

**Ultra-Low Power Circuit Design
for Cubic-Millimeter Wireless Sensor Platform**

by

Yoonmyung Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2012

Doctoral Committee:

Professor David Blaauw, Chair
Professor Dennis Michael Sylvester
Associate Professor Jerome P. Lynch
Assistant Professor David D. Wentzloff
Associate Professor Jae Yoon Sim, POSTECH

© Yoonmyung Lee

All Rights Reserved

2012

*To my God,
who has always directed my steps with His endless grace,
and Jiwon,
my best friend, supporter, counselor, and lovely wife,
and my beloved son and parents,
with love and gratitude*

ACKNOWLEDGEMENTS

By the grace of God, I have met so many wonderful people during my doctoral journey and I am sincerely grateful for the support, collaboration and encouragement I have received from them. Without their professional and personal supports, this dissertation would never have been completed.

First and foremost, I would like to offer my sincere gratitude to my advisor, professor David Blaauw, who has provided tremendous support, enthusiasm and motivation for my Ph.D. study. He has always thought about how I could learn and grow more, not just how to finish a work. He has also paid kind attention to my family and financial matters which made my life in Ann Arbor enjoyable yet challenging. I simply could not wish for a better or friendlier advisor.

For every research project throughout my graduate study, I have worked with professor Dennis Sylvester whose expertise, suggestions and encouragement have become a vital part of my research works. He really supported me as if I were his student. I appreciate him and my advisor for showing me an outstanding example of collaboration in academic research.

Professor David Wentzloff and Prabal Dutta have been great collaborators for joint cubic-millimeter sensor project. Their enthusiastic contribution made the joint project more valuable and successful. Professor Jerome Lynch and Jae-Yoon Sim graciously agreed to be on my dissertation committee.

I have been blessed to work with a friendly, enthusiastic and cheerful group of fellow students. Cubic-millimeter sensor node project could be successfully finished only with hard working and contributions from M3 team: Inhee Lee, Yejoong Kim, Gyouho Kim, Suyoung Bang. I have learned so much knowledge necessary for chip design from Zhiyoong Foo, David Fick, Carlos Tokunaga, and Jerry Kao. Collaboration with Bharan Giridhar, Mao-Ter Chen, Daeyeon Kim and Junsun Park was essential for some of my works and I also enjoyed co-working with Mingoo Seok,

Scott Hanson, Yu-Shiang Lin and Michael Wieckowski. I also enjoyed having research discussions with Greg Chen, Jonathan Brown, Kuo-Ken Huang, Prashant Singh, Narrachman Liu, Eric Karl, Brian Cline, Sudhir Satpathy, Yongjun Park, Sangwon Seo, Hyo Gyuem Rhew, Dongsuk Jeon, Dongmin Yoon, and Dongjin Lee. I also appreciate Jae-sun Seo and Mingoo Seok for being great mentors for me.

During my internships, I have gained invaluable experience and learned abundant knowledge that I can only learn from industry with kind guidances from my mentors. I would like to thank my mentors during my internships: Dr. Ram Krishnamurthy, Dr. Himanshu Kaul from Intel Corporation, and Dr. Leland Chang from IBM.

I have received financial support from Samsung Scholarship Foundation and Intel PhD Fellowship Program hence I would like to express my appreciation for their generous support during my graduate study. I also would like to thank the sponsors of the projects I have participated: the National Science Foundation (NSF), Defense Advanced Research Projects Agency (DARPA), the Multiscale System Center (MuSyC) of Semiconductor Research Corporation (SRC) and ARM.

Most importantly, I would like to express my deepest appreciation to my family – especially my wife, Jiwon Kim, parents and parents-in-law – for their unconditional support, love, prayer and encouragement throughout my long graduate life. I also would like to say thank you to my beloved son, Joshua Seungyun Lee, for so many “I love you daddy”s he has written, told and shown to me.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	x
CHAPTERS	
1 Introduction	1
1.1 Challenges for Cubic-Millimeter Sensor System	3
1.2 Contribution of This Work and Organization	6
2 A Modular 1.0mm ³ Die-Stacked Sensing Platform	9
2.1 Introduction	9
2.2 1mm ³ Sensing Platform Overview	9
2.3 Low Power I ² C Communication	14
2.3.1 Low Power I ² C Background	14
2.3.2 Low Power I ² C Implementation	15
2.3.3 Low Power I ² C Measurement Results	16
2.4 System Operation	16
2.5 Conclusion	20
3 Standby Power Reduction for Cubic-Millimeter Sensor Systems	22
3.1 Introduction	22
3.2 Standby Power Reduction for Logic Circuits	23
3.3 Standby Power Reduction for Memory Circuits	25
3.3.1 Leakage Reduction for Power Gated Blocks	26
3.3.2 Bit-line Leakage Reduction	26
3.3.3 Intra-cell Leakage	28
3.4 Ultra-Low Power Clock Generation	28
3.5 Measurement Results	30
3.5.1 Logic Circuits	30
3.5.2 Memory	32

3.5.3	Ultra-low Power Clock Generation	34
3.6	Conclusion	35
4	2T Dual V_{th} Gain Cell eDRAM for Cubic-Millimeter Sensor Systems	36
4.1	Introduction	36
4.2	2T eDRAM Gain Cell	37
4.2.1	Conventional 2T Gain Cell	37
4.2.2	2T Dual- V_{th} Gain Cell	38
4.3	Area-Efficient Single Inverter Sensing	39
4.4	Measurement Results	44
4.5	Conclusion	46
5	Low Power 7T SRAM with Heterojunction Tunneling Transistors (HETTs)	47
5.1	Introduction	47
5.1.1	HETT Device Characteristic	48
5.1.2	HETT Device Modeling	50
5.1.3	Asymmetric Current Flow of HETT	51
5.2	Limitations in Standard 6T SRAM	52
5.2.1	CMOS Standard 6T SRAM	52
5.2.2	HETT Standard 6T SRAM with Inward Access Transistors	53
5.2.3	HETT Standard 6T SRAM with Outward Access Transistors	55
5.3	Alternative SRAM Design with HETT	56
5.3.1	Read Structure for HETT SRAM	56
5.3.2	Write Structure for HETT SRAM	58
5.3.3	7T SRAM for HETT	59
5.4	Conclusion	63
6	A Sub-nW Gate-Leakage Based Temperature Compensated Timer for Cubic-Millimeter Sensor Systems	64
6.1	Introduction	64
6.1.1	Operation of Ultra-Low Power Wireless Sensor Node	64
6.1.2	Prior-art Timers for Ultra-Low Power Wireless Sensor Node	65
6.1.3	Metrics for Ultra-Low Power Timers	66
6.2	Multi-Stage Gate-Leakage Based Timer for Low Jitter	66
6.3	Temperature Compensation for Multi-Stage Gate-Leakage Based Timer	68
6.4	Measurement Results	69
6.4.1	Uncertainty Reduction	69
6.4.2	Long-term Uncertainty	76
6.4.3	Temperature Compensation	79
6.4.4	Die to Die Variation of Gate-Leakage-Based Timer	79
6.5	Conclusion	79
7	Conclusions	85
	BIBLIOGRAPHY	87

LIST OF FIGURES

Figure

1.1	Bell’s Law predicts continuous scaling of minimal-sized computing systems	2
1.2	Components of typical wireless sensor node	5
2.1	Stacked Die Structure and Dimension of 1mm ³ Sensing Platform	10
2.2	System Block Diagram of 1mm ³ Sensing Platform	11
2.3	Power Management States with Battery Voltage Changes	13
2.4	Conventional I ² C circuit diagram and data transfer waveform [19]	14
2.5	Proposed modified I ² C circuit diagram	15
2.6	Measured I ² C wavefrom and illustration of SCL sub-cycle	16
2.7	Proposed 1mm ³ die-stacked sensor platform.	17
2.8	Die micrograph of each layer in the 1.0mm ³ sensor platform.	18
2.9	An example of sensor platform system operation.	19
2.10	Standby power consumption by function unit for the 1.0mm ³ sensor platform.	19
3.1	Logic circuit standby power reduction by super cut-off.	24
3.2	Leakage paths in low leakage memory cell.	25
3.3	Proposed circuit for bit-line leakage reduction.	27
3.4	pW clock generator with current starved transistors and output waveform comparison between different starved transistor placement schemes.	29
3.5	Generated super cut-off voltage and power consumption of charge pump for logic circuit standby power reduction.	31
3.6	CPU leakage and charge pump operation power in standby mode.	31
3.7	Total standby power of CPU and charge pump with various CPU footer sizing.	32
3.8	Generated super cut-off voltage and power consumption of charge pump for memory standby power reduction.	33
3.9	Memory leakage and charge pump operation power in standby mode.	33
3.10	Optimal and generated clock frequency normalized at 40°C.	34
4.1	Structure of conventional 2T eDRAM cell and its three types of leakages that can destroy stored data.	37
4.2	2T dual-V _{th} eDRAM structure and possible data loss scenarios. Since data 1 loss is protected better than data 0 loss, the preferred state of the bit line for minimizing cell decay in standby time is 0.	38

4.3	Layout of the 2T dual- V_{th} eDRAM cell and its dimensions.	39
4.4	Single inverter sensing scheme and simulated waveforms with 32bits/bitline.	40
4.5	Block diagram of implemented eDRAM array.	41
4.6	Measured retention time for 16 dies and refreshing power for typical die versus temperature.	42
4.7	Distribution of measured cell retention time in typical die.	43
4.8	Array efficiency compared with other state-of-the-art eDRAM and low power memory: Boosted 3T [28], Gain Cell 2T [29], 3T Micro Sense-Amp [33], Pseudo-Two-Port [32]	45
5.1	Tunneling FET device concept as depicted by a) band diagrams in the source-to-drain direction, and b) qualitative current-voltage characteristics.	49
5.2	CMOS-compatible implementation of complementary tunneling FETs with type-II source-to-body hetero-junctions to improve device drive current.	50
5.3	Device symbols for (a) NHETT (b) PHETT.	51
5.4	Drain current of HETT device with $L=40\text{nm}$ (a) Forward bias (b) reverse bias.	52
5.5	Current flow paths in (a) read and (b) write operations in CMOS 6T SRAM.	53
5.6	Current flow paths in (a) read and (b) write operations in HETT 6T SRAM with inward direction access transistors.	54
5.7	Static noise margins of HETT 6T SRAM with (a) inward and (b) outward access transistor with $V_{DD}=0.5V$	55
5.8	Current flow paths in (a) read and (b) write operations in HETT 6T SRAM with outward direction access transistors.	56
5.9	Alternative read structures for HETT-based SRAM (a) Single HETT read (b) 8T read (c) reduced 8T read.	57
5.10	Read operation with reduced 8T read.	57
5.11	Alternative write structures for HETT-based SRAM (a) Two-side transmission gate write (b) One-side transmission gate write (c) Two-side NHETT pull down write (d) Two-side PHETT pull up write.	58
5.12	Proposed HETT 7T SRAM structure.	60
5.13	(a) 8T layout [50] and (b) corresponding HETT 7T layout.	60
5.14	Read/Write margin of 45nm commercial bulk CMOS 6T SRAM and HETT 7T SRAM.	61
5.15	Standby power of CMOS 6T and HETT 7T SRAM.	62
6.1	Power consumption of example ultra-low power wireless sensor module in various operation modes.	65
6.2	Proposed multi-stage gate-leakage based timer.	67
6.3	Effect of (a) multi-staging and (b) boosted charging.	67
6.4	Opposite temperature dependency of ZVTMOS and PMOS gate leakage current.	69
6.5	Circuit diagram of temperature compensated timer.	70
6.6	Controller for adaptive temperature compensation.	71
6.7	Duty cycle and period/stage change with number of stages.	71
6.8	Jitter and hourly clock uncertainty reduction with multi-staging.	72

6.9	Jitter and hourly clock uncertainty reduction with boosted charging.	73
6.10	Trade-off between various types of timers.	73
6.11	Power consumption of multi-stage gate-leakage based timer.	74
6.12	Distribution of period for 24 hour continuous measurement.	75
6.13	Distribution of error for measuring 1 hour synchronization cycle.	75
6.14	Theoretical and actual uncertainty with various timers.	76
6.15	Standard deviation of synchronization error.	77
6.16	Allan deviation of gate-leakage-based timers.	78
6.17	Period of temperature compensated timer with selected ZVTMOS/PMOS configurations.	80
6.18	Period of deviation vs temperature deviation for selected configurations.	81
6.19	Effective temperature sensitivity for -20°C - 60°C vs number of configurations.	81
6.20	Temperature profile used for testing closed loop temperature compensation.	82
6.21	Accumulated timing error with and without temperature compensation.	82
6.22	Distribution of gate-leakage-based timer period (number of active stages = 3).	83

LIST OF TABLES

Table

4.1	Comparison with other state-of-the art eDRAM and low power memories: Boosted 3T [28], Gain Cell 2T [29], Ultra-low Power SRAM [6]	44
4.2	Performance summary of 2T dual V_{th} eDRAM	45

CHAPTER 1

Introduction

Since the invention of transistors, continuous technology scaling has led to the integration of computational capabilities in an increasingly small volume. This has been leveraged to create both very small and yet highly capable systems, as well as new multi-core and/or networking technologies that push the upper limits of modern computing performance. This, in turn, has produced a diversification of computing platforms, ranging from portable handheld devices to building-scale data centers. According to Bell's Law, a new class of smaller computers is developed approximately every decade by using fewer components or fractional parts of state-of-the-art computing system [1]. First computers introduced in 1940s were as big as a room or even a small building, but the size of computers have continuously shrunk. Smaller and more affordable computers are made in forms of workstations in 1970s and personal computers in 1980s, and mobilities are added to computers with laptops in 1990s and portable handheld devices in 2000s as shown in Figure 1.1. As the next class of computers, increasing number of recent research works have shown the potential of mm³-scale wireless sensor nodes in wide range of applications [2][3][4][5].

Cubic-millimeter-scale wireless sensor nodes have a wide range of applications which take advantage of their small form factor, long device lifetime and reduced cost. Its low power operation enables flow rate monitoring in oil pipelines [2], or in heating, ventilation, and air conditioning(HVAC) systems [3] since a wireless sensor node can operate for years of lifetime without battery replacement which would require costly infrastructure disassembling. Cheap sensor cost allows distribution of many data collection points for a building's structural health monitoring [4] and environmental monitoring. A wireless sensor node can also take great advantage of its small

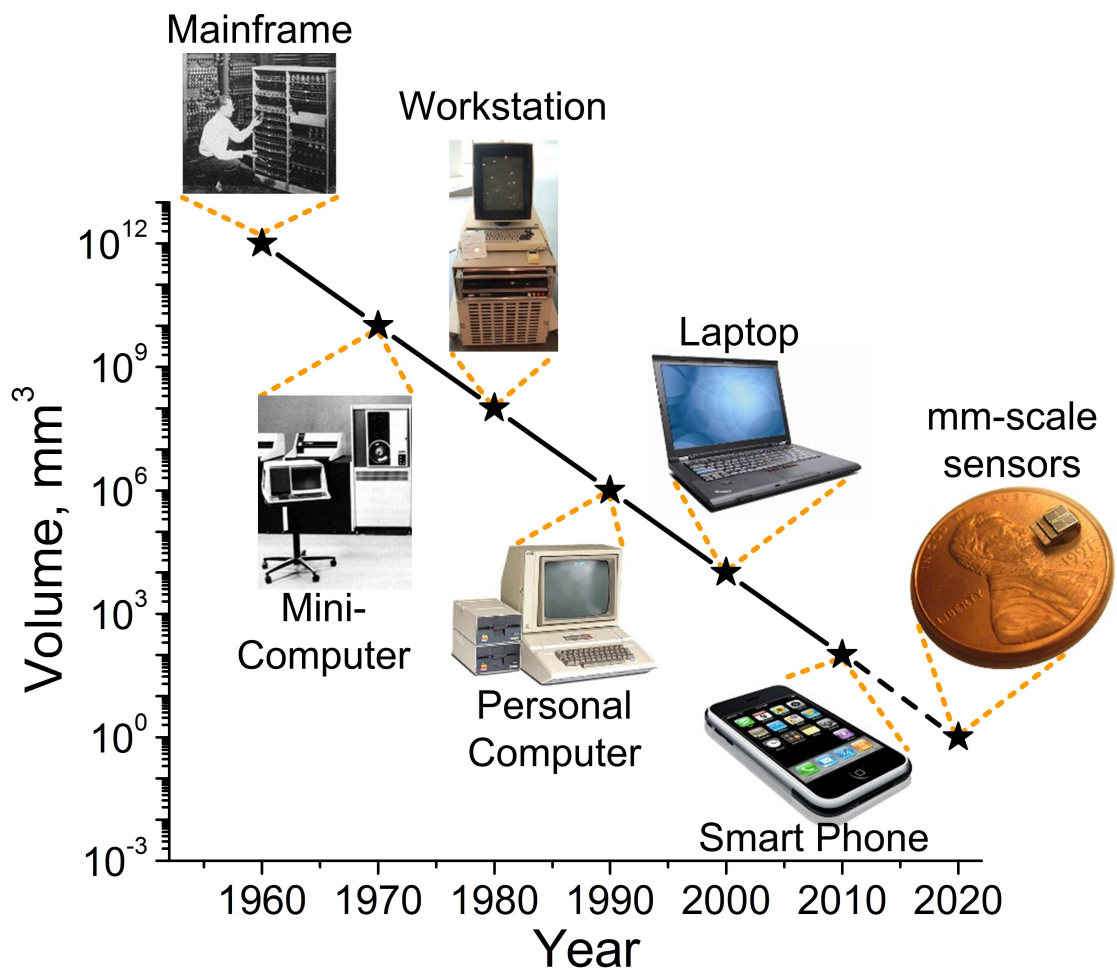


Figure 1.1: Bell's Law predicts continuous scaling of minimal-sized computing systems

size in implanted medical applications where the size of the sensor node is directly related to the invasiveness of implantation surgery [5]. However, today's wireless sensors are composed of multiple components on a printed circuit board (PCB). Bulky batteries are included in the system to power the circuit components with adequate lifetimes. The result is a milliwatt-powered system that is centimeters or tens of centimeters on a side. As Bell's law has predicted, advances in circuit and system design, packaging and battery technologies have created new exciting opportunities to dramatically reduce the size and cost of wireless sensors without affecting device lifetime. A new class of the miniature computing system is therefore poised to be unveiled : mm^3 -scale sensor nodes.

The stringent size limit on the form factor volume also limits the size of the battery. Therefore, batteries that are widely used, such as AA-sized or coin cell batteries, cannot be used for mm^3 -scale system. Recent prototype demonstrations of mm^3 -scale wireless sensor nodes [6][7][8] were based on a thin-film Li battery whose energy density is limited to $1\mu\text{Ah}/\text{mm}^2$ [9]. Therefore, average power consumption of a typical mm^3 -scale sensor node should be restricted to nanowatts or picowatts to sustain its functionality for months of lifetime. Energy harvesters such as photovoltaic, thermoelectric generator and microbial fuel cells can expand the average power budget, but sporadic availability of harvested energy can limit their use. To operate within such extremely constrained energy budget, ultra-low power operation of the sensor node is essential, which requires a new class of design techniques such as deep sub-threshold operation, extensive standby mode design, nanoamp-scale load power management, and aggressive power gating.

1.1 Challenges for Cubic-Millimeter Sensor System

Creating a mm^3 -scale sensor node with nanowatts of average power is a nontrivial challenge. To achieve a nanowatt-order average system power, average power consumption of each component in the system should be limited to nanowatts or less. Therefore, circuit design of each component should be carefully revisited to bring down the power consumption. This typically requires following three approaches.

Firstly, duty-cycling of each component should be maximized, especially with the components with high active power. Sensor nodes often require periodic operations and these periods of opera-

tions need to be determined by their applications. For example, monitoring intraocular pressure for glaucoma only requires pressure measurements every 10-15 minutes. Temperature monitoring for greenhouse control would need measurements every minute or so, whereas hourly measurements should be sufficient for reservoir water level monitoring. Therefore, a sensor system only has to be activated when it is required to take measurement, process measured data and store the processed data to its storage or transmit data to a base station/other sensor node. Moreover, each specific component only needs to be activated when its function is required. For example, a data processing microprocessor does not need to be turned on until the sensor measurement data is available. Since these operations can be performed in as little as ms, each component in the sensor system can be duty-cycled, with its own schedule, maximizing its savings of active power.

Secondly, the active power of each component has to be reduced with more energy-efficient operation. Although duty-cycling can put each component in a low-power consuming standby mode for the most of its lifetime, circuits that have several orders of magnitude higher active power than standby power can still consume a dominant portion of their total energy as active energy. Typical analog circuits require bias currents which can easily exceed microamps, and a large digital signal processor can consume significant active power as well. Moreover, some sensor components cannot be turned off at all. For example, SRAM for data retention cannot be turned off until the measured/processed data is collected by the user. Also the timer in a sensor node also has to be always on to track time for the next operation. For these components, significant effort is required to reduce the active power consumption.

Thirdly, standby power of each component has to be limited to nanowatts or less. Since the sensor system spends most of its time in standby mode, standby power of each component is of critical concern. The less frequent the sensor system operates, the more standby power is important. Our past work [6] suggests that standby power-oriented design strategy can significantly extend the lifetime of an infrequently activated sensor system.

These three approaches should be taken differently from component to component. Figure 1.2 shows the basic components of a typical wireless sensor system. There are one or more sensors in a sensor system and a microprocessor is required to process the raw measured data obtained by the sensors and extract useful information for the user. This processor or another processor can control the overall sensor operation sequence. After data processing, extracted information is

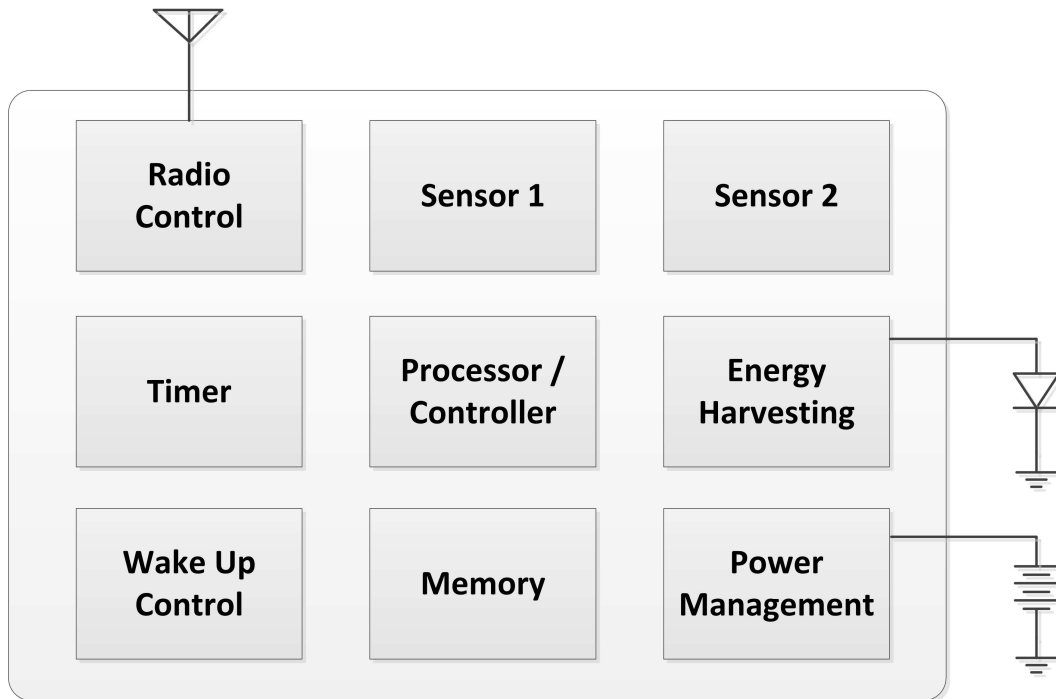


Figure 1.2: Components of typical wireless sensor node

saved to a memory. The memory is used as temporary storage of extracted data until the data is transmitted to an external communication device. The memory is also used as a scratchpad for data processing and storage for the execution program. After each measurement event, entire system is put in standby mode to save power. Before entering the standby mode, the processor sets up the next wake up time. During the standby mode, the timer is the only component that is active to track current time. When next designated wake up time is reached, the timer signals a wake up controller which then releases the power gating of each component with a required sequence and eventually hands over the control of entire system to the main processor. As the measured data is accumulated in the memory, data is transmitted to other sensor nodes or a base station where measurement data from multiple sensor nodes are collected and analyzed to obtain a larger picture of a monitored object. A power management unit provides efficient voltage regulation for both active and standby mode by utilizing two distinct configurations which can support microamp and nanoamp load current. Millimeter-scale energy harvesting can be employed, and the power management unit can switch to a battery charging mode when there is enough scavenged energy for both sensor operation and battery charging.

1.2 Contribution of This Work and Organization

This work proposes a number of new circuit techniques for designing various components for a mm^3 -scale wireless sensor node. We also investigate an effective standby power reduction scheme applicable to general circuits. Combining these techniques, we demonstrate a 1.0mm^3 sensor node system.

In Chapter 2, a modular 1.0mm^3 sensor system [16] is presented. This is the smallest complete sensor system with commercial microprocessor ever presented. Various challenges in realizing the 1.0mm^3 form-factor are addressed here. For example, to encapsulate as much functions as possible in 1.0mm^3 volume, a die-stacked structure with wirebonding is used, which also maximizes modularity of the system by allowing freedom of adding or removing IC layers. The limited number of bondwires on the sensor system only makes serial inter-layer communication feasible. With a limited power budget, conventional serial communication protocol is not feasible in 1.0mm^3 system and we propose a novel low power I²C scheme to overcome this issue. We also present an overview of the system operation of the 1.0mm^3 sensor node and other components in the system including two ARM®Cortex-M0 microprocessors, low power memory for data retention during standby, power management unit with adaptive multi-modal energy harvesting, battery voltage monitoring with brown-out detector, and optical communication which enables initialization, synchronization and re-programming of the sensor node.

In Chapter 3, standby power reduction schemes for circuit components in mm^3 -scale sensor nodes are investigated [11]. Typical mm^3 -scale sensor nodes are duty-cycled and spend most of their time on standby mode. The standby mode power consumption typically is significantly lower than active power. For example, a prototype sensor node in [6] has 35pW standby mode power, whereas active power is 220nW. However, overall energy budget can be dominated by standby power due to extremely low ratio of time spent in active mode to standby mode. For example, periodic temperature measurement and data compression in [6] only takes $\sim 100\text{ms}$ whereas measurement period can vary from 10 minutes to one hour which results in more than 75% of total energy consumed in standby mode. Therefore, optimizing circuits for low standby mode power consumption is a key approach for energy efficient sensor node. Two different approaches for logic circuits and memory circuits are investigated and an ultra-low power charge pump with pW-order

power consumption is proposed to generate additional bias voltage required for standby power reduction. With proposed strategies, standby power of logic circuits is reduced by up to $19\times$ and the memory circuits by 30%.

The most widely used method for reducing standby mode power is power-gating and sensors or processors can be completely power-gated during standby mode since releasing power-gating can completely restore the power-gated circuit's functions. However, memory in charge of retaining measured data or execution code cannot be power-gated since the data stored in the memory can be lost with power-gating. For this reason, power reduction in memory has to take a different approach. Chapter 4 and 5 discuss low leakage memory for mm^3 -scale sensor systems. We propose novel circuit designs for two different flavors of memories - embedded DRAM (eDRAM) and SRAM. In Chapter 4, a low leakage eDRAM design [12] is presented which has significantly smaller area than previously proposed low-leakage SRAM in [6][7]. By taking advantage of a 2T dual threshold voltage gain cell structure, refresh frequency is lowered by $8\times$ and retention power is reduced by more than $5\times$ compared to the state-of-the-art low power eDRAM [28]. In Chapter 5, a 7T SRAM using hetero-junction tunneling transistor (HETT) [13] is proposed. HETT is a CMOS compatible device developed for low subthreshold swing of $<60\text{mV/decade}$, which can significantly improve low power operation of circuits with low supply voltage. However, its asymmetric drain current limits its use on standard 6T SRAM. We propose a 7T SRAM structure which separates read and write paths in similar way with 8T SRAM [50], but by taking advantage of asymmetric nature of the HETT, 1 transistor in read structure of 8T SRAM is removed. The proposed 7T HETT-based SRAM reduces leakage power by $9\text{-}19\times$ with 15% area overhead over standard 6T SRAMs.

In Chapter 6, a novel low power timer which is a critical component for wireless sensor node synchronization is presented. To collect user-interested data measured from a sensor node, wireless communication should be activated. However, due to the high power consumption of wireless radios, periodic synchronization is required to duty-cycle the radios. This requires an accurate timing reference which stays within a power budget of the mm^3 -scale sensor node. By taking advantage of low gate-leakage current, a multi-stage temperature-compensated timer with reasonable accuracy [14][15] is proposed to consume 660pW . The standard deviation for measuring a one hour synchronization cycle was 196ms and temperature dependency could be reduced down

to 31ppm/°C with compensation.

All presented works are summarized and concluded in Chapter 7.

CHAPTER 2

A Modular 1.0mm³ Die-Stacked Sensing Platform

2.1 Introduction

Wireless sensor nodes have many compelling applications such as smart buildings, medical implants, and surveillance systems. However, existing sensor nodes are bulky, measuring $> 1\text{cm}^3$, and they are hampered by short lifetimes. These sensor nodes fail to realize the “smart dust” vision first proposed in [17]. Smart dust requires a mm^3 -scale, wireless sensor node with perpetual energy harvesting. Recently, two application-specific implantable microsystems [8][18] demonstrated the potential of a mm^3 -scale system in medical applications. However, [18] is not programmable and [8] lacks a method for re-programming or re-synchronizing once encapsulated. Other practical issues remain unaddressed, such as a means to protect the battery during the time period between system assembly and deployment and the need for flexible design to enable use in multiple application domains. To this end, in this chapter, we propose a 1mm^3 generic sensor platform, whose modularity allows easy combination with application-specific mm^3 -scale sensors, realizing the “smart dust” vision.

