ANALYSIS AND DESIGN OF LOW-ENERGY CLOCKED STORAGE ELEMENTS

by

Dejan Markovic

Memorandum No. UCB/ERL M00/64

11 December 2000

ANALYSIS AND DESIGN OF LOW-ENERGY CLOCKED STORAGE ELEMENTS

by

Dejan Markovic

Memorandum No. UCB/ERL M00/64

11 December 2000

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

Analysis and Design of Low-Energy Clocked Storage Elements

by

Dejan Marković

Master of Science in Electrical Engineering
and Computer Sciences

University of California at Berkeley

Proliferation of portable computing has promoted low energy consumption into the main requirement for successful digital system design. A key to minimize the energy consumption of any IC is a firm understanding of the tradeoffs related to the timing and energy specifications at both the circuit and system levels. The goal of this project is to develop a methodology for an aggressive energy reduction of the clocking subsystem and its constituent parts—memory elements as well as general clock distribution network. To achieve a minimum energy solution of the clocking subsystem for a given performance constraint, there is a systematic approach which involves: 1) proper selection and optimization of memory elements, 2) determination of interface parameters with respect to clock network and combinational logic designs, and 3) clock tree optimization and implementation based on these interface parameters. A new set synthetic metrics, providing full insight into the system level behavior of these crucial memory elements, is introduced and discussed. An exemplary synthetic metric is the *delay*, which takes into account the flip-flop clock-to-output delay and setup time, quantifying the flip-flop overall latency with respect to the system clock period. Other equally important synthetic metrics quantify the tolerance flip-flops have with regard to race conditions and provide better intuition into flip-flop energy consumption. Representative flip-flop topologies are analyzed and compared. More advanced issues such as glitching energy consumption and flip-flop timing and energy parameters under supply voltage scaling for various clock slopes are presented.

Table of Contents

List of Figures

List of Tables

Acknowledgments

As with any endeavor or hardship, rarely does one ever succeed alone. Gotham City is safe only because Batman has Robin at his side. Tom will never be as funny without Jerry, and Roadrunner will never sprint his fastest without Wile E. Coyote at his tail. This thesis is no exception to the rule. This work is an extensive collaboration between many of my colleagues at the Berkeley Wireless Research Center and myself. Many I must thank, and many more I will unfortunately forget to mention (Please forgive me if I leave you out!)

First and foremost, I would like to thank my advisors, Bob Brodersen and Bora Nikolić for all the personal and technical help they have offered and given me. I am extremely fortunate that Bob, the "big boss," was with me for the entire duration of the project with his guidance, patience, and invaluable advice. His constant drive for his students to succeed is truly exceptional, and if it were not for him, I would not be typing this now. Bora's, (the "little boss") understanding, technical support, and friendship over the last couple years beginning with my undergraduate has greatly transcended my knowledge about low-energy design and flip-flops. His initial inquisition about energy consumption in memory elements was enough to jump-start my work. I am indebted to both of them for their faith in my research and their tolerance of my weaknesses. I am stronger each day because of that.

I am thankful to all my fellow students from the BJgroup, especially my cubicle mates and good friends Fred Chen, Andy Klein and Johan Vanderhaegen who have collectively made my work environment much more pleasant. I must also thank Rhett "the button" Davis for mustering the patience to teach me how to use EVERY IC design tool. The SSHAFT group (Can you dig it?) also gave me reason to keep the ball rolling on my research project. Nathan Chan helped correct my English in the thesis. All the errors (or is it "all errors"?) remaining in the text are my fault. A special shout also goes out to all of my colleagues and friends in Cory Hall.

This work has been supported under the DARPA F30602-97-2-0346 grant and I gratefully acknowledge their generous support.

Finally, and most importantly, I must express gratitude to my parents and the rest of my family for everything they have done for me. They have been the greatest source of inspiration to me, and this work is a product of their endless love and continued support. I would have never imagined getting this far without my family's faith and confidence, and to them I dedicate this work.

Chapter 1

INTRODUCTION

The dominant source of energy consumption in digital VLSI systems is the switching component dissipated in nodes with high switching probabilities. The nodes of this kind are clocked nodes, so reducing energy consumed by the clocking subsystem can lead to a significant reduction in the overall system energy consumption. In a typical digital IC, about 30%-60% of the total energy is utilized by clocks. The clocking subsystem, as defined here, contains: 1) the clock distribution network, and 2) memory elements.

As technology scales down, energy breakdown between registers and the clock distribution network change because extra energy related to switching wire capacitances become increasingly important, as does the strategy in routing clock signals. Figure 1.1 shows an example of the energy breakdown in a chip implementing a decimation filter in 0.25μm CMOS.

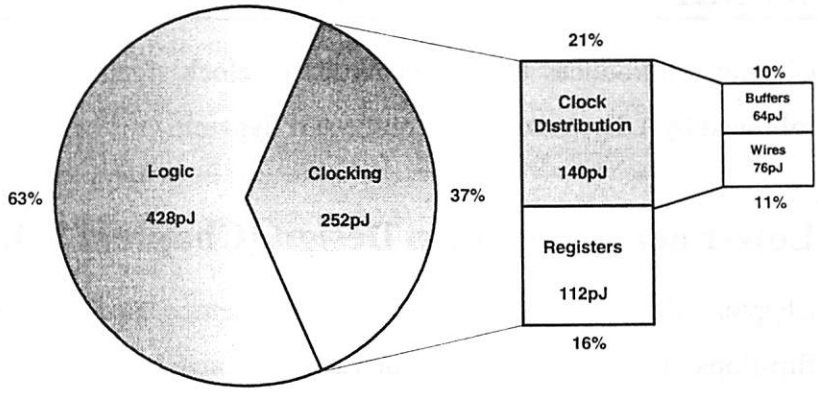


FIGURE 1.1 Breakdown of average energy consumption per clock cycle in decimation filter, containing 3360 2-phase FFs, implemented in 0.25μm. V_{DD} = 1V, f = 25MHz.

The chip has a single clock domain, driven by one master buffer. A two-phase non-overlapping clocking scheme is utilized which avoids timing problems associated with short paths due to a relaxed clock skew requirement. This also allows more freedom in the routing of clock signals. The relaxed timing requirements that come with a 2-phase clocking scheme incur energy overhead due to the global routing of an extra clock signal. As a result, there exists a tradeoff between timing robustness and energy efficiency. Besides that, there is also a tradeoff between circuit performance and energy efficiency. What one would like to know is:

For a given performance constraint, what is a minimum energy required by the clocking subsystem?

The goal of this project is to suggest an optimal strategy for the selection and placement of memory elements for minimum energy and maximum performance. A minimum energy solution of the clocking subsystem must also be robust against timing failures. In order to achieve this optimal strategy, one needs to understand the interaction between memory elements and the clock distribution network as well as provide the clock tree designer with relevant interface parameters.

1.1 OVERIVIEW OF THESIS

Chapter 2 introduces timing properties of clock signals and memory elements, followed by a discussion about static and dynamic latches and flip-flops.

1.1.1 Low-Energy Flip-Flop Design [Chapters 3,4,5]

Chapter 3 describes the timing and energy metrics used in the characterization of flip-flops. Timing parameters such as setup and hold times, delay and internal race immunity are introduced and discussed. In addition, energy-per-transition is introduced as a metric to evaluate energy consumption in flip-flops. This metric is used to compute energy dissipated in clocked and internal nodes in flip-flops as

well as energy expended to drive an external output load. Analyzing energy-per-transition therefore provides a basis for the exploration of energy saving capabilities of flip-flops with internal clock gating.

Chapter 4 discusses the mechanisms of energy reduction in flip-flops, based on the expression for switching component of energy. These mechanisms include: 1) minimizing switched capacitance, 2) supply voltage scaling, 3) low swing circuit techniques, and 4) minimizing switching activity. Sizing issues in flip-flop design will be described.

Low energy flip-flop design is presented in Chapter 5. It includes design and comparative analysis of various flip-flop techniques: master-slave, pulse-triggered, reduced clock swing, and flip-flops with internal clock gating.

1.1.2 Glitching Energy in Flip-Flops [Chapter 6]

The focus of Chapter 6 is on the analysis of the flip-flop glitch sensitivity. The glitching component of energy in various flip-flop topologies is explored and comparative results are presented and discussed.

1.1.3 Interfacing Issues Under Supply Voltage Scaling [Chapter 7]

Chapter 7 discusses interface parameters between flip-flops, clock network, and combinational logic. Impact of clock slope on the flip-flop internal race immunity and the system clock skew specification is elaborated. Energy consumption in clock buffers at various clock signal slopes and supply voltages is analyzed. Chapter 7 also provides a study of timing and energy parameters of flip-flops and combinational logic under supply voltage scaling conditions.

Chapter 2

CLOCK SUBSYSTEM DESIGN

Our approach to optimize the clocking subsystem consists of three steps, illustrated in the flow chart of Fig. 2.1. The first and most important step is the optimization of memory elements, which involves the selection of energy efficient flip-flop topologies. The characterization yields interface parameters relevant to clock network design. The third step is to optimize the clock distribution network based on the interface parameters revealed by the former step.

Prior to the selection and optimization of memory elements, there needs to be a firm understanding of the basic issues involved in the timing of these elements.

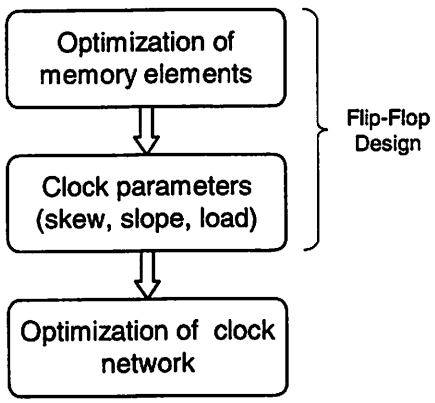


FIGURE 2.1 Clock subsystem design methodology.

2.1 TIMING

Figure 2.2 illustrates the terms used to describe signal transitions and delays between transitions. The terms associated with a single transition are measured when the signal has reached the critical points 10%, 50% and 90% of its final value. The relative timing (delay) between two transitions is measured from the 50% point of one transition to the 50% point of the other transition.

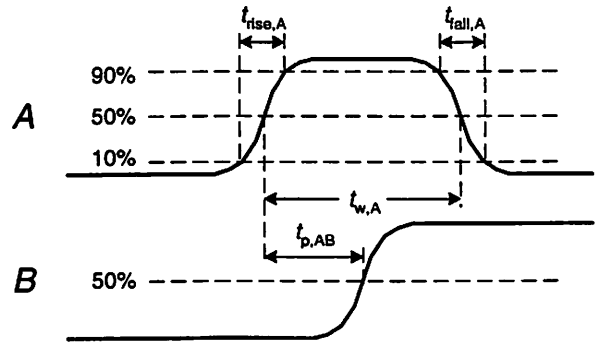


FIGURE 2.2 Timing fundamentals: rise time, fall time, pulse width, delay.

For example, the *rise time* of the signal A, $t_{rise,A}$ is the time from the start (10%) to the end (90%) of a rising transition. The *fall time* of the signal A, $t_{fall,A}$ is the time from the start (90%) to the end (10%) of a falling transition.

The *pulse width* of the signal A, $t_{w,A}$ is the time from one transition on A to the next transition on A. The pulse width often refers to the duration when the clock pulse is at its high value.

2.1.1 Timing Properties of Periodic Signals

Signal A is periodic of degree n if

$$T_{CY,An}(i) = t_{A(i+n)} - t_{Ai} \tag{2.1}$$

the delay between the i-th transition and the $(i+n)$-th transition of a signal A, is a constant for all i. A periodic signal of degree 1 is fully characterized by its transition period, $T_{CY,A1}$, or its transition frequency, f_{A1} as:

$$f_{A1} = \frac{1}{T_{CY,A1}} = \frac{1}{t_{A(i+1)} - t_{Ai}} \tag{2.2}$$

The *duty factor* of a periodic signal of degree 2 is the fraction of time the signal is high relative to its period:

$$d_A = \frac{t_{w,A}}{T_{CY,A}} \tag{2.3}$$

The periodic signals of degree 2 can be used to describe clocks which, for the purpose of this thesis, are the primary signals of interest.

2.1.1.1 Timing Uncertainties of Clock Signals

Different sources of clock uncertainties are described by jitter and clock skew.

Jitter

The jitter of a clock signal quantifies the *temporal uncertainty* between two *consecutive edges* and it is equal to the delay between the time of the expected transition of a signal and the time of its actual transition as depicted in Fig. 2.3. Jitter is a zero-mean random variable.

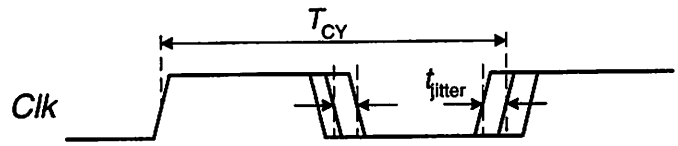


FIGURE 2.3 Jitter.

Skew

The skew between two clock signals describes the *spatial uncertainty* between two *temporally equivalent edges* of the clock signals, equal to the delay between the two signals, as shown in Fig. 2.4. Both jitter and skew affect the effective cycle time of a clock signal.

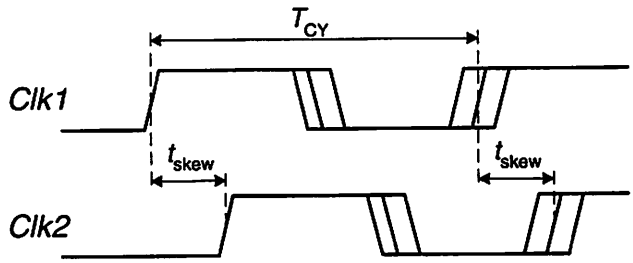


FIGURE 2.4 Skew.

2.2 FLIP-FLOPS VS. LATCHES

The memory elements can be implemented using either latches or flip-flops. The choice between latches and flip-flops is dependent upon the timing requirements of a digital system. Figure 2.5 shows symbol representations and operational differences between a latch and a flip-flop.

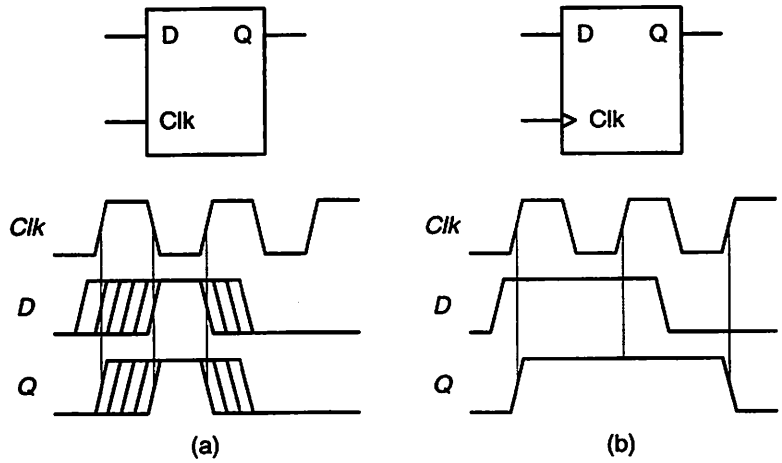


FIGURE 2.5 Symbol and timing properties of (a) latch, (b) flip-flop.

2.2.1 Level-Sensitive Latch

The level sensitive latch of Fig. 2.5.a is a clocked storage element that logically connects its input to its output when the clock is "high" and holds the output value stable when the clock is "low". In Fig. 2.5.a, the output Q changes according to input D when the clock *Clk* is "high", which means that unwanted transitions (glitches) at the input would also appear at the output. This improves high-performance designs, but introduces difficulty in timing verification.

2.2.2 Edge-Triggered Flip-Flop

The edge-triggered flip-flop, Fig. 2.5.b is a clocked storage element that logically connects its input to its output at the rising edge of the clock signal, and holds the output value stable until next rising clock edge. This way, the unwanted transitions of input D will not propagate to the output Q even if the clock is high. Therefore, flip-flops are preferred over latches for easy timing analysis and verification.

2.3 FLIP-FLOPS

Flip-flops are frequently built from latches. The main property of flip-flops is that they are *edge-triggered* storage elements. There are two main flip-flop styles: master-slave latch pairs, and pulse-triggered latches, as illustrated in Fig. 2.6.

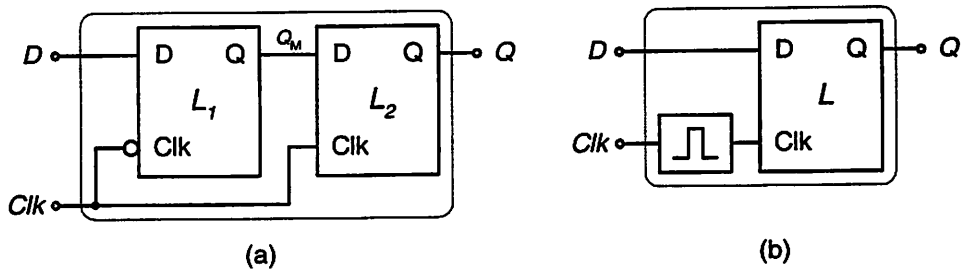


FIGURE 2.6 Flip-flop techniques: (a) master-slave latch pair, (b) pulse-triggered latch.

For the flip-flop in Fig. 2.6.b, the term "edge-triggered" stands for the clocking mechanism of the entire flip-flop, relative to the external clock, *Clk*. The term "pulse-triggered" describes the generation of internal pulse, which clocks the latch L used to store the data. In the master-slave latch pair in Fig. 2.6.a, the clocking of internal latches L_1 and L_2 is performed differently from the pulse-triggered latch.

2.3.1 Timing Operation of Master-Slave Latch-Pairs

A flip-flop of this style is essentially a latch pair, one sensitive to clock-high, and another sensitive to clock-low. Sometimes latches transparent during clock-high are called n-type, while latches transparent during clock-low are called p-type, [Sven98]. If a p-type latch follows an n-type latch, the resulting flip-flop is triggered by the rising edge of clock signal, as shown in Fig. 2.7.

Traditionally, the input latch is called the *master*, while the output latch is called the *slave*. The top three waveforms in Fig. 2.7 illustrate the timing operation of the master latch, transparent for clock-low. When the clock is "high," the slave latch is transparent and the master latch is opaque, which ensures that no glitch on D can propagate to the output Q.

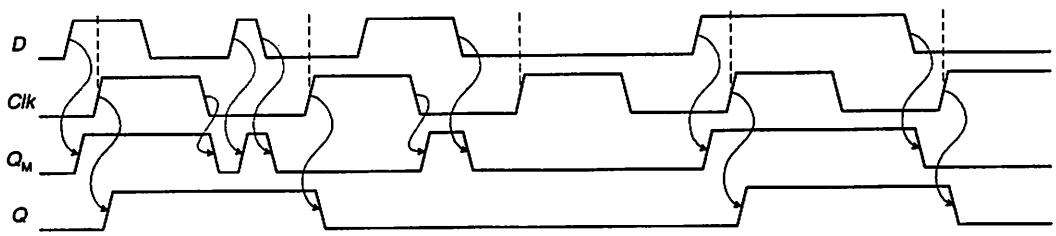


FIGURE 2.7 Timing diagrams of flip-flops (master-slave).

2.3.2 Timing Operation of Pulse-Triggered Latches

A pulse-triggered latch is also a two-stage flip-flop where the first stage is a pulse generator, and the second stage is a latch as shown in Fig. 2.6.b. If a *p*-type latch is used, the flip-flop is triggered by a rising edge of the clock. The pulse generator basically differentiates external clock signal, *Clk*, and generates an internal clock, Clk_{int}, which is used to clock the latch, *L*. The flip-flop actually operates as a latch, which means that it can still "catch" glitches from the input. However, shortened duty cycle of Clk_{int} reduces the probability of such undesired operation.

Clk_{int} is not only shortened but also a *delayed* version of the external clock, *Clk*, so the *Clk-Q* delay of the pulse-triggered latch will be longer by the delay of the pulse generator as opposed to a latch driven directly by *Clk*.

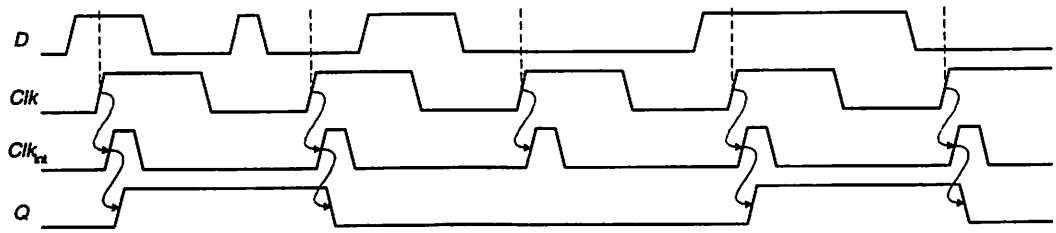


FIGURE 2.8 Timing diagrams of flip-flops (pulse-triggered).

2.4 LATCHES

The different latches that exist can strongly affect and dictate the properties of the flip-flops that contain them. It is therefore imperative to examine the different designs before that can be used to implement flip-flops.

2.4.1 Dynamic Latches

The simplest dynamic latch is shown in Fig. 2.9.a. It consists of a transmission gate cascaded by an inverter. When the clock, *CP*, is "high," transmission gate conducts and the storage node, *S*, follows the input *D* with a small RC delay. The inverter follows the signal on the storage node to generate the inverted output, $\bar{Q}$,

and isolates the storage node from the output node. When *CP* is "low," the transmission gate is *off* and the voltage on *S* remains constant. The value of the storage node is physically stored as charge on the parasitic capacitance C_S, which consists of the drain capacitance of the transmission gate, the parasitic wiring capacitance, and the gate capacitance of the inverter.

If the storage node is not stable at a digital level and is in the high-gain region of the inverter, large amounts of current will be drawn and any noise coupled into *S* will appear amplified on $\bar{Q}$. For this reason, *D* has to have stable digital level at least one setup time before and one hold time after the falling edge of *CP*. To charge *S* to a reliable "1" before *CP* falls, *D* must rise at least a setup time before *CP* falls. The setup time is usually a small multiple of the RC time constant set by the resistance of the transmission gate and the capacitance of the storage node. To prevent *S* from discharging until the transmission gate is completely *off*, *D* is not allowed to fall until after the hold time, after *CP* falls. The transmission gate is fully *off* once *CP* falls below V_{Tn}, so the hold time is largely determined by the fall time of the clock signal. It is not unusual for dynamic latches to have zero or even negative hold times as a result.

The storage node signal in a dynamic latch is by default dynamic in nature and must be periodically refreshed. If *CP* is "low" for a long period, charge will

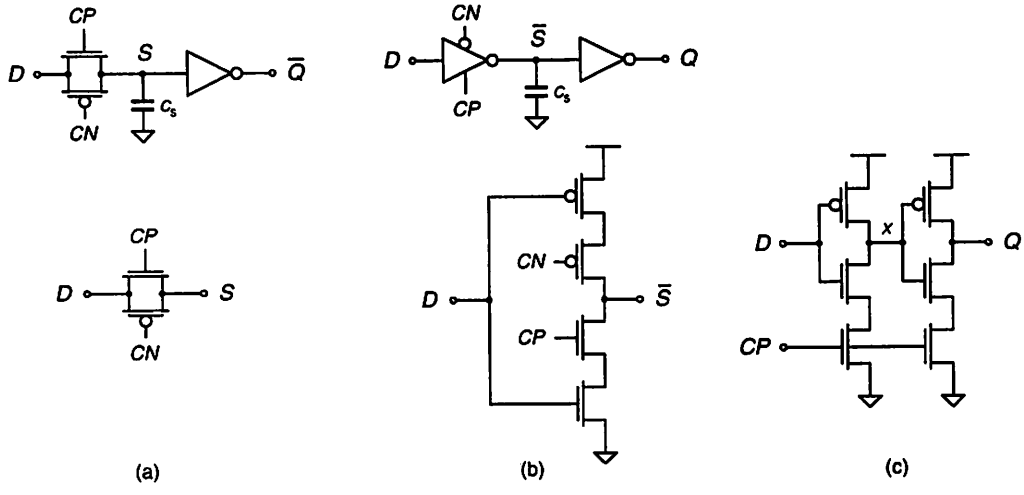


FIGURE 2.9 Dynamic latch circuits: (a) transmission gate latch, (b) clocked CMOS (C^2MOS) latch, (c) single-phase latch.

leak off of C_S, degrading the signal on *S*. There is a certain point at which the storage signal is degraded below the allowable noise margin and must be refreshed by raising *CP*. The primary mechanism for charge leakage is subthreshold conduction through the transmission gate.

A gate-isolated latch shown in Fig. 2.9.b is used whenever it is desirable to isolate the input node, *D*, from charge injection. When the simple dynamic latch switches on, the storage capacitor is connected to the input node. If the two nodes, *S* and *D*, are at different values, the resulting charge sharing disturbs node *D* by an amount dependent on the capacitance ratio.

Dynamic charge sharing is eliminated entirely by using a tri-state inverter for the input stage of the dynamic latch. This inverter operates like a normal inverter when *CP* is "high" and isolates its output when *CP* is "low." Thus, when *CP* rises, this inverter drives the storage node with no charge injection back to the input. The result is identical to adding an inverter to the input similar to Fig. 2.9.a for isolation.

A single-phase latch [Yuan89] provides gate isolation and requires only a single polarity of the clock signal *CP*. When *CP* is "high," the circuit simply acts as two back-to-back inverters, passing *D* to *Q*. When *CP* is "low," the intermediate signal, *X*, is monotonically rising – it can no longer fall. Thus, the output is stable. It cannot fall because the series NMOS is *off*, and it cannot rise, for that would require that *X* fall to switch the PMOS *off*. In practice, keepers on *X* and *Q* are added to this circuit to enhance its noise immunity.

2.4.1 Static Latches

Static latches have some form of positive feedback that refreshes charge on storage nodes. Some representative static latches [Sven98] are shown in Fig. 2.10. Figure 2.10.a shows single-rail input, dual-rail output latch based on the classical set-reset latch. It has relatively low clocking power due to only four clocked transistors. Figure 2.10.b shows single-rail input, single-rail output latch based on the classical transmission gate latch. It uses six clocked transistors. Both

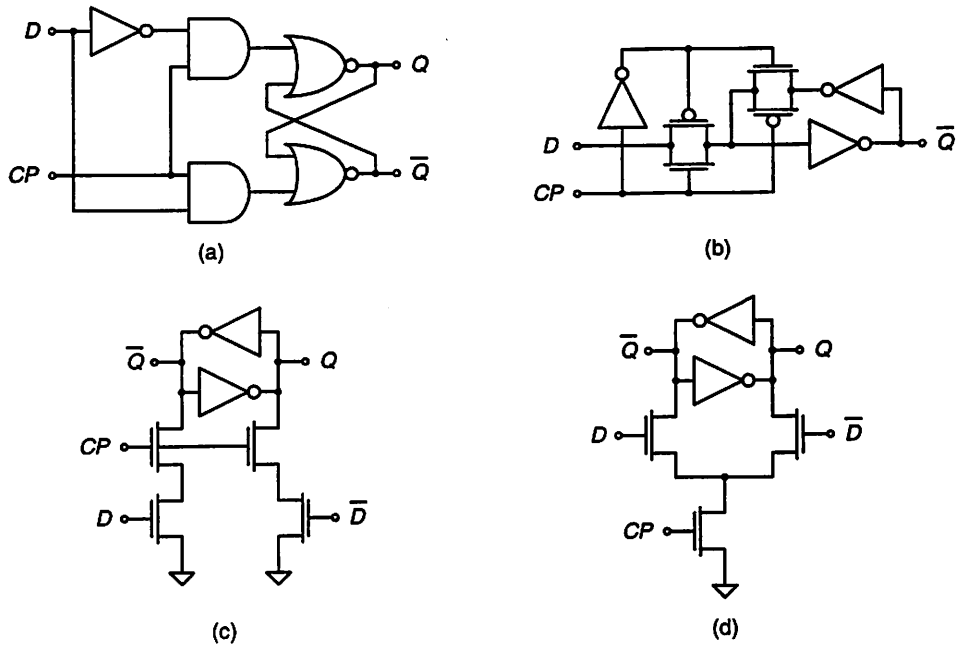


FIGURE 2.10 Static latch circuits: (a) 3-gate latch, (b) transmission gate latch, (c) *n*-RAM latch, (d) *n*-SSTC (static single-transistor clocked) latch. [Sven98]

of these latches are commonly used in standard CMOS cell libraries and are well established in the industry. RAM-type static latch, based on six-transistor SRAM memory cell is shown in Fig. 2.10.c. Figure 2.10.d shows static version of DSTC (dynamic single-transistor clocked) latch.

Static latches are more robust against noise than dynamic latches, so they are more suitable for sub-micrometer VLSI designs, when wire delays and signal cross-talk become important design issues. Static latches also consume less power than dynamic latches [Yuan97].

2.5 SUMMARY

Timing is a very important issue in system and sequential circuit design. The reason for this is that all realizable circuits have an inherent delay, meaning that the actual time of arrival for input signals may not be the same as the expected time in ideal circuits, which have outputs that instantaneously change with the input. Secondly, timing issues arise because the clock subsystem itself is subject to abnormalities such as clock skew and clock jitter. To combat these complex design issues, careful consideration must be given to the design of clock

dependent, or sequential circuits. The most important sequential circuits employed in VLSI systems today are memory elements.

Memory elements can be classified as being either a latch or a flip-flop, according to their triggering mechanism. A latch is an example of a level sensitive memory element. By level sensitive, we mean that the input signal is sampled when the clock is at a fixed logic level (either a "0" or "1"). A flip-flop on the other hand, is an example of an edge-triggered memory element that samples the input at either the rising or falling edge of the clock signal. It is important to note that edge-triggering behavior itself is a circuit "trick" employed using level sensitive latches. In a strict sense, every circuit employs level-sensitive devices to a certain degree. Designs today are largely composed of flip-flops as opposed to latches because flip-flops allow for easier timing analysis and verification than latches due to their inherent shorter sampling time. On the other hand, this gain is not without its disadvantages. Flip-flops are much more complex circuits and thus consume more energy than latches due to a larger area. However, the disadvantages are nonetheless outweighed by the excellent timing properties they possess.

There are two main flip-flop styles present in today's literature – master-slave latch pairs and pulse-triggered latches. Master-slave latch pairs are composed of two cascaded latches, each sensitive to opposite phases of the clock cycle. Pulse-triggered latches are built using a pulse generator cascaded by a latch. The pulse-generator generates an internal clock with a shortened duty cycle used to trigger the latch.

Latches used to build flip-flops can be static or dynamic in nature. In static logic designs, the output node is always tied to one of the supply rails, ensuring a solid logic level. Static flip-flop designs generally have better noise immunity, lower energy consumption, and better clock skew tolerance over their dynamic logic counterparts (Chapter 5). Dynamic logic involves a clock signal to incorporate two modes of operation to charge and evaluate the value of an output or intermediate node. Because of leakage currents and charge sharing, these

dynamic nodes must be periodically refreshed to restore original logic levels. Dynamic logic is often used in high speed, high performance circuits. However, the increased complexity in the design favors the use of static design. For this reason, low energy flip-flops are designed using static latches.

To conclude, it is extremely beneficial for the system designer to choose the correct topology for memory elements in order to satisfy the more stringent requirements of the clock subsystem design. The correct balance of low energy consumption and excellent timing properties allow more degrees of freedom for designers and system architects at the higher level of abstraction.

Chapter 3

FLIP-FLOP PERFORMANCE METRICS

To evaluate the performance of a flip-flop, it is essential to analyze its timing behavior and to measure its energy consumption. Section 3.1 introduces the basic timing parameters and metrics used to characterize a flip-flop. A study of the energy consumption in flip-flops, presented in Section 3.2, provides a solid foundation for understanding the system level issues related to minimizing the total energy consumed by the clocking subsystem. Based on these fundamental metrics, we derive synthetic timing parameters in Section 3.3, which will aid in the characterization of flip-flop performance with respect to the system clock. Interface parameters relevant for clock network design and combinational logic design are discussed in Section 3.4. Other commonly used metrics involved in the characterization of digital circuits such as *power* and *fanout-of-4 inverter* delay are presented to conclude the discussion.

3.1 TIMING METRICS

Data and *clock* inputs of a flip-flop have to satisfy basic timing restrictions to ensure correct operation of the flip-flop, as shown in Fig. 3.1. The flip-flop samples data from the input at the active clock edge and generates the appropriate output after the propagation delay, t_{pLH} if input undergoes a *0-1* transition between two consecutive clock edges, or t_{pHL} if input undergoes a *1-0* transition

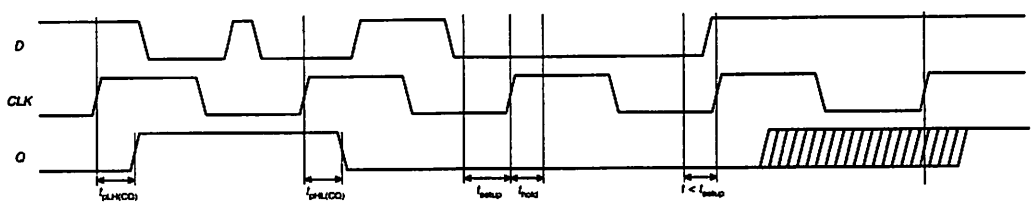


FIGURE 3.1 Basic timing diagrams of flip-flops.

between two consecutive clock edges, provided that there is no violation of timing constraints between data and clock inputs. Fundamental timing constraints between data and clock inputs are quantified with *setup* and *hold* times, as illustrated in Fig. 3.1. Setup and hold times define time intervals during which an input datum has to be stable to ensure correct flip-flop operation. Detailed definitions of setup and hold times are presented in the following two sections.

3.1.1 Setup Time

Setup time is frequently defined as the time interval before the active clock edge during which data is not allowed to change for correct operation of a flip-flop, [Hodges88], [Rabaey96], [Dally98]. This metric needs to be more precisely quantified because the setup time is characterized by an increase in clock-to-output delay of a flip-flop when data gets closer to clock as shown in Fig. 3.2.a and Fig. 3.3.

Alternative definition of setup time is as the data-to-clock offset which results in the minimal data-to-output delay. While it presents an essential measure of how fast a digital system can be clocked, it does not contain information about the variation of the flip-flop delay from its nominal value.

Frequently, ASIC standard cell libraries define setup time as specified increments in clock-to-output delay. The assumption made here is that an increase in the clock-to-output delay by 5% from its nominal value can be tolerated. The increase in the data-to-clock offset beyond the point, which is defined by the setup time results in a further increase in the clock-to-output delay

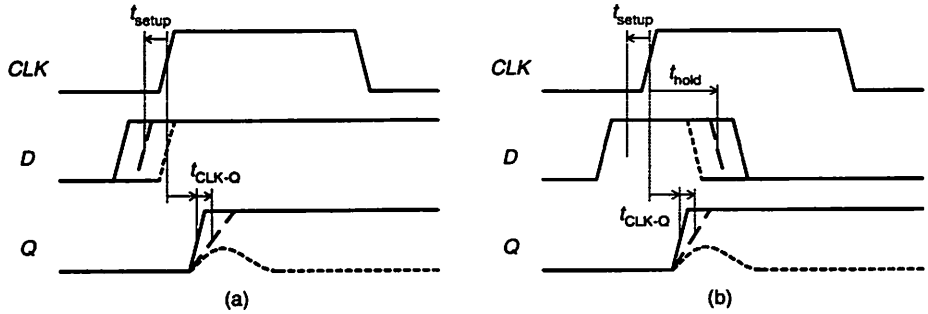


FIGURE 3.2 Degradation of clock-to-output delay when data gets closer to clock in: (a) setup region, (b) hold region.

as illustrated in Fig. 3.3. This can even result in circuit failure as shown in Fig. 3.2.a, when data arrives too late that the flip-flop is incapable of recording the input data transition.

> Setup time is the minimum data-to-clock offset that causes the clock-to-output delay to be 5% higher than its nominal value.
>
> 1) Setup of logic "1" is the setup time measured when data undergoes a *0-1* transition.
>
> 2) Setup of logic "0" is the setup time measured when data undergoes a *1-0* transition.

Setup time can be both positive (typical in master-slave latch pairs) and negative (typical in pulse-triggered latches) depending on the circuit topology, the supply voltage, and the simulation setup. In general, setup times of logic "1" and logic "0" are different. However, there are few cases when they are equal, for example, in fully balanced, fully differential flip-flops [Nikolic99].

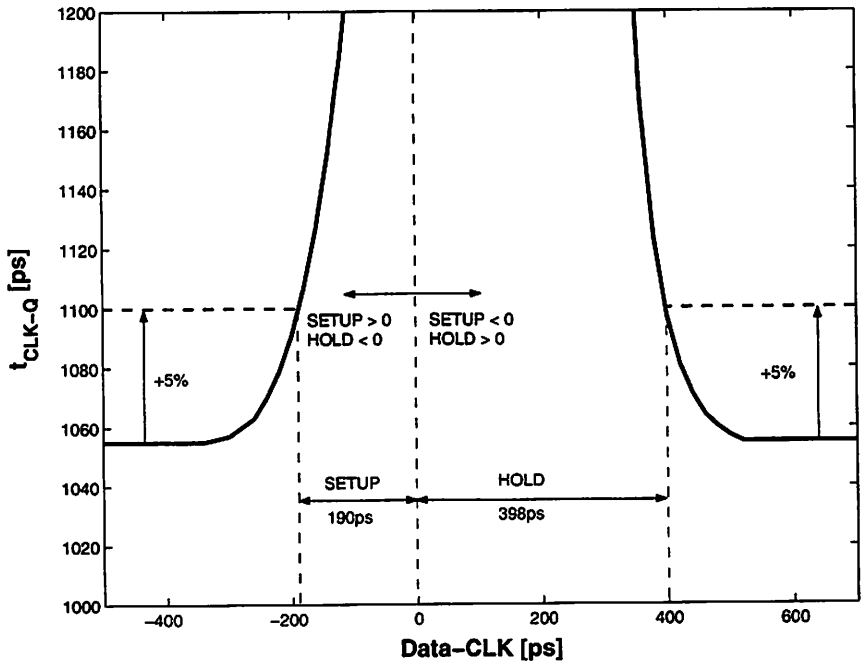


FIGURE 3.3 Definitions of setup and hold times.

3.1.2 Hold Time

If the new datum that arrived prior to the setup time changes back to its old value after the active clock edge, the clock-to-output delay of a flip-flop will be increased if this transition happens sooner after the active clock edge, as illustrated in Fig. 3.2.b and Fig. 3.3. As in the case of setup time, the definition of hold time is based on the tolerance to an increase in the clock-to-output delay.

> Hold time is the minimum time interval during which a new datum has to retain its value after the active clock edge so that clock-to-output delay is 5% higher than its nominal value.
>
> 1) Hold of logic "1" is the hold time measured when data undergoes a 0-1 transition
>
> 2) Hold of logic "0" is the hold time measured when data undergoes a 1-0 transition

If data changes too quickly, it can result in circuit malfunction, as shown in Fig. 3.2.b. Hold time can also be both negative (master-slave latch pairs) and positive (pulse-triggered latches). Like the setup time, hold time depends on the circuit topology, the supply voltage, and the simulation setup.

3.1.3 Clock-to-Output Delay

The clock-to-output delay is the *delay measured from the active clock edge to the output*. Besides the data-to-clock offset, as illustrated in Fig. 3.2 and Fig. 3.3, it also depends on the clock slope, the supply voltage, and the output load, which is shown later in Chapter 5.

Ideally it is desirable to have the delays through a flip-flop experienced by a rising or falling input transition to be similar. The delays cannot, in general, be identical because the two cases will experience a variety of rising and falling delays [Harris99]. This delay mismatch can sometimes be balanced with an adjusted output load, for example, in transmission gate flip-flop, as will be seen in Chapter 5.

3.1.4 Data-to-Output Delay

The data-to-output delay is the *delay measured from a 0-1 or 1-0 data transition to the output*, assuming that the flip-flop is clocked correctly. The data-to-output delay is not a good metric of a flip-flop performance because it depends on the arrival of new input datum relative to clock.

Simulation Setup

The simulation setup can greatly affect the setup and hold time for any given flip-flop. The test circuit shown in Fig. 3.4.a would result in a shorter setup time and longer clock-to-output delay compared to the setup time and clock-to-output delay obtained using the test circuit shown in Fig. 3.4.b. The mismatch between setup times and clock-to-output delays is equal to the delay of shaded inverter of Fig. 3.4.a. It is important to realize that the sum of the setup time and clock-to-output delay, $t_{setup}+t_{CLK\text{-}Q}$, is the same if the slope of the clock signal *CN* is equal in both cases. Similarly, the hold time obtained using setup from Fig. 3.4.a is longer, but the difference between clock-to-output delay and hold time, $t_{CLK\text{-}Q}$-t_{hold}, is the same in both cases.

Since most practical implementations have local clock buffering like that shown in Fig. 3.4.a, the setup of Fig. 3.4.a is used in simulations. As a result, setup and hold times might have an offset of one inverter delay, but this does not affect neither the flip-flop *delay* nor its inherent *race immunity*, which are the most important timing parameters of flip-flops as will be seen in Section 3.3.

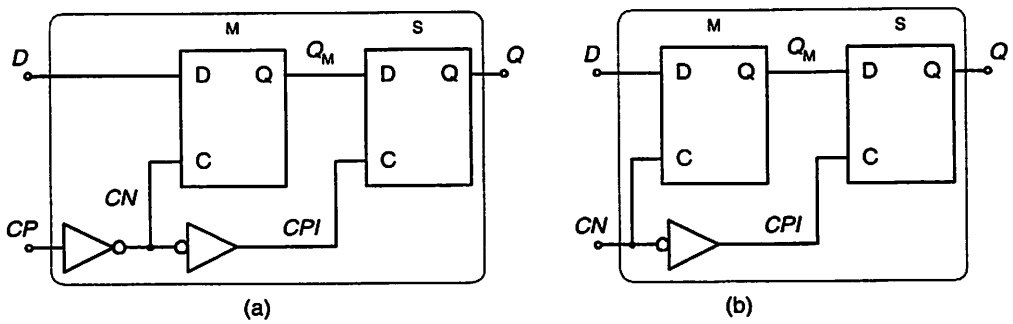


FIGURE 3.4 Comparison of simulation setups that result in different setup/hold times and different clock-to-output delays, but with the same $t_{CLK\text{-}Q}+t_{setup}$ and $t_{CLK\text{-}Q}$-t_{hold}.

3.1.5 Minimum Data Pulse Width

The minimum width of data pulse is the minimum time during which a new datum is required to be stable at a digital level to ensure correct operation of a flip-flop. It defines flip-flop sampling period and is ideally equal to *the sum of setup and hold times*. Illustration of the minimum pulse width over a range of supply voltages is shown in Fig. 3.5. This simplified definition neglects the fact that setup and hold times are defined with respect to the 50% point of the data waveform, so the minimum duration of a data pulse should be expressed as:

$$W_{\min} = t_{setup} + t_{hold} - t_{rise} \tag{3.1}$$

where $W_{\min}$ is specified from 10% or 90% point, since t_{rise} is specified between 10%-90% points, and the setup time is actually from $t_{\text{rise}}/2$ before the beginning of the $W_{\min}$. This metric is sometimes referred as the aperture time [Dally98].

A simple sum of the setup and hold times is more conservative, and the discrepancy from the exact definition is not significant unless one is designing the systems where the width of the aperture and uncertainty of delays are critical parameters.

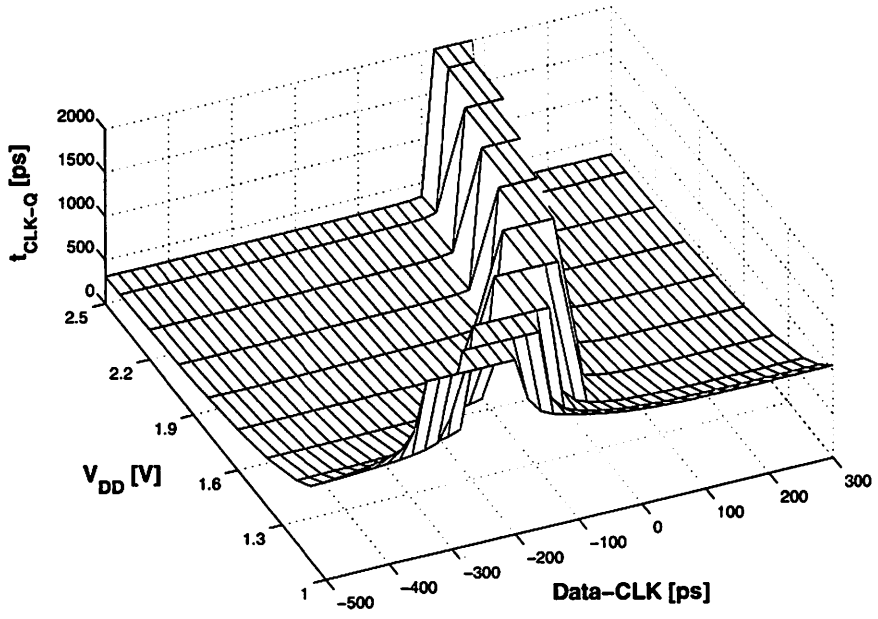


FIGURE 3.5 Illustration of variation of the minimum data pulse width with supply voltage.

