

DRIP CURRENT: MOORE'S LAW MEETS STANDING POWER**K.Niranjan Reddy**

Assoc.Professor&HOD,Dept.of ECE, CMR Institute Technology, Hyderabad.

ABSTARCT:

With technological down scaling, static power has become one of the main issues in a system on chip. One intriguing approach to overcoming this difficulty is the use of normally off computing with non-volatile (NV) sequential components. In general, many distinct designs for NV shadow flip-flops have been presented, with magnetic tunnel junction (MTJ) cells serving as redundant data storage mechanisms. Compared to its CMOS counterparts, MTJs are more prone to manufacturing faults because of the evolving fabrication procedures of magnetic layers. In addition, memory arrays are easily repairable because to well-established memory repair and coding systems, but flip-flops dispersed throughout the layout are more challenging to repair. As a result, NV flip-flops need robust defect and fault tolerance to provide a high production yield. In this research, we present a design for a fault-tolerant NV latch (FTNV-L) that is resistant to several types of MTJ cell failures. All single MTJ faults may be tolerated by our suggested FTVN-L with far less overhead than with conventional methods, as shown by our simulation findings. This sram design with FF implementation project has been extended to us.

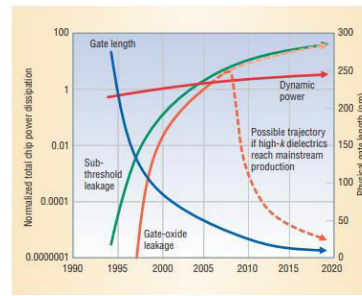
INTRODUCTION:

The semiconductor industry's primary technological challenge presently is power consumption. Intel chairman Andrew Grove raised the issue during the 2002 International Electron Devices Meeting, pointing specifically to off-state current leakage as a potential barrier to further microprocessor integration.¹ Static power, or the current that flows via off-state transistors, is known as off-state leakage. It is one of two primary causes of power dissipation in today's microprocessors. The other is dynamic power, which results from the constant charging and discharging of the capacitance at the chip's output due to the chip's hundreds of millions of gates. Only dynamic power has been a major contributor to power consumption until recently, and Moore's law has helped to rein it down. Below 100 nanometers, the technology used in processors has shrunk to the point that it requires a lower supply voltage. Because dynamic power is squared of the supply voltage, lowering the voltage drastically decreases power usage. Unfortunately, leakage is exacerbated with smaller geometries, and static power starts to dominate the power consumption equation when designing microprocessors.

MODELS OF PROCESS Historically, complementary metal-oxide semiconductor technology has dissipated far less power than older technologies such as transistortransistor and emitter-coupled

logic. In reality, CMOS transistors had almost little power loss while they weren't switching. With the rise in gadget speed and chip density, however, comes a commensurate rise in their power consumption. The academic world has long understood the importance of this rise. Using data from 2002 and normalizing it to the International Technology Roadmap for Semiconductors from 2001, Figure 1 displays the dynamics and statics of total chip power usage. According to the ITRS, dynamic power consumption will decrease over time. Total dynamic power per chip, however, will rise if we assume on-chip devices double in number every two years. Because of constraints like as high packaging and cooling costs and low battery capacity, this trend cannot be maintained indefinitely. Both subthreshold leakage (a weak inversion current across the device) and gate leakage (a tunnelling current through the gate oxide insulation) are expected to rise exponentially in the future, as shown in Figure 1. The ITRS anticipates a slowing in the pace of these gains in 2005, while they will still be significant. Although dynamic power dissipation is now greater, off-state subthreshold leakage is expected to grow to a greater proportion of overall power dissipation when feature sizes shrink below 65 nm. Emerging ways to reduce the gate-oxide tunnelling effect may be able to bring gate leakage under control by 2010. These methods focus on employing high-k dielectrics to better insulate the gate from the channel. The power equation that controls system performance, chip size, and cost must be rethought when leakage current becomes the dominant contributor to power consumption.

BASICS OF POWER The power-performance trade-offs of CMOS logic circuits are modeled by five equations. For the benefit of logic designers, architects, and system builders, we provide them here in simplified forms that convey the essentials. The first three are often seen in works dealing with little power.³ The latter two model subthreshold and gate-oxide leakage.



The two static power charts show ITRS forecasts for 2002 relative to 2001. Assuming a doubling of on-chip devices every two years, the dynamic power growth is substantial.

