Design and Implementation of Low-Power Cache Memory for Mission-Critical Embedded Systems

Anand Raj S.N\textsuperscript{1}, Dr. B. N. Jagadale\textsuperscript{2}
\textsuperscript{1}JNNCE, Dept. of Telecommunication Engineering, Shimoga, India
Email: anand6686@gmail.com
\textsuperscript{2}Dept. of Studies & Research in Electronics, Shankaragatta, India
Email: basujagadale@gmail.com

Abstract—Caches are mainly used to optimize the performance gap between the processor and main memory however accessing the cache more frequently causes the additional burden of increased power consumption by the embedded system. In many mission-critical embedded systems the power consumption is mainly due to memory access, based on the observation that an overwhelming majority of the values written to the cache are “0” as compared and contradicted to writing “1”, a zero-aware SRAM cell with an asymmetric inverter pair, called the ZA cell was implemented to minimize the cache power consumption in writing “0”. This cell saves power when zeros are stored. The pull-up circuit is also modified for the novel cache architecture.

Index Terms—SRAM, ZASRAM, asymmetric inverter pair.

I. INTRODUCTION

The power consumption has become an important consideration on the VLSI system design and microprocessor as the demand for the portable devices and embedded systems continuously increases. The on-chip caches can reduce the speed gap between the processor and main memory. These on-chip caches are usually implemented using arrays of densely packed static random access memories (SRAM) cells. Over the years, power reduction requirement in SRAM has undergone tremendous advancement. As the demand for the portable devices and battery powered embedded systems continuously increases, the power consumption has become an important consideration in the microprocessor and system designs. Since the on-chip caches can effectively reduce the speed gap between the processor and main memory, almost modern microprocessors employ them to boost system performance. These on-chip caches are usually implemented using arrays of densely packed SRAM cells for high performance. Studies show that the power dissipated by the caches is usually a significant part of the total chip power. Cache accesses include read and write operations. Because cache reads occur more frequently than writes, especially for the instruction cache, most low-power techniques focus on reducing the cache power dissipated in reading (i.e., cache read power). However, the write power is usually larger than the read power due to the large power dissipated in driving the cell bit lines to full swing. If a cache has a low hit ratio (that implies a large amount of cache writes), the power dissipated in the write operations become significant. Therefore, in this project mainly concentrating on reducing the RAM write power. Based on this observation, we propose a novel zero-aware SRAM cell, called ZA cell, which drastically reduces the cache power dissipated in writing “0.” The ZA cell consists of an asymmetric inverter pair, one
write port with single bit line and one read port with differential bit lines. The most important characteristics of the ZA cell are summarized as follows. First, in the conventional SRAM cell, because one of two bit lines must be discharged to low regardless of written value, the power consumption in both writing “0” and “1” are the same. In contrast, the ZA cell uses the complement of input data to perform the write operation that prevents the single write bitline from being discharged if then written value is “0”. Therefore, the write “0” power is far less than the write “1” power in the ZA cell. Second, writing cell state from low to high is considerably more difficult in single-bit line configuration because it presents conditions similar to that of the read mode. The boosted word line technique is a traditional solution to this problem, but it induces cell instability and hardware overheads. In random access memories the memory cells are identified by addresses, and the access to any memory cell under any address requires approximately the same period of time. A basic CMOS random access memory (RAM) consists of a Memory cell array, or matrix, or core. Sensing and writing circuit. Row or words address decoder. Column or bits address decoder and Operation control circuit. Memories can be bit-oriented or word-oriented. In a bit-oriented memory, each address accesses a single bit. Whereas in a word-oriented memory, a word consisting of n bits is accessed with each address. Column decoders addressed by Y address bits are often used to allow sharing of a single sense amplifier among 2, 4 or more columns. Most of modern SRAMs are self-timed. By minimizing the power dissipated in writing “0,” the ZA cell can reduce the average cache write power consumption up to 61% and 68%.

