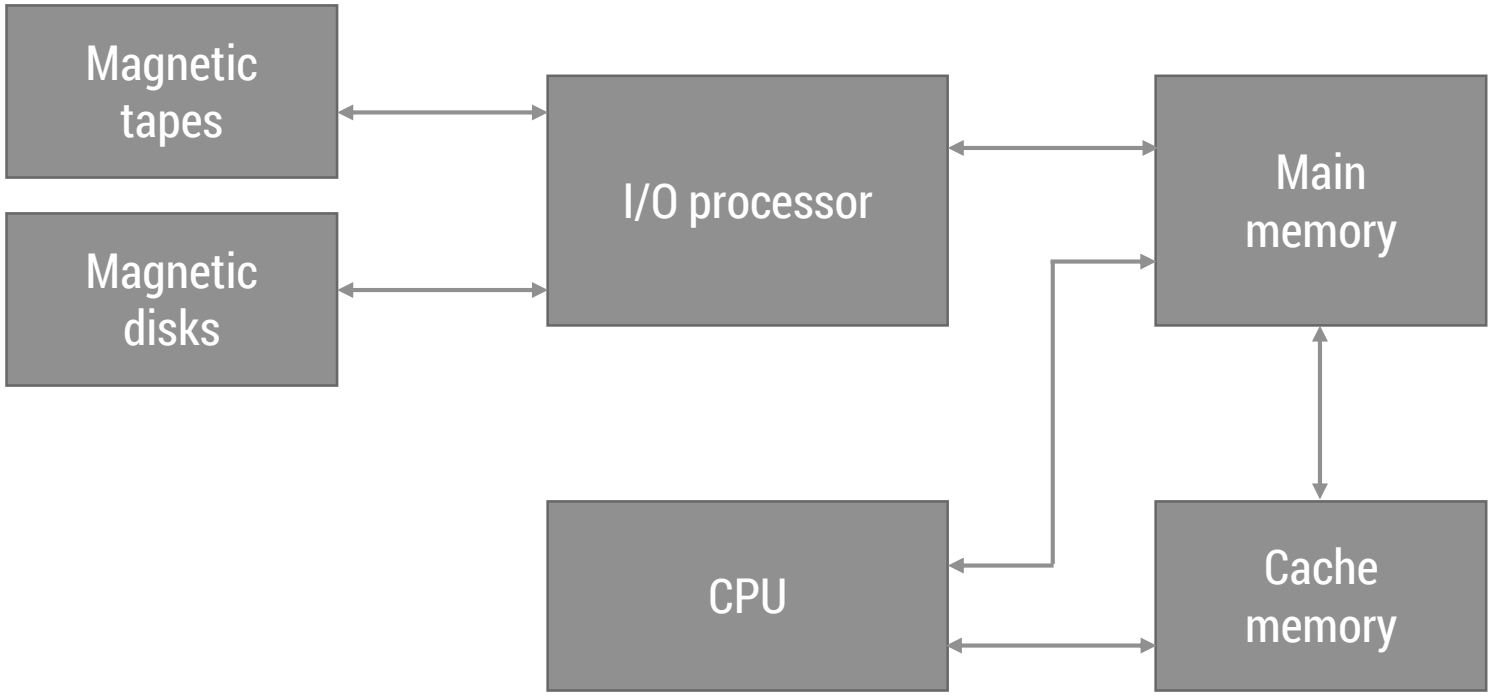
Unit-9

Memory Organization

Memory Hierarchy



Main Memory
Random access memory (RAM)

- Used in computers for the temporary storage of programs and data.
- Read and write both operations are performed by RAM which requires fast cycle times as not to slow down the computer operation.
- It is volatile and lose all stored information if power is interrupted or turned off.
- RAMs typically come with word capacities of 1K, 4K, 8K, 16K, etc.. and word sizes of 1, 4 or 8-bits.
- It can be expanded by combining several memory chips.

SRAM v/s DRAM

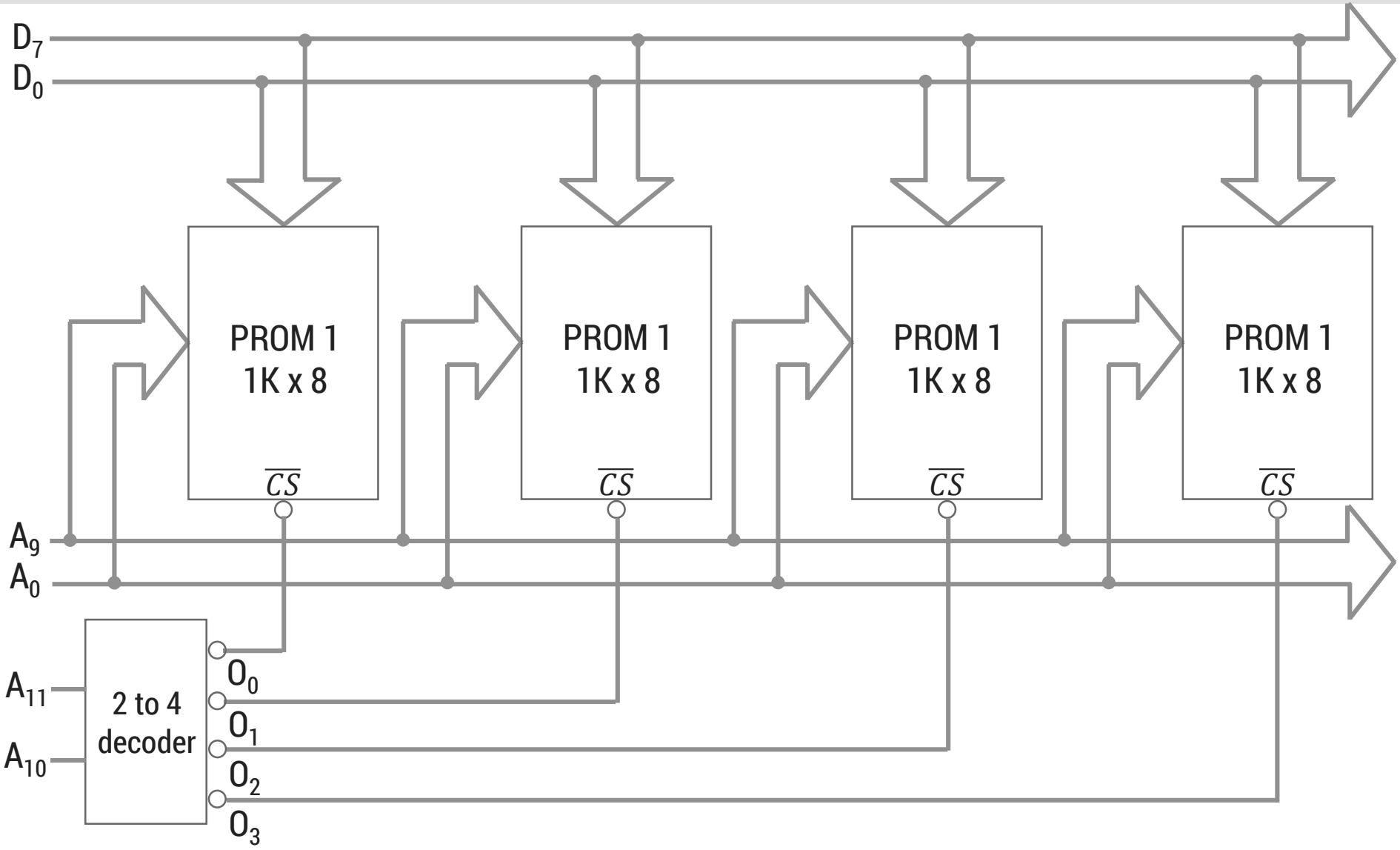
Static RAM	Dynamic RAM
1. SRAM has lower access time, so it is faster compared to DRAM.	1. DRAM has higher access time, so it is slower than SRAM.
2. SRAM is costlier than DRAM.	2. DRAM costs less compared to SRAM.
3.SRAM requires constant power supply, which means this type of memory consumes more power.	3. DRAM offers reduced power consumption, due to the fact that the information is stored in the capacitor.
4. Due to complex internal circuitry, less storage capacity is available compared to the same physical size of DRAM memory chip.	4. Due to the small internal circuitry in the one-bit memory cell of DRAM, the large storage capacity is available.
5. SRAM has low packaging density.	5. DRAM has high packaging density.
6. No need to refresh periodically.	6. Due to capacitor used as storage element, information may lose over period of time. So, need to refresh periodically.
7. Uses an array of 6 transistors for each memory cell.	7. Uses a single transistor and capacitor for each memory cell.

