Robust Circuit Design for Low-Voltage VLSI

by

Yejoong Kim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)
in The University of Michigan
2015

Doctoral Committee:

Professor David Blaauw, Chair
Associate Professor Kenn Richard Oldham
Professor Dennis Michael Sylvester
Assistant Professor Zhengya Zhang

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1

Introduction

The insatiable demand for more integration and performance recently resulted in a 15-core, 30-thread commercial processor with 4.31 billion transistors [1]. The clock frequency, one of the key indicators of chip performance, once reached at 4GHz in a 90nm CMOS process in 2004 [3]. However, it could not follow the trend observed in the transistor count and has remained near-constant over the years [2], where it seems to be saturated in the range of 5$\sim$6GHz. The main reason behind this is the "power-wall" where the excessive power density significantly limits chip reliability and yield as well as the performance and cooling expense [6]; this requires chip designers to consider the power consumption at all design levels.

At the other end of the spectrum lies portable hand-held devices and wireless sensor nodes. Their low power consumption requirement comes from the small form-factor where only a limited-sized battery is available. For example, an intraocular pressure monitoring system [4] shown in Figure 1.1 measures 1.5mm$\times$2mm$\times$0.5mm and includes an 1μAh thin-film battery. Due to this small capacity of the battery, every part of the system has been specifically designed for the target application.

A more general and modular approach to the wireless sensor nodes was introduced in [7], and the system photo is shown in Figure 1.2. It is a 1mm^3 wireless sensor node platform and limits its total volume within 1.4mm$\times$2.8mm$\times$1.6mm, hence only allowing a 0.6μAh thin-film battery on which two ARM Cortex^{TM}-M0 processors and other digital/analog circuits, including sensors, have to reliably operate. Although this system allows stacking many different IC-layers fabricated in different processes using a low-power inter-layer communication bus [5] making it

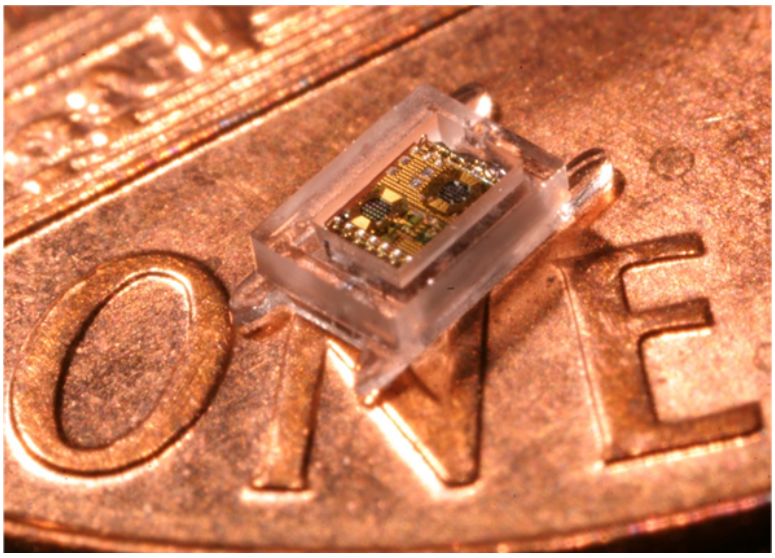


Figure 1.1: A cubic-millimeter intraocular pressure monitoring system [4]

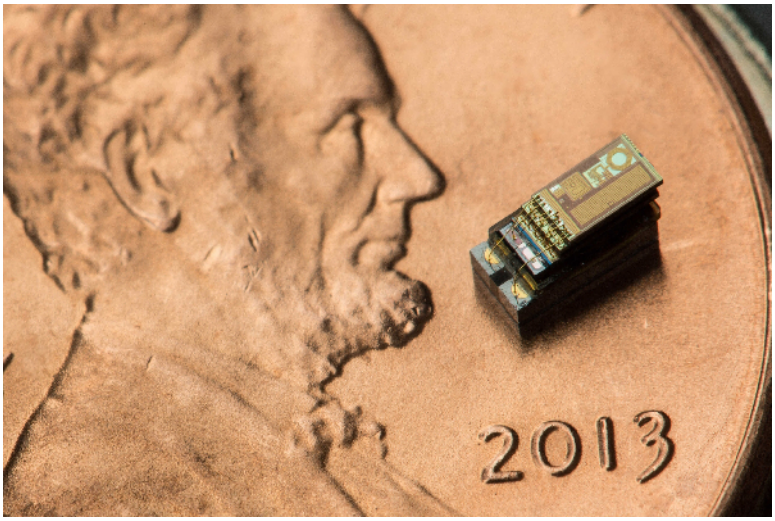


Figure 1.2: A modular 1mm^3 sensing platform [7]

easier to expand system functionality, the severe power constraint requires the entire system to consume less than 40μW active power while utilizing duty-cycled operations with extremely low sleep power (11nW). Therefore, every circuit component in this system must take into account the low-power concerns while still guaranteeing robust system functionality.

Generally, the dynamic power consumption of typical digital circuits can be found as below.

$$P_{dyn} = C_{eff} V_{DD}^2 f_{clk} \tag{1.1}$$

where C_{eff} indicates the effective switching capacitance, and V_{DD} and f_{clk} indicate the supply voltage and the operating clock frequency, respectively. While technology scaling helps reduce the intrinsic capacitance, many circuit techniques have been developed to utilize the quadratic relationship of V_{DD} for effective power reduction.

One of the widely used techniques is dynamic voltage and frequency scaling (DVFS) [8], where the supply voltage and the clock frequency become dynamically adjusted depending on load conditions or operation modes. The effectiveness of DVFS has made this technique quite popular, and many leading institutions and companies have applied DVFS in various types of designs [1][9][10][11][12], where the processors are aimed to achieve power savings without degrading the critical performance. In extremely power-constrained systems, further voltage scaling down to near- or sub-threshold level has been applied. An FFT processor in [13] achieves 90nW of FFT operations by lowering the supply voltage to 180mV, which is at the sub-threshold level in the standard 0.18μm CMOS logic process used in the work. Obviously, the lower supply voltage indicates lower power consumption as shown in Eq. (1.1). However, this lower power does not necessarily mean 'lower energy.' As the supply voltage becomes lower, the maximum achievable clock frequency becomes also degraded due to the reduced device on-current (I_{ON}). The slower operating frequency (i.e., a longer clock period) increases the leakage energy per cycle, hence reducing the ratio of the dynamic energy to the leakage energy. Therefore, there exists a minimum energy point where further voltage scaling does not reduce the overall energy consumption due to the dominating leakage energy. As a result, the FFT processor above achieves the minimum energy point at 350mV with 155nJ/FFT, whereas the minimum voltage point at 180mV consumes more than 1μJ/FFT.

This minimum energy point typically occurs at a voltage slightly lower than the device threshold voltage (hence, sub-threshold). However, researchers found that the energy reduction is only $\sim 2\times$ when V_{DD} is scaled from the near-threshold regime to the sub-threshold regime, whereas delay increases by 50–100$\times$ over the same region [14]. Thus, for many applications, the near-threshold regime can be a better choice than the sub-threshold in terms of an energy-delay trade-off, and the near-threshold computing (NTC) has become an attractive solution for low-power VLSI systems [15][16][17].

However, there are several issues in the NTC operations [14]. First, the lower supply voltage

significantly degrades the performance, although this could be compensated by parallelism to some extent. Second, NTC exhibits degraded process/voltage/temperature (PVT) variations. In the NTC region, the MOSFET drive current has an exponential dependency on the supply voltage (V_{DD}), device threshold voltage (V_{TH}), and temperature. Thus, even a small amount of variation can lead to a severe yield reduction especially in ratioed designs in which the circuit functionality depends on a relative device sizing. Therefore, proper circuit-level techniques have to be applied for low-voltage VLSI.

In this dissertation, we identify several circuit components that are critical to low-voltage VLSI operation and propose new and advanced techniques to improve their robustness and performance. A typical architecture of low-voltage VLSI systems is shown in Figure 1.3; level converter circuits, SRAM, and clocked sequential elements are highlighted, and each will be discussed in detail in the following chapters.

Level converters are one of the main concerns especially in aggressively voltage-scaled systems. Typically, digital cores operate at low supply voltages to save the power, but other peripherals are not always able to be run at such low voltages. For example, it is hard to apply the voltage scaling technique to analog circuits due to the reduced voltage headroom (hence reduced margins/offsets). Also, I/O voltages are not very well scalable due to the noise concerns. Thus, level converters are required at the interface between the low-voltage digital core and the high-voltage analog and peripherals. However, as the cores become deeply voltage-scaled, the voltage difference between the low voltage (V_{DDL}) and the high voltage (V_{DDH}) becomes larger. Especially for the core running in the NTC region, the reduced I_{ON}/I_{OFF} ratio makes it extremely difficult to achieve robust level conversions. The use of native-V_{TH} (or zero-V_{TH}) devices in [18] improves robustness by allowing to use thin gate-oxide devices (i.e., more stronger devices) for pull-down, but still, other techniques are required to further achieve a good performance, lower energy consumption, as well as a good yield. A well-known approach to improve the robustness is weakening the pull-up strength or strengthening the pull-down. For example, [19] uses PMOS diodes to weaken the pull-up strength, and [20] and [22] include reduced-swing inverters. A dynamic level converter can improve the speed and the robustness at the cost of extra power and a complicated synchronization circuit [21]. In Chapter 2, we will propose new static level converter circuits and a quantitative design method to guarantee robustness.

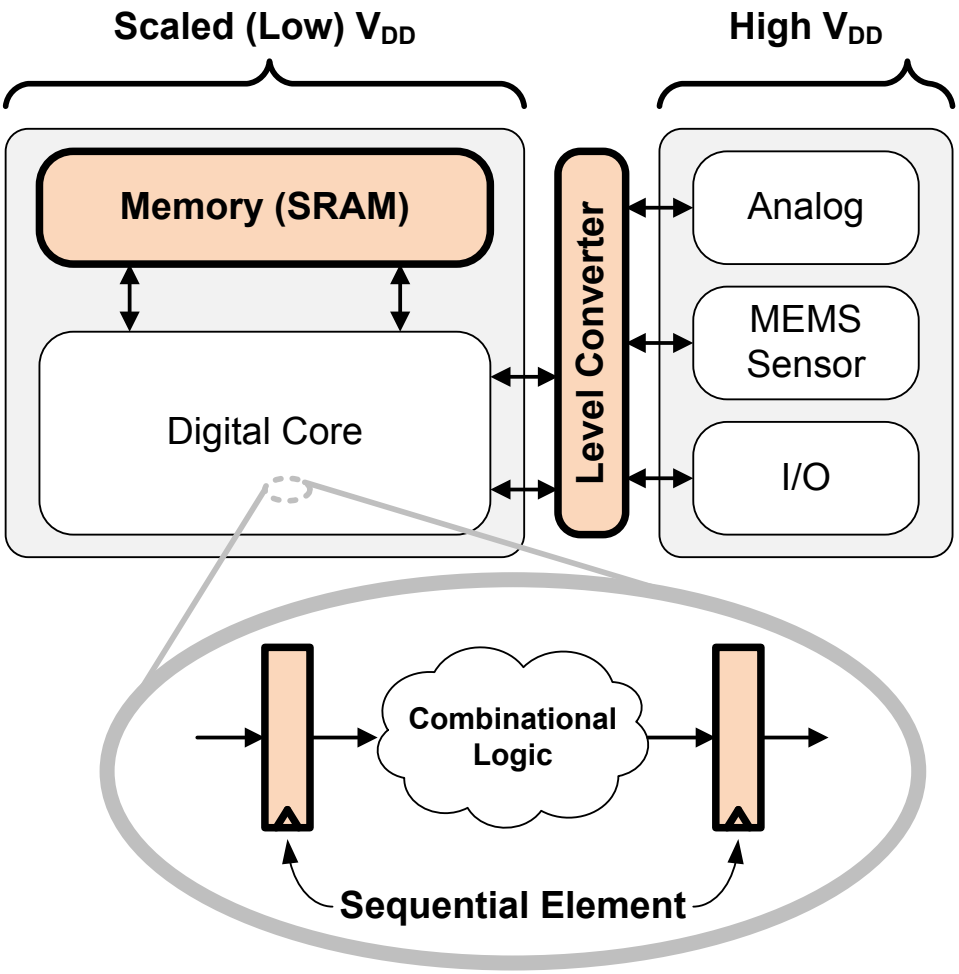


Figure 1.3: A typical architecture of low-voltage VLSI systems

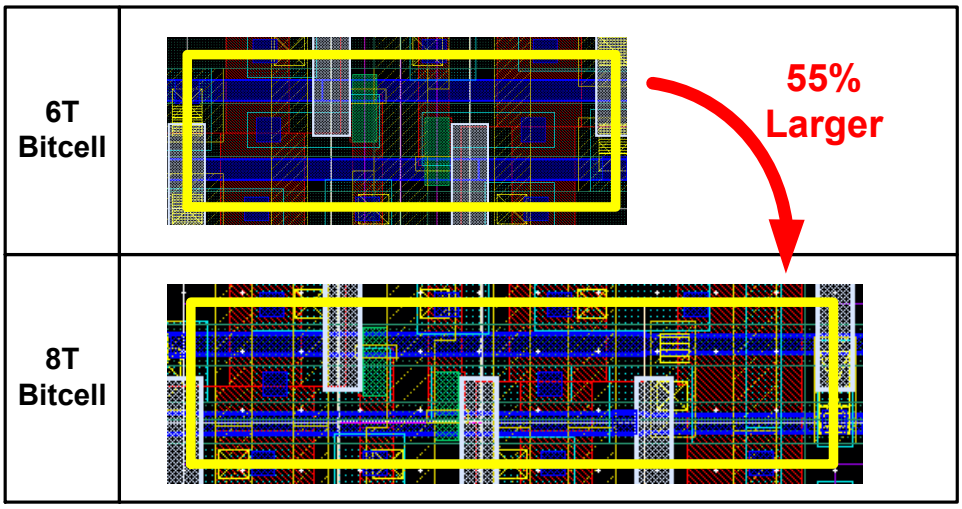


Figure 1.4: Bitcell size comparison between commercial 6T and 8T

SRAMs are one of the major bottlenecks in the voltage scaling [23]; the standard 6T bitcell requires the ratioed device sizing, and the two-sided constraint (READ and WRITE) significantly degrades the robustness at the low voltage regime. Using 8T bitcells decouples READ and WRITE operations, making it possible to separately optimize the two operations at the cost of a larger bitcell area [24][25][26]. Generally, 8T bitcells have a $30 \sim 55\%$ area penalty compared to the standard 6T bitcell, and one of the examples in an advanced technology node is shown in Figure 1.4. This significant area overhead makes the 8T bitcell unacceptable in severely area-constrained applications. In the NTC region, the functionality of the bitcell is further impacted due to the aggravated PVT variations. Thus, in this case, even the 8T requires assists from extra peripheral circuits for correct functionality [26][27], or a bitcell with more number of devices is preferred such as 10T bitcells in [7] and [28]. Recently, 7T bitcells have been proposed in [29] and [30]; they are supposed to have a smaller bitcell size than the 8T bitcell while still providing the similar robustness (i.e., decoupled READ and WRITE). In Chapter 3, we will address issues in the 7T structure and propose a new solution, still fully utilizing inherent advantages of the 7T.

The next key component is the clocked sequential element, called a flip-flop in short. Flip-flops are one of the critical components in today's digital processors. For example, both of POWER7^{TM}and SPARC T4 processors have more than 2 million flip-flops, taking up to 20% of the total core power [31][32] as shown in Figure 1.5. Mainly because of its importance in digital circuits, numerous flip-flop designs have been investigated and proposed [33][33]. The main issue

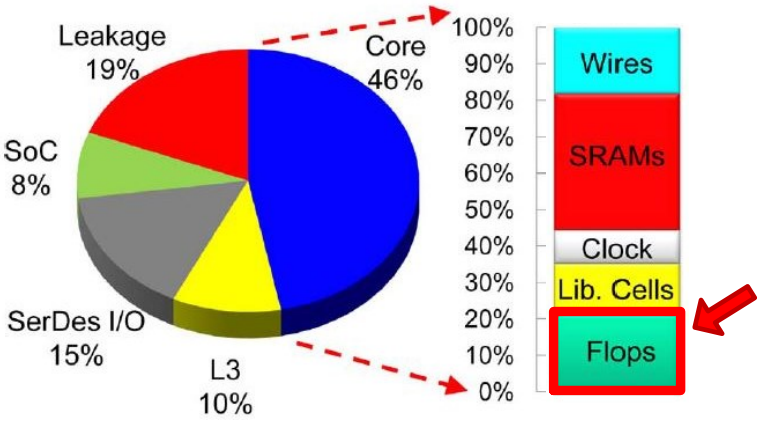


Figure 1.5: Power breakdown of SPARC T4 processor [32]

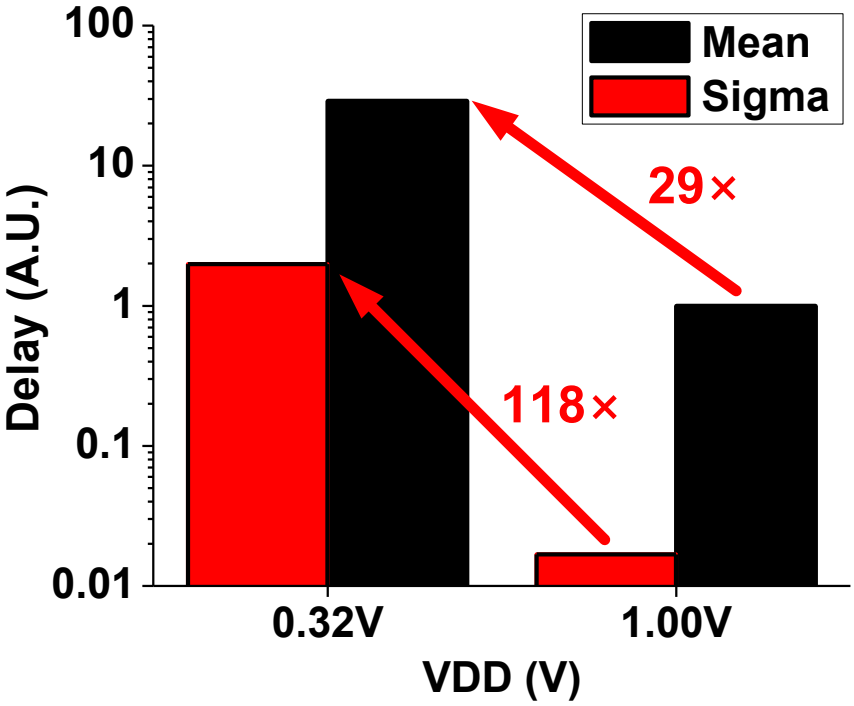


Figure 1.6: Normalized unit-FO4 delay measurement in 45nm

of the conventional flip-flops in the NTC region is the degraded hold-time variation [38], which requires excessive buffer insertions to meet the hold-time margin under severe PVT variations. In Chapter 4, we will further discuss issues in conventional flip-flops in literature [35][36][37], and propose a new flip-flop that is static, single-phase, and contention-free, which also provides a $\sim 40\%$ power reduction compared to the conventional flip-flop.

The last topic in this dissertation is a testing harness for flip-flop timing characterization. Rep-

resentative timing parameters of flip-flops are usually setup-time (T_{SETUP}), hold-time (T_{HOLD}), and C-Q delay (T_{CQ}). These parameters are usually in the range of $1 \sim 5$ FO4 delay, so an accurate Time-to-Digital Converter (TDC) is required to measure such a short delay. In addition, a more difficult problem arises in that those parameters are usually determined by mismatches in devices used to implement the flip-flops. At full V_{DD} level, those mismatches can be minimized by up-sizing transistors and careful layout techniques, but it is almost impossible to achieve the same measurement accuracy in low V_{DD} due to the severe variations mentioned earlier in this chapter. For example, Figure 1.6 shows that the standard deviation of measured unit-FO4 delays in 45nm degrades by $118\times$ when going from 1.0V to 0.32V, while the average (mean) value is increased by only $29\times$. These variations have severer effects in complicated circuits, and in the flip-flop timing characterization, they often cause a large offset in measurements. In Chapter 5, we will propose effective techniques to eliminate the measurement offsets incurred by the mismatches and provide setup/hold-time measurements at near-V_{TH} to demonstrate the benefit of the new flip-flop introduced in Chapter 4.

Finally, in Chapter 6, we will conclude this dissertation by summarizing the proposed circuits and discussing possible future works.

CHAPTER 2

Robust Level Converter Circuits for Wide-Range Voltage Conversion

2.1 Introduction

Low-voltage circuit design has been widely investigated for ultra-low power applications, reaching as low as 230mV in a recent multi-pipelined processor [39], and requiring wide-range level conversion for communication with I/O pads and high-voltage circuit blocks. In addition, cores on a chip multiprocessor are increasingly voltage scaled independently [9], necessitating level conversion between core voltage domains in high performance applications. Another example is a multi-core system in [41], which suggests an optimal voltage/frequency mapping among the cores and requires thousands of level converters (LCs).

LCs become more critical as the voltage difference grows, for instance, between aggressively voltage-scaled DSP accelerators [13] and I/O. An extreme case is the wireless sensor node platform in [7], where the core is operated at a sub-threshold level while sensors and radio use the battery voltage (3.6V). Due to such significant voltage differences, these applications require wide-range LCs with fast and low power operation. However, level conversion is challenging at reduced voltages since conventional approaches suffer from severe contention between weak pull-down devices and strong pull-up devices, making them vulnerable to process / voltage / temperature (PVT) variations. Also, LCs in many sensing applications, such as environmental monitoring, will be exposed to extreme conditions, exacerbating robustness challenges in the LCs.

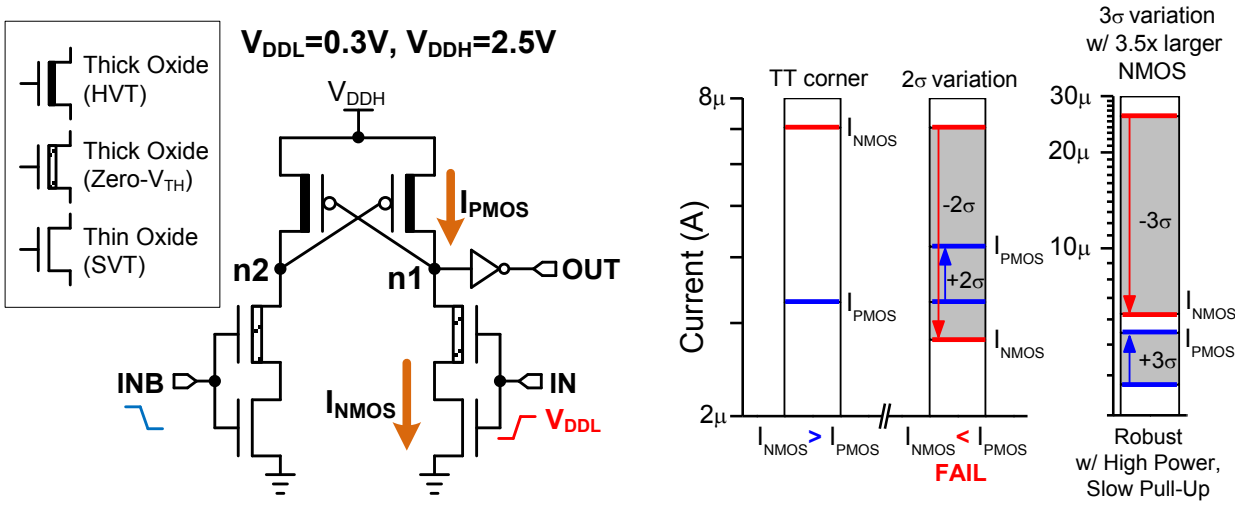


Figure 2.1: DCVS LC and its current margin plots

In this chapter, we will present two robust level converters, called Limited-Contention Level Converter (LC^2) and Split-control Level Converter (SLC), respectively. Operation details and measurement comparisons are following.

2.2 LC^2: Limited-Contention Level Converter

2.2.1 DCVS Level Converter and Its Current Margin

Figure 2.1 shows the operation of a conventional Differential Cascode Voltage Switch (DCVS) approach. A zero-V_{TH} device prevents oxide breakdown in the thin oxide devices, making it possible to use a fast standard-V_{TH} (SVT) pull-down device [18]. The DCVS LC suffers from a two-sided constraint on the PMOS device: if the PMOS is too weak, the pull-up transition becomes slow and the node may not be kept high, giving rise to performance and robustness issues; if the PMOS is too strong, the NMOS cannot overcome it and the circuit fails. The current margin plots in Figure 2.1 show that severe variations at the low voltages exacerbate this two-sided constraint. Although the circuit is designed such that $I_{NMOS} >> I_{PMOS}$ to discharge node *n1* or *n2*, as little as 2σ V_{TH} variation causes failure due to $I_{NMOS} < I_{PMOS}$. Increasing NMOS size by 3.5× guarantees 3σ robustness, but results in very large devices (W_{NMOS} = 105μm) with undesirable leakage (9nA). In addition, the increased diffusion capacitance slows the pull-up transition. This two-sided con-

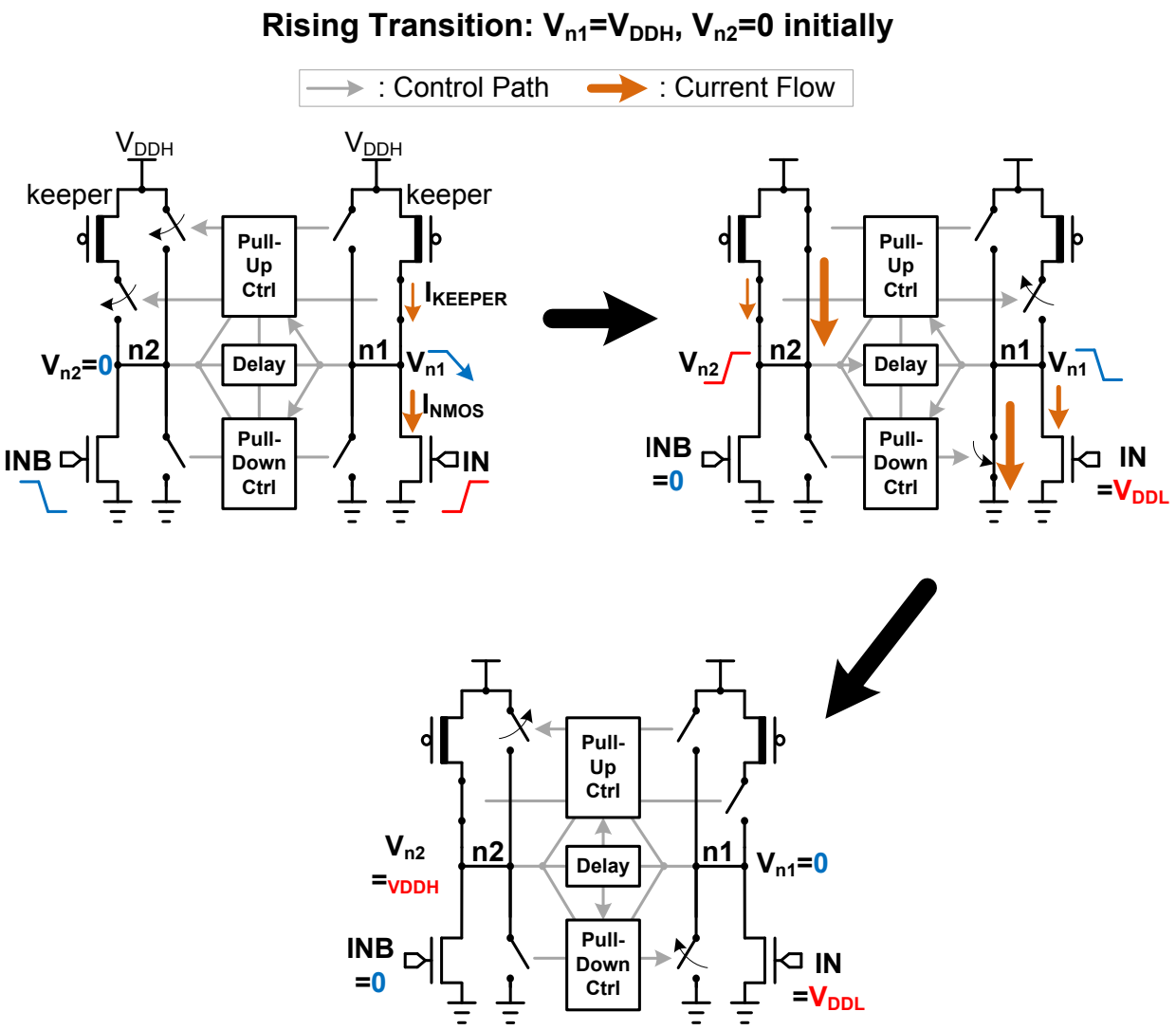


Figure 2.2: LC^2 operation

straint severely limits DCVS LC robustness under PVT variation. Multiple LC stages can improve robustness but introduce overhead due to intermediate supplies and increased latency. Other static LCs [19][20] have similar two-sided constraints and require precise transistor sizing, and have lacked silicon measurements. A recently proposed dynamic LC [21] uses a high-voltage clock, which improves robustness but increases layout size and power consumption. Furthermore, none of the previous LCs has demonstrated robustness through comprehensive silicon measurements.

2.2.2 Operation of LC^2

We propose a new approach called Limited Contention Level Converter (LCLC or LC^2) that eliminates the two-sided constraint without the use of high-voltage clocks. Figure 2.2 shows the conceptual operation of LC^2. Before the rising transition, node *n1* is held high by the weak keeper, which is sub-threshold-biased, while all other switches are off; hence $V_{n1} = V_{DDH}$ and $V_{n2} = 0$. Once V_{IN} rises to V_{DDL}, the pull-down driver starts to discharge *n1* and easily overcomes the weak keeper. This transition on *n1* causes "Pull-Up Control" to activate both the weak keeper and the strong switch on the other side, which quickly charges up *n2*. "Pull-Down Control" is then triggered to directly connect *n1* to ground, rapidly discharging it and completing the transition. Finally, a delay element turns off all switches (except the appropriate keeper) after all transitions are finalized. The next transition can then proceed such that the only contention is with the weak keeper. The use of separate and different strength pull-up devices for holding state and charging/discharging *n1* and *n2* substantially improves design robustness and performance.

Figure 2.3 shows the schematic of LC^2 with detailed timing waveforms. At the beginning of a rising transition, $V_{n1} = V_{n3} = V_{DDH}$ and $V_{n2} = V_{n4} = 0$, hence M6 and M11 are off and M1 contends only with the weak keeper Mx. Once M1 and M3 start to discharge *n1*, positive feedback from M10 and M7 boosts transition speed by pulling the gate of M7 to V_{DDH}. Thus, M10 can be sized for fast rising transitions on *n2* (using a min length device). In contrast, this transistor must remain weak in the conventional approach to minimize the contention, making it slower and less robust. Once the transition completes, M5 and M12 are turned off after an inverter chain delay to prepare for the next transition. Devices M5–M12 use minimum width, and the inverter chains simply require sufficient delay to fully charge *n1* or *n2*, simplifying device sizing. Although the pull-down drivers (M1 and M2) and keepers should be carefully sized, keeper size can be easily determined using known techniques [40], after determining M1 and M2 sizes based on the desired speed-power trade-off. A simple diode chain is used to generate the keeper voltage (V_{KEEPER}), setting the current supplied by the keeper. The current margin plot in Figure 2.4 shows that this design is robust to $> 3\sigma$ variation in simulation. Simulation results in Figure 2.5 indicate that DCVS is highly vulnerable to V_{TH} shifts, while LC^2 functions correctly within the entire process corner without significant delay change. Note that the vertices of polygons represent the pre-defined process corners (FF, FS,

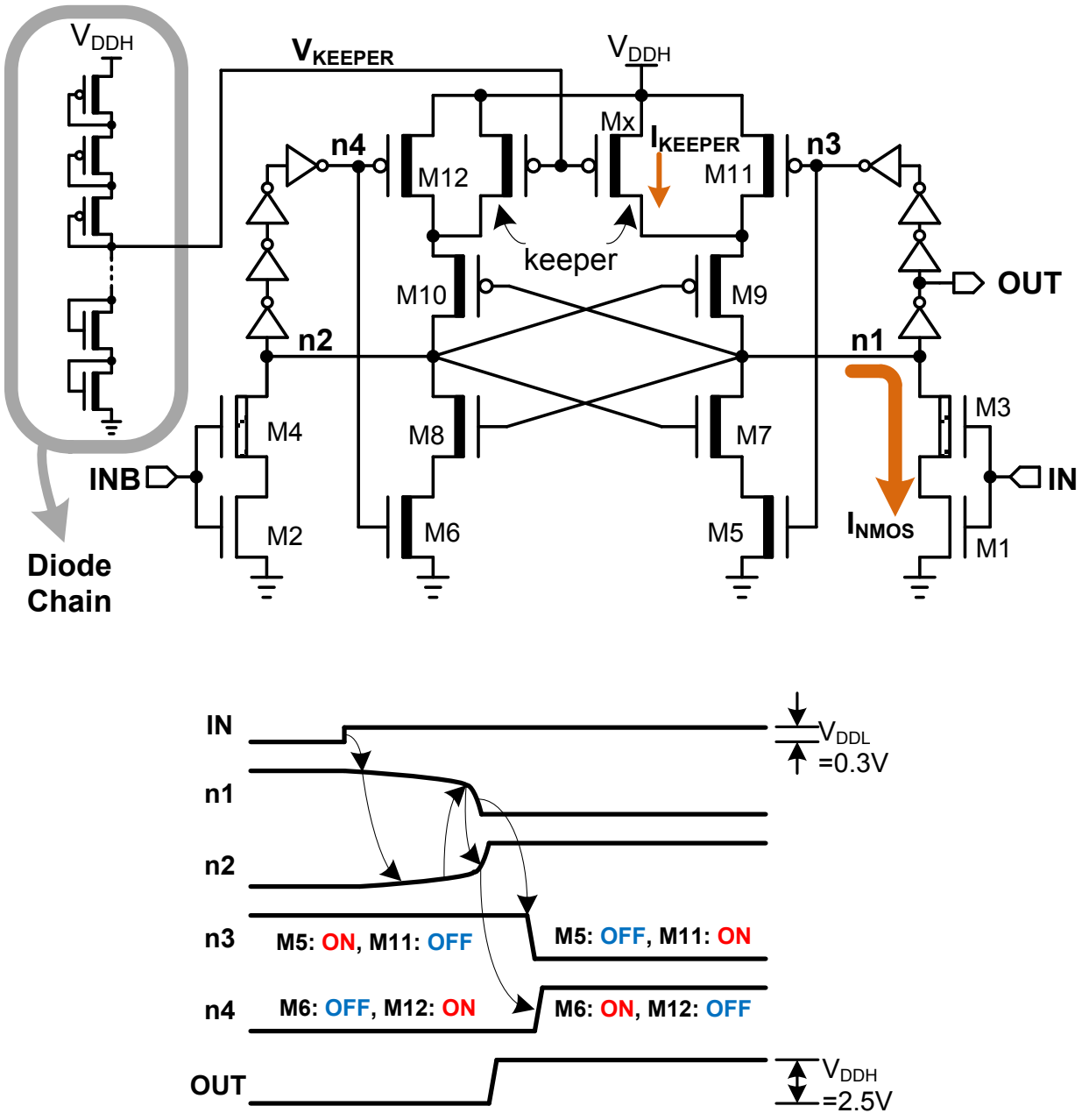


Figure 2.3: LC^2 schematic and its waveforms

SF, SS) of the specified devices in the figure. White regions indicate a delay larger than 10 FO4 or functional failure.

