SURVEY AND EVALUATION OF LOW-POWER FLIP-FLOPS

Ahmed Sayed and Hussain Al-Asaad
Department of Electrical & Computer Engineering
University of California
Davis, CA, U.S.A.

***Abstract*—*We survey a set of flip-flops designed for low power and high performance. We highlight the basic features of these flip-flops and evaluate them based on timing characteristics, power consumption, and other metrics.*

1 INTRODUCTION

Low power consumption has become a highly important design concern in this era and will become more and more important as we move to all mobile computing and communications. The transistor density of IC is growing at Moore's law rate and the incomparable battery advances will mandate lower power methodologies and designs.

Most of the current designs are synchronous which implies that flip-flops and latches are involved in one way or another in the data and control paths. One of the challenges of low power methodologies for synchronous systems is the power consumption of these flip-flops and latches. It is important to save power in these flip-flops and latches without compromising state integrity or performance.

Several researchers have worked on low power flip-flop design, but they are mostly focused on one or a few types of flip-flops or applications. The need for comparing different designs and approaches is the main motivation for this paper. The main trade-offs of any flip-flop are very important for a design engineer when designing a circuit or for a tool that automates the process of design. The rest of this paper is organized as follows. Section 2 presents background information about flip-flop design and characteristics. Section 3 presents the flip-flop circuits surveyed with a short description of each flip-flop and Section 4 presents the simulation and evaluation results of these flip-flops. Finally, Section 5 presents some remarks and conclusions.

2 BACKGROUND

2.1 SOURCES OF POWER CONSUMPTION

There are three major sources of power consumption in a digital CMOS circuit. These are summarized in the following equation:

$$P_{total} = p_t(C_L V V_{dd} f_{clk}) + I_{sc} V_{dd} + I_{leakage} V_{dd} \quad (1)$$

The first term represents the switching component of power, where C_L is the effective switched loading capacitance, f_{clk} is the clock frequency and p_t is the probability that a power consuming transition occurs (referred to as the activity factor in other publications). In most cases, the voltage swing V is the same as the supply voltage V_{dd}. However, in some logic design styles such as in pass-transistor logic, the voltage swing on some internal nodes may be slightly less. It is important to point out that the effect of internal glitching (to be discussed later) shows as a component of the switching power consumption. The second term is caused by the direct path short circuit current I_{sc}, which arises when both the NMOS and PMOS transistors or networks are simultaneously active or on, conducting current from the supply V_{dd} to ground. Finally, a factor that is growing more and more important as we develop deep submicron technologies, leakage current $I_{leakage}$, which can arise from substrate injection, gate leakage and sub-threshold effects. $I_{leakage}$ is primarily determined by the CMOS fabrication process technology characterization.

We can observe from (1) that the power consumption of a circuit depends strongly on its structure and input data statistics. All the nodes of a circuit contribute to the total power consumption of the circuit so (1) should be applied to each and every node at a micro scale. Or a designer might like to break it down to internal and external components, which identify the internal inherent power consumption of the circuit and the external effect of the load on the power consumption.

The dominant term in a well-designed circuit is the switching component, thus the low-power design goal becomes the task of minimizing $p_t(C_L V V_{dd} f_{clk})$, while retaining the required functionality and identifying the cost of such minimizations in terms of area and/or performance.

The power-delay product (PDP) can be viewed as the amount of energy expended in each switching event and is thus particularly important in comparing the power consumption of various circuits and design styles. Assuming that the full swing switching component of (1) is dominant, this metric becomes:

$$PDP = p_t(C_L V V_{dd} f_{clk}) / f_{clk} = p_t(C_L V_{dd}^2) \quad (2)$$

A more performance oriented metric for circuits and design styles would be the energy-delay product. This is considered if performance is of a higher importance and priority than power consumption. This will not be used here since low power is the main focus of this paper.

2.2 BASICS OF SEQUENTIAL ELEMENTS

Sequential elements are mainly used to store computation result values for future use. At the minimal level of storage an element should be able to store a logic "1" or "0" reliably.

Transitions on the inputs of a flip-flop may or may not lead to a state change. When input transitions do not change the state, the internal switching inside the flip-flop consumes some power. On the other hand, when the input transitions do change the state, a bigger amount of power is consumed.

Flip-flops can be classified in several ways: dynamic vs. static, square-wave vs. pulsed, conditional vs. non-conditional, and depending on the logic style used. In this paper we consider different flip-flops with different classifications.

