

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Precision On-Chip PVT sensors

Cristina Chen Cai



Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Prof. Vitor Tavares

July 24, 2026

Precision On-Chip PVT sensors

Cristina Chen Cai

Mestrado em Engenharia Eletrotécnica e de Computadores

July 24, 2026

Resumo

Com o crescimento exponencial da Internet das Coisas (IoT), os Sistemas-em-Chip (SoCs) de hoje em dia enfrentam desafios de fiabilidade devido a variações de Processo, Tensão e Temperatura (PVT). Esta dissertação aborda estes desafios através da realização de um estudo abrangente sobre técnicas de monitorização on-chip e da proposta de uma infraestrutura de sensorização PVT unificada, concebida para nós nanométricos avançados.

O trabalho inicia-se com uma análise detalhada do impacto das variações PVT no desempenho e no rendimento (*yield*) dos circuitos, demonstrando a necessidade de autocalibração adaptativa. Segue-se uma revisão crítica das topologias de sensores no estado da arte, avaliando os seus compromissos (*trade-offs*) em termos de potência, área e sensibilidade. Com base neste enquadramento teórico, é proposta uma arquitetura de monitorização PVT unificada, sequenciada por uma máquina de estados finitos (FSM), que utiliza uma Referência de Bandgap (BGR) de alta estabilidade para estabilizar os três blocos de sensorização.

O projeto proposto é desenvolvido e validado ao nível do transístor em tecnologia FinFET de 16-nm, com cada bloco de sensorização redesenhado a partir de princípios fundamentais e comparado quantitativamente com a topologia de referência correspondente do estado da arte. A infraestrutura integra três sub-sistemas de sensorização: um monitor de processo baseado em osciladores de anel com células de atraso (*Diode Connected Oscillator*), simplificado a partir das células ratioed NAND/NOR do estado da arte para uma estrutura de transístor único; um sensor de tensão digital por cadeia de atraso de transístores corrente-limitada (*current-starved*) com código termométrico, estabilizado por uma BGR nativa para FinFET; e um sensor de temperatura baseado num oscilador de anel, polarizado por um núcleo PTAT de transístores compostos selecionado pela sua rejeição de ruído na linha de alimentação normalizada. Acoplada a uma FSM hierárquica que sequencia a sensorização de processo, temperatura e tensão, esta abordagem visa mitigar a ambiguidade de diagnóstico transdomínio. Esta tese estabelece as restrições de projeto fundamentais, as escolhas arquiteturais, e os resultados de simulação obtidos para cada bloco individual, identificando também as limitações concretas e o trabalho futuro necessário para alcançar a integração em malha fechada da plataforma unificada proposta.

Abstract

With the exponential growth of the Internet of Things (IoT), modern Systems-on-Chip (SoCs) are facing critical reliability challenges due to Process, Voltage, and Temperature (PVT) variations. This dissertation addresses these challenges by conducting a comprehensive study of on-chip monitoring techniques and proposing a unified PVT sensing infrastructure designed for advanced nanometric nodes.

The work starts with an in-depth analysis of the impact of PVT variations on circuit performance and yield, demonstrating the necessity for adaptive self-calibration. It then provides a critical review of state-of-the-art sensing topologies, evaluating their trade-offs in terms of power, area, and sensitivity. Based on this theoretical framework, a unified PVT monitoring architecture is proposed, sequenced by a Finite State Machine (FSM).

The proposed design is developed and validated at the transistor level in 16-nm FinFET technology, with each sensing block re-derived from first principles and benchmarked quantitatively against the corresponding state-of-the-art topology. The infrastructure integrates three sensing subsystems: a process monitor built around a Diode Connected Oscillator, which simplifies the ratioed NAND/NOR delay cells of the state of the art into a single-transistor-type structure; a fully digital, thermometer-coded delay-chain voltage sensor built on a current-starved delay chain and stabilized by a FinFET-native N-well diode BGR; and a ring-oscillator-based temperature sensor biased by a composite-transistor PTAT current core, selected for its normalised supply-noise rejection. Coupled with a hierarchical FSM that sequences process, temperature, and voltage sensing, this approach is aimed at mitigating cross-domain diagnostic ambiguity. This thesis establishes the foundational design constraints, architectural choices, and simulation results obtained for each individual block, while also identifying the concrete limitations and future work required to achieve full integration of the proposed unified platform.

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework to achieve a better and more sustainable future for all. It includes 17 goals¹ to address the world's most pressing challenges, including poverty, inequality, climate change, environmental degradation, peace, and justice.

This dissertation addresses the study and integration of existing Process, Voltage, and Temperature (PVT) sensor topologies into a unified, complete on-chip sensing solution for sub-micron nodes, enabling lifetime IC characterization and reducing reliance on lengthy external ATE and wafer-level testing. As IoT and low-power device production scales up, this contributes to more efficient and resource-conscious semiconductor testing and manufacturing.

The specific Sustainable Development Goals addressed by this work have the following names:

SDG 9 Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation

SDG 12 Ensure sustainable consumption and production patterns

SDG	Target	Contribution	Performance Indicators and Metrics
9	9.5	Study, integration, and simulation of state-of-the-art PVT sensor topologies into a coordinated on-chip PVT sensing infrastructure, upgrading the technological capability available for IC test and characterization.	Sensor accuracy/resolution achieved (mV, °C); successful integration of sensors into a complete validated solution
12	12.2	On-chip PVT monitoring reduces reliance on external ATE test cycles, supporting more efficient use of test-equipment time and energy across the IC production and operation cycle.	Reduction in ATE test time/energy per chip; reduction in design over-margining

¹<https://sdgs.un.org/goals>

Acknowledgements

I would like to begin by thanking my boyfriend, Vasilen, for his endless support, patience, and encouragement throughout the course of this dissertation.

I am deeply grateful to my mother for her unwavering support and encouragement throughout my entire academic journey. None of this would have been possible without her.

I extend my sincere gratitude to my advisor, Professor Vítor Tavares, for his guidance, insightful feedback, and academic rigor throughout the development of this work.

I would also like to thank my supervisors at Synopsys, Engineer Ricardo Almeida and Engineer Daniele Marchetti, for their technical mentorship, their constant availability to discuss circuit-level design decisions, and for providing the tools, environment, and industrial perspective that made the simulation work in this dissertation possible.

Finally, I acknowledge that AI-based tools were used during the preparation of this document to assist with grammar and writing clarity. All technical content, analyses, design decisions, and results presented in this dissertation remain entirely my own.

Cristina Chen

Contents

1	Introduction	1
1.1	Scientific Context	2
1.2	System-Level Sequencing and FSM Architecture	3
1.3	Main Objectives of This Work	5
1.4	Research Contributions and Scope	5
1.5	Dissertation Structure	6
2	Background	7
2.1	Understanding PVT Variations in Modern CMOS	8
2.1.1	Process Variations and Manufacturing Corners	9
2.1.2	Voltage and Temperature Dynamics	11
2.2	Principles of PVT Sensing and the Bandgap Reference	13
2.2.1	Theoretical Foundations of the Bandgap Reference	13
2.2.2	Circuit Implementation and Operational Constraints	15
2.2.3	Low-Voltage Bandgap References	18
2.2.4	Supply-Independent Biasing and PTAT Current Generation	18
2.2.5	Speed, Noise, and Output Impedance	19
2.3	Ring Oscillators as PVT Transducers	20
2.4	16-nm FinFET Ring Oscillator Sensitivity Characterization	20
2.4.1	Characterization and Sensing Ambiguity	21
3	Related Work	23
3.1	Voltage References and Bandgap Architectures	23
3.1.1	High-Order Curvature Compensation	24
3.1.2	Amplifier-Free and Sub-Threshold Topologies	26
3.1.3	Sampled and Offset-Corrected BGRs	26
3.1.4	MOS-Only Architectures	27
3.1.5	Sub-1-V Current-Mode FinFET BGRs: PMOS and NMOS Sensing Variants	27
3.1.6	Precision Current References	31
3.2	Process Sensors	31
3.2.1	Ring Oscillator-Based Process Monitors	32
3.2.2	Sub-Threshold Leakage-Based Process Sensing	35
3.2.3	Within-Die Variability and Reconfigurable Monitors	36
3.3	Voltage Sensors	37
3.3.1	Reference-Free Voltage Monitoring	37
3.3.2	High-Bandwidth and Event-Driven Voltage Detection	38
3.3.3	Delay-Chain and Frequency-Domain Voltage Sensing	38
3.4	Temperature Sensors	42

3.4.1	Resistor-Based and BJT-Based Thermal Sensors	42
3.4.2	Ring-Oscillator-Based Thermal Sensor in 16-nm FinFET	43
3.4.3	Composite-Transistor PTAT Current Reference for Thermal Sensing . . .	45
3.4.4	N-well Diode Time-Domain Sensors in FinFET Nodes	47
3.4.5	Ultra-Compact CMOS-Only Temperature Sensor in 40-nm	48
3.4.6	Frequency-Locked Loop Architectures	49
3.4.7	Additional VCO-Based, Frequency-Domain, and Industrial-Grade Designs	50
3.5	Comparative Analysis and Design Choices	51
4	Implementation and Simulation Results	58
4.1	Process Sensor	61
4.1.1	Topological Overview and Structural Differences	61
4.1.2	Testbench Configuration and Frequency Constraints	65
4.1.3	Transient Analysis and Frequency Separation	66
4.1.4	Comparative Process Corner Variations at Nominal Conditions	67
4.1.5	Temperature Dependency	72
4.1.6	Supply Voltage Sensitivity	75
4.1.7	Architecture Selection and Design Justification	77
4.1.8	Monte Carlo Validation of the Diode Connected Load Ring Oscillator . .	80
4.2	Voltage Sensor	83
4.2.1	Foundational Baseline: The Bandgap Reference	83
4.2.2	Testbench Infrastructure and Circuit Components	86
4.2.3	Nominal Performance and Chain stabilization	88
4.2.4	Sequential Stage-Delay optimization	88
4.2.5	Environmental Robustness Under Temperature and Process Corner Variation	96
4.2.6	Simulation Results and Tabular Comparison	97
4.3	Temperature Sensor	100
4.3.1	PTAT Source Generation and BGR optimization	100
4.3.2	Testbench Infrastructure and Circuit Integration	102
4.3.3	Process Corner and Temperature Sweep Analysis	103
4.3.4	Supply Voltage Sensitivity and Final Topology Selection	106
4.4	Summary and Comparison with the State of the Art	111
4.4.1	Process Sensor Comparison	111
4.4.2	Voltage Sensor Comparison	112
4.4.3	Temperature Sensor Comparison	113
5	Conclusions	115
5.1	Summary of Contributions	115
5.2	Future Work	116
	References	118

List of Figures

1.1	Monitors and analytics throughout the silicon lifecycle, showing the data flow from design to in-field operation.	1
1.2	State diagram of the FSM controller, illustrating the hierarchical priority of process sensing followed by a cyclic environmental monitoring loop comprising reset, temperature, voltage, and compensation states.	3
2.1	The five classical CMOS process corners representing bounded statistical parameters.	11
2.2		11
2.3	First-order ideal temperature cancellation assuming perfectly linear CTAT and PTAT components.	15
2.4	Realistic BGR thermal behaviour exhibiting a residual parabolic error profile (“bandgap belly”).	15
2.5	Conceptual circuit illustrating the thermal compensation principle of the BGR [1].	16
2.6	Op-amp based implementation of the Bandgap Reference circuit enforcing fixed node potentials [1].	16
2.7	Testbench schematic of the basic enabled 7-stage ring oscillator configuration. . .	21
2.8	Simulated basic RO oscillation frequency versus operating junction temperature.	21
2.9	Simulated basic RO oscillation frequency versus supply voltage rail (V_{DD}).	21
3.1	Sub-1-V FinFET BGR (PMOS variant) by Low et al. [28].	29
3.2	Current-mode BGR employing diode-connected NMOS transistors with PMOS current mirror [29].	30
3.3	Ratioed-inverter RO process sensor by An et al. [14].	33
3.4	Proposed delay-cell design by Chandan et al. [32]: N-Ratioed NAND delay cell (left) and P-Ratioed NOR delay cell (right).	35
3.5	Inhomogeneous RO for WID variability characterization by Fujimoto et al. [33]. .	36
3.6	Proposed voltage sensor (ULV^2TC) by Chen et al. [4], showing the current-starved delay chain, the V_{REF}/V_{in} multiplexer, and the flip-flop tap array producing $V[4:0]$.	40
3.7	Controllable delay-chain voltage sensor from the on-chip CMOS corner detector of Lou et al. [10].	40
3.8	Dual voltage/temperature resistive sensor in 16-nm FinFET by Horng et al. [13]. .	41
3.9	RO-based thermal sensor in 16-nm FinFET by Lempel et al. [44].	45
3.10	Composite-transistor PTAT current reference by Hanumantha Rao et al. [31]. . .	46
3.11	Time-domain temperature sensor in 7-nm FinFET by Eberlein et al. [43].	48
3.12	Ultra-low-power 40-nm temperature sensor by Lu et al. [46].	49
3.13	FLL-based temperature sensor by Someya et al. [45].	50
4.1	Proposed top-level block diagram of the unified PVT monitoring system.	59

4.2	Finite State Machine control diagram, reproduced from Chapter 1, governing the sequential calibration routine of the unified PVT system.	59
4.3	Conceptual NMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.	62
4.4	Conceptual PMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.	62
4.5	DC sweep of the NMOS delay cell output voltage for three transistor sizing ratios, establishing the optimal driver-to-load ratio for the Diode Connected Load Ring Oscillator.	63
4.6	NMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.	64
4.7	PMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.	64
4.8	21-stage NMOS ring oscillator of the Diode Connected Load Ring Oscillator.	64
4.9	21-stage PMOS ring oscillator of the Diode Connected Load Ring Oscillator.	64
4.10	Shared testbench configuration used for evaluating both the Diode Connected Oscillator and the Ratioed Topology.	65
4.11	Transient output of the NMOS-path Ratioed Topology ring oscillator across all five process corners at 25°C and $V_{DD} = 0.8$ V.	66
4.12	Transient output of the PMOS-path Ratioed Topology ring oscillator across all five process corners at 25°C and $V_{DD} = 0.8$ V.	67
4.13	Two-dimensional process corner map of the Diode Connected Load Ring Oscillator at nominal conditions (25°C, $V_{DD} = 0.8$ V), plotting PMOS-path average frequency against NMOS-path average frequency for each of the five corners.	68
4.14	Two-dimensional process corner map of the Ratioed Topology at nominal conditions (25°C, $V_{DD} = 0.8$ V), included as a reference for direct comparison.	68
4.15	NMOS-path Diode Connected Load Ring Ring Oscillator oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8$ V).	73
4.16	PMOS-path Diode Connected Load Ring Oscillator oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8$ V).	73
4.17	NMOS-path Ratioed Topology oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8$ V), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.	74
4.18	PMOS-path Ratioed Topology oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8$ V), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.	74
4.19	NMOS-path Diode Connected Load Ring Oscillator frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$).	75
4.20	PMOS-path Diode Connected Load Ring Oscillator frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$).	76
4.21	NMOS-path Ratioed Topology oscillation frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.	76
4.22	PMOS-path Ratioed Topology oscillation frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.	77
4.23	Initial Monte Carlo distribution of the Diode Connected Load Ring Oscillator (pre-resizing).	80
4.24	Monte Carlo distribution of the Diode Connected Load Ring Oscillator after transistor resizing.	81

4.25	Simulated V_{ref} as a function of V_{DD} for the sized bandgap reference [29] (TT corner, 25°C).	84
4.26	Simulated V_{ref} as a function of temperature for the sized bandgap reference [29] ($V_{DD} = 0.8\text{ V}$, TT corner).	85
4.27	Simulated PTAT, CTAT, and total output current I_2 of the sized Moisel et al. [29] bandgap reference as a function of temperature ($V_{DD} = 0.8\text{ V}$, TT corner).	86
4.28	Voltage sensor testbench schematic.	87
4.29	Schematic of a single current-starved inverter (CSS) stage.	87
4.30	Per-stage propagation delay versus stage index for the 8-stage configuration at nominal conditions ($V_{DD} = 0.8\text{ V}$, TT corner, 25°C).	89
4.31	Voltage sensor output waveforms at nominal conditions ($V_{DD} = 0.8\text{ V}$, TT corner, 25°C).	89
4.32	Transient output waveforms for the 2-stage interval configuration. CLK and outputs D_1 – D_7 are shown for $V_{DD} \in \{650, 725, 800, 875, 950\}\text{ mV}$	91
4.33	Transient output waveforms for the 4-stage interval configuration. CLK and outputs D_1 – D_7 are shown for $V_{DD} \in \{650, 725, 800, 875, 950\}\text{ mV}$	92
4.34	Transient output waveforms for the 8-stage interval configuration. CLK and outputs D_1 – D_7 are shown for $V_{DD} \in \{650, 725, 800, 875, 950\}\text{ mV}$	94
4.35	Per-stage propagation delay versus stage transition across temperature ($V_{DD} = 0.8\text{ V}$, TT corner).	97
4.36	Per-stage propagation delay versus stage transition across process corners ($V_{DD} = 0.8\text{ V}$, 25°C).	98
4.37	Modified Bandgap Reference schematic used as the upgraded PTAT source.	101
4.38	Testbench for the native sensor configuration using its own PTAT core.	102
4.39	Testbench for the sub-threshold PTAT current reference integrated with the ring-oscillator sensor topology.	103
4.40	Testbench for the BGR-based PTAT source integrated with the ring-oscillator sensor topology.	103
4.41	Average output frequency versus temperature for the native sensor configuration across all five process corners.	104
4.42	Average output frequency versus temperature for the sub-threshold PTAT current reference integrated with the ring-oscillator sensor topology, across all five process corners.	104
4.43	Average output frequency versus temperature for the BGR-based PTAT source integrated with the ring-oscillator sensor topology, across all five process corners.	105
4.44	Average output frequency versus supply voltage V_{DD} at $T = 25^{\circ}\text{C}$ (TT corner) for all three PTAT configurations.	107

List of Tables

3.1	Comparison of Process Characterization Architectures (16-nm FinFET Relevant Subset).	51
3.2	Comparison of Voltage Reference and Bandgap Architectures (Anchored on the FinFET-Native Current-Mode Topologies).	52
3.3	Comparison of Voltage Sensor Architectures (16-nm FinFET Relevant Subset). .	53
3.4	Comparison of Temperature Sensor Architectures (16-nm FinFET Relevant Subset).	54
4.1	Comparison of delay cell behaviour across the three transistor sizing ratios. . . .	63
4.2	Simulated parameter variations across process corner pairs for the PMOS-path and NMOS-path oscillators of each topology.	70
4.3	Comparison of NMOS and PMOS process sensitivity coefficients between the Diode Connected Load Ring Oscillator and the Ratioed Topology, computed as $k = \frac{\Delta T/T_0}{\Delta V/V_0}$, where ΔT is the peak-to-peak period difference between the fastest and slowest corner on each path, T_0 is the period at the TT corner, and $\Delta V/V_0$ is the corresponding relative threshold-voltage deviation.	71
4.4	Monte Carlo mean and standard deviation of NMOS-path and PMOS-path oscillation frequency for the Diode Connected Load Ring Oscillator after transistor resizing, for each of the five process corners (200 runs per corner).	81
4.5	Clock delay variance Δt_{CLK} against the required encoding margin $ t_{D6}^{0.95V} - t_{D5}^{0.8V} $ for each evaluated stage-interval configuration.	95
4.6	Sampled logic states at monitored D-stage outputs after delay block optimization (8-stage configuration, TT corner, 25°C).	98
4.7	Voltage sensor encoded digital output after XOR network (8-stage configuration, TT corner, 25°C).	99
4.8	Comparison of supply-voltage sensitivity across the three PTAT testbenches, $V_{DD} = 0.65\text{--}0.95\text{ V}$, $T = 25^\circ\text{C}$, TT corner.	109
4.9	Comparison of process sensor parameters between this work and related ring-oscillator-based process sensing topologies.	111
4.10	Comparison of voltage sensor parameters between this work and related current starved delay chain voltage sensing topologies.	112
4.11	Comparison of temperature sensor parameters between this work and related time domain temperature sensing topologies. The supply sensitivity entries for Someya et al. [45] and Lu et al. [46] are not yet available and are marked accordingly. . .	113

Chapter 1

Introduction

With the relentless advance of semiconductor technology and the ubiquitous proliferation of the Internet of Things (IoT), the fundamental requirements for integrated circuits (ICs) have undergone a paradigm shift. In modern daily life, a single individual often interacts with a constellation of connected devices, from smartphones and smartwatches to wearable health monitors, all requiring high computational throughput within a constrained power envelope. This “low-power era” does not imply a compromise on functionality; rather, the industry demands higher performance-per-watt metrics, necessitating aggressive voltage scaling and tighter control over silicon behavior. To achieve these targets, modern ICs are no longer treated as static components but as evolving entities within a Silicon Lifecycle. This lifecycle encompasses everything from initial design and manufacturing to in-field operation and eventual decommissioning. Historically, ensuring reliability across this span relied on exhaustive external testing: wafer probe tests for manufacturing quality, Automated Test Equipment (ATE) for functional verification, and guard-banding to account for aging. However, as device volume scales and technology nodes shrink, these external phases have become a critical economic and technical bottleneck.

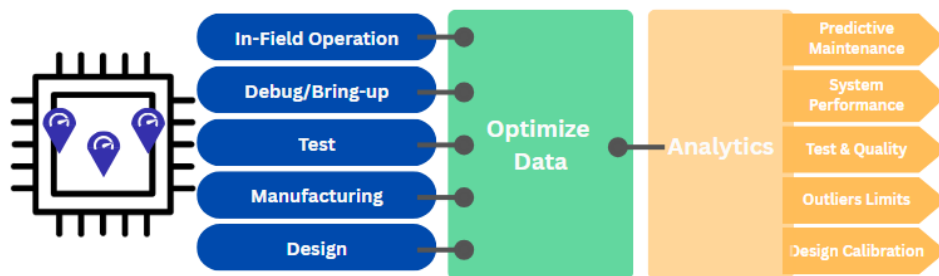


Figure 1.1: Monitors and analytics throughout the silicon lifecycle, showing the data flow from design to in-field operation.

As illustrated in Figure 1.1, data must flow bi-directionally between the physical chip and analytics platforms to optimize performance at every stage. Testing and calibration consume vast amounts of time and capital, limiting production throughput. Consequently, there is a burgeoning industrial need for Silicon Lifecycle Management (SLM), shifting the burden of characterization

from expensive external equipment to intelligent, on-chip self-characterization systems. By enabling an IC to autonomously detect its own environmental state and calibration needs, manufacturers can minimize dependency on expensive ATE cycles, yielding substantial economic benefits.

Consequently, this dissertation addresses this industrial need by proposing a comprehensive On-Chip PVT (Process, Voltage, Temperature) Sensing System. The primary objective is to develop the on-chip infrastructure necessary to detect and compensate for environmental and manufacturing variabilities in sub-micron technology nodes, moving the burden of metrology from external equipment to internal circuits.

1.1 Scientific Context

The research presented in this dissertation is grounded in the foundational principles of analog integrated circuit design [1], addressing the modern challenges of energy-quality scaling in the “twilight of Moore’s Law” [2], and utilizing advanced sensors for dynamic voltage scaling [3].

As technology nodes scale from FinFET to sub-nanometer regimes, the physical landscape of the silicon substrate changes drastically. While FinFET technology offers superior electrostatic control over the channel, it introduces significant secondary effects such as increased self-heating due to the three-dimensional fin structure and higher interconnect resistance. These advanced nodes are characterized by lower supply voltages (V_{DD}), and extreme power densities, leading to localized “thermal hotspots” and significant induced voltage drops. In this context, the traditional approach of worst-case guard-banding to cover process corners, from Fast-Fast to Slow-Slow, is no longer viable, as it consumes the very power-efficiency gains that FinFETs aim to provide.

Manufacturing variations in these regimes become statistically significant, rendering traditional yield learning vehicles, such as SRAM arrays, too complex to isolate specific logic failures effectively. While the industry has shifted toward Ring Oscillators (ROs) as primary proxies for process health [4], mere frequency monitoring is insufficient for advanced diagnosis. Modern architectures must employ failure isolation techniques to distinguish between catastrophic defects, such as interconnect shorts, and parametric shifts, such as threshold voltage variations [5]. Furthermore, sub-threshold leakage sensors have emerged as critical tools for monitoring static power variations in low-power applications, necessitating sensor architectures that can operate effectively in near-threshold regimes [3].

However, the efficacy of even the most sophisticated process sensor is rendered null if the voltage reference biasing is unstable. The reliability of any on-chip diagnostic system is anchored by its reference foundation. As established in classical theory by Razavi [1], the precision of mixed-signal systems is fundamentally limited by the stability of their reference generators. This creates a “recursive paradox” in self-characterization: the tool used to measure the chip is often susceptible to the same environmental noise as the chip itself.

Standard Bandgap Voltage References (BGR) are widely used to create temperature-independent

voltages, yet these first-order compensated circuits are typically limited by higher-order nonlinearities, restricting their accuracy across wide temperature ranges [6]. To support precise PVT monitoring in IoT devices, advanced curvature-compensated architectures are required. Recent works have demonstrated ultra-low drift performance using high-order compensation [7] and MOS-only topologies that minimize area [8]. Without such precision, temperature variations can be misinterpreted as voltage or process shifts, corrupting the self-characterization data.

1.2 System-Level Sequencing and FSM Architecture

To resolve these ambiguities and ensure accurate characterization, a system architecture that utilizes a centralized Finite State Machine (FSM) to manage a hierarchical sensing flow can be employed [3]. This operational sequence can be designed to decouple manufacturing variability from dynamic environmental fluctuations, an approach that has been refined in this work into a strictly sequential calibration scheme governing the interaction between the three sensing subsystems.

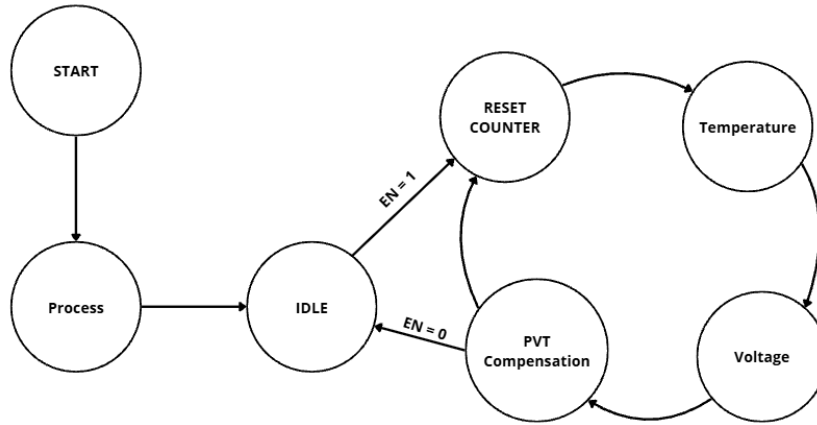


Figure 1.2: State diagram of the FSM controller, illustrating the hierarchical priority of process sensing followed by a cyclic environmental monitoring loop comprising reset, temperature, voltage, and compensation states.

As depicted in Figure 1.2, the controller transitions from an initial **START** state directly into the **Process** state upon power-up, before settling into an **IDLE** state that awaits an external enable signal. Once the enable condition is asserted ($EN = 1$), the FSM proceeds into a cyclic loop encompassing a **RESET COUNTER** state, followed sequentially by **Temperature** sensing, **Voltage** sensing, and finally **PVT Compensation**. The loop continues to iterate, returning to **RESET COUNTER**, until the enable signal is deasserted ($EN = 0$), at which point control returns to the **IDLE** state. This architecture enforces a strict operational hierarchy with three distinct stages of priority.

The first stage in this hierarchy is process monitoring. Upon system reset, the FSM prioritizes a single, upfront execution of the process sensing block before entering the cyclic environmental

loop. This stage reads the process sensor, realized as a Diode Connected Load Ring Oscillator (an all NMOS/PMOS ring oscillator), ahead of any voltage or temperature measurement. Because the silicon corner profile of a given die is fixed at fabrication and does not vary during operation, characterizing it once at the beginning of the sequence is sufficient to establish a static process baseline. This design choice is intended to reduce process dependency in subsequent temperature and voltage readings, so that the system does not need to fully disentangle process-induced frequency shifts from those caused by dynamic environmental fluctuations during every sensing cycle.

The second stage concerns the temperature sensing subsystem. The temperature sensor employed in this work is engineered around a Proportional-To-Absolute-Temperature (PTAT) core selected specifically for its reduced sensitivity to supply voltage fluctuations relative to its front end's native bias source. By relying on a PTAT component whose output tracks temperature with substantially attenuated dependence on V_{DD} , the temperature reading is made significantly more robust to the voltage domain, rather than fully independent of it. This is a deliberate architectural decision, as it limits the extent to which supply noise or IR-drop transients can be misinterpreted as thermal events, reducing, though not eliminating, the cross-compensation burden this would otherwise place on the calibration logic.

The third and final stage is the voltage and temperature compensation engine itself. Within this hierarchy, the role of the voltage sensor is to read the supply voltage, V_{DD} , once corrected for the process and temperature dependent offsets present in its own delay chain, providing the compensation engine with a direct measurement of the supply rail rather than acting as a generic environmental proxy. Because process variation has largely been characterized in the first stage, and the temperature block exhibits substantially reduced sensitivity to V_{DD} fluctuations by virtue of its PTAT architecture, the multi-sensor calibration engine that would otherwise be required to jointly resolve process, voltage, and temperature interactions can be considerably simplified. The remaining compensation task, executed in the `PVT Compensation` state, is reduced primarily to managing the interaction between voltage and temperature, with the process baseline acting as a static correction term rather than a third actively coupled variable. This reduction in dimensionality is the central contribution of the proposed sequencing strategy, as it allows the on-chip compensation logic to be implemented with substantially lower hardware complexity than a fully coupled three-variable calibration scheme would demand.

By logging the process characteristics first and structurally reducing the temperature path's exposure to supply variation, the system establishes a static reference that limits manufacturing dependency in subsequent dynamic measurements, while simultaneously reducing the dimensionality of the real-time compensation problem. This hierarchy allows the system to interpret real-time fluctuations against the established process baseline, culminating in the `PVT Compensation` state, where the system calculates and applies the necessary adjustments to operating parameters, enabling autonomous, closed-loop compensation for hotspots and droops without the need for external metrology.

The research methodology follows a standard industrial IC design flow utilizing Synopsys

proprietary tools, mainly focused on transistor-level schematic design and rigorous simulation by implementing a comprehensive analysis.

1.3 Main Objectives of This Work

The primary engineering goal of this dissertation is to propose a low-complexity PVT monitoring framework optimized for a 16 nm FinFET technology node operating under a 0.8 V nominal supply. In support of this overarching objective, the work pursues the following specific goals.

1. To design and characterize a process sensing block based on a Diode Connected Load Ring Oscillator topology, capable of establishing an accurate silicon corner baseline at start-up.
2. To develop a temperature sensing subsystem whose PTAT core architecture exhibits substantially reduced sensitivity to supply voltage fluctuations relative to its native bias source, minimizing cross-sensitivity between the temperature and voltage measurement paths.
3. To design a voltage sensing subsystem capable of reading the supply voltage (V_{DD}) directly.
4. To define and implement a sequential FSM control strategy that exploits the process-first, voltage-temperature sensing hierarchy in order to minimize the complexity of the final compensation engine.
5. To evaluate the resulting sensor blocks and control architecture through comprehensive transistor-level simulation across representative voltage and temperature skews.

Taken together, these objectives are intended to demonstrate that a carefully sequenced sensing architecture can achieve robust PVT awareness without resorting to the hardware overhead typically associated with fully coupled, simultaneous three-variable calibration engines.

1.4 Research Contributions and Scope

This thesis delivers several concrete contributions toward the goal of a self-calibrating, low-complexity PVT sensing system. At the circuit level, individual process, voltage, and temperature sensor blocks were designed and thoroughly verified at the transistor level in a native 16 nm FinFET process. Each sensor block was subjected to exhaustive standalone simulation, including corner analysis, supply and temperature sweeps, and Monte Carlo variation studies where applicable, in order to characterize its sensitivity and accuracy in isolation.

At the system level, a sequential FSM control scheme was proposed in in Figure 1.2, formalizing the process-first, voltage-temperature sensing hierarchy described Section 1.1 was not fully developed. However, a theoretical top-level integration scheme and a unified testbench framework was proposed (in Chapter 4), defining how the individual sensor blocks, the FSM controller, and the compensation engine are intended to interconnect within a single PVT monitoring architecture.

It is important to define transparently the practical scope of the simulation phase carried out in this work. Owing to the timeline constraints of the simulation environment, the exhaustive verification effort was concentrated on each individual sensor block operating independently, rather than on the full closed-loop simulation of the unified, multi-block architecture. The theoretical top-level integration scheme and the accompanying unified testbench framework should therefore be understood not as a fully simulated and silicon-validated subsystem, but as a validated architectural blueprint, grounded in the individually verified sensor behavior, that is intended to guide future integration efforts and eventual tape-out. This distinction is made explicit so that the reader correctly interprets the maturity level of the unified architecture relative to that of the individually characterized sensor blocks.

1.5 Dissertation Structure

In addition to the introduction, this document contains four more chapters.

The remainder of this document is organized as follows: Chapter 2 establishes the theoretical foundations of the study, detailing the physics of PVT variations and the fundamental operation of standard Ring Oscillators and Bandgap Voltage References. Chapter 3 presents a comprehensive review of the state-of-the-art, critically comparing recent advancements in on-chip monitoring circuits, logic-decomposed process sensing, and curvature-compensation topologies for voltage references. Chapter 4 presents the proposed master-satellite infrastructure, detailing its architectural layout, finite state machine control sequencing, and comprehensive simulation results characterized across independent voltage and temperature skews in a native 16-nm FinFET node. Finally, Chapter 5 summarizes the central insights of this dissertation, evaluates the performance metrics of the selected sensor blocks, and outlines the primary contributions of this work towards a robust, self-calibrating PVT sensing system.

Chapter 2

Background

The design of modern integrated circuits (ICs) is fundamentally constrained by the explosive magnification of Process, Voltage, and Temperature (PVT) variations alongside the aggressive scaling of feature sizes into the deep sub-micron and nanoscale regimes. As the semiconductor industry approaches the twilight of Moore’s Law, energy-quality scaling has forced operating points closer to the absolute boundaries of performance and energy efficiency [1], [2], [3]. The conventional approach of utilizing static, worst case design margins (commonly referred to as guard-banding) has become increasingly impractical in deep submicron regimes[2]. Historically, circuits were over-engineered to ensure robust functionality at absolute, worst case PVT corners, such as the SS process skew at minimum supply voltage and peak operating temperature [1], [3].

However, enforcing deterministic timing closure under these rare, multi-sigma conditions forces designers to aggressively upscale transistor dimensions, insert dense clock buffers, and inject redundant safety margins into critical paths [2], [9]. When the fabricated silicon operates under Typical-Typical (TT) nominal conditions, this structural over-engineering yields a massive, permanent penalty [2]. The artificially enlarged gate areas heavily inflate the parasitic node capacitances (C_L), driving up active dynamic power consumption, while simultaneously expanding the sub-threshold leakage paths that drain always-on power grids [1], [2].

Despite these clear inefficiencies, the underlying concept of over-provisioning remains deeply embedded and actively integrated within today’s mainstream silicon technologies. Because traditional safety margins provide a straightforward, localized guarantee against infant mortality and dynamic supply noise, contemporary industrial architectures still tolerate these area and power overheads out of sheer necessity. However, as operating margins continue to compress under modern sub-1-V headrooms, this legacy design framework has reached its scaling limit. The persistence of these heavy margins in current commercial processors directly highlights the critical industrial demand for a transition toward more intelligent systems.

Despite CMOS scaling facilitating notable improvements in integration density and raw performance, circuits have simultaneously become acutely sensitive to environmental variation. At nanoscale regimes, even atomic-level structural flaws can dramatically alter transistor characteristics, compromising consistent and predictable performance. Consequently, mitigating and com-

compensating for PVT variations has emerged as a primary challenge in semiconductor engineering, necessitating a meticulous balance between speed, power, and reliability [1]. A classic example of this critical tension can be seen in a modern multi-core application processor executing a high-throughput computational workload.

- **The Speed vs. Reliability Dilemma (Slow/Hot Corner):** If a localized hotspot develops on the die, pushing the temperature toward the high-temperature worst-case corner, carrier mobility degrades within the transistor channels [1]. This slowing down of the silicon gate propagation delay directly threatens timing paths; if a critical setup-time margin is violated, it results in data corruption and a catastrophic system crash, completely undermining operational *reliability* [3]. To maintain the targeted processing *speed*, the Power Management Unit (PMU) must respond by scaling up the supply voltage (V_{DD}) [10].
- **The Power vs. Reliability Dilemma (Fast/Cold Corner):** Conversely, if the chip operates at a cold temperature extreme at a Fast-Fast (FF) manufacturing process corner, the threshold voltage (V_{th}) drops significantly [1], [11]. While this scenario provides exceptional processing *speed*, the sub-threshold leakage current spikes exponentially [11]. This surge in leakage triggers a massive *power* penalty, rapidly draining battery-operated devices and threatening long-term reliability via accelerated electromigration along the power delivery lines [2].
- **The Dynamic Voltage Noise Bottleneck (Supply Droop):** Compounding these issues, if adjacent digital blocks switch simultaneously, the localized IR-drop can cause a transient supply voltage droop of 50–100 mV [12], [13]. To ensure the chip doesn't fail during this brief drop, engineers are forced to permanently lower the clock frequency (*speed*) or permanently over-volt the entire chip (*power*) to create a safe buffer [2].

This intricate, circular interdependency demonstrates that maximizing performance in any single dimension invariably degrades the other two, proving that static margins are no longer sufficient to manage nanometric silicon behavior.

2.1 Understanding PVT Variations in Modern CMOS

To systematically counteract these environmental and manufacturing deviations, modern Systems-on-Chip (SoCs) rely on dedicated on-chip process, voltage, and temperature sensors. Each sensor classification serves a distinct role within the holistic power and telemetry management infrastructure:

- **Process Sensors** evaluate local ring oscillators or replica circuits against a baseline clock to categorize a specific die or region as fast, typical, or slow relative to target specifications.
- **Voltage Sensors** monitor localized power delivery networks against an absolute reference to capture dynamic voltage drops (*IR* drop) and ground bounce.

- **Temperature Sensors** track localized thermal dissipation across the silicon die, providing the real-time feedback required for dynamic thermal throttling (DTT), performance regulation, and reliability enhancement.

2.1.1 Process Variations and Manufacturing Corners

Process variations are permanently embedded during fabrication and are broadly bifurcated into Within-Die (WID) variations and Die-to-Die (D2D) variations. D2D variations manifest as macroscopic parameter shifts affecting all transistors on a chip uniformly—such as global fluctuations in oxide thickness or channel doping concentrations across wafers and batches. Conversely, WID variations represent highly localized variations stemming from atomic-scale stochasticity, including random dopant fluctuation (RDF) and line-edge roughness (LER) between adjacent devices.

Because a transistor's threshold voltage (V_{th}) directly dictates its switching speed and sub-threshold leakage, these microscopic mismatches lead to wide variations in performance and power consumption. Historically, designers compensated for this uncertainty by applying conservative, static guard-bands to guarantee functionality under worst-case conditions. However, this strategy severely penalizes typical or superior silicon by burning excess energy under nominal operating conditions [1].

To illustrate this systemic inefficiency, consider the operational profile of a high-speed arithmetic processing block implemented in a 16-nm FinFET node. To guarantee functional reliability over a ten-year lifespan, engineers are forced to hardcode the global supply voltage and clock frequency to satisfy an absolute worst-case scenario: a combination of slow manufacturing variations, a severe peak temperature, and a transient voltage drop on the power delivery network [1], [2]. Under this rare multi-sigma condition, the degraded carrier mobility and compressed gate overdrive slow down the switching speed of the transistors, forcing the system to throttle its clock frequency to a conservative limit to avoid timing violations and data corruption [1], [9]. In reality, however, the vast majority of fabricated silicon dies operate closer to nominal parameters and rarely experience peak thermal stress and voltage drops at the exact same moment [2], [9]. Because the guard-bands are fixed at design time, these superior nominal chips are forced to run at a high, bloated supply voltage, even though their faster transistor characteristics mean they could easily achieve the exact same operational throughput at a significantly lower voltage rail [2], [3]. Because power consumption in digital circuits scales aggressively with the supply voltage, keeping the voltage locked at this worst-case baseline instead of lowering it dynamically forces the typical silicon to burn massive amounts of excess energy, which is dissipated entirely as wasted heat and directly limits the energy efficiency of the system.

To categorize these fabrication deviations, it is first necessary to further explore their physical origin. Even when two identical dies are fabricated from the same schematic and layout, they can exhibit noticeable differences in performance due to process variation. This variability manifests at multiple hierarchical levels.

At global scale, variations are observed across lot-to-lot, and die-to-die comparisons, reflecting systematic shifts in fabrication conditions. Moreover, even within a single wafer, electrical

characteristics can change from point to point, introducing spacial gradients in device behavior. These global variations primarily affect parameters such as oxide thickness, carrier mobility, and threshold voltage, thereby shifting the overall speed and power characteristics of a circuit. For instance, a slow process condition may correspond to thicker oxide, higher threshold voltage, and reduced mobility, ultimately increasing propagation delay in CMOS logic paths.

At the local scale, variations arise between nominally identical devices within the same die, leading to mismatches even for components placed in close proximity. These effects are often referred to as matching errors and originate from microscopic non-uniformities in material properties and geometric dimensions. As a result, quantities that are ideally identical (e.g., transistor currents or resistances) exhibit small statistical deviations, typically modeled using probabilistic distributions.

The semiconductor industry captures the impact of these variations, particularly the global shifts that define the overall performance envelope, through the concept of *Process Corners*, which represent representative extreme operating conditions under which a circuit must retain full functionality. Because digital and mixed-signal CMOS technologies rely on the matched interaction between N-channel (NMOS) and P-channel (PMOS) transistors, process corners are described using a two-letter nomenclature, where the first letter denotes NMOS speed characteristics and the second letter denotes PMOS speed characteristics:

- **Typical-Typical (TT):** Represents nominal processing conditions where both device types exhibit physical parameters (carrier mobility, oxide thickness, V_{th}) closest to the foundry's targeted design center.
- **Fast-Fast (FF):** Characterized by suppressed threshold voltages and elevated carrier mobilities (μ), causing both NMOS and PMOS transistors to switch significantly faster than nominal. While maximizing operational frequency, the FF corner severely penalizes static power metrics through exponentially magnified sub-threshold and gate leakage currents (I_{ddq}), threatening thermal stability.
- **Slow-Slow (SS):** Defines the lower performance boundary where devices exhibit elevated threshold voltages and degraded mobilities. Transistors require longer intervals to charge and discharge nodal capacitances, resulting in maximum propagation delay (t_p) across critical logic paths. The SS corner establishes the absolute boundary for timing closure and setup-time validation.
- **Fast-Slow (FS):** An asymmetrical or skewed corner pairing fast NMOS with slow PMOS devices. This imbalance alters logic switching thresholds, distorting noise margins and duty cycles within symmetric clock distribution networks.
- **Slow-Fast (SF):** The inverse of the FS condition, pairing slow NMOS transistors with fast PMOS devices. It introduces severe rise/fall time mismatches in digital paths and disrupts the precise current-matching requirements of analog structures, such as current mirrors or differential pairs.

