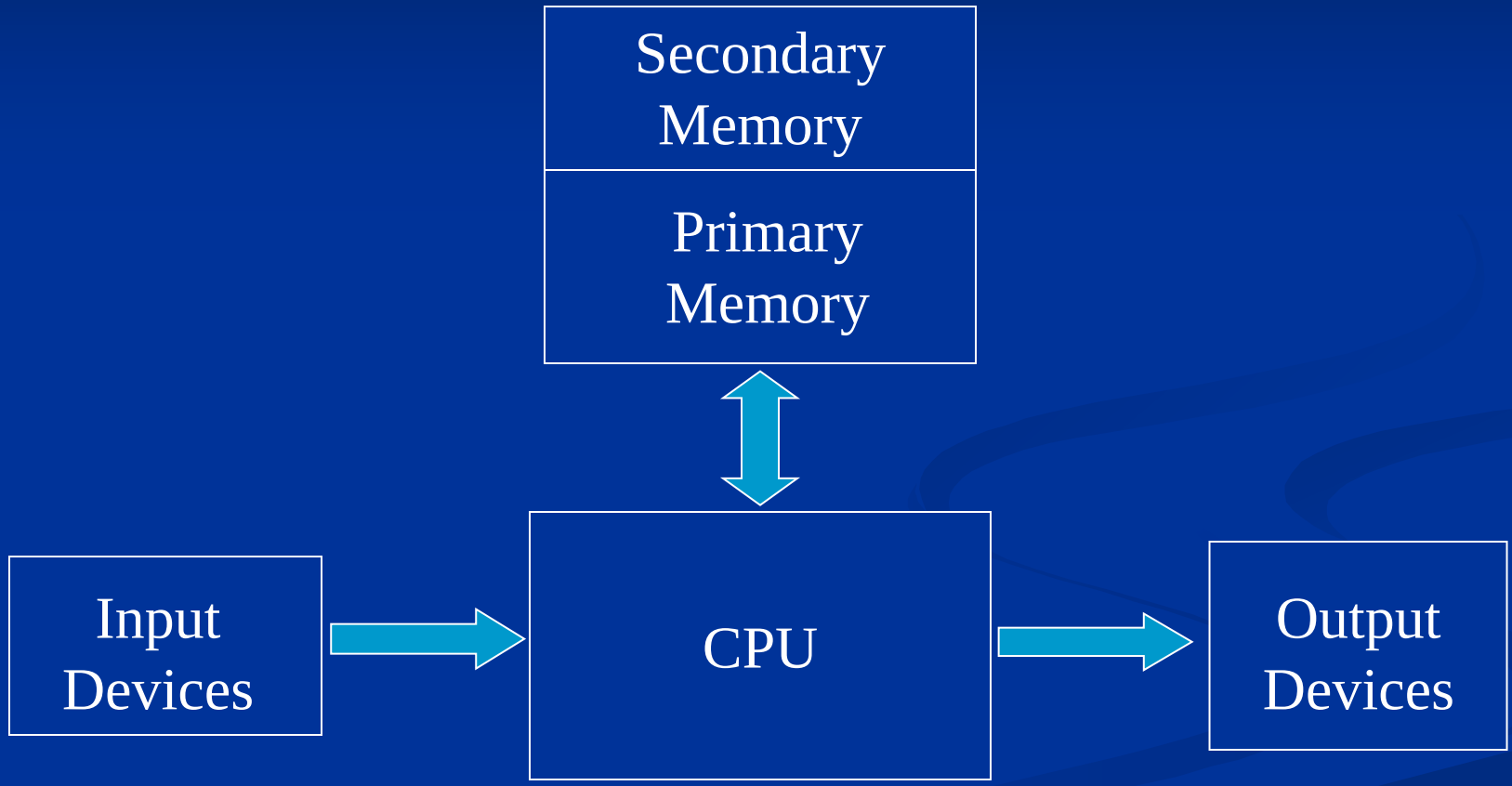
