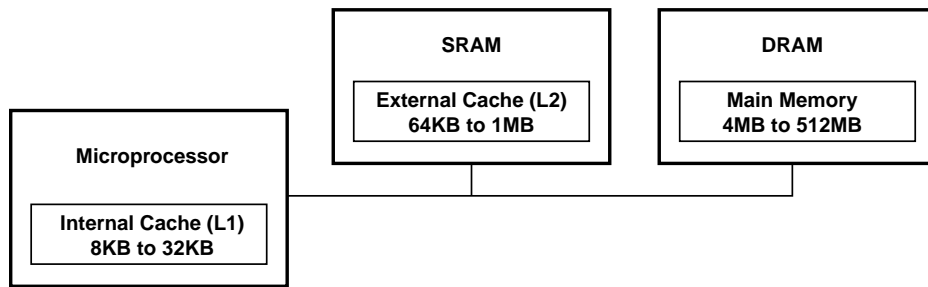# 8 SRAM TECHNOLOGY

OVERVIEW

An SRAM (Static Random Access Memory) is designed to fill two needs: to provide a direct interface with the CPU at speeds not attainable by DRAMs and to replace DRAMs in systems that require very low power consumption. In the first role, the SRAM serves as cache memory, interfacing between DRAMs and the CPU. Figure 8-1 shows a typical PC microprocessor memory configuration.


Source: Micron/ICE, "Memory 1997"   20812

Figure 8-1. Typical PC Microprocessor Memory Configuration

The second driving force for SRAM technology is low power applications. In this case, SRAMs are used in most portable equipment because the DRAM refresh current is several orders of magnitude more than the low-power SRAM standby current. For low-power SRAMs, access time is comparable to a standard DRAM. Figure 8-2 shows a partial list of Hitachi's SRAM products and gives an overview of some of the applications where these SRAMs are found.
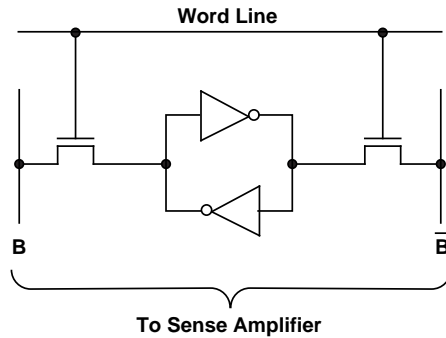
HOW THE DEVICE WORKS

The SRAM cell consists of a bi-stable flip-flop connected to the internal circuitry by two access transistors (Figure 8-3). When the cell is not addressed, the two access transistors are closed and the data is kept to a stable state, latched within the flip-flop.
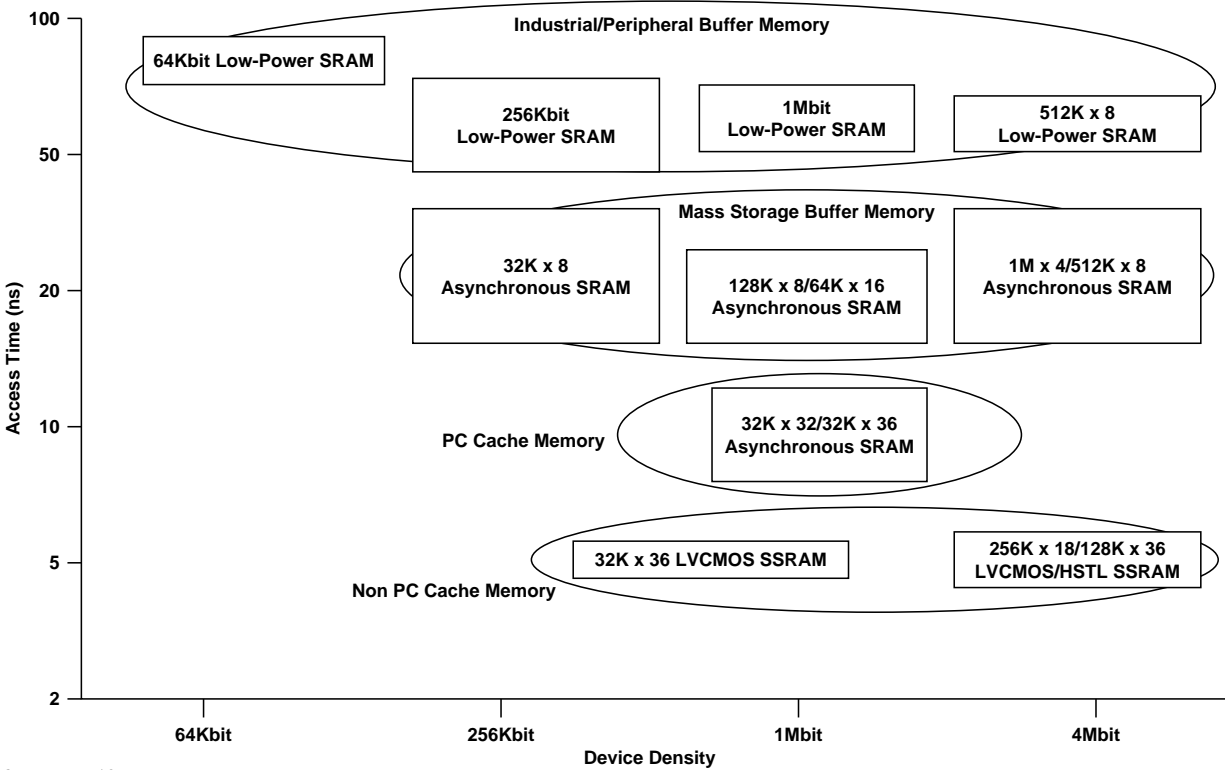
Figure 8-2. Hitachi's SRAM Products


Figure 8-3. SRAM Cell

The flip-flop needs the power supply to keep the information. The data in an SRAM cell is volatile (i.e., the data is lost when the power is removed). However, the data does not "leak away" like in a DRAM, so the SRAM does not require a refresh cycle.

Read/Write

Figure 8-4 shows the read/write operations of an SRAM. To select a cell, the two access transistors must be "on" so the elementary cell (the flip-flop) can be connected to the internal SRAM circuitry. These two access transistors of a cell are connected to the word line (also called row or X address). The selected row will be set at V_{CC}. The two flip-flop sides are thus connected to a pair of lines, B and $\overline{B}$. The bit lines are also called columns or Y addresses.
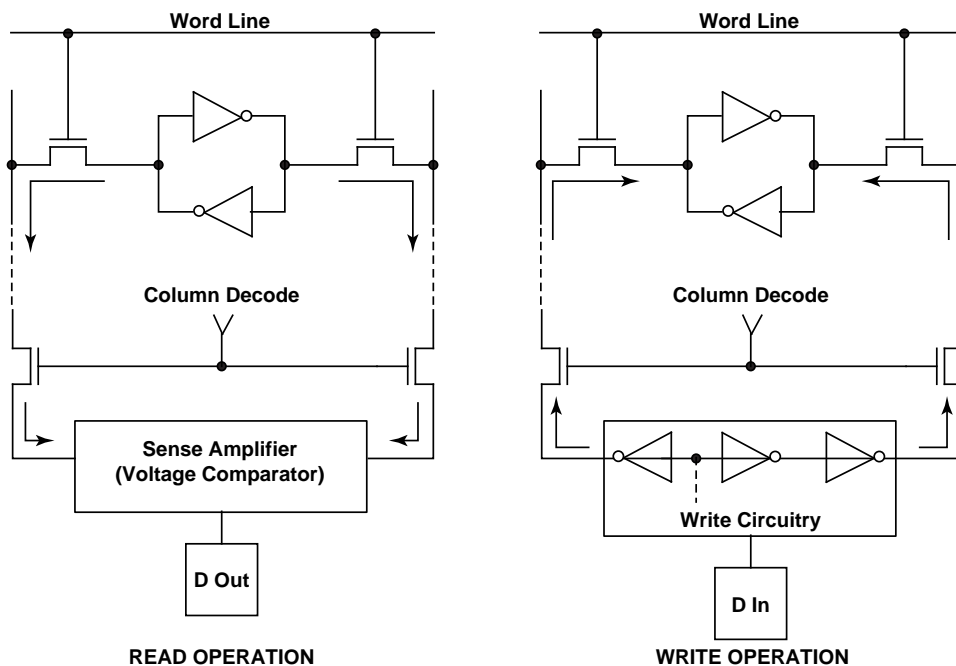
Source: ICE, "Memory 1997"   19952

Figure 8-4. Read/Write Operations

During a read operation these two bit lines are connected to the sense amplifier that recognizes if a logic data "1" or "0" is stored in the selected elementary cell. This sense amplifier then transfers the logic state to the output buffer which is connected to the output pad. There are as many sense amplifiers as there are output pads.

During a write operation, data comes from the input pad. It then moves to the write circuitry. Since the write circuitry drivers are stronger than the cell flip-flop transistors, the data will be forced onto the cell.

When the read/write operation is completed, the word line (row) is set to 0V, the cell (flip-flop) either keeps its original data for a read cycle or stores the new data which was loaded during the write cycle.

