

Doctoral thesis

Doctoral theses at NTNU, 2021:161

Even Låte

Low Voltage Logic and Memories on Silicon

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Electronic Systems



Norwegian University of
Science and Technology

Even Låte

Low Voltage Logic and Memories on Silicon

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU

Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

© Even Låte

ISBN 978-82-326-6512-9 (printed ver.)

ISBN 978-82-326-6971-4 (electronic ver.)

ISSN 1503-8181 (printed ver.)

ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:161

Printed by NTNU Grafisk senter

Abstract

Data centers, sensor nodes in the internet of things, and handheld devices are all systems that require energy efficiency and low power operation to reduce electricity cost, fit within power budgets of energy harvesters, avoid overheating or increase battery lifetime. These requirements can be achieved by lowering the supply voltage of their integrated circuits. The power consumption is reduced as the circuit supply voltage is lowered down towards and below the threshold voltage of the transistors; this comes at the cost of increased propagation delays. There exists an application-dependent sweet-spot for the supply voltage referred to as the minimum energy point, where the energy per operation of the circuit is at its lowest. Voltage scaling in integrated circuits requires considerations during the design phase. Even though down-scaling the supply voltage of integrated circuits is a popular and effective means of achieving these goals, the relative impact that process, voltage, and temperature variations have on the functionality of the circuits also tend to increase with decreasing supply voltage. For memory subsystems within low power and energy-efficient circuits, variations in strength ratios between transistors can lead to an inability to write, inability to read, or inability to hold logic values.

In this work, promising synchronous edge-triggered memory elements have been identified for low voltage operation. A comparative study of 9 data flip-flop circuit architectures at a target supply voltage of 200 mV has been performed. In the study all nine memory cells were sized and laid out using the same approach, setup time, hold time, minimum functional supply voltage, propagation delay, functionality in Monte Carlo simulations, power consumption and energy per cycle have been simulated for each data flip-flop on netlists extracted from the layout. The study revealed that the two most promising memory elements were the Pass Gate and the PowerPC flip-flops. The first had the lowest energy consumption per cycle for operating frequencies below 500 kHz, and for supply voltages below 400 mV, the lowest transistor count and the smallest layout footprint, it was also functional at the lowest supply voltage of 65 mV. The second, the PowerPC flip-flop, had the best power delay product, 26% of that of the worst flip-flop.

Compact near-threshold static random access memory cells have also been investigated. Many Low voltage SRAM cells have been reported in the literature. However, most of them deploy multiple extra transistors as write- and read- assists to the conventional 6 transistor SRAM cell. We bring the loadless 4 transistor cell down to the near-threshold voltage domain by adding a 2 transistor read buffer. The loadless nature of the 4 transistor core acts as a write assist while the 2 transistor read buffer acts as a read assist, ensuring read-ability and write-ability at a supply voltage of 350 mV in a 28 nm fully depleted silicon on insulator technology for a 6 T cell. High sigma simulation results on a 128 kb memory macro of the cell show that it performs on par with state of the art low voltage memory. Three variants of the macro were explored where the peripheral logic consisted of low threshold-voltage, high threshold-voltage devices, and high threshold voltage devices with up-scaled lengths. For the three variants at a supply voltage of 350 mV, the retention power consumption of the 128 kb system was found to range from 1.31 μW to 71.09 μW , the maximum operational frequency was found to lie within 1.87 MHz and 14.97 MHz while the read energy varied from 13.08 to 75.21 fJ/operation/bit. The minimum retention voltage of the loadless SRAM cell was found to be 230 mV in Monte Carlo simulations.

A study has been performed where we balance pull-up and pull-down networks in ultra-low-voltage logic gates to maximize noise margins and create robust low-power minority-3 and Boolean logic gates. To balance the drive strength of the complementary networks in the gates, design-phase gate-length biasing and runtime back-gate biasing was applied. These gates have been combined into ultra-low voltage 16- and 32- bit adders, and measurements were made on a physical prototype. The improvement in energy per 1-bit addition for the 32-bit adder was 37% at a supply voltage of 300 mV compared to the best reported adder in the same technology. For the 16 bit adder, the improvement was 25% at a supply voltage of 250 mV. In addition, the 32-bit adder was shown to be functional down to 125 mV for all ten chip-samples and down to 80 mV by tuning the drive strength of the pull-up networks, with back-gate biasing.

To interface ultra-low-voltage circuit blocks, subthreshold to superthreshold level shifters have been examined. By applying the principle of write assist circuitry in low voltage memory cells to low voltage level shifters, an ultra-low voltage level shifter was created. A physical prototype was implemented and manufactured in 130 nm bulk CMOS to demonstrate the principle. The assist circuit of the level shifter ensures a propagation delay of 21.08 ns, which results in an energy-efficient level shifter consuming 25.9 fJ per conversion when up-converting near-threshold input signals of 300 mV to regular supply voltage signals of 1.2V. Measurement results on ten samples show that the level shifter support up-conversion of logic

signals with mean value down to 31.1 mV. By tuning the balance between the pull-up and pull-down networks the mean value of the lower limit is decreased to 14.3 mV.

Macro-level schemes have been considered to reduce the power consumption of already voltage scaled low voltage memories. Low voltage memory cells are typically asymmetric in nature, which leads to state-dependent pros and cons. We have introduced a scheme that dynamically exploits content-dependencies in asymmetric memory cells where words are conditionally flipped during write operations to reach the more beneficial state for storage. A capacitive logic decision circuit is created to perform the word-flip operation, and the system is exemplified with a 1 kb static random access memory macro of asymmetric low voltage memory cells that have a state-dependent leakage quotient of 23. The memory array and the scheme was implemented and manufactured in a 130 nm bulk process. Measurement results, on a physical prototype of the exemplified memory macro, with the scheme activated, show an improvement of 11.93% in retention power consumption when storing 128 words concatenated from random bits that are uniformly distributed.

By applying voltage scaling on CMOS circuits, we have gained further insight into circuit architectures and schemes that ensures resilient and robust operation while yielding energy efficiency and low power consumption. The results of this work are identifications of promising circuit architectures as well as the introduction of new schematics and approaches to the design of low voltage memories and logic cells.

In memory of Knut Låte, my grandfather, who was forced to terminate his electronics study at NTNU when World War II broke out.

Preface

This thesis is the result of a four-year-long contract with the Department of Electronic Systems (IES) under the Faculty of Information Technology and Electrical Engineering (IE) at the Norwegian University of Science and Technology (NTNU). The work was carried out with Professor Snorre Aunet as the main supervisor. Professor Trond Ytterdal at IES has contributed to the research project as a co-supervisor.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors Snorre Aunet and Trond Ytterdal, for the periodic status meetings throughout the research stay at NTNU. Through these interactions, I have received a large amount of academic advice, as well as useful tips and tricks from their IC design experience. I want to express my gratitude to Professor Trond Ytterdal for being an active system administrator of the UNIX IC-design environment. Trond has also been very helpful in his role as a legal contact with CMP in France. In this role, Trond has provided me with access to the 28 nm FD-SOI design kit, and he has assisted me in acquiring and installing STM's 130nm BiCMOS design kit. Professor Snorre Aunet has, in a similar fashion, aided me in being a legal contact to ARM, from whom we received the register transfer level netlist to the lightweight ARM cortex M0 processor.

Ali Asghar Vatanjou has been a valuable technical office mate, advisor, and colleague to me, for which I am profoundly grateful. Together we have discussed design novelties, cooperated on physical prototyping, and reviewed state-of-the-art low voltage and energy-efficient electronics. I wish Ali the best of luck with his career as an analog designer at Nordic Semiconductor. Speaking of Nordic Semiconductor, I have to give Dr. Carsten Wulff, a big thumbs up for assisting both Ali and me in the fourth quarter of 2015 with his full design experience from an earlier prototype run in 28 nm FD-SOI. Being the guinea pig for new and advanced CMOS technologies is often a painful process. Industrial design experience with components of SAR ADCs as well as modeling experience in VerilogA was acquired during my summer internship with Atmel in 2015, thanks to everyone

who made this possible and to all colleagues I got to know there.

In 2016 I was fortunate enough to spend a month as a visiting researcher at the University of California, San Diego (UCSD). I am thankful to Gert Cauwenberghs and to all other persons that I got to know for inviting me over and providing me with a highly compelling application for my energy-efficient memory designs. A special thanks to Abraham Akinin for taking good care of me socially as well as technically with design kit installations.

Gratitude to Hourieh Atarzadeh Joozdani, Kangqiao Zhao, Harald Garvik, Peng Wang, and Rune Kaald for valuable technical discussions on design tools, concepts, and new ideas.

Last but not least, my supportive friends and family have provided me with the motivation I required to see this research project through. My mother, father, partner, and brother have shown their support through all highs and lows of this roller coaster experience.

Contents

1 Abstract	i
2 Preface	vii
Acronyms	xv
Symbols	xix

Part I Background

3 Introduction	3
3.1 Motivation	3
3.1.1 Sensor Nodes with Energy Harvesting	3
3.1.2 Handheld Devices	5
3.1.3 Data Centers	6
3.1.4 Voltage Scaling in CMOS Circuits	7
3.2 Research Objectives	7
3.3 Contributions	8
3.4 Thesis Structure	9
4 Low Voltage Logic and Memories on Silicon	11
4.1 Time-, Power- and Energy-quantities of Circuits	11
4.1.1 SI Units	11
4.1.2 Low Power Consumption of Circuits	12
4.1.3 Energy Efficiency of Circuits	13
4.2 Voltage Scaling for Low Power Consumption & Energy Efficiency	14
4.2.1 Operating Regions of CMOS Transistors	14
4.2.2 Weak and Moderate Inversion for Low Power Consumption and Energy Efficiency	16

4.2.3	Minimum Energy Point of CMOS Logic and the Activity Factor	16
4.2.4	Minimum Energy Point of CMOS SRAMs	20
4.3	Process, Voltage And Temperature Variations	24
4.3.1	Imperfect Devices in Low Voltage CMOS Circuits	24
4.3.2	Robustness, Resilience & Stability of Circuits	25
4.4	Low Voltage Memories	27
4.4.1	Data Retention in Low Voltage SRAM Cells	27
4.4.2	Read Operations in Low voltage SRAM Cells	28
4.4.3	Write Operations in Low voltage SRAM Cells	29
5	Methods	31
5.1	Research Time Line	31
5.2	CMOS Technologies	32
5.3	Design Software And Measurement Equipment	33
5.4	Manufactured Prototypes	33

Part II Author's Contribution

6	Research Questions	39
A	Extended Comparative Analysis of Flip-Flop Architectures for Sub-threshold Applications in 28 nm FD-SOI.	41
A.1	Abstract	41
A.2	Relevance to the Thesis	41
A.3	Declaration of Authorship	42
A.4	Manuscript	42
B	A Loadless 6T SRAM Cell for Sub-& Near-threshold Operation Implemented in 28 nm FD-SOI CMOS Technology	55
B.1	Abstract	55
B.2	Relevance to the Thesis	56
B.3	Declaration of Authorship	56
B.4	Manuscript	56

C	Ultra-low Voltage and Energy Efficient Adders in 28 nm FDSOI Exploring Poly-biasing for Device Sizing	67
C.1	Abstract	67
C.2	Relevance to the Thesis	67
C.3	Declaration of Authorship	68
C.4	Manuscript	68
D	An Energy Efficient Level Shifter Capable of Logic Conversion From Sub-15 mV to 1.2 V	79
D.1	Abstract	79
D.2	Relevance to the Thesis	79
D.3	Declaration of Authorship	80
D.4	Manuscript	80
E	Benefiting from State Dependencies in Asymmetric SRAM Cells through Conditional Word-flipping	87
E.1	Abstract	87
E.2	Relevance to the Thesis	87
E.3	Declaration of Authorship	88
E.4	Manuscript	88
I	Comparative Analysis of Flip-flop Architectures for Subthreshold Applications in 28nm FDSOI	95
I.1	Abstract	95
I.2	Relevance to the Thesis	95
I.3	Declaration of Authorship	96
I.4	Manuscript	96
II	Ultra-low Voltage Adders in 28 nm FDSOI Exploring Poly-biasing for Device Sizing	103
II.1	Abstract	103
II.2	Relevance to the Thesis	103
II.3	Declaration of Authorship	104
II.4	Manuscript	104

Part III Conclusions & Outlook

5	Concluding Remarks	113
6	Outlook	117

Acronyms

6T	6 Transistor (SRAM).
ADC	Analog to Digital Converter.
ARM	Advanced RISC Machines.
ASIC	Application Specific Integrated Circuit.
BiCMOS	Bipolar Complementary Metal-Oxide-Semiconductor.
BL	Bit Line.
BLR	Bit Line Read.
CMOS	Complementary Metal-Oxide-Semiconductor.
CMP	Circuits Multi-Projets.
CPU	Central Processing Unit.
CR	Cell Ratio in SRAM cells, also known as β ratio.
CUT	Circuit Under Test.
DAC	Digital to Analog Converter.
DCVS	Dfferential Cascode Voltage Switch.
DFF	Data Delay Flip-Flop.
DIBL	Drain Induced Barrier Lowering.
DITS	Drain Induced Threshold Shift.
DOI	Digital Object Identifier.
DRAM	Dynamic random Access Memory.
DVFS	Dynamic Voltage and Frequency Scaling.
DVS	Dynamic Voltage Scaling.
EKV	Enz, Krummenacher, Vittoz - model.
FA	Full Adder.
FBB	Forward Body-Biasing.
FD-SOI	Fully-Depleted Silicon On Insulator.
FF	Fast-NMOS Fast-PMOS (corner).
FIFO	First in, First Out.
FL	Footer Line.
FO4	Fanout of 4.
FS	Fast-NMOS Slow-PMOS (corner).
GPU	Graphics Processing Unit.
HA	Hold Assist.
HS	High Speed.
IC	Integrated Circuit.
IE	Faculty of Information Technology and Electrical Engineering.
IES	Department of Electronic Systems.

INV	Inverter.
INWE	Inverse Narrow Width Effect.
IO	Input/Output.
IOT	Internet Of Things.
LDS	Low-Discrepancy Sequence.
LL	Low Leakage.
LS	Level shifter.
LVT	Low-Threshold-Voltage (transistor).
MC	Monte Carlo (simulation).
MEP	Minimum Energy Point.
MiM	Metal-Insulator-Metal.
MKS	Meter, Kilogram, Second.
MoM	Metal-Oxide-Metal.
MOS	Metal-Oxide-Semiconductor.
MOSFET	Metal-Oxide-Semiconductor Field Effect Transistor.
MUX	Multiplexer.
NAND	Not AND.
NEC	Nippon Electric Company.
NFET	N-type Field Effect Transistor.
NMOS	N-type Metal-Oxide-Semiconductor.
NOR	NOT OR.
NORCAS	Nordic Circuits and Systems Conference.
NTNU	Norwegian University of Science and Technology.
NWE	Narrow Width Effect.
NWELL	N-type Well.
PCB	Printed Circuit Board.
PDF	Probability Density Function.
PDN	Pull-Down Network.
PDP	Power-Delay Product.
PFET	P-type Field Effect Transistor.
PG&A	Pulse generator and analysis (module in VerilogA).
PMOS	P-type Metal-Oxide-Semiconductor.
PMU	Power Management Unit.
PR	Pull up ratio in SRAM cells.
PSUB	P-type Substrate.
PUN	Pull-Up Network.
PVT	Process, Voltage and Temperature.
PWELL	P-type Well.
QFN	Quad-Flat No-leads.
QRC	Quantus Parasitic Extraction.

RA	Read Assist.
RAM	Random Access Memory.
RBB	Reverse Body-Biasing.
RCA	Ripple-Carry Adder.
RDF	Random Dopant Fluctuations.
RF	Radio Frequency.
RFID	Radio-frequency identification.
RISC	Reduced Instruction Set Computing.
RM	Read Margin.
RO	Research Objective.
RQ	Research Question.
RSCE	Reverse Short Channel Effect.
RVT	Regular-Threshold-Voltage (transistor).
SAR	Successive Approximation Register.
SCE	Short Channel Effect.
SEL	Single Event Latch-up.
SEU	Single Event Upset.
SF	Slow-NMOS Fast-PMOS (corner).
SHMAC	Single-ISA Heterogeneous MAny-core Computer.
SI	International System of Units.
SNM	Static Noise Margin.
SNW	Single N-type Well in a P-type Substrate.
SoC	System-on-Chip.
SPW	Single P-type Well in a Deep N-type Well in a P-type Substrate.
SRAM	Static Random Access Memory.
SS	Slow-NMOS Fast-PMOS (corner).
STI	Shallow Trench Isolation.
STM	STMicronics.
TDP	Thermal Design Power.
TT	Typical-NMOS Typical-PMOS (corner).
UCSD	University of California San Diego.
ULP	Ultra-Low Power.
ULV	Ultra-Low Voltage.
UNIX	Uniplexed Information Computing System.
UTBB	Ultra-thin Body and Buried Oxide.
VHDL	Very High Speed Integrated Circuit Hardware Description Language.
VLSI	Very-Large-Scale integration.
VTC	Voltage Transfer Curve.
WA	Write Assist.

WL	Word Line.
WM	Write Margin.
XCP	Cross-Coupled Pair.
XOR	Exclusive OR.

Symbols

α	Activity factor.
β	Ratio of width to length for the transistor (W/L).
C_{BL}	Bit Line to ground capacitance.
C_{BLW}	Write Bit Line to ground capacitance.
$C_{\overline{BLW}}$	Inverse Write Bit Line to ground capacitance.
C_D	Depletion Capacitance.
C_{db}	Drain-Bulk Capacitance.
C_{FL}	Footer Line to ground capacitance.
C_{gs}	Gate-Source Capacitance.
C_{OV}	Overlap Capacitance.
C_{OX}	Oxide Capacitance.
C_{sb}	Source-Bulk Capacitance.
C_{WL}	Word Line to ground capacitance.
C_{WLR}	Read Word Line to ground capacitance.
C_{WLW}	Write Word Line to ground capacitance.
E	Energy.
$E_{dynamic}$	Dynamic energy consumption.
E_{static}	Static energy consumption.
η	DIBL coefficient.
E_{total}	Total energy consumption.
f	Frequency.
GND	Ground.
I_{static}	Static current drawn.
I_{DS}	Drain to source current.
k	Boltzmann's constant.
L	Channel Length.
Ω	Resistance.
P	Power.
$P_{dynamic}$	Dynamic power consumption.
P_{static}	Static power consumption.
P_{total}	Total power consumption.
q	The amount of charge in 1 electron (elementary charge).
μ	Mean Micro Mobility.
σ	Standard Deviation Sigma.
T	Time.
T	Temperature in kelvin.

V_{DD}	Supply voltage.
V_{DDH}	Supply Voltage High.
V_{DDL}	Supply Voltage Low.
V_{DS}	Drain to source voltage.
V_{GS}	Gate to source voltage.
V_T	Thermal voltage.
V_{th}	Threshold voltage of a transistor.
W	Channel Width.

PART I

BACKGROUND

"Innovation is the only thing that will keep the company vital. If we stop innovating, it won't be long before we become the equivalent of the buggy whip industry."

-Gordon Moore

3.1 Motivation

For electronics, performance requirements are application dependent. In applications such as data centers and mining of cryptocurrency, operating speed is critical, and low power consumption is essential in satisfying requirements for thermal design power (TDP). Energy-efficiency these high-performance computers is also critical since it aids in reducing the cost of electricity and therefore the profitability of the operation. In off-grid applications such as sensor nodes, the sensors remain dormant for the majority of the time and wakes up only to sense or to send some sensed data and then go back to sleep again. In this case, low power consumption is crucial to satisfy the power budget of energy harvesting methods. Energy efficiency is here central in increasing the amount of work that can be executed before the node goes back to sleep.

Between the two extremes of high-performance computing and sensor nodes, there are handheld devices such as cellphones, where both performance and low power consumption are metrics required to give a responsive user experience and a prolonged battery lifetime. Battery dependent applications require energy-efficiency to maximize the number of operations per battery charge. For all three categories, there is a demand for low power and energy-efficient circuits.

3.1.1 Sensor Nodes with Energy Harvesting

Sensor nodes without energy harvesters are powered from a finite energy source and will require periodic replacement of the energy storage if a continuation of the sensing task is desired [1]. By adding an energy harvester to the sensing system, a prolonged or even perpetual operating life is achieved. Sensor nodes powered by energy harvesters are commonly divided into two categories, systems without

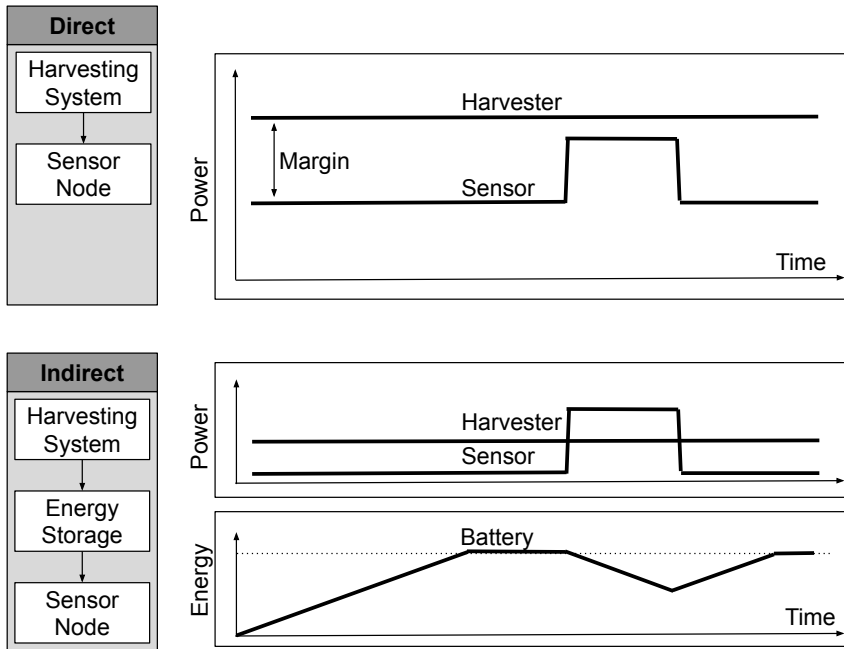


Figure 3.1: Energy harvesting in sensor systems without and with energy storage [1].

energy storage known as "Harvest-Use" systems, and systems with energy storage as shown known as "Harvest-Store-Use" systems both are shown in Fig. 3.1. The power consumption of the Harvest-Use sensor system, is limited to the output power of the energy harvester at all times, whereas the Harvest-Store-Use system can operate at a higher power consumption for as long as the stored energy lasts. Both systems are limited by power budgets that are given by either just the harvester or the combination of the harvester and the capacity of the battery.

For sensor nodes in the internet of things (IoT) [2] with energy-harvesting deployed, the power budget is found, in [3], to be dependent on the node size and the harvesting type. For thermo-electric (on-body patch), millimeter-sized photovoltaic (indoor), and airflow (indoor), harvesters can perpetually sustain a power consumption of several μW . To achieve a perpetual power budget of several tens of μW s for a millimeter-sized harvesting device, more abundant energy sources, such as photovoltaic (outdoor), thermo-electric (industrial machines) and body vibration (walking) needs to be harvested. For any of the above-mentioned types of energy harvesting, the power budget is proportional to the size of the node allowing perpetual life for power budgets in the hundreds of μW .

For sleeping sensor nodes with PMUs that are awake when the rest of the circuit is in a deep sleep mode or for always-on sensor nodes, the power consumption needs to fit in the power budget of the system. For the first example, the power consumption of the rest of the circuit is not that important as long as the energy it consumes when awake is less than the energy capacity that is accumulated in between sleep cycles. The circuit will be able to perpetually operate if the power consumed by the PMU is less than the power harvested. For the latter example, the average power dissipation of the entire sensor node needs to fit in the budget for the system to operate perpetually. In both examples, the lower the power consumption of the circuit, the more flexible is the choice of harvesting approach, and the smaller the required node size.

3.1.2 Handheld Devices

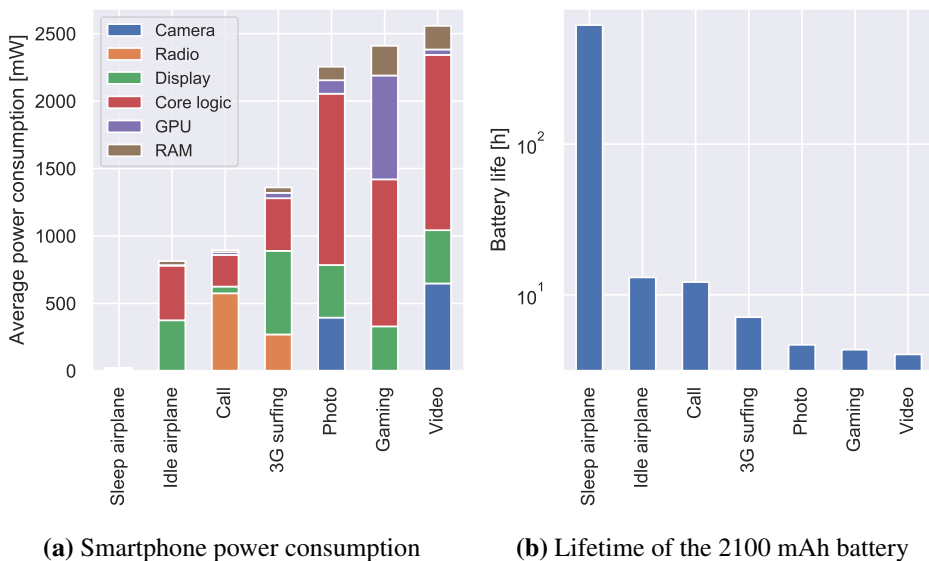


Figure 3.2: Average fractions of power consumption in a Samsung Galaxy S3 [4].

A study of the detailed power consumption of a smartphone is performed in [5], in [6] a similar study was performed on a Samsung Galaxy S3 where the detailed power consumption was measured for the circuit blocks, Camera, Radio, Display, GPU, external RAM, and Core Logic when using the cellphone for different applications. The results of the measurements are summarized in Fig. 3.2. It is evident from the stacked bar plot containing the power distribution, that the major contributor of the total power consumption is the core logic, except for the call action and the 3G web browsing where the radio and the display power dominates. By reducing the power consumption of the individual components, the overall bat-

tery lifetime is increased for a fixed battery size.

3.1.3 Data Centers

Moore's law [7], Dennard scaling[8], and advances in circuits and architectures has driven computer design in the last four decades and has exponentially increased the number of calculations per second in computers. Multi-core designs have lately been an effective means to continue the increase in performance in parallelizable applications. Historically, With each technology node comes an exponential increase in operating frequency and a slight decrease in supply voltage and average nodal capacitance. This *device scaling* leads to lower area and power consumption. The increasing complexity of the micro-architectures demand exponentially larger die sizes per operation, and thus an exponential increase in performance, this is referred to as *core scaling*. The multi-core trend is referred to as *multi-core scaling* and multiplies the circuit size with the number of cores. The product of these scaling effects is a multi-core speedup over time, which is limited by the thermal design power of the dice. That is, fractions of the silicon die cannot be powered on at the nominal operating voltage due to the constraint of overheating [9]. The dark silicon gap could be filled with advances in heterogeneous multi-core designs consisting of a mix of low power and high power cores. Solving this problem will increase the performance per silicon area of high-performance computing applications such as data centers. The need for efficient data centers has never been

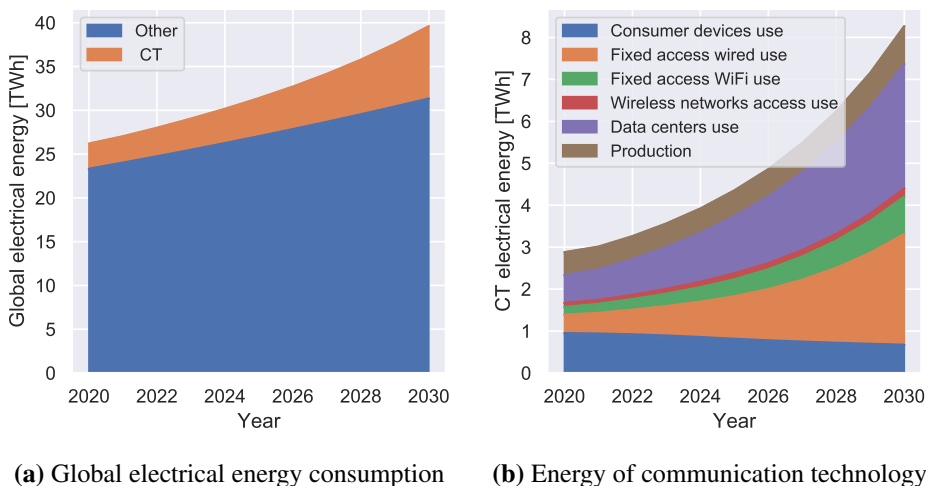


Figure 3.3: The expected fraction of global energy consumption consumed by communication technology and data centers in the next decade [4].

higher. The amount of data on the internet is exponentially increasing as computers

have become globally accessible, smartphones have gotten higher resolution camera sensors, and internet speeds have been increasing. In 2017 the yearly internet traffic passed the ZB mark, and the growth rate of the global of data is expected to increase in the future [10]. If the global, average, energy consumption per operation on the data remains constant, the energy demand for data centers will grow exponentially with the internet traffic, motivating energy efficient circuits. In [4], it is expected that the global electricity usage of data centers will increase with 450% from 2020 to 2030 to 7.5% of the worlds' global electricity usage. CT electricity usage could, in the worst case, stand for 23% of the globally released greenhouse gas emissions in 2030. The two area plots in figure Fig. 3.3 show the impact that data centers are expected to have in this decade. The majority of the energy consumed in data centers, around 40%, goes to cooling. The design of both energy-efficient and low power circuits is thus an important task towards directly and indirectly reducing the overall energy consumption of data centers.

3.1.4 Voltage Scaling in CMOS Circuits

For data centers, sensor nodes, and handheld devices, low power, and energy-efficient CMOS circuit design are vital enablers for future improvement. Supply voltage scaling is the most direct and dramatic knob for these metrics [11]. It can be used for adjusting propagation delay, switching current, and leakage current in static CMOS logic. Down-scaling supply voltage of CMOS circuits reduces the dynamic and static currents, which lowers dynamic and static power consumption at the cost of increased propagation delay for a fixed gate capacitance. Low power circuits are welcome, however, energy-efficiency is also often considered an even more important metric, and low power circuits are not necessarily energy-efficient. Energy-efficiency is often measured in energy per operation, and there are two main methods in achieving this, one can either reduce the time it takes to complete one unit of work, or one can reduce the power consumption at any time. These two methods are often complementary, and voltage reduction is only used to achieve the latter at the cost of the first.

3.2 Research Objectives

The underlying goal of this thesis is to study, develop, and prototype new concepts in low power and energy-efficient CMOS circuits. The research presented here has a special emphasis on low voltage memories such as data flip-flops (DFF)s and static random access memories (SRAM)s, but also on low voltage static CMOS logic.

Three research objectives have been identified:

- Innovate in low power circuit design to mitigate the Dark Silicon Effect.
- Find energy-efficient circuit design techniques to extend battery lifetime and cut energy consumption in battery driven devices and computationally intensive applications.
- Improve the stability of CMOS circuits at low supply voltages to achieve the aforementioned objectives, but also to widen the frequency and voltage solution space for system flexibility.

3.3 Contributions

This dissertation is compiled by a collection of already published research papers in the field of low voltage CMOS memories and logic. These publications are listed below in chronological order and grouped by primary and supplementary articles. Publications A, B, C, D and E are published journal articles, and the last two papers I and II, are published conference papers directly linked to paper A and C respectively.

- **Article A** (Published): E. Lâte, 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
- **Article B** (Published): E. Lâte, T. Ytterdal, and S. Aunet, “A loadless 6t sram cell for sub- & near-threshold operation implemented in 28 nm fd-soi cmos technology,” *Integration*, vol. 63, pp. 56–63, 2018
- **Article C** (Published): A. A. Vatanjou, E. Lâte, T. Ytterdal, and S. Aunet, “Ultra-low voltage and energy efficient adders in 28 nm fdsoi exploring poly-biasing for device sizing,” *Microprocessors and Microsystems*, vol. 56, pp. 92–100, 2018
- **Article D** (Published): E. Lâte, T. Ytterdal, and S. Aunet, “An energy efficient level shifter capable of logic conversion from sub-15 mv to 1.2 v,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. X, pp. Y–Z, 2020
- **Article E** (Published): E. Lâte, T. Ytterdal, and S. Aunet, “Benefiting from state dependencies in asymmetric sram cells through conditional word-flipping,” *An international journal*, vol. X, pp. Y–Z, Submitted

- Supporting paper I) (Published): E. Låte, A. A. Vatanjou, T. Ytterdal, and S. Aunet, “Comparative analysis of flip-flop architectures for subthreshold applications in 28nm fdsoi,” in *Nordic Circuits and Systems Conference (NORCAS): NORCHIP & International Symposium on System-on-Chip (SoC), 2015*. IEEE, 2015, pp. 1–4
- Supporting paper II) (Published): A. A. Vatanjou, E. Late, T. Ytterdal, and S. Aunet, “Ultra-low voltage adders in 28 nm fdsoi exploring poly-biasing for device sizing,” in *Nordic Circuits and Systems Conference (NORCAS), 2016 IEEE*. IEEE, 2016, pp. 1–4

3.4 Thesis Structure

The thesis is divided in to three parts:

Part I, "Background", describes the circuit design theory of CMOS memories and the methodology used throughout this work. The part consists of this introductory chapter as well as two other ones. Chapter 2 follows with details on the theoretical basis of the research that has been carried out. Chapter 3 contains the research method as well as a chronological overview of the conducted research. Papers, articles, prototypes, and other important milestones towards answering the above three research objectives, are all presented on a single timeline in the third chapter.

Part II, "Publications", describes the results from the four-year-long research period in the form of 5 published journal articles. Chapter 4 breaks the research objectives down to research questions and ties these to each publication. Then follows 7 alphabetised chapters, one for each publication. Subsections in these chapters include paper abstracts, relevance to the thesis, declarations of authorship, and the embedded manuscripts of the publications.

Part III, the chapters "Conclusion" and "Outlook", includes concluding remarks on the research findings and a discussion of work that remains to be investigated in the future.

Low Voltage Logic and Memories on Silicon

"We crave for new sensations but soon become indifferent to them. The wonders of yesterday are today common occurrences"

-Nicola Tesla

This chapter briefly presents an overview of the fundamental principles of energy-efficient and low power CMOS used in the articles of this compilation thesis. Since each article contains literature references, the information covered here will only cover the general ideas of the field that are not mentioned in detail in the articles.

4.1 Time-, Power- and Energy-quantities of Circuits

4.1.1 SI Units

The relationship between delay, power consumption, and energy consumption is often misunderstood, especially the difference between the latter two. All three are central metrics in digital circuit design and are used extensively throughout this thesis. Their quantity definitions are listed here according to the international system of units (SI) [19]:

- The time quantity [T] is measured in the SI base unit seconds [s], and is defined as: "The duration of 9192631770 periods of the radiation corresponding to the transition between the two hyper-fine levels of the ground state of the caesium-133 atom. It is defined by taking the fixed numerical value of the cesium frequency $\Delta\nu_{Cs}$, the unperturbed ground-state hyper-fine transition frequency of the cesium 133 atom, to be 9 192 631 770 when expressed in the unit Hz, which is equal to s^{-1} ."
- The energy quantity [E] is measured in the derived SI unit joule [$J = kgm^2s^{-2}$] and is defined to be: "The work done when the point of application of 1 MKS¹ unit of force [newton] moves a distance of 1 meter in the direction of the force. "

¹ MKS (meter, kilogram, second) system

- The power quantity [P] is measured in the derived SI unit watt [$W = J/s = kgm^2s^{-3}$] and is defined to be: "The watt is the power which in one second gives rise to energy of 1 joule."

There is a clear relationship between the three SI units in EQ.4.1.

$$P = \frac{E}{T} [kgm^2s^{-3}] \quad (4.1)$$

In circuit design it is more reasonable to express power, P, in terms of voltage and current. We have the following definitions of the electrical units [20]:

- "Ampere (unit of electric current). - The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} MKS unit of force [newton] per meter of length" [A].
- "Volt (unit of potential difference and of electromotive force). - The volt is the potential difference between two points of a conducting wire carrying a constant current of 1 ampere, when the power dissipated between these points is equal to 1 watt" [$V = kgm^2s^{-3}A^{-1} = \frac{W}{A}$].
- "Ohm (unit of electric resistance). - The ohm is the electric resistance between two points of a conductor when a constant potential difference of 1 volt, applied to these points, produces in the conductor a current of 1 ampere, the conductor not being the seat of any electromotive force" [$\Omega = kgm^2s^{-3}A^{-2}$].
- "Coulomb (unit of quantity of electricity). - The coulomb is the quantity of electricity carried in 1 second by a current of 1 ampere" [$C = As$].

The definition of the derived SI unit voltage and the SI base unit of current gives us the equation for electric power in EQ.4.2.

$$P = \frac{E}{T} = VI = VI [kgm^2s^{-3}] \quad (4.2)$$

4.1.2 Low Power Consumption of Circuits

Low power logic circuits have been a popular research topic for many decades [21], one of the early applications were low power binary counters for wrist watches [22] to ensure more than one year of battery lifetime with an 1.35V mercury button

battery as a consequence of volume constraints. The power consumption had to be reduced below $10\mu W$ [23]. Since then more complex circuits such as processors and memories have also been target to power minimization for reasons mentioned in the introduction chapter.

The total power consumption, P_{total} , of circuits is made up of a dynamic component, $P_{dynamic}$ and a static component P_{static} , and is shown in EQ. 4.3.

$$P_{total} = P_{static} + P_{dynamic} \quad (4.3)$$

Using EQ. 4.2, denoting the supply voltage of the circuit under investigation as V_{DD} , and the idle current of the circuit when no activity is performed as I_{static} , we have the static power consumption from EQ. 4.4. When a fraction of the circuit, equal to the activity factor (α), is switching with a frequency (f), from ground (GND) to the supply voltage (V_{DD}), and a total capacitance (C_{total}), the total dynamic power consumption of the circuit is given in EQ. 4.5 [24][25].

$$P_{static} = V_{DD}I_{static} \quad (4.4)$$

$$P_{dynamic} = \alpha f C_{total} V_{DD}^2 \quad (4.5)$$

From EQ. 4.3, EQ. 4.4 and EQ. 4.5, we can directly see the impact of supply voltage scaling in low power circuits.

4.1.3 Energy Efficiency of Circuits

Low power circuits are not necessarily energy efficient. Energy efficiency is measured in energy per operation [26]. The total energy of a circuit can also be expressed as a sum of a static element and a dynamic element, EQ. 4.6.

$$E_{total} = E_{static} + E_{dynamic} \quad (4.6)$$

If a low power circuit has a high propagation delay, its operating frequency has to be low to ensure functionality. The amount of energy it takes to complete an operation will then be dominated by the integration of its' static power consumption. The static energy consumption per cycle of a circuit is expressed in EQ. 4.7.

$$E_{static} = \frac{1}{f} V_{DD} I_{static} \quad (4.7)$$

Similarly, the dynamic energy component can be expressed as the dynamic power component of the total power consumption integrated over the time of a single clock cycle, as seen in EQ. 4.8

$$E_{dynamic} = \alpha \frac{1}{2} C_{total} V_{DD}^2 \quad (4.8)$$

Generally speaking according to 4.7, and assuming a constant I_{static} , the static energy consumption of a circuit is proportional to V_{DD} . As we will see in the following section however, I_{static} is dependent on V_{DD} as well. The dynamic power

consumption has a direct quadratic dependency on V_{DD} . Both these components show a clear benefit of scaling down V_{DD} to achieve energy efficiency from the given equations.

Since the static component is inversely proportional to the operating frequency, and the operating frequency is reduced as V_{DD} is reduced, the static component becomes more significant at low V_{DD} . The total energy consumption per operation, thus, has a minimum energy point (MEP), which is located somewhere on the V_{DD} axis dependent on the size of the leakage current of the circuit [27]. Targeting this MEP makes sense in circuit design when energy efficiency is desired without the constraints of power budgets from energy harvesting or thermal design power.

4.2 Voltage Scaling for Low Power Consumption & Energy Efficiency

4.2.1 Operating Regions of CMOS Transistors

MOSFET current characteristics are central throughout this thesis. There are three regions of operation related to the channel inversion. As the gate to source voltage (V_{GS}) increases, the electric field acts through the oxide and creates a layer of charge between the drain and source terminal, this layer of charge provides the channel that current can flow through. These three regions of operation are related to the threshold voltage of the transistor (V_{th}) and they are the weak inversion, also known as the deep subthreshold region where the $V_{GS} \ll V_{th}$. There is also the moderate inversion region, also known as the near-threshold region, where $V_{GS} \approx V_{th}$. And the strong inversion region also known as the upper superthreshold region where $V_{GS} \gg V_{th}$.

The three regions of operation, weak inversion, moderate inversion, and strong inversion are shown in Fig. 4.1. Moderate inversion is typically ± 100 mV from the V_{th} of the transistor. The device is in the subthreshold region when V_{GS} is below V_{th} , and the device is in the superthreshold region when V_{GS} is above the threshold voltage.

The analytical expression for the drain current in the subthreshold region is highly relevant throughout this thesis. It is given in EQ. 4.9[25][28][29].

$$I_{DS} = \mu_0 C_{ox} \frac{W}{L} (n - 1) V_T^2 e^{\frac{V_{GS} - V_{th}}{n V_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}} \right) \quad (4.9)$$

Here, μ_0 is the zero bias mobility, C_{ox} is the oxide capacitance, W and L is the effective width and length of the device respectively. The slope factor is denoted as n which is given by $(1 + \frac{C_d}{C_{ox}})$ for a depletion layer capacitance C_d . V_T is the thermal voltage $\frac{kT}{q}$, given the Boltzmann's constant (k) the temperature in kelvin

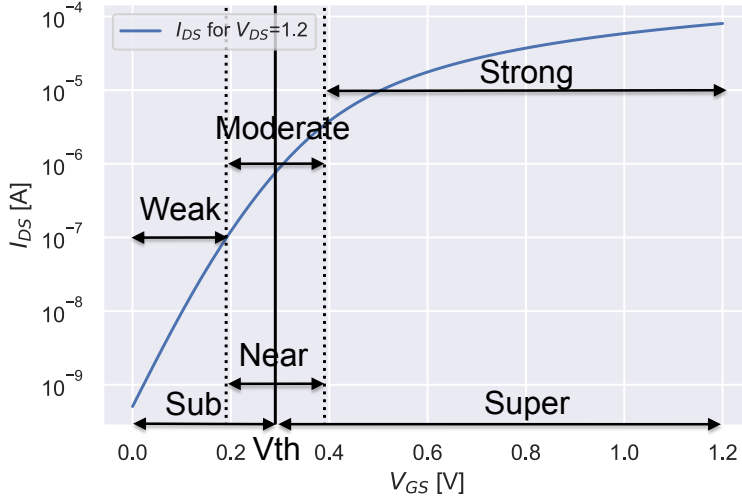


Figure 4.1: I_{DS} vs V_{GS} for a minimum sized low leakage device in 130 nm bulk CMOS where $V_{DS}=1.2V$.

(T) and the charge of an electron (q). EQ. 4.9 includes the most significant relationships for the subthreshold drain current that are used throughout this thesis. The expression can also be modified to include the reduction in threshold voltage as V_{DS} increases, i.e., the short channel effect called Drain Induced Barrier Lowering (DIBL) [29][30] as shown in EQ. 4.10. Where the positive length dependent DIBL coefficient η [31] is multiplied by the drain to source voltage.

$$I_{DS} = \mu_0 C_{ox} \frac{W}{L} (n-1) V_T^2 e^{\frac{V_{GS} - V_{th} + \eta V_{DS}}{n V_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}}\right) \quad (4.10)$$

From EQ.4.10, the subthreshold leakage current for an NMOS transistor is where $V_{DS}=V_{DD}$ and $V_{GS} = 0$. The resulting expression is given in EQ. 4.11 and shows the leakage current's dependency on the supply voltage as we discussed for EQ. 4.7.

$$I_{static} = \mu_0 C_{ox} \frac{W}{L} (n-1) V_T^2 e^{-\frac{V_{th} + \eta V_{DD}}{n V_T}} \left(1 - e^{-\frac{V_{DD}}{V_T}}\right) \quad (4.11)$$

Modeling the drain current in the near-threshold region is more complex and discrete. EKV is a model that is continuous and differentiable over all regions of operation. It was proposed in 1995 [28]. It is not repeated here, however the moderate inversion region of operation has been shown to be energy-efficient for circuits used in applications with low activity factors such as SRAMs [32].

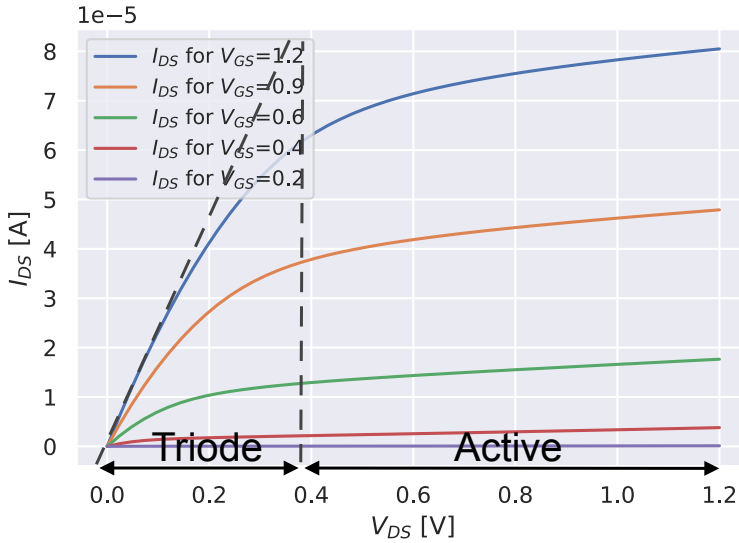


Figure 4.2: I_{DS} vs V_{DS} for a minimum sized low leakage device in 130 nm bulk CMOS for different values of V_{GS} .

The triode region and the saturation region are shown in Fig. 4.2. If the device is in strong inversion, the device is in the saturation region when the drain to source voltage is larger than the effective voltage ($V_{DS} \geq V_{GS} - V_{th}$). Furthermore, it is in the triode region if V_{DS} is below the effective voltage ($V_{DS} < V_{GS} - V_{th}$). If the MOSFET is in the weak inversion region, however, the saturation voltage is approximately independent of V_{GS} and can be approximated by $3V_T$.

4.2.2 Weak and Moderate Inversion for Low Power Consumption and Energy Efficiency

CMOS digital circuits operating in the Ultra-low-voltage domain were introduced in the late 1960's early 1970's [24][33]. Many research groups have since shown that operating circuits in the sub- and near-threshold regions yields low power and energy-efficient integrated circuits.

4.2.3 Minimum Energy Point of CMOS Logic and the Activity Factor

The benefits of voltage scaling in CMOS logic and the dependency on the activity factor are demonstrated here, in this subsection, with a circuit example. A self-timed NAND-based version of the [34] variable activity factor circuit is implemented in 130nm CMOS with the low leakage standard library. The circuit consists of 10 branches of 11 inverter-coupled NAND gates where one of the branches are is a ring oscillator driving the other branches if they are enabled. By disabling

a branch, the activity factor is reduced by 10%. The circuit is here used in V_{DD} sweep simulations to find the contributing components of the power- and energy-consumption of a circuit with a stage depth equivalent to a processor pipeline.

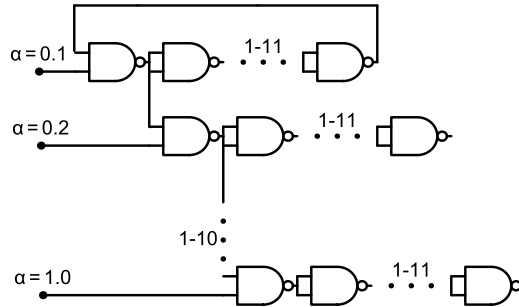


Figure 4.3: Schematic for the circuit with variable activity factor implemented in 130 nm.

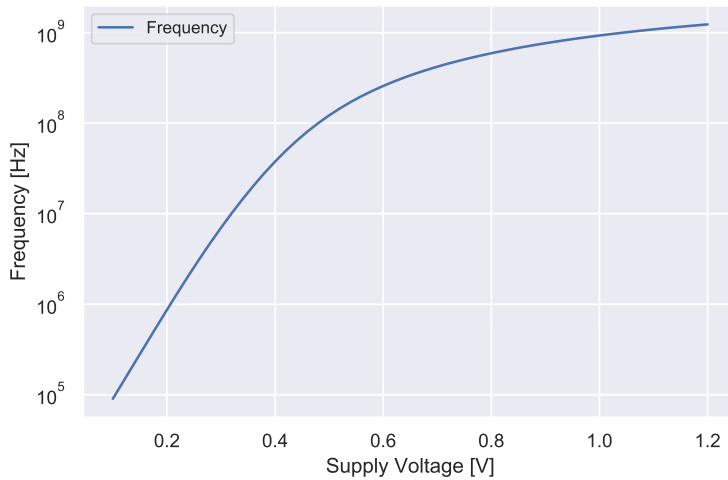


Figure 4.4: Operating frequency vs V_{DD} for the self-timed variable activity factor circuit.

The self-timed nature of the variable activity factor circuit is shown in Fig. 4.4. As the supply voltage is reduced, the propagation delay of each stage in the ring oscillator increases, resulting in a slower operating frequency.

