No noticeable IR drop was found on the supply nets, however some ground bounce was observed over the ground net as shown in figure 4.7. Explanation lies in the fact that the ground net is common to all supplies, experiencing larger currents than the separated supply nets. To deal with the issue, ground stripes were doubled in width and an extra core ground I/O pad was added on the top-right side of the chip.

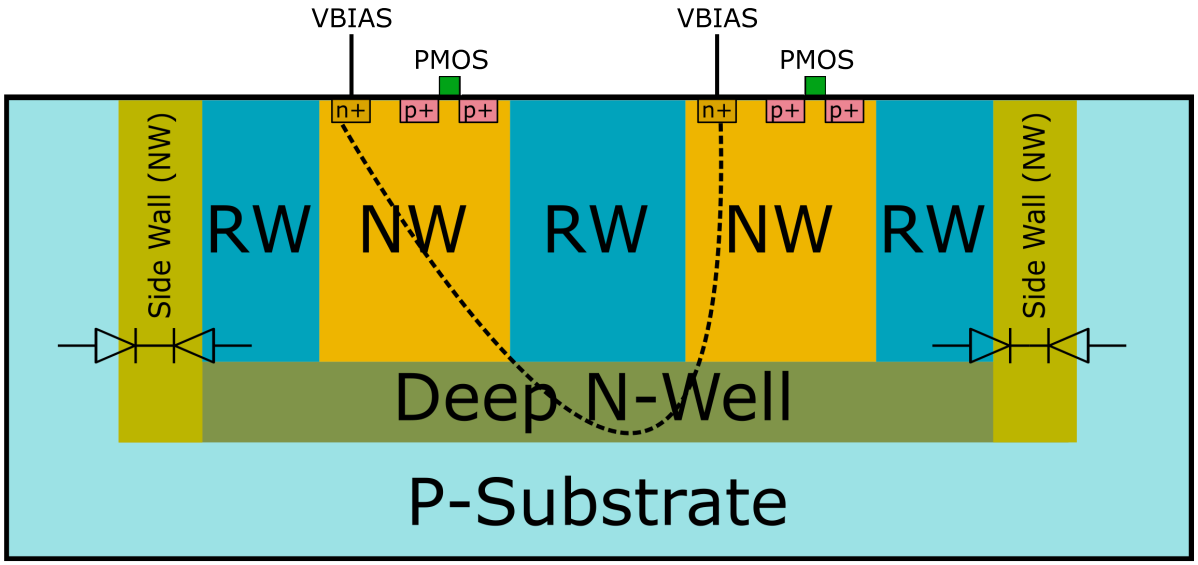



Figure 4.8: Isolation provided by DNW and short-circuit (dotted line) due to dynamic back-biasing when placed in a common DNW. RW stands for p-well in DNW.

## 4.4. Finalised Layout and Implementation Issues

The CooLib and TSMC40LP modules have incompatible back-bias supplies, i.e. if no care is taken significant current may start to flow through the substrate. TSMC40LP modules contain welltaps connected to ground (NMOS) and $V_{DD}$ (PMOS), while CooLib modules contain welltaps connected to VPW (NMOS) and VNW (PMOS). The solution is to encapsulate CooLib modules with a deep n-well, as this creates a back-to-back diode between neighbouring p-wells, this is shown in figure 4.8. Note that encapsulating TSMC40LP cells might work as well, but requires an additional deep n-well across the I/O cells. Additionally, the power gate cell contains biasing conditions similar to TSMC40LP (nominal), to make it predictable. These cells should thus be excluded from the deep n-well, or added into an additional deep n-well.

The deep n-wells cannot be properly added in the digital implementation software (Innovus). Consequently the layout should be exported to the analog environment (Virtuoso) to add the deep n-wells as a finalisation step, together with proper DRC and LVS checking. Two issues were encountered while adding the deep n-wells. Primarily, LVS identified a potential short-circuit through the deep n-well, caused by the dynamic back-bias of the HL shifter cells. This is displayed by the dashed line in figure 4.8. Normally, RW runs underneath NW which blocks this connection, however design rules and LVS errors showed that this certainly is not the case. Additionally, due to the freeze out of the substrate the resistance could be so high such that the short circuit would not cause harm. To avoid unnecessary risks, the choice was made to exchange the dynamic back-bias of the HL shifter for a nominal, static connection. The potential effect of this adjustment could be an increase in $V_{DD,min}$.

A second issue with respect to the deep n-wells was found while isolating the power gates. For a total of five modules too little space was present between power gates and surrounding logic as defined by design rules. To avoid complete re-implementation of the chip in the digital environment, a second flavor of the power gates was introduced without welltaps. These power gates are thus subject to back-biasing, making them harder to predict. Thankfully, three out of five scenarios apply to the back-up copies, minimising the potential effect on measurement results. On the contrary, this change allows measuring of the effect of back-biasing on power gates in terms of module performance (at 4.2 K).

As discussed in section 3.5, multiple iterations of cell layout and abstract view were required to get a DRC clean result after placement and routing. It turns out that, even though this worked fine for 'simple' blocks, several DRC errors remain when implementing a full chip. For this particular chip, about 80 DRC errors were found by Calibre. The digital environment was able to identify most of these, but was not able to fix them. Although the errors are easily solved manually, an experience similar to TSMC's standard cells should be obtained, where normally no errors are found after place and routing. This would require careful consideration of rare scenarios when updating the cell layout for a second version. For this tape-out, the errors were solved manually because of time constraints. Figure 4.9 shows the final layout (in the analog environment).

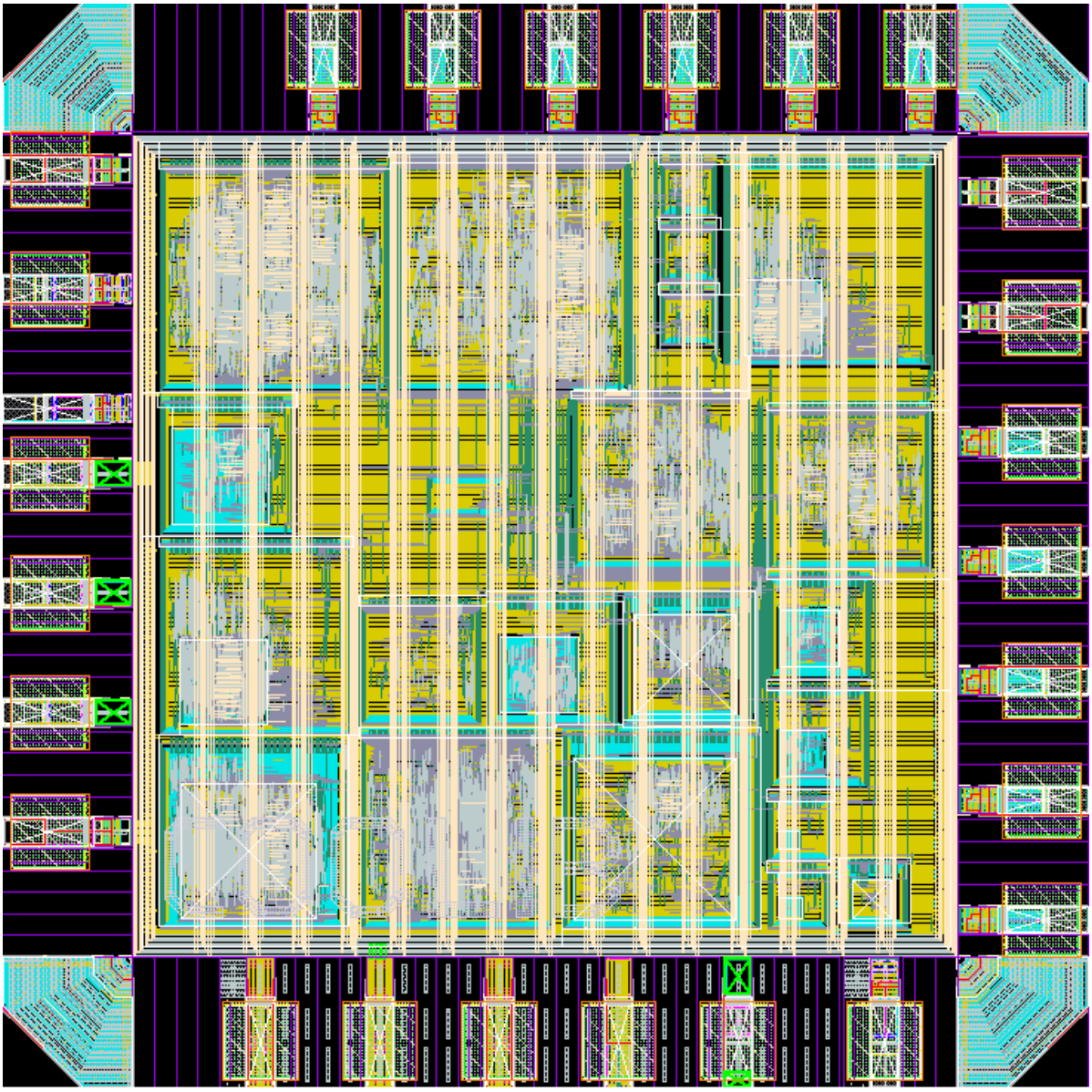


Figure 4.9: Full layout of the implemented test chip.

# 5

# Cryogenic RISC-V

The chip implemented in chapter 4 is aimed at quantization of the CooLib's performance and comparing it to a conventional standard cell library by implementation of several commonly encountered combinational blocks. This chip does not necessarily demonstrate the functionality of CooLib standard cells in more complex circuits. Therefore, a simple CPU is implemented in a second chip, to act as both a proof-of-concept and as the first CPU specifically implemented with 4.2 K operation in mind.

## 5.1. RISC-V

A popular open-source, RISC based instruction-set-architecture is RISC-V [64]. Given its popularity many different implementations are available, however the preferred implementation should focus on area due to constraints given on chip dimension. One such implementation is PicoRV32, which only implements the base instruction set with optional support for hardware multiplication and division, as well as interrupts. Additionally, the PicoRV32 is FPGA and even ASIC proven (albeit in an old process) and comes in an optional system-on-chip configuration with UART and execute-in-place SPI controller.
This PicoRV32 implementation without interrupts was sufficiently small to be implemented in the available core area of $450\mu m \times 450\mu m$, together with a 8 kB SRAM memory ($294\mu m \times 194\mu m$) serving as combined instruction and data memory. The SoC configuration was taken as starting point, hence no actual design was needed apart from logic that would allow SRAM programming (and readout). Furthermore, the XIP SPI controller was completely removed. Only CooLib-Static cells are used for simplicity required by time constraints. CooLib-Domino cells could in principle act as a viable accelerator for data-path elements such as a multiplier or divider.

## 5.2. SRAM Programming

For SRAM programming and readout, a simple structure similar to a boundary scan-chain is adopted. This structure is shown in figure 5.1. In addition to the rudimentary operations, this structure allows testing of the SRAM by i.e. march tests.
The programming and readout is done by the sequences defined in table 5.1. The SRAM itself operates at the nominal voltage of 1.1 V and is interfaced by CooLib level-shifters to communicate with the core running in a low-voltage power domain. Furthermore, the fastest type of SRAM offered by the foundry was chosen, inheriting the highest leakage power, to exploit the lower leakage currents at 4.2 K.
Auxiliary scripts and firmware for programming the RISC-V are provided in appendix B.

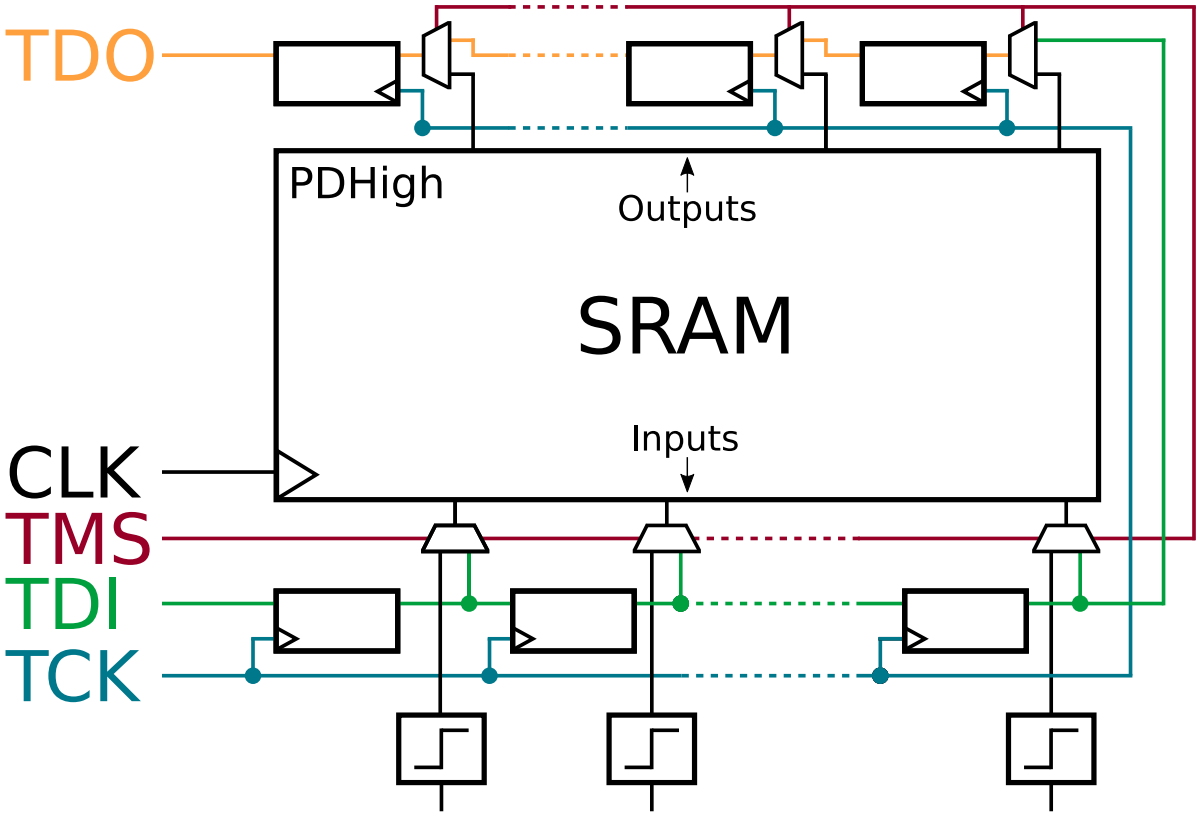



Figure 5.1: Logic surrounding the SRAM for programming and readout.

Table 5.1: SRAM read and write operation sequences.

| | Write Operation | Read Operation |
|---|---|---|
| | Pattern(90)=Data(32)&Addr(22)&WEN(4)&0s(32) | Pattern(90)=0s(32)&Addr(22)&WEN(4)&0s(32) |
| 1. | De-assert ResetN & Assert TMS | De-assert ResetN & Assert TMS |
| 2. | Shift 90*TCK (TDI) | Shift 90*TCK (TDI) |
| 3. | Disable TCK | Disable TCK |
| 4. | Pulse CLK ->SRAM[ADDR]=Data | Pulse CLK ->Q=SRAM[ADDR] |
| 5. | | De-assert TMS & Wait >= 1 Cycle(s) |
| 6. | | Assert TMS |
| 7. | | Shift 90*TCK (TDO) |

## 5.3. Implementation and Verification

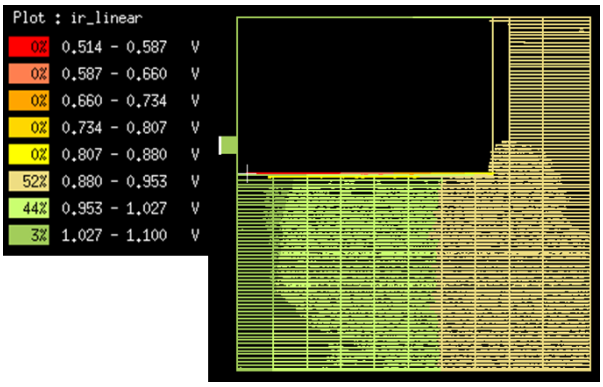



Figure 5.2: IR drop across the RISC-V core.