Data Retention

To work properly and to ensure that the data in the elementary cell will not be altered, the SRAM must be supplied by a V_{CC} (power supply) that will not fluctuate beyond plus or minus five or ten percent of the V_{CC}.

If the elementary cell is not disturbed, a lower voltage (2 volts) is acceptable to ensure that the cell will correctly keep the data. In that case, the SRAM is set to a retention mode where the power supply is lowered, and the part is no longer accessible. Figure 8-5 shows an example of how the V_{CC} power supply must be lowered to ensure good data retention.
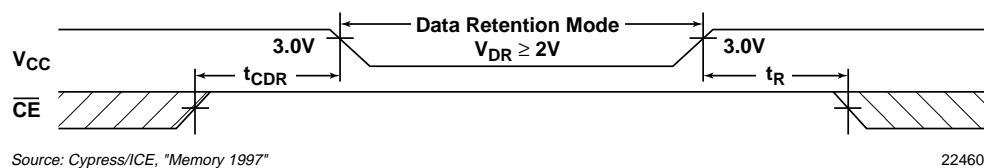

Source: Cypress/ICE, "Memory 1997"   22460

**Figure 8-5. SRAM Data Retention Waveform**

MEMORY CELL

Different types of SRAM cells are based on the type of load used in the elementary inverter of the flip-flop cell. There are currently three types of SRAM memory cells :

- The 4T cell (four NMOS transistors plus two poly load resistors)
- The 6T cell (six transistors—four NMOS transistors plus two PMOS transistors)
- The TFT cell (four NMOS transistors plus two loads called TFTs)

4 Transistor (4T) Cell

The most common SRAM cell consists of four NMOS transistors plus two poly-load resistors (Figure 8-6). This design is called the 4T cell SRAM. Two NMOS transistors are pass-transistors. These transistors have their gates tied to the word line and connect the cell to the columns. The two other NMOS transistors are the pull-downs of the flip-flop inverters. The loads of the inverters consist of a very high polysilicon resistor.

This design is the most popular because of its size compared to a 6T cell. The cell needs room only for the four NMOS transistors. The poly loads are stacked above these transistors. Although the 4T SRAM cell may be smaller than the 6T cell, it is still about four times as large as the cell of a comparable generation DRAM cell.
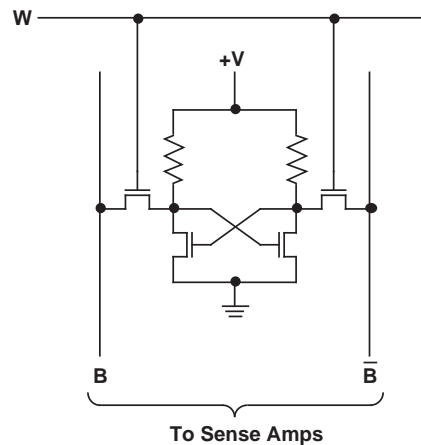
Figure 8-6. SRAM 4T (Four-Transistor) Cell

The complexity of the 4T cell is to make a resistor load high enough (in the range of giga-ohms) to minimize the current. However, this resistor must not be too high to guarantee good functionality.

Despite its size advantage, the 4T cells have several limitations. These include the fact that each cell has current flowing in one resistor (i.e., the SRAM has a high standby current), the cell is sensitive to noise and soft error because the resistance is so high, and the cell is not as fast as the 6T cell.

6 Transistor (6T) Cell

A different cell design that eliminates the above limitations is the use of a CMOS flip-flop. In this case, the load is replaced by a PMOS transistor. This SRAM cell is composed of six transistors, one NMOS transistor and one PMOS transistor for each inverter, plus two NMOS transistors connected to the row line. This configuration is called a 6T Cell. Figure 8-7 shows this structure. This cell offers better electrical performances (speed, noise immunity, standby current) than a 4T structure. The main disadvantage of this cell is its large size.

Until recently, the 6T cell architecture was reserved for niche markets such as military or space that needed high immunity components. However, with commercial applications needing faster SRAMs, the 6T cell may be implemented into more widespread applications in the future.

Much process development has been done to reduce the size of the 6T cell. At the 1997 ISSCC conference, all papers presented on fast SRAMs described the 6T cell architecture (Figure 8-8).
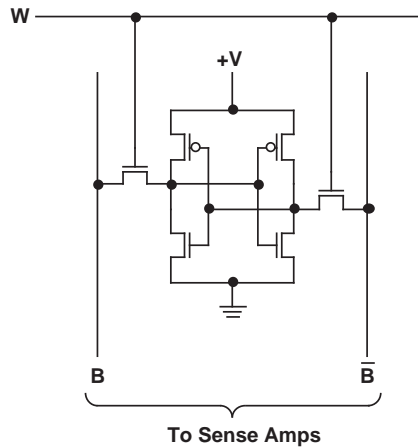
Source: ICE , "Memory 1997"   18471A

Figure 8-7. SRAM 6T (Six Transistor) Cell

Density	Company	Cell Type	Cell Size (μm²)	Process	Die Size (mm²)
4Mbit	NEC	6T	12.77	0.25μm	132
4Mbit	IBM	6T	18.77	0.3μm 0.2μm L$_{eff}$	145
128Kbit	Hitachi	6T	21.67	0.35μm	5.34

Source: ICE, "Memory 1997"   22459

**Figure 8-8. 1997 ISSCC Fast SRAM Examples**

TFT (Thin Film Transistor) Cell

Manufacturers have tried to reduce the current flowing in the resistor load of a 4T cell. As a result, designers developed a structure to change, during operating, the electrical characteristics of the resistor load by controlling the channel of a transistor.

This resistor is configured as a PMOS transistor and is called a thin film transistor (TFT). It is formed by depositing several layers of polysilicon above the silicon surface. The source/channel/drain is formed in the polysilicon load. The gate of this TFT is polysilicon and is tied to the gate of the opposite inverter as in the 6T cell architecture. The oxide between this control gate and the TFT polysilicon channel must be thin enough to ensure the effectiveness of the transistor.

The performance of the TFT PMOS transistor is not as good as a standard PMOS silicon transistor used in a 6T cell. It should be more realistically compared to the linear polysilicon resistor characteristics.
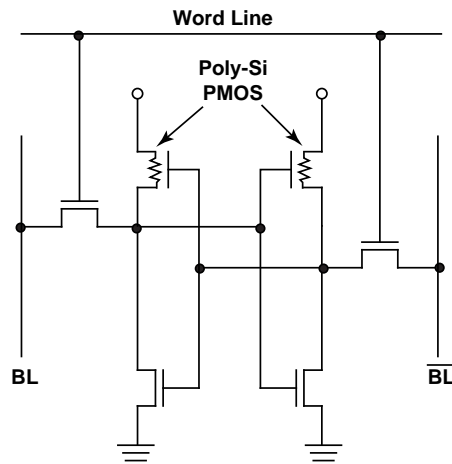
Figure 8-9 shows the TFT characteristics. In actual use, the effective resistance would range from about 11 x $10^{13}\Omega$ to 5 x $10^9\Omega$. Figure 8-10 shows the TFT cell schematic.
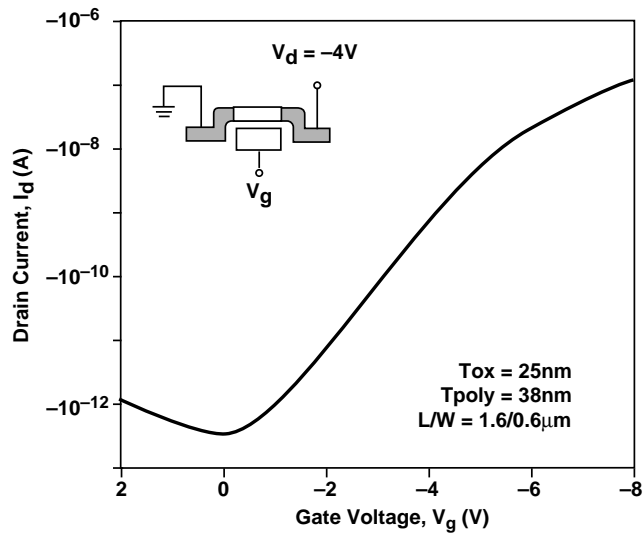
Source: Hitachi/ICE, Memory 1997"   19953