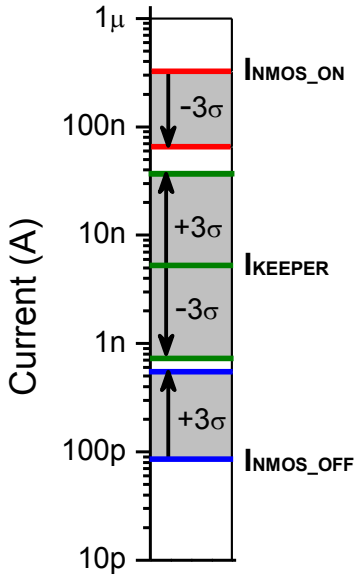


Figure 2.4: LC^2 current margin plot

2.2.3 Measurements

We measured 40 dies in 130nm CMOS; each die has two LC^{2}s and two DCVS LCs designed for 0.3V to 2.5V conversions (V_{DDL}=0.3V, V_{DDH}=2.5V) with a minimum-sized inverter as an output load. Figure 2.6 shows measured delay across temperature. LC^2 is 3.2× faster than DCVS with 2.38 FO4 delay at 25°C (FO4 measured at V_{DDL} supply and corresponding temperature). In addition, DCVS shows a 10.4× delay change across 10 ∼ 100°C, while LC^2 changes by only 4.3×. Normalizing to FO4 delays, LC^2 delay increases 18% from 10 to 100°C while DCVS worsens by 104%. This is due to the much reduced contention in LC^2. Figure 2.7 shows measured power consumption across temperature. While DCVS consumes 7.15nW static power, LC^2 consumes 15× less (475pW) at 25°C, mainly due to the smaller pull-down device (1.5μm). It consumes 2.29nW active power at 25°C which is 4.9× less than DCVS (11.21nW), as well as nearly constant active power over a wide temperature range. Due to the lack of contention, its active energy is dominated by charging of capacitances rather than short-circuit current as in DCVS, making it temperature insensitive. Active power changes only 2% (from 2.27nW to 2.32nW) in the 10 ∼ 100°C range while DCVS shows a 7.7× change (from 4.15nW to 31.88nW) and high power consumption at low temperature. Unlike LC^2, not all 80 DCVS LCs function below 10°C since the low temperature increases V_{TH}, weakening the NMOS exponentially and the PMOS linearly, exacerbating

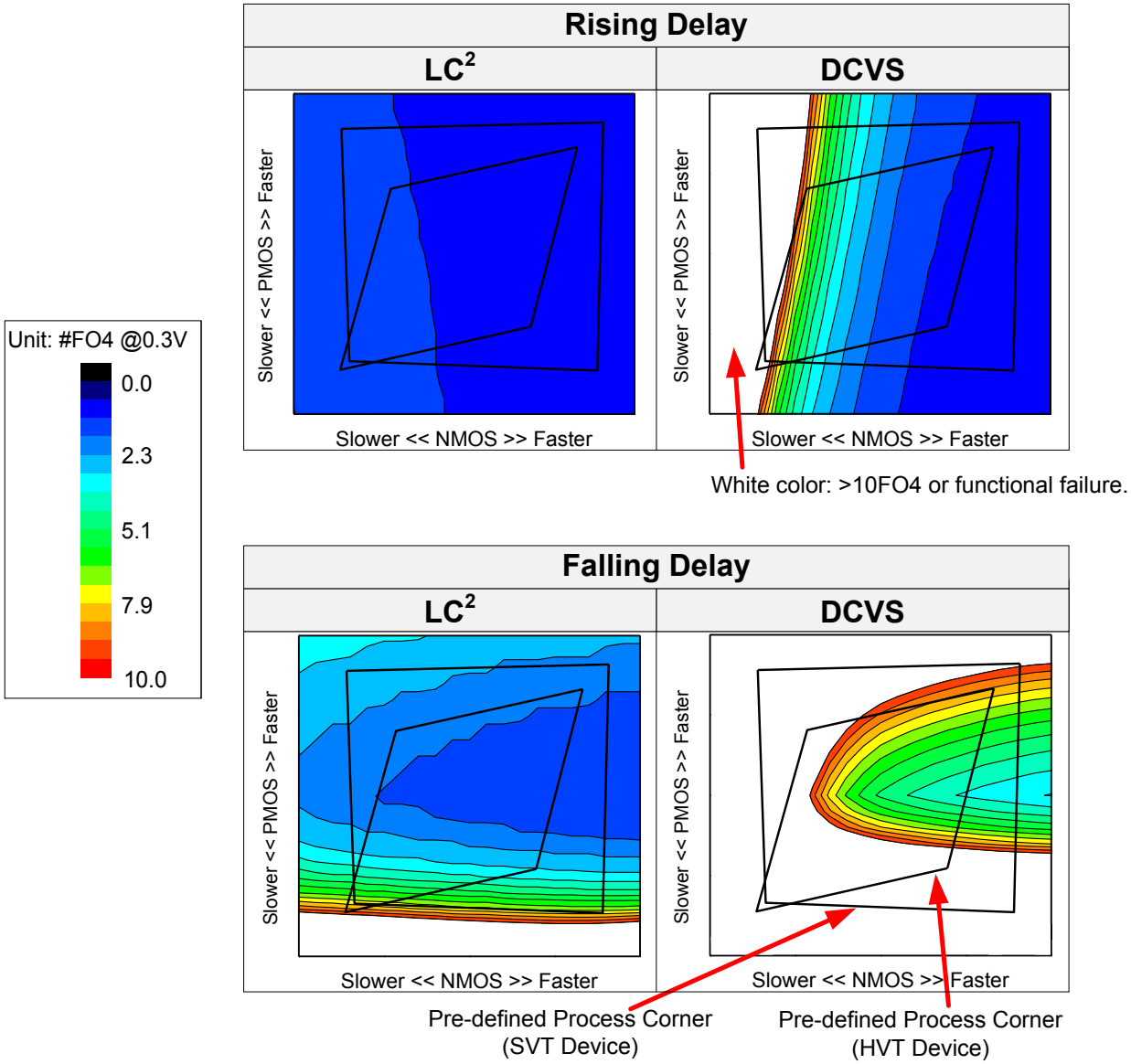


Figure 2.5: Simulation results of LC^2 and DCVS LC

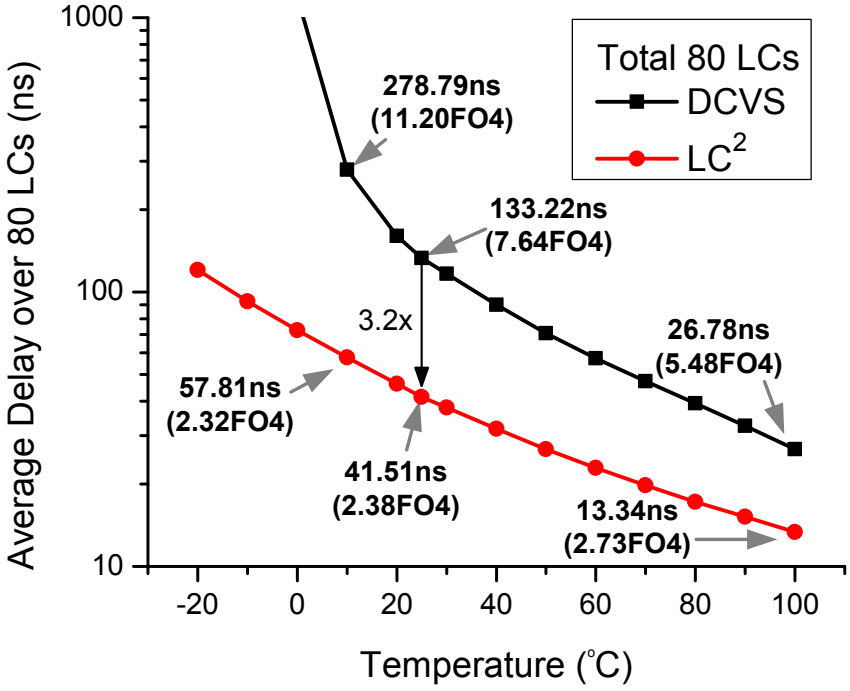


Figure 2.6: Measured delay compared to DCVS LC

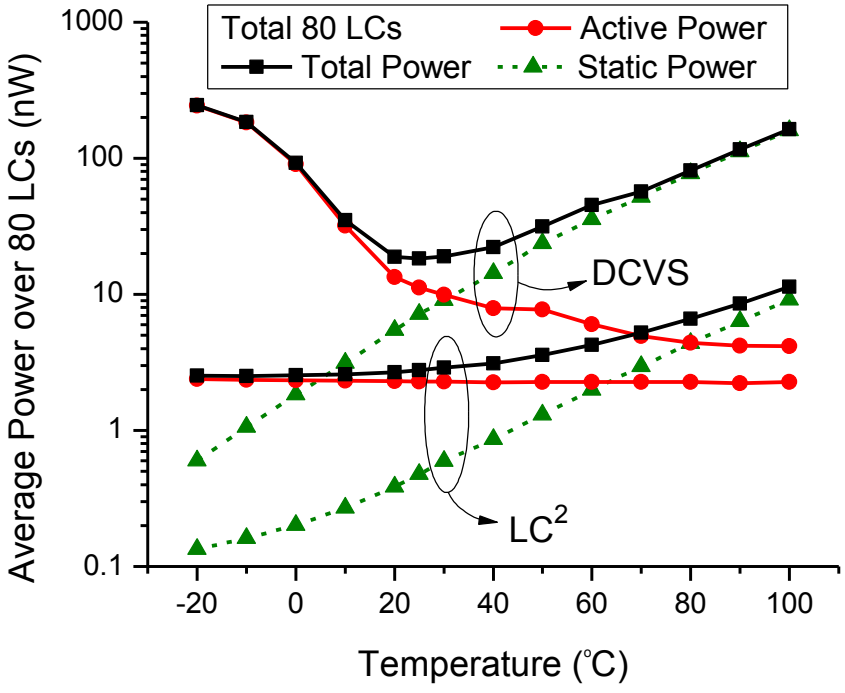


Figure 2.7: Measured power consumptions (freq=5kHz, α=2)

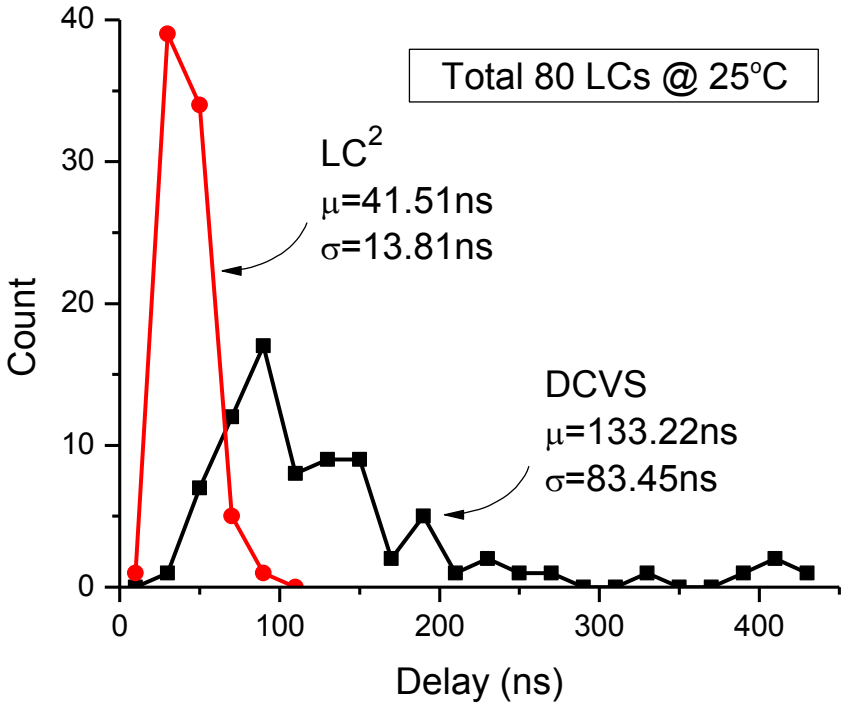


Figure 2.8: Measured delay variations

contention.

To show the impact of process variations, Figure 2.8 displays measured delay distributions for the LCs at 25°C. LC^2 shows 6× smaller standard deviation than DCVS. For voltage variations, Figure 2.9 shows performance degradations across voltage drop. While DCVS delay increases by 7.7× with 10% V_{DDL} drop, LC^2 slows by only 6% (normalized to FO4 delays at the corresponding voltages), indicating that the keeper sizing strategy is sufficiently robust to handle expected voltage variations. Figure 2.10 shows the number of operating LCs at 1MHz across temperature. DCVS was designed to operate as fast as 20MHz at 25°C, and the 1MHz clock allows 20× delay degradation. While all LC^2s operate reliably in the $-20 \sim 100$°C range, the first DCVS fails at 20°C, and only 5 of 80 work at −20°C, showing the robustness of LC^2 to PVT variations.

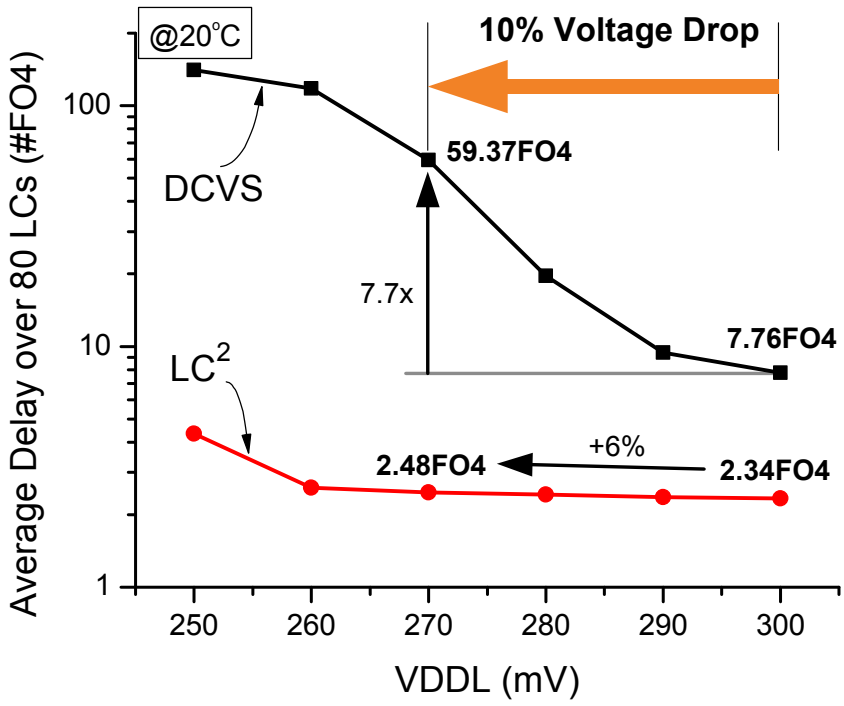


Figure 2.9: Impact of voltage fluctuations

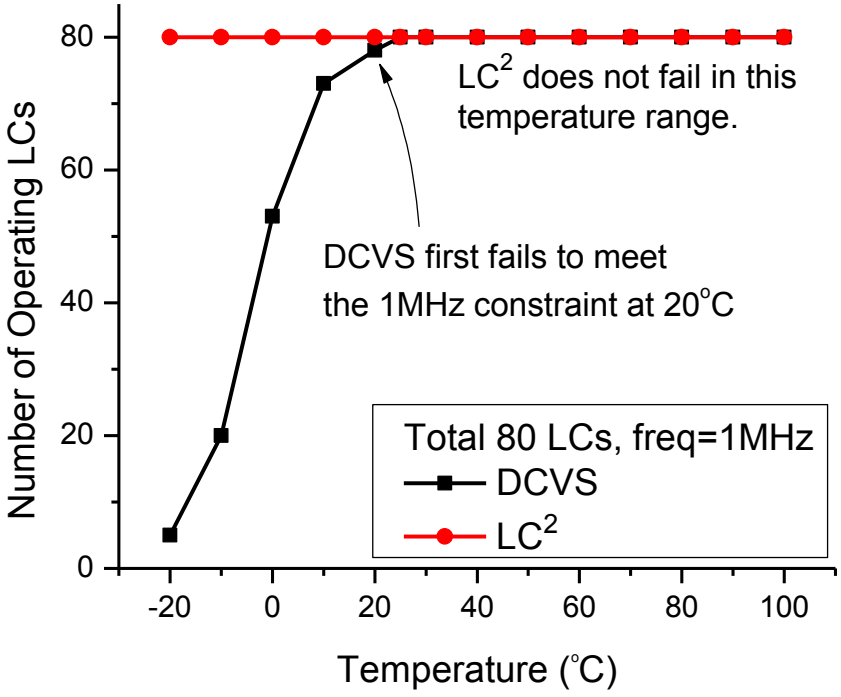


Figure 2.10: Number of operating LCs over temperature

2.3 SLC: Split-Control Level Converter

2.3.1 Previous Level Converters

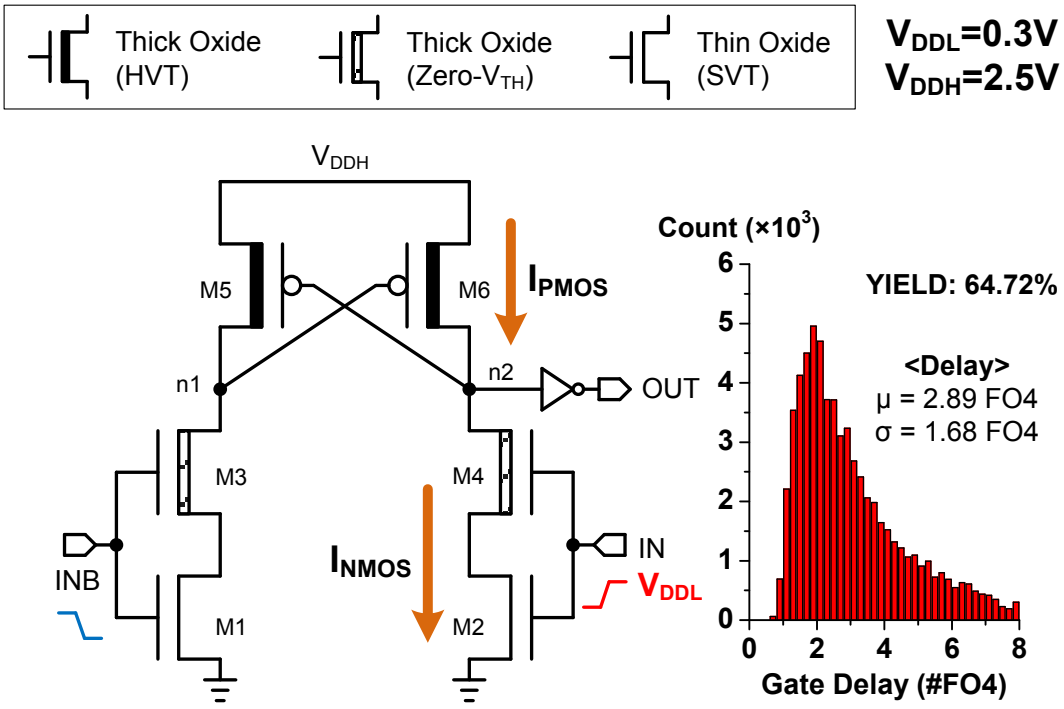
LC^2 introduced in the previous section shows robust level conversion with superior performance and power. However, in systems requiring thousands of LCs, the area of LC^2, which is comparable to the conventional DCVS LC, could become a limiting factor. Hence, a smaller (and probably simpler) LC can be beneficial in those applications.

As already discussed, DCVS LC shows poor robustness; in Figure 2.11(a), its yield is only 64.72% over 100,000 Monte Carlo simulations at 25°C even with very large pull-down devices of $(W/L)_{M1,M2} = 30\mu m/0.12\mu m$. The interrupted DCVS LC in Figure 2.11(b) has an additional PMOS M7 (or M8) that is expected to be weakened when $V_{INB} = V_{DDL}$ (or $V_{IN} = V_{DDL}$), thus reducing $I_{PMOS,ON}$. However, this is not effective for $V_{DDL} << V_{DDH}$ since $|V_{GS}|$ of M7 (or M8) remains close to V_{DDH}. Monte Carlo simulations show only marginal improvement over conventional DCVS in this case. Previously proposed LCs either use a sensitive sub-threshold analog circuit — i.e., a Reduced Swing Inverter — which has not been fully demonstrated in silicon [20][22], or a high voltage clock ($V_{CLK} = V_{DDH} = 2.5V$) that results in high power consumption and a complex synchronization circuit [21], causing 1016× larger layout size than the conventional DCVS LC.

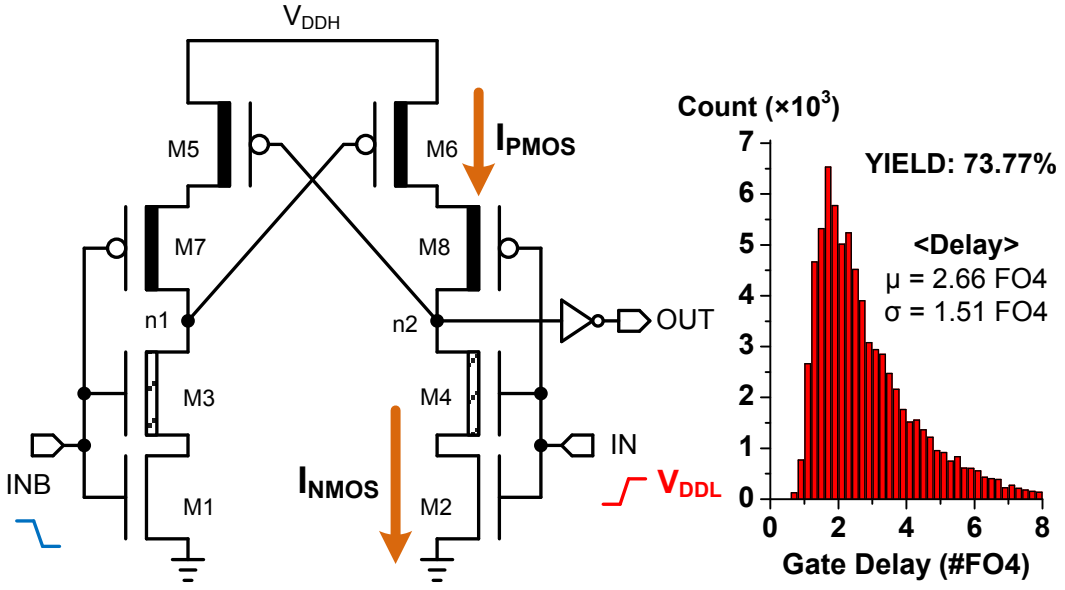
The LC in [19] is shown in Figure 2.12 and includes zero-V_{TH} devices and additional PMOS diodes to tolerate 0.3V to 2.5V conversion in 130nm CMOS. The diodes (M9–M12) serve as current limiters, effectively reducing $I_{PMOS,ON}$ and hence improving robustness. However, they also prevent nodes *n3* and *n4* from fully discharging to ground, hence this design requires additional pull-down devices (M5–M8) that add internal node capacitance. Thus, discharge speed at *n4* (or *n3*) is slow, causing short-circuit current in the output inverter. Also, *n1* (or *n2*) is never fully charged to V_{DDH} due to the diode voltage drop (V_D) and causes static near-threshold current as depicted in the figure.

2.3.2 Operation of SLC

Figure 2.13 shows the proposed LC, named Split-Control Level Converter (SLC). It includes a new output structure (M11 and M12) to avoid the aforementioned problems. At the beginning



(a)



(b)

Figure 2.11: (a) Conventional DCVS LC with Monte Carlo simulation result, (b) Interrupted DCVS LC with Monte Carlo simulation results

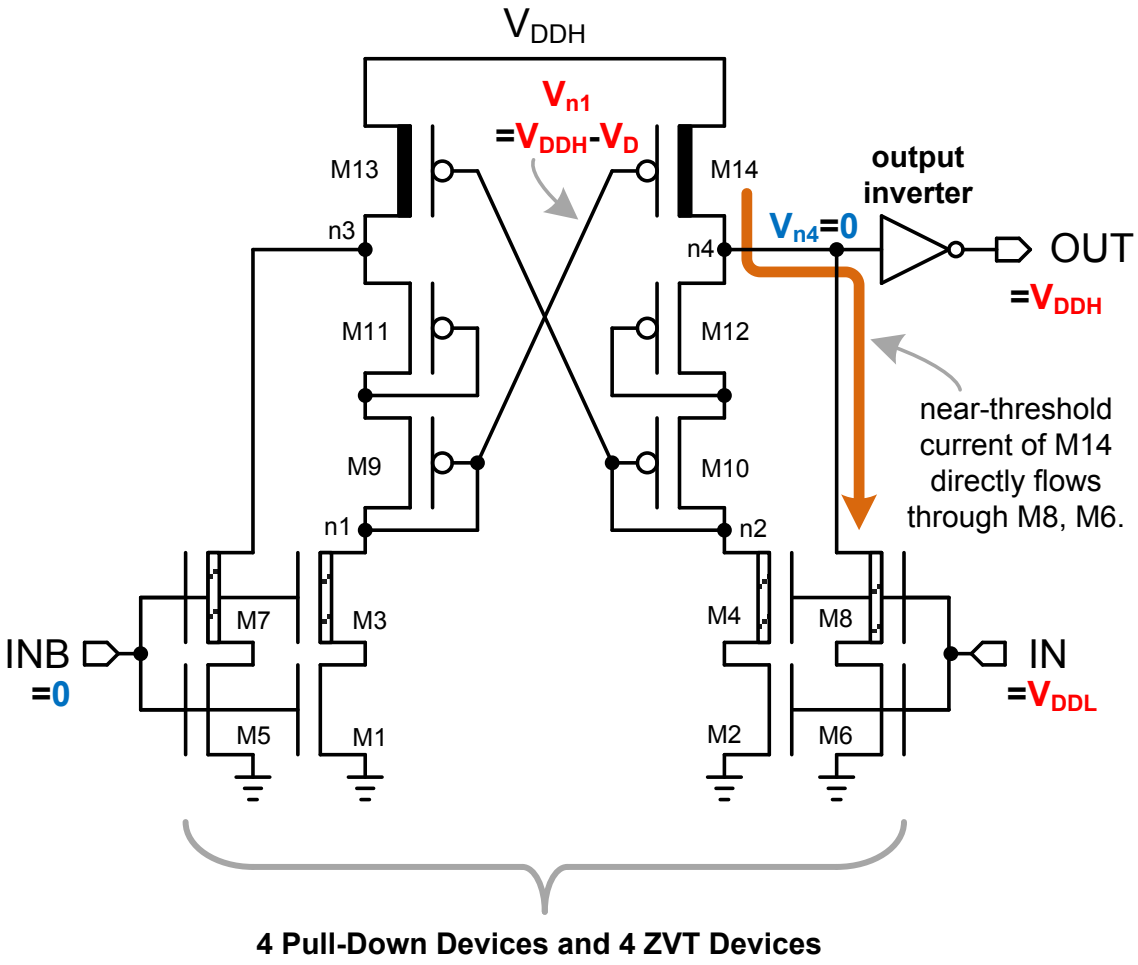


Figure 2.12: Level converter in [19]

of a rising transition at IN, $V_{n1} = 0$ and $V_{n2} = V_{DDH}V_D$, where V_D represents the diode voltage drop through M6/M8 (or M5/M7). Once V_{IN} goes high to V_{DDL}, M2 can easily discharge node *n2* because of the current-limiting diodes. Node *n4* is also discharged to V_D, and M11 is strongly on with a large $|V_{GS}|$, quickly charging up the output node while M12 is completely off. The circuit does not require the additional pull-down paths that contain the largest devices in the circuit, which results in at least 1.8× lower static power across process corners as shown in Figure 2.14(a). This also results in reduced internal loading at *n4* and *n3*, speeding transitions at these nodes. In addition, M11 and M12's gate voltages are separately controlled in the output buffer (hence the name Split-control LC). This configuration ensures that the transistor turning off in the M11–M12 stack always leads the transistor turning on, reducing short circuit current significantly and also improving the charging (or discharging) speed. Overall, Figure 2.14(b) shows that the circuit

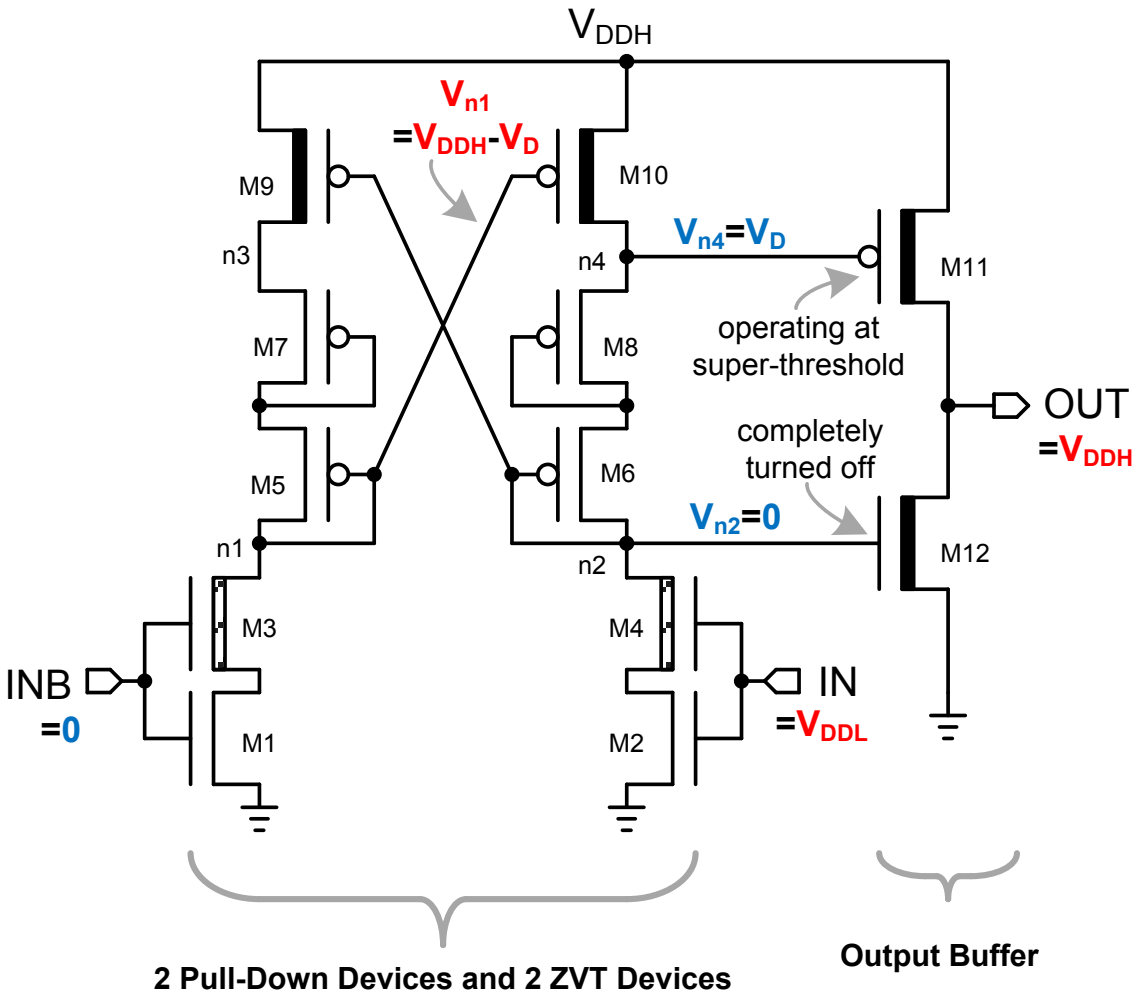


Figure 2.13: SLC schematic

provides a 3.8–12.9× reduction in short-circuit energy consumption across process corners. Monte Carlo simulations show high yield (98.93%) with much lower delay variability (Figure 2.14(c)). Compared to the LC in [19] which has μ = 2.02 FO4, σ = 0.79 FO4, SLC has improved the delay because of the output buffer.

