Memories

Sparsh Mittal

Background: Kilo vs Kibi

Multiples of bytes V · T · E				
SI decimal prefixes		Binary usage	IEC binary prefixes	
Name (Symbol)	Value		Name (Symbol)	Value
kilobyte (kB)	10^3	2^{10}	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	2^{20}	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	2^{30}	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	2^{40}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	2^{50}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	2^{60}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	2^{70}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	2^{80}	yobibyte (YiB)	2^{80}

RAM and ROM

There are two types of memories that are used in digital systems:
random-access memory (RAM) and read-only memory (ROM).

RAM stores new information for later use.

Storing new information into memory is referred to as a memory write operation.

The process of transferring the stored information out of memory is referred to as a memory read operation. RAM can perform both write and read operations.

ROM can perform only the read operation. This means that suitable binary information is already stored inside memory and can be retrieved or read at any time. However, that information cannot be altered by writing.

PLD

ROM is one example of programmable logic device (PLD).

The binary information that is stored within such a device is specified in some fashion and then embedded within the hardware by "programming" the device.

The word “programming” here refers to a hardware procedure which specifies the bits that are inserted into the hardware configuration of the device.

Other examples of PLD are programmable logic array (PLA), programmable array logic (PAL), and the field-programmable gate array (FPGA).

PLD programming

A PLD is an IC with internal logic gates connected through electronic paths that behave similarly to fuses.

In the original state of the device, all the fuses are intact.

Programming the device involves blowing those fuses along the paths that must be removed in order to obtain the particular configuration of the desired logic function.

Concise expression for array logic

(a) Conventional symbol (b) Array logic symbol

FIGURE 7.1
Conventional and array logic diagrams for OR gate

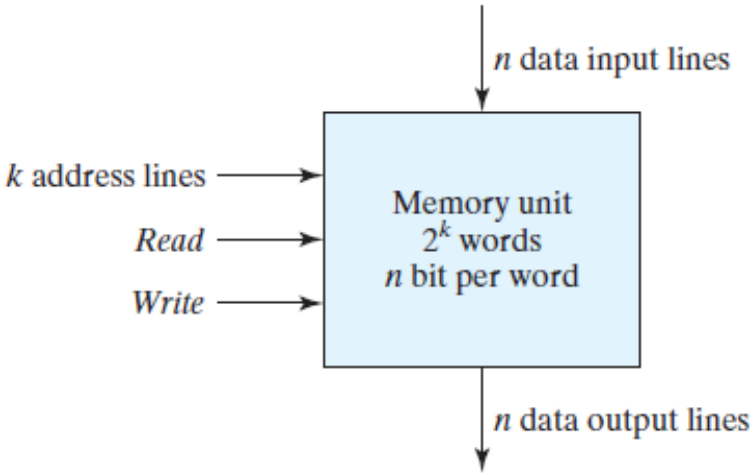
In a similar fashion, we can draw the array logic for an AND gate.

RAM

In a magnetic tape, the time required to retrieve information depends on the location of the data.

In RAM, information can be retrieved from any of its internal locations and retrieving information from any random location takes the same time.

The n data input lines provide the information to be stored in memory, and the n data output lines supply the information coming out of memory



RAM architecture

RAM Architecture

Each word in memory is assigned an identification number, called an *address*, starting from 0 up to 2^k - 1,

where k is the number of address lines.

The selection of a specific word inside memory is done by applying the k -bit address to the address lines.

An internal decoder accepts this address and opens the paths needed to select the word specified.

An Example of RAM

A memory with 1K words of
16 bits each
Total capacity = 2KiB

The words are recognized by their decimal address from 0 to 1,023. The equivalent binary address consists of 10 bits

Memory address		Memory content
Binary	Decimal	
0000000000	0	1011010101011101
0000000001	1	1010101110001001
0000000010	2	0000110101000110
⋮	⋮	⋮
1111111101	1021	1001110100010100
1111111110	1022	0000110100011110
1111111111	1023	1101111000100101

FIGURE 7.3
Contents of a 1024 × 16 memory

Write and Read Operations

Write

1. Apply the binary address of the desired word to the address lines.

2. Apply the data bits that must be stored in memory to the data input lines.

3. Activate the *write* input.

Read

1. Apply the binary address of the desired word to the address lines.

2. Activate the *read* input.

Table 7.1
Control Inputs to Memory Chip

Memory Enable	Read/Write	Memory Operation
0	X	None
1	0	Write to selected word
1	1	Read from selected word

Timing Waveforms

The CPU is usually synchronized by its own clock.

The memory does not employ an internal clock. Instead, its read and write operations are specified by control inputs.

Access time of memory is the time required to select a word and read it.

