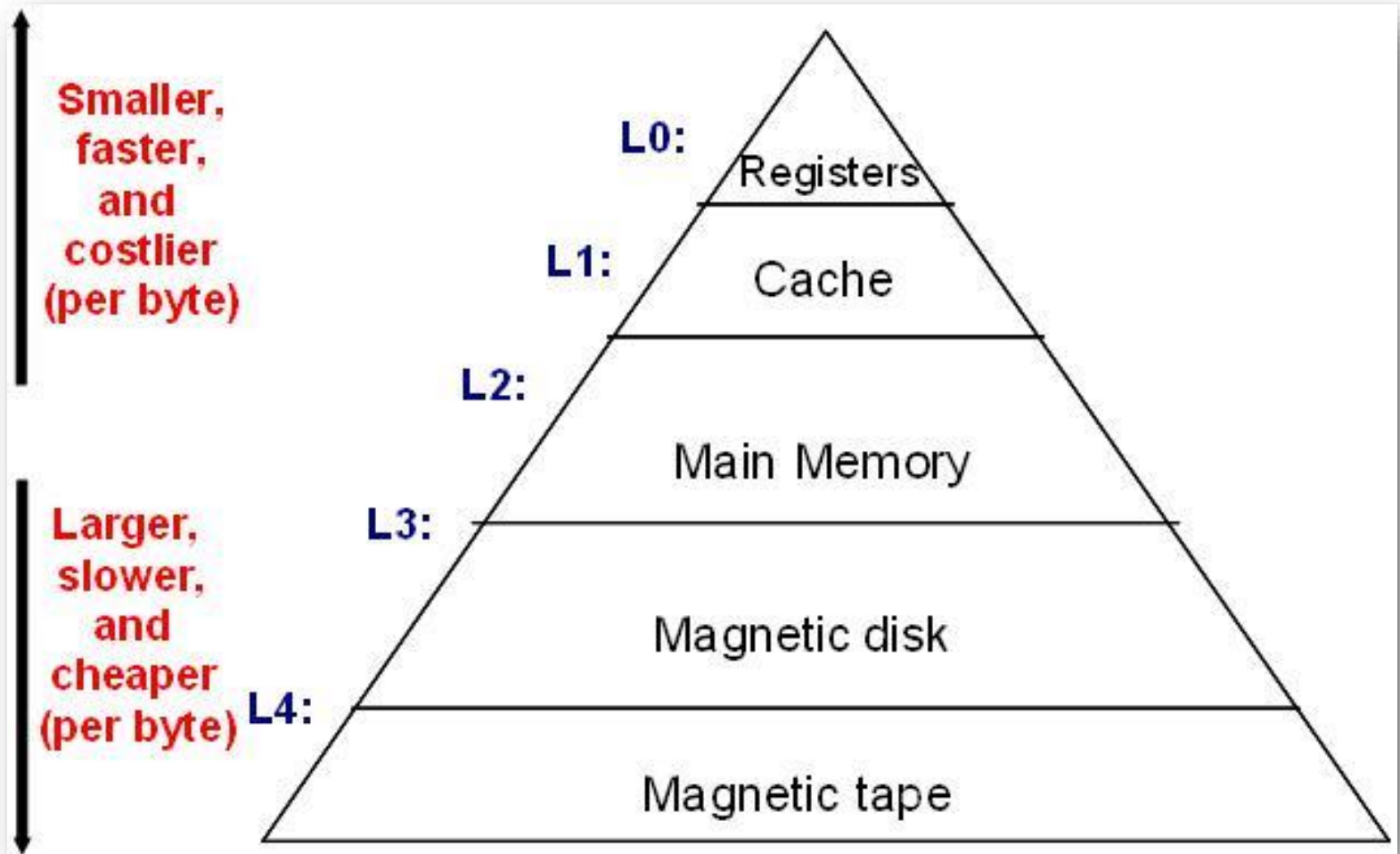# Unit – 2.b
Memory Organization
