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 provides cache memory, interfacing between DRAMs and the CPU. An SRAM can be accessed in as little as a few nanoseconds (ns), versus 50ns to 80ns for a DRAM.

The second application — low power — is found in most portable equipment where the DRAM refresh current is several orders of magnitude more than the low-power SRAM standby current. In the low-power parts, the access time is comparable to a DRAM.

How the Device Works

The SRAM cell consists of a bi-stable flip flop connected to two access transistors (Figure 8-1). The data is latched into the flip flop.

![SRAM Cell Diagram](source: ICE, "Memory 1996"

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.
Figure 8-2 shows the read/write operations of an SRAM. To select a cell, the word line is set to Vcc (X address). The B and B (bit lines) are connected to the sense amplifier or the write circuitry (depending on whether the device is in a read or a write mode) by the column decode transistors (Y address). In a read mode, the cell data is applied to the sense amplifier that recognizes the data (voltage comparator). In a write mode, the write circuitry forces the data onto the cell as the write circuitry drivers are stronger than the cell flip flop transistors.

MEMORY CELL

4T Cell

The most common SRAM cell consists of four transistors and two poly-load resistors (Figure 8-3). This design is called the 4T cell SRAM. The storage cell of a 4T cell SRAM is about four times as large as the cell of a comparable generation DRAM and contains four transistors.
4T cells have several limitations:

- Each cell has a current flowing in one resistor, i.e., high standby current.
- The cell is sensitive to noise and soft error rate because the resistors are so large.
- The cell is not the fastest cell design available.

6T Cell

A different cell design that eliminates these limitations is the 6T cell. This SRAM cell is composed of six transistors, two NMOS and two PMOS transistors connected as a flip flop, and two NMOS access transistors. This structure is shown in Figure 8-4. It offers better performance (speed, noise immunity, standby current) than a 4T structure. These devices are used when extremely low power consumption is mandatory, such as in palm top computers operating from AA batteries.

TFT (Thin Film Transistor) Cell

A mix of the 4T cell and the 6T cell structures has been developed. The load is formed by using polysilicon as a PMOS device. This PMOS transistor is called a Thin Film Transistor (TFT), and it is formed by depositing several layers of polysilicon above the silicon surface.
The performance of this TFT PMOS transistor is not as good as a standard PMOS silicon transistor used in a 6T cell. It is more realistically compared to the linear polysilicon resistor characteristics, as TFT cell technology is more an improvement and replacement of the 4T technology than the 6T technology. Process and cell size are closer to a 4T cell technology than a 6T cell technology because the area of the TFT transistors is above the NMOS transistors.

Figure 8-5 shows the TFT characteristics. In actual use, the effective resistance would range from about $11 \times 10^{13}\Omega$ to $5 \times 10^{9}\Omega$. Figure 8-6 shows the TFT cell schematic.
Figure 8-7 shows a cross section drawing of the TFT cell. The TFT technology requires the deposition of two more films and at least three more photolithography steps.

Figure 8-7. Cross Section of TFT SRAM Cell
Cell Size and Die Size

SRAM performance targets have a dramatic effect on cell size. Figure 8-8 shows cell sizes and other characteristics of SRAM parts analyzed in ICE’s laboratory in 1995. None of the devices analyzed made use of thin-film-transistor (TFT) pullups. Each supplier used the standard 4T cell using resistive pullups. It is interesting to note that the die size of the only 4Mbit part (Toshiba’s CMOS SRAM dated coded 9509) was larger than an NEC 4Mbit SRAM die analyzed in 1994. In fact, Toshiba’s 4Mbit cell size is actually larger than the cache SRAM on Intel’s Pentium microprocessor.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Die Size</th>
<th>Min Gate - (N)</th>
<th>Cell Pitch</th>
<th>Cell Area</th>
<th>Cell Type</th>
<th>VCC</th>
<th>Access Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOSHIBA TC554161FTL-70L 4Mb (x16) 9509</td>
<td>7.7 x 18.7mm (144mm²)</td>
<td>0.65µm</td>
<td>3.7 x 6.6µm (33mm²)</td>
<td>22µm²</td>
<td>4T</td>
<td>5V</td>
<td>70ns</td>
</tr>
<tr>
<td>SAMSUNG KM732V588 1Mb (x32) 1995</td>
<td>5 x 6.6mm (33mm²)</td>
<td>0.5µm</td>
<td>3.0 x 4.75µm</td>
<td>14.25µm²</td>
<td>4T</td>
<td>3.3V</td>
<td>—</td>
</tr>
<tr>
<td>GALVANTECH GVT7132C32Q7 1Mb (x32) 9524</td>
<td>4.5 x 6.8mm (31mm²)</td>
<td>0.4µm</td>
<td>3.4 x 4.9µm</td>
<td>16.5µm²</td>
<td>4T</td>
<td>3.3V</td>
<td>7ns</td>
</tr>
<tr>
<td>HITACHI HM67W1664JP-12 1Mb (x16) 9539</td>
<td>6.4 x 10.1mm (64mm²)</td>
<td>0.45µm</td>
<td>3.3 x 5.7µm</td>
<td>19µm²</td>
<td>4T</td>
<td>3.3V</td>
<td>12ns</td>
</tr>
<tr>
<td>NEC D461018LG5-A12 1Mb (x18) 9436</td>
<td>5.7 x 11.7mm (67mm²)</td>
<td>0.6µm</td>
<td>3.5 x 5.5µm</td>
<td>19µm²</td>
<td>4T</td>
<td>3.3V</td>
<td>12ns</td>
</tr>
<tr>
<td>MOTOROLA MCP67C618FN7 1Mb (x18) 9443</td>
<td>9.2 x 11.8mm (108mm²)</td>
<td>0.6µm</td>
<td>4.9 x 8.2µm</td>
<td>40µm²</td>
<td>4T</td>
<td>5V</td>
<td>7ns</td>
</tr>
</tbody>
</table>