A. Design and implementation of SRAM cell array

The SRAM-cell is the core of the memory and here the data bit is stored. In this generator Full CMOS SRAM cell model is used. In Full CMOS SRAM cell, CMOS inverters are used instead of a resistive load. The cell has therefore 6 transistors as shown in fig 1. This affects the cells area negatively, but are less of a problem with current processes. What makes this cell good is the large noise margins and operation at low supply voltage. [1] Memory cells are the key components of an SRAM serving for storage of binary information. A typical SRAM cell is comprised two cross-coupled inverters forming a latch and access transistors. Access transistors enable read and write access to the cell and cell isolation for the not accessed state. An SRAM cell has to provide non-destructive read access, write capability and infinite storage time provided the power is supplied to the cell. Hierarchically, memory cells are arranged in cores, which can be further divided into blocks and arrays depending on the system speed and power requirements. N1, N2, N3 and N4 are NMOS transistors, P1 and P2 are PMOS transistors. P1 and N1 will form a inverter pair named as Inv A. P2 and N2 will form a inverter pair named as Inv B. WL is write line, BL and BL bar are bit lines. [1]

![Figure 1. Basic Full CMOS SRAM Cell](image)

The circuit in Fig 2. makes it possible for write operation to a SRAM cell. In this generator a single ended write circuit is used. This circuit has two inputs, the data signal and an enable signal. The enable signal is active-low. When the data signal is ‘0’ the lower NOR gate will be high on the output and activate the transistor connecting BL to V\_GND. BL will be discharged and a ‘0’ is written in the cell. When the data signal is ‘1’ the output to the upper NOR gate will be high, which turns on the transistor connecting BL bar to V\_GND. BL bar is discharged and a logical ‘1’ is written to the cell. The output of the NOR gates can be connected to several BL or BL bar switches if necessary.[3]
Fig 3. Shows the 8 X 8 SRAM row decoder, it consists of two input nor gate with write_bar and data as input pins. The enable signal is active-low. Pins BL0, BL_bar0, BL1, BL_bar1, BL2, BL_bar2, BL3 and BL_bar3 are used.

The schematic circuits for all the circuit components like row decoder, column decoder, sense amplifier, read circuit, write circuit have to be designed to build a complete SRAM cell and the test circuit was designed. The power consumed by SRAM Cell Array(2 X 2) in writing 1 is found to be 3.536mw where as in writing 0 is found to be 2.964mw.
In this 2 X 2 SRAM cell array, 4 SRAM cells, two pull up circuits, a row decoder, a column decoder, a sense amplifier and 2 write circuits are used, which is shown in figure 4.

**Figure 5. Test Circuit of 2 X 2 SRAM Cell Array**

The test circuit of 2 X 2 SRAM cell array is shown in fig 5, which is designed using the symbol of the same schematic. The test circuit is tested with certain voltage values and delay.

**Figure 6. Simulation Waveform of 2 X 2 SRAM Cell Array for writing “0”**

The above simulation waveform shows the waveform obtained when zero is considered. That is when writing zero to the SRAM cell.

**Figure 7. Simulation Waveform of 2 X 2 SRAM Cell Array for writing “1”**
B. Implementation of ZARAM cell

In zero-aware asymmetric SRAM cell, same supporting blocks required as conventional SRAM. When writing to a memory, a zero is written between 69 and 94 percent of the time. Therefore the zero-aware SRAM cell using an asymmetric inverter pair, here called ZA cell, has been designed to reduce power when writing a zero. The basic idea is to reduce the number of times we have to recharge the bitlines in the memory array. [4]

Fig 4. shows the schematics of the ZA cell and its relative signals, where the write select (WS), write word line (WWL) are used to select a cell for writing, and the data line (WZ) is used when writing data instead of the bit lines. The ZA cell is further development from the full CMOS SRAM cell. By adding two transistors (N3, N4), only one bitline is needed when writing to the cell. [5]

