4-1-1996

On The Design of Adiabatic SRAMs

Yibin Ye
Purdue University School of Electrical and Computer Engineering

Dinesh Somasekhar
Purdue University School of Electrical and Computer Engineering

Kaushik Roy
Purdue University School of Electrical and Computer Engineering

On The Design of Adiabatic SRAMs

Yibin Ye
Dinesh Somasekhar
Kaushik Roy

TR-ECE 96-4
April 1996

School of Electrical
and Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285

On The Design of Adiabatic SRAMs *

Yibin Ye, Dinesh Somasekhar, and Kaushik Roy

Electrical Engineering

Purdue University

West Lafayette

IN 47907-1285, USA.

Ph: 317-494-2361

Fax: 317-494-3371

e-mail: kaushik@ecn.purdue.edu

Abstract

In the design of low-power circuits, adiabatic logic shows great promise. However, research till date have concentrated on adiabatic logic circuits/families. Today's *VLSI* systems integrate random logic, megamodules and memories. Hence, the success of adiabatic circuits will depend on the efficient implementation of not only random logic, but also the other components of a *VLSI* system. In this paper, we present a design of adiabatic *Static* RAM, which can be implemerited without greatly increasing area or circuit complexity. The design addresses the issue of building ultra-low power memory circuits in a VLSI system. Our results for a *4Kb* block of memory core indicates energy savings of approximately 75% for both read and write operations. Higher power savings are achieved in the address decoder and *I/O* drivers.

*The research was supported in part by NSF and by ARPA. A preliminary version of the paper appeared in the 1995 Low-Power Electronics Symposium.

1 Iritroduction

With the recent trend toward portable communication and computing, power dissipation has become ane of the major design concerns along with area and performance. Research to reduce power dissipation at various levels of design abstraction has started in earnest to achieve ultra-low power with optimum performance [4]. In this paper we will consider circuit level techniques for low-power SRAM design using the principles of adiabatic switching.

Power dissipation in static CMOS circuits can be best understood by considering the inverter of figure 1(a). A logic ONE to ZERO transition on input x turns the p-mos transistor on and the output node which is associated with capacitance C is charged from 0 to V_{dd}. With such a transition, $V_{dd}Q(= CV_{dd}^2)$ of energy is extracted from the supply, half of which, $\frac{1}{2}CV_{dd}^2$, is stored in the capacitance temporarily, and the other half is dissipated in the path. When the input experiences a LOW to HIGH transition, $\frac{1}{2}CV_{dd}^2$ of energy is again dissipated. Hence, CV_{dd}^2 of energy is dissipated in an entire cycle. It should be observed that whenever current experiences a voltage drop AV, energy is dissipated at the rate of $i\Delta V$ (instantaneous dissipative power), where i is the current. Such energy dissipation can be greatly minimized by considering adiabatic switching. Let us consider the circuit of figure 1(b), where the supply voltage Φ swings gradually from 0 to V_{dd} (evaluation period), stays at V_{dd} for some time (hold period) and then swings back from V_{dd} to 0 (restoration). If the output y at the beginning of the evaluation period is at logic ZERO and input x is valid and is also equal to logic ZERO then the output node y would follow Φ to a logic ONE in a way such that there will be very little voltage drop across the channel of the p-mos transistor. Hence, only a small amount of energy is dissipated. For a more detailed analysis and design of adiabatic circuits, the reader is referred to [6]. An effective model to estimate the power dissipation in adiabatic circuits is given by [6]

$$E_{dissipation} \approx \frac{1}{2}CV_t^2 + (\frac{RC}{T})CV_{dd}^2 \quad (1)$$

where R is the effective channel resistance, T is the transition time, and V_t is the threshold voltage of the MOSFET. The first term is referred to as the threshold loss, and the second term is referred to as the resistive loss. The threshold loss may not be present in some circuits if charging/discharging through MOSFET switches do not experience a threshold voltage. Since

$RC < 1ns$ for a moderate fanout, and $T \sim 1/f$ (f is the operating frequency), $E_{dissipation}$ is very small when the operating frequency $f \sim 10MHz$.

The power supply/clock waveforms can be generated using some simple schemes [1, 3, 5] which consist of two stages, as shown in figure 2. The first stage is the DC power supply, and the second stage generates alternating current/clock waveforms, which is controlled by external clock signal(s) to maintain the constant frequency. Multiple phases of supply voltage may be required to cascade such logic gates. If the circuit viewed by the power/clock generator is modeled as constant capacitive load, the entire system is effectively an RLC resonator. The dissipated energy is replenished by the first stage DC power supply by restoring the peak voltage of the second stage to V_{dd} (supply voltage) level in each cycle. The effective capacitance of the whole circuit is approximately a constant when the circuit size is sufficiently large. This can be justified as follows. If the average number of nodes switching in a circuit per cycle is N, then (for large N) from ***central limit theorem*** [10] the number of switching nodes has a nearly Gaussian distribution with the deviation of $K\sqrt{N}$. K is a constant which depends on the circuit structure and primary input signal patterns. Hence, the percentage deviation from the average is proportional to $1/\sqrt{N}$, which is negligible when N is large. The total load capacitance in a cycle is the sum of the capacitance associated with the switching nodes. Thus, the percentage deviation of the total capacitance from its average is also negligible for large circuits.

Numerous designs of adiabatic logic have been presented in [1, 3, 2, 5, 6], which have demonstrated the possibility of achieving ultra-low energy computing. Today's VLSI systems integrate both random logic and assorted memories. Hence, it is natural to apply the adiabatic switching; principle to memories to achieve similar large savings as in random logic. However the application of such methods should not cause drastic increases in either size or circuit complexity. We address this issue in this paper by presenting the design of a ***Static RAM,*** which is capable of recovering of the order of 75%of energy for both ***read*** and ***write*** operations. This is achieved without increasing the complexity of the memory cell and with a low area overhead over conventional ***SRAM.***

The rest of the paper is organized as follows. Section 2 describes the design of SRAM memory core working in adiabatic fashion. An adiabatic address decoding scheme is described

in section 3. Design of peripheral circuits is presented in section 4. A method to derive the optimal !supply voltage is derived in section 5. Quantitative results of the performance of our designs are detailed in section 6 based on layouts using $MOSIS$ $1.2\mu m$ CMOS NWELL process. In section 7, we summarize certain features of our adiabatic SRAM.

Adiabatic SRAM Core

In this section, we first briefly describe the organization of SRAM, then we describe the topology of the memory cell and SRAM core in details. Figure 3 shows the adiabatic SRAM organization. Compared to the standard CMOS SRAM, a row driver is inserted, which generates appropriate voltage signals to drive the memory core. Also note that sense amplifiers are replaced by the voltage level shifters.

Figure 4 shows the memory cell, which is identical in topology to the 6-transistor RAM cell used in the standard SRAM. The cell consists of a cross coupled inverter pair and a pair of read/write access transistors. The pair of access transistors is enabled by the word line. A block of SRAM core is composed of a multiplicity of these cells arrayed horizontally and vertically, and the memory core is again composed of multiple blocks. Within a block, the word lines of adjacent cells are connected along the horizontal axis, while the bit and $\overline{bit}$ lines are connected for all cells in a column.

With reference to figure 4, a conventional SRAM ties V_{hi} and V_{low} to the supply V_{dd} and ground. For a discussion on the operation of a six transistor SRAM, we refer readers to [7]. The design of the cell ratios the size of M3 and M5 (M4 and M6) so that the value stored in the cell is not upset. Layout of the memory core emphasizes compactness and is usually flipped for two adjacent rows to share the power supply. The dominant component of energy dissipation arises from switching the large capacitance on the bit lines and the $word$ lines. Other significant components of energy consumption are row and column address decoders and the sense amplifier. This paper concentrates on reducing the energy required in each component by applying the adiabatic switching principle. Two types of RAM cells are presented in this section, each of which has advantages depending on the overall architecture of the SRAM.

