

# A Comparative Study of Power Efficient SRAM Designs

Jeyran Hezavei, N. Vijaykrishnan, M. J. Irwin

Pond Laboratory,  
Department of Computer Science & Engineering,  
Pennsylvania State University

{hezavei, vijay, mji}@cse.psu.edu

## ABSTRACT

*This paper investigates the effectiveness of combination of different low power SRAM circuit design techniques. The divided bit line (DBL), pulsed word line (PWL) and isolated bit line (IBL) strategies have been implemented in a various size SRAM designs and evaluated using 0.35Micron technology and 3.3V VDD at 100MHz frequency. Different decoder structures have been investigated for their power efficiency as well. It is observed that the power reduces by 29%, 32% and 52% over an unoptimized SRAM design when (PWL+IBL), (PWL+DBL) and (PWL+IBL+DBL) are implemented in a 256\*2 size SRAM respectively.*

## Keywords

Low Power, SRAM, Decoder.

## 1. INTRODUCTION

In many current applications such as multimedia, the memory system has been demonstrated to be the main power-consuming unit. Thus, a significant effort has been invested in reducing the power of CMOS RAM chips using circuit and architectural techniques. For DRAM's, the main techniques used to reduce power are partial activation of multi-divided arrays (both for word-line and bit-line), half-VDD precharging, and lowering operating voltage using external power supply reduction [4][5]. The techniques used in SRAM's are of particular interest due to the large SRAM-based on-chip cache structures employed in current processors. Partially activating a divided bit and word lines, isolating the sense amplifier from the bit line, pulsing the word driver and column circuitry, reducing the bit/word-line swings, and charge recycling in the I/O buffer are some of the

techniques that can reduce the SRAM power consumption [4][5].

In [1], an automatic-power-save architecture, a pulsed word technique and an isolated bit line technique reduced the power dissipation of the cache memory to almost 60% at a frequency of 60MHz and to 20% at 10MHz by these techniques for 0.5Micron technology, 3.3V VDD. In [2], a novel hierarchical divided bit-line approach for reducing active power in SRAMs by reducing bit-line capacitance was introduced. This approach was found to reduce the power consumption by 50-60%. In this paper, we apply a combination of divided bit-line approach and the isolated bit-line approach along with the pulsed word-line technique [3] in reducing the power consumption in SRAM caches. In addition, we investigate the design of low power decoder designs for the cache. Various decoder designs have been investigated [6,7] for speed and power in the past. In this paper, we analyze six different decoder structures for their power efficiency.

The rest of the paper is organized as follows. Section 2 details the design of the different components of the SRAM structure for the different optimizations. The cache power result for the different combination of optimizations is presented in section 3. Finally, we close with conclusions.

## 2. SRAM POWER REDUCTION TECHNIQUES

The floorplan of basic cache block is shown in (Fig. 1). The core is a matrix of standard 6-transistor SRAM cells. Rows are the word lines and columns include the bit lines. In this work the core's columns were initially divided into smaller chunks (divided bit line (DBL) (Fig. 2))[2,6]; however, the rows remained continuous. DBL enhances the read/write speed of the circuit as the long bit lines are split into shorter ones and the bit line capacitances are decreased. This eventually reduces the power consumption significantly.

DBL also allows us to split the row decoders as only one segment needs to be activated at a time and the rest of the segments can be kept idle. It will be explained later in this paper that splitting the decoder will reduce the decoder power consumption. The decoder is a multi stage structure constructed of basic blocks, which were made of dynamic 3-input nand gates. It provides pulsed outputs. This way allows controlling the word line drivers, which enables

us to implement the Pulsed Word Line (PWL) scheme with no additional hardware overhead. The idea of PWL is to minimize the duration of active input on word lines by deactivating the word lines (and SRAM cells) before the bit line voltages make a full swing. This leads to a reduced power consumption and enhanced speed.

