Theory Article

ADVANCED TECHNIQUES FOR DYNAMIC POWER REDUCTION IN SRAM CIRCUITS: A PHYSICS-BASED APPROACH

Zh. Saipidnim[*]

Department of Physics, Ala-Too International University, Bishkek, Kyrgyzstan

*Corresponding Author E-mail: saipidinovjusup@gmail.com

ABSTRACT

Dynamic power reduction in Static Random Access Memory (SRAM) is essential for improving energy efficiency in modern digital systems, especially in processors and embedded devices. As SRAM power consumption contributes significantly to overall chip power, optimizing power in SRAM is critical. This thesis explores recent techniques for dynamic power reduction in SRAM, with a focus on architectural modifications, dynamic voltage scaling, adaptive voltage supply, and optimized circuit designs. This abstract outlines the need for low-power SRAM in reducing heat, noise, and electro-migration, particularly in high-performance computing. The research highlights how Multi-Tree Voltage Regulator (MTR) techniques and other dynamic power management methods contribute to significant reductions in power and heat generation, thereby enhancing both performance and reliability.

Keywords*: SRAM (Static Random Access Memory); Dynamic power reduction; Power consumption; Voltage scaling; Adaptive voltage supply; Multi-Tree Voltage Regulator (MTR); Low-power design; Power management techniques; Heat dissipation; Noise reduction; Electro-migration

INTRODUCTION

The most important memory in a computer is the Random Access Memory (RAM) that stores data or information temporarily and works in conjunction with the Central Processing Unit (CPU). RAM is divided into volatile and non-volatile memories, and volatile memory loses all data when the system is powered down. Non-volatile memory maintains its state when the system is powered down, with the classic example being the hard drive. The other memory types are called volatile memories because they lose their states as soon as the power supply is turned off. SRAM is a very popular semiconductor memory that can maintain its state without distortion even after disconnection from the power source. In addition, the response time of the SRAM is much faster than the response times of other semiconductor memories.

The turnaround rate from data input to output can reach five to ten times, according to the architecture and structure of the SRAM. However, the disadvantages associated with SRAM are hard to overcome due to its numerous and large transistors that compose a single 6-transistor unit. The first is that it consumes excessive power, which naturally contributes to a reduction in data density. The other is that the area occupied on the semiconductor chip is very large. As the demand for low power consumption in digital systems is increasing gradually, new techniques must be researched and implemented in order to allow for the control of the power caused by running the source for hundreds of thousands of SRAM cells. This text presents information about the most recent techniques for dynamic power reduction, effectively detailing three well-known techniques used for lowering consumption without a significant performance impact in SRAM circuits.

OVERVIEW OF SRAM TECHNOLOGY

SRAM technology, traditionally used for the storage of data in processor data caches, has evolved as the predominant high-speed random access memory in use today. External SRAM, used with processors since the earliest on-chip cache memories in microcontrollers, has been used for stand-alone processor data caches in many high-performance processor designs because it is typically cheaper, faster, and has lower access latencies than embedded DRAM used for processor instruction caches. The desire to have larger and faster memories with high bandwidth has driven the embedded vendor's development of larger SRAM-based memories with high bandwidths. Meanwhile, reducing access latencies has driven Dynamic Random Access Memory (DRAM) vendors to implement embedded type memories as last level caches. But the embedded DRAM cell efficiency advantages allow the embedded DRAM to be scaled smaller than SRAM and implemented at higher process nodes in gigabyte densities not feasible for SRAM, driving the fastest SRAM back into embedded designs. In summary, SRAM cell development has evolved to the minimum feature set for different SRAM developed for the fast external markets, while SRAM used in processors and other embedded designs is stuck in larger features to achieve deeper process module parity with DRAM. Failing device feature parity, the unity SRAM core modules just fall behind as a result in power and process modules. Successful SRAM module design will therefore have solutions for deep submicron process modules while achieving requirements for the external varieties. This thesis will begin with an overview of SRAM practices and power characteristics and outline their main shortcomings. Management of dynamic power will be discussed

and static power will be discussed in section.

Importance of dynamic power reduction

The three effects that emerge from the increase in power consumption are heat, noise, and electro-migration. The power consumption defines the overall implementation cost in microelectronic design, such as die area, performance, and reliability. Moreover, when the power consumption exceeds the power supply, the chip design stops working properly. However, in the case of SRAM, the power consumption dependence problem due to technology scaling is worsened. Approximately 60%-70% of the power ratio in a chip is related to the silicon area occupied by the cache memory. This silicon-based area occupied by memory is split into:

- The logic-based cell memory area.