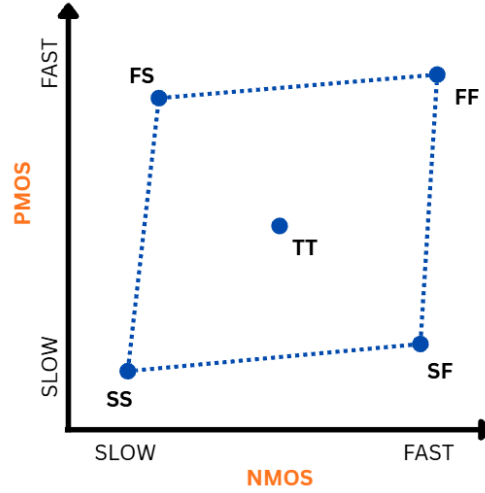


Figure 2.1: The five classical CMOS process corners representing bounded statistical parameters.

Figure 2.2

This graphical representation serves as a generic multi-dimensional map applicable across various physical and electrical device characteristics; the axes can map directly to raw transistor metrics like the threshold voltage (V_{th}), saturation current (I_{sat}), or sub-threshold leakage, as well as macroscopic performance targets like the localized maximum operating frequency. In nanoscale nodes like 16-nm FinFET, multi-corner safety margins have become profligately inefficient [2]. This reality underscores the absolute necessity of on-chip process sensors: by measuring the frequency or current ratios of localized structures, the control infrastructure can dynamically identify the exact operational corner of the fabricated silicon, allowing the system to de-embed manufacturing offsets and scale energy parameters accurately [3], [4], [14].

2.1.2 Voltage and Temperature Dynamics

Voltage variations arise from non-idealities within the power distribution network (PDN). When high-performance processing cores transition abruptly between idle and peak activity states, the rapid transient current demand (dI/dt) creates localized voltage droops across the finite resistance and inductance of the power delivery path. The resistive component of this droop is the classical IR-drop, defined as:

$$\Delta V_{IR} = I_{surge} \cdot R_{PDN} \quad (2.1)$$

where R_{PDN} is the effective resistance of the metal interconnect and package parasitics between the voltage regulator and the switching logic. In advanced nanometric nodes, where supply voltages have been scaled aggressively toward 0.7–0.9 V to limit dynamic power consumption, even a modest ΔV_{IR} of 50–100 mV represents a significant fraction of the available voltage headroom [10]. The inductive component of the PDN introduces a further transient perturbation: the voltage spike $V_L = L_{PDN} \cdot dI/dt$ that develops when switching activity decreases rapidly can temporarily

overshoot the nominal supply, stressing gate oxides and exacerbating reliability concerns at scaled technology nodes [1].

Because the propagation delay of digital logic gates depends non-linearly on the supply voltage through the Alpha-Power law [15]:

$$t_p \propto \frac{C_L \cdot V_{DD}}{(V_{DD} - V_{th})^\alpha} \quad (2.2)$$

even transient drops can corrupt timing paths, precipitating setup-time violations and system-level instability [1]. This non-linear relationship means that the delay sensitivity to supply droops is highest precisely when V_{DD} is closest to V_{th} —the near-threshold operating regime that modern energy-quality scaling designs intentionally target to minimize power [2]. Consequently, the same voltage scaling strategy that reduces dynamic power also maximizes the system’s vulnerability to PDN-induced timing failures, creating a fundamental tension at the heart of modern low-power design.

Compounding the purely resistive and inductive PDN effects is the phenomenon of ground bounce: simultaneous switching of large numbers of output drivers induces a transient voltage rise on the on-chip ground rail through the parasitic inductance of the ground return path. Ground bounce effectively reduces the gate-to-source overdrive of NMOS pull-down transistors, increasing their propagation delay in a manner that is indistinguishable from a supply droop from the perspective of downstream logic. In dense multi-core SoCs, where thousands of gates may switch simultaneously within a single clock cycle, ground bounce can add several tens of millivolts to the effective IR-drop experienced by critical timing paths [10].

Accurate supply voltage monitoring must therefore track not only the DC level of V_{DD} but also the transient envelope of these combined droops, overshoots, and ground bounce events across the bandwidth of the PDN. A sensor that captures only the average supply level will miss the worst-case transient conditions that actually threaten timing integrity.

This challenge is further compounded by localized thermal dynamics. Temperature gradients across the die are generated by both ambient conditions and the localized power dissipation of active switching paths. Transistors respond to thermal changes via two counteracting physical mechanisms: as temperature rises, carrier mobility decreases (degrading drive current), while the simultaneous reduction in threshold voltage increases sub-threshold leakage current [16]. Under standard operating conditions in strong inversion, the mobility degradation mechanism dominates, causing digital logic gates to slow down at elevated temperatures. The resulting delay increase is analogous in its effect on timing paths to a supply voltage droop, and the two phenomena can interact destructively: a thermal hotspot that slows down a critical path reduces the available timing margin, making that path more susceptible to a concurrent IR-drop event than it would be under nominal thermal conditions.

This behavior induces a problematic electro-thermal feedback loop: high computational workloads elevate localized temperatures, slowing down logic paths, which forces adaptive systems to

scale supply voltages higher to maintain clock frequencies, thereby accelerating thermal dissipation and static power leakage [1]. In nanometric FinFET nodes this loop is further tightened by the thermal isolation of the silicon fin from the substrate through shallow trench isolation oxide, which traps heat within the transistor channel and creates thermal hotspots at power densities that flat planar topologies never encountered [17]. The simultaneous presence of supply droops and thermal hotspots in the same spatial region of a high-performance processor therefore represents the most challenging operating condition for any on-chip monitoring infrastructure.

2.2 Principles of PVT Sensing and the Bandgap Reference

On-chip PVT sensors act as the primary telemetry apparatus of modern ICs, capturing the complex interactions between process parameters, supply fluctuations, and thermal gradients to convert them into standard digital formats.

Many integrated sensors monitor active structures whose behavioural variations correlate predictably with environmental parameters, such as the Ring Oscillator (RO). Comprising an odd number of inverting stages connected in a closed loop, an RO generates a periodic signal whose frequency is dictated by the collective propagation delays of its constituent inverters. Because this delay is highly dependent on process corners, current supply levels, and localised temperature, the oscillation frequency serves as a direct indicator of real-time operating conditions. By digitising this frequency against a stable reference clock, the system generates a localised telemetry metric that aligns tightly with actual logic paths [1].

However, raw frequency sensing is inherently ambiguous without a highly stable reference baseline. This requires a reference infrastructure decoupled from ambient fluctuations to ensure complete data integrity across downstream sensors [3]. Precision voltage references serve as the absolute baseline for the entire telemetry infrastructure, because any drift in the reference propagates directly into the error budget of every downstream data converter and sensor. The system's operational ceiling is therefore fundamentally bound by the thermal stability, line sensitivity, and Power Supply Rejection Ratio (PSRR) of the primary reference generator [1], [7], [18].

2.2.1 Theoretical Foundations of the Bandgap Reference

The fundamental principle of Bandgap Reference (BGR) design involves the weighted summation of a Proportional-to-Absolute-Temperature (PTAT) voltage and a Complementary-to-Absolute-Temperature (CTAT) voltage [1]. The objective is to achieve a theoretical Temperature Coefficient (TC) of zero at a specific reference temperature T_0 . The resulting reference voltage is typically expressed as:

$$V_{\text{REF}}(T) = V_{\text{BE}}(T) + K \cdot V_T(T) \quad (2.3)$$

where V_{BE} represents the forward-biased base-emitter voltage of a bipolar junction transistor (BJT) acting as the CTAT source, $V_T = kT/q$ represents the thermal voltage acting as the core PTAT component, and K is a scalable amplification constant.

Negative-TC Voltage (CTAT Source)

The base-emitter voltage V_{BE} of a bipolar transistor exhibits a negative temperature coefficient. Starting from $I_C = I_S \exp(V_{BE}/V_T)$, where $V_T = kT/q$, and using the temperature dependence of the saturation current [1]:

$$I_S = b T^{4+m} \exp\left[\frac{-E_g}{kT}\right] \quad (2.4)$$

where b is a proportionality constant, $E_g \approx 1.12$ eV is the bandgap energy of silicon, and $m \approx -3/2$ is the temperature exponent of minority-carrier mobility ($\mu \propto \mu_0 T^m$). Taking $\partial V_{BE}/\partial T$ with I_C held constant and substituting from (2.4) gives the intermediate result [1]:

$$\frac{V_T}{I_S} \frac{\partial I_S}{\partial T} = (4+m) \frac{V_T}{T} + \frac{E_g}{qT} \quad (2.5)$$

leading to the fundamental expression for the TC of V_{BE} [1]:

$$\frac{\partial V_{BE}}{\partial T} = \frac{V_{BE} - (4+m)V_T - E_g/q}{T} \quad (2.6)$$

With $V_{BE} \approx 750\text{--}800$ mV at $T = 300$ K, this evaluates to approximately -1.5 mV/K. Importantly, (2.6) shows that the TC of V_{BE} is not constant but itself depends on temperature, which is the physical origin of the residual curvature in first-order bandgap circuits [1].

Positive-TC Voltage (PTAT Source)

It was recognised by Hilbiber in 1964 that if two identical bipolar transistors Q_1 and Q_2 ($I_{S1} = I_{S2}$) are biased at collector currents of nI_0 and I_0 respectively, the difference in their base-emitter voltages is directly proportional to absolute temperature [1]:

$$\Delta V_{BE} = V_{BE1} - V_{BE2} = V_T \ln n \quad (2.7)$$

with TC:

$$\frac{\partial(\Delta V_{BE})}{\partial T} = \frac{k}{q} \ln n \quad (2.8)$$

Crucially, this TC is independent of temperature or the behaviour of the collector currents, making it an ideal positive-TC building block [1]. If Q_2 is realised as m unit cells each identical to Q_1 , the effective current-density ratio becomes nm , and $\Delta V_{BE} = V_T \ln(nm)$.

Combining CTAT and PTAT: The Bandgap Reference

Setting $\alpha_1 = 1$ and noting that $\partial V_{BE}/\partial T \approx -1.5$ mV/K and $\partial V_T/\partial T = k/q \approx +0.087$ mV/K at room temperature, zero TC requires [1]:

$$\alpha_2 \ln n \approx 17.2 \quad (2.9)$$

yielding the classical bandgap reference voltage:

$$V_{\text{REF}} \approx V_{BE} + 17.2 V_T \approx 1.25 \text{ V} \quad (2.10)$$

This is the origin of the term *bandgap reference*: setting $\partial V_{\text{REF}}/\partial T = 0$ and substituting from (2.6) yields [1]:

$$V_{\text{REF}} = \frac{E_g}{q} + (4 + m)V_T \quad (2.11)$$

As $T \rightarrow 0$, $V_{BE} \rightarrow E_g/q$, so $V_{\text{REF}} \rightarrow E_g/q \approx 1.12 \text{ V}$ — the bandgap voltage of silicon, hence the name [1].

While V_T scales linearly with temperature, V_{BE} exhibits an implicit, non-linear temperature dependence — termed “curvature” — which manifests as a characteristic $\ln(T)$ term in its temperature expansion. In a standard, first-order BGR architecture, only the linear temperature coefficients are cancelled out. The remaining uncompensated logarithmic term leaves a residual parabolic error profile — conceptually known as the “bandgap belly” — which fundamentally limits standard reference stability to approximately 20–50 ppm/°C [19].

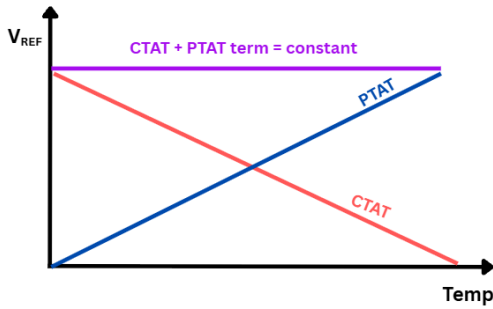


Figure 2.3: First-order ideal temperature cancellation assuming perfectly linear CTAT and PTAT components.

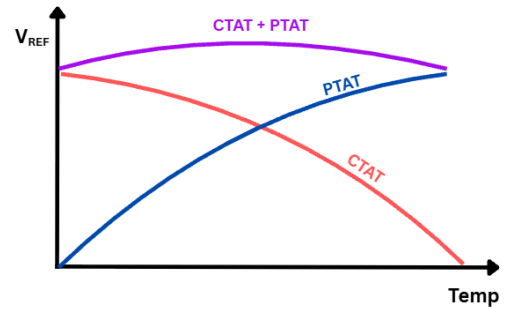


Figure 2.4: Realistic BGR thermal behaviour exhibiting a residual parabolic error profile (“bandgap belly”).

The individual components and the resulting non-linear deviation are illustrated in Figure 2.3 and Figure 2.4. To achieve parts-per-million precision across extended thermal envelopes, state-of-the-art designs must incorporate high-order curvature correction mechanisms to counteract this physical limitation [7], [20].

2.2.2 Circuit Implementation and Operational Constraints

The baseline circuit topology historically used to implement this compensation principle is shown in Figure 2.5.

The conceptual core in Figure 2.5 highlights the necessary combination of two discrete PN junction profiles [1]. The primary node, V_{BE1} , serves as the CTAT source to introduce the negative thermal slope. The complementary differential voltage, ΔV_{BE} , is extracted between two bipolar

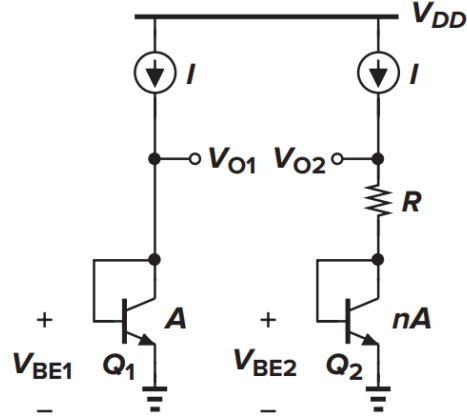


Figure 2.5: Conceptual circuit illustrating the thermal compensation principle of the BGR [1].

transistors operating at asymmetric current densities (Q_1 and Q_2), producing the linear PTAT slope as given in (2.7), where n is the emitter area ratio between the transistors. However, this standalone arrangement remains purely theoretical, because maintaining a precise, fixed current density ratio under changing supply voltages is impossible using passive components alone [1].

To enforce the strict current matching required for PTAT generation and isolate the output from V_{DD} supply fluctuations, practical BGR networks integrate high-gain operational amplifiers and active current-mirror loops, as illustrated in Figure 2.6.

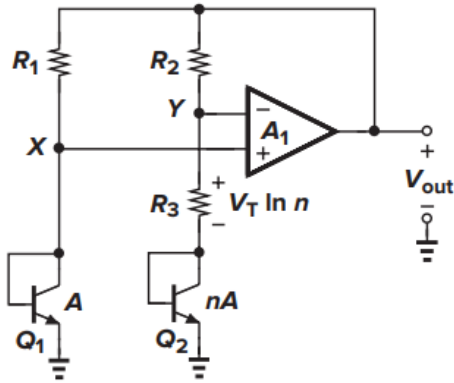


Figure 2.6: Op-amp based implementation of the Bandgap Reference circuit enforcing fixed node potentials [1].

In Figure 2.6, the high open-loop gain of the operational amplifier maintains equal base potentials at nodes X and Y, forcing the voltage drop across resistor R_1 to equal ΔV_{BE} and establishing a well-defined PTAT current $I_{PTAT} = \Delta V_{BE}/R_1$. This current is mirrored into the output branch and summed with the CTAT voltage of Q_3 , yielding [1]:

$$V_{out} = V_{BE2} + \left(1 + \frac{R_2}{R_3}\right) V_T \ln n \quad (2.12)$$

For a zero TC one requires $(1 + R_2/R_3) \ln n \approx 17.2$, which can be satisfied by, for example, choosing $n = 31$ and $R_2/R_3 = 4$ [1].

In the practical topology the collector currents are $I_C \approx (V_T \ln n)/R_3$, which are PTAT rather than constant. Repeating the TC derivation with $\partial I_C/\partial T = I_C/T$ modifies (2.6) to [1]:

$$\frac{\partial V_{BE}}{\partial T} = \frac{V_{BE} - (3 + m)V_T - E_g/q}{T} \quad (2.13)$$

indicating a slightly less-negative TC than -1.5 mV/K. In practice, accurate simulations are necessary to predict the TC precisely [1].

A key non-ideality is the input-referred offset voltage V_{OS} of the operational amplifier. Including V_{OS} in the analysis gives [1]:

$$V_{out} = V_{BE2} + \left(1 + \frac{R_2}{R_3}\right) (V_T \ln n - V_{OS}) \quad (2.14)$$

The offset is amplified by the factor $(1 + R_2/R_3)$, and since V_{OS} itself varies with temperature, it directly raises the output TC [1]. To reduce this effect one may: (i) use large transistor dimensions in the op amp to minimise random offset; (ii) ratio the collector currents of Q_1 and Q_2 by a factor m , replacing ΔV_{BE} with $V_T \ln(mn)$; or (iii) place two pn junctions in series in each branch, which doubles V_{BE} and thus reduces the relative contribution of V_{OS} — at the cost of raising V_{out} toward 2.5 V, which is difficult to sustain under low supply voltages [1].

In the op-amp-based topology, the feedback signal returns to both inputs. The negative-feedback factor and positive-feedback factor are respectively [1]:

$$\beta_N = \frac{1/g_{m2} + R_3}{1/g_{m2} + R_3 + R_2} \quad (2.15)$$

$$\beta_P = \frac{1/g_{m1}}{1/g_{m1} + R_1} \quad (2.16)$$

To guarantee overall negative feedback, β_P must be less than β_N , preferably by roughly a factor of two so that the circuit's transient response remains well-behaved with large capacitive loads [1].

The output voltage of the op-amp-based BGR is relatively independent of V_{DD} as long as the open-loop gain of the amplifier is sufficiently high. However, the circuit suffers from the start-up problem: if all transistors carry zero current when the supply is first applied, they may remain off indefinitely because the loop can support a zero-current degenerate bias point. A start-up device — typically a diode-connected MOSFET providing a current path from V_{DD} through the core — drives the circuit out of this state during power-on [1]. The supply must be swept from zero in both DC and transient simulations to verify correct start-up, as more than one degenerate point may exist in complex implementations [1].

Even after first-order cancellation, bandgap references exhibit a finite curvature: the TC is zero at one temperature T_0 but becomes positive or negative elsewhere [1]. This arises from the nonlinear temperature dependence of V_{BE} , producing the residual parabolic error profile — the bandgap belly — that limits standard first-order designs to approximately 20–50 ppm/°C.

Curvature-correction techniques have been proposed to suppress this variation, but in CMOS counterparts they are seldom used reliably because large device offsets and process variations shift the zero-TC temperature randomly from sample to sample, making systematic correction difficult [1].

2.2.3 Low-Voltage Bandgap References

The reference voltage of ≈ 1.25 V implied by (2.10) exceeds the supply voltages available in modern sub-1 V designs. An alternative approach, proposed by Banba et al., circumvents this constraint by summing PTAT and CTAT *currents* and converting the result to an arbitrarily scaled voltage [1]. Placing a resistor R_2 from node Y to ground in parallel with transistor Q_2 (and a matching R_3 in parallel with Q_1) yields, when $R_2 = R_3$:

$$|I_{D4}| = \frac{V_T \ln n}{R_1} + \frac{|V_{BE1}|}{R_2} \quad (2.17)$$

Choosing $(R_2/R_1) \ln n \approx 17.2$ renders $|I_{D4}|$ temperature-independent. This current is then copied and passed through an output resistor R_4 to produce a zero-TC voltage [1]:

$$V_{BG} = \frac{R_4}{R_2} \left[|V_{BE1}| + \frac{R_2}{R_1} V_T \ln n \right] \quad (2.18)$$

The factor R_4/R_2 can be chosen freely, enabling output voltages well below 1.25 V. The minimum V_{DD} required is $|V_{BE1}| + |V_{DS,sat}|$, which for a 10- μ A bias current can be as low as ≈ 0.75 V [1]. Including op-amp offset V_{OS} modifies the output to [1]:

$$V_{BG} = \frac{R_4}{R_2} \left[|V_{BE1}| + \frac{R_2}{R_1} V_T \ln n \right] - \frac{R_4}{R_1 \parallel R_2} V_{OS} \quad (2.19)$$

showing that the offset contribution is amplified by $R_4/(R_1 \parallel R_2)$, and can be minimised only by maximising n [1]. The circuit is implemented with a five-transistor OTA — designed with large dimensions to minimise flicker noise and offset — and must include a start-up mechanism, since the topology has a degenerate zero-current state analogous to the standard BGR [1].

2.2.4 Supply-Independent Biasing and PTAT Current Generation

Before a temperature-independent reference can be generated, a supply-independent bias current must first be established. Bootstrapping a self-referenced loop — where I_{REF} is derived from a copy of I_{out} — eliminates direct V_{DD} dependence. Adding a source-degeneration resistor R_S to break the indeterminate $I_{out} = K \cdot I_{REF}$ degeneracy yields, with K the current-mirror ratio and $\lambda \approx 0$ [1]:

$$I_{out} = \frac{2}{\mu_n C_{ox} (W/L)_N} \cdot \frac{1}{R_S^2} \left(1 - \frac{1}{\sqrt{K}} \right)^2 \quad (2.20)$$

This current is independent of supply voltage (but still process- and temperature-dependent). Long-channel transistors are used throughout to minimise channel-length modulation and reduce flicker noise [1].

A PTAT current is generated by introducing a bipolar core. With two identical PMOS mirrors and bipolar transistors Q_1 (area A) and Q_2 (area nA) in the two branches, forcing equal drain currents causes the circuit to establish [1]:

$$I_{D1} = I_{D2} = \frac{V_T \ln n}{R_1} \quad (2.21)$$

This PTAT current can be augmented with a CTAT component to directly generate a temperature-independent output voltage:

$$V_{\text{REF}} = |V_{BE3}| + \frac{R_2}{R_1} V_T \ln n \quad (2.22)$$

where the size of Q_3 and its V_{BE} are chosen so that the two terms sum to zero TC [1]. PMOS device mismatches introduce second-order errors in this topology.

2.2.5 Speed, Noise, and Output Impedance

Although reference generators are nominally low-frequency circuits, their finite output impedance can degrade the performance of the circuits they bias [1]. For a one-pole op-amp model $A(s) = A_0/(1 + s/\omega_0)$, the output impedance of the BGR core is [1]:

$$Z_{\text{out}}(s) = \frac{R_{\text{out}}}{1 + g_{mP}R_1A_0} \cdot \frac{1 + s/\omega_0}{1 + s/[(1 + g_{mP}R_1A_0)\omega_0]} \quad (2.23)$$

This expression exhibits a zero at ω_0 and a pole at $(1 + g_{mP}R_1A_0)\omega_0$. While $|Z_{\text{out}}|$ is small at low frequencies, it rises to approximately $R_{\text{out}}/\sqrt{2}$ at the pole frequency — close to the full open-loop impedance [1]. A large bypass capacitor at the output node can suppress mid-frequency disturbances, but its value must greatly exceed the coupling capacitance from nearby switching nodes, and its effect on loop stability must be verified [1].

Regarding output noise, the input-referred noise of the op amp $V_{n,\text{op}}$ appears essentially unattenuated at the reference output — since $g_{mP}A_0 \gg g_{mN} \gg R_1^{-1}$ — giving [1]:

$$|V_{n,\text{out}}| \approx V_{n,\text{op}} \quad (2.24)$$

This means that low-noise performance demands a low-noise amplifier in the BGR core, and that even a large bypass capacitor may not suppress low-frequency $1/f$ noise [1]. When the reference current is scaled by a factor N to bias downstream circuits, the noise power scales by N^2 , imposing strict noise requirements on the reference generator in precision systems such as high-resolution ADCs [1].

In summary, the Bandgap Reference provides the temperature- and supply-stable voltage baseline required for all downstream PVT telemetry. Its fundamental precision is bounded by first-order curvature (the bandgap belly), op-amp offset amplification, and output noise, each of which must be addressed through careful topology selection, device sizing, and — for the most demanding applications — higher-order curvature correction or trimming [1], [7], [20].

2.3 Ring Oscillators as PVT Transducers

While the Bandgap Reference provides a stable voltage baseline, active monitoring requires a transducer capable of converting physical variations into measurable digital signals. In modern high-density designs, the Ring Oscillator (RO) has emerged as the preferred sensing element due to its small footprint and high correlation with digital logic paths [1].

An RO consists of an odd number of inverting stages (N) connected in a feedback loop. The oscillation frequency (f_{osc}) is inversely proportional to the total propagation delay of the chain:

$$f_{osc} = \frac{1}{2 \cdot N \cdot t_p} \quad (2.25)$$

where t_p is the average propagation delay of a single inverter. In the deep sub-micron and FinFET regimes, t_p is a complex function of carrier mobility (μ), load capacitance (C_L), and the difference between the supply voltage (V_{DD}) and the threshold voltage (V_{th}). This dependency allows the RO to act as a multi-modal sensor:

- **Process Sensitivity:** Variations in the fin geometry or doping concentrations shift the V_{th} baseline, resulting in a permanent frequency offset for a specific die.
- **Voltage Sensitivity:** Changes in V_{DD} alter the drive current, causing the frequency to scale with supply fluctuations.
- **Temperature Sensitivity:** Thermal changes affect carrier mobility and threshold voltage. In FinFET nodes, mobility degradation typically dominates, meaning the frequency decreases as the temperature rises.

2.4 16-nm FinFET Ring Oscillator Sensitivity Characterization

To demonstrate the practical impact of multi-domain variations on logic gate propagation delay, a basic enabled ring oscillator structure was characterized and simulated within a native 16-nm FinFET process environment. As illustrated in the simulation testbench blueprint in Figure 2.7, the circuit is implemented as a classical homogeneous architecture consisting of six standard inverting buffer stages closed in a feedback loop with a single two-input NAND gate serving as the global synchronization enable control, providing a total of seven inverting delay stages. Using Synopsys PrimeSim/HSPICE tools, the baseline transient behavior of this basic topology was swept across independent voltage and temperature boundaries to map its raw frequency dependencies.

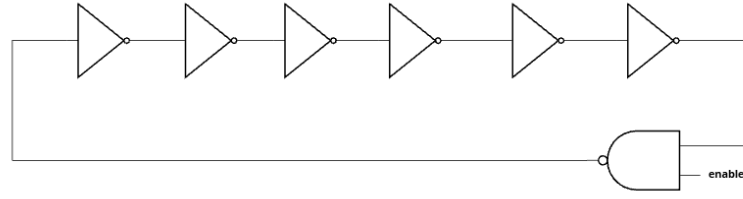


Figure 2.7: Testbench schematic of the basic enabled 7-stage ring oscillator configuration.

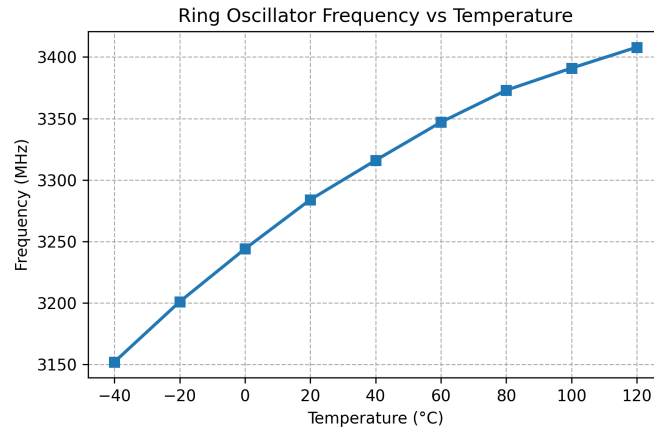
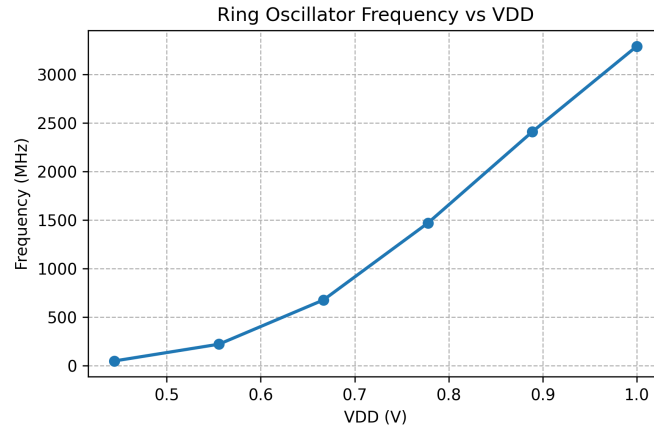


Figure 2.8: Simulated basic RO oscillation frequency versus operating junction temperature.

Figure 2.9: Simulated basic RO oscillation frequency versus supply voltage rail (V_{DD}).

2.4.1 Characterization and Sensing Ambiguity

The simulated electrical and thermal responses of the basic seven-stage topology are plotted in Figure 2.8 and Figure 2.9. As observed in the voltage swept domain of Figure 2.9, the ring oscillator exhibits a highly non-linear, upward-curving voltage-to-frequency translation profile as the supply rail ranges from 0.45 V up to 1.0 V, with the tracking output frequency scaling from ~ 50 MHz

up to approximately 3.3 GHz as dictated by nanometric drive current relationships [15]. Concurrently, the thermal swept response in Figure 2.8 reveals that the oscillation frequency scales from 3152 MHz at -40°C up to 3407 MHz at 120°C . Rather than displaying a conventional thermal deceleration, this positive frequency slope unmasks a clear temperature inversion characteristic typical of scaled nanometric nodes [16]. In this regime, the negative thermal coefficient of the device threshold voltage dominates over carrier mobility degradation, expanding the effective gate overdrive headroom ($V_{DD} - V_{th}$) at elevated temperatures and accelerating the transition delay of the logic loop [1], [16].

While these distinct sensitivities make the basic ring oscillator a highly responsive, compact transducer, they highlight the primary challenge in modern autonomous on-chip metrology: cross-domain sensing ambiguity. Because both an increase in operating temperature and a positive surge in the local supply rail trigger a major positive shift in the oscillator’s raw frequency, a standalone frequency reading becomes mathematically impossible to decipher. This diagnostic overlap is further compounded by localized manufacturing variations; a raw tracking frequency of 3.3 GHz could point to a nominal process baseline under peak voltage conditions, or a fast process corner operating under nominal conditions [9].

This total parametric overlap justifies the hierarchical architecture of the proposed monitoring infrastructure. By utilizing a stable, curvature-compensated voltage baseline to isolate and digitize the static process corner signature at system boot-up, the controller can de-embed manufacturing offsets from real-time tracking loops. Armed with this localized calibration factor, the system can systematically decouple dynamic voltage variations from slower thermal trends, allowing for high-precision distributed telemetry across the die without requiring expensive, per-chip external test equipment calibration.

Chapter 3

Related Work

The continuous scaling of complementary metal-oxide-semiconductor (CMOS) technology has historically driven unprecedented gains in computational density, operating frequency, and energy efficiency. However, as feature sizes descend into deep sub-micron and nanoscale regimes, the semiconductor industry has collided with severe physical boundaries, signalling what is widely characterised as the twilight of Moore’s Law [2]. In this modern era of energy-quality scaling, integrated circuits (ICs) are increasingly forced to operate at or near their absolute margins of performance and power efficiency [2]. Consequently, deterministic design paradigms have grown obsolete; the traditional approach of allocating static, worst-case guard-bands to absorb environmental and manufacturing fluctuations yields unsustainable over-engineering, penalising energy performance under nominal operating conditions [2]. To preserve the scaling benefits of advanced nodes, modern architectures must pivot toward adaptive, self-characterising systems capable of sensing and compensating for Process, Voltage, and Temperature (PVT) variations dynamically and in real time [3], [4].

This chapter provides a rigorous technical evaluation of the current state of the art across the three fundamental sensing domains of a unified PVT monitoring infrastructure. Section 3.1 examines the voltage reference architectures that form the absolute baseline of any sensing system. Sections 3.2, 3.3, and 3.4 then analyse the state of the art in process, voltage, and temperature monitoring respectively, culminating in a comparative synthesis that motivates the unified infrastructure proposed in this dissertation. Throughout, particular emphasis is placed on the recurring sub-micron trade-offs that govern every topology surveyed: quiescent power budget, subthreshold leakage behaviour at advanced nodes, temperature coefficient linearity, supply voltage sensitivity, and layout area consumption.

3.1 Voltage References and Bandgap Architectures

The architecture of an integrated PVT monitoring system is structured hierarchically, wherein individual sensing loops must be decoupled from ambient fluctuations to ensure data integrity [3]. Precision voltage references serve as the absolute baseline for the entire infrastructure, because

any drift in the reference propagates directly into the error budget of every downstream sensor. The operational ceiling of the system is therefore fundamentally bound by the thermal stability, line sensitivity, and Power Supply Rejection Ratio (PSRR) of the primary reference generator [1], [7], [18].

The foundational PTAT/CTAT summation principle, the curvature term introduced by the non-linear $V_{BE}(T)$ dependence, and the resulting first-order temperature coefficient (TC) limitation of 20 to 50 ppm/°C were established in Chapter 2 [1], [19]. This section surveys how the state of the art addresses that residual curvature and the associated PSRR, area, and power trade-offs across modern bandgap reference (BGR) topologies, before culminating in an in-depth examination of the FinFET-native, current-mode architectures selected for this dissertation.

3.1.1 High-Order Curvature Compensation

ZTC-point compensation [21] offers a particularly elegant route to high-order compensation by biasing a MOSFET stack at its Zero-Temperature-Coefficient (ZTC) point, where the opposing temperature dependencies of carrier mobility (μ) and threshold voltage (V_{th}) cancel in the drain current [16]. For a MOSFET in saturation,

$$I_D = \frac{1}{2} \mu(T) C_{ox} \frac{W}{L} (V_{GS} - V_{th}(T))^2, \quad (3.1)$$

with $\mu(T) = \mu(T_0)(T/T_0)^{-\kappa}$ and $V_{th}(T) = V_{th}(T_0) - \alpha_{V_{th}}(T - T_0)$. Differentiating I_D with respect to T and setting $\partial I_D / \partial T = 0$ yields the ZTC bias condition

$$V_{GS,ZTC} - V_{th}(T_0) = \frac{2\beta}{\eta} \cdot \frac{kT_0}{q}, \quad (3.2)$$

where β is the mobility temperature exponent ($\mu \propto T^{-\beta}$) and η is the temperature coefficient of V_{th} , following the notation of Jiang et al. [21]. Jiang et al. [21] exploited this in 65-nm CMOS by biasing a composite stack of four NMOS devices at $V_{gs,ZTC}$ and tapping V_{REF} from a resistive divider on the stack's terminal node. Because the reference is generated purely from MOS devices at a current-domain zero-TC point, no bipolar CTAT branch or active curvature-correction amplifier is required, and the design reported a TC of 5.6 ppm/°C together with an 87-dB PSRR. The trade-off is that the ZTC point is itself process-dependent through κ and $\alpha_{V_{th}}$, so V_{REF} accuracy is more exposed to die-to-die process spread than in a BJT-based design unless the bias current is itself trimmed.

$T \ln T$ compensation [6] retains the bipolar core but synthesises an explicit correction current proportional to the curvature term itself. Expanding $V_{BE}(T)$ as established in Chapter 2, the residual non-linear component scales as

$$\Delta V_{curv}(T) = -(\eta - x) \frac{kT}{q} \ln\left(\frac{T}{T_r}\right), \quad (3.3)$$

where x is the exponent characterising the temperature dependence of the collector current ($I_c = DT^x$), which a conventional PTAT/CTAT summation cannot cancel because it is neither linear in T nor constant. Lohitsiri and Panitantom [6] generates three branch currents from BJT pairs biased through op-amp feedback loops, replicates them through cascode current mirrors, and sums them across tapped resistors with ratios determined by a least-squares optimization over the full temperature range. This explicit, computationally optimised cancellation achieves a TC of approximately 2.39 ppm/°C from -40°C to 125°C , at the cost of additional mirror branches and associated bias current and area.

The compact curvature-corrected output driver of Zhuang et al. [20] pursues the same curvature-cancellation goal but minimises the area and complexity penalty typically incurred by multi-amplifier $T \ln T$ -type schemes. The architecture partitions the design into six blocks, including a PTAT generator with built-in over-temperature protection, a base-current compensation branch, and a compact output driver that injects a second-order correction current $I_{corr}(T)$ directly into the output node such that

$$V_{REF}(T) = V_{BE}(T) + K \cdot V_T(T) + I_{corr}(T) \cdot R_{out}, \quad (3.4)$$

with $I_{corr}(T)$ designed to approximate the quadratic and logarithmic residual of $V_{BE}(T)$ using a small number of transistors rather than a full optimised resistor network. The result is sub-5 ppm/°C stability with a layout footprint substantially smaller than conventional multi-amplifier $T \ln T$ topologies.

A line-sensitivity-enhanced BGR [7] addresses a complementary figure of merit, Line Sensitivity (LS), defined as

$$LS = \frac{1}{V_{REF}} \left. \frac{\partial V_{REF}}{\partial V_{DD}} \right|_{T=T_0}. \quad (3.5)$$

Rather than improving the loop gain of the bandgap core's own feedback amplifier, the architecture prepends an LS-enhancement block, a cascade of NMOS current-source stages, that pre-regulates an internal rail before it reaches the bandgap core. Because the core's LS scales with the inverse of the open-loop gain of this added stage, LS is suppressed without requiring a redesign of the BGR core's own amplifier, demonstrating that LS and TC can largely be decoupled and optimised independently.

A multi-output BGR [22] targets the system-level trade-off between reference precision and the silicon area cost of replicating a precision core for every analog subsystem. A single bandgap core drives a tapped resistive divider that generates five simultaneous output levels, so that each downstream block receives a dedicated, precisely ratioed rail from a common precision baseline while the core still achieves a TC of 1.12 ppm/°C from -40°C to 125°C . The trade-off is that the loading and switching behaviour of any one output can perturb the others through the shared divider.

3.1.2 Amplifier-Free and Sub-Threshold Topologies

A parallel research direction trades a fraction of accuracy for substantial reductions in quiescent power, supply voltage, and silicon area by eliminating the operational amplifier from the BGR core entirely. Lee [23] first showed that PTAT-CTAT balance can be enforced through an amplifier-free feedback loop, where a cross-coupled NMOS pair performs the implicit error correction by modulating the current split between two diode-connected BJT branches. Zou et al. [24] extended this concept with high-order compensation injected through a passive resistor network rather than an active multiplier, again avoiding any operational amplifier or its associated frequency-compensation overhead.

The deepest treatment in this amplifier-free family belongs to the sub-200 nW all-in-one BGR of Huang et al. [18]. The core employs two vertical NPN BJTs in an $N:1$ emitter-area ratio, with all MOSFETs designed to operate in the subthreshold region. The amplifier is replaced by two interacting transistor-level feedback loops, a Negative Feedback Loop and a Positive Feedback Loop, with the former dominant in loop gain to stabilise the circuit. The PTAT component is generated from

$$V_{R1} = V_{BE2} - V_{BE1} = V_T \ln N, \quad (3.6)$$

and the reference voltage is

$$V_{REF} = V_{BE2} + \frac{2R_2}{R_1} V_T \ln N. \quad (3.7)$$

The same LS-enhancement principle from Li et al. [7] is applied via a cascode pre-regulation stage ahead of the core. The measured total power of 192 nW, generating both a voltage and a current reference simultaneously, makes this topology directly relevant to always-on IoT monitoring nodes, at the cost of reduced design margin compared to amplifier-based loops. Also amplifier-free, Ding et al. [25] pushed the sub-threshold philosophy toward sub-1-V operation in a full CMOS process, generating a reference of $3\Delta V_{GS}$ from a thick-oxide and thin-oxide device pair sharing a common drain.

3.1.3 Sampled and Offset-Corrected BGRs

A related but distinct low-noise, high-precision strategy retains the operational amplifier but uses it either intermittently or as part of an in-situ self-correction loop, rather than as a continuous-time error amplifier biasing a fixed PTAT/CTAT network. The sampled BGR of Palani et al. [26] samples the ΔV_{BE} signal onto switched-capacitor pairs driven by non-overlapping clock phases, reconstructing the PTAT term at the amplifier's summing node only during the active sampling phase. This periodic refresh suppresses accumulated $1/f$ noise and long-term drift relative to a continuously-biased core, at the cost of a sampled-data output that must be held by an external buffer or switched-capacitor filter between refresh cycles.

The 0.8-V BGR with automatic current-controlled feedback of Cao et al. [27] instead targets process-induced offset directly rather than only temperature drift. Three loops are integrated: an Automatic Current-Controlled (ACC) feedback loop built around a Wilson-like mirror, a CTAT

current generator, and an output stage summing the corrected PTAT and CTAT branches. The ACC loop senses the mismatch-induced offset

$$\Delta V_{BE,os} = V_{BE1} - V_{BE2} \big|_{I_{C1}=I_{C2}} \neq 0 \quad (3.8)$$

and feeds back a correction current that re-balances I_{C1} and I_{C2} until $\Delta V_{BE,os}$ is nulled in situ, removing the need for external trimming or chopping. Operating at an 0.8-V supply with only $1.9 \mu\text{W}$ consumption, the design reports a 3σ output accuracy of 1.58% across process corners, demonstrating that closed-loop offset correction can substitute for the area and complexity of post-fabrication trimming at ultra-low supply voltages.

3.1.4 MOS-Only Architectures

As silicon area becomes the primary cost driver in nanometric nodes, and as native BJTs become increasingly costly to integrate in FinFET processes, a further research direction replaces the bipolar junction with MOSFET-only elements. Sahu et al. [8] generates both PTAT and CTAT currents from sub-threshold MOSFETs alone, with the PTAT term derived from $\Delta V_{GS} = V_T \ln((W_2/L_2)/(W_1/L_1))$ between two differently-sized NMOS devices and the CTAT term from the native negative TC of V_{th} itself. The fundamental limit of this family is Pelgrom's Law [9], $\sigma^2(\Delta V_{th}) = A_{V_{th}}^2/(WL) + S_{V_{th}}^2 D^2$, which forces transistors to be sized up to average out random dopant fluctuations, eroding the very area advantage the MOS-only approach sought. Eberlein and Pretl [17] surveys the structural difficulty of porting bipolar bandgap cores into 16-nm and 7-nm FinFET nodes given the absence of a native substrate PNP, motivating the use of N-well junction diodes as the CTAT element in FinFET-targeted designs, the approach adopted directly by the topology selected for this dissertation.

3.1.5 Sub-1-V Current-Mode FinFET BGRs: PMOS and NMOS Sensing Variants

The two architectures examined in this subsection form the anchor of the voltage reference comparison conducted later in this chapter and are the candidates carried forward into the simulation work of Chapter 4.