Cycle time of memory is the time required to complete a write operation.

The CPU must provide the memory control signals in such a way as to synchronize its internal clocked operations with the read and write operations of memory.

➔ The access time and cycle time of the memory must be within a time equal to a fixed number of CPU clock cycles.

Example

A CPU with clock frequency of 50 MHz ==> one cycle has a period of 20ns

This CPU communicates with a memory whose access time and cycle time do not exceed 50 ns.

==> Write cycle completes the storage of the selected word within a 50-ns interval and read cycle provides the output data of the selected word within 50 ns or less. (The two numbers are not always the same.)

Since the period of the CPU cycle is 20 ns, we must devote at least two-and-a-half, and possibly three, clock cycles for each memory request.

Write cycle

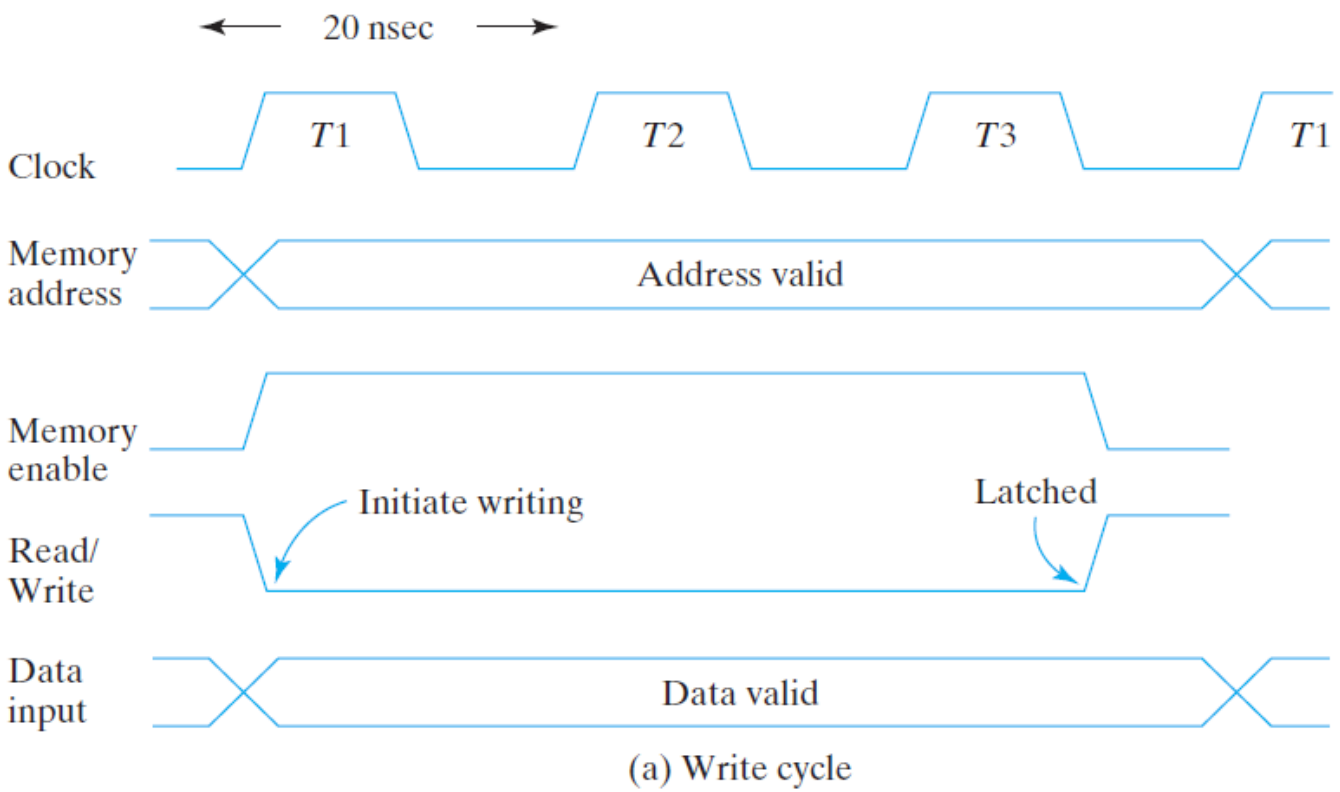
We have 50-MHz CPU and a memory with 50ns maximum cycle time.

There are three 20-ns cycles: T1, T2, and T3.

For a write operation, CPU must provide the address and input data to the memory.

This is done at the beginning of T1.

The memory enable and read/write signals must be activated after the signals in the address lines are stable in order to avoid destroying data in other memory words.