Figure 8-9. TFT (Thin Film Transistor) Characteristics

Source: ICE, "Memory 1997"   19954

Figure 8-10. SRAM TFT Cell

Figure 8-11 displays a cross-sectional drawing of the TFT cell. TFT technology requires the deposition of two more films and at least three more photolithography steps.
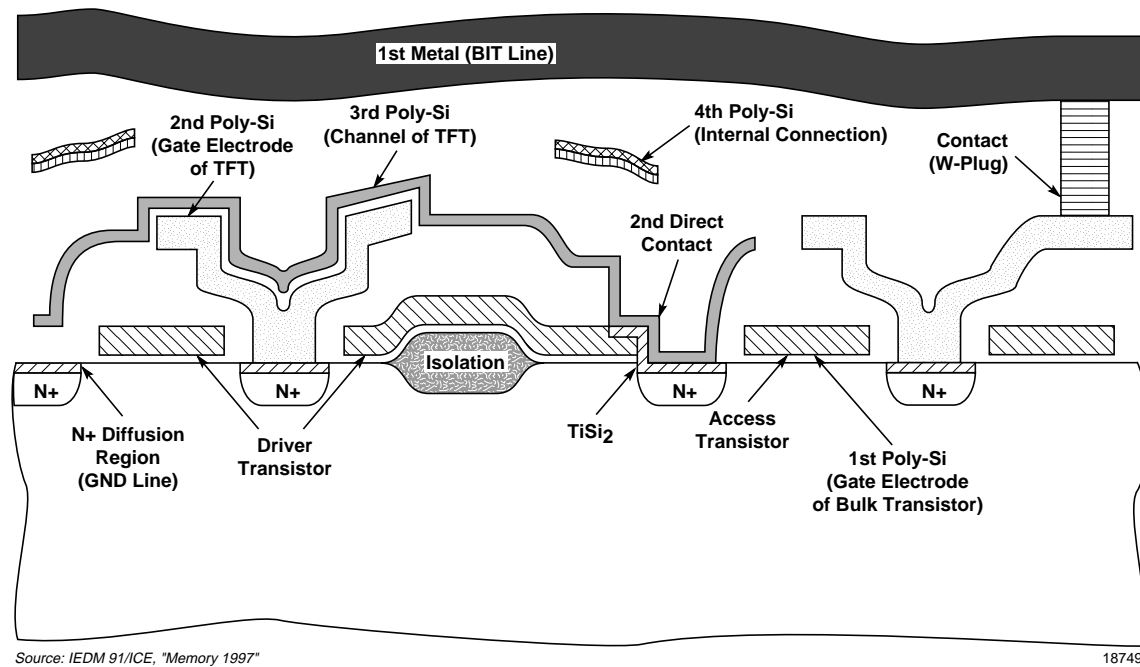
18749

Figure 8-11. Cross Section of a TFT SRAM Cell

Development of TFT technology continues to be performed. At the 1996 IEDM conference, two papers were presented on the subject. There are not as many TFT SRAMs as might be expected, due to a more complex technology compared to the 4T cell technology and, perhaps, due to poor TFT electrical characteristics compared to a PMOS transistor.

Cell Size and Die Size

Figure 8-12 shows characteristics of SRAM parts analyzed in ICE's laboratory in 1996 and 1997. The majority of the listed suppliers use the conventional 4T cell architecture. Only two chips were made with a TFT cell architecture, and the only 6T cell architecture SRAM analyzed was the Pentium Pro L2 Cache SRAM from Intel.

As indicated by the date code of the part and its technology, this study is a presentation of what is the state-of-the-art today. ICE expects to see more 6T cell architectures in the future.

Figure 8-13 shows the trends of SRAM cell size. Like most other memory products, there is a tradeoff between the performance of the cell and its process complexity. Most manufacturers believe that the manufacturing process for the TFT-cell SRAM is too difficult, regardless of its performance advantages.

	Date Code	Cell Type	Cell Size (μm^2)	Die Size (mm^2)	Min Gate (N) (μm)
Toshiba 4Mbit	9509	4T	22	144	0.65
Samsung 1Mbit	1995	4T	14.25	33	0.5
Galvantech 1Mbit	9524	4T	16.5	31	0.4
Hitachi 1Mbit	9539	4T	19	64	0.45
NEC 1Mbit	9436	4T	19	67	0.6
Motorola 1Mbit	9443	4T	40	108	0.6
Hualon 256Kbit	9523	4T	30	13.5	0.45
ISSI 1Mbit	9445	4T	27.5	50	0.5
Mosel-Vitelic 1Mbit	9409	4T	44	94.7	0.65
NEC 1Mbit	9506	4T	15.7	42.5	0.5
Samsung 4Mbit	9606	TFT	11.7	77.8	0.65
Sony 1Mbit	?	TFT	20	59	0.5
TM Tech 1Mbit	9530	4T	20	35	0.35
UMC 2Mbit	9631	4T	11.25	41	0.3
Winbond 1Mbit	9612	4T	10.15	32.5	0.5
Intel Pentium Pro L2 Cache	—	6T	33	—	0.35

Source: ICE, "Memory 1997"   22461

Figure 8-12. Physical Geometries of SRAMs

Figures 8-14 and 8-15 show size and layout comparisons of a 4T cell and a 6T cell using the same technology generation (0.3μm process). These two parts were analyzed by ICE's laboratory in 1996.

One of the major process improvements in the development of SRAM technology is the so called self aligned contact (SAC). This process suppresses the spacing between the metal contacts and the poly gates and is illustrated in Figure 8-16.
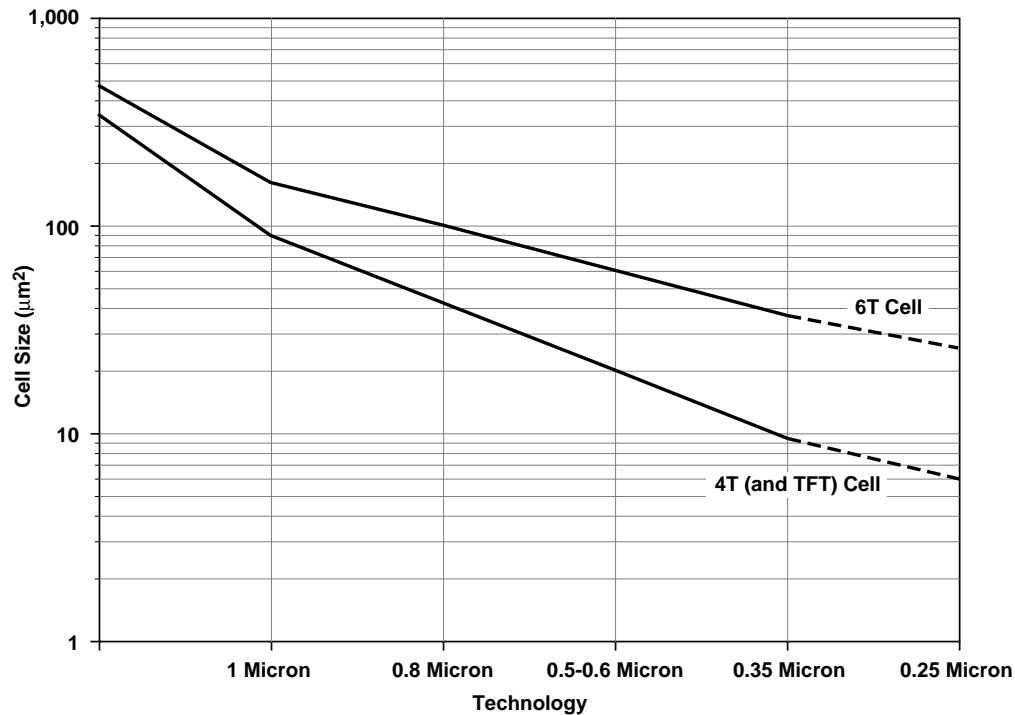
Figure 8-13. Trend of SRAM Cell Sizes

CONFIGURATION

As shown in Figure 8-17, SRAMs can be classified in four main categories. The segments are asynchronous SRAMs, synchronous SRAMs, special SRAMs, and non-volatile SRAMs. These are highlighted below.

Asynchronous SRAMs

Figure 8-18 shows a typical functional block diagram and a typical pin configuration of an asynchronous SRAM. The memory is managed by three control signals. One signal is the chip select (CS) or chip enable (CE) that selects or de-selects the chip. When the chip is de-selected, the part is in stand-by mode (minimum current consumption) and the outputs are in a high impedance state. Another signal is the output enable (OE) that controls the outputs (valid data or high impedance). Thirdly, is the write enable (WE) that selects read or write cycles.
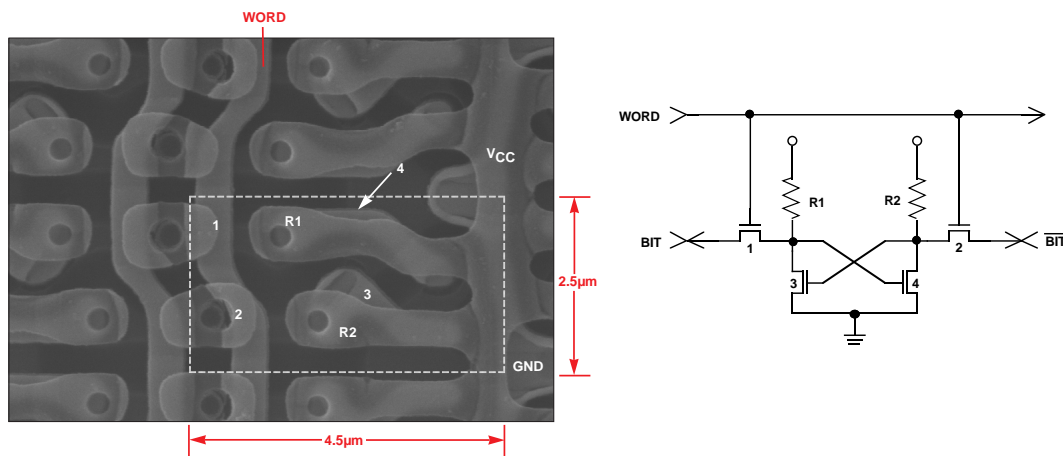
Synchronous SRAMs