The total power consumption of a logic circuit with an activity factor of 10% is shown in Fig. 4.5. The power consumption is broken down into a leakage component and a switch component. Where the leakage component is the power consumption of the circuit at an activity factor of 0, and the dynamic component of

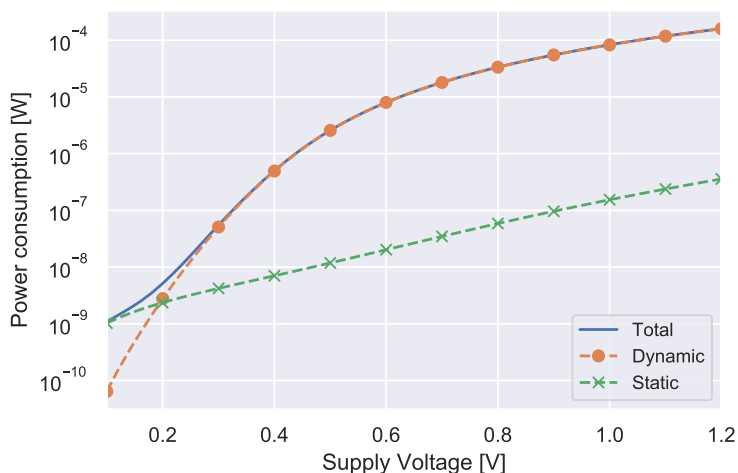


Figure 4.5: The total power consumption and its components vs V_{DD} for the simulated variable activity factor circuit with an activity factor of 10%.

the power consumption is the total power consumed at the listed activity factor minus the leakage component. The leakage component of the power consumption is log-linear, whereas the switching component is roughly quadratically dependent on V_{DD} according to the theoretical model.

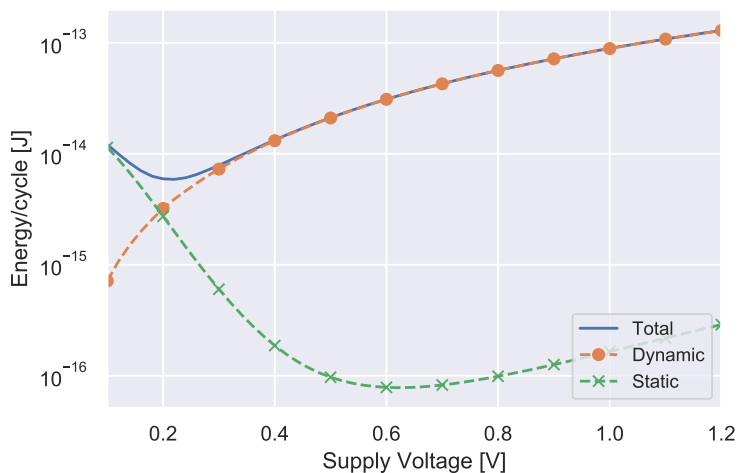


Figure 4.6: The total energy consumption and its components vs V_{DD} for the simulated 130 nm variable activity factor circuit with an activity factor of 10%.

The equation, breaking down energy consumption into a static and a dynamic com-

ponent, states that the dynamic part is strictly increasing with increasing V_{DD} . The simulation result of Fig. 4.6 confirms this where the switch line is exponentially dependent on the supply voltage. As the supply voltage is scaled down, the operating frequency decreases as we saw in Fig. 4.4, the slope of the operating frequency becomes steeper as the voltage is reduced. In Fig. 4.6, the static component of the energy consumption is increasing below 600 mV as the supply voltage is reduced. This is because there is an increasing gap between the leakage current and the delay, the leakage current decreases slower than the delay increases as the supply voltage is scaled down. The static leakage power is integrated over an increasing cycle period of the circuit. The dynamic switching energy consumption of the model is not dependent on the frequency since it is merely the energy that is dissipated to ground when discharging nodes that are switching. It does not matter how fast the node is discharged since there is a finite amount of energy stored in the nodal capacitance. Since the leakage energy becomes significant as the supply voltage is reduced, and since the switching energy becomes insignificant with the decreasing supply voltage, we find a sweet-spot in between the two, which is known as the minimum energy point (MEP). For some circuits, there are no MEPs since the balance between the two components of energy is not even; this is especially true for circuits with high activity factors and low leakage currents.

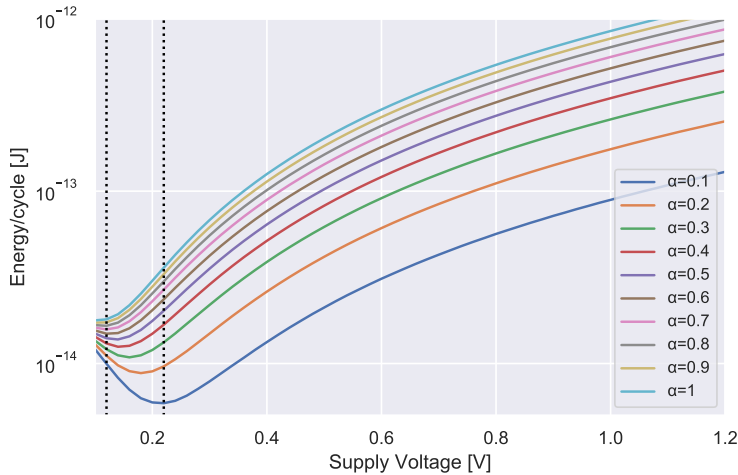


Figure 4.7: Total energy per cycle for the enable-controlled, variable activity factor circuit in Fig. 4.3. Each logic path in the circuit is cumulatively enabled to get the energy/cycle for the listed activity factors. The enabled paths are also operating at their maximum frequency, which is supply voltage dependent.

It can be seen from Fig. 4.7 that the minimum energy point, V_{DD} -wise, is skewed towards the near-threshold region as the activity factor is decreased. This is due to

the balance between integrated leakage current and the delay-independent switching energy. For an activity factor greater than 0.9, the minimum energy point is skewed below 100mV. Please note that this figure is only meant to illustrate the relationship between activity factors for CMOS logic and the minimum energy point, the swept activity factors are highly application specific.

4.2.4 Minimum Energy Point of CMOS SRAMs

Similar to the variable activity factor circuit shown in the previous section, arrays of SRAM cells, also inherit the same trade-off between leakage energy and switching energy. There is, however, a significant difference between the two circuit types. Since words in SRAM blocks share bit-lines (BL), the access frequency of each SRAM word is reduced by a factor of $1/\text{bits}/\text{BL}$. The overall activity factor of the SRAM block is therefore reduced by the same ratio.

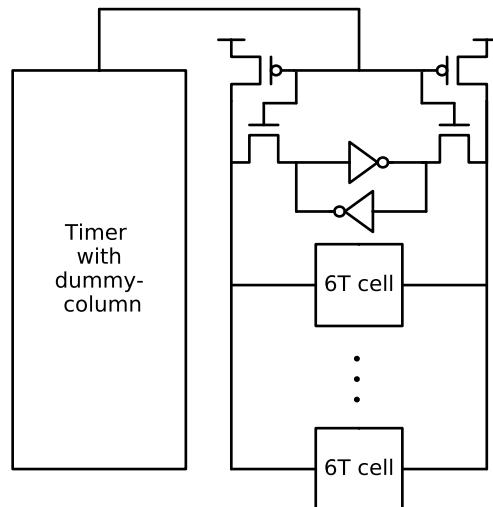


Figure 4.8: Schematic for the SRAM circuit implemented in 130 nm for multiple column heights.

To illustrate the impact that the number of bits per BL has on in the overall power and energy consumption of SRAMs, we modeled 6 self-timed SRAM columns of 16,32,64,128,256 and 512 bits per BL with identical 6-T SRAM cells of low leakage devices 130 nm technology. The schematic of the simulated circuit is shown in Fig. 4.8. The simulation setup in the figure has a precharge circuit that is active in the opposite phase of the pass-transistors of the SRAM cells in the activated read-out row. The simulations were performed on extracted cell-layouts to accurately

represent the significant BL parasitic capacitances in the read operations. For the simplified experiment, simple inverters were used as sense amplifiers. Spectre simulations were performed, sweeping the supply voltage of the CUTs. The simulated energy includes the energy consumed in precharging the BLs

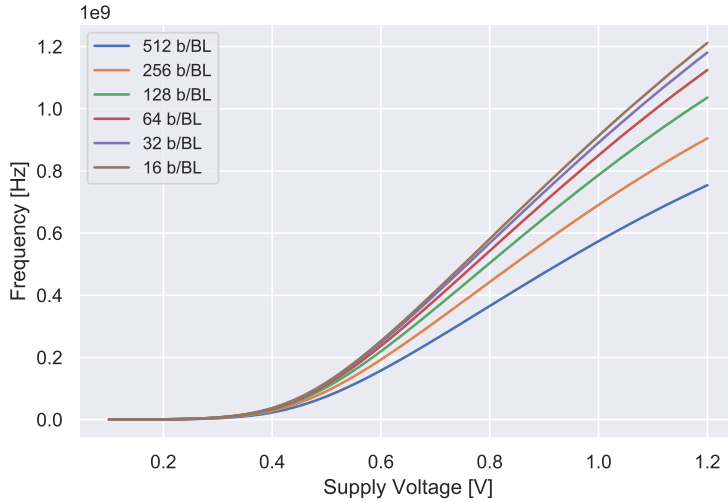


Figure 4.9: Operating frequency vs V_{DD} for the self-timed SRAM columns.

The timers of the 6 CUTs are self-timed based on a NAND ring oscillator like the one in the variable activity factor circuit. To adjust for the increased bit-line evaluation time of each column height, and to make the circuit self-timed in supply voltage scaling, a dummy column identical to the one to the right in the figure replaced one of the nodes in the ring oscillator. The result is that frequency is different for each CUT due to the column height and the identical SRAM cell across each CUT. One SRAM cell will need roughly double the time to pull-down an SRAM column with twice as many bits. The self-timed nature of the SRAM circuits are shown in Fig. 4.9, the frequency of the timer decreases with increasing bits/BL. Similar to the variable activity factor circuit, as the supply voltage is reduced, the propagation delay of each stage in the ring oscillator increases resulting in a slower operating frequency.

Fig. 4.10 shows that the leakage energy increases as V_{DD} is reduced below the threshold voltage of the transistor. The more bits in the column, the higher the leakage energy during a read operation, mainly since more cells are contributing with leakage currents while the cell is performing the write operation, but also since the read operation itself takes longer due to the BL capacitance. The dynamic energy per read operation per bit shown to increase with V_{DD} in roughly a

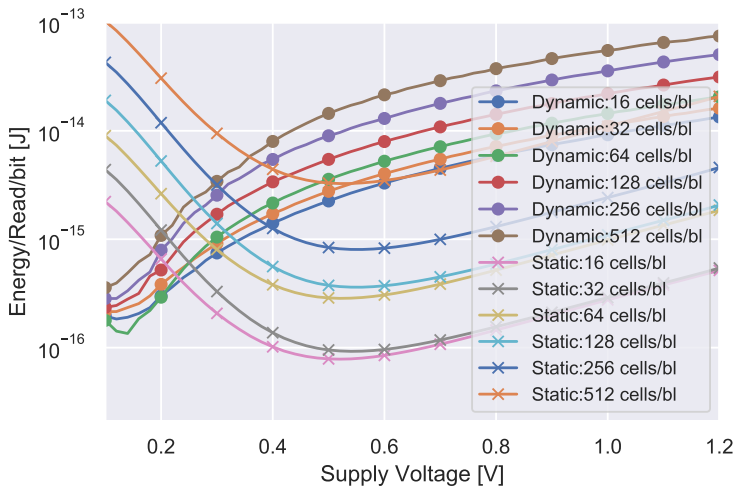


Figure 4.10: Dynamic read energy and leakage energy per operation per bit vs V_{DD} for each column height.

quadratic manner, fitting the theoretical model.

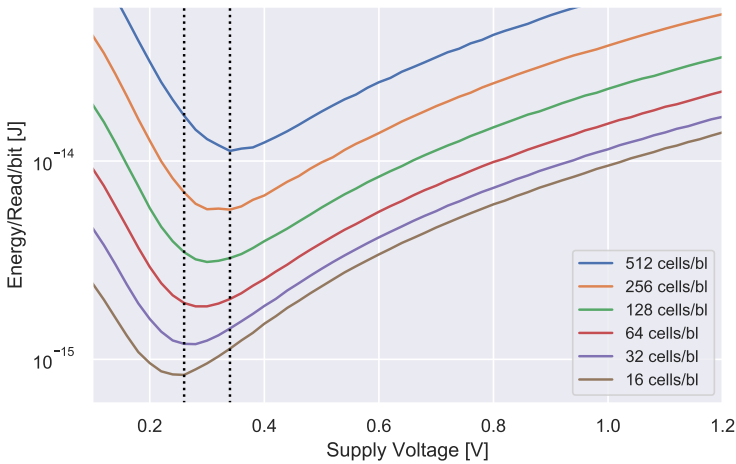


Figure 4.11: Total energy per bit-wise read operation vs V_{DD} for each column height.

The minimum energy point is shifted to the right in Fig. 4.11 as more bits are added to the BL. The behavior is analogous to that of the activity factor that we discussed in the previous section.

Fig. 4.12 shows that the leakage power of all SRAM columns increase with the

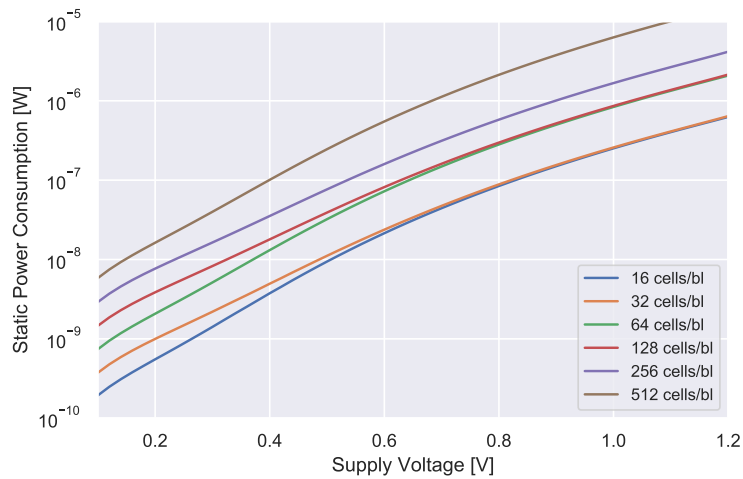


Figure 4.12: Static leakage power vs V_{DD} for each column height.

supply voltage in a close-to log-linear fashion.

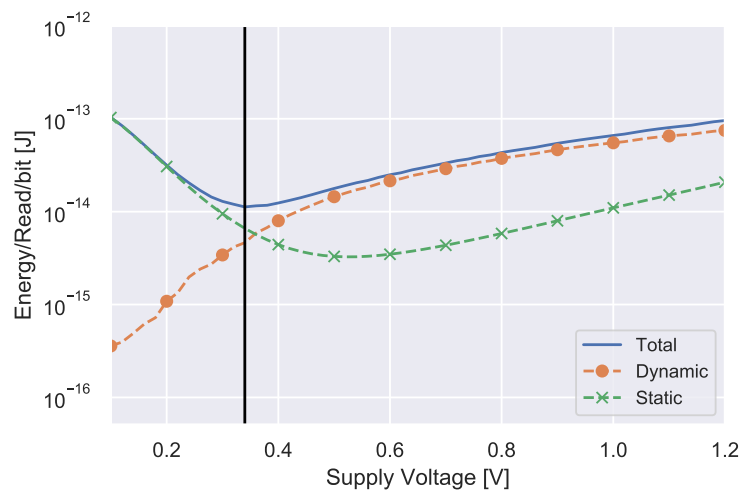


Figure 4.13: MEP and the energy components for the SRAM column with 512 b/BL.

Fig. 4.13 shows the read energy and its components for the simulated SRAM column of 512 cells/BL. The supply voltage that results in the minimum read energy per bit is, in this case, located in the near-threshold domain where $V_{DD} = 300$ mV.

4.3 Process, Voltage And Temperature Variations

According to the equations for drain current across all regions, the current through a MOSFET transistor is dependent on the threshold voltage of the transistor, supply voltages, and temperature (PVT). Variations in any of these parameters result in variations in drain current and internal node voltages, which again may result in incorrect behavior if the variations are significant enough. PVT variations that are likely to occur should be taken into account in the design phase of CMOS circuits. Due to low margins and exponential dependencies on voltages in circuits that operate in the Subthreshold and near-threshold region, PVT variations is a significant challenge to overcome to enable commercial ultra-low voltage circuits.

4.3.1 Imperfect Devices in Low Voltage CMOS Circuits

When manufacturing MOSFETs, process variation affects each device in a Gaussian fashion. Imperfect devices with random variations on the device length and width cause deviations in the threshold voltage (V_{th}) of each device. The standard deviation of the threshold voltage, $\sigma(V_{th})$, can be expressed analytically according to Pelgrom's equation [35][36] shown in EQ. 4.12.

$$\sigma(V_{th}) = \frac{c}{\sqrt{WL}} \quad (4.12)$$

Where c is a constant given a process technology, W is the channel width, L is the channel length.

Variations in threshold voltage imply variations in drain current, which are significant in the subthreshold region as the drain current is exponentially dependent on the threshold voltage, as seen in EQ. 4.9.

To illustrate the variation in drain current from process and mismatch variations, Monte Carlo (MC) simulations have been performed on a minimum sized NMOS transistor in the on-state and an NMOS transistor in the off-state. It can be seen from the sub-figures of Fig. 4.14 that I_{OFF} is log-normally distributed in process and mismatch variations for all values of V_{DD} . The I_{ON} current is normally distributed down to the threshold voltage of the transistor, once below the threshold voltage, the variation in drain current takes a log-normal distribution. Since the off current is log-normal for all values of V_{DD} , the ration I_{ON}/I_{OFF} is log-normally distributed.

Fig. 4.15 shows that the variability of the on-current increases exponentially as the supply voltage is lowered, the variability of the off current increases linearly with V_{DD} with a small slope.

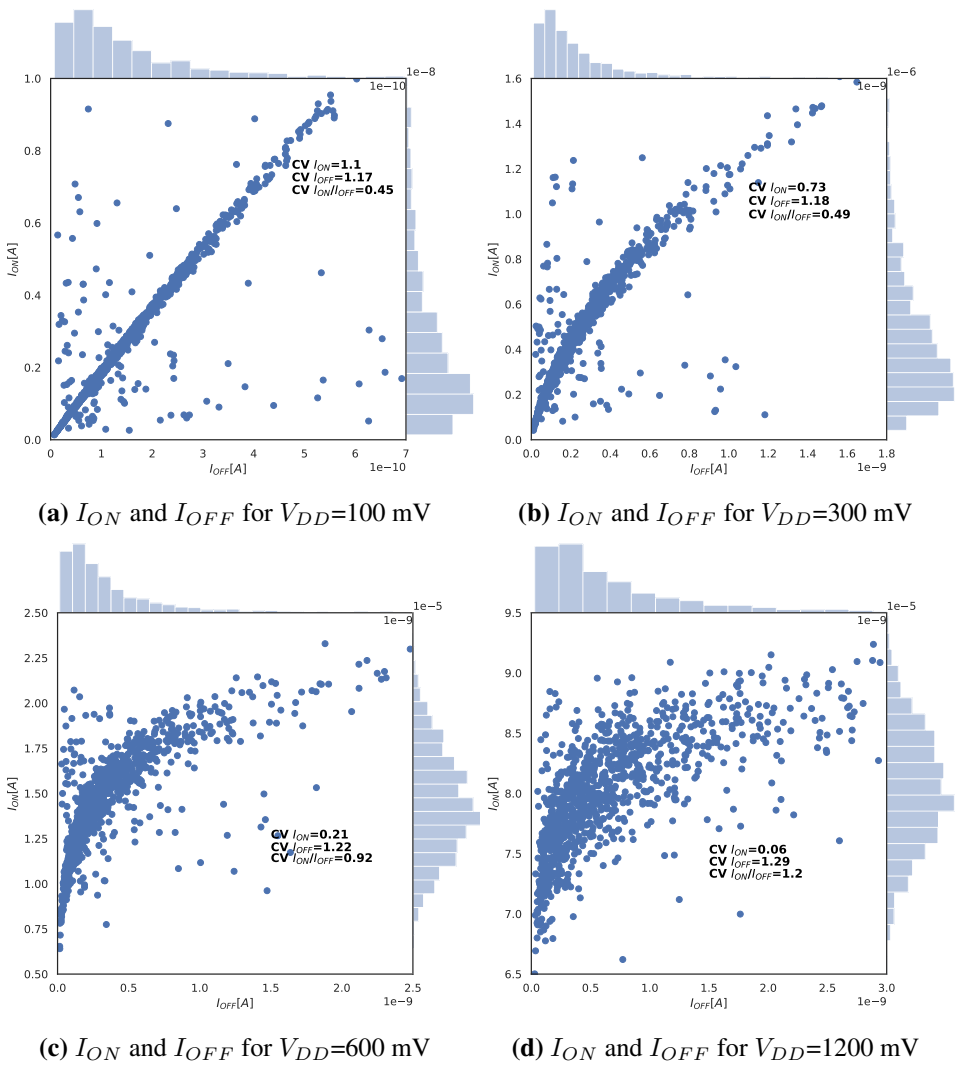


Figure 4.14: The 4 sub-figures a) b) c) d) show the variation in I_{ON} and I_{OFF} current for supply voltages of 100 mv, 300 mv, 600 mv, 1200 mV respectively of a minimum sized NMOS transistor in 130 nm CMOS.

4.3.2 Robustness, Resilience & Stability of Circuits

PVT variations on a device level require design considerations on a system level to ensure that circuit blocks function correctly; this is especially true for low voltage circuits. Three expressions that are commonly used in circuit design for variation tolerant circuits:

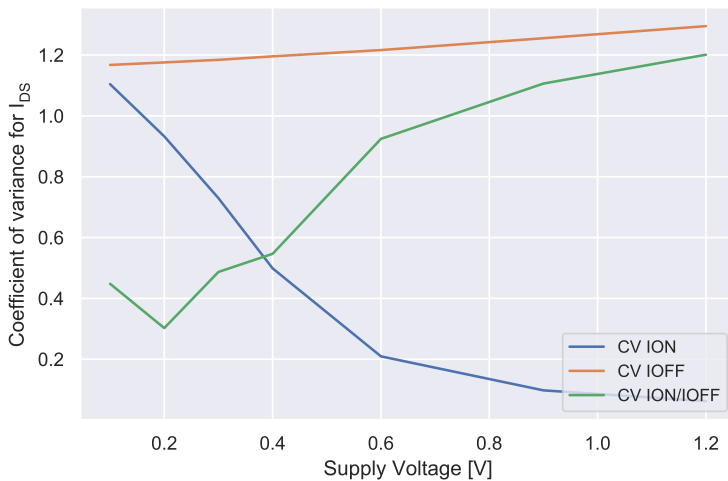


Figure 4.15: Variability of the drain currents from Fig. 4.14 as a function of supply voltage of a minimum sized NMOS transistor in 130 nm CMOS.

- **Robustness:** "The ability to withstand or overcome adverse conditions or rigorous testing" [37]. For the sake of circuit design: The circuits' ability to tolerate variations without erroneous behavior.
- **Resilience:** "the ability of a substance or object to spring back into shape; elasticity" [37]. For the sake of circuit design: The circuits' ability to recover from erroneous behavior.
- **Stability:** "The state of being stable" where the adjective stable is defined as "Not likely to give way or overturn" or "Not likely to change or fail" [37]. For the sake of circuit design: The stability of the circuit over time is dependent on the circuits' robustness and resilience abilities.

Sufficient design margins are needed to tolerate PVT variations and ensure robust operation. For circuits operating in environments with heavy radiation, such as space, methods of tolerating spikes in internal node voltages from particle strikes are required to ensure robustness from single event upsets (SEU) in memory logic and single event latch-up (SEL) in bulk CMOS [38].

Circuit design techniques such as redundancy with majority voters, check-sums, and automatic power-off - power-on functionality are examples of system design methods of achieving resilience against erroneous behavior. Circuit operation may continue despite the lack of robustness.

In low voltage circuit design where margins are generally small, a mix of robustness and resilience is often required to tolerate PVT variations and variations in internal node voltages from particle strikes to ensure correct functionality.

4.4 Low Voltage Memories

In the process of manufacturing integrated circuits, the production of each individual device can be considered as drawing multiple samples from individual statistical distributions as shown in Fig. 4.14. The more devices there are in a circuit, the more likely it is for a rare statistical sample to be drawn. Memory arrays such as DRAMs and SRAMs incorporate millions or even billions of transistor instances and should be designed to operate robustly across PVT variations. High sigma simulations and considerations are needed to ensure robust operation of memory arrays [39]. The larger the array, the higher the sigma required in Monte Carlo simulations to ensure none or a few defects. Special architectural techniques are required to correctly operate low voltage SRAMs, these are assist circuits for data retention, assist circuits for write operations and assist circuits for read operations and special care in SRAM sense amplifier design.

4.4.1 Data Retention in Low Voltage SRAM Cells

The core functionality of a memory cell is retaining data. As we scale the supply voltage down to save static and dynamic power and energy, retaining the binary state in the memory cell becomes increasingly tricky. The cell's ability to retain data in SRAMs is measured in static noise margin (SNM) [40]. The SNM is the maximum noise voltage that can be tolerated on the data nodes in the opposite direction of the data before the cell changes states [41]. Since an SRAM cell consists of two inverting elements where the output of one is connected to the input of the other and visa-versa, the SNM of the memory cell can be found graphically by looking at the voltage transition curves (VTC) of the inverting elements. The SNM is found by mirroring the VTC with respect to a 45° line and finding the length of the side of the largest square fitted inside the eyes of the two intersecting curves. Reducing the supply voltage SRAM cells, causes the eyes of the two intersecting VTCs to shrink drastically, creating a need for innovation in the core of the memory cell to achieve adequate noise margins.

Hold-assist circuits are used to make the cells robust against the transient voltage spikes that the SRAM cell is subject to in its retention state. Such circuits are adopted in low voltage SRAMs to prevent unintentional bit-flips while storing data, i.e., hold-assist circuits typically increase the hold static noise margin of the SRAM cells. A good example of a typical Hold-assist circuit is the addition of 4 transistors to the traditional 6 transistor (6T) SRAM cell, transforming each inverting element

of the cell into Schmitt triggers [42], thus steepening the VTC and opening up the SNM eyes. Hold-assist circuits such as the exemplified one drastically increase the transistor count of the cell.

4.4.2 Read Operations in Low voltage SRAM Cells

In read operations of traditional 6T SRAM cells, the pass-gates connecting the retention core to the bit lines (BLs) are opened to buffer the stored cell value to the readout circuitry. Current is flowing from the pre-charged BL through the pull-down transistor of the inverter holding a logic 0. As a consequence, the voltage on the data node is increased from the voltage drop across the effective resistance in the pull-down transistor. The increased voltage may cause the other inverter to evaluate its input as a logic high instead of a logic low flipping the state of the SRAM cell. This is known as a destructive read operation, and the data stored in the cell is lost. For regular-voltage SRAM cells, this problem may be solved by appropriately sizing the strength ratio between the pull-down devices and the pass-gate devices in the cell, also known as the cell ratio or the β ratio. The CR is given by EQ. 4.13. For read operations, the higher the β , the better the read margin [43].

$$\beta = CR = \frac{W_{Pull-down}}{L_{Pull-down}} \frac{W_{Pass-gate}}{L_{Pass-gate}} \quad (4.13)$$

Since BL evaluation time is part of the read access time sense amplifiers are commonly used to quickly evaluate cell values from BL voltage differentials. Sense amplifiers show poor performance in low voltage SRAMs due to variation in the weakened low-voltage transistors. Standard differential amplifiers are no longer practical as sense amplifiers for supply voltages below roughly twice the absolute threshold voltage of the transistors [44]. For low voltage memories special sense amplifier architectures have been proposed [44], [45], [46] as well as sense amplifier-less designs [47], [48], [49]. In this thesis we mainly delve in to cell architecture of SRAMs and use sense-amplifier less, full-swing, schemes for improved sensing margin and simplicity. There is still a need for good low-voltage sense amplifiers to improve the readout performance of SRAMs.

For read operations in low voltage SRAM cells, a popular approach is to introduce read-assist circuits. Read-assist circuits are used to make the cells robust against the transient voltage spikes that the SRAM cell is subject to in its' read state. Such circuits are used to reduce the chance of unintentional bit flips during read operations, i.e. Read-assist circuits typically increase the read static noise margin of the SRAM cell. A popular read-assist circuit will decouple the cell storage nodes from the BLs during read operation, making the SRAM cell as stable during

read operations as it is in data retention mode. This can be achieved by adding a 2 transistor read buffer to each data node [50].

4.4.3 Write Operations in Low voltage SRAM Cells

In write operations to conventional 6T SRAM cells, the intention is to flip the state of the cell if the state of the cell and the data to be written are binary opposites. Initially, in a write operation, one BL is pre-charged to V_{DD} and the other to GND , then the pass-gates are opened and current flows from the pull-up device in the inverter that holds a logic 1 through the pass-gate and to the pull-down device of the BL driver. The write margin (WM) of an SRAM cell is defined as the voltage of the low bit-line required to flip the state of the cell. The pull-up ratio (PR) of the cell, given in EQ. 4.14, has to be low enough to create a significant enough voltage drop across the pull-up device to have the data node holding a logic 1 take a logic 0 value successfully flipping the state of the cell [51]. Studying the equation for CR and the equation for PR shows a contradicting sizing objective for the write-ability of the SRAM cell as to the read stability of the same cell. For write operations, a strong pass-gate is required compared to the pull-up device, and for read operations, a weak pass-gate is required compared to the pull-down device.

$$PR = \frac{\frac{W_{Pass-gate}}{L_{Pass-gate}}}{\frac{W_{Pull-up}}{L_{Pull-up}}} \quad (4.14)$$

Write-assist circuits are used to ease write operations on the SRAM cell. Such circuits typically improve the write margin of the SRAM cell by weakening the pull-up device of the cell when the write operation occurs. Like the other assist circuits, a penalty for the write-assist circuit is an increased transistor count.

An interesting observation for the cell ratio and the pull-up ratio in SRAMs is that: By decoupling the retention core of the cell from the BL during read operations, the sizing requirement for cell read-ability can be ignored and the contradictive sizing requirement to achieve writability can be maximized alone.

"I have not failed. I've just found 10,000 ways that won't work."

-Thomas Edison

This chapter briefly gives an account of the particularities of how the research was undertaken. Since each article is stand-alone and contains its own method description, the information covered here will only cover the general research method that has been applied throughout this work.

5.1 Research Time Line

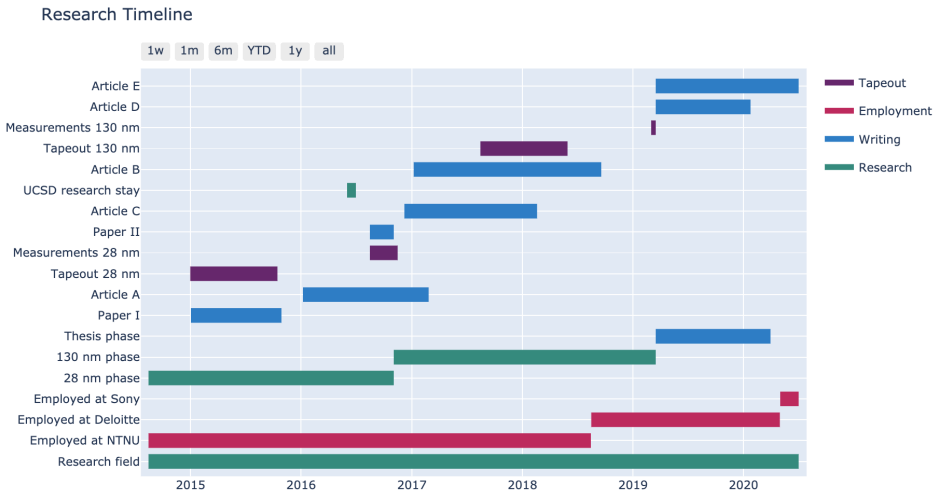


Figure 5.1: Timeline for the research conducted in this thesis.

The research presented in this thesis has been conducted in two major phases from 2014 to 2020. Each phase consisted of a literature study, brain-storming, design, tape out, measurements, and paper writing according to the timeline in Fig.5.1. The first cycle focused on DFF, and SRAM design in 28 nm Fully Depleted Silicon on insulator (FD-SOI), and the second phase targeted the design of SRAM schemes

and level shifter designs in a bulk 130 nm BiCMOS process.

5.2 CMOS Technologies

In the two above-mentioned design phases, the following design kits from STM through CMP were used:

- IC ST Microelectronics 28nm Advanced CMOS FDSOI 8 ML CMOS28 FD-SOI¹
 - 28nm drawn poly gate length
 - Triple well Fully Depleted SOI devices, with ultrathin BOX and Ground Plane
 - Body biasing
 - Dual Vt MOS transistors (LVT, RVT)
 - Dual gate oxide (1.0V for core and 1.8V for IO)
 - Temperature range: -40°C to 175°C
 - Dual-damascene copper for interconnect, 8 metal layers (8ML) for interconnect
 - 2 thick Cu top metal (0.880 micron)
 - Low k inter-level dielectric
 - Fringe MoM capacitors
 - Inductors
 - Analog / RF capabilities
 - Various power supplies supported: 1.8V, 1.0V
 - Standard cell libraries (more than 3Mgates/mm²)
 - On request: Embedded memory (Single port SRAM / ROM / Dual Port SRAM).
- IC STMicroelectronics 130nm BiCMOS SiGe 6 ML BiCMOS9MW²
 - 130nm drawn, 130nm effective
 - Deep Nwell and Deep Trench Isolation
 - Double Vt transistor offering (Low Leakage , High Speed)
 - Dual gate oxide (1.2V for core and 2.5V for IO)
 - Threshold voltages (for 2 families above): VTN = 450/340mV, VTP = 395/300mV
 - Isat (for 2 families above): TN @ 1.2V: 535/670uA/mic
 - TP @ 1.2V: 240/310uA/mic
 - Bipolar SiGe transistors: High Speed NPN
 - Medium VoltageNPN
 - Typical beta (for 2 families above): 1000/1000
 - Typical Ft (for 2 families above): 230/150GHz

¹ Kit Details from www.mycmp.fr

² Kit Details from www.mycmp.fr

- Power supply 1.2V
- Temperature range: -40°C to 175°C 6 Cu metal layers
- Low k inter-level dielectric
- MIM capacitors
- Standard cell libraries (more than 180kgates/mm²)
- Embedded memory (Single port RAM / ROM / dual port RAM).

5.3 Design Software And Measurement Equipment

- In each design phase, the same software tools have been used with slight variations in version numbers:
 - Cadence Virtuoso 6, Custom IC Design Environment (Schematic & Layout)
 - Cadence Spectre 16, Simulation Platform (Circuit simulator)
 - Cadence Innovus 16, Implementation system (Place & route)
 - Synopsys Design Compiler 2016, Synthesis solution (Logic synthesis)
 - Mentor Graphics Calibre 2017, nmDRC, nmLVS (Design rule checking & layout vs schematic comparisons)
 - Cadence Quantus 18, Extraction Solution (Layout extraction)
- During silicon measurements, the following tools instruments have been used:
 - Rigol DP 832A digital Power Supply
 - HP 6632A DC - digital Power Supply
 - Keithley 6485 - Picoammeter
 - Agilent 33522A - Function generator
 - Rohde & Schwarz RTE 1022 - Oscilloscope

5.4 Manufactured Prototypes

The final version of the circuit layout in Fig. 5.2 a) was sent to CMP in october 2015.

The final version of the circuit layout in Fig. 5.2 b) was sent to CMP in may 2018.

The micrograph of the manufactured circuit that was returned form the fab in august 2016 is shown in Fig. 5.2 c). The circuitry on the silicon is not visible since metal filling in the top layers blocks the light.

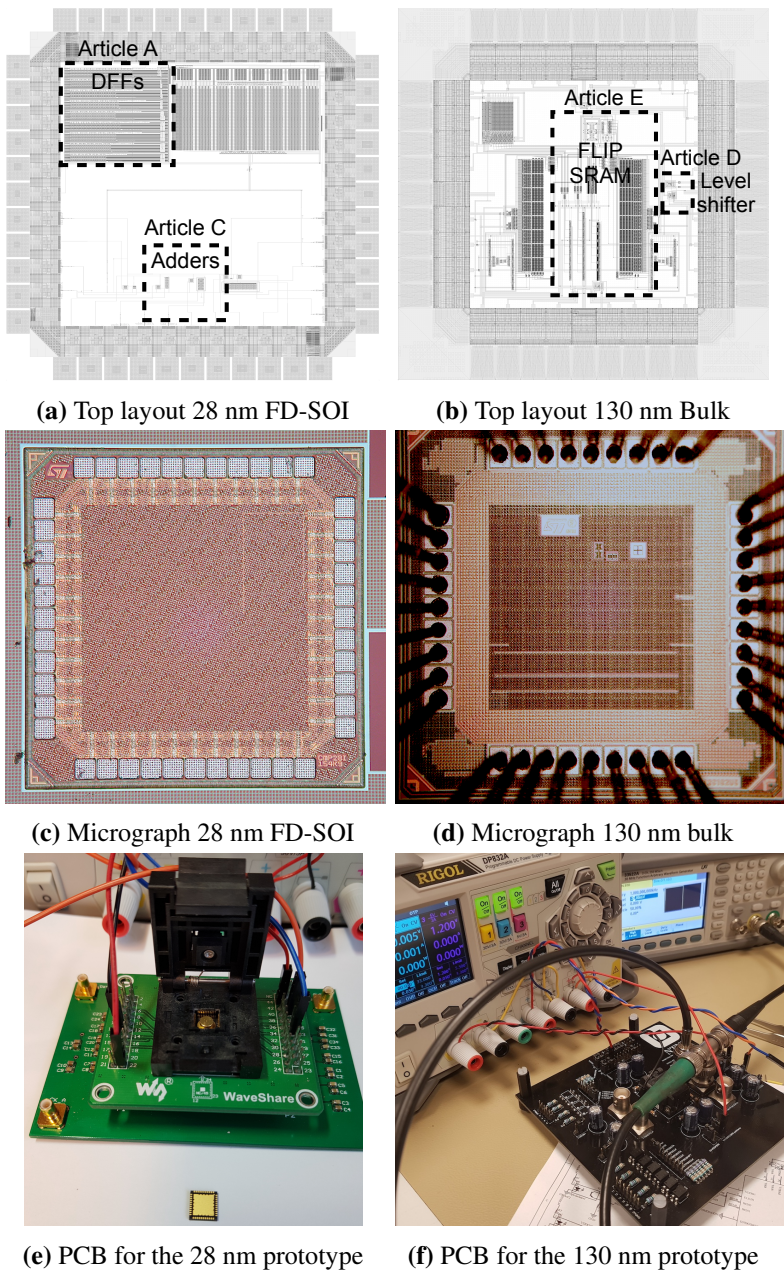


Figure 5.2: Top layout and micrographs for the two tapeouts of this thesis.

The micrograph of the manufactured circuit that was returned from the fab in march 2019 is shown in Fig. 5.2 d). The circuitry on the silicon is not visible

since metal filling in the top layers blocks the light.

The custom PCB used in measurements of the manufactured 28 nm circuit is shown in Fig. 5.2 e).

The custom PCB used in measurements of the manufactured 130 nm circuit is shown in Fig. 5.2 f).



PART II

AUTHOR'S CONTRIBUTION

Research Questions

"As we enjoy great advantages from the invention of others, we should be glad of an opportunity to serve others by any invention of ours, and this we should do freely and generously."

-Benjamin Franklin

In this thesis, low power and energy-efficient circuits and concepts are designed with voltage scaling as a key concept using modern IC tools in popular CMOS technologies. The results of this work are presented in 5 journal articles numbered chronologically from A-E with two supplementary conference papers I and II.

Five research questions have been answered throughout these publications.

- RQ 1) Is there a synchronous edge-triggered memory element that shows superior low voltage behavior? Answered by article A and supplementary article I, presented in the form of a comparative study.
- RQ 2) Many Low voltage SRAM cells have been reported. However, most of them deploy multiple extra transistors as write and read assists. Can a compact SRAM cell function in the near-threshold region? Answered by article B, where a 6T SRAM cell is proposed with a few re-arranged transistors.
- RQ 3) The drive strength ratio between uniformly sized NMOS and PMOS transistors differs significantly from 1 in subthreshold CMOS logic. How can back-gate biasing and the 28nm FDSOI poly-biasing option be used to balance this ratio towards 1. Answered by article C and supplementary paper II, where such techniques are used to manufacture low voltage adders.
- RQ 4) How can low voltage circuit blocks communicate with regular voltage modules in an energy-efficient manner in systems containing multiple voltage domains? Furthermore, what is the lower limit of the supply voltage of the low voltage circuit? Answered by article D, where a low voltage level shifter is proposed.

- RQ 5) Low voltage SRAM cells are often asymmetric in nature. They typically have a resulting state-dependent behavior that is either desired or undesired. Can this be exploited dynamically during write operations? Answered by article E, where a macro-level scheme for low voltage SRAMs is proposed.

Extended Comparative Analysis of Flip-Flop Architectures for Subthreshold Applications in 28 nm FD-SOI.

E. Lâte, A. A. Vatanjou, T. Ytterdal, and S. Aunet

Published in *Microprocessors and Microsystems*, vol. 48, pp. 11-20, 2017¹

DOI: <https://doi.org/10.1016/j.micpro.2016.07.016>

A.1 Abstract

Nine D-type Flip-Flop (DFF) architectures were implemented in 28 nm FDSOI at a target, subthreshold, supply voltage of 200 mV. The goal was to identify promising DFFs for ultra low power applications. The single-transistor pass gate DFF, the PowerPC 603 DFF and the C2MOS DFF are considered to be the overall best candidates of the nine. The pass gate DFF had the lowest energy consumption per cycle for frequencies lower than 500 kHz and for supply voltages below 400 mV. It was implemented with the smallest physical footprint and it proved to be functional down to the lowest operating voltage of 65 mV in the typical process corner. During Monte Carlo (MC) process and mismatch simulations it was also found that the pass gate DFF is least prone to variations in both minimal setup- and minimal hold-time. Race conditions, during mismatch variations, occurred for the flip-flop that is constructed from NAND and inverter based multiplexers. The pass gate DFF is outperformed slightly when it comes to D-Q-based power-delay product and more significantly when it comes to the maximum clock frequency. The flip-flops having the shortest D-Q delays were the PowerPC 603 and the transmission gate D flip-flop, these also had the lowest D-Q-based power-delay of 26% and 30% respectively of that of the worst-case S2CFF power-delay product.

A.2 Relevance to the Thesis

D Flip Flops are edge-triggered memory elements that are widely adopted in memory systems and commonly used as gates in pipeline stages of sequential computa-

¹References in article: [52] [53] [54] [55] [56] [57] [58] [59] [60] [61] [39] [62] [63] [64] [65] [66] [67] [68]

tional circuits such as data processors. This article presents the results from one of the very first circuit design projects that were performed as a PhD candidate at NTNU. The article analyzes various DFF architectures and identifies promising candidates for low voltage operation. The results from the analysis were intended to serve as a solid fundament for the circuit design projects that were to follow. The publication is also directly relevant to the thesis since it answers research objective O1, mentioned in chapter 3.

A.3 Declaration of Authorship

This article is an invited and extended version of the supporting paper I, that was presented at the Nordic Circuits and Systems Conference (NORCAS) in 2015. The research to the conference paper was conducted in the first 5 months of the PhD work. Ali Asghar Vatanjou was at the time familiar with low voltage design and contributed with valuable implementation techniques. I performed the literature study, implemented the circuits, and wrote the paper under the supervision of Trond Ytterdal and Snorre Aunet.

A.4 Manuscript



Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Extended Comparative Analysis of Flip-Flop Architectures for Subthreshold Applications in 28 nm FD-SOI[☆]



Even Låte^{1,*}, Ali Asghar Vatanjou², Trond Ytterdal³, Snorre Aunet⁴

Department of Electronics and Telecommunications, Norwegian University of Science and Technology, O.S. Bragstads plass 2a, 7034, Trondheim

ARTICLE INFO

Article history:

Received 22 February 2016
Revised 21 June 2016
Accepted 21 July 2016
Available online 22 July 2016

Keywords:

Subthreshold
28 nm FD-SOI
D Flip-flop
Comparison
Setup Time
Hold Time

ABSTRACT

Nine D-type Flip-Flop (DFF) architectures were implemented in 28 nm FDSOI at a target, subthreshold, supply voltage of 200 mV. The goal was to identify promising DFFs for ultra low power applications. The single-transistor pass gate DFF, the PowerPC 603 DFF and the C²MOS DFF are considered to be the overall best candidates of the nine. The pass gate DFF had the lowest energy consumption per cycle for frequencies lower than 500 kHz and for supply voltages below 400 mV. It was implemented with the smallest physical footprint and it proved to be functional down to the lowest operating voltage of 65 mV in the typical process corner. During Monte Carlo (MC) process and mismatch simulations it was also found that the pass gate DFF is least prone to variations in both minimal setup- and minimal hold-time. Race conditions, during mismatch variations, occurred for the flip-flop that is constructed from NAND and inverter based multiplexers. The pass gate DFF is outperformed slightly when it comes to D-Q-based power-delay product and more significantly when it comes to the maximum clock frequency. The flip-flops having the shortest D-Q delays were the PowerPC 603 and the transmission gate D flip-flop, these also had the lowest D-Q-based power-delay of 26% and 30% respectively of that of the worst-case S²CF power-delay product.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Ever since the Centre Electronique Horloger (CEH) in Switzerland succeeded with low power integrated circuits [2], scientists have strived to create new and improved methods for designing low power systems. Subthreshold CMOS operation is such an approach and is applied both in the analog and in the digital circuit design domain. By reducing the supply voltage below the threshold voltage of the transistors, the power consumption can be reduced by several orders of magnitude. The principle of operating transistors in weak inversion dates back to 1967 [3]. Despite its early introduction, only a few niche-products exploiting subthreshold operation have been established in the consumer-electronics market until now. The sensor-network trend known as the Internet

of Things (IoT) has re-opened the interest for the ultra low power (ULP) benefits of operating transistors in the subthreshold region. ULP Sensor nodes driven by coin-cell batteries will last for a long time, or perpetually, if energy harvesting is used in combination with subthreshold operation.

D flip-flops (DFFs) are widely used in digital applications such as parallel storage and to serially manipulate internal circuit nodes in SoCs. These edge triggered devices account for large fractions of circuit area and power consumption and are also amongst the most difficult digital building blocks to push to lower supply voltage levels without experiencing problems with data corruption. Either the output (Q) is subject to voltage drop so that it fails to take its correct logical value, or the internal delay of the master latch is too high to allow the slave latch to latch on to the data (D) signal.

The objective of this work is to explore multiple flip-flop architectures and identify promising candidates among these for subthreshold applications in 28 nm Fully Depleted Silicon On Insulator (FDSOI).

In this paper, section 2 introduces details on a schematic level for the chosen set of DFF architectures. The simulation-setup and performance metrics for evaluating these are explained in section 3, followed by results and an evaluation of the schematics in section 4 & 5 respectively.

[☆] This paper is an extended version of the already published conference paper "Comparative analysis of flip-flop architectures for subthreshold applications in 28nm FDSOI" [1].

* Corresponding author. Tel.: +4792893321.

E-mail addresses: even.late@ntnu.no, even.laate@iet.ntnu.no (E. Låte).

¹ PhD Candidate

² PhD Candidate

³ Professor & Supervisor

⁴ Professor & Supervisor

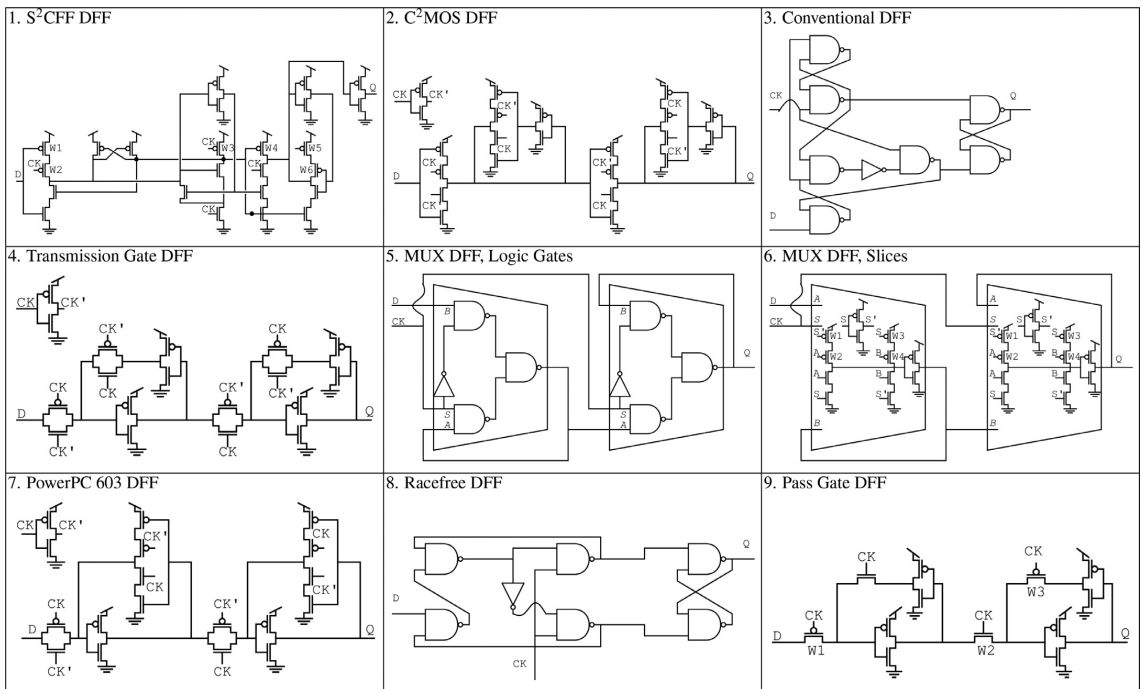


Fig. 1. Schematic overview of the flip-flop architectures that are considered in this study, the nand and inverter symbols in cell 3. and 5. are made with 4 and 2 LVT transistors respectively.