2.2 1mm³ Sensing Platform Overview

A 1.0mm^3 sensing platform is designed with stacked integrated circuit(IC) dies fabricated in three different technologies. Figure 2.1 shows the dimension of each dies and the wirebonding scheme of the sensor system. To enforce 1.0mm^3 volume, each layer measures less than

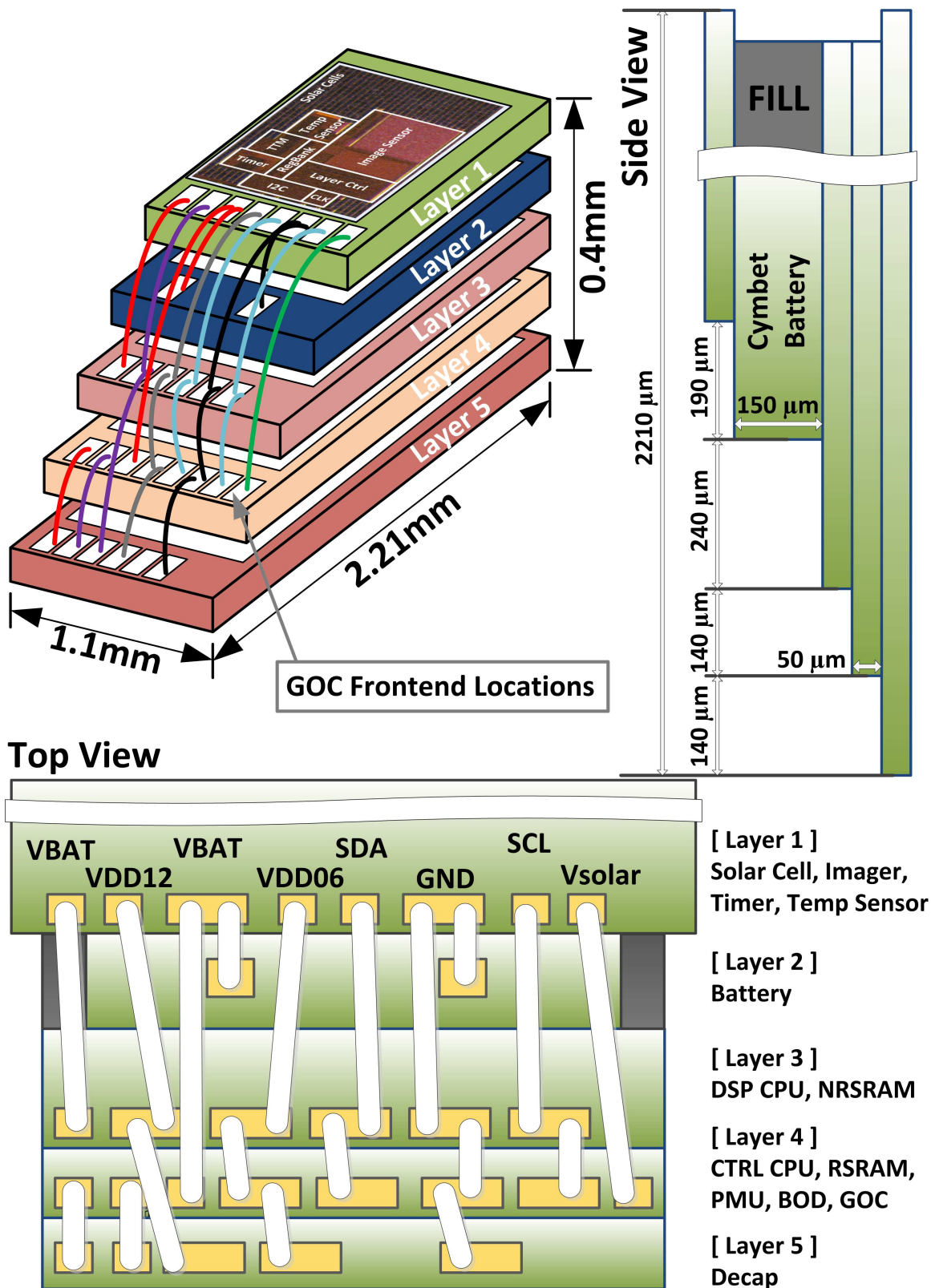


Figure 2.1: Stacked Die Structure and Dimension of 1mm³ Sensing Platform

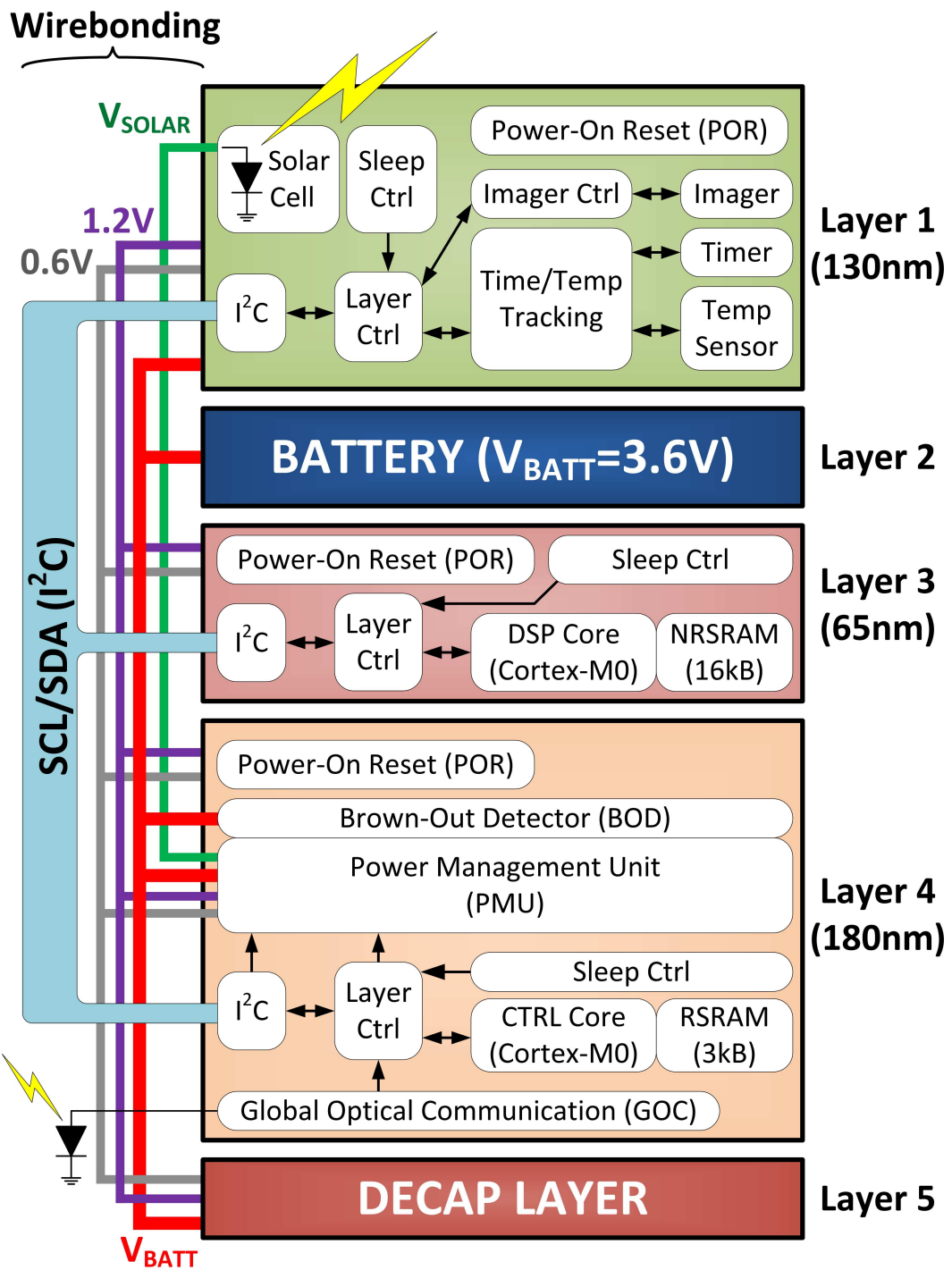


Figure 2.2: System Block Diagram of 1mm³ Sensing Platform

$2.21 \times 1.1\text{mm}$ and the length of each layer has to be reduced by more than $140\mu\text{m}$ compared to the lower layer to provide enough clearance for bond-wires. The height of each IC layer is thinned to $<50\mu\text{m}$ whereas custom made thin film Li battery is $150\mu\text{m}$ high. The die-stacked structure with wirebonding in proposed system not only provides maximum functionality or silicon area per volume but also enables easy expansion of the system with additional layers. End users can create a sensor system for new application by designing an application-specific layer in preferred technology, which complies with the system power and energy budget, and providing identical inter-layer communication interface.

Figure 2.2 shows the system block diagram. Two ARM®Cortex-M0 processors are located in separate layers: 1) The DSP CPU must efficiently handle data streaming from the imager (or other user-provided sensors), and is built in 65nm CMOS (Layer 3) with a large 16kB non-retentive SRAM (NRSRAM). 2) The CTRL CPU manages the system using an always-on 3kB retentive SRAM (RSRAM) to maintain the stored operating programs, and is built in low leakage 180nm CMOS (Layer 4). Solar cells for energy harvesting and a low power imager are placed in the top layer (Layer 1) for light exposure. A gate-leakage-based timer [14] and a temperature sensor is also implemented in Layer 1, which is fabricated in 130nm CMOS for gate-leakage current optimization for timer accuracy. Time tracking with temperature compensation is implemented using this timer, providing a timing reference to synchronize radios that will be attached to this modular platform in future work.

A switch-capacitor-network-based (SCN) PMU is implemented to provide efficient voltage down conversion from battery voltage (3.2V-4.1V) to two low supply voltages (1.2V and 0.6V) for low power operation. Flexible conversion rate reconfiguration allows efficient voltage down conversion for wide range of battery voltages. Power consumption of the system ranges from 11nW in sleep mode up to $\sim 40\mu\text{W}$ in active mode. Additional configurable SCN for harvesting allows energy harvesting from wide range of harvesting voltages, which allows energy harvesting in various environment typically determined by applications. For example, we have demonstrated harvesting within the system from Layer 1's 0.54mm^2 solar cell, as well as an external 125mm^2 thermoelectric generator, and an ocean microbial fuel cell.

Power management of the system is adaptively performed with battery voltage changes. Figure 2.3 shows the power management state change corresponding to the battery voltage change. A

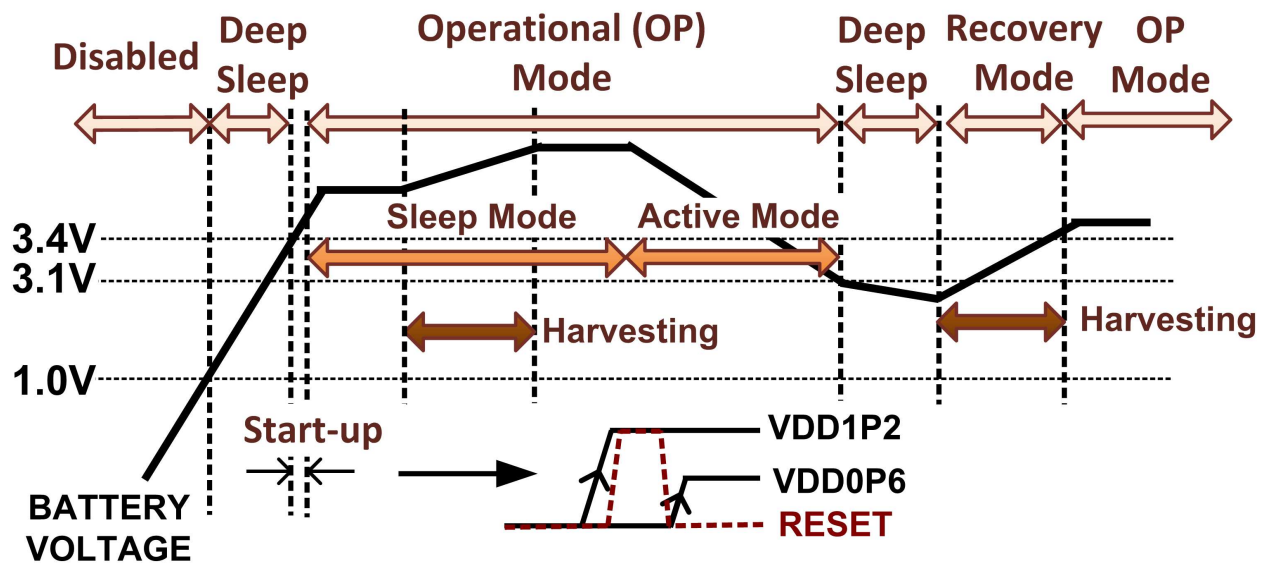


Figure 2.3: Power Management States with Battery Voltage Changes

low power (242pW) brown out detector (BOD) is used to detect initial battery contact (<1V) where the entire system is disabled. As the initial battery voltage ramps up, the PMU enters Deep Sleep mode until enough voltage (>3.4V) is measured. Once the battery voltage stabilizes above 3.4V, the PMU enters operational mode and the system is activated. With battery discharge, output voltage declines and this could incur permanent damage to the Li battery if discharged below 3.0V. To prevent this, the BOD detects voltage below 3.1V where all supplies are turned off and the system enters 185pW deep sleep mode. Availability of harvested energy is monitored in Deep Sleep mode and when sufficient light is detected, the system enters recovery mode, enabling energy harvesting. After the battery has reached a sufficient voltage (>3.4V), the system returns to its normal operational state and distributes a power on reset (POR) signals to all IC layers by sequentially releasing the 1.2V and 0.6V supplies. This sequence is detected by the POR circuit in each IC layer, which forces local reset for correct initialization.

Global optical communication (GOC) unit is provided to serve three critical purposes that enhance the usability of this sensing platform: initial programming after system assembly, re-synchronization during use, and re-programming out of Deep Sleep mode or when the program has become corrupted. The GOC front-end sensors are located in Layer 4 to enable direct communication with the PMU and programming RSRAM without utilizing inter-layer interface. Three photodiode front-ends with a majority-voted receiver is used for robustness. To prevent false trig-

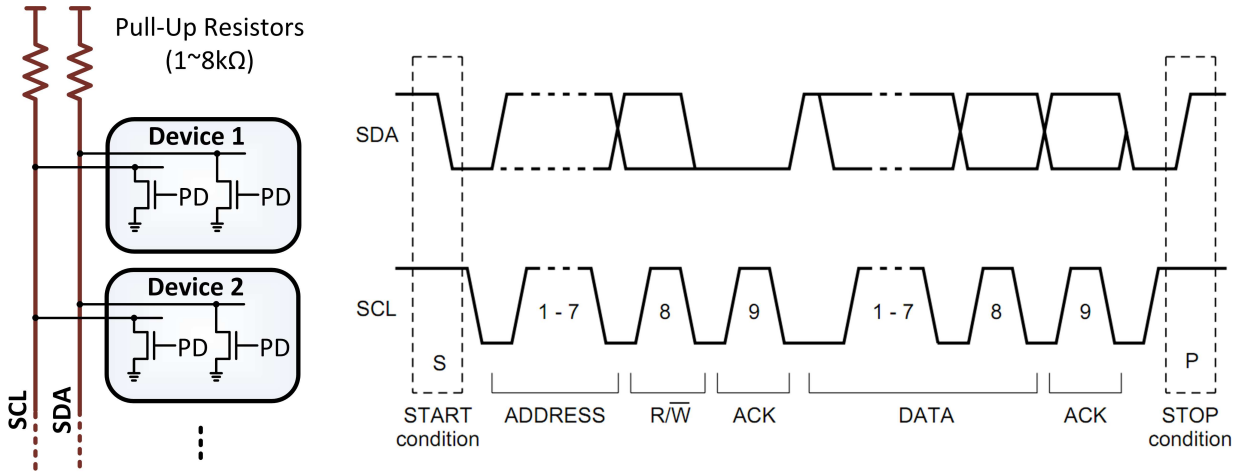


Figure 2.4: Conventional I²C circuit diagram and data transfer waveform [19]

gers from ambient light, a 16-bit predetermined bit pattern is used as a global passcode to initiate the GOC transition. Once the passcode is validated, GOC runs at an 8 times faster rate to enable higher transmission rate. In addition, a parity byte is used to detect any bit errors during data transmission. A local chip-ID with masking also allows selective batch-programming of multiple sensor nodes. GOC is measured to be operational up to 120bps, consuming 72pJ/bit.

2.3 Low Power I²C Communication

2.3.1 Low Power I²C Background

Die-stacked structure of proposed sensor platform requires communication among different layers. Due to the pad count limitations available in the 1mm³ form factor, number of wires used in communication is of critical concern. I²C [19] is a widely used industry standard serial communication protocol that only requires two wires - serial clock (SCL) and serial data (SDA) - and is easy to expand with any I²C-compatible devices. However, conventional I²C relies on pull up resistors, as shown in Figure 2.4, which consumes mW-order power when the wires are pulled down. Assuming 1.2V supply voltage and 1kΩ pull up resistance, average pull up current for both wires is 1.2mA, which results in 1.44mW power wasted just for pull up current, which does not include decoder and driver overhead. This is clearly not acceptable for a sensor platform with μW active power. Therefore, a modified communication protocol is required to meet the stringent

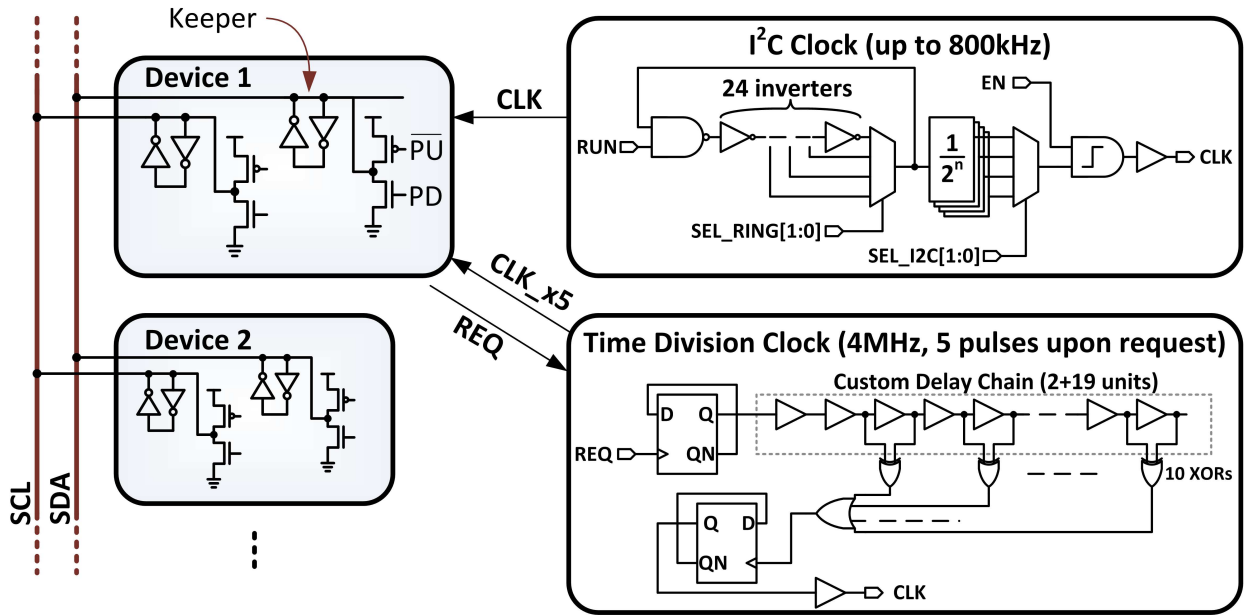


Figure 2.5: Proposed modified I²C circuit diagram

power budget of the system, while maintaining compatibility to standard I²C protocol to enable expansion with I²C compatible devices.

2.3.2 Low Power I²C Implementation

Figure 2.5 shows the circuit diagram for proposed low power I²C protocol. Pull up resistors in conventional I²C act as 1) a pull up device when the wire is at ground potential and pull down is released by attached devices and 2) a keeper for holding supply voltage once the wire is fully charged to supply voltage. To provide the pull up device function without pull up resistor, a SCL-low cycle is divided into five sub-cycles where a master device always pulls up SDA in the second sub-cycle and any attached device can pull down in the fourth, which complies with the I²C standard - SDA can change only when SCL is low and a layer pulling down has the higher priority. Length of the sub-cycle is determined by a sub-cycle clock generator (Figure 2.5) and marginal sub-cycles (first, third, fifth) provide margins for die to die sub-cycle length variations. To provide the keeper function of removed pull up resistor, a keeper is attached to each wire. The proposed low power I²C scheme allows communication between a low power I²C master device and standard I²C slave devices if proper power supply is provided. The only additional cost for such configuration is occasional short circuit current during second sub-cycle, when the

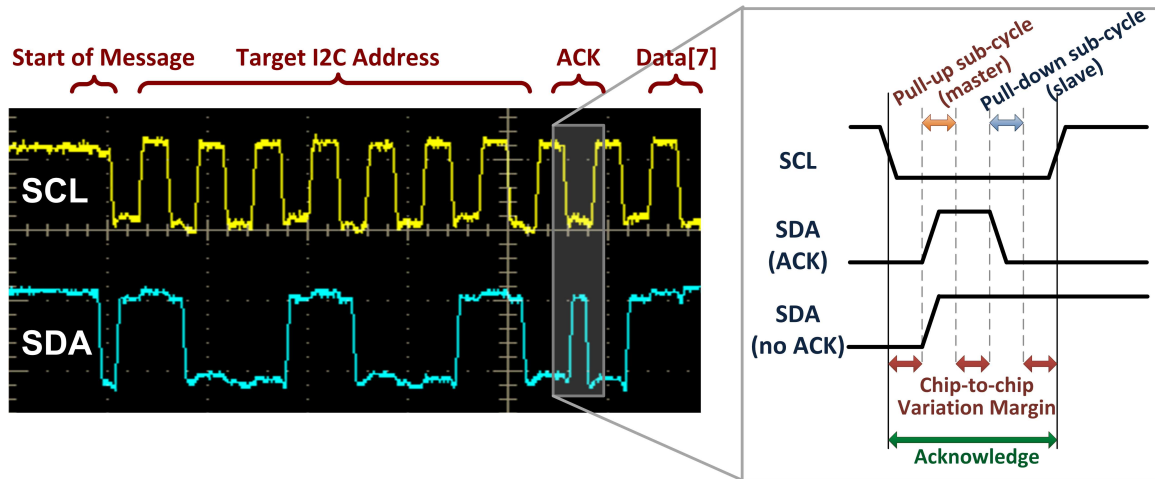


Figure 2.6: Measured I²C waveform and illustration of SCL sub-cycle

low power I²C master device pulls up SDA and the standard I²C slave device pulls down SDA for acknowledge. Energy overhead for such short circuit current would not be significant since it occurs only during one sub-cycle and only for acknowledge operation, which is once every 8 bit transmission. However, to tolerate current surge during this sub-cycle, external power supplies with higher current capacity could be required.

2.3.3 Low Power I²C Measurement Results

Figure 2.6 shows the measured low power I²C waveform. It is clearly shown that in each SCL-low cycle, SDA is raised only in the second sub-cycle and pulled down in the fourth sub-cycle, which is the most clear in acknowledge cycle where both pull up and pull down operation are performed. Measured energy consumption was 88pJ/bit which is more than an order of magnitude lower than 3.6nJ/bit, the theoretical minimum energy needed to drive the wires in standard I²C protocol excluding overhead for decoding and driving logic at 400kbps, which is the maximum data-rate in ‘fast mode’ I²C [19].

2.4 System Operation

Figure 2.7 shows a micrograph of the proposed 1.0mm³ sensor platform. Micrograph of each layer in the system is shown in Figure 2.8. An example system operation is performed as shown

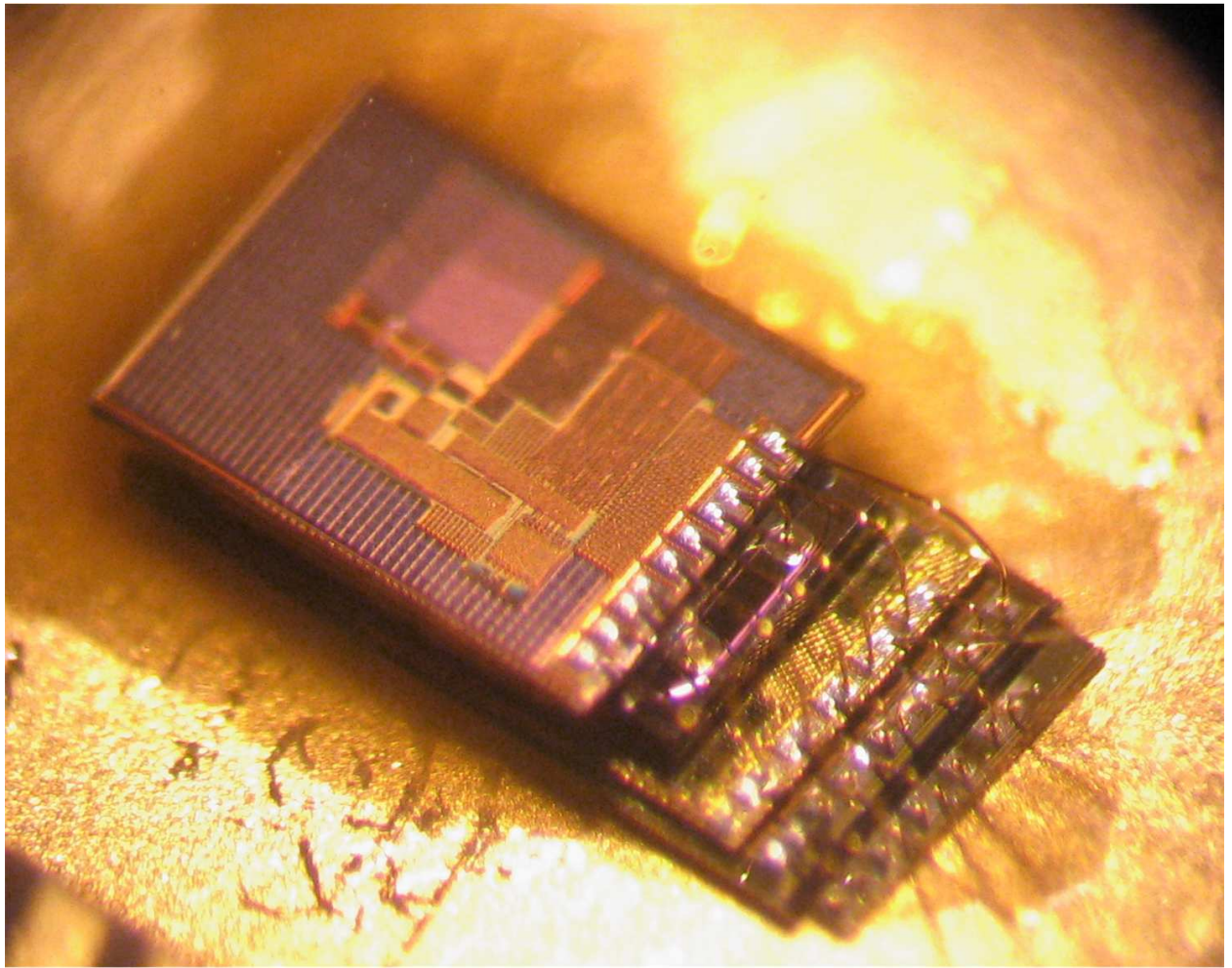


Figure 2.7: Proposed 1mm³ die-stacked sensor platform.

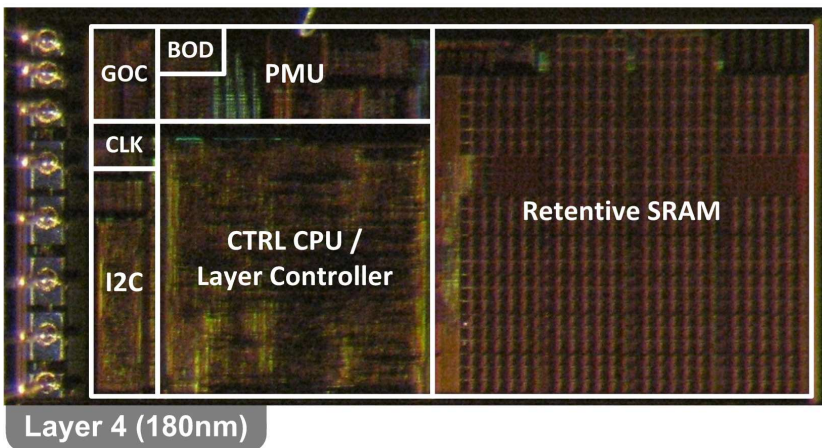
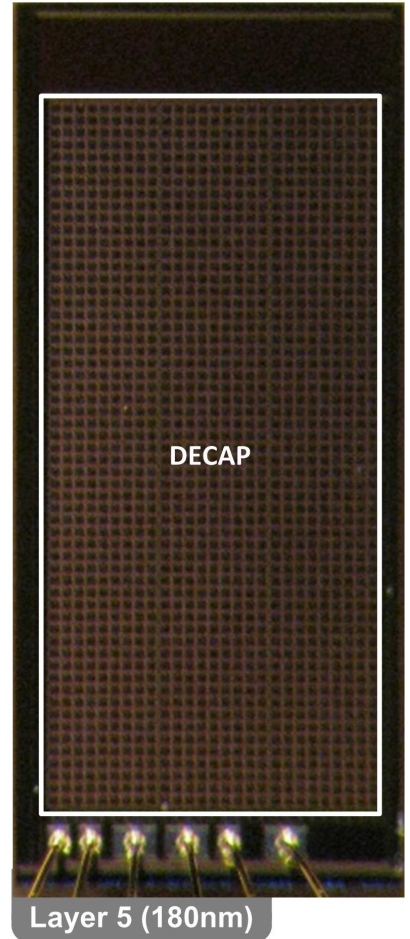
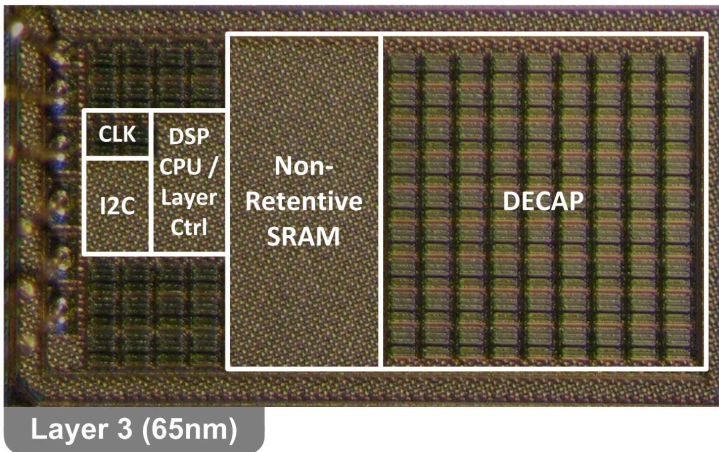
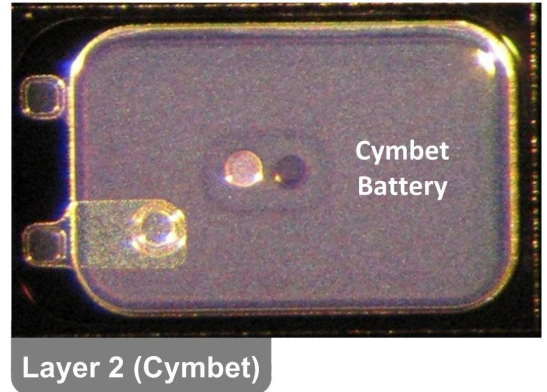
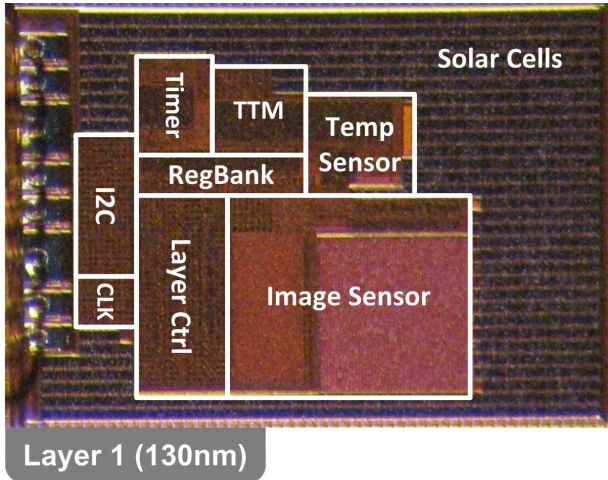


Figure 2.8: Die micrograph of each layer in the 1.0mm³ sensor platform.

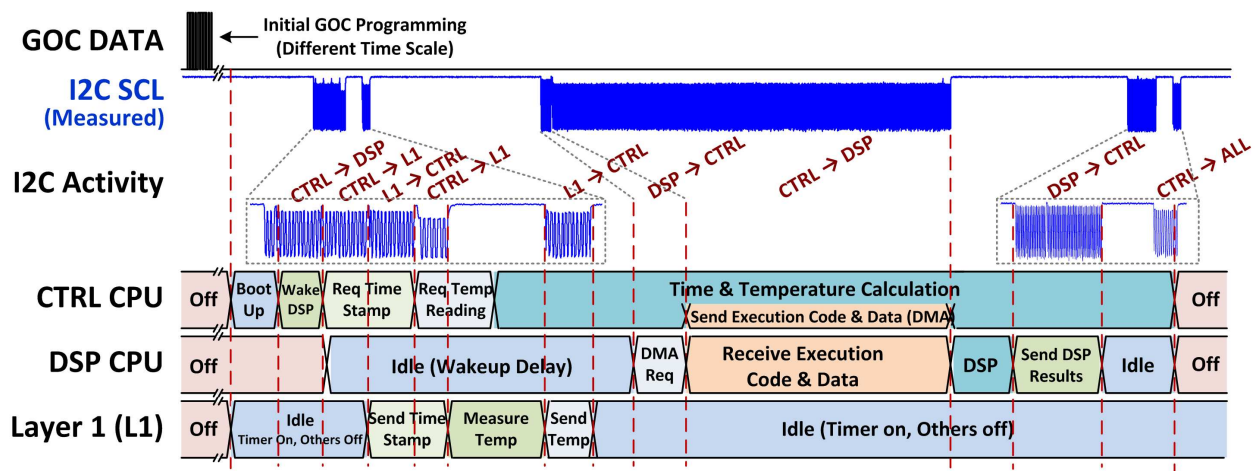


Figure 2.9: An example of sensor platform system operation.

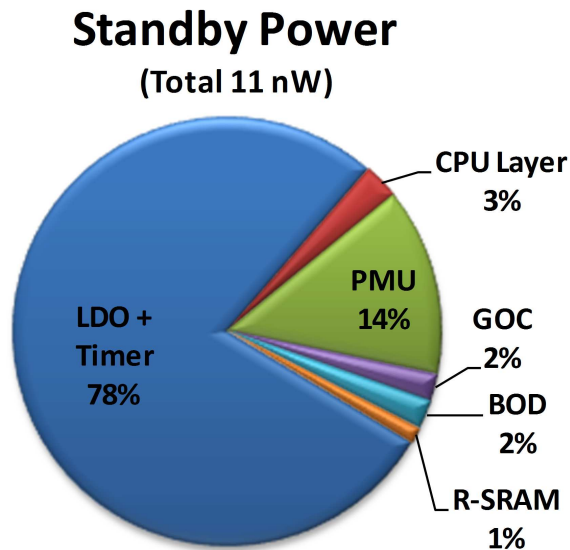


Figure 2.10: Standby power consumption by function unit for the 1.0mm³ sensor platform.

in Figure 2.9. Measured waveform of SCL wire of I²C shows the communication activity among the layers and state transition graph below describes the operation of each layer. After initial shipping from manufacturing, the sensor node is expected to be programmed through GOC as shown with ‘GOC DATA’ waveform in Figure 2.9. Control CPU Layer then initiates boot up sequence and also wakes up DSP (Digital Signal Processing) CPU Layer. Timestamp and temperature measurement request is sent to Layer 1 and results are transferred back to the Control CPU Layer for temperature-calibrated timestamp calculation. Meanwhile, DSP execution code and data is transferred to the DSP CPU Layer while the Control CPU Layer is performing its operation. This way, the Control CPU Layer and the DSP CPU Layer can concurrently process data. When the DSP is complete, result is sent back to the Control CPU Layer to be stored at a retentive memory located at the Control CPU Layer. The Control CPU then put the entire system into standby mode so that the entire system consumes minimum power until the next active operation for periodic sensor measurement. Figure 2.10 shows the power budget of the system in standby mode. Total standby power consumption was 11nW and dominant portion is power consumption for the gate-leakage-based timer [14], which is the only nW-level active unit in the entire system in standby mode to provide accurate timing reference. Without any harvesting, integrated 0.6 μ Ah thin-film battery can support the system in sleep mode for up to 2.3 days. For application where accurate timing is not required, standby power can be reduced down to 2.4nW without timer activation, allowing 10.5 days of sleep mode operation without energy harvesting.