In general, the RISC-V implementation is much more straightforward than the other chip due to the lack of complex power domains and the use of only a single standard cell library. Still, the chip contains two power domains; the SRAM and programming logic operate in the nominal voltage domain, while the core operates in the low-voltage domain. SRAM layout guidelines were followed to minimise IR drop with respect to the block, which effectively involves routing as much stripes over the block until the maximum metal density is achieved. The same software was used and the typical (digital) ASIC implementation was followed. Functional verification was performed using selectively QuestaSim at the various stages of implementation, by loading several provided test benches (assembly code) into the SRAM. The layout was verified by Calibre LVS (and DRC to some extent).

Similar to the other chip, IR drop analysis was performed which displayed a poor power grid at nominal voltage, which is shown in figure 5.2. Increasing the metal stripe width of the supplies and adding horizontal stripes in addition to selectively vertical stripes was found to solve the problem. Finally, the remaining free space was used to add 80 pF of decap to further mitigate IR drop.

## 5.4. Finalised Layout and Implementation Issues

Issues found in the chip implemented in chapter 4 apply also for this chip, a deep n-well was needed to isolate the core from the SRAM and I/O and a small number of DRC errors remained after implementation in the digital environment, which were fixed manually. Figure 5.3 shows the final layout.

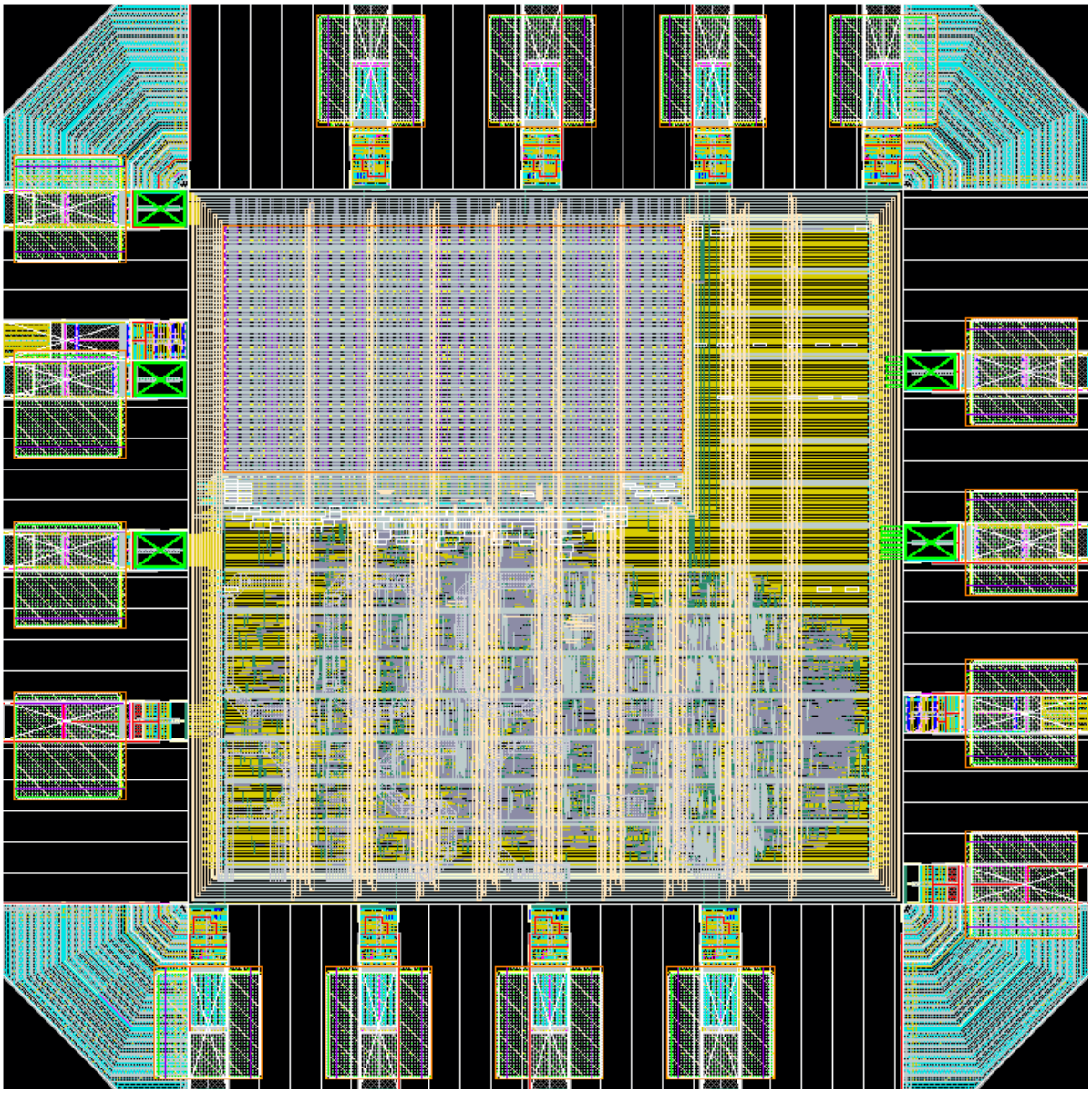


Figure 5.3: Finalised PicoRV32 implementation layout with CooLib-Static (and TSMC SRAM blackbox).

# 6 Testing & Results

## 6.1. Test Setup

Both of the chips require an FPGA, either for loading and capturing of patterns or for programming of the RISC-V chip. The work in [65] has demonstrated that certain Xilinx FPGAs operate without problems at the 4.2 K temperature. Furthermore, they even show improved characteristics such as a slight increase in overall performance and a reduction in clock jitter for clocks generated on-board. Placing an FPGA next to the chips to be tested at 4.2 K in a tank of liquid helium removes the need of long cables which inherit a large capacitance. Since the overall system of FPGA-to-chip is source-synchronous it is also desired to equalise wire lengths for clock and data signals, rendering long cables even less desired.

The particular FPGA that was tested to perform well is the Xilinx Artix-7, and is capable of generating clocks up-to 680 MHz [66]. In addition, the mixed-mode clock managers (MMCM) can be dynamically reconfigured to generate clock frequency sweeps on the FPGA itself. The last feature in conjunction with the other attractive properties of cryo-FPGA makes it the preferred choice. As PCB development with an FPGA is not considered trivial, it would be beneficial to develop a single FPGA motherboard, contrary to two complex PCBs, which contains general connections to a daughterboard. This FPGA motherboard was developed by H. Homulle, and acts as a development board for 4.2 K operation. The board holds two connectors are used to connect to a daughter-board which connect a total of 80 signals of several types (i.e. supplies and differential signals).

The daughter-board consists of a socket, decap for the chip and connectors for connection to the motherboard. The capacitors were chosen based on the work in [67], i.e. tantalum capacitors for large values ($> 4.7\mu F$) and NP0 capacitors for small values. For measurement of power and energy, $V_{DD}$ should be measured in terms of current consumption. As the range of $V_{DD}$ is expected to be large, the required (smallest) resolution for measuring is just as large. A Keithley 2636B source-meter unit was found to be appropriate for these constraints, as it can measure down to femtoampere range. The second channel of the SMU is used to measure the $V_{DDH}$, which is of interest for the RISC-V chip as to measure the power consumption of the SRAM. This particular SMU allows external triggering, meaning that the FPGA can indicate the SMU to do a measurement via a single wire. This feature enables automated measurement for the test chip. The remaining supplies, for both chip and FPGA, are provided with two 'simple' voltage sources. An overview of the test setup is displayed in figure 6.1.

## 6.2. Automated Test Pattern Generation

Combinational ATPG is performed using Cadence Modus, due to its integration with the synthesis tool that was used (Genus). Specifically, Genus generates the set-up files needed for Modus. Table 6.1 shows the fault coverage which is achieved and the space required for the patterns to be stored in FPGA memory. The total space required exceeds the FPGA's block RAM capabilities, which means that not all blocks can be tested in one sweep. Global fault coverage by just static combinational ATPG is close to 100% for most circuits, except for the int-to-float circuits. Since the patterns are primarily used for testing delay, this is not an concern.

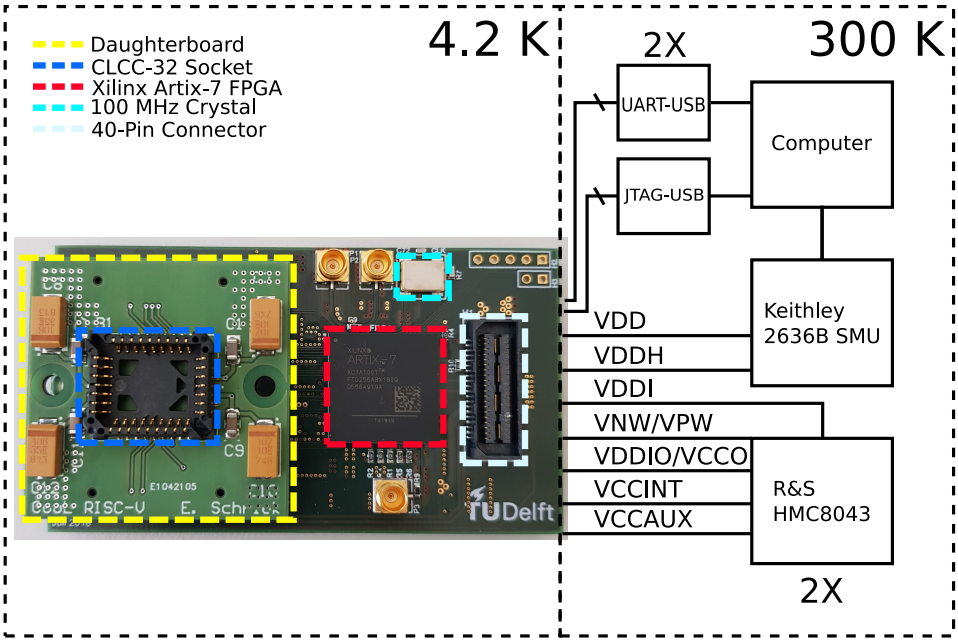



Figure 6.1: Complete test set-up. Supplies labelled with the VCC prefix are used for the FPGA.

## 6.3. Field-Programmable Gate Array Programming

The FPGA for the RISC-V chip is mainly used for SRAM programming and readout. The firmware does not require detailed explanation, as it implements the sequence(s) showed in table 5.1. The FPGA firmware for the test chip is much more involved, however. In detail, the FPGA completely automates the measurement of the chip; it triggers the source-meter unit for current (power) measurements, performs clock frequency sweeps while testing for delay and writes results and debug information via a serial interface. Dynamic power consumed by the test circuits is data dependent, which implies the test circuits should be exposed to randomised stimuli to remove the dependency by averaging. The FPGA implements a 21-bit linear-feedback shift register (LFSR) for providing these stimuli.
Figure 6.2 shows a simplified state diagram, in practice a total of approximately 50 states is used in various distinct state machines and clock domains. Debug information may consist of the captured pattern, clock information and the current state of testing (i.e. the pattern number. Communicating this information requires 38 bytes per clock frequency step, which boils down to 160 megabyte of data for a single voltage, bias and temperature test point. Furthermore, the test time overhead induced by the slow UART communication is excessive (several hours). Therefore, a second firmware is used outside of debugging, which writes just a single byte representing whether the received output is correct.

### 6.3.1. Dynamic Clock Generation

An important aspect of the FPGA is clock generation and reconfiguration for generating clock frequency sweeps. In theory, by just using integer divides of the MMCM, a total of 868,363 combinations can be synthesised [66]. In practice, overlap exists between many combinations, furthermore divider and multiplier values are bound by the input frequency of the MMCM, as the internal VCO has to maintain a certain range of frequencies. As a result, differences between certain neighbouring frequencies are not constant and should be analysed to calculate the error on the final delay measurement caused by this discreteness. Figure 6.3 shows a staircase plot of the synthesizable frequencies under the aforementioned constraints. The maximum error was found to be 1.7%, while the average error is 0.07%, which was found to be acceptable. Moreover, if the range turns out insufficient, two clock dividers can be cascaded to open up even more frequencies down to several kilohertz.

## 6.4. Results

Both chips were found to be functional at room temperature. The RISC-V chip has been tested by running (compiled) C code, with a relatively slow system clock of 100 MHz. The memory-mapped UART and most, if

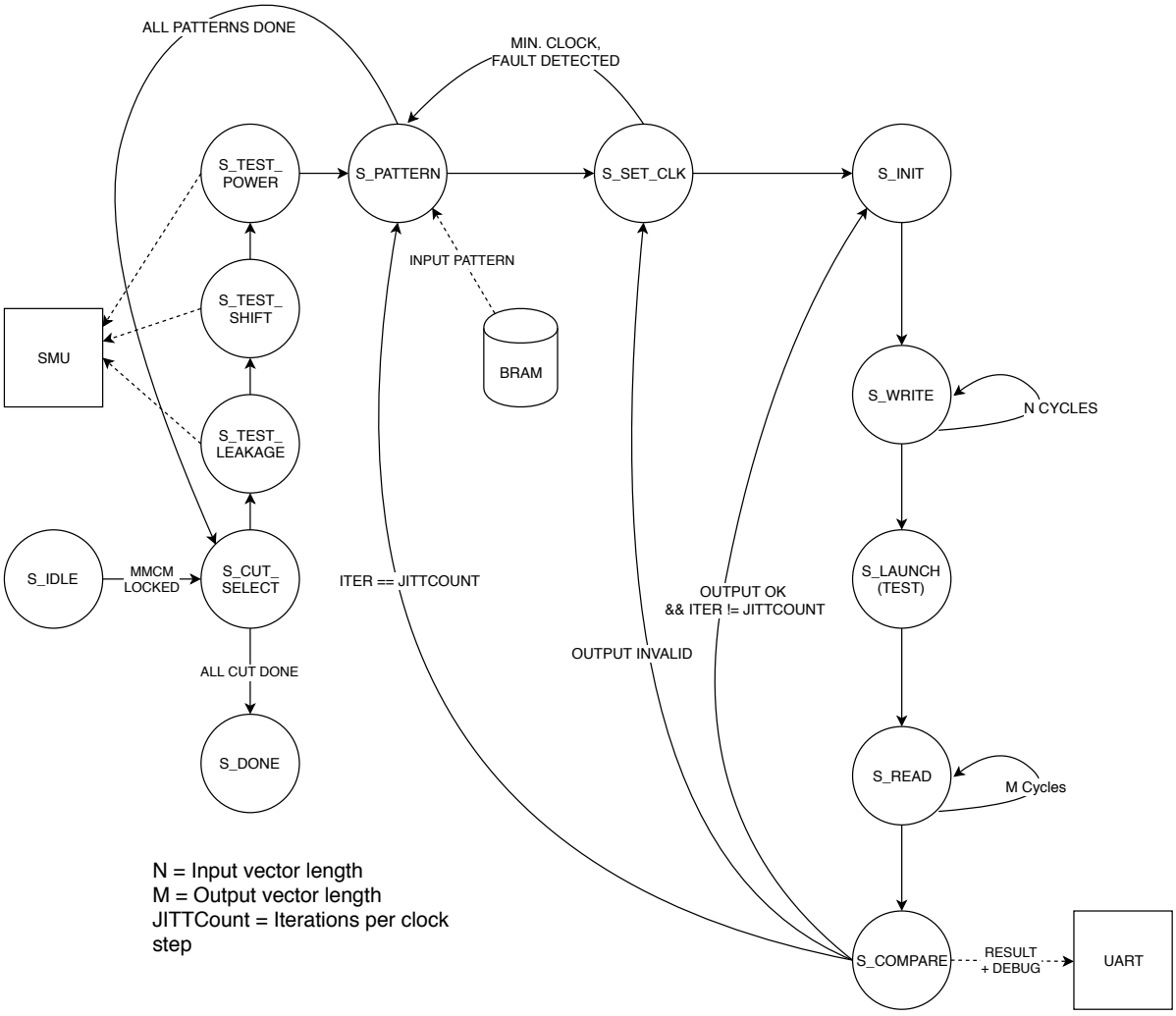



Figure 6.2: Simplified state diagram for the automated testing of the test chip.