2.3.3 Measurements

We compare SLC to the conventional DCVS rather than the design in [19], since the four zero-V_{TH} devices in the LC of [19] make it slower than conventional DCVS at $> 25°$C due to increased internal loading. The minimum size requirement of zero-V_{TH} devices also makes it comparable to the size of the large pull-down devices in DCVS, such that the LC in [19] has only 17% smaller

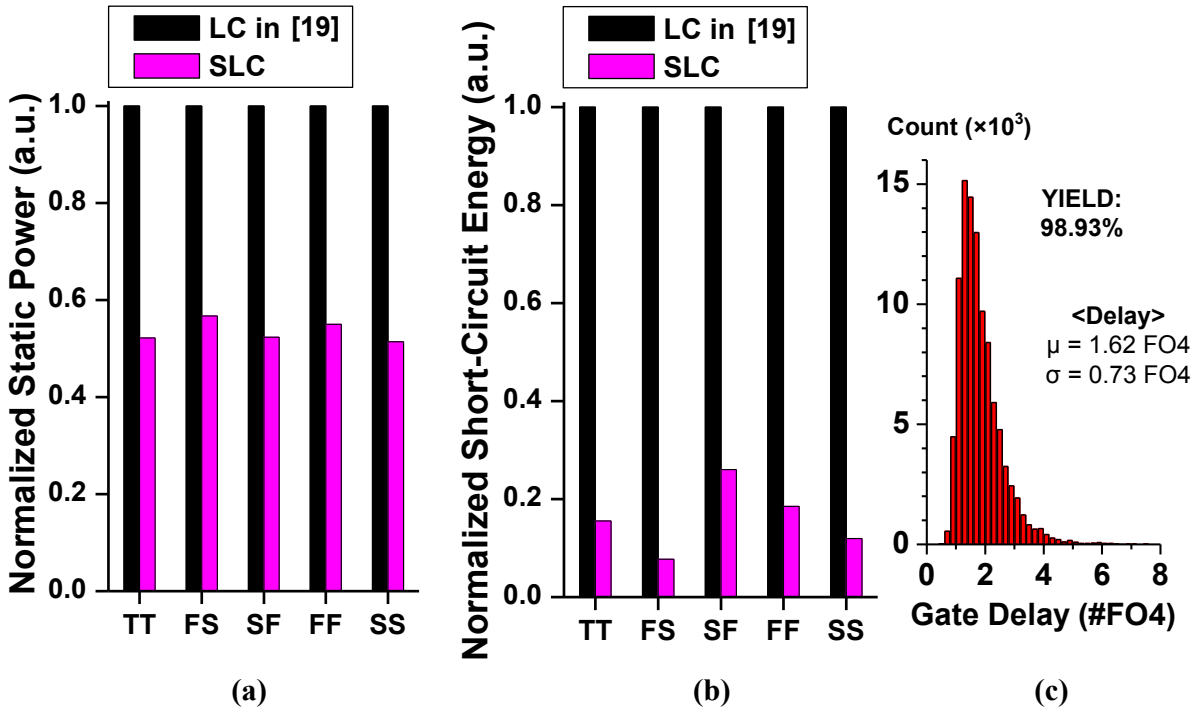


Figure 2.14: (a)(b) Comparisons between LC of [19] and SLC, (c) Monte Carlo simulations of SLC

layout size than DCVS despite the use of 15× smaller pull-down devices. Hence, DCVS provides a more challenging comparison point. We measured 40 dies in 130nm CMOS; each die had two DCVS LCs and two SLCs, providing 80 LCs for each type. The LCs were designed for 0.3V to 2.5V conversion. Also, we used the simulated unit-FO4 delay to convert measured delays into FO4 delays. The unit-FO4 delay was simulated at V_{DDL} and the corresponding temperature.

Figure 2.15(a) shows that SLC has a delay of 3.37 FO4 at 25°C, 2.3× faster than DCVS. Normalized to FO4 delay, SLC delay varies by only 9.5% over 10–100°C, while DCVS changes by more than 2×. In Figure 2.15(b), the new design has 9.9× lower static power at 25°C, mainly due to the smaller pull-down devices. Also, active power is 5.9× lower than DCVS, demonstrating the benefits of reduced contention. Across 10–100°C, the active power of SLC varies by 33%, while DCVS exhibits 7.7× variation over the same range.

Figure 2.15(c) shows that SLC has a 5.2× smaller standard deviation in measured delay at 25°C. The measured delay-power scatter plot in Figure 2.15(d) demonstrates much better robustness to process variations especially at the low temperature, since the exponential dependency of $I_{NMOS,ON}$ exacerbates the direct contention in DCVS.

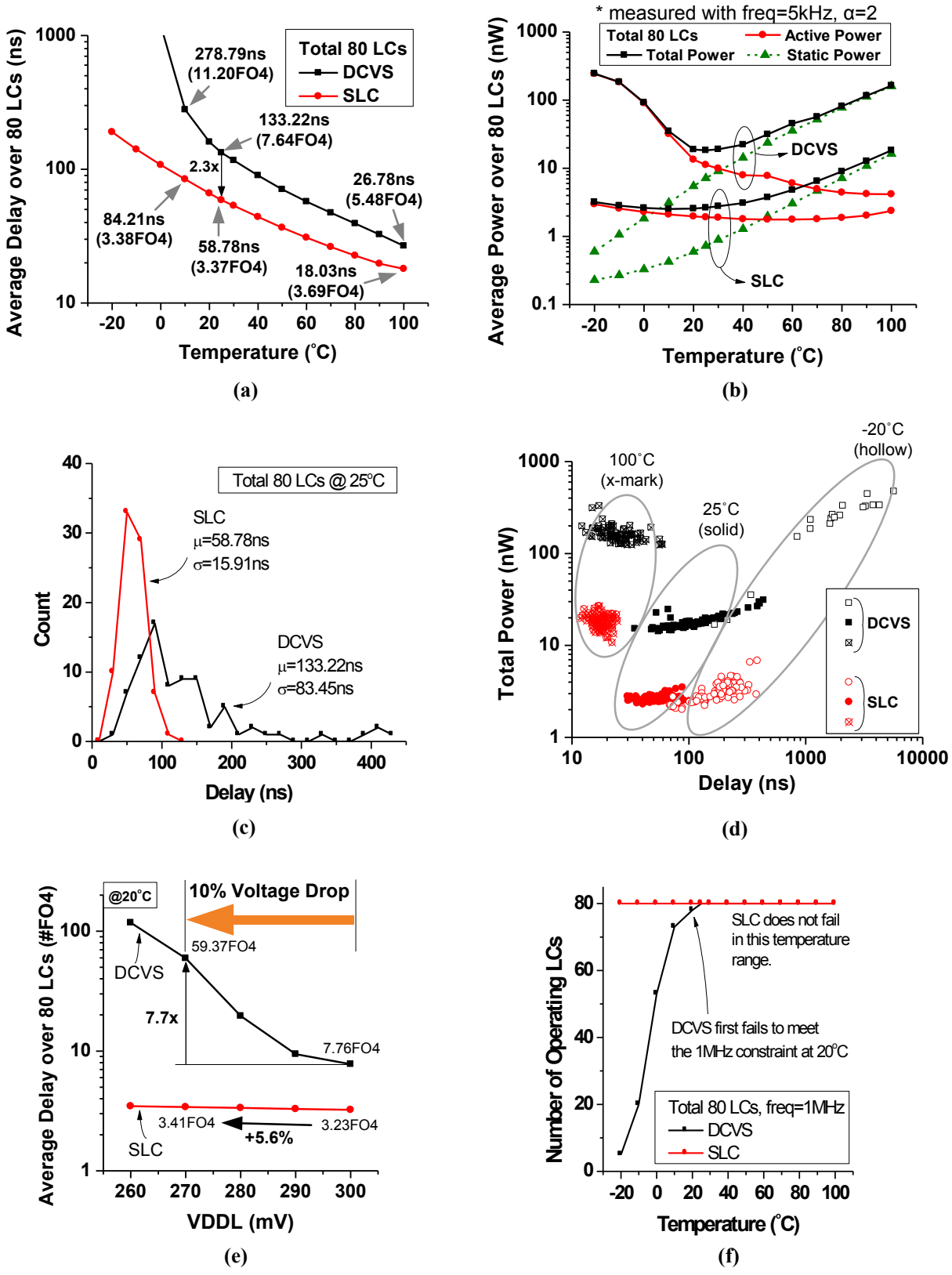


Figure 2.15: Measured result comparisons

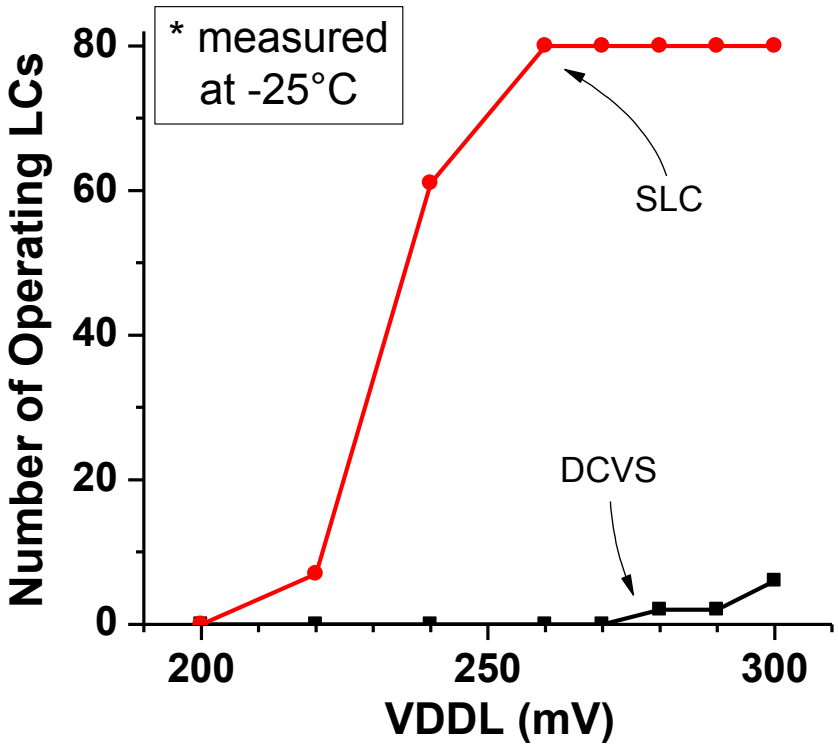


Figure 2.16: Yield comparison at very low temperature ($-25°$C)

Figure 2.15(e) and (f) show the effects of voltage/temperature variations. For a 10% V_{DDL} drop, DCVS LC delay degrades by 7.7×, while SLC speed reduces by only 5.6%. Although the DCVS LC is designed to operate at up to 20MHz at 25°C, some measured DCVS LCs fail to achieve 1MHz operation at 20°C and overall its functionality severely degrades as temperature is lowered. In contrast, SLC operates reliably over the full temperature range of -20 to 100°C. SLC robustness becomes more pronounced in severe conditions, as Figure 2.16 demonstrates all measured devices are functional even with $> 10\%V_{DDL}$ drop at very low temperature ($-25°$C), whereas DCVS LC is essentially non-functional at this condition. For sensor node applications, it is critical to work in a range of environments to enable true 'ubiquitous' networks; hence the robustness of SLC is a key advantage for such systems.

	LC^2	SLC	TVLSI'11 [21]	ESSCIRC'07 [19]
Technology	130nm	130nm	130nm	180nm
Conversion	0.3V to 2.5V	0.3V to 2.5V	0.3V to 2.5V	0.3V to 1.8V
Type	Static	Static	Dynamic (w/ 2.5V clock)	Static
Delay	41.51ns	58.78ns	125ns	∼600ns
Static Power	475pW	724pW	N/A	N/A
Energy per Transition	229fJ	191fJ	1.7pJ	∼20pJ
Area	$102.26\mu m^2$ (including the diode chain)	$71.94\mu m^2$	$0.1118mm^2$	No silicon implementation

Table 2.1: Comparison of wide-range LCs at 25°C

2.4 Conclusions

In this chapter, we presented new level converters and their measurements. Figure 2.17(a) shows the die photo and Table 2.1 shows comparisons to recent wide-range LCs.

Despite having more transistors than DCVS, LC^2 is smaller than DCVS in layout even including the extra diode chain, which can be shared among multiple LC^2s.

The static nature of LC^2 and SLC does not require clocks or complex synchronizing schemes, enabling 1093× and 1554× smaller area, respectively, compared to [21], which is also fabricated in 130nm CMOS. Compared to [21], LC^2 shows 7.4× lower energy per transition and 3× faster speed, while SLC has 8.9× lower energy per transition.

SLC is 35% smaller than the conventional DCVS, making it the smallest LC reported for wide-range (0.3V to 2.5V) conversions. We incorporated SLC in a previously reported low-power timer [42] and observed 15.8% reduction in switching energy; this improvement is conservative as the new timer includes overhead from an LDO regulator, which was not included in the previous design. Figure 2.17(b) shows the die photos of both timers. The new timer including SLC was successfully incorporated into the wireless sensor node system in the 130nm layer of [7]. This system also uses SLC (ported to 180nm CMOS) for its CPU, memory, and power management unit (PMU) interfaces. This SLC consists of thick-oxide I/O devices ($V_{TH} > 700mV$) and successfully operates for a 0.6V-3.6V conversion range.

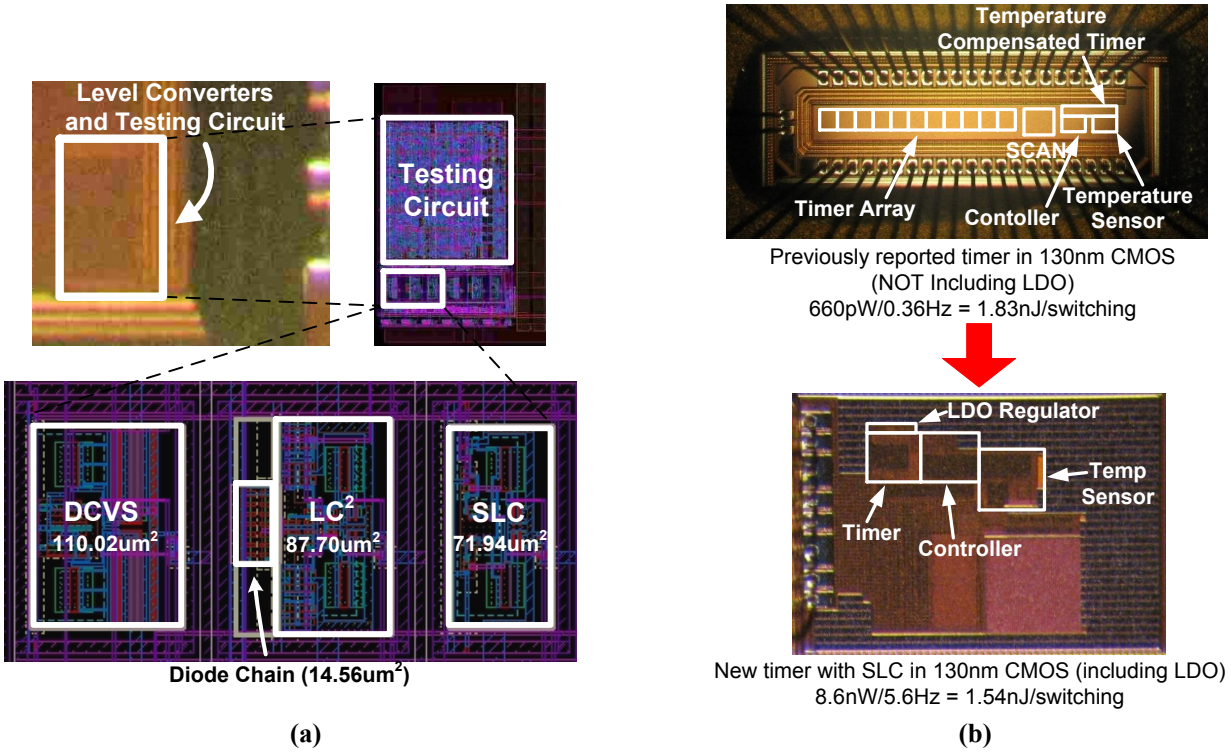


Figure 2.17: (a) Die photo of the test chip, (b) Die photos of low voltage timer designs [42][7]

CHAPTER 3

A Robust 7T SRAM Design

3.1 Introduction

SRAM suffers from reduced robustness due to severe process variation in nanoscale CMOS. In particular, it is challenging to jointly ensure reliable READ and WRITE operation in conventional 6T SRAM. As a result, 8T and even larger bitcells are widely used, particularly for low-voltage memories; they isolate READ and WRITE operations, so it is possible to separately optimize their robustness. However, this added robustness comes at the expense of density; 8T bitcells incur $\sim$30% area overhead compared to minimum achievable 6T bitcells [24][26][25]. In addition, 8T bitcells exhibit the so-called "Half-Select" problem making it difficult to apply column-muxing, as necessary for high array efficiency and SER robustness [25]. These issues are further complicated in emerging low power sensor systems due to ultra-low leakage requirements. For instance, the modular sensing system in [7] requires fW/bit standby power, necessitating the use of a 10T HVT bitcell (marked as 'K' in Figure 3.1) that is 4$\times$ larger than a commercial 6T SVT bitcell. Such area penalties are often not acceptable and hence there is a need for low leakage, low voltage tolerant designs that also achieve reasonable density.

3.2 Ultra Low-Leakage 7T SRAM

In this chapter, we propose a novel 7T SRAM that has decoupled READ/WRITE operation, similar to an 8T SRAM. It achieves robust operation at low voltage with 3.35 fW/bit standby

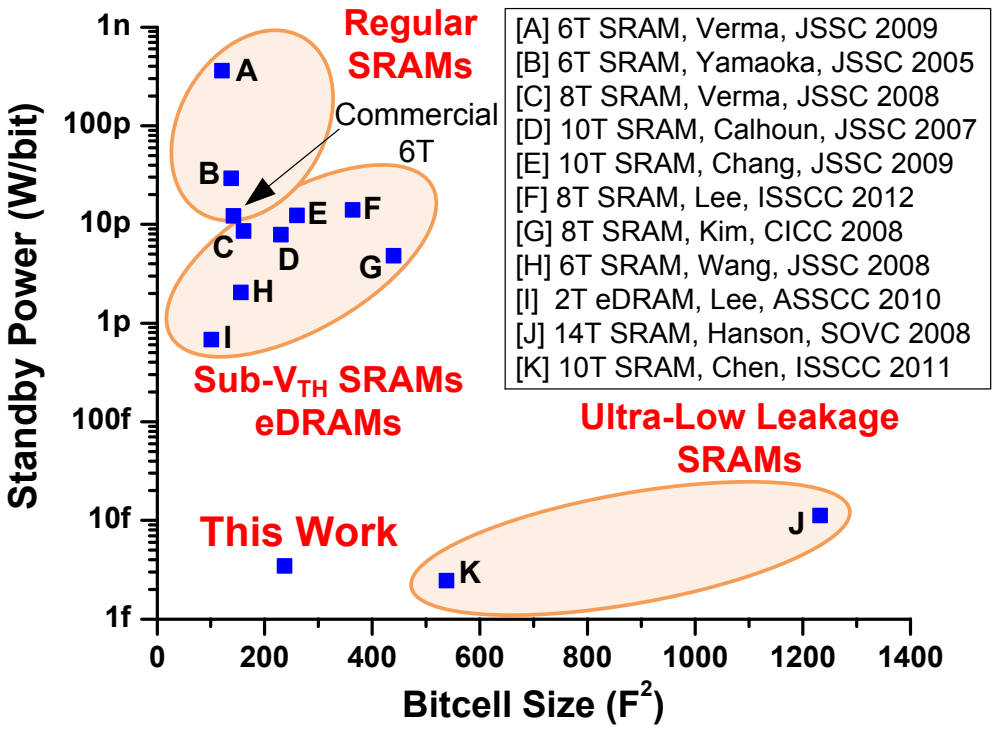


Figure 3.1: Bitcell size and standby power

power and reduces the area penalty of an 8T bitcell by 47%. It features a new dynamic read completion detection technique to avoid short-circuit current during READ and uses PMOS Pass Gate (PG) combined with dual supply voltages to mitigate the Half-Select problem and enable bit-interleaving. Prior 7T bitcells, using an L-shape layout, were presented in [29][30]. However, [29] uses tunneling FETs while [30] does not address the power overhead incurred by substantial short-circuit current during READ. Furthermore, [30] depends on Write-Back scheme to enable bit-interleaving, causing area/power overhead. The proposed 7T SRAM (8kB macro, 32-bit I/O with 2-way column-muxing) was fabricated in 180nm CMOS and addresses these issues while also providing extremely low leakage, making this SRAM applicable to low power applications without sacrificing area efficiency (Figure 3.1).

3.2.1 Auto-Shut-Off Sensing

Figure 3.2 shows the proposed 7T bitcell, which includes an HVT 6T portion and a single SVT READ Device (RD). As depicted in Figure 3.3, conventional READ in a 7T topology causes

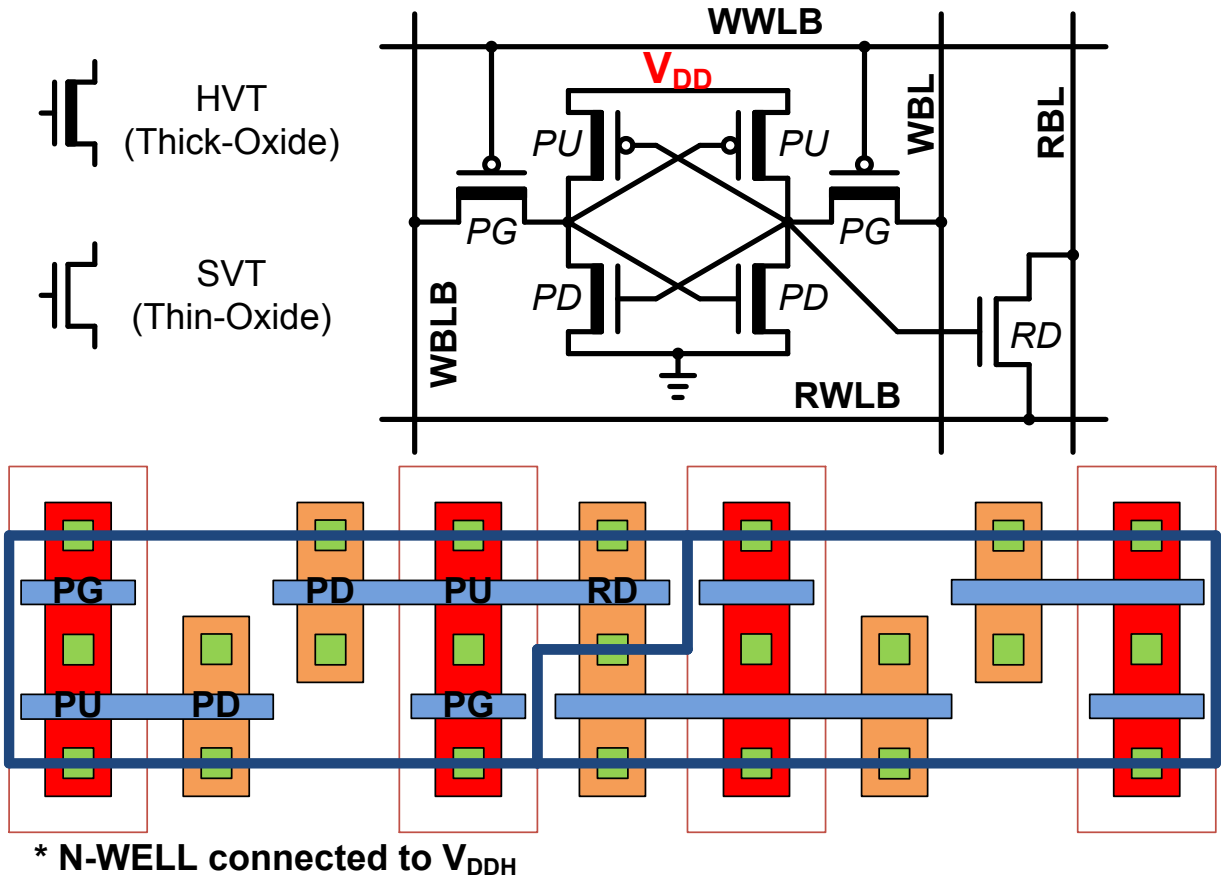


Figure 3.2: 7T bitcell schematic and the L-shaped layout

large short-circuit current from unselected cells (I_{UNSEL}) once V_{RBL} drops below $V_{DD} - V_{TH}$, turning on READ Devices (RD) along the column in bitcells storing Data1. This I_{UNSEL} limits the BL swing and incurs a large power penalty. The proposed 7T SRAM introduces an Auto-Shut-Off mechanism in which the selected READ wordline (RWLB) is automatically disabled during READ, thereby maintaining V_{RWLB} above $V_{DD} - V_{TH}$ and cutting off I_{UNSEL}. The READ wordline is not disabled if all selected bitcells store Data0. The proposed 7T SRAM uses dual voltages ($V_{DD} = 0.6V$, $V_{DDH} = 0.95V$) to provide a wider BL swing with negligible I_{UNSEL}. As shown in Figure 3.3, Auto-Shut-Off sensing with dual voltages reduces 7T READ energy by 6.8× (measured). The Auto-Shut-Off technique employs two sense amplifiers: coarse and fine (Figure 3.4). Once the fastest column discharges RBL sufficiently to trigger the coarse sense amp, RSTB (Reset Bar) is discharged, lowering RWL_EN to deactivate all wordlines so that all RBLs stop discharging and become floating. RSTB also asserts SAE, which fires the fine sense amp and isolates it

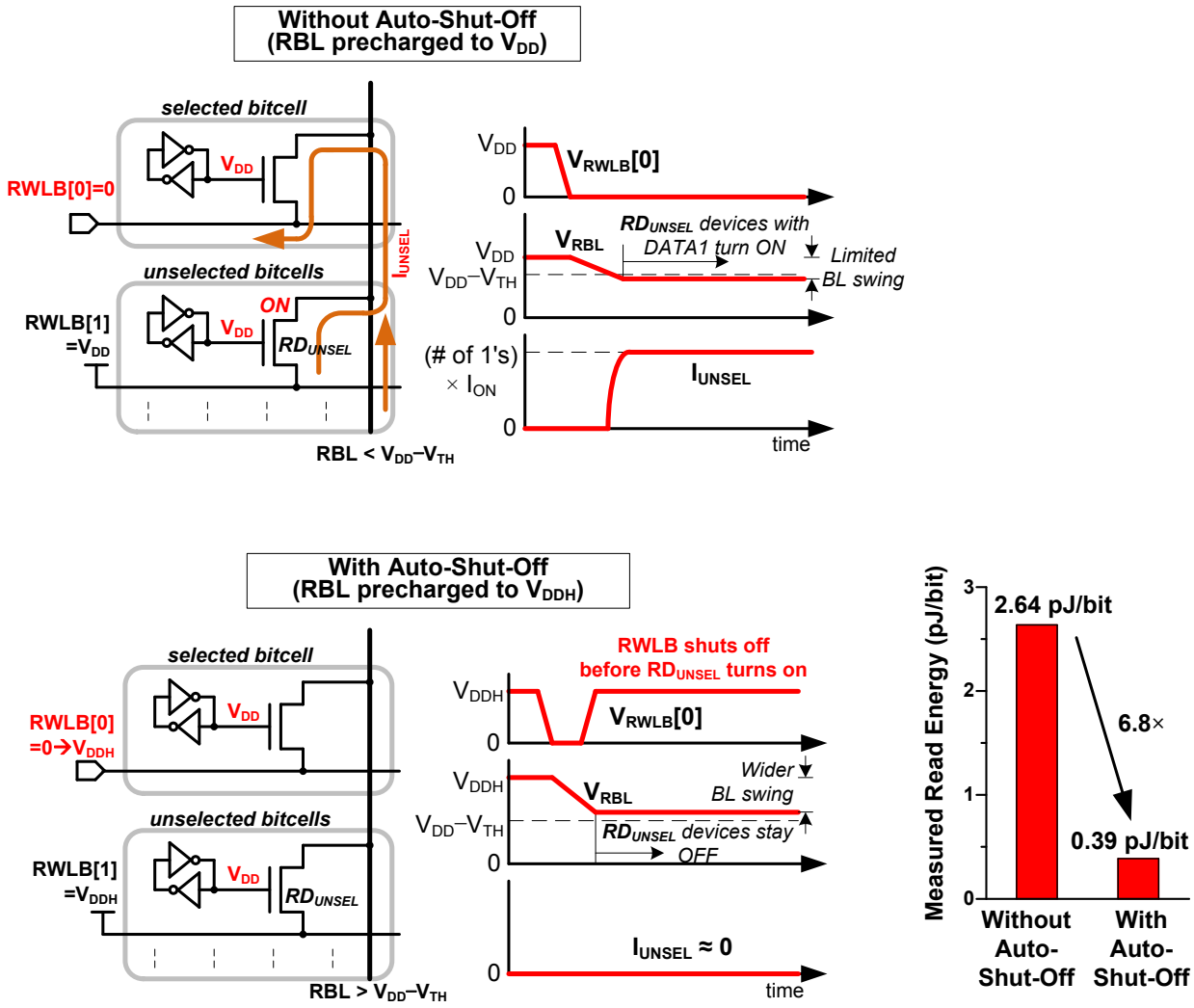


Figure 3.3: Auto-Shut-Off sensing and the measured improvement in READ energy

from RBL. Since the operation is stopped by the fastest column, the slowest column may have discharged a much smaller amount due to variations. To address this, the coarse sense amp must be margined to guarantee sufficient voltage differential for the fine sense amp to correctly detect the slowest RBL discharge. The fine sense amp is a biased topology designed to correctly detect voltage swings as small as 60mV. In the All-Data0 case, RBL remains high at V_{DDH}, as does RSTB. In this case, RWLB and RBL are reset at the falling edge of PULSE (Figure 3.4).