The sub-1-V FinFET BGR of Low et al. [28] adapts the current-mode summation architecture originally proposed by Banba et al. to 14-nm FinFET technology, replacing the bipolar elements of the original architecture with diode-connected FinFET devices biased in subthreshold. The circuit employs a three-branch PMOS current mirror of identical dimensions, driven by a single op-amp feedback loop, to force equal currents $I_1 = I_2 = I_3$ into three legs. In subthreshold operation, the drain current of a diode-connected FinFET follows an exponential dependence on gate-to-source voltage analogous to a BJT's I - V_{BE} law,

$$I_D = K e^{V_{GS}/nV_T}, \quad (3.9)$$

where K is a technology- and geometry-dependent constant, n the subthreshold slope factor, and V_T the thermal voltage. The equal mirrored currents are forced through two diode-connected FinFETs sized in a 1: N ratio, producing two distinct gate-source voltages whose difference is

$$\Delta V_f = V_{f1} - V_{f2} = V_T \ln(N), \quad (3.10)$$

which is linear in temperature and therefore PTAT. The op-amp forces its input nodes to the same potential, so ΔV_f appears directly across a resistor R_3 , generating a PTAT current $I_{PTAT} = \Delta V_f / R_3$. In parallel, the gate-source voltage of the diode-connected FinFET, which decreases with temperature, is dropped across R_2 , producing a CTAT current $I_{CTAT} = V_{f1} / R_2$. The PTAT and CTAT currents are summed and mirrored into a load resistor R_4 to produce the final reference voltage,

$$V_{REF} = R_4 \left(\frac{V_{f1}}{R_2} + \frac{\Delta V_f}{R_3} \right). \quad (3.11)$$

Unlike a classical bandgap design, the zero-TC point here is not solved analytically from a single closed-form resistor-ratio equation; instead, R_2 , R_3 , R_4 , and the device sizing ratio N are jointly optimised through iterative simulation and bias-current adjustment to balance the CTAT and PTAT contributions. Simulated in a 14-nm FinFET process at an 0.8-V supply, the design reports a reference voltage of 320.7 ± 3.5 mV over 0°C to 100°C , equivalent to a TC of approximately 222 ppm/ $^\circ\text{C}$, drawing 0.42 mA for a total static power of 336 μW . The residual upward slope of $V_{REF}(T)$ indicates that the PTAT term is not fully balanced against the CTAT term, leaving the design first-order under-compensated relative to the sub-10 ppm/ $^\circ\text{C}$ results achieved by the curvature- and offset-corrected topologies discussed earlier in this chapter. A supply sweep from 0.79 V to 0.81 V further reveals a line sensitivity of approximately 0.9 mV per mV of V_{DD} , since the architecture relies on a simple current-mirror bias rather than a dedicated LS-enhancement stage such as that of Li et al. [7] or Huang et al. [18]. The authors note that cascode current mirrors, regulated biasing, or explicit curvature compensation are natural directions to improve both line regulation and the residual TC, situating this topology as a deliberately simple, area-efficient, and first-order adaptation of the Banba architecture to FinFET technology.

By appropriately selecting the resistor ratios together with the sizing ratio m , the CTAT and PTAT contributions are balanced at first order, yielding a temperature-stable V_{REF} [29].

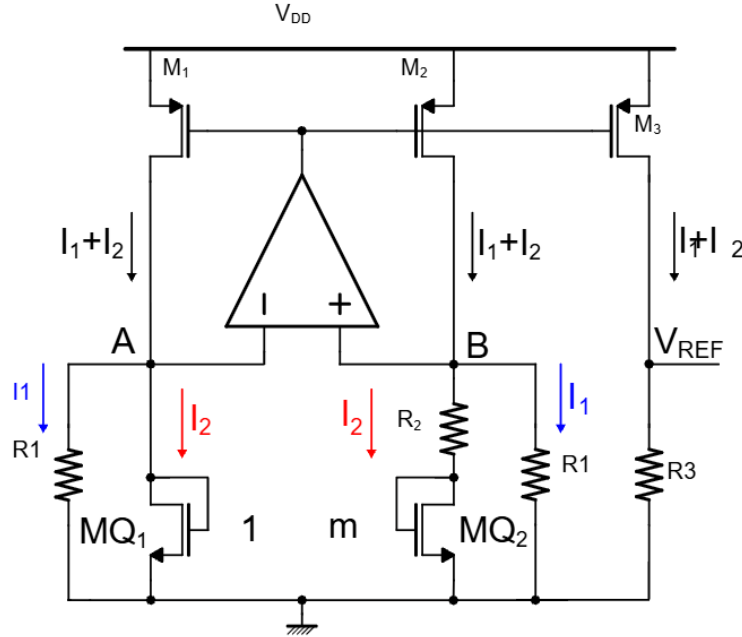


Figure 3.2: Current-mode BGR employing diode-connected NMOS transistors with PMOS current mirror [29].

Although the PMOS and NMOS variants are structurally analogous, the substitution of the sensing transistor type introduces non-trivial engineering trade-offs that affect temperature stability, minimum supply voltage, line sensitivity, and silicon area. Regarding minimum supply voltage, the PMOS variant stacks the diode-connected devices between the supply rail and the virtual ground enforced by the op-amp, requiring headroom of $V_{GS,PMOS} + V_{DS,mirror}$, whereas the NMOS variant connects its sensing devices between the summing nodes and ground, requiring $V_{GS,NMOS} + V_{DS,mirror}$ measured from V_{DD} to ground through the PMOS mirror. Since both variants operate in subthreshold, where V_{GS} is inherently low, the difference in minimum V_{DD} is modest but process-dependent, with the NMOS variant offering a marginal advantage in processes where $|V_{TH,PMOS}| > V_{TH,NMOS}$. Concerning the threshold voltage temperature coefficient, the CTAT term in both topologies originates from $\partial V_{TH}/\partial T$, whose magnitude differs between PMOS and NMOS devices by approximately 10% to 20% in typical 180-nm and 130-nm CMOS nodes, with NMOS devices generally exhibiting a slightly larger magnitude [29], requiring the resistor ratio to be re-optimised for each device type. With respect to sensitivity to process mismatch, both variants share the same structural sensitivity governed by Pelgrom's Law [9], but NMOS devices in standard planar CMOS typically exhibit smaller $A_{V_{TH}}$ mismatch coefficients than their PMOS counterparts, implying that the NMOS-based topology may achieve tighter V_{REF} spread for a given device area, an area benefit particularly attractive in the FinFET context studied here. In terms of flicker noise, PMOS devices exhibit lower flicker noise than NMOS devices in most standard CMOS processes, so the NMOS variant carries a higher intrinsic noise floor that

may require additional low-pass filtering or a higher-bandwidth correction loop. Finally, regarding the current mirror and amplifier topology, the NMOS configuration partially decorrelates the sensing and mirroring error sources relative to the PMOS configuration, and its input common-mode voltage sits at an intermediate level compatible with a rail-to-rail input stage, potentially relaxing op-amp design constraints relative to the PMOS configuration.

In summary, the NMOS-based topology of Moisello et al. [29] constitutes a structurally faithful implementation of the same current-mode bandgap-like architecture as the reference design of Low et al. [28], while offering the potential for smaller threshold voltage mismatch, lower minimum supply voltage in certain process corners, and relaxed op-amp common-mode requirements, at the cost of higher flicker noise and a requirement to re-optimize the resistor network and sizing ratio to account for NMOS-specific temperature coefficients. These two topologies are jointly carried forward as the anchor candidates for the voltage reference path of the proposed infrastructure.

3.1.6 Precision Current References

Closely related to voltage references, precision current references are equally critical for biasing temperature and process sensors in a distributed monitoring infrastructure. The sub-ranging PVT-insensitive current reference of Park et al. [30] is built around a Widlar-like current mirror using a BJT pair sharing an emitter resistor, with the loop closed by an operational amplifier enforcing the required ΔV_{BE} across that resistor,

$$I_{REF,raw}(T) = \frac{\Delta V_{BE}}{R_1} = \frac{1}{R_1} \cdot \frac{kT}{q} \ln(N), \quad (3.15)$$

which is inherently PTAT and, left uncorrected, would exhibit a TC dominated by the temperature coefficient of R_1 itself. The output current is instead drawn through a second resistor whose effective temperature coefficient is digitally tuned via a tapped series combination, with the tap ratio selected per temperature sub-band by a dedicated sub-range detection circuit that partitions the operating range into three segments. This piecewise-linear, digitally tap-selected correction achieves 11.4 ppm/°C from -20°C to 125°C without the silicon area or power overhead of a continuous non-linear compensation network, illustrating a complementary path to high precision relative to the continuous curvature-correction voltage references discussed above. The composite-transistor PTAT current reference of Hanumantha Rao et al. [31], which is structurally relevant to the temperature sensor bias chain selected for this dissertation, is treated separately in Section 3.4.3.

3.2 Process Sensors

As established in Chapter 2, on-chip process sensors are necessary because the global process corner no longer accurately characterises die-level behaviour in deep-submicron nodes, where within-die (WID) variability dominates [9]. The following subsections examine the principal architectural strategies for on-chip process characterization, from ring oscillator-based monitors to

sub-threshold leakage sensors and reconfigurable multi-modal systems, culminating in an exhaustive treatment of the three topologies selected for deepest analysis: the ratioed NAND/NOR sensor of Chandan et al. [32], the ratioed-inverter sensor of An et al. [14], and the inhomogeneous ring oscillator of Fujimoto et al. [33].

3.2.1 Ring Oscillator-Based Process Monitors

The Ring Oscillator (RO) is the dominant transducer for on-chip process characterization owing to its structural simplicity, compact footprint, and direct correlation with the propagation delay of real logic paths. An RO consists of an odd number N of inverting stages connected in a feedback loop, producing an oscillation frequency

$$f_{\text{osc}} = \frac{1}{2 \cdot N \cdot t_p}, \quad (3.16)$$

where t_p is the average propagation delay of a single inverter stage. Per the Alpha-Power law [15], this delay is

$$t_p \propto \frac{C_L \cdot V_{DD}}{(V_{DD} - V_{th})^\alpha}, \quad (3.17)$$

with C_L denoting the load capacitance, V_{DD} the supply voltage, V_{th} the threshold voltage, and α the velocity-saturation exponent, typically between 1 and 2 for short-channel devices. Because f_{osc} depends on V_{th} and the carrier mobility μ through t_p , the oscillation frequency constitutes a direct analogue of the local process corner.

A basic current-starved RO process monitor was proposed by López et al. [5], comprising a chain of N current-starved inverter stages followed by a frequency divider, where dedicated bias voltages applied to PMOS and NMOS cascode transistors restrict the maximum charging and discharging current of each stage. The principal engineering trade-off is sensitivity versus supply rejection, since any supply or temperature perturbation that shifts the bias voltages produces an erroneous process reading. Alternative low-complexity topologies in the current-starved family include [34], whose current-steering logic stages decouple the process measurement from supply voltage fluctuations at the cost of additional cascode loading that lowers the achievable oscillation frequency and dynamic range, and [35], an FD-SOI corner detector exploiting diode-connected complementary loads to isolate NMOS and PMOS sensitivity, but whose 28-nm FD-SOI body-bias dependence and modest 2% per-corner frequency spread constrain its applicability outside panel-driver SoCs.

3.2.1.1 Ratioed-Inverter RO Process Sensor

An et al. [14] developed an all-digital on-chip process sensor based on a ratioed-inverter RO. The key insight is the deliberate use of asymmetric NMOS and PMOS sizing within each delay cell. Each cell pairs a ratioed inverter, in which one device is intentionally oversized relative to its complementary device, with a standard symmetric CMOS inverter. Two complementary delay

cell variants are employed, an NMOS-type cell with an enlarged NMOS pull-down network and a PMOS-type cell with an enlarged PMOS pull-up network.

The propagation delay of the ratioed inverter is asymmetric by design. For the NMOS-type cell, the fall time t_{pHL} is shortened relative to the rise time t_{pLH} because the oversized NMOS transistor provides a larger pull-down current,

$$t_{pHL} \propto \frac{C_L \cdot V_{DD}}{k_n \cdot (V_{DD} - V_{th,n})^\alpha}, \quad t_{pLH} \propto \frac{C_L \cdot V_{DD}}{k_p \cdot (V_{DD} - V_{th,p})^\alpha}, \quad (3.18)$$

where k_n and k_p are the effective drive strengths of the NMOS and PMOS devices, respectively, set by the ratioed sizing. The average cell delay $t_p = (t_{pHL} + t_{pLH})/2$ is therefore weighted towards $V_{th,n}$ for the NMOS-type ring and towards $V_{th,p}$ for the PMOS-type ring. By comparing the frequencies of the two rings, the system decomposes the global process shift into NMOS-dominant and PMOS-dominant components, enabling the distinction of a Fast-Slow (FS) corner from a Slow-Fast (SF) corner, a capability entirely absent in symmetric inverter-chain designs. The architecture is fully digital and requires no analogue readout circuitry, keeping the area and power overhead low, though it remains susceptible to temperature cross-sensitivity because the mobilities μ_n and μ_p carry opposing temperature coefficients that can partially mask the process signature.

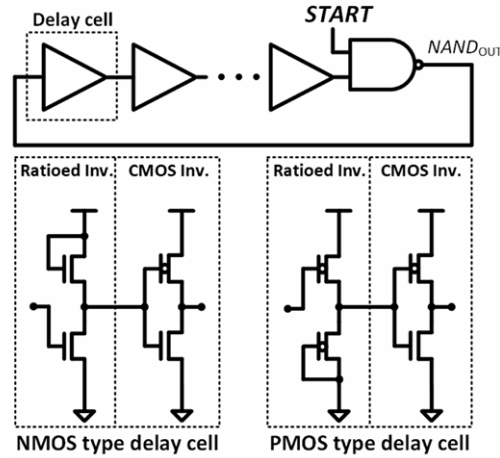


Figure 3.3: Ratioed-inverter RO process sensor by An et al. [14].

3.2.1.2 Ratioed NAND/NOR Process Sensor

Chandan et al. [32] proposed a process sensor topology built from two complementary ratioed delay cells, an N-Ratioed NAND cell and a P-Ratioed NOR cell, each designed to isolate the sensitivity of one transistor type while suppressing the other's contribution to delay. The motivation is that a conventional CMOS inverter-based RO tracks process variation only weakly, since both the pull-up and pull-down paths contribute to delay simultaneously and their variations partially

cancel. The ratioed cells instead force the signal-transition path through transistors of a single type, so that the threshold-voltage variation of that type dominates the measured delay.

The N-delay cell retains a standard NAND-style pull-down network gated by the cell input, but replaces the conventional PMOS pull-up with two diode-connected NMOS devices in parallel gated to V_{DD} . When the input is low, the NMOS pull-down switches off and the diode-connected NMOS pull-up weakly pulls the internal node high, passing a degraded logic-1 into the following CMOS NAND inverter stage, which restores it to logic-0. Because this logic-0 transition path is composed entirely of NMOS devices, the resulting delay,

$$t_{pHL}^{\text{N-cell}} = f(V_{th,n}), \quad (3.19)$$

is rendered more sensitive to a shift in the nMOS threshold voltage, while the logic-1 transition follows a conventional ring-oscillator path and is comparatively insensitive to $V_{th,n}$. The P-delay cell is constructed symmetrically, with two series PMOS devices forming the pull-up network and two diode-connected PMOS devices in parallel gated to ground forming the pull-down network, restored by a CMOS NOR stage, yielding a pull-up delay dominated by the PMOS threshold voltage,

$$t_{pLH}^{\text{P-cell}} = g(V_{th,p}). \quad (3.20)$$

Each ring is implemented as a 21-stage RO, twenty ratioed buffer or delay cells of the appropriate type followed by a single CMOS NAND or CMOS NOR closing stage. This final stage is required because a ring composed purely of ratioed-logic cells would have its oscillations converge rather than sustain; the closing stage restores adequate noise margin and produces steady oscillation. The use of twenty buffer stages also helps average out within-die variation, since additional stages reduce WID variance and make the spatially correlated process variation more dominant in the measured oscillation frequency, at the cost of increased area and power.

Sensitivity is quantified empirically through the coefficients

$$k_N = \frac{\Delta T/T_0}{\Delta V_n/V_{n0}}, \quad k_P = \frac{\Delta T/T_0}{\Delta V_p/V_{p0}}, \quad (3.21)$$

where T_0 is the oscillation period at the typical corner, ΔT is the period shift observed at a given process corner, and ΔV_n , ΔV_p are the corresponding threshold-voltage shifts from their typical-corner values. Simulated in a 45-nm technology, the proposed design achieves an N-sensitivity of 4.086 and a P-sensitivity of 7.223, compared to 1.934 and 3.189 respectively for the prior ratioed-inverter sensor of An et al. [14], more than double its sensitivity, obtained alongside a power consumption of 2.902 mW, approximately 12% lower than the ratioed-inverter design, though at the expense of additional layout area introduced by the extra ratioed devices per cell.

Frequency-versus-temperature sweeps across the five standard process corners confirm the intended separation between rings. For the N-delay-cell ring, the observed frequency ordering is FF greater than FS greater than TT greater than SS greater than SF, consistent with FF and FS both corresponding to a fast nMOS while SF and SS correspond to a slow nMOS. The P-delay-cell ring

instead orders as SF greater than FF greater than TT greater than SS greater than FS, since SF and FF share a fast pMOS while FS and SS share a slow pMOS. This separation is significant because conventional inverter-based sensors are typically more sensitive to the FF and SS corners while perceiving the mixed SF and FS corners as close to TT, making a highly ratioed approach necessary to reliably distinguish TT from SF and FS.

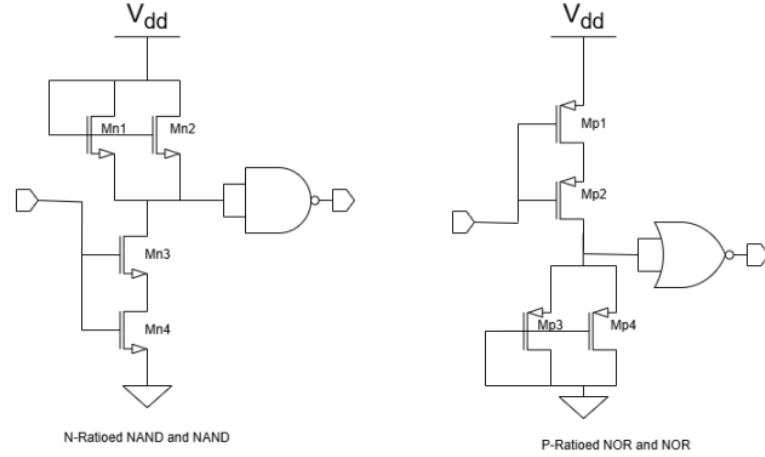


Figure 3.4: Proposed delay-cell design by Chandan et al. [32]: N-Ratioed NAND delay cell (left) and P-Ratioed NOR delay cell (right).

3.2.2 Sub-Threshold Leakage-Based Process Sensing

An alternative sensing paradigm exploits the exponential dependence of sub-threshold drain current on V_{th} to achieve far greater process sensitivity than any saturation-regime oscillator. In the weak inversion region, the drain current follows

$$I_{\text{sub}} \approx I_0 \cdot \frac{W}{L} \cdot \exp\left(\frac{V_{GS} - V_{th}}{nV_T}\right), \quad (3.22)$$

where $V_T = kT/q$ is the thermal voltage and n is the sub-threshold swing coefficient. This exponential dependence acts as a physical amplifier for process variations, a 10 to 20 mV shift in V_{th} , which manifests as only a 2% frequency deviation in a saturation-regime oscillator, produces a 20% to 50% change in sub-threshold leakage current. Lyu and Shi [11] exploits this amplification with a leakage inverter biased at $V_{GS} = 0$ V, in which the output node settles at the voltage where the off-state leakage currents of a series PMOS and NMOS pair balance, yielding an output swing in response to a process corner change that is exponentially amplified relative to the corresponding frequency shift in a linear-regime RO. The principal engineering trade-off is that leakage current is also a strong function of temperature through $V_T = kT/q$ and through the temperature dependence of V_{th} itself, making thermal management and cross-sensitivity correction essential for accurate process-only characterization [2]. Islam et al. [36] extended this leakage-based principle to a ratio-metric readout scheme effective from 0.5 V to 1.2 V without re-calibration, at the cost of long integration windows that introduce latency unsuitable for high-bandwidth real-time monitoring.

3.2.3 Within-Die Variability and Reconfigurable Monitors

As technology scales toward FinFET and Gate-All-Around (GAA) architectures, the assumption that a single oscillator accurately represents the full process state of a die becomes increasingly unreliable due to the growing stochastic component of WID variability.

Fujimoto et al. [33] addressed this limitation through the concept of an inhomogeneous ring oscillator (IHRO), in which each stage is composed of a structurally distinct transistor topology. The ring chain mixes six cell types: a standard CMOS inverter, an inverting NAND-based cell, a cell with an NMOS pass gate at the output, a cell with a PMOS pass gate at the output, a cell with an NMOS pass gate at the input, and a cell with a PMOS pass gate at the input. Each cell type exhibits a distinct sensitivity vector in the PVT space by virtue of its structural differences, including series and parallel stacking, pass-gate loading, and different V_{th} flavour assignments. The oscillation frequency of an IHRO stage is determined by the propagation delay weighted across all N distinct cell types,

$$f_{\text{IHRO}} = \frac{1}{2 \sum_{i=1}^N t_{p,i}}, \quad (3.23)$$

where $t_{p,i}$ is the delay of the i -th cell type. Because consecutive stages see different device geometries and V_{th} flavours, the total oscillation period integrates the stochastic process landscape of multiple adjacent device instances rather than accumulating the delay of a single repeated cell. This makes the IHRO frequency a fingerprint of the local process distribution rather than a global average, capturing the stochastic correlation structure of adjacent device mismatches in a manner that no homogeneous ring can replicate. The primary engineering trade-off is that matching the nominal delays of all six cell types to ensure balanced ring oscillation requires careful ratioed sizing, and the mixed-topology layout introduces additional systematic variability from cell-to-cell proximity effects.

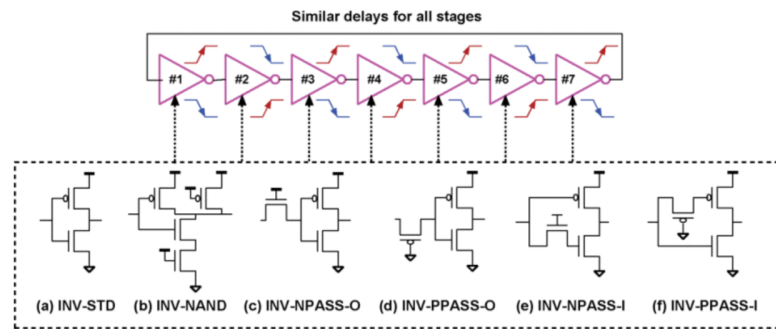


Figure 3.5: Inhomogeneous RO for WID variability characterization by Fujimoto et al. [33].

Kishimoto et al. [37] extended the multi-modal philosophy by implementing a reconfigurable ring oscillator that can be dynamically configured to monitor either temperature or process variations within a single hardware block, with configuration bits selecting between stages of high V_{th} sensitivity for process mode and stages of high mobility temperature coefficient for temperature

mode. The key engineering advantage is dual-mode operation within a single silicon block, reducing the total area overhead for combined PVT monitoring, at the cost of additional digital control logic and shift register overhead required to load the configuration bits.

3.3 Voltage Sensors

As established in Chapter 2, accurate on-chip voltage sensing is complicated by the inherent recursive dependency on a stable reference voltage, the sensitivity of gate propagation delay to supply fluctuations via the Alpha-Power law [15], and the wide-bandwidth nature of IR-drop and inductive ringing events in modern power delivery networks. This section surveys representative voltage sensor implementations across reference-free monitors, event-driven detectors, and delay-chain or frequency-domain encoders, before providing an exhaustive treatment of the three delay-chain and resistive architectures selected for deepest analysis: the delay-chain sensor of Chen et al. [4], the controllable delay-chain sensor of Lou et al. [10], and the dual-mode resistive sensor of Horng et al. [13].

3.3.1 Reference-Free Voltage Monitoring

Ituero et al. [38] provided a systematic taxonomy of on-chip power supply monitors, classifying existing designs according to three primary sensing principles: reference-based monitors, reference-free monitors, and event-driven monitors. This framework is fundamental for understanding the trade-offs between absolute accuracy, silicon area, and supply-noise bandwidth that govern the selection of a monitoring strategy within a given SoC context. Reference-based designs offer higher absolute accuracy but incur the area and power overhead of an on-chip voltage reference generator, whereas reference-free designs eliminate this dependency at the expense of accuracy that relies on the statistical matching of process-dependent circuit elements.

Aparicio et al. [39] demonstrated a 2.64 pJ reference-free power supply monitor capable of operating across a wide temperature range, using parallel Over/Undershoot Detector paths and a stacked voltage-to-delay tap network sized such that the delay exhibits a temperature coefficient that partially cancels the intrinsic temperature sensitivity of the delay chain. The key engineering trade-off is that eliminating the reference reduces area and power at the cost of making the trigger threshold process-dependent, requiring post-fabrication calibration to guarantee absolute accuracy. Lower-area alternatives in this same reference-free family include the thermometer-code ladder sensor of [40], whose comparator-driven resistive ladder taps V_{DD} directly but whose accuracy is limited by uncanceled comparator offset, and the ratiometric calibration-free monitor of [41], which achieves process independence by deriving both signal and reference nodes from a shared bandgap core at the cost of retaining a non-trivial bandgap area and quiescent current overhead.

3.3.2 High-Bandwidth and Event-Driven Voltage Detection

Vijayakumar et al. [12] addressed the challenge of high-frequency power supply noise detection in microprocessors by employing a voltage-divider front end whose output drives an inverter chain and a Schmitt trigger to produce a clean digital pulse whenever a noise event drives V_{DD} below a critical threshold, suitable for building a statistical histogram of supply noise occurrences via a digital event counter. The architecture achieves very high detection bandwidth with low power consumption since no continuous digitization is performed, but it provides only a binary threshold indicator rather than a quantitative voltage measurement.

Chiang and Tam [42] proposed a CMOS voltage sensor with a process-invariant threshold targeting passive UHF RFID transponders, comparing a resistively divided harvested input voltage against a stable V_{REF} generated by a bandgap reference block. Process invariance is achieved by biasing the threshold detection transistor at its ZTC point, where $\partial V_{th}/\partial T|_{ZTC} = 0$, ensuring the trigger voltage remains stable across fabrication corners and temperature excursions without laser trimming or digital calibration. The primary trade-off is that the ZTC bias point is process-node-specific and must be characterised prior to tapeout.

3.3.3 Delay-Chain and Frequency-Domain Voltage Sensing

The limitations of amplitude-based voltage sensing have motivated a transition toward frequency-domain and time-domain encoding of supply information. In a delay-chain-based sensor, the supply-dependent propagation time of a chain of logic stages converts V_{DD} fluctuations directly into timing deviations, governed by the same Alpha-Power law relationship that underlies process-sensing ring oscillators [15]: a supply droop lengthens propagation time while a supply spike shortens it, and the resulting timing deviation can be digitised without any reference voltage.

3.3.3.1 Delay-Chain Voltage Sensor

Chen et al. [4] demonstrated a fully on-chip PVT sensor suite in which voltage sensing is realised by an ultra-low-voltage voltage-to-time converter, denoted ULV²TC, that converts the monitored supply V_{in} into a digital code by measuring the propagation delay of a current-starved inverter and pass-gate chain. This design is selected as the voltage sensing topology for the present work and is therefore described in the greatest technical depth.

The sensor is built around a chain of current-starved delay stages, with the chain supplied either by a fixed voltage reference V_{REF} or by the monitored rail V_{in} , selected by an input multiplexer. The propagating clock edge is sampled by a bank of flip-flops tapped along the chain, producing a 5-bit digital output $V[4:0]$. A process-sensor-derived control bit is also applied to the chain to compensate the conversion for process-induced delay shifts. The principle of operation is that the propagation delay t_p of each current-starved stage is a monotonically decreasing function of the supply voltage applied to it; comparing the delay obtained with V_{in} against the delay obtained with the fixed V_{REF} therefore encodes the magnitude of V_{in} as the position of the logic transition within the flip-flop bank. The reported quantization step of the ULV²TC is approximately 50 mV.

The voltage sensor and its accompanying process sensor are both designed around the ZTC bias point of the MOS transistor, at which the mutual compensation of carrier-mobility degradation and threshold-voltage drift with temperature causes the delay of a ZTC-biased inverter stage to become approximately independent of temperature. Exploiting this property allows the voltage sensor's output code to be referenced primarily to V_{in} rather than to the ambient temperature, decoupling the two effects without requiring an explicit temperature correction term in the voltage-conversion path itself. Within the overall PVT suite, the principal role of the voltage sensor, together with the process sensor's output, is to compensate the temperature sensor's digital output for supply- and process-induced error, rather than to operate as a fully independent measurement decoupled from the other two sensors. This shared-compensation architecture is the primary reason the design is selected for integration in the present work, it allows a single low-overhead voltage-sensing block to simultaneously serve as a calibration input for the temperature channel.

Compared to a conventional voltage or current ADC-based approach, the ULV²TC offers several practical advantages. First, the sensor suite as a whole is explicitly positioned as requiring no bandgap reference, avoiding the high-headroom cascode amplifier stages that the authors note are difficult to implement at sub-1-V supplies; a dedicated voltage-reference block is still required to generate the sub-0.5-V V_{REF} supplied to the multiplexer, but this is a comparatively simple low-voltage reference rather than a bandgap circuit. Second, the converter itself is realised using current-starved inverters, flip-flops, and an XOR gate, standard-cell-compatible digital logic with no precision analog devices beyond that reference. Third, the converter operates over the full 0.3 to 1.0 V supply range of the overall PVT sensor while consuming no more than 3.7 μ W at 0.3 V and a sample rate of 10k samples per second.

The principal limitations are as follows. The 50 mV quantization step fixes the achievable voltage resolution for a given chain length, and finer resolution would require additional delay stages and flip-flop taps at increased area and power. The converter's accuracy depends on the process-compensation control bit supplied by the co-integrated process sensor; without this correction, process-induced delay shifts in the current-starved chain would directly degrade the mapping from delay to voltage code. Finally, because the voltage sensor's main system role is to correct the temperature sensor's output rather than to report an independently calibrated absolute voltage, its absolute accuracy as a standalone voltage reference is not separately characterised in the reported results.

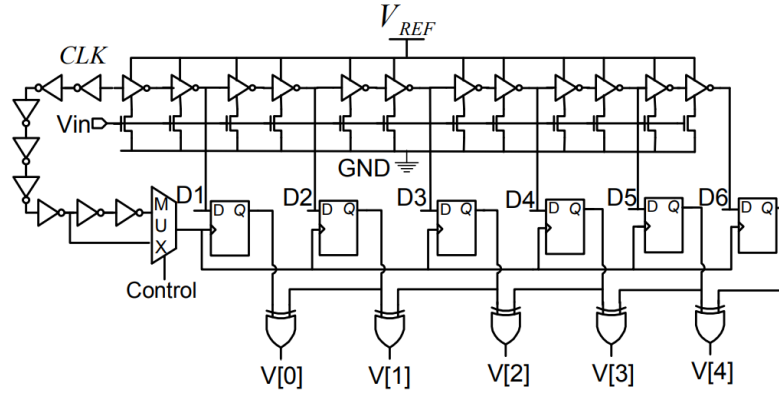


Figure 3.6: Proposed voltage sensor (ULV²TC) by Chen et al. [4], showing the current-starved delay chain, the V_{REF}/V_{in} multiplexer, and the flip-flop tap array producing $V[4:0]$.

3.3.3.2 Controllable Delay-Chain Voltage Sensor for Panel Driver SoCs

Lou et al. [10] presented a compact delay-chain voltage sensor as part of an on-chip CMOS corner detector targeting panel driver SoCs. The implementation uses a controllable delay block whose delay is tuned by a differential control voltage to track the supply level. This delay block drives a reference clock through a chain of six buffer and delay-stage pairs, each sampled by a dedicated D flip-flop, with exclusive-OR gates at the output producing the encoded supply metric $V[0]$ through $V[4]$. The XOR post-processing converts the raw thermometer code into a binary-weighted output, reducing the logic overhead of the subsequent decoder. Compared to the open-loop delay chain of Chen et al. [4], the controllable delay block adds a closed-loop tuning mechanism that improves linearity across the operating voltage range at the cost of additional control circuitry, and its co-integration alongside a matched NMOS- and PMOS-sensitive process monitor on the same production die (notably demonstrated within the Haswell processor environment described later in this chapter) is a direct architectural precedent for the shared master-satellite calibration strategy adopted in this dissertation.

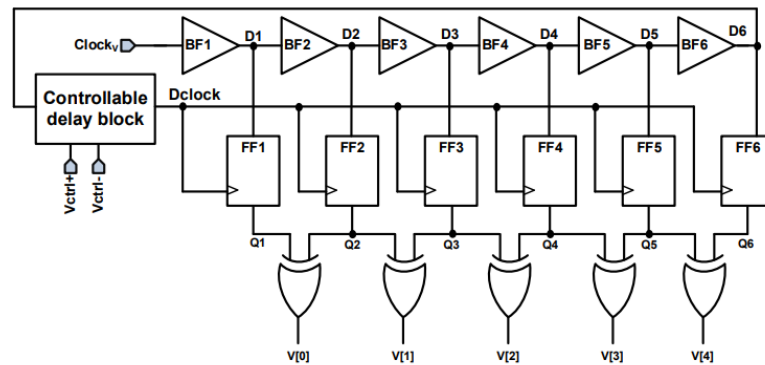


Figure 3.7: Controllable delay-chain voltage sensor from the on-chip CMOS corner detector of Lou et al. [10].

3.3.3.3 Dual Voltage and Temperature Resistive Sensor in 16-nm FinFET

Horng et al. [13] demonstrated a 0.7-V resistive sensor in 16-nm FinFET technology capable of detecting both temperature and voltage deviations through a shared charge and discharge circuit (CDC). The sensor provides two selectable sensing paths, a voltage sensing path and a temperature sensing path, each driving a resistor-based voltage divider whose taps are compared against a common threshold by two comparators connected to an SR flip-flop. The SR flip-flop controls the CDC, which generates a continuous charge and discharge oscillation whose frequency is set by the effective resistance and capacitance of the selected path,

$$f_{osc} = \frac{1}{2 \cdot R_{eff}(V_{DD}, T) \cdot C(T)}. \quad (3.24)$$

In voltage sensing mode, R_{eff} is modulated by the supply level through the MOSFET operating point, so that V_{DD} can in principle be recovered by inverting the frequency relationship, though this inversion is not analytically closed-form and in practice requires a pre-characterised look-up table or polynomial fit.

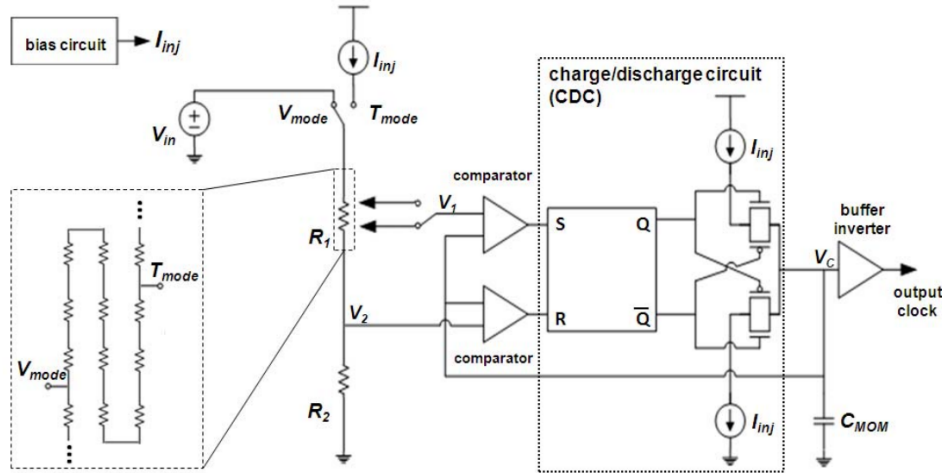


Figure 3.8: Dual voltage/temperature resistive sensor in 16-nm FinFET by Horng et al. [13].

A fundamental accuracy limitation arises from the fact that both R_{eff} and C are themselves temperature-dependent. Poly-silicon and diffusion resistors exhibit a positive temperature coefficient, so R_{eff} increases with die temperature, and MOS capacitances vary with both temperature and bias. The combined RC time constant therefore varies as

$$\tau(V_{DD}, T) = R_{eff}(V_{DD}, T) \cdot C(T), \quad (3.25)$$

meaning that a shift in f_{osc} cannot be unambiguously attributed to a change in V_{DD} without precise knowledge of the simultaneous die temperature. Since the shared CDC architecture permits only one sensing mode to be active at a time, voltage and temperature cannot be measured concurrently; the design must operate in a time-multiplexed fashion and assumes quasi-static variation between

successive measurement epochs, so any rapid thermal transient occurring during the voltage measurement window will corrupt the f_{osc} reading and produce an erroneous V_{DD} estimate.

Beyond this cross-sensitivity issue, the resistive sensing principle yields limited voltage sensitivity. The incremental change in oscillation frequency per unit change in supply voltage is

$$\frac{\partial f_{osc}}{\partial V_{DD}} = -\frac{1}{2C(T)} \cdot \frac{1}{R_{eff}^2} \cdot \frac{\partial R_{eff}}{\partial V_{DD}}, \quad (3.26)$$

and because $\partial R_{eff}/\partial V_{DD}$ is governed by the MOSFET channel resistance in the linear region, which is a relatively weak function of V_{DD} for devices biased well above threshold, the resulting frequency shift per millivolt of supply variation is small. This low supply-to-frequency gain necessitates either a high-resolution frequency counter or a wide integration window to resolve small supply deviations, both of which penalise detection latency. In temperature sensing mode, R_{eff} reflects the strongly temperature-dependent poly-silicon resistance, giving much higher sensitivity than in voltage sensing mode, revealing that the architecture is inherently better suited to temperature monitoring than to supply voltage tracking. The 0.7-V supply compatibility is a practical advantage for near-threshold SoC domains, and the 16-nm FinFET implementation confirms that resistive sensing scales to advanced nodes provided the resistance model is recalibrated for fin geometry; however, the thermal cross-sensitivity and limited supply sensitivity represent significant drawbacks when accurate, high-bandwidth voltage monitoring is the primary requirement, and these limitations are precisely what motivate the selection of the digital delay-chain approach of Chen et al. [4] over a resistive architecture for the present work.

3.4 Temperature Sensors

As power densities in sub-28-nm nodes continue to rise, thermal management has shifted from global die-level monitoring to localised hot-spot detection. Modern SoCs require dozens, if not hundreds, of thermal sensors distributed across the silicon surface to prevent localised thermal runaway and to enable the optimization of Dynamic Voltage and Frequency Scaling at the granularity of individual cores [10]. This section evaluates the evolution of thermal sensors from classic resistor- and BJT-based designs to modern, area-efficient frequency-locked loops, before providing an exhaustive treatment of the five architectures selected for deepest analysis and direct adoption in this dissertation: the N-well diode time-domain sensor of Eberlein et al. [43], the ring-oscillator thermal sensor of Lempel et al. [44], the frequency-locked loop sensor of Someya et al. [45], the ultra-compact 40-nm sensor of Lu et al. [46], and the composite-transistor PTAT current reference of Hanumantha Rao et al. [31].

3.4.1 Resistor-Based and BJT-Based Thermal Sensors

The classical approach to on-chip temperature sensing exploits the well-characterised temperature dependence of passive resistors or bipolar junction transistors. Park and Kim [47] demon-

strated a resistor-based temperature sensor in 65-nm CMOS achieving a supply sensitivity of only $0.28^\circ\text{C}/\text{V}$ at a 0.8-V supply, using a bridge transducer and 5-bit resistive DAC compared against a 5-bit interpolation DAC through a chopper-stabilised comparator, with SAR logic converging on a two-step coarse and fine code. The principal trade-off of this architecture is its reliance on an external linearization step that limits scalability for dense on-chip deployment, even though the analogue front end itself is supply-robust. Bass and Shor [48] demonstrated that the highest integration density for thermal monitoring in high-performance processors is achievable through PNP-based sensors, with substrate PNP transistors producing differential and absolute V_{BE} signals digitised by a sigma-delta modulator, achieving a sensing area of only 0.003 mm^2 in a 28-nm bulk CMOS process, making it well suited to dense CPU thermal arrays.

3.4.2 Ring-Oscillator-Based Thermal Sensor in 16-nm FinFET

Lempel et al. [44] addressed the specific challenge of building an ultra-compact thermal sensor for dense, multi-spot monitoring in 16-nm FinFET microprocessors, where a large number of sensors are required to track hotspots whose locations vary with workload and are difficult to predict at design time. The proposed sensor occupies only $700\text{ }\mu\text{m}^2$, reported by the authors as the smallest thermal sensor in its comparison class, while achieving a peak-to-peak inaccuracy of less than $\pm 1^\circ\text{C}$ near the throttle point and less than $\pm 1.5^\circ\text{C}$ across a 130°C range, with a conversion time of $6.9\text{ }\mu\text{s}$ and an energy consumption of approximately 290 pJ per conversion.

The sensor is built around a single current-controlled oscillator (CCO), implemented as a 15-stage CMOS inverter-based ring oscillator, whose oscillation amplitude V_{ccRO} is charged by a PMOS current source rather than connected directly to V_{DD} . The large stage count was deliberately chosen to keep the oscillation frequency low enough for the downstream level shifter and counter to operate correctly across all process corners, and to reduce the ripple on V_{ccRO} . The resulting frequency-to-current-to-voltage relationship is

$$F = \frac{I}{C_{ro}V}, \quad (3.27)$$

where F is the oscillation frequency, I is the bias current, C_{ro} is the total capacitance in the oscillator loop, and $V = V_{ccRO}$ is the oscillation amplitude set by that same current. This is a current-mode formulation rather than a conventional inverter-delay-chain expression, the PTAT current itself, not a temperature-dependent propagation delay, is the primary mechanism that encodes thermal information into F .

The current source is biased by an internal PTAT current reference operated in the sub-threshold regime, such that

$$I = \frac{\delta V_{gs}}{R} = \left(\frac{kT}{qR} \right) \ln(N), \quad (3.28)$$

with R a trimmable bias resistor and N the sizing ratio between the two reference transistors.

Combining the two preceding equations gives the overall sensor transfer function

$$F = \frac{kT \ln(N)}{qRC_{ro}V_{ccRO}}. \quad (3.29)$$

A key implementation detail enabling this compact PTAT source is the use of a Mixed- V_{th} FinFET stack for the current source and mirror devices. By stacking transistors of different threshold voltages, a high- V_{th} majority with a low- V_{th} device placed closest to the drain, the stack behaves as a self-biased cascode, raising the effective output resistance by approximately $8\times$ relative to a single device without requiring any additional cascode bias voltage or extra supply headroom. This directly improves both the accuracy of the current mirroring and the supply-voltage rejection of the sensor.

