Current and Future Technologies in Computer Memories

Bhavneet Soni

Author’s Note

In 1981, referring to computer memory, Bill Gates said, “640K (roughly 1⁄2 of a megabyte) ought to be enough for anybody.” Now consider this Computer used to write this report has a memory of 32 Giga Bytes and still can do with more. Computer memory has come a long way from vacuum tube based memory that could handle only a few bits of data a time, to terabytes of data in your palm. In addition, the physical size of memory has reduced from machines housed in big halls to small chips in your phones. In this report, we talk about some of the technological advancements that have taken in memory technologies over the years and future trend.

**Current and Future Technologies in Computer Memories**

Computer memory has seen a lot of growth over the years, storage capacity and reliability has increased tremendously over the years. Computers from back in 1950s to till date, as the amount and processing increased so did the need to handle and store more and more data. Earlier memory technologies used vacuum tube, punch cards, transistors, magnetic tapes and drums to handle the data with the advent of ICs and Microchips we have more storage than before. The 1st generation computers between 1940 and 1955 were powered by vacuum tubes and used magnetic drums; vacuum tube controlled the electric current through a sealed container made of thin transparent glass in a rough cylinder shape, magnetic drum cylinder coated with magnetic iron-oxide material, which stored data and programs. John Mauchly and John Presper Eckert created one of the earliest electric calculators ENIAC (Electric Numerical Integrator and Computer). It had 17,468 vacuum tubes, 70,000 resistors, 10,000 capacitors, 1,500 relays, 6,000 manual switches and 5 million soldered joints. It covered 1800 square feet of floor space and weighed 30 tons. ENIAC could workout 5,000 additions, 357 multiplications or 38 divisions in one second. It took weeks to reprogram the machine to undertake a different task and would require a lot of maintenance time. These machines were very expensive to operate and had limited application capabilities. They also huge power consumption and generated a lot of heat, the ENIAC used 160 kilowatts of electrical power. With advancement of electronics and advent of transistors and ICs the size and power consumption of these machines reduced dramatically and there capabilities increased exponentially. We have created worlds 90% data in last decade alone. Every day about 2.5 Quintillion bytes of data is created, which is equivalent to about 100 million Blu-ray discs, if stacked one top of another would be greater than the height of four Eiffel Towers staked up one top of another (Walker, 2015). How the data increased, it has changed the way we store and access data now, memory and storage have moved from big machines to small chips in computer to the cloud. By 2020, there will be 5200 GB of data per person and 15% of it will be stored on the cloud. (Mearian, 2012)

People often confuse between memory and storage, while both manage data in a computing environment they do not necessarily mean the same thing, while memory holds the data that the computer needs to work on and execute storage holds data for longer terms. Best analogy to understand the difference between Memory and Storage would be how things are at your office, say you are working in an office, and you have a number of files that your work on, however you can keep only a few files on your desk that you are currently working on and there are files in the filing cabinet you will work on later I the day or in future, so all the files on your desk are the files in memory and all the files in the filing cabinet are the files in storage. (The difference between memory and storage | Technick.net, 1998). Accessing data from memory is a lot faster than accessing from storage on an average it takes about 200 nanoseconds (ns) for the computer to access data from memory RAM while it takes about 12,000,000 ns to access the same data from storage (hard disk), which is equivalent to taking 4.5 months to complete a task that can be done in 3.5 minutes. (The difference between memory and storage | Technick.net, 1998). Memory is relatively more expensive than storage and not off of the data can be held in the memory Hence it created a need to differentiate between what CPU needs access to work and what needs to store. Figure 1.0, below shows how CPU requests and gets the data that it needs to process.