2.1 Operations of Adiabatic SRAM Core

In the adiabatic SRAM core of figure 4, V_{hi} and V_{low} are no longer static. V_{hi}, V_{low} and V_{word} are generated by the row driver circuitry, which is shown in figure 6. The row driver will be discussed, in detail in section 4. For the time being, it is sufficient for readers to realize the following: Row selection signals, $W_0, W_1, \cdots, W_{M-1}$, which are generated from row address decoder, enable the drivers for a particular row. V_{hi}, V_{low}, and V_{word} of the enabled row may now be controlled independently by global supply lines G_{hi}, G_{low}, and G_{word}, respectively. For the unselected rows, V_{hi}, V_{low} and V_{word} are connected to the static power supply lines S_{hi}, S_{low}, and ground, respectively. Figure 4 also shows the bit-line equalization transistor. Bit-line precharge circuits of the standard CMOS SRAM are not required in adiabatic SRAM.

We now show that by the proper application of stimulus at G_{hi}, G_{low}, and G_{word}, it is possible to operate the memory core in an adiabatic fashion. For the purpose of discussion, we shall assume the DC supply voltage of $S_{hi} = V_{dd} = 5\ volts$, $S_{low} = 2$ volts, and a transistor threshold $|V_t|$ of 1 volt. The SRAM core starts out in the rest state with all rows disabled. The row driver circuit ensures that V_{hi} is at $S_{hi} = 5$ volts, V_{low} is at $S_{low} = 2$ volts, and V_{word} is pulled to GND. The bit-line is assumed precharged midway to 2 volts. A read operation starts with the row selection being applied, and the V_{word} being smoothly ramped up to 3 volts by G_{word}. V_{hi} and V_{low} are now ramped down to 3 volts and 0 volt, respectively. The waveforms are shown in figure 7. The reader can also refer to circuit model of figure 13(b) for the read operation. If we assume that internal node A was LOW($=V_{low}$) and B was HIGH($=V_{hi}$), both M1 and M5 are ON and hence, bit follows V_{low} smoothly ramping down to 0. On the other hand, $\overline{bit}$ remains at 2 volts since node B is at $V_{hi}(\geq 3$ volts) and V_{word} is at 3 $volts$, which prevents access transistor M6 from turning on. The bit-line differential is amplified by an adiabatic level-shifter to generate the logic output. Unlike the conventional SRAM, where pre-charge circuitry is required to precharge the bit-lines after each operation, bit-lines are charged back to 2 volts by the same cells being read. Indeed, bit reverts back to the rest state through M5 and M1 when V_{hi} and V_{low} ramp up to their rest state. This process replenishes the charge on bit and hence, the bit-line peripheral circuitry can be eliminated in adiabatic SRAM. Subsequently, G_{word} is pulled low, turning OFF the row and equalizing the bit-lines. The stimulus applied on V_{hi} and V_{low} ensures a constant cell voltage all the times.

The write operation requires that the bit information stored in a cell be overwritten by signals carried in the bit-lines. This is accomplished by applying the row selection, pulling up V_{word} to enable the word line, and pulling down V_{hi} to 3 volts. Thus, the voltage difference across the cell is $V_{hi} - V_{low} = 1$ volt $= V_t$, which is high enough to ensure that cell state is held for columnns which are not being written, and low enough to easily flip the cell state for the selected columns. We now simultaneously and smoothly ramp down $\overline{bit}$ and V_{low} from 2 volts to 0 *volt*. B is then pulled down by $\overline{bit}$ and the cell state flips (as we have assumed in the read operation, A was LOW and B was HIGH). Thus, we write a 0 into B, as illustrated in figure 7. Returning to the rest state is accomplished in a fashion similar to the read operation when V_{low} and bit revert back to 2 volts and *word* is disabled.

2.2 Another Core Organization – Column Activated Memory Core

One phenomenon in the operation of SRAM is that all cells in the selected row within a memory block are activated. Although only the selected columns in a block are multiplexed to sense amplifiers and 1/0 buffers, all columns are active, discharging the bit-lines and charging them back to high subsequently. For instance, considering a 64-by-64 block of memory core in a SRAM chip with 8 1/O pins, a *read* only reads out 8 bit-lines while the rest of the *56* bit-lines also perform the read operations. The energy involved in charging these *56* bit-lines is actually wasted. If we only activate the selected columns, less energy will be dissipated in each operation. Motivated by this, we present another core organization which only enables the selected columns and consumes significantly less energy under certain memory organizations.

Figure 5 shows this new memory core organization. The unique modification from the core configuration of figure 4 is that V_{hi} and V_{low} now run vertically and are generated by column driver circuitry, which can be implemented analog to the row driver circuitry. The selection signals generated from the column address decoder enable the driver for the selected columns. V_{hi} and V_{low} for these columns are then controlled by G_{hi} and G_{low}, respectively. Row driver is still required because V_{word} runs horizontally. The cell structure remains the same. The memory core starts out in the rest state with all rows disabled. The column driver circuitry ensures that V_{hi} is at $S_{hi} = 5$ volts, V_{low} is at $S_{low} = 2$ volts and V_{word} is pulled down to GND by the row driver circuitry. A read operation starts with V_{word} in the selected row being

ramped up to *2.5 volts* by G_{word}. V_{hi} and V_{low} in the *selected columns* are now ramped down to 3 *volts* and 0 *volt,* respectively. The selected columns operate identically as described in section *2.1.* Let us consider the unselected columns. The gate terminals of access transistors *M5* and *M6* are tied to V_{word} and at *2.5 volts* during the read operation. However, V_{hi} and V_{low} of the cell are not ramping and stay at *5 volts* and *2 volts,* respectively. Assume that internal node A is at LOW $= V_{low}$ and *B* at HIGH $= V_{hi}$ (refer to the cell shown in figure *5).* Since V_{word} ramps up to *2.5 volts* only, neither of the access transistors *M5* and *M6* is turned on. Thus, the pair of bit-lines of the cell stays at rest state *2 volts.* We have obtained a memory core in which only the selected columns are active. *Write* operation is performed in a similar way such that only the writing columns are active.

In this core organization, each active column is driven by its own driver, and hence only small size transmission gates are required to drive V_{hi} and V_{low} for each column in the column driver circuitry. As a comparison, the previous core organization needs substantially larger transmission gates in the row driver circuitry because V_{hi} and V_{low} in one selected row are driving all the bit-lines (remember that every cell in the selected row is discharging and charging its bit-line). Cell size in this new core organization is slightly larger than thr: previous case. The actual power advantage depends on the overall SRAM memory organization, which will be further discussed in section *2.4*

2.3 Energy Dissipation in Memory Core

The choice of voltages which we have described earlier is governed by having a safe differential V_{keep} (selected *as 1 volt*) across the cell to hold its state. Assuming that V_{dd} is the supply voltage, V_{hi} swings between V_{dd} and $(V_{dd} + V_{keep})/2$, and V_{low} swings between $(V_{dd} - V_{keep})/2$ and 0. Very high energy recovery can be achieved if the read process is gradual enough. This is because the charging and discharging paths for read operation are identical, and no cell flips its state. Hence, the energy dissipation is dominated by resistive loss, wlhich is modeled by equation (1). Threshold loss is negligible due to the fact that $V_{word} - V_{low} \approx V_t$ at the beginning of read operation.

The write operation is not truly reversible. The bit of information stored in the cell is destroyed by switching the inverter loop through a voltage of V_{keep}. Thus, there is a certain

amount of energy dissipated in the cell being written, which is inevitable because of the irreversible nature of erasing information. However, the capacitances of tht: long bit-lines are much more significant, than the capacitance of a single cell. Hence, the major portion of charging and discharging is performed adiabatically. The above analysis concludes that the energy consumption of memory core for both read and write operations can be greatly minimized if the processes are sufficiently slow. Resistive loss is the dominant factor of energy dissipation, which is inversely proportional to the signal transition time T.

2.4 Comparison of two Memory Core Organizations

We now compare energy dissipation of the two memory core organizations, which might have a major impact on the total power dissipation and chip architecture. If the signal transition time T and the supply voltage V_{dd} are fixed and the same for both core organizations, the energy dissipation depends on the capacitances to be charged/discharged, which are modeled and compared below. The main difference between the two core organizations is that the second one activates selected columns only, which usually make up a small portion of the total number of columns in a memory core block. However, all cells on the selected columns participate in the voltage swings, while in the first core organization, only those cells in a single selected row are enabled. As a matter of fact, the total capacitance associated with cells in one column is more significant than the peripheral capacitance of the bit-line, which implies that the actual capacitance being charged during an operation depends on the geometry of the core block. Assume that both designs have the same recovery rate, and one column of the cells is associated with capacitance of C_{cell} and the bit,-line has the peripheral capacitance of C_{bit}. Further assume that a block of memory core consists of M-by-N bits, with n bits read or written in each operation. Then the ratio of effective capacitance of the two core organizations is given by:

$$\frac{C_{\text{1st organization}}}{C_{\text{2nd organization}}} = \frac{NC_{bit} + (N/M)C_{cell}}{n(C_{bit} + C_{cell})} \approx (\frac{N}{n})\frac{C_{bit}}{(C_{bit} + C_{cell})} \quad (2)$$

Figure 16 compares the SPICE simrilation results of the two core organizations for a 64by-64 block (with n = 8 1/0 pins). The column activated approach consumes slightly less energy than the regular row activated organization. However, as one can observe from equation 2, when the ratio of (M/n) is large, the second approach will be certainly superior to the first

one. e.g., in a 256-by-256 block, the row activated approach would consume approximately 4 times of the energy of the column activated counterpart. Thus, our designs provide a potential space for architectural level optimization of the SRAM.