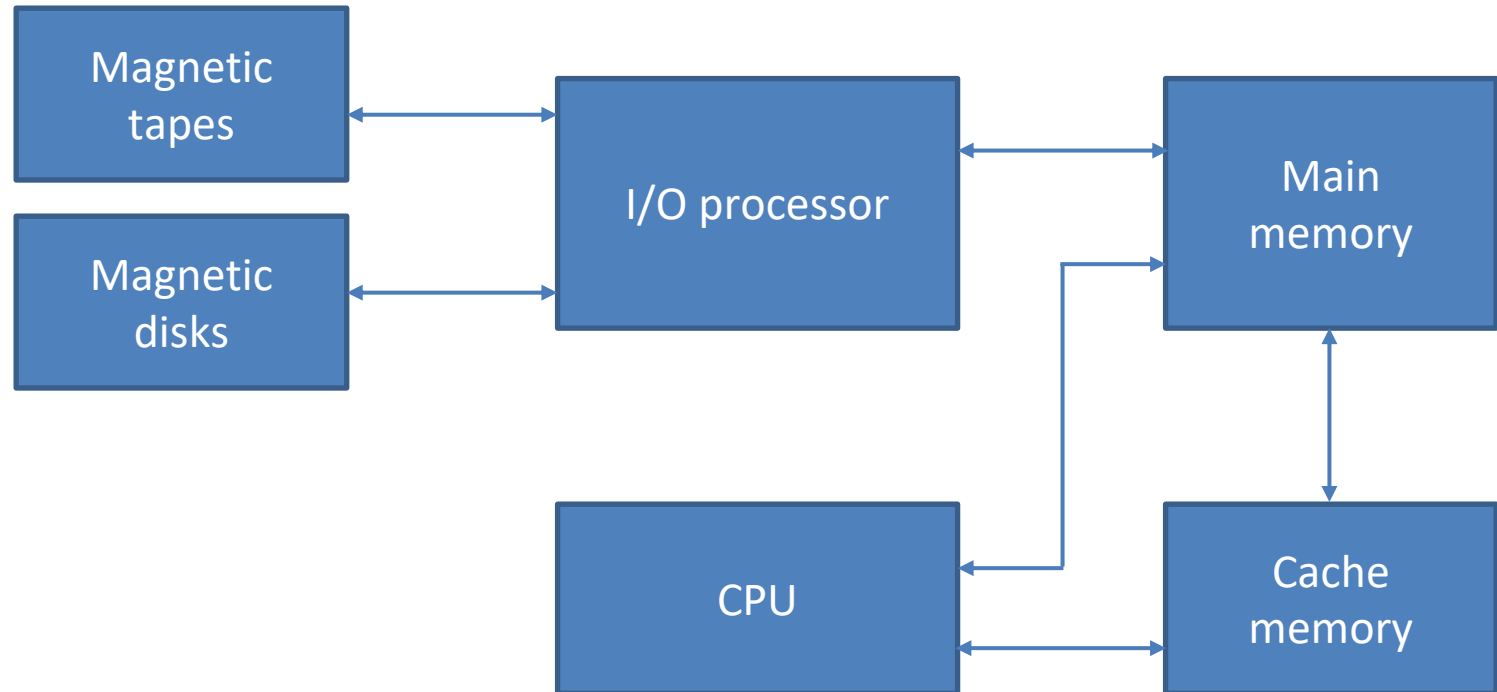
Topics to be covered