Table 6.1: Combinational ATPG results.

| | # Tests | Test-Mode | Global | # Tested | # Untested | Space (non-aligned) |
|---|---|---|---|---|---|---|
| Intfloat Static | 76 | 96.44% | 73.73% | 1788 | 624 | 171 byte(s) |
| Intfloat Domino | 64 | 95.21% | 67.30% | 1311 | 624 | 144 byte(s) |
| Intfloat TSMC40 | 80 | 96.25% | 72.70% | 1696 | 624 | 180 byte(s) |
| Intfloat TSMC40F | 71 | 96.39% | 71.42% | 1574 | 624 | 160 byte(s) |
| 16x16 mult Static | 53 | 99.48% | 95.21% | 12674 | 624 | 424 byte(s) |
| 16x16 mult Domino | 71 | 99.71% | 97.67% | 27342 | 635 | 568 byte(s) |
| 16x16 mult TSMC40 | 52 | 99.53% | 95.20% | 12490 | 624 | 416 byte(s) |
| 16x16 mult TSMC40F | 56 | 99.54% | 95.27% | 12696 | 624 | 448 byte(s) |
| Sine Static | 252 | 96.97% | 95.51% | 36226 | 1688 | 1544 byte(s) |
| Sine Domino | 252 | 97.94% | 97.09% | 64063 | 1850 | 1544 byte(s) |
| Sine TSMC40 | 260 | 96.62% | 95.18% | 36417 | 1808 | 1593 byte(s) |
| Sine TSMC40F | 265 | 96.88% | 95.43% | 36402 | 1707 | 1624 byte(s) |
| RR Arbiter Static | 1633 | 99.80% | 97.02% | 29614 | 624 | 78589 byte(s) |
| RR Arbiter Domino | 2504 | 99.67% | 98.38% | 43615 | 624 | 120505 byte(s) |
| RR Arbiter TSMC40 | 1676 | 99.64% | 97.77% | 29716 | 624 | 80658 byte(s) |
| RR Arbiter TSMC40F | 1723 | 99.61% | 97.76% | 30074 | 624 | 82920 byte(s) |

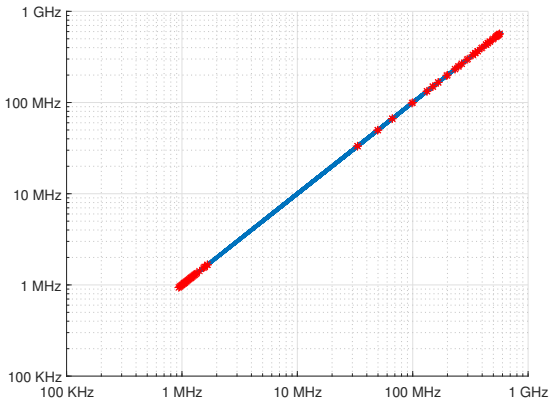



Figure 6.3: The range of synthesizable frequencies on the Xilinx Artix-7, with $F_{CRYSTAL} = 100MHz$. Points where the relative error between two neighbouring frequencies is above 0.5% are indicated by a red dot.

not all, instructions are verified this way. Some measurements of the test chip (chapter 4) have been carried out. Remaining measurements, specifically those at 4.2 K, are yet to be performed. An elaboration of the measurement results of the test chip will now follow.

### 6.4.1. 300 K Delay Measurements

Figure 6.4 shows the initial delay measurement of the test chip down to $V_{DDL} = 1.1V$, with $V_{NW} = V_{DDL}$ and $V_{PW} = 0V$. The BIST as proposed in chapter 4 appears to work as intended, although the test output contains 'bubbles'. To elaborate, results might show correct operation up to some frequency $F_{max}$, while the frequencies below $F_{max}$ might indicate failure. The results in figure 6.4 are filtered, using the highest frequency found up to the point where these bubbles start appearing. This point appears to be well below $F_{max}$ for $V_{DD} \geq 0.7$ for the multiplier and int-to-float circuits. For lower voltages this point appears to be equal to $F_{max}$.

A possible explanation is found in the clock frequency sweep provided by the FPGA, as described in section 6.3. The number of frequencies in this sweep has been heavily reduced, as the testing time and the size of the output data with a full sweep was found to be excessive. A second implication is a lower bound of 10 MHz for these measurements, which renders testing below $V_{DDL} = 0.4$ fruitless.

Domino logic test results were only found to be consistent in the case of the sine circuit, where it shows significant speedup over the other implementations. In detail, the other three circuits displayed substantial

fluctuations in the measured delay over the range $_VDDL$. A full clock frequency sweep is expected to provide more insight. Even though inconsistency was found in the three other cases, correct outputs were captured, verifying the implemented unate transformation and domino logic synthesis flow proposed in section 4.2.2. Finally, the proposed cells appear to outperform, in the region of interest, the foundry supplied standard cells. This is as expected, as the foundry supplied standard cells are implemented with standard threshold voltage devices, contrary to the proposed cells which make use of low threshold devices.

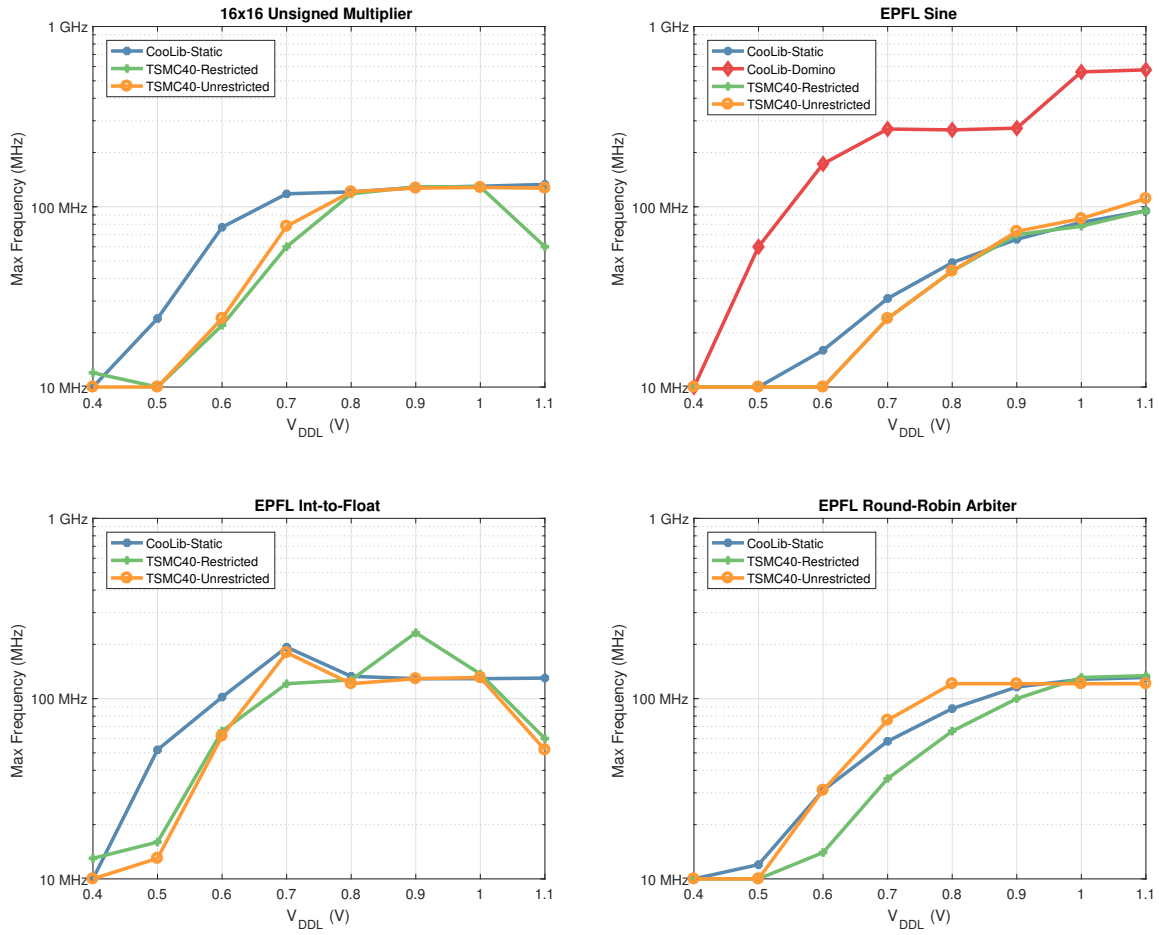



Figure 6.4: 300 K delay measurements for the test chip.

### 6.4.2. 300 K Leakage Power Measurements

Figure 6.5 shows the leakage power measurement results. For these measurements, all three clocks are disabled and the appropriate power domain is powered on, followed by a current measurement of several seconds. It should be noted that the results at $V_{DDL} \leq 0.2V$ are dominated by the leakage current of the capacitors on the PCB, and thus removed from the figure. Ideally, this point should be measured using a PCB without capacitors.

It is expected for the proposed cells to display the highest leakage power, due to the use of low threshold devices. This expectation is confirmed, as the proposed static and domino cells display a leakage power larger compared to TSMC implementations using standard threshold devices. The differences between TSMC implementations and CooLib-Static are small, however CooLib-Domino implementations inherit substantially more leakage power. CooLib-Domino implementations are substantially larger in sizes in most scenarios, as a result of the unate transformation (section 4.2.2), which appears to be the explanation.

Figure 6.6 shows the leakage (static) power consumption of the low-to-high level shifter. It was expected to show considerable static power consumption (as per section 3.2.1), but no consideration was taken towards scaling of this power consumption with respect to lowering the supply voltage. Clearly, as $V_{DDL}$ moves away

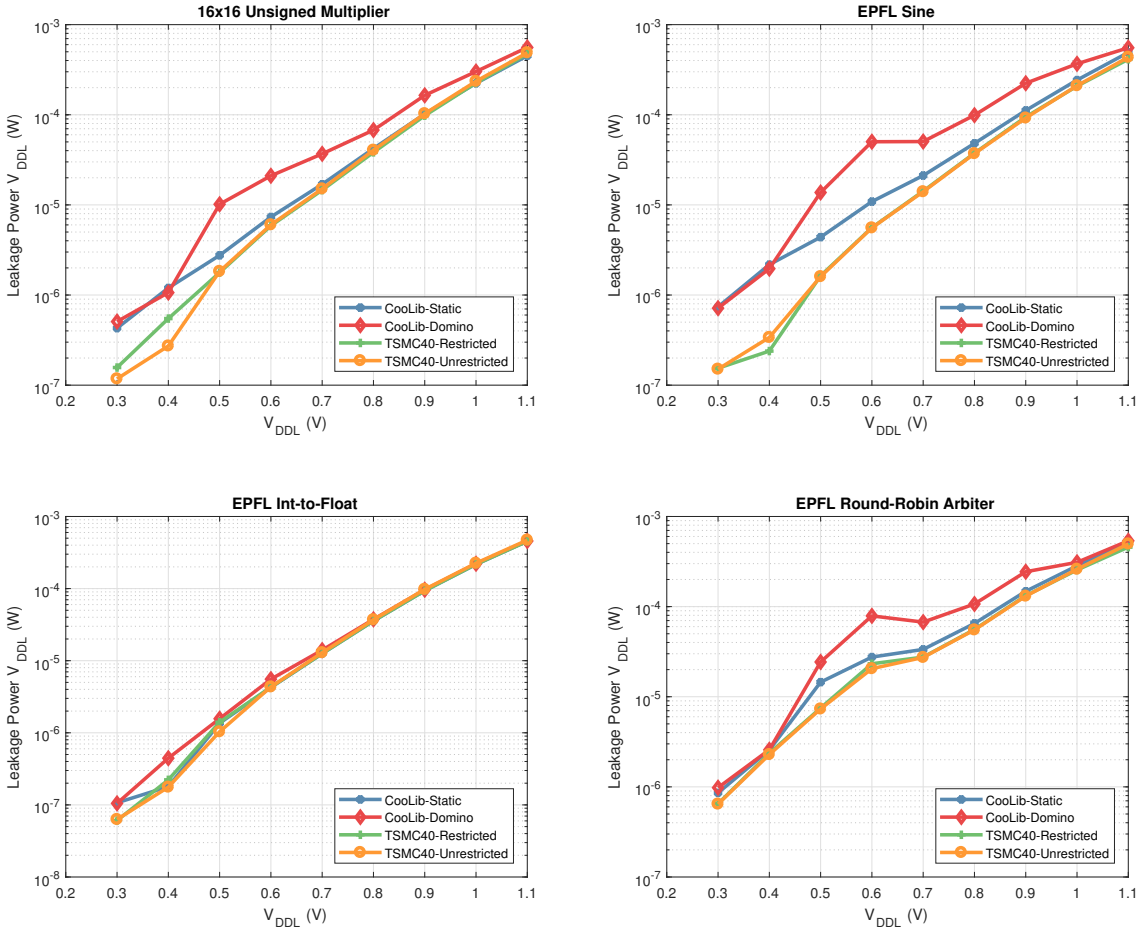



Figure 6.5: 300 K leakage power measurements for the test chip.

from $V_{DDH}$, which is fixed at 1.1 V, the static power consumption increases linearly.

An explanation is found in the PMOS input pair of the cell, as PMOS devices only partially turn off for gate voltages below 1.1 V. Hence, this is worsened as $V_{DDL}$ lowers, as this voltage effectively represents the gate voltage of that input pair.

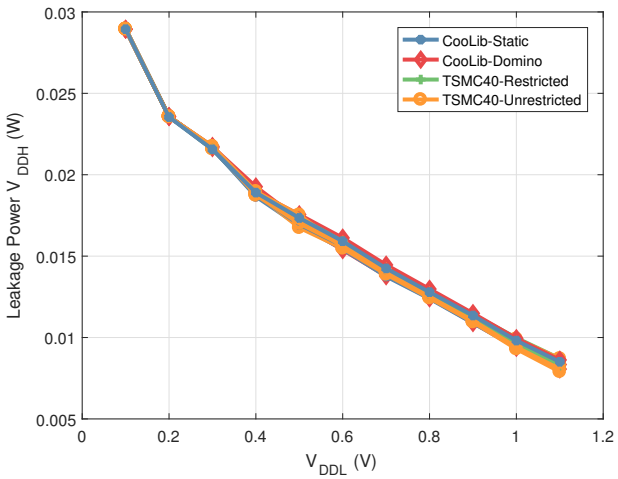



Figure 6.6: Low-to-high level shifter static power consumption at the $V_{DDH}$ supply, over the full range of $V_{DDL}$.