Because V_{ccRO} sits close to the threshold voltage to keep the CCO operating at very low power, it cannot be fed directly into the digital level shifter without corrupting the oscillator's effective load, so a unity-gain analogue buffer reproduces V_{ccRO} as a separate node before the signal is converted to a rail-to-rail digital output and passed to a counter. The dominant contribution to the temperature dependence of the closed-form expression comes from the PTAT current term in the numerator, which alone accounts for a 34% increase in frequency over a 120°C sweep, partially counteracted by V_{ccRO} itself, which drops by roughly 8% over the same range and causes F to increase by a further 8% since it appears in the denominator. The bias resistor contributes a smaller, opposing 1% increase with temperature. The authors identify V_{ccRO} , through its dependence on $V_{th}(T)$, as the principal source of the sensor's transfer-curve nonlinearity, since a temperature-dependent term in the denominator of F is intrinsically nonlinear; this curvature is repeatable across units and is corrected digitally in the back end via systematic curvature correction rather than additional analogue compensation.

In silicon, the sensor was integrated at two locations on a 16-nm FinFET SoC and measured across 16 instances. With simple two-point calibration the measured peak-to-peak inaccuracy was within $\pm 1.5^\circ\text{C}$ over the full range, falling to less than $\pm 1^\circ\text{C}$ at the hot calibration point after systematic curvature correction. The measured resolution was 0.32°C rms for a single $6.9\ \mu\text{s}$ conversion, improving further with moving-average filtering at the cost of conversion time. The sensor's overall power-supply sensitivity was measured at $6^\circ\text{C}/\text{V}$, attributed directly to the high output resistance furnished by the Mvt self-biased cascode current sources. The nominal simulated power consumption of the complete sensor is $41.8\ \mu\text{W}$, of which $36\ \mu\text{W}$ is attributed to the analogue circuitry and $5.8\ \mu\text{W}$ to the externally located counter logic.

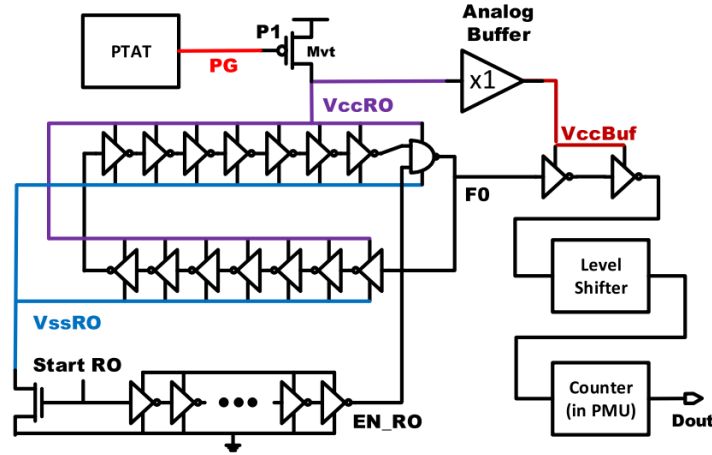


Figure 3.9: RO-based thermal sensor in 16-nm FinFET by Lempel et al. [44].

3.4.3 Composite-Transistor PTAT Current Reference for Thermal Sensing

Hanumantha Rao et al. [31] examined the fundamental challenge of generating stable PTAT current references in low-voltage CMOS processes, where the conventional beta-multiplier suffers from degraded supply rejection as V_{DD} approaches the sum of device threshold voltages. The proposed solution employs a composite transistor technique that modifies the effective threshold voltage of the current mirror devices to achieve improved supply voltage sensitivity at a 0.8-V supply, delivering a quiescent current of approximately 5 nA.

The canonical beta-multiplier current reference generates a PTAT current by exploiting the ΔV_{GS} mismatch between two identically biased NMOS branches operating in the subthreshold or near-threshold regime [1]. For devices in the subthreshold region, the drain current follows

$$I_D = I_0 \frac{W}{L} \exp\left(\frac{V_{GS} - V_{th}}{nV_T}\right) \left(1 - \exp\left(\frac{-V_{DS}}{V_T}\right)\right), \quad (3.30)$$

where $V_T = kT/q$ is the thermal voltage, n is the subthreshold slope factor, and I_0 is a process-dependent pre-exponential current. In the standard beta-multiplier topology with a mirror ratio of K and a resistor R in series with one branch, enforcing equal currents yields

$$I_{PTAT} = \frac{V_T \ln K}{R} = \frac{kT \ln K}{qR}, \quad (3.31)$$

whose proportionality to the thermal voltage is the source of the PTAT characteristic, with temperature coefficient

$$\frac{dI_{PTAT}}{dT} = \frac{k \ln K}{qR}, \quad (3.32)$$

positive and, to a first approximation, constant over temperature. The supply sensitivity, however, is governed by the output impedance of the mirror, with $\partial I_{PTAT} / \partial V_{DD}$ becoming significant as V_{DD} approaches the sum of the threshold voltages plus saturation overdrive.

The proposed circuit replaces the standard NMOS output devices with composite pairs, a stacked diode-configured NMOS pair in the signal path and a corresponding mirror leg. The stacking of a diode-connected MOSFET in series with a primary transistor creates an effective device whose gate-to-source voltage is $V_{GS,eff} = V_{GS,upper} + V_{GS,lower}$. For two identical devices in series, each operating in the subthreshold regime, the composite device exhibits an apparent threshold voltage

$$V_{th,eff} \approx 2V_{th} - V_T \ln\left(\frac{I_D}{I_0(W/L)}\right), \quad (3.33)$$

less sensitive to V_{DD} than either individual device, because the cascode-like stacking increases the effective output impedance of the mirror without requiring a dedicated cascode bias. If the intrinsic output resistance of a single device is r_o , the effective output resistance of the composite pair is approximately $r_{o,eff} \approx 2r_o + g_{m,lower}r_{o,upper}r_{o,lower}$, giving a significant boost in supply rejection without any additional bias circuitry.

The composite transistor approach achieves improved supply rejection at the cost of a reduced voltage headroom, stacking two devices in the signal path consumes an additional $V_{DS,sat}$, approximately 100 to 150 mV in near-threshold operation, which directly reduces the minimum operable V_{DD} . At 0.8 V, the design operates close to the minimum feasible supply for this class of reference. The 5 nA output current is well suited to driving subsequent temperature-to-frequency converters without significant self-heating, making this reference an appropriate building block for ultra-low-power sensing.

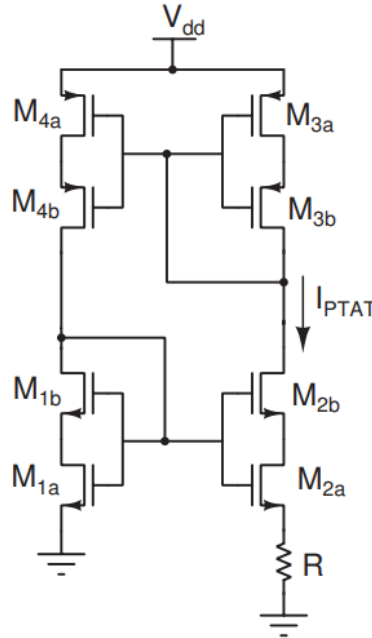


Figure 3.10: Composite-transistor PTAT current reference by Hanumantha Rao et al. [31].

This architecture is selected as the structurally necessary biasing companion to the ring-oscillator sensor of Lempel et al. [44] in the proposed infrastructure, since the PTAT-biased supply

voltage that modulates the VccRO oscillator must itself be generated by a current reference whose supply rejection is sufficiently high that supply perturbations do not reintroduce the common-mode coupling that the differential oscillator topology of Lempel et al. [44] is designed to cancel.

3.4.4 N-well Diode Time-Domain Sensors in FinFET Nodes

Eberlein et al. [43] pushed the scaling frontier of thermal sensing to the 7-nm FinFET node, realising a time-domain temperature sensor occupying only $820 \mu\text{m}^2$ using N-well diodes as the sensing element. In FinFET technologies, N-well diodes formed at the junction between a deep N-well and the P-substrate retain a predictable CTAT forward voltage characteristic despite the altered doping profiles associated with fin isolation,

$$V_D(T) = V_{g0} - \left(\frac{V_{g0} - V_{D0}}{T_0} \right) T - \eta V_T \ln \left(\frac{T}{T_0} \right), \quad (3.34)$$

where V_{g0} is the silicon bandgap voltage extrapolated to 0 K, V_{D0} is the diode forward voltage at reference temperature T_0 , and η accounts for the temperature-dependent ideality factor. This voltage is inherently CTAT with a typical slope of approximately $-1.5 \text{ mV}/^\circ\text{C}$.

The circuit uses two capacitors, C1 (1 pF) and C2 (4 pF), that are reset and then discharged by the sensing diode through a carefully timed three-phase sequence. In the reset phase, both capacitors are charged to V_{DD} . In the first discharge phase, C1 discharges through the CTAT diode while C2 remains held; the time t_1 required for the node voltage to cross a comparator threshold is

$$t_1 \approx \frac{C_1 \Delta V}{I_D(T)}, \quad (3.35)$$

where ΔV is the voltage excursion and $I_D(T)$ is the temperature-dependent diode current. The second discharge phase discharges a scaled reference capacitor over a reference interval. A comparator with an auto-zero capacitor resolves the voltage difference between the two nodes to determine the crossing time, and counter logic digitises the ratio t_1/t_r into a temperature code. Because the ratio is dimensionless and involves the same diode current in numerator and denominator, a large fraction of process-corner variation cancels, enabling the reported no-trim inaccuracy of $\pm 3.5^\circ\text{C}$ across the full process corner space without post-fabrication calibration, a significant advantage in high-volume manufacturing where individual sensor trimming represents a non-trivial test cost.

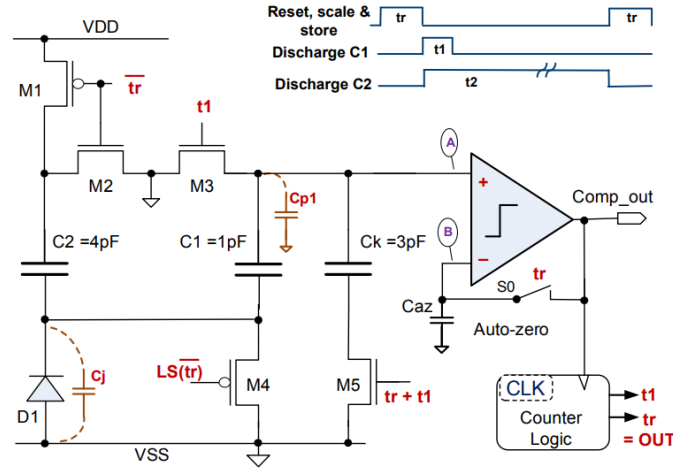


Figure 3.11: Time-domain temperature sensor in 7-nm FinFET by Eberlein et al. [43].

Eberlein and Pretl [49] complemented this result in 16-nm FinFET, employing the same bulk diode principle but extending the timing sequence to a three-phase SAR-assisted approach with an 8-bit SAR-controlled capacitive DAC replacing the fixed scaling capacitor to enable fine post-silicon trimming of the discharge ratio without analogue resistor or current trimming.

3.4.5 Ultra-Compact CMOS-Only Temperature Sensor in 40-nm

Lu et al. [46] developed a highly compact on-chip temperature sensor with ultra-low power consumption in a 40-nm CMOS process. The architecture centres on a sub-threshold PTAT current reference, implemented as a diode-connected NMOS transistor in series with a resistor, and regulated by an op-amp feedback loop against a fixed bias. A digital controller manages three measurement phases: in a first phase, the PTAT current charges a capacitor through a dedicated switch to establish a temperature-proportional voltage; in a second phase, this voltage is compared against the continuously generated PTAT reference current through a comparator, and the time required to cross the threshold, counted by the digital controller, encodes the temperature as a pulse width. The architecture is notable for its elimination of a dedicated VCO or counter clock, reducing area and quiescent power at the cost of a lower conversion rate relative to the frequency-domain designs surveyed elsewhere in this chapter, making it directly relevant as a low-area benchmark against the ring-oscillator-based approach of Lempel et al. [44] adopted in this dissertation.

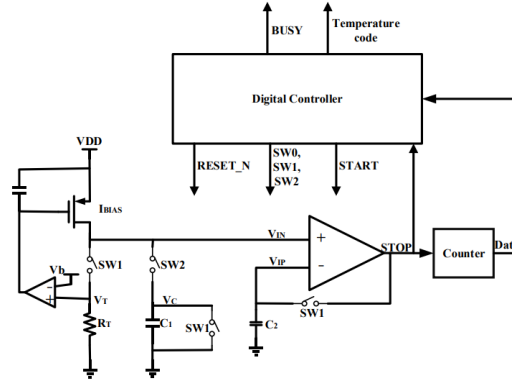


Figure 3.12: Ultra-low-power 40-nm temperature sensor by Lu et al. [46].

3.4.6 Frequency-Locked Loop Architectures

The transition from open-loop current-to-frequency conversion to closed-loop Frequency-Locked Loop (FLL) architectures represents the most significant recent advancement in temperature sensor design, providing simultaneous improvements in supply rejection, linearity, and noise immunity.

Someya et al. [45] introduced an FLL-based temperature sensor consuming only 6.4 nW with a relative inaccuracy of 1.7%. The architecture is distinguished by its use of the sub-thermal drain voltage (STV) regime, a biasing condition in which $V_{DS} < V_T = kT/q$, to achieve a temperature sensing mechanism that is inherently insensitive to supply fluctuations. The sensing core consists of two NMOS transistors both operating in the sub-threshold region, one biased in the STV regime generating a drain current I_1 , and the other biased in the above-thermal voltage (ATV) regime ($V_{DS} > 3V_T$) generating a reference current I_2 . When $V_{DS1} \ll V_T$, the current ratio approximates

$$\frac{I_2}{I_1} \approx \frac{U_T}{V_{DS1}} + C, \quad (3.36)$$

which is linear in U_T/V_{DS1} and therefore PTAT, where C is a temperature-independent offset removable by one-point calibration. The key property of the STV biasing is that the drain-source voltage of the STV-biased device remains well below the thermal voltage, suppressing drain-induced barrier lowering and rendering the drain current insensitive to fluctuations in V_{DS1} . Because V_{DS1} is set internally by the FLL rather than by an external supply rail, common-mode supply variations do not perturb I_1 , providing exceptional supply rejection without any power-hungry regulator.

The ratio $D_{PTAT} \propto f_2/f_1 = I_2/I_1$ is formed by two current-to-frequency converters whose output clocks are counted by two independent counters, and a state machine latches the ratio into the output word D_{PTAT} . The FLL ensures that V_{DS1} is continuously servoed to remain within the STV regime across temperature and process corners, the closed-loop stabilization function that distinguishes this topology from open-loop designs. At 6.4 nW, this sensor is among the lowest-power fully integrated thermal sensors reported and is well suited to near-threshold or sub-threshold SoC contexts, though its 0.32 mm² area footprint is substantially larger than the

compact ring-oscillator and N-well diode architectures surveyed above, illustrating directly the power-versus-area trade-off that governs the selection of a thermal sensing topology for a dense, replicated satellite network.

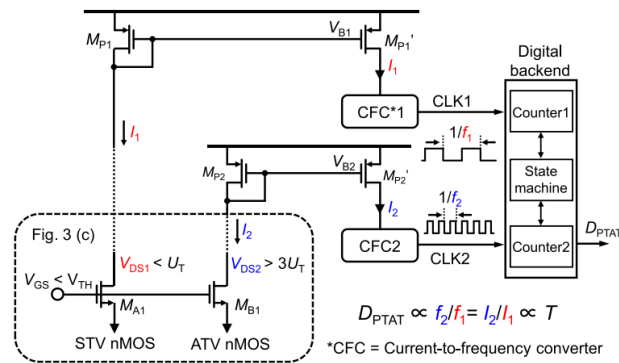


Figure 3.13: FLL-based temperature sensor by Someya et al. [45].

3.4.7 Additional VCO-Based, Frequency-Domain, and Industrial-Grade Designs

Beyond the five core architectures treated above, several further designs illustrate complementary points in the temperature sensor design space. Shui and Wang [50] achieves the current state of the art in area efficiency for sub-threshold CMOS sensors, reaching $173 \mu\text{m}^2$ with a Figure of Merit of $14.17 \text{ pJ}\cdot\text{K}^2$ through a compact 2T/4T sub-threshold structure, though its remote sensing path introduces additional accuracy degradation. Anand et al. [51] and Moisello et al. [52] pursue differential and self-regulating VCO architectures achieving supply sensitivities as low as $0.034^\circ\text{C}/\text{mV}$ by cancelling common-mode supply noise in a frequency ratio, while Kim et al. [53] demonstrates a 366 kS/s multiphase frequency-to-digital converter for capturing fast thermal transients at the cost of TSO/TIO mismatch-induced absolute precision loss. Lee and Hong [54] co-integrates a VCO directly with a bandgap reference source to make oscillation frequency proportional to a PTAT-to-CTAT current ratio rather than to the raw supply voltage. Karpuz et al. [55] demonstrates an ultra-compact single PTAT-CTAT current loop using a Wilson current mirror to minimise transistor count. For industrial and automotive contexts where guaranteed accuracy at temperature extremes outweighs area, Rajeswari et al. [56] validates a fully integrated 90-nm CMOS sensor across -40°C to 125°C using a self-biased beta-multiplier PTAT generator, Eberlein and Pretl [57] stabilises oscillator frequency in wireless sensor nodes via a PTAT-biased varactor oscillator, and Zhao et al. [58] explores a minimal four-transistor linear V_{th} -based transducer with $\text{sub-}0.5^\circ\text{C}/\text{V}$ supply sensitivity. Additionally, the Haswell processor [10] integrates a matched NMOS- and PMOS-sensitive process sensor alongside its voltage monitor, providing a production-proven precedent for the co-integration strategy of distinct sensing domains adopted in this dissertation.

3.5 Comparative Analysis and Design Choices

The preceding sections have examined the state of the art across four tightly coupled design domains: voltage references, process sensors, voltage sensors, and temperature sensors. Taken individually, each domain has produced implementations of remarkable precision or compactness, yet a consistent structural tension emerges when these results are viewed as a collective system. Designs that achieve the highest measurement accuracy invariably do so by consuming disproportionate area and power, rendering them incompatible with the density requirements of a distributed on-chip monitoring infrastructure. Conversely, designs that minimise footprint and quiescent current tend to sacrifice absolute accuracy or supply robustness in ways that become critical when sensors must operate continuously in the presence of power delivery network noise and across a wide range of process corners. The following tables compress the comparison to the six (process, voltage, temperature) or nine (bandgap reference) most relevant topologies for the 16-nm FinFET environment targeted by this work, anchored by the mandatory architectures identified throughout this chapter.

Table 3.1: Comparison of Process Characterization Architectures (16-nm FinFET Relevant Subset).

Reference	Approach	Key Strength	Limitation
Fujimoto [33]	Inhomogeneous RO	Captures within-die stochastic variability via six structurally distinct cell types	Complex post-silicon analysis and proximity-induced systematic variability
Chandan [32]	Ratioed NAND/NOR	4.086/7.223 N/P sensitivity, double that of [14], with two-dimensional corner separation	21-stage rings increase area; constrained to 45-nm validation
An [14]	Ratioed inverter RO	NMOS/PMOS corner decomposition; fully digital, no ADC required	Supply noise sensitivity; lower sensitivity than ratioed NAND/NOR
Kalarikkal [35]	Opposite-device RO	Compatible with 28-nm FD-SOI	Requires body-bias calibration; only 2% frequency spread per corner
Islam [36]	Wide-supply leakage	Ratiometric operation from 0.5 to 1.2 V without recalibration	Long integration window; temperature cross-sensitivity
Kishimoto [37]	Reconfigurable RO	Dual process and temperature sensing in one block	Mode-switching control overhead

Among process characterization architectures, ring-oscillator-based monitors occupy the dominant position in the literature owing to their structural simplicity and natural correlation with dig-

ital path delay. Supply coupling is the primary weakness shared across the RO family, partially addressed by current-steering decoupling at the cost of area [34]. The simplest current-starved configurations, such as that of [5], provide a direct frequency readout of the local process corner but treat NMOS and PMOS contributions as a single indistinguishable aggregate. Sub-threshold leakage sensors such as [11] exploit the exponential V_{th} dependence of off-state current to achieve far greater sensitivity to process shifts than any saturation-region oscillator, though this amplification is accompanied by equally strong sensitivity to temperature.

Table 3.2: Comparison of Voltage Reference and Bandgap Architectures (Anchored on the FinFET-Native Current-Mode Topologies).

Reference	Approach	Key Strength	Limitation
Low [28]	FinFET PMOS-sensing current-mode BGR	N-well CTAT; sub-1-V; natively compatible with 14-nm FinFET	First-order undercompensated; TC approximately 222 ppm/°C; LS approximately 0.9 mV/mV
Moisello [29]	FinFET NMOS-sensing current-mode BGR	Lower NMOS V_{th} mismatch coefficient; relaxed op-amp common-mode	Higher flicker noise than PMOS-sensing variant
Jiang [21]	ZTC-point BGR	5.6 ppm/°C; 87-dB PSRR; no active curvature amplifier	Process-dependent ZTC bias point
Lohitsiri [6]	$T \ln T$ BGR	2.39 ppm/°C; least-squares resistor optimization	Additional amplifier branch and area overhead
Khola [22]	Multi-output BGR	1.12 ppm/°C; five output rails from a single core	Output cross-loading through shared divider
Huang [18]	Amplifier-free all-in-one BGR	192 nW; simultaneous voltage and current output	Moderate TC; minimum supply 2 V
Cao [27]	0.8-V automatic-feedback BGR	3σ accuracy 1.58%; no external trimming required	Feedback loop stability at process corners
Sahu [8]	MOS-only BGR	No bipolar junction device; minimum silicon area	TC approximately 287 ppm/°C
Li [7]	LS-enhanced BGR	Pre-regulated supply decouples core from PDN noise	Cascode stage adds headroom requirement

The bandgap reference literature is characterised by a clear hierarchy of precision, with first-order designs achieving 20 to 50 ppm/°C against the single-digit ppm/°C achievable through high-order curvature compensation at the cost of additional amplifier or resistor-network complexity. Sampled and offset-corrected variants, such as [26], [20], [23], [24], and [25], alongside the sub-ranging current reference of [30], occupy intermediate positions on the area-precision-power trade-off curve and inform the design intuition behind the architectural choices made here, even though

they are not natively compatible with the FinFET, sub-1-V target environment. FinFET-native designs such as [28] navigate the absence of a native substrate PNP by substituting the BJT with a forward-biased N-well diode, recovering a usable CTAT characteristic without requiring process modifications, while the structurally analogous NMOS-sensing variant of [29] offers a complementary implementation that may yield tighter threshold voltage mismatch for a given device area.

Table 3.3: Comparison of Voltage Sensor Architectures (16-nm FinFET Relevant Subset).

Reference	Approach	Key Strength	Limitation
Chen [4]	Delay-chain thermometer-code sensor	Fully digital; absolute voltage readout; concurrent PVT sensing; no bandgap required	Temperature cross-sensitivity without process-bit correction
Lou [10]	Controllable-delay chain	CPU-grade integration; compact layout with XOR post-processing	Designed for specific in-chip deployment only
Hornig [13]	Resistive dual voltage/temperature sensor	Single block monitors both voltage and temperature in 16-nm FinFET	Time-multiplexed only; low $\partial f_{osc}/\partial V_{DD}$; thermal cross-sensitivity
Aparicio [41]	Ratiometric process-invariant monitor	1052 μm^2 ; calibration-free	Moderate detection bandwidth
Aparicio [39]	Reference-free delay monitor	2.64 pJ/conversion; wide temperature range	Process-dependent absolute accuracy
Vijayakumar [12]	Event-counting noise detector	Detects GHz-range supply transients	No continuous DC level sensing

The voltage sensing domain is characterised by a fundamental recursive constraint: accurate supply monitoring requires a reference that is itself immune to the supply fluctuations being measured. The classification framework established by [38] organises existing designs into reference-based, reference-free, and event-driven categories, with reference-free designs such as [40] and [42] occupying the low-area end of the spectrum at the cost of comparator-offset-limited or single-threshold accuracy. Among reference-free designs, [39] and [41] represent different points along the accuracy-versus-area trade-off, while event-driven monitors such as [12] fill a complementary role, detecting gigahertz-range transients that would be missed by any low-speed converter, but cannot supply the DC supply level information needed for guard-band elimination.

Table 3.4: Comparison of Temperature Sensor Architectures (16-nm FinFET Relevant Subset).

Reference	Approach	Key Strength	Limitation
Lempel [44]	Single CCO in 16-nm FinFET	$\pm 1.5^\circ\text{C}$ accuracy; $700\ \mu\text{m}^2$; 290 pJ/conversion; Mvt cas-code supply rejection	Nonlinearity corrected digitally in back end
Eberlein [43]	N-well diode charge-pump in 7-nm	$820\ \mu\text{m}^2$; scaling frontier; no post-silicon trim	$\pm 3.5^\circ\text{C}$ accuracy without calibration
Someya [45]	FLL with STV stabilization	6.4 nW; supply-independent; 1.7% relative inaccuracy	$0.32\ \text{mm}^2$; unsuitable for dense arrays
Lu [46]	Sub-threshold pulse-width 40-nm sensor	No dedicated VCO or counter clock; ultra-low quiescent power	Lower conversion rate
Hanumantha Rao [31]	Composite-transistor PTAT reference	Improved supply rejection without dedicated LDO regulator	Moderate absolute accuracy at 0.8-V supply
Shui [50]	2T/4T sub-threshold CCO	$173\ \mu\text{m}^2$; FoM of $14.17\ \text{pJ}\cdot\text{K}^2$	Remote sensing path degrades accuracy

Among the temperature sensing architectures reviewed, this domain exhibits the most diverse range of design methodologies. On the high-precision end of the spectrum, resistor-based sensors such as [47] and PNP-based topologies such as [48] provide exceptionally linear thermal characteristics and supply sensitivities below $0.3^\circ\text{C}/\text{V}$, but require substantial passive networks or dedicated bipolar devices that are difficult to integrate cleanly into scaled FinFET technologies. Alternative VCO-based and industrial-grade alternatives, including [51], [52], [53], [54], [55], [56], [57], [49], and [58], occupy intermediate points on the area, power, and supply-sensitivity trade-off space but are not carried forward into the final six-row comparison since their reported accuracy, area, or node compatibility is dominated by the topologies retained above for the 16-nm FinFET target environment.

Examining the four sensing domains collectively, a common structural bottleneck emerges. The designs that exhibit the lowest measurement uncertainty within each domain invariably depend on a stable, low-noise reference to function correctly, whether that reference is an explicit bandgap voltage, a stable bias current, or a thermally isolated calibration signal. The designs that attempt to operate without such a reference, either through ratiometric cancellation or reference-free delay encoding, achieve independence at the cost of absolute accuracy that becomes difficult to recover without per-chip calibration. This creates a circular dependency at the system level: each sensing domain requires environmental stability from the others in order to deliver reliable measurements of its own quantity. The fundamental implication is that the primary bottleneck in modern on-chip metrology is not the absence of sufficiently sensitive transducers. Adequate

sensitivity exists in every domain reviewed here. The bottleneck is rather the decoupling and synchronization of distributed monitors operating within the same noisy power delivery network, and it is this structural gap that the unified monitoring infrastructure developed in subsequent chapters of this dissertation is designed to address.

Against this analytical backdrop, the following paragraphs justify each of the specific topological selections made for simulation and implementation in this work, grounding each choice in the technical constraints imposed by the target 16-nm FinFET technology and the architectural requirements of the proposed Master-Satellite monitoring framework.

The delay-chain thermometer-code architecture of [4] is selected for the voltage sensing path. The primary objective of this work's voltage sensing block is to obtain a direct and accurate measurement of the absolute supply voltage, rather than a differential reading of voltage deviations relative to some previous state. The majority of alternative topologies reviewed in this chapter, including the event-driven detector of [12] and the ratiometric monitor of [41], are fundamentally oriented toward detecting supply excursions or encoding relative changes, and as a consequence they introduce systematic inaccuracies when the objective is continuous absolute voltage quantification. The architecture of [4] resolves this by directly converting the instantaneous supply level into a thermometer code whose transition position encodes the absolute voltage. Furthermore, this structure is fully digital by construction, requires no analogue-to-digital converter interface, and has been demonstrated in a concurrent multi-domain PVT sensing context where temperature and process correction coefficients are applied jointly to the raw voltage reading. Its principal limitation, cross-sensitivity to temperature through the carrier mobility dependence of inverter propagation delay, is manageable within the proposed unified infrastructure because an independently stabilised thermal reference is available from the co-integrated temperature sensor to supply the necessary correction. No other topology reviewed in this section combines absolute voltage encoding, fully digital output, and documented co-integration with process and temperature correction within a single coherent framework, which makes [4] the uniquely appropriate choice for this application.

The FinFET-native current-mode bandgap reference topologies of [28] and [29] are selected as the voltage reference components of the monitoring infrastructure. Both choices are driven by two hard constraints imposed by the 16-nm FinFET target process. First, sub-1-V operation is a fundamental requirement of the design, as the nominal supply voltage of advanced FinFET nodes falls below the headroom needed to sustain a classical bandgap reference employing a forward-biased substrate PNP device in series with an amplifier output stage. Second, the absence of a native substrate PNP transistor in planar-incompatible FinFET processes precludes the direct application of conventional bipolar-based PTAT-CTAT summation architectures, including the high-order designs of [6] and [21] that would otherwise be attractive for their temperature coefficient performance. The architecture of [28] resolves both constraints simultaneously by replacing the bipolar junction with a diode-connected PMOS FinFET pair biased in the subthreshold regime, generating the CTAT component from the native gate-to-source voltage without requiring any device not present in a standard digital FinFET process. The NMOS-sensing variant of [29] is included

as a complementary implementation of the same current-mode principle because the substitution of NMOS sensing devices offers potential advantages under specific process conditions, including smaller threshold voltage mismatch coefficients and a relaxed op-amp common-mode constraint. By evaluating both the PMOS and NMOS sensing variants, the design investigation in the following chapters can rigorously characterise the relative trade-offs between threshold voltage mismatch, flicker noise, supply headroom, and resistor network optimization requirements under 16-nm process conditions, and identify which configuration provides the most stable and manufacturable reference baseline for the proposed infrastructure.

The ratioed NAND/NOR process sensor of [32] is selected as the primary process sensing architecture. Among the ring-oscillator-based process monitors reviewed in this chapter, the ratioed NAND/NOR topology achieves the strongest demonstrated decomposition of NMOS-dominant and PMOS-dominant process shifts, more than doubling the sensitivity of the prior ratioed-inverter design of [14] while consuming approximately 12% less power, a combination that directly addresses the requirement of the proposed infrastructure to supply an accurate, low-overhead process-compensation bit to both the voltage sensor of [4] and the temperature sensor described below. The inhomogeneous ring oscillator of [33] remains a valuable reference point for characterising the within-die stochastic component of process variation during corner validation, even though its structural and post-silicon analysis complexity precludes its direct use as the lightweight, replicated satellite-node process sensor required here.

The pulsed dual-ring-oscillator thermal sensor of [44] is selected as the primary temperature sensing architecture, and the composite-transistor PTAT current reference of [31] is identified as the necessary complementary biasing block. The selection of [44] is governed by a hard architectural constraint: the proposed Master-Satellite framework requires that the temperature measurement be extracted from a frequency output rather than from an analogue voltage or a sigma-delta bitstream, because frequency-domain outputs can be directly counted and processed by the digital finite-state machine that governs the master node without any dedicated analogue interface circuitry. Among the temperature sensing architectures reviewed, only ring-oscillator-based designs produce the required frequency output natively. Within this subset, [44] is the most technically appropriate candidate for the 16-nm target node for several interconnected reasons. Its pulsed measurement strategy mitigates the self-heating problem that is endemic to ring-oscillator-based sensors in FinFET nodes, where the elevated thermal resistance of the fin dielectric stack means that continuous oscillator operation would raise the local junction temperature by a measurable amount and corrupt the reading. The differential frequency architecture, in which a PTAT-biased supply-rail ring and a ground-rail ring are counted against each other, suppresses common-mode supply-induced frequency shifts to first order, delivering the supply robustness that open-loop single-oscillator designs such as [46] and [50] cannot achieve. The reported no-trim inaccuracy of $\pm 2^\circ\text{C}$ across 16-nm process corners establishes a concrete accuracy baseline consistent with the requirements of distributed thermal monitoring for dynamic voltage and frequency scaling. The composite-transistor PTAT reference of [31] is a structurally necessary companion to [44] because the PTAT-biased supply voltage that modulates the VccRO oscillator in the [44] architecture must

itself be generated by a current reference whose supply rejection is sufficiently high that supply perturbations do not reintroduce the common-mode coupling that the differential oscillator topology was designed to cancel. Conventional beta-multiplier PTAT references lose supply rejection at sub-1-V supplies as the output impedance of the mirror devices degrades, while the composite-transistor stacking of [31] recovers this rejection through an effective output impedance boost without consuming additional voltage headroom. The combination of [31] as the bias generator and [44] as the frequency transducer therefore yields a temperature sensing path whose dominant error sources are independently identifiable, structurally decoupled, and addressable through the shared calibration resources of the master node. The alternative N-well diode time-domain approach of [43] and the sub-threshold pulse-width design of [46] remain valuable benchmarks for trim-free accuracy and ultra-low-area, ultra-low-power operation respectively, but neither natively produces the frequency-domain output required by the proposed digital master controller without an additional time-to-frequency conversion stage.

By grounding the process sensor in the NAND and NOR ratioed-gate architecture of [32], anchoring the voltage reference path in the FinFET-native current-mode topologies of [28] and [29], selecting the delay-chain thermometer-code approach of [4] for absolute voltage sensing, and pairing the PTAT-stabilised dual-ring thermal sensor of [44] with the supply-robust current reference of [31], the proposed infrastructure assembles a set of sensing blocks whose individual constraints and trade-offs are mutually complementary rather than conflicting. Each selected topology contributes a capability that the others cannot provide, and each limitation is compensated either by a co-integrated sensing block or by the centralised reference and calibration resources of the master node. The structural integration, simulation modelling, and corner validation of these topologies are comprehensively presented in Chapter 4.

Chapter 4

Implementation and Simulation Results

The preceding chapters established both the theoretical foundations of PVT sensing and a comprehensive survey of the state-of-the-art. From that analysis, a clear design challenge emerged: individually optimized sensors suffer from environmental cross-sensitivity and process-induced offsets that can only be resolved through a coordinated, hierarchically stabilized sensing infrastructure. This chapter presents the transistor-level design and simulation results of three sensing subsystems that collectively form the proposed unified PVT monitoring platform, implemented in a 16 nm FinFET technology node using Synopsys PrimeSim/HSPICE. Unless otherwise stated, all simulations throughout this chapter are performed at a nominal supply voltage of $V_{DD} = 0.8\text{ V}$ and a nominal temperature of 25°C , with temperature sweeps spanning the full industrial range of -40°C to 120°C .

A design constraint that recurs throughout this chapter is that the raw oscillation frequency of each sensing front-end should ideally remain below 1 GHz, since the frequency-domain output of each sensor is ultimately captured by a digital counter clocked within the FSM’s reference clock domain. Frequencies approaching or exceeding this threshold push the counter’s synchronization registers closer to their metastability margin and typically require an additional frequency-division stage purely to bring the signal into a countable range, adding jitter, power, and area. Keeping the raw frequency below 1 GHz allows the readout logic to sample the oscillator output directly, a property that distinguishes the two process sensor topologies compared in Section 4.1

Before presenting the individual sensor blocks in isolation, it is useful to first establish the complete architectural target that this dissertation works toward, so that the reader can situate each subsequent design decision within the context of the larger system it is intended to support. Figure 4.1 presents the proposed top-level block diagram of the unified PVT monitoring system, and Figure 4.2 reproduces the Finite State Machine control diagram first introduced in Chapter 1, which is intended to govern the sequencing between the three sensing phases.

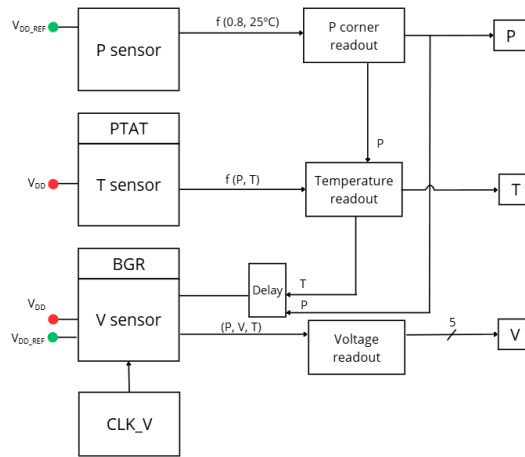


Figure 4.1: Proposed top-level block diagram of the unified PVT monitoring system.

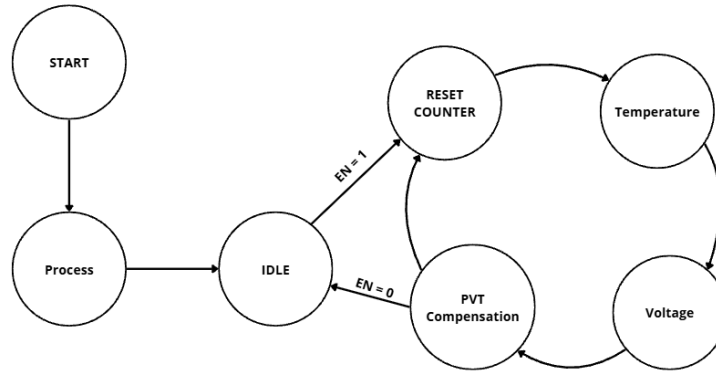


Figure 4.2: Finite State Machine control diagram, reproduced from Chapter 1, governing the sequential calibration routine of the unified PVT system.

As shown in Figure 4.1, the three sensor blocks are intended to be electrically interdependent at the system level, even though, for the reasons discussed in Chapter 5, they are designed and verified independently throughout the remainder of this chapter. The Process sensor accepts a fixed external reference supply, V_{DD_ref} , and produces a frequency output that is interpreted by the P corner readout block to yield the digital process code P . This code is intended to be fed forward into both the temperature sensing path and the voltage sensing path, establishing a data dependency that mirrors the sequential operating principle of the FSM shown in Figure 4.2. Upon reset, the FSM enters the `Process` state first, characterizing the silicon corner of the die under controlled laboratory conditions before any environmental measurement is taken. Once the enable condition is asserted, the FSM proceeds into a cyclic loop in which the `Temperature` state is visited before the `Voltage` state, reflecting the architectural intent that the T sensor, biased by its PTAT core and powered by the live supply rail V_{DD} , should receive the process code P as a correction at its readout, which is co-biased by the BGR reference V_{DD_ref} , the live supply V_{DD} ,

and a dedicated voltage sensor clock CLK_V , and is intended to receive both P and T through a delay block before its own voltage readout stage produces the final digital voltage code V . The cycle concludes at the `PVT Compensation` state, where the three resolved codes would, in the fully realized architecture, be combined to apply real-time operating point adjustments before the loop returns to `RESET COUNTER`.

This architectural target frames the design priorities pursued throughout the remainder of this chapter. Each sensor is evaluated not only on its standalone accuracy at nominal conditions, but specifically on the properties that the unified architecture in Figure 4.1 depends upon: the process sensor's ability to deliver an unambiguous corner classification that can be trusted as a static correction term for the other two blocks, the temperature sensor's independence from supply voltage so that its reading remains valid regardless of the process and voltage states being resolved elsewhere in the loop, and the voltage sensor's ability to accept process and temperature corrections so that its output reflects the true supply rail rather than a conflation of all three environmental variables. The detailed operational walkthrough of how each FSM state maps onto this hardware pipeline, together with an honest accounting of how far the present simulation work has progressed toward this full closed-loop target, is deferred to Chapter 5, once the individual sensor results that justify each design choice have been presented in full.

For each subsystem, the architecture proposed in this work is benchmarked directly against the corresponding state-of-the-art topology identified in Chapter 3, under a common simulation environment, so that the design decision motivating its selection is supported by quantitative evidence. For process monitoring, a diode-connected ratioed inverter oscillator is proposed and evaluated against the NAND/NOR Ratioed Topology of Chandan et al. [32]; although the latter achieves a marginally larger frequency separation at nominal conditions, the proposed topology offers superior robustness to temperature and supply variation and a raw frequency that remains below 1 GHz across all simulated corners without requiring a divide-by-two stage, and is selected on this basis. For supply voltage monitoring, the current-starved delay-chain topology of Chen et al. [4] is adopted as the sensing front-end and co-designed with the FinFET-compatible bandgap reference of Moisello et al. [29]; the inter-stage spacing is re-derived for the 16 nm node through a three-step optimization, since the 2-stage spacing adequate at 65 nm is shown to violate the clock-timing margin required for unambiguous encoding here, with an 8-stage interval ultimately adopted. For temperature sensing, the ring-oscillator architecture of Lempel et al. [44] is used as the fixed sensing front-end, into which three candidate PTAT bias sources are compared: the sensor's original core, the sub-threshold composite-transistor reference of Hanumantha Rao et al. [31], and an area-optimised extraction from the bandgap reference of Moisello et al. [29]; the BGR-based source is disqualified by excessive supply sensitivity despite its wide corner spread, and the composite-transistor reference of Hanumantha Rao et al. [31] is selected for its superior normalised supply rejection.

4.1 Process Sensor

The process sensor implemented in this work follows the ratioed diode-connected inverter principle introduced in Chapter 3, simplified from the NAND/NOR topology of Chandan et al. [32] into a single-transistor-type inverter chain. The primary objective of this section is to present and analyse the simulation results obtained for two distinct ring-oscillator (RO) based process-sensing architectures: the Ratioed Topology of Chandan et al. [32], whose schematic design and operating principle were established in detail in Chapter 3, and the architecture implemented in this work, formally designated here as the Diode Connected Load Ring Oscillator, whose structure, sizing rationale, and simulation results are presented in full for the first time in this chapter. Both topologies are evaluated under a common, rigorously controlled simulation environment so that every quantitative comparison is conducted on a strictly equitable basis. Both architectures are configured identically with 21 ring-oscillator stages to ensure an unbiased structural comparison that mirrors the 21-stage convention established by Chandan et al. [32].

4.1.1 Topological Overview and Structural Differences

Both architectures share the same two-path sensing philosophy. Rather than employing standard CMOS inverters in which both the NMOS pull-down and the PMOS pull-up networks contribute jointly to the propagation delay, each architecture isolates each device type into a dedicated ring oscillator. An NMOS-dominant oscillator is constructed from inverting delay cells whose propagation delay is governed primarily by the NMOS pull-down network, while a complementary PMOS-dominant oscillator is governed by the PMOS pull-up network. By reading the two oscillation frequencies independently and comparing them, the system obtains a two-dimensional process signature capable of distinguishing all five standard process corners (FF, FS, TT, SF, SS), a capability that a single symmetric inverter ring fundamentally cannot provide.

4.1.1.1 The Ratioed Topology

The Ratioed Topology of Chandan et al. [32] employs N-Ratioed NAND and P-Ratioed NOR delay cells as its fundamental building blocks. Its complete schematic design, delay-cell structure, working principle, and ring-oscillator connectivity were established comprehensively in Chapter 3 and are not reproduced here. In the context of this comparative study, the Ratioed Topology serves as the reference architecture against which the Diode Connected Load Ring Oscillator is evaluated.