Read-Only Memory (ROM)

- A read-only memory (ROM) is essentially a memory device in which permanent binary information is stored.
- The binary information must be specified by the designer and is then embedded in the unit to form the required interconnection pattern.
- Once the pattern is established, it stays within the unit when the power is turned off and on again.
- A ROM which can be programmed is called a PROM. The process of entering information in a ROM is known as programming.
- ROMs are used to store information which is of fixed type, such as tables for various functions, fixed data and instructions.
- ROMs can be used for designing combinational logic circuits.

Memory Address Map

- Example:- Show how to combine several 1K x 8 PROMs to produce 4K x 8 PROM.
- Solution:- 1K x 8 PROM has 10 number of address lines because 2^{10} = 1024 (1K).
- We need total 4 number of 1K x 8 PROM chips to make one 4K x 8 PROM chip.



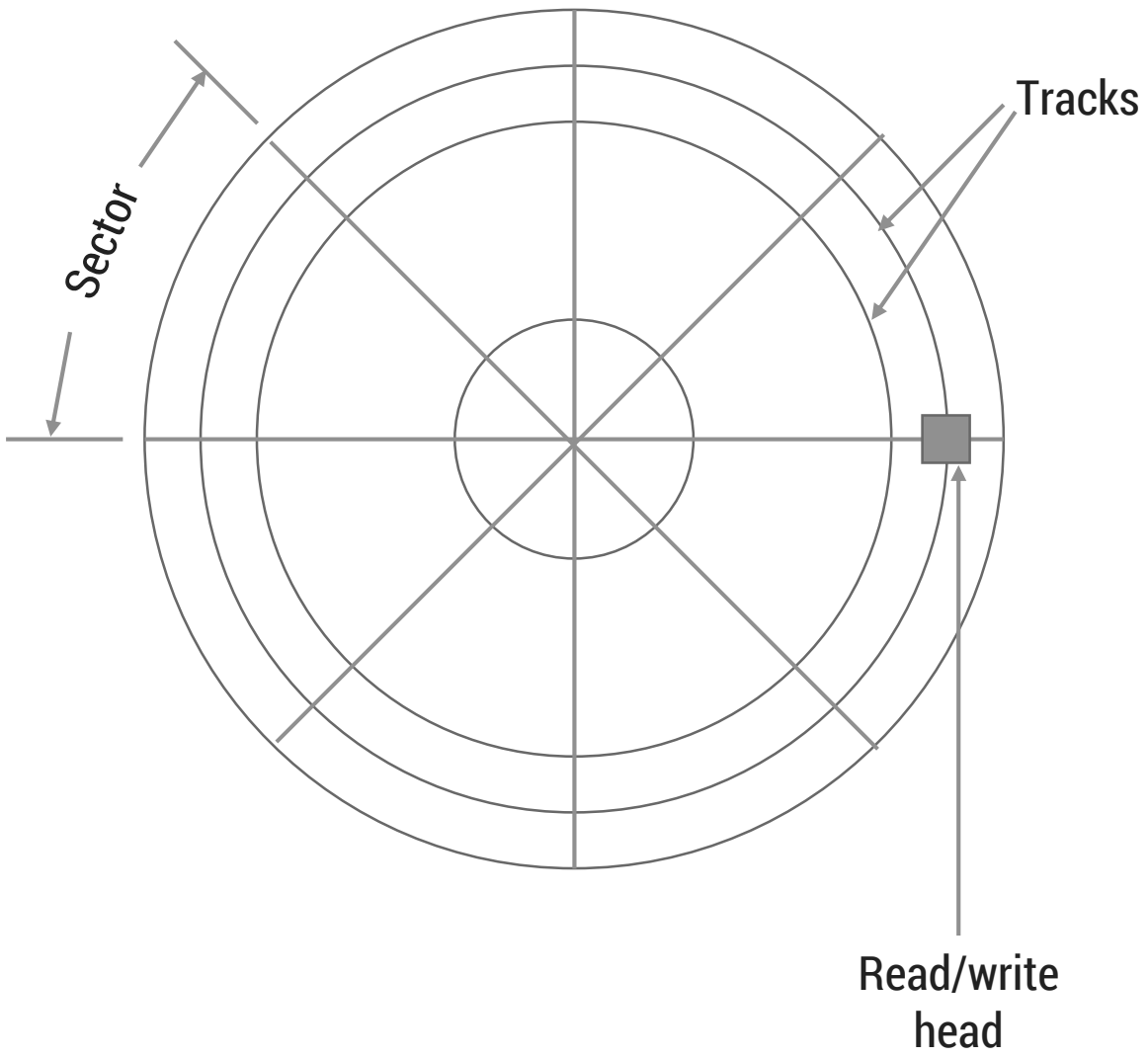
Memory Address Map

A_{11}	A_{10}	A_9	A_8	A_7	A_6	A_5	A_4	A_3	A_2	A_1	A_0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	1