2.5 Conclusion

A 1.0mm³ die-stacked sensor microsystem is demonstrated as a platform for future mm³-scale sensor applications. Its modular die-stacked structure allows easy extension to user-created micro-sensors. Its multi-modal energy harvesting scheme enables harvesting from wide range of energy sources and a battery-voltage dependent power management scheme allows safe operation without battery over-discharge. Low power I²C allows efficient communication among layers, while still leaving possibilities for integration with standard I²C devices with proper power supplies. Optical communication allows energy efficient programming and synchronization of sensor nodes. These circuit techniques together create an ultra-low power sensor platform, encapsulating two micro-

processors, 19kB memory, low power image sensors and timers in 1.0mm^3 volume, creating exciting new opportunities for future mm^3 -scale sensor applications including medical implantations, smart buildings, surveillance system and sensor-based oil well exploration. As a continuation of the Bell's Law, which predicted the emergence of smaller and more powerful microsystems, we believe that the 1.0mm^3 sensor platform is a significant step towards the 'smart dust' vision [17].

CHAPTER 3

Standby Power Reduction for Cubic-Millimeter Sensor Systems

3.1 Introduction

The size of ultra-low power sensor systems is a critical concern, especially for medical applications requiring implantation. Cost, which is related to system volume, is also an important limitation in sensor systems. Since the size of the power source is restricted in such applications, ultra-low power consumption on the order of nanowatts (nW) and picowatts (pW) is required for these sensor processors. One of the most promising approaches to achieving ultra-low power consumption is supply voltage scaling into the subthreshold regime [20] to minimize wake mode energy. However, many sensor systems spend much more time in standby mode than wake mode. Previous approaches have neglected the power consumed in this standby mode despite the fact that standby power can dominate the system budget [21]. Recent work [22] has shown that a better balance between wake mode power and standby mode power can be achieved by designing the system with standby power as a primary constraint. Careful technology selection for balancing active and standby power, stacking high- V_{th} transistors in memory cells for less subthreshold leakage, power gating for less standby power and other architectural/circuit techniques were shown to reduce standby power to tens of pW, giving 1 year lifetime with a 1mm^3 system size including battery. However, even with the sleep strategies presented in [22], standby power is still a dominant (>75%) source of total power consumption. Standby power consists of two components. The first component is the power consumed by circuits that are turned off (power gated) during standby mode. The second component is the power consumed by circuits that must retain state and remain

turned on (e.g., memory). The ratio between these two types of standby power can vary depending on the complexity of logic and amount of memory required, though the second type dominated the standby power in [22]. Therefore, developing different techniques for reducing each type of standby power is the key challenge for extending the lifetime of ultra-low power applications to the multi-year range. However, reducing the standby power for circuits that only consume tens of pW is very challenging for several reasons: 1) the power overhead for using any leakage reduction techniques must be a few pW in order to be beneficial, 2) since these systems are typically battery operated, only a single supply voltage is available, 3) any locally generated voltages for power reduction that are greater than power supply voltage (V_{DD}) or less than the ground voltage, should be controlled without level converters or other switches that introduce new leakage paths. In this chapter, we develop standby power reduction techniques that can be applied to ultra low power processors. First, we explore the use of super cut-off MTCMOS for reducing standby power in power gated blocks. Our key contribution is the development of an ultra-efficient charge pump and cut-off circuit designed for low frequency operation (1-10Hz). Next, we investigate leakage paths in memory and propose a leakage reduction strategy that uses a super cut-off voltage to reduce bit-line leakage. To support charge pump operation, a sub-pW clock generator with a unique current starving scheme is also introduced.

3.2 Standby Power Reduction for Logic Circuits

Large logic blocks in ultra low power processors, such as the CPU, are often power gated to minimize standby power. For such circuits, using super cut-off is a straightforward and effective method for further reducing standby power [23]. In the super cut-off technique a negative voltage is applied to the power gating NMOS footer or a voltage greater than V_{DD} is applied to the PMOS header. However, the power cost of generating this super cut-off voltage has been shown to be large (50nW in [23]) relative to the sub-nW standby power budget targeted in this work. Consequently, the application of this technique becomes challenging in ultra low power processors. To apply the super cut-off strategy to a block with tens of pW standby power, the generation of the super cut-off voltage must have a power overhead on the order of several pW, or 1000X lower than the results presented in [23].

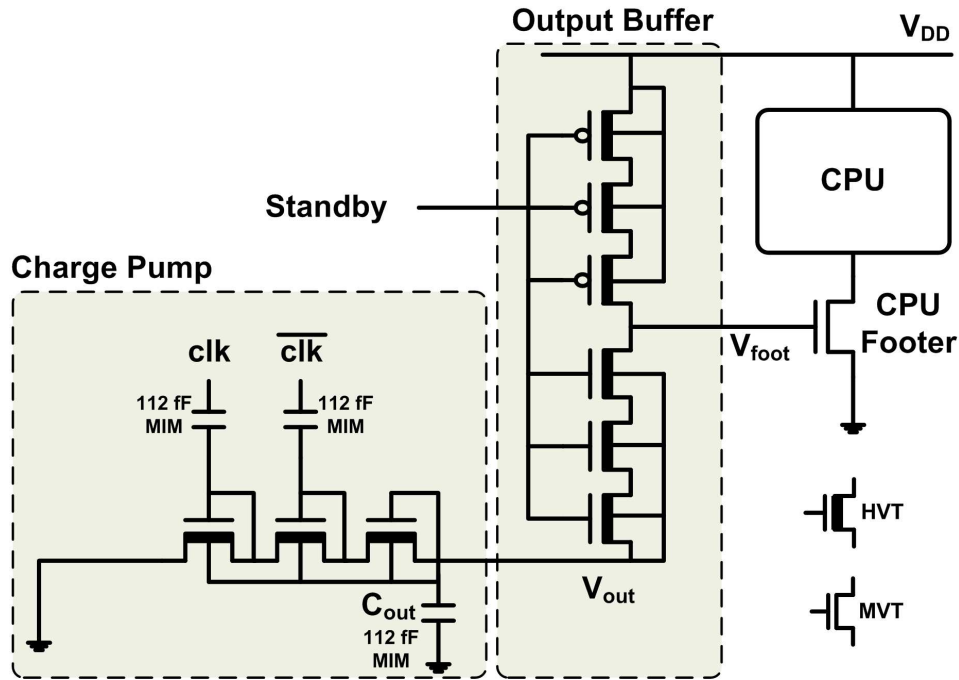


Figure 3.1: Logic circuit standby power reduction by super cut-off.

As shown in Figure 3.1., the proposed system includes a charge pump that generates the super cut-off voltage and an output driver to switch the gate voltage on the footer (V_{foot}) between the super cut-off voltage (V_{out}) in standby mode and V_{DD} in wake mode. The charge pump consists of three high- V_{th} NMOS transistors and three metal-insulator-metal (MIM) capacitors. Two clock signals with opposite phases (to be described further in 3.4) are applied to the pumping capacitors. To ensure maximum power efficiency, the clock must oscillate at the lowest possible frequency, so all leakage paths at V_{out} must be eliminated. Leakage is minimized along the pumping stack by using high- V_{th} devices and by reverse biasing the bodies of the pumping transistors using V_{out} . To further improve power efficiency, a triple stacked inverter is used for connecting V_{out} to the footer. The PMOS stack minimizes subthreshold leakage during standby mode thereby lessening the pumping overhead and the required pumping frequency, while the NMOS stack plays a critical role when switching from standby mode to wake mode. The long NMOS stack cuts the connection between V_{out} and the gate of the footer to eliminate contention between the PMOS stack and the charge pump. It is also crucial to bias the bodies of the entire NMOS stack with V_{out} to ensure that the NMOS stack is not forward biased during wake mode. The negative voltage developed at V_{out} is preserved during wake mode, which is typically very short (on the order of milliseconds) [21],

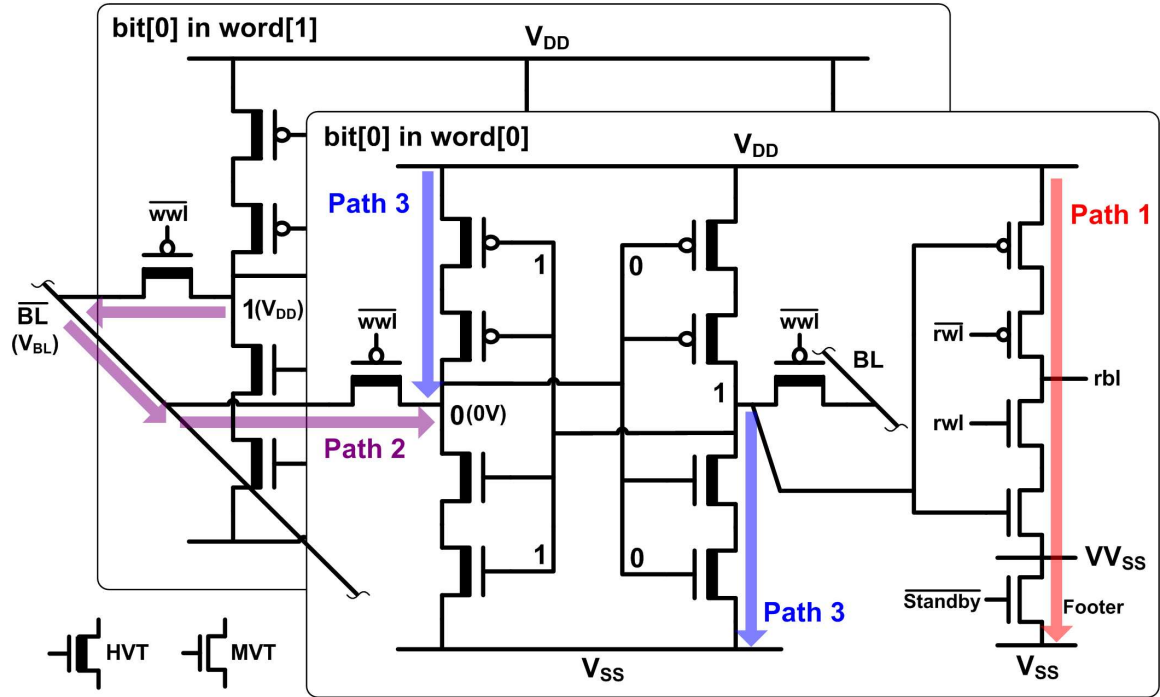


Figure 3.2: Leakage paths in low leakage memory cell.

thus minimizing the time and power overhead of switching back to standby mode. The carefully designed configuration described in this section allows the charge pump to be operated with low clock frequency (<10 Hz) and sub-pW power while guaranteeing sufficiently low (<-150 mV) super cut-off voltage at the output at room temperature (25°C).

3.3 Standby Power Reduction for Memory Circuits

Various SRAM structures, such as the modified-6T [24], 8T [25] and 10T [26] topologies, have been explored for low voltage applications. Despite obvious differences, each of these structures has similar components: a cross-coupled inverter pair, bit-lines, word-lines, access transistors and read buffers. Consequently, we can identify several sources of leakage that are common across all structures. To explore standby power reduction for memory, we study the low-leakage memory cell proposed in [22]. Given the general similarities between various SRAM structures, many of the conclusions in this work may be extended to other cells. As depicted in Figure 3.2, the memory cell under investigation uses cross-coupled inverters with stacked high- V_{th} transistors to minimize

the subthreshold leakage. A separate read buffer with medium- V_{th} transistors is used to boost the read performance and improve cell stability at low voltage.

3.3.1 Leakage Reduction for Power Gated Blocks

Figure 3.2 shows the most important leakage paths within and between memory cells. Path 1 is the leakage path for circuits that are power gated (i.e., turned off) during standby mode. Only the read buffer is shown in Figure 3.2, but this category of circuits also includes memory peripherals such as row/column decoders, bit-line drivers and other control logic. Since these circuits are all turned off by a footer, our analysis shows that Path 1 contributes only $\sim 2\%$ of the total standby power. A separate power gating transistor is used to ensure that the current drawn from other power intensive modules, such as the CPU, does not induce read/write errors during wake mode. However, the super cut-off voltage that is generated by the charge pump introduced in 3.2 can be shared with virtually no power overhead.

3.3.2 Bit-line Leakage Reduction

Path 2 in Figure 3.2 shows the bit-line leakage path in the array structure of the memory. During standby mode, the bit-lines (BL and \overline{BL} in Figure 3.2) float to some intermediate voltage, V_{BL} , between 0 and V_{DD} . The value of V_{BL} depends on the number of bit cells storing 0's and 1's in the bit-line column. As a result, the transistors that connect the bit-lines and the memory cell (pass transistors) will have a drain-source voltage of V_{BL} or $V_{DD} - V_{BL}$ when the cell stores 0 or 1 in the adjacent node, respectively. This drain-source voltage induces subthreshold leakage on the bit-line, which contributes 50% of total standby leakage. In order to reduce the bit-line leakage, a super cut-off voltage ($> V_{DD}$) can be applied to the gate of the pass transistors during standby mode. This can be achieved by using a charge pump to boost the power supply for the wordline driver connected to the pass transistor control. The basic concept of this strategy is similar to the strategy used with power gated logic blocks, but it raises the following new challenges: 1) a new power supply for the pass transistor control logic must be kept near V_{DD} or higher at all times since low voltage at the gate of pass transistors will turn on the transistors, resulting in data loss, 2) the new power supply should be able to supply enough current to meet the demands of the pass

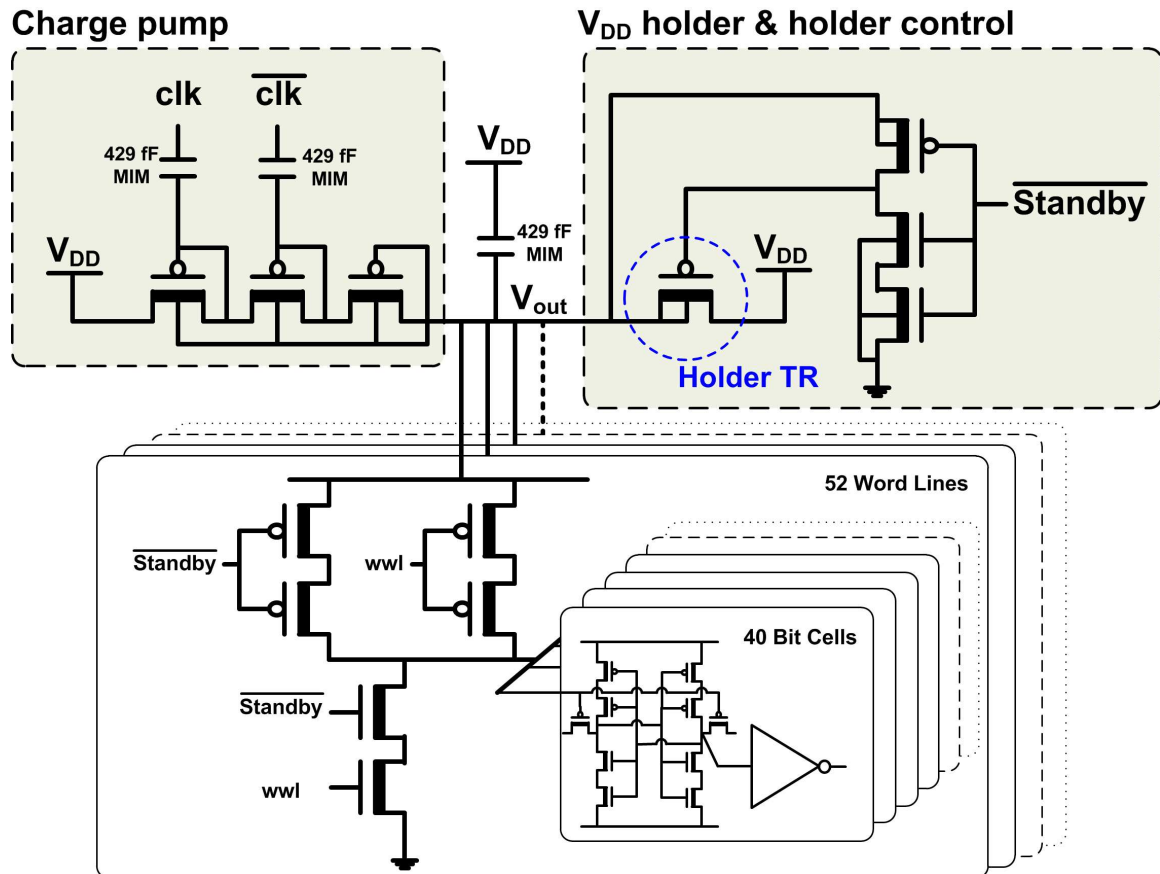


Figure 3.3: Proposed circuit for bit-line leakage reduction.

transistor control logic during a memory write operation, and 3) all these criteria should be met with a power budget on the order of pW.

The proposed circuit that meets these criteria is presented in Figure 3.3. An ultra-low power charge pump similar to the one presented in the previous section is used for boosting the power supply. PMOS transistors are used to generate a positive super cut-off voltage $V_{out} (> V_{DD})$. The output of this charge pump is tied to the power rail of the wordline drivers. Charge is continuously pumped into the output capacitor (C_{out}) to develop V_{out} . The wordline drivers are structured to always provide full V_{out} in standby mode while also enabling wordline control during the wake mode. However, there can be no direct connection to the power supply at the output node during wake mode because a direct connection to V_{DD} would prevent V_{out} from rising higher than V_{DD} in standby mode. As a result, write operations that lead to a transition at the output of the wordline drivers will consume the charge stored in C_{out} , thereby lowering V_{out} . Therefore, consecutive

write operations that occur between pumping cycles (due to the low pumping frequency) may bring V_{out} below V_{DD} . As the voltage reduces, the pass transistors of memory cells will be turned on, resulting in data loss. To prevent this data loss, a 'holder' transistor is introduced. The holder transistor indirectly connects V_{DD} with the output of the charge pump and is turned on during wake mode. When V_{out} drops below V_{DD} , the holder transistor is forward biased and can effectively 'hold' V_{out} near V_{DD} . A wide low- V_{th} transistor would be preferable for the holder transistor, but in standby mode, the holder transistor acts as a direct leakage path from the output of the charge pump to V_{DD} , thereby reducing pump efficiency. Thus, a moderately sized (W:0.55 μ m L:0.35 μ m) high- V_{th} transistor is chosen to alleviate this side effect. Worst case simulations show that this configuration maintains $V_{out} > 489\text{mV}$ at $V_{DD} = 0.5\text{V}$.

3.3.3 Intra-cell Leakage

Finally, Path 3 in Figure 3.2 shows the intra-cell subthreshold leakage path. In each cell, the primary leakage paths include a single NMOS stack and a single PMOS stack. For example, with a bit value of 1 stored in the front memory cell in Figure 3.2, the top left PMOS stack and bottom right NMOS stack will leak. Our analysis shows that this leakage amounts to 48% of total standby power. In order to suppress intra-cell subthreshold leakage, a reverse body bias can be applied to all transistors or high V_{th} transistors can be used. However, according to our analysis, the standby power of our target memory module was 60.5pW and the overhead of generating enough well bias current to compensate for junction leakage was greater than the projected leakage improvement. Therefore, our memory structure uses high- V_{th} transistors as in [22].

3.4 Ultra-Low Power Clock Generation

The clock generator is one of the most important elements in our proposed ultra-low leakage system. Without proper design, the clock generator can easily exceed the pW budget allotted. Figure 3.4 illustrates the proposed clock generator with a unique current starved inverter. In this inverter, current starved transistors are placed next to the output node whereas conventional design places them next to the power and ground rails. To achieve minimum power, the clock generator is

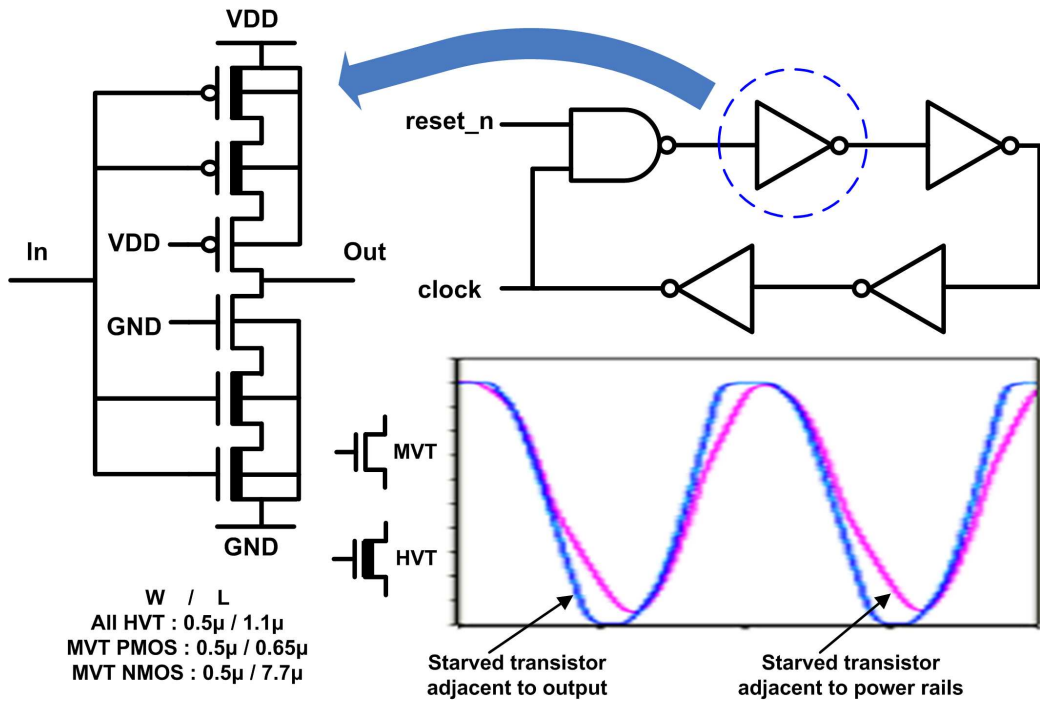


Figure 3.4: pW clock generator with current starved transistors and output waveform comparison between different starved transistor placement schemes.

designed for operation at very low frequencies (1 10Hz). Each inverter in the clock generator uses stacked high- V_{th} transistors adjacent to the power and ground rails and current-starved medium- V_{th} transistors in the off-state adjacent to the output node. In this configuration, the on-current of the inverter is determined by the subthreshold leakage of the starved medium- V_{th} transistors, which makes the current consumption very small. When the input is low, the NMOS stack is turned off and a small voltage is developed at the source of the starved NMOS due to stack effect. Thus, a reverse body bias is generated for the starved NMOS, making the off-current smaller and thereby improving the power efficiency over the case where the starved transistors are adjacent to the power and ground. The same effect can be observed in the PMOS stack. Our analysis shows that the 10%-90% rise/fall time can be reduced by 19.6% with our proposed design, making the clock generator more stable and robust.

3.5 Measurement Results

3.5.1 Logic Circuits

A large CPU block with 23,472 transistors has been tested using 4 different medium-V_{th} footer sizes at room temperature (25°C). Figure 3.5 shows the generated super cut-off voltage and charge pump power consumption as functions of the charge pump clock frequency. The charge pump clock was supplied externally in this specific experiment to give maximum tunability. Strong super cut-off voltages (<-150mV) are generated with low pumping frequency (<10Hz) and sub-pW power consumption. The leakage reduction achieved using super cut-off MTCMOS is shown in Figure 3.6. With a footer width of 17.16 μ m, the CPU block consumes 15.4pW in standby mode without super cut-off MTCMOS. For low pumping frequencies (<10Hz), increasing the pumping frequency reduces total standby power since the super cut-off voltage reduces. However, as frequency exceeds 10Hz, the charge pump overhead becomes dominant and increases total power consumption. Total standby power reaches a minimum of 0.8pW at 10Hz, a 19.3X reduction over normal operation.

Figure 3.7 shows the standby power reduction for different footer sizes. Despite different footer sizes, the standby power converges to 1pW for all cases at an optimal pumping frequency of 10Hz. Therefore, the power gain is largest (19.3X) with the widest footer and smallest (2.3X) with the narrowest, which suggests that this power reduction technique may also enable active power reduction by allowing more freedom when choosing the size of the power gating transistor. The size of the power gating transistor is constrained by the standby mode power budget and wake mode current demand. In wake mode, a wider power gating transistor is preferred to minimize the voltage drop across the power gating transistor. However, since the standby power of a circuit block is determined by the size of the power gating transistor, narrow width is preferred for minimum standby power. Energy consumption in standby mode dominates wake mode energy consumption for ultra-low power processors, so a power gating transistor with very narrow width is typically used (a footer width of only 0.66 μ m was used in [22]). The voltage drop across such a narrow power gating transistor effectively reduces VDD for the logic, making the circuit block slower, less robust and less energy efficient. In light of our measured results, a wider power gating transistor can be used with a minor standby power penalty and significant wake mode energy reduction

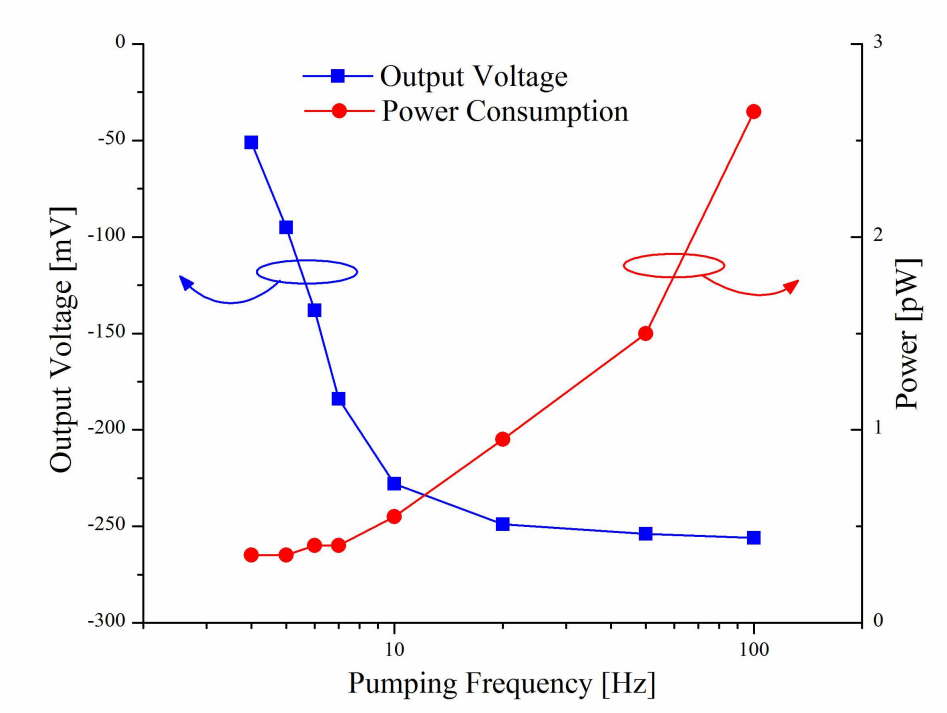


Figure 3.5: Generated super cut-off voltage and power consumption of charge pump for logic circuit standby power reduction.

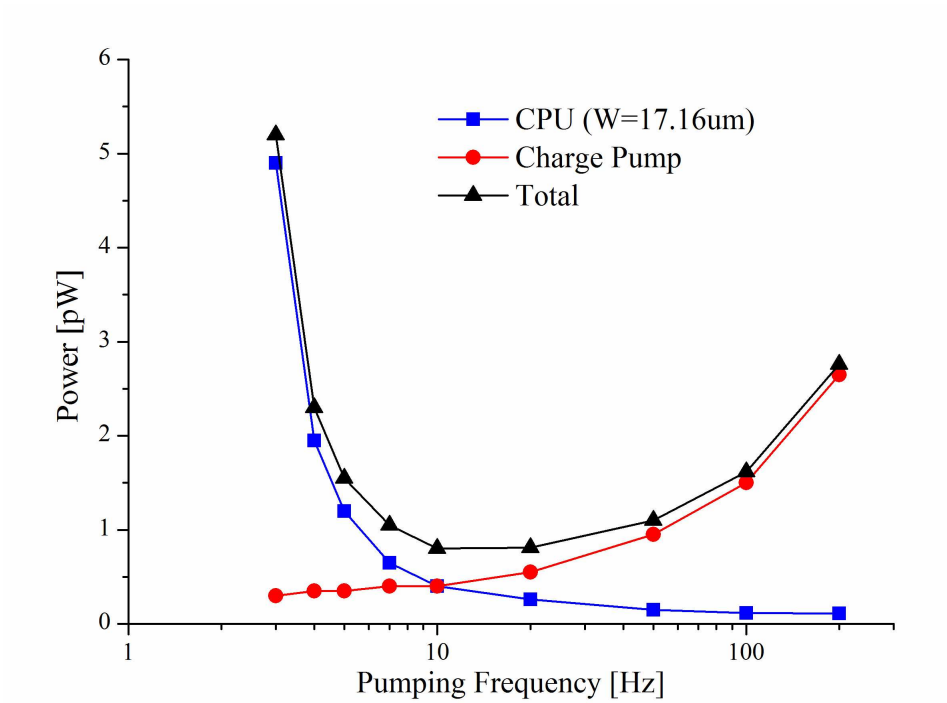


Figure 3.6: CPU leakage and charge pump operation power in standby mode.

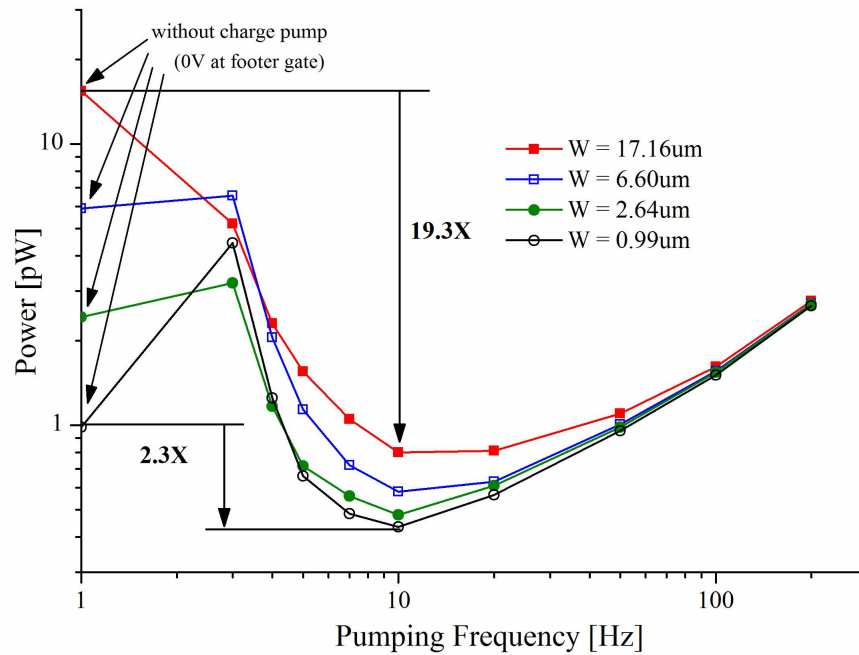


Figure 3.7: Total standby power of CPU and charge pump with various CPU footer sizing.

(estimated at 23% by eliminating 116mV out of 500mV)

3.5.2 Memory

A memory with 2,720 bit cells has been tested at room temperature (25°C). Figure 3.8 shows the generated super cut-off voltage and charge pump power consumption as functions of charge pump clock frequency. The power overhead for the charge pump is significantly higher than for the previous section due to the larger number of leakage paths such as the pass-transistor controllers and the holder transistor. At the power optimal pumping frequency of 20 Hz, the charge pump overhead is below 5% of original memory standby power. Total standby power is shown in Figure 3.9. At a pumping frequency of 20 Hz, standby power is reduced by 29.1% compared to normal operation. Note that power actually increases at low frequencies since the output of the charge pump can fall below V_{DD} (0.5V) in this region and cause increased leakage across pass transistors.

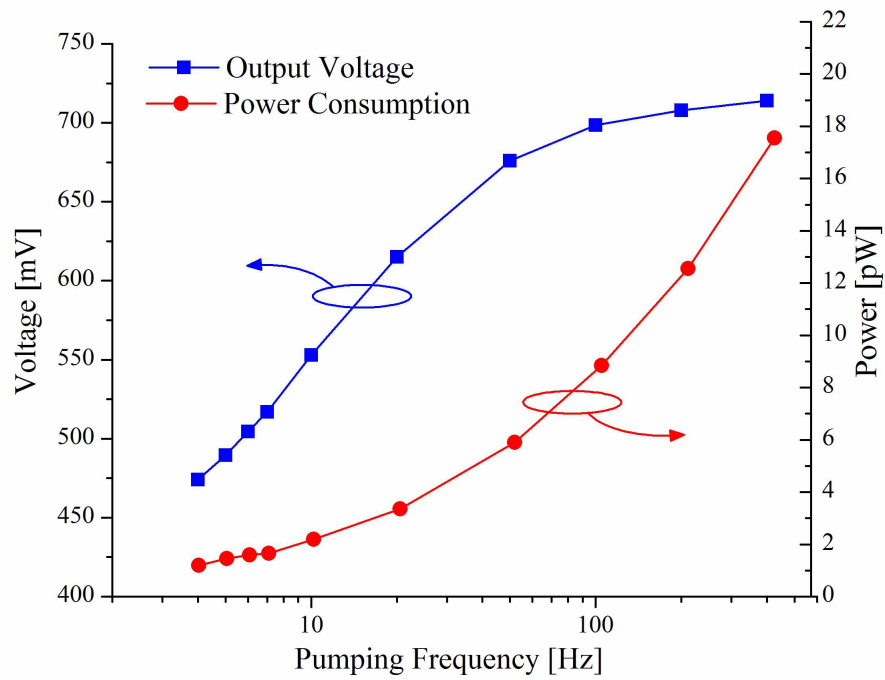


Figure 3.8: Generated super cut-off voltage and power consumption of charge pump for memory standby power reduction.

