Q1 Internal organization of RAM/memory chip/128bit memory chip?

- Memory-cells are organized in the form of array (Figure 8.2).
- Each cell is capable of storing 1-bit of information.
- Each row of cells forms a memory-word.
- All cells of a row are connected to a common line called as **Word-Line**.
- The cells in each column are connected to **Sense/Write** circuit by 2-bit-lines.
- The Sense/Write circuits are connected to data-input or output lines of the chip.
- During a write-operation, the sense/write circuit
  → receive input information &
  → store input info in the cells of the selected word.

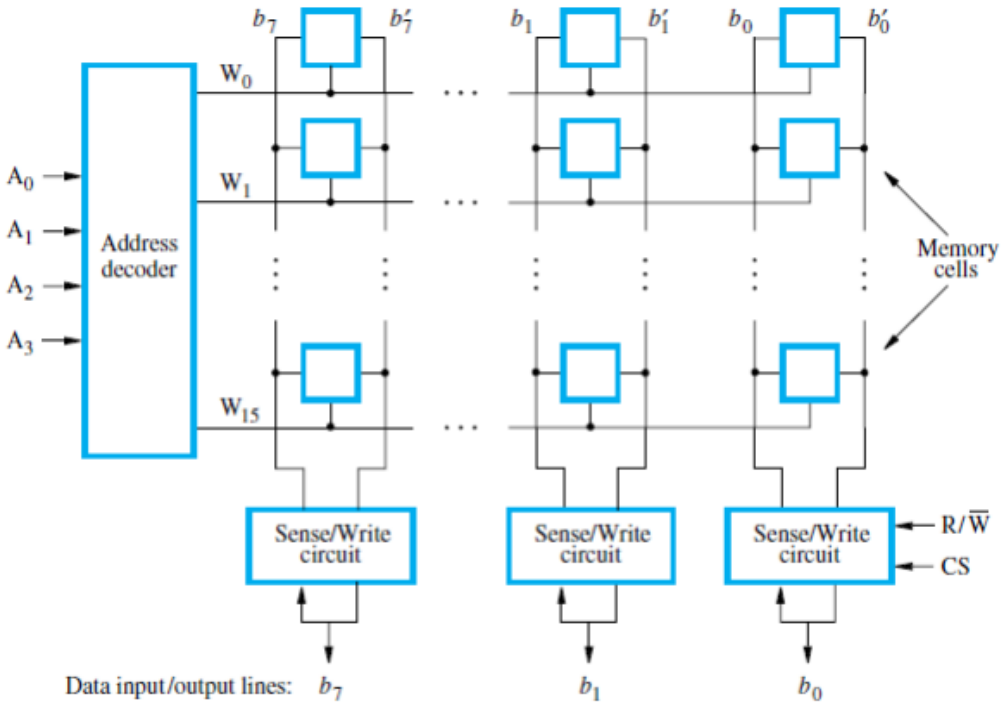


Figure 8.2 Organization of bit cells in a memory chip.

- The data-input and data-output of each Sense/Write circuit are connected to a single bidirectional data-line.
- Data-line can be connected to a data-bus of the computer.
- Following 2 control lines are also used:
  1) R/W' → Specifies the required operation.
  2) CS' → Chip Select input selects a given chip in the multi-chip memory-system.