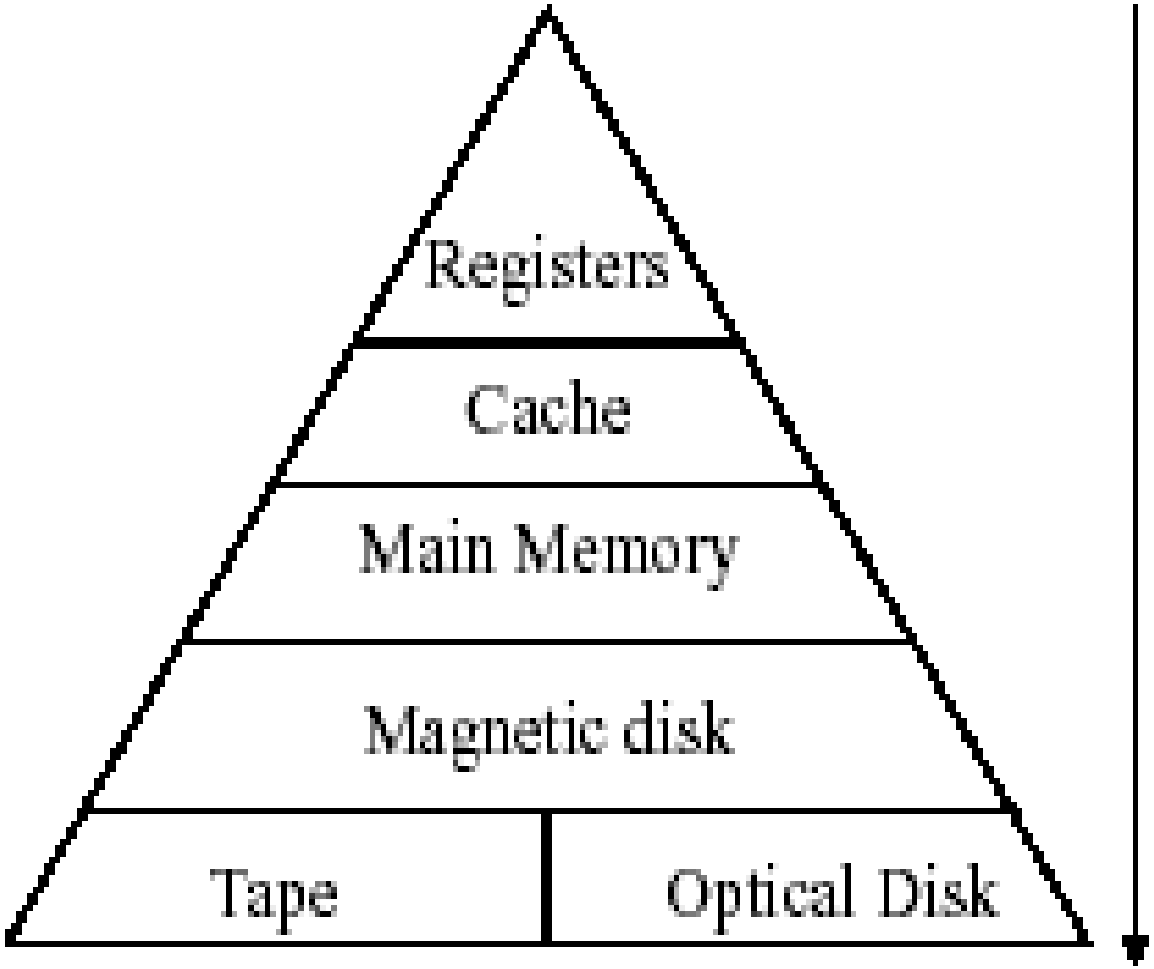
Memory