- Memory Hierarchy

- Main Memory

- Auxiliary Memory

- Cache Memory

- Virtual Memory

Memory Hierarchy

Smaller, faster, and costlier (per byte)

Larger, slower, and cheaper (per byte)

L0: Registers

L1: Cache

L2: Main Memory

L3: Magnetic disk

L4: Magnetic tape

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.

Auxiliary Memory

Magnetic Disks

Sector

Tracks

Read/write head
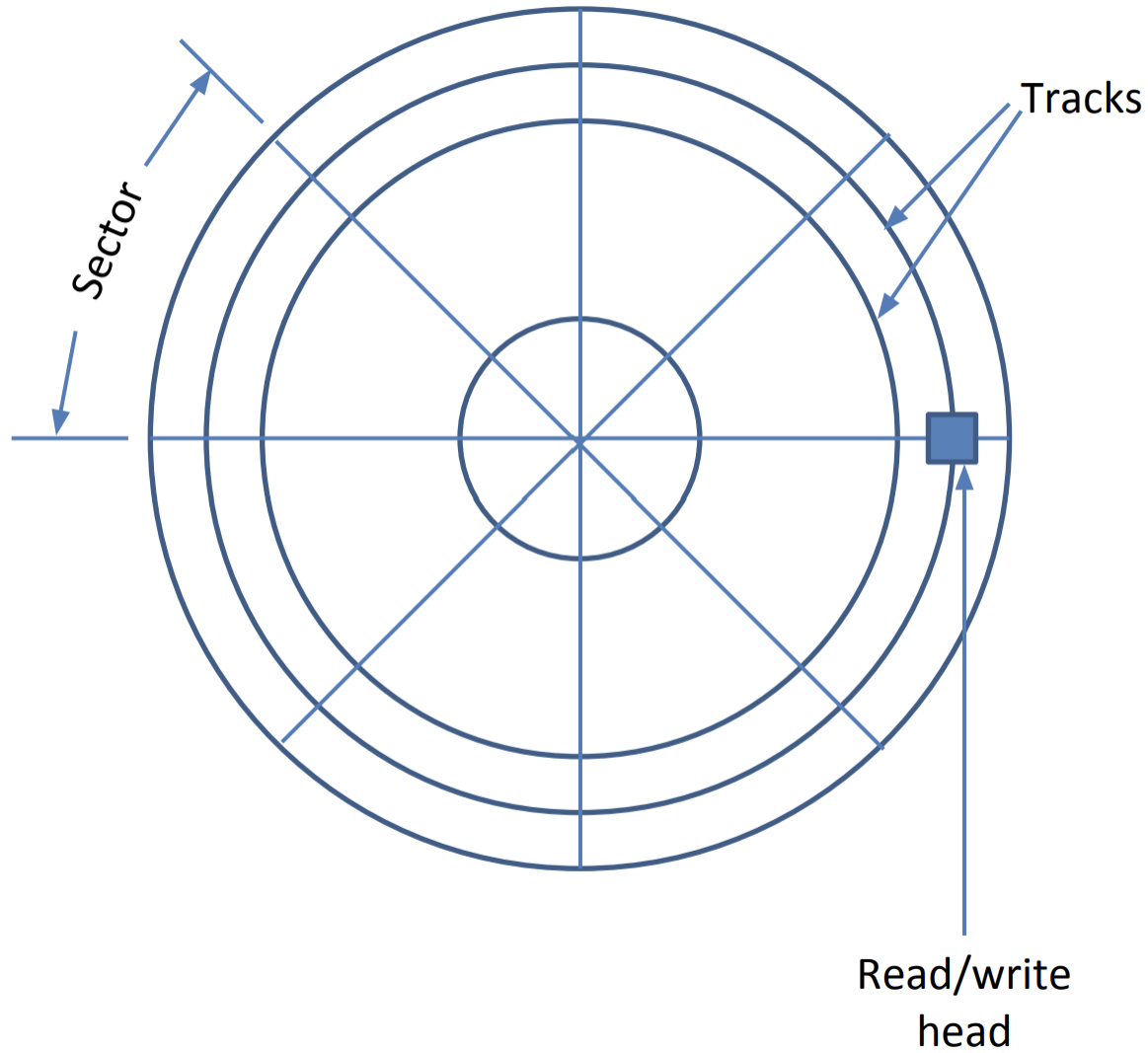
Magnetic Tape

- 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.

Cache 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.