The last technique used in this work is the Isolated Bit line Scheme (IBL). Sense amplifiers are included in the memory-read-circuitry to speed up the read operation. Here, the sense amplifiers attached to the bit lines are isolated after they detect a sufficient voltage difference on the bit line and bit line/. This prevents a full swing on the entire bit line and saves energy. The sense amplifiers are also isolated from the bit lines during the entire write operation.

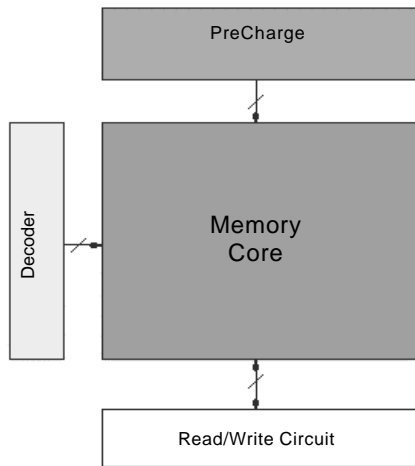


Fig. 1: Basic Memory Block Diagram

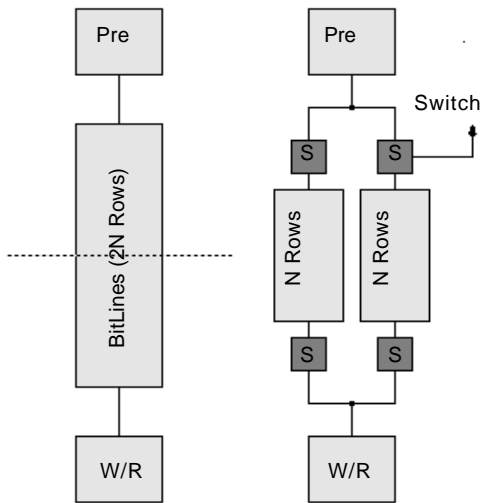


Fig. 2: Divided Bit line Scheme (DBL), Switches are Controlled by input address.

## 2.1 Decoder

The speed of the row decoder has a great impact on memory performance [6]. The row decoder drives the word lines of the SRAM array. At each read/write operation only one driver is active, making the SRAM cells connected to the corresponding word line accessible. The pulsed word line scheme (PWL) can be implemented by gating the outputs of standard decoders with a control enable pulse (from the sensing of voltage swings on the bit lines), which limits the duration of active output signals. In our design, the PWL control is integrated in the row decoder design by use of dynamic logic.

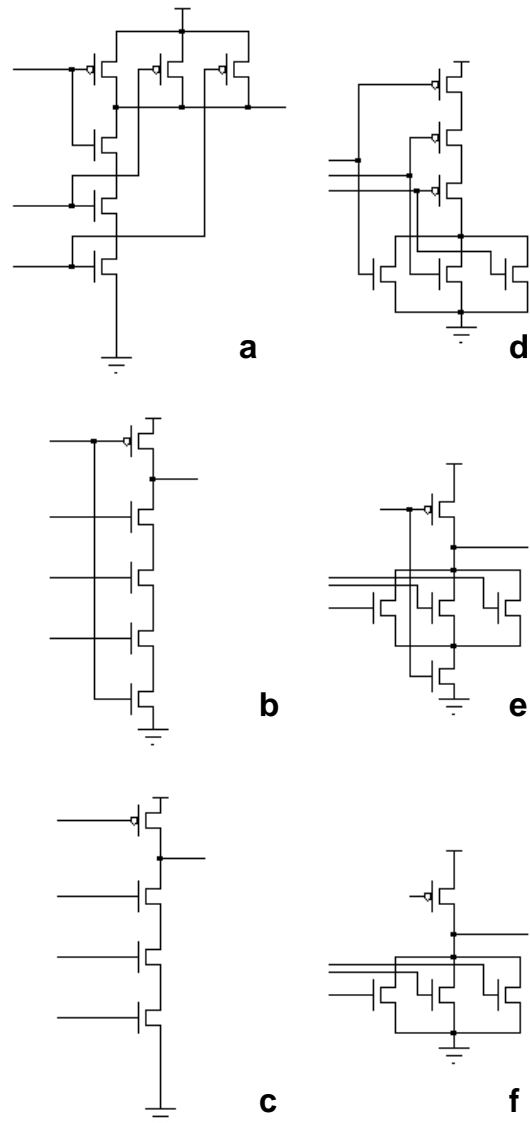


Fig. 3: Different 3 Input Basic Gates: a) CMOS Nand, b) Dynamic Nand, c) Skewed Nand, d) CMOS Nor, e) Dynamic Nor, f) Skewed Nor.