As computer system clocks increased, the demand for very fast SRAMs necessitated variations on the standard asynchronous fast SRAM. The result was the synchronous SRAM (SSRAM).
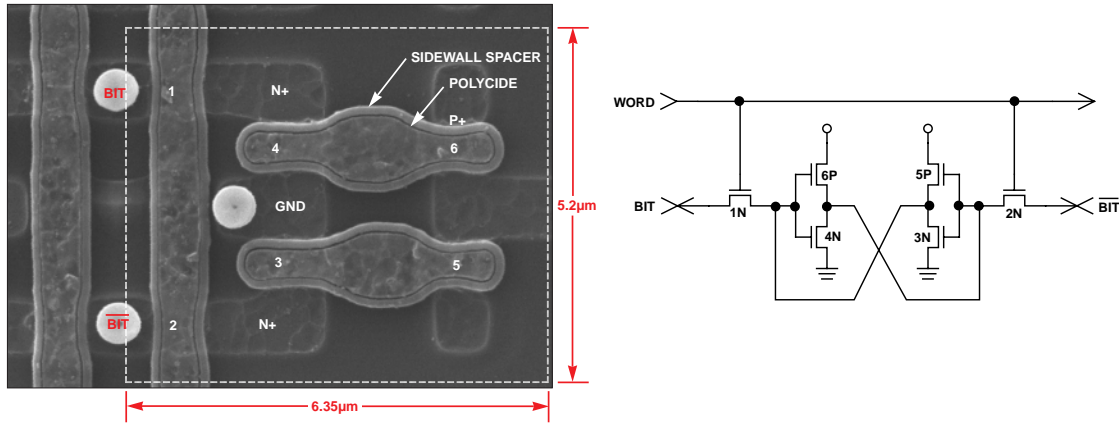
22172

Figure 8-14. 6T SRAM Cell

22171

Figure 8-15. 4T SRAM Cell

Synchronous SRAMs have their read or write cycles synchronized with the microprocessor clock and therefore can be used in very high-speed applications. An important application for synchronous SRAMs is cache SRAM used in Pentium- or PowerPC-based PCs and workstations. Figure 8-19 shows the trends of PC cache SRAM.

Figure 8-20 shows a typical SSRAM block diagram as well as a typical pin configuration. SSRAMs typically have a 32 bit output configuration while standard SRAMs have typically a 8 bit output configuration. The RAM array, which forms the heart of an asynchronous SRAM, is also found in SSRAM. Since the operations take place on the rising edge of the clock signal, it is unecessary to hold the address and write data state throughout the entire cycle.
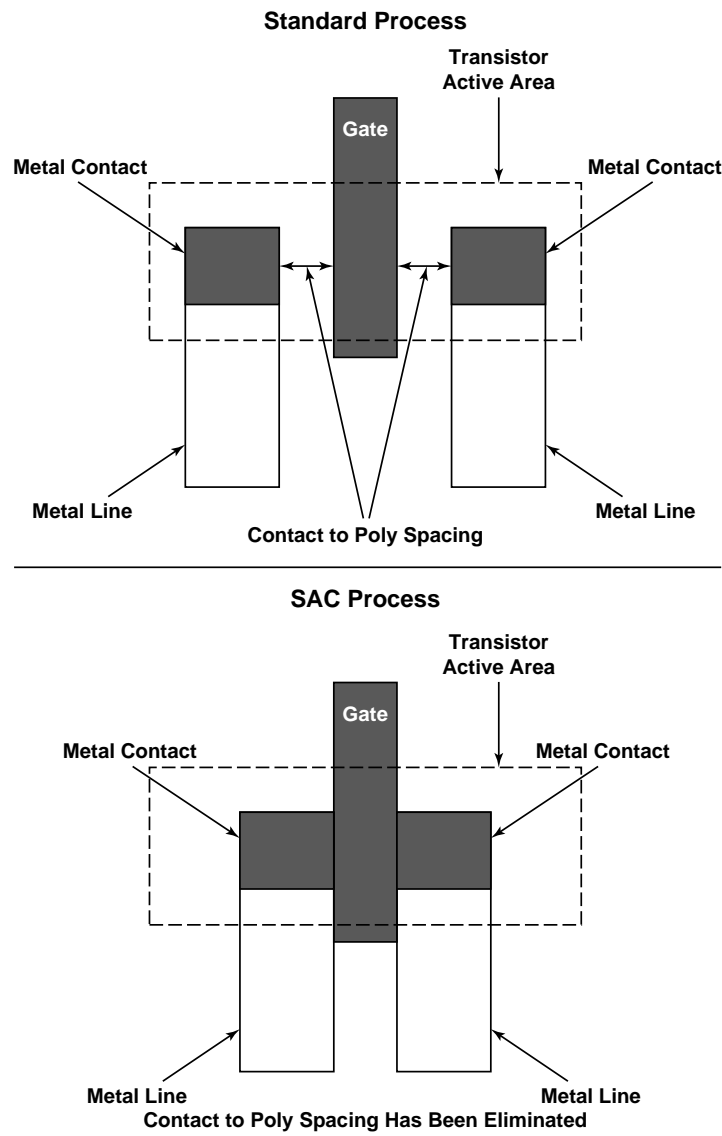
Standard Process

SAC Process

22456

**Figure 8-16. Self Aligned Contact (SAC) Process**

Burst Mode

The SSRAM can be addressed in burst mode for faster speed. In burst mode, the address for the first data is placed on the address bus. The three following data blocks are addressed by an internal built-in counter. Data is available at the microprocessor clock rate. Figure 8-21 shows SSRAM timing. Interleaved burst configurations may be used in Pentium applications or linear burst configurations for PowerPC applications.
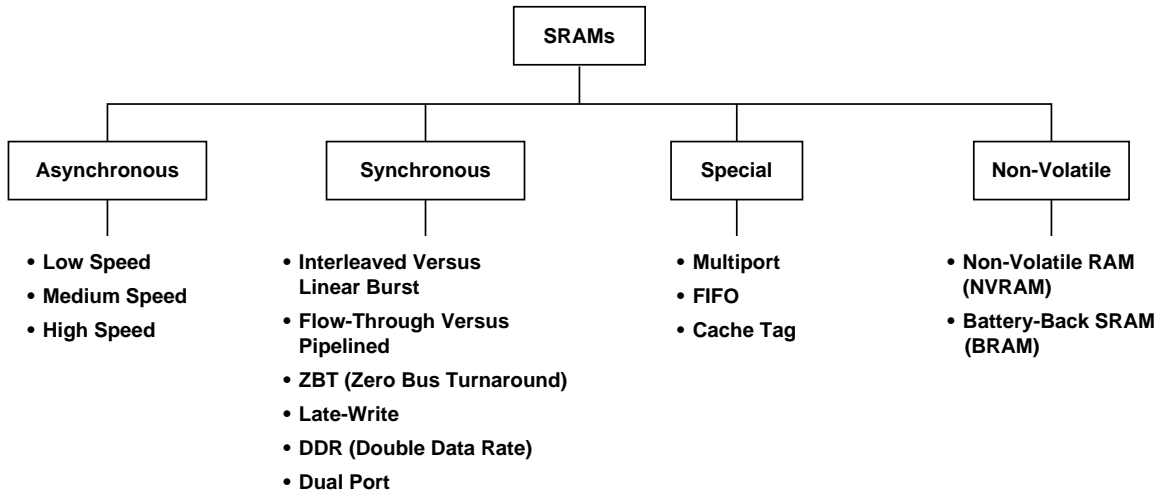
Source: ICE, "Memory 1997" 22454