Memory chip	Starting address	Ending address
PROM 1	000 H	3FF H
PROM 2	400 H	7FF H
PROM 3	800 H	BFF H
PROM 4	C00 H	FFF H

Auxiliary Memory

Magnetic Disks



- Circular plate of metal or plastic coated with magnetized material.
- Often both sides of the disk are used and several disks may be stacked on one spindle with read/write heads available on each surface.
- All disks rotate together at high speed and are not stopped or started for access purposes.
- Bits are stored in the magnetized surface in spots along concentric circles called track.
- The tracks are commonly divided into sections called sectors.

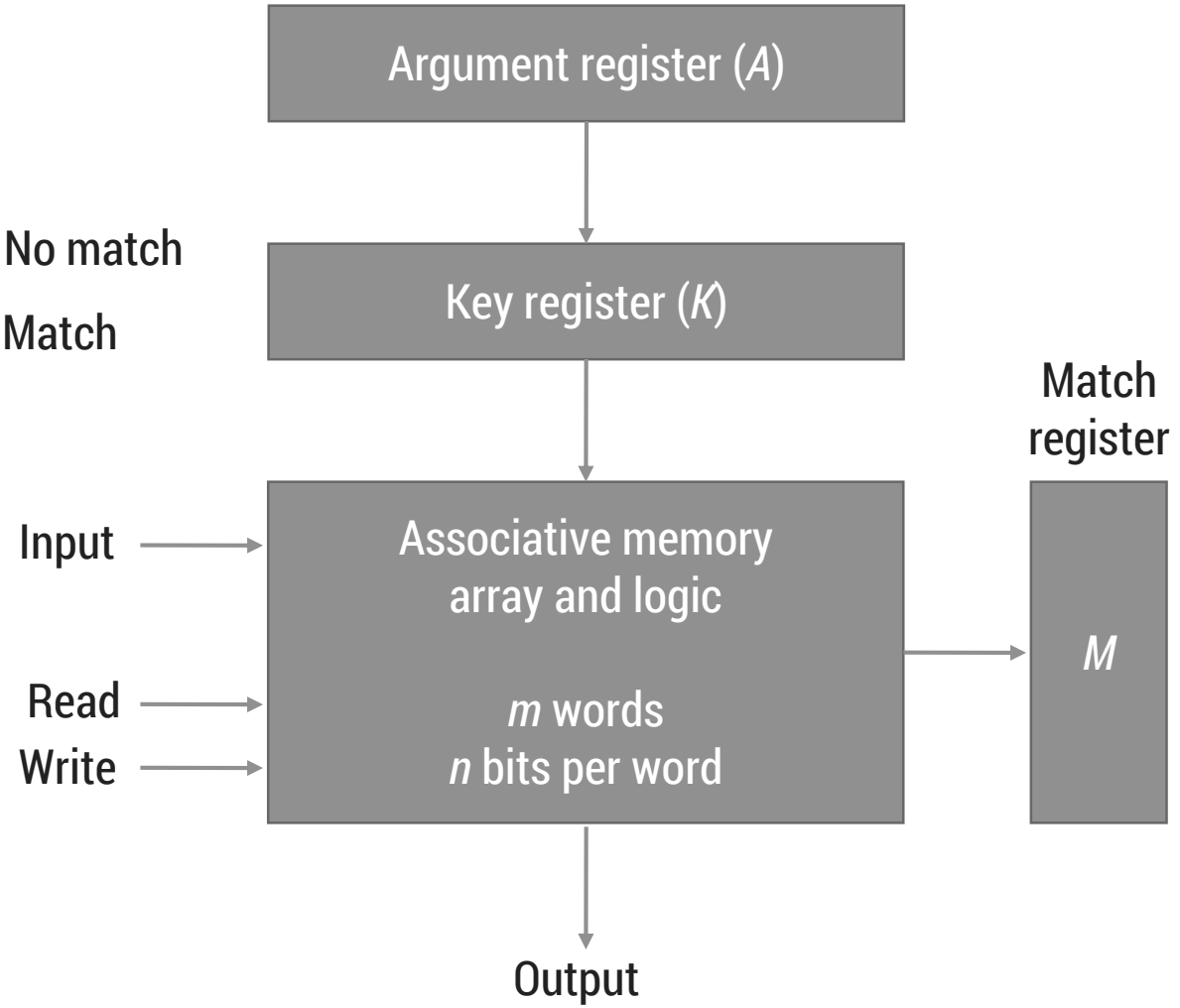
Magnetic Tape

- Magnetic tape transport consists of the electrical, mechanical and electronic components to provide the parts and control mechanism for a magnetic-tape unit.
- The tape itself is a strip of plastic coated with a magnetic recording medium.
- Bits are recorded as magnetic spots on the tape along several tracks.
- Magnetic tape units can be stopped, started to move forward or in reverse, or can be rewound. However, they cannot be started or stopped fast enough between individual characters.
- A tape unit is addressed by specifying the record number and the number of characters in the record.