2.3 Flip-Flop Comparison Metrics

There are several basic performance metrics that are used to qualify a flip-flop and compare it to other designs. These metrics are:

- Clock-to-Q delay: Propagation delay form the clock terminal to the output *Q* terminal. This is assuming that the data input *D* is set early enough with respect to the effective edge of the clock input signal.
- Setup time: The minimum time needed between the *D* input signal change and the triggering clock signal edge on the clock input. This metric guarantees that the output will follow the input in worst case conditions of process, voltage ad temperature (PVT). This assumes that the clock triggering edge and pulse has enough time to capture the data input change.
- Hold time: The minimum time needed for the *D* input to stay stable after the occurrence of the triggering edge of the clock signal. This metric guarantees that the output *Q* stays stable after the triggering edge of the clock signal occurs, under worst PVT conditions. This metric assumes that the *D* input change happened at least after a minimum delay from the previous *D* input change, this minimum delay is the setup time of the flip-flop.

Historically speaking, library developers try their best to minimize the setup time requirement of flip-flops and the Clock-to-Q delay since most synchronous designs are targeting the most design performance at hand. Specifically in pipelined designs, where flip-flop "Setup time + Clock-to-Q delay" is the main constraint of the maximum clock frequency of operation for a given function, which in turn mandates the number of stages needed to perform the required function and affects in turn the latency and throughput of the whole design.

Hold time was and still is not a main target of designers since it is not as critical as the setup time since it does not limit the speed of flip-flop based designs. On the other hand, hold time is very critical in latch-based designs.

2.4 Regions of Flip-Flop Operation

There are three regions of flip-flop operation, of which only one region is acceptable for a sequential design to function correctly. These regions are:

- Stable region: Where the setup and hold times of a flip-flop are met and the Clock-to-Q delay is not dependent on the D-to-Clock delay. This is the required region of operation.
- Metastable region: As D-to-Clock delay decreases, at a certain point the Clock-to-Q delay starts to rise exponentially and ends in failure. The Clock-to-Q delay is indeterministic and this might cause intermittent failures and behaviors which are very difficult to debug in real circuits.
- Failure region: Where changes in data are unable to be transferred to the output of the flip-flop.

Figure 1 illustrates the different regions of flip-flop operation. The optimal setup time noted on the graph would be the highest performance D-to-Clock delay to accomplish fastest D-to-Q delay. Due to the steep curve to the left of that point not all library developers would target this value. Instead, they would prefer adding guard bands to any library cell or design to guarantee stability and reliability.

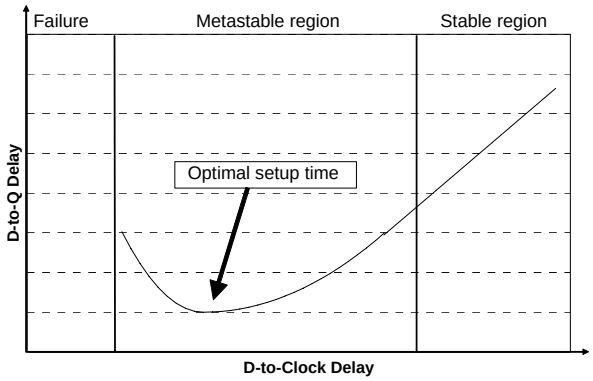


Figure 1 Flip-Flop regions of operation.

2.5 Hazards and Glitches

We define a glitch to be any spurious transient signal in combinational circuits. There are various phenomena that can cause glitches and the main one is hazards in combinational circuits. If the output signals for a combinational network depend on the internal circuit delays, elements and interconnects, as well as on input signals, the circuit is said to contain a hazard. There might be other causes of hazards in a circuit, for example the relative delays of the asynchronous inputs might exacerbate a hazard scenario which was not supposed to occur. Unequal delay paths in a circuit are a very common cause of hazards in combinational circuits. Interconnect delays are becoming more and more significant with submicron technologies and the balancing of different delay paths through the circuit is becoming a more important practice. There are several types of hazards [7] that can be classified as static and dynamic, or function and logic hazards. Function hazards are inherent to the function being implemented and occur in any implementation, logic hazards are specific to a particular implementation of a function and could still occur if function hazards are avoided.

Function hazards are avoided by restricting the inputs transitions to single variable changes, which is the fundamental mode of operation. Logic hazards are mainly avoided by choosing a different implementation or adding redundancy to the network used. Logic hazards are apparent in circuits with reconvergent signals. Some hazards could be removed by equalizing delay paths in the circuit at hand.

Hazards and glitches could be a nightmare to an asynchronous circuit designer, since they would cause misfiring of different events and cause system failure. Fortunately the synchronous design paradigm alleviates that headache and nightmare by giving enough settling time for all intermediate transient values called the setup time of a flip-flop or latch before the clock event, which brings that stable value to the outside world as output of a flip-flop or latch with no glitches. The unfortunate part of that is the all these glitches and hazards still cause unnecessary power consumption in the form of short circuit current. Reducing glitching and switching is an area of ongoing research.