4.1.1.2 The Diode Connected Load Ring Oscillator

The Diode Connected Load Ring Oscillator is the architecture implemented and proposed in this work. Rather than using standard CMOS inverters, in which both NMOS and PMOS transistors contribute to the propagation delay and therefore to the oscillation frequency, the proposed architecture isolates each device type into a dedicated ring oscillator. An NMOS-dominant oscillator is constructed entirely from inverters whose delay is governed by the NMOS pull-down network,

while a complementary PMOS-dominant oscillator is governed by the PMOS pull-up network. By reading the two oscillation frequencies independently and comparing them, the system obtains a two-dimensional process signature capable of distinguishing all five standard corners.

The fundamental building block of both oscillators is the diode-connected ratioed inverter. The NMOS-sensitive delay cell, illustrated conceptually in Figure 4.3, consists of a diode-connected NMOS transistor M_1 (gate tied to drain) acting as a high-impedance load, and an NMOS transistor M_2 driven by the input signal V_{IN} . The propagation delay of this cell is dominated by the NMOS discharge path: when V_{IN} switches high, M_2 pulls the output node toward ground at a rate determined by its overdrive ($V_{GS2} - V_{th,n}$). Because M_1 is diode-connected, it presents a non-linear resistance that is independent of V_{IN} , ensuring that the delay is predominantly set by the NMOS characteristics and is therefore directly sensitive to the NMOS process corner. The complementary PMOS-sensitive delay cell, shown in Figure 4.4, inverts this arrangement: the PMOS transistor M_1 is the active driver controlled by V_{IN} , while M_2 is a diode-connected PMOS acting as the load. In this case, the charging path through M_1 governs the delay, making the cell primarily sensitive to the PMOS process corner.

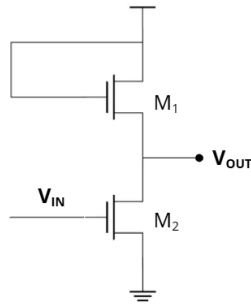


Figure 4.3: Conceptual NMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.

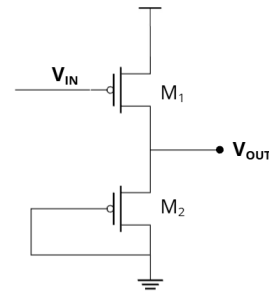


Figure 4.4: Conceptual PMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.

A critical design question for both cell types is the sizing ratio between the active transistor and the diode-connected load. If the load transistor is too strong relative to the driver, that is, if the load presents a low impedance, it will prevent the output node from being pulled fully to the supply rail or to ground, causing the ring oscillator to fail to sustain oscillation. To quantify this constraint, a DC sweep analysis was performed on a single NMOS delay cell across three sizing conditions: load M_1 larger than driver M_2 (the “TOP” case), equal sizes, and driver M_2 larger than load M_1 (the “BOT” case). The resulting output voltage versus input sweep curves are shown in Figure 4.5.

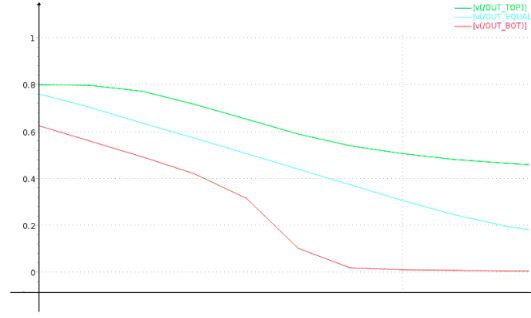


Figure 4.5: DC sweep of the NMOS delay cell output voltage for three transistor sizing ratios, establishing the optimal driver-to-load ratio for the Diode Connected Load Ring Oscillator.

The three sizing cases produce fundamentally different inverter behaviours, summarised in Table 4.1.

Table 4.1: Comparison of delay cell behaviour across the three transistor sizing ratios.

Feature	Top bigger ($M_1 > M_2$)	Equal ($M_1 = M_2$)	Bottom bigger ($M_2 > M_1$)
Logic strength	Weak pull-down	Balanced	Strong pull-down
Gain	Lowest	Medium	Highest
Saturation window	Small	Medium	Large
Oscillation viable	No	Marginal	Yes

The bottom-bigger configuration ($M_2 > M_1$) is confirmed as the optimal choice for ring oscillator operation. It achieves the highest gain, the lowest logic-low level (≈ 0.05 V), and the widest saturation window, all of which contribute to a sharp, well-defined switching transition and a full output voltage swing. This complete swing is the precondition for sustaining reliable oscillation across all process corners. While the top-bigger case yields a high trip point, its weak pull-down prevents the output from reaching a true logic low, making oscillation unreliable. The equal-sizing case sits between the two extremes but still falls short of producing the decisive switching behaviour required for a robust ring oscillator.

The actual schematic implementation of the NMOS delay cell is shown in Figure 4.6. An additional enable transistor (EN) is added in series with the diode-connected load M_1 : when EN is asserted high, the enable NMOS switches on and allows current to flow through the load path, activating the cell; when EN is deasserted, the load path is broken and oscillation is disabled. This enable functionality is essential for the system-level FSM described in Chapter 1, which must gate the oscillators during the process characterization phase to avoid spurious counts in the frequency readout path. The corresponding PMOS delay cell implementation is shown in Figure 4.7. The enable signal is applied to the bottom diode-connected PMOS load in this case, as enabling the PMOS path completes the pull-down current sink and allows the active PMOS driver to produce

a valid output swing. The symmetry of the enable placement between the two cell types ensures that both oscillators are activated and deactivated simultaneously by the same control signal.

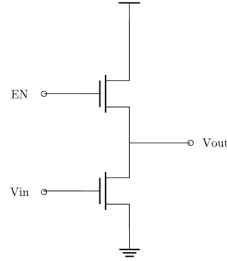


Figure 4.6: NMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.

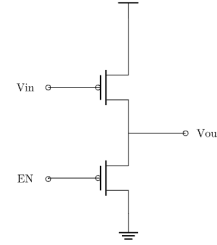


Figure 4.7: PMOS-sensitive delay cell of the Diode Connected Load Ring Oscillator.

Both delay cells are cascaded into 21-stage ring oscillators. The NMOS ring oscillator is shown in Figure 4.8: a first stage containing the diode-connected load and NMOS driver is followed by a cascade of `nmos_inv` cells, with the output of the last stage fed back to the input of the first. An odd number of inverting stages is used to guarantee oscillation, and the EN signal is routed to every cell simultaneously so the entire ring can be gated as a unit. The PMOS ring oscillator shown in Figure 4.9 follows the complementary structure, using `pmos_inv` cells throughout.

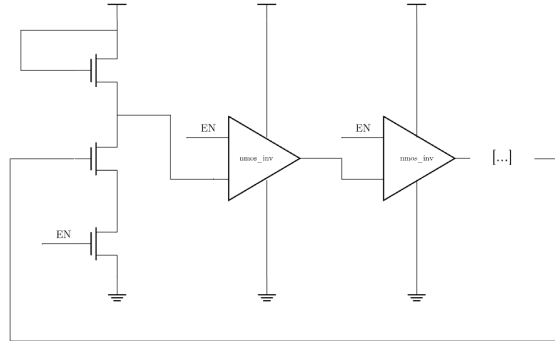


Figure 4.8: 21-stage NMOS ring oscillator of the Diode Connected Load Ring Oscillator.

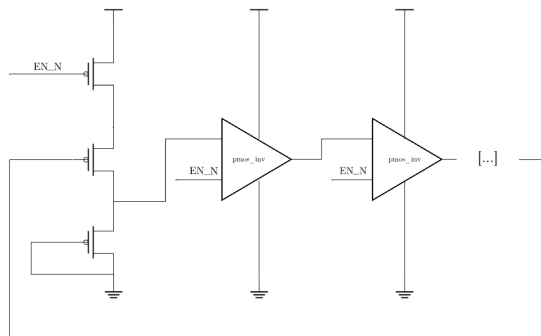


Figure 4.9: 21-stage PMOS ring oscillator of the Diode Connected Load Ring Oscillator.

The sensitivity of the oscillation period to process variation can be quantified through the sensitivity coefficients introduced by Chandan et al. [32]:

$$k_N = \frac{\Delta T / T_0}{\Delta V_n / V_{n0}} \quad \text{and} \quad k_P = \frac{\Delta T / T_0}{\Delta V_p / V_{p0}}$$

where k_N and k_P are the NMOS and PMOS process sensitivity coefficients respectively, ΔT is the deviation of the oscillation period from the typical-corner period T_0 , ΔV_n and ΔV_p are the deviations of the NMOS and PMOS threshold voltages from their respective typical-corner values V_{n0} and V_{p0} . A higher sensitivity coefficient implies that the same shift in threshold voltage produces a proportionally larger change in oscillation period, making the RO output a more discriminating indicator of the underlying process corner.

4.1.2 Testbench Configuration and Frequency Constraints

Both ring-oscillator topologies are evaluated within a common testbench environment to eliminate any instrumentation-related asymmetry from the comparative analysis. The testbench layout used for all process-sensor simulations is shown in Figure 4.10. The supply voltage is applied uniformly to all active cell instances, and the enable signal is asserted globally so that all 21 stages activate and deactivate simultaneously. Each oscillator path follows an identical three-stage readout chain: the raw oscillator output f_{osc} is first passed through a buffer stage to reshape the signal and prevent the input capacitance of downstream measurement instruments from loading the sensitive feedback loop and perturbing the oscillation frequency. The buffered signal f_{buff} is then routed through an inverter, which simultaneously reshapes the waveform into a clean digital transition and generates the complementary output $\bar{f}_{buff}$ required by the subsequent divide-by-two stage. The raw output of a ring oscillator built from diode-connected or ratioed delay cells exhibits a sinusoidal-like transition profile with finite rise and fall times that are asymmetric across process corners; the inverter restores the signal to a well-defined two-level waveform with sharp edges, ensuring that the frequency counter in the readout chain measures the true oscillation period rather than a corner-dependent distortion of it. The final divide-by-two stage, clocked by $\bar{f}_{buff}$, produces the output frequency f_{out} that is presented to the measurement instruments, and inherently restores a 50% duty cycle regardless of any duty-cycle unavoidable imbalance present at its input. This three-stage chain — buffer, inverter, and divide-by-two — is applied symmetrically to both the PMOS-path and NMOS-path oscillators, as illustrated in Figure 4.10.

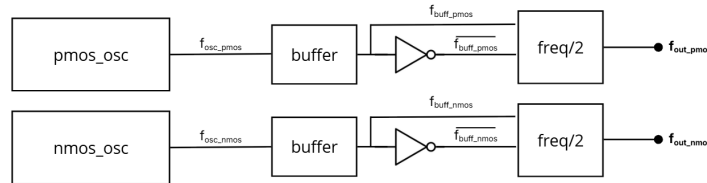


Figure 4.10: Shared testbench configuration used for evaluating both the Diode Connected Oscillator and the Ratioed Topology.

An important distinction between the two topologies emerges in the context of output frequency management. The Ratioed Topology of Chandan et al. [32] produces raw oscillation frequencies that can exceed the target ceiling of 1 GHz at fast process corners and elevated supply voltages, requiring the divide-by-two stage in its readout chain to bring the signal into a countable range. The Diode Connected Load Ring Oscillator does not face this constraint in the same way, since the diode-connected load arrangement inherently slows each stage transition relative to the NAND/NOR delay cells of the Ratioed Topology, keeping its raw output frequency naturally below the 1 GHz threshold across all simulated corners, temperatures, and supply voltages tested in this work. However, the divide-by-two stage is retained in the Diode Connected Load Ring Oscillator’s readout chain for the duty-cycle restoration reason described above, and the resulting halved output frequency is a secondary benefit that additionally keeps the readout requirements consistent with those of the Ratioed Topology.

4.1.3 Transient Analysis and Frequency Separation

Before comparing the two architectures quantitatively across environmental sweeps, it is instructive to examine the time-domain waveform behaviour of the Ratioed Topology to establish the baseline corner separation behaviour against which the Diode Connected Load Ring Oscillator is evaluated. The transient simulation results for the NMOS-path and PMOS-path ring oscillators of the Ratioed Topology, captured simultaneously across all five process corners at the nominal operating point of 25 °C and $V_{DD} = 0.8$ V, are shown in Figures 4.11 and 4.12 respectively.

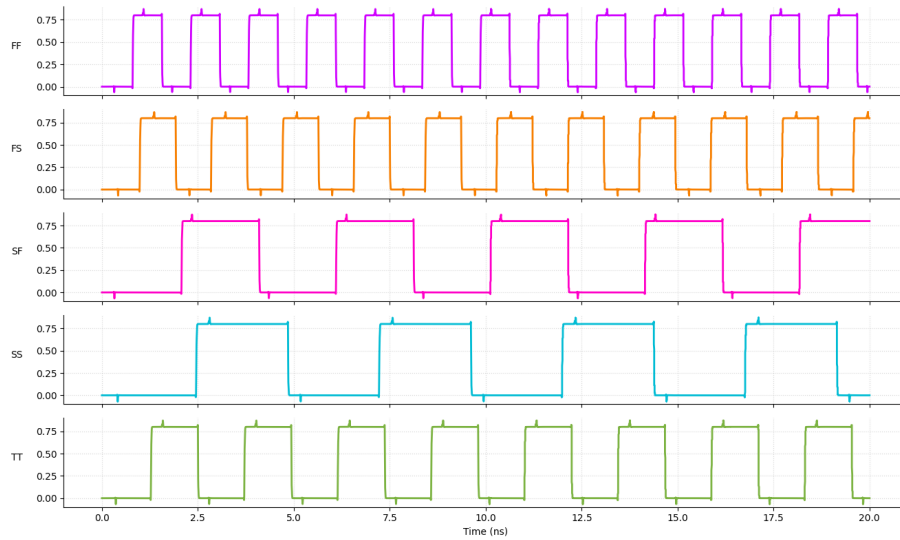


Figure 4.11: Transient output of the NMOS-path Ratioed Topology ring oscillator across all five process corners at 25 °C and $V_{DD} = 0.8$ V.

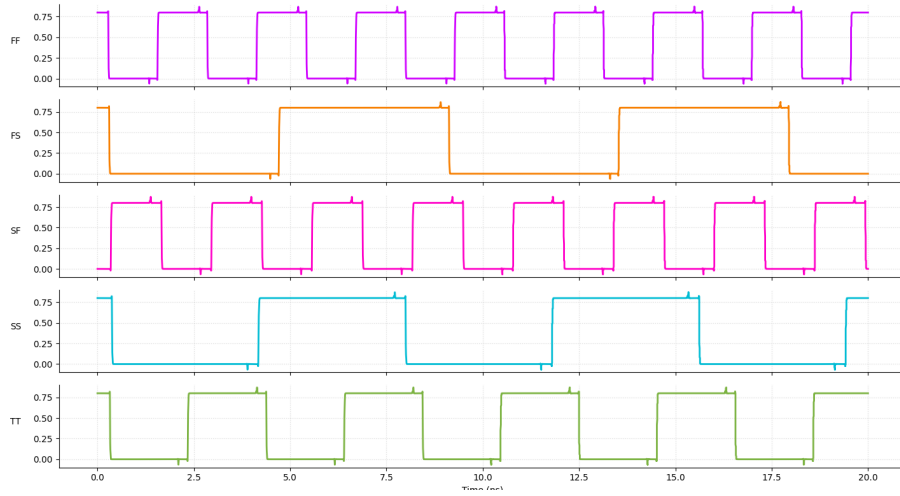


Figure 4.12: Transient output of the PMOS-path Ratioed Topology ring oscillator across all five process corners at 25 °C and $V_{DD} = 0.8$ V.

The transient plots provide a strong visual demonstration of the fundamental operating principle shared by both architectures: each process corner produces a distinctly different oscillation period, and these periods remain stable over the simulated time window, confirming that the ring oscillators sustain reliable oscillation across all five corners. In the NMOS-path oscillator of the Ratioed Topology (Figure 4.11), the FF corner produces the shortest period and therefore the highest frequency, since fast NMOS devices discharge the output node rapidly. The SF corner (slow NMOS, fast PMOS) produces the longest period in this path because the NMOS pull-down is sluggish regardless of the PMOS condition. The FS and SS corners occupy intermediate positions, with the clear separation between corner waveforms confirming that the NMOS-path oscillator is primarily governed by the NMOS threshold voltage.

The complementary behaviour is visible in the PMOS-path waveforms (Figure 4.12): here the SF corner yields the shortest period because the fast PMOS pull-up charges the output node quickly, while the FS corner (slow PMOS) is associated with the longest period. This complementary inversion of the corner ordering between the two oscillator paths is precisely the two-dimensional signature that enables unambiguous identification of all five process corners, including the skewed FS and SF corners that a single symmetric ring oscillator cannot distinguish. The same corner-separation behaviour is exhibited by the Diode Connected Load Ring Oscillator, as confirmed by the corner analysis presented in the following subsection.

4.1.4 Comparative Process Corner Variations at Nominal Conditions

With the baseline transient behaviour of the Ratioed Topology established, the analysis now turns to a direct side-by-side comparison of both architectures across all five process corners at the nominal operating point (25 °C, $V_{DD} = 0.8$ V). Figure 4.13 presents the corner-resolved oscillation frequencies for the Diode Connected Load Ring Oscillator plotted in the (F_{NMOS}, F_{PMOS}) plane,

and Figure 4.14 presents the corresponding two-dimensional corner map for the Ratioed Topology of Chandan et al. [32].

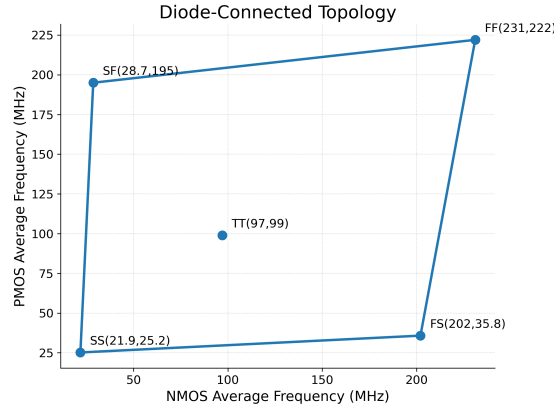


Figure 4.13: Two-dimensional process corner map of the Diode Connected Load Ring Oscillator at nominal conditions (25°C , $V_{DD} = 0.8\text{ V}$), plotting PMOS-path average frequency against NMOS-path average frequency for each of the five corners.

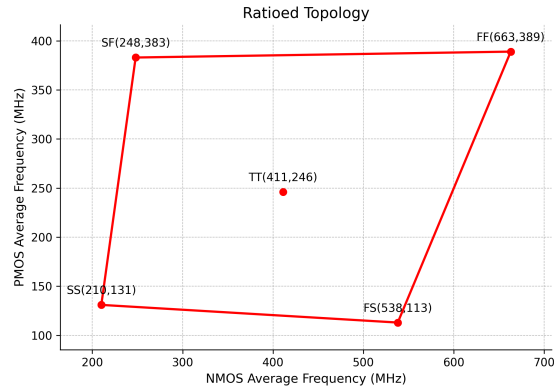


Figure 4.14: Two-dimensional process corner map of the Ratioed Topology at nominal conditions (25°C , $V_{DD} = 0.8\text{ V}$), included as a reference for direct comparison.

The two-dimensional corner maps reveal the full discriminability of each architecture more clearly than frequency bar charts: each process corner occupies a distinct position in the (F_{NMOS}, F_{PMOS}) plane, and the five positions form a convex polygon whose vertices and interior point correspond unambiguously to the five standard corners.

Examining the Diode Connected Load Ring Oscillator in Figure 4.13, the two-dimensional separation is immediately apparent. The SS corner sits at the origin-proximal vertex (43.7 MHz, 50.4 MHz), reflecting the combined effect of slow NMOS and slow PMOS devices depressing both path frequencies simultaneously. Moving along the PMOS axis, the SF corner (57.5 MHz, 391 MHz) occupies a position with a similarly low NMOS frequency but a markedly elevated PMOS frequency, as the fast PMOS pull-up accelerates the PMOS-path oscillator while the slow NMOS pull-down continues to retard the NMOS path. The FS corner (405 MHz, 71.5 MHz) is the

mirror image: a high NMOS frequency driven by the fast NMOS pull-down, paired with a suppressed PMOS frequency from the slow PMOS driver. FF (462 MHz, 444 MHz) occupies the high-frequency vertex in both dimensions, while TT (194 MHz, 198 MHz) falls near-symmetrically at the centre of the polygon. The near-equal TT coordinates reflect the balanced threshold voltages of the typical corner and confirm that the diode-connected load arrangement imposes comparable loading on both path types at nominal process conditions. The five corner positions are widely spaced in the plane, producing a large convex polygon that provides ample margin for the FSM classification logic.

One notable feature of the Diode Connected Load Ring Oscillator map is the pronounced asymmetry in the absolute frequency range of each path. The NMOS-path spans from 43.7 MHz (SS) to 462 MHz (FF), a ratio of approximately $10.6\times$, while the PMOS-path spans from 50.4 MHz (SS) to 444 MHz (FF), a ratio of approximately $8.8\times$. This large dynamic range, a direct consequence of the high impedance presented by the diode-connected load, means that even a moderate degradation from the FF corner towards TT produces a frequency shift that is large in absolute terms, facilitating robust digital counting within the readout window.

Comparing the two architectures, the Ratioed Topology in Figure 4.14 operates at substantially higher absolute frequencies: the NMOS path ranges from 210 MHz (SS) to 663 MHz (FF) and the PMOS path from 113 MHz (FS) to 389 MHz (SF), with TT at (411 MHz, 246 MHz). The overall shape of the corner polygon is qualitatively similar to that of the Diode Connected Load Ring Oscillator, confirming that both architectures encode the same two-dimensional process signature. The Ratioed Topology produces a marginally larger absolute frequency separation between adjacent corners at this specific nominal operating point, a predictable consequence of its NAND/NOR delay-cell architecture having been optimised to maximise process sensitivity at typical conditions. However, the Diode Connected Load Ring Oscillator produces a proportionally larger relative corner spread when expressed as a ratio of the inter-corner frequency gap to the centre frequency, and its raw frequencies remain below 1 GHz across all corners even before the divide-by-two stage is applied, as discussed in Section 4.1.2.

To provide a device-level grounding for the frequency differences visible in the corner maps, Table 4.2 decomposes the inter-corner variation into its constituent physical contributors across four representative corner pairs. Each percentage entry in the frequency columns is computed as the fractional deviation between the two corners under comparison:

$$\Delta_{\%} = \left(1 - \frac{F_{cornerA}}{F_{cornerB}} \right) \times 100 [\%],$$

where $F_{cornerA}$ and $F_{cornerB}$ are the oscillation frequencies measured at the two corners forming each pair, with $F_{cornerB} > F_{cornerA}$. For example, for the PMOS FF–SF pair of the Diode Connected Load Ring Oscillator, this evaluates to $(1 - 195/222) \times 100 \approx 14\%$.

Table 4.2: Simulated parameter variations across process corner pairs for the PMOS-path and NMOS-path oscillators of each topology.

Path	Corner Pair	ΔI_{DS}	ΔV_{th}	ΔC_g	$f_{osc,diode}$	$f_{buff,diode}$	$f_{out,diode}$	$f_{out,ratioed}$
PMOS	FF–SF	5%	7%	1%	14%	14%	14%	2%
	FS–SS	10%	4%	1%	42%	42%	42%	16%
NMOS	FF–FS	5%	6%	1%	14%	14%	14%	23%
	SF–SS	10%	4%	1%	31%	32%	32%	18%

The table serves two distinct analytical purposes. The first is to confirm that the inverter and divide-by-two stages in the readout chain do not distort the process signature. To establish this, the Diode Connected Load Ring Oscillator was measured at three successive points in its readout path: directly at the raw oscillator output ($f_{osc,diode}$), after the buffer and inverter stage ($f_{buff,diode}$), and after the divide-by-two stage ($f_{out,diode}$). The three columns are in perfect agreement across all four corner pairs (14%/14%/14%, 42%/42%/42%, 14%/14%/14%, and 31%/32%/32% for PMOS FF–SF, PMOS FS–SS, NMOS FF–FS, and NMOS SF–SS respectively), confirming that neither the inverter nor the frequency divider introduces any measurable distortion or asymmetry into the process signature. Having established this on the Diode Connected Load Ring Oscillator, the same inverter, buffer, and divide-by-two blocks were then inserted into the Ratioed Topology readout chain with confidence that any frequency variation observed at the output of that chain ($f_{out,ratioed}$) is attributable solely to the process sensitivity of the Ratioed Topology delay cells themselves, and not to artefacts introduced by the measurement infrastructure.

The second purpose is to evaluate, for each topology, how well its oscillator isolates the device type it is intended to monitor. Each oscillator pairs an NMOS-path and a PMOS-path ring built predominantly from one device type, and under ideal conditions its frequency should track the process corner of that device alone. Crucially, the corner pairs in each row of Table 4.2 are chosen to vary *only the corner letter of the device not nominally monitored by that path*, while holding the monitored device’s corner fixed: the PMOS-path rows (FF–SF, FS–SS) vary the NMOS corner letter while keeping the PMOS letter constant, and the NMOS-path rows (FF–FS, SF–SS) vary the PMOS corner letter while keeping the NMOS letter constant. Consequently, $f_{out,diode}$ and $f_{out,ratioed}$ in this table do not measure how strongly each oscillator responds to its own device corner; they measure *cross-sensitivity* – the unwanted residual frequency shift induced by the corner of the device the path is not meant to be sensing. Under this reading, a *lower* percentage indicates a cleaner, more device-selective sensor, since the oscillator’s frequency is less perturbed by variation in the device it should be largely insensitive to, whereas a higher percentage indicates that the path is partially confounded by the other device’s corner, which would impair its ability to attribute a frequency shift unambiguously to the intended process parameter.

Comparing $f_{out,diode}$ and $f_{out,ratioed}$ on this basis shows that the Ratioed Topology achieves lower cross-sensitivity than the Diode Connected Load Ring Oscillator in three of the four evaluated corner pairs. On the PMOS path, the Ratioed Topology is substantially more device-selective:

the FF–SF pair shows only 2% residual variation against 14% for the Diode Connected Oscillator, and the FS–SS pair shows 16% against 42%. This indicates that the diode-connected load arrangement is considerably more susceptible to being perturbed by the NMOS corner when it is meant to be isolating the PMOS corner, whereas the ratioed delay cell holds its PMOS-path frequency comparatively stable as the NMOS corner is swept. On the NMOS path, the picture is more mixed: the FF–FS pair favours the Diode Connected Load Ring Oscillator, which shows 14% cross-sensitivity against 23% for the Ratioed Topology, while the SF–SS pair reverses this, with the Ratioed Topology showing 18% against 32% for the Diode Connected Load Ring Oscillator. Averaged across all four pairs, the Ratioed Topology exhibits a mean cross-sensitivity of 14.75%, compared with 25.5% for the Diode Connected Load Ring Oscillator – roughly half the unwanted coupling.

To complement this cross-sensitivity analysis with a measure of each topology’s intrinsic responsiveness to its own monitored device, the process sensitivity coefficients k_N and k_P introduced in Section 4.1.1 were evaluated for both architectures using the peak-to-peak corner span on each path. These coefficients are summarised in Table 4.3.

Table 4.3: Comparison of NMOS and PMOS process sensitivity coefficients between the Diode Connected Load Ring Oscillator and the Ratioed Topology, computed as $k = \frac{\Delta T/T_0}{\Delta V/V_0}$, where ΔT is the peak-to-peak period difference between the fastest and slowest corner on each path, T_0 is the period at the TT corner, and $\Delta V/V_0$ is the corresponding relative threshold-voltage deviation.

Topology	k_N	k_P
Diode Connected Load Ring Oscillator	4.01	3.48
Ratioed Topology	1.34	1.54

The Diode Connected Load Ring Oscillator exhibits a substantially higher own-device sensitivity than the Ratioed Topology on both paths: k_N is roughly $3\times$ larger (4.01 vs. 1.34) and k_P is roughly $2.3\times$ larger (3.48 vs. 1.54). This is consistent with the wide relative frequency spans already noted for the Diode Connected Load Ring Oscillator earlier in this section (an approximately $10.6\times$ NMOS-path and $8.8\times$ PMOS-path range across corners, against the Ratioed Topology’s narrower ratios), and confirms that the diode-connected load arrangement converts a given threshold-voltage shift into a proportionally larger period change than the NAND/NOR ratioed load does.

This result, read alongside the cross-sensitivity figures discussed above, shows that k_N/k_P and cross-sensitivity are measuring two different, not necessarily aligned, properties. The Ratioed Topology’s lower mean cross-sensitivity (14.75% vs. 25.5%) reflects how little a path is disturbed by the *other* device’s corner, whereas its lower k_N/k_P here reflects how little that same path responds to its *own* device’s corner. A topology can therefore be simultaneously cleaner in isolation (Ratioed) and less sensitive overall (also Ratioed in this case) – the two metrics happen to favour the same topology here, but that need not be the case in general: a sensor with high own-device

sensitivity but poor device isolation, or vice versa, is conceivable. For the present comparison, the Diode Connected Load Ring Oscillator trades some of the Ratioed Topology's cleaner device separation for markedly higher raw sensitivity to its own target device, which in practice manifests as the larger absolute and relative frequency spreads already observed in the corner maps of Figure 4.13.

Neither topology is uniformly superior: the Diode Connected Load Ring Oscillator remains the more device-selective choice specifically on the NMOS path at the FF–FS corner pair, indicating that its weakness is concentrated rather than uniform. However, taken in aggregate, the Ratioed Topology demonstrates better isolation between the two monitored process parameters across the majority of the evaluated corner space, making it the preferable choice for a process sensor in which clean attribution of a frequency shift to a single device type is the primary design goal, as argued further in Section 4.1.7.

4.1.5 Temperature Dependency

As mentioned earlier, the supply voltage was held fixed at $V_{DD} = 0.8\text{ V}$ and the process corner fixed at typical (TT) for this topology, so that any variation in oscillation frequency across the temperature sweep could be attributed to temperature alone rather than confounded with supply or process variation.

In FinFET technologies, two competing mechanisms govern this temperature dependence. As temperature rises, V_{th} decreases (negative temperature coefficient), increasing overdrive and drive current and thus frequency; simultaneously, phonon scattering intensifies, degrading mobility and drive current and thus reducing frequency. Planar devices are typically mobility-dominated, giving a frequency decrease with temperature. The strong gate control of FinFETs, however, suppresses short-channel effects and lowers the intrinsic V_{th} temperature coefficient, so the balance between the two mechanisms can shift. In both topologies evaluated here, the V_{th} -reduction effect dominates, producing a monotonic frequency increase with temperature across the full industrial range.

4.1.5.1 Diode Connected Load Ring Oscillator Temperature Response

Figures 4.15 and 4.16 show the frequency versus temperature characteristics of the NMOS-path and PMOS-path Diode Connected Load Ring Ring Oscillator respectively, evaluated at the TT process corner.

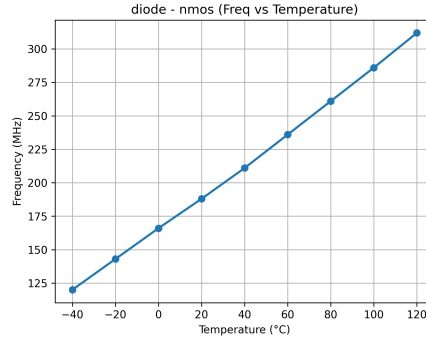


Figure 4.15: NMOS-path Diode Connected Load Ring Oscillator oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8\text{ V}$).

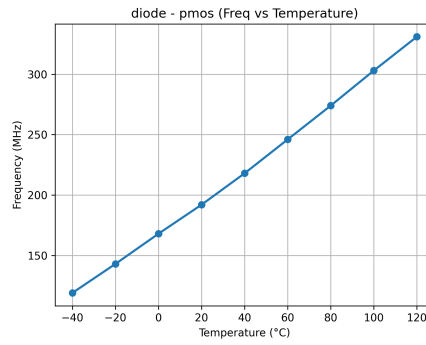


Figure 4.16: PMOS-path Diode Connected Load Ring Oscillator oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8\text{ V}$).

In both oscillator paths of the Diode Connected Load Ring Oscillator, the frequency increases monotonically with temperature, rising from approximately 120 MHz at -40°C to approximately 310 MHz and 330 MHz for the NMOS-path and PMOS-path respectively at 120°C , corresponding to a frequency excursion of roughly 190–210 MHz across the full industrial range. The response is approximately linear throughout the sweep. The relatively moderate absolute variation observed is a direct consequence of the diode-connected delay cell topology: because the driver and diode-connected load transistors of each stage are of the same device type and share the same fin geometry, their threshold-voltage and mobility temperature coefficients track closely with one another. The resulting partial cancellation of temperature-induced delay changes between the driver and load limits the net stage delay variation, and therefore constrains the overall frequency excursion to a smaller range than would be expected from an asymmetric load topology. This self-compensating behaviour is a significant advantage in a FinFET process sensor, as it reduces the frequency shift attributable to temperature alone and eases the discrimination requirement placed on the readout logic.

4.1.5.2 Ratioed Topology Temperature Response

Figures 4.17 and 4.18 present the temperature characteristics of the NMOS-path and PMOS-path Ratioed Topology oscillators respectively at the TT process corner, included here as a reference for direct comparison with the Diode Connected Load Ring Oscillator.

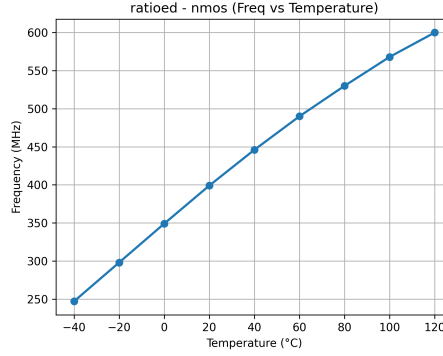


Figure 4.17: NMOS-path Ratioed Topology oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8\text{ V}$), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.

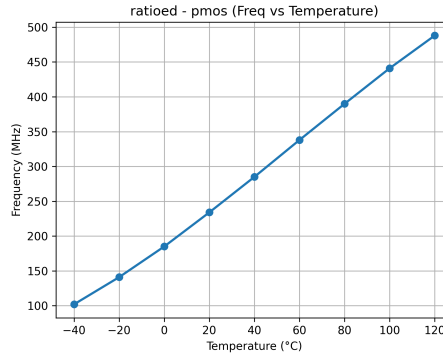


Figure 4.18: PMOS-path Ratioed Topology oscillation frequency versus temperature at the TT process corner ($V_{DD} = 0.8\text{ V}$), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.

The Ratioed Topology similarly exhibits a monotonically increasing and approximately linear frequency response with rising temperature. However, the absolute frequency excursion is considerably larger than that of the Diode Connected Load Ring Oscillator: the NMOS-path rises from approximately 250MHz at -40°C to approximately 600MHz at 120°C , a span of roughly 350MHz, while the PMOS-path spans approximately 390MHz over the same range. This greater temperature sensitivity arises from the inherent asymmetry of the ratioed delay cell: the NAND/NOR driver devices and the ratioed load devices differ in type and sizing, and in a Fin-FET process the discrete quantization of fin count means this sizing asymmetry is fixed in integer steps. As a result, the temperature coefficients of the driver and load do not cancel as effectively

as in the diode-connected case, and the net stage delay exhibits a larger variation per degree of temperature change. The wider frequency excursion imposes a more demanding constraint on the readout circuit, which must reliably resolve frequency differences in the presence of a larger temperature-induced background variation than is encountered with the Diode Connected Load Ring Oscillator.

4.1.6 Supply Voltage Sensitivity

Supply voltage variations directly modulate the overdrive voltage of every transistor in the delay cells, producing a significant effect on oscillation frequency; this sensitivity is particularly pronounced in FinFETs at low V_{DD} close to threshold, where drive current depends strongly on overdrive and small changes in V_{DD} yield disproportionately large changes in drive strength. Both topologies were swept over $V_{DD} = 0.65\text{--}0.95\text{ V}$, centred on the nominal 0.8 V , with temperature fixed at 25°C and the process corner fixed at TT, so that the resulting frequency variation could be attributed to supply voltage alone.

4.1.6.1 Diode Connected Load Ring Oscillator VDD Response

The oscillation frequency of the NMOS-path and PMOS-path Diode Connected Load Ring Oscillator as a function of supply voltage are shown in Figures 4.19 and 4.20 respectively.

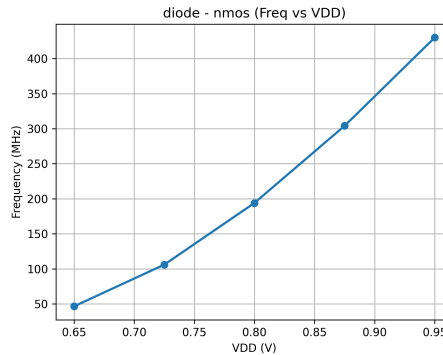


Figure 4.19: NMOS-path Diode Connected Load Ring Oscillator frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$).

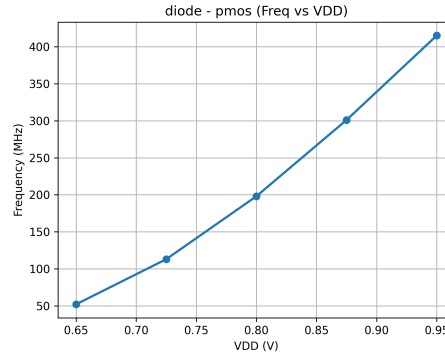


Figure 4.20: PMOS-path Diode Connected Load Ring Oscillator frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$).

The Diode Connected Load Ring Oscillator exhibits a strong positive and superlinear dependence of oscillation frequency on supply voltage in both paths. The NMOS-path frequency increases from approximately 50 MHz at 0.65 V to approximately 425 MHz at 0.95 V, while the PMOS-path rises from approximately 52 MHz to 415 MHz over the same interval. The superlinear character of the response is consistent with FinFET operation in the saturation regime, where drain current scales approximately with the square of the overdrive voltage; at lower supply voltages approaching the threshold, the sensitivity is further amplified as a greater proportion of the supply headroom is consumed by V_{th} itself. Despite the large total frequency span, the smooth and monotonic nature of the response means that supply-induced frequency shifts are well-behaved and can be accounted for through system-level calibration.

4.1.6.2 Ratioed Topology VDD Response

The corresponding VDD sweep results for the Ratioed Topology at the TT corner are shown in Figures 4.21 and 4.22.

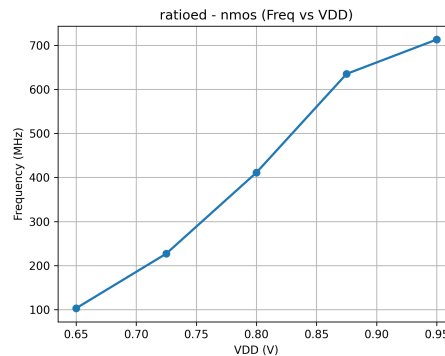


Figure 4.21: NMOS-path Ratioed Topology oscillation frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.

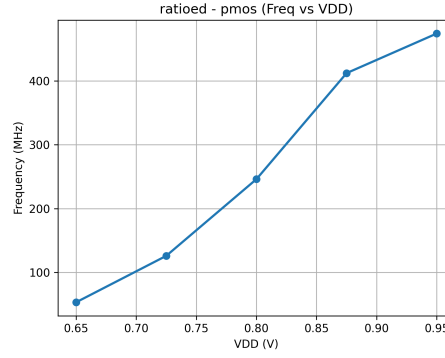


Figure 4.22: PMOS-path Ratioed Topology oscillation frequency versus supply voltage at the TT process corner ($T = 25^\circ\text{C}$), shown as a reference for comparison with the Diode Connected Load Ring Oscillator.

The Ratioed Topology exhibits a similarly positive and superlinear frequency response to supply voltage, but with a considerably steeper slope than the Diode Connected Load Ring Oscillator. The NMOS-path frequency rises from approximately 103 MHz at 0.65 V to approximately 713 MHz at 0.95 V, a total span of roughly 610 MHz, compared with approximately 375 MHz for the Diode Connected Load Ring NMOS-path over the same range. The PMOS-path similarly spans approximately 415 MHz. The heightened voltage sensitivity of the Ratioed Topology stems from its asymmetric driver-to-load structure: as V_{DD} increases, the overdrive margins of both the driver and ratioed load devices rise simultaneously but non-uniformly, compounding the rate of frequency increase relative to the diode-connected case where the driver and load track more symmetrically. In a FinFET process, where the discrete fin quantization constrains the available sizing ratios, this asymmetry is fixed and cannot be tuned continuously to reduce the V_{DD} sensitivity. As a practical consequence, even modest supply fluctuations due to load-dependent ripple, IR drop, or transient disturbances produce a more substantial absolute shift in the measured oscillation frequency in the Ratioed Topology than in the Diode Connected Oscillator, representing a more demanding requirement on supply regulation when this topology is deployed in an on-chip environment.

4.1.7 Architecture Selection and Design Justification

The simulation results presented in Sections 4.1.4 through 4.1.6 provide a comprehensive empirical basis for selecting one of the two architectures as the process sensor to be integrated into the final unified PVT monitoring platform. The following analysis weighs the relative merits of each topology across process corner discrimination, temperature stability, supply voltage robustness, and circuit-level implementation complexity, and articulates the reasoning behind the selection of the Diode Connected Load Ring Oscillator as the final design.

At the nominal operating point of 25°C and $V_{DD} = 0.8\text{ V}$, the Ratioed Topology produces a larger absolute frequency spread between adjacent process corners, reflecting the fact that its NAND/NOR delay-cell structure was specifically engineered to maximise process sensitivity at the

typical operating point. The Ratioed Topology NMOS-path spans 210 MHz to 663 MHz and the PMOS-path spans 113 MHz to 389 MHz across the five corners, while the Diode Connected Load Ring Oscillator operates at lower absolute frequencies, with the NMOS-path spanning 43.7 MHz to 462 MHz and the PMOS-path spanning 50.4 MHz to 444 MHz. However, this comparison of absolute frequency span is misleading when taken in isolation, since the two topologies operate at substantially different baseline frequencies. The process sensitivity coefficients k_N and k_P reported in Table 4.3 normalise this difference away by expressing the period change per unit relative threshold-voltage deviation, and on this basis the comparison reverses: the Diode Connected Load Ring Oscillator achieves $k_N = 4.01$ and $k_P = 3.48$, roughly $3\times$ and $2.3\times$ higher respectively than the Ratioed Topology's $k_N = 1.34$ and $k_P = 1.54$. This indicates that, relative to its own operating frequency, the Diode Connected Load Ring Oscillator is in fact the more sensitive of the two architectures to its targeted device corner – a given shift in V_{th} produces a proportionally larger period change in the diode-connected delay cell than in the ratioed NAND/NOR cell. The sensitivity decomposition in Table 4.2 additionally shows that the Ratioed Topology achieves better isolation between the two monitored device types (lower cross-sensitivity, Section 4.1.4), so the overall nominal-condition picture is one of a trade-off rather than a one-sided advantage: the Ratioed Topology is the cleaner, more device-selective sensor at the nominal corner, while the Diode Connected Load Ring Oscillator is the more sensitive one. In either case, the Diode Connected Load Ring Oscillator still achieves clear and unambiguous five-corner discrimination in the two-dimensional (F_{NMOS}, F_{PMOS}) frequency plane at nominal conditions, with all five corners well-separated and fully resolvable by the FSM comparison logic, so neither the lower cross-sensitivity of the Ratioed Topology nor the higher own-device sensitivity of the Diode Connected Load Ring Oscillator represents a capability the other architecture fundamentally lacks at this operating point.

