



Assessment 4:

Bachelor of Information Technology

Table of Contents

[1. Introduction 4](#_Toc124709066)

[2. Performance Characteristics of two-level Memories 4](#_Toc124709067)

[2.1. Computer memory 4](#_Toc124709068)

[2.2. Types of computer memory 4](#_Toc124709069)

[2.3. Memory Hierarchy 4](#_Toc124709070)

[2.3.1. CPU Register 5](#_Toc124709071)

[2.3.2. Caches Memory 5](#_Toc124709072)

[2.3.3. Primary Memory 5](#_Toc124709073)

[2.3.4. Secondary Memory 6](#_Toc124709074)

[2.3.5. Tertiary Storage 6](#_Toc124709075)

[2.4. Locality of Reference 6](#_Toc124709076)

[2.5. Operation of Two-Level Memory 6](#_Toc124709077)

[2.6. Performance analysis of two-level memory 7](#_Toc124709078)

[2.7. Virtual Memory 7](#_Toc124709079)

[2.8. Disk Cache 7](#_Toc124709080)

[3. Conclusion 8](#_Toc124709081)

[4. Reference 8](#_Toc124709082)

# Introduction

This report discusses the importance of a computer memory, types of memory, storage, its characteristics, and performance. Each type of computer memory differs from one another in terms of size, capacity, speed, and location. However, the two-level of memory consist of two different memories as computer components, an upper-level memory and lower-level memory.

The principle of locality of reference configures the two-level memory, where specified level has their own properties and speed when storing program’s data and instructions to be processed by processor. More content will be addressed through the report.

# Performance Characteristics of two-level Memories

## Computer memory

Memory is a computer component where instructions and data are stored for immediate use by the operating system.

Computer memory permit the users to store and access data when required. Hence, the data can either be stored temporarily, for a short time as Random Access Memory (RAM) or can be kept for a long term like a Read Only memory (ROM).

Also known as a primary memory, RAM is volatile, it does not retain the content if the computer loses power. On the other hand, the secondary memory is non-volatile which allows the computer to hold information in a long-term basis. Which means, that even when the power is off the data will be saved.

However, computer memory is essential for computer performance and memory size can differ depending on computer architecture and design.

## Types of computer memory

As mentioned before, memory can be classified as primary and secondary memory. Thus, the primary memory is a temporary memory which can be directly accessed by the processor. As an example, it can include RAM, cache memory, and CPU registers. In addition, ROM is a non-volatile memory and considered a primary memory.

Unlike primary memory, the secondary memory is a long-term memory, and it does not allow processor to directly access its memory. To do it so, firstly the data located in secondary memory is accessed and loaded by the main memory, and then used by processor. As an example of secondary memory can include hard disk drive, Optical disk, Magnetic Tape. There are different types of memory discussed through this report.

## Memory Hierarchy

Memory hierarchy has the purpose to organize memory inside of the computer. This mechanism aims to achieve operational efficiency through fast memory access. A computer can present different types of storage, and they can be arrange based on specific features such as costs, speed, sizes, and the responsibilities. Levels of different storage devices are displayed following speeds, from fastest memory to the lowest memory device.

The memory hierarchy design consists of different levels of memory, which include from the top to the bottom: CPU registers, Cache memory (SRAMs), Main Memory (DRAMs), Magnetic Disk (Disk Storage), Optical Disk and Magnetic Tape.

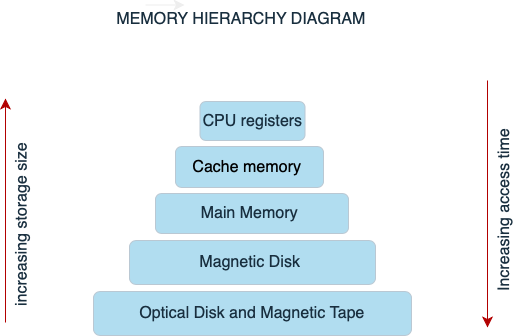


Figure 1. Memory Hierarchy Diagram

### CPU Register

A CPU register corresponds to a small part of the memory in a computer processor. It represents the faster memory compared with others and provides the quickest time to access data in a computer. In size, a CPU register is very small and can be the most expensive to implement in a computer.

