MEMORIES

Semiconductor memories are classified in different ways. A distinction is made between read-only (ROM) and read-write (RWM) memories. The contents RWMs can be changed in a short time for a virtually unlimited number of times and contents of ROMs are mostly useful for frequent reading and occasional writing. Since RWM memories use active circuitry (transistors) to store the information, they belong to the class of called volatile memories. This is because the data would be lost when the supply voltage is turned off. Read-only memories, on the other hand, encode information by the presence or absence of devices. Their data cannot be modified and they belong to the class of nonvolatile memories. That means the stored data is lost by the disconnection of supply voltage.

Table 1 : Classification Semiconductor Memories

RWM		**NVRWM**	**ROM**
Random Access	**Non Random Access**		
SRAM DRAM	FIFO Shift Register	EPROM E2PROM FLASH	Mask-programmed ROM Programmable ROM

Based on the access pattern, RWMs are classified as random access class and serial memories. FIFO (first-in-first-out) is an example for serial memories. Most memories belong to the random access class, which means memory locations can be read or written in random order. One would expect memories of this class to be called RAM (random access memory); nevertheless for historic reasons, RAM has been reserved for random access RWM memories. That means though most ROM units also provide random access, but the acronym RAM should not be used for them.

VOLATILE MEMORIES

Static Random Access Memory (SRAM) and Dynamic Random Access Memory (DRAM) are volatile memories. SRAM is used as a cache memory in computers since it offers the fastest write/read (~8ns) speed among all memories. Hardware design of a single SRAM cell consists of 6 transistors. A DRAM cell consists of one transistor and one capacitor and it is based on the charge stored in a capacitor. It is superior to SRAM because of its low cost per bit storage; nevertheless it is slower (` 50ns). In DRAM, the stored charge in the

capacitor can be maintained only for few milli-seconds and therefore, an extra hardware circuit is needed to periodically refresh the data periodically.

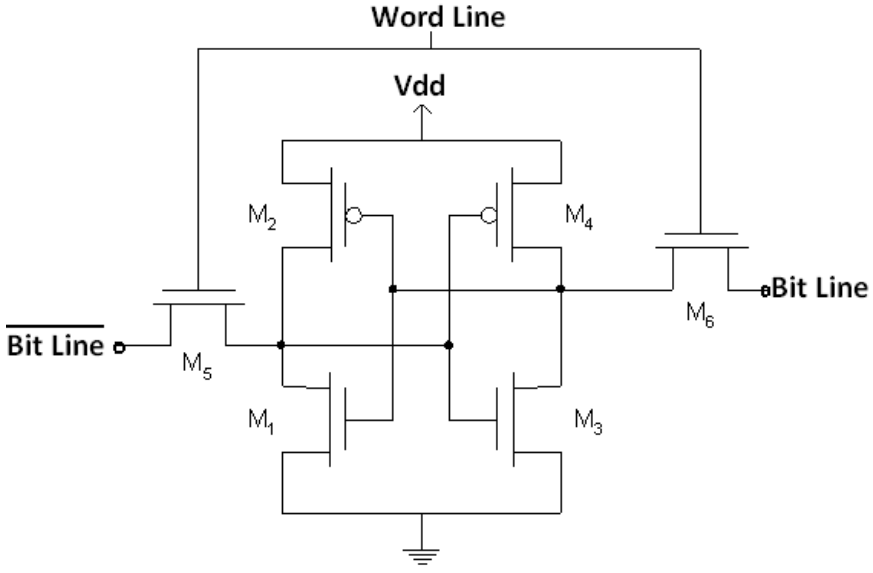
Static Random Access Memory (SRAM)

A single SRAM memory cell is shown in Fig. 5. Two NMOS and two PMOS transistors (M_1 to M_4) forms the simple latch to store the data and two pass NMOS transistors (M_5 and M_6) are controlled by Word Line to pass Bit Line and $\overline{\text{Bit Line}}$ into the cell.