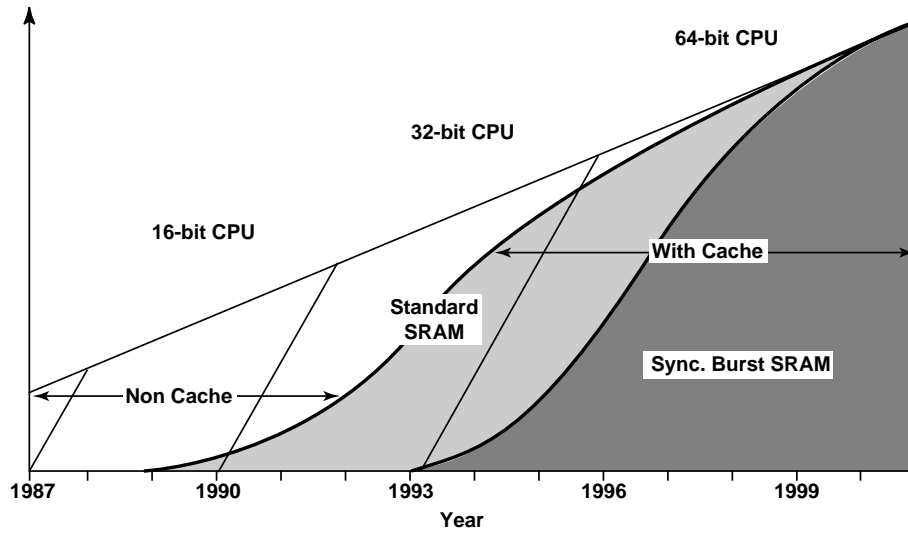**Figure 8-17. Overview of SRAM Types**

Logic Block Diagram

Pin Configuration
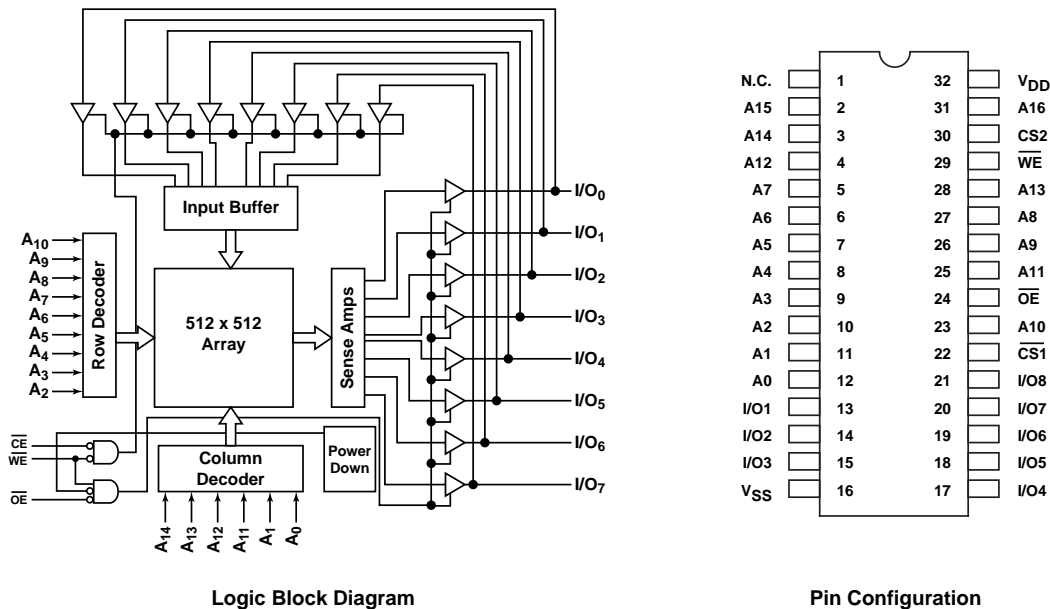
Source: Cypress/ICE, "Memory 1997" 22458
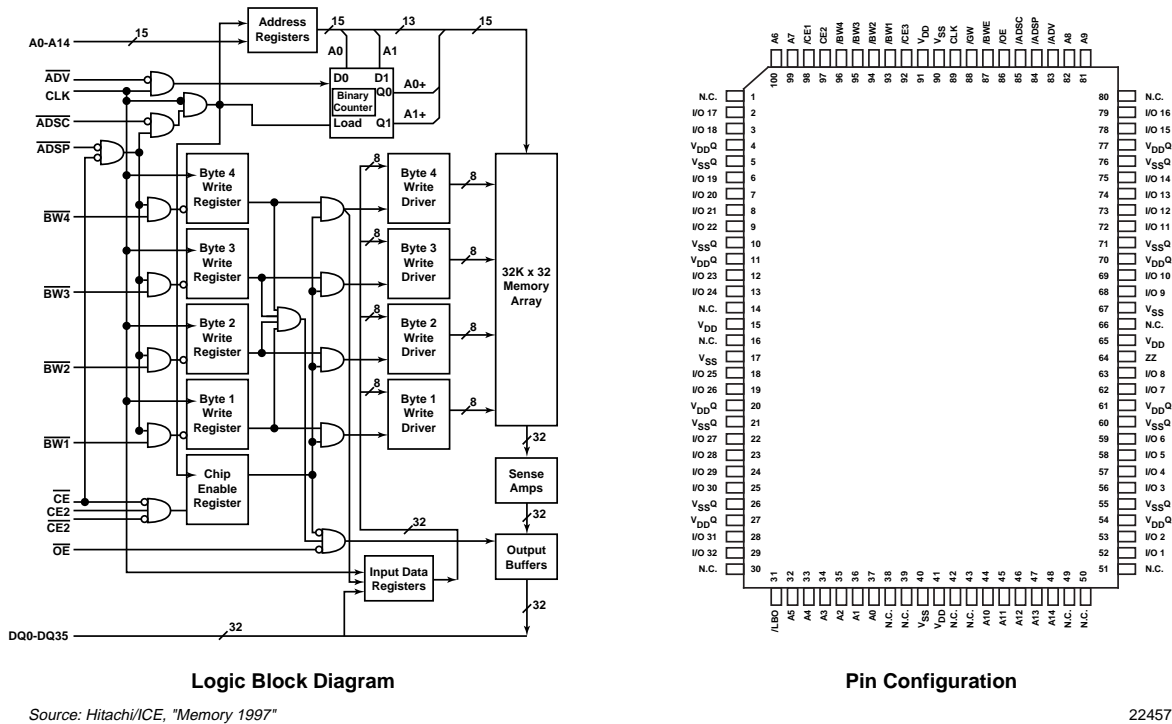
Figure 8-18. Typical SRAM

Flow-Through SRAM

Flow-through operation is accomplished by gating the output registers with the output clock. This dual clock operation provides control of the data out window.

Source: Mitsubishi/ICE, "Memory 1997"
20429A

Figure 8-19. Trend of PC Cache SRAM

Logic Block Diagram

Pin Configuration

Source: Hitachi/ICE, "Memory 1997"
22457

Figure 8-20. Typical SSRAM

SYNCHRONOUS MODE

CLOCK

Address

Output

BURST MODE
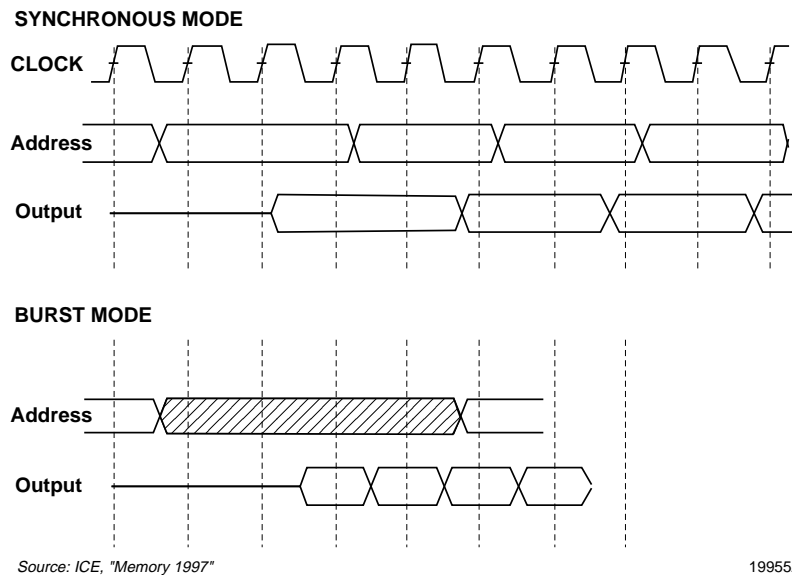
Address

Output

19955A

**Figure 8-21. SSRAM Timing**

Pipelined SRAMs

Pipelined SRAMs (sometimes called register to register mode SRAMs) add a register between the
memory array and the output. Pipelined SRAMs are less expensive than standard SRAMs for
equivalent electrical performance. The pipelined design does not require the aggressive manu-
facturing process of a standard SRAM, which contributes to its better overall yield. Figure 8-22
shows the architecture differences between a flow-through and a pipelined SRAM.

Figure 8-23 shows burst timing for both pipelined and standard SRAMs. With the pipelined
SRAM, a four-word burst read takes five clock cycles. With a standard synchronous SRAM, the
same four-word burst read takes four clock cycles.

Figure 8-24 shows the SRAM performance comparison of these same products. Above 66MHz,
pipelined SRAMs have an advantage by allowing single-cycle access for burst cycles after the first
read. However, pipelined SRAMs require a one-cycle delay when switching from reads to writes
in order to prevent bus contention.

