DEVELOPING AND ASSESSING AN ENERGY-EFFICIENT SRAM UTILIZING CMOS TECHNOLOGY

[#1]M.Anusha, Assistant Professor,

[#2]V.B.GopalaKrishna, Assistant Professor,

[#3]B.Narender, Associate Professor,

Department of Electrical and Communication Engineering,

SAI SPURTHI INSTITUTE OF TECHNOLOGY, SATHUPALLY, KHAMMAM.

ABSTRACT: Application-specific integrated circuits and commercial CPUs need less power when the size of technological nodes lowers. Implementing power-saving measures is universally acknowledged as the most successful way for VLSI systems, both now and in the future. These devices have large on-chip SRAM memories. In current systems, it is critical to reduce power consumption due by memory leakage while maintaining data integrity. Unfortunately, advanced approaches such as power-gating are limited to logic applications due to their ability to wipe the data stored in an SRAM device. Previous research has found significant time and space concentrations of data patterns in commercial processors and application-specific integrated circuits (ICs) that process audio, video, and image data. This work describes a new columnar Energy Compression technique for deactivating cells depending on a data pattern, hence preserving power in SRAM. This technique applies to commercial processors and is used to investigate energy conservation in application-specific integrated circuit SRAM memories. The study also assesses the impact of pre-storage photo processing and data cluster architectures on power reductions in order to optimise power consumption.

Index Terms: Low Power SRAM, SRAM, CMOS, PowerSaving.

1.INTRODUCTION

SRAM (Static Random Access Memory) is required for processing picture frame buffers. It is required to save several frames from videos. SRAM provides a high-speed mechanism for comparing master video frames to subsequent frames while minimizing delay. SRAM has a lower access time and can handle data requests more effectively than DRAM. Dynamic Random Access Memory (DRAM). Because of their on-chip design, they can be created using the same manufacturing procedure as the core and other on-chip circuits. Unlike DRAM, SRAM does not need to be refreshed on a regular basis. Because of the considerable temporal locality between consecutive video frames, the difference between the primary and subsequent frames is typically saved on a disk. A SRAM system that primarily fetches this frame stores the master frame.SRAM

memory systems in application-specific integrated circuits (ASICs) used to process photographs and videos waste up to 81% of the power in standby/leakage mode.

The chip's power dissipation is determined by the amount of power consumed per unit area. The system's total power can be separated into two categories: idle power and active power. During periods of heavy CPU utilization, active electricity is consumed. This electricity is used during data processing. There are periods of idleness, as evidenced by the usage patterns of commercial electronic gadgets, implying that a device does not have to be always busy. Transistors continually drain leakage power, also known as idle power. At sub-nanometer nodes, transistors' threshold voltage (Vth) lowers as leakage power increases. Even when the device is turned off, a lower threshold voltage (Vth) causes a rise in leakage

current in transistors due to the pressure differential between the Drain and Source voltages. As a result, idle power consumption rises, increasing the system's overall power consumption even while the CPU or application-specific integrated circuit is inactive.

SRAM accounts for a considerable amount of the on-chip area. This value has grown in tandem with the number of video processing engines. As the number of processor engines increases, so does SRAM capacity in order to enable greater throughput, as each processing engine can operate on its own independent section of memory. The amount of SRAM has a direct impact on system performance, as SRAM systems take up a large portion of the chip's surface area in most processors. The leakage power is governed by the number of SRAM cells. As the number of SRAM systems in commercial appliances grows, the importance of leakage power in total power consumption becomes clear.

Commercial processors and ASICs use techniques like power gating and frequency scaling to reduce power consumption when cores, processing engines, and other on-chip logic circuits are not in use. Power gating uses power gating transistors to increase impedance at the supply nodes (VDD/GND). This reduces the power supply to the logic circuitry, causing them to enter a tristate state. Information is stored in memory components, and the data's security cannot be compromised. To ensure data correctness and consistency, many big memory components, such as L3 (Level Three) caches, are outfitted with additional error-correcting circuits. Loss of data dependability in memory can lead to improper program execution, system crashes, or faulty picture and video frame processing by processor engines. This is not an issue that should be acknowledged or tolerated. As a result, we are unable to use power gating techniques on SRAM systems because tri-stating these elements causes data loss.