Memory Basics



Memory Hierarchies



- Decrease cost/bit
- Increase capacity
- Increase access time
- Decrease frequency of access via CPU
- Further away from CPU

Memory can be categorized into

- Registers
- primary memory (e.g. cache, main memory)
- secondary memory (e.g. disks)

- Registers and primary or main memory are the purely electronic and volatile memory that a computer uses to store current data and code while executing a program or application.

- Primary memory works in tandem with the CPU to store data, programs, and processed information that can be made immediately and directly accessible to the CPU or to other system devices.
- Primary memory also determines the size and number of programs that can be run simultaneously.

Memory Speed

- CPU registers can be accessed in a few nanoseconds (x10-9).
- Cache memories take a small multiple of CPU registers.
- Main memory accesses are typically a few tens of nanoseconds.
- Disk access are at least 10msec (x10-3)!
- Tape or optical disk access time can be measured in seconds.

Register Memory

Registers:

❖Are used to store temporary and partial results of calculations and sometimes system state information.

❖Reside on the CPU chip.

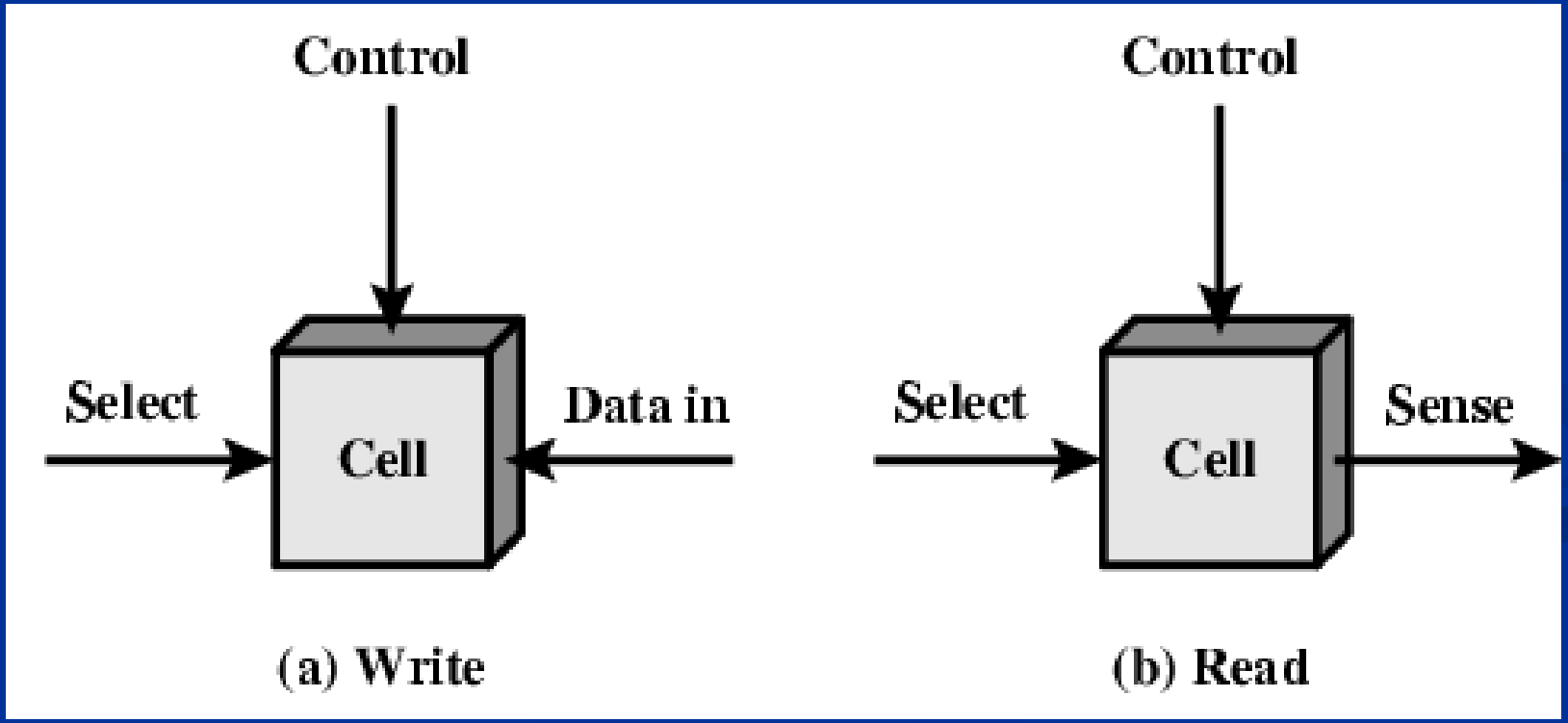
❖Are fast for CPU to access.

❖They may be as small as one bit or as large as several words depending upon the processor.

❖Because registers reside on the CPU chip itself there is no room for too many of them. Most modern processors have between 32 and 64 registers onboard.

Primary Memory

Computer’s primary memory is arranged as a matrix of"memory cells" laid out in rows and columns, like squares on a checkerboard.



- Write Operation
  - Select the Memory Cell :- Address Bus
  - Data-in :- Data Bus
  - Write Signal :- Control Bus
- Read Operation
  - Select the Memory Cell:- Address Bus
  - Read Signal:- Control Bus
  - Sense the data :- Data Bus

$$\text{Number of addresses} = 2^{\text{Address bus width}}$$

Capacity of memory = Number of addresses x Data bus width

Calculate the capacity (Size) of the memory

If, Address bus width= 8 bits

Data bus width= 8 bits

Calculate the capacity of the a memory

If, Address bus width= 16 bits

Data bus width= 8 bits

prefixes

$2^{10} = \mathrm{k}$

$2^{20} = \mathrm{M}$

Calculate the capacity of the a memory

If, Address bus width= 32 bits

Data bus width= 8 bits

- In a memory there are 16M address. What is the width of the address bus?