Late-Write SRAM

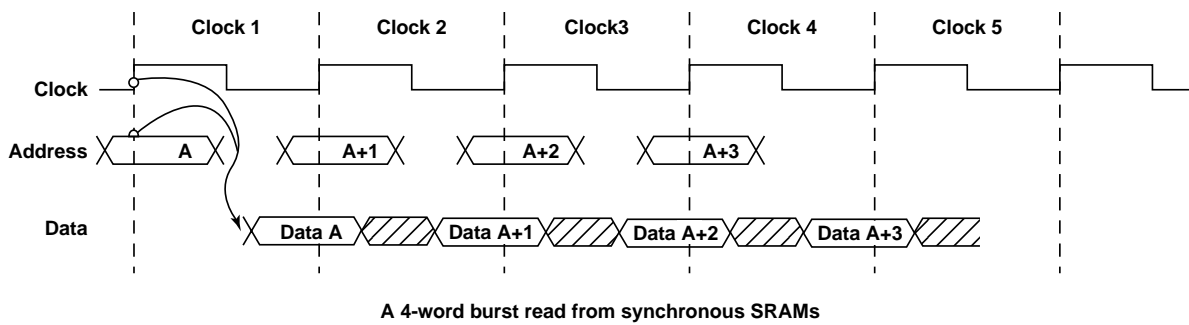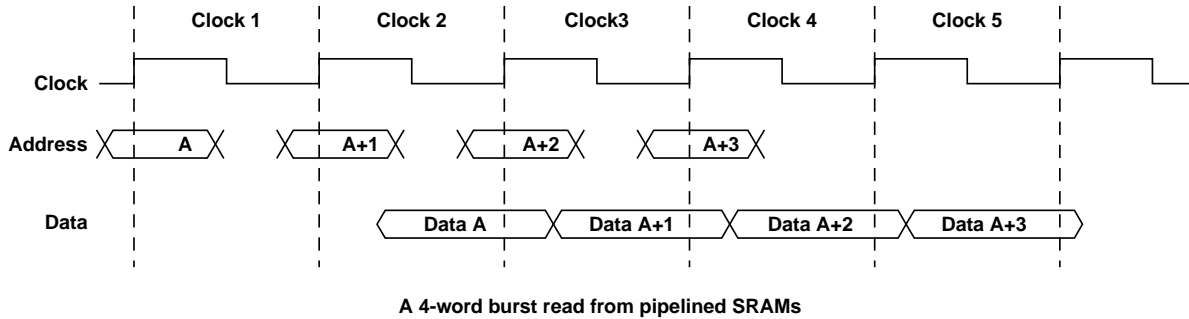Late-write SRAM requires the input data only at the end of the cycle.
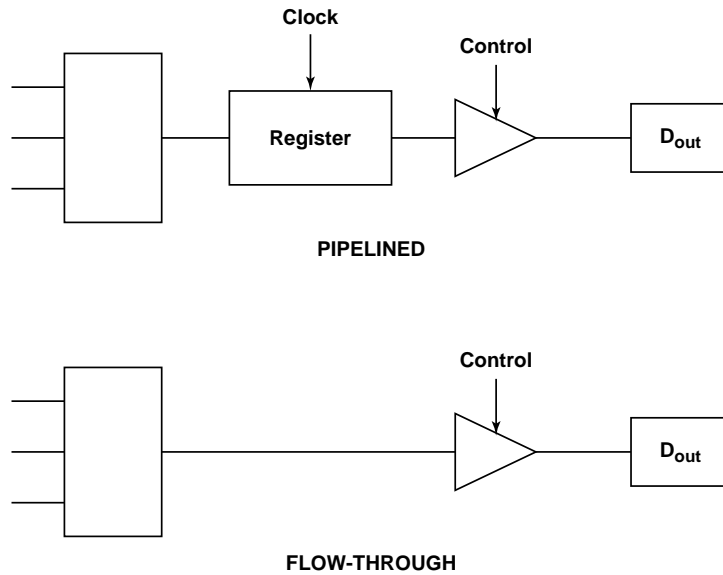
Source: ICE, "Memory 1997"   22608

Figure 8-22. Pipelined Versus Flow-Through Architectures

A 4-word burst read from pipelined SRAMs

A 4-word burst read from synchronous SRAMs

Source: Electronic Design/ICE, "Memory 1997"   20863

Figure 8-23. Pipelined Versus Non-Pipelined Timings

| Bus Frequency | 3.3V 32K x 8 | | | | 32K x 32 Pipelined | | | 32K x 32 Non-Pipelined | | | |
| | Speed (ns) | Banks | Performance | | Cycle Time | Performance | | Access Time | Cycle Time | Performance | |
			Read	Write		Read	Write			Read	Write
50	20	1	3-2-2-2	4-2-2-2	20	3-1-1-1	2-1-1-1	12	20	2-1-1-1	2-1-1-1
60	15	1 2	3-3-3-3 3-2-2-2	4-3-3-3 4-2-2-2	16.7	3-1-1-1	2-1-1-1	10	16.7	2-1-1-1	2-1-1-1
66	12 15	1 2	3-3-3-3 3-2-2-2	4-4-4-4 4-2-2-2	15	3-1-1-1	2-1-1-1	9	15	2-1-1-1	2-1-1-1
75	15	2	3-2-2-2	4-2-2-2	13.3	3-1-1-1	2-1-1-1	9	13.3	3-2-2-2	3-2-2-2
83	12	2	3-2-2-2	4-2-2-2	12	3-1-1-1	2-1-1-1	9	12	3-2-2-2	3-2-2-2
100	10	2	3-2-2-2	4-2-2-2	10	3-1-1-1	2-1-1-1	9	10	3-2-2-2	3-2-2-2
125	8	2	3-2-2-2	4-2-2-2	8	3-1-1-1	2-1-1-1	9	8	3-2-2-2	3-2-2-2

Source: Micron/ICE, "Memory 1997"   20864

Figure 8-24. SRAM Performance Comparison

ZBT (Zero Bus Turn-around)

The ZBT (zero bus turn-around) is designed to eliminate dead cycles when turning the bus around between read and writes and reads. Figure 8-25 shows a bandwidth comparison between the PBSRAM (pipelined burst SRAM), the late-write SRAM and the ZBT SRAM architectures.

| SRAM	Device Configuration	Clock Speed (MHz)	Bus Utilization	Bandwidth (Mbytes/sec)
PBSRAM	128K x 36 bits	100	50%	200
Late-Write SRAM	128K x 36 bits	100	67%	268
ZBT SRAM	128K x 36 bits	100	100%	400

Source: ICE, "Memory 1997"   22609

Figure 8-25. SSRAM Bandwidth Comparison

DDR (Double Data Rate) SRAMs

DDR SRAMs boost the performance of the device by transferring data on both edges of the clock.

Cache Tag RAMs

The implementation of cache memory requires the use of special circuits that keep track of which data is in both the SRAM cache memory and the main memory (DRAM). This function acts like a directory that tells the CPU what is or is not in cache. The directory function can be designed with standard logic components plus small (and very fast) SRAM chips for the data storage. An alternative is the use of special memory chips called cache tag RAMs, which perform the entire function. Figure 8-26 shows both the cache tag RAM and the cache buffer RAM along with the main memory and the CPU (processor). As processor speeds increase, the demands on cache tag and buffer chips increase as well. Figure 8-27 shows the internal block diagram of a cache-tag SRAM.
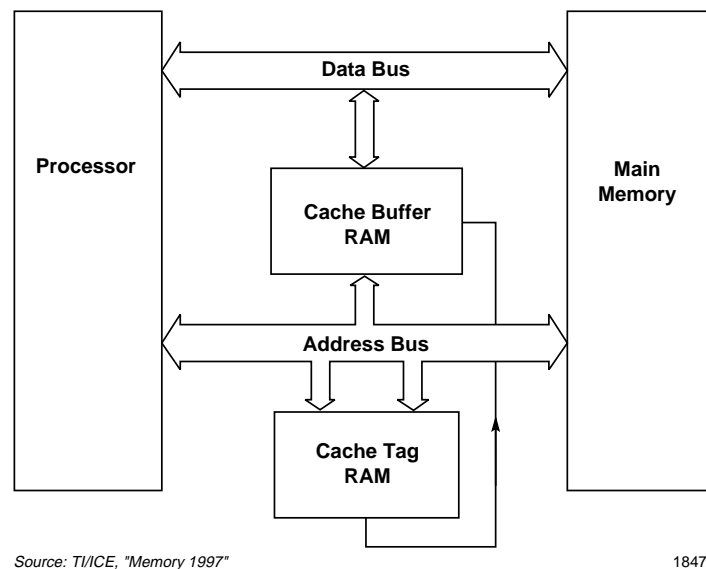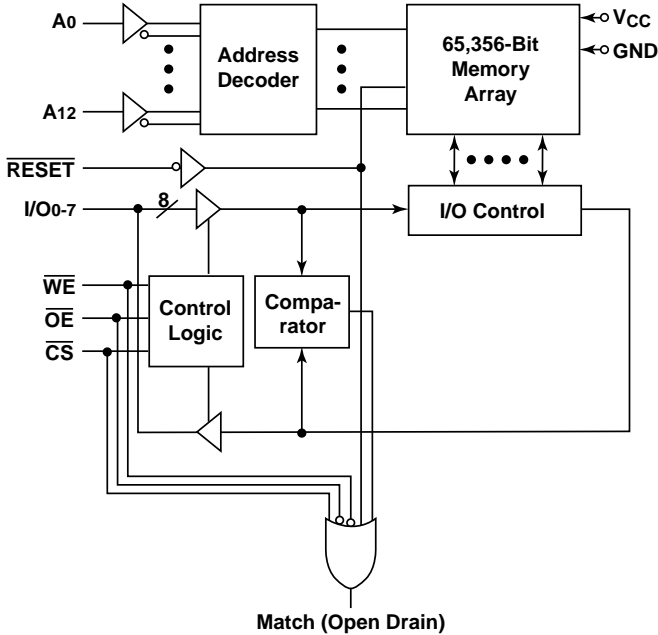