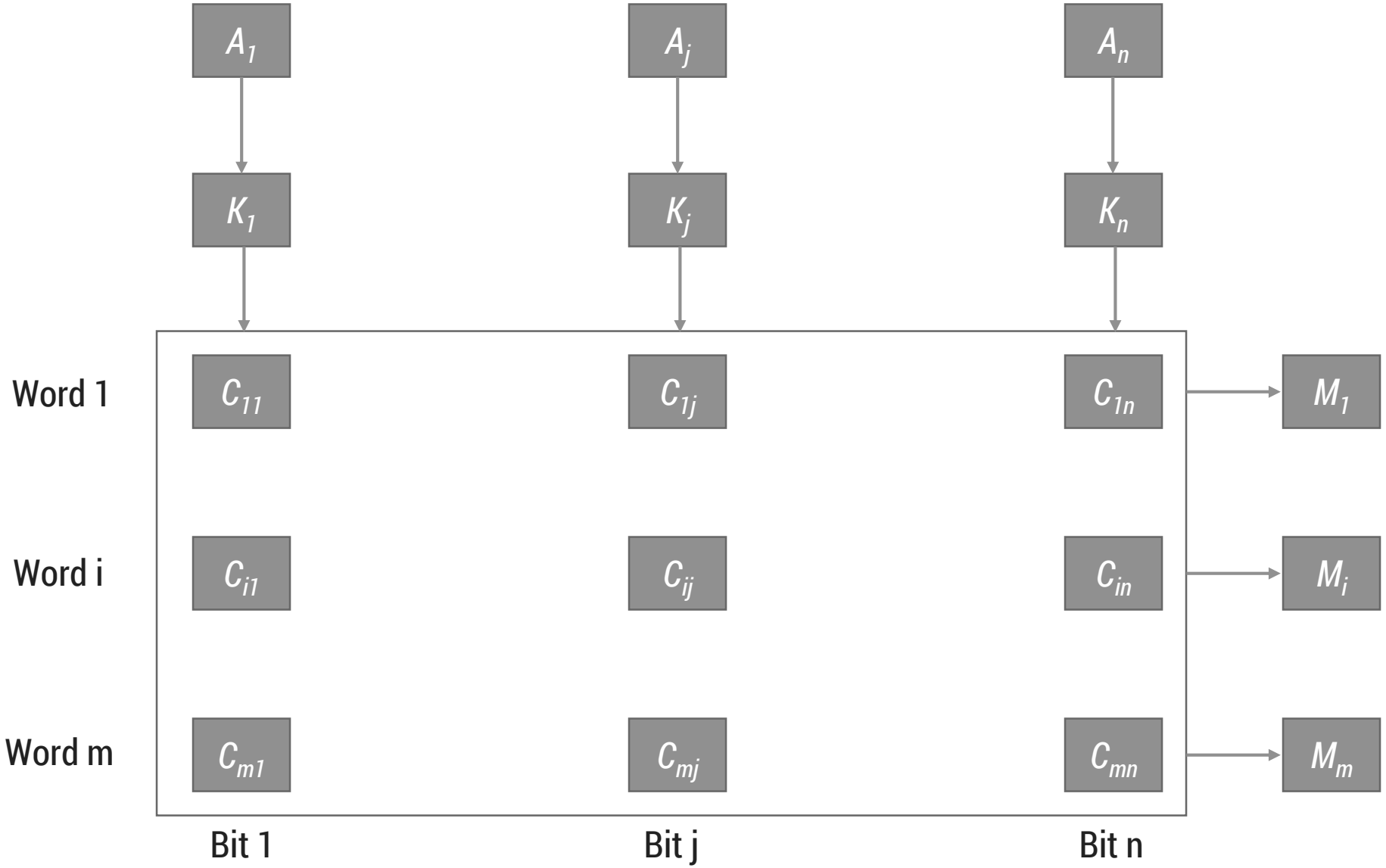
Associative Memory(Content Addressable Memory)

- The time required to find an item stored in memory can be reduced considerably if stored data can be identified for access by the content of the data itself rather than by an address.
- A memory unit accessed by content is called an associative memory or content addressable memory (CAM).
- This type of memory is accessed simultaneously and in parallel on the basis of data content rather than by specific address or location.

A	101	111100	
K	111	000000	
W1	100	111100	No match
W2	101	000001	Match



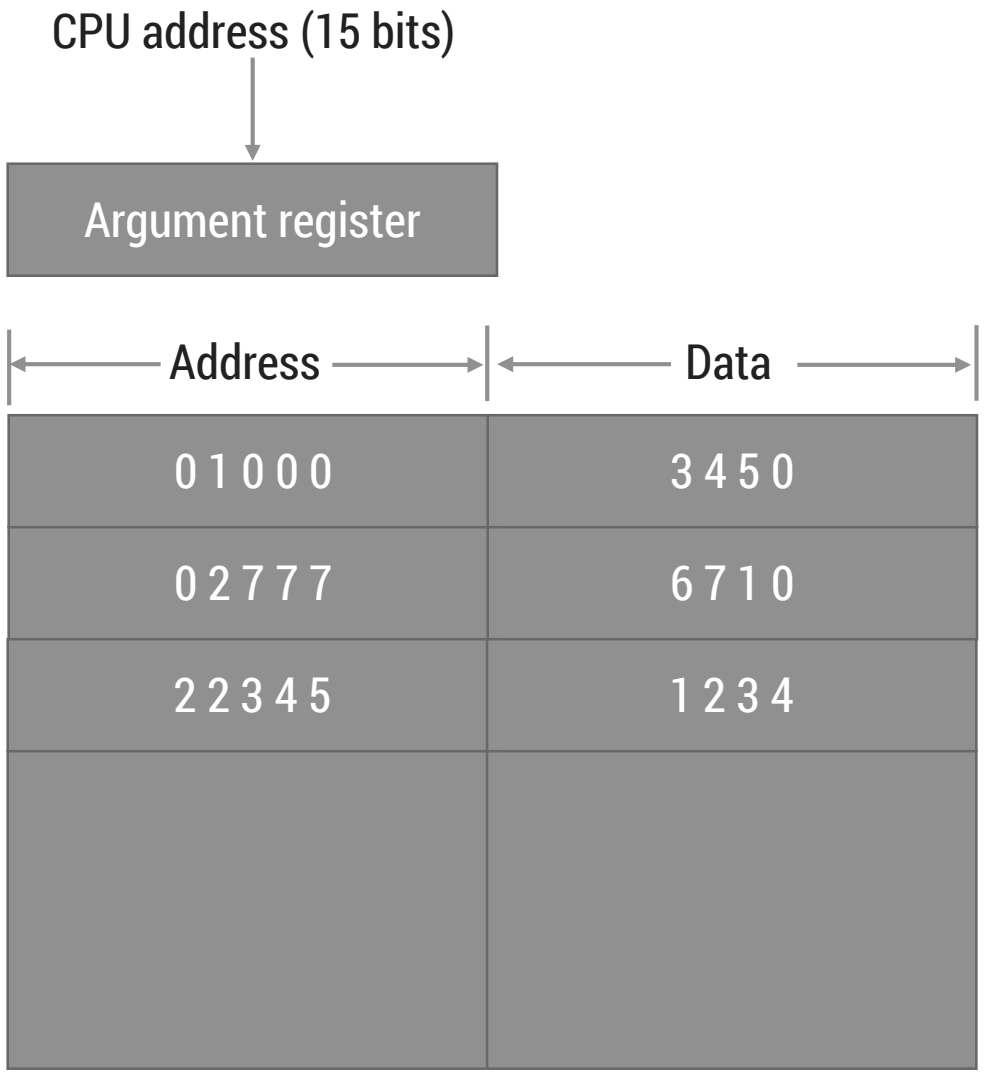
Associative Memory



Cache Memory

- Cache is a fast small capacity memory that should hold those information which are most likely to be accessed.
- The basic operation of the cache is, when the CPU needs to access memory, the cache is examined.
- If the word is found in the cache, it is read from the fast memory. If the word addressed by the CPU is not found in the cache, the main memory is accessed to read the word.
- The transformation of data from main memory to cache memory is referred to as a mapping process.
- The performance of the cache memory is frequently measured in terms of a quantity called *hit ratio*.
- When the CPU refers to memory and finds the word in cache, it is said to produce a *hit*.
- If the word is not found in cache, it is in main memory and it counts as a *miss.*
- The ratio of the number of hits divided by the total CPU references to memory (hits plus misses) is the *hit ratio.*
- Hit ratios of 0.9 and higher have been reported.