3 SURVEYED CIRCUITS

The flip-flop circuits shown in Figure 2 and Figure 3 are extracted from references [1][2][3][4][5]. They were built using cadence schematic capture Virtuoso tool and sized for minimum size to function correctly. The following is a short description of the flip-flop circuits.

F01 is the Power PC master-slave latch. It is one of the fastest classical structures and its main advantage is the short direct path and low power feedback. The large load on the clock will greatly affect the total power consumption of the flip-flop. This flip-flop is the transmission gate flip-flop, it has a fully static master–slave structure, which is constructed by cascading two identical pass gate latches and provides a short clock to output latency. It does have a bad data to output latency because of the positive setup time. And its sensitivity to clock signal slopes and data feed through is another concern when using it.

F02 is the modified standard dynamic C^2MOS master-slave latch that has shown good low power features, like small clock load and low power feedback. The modified C^2MOS is also robust to clock signal slopes.

F03 is the hybrid–latch flip-flop (HLFF) that is one of the fastest flip-flop structures. It is robust to clock signal slopes, but it does have a positive hold time. This is very suitable for high performance systems.

F04 is another hybrid flip-flop, the semi-dynamic flip-flop (SDFF). It is one of the fastest structures if not the fastest of all the flip-flops described in this paper. It does have a large clock load and large effective precharge capacitance which result in a slightly high power consumption. This is still best suited for high performance designs, though its power consumption is moderate.