3.2 ENERGY METRICS

The battery life in portable digital systems, which is desirable to be as long as possible, is proportional to the energy consumption. In the analysis of energy consumption in the clock subsystem, one is specifically interested in the energy consumed in flip-flops during one clock period, T. Total energy consumption in a flip-flop during one clock period, is obtained using Equation (3.2):

$$E = \int_{t}^{t+T} V_{DD} \cdot i_{V_{DD}}(\tau) \cdot d\tau \tag{3.2}$$

where t is the time point, conveniently chosen to include all relevant transitions – arrival of new datum, clock pulse, and output data transition.

There are four components of the energy consumption in a digital CMOS circuit:

$$E = E_{switching} + E_{short-circuit} + E_{leakage} + E_{static} \tag{3.3}$$

where

$$E_{switching} = \sum_{i=1}^{N} \alpha_{0\text{-}1}(i) \cdot C_i \cdot V_{swing}(i) \cdot V_{DD} \tag{3.4}$$

$$E_{short-circuit} = \int_{t}^{t+T} i_{short-circuit}(\tau) \cdot V_{DD} \cdot d\tau \tag{3.5}$$

$$E_{leakage} = \int_{t}^{t+T} i_{leakage}(\tau) \cdot V_{DD} \cdot d\tau \tag{3.6}$$

$$E_{static} = \int_{t}^{t+T} i_{static}(\tau) \cdot V_{DD} \cdot d\tau \tag{3.7}$$

In Equation (3.4), N is the number of nodes in a circuit, C_i is the capacitance loading node i, $\alpha_{0\text{-}1}(i)$ is the probability that an energy consuming transition occur at node i, and $V_{swing}(i)$ is voltage swing of node i. The switching component of energy is the main contributor to the overall energy consumption. The short-circuit and leakage components can be important as well, depending on the

supply voltage and input signal slope. The component due to static currents appears in two cases: 1) reduced voltage levels driving CMOS circuits, and 2) circuits with DC current (e.g. pseudo NMOS circuits). This component is negligible in most flip-flops.

Short Circuit Component of Energy

In the case of a CMOS inverter, due to finite rise and fall times of input waveform when $V_{Tn} < V_{in} < V_{DD} - |V_{Tp}|$ both the NMOS and PMOS transistors are *on*, causing short circuit energy consumption. During the pull-up operation, as shown in Fig. 3.6.a, it is desirable that all pull-up current of the PMOS transistor be delivered to C_L, in which case current of the NMOS transistor is short-circuit current. Similarly for the pull-down operation, current of the PMOS transistor is short-circuit current as shown in Fig. 3.6.b.

To measure the energy consumption due to short-circuit current, one needs to know what type of transition happens at every node in a circuit and during which time interval, and then integrate $i_{\text{short-circuit}}$ over the appropriate time interval:

$$E_{short-circuit} = \sum_{i=1}^{N} \int_{t_1(i)}^{t_2(i)} i_{short-circuit}(i,\tau) \cdot V_{DD} \cdot d\tau \qquad (3.8)$$

In Equation (3.8), N is the number of nodes in a circuit, while $t_1(i)$ and $t_2(i)$ define the time intervals during which there is a short-circuit current flow through the devices connected to node i.

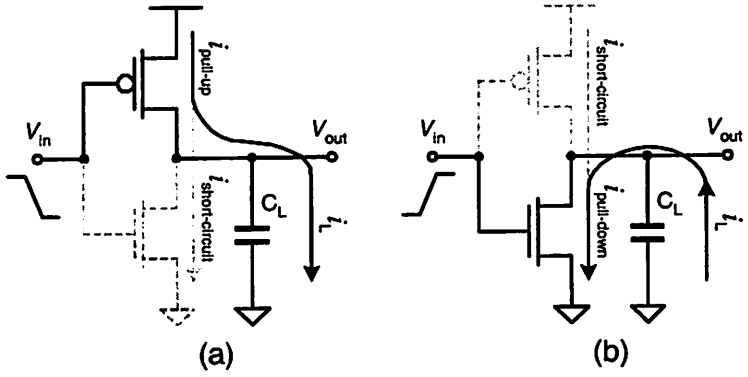


FIGURE 3.6 Short-circuit current in CMOS inverter during (a) pull-up, (b) pull-down operation.

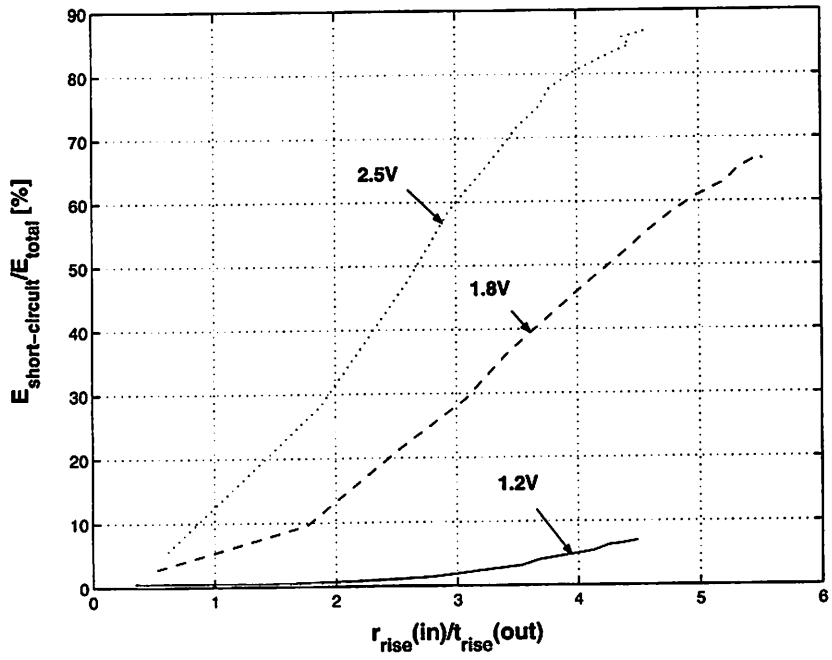


FIGURE 3.7 Short-circuit component of energy in a fanout-of-4 inverter.

The short circuit component of energy is typically less than 10% of the total energy. It depends on the input signal slope which defines the time interval during which both pull-up and pull-down devices are simultaneously *on*. The longer the *on* time, the greater the short-circuit component. Figure 3.7 shows the short-circuit component as percentage of the total energy versus the ratio of slopes of the input and output waveforms in a fanout-of-4 inverter.

For supply voltages $V_{DD} < V_{Tn} + |V_{Tp}|$, the short circuit energy is zero, while it becomes more significant for higher supply voltages because the voltage swing, $V_{DD} - |V_{Tp}| - V_{Tn}$, when both the NMOS and PMOS transistors are *on* increases. In addition, the short circuit component of energy as a portion of the total energy increases if the slope of input waveform is increased relative to the slope of output waveform. Therefore, the recommended region of operation for negligible short-circuit component of energy is in the range 0.8V-1.2V of supply voltage.

Leakage Component of Energy

Leakage component of energy arises from two types of leakage currents: 1) reverse-bias diode leakage at the transistor drains, and 2) sub-threshold leakage

through the channel of an *off* device. These components are determined by the process technology.

The diode leakage occurs when a transistor is *off*, and the drain-bulk or source-bulk diode is reverse biased so that it conducts current. The leakage current of the reverse biased diode is given by:

$$I_{leakage} = I_{sat} \cdot \left(e^{\frac{V}{V_t}} - 1 \right) \tag{3.9}$$

where V is the voltage of forward-biased diode. When the diode is reverse biased, its current is approximately equal to the reverse saturation current. This component is typically negligible compared to the subthreshold leakage component.

The subthreshold leakage component is due to carrier diffusion between the source and drain when the channel-to-substrate surface potential is $\phi_B < \phi_S < 2\phi_B$, which corresponds to the moderate inversion region, [Wolf95]. The drain-to-source current in the subthreshold region is exponentially proportional to the gate-to-source overdrive, $V_{GS} - V_T$, as given by Equation (3.10):

$$I_{DS,subthreshold} \propto e^{\frac{q\cdot(V_{GS}-V_T)}{n\cdot k\cdot T}} \cdot \left(1 - e^{\frac{q\cdot V_{DS}}{k\cdot T}} \right) \tag{3.10}$$

For $V_{DS} >> kT/q$, last term is approximately equal to 1, and I_{DS} is independent on V_{DS} which typically happens for V_{DS} larger than 0.1V [Chandr94]. This current is becoming increasingly important with the scaling of CMOS technology because the subthreshold slope increases due to an increase in gate-to-drain overlap capacitance [Wolf95]. The energy consumption related to the subthreshold current is given by Equation (3.11).

$$E_{subthreshold} = \sum_{i=1}^{N} \int_{t_1(i)}^{t_2(i)} i_{subthreshold}(i,\tau) \cdot V_{DD} \cdot d\tau \tag{3.11}$$

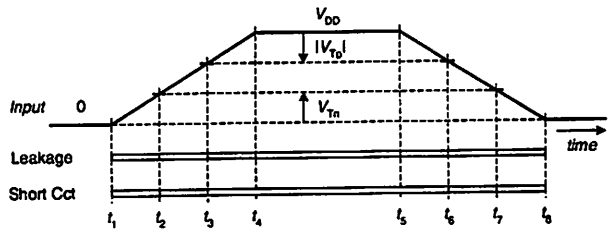


FIGURE 3.8 Short-circuit and leakage currents in CMOS inverter during pull-up/down.

where N is the number of nodes in a circuit, while $t_1(i)$ and $t_2(i)$ define time intervals during which there is a subthreshold current through the devices connected to node i.

Referring to Fig. 3.8, when the gate-to-source voltage of a MOSFET device is less than the threshold voltage leakage occurs: 1) during time intervals $\{t_3, t_4\}$ and $\{t_5, t_6\}$, where the current of the PMOS device contributes to the leakage component of energy, and 2) during $\{t_1, t_2\}$ and $\{t_7, t_8\}$, where the current of the NMOS device contributes to the leakage component of energy. Simulation results showing the subthreshold (leakage) component of energy in a fanout-of-4 inverter are shown in Fig. 3.9.

Figure 3.9 shows that the leakage component of energy is negligible at high supply voltages (Fig. 3.9.a), but that its value is almost constant over different supplies (Fig. 3.9.b). In contrast, both the switching and short-circuit components of energy increase with supply voltage. Therefore, the desirable region of operation is at low supply voltages.

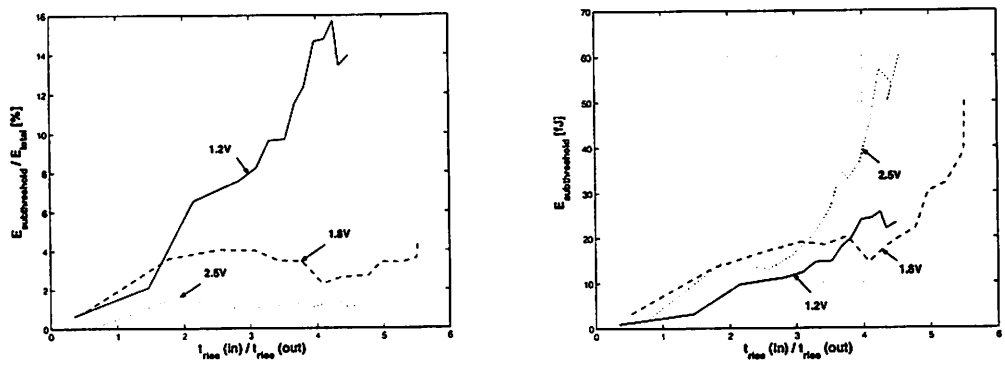


FIGURE 3.9 Subthreshold component of energy in a fanout-of-4-inverter: (a) as a fraction of the total energy, (b) absolute value.

Energy vs. Power-Delay-Product (PDP)

The product of power and flip-flop *Clk-Q* delay is commonly used to represent the energy consumption of a flip-flop. Comparing the expressions for power consumption (Equation (3.12)) and energy consumption (Equation (3.7)) in flip-flops, it turns out that $E = PDP$ only if the flip-flop *Clk-Q* delay is equal to the clock period.

$$P = \frac{1}{T} \cdot \int_{t}^{t+T} i_{V_{DD}}(\tau) \cdot V_{DD} \cdot d\tau \tag{3.12}$$

The condition of $t_{CLK\text{-}Q} = T$ is unrealistic because the minimal clock period required for correct flip-flop operation is limited by the flip-flop setup time and *Clk-Q* delay, Equation (3.13). A flip-flop, therefore, cannot operate at $f = 1/t_{CLK\text{-}Q}$ if the setup time is positive.

$$T \geq t_{CLK-Q} + t_{setup} \tag{3.13}$$

A better approach is to calculate the energy consumption of flip-flops as the power consumption at a certain frequency and for certain input data patterns [Stojan99]. This is accurate, but it requires knowledge about input switching patterns and operating frequency. Also, an additional simulation is required to measure the energy for each new input data pattern, increasing the complexity of the entire process.

3.2.1 Energy per Transition

Our methodology is to determine the flip-flop energy consumption according to Equation (3.7), for all input state transitions shown in Fig. 3.10. This method is simple because it is necessary to simulate only four transitions to obtain $E_{0\text{-}0}$, $E_{0\text{-}1}$, $E_{1\text{-}0}$, and $E_{1\text{-}1}$. If the average input signal transition probabilities are known, it is

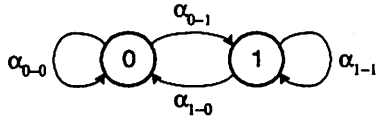


FIGURE 3.10 Input state transition diagram.

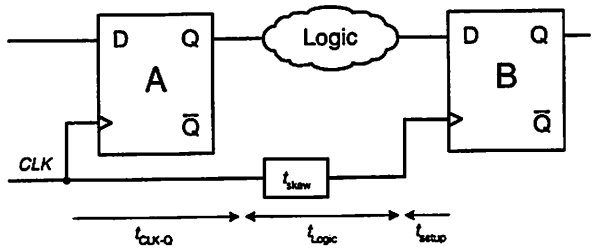


FIGURE 3.11 Flip-flop environment in a digital system (timing).

possible to calculate the average energy consumption of a flip-flop using Equation (3.14). Average power can be then obtained if the clock frequency is known.

$$E_{average} = \alpha_{0-0} \cdot E_{0-0} + \alpha_{0-1} \cdot E_{0-1} + \alpha_{1-0} \cdot E_{1-0} + \alpha_{1-1} \cdot E_{1-1} \qquad (3.14)$$

In addition, the energy per transition can be used to calculate energy breakdown in a flip-flop between clocked nodes, internal nodes, and energy in driving an external output load. Furthermore, if internal clock gating schemes are used, an overhead associated with the clock gating circuitry can be obtained which will aid in the estimation of energy saving capabilities of different clock gating schemes. Discussion about these components of energy will be presented in Section 3.3.2.

3.3 SYNTHETIC METRICS – A SYSTEM PERSPECTIVE

The flip-flop environment imposes additional constraints on the timing and energy parameters introduced thus far, meaning that there is still a need to introduce a few more synthetic metrics in order to effectively quantify the flip-flop behavior at the system-level.

3.3.1 Timing Metrics

Figure 3.11 depicts a model of a typical flip-flop environment in a digital system. The fundamental timing expression that this system has to satisfy for correct operation is given by:

$$T \geq 1.05 \cdot t_{CLK-Q,A} + t_{Logic} + t_{setup,B} + t_{skew} \qquad (3.15)$$

The first term of Equation (3.15) accounts for the worst-case clock-to-output delay, when data arrives exactly one setup time (5% increase in *Clk-Q* delay) before the active clock edge. The second term, t_{Logic}, defines what portion of the cycle time is consumed by combinational logic, while the third parameter, t_{skew}, describes the clock skew. The synthetic timing metrics can be derived based on Equation (3.15).

3.3.1.1 Delay

The flip-flop environment imposes a **maximum delay restriction** on the clock-to-output delay of a flip-flop given by Equation (3.16). Grouping the flip-flop parameters, $t_{CLK\text{-}Q}$ and t_{setup}, of Equation (3.16), the *delay* of a flip-flop is derived as given by Equation (3.17).

$$1.05 \cdot t_{CLK-Q} \leq T - t_{Logic} - t_{skew} - t_{setup} \tag{3.16}$$

$$D = 1.05 \cdot t_{CLK-Q} + t_{setup} \leq T - t_{Logic} - t_{skew} \tag{3.17}$$

> The *delay* of a flip-flop is the sum of its clock-to-output delay (measured at setup time) and setup time.

This is the only relevant flip-flop delay parameter with respect to system performance [Stojan99], since it describes how much of the clock cycle is occupied by the flip-flop. This is a robust metric, independent on simulation setup, so it avoids all sorts of misconceptions present in today's literature about flip-flop speed, often characterized as clock-to-output delay [Sven98], [Balsara00]. The flip-flop delay when data undergoes a *0-1* or *1-0* input transition is given by the following two equations. These definitions hold for both inverting and non-inverting flip-flops.

$$D_{0\text{-}1} = 1.05 \cdot t_{CLK-Q,0\text{-}1} + t_{setup,1} \tag{3.18}$$

$$D_{1\text{-}0} = 1.05 \cdot t_{CLK-Q,1\text{-}0} + t_{setup,0} \tag{3.19}$$

3.3.1.2 Internal Race Immunity

The maximal clock skew that a system can tolerate is determined by flip-flops. To quantify this last flip-flop timing metric, internal race immunity is introduced. Referring to Fig. 3.11, if the clock-to-output delay of flip-flop A is shorter than the hold time of flip-flop B and there is no logic in between, a race condition can occur. In other words, there is a **minimum delay restriction** on the clock-to-output delay given by Equation (3.20). If the flip-flop parameters of Equation (3.20) are grouped, internal race immunity of a flip-flop is derived as given by Equation (3.21).

$$t_{CLK-Q} \geq t_{hold} + t_{skew} \tag{3.20}$$

$$R = t_{CLK-Q} - t_{hold} \tag{3.21}$$

> Internal race immunity of a flip-flop is the difference between its clock-to-output delay and hold time.

The race immunity is a helpful metric that aids in a search for timing failures due to short-paths (races). Besides that, it also tells what the maximum clock skew a flip-flop can tolerate. Clock jitter does not affect the flip-flop race immunity. Equations (3.22) and (3.23) define the flip-flop internal race immunity for logic "0" and logic "1" in both inverting and non-inverting flip-flops.

$$R_0 = \begin{cases} t_{CLK-Q,0-1} - t_{hold,0} & (non-inverting) \\ t_{CLK-Q,1-0} - t_{hold,0} & (inverting) \end{cases} \tag{3.22}$$

$$R_1 = \begin{cases} t_{CLK-Q,1-0} - t_{hold,1} & (non-inverting) \\ t_{CLK-Q,0-1} - t_{hold,1} & (inverting) \end{cases} \tag{3.23}$$

The most common source of timing failures in high-speed digital systems is the lack of race immunity. Figure 3.12 depicts internal race immunity of several representative high-speed flip-flops [Partovi96], [Klass98], [Nikolic99]. It is interesting to note that the race immunity scales with supply voltage in similar fashion as the logic delay. Figure 3.12 explains why the clock skew requirement

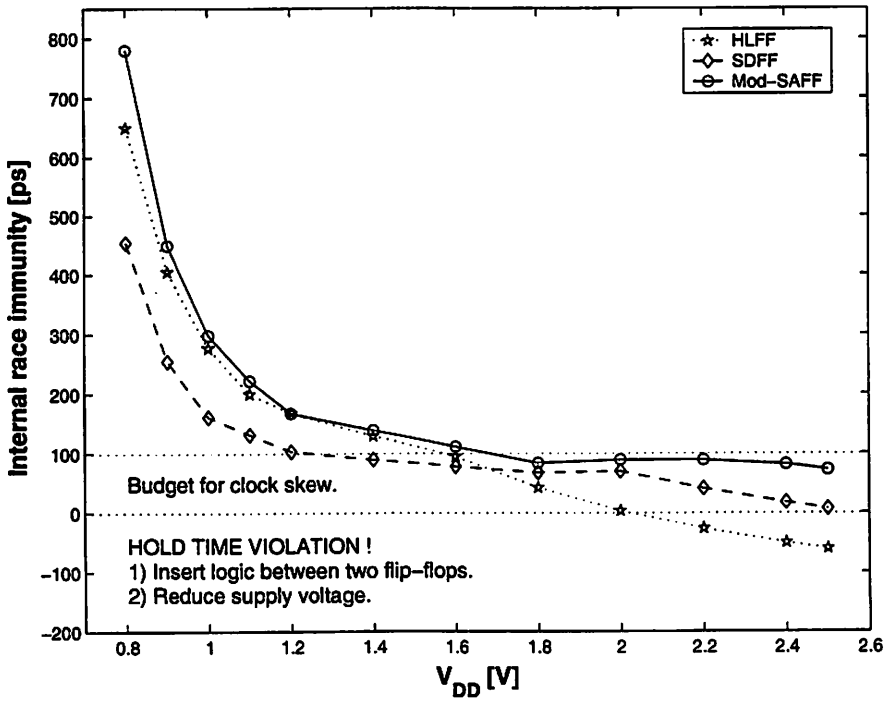


FIGURE 3.12 Race immunity of high-speed flip-flops under supply voltage scaling.

is so important in high-speed digital systems which operate at higher voltages. For example, clock skew in the 600MHz DEC Alpha [Bailey98] is only 75ps.

From the knowledge acquired so far about the flip-flop environment and internal race immunity, one apparent conclusion is that the design practices for good race immunity are: 1) insertion of logic in short paths, and/or 2) reduction of supply voltage. Depending on what is more important for a circuit designer, one of the two solutions is chosen: 1) if the speed is critical, the first alternative is better (at the expense of increased energy), while 2) energy conscious design requires reduced supply voltage (at the expense of reduced speed).

3.3.2 Energy Metrics

As a system-level designer, one would like to know how much energy is consumed in the clocking of flip-flops if the objective is to minimize energy of the clocking subsystem. The choice of flip-flop topologies that have low internal clock capacitance can save some clocking energy, but flip-flop efficiency is also important (i.e. what fraction of energy goes in the flip-flop internal nodes). This

motivates the idea to find the energy breakdown between: 1) clocked nodes in flip-flops, 2) internal nodes in flip-flops, and 3) external output load.

Table 3.1 summarizes which components of energy are contained in the energy-per-transition. The fields marked with both symbols denote that the specific component may be included or excluded, depending on the circuit structure. For example, E_{ext} is contained in $E_{0\text{-}1}$ in non-inverting flip-flops, and in $E_{1\text{-}0}$ in inverting flip-flops. Similarly, not all of the internal clock nodes in flip-flops are switched; for example in topologies which employ internal clock gating.

The flip-flop energy breakdown provides a good basis for the study of alternative circuit techniques that deal with internal clock gating and evaluation of the tradeoffs between a reduction in the clocking energy versus an overhead in the clock-gating logic, [Nogawa98], [Kuroda99], [Strollo00].

3.3.2.1 Clocking Energy in Flip-Flops

Clocking energy in a flip-flop is the energy consumed in its internal, clocked nodes even when the input data signal is not changing. In general $E_{0\text{-}0}$ and $E_{1\text{-}1}$ are not equal depending on the circuit structure. There are two cases to consider: 1) flip-flops without precharge nodes, and 2) flip-flops with precharge nodes.

Clocking Energy in Flip-Flops without Precharge Nodes

Clocking energy in flip-flops without precharge nodes [Suzuki73], [Gerosa94] is equal to the total flip-flop energy when the input data is at a constant logic level (input transitions *0-0* and *1-1*). Since there are no precharge (dynamic) nodes, all energy is consumed in switching of the internal clocked nodes, Equation (3.24).

$$E_{CLK} = E_{0\text{-}0} = E_{1\text{-}1} \tag{3.24}$$

TABLE 3.1. FLIP-FLOP ENERGY BREAKDOWN

	$E_{0\text{-}0}$	$E_{0\text{-}1}$	$E_{1\text{-}0}$	$E_{1\text{-}1}$
E_{CLK}	✓✗	✓	✓	✓✗
E_{int}	✗	✓	✓	✓✗
E_{ext}	✗	✓✗	✓✗	✗

Clocking Energy in Flip-Flops with Precharge Nodes

Flip-flops with precharge nodes operate in precharge-evaluate fashion and the precharge nodes need to be refreshed periodically by a regenerative circuit, typically implemented as cross-coupled inverters (pseudo-static operation).

In *single-input flip-flops* [Partovi96], [Klass98] $E_{0\text{-}0}$ and $E_{1\text{-}1}$ are different because the precharge node is precharged and discharged when the data input stays at logic "1", while it remains precharged when the data input stays at logic "0". The difference between $E_{1\text{-}1}$ and $E_{0\text{-}0}$ is equal to the energy consumed in the precharge and discharge of the internal precharge node, Equation (3.26). Since the precharge/discharge operation happens at every clock cycle when data stays at logic "1", $E_{1\text{-}1}$ also represents clocking energy, Equation (3.25). The equations below are valid for precharge circuits using NMOS pull-down logic, which is the case in all flip-flops analyzed in Chapter 5.

$$E_{CLK} = \{E_{0\text{-}0}, E_{1\text{-}1}\} \tag{3.25}$$

$$E_{pseudo\text{-}static\text{-}node} = E_{1\text{-}1} - E_{0\text{-}0} \tag{3.26}$$

In *differential-input flip-flops* [Matsui94], [Nikolic99] $E_{0\text{-}0}$ and $E_{1\text{-}1}$ are equal because of their differential nature – one differential-pair input is at logic "0" while the other input is at logic "1" so one of the output nodes in the differential-pair is precharged every clock cycle. Therefore, the clocking energy in differential-input flip-flops with precharge nodes is also defined by Equation (3.24). As opposed flip-flops without precharge nodes, where $E_{0\text{-}0}$ and $E_{1\text{-}1}$ describe the energy consumed only by the clocked transistors, in differential-input flip-flops there is an additional component of clocking energy – consumed in precharging/discharging of the precharge nodes.

3.3.2.2 Energy in Internal, non-Clocked Nodes in Flip-Flops

Calculation of energy consumed in internal, non-clocked nodes of a flip-flop is based on the measurement of energy consumed in the flip-flop when its data input undergoes a transition that results in a falling transition at the output – a *0-1*

transition at the input in inverting flip-flops, and a *1-0* transition at the input in non-inverting flip-flops, Equation (3.27). This equation holds both in flip-flops with and without precharge nodes.

$$E_{\text{int}} = \begin{cases} E_{0\text{–}1} - E_{CLK} & (non-inverting) \\ E_{1\text{–}0} - E_{CLK} & (inverting) \end{cases} \tag{3.27}$$

3.3.2.3 Energy in External Output Load

Energy consumed in charging of the external output load is equal to:

$$E_{ext} = C_{out} \cdot V_{swing} \cdot V_{DD} \tag{3.28}$$

Typically $V_{\text{swing}} = V_{\text{DD}}$ and Equation (3.28) reduces to a well-known quadratic formula. E_{ext} can also be calculated from the energy consumed in a flip-flop during one clock cycle when the data input undergoes a transition which results in a rising transition at the output, as given by Equation (3.30) for both inverting and non-inverting flip-flops.

$$E_{ext} = \begin{cases} E_{1\text{–}0} - E_{CLK} - E_{\text{int}} & (non-inverting) \\ E_{0\text{–}1} - E_{CLK} - E_{\text{int}} & (inverting) \end{cases} \tag{3.30}$$

3.3.3 Energy-Delay-Product

Energy consumption of a flip-flop can be approximated with the switching component which is proportional to $C \cdot V_{\text{swing}} \cdot V_{\text{DD}}$. Therefore, one can reduce the energy consumed by a flip-flop by reducing the supply voltage, reducing the signal swing, or decreasing the switched capacitance (circuit size). All of these techniques increase the delay of a flip-flop, so it is expected that the lowest energy flip-flop will also have lower performance. One usually wants to know the minimum energy for a given performance constraint or maximum performance for a given energy. Therefore, both quantities need to be considered simultaneously. The simplest way to quantify the compromise between energy and delay is to take the product of the two given by Equation (3.31).

$$EDP = E \cdot D \tag{3.31}$$

In Equation (3.31) the delay is the worst-case flip-flop delay given by Equation (3.32). The energy is average energy given by Equation (3.14).

$$D = \max\{D_{0-1}, D_{1-0}\} \tag{3.32}$$

3.4 DRIVING A FLIP-FLOP: INTERFACE WITH CLOCK NETWORK AND COMBINATIONAL LOGIC

The flip-flop metrics described thus far did not consider the input capacitance of the data and clock inputs. The clock network designer and logic designer need to be aware of these capacitances in order to design circuits that drive flip-flops.

3.4.1 Interface with Clock Network

The flip-flop parameters that affect timing specification of the clock distribution network are *clock skew* and *clock slope*. The important energy parameter is the total load of the clock distribution network which is defined by the input capacitance of the clock node and number of flip-flops on a chip.

3.4.1.1 Clock Skew

Besides the number of flip-flops on a chip that determines the clock load, the clock skew requirement is the reason why clock drivers have to consume energy in distributing the clock signal. The clock skew budget is determined by the flip-flop internal race immunity.

3.4.1.2 Clock Slope

Increase in *clock slope* results in degradation of the flip-flop performance, so clock network designer has to know what slopes flip-flops can tolerate. The clock slope also affects energy consumption of the clock distribution network. If larger clock drivers are used, the clock slope is smaller and flip-flop performance better, at the expense of an increase in energy consumption of the clock network. Optimal tradeoff is achieved with minimal energy consumption that delivers the desired flip-flop performance.

3.4.1.3 Clock Load

The clocking energy in a flip-flop, discussed in Section 3.3.2, is the amount of clocking energy expended in clocking of the flip-flop internal nodes. To evaluate the *total clocking energy* per clock cycle in the entire clock subsystem, one needs to add the energy consumed in the clock distribution network. The energy consumed in the clock distribution network depends on the total switched capacitance which is determined by the total number of flip-flops on a chip and the input capacitance of their clock inputs, the total wiring capacitance, and the total input capacitance of clock drivers as given by Equation (3.33).

$$C_{distrib-net} = N_{FF} \cdot C_{in-CLK,FF} + C_{wire} + C_{in-buff} \qquad (3.33)$$

The first term in Equation (3.33) is constant for a given flip-flop selection. The last two terms depend on buffer insertion/placement strategy and should be minimized. The shorter the total wire length, the smaller the wiring capacitance, C_{wire}. If wire lengths from clock drivers to flip-flops are not equal, there will be clock skew. The length of insertion delay from the root of the clock tree to the flip-flop clock inputs is not as important as it is important that these delays be balanced within clock skew specification. This places a constraint on how much extra wiring overhead one has to incur in order to keep the clock skew within a given margin. There also exists an energy-performance tradeoff between wide wires driving heavy nets, and narrow wires with buffer repeaters. Therefore, lower bound on the clock distribution energy consumption per clock cycle is imposed by $C_{\text{distrib-net}}$ and by the targeted clock slope at the inputs of flip-flops.

3.4.2 Interface with Combinational Logic

Similarly to driving the clock input of a flip-flop, one needs the parameters relevant for driving the flip-flop data input. The skew between the data inputs is not relevant as far as the data input signals arrive within setup/hold time specification. The relevant parameters to the combinational logic designer therefore are the input data slope and input data capacitance.

3.4.2.1 Data Slope

The data slope affects performance and energy consumption of both driving logic and flip-flops. Clock and data slopes are generally not equal.

3.4.2.2 Data Load

Energy consumption required to generate a signal that drives the flip-flop data input is not included in the energy budget of the flip-flop and is attributed to the logic network that computes the data. Typically the input load of the data input is very small.

3.5 OTHER COMMONLY USED METRICS

3.5.1 Power

Power is not a good metric to compare different flip-flops since it is proportional to the clock frequency. By simply reducing the clock speed one can reduce the power dissipated in a flip-flop. While the power decreases, the flip-flop does not really become "better" [Gonz96].

However, *average* power dissipation can be computed simply by multiplying average energy per clock cycle by the clock frequency. An average power is relevant in high-speed designs, when heat removal is an important issue. In portable systems that is not the case.

3.5.2 FO4 Inverter

Delays in CMOS circuits are decreasing with technology advancements. The FO4 inverter delay is a useful metric which normalizes process variations since delays of CMOS circuits normalized to the FO4 inverter delay have very small variation over several technology processes [Harris]. This does not include interconnect delays which are important in larger blocks when the variation of normalized delays is higher.

TABLE 3.2. SLOPE AND DELAY OF FO4 INVERTER OVER RANGE OF SUPPLY VOLTAGES

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4	2.5
Slope [ps]	1801	1015	690	523	362	287	245	219	202	190	179	176
t_P [ps]	1169	650	431	320	213	165	138	122	111	103	97	95

The FO4 inverter is a representative delay element because the fanout-of-4 is typically used in tapered buffers driving large loads. Figure 3.13 shows the simulation setup used to measure the FO4 inverter delay. The first stage provides a realistic slope for the input waveform. The last stage, 64*x* inverter, prevents excessive Miller multiplication of the third inverter's C_{gd}.

The FO4 inverter delay can be used to compare different architectures of the same circuit in different technologies when comparison in pS's or mW's is not fair. Table 3.2 shows how delay and the slope of output waveform in FO4 inverter scale with supply voltage. The slopes from the table are applied to the data and clock inputs in the simulations while the supply voltage is swept from 0.7V to 2.5V.

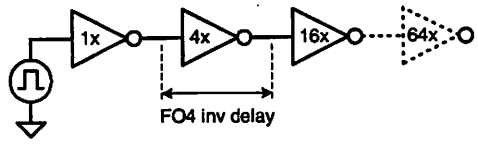


FIGURE 3.13 Fanout-of-4 inverter.

3.6 SUMMARY

Many textbooks on digital design often discuss the fundamental parameters that determine the properties of flip-flops. While these metrics often serve as a good starting point for simple designs, complex systems require more robust measures that encapsulate the results of these simple metrics and also offer novel insights to other system level parameters. Often, we can view a design involving flip-flops as having three important classes of metrics that help define the performance and energy consumption of a system – general (or "textbook") metrics, *synthetic* metrics, and *interface* metrics. For low-energy design, the latter two types are

extremely important and must be well understood in order to master superior design techniques.

Common metrics found in academic and general literature provide simple definitions for relevant timing and energy information. Examples of timing metrics widely used in flip-flop characterization include clock-to-output delay, setup time, hold time, and data-to-output delay. The clock-to-output delay is defined as the time in which a new output signal is stable relative to the active clock edge. Setup time refers to time period during which the signal at the input node must stay constant prior to the arrival of the active clock edge in order for the flip-flop to correctly sample its value. Its counterpart, hold time, on the other hand is the time period during which input must stay constant *after* the active clock edge to ensure proper input sampling. Most flip-flop designs today will ensure correct sampling even when the inputs do no remain stable for the entire duration of the time window dictated by the truest definition of setup and hold times. As a result, designers often define setup and hold times in terms of the tolerable degradation in clock-to-output delay relative to its nominal value (when the input node arrives and stabilizes much earlier than the arrival of the active clock edge). The data-to-output delay is the time period it takes a *0-1* or *1-0* input transition to propagate to the output. This, as we saw, is not a good performance metric because data arriving much earlier than the active clock edge adds to the data-to-output delay. This extra delay incurred has no physical significance in our timing analysis, which is why the metric is often ignored. It will also always be evident that useful timing metrics are those that affect and those that tightly relate to the energy that a flip-flop consumes. To this end, it is also no surprise that commonly used metrics also incorporate information about energy.

There are four components of energy consumption in digital circuits, consisting of switching, static, leakage, and short-circuit components. Good circuit designs today have about 90% of the total energy consumed during switching activity. The energy consumption of a circuit during one clock period, T, is given by:

$$E_{tot} = \int_{t_1}^{t_1+T} V_{DD} \cdot i_{V_{DD}}(t) \cdot dt \qquad (3.34)$$

For system designs a more useful energy metric is the *energy-per-transition.* Defined, the energy-per-transition is the total energy consumed in a flip-flop during one clock cycle for a specified input data transition (*0-0*, *0-1*, *1-0*, or *1-1*). This metric is extremely crucial in that it yields significant insight about circuit energy without the need for complex and intricate mathematical formulas. We can obtain this information empirically by running only four simulations to compute $E_{0\text{-}0}$, $E_{0\text{-}1}$, $E_{1\text{-}0}$, and $E_{1\text{-}1}$. Subsequently, these four values can be used to calculate the flip-flop energy consumption for any given input data pattern. Another commonly used energy related metric is the power-delay-product, used to estimate the energy consumption of flip-flops; however, this is an inexact energy metric because the flip-flop delay is in general not equal to the clock period. The timing and energy metrics described thus far ignore constraints imposed by the flip-flop environment, which raises the need for additional metrics that quantify flip-flop behavior at the system level.

The system level behavior of a flip-flop is quantified with synthetic timing and energy metrics that are derived from the simpler metrics introduced above. The timing parameters of flip-flops – clock-to-output delay, setup and hold times, and data-to-output delay, only characterize standalone flip-flops, often overlooking the system-level issues such as clock skew and flip-flop delay relative to the system clock. These issues are better understood with synthetic timing metrics such as *delay* and *internal race immunity*, which consider the clock-to-output delay and its specified variation from the nominal value for various data-to-clock offsets. The delay, D, is defined as $D = 1.05 \cdot t_{CLK\text{-}Q} + t_{setup}$. The internal race immunity, R, is defined as $R = t_{CLK\text{-}Q} - t_{hold}$. The critical assumption made here is that a system can tolerate a 5% increase in clock-to-output delay in the setup and hold time regions. The delay and internal race immunity are crucial parameters in the discovery of critical paths and timing failures due to short paths. While meeting timing requirements, it is also

important to minimize system energy consumption and to gain a firm understanding of where the energy is consumed.

By inspection of the node activity in a flip-flop for different input data transitions, the energy-per-transition can be utilized to obtain the energy breakdown between clocked nodes, internal nodes, and the external output load. This forms a good basis for the study of alternative circuit techniques that deal with internal clock gating. The energy breakdown information also offers valuable information about the tradeoffs associated with reduced clocking energy and the energy penalty incurred by the clock-gating logic, thus providing a better understanding of the optimization goals for the overall design. A reduction in energy consumption typically implies degradation in circuit performance, so the energy-delay-product is used to quantify a compromise between these two quantities. Besides careful flip-flop selection and optimization of its timing and energy parameters, it is equally important to understand the issues related to the interface between the clock distribution and combinational logic networks.

Designing different circuits that share common nodes requires a careful and deliberate consideration of the common properties exhibited in such nodes. These properties are best categorized as interface parameters. The input capacitance of the flip-flop clock and data inputs, as well as the external output capacitance, are interface parameters cogent to the design of clock and combinational logic networks. In addition, slopes of these waveforms impact the timing behavior immensely and must be taken into account as well. In our preceding analysis, the nominal values for data and clock slopes are estimated from the slopes of the output waveform of the fanout-of-4 (FO4) inverter.

The FO4 inverter delay is a useful metric that normalizes the variations in process technology since the delays of CMOS circuits normalized to the FO4 inverter delay exhibit inherently minute variations over different process technologies. This, as we recall, ignores the interconnect delays which are important in larger blocks in which case the variation of normalized delays will be larger. The FO4 inverter is a representative delay element because the fanout-

of-4 is typically used in tapered buffers driving larger load capacitances. The FO4 inverter delay can also be used to compare different architectures of the same circuit in different technologies when differential accuracy to the nearest picosecond or milliwatt is inadequate. Another commonly used metric in VLSI design is *average power consumption*; however, this is also not a favorable or useful metric since it is proportional to the clock frequency. By simply reducing the clock speed, one can reduce the power dissipated in a flip-flop. It is important to understand that reduced power dissipation in a flip-flop does not subsequently constitute a better design. In addition to lower energy designs, we must keep in mind that the performance specifications must also be met with an optimum speed-energy solution.

In conclusion, it is imperative to understand the different timing and energy metrics of flip-flops that quantify both circuit level and system level issues. A careful consideration given to these parameters enables the designer to prioritize and effectively optimize the essential parts of a circuit directly requiring special design attention. Being able to quickly analyze the consequences as well as the benefits simplifies the design process flow. More importantly, the system/circuit designer is able to correctly choose memory elements and a suitable clock tree early in the design phase that meet the stringent restraints in both the timing and energy domains.

Chapter 4

APPROACHES TO MINIMIZING ENERGY CONSUMPTION IN FLIP-FLOPS

The energy consumption in a flip-flop can be approximated by its switching component given by:

$$E_{switching} = \sum_{i=1}^{N} \alpha_{0\text{-}1}(i) \cdot C_i \cdot V_{swing}(i) \cdot V_{DD} \qquad (4.1)$$

where N is the number of nodes in a flip-flop, C_i is the capacitance loading node i, $\alpha_{0\text{-}1}(i)$ is the probability that an energy consuming transition occurs at node i, and V_{swing} is the voltage swing of node i. According to Equation (4.1), several techniques can be applied to minimize flip-flop energy consumption.

4.1 MINIMIZING SWITCHED CAPACITANCE (AREA)

Switched capacitances in flip-flops are clocked capacitances and capacitances of internal (logic) nodes, which both have to be minimized for low energy operation. Minimizing of the clocked capacitances is more important because they are switched every clock cycle as opposed to the capacitances of the internal nodes that are switched only when output changes. The switched capacitance consists of the diffusion, wire, and parasitic capacitance.

When the output of a flip-flop changes, the flip-flop drives the external output capacitance, which is determined by the flip-flop environment as shown in Fig. 4.1. This capacitance consists of several components given by:

$$C_l = C_{wire} + C_{in,Logic} + C_{off-path} + C_{parasitic,FF} \qquad (4.2)$$

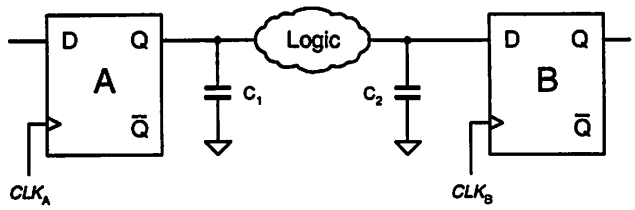


FIGURE 4.1 Flip-flop environment in a digital system (energy).

In Equation (4.1), $C_{off\text{-}path}$ is the input capacitance of other circuits that flip-flop A drives in addition to the logic that computes the logical input of flip-flop B.

It is therefore imperative to minimize all switched capacitances, but at times there is a lower limit to area reduction imposed by the requirements for circuit noise immunity. Standard cell libraries typically do not contain minimal sized transistors at the inputs of logic gates.

4.2 SUPPLY VOLTAGE SCALING

Energy consumption is a quadratic function of the supply voltage, so operating at reduced supply voltages leads to significant savings in the energy consumption. Lowering the supply voltage also helps reducing the total short-circuit and leakage components of energy. The reduction in supply voltage is limited by the required computational throughput and minimal acceptable ratio of switching component of energy to the total energy.

4.3 LOW SWING CIRCUIT TECHNIQUES

Sometimes a flip-flop operates with different logic levels for input and output signals. For example, in some memory designs, a low-swing output signal is amplified with sense-amplifiers.

Another low swing approach is the reduced swing clock operation, targeting savings in the clocking energy. The low swing clock can be generated with help of reduced-swing clock drivers [Sakurai98] or by powering up the clock buffers with separate supply voltage.