2. Materials and Methods

2.1. 9 D Flip-Flop Architectures

Nine static D flip-flop architectures were picked out for a comparative study of their subthreshold abilities. The schematics for each of them are illustrated in Fig. 1 and are further explained in the enumerations below:

1. The static single-phase contention-free flip-flop (S^2CFF) [4,5] is shown in cell No. 1 in Fig. 1. It is based on a dynamic true single-phase flipflop and is made static by including a slave latch and some additional transistors. The S^2CFF was designed in 2014 for near threshold operation on a 45 nm SOI technology.
2. The second structure is the well known Clocked CMOS flip-flop (C^2MOS) [6]. It consists of clocked 4T and basic 2T inverters and relies on two clock phases.
3. The conventional NAND-based design in cell 3 was used for its simplistic gate level structure together with Schmitt-trigger gates in [7] to push the lower limits of the supply voltage. The original design is made into a positive edge triggered DFF by removing one inverter on the clock (CK) input.
4. A flip-flop design consisting of transmission gates and inverters such as the one shown in cell number 4, is referred to as the Transmission Gate DFF. It is used extensively in sequential systems for its low power consumption and low transistor count. Note that two clock phases are used.
5. A static master slave flip-flop can be constructed using two multiplexers (MUXes) configured as latches. In this paper

two different MUX topologies are investigated for such a configuration. Cell number 5 shows such a DFF with MUXes built up from conventional logic gates.

6. Cell number 6 shows a second MUX-based flip-flop where the MUXes are built in a XNOR-like fashion with logic branches.
7. The PowerPC 603 flip-flop in cell number 7 was used in the PowerPC 603 microprocessor [8]. The PowerPC flip-flop has been modified slightly since then, the two transistors controlling the scan mode of the DFF have been removed. This design has in previous comparative studies been found to have great subthreshold abilities [9–11].
8. Cell number 8 contains another NAND and inverter based flip-flop that is race free [12]. This design differs from the conventional NAND based flip-flop by the fact that it has less logic gates and thus occupies a smaller amount of area.
9. Transmission gates are normally used to achieve symmetrical transmission of logic high and logic low signals. The single pass gate design in cell number 9 exploits the area and single clock-phase benefits of using single transistors as pass gates instead of the 2T transmission gates [13], the trade-off is loss in signal integrity. In 2015, this DFF architecture was simulated in 28 nm FDSOI with promising results [14].

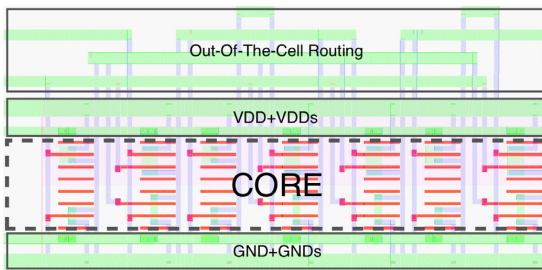
2.2. Flip-Flop Implementation, Sizing & Layout

For this study, all the DFF architectures were implemented for a supply voltage of 200 mV and with LVT transistors. The transistors were sized to give equal drive strength for pull up and pull down networks. Flip-flops consisting of logic building blocks were

Table 1

Transistor dimensions, transistor counts and area of the implemented layout for the 9 DFFs in Fig. 1.

DFF	Transistor dimensions (nm), $W_{\text{default}}=80$, $L_{\text{default}}=30$	Trans. Count	Area Full (μm^2)	Area Full (% of $S^2\text{CFF}$)	Area Core (μm^2)	Area Core (% of $S^2\text{CFF}$)
1. $S^2\text{CFF}$	$W_1 W_2 W_5 W_6 W_{\text{pINV}}=172, W_3=122, W_4=110$,	24	24.11	100.00	7.45	100.00
2. $C^2\text{MOS}$	$W_{\text{pCINV}}, W_{\text{pINV}}=172$	22	15.30	63.46	5.02	67.40
3. Conv.	$W_{\text{pNAND}}=162, W_{\text{pINV}}=172$	30	17.95	74.45	5.44	73.02
4. Transm.	$W_{\text{pTransm}}=200, W_{\text{pINV}}=172$	18	18.90	78.39	5.75	77.18
5. MUX-1	$W_{\text{pNAND}}=162, W_{\text{pINV}}=172$	28	17.65	73.21	5.35	71.81
6. MUX-2	$W_1, W_2, W_3, W_4=170, W_{\text{pINV}}=172$	24	19.67	81.58	6.00	80.53
7. PowerPC	$W_{\text{pTransm}}=200, W_{\text{pCINV}}, W_{\text{pINV}}=172$	18	16.34	67.78	5.39	72.34
8. Racefree	$W_{\text{pNAND}}=162, W_{\text{pINV}}=172$	26	15.81	65.57	4.75	63.76
9. Pass Gate	$W_1 W_2 W_3=200 W_{\text{pINV}}=172$	12	11.82	49.03	4.19	56.24

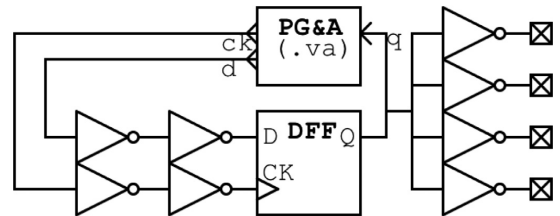
**Fig. 2.** Layout of the $C^2\text{MOS}$ DFF, width = 5.613 μm , full height = 2.726 μm .

sized from the bottom and up. The drawn channel lengths (L) of the transistors, in all flip-flops, were kept constant to minimize device mismatch and to simplify the architectural comparisons. All flip-flops were designed with drawn channel lengths equal to the 28 nm FDSOI process minimum at 30 nm.

For the basic 2T inverters, the width of the PMOS was found by locking the NMOS width to the process minimum while sweeping the input voltage such that the voltage transfer characteristic has an output equal to $V_{DD}/2$ for an input equal to $V_{DD}/2$. For gates with multiple inputs, multiple transfer curves were achieved by holding non switching inputs in the on-state. The average of the multiple transfer curves was used for $V_{DD}/2$. This approach is ideal at uniform distributions of data input patterns. The transmission gate that is used in some of the DFFs was sized to have equal drain current for both NMOS and PMOS while being in the on-state.

Table 1 lists the settled transistor dimensions in the second column from the left. The transistors that are not mentioned in the table have minimum process dimensions i.e. drawn channel lengths equal to 30 nm and channel widths equal to 80 nm. The table also lists the differences in transistor count and area consumption for the 9 flip-flops.

Fig. 2 shows the layout of the $C^2\text{MOS}$ DFF. A layout style where metal over gate-poly is avoided, is used for all DFF implementations in a similar fashion to make the comparison fair. The areas depend directly on the transistor counts and the regularity of the architectures, flip-flops with complex structures require extra area for routing outside of the cell. The out-of-the-cell routing makes it easier to visualize the differences in interconnect complexity for the flip-flops, which may ease the layout-process for the designer if taken into account during a flip-flop selection. See the four right-most columns of Table 1 for the area consumption of the flip-flops. The LVT transistors in 28 nm FDSOI are implemented in layout with flip well, N-channel devices reside in N-well and P-channel devices in P-well. The DFFs are compared with both the PMOS-network back-gate connection (VDDs), and the NMOS-network back-gate connection (GNDs), connected to the

**Fig. 3.** Test setup; A verilog A module for Pulse Generation and Analysis (PG & A), a FO4 load and realistic input signal drivers.

LVT-default ground potential. It could also be interesting to compare the DFFs to each other across the technology-given, feasible, range of body biasing but this is not done in this work.

All flip-flop layouts were analyzed with version 14.2 of the QRC parasitic extraction tool. Both the resulting post-layout-netlists and the pre-layout-schematics are used extensively in simulations throughout this paper.

2.3. Simulation Setup & Test Methodology

It is for subthreshold applications desired to compare the DFF topologies with respect to functionality, operating speed, power and robustness against process variations. The test-bench setup that is used in this comparative study is given for a single DFF instance in Fig. 3. The flip-flop architectures presented above were analyzed using similar comparative approaches as those proposed in [15]. Each flip-flop was simulated using double inverters to drive the inputs and a fan-out of 4 inverters as output loads. Version 14.1 of the Spectre circuit simulator was used for all analog simulations in this study.

The test setup in Fig. 3 also includes a programmable Pulse Generation and Analysis (PG&A) module. Since this module is written in verilogA, it is compatible with the Spectre circuit simulator.

2.3.1. Simulating Minimal Supply Voltage, Maximum Frequency & Functional Yield

The PG&A module generates data and clock signals such as the top two waves shown in Fig. 4. Functionality is verified by automatically probing the Q signal at the interval marked with circles in the on the Q-signal. To allow observation of functionality on each individual positive clock edge, the data signal toggles on every falling clock edge which gives symmetrical setup and hold times for the DFF. Functionality is then tested on a uniform output data distribution with a 50% switching activity-factor.

The lowest functional supply voltage was found for all flip-flops by introducing a minimal value for logic high equal to 75% of V_{DD} . A maximum value for logic low was also set to be 25% of V_{DD} . The supply voltage was then reduced to the point where the logic level

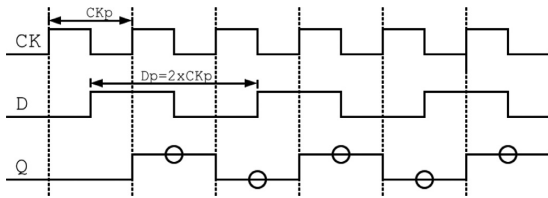


Fig. 4. Simulation stimuli used for verification of functionality.

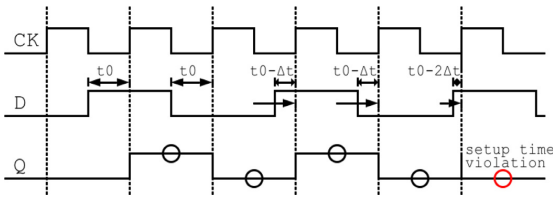


Fig. 5. Simulation stimuli used to map minimal setup time.

limits were breached by Q. For functionality comparisons at low operating voltages in the TT corner, the waveforms used in transient simulations were selected to allow settling of slow signals as a result of a low operating voltage. The lowest functional operating voltage was thus found with simulations at 1kHz clock-frequency.

Maximum operating frequencies at 200 mV in the TT corner was also mapped with the functionality-test stimuli in Fig. 4. Having a symmetrical hold & setup time, makes the PG&A platform scalable and thus suitable for frequency adjustments with parametric sweeps. The frequency was simply increased until the verilogA module prompted a faulty output for each of the flip-flop architectures.

Functional yield was also simulated for the DFF architectures for many supply voltage levels. Also here the limits of logic high and logic low were set to 75% of V_{DD} and to 25% of V_{DD} respectively. The functional yield was simulated using the suitable number of 1000 Monte Carlo(MC) simulation-points for 3σ accuracy [16] per DFF for every V_{DD} value.

2.3.2. Simulating the Limits of Setup- & Hold-Time

To map the minimal setup- and minimal hold- times of each flip-flop, the D signal transition was moved iteratively towards the CK edge until the DFF failed to propagate the change in the D input to the Q output on a positive clock edge. The minimal functional setup/hold time is then equal to the previous functional setup/hold time when such an event occurs. The PG&A module was made to sweep the hold and setup time and to detect events such as the one described above and then output the previous setup/hold time used. Using a variable in the delay parameter of the analog transition operator in verilogA allows for automatic sweeping of setup and hold time within the same transient simulation.

The waveform in Fig. 5 depicts a setup time sweep with a setup time violation on $t_0-2\Delta t$. The data signal transition is here moved from the left of the rising clock edge towards the edge.

The waveform in Fig. 6 shows us the corresponding hold time simulation waveform where a hold time violation is detected by the PG&A module yielding a minimal functional hold time equal to the initial hold time used minus one time quantum. The data signal transition is, for the hold time analysis, moved from the right of the clock cycle towards the clock cycle until the DFF fails to latch on to the previous data value.

By modifying the PG&A block to write each setup time or hold time value to a text file, statistical simulations for setup and hold

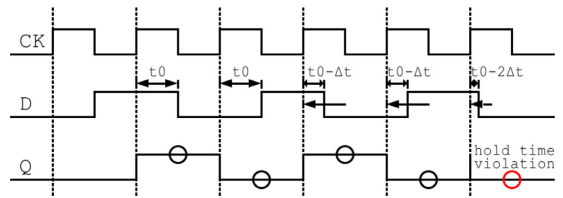


Fig. 6. Simulation stimuli used to map minimal hold time.

time were made possible. Once the output signal slips, the previously working time sample is written to a text file and a new statistical run can be initiated. For many Monte Carlo runs, the text file grows with delay samples for every transient run. This approach was used to map each DFF's timing-susceptibility to process variations which has a high correlation to yield in a larger system setting.

2.3.3. Simulating Power-Delay Products at Optimal Setup Times

Once the minimal setup time was identified for each DFF using the approach in subsection 2.3.2, the setup times were swept from this point and upwards, while observing the D-Q delay. The setup-time where the D-Q delay is minimal and optimal is often greater than the minimal setup time. This procedure was done for both rising and falling Q transitions, both for pure schematics netlists and for the extracted netlists from QRC. For rising and falling Q, the worst case optimal setup times of were chosen for PDP evaluation. This was done to account for asymmetry in the flip-flop schematics and to avoid functional errors deriving from optimistic clocking.

Using optimal setup times, and noting the minimal D-Q delay for a supply of 200 mV, average power consumption over one Q toggle period was probed for each DFF. Multiplying the average power consumptions per cycle by their corresponding D-Q delays gave the power-delay products for all flip-flops.

One should note here that the dual clock-phase flip-flops; C^2 MOS, transmission gate and PowerPC 603 DFF were all implemented with local clock inversion. The local clock inverters were included in the power calculations of flip-flops using two clock phases. The inverted clock signal could also have been generated globally, but the additional power consumption from routing two clock signals would be difficult to model accurately.

2.3.4. Simulating Power & Energy/Cycle in Different Operating Regions

It is in many systems desirable to perform voltage and frequency scaling. This requires that circuit components remain functional over the scaling intervals, and that their performance and power consumption is adequate for all operating regions. To get a comparative overview of the scaling performance of the DFFs the supply voltage and the frequency was swept from 150 mV and 1 kHz, up to the LVT supply limit and up towards the maximum operating frequency of the DFFs. Current drawn from the supply was probed and the circuits correct functionality was simultaneously ensured using the PG&A module that was created. The idle leakage power of each architecture was probed both for the case of storing a logic 0 and the case of storing a logic 1. The leakage metric used here is the average of these two assuming a uniform data pattern.

3. Simulation Results

3.1. Minimum Supply Voltages & Minimal Clock Periods

The bar plot in Fig. 7 depicts the minimal operating voltage where the circuits are still functional for the typical process corner.

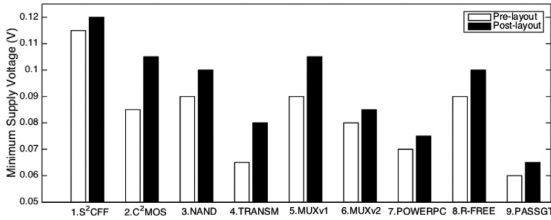


Fig. 7. Minimal functional supply voltage for the DFF architectures.

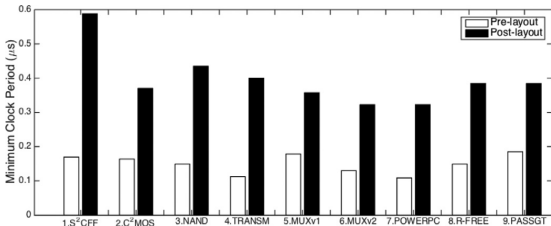


Fig. 8. Minimum functional clock period with symmetrical setup/hold time.

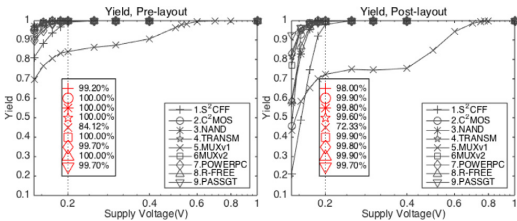


Fig. 9. Yield vs supply voltage for process & mismatch variations.

The bars in white represent simulation results from pre layout simulations, while the black bars represent results from simulations on extracted netlists. This colour encoding is repeated throughout the result section.

Fig. 8 shows the reciprocal of the maximum functional frequencies simulated with symmetrical hold and setup times also in the TT corner. Here, minimum clock period is used instead of maximum clock frequency to better visualize the differences in the post-layout results.

The yield curves in Fig. 9 are created from 1000 Monte Carlo Runs for various supply voltages. In the leftmost box in each plot window, the yield is listed for each DFF at the target supply voltage of 200 mV. The strong deviation in yield for the NAND&INV based mux DFF will be explained in the discussion section.

3.2. Setup Time and Hold Time Variations

In this subsection we will have a look at the results for the 9 DFF's timing simulations. Violation of both setup times and hold times leads to system failures so this aspect should be considered a top priority together with the yield aspect when selecting DFF topologies.

Fig. 10 shows us the variability of the minimal functional setup time in a box plot. The plot is created from 1000 time-samples per pre-layout and post-layout DFF architecture which is generated from 1000 Monte Carlo simulation points per test-bench, ref. Fig. 3. The central marks of the box plots are the median setup time for flip-flops exposed to process variations, the edges of the box are the 25th and 75th percentiles, while the whiskers extend to the

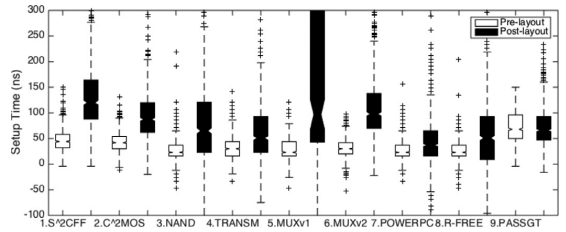


Fig. 10. Effect of process & mismatch on the minimal setup time.

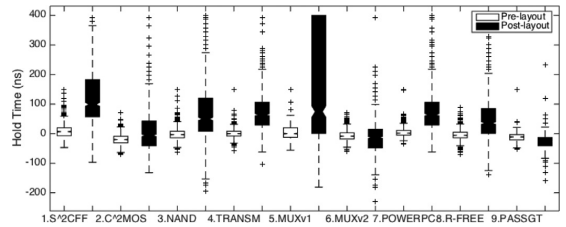


Fig. 11. Effect of process & mismatch on the minimal hold time.

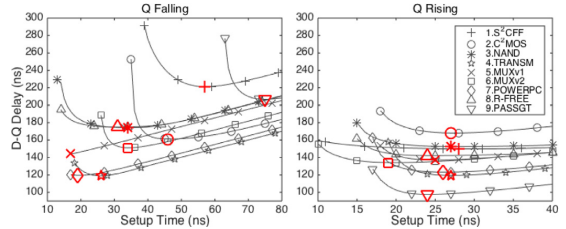


Fig. 12. D-Q delay varies with setup time, falling and rising Q for pre-layout netlist.

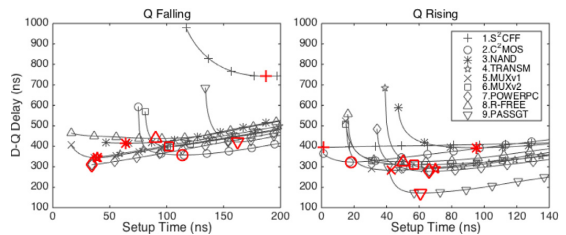


Fig. 13. D-Q delay varies with setup time, falling and rising Q for extracted netlists.

most extreme setup times that are not considered outliers of the setup-time data set. The crosses represent the outliers, that is any value that lies more than one and a half times the length of the box from either end of the box.

Similarly for the minimal functional hold time for each DFF, Fig. 11 is a box plot showing us how the minimal hold time is distributed for process and mismatch variations. This plot is also generated from 1000 Monte Carlo simulation points for each post and pre-layout DFF-netlist.

3.3. Comparative Power-Delay Products

The DFFs were compared with products of power consumption and D-Q delays in the typical corner for operation at optimal setup times. Fig. 12 and Fig. 13 show us how the D-Q delays vary for

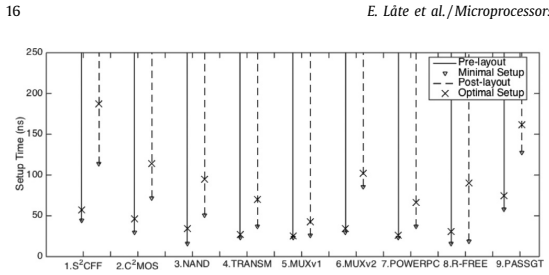


Fig. 14. Feasible ranges and optimal setup times for the 9 flip-flops in the typical process corner.

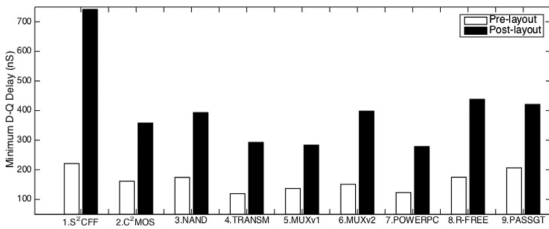


Fig. 15. D to Q delays using the optimal setup times in Fig. 14.

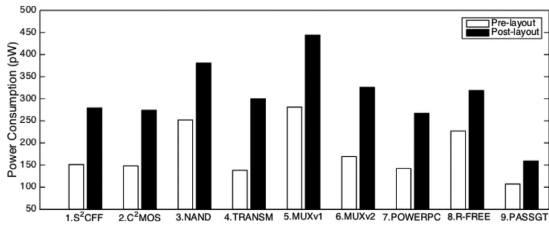


Fig. 16. Power consumptions at $V_{DD} = 200$ mV using optimal setup times.

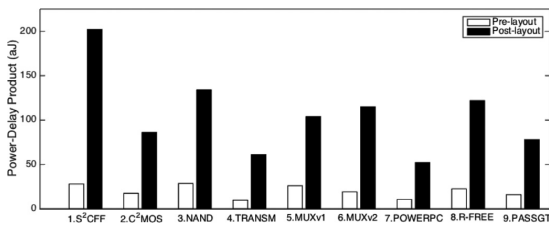


Fig. 17. Power-delay products (PDPs) for each flip-flop based on Fig. 16 and 15.

all flip-flop architectures as a function of setup-time. In the figures, the ideal setup times with respect to D-Q delays are marked with red. As a summary for the four plots, the feasible range of setup times and the optimal setup times for each architecture are given in Fig. 14. The corresponding minimal D-Q delays are plotted in Fig. 15. The average power consumed for one Q-toggle-cycle, of both rising and falling, for each of the DFFs, is given in Fig. 16. The two factors: Minimal D-Q delays and power consumptions, gives the power-delay bars in Fig. 17 when multiplied together.

3.4. Power and Energy in Different Operating Regions

For an idle DFF test-bench in the typical corner, the graphs in Fig. 18 shows us the leakage power that is consumed for different subthreshold supply voltages. The exponential nature of the leak-

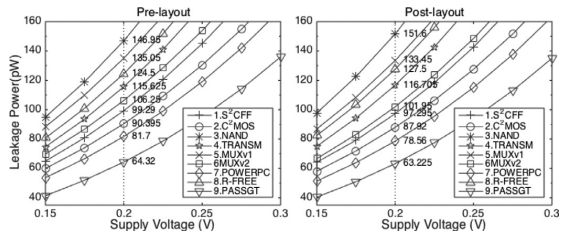


Fig. 18. Leakage power versus supply voltage level for each of the 9 DFFs on the extracted netlist.

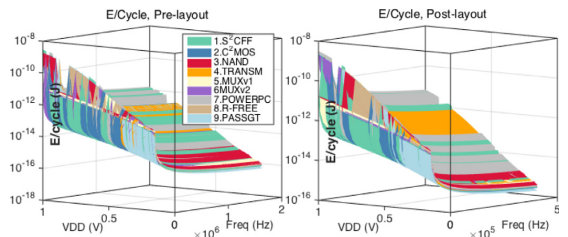


Fig. 19. Flip-flops and their energy consumption per clock cycle in the frequency- and supply-voltage- operation space. The best and the worst DFFs wrt. E/cycle, are further illustrated in Fig. 20, Fig. 21 and Fig. 22.

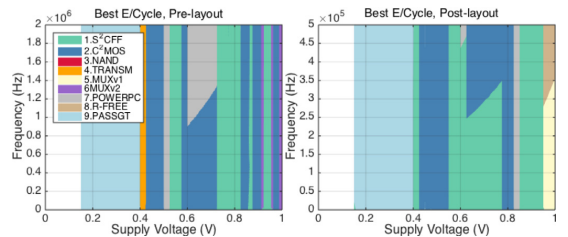


Fig. 20. The flip-flops having the least amount of energy consumed per cycle for given frequencies and supply voltages.

age power versus supply voltage plot continues all the way up to $V_{DD}=1V$ without any deviations from the depicted pattern.

Varying both the operating frequency and the supply voltage, changes what DFF is optimal for specific areas of operation. Each DFF has its own energy per cycle plane in the frequency and supply voltage space. Fig. 19 shows each energy per cycle -surface, with different colors for each DFF, in the same coordinate system. All DFFs are functional at the displayed supply voltages and frequencies. The most significant information is highlighted in Fig. 20 & Fig. 21. Fig. 20 is essentially the same MATLAB surface plot as that in Fig. 19, it is simply rotated to a view from below. Fig. 21 gives us a transposed and rotated view of the same surfaces to list the DFFs that are worst with respect to energy consumption per cycle.

Since this comparative study is mainly intended for DFFs in the sub-threshold region, the average value for each of the surfaces for $V_{DD} < 400$ mV are given in Fig. 22. It becomes evident that the pass gate DFF is most energy efficient in the subthreshold region.

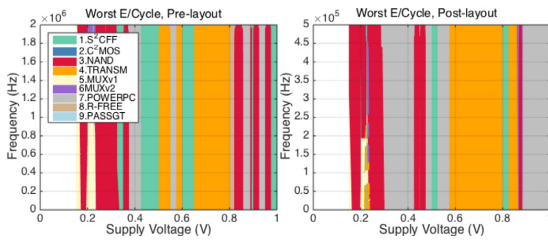


Fig. 21. The flip-flops having the most amount of energy consumed per cycle for given frequencies and supply voltages.

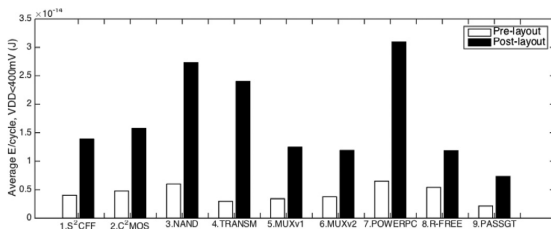


Fig. 22. Average energy per cycle in the sub-threshold region ($V_{DD} < 400$ mV) of the surfaces in Figs. 19, 21, 20.

4. Discussion

The results section in this paper contains comparative results from manifold aspects of DFF properties that can be weighted differently by circuit designers for diverse applications. Table 2 lists each DFF's rating relative to the other DFFs in the aspects of: Transistor Count, Minimal V_{DD} , minimal clock period (maximum frequency), functional yield, variation of minimal setup- & hold-time, power-delay product, leakage power and energy per cycle in the subthreshold region. The smallest values in the sum-column represent the best DFFs if each property is weighted uniformly, normally this is not the case, however it is up to the circuit designer to weigh them.

Generally, the performance variations for pre- and post-layout simulations appear rather large through the result section. This is as expected for subthreshold applications as changes in nodal capacitance are more noticeable with low drive currents. The differences between pre- and post-layout simulations exist for all flip-flops, but to various degrees. Since the layout is made using the same out-of-the-cell routing-style for all the DFFs, the parasitic effects of gate-capacitances should be fair for all, thus making the simulation results depend largely on the schematical advantages of each topology, their drive characteristics, their routing complexity and their transistor count.

Table 2

Summing up the findings for the extracted netlist-results with ratings, 1 = best, 9 = worst, the sum column weighs properties uniformly.

DFF type	# Trans.	Min. V_{DD}	Min. Clock Period	Yield, Single DFF	Min. Setup T. Var.	Min. Hold T. Var.	PDP	Leakage Power	Avg E/cycle, $V_{DD} < 400$ mV	SUM, Uniform Property Weight
1. S ² CFF	6	9	9	8	6	7	9	4	5	63
2. C ² MOS	4	7	3	3	3	4	4	3	6	37
3. Conv.	9	6	8	4	8	8	8	9	8	68
4. Transm.	3	3	7	7	4	5	2	6	7	44
5. MUX-1.	8	8	4	9	9	9	5	8	4	64
6. MUX-2.	5	4	1	1	7	2	6	5	3	34
7. PowerPC	2	2	2	5	2	3	1	2	9	28
8. Racefree	7	5	6	2	5	6	7	7	2	47
9. Pass Gate DFF	1	1	5	6	1	1	3	1	1	20 (best)

Of all the different topologies, the pass gate DFF has a superior area of its footprint compared to the other candidates as can be seen in Table 1. Using only 12 transistors, its footprint in this comparative implementation is only 49% as large as that of the worst case S²CFF. Depending on the circuit application, area is commonly a key criteria for topology selection.

All flip-flops are here designed for a supply voltage of 200 mV. Despite this fact, the lower the minimal operating voltage is, the greater the flexibility at a system level will be. A low limit for supply voltage reduction means having the possibility to perform voltage scaling over a larger voltage range, which again will increase the chances of finding a better global energy operating point, or lasting longer before a brown-out situation. Fig. 7 suggests that the pass gate design is best at this aspect, one should keep in mind that the minimal supply voltage levels found here are in the event of a typical corner.

The aspect of minimal functional clock period for symmetrical hold- & setup-times, also known as maximum clock frequency, is dominated by the XNOR-based MUX DFF and the PowerPC 603 DFF as Fig. 8 depicts. This is one of the weakest comparative checks for the pass-gate flip-flop since the n-channel and p-channel pass gates makes it highly asymmetrical when it comes to rise and fall times of the output, Q. The S²CFF is highly affected by parasitic capacitance from the out of the cell routing because of its routing complexity, it becomes the slowest DFF at 200 mV in the typical corner.

One of the most important comparison aspects is the functional yield, during a flip-flop topology selection this should have a high priority. Most DFFs have a yield in the top 90 percent category. This is, as mentioned section 3 and as depicted in Fig. 9, not the case for the MUX flip-flop that is built out of NANDs and inverters. It has at 200mV a yield of 84% for the pure schematic netlist and a yield of 72% for the post-layout netlist, which makes it useless for subthreshold applications as it stands. It can, however, be considered functional in the subthreshold region if the yield is improved to the top 90 percentage level. As can be seen of the schematic for the NAND&INV based MUX in Fig. 1, each MUX-latch has a feedback from the output of the MUX to either input A or input B for the slave and the master latch respectively. During a toggle of the clock signal, one logic branch is turned on and the other is turned off within one MUX. At the clock edge there is a race condition for the branches, and, depending on the mismatch of driving strengths, an erroneous output may be latched on to. By inserting delay elements in the feedback of each multiplexer-latch, the race condition is solved and, according to new simulations, the yield is improved to 99.2% for the pure schematic netlist. This increases the transistor count of the DFF by 8, making it the least desirable of the 9 with respect to transistor count.

Having a high functional yield as a solo device is important, however, in a system setting with propagation-delay variation, other aspects like resistance against minimal setup- and hold-time

variation, comes into play. Low variation in the limits for setup and hold time together with proper clock routing makes it more likely that no hold or setup time violation is caused by clock skew in the total system. As the box-plots in Figs. 10 and 11 show, the boxes having the least variation in minimal limits of setup time is the pass gate DFF, tightly followed by the PowerPC 603 and the C²MOS. The same is true for the hold time variation, in addition to the 3 DFFs mentioned for the setup time variation, the XNOR-based MUX DFF looks particularly promising. As for the single functional yield, the NAND&INV DFF sticks out as the worst flip-flop for the same reason. Functionality violations that are caused by mismatch in the Monte Carlo simulations are independent of the setup- and hold-time that is used. The minimal functional setup- and hold-time values that are registered for the functionality violations becomes the initial values used for the setup and hold time sweeps, resulting in the stretched 25&75 percentiles of the data-set.

The Power delay product used here for all architectures is the D-Q delay at the found optimal setup-time multiplied by the average power consumption of the switching event. The pass gate flip-flop is, for the same reason as for the minimal clock-period aspect, not at the very top of the list. The use of single transistor pass gates makes the flip-flop highly asymmetrical when it comes to propagation delays and setup times for both logic high and logic low signals. D-Q delay is in this paper found as the worst case D-Q of a falling and rising Q signal event. So even though the pass-gate flip-flop toggles Q to a logic high using a short D-Q delay, the flip-flop is evaluated using its worst case D-Q delay that occurs in the falling Q event as can be seen in the horizontal placement of the upside down delta symbols in Fig. 13. This makes the worst case PDP for the pass gate DFF about 4 times larger than what it would have been for the rising Q signal case. As a result, the two flip-flop designs that have better power-delay products than the pass gate are the PowerPC 603 and the transmission gate DFFs with 52aj and 61aj respectively.

According to Figs. 10 & 11, the MUX-based flip-flop constructed using logic gates is by far the most prone architecture to statistical variations. The 8 other flip-flops prove to be quite robust in this aspect of the analysis. To minimize hold time and setup time violations in a system setting one would normally prefer D flip-flops with small variations in hold- & setup-time. The smaller the variation, the better the functional yield on a system level. The box plots suggest that the pass gate and the PowerPC DFFs are most desired from a setup time variation viewpoint and that the pass gate-, the MUXv2-, the C²MOS-, the PowerPC and the R-Free DFFs are best with respect to hold time variation. Overall the pass gate proves to be surprisingly robust to process variations.

In the result section we listed the leakage power consumption and the energy per cycle of each DFF for multiple operation regions. The top three flip-flops, having the least leakage power consumption in the idle state for the entire swept range of supply voltage, is the pass gate, the PowerPC and the C²MOS -DFFs, the worst is the conventional NAND&INV based DFF most likely due to its high amount of gates and thus relatively low resistance between V_{DD} and GND. When it comes to energy per cycle, the pass gate is best while the conventional NAND&INV based DFF is worst for a supply voltage range of 150 mV to 400 mV. The flip-flops having the best energy consumption per cycle according to Fig. 20 proves to depend almost solely on the supply voltage and not that much on the frequency, with the exception on the voltage range from 0.6V to 0.8V, here the PowerPC 603 performs best for high frequencies while the C²MOS performs better for low frequencies on Pre-Layout simulations. For post layout simulations the S2CFF is best for low frequencies in this region while the C²MOS is best for high frequencies.

The simulation results found here, corresponds quite well to the earlier comparative studies in [9,10] and [11] where the PowerPC

603 DFF sticks out as the best architecture with respect to PDP. The results presented here supports these findings, but also broadens the DFF scope by shedding light on the pass gate flip-flop which performs better than the PowerPC 603 design on a broader basis except for maximum clock speed and energy per cycle outside of the subthreshold supply voltage region. The PDP of the PowerPC design in [10] is equal to 1 femtojoule at 200 mV post-layout, the lower PDP found here in 28 nm FDSOI may be explained by scaling advantages and the fact that the FDSOI process is more energy friendly than common bulk technologies.

5. Conclusion

In this paper various DFF structures are evaluated by their performance in the subthreshold region. The single-transistor pass gate flip-flop scores overall superior results compared to the other DFFs. It takes up silicon space equal to 49% of the S²CFF DFF area with the comparable layout approach described. Simulations point out that the pass gate design can be operated at lower voltages than the other DFFs. It is, however, worse than most DFFs for superthreshold operation, and it is not scoring very well on maximum clock frequency. The PowerPC 603 DFF is with a PDP of 52 attojoules, once again, found to be the flip-flop that has the lowest power-delay product followed by the transmission gate and the pass gate DFFs.

Energy efficiency is a significant challenge for IoT hardware, reducing the energy consumption of wireless sensor circuitry to the point where it can operate perpetually without the need for battery change is crucial to make the various IoT systems maintainable as they scale. Recent developments within the field are plentiful, for instance, in the 2016 January issue of IEEE JSSC, a subthreshold ARM processing sub-system was designed [17] with an energy consumption of 11.7 pJ/cycle as a result of system- and software-level optimization for energy consumption. The authors highlight the importance of low-power circuit and system components for IoT applications. The ARM processing sub-system was implemented with 6.5 % of its cells as state-retention flip-flops (RFF) that consists of a modified version of the conventional master-slave DFF that is analyzed here. According to their source of inspiration [18], the RFF may be implemented using any kind of master-slave DFF that is scan-testable, making each DFF discussed in the work performed here, eligible. By selecting another DFF topology to base the RFF on, they could, according to the study performed here, improve both the power consumption and the overall performance of the ARM processing subsystem even further.

All flip-flop architectures discussed in this paper are yet to be analyzed with silicon measurements, such investigations will reveal impact of any parasitic effects that are neglected by the QRC resistance and capacitance extraction tool used in this work. At the time of writing, the GDS files of a 1mm² test-die has been sent to Circuits Multi-Projects(CMP) for fabrication, the test chip completion- and delivery-date is scheduled to be in the summer of 2016.

References

- [1] E. Late, A.A. Vatanjou, T. Ytterdal, S. Aunet, Comparative analysis of flip-flop architectures for subthreshold applications in 28nm fdsoi, in: *Nordic Circuits and Systems Conference (NORCAS): NORCHIP & International Symposium on System-on-Chip (SoC)*, 2015, IEEE, 2015, pp. 1–4.
- [2] C. Piguet, The First Quartz Electronic Watch, in: B. Hocht, A.J. Acosta, M.J. Belido (Eds.), *Integrated Circuit Design, Power and Timing Modeling, Optimization and Simulation*, no. 2451 in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2002, pp. 1–15.
- [3] Y. Tsvitidis, Eric Vittoz and the Strong Impact of Weak Inversion Circuits, *IEEE Solid-State Circuits Society Newsletter* 13 (3) (2008) 56–58.

- [4] Y. Kim, W. Jung, I. Lee, Q. Dong, M. Henry, D. Sylvester, D. Blaauw, 27.8 A static contention-free single-phase-clocked 24t flip-flop in 45nm for low-power applications, in: Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International, 2014, pp. 466–467.
- [5] F. Klass, Semi-dynamic and dynamic flip-flops with embedded logic, in: VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on, IEEE, 1998, pp. 108–109.
- [6] Y. Suzuki, M. Hirasawa, K. Odagawa, Clocked CMOS calculator circuitry, in: Solid-State Circuits Conference. Digest of Technical Papers. 1973 IEEE International, XVI, 1973, pp. 58–59.
- [7] N. Lotze, Y. Manoli, A 62 mV 0.13 m CMOS Standard-Cell-Based Design Technique Using Schmitt-Trigger Logic, IEEE Journal of Solid-State Circuits 47 (1) (2012) 47–60.
- [8] G. Gerosa, S. Gary, C. Dietz, D. Pham, K. Hoover, J. Alvarez, H. Sanchez, P. Ippolito, T. Ngo, S. Litch, J. Eno, J. Golab, N. Vanderschaaf, J. Kahle, A 2.2 W, 80 MHz superscalar RISC microprocessor, IEEE Journal of Solid-State Circuits 29 (12) (1994) 1440–1454.
- [9] H. Alstad, S. Aunet, Seven subthreshold flip-flop cells, in: Norchip, 2007, 2007, pp. 1–4.
- [10] M. Voernes, T. Ytterdal, S. Aunet, Performance comparison of 5 subthreshold CMOS flip-flops under process-, voltage-, and temperature variations, based on netlists from layout, in: NORCHIP, 2014, 2014, pp. 1–6.
- [11] H. Alstad, S. Aunet, Three Subthreshold Flip-Flop Cells Characterized in 90 nm and 65 nm CMOS Technology, in: 11th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems, 2008. DDECS 2008, 2008, pp. 1–4.
- [12] S. Okugawa, N. Inoue, Systematic design of D flip-flops using two state variables, Electronics Letters 23 (17) (1987) 909–910.
- [13] H. Harris Weste Neil David M., Cmos Vlsi Design: A Circuits And Systems Perspective, 3d, Pearson Education India, 2006.
- [14] S. Farideh, in: Design of Reliable and Efficient Flip-Flops for Subthreshold Operation using Multi-Threshold MOSFETs and Transistor Sizing Technique., 2015.
- [15] V. Stojanovic, V. Oklobdzija, Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems, IEEE Journal of Solid-State Circuits 34 (4) (1999) 536–548.
- [16] T. McConaghy, K. Breen, J. Dyck, A. Gupta, Variation-Aware Design of Custom Integrated Circuits: A Hands-on Field Guide, Springer Science & Business Media, 2012.
- [17] J. Myers, A. Savanth, R. Gaddh, D. Howard, P. Prabhat, D. Flynn, A subthreshold arm cortex-m0+ subsystem in 65 nm cmos for wsn applications with 14 power domains, 10t sram, and integrated voltage regulator (2016).
- [18] M. Keating, D. Flynn, R. Aitken, A. Gibbons, K. Shi, Low power methodology manual: for system-on-chip design, Springer Publishing Company, Incorporated, 2007.



Even Låte graduated from the Norwegian University of Science and Technology (NTNU) with a M.Sc. degree in Electrical Engineering in 2014. His work on "Transaction Level Modeling of a PCI Express Root Complex" was awarded with the best master's thesis prize within the field of microelectronics at NTNU. Since the fall of 2014, Even has been employed as a PhD candidate at the Circuits and Systems Group. His research focus is mitigation of the dark silicon effect by designing ultra low power CMOS memory for use in the "Single-ISA Heterogeneous Many-Core Computer" project at NTNU.



Ali Asghar Vatanjou received his M.Sc. from University of Tabriz in 2010. During his M.Sc. degree, he worked towards pipeline ADC design. Since 2012, he started his PhD at Circuits and Systems Group, Norwegian University of Science and Technology (NTNU). He is currently working towards his PhD thesis at NTNU with main focus on ultra-low power and voltage digital circuit design. His main research interests includes ultra-low power and voltage VLSI and mixed-mode circuit design.



Trond Ytterdal received his M.Sc. and Ph.D. degrees in electrical engineering from the Norwegian Institute of Technology in 1990 and 1995, respectively. He is a Professor at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology. Current main research interests include design of analog integrated circuits and modeling of novel nanoscale transistors. He has authored and co-authored more than 180 scientific papers in international journals and conference proceedings. He is a co-author of three books. Prof. Ytterdal is a member of The Norwegian Academy of Technological Sciences and a Senior Member of IEEE.



Snorre Aunet received the Cand. Scient. degree in informatics from the University of Oslo (UiO) in 1993, and the Dr. Ing. degree in physical electronics from the Norwegian University of Science and Technology (NTNU), in 2002. He worked with ASIC design at Nordic VLSI from 1994 to 1997, prior to doctoral studies. Later, he worked at NTNU and UiO, and is currently a professor at NTNU. He has been a visiting researcher at the University of Paderborn (2004-2012). From August 2015 to July 2016 he stays at University of California, San Diego (UCSD). Research interests include ultra low-power mixed-signal circuits.





A Loadless 6T SRAM Cell for Sub- & near-threshold Operation Implemented in 28 nm FD-SOI CMOS Technology

E. Lâte, T. Ytterdal, and S. Aunet

Published in *Integration*, vol. 63, pp.56-63, 2018¹

DOI: <https://doi.org/10.1016/j.vlsi.2018.05.006>

B.1 Abstract

Most ultra-low power SRAM cells operating in the sub and near-threshold region deploy 8 or more transistors per storage cell to ensure stability. In this paper, we propose and design a low voltage, differential write, a single-ended read memory cell that consists of a total of 6 transistors. The innovative idea is to bring the loadless 4-transistor latch into the realm of low voltage memory cells by exploiting features of the 28nm FDSOI Process and by adding a 2-transistor read buffer with a footer line. Stand-alone and on a system level, the cell is stable during read, write and hold operations, and it has great write-ability due to its differential write and loadless nature. The single NWELL option in 28 nm FD-SOI allows the loadless core to have minimal device widths while greatly improving the time it takes to evaluate the read bit-line. The cell has, in this paper, been used in a 128 kb (217) SRAM in a 16 block configuration, exploring 3 different types of logic libraries for the peripheral logic of the system. Depending on the application, the IO-peripheral logic may be implemented using either high threshold voltage transistors or low threshold voltage transistors in where the power consumption of the 128 kb system was found to range from 1.31 μW to 71.09 μW . The maximum operational frequency lies within 1.87 MHz and 14.97 MHz, while the read energy varies from 13.08 to 75.21 fJ/operation/bit for a supply voltage of 350 mV. The minimum retention voltage of the loadless SRAM cell is found to be 230 mV covering 5σ of variation with Monte Carlo simulations.

¹References in article: [69] [70] [71] [72] [73] [74] [75] [76] [77] [9] [78] [79] [42] [80] [49] [81] [40] [33] [82] [45] [83]

B.2 Relevance to the Thesis

The 4 transistor loadless core together with the 2 transistor read buffer proves to be a viable alternative to state-of-the-art low voltage memory cells with high transistor counts. The high sigma simulation methodology used in this article tackles research-objective O3, mentioned in chapter 3. Stability for memory arrays at low supply voltages is vital to benefit from the energy efficiency for SRAMS at near-threshold operation.

B.3 Declaration of Authorship

The circuit design for the SRAM concept published in this article was initiated by me in early 2016 to interface with a subthreshold RISC-V processor synthesized by master student Åsmund Kvam Oma. At the time, two 2kB SRAMs were designed with the loadless core + read buffer concept to act as the data and instruction memory of the processor core. In 2017 the same SRAM cell concept was used to form the 16-block 2^{17} kb array for architectural comparison to state of the art low voltage SRAM in 28 nm FDSOI. Trond Ytterdal and Snorre Aunet initiated the cooperation with Åsmund and supervised the development of the paper.

B.4 Manuscript



Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

journal homepage: www.elsevier.com/locate/vlsi

A loadless 6T SRAM cell for sub- & near- threshold operation implemented in 28 nm FD-SOI CMOS technology

Even Låte*, Trond Ytterdal, Snorre Aunet

Department of Electronic Systems, Norwegian University of Science and Technology, O.S. Bragstads Plass 2a, 7034 Trondheim, Norway

ARTICLE INFO

Keywords:
Loadless
Near-threshold
Low-power
SRAM
Memory-cell

ABSTRACT

Most ultra low power SRAM cells operating in the sub and near threshold region deploy 8 or more transistors per storage cell to ensure stability. In this paper we propose and design a low voltage, differential write, single ended read memory cell that consists of a total of 6 transistors. The innovative idea is to bring the loadless 4-transistor latch into the realm of low voltage memory cells by exploiting features of the 28 nm FDSOI Process and by adding a 2-transistor readbuffer with a footer line. Stand-alone and on a system level, the cell is stable during read, write and hold operations and it has great write-ability due to its differential write and loadless nature. The single NWEELL option in 28 nm FD-SOI allows the loadless core to have minimal device widths while greatly improving the time it takes to evaluate the read bit-line. The cell has, in this paper, been used in a 128 kb (2^{17}) SRAM in a 16 block configuration exploring 3 different types of logic libraries for the peripheral logic of the system. Depending on the application, the IO-peripheral logic may be implemented using either high threshold voltage transistors or low threshold voltage transistors in where the power consumption of the 128 kb system was found to range from 1.31 μ W to 71.09 μ W, the maximum operational frequency lies within 1.87 MHz and 14.97 MHz while the read energy varies from 13.08 to 75.21 fJ/operation/bit for a supply voltage of 350 mV. The minimum retention voltage of the loadless SRAM cell is found to be 230 mV covering 5 σ of variation with Monte Carlo simulations.

1. Introduction

The number of devices connected to the Internet is predicted to increase from 12.5 billion in year 2010 to 50 billion within year 2020 by Cisco [1]. The explosive growth of on-line devices motivate advances in sensor node design as well as in data center efficiency. To enable this growth, an important area of research is energy efficiency for battery driven devices [2] and power efficiency in data center solutions to mitigate the dark silicon effect [3].

Static Random Access Memories (SRAMs) are heavily utilized in microprocessors as register files, instruction memories and data memories. Thus, SRAMs account for significant fractions of on-chip area and power & energy -consumption in microprocessors and therefore also in data centers and in sensor networks. To improve the energy & power efficiency in sensor networks and in data centers, innovations are needed within the field of SRAM design.

By designing digital circuits to operate in the sub- and near-threshold region one may trade operational speed for power consumption. Typically there is a sweet-spot for the supply voltage in terms of energy per operation that is normally located around the threshold

voltage of the transistor, it is called the minimum energy point (MEP) [4]. For SRAM memories, the voltage where the MEP is located is typically higher than that of static CMOS logic [5], yet it is so low that operation on the MEP raises quite a few challenges where ensuring stability of all bit-cells in the memory arrays during hold, read and write operations are among the main concerns.