### 6.4.3. 300 K Dynamic Power Measurements

Dynamic power is measured by shifting the LFSR's output at a frequency of 20 MHz into the launch register (as defined in section 4.1.1), while launching the data into the circuit under test when the associated clock signal is low (zero). The domino clock frequency is equal to the shift frequency. This was found to work well up to $V_{DD} = 0.5V$, for lower voltages this frequency appears to be too high, producing inconsistent results. Figure 6.7 displays the results. As expected, domino logic implementations display an increase in dynamic power consumption as a result of the domino logic clock. All other implementations are found to be in the same vicinity in terms of dynamic power consumption, with CooLib-Static generally showing a slight increase compared to TSMC implementations.

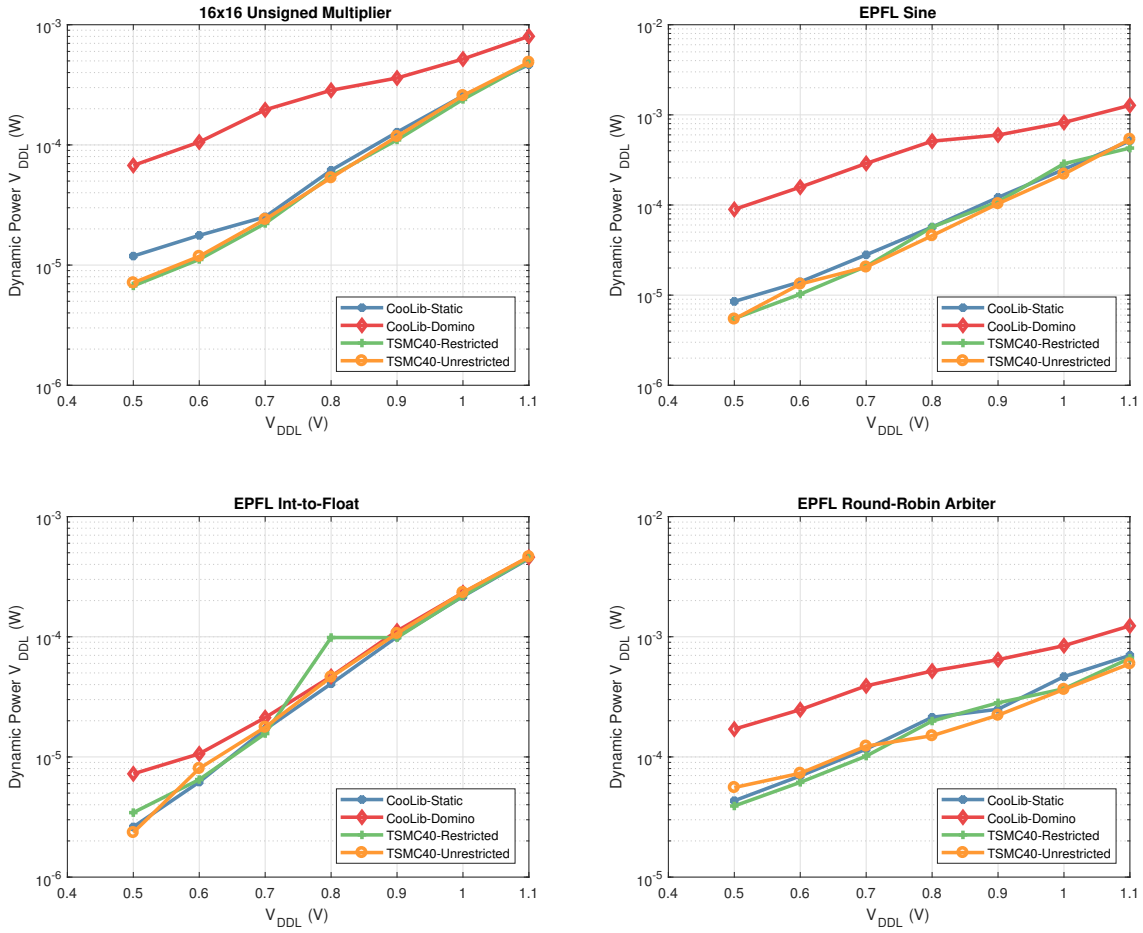



Figure 6.7: 300 K dynamic power measurements for the test chip.

# 7

# Conclusions and Future Work

## 7.1. Conclusions

This work has presented the first standard cell library, specifically aimed at operating at the 4.2 K operating temperature, in context of quantum computing. The library houses a basic set of cells available in two different logic families; static logic for its characteristics and its industry standard reputation, and domino logic as its issues are naturally solved at 4.2 K. Specific to the domino logic, necessary logic transformation algorithms have been implemented in C++, of which the correctness is verified in simulation and silicon. Two chips have been taped out using the proposed standard cell library. One chip was implemented for characterisation, most importantly in terms of power and delay, of common combinational circuits implemented in the proposed cells and conventional standard cells. A second chip implements an existing RISC-V design with the proposed static logic cells, to provide a proof-of-concept of the cell library's functionality. This chip presents the first ever CPU implemented specifically for 4.2 K operation.
The functionality of both chips produced in silicon is confirmed. Specifically, the BIST implementation found in the characterisation chip is proven to enable propagation delay measurements under FPGA supervision. The RISC-V chip has been verified by executing compiled C code on the device. Correct functioning of both chips justifies the flow used to construct the proposed standard cell library.
Future work comprises further testing of both chips, both at room temperature and the target 4.2 K operating temperature. Furthermore, the proposed library might receive continued development by providing a larger set of cells and eliminating poor performing cells.
The proposed standard cell library presented should enable future digital, or mixed-signal, designs to be implemented with the proposed standard cells. Moreover, the transition to using the proposed library should be seamless, due to full compatibility with commercial software tools.

## 7.2. Future Recommendations

A distinction is made between short-term and long-term future recommendations. To start with the short-term future recommendations:

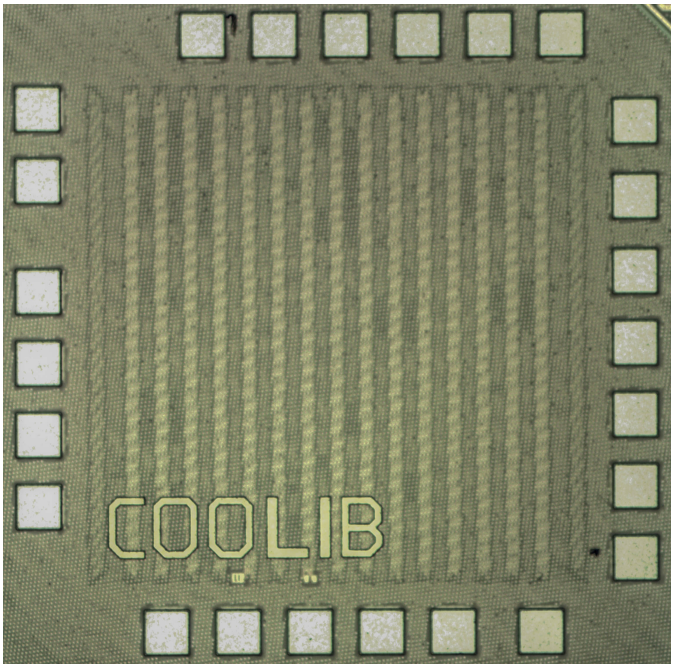
- The OR gate, which was not implemented due to subthreshold logic swing concerns, appears to be required by certain software tools (i.e. Synopsys DC) and should be provided for compatibility under a fair warning.
- Standard cell libraries offered by foundries often include several choices for $V_{TH}$, including ultra low $V_{TH}$. These should be compared to the proposed cell library, as only access to standard $V_{TH}$ cells was given for this work.
- The low-to-high level shifter should be replaced by an architecture which does not consume static power.
- Memories are not considered in this work. The same type of analysis performed in chapter two of this thesis can also be applied to static RAM (SRAM) cells. Additionally, dynamic RAM (DRAM) is expected to show significant increases in retention time at 4.2 K.

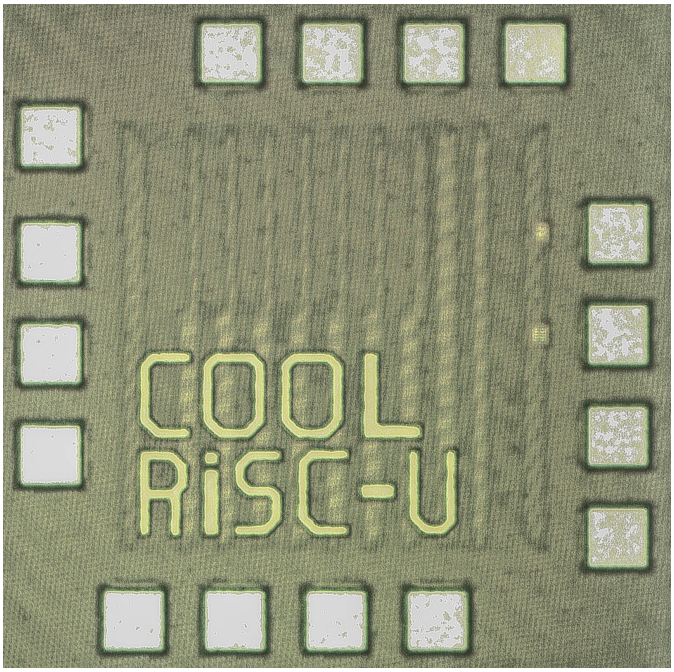
- State-of-the-art process nodes should be considered. For instance, the 14-nm node may present sufficient reductions in (dynamic) power to avoid operating in the subthreshold altogether.

As for the long-term future recommendations:

- Eventually, if a 4.2 K model is proposed for the used process node, the library can be re-characterised to generate PVT corners for this operating temperature. Furthermore, the model can be validated with the measurement results of the test chip.
- Migration of the proposed standard cell library to a more suitable, with respect to low-power, process node. In particular, FD-SOI is recommended.

# Chip Gallery



# Bibliography


[1] B. Patra, R. M. Incandela, J. P. G. van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, "Cryo-CMOS Circuits and Systems for Quantum Computing Applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, Jan 2018.

[2] L. Vandersypen and A. van Leeuwenhoek, "Quantum computing - the next challenge in circuit and system design," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2017, pp. 24–29.

[3] A. Montanaro, "Quantum algorithms: an overview," *npj Quantum Information*, vol. 2, p. 15023, 2016.

[4] B. Patra, R. M. Incandela, J. P. G. van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, "Cryo-cmos circuits and systems for quantum computing applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, Jan 2018.

[5] H. Homulle and E. Charbon, "Performance characterization of altera and xilinx 28 nm fpgas at cryogenic temperatures," in *2017 International Conference on Field Programmable Technology (ICFPT)*, Dec 2017, pp. 25–31.

[6] I. Kuon and J. Rose, "Measuring the Gap Between FPGAs and ASICs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 2, pp. 203–215, Feb 2007.

[7] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital integrated circuits*. Prentice hall Englewood Cliffs, 2002, vol. 2.

[8] S. Hanson, B. Zhai, K. Bernstein, D. Blaauw, A. Bryant, L. Chang, K. K. Das, W. Haensch, E. J. Nowak, and D. M. Sylvester, "Ultralow-voltage, minimum-energy CMOS," *IBM journal of research and development*, vol. 50, no. 4.5, pp. 469–490, 2006.

[9] M. Alioto, "Understanding DC behavior of subthreshold CMOS logic through closed-form analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1597–1607, 2010.

[10] D. Bol, R. Ambroise, D. Flandre, and J.-D. Legat, "Interests and limitations of technology scaling for subthreshold logic," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 10, pp. 1508–1519, 2009.

[11] L. A. Melek, M. C. Schneider, and C. Galup-Montoro, "Ultra-low voltage CMOS logic circuits," in *Micro-Nanoelectronics, Technology and Applications (EAMTA), 2014 Argentine Conference on*. IEEE, 2014, pp. 1–7.

[12] R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE Journal of Solid-State Circuits*, vol. 7, no. 2, pp. 146–153, 1972.

[13] G. Schrom and S. Selberherr, "Ultra-low-power CMOS technologies," in *Semiconductor Conference, 1996., International*, vol. 1. IEEE, 1996, pp. 237–246.

[14] A. Bryant, J. Brown, P. Cottrell, M. Ketchen, J. Ellis-Monaghan, and E. Nowak, "Low-power CMOS at Vdd= 4kT/q," in *Device Research Conference, 2001*. IEEE, 2001, pp. 22–23.

[15] N. Lotze and Y. Manoli, "A 62 mV 0.13um CMOS Standard-Cell-Based Design Technique Using Schmitt-Trigger Logic," *IEEE journal of solid-state circuits*, vol. 47, no. 1, pp. 47–60, 2012.

[16] L. A. P. Melek, A. L. da Silva, M. C. Schneider, and C. Galup-Montoro, "Analysis and Design of the Classical CMOS Schmitt Trigger in Subthreshold Operation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 4, pp. 869–878, April 2017.

[17] M. Alioto, "Impact of NMOS/PMOS imbalance in ultra-low voltage CMOS standard cells," in *Circuit Theory and Design (ECCTD), 2011 20th European Conference on*. IEEE, 2011, pp. 536–539.

[18] T. Kim, J. Keane, H. Eom, and C. H. Kim, "Utilizing Reverse Short-Channel Effect for Optimal Subthreshold Circuit Design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 15, no. 7, pp. 821–829, July 2007.

[19] J. Morris, P. Prabhat, J. Myers, and A. Yakovlev, "Unconventional Layout Techniques for a High Performance, Low Variability Subthreshold Standard Cell Library," in *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2017, pp. 19–24.

[20] N. Shigyo and T. Hiraoka, "A review of narrow-channel effects for STI MOSFET's: A difference between surface- and buried-channel cases," *Solid-State Electronics*, vol. 43, no. 11, pp. 2061 – 2066, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038110199001574

[21] E. Maricau and G. Gielen, *Analog IC reliability in nanometer CMOS*. Springer Science & Business Media, 2013.

[22] M. Alioto, "Ultra-low power VLSI circuit design demystified and explained: A tutorial," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 1, pp. 3–29, 2012.

[23] Massimo Alioto, *Enabling the Internet of Things: From Integrated Circuits to Integrated Systems*. Springer, 2017.

[24] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct 1989.

[25] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and Compact Modeling of Nanometer CMOS Transistors at Deep-Cryogenic Temperatures," *IEEE Journal of the Electron Devices Society*, pp. 1–1, 2018.

[26] H. Homulle, L. Song, E. Charbon, and F. Sebastiano, "The Cryogenic Temperature Behavior of Bipolar, MOS, and DTMOS Transistors in Standard CMOS," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 263–270, 2018.

[27] J. B. Burr, "Cryogenic ultra low power CMOS," in *Low Power Electronics, 1995., IEEE Symposium on*. IEEE, 1995, pp. 82–83.

[28] B. Szelag, F. Balestra, and G. Ghibaudo, "Comprehensive analysis of reverse short-channel effect in silicon MOSFETs from low-temperature operation," *IEEE Electron Device Letters*, vol. 19, no. 12, pp. 511–513, 1998.

[29] G. Ghibaudo and F. Balestra, "Device Physics and Electrical Performance of Bulk Silicon Mosfets," in *Device and circuit cryogenic operation for low temperature electronics*. Springer, 2001, pp. 3–35.