Q2 Different Memory organization chip diagrams

Eg:1K*1K

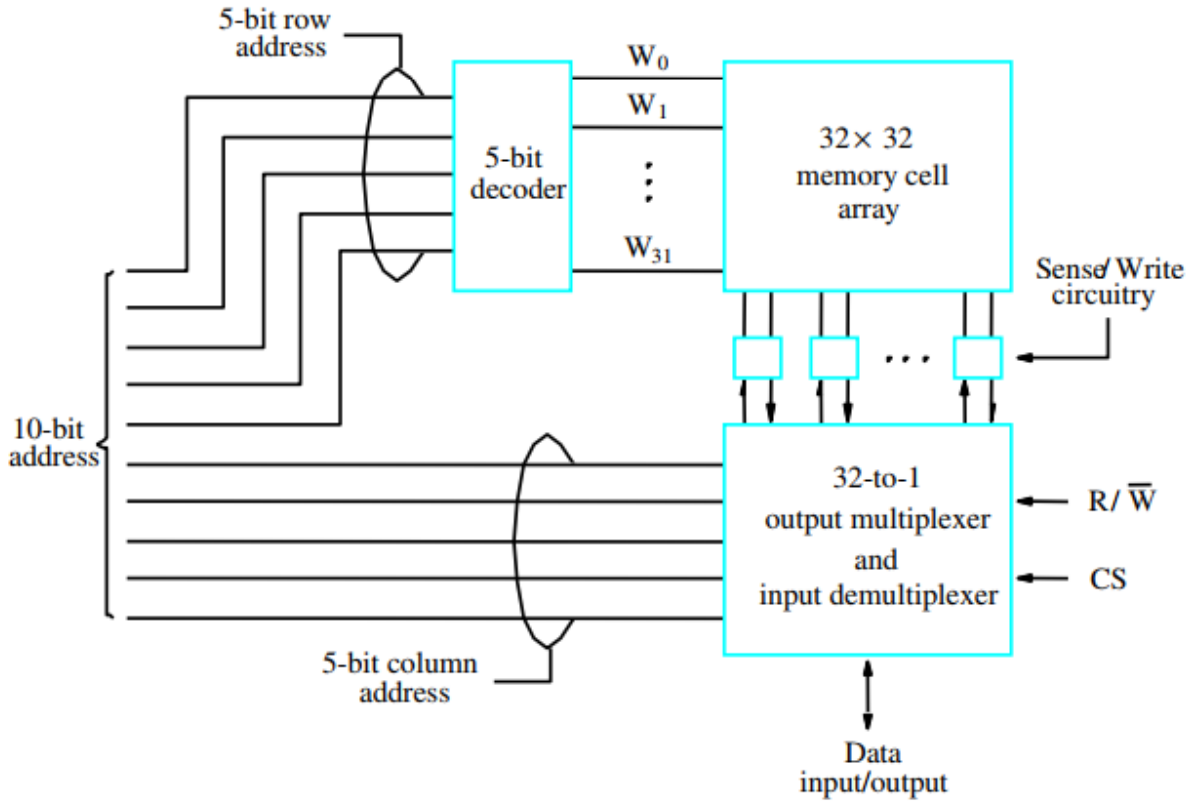


Figure 5.3. Organization of a 1K × 1 memory chip.

10- bit address line is needed, but there is only one data line resulting in 15 external connections. 10-bit address is divided into two groups of 5 bits each to form the row and the column addresses for the cell array. A row address selects a row of 32 cells, all of which are accessed in parallel. However, according to the column address, only one of these cells is connected to the external data line by output multiplexer and input demultiplexer.

Q3 Asynchronous and Synchronous DRAM

Asynchronous DRAM or 2M*8 Asynchronous DRAM

- Less expensive RAMs can be implemented if simple cells are used.
- Such cells cannot retain their state indefinitely. Hence they are called **Dynamic RAM (DRAM).**
- The information stored in a dynamic memory-cell in the form of a charge on a capacitor.
- This charge can be maintained only for tens of milliseconds.
- The contents must be periodically refreshed by restoring this capacitor charge to its full value.

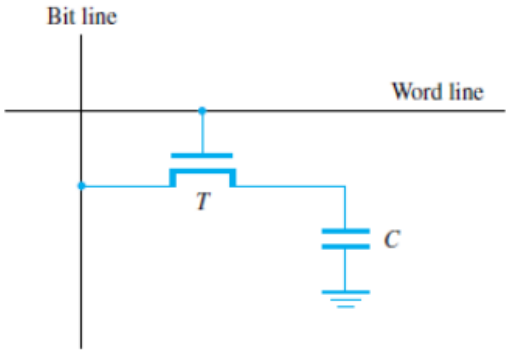


Figure 8.6 A single-transistor dynamic memory cell.

