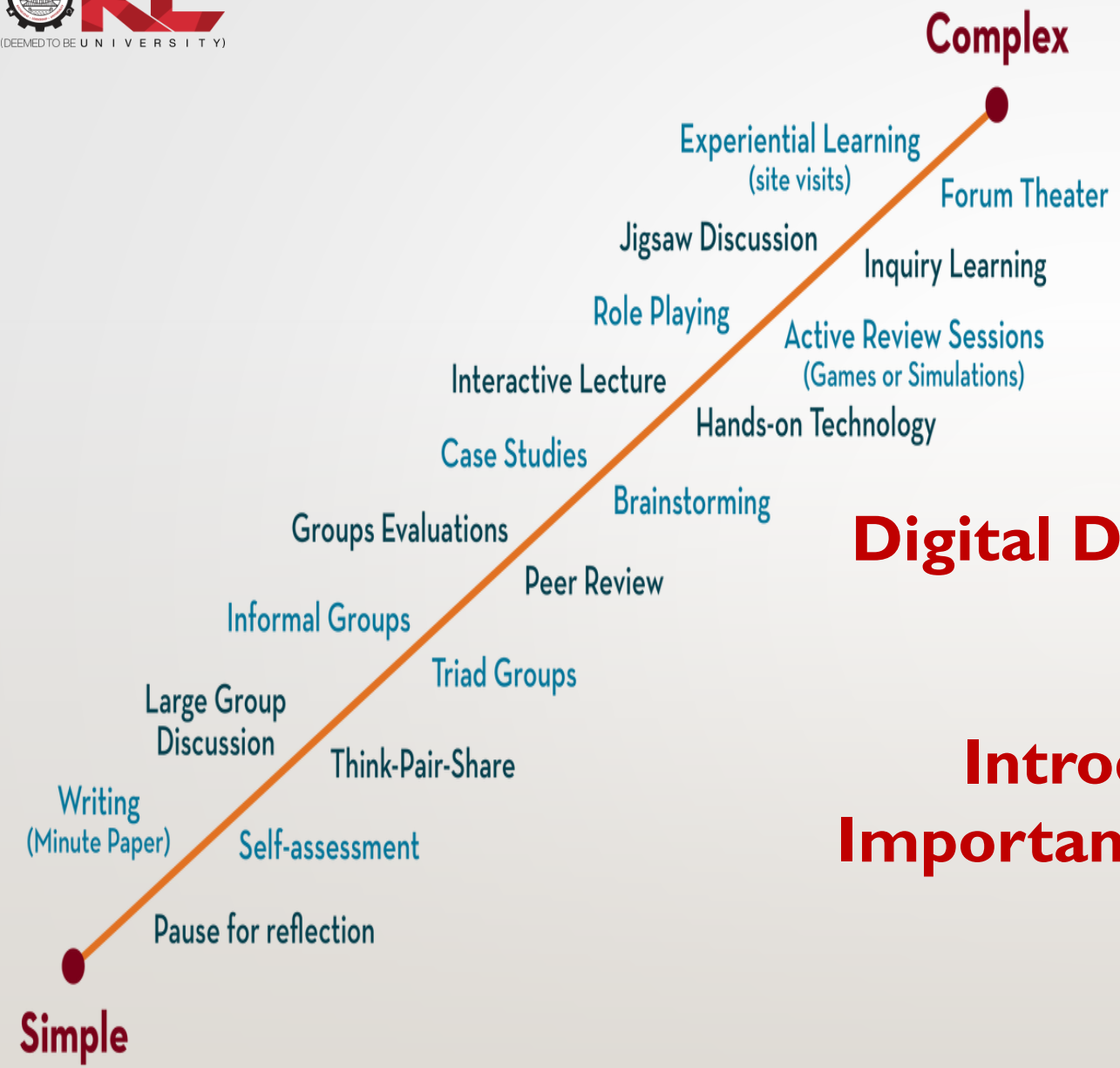


Department of BES-II

Digital Design and Computer Architecture

23EC1202

Topic:

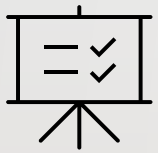
Introduction to Memory Hierarchy: Importance to temporal and spatial locality

Session No: 32

AIM OF THE SESSION

To familiarize students with the basic concept of Memory and its hierarchy with the importance of temporal and spatial locality.

INSTRUCTIONAL OBJECTIVES



This Session is designed to:

1. Demonstrate caches that can dramatically improve the speed of memory accesses.
2. Describe how virtual memory provides security and ease of programming.
3. List out the property of cache blocks and their advantage as a unit of cache organization.
4. Describe how the processors, memory and peripheral devices can be connected.

