Low-Power Sequential Access Memory Design

Joong-Seok Moon[1], William C. Athas[2], Peter A. Beerel[1], Jeffrey T. Draper[3]

[1]Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089
[2]Apple Computer Inc., Cuppertino, CA 95014
[3]Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292

Abstract

This paper presents the design and evaluation of a sequential access memory (SAM) that provides low power and high performance by replacing address decoders with special locally-communicating sequencers. A test chip containing one 16x16-b SAM and one 64x16-b SAM (consisting of four 16x16-b banks) has been designed, fabricated, and evaluated using a 0.25-µm CMOS process. With a clock frequency of 40MHz at 1.2V, the measured worst-case read power dissipations for the 16x16-b SAM and the 64x16-b SAM are 344µW and 358µW respectively, demonstrating power dissipation that is mostly independent of SAM size.

I. Introduction

Low power dissipation in portable battery-operated platforms has drawn considerable interest in order to increase available functionality and/or battery lifetimes. In many of these applications a large fraction of the power is consumed in memory accesses [1].

Numerous general-purpose low power and high performance SRAMs have been proposed, mostly for high-speed designs [2-3]. Self-resetting circuits triggered with matched delay lines implemented with replica (dummy) memory cells [4-6] often yield relatively low-power and high-speed by limiting the power dissipation in the bitlines and reducing precharge time. A few designs have been proposed for low-power memories with special emphasis on reduction of leakage current for low-threshold voltage devices [7-8].

In many DSP applications SRAM designs do not require random access and often have strictly sequential read and/or write access patterns. In particular, programmable FIR filters read/write coefficients and data in a first-in first-out pattern [9]. The naïve implementation of such structures involves the movement of data samples each clock cycle using shift registers. However, for low-power implementations, the memory can be configured as a circular buffer (with a sequential access pattern) in which pointers rather than data are moved [10]. In addition, for many digital communication channel decoders, interleavers used to store and re-organize blocks of data samples can be organized with memories that support random write and sequential reads (or vice versa). Moreover, sequential access of intermediate data within many channel decoders, including Fano decoders [11] and turbo decoders [12], is also typical. In all these cases, the naïve implementation involves using SRAMs despite the fact that the architecture often accesses data sequentially.

This motivates the design of sequential access memories to eliminate the power dissipation for address decoding.

This paper presents a novel sequential access memory (SAM) design where address sequencing logic and decoders are replaced with row sequencers to achieve high speed and low power. Most of the control signals are generated using efficient *sequencer* cells that communicate primarily with neighboring rows only, minimizing the power dissipation of wordline selection. When combined with typical bank structures that limit the amount of switched bit-line capacitance of large memories and efficient self-resetting postcharge logic, power dissipation is largely independent of memory size. This is in sharp contrast to conventional SRAM designs.

A test chip was fabricated in 0.25-µm CMOS technology to evaluate this design. The chip contains two different dual-port (one read port and one write port) SAM configurations: one16x16-b and one 64x16-b, consisting of four 16x16-b banks. The chip has been tested and is fully functional at operating voltages of 0.67V to 2.5V. The power dissipation of both SAMs was measured at different voltages and operating frequencies and found to be within 5% of each other, demonstrating that power dissipation is largely independent of memory size. With a clock frequency of 40MHz at 1.2V, the measured worst-case read power dissipation for the 16x16-b SAM is 344µW and for the 64x16-b SAM is 358µW.

The remainder of this paper is organized as follows. Section II describes the architecture of the proposed SAM followed by a detailed description of the circuit design in Section III. Section IV presents the simulation and experimental results followed by a brief summary and conclusion in Section V.

II. Architecture

Fig. 1 shows a block diagram of our proposed dual-port NxM-bit SAM allowing simultaneous read and write accesses. Two sequencers, one for read accesses and one for write accesses, are shown as well as controllers and I/O circuitry. Two reset sequencer signals (*RdRst, WrRst*) are asserted to independently initialize the read and write sequencers to point to the first row. To enable the read operation, the read enable signal (*RdEn*) must be asserted. Then, upon the rising edge of the clock signal, the read sequencer enable signal (*rseq_en)* is raised which triggers the sequencer cell associated with the current pointer location to assert its associated wordline signal (*rwl[k]*).