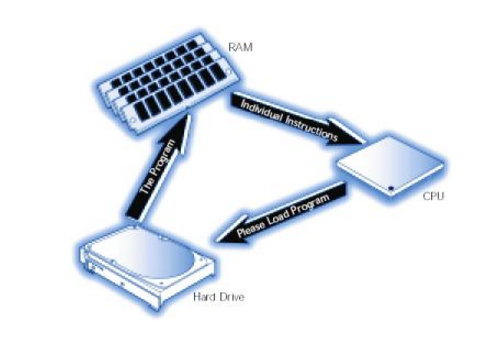


Figure Movement of data from Storage to CPU via memory

In reference to modern computers The term "memory", "primary storage" or "main memory", means addressable semiconductor memory, i.e. integrated circuits consisting of silicon-based transistors, used for example as primary storage but also other purposes in computers and other digital electronic devices. Data is organized into memory cells or bitable flip-flops, each storing one bit (0 or 1). Flash memory organization includes both one bit per memory cell and multiple bits per cell (called MLC, Multiple Level Cell). The memory cells are grouped into words of fixed word length, for example 1, 2, 4, 8, 16, 32, 64 or 128 bit. Each word can be accessed by a binary address of N bit, making it possible to store 2 raised by N words in the memory.

n early computer systems, programs typically specified the location to write memory and what data to put there. This location was a physical location on the actual memory hardware. The slow processing of such computers did not allow for the complex memory management systems used today. Also, as most such systems were single-task, sophisticated systems were not required as much.

This approach has its pitfalls. If the location specified is incorrect, this will cause the computer to write the data to some other part of the program. The results of an error like this are unpredictable. In some cases, the incorrect data might overwrite memory used by the operating system. Computer crackers can take advantage of this to create viruses and malware.Computer memories are broadly categorized as

The memory structure of PCs is often thought of as just main memory, but it's really a five or six level structure:

1. Primary Memory - Temporary or Volatile Memory
   1. Cache
   2. RAM
      1. Physical RAM
         1. DRAM
         2. SDRAM
         3. DDR RAM
         4. SRAM
         5. DRDRAM
      2. Virtual RAM
2. Secondary Memory also known as Permanent or Non Volatile Memory
   1. BIOS
   2. ROM
   3. Hard Disk
3. Auxiliary memory or Removable Memory

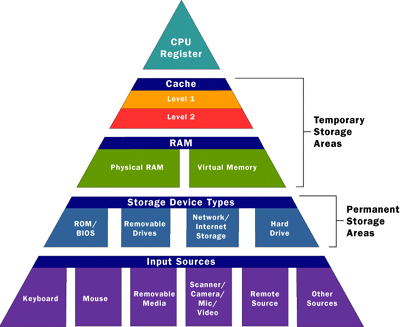


Figure 2. Various levels of Storages (Tyson, 2000)

### 1. Primary Memory / Volatile Memory:

Primary Memory is also known as volatile memory, that is the data stored in it discarded as soon as the power is turned off. If power failures happened in systems during memory access, then you will lose your data permanently. Examples of volatile memory are primary storage, which is typically dynamic random-access memory (DRAM), and fast CPU cache memory, which is typically static random-access memory (SRAM) that is fast but energy-consuming, offering lower memory areal density than DRAM. Although CPU registers are also a form of volatile data storage but registers are normally not considered as memory, as they only store one word and there is no addressing mechanism in CPU registers.

The fastest memory available to the computer is the Cache memory, it servers as buffer between the processor and RAM, as the cache name suggests the data is held in temporarily so that when CPU needs it does not have to go looking for it in the main memory, hence are a lot faster and can has a lot of impact on computer performance.

There are various levels of caches (L1, L2 and L3) available based on their location respective to CPU and main memory. L1 cache is present on th4e CPU itself whereas L@ cache can be on CPU or mother board depending on the computer architecture. Advantage of caches on the CPU is that it runs at the same speed as CPU and cost less as there is no need to set up an external bus and logic.

Main memory of a computer is called RAM (Random Access Memory), as the memory locations are selected randomly and the address is stored in register. It can perform read or write operations on memory in the same amount of time irrespective of the location of the data in the memory. RAM uses multiplexing and deplexing circuitry to connect the data lines to be addressed. categorized into following types (InformationQ, 2015).