POWER-EFFICIENT STRUCTURAL CHOICES With the exception of the additional gates introduced by latches to separate the pipe stages, the leakage introduced by a pipelined implementation is equivalent to the basic serial case since both subthreshold and oxide leakage rely on total gate width or, roughly, gate count. In comparison to the serial example, the voltage requirements of pipelined implementations are lower, which may result in reduced dynamic and static power usage. Parallel implementations can also operate at a lower voltage, but only by

nearly doubling the amount of hardware. As a result, the parallel scenario may leak more power than the serial example, thus nullifying the benefits of the latter's reduced dynamic power consumption. Therefore, pipelining is the efficient use of energy. Since it uses half as much hardware and operates at a lower voltage, it will always consume less power than the parallel case and be more efficient than the serial case. In reality, the total dynamic and static power loss associated with pipelining will be lower than in the serial situation.

STATISTICAL POWER USAGE DROP

Since leakage is not activity dependent like dynamic power, minimising node switching during idle times would not help save energy. While turning off the system's inactive components helps, doing so causes data loss. CHANGES IN RETENTION Assuming the save/restore cost is negligible in comparison to the power saved by the snooze duration, a "snooze" mode may be useful when a device is dormant for an extended length of time. Researchers came up with "balloon" logic, also known as retention flip-flops, for shorter idle times. The goal is to make a copy of state-preserving latches and store it in a high- V_{th} lock. Equation 1 demonstrates how much lower the sub threshold leakage is in the high- V_{th} latches. While sleeping, the main latch is turned off and its state is copied to the retention latch. The threshold voltage of a transistor is determined by its layout and manufacturing method. In 180 nm procedures, typical values are 450 mV, whereas in 130 nm technologies, typical values are 350 mV. The substrate's threshold voltage may be raised by 100 mV using doping or bias voltage. This results in a 10-

fold reduction in leakage but a 15% increase in switching time. Since using low-leakage retention flops on the processor's critical route would slow it down, they are only effective for conserving state energy efficiently. Additionally, they may see substantial die area growth due to the addition of extra latches. It is typical for tiny embedded cores to take up just 10% of the entire chip space and 20% of the logic size.

STOP THE DRAIN ON MEMORY

Leakage is mostly attributable to the large amount of transistors used for the processor's on-chip caches. According to estimates, leakage will use almost 70% of the cache's power budget at 70 nm.⁶ Leakage current channels in a typical memory cell are shown in Figure 2. Bit line leakage is the current that flows via access transistor N3 from the bit line, whereas cell leakage is the current that flows through transistors P1 and N2. Sub threshold conduction, or current flowing from source to drain even when gate-source voltage is below the threshold voltage, is the root cause of both bitline and cell leakage. In addition, the transistor gates are experiencing leakage current due to gate oxide degradation.

CONNECTIONS AND SYSTEMS. To minimize leakage, designers of circuits might use either state-destructive or state-preserving methods. Ground gating, commonly known as gated- V_{dd} , is used in state-destructive methods. To implement ground gating, a sleep transistor based on NMOS (n-channel metal-oxide semiconductor) is added between the memory cell and the ground of the power supply.⁷⁻⁹ the greatest amount of leakage

power can be saved by turning off a cache line, but the system is vulnerable to wrong turn-off choices because of the loss of state. Because of the extra cache misses that must be satisfied by off-chip memory, such choices may result in substantial power and performance overhead. Different state-preserving methods exist. The supply voltages for drowsy caches are multiplexed according on the status of each cache line and cache block. Caches in sleep mode employ a low retention voltage level, storing data in a cache area that must be accessed with a high voltage.¹⁰ There is a one-cycle penalty for reawakening the sleeping cache lines since this action is seen as a faux cache miss. Techniques that maintain state have included lowering threshold voltages gradually¹¹ and employing preferred data values.¹² Accessing the low-leakage state incurs substantially less penalty, although all of these methods decrease leakage less than totally turning off a cache line. Furthermore, although destructive approaches may decrease leakage by more than a factor of 1,000 and state-preserving techniques can only lower it by around a factor of 10, the net difference in power consumption is less than 10 percent. State-preserving methods often have superior performance when the shorter wake-up time is included in the total programme duration. They also benefit from not needing a level-2 cache.

Methods of REGULATION. For leakage-reducing capabilities, there are two main types of control methods:

- Application-insensitive controls, which regularly disable cache lines,^{9,13} and
- application-sensitive controls, which are dependent on runtime performance input

from the programme itself.^{6,7,10} Control is quite different amongst the different categories.