This activates the associated memory cells and dummy cell, driving the read bitlines (*rbit[M-1:0]*) and the dummy bitline (*dumbit*), respectively. The sequencer cell also asserts a trigger signal (*trig[k+1]*) which is combined with the *rseq_en* signal to activate the next sequencer, moving the current pointer location to the next row. The reset of the current sequencer is triggered by the assertion of the next wordline. The write operation is similar except the write bitlines (*wbit[M-1:0]*) are not precharged.
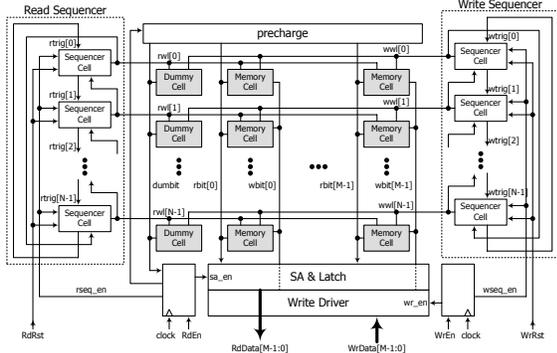
Fig. 1: A block diagram of an *N* x *M* SAM

For large *N*, driving long bitlines leads to increased power dissipation and latency. A memory bank structure can be easily applied to this SAM structure because most of the control signals are locally generated. Fig. 2 shows a block diagram of a 64x16-b SAM composed of four 16x16-b SAM banks.
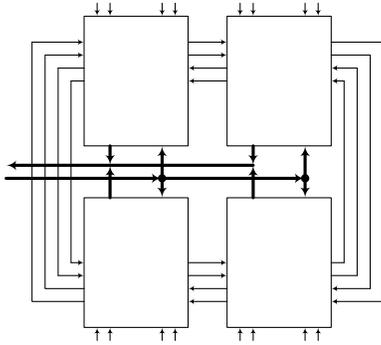

Fig. 2: A block diagram of a 64x16-b SAM composed of four 16x16-b SAM banks

The banks are daisy-chained so that the current bank generates a trigger signal to enable the next bank when a sequencer pointer has reached the last row in the current bank. In the subsequent cycle, the first wordline of the next bank is fed back to reset the trigger signal asserted in the previous cycle. Tri-state buffers are used for I/O circuitry of each bank so that only one of the banks is connected to the common input/output buses.

III. Circuits

A. Sequencer Cell

Fig. 3 and Fig. 4 show a schematic of the sequencer cell and its timing diagram. Initially, for the current sequencer cell

trig[k] is high while all other trigger signals are low. In addition, both *triggen* and *wl[k]* are low. Upon the assertion of the sequencer enable (*seq_en*), the wordline signal (*wl[k]*) is asserted via a dynamic AND gate. When *seq_en* is de-asserted, *wl[k]* is de-asserted and a short pulse (*triggen*) is generated by a NOR gate to assert the subsequent trigger signal (*trig[k+1]*), using a jam-latch (pulse-to-level converter). Notice that *trig[k+1]* is asserted approximately three gate delays after *seq_en* goes low to avoid two wordline signals being activated simultaneously. The *trig[k+1]* signal is reset by the assertion of the wordline signal *wl[k+1]* at the next read cycle.