Associative Mapping

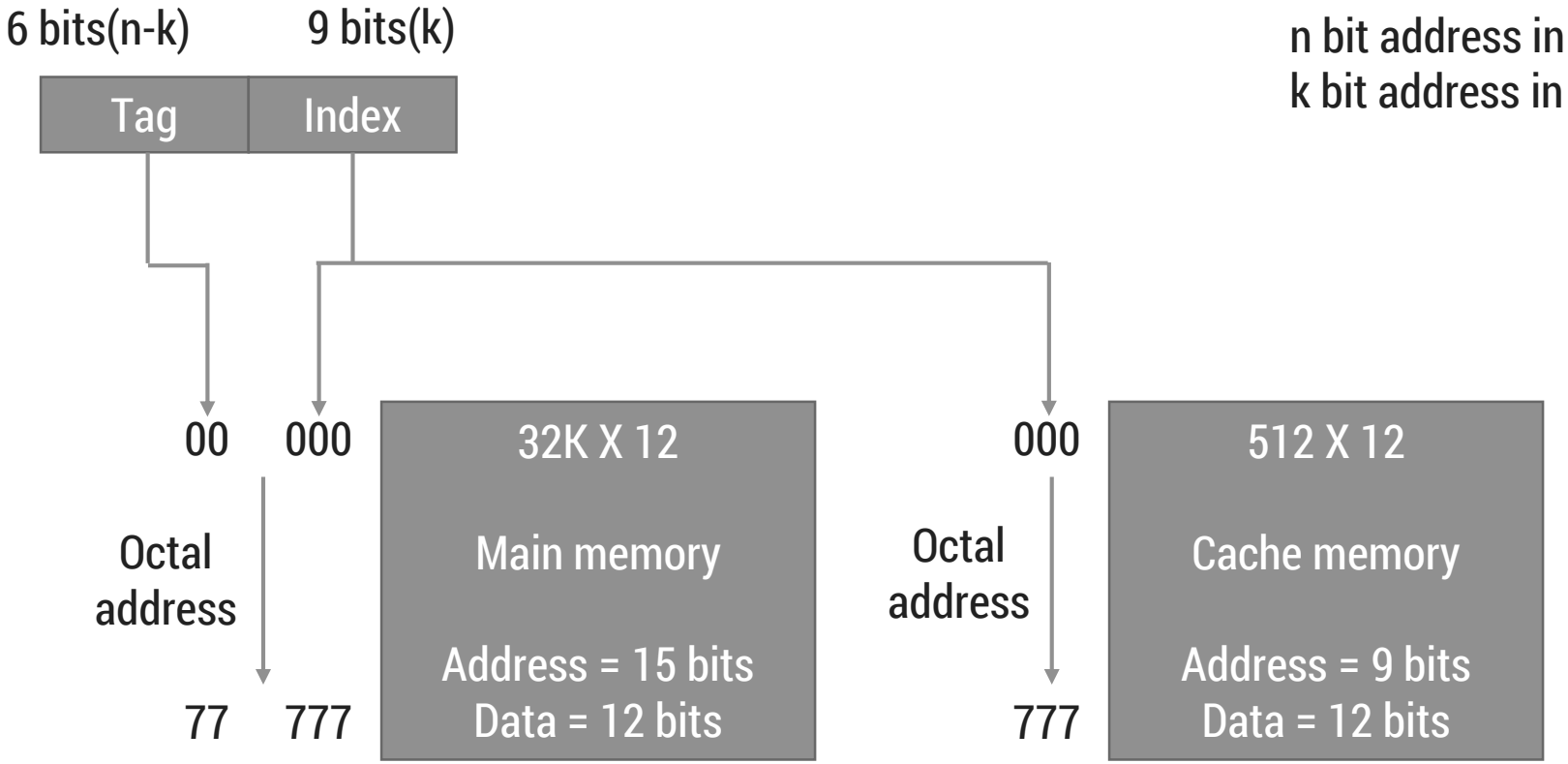


Associative memory
FIFO replacement policy
Address and data in octal
Round robin scheduling