[30] P. Martin, M. Cavelier, R. Fascio, and G. Ghibaudo, "MOSFET Compact Modeling Issues for Low Temperature (77 K-200 K) Operation," in *NSTI-Nanotech Conference, Workshop on Compact Modeling*, vol. 3, 2008, pp. 865–868.

[31] F. Frustaci, P. Corsonello, and S. Perri, "Analytical Delay Model Considering Variability Effects in Subthreshold Domain," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 3, pp. 168–172, March 2012.

[32] P. A. 't Hart, J. P. G. van Dijk, M. Babaie, E. Charbon, A. Vladimirescu and F. Sebastiano, "Characterization and Model Validation of Mismatch in Nanometer CMOS at Cryogenic Temperatures."

[33] N. D. Arora and M. S. Sharma, "MOSFET Substrate Current Model for Circuit Simulation," *IEEE Transactions on Electron Devices*, vol. 38, no. 6, pp. 1392–1398, June 1991.

[34] L. Deferm, E. Simoen, B. Dierickx, and C. Claeys, "Anomalous latch-up behaviour of CMOS at liquid helium temperatures," *Cryogenics*, vol. 30, no. 12, pp. 1051 – 1055, 1990. [Online]. Available: http://www.sciencedirect.com/science/article/pii/001122759090206R

[35] C. J. Marshall, P. W. Marshall, R. L. Ladbury, A. Waczynski, R. Arora, R. D. Foltz, J. D. Cressler, D. M. Kahle, D. Chen, G. S. Delo, N. A. Dodds, J. A. Pellish, E. Kan, N. Boehm, R. A. Reed, and K. A. LaBel, "Mechanisms and Temperature Dependence of Single Event Latchup Observed in a CMOS Readout Integrated Circuit From 16–300 K," *IEEE Transactions on Nuclear Science*, vol. 57, no. 6, pp. 3078–3086, Dec 2010.

[36] E. A. Gutierrez-D, J. Deen, and C. Claeys, *Low temperature electronics: physics, devices, circuits, and applications*. Elsevier, 2000.

[37] R. Hossain, *High performance ASIC design: using synthesizable domino logic in an ASIC flow*. Cambridge University Press Cambridge, 2008.

[38] H. Soeleman, K. Roy, and B. Paul, "Sub-domino logic: ultra-low power dynamic sub-threshold digital logic," in *VLSI Design, 2001. Fourteenth International Conference on*. IEEE, 2001, pp. 211–214.

[39] R. Kumar and G. Hinton, "A family of 45nm IA processors," in *2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*, Feb 2009, pp. 58–59.

[40] H. Soeleman, K. Roy, and B. C. Paul, "Robust subthreshold logic for ultra-low power operation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 9, no. 1, pp. 90–99, Feb 2001.

[41] W. Elgharbawy and M. Bayoumi, "New bulk dynamic threshold NMOS schemes for low-energy subthreshold domino-like circuits," in *IEEE Computer Society Annual Symposium on VLSI*, Feb 2004, pp. 115–120.

[42] N. Reynders and W. Dehaene, *Ultra-Low-Voltage Design of Energy-Efficient Digital Circuits*. Springer, 2015.

[43] M. W. Allam and M. I. Elmasry, "Dynamic current mode logic (dycml): a new low-power high-performance logic style," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 3, pp. 550–558, March 2001.

[44] J. Zhou, S. Jayapal, B. Busze, L. Huang, and J. Stuyt, "A 40 nm inverse-narrow-width-effect-aware subthreshold standard cell library," in *2011 48th ACM/EDAC/IEEE Design Automation Conference (DAC)*, June 2011, pp. 441–446.

[45] J. Kwong and A. P. Chandrakasan, "Variation-Driven Device Sizing for Minimum Energy Sub-threshold Circuits," in *ISLPED'06 Proceedings of the 2006 International Symposium on Low Power Electronics and Design*, Oct 2006, pp. 8–13.

[46] S. Timarchi and M. Alioto, "Ultra-low voltage standard cell libraries: Design strategies and a case study," in *2016 IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Dec 2016, pp. 520–523.

[47] M. Arif, M. Bhuiyan, A. Khairul, R. Rosli, N. Mohd, R. Hafizi, M. Daud, M. B. I. Reaz, S. Arif, and B. , "A comparative study on SOI MOSFETS for low power applications," vol. 5, pp. 2586–2591, 03 2013.

[48] P. Magarshack, P. Flatresse, and G. Cesana, "UTBB FD-SOI: A process/design symbiosis for breakthrough energy-efficiency," in *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2013, pp. 952–957.

[49] N. Planes, O. Weber, V. Barral, S. Haendler, D. Noblet, D. Croain, M. Bocat, P. . Sassoulas, X. Federspiel, A. Cros, A. Bajolet, E. Richard, B. Dumont, P. Perreau, D. Petit, D. Golanski, C. Fenouillet-Béranger, N. Guillot, M. Rafik, V. Huard, S. Puget, X. Montagner, M. . Jaud, O. Rozeau, O. Saxod, F. Wacquant, F. Monsieur, D. Barge, L. Pinzelli, M. Mellier, F. Boeuf, F. Arnaud, and M. Haond, "28nm FDSOI technology platform for high-speed low-voltage digital applications," in *2012 Symposium on VLSI Technology (VLSIT)*, June 2012, pp. 133–134.

[50] E. Låte, A. A. Vatanjou, T. Ytterdal, and S. Aunet, "Extended Comparative Analysis of Flip-Flop Architectures for Subthreshold Applications in 28 nm FD-SOI," *Microprocessors and Microsystems*, vol. 48, pp. 11 – 20, 2017, extended papers from the 2015 Nordic Circuits and Systems Conference. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0141933116301004

[51] V. Stojanovic, V. Oklobdzija, and R. Bajwa, "Comparative analysis of latches and flip-flops for high-performance systems," in *Proceedings International Conference on Computer Design. VLSI in Computers and Processors (Cat. No.98CB36273)*, Oct 1998, pp. 264–269.

[52] S. Lutkemeier and U. Ruckert, "A Subthreshold to Above-Threshold Level Shifter Comprising a Wilson Current Mirror," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 9, pp. 721–724, Sept 2010.

[53] B. Razavi, *Design of Analog CMOS Integrated Circuits*, 1st ed. New York, NY, USA: McGraw-Hill, Inc., 2001.

[54] X. Meng, R. Saleh, and K. Arabi, "Layout of Decoupling Capacitors in IP Blocks for 90-nm CMOS," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 16, no. 11, pp. 1581–1588, Nov. 2008. [Online]. Available: https://doi.org/10.1109/TVLSI.2008.2001240

[55] A. B. Kahng, "New game, new goal posts: A recent history of timing closure," in *Design Automation Conference (DAC), 2015 52nd ACM/EDAC/IEEE*. IEEE, 2015, pp. 1–6.

[56] J. Bhasker and R. Chadha, "Standard cell library," in *Static Timing Analysis for Nanometer Designs*. Springer, 2009, pp. 43–100.

[57] "SiliconSmart ACE User Guide," *Synopsys Inc.*, 2016.

[58] "LEF/DEF Language Reference Version 5.7," *USA: Cadence Design System Inc*, vol. 56, no. 5, pp. 57–62, 2003.

[59] J. S. J. Wong, P. Sedcole, and P. Y. K. Cheung, "Self-Measurement of Combinatorial Circuit Delays in FPGAs," *ACM Trans. Reconfigurable Technol. Syst.*, vol. 2, no. 2, pp. 10:1–10:22, Jun. 2009. [Online]. Available: http://doi.acm.org/10.1145/1534916.1534920

[60] L. Amarù, P.-E. Gaillardon, and G. De Micheli, "The EPFL Combinational Benchmark Suite," *Proceedings of the 24th International Workshop on Logic & Synthesis (IWLS)*, 2015.

[61] Cadence Design Systems, Inc. (2017) Innovus Implementation System. [Online]. Available: https://www.cadence.com/content/dam/cadence-www/global/en_US/documents/tools/digital-design-signoff/innovus-implementation-system-ds.pdf

[62] V. Yuzhaninov, I. Levi, and A. Fish, "Design Flow and Characterization Methodology for Dual Mode Logic," *IEEE Access*, vol. 3, pp. 3089–3101, 2015.

[63] A. Cao and C.-K. Koh, "Post-layout Logic Optimization of Domino Circuits," in *Proceedings of the 41st Annual Design Automation Conference*, ser. DAC '04. New York, NY, USA: ACM, 2004, pp. 820–825. [Online]. Available: http://doi.acm.org/10.1145/996566.996786

[64] RISC-V Foundation. Risc-v. [Online]. Available: https://riscv.org/

[65] H. Homulle, S. Visser, B. Patra, G. Ferrari, E. Prati, F. Sebastiano, and E. Charbon, "A reconfigurable cryogenic platform for the classical control of quantum processors," *Review of Scientific Instruments*, vol. 88, no. 4, p. 045103, 2017. [Online]. Available: https://doi.org/10.1063/1.4979611

[66] Xilinx Inc. (2018) 7 Series FPGAs Clocking Resources. [Online]. Available: https://www.xilinx.com/support/documentation/user_guides/ug472_7Series_Clocking.pdf

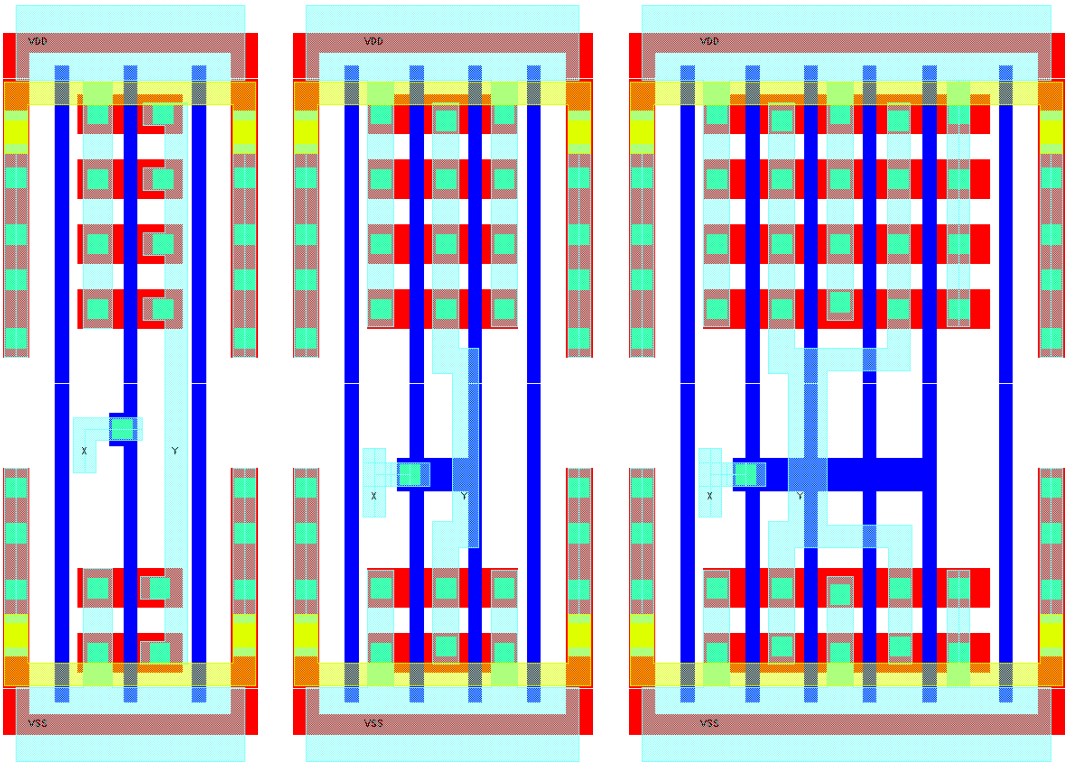
[67] H. Homulle, S. Visser, B. Patra, and E. Charbon, "Design techniques for a stable operation of cryogenic field-programmable gate arrays," *Review of Scientific Instruments*, vol. 89, no. 1, p. 014703, 2018. [Online]. Available: https://doi.org/10.1063/1.5004484

# A

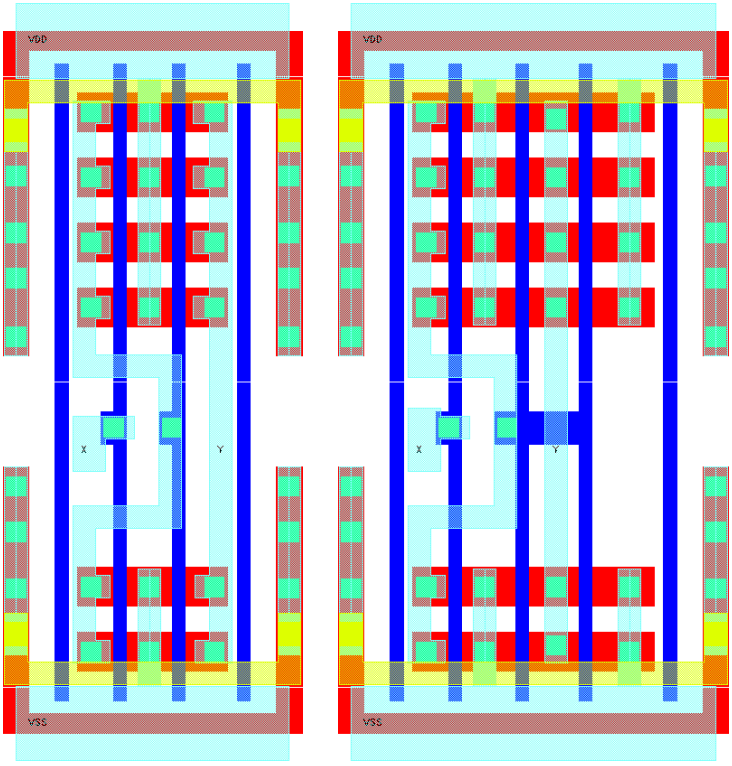
# Standard Cell Layouts

This appendix contains the layout of all the cells present in the proposed standard cell library.