According to studies, processors and engines built

for image, sound, or video processing work with data that is biased toward a specific value, such as zero or one.

As a result, SRAM memories include more data cells than those with distorted values. As a result, earlier research has suggested the construction of memory systems with predisposed properties for storing predetermined values. One method for addressing this issue is to reduce the supply voltage delivered to the SRAM cells. The phenomenon of lower idle power happens when the leakage current reduces exponentially as the supply voltage drops. The data retention voltage (DRV) of an SRAM cell is the lowest voltage required to produce a flip in the cell. This voltage limits the amount that the supply voltage can be reduced. Process parameters create considerable fluctuations in the chip's dynamic random access voltage (DRV)[9]. This thesis describes a novel implementation of a 7T SRAM cell and compares the performance of typical 6T SRAM designs in terms of image storage, particularly during extended periods of inactivity. The technique, focusing on sub-22nm nodes, aims to reduce system power consumption by 15% through the use of an alternate design known as the "Energy Compressed SRAM system." These proposals are aimed at large-scale, low-activity memory devices that demand huge amounts of electricity when idle. The thesis will largely focus on Application-Specific Integrated Circuits (ASICs) that are specifically developed for image processing. This study offers a revolutionary SRAM technique designed to reduce power loss due to leakage in image and video processing applications.

2.BACKGROUND AND MOTIVATION

SRAM cells use a cross-coupled inverter arrangement to store data. Figure 1 shows the cross-coupled inverter structure and its corresponding 6T (6 transistor) circuit. Because of its simple architecture, this structure is widely

seen in ASIC memory and CPU cache.

Memory System using SRAM: Organization

SRAM memory systems are made up of numerous banks, each including several subarrays. The subarrays are grouped into a grid-like layout with rows and columns. The SRAM cells are arranged in vertical columns and connected by two bitlines. The sense amplifier receives multiplexed and grouped columns. Rows are made up of SRAM cells that all share the word "Line". Figure 2 shows the entire system powered by SRAM. To read data, the two bit lines must first be precharged. To access an SRAM cell, decode the address up to the subarray level, assert the correct word line, and select one column from the group. The column address is used to decide which pair of column bit lines should be connected to the sensing amplifier. Upon cell activation, one of these bitlines will begin to discharge, based on the stored data value. The sense amplifier's purpose is to magnify voltage differences in order to detect data. The sensing amplifier's output, which might have voltage levels of VDD (representing logic 1) or Zero (representing logic 0), displays the stored bit value.

Leakage Power in SRAM

Leakage power is the primary cause of power consumption in a big memory system with SRAM at nanoscale nodes. Memory found in ASICs used in cameras and commercial CPUs, particularly in last level caches, has the qualities of long-term data storage and low activity.

Figure 1: A 6T SRAM Cell
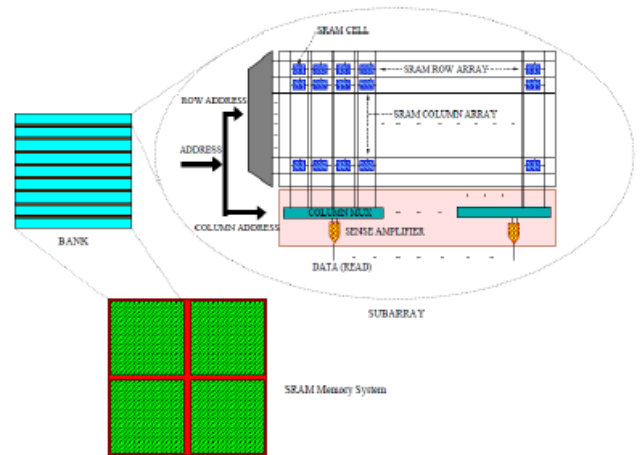
Figure 2: SRAM Memory System

Leakage currents cause the dissipation of idle electricity, increasing their power consumption. Nanoscale nodes show subthreshold, gate, and reversed-biased junction leakage currents. A transistor's threshold voltage (Vth) drops as the technological node decreases. This effect has a large impact on total leakage while amplifying sub-threshold leakage in nanoscale nodes. Subthreshold leakage in SRAM is made up of two components: bit line leakage and cell leakage. Figure 1 depicts the orientation of the bit line and cell leakage. Bit line leakage can happen in any direction, depending on the data in the cell. Cell leakage flows from the power supply voltage (VDD) to ground (GND). The difference in supply voltage across these SRAM cells is what causes the leakage. Bit line leakage, which occurs less frequently than cell leakage, is caused by a voltage difference between the bit lines and the storage nodes. One way for reducing bit line leakage is to use body biasing access transistors. Minimal power The basic goals of SRAM schemes are to reduce supply voltage or enhance threshold voltage for each transistor in the SRAM cell.