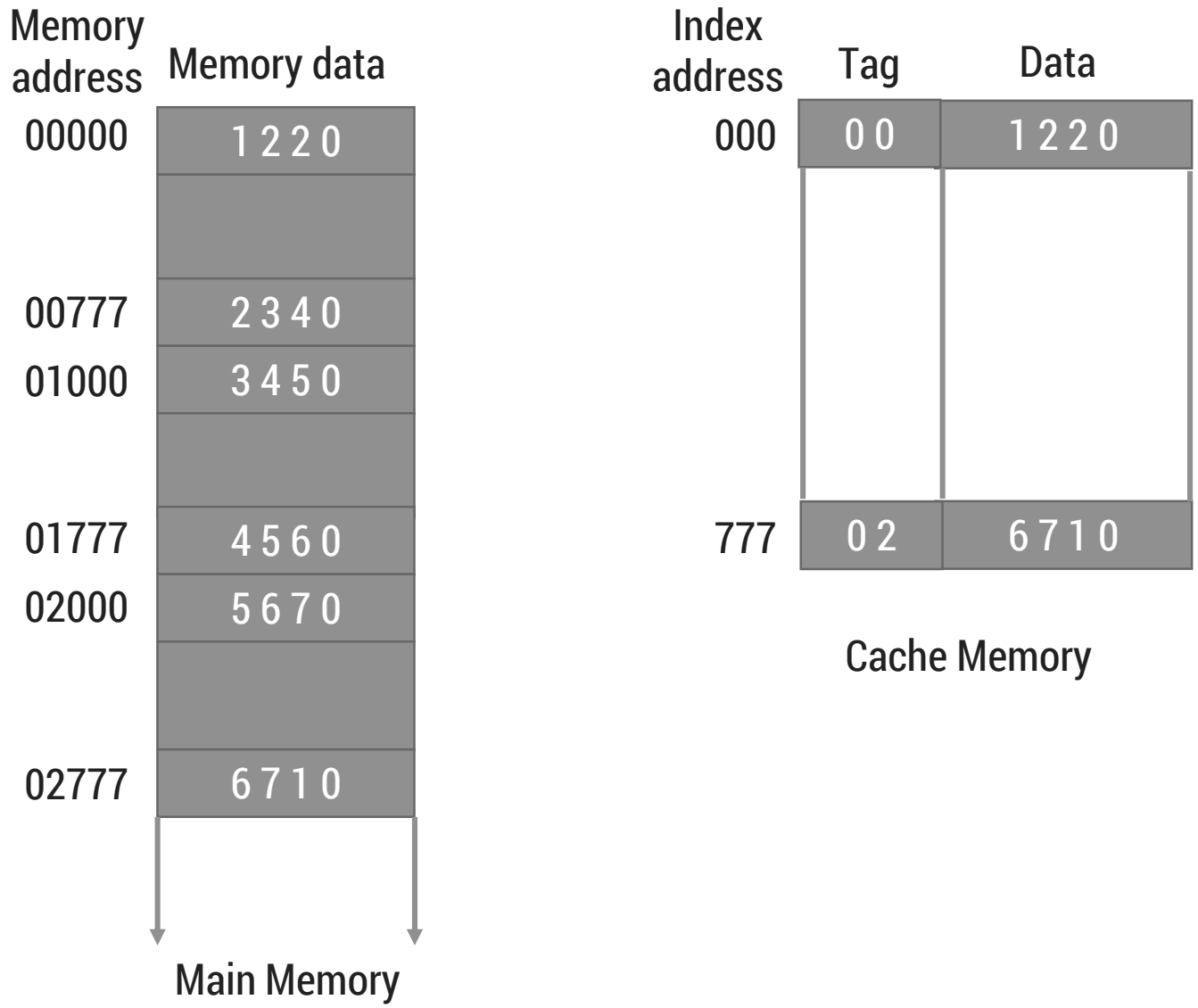
Direct Mapping

Associative memory is costly, RAM is used in direct mapping

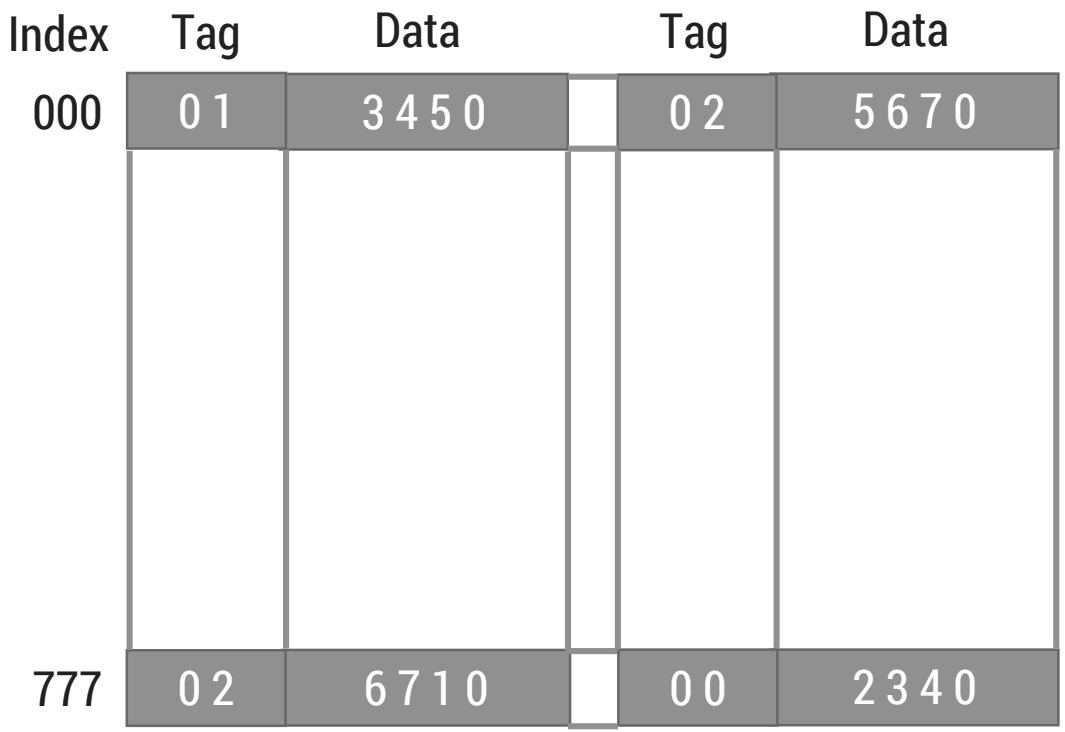
2^n words in RAM
2^k words in cache
n bit address in RAM
k bit address in cache