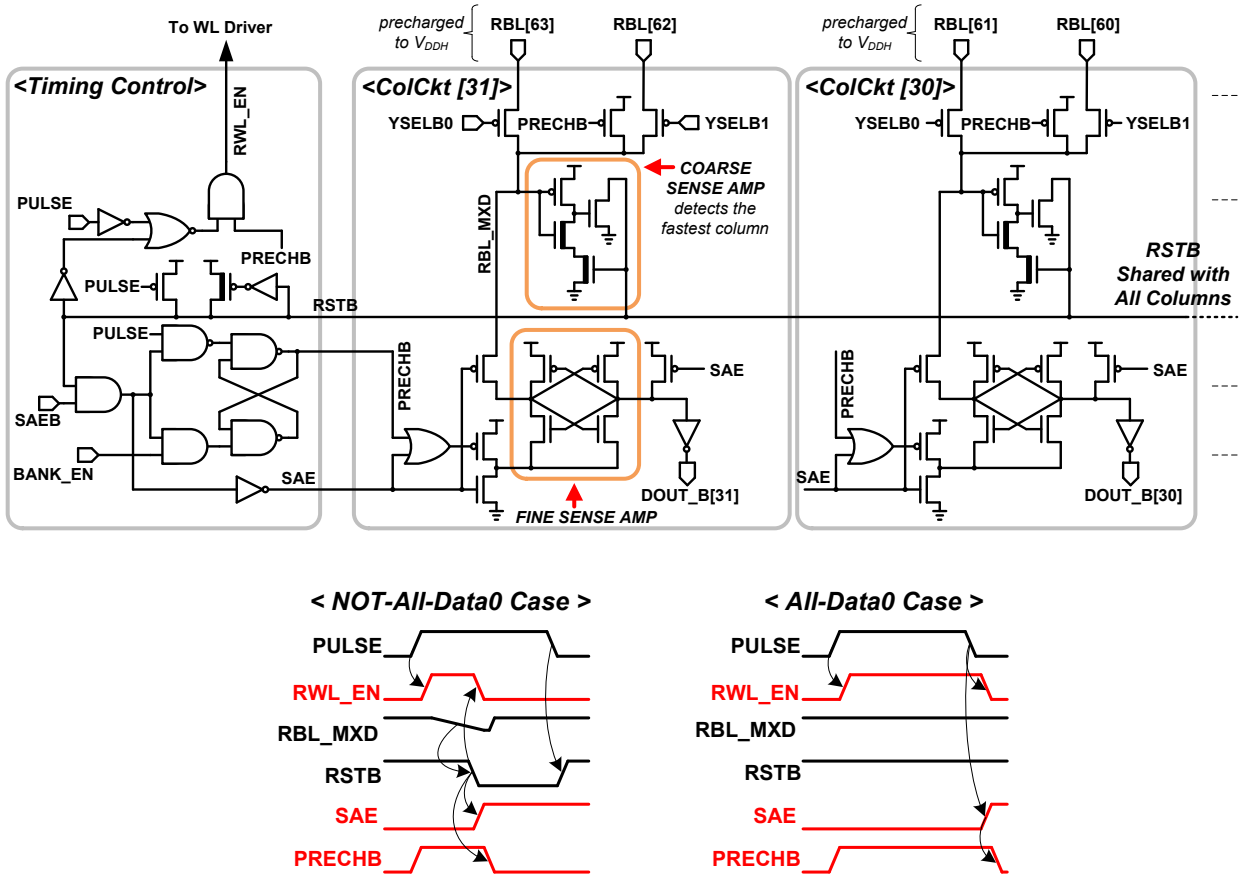


Figure 3.4: Circuit implementation of Auto-Shut-Off sensing

3.2.2 Quasi-Static READ

The proposed dual-V_{DD} 7T SRAM exhibits an innate bitline leakage suppression effect in unselected bitcells resulting from negative V_{GS} on their READ devices. When reading Data0 as in Figure 3.5, the worst-case scenario in 8T occurs when all unselected bitcells on a column have Data1, maximizing bitline leakage current. In contrast, I_{LEAK} from unselected cells in the 7T topology flows in the opposite direction, and therefore can help keep RBL high. Thus, the worst-case in a 7T occurs when all unselected bitcells also have Data0, creating a larger negative V_{GS} in unselected bitcells and thus reducing the beneficial I_{LEAK} while increasing I_{GATE}. However, I_{GATE} is significantly smaller than I_{LEAK} especially at high temperature. Also, due to the negative V_{GS} ($= V_{DDL} - V_{DDH}$ or $-V_{DDH}$, depending on cell data), I_{LEAK} is greatly suppressed and becomes negligible. Simulation shows that 7T bitline leakage is 113× smaller than in an 8T, such that the design shows quasi-static READ behavior. 8T SRAM generally requires a bitline keeper at low

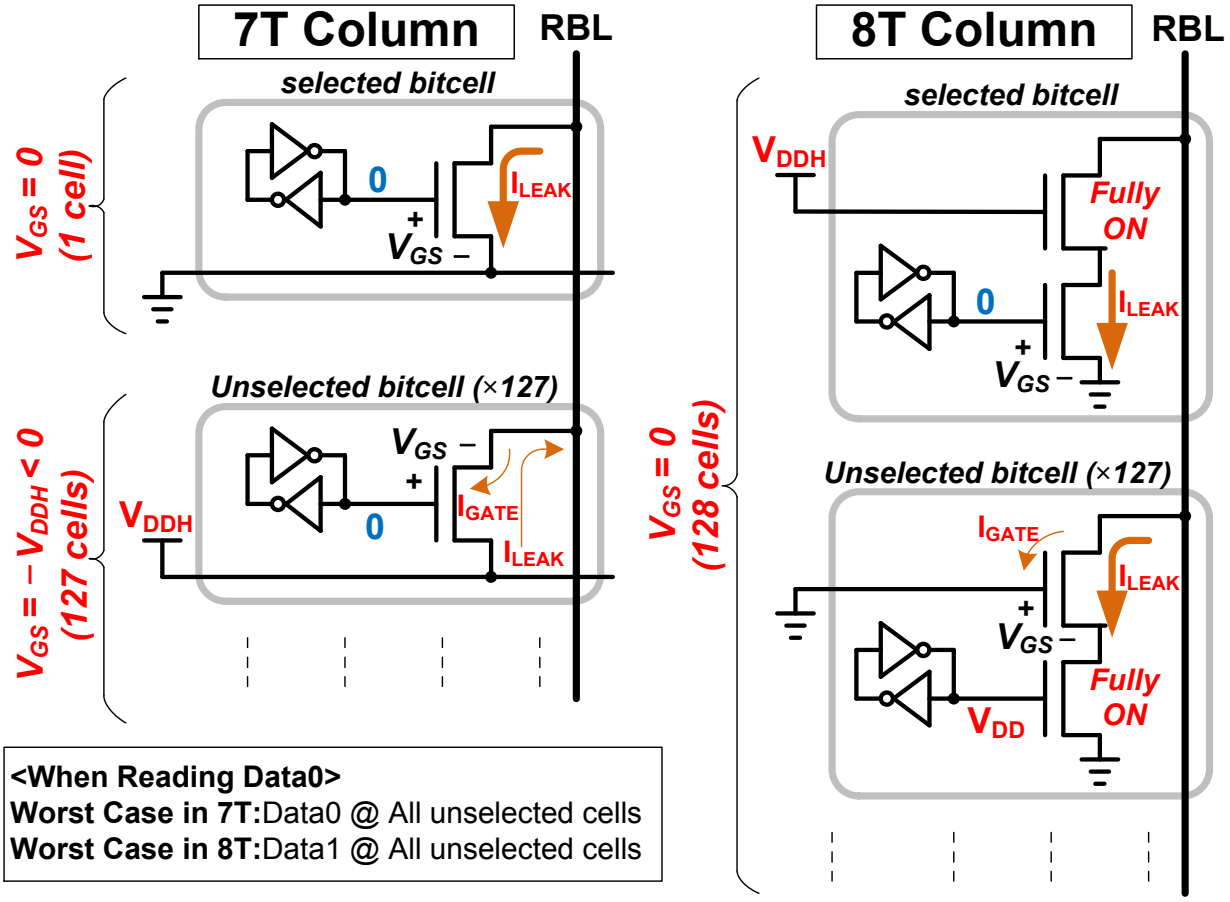


Figure 3.5: Quasi-Static READ

frequencies, which creates additional complexity, requires margining, and reduces robustness at low V_{DD}. The proposed 7T maintains robust operation without the keeper across supply voltages, as shown by measured results in Figure 3.6.

3.2.3 Bit-Interleaving with PMOS Pass-Gate

The use of conventional NMOS PG devices makes bit-interleaving difficult in low-voltage memories. As shown in Figure 3.7, $V_{GS} = V_{WWL}$ and is V_{DDH} for both written and the half-selected cells. Reducing V_{WBL} in the half-select cells does not improve the margin substantially between I_{PG} (WRITE) and I_{PG} (Half-Select), causing the PG device to fully transfer $V_{WBL(B)}$ $(= V_{DD})$ to the internal node during Half-Select. Several techniques [30][44] have been proposed to address this problem, resulting in significant complexity and area overhead. The proposed dual-voltage 7T

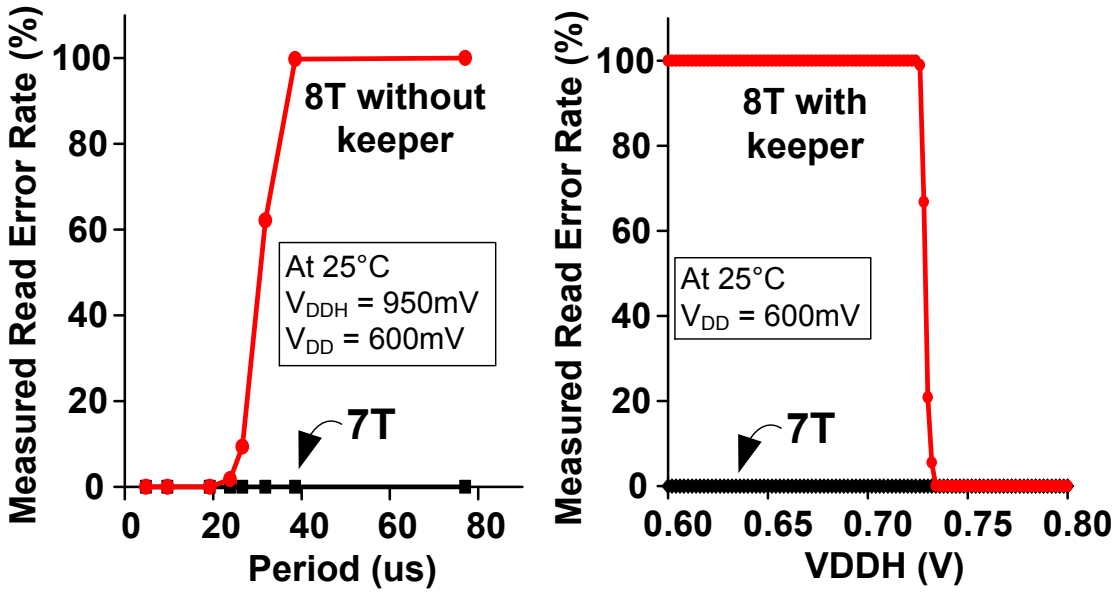


Figure 3.6: Measured improvement in read error rate due to Quasi-Static READ

instead uses PMOS PG such that $|V_{GS}| = V_{WBL(B)}$ and PG strength can be modulated by applying different $V_{WBL(B)}$ in WRITE and Half-Select columns. Also, the PMOS PG is reverse body-biased during Half-Select ($V_{BS} = V_{DDH} - V_{DD}$), increasing V_{TH} of these HVT devices such that the PG operates in the near-V_{TH} regime. This increases sensitivity of the PG to V_{GS} through $V_{WBL(B)}$ modulation, allowing us to further separate the Half-Select and WRITE PG currents as shown in Figure 3.7, in which a 0.35V change in V_{WBL} between WRITE ($V_{WBL} = V_{DDH}$) and Half-Select ($V_{WBL} = V_{DD}$) changes drain current by $\sim 104\times$ at TT corner. This controllability enables true column multiplexing without area overhead. Measurements in Figure 3.8 show that $V_{DDH} - V_{DD} > 100mV$ is sufficient to create enough V_{GS} sensitivity of the PG, resulting in no READ error from Half-Select columns. Since NWELL is biased at V_{DDH}, this reverse body-biasing also reduces standby power, which is minimized at $V_{DDH} - V_{DD} = 200mV$.

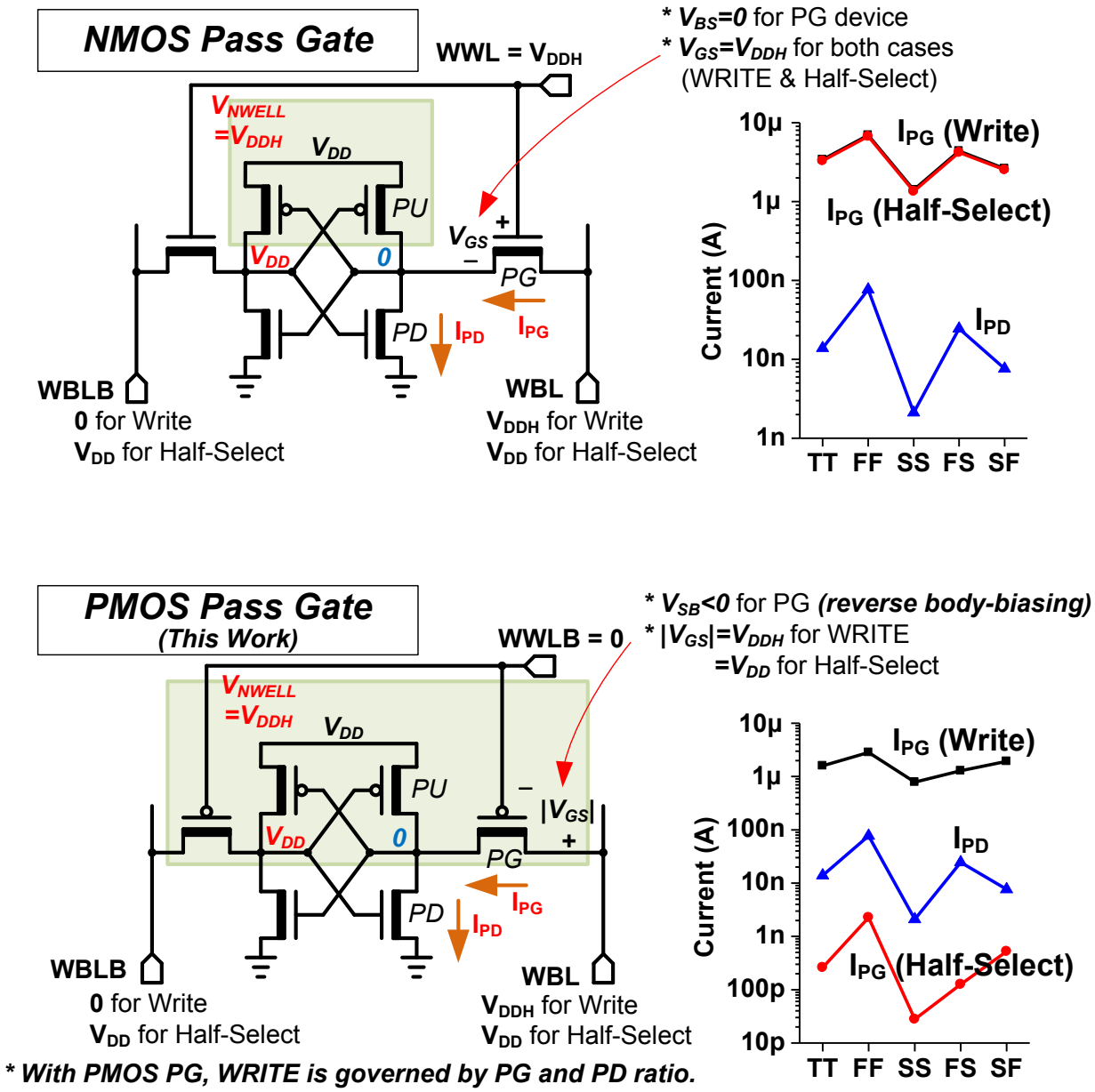


Figure 3.7: Bit-interleaving with PMOS pass-gate

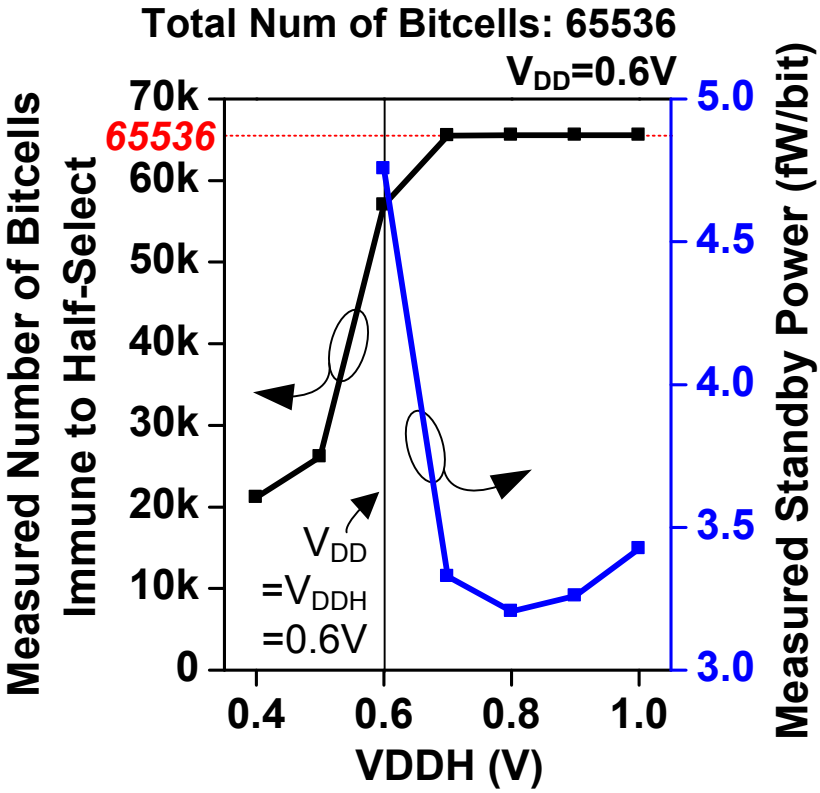


Figure 3.8: Effects of body biasing

3.3 Conclusions

A new 7T SRAM was fabricated in 180nm CMOS, and the 8kB macro shows the benefits from the novel Auto-Shut-Off sensing, Quasi-Static READ, and the bit-interleaving with PMOS PG devices. This 7T cell is 2.3× smaller than the 10T bitcell in [7], but still enables fW/bit standby power (3.35 fW/bit). It shows $>$ 3500× reduction in standby power compared to a commercial 6T bitcell. Figure 3.9 is a Shmoo plot showing V_{MIN} of 320mV. Table 3.1 shows a comparison with other low-power SRAMs, where the lowest bitcell leakage power and the column-muxing without extra circuit overhead (e.g., Write-Back) of the proposed 7T are clearly noticeable. The proposed bitcell is 20% larger than the 6T bitcell, while the 8T in [27] and the 10T in [43] have more than 60% increase in bitcell size. The die photo is shown in Figure 3.10.

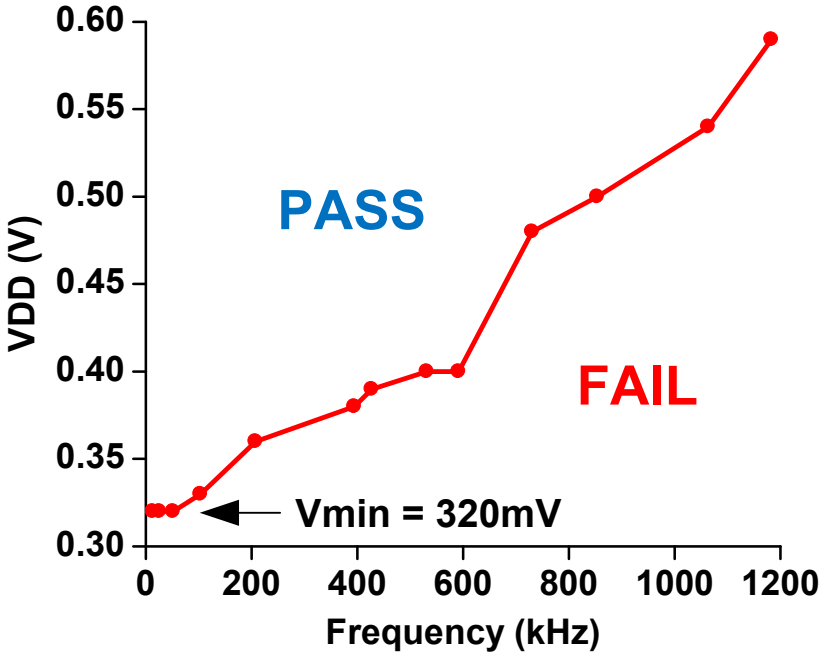


Figure 3.9: Shmoo plot

	This Work	JSSC'13 [30]	ISSCC'06 [43]	JSSC'09 [27]
Devices	7T (HVT)	7T (SVT)	10T (SVT)	8T (SVT)
Process	180nm	65nm	65nm	130nm
Voltage	Nominal $V_{DD} = 0.6V$ $V_{MIN} = 0.32V$	$V_{MIN} = 0.26V$	$0.4V$	Nominal $V_{DD} = 1.2V$ $V_{MIN} = 0.23V$
Bitcell Size (Normalized by 6T)	$7.75\mu m^2$ ($239F^2$) = 1.20× 6T (HVT) = 1.66× 6T (SVT)	<1.15× 6T (SVT)	1.66× 6T (SVT)	$6.36\mu m^2$ ($442F^2$) = 3.12× 6T (SVT)
#Bitcells/bitline	128	256	256	512
Column Muxing	2:1 (w/o assist)	8:1 (w/ Write-Back)	No	Not Reported
Energy	390 fJ/bit @ 0.6V	350 fJ/bit @ 0.26V	54 fJ/bit	2.69 pJ/bit @ 0.23V
Leakage/Bit	3.35 fW/bit	Not Reported	7.6 pW/bit	45 pW/bit

Table 3.1: Comparison of low-power SRAMs

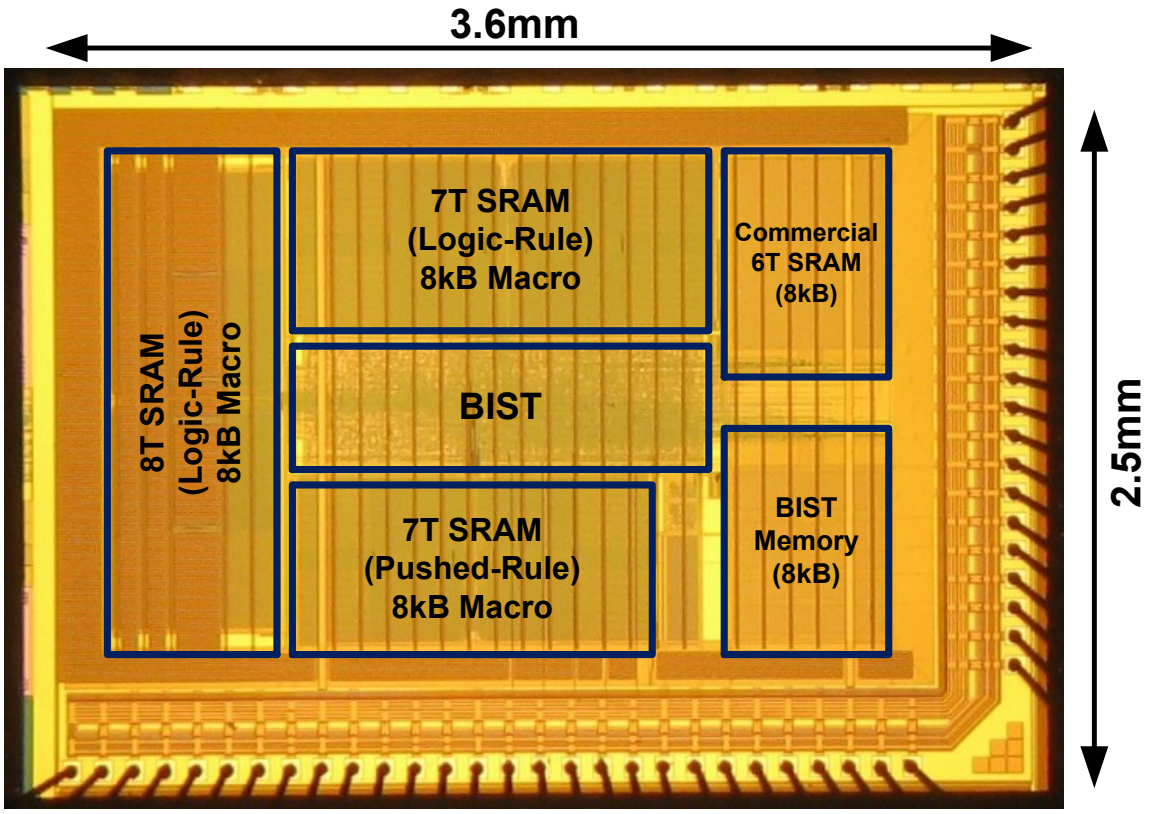


Figure 3.10: Die photo

CHAPTER 4

A Static Single-Phase Contention-Free Flip-Flop

4.1 Introduction

Near-threshold computing (NTC) is an attractive solution to stagnating energy efficiencies in digital integrated circuits, arising from slowed voltage scaling in nanometer CMOS [15][45]. However, the design of sequential elements for NTC, as well as in voltage-scaled systems operating at both near-threshold and super-threshold, has not been extensively studied; a recent study analyzes and compares many existing flip-flop topologies [33][34], but it is limited to the full V_{DD} (i.e., super-threshold) operations and does not take into account process / voltage / temperature (PVT) variations. In NTC, these variations become a critical concern for circuit robustness, and a correct operation at one PVT corner does not necessarily guarantee functional correctness at other PVT corners. The design of sequential elements is not an exception, and it is well known that they have a strong sensitivity to process variations in NTC [45], which can have a significant impact on system yield and power consumption. In order to achieve reliable energy-efficient operation across a wide operating voltage range, a flip-flop should have the following attributes: *a*) static operation, since dynamic nodes are highly susceptible to PVT variations at low voltage; *b*) contention-free transitions, since ratioed logic has poor robustness across the wide range of device I_{ON}/I_{OFF} ratios incurred with voltage scaling; *c*) single-phase clocking, which avoids toggling of internal clock inverters and incurs a corresponding power penalty; *d*) minimum or no area penalty compared to conventional ones.

While many flip-flops have been proposed, no prior design meets all these requirements for an

	TGFF	ACFF	TGPL	TSPC
Static Operation	YES	YES	YES	NO
Single-Phase Clock	NO	YES	NO	YES
Contention-Free	YES	NO	YES	YES
Device Count	24	22	28	11

Table 4.1: Comparison of conventional flip-flops

energy-efficient, highly voltage-scalable sequential element [33][34][35][36][37]. In the following sections, we will briefly discuss the issues with the conventional flip-flops, and then present a new flip-flop which owns all the above-mentioned characteristics. Details on operations and a beneficial "simple hold time path" will be presented, followed by measured data and comparisons with conventional ones.

4.2 Previous Flip-Flops

Figures 4.1 4.2 show schematics of several common flip-flop designs: transmission-gate flip-flop (TGFF), which is widely used in commercial standard-cell libraries; adaptive-coupling flip-flop (ACFF) [35]; transmission-gate pulsed-latch (TGPL) [36]; and true single-phase clock flip-flop (TSPC) [37]. Shortcomings of these flip-flops are summarized in Table 4.1.

The conventional TGFF is completely static and contention-free thus showing robust operations with voltage scaling. Its robustness and a highly-optimized cell layout with 24 transistors make it a de facto standard in commercial standard-cell libraries. However, it exhibits high power consumption due to a large number of clocked nodes (i.e., not single-phase clocked). It is possible to remove the two clock inverters from TGFF and distribute both CK and CKB through a clock tree design; this reduces the number of the always-toggling clock nodes in the flip-flop, but handling both polarities with ever-increasing clock skew is not an attractive option for voltage-scaled designs.

ACFF [35] is a static flip-flop which also incorporates single-phase clocking operation and has fewer devices than TGFF. The single-phase clock and the fewer device count results in lower energy consumption at low activity ratio at super-threshold regime. However, it has a degraded state-holding in the master latch. For example, suppose that FN=0 and F=1 right before the positive

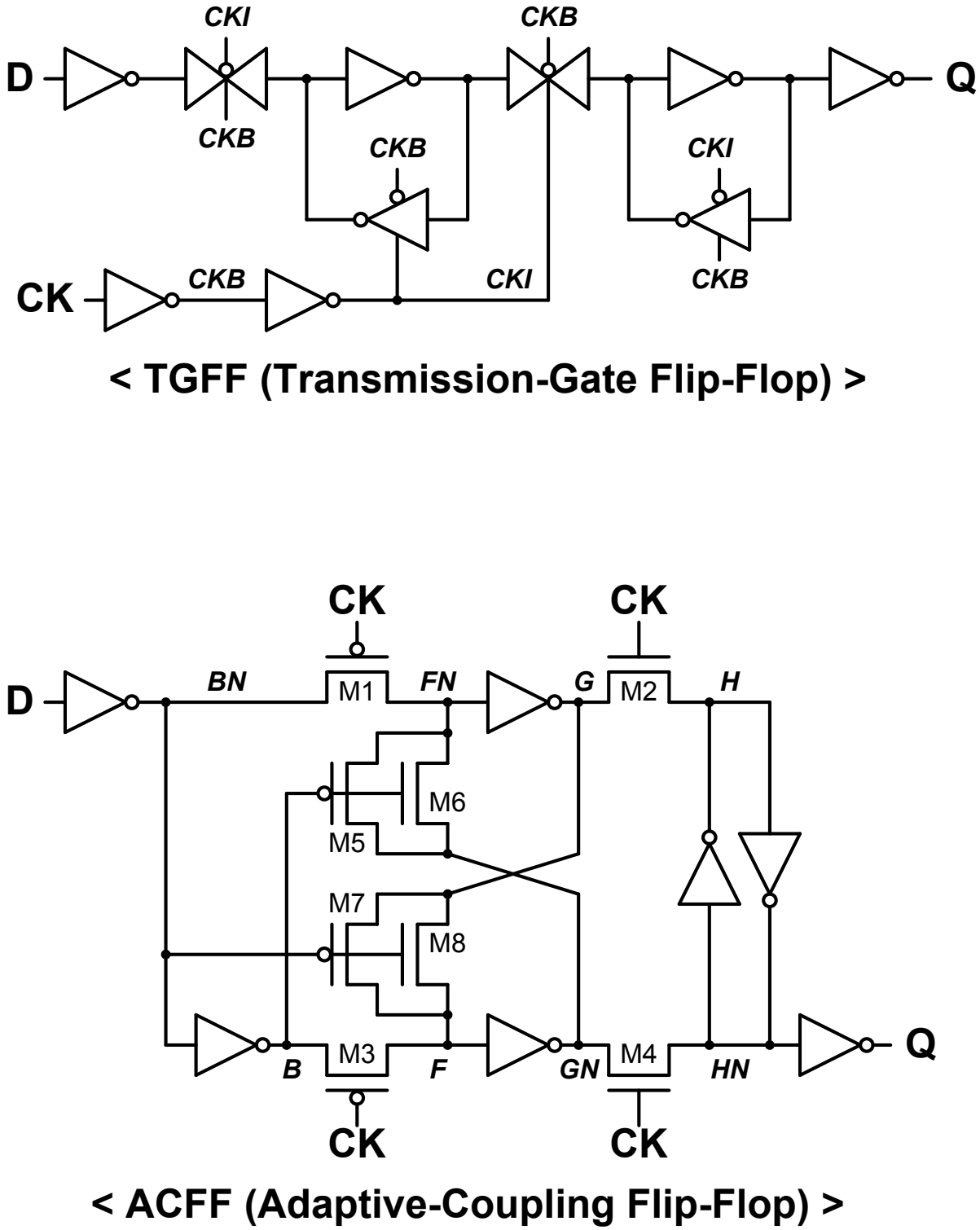
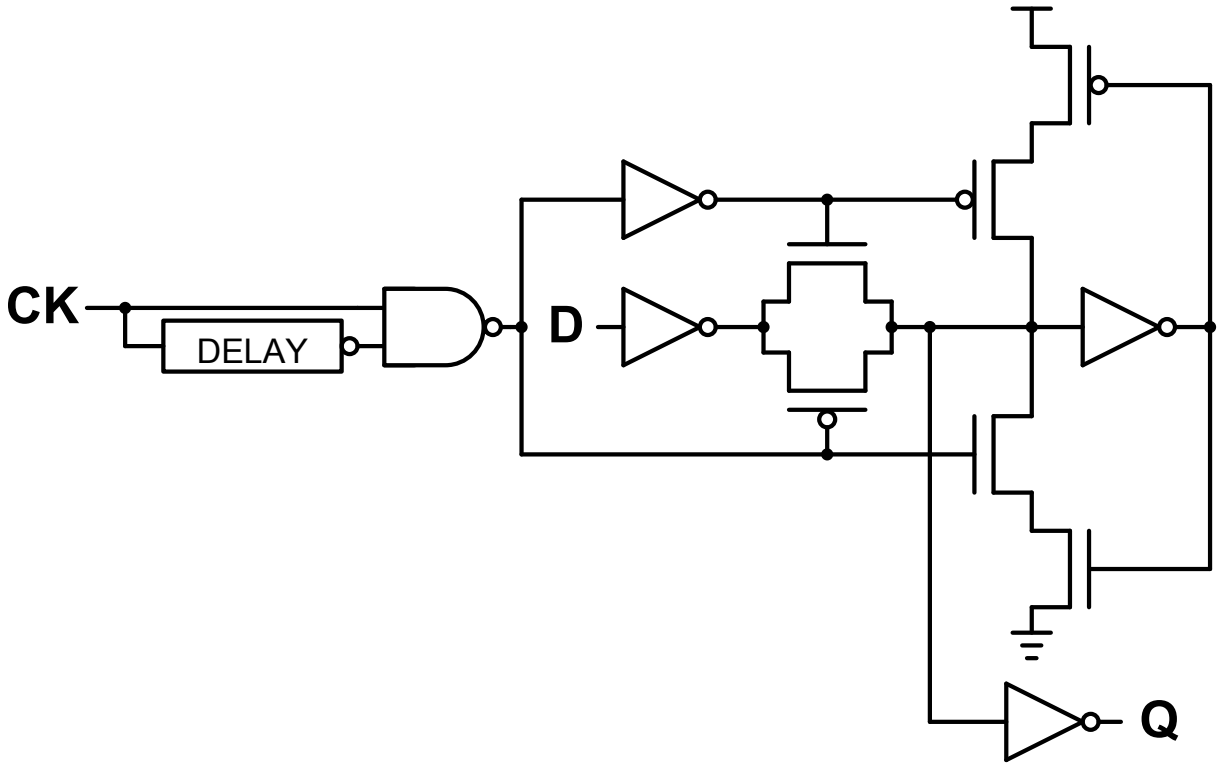
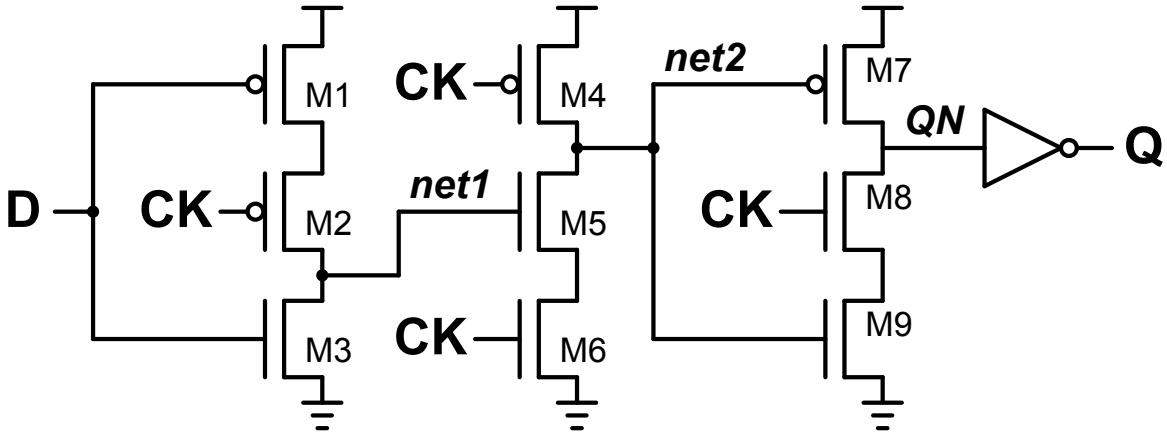


Figure 4.1: Schematics of TGFF and ACFF [35]



< TGPL (Transmission-Gate Pulsed-Latch) >



< TSPC (True Single-Phase Clock Flip-Flop) >

Figure 4.2: Schematics of TGPL [36] and TSPC [37]

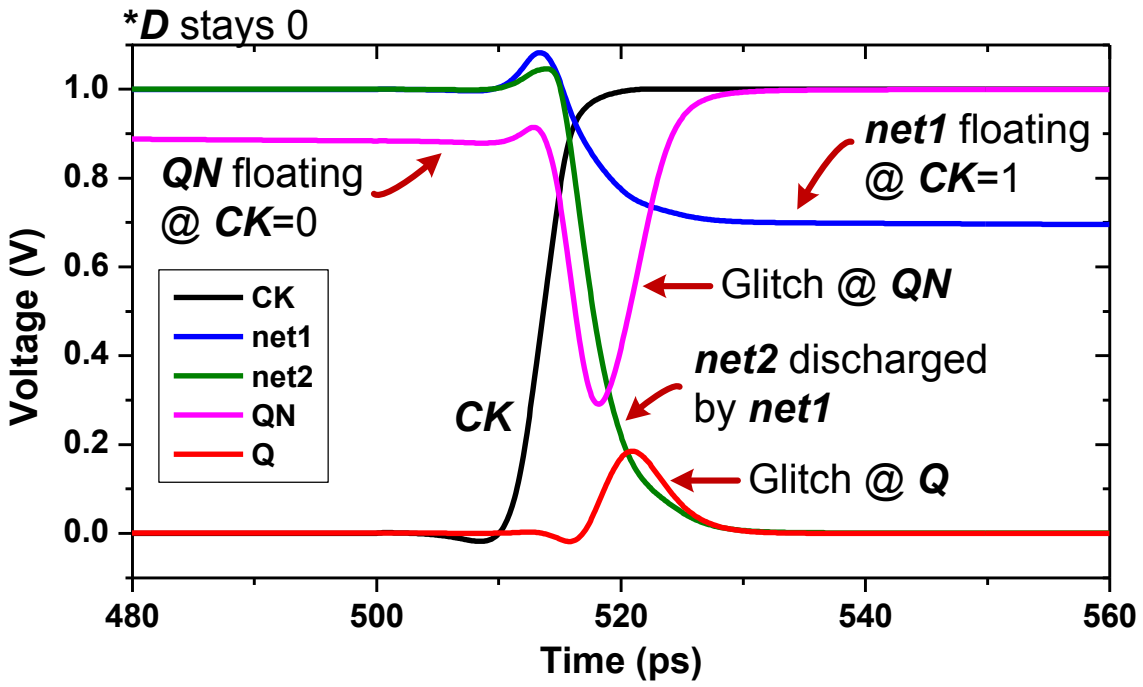


Figure 4.3: Waveforms in TSPC when D stays 0 for consecutive cycles

edge of CK, which also means BN=0, B=1, GN=0, and G=1. With the CK rising transition, M1 and M3 becomes turned off, and FN is held low by GN node through M6, while F is held high by G node through M7. If D changes during CK=1 phase, BN and B will change their values (i.e., it becomes BN=1 and B=0), thus turning off M6/M7 and turning on M5/M8. This causes FN kept low through a PMOS (M5) and F kept high through an NMOS (M8), which is undesirable for low voltage operation. ACFF also experiences current contention in the slave latch when updating H and HN nodes through M2 and M4; this causes rapidly increasing active power with higher activity ratio as well as functional failures at low voltage operation. This contention can be suppressed at the expense of additional devices, which then requires 26 transistors in total.