Early SRAM low-power solutions included a variety of decoding innovations, including the use of dual Vt to reduce leakage power. Despite the use of these approaches, most modern SRAM
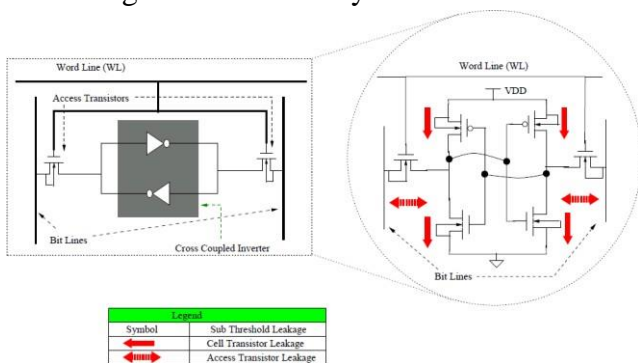
16364

cells still have large leakage currents. This is because the vast majority of leakage control techniques are useless at nanoscale nodes.

Body biasing the access transistors and lowering the gate voltage can reduce leakage power in 6T SRAM cells.

PMOS transistors and dual biasing are further techniques for reducing leakage.

To reduce power consumption while the system is inactive, the SRAM can be turned off selectively at the sub-array level with a low supply voltage. This technique makes use of the data retention features of SRAM cells to keep the supply voltage above the majority of their data retention voltages (DRVs). The data retention voltage is the voltage at which an SRAM cell's data integrity is most likely to be maintained. However, the DRV value frequently varies within a large population of cells. Retention failures occur due to bit flips in SRAM cells caused by an increase in DRV (Data Retention Voltage) when technology nodes drop. As a result, modern last-level caches feature ECC to ensure error-checking security. Many manufacturers use fault-tolerant SRAM systems at the system level, taking advantage of DRV features.

Image Storage for Application Specific Integrated Circuits

Image benchmarks are assessed by evaluating unaltered raw pictures to aid in the initial design phase. The benchmarks used in image and video processing research are available to the public. Figure 3 shows the needed size of the SRAM system for storing a single image. The average benchmark picture size for raw bitmaps is around 1MB.

Figure 4 shows that, on average, over half of all cells have the logical value "0." Because neighboring pixels in pictures have a strong association, it is possible to aggregate adjacent data sets and analyze them together. This thesis uses data storage and inversion techniques to

assure compatibility with a wide range of systems, including those utilized by commercial processors. This thesis looks at data sets that have zeros between consecutive pixels in the same position. Figure 5 shows groups of 2, 4, 8, and 16 data bytes. It also displays the percentage of these groups, known as compress-groups, in which all scanning bit positions for each byte in the group are zeros. For example, suppose two neighboring pixels, each consisting of a single byte with zeros at the most significant bit (MSB) and least significant bit (LSB) places, are merged to make a group. Every alternate position has one. Given that, on average, two of every eight bits (the least significant bit and the most significant bit) are zeros, the image's compress-group to total group ratio is 0.25. It is worth noting that if all pixel elements have a most significant bit (MSB) and a least significant bit (LSB) position with a value of zero, the value will remain constant regardless of whether the group is formed with 4, 8, or 16 neighboring pixels.