- In order to store information in the cell, the transistor T is turned 'ON' (Figure 8.6).
- The appropriate voltage is applied to the bit-line which charges the capacitor.
- After the transistor is turned off, the capacitor begins to discharge.
- Hence, info. stored in cell can be retrieved correctly before threshold value of capacitor drops down.
- During a read-operation,
  - → transistor is turned 'ON'
  - → a sense amplifier detects whether the charge on the capacitor is above the threshold value.
    - ➢ If (charge on capacitor) > (threshold value) → Bit-line will have logic value '1'.
    - ➢ If (charge on capacitor) < (threshold value) → Bit-line will set to logic value '0'.

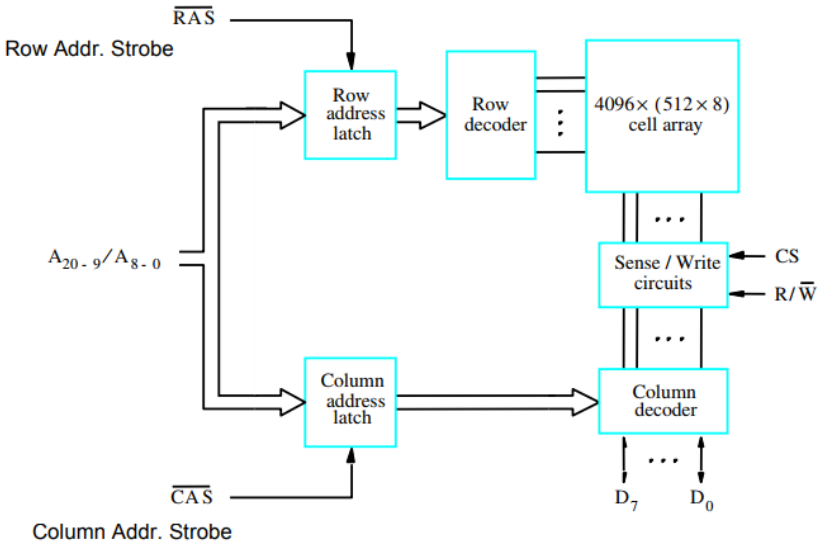


Figure 5.7. Internal organization of a 2M × 8 dynamic memory chip.

- During Read/Write-operation,
  - → row-address is applied first.
  - → row-address is loaded into row-latch in response to a signal pulse on **RAS'** input of chip. (RAS = Row-address Strobe CAS = Column-address Strobe)
- When a Read-operation is initiated, all cells on the selected row are read and refreshed.
- Shortly after the row-address is loaded, the column-address is

- 21 bit address is needed to access a byte in the memory. 21 bit is divided as follows:
  1) 12 address bits are needed to select a row.
     i.e. A_{8-0} → specifies row-address of a byte.
  2) 9 bits are needed to specify a group of 8 bits in the selected row.
     i.e. A_{20-9} → specifies column-address of a byte.

FAST PAGE MODE:
When DRAM in the above diagram is accessed, the contents of all 4096 cells in the selected row are sensed, but only 8 bits are placed on the data lines D_{7-0}, as selected by A_{8-0}. Fast page mode makes it possible to access the other bytes in the same row without having to reselect the row.
A latch is added at the output of the sense amplifier in each column.