Direct Mapping



Set-Associative Mapping



Virtual Memory

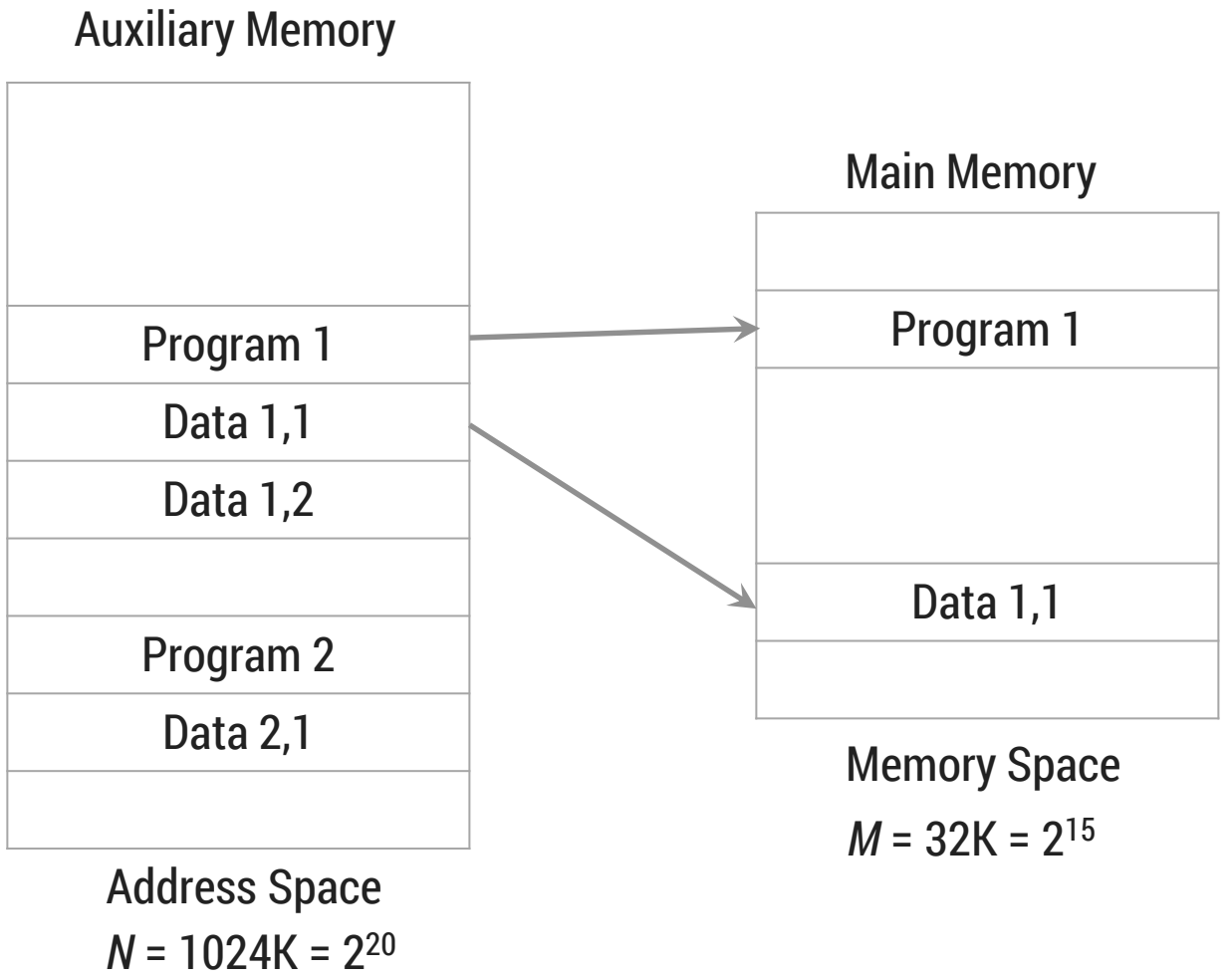
- Virtual memory is used to give programmers the illusion that they have a very large memory at their disposal, even though the computer actually has a relatively small main memory.
- A virtual memory system provides a mechanism for translating program-generated addresses into correct main memory locations.
- **Address space**

  An address used by a programmer will be called a virtual address, and the set of such addresses is known as address space.
- **Memory space**

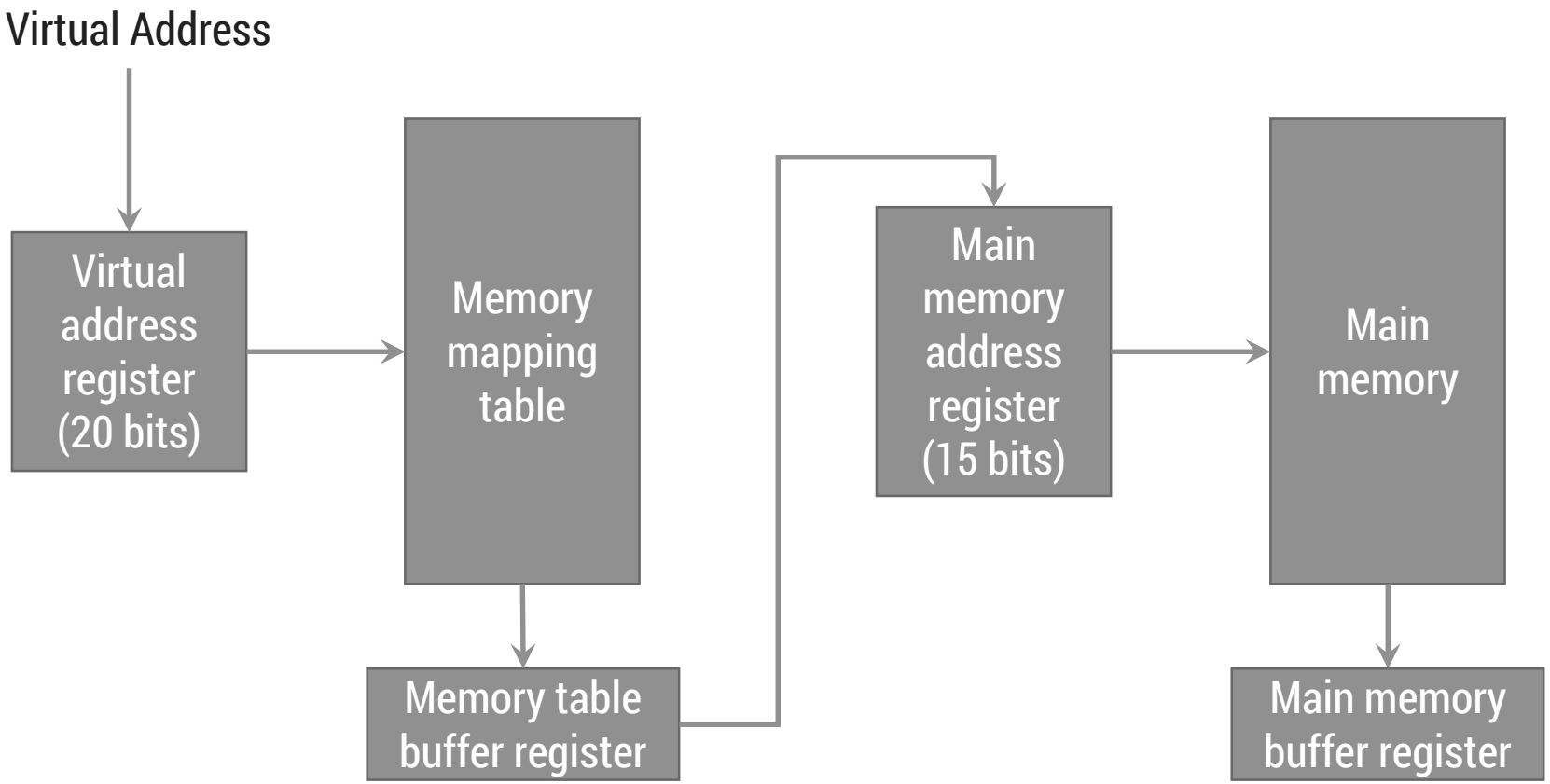
  An address in main memory is called a location or physical address. The set of such locations is called the memory space.