When implementing zero-aware SRAM cell array we made some changes in the pull up circuit. The pull-up circuit has been changed from a “simple pull-up circuit” to a “switchable pull-up circuit”. The new pull-up circuit works much better than the old one. With an enable signal the circuits will not be shorted and unnecessary power loss is avoided. It may however turn out to be necessary to change the transistor sizes to achieve a satisfying speed of the memory.[6]

Decoder is used to select the row (Word lines) of SRAM cell and column decoder is used to select the column of SRAM cell array which are connected the bit lines. Precharge circuit is used to store the previous value. It stores the voltage \( V_{dd}/2 \). Precharge circuit connects with the bit line and bit bar line. When precharge signal is high, it will reset the old value. SRAM cell is designed by using two cross coupled inverters. Inverter using by PMOS and NMOS transistors. SRAM cell has two pass transistors which are used to select the bit lines. RAM cell is used to store the one bit. When row and column decoder selects the one SRAM cell then precharge circuit gives the difference voltage of bit lines, it store the one bit in SRAM cell.

Sense amplifier uses for amplify the voltage. Sense amplifier is connected the bit lines. Bit line will as act one input of sense amplifier and bit bar lines will act as another input of sense amplifier. Output of sense amplifier is the output signal of SRAM cell. In this 2 X 2 SRAM cell array circuit,4 SRAM cells, two pull up circuits, a row decoder, a column decoder, a sense amplifier and 2 write circuits are used. The schematic design shown in fig 9. contains array of zero-aware SRAM cells and supporting circuits as shown for SRAM cell array design. For the design of zero-aware SRAM cell array only change is the zero-aware SRAM cell in place of SRAM cell. The test circuit for the above schematic is shown in figure 6.14, which is designed using the symbol of that schematic. The test circuit is tested with certain voltage values and delay.

The power consumption in SRAM cell array is more than in zero-aware SRAM cell array. From the waveforms, the negatives are obtained. Magnitude of are taken in order to get power consumption value. Here the power consumption calculation for 2 X 2 is explained. Further SRAM and zero-aware SRAM are compared. In the earlier work on low power cache design techniques single 6 transistor SRAM cell consumed 5.6576mw, for writing "1" and 4.7424mw for writing "0" which is 60% more than the proposed work[2].

268
TABLE I. COMPARISON OF POWER CONSUMPTION FOR SRAM AND ZERO-AWARE SRAM CELL ARRAY

<table>
<thead>
<tr>
<th>SRAM Cell Array (2 X 2)</th>
<th>Zero-Aware SRAM Cell Array (2 X 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Power Consumption</td>
</tr>
<tr>
<td>1</td>
<td>3.536 mw</td>
</tr>
<tr>
<td>0</td>
<td>2.964 mw</td>
</tr>
</tbody>
</table>

II. CONCLUSIONS

Most low-power SRAM techniques only reduce read power, but generally write power is larger than read power. The implemented ZA cell differs slightly from the traditional Full CMOS cell in such way that it has two more transistors which controls whether the stored value should be changed or not. Drawbacks with this cell are that the size of the cell is increased due to the two transistors added. The memory cell itself has a power reduction up to 95.7% when writing to zero compared with the full CMOS cell. This is when all circuitry has been taken into account. Even though the ZA-cell needs more area it is more power efficient than the full CMOS cell. It is however recommended to use as large words as possible to achieve the lowest power consumption. Finally by comparing the power consumption of both SRAM and zero-aware SRAM cell array we can construct a table. For both writing zero and one. Because of the lower power consumption during write operation, the zero-aware SRAM cell can be used in many applications of CMOS circuit designs and in battery powered embedded systems. Circuit power consumption can be reduced by implementing test circuit from schematic. In future this zero-aware SRAM can be implemented with BiCMOS technology and also thesis has to be developed in order to reduce silicon area, so that number of transistors required can be minimized.

REFERENCES


J. Montanaro et al., “A 160 MHz, 32 b 0.5 W CMOS RISC microprocessor,” in IEEE ISSCC 1996 Dig. Papers, 1996.