3 Adiabatic Address Decoder

Another major component of power dissipation in *SRAMs* is due to the address decoder. In order to reduce the total power consumption, it is desirable to design the address decoder which can be operated adiabatically as well. Figure 8 shows two implementations of 12-to-4 adiabatic address decoder– NAND and NOR implementation. The only difference in the configuration of the adiabatic decoder from the conventional dynamic decoder is that the precharge p-mos transistors are replaced by n-mos transistors, which no longer function as precharge transistors. Let us first show how the NOR decoder can operate in the adiabatic fashion. The decoder starts out in the rest state with V_{low} at V_{dd} and all row lines, W_0, W_1, W_2, W_3, at $V_{dd} - V_t$ (guaranteed by the transistors at the left-most column). After the address signals settle down, V_{low} gradually swings down from V_{dd} to 0, and all the row lines follow V_{low} down to 0 except the selected row staying at $V_{dd} - V_t$. Note that the transistors in the left-rnost column are disabled in this discharging process since their gate terminals are also tied to V_{low}. During the period when V_{low} stays at 0 volt, the row selection signals, W_0, W_1, W_2, W_3, are valid and are sampled by the row driver, which then enables the selected row in the memory core. Reverting back to the rest state is accomplished by ramping up V_{low} to V_{dd}, which pulls up all row lines to their rest state at $V_{dd} - V_t$. Subsequently all address signals return to zero. The waveforms of the adiabatic NOR decoder are shown in figure 9.

In a similar fashion, the adiabatic decoder can be implemented in a NAND array, which is shown in figure 8(b). Again, the precharge transistor in each row of a clynamic NAND decoder is replaced by an n-mos transistor, which ensures that every row line returns to HIGH at $V_{dd} - V_t$ after a decoding operation. The operation process is similar to the NOR decoder. In the NAND decoder, however, only one row is selected each time. The selected row line follows V_{low} ramping down to 0 while all the other row lines stay at the rest state voltage of $V_{dd} - V_t$. Because of this, adiabatic NAND decoder consumes much less energy than the NOR decoder. Results from the SPICE simulations of 6-to-64 adiabatic NOR and NAND

decoders are detailed in section 6. In a NAND structure decoder, transistors in each row are serially connected, which makes it inappropriate to be used in large decoders. Hence, the choice between the NOR and NAND structure depends on the size of decoder and the speed required.

Adiabatic decoders can also be realized in two stages with *pre-decodingr,* in which less number of transistors are used. However, two phases of decoding have to be introduced as well, hence the total time required for address decoding is the twice of the transition time of V_{low}. By using the pre-decoding stage, the total effective capacitance of the decoder is substantially smaller since less number of transistors are required. Moreover, the RC time constant in each stage is reduced, which results in higher percentage of energy recovery. Two level implementation is a better choice for adiabatic decoders of large size.

Unlike the standard CMOS SRAM, where row decoder and column decoder are usually implemented in different styles, the same decoding scheme is adopted for both row and column decoders in our adiabatic SRAM design. It is also interesting to note that, the geometric structure of address decoder is identical to the ROMs, differing only in the data pattern. Thus, the same designs for address decoder can be applied to adiabatic ROMs.

Design of Peripheral Circuits

4.1 Adiabatic Level-Shifter and 1/0 Buffer

The pair of bit lines from the memory core has a voltage difference of 2 *volts* in our designs, which need to be amplified and buffered in order to drive the 1/0 bus lines. In standard CMOS *SRAMs*, *sense amplifiers* (SA) are used to amplify the small voltage difference into full scale. The sense amplifier is usually clocked by an *enable* signal generated by the Address Transition Detection circuitry (ATD) to reduce the energy loss due to *short circuit current.* We use *a* slightly different approach in order to minimize the short circuit current and avoid the use of ATD circuitry, which consumes considerable amount of power [8].

The ;approach we use is shown in figure 10, which consists of two stages, a *voliage level-shifter* and a *buffer.* The level-shifter shown is essentially a p-mos cross-coupled sense amplifier with two access transistors. It functions as follows: initially V_{ls} is at 0 *volt* and the shifter is disabled. The two internal nodes A and $\overline{A}$ are also equalized. After the access transistors are

turned on and bit and $\overline{bit}$ arrive, the voltage difference between two sides of the cross-coupled sense amplifier is built. Then we ramp V_{ls} gradually from 0 *volts* to V_{dd}. Since the voltage difference has been built before the sense amplifier is enabled, short circuit current is not significant due to the positive feedback effect. As we can see from figure 11, node A is pulled up very rapidly and the stable state is achieved immediately, hence the short circuit current is negligible. Subsequently, the two access transistors are turned OFF to isolate the level-shifter from the bit lines such that A and $\overline{A}$ hold their states while the bit lines return to the rest state. The buffer is now ready to drive the 1/0 bus line. The transmission gate constructs of the buffer ensures that the charging and discharging processes are performed adiabatically, and simulation results indicate that more than 00% of energy recovery can be achieved. Due to the threshold voltage of the transistors, A is not able to revert back to 0 *volt* (stops at V_t instead) when V_{ls} returns to 0 *volt.* Equalization is then applied to A and $\overline{A}$. However, the energy loss in the level-shifter is not significant, since it only drives a buffer. The major portion of charging and discharging is performed adiabatically.

4.2 Row Driver Circuitry

We now consider another interface circuitry, the row driver, which uses the signals from the row address decoder to enable a particular row of the memory core. Each row of the row driver consists of two stages, a standard CMOS *tri-state* buffer and adiabatic transmission gate drivers, respectively. Let us consider the tri-state buffer first and assume that the NAND structure decoding scheme is used. The decoded signal W for the selected row is LOW at 0 *volt* when it is in the valid period. Then the *enable signal* WEN switches from 0 to V_{dd}, which enables the tri-state buffer and results in $SEL = V_{dd}$ and $\overline{SEL} = 0$ *volt.* Unlike other signals which switch gradually during a transition, WEN switches in step function to avoid the short circuit current in both NAND and NOR gates. When W is still valid, WEN switches back to 0, which simultaneously turns off the p-mos and the n-mos transistors. Thus, SEL and $\overline{SEL}$ are disconnected from their input and kept at V_{dd} and 0, respectively. The reason for using the standard CMOS tri-state buffer is as follows: Only one of the M rows is selected at each time, hence two rows have transition at most, i.e, the selected row in last cycle is unselected, and a new row is selected. Therefore, the rest of $M - 2$ rows do not have transition by using

the tri-state buffer, and hence do not consume energy (ignoring the leakage current).

The second stage of the row driver is three transmission gate drivers for V_{hi}, V_{low} and V_{word}, respectively. For the particular row which is selected, SEL = V_{dd} and $\overline{\text{SEL}}$ = 0. Hence, V_{hi}, V_{low}, ancl V_{word} are controlled by G_{hi}, G_{low}, and G_{word}, respectively. For all otlher M − 1 rows, V_{hi}, V_{low}, and V_{word} are connected to the DC supplies S_{hi}, S_{low}, and GND, respectively.

At this point, we have discussed our adiabatic SRAM from address input to data output. The overall clocking scheme for the SRAM is shown in figure 12.

5 Optimal Voltage Selection

In conventional CMOS digital systems, energy consumption decreases in proportion to the square of'the supply voltage V_{dd}. This is NOT the case for adiabatic circuits in general. Instead, there might exist an optimal voltage swing which leads to the minimal energy dissipation in adiabatic circuits with certain constructs [9]. In this section, we re-examine the optimal voltage swing problem and derive a method to find the optimal voltage for the adiabatic *SRAM*s, as well as for general adiabatic circuits.