A **Write** operation is performed by first charging the Bit Line and $\overline{\text{Bit Line}}$ with values that are desired to be stored in the memory cell. Setting the Word Line high performs the actual write operation, and the new data is latched into the circuit.

A **Read** operation is initiated by pre-charging both Bit Line and $\overline{\text{Bit Line}}$ to logic 1. Word Line is set high to close NMOS pass transistors to put the contents stored in the cell on the Bit Line and $\overline{\text{Bit Line}}$.

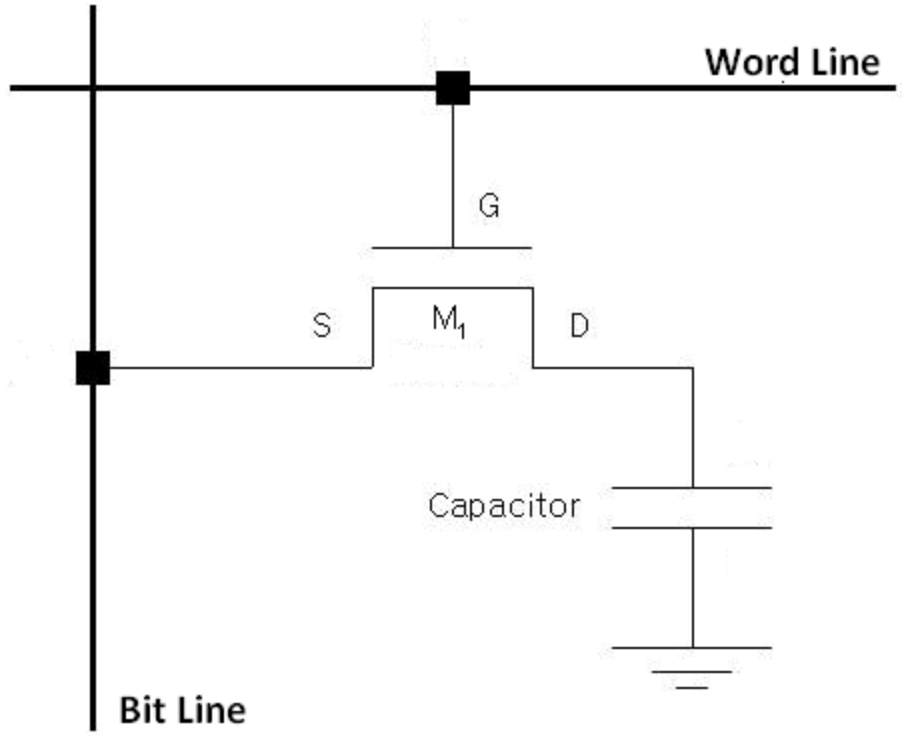
Transistors M_1 to M_4 constitute the latch and are constantly toggling back and forth. During these switching the power consumption in CMOS circuits takes place and therefore, the sizes of these transistors are kept as small as possible. NMOS transistors are basically switches opening and closing access to the SRAM cell. To minimize the propagation delay caused by these transistors their sizes are kept relatively larger.



Dynamic Random Access Memory (DRAM)

DRAM stores each bit in a storage cell consisting of a capacitor and a transistor. Capacitors tend to lose their charge rather quickly; thus, the need for recharging. The presence or absence of charge in the capacitor determines whether the cell contains a '1' or a '0'.

The **Read** operation begins by precharging the bit line to an intermediate value, $\frac{V_{DD}}{2}$. The word line is raised to a high potential and the charge stored on capacitor is shared with $\frac{V_{DD}}{2}$ that on the bit line. The change in the bit line voltage is given by the change on the bit line capacitor when the charge stored on capacitor C is shared with the bit line.



During the ' **Write 1** ' operation, the word line is driven high, the bit line is tied to V_{DD} corresponding to a 1, and voltage across C rises toward V_{DD} - V_{tn} . To ' **Write 0**' in the cell, the bit line is grounded during the write operation and 0V is stored on capacitor C.