(a) Write cycle

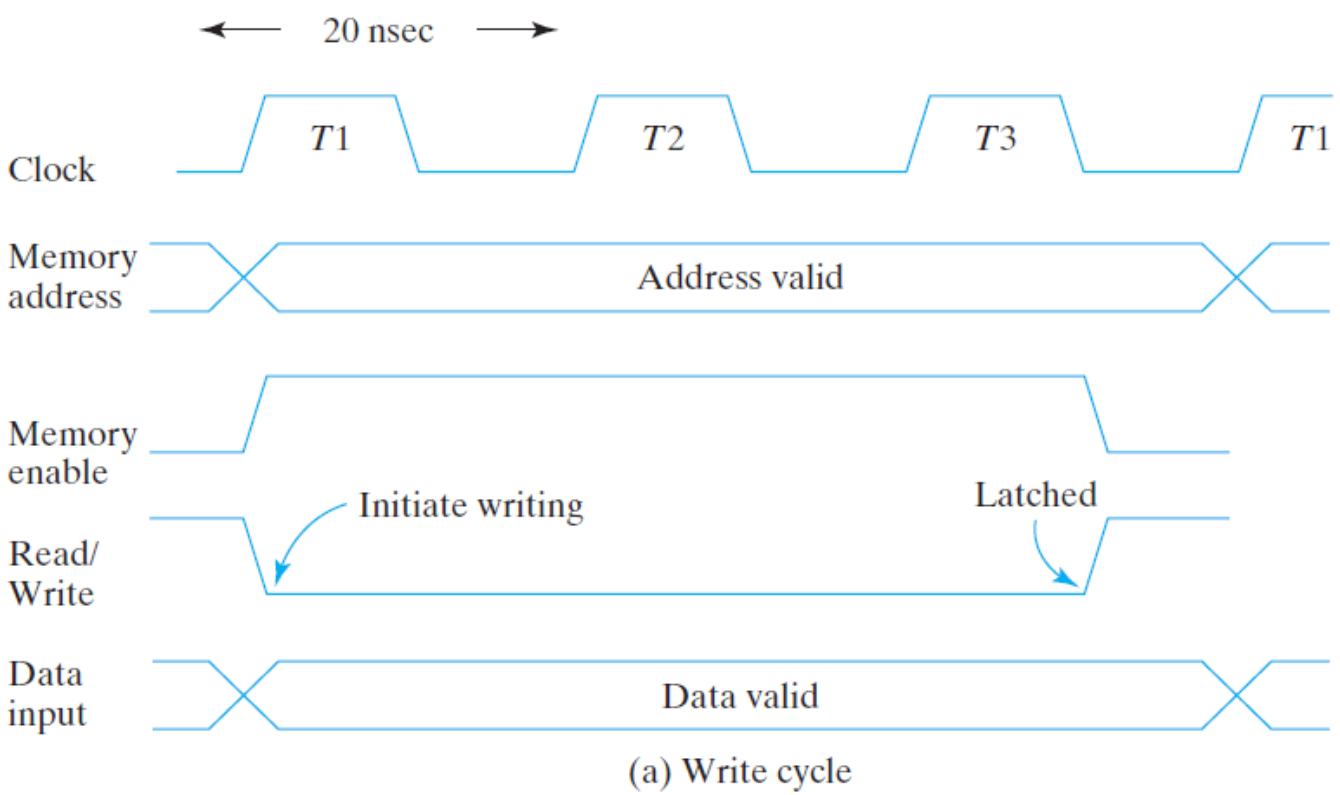
Write cycle

The memory enable signal switches to high level and read/write signal switches to the low level to indicate a write operation.

The two control signals must stay active for at least 50 ns.

The address and data signals must remain stable for a short time after the control signals are deactivated.

At the completion of the third clock cycle, the memory write operation is completed and the CPU can access the memory again with the next T1 cycle.



