Energy Recovering Static Memory

Joohee Kim   Conrad H. Ziesler   Marios C. Papaefthymiou

Advanced Computer Architecture Laboratory
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109
{jooheek, cziesler, marios}@eecs.umich.edu

ABSTRACT

This paper proposes an energy-recovering (a.k.a. adiabatic) static RAM with a novel driver that reduces power dissipation by efficiently recovering energy from the bit/word line capacitors. Powered by a single-phase sinusoidal power-clock, our SRAM delivers read and write operations with single-cycle latency. To that end, a precharge-low scheme is employed along with a modified sense amplifier design that achieves high efficiency at differential voltages near V_{SS}. A simple control circuit is used to maintain driver operation in synchrony with the power-clock waveform. Feedback circuitry from the driver output to the control circuit ensures that our driver remains efficient, independent of the access pattern.

Our energy recovering SRAM functions correctly while achieving substantial energy savings over a wide range of supply voltages and operating frequencies. Hspice simulations of a simple full-custom adiabatic 256x256 SRAM, that includes the energy recovering bit/word line drivers, the cell array, and the sense amplifiers, show over 2.6x energy savings at 3V, 300MHz in comparison with its conventional counterpart.

Categories and Subject Descriptors

B.3.1 [**Memory Structures**]: Semiconductor Memories— *Static memory (SRAM)*

General Terms

Design, Performance

Keywords

Adiabatic circuitry, charge recovery, cache memories, on-chip memories, low-energy design, low-power computing.

1. INTRODUCTION

Static RAMs are used extensively in modern processors as on-chip memories due to their large storage density and small access latency. Low power on-chip memories have become the topic of

substantial research as they can account for almost half of total CPU dissipation, even for extremely power-efficient designs [1].

Energy recovery is a particularly attractive approach to the reduction of power dissipation in high-density memories with large switching capacitance. Energy recovery schemes reduce energy dissipation by limiting voltage differences across conducting devices and by recovering charges from the load capacitors. This controlled mode of operation is typically accomplished through the coordination of time-varying voltage waveforms, called power-clocks. Previous energy recovery approaches for static memories achieved considerable energy savings over conventional SRAMs [2, 3, 4, 5, 6, 7, 8]. These schemes required multiple-phase power-clocks, however, and experienced a variety of drawbacks, including relatively low operating frequencies, long latencies, non-trivial area overheads, and access-pattern dependent energy savings.

In this paper, we propose a novel energy recovering static RAM that achieves substantial energy savings. With its fast operation and low overhead, our memory is suitable for on-chip caches. In particular, our SRAM provides single-cycle latency read and write operations, if decoding is pipelined, while avoiding the shortcomings of previous energy recovering approaches. It is powered by a single-phase sinusoidal power-clock to minimize power-clock generator dissipation, coupling noise, and area overhead.

The main feature of our SRAM is an energy recovering driver that reclaims energy from the capacitors of the bit/word lines. A small control circuit embedded to the driver keeps its operation in synchrony with the power-clock. Through the use of feedback from the driver output to the control circuit, the operation of our driver remains efficient, independent of the operation sequence. To provide single-cycle read with a single-phase power-clock, a precharge-low scheme is employed in conjunction with a current-mode sense amplifier that is modified to operate efficiently near V_{SS}. With the exception of the energy recovering drivers and the modified sense amplifiers, the structure of our static memory is identical to that of conventional SRAMs.

In Hspice simulations of a 256x256 SRAM in 0.35μm TSMC process, our energy recovering memory achieves energy savings in excess of 2.6x in comparison with a conventional counterpart at 3V, 300MHz. Our SRAM functions correctly over a wide range of supply voltages and operating frequencies. Maximum operating frequency ranges from 1MHz at 0.7V to over 500MHz at 2.75V.