The more decisive advantages of the Diode Connected Load Ring Oscillator emerge when the evaluation is extended beyond the nominal operating point to the full environmental envelope that a deployed process sensor must survive. With respect to temperature, both topologies exhibit a monotonically increasing frequency response with rising temperature across the industrial range of -40°C to 120°C , a behaviour governed in this 16 nm FinFET technology by the dominance of threshold-voltage reduction over mobility degradation. However, the magnitude of the resulting frequency excursion differs substantially between the two architectures. The Diode Connected Load Ring Oscillator spans approximately 190–210 MHz across the full temperature range, while the Ratioed Topology spans approximately 350–390 MHz over the same interval, a factor of roughly $1.8\times$ larger. This greater temperature-induced frequency variation is a direct consequence of the asymmetric driver-to-load structure of the NAND/NOR delay cells, in which the temperature coefficients of the driver and load devices do not cancel effectively. In the Diode Connected Load Ring Oscillator, the driver and diode-connected load share the same device type and fin geometry, so their threshold-voltage and mobility shifts track closely with temperature and partially cancel at each stage, constraining the net frequency excursion to a smaller range. The practical implication is significant: a larger temperature-induced frequency variation increases the risk that a temperature change alone causes the readout logic to perceive a shift in process state,

degrading the diagnostic integrity of the sensor across the operating life of the device. The Diode Connected Load Ring Oscillator's lower temperature coefficient therefore directly translates into more reliable process classification wherever the die operates away from the nominal temperature.

An equivalent argument applies to supply voltage robustness. Over the VDD sweep from 0.65 V to 0.95 V at the TT corner, the Diode Connected Oscillator NMOS-path spans approximately 375 MHz, while the Ratioed Topology NMOS-path spans approximately 610 MHz, a factor of roughly $1.6\times$ larger across the same supply range. The heightened VDD sensitivity of the Ratioed Topology stems from its asymmetric driver-to-load structure, in which the overdrive margins of the driver and ratioed load devices increase non-uniformly with supply voltage, compounding the rate of frequency change. In an on-chip environment subject to load-dependent supply ripple, IR drop across the power grid, or transient disturbances from neighbouring circuit blocks, this steeper VDD sensitivity means that supply perturbations produce proportionally larger absolute frequency shifts, increasing the probability of a spurious corner classification driven by supply noise rather than the true process state of the silicon. The Diode Connected Load Ring Oscillator's shallower and more predictable VDD response makes it substantially more resilient to such artefacts, and its process corner separation remains clearly resolved across the full supply sweep without requiring additional supply filtering or compensation circuitry.

Beyond environmental robustness, the Diode Connected Load Ring Oscillator also offers a significant circuit-level implementation advantage. Its raw oscillation frequency remains naturally below 1 GHz across all evaluated process corners, temperatures, and supply voltages, a direct consequence of the higher impedance presented by the diode-connected load relative to the NAND/NOR cells of the Ratioed Topology. This eliminates the strict frequency division requirement that constrains the Ratioed Topology readout chain, reducing the criticality of the divide-by-two stage from a functional necessity to a duty-cycle correction block, and removing a potential source of duty-cycle asymmetry and phase jitter from the measurement path.

In summary, while the Ratioed Topology achieves better device isolation at the exact nominal operating point, the Diode Connected Load Ring Oscillator is in fact the more sensitive of the two to its own targeted device corner once frequency-normalised through k_N and k_P , and provides fully sufficient five-corner resolution at nominal conditions regardless of which metric is emphasised. Beyond the nominal point, the Diode Connected Load Ring Oscillator offers demonstrably superior robustness to temperature variation, reduced sensitivity to supply voltage fluctuations, and a simpler readout architecture requiring no frequency division for range management. In a real deployed sensor that must operate reliably across the full industrial temperature and supply voltage envelope, these environmental robustness advantages – together with its higher intrinsic process sensitivity – outweigh the nominal-condition device-isolation margin of the Ratioed Topology. The Diode Connected Load Ring Oscillator is therefore selected as the process sensor for all subsequent system-level integration and co-design described in the remainder of this chapter, a selection further substantiated by the Monte Carlo robustness analysis presented in Section 4.1.8.

4.1.8 Monte Carlo Validation of the Diode Connected Load Ring Oscillator

Having selected the Diode Connected Load Ring Oscillator as the process sensor architecture in Section 4.1.7, its manufacturability was further validated through Monte Carlo (MC) simulation, which superimposes random local mismatch and global process variation on top of the deterministic corner models used so far. Robustness is assessed throughout this section at the 3σ level, since a process sensor must reliably classify the corner for essentially the entire manufactured population, and this is the criterion applied consistently below. An equivalent Monte Carlo characterization of the Ratioed Topology was attempted for direct comparison; however, the simulation environment did not execute correctly for this topology within the available project timeframe, and the following analysis is therefore limited to the Diode Connected Load Ring Oscillator.

4.1.8.1 Initial Sizing: Poor Corner Separability

Figure 4.23 shows the first Monte Carlo run, performed on the initial transistor sizing of the diode-connected delay cells (200 runs per corner, 5 corners, plotted in the (F_{NMOS}, F_{PMOS}) plane). At the 3σ level, the five corner envelopes already overlap substantially: the TT cluster (green) overlaps with the tails of both the SF (blue) and FS (orange) clusters, and several SF, FS, and TT samples fall well outside even their respective 3σ envelopes – visible as isolated outlier points extending toward the cluster boundaries below 75 MHz and above 100 MHz on the NMOS axis. The FF cluster (red) shows the largest absolute spread, with NMOS frequency samples ranging from approximately 120 MHz to 205 MHz, a span of nearly 80 MHz around a centroid near 165 MHz. This sizing was therefore judged insufficiently robust: a real die exhibiting process variation under the TT corner could, at the 3σ tail, produce a frequency reading that overlaps the SF or FS region, causing the FSM corner classification logic to misidentify the process state.

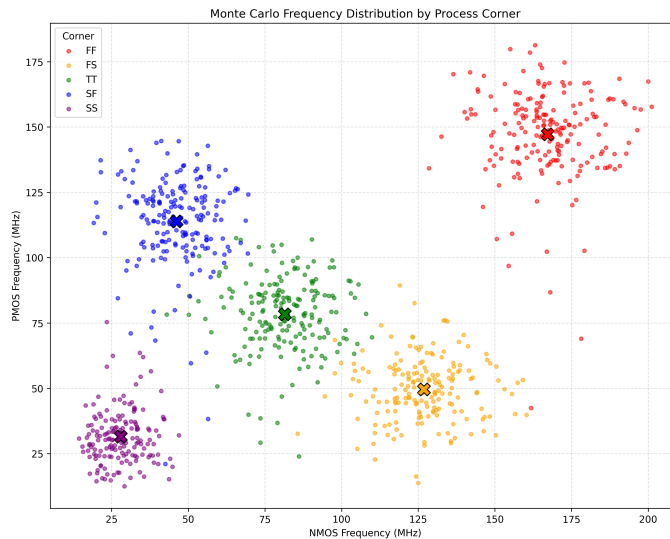


Figure 4.23: Initial Monte Carlo distribution of the Diode Connected Load Ring Oscillator (pre-resizing).

4.1.8.2 Resized Cells: Improved but Imperfect Robustness

To address this, the driver-to-load transistor ratio within each diode-connected delay cell was resized, following the bottom-bigger sizing principle established in Section 4.1.1 (Table 4.1), but with the absolute device widths increased to reduce the relative impact of local mismatch on the effective overdrive voltage of each stage. Figure 4.24 shows the resulting Monte Carlo distribution under the resized configuration, and Table 4.4 reports the corresponding per-corner mean and standard deviation for both oscillator paths.

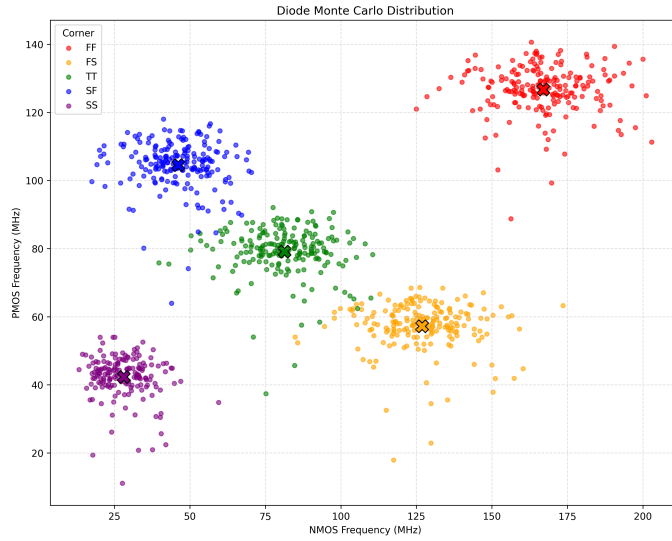


Figure 4.24: Monte Carlo distribution of the Diode Connected Load Ring Oscillator after transistor resizing.

Table 4.4: Monte Carlo mean and standard deviation of NMOS-path and PMOS-path oscillation frequency for the Diode Connected Load Ring Oscillator after transistor resizing, for each of the five process corners (200 runs per corner).

Corner	NMOS μ (MHz)	NMOS σ (MHz)	PMOS μ (MHz)	PMOS σ (MHz)
FF	167.052	13.915	126.745	7.094
FS	126.846	13.705	57.244	7.032
TT	81.305	12.539	79.149	7.152
SF	46.072	10.698	104.528	7.151
SS	28.067	7.526	42.287	6.257

The resized design shows a clear improvement in cluster separability. Visually, each corner now occupies a more compact and better-isolated region of the plane, and the previously overlapping TT/SF and TT/FS 3σ boundaries in Figure 4.23 are still in contact, but less in Figure 4.24. Quantitatively, Table 4.4 shows that the coefficient of variation (σ/μ) on the NMOS path improves from corner to corner but remains corner-dependent: it is smallest at FF ($13.915/167.052 \approx 8.3\%$)

and largest at SS ($7.526/28.067 \approx 26.8\%$), reflecting the fact that absolute frequency – and therefore absolute timing margin – shrinks faster than the underlying mismatch-induced jitter as the corner slows down. The same pattern holds on the PMOS path, where the SS corner again shows the highest relative spread ($6.257/42.287 \approx 14.8\%$) against $7.094/126.745 \approx 5.6\%$ at FF.

This asymmetry is the most important remaining weakness of the design. The 3σ envelope of the SS corner (NMOS: 28.07 ± 22.58 MHz, i.e. approximately 5.5–50.6 MHz) sits closer, in relative terms, to its nearest neighbouring cluster than the FF 3σ envelope does to its neighbours, because the SS corner combines the lowest mean frequency with a standard deviation that, while smaller in absolute terms than at FF, is large relative to that mean. On the slow side, the SS corner is therefore the one most exposed to encroaching on the noise floor of the frequency counter in the readout chain, and its 3σ tail on the NMOS axis alone would overlap the 3σ tail of the neighbouring SF cluster if classification relied on a single axis. Joint two-dimensional separation in the (F_{NMOS}, F_{PMOS}) plane keeps the five corners distinguishable at 3σ even so, but the margin at the slow corners is markedly tighter than at the fast corners. Practically, this means the resizing exercise improved the global picture without fully equalizing robustness across corners: the design is now significantly more reliable than the initial sizing, but a further iteration – for example, corner-dependent sizing, or a non-uniform stage count between the NMOS and PMOS rings – could still narrow the relative 3σ spread at the slow corners specifically, rather than uniformly across all five. Given the time constraints of this work, this is identified as a direction for future refinement rather than a flaw requiring re-design at this stage, since the 3σ envelopes of all five corners remain jointly non-overlapping in the two-dimensional plane, which is sufficient for reliable classification by the FSM comparison logic under the assumptions of this work.

4.2 Voltage Sensor

This section presents the complete simulation characterization of the voltage sensor subsystem, proceeding from the foundational reference circuit that underpins the entire measurement chain through to the final encoded digital output and its comparison with the source architecture of Chen et al. [4]. The analysis is structured in a deliberate chronological order that mirrors the design flow: the reference voltage is established first, the testbench infrastructure is described second, the nominal delay behaviour is characterised third, the stage interval optimization is derived fourth through a systematic progression of three design iterations, the environmental robustness under process and temperature variation is examined fifth, and the final simulation results and tabular comparison close the section.

4.2.1 Foundational Baseline: The Bandgap Reference

Before any voltage sensor measurement can be interpreted, the stability of the voltage reference that biases the current-starved delay chain must be rigorously established. The bandgap reference (BGR) circuit employed in this work is the topology proposed by Moisel et al. [29], which was presented in full schematic detail in Chapter 3. To avoid redundancy, the schematic of the BGR is not reproduced here; instead, this subsection reports the post-sizing simulation results that validate the reference output quality under supply and temperature variations, thereby establishing the performance baseline upon which all subsequent voltage sensor measurements depend.

4.2.1.1 Supply Voltage Sensitivity

Figure 4.25 shows the evolution of V_{ref} as V_{DD} is swept across the full operating range of 0.65 V to 0.95 V at the nominal conditions of $T = 25^\circ\text{C}$ and the TT process corner. The reference voltage rises from approximately 470 mV at $V_{DD} = 0.65$ V to approximately 484 mV at $V_{DD} = 0.95$ V, yielding a total peak-to-peak variation of 14 mV across the entire supply range. Although V_{ref} is not perfectly flat, this 14 mV spread represents a variation of approximately 2.9% relative to the nominal output of 480 mV at $V_{DD} = 0.8$ V, which remains well within the tolerance budget required for correct operation of the current-starved delay chain. The line regulation figure of merit, defined as the change in V_{ref} per unit change in V_{DD} , is computed over the operating window as follows:

$$\text{Line Regulation} = \frac{\Delta V_{ref}}{\Delta V_{DD}} = \frac{14 \text{ mV}}{300 \text{ mV}} \approx 46.7 \text{ mV/V} \quad (4.1)$$

The concave shape of the curve indicates that the sensitivity of V_{ref} to supply perturbations is highest at lower supply voltages, where the headroom available to the internal current mirror stack is most constrained, and progressively diminishes as V_{DD} rises above 0.8 V and the circuit enters a more comfortable operating regime.

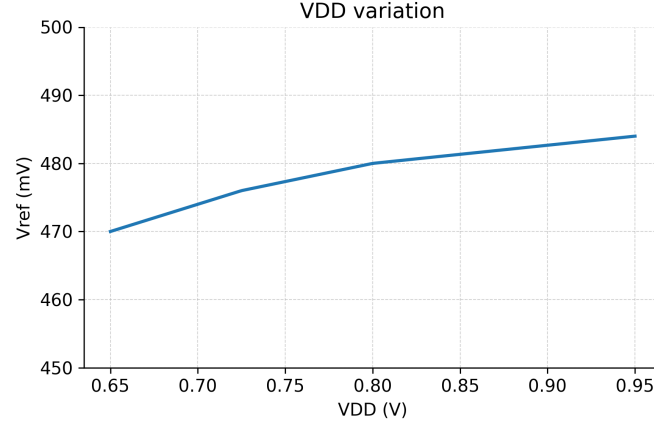


Figure 4.25: Simulated V_{ref} as a function of V_{DD} for the sized bandgap reference [29] (TT corner, 25°C).

4.2.1.2 Temperature Coefficient

Figure 4.26 presents the temperature sweep of V_{ref} from -40°C to 120°C at the nominal supply of $V_{DD} = 0.8\text{ V}$ and the TT process corner. The reference voltage remains remarkably stable across this 160°C span, exhibiting a total peak-to-peak variation of only 3 mV, from a minimum of approximately 479 mV near -40°C to a maximum of approximately 482 mV near 120°C . The temperature coefficient (TC) of the reference, computed over the full temperature range, is given by:

$$TC = \frac{\Delta V_{ref}}{V_{ref,nom} \cdot \Delta T} = \frac{3\text{ mV}}{480\text{ mV} \times 160^{\circ}\text{C}} \approx 39.04\text{ ppm}/^{\circ}\text{C} \quad (4.2)$$

This result confirms that the PTAT and CTAT current branches within the bandgap core of Moisello et al. [29] are well balanced after sizing, with their opposing temperature slopes cancelling to first order over the target operating range. Unlike the classical BJT-based bandgap topology where the PTAT component originates from the base-emitter junction voltage V_{BE} , the PTAT branch in this implementation is governed by the gate-to-source voltage V_{GS} of diode-connected NMOS transistors (N1, N2) biased in the subthreshold regime, where the effective thermal voltage is scaled by the subthreshold slope factor n , giving nV_T in place of V_T . The residual curvature visible in Figure 4.26, which manifests as the slight upward drift above 60°C , is attributable to the higher-order temperature dependence of $V_{GS}(T)$ itself. Specifically, both the threshold voltage $V_{th}(T)$ and the carrier mobility $\mu(T) \propto T^{-3/2}$ exhibit nonlinear temperature characteristics that are not cancelled by the first-order PTAT-CTAT compensation scheme of the Moisello et al. [29] architecture. Nevertheless, a temperature coefficient of 39.04 ppm/ $^{\circ}\text{C}$ is a strong result for a 16 nm FinFET implementation operating at a sub-1 V supply, and it establishes V_{ref} as a sufficiently stable bias source for the voltage sensor delay chain across the full intended operating envelope.

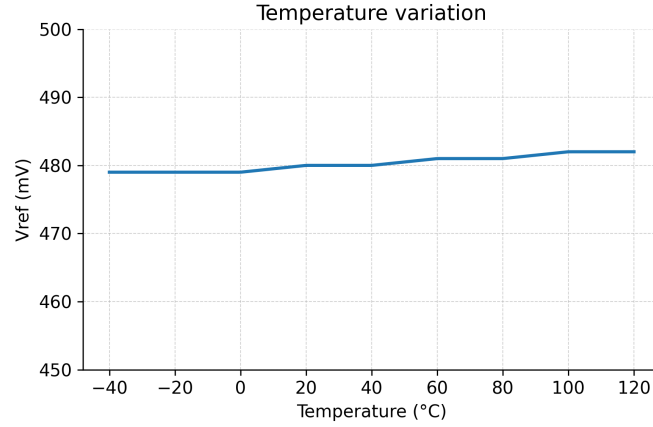


Figure 4.26: Simulated V_{ref} as a function of temperature for the sized bandgap reference [29] ($V_{DD} = 0.8$ V, TT corner).

4.2.1.3 PTAT and CTAT Current Verification

To further validate the internal operating mechanism of the sized reference, Figure 4.27 plots the three key currents within the bandgap core of Moisello et al. [29] as a function of temperature: the complementary-to-absolute-temperature component I_{CTAT} , the proportional-to-absolute-temperature component I_{PTAT} , and the total summed output current I_2 . The CTAT current decreases monotonically from approximately $3.5 \mu\text{A}$ at -40°C to approximately $2.6 \mu\text{A}$ at 120°C , while the PTAT current increases from approximately $1.9 \mu\text{A}$ to approximately $3.1 \mu\text{A}$ over the same range. The two branches cross at approximately 80°C , which is the equilibrium point at which neither component dominates. Their sum, I_2 , rises gently from approximately $5.45 \mu\text{A}$ to approximately $5.72 \mu\text{A}$, a variation of only $0.27 \mu\text{A}$, confirming that the partial cancellation between the two branches is effective across the full temperature range. The mild positive slope of I_2 is consistent with the slight upward drift observed in V_{ref} at high temperatures in Figure 4.26, and reflects the residual higher-order curvature of $V_{GS}(T)$ noted above. Taken together, these results confirm that the sizing procedure has correctly established the intended PTAT-CTAT compensation mechanism, and that the Moisello et al. [29] bandgap reference delivers a stable and well-characterised V_{ref} suitable for use as the biasing source of the voltage sensor current-starved delay chain.

It should be noted that all of the bandgap reference results presented in this subsection, and by extension the voltage sensor results depend on the V_{ref} through the remainder of this section, were obtained using an ideal operational amplifier model for the error amplifier at the core of the Moisello et al. [29] topology, rather than a transistor-level implementation. An ideal op-amp model assumes infinite open-loop gain, zero input-referred offset voltage, and unlimited bandwidth, none of which hold in a real, fabricated amplifier. A physically implemented op-amp would introduce a finite open-loop gain that degrades the line and load regulation reported in Section 4.2.1.1 and Section 4.2.1.2, a non-zero input-referred offset voltage that directly perturbs V_{ref} and would need to be absorbed into the resistor-ratio sizing, and a finite bandwidth and slew rate that would slow

the loop's transient response and could affect the settling behaviour assumed throughout the delay-chain characterization from Section 4.2.3 onward. The reported V_{ref} stability figures (14 mV line regulation, 39.04 ppm/°C temperature coefficient) should therefore be interpreted as a best-case bound rather than a silicon-representative result; replacing the ideal op-amp with a fully sized transistor-level implementation is left as future work, and would likely yield somewhat degraded, though not qualitatively different, reference performance.

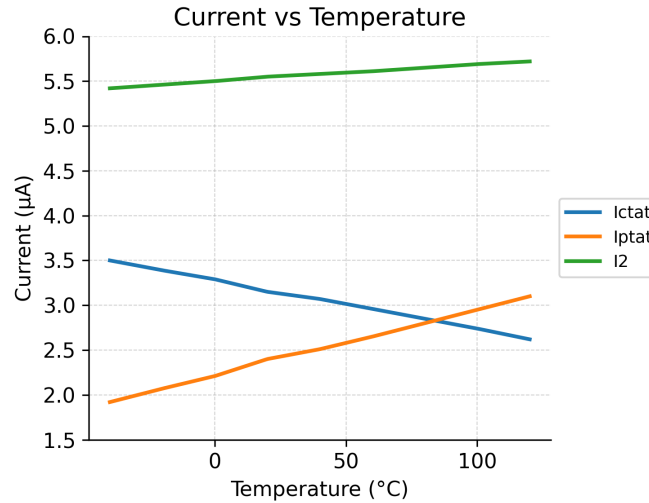


Figure 4.27: Simulated PTAT, CTAT, and total output current I_2 of the sized Moisel et al. [29] bandgap reference as a function of temperature ($V_{DD} = 0.8$ V, TT corner).

4.2.2 Testbench Infrastructure and Circuit Components

The simulation environment constructed to evaluate the voltage sensor is shown in Figure 4.28. The testbench comprises eight cascaded current-starved inverter stages, labelled CSS_1 through CSS_8, connected in a serial delay chain that is jointly powered by V_{DD} and biased by the reference voltage V_{ref} sourced from the Moisel et al. [29] bandgap reference block. A clock generator of 71.4 MHz drives the chain at its input, and the signal at each stage output node is routed to a D-type flip-flop (FF1 through FF6). Each flip-flop captures the logic state of its respective stage output relative to a delayed version of the input clock, and its output is then passed through an XOR gate with a neighbouring flip-flop output, producing the final encoded sensor bits $V[0]$ through $V[4]$. The delay block interposed between the clock source and the flip-flop clock inputs introduces a fixed offset, ensuring that the sampling edge arrives after the slowest anticipated propagation transition through the chain. The testbench architecture thereby converts the analog quantity of interest, namely the supply-voltage-dependent propagation delay through the current-starved stage chain, into a thermometer-coded digital word that directly encodes V_{DD} .

The modular building block forming the core of this delay chain is the current-starved inverter stage, whose schematic is shown in Figure 4.29. Each CSS stage is implemented as a standard CMOS inverter composed of a PMOS pull-up transistor and an NMOS pull-down transistor, with

an additional NMOS device connected in series with the pull-down path and controlled by V_{ref} . This third transistor acts as a current limiter: because its gate is held at the fixed bandgap reference voltage, it restricts the maximum current that can flow through the pull-down network to a value governed by V_{ref} , and not by V_{DD} directly. Crucially, as V_{DD} increases, the overdrive voltage across the main inverter transistors increases, which would normally accelerate the output transition; however, the current-starving device clamps the discharge current, so the net propagation delay through the stage still exhibits a strong monotonic dependence on V_{DD} . It is this voltage-to-delay transduction mechanism that forms the physical basis for supply voltage sensing. All eight stages in the testbench are instantiations of this same schematic, designed in the 16 nm FinFET technology node [4].

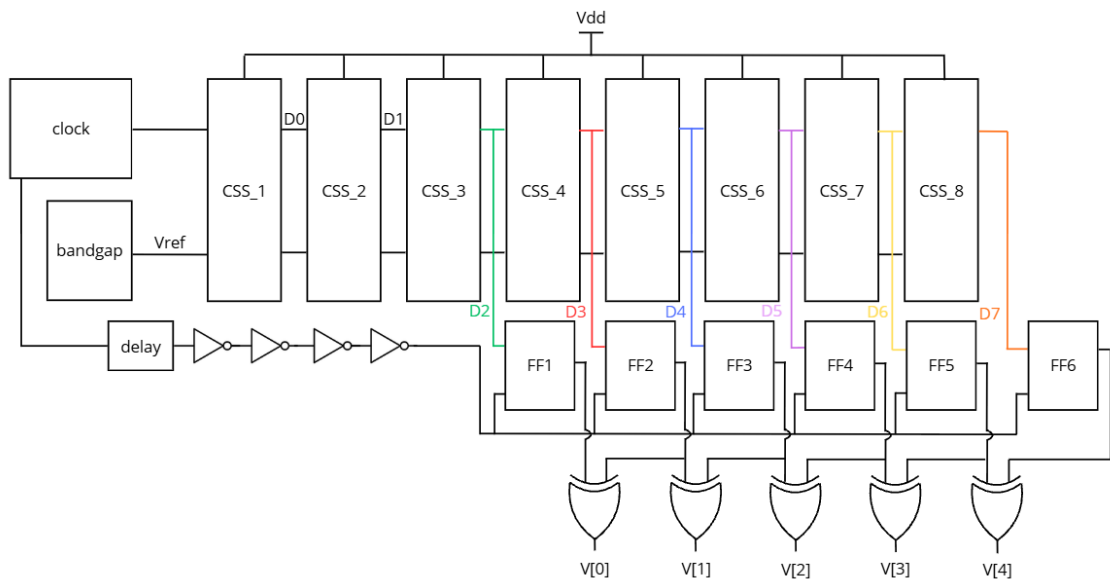


Figure 4.28: Voltage sensor testbench schematic.

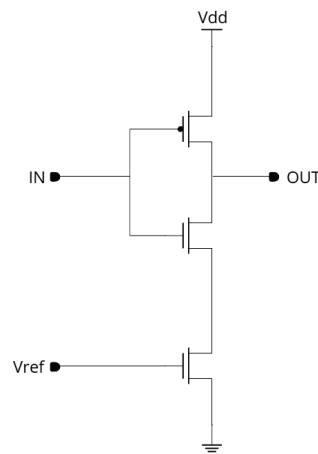


Figure 4.29: Schematic of a single current-starved inverter (CSS) stage.

4.2.3 Nominal Performance and Chain stabilization

Before the inter-stage delay behaviour under varying supply conditions can be meaningfully interpreted, it is necessary to characterise the transient initialization of the chain at the nominal operating point of $V_{DD} = 0.8 \text{ V}$, $T = 25^\circ\text{C}$, and the Typical-Typical (TT) process corner. The propagation delay measured from stage to stage is not immediately uniform across the full chain; instead, the earliest stages exhibit a markedly different delay characteristic compared to those deeper in the structure, and this behaviour must be understood before any measurement nodes are designated.

Figure 4.30 plots the per-stage propagation delay as a function of stage index for the 8-stage configuration under nominal conditions. The first transition, from D_0 to D_1 , yields a delay of approximately 0.046 ns, which is strikingly lower than the delays observed at all subsequent stages. From D_1 onward, the per-stage delay rises sharply and settles to a stable plateau in the vicinity of 0.110 ns, remaining essentially constant across transitions D_2 through D_6 before showing a minor deviation at D_7 .

This non-uniform initialization at D_0 is a direct consequence of the injection and driving influence of the input clock on the very first stage. At D_0 , the CSS stage is not propagating a signal that has already traversed the current-starving network under steady-state conditions; rather, it is directly responding to the hard-driven clock edge. The current-starving mechanism has not yet established full control over the discharge current at this earliest node, which results in an anomalously fast and therefore unrepresentative transition. The ring oscillator, in effect, requires a brief stabilization interval during which the dominant influence transitions from the injected clock front to the biasing condition imposed by V_{ref} . By the second stage output D_1 , this injection effect has largely dissipated, and by D_2 the chain has fully stabilised, producing delay steps that are genuinely governed by the CSS biasing condition and not by the residual driving strength of the clock source. For this reason, the design deliberately avoids using outputs D_0 and D_1 as sensor measurement nodes. All five output bits $V[0]$ through $V[4]$ are derived exclusively from stages D_2 through D_7 , where delay uniformity guarantees that each step faithfully reflects a fixed, supply-determined time increment rather than an artefact of the clock injection front.

The nominal single-stage delay at the stable plateau is:

$$t_p \approx 0.110 \text{ ns} \quad (\text{stable region, } D_2 \text{ to } D_6, \text{ TT, } V_{DD} = 0.8 \text{ V, } 25^\circ\text{C}) \quad (4.3)$$

This value serves as the reference delay increment from which the inter-stage spacing in each design iteration is calculated, and it provides the upper bound on the quantization resolution achievable per encoded voltage step.

4.2.4 Sequential Stage-Delay optimization

With the per-stage delay behaviour established, the central architectural challenge becomes one of selecting the correct number of CSS stages to interpose between successive monitored output nodes D_x . Figure 4.31 presents the time-domain waveforms of CLK and outputs D_1 through D_7 at

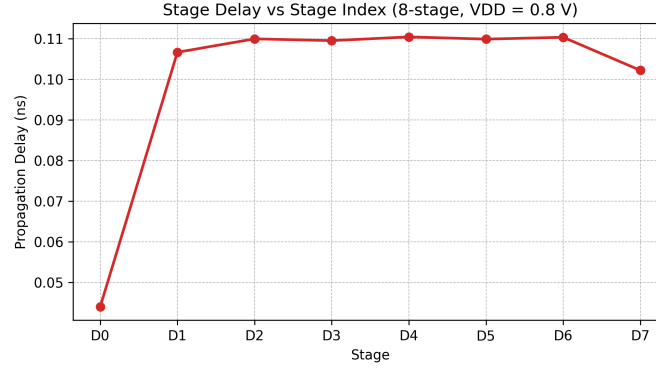


Figure 4.30: Per-stage propagation delay versus stage index for the 8-stage configuration at nominal conditions ($V_{DD} = 0.8$ V, TT corner, 25°C).

the nominal operating point ($V_{DD} = 0.8$ V, TT corner, 25°C), illustrating the sequential progression of transitions through the chain and the inter-stage temporal spacing that the encoding logic must correctly resolve.

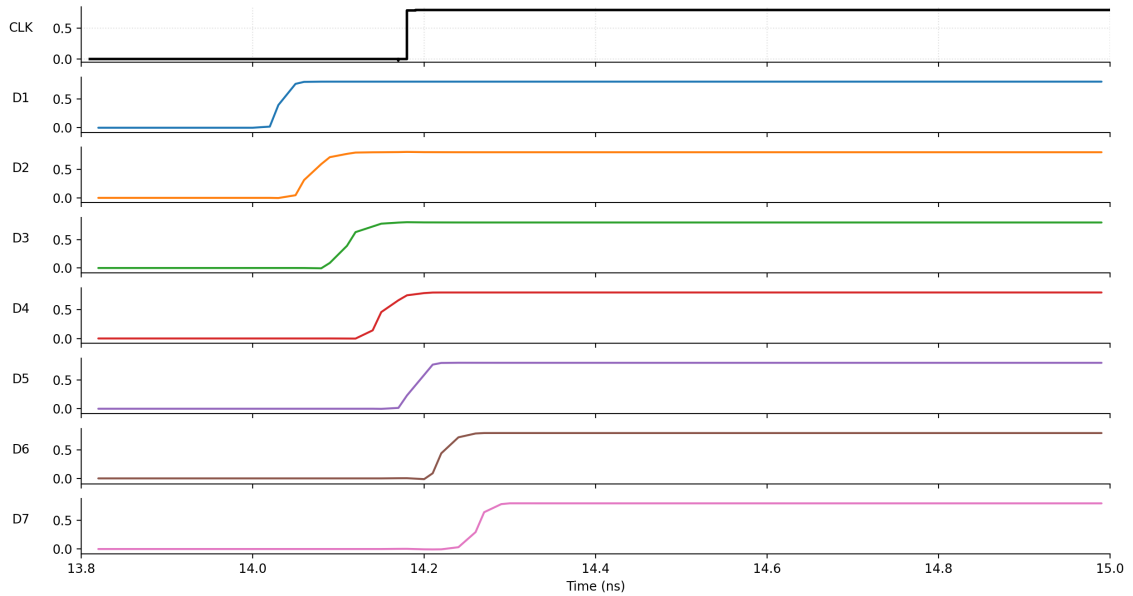


Figure 4.31: Voltage sensor output waveforms at nominal conditions ($V_{DD} = 0.8$ V, TT corner, 25°C).

The architecture adapted from Chen et al. [4] encodes V_{DD} into a digital word by comparing the logic states of the monitored outputs against a shared clock signal edge. For the encoding methodology to remain unambiguous across all simulated supply conditions, the clock edge must arrive at the flip-flop inputs at a time that lies strictly between the 50% crossing of consecutive output nodes for two distinct supply voltages.

A critical complication arises from the fact that the clock delay is itself a function of supply voltage, because the delay buffer chain driving the flip-flop clock inputs also propagates faster at higher V_{DD} . Let $t_{CLK,min}$ and $t_{CLK,max}$ denote, respectively, the earliest and latest 50% threshold

crossing of the clock signal across all simulated V_{DD} values. The resulting clock delay variance is therefore:

$$\Delta t_{CLK} = t_{CLK,max} - t_{CLK,min} \quad (4.4)$$

For unambiguous codification to be maintained across the full V_{DD} spectrum, this variance must be strictly smaller than the temporal interval separating the 50% threshold of output D_6 at $V_{DD} = 0.95$ V from the 50% threshold of output D_5 at $V_{DD} = 0.8$ V. Denoting these two crossing times as $t_{D_6}^{0.95V}$ and $t_{D_5}^{0.8V}$, the required timing constraint is:

$$\Delta t_{CLK} < |t_{D_6}^{0.95V} - t_{D_5}^{0.8V}| \quad (4.5)$$

If the inter-stage interval is too narrow, the time separation between the transitions of adjacent outputs at different supply voltages becomes comparable to Δt_{CLK} , rendering the encoding ambiguous and making correct codification impossible. Conversely, if the interval is too wide, the number of resolved voltage bins across the target V_{DD} range decreases, degrading the quantization resolution of the sensor. Three distinct design iterations were evaluated, using stage intervals of 2, 4, and 8 CSS stages between each pair of monitored outputs D_x .

4.2.4.1 2-Stage Interval Configuration

The evaluation begins with the 2-stage interval configuration, which is the same stage spacing employed in the original 65 nm implementation of Chen et al. [4]. The transient waveforms of the CLK signal and outputs D_1 through D_7 for all five tested supply voltages (650 mV, 725 mV, 800 mV, 875 mV, and 950 mV) are shown in Figure 4.32.

Inspection of Figure 4.32 reveals that the 2-stage interval configuration fails the timing constraint of Equation (4.5), as indicated later in Table 3.3. The output transitions at adjacent nodes are separated by such a small time window that the supply-voltage-induced variation in clock delay Δt_{CLK} exceeds the target interval. The CLK waveform in this configuration crosses the 50% threshold at a time that, depending on V_{DD} , is indistinguishable from the transition timing at two or more D_x outputs simultaneously, making correct encoding impossible. It is important to note that the 2-stage spacing was adequate in the original 65 nm bulk CMOS implementation of Chen et al. [4] because the absolute stage delays in that technology were considerably larger, providing a more generous encoding window. Under the compressed propagation delays characteristic of the 16 nm FinFET node, the same 2-stage interval is fundamentally insufficient.

4.2.4.2 4-Stage Interval Configuration

Having established that the 2-stage configuration is insufficient, the evaluation progresses to a 4-stage interval as a systematic intermediate expansion step. The corresponding transient waveforms are presented in Figure 4.33.

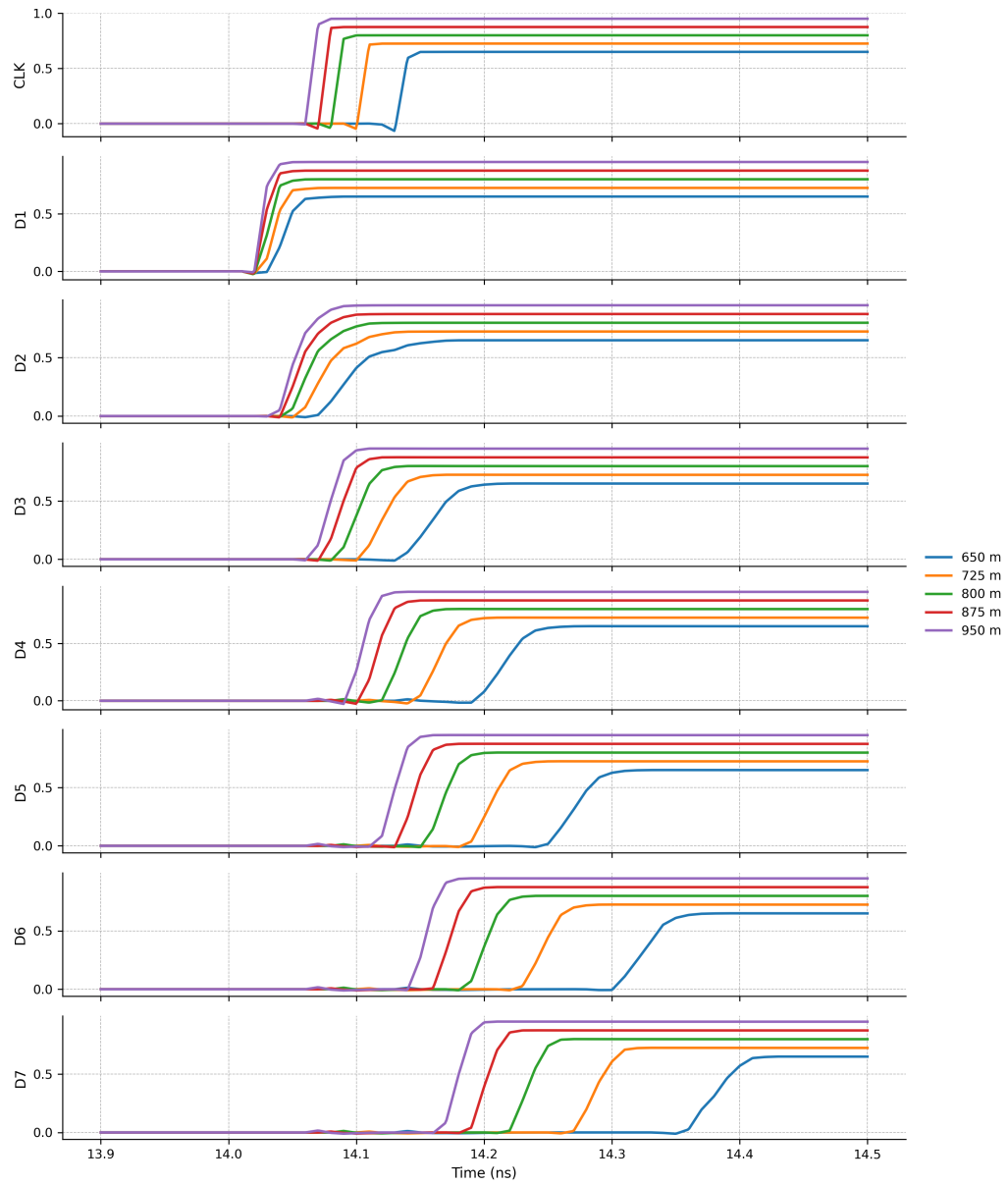


Figure 4.32: Transient output waveforms for the 2-stage interval configuration. CLK and outputs D_1 – D_7 are shown for $V_{DD} \in \{650, 725, 800, 875, 950\}$ mV.

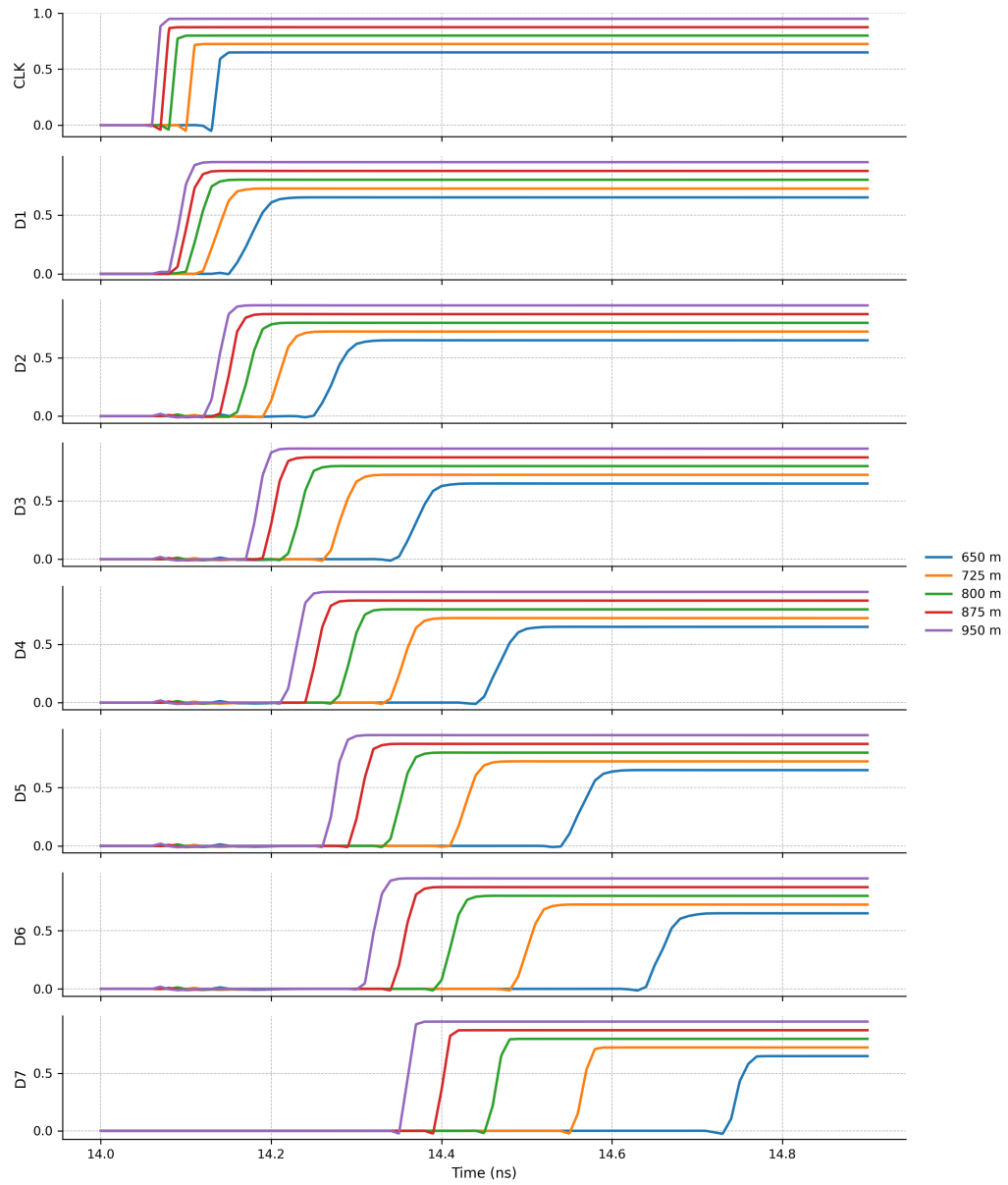


Figure 4.33: Transient output waveforms for the 4-stage interval configuration. CLK and outputs D_1 – D_7 are shown for $V_{DD} \in \{650, 725, 800, 875, 950\}$ mV.