Source: TI/ICE, "Memory 1997"   18472

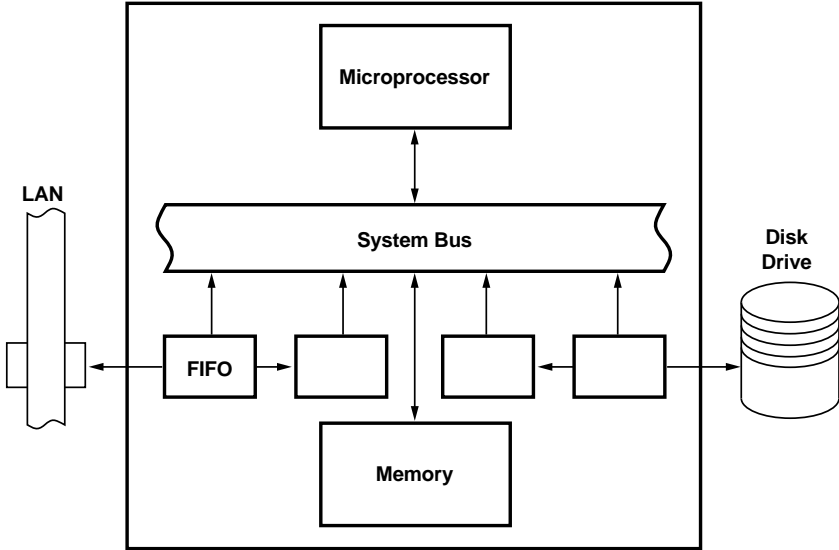**Figure 8-26. Typical Memory System With Cache**

FIFO SRAMs

A FIFO (first in, first out) memory is a specialized memory used for temporary storage, which aids in the timing of non-synchronized events. A good example of this is the interface between a computer system and a Local Area Network (LAN). Figure 8-28 shows the interface between a computer system and a LAN using a FIFO memory to buffer the data.

Synchronous and asynchronous FIFOs are available. Figures 8-29 and 8-30 show the block diagrams of these two configurations. Asynchronous FIFOs encounter some problems when used in high-speed systems. One problem is that the read and write clock signals must often be specially shaped to achieve high performance. Another problem is the asynchronous nature of the flags. A synchronous FIFO is made by combining an asynchronous FIFO with registers. For an equivalent level of technology, synchronous FIFOs will be faster.
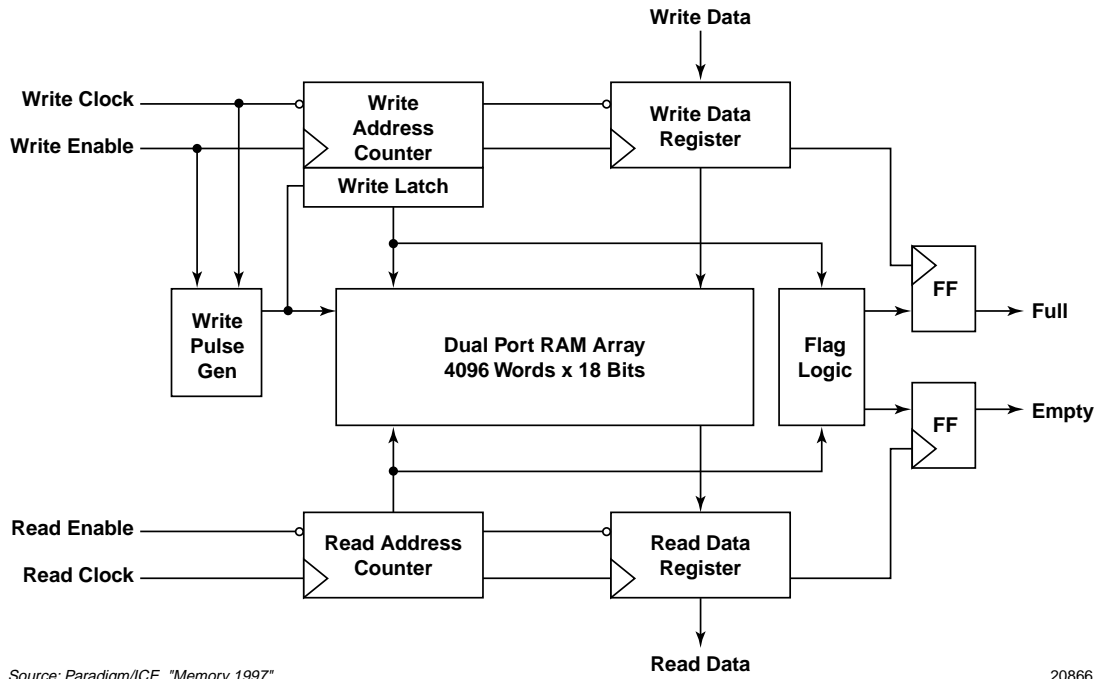
Match (Open Drain)

20865

Figure 8-27. Block Diagram of Cache-Tag SRAM

18804

Figure 8-28. FIFO Memory Solution for File Servers

Figure 8-29. Synchronous FIFO Block Diagram
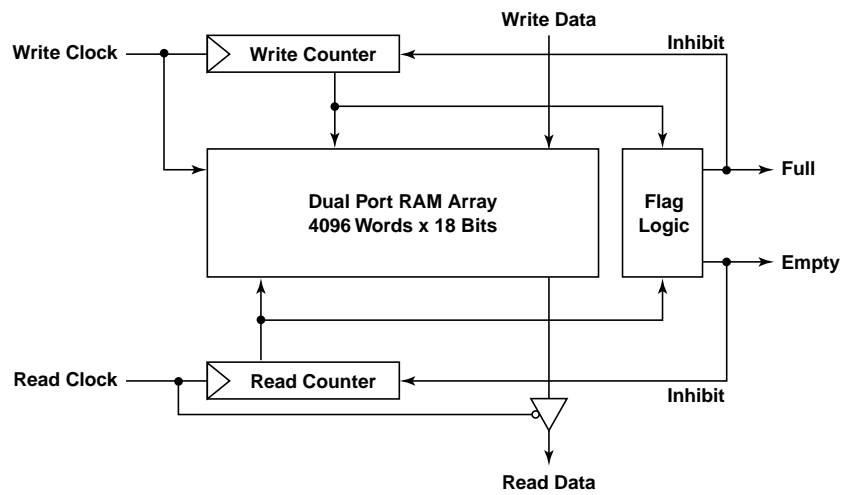
Figure 8-30. Asynchronous FIFO Block Diagram

Multiport SRAMs

Multiport fast SRAMs (usually two port, but sometimes four port) are specially designed chips using fast SRAM memory cells, but with special on-chip circuitry that allows multiple ports (paths) to access the same data at the same time.

INTEGRATED CIRCUIT ENGINEERING CORPORATION

Figure 8-31 shows such an application with four CPUs sharing a single memory. Each cell in the memory uses an additional six transistors to allow the four CPUs to access the data, (i.e., a 10T cell in place of a 4T cell). Figure 8-32 shows the block diagram of a 4-port SRAM.
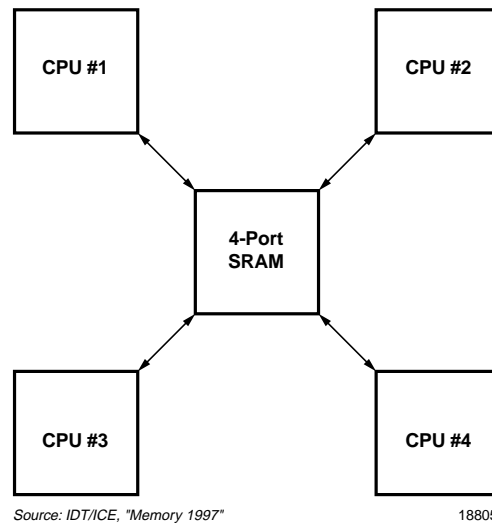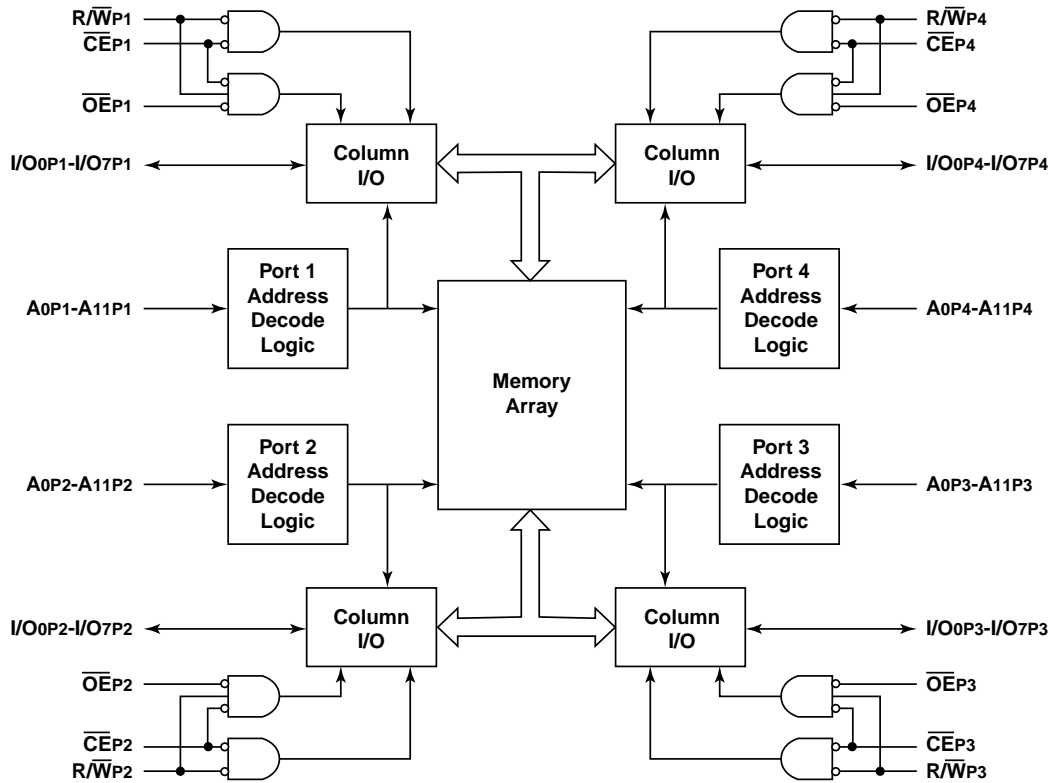