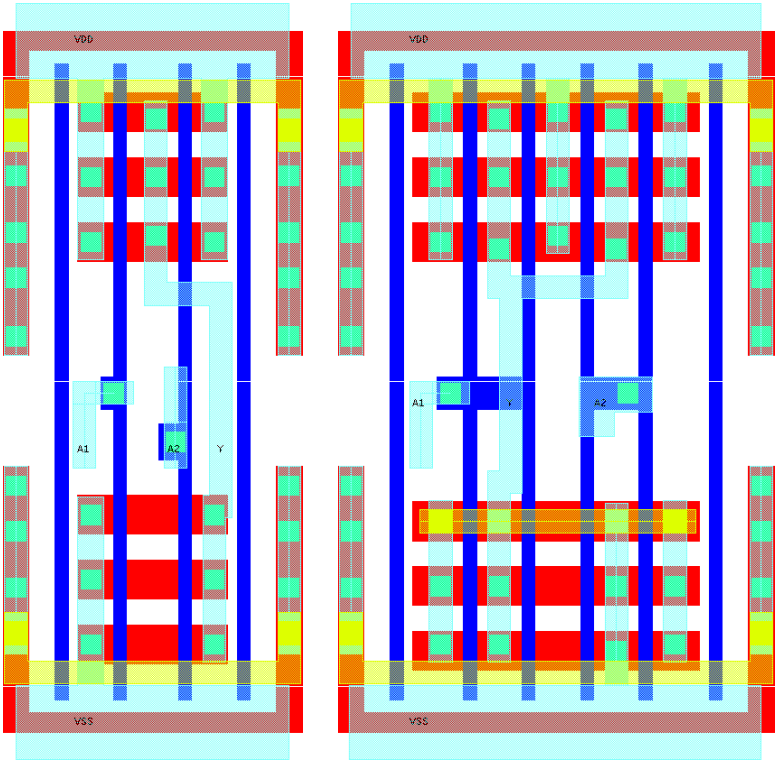
## Static Inverters



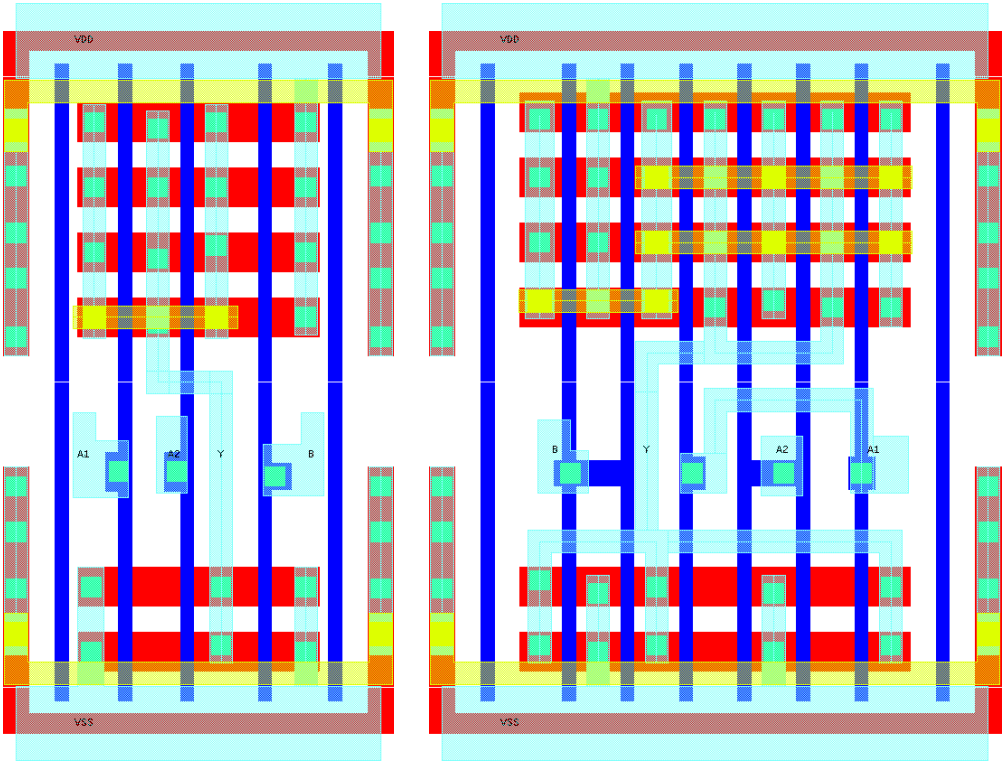
## Static Buffers



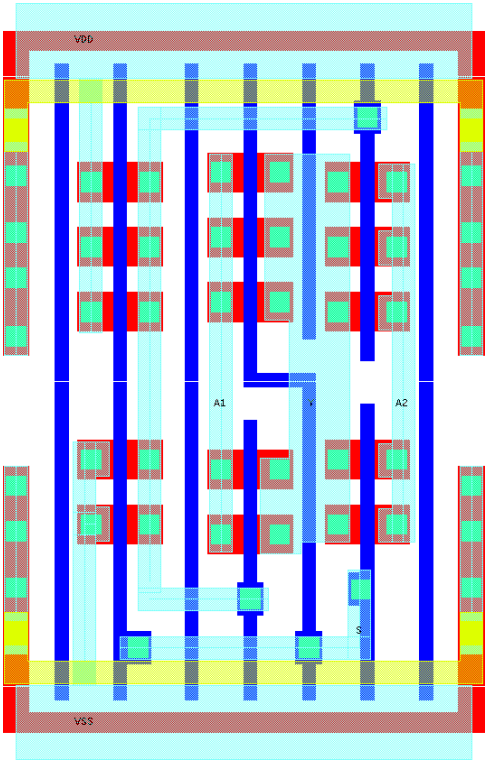

## Static NANDs



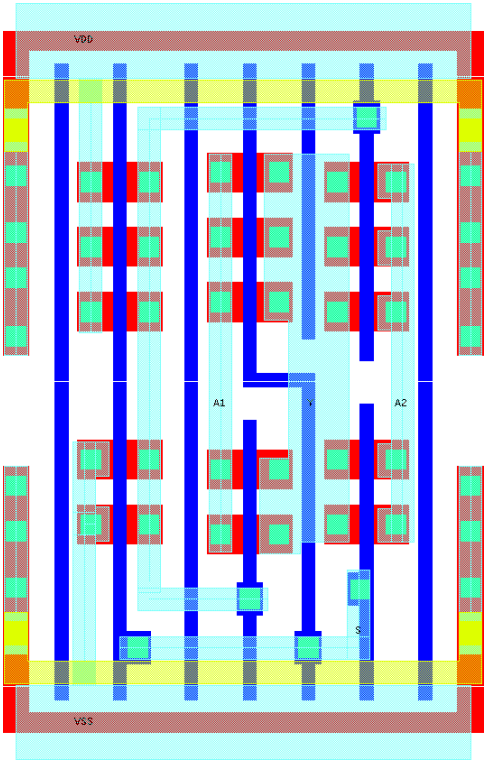
## Static AOIs



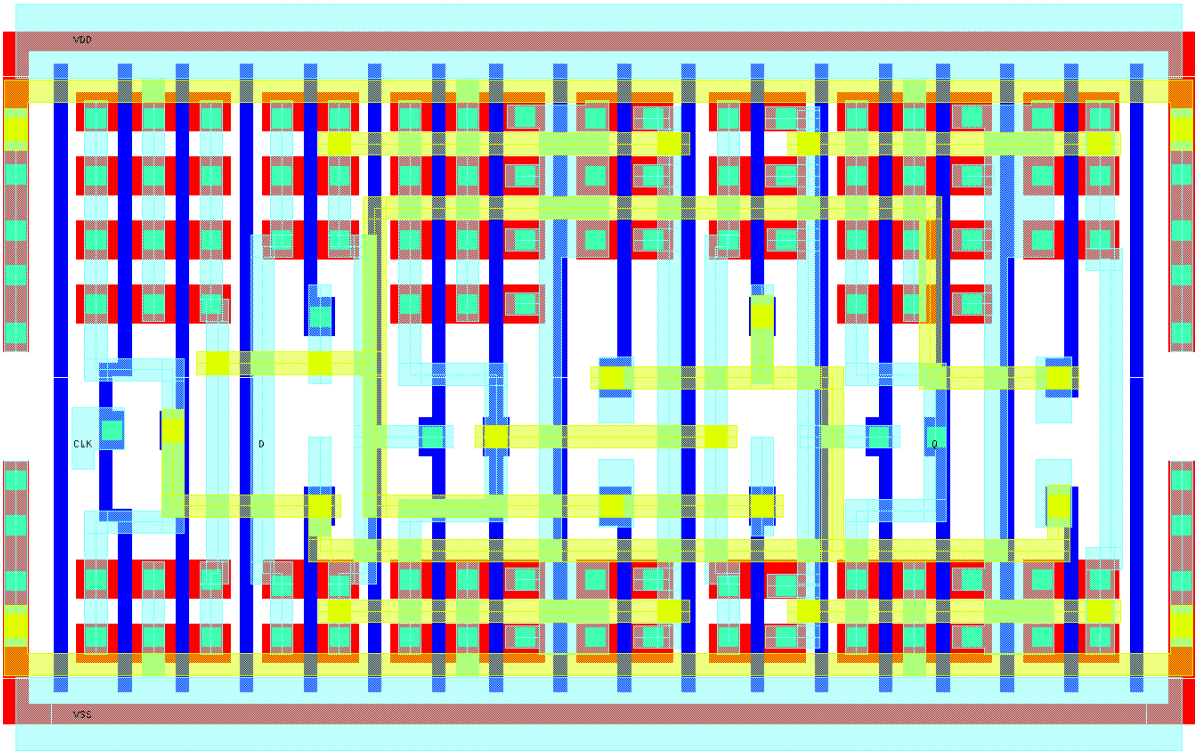

## Transmission Gate Mux



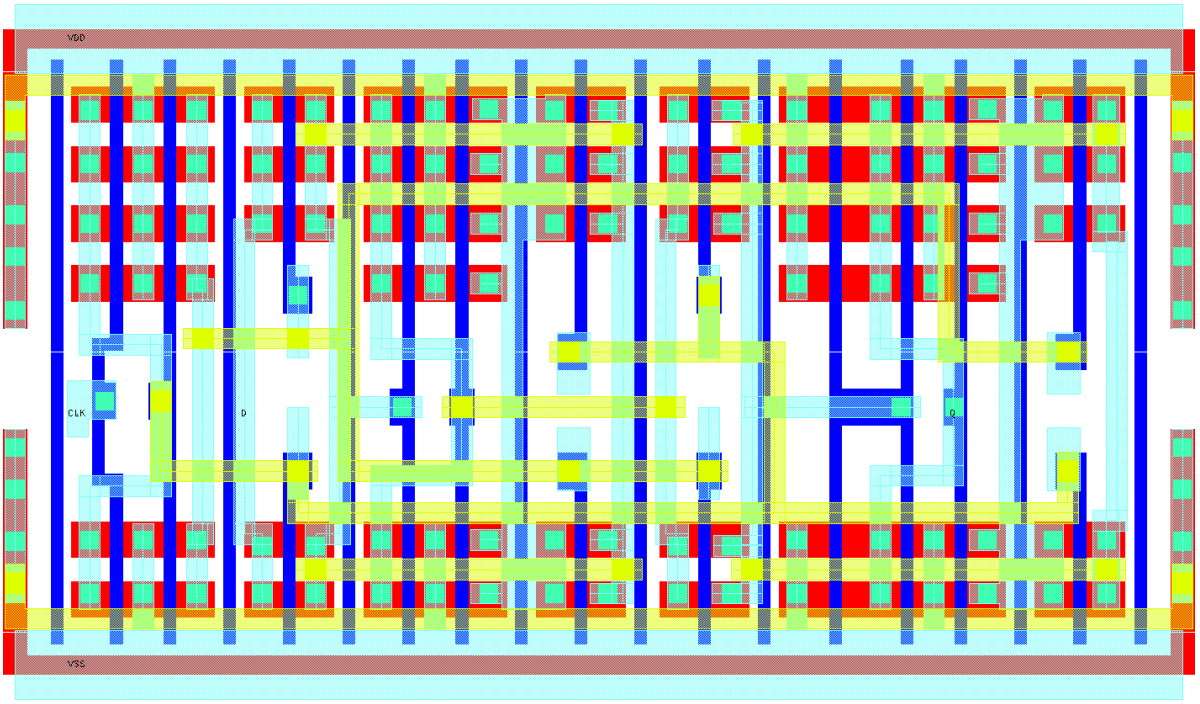
## Transmission Gate Mux



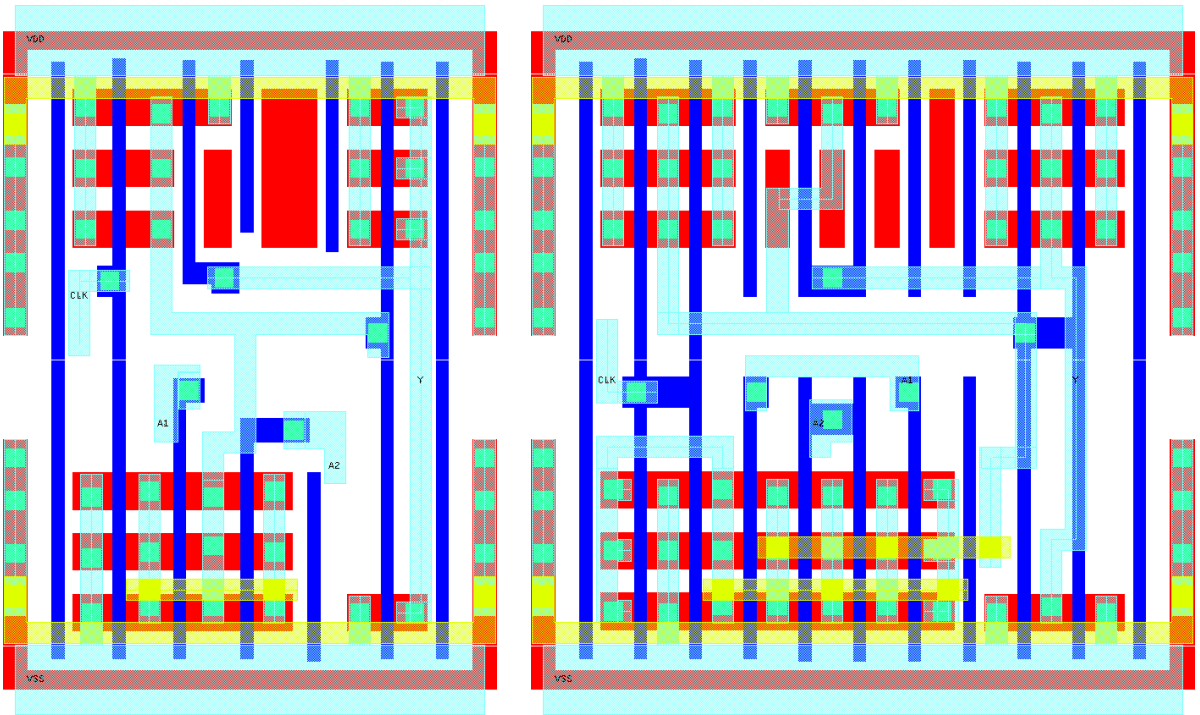

## Positive Triggered D-type Flip-flops (No Reset) Single Strength



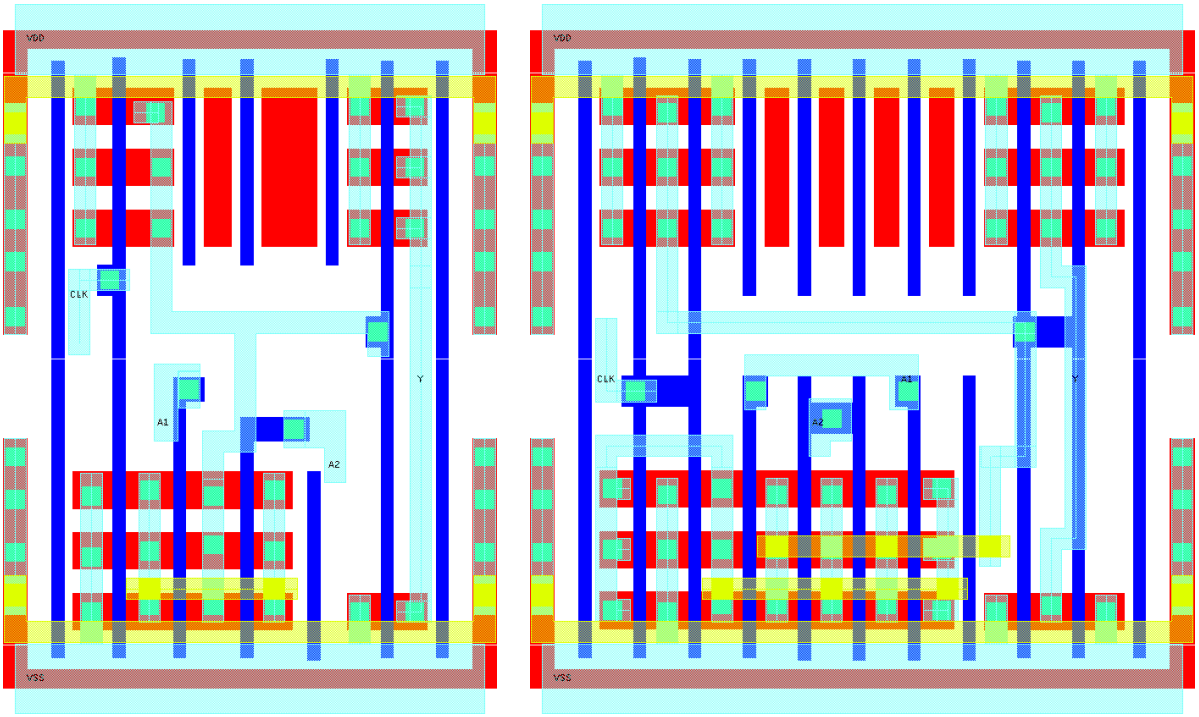
## Positive Triggered D-type Flip-flops (No Reset) Double Strength