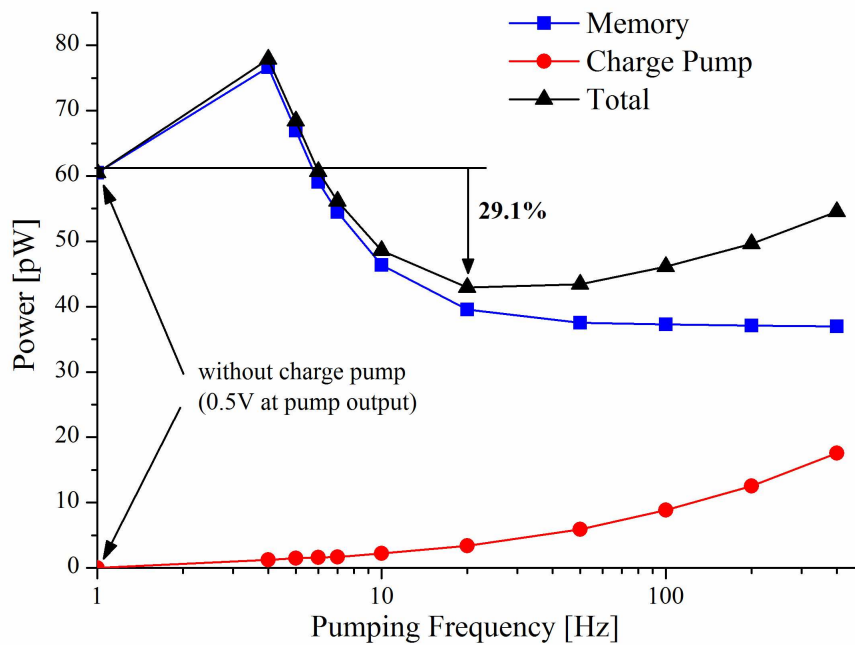


Figure 3.9: Memory leakage and charge pump operation power in standby mode.

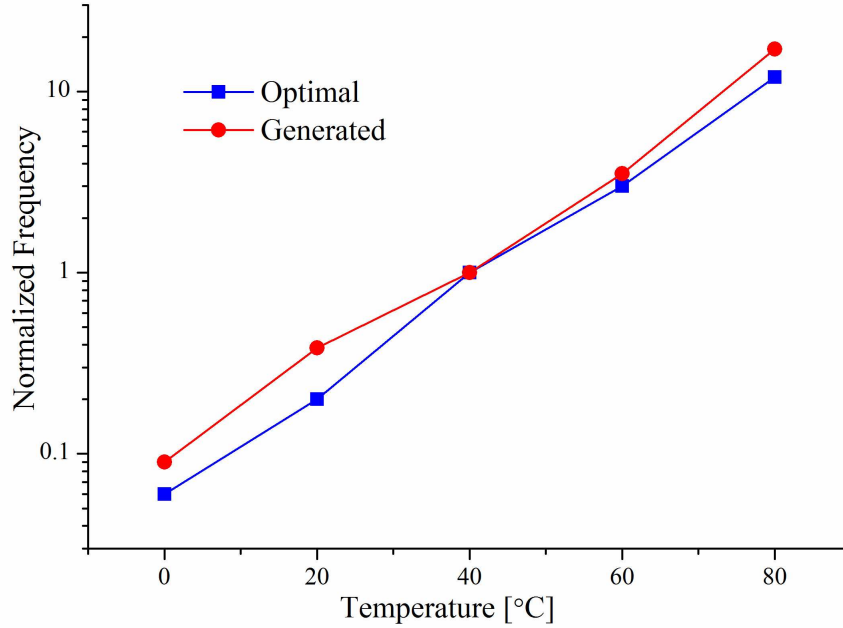


Figure 3.10: Optimal and generated clock frequency normalized at 40°C.

3.5.3 Ultra-low Power Clock Generation

Testing of the low power clock proposed in 3.4 shows an average oscillating frequency of 4.6 Hz with a power consumption of only 0.64pW. Simple calculations suggest that, at the optimal frequency for the two previously described charge pumps (10Hz, 20Hz), clock power can be maintained below 3pW. Since the power characteristic in Figure 3.6 is flat near the minimum, applying the memory-optimal clock frequency of 20Hz to the CPU charge pump results in a negligible power penalty of only 1.3%. This result suggests that a single clock generator can be shared between the memory and CPU. Measurements at temperatures ranging from 0-80°C reveal that the low power clock tracks the power optimal frequency well. Figure 3.10 shows the power optimal charge pump clock frequency for CPU and generated frequency, both normalized at 40°C. Over this temperature range, discrepancies between the optimal frequency and the generated frequency result in a maximum power penalty of only 14% compared to the optimal operation point

3.6 Conclusion

Super cut-off circuit techniques for reducing the standby power of ultra-low power processors have been presented along with a supporting low power clock generator. A standby power reduction of 2.3-19.3X is achieved for power gated logic blocks, while standby power is reduced by 29.1% for memory using the proposed techniques.

CHAPTER 4

2T Dual V_{th} Gain Cell eDRAM for Cubic-Millimeter Sensor Systems

4.1 Introduction

Battery-operated ultra-small sensing systems have wide applications ranging from implantable medical devices to pervasive environmental monitors. With limitations on battery size, these systems are severely energy constrained; therefore, managing power consumption is of critical concern for reasonable lifetime. Recent work [6] has shown that retentive memory dominates the power budget for such systems, making low leakage memory design indispensable [27]. An ultra-low leakage SRAM was proposed to mitigate this issue at the cost of a large area penalty ($1230F^2$) [6]. Flash memory can also serve as retentive memory and offers near-zero standby power, but it requires additional cost for process/masks and charge pumps, and also incurs very large write power that quickly dominates total sensor power consumption. In this section, a logic-compatible embedded DRAM (eDRAM) with a 2T dual- V_{th} gain cell is proposed, which has $12\times$ smaller cell area than a previously proposed ultra-low leakage SRAM [6] and 5 lower retention power than the best previously reported low-power eDRAM [28]. Conventional eDRAM designs [28][29] are optimized for read/write (R/W) speed at the cost of retention power and hence far exceed the power and performance requirements of typical sensor applications. Instead, the proposed design intentionally exploits the low processor speeds of sensor nodes (commonly 0.1-1MHz) to drastically reduce the retention power of eDRAM, which is dominant in these systems due to long standby

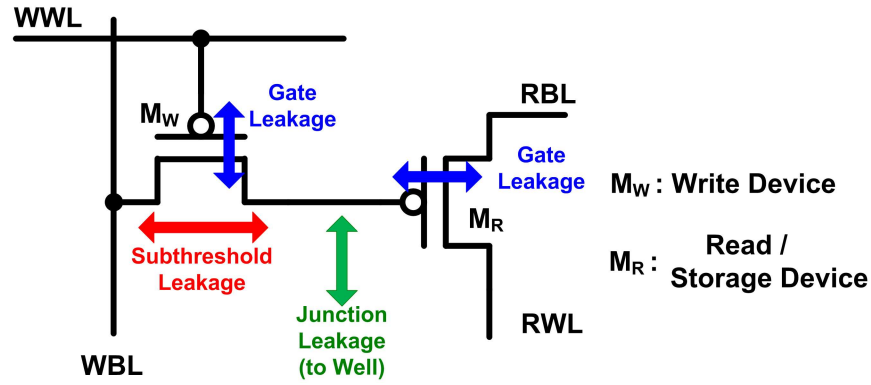


Figure 4.1: Structure of conventional 2T eDRAM cell and its three types of leakages that can destroy stored data.

times. Among the various eDRAM bit cells, a 2T eDRAM is used because of its small cell area [30]. Using a novel dual- V_{th} approach, retention time is increased by $8\times$ without an explicit capacitor in the cell. The proposed 2T dual- V_t gain cell-based eDRAM is implemented in 180nm CMOS technology at 0.75V, which provides an optimal tradeoff between standby and active mode power for ultra-small sensor systems [31]. At cubic millimeter volumes, even the relatively small memory sizes of these sensor systems (as low as several kb) can be a large fraction of system size. Hence, the area overhead of sense amplifiers is difficult to amortize over the small number of bits per bitline. A single inverter sensing scheme is proposed to greatly reduce the sense amplifier area overhead and achieve high array efficiency for an 8×2 kb array, reducing overall sensor node size and cost.

4.2 2T eDRAM Gain Cell

4.2.1 Conventional 2T Gain Cell

Figure 4.1 shows the conventional 2T eDRAM cell structure and three types of leakage current that can destroy the data stored in the cell during retention, namely gate leakage, junction leakage and subthreshold leakage. With the 180nm technology, gate leakage current is negligible compared to the subthreshold leakage due to its thick gate oxide. Since the junction leakage is also negligible compared to subthreshold leakage, reducing subthreshold leakage in the 2T cell is key to extending its refresh cycle time. This will also reduce overall retention power with less frequent refresh

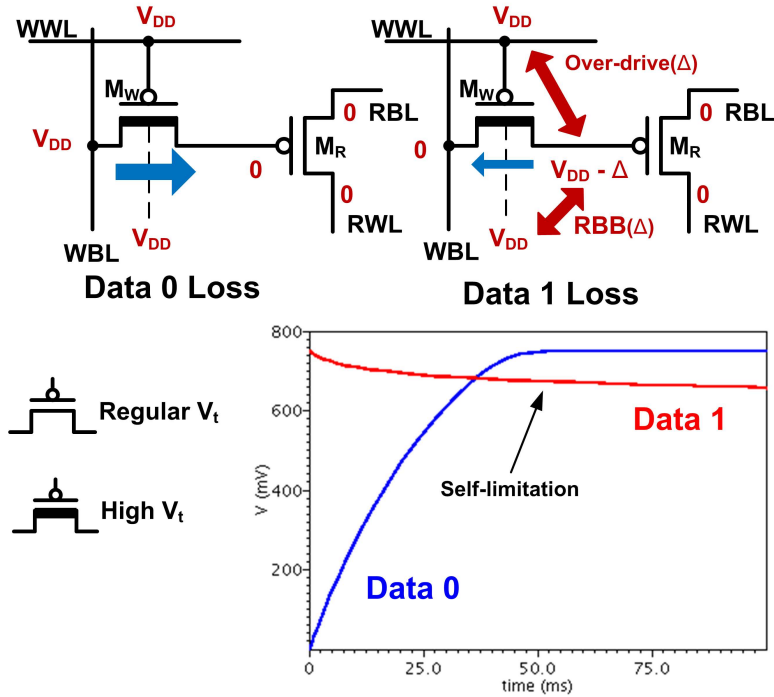


Figure 4.2: 2T dual- V_{th} eDRAM structure and possible data loss scenarios. Since data 1 loss is protected better than data 0 loss, the preferred state of the bit line for minimizing cell decay in standby time is 0.

operation. In the 2T cell, threshold voltage (V_{th}) of the write device (M_W) is bounded by the overall system speed since it has a direct effect on write speed. With loose constraints on system speed in typical sensor node processors, a high V_{th} transistor can be used as a write device to drastically reduce subthreshold leakage while maintaining reasonable sufficient write speed.

4.2.2 2T Dual- V_{th} Gain Cell

Figure 4.2 shows the structure of the 2T PMOS dual- V_{th} cell and two possible data loss scenarios. The write device (M_W) is a minimum length thick-oxide high- V_{th} transistor and the read/storage device (M_R) is a minimum length standard- V_{th} transistor. Given that gate oxide leakage of PMOS with $V_G \sim 0.75V$ is negligible in this technology, data written on the storage device is predominantly lost through subthreshold leakage in two different scenarios as shown in Figure 4.2. In the first scenario, when data 1 is stored with write bitline (WBL) grounded, charge leakage to WBL can destroy data 1. However, this subthreshold leakage is self-limited [29]: as the stored voltage decays by Δ , M_W is both super cut-off and reverse-body biased (RBB) by Δ suppressing

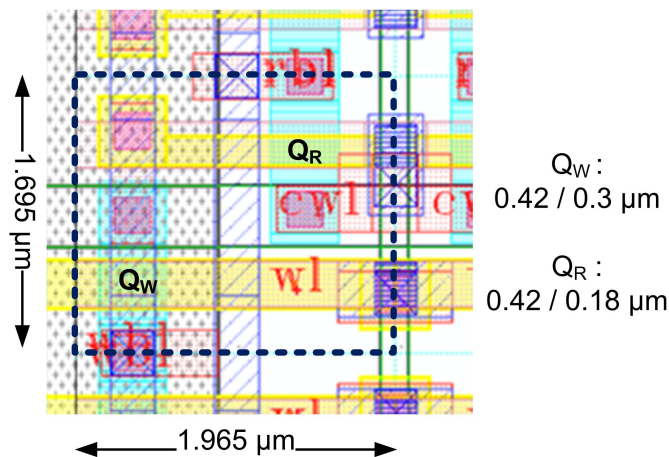


Figure 4.3: Layout of the 2T dual- V_{th} eDRAM cell and its dimensions.

leakage harder. In the second scenario, when data 0 is stored with write wordline (WWL) disabled (high), subthreshold leakage can raise the stored 0. In this case there is no super cut-off or RBB condition, however this scenario is largely avoided by employing a ground pre-charge scheme for WBL during write operations and idle time. In both cases, the high- V_{th} write transistor reduces subthreshold leakage by more than 2 orders of magnitude at the cost of slow write times of up to 30ns at 85°C and 1s at 25°C, which meets typical sensor node operating frequencies of ≤ 1 MHz. To aid in writing 0's with a PMOS pass transistor, the gate of M_W is driven to a negative voltage (-550mV) which is common in 2T gain cell design [29][30].

The layout of the cell is shown in Figure 4.3 with an area of $103F^2$ ($3.33 \mu\text{m}^2$ in 180nm process), which is 28% smaller than a push-rule 6T SRAM in this technology. The cell is made with logic design rules that impose a spacing requirement between thick-oxide high- V_{th} and regular- V_{th} devices. Therefore cell area is 56% larger than a previously demonstrated high-performance 2T gain cell [29], but 30% smaller than a recently proposed long retention time 3T cell [28] after normalizing to process technology.

4.3 Area-Efficient Single Inverter Sensing

Figure 4.4 shows the proposed single inverter sensing scheme with simulated waveforms. The read wordline (RWL) and read bitline (RBL) connected to the PMOS read transistor are both pre-

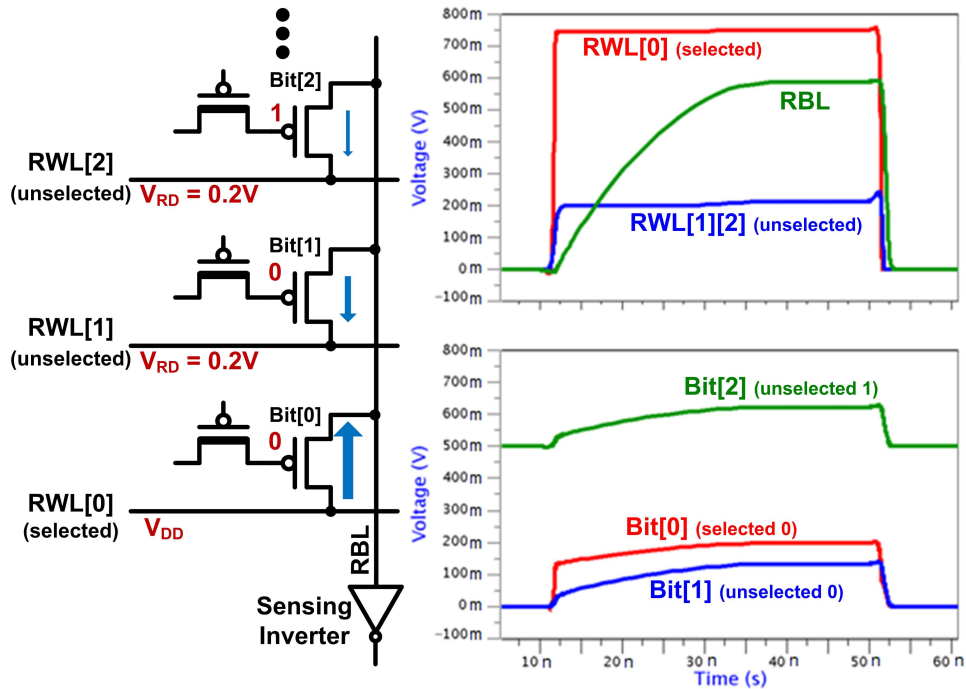


Figure 4.4: Single inverter sensing scheme and simulated waveforms with 32bits/bitline.

discharged to ground for a read operation. As RWL of the selected word is raised to V_{DD} , the read transistor of the selected row either charges up or holds the RBL voltage low, depending on the value the cell stores. When the cell stores a 1, the RBL voltage remains low since the read transistor of the selected word is turned off with a stored 1, and all the other read transistors connected to the RBL can only leak to pre-discharged RWLs. When the cell stores a 0, current flowing from the selected RWL to RBL will charge up the capacitance of the RBL. As the RBL voltage increases, the read transistors in other cells storing 0 begin to leak charge from RBL to unselected RWLs. However, the bodies of these transistors are tied to V_{DD} , which leads to reverse body bias (and low leakage) for low RBL voltages and allows the selected cell to pull the RBL voltage up sufficiently to flip the inverter attached to the RBL. A small positive voltage ($V_{RD}=0.2V$) is applied to unselected RWLs instead of ground, which 1) accelerates initial voltage development on RBL and 2) couples up the storage node voltage of unselected cells (Bit[2] in Figure 4.4) to reduce unwanted charge leakage to RWL after high RBL voltage development, improving read 0 margin.

The implemented 16kb array consists of eight 2kb banks with 32 rows \times 64 columns (Figure 4.5). The number of bits per bitline (32) is chosen to demonstrate high array efficiency with low bits per bitline and reduce the overhead for driving the unselected RWL to V_{RD} by only raising

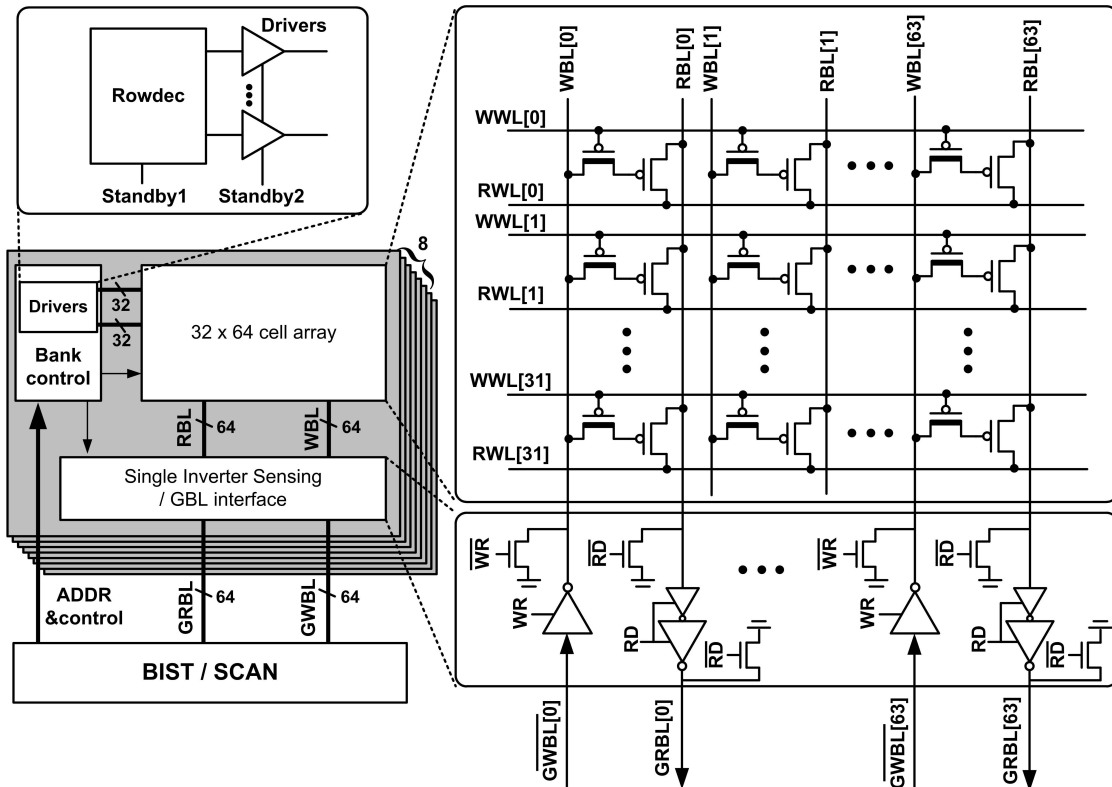


Figure 4.5: Block diagram of implemented eDRAM array.

RWLs of the selected bank. This configuration also helps avoiding the data 0 loss scenario in Figure 4.1. With the area-efficient single inverter sensing scheme, an array efficiency of 58% is maintained despite the small number of bits per bitline (32) and small size (2kb). In contrast, an array efficiency of only 28% could be expected if the array is paired with a conventional sense amplifier design. A standby mode is employed for all decoders/peripherals where WWL is gated such that it remains at V_{DD} to maintain the data in cells while WBL/RBL/RWL are grounded to achieve minimum leakage, as discussed in the previous section. Transitions from standby to active mode and vice versa are completed in 400ns, which is within a clock cycle for typical low power sensor systems, and in two stages (standby1/2 signals in Figure 4.5) so that the voltages on WLs and BLs can be hold stably during transition.

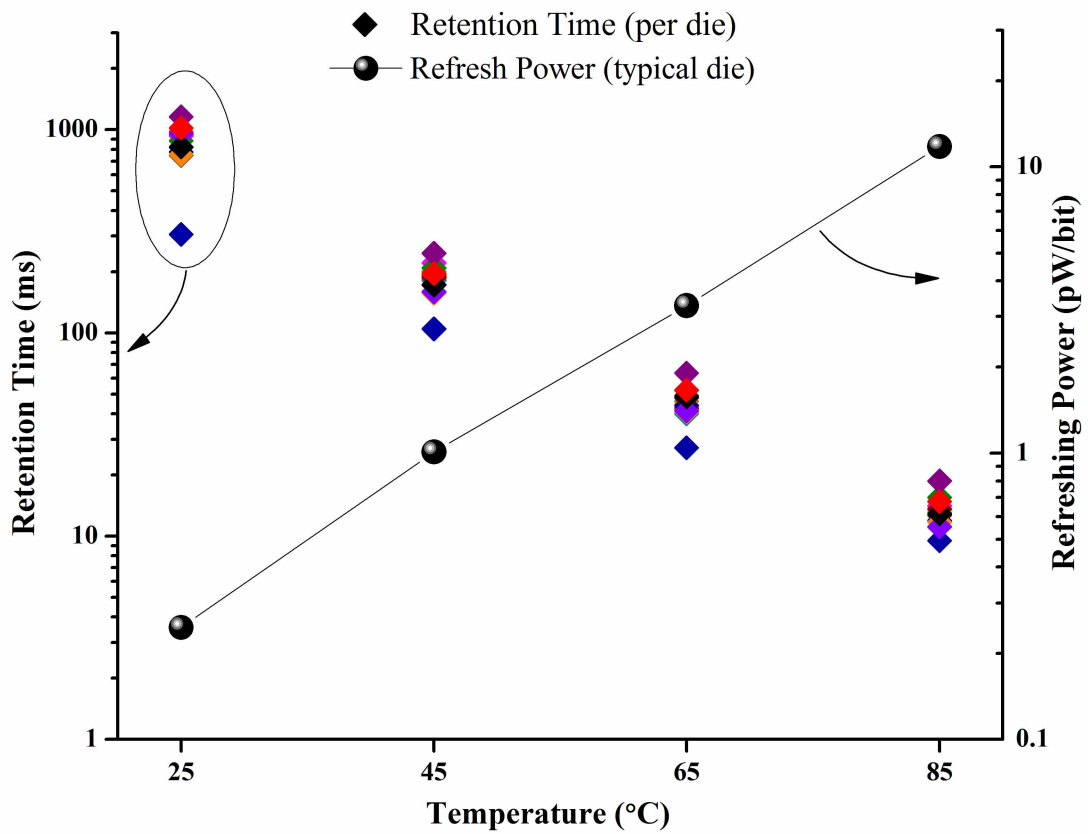


Figure 4.6: Measured retention time for 16 dies and refreshing power for typical die versus temperature.

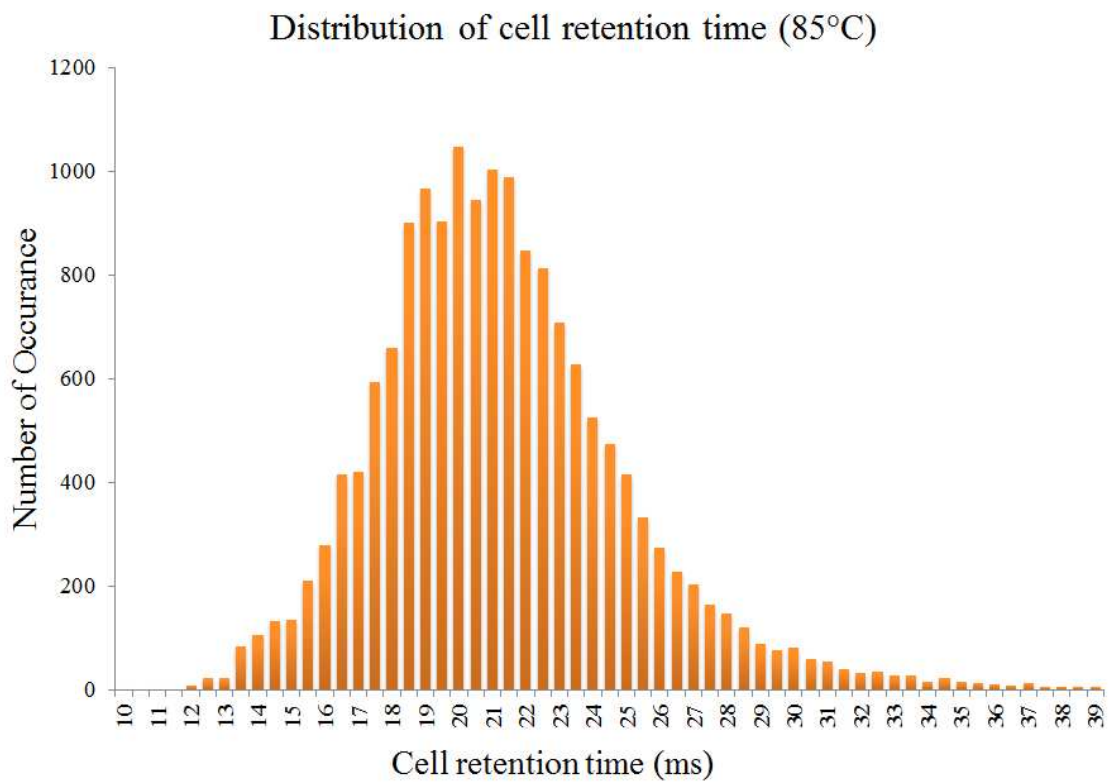


Figure 4.7: Distribution of measured cell retention time in typical die.

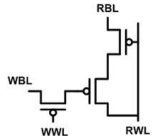
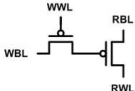
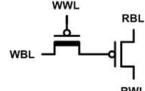
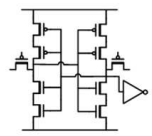
	Boosted 3T	Gain Cell 2T	Proposed 2T Dual V_t	Ultra-low Power SRAM
Cell Structure				
Cell Size	148 F ²	66 F ²	103 F ²	1230 F ²
Target V _{DD}	0.9V	1.1V	0.75V	0.5V
Retention Time	(worst die) 1.25ms @ 85°C	10μs @ 25°C	(worst die) 306 ms @ 25°C 9.5 ms @ 85°C	-
Retention Power	87.1pW/bit @85°C	Not reported	662fW/bit @25°C 16pW/bit @85°C	11fW/bit @25°C

Table 4.1: Comparison with other state-of-the art eDRAM and low power memories: Boosted 3T [28], Gain Cell 2T [29], Ultra-low Power SRAM [6]

4.4 Measurement Results

Figure 4.6 shows the measured worst-case retention time for the fabricated prototype of 9.5ms at 85°C and 306ms at 25°C with refresh power of 16pW/bit and 662fW/bit, respectively. Refresh power is measured for an average performing die refreshing with 10% margin applied to the measured retention time of the worst die. Note that the target application space of ultra-low power sensor systems tend to experience lower temperature ranges than high performance ICs, making realistic retention times for the proposed eDRAM in the 100ms range. Figure 4.7 shows the distribution of cell retention time in 16kb macro in a typical die at 85°C, with an average of 21ms with a standard deviation of 4ms. For this die, minimum retention time was 12ms; this could be further lengthened to 12.5ms with 99.9% bit yield.

A comparison between this work and other recent eDRAM/low power memory designs is provided in Table 1 and Figure 4.8. Table 1 shows that the proposed eDRAM achieves 7.6× longer retention time than any of the current eDRAM designs and with 12× smaller area than the lowest power SRAM reported. With recently published eDRAM designs, a large number of bits per bitline ranging from 128 to 1024 is needed to maintain a reasonable array efficiency around 60% as shown in Figure 4.8. The proposed eDRAM maintains comparable array efficiency with many fewer bits per bitline, down to 32, allowing efficient implementation of the smaller arrays that are common in sensor applications.

Performance of the 2T dual V_t eDRAM is summarized in Table 2. Measured read delay (RWL

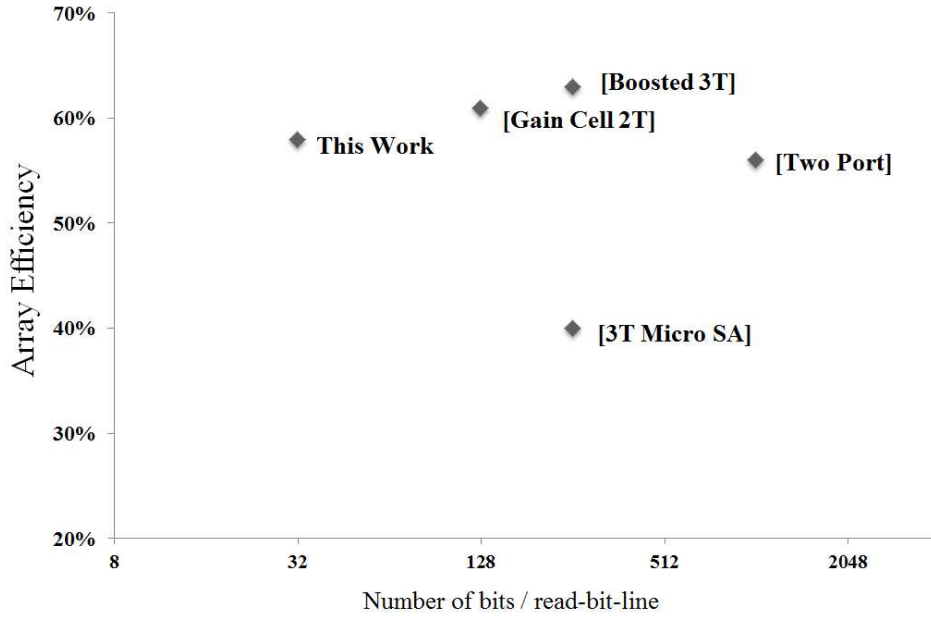


Figure 4.8: Array efficiency compared with other state-of-the-art eDRAM and low power memory: Boosted 3T [28], Gain Cell 2T [29], 3T Micro Sense-Amp [33], Pseudo-Two-Port [32]

Technology	180nm Standard Logic CMOS
Cell Area	3.33 μm^2 (103 F ²)
DRAM / SRAM Cell area ratio	0.72X
Supply Voltage	0.75 V / -0.55V
Circuit Dimension	556 \times 169 μm^2
Macro Size	16 kb
Retention Time (worst / average)	306ms / 861ms @ 25°C 9.5ms / 13ms @ 85°C
Refresh Power	5.42nW/ kB @ 25°C 131nW/ kB @ 85°C
Read Delay	3 μs @ 25°C 300ns @ 85°C
Write Delay	1 μs @ 25°C 30ns @ 85°C

Table 4.2: Performance summary of 2T dual V_{th} eDRAM

to RBL) is 300ns at 85°C and 3 μ s at 25°C whereas measured write delay (WWL to storage node) is 30ns at 85°C and 1 μ s at 25°C which is all acceptable for speed of sensor node processor.

4.5 Conclusion

In summary, eDRAM using a 2T Dual- V_{th} gain cell is demonstrated and 5.42nW/kB retention power with 306ms retention time at 25°C and 131nW/kB retention power with 9.5ms retention time at 85°C is achieved with 103F² cell area. With area efficient single inverter sensing scheme, 58% array efficiency could be achieved for 2kb memory array with 32 bits per bitline.

CHAPTER 5

Low Power 7T SRAM with Heterojunction Tunneling Transistors (HETTs)

5.1 Introduction

Low voltage operation is one of the most effective low power design techniques due to its quadratic dynamic energy savings. Recently, a number of works [34][35][36] have shown aggressive supply voltage reduction to near or below the threshold voltage (V_{th}) of MOSFET devices with considerable reduction in power consumption. However, this power improvement has come at the cost of operation speed (typically <10 MHz). At such low supply voltages, ON current drops dramatically due to lack of gate overdrive resulting in large signal transition delays. To regain this performance loss it is possible to reduce the threshold voltage. However, this exponentially increases OFF current, which is particularly problematic in applications that spend significant time in standby mode [37]. For instance, lowering the supply voltage from 500mV to 250mV while enforcing iso-performance by reducing the V_{th} increases leakage power by $275\times$ in a commercial bulk-CMOS 45nm technology, which is unacceptable. To address this dilemma, there has been recent interest in new devices with significantly steeper subthreshold slopes than traditional MOSFETs [38][39][40][41][42]. A steep subthreshold slope enables operation with a much lower threshold voltage while maintaining low leakage. In turn, a low V_{th} enables low voltage operation while maintaining performance. Hence, steep subthreshold slopes can provide power efficient operation without loss of performance.