Synchronous DRAM

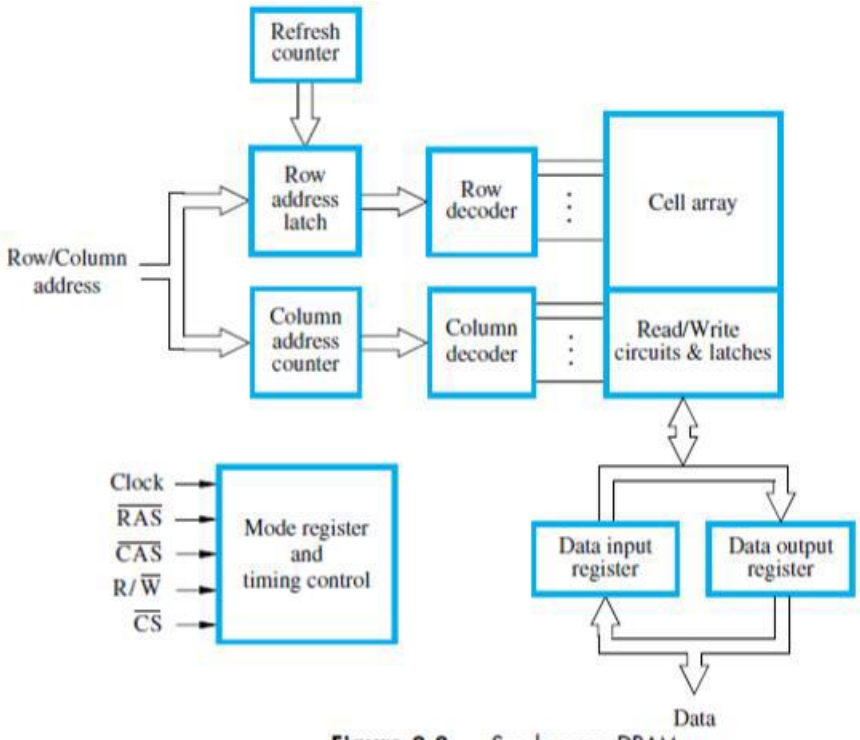


Figure 8.8 Synchronous DRAM.

The operations are directly synchronized with clock signal
 The address and data connections are buffered by means of registers
 The output of each sense amplifier is connected to a latch.
 During a Read operation, the contents of the cells in a row are loaded onto the latches.

 During a refresh operation, the contents of the cells are refreshed without changing the contents of the latches. Refresh counter provides the address of the rows that are selected for refreshing. (generally, each row must be refreshed at least every 64 ms)

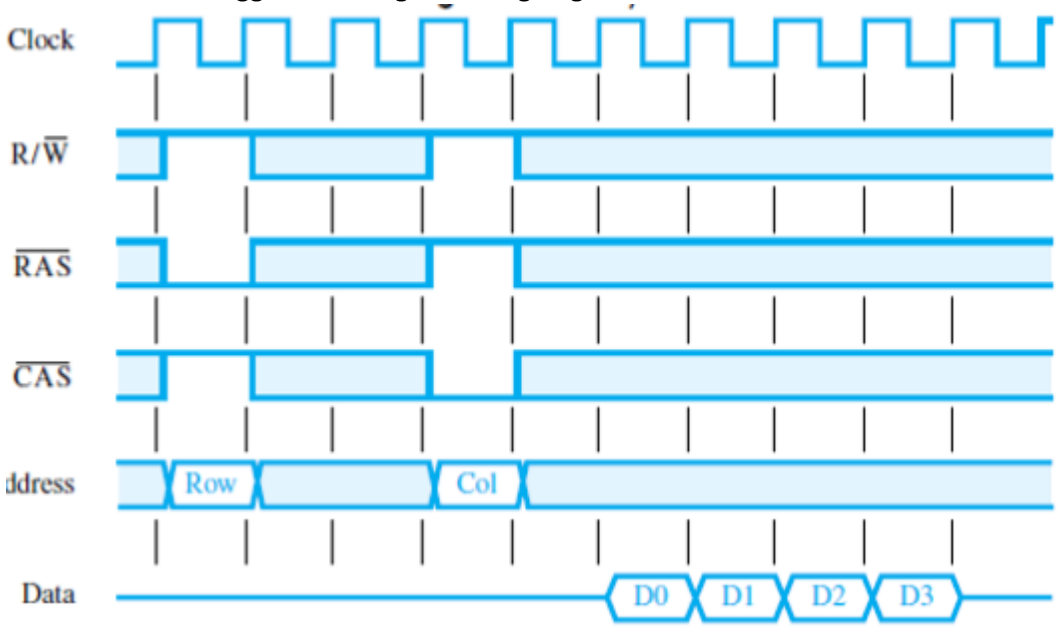
 Data held in latches that correspond to selected columns are transferred into data-output register. Thus, data becoming available on the data-output pins.