Let us consider charging a load capacitance of C_L through a MOSFET to (delivera charge of $C_L V_{dd}$ over a time period T. The energy dissipation through the channel of the MOSFET is given by:

$$E_{diss} = \int_0^T \Delta V i dt = \int_0^T (\Delta V)^2 g dt = T<(\Delta V)^2 g> \tag{3}$$

Where g is the MOSFET channel conductance, AV is the voltage drop across the channel and <.> denotes the average over time period T. On the other hand,

$$C_L V_{dd} = \int_0^T i dt = \int_0^T \Delta V g dt = <\Delta V g> T. \tag{4}$$

Hence,

$$<\Delta V g> = C_L V_{dd}/T. \tag{5}$$

We use the following approximation to simplify E_{diss},

$$<(\Delta V)^2 g><g> \approx <\Delta V g>^2 \tag{6}$$

Substitute equation 5 and 6 into equation 3,

$$E_{diss} = T(\frac{C_L V_{dd}}{T})^2 \frac{<(\Delta V)^2 g>}{<\Delta V g>^2} = (\frac{C_L}{<g>T}) C_L V_{dd}^2. \tag{7}$$

The approximation of equation 6 turns out to be exact if ΔV is a constant throughout the charging process. More detailed analysis and simulations suggest only a small variance of ΔV, which promises the accuracy of the approximation.

We now consider a transmission gate shown in figure 13(a), which is a fundamental circuit construct to many adiabatic approaches, to drive a capacitive load C_L. The conductance of the n-MOS channel is

$$g_n = \begin{cases} k_n(V_{dd} - V_x - V_t), & V_x < V_{dd} - V_t \\ 0, & V_x \geq V_{dd} - V_t \end{cases}, \tag{8}$$

where k_n is constant, $(\mu_n \epsilon / t_{ox})(W/L)$. The average conductance is then given by:

$$\begin{aligned} <g_n> &= \frac{1}{T}\int_0^T g_n dt \\ &= \frac{1}{V_{dd}}\int_0^{V_{dd}-V_t} k_n(V_{dd} - V_x - V_t)dV_x \\ &= \frac{k_n}{2V_{dd}}(V_{dd} - V_t)^2 \end{aligned} \tag{9}$$

The waveform has been assumed to be switching linearly so that the average over time can be obtained by integrating over voltage. Similarly,

$$<g_p> = \frac{k_p}{2V_{dd}}(V_{dd} - V_t)^2. \tag{10}$$

Thus, we have

$$<g> = <g_n> + <g_p> = \frac{k_n + k_p}{2V_{dd}}(V_{dd} - V_t)^2, \tag{11}$$

and hence, the energy dissipation is:

$$E_{T-gate} = \frac{2C_L^2}{(k_n + k_p)T}\,\frac{V_{dd}^3}{(V_{dd} - V_t)^2}. \tag{12}$$

The energy dissipation E_{T-gate} has a minimum at $V_{dd} = 3V_t$. When the difference in threshold voltage between n-MOS and p-MOS is accounted for, the optimal V_{dd} lies between $3V_{tn}$ and $3V_{tp}$. Although the second order effects (e.g. the body effect,) have not been taken into account, SPICE simulations verify that $3V_t$ is close to the minimum energy dissipation. In our SRAM designs, transmission gate drivers have been used in I/O buffers, the supply drivers for V_{hi} and V_{low} of the memory core, the word-line drivers and the address signal drivers.

Let us now consider another major component of energy dissipation–the bit lines in the memory core. Circuit model of discharging a bit-line capacitance is shown in figure 13(b). The

conductance of transistor M1 and M5 are

$$g_{M1} = k_n(V_{hi} - V_{low} - V_t) = k_n(S_{word} - V_t),$$
$$g_{M5} = k_n(S_{word} - V_A - V_t),$$

where $V_A \approx V_{low}$. Set $V_{keep} = V_t$. We have $S_{word} - V_{keep} = S_{word} - V_t = S_{low}$. The serial connection of M1 and M5 gives:

$$\frac{1}{g} = \frac{1}{g_{M1}} + \frac{1}{g_{M5}} \tag{13}$$

Hence,

$$\begin{aligned} g &= \frac{g_{M1}g_{M5}}{g_{M1} + g_{M5}} \\ &= \frac{k_n(S_{word} - V_t)(S_{low} - V_{low})}{\text{Sword} + \text{Slow} - V_t - V_{low}} \\ &= k_n[(S_{word} - V_t) - \frac{(\text{Sword} - V_t)^2}{\text{Sword} + \text{Slow} - V_t - V_{low}}]. \end{aligned} \tag{14}$$

$$\begin{aligned} <g> &= \frac{1}{S_{low}} \int_0^{S_{low}} g dV_{low} \\ &= k_n(S_{word} - V_t) - \frac{k_n}{\text{slow}} \int_0^{S_{low}} \frac{(\text{Sword} - V_t)^2}{S_{word} + \text{Slow} - V_t - V_{low}} dV_{low} \\ &= k_n(S_{word} - V_t) - k_n \frac{(\text{Sword} - V_t)^2}{S_{low}} ln(\frac{S_{word} + S_{low} - V_t}{S_{word} - V_t}) \end{aligned} \tag{15}$$

On the other hand, we have $S_{word} = (V_{dd}+V_{keep})/2 = (V_{dd}+V_t)/2$ and $S_{low} = (V_{dd}-V_{keep})/2 = (V_{dd} - V_t)/2$. Substituting S_{low} and S_{word} into the above equation gives,

$$\begin{aligned} <g> &= k_n(\frac{V_{dd} - V_t}{2}) - k_n(\frac{V_{dd} - V_t}{2}) ln2 \\ &= \frac{k_n(1 - ln2)}{2}(V_{dd} - V_t). \end{aligned} \tag{16}$$

Thus, we obtain the energy dissipation on a bit-line to be

$$\begin{aligned} E_{bit} &= (\frac{C_{bit}}{<g>T}) C_{bit} S_{low}^2 \\ &= \frac{2C_{bit}}{k_n(1 - ln2)T} \frac{C_{bit}(\frac{V_{dd} - V_t}{2})^2}{(V_{dd} - V_t)} \\ &= \frac{C_{bit}}{2k_n(1 - ln2)T} C_{bit}(V_{dd} - V_t), \end{aligned} \tag{17}$$

which is proportional to $V_{dd} - V_t$. Hence, the smaller the voltage swing, the smaller the energy dissipation. However, there are other constraints on the voltage swing. With the level-shifter

we designed, $S_{low} > V_t$ must hold to make it work properly, which results in $V_{dd} > 3V_t$. Therefore, for our design of the SRAM, the optimal voltage swing lies between 314 and $4V_t$. The more accurate value can be determined by simulations and also depends on the particular design of the level shifter (remember that there are plenty of choices for the level-shifter).

6 Implementation and Results

We implemented the circuits using $MOSIS$ $1.2\mu m$ CMOS process. The layout of the cell of the two core organizations (referred to as design 1 and design 2) is shown in figure 14. The cell size of the two designs is 27λx40λ and 32λx38λ, respectively, where $\lambda = 0.6\mu m$. Unlike standard RAM cells, a memory cell working in an adiabatic fashion does not require the high aspect ratio between the pass and the pull down transistor. The ratio of transistors h13 and M5 (M4 and M6) is unity in our designs. Although the topology of the cell in design 1 is identical to a standard 6-transistor RAM cell, the fact that separate V_{hi} and V_{low} lines are needed for each row results in additional area overhead. For comparison, a conventional cell takes 27λx36.6λ area. The cell of design 2 takes the largest area due to the vertically laying of V_{hi} and V_{low} lines.

Spice simulations were carried out on extracted sub-circuits for a group of four SRAM cells for both core organizations. Full parasitic extraction was performed. We simulated a block of $SRAM$ core of 64 rows by 64 columns by accounting for various capacitive loads on lines. Simulations using level three SPICE models from a recent MOSIS run verified the operations of the memory core. Waveforms from simulations for both core organization are similar and results for the first core organization are shown in figure 7. The first operation was a *read* followed by a *write*. Figure 15 shows the fraction of energy recovered at various speed of operation. For the stimulus with transition time of $10ns$, energy recovery was .around 50% for both read and write operations.