To reset the SAM, the *trig[0]* signal should be the only asserted trigger signal. Thus, all sequencer cells except the last sequencer cell, should have a reset NMOS transistor controlled by the *Rst* signal attached to the jam-latch as shown in Fig. 3. In contrast, the last sequencer should have the reset transistor attached to the same side of the jam-latch as *trigen* to assert *trig[0]*.
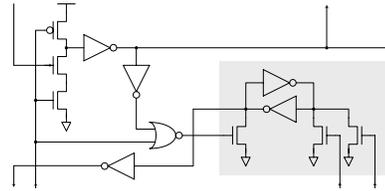

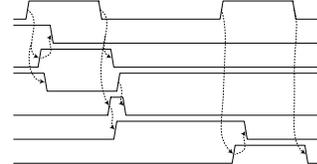Fig. 3: A schematic of the sequencer cell

Fig. 4: Timing diagram of the sequencer cell

B. Bank Sequencer

In conventional banked memory designs, the current memory bank is enabled directly by the address decoder. For the proposed SAM design, however, the active memory bank should notify the next memory bank as soon as read and/or write operations are completed in the current bank. A special bank sequencer, shown in Fig. 5, replaces the read/write controllers to achieve this goal.
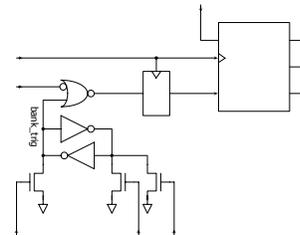

Fig. 5: A schematic of the bank sequencer

It is triggered by the last trigger signal (*trig_start*) of the previous bank and reset by the first wordline signal (*trig_end*) of the next bank. The bank trigger signal (*bank_trig*) is combined with the global read or write enable signal (*RdEn/WrEn*) to generate the bank enable signal (*bank_en*). The timing diagram of the bank sequencing circuit is depicted in Fig. 6. Note that, similar to the sequencer cell, the location of the *RdRst (WrRst)* transistor for the last bank is changed to the same side of the jam-latch as the *trig_start* signal.
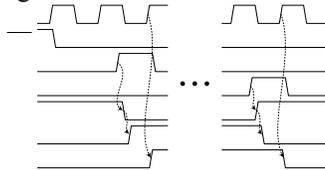

Fig. 6: Timing diagram of the bank sequencer

C. Read and Write Controllers

Three control signals - *precharge*, *sa_en*, *rseq_en* - are generated by the read controller. Self-resetting postcharge bit-lines [4-5] are used to limit the power dissipation and reduce precharge time. This, in turn, improves voltage scalability by enabling the use lower supply voltages while still meeting desired access times. Notice that the controller has no dependency on the falling edge of the clock signal so that read operations can be completed any time within the clock cycle. In particular, a larger portion of the clock cycle time can be used for the read operation plus a significant amount of subsequent combinational logic. Notice that like some other designs [4-6] we used dummy memory cells to simulate the bit-line discharge timing to simplify the control signal generation. Fig. 7 and Fig. 8 show the circuit and its corresponding timing diagram.
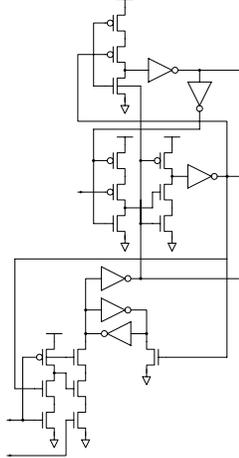

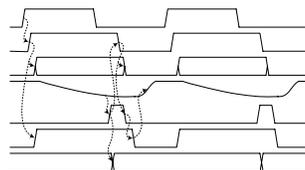Fig. 7: A schematic of read controller

Fig. 8: Timing diagram of read controller

For a write operation, it is required to enable the current sequencer cell and to drive bitlines, which can be done by one control signal (*wseq_en*). Write operations are processed during the high phase of the clock. Therefore, a dynamic flip-flop (output resets to zero at the falling edge of the clock) with buffers is used for the write controller.