### 2.1.1 Basic Decoder

Using fast, low power basic gates, can greatly optimize the delay and power consumption of the decoder. (Fig. 3) shows schematics of different core 3-input gates, which were used to build basic 3x8 decoders. In (Fig. 4) comparison between the hspice simulation results of all schematics is presented. The schemes used are as follows: a) CMOS Nand, b) Dynamic Nand, c) Skewed Nand, d) CMOS Nor, e) Dynamic Nor, f) Skewed Nor. Dynamic Nands with shared NMOS and PMOS transistors consume the least amount of power as they disconnect the direct VDD/GND path all the time [9]. Based on these results and according to dynamic gates' behavior 3X8 basic decoders were built using 3-input dynamic nand gate.



Fig. 4: Average Power Consumption & Worst Case Delay of Circuits in Fig. 3.

### 2.1.2 Total Structure

The complete decoder can be implemented by cascading the basic 3X8 decoders in multi stages (based on the size of the decoder). (Fig. 5) shows a sample 6x64 decoder. Basic decoders are 3X8 decoders explained in the previous part. A zero voltage on precharge input of the first stage grounds all the outputs of the first stage connected to the precharge input of the second stage decoders and consequently all the final outputs. When the precharge is at high voltage only one output of the first stage is going high. Hence, only one single output of the final stage will be active. This keeps the other components of the second stage idle and as no switching occurs in them they consume very little power. An example is shown in (Fig. 5). For an address ranging between 000000 and 000111 only the mentioned area is switching and the rest of the circuit sits idle. This way the dynamic power is minimized, while controllable pulses are generated at the output (for PWL).

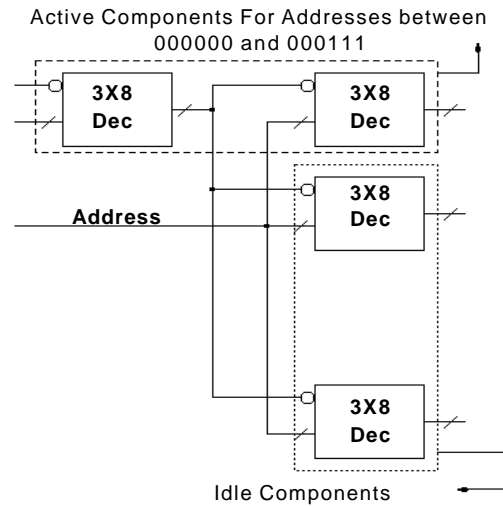


Fig. 5: Two Stage 6X64 Decoder Showing Active and Idle Areas for Addresses Ranging between 000000 & 000111.