To lower the minimum operating voltage (V_{min}) of SRAM cells, architectural innovations are needed. Today most SRAM cells that are made for low supply voltages are made stable by adding read and write assists to the conventional 6T SRAM cells thereby increasing the number of transistors per cell to 8T [6], 9T [7], 10T [8], 11T [9] 12T [10], 14T [11] and even higher numbers. In applications where large amounts of embedded memory is needed, having such large transistor count per memory cell and thus proportionally large area per memory bit leads to infeasible total area consumption. In this paper we therefore propose a 6T SRAM cell architecture that is suitable for low voltage applications. The cell has a retention voltage of 230 mV, and for a 2^{17} bit (128 kb), 16 block, SRAM configuration with I/O peripheral logic made from a custom, low leakage library, a leakage per bit of 9.99 pW and energy per bit accessed in a 64 bit word read-operation of

* Corresponding author.
E-mail address: even.late@ntnu.no (E. Låte).

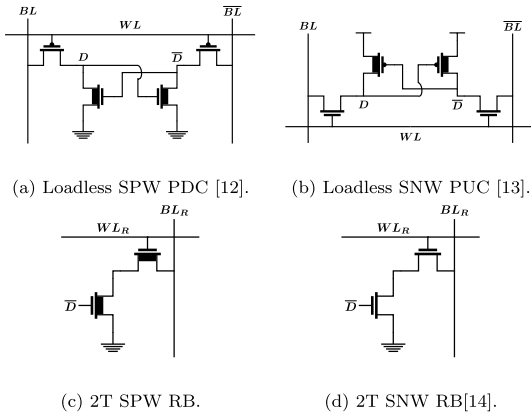


Fig. 1. a & b): Loadless single p-well (SPW) pull down (PD) and single n-well (SNW) pull up (PU) cores that are prone to read failures at low V_{DD} . c) & d): read buffers to mitigate the problem.

13.08 fJ.

In section 2, the proposed cell architecture is explained in detail, the cell sizing is elaborated with respect to the cell's hold and read stability of the cell together with the cell's write ability. The simulation setup is also listed in the same section. Following is the results in section 3 in terms of well established figures of merit for memory systems. The results are discussed in section 4 and conclusions are drawn in the same section.

2. Methodology

2.1. Memory cell design, stability & reliability

2.1.1. Memory cell design

To achieve a low transistor count a loadless 4T cell core was created as shown in Fig. 1(a) by the Nippon Electric Company (NEC) for ultra-high density SRAM macros [12]. NEC based their work on the 4T NMOS memory cell with pull up resistors that has dominated the standalone SRAM market. By removing the resistive loads that requires a dedicated process step, NEC introduced the 4T loadless cell for embedded SRAM applications. The loadless pull down core structure in Fig. 1 (a) operates on a principle of high leakage PMOS pass-gates and low leakage cross coupled NMOS pull down transistors. Data retention is achieved by pre-charging the bitlines to VDD whenever the SRAM is idle. The high leakage PMOS pass gates then provides the cross coupled NMOS pair with supply current. Two design criteria for robust data retention in loadless pull down SRAM cells exist:

- The off current of the NMOS transistors should be significantly lower than the off current of the PMOS transistors in order to store a logic high.
- The on current for the NMOS should be significantly higher than the off current of the PMOS transistors in order to store a logic low.

Similarly for the loadless pull up core in Fig. 1 (b) [13], the bitlines are pre-charged to ground potential whenever the SRAM cell is in the hold state. To retain the stored data, the high leakage NMOS pass gates pulls sufficient supply current from the cross coupled PMOS pair. We have the following two design criteria for data retention:

- The off current of the PMOS transistors should be significantly lower than the off current of the NMOS transistors in order to store a logic high.

- The on current for the PMOS should be significantly higher than the off current of the NMOS transistors in order to store a logic low.

Both cell cores in Fig. 1 (a) and (b) were at first implemented on a schematic level. To fulfill the design criteria itemized above: Low threshold voltage PMOS devices (LVTNPFETs) were used as access gates and regular threshold voltage devices (RVTNPFETs) were used as pull down devices for the loadless pull-down cell. And for the pull-up cell, LVTNPFETs were used as access gates and RVTNPFETs were used as pull up devices. Exploiting the flip-well structure of the LVT devices in 28 nm FD-SOI allows the entire bitcell to be placed in a single P-well and a single N-well for the pull-down SRAM cell and for the pull-up cell respectively. Simulations proved that both loadless cores failed to retain their memory state during read operations for a supply voltage of 350 mV. The differential read, differential write -cores were thus made into single ended read, differential write -cores by adding the single ended read buffers in Fig. 1 (c) and (d) [14]. The transistor count is increased from 4T to 6T to completely isolate the memory cells during read operations, equating the read margin to the hold margin for both cores.

The main issue with operating the conventional read buffer architectures in the near-threshold and subthreshold -regions, is the low I_{on}/I_{off} ratios of the transistors in the read buffers. The low on-current and relatively high off-current makes it difficult to distinguish the voltage differential on the bitlines caused by the non-selected rows in the same column from the read voltage differential generated by the accessed read buffer. Fig. 2 illustrates the bitline leakage problem. The higher the bitline is, in terms of column bitcells, the more cells are not accessed but still contribute to pulling down the pre-charged capacitance which again makes it more difficult to distinguish between logic 1s and 0s for the read-out logic. Since the readout buffer is single ended and, let us say that it is connected to the \bar{D} node in the loadless SRAM core, the worst case scenario occurs when the accessed cell has $D = High$ with a corresponding $\bar{D} = Low$ and if all other cells in the column have $D = Low$, $\bar{D} = High$ ie. open buffer transistors and closed access transistors. The bitline is in this case supposed to retain its value since the accessed read buffer transistor is turned off, however BL_R ends up getting discharged by the sum of leakage currents from the unaccessed cells. This is also furthermore amplified by the impact on the I_{on}/I_{off} ratios from process voltage and temperature (PVT) variations. Fig. 3(a) depicts that the I_{on}/I_{off} ratio of an LVTNPFET transistor varies by several orders of magnitude over the industrial temperature range from $T = -40^\circ C$ to $T = 85^\circ C$ with a supply voltage variation range of $\pm 10\%$. Process and mismatch variation adds even more variation to the I_{on}/I_{off} ratio which can be seen in (b).

To be able to perform read operations in the worst case read

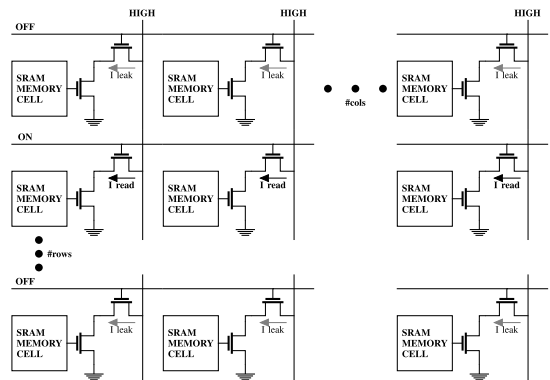
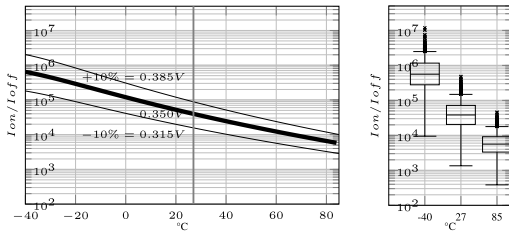
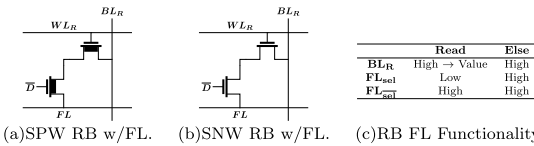


Fig. 2. Read BL problem for conventional 2T read buffers.



(a) I_{on}/I_{off} , T & V -variation. (b) I_{on}/I_{off} , P-variation

Fig. 3. PVT variations of the I_{on}/I_{off} ratio for an LVTNFET. In a) V_{DS} & $V_{GS,off}$ varies $\pm 10\%$ from the supply voltage of 350 mV, the bold line in a) and for plot b).



(a) SPW RB w/ FL. (b) SNW RB w/ FL. (c) RB FL Functionality.

Fig. 4. Used read buffer architecture and its functionality [15].

situation mentioned above, a dynamic read footer line (FL) is used instead of the constant ground potential in the read buffer scheme to achieve a V_{DS} of 0 for all un-accessed read buffers. The footer-line read buffer scheme is depicted in Fig. 4. FL is, during the bitline evaluation phase of a read cycle, pulled to ground potential for the selected (sel) word-line. This eliminates the leakage from unaccessed readbuffers (sel) and enables fast decision making in read operations [6].

Since the carrier mobility of N-channel devices is greater than that of P-channel devices, the loadless SPW pull-down core with LVTNPFET passgates, would require very wide pmos access transistors, according to simulations 7.5 times the minimum width to satisfy the leakage design criteria introduced above. Instead, the SNW solution utilize the LVTNFET devices as the high leakage passgate allowing it to have minimum width, The SPW cell is therefore discarded. The remaining SNW loadless SRAM cell is not just smaller than the SPW, it also has the beneficial advantage of having LVTNFETs in the SNW readbuffer as well, greatly improving the bitline pulldown time. The complete schematic of the proposed 6T SNW loadless SRAM cell is shown in Fig. 5. It is realized with all minimum channel widths and with gate lengths equal to 48 nm reducing cell leakage. With this configuration, the cell outperforms an implemented reference version of the 8T conventional subthreshold SRAM cell [6], both on the bitcell's minimum retention V_{DD} by 9% and on the bitcell leakage by 51.8% at $V_{DD} = 0.35$ V. This is despite the reduction in transistor count from 8T to 6T. The reference cell was also made of all minimum width, 48 nm length mixed V_{th} transistors in a SNW, and the comparison was done by running transient simulations with the write bitlines of both cells set to ground potential.

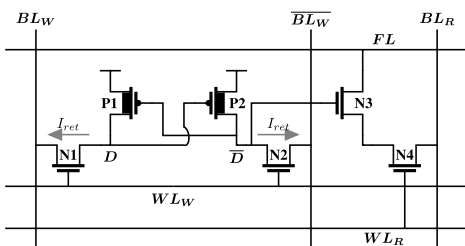


Fig. 5. The proposed 6T loadless SRAM cell.

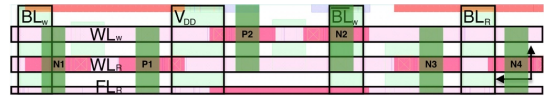


Fig. 6. Layout of the proposed 6T loadless SRAM cell with shared RB footer line. Red = poly/metal gate, green = active area, light-blue = metal 1, light-green = metal 2, salmon = metal 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The memory cell is laid out in a flat, twisted and skewed layout scheme shown in Fig. 6. By twisted we mean that the area consuming cross coupling of the cell's core is avoided by twisting the right side 180° around the horizontal axis, the halves of the SRAM cell are then skewed relative to each other to allow the gate of the right pull-up device to be connected to the drain of the left pull up device with a straight wire and vice versa. Also by making the layout in a flat manner with a horizontal row of transistors we avoid RVT to LVT spacing rules in the vertical direction, saving area directly. The height of the SRAM cell is limited by 3 minimum sized metal lines and 3 minimum sized line distances. To save one horizontal poly-poly spacing, the cell is flipped horizontally and the readbuffer pass gate poly is shared between neighboring columns in Fig. 7, in addition, the cell is flipped vertically to share supply rail connections between neighboring rows. By grouping 4 bitcells in one layout cell, memory matrices may be created with the mosaic functionality in the virtuoso layout editor achieving a SRAM cell area of $0.3 \mu\text{m} \times 1.824 \mu\text{m} = 0.5472 \mu\text{m}^2/\text{bit}$. Again, to provide a comparative metric, the layout of the conventional subthreshold 8T cell [6], was drawn using the single poly row approach. The area of this cell was found to be $0.3 \mu\text{m} \times 2.306 \mu\text{m} = 0.6918 \mu\text{m}^2/\text{bit}$, which reveals an area reduction for the loadless 6T core proposed in this paper of 20.90% in comparison. A benefit of choosing such a wide bitcell scheme is reduced bitline capacitance. Nodal capacitance for both pre and post-layout netlists are given in Table 1, the post-layout capacitances are generated using the Cadence QRC layout extraction tool with the decoupled option activated for nominal, minimum and maximum corner scenarios. The above assumptions of low bitline capacitances for wide SRAM cells is confirmed by the difference between the post-layout BL capacitance and the post layout FL capacitance listed in the table. The corresponding penalty of choosing the flat layout scheme is increased WL capacitance which is also shown in the same table. It is however better to drive higher WL capacitances from a stronger row-decoder rather than increasing the size of the read buffer to pull down the higher BL capacitances through the area- and leakage- constrained memory cell. The post layout parasitic capacitance was also found for the 8T conventional subthreshold reference cell. The loadless 6T SNW cell has, with its retention current to the write bitlines, no need for vertical ground lines, reducing the C_{FL} by 21.3%, C_{WL_R} by 17.6% and C_{WL_W} by 16.3% in comparison to the 8T cell. This again contributes to a lower dynamic power consumption since less parasitic capacitance is switched on memory operations.

2.1.2. Biasing the SRAM cell's SNW

Since all devices in the SRAM cell are placed in a single well, they all inherit the same back gate bias conditions. By biasing the back gate of the entire SRAM cell, the threshold voltage of the RVTPFETs increases with increasing back gate voltage while the threshold voltage of the

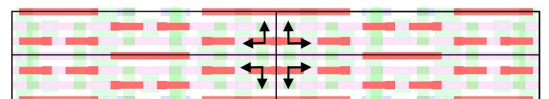
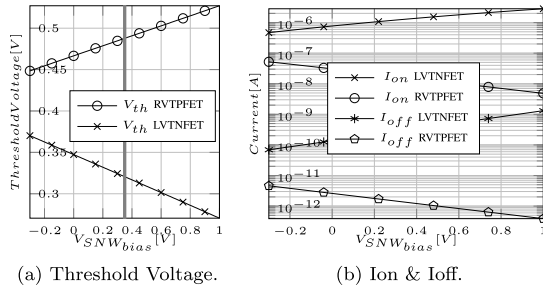


Fig. 7. Layout of 4 Cells flipped horizontally and vertically to create mosaicable cells.

Table 1
 Decoupled nodal capacitances on bitcell from extracted net-list.

	C_{BLW} [aF]	C_{BLW}^- [aF]	C_{WLW} [aF]	C_{BLR} [aF]	C_{WLR} [aF]	C_{FLR} [aF]
Pre-Layout	41.84	41.84	135.91	41.8366	67.95	41.84
Post-Layout nom	102.60	116.77	478.36	113.52	433.25	329.17
Post-Layout max	120.22	136.59	567.47	132.44	530.91	389.91
Post-Layout min	89.12	101.87	410.64	99.74	361.69	283.05


Fig. 8. Design knob, biasing the SNW from -0.3V to 1V at $V_{DD} = 350\text{ mV}$.

LVTNFET decreases as Fig. 8 (a) shows. This again affects the on and off currents of the RVT and LVT devices in the cell in an inverse relationship as shown in (b). I_{on} LVTNFET & I_{off} RVTPEFET increases with increasing $V_{SNWbias}$, while I_{on} RVTPEFET & I_{off} LVTNFET decreases with increasing $V_{SNWbias}$. Since the LVT pass gates are used as SRAM load, and since the pull up devices react inversely to a change in the bias condition, the bias voltage therefore impact the logic high voltage level of the memory cell. A logic high voltage close to that of the pull up voltage is desired for data retention.

2.1.3. Hold & read stability

To maximize the stability of SRAM cells in low supply voltages, it is desirable to strengthen the static noise margin (SNM) of the internal latch, both while retaining and reading data. The hold SNM for a conventional SRAM cell is defined as the maximum value of V_n in Fig. 9(a) that the cell may tolerate without flipping its stored state unintentionally [16]. The hold SNM definition of the loadless 6T SRAM cell is directly derived from that of the conventional SRAM latch and is shown as V_n in Fig. 9(b), note that the latch is disconnected from the read bitline when retaining data. Since this also is the case during read operations on the SRAM cell, the same definition may be used for the SNM in the read state making the SRAM cell implemented here as read stable as it is hold stable. We can model static noise margins of SRAMs analytically by finding the voltage transfer curve (VTC) of each of the SRAM's compromising inverters, and then, after inverting one of the VTCs, by finding the length of the side of the maximal square fitted inside each eye of the resulting butterfly plot.

For SRAMs operating below the threshold voltage the VTCs of the inverters may be derived using the widely adopted model for

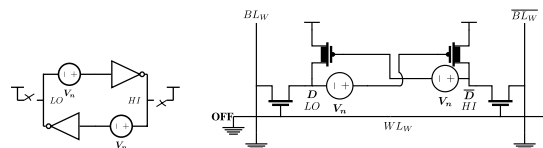

(a) General SNM. **(b)** Loadless Core SNM.

Fig. 9. Hold and read SNM for the loadless 4T pull up core.

subthreshold drain current in MOSFETs and Kirchoff's current law [17].

$$I_{DS} = I_S e^{\frac{V_{GS}-V_{th}}{nV_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}}\right) (1 + \lambda V_{DS}) \quad (2.1)$$

The exponential relationship between the drain current and node voltages for an NMOS transistor in the subthreshold region is given by Eq. (2.1). V_{th} , V_{GS} and V_{DS} is the threshold voltage, the gate to source and the drain to source voltage respectively. V_T is the thermal voltage given by $V_T = kT/q$ for an absolute temperature T , with the Boltzmann constant, k , and the elementary charge q . n is the subthreshold slope factor that is equal to $1 + C_D/C_{ox}$. I_S is the current that flows through the transistor for a $V_{GS} = V_{th}$. We included a model for the Drain Induced Barrier Lowering (DIBL) effect to get a more realistic slope on the VTCs, the incorporated factor is $1 + \lambda V_{DS}$, where λ is the output impedance constant of the transistor.

$$I_{DSn} = I_{Dsp} \quad (2.2)$$

For an arbitrary loadless inverter in the loadless SRAM cell, the current through the pullup PMOS is equal to the leakage current through the LVT NMOS in both the retention state and in the read state, Eq. (2.2). Inserting Eq. (2.1) into Eq. (2.2) yields Eq. (2.3), it is further transformed into the closed-form expression in to obtain the VTC for the loadless inverter.

$$\begin{aligned} I_{Sn} e^{\frac{-V_{in}}{n_n V_T}} \left(1 - e^{-\frac{V_{out}}{V_T}}\right) (1 + \lambda_n V_{out}) \\ = I_{Sp} e^{\frac{V_{DD}-V_{in}-|V_{thp}|}{n_p V_T}} \left(1 - e^{-\frac{V_{DD}+V_{out}}{V_T}}\right) (1 + \lambda_p (V_{DD} - V_{out})) \end{aligned} \quad (2.3)$$

$$\begin{aligned} V_{in} = V_{DD} - |V_{thp}| - \\ \ln \left(\frac{I_{Sn} e^{\frac{-V_{in}}{n_n V_T}} \left(1 - e^{-\frac{V_{out}}{V_T}}\right) (1 + \lambda_n V_{out})}{I_{Sp} e^{\frac{V_{DD}-V_{in}-|V_{thp}|}{n_p V_T}} \left(1 - e^{-\frac{V_{DD}+V_{out}}{V_T}}\right) (1 + \lambda_p (V_{DD} - V_{out}))} \right) n_p V_T \end{aligned} \quad (2.4)$$

Parameters for the minimum sized LVTNFET and the minimum sized RVTPEFET are extracted from the 28 nm FDSOI kit and are listed in Fig. 10(a). These are used to generate the butterfly plots in Fig. 10(b). It can be observed from the figure that the plots have large eyes, for the targeted supply voltage of 350 mV the SNM for read and hold is equal to 125 mV.

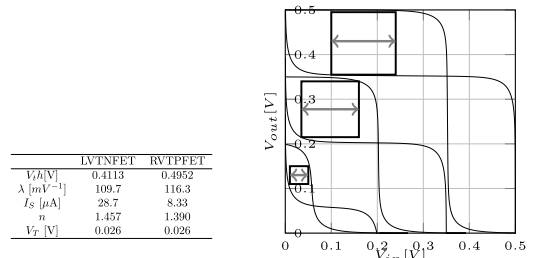

(a) Model Parameters. **(b)** Butterfly Plots varying V_{DD} .

Fig. 10. Model Parameters from the targeted area of operation with 0V back-gate bias and butterfly plots for an analytical evaluation of operation read & hold SNM.

E. Lâte et al.

INTEGRATION, the VLSI journal 63 (2018) 56–63

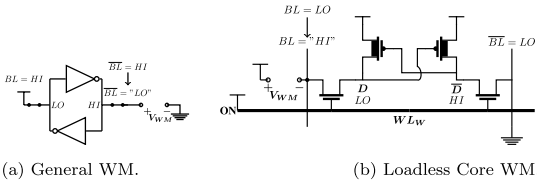


Fig. 11. Write margin for the loadless 6T SRAM cell.

In addition to the analytical analysis of the read and hold SNM, transient simulations were performed with the data node D and \bar{D} in Fig. 9(b) initialized to a logic low and a logic high respectively, this is given by the direction of the static noise sources in the figure. As transient static noise sources, verilog-A voltage ladders were used with long enough settling times for the circuit to settle to its steady state in between voltage increments. The read/hold-SNM is compared that of an implemented reference 6T cell in additional simulations.

2.1.4. Writability

For conventional 6T SRAM cells the bit-lines are pre-charged to V_{DD} in data retention mode and in read mode. To write to the bit-cell, one of the bit-lines is pulled low while the other one remains at V_{DD} . Fig. 11(a) depicts the write operation of the conventional SRAM latch, here the write margin, V_{WM} , is defined as the largest logic low bit-line voltage that intentionally flips the state of the SRAM cell [18]. The higher the write margin, that is, the less one of the bitline has to be pulled down in order to write data to the cell, the greater the write-ability of the cell.

An analytical expression for the write margin of the cell can not be obtained with a similar methodology as in the previous section, this is due to the fact that there are no pull-down transistors in the latch active in the write phase, greatly improving the writability. The loadless cell proposed here has a default precharged bit-line voltage at ground potential to retain the data within the cells in hold mode and in read mode. During a write operation, the desired bit-line should be pulled up and not down in order to flip the state of the cell, this implies a corresponding change in the write margin definition. The write margin of the loadless pull-up cell is equal to V_{DD} minus the required voltage increment on the bit-line to flip the cell. The less the bit-line has to be pulled up for a write operation, the greater the writability is for the cell.

To find the write margin with simulations for the loadless 6T SRAM cell, transient write simulations are performed with the data node D and \bar{D} in Fig. 11(b) initialized to a logic low and a logic high respectively. The dc value of the bit-line BL is incremented in consecutive transient simulations until the cell succeeds to toggle and store the state of the cell after WL_W goes low again and closes the pass-gates. Also in the case of writability, the loadless core is compared to that of an implemented conventional 6T cell.

2.1.5. Write stability for unaccessed cells in the accessed column

The loadless core retains its data by holding the write bitlines at ground potential, forcing a voltage across the series RVT/PFET and the LVT/NFET of VDD, this gives origin to the retention current from VDD to the write bitlines. During a write operation, all memory nodes that are connected to the bitline that is pulled up, and that stores a logic 0 on the pulled up side, relies on their internal data node capacitance to retain the stored data during the column write access. This implies a maximum time that a bitline may be kept high, and therefore a timing requirement for write operations. Monte Carlo simulations of the nodal voltage-increase with a defined upper limit of the logic low node of 50% give rise to a maximum allowed bitline high-time of 6.077 μ s as Fig. 12(b) shows. The Monte Carlo run is here performed on the worst case internal nodal capacitance, that is the lowest value of the extraction corners listed in Fig. 12(a).

	C_D [aF]	$C_{\bar{D}}$ [aF]
Pre-Layout	146.74	214.69
Post-Layout nom	169.17	219.75
Post-Layout max	198.47	258.04
Post-Layout min	147.44	192.13

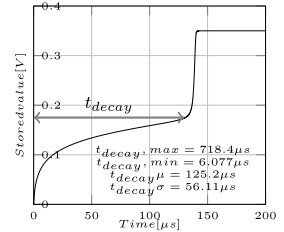


Fig. 12. Maximum allowable time a bitline may be pulled up during a memory write operation for a full write-bitline swing.

2.2. Memory matrix, peripheral logic & timing

2.2.1. Choosing the memory matrix dimensions

The number of rows in an array of SRAM cells significantly impact the time it takes to evaluate the read bit lines. To illustrate this aspect of the dimensioning of the SRAM configuration, the time it takes to discharge the read bitlines for 16 to 512 words is given in Fig. 13. The worst case extracted capacitance corner is used in the Monte Carlo analysis.

For this work, to enable comparison to existing, state of the art, ultra low voltage SRAMs [19–21], a 128 kb SRAM (2^{17}) is built up from 16×8 kb SRAM sub-arrays of the loadless 6T memory cell. The dimensions of each sub-array or “block” is set to be 128 rows of 64-bit words to fit in modern 64 bit CPU architectures.

The $128 \times 64 \times 16$ system configuration that is used as a benchmark for the 6T loadless SRAM cell is shown in Fig. 14. It is implemented with 3 different custom logic libraries designed for a supply voltage of 350 mV. The first using minimum sized LVT transistors, the second using minimum sized RVT transistors and the third using RVT transistors with gate lengths of 48 nm trading off speed for power consumption. All libraries are implemented with switching points of $V_{DD}/2$. The timing diagram for read and write operations on the 2^{17} SRAM system is shown in Fig. 15. The timing module is written as a 16-state state-machine in VHDL and is synthesized using Synopsys design compiler.

2.3. Simulation setup

All simulations are performed with $gmin = 10^{-18}$, $reitol = 10^{-9}$, $vabstol = 10^{-9}$ & $iabstol = 10^{-15}$ to achieve accurate current simulation accuracy in the femto-amp range. Mismatch and process Monte Carlo simulations with specific sigma accuracy are performed using the Low-Discrepancy Sequence (LDS) sampling algorithm with yield targets set to the desired expected fraction of population inside range. All Monte Carlo simulations that imply functionality and yield are performed with 5σ accuracy. A sigma accuracy of 5 corresponds to a LDS

#WLS	μ	σ	min	max
16	6.9	3.36	1.7	28.2
32	13.6	6.7	3.4	55.9
64	27.1	13.2	6.7	111.2
128	54.0	26.4	13.3	221.8
256	107.9	52.7	26.6	443.1
512	215.7	105.3	53.1	885.6

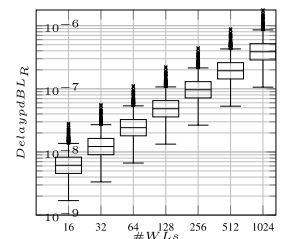


Fig. 13. Read Delay vs # of WLS in 3σ MC simulations.

E. Lâte et al.

INTEGRATION, the VLSI journal 63 (2018) 56–63

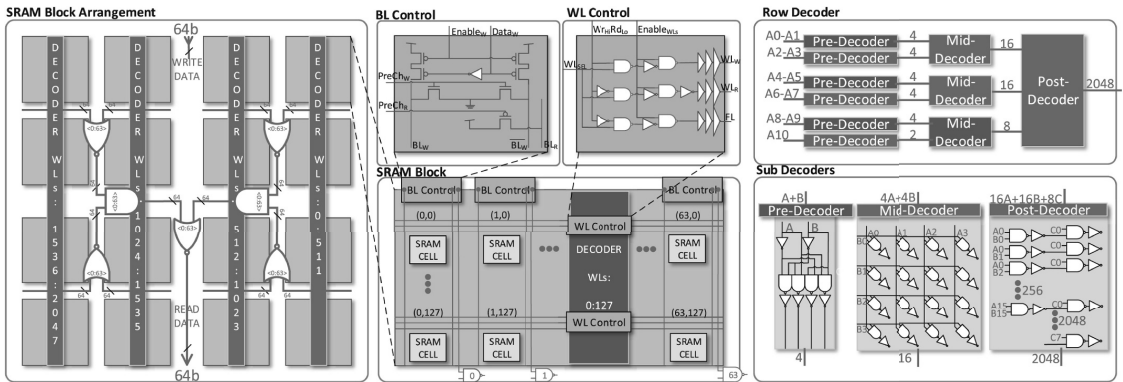


Fig. 14. 2¹⁷ bit SRAM array structure for a configuration of 128 rows × 64 columns × 16 blocks.

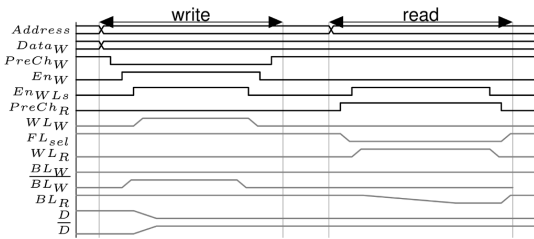


Fig. 15. Timing diagram for a write and a read operation on the SRAM system in Fig. 14.

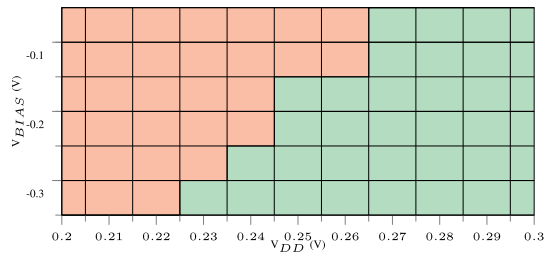


Fig. 18. 5 σ shmoo plot of data retention for V_{BIAS} vs V_{DD} .

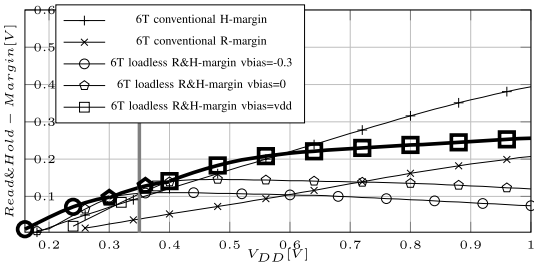


Fig. 16. Read and hold margin for the loadless and conventional SRAM cells.

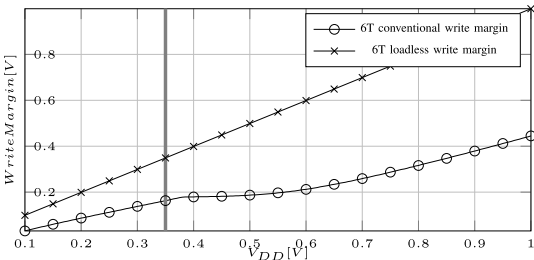


Fig. 17. Write margin for the loadless and conventional SRAM cells, loadless is independent of v_{bias}.

yield target of 99.99994267%, that is, 1 in 1744278 samples lies outside of the simulation results which is a suitable number for a 2¹⁷ bits SRAM.

All experiments on the total 2¹⁷ SRAM configuration are performed with a reduced netlist where only the accessed row and column are instantiated with extracted models, this to reduce the simulation time to a feasible level. To achieve an accurate energy-per-operation metric, the read-time is multiplied by the 5 σ mean idle leakage power of all removed non-activated cells and the resulting energy is then added to the system's simulation result. This was done for the 15 non-selected SRAM blocks and for the SRAM cells that are not connected to the activated bit-line or the activated word/foot-lines of the selected block.

3. Simulation results

3.1. Robustness of the loadless 6T SRAM cell

The hold and read stability of the SRAM cell is given in Fig. 16 as a function of supply voltage for different back gate bias potentials. In the same figure the read and hold SNM of a conventional 6T SRAM cell are used as a comparative metric. The bold line indicates the best biasing condition for the loadless SRAM cell at given values of V_{DD}. It can be noted that the bold line is superior in comparison to the read margin of the conventional SRAM cell for all values of the supply voltage. For a V_{DD} of 350 mV a bias voltage of 0 V gives the best hold and read margin for the loadless core.

The write margin of the loadless core is quite interestingly found to be close to V_{DD} for all values of V_{DD} as Fig. 17 shows. For a 350 mV voltage supply, the write margin is, thanks to the loadless nature of the cell, found to be approximately equal to 349 mV.

3.2. Minimum supply voltage

Fig. 18 is made from multiple 5 σ monte carlo simulations revealing that the minimum retention voltage of the SRAM is 230 mV for a single NWELL bias of -300 mV.

Table 2
Systems with different I/O implementations.

I/O Logic Library	Idle Leakage				Active Speed		Active Energy	
	Leakage SRAM Cells [μ W]	Leakage I/O [μ W]	Total leakage [μ W]	Leakage [pW/bit]	Delay Read [ns]	Max. Rd. Freq [MHz]	Read Energy [fJ/OP]	Read Energy per bit [fJ/OP/bit]
LVT30	1.07	70.02	71.09	542.37	66.79	14.97	4813.20	75.21
RVT30	1.07	6.57	7.64	58.29	109.75	9.11	908.56	14.19
RVT48	1.07	0.24	1.31	9.99	535.00	1.87	837.52	13.08

Table 3
Comparison table to state of the art SRAMs. * = simulated results. @ = lack of access to SRAM transistors.

Ref.	Tech. Node	T.-count [T/bit]	Cell Area [μ m ² /bit]	BL Height [Cells/BL]	Min. V _{DD} [mV]	Max. R-Freq. [Hz]	Lk. Pwr. [pW/b]	Rd. Energy [fJ/OP/b]
*@ LVT30	28 nm FD-SOI	6	0.547	128	230	14.97 M (0.35 V)	542.37 (0.35 V)	75.2 (0.35 V)
*@ RVT30	28 nm FD-SOI	6	0.547	128	230	9.11 M (0.35 V)	58.29 (0.35 V)	14.2 (0.35 V)
*@ RVT48	28 nm FD-SOI	6	0.547	128	230	1.87 M (0.35 V)	9.99 (0.35 V)	13.1 (0.35 V)
[19]	28 nm FD-SOI	10	0.384	64	350	13 M (0.35 V)	N/A	71.8 (0.35 V)
[20]	28 nm FD-SOI	6	0.232	32	360	9 M (0.36 V)	N/A	52.5 (0.45 V)
[21]	28 nm FD-SOI	7	0.261	64	200	90 M (0.30 V)	0.9 (0.30 V)	8.4 (0.30 V)

3.3. Power and energy consumption on a system level

Table 2 shows the idle leakage power for the 128 kb SRAM and the maximum frequency & energy consumed for read operations. Results for the entire SRAM structure with I/O peripheral circuitry implemented with the LVT30 library, the RVT30 library and the RVT48 library are given.

4. Discussion & conclusion

In this paper we have proposed a 6T SRAM cell for low voltage applications. The construction of the cell is reasoned for in the methodology section, it consumes 6 transistors and is built up from a loadless 4T core and a 2T read buffer. Fig. 16 reveals that the loadless structure dominates the read and hold margin of the conventional 6T cell for low voltages. For higher voltages the opposite is true unless a positive bias voltage is applied equal to that of the supply voltage. This is to correct for the change in the leakage ratios of the RVT pull up devices and the LVT access devices. Fig. 18 confirms the effect of reducing the SNW bias voltage on the hold margin. The cell retains its data at supply voltages as low as 230 mV over 5 σ variation with process & mismatch variation taken into account.

It can furthermore be observed from Fig. 17 that the proposed cell inherits close to perfect write margins, this is due to its loadless nature. Since the loading of the loadless SRAM cell occurs as off-state leakage through the LVTNFET pass gates, once the pass gates are opened, the content of the cell is forced to take the differential value of the differential bitlines without any load working against the write operation. The loadless nature of the loadless core acts as a write-assist, and allows very small differential voltages to be used for write operations, potentially saving power and write-delay since a full swing bit-line scheme may be relaxed.

3 different SRAM implementations were made using 3 classes of logic libraries varying in terms of speed and leakage. From Table 2, one can see that the choice of logic libraries for the I/O peripherals greatly impacts all major SRAM performance metrics. The library that yields the best energy per operation per bit has a low operational speed. Choosing an I/O library is therefore highly application specific. Slow applications that infrequently performs memory operations, may benefit from the RVT48 library while fast applications that writes and reads in a high frequent manner may be forced to choose the implementation with the LVT peripheral circuitry.

A comparison to state of the art low voltage SRAMs is provided in Table 3. Finding the best SRAM out of a few is ambiguous, especially

when one compares conservative simulated results to results obtained through physical measurements. Even though such a comparison is not considered fair, it does provide enough information to decide whether or not one should pursue the idea further.

Slight variations in system implementations may result in great differences in well known metrics. For this implementation conventional RVT and LVT transistors are used, in the 28 nm FD-SOI technology a set of more restricted dedicated SRAM transistors are also available that have relaxed layout rules and thicker oxide. The authors of [21] and the authors of [20] collaborate with ST Microelectronics while the author of [19] worked for STM at the time of writing and it is unclear from their papers whether they have access to the dedicated SRAM transistors or not. This may be a reason for their low reported cell-area. The purpose of bringing this up here is that greater leakage and area improvements is still possible for the proposed memory cell by porting the design to the dedicated SRAM transistors. Another reason for the difference in area consumption is the poly strip configuration, in this paper a flat layout scheme was chosen where only a single horizontal row of 6 poly strips makes up the memory cell, this was to minimize the bitline capacitance in order to reduce the read out time which is limited by the read buffer. It is possible to create a two-row configuration with 3 polystrips in each so that the following pins are shared with the neighbour cells above and below: top: BL_W , BL_W , BL_R bot: V_{DD} , V_{DD} , FL . Such a physical configuration should have less area at the cost of a greater read out delay. As we saw in section 2, the area of the proposed cell is 20.7% smaller than the implemented 8T reference cell.

The 2¹⁷-bit array of loadless 6T SRAM Cells has performance and energy metrics according to the rightmost columns of Table 3. Using the LVT30 library for the peripheral logic results in a maximum frequency of 14.97 Mhz at 0.35 V and a read energy of 75.20 fJ/OP/bit, however the LVT30 implementation has a very high retention leakage and the peripheral logic should be power gated. Compared to the LVT30 design, the RVT48 memory system consumes roughly 1/54th of the retention power per bit with a speed reduction of only one 8th which results in the highest energy efficiency of the designs proposed here. If a charge pump scheme such as the one in Ref. [21] was applied to the read word-line of the SRAM proposed here, a great reduction of the read out delay and thus a higher energy efficiency on read operations would be expected at the cost of higher complexity.

The true yield of the SRAM remains to be found in a physical prototype of the system, but according to statistical simulations with 5 σ accuracy, the proposed loadless 6T cell performs on par with state of the art low voltage SRAMs using conventional figures of merit.

E. Låte et al.

INTEGRATION, the VLSI journal 63 (2018) 56–63

References

- [1] D. Evans, The Internet of Things, Whitepaper, Cisco Internet Business Solutions Group (IBSG) 1 (2011), pp. 1–12.
- [2] L. Atzori, A. Iera, G. Morabito, The internet of things: a survey, *Comput. Netw.* 54 (15) (2010) 2787–2805.
- [3] H. Esmailzadeh, E. Blem, R.S. Amant, K. Sankaralingam, D. Burger, Dark silicon and the end of multicore scaling, *ISCA* 2011, *IEEE*, 2011, pp. 365–376.
- [4] B.H. Calhoun, A. Wang, A. Chandrakasan, Modeling and sizing for minimum energy operation in subthreshold circuits, *IEEE JSSC* 40 (9) (2005) 1778–1786.
- [5] G. Chen, D. Sylvester, D. Blaauw, T. Mudge, Yield-driven near-threshold sram design, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 18 (11) (2010) 1590–1598.
- [6] N. Verma, A.P. Chandrakasan, A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy, *IEEE JSSC* 43 (1) (2008) 141–149.
- [7] M.-H. Tu, J.-Y. Lin, M.-C. Tsai, C.-Y. Lu, Y.-J. Lin, M.-H. Wang, H.-S. Huang, K.-D. Lee, W.-C. Shih, S.-J. Jou, et al., A single-ended disturb-free 9t subthreshold sram with cross-point data-aware write word-line structure, negative bit-line, and adaptive read operation timing tracing, *IEEE JSSC* 47 (6) (2012) 1469–1482.
- [8] J.P. Kulkarni, K. Kim, K. Roy, A 160 mv robust schmitt trigger based subthreshold sram, *IEEE JSSC* 42 (10) (2007) 2303–2313.
- [9] J. Chen, L.T. Clark, T.-H. Chen, An ultra-low-power memory with a subthreshold power supply voltage, *IEEE JSSC* 41 (10) (2006) 2344–2353.
- [10] Y.-W. Chiu, Y.-H. Hu, J.-K. Zhao, S.-J. Jou, C.-T. Chuang, A subthreshold sram with embedded data-aware write-assist and adaptive data-aware keeper, *ISCAS* 2016, *IEEE*, 2016, pp. 1014–1017.
- [11] S. Hanson, M. Seok, Y.-S. Lin, Z. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, D. Blaauw, A low-voltage processor for sensing applications with picowatt standby mode, *IEEE JSSC* 44 (4) (2009) 1145–1155.
- [12] K. Noda, K. Matsui, K. Imai, K. Inoue, K. Tokashiki, H. Kawamoto, K. Yoshida, K. Takeda, N. Nakamura, T. Kimura, et al., A 1.9 μm^2 loadless cmos four-transistor sram cell in a 0.18 μm logic technology, 1998. *IEDM* 1998, *IEEE*, 1998, pp. 643–646.
- [13] X. Deng, T.W. Houston, Loadless 4t sram cell with pmos drivers, *uS Patent* 6,731, 533 (May 4 2004).
- [14] L. Chang, D.M. Fried, J. Hergenrother, J.W. Sleight, R.H. Dennard, R.K. Montoye, L. Sekaric, S.J. McNab, A.W. Topol, C.D. Adams, et al., Stable sram cell design for the 32 nm node and beyond, *VLSI Technology*, 2005. *Digest of Technical Papers. 2005 Symposium on*, *IEEE*, 2005, pp. 128–129.
- [15] J. Kwong, Y.K. Ramadass, N. Verma, A.P. Chandrakasan, A 65 nm sub-micro-controller with integrated sram and switched capacitor dc-dc converter, *IEEE JSSC* 44 (1) (2009) 115–126.
- [16] E. Seevinck, F.J. List, J. Lohstroh, Static-noise margin analysis of mos sram cells, *IEEE JSSC* 22 (5) (1987) 748–754.
- [17] R.M. Swanson, J.D. Meindl, Ion-implanted complementary mos transistors in low-voltage circuits, *IEEE J. Solid State Circ.* 7 (2) (1972) 146–153.
- [18] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, M. Bohr, A 3-ghz 70-mb sram in 65-nm cmos technology with integrated column-based dynamic power supply, *IEEE JSSC* 41 (1) (2006) 146–151.
- [19] F. Abouzeid, A. Bienfait, K.C. Akyel, A. Feki, S. Clerc, L. Ciampolini, F. Giner, R. Wilson, P. Roche, Scalable 0.35 v to 1.2 v sram bitcell design from 65 nm cmos to 28 nm fdsol, *IEEE JSSC* 49 (7) (2014) 1499–1505.
- [20] A. Biswas, A.P. Chandrakasan, A 0.36 v 128kb 6t sram with energy-efficient dynamic body-biasing and output data prediction in 28nm fdsol, *ESSCIRC Conference* 2016: 42nd, *IEEE*, 2016, pp. 433–436.
- [21] B. Mohammadi, O. Andersson, J. Nguyen, L. Ciampolini, A. Cathelin, J.N. Rodrigues, A 128 kb single-bitline 8.4 fJ/bit 90mhz at 0.3 v 7t sense-amplifierless sram in 28 nm fd-soi, *ESSCIRC Conference* 2016: 42nd, *IEEE*, 2016, pp. 429–432.



Even Låte graduated from the Norwegian University of Science and Technology (NTNU) with a M.Sc. degree in Electrical Engineering in 2014. His work on “Transaction Level Modeling of a PCI Express Root Complex” was awarded with the best master’s thesis prize within the field of microelectronics at NTNU. Since the fall of 2014, Even has been employed as a PhD candidate at the Circuits and Systems Group. His research focus is mitigation of the dark silicon effect by designing ultra low power CMOS memory for use in the “Single-ISA Heterogeneous Many-Core Computer” project at NTNU.



Trond Ytterdal received his M.Sc. and Ph.D. degrees in electrical engineering from the Norwegian Institute of Technology in 1990 and 1995, respectively. He is a Professor at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology. Current main research interests include design of analog integrated circuits and modeling of novel nanoscale transistors. He has authored and co-authored more than 180 scientific papers in international journals and conference proceedings. He is a co-author of three books. Prof. Ytterdal is a member of The Norwegian Academy of Technological Sciences and a Senior Member of IEEE.



Snorre Aunet received the Cand. Scient. degree in informatics, from the University of Oslo (UiO) in 1993, and the Dr.Eng. degree in physical electronics from the Norwegian University of Science and Technology (NTNU), in 2002. He worked with ASIC design at Nordic VLSI from 1994 to 1997. Later, he worked at NTNU and UiO, and is currently a professor at NTNU. He has been a visiting researcher at the University of Paderborn, Germany, for shorter periods, between 2004 and 2012. Between July 2015 and July 2016 he had a sabbatical at UCSD. Research interests include ultra low-power mixed-signal circuits.



Ultra-low Voltage and Energy Efficient Adders in 28 nm FDSOI Exploring Poly-biasing for Device Sizing

A. A. Vatanjou, E. Lâte, T. Ytterdal, and S. Aunet

Published in *Microprocessors and Microsystems*, vol. 56, pp. 92-100, 2018¹

DOI: <https://doi.org/10.1016/j.micpro.2017.11.002>

C.1 Abstract

Balancing the PMOS/NMOS strength ratio is a key issue to maximize the noise margin, and hence, the functional yield of CMOS logic gates and minimize the leakage energy per cycle in the subthreshold region. In this work, the PMOS/NMOS strength ratio was balanced using a poly-biasing technique in conjunction with back-gate biasing provided in a 28 nm fully depleted silicon on insulator (FDSOI) CMOS technology. A 32-bit adder based on minority-3 (min-3) gates and a 16-bit adder based on Boolean gates have been implemented. Chip measurement results of nine samples show highly energy-efficient adders. The 32-bit and 16-bit adders achieved mean minimum energy points (MEP) of 20.8 fJ at 300 mV and 12.34 fJ at 250 mV, respectively. In comparison to adders reported in other works in the same technology, the energy per 1-bit addition of the 32-bit adder is improved by 37%. This improvement in energy consumption is 25% for the 16-bit adder. According to the measurement results of ten chips, the designed adders exhibited functionality down to supply voltages of 110 mV-125 mV, without body biasing. Additionally, the minimum V_{dd} of all the 32-bit adders based on minority-3 gates decreased to 80 mV by applying a reverse back-bias voltage to the PMOS devices. One sample was functional at 79 mV with a 430 mV reverse back-bias voltage applied to its PMOS devices.

C.2 Relevance to the Thesis

Memory systems commonly contain logic blocks to interact with the core memory cells. The paper is relevant to this research on low voltage memory systems by the fact that it contains a method to create logic libraries that are robust at low supply

¹References in article: [84] [85] [18] [86] [87] [88] [89] [90] [91] [92] [93] [94] [95] [96] [97] [98] [99] [100] [101] [102] [103] [104]

voltages. Silicon measurements verify the method.

C.3 Declaration of Authorship

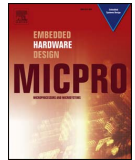
This article is an invited and extended version of the supporting paper that was presented at the Nordic Circuits and Systems Conference (NORCAS) in 2016. The paper contains measurements of the silicon die that Ali Asghar Vatanjou and I designed in 2015. Ali performed the literature study, implemented the adders, and wrote the paper together with me under the supervision of Trond Ytterdal and Snorre Aunet.

C.4 Manuscript



Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

Ultra-low voltage and energy efficient adders in 28 nm FDSOI exploring poly-biasing for device sizing

Ali Asghar Vatanjou*, Even Låte, Trond Ytterdal, Snorre Aunet

Department of Electronic Systems, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), O.S. Bragstads plass 2a, Trondheim, 7491, Norway

ARTICLE INFO

Keywords:
Subthreshold
28 nm FD-SOI
CMOS
Full-adder
Ultra-low power

ABSTRACT

Balancing the PMOS/NMOS strength ratio is a key issue to maximize the noise margin, and hence, the functional yield of CMOS logic gates and minimize the leakage energy per cycle in the subthreshold region. In this work, the PMOS/NMOS strength ratio was balanced using a poly-biasing technique in conjunction with back-gate biasing provided in a 28 nm fully depleted silicon on insulator (FDSOI) CMOS technology. A 32-bit adder based on minority-3 (min-3) gates and a 16-bit adder based on Boolean gates have been implemented. Chip measurement results of nine samples show highly energy efficient adders. The 32-bit and 16-bit adders achieved mean minimum energy points (MEP) of 20.8 fJ at 300 mV and 12.34 fJ at 250 mV, respectively. In comparison to adders reported in other works in the same technology, the energy per 1-bit addition of the 32-bit adder is improved by 37%. This improvement in energy consumption is 25% for the 16-bit adder. According to the measurement results of ten chips, the designed adders exhibited functionality down to supply voltages of 110 mV–125 mV, without body biasing. Additionally, the minimum V_{dd} of all the 32-bit adders based on minority-3 gates decreased to 80 mV by applying a reverse back bias voltage to the PMOS devices. One sample was functional at 79 mV with a 430 mV reverse back bias voltage applied to its PMOS devices.

1. Introduction

Reducing power consumption to ultimately break heat walls or to reduce energy consumption to prolong battery lifetime are key issues in present day microelectronics. The growing number of battery driven wireless devices, as well as the increasing energy demand of the applications to which these are intended, create a need for innovations in ultra-low power (ULP) circuitry. Subthreshold operation, potentially providing higher energy efficiency than other known low power circuits, has been known since the 1960s [1]. However the increasing demand mentioned above, and new features of modern process technologies, brings relevance of subthreshold design back.