Source: ICE, “1996 Successful Technologies Review”

Figure 8-8. Physical Geometries of SRAMs

Figure 8-9 shows the trends of the SRAM cell size. There is a tradeoff between the performance of the cell and the process complexity. Most manufacturers believe that the TFT-cell SRAM manufacturing process is too difficult, regardless of its advantages.

The 6T cell gives a better performance but has a much larger cell size. This cell will only be used in high performance and ultra low-power SRAMs. As we will see in Section 13 (embedded memories), the 6T cell is very common for microprocessor or on-chip cache applications. The design offers extremely high performance and the process is virtually identical to the microprocessor process. Cell size remains an issue, however, so some microprocessor manufacturers utilize the 4T cell. Figure 8-10 shows the field applications of SRAMs.
Figure 8-9. Trend of SRAM Cell Sizes

Figure 8-10. Application of DRAM/SRAM
CONFIGURATION

The SRAM can be classified in three main categories:

• Asynchronous SRAMs
  Low speed and high speed
• Synchronous SRAMs
  Standard and pipelined
• Special SRAMs
  Cache Tag, FIFO, and Multiports

Figure 8-11 shows the SRAM classification.

Asynchronous SRAMs

Figure 8-12 shows a functional block diagram of an asynchronous SRAM. The memory is controlled by three clocks:

- Chip Select (CS) that selects or de-selects the chip. When the chip is deselected, the part is stand-by mode (minimum current consumption) and the outputs are in a high impedance state.

- Output Enable (OE) that controls the outputs (valid data or high impedance).

- Write Enable (WE) that allows a read or a write cycle.
Low-Speed SRAM

Low-speed SRAM devices offer low power and low cost. The main applications are high density hard disk drive (HDD) and industrial control systems. The speed of these devices is greater than 55ns.

High-Speed SRAM

Definitions vary, but high-speed SRAM devices have access time generally ranging from 35ns to 15ns. Applications are local storage in telecommunications systems and cache memory in computer systems.

Synchronous SRAM

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

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 growing application is the cache SRAM used in Pentium- or PowerPC-based PCs and workstations.
Figure 8-13 shows the configuration of an SSRAM. 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 unnecessary to hold the address and write data state throughout the entire cycle.

**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. These devices are offered in a wide word organization such as a 1Mbit device in a 32Kbit x 32 organization. Figure 8-14 shows 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 performances. The pipelined design does not require the aggressive manufacturing process of a standard SRAM, which contributes to its better overall yield.

Figure 8-15 shows a 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 8-16 shows the SRAM performance comparison of these same products. Above 66 MHz, 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.

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 8-17 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 8-18 shows the internal block diagram of a cache-tag SRAM.

**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 8-19 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 8-20 and 8-21 shows 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.
## SRAM Performance Comparison

<table>
<thead>
<tr>
<th>Bus Frequency (ns)</th>
<th>3.3V 32K x 8</th>
<th>32K x 32 Pipelined</th>
<th>32K x 32 Non-Pipelined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance</td>
<td>Cycle Time</td>
<td>Performance</td>
</tr>
<tr>
<td></td>
<td>Read</td>
<td>Write</td>
<td>Read</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Micron

### Figure 8-16. SRAM Performance Comparison

### Typical Memory System With Cache

Source: TI
Figure 8-18. Block Diagram of Cache-Tag SRAM

Source: IDT

Figure 8-19. FIFO Memory Solutions for File Servers

Source: IDT
Figure 8-20. Synchronous FIFO Block Diagram

Figure 8-21. Asynchronous FIFO Block Diagram
Multiport SRAMs

Multiport fast SRAM (usually two port, but sometimes four port) memories 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.

Figure 8-22 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 8-23 shows the block diagram of a 4-port SRAM.

![Figure 8-22. Shared Memory Using 4-Port SRAM](source: IDT 18805)

**RELIABILITY CONCERNS**

SRAMs are susceptible to alpha particle radiation. As designs have reduced the load currents in the 4T cell structures by raising the value of the load resistance, the energy required to switch the cell to the opposite state (soft error) has been reduced. This, in turn, has made the devices more sensitive to radiation. 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 8-23. Block Diagram of a 4-Port DRAM