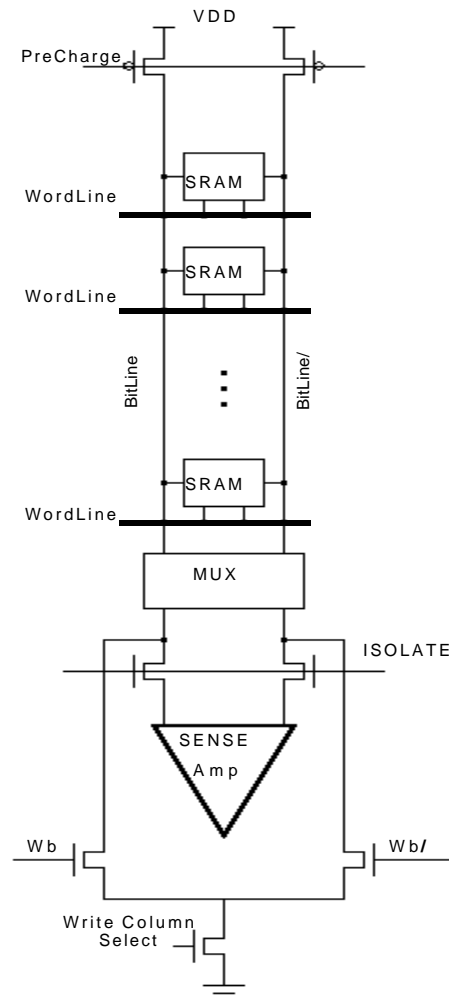


Fig. 6: Overall View of Bit Line Structure and Connected Circuits.

## 2.2 Memory Core: Bit Line Architecture

An overall view of the bit line structure used in this work is shown in (Fig. 6). Each column includes the following parts:

### 2.2.1 SRAM

Generic 6-transistor SRAM cells (Fig. 7) [9] were used as memory core cells. They are costly and occupy large areas, but they are widely used due to important advantages including higher speed and less static current (consequently less power consumption).

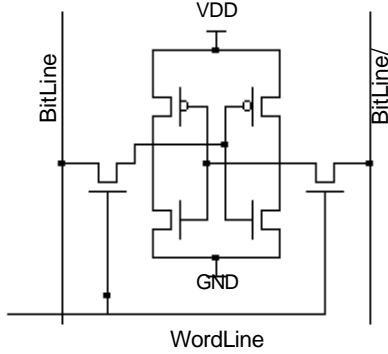


Fig. 7: Six-Transistor SRAM Cell

### 2.2.2 Pre charge Logic and Column MUX

A pair of pull up transistors per each column was used in the design [8]. The pull up transistors are turned on before each read/write operation precharging both the bit line and bit line/ to a high voltage and then will be switched off at the beginning of the operation leaving the bit lines at the same voltage. During the operation either bit line or bit line/ will have a voltage swing while the other one will stay at precharge level voltage.

While a precharge operation is performed the column MUX's disconnect the bit lines from the read/write circuitry. This minimizes the occurrence of direct Vdd/Gnd paths in the column, which is one of the major sources of power consumption in every circuit. However, this is not the major task of the column MUX. Column MUX's are basically used to avoid duplicating the read/write circuitry for all bit lines. A simple tree MUX was used in this work.

### 2.2.3 Read Circuitry

**Sense Amplifier:** During the read operation the voltage on one of the bit line or bit line/ will slightly start dropping. A full swing on the line is rather a slow operation. Hence, sense amplifiers are used to sense the slight voltage difference and amplify it to a correct data value. Additional stages can boost the speed of read operation significantly [8]. A two stage sense amplifier was used (Fig. 8) in this work.

**Isolation Transistors:** In IBL scheme a pair of isolation transistors are used to disconnect the sense amplifier from the bit lines by the time the correct data is detected. This technique reduces the read power as it prevents the complete swing on the bit lines. It also disconnects the sense amplifiers from the bit lines during write operation as they are not needed. Generally the

isolation transistors can be turned off after a minimum 10% voltage difference is sensed between the lines.

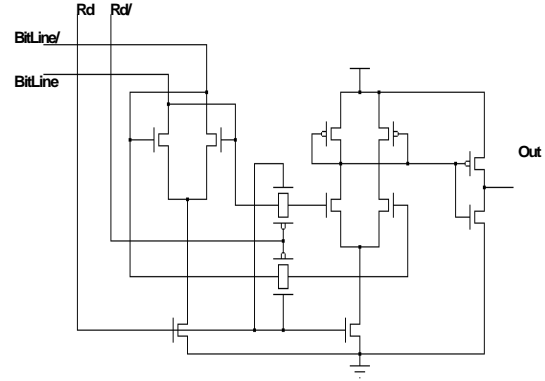


Fig. 8: Two Stage Sense Amplifier

### 2.2.4 Write Circuitry

Two pass transistors controlled by WB and WB/ signals are the devices used to control write operation. The two mentioned signals are assumed to be generated by a control unit not included in memory circuitry [8]. As mentioned, before the beginning of write operation both bit line and bit line/ are precharged and all the word lines are grounded. During the data write operation only one of the two signals is active and the connected pass transistor grounds the relative line forcing a zero voltage to that line. Depending on whether bit line or bit line/ is grounded a “zero” or “one” value will be written into the active SRAM cell respectively.