Figure 5 shows that the percentage of compress-groups decreases as the group size grows. This thesis focuses on researching a compressed group of size eight, as the benefits are greatest at this number. Use data patterns to selectively deactivate clusters of SRAM cells, reducing leakage power and consolidating energy in the En-Com system.
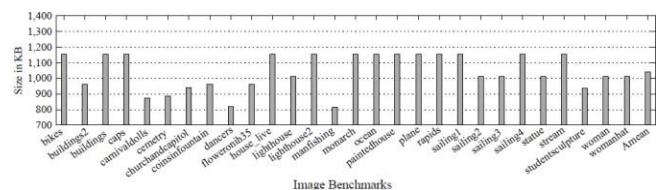
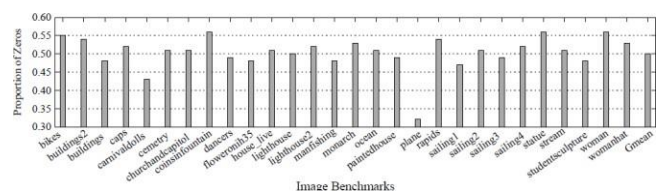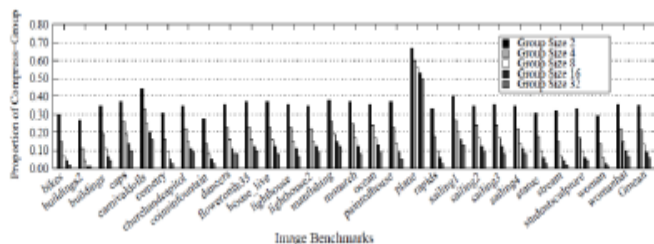

Figure 3: The size of memory occupied by the raw-image

Figure 4: The proportion of SRAM cells that store a value of '0' to total number of SRAM cells

16365

Figure 5: The size of the group is varied and the proportion of compress-groups is compared to total groups. The proportion of compress group is found to exponentially decrease with increasing group size

Because of the rising dominance of leakage power at these nodes, the suggested En-Com System should be used in designs at low nanoscale levels (less than 22 nm). The high Dynamic Random Variation (DRV) values, as well as the ineffectiveness of the DRV technique when combined with Single Error Correction Double Error Detection (SECDED), support this recommendation. This study will evaluate the design using IBM 130nm technology.

3.SYSTEM DESIGN CONSIDERATION AND IMPLEMENTATION

Video and photo data exhibit significant temporal and geographical localization. These structures allow for data reduction by grouping neighboring pixels. At the system level, these pixels can be stored in adjacent SRAM columns. Each row in SRAM can represent a single line of pixels. Each picture pixel can be assigned one-eighth of a bit. The use of a single bit enables the compression of a frequently occurring pattern, such as 00000000, which contains information for all eight bits. Deactivating these eight bits will lower the leakage power. There are two techniques to accomplishing this: finding the recurrent pattern on a row or column level. Each of these options has tradeoffs in terms of layout and implementation
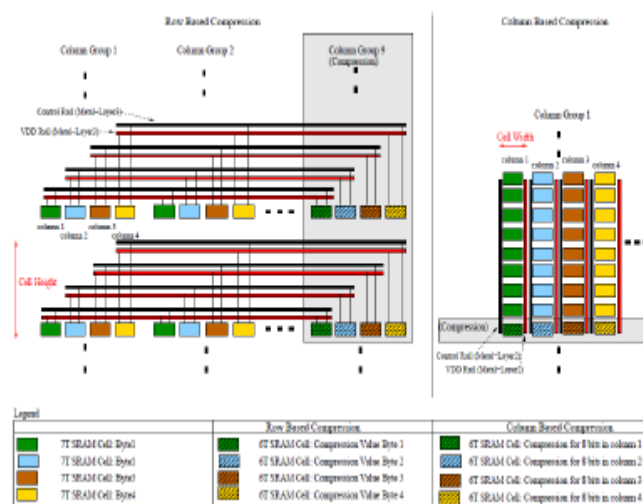
Figure 6: Row Based Compression increases the per cell height by four 'Layer 3' metal lanes per cell. ColumnBased Compression increases the per cell width by two Layer 2 metal layers only

Rather than the previously described row-based approaches, the En-Com system employs a novel columnar data compression methodology. Row-based approaches necessitate the connecting of a large number of metal wires to each SRAM cell. As seen in Figure 6, this results in an increased space required when creating the SRAM cell. The column-based technique employs parallel metal lines organized in columns. The row-based compression strategy adds four "Layer 3" metal lanes to the cell height. In contrast, the column-based technique effectively conserves layout space in this article's design by increasing cell width by only two "Layer 2" metal lanes.