4.4 MINIMIZING SWITCHING ACTIVITY

There are two techniques to reduce switching activity in flip-flops: 1) proper selection of flip-flop topology that has inherently low switching activity of its internal nodes or small number of clocked nodes, and 2) clock gating which can be global or local.

4.4.1 Circuit Topology

Circuit optimization is related to the minimization of the total switched capacitance and the signal activity factor, α. Internal nodes of the static circuits in general have inherently lower switching activity than the internal nodes of the dynamic circuits due to the absence of dynamic nodes that need to be precharged every clock cycle. Dynamic operation also incurs overhead in energy, consumed in the logic which generates the clock signals that drive dynamic nodes, which is not required if static logic is used.

4.4.2 Clock Gating

Clock gating is an efficient way of reducing the overall energy consumption in flip-flop-intensive digital systems when energy in the clocking subsystem is significant part of the overall system energy, or when the flip-flop data inputs have very small switching activities. The mechanism behind clock gating is to allow clocking of a flip-flop only if a new arriving datum is different from the current flip-flop output.

The clock gating can be *global* when the gating logic is shared between several flip-flops, or *local* when the gating logic is embedded in each flip-flop. In both cases, the design of the clock gating circuitry needs to be carried out carefully so that the savings in the clocking energy be bigger than the overhead incurred by the clock gating logic, for given input data statistics.

4.5 SIZING ISSUES IN FLIP-FLOP DESIGN

The imperative is to optimize the size of a flip-flop for minimal energy-delay-product which can be achieved by increasing speed or reducing area. To find an optimal compromise both options need to be considered. Majority of logic gates on a digital chip (typically over 80%) drive a fanout of four or less. Since different logic gates have different input capacitances, our assumption is that the external output load presented to the flip-flops is equal to four "standard loads" which is about 30fF. One "standard load" corresponds to the input capacitance of an inverter which has two times better driving capability than a minimum sized inverter.

Our approach is to reduce a flip-flop size as much as possible so that the minimum amount of energy is consumed when driving an external output load equivalent to four "standard loads." This implies the reduction of sizes of the clocked transistors and size optimization of non-clocked transistors that are responsible for circuit performance. The flip-flop size optimization depends on the flip-flop topology and is discussed individually for each of the flip-flops presented in Chapter 5.

4.5.1 The Method of Logical Effort

The flip-flop size optimization is carried out using the method of logical effort, which is an equivalent RC model based circuit delay optimization, [Harris99]. The RC delay model describes delays caused by the capacitive load that the logic gate drives and by the topology of the logic gate. Inverters, as simplest logic gates, drive loads best. More complex logic gates often require more transistors, some of which are connected in series, making them poorer than inverters at driving current. For example, a NAND gate has more delay than an inverter with similar transistor sizes when driving the same load. The method of logical effort quantifies these effects to simplify delay analysis for individual logic gates and complex multistage logic networks.

The delay of a logic gate has two components: 1) a fixed component called the parasitic delay p, and 2) a component that is proportional to the gate's output load called the effort delay of stage effort f. The total delay, measured in units of τ which is technology dependent parameter, is the sum of the effort and parasitic delays:

$$d = f + p \tag{4.3}$$

The effort delay depends on the load and on properties of the logic gate driving the load, which can be described by two terms – *logical effort* and *electrical effort*. The logical effort, g, describes the effect of the logic gate's topology on its ability to produce output current. It is independent of the size of the transistors in the circuit. The electrical effort, h, characterizes the load and describes how the size in the transistors in the gate affects its driving capability. It is defined by:

$$h = \frac{C_{out}}{C_{in}} \tag{4.4}$$

4.5.1.1 Multistage Logic Networks

In multistage logic networks, the method of logical effort reveals the best number of stages and the least overall delay by balancing the delay among the stages. The notions of logical and electrical effort can be generalized from individual gates to multistage paths.

The logical effort, G, along a path is the product of the logical efforts of all the logic gates along the path.

The electrical effort, H, of a multistage logic network is the ratio of the load capacitance at the last stage in the path to the input capacitance of the first logic gate in the path.

A new kind of effort, named *branching effort*, has to be introduced to account for fanout within a logic network. The branching effort, b, at the output of a logic gate is defined as:

$$b = \frac{C_{on-path} + C_{off-path}}{C_{on-path}} = \frac{C_{total}}{C_{useful}} \tag{4.5}$$

where $C_{\text{on-path}}$ is load capacitance along the path one is analyzing and $C_{\text{off-path}}$ is the capacitance of connections that lead off the path. If the path does not branch, the branching effort is equal to one. The branching effort, B, along the entire path is the product of branching efforts of all the gates along the path.

Similarly to the stage effort of individual logic gates, the *path effort, F,* is defined in the multistage logic networks as the product of the logical, electrical, and branching efforts: $F = G \cdot B \cdot H$.

Optimizing Number of Stages for Minimal Delay

Minimum delay along an N-stage logic network is achieved when each of the stages in the path bears the same stage effort (for more details, refer to Chapter 3 of [Harris99]). The minimal delay is achieved when the stage effort is:

$$\hat{f} = g_i \cdot h_i = F^{1/N} \tag{4.6}$$

In Equation (4.6), subscript i denotes the i-th stage on the path. The minimal path delay equals:

$$\hat{D} = N \cdot F^{1/N} + P \tag{4.7}$$

where P accounts for the total parasitic delay along the path. If $N = 1$, this equation reduces to Equation (4.3).

To equalize the effort borne by each stage on a path, appropriate transistor sizes for each stage of logic along the path have to be chosen. From Equation (4.6), starting at the end of the path and working backwards, applying the capacitance transformation:

$$C_{in}(i) = \frac{g_i \cdot C_{out}(i)}{\hat{f}} \tag{4.8}$$

the input capacitance, $C_{\text{in}}(i)$, of each logic gate on the path can be determined and appropriately distributed among the transistors connected to the input.

4.5.1.2 Logical Effort of Logic Gates Used in Flip-Flops

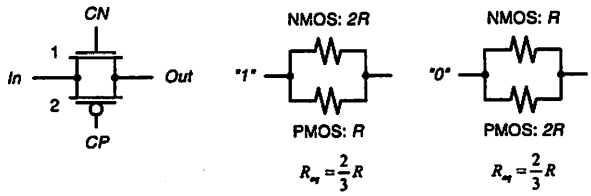


FIGURE 4.2 A transmission gate with PMOS and NMOS transistors sized 2:1, and its equivalent model for rising and falling transitions.

A CMOS *inverter* has the logical effort equal to one, by definition. The logical of a transmission gate is determined by comparing its equivalent resistance, when turned *on*, to the *on* resistance of a minimum-sized inverter, R. If the transistors in a transmission-gate are the same size as the transistors in a minimum sized inverter, each has equivalent resistance, R, when turned *on*. However, only one of the transistors is driving its good side (NMOS pulling down, or PMOS pulling up). If the transistor pulls in its poor direction, we assume its resistance is approximately doubled compared to the resistance when it pulls in its good direction as shown in Fig. 4.2. Therefore, the equivalent resistance of conducting transmission-gate is $R_{eq} = 2/3R$. The transmission-gate is passing a signal which is generated by some other logic circuit that can be modeled with its R_{on}. For simplicity, the logical effort of a transmission gate is approximately equal to one.

A 2-input NAND gate that has the same drive characteristics as an inverter with a pull-down transistor of width 1 and a pull-up transistor of width 2 is shown in Fig. 4.3.b. Because the two pull-down transistors are in series, each must have twice the conductance of the inverter pull-down transistor. Similarly, the pull-up

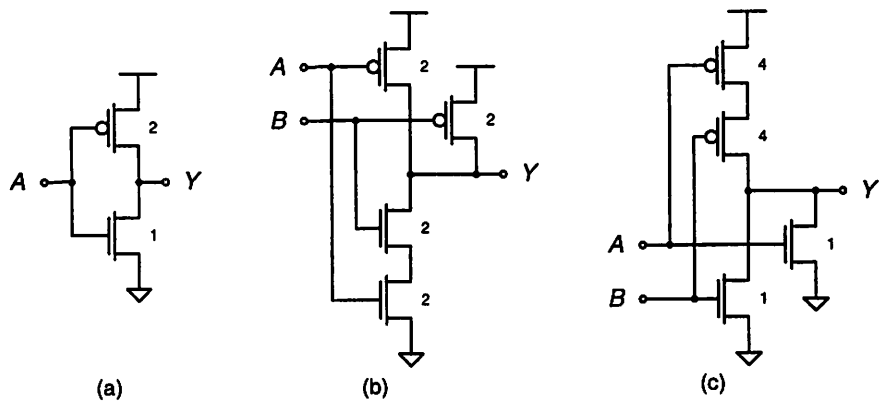


FIGURE 4.3 Simple logic gates: (a) reference inverter, (b) 2-input NAND gate, (c) 2-input NOR gate. [Harris99]

transistors of a 2-input NOR gate must have twice the conductance of the inverter pull-up transistor as shown in Fig. 4.3.c. The logical effort of these gates is found by extracting the input capacitances from the circuit schematics. One input of the NAND gate "sees" total transistor width of 4, while one input of the NOR gate "sees" total transistor width of 5. Since input of the inverter "sees" total transistor width of 3, the logical efforts of the NAND and NOR gates are 4/3 and 5/3, respectively.

4.5.1.3 Limitations of Logical Effort

The RC delay model fails to capture the effects of velocity saturation and of variable rise times. However, the rise times tend to be equal in well-designed circuits with equal effort delay and, fortunately, velocity saturation can be handled by characterizing the logical effort of gates through simulation.

The logical effort does not account for the impact of interconnect delay on circuit performance and it is therefore not applicable in submicrometer designs when interconnect delays become important. Low energy circuits for portable applications typically operate at reduced supply voltages, in which case the interconnect delay is insignificant compared to the delay of logic gates as shown in Fig. 4.4. In Fig. 4.4, logic delay is estimated as the delay of a FO4 inverter, while wire delay is estimated as the delay of a 7-segment distributed RC network that loads the FO4 inverter.

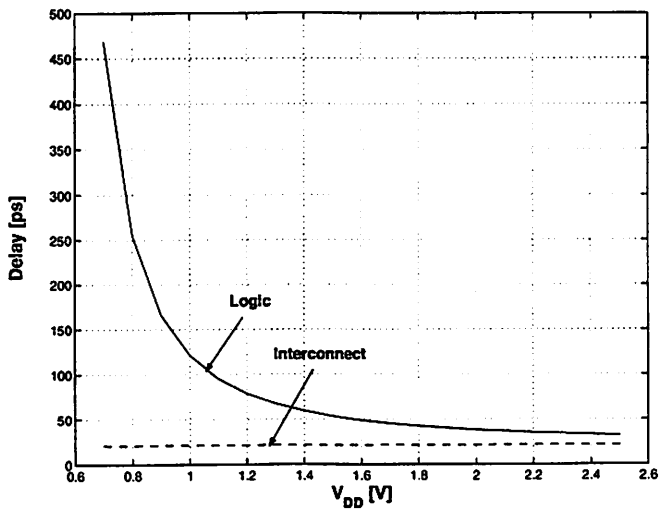


FIGURE 4.4 Scaling of interconnect and logic delay with supply voltage.

4.6 SUMMARY

A common design approach to minimize the energy consumption in flip-flops is to concentrate on reducing the switching component of energy given by $E = \alpha \cdot C_{sw} \cdot V_{swing} \cdot V_{DD}$. Based on this simple formula, the guidelines to reducing energy consumption are simply to minimize each of the terms in the product expression. The most efficient way to minimize energy, as it should be obvious, is aggressive voltage scaling because the energy has, to the first order, a quadratic dependency on supply voltage. Minimal switching activity can be achieved by circuit optimization and careful physical design as described earlier in the chapter. These are rather easy guidelines to follow because every circuit can generally be scaled to a lower voltage and the choice of circuit topology is always an option left available to the designer. In addition, transistors can be arbitrarily sized. However, the goal here is not to adjust each product term in the energy expression in a desultory fashion, but to find a systematic way to arrive at an optimal energy-performance solution that satisfies the design criteria.

The more advanced issue in minimizing energy consumption in flip-flops pertains to the circuit sizing that yields an optimal energy-performance tradeoff. Ideally, we would like to have the lowest possible energy and the highest level of performance. However, the two dimensions are often orthogonal quantities in circuit design. Clocked nodes should be made minimum size in order to compensate for the increased switching activity. The total circuit area ultimately depends on the size of load that it needs to drive, implying that larger loads may need a design of larger proportions to maintain acceptable driving strength; however, it is not desirable to size transistors in such a way that they are overly robust. This means that a circuit with transistors of fixed size cannot optimally drive various output loads, and that the extra area spent in designing them to support the largest load capacitance is really wasted since the performance upgrade here serves only to alleviate an issue that may manifest itself infrequently. It is more important from an energy standpoint that circuits are sized to satisfy the constraints in the most common cases, which often happens to

be approximately a four-fold increase over standard inverter loads. In more advanced VLSI designs today, the output loads are even lower 80% of the time.

The method introduced to aggressively optimize the transistor sizes is the method of *logical effort*. Logical effort is an RC-delay model based approach that describes the cost of computation inherent in the circuit topology that implements any given logic function. It quantifies the driving capability of a logic gate relative to a standard inverter, so that a valid correlation can be established between the required transistor sizes and the computed logical effort. Unfortunately, logical effort analyses ignore wire delay. As a consolation, this shortcoming can usually be well managed and only becomes a problem on large scale designs where the interconnect loads become significant and cannot be neglected.

Being able to systematically approach the transistor sizing issue completes the discussion of basic principles in low energy flip-flop design for digital systems. At this point in time, the designer is able to clearly discern the interdependencies that timing, energy, and sizing properties exhibit. Gaining a clear understanding of the tradeoffs associated with the various methods of optimization greatly reduces the effective design time and increases the likelihood of a more elegant, robust circuit solution. Keen design decisions made at the circuit level ultimately expedite the design at the system level with little or no sacrifice to energy and performance. Many widely accepted flip-flop designs that yield low-energy and above average performance characteristics can be derived and analyzed using the aforementioned methodologies. In the next chapter, we look at a number of practical designs as a case study for the application of many of these principles that help improve digital system designs.

Chapter 5

LOW-ENERGY FLIP-FLOP DESIGN

Two dominant architectures in flip-flop design are master-slave latch pairs and pulse triggered latches. The objective of this chapter is to present a methodology for the design of flip-flops that consume a minimal amount of energy for a given computational throughput.

In flip-flop-intensive designs such as DSP structures [Kuroda99], flip-flops consume most of the overall system energy, so the minimal flip-flop energy implies a reduction in the overall energy consumption. A plethora of recent research efforts are concerned with the optimization of the clocking energy in latches and flip-flops, and several flip-flop techniques that deal with these issues are introduced. The new flip-flops are often derived from master-slave latch-pairs or pulse-triggered latches. An example of such flip-flops is reduced clock swing flip-flops in which the clock signal does not swing from rail-to-rail [Sakurai98]. Other techniques employ internal clock gating when the input data signal is not changing (data-transition look-ahead structures), or triggering off of both clock edges (double edge-triggered flip-flops).

5.1 MASTER-SLAVE LATCH PAIRS

This section describes design and characterization of representative master-slave latch pairs: transmission gate based flip-flops and C^2MOS flip-flop.

5.1.1 Transmission-Gate Flip-Flop

The transmission gate flip-flop (TGFF) is built from two transmission gate latches. There are several latch circuits that can be used in the implementation [Xiaohai98]. The simplest TG latch is the latch shown in Fig. 5.1.a. The problem

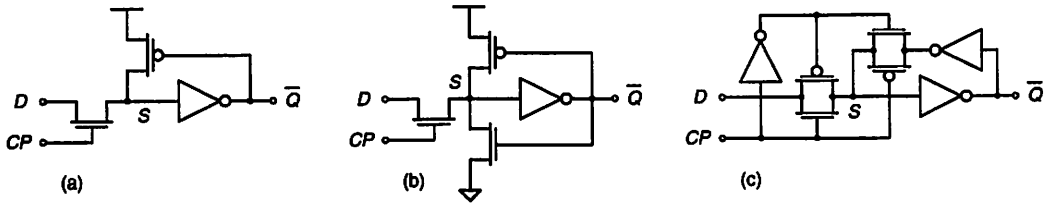


FIGURE 5.1 Transmission gate latches.

with this latch is that its storage node, *S*, appears dynamic because of the absence of pull-down transistor, rendering the latch susceptible to noise. A basic static version of this latch is shown in Fig. 5.1.b, where the pull-down NMOS device is added to the latch of Fig. 5.1.a. The NMOS transistor of the TG is weak pull-up device since a logic "1" is limited by the threshold voltage. Also, there is a simultaneous pull-up and pull-down between the NMOS transistor of the TG and the feedback transistors during both pull-up and pull-down on *S*. These problems are remedied in the circuit shown in Fig. 5.1.c. An extra feedback TG avoids the simultaneous pull-up/down problem, while an additional PMOS transistor of the input TG enables good, full swing, pull-up on node *S*. Robustness to noise of the latch in Fig. 5.1.c is therefore traded off for its increased clocking energy.

If the latch from Fig. 5.2.c is used, conventional TGFF as shown in Fig. 5.2 is obtained. It is a single-phase positive-edge triggered flip-flop with local clock generation (clock signals *CN*, and *CPI*). The energy consumption of this flip-flop can be reduced if the wire connecting drains of the top PMOS and bottom NMOS transistors is removed as shown in Fig. 5.3. Removal of the wire allows for more efficient layout due to the reduction of contact holes when the transmission gates

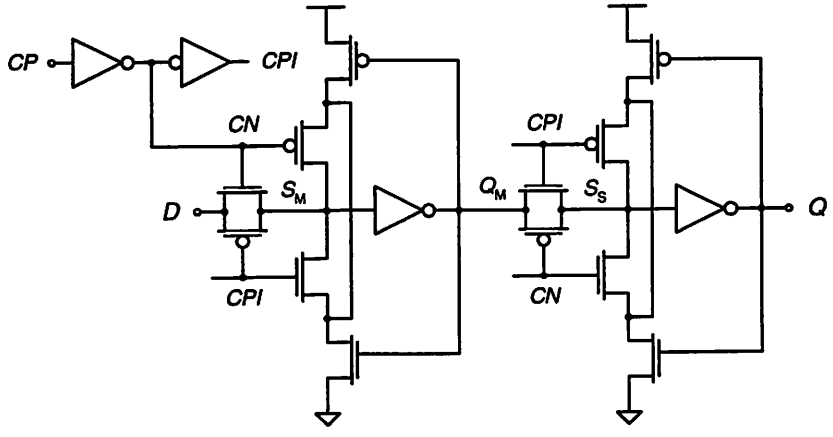


FIGURE 5.2 Conventional transmission gate flip-flop.

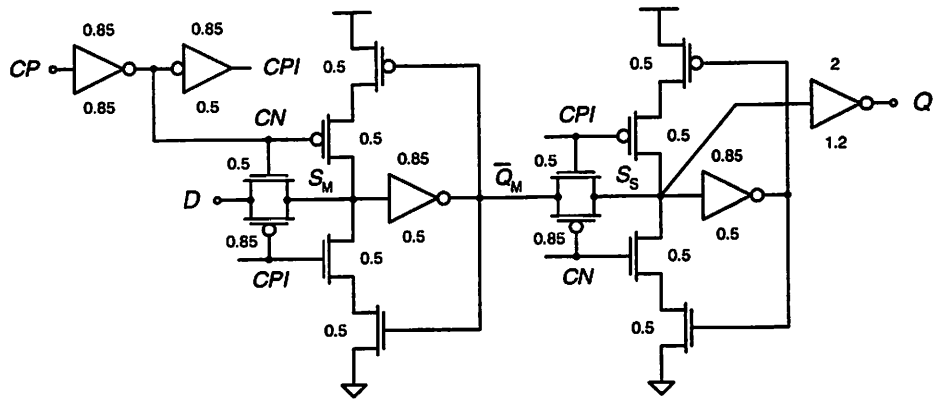


FIGURE 5.3 Transmission gate flip-flop (4-stage realization).

are replaced with series connected switches [Suzuki73]. The slave latch has the same structure as the master latch with the addition of an extra inverter that drives the output and prevents loading of the feedback loop by the output capacitance.

5.1.1.1 Circuit Operation

When the clock, *CP*, is "low" the transmission gate of the master latch is transparent and an input datum, *D*, is stored on the master's latch storage node, S_M (master latch is in *write* mode). The output, $\overline{Q}_M$, of the master latch follows S_M and stores its inverse. The feedback of the master latch is electrically open, while the feedback of the slave latch is electrically closed keeping the previously stored value at the slave's storage node, S_S.

When the clock, *CP*, goes from "low" to "high" the transmission gate of the master latch is opaque, the feedback of the master latch closes up keeping the stored value of $\overline{Q}_M$. The slave latch is in *write* mode, and the output of the master latch, $\overline{Q}_M$ is passed through the transmission gate of the slave latch and stored on the storage node, S_S, of the slave latch. This newly stored value of S_S is inverted and passed to the output, *Q*, of the flip-flop.

5.1.1.2 Sizing

To optimize a flip-flop for minimal energy-delay-product (*EDP*), one first needs to find an active circuit that is responsible for output transitions after the arrival of the active clock edge. The active circuit of the 4-stage TGFF is shown in

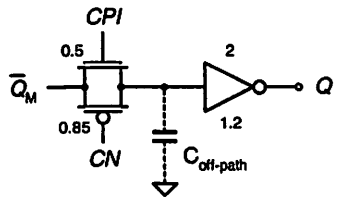


FIGURE 5.4 Active circuit that limits clock-to-output delay of the TGFF.

Fig. 5.4. Since $\overline{Q}_M$ does not change at the active clock edge, the critical path is from the clock bundle {*CN, CPI*} to the output *Q*. This active circuit is sized according to the method of logical effort, constrained with the requirement for smallest possible circuit area. The area is mainly decreased by the reduction of clocked capacitances and by sizing the flip-flop's combinational logic to drive relatively small output load – four "standard loads" – at the output with minimal delay.

Small clock load requirement determines the sizes of the transistors driven by {*CN, CPI*}. The transmission-gate of Fig. 5.4 is sized to have equivalent drive as a minimum sized inverter. The transmission-gate drives the input capacitance of the output inverter (*on-path* capacitance) and the input capacitance of the feedback inverter (*off-path* capacitance). The on-path capacitance is *x* times larger than the input capacitance of a minimum sized inverter. The off-path capacitance is determined by the input capacitance of the inverter in the feedback path which is equivalent to a minimum size inverter, so $C_{\text{off-path}} = C_{\text{min}}$. Therefore, the branching effort at the output of the 1st stage is given by:

$$b_1 = \frac{x+1}{x} \tag{5.1}$$

and it determines the total branching effort, *B*, since there is no branching at the output of the 2nd stage:

$$B = b_1 = \frac{x+1}{x} \tag{5.2}$$

The resulting logical effort *G*, electrical effort *H*, total path effort *F*, and optimal stage effort $\hat{f}$, are given by equations (5.3)-(5.6).

$$G = 1 \tag{5.3}$$

$$H = \frac{C_{out}}{C_{in}} = 8 \tag{5.4}$$

$$F = G \cdot B \cdot H = 8 \cdot \frac{x+1}{x} \tag{5.5}$$

$$\hat{f} = \sqrt{F} = \sqrt{8 \cdot \frac{x+1}{x}} \tag{5.6}$$

In Equation (5.4), C_{in} is the input capacitance looking into {*CN, CPI*} of the circuit shown in Fig. 5.4 and is equal to the capacitance of a minimum sized inverter, C_{min}. C_{out} is the output capacitance equivalent to four "standard loads" and is equal to $8 \cdot C_{min}$.

The input capacitance of the 2nd stage then is:

$$C_{in_2} = \frac{g_2 \cdot C_{out_2}}{\hat{f}} = \frac{8 \cdot C_{min}}{\sqrt{8 \cdot \frac{x+1}{x}}} = x \cdot C_{min} \tag{5.7}$$

which gives $x = 2.4$ and the size of inverter in Fig. 5.4 of $W_p/W_n = 2\mu/1.2\mu$.

All the transistors in feedback loop (keeper transistors) in both master and slave stage are minimum width. Since the master stage drives small load – the input capacitance of its feedback transistors and the input capacitance of the transmission gate in the slave stage – its gates have driving strength similar to a minimum sized inverter.

5.1.1.3 Characterization

Timing Parameters

Flip-flop timing parameters are summarized in Table 5.1. This flip-flop is an example of unequal setup and hold times of logic "0" and logic "1" as a consequence of the circuit structure. It is also an example of the setup and hold times that can take both positive and negative values depending on the supply voltage. The difference between setup and hold times is illustrated in Fig. 5.5 which shows the flip-flop master stage responsible for the setup and hold times.

TABLE 5.1 TIMING PARAMETERS OF 4-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	$t_{C\text{-}Q,0\text{-}1}$ [ps]	$t_{C\text{-}Q,1\text{-}0}$ [ps]	$t_{setup,0}$ [ps]	$t_{setup,1}$ [ps]	$t_{hold,0}$ [ps]	$t_{hold,1}$ [ps]	D [ps]	R [ps]
1.0	1159	1280	220	100	360	-30	1564	799
1.2	726	788	150	70	180	-20	977	546
1.8	263	376	100	40	70	-10	495	293
2.5	254	257	70	30	50	0	340	204

At the rising edge of *CP* the transmission gate of Fig. 5.5 switches *off*. The PMOS transistor of the transmission gate switches *off* Δt later then its NMOS transistor, so the *pull-up* side conducts Δt longer than the *pull-down* side. A new input datum, *D*, can therefore setup to logic "1" Δt later than to logic "0" leading to a shorter setup time of logic "1". In addition, at the rising edge of *CP*, the pull-up side of the local feedback in the slave latch is enabled Δt earlier than its pull-down side which helps a *0-1* transition at the input of the transmission-gate to propagate faster, further shortening the setup time of logic "1". Similarly, the hold time of logic "0" is longer than the hold time of logic "1".

This flip-flop exhibits good internal race immunity especially at reduced supplies when an increase in $t_{CLK\text{-}Q}$ is greater than the decrease in t_{hold}, Table 5.1.

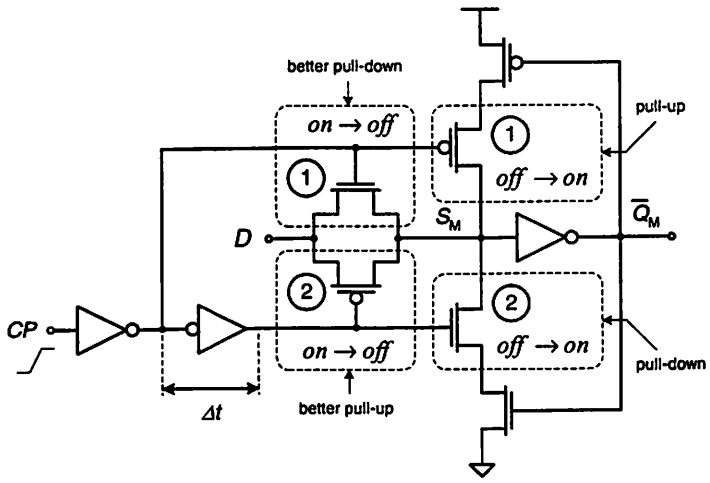


FIGURE 5.5 Active circuit that defines setup and hold times of the TGFF.

Energy Parameters

TABLE 5.2 BREAKDOWN OF ENERGY-PER-TRANSITION IN 4-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]		1.0	1.2	1.8	2.5
$E_{0\text{-}0}$ [fJ]	E_{CLK} [fJ]	12.5	18.4	45.9	100
$E_{0\text{-}1}$ [fJ]	E_{CLK} [fJ]	12.5	18.4	45.9	100
	E_{int} [fJ]	9.3	14.0	35.8	83.5
	E_{ext} [fJ]	30	43.2	97.2	187.5
$E_{1\text{-}0}$ [fJ]	E_{CLK} [fJ]	11.9	17.6	43.2	97.1
	E_{int} [fJ]	17.1	25.4	63.4	137.9
$E_{1\text{-}1}$ [fJ]	E_{CLK} [fJ]	11.9	17.6	43.2	97.1

Breakdown of energy-per-transition is presented in Table 5.2. From the table, it is interesting that E_{int}(1-0) is higher then E_{int}(0-1). This is because the total capacitance of the internal nodes that is charged up when *D* undergoes a *1-0* transition is higher than the total capacitance charged up when *D* undergoes a *0-1* transition as illustrated in Fig. 5.6.

In Fig. 5.6 C_1 is the sum of the output parasitic capacitance of the master stage and the input capacitance of the slave latch; C_2 is the sum of the output parasitic capacitance of the slave latch and the gate capacitance of feedback transistors; $C_{off\text{-}path}$ is the input capacitance of the feedback inverter.

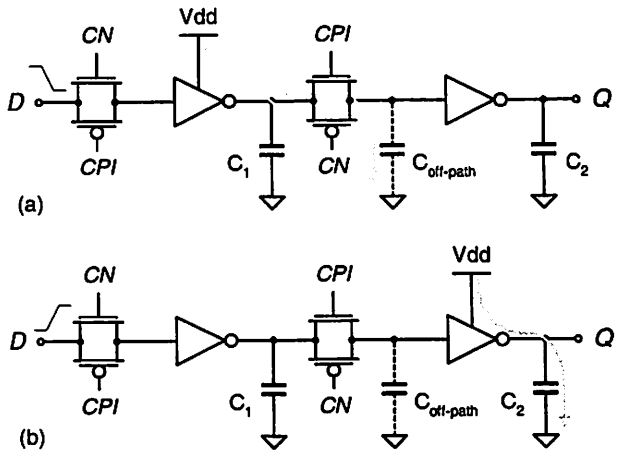


FIGURE 5.6 Energy consuming internal nodes when data undergoes: (a) *1-0*, (b) *0-1* transition.

As discussed in Chapter 3, the energy consumed in driving the flip-flop data input is attributed to the combinational logic that drives *D*. Therefore, the energy consumed in charging up the input capacitance of the master stage of the 4-stage TGFF is not included in $E_{0\text{-}1}$.

Throughout simulations performed to obtain the results in Table 5.1 and Table 5.2, the clock and data signal slopes are adjusted with supply voltage accordingly using the slopes of output waveform in a FO4 inverter, Table 3.1.

Interface Parameters

Clock slope: An increase in the slope of *CP* results in a very small increase in the slopes of internal clocks *CN* and *CPI*, so the flip-flop delay with respect to the internal clocks shows little variation. Therefore, increase in the flip-flop clock-to-output delay comes from the increase in slope of the external clock, *CP*. This is correct only at low supply voltages, as shown in Fig. 5.7. At higher supply voltages, the delay reference point $V_{DD}/2$ is well above V_{Tn}, meaning that the output transition is well under way when *CP* reaches 50% of its final value. As a result, the flip-flop clock-to-output delay appears shorter, Fig. 5.7.

The setup and hold times are affected in a similar fashion. Table 5.3 shows the variation in setup and hold times as a function of the slope of *CP* for supply voltage $V_{DD} = 1\text{V}$ and external output load $C_{out} = 4\text{SL}$. The race immunity appears slightly better at higher clock slopes because an increase in *Clk-Q* delay (450ps) is higher than the increase in the hold time (290ps). For more detailed results refer to Appendix A.1.

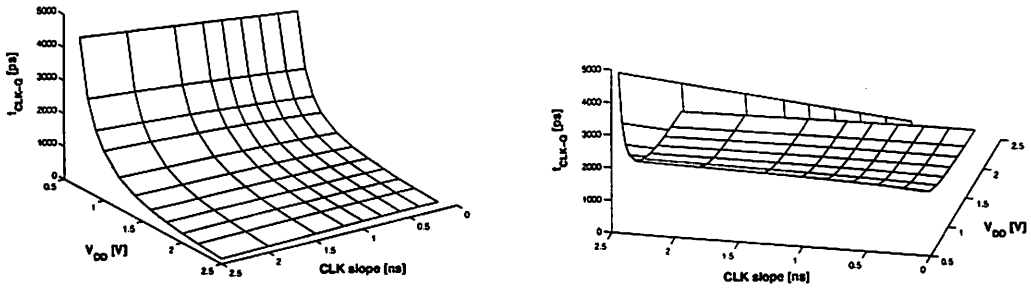


FIGURE 5.7 Dependency of *Clk-Q* delay on clock slope and supply voltage. $C_{out} = 4\text{SL}$.

TABLE 5.3 IMPACT OF CLOCK SLOPE ON TIMING OF TGFF. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	$t_{C\text{-}Q}$ [ps]	t_{setup} [ps]	t_{hold} [ps]	D [ps]	R [ps]
0.2	1109	250	190	1414	795
0.8	1269	130	270	1462	877
1.5	1408	60	340	1538	946
2.5	1561	0	480	1639	961

When the slope of *CP* is increased the short-circuit component of energy consumed by the flip-flop is increased at high supply voltages (see Chapter 3). Fig. 5.8 shows the flip-flop energy consumption during one clock cycle when *D* undergoes a *0-1* transition.

Output load: Variation of output load affects the flip-flop clock-to-output delay and $E_{0\text{-}1}$ due to a variation in the E_{ext} component. The setup and hold times are unaffected because they depend only on the input stage which is decoupled from the output (load at the output does not affect the load presented to the output of the master stage). Fig. 5.9 shows the variation in clock-to-output delay (computed as the average of $t_{CLK\text{-}Q,0\text{-}1}$ and $t_{CLK\text{-}Q,1\text{-}0}$) as a function of the output load. The figure on the right shows that the rising and falling output delays, which are not equal in general, can be equal for a certain output load.

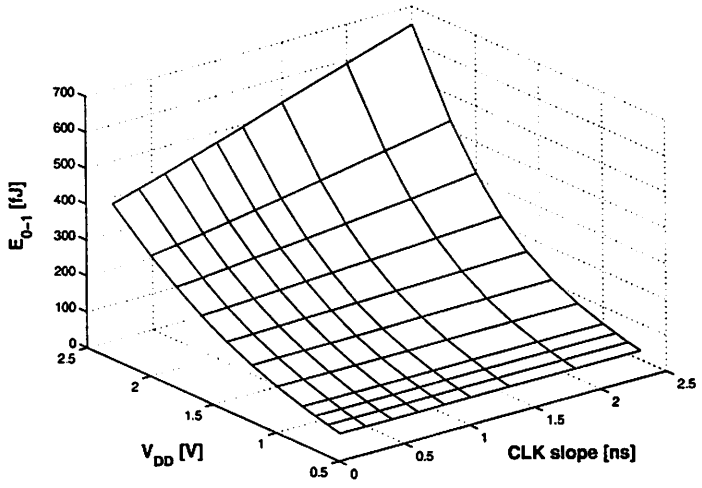


FIGURE 5.8 Energy per transition vs. clock slope and supply voltage. C_{out} = 4SL.

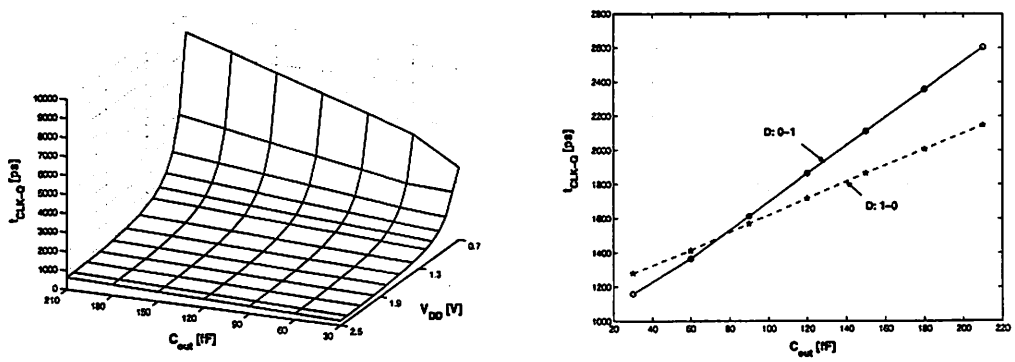


FIGURE 5.9 (a) Variation of clock-to-output delay with output load and supply. (b) Comparison of rising and falling delays as a function of output load at V_{DD} = 1V.

Physical parameters: Input capacitances of *D* and *CP* inputs are C_{in}(D) = 6.2fF and C_{in}(CP) = 2.45fF. Total gate width of this flip-flop is 15.65μm with all transistors minimum length, L_{min} = 0.5μm.

5.1.2 Transmission-Gate Flip-Flop with Input Gate Isolation

The master latch of the 4-stage TGFF is susceptible to the input charge injection. Noise sources that affect the latch state node are illustrated in Fig. 5.10.

The TGFF is susceptible to the first two noise sources which can be overcome by the input gate isolation as shown in Fig. 5.11 (shaded inverter). The noise sources arising from unrelated signal coupling (cross-talk) and power supply noise are attenuated by the latch feedback that makes *S* pseudo-static. There is an additional inverter at the output of the flip-flop in Fig. 5.11 for non-inverting operation (shaded inverter). The complementary output can be easily generated by addition of one extra inverter as shown in Fig. 5.11 (dashed inverter).

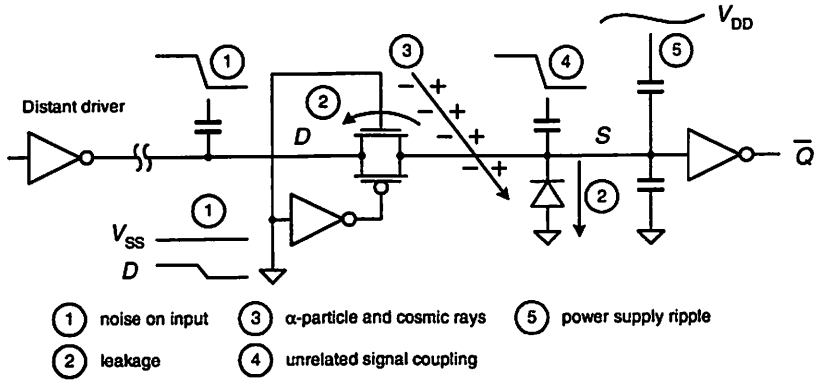


FIGURE 5.10 Sources of noise affecting latch state node. [Partovi, SSTC Workshop '00]

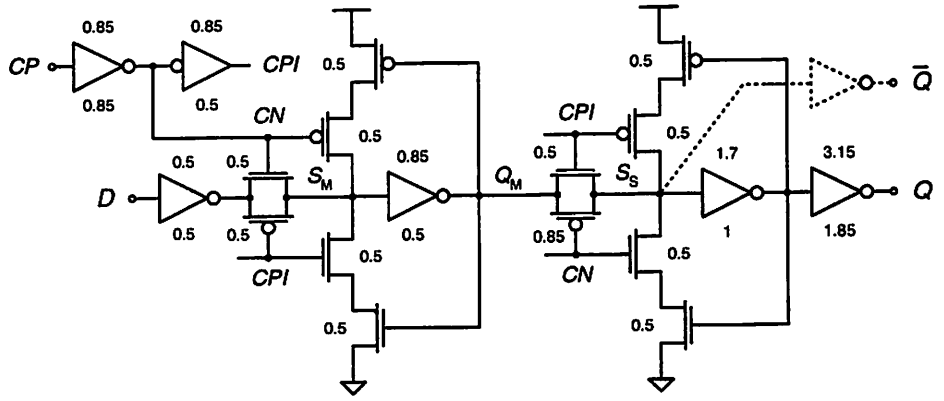


FIGURE 5.11 Transmission gate flip-flop with input gate isolation. (6-stage realization)

5.1.2.1 Circuit Operation

The flip-flop operates in standard master-slave fashion similar to the 4-stage TGFF shown in Fig. 5.3. It is also a positive edge triggered.

5.1.2.2 Sizing

The active circuit responsible for the flip-flop performance is illustrated in Fig. 5.12. Compared to the circuit in Fig. 5.4, there is an extra inverter at the output of the circuit in Fig. 5.12. Insertion of an inverter into the path does not change the path's *logical effort* which is in this case $G = 1$.

The off-path capacitance driven by the second stage is equal to the gate capacitance of two minimum width feedback transistors, $C_{\text{off-path}} = 0.74 \cdot C_{\text{min}}$. Assuming the size of the shaded inverter x times larger then a minimum sized inverter, the total *branching effort* is:

$$B = b_2 = \frac{x + 0.74}{x} \tag{5.8}$$

The total path effort F, and optimal stage effort $\hat{f}$, are given by Equations (5.9)-(5.10).

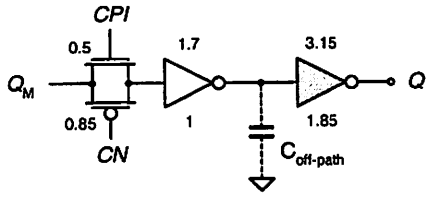


FIGURE 5.12 Active circuit that limits clock-to-output delay of the gate isolated TGFF.

$$F = G \cdot B \cdot H = 8 \cdot \frac{x + 0.74}{x} \tag{5.9}$$

$$\hat{f} = \sqrt[3]{F} = \sqrt[3]{8 \cdot \frac{x + 0.74}{x}} \tag{5.10}$$

The input capacitance of the 3rd stage is determined as:

$$C_{in_3} = \frac{g_3 \cdot C_{out}}{\hat{f}} = \frac{8}{\sqrt[3]{8 \cdot \frac{x + 0.74}{x}}} = x \cdot C_{\min} \tag{5.11}$$

Solving Equations (5.10)-(5.11) yields $x = 3.7$ and $\hat{f} = 2.16$.

Therefore, $(W_p/W_n)_3 = 3.15\mu/1.85\mu$. Applying the capacitance transformation backwards, the size of the 2nd stage is determined by:

$$C_{in_2} = \frac{g_2 \cdot C_{out_2}}{\hat{f}} = \frac{(3.7 + 0.74) \cdot C_{\min}}{2.16} = 2.06 \cdot C_{\min} \tag{5.12}$$

and is equal to $(W_p/W_n)_2 = 1.7\mu/1.0\mu$.

All other gates along the path are sized to have driving capability equivalent to a minimum sized inverter while all the feedback transistors are minimal width. Because of the intrinsically better pull-up side of the input transmission gate, the PMOS transistor of the input inverter and the PMOS transistor of the input transmission gate are made weaker (minimal width) to save some circuit area. This reduction impairs the circuit noise immunity, but this degradation is not considered as critical parameter because the fully static circuit design is assumed.

5.1.2.3 Characterization

Timing Parameters

Timing parameters of the transmission-gate flip-flop with input gate isolation (6-stage TGFF) are summarized in Table 5.4. Like a 4-stage TGFF, this flip-flop can tolerate relatively large clock skew because of its good internal race immunity, which makes it suitable for larger designs with large clock skew. The

TABLE 5.4 TIMING PARAMETERS OF 6-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	$t_{C\text{-}Q,0\text{-}1}$ [ps]	$t_{C\text{-}Q,1\text{-}0}$ [ps]	$t_{setup,0}$ [ps]	$t_{setup,1}$ [ps]	$t_{hold,0}$ [ps]	$t_{hold,1}$ [ps]	D [ps]	R [ps]
1.0	1242	1123	400	390	-210	-110	1694	1233
1.2	767	703	270	260	-140	-80	1065	783
1.8	373	353	150	140	-80	-50	531	403
2.5	258	250	110	100	-50	-40	373	290

6-stage TGFF exhibits positive setup and negative hold time. More detailed results are listed in Appendix A.2.

Energy Parameters

TABLE 5.5 BREAKDOWN OF ENERGY-PER-TRANSITION IN 6-STAGE TGFF, C_{OUT} = 4SL

V_{DD} [V]		1.0	1.2	1.8	2.5
$E_{0\text{-}0}$ [fJ]	E_{CLK} [fJ]	12.5	18.4	45.5	102
$E_{0\text{-}1}$ [fJ]	E_{CLK} [fJ]	12.5	18.4	45.5	102
	E_{int} [fJ]	22.7	32.3	81.2	163.6
	E_{ext} [fJ]	30	43.2	97.2	187.5
$E_{1\text{-}0}$ [fJ]	E_{CLK} [fJ]	12.5	18.2	45.4	101
	E_{int} [fJ]	25.2	37.1	89.5	187.1
$E_{1\text{-}1}$ [fJ]	E_{CLK} [fJ]	12.5	18.2	45.4	101

The energy consumed in internal flip-flop nodes in the 6-stage TGFF is higher than in the 4-stage TGFF because of the additional stages and increased total clocked capacitance. For details, refer to Fig. A.2.2 in Appendix A.2.