TGPL [36] is based on pulsed operation and achieves high performance at full V_{DD} but has poor robustness at low V_{DD} due to increased process variation sensitivity in pulse generation. Its hold time requirement is determined by the pulse width, hence the hold time of TGPL is positive unlike the above-mentioned flip-flops. At low V_{DD}, the pulse width becomes unpredictable, so does the hold time, because the delay element used for the pulse generation becomes quite susceptible to PVT variations. This often results in an excessive hold time margining during the design time, which causes power and area overhead.

TSPC [37] employs single-phase clock operation and uses only 11 devices. However, its dynamic operation degrades robustness, especially at low V_{DD}. In addition, Figure 4.3 illustrates a non-negligible glitch at node *QN* in TSPC whenever CK goes high while D remains 0. This arises since precharged *net2* begins to discharge *QN* before M5/M6 can pull *net2* low. Although *QN* will be eventually recovered back to the correct state (=high) by the discharged *net2* and M7, this glitch results in unnecessary power consumption or even system malfunction. From Monte Carlo simulations in 45nm SOI, the glitch at Q exceeds $V_{DD}/2$ with $\sim$1% probability (92/10,000 Monte Carlo simulations, V_{DD}=1.0V), potentially allowing for propagation to subsequent logic.

4.3 S^{2}CFF (Static Single-phase Contention-free Flip-Flop

4.3.1 Schematic and Operation Details

This work presents a new flip-flop, referred to as S^{2}CFF (Static Single-phase Contention-free Flip-Flop) that meets all the requirements mentioned in the introduction; it is static, completely contention-free, and uses single-phase clocking. It has the same device count as a TGFF, with only a 7% increase in layout size that corresponds to one poly-pitch increase in 45nm technology where fixed poly-pitch is enforced. Figure 4.4 shows the S^{2}CFF schematic, and the detailed operations are described in Figure 4.5 where grayed-out devices indicate OFF devices while others are ON.

In the schematic, M1$\sim$M4 becomes an inverter during CK=0 phase. Hence, *net1* holds an inverted D value when CK=0. Since M3 is fully turned on by the precharged *net2* (precharged through M8), any change in D can propagate to *net1*, i.e., it is transparent, and both *net1* and *net2* are static during CK=0 phase. At the positive edge of CK, depending on the *net1* value, *net2* will be staying high or discharged through M9$\sim$M10. This will update the slave latch (M17$\sim$M22); *QN* will be charged up by M13 if *net2* becomes low; otherwise, *QN* will be discharged through M14$\sim$M16 if *net2* stays high. In this CK=1 phase, M22 is conditionally turned on/off depending on *net2* value (data-dependent), while M19 is always off. M3 is an isolation device that prevents a change in D from affecting *net1*. M5$\sim$M7 are keeper devices and make *net1*/*net2* fully static. M11$\sim$M12 generates *net1b* signal that controls the keeper (M7) as well as the glitch prevention device (M15), which will be explained later.

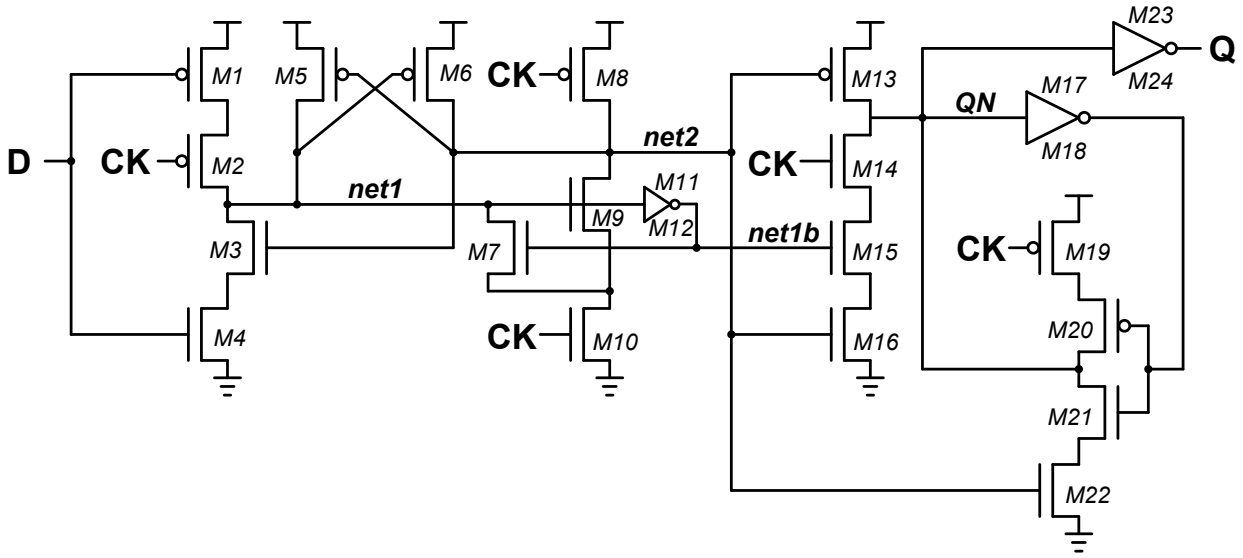


Figure 4.4: Schematic of S^2CFF

If D=0, *net1* holds an inverted D value (=high) and *net2* precharges through M8 while CK=0. In this state, there is no keeper needed; the keepers M5 and M6 are off because both *net1* and *net2* are high, and the keeper M7 is also off since *net1b* is low. The slave latch (M17∼M22) stores the previous data and is isolated from the previous stage because M13 and M14 are turned off. At the positive edge of CK, the high *net1* starts discharging *net2* through M9 and M10. Then, the discharged *net2* turns off M3, completely isolating the circuit from changes in D. Also, the low *net2* charges *QN* through M13, updating the data in the slave latch. The low *net2* activates the keeper M5, which holds *net1* high. M9 and M10 keep *net2* low during CK=1 phase.

If D=1, *net1* holds an inverted D value (=low) and *net2* precharges through M8 while CK=0, as same as before. However, the positive edge of CK does not generate any dynamic transitions at *net1* and *net2* since the low *net1* turns off M9 so that *net2* just stays at the precharged state (=high) after the clock rising transition. Note that *net1* is kept low by M7/M10, and M6 holds *net2* high during CK=1 phase. If the previous Q value is same as the current D input (i.e, Q=1, *QN*=0), there is also no transition at *QN*. Otherwise, *QN* discharges through M14∼M16. Although M3 stays on during CK=1 phase due to the high *net2*, it does not affect the *net1* state (=low). If D changes from 1 to 0 during CK=1 phase, it cuts off the discharging path (M3∼M4) by turning off M4; however, *net1* is still held low by M7 and M10, so it still remains static.

Signal *net1b* is also used to control M15 to prevent glitches; without this sub-circuit, *QN* will

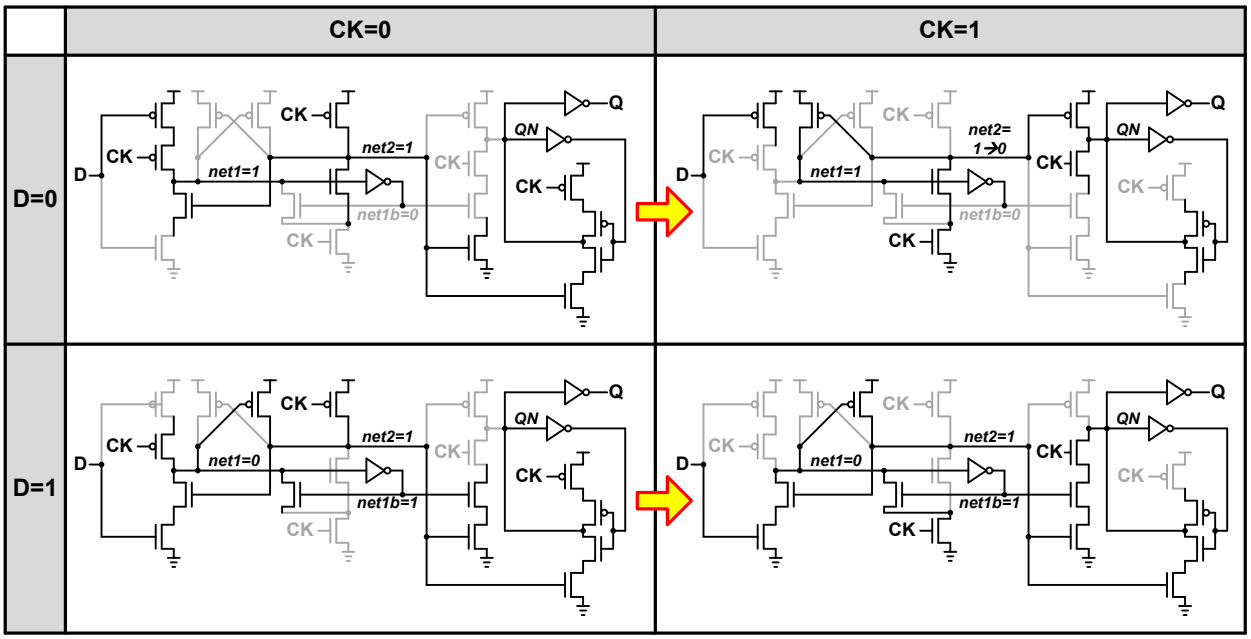


Figure 4.5: Operation of S^{2}CFF

glitch when CK rises with D staying low in consecutive cycles, similar to TSPC. M15 eliminates this glitch by cutting off the discharge path (M14∼M16) depending on *net1*s value; it turns off M15 if *net1* is high (i.e., D=0, *net1b*=0), hence *QN* can stay high without a glitch. M15 stays on if *net1* is low (i.e., D=1, *net1b*=1). *QN* can be discharged as intended through M14∼M16 in this case.

It should be noted that there is no contention throughout the operation, all internal nodes are fully static, and only one clock phase (CK) is used. Moreover, all of these are achieved with 24 transistors, which is same as in TGFF. This implies that the area penalty is just negligible, if not zero.

4.3.2 Hold Time Path

An additional benefit of the S^{2}CFF topology is that it simplifies the "hold time path" compared to a regular TGFF. Figure 4.6 shows the hold time paths of TGFF and S^{2}CFF. As described in [38], the worst-case hold time in a TGFF is when D changes from 1 to 0 just after the CK rising transition. Due to clock inversion in I4, the PMOS in I2 always turns off later than its NMOS. The 0-to-1 transition at node *DN* (1-to-0 at D) has more time to propagate through I2 compared

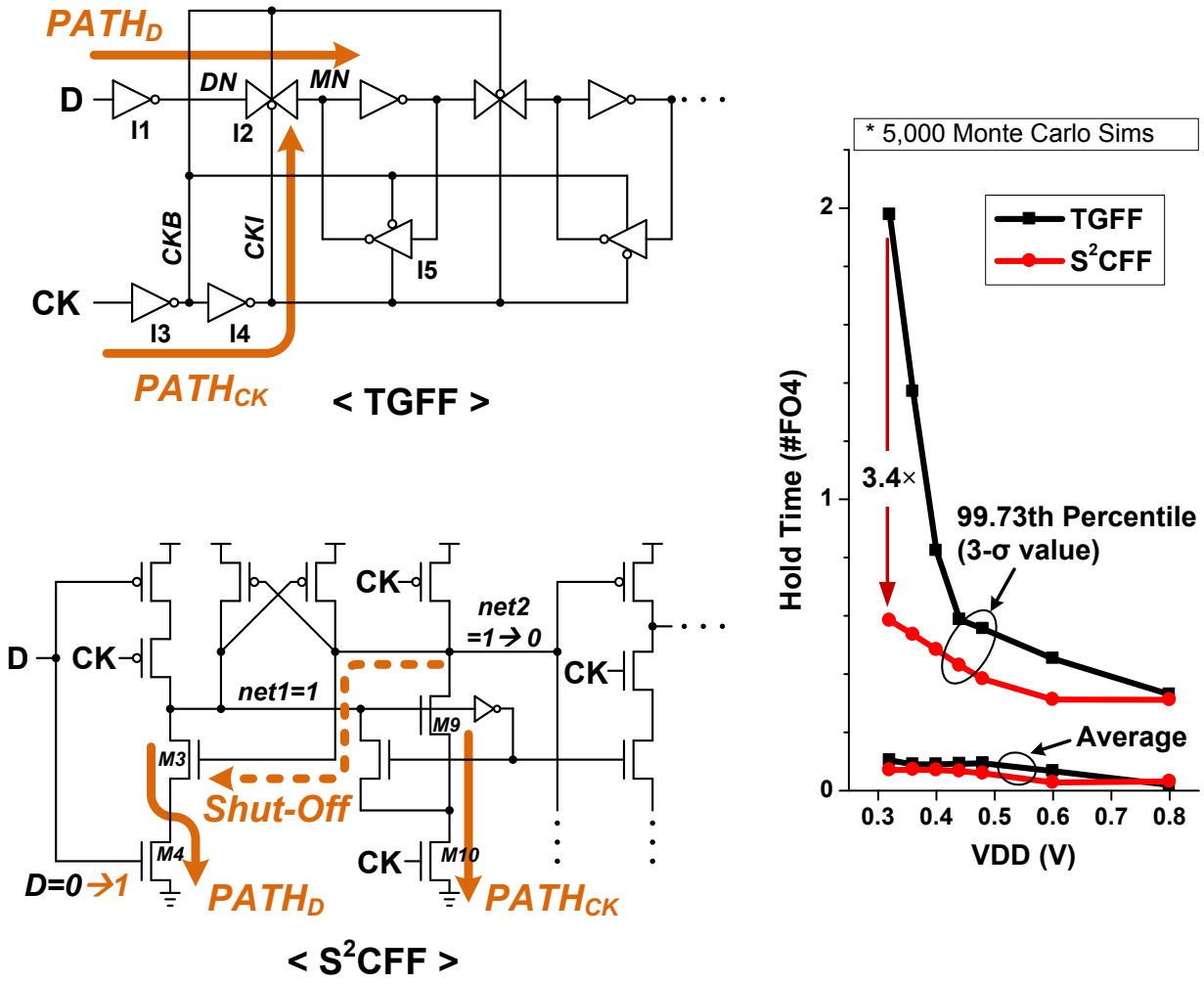


Figure 4.6: Hold time paths in TGFF and S^{2}CFF

to the 1-to-0 transition at node *DN* (0-to-1 at D). Also, the clocked PMOS in I5 always turns on earlier than its NMOS counterpart, thereby weakening the pull-down strength at node *MN*. Hence, node *MN* becomes more vulnerable to the 0-to-1 transition (1-to-0 at D) around the positive edge of CK. In addition, the data arrival time at *DN* is dictated by I1, while the clock arrival time at I2 is determined by I3 and I4. Thus, in sum, TGFF hold time is dictated by the mismatch among the clock/data inverters (I1, I3, I4), causing a severe hold time degradation at low V_{DD} where mismatch is accentuated.

On the contrary, the worst-case hold time in S^{2}CFF occurs when D changes from 0 to 1 just after the CK rising transition. The high *net1* starts discharging *net2*, and the discharged *net2* turns off M3, isolating the D input. A hold failure may occur, if D becomes 1 before *net2* shuts off M3,

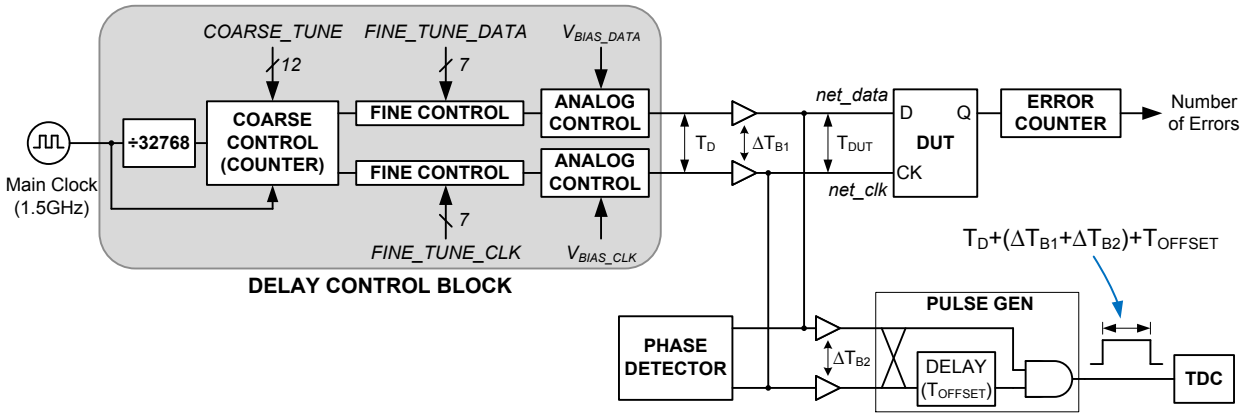


Figure 4.7: Setup/hold time measurement circuit

and thus discharges *net1*. Only the discharging speed of *net2* through $PATH_{CK}$ (M9 and M10) dictates the hold time. It should be noted that $PATH_D$ (M3 and M4) delay does not affect the worst-case hold time mentioned above, because: if $PATH_D$ is faster than $PATH_{CK}$, there is always a hold violation, so the (required) hold time must be the $PATH_{CK}$ delay (or less); if $PATH_D$ is slower than $PATH_{CK}$, there is no hold violation at all. As a result, the hold time of S^2CFF is determined solely by the discharging speed through $PATH_{CK}$ thus is much less prone to variability compared to a TGFF, which involves the time difference of several gate delays. The plot in Figure 4.6 shows a substantial reduction (3.4×) in hold time at the 3-σ value at 0.32V for S^2CFF (Monte Carlo simulations). This suggests large potential benefit for NTC, since small hold time variation reduces buffer-insertion overhead, reducing power and improving system yield.

4.4 On-Chip Testing Circuits

On-chip testing circuits are required to accurately measure sequential elements' timing characteristics, such as setup/hold time and C-Q delay. It is also important to measure flip-flops' power in various conditions. This section discusses each testing circuits in the following sub-sections.

4.4.1 Setup/Hold Time

The on-chip setup/hold time measurement circuit is shown in Figure 4.7, which is based on the structure in [46]. The fast main clock (∼1.5GHz) is divided by 32768 to generate a suf-

ficiently slow periodic signal. Coarse Control block generates two periodic signals based on the divided clock, and one signal can be made lagged or led by the other signal using COARSE_TUNE bits. One signal becomes a data input to DUT (*net_data*), while the other becomes a clock input (*net_clk*). This Coarse Control block is basically a counter operated by the fast main clock, so the control resolution is determined by the main clock frequency. Fine Control block is a long inverter chain. The data path and the clock path have its own Fine Control block, so that the delays are separately controlled using tuning bits (FINE_TUNE_DATA and FINE_TUNE_CLK), and the control resolution is one FO1 delay. Finally, Analog Control consists of current-starved inverter chains where the delay can be controlled using analog voltages (V_{BIAS_DATA} and V_{BIAS_CLK}) which provides a further fine resolution (<1ps). This Delay Control Block makes a delay difference (T_D), and the two signals are delivered into DUT through buffers. Phase Detector is used to align the edges of data/clock signals on *net_data* and *net_clk*. Based on this alignment, which indicates $T_{DUT} = 0$, a slight time difference can be made by changing the tuning bits or the bias voltages in Delay Control Block, while Error Counter determines whether there is a setup/hold failure by checking the DUT output. Pulse Gen generates a pulse whose pulse width corresponds to $T_{DUT} + T_{OFFSET}$. This pulse width is then measured using the sub-1ps resolution TDC [47]. At full V_{DD}, buffer mismatch (ΔT_{B1} and ΔT_{B2}) is negligible compared to T_D, and setup/hold times can be accurately measured.

4.4.2 C-Q Delay

The C-Q delay measurement circuit is shown in Figure 4.8. It incorporates a new flip-flop ring, where a short pulse at EN input triggers the oscillation of DUT Ring with a period that is proportional to T_{CQ} with an offset value.

$$T_{P,OSC} = 2N \times T_{CQ} + 2N \times T_M + N \times (T_B + T_I) \tag{4.1}$$

where N is the number of Unit Cells in a ring, T_{CQ} is the C-Q delay of DUT. T_M, T_B, and T_I represent the mux, buffer, and inverter delays in Unit Cell, respectively. The offset value can be measured using Reference Ring. The period of the oscillation in Reference Ring is:

$$T_{P,REF\_OSC} = 2N \times T_M + N \times (T_B + T_I) \tag{4.2}$$

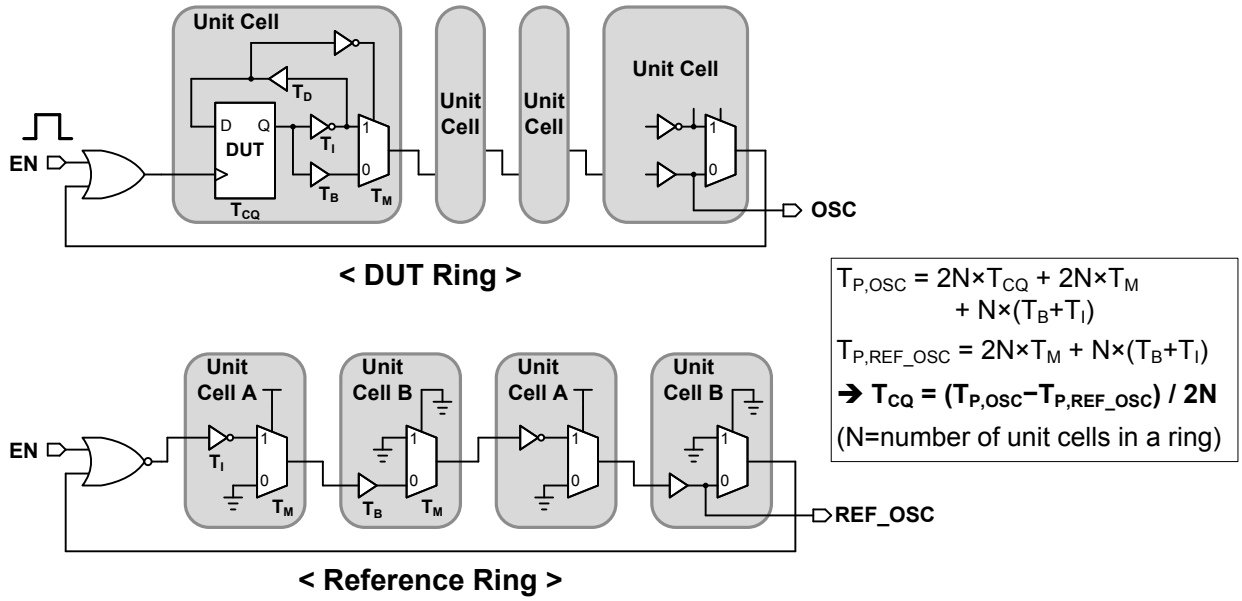


Figure 4.8: C-Q delay measurement circuit

Thus, the average C-Q delay can be obtained by subtracting T_{P,REF_OSC} from $T_{P,OSC}$.

$$T_{CQ} = (T_{P,OSC} - T_{P,REF\_OSC})/2N \tag{4.3}$$

With a large N value, local mismatch is effectively cancelled out making it possible to obtain accurate C-Q delays. While only 4 unit cells are shown in the figure for simplicity, the actual test chip implementation includes 100 Unit Cells in DUT Ring ($N = 100$). Reference Ring alternates Unit Cell A and Unit Cell B, with 50 of each in the full ring. The DUT Ring also gives insight on DUT yield, since oscillation stops unless all 100 DUTs in the ring are functional.

4.4.3 Power

Figure 4.9 shows the power measurement circuit where the activity ratio is controlled from 0% to 100% by loading the 20-bit INITIAL_PATTERN, as shown in Table 4.2. In order to mimic a realistic scenario, it has one clock buffer driving 10 DUTs. The current flowing into 'CLKBUF + 10 DUTs' is measured and then divided by 10. Hence, measured power consumptions in this paper also take into account the clock driving power.

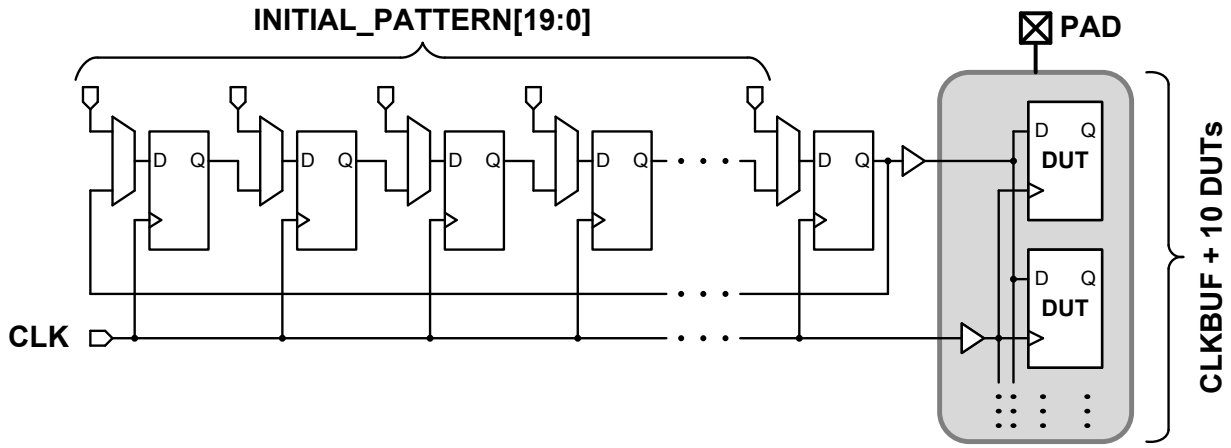


Figure 4.9: Power measurement circuit

INITIAL_PATTERN[19:0]	Activity Ratio
0000 0000 0000 0000 0000	0%
1000 0000 0000 0000 0000	10%
1010 0000 0000 0000 0000	20%
1010 1000 0000 0000 0000	30%
1010 1010 0000 0000 0000	40%
1010 1010 1000 0000 0000	50%
1010 1010 1010 0000 0000	60%
1010 1010 1010 1000 0000	70%
1010 1010 1010 1010 0000	80%
1010 1010 1010 1010 1000	90%
1010 1010 1010 1010 1010	100%

Table 4.2: Setting activity ratio in power measurement circuit

4.5 Measurements

S^2CFF was characterized in a 45nm SOI test chip, and TGFF, ACFF, and TGPL were also implemented in the same test chip for fair comparisons; 50 dies were fabricated and measured.

Figures 4.10 and 4.10 show measured total power and energy. S^2CFF does not require internal clock inverters, and this enables a clock power reduction, where the clock power is defined as total power at 0% activity ratio with D=0. From the power measurement, S^2CFF shows a clock power reduction of 41% and 40% at 1V/1GHz and 0.4V/200MHz operations, respectively, compared to TGFF. Assuming that flip-flops in a typical system have 20% activity ratio, S^2CFF provides 39% and 38% improvement in total sequential power at 1V/1GHz and 0.4V/200MHz, respectively, compared to TGFF. ACFF also has single-phase clocking operation thus showing a similarly low clock power as S^2CFF. However, the total power of ACFF increases rapidly as activity rises due to contention in the slave latch; this makes S^2CFF the lowest power flip-flop at any activity ratio. TGPL has a delay element, which leads to higher total power consumption even at 0% activity ratio. In terms of active energy consumption, S^2CFF shows 32% and 34% reduction at 1.0V and 0.4V, respectively, compared to TGFF. S^2CFF is the lowest energy flip-flop due to the static, single-phase clock, and contention-free operation.