LEARNING OUTCOMES

At the end of this session, you should be able to:

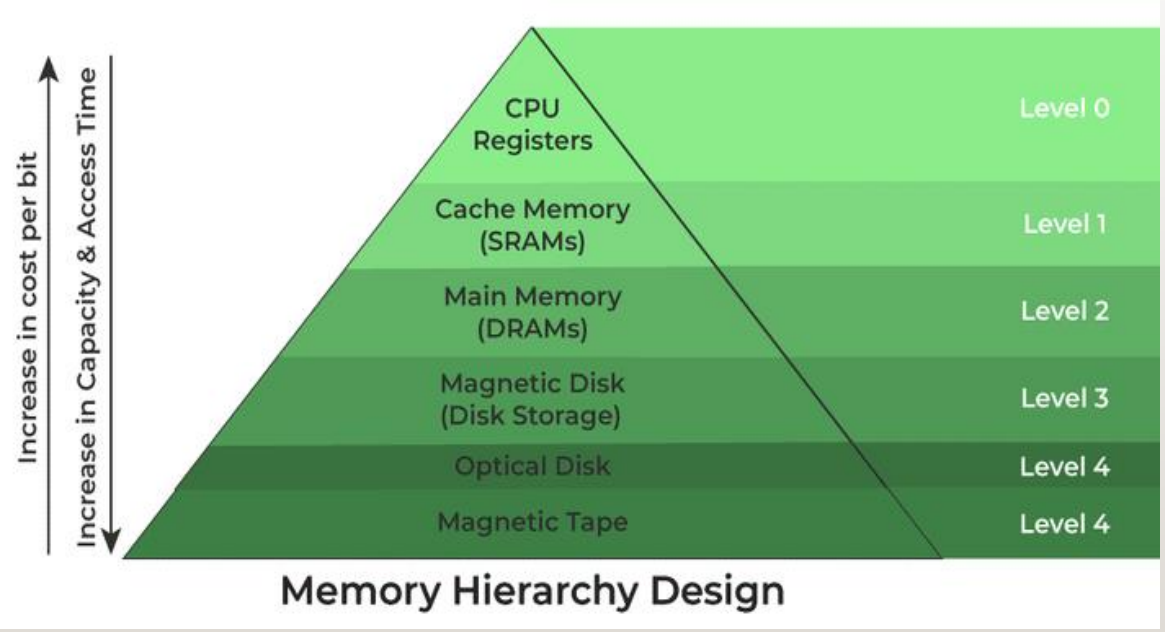
1. Define cache.
2. Describe hierarchy of memory.
3. Summarize importance of temporal and spatial locality in memory hierarchy.

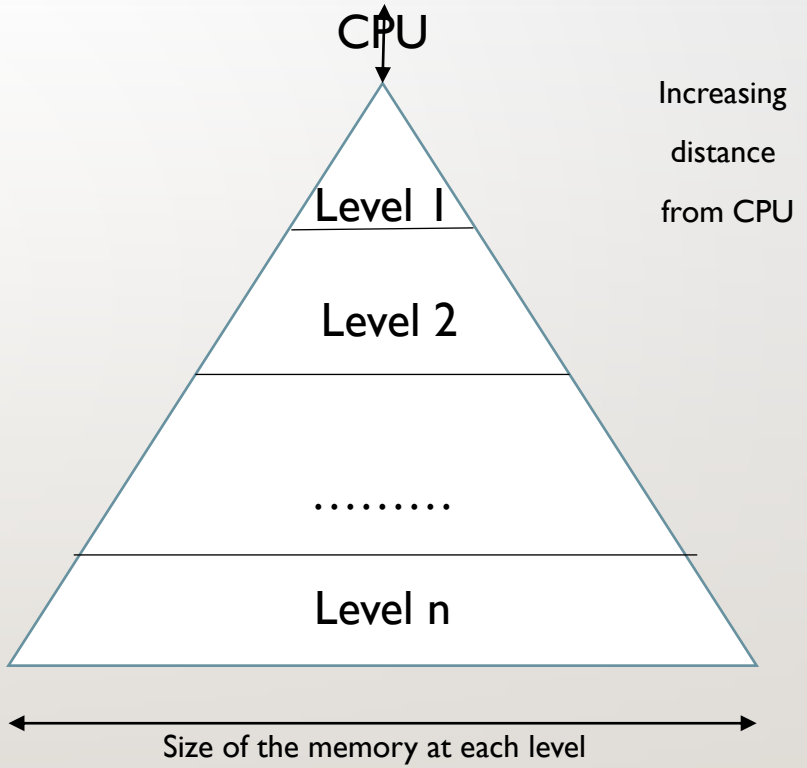
Memory Hierarchy

- Memory hierarchy is the goal of presenting the processor with the data it needs at the right time, as efficiently as possible.
- The memory hierarchy is typically structured from the fastest, least voluminous, and most expensive per bit storage to the slowest, most voluminous, and cheapest per bit storage.
- This structure includes registers at the top, followed by cache memory, primary memory (RAM), secondary storage (hard drives, SSDs), and even tertiary storage for large-scale data storage solutions.

