

Scaling Consideration of BiCMOS SRAMs

J. J. Tsaur, C. W. Jih, H. W. Tsaur, J. B. Kuo

Rm. 526, Dept. of Electrical Eng., National Taiwan University
Roosevelt Rd., Sec. 4, #1, Taipei, Taiwan 107
FAX:886-2-363-8247, Telephone:886-2-363-5251, E-mail:t7503007@twntucc1.bitnet

Abstract

This paper presents scaling consideration of BiCMOS SRAMs based on an experimental 4K ECL I/O BiCMOS SRAM design. According to analysis, further integration of the BiCMOS SRAMs over 1Mbits is limited by the evolution of the CMOS processing technology and the BiCMOS circuit design techniques, and BiCMOS circuits shrink the access time without a power penalty.

Summary

Recently, BiCMOS technology has been becoming one of the major VLSI technologies. BiCMOS SRAMs, combining CMOS memory cells, bipolar sense amps and ECL I/O buffers, have shown their capabilities for high-speed and large-size memory applications. Scaling of CMOS technologies, devices, and SRAM circuits has been intensively reported. Scaling of bipolar ECL devices and circuits has also been discussed. However, there are few reports on scaling consideration of BiCMOS devices and SRAM circuits. In this paper, scaling of the BiCMOS SRAMs will be described.

Fig. 1 shows the performance of the recent BiCMOS, CMOS, and ECL SRAMs in terms of per-bit power access time product as a function of size of integration. For CMOS SRAMs, the per-bit power access time product has decreased from 20fJ/bit for a 64K SRAM to 7fJ/bit for a 1M SRAM. For ECL SRAMs, the per-bit power access time product stays around 1000fJ/bit for below-256K SRAMs. As for BiCMOS SRAMs, the per-bit consumed energy is between CMOS and ECL ones. The continuing reduction on the per-bit consumed energy for the CMOS and Bi-

MOS SRAMs is owing to the scaling of the CMOS technology. As a matter of fact, the per-bit consumed energy of the recent BiCMOS SRAMs is close to CMOS ones as a result of the dominance of the CMOS memory cell array in a SRAM chip. On the other hand, as shown in Fig. 2, the access time of the BiCMOS SRAM, which is comparable to ECL one, is much smaller as compared to CMOS one. Consequently, the BiCMOS SRAM has the ECL speed and the CMOS low power capabilities. In order to investigate the role of the BiCMOS circuit techniques using scaled BiCMOS technology, Figs. 1 and 2 have been reorganized as shown in Figs. 3 and 4. The normalized per-bit consumed energy is defined as the per-bit power access time product divided by the square of the basic unit of the design rule, e.g. $L=1.2$ for the 4K SRAM [14] using a $1.2\mu\text{m}$ BiCMOS technology. Using the normalized per-bit consumed energy definition, the figure-of-merit of each SRAM can be evaluated in terms of the circuit performance only, excluding the benefits coming from the technology scaling. As shown in Fig. 4, the normalized access time, which is defined as the access time divided by the basic unit of the design rule, for the CMOS SRAM increases for every SRAM generation. On the other hand, the normalized access time of the BiCMOS SRAM stays about the same, which implies that the bipolar devices have been utilized effectively in the BiCMOS circuit design. The usefulness of the BiCMOS circuit design techniques is especially visible for large-size SRAMs. Fig. 5 shows the normalized size of a memory cell. For BiCMOS and CMOS SRAMs, the normalized size of the memory cell stays about the same in spite of scaling of the technology. As shown in Fig. 6, the percentage area of the memory cell array in a BiCMOS SRAM, which is defined as the product of the memory cell size and the cell number divided by the overall die area, is also comparable to CMOS, which implies the inclusion of the bipolar devices in the BiCMOS SRAMs does not lower the efficiency of the layout area.