Interface Parameters

Clock slope: Variation of timing parameters with clock slope is given in Table 5.6. Figure 5.13 shows variation of timing parameters in 6-stage TGFF over various clock slopes. Since tradeoff between *Clk-Q* delay and setup time in

TABLE 5.6 IMPACT OF CLOCK SLOPE ON TIMING OF 6-STAGE TGFF. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]
0.2	490	-150	1577	1283
0.6	370	-50	1593	1334
1.2	250	70	1630	1384
2.5	160	240	1762	1466

D ($t_{CLK\text{-}Q}$ ↑, t_{setup} ↓) and tradeoff between *Clk-Q* delay and hold time in R ($t_{CLK\text{-}Q}$ ↑, t_{hold} ↑) change, D and R exhibit very small variation with clock slope.

Output load: Variation of *Clk-Q* delay and rise and fall time with output load is shown in Fig. 5.14. Figure 5.14.a shows that *0-1* and *1-0 Clk-Q* delays, which are generally not equal, do not necessarily have to be equal for a certain output load.

Physical parameters: Input capacitances of D and CP inputs are $C_{in}(D)$ = 1.96fF and $C_{in}(CP)$ = 2.45fF. Total gate width of this flip-flop is 20.15μm with all transistors minimum length, L_{min} = 0.5μm.

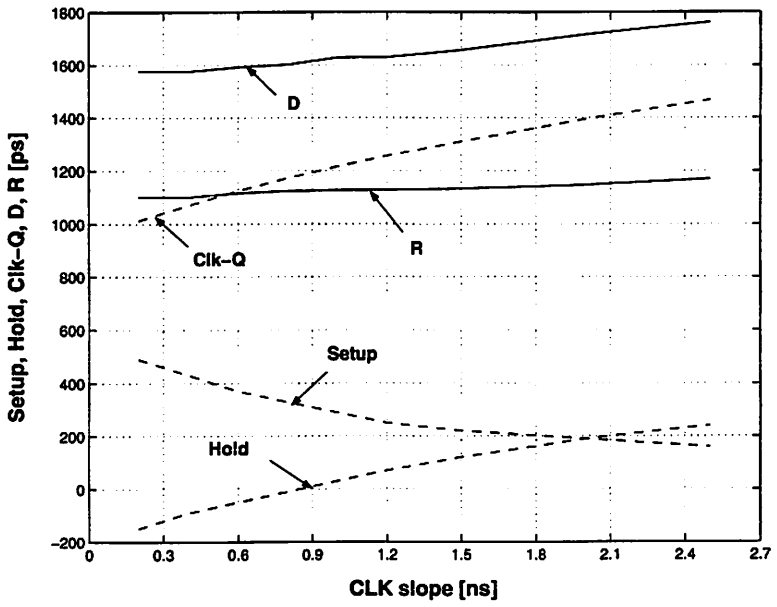


FIGURE 5.13 Variation of timing parameters in 6-stage TGFF with clock slope. Data slope is equal to the slope of the output waveform in a FO4 inverter. V_{DD} = 1V, C_{out} = 4SL.

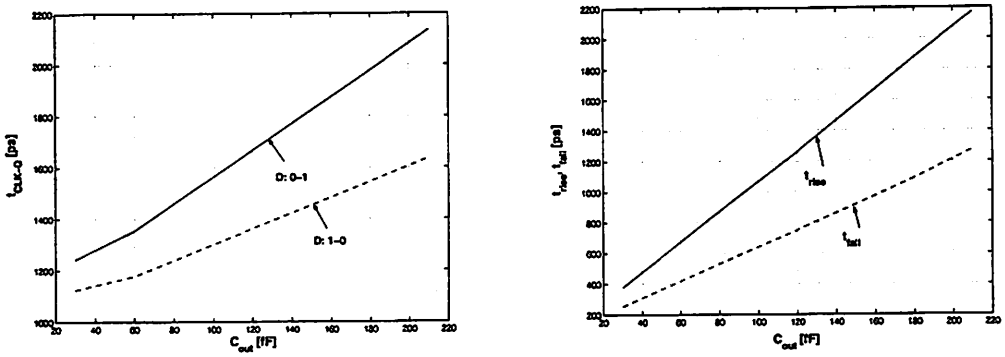


FIGURE 5.14 Variation of (a) rising and falling *Clk-Q* delay, (b) output rise/fall time, with output load. $V_{DD} = 1V$.

5.1.3 C²MOS Flip-Flop

The C²MOS flip-flop is based on a C²MOS latch shown in Fig. 5.15.b. The C²MOS latch is obtained from the transmission gate latch in Fig. 5.15.a when the wire that connects drains of the top PMOS and bottom NMOS transistors is removed from the transmission-gate latch. The pseudo-static C²MOS flip-flop [Suzuki73] of Fig. 5.16 is obtained by the addition of a weak C²MOS feedback at the outputs of the master and the slave latches in a dynamic C²MOS flip-flop. The C²MOS flip-flop of Fig. 5.16 is positive edge triggered.

5.1.3.1 Circuit Operation

A C²MOS flip-flop operates in standard master-slave fashion. It is insensitive to overlap of *CN* and *CPI*, as long as the rise and fall times of *CN* and *CPI* are sufficiently small [Rabaey96], which is ensured by local clock generation.

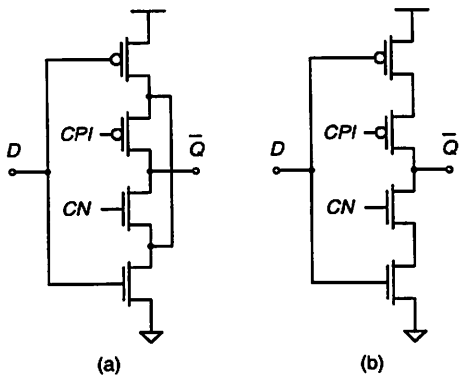


FIGURE 5.15 Gate isolated dynamic latches: (a) transmission-gate, (b) C²MOS.

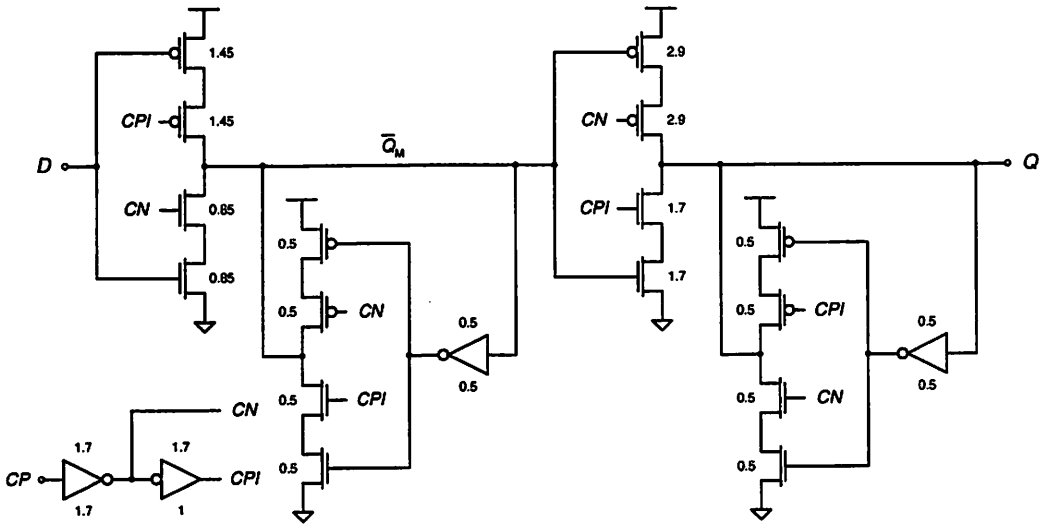


FIGURE 5.16 C^2MOS flip-flop.

5.1.3.2 Sizing

The active circuit that limits performance of the C^2MOS flip-flop reduces to the slave latch which is single stage with unknown input capacitance (size), so the method of logical effort cannot be used to resolve sizing. The slave latch has better driving capability if wider transistors are used at the cost of increased clocking energy. Therefore, one needs to find the best compromise between good driving strength of the output stage and small energy consumption in clocked nodes. Area optimization of this circuit is carried out the following way:

1. The driving capability of both master and slave latches are kept the same, W times stronger than a minimum sized inverter while all feedback transistors are minimal width. Parameter W is swept and the W_{opt} that offers minimal EDP is found, as shown in Fig 5.17.

2. From W_{opt}, the input capacitance of the slave latch $C_{in\text{-}slave}$, is obtained. The master stage is then scaled down by $k = 4SL/C_{in\text{-}slave}$ which reduces total circuit area and balances the driving capabilities of master and slave stages.

This procedure gave $W_{opt} = 3.4$, so size of the transistors in the slave stage is $(W_p/W_n)_{slave} = 2.9\mu/1.7\mu$. The input capacitance of the slave stage is then $3.4C_{min} + 0.74C_{min} = 4.14C_{min}$, where the second term is the input capacitance of the feedback inverter. Total output capacitance is $8C_{min} + 0.74C_{min} = 8.74C_{min}$.

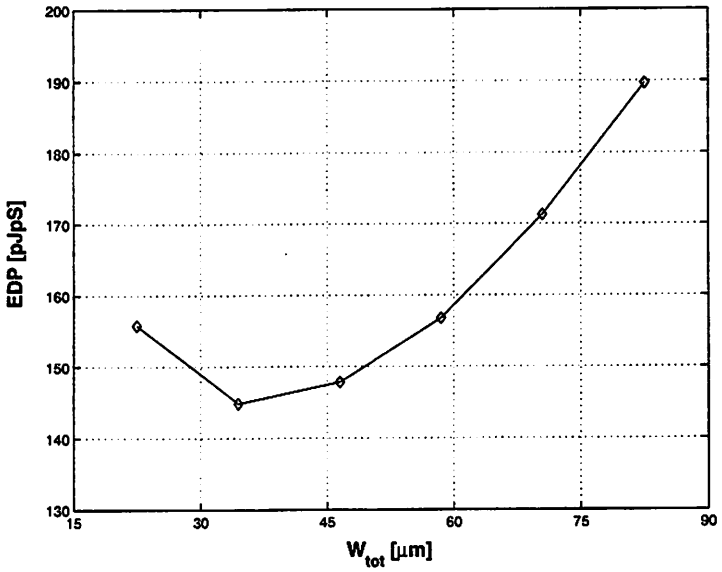


FIGURE 5.17 Sizing of C²MOS flip-flop for minimal EDP, V_{DD} = 1V.

Therefore, the size of the master stage is 8.74/4.14 = 2.1 times smaller than the size the slave stage, $(W_p/W_n)_{master}$ = 1.45µ/0.85µ. This reduction does not affect delay and achieves about 10% reduction in flip-flop energy consumption.

5.1.3.3 Characterization

Timing Parameters

Timing parameters of the C²MOS flip-flop are summarized in Table 5.7. For more details refer to Appendix A.3. Logic "0" has longer setup time because the pull-up side of the master stage conducts one inverter delay longer than the pull-down side (*CPI* goes "high" one inverter delay after *CN* became "low").

TABLE 5.7 TIMING PARAMETERS OF C²MOS FLIP-FLOP. C_{OUT} = 4SL

V_{DD} [V]	$t_{C\text{-}Q,0\text{-}1}$ [ps]	$t_{C\text{-}Q,1\text{-}0}$ [ps]	$t_{setup,0}$ [ps]	$t_{setup,1}$ [ps]	$t_{hold,0}$ [ps]	$t_{hold,1}$ [ps]	D [ps]	R [ps]
1.0	868	758	600	450	-150	-30	1396	908
1.2	529	467	390	310	-110	-20	880	549
1.8	303	275	210	180	-80	-20	529	355
2.5	195	185	150	130	-50	-10	364	235

Energy Parameters

TABLE 5.8 BREAKDOWN OF ENERGY-PER-TRANSITION IN C²MOS FF. C_{OUT} = 4SL

V_{DD} [V]		1.0	1.2	1.8	2.5
$E_{0\text{-}0}$ [fJ]	E_{CLK} [fJ]	24.1	35.9	89.2	197
$E_{0\text{-}1}$ [fJ]	E_{CLK} [fJ]	24.1	35.9	89.2	197
	E_{int} [fJ]	18.5	26.9	67.2	138.5
	E_{ext} [fJ]	30	43.2	97.2	187.5
$E_{1\text{-}0}$ [fJ]	E_{CLK} [fJ]	24.5	35.7	89.4	201
	E_{int} [fJ]	20.4	29.9	73.1	147.6
$E_{1\text{-}1}$ [fJ]	E_{CLK} [fJ]	24.5	35.7	89.4	201

Interface Parameters

Clock slope: Impact of the clock slope on flip-flop timing parameters is summarized in Table 5.9. Similarly to the 6-stage TGFF, *D* and *R* show very little variation with clock slope.

Output load: Impact of output load on flip-flop timing parameters is summarized in Table 5.10. The *Clk-Q* delay is increased because of an increase in the output rise and fall times. The hold time depends on the slope of input *D*, so increase in the *Clk-Q* delay does not imply an improvement in the flip-flop internal race immunity unless the slope of input waveform is smaller than the slope of the output waveform. For more details about the variation of the *Clk-Q* delay over a range of supply voltages and output loads, refer to Fig. A.3.2 in Appendix A.3

TABLE 5.9 IMPACT OF CLK SLOPE ON TIMING OF C²MOS FF. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]
0.2	710	-90	1320	731
0.6	570	10	1301	786
1.2	430	130	1325	818
2.5	240	340	1474	835

TABLE 5.10 TIMING PARAMETERS OF C^2MOS FLIP-FLOP. V_{DD} = 1V

C_{out} [fF]	t_{rise} [ps]	t_{fall} [ps]	$t_{C\text{-}Q,0\text{-}1}$ [ps]	$t_{C\text{-}Q,1\text{-}0}$ [ps]
30	931	506	868	758
60	1627	860	1197	904
120	3005	1575	1945	1293
210	5051	2639	3064	1875

Physical parameters: Input capacitances of *D* and *CP* inputs are $C_{in}(D) = 3.04fF$ and $C_{in}(CP) = 4.91fF$. Total gate width of this flip-flop is 25.9μm with all transistors minimum length, $L_{min} = 0.5\mu m$.

5.1.4 Summary of Master-Slave Latch Pairs

The flip-flop selection for good energy-performance tradeoff and robust timing is dictated by the flip-flop delay, internal race immunity and energy consumption as well as noise robustness.

Figure 5.18 shows the comparison of timing and energy parameters in the 6-stage TGFF and C^2MOS flip-flop. The 4-stage TGFF is out of competition because of its poor noise immunity due to charge injection problems in the master stage. Figure 5.18.a shows that the TGFF has about 40% better internal race immunity and about 20% worse delay. This clearly favors the TGFF. Figure 5.17.b shows that the 6-stage TGFF is more energy efficient than the C^2MOS flip-flop for all types of input transitions, especially *0-0* and *1-1* which show that

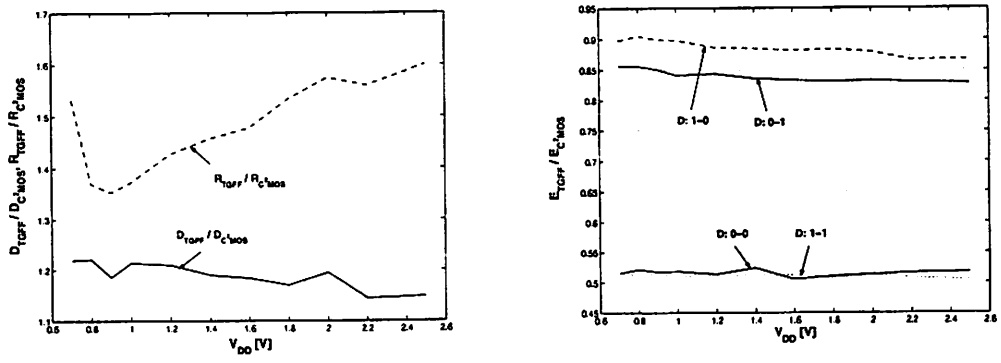


FIGURE 5.18 Comparison of (a) delay and internal race immunity, and (b) energy efficiency of master-slave flip-flops.

TABLE 5.11 PHYSICAL PARAMETERS OF MASTER-SLAVE LATCH PAIRS

	C_{in}(D) [fF]	C_{in}(CP) [fF]	W_{tot} [μm]
4s-TGFF	6.2	2.45	15.65
6s-TGFF	1.96	2.45	20.15
C^2MOS	3.04	4.91	25.9

the clocking energy of the 6-stage TGFF is half the clocking energy of the C^2MOS flip-flop.

Table 5.11 summarizes physical parameters of the TGFFs and C^2MOS flip-flop and shows that the 6-stage TGFF has lower C_{in}(D) and C_{in}(CP) meaning a further energy reduction in the circuitry that drive *D* and *CP*. It is interesting that the 4-stage TGFF has C_{in}(D) higher than the 6-stage TGFF because the input of the TG at its input, when *on*, "sees" the capacitance of the master latch state node.

Among the master-slave latch-pair topologies presented in Section 5.1, the TGFF with input gate isolation is the best choice for low energy digital design due to its good energy-delay tradeoff, race margin and noise robustness.

5.2 PULSE-TRIGGERED LATCHES

This section describes design and analysis of representative pulse-triggered latch topologies with single-input single-output, and differential-input differential-output.

5.2.1 Semi-Dynamic Flip-Flop (SDFF)

The SDFF is shown in Fig. 5.19. It is a single-input single-output, positive edge triggered *inverting* flip-flop.

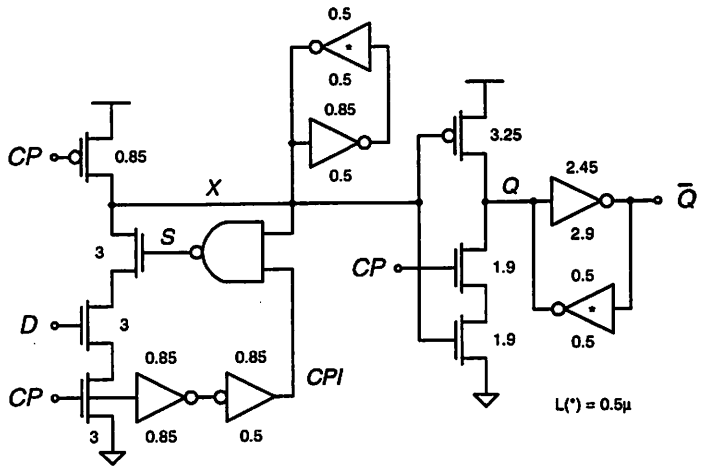


FIGURE 5.19 Semi-dynamic flip-flop (SDFF).

5.2.1.1 Circuit Operation

The semi-dynamic flip-flop [Klass98] of Fig. 5.19 is composed of a dynamic front-end and a static back-end. When *CP* is "low" the flip-flop is in the *precharge phase*. Node *X* is precharged "high" and node *Q* is decoupled from the precharge phase. The output inverters hold the previous value of *Q* and $\overline{Q}$. At the rising edge of *CP* the *evaluation phase* begins. If *D* is "low", *X* would remain "high", if *D* is high *X* would be pulled-down, driving $\overline{Q}$ "high". Three gate delays after the rising edge of *CP*, the output of the NAND gate goes "low", preventing discharge of node *X* by subsequent *0-1* transitions on *D*. This narrow sampling window makes this flip-flop appear edge-triggered.

5.2.1.2 Sizing

The *0-1* and *1-0* input transitions propagate along different paths to the output as illustrated in the active circuits of Fig. 5.20.

The 1st stage of the circuit in Fig. 5.20.a pulls down against feedback, and a *0-1* transition on *D* propagates through 3 stages as opposed to a *1-0* transition which propagates through 2 stages. Because of the additional stage in the circuit of Fig. 5.20.a and a "fight" against a weak inverter, the 1st stage is sized to be two times stronger than a minimum sized inverter, which is equivalent to a 1µm wide pull-down NMOS. Since there are 3 stacked transistors, each is 3µm wide.

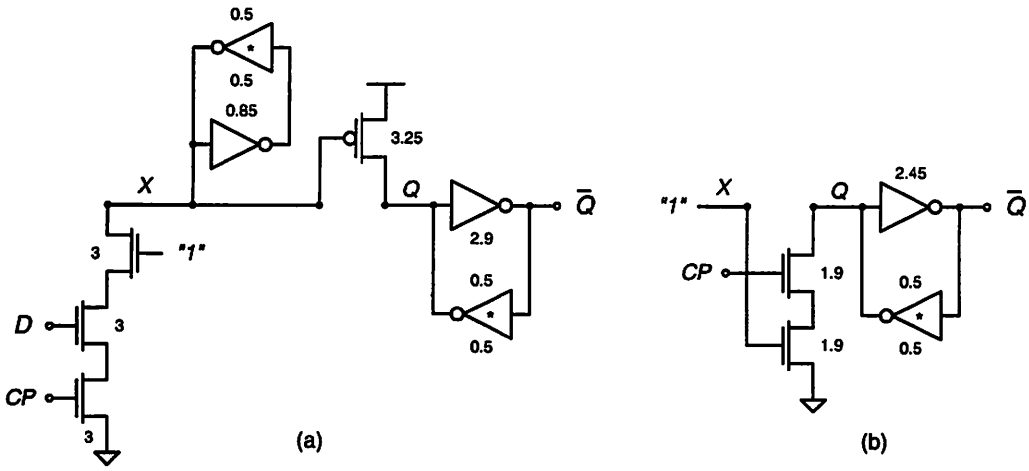


FIGURE 5.20 Active circuit in SDFF responsible for (a) *0-1*, (b) *1-0* input transition.

The pull-up (precharge) of the 1st stage is assumed to be *2x* weaker than the pull-down because the pull-up propagates to the output through smaller number of stages, so the precharge PMOS device is 0.85μm wide. This gives C_{in}(CP) of about $3.5C_{min}$ neglecting the load on *CP* from the 2nd stage.

Accounting for small branching at the output of the 1st stage, the total path effort, F, is about 3.5 which gives an optimal stage effort of $\hat{f} = 1.52$. This then determines the size of the output stage to be 5.8*x* stronger than a minimum sized inverter, which is $(W_p/W_n)_{out} = 4.9\mu/2.9\mu$. Since the rising transition at the input propagates through two stages to the output, while the falling transition at the input propagates through 2 stages to the output, the width of the pull-up PMOS transistor is reduced by *2x* and is equal to 2.45μm as depicted in Fig. 5.20.b.

From the size of the 3rd stage, applying the capacitance transformation backwards, one can derive that size of the 2nd stage has to be *3.8x* bigger than a minimum sized inverter, which is equivalent to an inverter of size $(W_p/W_n) = 3.25\mu/1.9\mu$. The pull-down side is made *2x* weaker because a *1-0* input transition propagates faster, so each of the two stacked NMOS devices is 1.9μ wide.

5.2.1.3 Characterization

Important timing and energy parameters are summarized in Table 5.12.

TABLE 5.12 TIMING AND ENERGY PARAMETERS OF SDFF. C_{OUT} = 4SL

V_{DD} [V]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]	E_{0-0} [fJ]	E_{0-1} [fJ]	E_{1-0} [fJ]	E_{1-1} [fJ]
1.0	0	590	796	168	18.1	79.4	65.0	69.2
1.2	20	410	551	96	26.7	119	94.5	105
1.8	50	210	310	38	67.9	307	226	267
2.5	70	130	237	29	167	730	485	671

E_{1-1} is higher than E_{0-0} because of precharge/discharge of node X. This flip-flop is a representative example of unequal paths of *0-1* and *1-0* input transitions, but the delays can be balanced for a range of output loads. The rising and falling transitions have equal delay for C_{out} = 60fF as shown in Fig. 5.21.

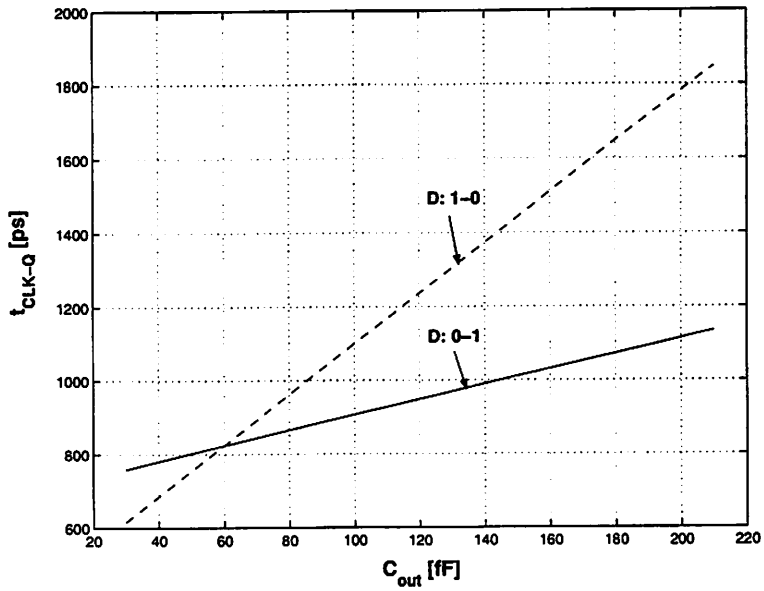


FIGURE 5.21 Impact of output load on rising and falling *Clk-Q* delays in SDFF. V_{DD} = 1V.

5.2.2 Hybrid Latch-Flip-Flop (HLFF)

The HLFF [Partovi96] is shown in Fig. 5.22. It is a single-input single-output, positive edge triggered *inverting* flip-flop.

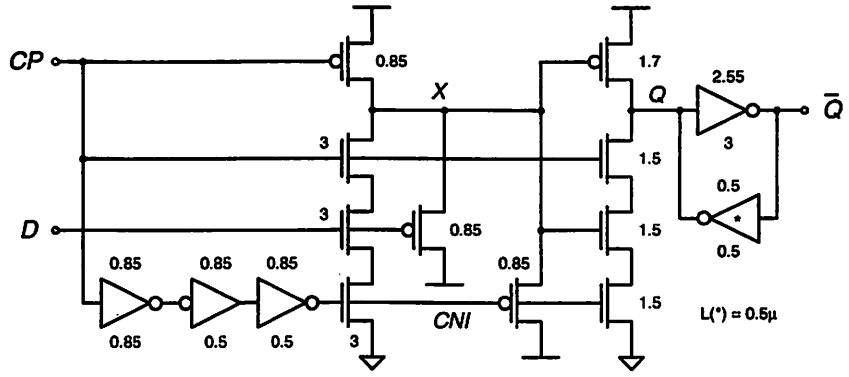


FIGURE 5.22 Hybrid latch-flip-flop (HLFF).

5.2.2.1 Circuit Operation

Prior to the rising edge of *CP*, the flip-flop is in the *precharge phase*, node *X* is precharged to "high" and the output inverters hold previously stored logical value on Q and $\overline{Q}$ which is decoupled from *X* because the 2nd stage is *off*.

At the rising edge of *CP*, the pull-down side of both the 1st and the 2nd stage is enabled for a period of time defined by the 3-inverter delay chain. During this period the latch is transparent and *D* can be sampled into the latch. Once *CNI* goes "low" node *X* is decoupled from D and is held at "high" or begins to precharge to "high" by the PMOS device that is driven by *CNI*.

At the falling edge of *CP*, node *X* is precharged to "high" by the PMOS transistor that is driven by *CP* and *X* remains "high" as long as *CP* stays "low".

5.2.2.2 Sizing

The circuits shown in Fig. 5.23 limit the performance of HLFF. These circuits are very similar to the circuits of Fig. 5.20 so is the sizing methodology. The pull-down of the 1st stage is *2x* stronger than the pull-up of the 1st stage because the pull-down propagates to the output through 3 stages, while the pull-up propagates to the output through 2 stages. Because of this mismatch, the pull-down of the flip-flop's 2nd stage (shown as the 1st stage in Fig. 5.23.b) and the pull-up of the 3rd stage are 2*x* weaker than the pull-up of the 2nd stage and pull-down of the 3rd stage, respectively. The PMOS transistors that precharge node *X* have equivalent driving strength as a unit inverter, and are 0.85µm wide.

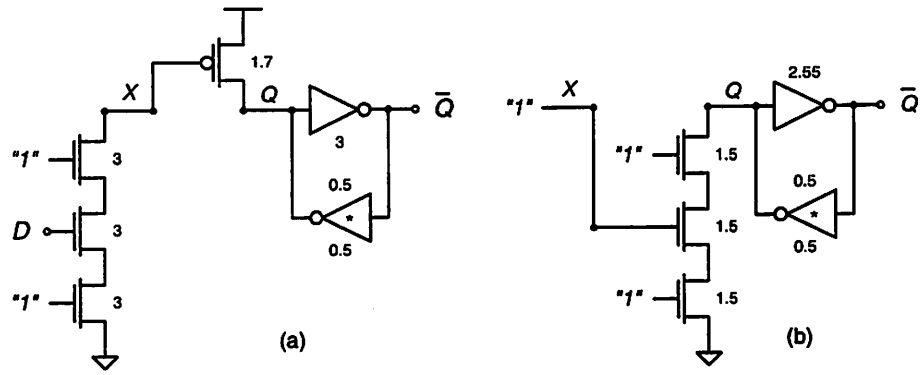


FIGURE 5.23 Active circuit in HLFF responsible for (a) *0-1*, (b) *1-0* input transition.

5.2.2.3 Characterization

The important timing and energy parameters at several supply voltages are summarized in Table 5.13.

TABLE 5.13 TIMING AND ENERGY PARAMETERS OF HLFF. C_{OUT} = 4SL

V_{DD} [V]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]	E_{0-0} [fJ]	E_{0-1} [fJ]	E_{1-0} [fJ]	E_{1-1} [fJ]
1.0	70	460	751	188	29.9	101	89.4	94.2
1.2	80	340	551	109	44.9	151	132	144
1.8	150	180	389	47.7	114	373	320	374
2.5	210	180	358	-39.3	262	814	666	840

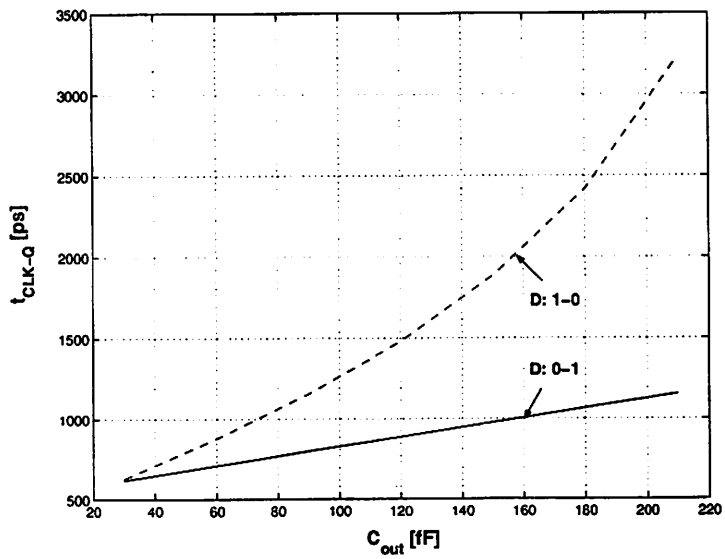


FIGURE 5.24 Rising and falling *Clk-Q* delays in HLFF as a function of output load, C_{out},. V_{DD} = 1V.

This flip-flop has equal rising and falling *Clk-Q* delays when driving $C_{out} = 4SL$, whereas these delays are imbalanced at higher C_{out}, as shown in Fig. 5.24. Similarly to SDFF, it also has higher $E_{1\text{-}1}$ than $E_{0\text{-}0}$ due to precharge/discharge operation on node *X*.

5.2.3 Differential-Input Differential-Output Flip-Flops

This section presents two sense amplifier based flip-flops.

5.2.3.1 Sense Amplifier Based Flip-Flop (SAFF)

The SAFF [Matsui94] is shown in Fig. 5.25. It is differential-input differential-output, positive edge triggered flip-flop.

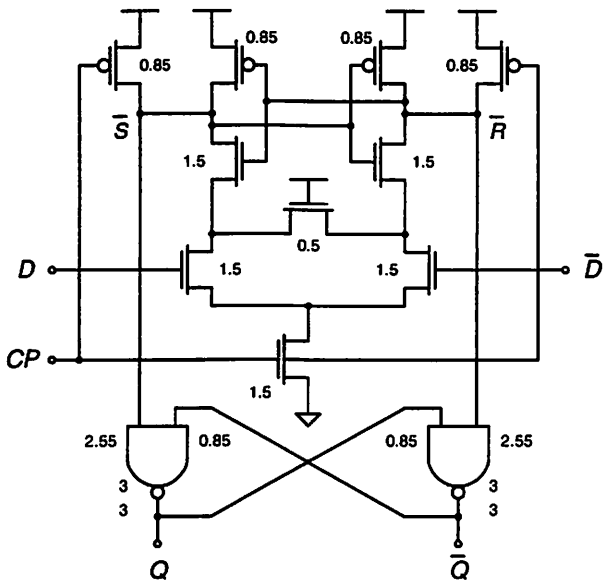


FIGURE 5.25 Sense amplifier based flip-flop (SAFF).

Circuit Operation

This flip-flop operates in precharge/evaluate mode. When *CP* is "low" the flip-flop is in the *precharge phase*, the input differential-pair is *off* and the cross-coupled NAND gates (*S-R* latch) in the output stage hold previously stored logic value at Q and $\overline{Q}$.

At the rising edge of *CP*, the differential-pair senses complementary inputs, *D* and $\overline{D}$, generates pulses at $\overline{S}$ and $\overline{R}$, and the *S-R* latch captures logic values of the differential inputs and holds them until the next rising edge of *CP*.

Sizing

The pull-down of the 1st stage is sized for the driving capability of a unit inverter. The keeper PMOS transistors in the *S-R* latch are unit size, 0.85μm wide. The output capacitance therefore is $8.67C_{min}$. The input capacitance of the active circuit is $1.1C_{min}$ so the total path effort is $F = 7.9$. The optimal stage effort is $\hat{f} = 2.8$ which gives required driving strength of the 2nd stage to be $3x$ better than a unit inverter.

Characterization

The important timing and energy parameters are summarized in Table 5.14.

TABLE 5.14 TIMING AND ENERGY PARAMETERS OF SAFF. C_{OUT} = 4SL

V_{DD} [V]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]	$E_{0\text{-}0}$ [fJ]	$E_{0\text{-}1}$ [fJ]	$E_{1\text{-}0}$ [fJ]	$E_{1\text{-}1}$ [fJ]
1.0	-20	710	1247	240	21.7	77.7	69.7	23.5
1.2	10	520	844	129	32.3	114	101	35.0
1.8	70	280	486	58	79.7	272	240	88.6
2.5	90	180	370	51	179	566	499	196

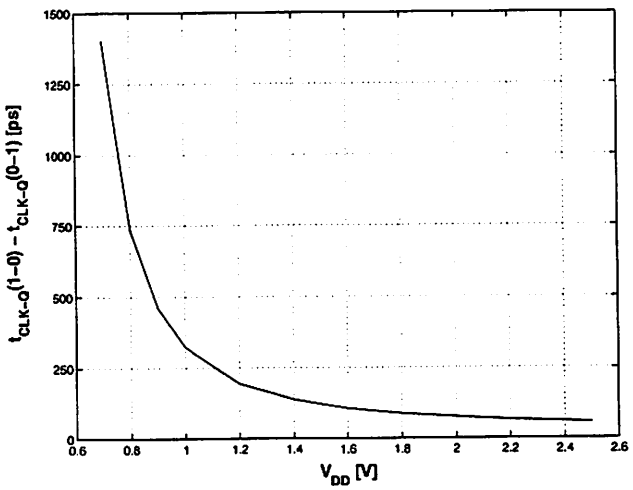


FIGURE 5.26 Variation of mismatch in *Clk-Q* delay with supply voltage in SAFF.

This flip-flop has imbalanced rising and falling *Clk-Q* delays – *1-0* transition on *D* takes longer to propagate to *Q* than a *0-1* transition as shown in Fig. 5.26, because of an additional delay through the NAND circuit. Likewise SDFF and HLFF, the SAFF has relatively small delay but also a small race margin.

5.2.3.2 Modified SAFF

The mismatch between rising and falling *Clk-Q* delays of the SAFF is solved in the modified SAFF [Nikolic99] where the output stage is implemented in pass-transistor circuit technique instead of complementary CMOS. There is an overhead of two minimum sized inverters that generate signals *S* and *R* from $\overline{S}$ and $\overline{R}$ in the circuit of Fig. 5.27, while all other complementary signals are freely available, which makes the pass transistor technique affordable in terms of energy efficiency.

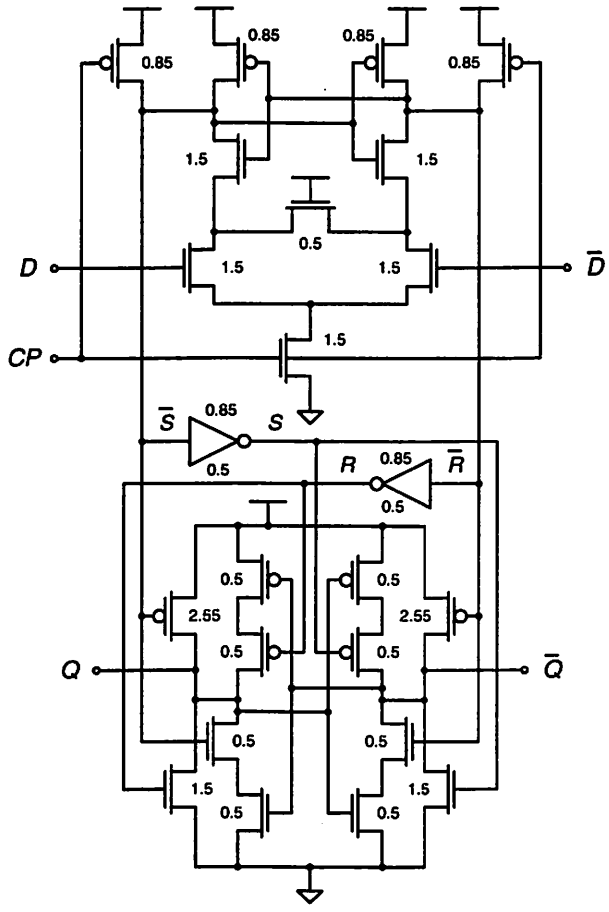


FIGURE 5.27 Modified sense amplifier based flip-flop.

Circuit Operation

The overall circuit operation is identical to the operation of SAFF. Inputs $\overline{S}$ and $\overline{R}$ are set and reset inputs of the *S-R* latch, respectively. The "low" level at both of these inputs is not allowed which is ensured by the sense amplifier.

Sizing

The same sizing equations as in the case of SAFF hold here, the only difference being the implementation of the output stage. The keeper transistors in the output stage have minimum width.

Characterization

The important timing and energy parameters of the modified SAFF are summarized in Table 5.15.

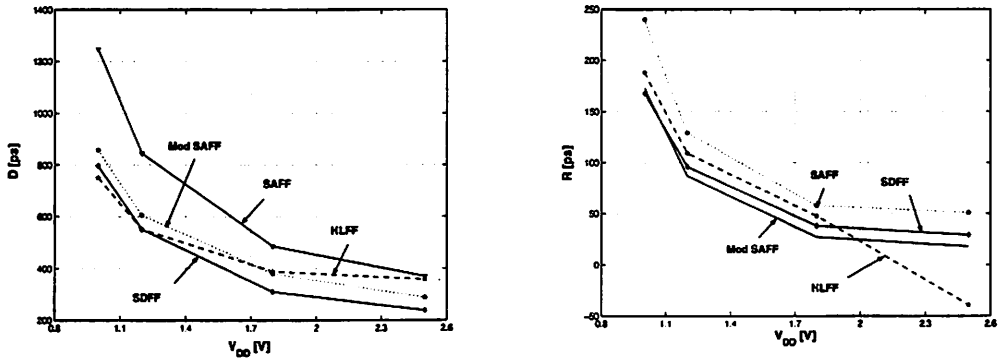
This flip-flop has balanced timing and energy parameters due to its fully differential structure and fully balanced output stage.

TABLE 5.15 TIMING AND ENERGY PARAMETERS OF MODIFIED SAFF. C_{OUT} = 4SL

V_{DD} [V]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]	$E_{0\text{-}0}$ [fJ]	$E_{0\text{-}1}$ [fJ]	$E_{1\text{-}0}$ [fJ]	$E_{1\text{-}1}$ [fJ]
1.0	-20	670	856	173	26.6	67.8	67.9	27.6
1.2	20	480	605	87	39.8	99.2	99.5	41.2
1.8	80	260	381	27	99.4	234	235	104
2.5	100	170	287	18	229	493	499	237

5.2.3 Summary of Pulse-Triggered Latches

Figure 5.28 presents comparison of delay and internal race immunity of pulse-triggered flip-flops. All of these flip-flops show very small race margin that imposes tight constraints on allowable clock skew. Aggressive clock skew requirement requires careful clock distribution and de-skewing circuits leading to an increased energy consumed in the clock distribution network. The energy



(a) (b)

FIGURE 5.28 Timing parameters of pulse-triggered latches: (a) delay, (b) internal race immunity.

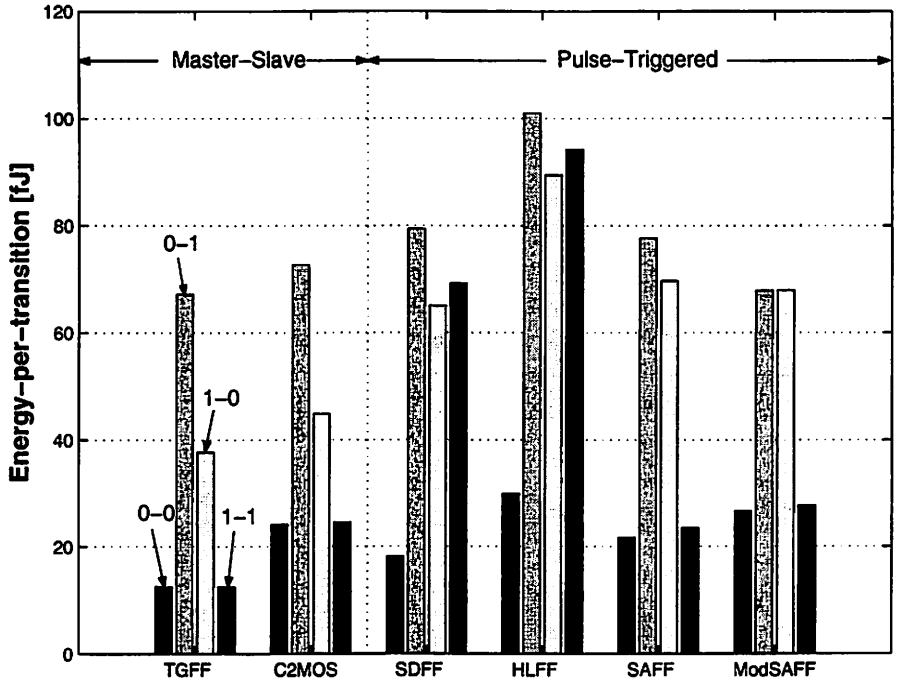


FIGURE 5.29 Comparison of energy-per-transition in master-slave and pulse-triggered flip-flops. V_{DD} = 1V, C_{out} = 4SL.

consumed in flip-flops themselves is substantially higher in the master-slave latch pairs as illustrated in Fig. 5.29.

Table 5.16 summarizes physical parameters of the pulse-triggered latches.

Better internal race immunity and lower energy consumption of master-slave latch pairs favor them over the pulse-triggered latches at the expense of little increase in delay.

TABLE 5.16 PHYSICAL PARAMETERS OF PULSE-TRIGGERED LATCHES

	C_{in}(D) [fF]	C_{in}(CP) [fF]	W_{tot} [µm]
SDFF	4.2	10	33.2
HLFF	4	7.2	28.7
SAFF	2.1	4	30.2
Mod SAFF	2.1	4	26.2

5.3 FLIP-FLOPS WITH REDUCED SWING CLOCK

A reduced swing clock flip-flop technique targets the energy savings in the clocking of a flip-flop. This section presents two flip-flops with reduced swing clock operation: a pulse-triggered latch, and a master slave latch pair.