D. Overall Read Operation Summary

Fig. 9 summarizes the signal transitions for a read operation. Each transition is identified and numbered in increasing order of operation. Transitions with the same number are concurrent.
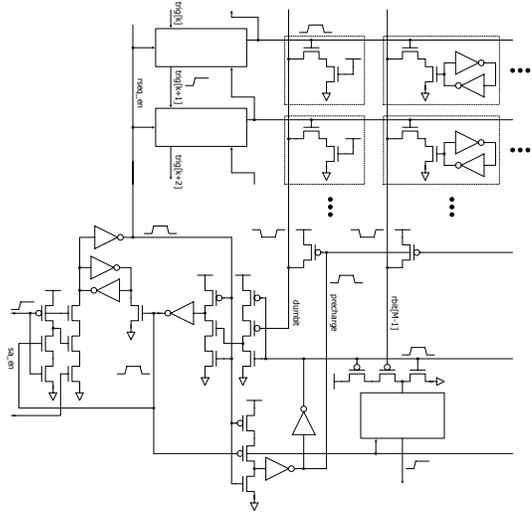

Fig. 9: Signal transitions of read operation

III. Test Chip and Measurement Results

A test chip was fabricated in the TSMC 0.25-µm n-well CMOS process offered through MOSIS. A microphotograph of the test chip is shown in Fig. 10. Three metal layers are used for the memory core while all five metal layers are used for I/O pads. The 16x16-b and 64x16-b SAMs occupy an area of 197.3µm x 131.90µm and 366.2µm x 265.4µm, respectively. Table I summarizes the characteristics of the process technology and test chip.
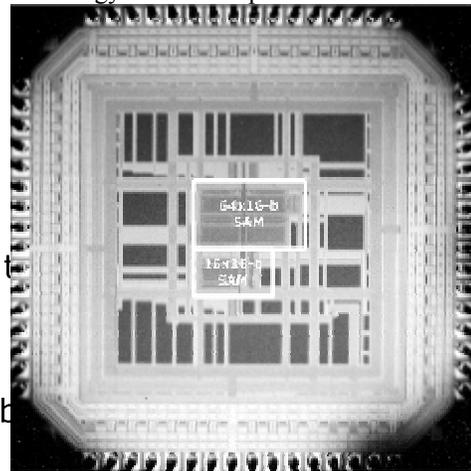

Fig. 10: A microphotograph of the test chip

Table I: Summary of the process technology and the test chip

Technology	0.25-μm n-well CMOS Logic Process (Five metal layers, single poly)
V_{th}	0.51V (NMOS), -0.52V (PMOS)
$f_{ring\ oscillator}$	273MHz (31 stages)
FO4 delay	174ps
Number of transistors	24361
Operating voltage	0.67V-2.5V (core), 1.5V-3.3V (I/O PAD)
Die size	4.26 mm^2

The minimum operating core supply voltage was measured to be 0.67V with a corresponding maximum frequency of 34MHz. Fig. 11 and Fig. 12 show the measured worst-case power dissipation of two SAMs (16x16-b and 64x16-b) for read and write operations for a variety of supply voltages and frequencies. Note that due to the limitations of available test equipments, testing at frequencies higher than 40MHz was not possible. However, HSPICE simulation shows that the worst-case delay is 5.3ns with a 1.2V supply voltage and 2.1ns with a 2.5V supply voltage.
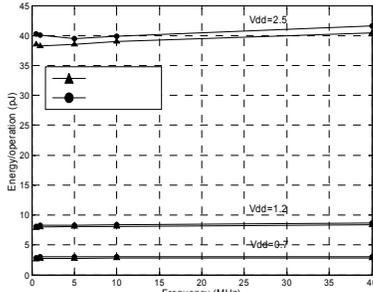

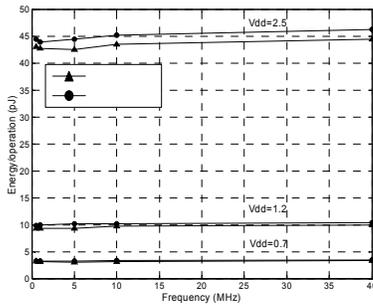Fig. 11: Measured power dissipation of the worst-case read operations

Fig. 12: Measured power dissipation of the worst-case write operations

The measured power dissipation for the 64x16-b SAM read operation is 358μW (8.95pJ*40MHz) at 40MHz and 1.2V and 344μW for the 16x16-b SAM. Power dissipation for write operations is higher than read operations (416μW for the 64x16-b SAM, 396μW for the 16x16-b SAM). The average power dissipation was also measured using random vectors with simultaneous read and write operations. For this case, the power dissipation for the 64x16-b SAM is 467μW at 40MHz and 1.2V and 446μW for the 16x16-b SAM. The differences in power dissipation between these two SAMs are less than 5% for all conditions. The independence of energy per operation with respect to frequency in the above graphs suggests that there is neglibible static current in the proposed design. For both of the operations, additional test

measurements suggest that approximately one third of the power dissipation is consumed by the input and output flip-flops connected to I/O pads. The standby current was negligible (less than 0.1μA) for all operating voltages and frequencies.