Source: IDT/ICE, "Memory 1997"   18805

Figure 8-31. Shared Memory Using 4-Port SRAM

Shadow RAMs

Shadow RAMs, also called NOVROMs, NVRAMs, or NVSRAMs, integrate SRAM and EEPROM technologies on the same chip. In normal operation, the CPU will read and write data to the SRAM. This will take place at normal memory speeds. However, if the shadow RAM detects that a power failure is beginning, the special circuits on the chip will quickly (in a few milliseconds) copy the data from the SRAM section to the EEPROM section of the chip, thus preserving the data. When power is restored, the data is copied from the EEPROM back to the SRAM, and operations can continue as if there was no interruption. Figure 8-33 shows the schematic of one of these devices. Shadow RAMs have low densities, since SRAM and EEPROM are on the same chip.

Battery-Backed SRAMs

SRAMs can be designed to have a sleep mode where the data is retained while the power consumption is very low. One such device is the battery-backed SRAM, which features a small battery in the SRAM package. Battery-backed SRAMs (BRAMs), also called zero-power SRAMs, combine an SRAM and a small lithium battery. BRAMs can be very cost effective, with retention times greater than five years. Notebook and laptop computers have this "sleep" feature, but utilize the regular system battery for SRAM backup.

Figure 8-32. Block Diagram of a 4-Port DRAM

Figure 8-34 shows a typical BRAM block diagram. A control circuit monitors the single 5V power supply. When V_{CC} is out of tolerance, the circuit write protects the SRAM. When V_{CC} falls below approximately 3V, the control circuit connects the battery which maintains data and clock operation until valid power returns.

RELIABILITY CONCERNS

For power consumption purposes, designers have reduced the load currents in the 4T cell structures by raising the value of the load resistance. As a result, the energy required to switch the cell to the opposite state is decreased. This, in turn, has made the devices more sensitive to alpha particle radiation (soft error). The TFT cell reduces this susceptibility, as the active load has a low resistance when the TFT is "on," and a much higher resistance when the TFT is "off." Due to process complexity, the TFT design is not widely used today.
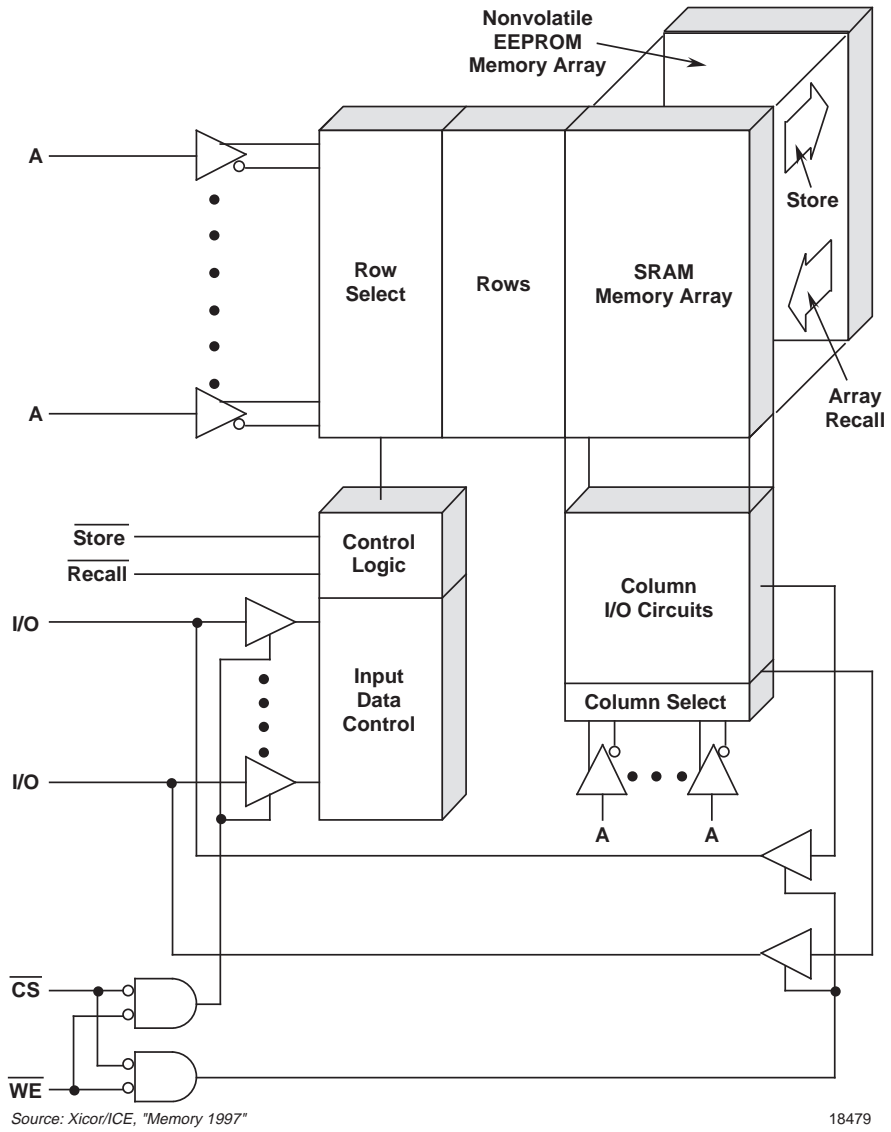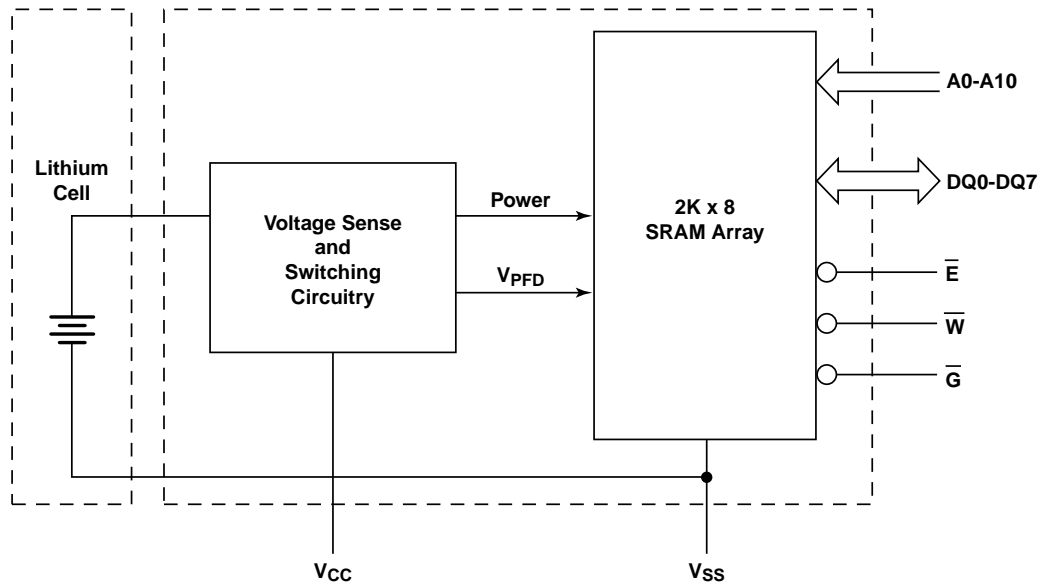
Figure 8-33. Block Diagram of the Xicor NOVRAM Family

Source: SGS-Thomson/ICE, "Memory 1997"   20831A

Figure 8-34. Block Diagram of a Typical BRAM