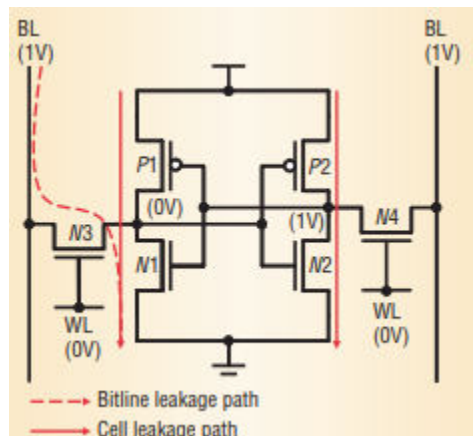
While an application-aware technique⁹ would reveal that the CPU can disable 25% of the cache due to an extremely high hit rate, this information would tell us nothing about which 75% of the cache lines will be utilised in the near future. The cache-decay algorithm⁷, on the other hand, maintains track of statistics for each cache line and may thus give superior predictive behaviour due to its insensitivity. Algorithms that are "application-agnostic" aren't tuned to perform optimally in any given circumstance. Instead, they try to be generally reliable and avoid or at least mitigate the negative consequences of being wrong. For instance, periodically disabling a cache line is one such method.¹⁰ To be effective, the chosen period must accurately represent the pace of change in the instruction or data working set. In particular, the ideal time frame may vary not just across applications but also between stages of the same application. This causes the algorithm to squander energy by keeping the cache lines alive for longer than required or shutting down the lines that contain the active set of instructions. If we try to fix the first issue by shortening the time, we'll just make the second one worse. This method is advantageous for data caches since it is easy to implement and causes little overhead. Pathologically awful behaviour on certain code types may also be shown via application-agnostic approaches. One method⁶ involves reactivating dormant cache subbanks at the appropriate times during execution. The method may produce frequent drowsy-awake transitions, which

degrade the performance of the loop, if the initial half of the loop is in one bank and the second part is in another. Furthermore, substantial energy might be lost due to leakage current wasted by lines that are awake but not being accessed. Reducing node switching even when there is no work does not assist decrease energy consumption due to leakage since it is not hobby mainly focused like dynamic power. While turning off the device's inactive components helps, doing so reduces its effectiveness. A "snooze" mode may be useful when a device is going to be dormant for an extended period of time, but only if the cost of maintenance and repair is negligible compared to the energy savings achieved. Scientists have developed "balloon" not unique sensation, also known as retention flip flops, for use during relatively brief periods of inactivity. The plan is to use excessive locks to replicate the latches of the nation-keepers. Equation four demonstrates that the sub threshold leakage is drastically reduced due to the excessive latches. In nap mode, the manage common feel switches off the primary latch and duplicates its state to the secondary latch in order to save power. Leakage is caused in large part by the on-chip caches that make up the bulk of the processor's transistor cost. According to estimates, leakage will be responsible for as much as 70% of the cache power fee range in the 70-nm generation. Leakage contemporary routes in a commonly used memory cell are shown in Figure 2. Bitline leakage refers to current leaking from the bitline via access transistor N3, whereas cellular leakage refers to current leaking from cells through transistors P1 and N2. Sub threshold conduction, or current flowing

from the deliver to drain even when the gate-deliver voltage is below the threshold voltage, is the root cause of both bitline and cell leakage. In addition, there is a current leakage of gate-oxide via the transistor gates.