1. DRAM – Dynamic RAM, each bit of data is stored on a separate capacitor, and they get discharged when power is removed so they have to be refreshed periodically hence the name dynamic. These are structurally simple one transistor and a capacitor is required to store one bit which allows for relatively low cost and have high memory capacity. DRAM is usually used as the main memory in computers and other components such as graphic cards. Due to dynamic refresh of the data these consume large amounts of power. Modern day DRAMs are setup on a series of dynamic RAM memory modules known as dual in line memory module (DIMM) which replaced older single line memory modules (SIMM) which were predominantly used in Pentium based computer sets. DIMMs have a 64-bit bus width installed in matched pairs giving access to two modules in parallel
   1. SDRAM – Synchronous DRAM Synchronous DRAM technology by the end of 1998, 100MHz SDRAM (PC100) became the industry standard for mainstream PC and servers. PC133 SDRAM – a straightforward incremental speed upgrade to PC100 memory. This has the advantage that modules can be used in existing 100MHz systems. In the autumn of 1999 Intel finally agreed to adopt the PC133 standard and announced the intention to produce a PC133 chipset in the first half of 2000.
   2. Double Data Rate SDRAM (DDR-SDRAM), runs at the same speed ass 100 or 133 MHz as SDRAM but uses a more advanced synchronization which allows for double the available bandwidth to 2.1 GBps. has the advantage of running at the same bus speed (100MHz or 133MHz) as SDRAM, but by using more advanced synchronisation is able to double the available bandwidth to 2.1GBps. A development of this approach is DDR2-SDRAM, offering up to 4.8 GBps, which is backed by a new consortium of DRAM manufacturers known as Advanced Memory International. This group is effectively a rechartered incarnation of the SyncLink DRAM (SLDRAM) Consortium, which in 1998 promoted a scaleable packet-based technology very similar to Direct Rambus, but as a royalty-free standard.
2. SRAM – Static RAM faster and used in Cache memories 4 or 6 capacitors and transistors are required for storing 1 bit
3. DRDRAM - A California based company Rambus invented an alternative to SDRAM, Direct Rambus it was a highly scale able and pipelined memory architecture and had backing from the industry giant Intel. However due to high royalty costs to the paid to Rambus technology and up to 50% price premium for DRDRAM modules (RIMMs) over SDRAM it didn’t gained much traction and PC manufacturers looked for an alternate to an interim higher-bandwidth, lower latency memory, particularly for use in servers and workstations.

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **Machine** | **Typical RAM** | **× Apple I** |
| 1976 | Apple I | 8KB | 1 |
| 1977 | Apple ][ | 24KB | 3 |
| 1980 | Apple III | 128KB | 16 |
| 1984 | Macintosh | 256KB | 32 |
| 1986 | Mac Plus | 1MB | 125 |
| 1992 | Mac LC | 10MB | 1250 |
| 1996 | PowerMac | 16MB | 2000 |
| 1998 | iMac | 32MB | 4000 |
| 2007 | iPhone | 128MB | 16000 |
| 2010 | iPhone 4 | 512MB | 64000 |
| 2016 | iPhone 7 | 3GB | 375000 |

Table Evolution of RAM capacity over the years (Woodford, 2010)

Due to high cost of RAM operating systems makes use of page files and control physical memory creating Virtual memory. When a program needs memory, it requests it from the operating system. The operating system then decides what physical location to place the memory in.

This offers several advantages. Computer programmers no longer need to worry about where the memory is physically stored or whether the user's computer will have enough memory. It also allows multiple types of memory to be used. For example, some memory can be stored in physical RAM chips while other memory is stored on a hard drive. This drastically increases the amount of memory available to programs. The operating system will place actively used memory in physical RAM, which is much faster than hard disks. When the amount of RAM is not sufficient to run all the current programs, it can result in a situation where the computer spends more time moving memory from RAM to disk and back than it does accomplishing tasks; this is known as thrashing.

Virtual memory systems usually include protected memory, but this is not always the case.