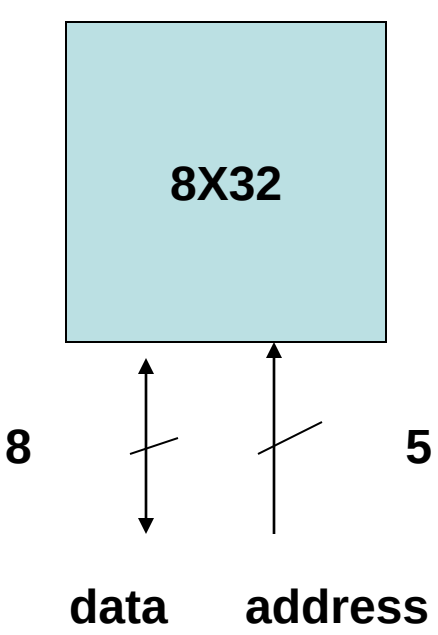
Calculate the capacity of the a memory

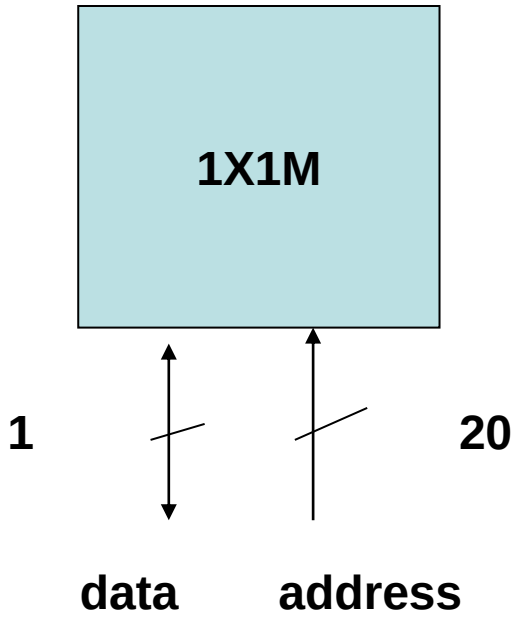
If, No of address= 8M

Data bus width= 8 bits

RAM/ROM Naming Convention

- Examples 8X32, 8 by 32➔ 32 8-bit words