 SDRAMs have several different modes of operation, which can be selected by writing control information into a mode register.

 In SDRAMs, it is not necessary to provide externally generated pulses on CAS line to select successive columns, The necessary control signals are provided internally using a column counter and the clock signal.

 New data is placed on the data lines in each clock cycle.

 All actions are triggered during the rising edge of the clock.



Q4 Explain in detail Types of ROM?

READ-ONLY MEMORIES

 SRAM and SDRAM chips are volatile:

 Lose the contents when the power is turned off.

 Many applications need memory devices to retain contents after the power is turned off.

 For example, computer is turned on, the operating system must be loaded from the disk into the memory which is known as booting the OS.

 Memory is needed to store instructions which would load the OS from the disk which means instructions that would to load boot program into memory.

 These instructions should not be lost after the power is turned off.

 We need to store the instructions into a non-volatile memory.

 Non-volatile memory is read in the same manner as volatile memory.

 However, separate writing process is needed to place information in this memory

- **At Logic value '0'** → Transistor(T) is connected to the ground point (P).
  Transistor switch is closed & voltage on bit-line nearly drops to zero (Figure 8.11).
- **At Logic value '1'** → Transistor switch is open.
  The bit-line remains at high voltage.

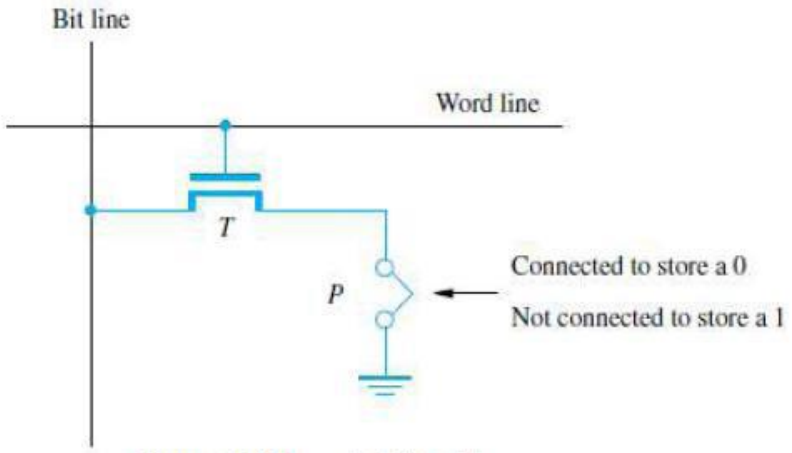


Figure 8.11 A ROM cell.

- To read the state of the cell, the word-line is activated.
- A Sense circuit at the end of the bit-line generates the proper output value.

PROM(Programmable Read Only Memory)

- PROM allows the data to be loaded by the user.
- Programmability is achieved by inserting a 'fuse' at point P in a ROM cell.
- Before PROM is programmed, the memory contains all 0's.
- User can insert 1's at required location by burning-out fuse using high current-pulse.
- This process is irreversible.
- **Advantages:**
  1) It provides flexibility.
  2) It is faster.
  3) It is less expensive because they can be programmed directly by the user.

EPROM (Erasable Reprogrammable Read Only Memory)

- EPROM allows
  → stored data to be erased and
  → new data to be loaded.
- In cell, a connection to ground is always made at 'P' and a special transistor is used.
- The transistor has the ability to function as
  → a normal transistor or
  → a disabled transistor that is always turned 'off'.
- Transistor can be programmed to behave as a permanently open switch, by injecting charge into it.
- Erasure requires dissipating the charges trapped in the transistor of memory-cells.
  This can be done by exposing the chip to ultra-violet light.
- **Advantages:**
  1) It provides flexibility during the development-phase of digital-system.
  2) It is capable of retaining the stored information for a long time.
- **Disadvantages:**
  1) The chip must be physically removed from the circuit for reprogramming.
  2) The entire contents need to be erased by UV light.