CPU registers helps computers to improve computer performance, working in a high speed and empowering the processor to rapidly access data, instructions or other information when needed. Therefore, when the computer wants to execute a program, the data is loaded into the main memory, which addresses the memory location.

### Caches Memory

Caches memory can be localized inside the processor, or between CPU register and main memory. Because of its position cashes memory can be faster and have a high performance which is considered essential as a computer component. The caches memory stores data of programs recently used by processor. Therefore, when the processor wants to review the same data, the cache memory allows quick access to it. If the data cannot be found in the cache memory, the operational system search for the data in the main memory.

### Primary Memory

Main memory is the most important storage unit in computer. The main memory directly exchanges information with the CPU, and peripheral devices using Input/Output processor.

It is considered larger in size and less expensive when compared with CPU registers and cache memory. Speed is also relatively slower than CPU registers and cache memory because primary memory is apart from CPU.

### Secondary Memory

Is a computer memory that permit users to store data or information such as files, programs, etc. that is not frequently accessed by processor. Information is indirectly transferred between RAM and Secondary memory through Input/Output programs. Secondary memory has a much slower speed capacity when compared to the main memory. It’s also cheaper in price than RAM. As larger in size, this memory can have capacity up to 20 terabytes.

### Tertiary Storage

This type of storage can be a magnetic tape or optical disk. They are used to store archive records of data, when it’s not required for use, and can be stored for a long period of time. They represent the cheapest memory choice, also presents the lowest performance as well. Their capacity is between 1TB to 20TB.

## Locality of Reference

According to Stallings (p.62, 2018) “The Principle of Locality states that memory references tend to cluster.” It describes a processor’s tendency to access a specific set of memory locations for a given period repeatedly. In addition, the result of a program’s language studies shows that most programs tend to execute sequentially; it can present few instructions that repeats all the time, and it can be closely located to one another. Locality predicts behaviours that happens within the computer system, and consequently can contribute to performance optimization between memory and processor operation using some techniques.

There are two distinct types of localities, Spatial locality, and Temporal locality.

Spatial location is based on the idea that if a current memory location has been referenced, then it is likely that nearby data or instructions’ storage locations will be referenced in the future.

Temporal locality defends the idea that current data or instructions that’s been executed by processor will be referenced again in the future. Therefore, to avoid accessing the main memory again, the data should be stored in the memory cache.

## Operation of Two-Level Memory

Stallings (2018) mentioned the two-level memory as a result of the property of locality, where two level of memory is formed by locality reference.

The upper-level memory is nominated as M1 and the lower-level is known as M2. M1shows a higher performance than the lower-level memory M2. It happens because M1 is faster and smaller in size, which make it more expensive as well.

M1 acts as a temporary storage for M2, which is larger in size than M1. When a memory reference is made from coping a block of memory locations from M2 to M1, it allows the processor to quickly access data from M1. In overall, this process increases services within the computer.

The overage time to access an item can be expressed by the expression bellow:

Ts = H X T1 + (1-H) X (T1+T2)

Ts = average access time

H = hit ratio

T1 = M1 access time

T2 = M2 access time

Therefore, access efficiency can be expressed by:

T1 = 1 .

Ts 1 + (1 - H) T2/T1

When hit ratio is close to 1 it means a positive improvement.

## Performance analysis of two-level memory

To analyse performance of a two-level memory mechanism, it is a must to consider costs of operation. There is an expression that allows to calculate average costs for two level memory as shown below:

Cs = C1S1 + C2S2

S1 + S2

Cs = average cost per bit

C1 = average cost per bit for M1

C2 = average cost per bit for M2

S1 = size of M1

S2 = size of M2

Based on studies, Stallings (2018) affirms that when a strong locality is in place, even if the upper level (M1) has a small memory size it is still possible to achieve high hit ratio. In addition, it’s shown that small cache sizes can have a hit ratio of 0.75 independent of main memory size.