## 2. Secondary Memory / Non Volatile Memory:

Secondary memory is external and permanent memory that is useful to store the external storage media such as floppy disk, magnetic disks, magnetic tapes, CDs, DVDs, flash drives etc. Secondary memory deals with following types of components. Examples of non-volatile memory are flash memory (used as secondary memory) and ROM, PROM, EPROM and EEPROM memory (used for storing firmware such as BIOS).

1. Read Only Memory (ROM) ROM is permanent memory location that offer huge types of standards to save data. But it work with read only operation. No data lose happen whenever power failure occur during the ROM memory work in computers. hey're preprogrammed with information in the factory and used to store things like the computer's BIOS (the basic input/output system that operates fundamental things like the computer's screen and keyboard).
2. ROM memory has several models such names are following.

**1. PROM:** Programmable Read Only Memory (PROM) maintains large storage media but can’t offer the erase features in ROM. This type of RO maintains PROM chips to write data once and read many. The programs or instructions designed in PROM can’t be erased by other programs.

**2. EPROM :** Erasable Programmable Read Only Memory designed for recover the problems of PROM and ROM. Users can delete the data of EPROM thorough pass on ultraviolet light and it erases chip is reprogrammed.

**3. EEPROM:** Electrically Erasable Programmable Read Only Memory similar to the EPROM but it uses electrical beam for erase the data of ROM.

**Cache Memory:** Mina memory less than the access time of CPU so, the performance will decrease through less access time. Speed mismatch will decrease through maintain cache memory. Main memory can store huge amount of data but the cache memory normally kept small and low expensive cost. All types of external media like Magnetic disks, Magnetic drives and etc store in cache memory to provide quick access tools to the users.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **PC100** | **PC133** | **DDR**  **SDRAM** | **SLDRAM** | **Base**  **Rambus** | **Concurrent**  **Rambus** | **Direct**  **Rambus** |
| **Announced**  **Frequency**  **(MHz)** | 100 | 133 | 200/266 | 800 | 700 | 700 | 600/800 |
| **Maximum**  **Bandwidth**  **(GBps)** | 0.80 | 1.00 | 1.6/2.1 | 1.60 | 0.70 | 0.70 | 1.2/1.6 |
| **Expected**  **Bandwidth**  **(GBps)** | 0.50 | 0.60 | 0.9/1.2 | unknown | 0.40 | 0.50 | 1.1/1.5 |
| **Efficiency (%)** | 65 | 60 | 60 | unknown | 60 | 80 | 97 |
| **Data Width**  **(bits)** | 64 | 64 | 64 | 16 | 8/9 | 8/9 | 16/18 |

Volatile memory[[edit](https://en.wikipedia.org/w/index.php?title=Computer_memory&action=edit&section=2)]

Various memory modules containing different types of DRAM (from top to bottom): DDR SDRAM, SDRAM, EDO DRAM, and FPM DRAM

*Main article:*[*Volatile memory*](https://en.wikipedia.org/wiki/Volatile_memory)

Volatile memory is computer memory that requires power to maintain the stored information. Most modern [semiconductor](https://en.wikipedia.org/wiki/Semiconductor) volatile memory is either static RAM ([SRAM](https://en.wikipedia.org/wiki/Static_random_access_memory)) or dynamic RAM ([DRAM](https://en.wikipedia.org/wiki/DRAM)). SRAM retains its contents as long as the power is connected and is easy for interfacing, but uses six transistors per bit. Dynamic RAM is more complicated for interfacing and control, needing regular refresh cycles to prevent losing its contents, but uses only one transistor and one capacitor per bit, allowing it to reach much higher densities and much cheaper per-bit costs.

SRAM is not worthwhile for desktop system memory, where DRAM dominates, but is used for their cache memories. SRAM is commonplace in small embedded systems, which might only need tens of kilobytes or less. Forthcoming volatile memory technologies that aim at replacing or competing with SRAM and DRAM include [Z-RAM](https://en.wikipedia.org/wiki/Z-RAM) and [A-RAM](https://en.wikipedia.org/wiki/A-RAM).