EEPROM (Electrically Erasable ROM)

A significant disadvantage of EPROMs is that a chip must be physically removed from the circuit for reprogramming and that its entire contents are erased by the ultraviolet light. It is possible to implement another version of erasable PROMs that can be both programmed and erased electrically. Such chips, called EEPROMs, do not have to be removed for erasure. Moreover, it is possible to erase the cell contents selectively. The only disadvantage of EEPROMs is that different voltages are needed for erasing, writing, and reading the stored data.

FLASH MEMORY

1) Flash Cards

- One way of constructing larger module is to mount flash-chips on a small card.
- Such flash-card have standard interface.
- The card is simply plugged into a conveniently accessible slot.
- Memory-size of the card can be 8, 32 or 64MB.
- Eg: A minute of music can be stored in 1MB of memory. Hence 64MB flash cards can store an hour of music.

2) Flash Drives

- Larger flash memory can be developed by replacing the hard disk-drive.
- The flash drives are designed to fully emulate the hard disk.
- The flash drives are solid state electronic devices that have no movable parts.

Advantages:

1) They have shorter seek & access time which results in faster response.
2) They have low power consumption. .'. they are attractive for battery driven application.
3) They are insensitive to vibration.

Disadvantages:

1) The capacity of flash drive (<1GB) is less than hard disk (>1GB).
2) It leads to higher cost per bit.
3) Flash memory will weaken after it has been written a number of times (typically at least 1 million times).

Q5 Memory hierarchy/Speed, Size, Cost

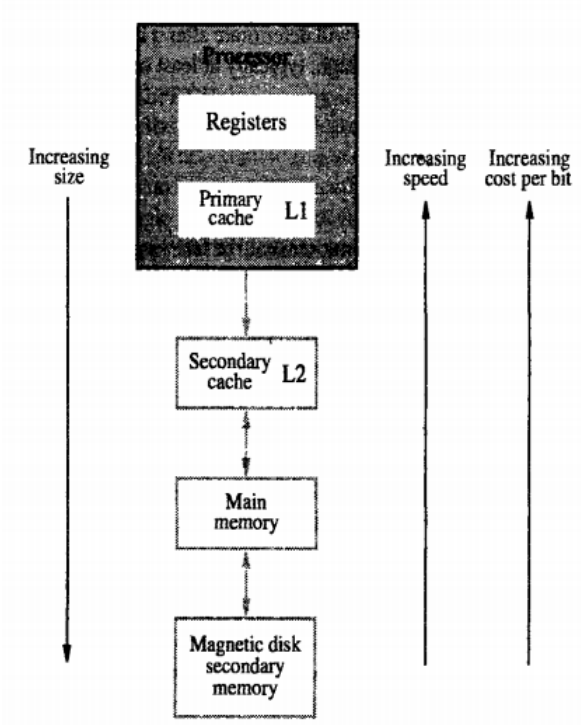


Figure 5.13 Memory hierarchy.

- Fastest access is to the data held in processor registers. Registers are at the top of the memory hierarchy.
- Relatively small amount of memory that can be implemented on the processor chip. This is processor cache. Usually implemented as SRAM.
- Two levels of cache.
Level 1 (L1) cache is on the processor chip.
Level 2 (L2) cache is in between main memory and processor.
- Next level is main memory, implemented as DRAM (SIMMs,RIMM,DIMM). Much larger, but much slower than cache memory.
- Next level is magnetic disks. Huge amount of inexepensive storage.
- Speed of memory access is critical, the idea is to bring instructions and data that will be used in the near future as close to the processor as possible.

Q6 Explain three types of mapping functions for cache memory.