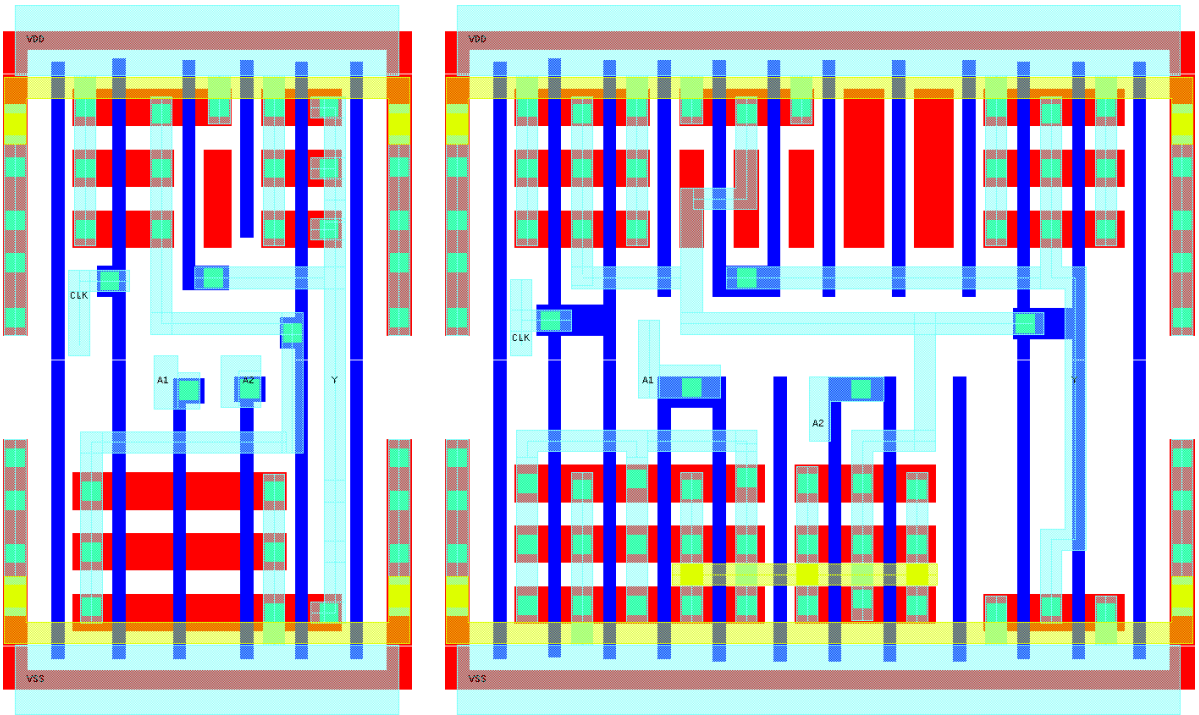

## Domino ORs



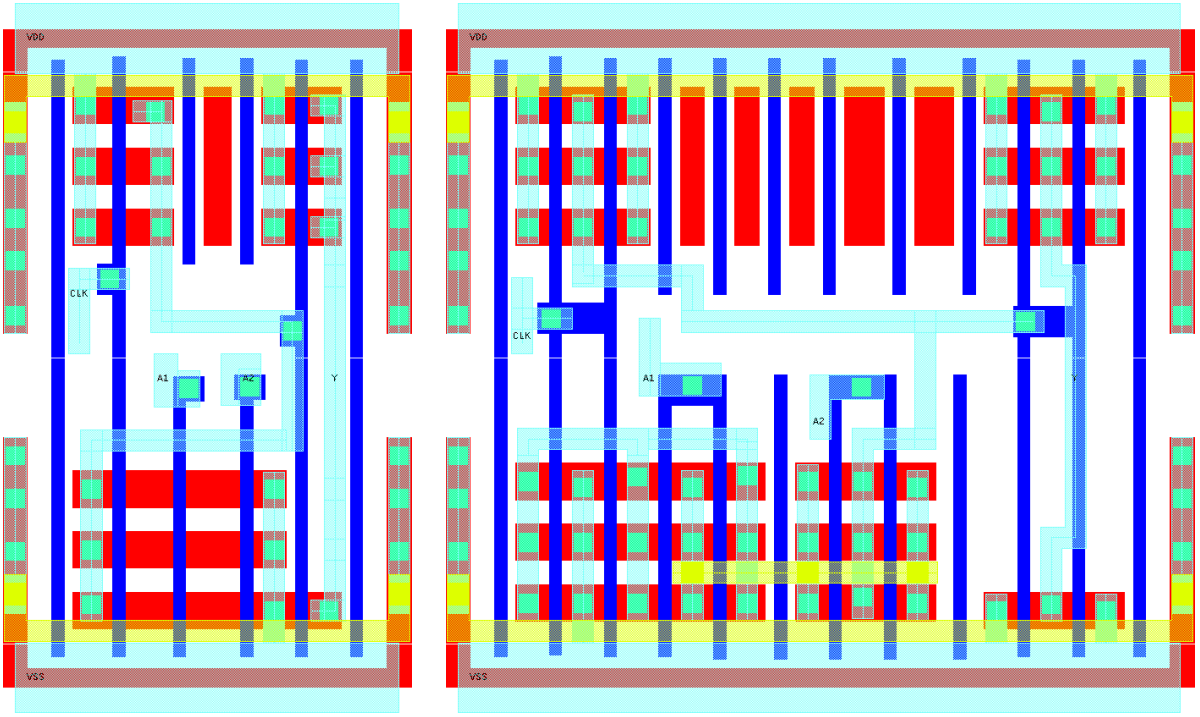
## Domino ORs (No Weak-Keeper)



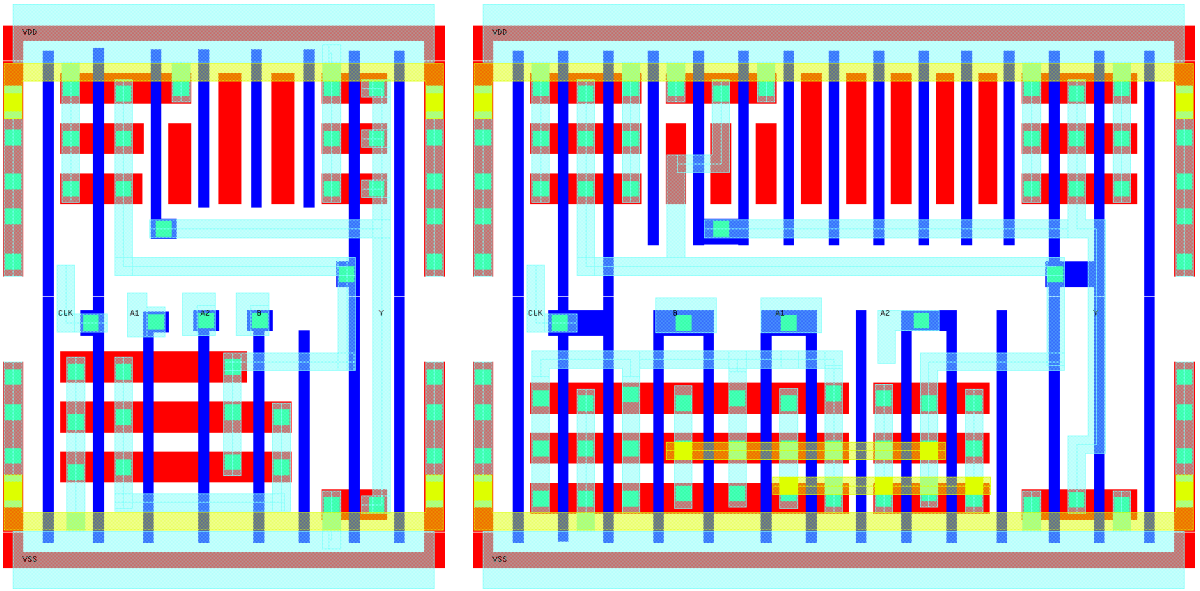

## Domino ANDs



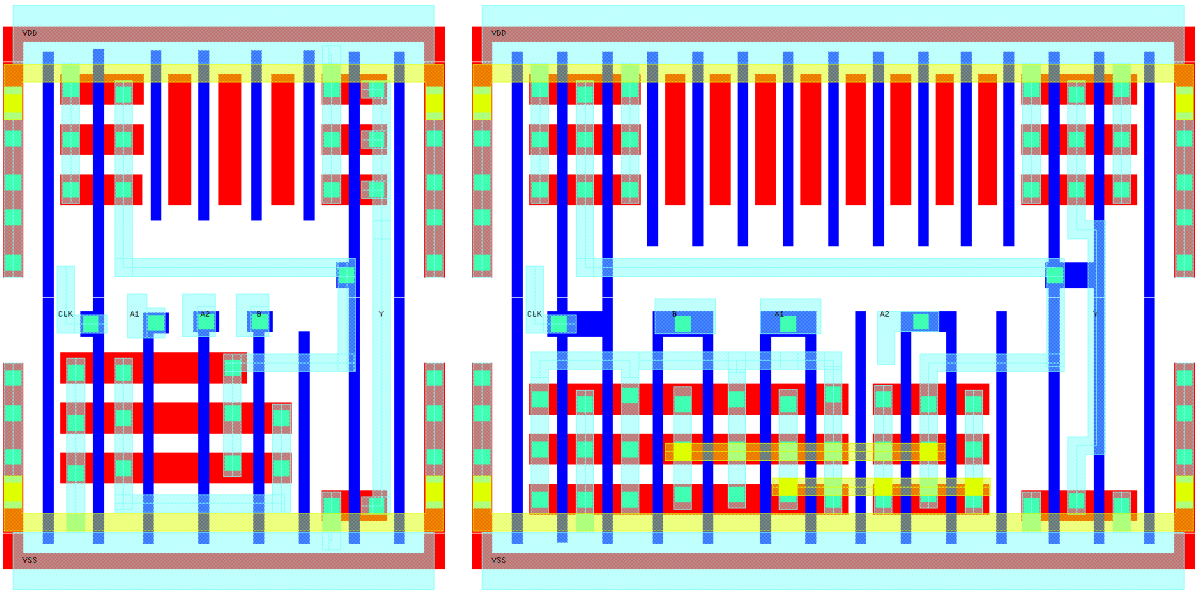
## Domino ANDs (No Weak-Keeper)