## Virtual Memory

Virtual memory is a logical storage location that behaves as an extension of the main memory using a secondary memory. It works as a memory management technique to increase capacity of main memory and it’s managed by the operational system.

Virtual memory allows to run multiple programs at the same time, or to execute a large program even when program’s size is bigger than the main memory.

Virtual memory uses a mapping structure that stores virtual address and corresponding physical address.

When an application is executed, main memory stores data from a program to a physical address. Then MMU - memory management unit translates addresses to virtual addresses using physical addresses as reference. Data can be swapped in and out between RAM and virtual memory, it makes RAM available to next processes to be executed.

## Disk Cache

A disk cache is a software mechanism which is part of the hard disk. It allows to accelerate the process of storing and accessing data from a hard disk driver. When data is accessed by the system from a hard disk, a copy of the previous data is kept in the disk cache. That way, when the operating system requests the same data, it firstly checks in the disk cache. This process can provide a fast reading without accessing the hard disk which would be much slower, taking more time.

There is also a soft disk cache which is implemented on the Main memory instead of hard disk. Disk cache increases performance through the speed when transferring data along hard disk and computer system.

# Conclusion

In conclusion, the memory is a vital component of the computer. It enables the operational system to improve computer performance through a fast-processing speed. Depending on the memory characteristics, the computer can transfer, store and access data when needed in a more efficient way. The principle of locality allows two-level of memory to identify the tendency of processors to access the same data and instruction from a memory location when executing a program in a computer. This mechanism can help to improve efficiency of the computer performance as well. The primary memory and the secondary memory support the operational system to store and access data quickly and easily. As shown by the memory hierarchy there are different levels of memory devices inside the computer that can be organized to increase access time and to improve storage capacity. Depending on the characteristics of each type of memory device you can achieve different results.

# Reference

Stallings , W. (2017) *Operating System - Internal and Design Principles*. 9th edn. Edinburgh Gate, Harlow: Pearson Education.

*Levels of memory in operating system* (2020) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/levels-of-memory-in-operating-system/ (Accessed: January 5, 2023).

*Memory hierarchy design and its characteristics* (2018) *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/memory-hierarchy-design-and-its-characteristics/ (Accessed: January 5, 2023).

*Computer - memory* (no date) *Tutorials Point*. Available at: https://www.tutorialspoint.com/computer\_fundamentals/computer\_memory.htm (Accessed: January 6, 2023).

Gillis, A.S. (2020) *What is computer memory and what are different types?*, *WhatIs.com*. TechTarget. Available at: https://www.techtarget.com/whatis/definition/memory (Accessed: January 6, 2023).

Lutkevich, B. (2020) *What is cache memory? cache memory in computers, explained*, *Storage*. TechTarget. Available at: https://www.techtarget.com/searchstorage/definition/cache-memory (Accessed: January 7, 2023).

*What is a disk cache? - definition from Techopedia* (no date) *Techopedia.com*. Available at: https://www.techopedia.com/definition/6867/disk-cache (Accessed: January 9, 2023).

*Locality of reference and cache operation in Cache Memory* (2019) *GeeksforGeeks*. GeeksforGeeks. Available at: https://www.geeksforgeeks.org/locality-of-reference-and-cache-operation-in-cache-memory/ (Accessed: January 9, 2023).

J. Denning, P. (no date) *The locality principle - denning institute*. Available at: https://denninginstitute.com/pjd/PUBS/CACMcols/cacmJul05.pdf (Accessed: January 10, 2023).

*Difference between cache memory and virtual memory - javatpoint* (no date) *www.javatpoint.com*. Available at: https://www.javatpoint.com/cache-memory-vs-virtual-memory (Accessed: January 10, 2023).

*Virtual memory* (no date) *Virtual Memory - an overview | ScienceDirect Topics*. Available at: https://www.sciencedirect.com/topics/computer-science/virtual-memory (Accessed: January 11, 2023).