5.3.1 Reduced Clock-Swing Flip-Flop (RCSFF)

The RCSFF [Sakurai98] is shown in Fig. 5.30. It is derived from the SAFF aiming the reduction in the clocking energy due to reduced swing clock operation. The RCSFF is a positive edge triggered flip-flop.

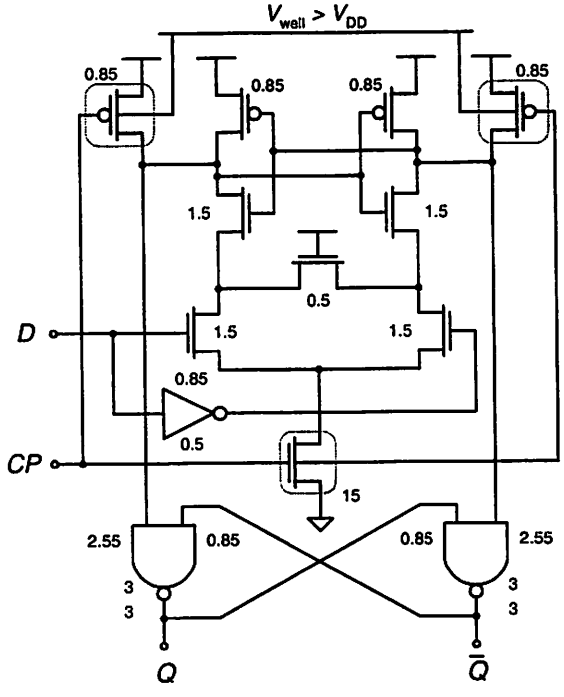


FIGURE 5.30 Reduced clock-swing flip-flop (RCSFF).

5.3.1.1 Circuit Operation

Operational behavior of RCSFF is very similar to the behavior of SAFF. The differences are in the reduction of the clock swing and in the biasing of the precharge PMOS transistors in the RCSFF. The reduction in the clock swing leads to an increase in leakage current of the precharge PMOS transistors in RCSFF so their wells have to be tied to a voltage higher than supply.

The RCSFF requires a reduced swing clock driver. The driver can be a specific circuit that converts rail-to-rail swing at its input to a lower swing at the output [Sakurai98], or it can be a simple buffer powered up by an on-chip dc-dc converter or an external voltage supply.

5.3.1.2 Sizing

Circuit sizing is similar to the SAFF. The width of the NMOS transistor in the tail current source of differential-pair is determined by the acceptable variation in the *Clk-Q* delay as a function of the clock swing [Sakurai98]. All other transistors are of identical size in the SAFF.

5.3.1.3 Characterization

Comparison with conventional flip-flop: Main feature of the RCSFF reported in [Sakurai98] is its reduced clocking energy; however, there need to exist appropriate reference circuit for comparison to support this claim. In [Sakurai98], comparison of the RCSFF with the conventional transmission-gate based flip-flop is presented. It is hard to make fair comparison of these two circuits because they are structurally different. In addition, there is a large mismatch in circuit areas in favor of the RCSFF. Its total transistor gate width is 48μm, as opposed to 90μm in the conventional flip-flop. More importantly from the clocking energy standpoint, total transistor gate width of the clocked transistors in the RCSFF is 11μm, while the total transistor gate width in the conventional flip-flop is 37.5μm, Figure 5.31. This clearly favors the RCSFF over the conventional flip-flop and makes comparison of these two circuits inappropriate.

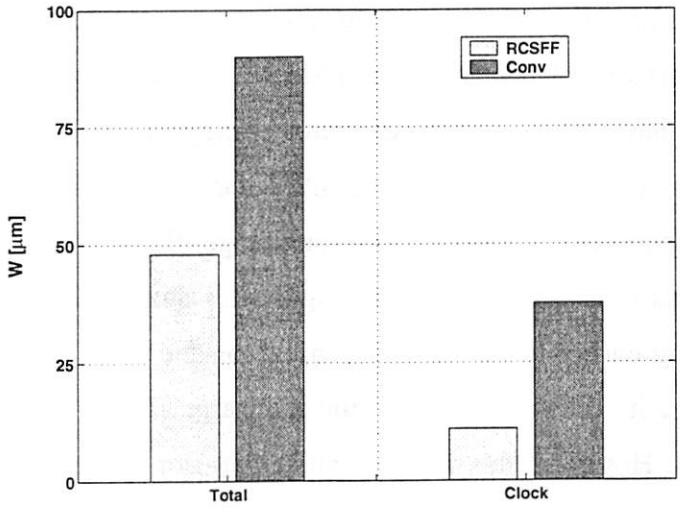
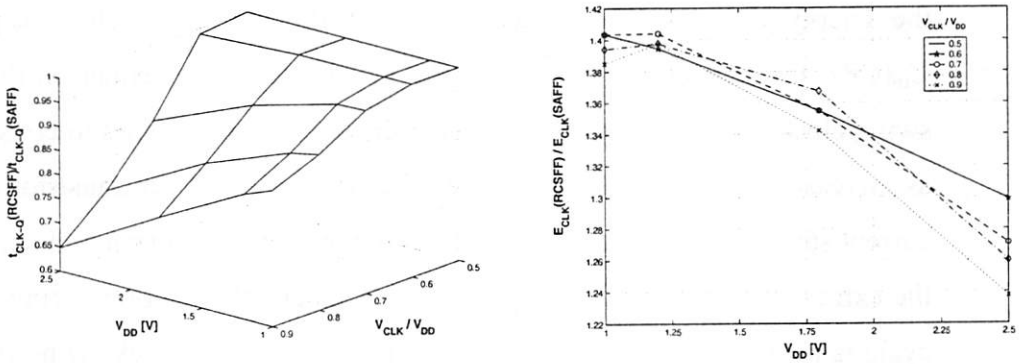


FIGURE 5.31 Illustration of mismatch in total transistor width and total width of clocked transistors in RCSFF and conventional flip-flop.



(a) (b)

FIGURE 5.32 Comparison of RCSFF and SAFF: (a) performance (ratio of delays is rounded to 1), (b) clocking energy. The well of the precharge PMOS transistors of the RCSFF is biased with $V_{WELL} = 2V_{DD}$.

Comparison with SAFF: More appropriate benchmark circuit for the evaluation of energy saving capabilities of the RCSFF is the SAFF which is structurally identical to the RCSFF. The RCSFF has larger area due to a wider NMOS transistor in the tail current source of differential-pair that is necessary to accommodate for the performance loss arising from its lower swing clock. This increased size of the NMOS transistor does not come without cost because it compromises flip-flop performance and the clocking energy. For example, if the NMOS transistor in the tail current source is made $10x$ wider, the RCSFF offers similar performance as SAFF for clock swing of about $0.6V_{DD}$ as shown in

Fig. 5.32.a. Since wider transistor provides larger bias current for the differential-pair, this directly results in an increase of the clocking energy. Our simulation results have shown that the clocking energy of the RCSFF is up to 40% higher than the clocking energy of the SAFF for the same performance, Fig. 5.32.b. In this experiment, the wells of the precharge PMOS transistors are tied to $2V_{DD}$ to suppress their leakage current. Fig. 5.32.b shows also that the overhead in the clocking energy is almost independent on the clock signal swing because of an increase in leakage current of the precharge PMOS transistors at reduced clock swings. However, this impairment in flip-flop internal clocking energy may be overcome in the reduced swing clock distribution network.

The internal flip-flop clocking energy is estimated as the total flip-flop energy consumption when input data undergoes a *0-0* transition. This does not include the energy consumed in switching of the flip-flop input clock capacitance, C_{in}(CP). The true energy savings of the RCSFF therefore remain in the reduced swing clock distribution network. However, there are still issues to consider, such as increased C_{in}(CP) arising from the increase of the NMOS transistor of the tail current source, and routing of an extra wire for the reduced swing clock. Ignoring the extra wiring problem, energy required to clock one flip-flop during one clock cycle is about 2 times higher in the RCSFF with $0.6V_{DD}$ clock swing than in the rail-to-rail clocked SAFF. This comes to the only remaining option for energy saving – saving in the internal nodes of clock distribution. Given the energy loss in the flip-flop itself (40%) and the energy loss in clocking capacitance of the flip-flop clock input (2x), it is hard to believe that all this can be overcome by the energy reduction only in internal nodes of the clock distribution network. To summarize, the RCSFF does not seem to be a viable candidate for energy efficient clock subsystem design.

5.3.2 Low Power Flip-Flop

Figure 5.33 shows the static random access memory (SRAM) type low power flip-flop [Landers98]. It is a negative edge triggered flip-flop.

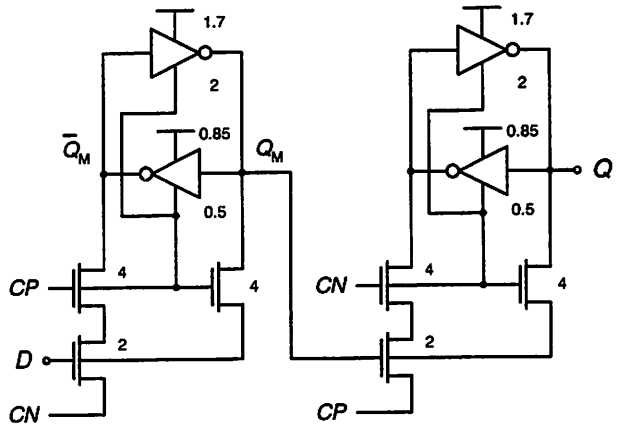


FIGURE 5.33 Low power flip-flop. [US Patent 5,789,956]

5.3.2.1 Circuit Operation

The clock signals *CP* and *CN* operate with reduced swing ($V_{DD}/2$) and are coupled into ground terminals of the cross-coupled inverters in the master and slave stage. This provides a negative gate to source voltage, thereby ensuring full turn off of the inverter transistors during their *off* periods. When *CP* goes "high", the signal on the data line is fed to one side of the master latch and the other side of the latch is coupled to ground or reference voltage.

When *CP* signal goes "low", *CN* concurrently goes "high" and the signal latched into the master latch is fed to the slave latch. The slave latch is identical to the master latch except that it operates at the opposite phase of the clock. The signal stored in the slave latch is the output of the flip-flop.

5.3.2.2 Sizing

The circuits responsible for pulling the output *Q* up and down are shown in Fig. 5.34. They are sized to be two times stronger than a unit inverter. Transistors driven with half-swing signals at their gates are $2x$ wider to provide sufficient current. The clocking scheme of this flip-flop makes the circuit sizing difficult because the pass signals that drive the slave latch are not V_{DD} or ground as in master-slave latches introduced thus far. The pass signals are generated by the master stage instead, with the entire pass transistor chain terminating to ground terminal of the clock buffer that generates *CP*. For example, the output of the

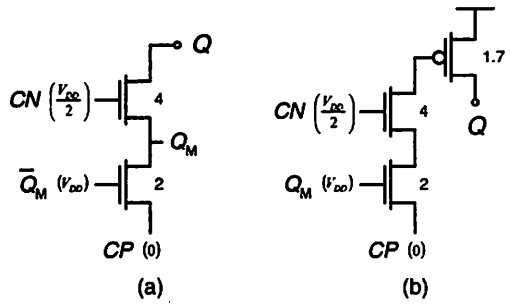


FIGURE 5.34 Active circuit in LPFF responsible for (a) pull-down, (b) pull-up operation.

transistor that pulls down the output Q is tied to the output of the master latch Q_M and terminates to ground through another two pass transistors. This coupling back to the previous stage makes cascading of the LPFFs difficult and for that reason the master and slave stages are made identical.

5.3.2.3 Characterization

The operation waveforms of the LPFF are shown in Fig. 5.35. When *CP* is "low" there exists a "fight" between the transistors M_1 and M_2 as illustrated in Fig. 5.36. During the "fight" there exists static current flow adding up to the energy consumption. To avoid the "fight," M_1 has to be much wider than M_2, but on the other hand M_2 has to be wide to provide sufficient pull-up on Q.

Figure 5.36 also explains that there is a limit in scaling of supply voltage – gate-to-source overdrive of M_1 and M_2 gets smaller with reduction in supply voltage because their gates are driven by the half-swing signals. For the best performance, it is desirable that ΔV_1 be large and ΔV_2 be small but unfortunately

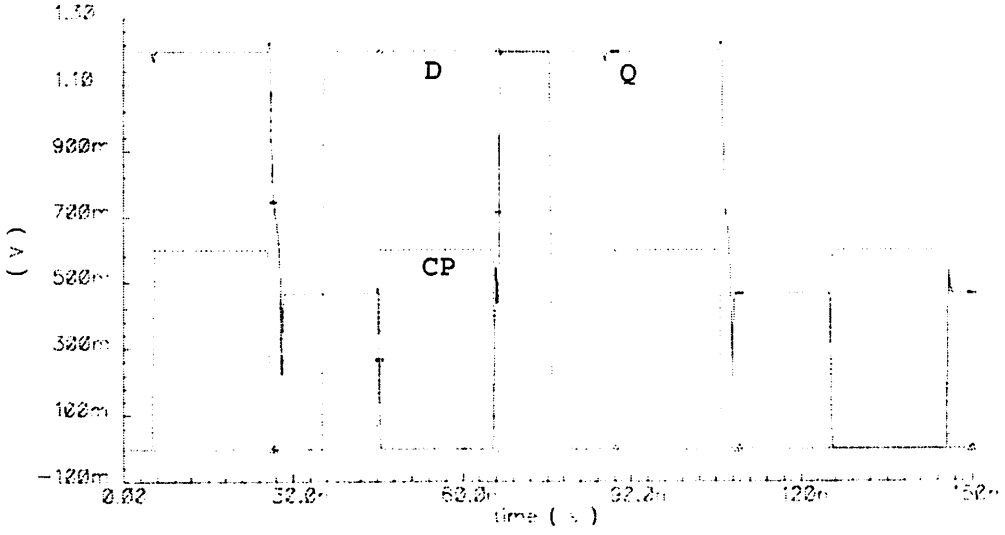


FIGURE 5.35 Operation waveforms of LPFF. V_{DD} = 1.2V, V_{DD}(CP) = 0.6V, C_{out} = 4SL.

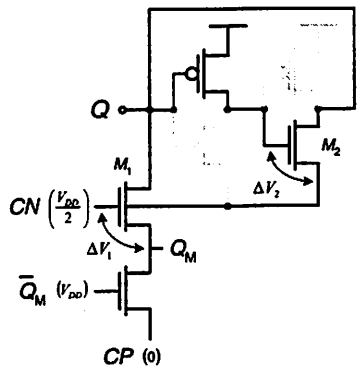


FIGURE 5.36 Illustration of a "fight" between pull-down (M_1) and pull-up (M_2) during the pull-down operation on Q.

they are nearly equal. The pull-down delay gets worse than the pull-up delay at reduced supplies as shown in Fig. 5.37.

Therefore, the LPFF has very low performance at low supply voltages, and relatively high energy consumption at high supply voltages because the static component of energy become pronounced.

5.3.3. Summary of Flip-Flops with Reduced Swing Clock

Table 5.17 summarizes the physical parameters of the RCSFF and LPFF. The RCSFF have shown inefficient in terms of performance-energy tradeoff compared to the SAFF. The LPFF exhibits timing problems and an increase in energy due to static current during the pull-down operation.

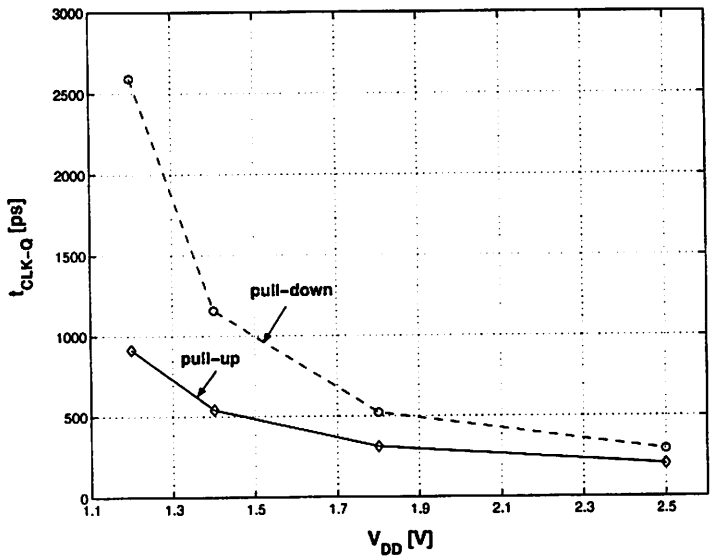


FIGURE 5.37 Degradation of pull-down delay of the LPFF at reduced supply voltages.

TABLE 5.17 PHYSICAL PARAMETERS OF FFS WITH REDUCED CLK SWING

	C_{in}(D) [fF]	C_{in}(CP) [fF]	W_{tot} [μm]
RCSFF	4.1	23.3	45.05
LPFF	10.1	4.9	30.1

Reduced swing clock flip-flops themselves have worse energy efficiency than conventional, full swing clock, flip-flops. Sustaining same performance requirement with the reduced swing clock operation typically results in increased input capacitance of the flip-flop clock node, reducing the capability of energy savings in the clock tree.

5.4 FLIP-FLOPS WITH INTERNAL CLOCK GATING

The main feature of this flip-flop family is a mechanism for predictive turn *off* the internal clock when input and output data are equal. Since there exists an overhead associated with local clock gating, the use of these flip-flops is justified for low switching activities of the input data. The internal clock gating technique can be applied to all circuits introduced thus far.

5.4.1 Data-Transition Look-Ahead D Flip-Flop (DL-DFF)

The DL-DFF [Nogawa98] is shown in Fig. 5.38. It is a positive edge triggered, non-inverting flip-flop.

5.4.1.1 Circuit Operation

The DL-DFF shown in Fig. 5.38 is derived from a conventional flip-flop that is similar to the 6-stage TGFF shown in Fig. 5.11. The circuits enclosed with dashed lines show overhead associated with internal clock gating.

The *data-transition look-ahead* (DL) circuit performs an XNOR function on D and Q. When $D = Q$, the DL circuit produces a logic "0" at P_1 and generation

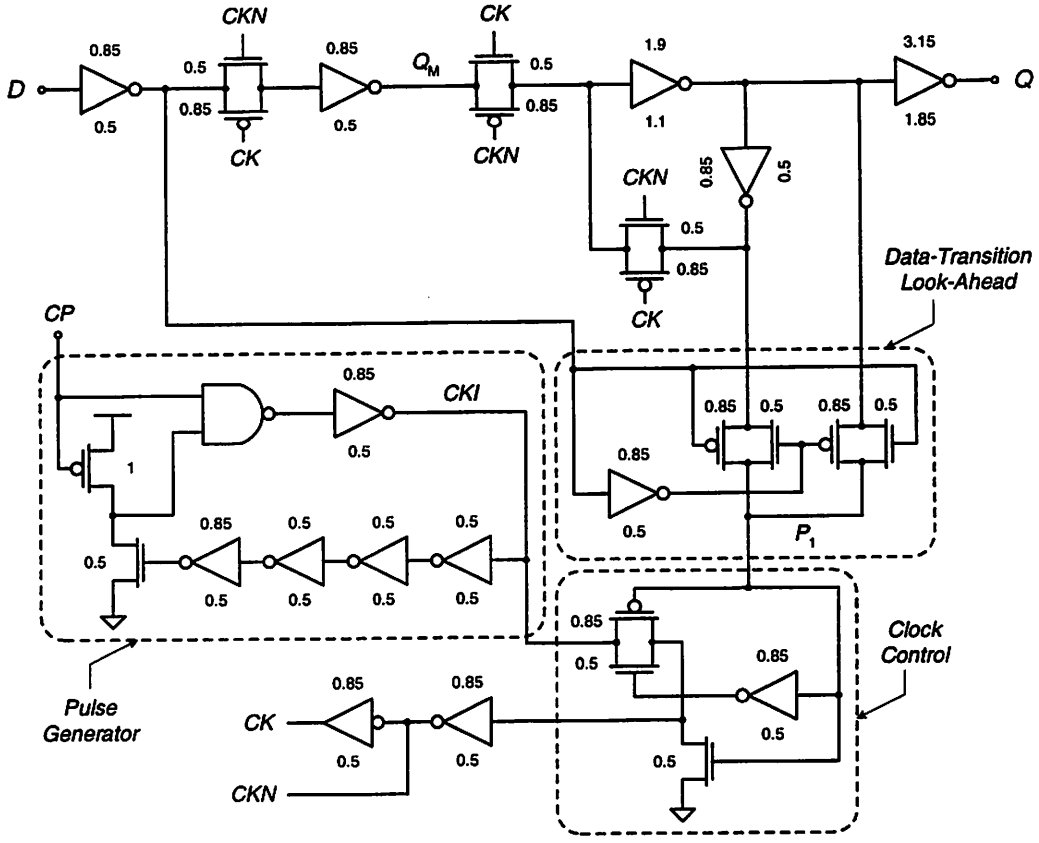


FIGURE 5.38 Data-transition look-ahead D flip-flop (DL-DFF).

of the internal clock {*CK, CKN*} is disabled. When $D \neq Q$, P_1 is "high" and the clock control (CC) circuit enables generation of {*CK, CKN*}.

The *pulse generator* (PG) circuit generates a short pulse, *CKI*, at every rising edge of the external clock, *CP*. *CKI* then triggers the flip-flop if $D \neq Q$. If there was no pulse generator, this flip-flop could be triggered by data instead of the clock. For example, if $D \neq Q$ and the rising edge of *CKI* arrives, then the clock pulse, *CK*, is generated and Q changes. However, if D changes again while the clock is still high and becomes different from Q, then another pulse of internal clock, *CK*, would be generated and the flip-flop would be actually triggered by the data.

5.4.1.2 Sizing

The active circuit responsible for output transitions is shown in Fig. 5.39. The circuit shown in Fig. 5.39 is identical to the circuit in Fig. 5.11, so the same equations apply as in case of 6-stage TGFF. The only difference is that now C_{off}-

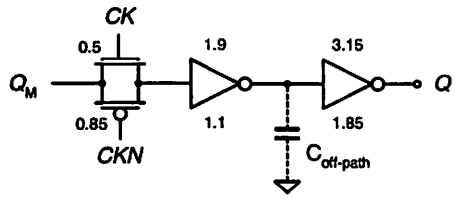


FIGURE 5.39 Active circuit that limits performance of the DL-DFF.

$_{path} = C_{min}$. Solving equations (5.8)-(5.12) with new value of $C_{off\text{-}path}$ gives $x = 3.7$ and $\hat{f} = 2.17$. The sizes of the 2nd and 3rd stages are then $(W_p/W_n)_3$ = 3.15μ/1.85μ, and $(W_p/W_n)_2$ = 1.9μ/1.1μ.

Devices in DL, pulse generator, and clock gating circuits are chosen to be minimum sized in order to minimize the energy penalty associated with clock gating. All the transistors of 2-input NAND gate in a DL circuit are 1μ wide, and the NAND gate is implemented in complementary CMOS.

5.4.1.3 Characterization

Timing Parameters

Timing and energy parameters of the DL-DFF are given in Table 5.18. Although the input of the DL-DFF is similar to the input of the 6-stage TGFF, the setup time of both logic "0" and logic "1" in the DL-DFF is shorter than the setup time in the 6-stage TGFF because the internal clock, *CK*, in DL-DFF has an extra delay with respect to *CP* as it propagates through DL and clock control logic. The DL-DFF has *small race margin* because it is actually a pulse-triggered flip-flop. Compared to conventional pulse-triggered flip-flops, race margin of the DL-DFF reduces at reduced supply voltages because of an increase in hold time due to

TABLE 5.18 TIMING PARAMETERS OF DL-DFF. C_{OUT} = 4SL

V_{DD} [V]	$t_{C\text{-}Q,0\text{-}1}$ [ps]	$t_{C\text{-}Q,1\text{-}0}$ [ps]	$t_{setup,0}$ [ps]	$t_{setup,1}$ [ps]	$t_{hold,0}$ [ps]	$t_{hold,1}$ [ps]	D [ps]	R [ps]
1.0	1957	1880	330	320	2040	1800	2375	-160
1.2	1236	1210	240	230	1180	1070	1528	30
1.8	590	600	140	130	520	490	770	80
2.5	410	427	100	100	320	310	548	100

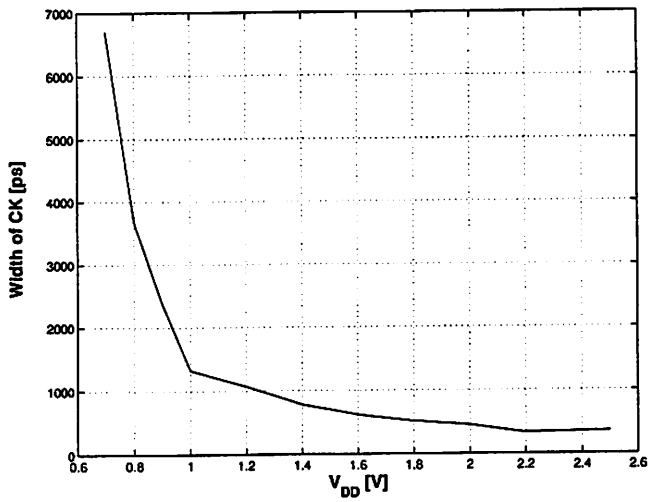


FIGURE 5.40 Width of internal clock, *CK*, of the DL-DFF as a function of supply voltage.

increase in the width of *CK* as shown in Fig. 5.40. The width of *CK* is the sum of *CK-Q* delay, delay of DL logic, and delay of CC logic. The delay of DL and CC gets worse at lower supply voltages, and the hold time is dependent on the *CK-Q* delay. As a result, the internal race immunity is worse at reduced supplies.

Energy Parameters

In flip-flops with internal clock gating, it is essential to know what the overhead is in internal clock gating circuitry and which input data activities offer better energy efficiency than conventional master-slave flip-flops. To find that out, a portion of DL-DFF circuit shown in Fig. 5.41 is analyzed.

When the energy-per-transition of the circuit in Fig. 5.41 is subtracted from the energy-per-transition of the circuit in Fig. 5.38, the energy overhead in data look-ahead (DL), clock control (CC) and pulse generator (PG) is obtained, as

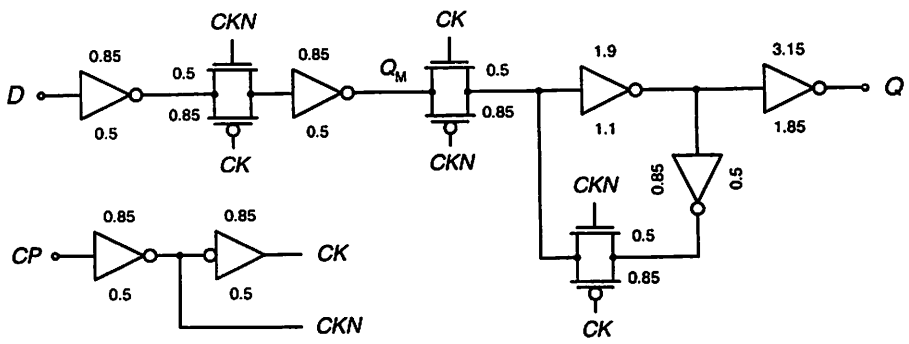


FIGURE 5.41 DL-DFF without internal clock gating.

TABLE 5.19 ENERGY OVERHEAD (PG + CC + DL) IN DL-DFF

V_{DD} [V]	1.0	1.2	1.8	2.5
$E_{0\text{-}1}$ [fJ]	58.8	86.9	210	447
$E_{1\text{-}0}$ [fJ]	57.4	85.2	209	444

TABLE 5.20 ENERGY CONSUMPTION IN PULSE GENERATOR

V_{DD} [V]	1.0	1.2	1.8	2.5
E_{PG} [fJ]	35.5	52.8	130	274
E_{Cin} [fJ]	3.5	5.0	11.3	21.9

listed in Table 5.19. This only applies to *0-1* and *1-0* input transitions, because only then are all sub-circuits in Fig. 5.38 and Fig. 5.41 active. For *0-0* and *1-1* input transitions, the internal clock (shaded inverters) is activated in the circuit of Fig. 5.41, and deactivated in the circuit in Fig. 5.38.

The energy overhead of Table 5.19 is independent of C_{out}. The PG is commonly shared among several flip-flops, so further energy breakdown is needed to understand where the energy goes exactly. Energy consumed by the PG is estimated by simulation of a stand-alone PG loaded with capacitance C_{in}(CC) that *CKI* "sees" when "looking" into the CC circuit as shown in Fig. 5.42. This capacitance is measured by simulation and found to be 3.5fF. Energy associated with the PG is summarized in Table 5.20, where E_{PG} is the energy consumed in the pulse generator itself, and E_{Cin} is the energy consumed in driving of C_{in}(CC). The PG consumes energy regardless of what input transition occurs.

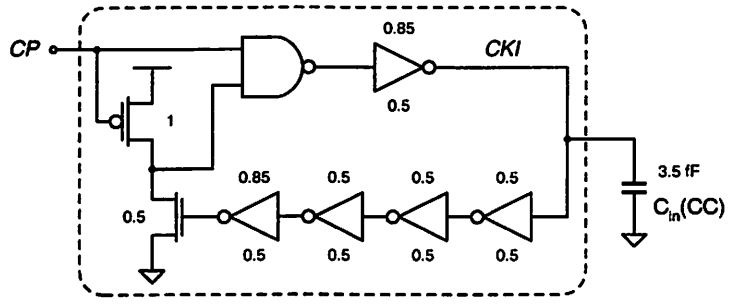


FIGURE 5.42 Estimation of energy consumption in pulse generator.

TABLE 5.21 ENERGY BREAKDOWN IN DL-DFF. C_{OUT} = 4SL

V_{DD} [V]		1.0	1.2	1.8	2.5
E_{0-0} [fJ]	E_{PG} [fJ]	35.5	52.8	130	274
	E_{Cin} [fJ]	3.5	5.0	11.3	21.9
E_{0-1} [fJ]	E_{CLK} [fJ]	11.4	16.8	41.3	91.5
	E_{int} [fJ]	22.2	32.0	78.9	174
	E_{ext} [fJ]	30	43.2	97.2	187.5
	E_{DL+CC} [fJ]	19.8	29.1	69	151
	E_{PG} [fJ]	35.5	52.8	130	274
	E_{Cin} [fJ]	3.5	5.0	11.3	21.9
E_{1-0} [fJ]	E_{CLK} [fJ]	11.4	16.8	41.5	91.3
	E_{int} [fJ]	26.1	38.2	92.1	195
	E_{DL+CC} [fJ]	18.4	27.4	67	148
	E_{PG} [fJ]	35.5	52.8	130	274
	E_{Cin} [fJ]	3.5	5.0	11.3	21.9
E_{1-1} [fJ]	E_{PG} [fJ]	35.5	52.8	130	274
	E_{Cin} [fJ]	3.5	5.0	11.3	21.9

Detailed energy breakdown in DL-DFF is given in Table 5.21. In Table 5.21, E_{CLK}(0-1,1-0) is obtained from the circuit in Fig. 5.40 (shaded inverters). For more details refer to Appendix B.1.

Analysis of Energy Efficiency of the DL-DFF

Energy saving capabilities of DL-DFF depend on two parameters: 1) number of flip-flops, N, driven by a single PG, and 2) input data transition probability. The energy consumed per flip-flop during one clock cycle, when $D = Q$ is given by:

$$E_{D-idle} = E_{0-0} = E_{1-1} = \frac{E_{PG}}{N} + E_{Cin} \tag{5.12}$$

The energy consumption when D undergoes a *0-1* or *1-0* transition is given by Equations (5.13)-(5.14).

$$E_{0-1} = E_{D-idle} + E_{CLK} + E_{DL+CC} + E_{\text{int}} + E_{ext} \tag{5.13}$$

$$E_{1-0} = E_{D-idle} + E_{CLK} + E_G + E_{\text{int}} \tag{5.14}$$

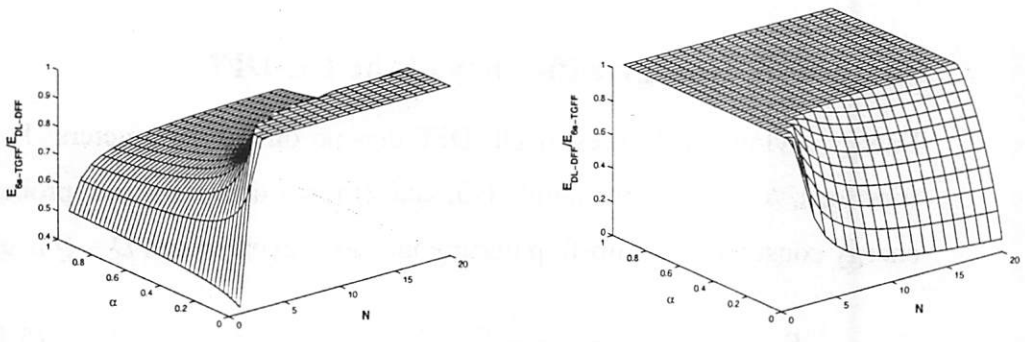
Comparison with 6-stage TGFF: Assuming there is no glitching at input D, $\alpha_{0-1} = \alpha_{1-0} = \alpha/2$, where α is data transition probability. Under this postulation, average energy consumption of the DL-DFF and the 6-stage TGFF are:

$$E_{DL-DFF} = \frac{\alpha}{2} \cdot (E_{0-1} + E_{1-0}) + \frac{1-\alpha}{2} \cdot E_{D-idle} \tag{5.15}$$

$$E_{6stage-TGFF} = \frac{\alpha}{2} \cdot (E_{0-1} + E_{1-0}) + \frac{1-\alpha}{2} \cdot E_{CLK} + E_{Cin} \tag{5.16}$$

In Equation (5.16), E_{Cin} is the energy consumed in switching C_{in}(CP) of the 6-stage TGFF. This term represents the energy consumed by a simple clock buffer that drives C_{in}(CP) and is included for a fair comparison with the DL-DFF where E_{PG}/N represents the energy consumption in PG per flip-flop.

Figure 5.43 compares energy consumption in DL-DFF and 6-stage TGFF as a function of N and α. Energy saving characteristic of the DL-DFF over the 6-stage TGFF is shown in Fig. 5.44. Figure 5.44.a shows the area in the N-α plane where the DL-DFF has better energy efficiency than the 6-stage TGFF which is $N > 2$ and $\alpha < 0.25$. Figure 5.44.b shows how much energy saving could be achieved if



(a) (b)

FIGURE 5.43 Ratio of energy consumption (rounded to 1) in DL-DFF and 6-stage TGFF, over range of switching activities and number of flip-flops driven by a single pulse generator in DL-DFF. (a) $E_{6s\text{-}TGFF}/E_{DL\text{-}DFF}$, (b) $E_{DL\text{-}DFF}/E_{6s\text{-}TGFF}$. V_{DD} = 1V.

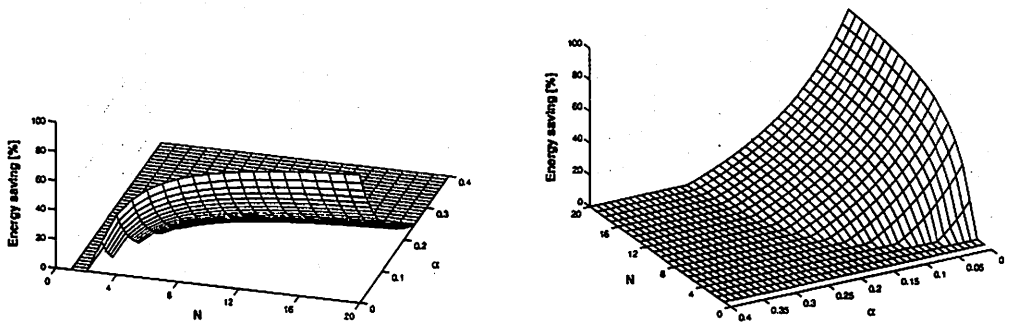


FIGURE 5.44 Energy saving characteristics of DL-DFF compared to 6-stage TGFF.

the DL-DFF is utilized.

Interface Parameters

Clock slope: Dependency of the variation of *Clk-Q* delay as function of clock slope and supply voltage is illustrated in Fig. 5.45. Figure 5.45.b shows almost no variation in the *Clk-Q* delay as a function of the clock slope and supply voltage as opposed to the 6-stage TGFF.

Output load: Figure 5.46 shows the *Clk-Q* (taken as the average of $t_{CLK\text{-}Q,0\text{-}1}$ and $t_{CLK\text{-}Q,1\text{-}0}$) delay as a function of the output load and supply voltage. Figure 5.46.b shows that the coefficients a and b in a "linear delay model" expressed in a form $y = a \cdot x + b$ do depend on the supply voltage.

Physical parameters: Input capacitance of *D* and *CP* inputs are $C_{in}(D) = 2fF$ and $C_{in}(CP) = 3.2fF$. The total gate width of this flip-flop is 34.25μm (18.8μm flip-flop, 4.05μm DL, 3.2μm clock control, 11.2μm pulse generator), with all gates minimum length, $L_{min} = 0.5\mu m$.

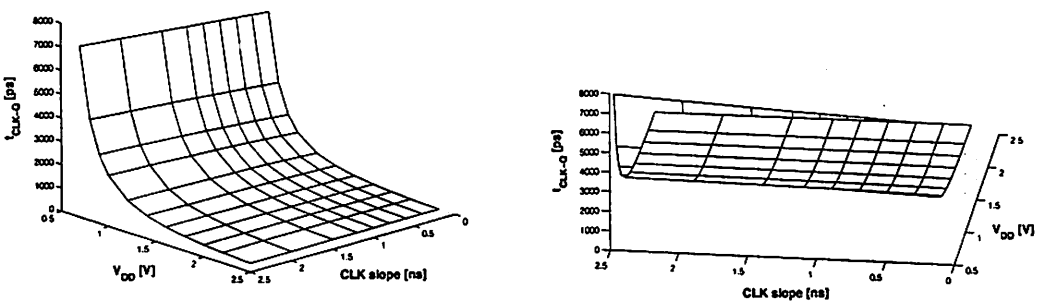
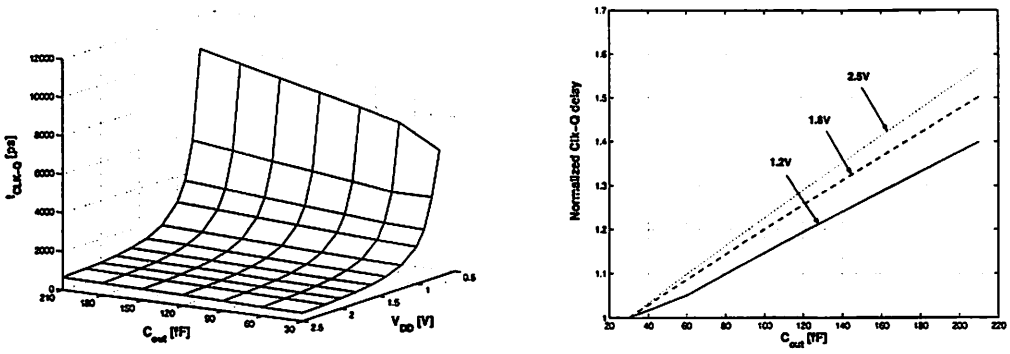


FIGURE 5.45 Dependency of *Clk-Q* delay on clock slope and supply voltage.



(a) (b)

FIGURE 5.46 (a) Variation of *Clk-Q* delay with output load and supply, (b) Variation in linearity of Clk-Q delay as a function of output load.

5.4.2 Clock-on-Demand Flip-Flop (COD-FF)

The COD-FF [Kuroda99] is shown in Fig. 5.47. It is a positive edge triggered, non-inverting flip-flop. Circuits enclosed with dashed lines show the overhead associated with pulse generation and data-transition look-ahead.

5.4.2.1 Circuit Operation

As in the DL-DFF, the *data-transition look-ahead* (DL) circuit also performs an XNOR function on D and Q. When $D = Q$, XNOR = 0 and *CKI* is disabled.

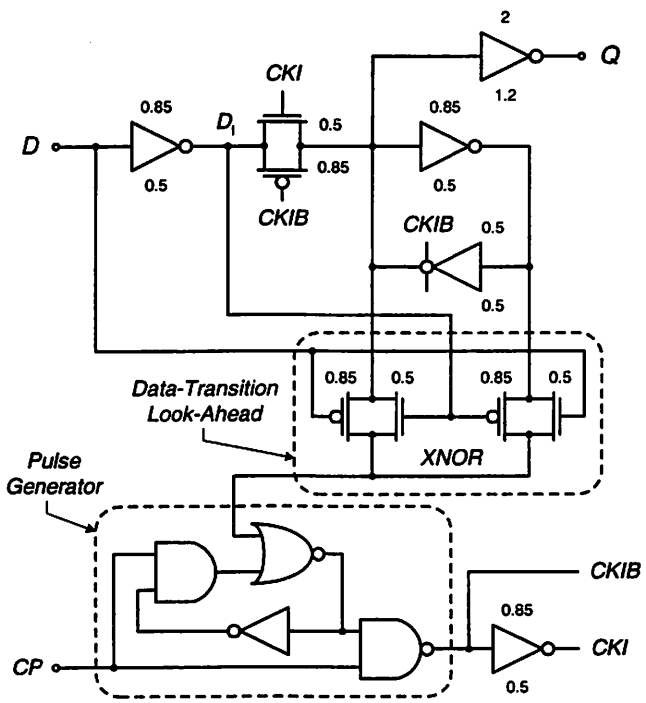


FIGURE 5.47 Clock-on-demand flip-flop (COD-FF).

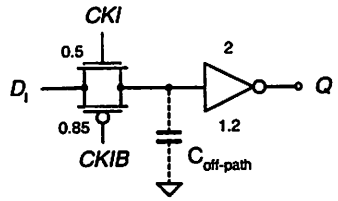


FIGURE 5.48 Active circuit that limits performance of the COD-FF.

When $D \neq Q$, pulse generator circuit (PG) generates a short pulse, *CKI*, at every rising edge of the external clock, *CP*.

Unlike the DL-DFF, the COD-FF has its local pulse generation. As pointed out in [Kuroda99], this helps avoid problems with distortion of the pulse in the clock distribution and power penalty of the pulse clock generator. The clock control function is integrated in the internal pulse generator of the COD-FF as opposed to the DL-DFF. This reduces area overhead of the COD-FF and promises better energy efficiency than in the DL-DFF.

5.4.2.2 Sizing

The active circuit responsible for output transitions is shown in Fig. 5.48. The circuit shown in Fig. 5.48 is identical to the circuit shown in Fig. 5.4, so the same equations apply as in case of the 4-stage TGFF. As in the 4-stage TGFF, $C_{\text{off-path}} = C_{\text{min}}$ and the size of the 2nd stage is $(W_p/W_n)_2 = 2\mu/1.2\mu$.

Design of the Pulse Generator: Minimizing Energy Overhead

Careful optimization of PG is key to the minimization of energy overhead associated with internal clock gating. If the PG circuit were implemented in complementary CMOS, there would be energy consumption in the PG even when *D* is idle, as illustrated in Fig. 5.49. In order to avoid this unnecessary energy

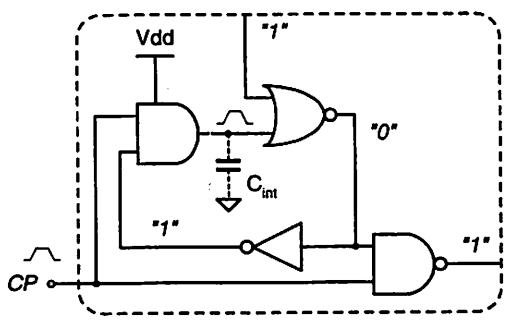


FIGURE 5.49 Illustration of energy consumption in internal nodes of PG circuit of the COD-FF when $D = Q$ (std. cell CMOS realization).