In order to investigate the future trend on scaling BiCMOS SRAMs, an experimental 4K ECL I/O compatible BiCMOS SRAM using a $3\mu\text{m}$ BiCMOS technology has been designed. Fig. 7 shows the layout, the floor plan and the block diagram of the BiCMOS SRAM. The 6T memory

cell occupies about an area of $44\mu\text{m} \times 44\mu\text{m}$ with a proper choice of aspect ratio for each transistor for an optimized signal-to-noise margin. This chip, which occupies an area of $0.44\text{cm} \times 0.43\text{cm}$, is divided into two halves with X decoders in between. The pitch of each X decoder matches the size of each memory cell. Each half has 64 word lines and 32 bit lines. There are 15 ECL input buffers for the 12 address, chip select, write enable, and data-in inputs followed by the ECL/CMOS translator and the X and Y BiCMOS decoders. The outputs of the X decoders are used to drive 64 word lines, which control the access transistors of each memory cell. The outputs of the Y decoders are used to control the 64 bit-line pairs. The selected bit-line pair is connected to the sense amps, followed by the output buffer. Fig. 8 shows the input buffer circuit, an ECL/CMOS translator, and a BiCMOS buffer for driving a large capacitive load. As also shown in Fig. 8, a NAND/NOR structure has been used in the decoder. Fig. 9 shows the bit-line related circuit. A variable impedance bit line load has been used to facilitate an efficient read and write operations. Together with write-enable, the outputs of the decoders are used to control the write and read access of the bit lines. Fig. 10 shows the two-stage sense amp structure. In the output buffer, a 4X ECL buffer has been used, followed by a 90X open-emitter transistor to drive an output load of 50ohm and 20pF. For the overall layout consideration, power lines have been arranged such that they can be accessed from every part of the chip. Input buffers are in the upper portion of the chip and output buffers are at bottom. The power lines of the input and output buffers are isolated from the rest of the circuits to reduce noise. Each input and output buffer consumes 20mW and the second-stage sense amp consumes 36mW. The total power dissipation of the chip is about 350mW. Fig. 11 shows the transient performance during the read and write operations for the experimental SRAM. During read, applying a voltage step from -1V to -1.8V at the address input, the access time is 13ns. During write, a write pulse of 10ns is sufficient. Compared to other large-size SRAMs, the normalized access time ($1.5\text{ns}/\mu\text{m}^2$) and the normalized per-bit power access time product ($100\text{fJ}/\text{bit}/\mu\text{m}^2$) are competitive. The percentage area of the memory cell array is comparable to other large-size SRAMs.

In order to evaluate the performance of the scaled BiCMOS SRAMs, the access time of a BiCMOS SRAM is analyzed based on the design of the experimental 4K BiCMOS SRAM. The read access time of a BiCMOS SRAM is determined by the ECL-related and BiCMOS-related delays. The ECL-related delays include the input and output-buffer delays and sense-amp delays. The BiCMOS-related delays are determined by the decoder delay, the word-line delay, the bit-line delay, and the delays due to the parasitics between input buffer and decoder and between sense amp and output buffer. Currently, scaling of BiCMOS technology and device is emphasized on shrinking the CMOS devices. The scaling of the bipolar device in the BiCMOS structure has not reached an agreement due to the vertical