Virtual Memory

- Relation between Address space & Memory space

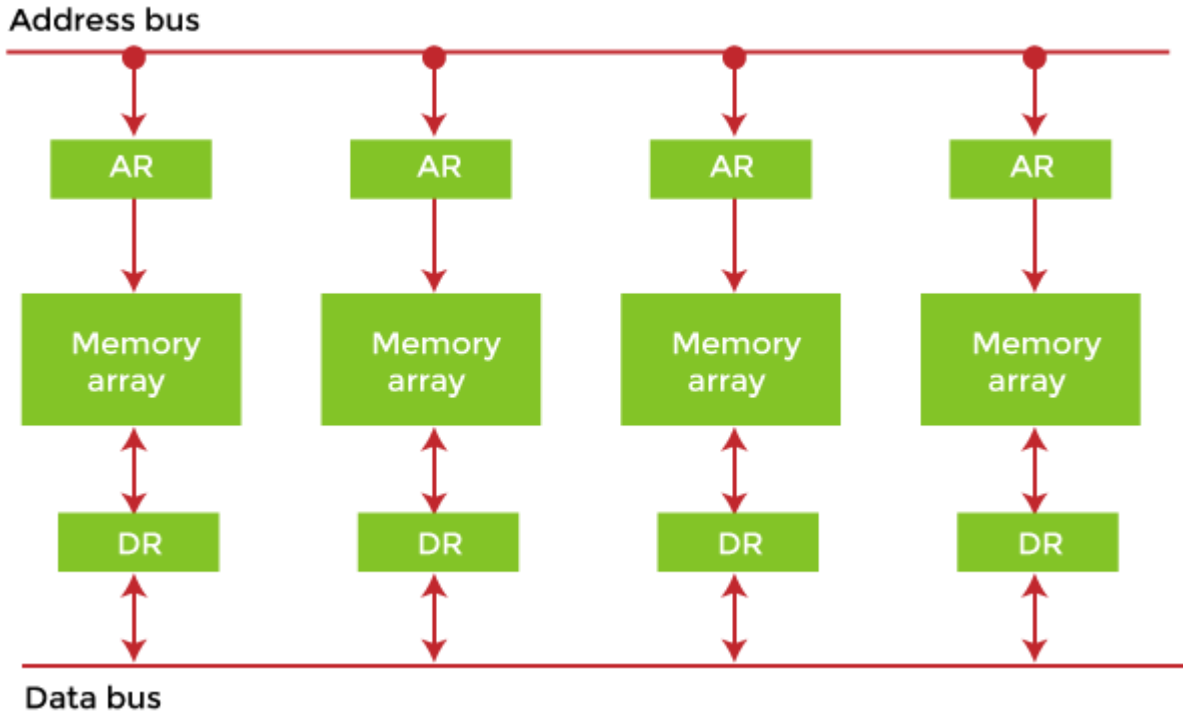


Virtual Memory

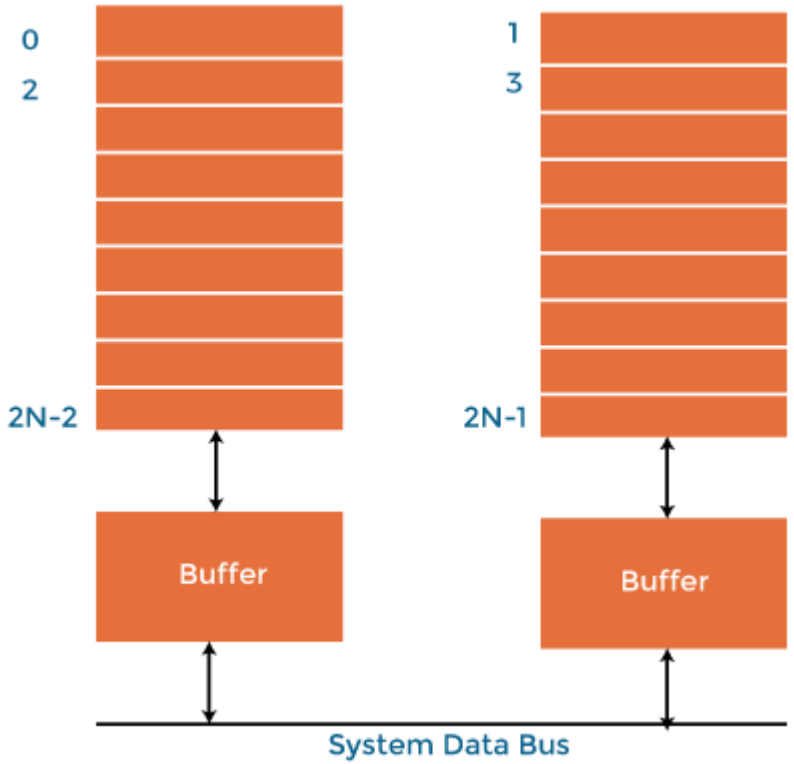


Memory Interleaving / Interleaved Memory

- Interleaved memory is designed to compensate for the relatively slow speed of dynamic random-access memory (DRAM) or core memory by spreading memory addresses evenly across memory banks. In this way, contiguous memory reads and writes use each memory bank, resulting in higher memory throughput due to reduced waiting for memory banks to become ready for the operations.



- In the interleaved bank representation below with 2 memory banks, the first long word of bank 0 is flowed by that of bank 1, followed by the second long word of bank 0, followed by the second long word of bank 1 and so on.
- The following image shows the organization of two physical banks of n long words. All even long words of the logical bank are located in physical bank 0, and all odd long words are located in physical bank 1.



- Interleaved memory results in contiguous reads (which are common both in multimedia and execution of programs) and contiguous writes (which are used frequently when filling storage or communication buffers) actually using each memory bank in turn, instead of using the same one repeatedly. This results in significantly higher memory throughput as each bank has a minimum waiting time between reads and writes.

Questions asked in GTU exam

1. For cache memories explain:
    1. Direct Mapping algorithm
    2. Set Associative Mapping
    3. Associative Mapping
2. What is virtual memory? Explain relation between address space and memory space in virtual memory system.
3. How main memory is useful in computer system? Explain the memory address map of RAM and ROM.
4. Write a short note on associative memory.
5. Content Addressable Memory.
6. Compare SRAM and DRAM.
7. Explain what do you understand by Cache memories
8. What is cache miss and cache hit?
9. Define hit ratio.