The remainder of this paper has four sections. Section 2 describes the energy recovering driver of our SRAM. The architecture and operation of our energy recovering SRAM is explained in Section 3. Hspice simulation results are presented in Section 4. Section 5 concludes with a summary of our contribution and ongoing research.
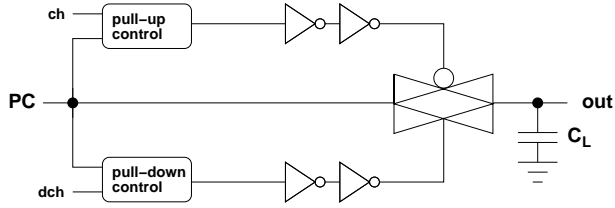
2. ENERGY RECOVERING DRIVER

Figure 1: Core structure of energy recovering driver.

Figure 1 shows the core organization of our energy recovering driver. The driver is composed of a transmission gate that drives the load in an energy recovering manner and control circuitry that synchronizes driver operation with a single-phase sinusoidal power-clock PC.

The transmission gate enables the correct operation of our energy recovering driver over a wide range of supply voltages and operating frequencies. The buffers driving the large PMOS device are carefully ratioed to minimize the potentially large power dissipation associated with driving heavy loads. An alternative to the transmission gate would have been bootstrapped NMOS, as in [3]. This approach would have yielded lower dissipation than the transmission gate by enabling the use of smaller buffers. The resulting driver would operate correctly only over a narrow range of supply voltages and operating frequencies, however, since the boot capacitance, on which the correct operation of the bootstrapped NMOS depends, varies with time and supply voltage [9].
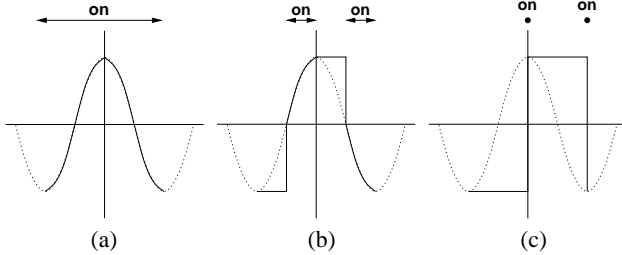

Figure 2: Various output waveforms of energy recovering driver with (a) full gradual transition, (b) partial gradual transition, and (c) abrupt transition. Power-clocks are denoted by dotted lines and driver outputs are denoted by solid lines.

The energy efficiency and the maximum operating frequency of an energy recovering driver such as ours depend on the timing of its operation with respect to the power-clock. Accordingly, energy efficiency can be traded off for speed by appropriately controlling the on-time of the driver. The driver output in Figure 2(a) results when the driver is turned on for the entire power-clock period, allowing the output to track the gradually changing power-clock waveform. This approach is used in most of previous energy recovering drivers and yields high energy efficiency, since driver output changes smoothly throughout. It inevitably results in the lowest operating frequency, however, due to the short stay of the driver output at the peak value.

The result of turning the driver on only at the peaks (positive or negative) of the power-clock waveform is shown in Figure 2(c). This mode of operation yields maximum speed, since the driver output stays at its peak value for about a half clock cycle. It also results in the lowest energy efficiency, however, due to the abruptness of the output transitions.

Our driver operates at an intermediate condition with partial gradual transitions, shown in 2(b), resulting in both high efficiency and high speed. Another advantage of our partial approach over full gradual transition is that the driver output does not need to be pulled down to V_{SS} after each operation. Consequently, unlike other energy recovering drivers, our driver does not dissipate energy during consecutive operations of the same kind. as evidenced by our simulation results in Section 4.
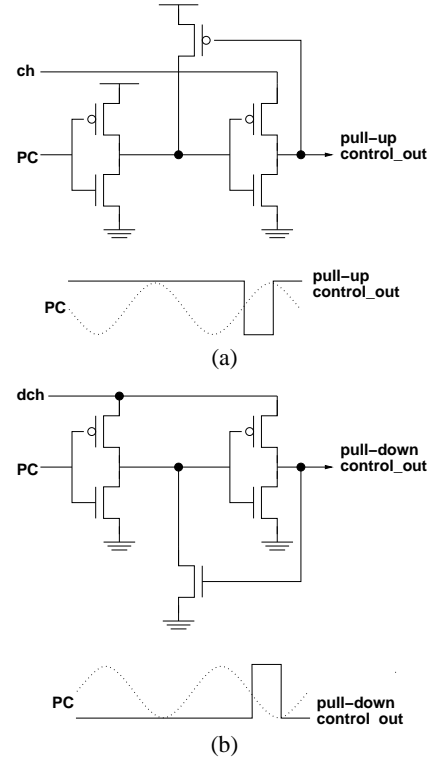

Figure 3: Synchronizing circuitry and associated waveforms for the driver's (a) pull up and (b) pull down control.