F05 is the K6 edge-triggered latch (ETL) with the reset

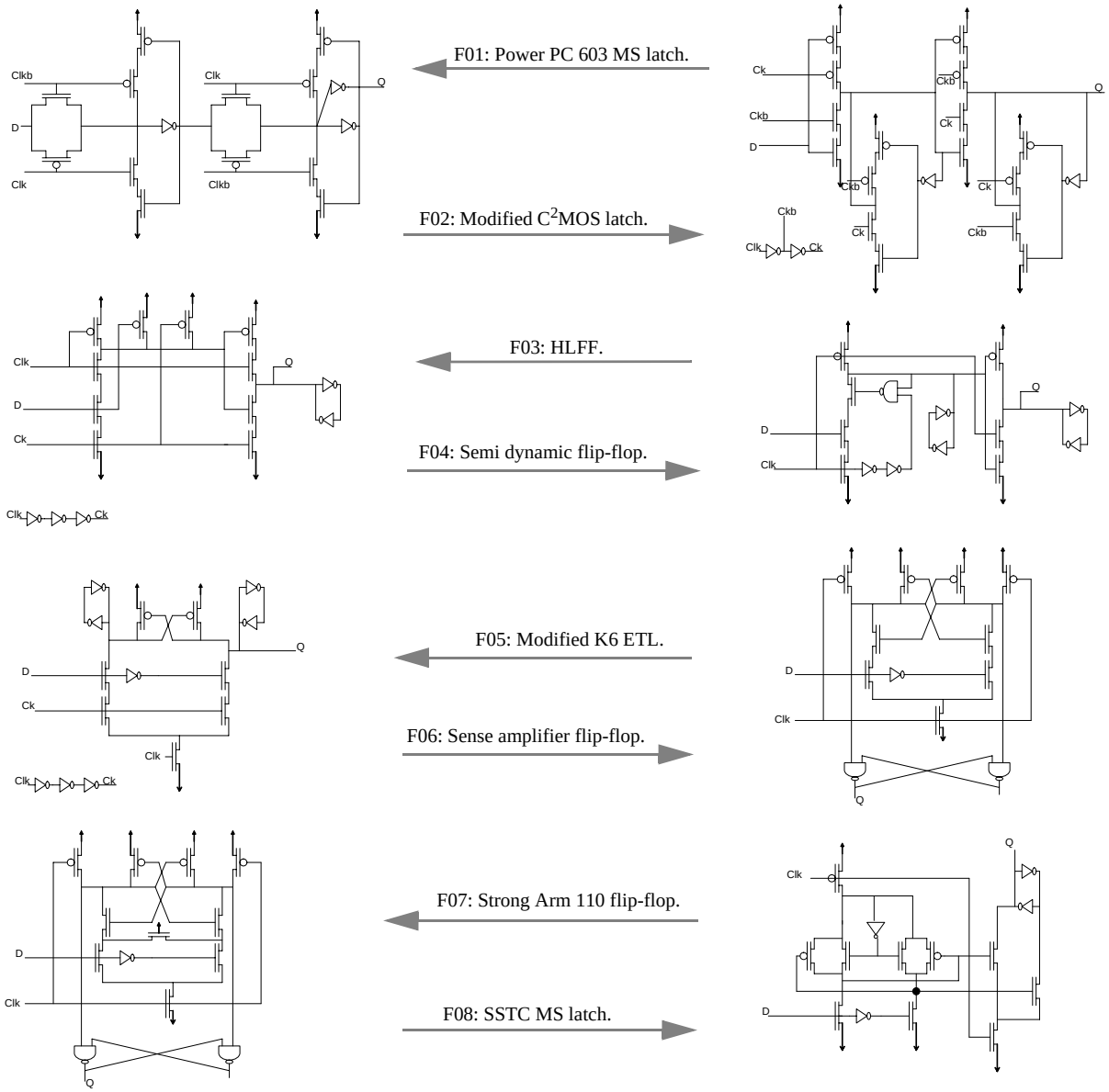


Figure 2 The first set of surveyed flip-flop circuits.

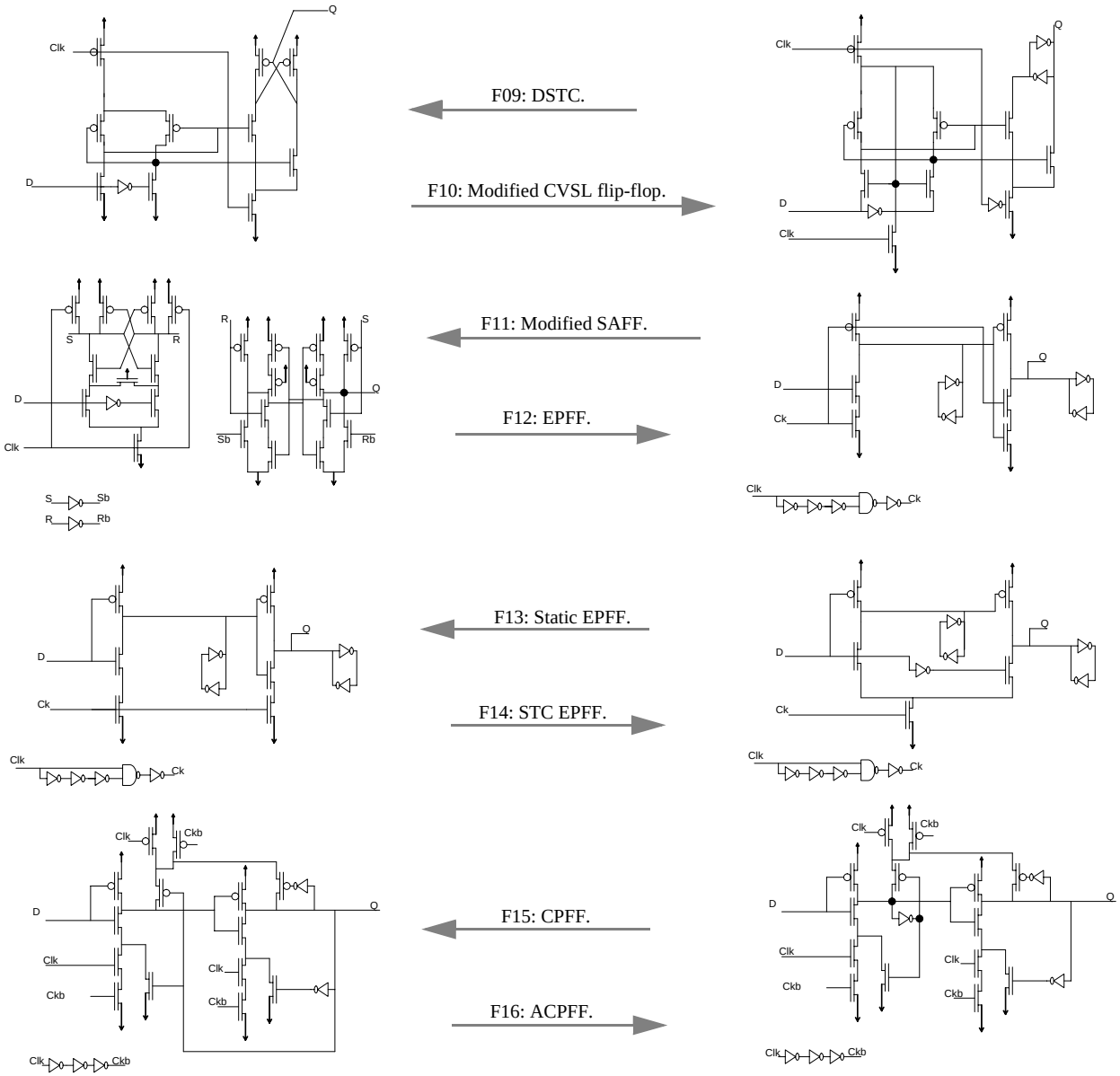


Figure 3 The second set of surveyed flip-flop circuits.

circuitry removed. It is very fast but its differential structure along with the precharge would cause a slight increase in power consumption.