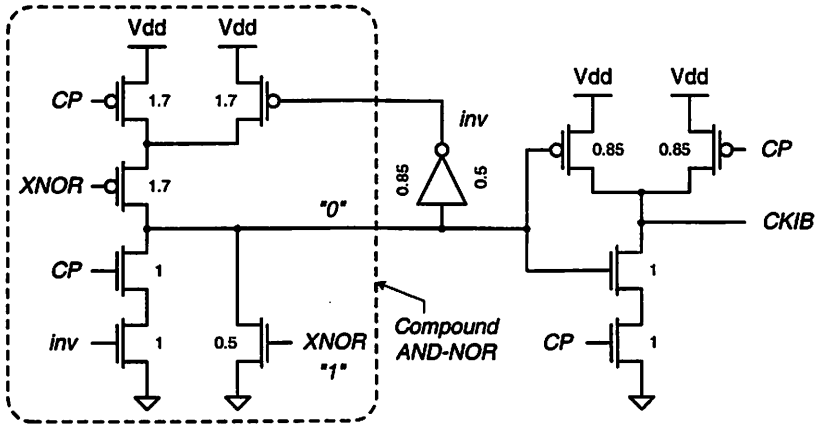


FIGURE 5.50 Energy efficient implementation of PG in COD-FF, with no energy consumption in internal nodes when $D = Q$ (realization with compound AND-NOR gate).

consumption, AND and NOR circuits are implemented as a single compound gate shown in Fig. 5.50. Logic gates are sized to have similar driving capability as a minimum sized inverter. Sizes of all devices are marked in Fig. 5.50.

5.4.2.3 Characterization

DL circuit of the DL-DFF, although logically identical to the DL circuit of the COD-FF, consumes more energy than the DL circuit of the COD-FF because of an extra inverter which generates data signal, as opposed to the COD-FF where data signal is directly wired from the input of the flip-flop. This discrepancy makes a comparison of these two circuits difficult, emphasizing the importance of interface parameters, specifically C_{in}(D). The D input of COD-FF is loaded with an additional capacitance – the input capacitance of the DL circuit – which causes logic gate that drives D to consume more energy.

Timing Parameters

TABLE 5.22 TIMING PARAMETERS OF COD-FF. C_{OUT} = 4SL

V_{DD} [V]	$t_{C\text{-}Q,0\text{-}1}$ [ps]	$t_{C\text{-}Q,1\text{-}0}$ [ps]	$t_{setup,0}$ [ps]	$t_{setup,1}$ [ps]	$t_{hold,0}$ [ps]	$t_{hold,1}$ [ps]	D [ps]	R [ps]
1.0	1216	1612	800	570	1440	1150	2493	66
1.2	779	1008	540	390	900	740	1598	39
1.8	375	463	260	210	400	330	746	45
2.5	261	314	200	150	280	230	530	31

TABLE 5.23 ENERGY BREAKDOWN IN COD-FF. C_{OUT} = 4SL

V_{DD} [V]		1.0	1.2	1.8	2.5
E_{0-0} [fJ]		0.0	0.0	0.0	0.0
E_{0-1} [fJ]	E_{CLK} [fJ]	9.1	13.4	33.8	77.1
	E_{int} [fJ]	11.3	16.6	41.3	91.5
	E_{ext} [fJ]	30.0	43.2	97.2	188
	E_{OH} [fJ]	28.6	42.1	100	207
E_{1-0} [fJ]	E_{CLK} [fJ]	9.1	13.5	34.1	77.9
	E_{int} [fJ]	12.9	19.2	49.1	112
	E_{OH} [fJ]	31.2	46.2	118	269
E_{1-1} [fJ]		0.0	0.0	0.0	0.0

The COD-FF is pass-gate based flip-flop with a pulse-triggered clocking mechanism. Therefore, it exhibits *small race margin.*

Energy Parameters

Energy overhead is examined by analysis of the COD-FF without internal clock gating as shown in Fig. 5.51. An extra (shaded) inverter is inserted into local clock generation to make clock slopes of internal signals realistic and to account for energy consumption in driving *CKIB*.

Breakdown of the energy-per-transition of the COD-FF is given in Table 5.23. For more details refer to Table B.2.4 in Appendix B.2.

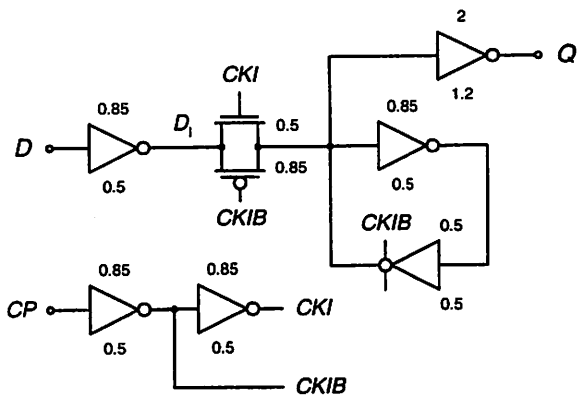
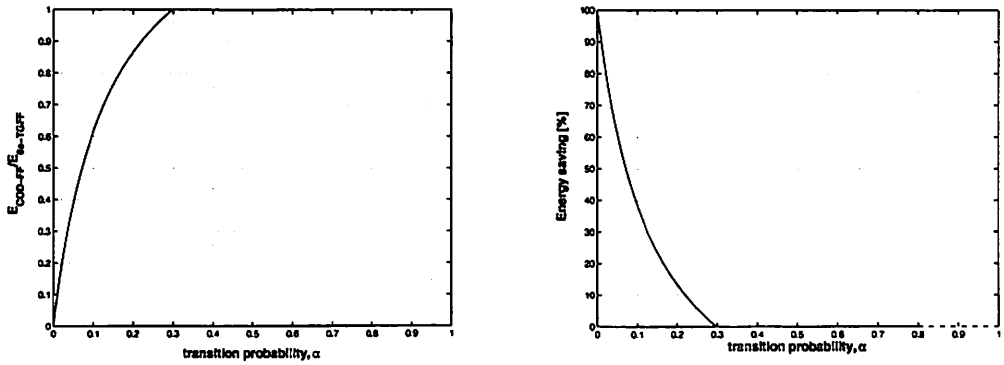


FIGURE 5.51 COD-FF without internal clock gating.



(a) (b)

FIGURE 5.52 Ratio of energy consumption (truncated to 1) in COD-FF and 6-stage TGFF as a function of switching activity: (a) $E_{COD-FF}/E_{6s-TGFF}$, (b) $E_{6s-TGFF}/E_{CODFF}$. C_{out} = 4SL, V_{DD} = 1V.

$E_{0\text{-}0}$ and $E_{1\text{-}1}$ of the circuit in Fig. 5.51 represent clocking energy for *0-1* and *1-0* input data transitions. In the table above, the E_{OH} component of $E_{0\text{-}1}$ and $E_{1\text{-}0}$ denotes the energy overhead – energy consumed by the DL and the PG. E_{PG} is the energy consumption in the pulse generator for *0-0* and *1-1* input transitions.

Analysis of Energy Efficiency of the COD-FF

Energy saving capability of DL-DFF depends on its input data transition probability.

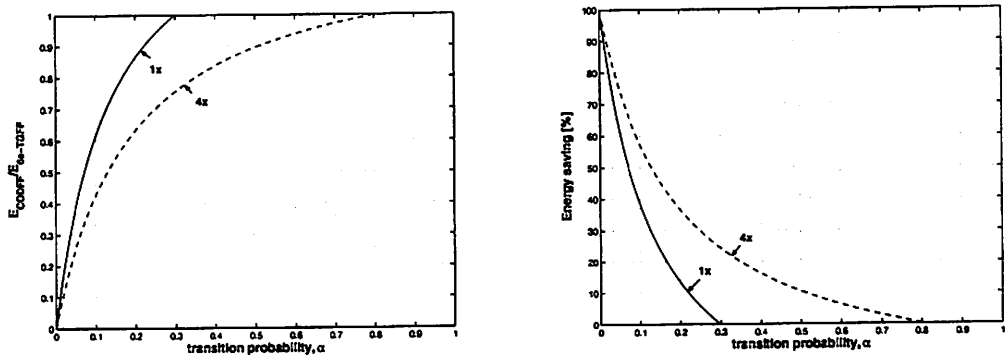
Comparison with 6-stage TGFF: Assuming no glitching at the input, $\alpha = \alpha_{0\text{-}1} = \alpha_{1\text{-}0}$ and average energy consumption of the CODFF and the 6-stage TGFF is given by Equations (5.17)-(5.18).

$$E_{COD-FF} = \frac{\alpha}{2} \cdot (E_{0-1} + E_{1-0}) + \frac{1-\alpha}{2} \cdot E_{D-idle} \qquad (5.17)$$

$$E_{6stage-TGFF} = \frac{\alpha}{2} \cdot (E_{0-1} + E_{1-0}) + \frac{1-\alpha}{2} \cdot E_{CLK} \qquad (5.18)$$

Figure 5.52.a shows that the DL-DFF is more energy efficient than the 6-stage TGFF for $\alpha < 0.3$. Figure 5.52.b shows that for $\alpha > 0.3$ conventional 6-stage TGFF consumes up to 22% less energy than the COD-FF.

The real issue in energy saving capabilities of flip-flops with internal clock gating is the understanding of the energy overhead in the internal clock gating



(a) (b)

FIGURE 5.53 Ratio of energy consumption (truncated to 1) in COD-FF and 6-stage TGFF as a function of switching activity and size of the clocked devices: (a) $E_{CODFF}/E_{6s\text{-}TGFF}$, (b) $E_{6s\text{-}TGFF}/E_{COD\text{-}FF}$. $C_{out} = 4SL$, $V_{DD} = 1V$.

circuitry compared to the clocking energy in the flip-flops. For example, if sizes of clocked transistors of both the 6-stage TGFF and the COD-FF are increased four times, then the energy saving capability of the COD-FF looks different as shown in Fig. 5.53.

Therefore, utilization of internal clock gating makes sense when sizes of the clocked transistors are large and/or for low data activity rates.

Interface Parameters

Clock slope: The *Clk-Q* delay as a function of the clock slope and supply voltage is very similar to the case of the DL-DFF. For more details refer to Fig. B.2.1 in Appendix B.2.

Output load: The COD-FF shows similar behavior to DL-DFF.

Physical parameters: Input capacitance of *D* and *CP* inputs are $C_{in}(D) = 3.49fF$ and $C_{in}(CP) = 4.95fF$. Total gate width of this flip-flop is 24.9μm with all gates minimum length, $L_{min} = 0.5\mu m$.

5.4.2 Conditional-Capture Flip-Flop (CCFF)

The CCFF [Kong00] is shown in Fig. 5.54. It is a positive edge triggered differential input differential output flip-flop.

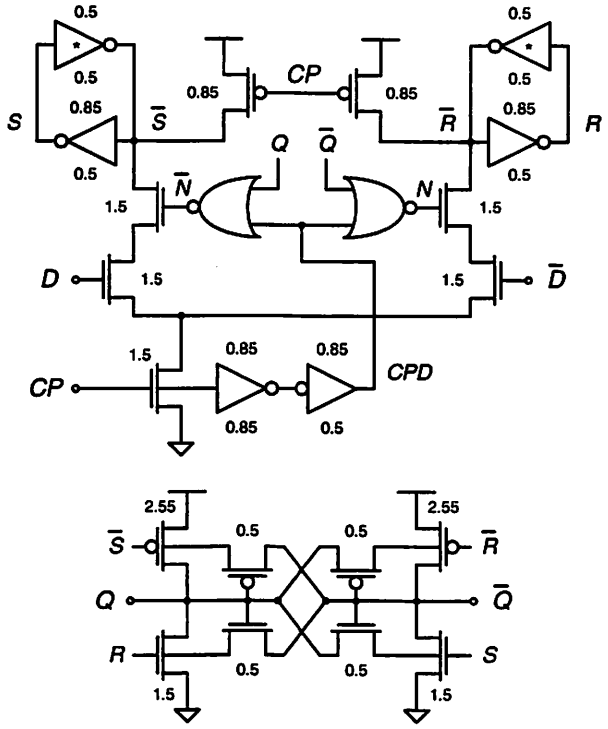


FIGURE 5.54 Conditional-capture flip-flop (CCFF).

5.4.3.1 Circuit Operation

The CCFF is similar to the modified SAFF [Nikolic99], with modifications in the output stage and addition of the internal clock gating. When *CP* is "low", the flip-flop is in the *precharge* phase, $\overline{S}$ and $\overline{R}$ are precharged "high" and the SR latch is disabled. At the rising edge of *CP*, behavior of the CCFF depends on the incoming data value – if new datum is not equal to the previously recorded output datum, one of the outputs of the NOR gates is "high" enabling pull-down of $\overline{S}$ or $\overline{R}$. The transparency period of the diff-pair is two inverter delays long because *N* and $\overline{N}$ both go "low" when *CPD* is "high". During this short transparency period a new datum is latched by the SR latch at the output.

5.4.3.2 Sizing

Transistors of the CCFF are sized to achieve same driving capability as in the case of the SAFF and the modified SAFF. Transistor widths are denoted on the circuit schematic in Fig. 5.54.

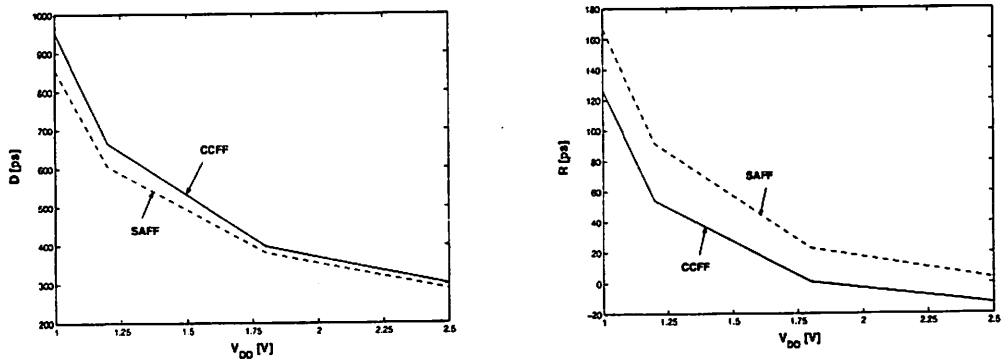
TABLE 5.24 TIMING AND ENERGY PARAMETERS OF CCFF. C_{OUT} = 4SL

V_{DD} [V]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]	E_{0-0} [fJ]	E_{0-1} [fJ]	E_{1-0} [fJ]	E_{1-1} [fJ]
1.0	-50	710	953	126	23.0	109	109	23.0
1.2	-20	510	666	54	33.6	161	162	34.1
1.8	30	310	399	1	83.4	391	395	84.0
2.5	60	230	301	-13	192	842	848	192

5.4.3.3 Characterization

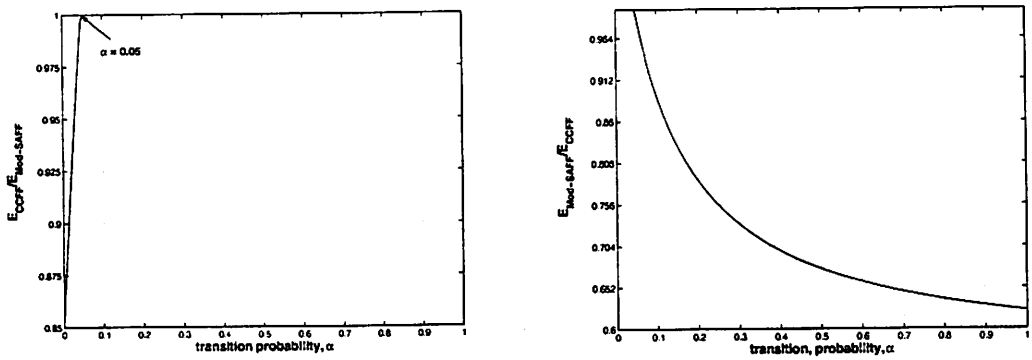
Timing and energy parameters of the CCFF are summarized in Table 5.24. Relative to the modified SAFF, the CCFF has longer delay and smaller internal margin as shown in Fig. 5.55.

Energy wise, the CCFF is worse than the modified SAFF as shown in



(a) (b)

FIGURE 5.55 Comparison of timing parameters of CCFF and modified SAFF: (a) delay, (b) internal race immunity.



(a) (b)

FIGURE 5.56 Energy saving capabilities of CCFF compared to modified SAFF: (a) $E_{CCFF}/E_{Mod-SAFF}$, (b) $E_{Mod-SAFF}/E_{CCFF}$. V_{DD} = 1V, C_{out} = 4SL.

Fig. 5.56 for the circuit sizes indicated in the schematics. This means that the internal clock gating is not applicable if the transistors in a flip-flop are small so that the gating circuit incurs relatively big overhead in energy consumption.

Physical parameters: Input capacitance of the data and clock inputs are $C_{in}(D) = 2.1fF$ and $C_{in}(CP) = 6fF$. Total gate width of this flip-flop is 24.9μm with all gates minimum length, $L_{min} = 0.5\mu m$.

5.4.4 Gated TGFF

The flip-flops with internal clock gating presented thus far were all pulse triggered and thus suffered from very small intrinsic race margin. Now, the internal clock gating is applied to the 6-stage TGFF, master-slave latch-pair, which should gain its low energy consumption and offer good internal race immunity. The TGFF with internal clock gating in the master stage is shown in Fig. 5.57.

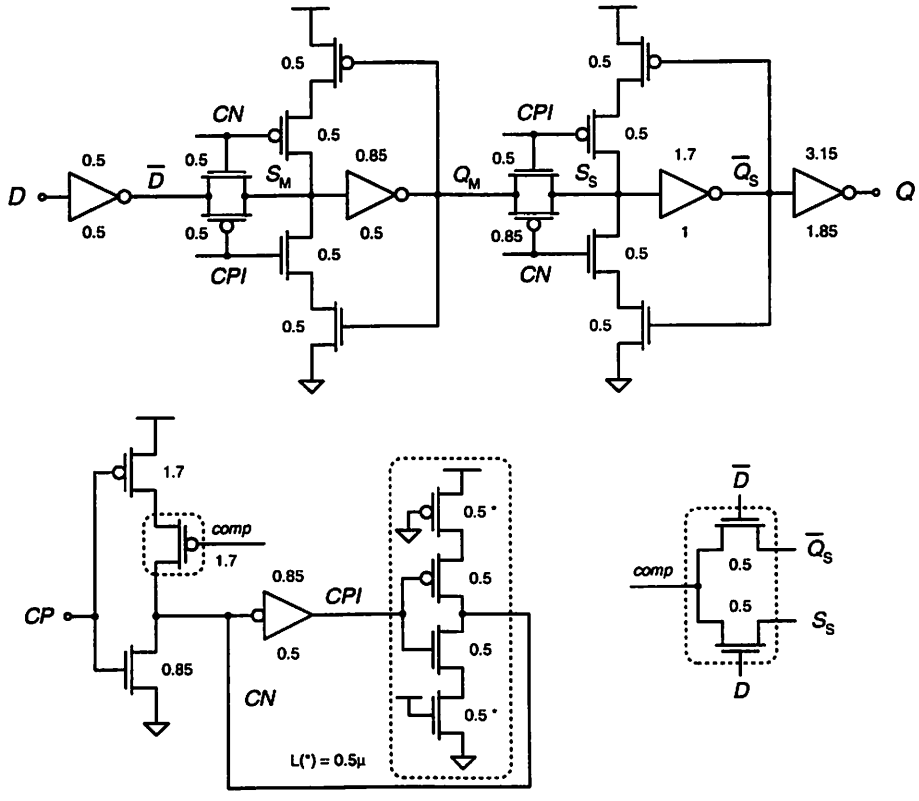


FIGURE 5.57 TGFF with internal clock gating in the master stage.

5.4.4.1 Circuit Operation

This flip-flop is derived from the 6-stage TGFF shown in Fig. 5.11. The circuitry for internal clock gating, which is a modification of the clock gating circuitry presented in [Strollo00], is encircled with dashed lines in Fig. 5.57. Comparator performs an XNOR operation on D and Q. The comparator is implemented in complementary pass-transistor logic (CPL) technique [Yano90] taking advantage of freely available true and complementary signals. This reduces transistor count of the clock gating circuitry. When $D \neq Q$, output of the comparator, *comp*, is "low" and enables external clock *CP* to propagate through the internal clock generation circuits which generate internal clocks *CN* and *CPI*.

The pull-up side of the input clock inverter is chosen to be gated because the CPL realization of an XNOR has better pull-down allowing for the generation of the internal clocks faster than if pull-down side of the input inverter was gated. Weak feedback is added around the inverter that outputs *CPI* for pseudo-static operation.

5.4.4.2 Sizing

The transistors of the comparator are minimal size. The pull-up of the input inverter is sized for equivalent driving capability as the input inverter of the TGFF without internal clock gating shown in Fig. 5.11. Weak transistors in the feedback are all minimum width. The size of all other transistors is identical to the size of corresponding transistors in the TGFF without internal clock gating.

5.4.4.3 Characterization

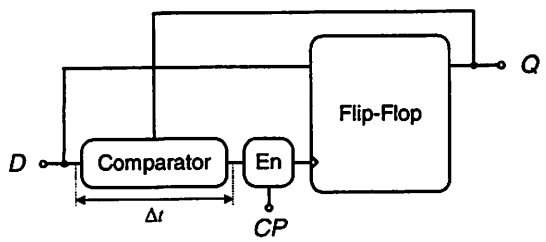


FIGURE 5.58 Impact of internal clock gating circuitry on flip-flop setup time.

TABLE 5.25 TIMING AND ENERGY PARAMETERS OF GATED TGFF. C_{OUT} = 4SL

V_{DD} [V]	t_{su} [ps]	t_{hd} [ps]	D [ps]	R [ps]	$E_{0\text{-}0}$ [fJ]	$E_{0\text{-}1}$ [fJ]	$E_{1\text{-}0}$ [fJ]	$E_{1\text{-}1}$ [fJ]
1.0	990	-80	2380	1280	0	79.8	52.6	0
1.2	730	-50	1710	928	0	116	77.4	0
1.8	370	-40	855	502	0	280	194	0
2.5	260	-10	603	337	0	576	407	0

Timing and energy parameters of the gated TGFF are listed in Table 5.25.

Compared to the TGFF without gating, delay of the gated TGFF is longer due to increased setup time. An increase in the setup time of the gated TGFF is explained by an additional delay Δt through the comparator as illustrated in Fig. 5.58. The hold times on the other hand are nearly identical because holding an input datum is decoupled from enabling of the clock. When a new input datum is being recorded, *CP* is already enabled so there is no additional propagation delay through the comparator. Figure 5.59 shows the difference between setup and hold times of the gated and conventional TGFF.

$E_{0\text{-}1}$ and $E_{1\text{-}0}$ are increased in the TGFF with internal clock gating because there exists an overhead in clock gating logic, but the clocking energy is not consumed when input *D* is not changing. Overall flip-flop with internal clock

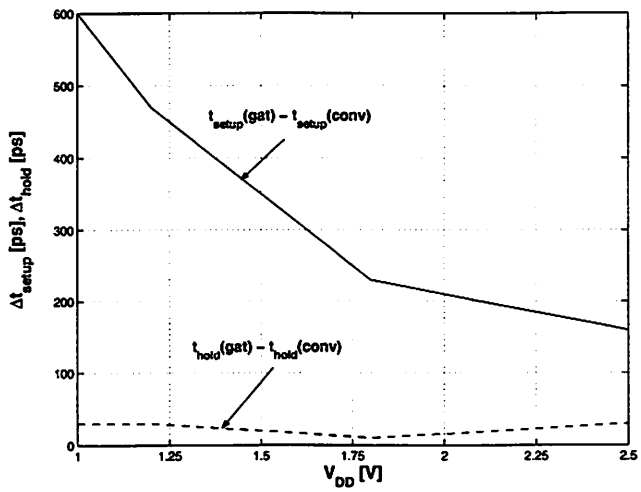
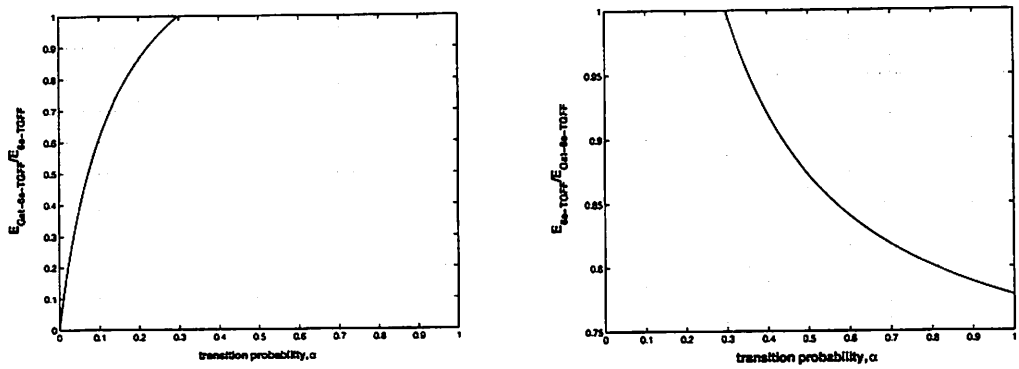


FIGURE 5.59 Impact of internal clock gating on setup and hold times in TGFF: t_{setup}(gat) – t_{setup}(conv) and t_{hold}(gat) – t_{hold}(conv) as a function of supply voltage.



(a) (b)

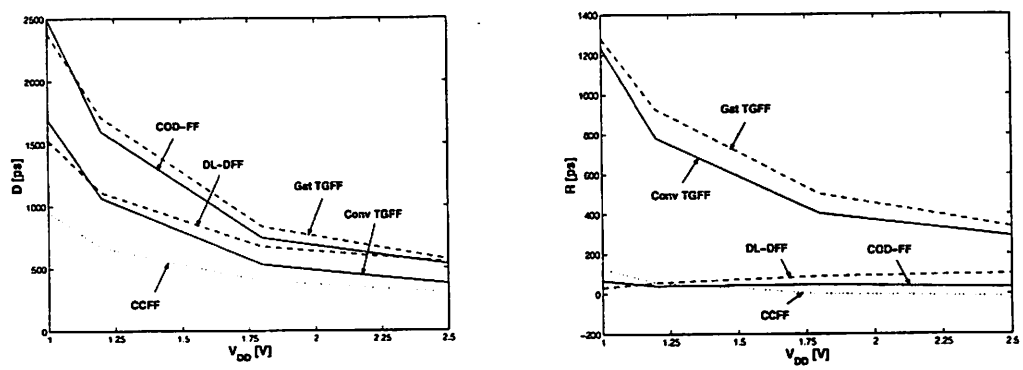
FIGURE 5.60 Illustration of energy efficiency of the gated TGFF compared to a conventional TGFF: (a) $E_{Gat\text{-}6s\text{-}TGFF}/E_{6s\text{-}TGFF}$, (b) $E_{6s\text{-}TGFF}/E_{Gat\text{-}6s\text{-}TGFF}$.

gating has better energy efficiency than the conventional flip-flop when input data activity is $\alpha < 0.3$ as shown in Fig. 5.60.

Physical parameters: Input capacitance of the data and clock inputs are $C_{in}(D) = 2.6fF$ (2fF in TGFF without gating) and $C_{in}(CP) = 3.4fF$ (2.45fF in the TGFF without gating). Total gate width of this flip-flop is 23.65μm.

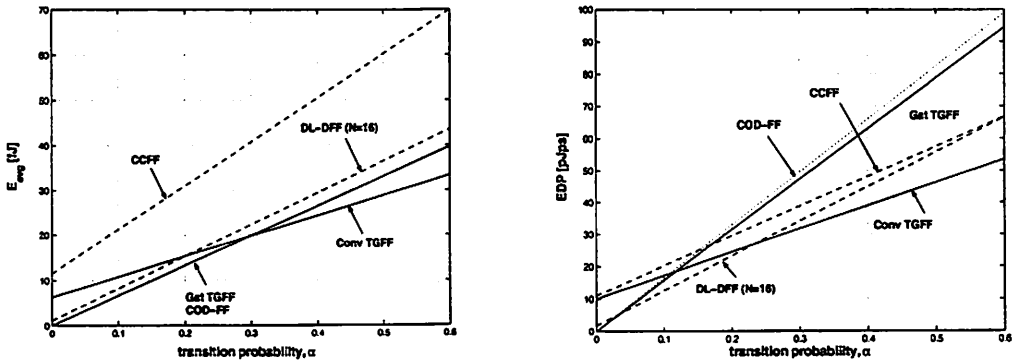
5.4.5 Summary of Flip-Flops with Internal Clock Gating

Delay of the gated TGFF is increased relative to the conventional TGFF as shown in Fig. 5.61.a due to an increase in setup time shown in Fig. 5.58. In the DL-DFF and COD-FF, setup and hold times are coupled with delay of internal gating logic and the internal clock pulse width, so these flip-flops have longer delay than, for example, CCFF where there are no such constraints on setup and hold times. Figure 5.61.b shows comparison of internal race immunity of the



(a) (b)

FIGURE 5.61 Comparison of timing parameters in flip-flops with internal clock gating as a function of supply voltage: (a) delay, (b) internal race immunity. $C_{out} = 4SL$.



(a) (b)

FIGURE 5.62 Comparison of flip-flops with internal clock for various input data activities (a) average energy, (b) energy-delay-product. $V_{DD} = 1V$, $C_{out} = 4SL$.

gated flip-flops and the conventional TGFF. The gated TGFF has even better internal race margin than the conventional TGFF because of increased *Clk-Q* delay. Figure 5.61.b shows that the pulse-triggered latches have very small race margin.

Average energy consumption of the flip-flops with internal clock gating as a function of input data activity is shown in Fig. 5.62.a. Since it is not fair to compete pulse-triggered latches and master-slave latch-pairs in terms of energy efficiency because pulse-triggered latches have higher performance, comparison of energy-delay-product is given in Fig. 5.62.b. From Fig. 5.62.b it seems that the DL-DFF is the best flip-flop to use for $\alpha \in [0.03, 0.23]$. This is not correct because of its poor race margin. The conventional TGFF offers best energy-delay tradeoff and also good race immunity for $\alpha > 0.12$, while its gated version has the best EDP for $\alpha < 0.12$.

Table 5.26 summarizes physical parameters of FFs with internal clock gating.

TABLE 5.26 PHYSICAL PARAMETERS OF FLIP-FLOPS WITH INTERNAL CLOCK GATING

	C_{in}(D) [fF]	C_{in}(CP) [fF]	W_{tot} [μm]
DL-DFF	2	3.2	23.1
COD-FF	3.5	5	24.9
CCFF	2.1	6	35.85
Gat-TGFF	2.6	3.4	23.65

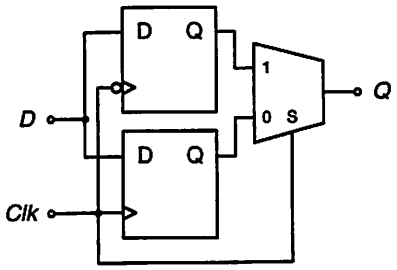


FIGURE 5.63 Block diagram of double-edge-triggered flip-flop.

5.5 DOUBLE-EDGE-TRIGGERED FLIP-FLOPS

Double-edge-triggered (DET) flip-flops sample their inputs and update their outputs on both the rising and falling edges of the clock. With this approach the maximum toggle frequency of the clock is identical to the maximum toggle frequency of the data. In contrast with conventional single-edge-triggered flip-flops, the clock frequency is twice the highest data frequency, which, to first approximation, doubles the clock energy.

Double-edge triggering requires careful control of the clock's duty factor to ensure that the combinational logic has adequate time to operate during both the clock "high" and the clock "low" cycles. The DET-FF is constructed from two (level-sensitive) latches and a multiplexer as shown in Fig. 5.63. The top latch samples data on the rising edge of the clock, and the bottom latch samples data on the falling edge of the clock. The multiplexer is switched by the clock always to select the latch that is holding sampled data, not the latch that is passing its input to its output. Figure 5.64 shows the timing waveforms of the DET-FF.

5.5.1 DET-FF Based on a Transmission Gate Latch

The DET-FF based on a TG latch is shown in Fig. 5.65.

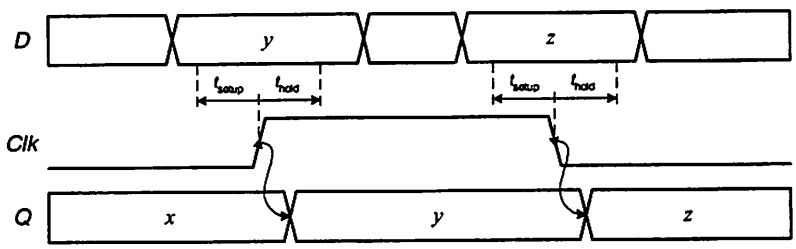


FIGURE 5.64 Timing waveforms of DET-FF.

5.5.1.1 Circuit Operation

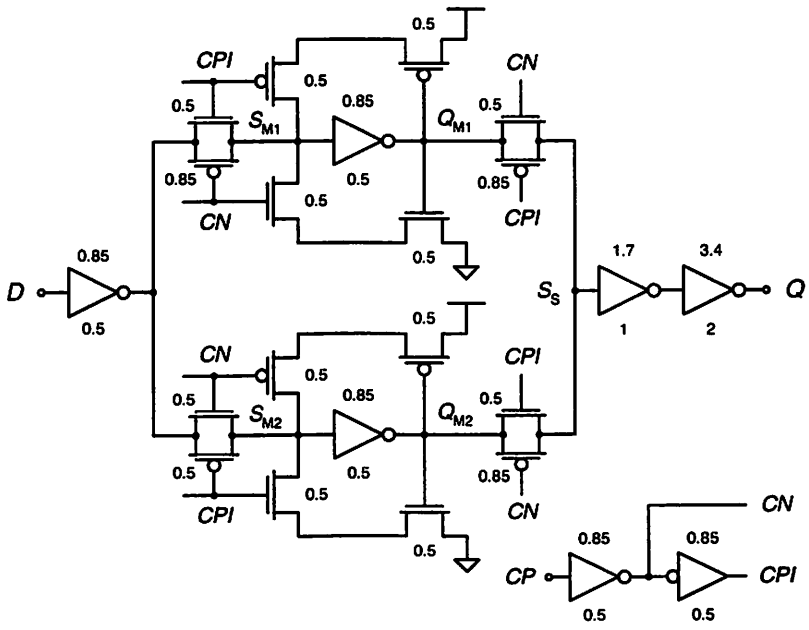
The DET-FF has an input inverter for input gate isolation of the latches in parallel branches. The latch outputs are selected by one of the two transmission gates. Output consists of two cascaded driving inverters. Circuit of Fig. 5.65 operates as described by timing diagrams of Fig. 5.64.

5.5.1.2 Sizing

Active circuit that limits *Clk-Q* delay is very similar to the active circuit of the 6-stage TGFF, so similar sizing equations apply to the DET-FF.

5.5.1.3 Characterization

Timing parameters of the DET-FF are similar to the 6-stage TGFF. Main feature of the DET-FF is energy saving in clocking of the flip-flop. Total clocked capacitance in each of the parallel branches in the DET-FF is smaller than the total clocked capacitance of the 6-stage TGFF, which leads to savings in E_{CLK} of the DET-FF as shown in Fig. 5.66. The input clock capacitances of these two flip-flops are equal and the clock frequency of the DET-FF is two times smaller so the DET-FF achieves two times smaller energy consumption in the clock distribution network.



FIUGRE 5.65 DET-FF based on a TG latch.

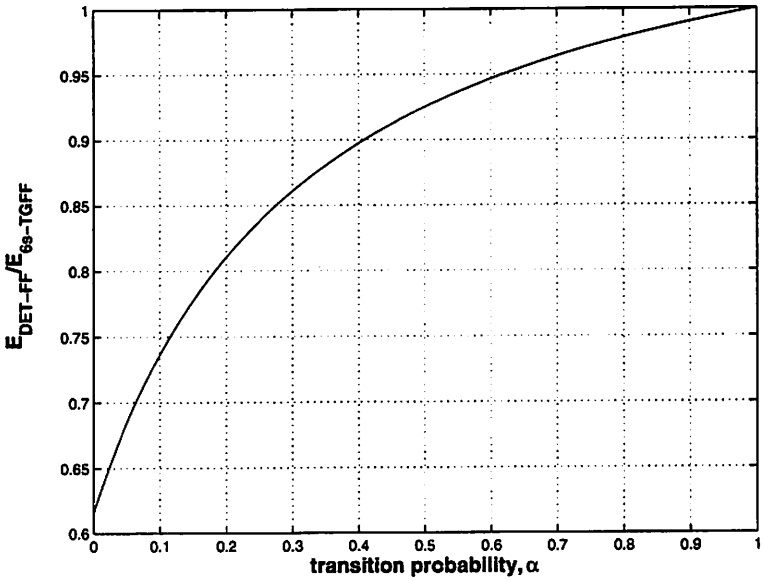


FIGURE 5.66 Comparison of average energy consumption of the DET-FF and 6-stage TGFF. Both flip-flops have same C_{in}(CP) and same driving capability. The DET-FF has 20% larger gate width. $V_{DD} = 1V$, $C_{out} = 4SL$.

In the DET-FF, E_{CLK} is different when the flip-flop is triggered by positive and negative edges of the clock due to difference in the total charged capacitance. This difference is equal to 2.7fJ at $V_{DD} = 1V$, which corresponds to $\Delta C = 2.7fF$. The clocking energy is estimated as the average clocking energy for both clock edges.

An extra energy is consumed in the DET-FF by sneak current that exists during short time period Δt, as illustrated in Fig. 5.67.

Physical parameters: Input capacitance of *D* and *CP* inputs are $C_{in}(D) = 3.5fF$

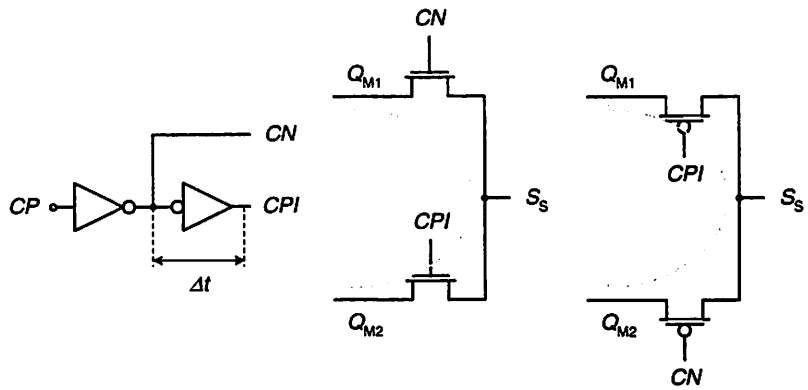


FIGURE 5.67 Illustration of sneak current path in DET-FF.

and $C_{in}(CP) = 3.5fF$. Total gate width of this flip-flop is 23.9μm with all gates minimum length, $L_{min} = 0.5\mu m$.

5.6 SUMMARY

From the examples presented, we can summarize some of the main methods used in the design process to achieve low-energy consumption in memory elements. The preceding examples have indicated that low-energy design can be done in a systematic fashion to a certain degree, implying that the designer still has various degrees of freedom. Low energy designs can be effectively concocted with methods that employ clock gating, reduced swing clocking, or dual-edge-triggering. While there are two types of flip-flops (master-slave and pulse-triggered), the principles used to optimize them are essentially the same, differing only in the physical realization and in the resulting tradeoffs made with regards to circuit delay and internal race immunity.

Internal clock gating is an efficient design technique used to reduce energy consumption in flip-flops by virtually disabling the internal clock when the input data pattern is constant. The advantage here is that energy is not expended during periods of input inactivity. There are several design issues associated with this technique. One major issue is the additional energy consumed in the newly added clock gating logic. In order for this technique to be considered efficient, the additional energy usually must be less than the energy in the internal clocked nodes. It must also be stressed that the switching activity of the input must be considered when deciding on the feasibility of internal clock gating. Clock gating should be used on low input data rates, otherwise the energy consumed by the clock gating logic will dominate the total energy measure. In other words, the implementation of the clock gating logic incurs unnecessary overhead, opting for the designer to find another circuit optimization technique. The delay typically increases as a result of an increase in clock-to-output delay. The clock-to-output delay reflects an extra propagation delay from the additional clock gating logic. While detrimental for delay, this increase in clock-to-output delay can sometimes

improve the flip-flop internal race margin. When the input pattern switches with high probability, another technique must be used to reduce the energy consumption. Many designs today resort to using the method of *reduced swing clocking*.

As the name implies, reduced swing clocking means that the clock signal does not complete a full rail-to-rail cycle. At a first glance, this technique appears to offer tremendous savings in energy because the energy expended in the clocked nodes decreases quadratically with the maximum clock output level. This does not necessarily lead to an improvement in energy consumption because in typical circuits, reduced swing operation sometimes requires exclusive transistor sizing and substrate biasing in order to suppress excessive leakage currents.

One last technique introduced in this chapter is the use of *double-edge triggering* (DET). DET offers many wonderful advantages to deal with the energy-performance tradeoff. Because both edges of the clock are utilized, there is an increase in energy efficiency. To achieve this efficiency, we therefore must reduce the clock frequency to lower the energy and to maintain the same level of performance. There are some very important design issues regarding the use of DET. A careful control of the clock's duty factor is crucial to ensure that the combinational logic has sufficient time to function correctly during both phases of a clock cycle.

A careful examination of the various circuits presented in this chapter reveal, in addition to the techniques described above, an important abstraction that can be identified and described in each flip-flop design. This abstraction is an active circuit that dictates the performance of the flip-flop. For a given flip-flop class (master-slave or pulse-triggered), we find that the active circuits are very similar. The active circuit is important because it is the driving force involved in the most complex computations within the flip-flop. The ability to recognize this active circuit allows the designer to make the appropriate changes to enhance the properties of the flip-flop with respect to energy and performance.

On a final note, the choice of flip-flop topology depends on the target application. Flip-flops based on master-slave latch-pairs in general have several advantages over those based on pulse-triggered latches. Master-slave latch-pairs tend to have better race immunity at the expense of increased delay. Pulse-triggered latches generally possess higher energy consumption, and therefore practical applications where low energy is primary concern would involve master-slave topologies. Likewise, master-slave latch-pairs are used over pulse-triggered latches when performance is not the main design goal. Once a flip-flop topology has been chosen for the specific application, further effort to reduce the overall energy is essentially a systematic procedure. The overall result, therefore is a energy conservative flip-flop that possesses desirable properties from a system perspective.

Chapter 6

GLITCHING ENERGY IN FLIP-FLOPS

This chapter analyzes the energy consumed by dynamic hazards that are generated by the unintended transitions propagating from the fan-in gates, often termed *propagating glitches* [Hash98]. Glitches produced by non-glitch transitions at the inputs, called *generated glitches*, are neglected in our analysis because we deal only with static flip-flops and hence this problem is eliminated.

There are four types of glitches that can occur in flip-flops as shown in Fig. 6.1. Average flip-flop glitching energy is determined by the glitching probability and the energy that flip-flop consumes during glitching, Equation (6.1).

$$E_{avg-glitch} = \sum_{i=1}^{4} \beta_i \cdot E_{g_i} \qquad (6.1)$$

Flip-flop glitch sensitivity depends on its structure. In general, the pulse-triggered latches (SDFF, HLFF, modified SAFF) exhibit better glitch immunity than the master-slave latch pairs (TGFF, C^2MOS). This is because internal nodes in the pulse-triggered latches are coupled with *D* input only during narrow period when flip-flops sample input data; while in master-slave latch pairs, the master latch, when transparent, is sensitive to glitches half of the clock period. Flip-flops with internal clock gating are susceptible to glitches the most because the glitches

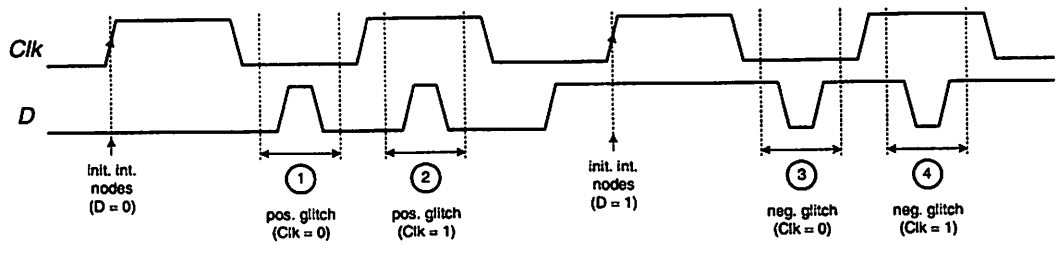


FIGURE 6.1 Types of glitches in flip-flops.