Figure 3 shows the schematics and waveforms of the control circuitry for the PMOS and NMOS devices of the transmission gate. The control circuitry resembles a Schmitt trigger and returns sharp transitions from slow power-clock transitions. The transition point with respect to the power-clock and the pulse width can be controlled by ratioing the transistors of the first inverter and the stand alone PMOS (for pull-up control) or NMOS (for pull-down control) in a way similar to the one shown for CMOS Schmitt triggers in [10]. The signals ch and dch selectively enable the control circuitry, minimizing idle power dissipation. As evidenced by our simulations in Section 4, the simple structure of our driver ensures correct operation of our SRAM for a broad range of supply voltages and operating frequencies.

As the supply voltage changes, the delay through different paths changes non-linearly causing variance in timing of the control signals. Since the PMOS and NMOS control signals must stay in synchrony with the power-clock despite changes in the supply voltage, the control circuit must be tolerant to the variance in the timing of the driver control signals. Figure 4 shows the timing of the driver control signals ch and dch for correct driver operation. Significant variance in the timing of the control signals can be tolerated, making it possible for the driver to operate correctly over a wide range of supply voltages.
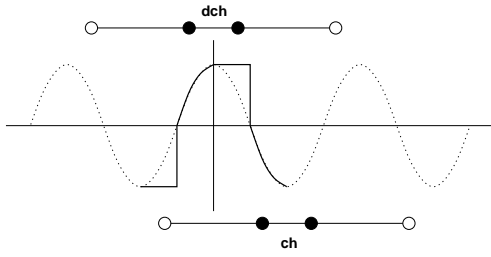
Figure 4: Timing margin for control signals ch **and** dch**. For correct driver operation, these signals must cover the intervals between the closed circles. To prevent faulty operation, they should not be asserted beyond the open circles.**
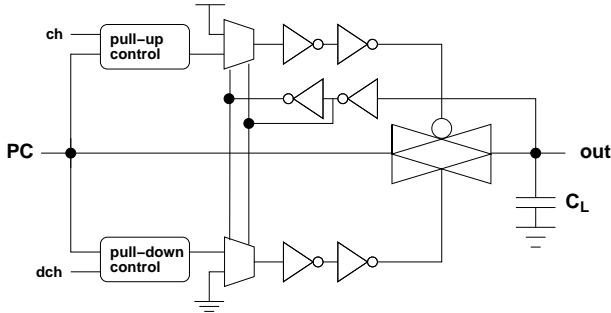

Figure 5: Complete structure of energy recovering driver with feedback.

The driver core in Figure 1 dissipates unnecessarily when the driver performs the same function in consecutive cycles. In such cases, the internal nodes and buffer stages dissipate energy although the driver output stays the same. The dissipation of internal nodes can be minimized by adding feedback circuitry that prevents their unnecessary switching. Figure 5 shows our complete energy recovering driver with the feedback path. Two multiplexors selectively pass the pull-up and pull-down control signals to the driver and hold the transmission gate off if the operation is not necessary. Due to this feedback, the energy efficiency of our driver does not suffer from the operation sequence dependency that previous energy recovering memories have experienced [3].

Figure 6 shows the operation of our energy recovering driver compared to that from [3], which uses a two-phase power-clock, at 3V, 300MHz with a 1pF load. The output of our driver stays at peak value for longer periods of time, making it suitable for high-speed SRAM applications. Moreover, our driver does not switch its load unnecessarily during consecutive operations of the same kind, since its output does not need to be pulled down after each operation. Hence, for the first two consecutive charge operations, our driver stops dissipating once the output reaches V_{DD}. The other driver, however, dissipates energy on every cycle.