Figures 4.12 and 4.12 show measured C-Q delays and leakage power. The C-Q delay in S^2CFF is determined by *net2* being staying precharged or discharged depending on the *net1* value at the positive edge of CK, followed by updating *QN* (thus Q) node. Compared to TGFF where the C-Q delay is determined by the delay through one transmission-gate and two inverters, S^2CFF shows modest improvement across V_{DD} with 14.8% faster C-Q delay at 1.0V. ACFF has the fastest C-Q delay by placing the output inverter right after the passgate (M4 in Figure 4.1). However, it should be noted that the missing points in the plot indicate that ACFF fails to have 100% yield at 0.4V. This is due to the current contention in the slave latch as well as the degraded state-holding in the master latch, as described earlier. Similarly, TGPL fails at $V_{DD} \leq 0.6V$, mainly due to hold time failures; it has a positive hold time constraint because of the pulsed operation, and the pulse width becomes very sensitive to PVT variations especially at low V_{DD}. This illustrates the importance of static and contention-free operation at low V_{DD}, since only TGFF and S^2CFF show 100% yield across the wide V_{DD} range. From the leakage measurement, S^2CFF has 35% and 37% lower leakage power

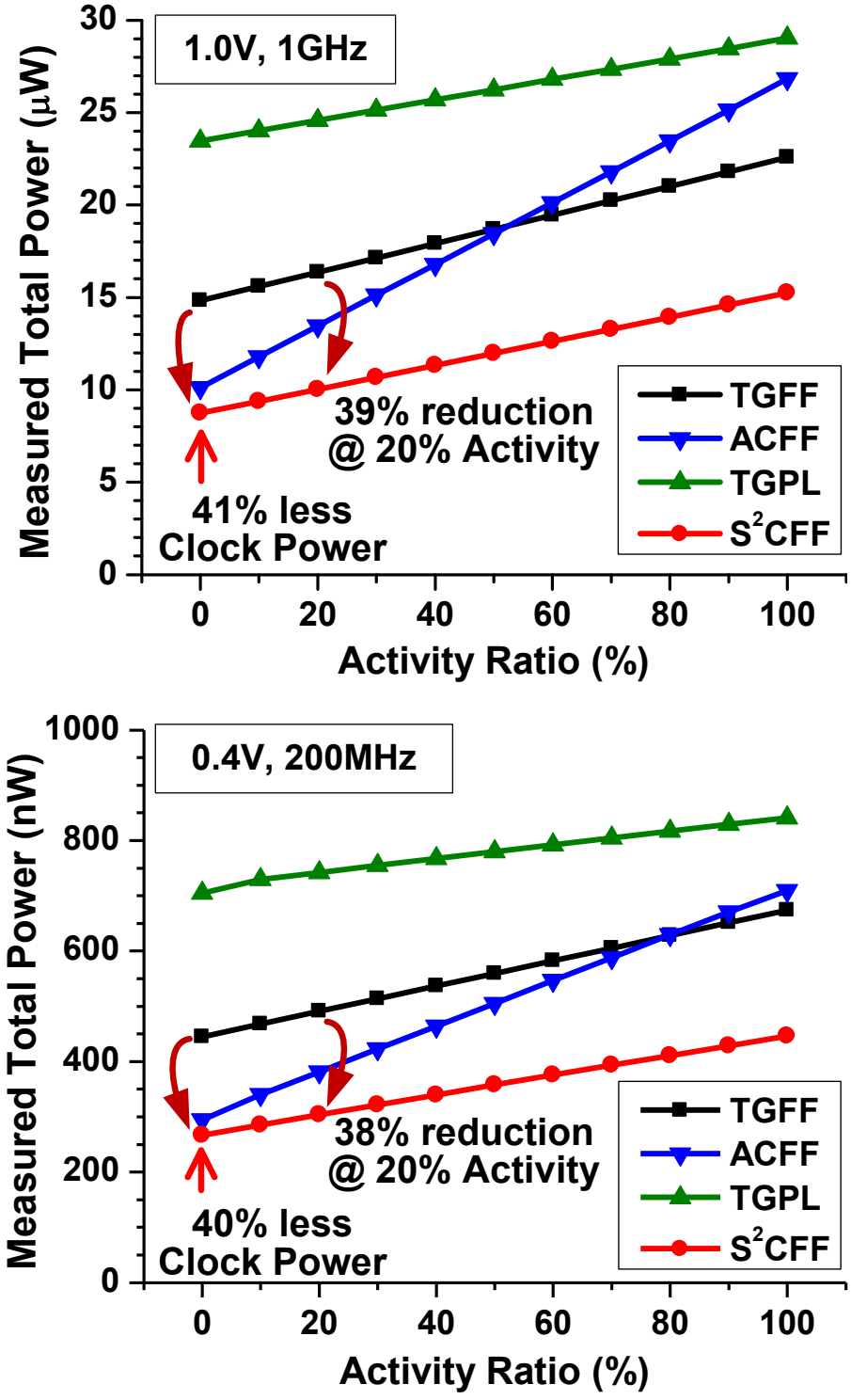


Figure 4.10: Measured total power

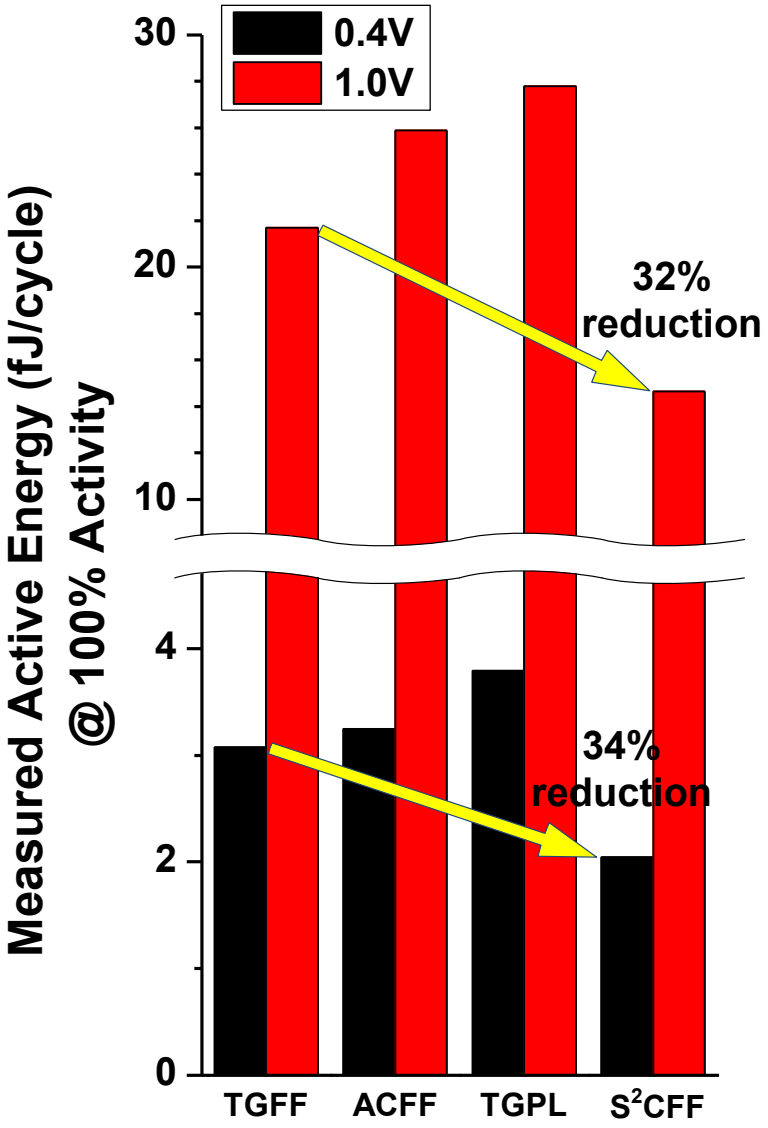


Figure 4.11: Measured energy

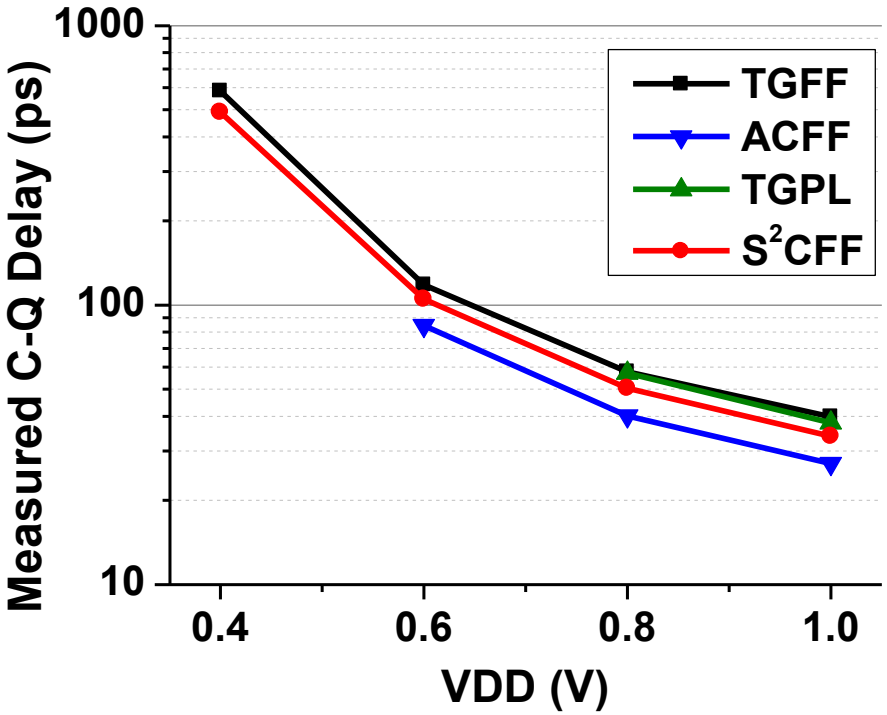


Figure 4.12: Measured C-Q delay

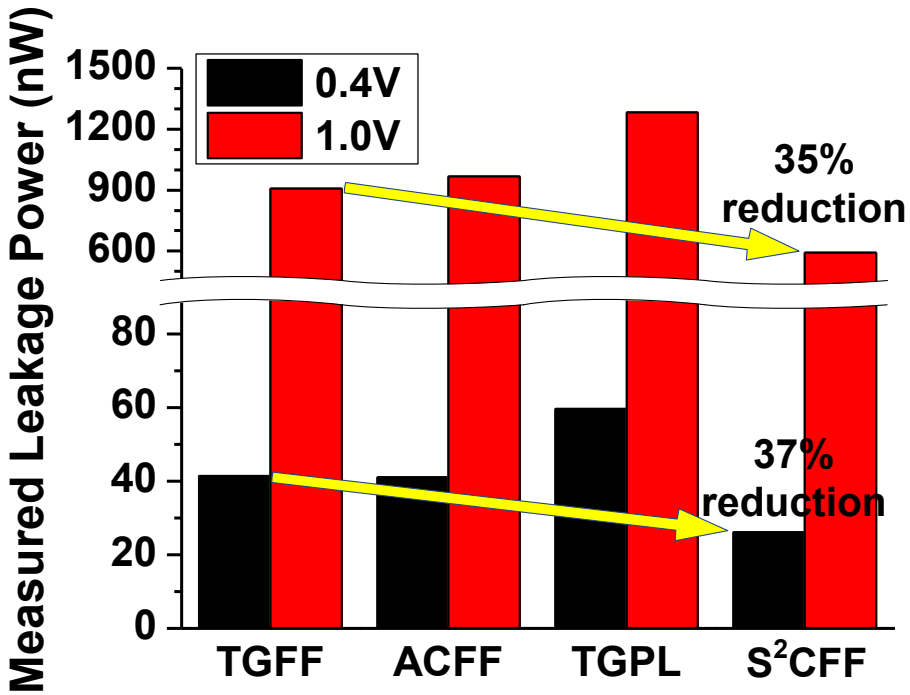


Figure 4.13: Measured leakage power

	S^2CFF (This Work)	**TGFF** Standard Cell Lib.	**ACFF** Teh, ISSCC'11	**TGPL** Naffziger, JSSC'02	**CSP^3L*** Consoli, ISSCC'12	**DMFF*** Nomura, ISSCC'08	**CPSA*** Ueda, ISSCC'06	**CCFF*** Kong, JSSC'01	**HLFF*** Partovi, ISSCC'96
Type	Static	Static	Static	Pulsed	Pulsed	Pulsed	Static	Pulsed	Pulsed
Contention-Free	Yes	Yes	No	Yes	Yes	No	No	No	No
Single Phase Clock	Yes	No	Yes	No	No	No	Yes	No	No
Number of Transistors	24	24	22[1)]	28[2)]	42[3)]	24[4)]	28	35	20
Normalized Layout Size	1.07	1.00	1.13	1.40					
Measured C-Q Delay @ 1.0V	33.9ps	39.8ps	27.1ps	37.9ps					
Measured Setup Time @ 1.0V	34.0ps	40.6ps	77.8ps	8.5ps					
Measured Hold Time @ 1.0V	-25.7ps	-31.4ps	-66.1ps	1.28ps					
Measured Total Power @ 1.0V, 1GHz, 20% Activity	10.02µW	16.36µW	13.45µW	24.57µW					
Measured Leakage @ 1.0V	592nW	909nW	967nW	1283nW					

1) It becomes 26 if ACE (Adaptive-Coupling Element) is added to the slave latch for low-voltage robustness
2) Delay element has 5 inverters to generate a pulse
3) 16 transistors (pulse generator) can be shared among multiple flip-flops.
4) Assuming 3 inverters are used for delay generation

*** CSP^3L, DMFF, CPSA, CCFF, HLFF are not implemented in this test chip.**

Table 4.3: Measurement and topology comparison of flip-flops

than TGFF at 1.0V and 0.4V, respectively. This is because S^2CFF has a fewer number of leakage paths than TGFF.

Finally, Table 4.3 includes the measured setup/hold time as well as the comparisons with other recently proposed flip-flops. S^2CFF has 15.5% faster 'setup time + C-Q delay' at 1.0V compared to TGFF, with the lowest power consumption among the compared flip-flops. The table also shows that S^2CFF is the only flip-flop that provides static, contention-free, and single-phase clock operations without increasing the device count compared to the conventional TGFF. While TGFF, ACFF, and TGPL have been already discussed in detail in the previous sections, other flip-flops also fail to meet these requirements: CSP^3L [48] is based on pulsed operation and does not provide single-phase clocking, while the device count exceeds that of TGFF; DMFF [49] has the same device count as TGFF, but it requires an clock inverter and Q node suffers from contention; CPSA [50] is a static, single-phase clocking flip-flop, but internal nodes suffer from contention; CCFF [51] also suffers from contention and area penalty (35 devices), and it is based on pulsed operation; HLFF [52] also has pulsed operation and requires clock inverters, and the output is not contention-free. The S^2CFF layout size is only 7% larger than TGFF, which corresponds to one poly-pitch increase in 45nm technology. The die photo of the test chip is shown in Figure 4.14 with the locations of the testing circuits annotated.

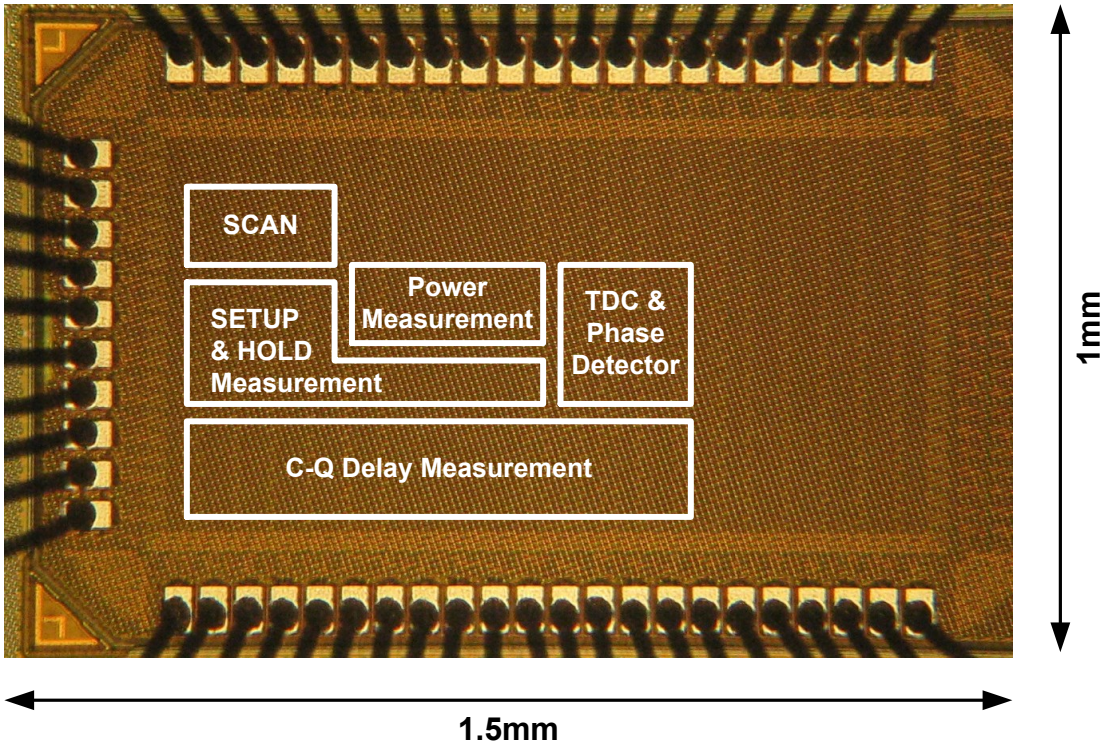


Figure 4.14: Die photo of the test chip fabricated in 45nm SOI

4.6 Conclusions

We presented a new flip-flop named S^{2}CFF which incorporates all the characteristics that an energy-efficient, highly voltage-scalable sequential element requires: static operation, contention-free transitions, single-phase clocking, and minimum or no area penalty compared to conventional ones. The robust operation with the lowest power consumption is demonstrated from the silicon measurements using the test chips fabricated in 45nm SOI. S^{2}CFF is reliably operating at near-threshold voltage (0.4V) and is one of the only two flip-flops that shows 100% yield across the wide V_{DD} range. The other flip-flop with the 100% yield is TGFF, but S^{2}CFF further reduces the power and energy consumptions, demonstrating 32% less active energy, 41% less clock power, and 35% less leakage power. It also improves 'setup time + C-Q delay' by 15.5%, and more importantly, all of these are achieved using the same device count as in TGFF, which implies that the area penalty is just negligible, if not zero. In this implementation, compared to the commercial TGFF, S^{2}CFF has only 7% larger layout size, which corresponds to one poly-pitch increase in

45nm SOI. It is also shown that the simple hold time path in S^2CFF results in a 3.4× reduction in hold time at the 3-σ value at near-threshold voltage (0.32V). All of these suggest that S^2CFF is an attractive candidate for sequential elements for low-power and highly voltage-scalable systems.

CHAPTER 5

A Testing Harness for Low-Voltage Flip-Flop Timing Characterization

5.1 Introduction

Electronic design automation (EDA) tools are indispensable in today's VLSI designs. The reliability of these tools depends on how accurate the devices and gates have been modeled. For example, more accurate MOSFET I-V characteristics in a SPICE model file can lead to more accurate simulation results.

If the modeling is not accurate, automatic place-and-route (APR) tools, for example, could insert unnecessarily many buffers for fixing the hold-time margin of flip-flops. While the functionality of the system remains same, in a large system where millions of flip-flops are used, these extra buffers would take up a significant portion of the total power. In addition, as the supply voltage becomes scaled down, the effects from any kind of variations, including the hold-time variation, can negatively impact the system yield and performance. Therefore, these variations must be addressed with a special concern at low V_{DD}. Although conventional flip-flops at low V_{DD} have been studied through simulations [38], it is hard to find any on-chip testing circuits aimed for actual silicon measurements at low V_{DD}. There are on-chip testing circuits proposed for accurate flip-flop measurements [53], but it is limited to the full (i.e., nominal) V_{DD} measurements.

In this chapter, we will discuss the issues in on-chip testing circuits for flip-flop timing characterizations, mainly focused on wide-range V_{DD} measurements, and then we will propose a new test-

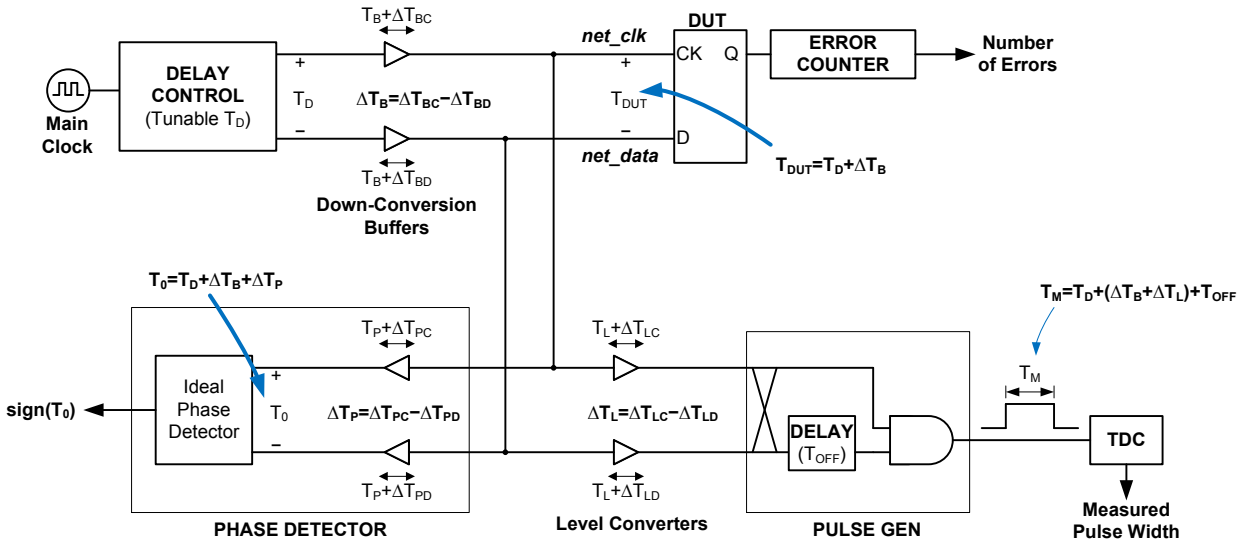


Figure 5.1: Mismatch sources in a setup/hold-time measurement circuit

ing harness for accurate low-voltage measurements. This technique will be demonstrated through silicon measurements.

5.2 Issues in Low V_{DD} Flip-Flop On-Chip Measurements

Figure 5.1 shows possible mismatch sources in the setup/hold-time measurement circuit used in Chapter 4 for timing characterization. The basic operation is explained in Section 4.4.1. The Delay Control runs with the Main Clock, and it generates CK and D signals depending on the tuning bits and bias voltages. The time difference between CK and D at the Delay Control output is T_D. Note that the Delay Control is running at the full voltage (V_{DD}). However, the DUT must be at a separate voltage domain (V_{DDL}), and this voltage could be lower than V_{DD} in order to measure the voltage dependency of setup/hold-time. Thus, there must be down-conversion buffers between the Delay Control and the DUT. Since there are two separate paths (CK and D), and each path has its own down-conversion buffer, there is a mismatch between those buffer delays. In Figure 5.1, each buffer's delay is T_{BC} and T_{BD} in the clock path and the data path, respectively, but each has its own delay variation at a lower voltage, which is indicated by ΔT_{BC} and ΔT_{BD}, respectively. Thus, they are combined together to generate the relative mismatch, $\Delta T_B = \Delta T_{BC} - \Delta T_{BD}$, and this mismatch appears at the DUT input (i.e., $T_{DUT} = T_D + \Delta T_B$) on *net_clk* and *net_data* as shown in

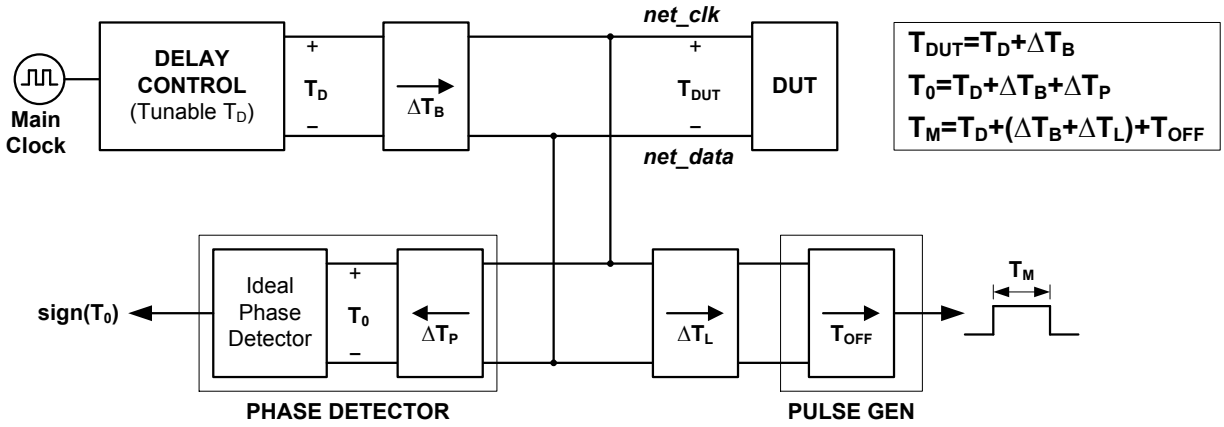


Figure 5.2: A simplified diagram of the mismatch sources in a setup/hold-time measurement circuit

the figure.

There are other in-accuracies involved in this testing circuit, too. Since the DUT is running at V_{DDL}, there must be level converters to generate pulses to be measured with the TDC. For accurate measurements, the TDC must be operating at the full voltage (V_{DD}). These level converters themselves also have mismatches as indicated by ΔT_L in the figure. However, the sum of this mismatch and the offset from the Pulse Gen ($\Delta T_L + T_{OFF}$) can be measured as long as the edges of *net_clk* and *net_data* are accurately aligned. The perfect alignment of *net_clk* and *net_data* indicates that the pulse width at the TDC input is just a sum of the level converter mismatch and the Pulse Gen offset (i.e., $T_M = \Delta T_L + T_{OFF}$), and this can be measured using the TDC. A Phase Detector shown in the figure is used to align those edges. However, this Phase Detector is not ideal, too. It can be modeled as a 'Ideal Phase Detector' and 'non-ideal input buffers' as shown in the figure. The 'Ideal Phase Detector' is assumed to have 'zero' mismatch, but now the 'non-ideal input buffers' have ΔT_P causing imperfect alignments of the *net_clk* and *net_data* signals. Thus, in real measurements, T_0 can be made zero by tuning the delays in the Delay Control, but this does not necessarily mean $T_{DUT} = 0$ due to the non-ideality (ΔT_P) of the Phase Detector.

All of these mismatch components are summarized and shown in Figure 5.2. Note that the mismatches can be effectively alleviated at the full voltage (V_{DD}) through device up-sizing and a careful layout. However, this becomes almost impossible at lower V_{DD} due to the severe variations.

5.3 A New Phase Detection Circuit for Low V_{DD} Operation

We discussed that the mismatches in the Phase Detector can result in inaccurate measurement. In other words, if the perfect alignment between the CK and D edges is guaranteed, the sum of the level converter mismatch (ΔT_L) and the Pulse Gen offset (T_{OFF}) can be measured and subtracted out from the final T_M value since T_M is given by the following equation:

$$T_M = T_D + \Delta T_B + \Delta T_L + T_{OFF} \tag{5.1}$$

Since T_{DUT} is a sum of T_D and ΔT_B,

$$T_{DUT} = T_D + \Delta T_B \tag{5.2}$$

Eq. (5.1) can be written as following:

$$T_M = T_{DUT} + \Delta T_L + T_{OFF} \tag{5.3}$$

We are interested in finding out T_{DUT} at which the DUT starts having setup/hold failures.

$$T_{DUT} = T_M - (\Delta T_L + T_{OFF}) \tag{5.4}$$

Since T_M can be measured using the TDC, the remaining unknown is $(\Delta T_L + T_{OFF})$. The only way to measure this is to perfectly align the CK and D edges on *net_clk* and *net_data*. This also must be done in a wide voltage-range to provide an accurate setup/hold-time measurement at low voltages. Therefore, this problem is narrowed down to a design of an accurate phase detector for a wide voltage-range.

Note that if D changes from 0 to 1 around the CK rising edge, there is no need to have a decent phase detector since just shorting *net_clk* and *net_data* will provide the perfect alignment, as suggested in Figure 5.3. However, it is difficult to align a D falling edge with a CK rising edge, and this is where the accurate phase detector is required. Since CK and D have the opposite directions, the traditional D-flip-flop or SR-latch approach incurs inaccuracies because at least one input from the two paths (CK or D) must have an additional inverter, hence causing imbalanced

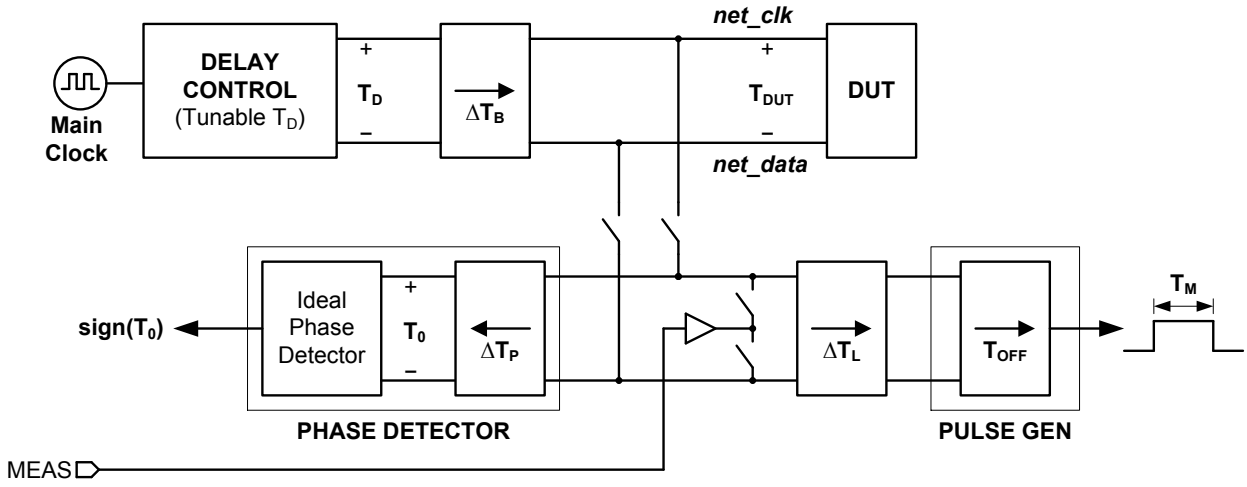


Figure 5.3: Edge alignment and offset ($\Delta T_L + T_{OFF}$) measurement when D rises

delays.

In order to solve this issue, we adopt an alternate approach, where one circuit detects "non-overlapping" of CK and D while the other circuit detects "overlapping". The key components of these approaches are shown in Figure 5.4, where the non-overlapping detector and the overlapping detector are shown. They are based on the dynamic NOR/NAND structures, and the example waveforms are shown in the figure. The reason of using the dynamic structures is that, if there is only a slight non-overlap (or a slight overlap), then *net0* would see just a small glitch, but the phase detector should be able to detect it. The conventional static-approaches (D-flip-flop or SR-latch) cannot do this because they require a voltage rise to be larger than their trip point, which is usually around the half-V_{DD}. From corner simulations, the worst-case error of these dynamic structures is 0.061 FO4 and 0.057 FO4 at 1.0V and 0.3V, respectively. In addition, by using periodic CK and D signals and running this circuit for many cycles, it can tolerate more non-idealities.

Figure 5.5 shows the whole circuit diagram of the phase detector. The Non-Overlapping and Overlapping detection circuits are at the core of this circuit, and the SR-latches and the flip-flops are sampling the output from the detection circuits which are then fed into a controller circuit that counts the number of CK_LEAD and D_LEAD. There are also static NOR and AND gates at the bottom of the circuit diagram for trivial cases where the amount of the non-overlapping (or overlapping) duration is sufficiently long enough to trip the static gate's output. The Disable Control block resets the Non-Overlapping/Overlapping detection circuits after some amount of

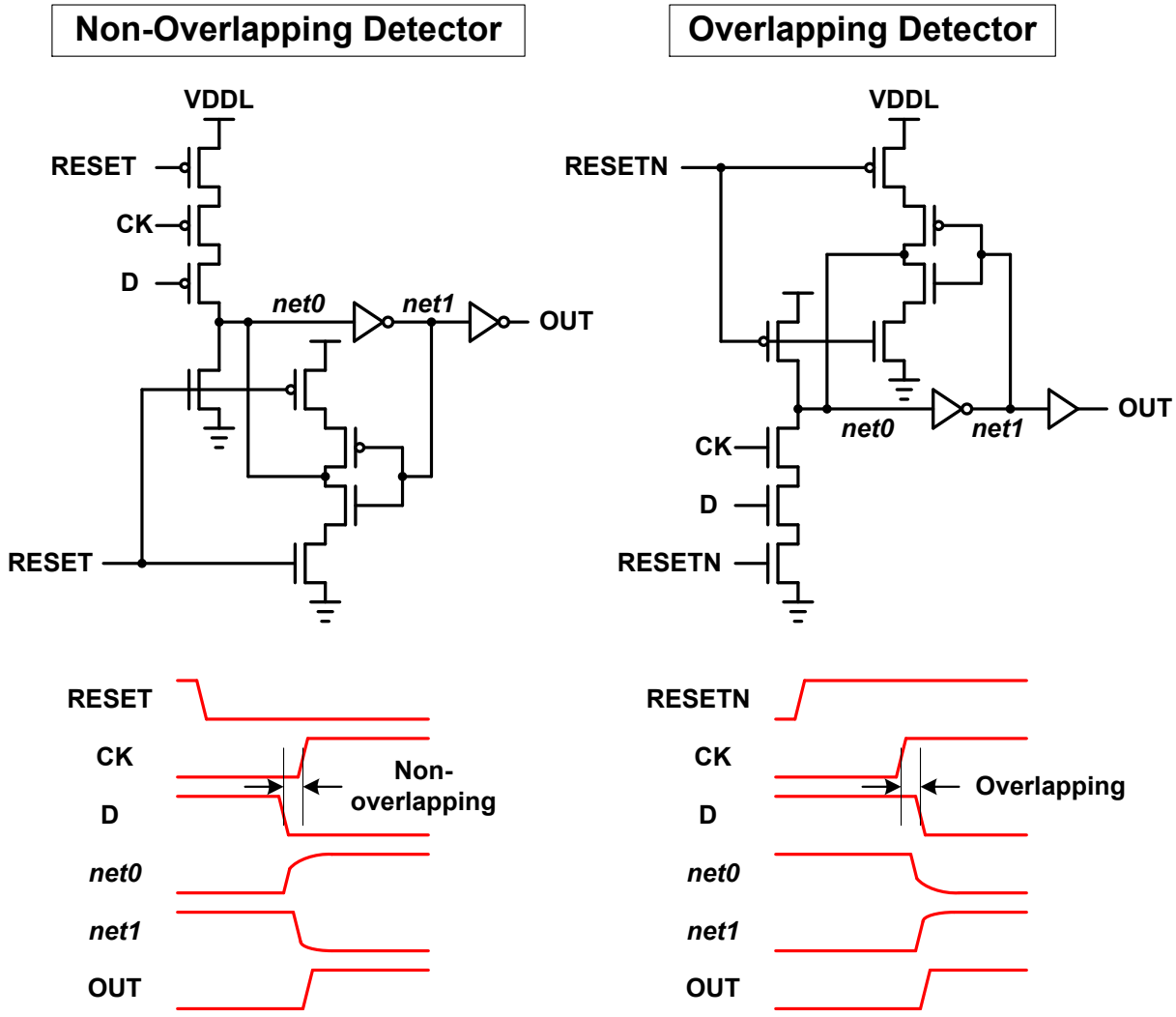


Figure 5.4: Dynamic NAND/NOR structures for edge alignment

delay from the CK rising edge; this is to prevent a false trigger of the dynamic circuits due to the leakage current. All of this operation is repeated many times, and the outputs from the detection circuits increment counters, which can be then used to determine the edge alignment.