NON-VOLATILE MEMORIES

Based on the programmability of the devices non-volatile memories are categorized as follows. Writing data into ROMs is possible only at the time of manufacturing the devices and used only for reading the data stored. Even though these devices are less in cost the constraint that they are to be programmed at the time of manufacturing is an inconvenience. PROM devices are one time programmable ROM. At the time of device manufacturing every cell is stored with "1" and can be programmed by customer once. But, single write phase makes them unattractive. For instance, a single error in the programming process or application makes the device unusable.

EPROM is Erasable PROM. Multiple times programming feature is added in EPROM. In this case, first whole memory is to be erased by shining ultraviolet light. The erase process is slow and can take from seconds to several minutes, depending on the intensity of the UV source. Programming takes several (5-10) μs /word. EPROM cell is extremely simple and dense, making it possible to fabricate large memories at a low cost. EPROMs were therefore attractive in applications that not require frequent programming. Electrically-Erasable PROM (EEPROM) can be erased without removing from board, unlike in UV erasable where memory must be removed from the board. The voltage approximately applied for programming is 18V. In addition, it is a reverse process; means by applying high negative voltage at gate can erase the cell. Another advantage over EPROM is that EEPROM can be programmed for 10^5 cycles.

Flash Electrically Erasable PROM Technically the Flash EEPROM is a combination of the EPROM and EEPROM approaches. The main difference is that erasure can be performed for the complete chip, or for a sub-section of the memory. The control circuits on the memory chip can be regularly checked for the value of the threshold during erasure, and the erasure time can be adjusted dynamically. Flash technology has three basic weaknesses. First, its bulk erase nature prevents the use of normal byte-oriented update. Second, Based on the architecture used write and erase operations take different time and consume more power than read. Finally, each flash-memory block has a limitation on the erase cycle count.

Although transistors are used for realization of Read Only Memory, the functioning of Rom can be easily understood by diode matrix network depicted in Fig. 6. In this network, whichever switch is closed, those diodes will conduct and the output will be high (logic 1), sections where there is no diode connected there will be no current flowing and the output will be low (logic 0). For instance, when switch S5 is closed, the diodes D6 and D7 are on

and therefore both output 1 and output 3 are at logic 1 and both output 2 and output 4 are at logic 0. Hence the corresponding binary number is 0101 and its decimal value is 5.

The disadvantage of a diode cell is that it does not isolate the bit line from the word line. For better isolation the diode can be replaced by gate-source connection of a NMOS transistor. Moreover, in order to achieve the programmability i.e. for multiple read write capability a modified transistor known as Floating Gate (FG) Transistor is employed. The structure is similar to a traditional MOS device, except that an extra gate is inserted between gate and channel. The threshold voltage of the FG is programmable and corresponding to its different values the level 0 and level 1 can be identified.

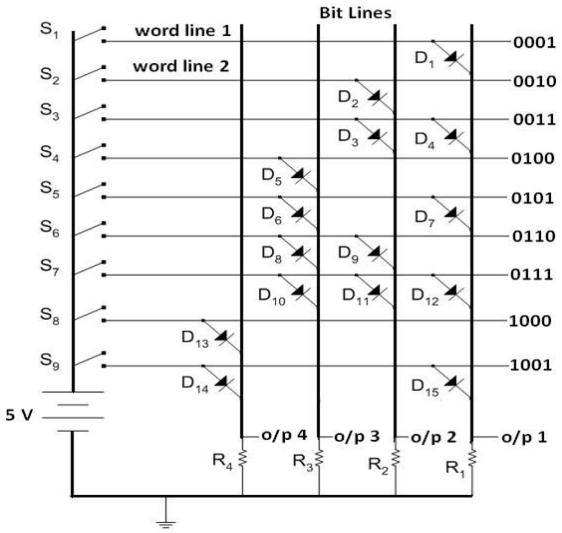


Fig. 6 : A Diode ROM Matrix

Flash memory cells can be arranged in two popular architectures; the NOR and the NAND architectures as explained in the following sections.