3. SRAM ARCHITECTURE AND OPERA-TION

This section describes the architecture, timing requirements, and operation modes of our energy recovering SRAM. As can be seen in Figure 7, our energy recovering SRAM uses general 6T SRAM cells and has similar architecture to a conventional SRAM. Except for the drivers and the sense amplifiers, all the components of our adiabatic SRAM are the same as their conventional counterparts.
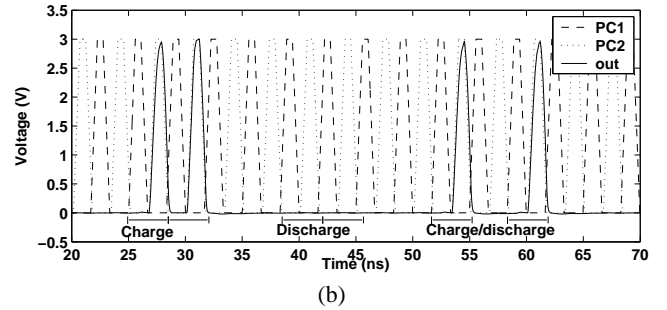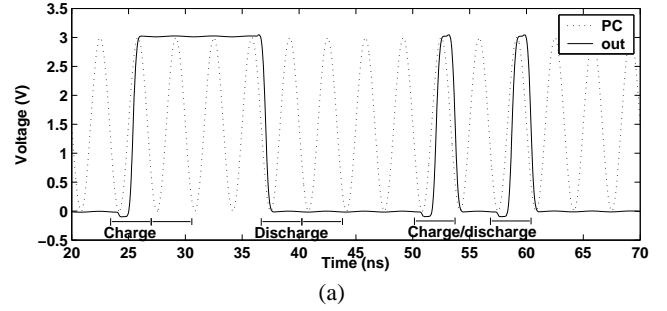

(a)

(b)

Figure 6: Operation of energy recovering drivers for 2 cycles of charging, discharging, and back-to-back charging and discharging, with idle cycles in-between. (a) Our driver. (b) Driver from [3].
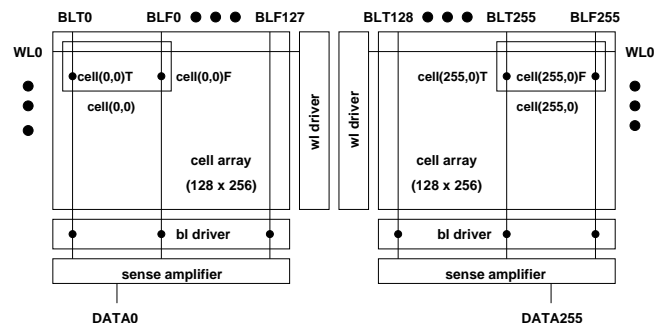

Figure 7: Architecture of a 256x256 energy recovering SRAM.

Our modifications ensure that the word lines and bit lines are powered only by the power-clock to minimize energy dissipation.

Correct timing of the bit/word lines is especially important in energy recovering SRAMs, since the power-clock can provide peak voltages only at the peaks of its waveform. Hence, charge and discharge of the load capacitor cannot occur at the same time. The timing requirements and necessary modes of operation enabling single-cycle operations with a single-phase power-clock are described in the following paragraphs.
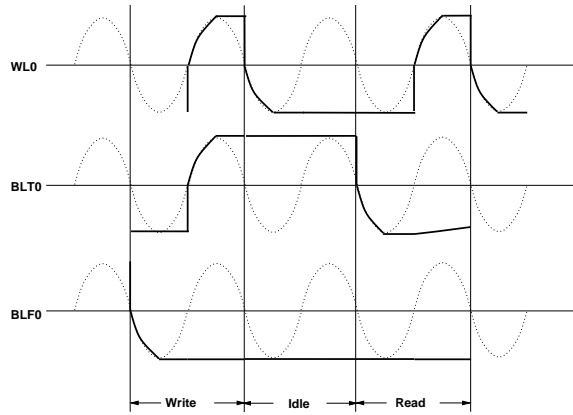

Figure 8: Waveforms of bit/word lines for writing and reading a "1". An idle cycle is added in between for clarity.