## 2.3 Cache Power Characterization

The layout of different memory power optimizations were done using the magic design tools using 0.35Micron technology and simulated using Avant! Hspice (100 MHz frequency and power supply voltage of 3.3 V). Initially DBL, IBL and PWL schemes were implemented. (Fig. 9) shows the percentage of average power consumed by each subcomponent during different operations. Operations are listed in (Table 1). The results shown were averaged over simulations performed on different SRAM configurations ranging between (64, 128, 256 bits bit line size and 1, 2 and 4 bits word line size).

	<i>Name</i>	<i>Operation</i>
1	<i>H0-W0</i>	<i>Cell Holds 0 a 0 is written into that</i>
2	<i>H1-W1</i>	<i>Cell Holds 1 a 1 is written</i>
3	<i>H0-W1</i>	<i>Cell Holds 0 a 1 is written</i>
4	<i>H1-W0</i>	<i>Cell Holds 1 a 0 is written</i>
5	<i>R0</i>	<i>A 0 is read</i>
6	<i>R1</i>	<i>A 1 is read</i>

Table 1: Different Memory Operations

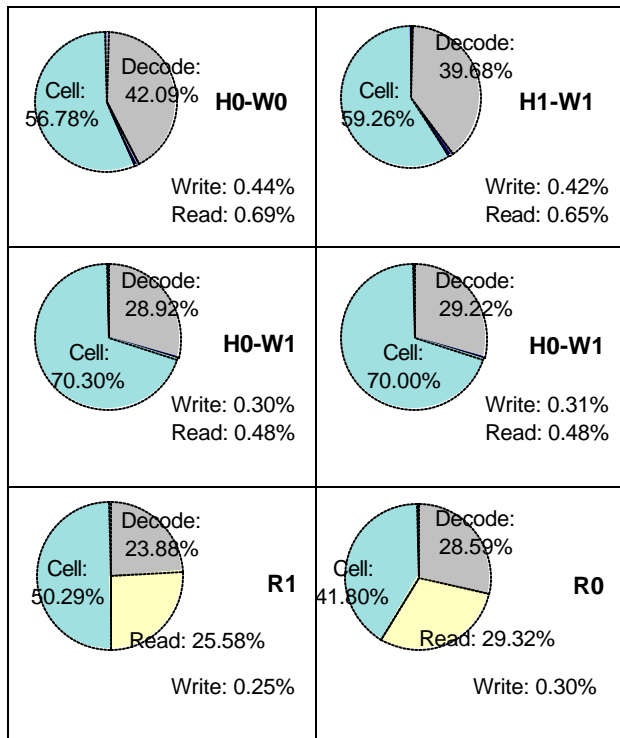


Fig. 9: Average Power Consumption Percentage of Different Memory Components During Different Operations (Table 1).

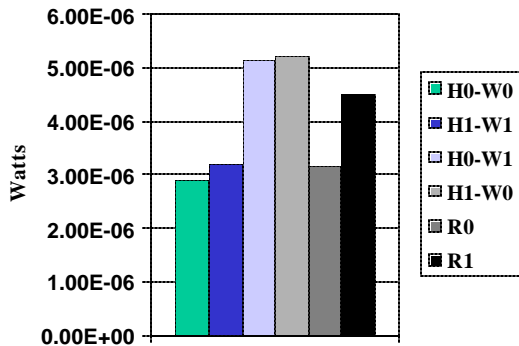


Fig. 10: Average Power Comparison of Single Memory Cell For Operations Listed in Table 1.