F06 and F07 are two flip-flops that are very close to one another. The precharged sense-amplifier stage is very fast, but the set-reset latch almost doubles the delay due to unequal rise and fall times. This might cause glitches in succeeding logic stages and increasing the power consumption of these stages. F06 has better delay performance but suffers from floating output node of the sense amplifier stage if the data changes during the high phase of the clock, but still it has very low clock load which is an advantage in power consumption. F07 improves on the leakage power consumption.

F08 and F09 are again two single transistor clocked (STC) flip-flops that are very similar. They suffer from substantial voltage drop at the outputs due to the capacitive coupling effect between the common node of the slave latch and the floating output driving node of the master latch. This effect takes place at the rising edge of the clock and causes an increase in delay and short circuit power consumption in the slave latch which could dominate the dynamic power consumption. The capacitive coupling, floating node and data input signal glitches result in these flip-flops having lower driving capabilities than the rest of the flip-flop circuits used in this paper. This should be taken into account by adding the power consumption of the dummy loads into the power measurements/calculations.

F10 is a modified cascode voltage switch logic (CVSL) flip-flop. One of its advantages is using less transistors than other flip-flops. No floating nodes but still only one of the output nodes of the input stage can be fully pulled to a

weak “0” which might cause more short circuit power consumption.

F11 is the modified sense amplifier flip-flop (SAFF). It incorporates a precharge sense amplifier and a set and reset latch to hold the data. SAFF’s latency is a little higher than other flip-flops due to the delay of one output from the other in the output stage. This drawback is avoided in the modified design, where it supports fully symmetric output transitions.

F12 is the explicitly pulsed flip-flop (EPFF). It consists of a two stage dynamic structure, which has its effect on the power consumption. Noise immunity is another concern with any dynamic design style.

F13 transformed the first stage of the EPFF to a static stage, reducing its power consumption that is caused by precharging, switching and eliminating glitching. It also reduced the clock load. The pulse generator could be shared among several flip-flops which amortizes the cost in term of area and power consumption. Charge sharing is another concern in using any dynamic output stage, which might cause glitches in the succeeding circuits. However, a jam-latch (keeper structure) alleviates this concern.

F14 is the single transistor clocked EPFF. It uses two static latch stages sharing one clock transistor. Pulse width is a very important design parameter for circuits F12, F13, and F14, since it is sensitive to PVT variations and necessary for correct flip-flop functionality.

F15 is the conditionally precharged flip-flop (CPFF). Due to the notoriety of dynamic circuits for high power consumption, the CPFF adds conditional logic for the gate to precharge, otherwise the precharge step is skipped saving its power. It does come with a cost to it, which is higher setup time for the conditional logic to evaluate and give an output to the rest of the flip-flop. F15 has the disadvantage of the transparency of the first stage to glitches on the inputs when the output is high.

F16 is the alternative CPFF where the transparency to input glitches is avoided by using an inverter which prevents the propagation of any glitches during the transparency period.

4 Simulation Results

4.1 Simulation Model

All flip-flop circuits were sized for minimum size transistors of a 90nm technology initially, and sized up iteratively for correct functionality. Performance was not a sizing criterion and the idea behind this is that our goal is the lowest power possible, which implies reduction in loading effects. We did see failures at certain clock frequencies and that is the only performance sizing effort that was done, improving performance was not one of our goals in this paper. For a general design situation, the inputs were driven with minimum size buffers and the outputs were captured after a minimum size buffer stage as well. Figure 4 shows the basic model used for all simulation results presented in this paper.

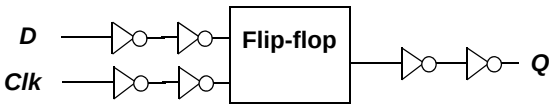


Figure 4 Simulation setup for flip-flops.

All the circuitry power consumption was included in the measurement of max power, due to the fact that this is the real maximum power that will be consumed if the circuit is used as part of a system. This is to account for the effects of the inputs, the driving capabilities and glitches –if any, on the flip-flop outputs.