The 4-stage interval configuration produces a larger inter-output temporal interval than the 2-stage case, reflecting the additional propagation delay accumulated over the doubled stage count. However, examination of Figure 4.33 shows that the margin remains insufficient: the spread of the CLK crossing across the simulated V_{DD} range still encroaches upon the required separation window defined by Equation (4.5). While the 4-stage interval reduces the severity of the overlap problem relative to the 2-stage case, it does not eliminate it. Across the extremes of the supply voltage range, the CLK edge still cannot be guaranteed to land unambiguously within the encoding window defined by adjacent D_x transitions, meaning that certain V_{DD} levels would be incorrectly codified under realistic operating conditions.

4.2.4.3 8-Stage Interval Configuration

Proceeding from the conclusions of the 4-stage evaluation, the design is advanced to an 8-stage interval, which doubles the inter-output delay spacing once more. The transient waveforms for this configuration are presented in Figure 4.34.

The 8-stage interval configuration provides a substantially expanded inter-output time spacing relative to both prior iterations. The simulation data confirm that under this configuration the clock delay range δt_{CLK} satisfies Equation 4.5 across all five tested VDD levels, meaning the clock edge consistently falls within the correct timing window to unambiguously capture the intended thermometer-coded output state. Expressed formally, the verification condition is satisfied when:

$$\Delta t_{CLK} < |t_{D6}^{0.95V} - t_{D5}^{0.8V}| \quad (4.6)$$

which the 8-stage simulation results confirm to hold with a demonstrably safe margin across the entire operating range. The 8-stage configuration was therefore adopted as the final design. It is the only iteration among the three evaluated in which the timing constraint is unambiguously satisfied, and it is the configuration from which all subsequent non-ideality analyses are conducted.

The 8-stage interval configuration provides a substantially expanded inter-output time spacing relative to both prior iterations. The simulation data confirm that under this configuration the clock delay range Δt_{CLK} satisfies Equation (4.5) across all five tested V_{DD} levels, meaning the clock edge consistently falls within the correct timing window to unambiguously capture the intended thermometer-coded output state. Expressed formally, the verification condition is satisfied when:

$$\Delta t_{CLK} < |t_{D6}^{0.95V} - t_{D5}^{0.8V}| \quad (4.7)$$

which the 8-stage simulation results confirm to hold with a demonstrably safe margin across the entire operating range. The 8-stage configuration was therefore adopted as the final design. It is the only iteration among the three evaluated in which the timing constraint is unambiguously satisfied, and it is the configuration from which all subsequent non-ideality analyses are conducted.

Table 4.5 makes this progression quantitatively explicit by directly tabulating the two quantities compared in Equation (4.5) for each of the three evaluated stage intervals: the clock delay variance Δt_{CLK} and the required encoding margin $|t_{D6}^{0.95V} - t_{D5}^{0.8V}|$.

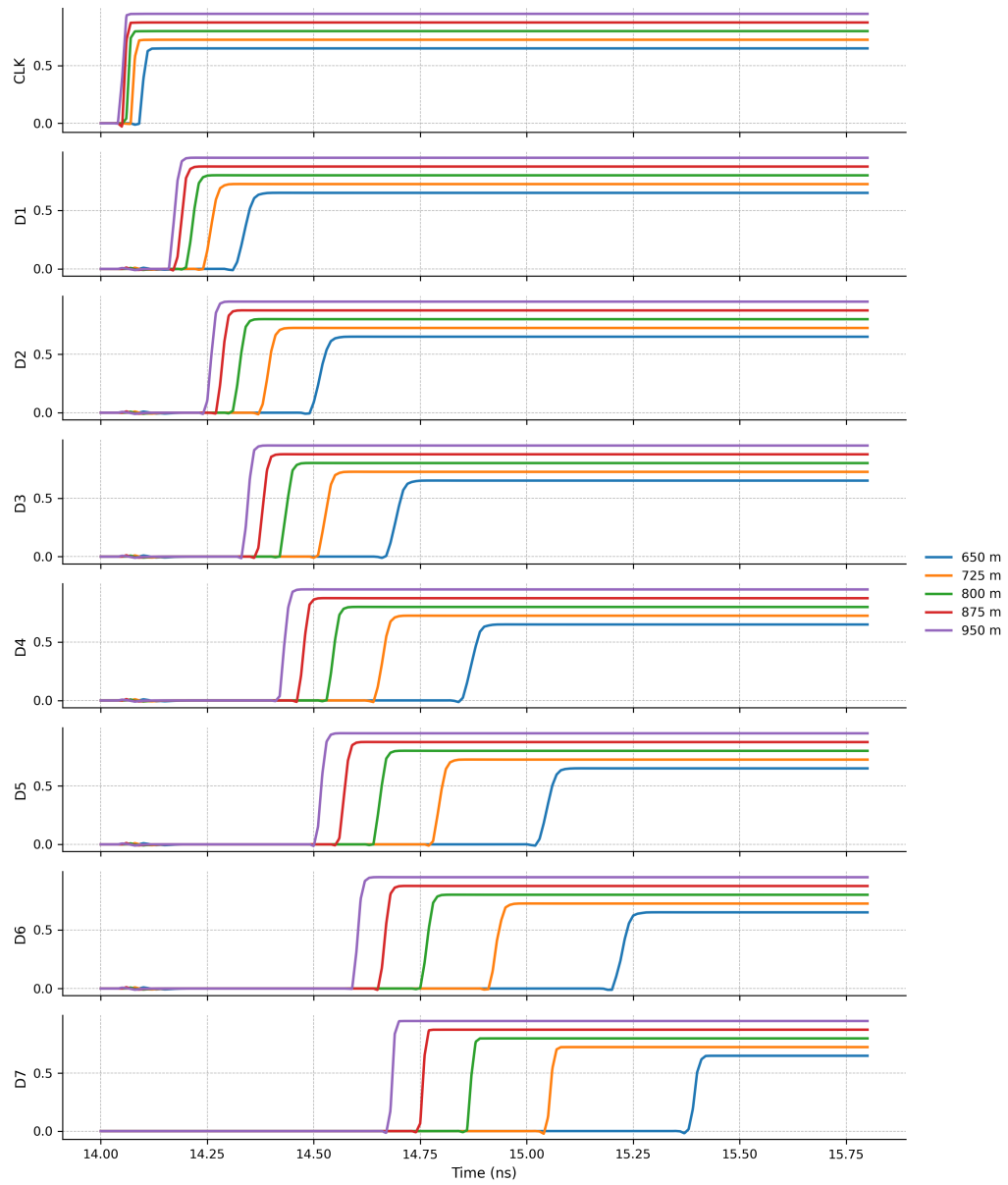


Figure 4.34: Transient output waveforms for the 8-stage interval configuration. CLK and outputs D_1 – D_7 are shown for $V_{DD} \in \{650, 725, 800, 875, 950\}$ mV.

Table 4.5: Clock delay variance Δt_{CLK} against the required encoding margin $|t_{D6}^{0.95V} - t_{D5}^{0.8V}|$ for each evaluated stage-interval configuration.

Stage Interval	Δt_{CLK}	$ t_{D6}^{0.95V} - t_{D5}^{0.8V} $
2	47.2 ps	11.5 ps
4	47.0 ps	33.3 ps
8	47.2 ps	65.5 ps

Because the clock delay buffer chain that sets Δt_{CLK} is not altered between the three iterations – only the inter-stage spacing of the CSS chain itself – Δt_{CLK} remains essentially constant at approximately 47 ps across all three rows, with the small variation between 47.0 and 47.2 ps attributable to simulation noise rather than any systematic trend. The encoding margin, by contrast, scales directly with the stage interval, since a wider spacing between monitored nodes proportionally increases the inter-stage propagation delay that separates D6 and D5. At the 2-stage interval, the margin of 11.5 ps is more than $4\times$ smaller than Δt_{CLK} , decisively violating Equation (4.5); at the 4-stage interval, the margin grows to 33.3 ps but still falls short of the 47.0 ps clock delay variance, so the constraint remains violated, consistent with the qualitative observations in Section 4.2.4.1. Only at the 8-stage interval does the margin of 65.5 ps exceed $\Delta t_{CLK} = 47.2$ ps, satisfying Equation (4.5) with a margin of approximately 18.3 ps, which quantitatively confirms the safe margin referenced above and justifies the selection of the 8-stage interval as the final design.

4.2.4.4 Delay Block optimization and Its Inherent Limitation

Once the 8-stage interval was confirmed as the viable configuration, a secondary design effort was directed at tuning the delay block that is interposed between the clock source and the flip-flop clock inputs in the testbench (see Figure 4.28). This block, implemented as a chain of standard inverters, introduces the fixed temporal offset that ensures the sampling clock edge arrives after the slowest anticipated propagation transition through the CSS chain, thereby guaranteeing a valid setup window for all five flip-flops across the full V_{DD} range.

The sizing of this inverter chain, however, exposes a fundamental tension inherent to the architecture. Because the delay block is powered by the same supply rail as the rest of the circuit, its propagation delay is itself a function of V_{DD} : adding inverters to the delay block extends the clock offset but simultaneously increases Δt_{CLK} , the clock delay variance across the supply range, since each additional inverter contributes a further V_{DD} -dependent delay increment. Conversely, reducing the number of inverters in the delay block decreases Δt_{CLK} but may cause the sampling edge to arrive before the chain transition has completed at the lowest supply voltages, resulting in a metastable or incorrect flip-flop capture. This trade-off means that the delay block length cannot be optimised independently of the inter-stage interval: any change to the inverter count in the clock path directly modulates the clock delay spread and therefore the margin available to satisfy Equation (4.5). This coupling between the clock delay block and the encoding constraint represents a notable limitation of the present implementation. In a more robust future design, a

supply-independent delay element, for instance one based on a current-starved topology biased by the same V_{ref} as the CSS chain, could be employed to decouple the clock offset from supply variation and thereby eliminate this source of measurement uncertainty.

4.2.5 Environmental Robustness Under Temperature and Process Corner Variation

Having established the 8-stage interval as the design-compliant configuration, the analysis now turns to the robustness of the circuit under realistic manufacturing and operational variations. Two separate sweep campaigns were conducted: one spanning nine temperature points from -40°C to 120°C at the TT corner and $V_{DD} = 0.8\text{ V}$, and a second spanning the five standard process corners (FF, TT, FS, SF, SS) at 25°C and $V_{DD} = 0.8\text{ V}$.

4.2.5.1 Temperature Variation

Figure 4.35 plots the per-stage propagation delay, expressed in picoseconds, as a function of stage transition across the full temperature range. Several important observations emerge from this characterization.

First, the delay at the very first transition (D_0 to D_1) is lower than the delay at subsequent transitions for all temperature values, confirming that the initialization phenomenon identified in Section 4.2.3 persists uniformly across temperature. This establishes the phenomenon as a structural property of the chain arising from clock injection rather than a consequence of specific biasing at a single operating point.

Second, a notable transient overshoot is visible at the D_2 -to- D_3 transition for all temperatures: the per-stage delay rises above its final plateau value before settling at the D_3 -to- D_4 transition. This behaviour is consistent with residual injection-induced ringing propagating into the third inter-stage measurement window; by D_3 -to- D_4 the chain has fully equilibrated, and the delay remains effectively constant through D_6 before showing a slight reduction at D_6 -to- D_7 .

Third, and most significantly for sensor functionality, the per-stage delay exhibits a clear and monotonic inverse relationship with temperature across the stabilised region (D_3 to D_7). At -40°C the plateau delay is approximately 187 ps, while at 120°C it drops to approximately 155 ps. This reflects the well-understood temperature dependence of FinFET drain current: at lower temperatures, carrier mobility is higher, producing faster drive currents and therefore shorter propagation delays [4]. Because the voltage sensor reads out V_{DD} at a fixed moment in time relative to a common clock edge, this temperature-induced shift in absolute delay represents a source of cross-sensitivity that must be accounted for when the full PVT sensor system combines the voltage sensor output with temperature compensation data from the temperature sensor block.

4.2.5.2 Process Corner Variation

Figure 4.36 plots the per-stage propagation delay for the five process corners (FF, TT, FS, SF, and SS) at $V_{DD} = 0.8\text{ V}$ and 25°C . The spread in delay across corners is substantial. The FF corner

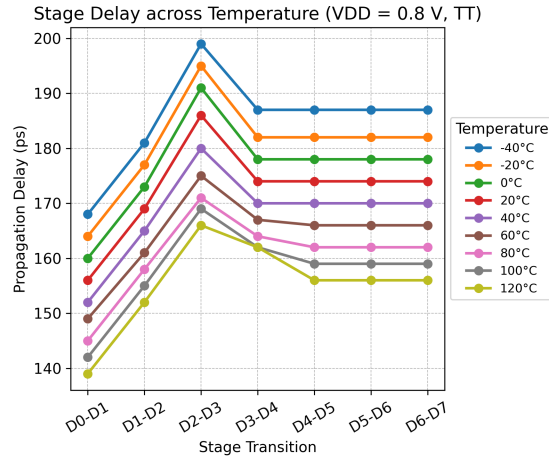


Figure 4.35: Per-stage propagation delay versus stage transition across temperature ($V_{DD} = 0.8$ V, TT corner).

produces the shortest stable per-stage delay, settling at approximately 113 ps in the stabilised region (D_3 onward), while the SS corner produces the longest, plateauing at roughly 305 ps. The TT corner occupies an intermediate position at approximately 173 ps. The mixed corners FS and SF cluster close to the TT value, reaching plateau delays of approximately 188 ps and 187 ps respectively, reflecting the partial compensation that occurs when one carrier type is fast and the other is slow within the same device.

The characteristic initialization signature is again present at D_0 -to- D_1 for all corners, with every corner exhibiting a depressed delay at the first transition that rapidly climbs toward its respective plateau by D_2 -to- D_3 . From D_3 onward, all five corner trajectories are essentially flat, confirming that the plateau region of the chain is stable and free of the injection transient.

The corner-induced delay range of approximately 192 ps (from 113 ps to 305 ps) across the full process spread is considerably larger than the temperature-induced range of approximately 32 ps, underscoring that process variation is the dominant non-ideality in this voltage sensor design. The successful resolution of V_{DD} under the worst-case process deviation therefore depends critically on the system-level correction provided by the process sensor block, whose output $P[3:0]$ is used in the full PVT compensator to adjust the voltage sensor digital code accordingly [4].

4.2.6 Simulation Results and Tabular Comparison

As established in Section 4.2.4 and confirmed by the nominal output waveforms presented in Figure 4.31, the 8-stage configuration produces a sufficiently wide and uniform inter-stage temporal spacing to ensure that the shared clock edge lands unambiguously within the correct encoding window across the full V_{DD} range. The complete simulation results are presented as a two-stage readout: first, the raw thermometer-coded states captured at the monitored D-stage outputs, and second, the encoded output produced after those states pass through the XOR network.

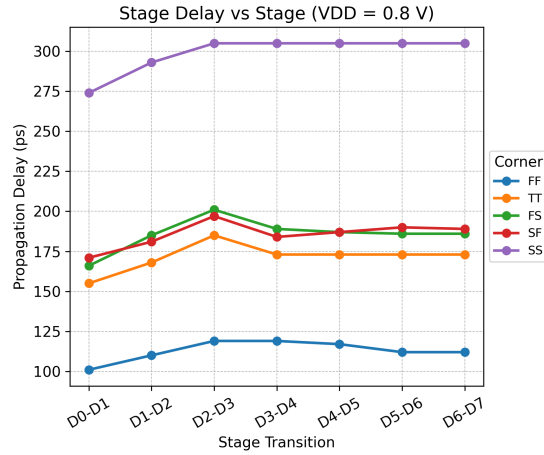


Figure 4.36: Per-stage propagation delay versus stage transition across process corners ($V_{DD} = 0.8$ V, 25°C).

Table 4.6 reports the logic state sampled at each monitored output node D_2 through D_7 for the five tested supply voltages, following the delay block tuning that correctly positions the clock sampling edge. The table exhibits the expected thermometer-code behaviour: at $V_{DD} = 0.65$ V, only D_2 has transitioned by the time the clock edge arrives, reflecting the longest propagation delay at the lowest supply, whereas at each successive supply voltage an additional output node has completed its transition, advancing the thermometer boundary by one position. At $V_{DD} = 0.95$ V, outputs D_2 through D_6 are all asserted while D_7 remains low, confirming that the clock edge is correctly positioned to capture five distinct and unambiguous supply voltage states across the full 300 mV operating window.

Table 4.6: Sampled logic states at monitored D-stage outputs after delay block optimization (8-stage configuration, TT corner, 25°C).

V_{in} (V)	D-Stage Output					
	D_2	D_3	D_4	D_5	D_6	D_7
0.65	1	0	0	0	0	0
0.725	1	1	0	0	0	0
0.8	1	1	1	0	0	0
0.875	1	1	1	1	0	0
0.95	1	1	1	1	1	0

Table 4.7 presents the final encoded output word $V[4 : 0]$, produced by passing the adjacent D-stage pairs through the XOR gate network. Because each XOR gate asserts a logic high only when its two inputs differ, the encoded output bit transitions from zero to one precisely at the thermometer boundary, identifying the single supply voltage bin in which V_{DD} falls. This follows the tabular format of Chen et al. [4] and maps each V_{DD} setting to its corresponding encoded output.

Table 4.7: Voltage sensor encoded digital output after XOR network (8-stage configuration, TT corner, 25 °C).

V_{in} (V)	Digital Code				
	$V[0]$	$V[1]$	$V[2]$	$V[3]$	$V[4]$
0.65	1	0	0	0	0
0.725	0	1	0	0	0
0.8	0	0	1	0	0
0.875	0	0	0	1	0
0.95	0	0	0	0	1

The encoding scheme is strictly one-hot: for any given V_{DD} level, exactly one output bit is asserted high and all others remain low. The quantization step between adjacent voltage levels is 75 mV, determined by the product of the per-stage delay increment introduced by each CSS stage and the supply-voltage sensitivity of that delay. The encoded range spans from 650 mV to 950 mV, covering a 300 mV window centred near the nominal V_{DD} of 0.8 V, which is the operating range of interest for the target 16 nm FinFET design.

4.3 Temperature Sensor

This section presents a systematic simulation-based evaluation of the proposed temperature sensor sub-system implemented in 16 nm FinFET technology. The analysis proceeds through four logical stages: the architectural motivation and PTAT source selection, the testbench infrastructure assembled to evaluate each candidate, the full process-corner temperature sweep results, and finally a supply-voltage sensitivity study that drives the conclusive topology selection.

4.3.1 PTAT Source Generation and BGR optimization

The fundamental operating principle of the temperature sensor architecture rests on the generation of a high-linearity Proportional-to-Absolute-Temperature (PTAT) voltage or current signal. In the current-controlled-oscillator (CCO) based temperature sensing scheme adopted from Lempel et al. [44], the oscillation frequency F_{out} is set by a PTAT bias current $I_{\text{PTAT}}(T)$ charging a fixed oscillator capacitance C_{ro} against the supply-like control voltage V_{ccro} , so that the overall transfer characteristic can be expressed, to a first-order approximation, as

$$F_{\text{out}}(T) \approx \frac{I_{\text{PTAT}}(T)}{C_{\text{ro}} \cdot V_{\text{ccro}}} = \frac{kT \ln(N)}{qRC_{\text{ro}} V_{\text{ccro}}},$$

where k is the Boltzmann constant, q is the electron charge, R is the PTAT bias resistor, N is the transistor sizing ratio that sets the PTAT current density, and V_{ccro} is the supply-like control voltage derived from the PTAT source. The linear rise of $I_{\text{PTAT}}(T)$ with absolute temperature T is the primary mechanism encoding thermal information into the output frequency. The linearity and supply independence of V_{ccro} therefore directly govern the sensor accuracy and the complexity of subsequent digital calibration.

In the course of this work, a structural modification was applied to a standard Bandgap Reference (BGR) topology proposed by Moisello et al. [29] with the specific objective of adapting its PTAT extraction node, V_{ptat} , to interface directly with the gate-driven PMOS pass transistor used in the ring-oscillator-based temperature sensor topology adopted from Lempel et al. [44].

The modified BGR schematic, illustrated in Figure 4.37, retains the core current-mode summation structure: an operational amplifier enforces the virtual short $V_a = V_b$, with V_a set by R1 and the diode-connected device N1, while V_b splits into two parallel paths feeding R2 and R3. The R2 path reproduces $V_a (= V_b)$ across R_2 , generating the CTAT current $I_{\text{CTAT}} = V_{f1}/R_2$, while the R3 path drives a second diode-connected device, N2, sized in a 1: N ratio relative to N1. Because N1 and N2 carry the same current but differ in size, their gate-source voltages differ by $\Delta V_f = V_T \ln(N)$, which appears across R_3 and produces the PTAT current $I_{\text{PTAT}} = \Delta V_f/R_3$, mirroring exactly the operating principle established for the core BGR topology in Chapter 3. As in the original architecture, the two currents are summed and mirrored into R_4 via the P3/N3 branch to produce the primary reference output V_{ref} , which remains unaffected by the PTAT extraction modification described below.

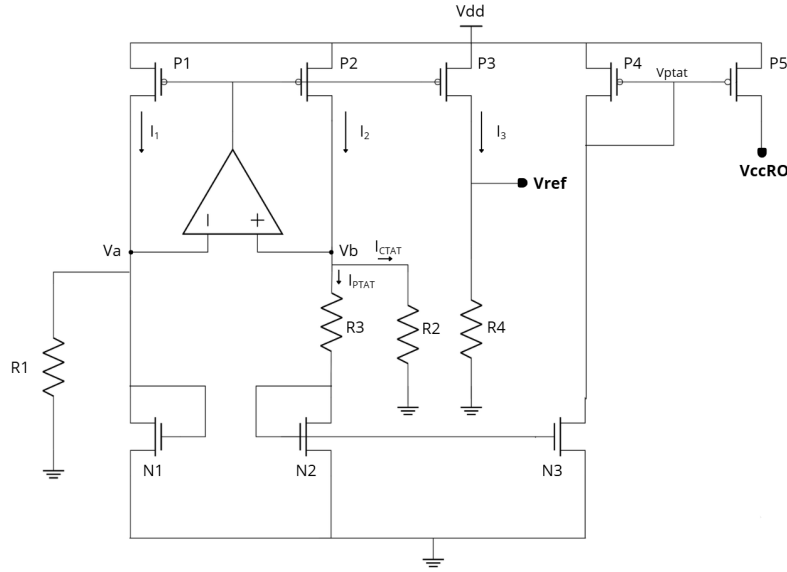


Figure 4.37: Modified Bandgap Reference schematic used as the upgraded PTAT source.

The original BGR topology presents V_{ptat} as a resistor-referenced voltage, which is poorly suited to driving the gate of the PMOS pass transistor that regulates V_{ccRO} in the ring-oscillator sensor core, since that interface requires a low-impedance, diode-referenced bias rather than a high-impedance resistive node. To meet this requirement, the modification reuses the gate node of the diode-connected device N2 – which already carries the full information of I_{PTAT} – as the input to an additional NMOS device, N3, configured as a current mirror of N2. N3 thus conducts a current $I_{N3} = a \cdot I_{PTAT}$, where a is the N3-to-N2 sizing ratio. This mirrored current is pulled from V_{DD} through a diode-connected PMOS device, P4, whose gate-source voltage self-adjusts to support I_{N3} according to the same subthreshold exponential law described above. The resulting gate bias of P4 is shared directly with P5, forming a second PMOS current mirror; V_{ptat} is taken at this common P4/P5 gate node, which presents a diode-referenced voltage of exactly the type required to drive the gate of the downstream PMOS pass transistor, with P5 itself sized to match the replica device used in that downstream stage. This N2→N3→P4/P5 path therefore tailors the PTAT extraction node to the specific interface demanded by the ring-oscillator sensor topology, at the cost of introducing two additional mirror stages between the PTAT-generating branch and the sensor interface, a trade-off whose impact on supply sensitivity is examined in Section 4.3.4.

This section evaluates and compares three distinct PTAT cores derived from the literature. The first candidate is the sub-threshold PTAT current reference with composite transistors proposed by Hanumantha Rao et al. [31]. The second candidate is the temperature sensor core proposed by Lempel et al. [44], a compact current-controlled-oscillator sensor whose PTAT current source uses a Mixed- V_{th} FinFET stack to act as a self-biased cascode, boosting output resistance and supply rejection without an explicit cascode bias generator, and whose ring-oscillator topology forms the fixed sensing front-end into which all three PTAT sources are integrated. The third candidate is the topology-matched BGR-based PTAT extraction described above, proposed by Moisello et al. [29].

The full schematics and operating equations for each of these cores were thoroughly presented and analysed in Chapter 3 of this dissertation and are therefore not repeated here. The evaluation that follows focuses exclusively on post-integration simulation behaviour when each PTAT source is embedded within the common temperature sensor signal chain.

4.3.2 Testbench Infrastructure and Circuit Integration

Three separate simulation testbenches were constructed, one for each candidate PTAT core, all sharing an identical downstream sensor signal chain so that any difference in the measured output frequency is attributable solely to the PTAT source under evaluation and not to disparities in the readout path.

The first testbench, shown in Figure 4.38, integrates the temperature sensor core proposed by Lempel et al. [44] with its original internal PTAT source. This sensor core is built around a current-controlled oscillator (CCO) charged by a PMOS current source, whose bias is set by an internal PTAT current reference, so that the PTAT voltage V_{ptat} is generated internally and applied directly to the gate of a PMOS pass transistor that regulates V_{ccro} , the effective supply voltage of the temp_ring_osc oscillator block. Because V_{ccro} sits close to the threshold voltage, the oscillator output frequency F_{out} is routed through a unity-gain buffer and a level shifter to produce the rail-to-rail digital output D_{out} , mirroring the buffer-and-level-shifter arrangement used by Lempel et al. [44] to prevent the supply-dependent input capacitance of the level shifter from corrupting the oscillator's effective load. This testbench therefore represents the unmodified reference point against which the two alternative PTAT sources are benchmarked.

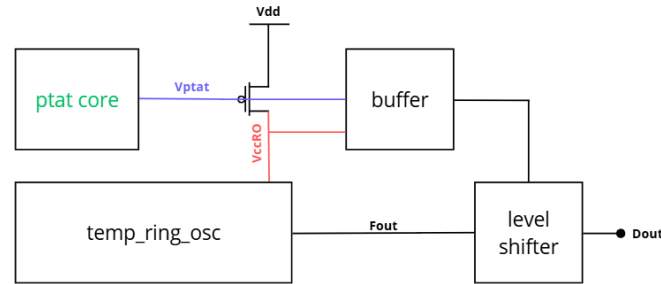


Figure 4.38: Testbench for the native sensor configuration using its own PTAT core.

The second testbench, illustrated in Figure 4.39, replaces the internal PTAT source of the sensor core with the external sub-threshold PTAT current reference proposed by Hanumantha Rao et al. [31]. In this configuration the current reference presents V_{ptat} through a two-terminal output that drives a stacked arrangement of two series PMOS devices before feeding V_{ccro} to the ring oscillator. The stacked drive topology is necessary to satisfy the compliance headroom required for the sub-threshold current reference to remain in its correct operating region while simultaneously presenting a gate drive voltage compatible with the gate-oxide limits of the 16 nm FinFET process.

The third testbench, depicted in Figure 4.40, substitutes the internal PTAT source with the upgraded BGR block proposed by Moisello et al. [29]. Here V_{ptat} is extracted directly from the

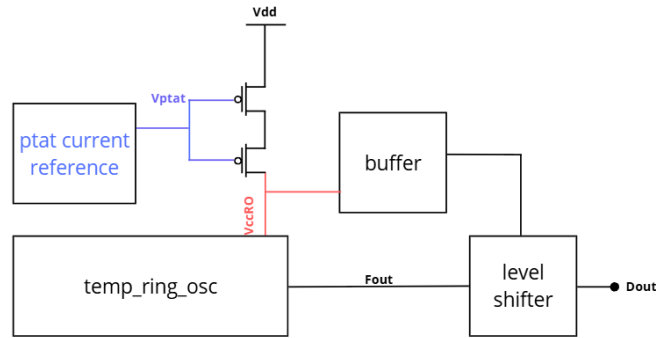


Figure 4.39: Testbench for the sub-threshold PTAT current reference integrated with the ring-oscillator sensor topology.

dedicated BGR output node and connected to the gate of a single PMOS device in the same arrangement as the baseline testbench, reflecting the fact that the BGR-sourced V_{ptat} already presents an appropriate voltage level and output impedance for direct gate drive without additional stacking. The ring oscillator, buffer, and level shifter stages downstream are held constant and identical across all three testbenches.

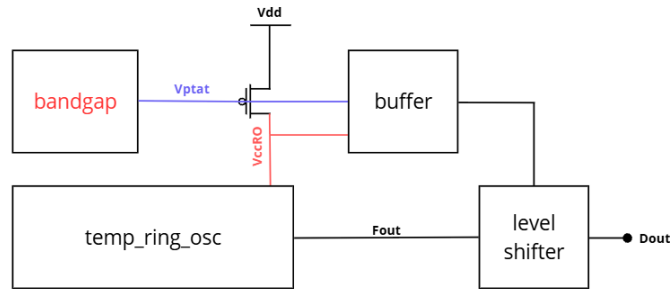


Figure 4.40: Testbench for the BGR-based PTAT source integrated with the ring-oscillator sensor topology.

Process-corner simulations are executed across all five standard corners: slow-slow (SS), fast-fast (FF), typical-typical (TT), slow-fast (SF), and fast-slow (FS).

4.3.3 Process Corner and Temperature Sweep Analysis

Figures 4.41, 4.42, and 4.43 present the simulated average output frequency F_{out} as a function of temperature across all five process corners for the native sensor baseline, the sub-threshold PTAT current reference integration, and the BGR-based integration respectively.

Across all three configurations the output frequency increases monotonically with temperature in every process corner, confirming that the PTAT-to-frequency transduction mechanism is fundamentally functional for all three candidate PTAT sources. This monotonic behaviour is a necessary prerequisite for unambiguous temperature encoding, since a non-monotonic transfer

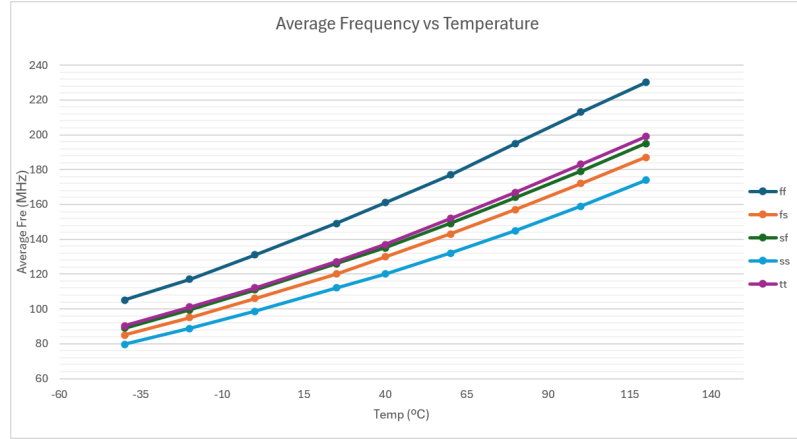


Figure 4.41: Average output frequency versus temperature for the native sensor configuration across all five process corners.

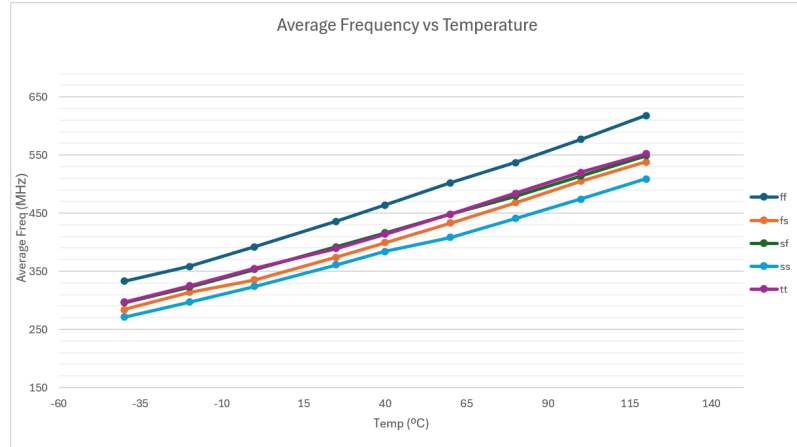


Figure 4.42: Average output frequency versus temperature for the sub-threshold PTAT current reference integrated with the ring-oscillator sensor topology, across all five process corners.

function would introduce temperature ambiguities that no amount of digital post-processing could resolve.

A first important observation is that the inter-corner spread differs markedly between the three configurations, and this spread has a direct bearing on the required range and complexity of the digital calibration algorithm. In the native sensor configuration (Figure 4.41) the five corner curves are comparatively well-clustered. The FF and SS corners bound a frequency window of approximately 25 MHz and at -40°C , and this spread remains similarly contained at approximately 56 MHz at 120°C . The TT, SF, and FS curves are visually difficult to distinguish from one another throughout the sweep, indicating that the native PTAT source of this sensor core introduces only a modest process-induced scatter in the oscillation frequency at any given temperature point. The temperature sensitivity of the TT corner in this baseline configuration can be estimated as

$$S_{F,\text{native}} = \frac{\Delta F_{\text{out}}}{\Delta T} \approx \frac{(200 - 90) \text{ MHz}}{(120 - (-40))^{\circ}\text{C}} = 0.688 \text{ MHz}/^{\circ}\text{C}.$$

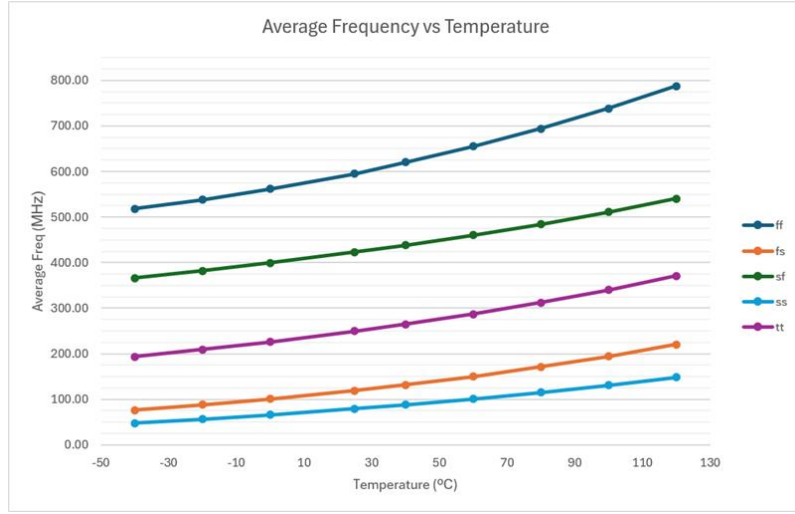


Figure 4.43: Average output frequency versus temperature for the BGR-based PTAT source integrated with the ring-oscillator sensor topology, across all five process corners.

For the sub-threshold current reference integration (Figure 4.42) the corner spread is similarly compact. At -40°C the SS and FF corners differ by approximately 50 MHz, and this spread grows modestly to approximately 100 MHz at 120°C . The TT, SF, and FS curves remain tightly grouped throughout, suggesting that the composite-transistor architecture employed by Hanumantha Rao et al. [31] is effective at suppressing sensitivity to individual transistor threshold voltage and carrier mobility variations, as its elevated effective output resistance $R_{\text{out}} = g_{m2}r_{o2}r_{o1}$ inherently attenuates the gain from process perturbations to the reference current. The TT-corner temperature sensitivity for this configuration is

$$S_{F,\text{ref}} = \frac{\Delta F_{\text{out}}}{\Delta T} \approx \frac{(553 - 297) \text{ MHz}}{(120 - (-40))^{\circ}\text{C}} = 1.6 \text{ MHz}/^{\circ}\text{C},$$

which is approximately $2.3\times$ the sensitivity of the native baseline, reflecting the higher absolute PTAT swing generated by this current reference across the same temperature range.

The BGR-based integration (Figure 4.43) is visually striking in comparison. The BGR-sourced PTAT extraction generates a substantially wider peak-to-peak frequency range across the temperature sweep. The FF corner reaches approximately 790 MHz at 120°C while the SS corner registers only approximately 150 MHz at the same temperature, a corner-to-corner spread exceeding 640 MHz. At the cold extreme of -40°C the FF-to-SS frequency difference already exceeds 475 MHz. The TT-corner sensitivity evaluates to

$$S_{F,\text{bgr}} = \frac{\Delta F_{\text{out}}}{\Delta T} \approx \frac{(370 - 200) \text{ MHz}}{(120 - (-40))^{\circ}\text{C}} = 1.063 \text{ MHz}/^{\circ}\text{C}.$$

Although this figure is nominally comparable to that of the sub-threshold current reference integration, the critical distinction lies in the dramatically wider inter-corner spread. This pronounced process sensitivity arises because the BGR loop, while generating a well-defined temperature-compensated reference voltage V_{ref} at its primary output, exposes the PTAT extraction branch

formed by P4, P5, and R5 to direct process scatter. The mirror ratio of P4 and P5 and the absolute value of R5 are both subject to process-induced variation, and any mismatch between these devices directly perturbs V_{ptat} and therefore V_{ccro} . Although the BGR-based configuration nominally offers the widest raw PTAT swing and thus the highest theoretical temperature sensitivity under the FF corner, this advantage comes at the cost of a significantly expanded inter-corner frequency spread that complicates multi-point digital calibration and substantially degrades worst-case uncalibrated accuracy.

4.3.4 Supply Voltage Sensitivity and Final Topology Selection

Temperature sensors embedded in complex mixed-signal or digital Systems-on-Chip (SoCs) are invariably subject to supply voltage fluctuations arising from simultaneous switching activity in neighbouring logic blocks, resistive IR-drop across the power distribution network, and the finite transient response of on-chip voltage regulators. When the output frequency of a temperature sensor varies with supply voltage V_{DD} independently of temperature, the resulting frequency shift is indistinguishable from a genuine temperature change at the digital readout stage. This cross-axis sensitivity introduces a systematic bias into the temperature estimate that cannot be removed by a conventional single-point temperature calibration. In the context of a multi-sensor PVT compensation scheme, where a voltage sensor, a process monitor, and the temperature sensor are all read simultaneously, any residual VDD dependence in the temperature measurement path forces the calibration algorithm to solve a coupled system of equations rather than treating each physical quantity independently, adding substantial algorithmic complexity and reducing convergence robustness. Minimizing VDD sensitivity in the temperature sensor path is therefore not merely a performance consideration but a fundamental requirement for system-level design closure.

The VDD sensitivity of each configuration was evaluated by sweeping the supply voltage from 0.65 V to 0.95 V at a fixed ambient temperature of 25°C under the TT corner while recording the average output frequency. The results for all three configurations are collected in Figure 4.44.

The VDD sweep results reveal a decisive performance differentiation between the three configurations. The BGR-based topology displays a strongly rising frequency response with increasing supply voltage, climbing from approximately 206 MHz at $V_{\text{DD}} = 0.65$ V to approximately 296 MHz at $V_{\text{DD}} = 0.95$ V, a total variation of 90 MHz across a supply window of only 0.30 V. The equivalent VDD sensitivity is

$$S_{\text{VDD}, \text{bgr}} = \frac{\Delta F_{\text{out}}}{\Delta V_{\text{DD}}} \approx \frac{90 \text{ MHz}}{0.30 \text{ V}} = 300 \text{ MHz/V}.$$

Placing this figure in the context of the TT-corner temperature sensitivity derived in Section 4.3.3, a VDD perturbation of as little as 10 mV produces a frequency shift that a single-axis calibration algorithm – one that only knows the $T \rightarrow F_{\text{out}}$ mapping and has no visibility into V_{DD} – would misinterpret as a genuine temperature change of approximately

$$\Delta T_{\text{equiv}, \text{bgr}} \approx \frac{S_{\text{VDD}, \text{bgr}}}{S_{F, \text{bgr}}} \cdot \Delta V_{\text{DD}} = \frac{300 \text{ MHz/V}}{1.063 \text{ MHz/}^\circ\text{C}} \cdot 10 \text{ mV} \approx 2.8^\circ\text{C},$$

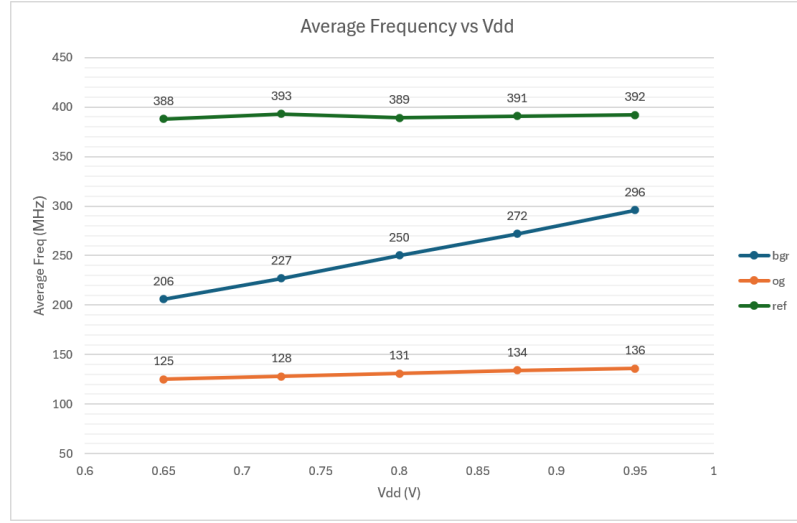


Figure 4.44: Average output frequency versus supply voltage V_{DD} at $T = 25^\circ\text{C}$ (TT corner) for all three PTAT configurations.

even though the true die temperature has not changed at all. This is therefore a pure readout error injected by supply noise, and it alone exceeds the target accuracy budget of the sensor sub-system. The physical origin of this sensitivity lies in the PTAT extraction branch of the BGR: the drain current of P5, and therefore the voltage developed across R5, carries a residual V_{DD} dependence because the output resistance of the P4/P5 current mirror is finite at 16 nm channel lengths. Even when the BGR core feedback loop remains correctly settled, the effective mirroring ratio seen at the P5 drain shifts with V_{DD} , causing V_{ptat} and consequently V_{ccRO} to track supply perturbations. This constitutes a fundamental limitation of extracting a second functional output from a BGR core in an advanced FinFET node where the intrinsic transistor output resistance is substantially degraded relative to planar CMOS processes, and it cannot be fully resolved without cascoding or a dedicated pre-regulator, either of which would reintroduce the area overhead that the modification sought to eliminate.