A commercially available 28 nm FDSOI CMOS process tackles some of the key challenges faced by the scaling of conventional bulk silicon devices better, and so enables prolongation of Moore's law [2]. In this technology, the junction capacitance is significantly reduced and the drain induced barrier lowering (DIBL) effect is also lowered due to the ultra thin body and buried oxide [2], this while increasing the efficiency of back-gate biasing. In addition, the random dopant fluctuation is suppressed because threshold voltage adjustments do not depend on doping levels.

This manuscript present several extensions to the contents in [3] with respect to both statistical simulations and new measurements on additional chip samples. The extensions to simulations include variability in on-currents and delay as functions of sizing, and how sizing can affect energy per cycle for the designed logic cells. Post-layout simulations were also added, and compared to measurements results. Additionally, new measurements results have been added, which include effects of applying back-gate biasing voltages. Further extensions to [3] include measurements of static power consumption for nine 16-bit and 32-bit adder samples.

We here report highly energy efficient logic gates by exploiting both poly-biasing [4] as well as back-gate biasing techniques. Both techniques allow us to reduce the size of the pull up devices, thus reducing the leakage current and parasitic capacitance of the cells. This in turn leads to reduced power consumption. The poly-biasing technique increases the effective channel gate length without increasing the active area of the devices. Applying poly-biasing while connecting the PWELL/NWELL ties to ground improves the strength of the PMOS transistors. Moreover, the regular threshold voltage (RVT) devices have been chosen to reduce the leakage current further, compared to low threshold voltage (LVT) devices. Although higher threshold voltage

* Corresponding author.

E-mail addresses: ali.vatanjou@ntnu.no (A.A. Vatanjou), even.laate@iet.ntnu.no (E. Låte), trond.ytterdal@ntnu.no (T. Ytterdal), snorre.aunet@ntnu.no (S. Aunet).

leads to slower circuits, it can reduce the energy consumption of circuits with more relaxed throughput requirements [5]. The computational demands of wireless sensor networks (WSNs) have been discussed in [6], and the sensor-network applications were categorized into three groups: 1- low-bandwidth rates (sample rate less than 100 Hz), 2- mid-bandwidth rates (sample rate 100 Hz-1 KHz) and 3- high-bandwidth rates (sample rate higher than 1 KHz). It was shown in [6] that the performance of a processor with 168 KHz clock was more than four times of the desired performance for mid-bandwidth applications. For such systems with relaxed throughput requirements, a lower static power and robust functionality at ultra-low voltages translates into a lower energy consumption. Moreover, the use of RVT devices enables applying extreme reverse back-bias schemes to save leakage energy in the sleep mode, and avoid costly power-gating.

To validate the proposed sizing techniques, we have implemented a 32-bit adder based on min-3 gates [7] and a 16-bit adder based on traditional Boolean gates. The ripple carry adder (RCA) topology has been chosen because the energy consumption of the serial adders may be lower than the parallel adders while maintaining the same speed, when operated in subthreshold [8]. Comparing to adders reported in the same 28 nm FDSOI process technology [9], the minimum energy per 1-bit addition of the 32-bit and 16-bit adders have been improved by 37% and 25%, respectively.

To the knowledge of the authors, the 32-bit adder provides the lowest energy per 1-bit addition and the lowest functional V_{dd} among static CMOS circuits with more complex functionalities than simple inverters or ring-oscillators.

The remainder of the paper is organized as follows: The device sizing is discussed in Section 2. Two implementations of subthreshold adders are presented in Section 3. Measurement results are shown in Section 4, and is followed by a discussion in Section 5, before the final conclusions in Section 6.

2. Device sizing for balanced PMOS/NMOS

2.1. Transistor dimensions

Balancing the driving strength of the pull-up/pull-down networks (PUN/PDN) is a key issue to properly operate logic gates in the ultra-low voltage (ULV) domain. In addition, unbalanced PMOS/NMOS increases the leakage energy of subthreshold circuits [10]. Nevertheless as shown in Fig. 1, the on-current ratio of the PMOS with respect to NMOS is too small in the ULV domain. In Fig. 1, the NMOS device width is 200 nm, $L = 30$ nm and the drain-source voltage is set equal to the supply voltage. For example, the PMOS gate width should be upsized by more than 4X to have a balanced PUN/PDN at a supply voltage of 150 mV. This causes asymmetry in the layout of the cells. Additionally, both leakage current and parasitic capacitances, and consequently the power consumption of the circuit, increase with wider PMOS

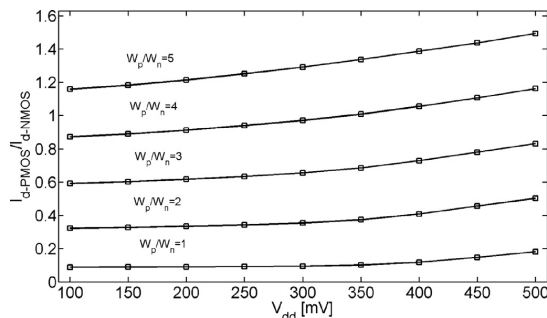


Fig. 1. PMOS/NMOS on-current ratio versus V_{dd} for $W_n = 200$ nm and $L_n = L_p = 30$ nm.

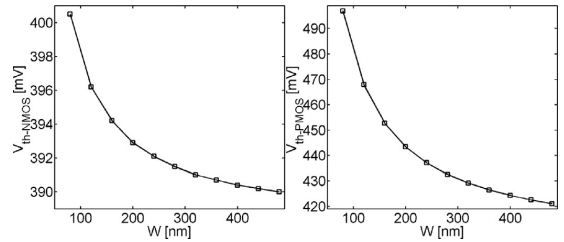


Fig. 2. Dependency of the threshold voltage on channel width (a) NMOS (b) PMOS.

transistors. Therefore using techniques to improve the strength ratio between PMOS and NMOS, can be an effective approach to reduce both power consumption and minimum operating supply voltage of sub-threshold circuits.

In this work we increased the effective gate length of the NMOS transistors by 16 nm using poly-biasing to balance NMOS and PMOS driving strengths. We also connected both NWEEL and PWEEL ties to ground, thus the bulk-source junction of PMOS transistors were forward biased. In order to reduce the leakage current further, the effective PMOS gate length was also increased by 4 nm using poly-biasing, while maintaining the switching voltage at $V_{dd}/2$. As can be observed from Fig. 2, the threshold voltage of this 28 nm FDSOI technology is inversely proportional to the gate width of the devices. In Fig. 2, $L_{nmos} = L_{min} + 16$ nm, $L_{pmos} = L_{min} + 4$ nm and $V_{dd} = 200$ mV. The NMOS and PMOS transistor channel widths were chosen taking into account the channel width dependency of the threshold voltage. The minimum width of the NMOS devices was chosen to be 200 nm and the corresponding PMOS width that resulted in a switching voltage of $V_{dd}/2$ for an inverter was 300 nm. Although a W_p/W_n ratio of 1.5 is required for balancing the PMOS and NMOS transistors, this is a reasonable ratio comparing with for example [11], where for $W_p/W_n = 5$, PMOS/NMOS strength ratio is balanced and for $W_p/W_n = 2$, the PMOS and NMOS transistors are still unbalanced. Fig. 3 shows the schematics and device sizes for the basic logic cells that were used to implement the 1-bit full adders (FA).

2.2. Exploring variability and energy efficiency

In order to explore the robustness of the proposed sizing approach, the drain current variability of an inverter (INV3) based on the proposed sizing approach was simulated and compared to two other inverters. The drive strength of the pull-up/down devices of the first inverter (INV1) were balanced and the active area of the PMOS was more than 5X the NMOS active area with $W_p = 420$ nm, $W_n = 80$ nm and $L_p = L_n = 30$ nm. The second inverter (INV2) had the same total area as INV1 but the PMOS/NMOS drive strengths are not balanced ($W_p = 300$ nm, $W_n = 200$ nm and $L_p = L_n = 30$ nm). The active area of the PMOS in INV3 is only 1.1X the NMOS active area where $W_p = 300$ nm, $W_n = 200$ nm, $L_p = 30 + 4$ nm and $L_n = 30 + 16$ nm.

Fig. 4 shows the NMOS/PMOS on-currents variability for the above discussed inverters as a function of supply voltage. Monte Carlo

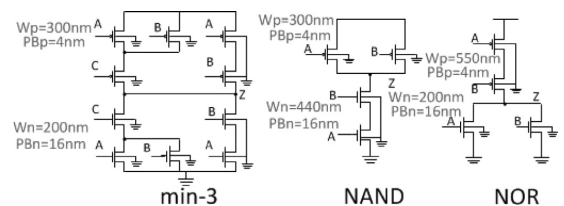


Fig. 3. Schematic view and corresponding device sizes of the basic logic gates used to implement FAs.

A.A. Vatanjou et al.

Microprocessors and Microsystems 56 (2018) 92–100

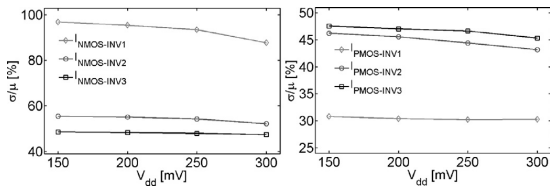


Fig. 4. Percentage variability of the NMOS/PMOS on-currents for different inverter implementations.

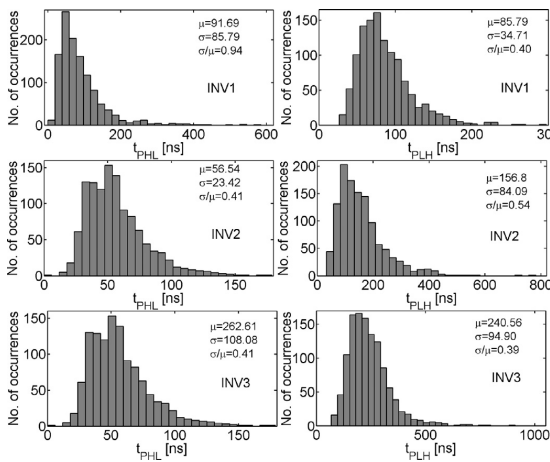


Fig. 5. Rise and fall FO4 delay variability of different sizing approaches at a supply voltage of 200 mV.

simulations have been carried out at the typical process corner (TT) considering mismatch and process variations. The NMOS on-current variability in INV1 is much higher than that of the PMOS. In INV2, the variability of the PMOS on-current is slightly lower than the NMOS on-current variability, but the NMOS is stronger than the PMOS for this inverter. For INV3, the on-current variability of the NMOS and PMOS are almost the same and the drive strength of the transistors are balanced. Therefore, logic cells with relatively equal active areas of PMOS and NMOS devices might benefit from similarity in rise and fall delays variability. Fig. 5 shows Monte Carlo simulations of the rise/fall FO4 delays of INV1, INV2 and INV3 with $3\text{-}\sigma$ accuracy (99.73% target yield) at the TT process corner and 200 mV taking mismatch and process variations into account. It can be observed from Fig. 5 that the variability of the fall delay of INV1 is 2.3X the variability of its rise delay. Despite the small difference in rise and fall delays variability in INV2, its rise delay mean value is 2.8X fall delay mean value. However, the difference in fall and rise delay variability is only 2% for INV3, and the fall delay mean value is comparable to the rise delay mean value. Increasing the channel length of transistors results in longer delays. Nevertheless, as discussed previously, it reduces the delay variability and improves the functional yield of the logic cells. In order to investigate if the drop in the speed of logic cells can be compensated without energy overhead by increasing the supply voltage, the energy of inverter chains with 20 FO4 delay logic depth was simulated versus V_{dd} . The logic depth of 20 was chosen because as it is discussed in [12] and [13], a logic depth of 20 to 17 results in a balance between the path delay variability and the energy efficiency. Fig. 6 shows the total energy per cycle of 20 FO4 delay inverter chains consist of INV1, INV2 and INV3 versus supply voltage, with a switching activity of 0.25. For example at a target operating frequency of 1 MHz, the energy per operation of INV3 chain is 6.08 fJ (@283 mV) which is 9% lower than the

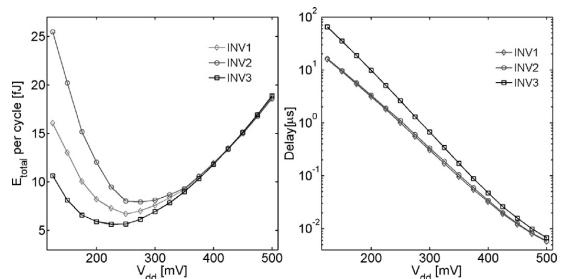


Fig. 6. Energy per cycle and delay of the 20 FO4 delay inverter chains.

energy per operation of INV1 chain (6.68 fJ @ 250 mV) and 24% lower than the energy per operation of the INV2 chain (7.97 fJ @ 254 mV). Therefore, the resultant drop in the speed due to the larger channel lengths can be compensated by increasing the supply voltage without increasing the energy consumption.

2.3. Back-gate biasing effects

The FDSOI technology offers a relatively wide range of back-gate biasing voltages, due to the buried oxide, that regulates the threshold voltages of the transistors. The effect of back-gate biasing voltage on the delay and leakage of subthreshold circuits can be estimated taking the exponential dependency of the drain-source current on V_{th} ($I_{ds} \propto \exp((V_{GS} - V_{th})/nV_t)$ [14,15]) and linear dependency of V_{th} on back-bias voltage [16] into account.

The linear relationship between the threshold voltage, V_{th} , and the back-gate voltage, V_{BS} , of FDSOI devices, can be expressed as [17]:

$$V_{th} = V_{th0} - \gamma V_{BS} \quad (1)$$

In [17], γ was approximated to be 85 mV/V.

Considering the exponential dependency of the drain-to-source current on the threshold voltage in the subthreshold region, we can write the following expressions for the delay ratio and the leakage current ratio for different back-gate biasing voltages with a fixed supply voltage:

$$\frac{t_{D1}}{t_{D0}} \propto \exp((V_{th1} - V_{th0})/nV_t) = \exp(-\gamma V_{BS}/nV_t) \quad (2)$$

$$\frac{I_{leak1}}{I_{leak0}} \propto \exp((V_{th0} - V_{th1})/nV_t) = \exp(\gamma V_{BS}/nV_t) \quad (3)$$

In Eqs. (2) and (3), n is the inverse slope of the subthreshold current, V_{th0} is the threshold voltage of the device with no body biasing voltage, and V_t is the thermal voltage ($4kT/q \approx 26$ mV at room temperature). We will use these equations later, in Section 4, to compare our calculations to the measurement results.

3. Adder circuit design

Fig. 7 illustrates 1-bit FA cells of the two different logic styles. The FA based on min-3 gates [7] was used to implement a 32-bit RCA [8]. A 16-bit RCA was also implemented using the FA based on Boolean logic gates. The layout views of the FAs are shown in Fig. 8. We focused on layout regularity to increase the matching properties of the devices and alleviate local systematic matching errors [18]. There is no rounding and routing in the poly layer. All the poly polygons have single direction, and 106 nm poly pitch was used to be able to apply poly-biasing on transistors gates. As a result, the mismatch between transistors gate lengths should decrease [18], as well as leakage currents and $I_{on-NMOS}/I_{on-PMOS}$. Back-bias voltage rails are routed in parallel with the supply and ground rails and well ties are placed on top and bottom of each cell to reduce the device's active area distance to the well ties.

A.A. Vatanjou et al.

Microprocessors and Microsystems 56 (2018) 92–100

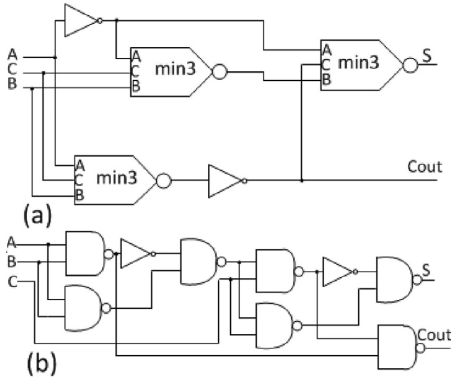


Fig. 7. Block-diagram of 1-bit FA (a) implemented with min-3 gates (b) implemented with Boolean gates.

4. Simulation and measurement results

Fig. 9 shows the PCB and QFN44 socket that was used for measuring the ten chips. Fig. 10 illustrates the test chip block diagram. A toggling input carry was applied to the adders while all the A inputs were shorted to V_{dd} , and the B inputs connected to ground. This exercised the critical path and caused the longest delay [8]. The RVT devices that were used in the core circuitry had threshold voltages of around 430 mV. The buffers were implemented with LVT transistors to make the delay through the buffers and I/O pads negligible compared to that of the adders. The absolute value of the threshold voltages of the LVT transistors used in the buffers is around 390 mV. An HP 6632A DC power supply was used for the supply voltages of the adders. The input carry signal was generated by an Agilent 33522A function generator. To measure the currents, a Keithley 6485 Picoammeter was used. The output/input waveforms were captured by a ROHDE & SCHWARZ RTE 1022 oscilloscope.

Fig. 11 (a) shows measured energy per operation of the 32-bit adder versus supply voltage at maximum operating speed. The delay of the 32-bit adder versus V_{dd} is also shown in Fig. 11 (b). The MEP of nine samples was measured for the 32-bit and 16-bit RCAs. The 32-bit RCA achieved a minimum energy point of 20.8 fJ on average at a 300 mV V_{dd} and the mean delay at this V_{dd} was 3.2 μ s. This means that the mean MEP per 1-bit addition is 0.65 fJ for this adder. Measured energy and delay of the 16-bit adder versus supply voltage are depicted in Fig. 12(a) and (b). The average of the minimum energy point of the 16-

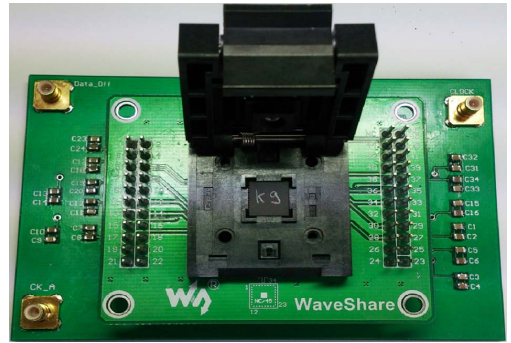


Fig. 9. The PCB and QFN44 socket for measuring chip samples.

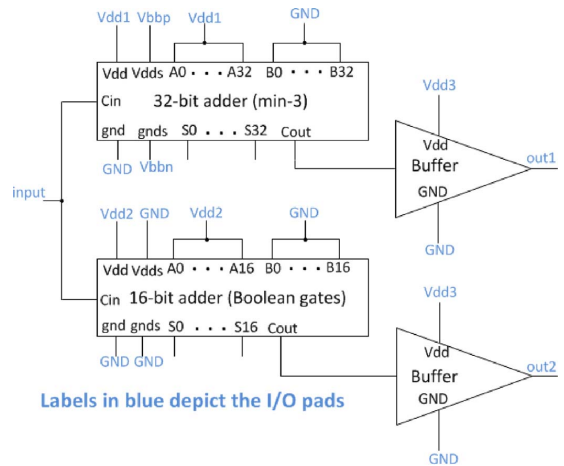


Fig. 10. Block-diagram of the test chip.

bit adder for nine samples was measured to be 12.34 fJ at 250 mV. Thus, the minimum required energy per 1-bit addition of this adder is 0.77 fJ on average. The mean delay of the 16-bit adder was 6.2 μ s at the MEP (@250 mV).

Fig. 13 shows the measured static power consumption of nine 32-bit RCA and 16-bit RCA samples. During leakage current measurements,

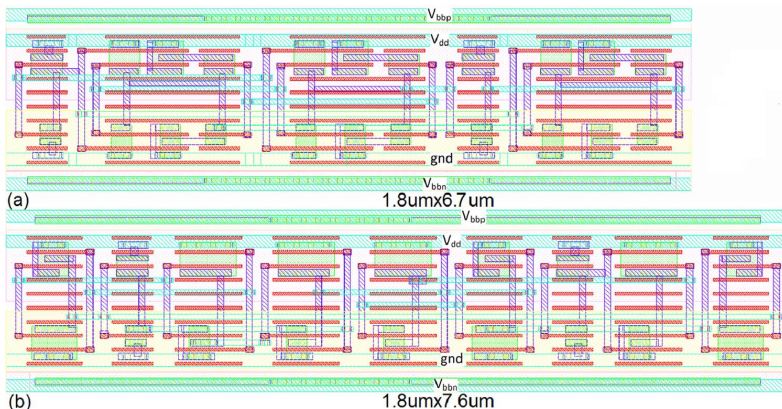


Fig. 8. Layout view of the 1-bit full adder (a) implemented with min-3 gates (b) implemented with Boolean gates.

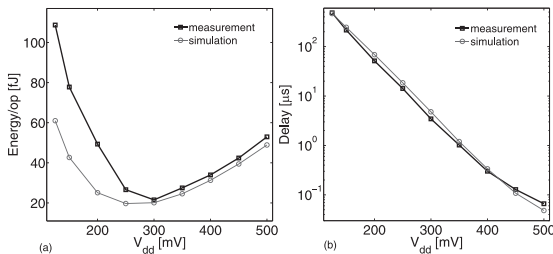


Fig. 11. Measured energy and delay of the 32-bit adder versus V_{dd} .

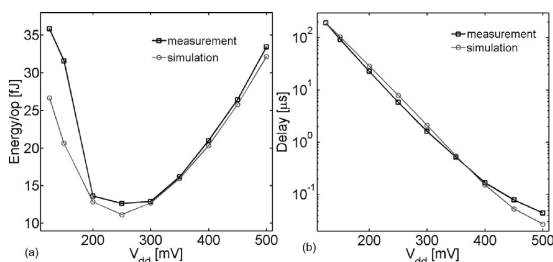


Fig. 12. Measured energy and delay of the 16-bit adder versus V_{dd} .

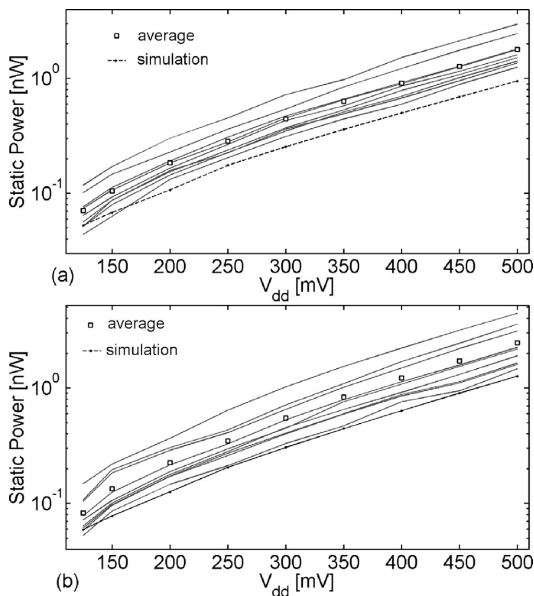


Fig. 13. Measured static power consumption of nine (a) 32-bit adder samples (b) 16-bit adder samples.

using a Keithley 6485 Picoammeter, the carry in of both adders was grounded. Chip number 8 was damaged during these measurements, so its static power consumption is not included in Fig. 13. The simulated and average measured static power consumptions of nine samples is also depicted in Fig. 13 (with square markers).

To draw a fair comparison between the static power of the FAs based on min-3 and Boolean logic gates, the static power consumption of 1-bit FAs were extracted from the mean measured static power in Fig. 13, and is illustrated in Fig. 14. As can be observed from Fig. 14,

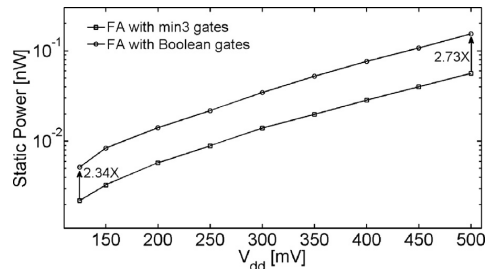


Fig. 14. Static power consumption of the 1-bit FAs.

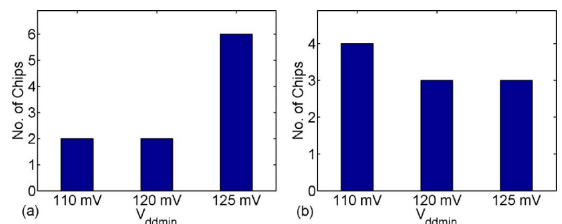


Fig. 15. Minimum V_{dd} distribution of (a) 32-bit adders implemented with min-3 gates (b) 16-bit adders implemented with Boolean gates.

the static power consumption of the 1-bit FA based on Boolean gates is more than 2X of the static power of the 1-bit FA implemented with min-3 gates. At the supply voltages of 125 mV and 500 mV, the static power consumption of the FA based on Boolean gates is 2.34X and 2.73X the static power of the FA implemented with min-3 gates, respectively.

We measured ten chips to explore the functionality of the adders at extremely low supply voltages. Fig. 15 shows the minimum V_{dd} of the adders as a histogram for ten measured chips with both PMOS and NMOS back-gates (V_{bbp} and V_{bbn}) shorted to ground. The minimum V_{dd} s were all between 110 and 125 mV for both types of adders, when having their back-gates connected to ground. In the 32-bit adder case, two chips out of ten measured samples were functional down to 110 mV. For the 16-bit adder, four chip samples out of ten were functional down to 110 mV.

As shown in Fig. 10, the PWELL/NWELL contacts of the 32-bit adder are routed to the I/O pads, which enables applying off-chip back bias voltages to this adder. The V_{bbp} and V_{bbn} voltages were generated by off-chip MASCOT 719 DC power supplies. We were able to decrease the minimum V_{dd} of the first 32-bit adder sample down to 79 mV by applying a 430 mV positive V_{bbp} to the PMOS transistors, i.e. applying a reverse back bias voltage to the PMOS transistors. The oscilloscope plots of the input/output carries of the 32-bit adder at a supply voltage of 79 mV with $V_{bbp} = 430$ mV and $V_{bbn} = 0$ V are shown in Fig. 16. Fig. 17 depicts the input/output carries of the 16-bit adder for a V_{dd} of 110 mV. The minimum V_{dd} of all the 32-bit adder samples decreased by applying a reverse back bias (RBB) to the PMOS transistors. The required back-gate voltages for PMOS devices of the 32-bit adders to achieve a minimum functional V_{dd} of 80 mV, and corresponding leakage currents of the 32-bit adder samples at 80 mV, are listed in Table 1.

Unlike the conventional back biasing of the RVT transistors where $V_{bbp} = V_{dd} - V_{bb}$ and $V_{bbn} = V_{bb}$ [20], the absolute values of the PMOS and NMOS back bias voltages are equal in the logic cells we developed, i.e. $V_{bbp} = -V_{bb}$ and $V_{bbn} = V_{bb}$. In order to decrease the delay of the 32-bit RCA, we applied a forward back bias (FBB) voltage equal to 0.25 V ($V_{bbp} = -0.25$ V and $V_{bbn} = 0.25$ V). Fig. 18(a) shows the measured delay of the 32-bit adder with three different back biasing schemes as a function of supply voltage. The delay of the 32-bit adder decreased by

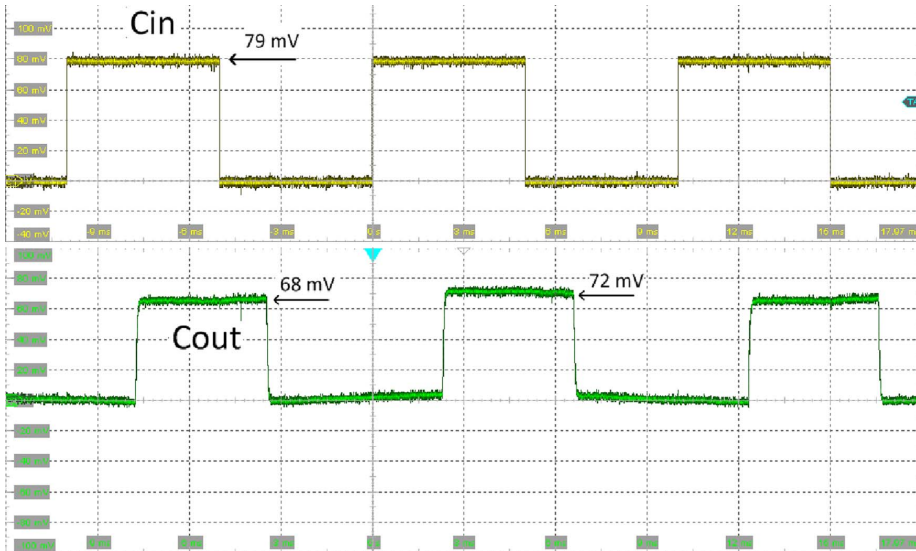


Fig. 16. Oscilloscope plots proving that the 32-bit adder is functional at 79 mV with $V_{bbp} = 430$ mV and $V_{bbn} = 0$ V.

39% at 125 mV and 23% at 500 mV when a 0.25 V forward back bias was applied to this adder.

A 0.5 V reverse back bias voltage ($V_{bbp} = 0.5$ V and $V_{bbn} = -0.5$ V) was also applied to the adder to reduce the leakage current of the circuit. As can be observed from Fig. 18(b), a static power reduction of 70% at 125 mV and 68% at 500 mV was achieved by applying a 0.5 V back-gate bias voltage to the 32-bit adder. For a supply voltage of 200 mV and 0.25 V FBB, the resulting reduction in delay was measured to be 1.8X, while the RBB of 0.5 V reduced the leakage current by a factor of 3.3X. Eqs. (2) and (3) combined with extracted values based on simulations, for a V_{ad} of 200 mV, were used for comparisons with

measured data. The inverse slope of the subthreshold current, n , was found to be 1.22 for the NMOS and 1.18 for the PMOS. The body factor [17] was found to be 64 mV/V for the NMOS, and 71 mV/V for the PMOS, respectively. Using the average value of the parameters for the PMOS and NMOS transistors, the measured reduction in delay of 1.8X was estimated as 1.7X based on Eq. (2), while the measured reduction in the leakage current (RBB of 0.5 V) of 3.3X was estimated to be 2.9X, based on Eq. (3).

Fig. 19 shows the layout view of the 32-bit RCA and the parasitic PWELL/NWELL diodes. A deep-NWELL (D-NWELL) and an NWELL guard-ring was used to isolate the PWELL of the adder from the p-

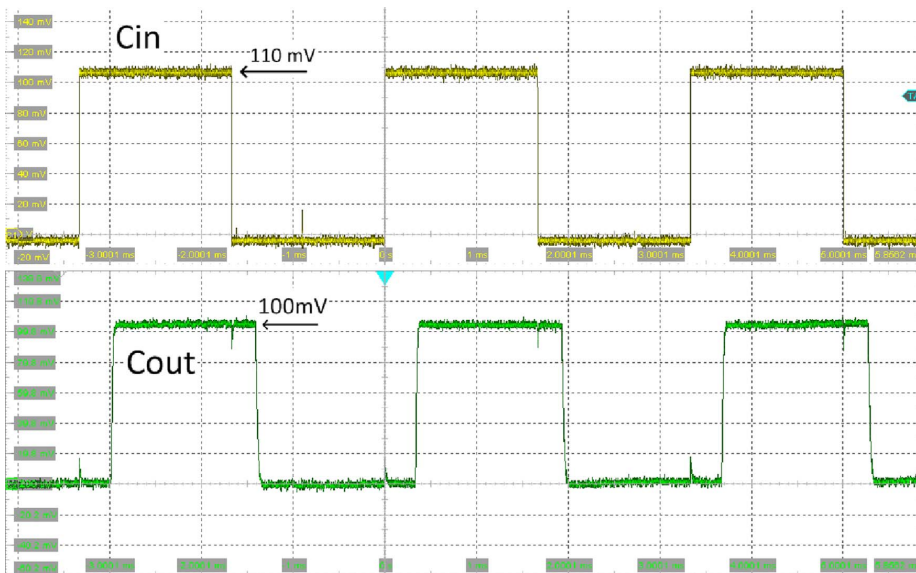


Fig. 17. Oscilloscope plots proving that the 16-bit adder is functional at 110 mV with $V_{bbp} = V_{bbn} = 0$ V.

Table 1

Required back-bias voltage for the PMOS devices of ten 32-bit adder samples to reduce the minimum functional V_{dd} of the adders down to 80 mV and the corresponding measured leakage current of each sample at 80 mV.

	Chip1	Chip2	Chip3	Chip4	Chip5	Chip6	Chip7	Chip8	Chip9	Chip10
V_{bbp} [mV]	420	355	530	315	625	535	355	685	610	475
I_{leak} [nA]	0.27	0.25	0.15	0.21	0.29	0.31	0.23	N/A	0.17	0.24

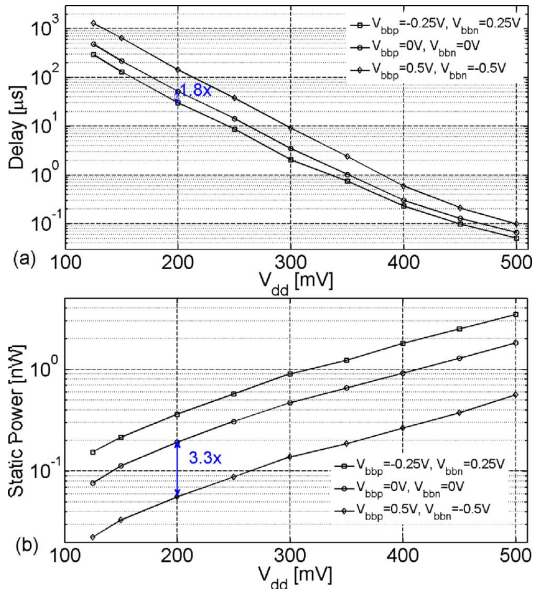


Fig. 18. Measured (a) Delay and (b) static power of the 32-bit adder (first sample) for three back-biasing schemes.

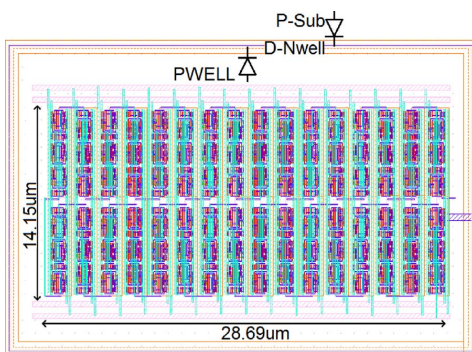


Fig. 19. Layout view of the 32-bit adder and corresponding parasitic PWELL/NWELL diodes.

substrate. The breakdown voltage of the PWELL/NWELL diode restricts the reverse back bias voltage range. For the forward back bias scenario, the limiting factor comes from the threshold voltage of these diodes. Therefore, the applicable reverse back bias voltage to the RVT devices is wider than the forward back bias voltage range. The leakage current of the PWELL/NWELL diode was measured to be 12.6 nA when a 0.25 V forward back bias was applied to the adder. This leakage current increased to 1.44 μ A when the applied forward back bias voltage was increased to the 0.3 V. Thus, we limited the forward back bias voltage to be 0.25 V to reduce the leakage current through the parasitic diodes.

5. Discussion

Table 2 compares the measurement results of the designed adders with the post-layout simulation results of existing works in terms of energy, delay and area. The 32-bit RCA presented here achieved 0.65 fJ per 1-bit addition. This is the lowest energy per 1-bit addition among the works listed in Table 2. In comparison to [9] which has the lowest reported energy per 1-bit addition (1.03 fJ) among the other works, the energy per 1-bit addition of the 32-bit and 16-bit adders are improved by 37% and 25%, respectively.

Comparing the 1-bit addition delay times at 300 mV, the adder in [9] has the shortest delay of 1.7 ns, thanks to the LVT devices and forward body biasing. The 1-bit addition delay times the 32-bit and 16-bit adders in [19] are 28 ns. This means that the adders in [19] are 3.8X and 3.6X faster than the presented 32-bit and 16-bit adders, respectively, when $V_{bbp} = V_{bbn} = 0$. As listed in Table 2, the static power consumption of the reported 32-bit RCA in [19] is 15.9X higher than the static power of the 32-bit adder and the 16-bit adder in [19] has 6.8X higher static power consumption than the 16-bit adder in this work. The low speed of the designed adders was not unexpected considering RVT transistors and longer channel lengths of the devices. Nevertheless, the delay of the 32-bit adder at 300 mV decreased to 2.04 μ s (i.e. 63.75 ns per 1-bit addition) when a 0.25 V forward back bias was applied. Moreover, there are many applications with very low operating frequency requirements where wireless sensor networks are utilized to monitor phenomena with low sample rates [6,21].

The minimum supply voltage of the 32-bit adder decreased by applying a reverse back-bias voltage to the PMOS devices, and the same trend was observed for all the samples (Table 1). Therefore, the dies position might be towards the slow-NMOS fast-PMOS (SF) process corner.

Despite the fact that the FA based on min-3 gates has 34 transistors and the FA based on Boolean gates has 32 transistors, the footprint area of the FA based on min-3 gates (12.06 μm^2) is smaller than the area of FA based on Boolean gates (13.68 μm^2). The area of the 32-bit adder is 3% smaller than the 32-bit adder in [19], while the 16-bit adder has 9% more area in comparison with the 16-bit adder in [19].

The leakage power of the 32-bit RCA was lower than the leakage power of the 16-bit RCA. From Fig. 14, the static power of the 1-bit FA based on Boolean gates is more than 2X of that of the 1-bit FA implemented with min-3 gates. The static power consumption accounts for a considerable part of the total power consumption at extremely low supply voltages where the operating frequency of the system is relatively low. Thus, the leakage current plays a significant role on energy consumption at ultra-low V_{dd} s and a high leakage current deteriorates the energy consumption of the circuit. Thereby, a lower MEP at a lower supply voltage could be achieved by using FAs based on min-3 gates in the datapath blocks of the system.

Schmitt-Trigger logic cells functioning down to a supply voltage of 62 mV have been reported in [22]. However, these logic cells were not energy and area efficient. An ultra-low voltage subthreshold adder based on static CMOS logic cells, implemented in 65 nm CMOS, have been reported in [23], which was capable of working correctly down to a supply voltage of 84 mV. LVT devices have been used in [23], hence the leakage current is relatively high. Although the minimum reported V_{dd} is higher compared to [22], the logic cells are more area and power efficient. Thereby, the proposed approach can also be useful for applications powered by an energy harvesting source where the limiting factor comes from the power consumption of the system [24].

Table 2
Comparison with existing works in 28 nm FDSOI technology.

	This work 32-bit adder	This work 16-bit adder	[19] 32-bit adder	[19] 16-bit adder	[9] 9-bit adder
E_{min}	20.8* fJ	12.34* fJ	207.4 fJ	74.9 fJ	9.27 fJ
P_{static} @ 0.3 V	0.446* nW	0.549* nW	7.10 nW	3.74 nW	N.A.
Delay @ MEP	3.24* μ s	6.2* μ s	0.898 μ s	0.448 μ s	0.555 μ s
V_{dd} @ MEP	300 mV	250 mV	300 mV	300 mV	240 mV
Devices	RVT	RVT	RVT/LVT	RVT/LVT	LVT
Area	406 μ m ²	230.3 μ m ²	418.56 μ m ²	209.28 μ m ²	N.A.
Results	Measurement	Measurement	Post-layout	Post-layout	Post-layout

* Average of nine measured samples.

6. Conclusion

The development of ultra-low energy subthreshold adders in a 28 nm FDSOI CMOS technology has been presented. Poly and back-gate biasing techniques has been used to balance PMOS/NMOS strength ratio. This combined with selected building block topology and regular layout which in turn resulted in a robust and energy efficient adder. To the knowledge of the authors, the measured 0.65 fJ energy per 1-bit addition of the 32-bit adder is the lowest measured energy reported so far. Both the 32-bit and the 16-bit adders were functional in the deep subthreshold region during measurements. The minimum functional supply voltage of the 32-bit adder was found to be 79 mV with a 430 mV reverse back-gate bias applied to PMOS transistors. The back biasing knob has been utilized to enhance the performance of the 32-bit adder or reduce its leakage power consumption.

References

- [1] Y. Tsvitdis, Eric vittoz and the strong impact of weak inversion circuits, *IEEE Solid-State Circuits Soc. Newslett.* 13 (3) (2008) 56–58.
- [2] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Boeuf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J.P. Schoellkopf, E. Perea, R. Ferrant, H. Mingam, Innovative materials, devices, and cmos technologies for low-power mobile multimedia, *IEEE Trans. Electron Devices* 55 (1) (2008) 96–130.
- [3] A.A. Vatanjou, E. Late, T. Ytterdal, S. Aunet, Ultra-low voltage adders in 28 nm fdsOI exploring poly-biasing for device sizing, 2016 IEEE Nordic Circuits and Systems Conference (NORCAS), (2016), pp. 1–4.
- [4] D. Jacquet, F. Hasbani, P. Flatresse, R. Wilson, F. Arnaud, G. Cesana, T.D. Gilio, C. Lecocq, T. Roy, A. Chhabra, C. Grover, O. Minez, J. Uginet, G. Durieu, C. Adobati, D. Casalotto, F. Nyer, P. Menut, A. Cathelin, I. Vongsavady, P. Magarshack, A 3 ghz dual core processor arm cortex tm -a9 in 28 nm utbb fd-soi cmos with ultra-wide voltage range and energy efficiency optimization, *IEEE J. Solid-State Circuits* 49 (4) (2014) 812–826.
- [5] D. Bol, R. Ambroise, D. Flandre, J.D. Legat, Analysis and minimization of practical energy in 45nm subthreshold logic circuits, 2008 IEEE International Conference on Computer Design, (2008), pp. 294–300.
- [6] B. Zhai, S. Pant, L. Nazhandali, S. Hanson, J. Olson, A. Reeves, M. Minuth, R. Helfand, T. Austin, D. Sylvester, D. Blaauw, Energy-efficient subthreshold processor design, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 17 (8) (2009) 1127–1137.
- [7] S. Aunet, Y. Berg, Three sub-fJ power-delay-product subthreshold cmos gates, 2005 IFIP International Conference on Very Large Scale Integration, (2005), pp. 465–470.
- [8] V. Beiu, A. Djupdal, S. Aunet, Ultra Low-Power Neural Inspired Addition: When Serial Might Outperform Parallel Architectures, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- [9] H. Attarzadeh, S. Aunet, T. Ytterdal, An ultra-low-power/high-speed 9-bit adder design: analysis and comparison vs. technology from 130nm-lp to utbb fd-soi-28 nm, 2015 Nordic Circuits and Systems Conference (NORCAS): NORCHIP International Symposium on System-on-Chip (SoC), (2015), pp. 1–4.
- [10] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, D. Blaauw, Exploring variability and performance in a sub-200-mv processor, *IEEE J. Solid-State Circuits* 43 (4) (2008) 881–891.
- [11] M.Z. Li, C.I. Jeong, M.K. Law, P.I. Mak, M.I. Vai, S.H. Pun, R.P. Martins, Energy optimized subthreshold vlsi logic family with unbalanced pull-up/down network and inverse narrow-width techniques, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 23 (12) (2015) 3119–3123.
- [12] B. Zhai, S. Hanson, D. Blaauw, D. Sylvester, Analysis and mitigation of variability in subthreshold design, ISLPED '05. Proceedings of the 2005 International Symposium on Low Power Electronics and Design, 2005. (2005), pp. 20–25.
- [13] M. Seok, D. Jeon, C. Chakrabarti, D. Blaauw, D. Sylvester, Pipeline strategy for

- improving optimal energy efficiency in ultra-low voltage design, 2011 48th ACM/EDAC/IEEE Design Automation Conference (DAC), (2011), pp. 990–995.
- [14] M. Belleville, O. Faynot, Evolution of deep submicron bulk and soi technologies, in: C. Pignat (Ed.), *Low-Power CMOS Circuits Technology, Logic Design and CAD Tools*, Taylor & Francis Group, 2006.
- [15] A. Ferré, J. Figueras, Leakage in cmos nanometric technologies, in: C. Pignat (Ed.), *Low-Power CMOS Circuits Technology, Logic Design and CAD Tools*, Taylor & Francis Group, 2006.
- [16] J.P. Noel, O. Thomas, M.A. Jaud, O. Weber, T. Poiroux, C. Fenouillet-Beranger, P. Rivallin, P. Scheiblin, F. Andrieu, M. Vinet, O. Rozeau, F. Boeuf, O. Faynot, A. Amara, Multi-v_t utbb fdsoi device architectures for low-power cmos circuit, *IEEE Trans. Electron Devices* 58 (8) (2011) 2473–2482.
- [17] E. Beigné, A. Valentian, I. Miro-Panades, R. Wilson, P. Flatresse, F. Abouzeid, T. Benoist, C. Bernard, S. Bernard, O. Billoint, S. Clerc, B. Giraud, A. Grover, J.L. Coz, J.P. Noel, O. Thomas, Y. Thonnart, A 460 mhz at 397 mv, 2.6 ghz at 1.3 v, 32 bits vliw dsp embedding f max tracking, *IEEE J. Solid-State Circuits* 50 (1) (2015) 125–136.
- [18] L.L. Lewyn, T. Ytterdal, C. Wulff, K. Martin, Analog circuit design in nanoscale cmos technologies, *Proc. IEEE* 97 (10) (2009) 1687–1714.
- [19] R. Taco, I. Levi, M. Lanuzza, A. Fish, Extended exploration of low granularity back biasing control in 28nm utbb fd-soi technology, 2016 IEEE International Symposium on Circuits and Systems (ISCAS), (2016), pp. 41–44.
- [20] A. Bryant, J. Brown, P. Cottrell, M. Ketchen, J. Ellis-Monaghan, E.J. Nowak, Low-power cmos at v_{dd} = 4kt/q, Device Research Conference. Conference Digest, (2001), pp. 22–23.
- [21] L. Nazhandali, M. Minuth, T. Austin, Sensebench: toward an accurate evaluation of sensor network processors, *IEEE International. 2005 Proceedings of the IEEE Workload Characterization Symposium*, 2005. (2005), pp. 197–203.
- [22] N. Lotze, Y. Manoli, A 62 mv 0.13 μ m cmos standard-cell-based design technique using Schmitt-Trigger logic, 2011 IEEE International Solid-State Circuits Conference, (2011), pp. 340–342.
- [23] A.A. Vatanjou, T. Ytterdal, S. Aunet, Energy efficient sub/near-threshold ripple-carry adder in standard 65nm cmos, 2015 6th Asia Symposium on Quality Electronic Design (ASQED), (2015), pp. 7–12.
- [24] W. Lim, I. Lee, D. Sylvester, D. Blaauw, 8.2 batteryless sub-nw cortex-m0+; processor with dynamic leakage-suppression logic, 2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers, (2015), pp. 1–3.



Ali Asghar Vatanjou received his M.Sc. from University of Tabriz in 2010. During his M.Sc. degree, he worked towards pipeline ADC design. Since 2012, he started his PhD at Circuits and Systems Group, Norwegian University of Science and Technology (NTNU). He is currently working towards his PhD thesis at NTNU with main focus on ultra-low power and voltage digital circuit design. His main research interests includes ultra-low power and voltage VLSI and mixed-mode circuit design.



Even Låte graduated from the Norwegian University of Science and Technology (NTNU) with a M.Sc. degree in Electrical Engineering in 2014. His work on “Transaction Level Modeling of a PCI Express Root Complex” was awarded with the best master’s thesis prize within the field of microelectronics at NTNU. Since the fall of 2014, Even has been employed as a PhD candidate at the Circuits and Systems Group. His research focus is mitigation of the dark silicon effect by designing ultra-low power CMOS memory for use in the Single-ISA Heterogeneous Many-Core Computer* project at NTNU.

A.A. Vatanjou et al.



Trond Ytterdal received his M.Sc. and Ph.D. degrees in electrical engineering from the Norwegian Institute of Technology in 1990 and 1995, respectively. He is a Professor at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology. Current main research interests include design of analog integrated circuits and modeling of novel nanoscale transistors. He has authored and co-authored more than 180 scientific papers in international journals and conference proceedings. He is a co-author of three books. Prof. Ytterdal is a member of The Norwegian Academy of Technological Sciences and a Senior Member of IEEE.



Snorre Aunet received the Cand. Scient. degree in informatics from the University of Oslo (UiO) in 1993, and the Dr. Ing. degree in physical electronics from the Norwegian University of Science and Technology (NTNU), in 2002. He worked with ASIC design at Nordic VLSI from 1994 to 1997, prior to doctoral studies. Later, he worked at NTNU and UiO, and is currently a professor at NTNU. He has been a visiting researcher at the University of Paderborn (2004–2012) and at University of California, San Diego (UCSD), during the academic year 2015/2016. Research interests include ultra low-power mixed-signal circuits.

Microprocessors and Microsystems 56 (2018) 92–100



Article D: An Energy Efficient Level Shifter Capable of Logic Conversion From Sub-15 mV to 1.2 V

E. Låte, T. Ytterdal, and S. Aunet

Published in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. X, pp. Y-Z, 2020¹

DOI: <https://doi.org/10.1109/TCSII.2020.2966654>

D.1 Abstract

We propose architectural advances in low voltage, energy-efficient, level shifters. A write assist circuit is introduced to support the up-conversion of deep sub-threshold inputs. We also present an approach to reduce the leakage current in split signal output stages. A prototype was created in a 130 nm bulk CMOS process, and some samples were successfully tested for input voltages as low as 5 mV. For 10 measured samples, the mean functional, minimum, input voltage was 31.1 mV. By applying body bias to selected NMOS transistors to compensate for process and mismatch variation, the measured mean minimum input voltage was lowered to 14.3 mV. The leakage reduction in the split control output driver reduces the driver leakage by a factor of 8. The designed level shifter was found to be energy efficient compared to published structures, it consumed an average of 25.9 fJ per conversion in post-layout simulations, and the delay was measured to 21.08 ns at 300 mV input signals.