In this Chapter, SRAM design using the recently proposed Si/SiGe HETEROJUNCTION Tunneling Transistor (HETT) [43] is investigated. The Si/SiGe heterostructure uses gate-controlled modulation of band-to-band tunneling to obtain subthreshold swings of less than 30 mV/decade with a large ON current of 0.42mA/ μm at $V_{ds} = 0.5\text{V}$. Furthermore, Si/SiGe heterostructures are fully compatible with current MOSFET fabrication process and can leverage the extensive prior investment in CMOS fabrication technology. Currently, several industry and university teams are actively developing Si/SiGe HETT type transistor structures, and initial devices have been experimentally demonstrated [44][45]. The key differences between HETTs and traditional MOSFETs that must be considered in the design of SRAM circuits using these new devices is asymmetric conductance. In MOSFETs, the source and drain are interchangeable, with the distinction only determined by the voltages during operation. However, in HETTs, the source and drain are determined at the time of fabrication, and the current flow for $V_{ds} < 0$ is substantially less than for $V_{ds} > 0$ (in an NHETT). Hence, HETTs can be thought to operate ‘uni-directionally’, passing logic values only in one direction. The unidirectional characteristic of HETTs can actually be exploited in SRAM design to enable a novel 7T SRAM cell.

5.1.1 HETT Device Characteristics

The 60 mV/decade subthreshold slope limitation of conventional MOSFETs arises due to the thermionic nature of the turn-on mechanism. Tunneling transistors do not suffer from this fundamental limitation, since the turn-on in these devices is not governed by thermionic emission over a barrier.

Figure 5.1 illustrates the basic concept of tunneling transistor operation. In an n-type tunneling transistor, the source is doped p-type, the channel is undoped or lightly doped, and the drain is n-type. As shown in Figure 5.1, when the gate is biased positively the device is turned on because electrons in the valence band of the p-type source can tunnel into the conduction band of the channel. If the Fermi level in the source is less than a few thermal voltages (kT) below the valence band edge, the bandgap acts as an “energy filter”, precluding tunneling from the exponential portion of the Fermi-Dirac distribution. If the gate bias is reduced sufficiently so that the bottom of the conduction band in the channel rises above the top of the valence band in the source, the tunneling

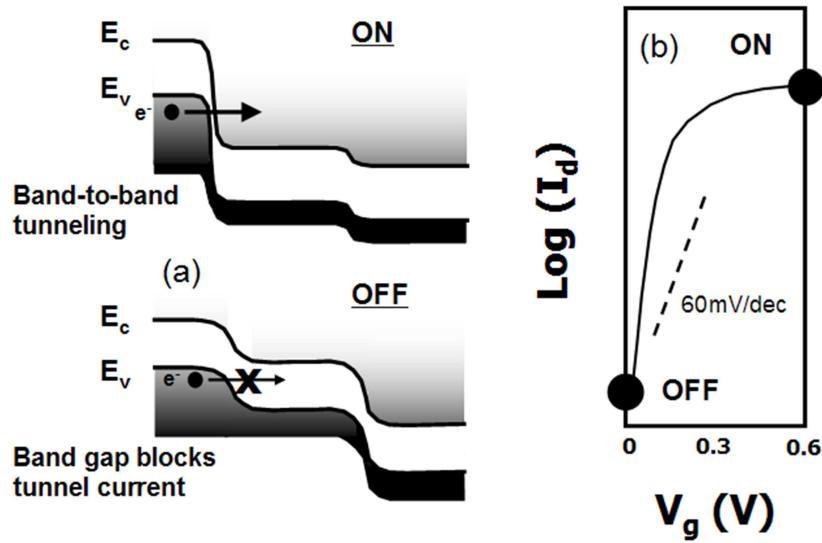


Figure 5.1: Tunneling FET device concept as depicted by a) band diagrams in the source-to-drain direction, and b) qualitative current-voltage characteristics.

abruptly shuts off. Due to this filtering of the Fermi-Dirac distribution function by the bandgap, the subthreshold slopes can be significantly less than 60 mV/decade.

A potential problem with tunneling transistors is that a very narrow bandgap semiconductor must be used to obtain sufficiently high ON current. However, narrow bandgap materials also lead to higher OFF currents, and are often incompatible with standard CMOS processing. To avoid this problem, a type-II hetero-junction tunneling transistor (HETT) can instead be employed. In such a case, the source-to-body contact has a staggered band lineup that creates an effective tunneling band gap, E_{geff} , which is smaller than that of the constituent materials. Such a band structure can also be realized in the Si/SiGe heterostructure material system, and complementary N- and P-HETT can be fabricated, making this technology fully CMOS compatible. Figure 5.2 shows a schematic diagram of a complementary Si/SiGe HETT technology. For the circuit simulations in this work, an optimized device structure was used. The simulated HETT devices have a gate length of 40 nm, and a high-k gate dielectric with effective gate oxide thickness of 1.2 nm. For NHETT, the source consists of pure Ge, with 3% biaxial compressive strain, and Si channel with 1% biaxial tensile strain. The complementary PHETT design includes a strained Si source and pure Ge channel. Using band offsets from [46], the effective bandgap for this structure is 0.22 eV. For the transport calculations, a non-local tunneling model [47] with a 2-band dispersion relationship

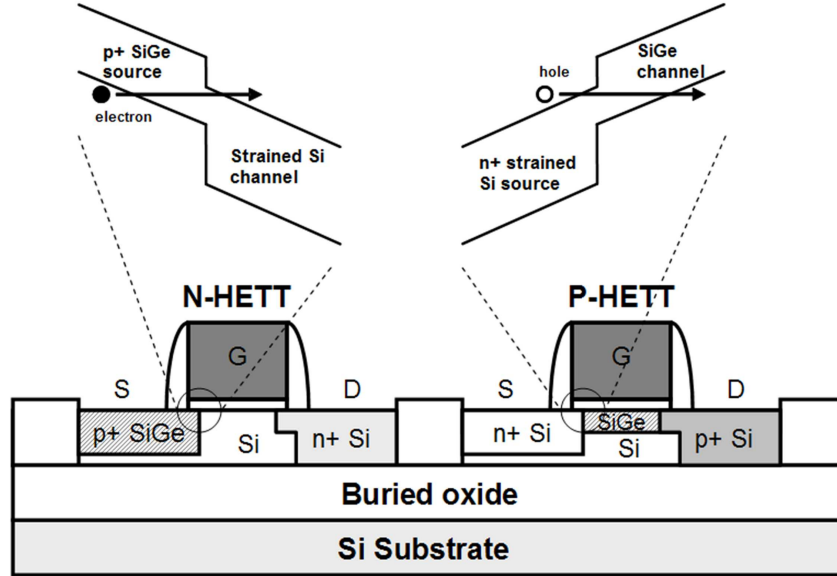


Figure 5.2: CMOS-compatible implementation of complementary tunneling FETs with type-II source-to-body hetero-junctions to improve device drive current.

within the gap was used. Effective masses are $0.17m_0$ near the conduction band and $0.105m_0$ near the valence band in the silicon channel, and $0.10m_0$ near the conduction band and $0.055m_0$ near the valence band in the pure Ge source [48]. The device has a 2nm gate overlap of the source and an abrupt source doping profile. A gate work function of $\sim 4.4\text{eV}$ is used to set the OFF current to $< 1\text{pA}/\mu\text{m}$.

5.1.2 HETT Device Modeling

Since accurate analytical models for HETTs are not available, we first built a look-up table based model using Verilog-A to enable circuit simulations. This technique is a simple and accurate way of compact modeling for emerging devices [49] where analytical expressions for the I-V characteristics are not well established. A look-up table model is built for I-V and C-V characteristics using T-CAD simulation data based on the device parameters described in the above section. The HETT is modeled as a three-terminal device (source, gate, and drain) and current is assumed to flow only between source and drain since gate leakage is negligible with high-k gate dielectrics. Two parasitic capacitors are modeled; C_{gd} and C_{gs} , which include inner fringing capacitance and overlap capacitance between gate and drain and between gate and source, respectively. Channel

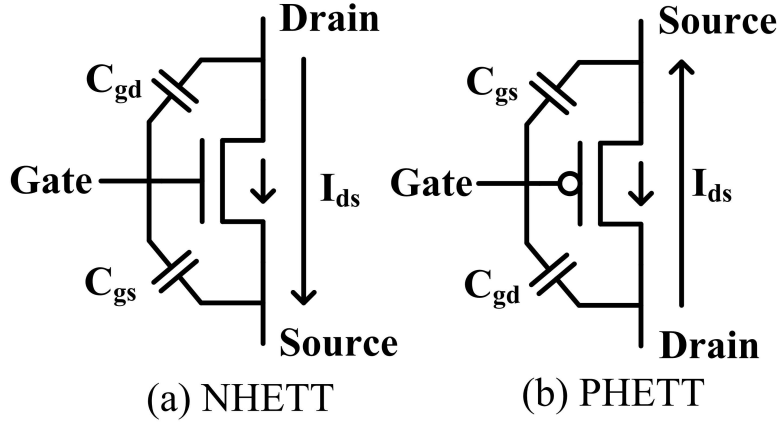


Figure 5.3: Device symbols for (a) NHETT (b) PHETT.

capacitance is negligible because the device has a fully-depleted channel and junction capacitance is also negligible due to its SOI-type substrate. As a result, we build three two-dimensional tables that are functions of two input voltages, V_{gs} and V_{ds} , for modeling HETTs: $I_{ds}(V_{gs}, V_{ds})$, $C_{gd}(V_{gs}, V_{ds})$, and $C_{gs}(V_{gs}, V_{ds})$. V_{gs} and V_{ds} are swept in 50mV steps in general, however in the slightly reverse biased region ($-0.2V < V_{ds} < 0V$) where I_{ds} transition is rapid V_{ds} steps are 10mV for the I_{ds} tables. In Figure 5.3, new symbols for NHETT and PHETT are presented. An arrow inside the conventional MOSFET symbol denotes the direction of forward biased current, which is from drain to source for NHETT and vice versa for PHETT.

5.1.3 Asymmetric Current Flow of HETT

HETT source and drain are determined at fabrication time and current flow between the two nodes is not symmetric. Figure 5.4 demonstrates this asymmetric current flow in an NHETT. We assume that the nominal voltage of HETTs will be $< 0.5V$ as HETTs target ultra-low voltage applications and are well suited for this voltage regime. Figure 5.4(a) shows forward bias current with V_{gs} swept from 0V to 0.5V. The drain current curves look similar to CMOS devices. However, reverse bias current, where the voltage across the drain and source is negative, differs from CMOS devices as shown in Figure 5.4(b). Note that I_{ds} is negative in Figure 5.4(b). For most regions of V_{ds} , drain current is several orders of magnitude smaller than forward current. However, there are two cases where the reverse bias current becomes non-negligible. First is when V_{ds} is approximately 0.5V, at which point drain current become non-negligible regardless of V_{gs} . The

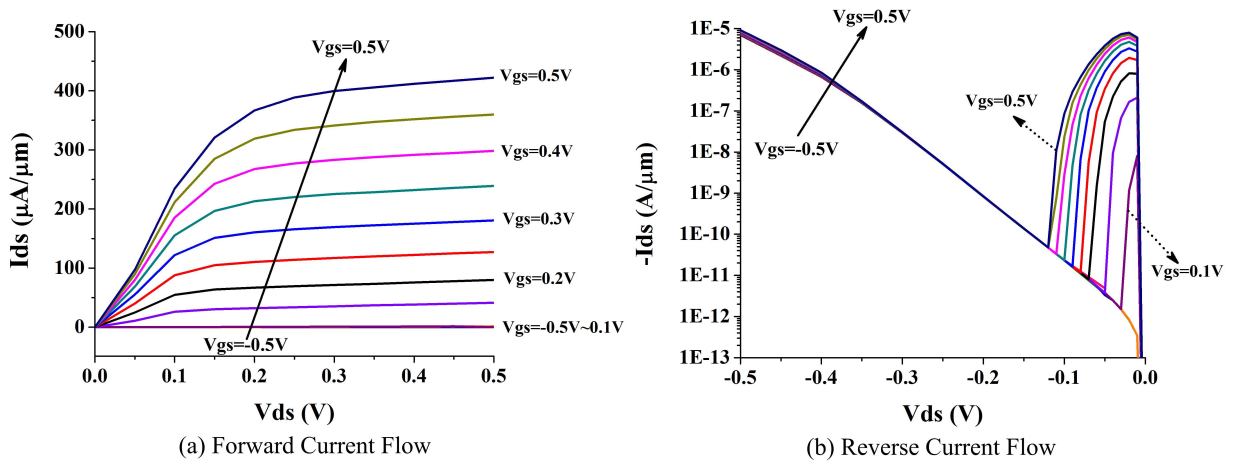


Figure 5.4: Drain current of HETT device with $L=40\text{nm}$ (a) Forward bias (b) reverse bias.

second case occurs for positive V_{gs} combined with a small negative V_{ds} . PHETTs exhibit similar asymmetry in their current flow.

The asymmetric current flow does not restrict the use of traditional static CMOS logic circuits with pull-up network (PUN) and the pull-down network (PDN) because the current flow of each device in the PUN and PDN is uni-directional. However, pass-transistor and transmission-gate operation is limited since they require current flow in both directions. This asymmetric current flow also limits the use of the standard 6T SRAM cell and static latches/registers, which exploit pass-gates and transmission-gates as key components. To make low power SRAM feasible with HETT, limitation in standard 6T SRAM is analysed in Section 5.2, and based on this analysis, novel 7T SRAM for HETT is proposed in Section 5.3.

5.2 Limitations in Standard 6T SRAM

5.2.1 CMOS Standard 6T SRAM

To understand the difference between HETT-based 6T SRAM and CMOS-based 6T SRAM, we trace current flow paths in read and write operations. Figure 5.5 shows a CMOS 6T SRAM cell storing 0. To read the stored value, bit lines (BIT, BIT_B) are pre-charged to V_{DD} and as word line (WL) is driven high, NPDL pulls down the voltage at BIT as shown in Figure 5.5(a). This pull down current or voltage can be sensed by a sense amplifier to determine the stored value. For

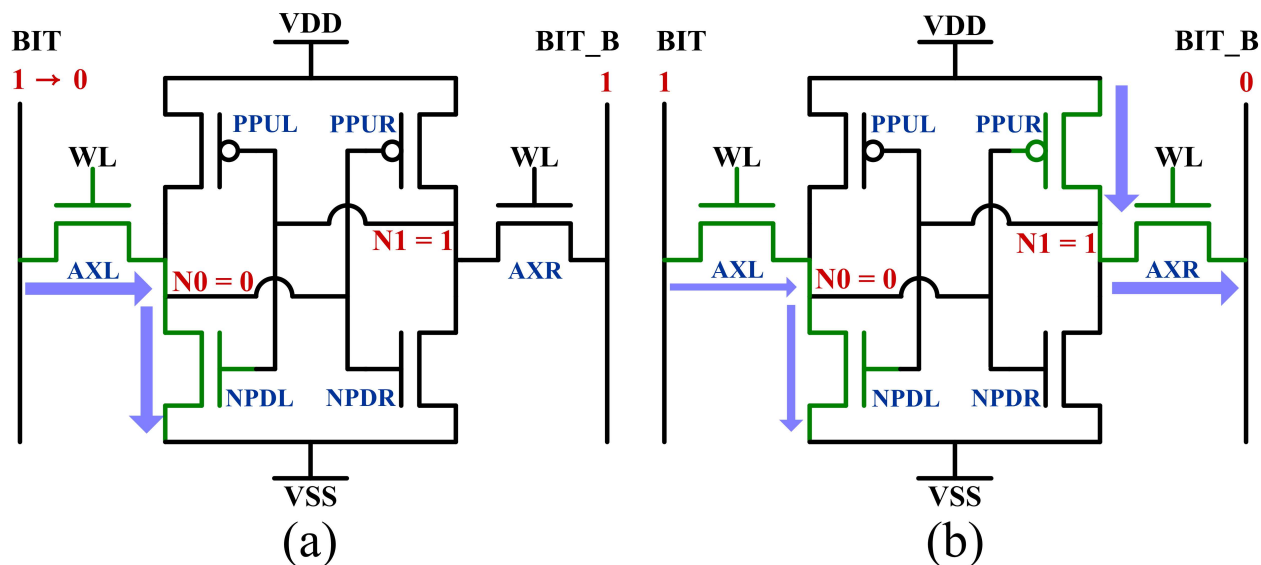


Figure 5.5: Current flow paths in (a) read and (b) write operations in CMOS 6T SRAM.

writing a value 1, as shown in Figure 5.5(b), AXL pulls up internal node N0 while AXR pulls down internal node N1. However, since both access transistors are NMOS, which are better at pulling low, AXR plays the major role in write 1 operation. AXL aids in writing a 1 by pulling up N0 to a certain extent and making the bit flip more easily. For this type of SRAM, read stability can be improved by increasing the sizing ratio of NPDL to AXL (or NPD to AX), which is commonly referred to as the cell β -ratio. As cell β -ratio increases, NPDL in Figure 5.5(a) holds the voltage at node N0 to ground more strongly during read, making it more stable. At the same time, this worsens writeability of the cell by making it more difficult to change the voltage at node N0. However as shown in Figure 5.5(b), since the pull down current path (AXR) plays the major role in writing, the size ratio of AXR to PPUR, or AX to PPU, is the critical one for writeability and can be improved by increasing this ratio. This implies that, up to a point, readability and writeability in CMOS 6T SRAM can be improved individually at the cost of larger area.

5.2.2 HETT Standard 6T SRAM with Inward Access Transistors

Due to its uni-directional nature, access transistors in HETT 6T SRAM can drive current either inward or outward only. Figure 5.6 shows a HETT 6T SRAM structure with inward current flow configuration and storing 0. Read operation for this SRAM is similar to a CMOS 6T SRAM. Bit-

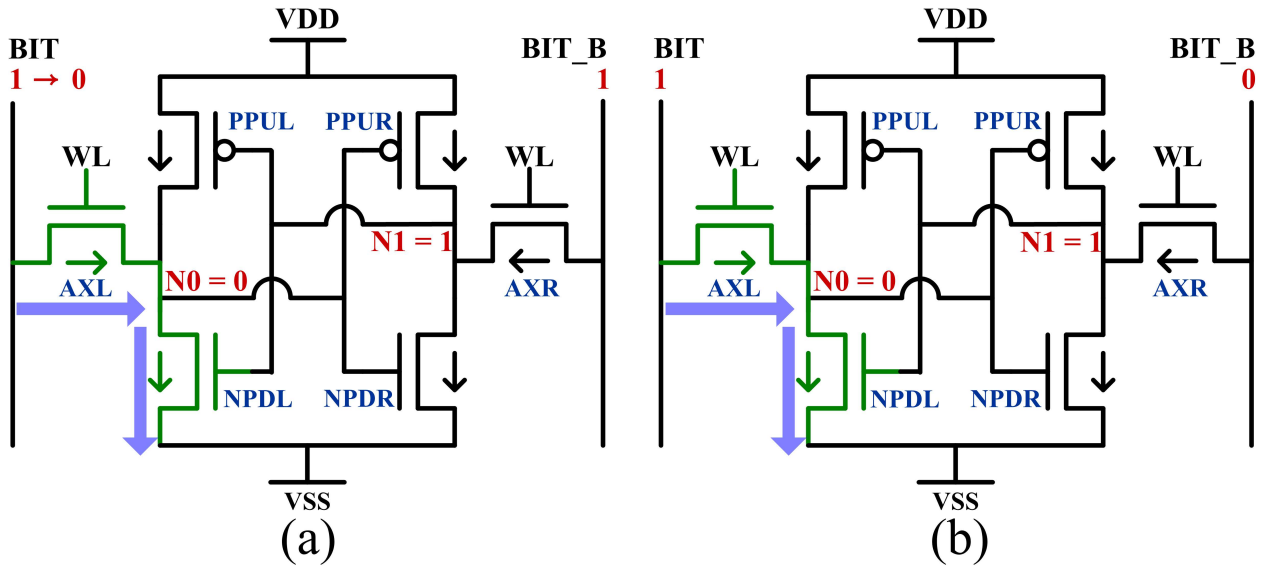


Figure 5.6: Current flow paths in (a) read and (b) write operations in HETT 6T SRAM with inward direction access transistors.

lines are precharged and current flows through AXL and NPDL. Therefore, similar to CMOS 6T SRAM, higher cell β -ratio is preferred for preventing read upset. However, to write 1 to this cell, AXR cannot pull down the voltage at N1 since it can only conduct current inward, implying that AXL must pull up the voltage at N0 without differential aid, as shown in Figure 5.6(b). Therefore, the write operation is performed only by one side and the stronger current path is removed in HETT 6T SRAM. Since we are relying on an N-type transistor to drive the internal node voltage high, writeability of this cell is substantially worse than a CMOS 6T SRAM. To overcome poor writeability, AXL should be strengthened compared to NPDL, i.e., the cell β -ratio should be decreased. However, decreasing the cell β -ratio negatively affects the read margin.

This tradeoff between readability and writeability can be clearly seen if we plot static noise margin (SNM) of read and write operation versus cell β -ratio, as shown in Figure 5.7(a). SNM is the maximum DC voltage of the noise that can be tolerated by the SRAM and it is widely used for modeling stability of SRAM cells. SNM can be defined for three different operations - read, write, and standby (hold) - but only read and write margins are compared here since they limit SRAM stability. In SNM analysis for HETT-based SRAMs, all simulations use $V_{DD} = 0.5V$ since HETTs are aimed at this voltage regime. For HETT 6T SRAM with inward access transistors with cell β -ratio of 1, read margin is 34mV but write margin is 0V, meaning that write operation is

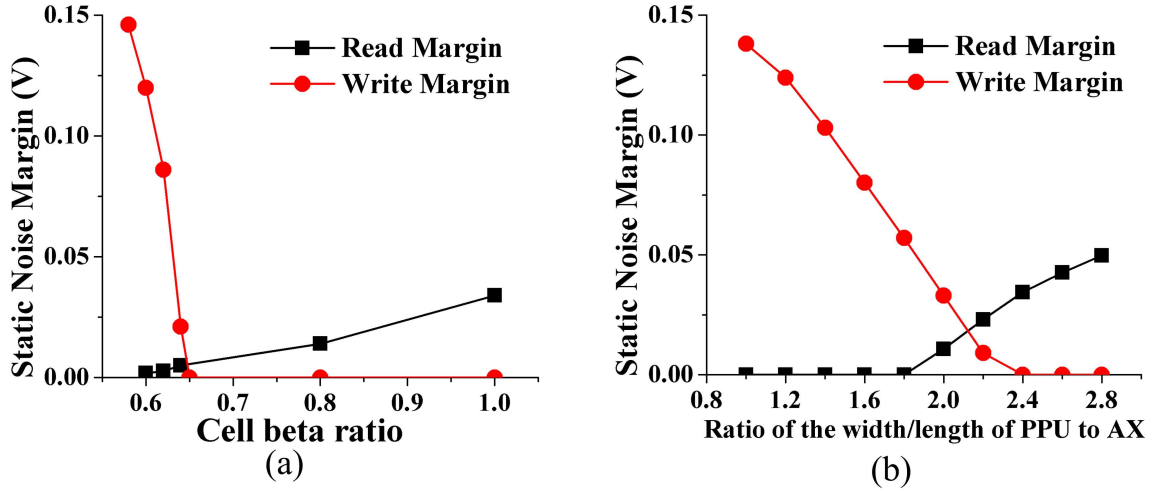


Figure 5.7: Static noise margins of HETT 6T SRAM with (a) inward and (b) outward access transistor with $V_{DD}=0.5V$.

impossible. As we decrease the cell β -ratio to improve writeability, write margin becomes positive at a cell β -ratio of 0.64, however read margin at this point has degraded to <3 mV, indicating that the cell is highly vulnerable to read upset at this design point. From this we conclude that HETT 6T SRAM with inward access transistors is not feasible.

5.2.3 HETT Standard 6T SRAM with Outward Access Transistors

HETT 6T SRAM with outward access transistors has a similar limitation. Figure 5.8(a) shows a read operation, where bit lines (BIT BIT_B) are pre-discharged and BIT_B is charged through AXR and must be sensed. For writing, AXR must drive internal node N1 to ground and flip the stored value without differential assistance from AXL. Since both of these operations involve PPUR and AXR, adjusting the ratio of PPUR to AXR strengths will improve one operation and worsen the other. This tradeoff can be clearly seen in Figure 5.7(b). The read operation requires PPUR to AXR ratio higher than 1.8, while the write operation malfunctions when the ratio is higher than 2.4. In the remaining design space the SNM for read/write operations is limited to ≈ 50 mV, which is insufficient. Therefore, an alternative SRAM topology is needed to achieve robust low leakage SRAM with HETT's.

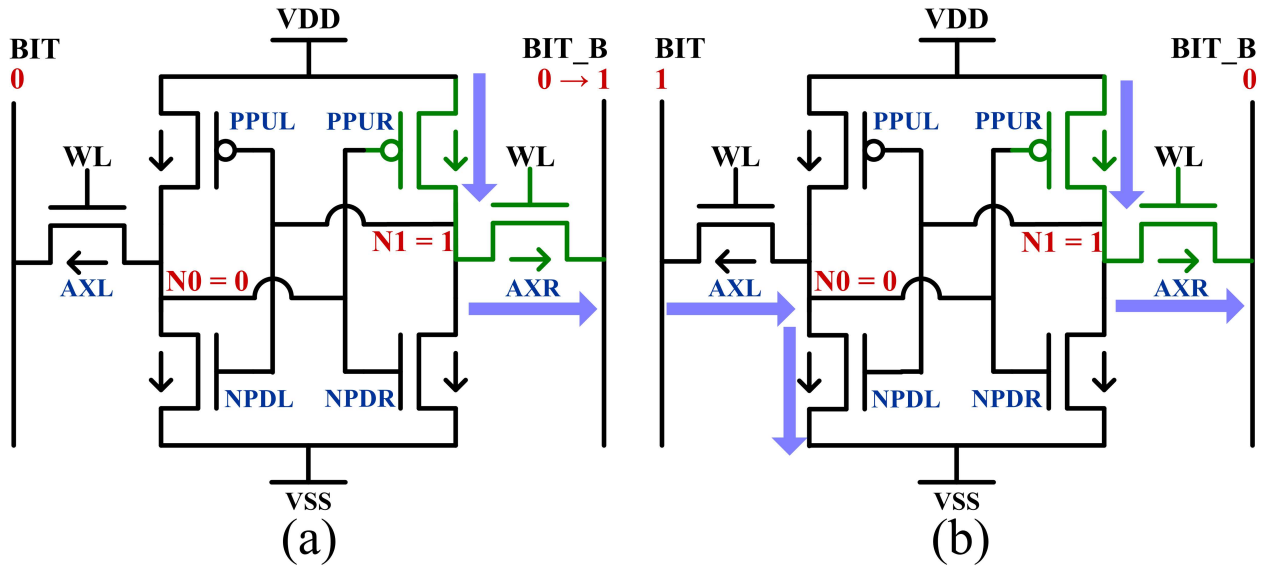


Figure 5.8: Current flow paths in (a) read and (b) write operations in HETT 6T SRAM with outward direction access transistors.

5.3 Alternative SRAM Design with HETT

A fundamental trade-off between readability and writeability limits the implementation of 6T HETT SRAM. This trade-off can be avoided by separating read and write current flow path at the cost of a few additional transistors. In this section, various possible read and write structures for HETT-based SRAM are compared. Then a 7T HETT SRAM is proposed and analyzed in detail.

5.3.1 Read Structure for HETT SRAM

In 6T SRAM, back-to-back inverters are the components that store the value and two access transistors (AXL/AXR in Figure 5.5) are used as read structure and write structure at the same time. To separate read and write path, three possible read structures are shown in Figure 5.9.

Figure 5.9(a) shows a single HETT read structure where an additional HETT dedicated to read operation is attached to back-to-back inverter pair. With this structure, inward NHETT configuration is preferred to outward configuration to minimize chance of read upset. The benefit of this separate read structure is that separate cell β -ratio can be obtained for read and write operation. By utilizing weaker NHETT just for read operation, better read margin can be obtained while maintaining same write margin. Figure 5.9(b) shows the read structure widely used in CMOS

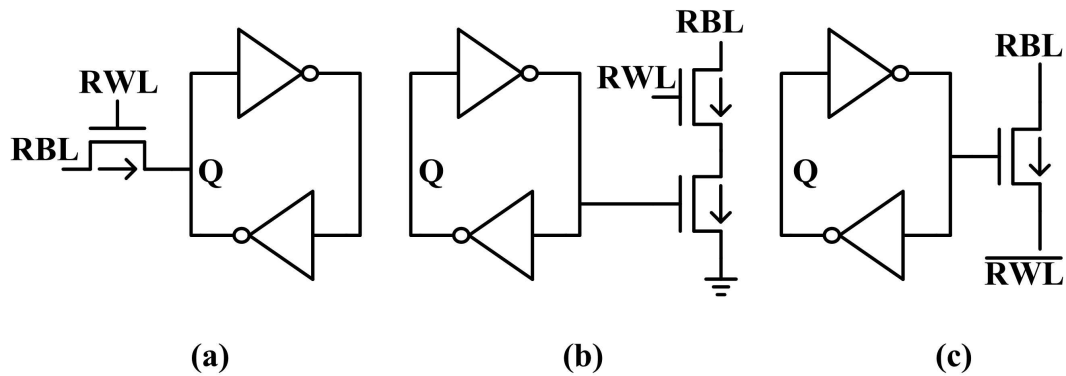


Figure 5.9: Alternative read structures for HETT-based SRAM (a) Single HETT read (b) 8T read (c) reduced 8T read.

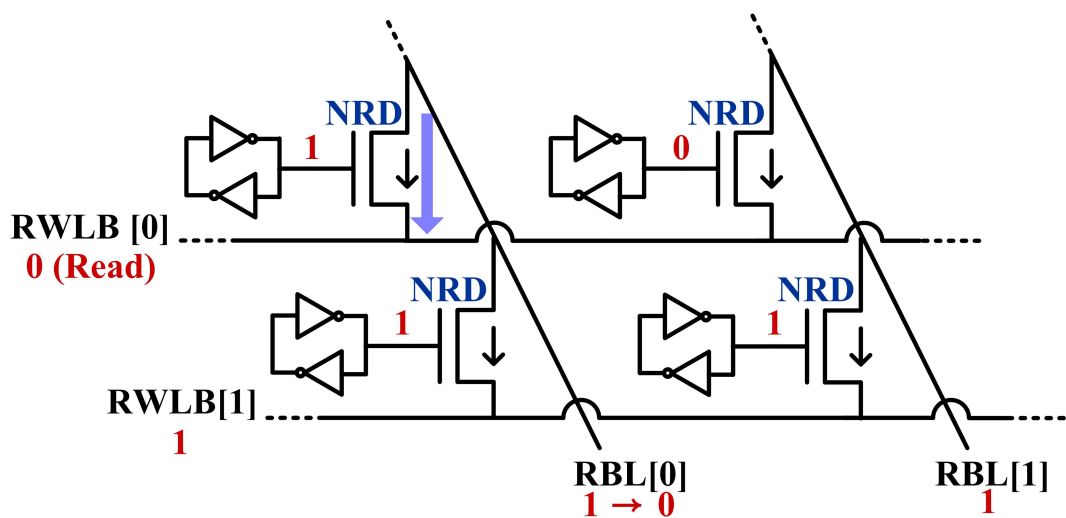


Figure 5.10: Read operation with reduced 8T read.

8T SRAM [10], where transistors are replaced with HETTs. This structure implements voltage sensing of the stored value, eliminating the current flow path through the back-to-back inverters. Therefore, the possibility of read upset is virtually eliminated at the cost of two additional transistors. In this structure, bottom NHETT senses the stored voltage and top NHETT selects the word to be read. However, by taking advantage of HETT's asymmetric current flow, voltage sensing and word selection can be done with one NHETT as shown in 5.9(c). Instead of grounding source of the sensing (bottom) HETT, an inverted RWL is connected to the source so that only the selected word can drain current through the sensing HETT.