Figure 8 shows the timing necessary for enabling single-cycle latency operations with a single-phase power-clock. Write operations occur in a manner similar to that of conventional SRAMs. First, the bit line BLF0 storing "0" is pulled down. Then, both the word line WL0 and the bit line BLT0 storing "1" are pulled up, storing data into the cell.

For every memory access, only one selected word line is pulled up, and all other word lines are pulled down. In a conventional SRAM, pulling down the unselected word lines is not dissipative, since V_{SS} level is always available. In our SRAM, however, this operation dissipates power. Since the pull down starts when the power-clock is above V_{SS}, the word lines are actually pulled up above V_{SS} and then pulled down. Hence, in our energy recovering SRAM, the selected word line needs to be pulled down explicitly after each access.

Read operations are different from conventional SRAM. Since precharge must precede the assertion of the word line, all bit line pairs must be precharged low for the read to occur in a single cycle. After precharge, the word line is charged, and the cell nodes cause a voltage difference between each pair of bit lines. Precharging low is more energy-efficient in our energy recovering SRAM than precharging high, since the charge pumped from the cell to the bit line can be recovered through the bit line driver. This precharge-low scheme necessitates the modification of conventional sense amplifiers to make them more sensitive to the voltage difference near V_{SS} as opposed to V_{DD}. Precharge-low has already been proposed for low power designs resulting in the modification of other components such as the memory cells [11] in conventional SRAM. A previously reported adiabatic SRAM with full gradual transition driver used precharge-low and a current-mode sense amplifier to reduce dissipation during consecutive read and write operations [8].

To enable the operation of our SRAM with precharge-low, we modified a previously reported high-speed sense amplifier [12]. Our sense amplifier circuitry, shown in Figure 9, is a 2-stage de-
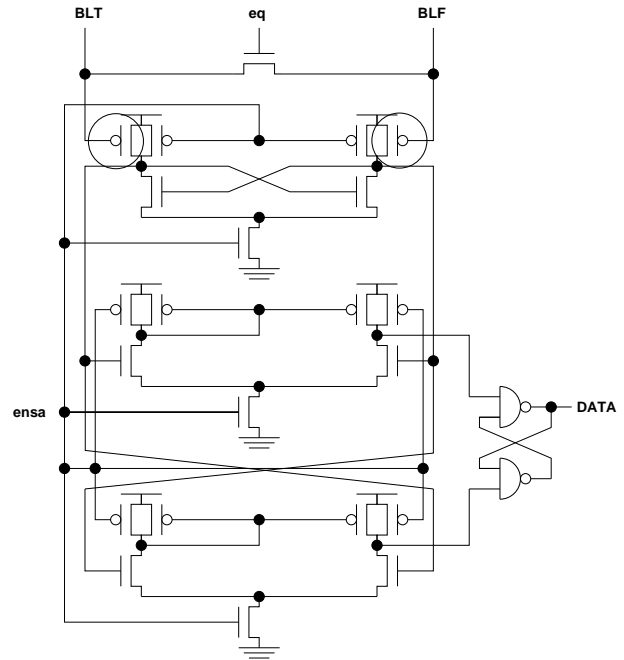

Figure 9: Structure of modified sense amplifier.

sign composed of general cross-coupled and current-mirror sense amplifiers. The only difference of our modified version from the original one is the input PMOS pair in circles, which is NMOS in [12]. We did not select the latch type sense amplifier with two cross-coupled inverters, since it drives the bit line pair during amplification in a non-energy recovering manner.

The varying load capacitance between different operations in our SRAM causes load dependent voltage and timing variations. For consecutive operations, word lines cause insignificant variations, regardless of the operation, since word line capacitances are designed equal. Bit lines cause more significant variations between write and read operations, however, due to the discharging of a bit line in each pair during precharge-low. These voltage and timing variations can be reduced by adding enough redundant capacitance to the power-clock generator. A resonant single-phase power-clock generator such as the one in [13] can be used with our SRAM.