nature of the BJT. The lateral dimension of the bipolar device also affects device performance, which is critical for BiCMOS circuits. In addition, the integration density of BiCMOS circuits is affected by the lateral dimensions of the bipolar device. On the other hand, for a BiCMOS SRAM, since bipolar devices have been used only in I/O buffers, sense amps, the driving capabilities of the ECL-related circuits is less sensitive to scaling of the BiCMOS technology. Therefore, the impact of scaling on the performance of BiCMOS SRAM is basically coming from the down-scaling of the size of the CMOS memory cell and the up-scaling of the integration size, which determines the parasitic capacitive load of the word lines and the bit lines. As a result, the access time of a scaled BiCMOS SRAM is determined mainly by the word-line and the bit-line delays. As for the ECL circuits, the scaled delay is mostly influenced by the reduction in the allocated power. In the analysis, both CMOS and BiCMOS technologies are assumed to scale down by a factor of two for every other generation of the SRAMs, i.e., a $3\mu\text{m}$ technology for the 4K SRAM, a $1.5\mu\text{m}$ technology for 16K and 64K SRAMs, a $0.75\mu\text{m}$ technology for 256K and 1M SRAMs, and a $0.4\mu\text{m}$ technology for 4M and 16M SRAMs. Based on the above analysis, the access time of the scaled BiCMOS SRAM is shown in Fig. 12(a). The access time increases for up-scaled SRAMs. Also shown in the figure is the scaled access time based on the partitioned memory cell array structure and tree-structure decoders. Thanks to the BiCMOS drivers, the access time of the scaled BiCMOS SRAMs using partitioned memory array and tree-structure decoders can be maintained in spite of the increase in the size of integration. The normalized per-bit consumed energy of the scaled BiCMOS SRAM as shown in Fig. 12 is similar to the CMOS curve. Consequently, the performance of down-scaled BiCMOS SRAMs is determined by the BiCMOS circuit designs. In conclusion, further integration of the BiCMOS SRAM is limited by the evolution of the CMOS processing technology and the BiCMOS circuit design techniques, which can help resolve the access time disadvantage resulting from the increased size without a substantial power penalty.

References

- [1] Y. Makie et. al., "A 6.5ns 1Mb BiCMOS ECL SRAM," *ISSCC Digest* 2/90
- [2] M. Takada et. al., "A 5ns 1Mb ECL biCMOS SRAM," *ISSCC Digest* 2/90
- [3] J. J. Tsaur et. al., "A 13ns BiCMOS SRAM using $3\mu\text{m}$ BiCMOS Technology," to be published.
- [4] F. Towler, et. al., "A 128K 6.5ns Access/5ns Cycle CMOS ECL Static RAM," *ISSCC Digest*, 2/89
- [5] M. Suzuki, et. al., "A 3.5ns 500mW 16Kb BiCMOS ECL RAM," *ISSCC Digest*, 2/89
- [6] K. Sasaki et. al., "A 9ns 1Mb CMOS SRAM," *ISSCC Digest*, 2/89

[7] H. Tran et. al., "An 8ns BiCMOS 1Mb ECL SRAM with a Configurable Memory Array Size," *ISSCC Digest*, 2/89

[8] M. Matsui et. al., "An 8ns 1Mb ECL BiCMOS SRAM," *ISSCC Digest*, 2/89

[9] H. Shimada et. al., "An 18ns 1Mb CMOS SRAM," *ISSCC Digest*, 2/88

[10] F. List et. al., "A 25ns Full-CMOS 1Mb SRAM," *ISSCC Digest*, 2/88

[11] H. S. Lee et. al., "An Experimental 1Mb CMOS SRAM with Configurable Organization and Operation," *ISSCC Digest*, 2/88

[12] S. Flannagan et. al., "A 16ns 256Kx1 CMOS SRAM," *ISSCC Digest*, 2/88

[13] H. V. Tran et. al., "An 8ns Battery Back-Up Submicron BiCMOS 256K ECL SRAM," *ISSCC Digest*, 2/88

[14] T. S. Yang et. al., "A 4ns 1-bit Two-Port BiCMOS SRAM," *IEEE J. of Solid - State Ckts*, 10/88

[15] N. Okazaki et. al., "A 30ns 256K Full CMOS SRAM," *ISSCC Digest*, 2/86

[16] K. Ogiue et. al., "A 13ns/500mW 64Kb ECL RAM," *ISSCC Digest* 2/86

[17] K. Yamaguchi et. al., "A 3.5ns 2W 20mm² 16Kb ECL Bipolar RAM," *ISSCC Digest* 2/86