Figure 5.10 illustrates how the NHETT of a reduced 8T read structure (NRD) in each cell is connected in the array structure. The source of NRD is connected to that of other cells in the

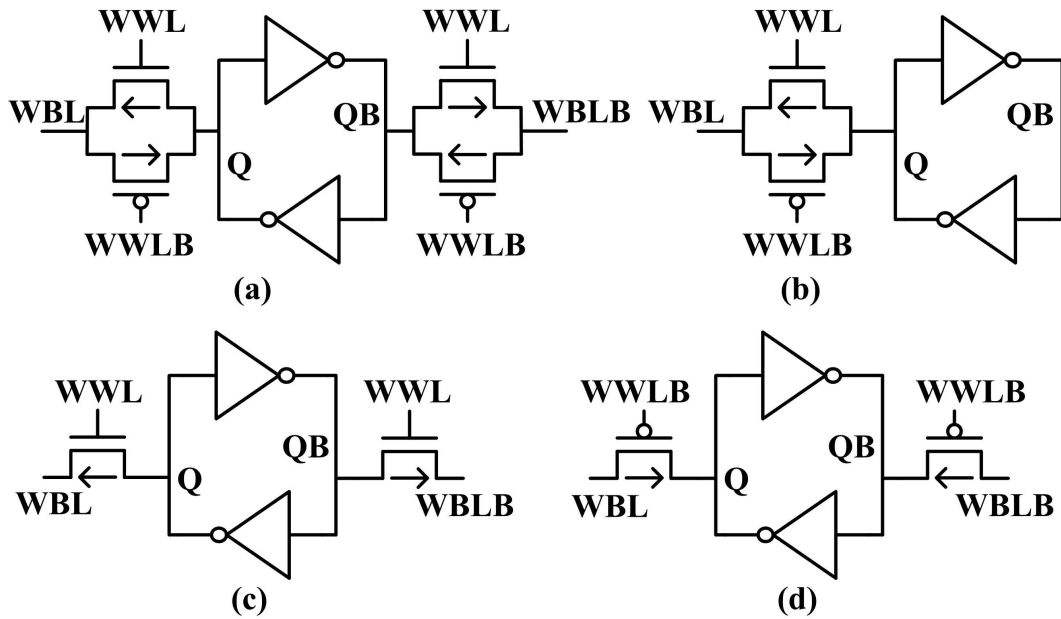


Figure 5.11: Alternative write structures for HETT-based SRAM (a) Two-side transmission gate write (b) One-side transmission gate write (c) Two-side NHETT pull down write (d) Two-side PHETT pull up write.

same word (RWLB), while the drain is connected to that of other cells in same column (RBL). To read values in word[0] (top row of Figure 5.10), bit-lines (RBL[0], RBL[1]) are pre-charged and RWLB[0] is asserted (driven to ground) while all other RWLBs are set to VDD. Since the source of the NRDs in word[0] are set to ground, cells that store value ‘1’ can discharge the bit line, as depicted with the thick arrow in Figure 5.10. With CMOS transistors, this read scheme does not work because, as RBL[0] is discharged, other cells storing ‘1’ on the same bit line can start charging up RBL[0] as in the case of the bottom-left cell in Figure 5.10. However, by leveraging the asymmetric nature of HETTs, this unwanted reverse-direction charging current is eliminated without the cost of an additional transistor. Therefore, reduced 8T read can achieve robust read operation as robust as 8T read with the same HETT count with single HETT read.

5.3.2 Write Structure for HETT SRAM

Figure 5.11 shows four of possible HETT write structures. The trade-off between readability and writeability originates from asymmetric current flow of access transistors (AXL/AXR in Figure 5.5). Therefore, allowing bidirectional current flow by replacing access transistors with

transmission gates (Figure 5.11(a)) can eliminate this trade-off. Although this scheme allows both read and write access through transmission gates, it requires 8 HETTs which can be reduced by more advanced read and write structures. To reduce HETT count, single-ended access can be used where transmission gate on one side can be eliminated (Figure 5.11(b)). However, this requires the PHETT in the transmission gate to be sized up by $1.55\times$ since PHETT has weaker current driving capability.

The non-uniform sizing of NHETT and PHETT in the transmission gate can result in irregular layout especially when the size difference is as high as 55%. To avoid this, an identical type of HETT can be used as access transistor, just as in standard 6T SRAM, but only for write operation. Figure 5.11(c) shows two-side NHETT pull down write where value is written by pulling down one of the storage nodes (Q and QB). For writing with NHETT, outward configuration is preferred. If we assume back-to-back inverters are min-sized for minimum cell area, size of NHETT and PHETT should be identical. With this assumption, Figure 5.7 shows that writing with outward minimum sized NHETT is robust with noise margin of 143mV, whereas inward NHETT has to be widened by 1.4 just to be functional. For the same reason, inward configuration is better with PHETT write (Figure 5.11(d)). However, this scheme also requires 1.55 times larger PHETT to achieve comparable write noise margin with NHETT write.

Two-side NHETT pull down writing structure (Figure 5.11(c)) also can benefit from the uni-directional current flow, which mitigates the half select disturbance in a bit-interleaved array. The half select disturbance accidentally flips internal data in half selected bitcells which share the same write word line with targeted bitcells for write operation [51]. With two-side NHETT pull down write structure, if the write bit lines of half selected bitcells are kept at VDD, the amount of current flow via access transistors is limited to the leakage current level. Therefore, two-side NHETT pull down write structures have improved immunity during half select accesses.

5.3.3 7T SRAM for HETT

Based on previous discussion, 7T SRAM optimized for HETT is proposed as shown in Figure 5.12. In this topology, readability/writeability tradeoffs in HETT-based 6T SRAM is overcome by utilizing separate read and write structure. The reduced 8T read enables extremely robust read with

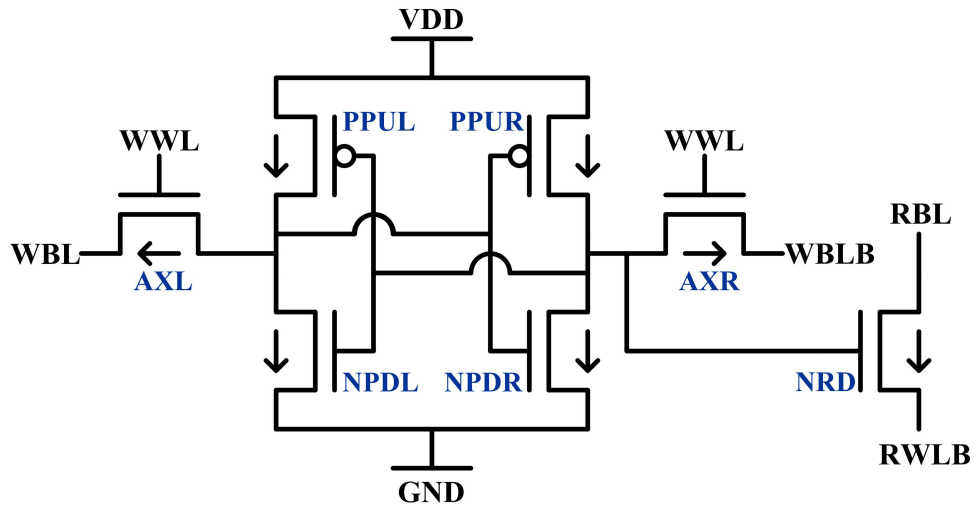


Figure 5.12: Proposed HETT 7T SRAM structure.

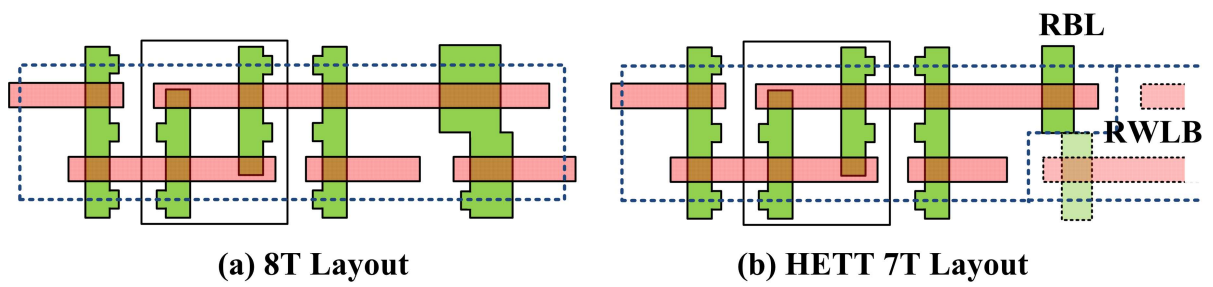


Figure 5.13: (a) 8T layout [50] and (b) corresponding HETT 7T layout.

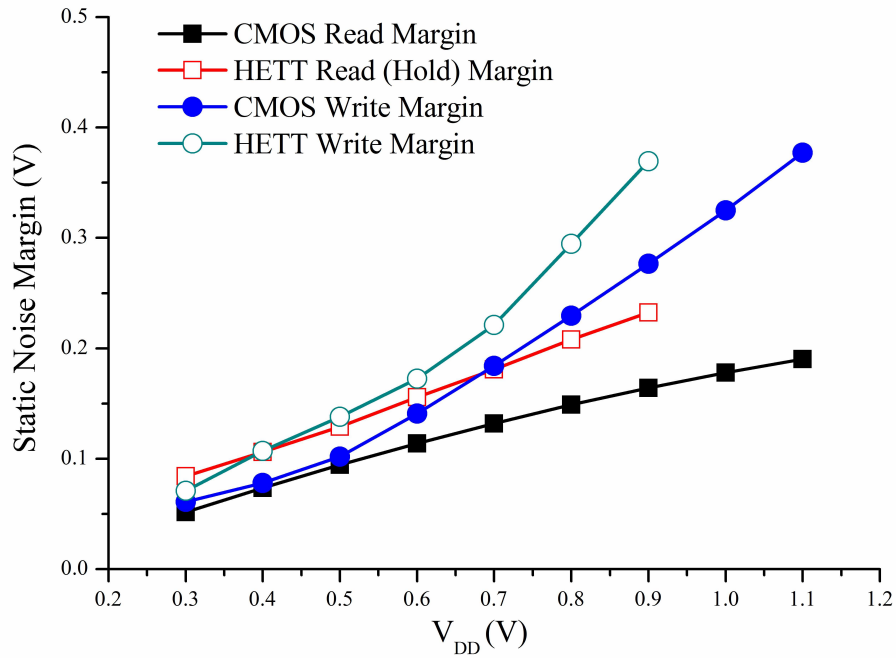


Figure 5.14: Read/Write margin of 45nm commercial bulk CMOS 6T SRAM and HETT 7T SRAM.

minimal additional number of HETT and two-side NHETT pull down write enables robust write with cell β -ratio of 1, where all HETT sizes can be minimum.

The HETT 7T SRAM is estimated to have <15% area overhead over a standard 6T while 8T SRAM exhibits 29% cell area overhead [10]. Figure 5.13 shows that two read transistors (NRD in Figure 5.12) from adjacent cells can be abutted in 7T SRAM, making the overhead for two 7T cells equal to that of one 8T cell. Moreover, as will be shown below the 7T cell with all transistors at minimum size shows improved robustness over 6T at low voltage, hence if an upsized 6T were used to achieve iso-robustness the area penalty would be much smaller than 15%.

A write operation in this 7T structure is equivalent to the HETT 6T SRAM with outward access transistors. However, since the read/write operations are performed by separate current paths, device sizes for all transistors other than NRD can be chosen to favor writeability.

We compare SNM of HETT-based 7T SRAM to a 45nm commercial bulk CMOS 6T SRAM cell provided by a foundry. All HETT devices are set to equal (minimum) width for maximum density. Read and write margins of both types of SRAMs across a range of supply voltages are plotted in Figure 5.14. SNM for HETT is analyzed with supply voltages up to 0.9V only since

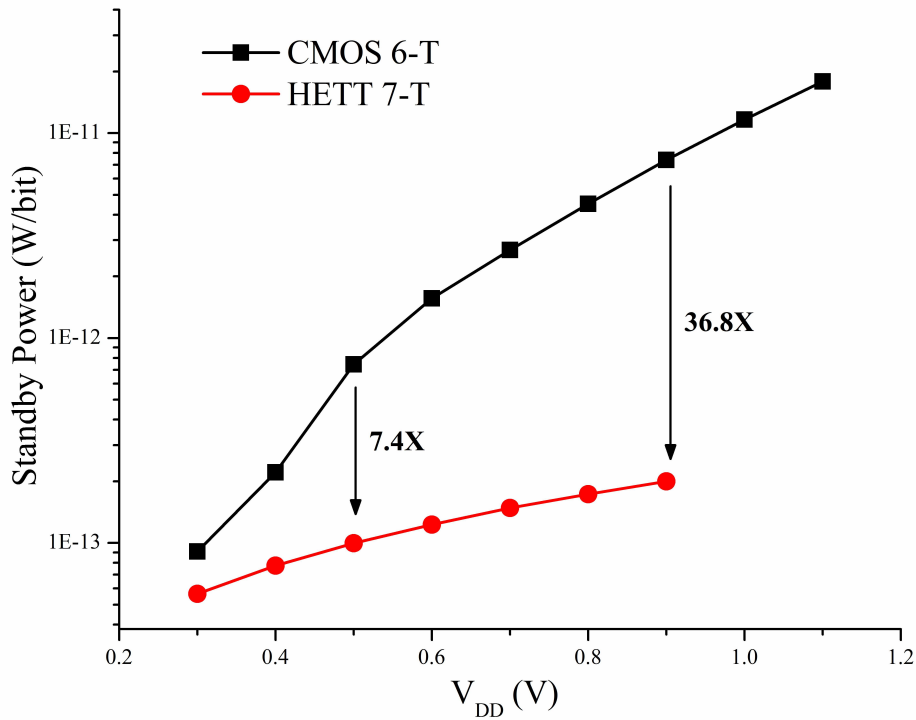


Figure 5.15: Standby power of CMOS 6T and HETT 7T SRAM.

HETT is designed for low voltage ($\sim 0.5V$) operation. Write margins of HETT 7T SRAM are more than 30% higher than CMOS 6T SRAM for supply voltages of $>0.4V$ as shown in Figure 5.14.

Since the read operation uses an additional read transistor in the HETT 7T SRAM and all other transistors are in standby (hold) state during read operation, hold margin is equivalent to read margin in HETT 7T SRAM. Given this, HETT 7T read margin is 232 mV at $V_{DD}=0.9V$ and 129 mV at 0.5V, which is 41% and 37% higher than commercial bulk CMOS 6T SRAM, respectively. Such improvements in read/write margin can be observed for V_{DD} down to 0.3V, suggesting that improved read/write robustness can be achieved with HETT 7T SRAM over traditional CMOS at low voltage.

Finally, HETT-based SRAM standby power is significantly reduced compared to CMOS 6T SRAM, as seen in Figure 5.15. At a supply voltage of 0.9V, standby power is reduced by 36.8 \times and at 0.5V, by 7.4 \times . This clearly shows the promising low-leakage properties of HETT devices for future memory-dominated low-power applications.

5.4 Conclusion

A circuit perspective of a new promising tunneling transistor, HETT, with steep subthreshold swing for extremely low power applications was presented in this section. 9-19× dynamic power reduction is observed with HETT-based circuits due to their improved voltage scalability. The limitations of HETTs are examined as they relate to circuit operation. To overcome and exploit the inherent device asymmetry, a new HETT-based SRAM cell topology was presented with 7-37× leakage power reduction.

CHAPTER 6

A Sub-nW Gate-Leakage Based Temperature Compensated Timer for Cubic-Millimeter Sensor Systems

6.1 Introduction

6.1.1 Operation of Ultra-Low Power Wireless Sensor Node

Recent work in ultra-low-power sensor platforms has enabled a number of new applications in medical, infrastructure, and environmental monitoring. To sustain its functionality with its limit on the volume for energy source, these sensor node maximizes the use of duty-cycling and operates in various operation modes - where each operation modes has wide range of power budgets depending on the activated functions. Figure 6.1 shows various ultra-low power wireless sensor node operation modes and its power consumption in each mode. The sensor nodes typically operate with long idle times and ultra-low standby power ranging from 10s of nW down to 100s of pW [52][7]. In sensor measurement mode, sensor and data processing unit is activated and it can consume 10n-10 μ W of power. In radio transmission/reception (TX/RX) mode, the sensor node consumes significantly larger power since radio transmission/reception is relatively expensive, even at the lowest reported power of 0.2mW [53]. Since the radio TX/RX mode power consumption is 5-6 orders of magnitude larger than idle mode power consumption, wireless communication between sensor nodes must be performed infrequently. Even with infrequent wireless communication, radio can still dominate sensor node energy budget if there is large uncertainty with radio TX/RX timing. Figure 6.1 shows an example operation scenario of an ultra-low power wireless sensor node. In this scenario, the

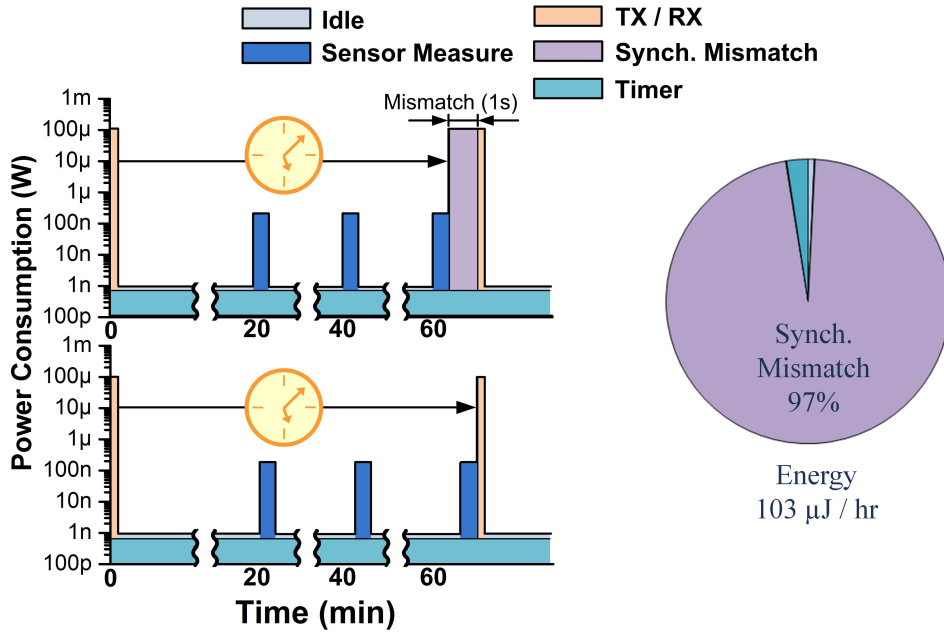


Figure 6.1: Power consumption of example ultra-low power wireless sensor module in various operation modes.

sensor node takes measurement every 20 minutes for 100ms, and transmits accumulated data once every hour. To transmit 1kb data at 1Mbps radio, it only takes 1ms. However, if there is mismatch in communication interval time (i.e. synchronization cycle) measurement for two communicating sensor nodes, this results in synchronization uncertainty (Figure 6.1) and the sensor node which activated radio first has to wait until the second sensor node activates its radio. This radio activation for synchronization uncertainty is as expensive as radio communication in terms of power but can be much more expensive in terms of energy. The uncertainty of timers with <1 nW power budget can be > 1 s per hour [55][57] and 1s uncertainty will result in radio dominating 97% of energy budget as shown in Figure 6.1, which will significantly limit the lifetime of battery-operated wireless sensor nodes. Therefore, accurate measurement of synchronization cycle is of great importance.

6.1.2 Prior-art Timers for Ultra-Low Power Wireless Sensor Node

Quartz crystal oscillators and CMOS harmonic oscillators exhibit very small sensitivity to supply voltage and temperature [54] but cannot be used in the target application space since they

operate at very high frequencies and exhibit power consumption that is several orders of magnitude larger ($>300\text{nW}$) than the needed idle power. Moreover, with mm-scale sensors, physical volume of quartz crystal can be significant portion of the system which will further reduce the volume for energy storage (i.e. battery). To reduce power consumption of timer, a gate-leakage based timer was proposed [55] that leveraged small gate leakage currents as small as 10s of $\text{pA}/\mu\text{m}^2$ [56] to achieve power consumption within the required budget ($< 1\text{nW}$). However, this timer incurs high RMS jitter (1400ppm) and temperature sensitivity ($0.16\%/^{\circ}\text{C}$). A 150pW program-and-hold timer was proposed [57] to reduce temperature sensitivity but its drifting clock frequency limits its use for synchronization.

6.1.3 Metrics for Ultra-Low Power Timers

The quality of a timer is not captured well by RMS jitter since it ignores the averaging of jitter over multiple timer clock periods in a single synchronization cycle. Instead, the uncertainty in a single synchronization cycle of length T is proposed as new metric and use this synchronization uncertainty (SU) to evaluate different timer approaches. The timer period is a random variable $X(n)$, with mean and sigma, μ and σ . Given a synchronization cycle time T , consisting of N timer periods, SU is defined as the standard deviation of T as given by

$$SU = \sqrt{(T/\mu)} \times \sigma = \sqrt{N} \times \sigma = \sqrt{\mu T} \times \sigma/\mu \quad (6.1)$$

assuming $X(n)$ is Gaussian. Note that a smaller clock period increases N and results in more averaging and a lower SU with fixed jitter (σ/μ).

6.2 Multi-Stage Gate-Leakage Based Timer for Low Jitter

The timer in [55] has a high SU since it is triggered with a low gain Schmitt trigger and it has a long period ($\sim 10\text{s}$). To combat this, followings are introduced: 1) a multi-stage structure with a high-gain triggering buffer, 2) boosted capacitance charging, 3) the use of zero threshold voltage transistor (ZVT) for faster gate leakage discharge.

The structure of the proposed multi-stage gate-leakage based timer and its waveforms are

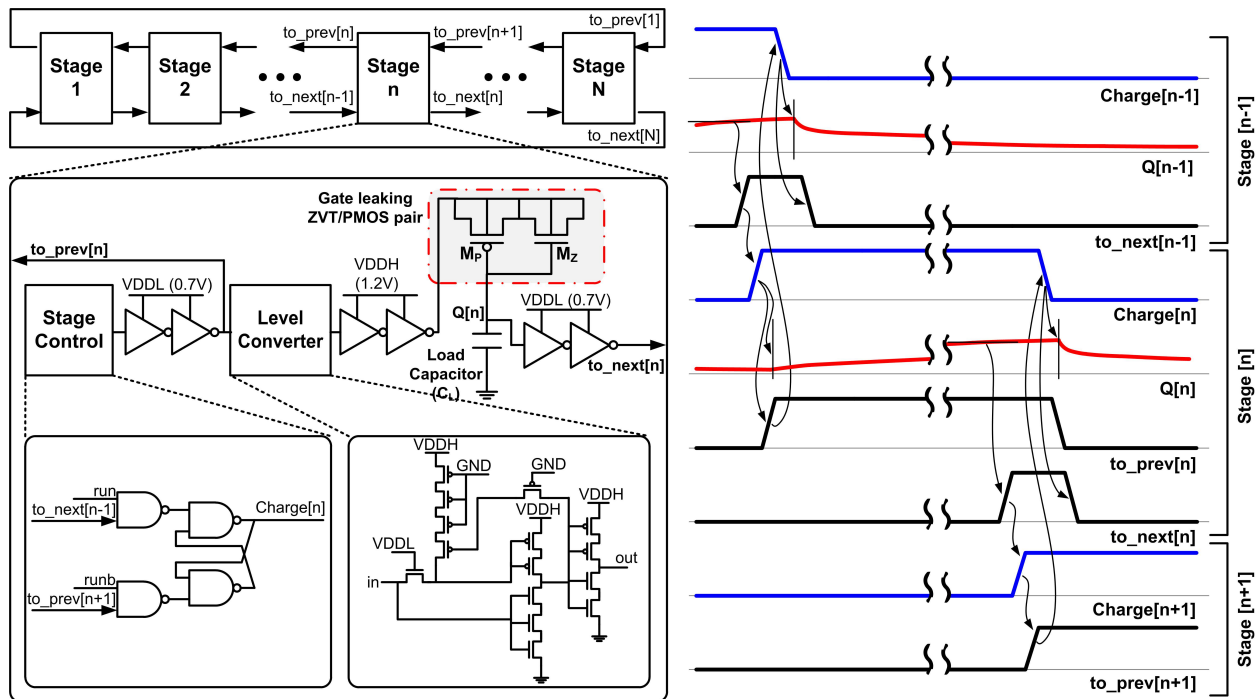
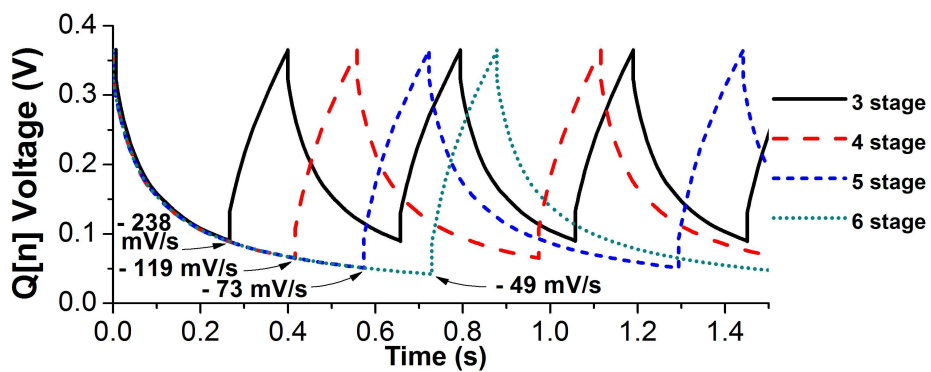
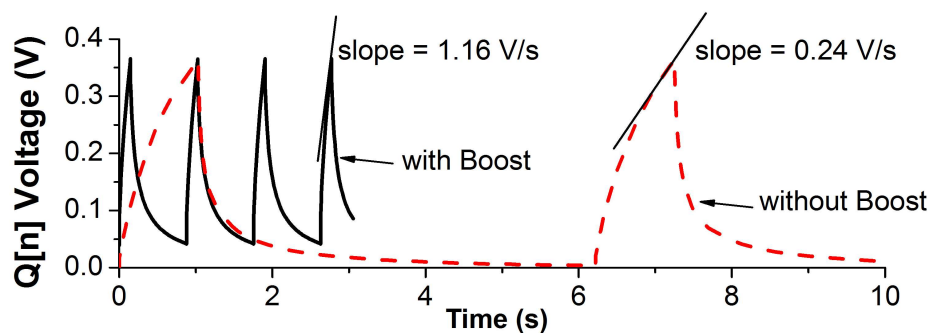


Figure 6.2: Proposed multi-stage gate-leakage based timer.



(a)



(b)

Figure 6.3: Effect of (a) multi-staging and (b) boosted charging.

shown in Figure 6.2. In a stage, a load capacitor (C_L) is charged with the combined gate leakage current of a ZVT and a PMOS transistor. As C_L is charged, the output driving the next stage is triggered by a buffer stage, which shows higher gain than a traditional Schmitt trigger previously used [55]. This places the next stage in a charging state while the current stage discharges. At any given time, only one stage is in a charging state while all others discharge. This allows $n-1$ more discharging time than charging time in an n stage timer and increases the voltage swing on C_L ($Q[n]$). Figure 6.3(a) clearly shows this benefit of multi staging. Longer discharge time lowers the slope at node $Q[n]$ at the end of discharging state (from -238mV/s to 20mV/s for n from 3 to 10), which makes the initial capacitor node voltage for next following charging stage less sensitive to uncertainty. To reduce the uncertainty at the triggering point, boosted charging is introduced. Each stage has low and high supply voltage domains. Low voltage domain is used for the most of the circuits to minimize the leakage power. High voltage domain is used to boost the gate-leakage current which steepens the charging transition on $Q[n]$ and reduces uncertainty at the triggering point. The simulation result shows that boosting 0.7V supply voltage to 1.2V steepens the charging transition by $5\times$ as shown in Figure 6.3(b).

6.3 Temperature Compensation for Multi-Stage Gate-Leakage Based Timer

Temperature compensation for gate-leakage based timer can be achieved by exploiting the opposite temperature dependencies of gate leakage current (I_{gate} of ZVTMOS and PMOS. For fixed gate voltage, simulation result shows that I_{gate} of ZVTMOS and PMOS has linear dependency on temperature (Figure 6.4). Therefore, by selecting appropriate sizing ratio between ZVTMOS and PMOS, linear temperature dependency can be eliminated. Figure 6.5 shows the block diagram of temperature compensated I_{gate} based timer. Instead of single pair of ZVTMOS and PMOS, array of ZVTMOS and PMOS are deployed and proper sizing combination can be chosen for linear dependency elimination. However, this compensation scheme results in a residual second order dependency. To minimize the impact of this second order dependency, an adaptive scheme is proposed in which, for each temperature range, a controller automatically selects a pre-stored

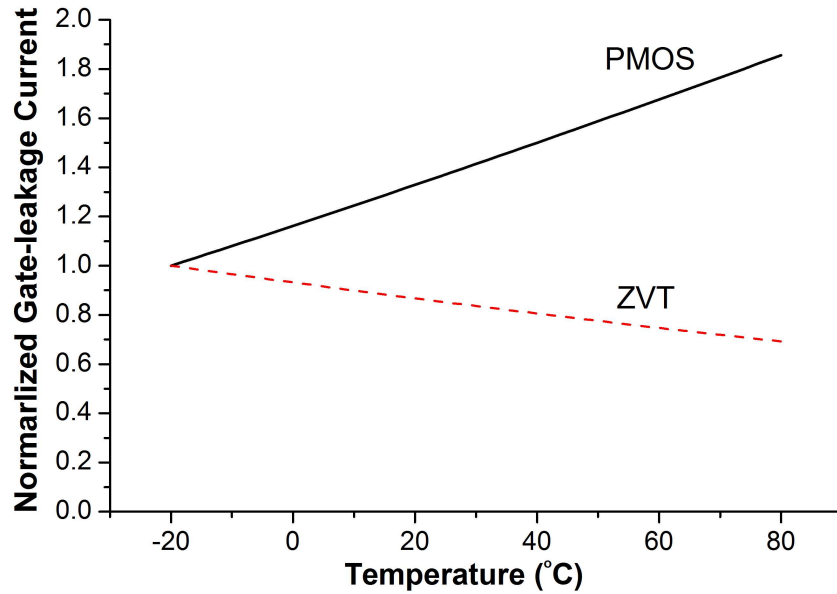


Figure 6.4: Opposite temperature dependency of ZVTMOS and PMOS gate leakage current.

transistor size configuration which minimizes the second order dependency (Figure 6.6). The optimal configurations are determined and stored during post-silicon testing. Each time when the sensor node processor wakes up, it computes time by calculating the elapsed time using the stored period for proceeding configuration and the number of cycles during the last standby state. The transition between configurations occurs synchronously when the first stage starts a new charging state; this allows an exact period calculation and prevents noise injection during capacitor charging. Un-selected ZVTMOS transistors are driven to 400mV to minimize leakage by placing them in accumulation mode.

6.4 Measurement Results

6.4.1 Uncertainty Reduction

A test chip was designed and fabricated in 0.13 μ m CMOS with the proposed multi-stage gate-leakage timer (MGT). As number of stages (N) increases, duty cycle decreased, inverse proportional to number of stages (Figure 6.7). Lower duty cycle implies the longer discharging time with

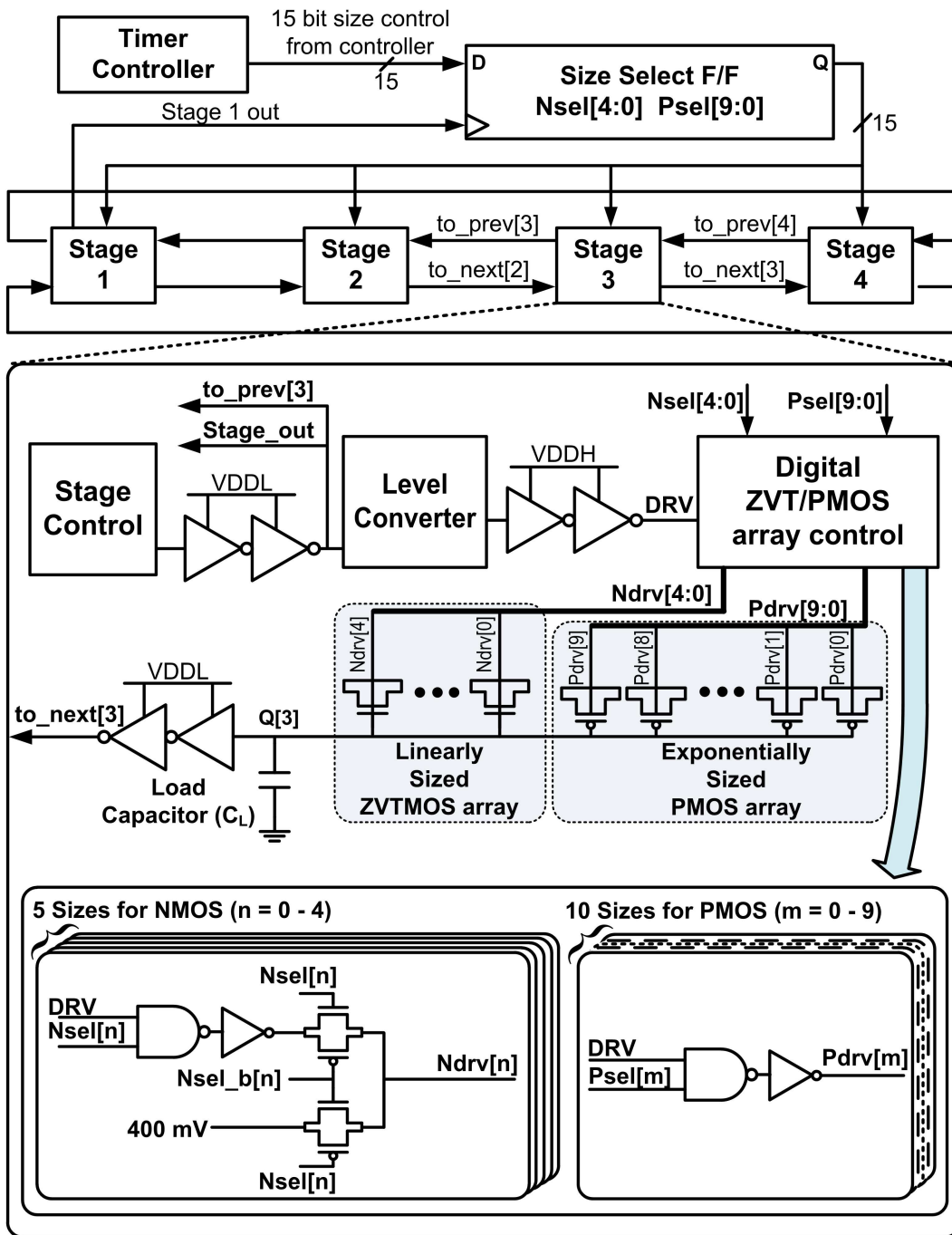


Figure 6.5: Circuit diagram of temperature compensated timer.