4. SIMULATION RESULTS

We have designed an energy recovering 256x256 SRAM using MOSIS SCMOS4M SUBM design rules for the 0.35μm TSMC process. For comparison, we designed a conventional SRAM with the same components, except for the drivers and the sense amplifiers. Figure 10 shows the structure of the simple and low-power conventional bit/word line drivers we implemented for comparison. The sense amplifier we used in the conventional SRAM was borrowed from [12].

Hspice simulations were performed to study the energy efficiency of our SRAM. To reduce simulation time, actual simulations were done on a subset of the SRAM that includes a 2x2 cell array, 2 word line drivers, 4 bit line drivers, and 2 sense amplifiers. Lumped capacitors, measured from driving a 128-cell row and a 256-cell column, were added to word lines and bit lines to match the bit/word line capacitance of a 256x256 SRAM. The energy dissipation of each module was measured, and the energy dissipation of the entire 256x256 SRAM array was calculated.
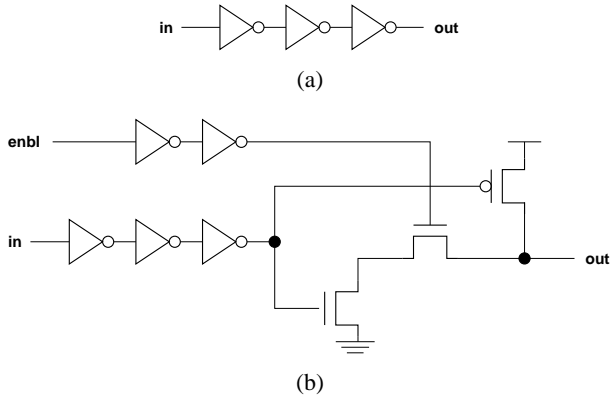
Figure 10: Structure of conventional (a) word line driver and (b) bit line driver.
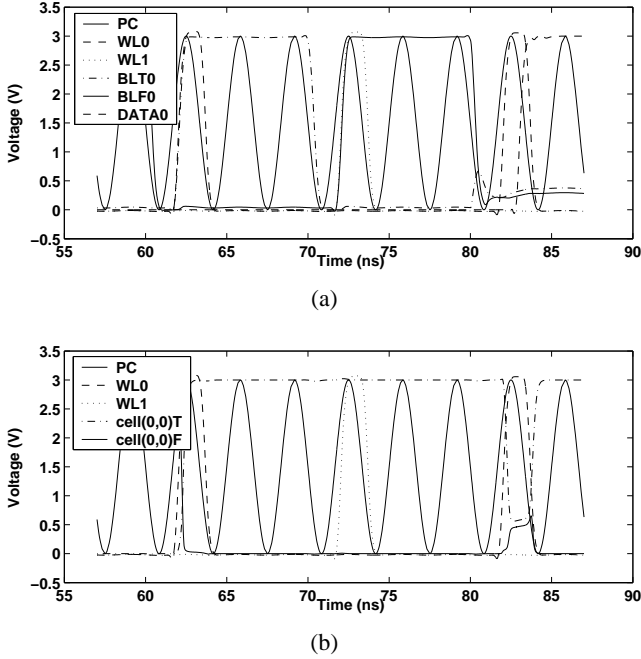

(a)

(b)

Figure 11: Hspice waveforms of energy recovering SRAM for 2 consecutive writes and a read. Idle cycles are inserted in between for clarity. (a) Bit/word lines and data output. (b) Cell's internal nodes.

Figure 11(a) shows the Hspice waveforms of the bit/word lines and the data output of the energy recovering SRAM for the following operations at 3V, 300MHz: Write "1" to cell (0,0), write "0" to cell (0,1), and read cell (0,0). Figure 11(b) shows the waveforms of the cell's internal nodes during read. As can be seen, the read operation with precharge-low is nondestructive. Initially, as the access transistors are turned on, the cell node at high is pulled down below V_{DD}. However, it recovers to high as the access transistors are turned off, since the cell node at low is clamped by the bit line.