- 1X1M, 1 by 1M➔ 1M 1-bit words





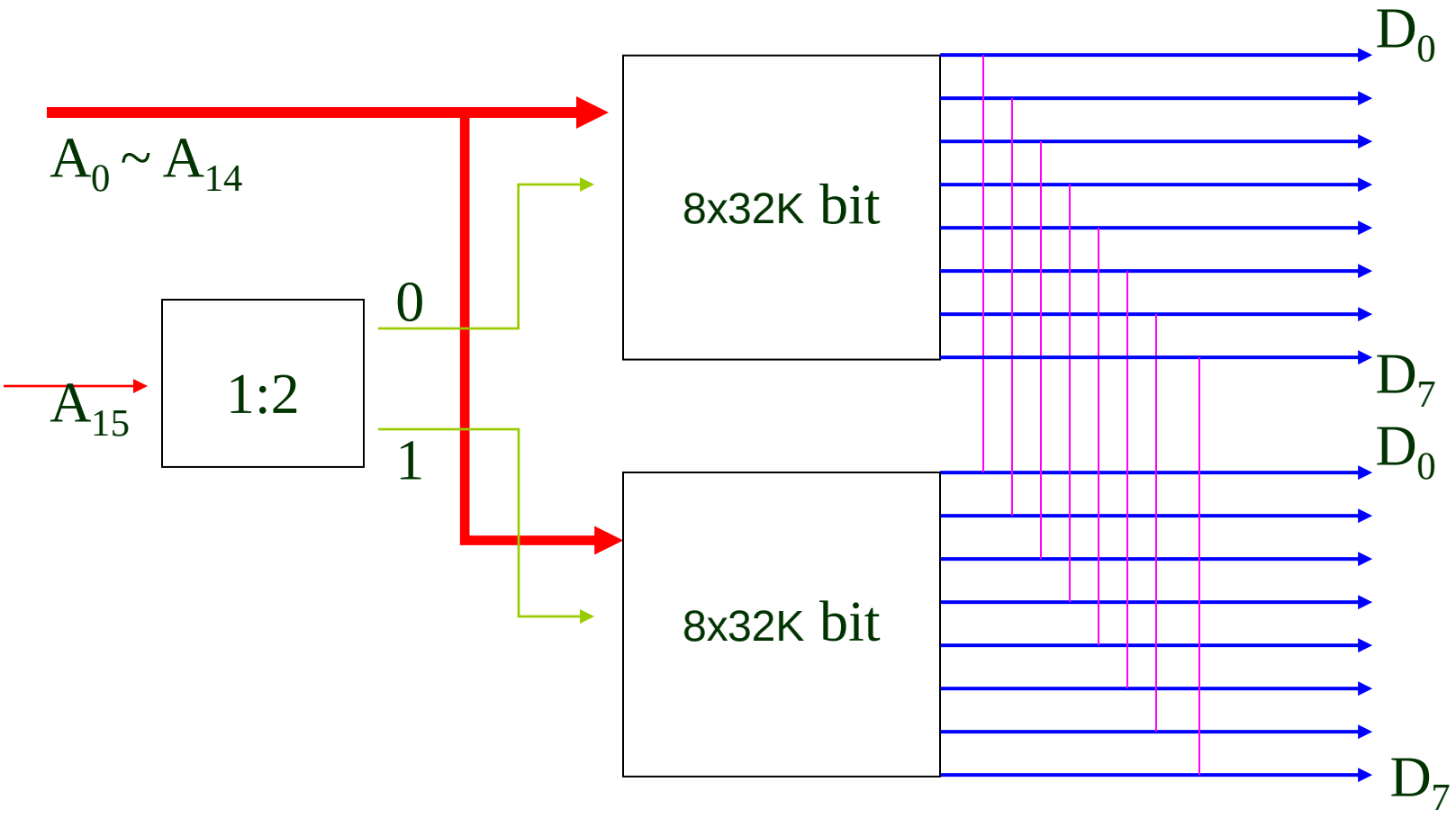
Memory Expansion

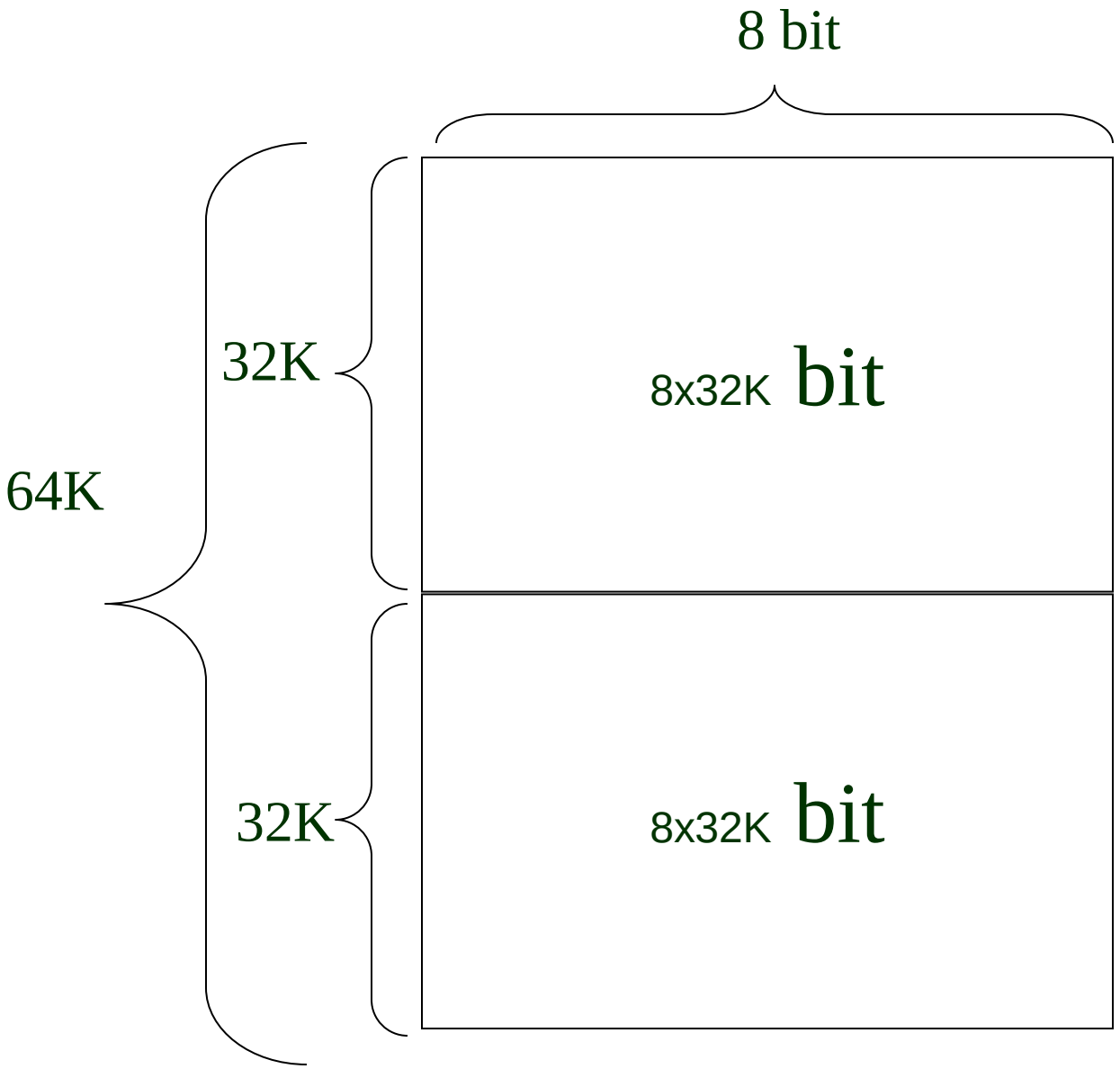
This refers to the construction of a larger memory in the sense of memory address space (No of memory locations) or word length