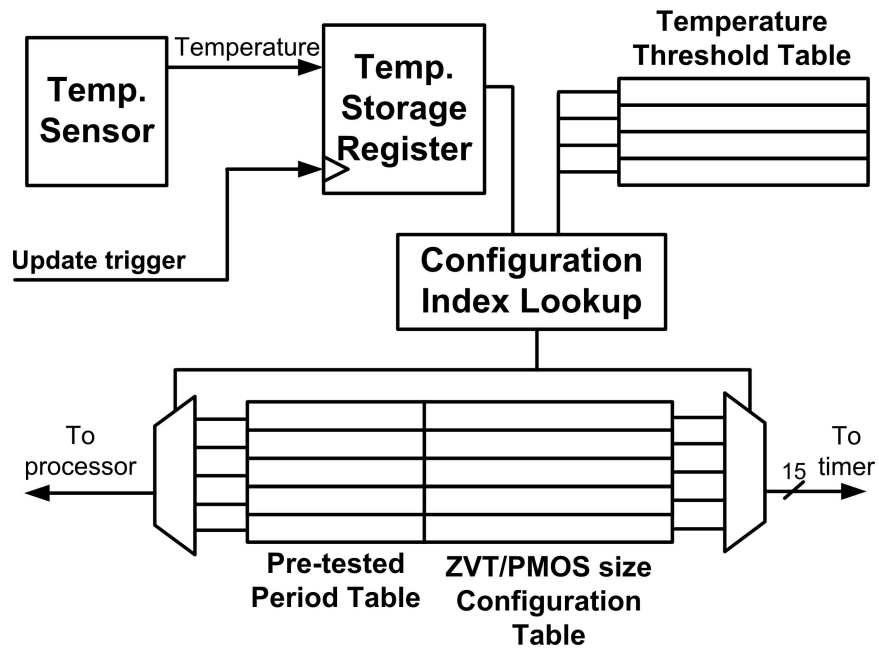


Figure 6.6: Controller for adaptive temperature compensation.

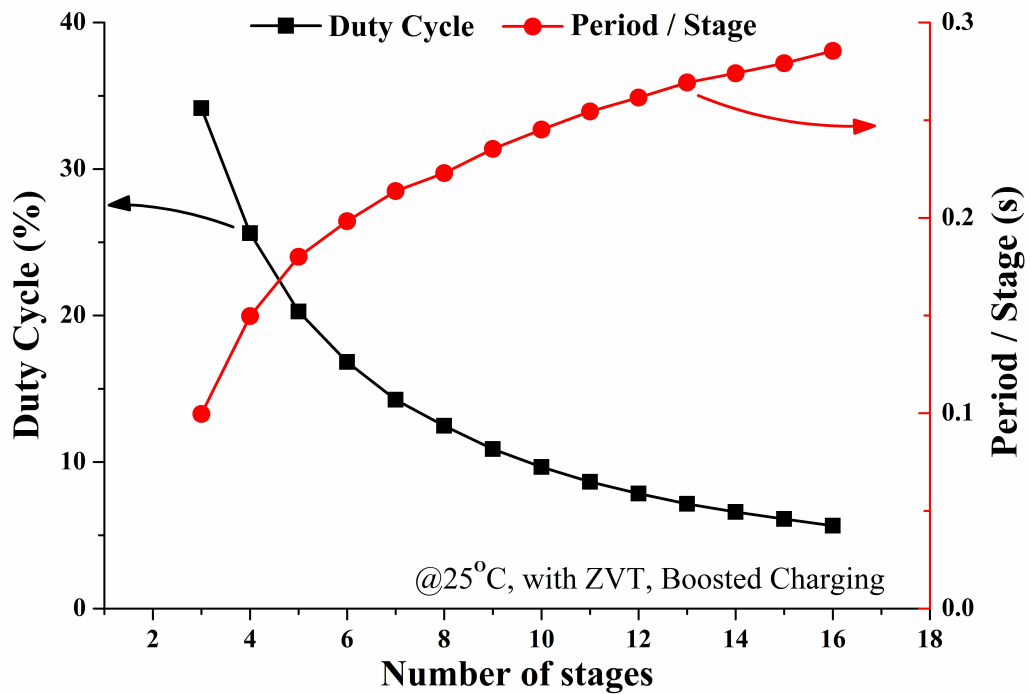


Figure 6.7: Duty cycle and period/stage change with number of stages.

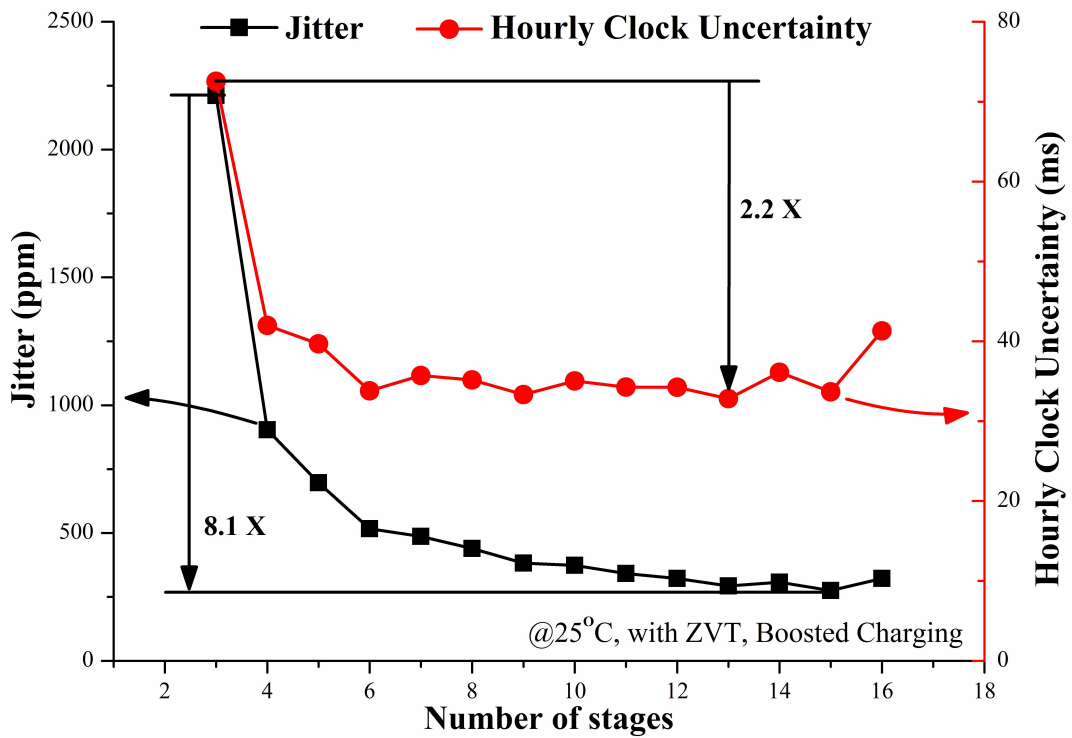


Figure 6.8: Jitter and hourly clock uncertainty reduction with multi-staging.

larger N , which lowers the initial C_L voltage at the beginning of charging. This also makes voltage swing wider hence period per number of stages larger as shown in Figure 6.7.

The jitter and uncertainty reduction with multi-stage timer is shown in Figure 6.8. With increasing number of stages, RMS jitter and hourly clock uncertainty is reduced by $8.1\times$ and $2.2\times$ respectively.

Figure 6.9 shows the jitter and uncertainty reduction by boosted charging. With large number of stages (>6), RMS jitter reduction was less than 50%. However, hourly clock uncertainty reduction was more than $3\times$, which is due to enhanced statistical averaging with shorter period.

Figure 6.10 clearly shows the trade-off between multi-stage and boosted timers. As number of stages is increased from 3 to 9, and boosted charging is utilized, hourly uncertainty is reduced by $3.6\times$ whereas power consumption is increased to 660pW. However, this power consumption is still well within idle mode power budget of state-of-the-art low power sensor nodes [7].

Figure 6.11 shows the power consumption of multi-stage gate-leakage based timer. With small stage counts (<5), power consumption increases with smaller number of stages. This is due to the

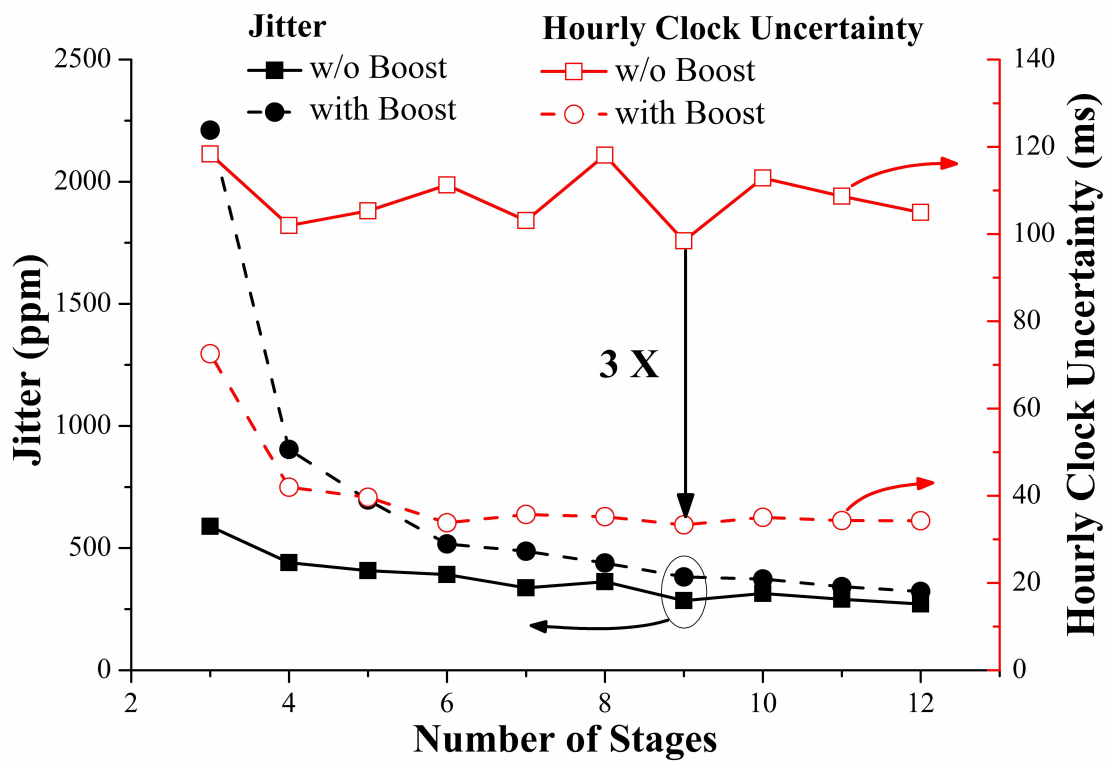


Figure 6.9: Jitter and hourly clock uncertainty reduction with boosted charging.

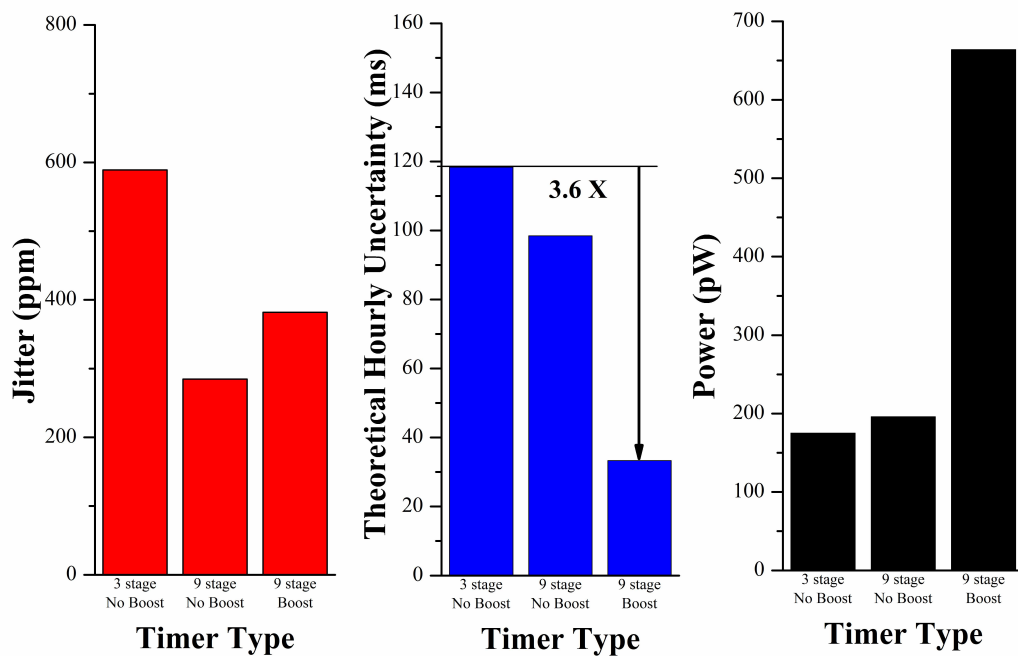


Figure 6.10: Trade-off between various types of timers.

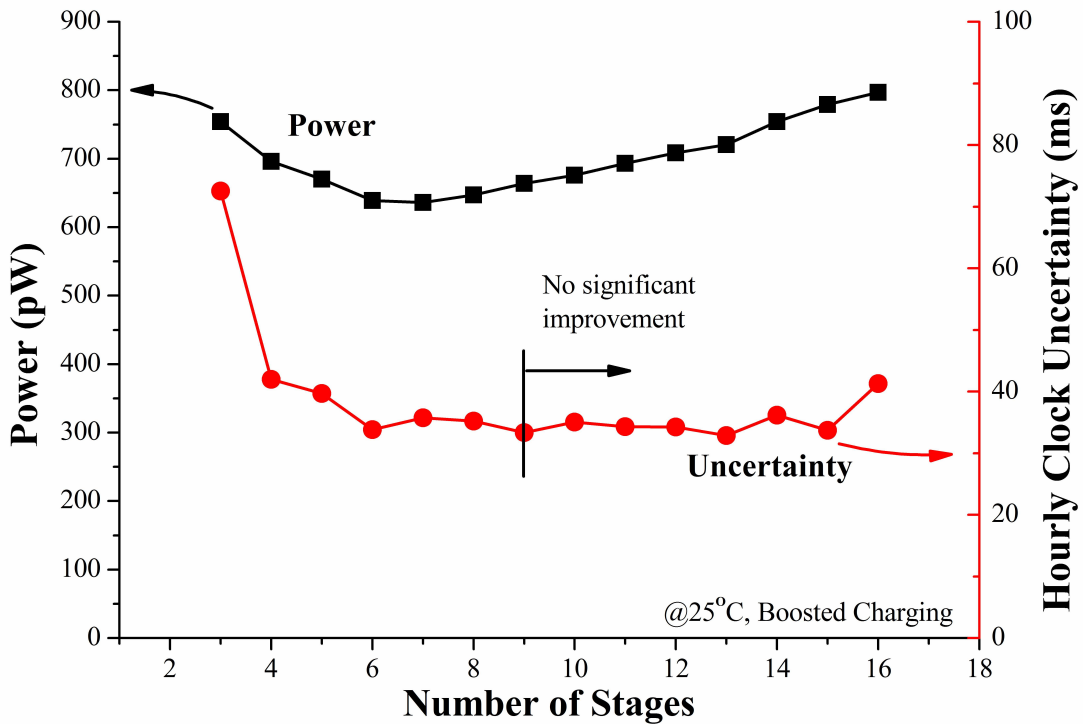


Figure 6.11: Power consumption of multi-stage gate-leakage based timer.

higher average node voltage of $Q[n]$ resulting in higher leakage current for the triggering buffer. With high stage counts (>7), power increases again due to additional static leakage of added stages. Having more stages than 9 does not significantly reduce the uncertainty while power consumption increases steadily with additional states. Therefore, a proposed MGT with 9 stages was chosen and tested for 24 hours and the SU for a large number of synchronization intervals are computed. A 3-stage MGT without boosted charging or ZVT transistors is also measured as a baseline timer.

Figure 6.12 shows the distribution of clock period for 24 hours. With this measurement, error for measuring 1 hour is computed which is equal to SU for 1 hour synchronization cycle time. The SU distribution had expected value of 196ms for 1 hour synchronization intervals (Figure 6.13). The theoretical uncertainty estimated by (6.1) and the actual uncertainty is compared in Figure 6.14. The proposed timer reduced the expected SU by $3.6\times$ compared to the baseline. Since the period of the timer is not truly Gaussian, the measured SU was larger than the theoretical calculation based on jitter. Measured SU is also reduced by $4.1\times$, confirming the effectiveness of proposed approach. The power supply sensitivity was $0.42\%/mV$ from 650mV to 750mV for low supply

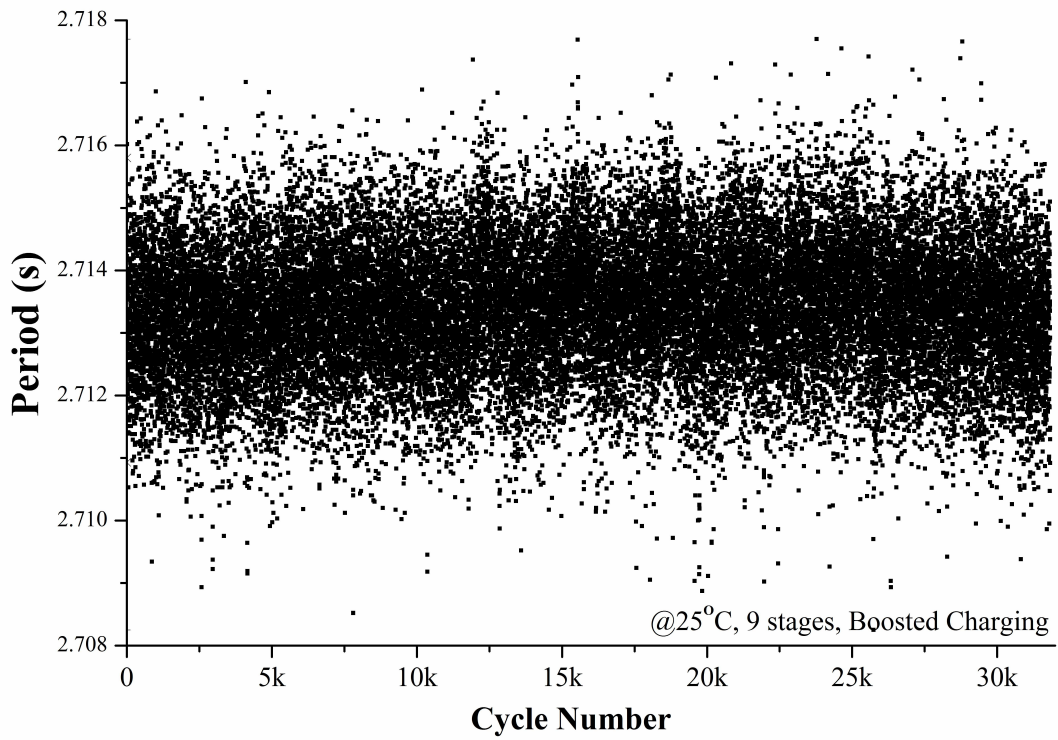


Figure 6.12: Distribution of period for 24 hour continuous measurement.

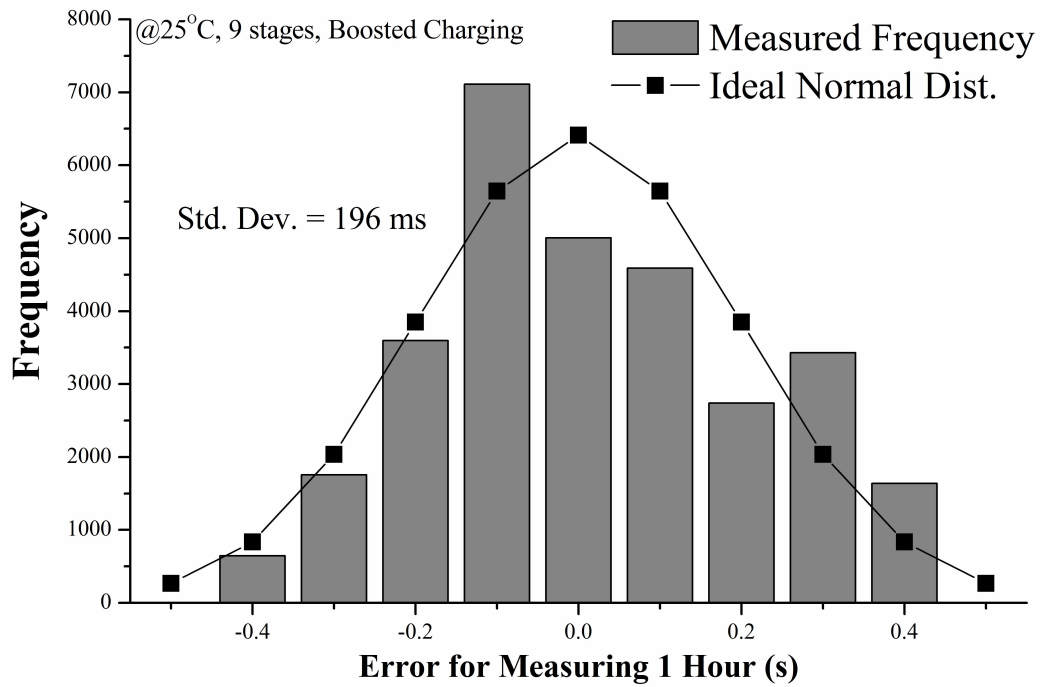


Figure 6.13: Distribution of error for measuring 1 hour synchronization cycle.

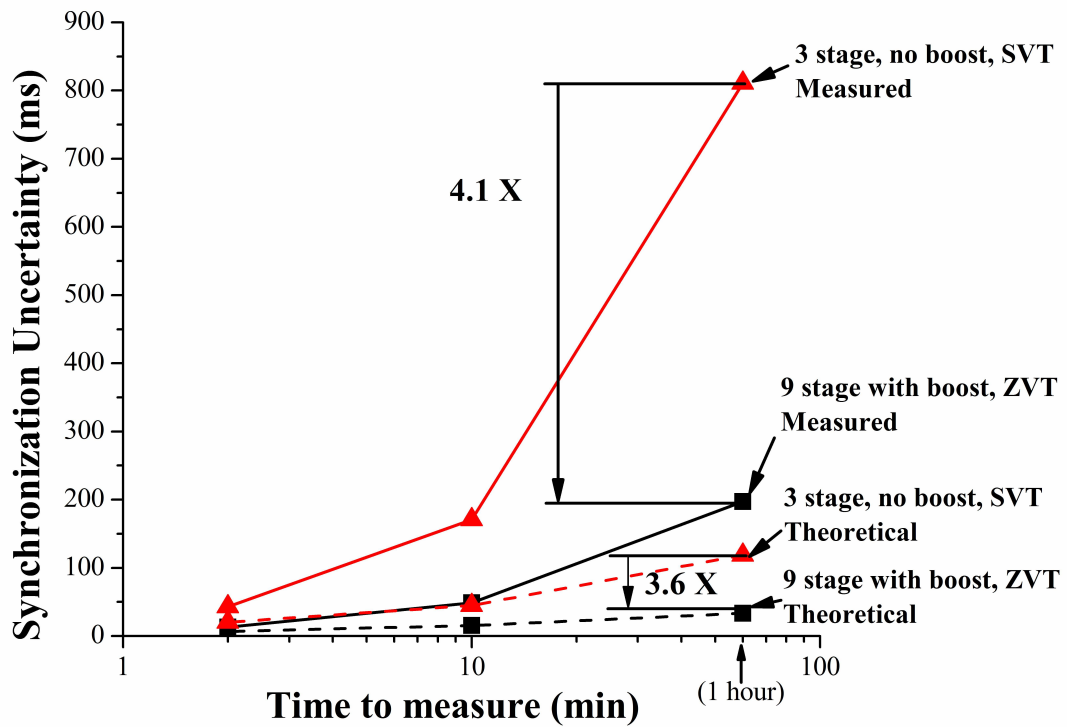


Figure 6.14: Theoretical and actual uncertainty with various timers.

and was 0.49%/mV from 1.15V to 1.25V for high supply. This necessitates the voltage regulation using an ultra-low power voltage reference such as the one proposed in [58].

6.4.2 Long-term Uncertainty

To verify effectiveness of multi-stage gate-leakage-based timer approach on sensor node synchronization timing, single and multi-stage timers are compared for long-term measurements. Figure 6.15 shows the standard deviation of error for measuring synchronization period for each type of timers. For measuring one hour synchronization period, single stage gate-leakage-based timer exhibits standard deviation of 913ms whereas multi-stage timer has 196ms. Multi-stage approach clearly improved the accuracy of the timer for entire measured range up to 1 hour. Uncertainty of timers can also be well characterized with Allan deviation [59] which is standard accuracy metric for fast oscillators. Figure 6.16 shows the Allan deviation of gate-leakage-based timers which also clearly shows that multi-stage approach significantly improve accuracy of the timers for examined time scope up to 1 hour.

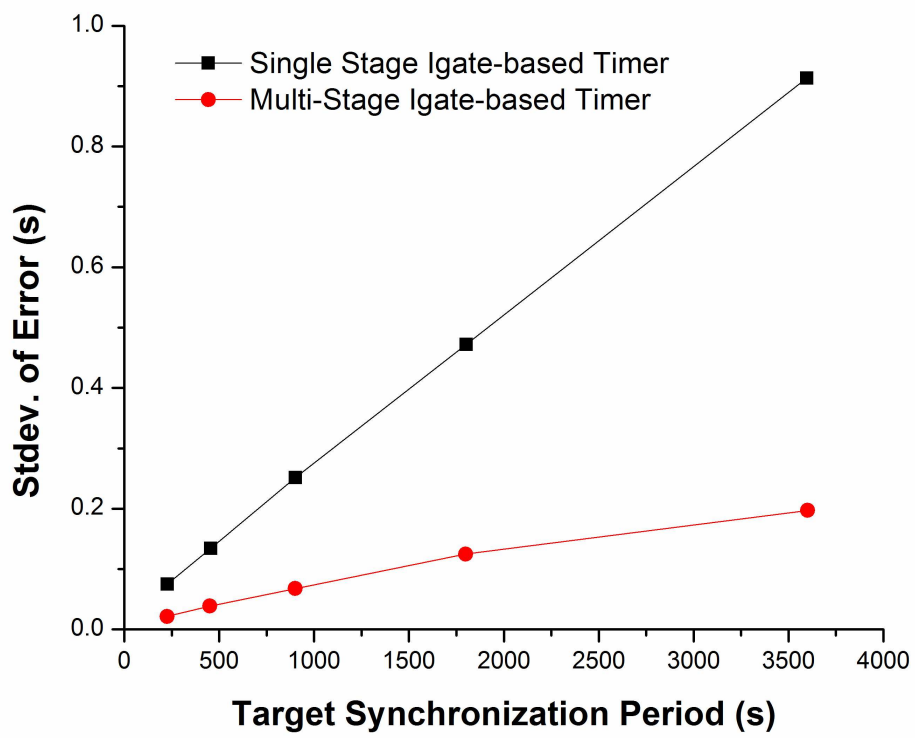


Figure 6.15: Standard deviation of synchronization error.

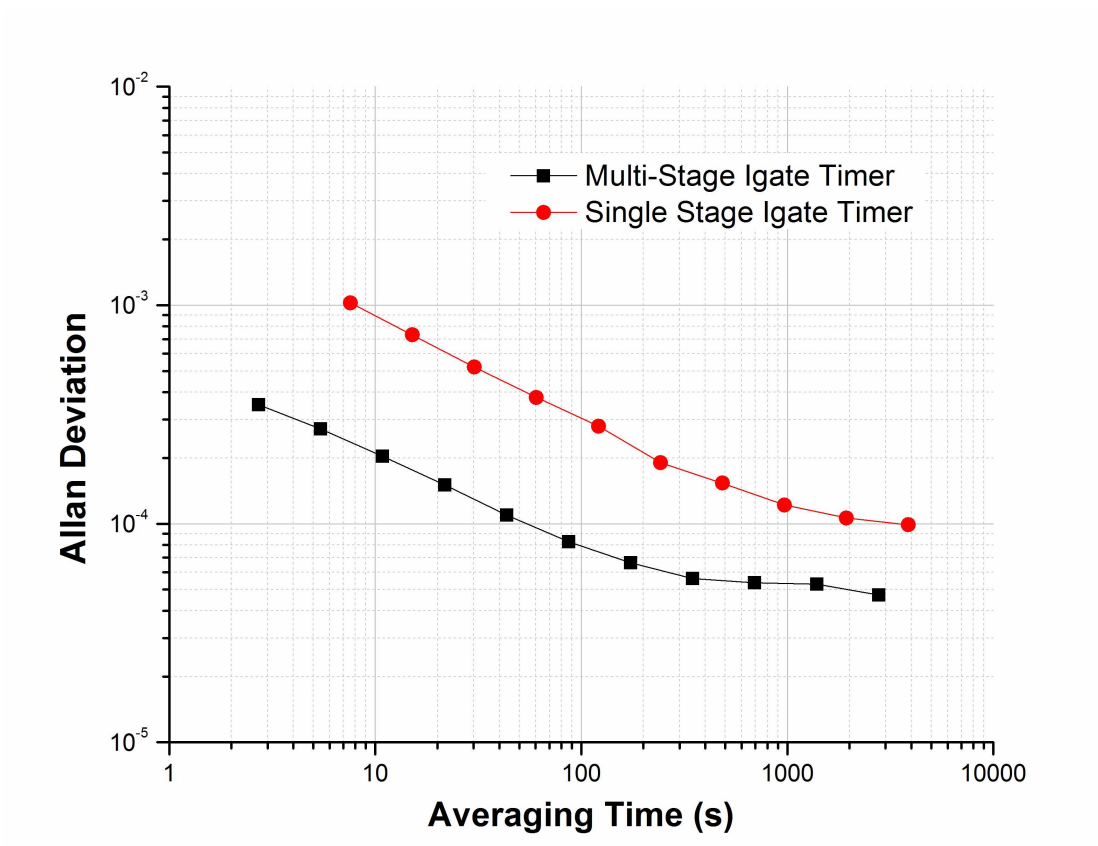


Figure 6.16: Allan deviation of gate-leakage-based timers.

6.4.3 Temperature Compensation

The period of the temperature compensated MGT for -20 – 60°C temperature range with selected configurations is shown in Figure 6.17. A five configuration scheme and its temperature range is shown as an example (Config.1-5 in Figure 6.17. For each configuration, period deviation as a function of temperature is shown in Figure 6.18 and worst period deviation was 0.28%. With a single configuration, the maximum deviation in period over -20 – 60°C was 3% with Config. 3 in Figure 6.17. With 5 configuration example, maximum deviation is reduced to 0.28%, and 10 configurations reduced this to 0.25%, giving an effective temperature sensitivity of 31ppm/ $^{\circ}\text{C}$ (Figure 6.19).

A closed loop timer control is tested with temperature profile as shown in Figure 6.20. As temperature changes between 20 – 30°C , closed loop controller switched among 4 different configurations with pre-set switching thresholds. Figure 6.21 shows the measured accumulated timing error with and without temperature compensation. With temperature compensation, maximum error was reduced by $4.8\times$.

6.4.4 Die to Die Variation of Gate-Leakage-Based Timer

For periodic synchronization of multiple wireless sensor nodes, timers in each sensor node should generate timestamps that agree with each other with acceptable error. However, due to the exponential dependency of gate leakage current on the gate oxide thickness, the period of a gate-leakage-based timer can significantly vary from die to die. Figure 6.22 shows the period distribution of a gate-leakage-based timer for 40 dies, when 3 stages are activated. With an average period of 304ms, a relatively large standard deviation of 86ms is observed. Therefore, the proposed gate-leakage-based timer requires die to die trimming or correction calibration to be used as valid time reference for synchronization among multiple sensor nodes.