The En-Com system stores the compressed value in a spare cell known as the Zero Switch Cell. The compression pattern for this system is 00000000. The compress group is turned off when the Zero-Switch Cell saves the compressed value (data value 0). The 7T SRAM cells within the compress-group keep their values (data value 0) even when powered off. The read circuitry does not require any modifications to read these deactivated cells; they can be read while remaining in an off state. En-Com is an excellent

16366

framework for programs that analyze images and movies, especially those that frequently compare master frames to succeeding ones. The comparison includes SRAM reads. En-Com can reduce leakage power in particular applications by selectively deactivating it. The Zero-Switch Cell cannot disable the compress-group for any other pattern. To facilitate implementation, the Zero-Switch Cells are initially set to "0". This ensures that the majority of the SRAM is used initially. In addition to the original cell, a "1" is written into the Zero-Switch Cell.

En-Com's use of only one cell per compress-group causes an increase in area overhead as the group size shrinks. An additional burden exists because each compress-group requires an additional SRAM cell to hold its data. Table 1 shows the increase in area required in response to the size of the groups.

Table 1: Compress-Group Area Overhead

Compress-Group Size	Area Overhead
2	50%
4	25%
8	**12.5%**
16	6.25%
32	3.125%

Table 2: Parameters of the Energy Compressed System

Feature	Specification
Supply Voltage	1.2V
Row Decoder	4:16 working at 500MHz (Dynamic)
Column Decoder	2:4 working at 500 MHz (Dynamic)
Sense Amplifer	voltage based
Data Granularity	1 Byte access/cycle
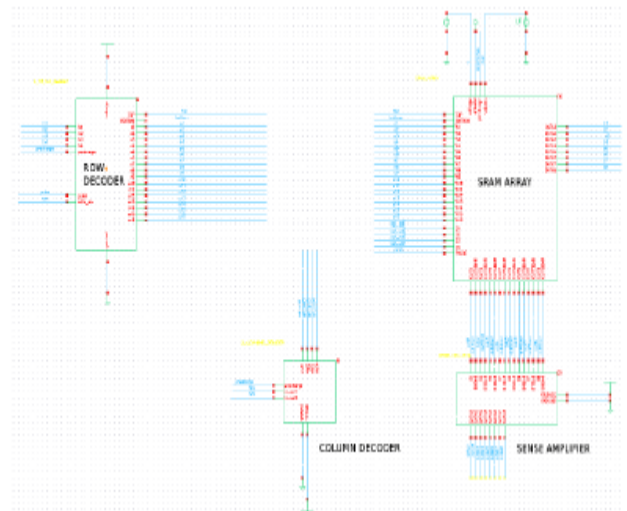
Figure 7: 'En-Com' implementation (Schematic). The topand bottom pins in the decoder and the SRAM array represent the lines to the Zero-Switch Cells

The En-Com SRAM System's write circuits, SRAM column array, and decoders all need to be adjusted somewhat. To allow the double write signal during the process of writing the binary value "1" into the array, En-Com needs make architectural changes. An additional repetition is required to encode the binary value "1". The decoder tree of a huge SRAM memory system encounters significant clock cycle delays, hence surrendering an extra cycle is a realistic concession to improve energy efficiency. Figure 7 shows a schematic of the En-Com system.

The SRAM-based memory system at the 130nm node is built with IBM technology. Dynamic NAND-based decoders allow a 16x32 SRAM subarray to operate at 500 MHz. Table 2 lists the system's specifications.

Design Consideration and Working

The En-Com System compresses energy using a frequently occurring data pattern. We use the data pattern 0000000 (eight zeros) to aid compression. Table 1 and Figure 5 were used to select this figure. This eight-zero pattern is kept in cells that form the design's compress-group. The compress group consists of every SRAM cell with seven transistors. The design also requires the Zero-

Switch Cell, which is a pivot cell that disables the compress-group and records its value. Even when photos only have "1"s in their compress groups, the En-Com system saves inverted data for them. This ensures design uniformity in commercial processors with more "0"s.