D.2 Relevance to the Thesis

It has been shown that energy efficiency and low power consumption can be achieved by lowering the supply voltage. Still, as a consequence of performance requirements or robustness, some circuit blocks require a certain voltage level to function adequately. Large, single-chip systems often deploy multiple voltage domains for reasons such as energy efficiency and mitigation of the dark silicon effect. When interfacing low voltage and regular voltage digital circuits, level shifters are required to enable communication between the different voltage domains. In the

¹References in article: [102] [105] [106] [107] [108] [14] [109] [110] [111] [112] [113] [114] [115] [116] [117] [118] [119] [120] [121]

process of developing subthreshold CMOS -memories and -adders throughout this thesis, subthreshold to above-threshold level shifters make these blocks compatible with standard CMOS libraries. In heterogeneous architectures, such as the SHMAC² project at NTNU, low voltage tiles, and level shifters may enable the continuation of multi-core scaling by mitigating the dark silicon effect by reducing thermal power dissipation.

D.3 Declaration of Authorship

I initiated the development of the concept and the design of the level shifter in august 2017. The idea of applying a write assist similar to that of the write operation for differential SRAM cells to support ultra-low voltage input signals came to me after having worked with SRAM cell design for three years. I made the prototype schematic, layout, and PCB design, as well as the chip-measurements reported in the paper. Trond Ytterdal and Snorre Aunet participated in technical discussions and supervised the development of the paper.

D.4 Manuscript

²The Single-ISA Heterogeneous MAny-core Computer (SHMAC) <https://www.ntnu.edu/ie/eecs/shmac/>

An Energy Efficient Level Shifter Capable of Logic Conversion From Sub-15 mV to 1.2 V

Even Låte, *Member, IEEE*, Trond Ytterdal, *Senior Member, IEEE*, and Snorre Aunet, *Senior Member, IEEE*

Abstract—We propose architectural advances in low voltage, energy-efficient, level shifters. A write assist circuit is introduced, to support the up-conversion of deep subthreshold inputs. We also present an approach to reduce the leakage current in split signal output stages. A prototype was created in a 130 nm bulk CMOS process, and some samples were successfully tested for input voltages as low as 5 mV. For 10 measured samples, the mean functional, minimum, input voltage was 31.1 mV. By applying body bias to selected NMOS transistors to compensate for process and mismatch variation, the measured mean minimum input voltage was lowered to 14.3 mV. The leakage reduction in the split control output driver reduces the driver leakage by a factor of 8. The designed level shifter was found to be energy efficient compared to published structures, it consumed an average of 25.9 fJ per conversion in post-layout simulations and the delay was measured to 21.08 ns at 300 mV input signals.

Index Terms—Level shifter, subthreshold, energy-efficient.

I. INTRODUCTION

ENERGY-efficiency continues to gain momentum with analog and digital IC designers. With a dynamic power consumption that is quadratically dependent on the supply voltage (VDD), voltage reduction on circuit blocks is an effective method to achieve global energy efficiency. In wake-up/surveillance circuit-blocks or state-holding elements, where the static energy dominates, VDD minimization is meaningful [1] since the minimum energy point is dependent on the activity factor. The same goes for sensor nodes in the internet of things [2]. Consequently, ultra-low voltage (ULV) level shifters (LS) are required in-between modules in systems of heterogeneous voltage domains.

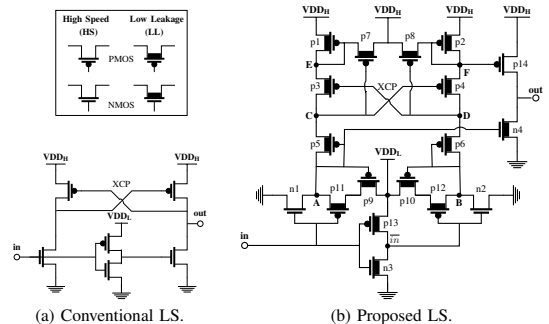
Fig. 1 (a) shows the commonly adopted differential cascode voltage switch (DCVS) LS. When VDD_L in the LS is lowered to interface deep subthreshold and regular superthreshold voltage domains, the strength of the subthreshold input gates decreases exponentially according to the approximated subthreshold drain current in Eq. I.1 [3]. A strength gap is formed since the strength of the current through the superthreshold cross-coupled pair (XCP) remains unchanged.

$$I_{DS} = I_{se} \frac{V_{GS} V_{th}}{nV_T} \left(1 - e^{-\frac{V_{DS}}{V_T}} \right) \quad (I.1)$$

Several LS architectures have been proposed to balance the strength of the pull-up network (PUN) and the pull-down network (PDN) in low voltage LSs. Building on the DCVS, transistor stacking, and diode coupling have shown to be energy efficient methods used to weaken the PUN [4]–[6].

The authors are with the Department of Electronic Systems, Norwegian University of Science and Technology (NTNU), O.S. Bragstads plass 2a, 7034 Trondheim Norway. E-mail: (even.laate@iet.ntnu.no).

Manuscript received Month DD, YYYY; revised Month DD, YYYY.



(a) Conventional LS.

(b) Proposed LS.

Fig. 1: Schematic of conventional and proposed LS.

Body biasing (BB) has also been used in [5], [6] to operate the LSs efficiently after production. Split gate-signal output drivers are used to reduce power in the output stage of multi-stage LSs [4], [7]. However, the transistors in the split signal driver are not turned on in a mutually exclusive fashion due to the diode voltage drop, opening for further power improvements. Subthreshold design methods, as well as flip-well structures and poly-biasing, were used in [6] to bring the mean minimum value of VDD_L to 43.5 mV. Architectural advances to lowering VDD_L to this point is needed for regular bulk CMOS processes where poly biasing and flip-well structures are unavailable.

In this paper, we introduce two concepts to ultra-low voltage and energy-efficient LS design. The first is to apply a voltage differential, during logic transitions, between the drains of the PMOS stacks in stage 1, similar to that of the bit-line differential required to write to SRAM cells. This allows robust level conversion at ultra-low input voltages. The latter concept is a method to reduce leakage in the output stage by pulling the low swing node of the PMOS up to the VDD rail cutting the leakage through it during logic low outputs. An LS deploying these two principles was manufactured in 130 nm CMOS. Measurement results reveal that the proposed LS is functional for a mean value of VDD_L as low as 31.1 mV with a VDD_H of 1.2 V. Using P-well bias to compensate for process and mismatch variation, the min. mean VDD_L was reduced to 14.3 mV with two samples working at $VDD_L=5$ mV. The LS has a measured conversion delay of 21.98 ns at $VDD_L=300$ mV and it consumes 25.9 fJ per transition in post-layout simulations.

Section II describes the LS. Results are presented in section III, discussions in IV and conclusions in V.

II. METHODOLOGY

A. The Proposed Level Shifter Architecture

1) *Functionality*: The schematic of the proposed two-stage LS is depicted in Fig. 1 (b). The input stage, has a cross-

coupled pair {p3,p4} at its core and a write assist scheme at the foot of each superthreshold PMOS stack {p1,p3,p5} & {p2,p4,p6}. The output stage, is a split gate signal driver inspired by [4] where the gate-signal of the pull-up device, node F, differs from the gate-signal of the pull-down device, node A. The cross-coupled pair is wrapped in between current limiting diodes {p1,p2} & {p5,p6} in a similar fashion to [4], [8] to generate low swing nodes, A and F, for the split signal output driver.

The write assist scheme of the LS consists of transistors {p11,p9} & {p12,p10} and is here explained for the left branch of the first stage only since the two branches are symmetrical. For the LS without the write assist transistors, when the input signal "in" transitions from high to low, transistor {n1} turns off and node A is initially left to float until the XCP is flipped and {p3} turns on. The state flipping is then performed by the pull-down device in the right branch only. By adding transistors {p11,p9} the left branch assists the right branch in flipping the state of the XCP, {p11} turns on when "in" goes low, since {p9} is on when node A is low, the write assist scheme immediately starts to increase the voltage of node A, even before the XCP is flipped. As the voltage of A increases, so does the voltage of node C and the drive current of transistor {p4} is reduced which assists transistor {n2} in overcoming it thus flipping the state of the XCP. Switch {p9} is only on during the transition itself, once node A reaches V_{DD_L} , V_{SG} of {p9} is 0 V, and the switch turns off, ensuring that no current flows between V_{DD_H} and V_{DD_L} once the state of the cross-coupled pair has toggled and transistor {p3} is turned on. By adding the write assist scheme to the LS the minimum functional V_{DDL} is lowered since the required strength subthreshold pull-down transistors {n1,n2} is relaxed. Simulations without and with the write assist scheme in the LS, transistors {p11,p9} & {p12,p10}, reveal an improvement in minimum functional V_{DDL} from 60 mV to 20 mV.

The main innovation related to the split signal driver, {p14,

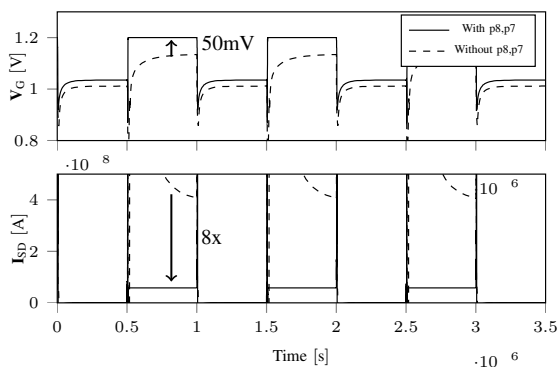


Fig. 2: Transients of {p14} w & w/o {p7,p8} at 1 MHz.

n4} in the output stage, is the reduction in short circuit current using transistors {p7, p8} to completely turn off {p14} for logic low output voltages. The input signals to the second stage, nodes A and F, are split and retrieved from low swing nodes in the first stage to reduce the leakage current through the off state device in the driver. The effect that the added

transistors have is illustrated in Fig. 2. The top waveform shows the gate signal of {p14} with and without transistors {p7,p8}, and the bottom waveform shows the corresponding source to drain current on the same time axis. Diode coupled MOS devices have a voltage drop from source to drain when they are conducting; this is the case for transistors {p1,p2, p5, p6}. The voltage drop across {p2} is undesired when we want to pull the drain of p2, node F, high to completely turn off {p14} when the voltage of the input node is low. This is the task of transistor {p8}; it pulls F up to V_{DD_H} whenever the drain of {p4}, node D, is low. {p7} is added to maintain the symmetry in the differential stacks in the first stage. The source to drain current of {p14} is reduced by a factor of 8 when {p14} is leaking in the off-state.

Device	Type	Well	Width	Length
p1,p2	PLLMOS4	N-well 1 V_{DD_H}	300 nm	130 nm
p3,p4	PLLMOS4	N-well 1 V_{DD_H}	300 nm	390 nm
p5,p6	PLLMOS4	N-well 1 V_{DD_H}	300 nm	130 nm
p7,p8	PLLMOS4	N-well 1 V_{DD_H}	300 nm	130 nm
p9,p10	PLLMOS4	N-well 2 V_{DD_L}	300 nm	130 nm
p11,p12	PLLMOS4	N-well 2 V_{DD_L}	300 nm	130 nm
p13	PLLMOS4	N-well 2 V_{DD_L}	1.75 μm	130 nm
p14	PHSMOS4	N-well 1 V_{DD_H}	600 nm	130 nm
n1,n2	NHSMOS4	P-well GND	825 nm	200 nm
n3	NLLMOS4	P-well GND	300 nm	200 nm
n4	NLLMOS4	P-well GND	300 nm	260 nm

TABLE I: Transistor types, well placement and dimensions.

2) *Transistor Type Selection & Sizing*: Table I shows the transistor types and the drawn dimensions of each device in the LS. LL transistor types are chosen by default to obtain a high energy efficiency of the circuit and HS transistors are chosen where strong devices are needed, ie. HS devices are chosen for the subthreshold pull-down transistors {n1, n2} to achieve ultra low voltage operation since they have to overcome the superthreshold pull-up devices. HS is also chosen for {p14} to decrease the delay of low to high transitions, node F is of particularly low amplitude since it is connected to the drain of the diode connected device p2. Comparative spectre simulations of the circuit with {n1, n2, p14} as both LL and HS show that the minimum functional V_{DDL} is improved from 120 mV to 20 mV using the HS devices as indicated, at the same time the propagation delay of the LS is improved from 1.5 us to 9.5 ns both at the cost of a large penalty in retention power of 3 x. The LS is mostly designed with transistors of minimum lengths and up-sized widths for robustness against mismatch and process variations. Exceptions in bold are sized differently for balance in the pull-up and pull-down network at low input voltages.

3) *Body Biasing as Device Variation Compensation*: To robustly up-convert logic values in the deep subthreshold domain, device up-sizing is applied to reduce device variation. Also, body biasing prior to manufacturing is applied to compensate for the incurred device variation. In a bulk MOSFET, adjusting the well bias modifies the depletion-charge which results in an altered threshold voltage of the device. The well-known body effect can be approximated by Eq. II.1 for the threshold voltage of an NMOS device located in a P-well.

$$V_{tn} = V_{tn0} + \gamma(\sqrt{V_{SB} + |2\phi_F|} - \sqrt{2\phi_F}) \quad (\text{II.1})$$

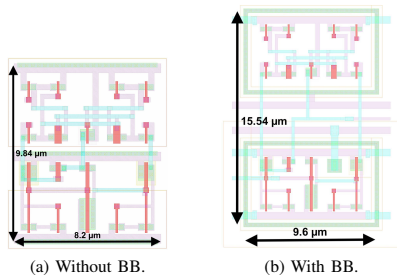


Fig. 3: Cell layout of the LS without and with BB support.

Here, V_{t0} is the threshold voltage with $V_{SB} = 0$, or a body bias equal to the source voltage and γ is the body-effect constant. For this implementation the bodies of transistors {p1,p2,p3,p4,p5,p6,p7,p8,p14} reside in a common high voltage N-well tied to V_{DDH} , transistors {p9,p10,p13} are placed in a low voltage N-well tied to V_{DDL} whereas the bodies of transistors {n1,n2,n3,n4} reside in an isolated P-Well that functions as a device variation compensation knob. By adjusting the voltage of the isolated P-well, the pull-down network and pull-up network may be balanced to make the LS functional at ultra-low input voltage levels.

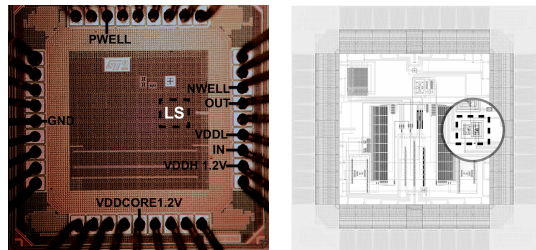
4) *Layout and Well Ties*: The resulting layout with two N-wells is depicted to the left in Fig. 3. The layout of the proposed LS is drawn with a height equal to $8.2 \mu\text{m}$, which is two times the height of the standard cells of the technology. To also be compatible with the standard cells in the horizontal direction, the width of the LS is $9.84 \mu\text{m}$, which is a multiple of the $0.41 \mu\text{m}$ horizontal routing grid of the standard cells in the 130 nm kit. For designs that are not incorporating standard cells from the design kit, the area of the LS can be shrunk to $6.44 \mu\text{m} \times 7.42 \mu\text{m} = 47.78 \mu\text{m}^2$.

To enable body biasing of the low voltage P-well and the high voltage N-well, the layout of the LS can be changed as depicted in Fig. 3 (b). Since the 130 nm technology used in this work supports isolated P-wells, an N-type guard ring and a deep buried N-well layer is placed around the bottom N-well to isolate the P-well from the substrate, enabling body biasing of the NMOS devices.

B. Simulation & Measurement Setup

Simulation results presented in this paper are obtained with models from the foundry-supplied 130 nm BiCMOS design kit. All Monte Carlo simulations are performed with 3σ accuracy using low discrepancy sampling. Simulated power and energy data for the LS is here reported with a single minimum sized inverter as load using netlists from the extracted layout in the typical extraction corner. The power consumption of the minimum sized inverter is not included in the results since the split signal output rails the output signal.

The layout in Fig. 3 (b) was inserted into a 1mm^2 pad-frame and manufactured in CMP's S13S18_2 run. 10 samples packaged in QFN36 open-lid packages were used to measure statistical spread. The chip micrograph and the chip layout of the circuit are shown in Fig. 4. A QFN36 test and burn-in clam-shell socket from Plastronics was mounted on a custom



(a) Chip micrograph. (b) Chip layout. Fig. 4: Manufactured circuit on a 1.25mm^2 die.

PCB to enable LS measurements of inter-die spread.

The power supply nodes, V_{DDL} , V_{DDH} , $V_{P\text{-well}}$ and V_{core} , were generated by a Rigol DP 832A digital PSU. Current drawn from the power supply nodes was measured with a Keithley 6485 Picoammeter. The input from the Agilent 33522A function generator was directly driving the input of the LS through the pad. The output of the LS was connected to the pad through a 4 stage, on-chip, pad-driver, and to the input of a Rohde & Schwarz RTE 1022 oscilloscope.

III. RESULTS

A. Simulation Results

1) *Post-layout Simulated Maximum Frequency and Propagation Delay*: Post layout simulation results show that the implemented LS can operate at up to 178 MHz at $V_{DDL}=600 \text{mV}$, at $V_{DDL}=300 \text{mV}$, the maximum conversion frequency is 43 MHz. Fig. 5 shows the simulated conversion delay as

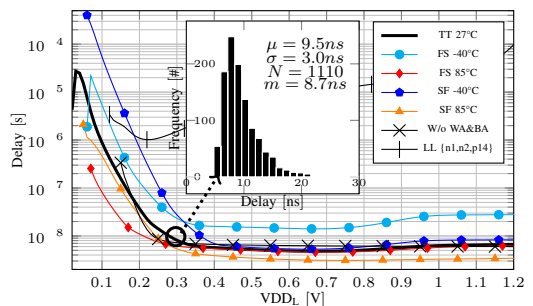
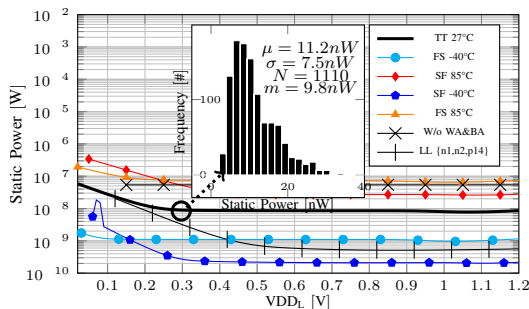


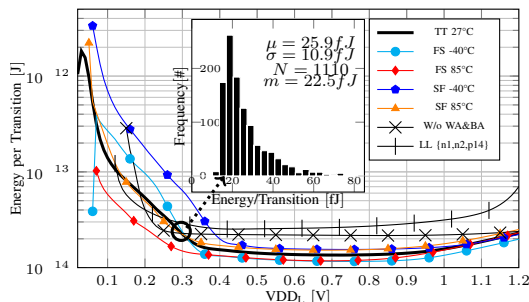
Fig. 5: Propagation delay as a function of V_{DDL} .

a function of V_{DDL} , i.e. the maximum delay between rising and falling propagation delay based on 25% and 75% logic limits. Process and temperature corners are also included. The nested plot shows the 3σ Monte Carlo distribution of conversion delays at $V_{DDL}=300 \text{mV}$, including both mismatch and process variations. Simulation results are also shown for the level shifter without the write assist and buffer pull-up circuit (W/o WA&BA), and for the original level shifter with transistors {n1,n2,p2} as LL devices (LL {n1,n2,p14}).

2) *Power Consumption*: The plots in Fig. 6 contain the standby power consumption for the LS as a function of V_{DDL} . The power that is drawn from both V_{DDL} and V_{DDH} are included in each graph. The nested plot shows the 3σ Monte Carlo distribution of the static power consumption

Fig. 6: Static leakage for $V(\text{in})=V_{\text{DD}_L}$ as a function of V_{DD_L} .

at $V_{\text{DD}_L}=300$ mV, including both mismatch and process variations. Static power is also shown for the LS without the write assist and buffer assist circuit as well as for the original LS circuit with LL devices for transistors {n1,n2,p14}.

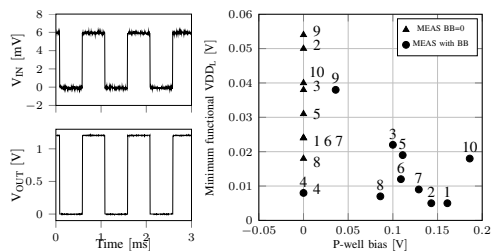
Fig. 7: Energy per transition as a function of V_{DD_L} .

3) *Energy Consumption*: The plots in Fig. 7 depicts the energy per transition for the LS as a function of V_{DD_L} at an input frequency of 1MHz. The active power that is drawn from both V_{DD_L} and V_{DD_H} during a conversion, is multiplied by the duration of the conversion. The nested plot shows the 3σ Monte Carlo distribution of the energy consumption per transition at $V_{\text{DD}_L}=300$ mV, including both mismatch and process variations. The energy is also shown for the LS without assist circuits and for the original LS with all LL devices.

B. Measurement Results

1) *Reliability at low V_{DD_L}* : The output of chip-sample number 1 is shown in the bottom waveform of Fig 8 (a) for a $V_{\text{DD}_L}=6$ mV and $V_{\text{P-well}} = 160$ mV and an input signal frequency of 1 kHz and 6 mV as shown in the top waveform. The LS successfully converts the 6 mV square signal to a 1.2 V square signal. The minimum functional V_{DD_L} for each manufactured chip sample is shown in Fig. 8 (b). The same parameter is also reported for the same dies after applying post-variation P-well bias compensation. Post-variation P-well biasing reduces the mean value of the minimum functional V_{DD_L} from 31.1 mV to 14.3 mV.

2) *Performance*: The mean value of the static power consumption for the 10 samples was measured to be 9.87 nW at

Fig. 8: Measured waveform of sample 1 at $V_{\text{DD}_L}=6$ mV and $V_{\text{P-well}} = 160$ mV at 1 kHz and V_{DD_L} for all 10 samples.

$V_{\text{DD}_L}=300$ mV. The maximum operating frequency measured with a functional chip output was 5.75 MHz at $V_{\text{DD}_L}=300$ mV. Similarly, the propagation delay from input to output of the chip was measured at a mean value of 21.98 ns at 300 mV. Note that this measurement includes the delay of the pad-driver, which is found to be 12.36 ns in spectre simulations, in the typical corner, with a simulated pad capacitance of 15 pF.

IV. DISCUSSION

Table II compares the performance of the proposed LS against published LSs in 65 nm, 90 nm, 130 nm, and 180 nm CMOS. The bold numbers highlight the best performing LS within each technology node for the column metric.

Area wise, the proposed LS takes up silicon area on par with the published LSs in 130 nm, the area consumption is also lower than all published LSs in 180 nm CMOS but more significant than most published 90 nm designs. The area of the proposed circuit may be reduced to $47.78\mu\text{m}^2$ if the circuit is not to be used with technology standard cells.

The propagation delay value for the proposed LS is 21.98 ns, which is an average measurement of the 10 chip samples that includes the propagation delay of the pad driver. In comparison, post-layout Monte Carlo average was found to be 9.5 ns. The proposed LS is approximately twice as fast as the listed LS in 130 nm. Versus the other technology nodes, the proposed LS has a lower propagation delay than all but [17] which is based on schematic simulations.

The mean measured static power consumption of the proposed LS topology is higher than most other reported LSs. This is a direct result from trading off static power consumption to lower the minimum functional V_{DD_L} to the ultra-low voltage domain. Using HS devices for transistors for {n1,n2,p14} further increases the leakage by 3 x. Actions to improve the leakage power while maintaining a low propagation delay and low min.- V_{DD_L} may in future versions of the design include sleep transistors to reduce static power consumption by 10x [18] or sleep modes using body biasing.

Energy efficiency is a key LS metric for high-speed applications in the low voltage domain. The 3σ Monte Carlo mean at 0.3 V was reported here to include the spike in current during the logic transitions. The energy per conversion of 25.9 fJ for this LS is a quarter of the reported conversion energies for the other LS' in the 130 nm technology. The proposed LS performs on par with the top reported values in the 65 nm node

	Tech.	Chip	T-count [#T]	Area [μm^2]	VDD _H [V]	VDD _L min [mV]	Delay [ns, mV]	Static Pwr [nW, mV]	Energy/tr [fJ/tr, mV]
This	130 nm	yes	18	80.69	1.2	31.1	21.98, 300	9.87, 300	25.9*, 300
This BB	130 nm	yes	18	178.40	1.2	14.3	21.98, 300	9.87, 300	25.9*, 300
[9]	65 nm	yes	16	16.80	1.2	165	52***, 300	5.44	-
[5]	65 nm	yes	12	7.8	1.2	120	66, 200	0.64 , 120	28 , 300
[10]	65 nm	yes	27	17.6	1.2	100	25.1 , 300	2.5, 300	30.7* , 300
[4]	90 nm	no	17	90.16	1.0	180*	32.00*, 180	2.5* , 180	17* , 180
[8]	90 nm	no	13	36.50	1.0	180*	21.80*, 200	6.4*, 200	74*, 200
[11]	90 nm	no	15	37.30	1.0	100*	16.6* , 200	8.7*, 200	77*, 200
[12]	130 nm	yes	10	-	1.2	100	40.00 , 200	8, 200	25000, 200
[13]	130 nm	yes	35	102.26	2.5 IO	300	41.50, 300	0.475 , 300	229, 300
[7]	130 nm	yes	12	71.94	2.5 IO	300	58.78, 230	0.724, 300	191 , 300
[14]	180 nm	yes	17	153.01	1.8	210	166.86***, 300	0.160, 300	39 , 300
[15]	180 nm	yes	14	108.80	1.8	100	31.7 , 400	0.055 , 400	173*, 400
[16]	180 nm	yes	12	160.00	1.8	80	95, 400	1.8, 100	118, 400

TABLE II: Comparison to top 3 published LSs in 65 nm, 90 nm, 130 nm and 180 nm CMOS. * denotes post-layout simulation results, ** denotes schematic simulation results and *** denotes results calculated from reported numbers.

and is all over the third lowest reported energy per conversion value listed in the table. Factors contributing towards the low energy consumption include the small propagation delay and low short circuit current during a logic value transition as a result of the split control buffer.

Finally, the minimum VDD_L reported here, a mean value of 31.1 mV or a mean value of 14.3 mV with body bias as post-variation compensation, is to the knowledge of the authors, the lowest reported value at the time of writing. The lowest reported minimum VDD_L was, until now, an LS designed and measured in 28 nm FDSOI [6]. Meindl and Davis have derived in [19] that the theoretical limit for the minimum supply voltage of an ideal conventional CMOS inverter at room temperature, where the subthreshold swing is 60 mV/dec, is 36 mV. The LS proposed here is functional for digital signals below this limit, which opens up new possibilities for voltage domain bridging in heterogeneous digital architectures.

V. CONCLUSION

We have presented a subthreshold to above-threshold LS with write assist for ultra-low-voltage operation and a leakage reduction technique in the split signal output driver. Simulation results and measurement results in a 130 nm process technology show that this circuit topology is energy efficient with transition energy of 25.9 fJ per transition for a 300 mV input signal. The functionality of the circuit is proved with measurements for input signals down to a mean value of 31.1 mV over 10 chip samples. The same architecture is functional down to a mean input voltage value of 14.3 mV when body bias is applied to the P-Well to compensate for process and mismatch variation.

REFERENCES

- [1] N. Lotze and Y. Manoli, "A 62 mV 0.13 μm cmos standard-cell-based design technique using schmitt-trigger logic," *IEEE journal of solid-state circuits*, vol. 47, no. 1, pp. 47–60, 2011.
- [2] —, "Ultra-sub-threshold operation of always-on digital circuits for iot applications by use of schmitt trigger gates," *IEEE Transactions on Circuits and Systems I*, vol. 64, no. 11, pp. 2920–2933, 2017.
- [3] E. A. Vittoz, "Very low power circuit design: fundamentals and limits," in *1993 IEEE International Symposium on Circuits and Systems*. IEEE, 1993, pp. 1439–1442.
- [4] A. Hasanbegovic and S. Aunet, "Low-power subthreshold to above threshold level shifter in 90 nm process," in *NORCHIP*. IEEE, 2009.
- [5] B. Mohammadi and J. N. Rodrigues, "A 65 nm single stage 28 fJ/cycle 0.12 to 1.2 v level-shifter," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 990–993.
- [6] A. A. Vatanjou, T. Ytterdal, and S. Aunet, "An ultra-low voltage and low-energy level shifter in 28 nm utbb-fdsoi," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 6, 2019.
- [7] Y. Kim, Y. Lee, D. Sylvester, and D. Blaauw, "Slc: Split-control level converter for dense and stable wide-range voltage conversion," in *ESSCIRC, 2012 Proceedings of the*. IEEE, 2012, pp. 478–481.
- [8] M. Lanuzza, P. Corsonello, and S. Perri, "Low-power level shifter for multi-supply voltage designs," *IEEE Transactions on Circuits and Systems II*, vol. 59, no. 12, pp. 922–926, 2012.
- [9] S.-C. Luo, C.-J. Huang, and Y.-H. Chu, "A wide-range level shifter using a modified wilson current mirror hybrid buffer," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 1656–1665, 2014.
- [10] W. Zhao, A. B. Alvarez, and Y. Ha, "A 65-nm 25.1-ns 30.7-fJ robust sub-threshold level shifter with wide conversion range," *IEEE Transactions on Circuits and Systems II*, vol. 62, no. 7, pp. 671–675, 2015.
- [11] M. Lanuzza, P. Corsonello, and S. Perri, "Fast and wide range voltage conversion in multisupply voltage designs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, pp. 388–391, 2015.
- [12] T.-H. Chen, J. Chen, and L. T. Clark, "Subthreshold to above threshold level shifter design," *Journal of Low Power Electronics*, vol. 2, no. 2, pp. 251–258, 2006.
- [13] Y. Kim, D. Sylvester, and D. Blaauw, "Lc 2: Limited contention level converter for robust wide-range voltage conversion," in *VLSI Circuits (VLSIC), 2011 Symposium on*. IEEE, 2011, pp. 188–189.
- [14] J. Zhou, C. Wang, X. Liu, X. Zhang, and M. Je, "An ultra-low voltage level shifter using revised wilson current mirror for fast and energy-efficient wide-range voltage conversion from sub-threshold to i/o voltage," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 3, pp. 697–706, 2015.
- [15] M. Lanuzza, F. Crupi, S. Rao, R. De Rose, S. Strangio, and G. Iannaccone, "An ultralow-voltage energy-efficient level shifter," *IEEE Transactions on Circuits and Systems II*, vol. 64, no. 1, pp. 61–65, 2017.
- [16] R. Lotfi, M. Saberi, S. R. Hosseini, A. R. Ahmadi-Mehr, and R. B. Staszewski, "Energy-efficient wide-range voltage level shifters reaching 4.2 fJ/transition," *IEEE Solid-State Circuits Letters*, vol. 1, no. 2, pp. 34–37, 2018.
- [17] M. Ashouei, H. Luijmes, J. Stuijt, and J. Huisken, "Novel wide voltage range level shifter for near-threshold designs," in *Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 285–288.
- [18] A. Hasanbegovic and S. Aunet, "Low-power subthreshold to above threshold level shifters in 90 nm and 65 nm process," *Microprocessors and Microsystems*, vol. 35, no. 1, pp. 1–9, 2011.
- [19] J. D. Meindl and J. A. Davis, "The fundamental limit on binary switching energy for terascale integration (tsi)," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 10, pp. 1515–1516, 2000.

Benefiting from State Dependencies in Asymmetric SRAM Cells through Conditional Word-flipping

E. Låte, T. Ytterdal, and S. Aunet

Published in *IEEE TRANSACTIONS on Very Large Scale Integration (VLSI) Systems*, vol. 28, pp. 2223-2227, October 2020¹

DOI: <https://doi.org/10.1109/TVLSI.2020.3013139>

E.1 Abstract

This brief presents an approach that dynamically exploits content dependencies in asymmetric memory cells. By using a capacitive, logic-value majority circuit and an extra column of memory cells, words are conditionally flipped during write operations to reach the more beneficial state for storage. A 1-kb SRAM block of low-voltage memory cells was implemented and manufactured in a 130-nm CMOS. The memory cells were made writable and read-stable at low supply voltages with a single-ended write and single-ended read structure using six multithreshold transistors that give rise to an asymmetric retention power. At a supply voltage of 350 mV, the content-dependent leakage power in the asymmetric memory cell is 23 times smaller when storing logic '1's compared with logic '0's. A derived statistical model suggests that the mean, wordwise, static power savings of the word-flipping scheme become 15.70% for 8-bit words of uniform bit probability. For the implemented SRAM macro, the mean improvement for the retention power is, including flip logic and decoder and driver overheads, found to be 14.69%. In boundary tests, by writing all words full of undesired values, the power saving becomes 80.37%, while writing all words full of desired values causes a power penalty of 6.14%. Measurement results confirm the improvement in retention power with a ten-chip mean improvement of 11.93% for the same data set.

E.2 Relevance to the Thesis

A scheme that targets SRAM macros consisting of asymmetric SRAM cells to exploit state dependencies is directly relevant to this thesis since asymmetries are

¹References in article: [122] [49] [123] [124] [125] [126] [127] [128]

commonly introduced in low-voltage SRAM design as write- and read- assists. Depending on the state dependency of the memory cell, the scheme may be used to improve various performance metrics. However, the most intuitive and measurable metric is retention power consumption.

E.3 Declaration of Authorship

The idea of conditionally flipping written words in write operations to SRAM macros of asymmetric memory cells came to me after re-creating asymmetric memory cells from an article observing that the reported SRAM leakage was reported for the macro with all cells in the cheapest state retention-current-wise. The majority decision circuit was invented by me with inspiration in SAR-ADCs. I made the prototype schematic, layout, and PCB design, as well as the chip-measurements reported in the paper. Trond Ytterdal and Snorre Aunet participated in technical discussions and supervised the development of the paper.

E.4 Manuscript

Transactions Briefs

Benefiting From State Dependencies in Asymmetric SRAM Cells Through Conditional Word-Flipping

Even Låte^{1b}, Trond Ytterdal^{1b}, and Snorre Aunet^{1b}

Abstract—This brief presents an approach that dynamically exploits content dependencies in asymmetric memory cells. By using a capacitive, logic-value majority circuit and an extra column of memory cells, words are conditionally flipped during write operations to reach the more beneficial state for storage. A 1-kb SRAM block of low-voltage memory cells was implemented and manufactured in a 130-nm CMOS. The memory cells were made writable and read-stable at low supply voltages with a single-ended write and single-ended read structure using six multithreshold transistors that give rise to an asymmetric retention power. At a supply voltage of 350 mV, the content-dependent leakage power in the asymmetric memory cell is 23 times smaller when storing logic “1”s compared with logic “0”s. A derived statistical model suggests that the mean, wordwise, static power savings of the word-flipping scheme become 15.70% for 8-bit words of uniform bit probability. For the implemented SRAM macro, the mean improvement for the retention power is, including flip logic and decoder and driver overheads, found to be 14.69%. In boundary tests, by writing all words full of undesired values, the power saving becomes 80.37%, while writing all words full of desired values causes a power penalty of 6.14%. Measurement results confirm the improvement in retention power with a ten-chip mean improvement of 11.93% for the same data set.

Index Terms—Majority voter, memory, word-flip.

I. INTRODUCTION

Bus-invert coding [1], i.e., the reduction of hamming distance between the present state of the bus and the next state of the bus by conditionally flipping the bits in the next state’s data, was proposed as a method of decreasing the activity factor on the high-capacitance nodes in buses, thus decreasing the I/O peak power consumption. More recently, numerous published research articles motivate the design of asymmetric SRAM cells for write-assist properties in low-voltage operation [2], [3], data-dependent reinforcement against soft errors in content-addressable memories [4], [5], and data-dependent leakage improvements in caches where the data stream contains a majority of either of the binary logic values [6]. Chang *et al.* [7] researched the distribution of binary values for both the instruction and the data cache using the SPEC2000 benchmark suite, and they report a majority of logic 0’s that they statically exploit using asymmetric SRAM cells for reduction of write power. A dynamic solution is reported in [8], where a majority circuit is used to reduce bitline switching power for symmetric memory cells.

In this brief, the concepts of bus-invert coding and state optimization of asymmetric memory cells are combined. We propose a memory system of asymmetric memory cells where a digital-to-analog converter (DAC)-inspired majority-voter constructed from capacitors is used to flip written words conditionally, hereby referred to as the word-flipping scheme. Demonstrating the word-flipping system using a near-threshold SRAM cell with content-dependent retention power

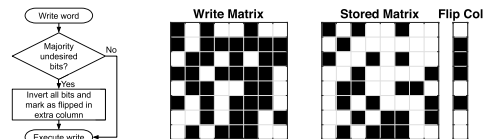


Fig. 1. Conditional word-flipping for an array of eight words of 8 bits. White box indicates the desired logic value.

in a 1-kb macro shows a simulated improvement in system leakage power of 14.69% and a measured mean improvement of 11.93%.

II. WORD-FLIPPING SCHEME

By combining the state dependence of asymmetric SRAM cells common in low-voltage memories [2], [6] and the concept of actively flipping words dependent on the data [1], [8], we obtain the word-flipping scheme shown in Fig. 1. Using a memory cell in a memory matrix that has a state-dependent benefit allows for a greedy, word-oriented, bit flipping system by including extra peripheral logic and an additional column of memory cells to keep track of the flipped words. Words are evaluated during write operations and are conditionally flipped if such an operation results in a more desired outcome.

A. Statistical Model

Let us assume for now that the bit concatenation for a write operation in an SRAM consists of several repeated, independent trials equal to the width of the SRAM and that each trial has a binary outcome. Then, the number of logic “1”s and the number of logic “0”s in a single concatenated word follow a binomial probability distribution. If the probability of each binary outcome for a trial is uniform, then the binomial probability distribution for the number of a given logic value in a word becomes a symmetric bell-shape. Fig. 2 shows the probability distributions for the number of undesired logic values in a word. For each word length, the probability that the input data contain a majority of the undesired binary value is then equal to the sum of the probabilities represented by the black bars. Since conventional word widths are powers of two, they all have a single most likely configuration equal to their mean value marked in gray. There is no point in flipping a word with equal amounts of zeros and ones, so the majority circuit should be configured not to flip the word in this scenario, primarily if the undesired state is used to mark a flip in the additional flip column. The flip balance may be further skewed if the flip operation, itself, is costly regarding overhead.

The probability of a word-flip decision to occur during a write operation, p_{flip} , is equal to the sum of the black bars shown in Fig. 2. Denoting the number of undesired bits in the input vector x , an even SRAM word width n , and a bitwise probability for observing an undesired logic value $p_{\bar{d}}$ gives us the upper cumulative distribution function the following equation:

$$p_{\text{flip}} = p \left(x > \frac{n}{2} \right) = \sum_{k=\frac{n}{2}+1}^n \binom{n}{k} (p_{\bar{d}})^k (1 - p_{\bar{d}})^{n-k}. \quad (1)$$

Manuscript received January 18, 2020; revised April 21, 2020 and June 15, 2020; accepted July 10, 2020. Date of publication August 14, 2020; date of current version September 25, 2020. (Corresponding author: Even Låte.)

The authors are with the Department of Electronics and Telecommunications, Norges teknisk-naturvitenskapelige universitet, 7030 Trondheim, Norway (e-mail: even.late@iet.ntnu.no).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVLSI.2020.3013139

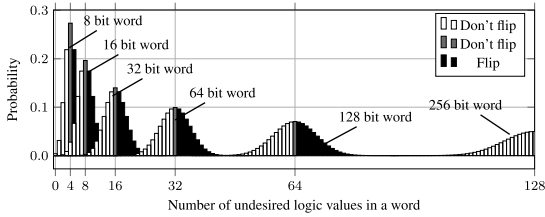


Fig. 2. Probability density functions (pdfs) of binary configurations for different word widths.

As a baseline throughout this brief, we compare the performance benefits of deploying the flip scheme on a memory matrix of asymmetric memory cells to the performance of the same memory matrix without the flip scheme. For the comparative baseline, the expected number of undesired and desired binary values in a word (x and \bar{x}) is given by the expected value of a binomial process

$$E[x] = p_{\bar{d}} \cdot n \quad E[\bar{x}] = n - E[x]. \quad (2)$$

For the memory matrix with the majority bit-flipping scheme implemented, the expected number of the undesired and desired logic values in a word, included the flip indication bit, given that a flip decision is made (*flipped*) and the word is not flipped (*flipped*), becomes the expressions given in

$$E[x|\overline{\text{flipped}}] = \frac{\sum_{k=0}^{\frac{n}{2}} k \cdot \binom{n}{k} (p_{\bar{d}})^k (1 - p_{\bar{d}})^{n-k}}{X \sum_{k=0}^{\frac{n}{2}} \binom{n}{k} (p_{\bar{d}})^k (1 - p_{\bar{d}})^{n-k}}$$

$$E[\bar{x}|\overline{\text{flipped}}] = 1 + n - E[x|\overline{\text{flipped}}]$$

$$E[x|\text{flipped}] = 1 + \frac{\sum_{k=\frac{n}{2}+1}^n (n-k) \cdot \binom{n}{k} (p_{\bar{d}})^k (1 - p_{\bar{d}})^{n-k}}{\sum_{k=\frac{n}{2}+1}^n \binom{n}{k} (p_{\bar{d}})^k (1 - p_{\bar{d}})^{n-k}}$$

$$E[\bar{x}|\text{flipped}] = n - (E[x|\text{flipped}] - 1). \quad (3)$$

Here, the flip column holds the undesired state for flipped words and the desired state for unflipped words.

B. Majority Flipping for Leakage Power Optimization

One scenario where the majority inversion scheme can be applied is to minimize retention power consumption for an array of asymmetric memory cells. Each memory cell has a desired state and an undesired state, with a corresponding power consumption of P_d and $P_{\bar{d}}$, respectively. The expected intraword power consumption $E[P_{\text{word}}]$ of the baseline, i.e., a word without an implemented flip scheme, is then given by the following equation:

$$E[P_{\text{word}}] = P_{\bar{d}} E[x] + P_d E[\bar{x}]. \quad (4)$$

By deploying the flipping scheme, the expected intraword retention power becomes the weighted sum of both the expected power consumption in the case of a flipped word and the expected power consumption when no word flipping is performed as follows:

$$E[P_{\text{word}}] = p_{\text{flip}} (P_{\bar{d}} E[x|\text{flipped}] + P_d E[\bar{x}|\text{flipped}]) + (1 - p_{\text{flip}}) (P_{\bar{d}} E[x|\overline{\text{flipped}}] + P_d E[\bar{x}|\overline{\text{flipped}}]). \quad (5)$$

The effectiveness of the flipping scheme varies with the word length n and the bitwise, leakage quotient, $P_{\bar{d}}/P_d$. The impact that these two factors make is shown in Fig. 3 with n ranging from 4 to 1024 and with $P_{\bar{d}}/P_d$ ranging from 1 to 1024. From the

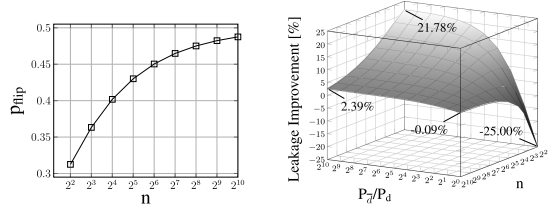


Fig. 3. Impact of n and $P_{\bar{d}}/P_d$ on p_{flip} and the improvement in $E[P_{\text{word}}]$ from deploying the word-flipping scheme.

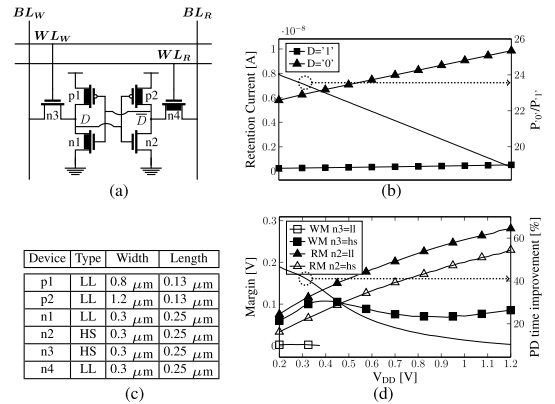


Fig. 4. (a) Cell schematic (asymmetric SRAM cell). (b) Cell leakage (state dependency versus V_{DD}). (c) Device sizing (transistor sizes). (d) Write/read margins as well as read bitline pull-down time (margins and read delay).

figure, it is evident that the relative impact of the majority flip scheme is more significant for short words and high-power quotients.

C. Asymmetric Memory Cell

A near-threshold SRAM cell may be realized in a 130-nm technology with a single-ended write and single-ended read structure, as shown in Fig. 4(a). By using a low-threshold pass-gate (hs) as device n3, the cell is made writeable compared with using a low-leakage (ll) device, as seen in the curve in Fig. 4(d). Similarly, by using a low-threshold device as transistor n2, the read margin is improved, and the bitline pull-down time is improved by 40%, as shown in Fig. 4(b), according to Spectre simulations. These asymmetric characteristics give rise to the data dependence that we will exploit with the majority bit-flipping scheme. For a supply voltage of 350 mV, $P_d = P_{1'} = 278$ pA and $P_{\bar{d}} = P_{0'} = 6.4$ nA for the memory cell. The difference in the data retention current is mainly caused by the OFF-state leakage through the low-threshold voltage device n2 when the high threshold voltage device p2 is ON, and vice versa. With a statewise retention quotient of 23 for storing a logic "0" to storing a logic "1" at the data node (D), the system should, thus, aim to minimize the number of logic "0"s in each stored word.

Since the relative impact of the majority flip scheme is more significant for short word widths, as shown in Fig. 3, we implement the proof of concept with a word width of 8 bits. Wide SRAMs may be created by instantiating multiple majority-flip blocks in parallel. The expected number of logic values in the 8-bit word and the corresponding expected retention power for both the baseline and the proposed scheme is shown in Table I. For a $p_{\bar{d}} = 0.5$, the word-flipping scheme saves 15.70% per word of the baseline's retention power.

TABLE I
EXPECTED IMPROVEMENT, $n = 8$, $V_{DD} = 350$ mV

$P_{\bar{q}}$	$E_{[x]}$	$E_{[\bar{x}]}$	$E_{[Pwr]}$	$E_{[x \bar{x}]}$	$E_{[\bar{x} \bar{x}]}$	$E_{[x f]}$	$E_{[\bar{x} f]}$	$E_{[Pwr]}$	P_{saved}
0.00	0.00	8.00	0.78 nW	0.00	9.00	1.00	8.00	0.88 nW	-12.50 %
0.25	2.00	6.00	5.06 nW	1.91	7.09	2.71	6.29	5.02 nW	0.92 %
0.50	4.00	4.00	9.35 nW	3.14	5.86	3.50	5.51	7.88 nW	15.70 %
0.75	6.00	2.00	13.64 nW	3.72	5.28	3.83	5.17	9.06 nW	33.58 %
1.00	8.00	0.00	17.92 nW	8.00	1.00	1.00	8.00	3.02 nW	83.16 %

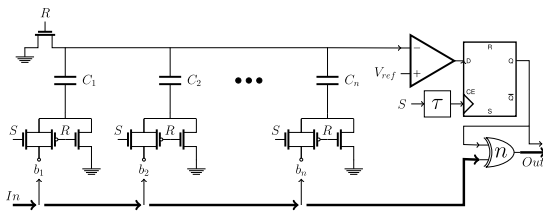


Fig. 5. Proposed majority-1-based flip decision circuit.

D. Majority Decision Circuitry

For the flip scheme, a majority voter is placed in the write path of the SRAM, and the flip-back circuitry is placed in the read path of the SRAM to evaluate the flip-state column and conditionally flip back the word. Low-voltage, digital majority voters can be synthesized with a low-voltage CMOS library consisting of two-input logic gates. However, this does not scale well with the word width, primarily when the main cost of the word-flipping scheme already is a larger die area. Multiple logic synthesis runs, and curve fitting with MATLAB shows a quadratic polynomial trend for the gate count, where a word length of 8 takes up 44 logic gates, a word length of 16 needs 221 gates, and a word length of 32 bits requires 1022 gates. An alternative implementation of the majority voter is a sequential bubble sort algorithm into a logic vector that is thermometer encoded. However, this approach requires too many cycles to complete to be in the write path of an SRAM. Consequently, we created the two-phase, DAC-inspired, majority circuit shown in Fig. 5.

The functionality of the majority circuit is similar to that of a capacitive DAC. The bottom plates of the capacitor array are charged to V_{DD} or V_{SS} according to the input when new data are applied to the SRAM, and the top plates are simultaneously set to V_{SS} potential. As soon as the signal S is set high, the charge on the capacitors is redistributed to the capacitors, and the top plate voltage settles somewhere between V_{DD} or V_{SS} depending on the input data b_1 – b_n during the sample phase.