6.5 Conclusion

Accurate measuring of radio communication interval is of critical concern for ultra-low power wireless sensor nodes. A multi-stage gate-leakage based timer is proposed with 660pW power con-

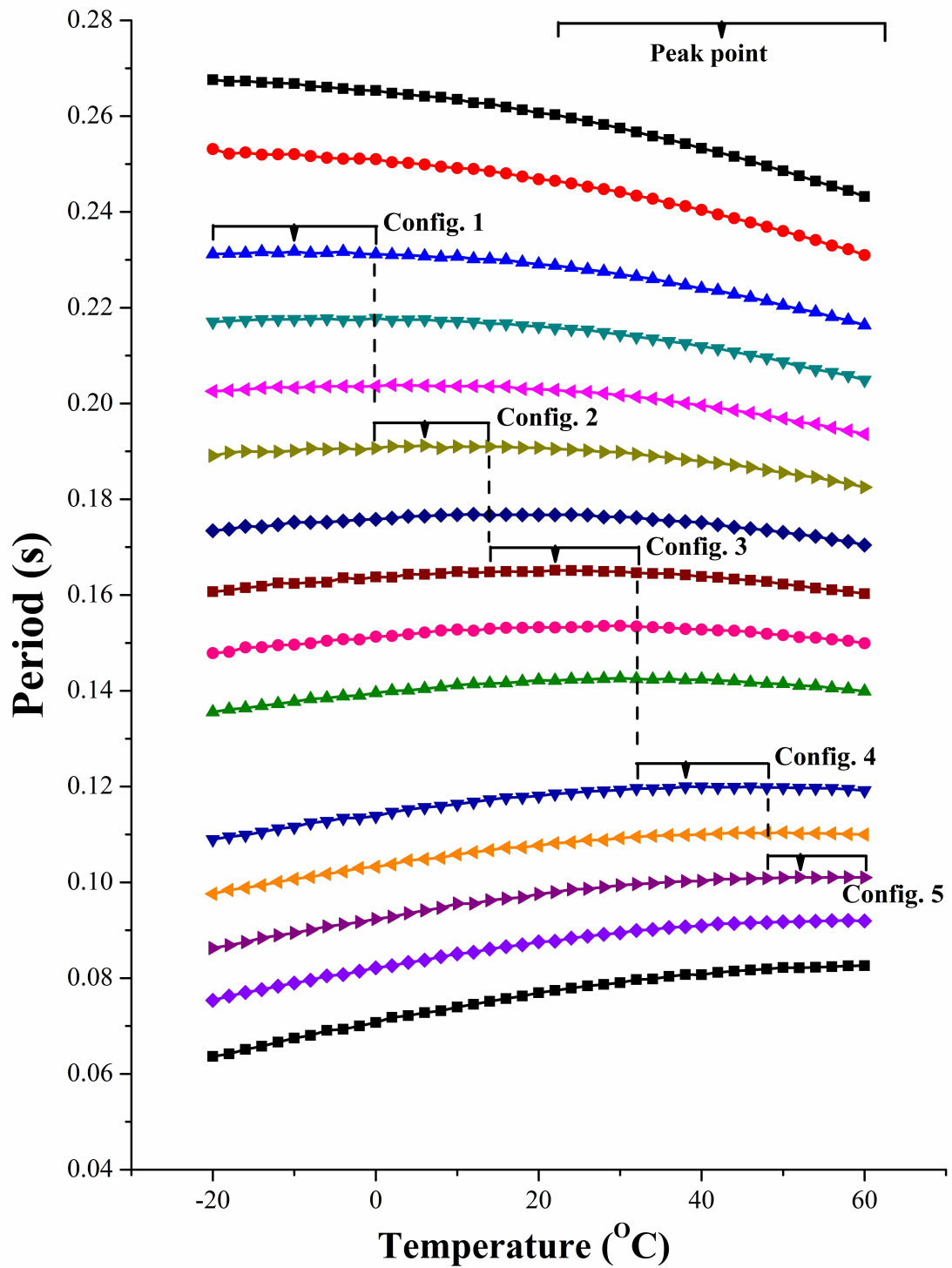


Figure 6.17: Period of temperature compensated timer with selected ZVTMOS/PMOS configurations.

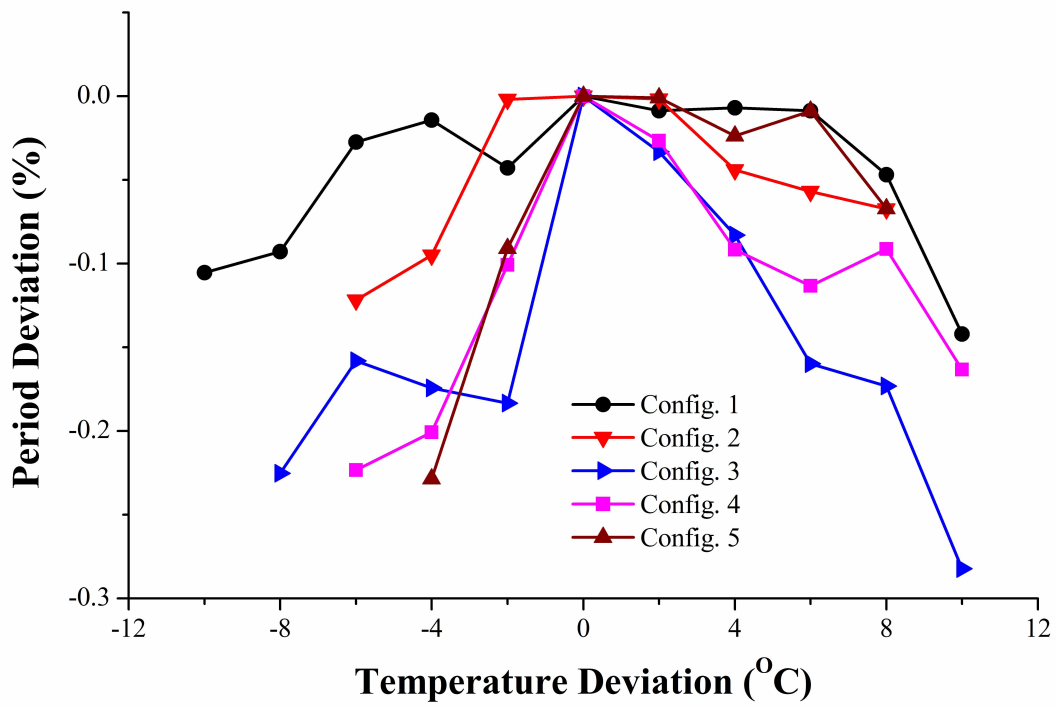


Figure 6.18: Period of deviation vs temperature deviation for selected configurations.

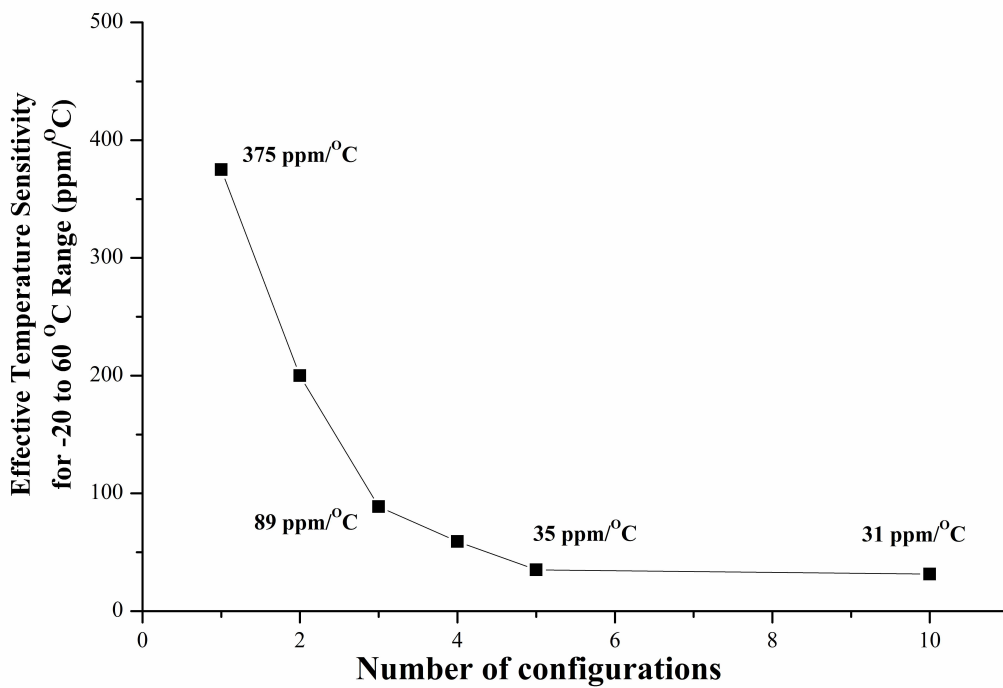


Figure 6.19: Effective temperature sensitivity for -20°C - 60°C vs number of configurations.

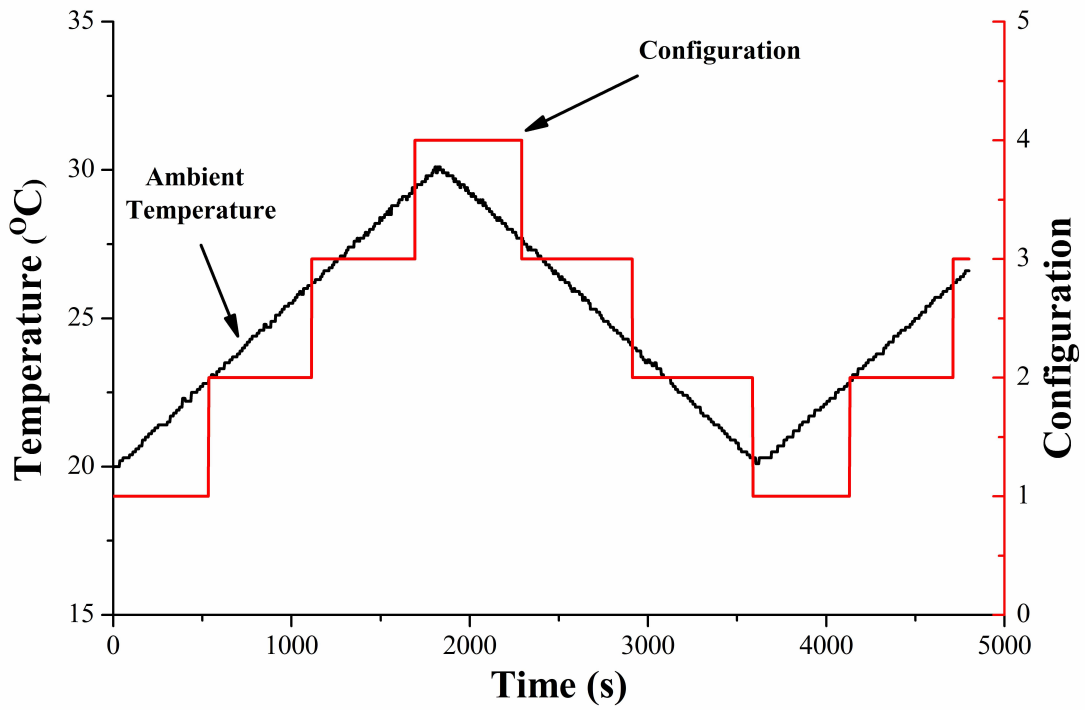


Figure 6.20: Temperature profile used for testing closed loop temperature compensation.

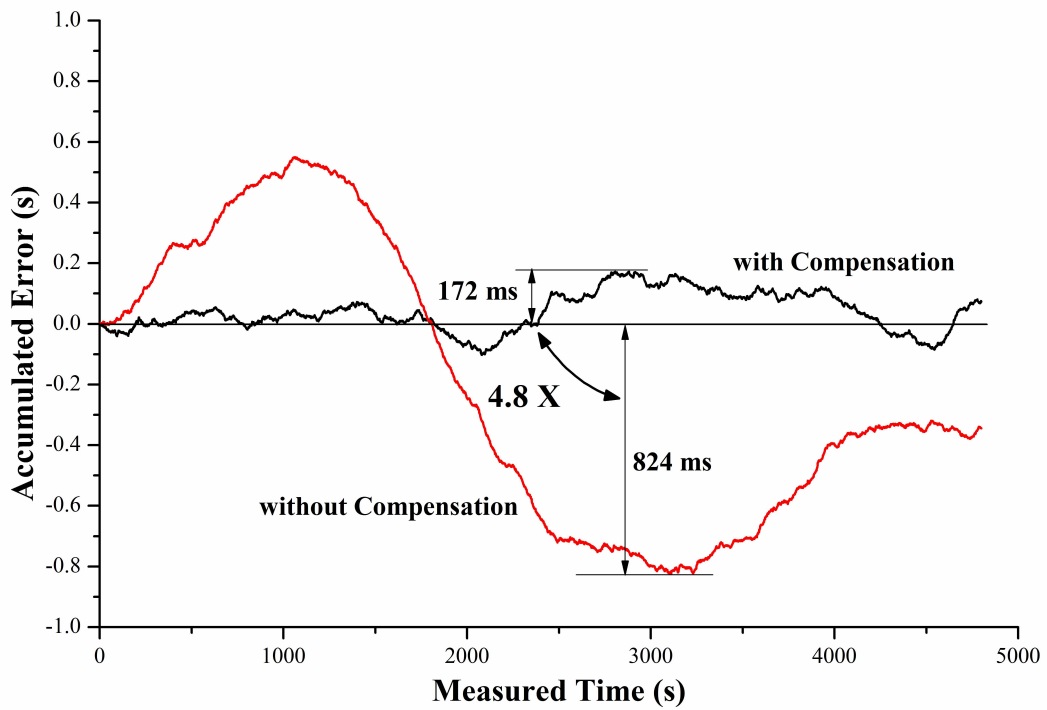


Figure 6.21: Accumulated timing error with and without temperature compensation.

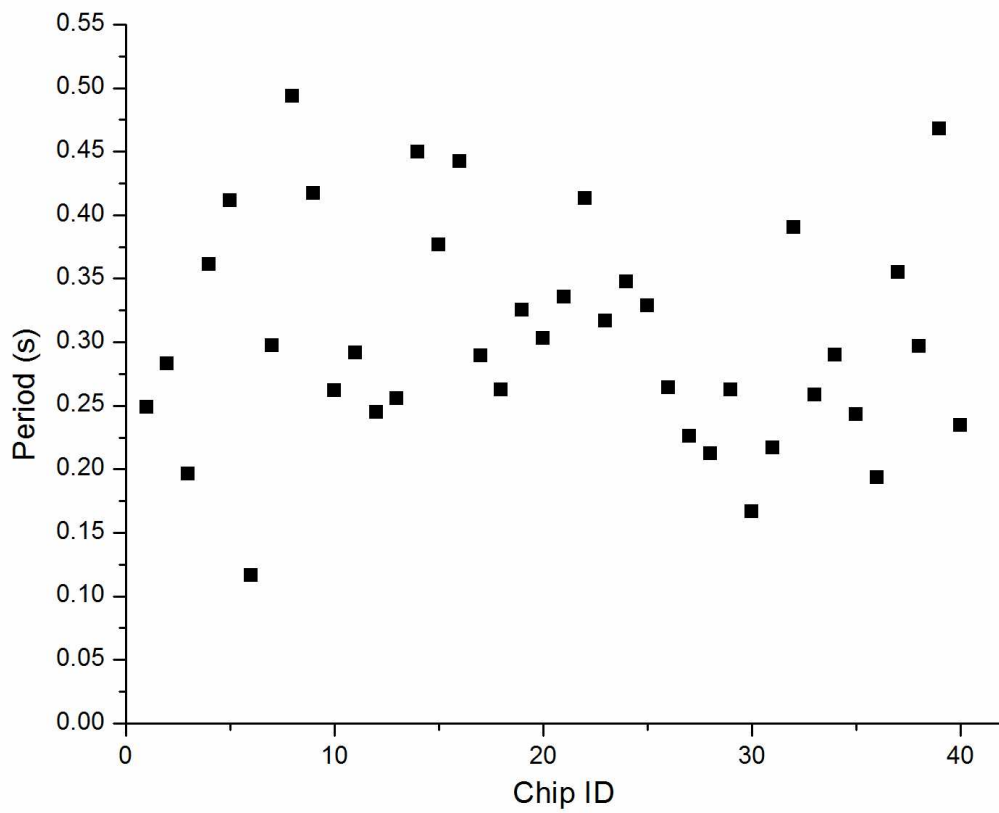


Figure 6.22: Distribution of gate-leakage-based timer period (number of active stages = 3).

sumption. With multi-stage structure and boosted charging, synchronization uncertainty has been reduced by $4.1\times$. An adaptive temperature compensation scheme that exploits the opposite temperature dependency of ZVTMOS and PMOS gate leakage current is demonstrated. With closed loop ZVTMOS/PMOS sizing control with 10 configurations, effective temperature dependency was $31\text{ppm}/^\circ\text{C}$.

CHAPTER 7

Conclusions

Modern daily life is surrounded by smaller and smaller computing devices. Laptops, for the first time, have provided mobility to computing devices and recent handheld devices drastically improved availability of computing power with their ‘hand’-sized form-factors. As Bell’s Law predicts, the research community is now looking at even smaller computing platforms and the mm³-scale sensor systems are drawing an increasing amount of attention since they can create a whole new computing environment as the next generation of smaller computers. Designing mm³-scale sensor nodes raises various circuit and system level challenges and we have addressed and proposed novel solutions for many of these challenges to create the first complete 1.0mm³ sensor system including commercial microprocessors as presented in Chapter 2.

We demonstrate a 1.0mm³ form factor sensor whose modular die-stacked structure allows maximum volume utilization. Low power I²C communication enables inter-layer serial communication with 88pJ/bit energy consumption without losing compatibility to standard I²C communication protocol. Dual ARM®Cortex-M0 microprocessors enable concurrent computation for the sensor node control and measurement data processing. A multi-modal power management unit allowed energy harvesting from various harvesting sources. An optical communication scheme is provided for initial programming, synchronization and re-programming after recovery from battery discharge. By adding an additional layer with application-specific sensor and/or wireless radio, the sensor system can be used for various applications.

Standby power reduction techniques are investigated in Chapter 3. A super cut-off power gating scheme with an ultra-low power charge pump reduces standby power of logic circuits by 2-19×

and memory by 30%. Different approaches for designing low-power memory for mm³-scale sensor nodes are presented in Chapter 4 and Chapter 5. A dual threshold voltage gain cell eDRAM design achieves the lowest eDRAM retention power and a 7T SRAM design based on hetero-junction tunneling transistors reduces the standby power of SRAM by 9-19× with only 15% area overhead.

We have paid special attention to the timer for the mm³-scale sensor systems, since a timer is a critical element to enable wireless sensor node synchronization and it has to be always on to track accurate timing. We propose a multi-stage gate-leakage-based timer in Chapter 6 to limit the standard deviation of the error in hourly measurement to 196ms and a temperature compensation scheme reduces temperature dependency to 31ppm/°C.

These techniques for designing ultra-low power circuits for a mm³-scale sensor enable implementation of a 1.0mm³ sensor node, which can be used as skeleton for future micro-sensor systems in variety of applications. These microsystems imply the continuation of the Bell's Law, which also predicts the massive deployment of mm³-scale computing systems and emergence of even smaller and more powerful computing systems in the near future. With ultra-low power circuit design, our daily life will eventually be surrounded by such microsystems, creating more convenient and abundant life with ubiquitous and pervasive computing.

BIBLIOGRAPHY

- [1] G. Bell, "Bell's Law for the Birth and Death of Computer Classes," *Communications of the ACM*, Vol 51, No. 1, pp. 86-94, Jan. 2008.
- [2] N. Mohamed, I. Jawhar, "A Fault Tolerant Wired/Wireless Sensor Network Architecture for Monitoring Pipeline Infrastructures," *International Conference on Sensor Technologies and Applications*, pp. 179-184, Aug. 2008.
- [3] Y. Tachwali, H. Refai, J. Fagan, "Minimizing HVAC Energy Consumption Using a Wireless Sensor Network," *33rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 439-444, Nov. 2007.
- [4] N. Elvin, N. Lajnef, A. Elvin, "Feasibility of structural monitoring with vibration powered sensors," *Smart Materials and Structures*, vol. 15, pp. 977-986, June 2006.
- [5] L. Schwiebert, S. Gupta, J. Weinmann, "Research challenges in wireless networks of biomedical sensors," *International Conference on Mobile Computing and Networking*, 2001.
- [6] S. Hanson, M. Seok, Y.-S. Lin, Z.Y. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, D. Blaauw, "A low-voltage processor for sensing applications with picowatt standby mode," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1145-1155, April 2009.
- [7] G. Chen, M. Fojtik, D. Kim, D. Fick, J. Park, M. Seok, M. Chen, Z. Foo, D. Sylvester, D. Blaauw, "Millimeter-scale nearly perpetual sensor system with stacked battery and solar cells," *IEEE International Solid-State Circuits Conference*, pp. 288-289, Feb. 2010.
- [8] G. Chen, H. Ghaed, R. Haque, M. Wieckowski, Y. Kim, G. Kim, D. Fick, D. Kim, M. Seok, K. Wise, D. Blaauw, D. Sylvester, "A cubic-millimeter energy-autonomous wireless intraocular pressure monitor," *IEEE International Solid-State Circuits Conference*, pp. 310-312, Feb. 2011.
- [9] Cymbet Corporation, "Rechargeable Thin Film Battery 12Ah, 3.8V," EnerChip CBC012 datasheet, 2009.
- [10] L. Chang, R.K. Montoye, Y. Nakamura, K.A. Batson, R.J. Eickemeyer, R.H. Dennard, W. Haensch, and D. Jamsek, "An 8T-SRAM for variability tolerance and low-voltage operation in high-performance caches," *IEEE Journal of Solid-State Circuits*. vol. 43, no. 4, pp.956-963, Apr. 2008.
- [11] Y. Lee, M. Seok, S. Hanson, D. Blaauw, D. Sylvester, "Standby power reduction techniques for ultra-low power processors," *IEEE European Solid-State Circuits Conference*, pp. 186-189, Sep. 2008.
- [12] Y. Lee, M.-T. Chen, J. Park, D. Sylvester, D. Blaauw, "A 5.42nW/kB retention power logic-compatible embedded DRAM with 2T dual-Vt gain cell for low power sensing applications," *IEEE Asian Solid-State Circuits Conference*, Nov. 2010.
- [13] D. Kim, Y. Lee, J. Cai, I. Lauer, L. Chang, S.J. Koester, D. Sylvester, D. Blaauw, "Low power circuit design based on heterojunction tunneling transistors (HETTs)," *ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED)*, pp. 219-224, Aug. 2009.

- [14] Y. Lee, B. Giridhar, Z. Foo, D. Sylvester, D. Blaauw, "A 660pW multi-stage temperature-compensated timer for ultra-low-power wireless sensor node synchronization," *IEEE International Solid-State Circuits Conference*, pp. 46-48, Feb. 2011.
- [15] Y. Lee, D. Sylvester, D. Blaauw, "Synchronization of ultra-low power wireless sensor nodes," *IEEE International Midwest Symposium on Circuit and Systems*, Aug. 2011.
- [16] Y. Lee, G. Kim, S. Bang, Y. Kim, I. Lee, P. Dutta, D. Sylvester, D. Blaauw, "A Modular 1mm³ die-stacked sensing platform with optical communication and multi-modal energy harvesting," *IEEE International Solid-State Circuits Conference*, to appear, Feb. 2012.
- [17] B. Warneke, M. Last, B. Liebowitz, K. Pister, "Smart Dust: communicating with a cubic-millimeter computer," *IEEE Computer*, vol. 34, pp. 44-51, Jan. 2001.
- [18] E.Y. Chow, S. Chakraborty, W.J. Chappell, P.P. Irazoqui, "Mixed-signal integrated circuits for self-contained sub-cubic millimeter biomedical implants," *IEEE International Solid-State Circuits Conference*, pp. 236-237, Feb. 2010.
- [19] NXP Semiconductors, "I²C-bus specification and user manual" UM10204 datasheet, Rev. 03, Jun. 2007.
- [20] A. Wang, A. Chandrakasan, "A 180mV FFT processor using subthreshold circuit techniques," *IEEE International Solid-State Circuits Conference*, pp. 292-293, Feb. 2004.
- [21] M. Seok, S. Hanson, D. Sylvester, D. Blaauw, "Analysis and optimization of sleep modes in subthreshold circuit design," *Design Automation Conference*, pp. 694-699, 2007.
- [22] M. Seok, S. Hanson, Y. Lin, Z. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, D. Blaauw, "The Phoenix Processor: A 30pW platform for sensor applications," *IEEE Symposium on VLSI Circuits*, pp. 188-189, June 2008.
- [23] H. Kawaguchi, K.-I. Nose, T. Sakurai, "A super cut-off CMOS (SCCMOS) scheme for 0.5-V supply voltage with picoampere stand-by current," *IEEE International Solid-State Circuits Conference*, pp. 192-193, Feb. 1998.
- [24] Bo Zhai, D. Blaauw, D. Sylvester, S. Hanson, "A Sub-200mV 6T SRAM in 0.13nm CMOS," *IEEE International Solid-State Circuits Conference*, pp. 332-333, Feb. 2007.
- [25] 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, K.W. Guarini, W. Haensch, "Stable SRAM cell design for the 32 nm node and beyond," *IEEE Symposium on VLSI Circuits*, pp. 128-129, June 2005.
- [26] B.H. Calhoun, A. Chandrakasan, "A 256kb Sub-threshold SRAM in 65nm CMOS," *IEEE International Solid-State Circuits Conference*, pp. 2592-2601, Feb. 2006.
- [27] N. Verma, A. Chandrakasan, "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy," *IEEE Journal of Solid-State Circuits*, pp. 141-149, Jan. 2008.

- [28] K. Chun, P. Jain, J. Lee, C. Kim, "A sub-0.9V logic-compatible embedded DRAM with boosted 3T gain cell, regulated bit-line write scheme and PVT-tracking read reference bias," *IEEE Symposium on VLSI Circuits*, pp. 134-135, June 2009.
- [29] D. Somasekhar, Y. Ye, P. Aseron, S. Lu, M. Khellah, J. Howard, G. Ruhl, T. Karnik, S.Y. Borkar, V. De, A. Keshavarzi, "2GHz 2Mb 2T Gain-Cell Memory Macro with 128GB/s Bandwidth in a 65nm Logic Process," *IEEE International Solid-State Circuits Conference*, pp. 274-275, Feb. 2008.
- [30] K. Chun, P. Jain, J. Lee, C. Kim, "An experimental 2T cell RAM with 7 ns access time at low temperature," *IEEE Symposium on VLSI Circuits*, pp. 134-135, June 1990.
- [31] M. Seok, D. Sylvester, D. Blaauw, "Optimal technology selection for minimizing energy and variability in low voltage applications," *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, pp. 9-14, Aug 2008.
- [32] M. Kaku, H. Iwai, T. Nagai, M. Wada, A. Suzuki, T. Takai, N. Itoga, T. Miyazaki, T. Iwai, H. Takenaka, T. Hojo, S. Miyano, N. Otsuka, "An 833MHz Pseudo-Two-Port Embedded DRAM for Graphics Applications," *IEEE International Solid-State Circuits Conference*, pp. 276-613, Feb. 2008.
- [33] J. Barth, W. Reohr, P. Parries, G. Fredeman, J. Golz, S. Schuster, R. Matick, H. Hunter, C. Tanner, J. Harig, H. Kim, B. Khan, J. Griesemer, R. Havreluk, K. Yanagisawa, T. Kirihata, S. Iyer, "A 500MHz Random Cycle 1.5ns-Latency, SOI Embedded DRAM Macro Featuring a 3T Micro Sense Amplifier," *IEEE International Solid-State Circuits Conference*, pp. 486-617, Feb. 2007.
- [34] I.J. Chang, J.-J. Kim, S.P. Park, K. Roy, "A 32kb 10T Subthreshold SRAM Array with Bit-Interleaving and Differential Read Scheme in 90nm CMOS," *IEEE International Solid-State Circuits Conference*, pp. 388-389, Feb. 2008.
- [35] Y. Pu, J.P. de Gyvez, H. Corporaal, Y. Ha, "An Ultra-Low-Energy/Frame Multi-Standard JPEG Co-Processor in 65nm CMOS with Sub/Near-Threshold Power Supply," *IEEE International Solid-State Circuits Conference*, pp. 146-147, Feb. 2009.
- [36] H. Kaul, M. Anders, S. Mathew, S. Hsu, A. Agarwal, R. Krishnamurthy, S. Borkar, "A 300mV 494GOPS/W Reconfigurable Dual-Supply 4-Way SIMD Vector Processing Accelerator in 45nm CMOS," *IEEE International Solid-State Circuits Conference*, pp. 260-261, Feb. 2009.
- [37] B. Zhai, D. Blaauw, D. Sylvester, K. Flautner, "Theoretical and Practical Limits of Dynamic Voltage Scaling," *Design Automation Conferen*, pp. 868-873, 2004.
- [38] C. Hu, P. Patel, A. Bowonder, K. Jeon, S.H. Kim, W.Y. Loh, C.Y. Kang, J. Oh, P. Majhi, A. Javey, T.-J.K. Liu, R. Jammy, "Prospect of tunneling green transistor for 0.1V CMOS," *IEEE International Electron Device Meeting*, pp. 16.1.1-16.1.4, 2005.

- [39] L. Leem, A. Srivastava, S. Li, B. M.-K., G. Iannaccone, J.S. Harris, G. Fiori, "Multi-scale simulation of partially unzipped CNT hetero-junction Tunneling Field Effect Transistor," *IEEE International Electron Device Meeting*, pp. 32.5.1-32.5.4, 2005.
- [40] W. Choi, J. Song, J. Lee, Y. Park, B. Park, "70-nm Impact-Ionization Metal-oxide-semiconductor (I-MOS) Devices Integrated with Tunneling Field-Effect Transistors (TFETs)," *IEEE International Electron Device Meeting*, pp. 955-958, 2005.
- [41] J. Knoch, S. Mantl, J. Appenzeller, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics*, pp. 572-578, 2007.
- [42] E. Tho, G. Wang, L. Chan, G. Lo, G. Samudra, Y. Yeo, "I-MOS Transistor With an Elevated Silicon-Germanium Impact-Ionization Region for Bandgap Engineering," *IEEE Electron Device Letters*, pp. 975-977, Dec. 2006.
- [43] O. Nayfeh, C. Chleirigh, J. Hennessy, L. Gomez, J. Hoyt, D. Antoniadis, "Design of Tunneling Field-Effect Transistors Using Strained-Silicon/Strained-Germanium Type-II Staggered Heterojunctions," *IEEE Electron Device Letters*, pp. 1074-1077, Sep. 2008.
- [44] T. Krishnamohan, D. Kim, S. Raghunathan, K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With Record High Drive Currents and ≈ 60 mV/dec Sub-threshold Slope," *IEEE International Electron Device Meeting*, pp. 947-949, 2008.
- [45] F. Mayer, C. Royer, J. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, S. Deleonibus, "Impact of SOI, Si_{1-x}Ge_xOI and GeOI substrates on CMOS compatible Tunnel FET performance," *IEEE International Electron Device Meeting*, pp. 163-166, 2008.
- [46] M. M. Rieger, P. Vogl, "Electronic-band parameters in strained Si_{1-x}Ge_x alloys on Si_{1-y}Ge_y substrates," *Physical Review B*, pp. 14276-14287, Nov. 1993.
- [47] M. Jeong, P. Solomon, S. Laux, H. Wong, D. Chidambarao, "Comparison of raised and Schottky source/drain MOSFETs using a novel tunneling contact model," *IEEE International Electron Device Meeting*, pp. 733-736, 1998.
- [48] M. V. Fischetti, S. E. Laux, "Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys," *Journal of Applied Physics*, pp. 2234-2252, Aug. 1996.
- [49] J. Lin, E.H. Toh, C. Shen, D. Sylvester, C.H. Heng, G. Samudra, Y.C. Yeo, "Compact HSPICE model for IMOS device," *Electronics Letters*, pp. 91-92, Jan. 2008.
- [50] L. Chang, Y. Nakamura, R.K. Montoye, J. Sawada, A.K. Martin, K. Kinoshita, F.H. Gebara, K.B. Agarwal, D.J. Acharyya, W. Haensch, K. Hosokawa, D. Jamsek, "A 5.3GHz 8T-SRAM with Operation Down to 0.41V in 65nm CMOS," *IEEE Symposium on VLSI Circuits*, pp. 252-253, Jun. 2007.
- [51] D. Kim, V. Chandra, R. Aitken, D. Blaauw, D. Sylvester, "Variation-Aware Static and Dynamic Writability Analysis for Voltage-Scaled Bit-Interleaved 8-T SRAMs" *IEEE International Symposium on Low Power Electronic Designs*, pp. 145-150. 2011.

- [52] B.A. Warneke, K.S.J. Pister, "An ultra-low energy microcontroller for Smart Dust wireless sensor networks," *IEEE International Solid-State Circuits Conference*, pp. 316-317, Feb. 2004.
- [53] M. Crepaldi, C. Li, K. Dronson, J. Fernandes, P. Kinget, "An Ultra-Low-Power interference-robust IR-UWB transceiver chipset using self-synchronizing OOK modulation," *IEEE International Solid-State Circuits Conference*, pp. 226-227, Feb. 2010.
- [54] M.S. McCorquodale, S.M. Pernia, J.D. O'Day, G. Carichner, E. Marsman, N. Nguyen, S. Kubba, S. Nguyen, J. Kuhn, R.B. Brown, "A 0.5-to-480MHz Self-Referenced CMOS Clock Generator with 90ppm Total Frequency Error and Spread-Spectrum Capability," *IEEE International Solid-State Circuits Conference*, pp. 350-619, Feb. 2008.
- [55] Y. Lin, D. Sylvester, D. Blaauw, "A sub-pW timer using gate leakage for ultra low-power sub-Hz monitoring systems," *IEEE Custom Integrated Circuits Conference*, pp. 397-400, Sep. 2007.
- [56] C.-H. Choi, Ki-Y. Nam, Z. Yu, R.W. Dutton, "Impact of gate direct tunneling current on circuit performance : a Simulation Study," *IEEE Transactions on Electron Devices*, Vol. 48, No. 12, pp. 2823-2829, Dec. 2001.
- [57] Y. Lin, D. Sylvester, D. Blaauw, "A 150pW program-and-hold timer for ultra-low-power sensor platforms," *IEEE International Solid-State Circuits Conference*, pp. 326-327, Feb. 2009.
- [58] M. Seok, G. Kim, D. Sylvester, D. Blaauw, "A 0.5V 2.2pW 2-transistor voltage reference," *IEEE Custom Integrated Circuits Conference*, pp. 577-580, Sep. 2009.
- [59] D.W. Allan "Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-34, No. 6, pp. 647-654, Nov. 1987.