All the numbers and results presented here are from simulations done at 25 degrees Celsius, with a 1.2 volts V_{dd} power supply and at the target process corner. We simulated all circuits at 10, 25, 50 and 100 MHz. This is done with relative schmooing of the data input relative to the clock with equidistant increments which leads to 6 steps for each schmoo at each frequency. This all results in $4 \times 6 = 24$ simulations for each flip-flop, gathering worst case delay (Clock-to-Q & D-to-Q) and power. In total there were $16 \times 24 = 384$ simulations to get the results and many more for debugging purposes and sizing iterations.

4.2 Timing and Performance

The charts shown in Figure 5 display the Clock-to-Q (Clk2Q) delay behavior at different Data-to-Clock delay values and different clock frequencies. Also, the charts shown in Figure 6 display the Data-to-Q (D2Q) delay behavior at different Data-to-Clock delay values and different clock frequencies. It is worth noting that Clk2Q results should match the D2Q results in the sense that if we see an increased delay in Clk2Q we should see a corresponding increased D2Q delay which caused it.

From Clk2Q and D2Q charts, we notice that at 100 MHz flip-flops F03, F10, F13 and F14 have bad delays at 2ns data input delay. We further notice that flip-flops F10, F13 and F14 have bad delays at 4ns as well. Both of these notes are attributed to the fact that these flip-flops have high setup time and could be seen from the D2Q charts.

At 50 MHz, we again notice that flip-flops F03, F10, F13 and F14 have bad delays at 4ns data input delay, F03 and F10 still have bad delays at 8ns data input delay and F03 still has bad delays at 12ns data input delay. These again are attributed to the fact that these flip-flops have high setup time and could be seen from the D2Q charts.

At 25 MHz, we notice that flip-flops F04 and F10 have bad delays at 8 and 16ns of data input delay and F04 continues to have bad delays at 24ns data input delay.

At 10 MHz, we notice that flip-flop F10 has bad delays at 20ns and 40ns data input delays. Another note is F02, which has high delay for 2ns but decreased delays for all other delay values.

The outlier performance of certain flip-flops at certain frequencies can be attributed to specific race conditions that may occur when a circuit has internal feedback paths that are racing with input or clock to output paths.

4.3 Power and Power-Delay-Product

We chose the peak power consumption to be measured because this is really the parameter to be concerned with during the design phase of a system. The clock and power delivery networks should be capable of withstanding the peak power consumption of the system without failing. Average power is a good metric for the goodness of the circuit and how much power would be used on average, but is dependent on activity and switching probabilities, which in turn are very dependent on the application.

The peak power measurement is quite problematic, the reason behind this statement is the difficulty of establishing

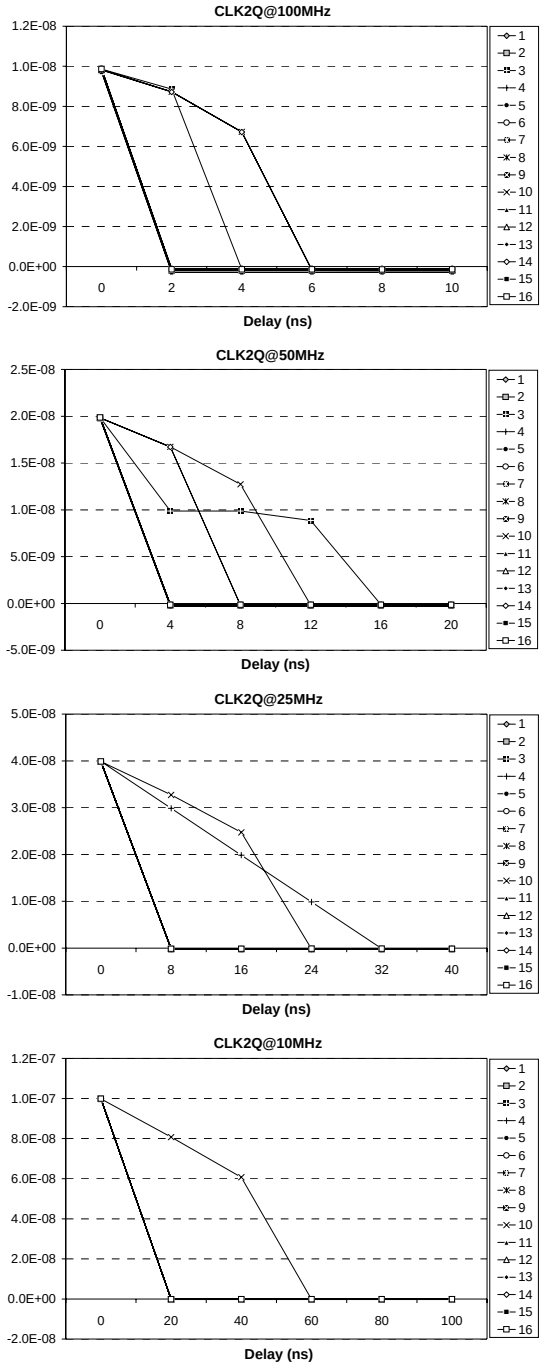


Figure 5 CLK2Q simulation results.