According to (Fig 9) the major amount of power is consumed by cell array. The amount of cell power consumption is different in each access according to the operation (Read/Write) and the values ('0'/'1'). (Fig. 10) shows the comparison between power consumed by each single cell (6-transistor SRAM) in the core during each operation. The characterizing information provided in this figure can serve as useful information for modeling high-level cache power models for the 0.35Micron designs.

### 3. POWER SAVINGS

(Fig.11) shows the average power consumed by a 256\*2 SRAM with different memory power reduction schemes and compare them for different operations listed in (Table1). PWL is initially implemented in all schemes. The average power savings gained by different schemes are 29%, 32% and 52% using (IBL+PWL), (DBL+PWL) and (DBL+IBL+PWL) respectively. The power savings obtained using the IBL and DBL techniques are much less as compared to [1] and [2] due to the smaller size of our cell array and the decoder power is more significant in our design. Also, we observe a potential for more optimizations by combination of different circuit optimizations.

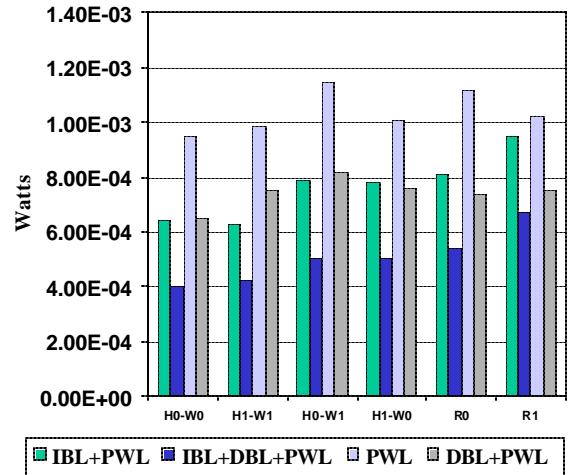


Fig. 11: Memory Average Power Consumption Using Combinations of Different Power Saving Schemes.

### 4. CONCLUSION

The effectiveness of a combination of different low power RAM design circuit techniques that included the divided bit line (DBL), pulsed word line (PWL) and isolated bit line (IBL) strategies was investigated. A series of different size SRAM's using these techniques was implemented in 0.35Micron, 3.3V technology and simulated at 100MHz frequency. It is observed that the power reduces by 29%, 32% and 52% over an unoptimized SRAM design when (PWL+IBL), (PWL+DBL) and (PWL+IBL+DBL) are used in a 256\*2 size SRAM.

### 5. REFERENCES

- [1] Shimazi Y., et. al., An automatic-power-save cache memory for low-power RISC processors, IEEE Symposium on Low Power Electronics, pp. 58-59, 1995.
- [2] A. Karandikar and K. K. Parhi, Low power SRAM design using hierarchical divided bit-line approach, Proc. of International Conference on Computer Design, pp.82-88, 1998.

- [3] D. T. Wong, An 11-ns 8K\*18 CMOS static RAM with 0.5- $\mu$ m devices, IEEE Journal of Solid-State Circuits, 23(5):1095-1103, Oct. 1998.
- [4] K. Itoh, K. Sasaki, Y. Nakagome, Trends in low-power RAM circuit technologies, Proceedings of the IEEE, 83(4):524-543, April 1995.
- [5] M. B. Kamble, K. Ghose, Energy Efficiency of VLSI Caches: A Comparative Study, Proc. IEEE International Conference on VLSI Design, pp. 261-267, 1997.
- [6] B. Amrutur, Design and Analysis of Fast Low Power SRAMs, Ph.D. Thesis, Department of Electrical Engineering, Stanford University, 1998.
- [7] B. Bhaumik, et. al., A low power 256 KB SRAM design, Proc. 12th International Conference on VLSI Design, pp. 67-70, 1999.
- [8] S. M. Kang, Y. Leblebici, CMOS Digital Integrated Circuits Analysis and Design, Second Edition, McGraw-Hill, 1999.
- [9] J. M. Rabaey, Digital Integrated Circuits a Design Perspective, Prentice Hall, 1996.