Cache Memory

▪ 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.
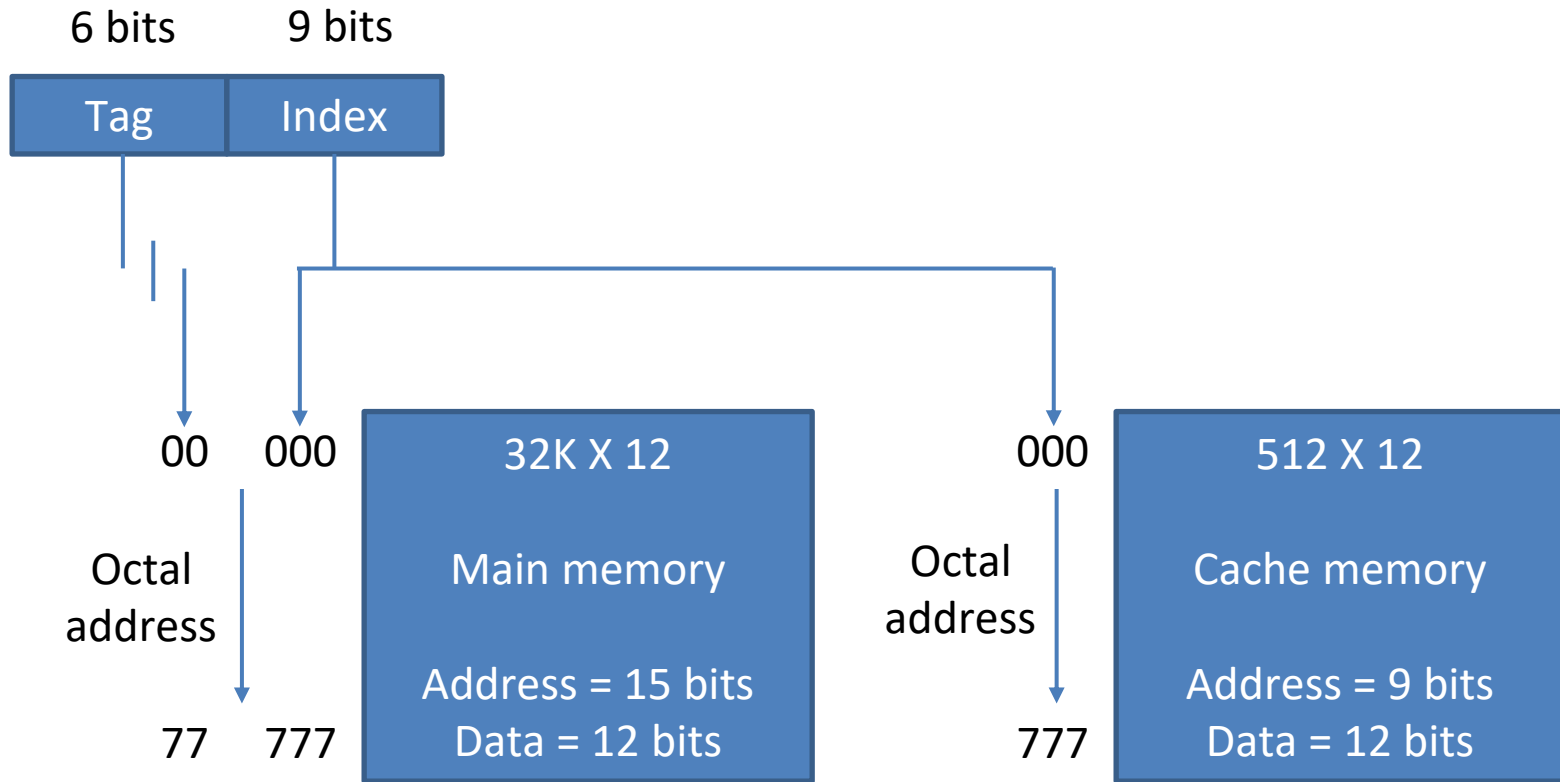
Associative Mapping
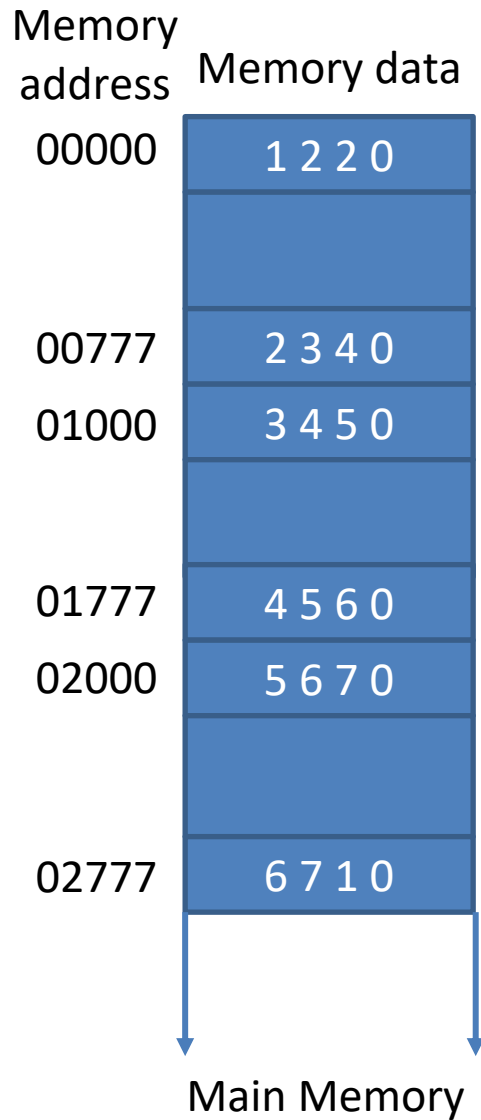
CPU address (15 bits)