Memory Hierarchy

- In the memory hierarchy :
  - ➢ Small and fast memories are put closer to CPU
  - ➢ Large and slow memories further away.





Temporal Locality

- Temporal locality refers to the principle that if a particular storage location was accessed recently, it is likely to be accessed again in the near future.
- This pattern of data access is leveraged by caching mechanisms within the memory hierarchy.
- By keeping recently accessed data in faster storage layers (like CPU caches), the system can significantly reduce access times for these repeatedly accessed data points, leading to performance improvements.

Spatial Locality

- Spatial locality is based on the principle that if a storage location is accessed, locations whose addresses are close by are likely to be accessed soon.
- This is due to the way programs are typically structured and accessed; for example, when a program accesses an array of data, it's likely to access multiple items of that array in sequence.
- Taking advantage of spatial locality means that when data is fetched into faster storage layers, not just the specifically requested data is fetched, but also the adjacent data.
- This preemptive action reduces the need for future accesses to slow storage layers for nearby data.

Why Temporal and Spatial Matters?

- **Performance Enhancement:** By strategically caching data that exhibits temporal and spatial locality, systems can dramatically reduce the average time to access data. This leads to faster program execution and an overall performance boost, as the processor spends less time waiting for data.

- **Efficiency and Energy Savings:** Accessing data from higher levels of the memory hierarchy (like RAM and caches) consumes less energy compared to fetching data from disk storage. Efficient use of the memory hierarchy can lead to significant energy savings, especially in large-scale data centers and in battery-operated devices.

Why Temporal and Spatial Matters?

- **Effective Resource Utilization:** By understanding and designing systems that leverage the principles of temporal and spatial locality, architects can make more effective use of the available memory resources. This includes deciding the sizes of caches, the organization of data in memory, and the algorithms for prefetching data.

Temporal Vs Spatial

Temporal	Spatial
A recently executed instruction is likely to be executed again very soon.	Nearby instructions to recently executed instruction are likely to be executed soon.
It refers to the tendency of execution where memory location that have been used recently have a access.	It refers to the tendency of execution which involve a number of memory locations .
It is also known as locality in time.	It is also known as locality in space.
It repeatedly refers to same data in short time span.	It only refers to data item which are closed together in memory.

SELF-ASSESSMENT QUESTIONS

1. What does temporal locality refer to in the context of memory hierarchy?

A) The tendency of a program to access data locations at regular intervals.
B) The principle that if a data location is accessed, it is likely to be accessed again in the near future.
C) The practice of accessing data from secondary storage devices.
D) The likelihood of accessing data from the same memory module.

2. Spatial locality in memory access patterns implies that:

A) Data accessed at one moment is unlikely to be accessed again soon.
B) If a storage location is accessed, locations with significantly different addresses are likely to be accessed next.
C) When a storage location is accessed, nearby locations are likely to be accessed in the near future.
D) Data once written to a memory location is not modified or accessed again.

SELF-ASSESSMENT QUESTIONS

3. Which level of memory hierarchy is directly accessible by the CPU and provides the fastest access to data?

A) Hard Disk Drives (HDD)
B) Solid State Drives (SSD)
C) Cache memory
D) Random Access Memory (RAM)

4. Why is the concept of memory hierarchy important in computer architecture?

A) It ensures that all data is stored on the hard drive for permanent storage.
B) It allows the CPU to bypass the RAM and access the hard disk directly for faster data retrieval.
C) It organizes storage media in a way that balances access speed and storage capacity, making efficient use of temporal and spatial locality.
D) It reduces the overall cost of the computer by using cheaper storage options.

TERMINAL QUESTIONS

Short answer questions:

1. Draw the memory hierarchy diagram with increasing distance from CPU.

Long answer questions:

1. Differentiate between the concepts of temporal and spatial locality in memory access patterns.
2. Describe the importance of temporal and spatial locality with respect to memory.

REFERENCES FOR FURTHER LEARNING OF THE SESSION

Reference Books:

1. Computer Organization by Carl Hamacher, Zvonko Vranesic and Saftwat Zaky.
2. Computer System Architecture by M. Morris Mano
3. Computer Organization and Architecture by William Stallings

Sites and Web links:

1. https://www.geeksforgeeks.org/difference-between-spatial-locality-and-temporal-locality/

THANK YOU

(DEEMED TO BE UNIVERSITY)

Team – Digital Design & Computer Architecture