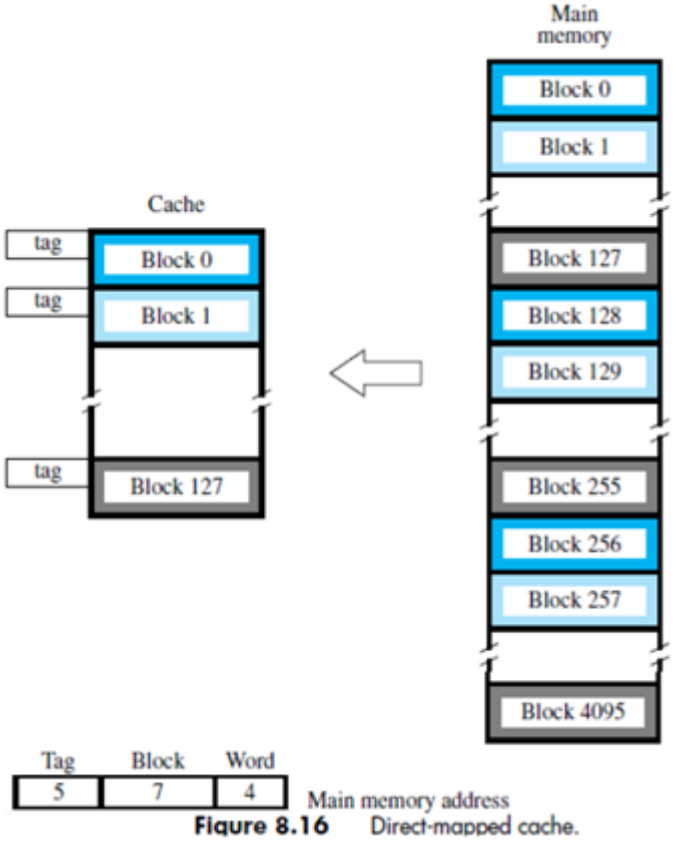


Figure 8.16 Direct-mapped cache.

DIRECT MAPPING

This technique is easy to implement but not very flexible.

Block j of the main memory maps onto j modulo 128 of the cache. For example, whenever **one of** the main memory blocks 0, 128, 256, ... Is loaded in the cache, it is stored in cache block 0 . Main memory blocks 1, 129, 257, ... are stored in cache block 1 (one at a time), and so on. Contention may occur for a single cache block required by multiple memory blocks. E.g when for program execution both memory block 1 and 129 are required but cache block 1 can only store one memory block. To resolve this, new blocks are allowed to overwrite the currently resident block.

From example,

4096 memory blocks need to be mapped to 128 cache blocks. i.e, each cache block identified 32 memory blocks(4096/128).

Main memory address is divided into three parts:

Tag (5 bits): identify which memory block (out of 32 in this case) is currently resident in the cache

Block (7 bits): cache block position where the new memory block must be stored

Word (4 bits): selects one of the words of the memory block (out of 16 words per block in this case)