The caches include "1"s. Business processors can disregard data inversion. Additional important design aspects are::

- ➤ To function on the compress group, the Zero-Switch Cell requires a small driver and a power-gate transistor.
- ➤ The size of the power gate transistor affects the 8-cell turn-on/off time.
- ➤ The compress group requires a 7T SRAM structure to store data values (0 per cell).
- ➤ This ensures that reads to the compress group function do not change the SRAM's read circuitry.
- ➤ A number of system-level adjustments are required to implement write and read logic for the SRAM Array. The pulse generator needs to be active for one more cycle in order to enable dual write, whereas it needs to be dormant for reads.

The 16x32 array is divided into two vertical portions of eight rows each. In each segment, there is a Zero-Switch Cell in every column. All Zero-Switch Cells are initially set to "0". The Zero-Switch Cell of the SRAM is also written to whenever a 1 is written to the sector. The Zero-Switch Cell only turns off the compress group when all 8 cells in the column have "0's," or when the compress group has the data pattern 00000000.

7T SRAM Cell Design

The compress-group is made up of cells that include seven transistor SRAM. Storing the compressed data value instead of putting the SRAM in a high impedance condition is an improvement over regular 6T SRAM. When the compress-group is activated, the 7T SRAM cell actively maintains the logic zero state. The highlighted transistor (Figure 8) is enabled to prevent the SRAM from entering the high impedance condition when the cell is turned off. If the cell is turned off, the absence of this transistor may cause the sensing amplifier to provide an incorrect signal. When the cell is activated, the ultimate outcome becomes unclear due to positive feedback. To address these issues, 7T SRAM includes a robust '0' storing mechanism. Figure 8 shows the logical form of a 7T SRAM cell.

Compress Group

A "compress group" is a grouping of eight cells with a Zero-Switch Cell attached to each column. To power gate the compress group, enter a '0' into the Zero-Switch Cell. Despite power gating, all cells inside the compress group will preserve the data value '0'. Figure 9 shows how to use the compress group with Zero-Switch Cells.

Zero-Switch Cell

The Zero-Switch Cell consists of an inverter driver and a 6T SRAM cell. The primary use of this entity is to store data about the compress group. The Zero-Switch Cell is engaged when all of the data in the compress group is '0'. A "0" in the Zero-Switch Cell indicates a compressed group sequence that is entirely made up of "0s". For all other patterns, the Zero-Switch Cell holds the value '1'.

Dual Write Pulse Generator

The pulse generator must be activated when dual write is required. The timed signal produced picks the Zero-Switch Cell from the Dual Write Pulse Generator section. If the data value is "1" following a write operation in the original cell, the Dual Write Pulse will also write to the Zero-Switch Cell.

Modification in Row Decoder

The row decoder can be adjusted to accept an additional cell per group. This necessitates the employment of specific signals generated by isolating the Most Significant Bit of the 4:16 row
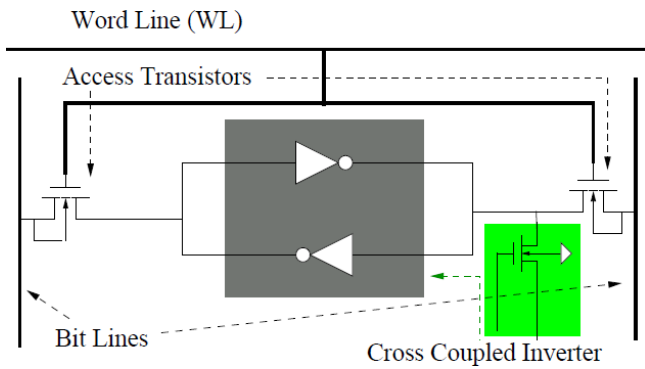
decoder.