We also simulated the extracted circuits of 6-to-64 NOR and NAND decoders implemented in the same technology, which is shown in figure 17. Results indicate that approximately 90% of energy. can be recovered for both organizations when the stimulus has a transition time of $10ns$. However, as figure 18 suggested, the energy dissipation of the NAND-array decoder is far less than NOR-array decoder. The reason is due to the fact that only a single row is pulled

down in the NAND decoder while N − 1 rows are pulled down in the NOR decoder, which has been discussed in section 3.

Although level-shifter with high energy recovery is difficult to design, it should be noted that it only drives a buffer. The energy loss at voltage-level-shift stage is insignificant. The I/O buffer, which has heavy capacitive load, can perform in adiabatic fashion and can recover most of the energy by using the transmission gate construct. Our simulation results indicate that more than 90% of energy recovery can be achieved with $1pF$ capacitive load at $10ns$ of stimulus transition time.

7 Summary

In this paper, we have demonstrated a novel application of the principle of adiabatic switching to the design of *Static* RAM, which can be implemented without significantly increasing area or circuit, complexity. Our design also provides the flexibility for possible optimizations of the SRAM from overall architectural considerations. Results indicate that the essential advantage of adiabatic logic, that of low-power, is achievable in SRAM.

Certain features of our adiabatic SRAM designs are summarized as follows:

- Very small standby current since 6-transistor cell organization is used.
- Bit-line peripheral circuitry has been eliminated except for the equalization transistor. There is no need for pre-charge since bit-lines revert to their rest state voltage level at the end of each operation.
- Tht: sense amplifier has been replaced by the adiabatic voltage level shifter and hence, short circuit current has been minimized.
- Address transition detection circuitry (ATD) is not adopted, which is used to reduce the short circuit current of the sense amplifiers in the standard CMOS *SRAM*. However, ATD circuitry itself consumes considerable amount of energy[8]. Fully global controls are required in adiabatic designs for energy recovery, and hence, self-timed approaches are not used in our design. The overall supply schemes for the SRAM is shown in figure 12.

Control circuits are not addressed in this paper because they are not among the main components of energy dissipation in *SRAMs* and can be implemented in stantlard CMOS.

In this paper the issue of the design of power supplies required by the SRAM has not been addressed. The possibility of the above designs to be used in actual practice, hinges largely on the ability to ensure proper supplies to the SRAM. At this moment, on-chip methods for generating those supply waveforms are not present. It is however feasible to generate off-chip supplies[1, 3, 5] to make adiabatic operation of memories a reality.

References

[1] W.C.Athas, L."J".Svensson, J.G.Koller, N.Tzartzanis, and E.Chou, " A Framework for Practical Low-Power Digital CMOS Systems Using Adiabatic-Switching P'rinciplea," *Intl. Workshop* on Low Power *Design*,Napa Valley, California, 19'34, pp. 189-194.

[2] J.S.Denker, S.C.Avery, A.G.Dickinson, A.Kramer, and T.R.Wik, "Adiabatic Computing with the 2N-2N2D logic Family," Inil. Workshop on *Low Power Design*, Napa Valley, California, 1994, pp. 183-187.

[3] S.G.Younis and T.F.Knight, "Asymptotically Zero Energy Split-Level Charge Recovery Logic," Inil. *Workshop* on Low Power Design, Napa Valley,California, 1994, pp. 177-182.

[4] A.P. Chandrakashan, S. Sheng, and R. Brodersen, "Low Power CMOS Digital Design," IEE'E Journal of *Solid-State* Circuits., vol. 27, No. 4, Apr. 1992, pp. 473-483.

[5] A.G. Dickinson and J.S.Denker, "Adiabatic Dynamic Logic," IEEE Journal of Solid-Siaie *Circuits*, Vol.30, No.3, March 1995, pp.311-315.

[6] Y. Ye and K. Roy, "Energy Recovery Circuits Using Reversible and Partially Reversible Logic," IEEE *Trans.* on Circuits and Sysiems I, to appear.

[7] L.A. Glasser and D.W.Dobberpuhl, The design and *Analysis* of VLSI *Circuits*, pp.390-393.

[8] S.T. Flannagan, P.H. Pelley, N. Herr, B.E. Engles, T. Feng, S.G. Nogle, J.W. Eagan, R.J. Dunnigan, L.J. Day and R.I. Kung, "8-ns CMOS 64K x 4 and 256K x 1 SRAM's," IEEE *Journal* of *Solid-State Circuits*, Vol.25, No.5, Oct. 1'3'35, pp.1049-1056.

[9] W.C.Athas, L."J".Svensson, J.G.Koller, N.Tzartzanis, and E.Chou, "Low-Power Digital Systems Based on Adiabatic-Switching Principles,"IEEE Trans. on VLSI Systems, Vo1.2, No.4, Dec. 1994, pp.398-407.

[10] A. Papoulis, Probability, Random variables, *and Stochastic* Processes, Third edition, Mc-Graw Hill.