Argument register

Address	Data
0 1 0 0 0	3 4 5 0
0 2 7 7 7	6 7 1 0
2 2 3 4 5	1 2 3 4

Direct Mapping

Direct Mapping

Memory address	Memory data
00000	1 2 2 0
00777	2 3 4 0
01000	3 4 5 0
01777	4 5 6 0
02000	5 6 7 0
02777	6 7 1 0

Main Memory

Index address	Tag	Data
000	0 0	1 2 2 0
777	0 2	6 7 1 0

Cache Memory

Set-Associative Mapping

Index	Tag	Data		Tag	Data
000	0 1	3 4 5 0		0 2	5 6 7 0
777	0 2	6 7 1 0		0 0	2 3 4 0

Virtual Memory

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.
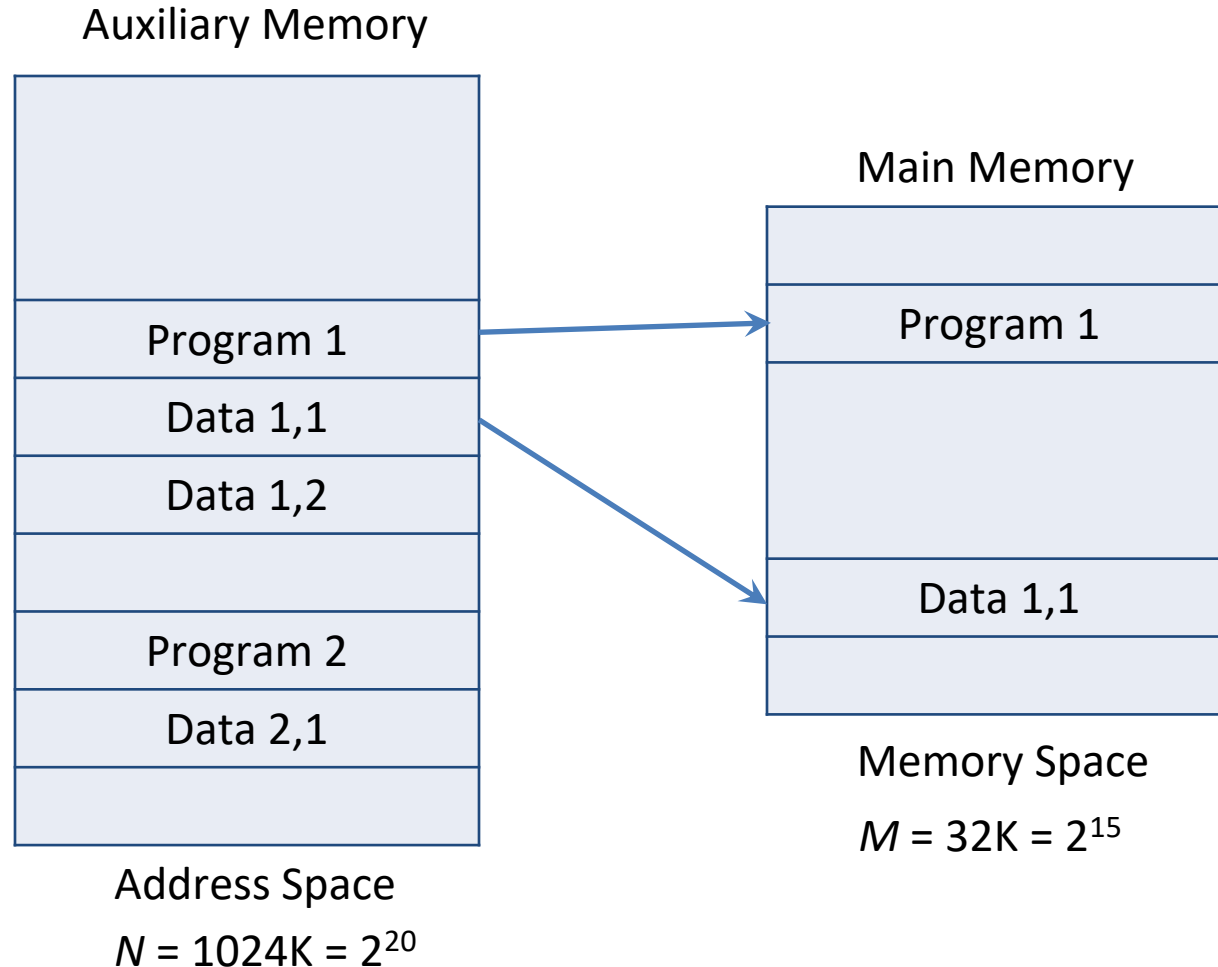
Virtual Memory