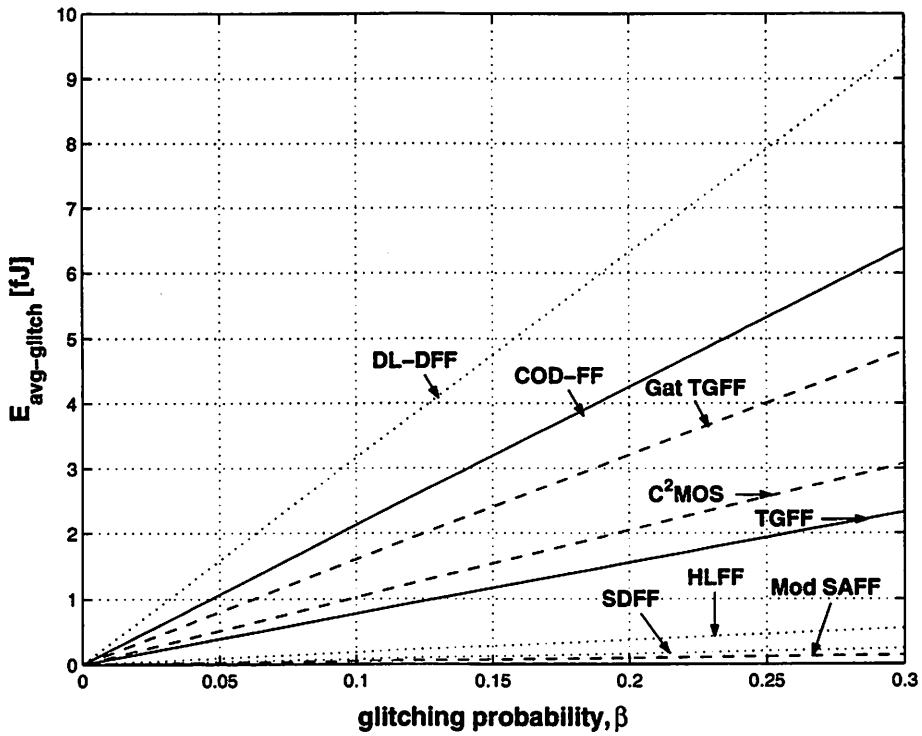


FIGURE 6.2 Average glitching energy in flip-flops as a function of glitching probability, β. V_{DD} = 1V.

affect both their internal nodes and nodes inside the clock gating logic.

Figure 6.2 shows comparison of average glitching energy in different flip-flops. The results are obtained under the assumption that all glitches may appear with the same probability:

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \frac{\beta}{4} \tag{6.2}$$

Fig. 6.2 would change if the probabilities were not equal because the glitching component of energy depends on which glitch occurs at the input. Figure 6.3 illustrates energy-per-transition and glitching energy in several representative flip-flops: TGFF, SDFF, COD-FF and gated TGFF. The results showing glitching energy in SDFF can be generalized to HLFF and SAFF because the portion of the circuit that is affected by glitches is very similar in all these flip-flops, Fig. 6.5. Equivalent circuits that describe glitching activity of the internal nodes in master-slave latch pairs, pulse-triggered latches, and gated flip-flops are shown in Fig. 6.4-6.6.

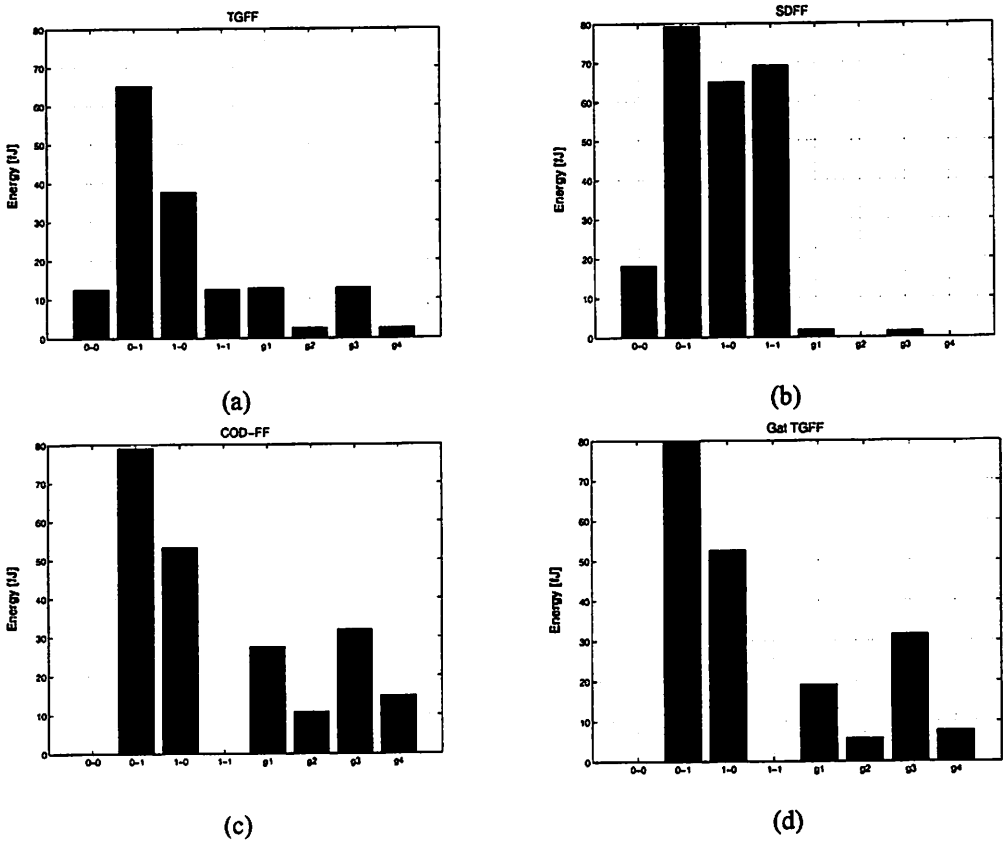


FIGURE 6.3 Comparison of energy-per-transition and glitching energy in flip-flops. (a) TGFF, (b) SDFF, (c) COD-FF, (d) gated TGFF. $V_{DD} = 1V$, $C_{out} = 4SL$.

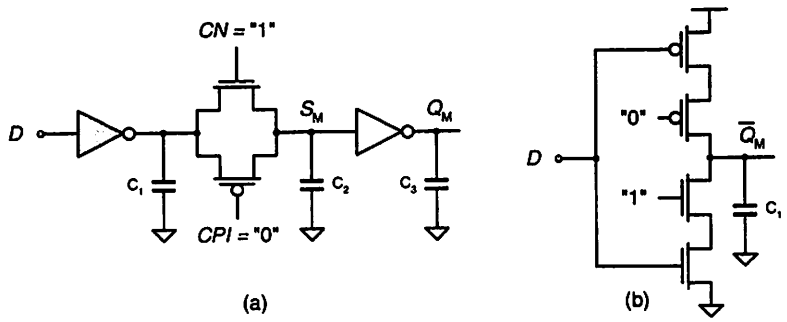


FIGURE 6.4 Equivalent circuits that describe glitching activity of internal nodes in master-slave latch pairs: (a) TGFF, (b) C^2MOS flip-flop.

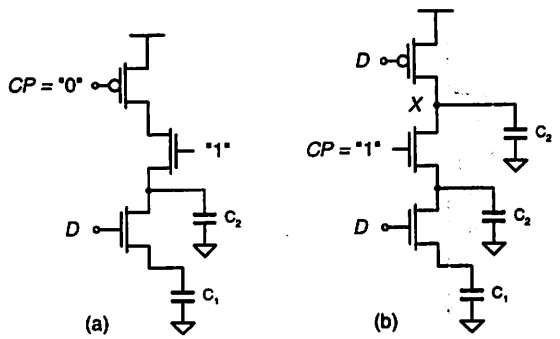


FIGURE 6.5 Equivalent circuits that describe glitching activity of internal nodes in pulse-triggered latches: (a) SDFF, SAFF, and modified SAFF, (b) HLFF.

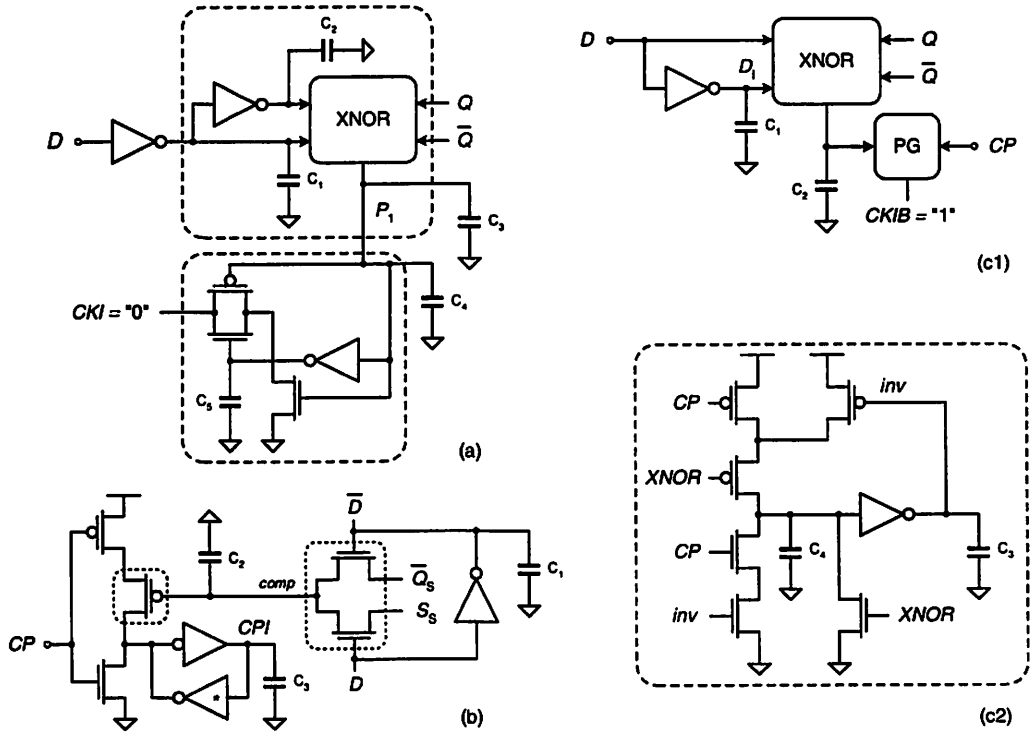


FIGURE 6.6 Equivalent circuits that describe glitching activity of internal nodes in flip-flops with internal clock gating: (a) DL-DFF, (b) gated TGFF, (c1) COD-FF, (c2) pulse generator of the COD-FF.

Energy consumption of the circuits in Fig. 6.5 is small for two reasons. First, D input is decoupled from internal nodes during clock "low". Second, capacitances C_1 and C_2 are very small transistor parasitic capacitances. By circuit inspection, one can see that C_2 is always precharged to $V_{DD} - V_{Tn}$, while C_1 is charged-up if positive glitch appear at D during clock "high".

The TGFF and C^2MOS flip-flop reduce to the master latch during clock "low" and input glitches propagate to the master latch output. During clock "high" only shaded inverter of Fig. 6.4.a consumes energy.

Most of the nodes in clock gating circuitry of the flip-flops with internal clock gating are coupled with input D for both logic levels of the clock. Typical representative of such undesired behavior is the DL-DFF, Fig. 6.6.a. The COD-FF and gated TGFF have somewhat lower glitching activity of their internal nodes. For example, C_3 of the gated TGFF is switched only when input glitches arrive during clock "low". Also, C_3 and C_4 of the COD-FF are switched only when input glitches arrive during clock "high".

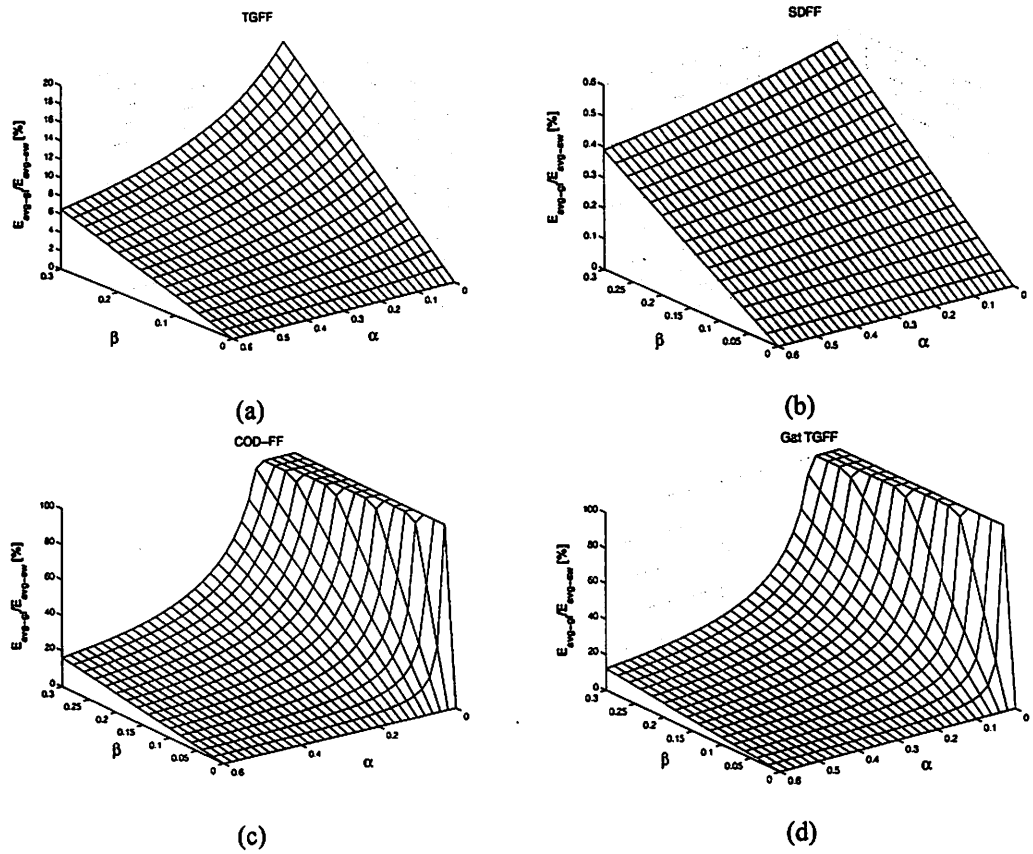


FIGURE 6.7 Comparison of average glitching energy and average switching energy in representative flip-flops: (a) TGFF, (b) SDFF, (c) COD-FF, (d) gated TGFF. V_{DD} = 1V, C_{out} = 4SL.

Comparison of average glitching energy and average switching energy as a function of input signal transition probability, α, and input signal glitching probability, β, is depicted in Fig. 6.7. Glitching probability, β, is assumed to be at most 30%.

The TGFF can consume about 15% of energy in propagating input glitches, depending on the input glitching activity. Glitching component of energy is negligible in SDFF (and other pulse-triggered latches), while the glitching component of energy in the flip-flops with internal clock gating can even exceed switching component at low α.

6.1 SUMMARY

Detrimental effects that impact energy consumption in registers are spurious input data transitions, often referred to as *glitches*, or dynamic hazards.

The flip-flops that have proven the best thus far in terms of energy efficiency and wide race margins are susceptible to propagating glitches that are present at the data input nodes. For example, for very low input data transition probabilities (less than 0.1) and relatively high glitching probabilities (greater than 0.1) the glitching component of energy in the clock gated transmission gate flip-flop can become equal or even greater than its switching component counterpart. Therefore, the glitching activity is another important parameter to consider prior to the selection of a suitable flip-flop topology that aims to minimize the overall average energy consumption of the clocking subsystem.

In prior chapters, we examined the different flip-flop topologies and classified them according to their energy dissipation factor. When we add glitching to the picture, we generally arrive at opposite conclusions. Specifically, the master-slave latch-pairs exhibited lower energy consumption than their pulse-triggered analogues. However, the pulse-triggered latches are less prone to anomalies in the input data patterns. Master-slave latch pairs, on the other hand suffer from increased sensitivity to these dynamic hazards which ultimately affect the robustness of the design. It is therefore, imperative not to overlook the possible degradation in signal integrity due to the presence of glitches.

Chapter 7

INTERFACING ISSUES UNDER SUPPLY VOLTAGE SCALING

This chapter discusses impact of clock slope on the clock skew requirement and the energy consumption in the clock distribution network. Section 7.3 discusses the issues involved in the scaling of the flip-flop and logic delay over a range of supply voltages. We also examine the impact of the short-circuit and subthreshold components of energy consumption in flip-flops.

7.1 CLOCK SKEW AS A FUNCTION OF CLOCK SLOPE

The clock skew is limited by the clock-to-output delay of flip-flop A and hold time of flip-flop B shown in Fig 7.1. Neglecting delay of the interconnect between Q_A and D_B, the clock skew of Fig. 7.1 is constrained by:

$$t_{skew} < t_{C-Q,A}(t_{rise}(Clk_A)) - t_{hd,B}(t_{rise}(Clk_B)) \tag{7.1}$$

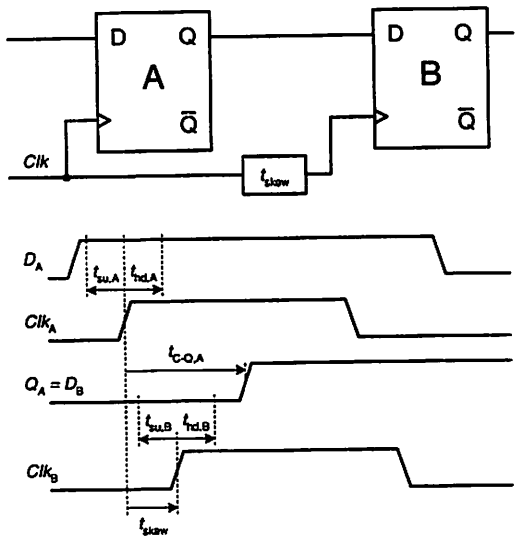


FIGURE 7.1 Illustration of clock skew specification.

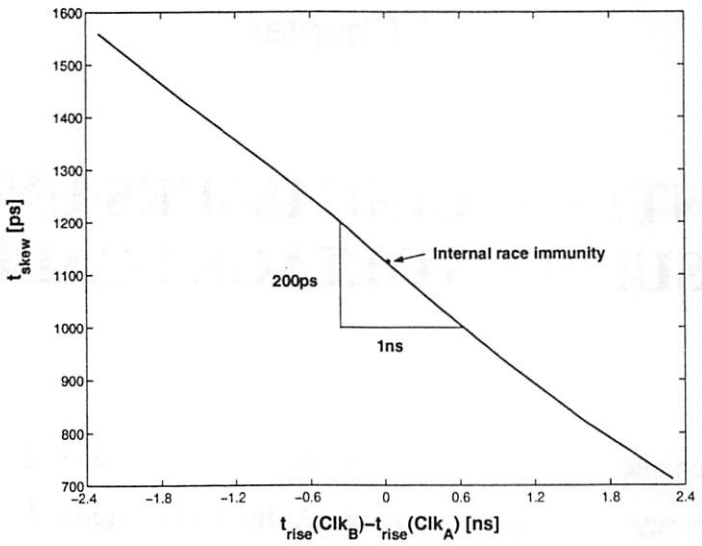


FIGURE 7.2 Impact of mismatch in the slopes of two clock signals, driving two 6-stage TGFFs, connected as shown in Fig. 7.1, on the clock skew specification. V_{DD} = 1V.

In Equation (7.1), clock-to-output delay of flip-flop A, $t_{C\text{-}Q,A}$, and hold time of flip-flop B, $t_{hd,B}$, are functions of the clock slopes, $t_{rise}(Clk_A)$ and $t_{rise}(Clk_B)$, respectively. Worst-case scenario is when $t_{rise}(Clk_A) < t_{rise}(Clk_B)$. Figure 7.2 illustrates the flip-flop internal race immunity and clock skew requirement as a function of mismatch between the slopes of clock signals Clk_B and Clk_A that drive two 6-stage TGFFs connected as shown in Fig. 7.1.

Mismatch in the clock slopes of about 2ns can occur if one clock buffer drives about 3,000 flip-flops. Figure 7.3 shows post-layout simulation results of

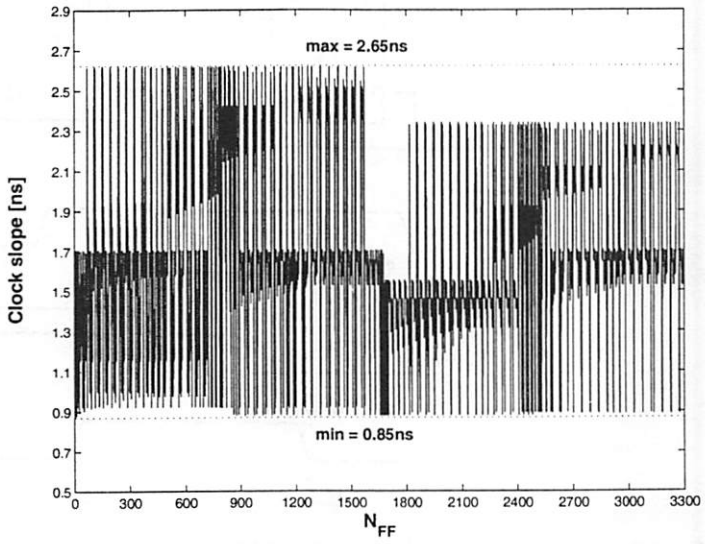


FIGURE 7.3 Distribution of clock slopes in a decimation filter. V_{DD} = 1V.

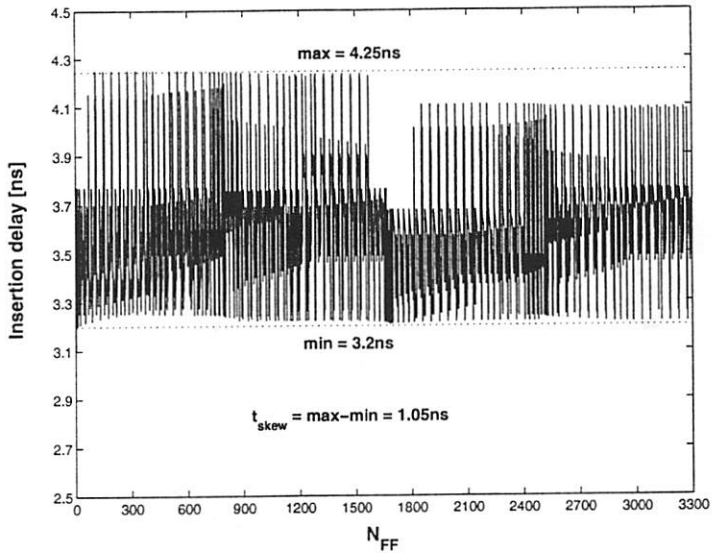


FIGURE 7.4 Distribution of insertion delays in a decimation filter. V_{DD} = 1V.

distribution of the clock slopes of one clock phase, across a 3mm×3mm chip implementing a decimation filter, which contains 3,362 flip-flops with two-phase non-overlapping clocks. Total clock input flip-flop capacitance is 35pF, and estimated wire capacitance is about 30pF. Spectre simulation results have shown worst-case mismatch of 1.8ns.

If the single-phase 6-stage TGFF is used in the filter design, it could tolerate about 750ps of clock skew worst-case. Figure 7.4 shows distribution of insertion delays of one clock phase in the decimation filter chip.

7.2 ENERGY AS A FUNCTION OF CLOCK SLOPE

The clock slope can significantly affect energy consumption of clock drivers. Undesired components of energy, subthreshold and short-circuit, are minimized in clock drivers that operate at low supply voltages up to 1.4V and with rise/fall times of their input waveforms less than about 2ns. Figures 7.5-7.9 show comparison of short-circuit, subthreshold, switching and total energy consumption per clock cycle in a FO4 inverter as function of supply voltage and input signal rise/fall times.

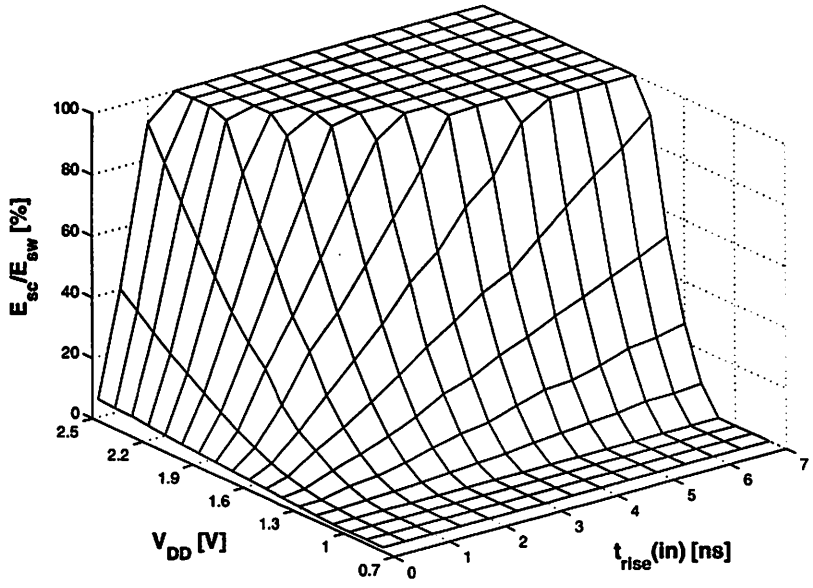


FIGURE 7.5 Percentage ratio of short-circuit and switching components of energy in a FO4 inverter (rounded to 100%) as function of supply voltage and input signal rise time.

The short-circuit component of energy is more sensitive to input signal slopes at higher supply voltages, while the subthreshold component is more sensitive to input signal slopes at lower supply voltages.

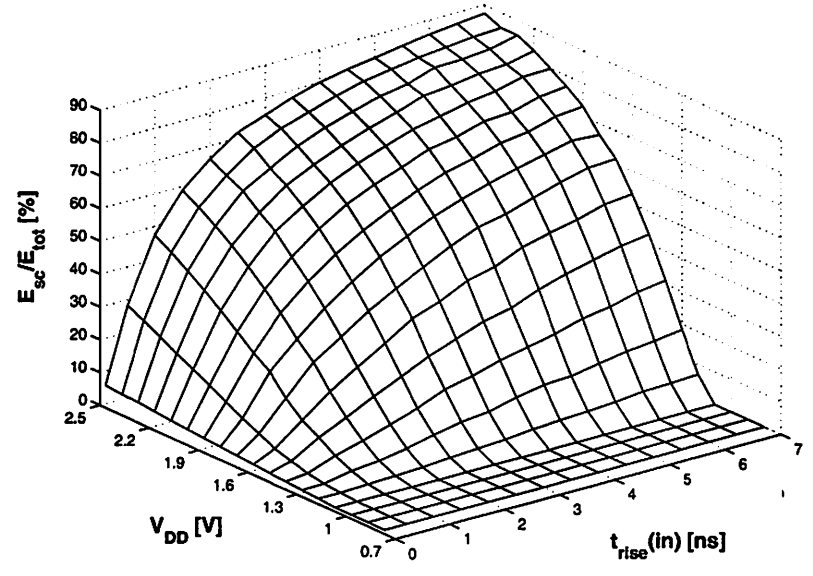


FIGURE 7.6 Percentage ratio of short-circuit and total energy in a FO4 inverter as function of supply voltage and input signal rise time.

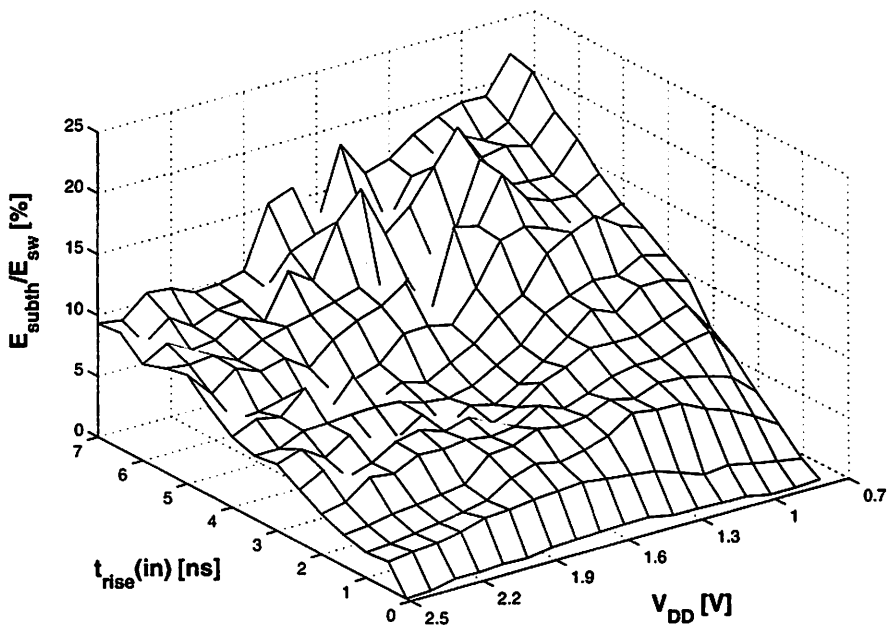


FIGURE 7.7 Percentage ratio of subthreshold and switching components of energy in a FO4 inverter as function of supply voltage and input signal rise time.

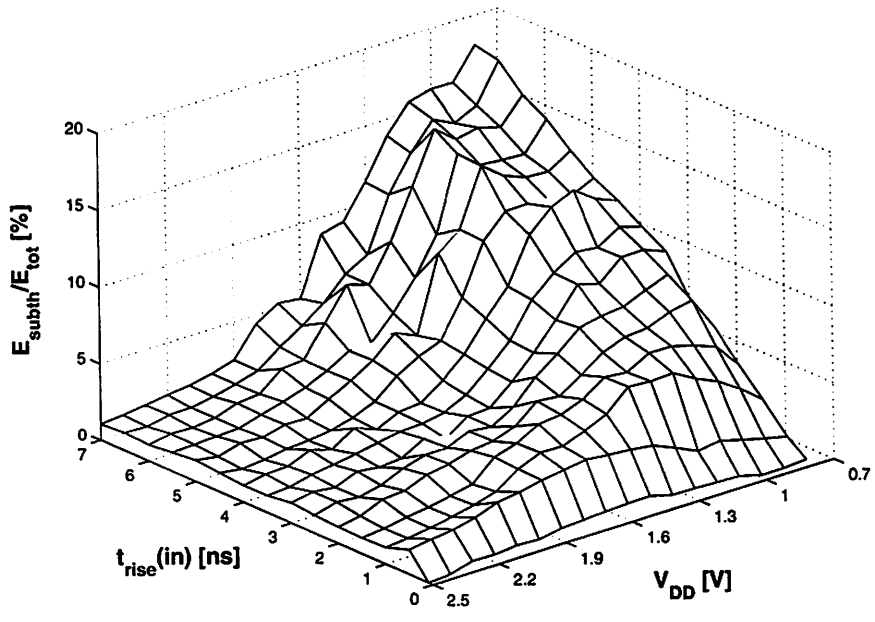


FIGURE 7.8 Percentage ratio of subthreshold and total energy in a FO4 inverter as function of supply voltage and input signal rise time.

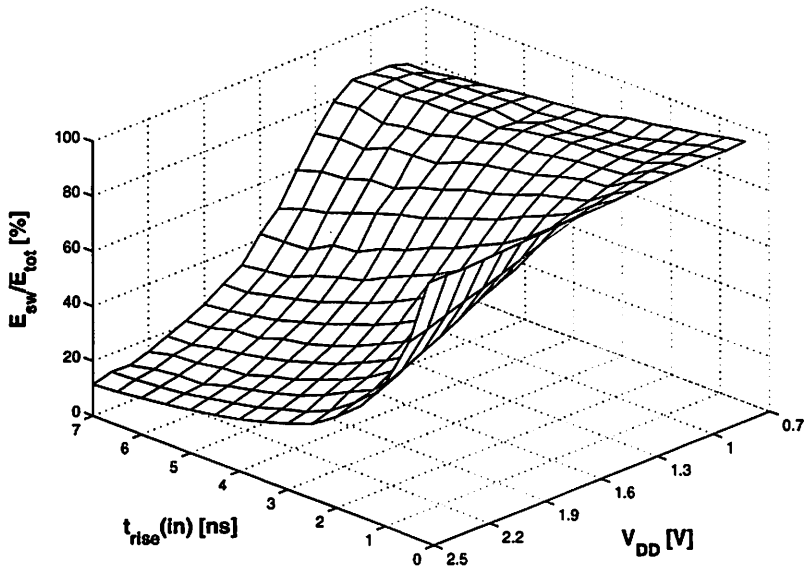


FIGURE 7.9 Percentage ratio of switching and total energy in a FO4 inverter as function of supply voltage and input signal rise time.

7.3 FLIP-FLOPS VS. LOGIC UNDER SUPPLY VOLTAGE SCALING

In digital systems that deal with supply voltage scaling [Burd00] it is desirable that important timing relationships scale with supply voltage. It is thus important that the flip-flop and combinational logic delay scale in the same fashion.

Figure 7.10 depicts delay of the flip-flops and two input NAND and NOR gates relative to the delay of 1*x* inverter, all driving output load equivalent to 4SL. The flip-flops preferred thus far, TGFF and gated TGFF, scale almost the same way as the delay of an inverter. There is slight increase in relative delay of the TGFF at reduced voltages due its transmission gate properties and the fact that pass-transistor logic style does not scale the same way as complementary CMOS. The modified SAFF possesses the best scaling capabilities, since its relative delay decreases with reduction in the supply voltage, which allows for longer delay of combinational logic and versatility in combinational logic design (e.g. utilization of pass-transistor logic style).

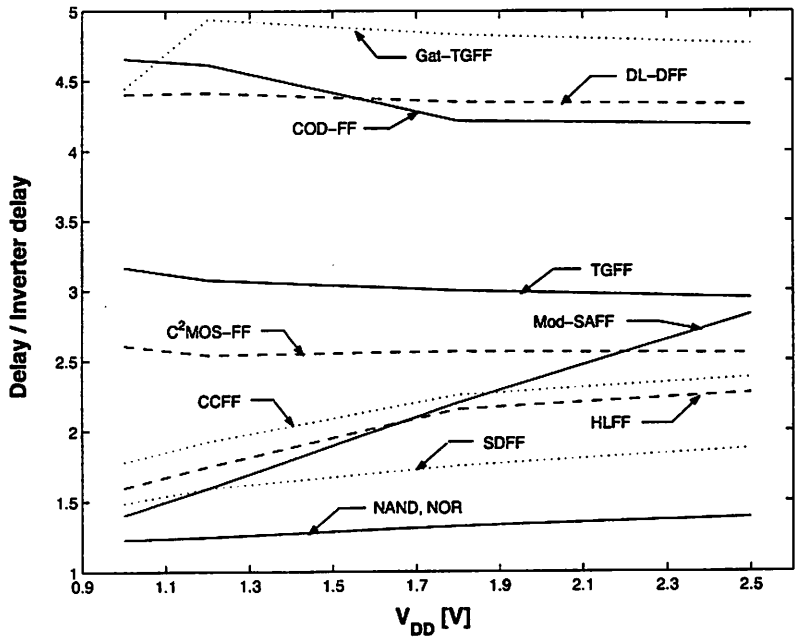


FIGURE 7.10 Voltage scaling of delay in flip-flops and 2-input NAND and NOR logic gates relative to delay of the minimum sized inverter. C_{out} = 4SL.

As far as the race immunity is concerned, it is desirable that it does not decrease, relative to the delay of logic. Figure 7.11 shows that most of the flip-flops exhibit such a desired behavior, except the DL-DFF.

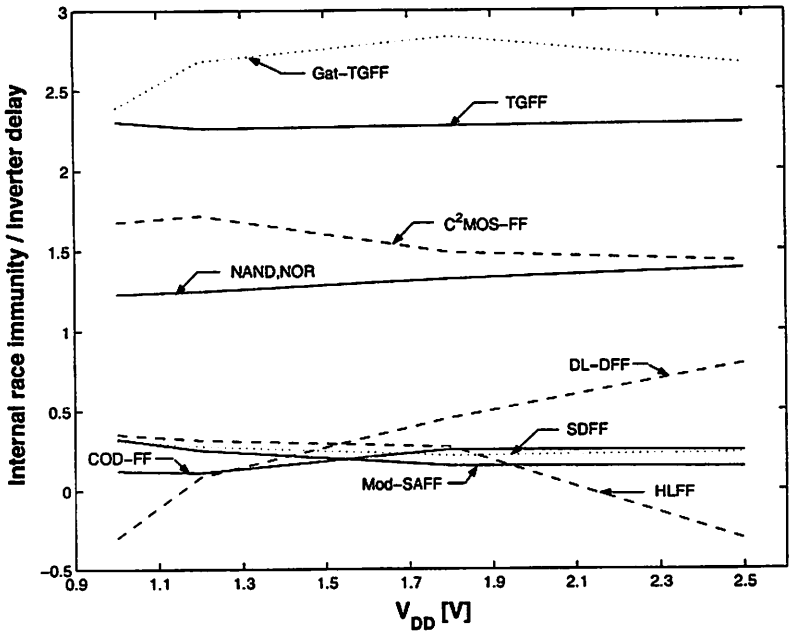


FIGURE 7.11 Voltage scaling of internal race immunity in flip-flops relative to delay of the minimum sized inverter. Delay of 2-input NAND and NOR logic gates is given for reference.

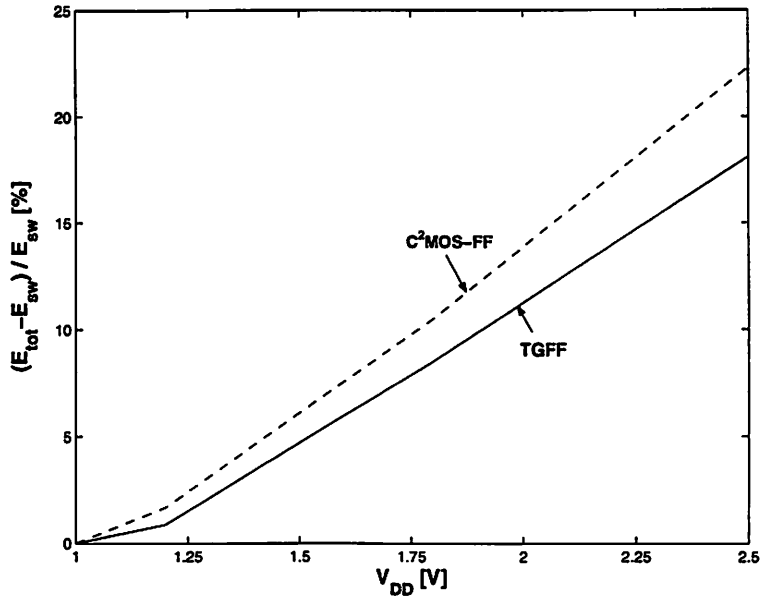


FIGURE 7.12 Short-circuit and subthreshold components of energy in master-slave latch pairs as a function of supply voltage.

Impact of the short-circuit and subthreshold components of energy in flip-flops under supply voltage scaling is shown in Fig. 7.12. In Fig. 7.12, E_{tot} is the average energy consumption if flip-flops assuming glitch free operation and input data transition probability $\alpha = 0.5$. The E_{sw} is calculated as $C_{tot} \cdot V_{DD}^2$, where C_{tot} is the total equivalent switching capacitance estimated as $E_{tot}(1V)$. This assumption is fair because the short-circuit and subthreshold components of energy are negligible at 1V, as shown in Fig. 7.5-7.9.

Figure 7.13 shows the subthreshold and short-circuit components of energy in pulse-triggered latches. These unwanted components of energy are higher in the pulse-triggered latches than in the master-slave latches of Fig. 7.12. For example, in SDFF and HLFF there exists a "fight" between pull-up/down of the precharge node and cross coupled-inverters that keep the precharge node pseudo-static, during precharge/discharge operation of the precharge node.

In the flip-flops with internal clock gating, smallest overhead in short-circuit and subthreshold is found in the gated TGFF and DL-DFF.

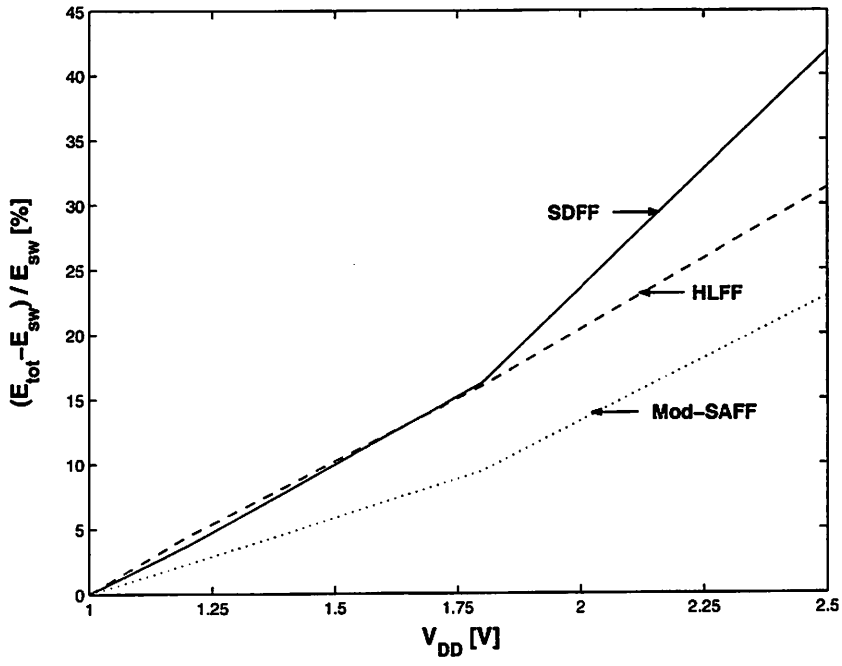


FIGURE 7.13 Short-circuit and subthreshold components of energy in pulse-triggered latches as a function of supply voltage.

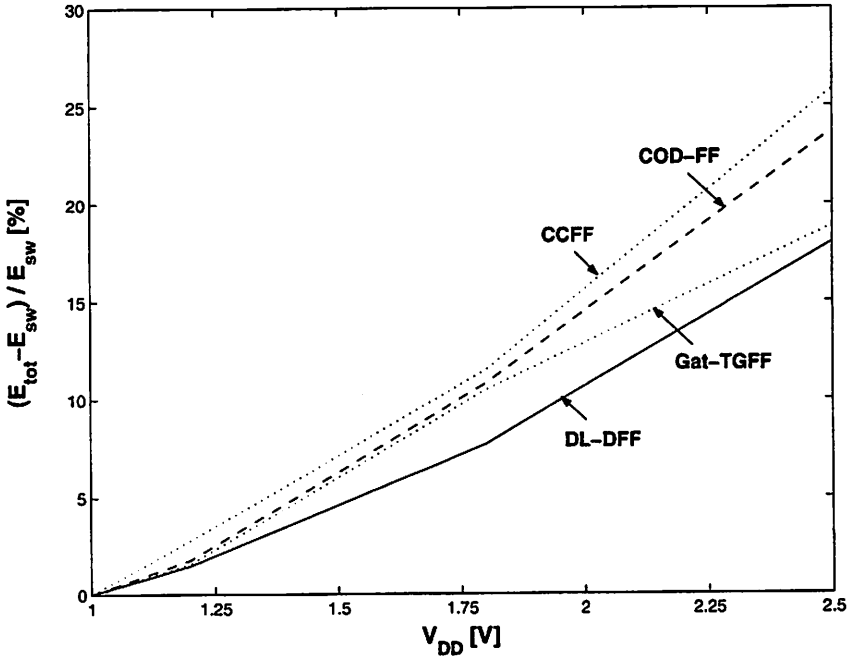


FIGURE 7.14 Short-circuit and subthreshold components of energy in flip-flops with internal clock gating as a function of supply voltage.

7.4 SUMMARY

Often times, designers find it difficult to engender a piece of hardware that interfaces with other subsystems on a chip. In this chapter we push the flip-flop designs up a level of abstraction and examine the impact of supply voltage scaling on flip-flop timing and energy parameters. An important result of this analysis shows that the behavior of the system presents even more complex issues that were not present in the original design of the flip-flops. In addition to the supply voltage, we also examined the impact of the clock-tree on the clock slopes. These two important features of the system greatly alter the energy-performance tradeoff profile derived in previous chapters.

Supply voltage scaling is the most efficient way to minimize energy consumption in digital circuits. Low voltage operation is desirable not only because the switching component of energy is reduced, but also because the short-circuit and subthreshold components are reduced *relative* to the switching component. The timing behavior of flip-flops and combinational logic scale the same way with supply voltage, so it is desirable and imperative to operate at the lowest supply voltage that yields the required computational throughput.