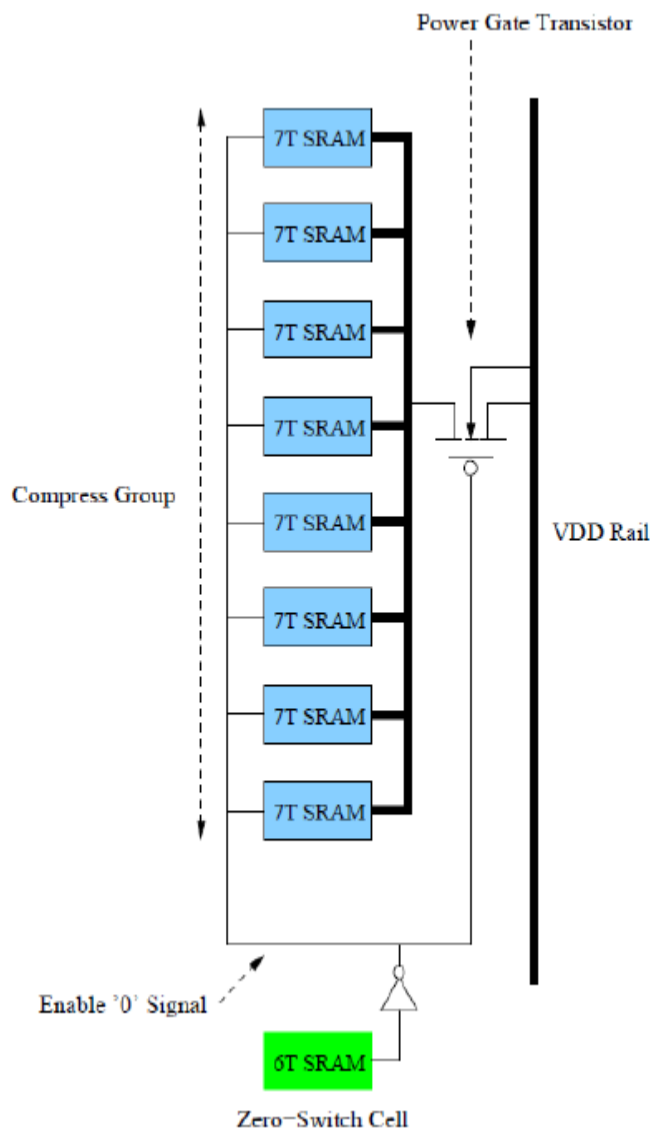**Figure 8: 7T SRAM cell (Logical View)**

Figure 9: An 8 Cell Compress Group

The assertion of the Dual Write Pulse Generator effects cell selection. The two selection signals are

implemented as shown in Figure 10.

Modification in the Write Circuitry

When writing the number "1," the circuitry responsible for writing is tweaked to perform two writing operations. To perform a dual write, the isolation must be maintained for two cycles, as writing entails isolating the sensing amplifier circuitry.
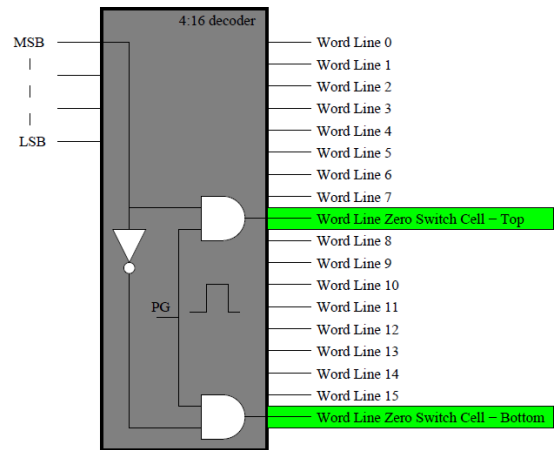

Figure 10: En-Com requires one NOT gate and twoAND gates to select between the top or bottom
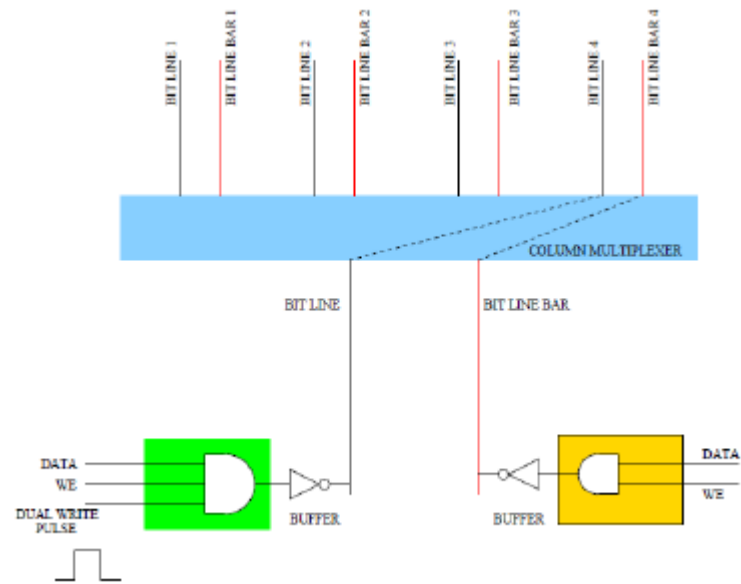
Zero-Switch Cell in case of a write