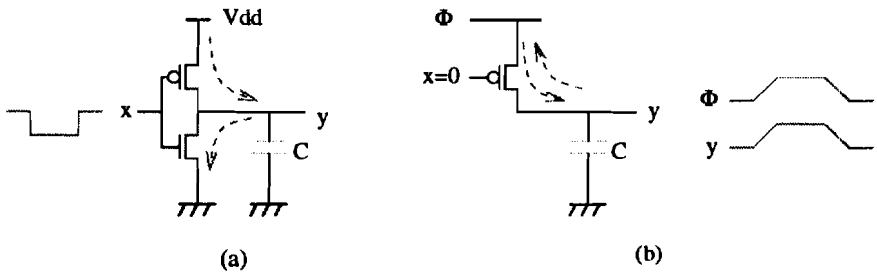


Figure 1: Charing and discharging in standard CMOS and adiabatic fashion

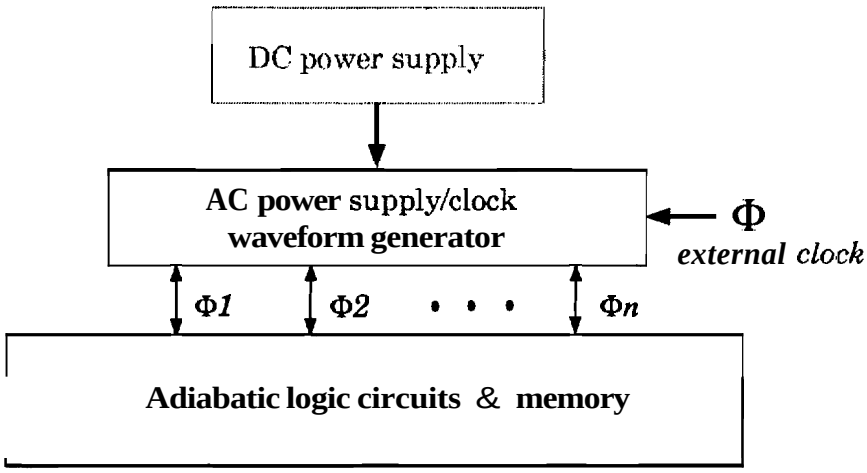


Figure 2: A framework for power supply and circuit

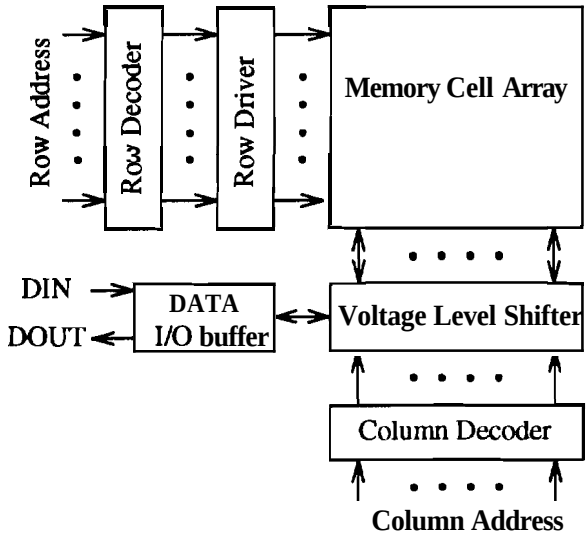


Figure 3: SRAM organization

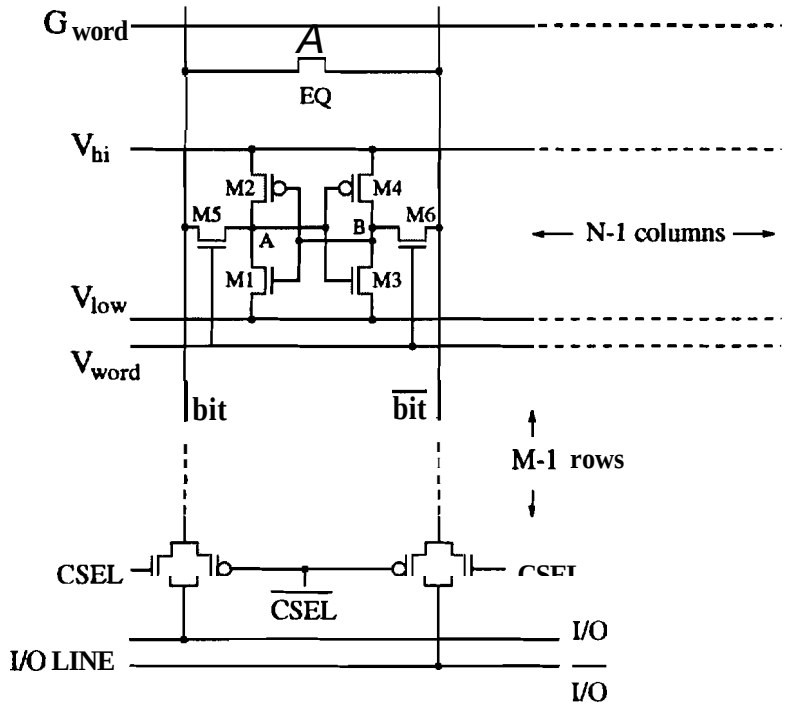


Figure 4: Memory core organization

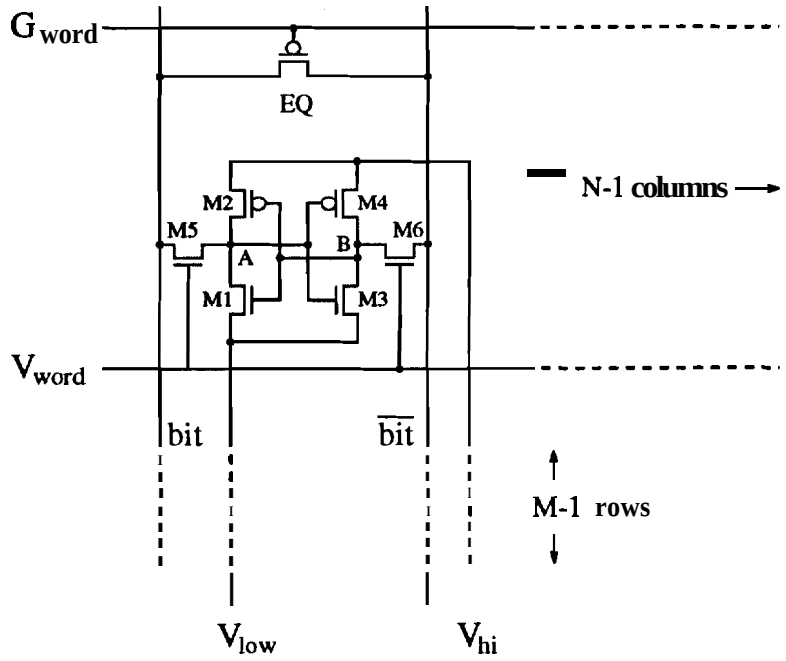


Figure 5: Another memory core organization. V_{hi} and V_{low} run vertically and are generated by column drivel: circuitry. In this core organization, only selected columns are activated.

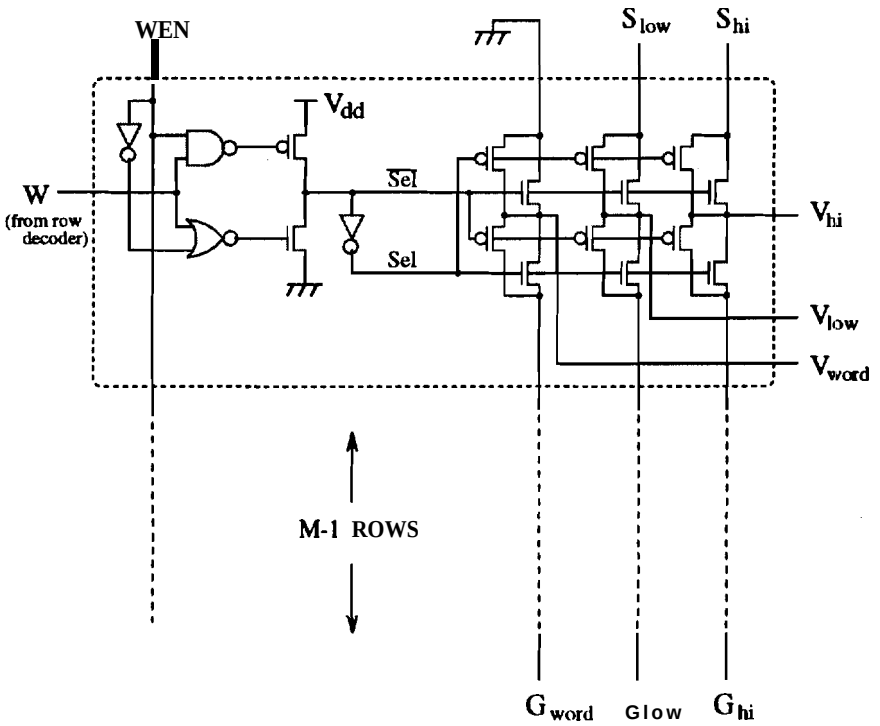


Figure 6: The row driver circuit

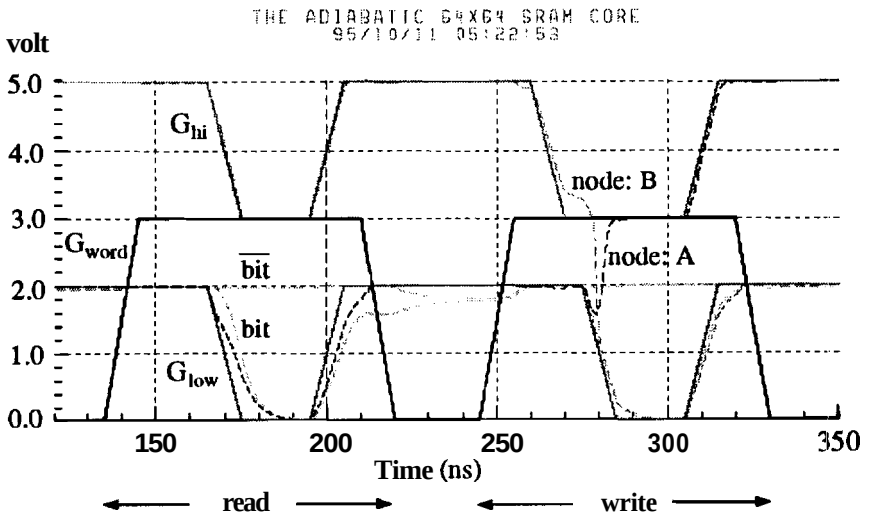
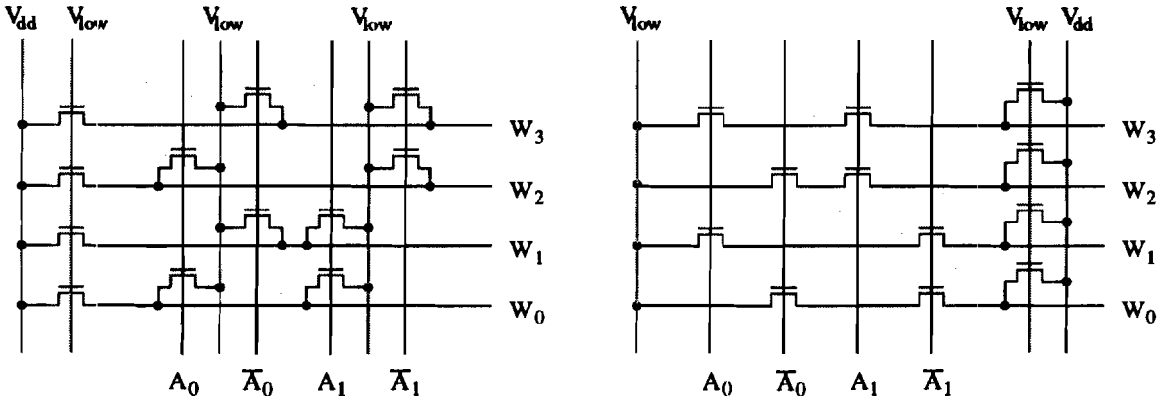