III. Conclusion

A novel low-power sequential access memory has been designed, fabricated, and tested using a 0.25-μm CMOS technology. In particular, our design demonstrates the unique feature of having power dissipation that is largely independent of memory size. The proposed design demonstrates the low-power potential of sequential access memories for applications requiring non-random access patterns. Future work can include the development of hybrid designs that support both sequential and random accesses to increase the flexibility of the memory access.

Acknowledgement

References

[1] W. Shiue, C. Chakrabarti, "Memory Design and Exploration for Low Power, Embedded Systems", in Proc. 36th ACM/IEEE Design Automation Conference, pp. 140-145, Jun. 1999

[2] W. Hwang, R. Joshi, and W. Henkels, "A 500-MHz, 32-Word x 64-Bit, Eight-Port Self-Resetting CMOS Register File", IEEE J. Solid-State Circuits, vol. 34, pp. 56-67, Jan. 1999

[3] R. Krishnamurthy, A. Alvandpour, G. Balamurugan, N. Shanbhag, K. Soumyanath, and S. Borkar, "A 0.13um 6GHz 256x32b Leakage-tolerant Register File", in Symp. VLSI Circuits Dig., pp. 25-26, Jun. 2001

[4] M. Golden and H. Partovi, "A 500 MHz, Write-bypasswd, 88-entry, 90-bit Register File", in Symp. VLSI Circuits Dig., pp. 105-108, Jun. 1999

[5] B. Amrutur, M. Horowitz, "A Replica Technique for Wordline and Sense Control in Low-power SRAMs", IEEE J. Solid-State Circuits, vol. 33, pp. 1208-1219, Aug. 1998

[6] K. Osada, J. L. Shin, M. Khan, Y. Liou, K. Wang, K. Shoji, K. Kuroda, S. Ikeda, K. Ishibashi, "Universal-V_{dd} 0.65-2.0-V 32-kB Cache Using a Voltage-Adapted Timing-Generation Scheme and a Lithographically Symmetical Cell", IEEE J. Solid-State Circuits, vol. 36, pp. 1738-1744, Nov. 2001

[7] H. Kawaguchi, Y. Itaka, T. Sakurai, "Dynamic Leakage Cut-off Scheme for Low-Voltage SRAM's", in Symp. VLSI Circuits Dig., pp. 140-141, Jun. 1998

[8] K. Itoh, A. Fridi, A. Bellaouar, M. Elmasry, "A Deep Sub-V, Single Power-Supply SRAM Cell with Multi-V_T, Boosted Storage Node and Dynamic Load", in Symp. VLSI Circuits Dig., pp. 132-133, Jun. 1996

[9] M. Mehendale, S. Sherlekar, G. Venkatesh, "Low-Power Realization of FIR Filters on Programmable DSP's", IEEE Tran. VLSI Systems, vol. 6, pp. 546-553, Dec. 1998

[10] TMS320C2x/C5x User's Guide, Texas Instruments, 1993

[11] S. Singh, P. Thiennviboon, R. Ozdag, S. Tugsinavisut, P. Beerel, K. Chugg, "Algorithm and Circuit Co-Design for a Low-Power Sequential Decoder," in Proc. 33rd Asilomar Conference on Signals, Systems, and Computers, vol. 1, pp. 389 -394, Nov. 1999

[12] G. Masera, G. Piccinini, M. Roch, M. Zamboni, "VLSI Architecture for Turbo Codes", IEEE Tran. VLSI Systems, vol. 7, pp. 369-379, Sep. 1999