NOR ARCHITECTURE

Here every cell is connected in NOR fashioned manner as shown in Fig. 4(a). Note that the transistors used are FG type and two gates can be seen in their symbols. Every source terminal of the transistor is connected to ground in NOR architecture. Metal lines are required between each individual cell to run the ground in NOR architectures and therefore they occupy more area. NOR-based flash has long erase and write times, but has a full address/data (memory) interface that allows random access to any location. This makes it suitable for storage of program code that needs to be infrequently updated, such as computers' BIOS.

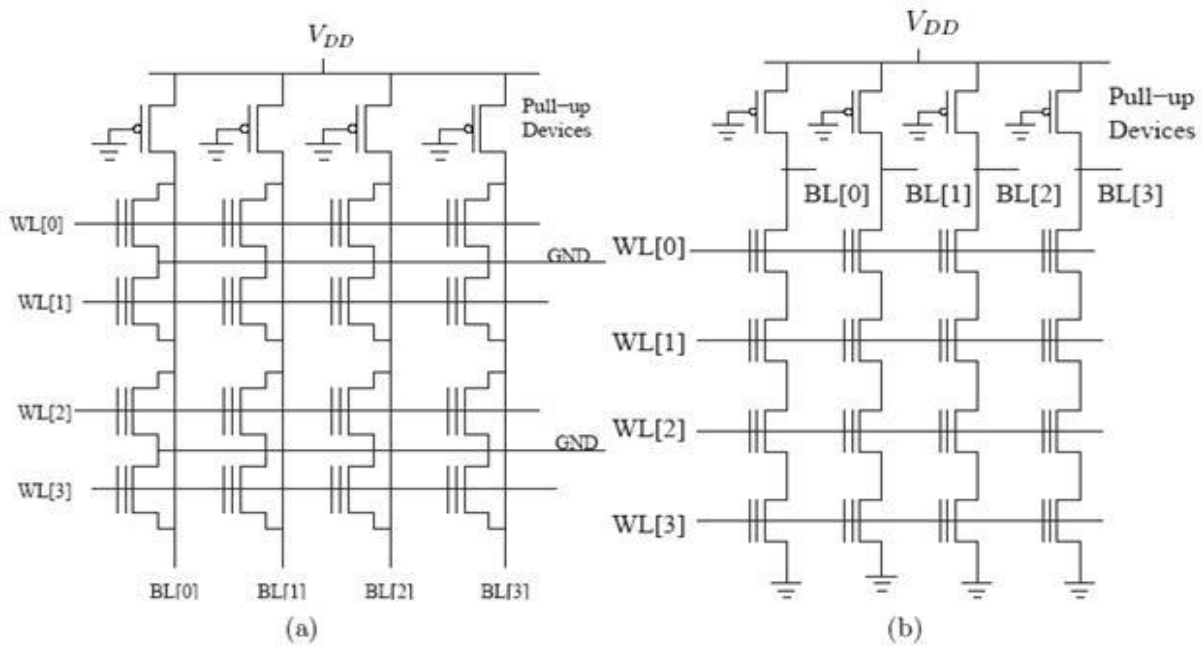


Figure 1.1: (a) NOR architecture (b) NAND architecue

NAND ARCHITECTURE

In the NAND structure, a series of floating gate transistors are connected between the bit line and ground line. This organization allows the elimination of all contacts to ground line and thus reducing the area by 40% compared to NOR architecture. It has faster erase and write times, higher density, and lower cost per bit than NOR flash. This can be obtained by arranging 8 to 16 floating gate transistors connected in series as shown in the Fig. 4(b). However, its I/O interface allows only sequential to data. This makes it suitable for mass-storage devices such as PC cards and various memory stick cards.