- The Dynamic power loss represented by the Voltage

multiplied by the supply voltage and the bus capacitance.

- The static losses represented by Ioff, which is proportional to the cache leakage and voltage dependent. With the aim of reducing heat, in recent years, there has been an intensification of research aimed at designing a logic circuit or computer-based system that would be functional for some typical scenarios with uncertain boundary conditions within a safe operational area.

In particular, the power consumed by the CPU is reduced by new commercial high-density low-leakage SRAM products that take advantage of the potential energy gains due to reduced power consumption, improving the logic gate circuits that are close to the CPU within the power delivery network. These voltages are supplied by the BIOS, which presents adaptive voltage scaling equipped with a power supply, maintaining prescribed electrical performances, such as the key hard-task execution time delay in computation-intensive applications, including the calibration of models. In the case of overheating, the advancement of the critical hot spot node in a three-dimensional stacked wafer becomes an additional problem not only in the full integration of the SoCs in the desired volume and complexity but also in the power supply. From the analysis of the data obtained, the team of researchers has identified an interesting correlation between power density and hot-spot pixel perimeter. They notice a linear equation if they correlate it to the empirical quadratic equation for the instantaneous current drawn by the power density, depending on typical circuits: an on-chip SRAM comparator for both correlations with the highest goodness of fit.

According to their findings, dedicated design for variable and dynamic power supply at the node, called Multi-Tree VR, reduces simultaneously the power and heat generated by SRAM. In particular, the MTR power supply design will reduce the SLR energy dissipation, resulting in a 25% reduction in power and a 9 degrees Celsius reduction in temperature. In the case of dynamic operation voltage, the golden technology node is intriguing for the commercial high-density low-leakage products that are used today in consumer electronics with power-saving features when in optimal working conditions. These conditions, since they are typical of what we call 'power-on condition', are already requested by an electronic device for operation; in particular, the device must guarantee that a stream of bytes at an operational speed is generated.

Dynamic power reduction techniques in SRAM

Basic SRAM architecture

The extremely basic SRAM cell consists of two cross-coupled CMOS inverters, with two access transistors for each inverter. The two-bit cell has a read/write and a complementary read/write path. Dynamic power is dissipated during read and writes operations. The read and write paths both rely on the bitlines. Normally, one of them is enabled by the wordline of the access transistors for the row cell. During a read operation, both bitlines are pre-charged. After this pre-charge is completed, a column decoder selects the wordline of the access transistors of the column of SRAM cells that a read operation is being performed on, asserting that wordline. The corresponding two transistors get activated. Then the cell performs the retrieval of the content and transfers it to the bitline in a destructive manner [1].

The state of the cell is sensed by both bitlines, and a sense amplifier at the end of the bitlines latches the data and drives the pair, thus pre-charging the bitlines to their logical value. A write operation is similar but has the inverse behavior. Many cells form an array with a control word line. For certain timing relationships, both bitlines get connected to the data bus [2]. If the data bus is slow, both source line and drain line or bitlines can get shorted. To avoid the races, stochastic pre-charging is used. If the write and read wordline enable times overlap, erroneous data would be read [3].

Physics behind power consumption

Dynamic power dissipation is fundamentally governed by electromagnetic theory and the behavior of capacitive circuits. When an SRAM cell switches state, it charges and discharges the parasitic capacitances associated with transistors and interconnects. The energy required.

$$E = \frac{1}{2} C_{load} V^2{}_{DD}$$

for this process is given by:

Over a cycle at frequency if, the dynamic power is:

$$P_{dynamic} = f.E = \frac{1}{2} C_{load} V^2{}_{DD}.f$$

The resistive heating caused by current flow in transistors leads to power loss, which is a key concern in device miniaturization. Furthermore, heat generation impacts device reliability through thermal noise and electromigration, which degrade material integrity over time.