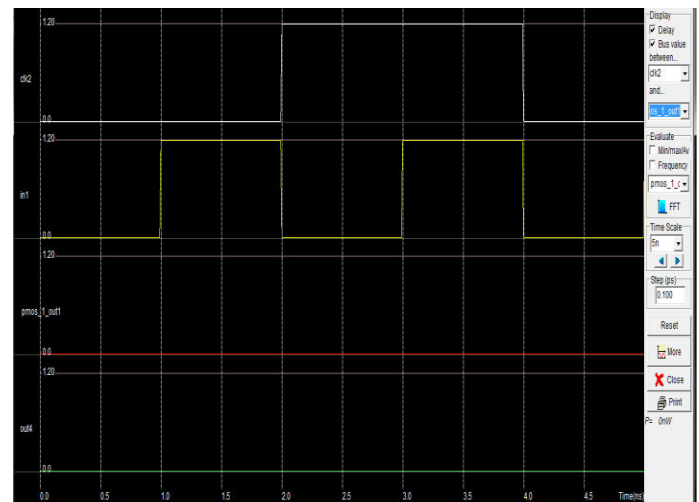
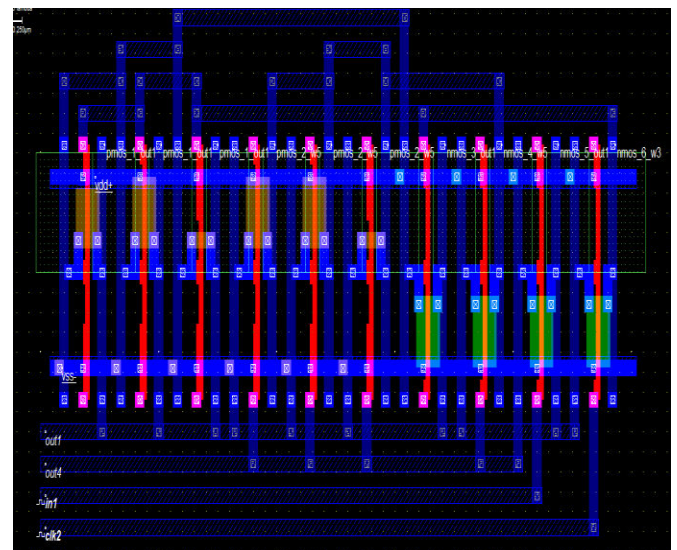
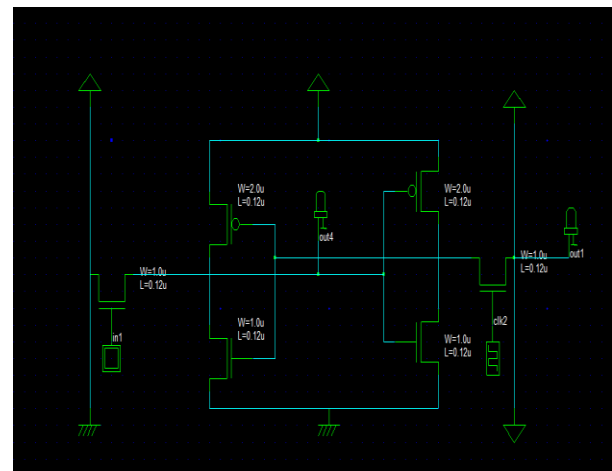
Methods for holding onto a state include gradually lowering threshold voltages¹¹ and selecting preferred data values.¹² Without a question, shutting off a cache line is the most effective way to reduce leakage, but all the other methods incur far less of a cost than acquiring access to the low-leakage U.S. The net difference in strength fed on via the two strategies is much less than 10 percentage, even though nation-keeping techniques can quality lessen leakage with the useful resource of use of approximately a component of 10 in contrast to more than an aspect of 1,000 for damaging strategies. State-maintaining techniques often perform better in current application runtime due to the shorter wake-up time. Within each company, the level of control varies considerably. If the processor has a very high hit rate, the software-sensitive technique⁹ may propose that 25% of the cache be disabled, but it will not indicate which 75% of the cache capacity will be utilised in the near future. However, an approach that is not as sensitive as the cache-decay algorithm⁷ could be able to provide more accurate predictions since it keeps track of statistics for each cache line. Application-agnostic algorithms no longer tailor their performance to each individual task. Instead, they argue for socially acceptable behavior and mitigate the negative impact of wrong predictions. One such method is the periodic disabling of a cache line.¹⁰ How well the selected time

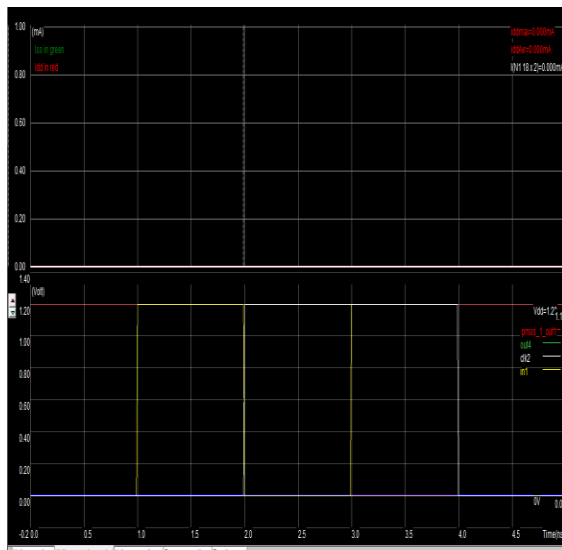
frame depicts the rate at which the guiding or statistical working set changes will largely determine its success. In particular, the most impressive threshold modern-day makes use of not one but two threshold voltage period. Threshold voltages are a common feature of modern approaches. The designers set the threshold voltage high for the majority of the transistors that are not time-critical but have a lower priority than the ones that are essential to basic performance. For current performance-critical transistors, this method results in considerable sub threshold leakage, but it has the potential to significantly lower total leakage.