In previous analyses, we did not take into account the insertion delays for the flip-flops in the system. This is not something that designer can easily control because it is extremely difficult to obtain information about the clock tree network during the circuit design phase. The flip-flop behavior obtained during the circuit design phase assumes that the clock slopes are uniform throughout the circuit. However, when the clock tree is inserted into the system, irregularities in the clock tree network (RC wire delays, imbalanced loads, etc) create mismatches in the clock slopes which directly affects the system clock skew specifications. This information about the impact of mismatched clock slopes is essential to the designer of the clock network in order to satisfy the system timing requirements. An extremely important result was derived in this chapter, and that is that the system clock skew specification is a linear function of the mismatch in clock slopes. The impact of this result is that now, the clock network designer and the

circuit designer have a direct link with regards to system specifications and meeting the stringent energy-performance requirements. With information about the clock slope variation from the flip-flop designer, the clock network can be suitably designed to ensure that clock skew and other timing requirements are fully satisfied. Another important result from this analysis is that low voltage operation is beneficial for the clock network because subthreshold and short-circuit components of energy are minimized.

Chapter 8

CONCLUSION

In VLSI designs of yesterday, the roles of the system designer and circuit designer were clearly defined. A system designer worked mainly at the architectural level, whereas the circuit designer worked with transistors and gates. Any optimization would be done at strictly their appropriate levels of hierarchy. Today, however the roles have changed greatly and the distinction between the two slowly disappears. Circuit designers today must understand the critical issues at the system level in order to properly design their circuits for optimal interfacing, yielding the desired performance-energy characteristics.

A prime example of the need to combine the roles of system and circuit designers is apparent in the design of the clocked storage elements in various digital systems. With the ever-increasing complexity of digital systems and the push to densely pack many different logic blocks on the same chip, designers of memory elements must be able to produce a device that is capable of providing a proper communication channel between these blocks in a way that incurs minimal overhead and yields the optimal energy-performance numbers.

Memory elements are prevalent in virtually all useful circuits in various disciplines, ranging from microprocessors, communication equipment, etc. New products that involve highly sophisticated circuitry require well designed memory elements that ultimately interface other modules in the design. The idea, therefore, is to design one that allows the system architects more flexibility to design the clock system that brings the individual elements of the system together. The design process involves a deep understanding of the issues related to low-level devices and system architectures. For example, a superior flip-flop design is considered "superior" only if it offers low energy and good performance

not by itself, but when it is actively used in conjunction with other system level blocks.

To achieve a minimum energy solution of the clocking subsystem for a given performance constraint, there is a systematic approach which involves: 1) proper choice and optimization of memory elements, 2) determination of interface parameters with clock network and combinational logic designs, 3) clock tree optimization and implementation based on these interface parameters. The first step is the most important in the entire design process because the clock tree is always designed in accordance with the parameters obtained when a memory element topology has been selected. The proper topology really depends on the specific application. However, in all applications, the desire to minimize energy is the same. The performance, on the other hand, can be obtained at the expense of energy. Design of memory elements requires a solid understanding of the metrics used to characterize them.

There are many forms of metrics that offer insight into the circuit characteristics. Some are more relevant than others depending on the design goals. Metrics commonly found in literature often prove themselves to be insufficient in providing relevant information at the system level. For this reason, it is necessary to create new metrics that offer more insight about different aspects of a circuit. An example of a synthetic metric introduced is the *delay*, which is the only relevant flip-flop delay metric relative to the system clock period. Other synthetic metrics quantify the tolerance flip-flops have with regard to race conditions and provide better insight into flip-flop energy consumption.

An excellent flip-flop must be soundly sized to provide minimal energy consumption while maintaining an acceptable level of performance. In the past, this usually entails trial-and-error type sizing, which offers the designer very little insight on the important issues that largely dictate the energy consumption. However, designs today can be approached in an extremely procedural manner. Aggressive sizing to achieve the optimal solution is a reality in today's system designs, and the methods used to achieve this solution are highly systematic.

Time and time again, designs ascertain this fact because virtually all designs can be broken down into the same fundamental units that limit circuit performance. By approaching this limiting factor directly, we can employ a variety of techniques to optimize the circuit to achieve better results.

Understanding the various issues involved in circuit and system design such as glitching and interface parameters are priceless in the design process. Low energy design has and always will involve decisions that enhance one aspect of operation and degrade one aspect of another. While it is nearly impossible to know precisely everything that is occurring in large (and highly dynamic) systems, understanding just a few of the dominant parameters greatly eases the design process. The most important concept is that these known parameters can help designers make more informed choices about the tradeoffs associated at the system and circuit levels.

8.1 FUTURE RESEARCH DIRECTIONS

The results arrived thus far are by no means complete. The material presented here merely opens up a larger window of opportunity to explore other methods to minimize the overall energy consumption by the efficient utilization of multiple voltage and multiple threshold technologies.

There are certainly many more circuit topologies that have yet to be discovered. The circuits detailed in this work offer valuable insight about the importance of various parts of circuits. Another idea is to investigate the possibilities of integrating logic functions into memory elements which help minimize circuit area and relax timing constraints. Perhaps this direction will further open up a field of study that achieves even lower energy and better performance.

BIBLIOGRAPHY

[Bailey98] D. W. Bailey and B. J. Benschneider, "Clocking Design and Analysis for a 600-MHz Alpha Microprocessor," *IEEE J. Solid-State Circuits*, vol. 33, pp. 1627-1633, Nov. 1998.

[Balsara00] U. Ko and P. T. Balsara, "High-Performance Energy-Efficient D-Flip-Flop Circuits," *IEEE Trans. VLSI*, vol. 8, pp. 94-98, Feb. 2000.

[Burd00] T. Burd, T. Pering, A. Stratakos, and R. Brodersen, "A Dynamic Voltage-Scaled Microprocessor System," in *ISSCC Dig. Tech. Papers,* Feb. 2000, pp. 294-295.

[Chandr94] A. Chandrakasan, "Low-Power CMOS Design," Ph.D. Thesis, University of California at Berkeley, Aug. 1994.

[Chandr98] A. Chandrakasan and R. Brodersen, Eds. *Low-Power CMOS Design*, Piscataway, NJ: IEEE Press, 1998.

[Dally98] W. Dally and J. Poulton, *Digital Systems Engineering*, Cambridge Univ. Press, 1998.

[Gerosa94] G. Gerosa *et al.*, "A 2.2W, 80 MHz Superscalar RISC Microprocessor," *IEEE J. Solid-State Circuits*, vol. 29, pp. 1440-1454, Dec. 1994.

[Gonz96] R. Gonzalez and M. Horowitz, "Energy dissipation in general purpose microprocessors," *IEEE J. Solid-State Circuits*, vol. 31, pp. 1277-1284, Sept. 1996.

[Harris99] I. Sutherland, B. Sproul, and D. Harris, *Logical Effort: Designing Fast CMOS Circuits*, San Francisco, CA: Morgan Kaufmann, 1999.

[Harris] http://www3.hmc.edu/~harris/research/FO4.pdf

[Hash98] M. Hashimoto, H. Onodera, and K. Tamaru, "A power optimization method considering glitch reduction by gate sizing," in *ISPLED Dig. Tech. Papers,* Aug. 1998, pp. 221-226.

[Hodges88] D. Hodges and H. Jackson, *Analysis and Design of Digital Integrated Circuits*, McGraw Hill, 1988.

[Klass98] F. Klass, "Semi-Dynamic and Dynamic Flip-Flops with Embedded Logic," in *Symp. VLSI Circuits Dig. Tech. Papers*, June 1998, pp. 108-109.

[Kong00] B.-S. Kong, S.-S. Kim, and Y.-H. Jun, "Conditional Capture Flip-Flop for Statistical Power Reduction," in *ISSCC Dig. Tech. Papers*, Feb. 2000, pp. 290-291.

[Kuroda99] M. Hamada *et al.*, "Flip-Flop Selection Technique for Power-Delay Trade-off," in *ISSCC Dig. Tech. Papers*, Feb. 1999, pp. 270-271.

[Landers98] S. S. Mahant-Shetti and R. J. Landers, "Low Power Flip-Flop", *US Patent*, number 5,789,956, Aug. 1998.

[Matsui94] M. Matsui *et al.*, "A 200 MHz 13 mm^2 2-D DCT Macrocell Using Sense-Amplifying Pipeline Flip-Flop Scheme", *IEEE J. Solid-State Circuits*, vol. 29, pp. 1482-1490, Dec. 1994.

[Nikolic99] B. Nikolic *et al.*, "Sense Amplifier-Based Flip-Flop," in *ISSCC Dig. Tech. Papers*, Feb. 1999, pp. 282-283.

[Nogawa98] M. Nogawa and Y. Ohtomo, "A Data-Transition Look-Ahead DFF Circuit for Statistical Reduction in Power Consumption," *IEEE J. Solid-State Circuits*, vol. 33, pp. 702-706, May 1998.

[Partovi96] H. Partovi *et al.*, "Flow-Through Latch and Edge-Triggered Flip-flop Hybrid Elements," in *ISSCC Dig. Tech. Papers*, Feb. 1996, pp. 138-139.

[Rabaey96] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*, Upper Saddle River, NJ: Prentice Hall, 1996.

[Sakurai98] H. Kawaguchi and T. Sakurai, "A Reduced Clock-Swing Flip-Flop (RCSFF) for 63% Power Reduction," *IEEE J. Solid-State Circuits*, vol. 33, pp. 807-811, May 1998.

[Stojan99] V. Stojanovic and V. G. Oklobdzija, "Comparative Analysis of Master-Slave Latches and Flip-Flops for High-Performance and Low-Power Systems," *IEEE J. Solid-State Circuits*, vol. 34, pp. 536-548, Apr. 1999.

[Strollo00] A. G. M. Strollo, E. Napoli, and D. De Caro, "New Clock-Gating Technique for Low-Power Flip-flops," in *ISPLED Dig. Tech. Papers,* July 2000, pp. 114-119.

[Suzuki73] Y. Suzuki, K. Odagawa, and T. Abe, "Clocked CMOS calculator circuitry," *IEEE J. Solid-State Circuits*, vol. SC-8, pp. 462-469, Dec. 1973.

[Sven98] C. Svensson and J. Yuan, "Latches and Flip-flops for Low Power Systems," in *Low Power CMOS Design*, A. Chandrakasan and R. Brodersen, Eds. Piscataway, NJ: IEEE Press, 1998, pp. 233-238.

[Weste93] N. Weste and K. Eshragian, *Principles of CMOS VLSI Design: A Systems Perspective*, second ed. Reading, MA: Addison-Wesley, 1993.

[Wolf95] S. Wolf, *Silicon Processing for the VLSI Era, Volume 3 – The Submicron MOSFET*, Sunset Beach, CA: Lattice Press, 1995.

[Xiaohai98] Q. Xiaohai and C. Hongyi, "Discussion on the Low-power CMOS Latches and Flip-flops," in *Proc. Int. Conf. on Solid-State and Integrated Circuit Technology*, Oct. 1998, pp. 477-480.

[Yuan89] J. Yuan and C. Svensson, "High-speed CMOS circuit technique," *IEEE J. Solid-State Circuits*, vol. 24, pp. 62-70, Feb. 1989.

[Yuan97] J. Yuan and C. Svensson, "New Single-Clock CMOS Latches and Flipflops with Improved Speed and Power Savings," *IEEE J. Solid-State Circuits*, vol. 32, pp. 62-69, Jan. 1997.

[Yano90] K. Yano *et al.*, "A 3.8 ns CMOS 16x16-b Multiplier Using Complementary Pass-Transistor Logic," *IEEE J. Solid-State Circuits*, vol. 25, pp. 388-385, Apr. 1990.

Appendix A

MASTER-SLAVE LATCH-PAIRS

A.1 Transmission Gate Flip-Flop

TABLE A.1.1 TIMING PARAMETERS OF 4-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	3540	2618	1627	1159	726	539	430	363	318	287	254
$t_{C\text{-}Q,1\text{-}0}$ [ps]	4063	2907	1808	1280	788	574	451	376	326	292	257
$t_{setup,0}$ [ps]	640	520	320	220	150	130	110	100	80	80	70
$t_{setup,1}$ [ps]	230	180	140	100	70	60	50	40	40	40	30
$t_{hold,0}$ [ps]	800	640	480	360	180	130	90	70	60	60	50
$t_{hold,1}$ [ps]	-160	-100	-40	-30	-20	-20	-10	-10	-10	-10	0

TABLE A.1.2 SYNTHETIC TIMING PARAMETERS OF 4-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$D_{0\text{-}1}$ [ps]	3947	2929	1848	1317	832	626	501	421	374	341	297
$D_{1\text{-}0}$ [ps]	4906	3572	2218	1564	977	733	584	495	422	387	340
R_0 [ps]	2740	1978	1147	799	546	409	340	293	258	227	204
R_1 [ps]	4223	3007	1848	1310	808	594	461	386	336	302	257

TABLE A.1.3 BREAKDOWN OF ENERGY-PER-TRANSITION IN 4-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5

0-0

E_{CLK} [fJ]	6.1	7.8	10.0	12.5	18.4	25.8	35.0	45.9	58.6	73.3	100
0-1											
E_{CLK} [fJ]	6.1	7.8	10.0	12.5	18.4	25.8	35.0	45.9	58.6	73.3	100
E_{int} [fJ]	4.3	5.8	7.3	9.3	14.0	19.8	25.8	35.8	46.2	60.4	83.5
E_{ext} [fJ]	14.7	19.2	24.3	30	43.2	58.8	76.8	97.2	120	145.2	187.5
1-0											
E_{CLK} [fJ]	5.8	7.4	9.4	11.9	17.6	24.5	32.4	43.2	54.9	69.9	97.1
E_{int} [fJ]	8.0	10.6	13.7	17.1	25.4	36.3	48.4	63.4	81.2	101.5	137.9
1-1											
E_{CLK} [fJ]	5.8	7.4	9.4	11.9	17.6	24.5	32.4	43.2	54.9	69.9	97.1

TABLE A.1.4 IMPACT OF CLOCK SLOPE ON TIMING PARAMETERS OF TGFF. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	0.2	0.4	0.6	0.8	1.0	1.2	1.5	2.0	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	985	1042	1098	1147	1191	1232	1286	1369	1441
$t_{C\text{-}Q,1\text{-}0}$ [ps]	1109	1167	1221	1269	1312	1352	1408	1490	1561
$t_{setup\text{-}0}$ [ps]	250	190	150	130	110	90	60	30	0
$t_{setup\text{-}1}$ [ps]	150	90	50	-10	-50	-90	-140	-210	-260
$t_{hold\text{-}0}$ [ps]	190	230	250	270	290	310	340	410	480
$t_{hold\text{-}1}$ [ps]	-50	10	50	90	130	170	220	290	360

TABLE A.1.5 DELAY AND INTERNAL RACE IMMUNITY VS. CLOCK SLOPE. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	0.2	0.4	0.6	0.8	1.0	1.2	1.5	2.0	2.5
$D_{0\text{-}1}$ [ps]	1184	1184	1203	1194	1201	1204	1210	1227	1253
$D_{1\text{-}0}$ [ps]	1414	1415	1432	1462	1488	1510	1538	1595	1639
R_0 [ps]	1159	1157	1171	1179	1182	1182	1188	1200	1201
R_1 [ps]	795	812	848	877	901	922	946	959	961

A.2 Transmission Gate Flip-Flop with Input Gate Isolation

TABLE A.2.1 TIMING PARAMETERS OF 6-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	3966	2832	1757	1242	767	563	444	373	325	292	258
$t_{C\text{-}Q,1\text{-}0}$ [ps]	3425	2549	1583	1123	703	523	417	353	310	280	250
$t_{setup,0}$ [ps]	1380	880	580	400	270	210	170	150	130	120	110
$t_{setup,1}$ [ps]	1120	820	480	390	260	200	160	140	130	110	100
$t_{hold,0}$ [ps]	480	460	280	210	140	110	-90	-80	-70	-60	-50
$t_{hold,1}$ [ps]	240	240	150	110	-80	-70	-60	-50	-50	-40	-40

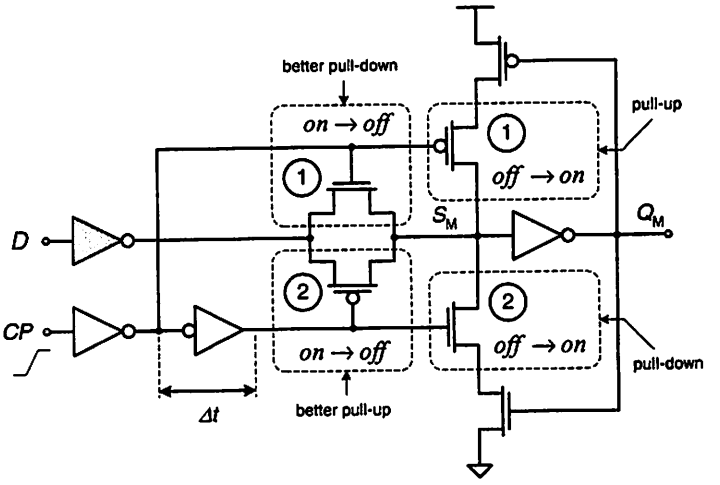


FIGURE A.2.1 Active circuit that defines setup and hold times in gate isolated TGFF.

In this flip-flop the hold time of logic "1" is greater than the hold time of logic "0", Fig. A.2.1. An additional inverter at the input makes the input *D* to favor a *1-0* transition because the transmission gate favors a *0-1* at its input. Another difference from 4-stage realization is that both master and slave latches are now non-inverting. As a consequence, when *D* switches *0-1* S_M pulls down against weak pull-up of the feedback circuit until S_M pulls down enough such that Q_M turns *off* pull-down path of the feedback circuit and turns *on* the pull-up feedback path. When *D* switches *1-0* S_M pulls up against weak pull-down of the feedback circuit.

Output of the master latch, Q_M, faces the input of transmission gate in the slave latch which captures a *0-1* transition faster than a *1-0* transition. Master latch, on the other hand, captures a *1-0* transition faster than a *0-1* transition. Therefore for both *0-1* and *1-0* transition at the input, *D*, one latch transmits its good side and another latch transmits its bad side, which results in longer overall setup time of logic "0" and logic "1" at the input.

TABLE A.2.2 SYNTHETIC TIMING PARAMETERS OF 6-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$D_{0\text{-}1}$ [ps]	5284	3794	2325	1694	1065	791	626	531	471	416	371
$D_{1\text{-}0}$ [ps]	4976	3556	2242	1579	1008	760	608	521	456	414	373
R_0 [ps]	4446	3292	2037	1452	907	673	534	453	395	352	308
R_1 [ps]	3665	2789	1733	1233	783	593	477	403	360	320	290

TABLE A.2.3 BREAKDOWN OF ENERGY-PER-TRANSITION IN 6-STAGE TGFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
0-0											
E_{CLK} [fJ]	6.1	7.8	10.0	12.5	18.4	25.9	34.6	45.5	58.6	73.5	102
0-1											
E_{CLK} [fJ]	6.1	7.8	10.0	12.5	18.4	25.9	34.6	45.5	58.6	73.5	102
E_{int} [fJ]	10.6	14.1	18.1	22.7	32.3	44.9	60.3	81.2	101.0	122.6	163.6
E_{ext} [fJ]	14.7	19.2	24.3	30	43.2	58.8	76.8	97.2	120	145.2	187.5
1-0											
E_{CLK} [fJ]	6.1	7.8	10.0	12.5	18.2	25.7	34.7	45.4	58.4	73.3	101
E_{int} [fJ]	11.8	15.7	20.2	25.2	37.1	51.7	68.8	89.5	112.8	140.1	187.1
1-1											
E_{CLK} [fJ]	6.1	7.8	10.0	12.5	18.2	25.7	34.7	45.4	58.4	73.3	101

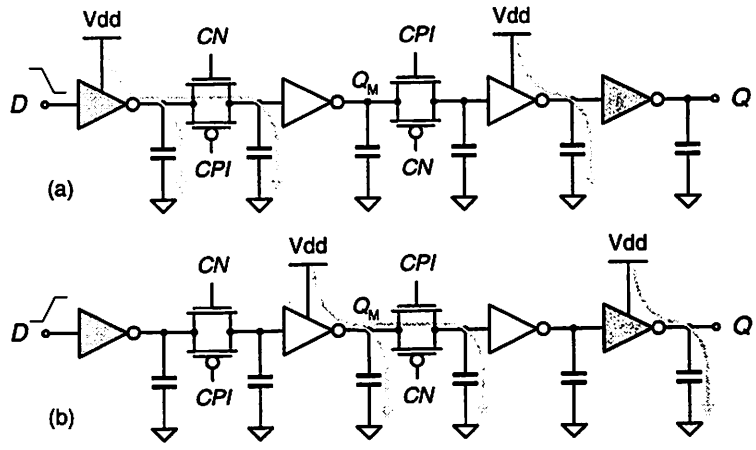


FIGURE A.2.2 Energy consuming internal nodes for (a) 1-0, (b) 0-1 input data transition.

TABLE A.2.4 IMPACT OF CLOCK SLOPE ON TIMING PARAMETERS OF TGFF. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	0.2	0.4	0.6	0.8	1.0	1.2	1.5	2.0	2.5
$t_{setup-0}$ [ps]	490	430	370	330	290	250	200	110	40
$t_{setup-1}$ [ps]	450	390	350	310	290	250	220	190	160
t_{hold-0} [ps]	-210	-170	-150	-110	-90	-70	-60	-10	60
t_{hold-1} [ps]	-150	-90	-50	-10	30	70	120	190	240

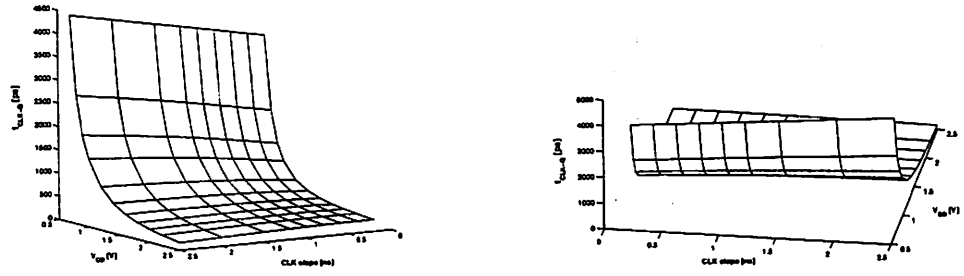


FIGURE A.2.3 Dependency of *Clk-Q* delay on clock slope and supply voltage.

TABLE A.2.5 DELAY AND INTERNAL RACE IMMUNITY VS. CLOCK SLOPE. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	0.2	0.4	0.6	0.8	1.0	1.2	1.5	2.0	2.5
D_{0-1} [ps]	1577	1577	1593	1603	1628	1630	1656	1716	1762
D_{1-0} [ps]	1489	1492	1489	1500	1504	1508	1516	1514	1521
R_0 [ps]	1283	1300	1334	1341	1364	1384	1428	1463	1466
R_1 [ps]	1101	1101	1116	1124	1126	1128	1133	1147	1170

A.3 C²MOS Flip-Flop

TABLE A.3.1 TIMING PARAMETERS OF C²MOS FLIP-FLOP. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	2553	2072	1251	868	529	529	387	303	253	219	195
$t_{C\text{-}Q,1\text{-}0}$ [ps]	2071	1806	1086	758	467	467	348	275	233	204	185
$t_{setup,0}$ [ps]	2160	1210	820	600	390	300	240	210	180	170	150
$t_{setup,1}$ [ps]	1370	940	610	450	310	240	200	180	160	150	130
$t_{hold,0}$ [ps]	-320	-230	-200	-150	-110	-110	-90	-80	-60	-60	-50
$t_{hold,1}$ [ps]	-30	-30	-30	-30	-20	-20	-20	-20	-10	-10	-10

TABLE A.3.2 SYNTHETIC TIMING PARAMETERS OF C²MOS FLIP-FLOP. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$D_{0\text{-}1}$ [ps]	3831	3096	1924	1362	865	865	647	518	445	390	355
$D_{1\text{-}0}$ [ps]	4335	3106	1960	1396	880	880	665	529	454	394	364
R_0 [ps]	2391	2036	1286	908	577	577	438	355	293	264	235
R_1 [ps]	2633	2122	1281	898	549	549	407	323	263	229	205

TABLE A.3.3 BREAKDOWN OF ENERGY-PER-TRANSITION IN C²MOS FF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
0-0											
E_{CLK} [fJ]	11.9	15.0	19.4	24.1	35.9	49.4	68.5	89.2	114	142	197
0-1											
E_{CLK} [fJ]	11.9	15.0	19.4	24.1	35.9	49.4	68.5	89.2	114	142	197
E_{int} [fJ]	8.4	11.3	14.6	18.5	26.9	38.3	49.4	67.2	83.7	106.4	138.5

E_{ext} [fJ]	14.7	19.2	24.3	30	43.2	58.8	76.8	97.2	120	145.2	187.5
1-0											
E_{CLK} [fJ]	11.9	15.2	19.5	24.5	35.7	50.1	67.8	89.4	114	144	201
E_{int} [fJ]	9.1	12.2	16.0	20.4	29.9	42.5	56.5	73.1	92.3	112.8	147.6
1-1											
E_{CLK} [fJ]	11.9	15.2	19.5	24.5	35.7	50.1	67.8	89.4	114	144	201

TABLE A.3.4 IMPACT OF CLOCK SLOPE ON TIMING OF C²MOS FF. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	0.2	0.4	0.6	0.8	1.0	1.2	1.5	2.0	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	710	630	570	530	470	430	380	290	220
$t_{C\text{-}Q,1\text{-}0}$ [ps]	510	450	410	390	350	330	320	270	240
$t_{setup\text{-}0}$ [ps]	-150	-130	-90	-70	-50	-40	-20	10	40
$t_{setup\text{-}1}$ [ps]	-90	-30	10	50	90	130	180	270	340
$t_{hold\text{-}0}$ [ps]	710	630	570	530	470	430	380	290	220
$t_{hold\text{-}1}$ [ps]	510	450	410	390	350	330	320	270	240

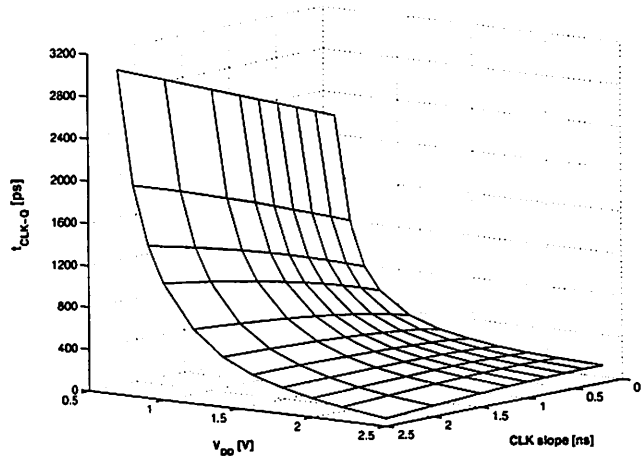


FIGURE A.3.1 Dependency of *Clk-Q* delay on clock slope and supply voltage.

TABLE A.3.5 DELAY AND INTERNAL RACE IMMUNITY VS. CLOCK SLOPE. V_{DD} = 1V, C_{OUT} = 4SL

CLK slope [ns]	0.2	0.4	0.6	0.8	1.0	1.2	1.5	2.0	2.5
$D_{0\text{-}1}$ [ps]	1233	1237	1257	1291	1300	1325	1377	1423	1474
$D_{1\text{-}0}$ [ps]	1320	1302	1301	1313	1302	1307	1320	1322	1331
R_0 [ps]	731	770	786	816	842	875	915	973	1018
R_1 [ps]	778	779	797	808	815	818	827	828	835

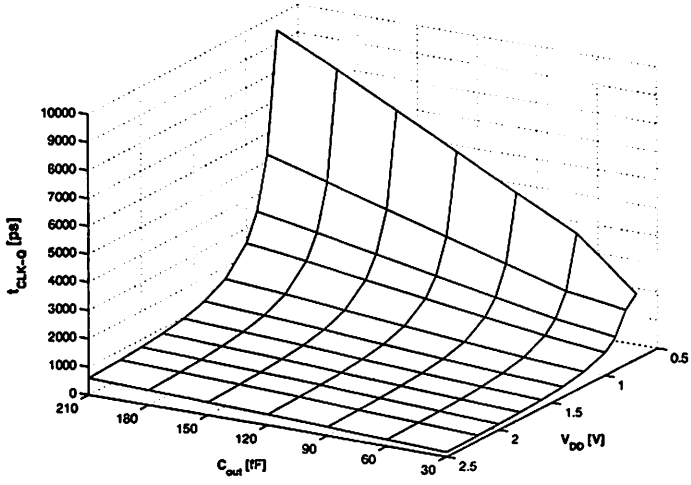


FIGURE A.3.2 Variation of clock-to-output delay with output load.

Appendix B

FLIP-FLOPS WITH INTERNAL CLOCK GATING

Local vs Global Clock Gating: Data look-ahead flip-flop technique is an attractive choice when there is a wide range of switching probabilities within a single clock domain. When a global clock is enabled, data look-ahead flip-flops can be placed in local sub-domains that have low data switching probability, while conventional flip-flops with global clock gating are placed only in sub-domains with high data switching probabilities.

B.1 Data-Transition Look-Ahead D Flip-Flop (DL-DFF)

TABLE B.1.1 TIMING PARAMETERS OF DL-DFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	6839	4371	2751	1957	1236	889	703	590	514	462	410
$t_{C\text{-}Q,1\text{-}0}$ [ps]	6256	4110	2615	1880	1210	885	708	600	528	478	427
$t_{setup,0}$ [ps]	-	640	440	330	240	180	150	140	120	110	100
$t_{setup,1}$ [ps]	930	600	420	320	230	170	140	130	110	100	100
$t_{hold,0}$ [ps]	-	4960	2960	2040	1180	830	640	520	440	380	320
$t_{hold,1}$ [ps]	-	4420	2640	1800	1070	760	590	490	410	360	310

Energy consumption of the entire flip-flop is summarized in Table B1.2.

TABLE B.1.2 ENERGY-PER-TRANSITION OF ENTIRE DL-DFF WITH GATING. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}0}$ [fJ]	34.4	23.7	30.9	39.0	57.8	81.0	109	141	179	222	296

$E_{0\text{-}1}$ [fJ]	72.5	76.3	97.9	122	179	251	333	428	538	660	900
$E_{1\text{-}0}$ [fJ]	60.7	58.7	76.0	94.9	140	196	263	342	435	542	730
$E_{1\text{-}1}$ [fJ]	34.4	23.7	30.9	39.0	57.8	81.0	109	141	179	222	296

TABLE B.1.3 ENERGY-PER-TRANSITION OF DL-DFF WITHOUT GATING. $C_{OUT} = 4SL$

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}0}$ [fJ]	5.5	7.1	9.1	11.4	16.8	23.5	31.6	41.3	52.8	66.4	91.5
$E_{0\text{-}1}$ [fJ]	30.9	40.1	51.2	63.6	92.1	127	168	217	273	338	453
$E_{1\text{-}0}$ [fJ]	17.9	23.4	30.0	37.5	55.0	76.7	103	134	169	212	287
$E_{1\text{-}1}$ [fJ]	5.6	7.1	9.1	11.4	16.8	23.6	31.9	41.5	52.9	66.6	91.3

TABLE B.1.4 ENERGY OVERHEAD (PG + CC + DL) IN DL-DFF

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}1}$ [fJ]	41.5	36.1	46.7	58.8	86.9	123	165	210	265	323	447
$E_{1\text{-}0}$ [fJ]	42.7	35.3	46.0	57.4	85.2	119	160	209	266	330	444

This is the total energy overhead in DL-DFF and it is independent of C_{out}. A pulse generator (PG) is commonly shared among several flip-flops, so further energy breakup is needed to understand where the energy goes exactly. Energy consumed by the PG is estimated by simulation of a stand-alone PG loaded with capacitance C_{in}(CC) that *CKI* "sees" when "looking" into the clock control circuit (CC). This capacitance is measured by simulation and found to be 3.5fF.

TABLE B.1.5 ENERGY CONSUMPTION IN PULSE GENERATOR (PG)

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
E_{PG} [fJ]	32.7	21.4	28.1	35.5	52.8	74.2	100	130	165	205	274
E_{Cin} [fJ]	1.7	2.2	2.8	3.5	5.0	6.9	9.0	11.3	14.0	16.9	21.9

TABLE B.1.6 ENERGY OVERHEAD (CC + DL) IN DL-DFF

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}1}$ [fJ]	7.1	12.4	15.8	19.8	29.1	42	56	69	86	101	151
$E_{1\text{-}0}$ [fJ]	8.3	11.7	15.1	18.4	27.4	38	52	67	87	109	148

TABLE B.1.7 ENERGY CONSUMPTION IN PULSE GENERATOR (PG)

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
E_{PG} [fJ]	32.7	21.4	28.1	35.5	52.8	74.2	100	130	165	205	274
E_{Cin} [fJ]	1.7	2.2	2.8	3.5	5.0	6.9	9.0	11.3	14.0	16.9	21.9

E_{PG} is the energy burned in the pulse generator itself, and E_{Cin} is the energy burned in driving C_{in}(CC).

TABLE B.1.8 SYNTHETIC TIMING PARAMETERS OF DL-DFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$D_{0\text{-}1}$ [ps]	8111	5190	3309	2375	1528	1104	878	749	650	586	531
$D_{1\text{-}0}$ [ps]	-	4956	3186	2304	1511	1109	893	770	674	612	548
R_0 [ps]	-	-850	-345	-160	30	55	68	80	88	98	107
R_1 [ps]	-	-49	111	157	166	129	113	100	104	102	100

TABLE B.1.9 ENERGY BREAKDOWN IN DL-DFF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
0-0											
E_{PG} [fJ]	32.7	21.4	28.1	35.5	52.8	74.2	100	130	165	205	274
E_{Cin} [fJ]	1.7	2.2	2.8	3.5	5.0	6.9	9.0	11.3	14.0	16.9	21.9
0-1											
E_{CLK} [fJ]	5.5	7.1	9.1	11.4	16.8	23.5	31.6	41.3	52.8	66.4	91.5

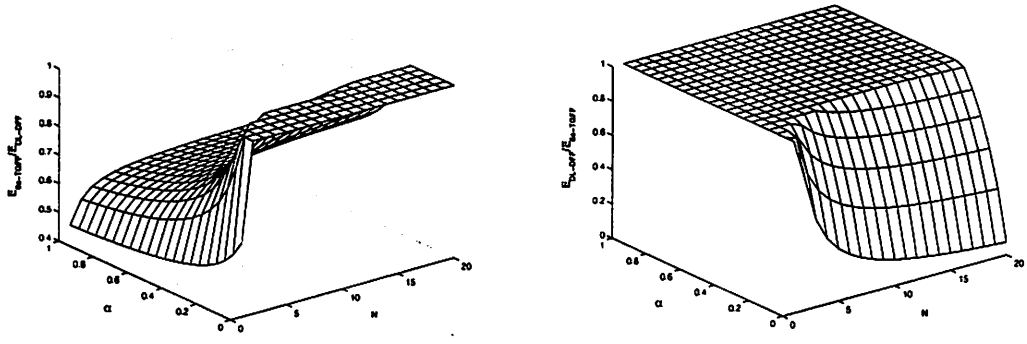
E_{int} [fJ]	10.7	13.8	17.8	22.2	32.0	45.0	59.9	78.9	100	126	174
E_{ext} [fJ]	14.7	19.2	24.3	30	43.2	58.8	76.8	97.2	120	145.2	187.5
E_G [fJ]	7.1	12.4	15.8	19.8	29.1	42	56	69	86	101	151
E_{PG} [fJ]	32.7	21.4	28.1	35.5	52.8	74.2	100	130	165	205	274
E_{Cin} [fJ]	1.7	2.2	2.8	3.5	5.0	6.9	9.0	11.3	14.0	16.9	21.9

1-0

E_{CLK} [fJ]	5.6	7.1	9.1	11.4	16.8	23.6	31.9	41.5	52.9	66.6	91.3
E_{int} [fJ]	12.3	16.3	20.9	26.1	38.2	53.2	70.7	92.1	116	145	195
E_G [fJ]	8.3	11.7	15.1	18.4	27.4	38	52	67	87	109	148
E_{PG} [fJ]	32.7	21.4	28.1	35.5	52.8	74.2	100	130	165	205	274
E_{Cin} [fJ]	1.7	2.2	2.8	3.5	5.0	6.9	9.0	11.3	14.0	16.9	21.9

1-1

E_{PG} [fJ]	32.7	21.4	28.1	35.5	52.8	74.2	100	130	165	205	274
E_{Cin} [fJ]	1.7	2.2	2.8	3.5	5.0	6.9	9.0	11.3	14.0	16.9	21.9



(a) (b)

FIGURE B.1.1 Ratio of energy consumption (rounded to 1) in DL-DFF_4x and 6-stage TGFF_4x, over range of switching activities and number of flip-flops driven by a single pulse generator in DL-DFF. (a) $E_{6s\text{-}TGFF}/E_{DL\text{-}DFF}$, (b) $E_{DL\text{-}DFF}/E_{6s\text{-}TGFF}$. V_{DD} = 1V.

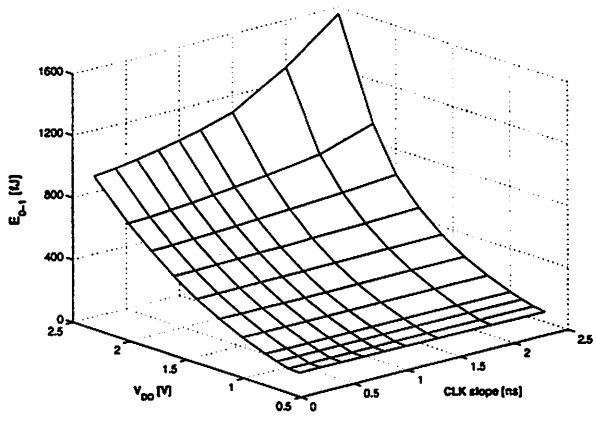


FIGURE B.1.2 Energy-per-transition as a function of the clock slope and supply voltage.

B.2 Clock-on-Demand Flip-Flop (COD-FF)

TABLE B.2.1 TIMING PARAMETERS OF COD-FF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$t_{C\text{-}Q,0\text{-}1}$ [ps]	3845	2732	1709	1216	779	561	445	375	328	294	261
$t_{C\text{-}Q,1\text{-}0}$ [ps]	5500	3629	2277	1612	1008	715	558	463	400	357	314
$t_{setup,0}$ [ps]	2840	1720	1120	800	540	380	320	260	240	220	200
$t_{setup,1}$ [ps]	1930	1040	770	570	390	290	240	210	180	170	150
$t_{hold,0}$ [ps]	4820	3560	2100	1440	900	620	480	400	340	320	280
$t_{hold,1}$ [ps]	3300	2830	1670	1150	740	510	400	330	280	250	230

Energy consumption of the entire COD-FF is summarized in table B.2.2.

TABLE B.2.2 ENERGY-PER-TRANSITION OF COD-FF. C_{OUT} = 4SL (COMPOSITE AND-NOR)

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}0}$ [fJ]	0	0	0	0	0	0	0	0	0	0	0
$E_{0\text{-}1}$ [fJ]	37.8	49.6	63.4	79.0	115	159	211	273	343	423	563

$E_{1\text{-}0}$ [fJ]	24.8	33.1	42.6	53.3	78.9	111	152	202	261	332	460
$E_{1\text{-}1}$ [fJ]	0	0	0	0	0	0	0	0	0	0	0

TABLE B.2.3 ENERGY-PER-TRANSITION OF COD-FF WITHOUT GATING. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}0}$ [fJ]	4.5	5.7	7.3	9.1	13.4	19.1	25.8	33.8	44.0	55.8	77.1
$E_{0\text{-}1}$ [fJ]	24.6	32.0	40.7	50.4	73.2	101	134	172	217	268	356
$E_{1\text{-}0}$ [fJ]	10.7	13.8	17.6	22.1	32.6	46.1	62.1	83.1	108	137	190
$E_{1\text{-}1}$ [fJ]	4.5	5.7	7.3	9.1	13.5	18.9	25.6	34.1	44.2	56.2	77.9

TABLE B.2.4 ENERGY OVERHEAD (DL + PG) IN COD-FF

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$E_{0\text{-}1}$ [fJ]	13.2	17.6	22.7	28.6	42.1	58.4	77.5	100	126	155	207
$E_{1\text{-}0}$ [fJ]	14.1	19.3	24.9	31.2	46.2	65.2	90.1	118	153	195	269

TABLE B.2.5 SYNTHETIC TIMING PARAMETERS OF COD-FF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
$D_{0\text{-}1}$ [ps]	5967	3909	2564	1847	1208	879	707	603	524	479	424
$D_{1\text{-}0}$ [ps]	8615	5530	3511	2493	1598	1131	906	746	660	595	530
R_0 [ps]	680	69	177	172	108	95	78	63	60	37	34
R_1 [ps]	545	-98	39	66	39	51	45	45	48	44	31

TABLE B.2.6 ENERGY BREAKDOWN IN COD-FF. C_{OUT} = 4SL

V_{DD} [V]	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.5
0-0											
	0	0	0	0	0	0	0	0	0	0	0

0-1

E_{CLK} [fJ]	4.5	5.7	7.3	9.1	13.4	19.1	25.8	33.8	44.0	55.8	77.1
E_{int} [fJ]	5.4	7.1	9.1	11.3	16.6	22.9	31.3	41.3	52.8	66.8	91.5
E_{ext} [fJ]	14.7	19.2	24.3	30.0	43.2	58.8	76.8	97.2	120	145	188
E_{OH} [fJ]	13.2	17.6	22.7	28.6	42.1	58.4	77.5	100	126	155	207

1-0

E_{CLK} [fJ]	4.5	5.7	7.3	9.1	13.5	18.9	25.6	34.1	44.2	56.2	77.9
E_{int} [fJ]	6.2	8.0	10.3	12.9	19.2	27.2	36.6	49.1	63.9	81.2	112
E_{OH} [fJ]	14.1	19.3	24.9	31.2	46.2	65.2	90.1	118	153	195	269

1-1

	0	0	0	0	0	0	0	0	0	0	0

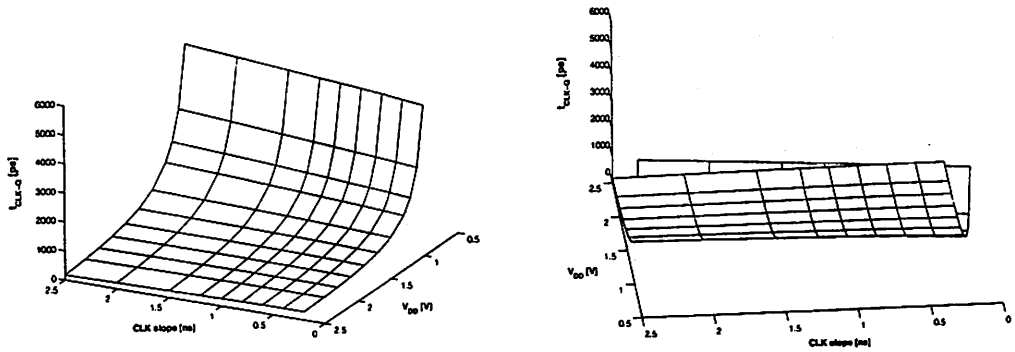


FIGURE B.2.1 Dependency of *Clk-Q* delay on clock slope and supply voltage.

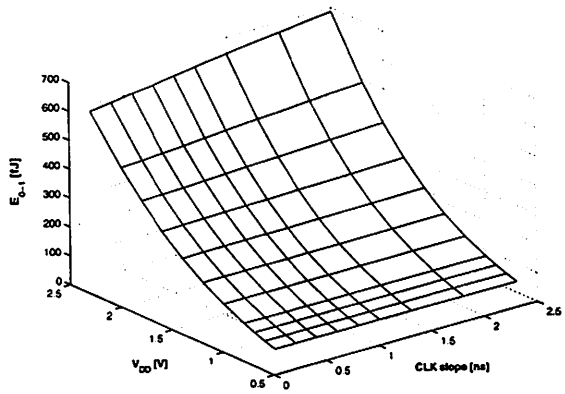


FIGURE B.2.2 Energy per transition vs. clock slope and supply voltage.