ASSOCIATIVE MAPPING

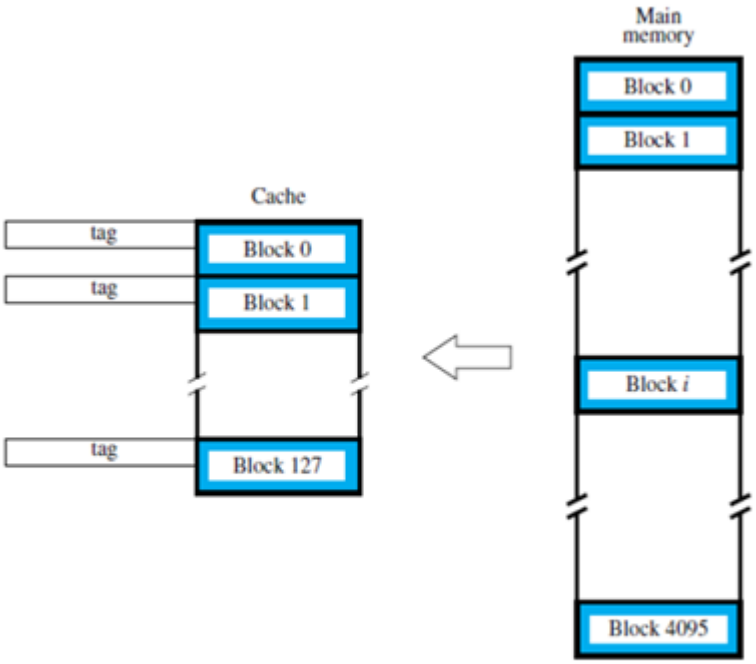


Figure 8.17 Associative-mapped cache.

- It is more flexible than direct mapping technique but more expensive. Main memory block can be placed into any cache block position.
- Memory address is divided into two fields:
  - Low order 4 bits identify the memory word within a block.
  - High order 12 bits or tag bits identify a memory block when residing in the cache.
- Flexible, and uses cache space efficiently.
- Replacement algorithms can be used to replace an existing block in the cache when the cache is full.
- Cost is higher than direct-mapped cache because of the need to search all 128 patterns to determine whether a given block is in the cache.

SET-ASSOCIATIVE MAPPING

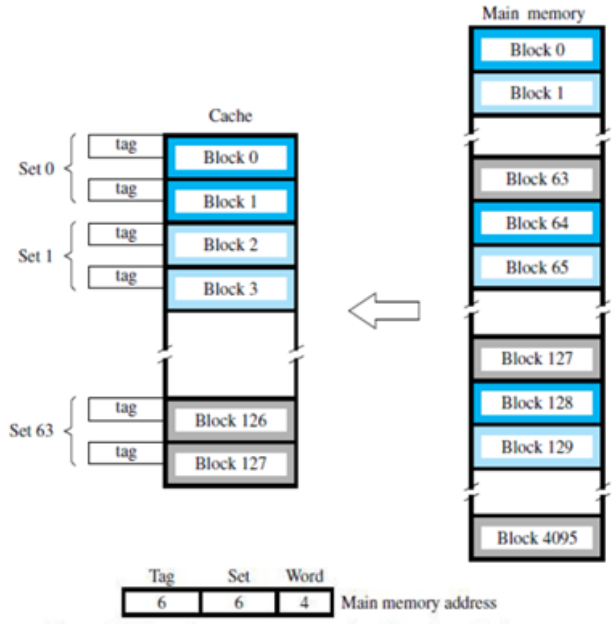


Figure 8.18 Set-associative-mapped cache with two blocks per set.

It is a **combination** of direct mapping and associative mapping techniques.
Blocks of the cache are grouped into **sets**, and the mapping allows a block of the main memory to reside in **any block of a specific set**.
Contention problem of direct mapping is eased by having a few choices for block placement. Hardware cost is reduced by decreasing the associative search.

Q8 Briefly explain replacement algorithms for cache memory.

Example:
Consider cache controller is tracking a set of four blocks in a set associative cache.
A 2-bit counter is used for each block (00=0, 01=1, 10=2, 11=3)

a) Hit occurs:
The counter of the block that is referenced is set to 0. Other block counters having value less than the value of the referenced counter are incremented by 1. Block counters having value greater than the value of referenced counter are remain unchanged.
Initially: 2, 3, 0, 1
Hit occurs for 2
Finally: 0 (after set to 0), 3 (unchanged), 1 (after increment), 2 (after increment)

b) Miss occurs (Set not full):
The counter of the block where new block is loaded from memory is set to 0. All other block counters value is incremented by 1.
Initially: 2, 1, 0, _
Miss occurs
Finally: 3, 2, 1, **0** (new)

c) Miss occurs (Set full):
The block with the highest counter value i.e, 3 is removed. The new block is put into its place and its counter value is set to 0. All other three blocks counter values are incremented by 1.
Initially: 2, 3, 0, 1
Miss occurs
Finally: 3, 0 (replaced), 1, 2

MODULE 3 –important questions