[18] M. Isobe et. al., "A 46ns 256K CMOS RAM," *ISSCC Digest* 2/84

[19] K. Hardee et. al., "A 30ns 64K CMOS RAM," *ISSCC Digest*, 2/84

[20] Ozawa et. al., "A 25ns 64K SRAM," *ISSCC Digest*, 2/84

[21] F. Tokuyoshi et. al., "A 2.3ns Access Time 4K ECL RAM," *ISSCC Digest* 2/84

[22] O. Minato et. al., "A 20ns 64K CMOS SRAM," *ISSCC Digest*, 2/84

[23] S. Schuster et. al., "A 20ns 64K NMOS RAM," *ISSCC Digest*, 2/84

[24] R. A. Kertis, et. al., "A 12ns ECL I/O 256Kx1 SRAM using a 1 μ m BiCMOS Technology," *IEEE J. of Solid - State Ckts*, 10/88

[25] N. Tamba et. al., "An 8ns 256K BiCMOS SRAM," *IEEE J. of Solid - State Ckts* 8/89

[26] E. Seevinck et. al., "Static-Noise Margin Analysis of MOS SRAM Cells," *IEEE J. of Solid - State Ckts*, 10/87

[27] J. Lohstroh et. al., "Worst-Case Static Noise Margin Criteria for Logic Circuits and Their Mathematical Equivalence," *IEEE J. of Solid - State Ckts*, 12/83

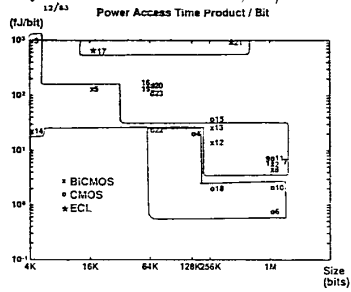


Fig. 1.

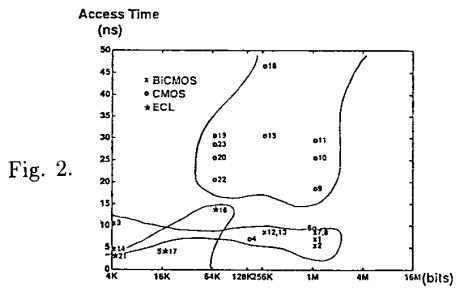


Fig. 2.

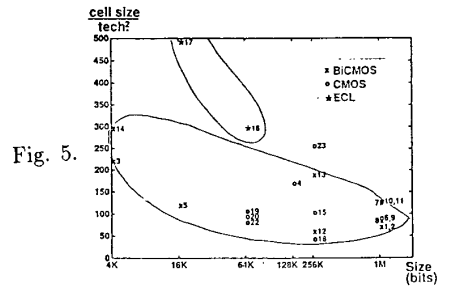


Fig. 5.

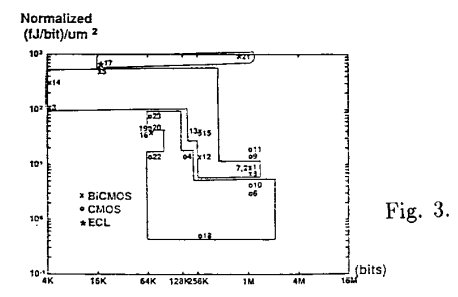


Fig. 3.

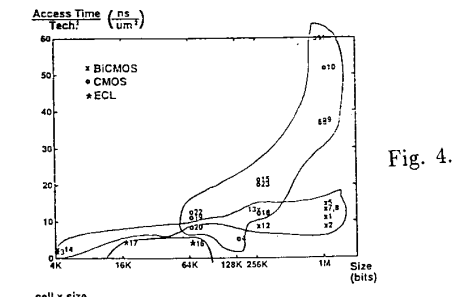


Fig. 4.

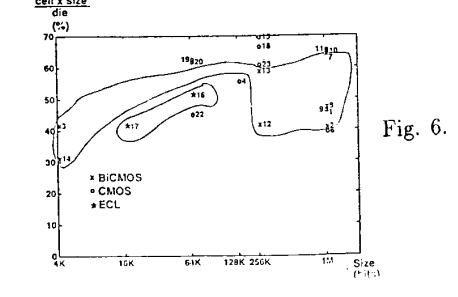


Fig. 6.