Figure 7: Waveforms for the SRAM core



(a) 2-to-4 adiabatic NOR decoder (b) 2-to-4 adiabatic NAND decoder

Figure 8: Configurations of adiabatic address decoder

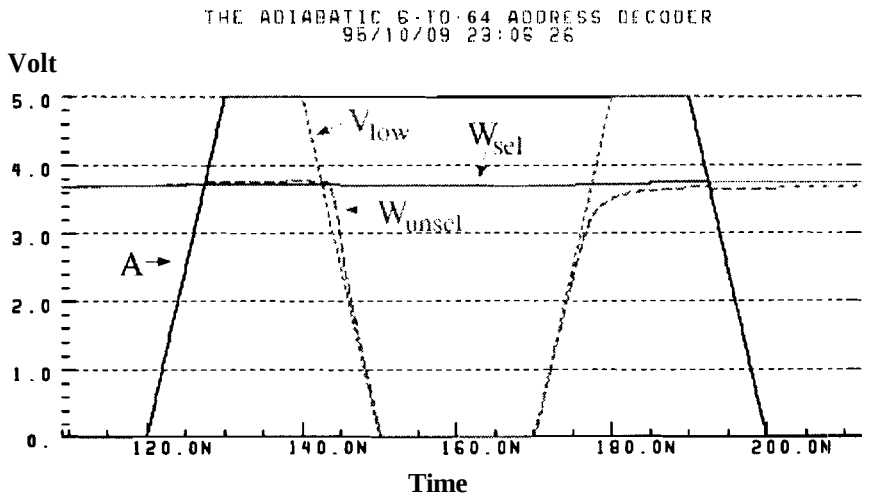


Figure 9: Waveforms of adiabatic address decoder

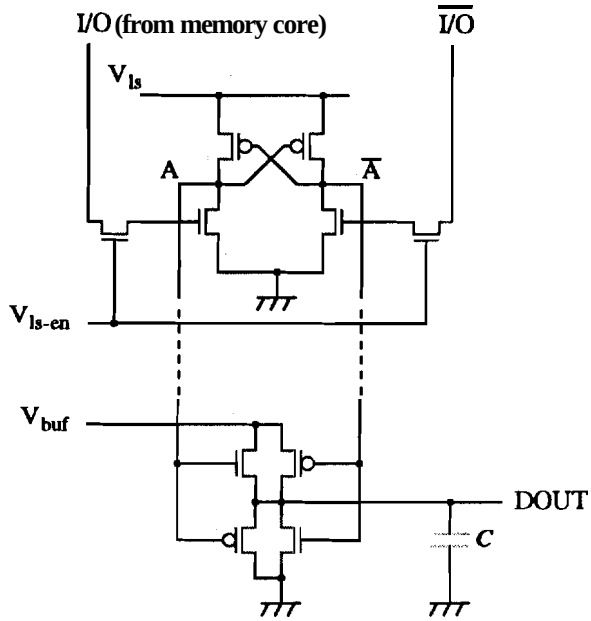


Figure 10: Adiabatic level-shifter and I/O buffer

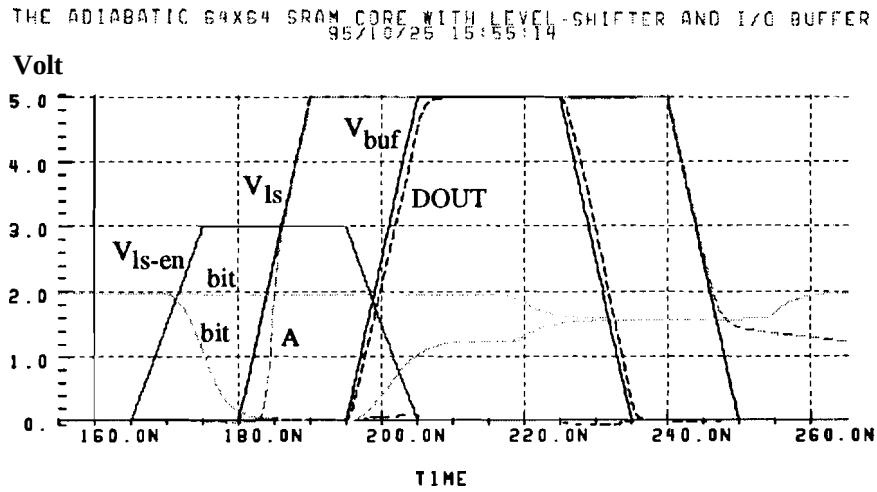


Figure 11: Waveforms for the adiabatic level-shifter

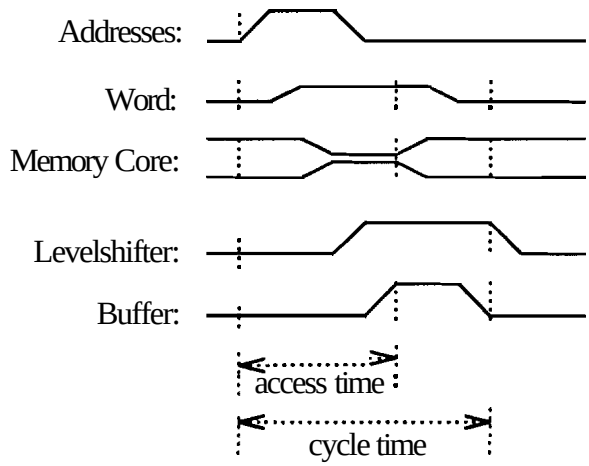


Figure 12: Overall supply schemes for the adiabatic *SRAM*

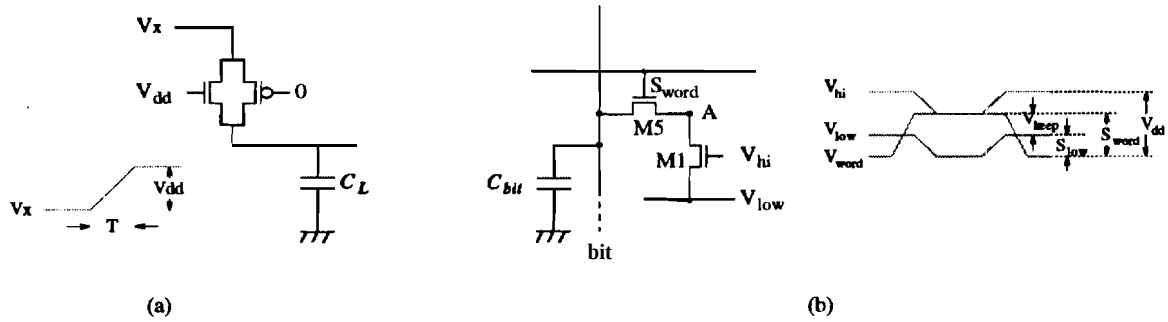


Figure 13: Circuit models of adiabatic charging/discharging

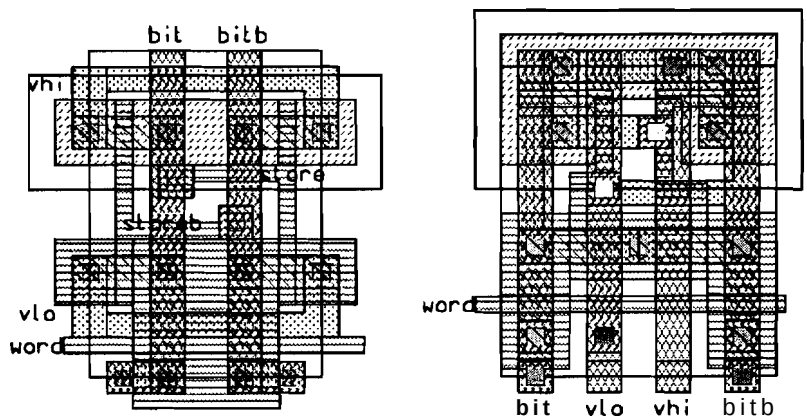


Figure 14: Layout of two *SRAM* cell configurations. The cell shown on the left side is for row activated memory core organization, and the cell shown on the right side is for column activated memory core organization, in which only the selected columns are active. The supply lines V_{hi} and V_{low} run horizontally in the left cell and vertically in the right cell.