5.4 A Setup/Hold-Time Measurement Circuit for Wide Voltage-Range Operation

Figure 5.6 shows an overall circuit diagram of the proposed setup/hold-time measurement circuit. The Delay Control, Down-Conversion Buffers, Level Converters, and the Pulse Gen are same

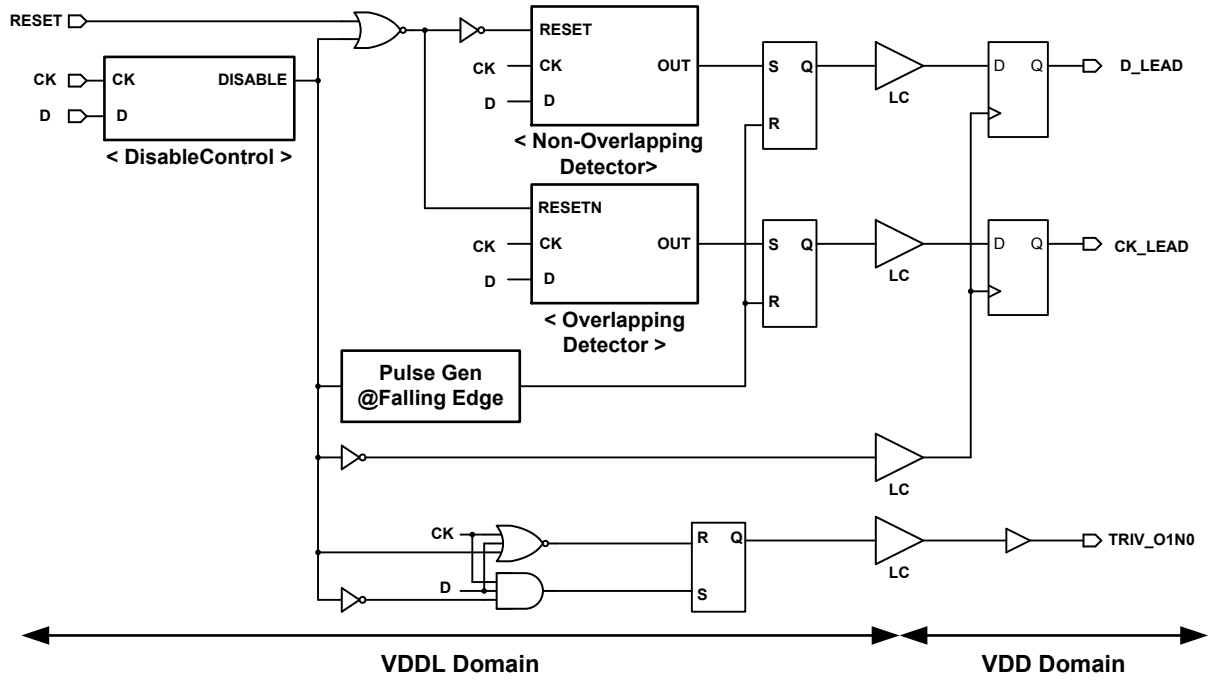


Figure 5.5: Phase detector circuit diagram

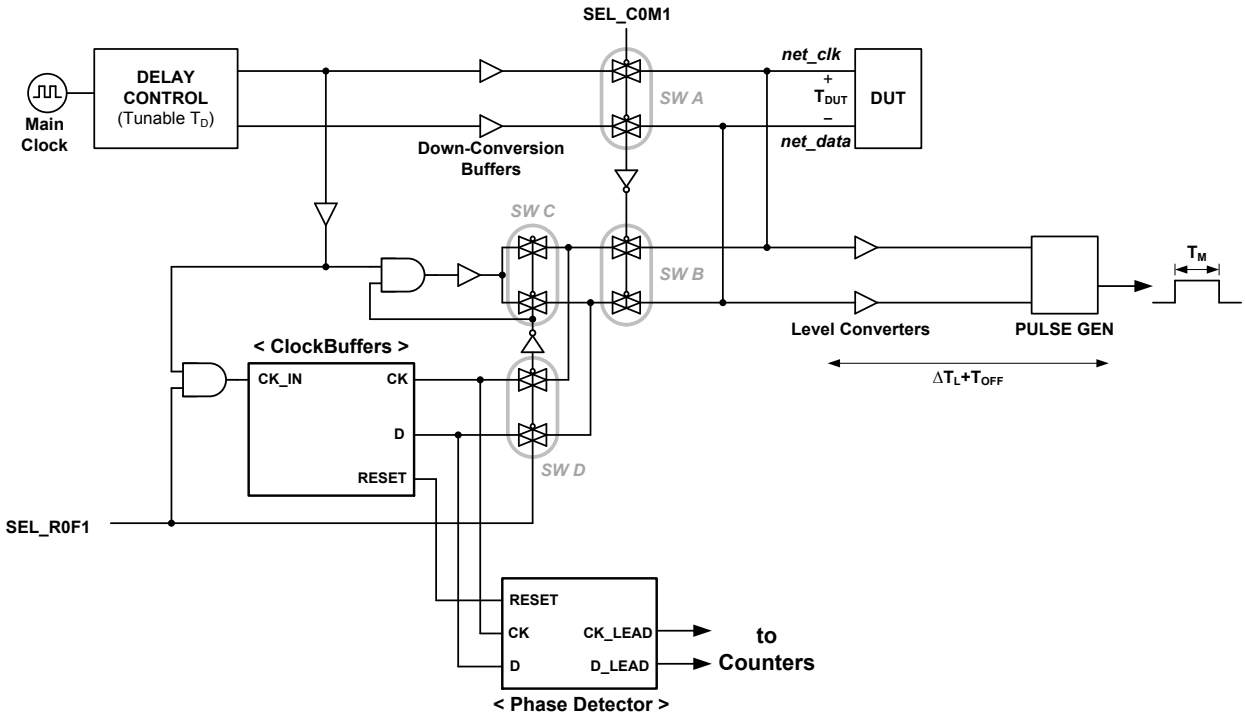


Figure 5.6: Setup/hold-time measurement circuit

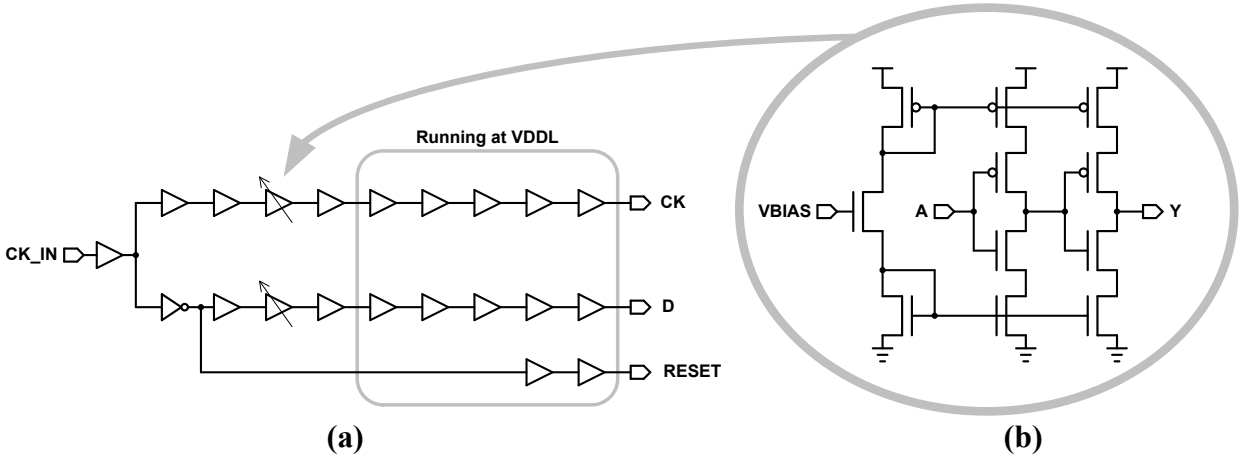


Figure 5.7: (a) Clock Buffer schematic (b) Current-starved buffer for delay tuning

as in the previous circuit shown in Figure 5.1. The Phase Detector is the one described in Section 5.3. There are four pairs of switches (transmission-gates) and they are controlled at the full voltage (V_{DD}) to minimize their channel resistance. When measuring the offset value ($\Delta T_L + T_{OFF}$) for a D-rising edge, SW B and SW C are on, while SW A and SW D are off. This provides a short between the inputs of the two level converters, so the perfect alignment of *net_clk* and *net_data* is guaranteed. Then, the Main Clock provides a periodic signal, and the corresponding pulse width ($T_M = \Delta T_L + T_{OFF}$) is measured. When measuring the offset value for a D-falling edge, SW B and SW D are on, while SW A and SW C are off. Then, the Main Clock provides a periodic signal to the Clock Buffer. The schematic of this Clock Buffer is shown in Figure 5.7, which generates the CK and D signals as well as the RESET signal for the Phase Detector. The analog bias voltage (V_{BIAS}) in Figure 5.7(b) shall be kept being changed until the Phase Detector outputs indicate that there is a good alignment between the CK and D edges. At this point, the corresponding offset value ($T_M = \Delta T_L + T_{OFF}$) can be measured. Finally, in order to check setup/hold-time failures, SW A is on, while the others are off. The Delay Control tuning bits and voltages shall be kept being changed until the DUT fails, and then the corresponding pulse width ($T_M = T_{DUT} + \Delta T_L + T_{OFF}$) can be measured. Once $\Delta T_L + T_{OFF}$ is subtracted from T_M, the remaining T_{DUT} will be the final setup- (or hold-) time.

It should be noted that all the important mismatch values, such as ΔT_L, can be subtracted out using the provided switches. Also, the reliable operation of the Phase Detector allows a wide voltage-range timing characterization of flip-flops.

V_{DD}		TGFF	S^{2}CFF	Improvement
1.00V	Mean	6.30ps	5.62ps	-
	Sigma	4.84ps	2.11ps	2.3×
	Maximum	24.38ps	11.14ps	2.2×
	Minimum	-3.28ps	0.33ps	-
0.40V	Mean	3.66ps	22.63ps	-
	Sigma	40.72ps	23.23ps	1.8×
	Maximum	155.27ps	82.49ps	1.9×
	Minimum	-84.91ps	-35.61ps	-
0.35V	Mean	31.42ps	37.46ps	-
	Sigma	97.69ps	46.33ps	2.1×
	Maximum	351.21ps	167.10ps	2.1×
	Minimum	-184.17ps	-74.87ps	-
0.32V	Mean	11.38ps	51.48ps	-
	Sigma	111.69ps	74.86ps	1.5×
	Maximum	486.52ps	276.03ps	1.8×
	Minimum	-217.53ps	-130.53ps	-

Table 5.1: Comparison of the hold-time variations of TGFF and S^{2}CFF (172 flip-flops of each type)

5.5 Measurements

Test chips were fabricated in a 45nm SOI technology. Each test chip contains 4 TGFFs and 4 S^{2}CFFs. 43 chips have been measured using the proposed timing characterization circuit, thus the sample size is 172 for each flip-flop. Hold-time distributions are measured at the full V_{DD} (=1.0V), 0.40V, 0.35V, and 0.32V, where ∼0.35V indicates the near-V_{TH}. Histograms from the 172 flip-flops of each type are shown in Figure 5.8 and Figure 5.9, measured at each specified voltage, and the statistical results are summarized in Table 5.1. Also, an average value from each chip (i.e., an average value of the hold-time of the 4 flip-flops of each type in the same chip) is calculated, hence total 43 average values, and shown as histograms in Figure 5.10 and Figure 5.11, measured at each specified voltage. This is to observe chip-to-chip variations while reducing effects from within-die variations. Statistical results from these distributions are summarized in Table 5.2.

From these measurements, it is obvious that S^{2}CFF provides much less hold-time variations. In Figure 5.8 and Figure 5.9, also summarized in Table 5.1, it shows 2.3× and 2.1× less sigma values at 1.0V and 0.35V, respectively, mainly because of the simple hold-time path described in Section 4.3.2. The most critical measurement is the 'Maximum' value of the hold-time, since a hold-

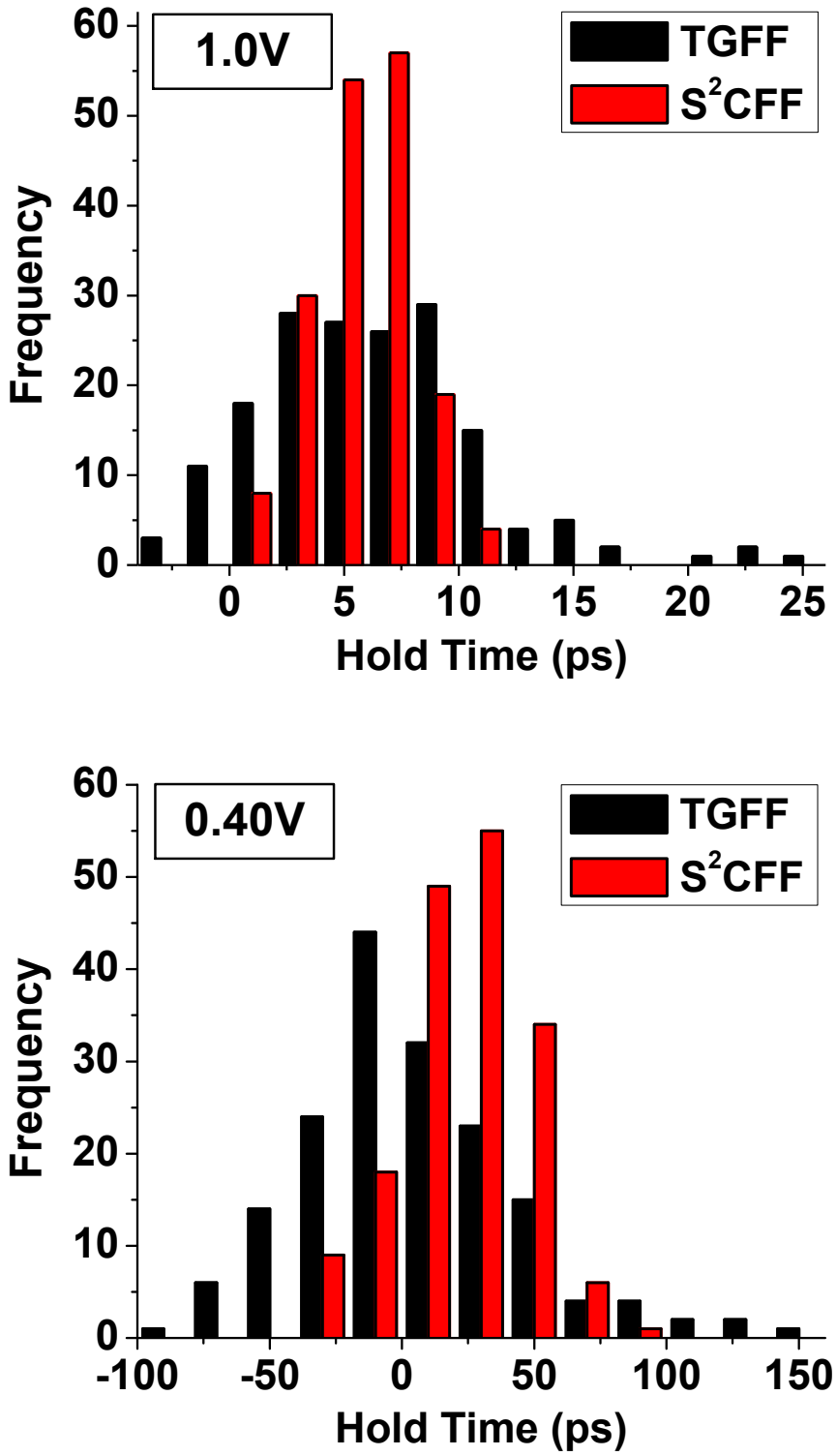


Figure 5.8: Hold-time distribution of TGFF and S^2CFF at 1.0V and 0.4V (172 flip-flops of each type)

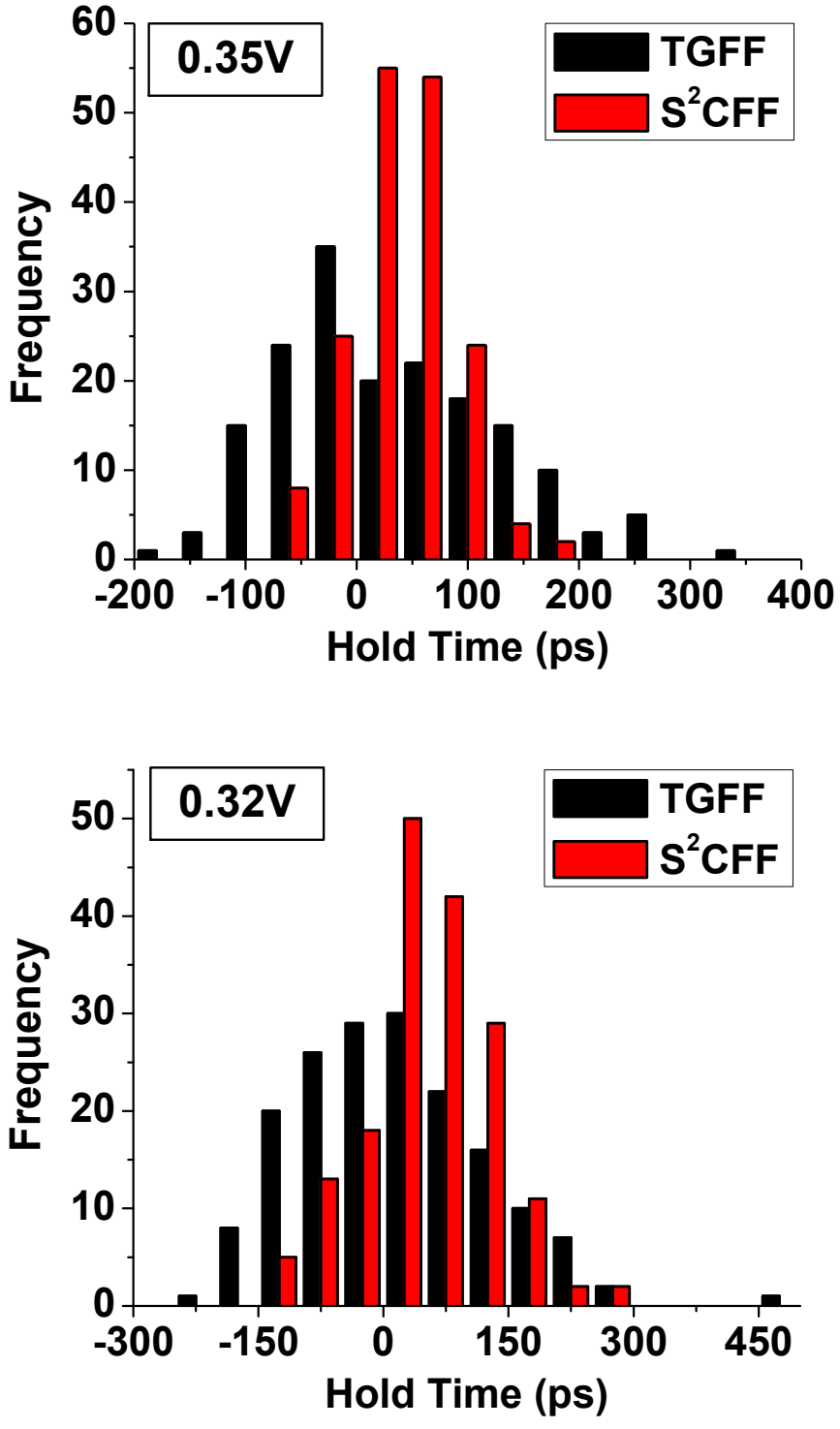


Figure 5.9: Hold-time distribution of TGFF and S^2CFF at 0.35V and 0.32V (172 flip-flops of each type)

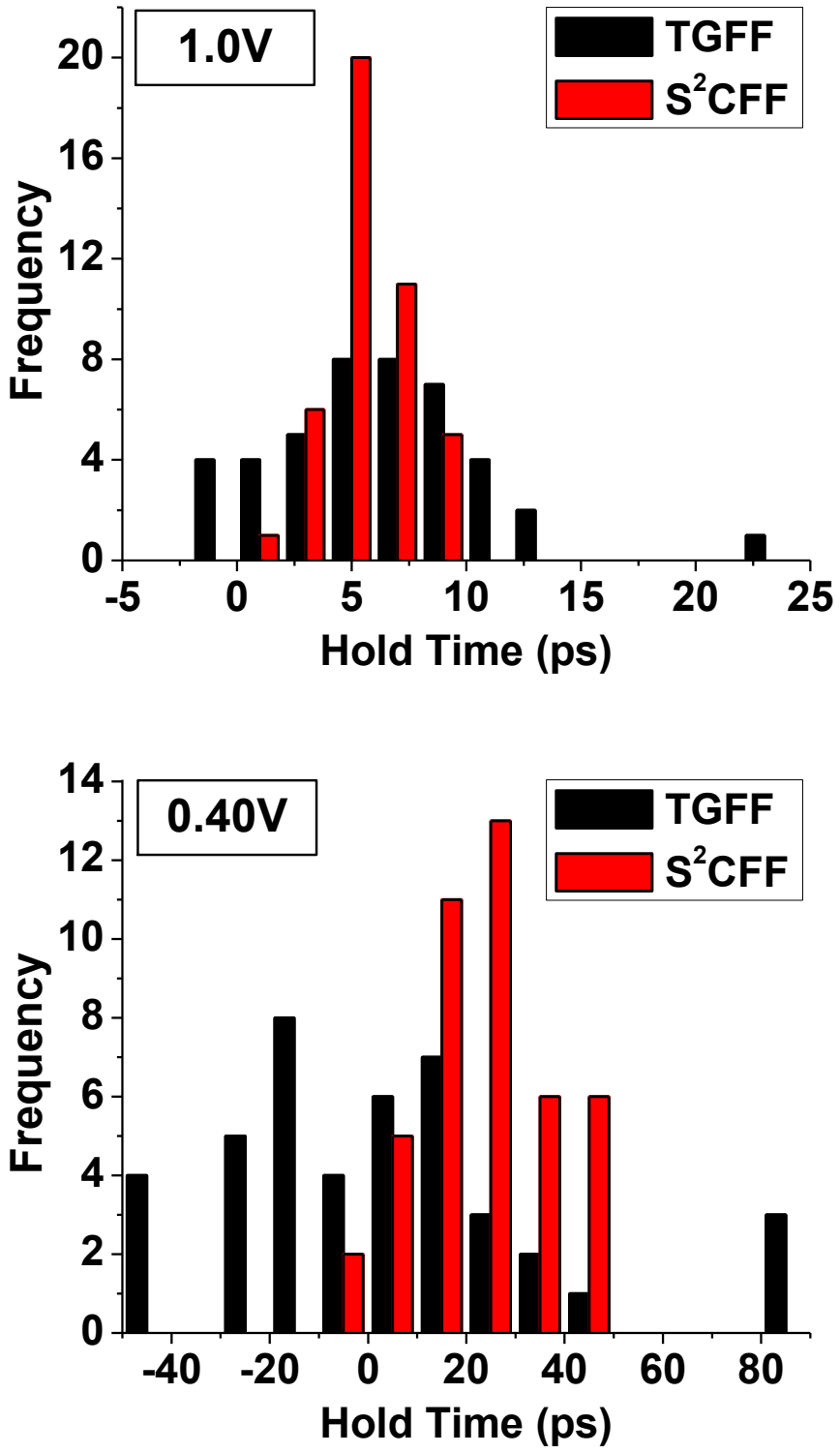


Figure 5.10: Hold-time distribution of TGFF and S^2CFF at 1.0V and 0.4V (43 chips)

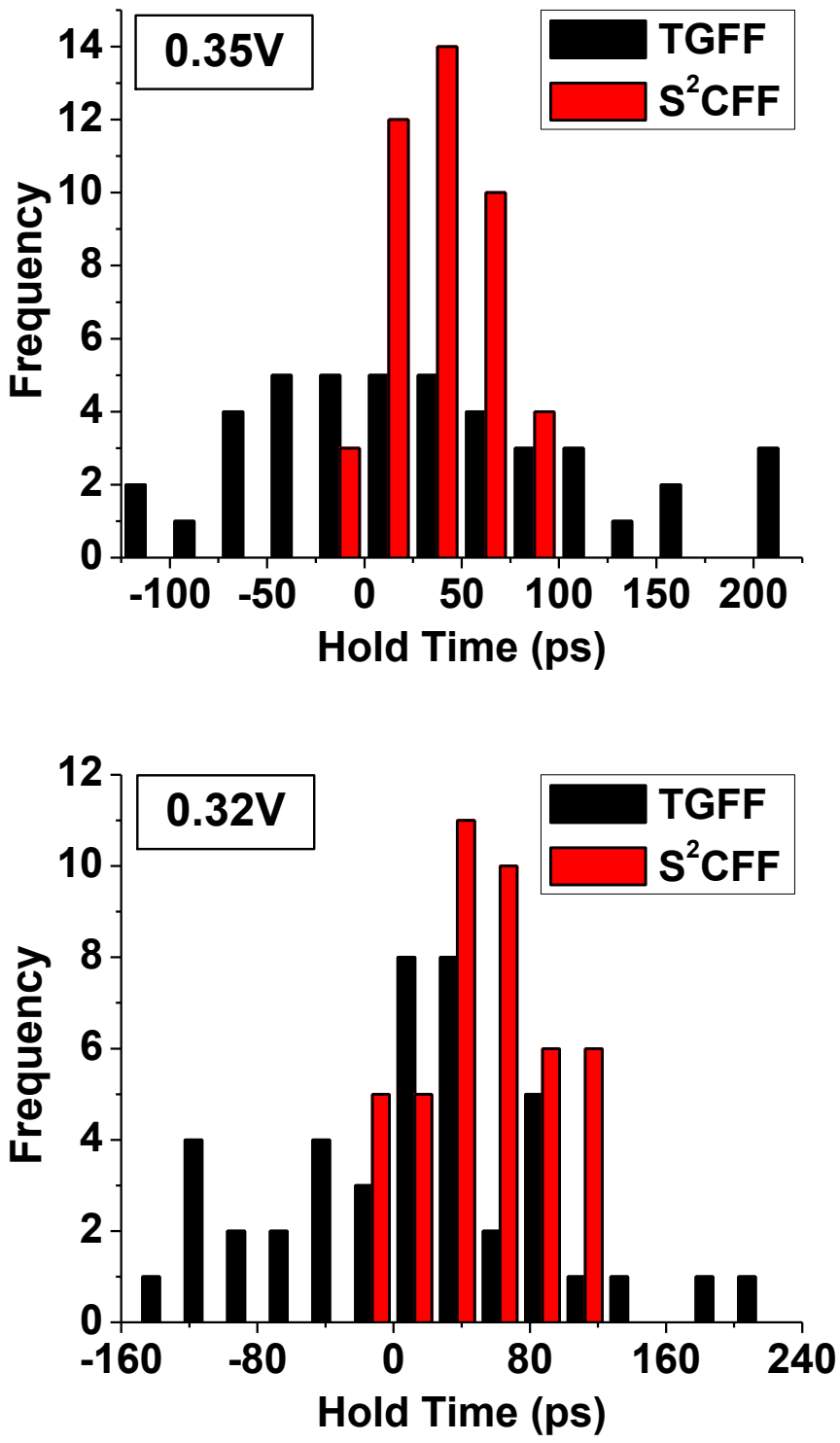


Figure 5.11: Hold-time distribution of TGFF and S^2CFF at 0.35V and 0.32V (43 chips)

V_{DD}		TGFF	S^{2}CFF	Improvement
1.00V	Mean	6.30ps	5.62ps	-
	Sigma	4.53ps	1.65ps	2.7×
	Maximum	22.89ps	8.81ps	2.6×
	Minimum	-1.20ps	1.93ps	-
0.40V	Mean	3.66ps	22.63ps	-
	Sigma	31.40ps	12.43ps	2.5×
	Maximum	88.71ps	46.44ps	1.9×
	Minimum	-48.34ps	-4.54ps	-
0.35V	Mean	31.42ps	37.46ps	-
	Sigma	83.07ps	24.68ps	3.4×
	Maximum	218.98ps	85.44ps	2.6×
	Minimum	-103.07ps	-6.39ps	-
0.32V	Mean	11.38ps	51.48ps	-
	Sigma	76.75ps	40.27ps	1.9×
	Maximum	202.51ps	124.96ps	1.6×
	Minimum	-138.15ps	-23.36ps	-

Table 5.2: Comparison of the hold-time variations of TGFF and S^{2}CFF (43 chips)

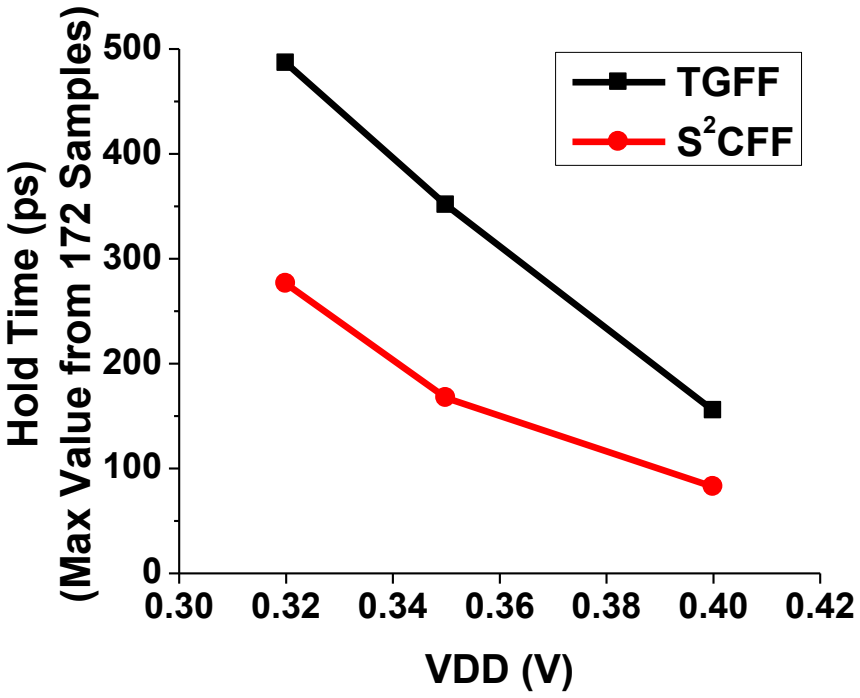


Figure 5.12: Maximum hold-time value from the measured 172 flip-flops of each type

time fix process in a system design must take the worst-case value of the hold-time into account, adding buffers in order to make the shortest path delay exceed the worst-case hold-time. It is

clearly shown that S^{2}CFF provides 2.2$\times$ and 2.1$\times$ reduction in the maximum hold-time at 1.0V and 0.35V, respectively, implying that it can reduce the number of the hold-time fixing buffers by $> 2\times$, followed by overall system power reduction and yield improvement.