The energy breakdown for each component of the conventional SRAM and our SRAM running worst-case writes and reads at 3V, 300MHz is given in Table 1 and Table 2, respectively. In Table 2,

Table 1: Energy breakdown of conventional SRAM (pJ/cycle at 3V, 300MHz).

	Write	Read	Mean
Wl drivers	27	27	27
Cell array	32	0.64	16.3
Bl drivers	4950	5225	5088
Sense amps	-	525	262.5
Total	5009	5778	5394

Table 2: Energy breakdown of energy recovering SRAM (pJ/cycle at 3V, 300MHz).

	Write	Read	Mean	E_{conv}/E_{er}
Wl drivers	19.6	19.6	19.6	1.38
Cell array	41.5	214.5	128	0.13
Bl drivers	3050	64	1557	3.27
Sense amps	-	650	325	0.81
Total	3111	948	2030	2.66

the relative energy savings E_{conv}/E_{er} are also shown. The cell array of our SRAM dissipates more energy during reads, because it charges the precharged-low bit lines. This charge is recovered through the bit line drivers, however, thus reducing bit line driver dissipation. Our sense amplifiers are more dissipative as a result of their modification. However, our energy recovering bit/word line drivers, which are responsible for the bulk of total energy dissipation, dissipate far less energy than their conventional counterparts. Overall energy savings in excess of 2.6x are achieved under these conditions, assuming an ideal power-clock generator with 100% energy recovery.
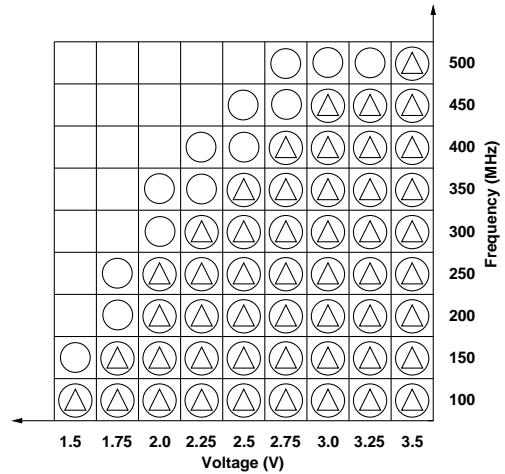

Figure 12: Schmoo plot of SRAM in 1.5V-3.5V range. Only conditions that result in correct operation are marked. Our SRAM is denoted by a circle. Conventional SRAM is denoted by a triangle.

We studied the operation of our SRAM when voltage and frequency are scaled. As can be seen in Figure 12, our SRAM functions correctly over a wide range of supply voltages and operating

frequencies and is thus suitable for general on-chip caches. Both SRAMs eventually fail due to incorrect timing of the sense amplifier enable signal `ensa` which is set fixed with respect to the clock. However, this speed limitation can be relaxed by implementing timing control circuitry [14].
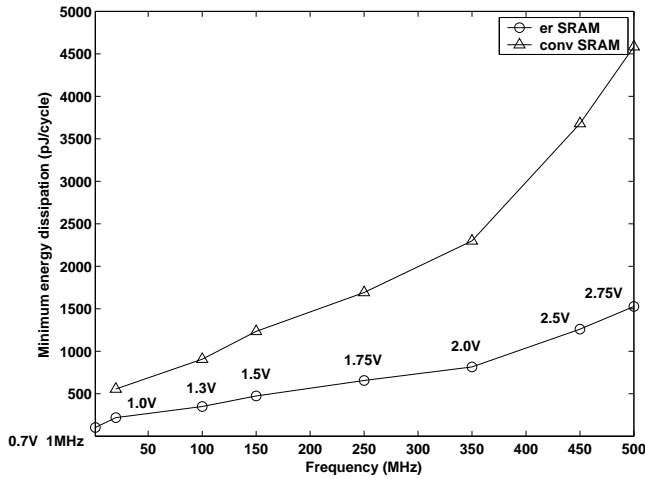

Figure 13: Minimum energy dissipation versus operating frequency for energy recovering SRAM and conventional SRAM.