Dynamic power consumption in SRAM

Under dynamic dissipation occurs mainly in two components: the bit lines and the evaluate circuit. Dynamic power is used to charge and discharge the bit lines during the word line perturbation time. Dynamic power is also used in the read compare path to sense the signal level at the bit lines and force the bit lines to their recharge level [4]. Although most of the SRAM uses state space flip-flops to store the data, the dissipation in them is usually just a fraction of the dynamic power of the SRAM [5]. Hence, bit line power is also known as dynamic dissipation as it occurs only when fetching the data from or writing the data to SRAM. The charging and discharging of the bit lines is the most important factor, contributing up to 60%-70% of the total SRAM power consumption. Such high power consumption has drawn everyone's attention to find out various techniques which could reduce the dynamic power [6]. One of the methods is to reduce the bit line coupling capacitance. The other one is to reduce the amplitude of the bit lines before word line perturbation and detect the signal with enhanced sense amplifiers which use low power decoder [5].

Sample calculation: Dynamic power in an SRAM cell

Consider an SRAM chip operating at a supply voltage $V_{DD} = 1.2V$ with load capacitance C_{load}=10fF, switching activity factor α =0.5, and frequency f =1GHz. The dynamic power is calculated as:

$$P_{dynamic} = \alpha . C_{load} . V^2{}_{DD}.f$$

$$P_{dynamic} = 0.5 \cdot 10 \cdot 10^{-15} \cdot (1.2)^2 \cdot 10^9$$

$$P_{dynamic} = 7.2 \mu W$$

Dynamic power dissipation: Detailed calculation

Dynamic power dissipation arises due to the repeated charging and discharging of parasitic μcapacitances during transitions [6]. The energy required per transition is:

$$E = \frac{1}{2} C_{load} V^2 DD$$

If we consider an SRAM array with N=1×10⁶ cells, and each cells has a parasitic capacitance of C_{load}=10fF, operating at $V_{DD} = 1.2V$, and an activity factor α =0.2, we can calculated the total dynamics power [7].

Energy per transition for one cell:

$$E_{cell} = \frac{1}{2} \cdot 10 \times 10^{-15} \cdot (1.2)^2 = 7.2\text{fJ}$$

Energy for all cells per cycles:

$$E_{total} = N \cdot \alpha \cdot E_{cell} = 1 \times 10^6 \cdot 0.2 \cdot 7.2 \times 10^{-15} = 1.44\mu\text{J}$$

Power at f =1 GH$_Z$:

$$P_{dynamic} = E_{total} \cdot f = 1.44 \times 10^{-6} \cdot 10^9 = 1.44\text{W}$$

CONCLUSION AND FUTURE DIRECTIONS

In this paper, we presented the related work and potential design techniques1 in the direction of dynamic power reduction in standby mode in SRAM arrays. Dynamic power reduction in standby mode is especially interesting in conjunction with emerging approaches targeting redundancy in caches. These seem to be a natural fit for having a cache that consumes low static current while providing acceptable standby and evaluation energy. The approaches could target various sub-arrays within the cache, such as those for the tags or the data. Alternatively, grouping some of the various cache chips together and providing redundancy at the chip level, inclusive of cache and interface to that chip, could also be manageable in existing process technologies. We hypothesize that tag array access savings would be substantial, and a solution designed to target just the tag array would be energy-efficient. Dynamic power reduction techniques for data arrays and circuits of a cell in SRAM affect performance and leakage savings directly.

REFERENCES

1. Prakash, G., Dutta, U., & Khan, M. T. "Dynamic Power Reduction Techniques in SRAM: An Overview." J. Low-Power Electron. Appl.
2. Khan A, Mekuria RR, Isaev R. Applying Machine Learning Analysis for Software Quality Test. IEEE. 1-15(2023)
3. Rabaey JM, Chandrakasan A, Nikolic B. Digital integrated circuits. Englewood Cliffs: Prentice hall; (2002)
4. Calhoun BH, Chandrakasan AP. Ultra-dynamic voltage scaling (UDVS) using sub-threshold operation and local voltage dithering. IEEE J. Solid-State Circuits. 41(1):238-245(2005)
5. Mai KW, Mori T, Amrutur BS, et.al. Low-power SRAM design using half-swing pulse-mode techniques. IEEE J. Solid-State Circuits. 33(11):1659-1671(1998).
6. Sylvester D, Keutzer K. Getting to the bottom of deep submicron. 203-211(1998
7. Roy K, Prasad SC. Low-power CMOS VLSI circuit design. John Wiley Sons; (2009).

How to cite this article:

Saipidinov *et al.* Advanced techniques for dynamic power reduction in sram circuits: a physics-based approach J Pharm Sci Innov. 2024;13(4):1-3 .