S^{2}CFF shows much more improvements in the hold-time variations when it comes to chip-to-chip variations. In Figure 5.10 and Figure 5.11, also summarized in Table 5.2, S^{2}CFF shows 2.7$\times$ and 3.4$\times$ less sigma values at 1.0V and 0.35V, respectively. The figures suggest that TGFF has significantly degraded variations especially at low voltages, whereas S^{2}CFF still maintains good spreads at low voltages. As explained in Section 4.3.2, TGFF's hold-time is mainly determined by the mismatches among several gates. Since it is prone to any kind of variations, it is not unexpected that the global variations (i.e., chip-to-chip variations) have more effects compared to the local variations (i.e., within-die variations). In contrast, S^{2}CFF's hold-time is mainly determined by the discharging speed through PATH_{CK} (Figure 4.6), so it shows smooth bell-shaped distributions in all the measurements even at near-V_{TH}.

The maximum hold-time values from the 172 flip-flops of each type are also plotted in Figure 5.12 to show a trend. The maximum hold-time value of S^{2}CFF at 0.32V is even shorter than the maximum hold-time of TGFF at a higher voltage (0.35V). Therefore, S^{2}CFF can provide either: 1) a smaller number of buffers added for hold-time fix; 2) a lower V_{MIN}. Both benefits can lead to an overall system power reduction, while still guaranteeing the system robustness (i.e., no hold-time failure).

A die photo is provided in Figure 5.13.

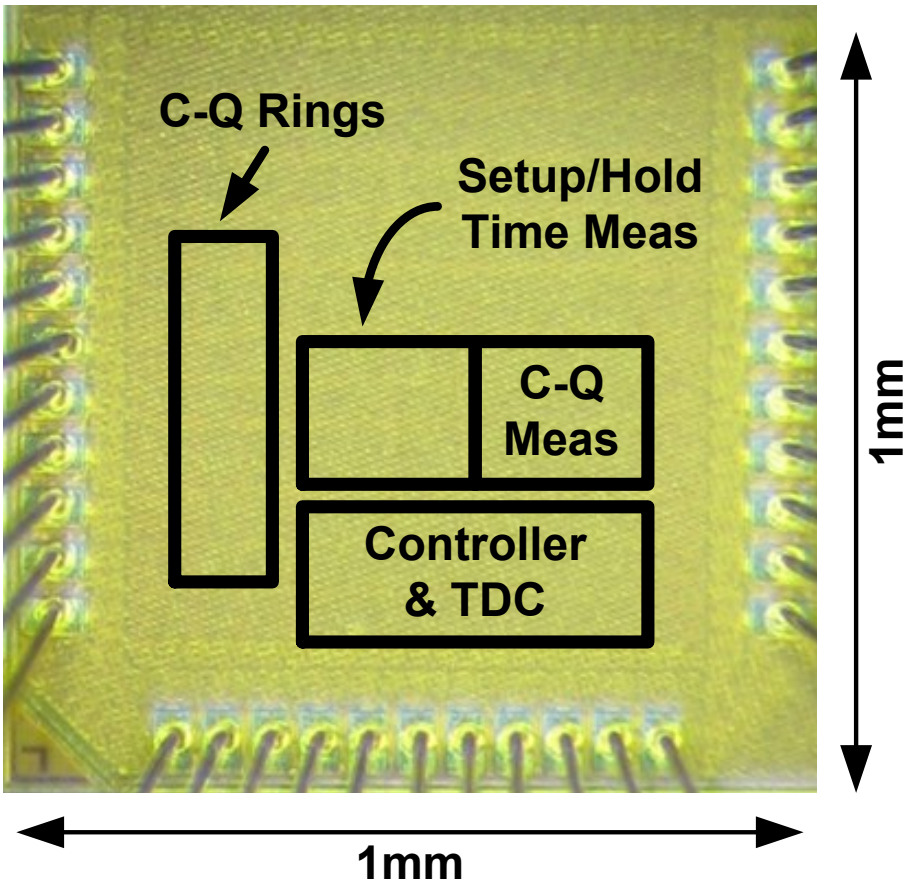


Figure 5.13: Die photo of the test chip fabricated in 45nm SOI

CHAPTER 6

Conclusion

The on-going demand for achieving faster computing speed has met a major huddle in increasing the clock frequency due to the excessive power consumption. Thus, in recent years, low-power design is not optional anymore; it has become one of the most important design criteria that virtually all digital/analog circuits should meet. Voltage scaling is an effective way to reduce the overall power consumption, but the major challenges in sub- or near-V_{TH} operations include performance degradation and reliability issues due to PVT variations. Although the performance degradation could be compensated by utilizing more parallelism (e.g., multi-core systems), the reliability concerns must be correctly addressed during design phase in order to avoid serious system failure.

In this dissertation, we identified several important circuit components that are prone to such variations in NTC, proposed new techniques to improve robustness, and demonstrated the effectiveness through silicon measurements.

Level converters are critical components in voltage-scaled VLSI systems in that they must provide a reliable interface between two different voltage domains. Digital cores tend to run at severely voltage-scaled domains, while other analog/peripheral circuits still require a high voltage, and especially in the NTC region, the reduced I_{ON}/I_{OFF} ratio makes it extremely difficult to achieve robust level conversions. In Chapter 2, we proposed two static level converter designs called LC^2 and SLC. LC^2 adopts a novel thyristor and pulsed-operation and modulates its pull-up strength depending on its state. During idle state where there is no input change, it holds the internal state through the week keepers, whereas the strong devices running at V_{DDH} participate in actual signal transitions when the input changes. The device sizing of the keepers are the most important

design criteria in LC^2. We demonstrated that it can easily meet the 3σ robustness requirement through the systematic approach using the current margin plot. Because the actual transitions are handled by the strong devices, LC^2 provides the fastest performance compared to other designs, demonstrating 3.2× speed improvement over DCVS. SLC inherently reduces the contention by incorporating diodes in the stack, so that the pull-down devices are fighting with the diode whose $|V_{GS}|$ corresponds to the diode voltage-drop (V_D). Compared to other designs where the pull-down devices contends with a strong PMOS device whose $|V_{GS}|$ is usually $\sim V_{DDH}$, SLC provides a great improvement in the robustness resulting in 98.93% yield from Monte-Carlo simulations as well as no failure in a wide temperature range during silicon measurements. Moreover, the simple schematic and the small layout size of SLC make it suitable to fit in standard-cell libraries and could streamline the system design process.

SRAMs exist in virtually all processors. However, they are also a major bottleneck in voltage-scaling due to its inherent ratioed bitcell design. Widely-used 8T bitcells decouples READ and WRITE operations, eliminating the two-sided constraint, at the cost of a larger bitcell size. Usually, the area overhead is in a $30 \sim 55\%$ range, thus sometimes preventing it to be used in severely area-constrained applications. In Chapter 3, we proposed a novel 7T SRAM bitcell and the peripherals, in order to alleviate the area overhead and provide a robust operation. The Auto-Shut-Off sensing effectively eliminates the short-circuit current from unselected cells, resulting in a 6.8× READ energy reduction. Also, the 7T bitcell's innate bitline leakage suppression effect in un-selected bitcells resulting from negative V_{GS} on their READ device provides the 113× less bitline leakage compared to the conventional 8T memory through the simulation. This Quasi-Static READ has been also demonstrated through the silicon measurement which shows the much improved READ error rate. In addition, the use of PMOS transistors as Pass-Gate devices improves the half-select robustness by directly modulating the transistor $|V_{GS}|$ through the WRITE bitline voltage. The silicon measurement shows a robust bit-interleaved operation and achieves the 3.35fW/bit leakage power.

The clocked sequential element, a flip-flop in short, is ubiquitous in today's digital systems. While many flip-flop designs have been proposed, the main issue has still remained same: the hold-time variation. This often causes unnecessarily excessive buffer insertions to meet the hold time margin under the severe PVT variations. Also, in terms of robustness and design-overhead, it

is very hard to find a flip-flop that is static and contention-free with negligible or no area overhead compared to the widely-used TGFF. In Chapter 4, we proposed a new flip-flop called S^2CFF. It is single-phase, meaning that it does not require the inverted clock signal. It is static and contention-free, and it also has the same number of devices (24 transistors) as in the TGFF. This makes the area overhead of S^2CFF quite negligible. It is the only flip-flop that meets all of these requirements (single-phase, static, contention-free, same device count) among the compared baseline designs. Mainly due to the single-phase operation, S^2CFF shows a $\sim 40\%$ power reduction compared to the TGFF through silicon measurements. In addition, due to its static and contention-free operation, it demonstrates the robust low-voltage operations similar to TGFF, reliably running at 0.4V, while other designs fail. Another benefit of S^2CFF is its simple hold-time path. This reduces its mismatches that determine the hold-time, followed by $3.4\times$ improvement in 3σ hold-time compared to TGFF.

The flip-flop testing harness for the timing characterization was also discussed and demonstrated through the silicon measurements. This testing harness incorporates the dynamic NAND/NOR structures and many-cycle operations, in order to more accurately align the CK and D edges. This makes it easy to measure the offset caused by the severe mismatches in low V_{DD} operations, so the offset can be easily subtracted out through a simple calculation. By measuring the testchips, it was demonstrated that S^2CFF has up to $3.4\times$ reduction in the standard deviation of the measured hold-time at 0.35V, compared to the TGFF. It was also showed that S^2CFF at 0.32V has a better worst-case hold-time, even when compared to TGFF at a higher voltage (0.35V).

All of these new circuit techniques proposed in this dissertation can be extensively used in most VLSI systems. Especially, the NTC operations could benefit more from the proposed techniques, since the new circuits are targeted for much improved robustness while still providing excellent performance and low power consumption. The wireless sensor node platform [7] mentioned in Chapter 1 already uses SLC as its standard level conversion circuits and demonstrates robust and power-efficient operations with three different voltage domains (0.6V/1.2V/3.6V), while the 7T SRAM and S^2CFF are also planned to be implemented in future-version of the system. We expect that these robust circuit designs for low-voltage VLSI can foster the development of future low-power system designs.

6.1 Future Works

Based on the circuit techniques presented in this dissertation, there are other possibilities to further improve circuit robustness and performance in low-voltage VLSI. As mentioned before, the 7T SRAM is planned to be implemented in the wireless sensor node platform [7], which currently has only a 3kB SRAM, and this SRAM capacity is a limiting factor in achieving more flexible system functionality. The bitcell size of the 10T bitcell used in the current version of the sensor node is almost $2\times$ larger than the 7T bitcell size, so from a simple estimation, it is expected to have at least $\sim$6kB of SRAM capacity in the future version by having the 7T SRAM. One more advantage of the 7T SRAM is that, it provides a much more robust bit-interleaving capability, and this will further improve the array efficiency. There are other concerns specifically related to this sensor node platform; for example, its extremely low sleep power requirement enforces a use of HVT (I/O) devices in the bitcells. In order to achieve a reliable operation with these HVT devices, the decoupled READ and WRITE is a must. This necessitates a use of bitcells that have $>$6 devices, unless a peripheral assist circuit is also implemented. Most of conventional peripheral assist techniques, such as [54][55][56][57], usually incur a non-negligible area/power overhead. In addition, it is hard to find a decent assist circuit that is very effective under severe variations at such a low voltage [58]; note that the supply voltage used in the sensor node system [7] is 0.6V which is a sub-V_{TH} regime since the HVT devices' threshold voltage is in the range of 0.7V $\sim$ 0.8V. One of the ways to assist the HVT bitcells is to utilize their extremely low leakage. For example, even if the supply voltage becomes lost, these HVT bitcells could retain their data for a limited amount of time, and it is expected to survive a longer power-loss duration compared to standard SVT bitcells. A similar approach has been presented in an advanced process node (hence with more device variations) [59]. It is interesting that the most advanced process nodes and the sub-V_{TH} operations in old (hence mature) process nodes bear a similarity in that both are prone to variations.

The small size and the low power consumption of the sensor node platform will enable many new applications most of which have been regarded impossible due to their size and power limitations. Some examples found in recent literature include glucose monitoring systems [60][61] and other bio applications [62][63][64]. However, most of them still suffer from limited battery-life

and system functionality. Developing a robust and flexible sensing node platfrom through further circuit innovations are the most primary future goal of this dissertation.

6.2 Related Publications and Patents

- Yejoong Kim, Dennis Sylvester, and David Blaauw, "LC^2: Limited Contention Level Converter for Robust Wide-Range Voltage Conversion," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2011, pp. 188–89.

- Yejoong Kim, Yoonmyung Lee, Dennis Sylvester, and David Blaauw, "SLC: Split-Control Level Converter for Dense and Stable Wide-Range Voltage Conversion," in *Proc. European Solid-State Circuits Conf.*, Sep. 2012, pp.478–481.

- Yejoong Kim, Dennis Sylvester, and David Blaauw, "A 3.35fW/bit Bit-Interleaved 7T SRAM with Quasi-Static Read and Auto-Shut-Off Sensing," planned to be submitted to *IEEE J. Solid-State Circuits*, 2015.

- Yejoong Kim, Wanyeong Jung, Inhee Lee, Qing Dong, Michael Henry, Dennis Sylvester, and David Blaauw, "A Static Contention-Free Single-Phase-Clocked 24T Flip-Flop in 45nm for Low Power Applications," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2014, pp. 466–467.

- Yejoong Kim, Michael Brewer Henry, Dennis Michael Sylvester, David Theodore Blaauw, "Static Signal Value Storage Circuitry Using a Single Clock Signal," US Patent 13/860,756, filed on April 11, 2013.

- Yejoong Kim, Dennis Michael Sylvester, David Theodore Blaauw, Brian Tracy Cline, "Measurement Circuitry and Method for Measuring a Clock Node to Output Node Delay of a Flip-Flop," US Patent 14/175,015, filed on February 3, 2014.

BIBLIOGRAPHY

[1] S. Rusu, H. Muljono, D. Ayers, S. Tam, W. Chen, A. Martin, S. Li, S. Vora, R. Varada, and E. Wang, "Ivytown: A 22nm 15-Core Enterprise Xeon® Processor Family," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2014, pp. 102–103.

[2] A. Wang, K. C. Smith, and L. C. Fujino. (2013, Nov. 1). *ISSCC 2014 Trends* [Online]. Available: http://www.isscc.org/doc/2014/2014_Trends.pdf

[3] S. Mathew, M. Anders, B. Bloechel, T. Nguyen, R. Krishnamurthy, and S. Borkar, "A 4GHz 300mW 64b Integer Execution ALU with Dual Supply Voltages in 90nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2004, pp. 162–163.

[4] G. Chen, H. Ghaed, R. Haque, M. Wieckowski, Y. Kim, G. Kim, D. Fick, D. Kim, M. Seok, K. Wise, D. Blaauw, and D. Sylvester, "A Cubic-Millimeter Energy-Autonomous Wireless Intraocular Pressure Monitor," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2011, pp. 310–311.

[5] Y.-S. Kuo, P. Pannuto, G. Kim, Z. Foo, I. Lee, B. Kempke, P. Dutta, D. Blaauw, and Y. Lee, "MBus: A 17.5pJ/bit/chip Portable Interconnect Bus for Millimeter-Scale Sensor Systems with 8nW Standby Power," in *Proc. IEEE Custom Integrated Circuits Conference*, Sep. 2014.

[6] R. Viswanath, V. Wakharkar, A. Watwe, and V. Lebonheur, "Thermal Performance Challenges from Silicon to Systems," *Intel Tech. J.*, Q3 2000.

[7] Y. Lee, S. Bang, I. Lee, Y. Kim, G. Kim, M. H. Ghaed, P. Pannuto, P. Dutta, D. Sylvester, and D. Blaauw, "A Modular 1 mm^3 Die-Stacked Sensing Platform With Low Power I^2C Inter-Die Communication and Multi-Modal Energy Harvesting," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 229–243, Jan. 2013.

[8] T. D. Burd, T. A. Pering, A. J. Stratakos, and R. W. Brodersen, "A Dynamic Voltage Scaled Microprocessor System," *IEEE J. Solid-State Circuits*, vol. 35, no. 11, pp. 1571–1580, Nov. 2000.

[9] J. Howard, S. Dighe, Y. Hoskote, S. Vangal, D. Finan, G. Ruhl, D. Jenkins, H. Wilson, N. Borkar, G. Schrom, F. Pailet, S. Jain, T. Jacob, S. Yada, S. Marella, P. Salihundam, V. Erraguntla, M. Konow, M. Riepen, G. Droege, J. Lindemann, M. Gries, T. Apel, K. Henriss, T. Lund-Larsen, S. Steibl, S. Borkar, V. De, R. Van Der Wijngaart, and T. Mattson, "A 48-Core IA-32 Message-Passing Processor with DVFS in 45nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2010, pp. 108–109.

[10] E. J. Fluhr, J. Friedrich, D. Dreps, V. Zyuban, G. Still, C. Gonzalez, A. Hall, D. Hogenmiller, F. Malgioglio, R. Nett, J. Paredes, J. Pille, D. Plass, R. Puri, P. Restle, D. Shan, K. Stawiasz, Z. T. Deniz, D. Wendel, and M. Ziegler, "POWER8™: A 12-Core Server-Class Processor in 22nm SOI with 7.6Tb/s Off-Chip Bandwidth," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2014, pp. 96–97.

[11] M. Putic, L. Di, B. H. Calhoun, and J. Lach, "Panoptic DVS: A Fine-Grained Dynamic Voltage Scaling Framework for Energy Scalable CMOS Design," in *Proc. IEEE Int. Conf. Computer Design*, Oct. 2009, pp. 491–47.

[12] A. Muramatsu, T. Yasufuku, M. Nomura, M. Takamiya, H. Shinohara, and T. Sakurai, "12% Power Reduction by Within-Functional-Block Fine-Grained Adapative Dual Supply Voltage Control in Logic Circuits with 42 Voltage Domains," in *Proc. European Solid-State Circuits Conf.*, Sep. 2011, pp.191–194.

[13] A. Wang and A. Chandrakasan, "A 180-mV Subthreshold FFT Processor Using a Minimum Energy Design Methodology," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, Jan. 2005.

[14] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-Threshold Computing: Reclaiming Moores Law through Energy Efficient Integrated Circuits," *Proc. IEEE*, vol. 98, no. 2, pp. 253–266, Feb. 2010.

[15] B. Zhai, R. G. Dreslinski, D. Blaauw, T. Mudge, and D. Sylvester, "Energy Efficient Near-threshold Chip Multi-processing," in *Proc. ACM/IEEE Int. Symp. Low Power Electronics and Design*, Aug. 2007, pp. 32–37.

[16] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw, "Performance and Variability Optimization Strategies in a Sub-200mV, 3.5pJ/inst, 11nW Subthreshold Processor," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2007, pp. 152–153.

[17] M. Seok, S. Hanson, Y. Lin, Z. Foo, D. Kim, Y. Lee, N. Liu, D. Sylvester, and D. Blaauw, "The Phoenix Processor: A 30pW Platform for Sensor Applications," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2008, pp. 188–189.

[18] W.-T. Wang, M.-D. Ker, M.-C. Chiang, C.-H. Chen, "Level Shifters for High-Speed 1V to 3.3V Interfaces in a 0.13 μm Cu-Interconnection/Low-k CMOS Technology," in *Proc. VLSI Technology, Systems, and Applications*, Apr. 2001, pp.307–310.

[19] H. Shao and C.-Y. Tsui, "A Robust, Input Voltage Adaptive and Low Energy Consumption Level Converter for Sub-threshold Logic," in *Proc. European Solid-State Circuits Conf.*, Sep. 2007, pp.312–315.

[20] I. J. Chang, J.-J. Kim, and K. Roy, "Robust Level Converter Design for Sub-threshold Logic," in *Proc. Int. Low Power Electronics and Design*, Oct. 2006, pp.14–19.

[21] I. J. Chang, J.-J. Kim, K. Kim, and K. Roy, "Robust Level Converter for Sub-Threshold/Super-Threshold Operation: 100mV to 2.5V," *IEEE Trans. Very Large Scale Integration Systems*, vol. 19, no. 8, pp.1429–1437, Aug. 2011.

[22] Y. Lin and D. Sylvester, "Single Stage Static Level Shifter Design for Subthreshold to I/O Voltage Conversion," in *Proc. ACM/IEEE Int. Symp. Low Power Electronics and Design*, Aug. 2008, pp. 197–200.

[23] H. Kaul, M. Anders, S. Hsu, A. Agarwal, R. Krishnamurthy, and S. Borkar, "Near-threshold Voltage (NTV) Design - Opportunities and Challenges," in *Proc. ACM/IEEE Design Automation Conference*, Jun. 2012, pp.1149–1154.

[24] L. Chang, D. Fried, J. Hergenrother, J. Sleight, R. Dennard, R. Montoye, L. Sekaric, S. McNab, A. Topol, C. Adams, K. Guarini, and W. Haensch, "Stable SRAM Cell Design for the 32nm Node and Beyond," in *Symp. VLSI Technology Dig. Tech. Papers*, Jun. 2005, pp. 128–129.

[25] L. Chang, R. Montoye, Y. Nakamura, K. Batson, R. Eickemeyer, R. Dennard, W. Haensch, and D. Jamsek, "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 956–963, Apr. 2008.

[26] J. Kulkarni, B. Geuskens, T. Karnik, M. Khellah, J. Tschanz, and V. De, "Capacitive-Coupling Wordline Boosting with Self-Induced V_{CC} Collapse for Write V_{MIN} Reduction in 22-nm 8T SRAM," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2012, pp. 234–235.

[27] T. Kim, J. Liu, and C. Kim, "A Voltage Scalable 0.26V, 64kb 8T SRAM with V_{MIN} Lowering Techniques and Deep Sleep Mode," *IEEE J. Solid-State Circuits*, vol. 44, no. 6, pp. 1785–1795, Jun. 2009.

[28] B. H. Calhoun, A. P. Chandrakasan, "A 256-kb 65-nm Sub-threshold SRAM Design for Ultra-Low-Voltage Operation," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 680–688, Mar. 2007.

[29] Y. Lee, D. Kim, J. Cai, I. Lauer, L. Chang, S. J. Koester, D. Blaauw, D. Sylvester, "Low-Power Circuit Analysis and Design Based on Heterojunction Tunneling Transistors (HETTs)," *IEEE Trans. Very Large Scale Integration Systems*, vol. 21, no. 9, pp. 1632–1643, Sep. 2013.

[30] M. Chang, M. Chen, L. Chen, S. Yang, Y. Kuo, J. Wu, H. Su, Y. Chu, W. Wu, T. Yang, and H. Yamauchi, "A Sub-0.3V Area-Efficient L-shaped 7T SRAM with Read Bitline Swing Expansion Schemes Based on Boosted Read-Bitline, Asymmetric-V_{TH} Read-Port, and Offset Cell V_{DD} Biasing Techniques," *IEEE J. Solid-State Circuits*, vol. 48, no. 10, pp. 2558–2569, Oct. 2013.

[31] D. F. Wendel, R. Kalla, J. Warnock, R. Cargnoni, S. G. Chu, J. G. Clabes, D. Dreps, D. Hrusecky, J. Friedrich, S. Islam, J. Kahle, J. Leenstra, G. Mittal, J. Paredes, J. Pille, P. J. Restle, B. Sinharoy, G. Smith, W. J. Starke, S. Taylor, J. Van Norstrand, S. Weitzel, P. G. Williams, and V. Zyuban, "POWER7$^{\text{TM}}$, a Highly Parallel, Scalable Multi-Core High End Server Processor," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 145–161, Jan. 2011.

[32] J. L. Shin, R. Golla, H. Li, S. Dash, Y. Choi, A. Smith, H. Sathianathan, M. Joshi, H. Park, M. Elgebaly, S. Turullols, S. Kim, R. Masleid, G. K. Konstadinidis, M. J. Doherty, G. Grohoski, and C. McAllister, "The Next Generation 64b SPARC Core in a T4 SoC Processor," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 82–90, Jan. 2013.

[33] M. Alioto, E. Consoli, and G. Palumbo, "Analysis and Comparison in the Energy-Delay-Area Domain of Nanometer CMOS Flip-Flops: Part I — Methodology and Design Strategies," *IEEE Trans. Very Large Scale Integration Systems*, vol. 19, no. 5, pp. 725–736, May 2011.

[34] M. Alioto, E. Consoli, and G. Palumbo, "Analysis and Comparison in the Energy-Delay-Area Domain of Nanometer CMOS Flip-Flops: Part II — Results and Figures of Merit," *IEEE Trans. Very Large Scale Integration Systems*, vol. 19, no. 5, pp. 737–750, May 2011.

[35] C. K. Teh, T. Fujita, H. Hara, and M. Hamada, "A 77% Energy-Saving 22-Transistor Single-Phase Clocking D-Flip-Flop with Adaptive-Coupling Configuration in 40nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2011, pp. 338–339.

[36] S. D. Naffziger, G. Colon-Bonet, T. Fischer, R. Riedlinger, T. J. Sullivan, and T. Grutkowski, "The Implementation of the Itanium 2 Microprocessor," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1448–1460, 2002.

[37] J. Yuan and C. Svensson, "High-Speed CMOS Circuit Technique," *IEEE J. Solid-State Circuits*, vol. 24, no. 1, pp. 62–70, 1989.

[38] C.-H. Chen, K. Bowman, C. Augustine, Z. Zhang, and J. Tschanz, "Minimum Supply Voltage for Sequential Logic Circuits in a 22nm Technology," in *Proc. ACM/IEEE Int. Symp. Low Power Electronics and Design*, Sep. 2013, pp. 181–186.

[39] H. Kaul, M. A. Anders, S. K. Mathew, S. K. Hsu, A. Agarwal, R. K. Krishnamurthy, and S. Borkar, "A 300mV 494GOPS/W Reconfigurable Dual-Supply 4-Way SIMD Vector Processing Accelerator in 45nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2009, pp. 260–261.

[40] M. Seok, S. Hanson, J. Seo, D. Sylvester, and D. Blaauw, "Robust Ultra-Low Voltage ROM Design," in *Proc. IEEE Custom Integrated Circuits Conference*, Sep. 2008, pp. 423–426.

[41] S. Dighe, S. Gupta, V. De, S. Vangal, N. Borkar, S. Borkar, and K. Roy, "A 45nm 48-Core IA Processor with Variation-Aware Scheduling and Optimal Core Mapping," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2011, pp. 250–251.

[42] Y. Lee, B. Giridhar, Z. Foo, D. Sylvester, and D. Blaauw, "A 660pW Muti-Stage Temperature-Compensated Timer for Ultra-Low-Power Wireless Sensor Node Synchronization," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2011, pp. 46–47.

[43] B. Calhoun and A. Chandrakasan, "A 256kb Sub-threshold SRAM in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2006, pp. 2592–2593.

[44] M. Chang, J. Wu, K. Chen, Y. Chen, Y. Chen, R. Lee, H. Liao, and H. Yamauchi, "A Differential Data-Aware Power-Supplied (D^2AP) 8T SRAM Cell with Expanded Write/Read Stabilities for Lower VDDmin Applications," *IEEE J. Solid-State Circuits*, vol. 45, no. 6, pp. 1234–1245, Jun. 2010.

[45] S. Jain, S. Khare, S. Yada, V. Ambili, P. Salihundam, S. Ramani, S. Muthukumar, M. Srinivasan, A. Kumar, S. K. Gb, R. Ramanarayanan, V. Erraguntla, J. Howard, S. Vangal, S. Dighe, G. Ruhl, P. Aseron, H. Wilson, N. Borkar, V. De, and S. Borkar, "A 280mV-to-1.2V Wide-Operating-Range IA-32 Processor in 32nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2012, pp. 66–67.

[46] B. Giridhar, M. Fojtik, D. Fick, D. Sylvester, and D. Blaauw, "Pulse Amplification Based Dynamic Synchronizers with Metastability Measurement Using Capacitance De-rating," in *Proc. IEEE Custom Integrated Circuits Conference*, Sep. 2013.

[47] D. Fick, N. Liu, Z. Foo, M. Fojtik, J. Seo, D. Sylvester, and D. Blaauw, "In Situ Delay-Slack Monitor for High-Performance Processors Using an All-Digital Self-Calibrating 5ps Resolution Time-to-Digital Converter," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2010, pp. 188–189.

[48] E. Consoli, M. Alioto, G. Palumbo, and J. Rabaey, "Conditional Push-Pull Pulsed Latches with 726fJ·ps Energy-Delay Product in 65nm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2012, pp. 482–483.

[49] S. Nomura, F. Tachibana, T. Fujita, C. K. Teh, H. Usui, F. Yamane, Y. Miyamoto, C. Kumtornkittikul, H. Hara, T. Yamashita, J. Tanabe, M. Uchiyama, Y. Tsuboi, T. Miyamori, T. Kitahara, H. Sato, Y. Homma, S. Matsumoto, K. Seki, Y. Watanabe, M. Hamada, and M. Takahashi, "A 9.7mW AAC-Decoding, 620mW H.264 720p 60fps Decoding, 8-Core Media Processor with Embedded Forward-Body-Biasing and Power-Gating Circuit in 65nm CMOS Technology," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2008, pp. 262–263.

[50] Y. Ueda, H. Yamauchi, M. Mukuno, S. Furuichi, M. Fujisawa, F. Qiao, and H. Yang, "6.33mW MPEG Audio Decoding on a Multimedia Processor," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2006, pp. 1636–1637.

[51] B.-S. Kong, S.-S. Kim, and Y.-H. Jun, "Conditional-Capture Flip-Flop for Statistical Power Reduction," *IEEE J. Solid-State Circuits*, vol. 36, no. 8, pp. 1263–1271, Aug. 2001.

[52] H. Partovi, R. Burd, U. Salim, F. Weber, L. DiGregorio, and D. Draper, "Flow-through Latch and Edge-Triggered Flip-Flop Hybrid Elements," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 1996, pp. 138–139.

[53] N. Nedovic, W. W. Walker, and V. G. Oklobdzija, "A Test Circuit for Measurement of Clocked Storage Element Characteristics," *IEEE J. Solid-State Circuits*, vol. 39, no. 8, pp. 1294–1304, Aug. 2004.

[54] M. Yabuuchi, K. Nii, Y. Tsukamoto, S. Ohbayachi, Y. Nakase, and H. Shinohara, "A 45nm 0.6V Cross-Point 8T SRAM with Negative Biased Read/Write Assist," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2009, pp. 158–159.

[55] M. Sinangil, H. Mair, and A. P. Chandrakasan, "A 28nm High-Density 6T SRAM with Optimized Peripheral-Assist Circuits for Operation Down to 0.6V," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2011, pp. 260–262.

[56] A. Bhavnagarwala, S. Kosonocky, C. Radens, Y. Chan, K. Stawiasz, U. Srinivasan, S. P. Kowalczyk, and M. M. Ziegler, "A sub-600mV, Fluctuation Tolerant 65nm CMOS SRAM Array with Dynamic Cell Biasing," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2007, pp. 78–79.

[57] H. Pilo, I. Arsovski, K. Batson, G. Braceras, J. Gabric, R. Houle, S. Lamphier, C. Radens, and A. Seferagic, "A 64Mb SRAM in 32nm High-k Metal-Gate SOI Technology with 0.7V Operation Enabled by Stability, Write-Ability and Read-Ability Enhancements," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 97–106, Jan. 2012.

[58] B. Zimmer, S. O. Toh, H. Vo, Y. Lee, O. Thomas, K. Asanovic, and B. Nikolic, "SRAM Assist Techniques for Operation in a Wide Voltage Range in 28nm CMOS," *IEEE Trans. Circuits and Systems – II: Express Briefs*, vol. 59, no. 12, pp. 853–857, Dec. 2012.

[59] E. Karl, Y. Wang, Y.-G. Ng, Z. Guo, F. Hamzaoglu, U. Bhattacharya, K. Zhang, K. Mistry, and M. Bohr, "A 4.6GHz 162Mb SRAM Design in 22nm Tri-Gate CMOS Technology with Integrated Active V_{MIN}-Enhancing Assist Circuitry," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2012, pp. 230–231.

[60] A. D. Dehennis, M. Mailand, D. Grice, S. Getzlaff, and A. E. Colvin, "A Near-Field-Communication (NFC) Enabled Wireless Fluorimeter for Fully Implantable Biosensing Applications," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2013, pp. 298–299.

[61] S. Tankiewics, J. Schaefer, and A. Dehennis, "A Co-Planar, Near Field Communication Telemetry Link for a Fully-Implantable Glucose Sensor Using High Permeability Ferrites," in *Proc. IEEE Sensors*, Nov. 2013.

[62] D. Jeon, Y.-P. Chen, Y. Lee, Y. Kim, Z. Foo, G. Kruger, H. Oral, O. Berenfeld, Z. Zhang, D. Blaauw, and D. Sylvester, "An Implantable 64nW ECG-Monitoring Mixed-Signal SoC for Arrhythmia Diagnosis," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2014, pp. 416–417.

[63] S.-Y. Hsu, Y. Ho, Y. Tseng, T.-Y. Lin, P.-Y. Chang, J.-W. Lee, J.-H. Hsiao, S.-M. Chuang, T.-Z. Yang, P.-C. Liu, T.-F. Yang, R.-J. Chen, C. Su, and C.-Y. Lee, "A Sub-100μW Multi-Functional Cardiac Signal Processor for Mobile Healthcare Applications," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2012, pp. 156–157.

[64] S. Kim, L. Yan, S. Mitra, M. Osawa, Y. Harada, K. Tamiya, C. van Hoof, and R. F. Yazicioglu, "A 20μW Intra-Cardiac Signal-Processing IC with 82dB Bio-Impedance Measurement Dynamic Range and Analog Feature Extraction for Ventricular Fibrillation Detection," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2013, pp. 302–303.