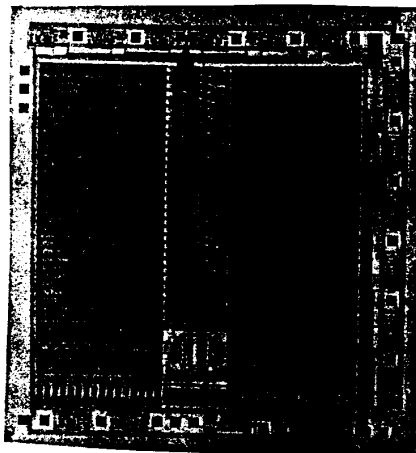


Fig. 7.

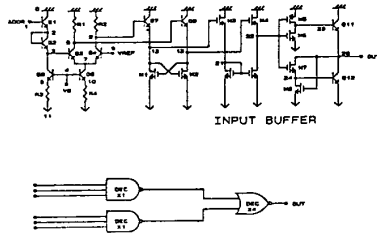
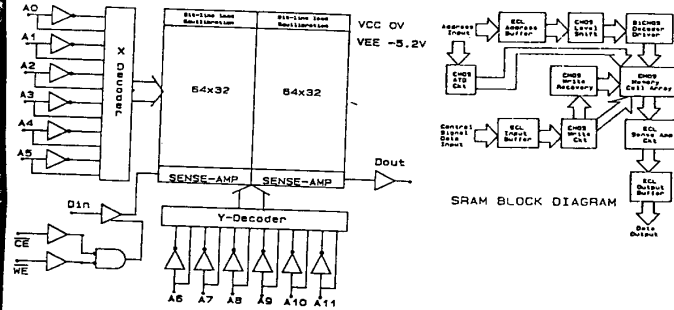


Fig. 8.

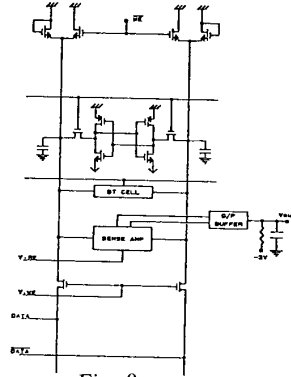


Fig. 9.

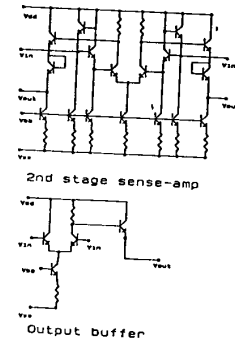


Fig. 10.

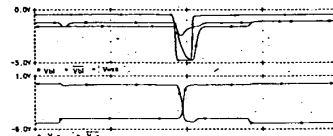
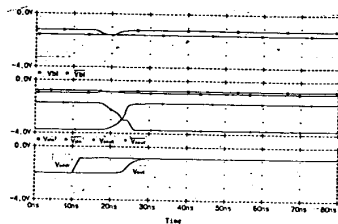


Fig. 11.

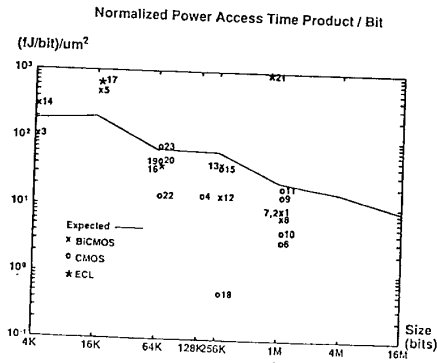
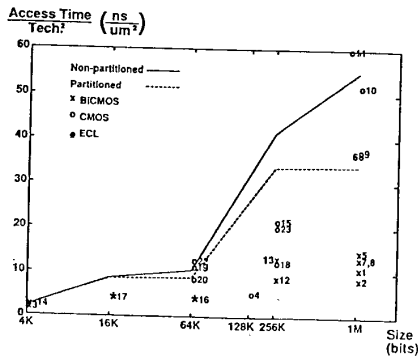


Fig. 12.