Since the switching threshold for the majority circuit lies just below $V_{DD}/2$, a Schmitt trigger is carefully sized to act as the comparator and reference voltage. Using a Schmitt trigger instead of a regular buffer gives a steeper transition curve, allowing for a more precise top-plate evaluation. The unit capacitors in the majority decision capacitor array were, in this brief, chosen to be the smallest available metal–insulator–metal (MIM) capacitors available in the design kit, i.e., 29.12 fF. The top plate of the capacitor array, which is common to all eight capacitors, was laid out with equal distance from each top plate to the input of the comparator to weigh each input equally in a majority vote decision. The 8-bit design was verified at the typical corner for mismatch and process variations. The energy that is consumed per write operation, in the majority decision circuit, was found to be 35.44 fJ/operation at a frequency of 5 MHz. The majority voter consists of both capacitive and static components; for high frequencies, the energy consumption is dominated by the

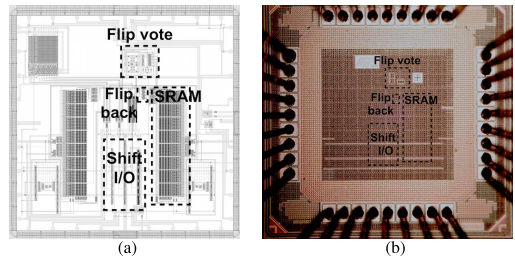


Fig. 6. Manufactured circuit on a $1.25\text{-}\mu\text{m}^2$ die. (a) Chip layout. (b) Chip micrograph.

charging and discharging of capacitors, whereas, for low frequencies, the energy consumption is dominated by the leakage component.

E. Simulation and Measurement Setup

The $128 \times (8 + 1)$ proof of concept extends the baseline matrix with one column and includes the majority flip logic on the input as well as the flip-back circuitry on the output. The postlayout netlist was extracted and was then used in system-level simulations with Spectre. Simulation results presented in this brief were obtained with models from the foundry-supplied 130-nm bipolar CMOS (BiCMOS) design kit.

A chip-level layout of the example system was created, as shown in Fig. 6(a). To write and read data to and from the pad constrained SRAM prototype, the data were serially shifted in and out from the circuit with shift registers that are shared with other modules on the chip. To compare the retention power of the array with the word-flip functionality enabled against the baseline array, the flip decision was made overridable by routing an “enable-signal” to the pad.

The layout shown in Fig. 6(a) was manufactured in CMP’s S13S18_2 run and then packaged in QFN36 open-lid packages. The chip micrograph is shown in Fig. 6(b). A QFN36 test and burn-in clam-shell socket from Plastronics was soldered on to a custom PCB to allow measurements of inter die spread of the SRAM retention current.

The power supplies for the I/O, flip vote, flip back, and the SRAM were generated by a Rigol DP 832A digital PSU. Current drawn from the power supply nodes was measured with a Keithley 6485 Picoammeter. The input to the memory macro was shifted in using an STM Nucleo L152RE microcontroller via optocouplers for voltage conversion. The microcontroller was also used to trig an Agilent 33522A function generator for the SRAM clock. The output of the readout shift-register was connected to the microcontroller’s analog-to-digital converter (ADC) to verify the read and write functionality of the macro. The oscilloscope used was a Rohde & Schwarz RTE 1022.

A 128×8 binary matrix was randomly drawn once and applied to the memory matrix in both simulations and tests for fair comparisons. The probability of a logic “1” was set equal to that of a logic “0” in the stochastic process, ensuring a realistic set of test data that are uniformly binomially distributed.

III. RESULTS

A. Simulation Results

The simulated retention power consumption for both the 1-kb (128×8) baseline (B) block and the 1-kb ($128 \times (8 + 1)$) proposed scheme (S) is listed in Table II. The word in the leftmost column was written to all words in the SRAM blocks. The baseline memory stored the written data as the column $\text{Stored}_{\text{Baseline}}$, and the proposed memory system conditionally flipped all words and stored the words

TABLE II
SPECTRE SIMULATION RESULTS FOR THE 128 × 8 AND 128 × (8 + 1)
MEMORY SYSTEM AT VDD = 350 mV

Data in All words	Stored _{Baseline}	P _{Baseline}	Stored _{Scheme}	P _{Scheme}	(P _S -P _B)/P _B
1111 1111 (0x0F)	1111 1111 (0x0F)	202.83 μ W	1 1111 1111 (1+0x0F)	215.28 μ W	-6.14%
1111 1110 (0x0F)	1111 1110 (0x0F)	479.99 μ W	1 1111 1110 (1+0x0F)	492.45 μ W	-2.59%
1111 1100 (0x0F)	1111 1100 (0x0F)	757.55 μ W	1 1111 1100 (1+0x0F)	770.00 μ W	-1.64%
1111 1000 (0x0F)	1111 1000 (0x0F)	1032.99 μ W	1 1111 1000 (1+0x0F)	1045.45 μ W	-1.21%
1111 0000 (0x0F)	1111 0000 (0x0F)	1312.30 μ W	1 1111 0000 (1+0x0F)	1324.75 μ W	-0.95%
110 0000 (0x70)	110 0000 (0x70)	1571.78 μ W	0 0001 1111 (0+0x0F)	1304.10 μ W	17.03%
110 0000 (0x30)	1100 0000 (0x30)	1850.38 μ W	0 0011 1111 (0+0x0F)	1025.50 μ W	44.58%
1000 0000 (0x10)	1000 0000 (0x10)	2127.58 μ W	0 0111 1111 (0+0x0F)	750.05 μ W	64.75%
0000 0000 (0x00)	0000 0000 (0x00)	2405.13 μ W	0 1111 1111 (0+0x0F)	472.15 μ W	80.37%
Random w $p_T=0.5$	Same words	1346.95 μ W	Cond. flipped words	1149.05 μ W	14.69%

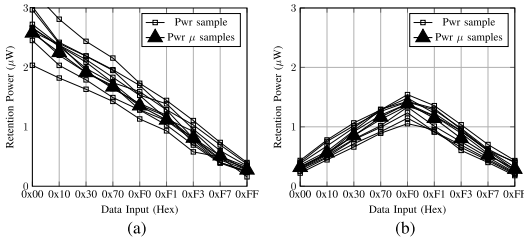


Fig. 7. Measured retention power for ten chip samples when writing the words on the x-axis to the entire memory macro. (a) Without cond. word-flip. (b) With cond. word-flip.

TABLE III
CHIP MEAN MEASUREMENT RESULTS FOR THE 128 × (8 + 1)
MEMORY SYSTEM AT VDD = 350 mV

Data in all words	Stored _{Enabled}	μ P _{Enabled}	Stored _{Enabled}	μ P _{Enabled}	(P _E -P _D)/P _D
1111 1111 (0x0F)	1 1111 1111 (0x0F)	274.18 μ W	1 1111 1111 (1+0x0F)	287.42 μ W	-4.83%
1111 1110 (0x0F)	1 1111 1110 (0x0F)	512.42 μ W	1 1111 1110 (1+0x0F)	525.48 μ W	-2.55%
1111 1100 (0x0F)	1 1111 1100 (0x0F)	810.67 μ W	1 1111 1100 (1+0x0F)	804.83 μ W	0.72%
1111 1000 (0x0F)	1 1111 1000 (0x0F)	1115.07 μ W	1 1111 1000 (1+0x0F)	1143.09 μ W	-2.51%
1111 0000 (0x0F)	1 1111 0000 (0x0F)	1353.76 μ W	1 1111 0000 (1+0x0F)	1389.43 μ W	-2.64%
110 0000 (0x70)	1 1110 0000 (0x70)	1671.61 μ W	0 0001 1111 (0+0x0F)	1164.36 μ W	30.03%
110 0000 (0x30)	1 1100 0000 (0x30)	1914.20 μ W	0 0011 1111 (0+0x0F)	850.01 μ W	55.39%
1000 0000 (0x10)	1 0000 0000 (0x10)	2251.81 μ W	0 0111 1111 (0+0x0F)	563.49 μ W	74.97%
0000 0000 (0x00)	1 0000 0000 (0x00)	2579.57 μ W	0 1111 1111 (0+0x0F)	317.71 μ W	87.68%
Random w $p_T=0.5$	Same words	1393.95 μ W	Cond. flipped words	1227.05 μ W	11.93%

listed in column Stored_{Scheme}. The bottom row in the table shows the simulated retention power for both the baseline macro and the word-flipping macro for the random test data set. For each data input, the scheme's percentage of improvement in retention power (P_S-P_B)/P_B is listed to the right. Negative percentages indicate that for the given input word, the overhead power of the flip logic is greater than the benefit of the scheme.

B. Measurement Results

Measurement results of the retention power consumption of the manufactured SRAM macro are shown in Fig. 7. Fig. 7(a) presents the power consumption of the circuit with the word-flipping scheme disabled (D), while Fig. 7(b) shows the same metric with the scheme enabled (E). Both measurement results are presented as a function of the data input that is written to the entire memory array for each retention measurement sample point. The measurement results for ten chip samples are shown in each plot, with the mean retention power for each word marked with triangles. The measured mean retention power of the 1-kb memory macro is listed in Table III, both with the word-flipping scheme disabled and enabled. The mean retention power from writing 128 random words to the entire memory macro is shown in the bottom row.

IV. DISCUSSION

The theoretical model that was developed for the expected, word-wise, retention power of the word-flipping scheme suggested an

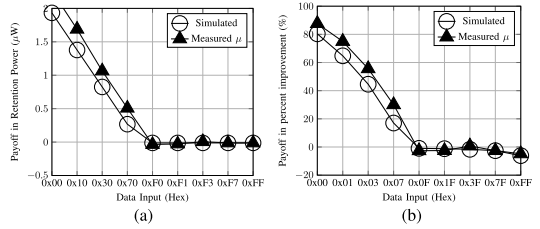


Fig. 8. Improvement in retention power consumption as a function of input data. (a) Scheme payoff in μ W. (b) Scheme payoff in %.

improvement of 15.70% in retention power from deploying the word-flipping scheme on the designed SRAM macro assuming uniformly binomially distributed data. The expected retention power of a word in the cell array was 7.88 nW per word using the scheme compared with 9.35 nW per baseline word, neglecting peripheral circuitry.

The simulation results in Table II illustrate the benefit of the word-flipping scheme for memory macros that consist of asymmetric memory cells. The higher the number of undesired bits, the greater the improvement of the proposed scheme. For data that are uniformly binomially distributed, the total improvement in the retention power of the scheme compared with the baseline is 14.69%. The average retention power per word becomes 8.98 nW per word instead of 10.52 nW for the binomially distributed data. The simulated percentage improvement in consumed retention power is lower than the percentage improvement of the theoretical model since the simulations are performed on a macro level with peripheral circuitry and flip logic.

For the measurement results, the proposed scheme was implemented with an enable switch, so the comparison is between having the scheme enabled or disabled, i.e., having the flip status column idle in the cheapest state and the word-flip decision circuit bypassed when the system is disabled. Apart from that, the SRAM macro that was measured is equivalent to the simulated one. Also, for the measured results, the advantage of the scheme is evident from the percentage improvement in Table III. The numbers reported here are the mean values of the ten measured chip samples for each combination of data input. The mean measurement results confirm that the percentage improvement in retention power grows with the number of undesired bits contained within a word. The mean percentage improvement of the ten chip samples with and without the scheme enabled was measured to be 11.93%.

The simulated and measured reduction in retention power consumption of the 1-kb SRAM from applying the word-flipping scheme is shown in Fig. 8. The measured mean power improvement follows the simulated power improvement with small deviations that could be caused by leakage variation in the SRAM cells or the peripheral logic. The upside of having the flip scheme implemented for input words with a majority of logic 0's is significantly greater than the overhead of having the flip scheme implemented for input words consisting of a majority of logic 1's. This is also explained in the theoretical model. With an expected fraction of logic 0's in a word equal to 50%, the retention-current overhead of implementing the flip scheme will pay off for large retention current quotients. The approach of statically exploiting nonuniform data probability of words in caches [6] was shown to have a 4.5 \times and 3.8 \times leakage reductions for a 78.8% majority and a 62.9% majority of zeroes, respectively. For uniformly distributed data, the asymmetric cells alone have no leakage reduction at all, for data that is distributed with a majority of logic ones the

scheme in [6] actually has a high leakage penalty. It could, therefore, benefit from adopting the dynamic flip approach that we present here to save retention power both for uniform data and data that are skewed in the undesired direction.

The word-flipping scheme is insensitive to errors in the majority voter. If an erroneous flip decision occurs, which is most likely to occur when the number of zeroes is close to the number of ones in a word, the memory simply stores the word in the nonideal state. For the exemplified memory array, the nonideal state for a word simply consumes more power. Such an error is not critical to the block functionality; it only leads to increased power consumption.

Introducing asymmetries in SRAM cells has become a popular approach for increased low-voltage performance. Along with improving retention power for memory macros consisting of cells with asymmetric leakage, a dynamic word-flipping scheme, such as the one presented here, may also be used to increase noise margins and reduce switching energy depending on the underlying, asymmetric, memory cell.

V. CONCLUSION

The concept of benefiting from state dependencies in asymmetric SRAM cells through conditional word-flipping is illustrated through a manufactured a 1-kb SRAM macro of near-threshold single-ended read and single-ended write cells in 130-nm bulk CMOS. The asymmetric SRAM cell of choice uses an LVT transistor to pull down the read-BL and an LVT transistor as a pass gate to the write-BL, trading off retention power for a faster BL_R evaluation time and a higher write margin. The resulting memory cell had an asymmetric retention power, which was exploited using the word-flipping scheme.

Word-flipping schemes can be applied to any matrices of asymmetric memory cells to exploit their state dependencies. In the exemplified case in this brief, the scheme enables the use of six-transistor memory cells for low-voltage operation by compensating for their

state-dependent leakage quotients. The word-flipping system was in simulations on the 1-kb, 350-mV, SRAM macro, found to improve the retention power of the macro by 14.69% for uniformly distributed data. The scheme payoff was measured to be 11.93%, on average, over ten chip samples for the same random data set.

ACKNOWLEDGMENT

The authors would like to thank the Norwegian PhD Network on Nanotechnology for Microsystems for funding the silicon prototype.

REFERENCES

- [1] M. R. Stan and W. P. Burleson, "Bus-invert coding for low-power I/O," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 3, no. 1, pp. 49–58, Mar. 1995.
- [2] B. Mohammadi, O. Andersson, J. Nguyen, L. Ciampolini, A. Cathelin, and J. N. Rodrigues, "A 128 kb single-bitline 8.4 fI/bit 90MHz at 0.3 V 7T sense-amplifierless SRAM in 28 nm FD-SOI," in *Proc. 42nd Eur. Solid-State Circuits Conf.*, Sep. 2016, pp. 429–432.
- [3] S. Nalam and B. H. Calhoun, "5T SRAM with asymmetric sizing for improved read stability," *IEEE J. Solid-State Circuits*, vol. 46, no. 10, pp. 2431–2442, Oct. 2011.
- [4] V. Gherman and M. Cartron, "Soft-error protection of TCAMs based on ECCs and asymmetric SRAM cells," *Electron. Lett.*, vol. 50, no. 24, pp. 1823–1824, Nov. 2014.
- [5] I. Alouani, W. M. Elsharkasy, A. M. Eltawil, F. J. Kurdahi, and S. Niar, "AS8-static random access memory (SRAM): Asymmetric SRAM architecture for soft error hardening enhancement," *IET Circuits, Devices Syst.*, vol. 11, no. 1, pp. 89–94, Jan. 2017.
- [6] N. Azizi, F. N. Najm, and A. Moshovos, "Low-leakage asymmetric-cell SRAM," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 11, no. 4, pp. 701–715, Aug. 2003.
- [7] Y.-J. Chang, F. Lai, and C.-L. Yang, "Zero-aware asymmetric SRAM cell for reducing cache power in writing zero," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 12, no. 8, pp. 827–836, Aug. 2004.
- [8] H. Fujiwara *et al.*, "Novel video memory reduces 45% of bitline power using majority logic and data-bit reordering," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 6, pp. 620–627, Jun. 2008.

Comparative Analysis of Flip-flop Architectures for Subthreshold Applications in 28nm FDSOI

The conference version of article A.

E. Lâte, A. A. Vatanjou, T. Ytterdal, and S. Aunet

Published in *Nordic Circuits and Systems Conference (NORCAS), IEEE, 2015*, pp. 1-4¹

DOI: <https://doi.org/10.1109/NORCHIP.2015.7364372>

I.1 Abstract

Nine D flip-flop architectures were implemented in 28nm FDSOI at a target, sub-threshold, supply voltage, of 200mV. The goal was to identify promising D flip-flops for ultra low power applications. The pass gate flip-flop was implemented using 49% of the S2CFF's area and was functional at the lowest operating voltage of 65mV in the typical process corner. At the targeted supply voltage of 200 mV the race-free DFF gives the best functional yield of 99.8%. The flip-flops having the shortest D-Q delays were the PowerPC 603 and the transmission gate D flip-flop. These also had the lowest power delay products of 52.08aJ and 61.09aJ, respectively.

I.2 Relevance to the Thesis

D Flip-Flops are edge-triggered memory elements that are widely adopted in memory systems and commonly used as gates in pipeline stages of sequential computational circuits such as data processors. This article presents the results from one of the very first circuit design projects that were performed as a PhD candidate at NTNU. The article analyzes various DFF architectures and identifies promising candidates for low voltage operation. The results from the analysis were intended to serve as a solid fundament for the circuit design projects that were to follow. The publication is also directly relevant to the thesis since it answers research objective O1, mentioned in chapter 3.

I.3 Declaration of Authorship

The article was presented at the Nordic Circuits and Systems Conference (NOR-CAS) in 2015. The research to the conference paper was conducted in the first 5 months of the PhD work. Ali Asghar Vatanjou was at the time familiar with low voltage design and contributed with valuable implementation techniques. I performed the literature study, implemented the circuits, and wrote the paper under the supervision of Trond Ytterdal and Snorre Aunet.

I.4 Manuscript

Comparative Analysis of Flip-Flop Architectures for Subthreshold Applications in 28nm FDSOI

Even Låte

Department of Electronics and Telecommunications
Norwegian University of Science and Technology
O.S. Bragstads plass 2a, 7034 Trondheim
Email: even.late@iet.ntnu.no

Ali Asghar Vatanjou, Trond Ytterdal, Snorre Aunet
Department of Electronics and Telecommunications
Norwegian University of Science and Technology
O.S. Bragstads plass 2a, 7034 Trondheim

Abstract—Nine D flip-flop architectures were implemented in 28nm FDSOI at a target, subthreshold, supply voltage of 200mV. The goal was to identify promising D flip-flops for ultra low power applications. The pass gate flip-flop was implemented using 49% of the S²CFF's area and was functional at the lowest operating voltage of 65mV in the typical process corner. At the targeted supply voltage of 200mV the racefree DFF gives the best functional yield of 99.8%. The flip-flops having the shortest D-Q delays were the PowerPC 603 and the transmission gate D flip-flop. These also had the lowest power delay products of 52.08aJ and 61.09aJ respectively.

I. INTRODUCTION

Ever since the Centre Electronique Horloger(CEH) in Switzerland succeeded with low power integrated circuits [1], scientists have strived to create new and improved methods for designing low power systems. Subthreshold CMOS operation is such an approach and is applied both in the analog and in the digital circuit design domain. By reducing the supply voltage below the threshold voltage of the transistors, the power consumption is reduced by several orders of magnitude. The principle of operating transistors in weak inversion dates back to 1967 [2]. Despite its early introduction, only few niche-products have been established in the consumer-electronics market until now. The sensor-network trend known as the Internet of Things(IoT) has re-opened the interest for the ultra low power(ULP) benefits of operating the transistor in the subthreshold region. ULP Sensor nodes driven by coin-cell batteries will last for a long time, or perpetually, if energy harvesting is used in combination with subthreshold operation.

D Flip-Flops(DFFs) are widely used in digital applications such as parallel storage and to serially manipulate internal circuit nodes in SoCs. These edge triggered devices account for large fractions of circuit area and power consumption and are also amongst the most difficult digital building blocks to push to lower supply voltage levels without experiencing problems with data corruption. Either the output(Q) is subject to voltage drop so that it fails to take its correct logical value, or the internal delay of the master latch is too high to allow the slave latch to latch on to the data(D) signal.

The objective of this work is to explore multiple flip-flop architectures and identify promising candidates among these for subthreshold applications in 28nm Fully Depleted Silicon On Insulator(FDSOI). Section II introduces details on a schematic level for the chosen set of DFF architectures. The simulation-setup and performance metrics for evaluating these are explained in section III, followed by results and an evaluation of the schematics in sections IV & V respectively.

II. FLIP-FLOP ARCHITECTURES

Nine static D flip-flop architectures were picked out for a comparative study of their subthreshold abilities. The schematics for each of them are illustrated in figure 1 and are further explained in the enumerations below:

- 1) The static single-phase contention-free flip-flop(S²CFF) [3] is shown in cell nr 1 in figure 1. It is based on a dynamic true single-phase flipflop and is made static by including a slave latch and some additional transistors. The S²CFF was designed in 2014 for near threshold operation on a 45nm SOI technology.
- 2) The second structure is the more well known Clocked CMOS flip-flop (C²MOS) [4]. It consists of clocked 4T and basic 2T inverters and relies on two clock phases.
- 3) The conventional nand-based design in cell 3 was used for its simplistic gate level structure together with schmitt-trigger gates in [5] to push the lower limits of the supply voltage. The original design is made into a positive edge triggered DFF by removing one inverter on the CK input.
- 4) A flip-flop design consisting of transmission gates and inverters such as the one shown in cell number 4, is referred to as the Transmission Gate DFF. It is used extensively in sequential systems for its low power consumption and low transistor count. Note that two clock phases are used.
- 5) A static master slave flip-flop can be constructed using two multiplexers(MUXes) configured as latches. In this paper two different MUX topologies are investigated for such a configuration. Cell number 5 shows such a DFF with MUXes built up by conventional logic gates.
- 6) Cell number 6 shows the second MUX-based flip-flop where the MUXes are built in a xnor-like fashion with logic branches
- 7) The PowerPC 603 flip-flop in cell number 7 was used in the PowerPC 603 microprocessor[6]. The Powerpc flip-flop has been modified slightly since then, the two transistors controlling the scan mode of the DFF have been removed. This design has in previous comparative studies been found to have great subthreshold abilities[7][8][9].
- 8) Cell number 8 contains another nand and inverter based flip-flop, this design differs from the other nand based flip-flop by the fact that it is race free and that it consumes less area. It was introduced in [10].
- 9) Transmission gates are normally used to achieve symmetrical transmission of logic high and logic low signals. The pass gate design in cell number 9 exploits the area and single clock-phase benefits of using single transistors as pass gates instead of the 2T transmission gates [11], the trade-off is loss in signal integrity. In 2015, this design was simulated in 28nm FDSOI [12].

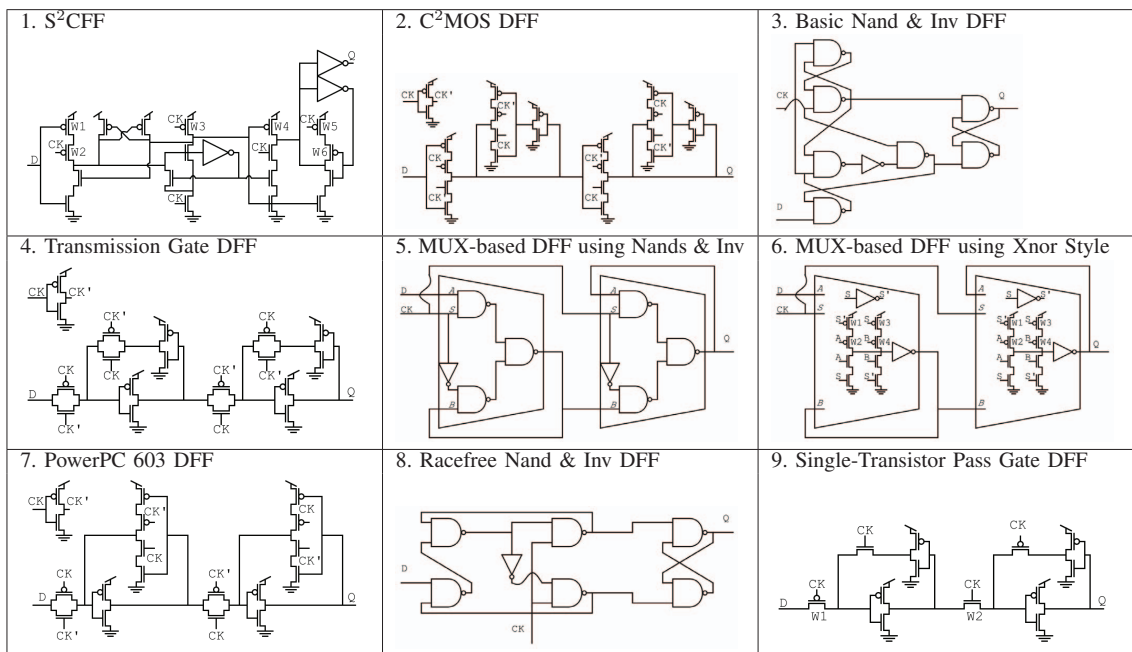


Fig. 1: Schematic overview of the selected flip-flop architectures

The DFFs were designed for a supply voltage of 200mV using LVT transistors. The transistors were sized to have equal drive strength for pull up and pull down networks. Flip-flops consisting of logic building blocks were sized from the bottom and up. The transistor lengths (L) of all flip-flops were kept constant to minimise mismatch and to simplify the architectural comparisons. All flip-flops were designed with lengths equal to the process minimum at 30nm.

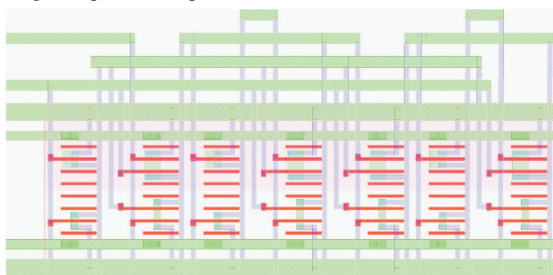


Fig. 2: Layout of the C^2 MOS DFF, width=5.613 μ m, height=2.726 μ m

For the basic 2T inverter, the width of the PMOS was found such that the voltage transfer characteristic had an output equal to $V_{DD}/2$ for an input equal to $V_{DD}/2$. For gates with multiple inputs, multiple transfer curves were achieved by holding non switching inputs in the on-state. The average of the multiple transfer curves was used for $V_{DD}/2$. This approach is optimal at uniform distributions of data input patterns. The transmission gate was sized to have equal drain current for both NMOS and PMOS in the on state. Transistor dimensions that are used in the various DFFs are listed in table I.

Figure 2 shows the layout of the C^2 MOS DFF. A layout style where metal over gate-poly is avoided, is used in all DFF

implementations to make the comparison fair. This is also done to better visualise the differences in interconnect complexity for the different flip-flops. LVT transistors in 28nm FDSOI are implemented in layout with flip well. Transistor bulk nodes are implemented using a hybrid layer to create holes in the insulator for the body contacts. The possibility of body biasing is not investigated in this paper, so both the PMOS, and the NMOS have grounded body contacts even though the layout is made in such a way that individual p- and n- network body biasing is possible.

III. SIMULATION SETUP

The flip flop architectures presented above were analysed using similar comparative approaches as those proposed in [13]. Each flip-flop was simulated using double inverters to drive the inputs and a fan-out of 4 inverters as output loads. All circuit elements that are used externally of the flip-flops have their own power grids to allow measurements of the exact current consumption of the DFFs. The dual clock-phase flip-flops; C^2 MOS, transmission gate and PowerPC 603 DFF were all implemented with local clock inversion. The local clock inverters were included in the power calculations of flip-flops using two clock phases. The inverted clock signal could also have been generated globally, but the additional power consumption from routing two clock signals would be difficult to model accurately.

The waveforms used in transient simulations were selected to allow settling of slow signals as a result of a low operating voltage. A signal pulse period of 4 μ s was used as the D input, while a signal pulse period of 2 μ s was used as the clock signal in 5 μ s transient simulations. To measure the minimal setup time of each flip-flop, the D signal was moved iteratively towards the CK edge until the DFF failed to propagate the change in the D input to the Q output on the same clock edge.

DFF	Transistor widths (nm), 80nm if nothing is specified	Transistor Count	Area of Layout (μ^2)
1. S ² CFF	w1=172, w2=172, w3=122, w4=110, w5=172, w6=172	24	24.11
2. C ² MOS DFF	wp clocked inv=172, wp inv=172	22	15.30
3. Basic Nand & Inv DFF	wp nand gate=162, wp inv=172	30	17.95
4. Transmission Gate DFF	wp transmission=200, wp inv=172	18	18.90
5. MUX-based DFF using Nands & Inv	wp nand gate=162, wp inv=172	28	17.65
6. MUX-based DFF using Xnor Style	w1=170, w2=170, w3=170, w4=170	24	19.67
7. PowerPC 603 DFF	wp transmission=200, wp clocked inv=172, wp inv=172	18	16.34
8. Racefree Nand & Inv DFF	wp nand gate=162, wp inv=172	26	15.81
9. Single-Transistor Pass Gate DFF	w1=200, w2=200	12	11.82

TABLE I: Transistor dimensions, transistor counts and layout area for the 9 DFF schematics in figure 1

The minimal setup time is equal to the setup time where the D transition is first discovered on the following clock edge. Once the minimal setup time was identified the setup time was swept from this point and upwards while observing the D-Q delay. The point where the D-Q delay is minimal and optimal is always located a small period of time after the minimal setup time due to the increasing CK-Q delay for lower setup times. This procedure was done for both rising and falling Q transitions and the worst case minimal and optimal setup times were chosen to account for asymmetry in the flip-flop schematics and to avoid functional errors deriving from optimistic clocking.

The lowest functional supply voltage was found for all flip-flops by introducing a minimum value for logic high equal to 75% of VDD. A maximum value for logic low was also set to be 25% of VDD. The supply voltage was then reduced to the point where the logic level limits were breached by Q.

In a similar fashion to the minimum supply voltage testing-methodology, functional yield was also simulated for the DFF architectures using 1000 Monte Carlo(MC) simulation-points. A lower logic high limit of 75% and an upper logic low limit of 25% was defined in concert with a pre-known Q-waveform. A deviation from the multiple pre-defined data-points counts as a failure caused by statistical variations. The MC analysis was performed on the extracted netlists.

IV. RESULTS

Table I lists the settled transistor dimensions, the transistor has minimum dimensions ie. a length equal to 30nm and a width equal to 80nm. It also lists the differences in transistor count and area consumption for the 9 flip-flops. The areas depend directly on the transistor counts and the regularity of the architectures, flip-flops with complex structures require extra area for routing outside of the cell.

The bar plot in figure 3 depicts the minimum operating

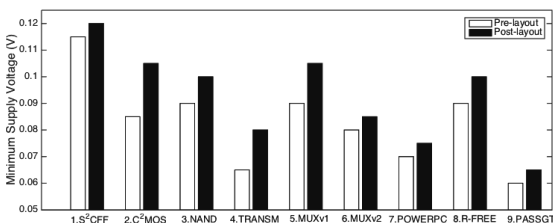


Fig. 3: Minimal functional supply voltage for the DFF architectures voltage where the circuit is still functional.

Figure 4 shows the range of valid setup times for both raw schematic and netlists generated from extracted cell layouts. Each range is marked with an X indicating the optimal setup time. Using optimal setup times results in the minimum D-Q delays illustrated in figure 5.

At optimal setup times and with supply voltages equal to

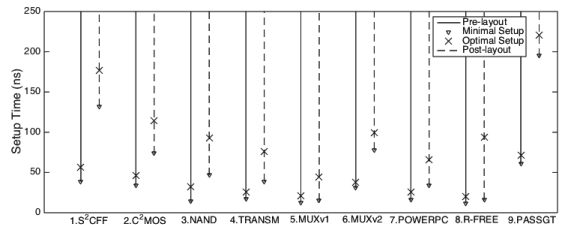


Fig. 4: Feasible ranges and optimal setup times for the 9 flip-flops

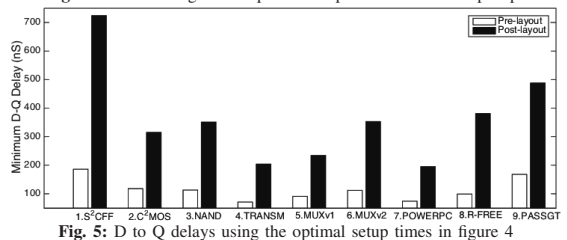


Fig. 5: D to Q delays using the optimal setup times in figure 4

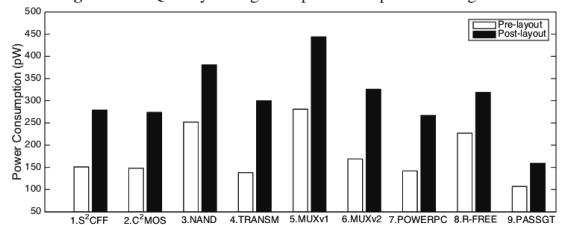


Fig. 6: Power consumptions at VDD=200mV using optimal setup times

200mV the average current drawn gives the power consumption of the DFFs, these are plotted in figure 6.

The two factors D-Q delays and power consumptions gives

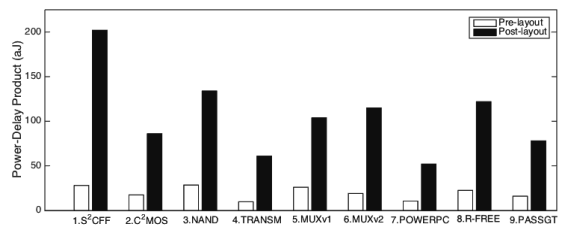


Fig. 7: Power delay products(PDPs) for each flip-flop based on figure 6 and 5 the power delay products in figure 7 when multiplied together.

High functional yields are found for most architectures except from number 5 in table II.

V. DISCUSSION

The plot of minimum functional supply voltage in figure 3 shows a solid trend of a higher minimum VDD for the

DFF	Functional Yield(%)
1. S ² CFF	98.3
2. C ² MOS DFF	97.4
3. Basic Nand & Inv DFF	98.9
4. Transmission Gate DFF	98.3
5. MUX-based DFF using Nands & Inv	65.6
6. MUX-based DFF using Xnor Style	99.3
7. PowerPC 603 DFF	96.0
8. Racefree Nand & Inv DFF	99.8
9. Single-Transistor Pass Gate DFF	92.7

TABLE II: Functional Yield, found from 1000 MC simulation points

netlist based on extracted layout. This seems to be out of the ordinary at first glance, taking only the parasitic capacitors into account. The trend does however make more sense when the resistivity in the routing is incorporated in the calculations. Voltage drops across resistive interconnects can explain the distorted output signals.

Generally, the performance variations for pre- and post-layout simulations appear rather large. This is as expected for subthreshold applications as changes in nodal capacitance are more noticeable with low drive currents. The differences between pre- and post-layout simulations exist for all flip-flops, but to various degrees. Since the layout is made using the same out-of-the-cell routing-style for all the DFFs, the parasitic effects should be fair for all, thus making the simulation results depend largely on the schematical advantages of each topology, their drive characteristics, their routing complexity and their transistor count.

All flip-flops are designed for a supply voltage of 200mV. Despite this fact, the lower the minimum operating voltage is, the greater the flexibility at a system level will be. A low limit for supply voltage reduction means having the possibility to perform voltage scaling over a larger voltage range, which again will increase the chances of finding a better global energy operating point, or lasting longer before a brown-out situation. Figure 3 suggests that the pass gate design is best at this aspect. The pass gate DFF also has a superior area compared to the other candidates as can be seen in table I. The pass gate flip-flop is, however, not at the top of the list when it comes to the power-delay product. A reason for this is that the use of single transistor pass gates makes the flip-flop highly unsymmetrical when it comes to propagation delays and setup times for both logic high and logic low signals. D-Q delay is in this paper found as the worst case of a falling and rising Q delay, so even though the flip-flop conducts a logic high using a D-Q delay of 127ns, the flip-flop is evaluated using its worst case D-Q delay of 488ns for a logic low propagation, making the worst case PDP about 4 times larger. The flip-flop designs that have the better power delay products are the PowerPC 603 and the Transmission gate DFFs, these are also the fastest flip-flops according to figure 5 with symmetrical logic high and logic low D-Q delays. Their power delay products are 52.08aJ and 61.09aJ respectively.

According to table II, the MUX-based flip-flop constructed using logic gates is by far the most prone architecture to statistical variations. The 8 other flip-flops prove to be quite robust in this aspect of the analysis. The DFF that consists of pass-gates is the weakest among the top 8 due to its asymmetric nature. The racefree flip flop is the most robust of them all, its racefree and simple nature could support this finding.

The simulation results found here, corresponds quite well to the earlier comparative studies in [7], [8] and [9] where the PowerPC 603 DFF sticks out as the best architecture

with respect to PDP. The results presented here supports these findings, but also broadens the DFF scope by shedding light on the pass-gate flip-flop which performs comparable to the PowerPC 603 design. The PDP of the PowerPC design in [8] is equal to 1fJ at 200mV post-layout, the lower PDP found here in 28nm FDSOI may be explained by scaling advantages and the fact that the FDSOI process is more energy friendly than common bulk technologies. All flip-flop architectures discussed in this paper are yet to be implemented on silicon and compared with measurements. Such investigations will reveal the true yield aspect of the D flip-flop analysis and the impact of parasitic effects that are neglected by the QRC resistance and capacitance extraction tool used in this work.

VI. CONCLUSION

The single-transistor pass gate flip-flop consumes less area than all other designs. It takes up silicon space equal to 49% of the S²CFF DFF area with the comparable layout approach described. Simulations also point out that the pass gate design can be operated at lower voltages than the other DFFs. It is however subject to a relatively low functional yield. The PowerPC 603 DFF is with a PDP of 52.08aJ, once again, found to be the flip-flop that has the lowest power-delay product followed by the transmission gate and the pass gate DFFs.

REFERENCES

- [1] C. Piguat, "The First Quartz Electronic Watch," in *Integrated Circuit Design. Power and Timing Modeling, Optimization and Simulation* (B. Hochet, A. J. Acosta, and M. J. Bellido, eds.), no. 2451 in Lecture Notes in Computer Science, pp. 1–15, Springer Berlin Heidelberg, 2002.
- [2] Y. Tsvividis, "Eric Vittoz and the Strong Impact of Weak Inversion Circuits," *IEEE Solid-State Circuits Society Newsletter*, vol. 13, no. 3, pp. 56–58, 2008.
- [3] Y. Kim, W. Jung, I. Lee, Q. Dong, M. Henry, D. Sylvester, and D. Blaauw, "27.8 A static contention-free single-phase-clocked 24t flip-flop in 45nm for low-power applications," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pp. 466–467, Feb. 2014.
- [4] Y. Suzuki, M. Hirasawa, and K. Odagawa, "Clocked CMOS calculator circuitry," in *Solid-State Circuits Conference. Digest of Technical Papers. 1973 IEEE International*, vol. XVI, pp. 58–59, Feb. 1973.
- [5] N. Lotze and Y. Manoli, "A 62 mV 0.13 m CMOS Standard-Cell-Based Design Technique Using Schmitt-Trigger Logic," *IEEE Journal of Solid-State Circuits*, vol. 47, pp. 47–60, Jan. 2012.
- [6] G. Gerosa, S. Gary, C. Dietz, D. Pham, K. Hoover, J. Alvarez, H. Sanchez, P. Ippolito, T. Ngo, S. Litch, J. Eno, J. Golab, N. Vanderschaaf, and J. Kahle, "A 2.2 W, 80 MHz superscalar RISC microprocessor," *IEEE Journal of Solid-State Circuits*, vol. 29, pp. 1440–1454, Dec. 1994.
- [7] H. Alstad and S. Aunet, "Seven subthreshold flip-flop cells," in *Norchip*, 2007, pp. 1–4, Nov. 2007.
- [8] M. Voernes, T. Ytterdal, and S. Aunet, "Performance comparison of 5 subthreshold CMOS flip-flops under process-, voltage-, and temperature variations, based on netlists from layout," in *NORCHIP, 2014*, pp. 1–6, Oct. 2014.
- [9] H. Alstad and S. Aunet, "Three Subthreshold Flip-Flop Cells Characterized in 90 nm and 65 nm CMOS Technology," in *11th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems, 2008. DDECS 2008*, pp. 1–4, Apr. 2008.
- [10] S. Okugawa and N. Inoue, "Systematic design of D flip-flops using two state variables," *Electronics Letters*, vol. 23, pp. 909–910, Aug. 1987.
- [11] H. Harris, Weste Neil, David M., *Cmos Vlsi Design: A Circuits And Systems Perspective*. Pearson Education India., 3d ed., 2006.
- [12] S. Farideh, "Design of Reliable and Efficient Flip-Flops for Subthreshold Operation using Multi-Threshold MOSFETs and Transistor Sizing Technique," 2015.
- [13] V. Stojanovic and V. Oklobdzija, "Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 536–548, Apr. 1999.





Ultra-low Voltage Adders in 28 nm FDSOI Exploring Poly-biasing for Device Sizing

The conference version of article C.

A. A. Vatanjou, E. Lâte, T. Ytterdal, and S. Aunet

Published in *Nordic Circuits and Systems Conference (NORCAS), 2016 IEEE*, pp. 1-4¹

DOI: <https://doi.org/10.1109/NORCHIP.2016.7792895>

II.1 Abstract

Balancing the PMOS/NMOS strength ratio is a key issue to maximize the noise margin, and hence, the functional yield of CMOS logic gates in the subthreshold region. In this work, the PMOS/NMOS strength ratio was balanced using a poly-biasing technique in conjunction with back-gate biasing provided in a 28 nm fully depleted silicon on insulator (FDSOI) technology. A 32-bit adder based on minority-3 (min-3) gates and a 16-bit adder based on Boolean gates have been implemented. Chip measurement results show highly energy-efficient adders, so that the 32-bit and 16-bit adders achieved minimum energy point (MEP) of 21.5 fJ at 300 mV and 12.62 fJ at 250 mV, respectively. In comparison to adders reported in other works in the same technology, the energy per 1-bit addition of the 32-bit adder is improved by 35%, and for the 16-bit adder, this improvement in energy consumption is 23%. The designed adders were functional down to a supply voltage of 110 mV. Additionally, the minimum V_{dd} of the 32-bit adder decreased to 79 mV by applying a reverse back-bias voltage to the PMOS devices.

II.2 Relevance to the Thesis

Memory systems commonly contain logic blocks to interact with the core memory cells. The paper is relevant to this research on low voltage memory systems by the fact that it contains a method to create logic libraries that are robust at low supply voltages. Silicon measurements verify the method.

II.3 Declaration of Authorship

The article was presented at the Nordic Circuits and Systems Conference (NOR-CAS) in 2016. The paper contains measurements of the silicon die that Ali Asghar Vatanjou and I designed in 2015. Ali performed the literature study, implemented the adders, and wrote the paper together with me under the supervision of Trond Ytterdal and Snorre Aunet.

II.4 Manuscript

Ultra-Low Voltage Adders in 28 nm FDSOI Exploring Poly-Biasing for Device Sizing

Ali Asghar Vatanjou, Even Låte, Trond Ytterdal and Snorre Aunet

Department of Electronics and Telecommunications

Norwegian University of Science and Technology, O. S. Bragstads Plass 2 A, 7491 Trondheim, Norway

Email: ali.vatanjou@iet.ntnu.no

Abstract—Balancing the PMOS/NMOS strength ratio is a key issue to maximize the noise margin, and hence, functional yield of CMOS logic gates in the subthreshold region. In this work, the PMOS/NMOS strength ratio was balanced using a poly-biasing technique in conjunction with back-gate biasing provided in a 28 nm fully depleted silicon on insulator (FDSOI) technology. A 32-bit adder based on minority-3 (min-3) gates and a 16-bit adder based on Boolean gates have been implemented. Chip measurement results show highly energy efficient adders, so that the 32-bit and 16-bit adders achieved minimum energy point (MEP) of 21.5 fJ at 300 mV and 12.62 fJ at 250 mV, respectively. In comparison to adders reported in other works in the same technology, the energy per 1-bit addition of the 32-bit adder is improved by 35% and for the 16-bit adder this improvement in energy consumption is 23%. The designed adders were functional down to a supply voltage of 110 mV. Additionally, the minimum V_{dd} of the 32-bit adder decreased to 79 mV by applying a reverse back bias voltage to the PMOS devices.

I. INTRODUCTION

Reducing power consumption to ultimately break heat walls and to reduce energy consumption to prolong battery lifetime are key issues in modern VLSI technology. The growing number of battery driven wireless devices as well as the increasing energy demand of the applications to which these are intended, creates a need for innovation in ultra-low power (ULP) circuitry. The already established subthreshold design principle has been known to circuit designers since the 1960s, however the increasing demand mentioned above and new features of modern process technologies brings relevance of subthreshold design back.

A commercially available 28 nm FDSOI process tackles some of the key challenges faced by the scaling of conventional bulk silicon devices, and so enables prolongation of Moore's law [1]. In this technology the junction capacitance is significantly reduced and the drain induced barrier lowering (DIBL) effect is also lowered due to the ultra thin body and buried oxide [1], this while increasing the efficiency of back-gate biasing. In addition, the random dopant fluctuation is suppressed because the threshold voltage adjustment does not require doping.

In this paper we report highly energy efficient logic gates by exploiting both poly-biasing [2] and back-gate biasing techniques. Both biasing techniques allow us to reduce the size of the pull up devices, thus reducing the leakage current and parasitic capacitance of the cells. The poly-biasing technique increases the effective channel gate length but not the active

area of the devices. Applying poly-biasing while connecting the PWELL/NWELL ties to ground improves the strength of the PMOS transistors. Moreover, the regular threshold voltage (RVT) devices have been chosen to reduce the leakage current further. Although higher threshold voltage leads to slower circuits, it reduces the energy consumption of circuits with more relaxed throughput requirements [3].

To validate the proposed sizing techniques, we have implemented a 32-bit adder based on min-3 gates [4] and a 16-bit adder based on Boolean gates. The ripple carry adders (RCA) has been chosen because the energy consumption of the serial adders may be lower than the parallel adders at the same speed in subthreshold [5]. Comparing with the adders reported in the same 28 nm FDSOI process technology [8], the minimum energy per 1-bit addition of the 32-bit and 16-bit adders are improved by 35% and 23%, respectively.

Schmitt-Trigger logic cells functioning down to 62 mV have been reported in [6]. However, these logic cells were not energy and area efficient. An ultra-low voltage subthreshold adder based on static CMOS logic cells has been reported in [7], which was capable of working correctly down to an 84 mV supply voltage. Apart from Schmitt-Trigger logic cells reported in [6], the minimum V_{dd} of 79 mV reported in this work is, to the knowledge of the authors, the lowest supply voltage for CMOS logic cells.

II. DEVICE SIZING FOR BALANCED PMOS/NMOS

Balancing the driving strength of the pull-up/pull-down networks (PUN/PDN) is a key issue to properly operate logic gates in the ultra-low voltage (ULV) domain. However as shown in Figure. 1, the on-current ratio of the PMOS with respect to NMOS is too small in the ULV domain. In Figure. 1, the NMOS device width is 200 nm, $L=30$ nm and the drain-source voltage is set equal to the supply voltage. For example, the PMOS gate width should be upsized by more than 4x to have a balanced PUN/PDN at a supply voltage of 150 mV. This causes asymmetry in layout of the cells. Additionally, both leakage current and parasitic capacitance, and consequently the power consumption of the circuit, increase with wider PMOS transistors. Therefore using techniques to improve the strength ratio between PMOS and NMOS, is an effective approach to reduce both power consumption and minimum operating supply voltage of some subthreshold circuits.

In this work we increased the gate length of the NMOS

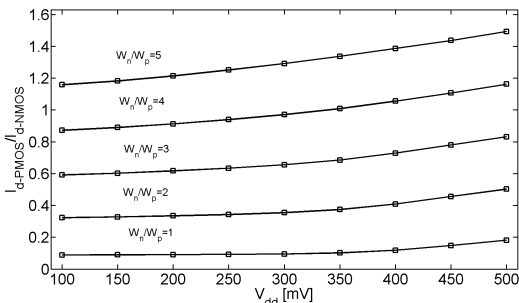


Fig. 1. PMOS/NMOS on-current ratio versus V_{dd} for $W_n = 200$ nm and $L_n = L_p = 30$ nm.

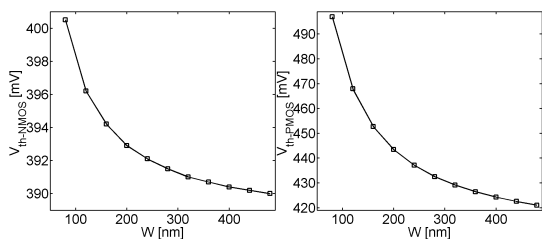


Fig. 2. Dependency of the threshold voltage on channel width (a) NMOS (b) PMOS.

transistors by 16 nm using poly-biasing to balance NMOS and PMOS driving strengths. We also connected both NWELL and PWELL ties to ground, thus the bulk-source junction of PMOS transistors were forward biased. In order to reduce the leakage current further, the PMOS gate length was also increased by 4 nm using poly-biasing, while maintaining the switching voltage at $V_{dd}/2$. As can be observed from Figure. 2, the threshold voltage of this 28 nm FDSOI technology is inversely proportional to the gate width of the devices. In Figure. 2, $L_{nmos} = L_{min} + 16$ nm, $L_{pmos} = L_{min} + 4$ nm and $V_{dd} = 200$ mV. The NMOS and PMOS transistor channel widths were chosen taking to account the channel width dependency of the threshold voltage. The minimum width of the NMOS devices was chosen to be 200 nm and the corresponding PMOS width that resulted in a switching voltage of $V_{dd}/2$ for an inverter was 300 nm. Although a W_n/W_p ratio of 1.5 is required for balancing PMOS and NMOS transistors, this is a reasonable ratio comparing with for example [9] where with $W_n/W_p = 5$ PMOS/NMOS strength ratio is balanced and with $W_n/W_p = 2$, the PMOS and NMOS transistors are still unbalanced. Figure. 3 shows the schematic and device sizes for the basic logic cells that were used to implement RCAs.

III. ADDER CIRCUIT DESIGN

Figure. 4 illustrates 1-bit FA cells of the two different logic styles. The FA based on min-3 gates [4] was used to implement a 32-bit RCA. A 16-bit RCA was also implemented using the FA based on Boolean logic gates.