Figure 11: The write circuit is modified to support dualwrite in case of data value '1'. The pulse generator enables the dual write

Figure 11 demonstrates the argument for

16369

implementing dual writing. To ensure that dual write is only active when a dual write pulse is created, a three-input AND gate is included for writing into the "BIT" position.

4.CONCLUSION AND FUTURE SCOPE

The principal source of leakage in SRAM systems is cell leakage. When these systems are employed to store data, they typically face high power waste due to leakage currents. Typically, they are used in ASICs to store images or video frames, or in commercial processors to save application data. ASICs use processing engines to handle this data. Because of their higher level of complexity, these processing engines perform better when applied to larger datasets. The size and leakage power of these ASICs' SRAMs have increased to match the large amount of data they provide. As these SRAM systems maintain images and video frames for lengthy periods of time, minimizing leakage power is an important consideration when building them.

To reduce power leakage, the study employed an Energy Compression System (En-Com), which deactivates cells using patterns. En-Com is designed as a columnar approach rather than a row-based architecture. En-Com eliminates the need to activate the deactivated cells and allows reading from the SRAM without requiring any changes. To allow these readings, a 7T configuration is constructed in a normal 6T SRAM. En-Com requires an additional Zero-Switch Cell, which is engaged when a cell is compressed and a group switch is triggered.

REFERENCES

1. H. Fujiwara, K. Nii, J. Miyakoshi, Y. Murachi, Y. Morita, H. Kawaguchi, and M. Yoshimoto, "A two-port sram for real-time video processor saving 53bitline power with majority logic and data-bit reordering," in Low Power Electronics and Design, 2006. ISLPED"06. Proceedings of the 2006 International Symposium on, 2006, pp. 61–66.

2. Y. Murachi, T. Kamino, J. Miyakoshi, H. Kawaguchi, and M. Yoshimoto, "A power-efficient sram core architecture with segmentation-free and rectangular accessibility for super-parallel video processing," in VLSI Design, Automation and Test, 2008. VLSI-DAT 2008. IEEE International Symposium on, 2008, pp. 63–66.

3. M. Cho, J. Schlessman, W. Wolf, and S. Mukhopadhyay, "Reconfigurable sram architecture with spatial voltage scaling for low power mobile multimedia applications," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on, vol. 19, no. 1, pp. 161–165, 2011.

4. H. Noguchi, Y. Iguchi, H. Fujiwara, Y. Morita, K. Nii, H. Kawaguchi, and M. Yoshimoto, "A 10t non-precharge two-port sram for 74video processing," in VLSI, 2007. ISVLSI "07. IEEE Computer Society Annual Symposium on, 2007, pp. 107–112.

5. S. Naffziger, B. Stackhouse, T. Grutkowski, D. Josephson, J. Desai, E. Alon, and M. Horowitz, "The implementation of a 2-core, multi-threaded itanium family processor," Solid-State Circuits, IEEE Journal of, vol. 41, no. 1, pp. 197–209, 2006.

6. N. Azizi, A. Moshovos, and F. N. Najm, "Low-leakage asymmetric-cell sram," in Proceedings of the 2002 international symposium on Low power electronics and design, ser. ISLPED "02. New York, NY, USA: ACM, 2002, pp. 48–51.

7. Y.-J. Chang and F. Lai, "Dynamic zero-sensitivity scheme for low-power cache memories," Micro, IEEE, vol. 25, no. 4, pp. 20–32, 2005.

8. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local

binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, pp. 971–987, 2002.

9. M. Qazi, M. Sinangil, and A. Chandrakasan, "Challenges and directions for low- voltage sram," Design Test of Computers, IEEE, vol. 28, no. 1, pp. 32–43, 2011.

10. U. B. et. al, "45nm sram technology development and technology lead vehicle," Intel Technology Journal, vol. 12, no. 2, pp. 111–120, 2008.