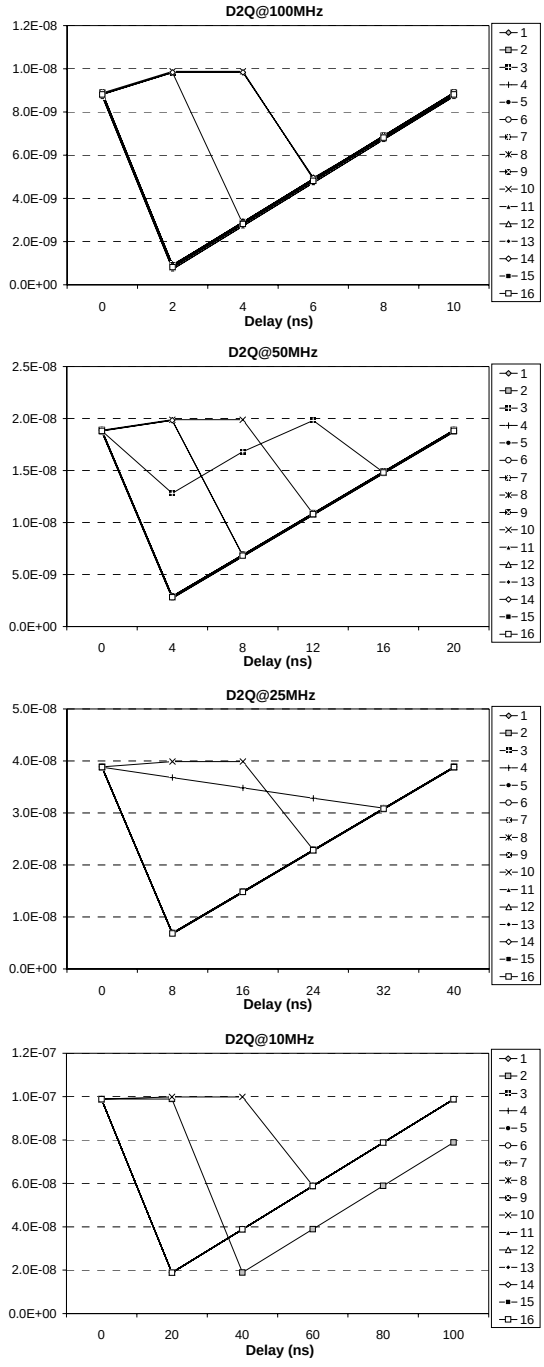


Figure 6 D2Q simulation results.

and qualifying the set of input transitions i.e. vectors and relative timings that cause the circuit to consume most power. This is a very tough issue to solve in generic designs or circuits, but not that bad for flip-flop circuits as the number of inputs is limited and the relative timings are direct forward, i.e. within the clock period of operation.

From our experiments, we noticed that the maximum power or peak power is not dependent on frequency and very slightly dependent on the delay of the data input (Data-to-Clock) which actually met our expectations. From our simulation results (A sample is shown in Figure 7), we notice that flip-flop F03 has a maximum power which peaks above and beyond the other designs, this is attributed to the structure of the circuit and makes great sense when looking back at its performance. We further noticed that some flip-flops are more sensitive to delay than others, this is due to the structure and internal organization of the flip-flops themselves. Finally, we have noticed that F05 (K6 ETL) without the reset circuitry is not that impressive power wise.