To isolate the root cause of this behaviour from any loading effect introduced by the downstream ring-oscillator sensor, V_{ptat} was additionally probed directly at the BGR output node, with the ring-oscillator sensor core disconnected from the bias network. This isolated measurement still exhibited a supply-induced variation considerably larger than the corresponding PTAT bias variation observed for the native Lempel et al. [44] source and the sub-threshold current reference of Hanumantha Rao et al. [31] under the same sweep, confirming that the sensitivity originates within the BGR core and its extraction mirror itself rather than being introduced or amplified by the ring-oscillator load. Two candidate explanations were considered for this behaviour: either the P4/P5 mirror is a comparatively low-output-impedance structure that does not, by itself, provide sufficient robustness to hold the PTAT current constant against V_{DD} variation, consistent with the finite output-resistance argument given above, or the core devices require further resizing to tighten this regulation. Pursuing the resizing route, however, exposed a structural conflict: the same core devices, most notably N2, jointly generate the primary PTAT current that is summed

into the V_{ref} output used to bias the voltage sensor's current-starved delay chain (Section 4.2.1) and feed the N3-mirrored extraction branch that produces V_{ptat} . Any resizing attempted to improve the supply rejection of V_{ptat} was found to perturb the balance of the PTAT and CTAT branches at the primary output, degrading the line regulation and temperature coefficient of V_{ref} reported in Section 4.2.1, and conversely, re-tuning the core for the tightest possible V_{ref} performance widened the V_{ptat} excursion. No sizing point was found within the explored design space that simultaneously satisfied both requirements. This strict trade off between the two outputs, a direct consequence of reusing a single BGR core to serve both the voltage sensor's reference and the temperature sensor's PTAT bias, is identified as the most probable structural cause of the BGR-based configuration's poor supply rejection, and it further reinforces the decision, discussed below, to disqualify this topology as the temperature sensor's PTAT source in favour of a dedicated, single-purpose reference.

The native sensor configuration behaves more favourably than the BGR-based topology under the same VDD sweep. Its output frequency changes from approximately 125 MHz to 136 MHz across the 0.65 V to 0.95 V range, a total absolute variation of 11 MHz – the smallest raw excursion of the three candidates – yielding

$$S_{\text{VDD}, \text{native}} \approx \frac{11 \text{ MHz}}{0.30 \text{ V}} = 36.7 \text{ MHz/V}.$$

However, the meaningful metric for comparing supply rejection across configurations operating at different absolute frequency levels is the normalised fractional variation, defined as

$$\delta_{\text{VDD}} = \frac{\Delta F_{\text{out}}}{F_{\text{out}, \text{nom}}} \times 100 [\%],$$

where $F_{\text{out}, \text{nom}}$ is the output frequency at the nominal supply of 0.8 V. Evaluating this metric for the native sensor configuration, whose nominal output frequency is only ≈ 131 MHz, gives $\delta_{\text{VDD}, \text{native}} \approx 8.4\%$ across the full supply window – noticeably worse, on a normalised basis, than the sub-threshold current reference integration discussed below, despite having the smallest absolute frequency swing of the three candidates.

The sub-threshold PTAT current reference integration produces the flattest normalised response of the three candidates by a wide margin. Its output frequency varies only from approximately 388 MHz to 392 MHz across the same VDD range, a total absolute variation of only 4 MHz, yielding

$$S_{\text{VDD}, \text{ref}} \approx \frac{4 \text{ MHz}}{0.30 \text{ V}} = 13.3 \text{ MHz/V}.$$

Its nominal output frequency at $V_{\text{DD}} = 0.8 \text{ V}$ is ≈ 389 MHz, so the normalised fractional variation evaluates to $\delta_{\text{VDD}, \text{ref}} \approx 1.0\%$, an order of magnitude lower than either of the other two candidates. This superior relative supply rejection is a direct consequence of the composite transistor structure employed by Hanumantha Rao et al. [31]: the elevated effective output resistance

$R_{\text{out}} = g_{m2} r_{o2} r_{o1}$ of the composite device acts as an inherent cascode, shielding the reference current from supply perturbations without consuming additional static current or requiring a dedicated cascode supply rail. As the reference current, and therefore V_{ptat} , remains essentially constant with VDD, the ring-oscillator supply V_{ccRO} is similarly stabilised, and the fractional change in F_{out} is minimised relative to its nominal value. The near-flat normalised profile of this integration in Figure 4.44 confirms that the gain from VDD variations to the PTAT current is substantially attenuated compared to both the native sensor topology and the BGR-based extraction, as reported by Hanumantha Rao et al. [31].

Table 4.8 consolidates the VDD sensitivity figures for all three topologies, alongside the TT-corner temperature sensitivity S_F derived in Section 4.3.3 and the resulting temperature-equivalent error for a representative 10mV supply perturbation. This combined view makes clear that the ranking of the three topologies is unchanged whether expressed in absolute or normalised terms: the BGR-based configuration is the worst-performing candidate by a wide margin in every column, the native sensor configuration is intermediate, and the sub-threshold current reference integration is the best-performing candidate once the comparison is normalised to each configuration's own operating frequency – which is the only fair basis for comparison given the very different absolute frequency ranges of the three testbenches.

TECHNICAL NOTE: The 10mV step is not arbitrary: Eberlein et al. [43] report their sensor's supply sensitivity as $\pm 0.2^\circ\text{C}/10\text{mV}$, and the same convention is used here so Table 4.8 and Table 4.11 compare directly against the literature.

Table 4.8: Comparison of supply-voltage sensitivity across the three PTAT testbenches, $V_{\text{DD}} = 0.65\text{--}0.95\text{ V}$, $T = 25^\circ\text{C}$, TT corner.

Configuration	Native [44]	Sub-threshold [31]	Updated BGR [29]
$F_{\text{out}} @ 0.65\text{ V (MHz)}$	125	388	206
$F_{\text{out}} @ 0.95\text{ V (MHz)}$	136	392	296
$F_{\text{out,nom}} @ 0.80\text{ V (MHz)}$	131	389	250
$\Delta F_{\text{out}} \text{ (MHz)}$	11	4	90
$S_{\text{VDD}} \text{ (MHz/V)}$	36.7	13.3	300.0
$\delta_{\text{VDD}} \text{ (\%)}$	8.4	1.0	36.0
$S_F, \text{ TT corner (MHz/}^\circ\text{C)}$	0.688	1.6	1.063
$\Delta T_{\text{equiv}} \text{ for } 10\text{mV (}^\circ\text{C)}$	0.53	0.08	2.8

Taken together, the process-corner analysis of Section 4.3.3 and the VDD sensitivity results of this subsection define a clear engineering trade-off across the three configurations, summarised in Table 4.8. The BGR-based topology, despite offering the widest raw inter-corner frequency excursion under the FF process corner, exhibits a fractional VDD sensitivity of approximately

36.0% across the evaluated supply window, rendering it incompatible with the strict supply decoupling requirement of this application. The native sensor configuration presents the most compact inter-corner process spread and the smallest absolute frequency excursion with VDD, but once normalised to its own (comparatively low) nominal operating frequency its fractional sensitivity of 8.4% is in fact almost an order of magnitude higher than that of the sub-threshold current reference integration. The sub-threshold PTAT current reference integration achieves by far the best normalised VDD rejection of the three candidates, at approximately 1.0%, while simultaneously providing a wider temperature-sensitive frequency range than the native baseline, as demonstrated by the higher $S_{F,\text{ref}}$ derived in Section 4.3.3, and the lowest temperature-equivalent error for a given supply perturbation, as shown in the bottom row of Table 4.8.

Given that the primary design objective of this work is to strictly remove VDD dependence from the temperature measurement path so as to reduce the global PVT compensation algorithm to a set of decoupled, single-axis regression operations, the ring-oscillator sensor topology integrated with the sub-threshold PTAT current reference is selected as the final architecture for all subsequent implementation and layout steps. This configuration satisfies the VDD independence criterion most rigorously on a normalised basis, benefits from the supply rejection mechanism inherent to the composite-transistor structure proposed by Hanumantha Rao et al. [31], and delivers a wider temperature-to-frequency conversion range than the native baseline, improving the effective resolution of the downstream digital frequency counter without any additional power or area penalty at the system level.

4.4 Summary and Comparison with the State of the Art

Having presented the simulation results for the three sensing subsystems individually in Sections 4.1, 4.2, and 4.3, this closing section consolidates the key performance parameters of each proposed sensor and places them directly alongside the corresponding topologies surveyed in Chapter 3. Three comparison tables are presented in turn, one for each sensing modality, followed by a discussion of how far this work has progressed toward the unified PVT monitoring platform introduced at the start of this chapter, and what remains as future work.

4.4.1 Process Sensor Comparison

Table 4.9 positions the Diode Connected Load Ring Oscillator and the Ratioed Topology of Chandan et al. [32] against three further process sensors drawn from the literature.

Table 4.9: Comparison of process sensor parameters between this work and related ring-oscillator-based process sensing topologies.

Work	[33]	[14]	[32]	Diode (this work)	Ratioed (this work)
Year	2013	2016	2025	2026	2026
Process	65 nm	130 nm	–	16 nm	16 nm
V_{DD} (V)	1.1	1.2	1.8	0.8	0.8
Stages	7	35	21	21	21
Range (°C)	–	–	10–100	–40 to 120	–40 to 120
k_N	0.865	1.934	4.086	3.48	1.54
k_P	1.335	3.189	7.223	4.01	1.34
Power (mW)	0.291	3.288	2.902	1.544	1.860

The table places the proposed Diode Connected Load Ring Oscillator alongside both the older literature designs of Fujimoto et al. [33] and An et al. [14] and the more recent Ratioed Topology of Chandan et al. [32] introduced in Chapter 3. Several trends are visible. The two older works operate at considerably higher supply voltages, between 1.1 V and 1.2 V, reflecting the larger technology nodes in which they were implemented, while both topologies evaluated in this dissertation operate at the 0.8 V supply native to the 16 nm node. The stage count varies widely across the literature, from 7 to 35, with no clear correlation to sensitivity, suggesting that the sensitivity coefficients k_N and k_P are governed primarily by the delay cell topology itself rather than by the number of cascaded stages. Among all five entries, the Ratioed Topology of Chandan et al. [32] reports the highest own device sensitivity, with k_N and k_P both exceeding 4, while the proposed Diode Connected Load Ring Oscillator of this work occupies an intermediate position, ahead of the two earlier works of Fujimoto et al. [33] and An et al. [14] but below the most recent reference design. This is consistent with the corner level analysis already presented in Section 4.1.4, where the Diode Connected Load Ring Oscillator was shown to outperform the same Ratioed Topology

once frequency normalised sensitivity, temperature stability, and supply robustness are considered jointly rather than in isolation. The power figures further reinforce the suitability of the proposed design for the unified platform, with the Diode Connected Load Ring Oscillator consuming less power than the Ratioed Topology implemented in the same technology node under the same test-bench, despite its comparable structural complexity.

4.4.2 Voltage Sensor Comparison

Table 4.10 compares the voltage sensor implemented in this work against three current starved delay chain based voltage sensors from the literature, including the original architecture of Chen et al. [4] on which this design is based.

Table 4.10: Comparison of voltage sensor parameters between this work and related current starved delay chain voltage sensing topologies.

Work	[4]	[13]	[10]	This work
Year	2010	2014	2021	2026
Process	65 nm	16 nm	180 nm	16 nm
V_{DD} (V)	0.5	0.7	3.3	0.8
Resolution (mV)	50 (5%)	4 (5.7%)	16.5 (0.5%)	75 (5%)
Range (°C)	0–100	–10 to 90	–40 to 80	–40 to 120
Power	3.7 μ W	70 μ W	48.5 mW	12.8 μ W

The voltage sensor comparison highlights the trade off between resolution and operating range that characterises this class of sensor. The work of Horng et al. [13] achieves the finest resolution of the four entries, at 4 mV, but this is reported over a comparatively narrow temperature range of -10°C to 90°C . The implementation of Lou et al. [10] achieves a resolution of 16.5 mV, but does so at a supply voltage of 3.3 V in a much older 180 nm node and at a power cost of 48.5 mW, several orders of magnitude above every other entry in the table. The voltage sensor proposed in this work reports a resolution of 75 mV, the coarsest of the four entries, while extending the validated temperature range to the full industrial span of -40°C to 120°C and operating at a power consumption of 12.8 μ W, comfortably between the two low power literature designs and far below the high resolution but power hungry implementation of Lou et al. [10]. The coarser resolution relative to Chen et al. [4] and Horng et al. [13] is acknowledged as a limitation of the present design rather than a deliberately targeted trade off, and improving it, for example through finer current-starved delay cell sizing or an increased number of delay stages, is identified as a direction for future refinement, to be weighed against the corresponding impact on power consumption and range.

4.4.3 Temperature Sensor Comparison

Table 4.11 compares the temperature sensor configuration selected in Section 4.3.4, here labelled Ratioed in line with its sub-threshold composite transistor PTAT source, against three time domain temperature sensors from the literature.

Table 4.11: Comparison of temperature sensor parameters between this work and related time domain temperature sensing topologies. The supply sensitivity entries for Someya et al. [45] and Lu et al. [46] are not yet available and are marked accordingly.

Work	[43]	[44]	[45]	[46]	This work
Year	2019	2021	2020	2022	2026
Process	7 nm	16 nm	65 nm	40 nm	16 nm
V_{DD} (V)	1	1.2	0.9	1	0.8
Range ($^{\circ}\text{C}$)	−20 to 120	−20 to 110	−45 to 85	−40 to 85	−40 to 120
Power (μW)	12.5	41.8	47.2	30	16.32
Supply Sensitivity ($^{\circ}\text{C}/10\text{ mV}$)	0.200	0.060	0.028	N/A	0.083

Among the four reference designs, the sensor of Lu et al. [46] reports the lowest supply sensitivity of the comparison set, at 0.028°C per 10 mV (equivalently $2.8^{\circ}\text{C}/\text{V}$ as reported in the original work). The sensor of Lempel et al. [44], which forms the fixed ring oscillator front end adopted in this work, reports the second-best supply sensitivity at 0.060°C per 10 mV ($6^{\circ}\text{C}/\text{V}$) when paired with its own native PTAT source. The sensor of Eberlein et al. [43] reports the third-best supply sensitivity of the group at 0.20°C per 10 mV , but achieves the lowest power consumption of the comparison set at $12.5\text{ }\mu\text{W}$, despite being implemented in the most advanced 7 nm node of the group. The temperature sensor proposed in this work, built on the same Lempel et al. [44] front end but reconfigured with the composite transistor PTAT source of Hanumantha Rao et al. [31], achieves a supply sensitivity of 0.083°C per 10 mV (equivalently $8.3^{\circ}\text{C}/\text{V}$), following a refinement of the composite-transistor bias network that rebalanced the current ratios and overdrive margins of the PTAT branch. This is the third most favourable figure of the five entries in Table 4.11, behind the frequency-locked-loop sensor of Lu et al. [46] and the unmodified Lempel et al. [44] front end with its own native PTAT source, and it comfortably outperforms the N-well diode charge-pump sensor of Eberlein et al. [43] (0.200°C per 10 mV) despite that design being implemented in the more advanced 7 nm node. Relative to that unmodified Lempel et al. [44] front end (0.060°C per 10 mV , equivalently $6^{\circ}\text{C}/\text{V}$), the proposed configuration remains approximately 38% higher, and it is still roughly $3\times$ higher than the best-in-class design of Lu et al. [46] (0.028°C per 10 mV). This residual gap is also achieved despite this work operating at the lowest supply voltage of the five entries in Table 4.11: $V_{DD} = 0.8\text{ V}$, against 1.2 V for Lempel et al. [44], 1 V for both Eberlein et al. [43] and Someya et al. [45], and 0.9 V for Lu et al. [46]. A lower supply voltage generally leaves less headroom for the cascoding or dedicated pre-regulation stages that the highest-PSRR designs rely on to suppress VDD sensitivity at the transducer or bias-generation

level, so narrowing the gap to Lempel et al. [44]’s own native-source figure while simultaneously operating at a 0.4 V lower supply is a comparatively favourable outcome, even though it does not fully close that gap. Part of the remaining gap is likely attributable to factors beyond the choice of PTAT topology and supply voltage alone. Lu et al. [46] employs a closed-loop frequency-locked architecture with sub-thermal drain voltage stabilization explicitly engineered to suppress VDD sensitivity at the transducer level rather than only at the bias-generation level, as in the present design, and Lu et al. [46] and Eberlein et al. [43] are additionally implemented in different process nodes (65 nm and 7 nm respectively), so the comparison with this work and Lempel et al. [44], both evaluated in the same 16 nm node, is not fully like-for-like and intrinsic device-level effects that vary with node are convolved with the architectural differences being compared. The proposed sensor also retains the widest validated temperature range of the comparison set, from -40°C to 120°C , and a power consumption of $16.32\text{ }\mu\text{W}$ that remains competitive with the rest of the group. The supply-sensitivity result is therefore reported here as a substantially narrowed gap relative to the best literature designs – a clear improvement over the earlier, unrefined biasing of the same composite structure – rather than as the standalone limitation it would otherwise represent.

The three comparison tables presented above confirm that the process and voltage sensing blocks developed in this chapter perform competitively against the corresponding state of the art topologies surveyed in Chapter 3, while also identifying two concrete, honest limitations of the present design: the coarser voltage resolution relative to Chen et al. [4] and Horng et al. [13], discussed in Section 4.4.2, and the residual supply sensitivity gap of the temperature sensor relative to the two best literature designs, discussed in Section 4.4.3.

Chapter 5

Conclusions

This dissertation set out from the premise that individually optimized process, voltage, and temperature sensors are insufficient on their own, since each remains exposed to cross sensitivity from the other two environmental axes, and that a coordinated sensing infrastructure is required to resolve this. Chapter 4 pursued that premise through the transistor level design, simulation, and benchmarking of three sensing subsystems in a 16 nm FinFET technology node, each evaluated directly against the corresponding state of the art topology identified in Chapter 3, with the detailed comparative results discussed in Section 5.2 alongside the limitations and future work arising from them. This closing chapter steps back from those details to summarise, at a higher level, what this dissertation has established and what it means for the unified PVT monitoring platform proposed as its architectural target.

5.1 Summary of Contributions

Across the three sensing subsystems, this work consistently found that state of the art topologies designed for older or larger process nodes do not transfer unmodified to 16 nm FinFET, and that re-deriving each design from first principles at the target node exposes both new limitations and new opportunities. The Diode Connected Load Ring Oscillator proposed for process sensing simplifies the ratioed NAND and NOR delay cells of Chandan et al. [32] into a single transistor type structure that, once properly frequency normalised through the sensitivity coefficients k_N and k_P , proved more sensitive to its own targeted device corner while also being substantially more robust to temperature and supply variation. The voltage sensor built on the current starved delay chain of Chen et al. [4] and the bandgap reference of Moisello et al. [29] required a complete re-derivation of the inter-stage timing margin, revealing that the 2-stage spacing adequate at 65 nm violates the encoding constraint once propagation delays are compressed at 16 nm, a scaling limitation that an 8-stage interval was shown to resolve. The temperature sensor, built around the ring oscillator front end of Lempel et al. [44], demonstrated that supply rejection in this class of sensor is governed primarily by the choice of PTAT bias source rather than by the oscillator topology itself: the composite transistor reference of Hanumantha Rao et al. [31] reduced supply sensitivity by more

than a factor of six relative to the sensor's native source, and did so while operating at a 0.8 V supply, 0.4 V below the 1.2 V used by Lempel et al. [44]'s own reported implementation, leaving comparatively less headroom for supply-rejection circuitry. A third candidate PTAT source, extracted from the FinFET-native bandgap reference of Moisello et al. [29], was disqualified on the same basis; isolating its PTAT node from the sensor load confirmed that its excess supply sensitivity originated within the bandgap core itself rather than the ring-oscillator load, and that it could not be resolved through transistor resizing alone without degrading the reference voltage that the same core supplies to the voltage sensor, an inherent conflict of reusing a single bandgap core for two independent PVT sensing paths.

Taken together, these results constitute the principal contribution of this dissertation: a complete, transistor level, simulation validated set of process, voltage, and temperature sensing building blocks, each re-optimized for the 16 nm FinFET node and each benchmarked quantitatively against the relevant state of the art, as detailed throughout Chapter 4.

5.2 Future Work

It is important to revisit the architectural target set out at the beginning of Chapter 4 in Figure 4.1 and Figure 4.2, where the three sensors were proposed not as standalone blocks but as electrically interdependent stages of a single unified PVT monitoring system, in which the process code P would feed forward into both the temperature and voltage sensing paths, and the temperature code T would in turn correct the voltage sensor ahead of the final compensation stage governed by the FSM.

Due to the time constraints of this dissertation, this full closed loop integration was not reached. Each sensor in this chapter was designed, sized, and validated independently, under its own dedicated testbench and its own fixed bias conditions, rather than within the shared, sequentially dependent environment that the FSM of Figure 4.2 describes. The process correction input to the temperature sensor, the combined process and temperature correction inputs to the voltage sensor, and the final PVT compensation logic itself were therefore not implemented or simulated as part of this work. This is acknowledged here as an honest limitation of the present results rather than as a claim of full system closure, and it is identified as the most significant remaining piece of future work arising from this dissertation. A natural continuation of this work would begin by formalizing the digital interface between the three sensor codes, designing the correction logic implied by the FSM states `Temperature` and `Voltage` in Figure 4.2, and validating the complete closed loop system under the same 16 nm FinFET testbench conditions used throughout this dissertation, so that the standalone performance figures reported in Tables 4.9, 4.10, and 4.11 can ultimately be re-evaluated under realistic, fully coupled operating conditions.

Beyond this system level integration effort, three block level refinements identified in Chapter 4 also remain as future work. First, the resolution of the voltage sensor, currently the coarsest of the four entries compared in Table 4.10, could be improved through finer current-starved delay cell sizing or an increased number of delay stages. Second, and still relevant given its role

within the unified platform, the residual supply sensitivity of the temperature sensor could be narrowed further: following the refinement of the composite-transistor bias network reported in Section 4.4.3, the proposed configuration achieves 0.083°C per 10 mV, the third most favourable figure of the five entries in Table 4.11, but it remains approximately $3\times$ higher than the best-in-class design of Lu et al. [46] and modestly higher than the unmodified Lempel et al. [44] baseline it was built from. Since the temperature code T is intended to correct the voltage sensing path in the unified architecture of Figure 4.2, a temperature sensor whose own output varies with supply voltage would still propagate a residual, though now considerably smaller, error into the voltage correction stage, partially limiting the closed loop compensation the platform is meant to provide. Closing the remaining gap is therefore not only a further standalone performance improvement but would strengthen the precondition for the system level integration described above, and a natural next step would be to investigate alternative bias generation or transducer level techniques, such as those used by Lu et al. [46] and Eberlein et al. [43], that suppress VDD sensitivity directly at the sensing element rather than only at the PTAT bias path. Third, the bandgap reference results reported in Section 4.3.1, and consequently the voltage sensor and BGR-based temperature sensor results that depend on them, were obtained using an ideal operational amplifier model rather than a transistor-level implementation; replacing this ideal model with a fully sized op-amp would be a necessary step before tape-out, and is expected to modestly degrade, though not qualitatively change, the reported line regulation, temperature coefficient, and settling-time margins.

The individual results presented throughout this dissertation nonetheless provide a solid and independently validated foundation for that future integration effort.

References

- [1] B. Razavi, *Design of Analog CMOS Integrated Circuits*, 2nd. New York, NY, USA: McGraw-Hill Education, 2017.
- [2] M. Alioto, V. De, and A. Marongiu, “Energy-quality scalable integrated circuits and systems: Continuing energy scaling in the twilight of moore’s law,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 4, pp. 653–678, 2018. DOI: [10.1109/JETCAS.2018.2881461](https://doi.org/10.1109/JETCAS.2018.2881461)
- [3] M.-H. Chang et al., “Near-/sub-vth process, voltage, and temperature (pvt) sensors with dynamic voltage selection,” in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 133–136. DOI: [10.1109/ISCAS.2013.6571800](https://doi.org/10.1109/ISCAS.2013.6571800)
- [4] S.-W. Chen, M.-H. Chang, W.-C. Hsieh, and W. Hwang, “Fully on-chip temperature, process, and voltage sensors,” in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010, pp. 897–900. DOI: [10.1109/ISCAS.2010.5537410](https://doi.org/10.1109/ISCAS.2010.5537410)
- [5] E. C. López, J. Montahuano, E. Astudillo, and E. Holguín, “Process variation sensing and compensation employing a ring oscillator-based architecture in skywater 130nm technology,” in *2025 38th SBC/SBMicro/IEEE Symposium on Integrated Circuits and Systems Design (SBCCI)*, 2025, pp. 1–5. DOI: [10.1109/SBCCI66862.2025.11218682](https://doi.org/10.1109/SBCCI66862.2025.11218682)
- [6] M. Lohitsiri and N. Panitantum, “A robust bandgap voltage reference using tlnt compensation technique for a low temperature coefficient,” in *2025 22nd International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2025, pp. 1–6. DOI: [10.1109/ECTI-CON64996.2025.11101119](https://doi.org/10.1109/ECTI-CON64996.2025.11101119)
- [7] Z. Li, Y. . Xu, K. Cao, W. . Li, X. Hu, and X. Dong, “A high-order temperature compensation bandgap voltage reference with line sensitivity enhanced,” in *2024 9th International Conference on Integrated Circuits and Microsystems (ICICM)*, 2024, pp. 757–761. DOI: [10.1109/ICICM63644.2024.10814172](https://doi.org/10.1109/ICICM63644.2024.10814172)
- [8] T. P. Sahu, R. K. Tiwari, P. K. Misra, M. Goswami, and K. Kandpal, “Design and analysis of low-voltage, mos-only bandgap reference circuit,” in *2023 IEEE International Symposium on Smart Electronic Systems (iSES)*, 2023, pp. 283–286. DOI: [10.1109/iSES58672.2023.00064](https://doi.org/10.1109/iSES58672.2023.00064)
- [9] M. J. Pelgrom and A. C. Duinmaijer, “Matching properties of mos transistors,” in *ESSCIRC ’88: Fourteenth European Solid-State Circuits Conference*, 1988, pp. 327–330. DOI: [10.1109/ESSCIRC.1988.5468276](https://doi.org/10.1109/ESSCIRC.1988.5468276)
- [10] P.-Y. Lou, Y.-X. Chen, and C.-C. Wang, “On-chip cmos corner detector design for panel drivers,” in *2021 18th International SoC Design Conference (ISOCC)*, 2021, pp. 11–12. DOI: [10.1109/ISOCC53507.2021.9613965](https://doi.org/10.1109/ISOCC53507.2021.9613965)

- [11] S. Lyu and Z. Shi, "On-chip process variation sensor based on sub-threshold leakage current with weak bias voltages," in *2019 International Conference on IC Design and Technology (ICICDT)*, 2019, pp. 1–4. DOI: [10.1109/ICICDT.2019.8790891](https://doi.org/10.1109/ICICDT.2019.8790891)
- [12] A. Vijayakumar, R. Kumar, and S. Kundu, "On design of low cost power supply noise detection sensor for microprocessors," in *2012 IEEE Computer Society Annual Symposium on VLSI*, 2012, pp. 120–125. DOI: [10.1109/ISVLSI.2012.32](https://doi.org/10.1109/ISVLSI.2012.32)
- [13] J.-J. Horng et al., "A 0.7v resistive sensor with temperature/voltage detection function in 16nm finfet technologies," in *2014 Symposium on VLSI Circuits Digest of Technical Papers*, 2014, pp. 1–2. DOI: [10.1109/VLSIC.2014.6858376](https://doi.org/10.1109/VLSIC.2014.6858376)
- [14] Y.-J. An, D.-H. Jung, K. . Ryu, H. S. Yim, and S.-O. Jung, "All-digital on-chip process sensor using ratioed inverter-based ring oscillator," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 11, pp. 3232–3242, 2016. DOI: [10.1109/TVLSI.2016.2550603](https://doi.org/10.1109/TVLSI.2016.2550603)
- [15] T. Sakurai and A. Newton, "Alpha-power law mosfet model and its applications to cmos inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, 1990. DOI: [10.1109/4.52187](https://doi.org/10.1109/4.52187)
- [16] M. Coelho et al., "Revisiting ztc and gztc in cmos finfet using a five-dc-parameter mosfet model," *IEEE Access*, vol. 13, pp. 189 959–189 978, 2025. DOI: [10.1109/ACCESS.2025.3629328](https://doi.org/10.1109/ACCESS.2025.3629328)
- [17] M. Eberlein and H. Pretl, "Recent developments in bandgap references for nanometer cmos technologies," in *2020 Austrochip Workshop on Microelectronics (Austrochip)*, 2020, pp. 42–46. DOI: [10.1109/Austrochip51129.2020.9232986](https://doi.org/10.1109/Austrochip51129.2020.9232986)
- [18] W. Huang, L. Liu, and Z. Zhu, "A sub-200nW all-in-one bandgap voltage and current reference without amplifiers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 1, pp. 121–125, 2021. DOI: [10.1109/TCSII.2020.3007195](https://doi.org/10.1109/TCSII.2020.3007195)
- [19] P. Mok and K. N. Leung, "Design considerations of recent advanced low-voltage low-temperature-coefficient cmos bandgap voltage reference," in *Proceedings of the IEEE 2004 Custom Integrated Circuits Conference (IEEE Cat. No.04CH37571)*, 2004, pp. 635–642. DOI: [10.1109/CICC.2004.1358907](https://doi.org/10.1109/CICC.2004.1358907)
- [20] H. Zhuang, Y. Li, X. Wang, L. Huang, and Q. Li, "A high-accuracy bandgap reference with compact output driver," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 72, no. 3, pp. 464–468, 2025. DOI: [10.1109/TCSII.2025.3531922](https://doi.org/10.1109/TCSII.2025.3531922)
- [21] J. Jiang, W. Shu, and J. S. Chang, "A 5.6 ppm/°C temperature coefficient, 87-db psrr, sub-1-v voltage reference in 65-nm cmos exploiting the zero-temperature-coefficient point," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 3, pp. 623–633, 2017. DOI: [10.1109/JSSC.2016.2627544](https://doi.org/10.1109/JSSC.2016.2627544)
- [22] R. Khola, A. P. Shah, and A. Shukla, "A −40°C to 125°C, 1.12 ppm/°C multiple voltage bandgap reference circuit," in *2024 International VLSI Symposium on Technology, Systems and Applications (VLSI TSA)*, 2024, pp. 1–2. DOI: [10.1109/VLSITSA60681.2024.10546419](https://doi.org/10.1109/VLSITSA60681.2024.10546419)
- [23] E. K. Lee, "A low voltage cmos bandgap reference without using an opamp," in *2009 IEEE International Symposium on Circuits and Systems*, 2009, pp. 2533–2536. DOI: [10.1109/ISCAS.2009.5118317](https://doi.org/10.1109/ISCAS.2009.5118317)

- [24] X. Zou, Y. Zhang, X. Wu, and G. Cong, "A high stability bandgap reference circuit without amplifier structure," in *2024 3rd International Conference on Energy, Power and Electrical Technology (ICEPET)*, 2024, pp. 968–971. DOI: [10.1109/ICEPET61938.2024.10626027](https://doi.org/10.1109/ICEPET61938.2024.10626027)
- [25] Q. Ding, P. Yuan, D. Li, and Z. Wang, "A sub-1-v ultra-low power full cmos bandgap reference working in subthreshold region," in *2014 IEEE International Conference on Electron Devices and Solid-State Circuits*, 2014, pp. 1–2. DOI: [10.1109/EDSSC.2014.7061222](https://doi.org/10.1109/EDSSC.2014.7061222)
- [26] R. K. Palani, S. Bhagavatula, and D. K. Yuen, "A sub-1-v 8.5-ppm/°C sampled bandgap voltage reference," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 10, pp. 4153–4157, 2022. DOI: [10.1109/TCSII.2022.3184816](https://doi.org/10.1109/TCSII.2022.3184816)
- [27] Y. Cao, H. Zhuang, and Q. Li, "A 0.8-v supply, 1.58% 3σ -accuracy, 1.9- μ W bandgap reference in 0.13- μ m cmos," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 71, no. 4, pp. 1884–1888, 2024. DOI: [10.1109/TCSII.2023.3339236](https://doi.org/10.1109/TCSII.2023.3339236)
- [28] K. L. Low, N. Ahmad, R. Hussin, and J. Wen, "Design and optimization of sub-1v low-power bandgap reference circuit using finfet technology," in *2025 IEEE International Conference on Sensors and Nanotechnology (SENNANO)*, 2025, pp. 234–237. DOI: [10.1109/SENNANO64909.2025.11472791](https://doi.org/10.1109/SENNANO64909.2025.11472791)
- [29] E. Moisello, E. Bonizzoni, and P. Malcovati, "Mosfet-based voltage reference circuits in the last decade: A review," *Micromachines*, vol. 15, no. 12, p. 1504, 2024. DOI: [10.3390/mi15121504](https://doi.org/10.3390/mi15121504)
- [30] P. Park, J. Lee, and S. Cho, "A pvt-insensitive sub-ranging current reference achieving 11.4-ppm/°C from -20°C to 125°C ," *IEEE Journal of Solid-State Circuits*, vol. 59, no. 12, pp. 4057–4067, 2024. DOI: [10.1109/JSSC.2024.3450950](https://doi.org/10.1109/JSSC.2024.3450950)
- [31] G. Hanumantha Rao, C. B. Muhammed Mansoor, and S. Rekha, "A 0.8 v, 5 na ptat current reference circuit with improved supply voltage sensitivity," in *2019 Global Conference for Advancement in Technology (GCAT)*, 2019, pp. 1–4. DOI: [10.1109/GCAT47503.2019.8978311](https://doi.org/10.1109/GCAT47503.2019.8978311)
- [32] M. Chandan, S. S. Rekha, and K. Sudeendra Kumar, "Ratioed nand and nor based on-chip process sensor," in *2025 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, 2025, pp. 1–6. DOI: [10.1109/CONECCT65861.2025.11306891](https://doi.org/10.1109/CONECCT65861.2025.11306891)
- [33] S. Fujimoto, A. K. M. M. Islam, T. Matsumoto, and H. Onodera, "Inhomogeneous ring oscillator for within-die variability and rtn characterization," *IEEE Transactions on Semiconductor Manufacturing*, vol. 26, no. 3, pp. 296–305, 2013. DOI: [10.1109/TSM.2013.2265702](https://doi.org/10.1109/TSM.2013.2265702)
- [34] S. Singhai, A. Kharalkar, and R. Zele, "Process sensor using current-steering logic inverter based ring oscillator," in *2022 IEEE 19th India Council International Conference (INDICON)*, 2022, pp. 1–5. DOI: [10.1109/INDICON56171.2022.10039927](https://doi.org/10.1109/INDICON56171.2022.10039927)
- [35] G. P. Kalarikkal, R. Goel, and H. Shrimali, "Design of cmos device process sensor in 28 nm fd-soi with 2 % of frequency spread," in *2021 28th IEEE International Conference on Electronics, Circuits, and Systems (ICECS)*, 2021, pp. 1–6. DOI: [10.1109/ICECS53924.2021.9665465](https://doi.org/10.1109/ICECS53924.2021.9665465)

- [36] A. K. M. M. Islam, J. Shiomi, T. Ishihara, and H. Onodera, "Wide-supply-range all-digital leakage variation sensor for on-chip process and temperature monitoring," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 11, pp. 2475–2490, 2015. DOI: [10.1109/JSSC.2015.2461598](https://doi.org/10.1109/JSSC.2015.2461598)
- [37] T. Kishimoto, T. Ishihara, and H. Onodera, "On-chip temperature and process variation sensing using a reconfigurable ring oscillator," in *2017 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, 2017, pp. 1–4. DOI: [10.1109/VLSI-DAT.2017.7939649](https://doi.org/10.1109/VLSI-DAT.2017.7939649)
- [38] P. Ituero, M. Lopez-Vallejo, H. Aparicio, and F. Garcia-Redondo, "Taxonomy of power supply monitors and integration challenges," in *2016 IEEE 21st International Mixed-Signal Testing Workshop (IMSTW)*, 2016, pp. 1–6. DOI: [10.1109/IMS3TW.2016.7524237](https://doi.org/10.1109/IMS3TW.2016.7524237)
- [39] H. Aparicio, P. Ituero, and M. Lopez-Vallejo, "2.64 pj reference-free power supply monitor with a wide temperature range," in *2015 International Workshop on CMOS Variability (VARI)*, 2015, pp. 9–12. DOI: [10.1109/VARI.2015.7456555](https://doi.org/10.1109/VARI.2015.7456555)
- [40] C. Benito, P. Ituero, and M. López-Vallejo, "A low-area reference-free power supply sensor," in *2013 Euromicro Conference on Digital System Design*, 2013, pp. 728–733. DOI: [10.1109/DSD.2013.83](https://doi.org/10.1109/DSD.2013.83)
- [41] H. Aparicio, P. Ituero, and M. López-Vallejo, "Calibration-free 1052 μm^2 power supply monitor," in *2016 12th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, 2016, pp. 1–4. DOI: [10.1109/PRIME.2016.7519495](https://doi.org/10.1109/PRIME.2016.7519495)
- [42] C.-P. Chiang and K.-W. Tam, "Novel cmos voltage sensor with process invariant threshold for passive uhf rfid transponders," in *2008 Asia-Pacific Microwave Conference*, 2008, pp. 1–4. DOI: [10.1109/APMC.2008.4958347](https://doi.org/10.1109/APMC.2008.4958347)
- [43] M. Eberlein, H. Pretl, and T. Krebs, "A time-domain temperature sensor using nwell diodes on 820 μm^2 in 7nm finfet," in *2019 Austrochip Workshop on Microelectronics (Austrochip)*, 2019, pp. 6–9. DOI: [10.1109/Austrochip.2019.00013](https://doi.org/10.1109/Austrochip.2019.00013)
- [44] Y. Lempel, R. Breuer, and J. Shor, "A 700- μm^2 , ring-oscillator-based thermal sensor in 16-nm finfet," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 30, no. 2, pp. 248–252, 2022. DOI: [10.1109/TVLSI.2021.3137338](https://doi.org/10.1109/TVLSI.2021.3137338)
- [45] T. Someya, A. K. M. M. Islam, and K. Okada, "A 6.4 nw 1.7% relative inaccuracy cmos temperature sensor utilizing sub-thermal drain voltage stabilization and frequency-locked loop," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 458–461, 2020. DOI: [10.1109/LSSC.2020.3025962](https://doi.org/10.1109/LSSC.2020.3025962)
- [46] S. Lu, D. Wu, and L. Wu, "Design of a highly compact on-chip temperature sensor with ultra-low power based on 40nm cmos," in *2022 IEEE 16th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)*, 2022, pp. 1–3. DOI: [10.1109/ICSICT55466.2022.9963407](https://doi.org/10.1109/ICSICT55466.2022.9963407)
- [47] H. Park and J. Kim, "A 0.8-v resistor-based temperature sensor in 65-nm cmos with supply sensitivity of 0.28 $^{\circ}\text{C}/\text{V}$," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 3, pp. 906–912, 2018. DOI: [10.1109/JSSC.2017.2788878](https://doi.org/10.1109/JSSC.2017.2788878)
- [48] O. Bass and J. Shor, "A miniaturized 0.003 mm^2 pnp-based thermal sensor for dense cpu thermal monitoring," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 9, pp. 2984–2992, 2020. DOI: [10.1109/TCSI.2020.2987595](https://doi.org/10.1109/TCSI.2020.2987595)

- [49] M. Eberlein and H. Pretl, "A no-trim, scaling-friendly thermal sensor in 16nm finfet using bulk diodes as sensing elements," *IEEE Solid-State Circuits Letters*, vol. 2, no. 9, pp. 63–66, 2019. DOI: [10.1109/LSSC.2019.2938140](https://doi.org/10.1109/LSSC.2019.2938140)
- [50] Y. Shui and A. Wang, "A $14.17\text{pJ} \cdot \text{K}^2$ fom cmos temperature sensor with $173\text{ }\mu\text{m}^2$ sensing core for remote sensing in 65-nm cmos," *IEEE Sensors Journal*, vol. 23, no. 22, pp. 27 059–27 067, 2023. DOI: [10.1109/JSEN.2023.3321741](https://doi.org/10.1109/JSEN.2023.3321741)
- [51] T. Anand, K. A. A. Makinwa, and P. K. Hanumolu, "A vco based highly digital temperature sensor with $0.034^\circ\text{C}/\text{mV}$ supply sensitivity," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 11, pp. 2651–2663, 2016. DOI: [10.1109/JSSC.2016.2598765](https://doi.org/10.1109/JSSC.2016.2598765)
- [52] E. Moisello, P. Malcovati, and E. Bonizzoni, "A compact-area low-power temperature sensor featuring high supply voltage scalability," in *2023 30th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, 2023, pp. 1–4. DOI: [10.1109/ICECS58634.2023.10382756](https://doi.org/10.1109/ICECS58634.2023.10382756)
- [53] K. Kim, H. Lee, and C. Kim, "366-ks/s 1.09-nj 0.0013mm^2 frequency-to-digital converter based cmos temperature sensor utilizing multiphase clock," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 10, pp. 1950–1954, 2013. DOI: [10.1109/TVLSI.2012.2220389](https://doi.org/10.1109/TVLSI.2012.2220389)
- [54] M.-C. Lee and S.-Q. Hong, "Design and implementation of a voltage-controlled oscillator with bandgap voltage reference source and temperature sensing," in *2017 International Conference on Green Energy and Applications (ICGEA)*, 2017, pp. 39–45. DOI: [10.1109/ICGEA.2017.7925452](https://doi.org/10.1109/ICGEA.2017.7925452)
- [55] T. Karpuz, K. Ozanoglu, and S. Can, "A low-voltage, low-power cmos temperature sensor utilizing a single ptat-ctat current loop," in *2025 21st International Conference on Synthesis, Modeling, Analysis and Simulation Methods, and Applications to Circuits Design (SMACD)*, 2025, pp. 1–4. DOI: [10.1109/SMACD65553.2025.11092131](https://doi.org/10.1109/SMACD65553.2025.11092131)
- [56] K. Rajeswari, V. Rahul, B. S. Thrinath, K. Geetha, N. Vishnu, and A. J. Ghorpade, "A fully integrated 90nm cmos temperature sensor for robust industrial applications," in *2025 12th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2025, pp. 1–6. DOI: [10.1109/ICRITO66076.2025.11241543](https://doi.org/10.1109/ICRITO66076.2025.11241543)
- [57] M. Eberlein and H. Pretl, "A current-mode temperature sensor with a $\pm 1.56^\circ\text{C}$ raw error and duty-cycle output in 16nm finfet," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5. DOI: [10.1109/ISCAS51556.2021.9401138](https://doi.org/10.1109/ISCAS51556.2021.9401138)
- [58] C. Zhao, J. He, S.-H. Lee, K. Peterson, R. Geiger, and D. Chen, "Linear vt-based temperature sensors with low process sensitivity and improved power supply headroom," in *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, 2011, pp. 2553–2556. DOI: [10.1109/ISCAS.2011.5938125](https://doi.org/10.1109/ISCAS.2011.5938125)