Figure 13 compares the trends of minimum energy dissipation versus operating frequency between our SRAM and its conventional counterpart. Although the conventional SRAM failed at lower frequencies, extrapolated dissipation from closest working condition was plotted, since the failure is caused by the mistiming of sensing and is not inherent to its structure. Our results show that our SRAM achieves significant energy savings over a wide frequency range from 1MHz at 0.7V to 500MHz at 2.75V.

5. CONCLUSION

This paper describes a novel energy recovering SRAM with substantially lower power dissipation than conventional SRAM. Our SRAM runs on a single-phase power-clock and can deliver single-cycle latency write and read operations. A novel energy recovering driver reduces SRAM dissipation by efficiently recovering the energy from the capacitors of the bit/word lines. Simple synchronizing circuit enables correct and efficient operation of our driver without introducing additional control signals. Feedback from the driver output keeps the energy efficiency independent of the operation sequence. To enable single-cycle reads with a single-phase power-clock, a precharge-low approach is used. Accordingly, a modified sense amplifier has been implemented that operates efficiently at bit line voltages near V_{SS}. The structure of our energy recovering SRAM is very close to conventional SRAM, thus enabling the application of other low-power SRAM techniques to further reduce its power dissipation.

Simulations of 256x256 SRAMs show that our energy recovering memory achieves energy savings in excess of 2.6x at 3V, 300MHz. Moreover, our SRAM functions correctly over a wide range of supply voltages and operating frequencies. We are currently in the process of fabricating our energy recovering SRAM to validate the efficiency and robustness of our design through actual measurements.

6. REFERENCES

[1] J. Montanaro et al., "A 160-MHz, 32-b, 0.5-W CMOS RISC microprocessor," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 11, pp. 1703–1714, November 1996.

[2] D. Somasekhar, Y. Ye, and K. Roy, "An energy recovery static RAM memory core," in *IEEE Symposium on Low Power Electronics*. IEEE, 1995, pp. 62–63.

[3] N. Tzartzanis and W.C. Athas, "Energy recovery for the design of high-speed, low-power static RAMs," in *International Symposium on Low Power Electronics and Design*. IEEE, 1996, pp. 55–60.

[4] Y. Moon and D.K. Jeong, "A 32 x 32-b adiabatic register file with supply clock generator," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 5, pp. 696 –701, May 1998.

[5] S. Avery and M. Jabri, "A three-port adiabatic register file suitable for embedded applications," in *International Symposium on Low Power Electronics and Design*. IEEE, 1998, pp. 288–292.

[6] J.H. Kwon, J. Lim, and S.I. Chae, "A three-port nRERL register file for ultra-low-energy applications," in *International Symposium on Low Power Electronics and Design*. IEEE, 2000, pp. 161–166.

[7] K.W. Ng and K.T. Lau, "A novel adiabatic register file design," *Journal of Circuits, Systems, and Computers*, vol. 10, no. 1, pp. 67–76, 2000.

[8] N. Tzartzanis, W.C. Athas, and L. Svensson, "A low-power SRAM with resonantly powered data, address, word, and bit lines," in *European Solid-State Circuits Conference*, 2000.

[9] L.A. Glasser and D.W. Dobberpuhl, *The design and analysis of VLSI circuits*, Addison Wesley, 1988.

[10] J.M. Rabaey, *Digital integrated circuits*, Prentice Hall, 1996.

[11] A.J. Bhavnagarwala, A. Kapoor, and J.D. Meindl, "Source-pulsed dynamic-threshold CMOS SRAMs for fast, portable applications," in *European Solid State Circuits Conference*, 2000, pp. 183–186.

[12] H. Nambu et al, "A 1.8-ns access, 550-MHz, 4.5-Mb CMOS SRAM," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 11, pp. 1650–1658, November 1998.

[13] C.H. Ziesler, Suhwan Kim, and M.C. Papaefthymiou, "A resonant clock generator for single-phase adiabatic systems," in *International Symposium on Low Power Electronics and Design*. IEEE, 2001, pp. 159–164.

[14] B.S. Amrutur and M.A. Horowitz, "A replica technique for wordline and sense control in low-power SRAM's," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 8, pp. 1208–1219, August 1998.