Memory cell leakage current routes are shown in Figure 2. Current leaks from the bitline into access transistor N3, whereas current leaks from the cells into P1 and N2.

SIMULATION RESULTS:





CONCLUSION

Nowadays, spintronic-based entirely shadow latches are getting attention as the ones are notably advantageous for leakage bargain. This is because the latches' storage devices, most likely MTJ cells, offer desirable properties like as zero leakage and fast access. This makes the latches ideal for the kind of instant-on/generally-off processing typical of a SoC. Short oxide, open vias, and the inability to transmit magnetic orientation of the FL are just a few of the manufacturing problems that are particularly common in MTJ cells. Since a single faulty flip-flop may bring down the whole redundancy system, this has a negative impact on layout yield. In order to tolerate MTJ-related failures, we developed an FTNV-L in which MTJs are serially and parallelly coupled in a novel fashion. We have proved the capacity of our recommended layout in the presence of all MTJ faults beneath the have an effect on of device model and going for walks temperature. Furthermore, any single defect consistent with latch may be tolerated using

the FTNV-L configuration at significantly reduced costs in comparison to the typical method based on 3MR.

REFERENCES

1. R. Wilson and D. Lammers, "Grove Calls Leakage Chip Designers' Top Problem," EE Times, 13 Dec. 2002; www.eetimes.com/story/OEG20021213S0040.
2. Semiconductor Industry Assoc., International Technology Roadmap for Semiconductors, 2002 Update; <http://public.itrs.net>.
3. T. Mudge, "Power: A First-Class Architectural Design Constraint," Computer, Apr. 2001, pp. 52-58.
4. K. Nowka et al., "A 0.9V to 1.95V Dynamic Voltage-Scalable and Frequency-Scalable 32b PowerPC Processor," Proc. Int'l Solid-State Circuits Conf. (ISSCC), IEEE Press, 2002, pp. 340-341.
5. A. Chandrakasan, W. Bowhill, and F. Fox, Design of High-Performance Microprocessor Circuits, IEEE Press, 2001.
6. N. Kim et al., "Drowsy Instruction Caches: Leakage Power Reduction Using Dynamic Voltage Scaling and Cache Sub-bank Prediction," Proc. 35th Ann. Int'l Symp. Microarchitecture (MICRO-35), IEEE CS Press, 2002, pp. 219-230.
7. S. Kaxiras, Z. Hu, and M. Martonosi, "Cache Decay: Exploiting Generational Behavior to Reduce Cache Leakage Power," Proc. 28th Int'l Symp. Computer Architecture (ISCA 28), IEEE CS Press, 2001, pp. 240-251.

8. M. Powell et al., "Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep-Submicron Cache Memories," Proc. Int'l Symp. Low-Power Electronics and Design (ISLPED 00), ACM Press, 2000, pp. 90-95.
9. M. Powell et al., "Reducing Leakage in a High-Performance Deep-Submicron Instruction Cache," IEEE Trans. VLSI, Feb. 2001, pp. 77-89.
10. K. Flautner et al., "Drowsy Caches: Simple Techniques for Reducing Leakage Power," Proc. 29th Ann. Int'l Symp. Computer Architecture (ISCA 29), IEEE CS Press, 2002, pp. 148-157.
11. H. Kim and K. Roy, "Dynamic Vth SRAMs for Low Leakage," Proc. Int'l Symp. Low-Power Electronics and Design (ISLPED 02), ACM Press, 2002, pp. 251-254.
12. N. Azizi, A. Moshovos, and F.N. Najm, "Low-Leakage Asymmetric-Cell SRAM," Proc. Int'l Symp. Low-Power Electronics and Design (ISLPED 02), ACM Press, 2002, pp. 48-51.
13. W. Zhang et al., "Compiler-Directed Instruction Cache Leakage Optimization," Proc. 35th Ann. Int'l Symp. Microarchitecture (MICRO-35), IEEE CS Press, 2002, pp. 208-218.