Non-volatile memory[[edit](https://en.wikipedia.org/w/index.php?title=Computer_memory&action=edit&section=3)]

[Solid-state drives](https://en.wikipedia.org/wiki/Solid-state_drive) are one example of non-volatile memory.

*Main article:*[*Non-volatile memory*](https://en.wikipedia.org/wiki/Non-volatile_memory)

Non-volatile memory is computer memory that can retain the stored information even when not powered. Examples of non-volatile memory include read-only memory (see [ROM](https://en.wikipedia.org/wiki/Read-only_memory)), [flash memory](https://en.wikipedia.org/wiki/Flash_memory), most types of magnetic computer storage devices (e.g. [hard disk drives](https://en.wikipedia.org/wiki/Hard_disk_drive), [floppy disks](https://en.wikipedia.org/wiki/Floppy_disk) and [magnetic tape](https://en.wikipedia.org/wiki/Magnetic_tape)), [optical discs](https://en.wikipedia.org/wiki/Optical_disc), and early computer storage methods such as [paper tape](https://en.wikipedia.org/wiki/Paper_tape) and [punched cards](https://en.wikipedia.org/wiki/Punched_card).

Forthcoming non-volatile memory technologies include [FeRAM](https://en.wikipedia.org/wiki/FeRAM" \o "FeRAM), [CBRAM](https://en.wikipedia.org/wiki/CBRAM), [PRAM](https://en.wikipedia.org/wiki/Parallel_Random_Access_Machine), [SONOS](https://en.wikipedia.org/wiki/SONOS), [RRAM](https://en.wikipedia.org/wiki/RRAM), [racetrack memory](https://en.wikipedia.org/wiki/Racetrack_memory), [NRAM](https://en.wikipedia.org/wiki/Nano-RAM), [3D XPoint](https://en.wikipedia.org/wiki/3D_XPoint), and [millipede memory](https://en.wikipedia.org/wiki/Millipede_memory).

Modern general-purpose computer systems manage memory at two levels: at the system level (see [memory management (operating systems)](https://en.wikipedia.org/wiki/Memory_management_(operating_systems))); and at the application level (as discussed in this article). Application-level memory management is generally categorized as either automatic memory management, usually involving [garbage collection (computer science)](https://en.wikipedia.org/wiki/Garbage_collection_(computer_science)), or[manual memory management](https://en.wikipedia.org/wiki/Manual_memory_management).

Several methods have been devised that increase the effectiveness of memory management. [Virtual memory](https://en.wikipedia.org/wiki/Virtual_memory) systems separate the memory addresses used by a process from actual physical addresses, allowing separation of processes and increasing the size of the [virtual address space](https://en.wikipedia.org/wiki/Virtual_address_space) beyond the available amount of [RAM](https://en.wikipedia.org/wiki/Random-access_memory) using [paging](https://en.wikipedia.org/wiki/Paging) or swapping to [secondary storage](https://en.wikipedia.org/wiki/Secondary_storage). The quality of the virtual memory manager can have an extensive effect on overall system performance.

References

InformationQ. (2015, May 19). *Two types Computer Memory | Primary and Secondary Memory |*. Retrieved from http://www.informationq.com: http://www.informationq.com/computer-memory/

Mearian, L. (2012, December 11). *By 2020, there will be 5,200 GB of data for every person on Earth | Computerworld*. Retrieved from Computerworld.com: http://www.computerworld.com/article/2493701/data-center/by-2020--there-will-be-5-200-gb-of-data-for-every-person-on-earth.html

*The difference between memory and storage | Technick.net*. (1998, March 01). Retrieved from technick.net: https://technick.net/guides/hardware/umg/01\_003/

Tyson, J. (2000, August 23). *How Comuter Memory works*. Retrieved from http://howstuffworks.com: http://computer.howstuffworks.com/computer-memory.htm

Walker, B. (2015, April 5). *Every Day Bid Data*. Retrieved from http://www.vcloudnews.com/: http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily

Woodford, C. (2010). *Computer memory*. Retrieved from www.explainthatstuff.com: http://www.explainthatstuff.com/how-computer-memory-works.html