The layout views of the FAs are shown in Figure. 5. We

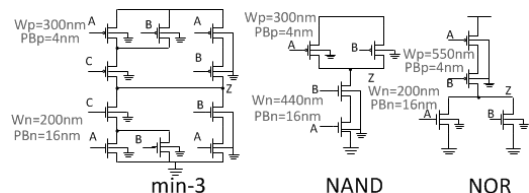


Fig. 3. Schematic view and corresponding device sizes of the basic logic gates used to implement FAs.

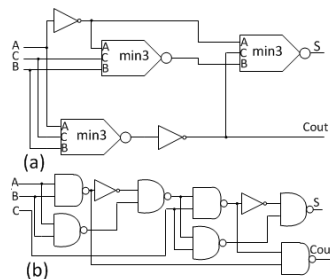


Fig. 4. Block-diagram of 1-bit FA (a) implemented with min-3 gates (b) implemented with boolean gates.

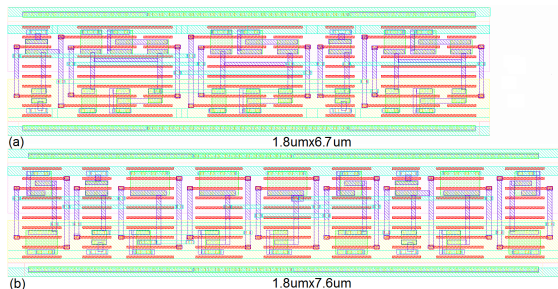


Fig. 5. Layout view of the 1-bit full adders (a) implemented with min-3 gates (b) implemented with Boolean gates.

focused on layout regularity to increase the matching properties of the devices [10]. There is no rounding and routing in the poly layer. All the poly polygons have single direction and 106 nm poly pitch was used to be able to apply poly-biasing on transistors gates. As a result, the mismatch between transistors gate lengths should decrease [10], as well as leakage currents and strength ratios between PMOS and NMOS transistors.

IV. MEASUREMENT RESULTS

Figure. 6 shows the PCB and QFN44 socket we used for measuring eight chips. Figure. 7 illustrates the test chip block diagram. A toggling input carry was applied to the adders while all the A inputs are shorted to V_{dd} and the B inputs are connected to ground. This exercises the critical path and causes the longest delay [5]. The RVT devices that were used in the core circuitry have threshold voltages of around 430 mV. The buffers were implemented with LVT (low threshold) transistors, so the delay through the buffers and the I/O pads

could be negligible compared to the delay of the adders. The absolute threshold voltage of the LVT transistors is around 390 mV. Figure. 8 (a) shows measured energy per operation of the 32-bit adder versus supply voltage at maximum operating speed. The delay of the 32-bit adder versus V_{dd} is also shown in Figure. 8 (b). The 32-bit RCA achieved a minimum energy point of 21.5 fJ at 300 mV and delay at this V_{dd} was 3.45 μ s. This means that minimum energy per 1-bit addition is 0.67 fJ for this adder. Measured energy and delay of the 16-bit adder versus supply voltage are depicted in Figure. 9 (a) and (b). The minimum energy per operation of the 16-bit adder occurs at 250 mV and is 12.62 fJ. Thus, minimum required energy per 1-bit addition is 0.79 fJ for this adder. The delay of the 16-bit adder is 5.8 μ s at minimum energy point (@250 mV).

We measured eight chips to explore the functionality of the chips for different supply voltages. Figure. 10 shows the minimum V_{dd} of the adders for eight measured chips with both PMOS and NMOS back-gates (V_{bbp} and V_{bbn}) shorted to ground. The worst samples of the eight, operated properly down to 125 mV for both adders. In the 32-bit adder case, two chips out of eight were functional down to 110 mV. For the 16-bit adder, three chips out of eight were functional down to 110 mV.

As shown in Figure. 7, the PWELL/NWELL contacts of the 32-bit adder are routed to the I/O pads, which enables applying different back bias voltages to this adder. We were able to

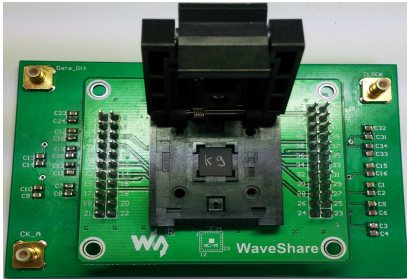


Fig. 6. The PCB and QFN44 socket for measuring eight chips.

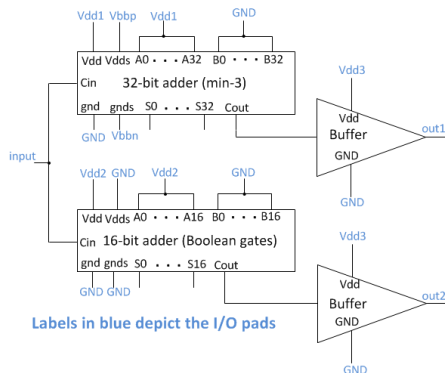


Fig. 7. Test chip block-diagram.

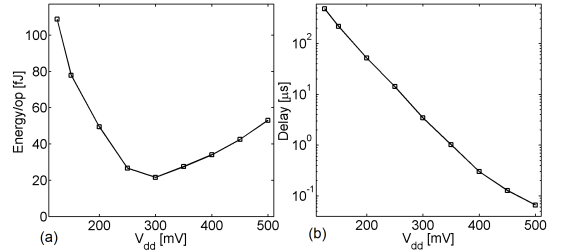


Fig. 8. Measured energy and delay of 32-bit adder versus V_{dd} .

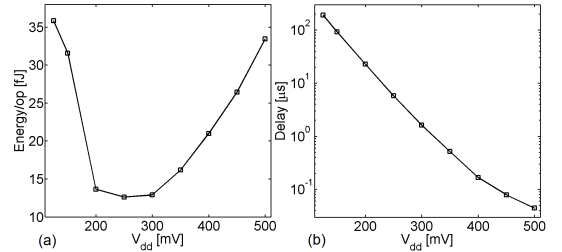


Fig. 9. Measured energy and delay of 16-bit adder versus V_{dd} .

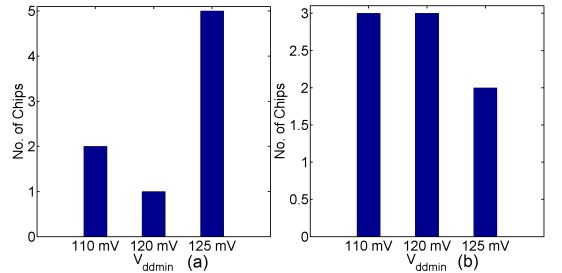


Fig. 10. Minimum V_{dd} of (a) 32-bit adders implemented with min-3 gates (b) 16-bit adders implemented with Boolean gates.

decrease the minimum V_{dd} of the 32-bit adder to 79 mV by applying a 430 mV positive V_{bbp} to the PMOS transistors, i.e. applying a reverse back bias voltage to the PMOS transistors. The oscilloscope plots of the input/output carries of the 32-bit adder at a supply voltage of 79 mV with $V_{bbp} = 430$ mV and $V_{bbn} = 0$ V are shown in Figure. 11. Figure. 12 depicts the input/output carries of the 16-bit adder at 110 mV.

V. DISCUSSION

TABLE I compares the measurement results of the designed adders with the post-layout simulation results of existing works in terms of energy, delay and area. The 32-bit RCA presented here achieves 0.67 fJ per 1-bit addition. This is the lowest energy per 1-bit addition among the works listed in TABLE I. In comparison to [8] which has the lowest reported energy per 1-bit addition (1.03 fJ) among the other works, the energy per 1-bit addition of the 32-bit and 16-bit adders are improved by 35% and 23%, respectively.

Comparing the 1-bit addition delay times at 300 mV, the adder

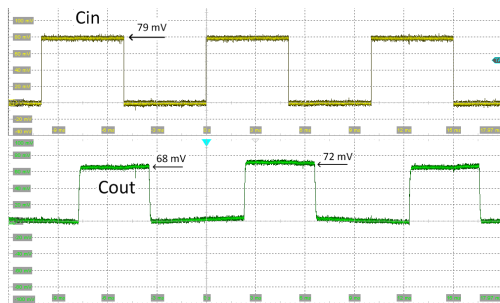


Fig. 11. Oscilloscope plots proving the 32-bit adder is functional at 79 mV with $V_{bbp} = 430$ mV and $V_{bbn} = 0$ V.

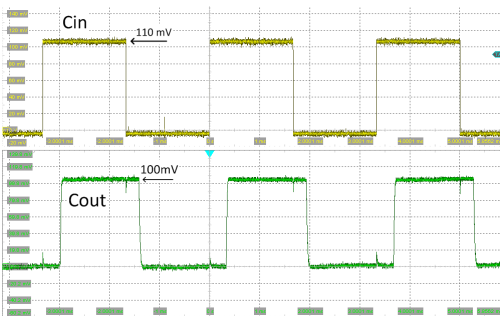


Fig. 12. Oscilloscope plots proving the 16-bit adder is functional at 110 mV.

in [8] has the shortest delay of 1.7 ns thanks to the LVT devices and forward body biasing. The 1-bit addition delay times of the 32-bit and 16-bit adders in [11] are 28 ns. This means that the adders in [11] are 3.8x and 3.6x faster than the presented 32-bit and 16-bit adders, respectively. The lower speed of the designed adders were expected considering RVT transistors and longer channel lengths. Nevertheless, since the adders have low energy and small area, they are suitable for applications with more relaxed throughput requirements.

The minimum supply voltage of the 32-bit adder decreased by applying a reverse back-bias voltage to the PMOS devices, and the same trend was observed for all the samples. Therefore, the dies might be towards the slow-NMOS fast-PMOS (SF) process corner.

Despite the fact that the FA based on min-3 gates has 34 transistors and the FA based on Boolean gates has 32 transistors, the footprint area of the FA based on min-3 gates ($12.06 \mu\text{m}^2$) is smaller than the area of FA based on Boolean gates ($13.68 \mu\text{m}^2$). The area of the 32-bit adder is 3% smaller than the 32-bit adder in [11], while the 16-bit adder has 9% more area in comparison with the 16-bit adder in [11].

VI. CONCLUSION

The development of ultra-low energy subthreshold adders in 28 nm FDSOI technology has been presented. Poly and back-gate biasing has been used to balance PMOS/NMOS strength ratio which in turn results in less parasitic capacitance and

TABLE I
COMPARISON WITH EXISTING WORKS IN 28 NM FDSOI TECHNOLOGY.

	This work 32-bit adder	This work 16-bit adder	[11] 32-bit adder	[11] 16-bit adder	[8] 9-bit adder
E_{min} [fJ]	21.5	12.62	207.4	74.9	9.27
Delay @ MEP [μs]	3.45	5.8	0.898	0.448	0.555
V_{dd} @ MEP [mV]	300	250	300	300	240
Devices	RVT	RVT	RVT/LVT	RVT/LVT	LVT
Area [μm^2]	406	230.3	418.56	209.28	N.A.
Results	Meas.	Meas.	Post-lay.	Post-lay.	Post-lay.

leakage current. To the knowledge of the authors, the measured 0.67 fJ energy per 1-bit addition of the 32-bit adder is the lowest measured energy reported so far. Both the 32-bit and the 16-bit adders were functional in the deep subthreshold region during measurements. The minimum functional supply voltage of the 32-bit adder was found to be 79 mV with a reverse back-gate bias for the PMOS transistors.

REFERENCES

- [1] T. Skotnicki et al., "Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia," in IEEE Transactions on Electron Devices, vol. 55, no. 1, pp. 96-130, Jan. 2008.
- [2] D. Jacquet et al., "A 3 GHz dual core processor ARM cortex TM-A9 in 28 nm UTBB FD-SOI CMOS with ultra-wide voltage range and energy efficiency optimization", IEEE J. Solid-State Circuits, vol. 49, no. 4, pp. 812-826, 2014.
- [3] D. Bol, R. Ambroise, D. Flandre and J. D. Legat, "Analysis and minimization of practical energy in 45nm subthreshold logic circuits," Computer Design, 2008. ICCD 2008. IEEE International Conference on, Lake Tahoe, CA, 2008, pp. 294-300.
- [4] S Aunet and Y. Berg, "Three Sub-fJ Power-Delay-Product Subthreshold CMOS Gates", Very Large Scale Integration, 2005 IFIP International Conference on, pp. 465-470.
- [5] Beiu, V.; Djupdal, A.; Aunet, S., Ultra Low-Power Neural Inspired Addition: When Serial Might Outperform Parallel Architectures, Lecture Notes in Computer Science, Computational Intelligence and Bioinspired Systems, Springer, vol. 3512, pp. 486, 493, 2005.
- [6] N. Lotze and Y. Manoli, "A 62mV 0.13m CMOS standard-cell-based design technique using schmitt-trigger logic," 2011 IEEE International Solid-State Circuits Conference, San Francisco, CA, 2011, pp. 340-342.
- [7] A. A. Vatanjou, T. Ytterdal and S. Aunet, "Energy efficient sub/near-threshold ripple-carry adder in standard 65 nm CMOS," Quality Electronic Design (ASQED), 2015 6th Asia Symposium on, Kula Lumpur, 2015, pp. 7-12.
- [8] H. Attarzadeh, S. Aunet and T. Ytterdal, "An ultra-low-power/high-speed 9-bit adder design: Analysis and comparison Vs. technology from 130nm-LP to UTBB FD-SOI-28nm," Nordic Circuits and Systems Conference (NORCAS): NORCHIP & International Symposium on System-on-Chip (SoC), 2015, Oslo, pp. 1-4.
- [9] M. Z. Li et al., "Energy Optimized Subthreshold VLSI Logic Family With Unbalanced Pull-Up/Down Network and Inverse Narrow-Width Techniques," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 12, pp. 3119-3123, Dec. 2015.
- [10] L. L. Lewyn, T. Ytterdal, C. Wulff and K. Martin, "Analog Circuit Design in Nanoscale CMOS Technologies," in Proceedings of the IEEE, vol. 97, no. 10, pp. 1687-1714, Oct. 2009.
- [11] R. Taco, I. Levi, M. Lanuzza and A. Fish, "Extended exploration of low granularity back biasing control in 28nm UTBB FD-SOI technology," 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, QC, 2016, pp. 41-44.

PART III

CONCLUSIONS & OUTLOOK

Concluding Remarks

"What I cannot create, I do not understand."

-Richard Feynman

In this work, the concept of voltage scaling of CMOS circuits has been explored in the implementation of low-power and energy-efficient edge-triggered memories, general-purpose logic libraries, and static random access memories.

- **RQ 1)** Is there a synchronous edge-triggered memory element that shows superior low voltage behavior? Answered by article A and complementary paper F present in the form of a comparative study. For the comparative study of published DFF architectures, it is found that:
 - At a supply voltage of 200 mV, the characteristics of the 9 different DFF architectures that were studied, vary significantly as shown in Figs 15-17 in manuscript A. If the supply voltage is reduced further down below 200 mV, the differences are magnified.
 - By uniformly weighing all metrics, the most promising DFF of the 9 studied architectures was found to be the low-transistor-count Pass-gate DFF as can be seen in Table 2 of manuscript A. However, yield-wise at ultra-low voltages, the structure suffers from low functional yield.
 - The PowerPC DFF is a better choice in terms of power-delay product and functional yield at ultra-low voltages.
- **RQ 2)** Many Low voltage SRAM cells have been reported. However, most of them deploy multiple extra transistors as write and read assists. Can a compact SRAM cell function in the near-threshold region? Answered by article B, where a 6T SRAM cell is proposed with a few re-arranged transistors. For the study of compact, low voltage SRAM cells, it is found that:
 - Most SRAM cells add transistors to the conventional 6T SRAM cell to support low voltage operation resulting in low bit/area ratios.

- The loadless 4T SRAM cell can be combined with a conventional 2T read buffer to create a compact 6T SRAM cell. The loadless core acts as a write assist, and the read buffer acts as a read assist in supporting operation with near-threshold low supply voltages.
 - The cell is combined into arrays of 128 x 64-bit words. 16 such arrays are combined into a 128 kb SRAM to demonstrate the system functionality of the cell and make it comparable to other published 128 kb macros.
 - Post layout simulation results, listed in Table 3 of manuscript B, show that the cell performs on par with reported measured results of other low voltage SRAM cells in the same CMOS technology in terms of min functional V_{DD} , max read frequency, leakage power per bit and read energy per operation per bit having a low transistor count for a low voltage memory cell.
- **RQ 3)** The drive strength ratio between uniformly sized NMOS and PMOS transistors differs significantly from 1 in subthreshold CMOS logic. How can back-gate biasing and the 28nm FDSOI poly-biasing option be used to balance this ratio towards 1? Answered by article C and complementary paper G, where such techniques are used to manufacture low voltage adders. To balance the drive-strength of deep subthreshold NMOS- and PMOS-networks in CMOS logic cells, it is found that:
 - By applying the 28nm FD-SOI specific poly biasing option, and a regular layout technique to the devices in minority-3 and boolean logic gates. Deep subthreshold logic libraries are created.
 - By combining the special building blocks into 16 bit and 32-bit adders deep subthreshold energy-efficiency is confirmed with measurements on physical prototypes.
 - The 32-bit minority-3 adder functions down to a supply voltage of 79 mV across chip samples using a reverse back-gate bias on the PMOS transistors as shown in Fig.16 of manuscript C.
- **RQ 4)** How can low voltage circuit blocks communicate with regular voltage modules in an energy-efficient manner in systems containing multiple voltage domains? Furthermore, what is the lower limit of the supply voltage of the low voltage circuit? Answered by article D where a low voltage level shifter is proposed. To interface deep-subthreshold circuit-blocks with modules operating in the regular-voltage domain, it is found that:

- The concept of a differential write operation of a regular 6T SRAM cell can be used as a write assist in level shifters to support deep sub-threshold to above threshold level conversion.
 - The write assist results in a faster write operation to the level shifter, which in turn gives a low energy per transition of 25.9 fJ per transition as shown in Table II of manuscript D.
 - A manufactured prototype in 130 nm bulk CMOS of the new level shifter architecture is measured to be functional at a mean lower supply voltage of 15 mV as shown in Fig. 8b of manuscript D. This is, to the author's knowledge and at the time of writing, the lowest reported V_{DDL} for a published level shifter.
- **RQ 5)** Low voltage SRAM cells are often asymmetric in nature. They typically have a resulting state-dependent behavior that is either desired or undesired. Can this be exploited dynamically during write operations? Answered by article E, where a macro level scheme is proposed. To exploit state dependencies in asymmetric memory cells, it is found that:
- In a macro of asymmetric memory cells, a binary majority decision circuit and an extra column of memory cells can be used to exploit the content dependent benefits of the asymmetric cells.
 - As a majority decision circuit in such a scheme, a capacitor array can be used to conditionally flip each written word to the macro of asymmetric memory cells.
 - For an asymmetric memory cell that has a content-dependent differential in retention current of 23x, the word-flip scheme can be used to reduce the expected leakage current of the memory cell macro by 15.70% given by the statistical model in EQ. 5 of manuscript E.

These conclusions are drawn from simulations in modern IC design software and measurements on physical silicon prototypes. Promising logic cells and 1-bit memory cells have been identified to be suitable for sensor node applications and low power cores in heterogeneous multi-core computing. Furthermore, the conclusions confirm the functionality of the introduced circuit architectures, new designs and concepts have been added to the toolbox of engineers and researchers out there that want to design low voltage circuits.

Overall, these findings confirm that voltage scaling is an attractive approach to achieving low power consumption and energy-efficiency in the design of modern CMOS circuits. The findings also confirm that low voltage circuits require additional architectural modifications to be viable.

Outlook

"If you get up in the morning and think the future is going to be better, it is a bright day. Otherwise, it's not."

-Elon Musk

In the future, our society will become dependent on interconnected, autonomous, and maintenance-free sensors. The sampled environmental data will be used to optimize food production, control traffic, report natural disasters before they occur, and save lives by providing preventative feedback to medical patients. Energy-efficient and low power circuits are essential to prolong battery lifetime in battery-driven sensors, or even allow perpetual operation with the assistance of energy harvesting.

At the time of writing, in 2020, 60% of the world's population are users of mobile phones, and 72% of these cellphones are smartphones, both these fractions have increased yearly, and are expected to converge towards 100% as the previous generation in the population fades out. With an average smartphone usage of 3 hours per day per person, there is a global demand for energy-efficient smartphone circuitry to reduce the frequency of battery re-charge.

The use of handheld devices has evolved over the past few decades towards a usage where data-traffic is at its center. Smartphone cameras and online applications gain popularity, which results in an exponential growth in global internet traffic. According to a forecast made by Ericsson, as 5G is made accessible to the public, each smartphone user will generate, on average, 24 GB of data monthly by 2025, putting a severe load on data centers to handle the traffic. With the recent pandemic of the COVID-19 virus, social distancing has drastically increased the usage of virtual meeting applications and on-demand entertainment that has come to stay. The need for low power and energy-efficient circuitry in data centers to deal with the dark silicon effect and cut the cost of electricity is correlated to the rapidly increasing internet traffic.

The engineering of CMOS circuit architectures that support low voltage operation has shown energy efficiency and low power consumption. Voltage scaling above

the threshold voltage of the transistor is already commonly adopted in commercial electronics. Many research problems remain to be solved before the adoption of subthreshold and near-threshold circuits can be widespread in the internet of things, data centers, and handheld devices. The key challenge that needs to be solved to enable commercialisation of low voltage CMOS is robustness or resilience against device variations in large scale integrated circuits. The sub 100 mV design space deserves attention once the problem of device variation is solved. Spin-transfer torque RAM (STT-RAM) might eventually replace the need for low voltage SRAMs, but this technology remains immature both in terms of endurance and expensive process steps.

Overall, we hope that the work presented in this thesis inspires further work within the field of low voltage memories and logic on silicon.

Bibliography

- [1] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2010.
- [2] K. Ashton *et al.*, "That "Internet of things" thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.
- [3] M. Alioto, *Enabling the Internet of Things: From Integrated Circuits to Integrated Systems*. Springer, 2017.
- [4] A. S. Andrae and T. Edler, "On global electricity usage of communication technology: trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.
- [5] A. Carroll, G. Heiser *et al.*, "An analysis of power consumption in a smart-phone." in *USENIX annual technical conference*, vol. 14. Boston, MA, 2010, pp. 21–21.
- [6] S. A. Carvalho, D. C. Cunha, and A. G. Silva-Filho, "Autonomous power management in mobile devices using dynamic frequency scaling and reinforcement learning for energy minimization," *Microprocessors and Microsystems*, vol. 64, pp. 205–220, 2019.
- [7] G. E. Moore *et al.*, "Cramming more components onto integrated circuits," 1965.
- [8] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted mosfet's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.
- [9] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *ISCA 2011*. IEEE, 2011, pp. 365–376.
- [10] N. Jones, "How to stop data centres from gobbling up the world's electricity," *Nature*, vol. 561, no. 7722, pp. 163–167, 2018.
- [11] J. M. Rabaey and M. Pedram, *Low power design methodologies*. Springer Science & Business Media, 2012, vol. 336.

- [12] 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.
- [13] E. Låte, T. Ytterdal, and S. Aunet, "A loadless 6t sram cell for sub- & near-threshold operation implemented in 28 nm fd-soi cmos technology," *Integration*, vol. 63, pp. 56–63, 2018.
- [14] A. A. Vatanjou, E. Låte, T. Ytterdal, and S. Aunet, "Ultra-low voltage and energy efficient adders in 28 nm fdsoi exploring poly-biasing for device sizing," *Microprocessors and Microsystems*, vol. 56, pp. 92–100, 2018.
- [15] E. Låte, T. Ytterdal, and S. Aunet, "An energy efficient level shifter capable of logic conversion from sub-15 mv to 1.2 v," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. X, pp. Y–Z, 2020.
- [16] E. Låte, T. Ytterdal, and S. Aunet, "Benefiting from state dependencies in asymmetric sram cells through conditional word-flipping," *An international journal*, vol. X, pp. Y–Z, Submitted.
- [17] E. Låte, A. A. Vatanjou, T. Ytterdal, and S. Aunet, "Comparative analysis of flip-flop architectures for subthreshold applications in 28nm fdsoi," in *Nordic Circuits and Systems Conference (NORCAS): NORCHIP & International Symposium on System-on-Chip (SoC), 2015*. IEEE, 2015, pp. 1–4.
- [18] A. A. Vatanjou, E. Late, T. Ytterdal, and S. Aunet, "Ultra-low voltage adders in 28 nm fdsoi exploring poly-biasing for device sizing," in *Nordic Circuits and Systems Conference (NORCAS), 2016 IEEE*. IEEE, 2016, pp. 1–4.
- [19] B. N. Taylor and A. Thompson, "Nist special publication 330: The international system of units (si)," Technical report, National Institute of Standards and Technology, 2008 – 2019, Tech. Rep., 1991.
- [20] D. B. Newell and E. Tiesinga, "The international system of units (si)," Tech. Rep., 2019.
- [21] F. Wanlass and C. Sah, "Nanowatt logic using field effect mos triodes," in *IEEE Solid State Circuits Conf.*, 1963, p. 32.
- [22] A. Rapp, L. Wennick, H. Borkan, and K. Keller, "Complementary-mos integrated binary counter," in *1967 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, vol. 10. IEEE, 1967, pp. 52–53.

- [23] E. A. Vittoz, "The electronic watch and low-power circuits," *IEEE Solid-State Circuits Society Newsletter*, vol. 13, no. 3, pp. 7–23, 2008.
- [24] F. Leuenberger and E. Vittoz, "Complementary-mos low-power low-voltage integrated binary counter," *Proceedings of the IEEE*, vol. 57, no. 9, pp. 1528–1532, 1969.
- [25] E. A. Vittoz, "Weak inversion for ultimate low-power logic," in *Low-Power Electronics Design*. CRC Press, 2004, pp. 270–287.
- [26] T. D. Burd and R. W. Brodersen, "Energy efficient cmos microprocessor design," in *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences*, vol. 1. IEEE, 1995, pp. 288–297.
- [27] J. Burr and A. Peterson, "Ultra low power cmos technology," 1991.
- [28] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical mos transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog integrated circuits and signal processing*, vol. 8, no. 1, pp. 83–114, 1995.
- [29] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 305–327, 2003.
- [30] A. Tajalli and Y. Leblebici, *Extreme low-power mixed signal IC design: subthreshold source-coupled circuits*. Springer Science & Business Media, 2010.
- [31] Z.-H. Liu, C. Hu, J.-H. Huang, T.-Y. Chan, M.-C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer mosfets," *IEEE Transactions on electron devices*, vol. 40, no. 1, pp. 86–95, 1993.
- [32] G. Chen, D. Sylvester, D. Blaauw, and T. Mudge, "Yield-driven near-threshold sram design," *IEEE transactions on very large scale integration (VLSI) systems*, vol. 18, no. 11, pp. 1590–1598, 2009.
- [33] 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.
- [34] A. Wang, B. H. Calhoun, and A. P. Chandrakasan, *Sub-threshold design for ultra low-power systems*. Springer, 2006, vol. 95.

- [35] M. J. Pelgrom, A. C. Duinmaijer, and A. P. Welbers, "Matching properties of mos transistors," *IEEE Journal of solid-state circuits*, vol. 24, no. 5, pp. 1433–1439, 1989.
- [36] P. A. Stolk, F. P. Widdershoven, and D. Klaassen, "Modeling statistical dopant fluctuations in mos transistors," *IEEE Transactions on Electron devices*, vol. 45, no. 9, pp. 1960–1971, 1998.
- [37] A. Stevenson, *Oxford dictionary of English*. Oxford University Press, USA, 2010.
- [38] P. E. Dodd, J. R. Schwank, M. R. Shaneyfelt, J. A. Felix, P. Paillet, V. Ferlet-Cavrois, J. Baggio, R. Reed, K. Warren, R. Weller *et al.*, "Impact of heavy ion energy and nuclear interactions on single-event upset and latchup in integrated circuits," *IEEE Transactions on Nuclear Science*, vol. 54, no. 6, pp. 2303–2311, 2007.
- [39] T. McConaghy, K. Breen, J. Dyck, and A. Gupta, *Variation-aware design of custom integrated circuits: a hands-on field guide*. Springer Science & Business Media, 2012.
- [40] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of mos sram cells," *IEEE JSSC*, vol. 22, no. 5, pp. 748–754, 1987.
- [41] J. Lohstroh, E. Seevinck, and J. De Groot, "Worst-case static noise margin criteria for logic circuits and their mathematical equivalence," *IEEE Journal of Solid-State Circuits*, vol. 18, no. 6, pp. 803–807, 1983.
- [42] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mv robust schmitt trigger based subthreshold sram," *IEEE JSSC*, vol. 42, no. 10, pp. 2303–2313, 2007.
- [43] E. Grossar, M. Stucchi, K. Maex, and W. Dehaene, "Read stability and write-ability analysis of sram cells for nanometer technologies," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 11, pp. 2577–2588, 2006.
- [44] J. S. Caravella, "A low voltage sram for embedded applications," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 3, pp. 428–432, 1997.
- [45] N. Verma and A. P. Chandrakasan, "A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy," *IEEE JSSC*, vol. 43, no. 1, pp. 141–149, 2008.
- [46] M. Khayatzadeh and Y. Lian, "Average-8t differential-sensing subthreshold sram with bit interleaving and 1k bits per bitline," *IEEE Transactions on*

- Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 5, pp. 971–982, 2013.
- [47] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, “A variation-tolerant sub-200 mv 6-t subthreshold sram,” *IEEE Journal of Solid-State Circuits*, vol. 43, no. 10, pp. 2338–2348, 2008.
- [48] T.-H. Kim, J. Liu, J. Keane, and C. H. Kim, “A high-density subthreshold sram with data-independent bitline leakage and virtual ground replica scheme,” in *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*. IEEE, 2007, pp. 330–606.
- [49] B. Mohammadi, O. Andersson, J. Nguyen, L. Ciampolini, A. Cathelin, and J. N. Rodrigues, “A 128 kb single-bitline 8.4 fj/bit 90mhz at 0.3 v 7t sense-amplifierless sram in 28 nm fd-soi,” in *ESSCIRC Conference 2016: 42nd*. IEEE, 2016, pp. 429–432.
- [50] I. J. Chang, J.-J. Kim, S. P. Park, and K. Roy, “A 32 kb 10t sub-threshold sram array with bit-interleaving and differential read scheme in 90 nm cmos,” *IEEE journal of Solid-State circuits*, vol. 44, no. 2, pp. 650–658, 2009.
- [51] A. Pavlov and M. Sachdev, *CMOS SRAM circuit design and parametric test in nano-scaled technologies: process-aware SRAM design and test*. Springer Science & Business Media, 2008, vol. 40.
- [52] H. Alstad and S. Aunet, “Seven subthreshold flip-flop cells,” in *Norchip, 2007*, Nov. 2007, pp. 1–4.
- [53] H. Alstad and S. Aunet, “Three Subthreshold Flip-Flop Cells Characterized in 90 nm and 65 nm CMOS Technology,” in *11th IEEE Workshop on Design and Diagnostics of Electronic Circuits and Systems, 2008. DDECS 2008*, Apr. 2008, pp. 1–4.
- [54] S. Farideh, “Design of Reliable and Efficient Flip-Flops for Subthreshold Operation using Multi-Threshold MOSFETs and Transistor Sizing Technique.” 2015.
- [55] G. Gerosa, S. Gary, C. Dietz, D. Pham, K. Hoover, J. Alvarez, H. Sanchez, P. Ippolito, T. Ngo, S. Litch, J. Eno, J. Golab, N. Vanderschaaf, and J. Kahle, “A 2.2 W, 80 MHz superscalar RISC microprocessor,” *IEEE Journal of Solid-State Circuits*, vol. 29, no. 12, pp. 1440–1454, Dec. 1994.
- [56] H. Harris, Weste Neil. David M., *Cmos Vlsi Design: A Circuits And Systems Perspective*, 3rd ed. Pearson Education India,, 2006.

- [57] M. Keating, D. Flynn, R. Aitken, A. Gibbons, and K. Shi, *Low power methodology manual: for system-on-chip design*. Springer Publishing Company, Incorporated, 2007.
- [58] Y. Kim, W. Jung, I. Lee, Q. Dong, M. Henry, D. Sylvester, and D. Blaauw, “27.8 A static contention-free single-phase-clocked 24t flip-flop in 45nm for low-power applications,” in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, Feb. 2014, pp. 466–467.
- [59] F. Klass, “Semi-dynamic and dynamic flip-flops with embedded logic,” in *VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on*. IEEE, 1998, pp. 108–109.
- [60] E. Late, A. A. Vatanjou, T. Ytterdal, and S. Aunet, “Comparative analysis of flip-flop architectures for subthreshold applications in 28nm fdsoi,” in *Nordic Circuits and Systems Conference (NORCAS): NORCHIP & International Symposium on System-on-Chip (SoC), 2015*. IEEE, 2015, pp. 1–4.
- [61] N. Lotze and Y. Manoli, “A 62 mV 0.13 m CMOS Standard-Cell-Based Design Technique Using Schmitt-Trigger Logic,” *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 47–60, Jan. 2012.
- [62] J. Myers, A. Savanth, R. Gaddh, D. Howard, P. Prabhat, and D. Flynn, “A subthreshold arm cortex-m0+ subsystem in 65 nm cmos for wsn applications with 14 power domains, 10t sram, and integrated voltage regulator,” 2016.
- [63] S. Okugawa and N. Inoue, “Systematic design of D flip-flops using two state variables,” *Electronics Letters*, vol. 23, no. 17, pp. 909–910, Aug. 1987.
- [64] C. Piguet, “The First Quartz Electronic Watch,” in *Integrated Circuit Design. Power and Timing Modeling, Optimization and Simulation*, ser. Lecture Notes in Computer Science, B. Hochet, A. J. Acosta, and M. J. Bellido, Eds. Springer Berlin Heidelberg, 2002, no. 2451, pp. 1–15.
- [65] V. Stojanovic and V. Oklobdzija, “Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems,” *IEEE Journal of Solid-State Circuits*, vol. 34, no. 4, pp. 536–548, Apr. 1999.
- [66] Y. Suzuki, M. Hirasawa, and K. Odagawa, “Clocked CMOS calculator circuitry,” in *Solid-State Circuits Conference. Digest of Technical Papers. 1973 IEEE International*, vol. XVI, Feb. 1973, pp. 58–59.

- [67] Y. Tsvividis, “Eric Vittoz and the Strong Impact of Weak Inversion Circuits,” *IEEE Solid-State Circuits Society Newsletter*, vol. 13, no. 3, pp. 56–58, 2008.
- [68] M. Voernes, T. Ytterdal, and S. Aunet, “Performance comparison of 5 sub-threshold CMOS flip-flops under process-, voltage-, and temperature variations, based on netlists from layout,” in *NORCHIP, 2014*, Oct. 2014, pp. 1–6.
- [69] F. Abouzeid, A. Bienfait, K. C. Akyel, A. Feki, S. Clerc, L. Ciampolini, F. Giner, R. Wilson, and P. Roche, “Scalable 0.35 v to 1.2 v sram bitcell design from 65 nm cmos to 28 nm fdsoi,” *IEEE JSSC*, vol. 49, no. 7, pp. 1499–1505, 2014.
- [70] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [71] A. Biswas and A. P. Chandrakasan, “A 0.36 v 128kb 6t sram with energy-efficient dynamic body-biasing and output data prediction in 28nm fdsoi,” in *ESSCIRC Conference 2016: 42nd*. IEEE, 2016, pp. 433–436.
- [72] B. H. Calhoun, A. Wang, and A. Chandrakasan, “Modeling and sizing for minimum energy operation in subthreshold circuits,” *IEEE JSSC*, vol. 40, no. 9, pp. 1778–1786, 2005.
- [73] L. Chang, D. M. Fried, J. Hergenrother, J. W. Sleight, R. H. Dennard, R. K. Montoye, L. Sekaric, S. J. McNab, A. W. Topol, C. D. Adams *et al.*, “Stable sram cell design for the 32 nm node and beyond,” in *VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on*. IEEE, 2005, pp. 128–129.
- [74] J. Chen, L. T. Clark, and T.-H. Chen, “An ultra-low-power memory with a subthreshold power supply voltage,” *IEEE JSSC*, vol. 41, no. 10, pp. 2344–2353, 2006.
- [75] G. Chen, D. Sylvester, D. Blaauw, and T. Mudge, “Yield-driven near-threshold sram design,” *IEEE transactions on very large scale integration (VLSI) systems*, vol. 18, no. 11, pp. 1590–1598, 2010.
- [76] Y.-W. Chiu, Y.-H. Hu, J.-K. Zhao, S.-J. Jou, and C.-T. Chuang, “A sub-threshold sram with embedded data-aware write-assist and adaptive data-aware keeper,” in *ISCAS 2016*. IEEE, 2016, pp. 1014–1017.
- [77] X. Deng and T. W. Houston, “Loadless 4t sram cell with pmos drivers,” May 4 2004, uS Patent 6,731,533.

- [78] D. Evans, "The internet of things," *Whitepaper, Cisco Internet Business Solutions Group (IBSG)*, vol. 1, pp. 1–12, 2011.
- [79] S. Hanson, M. Seok, Y.-S. Lin, Z. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, and D. Blaauw, "A low-voltage processor for sensing applications with picowatt standby mode," *IEEE JSSC*, vol. 44, no. 4, pp. 1145–1155, 2009.
- [80] J. Kwong, Y. K. Ramadass, N. Verma, and A. P. Chandrakasan, "A 65 nm sub-microcontroller with integrated sram and switched capacitor dc-dc converter," *IEEE JSSC*, vol. 44, no. 1, pp. 115–126, 2009.
- [81] K. Noda, K. Matsui, K. Imai, K. Inoue, K. Tokashiki, H. Kawamoto, K. Yoshida, K. Takeda, N. Nakamura, T. Kimura *et al.*, "A $1.9 \mu\text{m}^2$ loadless cmos four-transistor sram cell in a $0.18 \mu\text{m}$ logic technology," in *1998. IEDM 1998*. IEEE, 1998, pp. 643–646.
- [82] M.-H. Tu, J.-Y. Lin, M.-C. Tsai, C.-Y. Lu, Y.-J. Lin, M.-H. Wang, H.-S. Huang, K.-D. Lee, W.-C. Shih, S.-J. Jou *et al.*, "A single-ended disturb-free 9t subthreshold sram with cross-point data-aware write word-line structure, negative bit-line, and adaptive read operation timing tracing," *IEEE JSSC*, vol. 47, no. 6, pp. 1469–1482, 2012.
- [83] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, D. Murray, N. Vallepalli, Y. Wang, B. Zheng, and M. Bohr, "A 3-ghz 70-mb sram in 65-nm cmos technology with integrated column-based dynamic power supply," *IEEE JSSC*, vol. 41, no. 1, pp. 146–151, 2006.
- [84] Y. Tsvividis, "Eric vittoz and the strong impact of weak inversion circuits," *IEEE Solid-State Circuits Society Newsletter*, vol. 13, no. 3, pp. 56–58, 2008.
- [85] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Boeuf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud *et al.*, "Innovative materials, devices, and cmos technologies for low-power mobile multimedia," *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 96–130, 2007.
- [86] D. Jacquet, F. Hasbani, P. Flatresse, R. Wilson, F. Arnaud, G. Cesana, T. Di Gilio, C. Lecocq, T. Roy, A. Chhabra *et al.*, "A 3 ghz dual core processor arm cortex tm-a9 in 28 nm utbb fd-soi cmos with ultra-wide voltage range and energy efficiency optimization," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 4, pp. 812–826, 2014.

- [87] D. Bol, R. Ambroise, D. Flandre, and J.-D. Legat, "Analysis and minimization of practical energy in 45nm subthreshold logic circuits," in *2008 IEEE International Conference on Computer Design*. IEEE, 2008, pp. 294–300.
- [88] B. Zhai, S. Pant, L. Nazhandali, S. Hanson, J. Olson, A. Reeves, M. Minuth, R. Helfand, T. Austin, D. Sylvester *et al.*, "Energy-efficient subthreshold processor design," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 8, pp. 1127–1137, 2009.
- [89] S. Aunet and Y. Berg, "Three sub-fj power-delay-product subthreshold cmos gates," *IFIP VLSI SoC, Perth, Australia*, vol. 1719, 2005.
- [90] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin *et al.*, "Exploring variability and performance in a sub-200-mv processor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 881–891, 2008.
- [91] M.-Z. Li, C.-I. Jeong, M.-K. Law, P.-I. Mak, M.-I. Vai, S.-H. Pun, and R. P. Martins, "Energy optimized subthreshold vlsi logic family with unbalanced pull-up/down network and inverse narrow-width techniques," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 12, pp. 3119–3123, 2015.
- [92] B. Zhai, S. Hanson, D. Blaauw, and D. Sylvester, "Analysis and mitigation of variability in subthreshold design," in *Proceedings of the 2005 international symposium on Low power electronics and design*, 2005, pp. 20–25.
- [93] M. Seok, D. Jeon, C. Chakrabarti, D. Blaauw, and D. Sylvester, "Pipeline strategy for improving optimal energy efficiency in ultra-low voltage design," in *Proceedings of the 48th Design Automation Conference*, 2011, pp. 990–995.
- [94] M. Belleville and O. Faynot, "Evolution of deep submicron bulk and soi technologies," in *Low-power Electronics Design*. CRC Press, 2004, pp. 32–47.
- [95] A. Ferré and J. Figueras, "Leakage in cmos nanometric technologies," in *Low-Power Electronics Design*. CRC Press, 2004, pp. 48–66.
- [96] E. Beigne, A. Valentian, I. Miro-Panades, R. Wilson, P. Flatresse, F. Abouzeid, T. Benoist, C. Bernard, S. Bernard, O. Billoint *et al.*, "A 460 mhz at 397 mv, 2.6 ghz at 1.3 v, 32 bits vliw dsp embedding f max tracking," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 125–136, 2014.

- [97] J.-P. Noel, O. Thomas, M.-A. Jaud, O. Weber, T. Poiroux, C. Fenouillet-Beranger, P. Rivallin, P. Scheiblin, F. Andrieu, M. Vinet *et al.*, “Multi- v_{T} utbb fdsoi device architectures for low-power cmos circuit,” *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2473–2482, 2011.
- [98] L. L. Lewyn, T. Ytterdal, C. Wulff, and K. Martin, “Analog circuit design in nanoscale cmos technologies,” *Proceedings of the IEEE*, vol. 97, no. 10, pp. 1687–1714, 2009.
- [99] R. Taco, I. Levi, M. Lanuzza, and A. Fish, “Extended exploration of low granularity back biasing control in 28nm utbb fd-soi technology,” in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2016, pp. 41–44.
- [100] A. Bryant, J. Brown, P. Cottrell, M. Ketchen, J. Ellis-Monaghan, and E. Nowak, “Low-power cmos at $v_{dd} = 4kT/q$,” in *Device Research Conference. Conference Digest (Cat. No. 01TH8561)*. IEEE, 2001, pp. 22–23.
- [101] L. Nazhandali, M. Minuth, and T. Austin, “Sensebench: Toward an accurate evaluation of sensor network processors,” in *IEEE International. 2005 Proceedings of the IEEE Workload Characterization Symposium, 2005*. IEEE, 2005, pp. 197–203.
- [102] N. Lotze and Y. Manoli, “A 62 mv 0.13 μ m cmos standard-cell-based design technique using schmitt-trigger logic,” *IEEE journal of solid-state circuits*, vol. 47, no. 1, pp. 47–60, 2011.
- [103] A. A. Vatanjou, T. Ytterdal, and S. Aunet, “Energy efficient sub/near-threshold ripple-carry adder in standard 65 nm cmos,” in *2015 6th Asia Symposium on Quality Electronic Design (ASQED)*. IEEE, 2015, pp. 7–12.
- [104] W. Lim, H. Lee, H. Kang, S. Cho, and Y. Yang, “Power amplifier based on 50 ω load impedance using boost dc-dc converter for apt,” in *2015 Asia-Pacific Microwave Conference (APMC)*, vol. 1. IEEE, 2015, pp. 1–3.
- [105] N. Lotze and Y. Manoli, “Ultra-sub-threshold operation of always-on digital circuits for iot applications by use of schmitt trigger gates,” *IEEE Transactions on Circuits and Systems I*, vol. 64, no. 11, pp. 2920–2933, 2017.
- [106] E. A. Vittoz, “Very low power circuit design: fundamentals and limits,” in *1993 IEEE International Symposium on Circuits and Systems*. IEEE, 1993, pp. 1439–1442.

-
- [107] A. Hasanbegovic and S. Aunet, "Low-power subthreshold to above threshold level shifter in 90 nm process," in *NORCHIP*. IEEE, 2009.
- [108] B. Mohammadi and J. N. Rodrigues, "A 65 nm single stage 28 fj/cycle 0.12 to 1.2 v level-shifter," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 990–993.
- [109] Y. Kim, Y. Lee, D. Sylvester, and D. Blaauw, "Slc: Split-control level converter for dense and stable wide-range voltage conversion," in *ESSCIRC, 2012 Proceedings of the*. IEEE, 2012, pp. 478–481.
- [110] M. Lanuzza, P. Corsonello, and S. Perri, "Low-power level shifter for multi-supply voltage designs," *IEEE Transactions on Circuits and Systems II*, vol. 59, no. 12, pp. 922–926, 2012.
- [111] S.-C. Luo, C.-J. Huang, and Y.-H. Chu, "A wide-range level shifter using a modified wilson current mirror hybrid buffer," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, pp. 1656–1665, 2014.
- [112] W. Zhao, A. B. Alvarez, and Y. Ha, "A 65-nm 25.1-ns 30.7-fj robust sub-threshold level shifter with wide conversion range," *IEEE Transactions on Circuits and Systems II*, vol. 62, no. 7, pp. 671–675, 2015.
- [113] M. Lanuzza, P. Corsonello, and S. Perri, "Fast and wide range voltage conversion in multisupply voltage designs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, pp. 388–391, 2015.
- [114] T.-H. Chen, J. Chen, and L. T. Clark, "Subthreshold to above threshold level shifter design," *Journal of Low Power Electronics*, vol. 2, no. 2, pp. 251–258, 2006.
- [115] Y. Kim, D. Sylvester, and D. Blaauw, "Lc 2: Limited contention level converter for robust wide-range voltage conversion," in *VLSI Circuits (VLSIC), 2011 Symposium on*. IEEE, 2011, pp. 188–189.
- [116] J. Zhou, C. Wang, X. Liu, X. Zhang, and M. Je, "An ultra-low voltage level shifter using revised wilson current mirror for fast and energy-efficient wide-range voltage conversion from sub-threshold to i/o voltage," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 3, pp. 697–706, 2015.
- [117] M. Lanuzza, F. Crupi, S. Rao, R. De Rose, S. Strangio, and G. Iannaccone, "An ultralow-voltage energy-efficient level shifter," *IEEE Transactions on Circuits and Systems II*, vol. 64, no. 1, pp. 61–65, 2017.

- [118] R. Lotfi, M. Saberi, S. R. Hosseini, A. R. Ahmadi-Mehr, and R. B. Staszewski, "Energy-efficient wide-range voltage level shifters reaching 4.2 fj/transition," *IEEE Solid-State Circuits Letters*, vol. 1, no. 2, pp. 34–37, 2018.
- [119] M. Ashouei, H. Luijmes, J. Stuijt, and J. Huisken, "Novel wide voltage range level shifter for near-threshold designs," in *Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 285–288.
- [120] A. Hasanbegovic and S. Aunet, "Low-power subthreshold to above threshold level shifters in 90 nm and 65 nm process," *Microprocessors and Microsystems*, vol. 35, no. 1, pp. 1–9, 2011.
- [121] J. D. Meindl and J. A. Davis, "The fundamental limit on binary switching energy for terascale integration (tsi)," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 10, pp. 1515–1516, 2000.
- [122] M. R. Stan and W. P. Burleson, "Bus-invert coding for low-power i/o," *IEEE Transactions on VLSI systems*, vol. 3, no. 1, pp. 49–58, 1995.
- [123] S. Nalam and B. H. Calhoun, "5t sram with asymmetric sizing for improved read stability," *IEEE journal of solid-state circuits*, vol. 46, no. 10, pp. 2431–2442, 2011.
- [124] V. Gherman and M. Cartron, "Soft-error protection of tcams based on eccs and asymmetric sram cells," *Electronics Letters*, vol. 50, no. 24, pp. 1823–1824, 2014.
- [125] I. Alouani, W. M. Elsharkasy, A. M. Eltawil, F. J. Kurdahi, and S. Niar, "As8-static random access memory (sram): asymmetric sram architecture for soft error hardening enhancement," *IET Circuits, Devices & Systems*, vol. 11, no. 1, pp. 89–94, 2017.
- [126] N. Azizi, F. N. Najm, and A. Moshovos, "Low-leakage asymmetric-cell sram," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 4, pp. 701–715, 2003.
- [127] Y.-J. Chang, F. Lai, and C.-L. Yang, "Zero-aware asymmetric sram cell for reducing cache power in writing zero," *IEEE transactions on VLSI systems*, vol. 12, no. 8, pp. 827–836, 2004.
- [128] H. Fujiwara, K. Nii, H. Noguchi, J. Miyakoshi, Y. Murachi, Y. Morita, H. Kawaguchi, and M. Yoshimoto, "Novel video memory reduces 45% of

bitline power using majority logic and data-bit reordering," *IEEE transactions on VLSI systems*, vol. 16, no. 6, pp. 620–627, 2008.

ISBN 978-82-326-6512-9 (printed ver.)
ISBN 978-82-326-6971-4 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)



NTNU

Norwegian University of
Science and Technology