- **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.
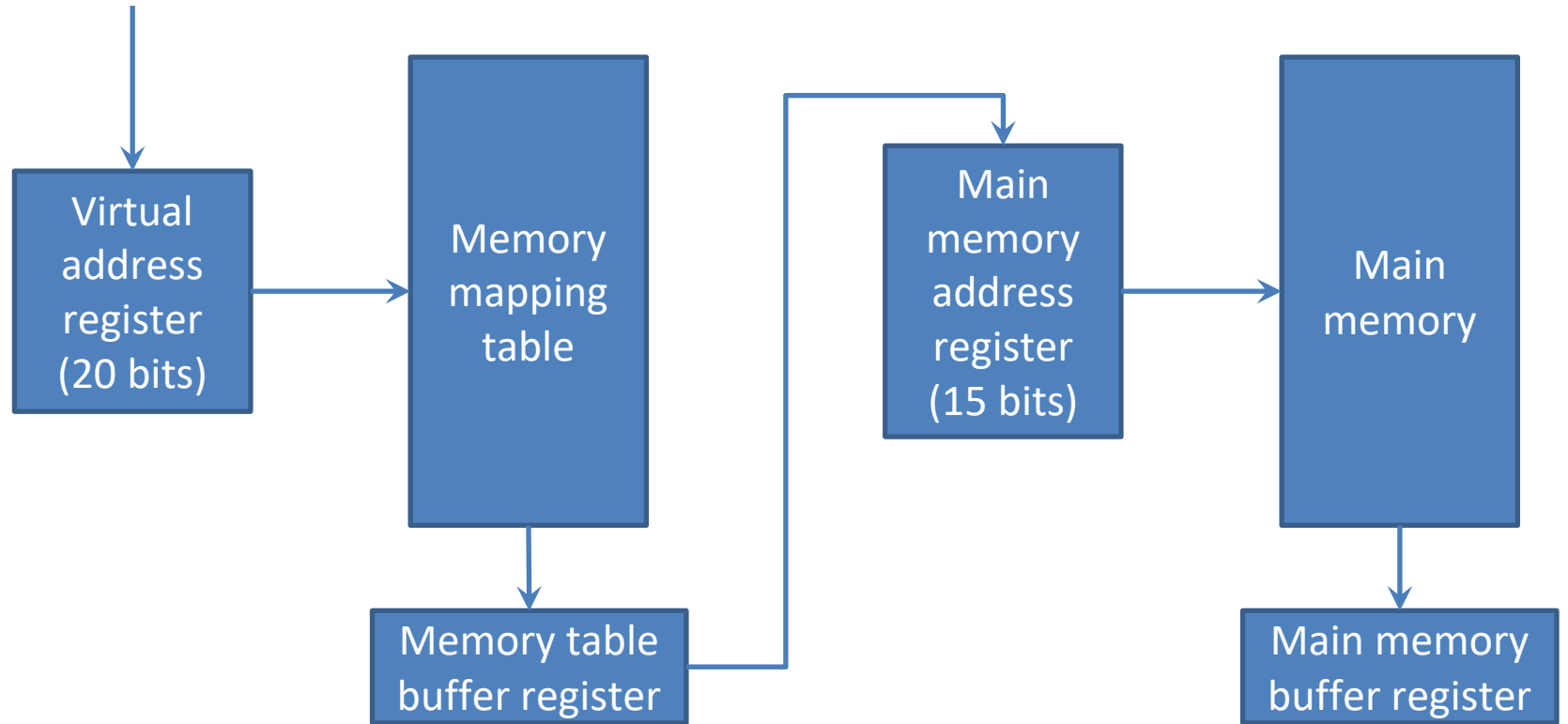
Virtual Memory

- Relation between Address space & Memory space

Auxiliary Memory

Program 1
Data 1,1
Data 1,2
Program 2
Data 2,1

Address Space
N = 1024K = 2^{20}

Main Memory

Program 1
Data 1,1

Memory Space
M = 32K = 2^{15}

Virtual Memory

Virtual Address

Virtual address register (20 bits)

Memory mapping table

Main memory address register (15 bits)

Main memory

Memory table buffer register

Main memory buffer register

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 vitual 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.

Questions asked in GTU exam

6. Explain paging and address translation with example.

7. Compare SRAM and DRAM.

8. Explain what do you understand by Cache memories

9. What is cache miss and cache hit?

10. Define hit ratio.

11. What is page fault?