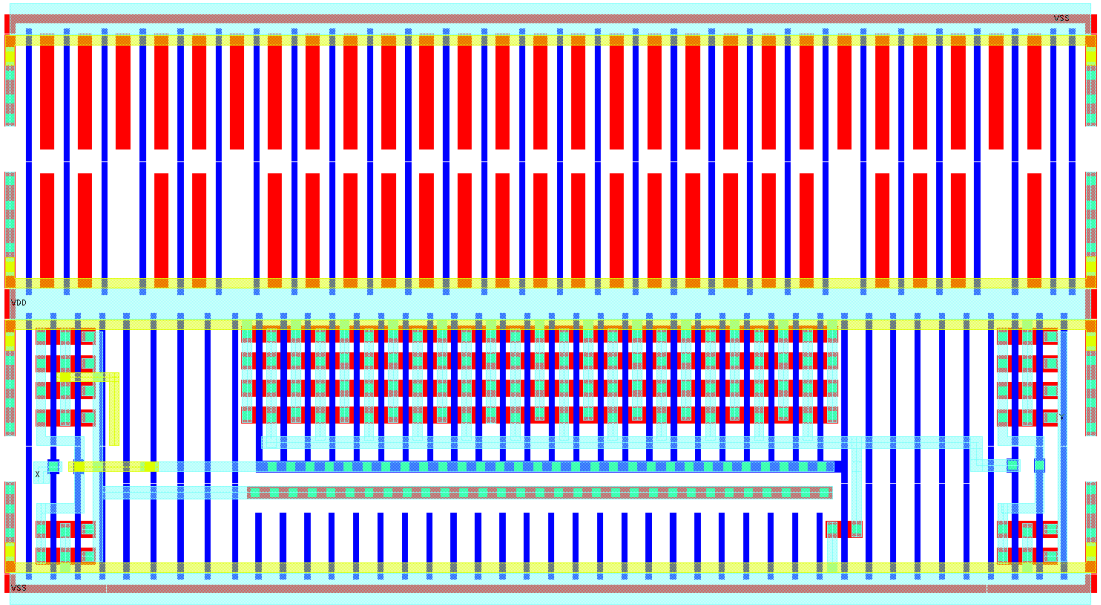

## Domino AOs



## Domino AOs (No Weak-Keeper)

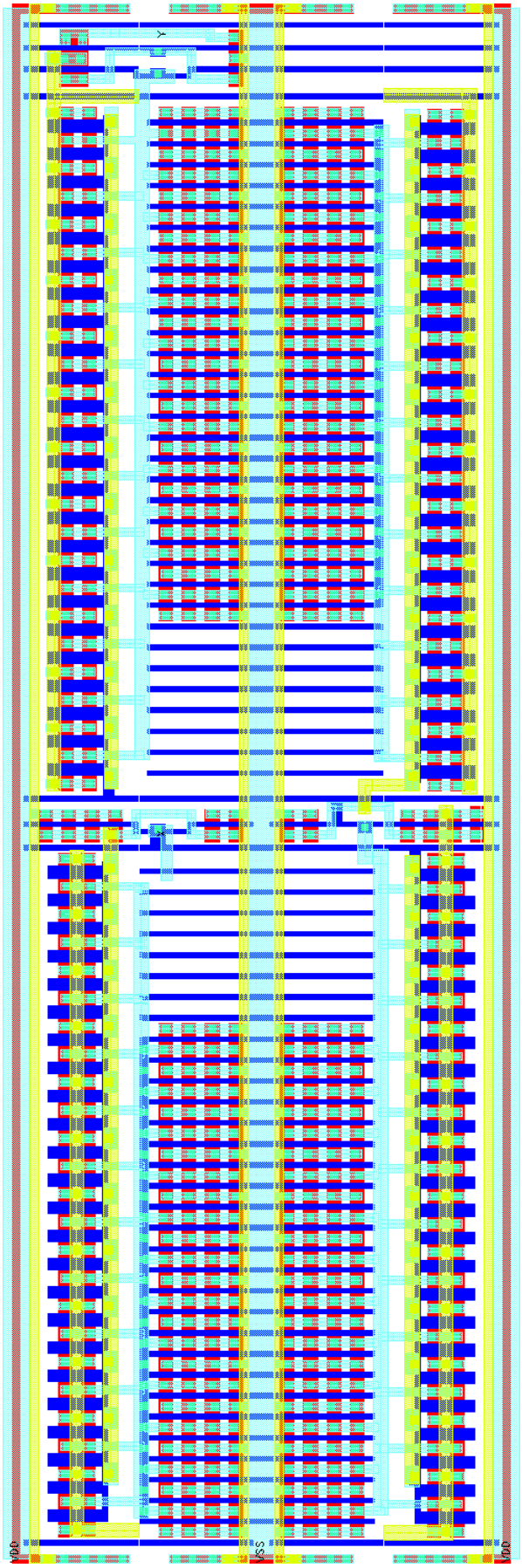



## Level-Down Shifter

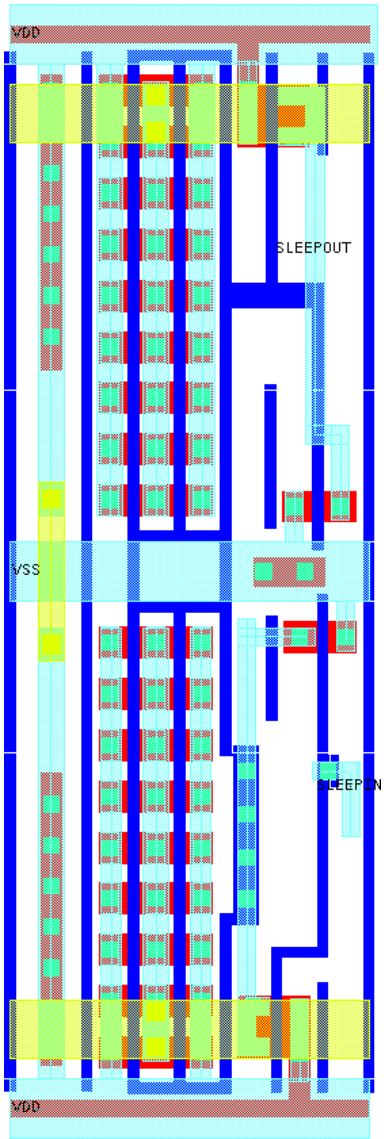



## Level-Up Shifter

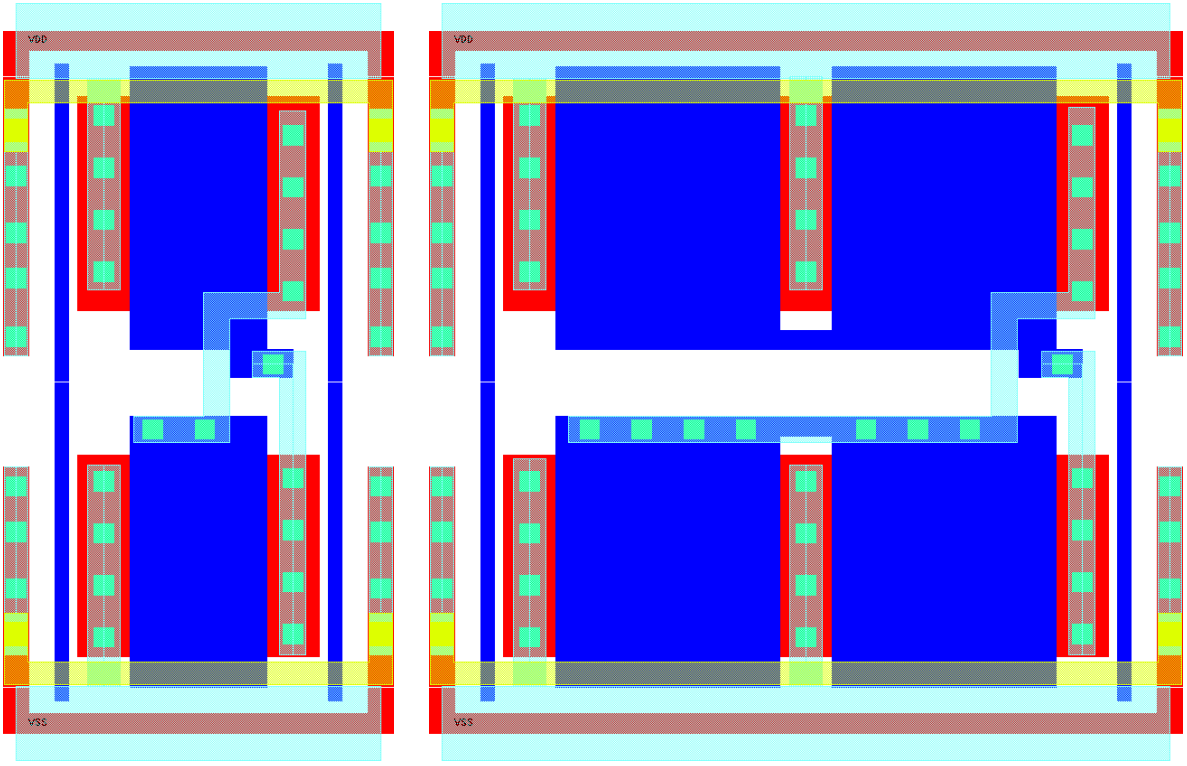
Rotated for clearer view.


Y
X
VDD
VSS
VDD

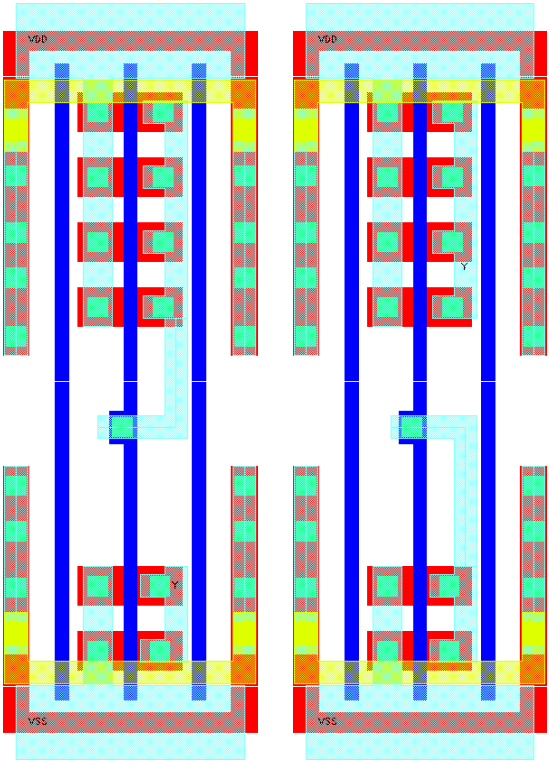
## Power Gate (Header)



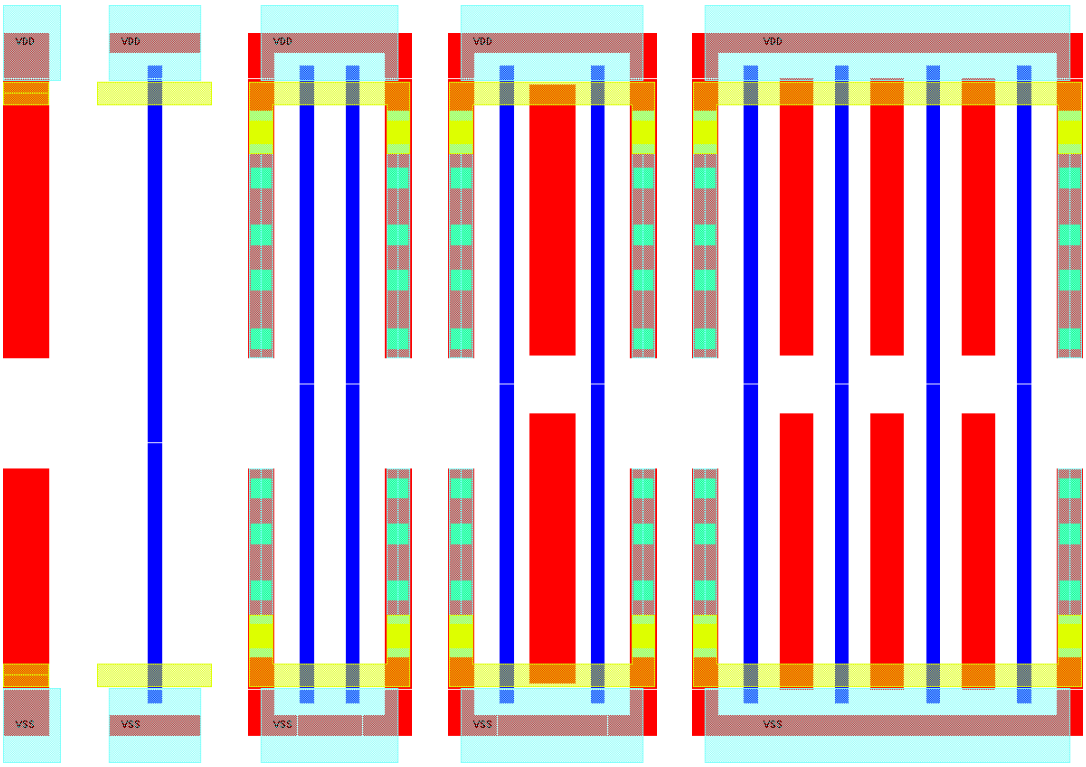
## Decap cells



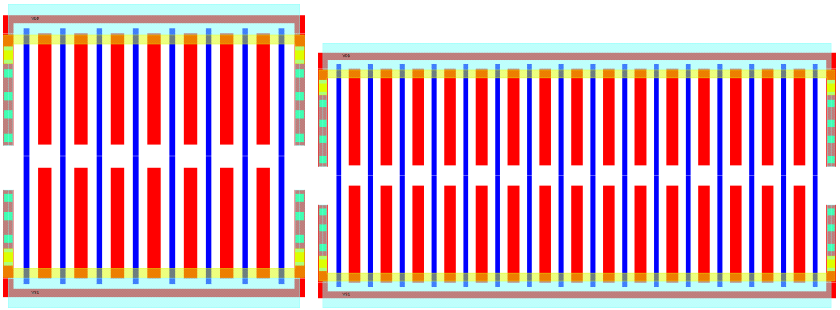

## Tie cells



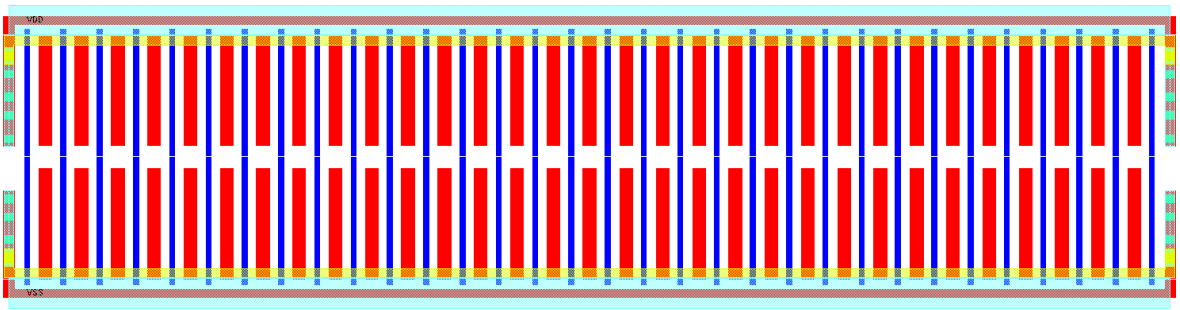
## Filler cells (1, 2, 3, 4, 8 Tracks)




## Filler cells (16, 32 Tracks)



## Filler cells (64 Tracks)



# B

# RISC-V Programming & Firmware

To load firmware into the SRAM found in the RISC-V chip, the sequences stated in chapter 5 should be implemented by an FPGA or similar. This appendix provides the auxiliary scripts needed to convert C(++) code to a coefficient file required by Xilinx Vivado.

## Startup.S

Initialises registers and calls main.

```
1
2  .section .text
3
4  start:
5
6  # zero-initialize register file
7  addi x1, zero, 0
8  # x2 (sp) is initialized by reset
9  addi x3, zero, 0
10 addi x4, zero, 0
11 addi x5, zero, 0
12 addi x6, zero, 0
13 addi x7, zero, 0
14 addi x8, zero, 0
15 addi x9, zero, 0
16 addi x10, zero, 0
17 addi x11, zero, 0
18 addi x12, zero, 0
19 addi x13, zero, 0
20 addi x14, zero, 0
21 addi x15, zero, 0
22 addi x16, zero, 0
23 addi x17, zero, 0
24 addi x18, zero, 0
25 addi x19, zero, 0
26 addi x20, zero, 0
27 addi x21, zero, 0
28 addi x22, zero, 0
29 addi x23, zero, 0
30 addi x24, zero, 0
31 addi x25, zero, 0
32 addi x26, zero, 0
33 addi x27, zero, 0
```

```
34 addi x28, zero, 0
35 addi x29, zero, 0
36 addi x30, zero, 0
37 addi x31, zero, 0
38
39 call main
40
41 loop:
42 j loop
```

## Sections.lds

Most basic linker script possible.

```
1  /*
2  This is free and unencumbered software released into the public domain.
3
4  Anyone is free to copy, modify, publish, use, compile, sell, or
5  distribute this software, either in source code form or as a compiled
6  binary, for any purpose, commercial or non-commercial, and by any
7  means.
8  */
9
10 MEMORY {
11         /* SRAM size */
12         mem : ORIGIN = 0x00000000, LENGTH = 8K
13 }
14
15 SECTIONS {
16         .memory : {
17                 . = 0x000000;
18                 start*(.text);
19                 *(.text);
20                 *(*);
21                 end = .;
22         } > mem
23 }
```

## Makefile

Basic makefile, requires the makehex python script available in the PicoRV32 Github repository.

```
1
2  ARCH=i
3  COMPILER_DIR=/opt/riscv32$(ARCH)/
4
5  firmware.elf: sections.lds start.s firmware.c
6          $(COMPILER_DIR)/bin/riscv32-unknown-elf-gcc -march=rv32$(ARCH) -Wl,-Bstatic
               ↪ ,-T,sections.lds,--strip-debug -ffreestanding -nostdlib -Wl,--no-
               ↪ relax -o firmware.elf start.s firmware.c
7
8  firmware.bin: firmware.elf
9          $(COMPILER_DIR)/bin/riscv32-unknown-elf-objcopy -O binary firmware.elf
               ↪ firmware.bin
10
11 firmware.coe: firmware.elf
```

```
12          python3 makehex.py firmware.bin 2048 > firmware.coe
13
14 clean:
15         rm -f firmware.elf firmware.hex firmware.bin
16
17 all : firmware.elf firmware.bin firmware.coe
```

## Makehex.py

Simple Python script used to convert the binary into coefficient file format.

```
1
2 #!/usr/bin/env python3
3 #
4 # This is free and unencumbered software released into the public domain.
5 #
6 # Anyone is free to copy, modify, publish, use, compile, sell, or
7 # distribute this software, either in source code form or as a compiled
8 # binary, for any purpose, commercial or non-commercial, and by any
9 # means.
10
11 from sys import argv
12
13 binfile = argv[1]
14 nwords = int(argv[2])
15
16 with open(binfile, "rb") as f:
17     bindata = f.read()
18
19 assert len(bindata) < 4*nwords
20
21 print("memory_initialization_radix=16;")
22 print("memory_initialization_vector=")
23
24 for i in range(nwords):
25     if i < len(bindata) // 4:
26         w = bindata[4*i : 4*i+4]
27
28         print("%02x%02x%02x%02x," % (w[3], w[2], w[1], w[0]))
29     else:
30         print("00000000,")
```