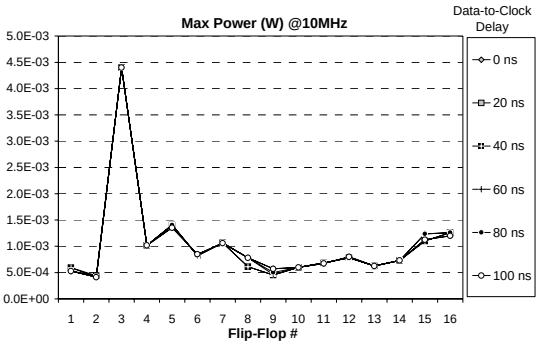
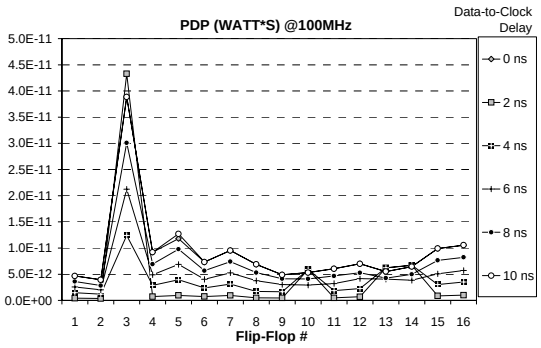
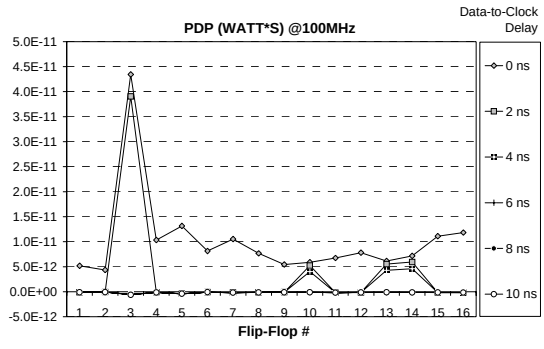


Figure 7 Maximum power simulations.



(a) Using D2Q delay.



(b) Using Clk2Q delay.

Figure 8 PDP at 100 MHz.

In addition to the above results, we constructed the power delay product charts for finding out the trade-offs between power consumption and delays for the different flip-flops. A sample of these charts is shown in Figure 8 where the PDP is graphed using (a) D2Q delay and (b) Clk2Q delay. We noticed from the graphs that if the flip-flop is not in the stable operating region, then its delay will dominate the PDP graphs as shown in Figure 8. But as we might further notice that PDP trends for stable regions of operations and across frequencies is a fairer comparison.

5 REMARKS AND CONCLUSIONS

On timing and performance, flip-flops F03, F10, F13 and F14 seem to have a higher optimal setup time than the rest of the flip-flops. F04 seems to have a particularly bad performance at 25 MHz which is inherent to the internals of the design itself and dependent on the technology used as well. F02 has a sweet spot at 10 MHz.

Table 1 Transistor count of flip-flop circuits.

F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12	F13	F14	F15	F16
16	24	20	23	23	19	20	18	12	17	28	26	26	27	24	26

On the power-delay front, we have noticed, as mentioned above, that F03 has high peak power consumption than the rest, even if in the stable performance region of operation. F05, F15 and F16 are next in line. This makes F05 the best of all for high performance systems where the trade-off between power and performance are very obvious. The other flip-flops are comparable regarding power consumption and performance. It is worth mentioning that F02 is best peak power consumption followed by F01 and F09.

If we would consider the number of transistors as a rough metric of area, given that minimizing the size of transistors was one of the main goals, then Table 1 shows the comparison of the different flip-flops in the area dimension. From this table, we notice that the best area is F09 and worst is F11, but since the transistors are quite small in area, this difference effect is diminished in larger designs where flip-flops and latches are a lower percentage of the gate count due to the large combinational logic blocks used to perform the main function needed.

In summary, we have presented a survey and evaluation of low-power flip-flop circuits. Our experimental results enabled us to identify the power and performance trade-offs of existing flip-flop designs.

REFERENCES

[1] V. Stojanovic and V. G. Oklabdzija, "Comparative analysis of master-slave latches and flip-flops for high-performance and low-power systems," *IEEE Journal of Solid State Circuits*, Vol. 34, pp. 536-548, April 1999.

[2] T. Yalcin and N. Ismailoglu, "Design of a fully-static differential low-power CMOS flip-flop", *Proc. International Symposium on Circuits and Systems*, 1999, pp. 331 - 333.

[3] A. Ghadiri and H. Mahmoodi-Meimand, "Comparative energy and delay of energy recovery and square wave clock flip-flops for high-performance and low-power applications", *Proc. International Conference on Microelectronics*, 2003, pp. 89 - 92.

[4] P. Zhao, T. Darwish, and M. Bayoumi, "Low power and high speed explicit-pulsed flip-flops", *Proc. Midwest Symposium on Circuits and Systems*, 2002, pp. 477-480.

[5] N. Nedovic, M. Aleksic, and V. G. Oklabdzija, "Conditional techniques for low power consumption flip-flops," *Proc. International Conference on Electronics, Circuits and Systems*, 2001, pp. 803-806.

[6] R. H. Katz, *Contemporary Logic Design*, Benjamin/ Cummings Publishing Company, Inc. 1994.

[7] E. J. McCluskey, *Logic Design Principles*, Prentice Hall, 1986.