(a) Write cycle

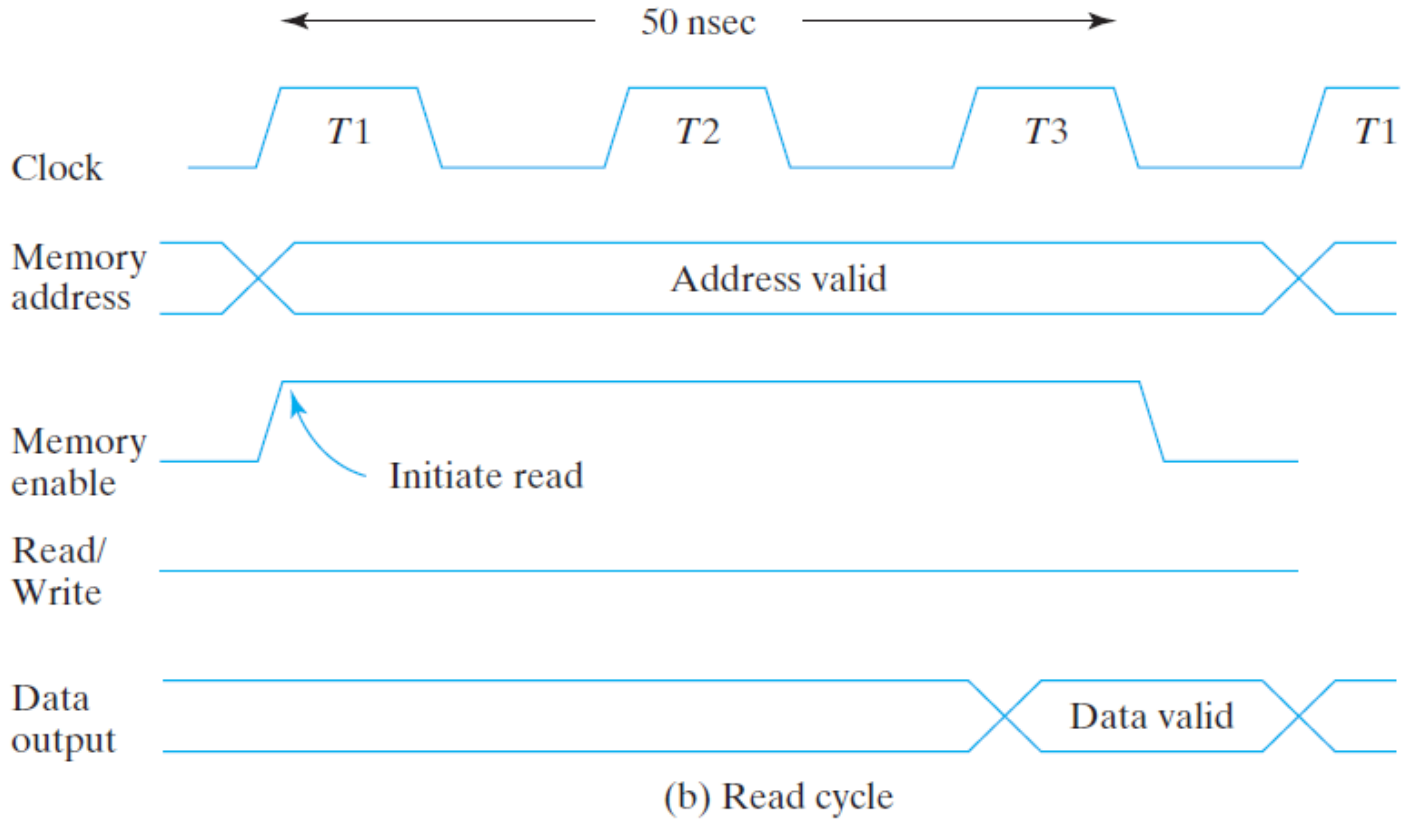
Read cycle

A memory address is provided by the CPU.

The memory-enable and read/write signals must be in their high level.

The memory places the data of the word selected by the address into the output data lines within a 50-ns interval from the time that the memory enable is activated.

The CPU can transfer the data into one of its internal registers during the negative transition of T3.



(b) Read cycle

SRAM and DRAM

SRAM and DRAM

- Static RAM (SRAM) stores data through internal latches.
- The stored information remains valid as long as power is applied to the unit.
- Dynamic RAM (DRAM) stores data in the form of electric charges on capacitors provided inside the chip by MOS transistors.
- The stored charge on the capacitors tends to discharge with time, and the capacitors must be periodically recharged by refreshing the dynamic memory.
- Refreshing is done by cycling through the words every few milliseconds to restore the decaying charge.
- DRAM offers reduced power consumption and larger storage capacity in a single memory chip.
- SRAM is easier to use and has shorter read and write cycles.

SRAM and DRAM

SRAM has 6 transistors (or more)

DRAM has 1 transistor and 1 capacitor.
Because of their simple cell structure, DRAMs typically have four times the density of SRAMs.

A further cost savings is realized because of the lower power requirement of DRAM cells.

Volatile and Non-volatile

Both SRAM and DRAM are volatile: they lose stored information when power is turned off.

Example of nonvolatile memory: magnetic tape and ROM.

A nonvolatile memory enables digital computers to store programs that will be needed again after the computer is turned on.

Programs and data that cannot be altered are stored in ROM, while other large programs are maintained on magnetic disks.

The latter programs are transferred into the computer RAM as needed.

Before the power is turned off, the binary information from the computer RAM is transferred to the disk so that the information will be retained.