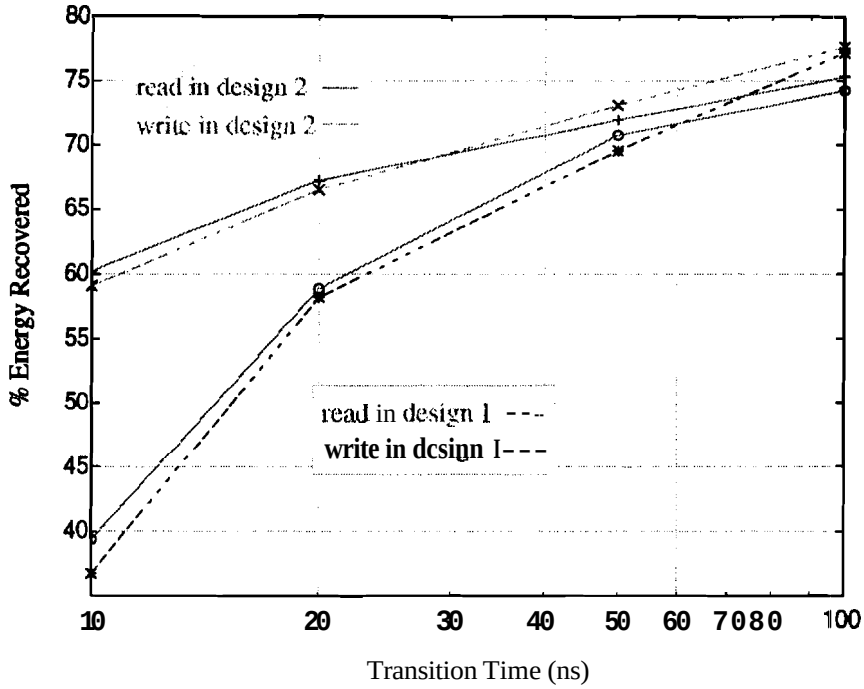


Figure 15: Energy recovery in adiabatic memory core

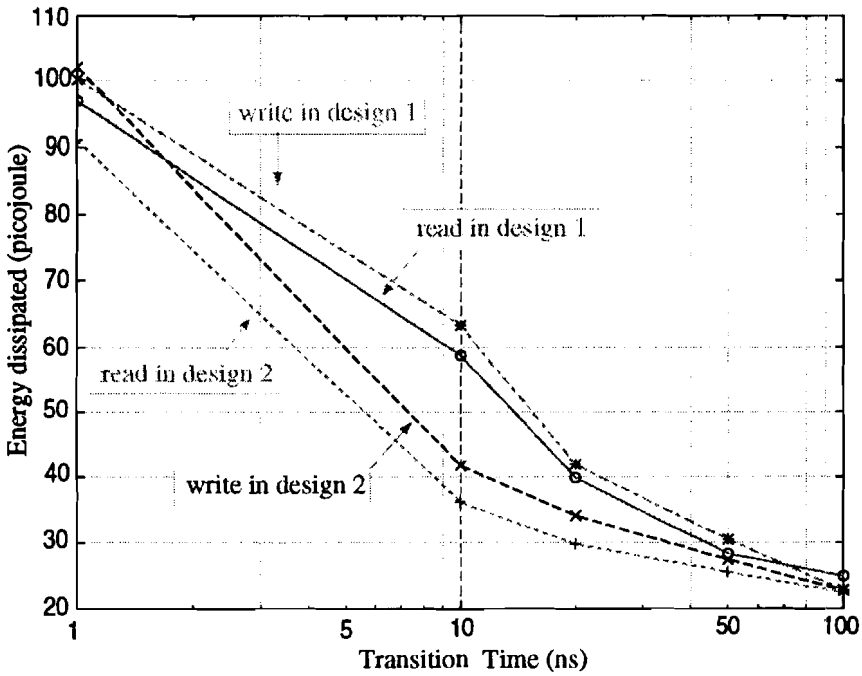


Figure 16: Comparison of energy dissipation in two memory core organizations

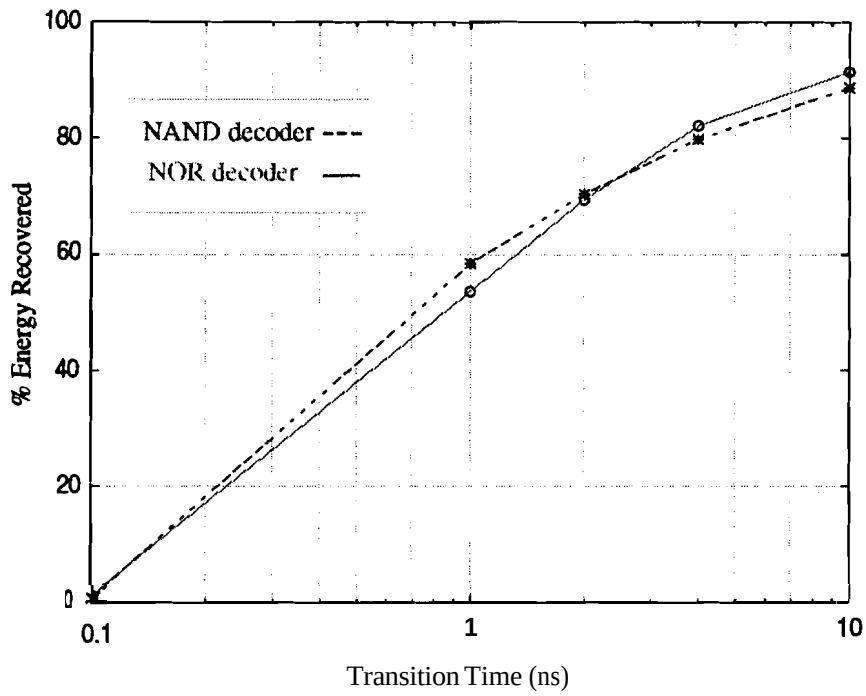


Figure 17: Energy recovery in 6-to-64 adiabatic NAND and NOR decoders

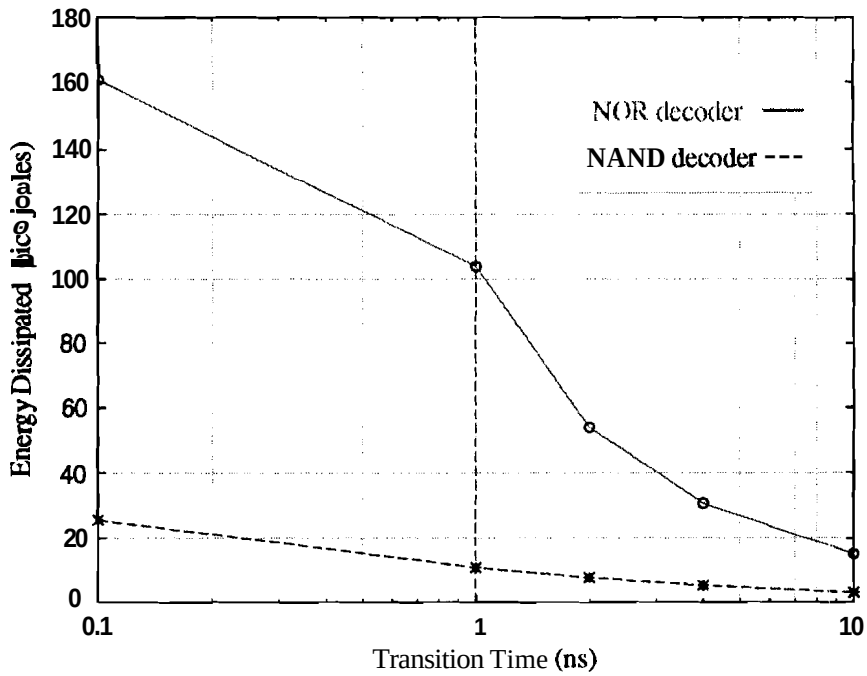


Figure 18: Comparison of energy dissipation in adiabatic NAND and NOR decoder