Address Expansion.

Ex1:

Construct 8x64K memory using 8x32K units


D_0
$A_0 \sim A_{14}$
8x32K bit
0
1:2
A_{15}
D_7
D_0
1
8x32K bit
D_7


8 bit
64K
32K
8x32K bit
32K
8x32K bit

Ex2:

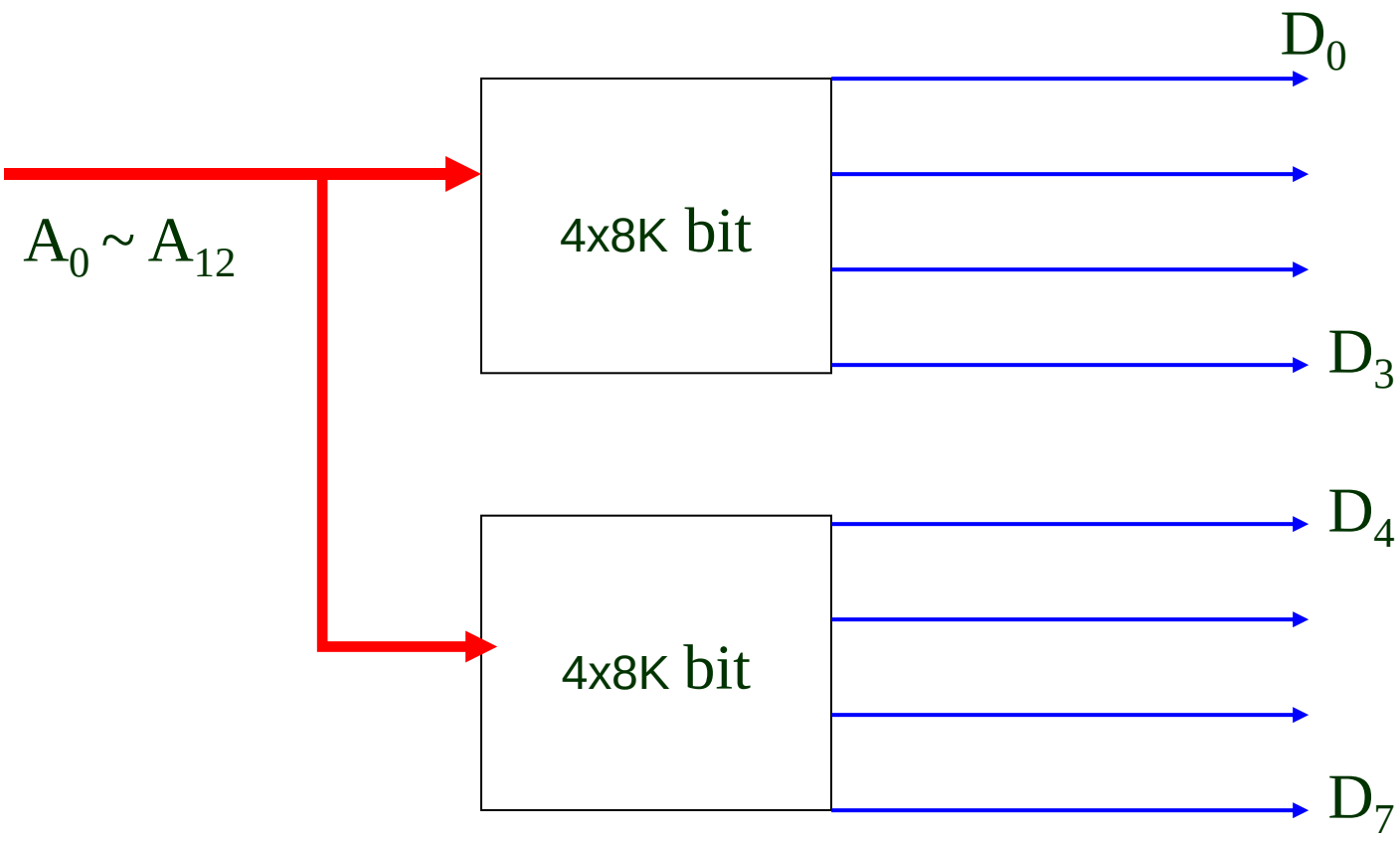
Construct 4x32K memory using 4x8K units

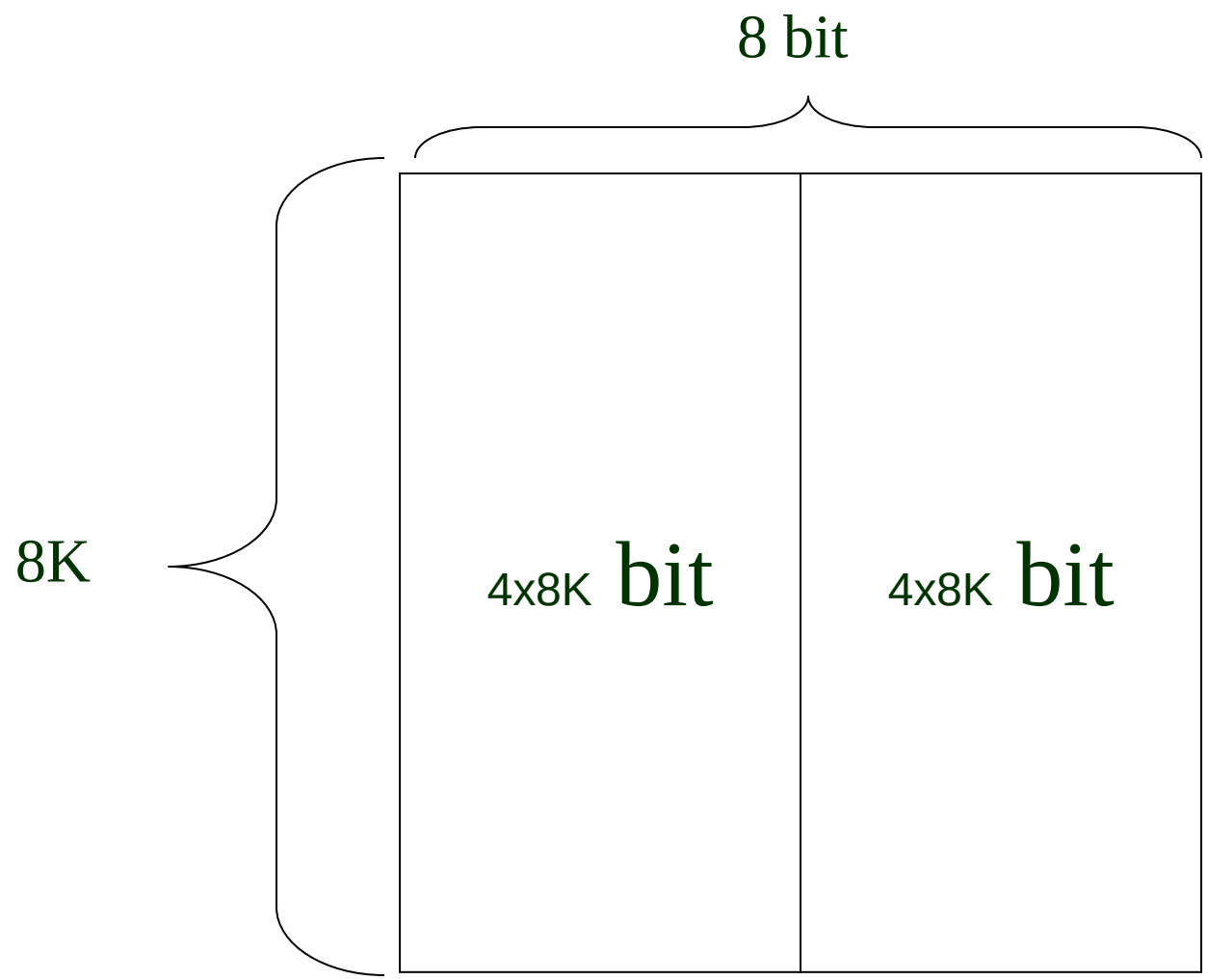
Ex3:

Construct 2x64K memory using 2x8K units

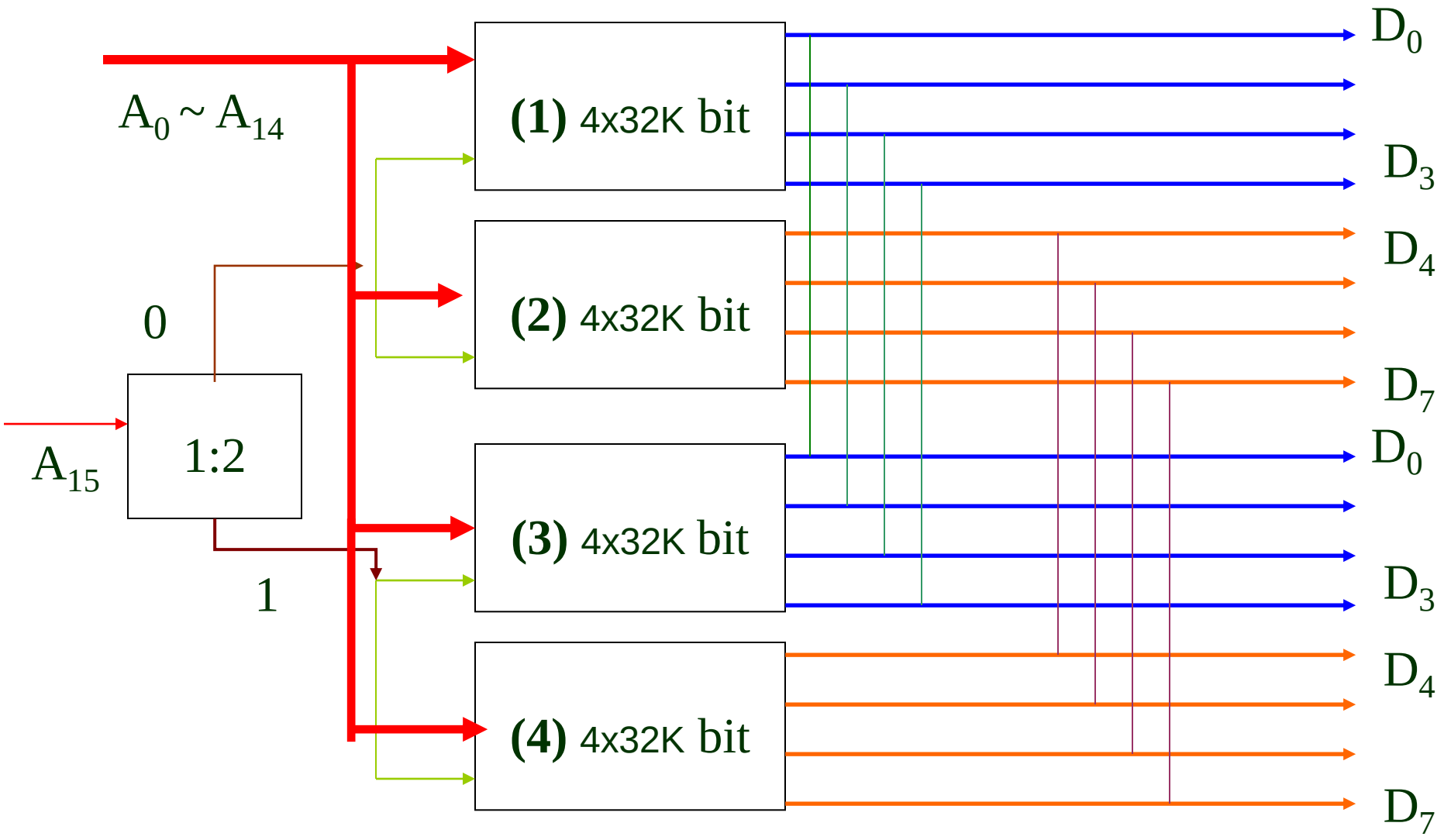
Word expansion.

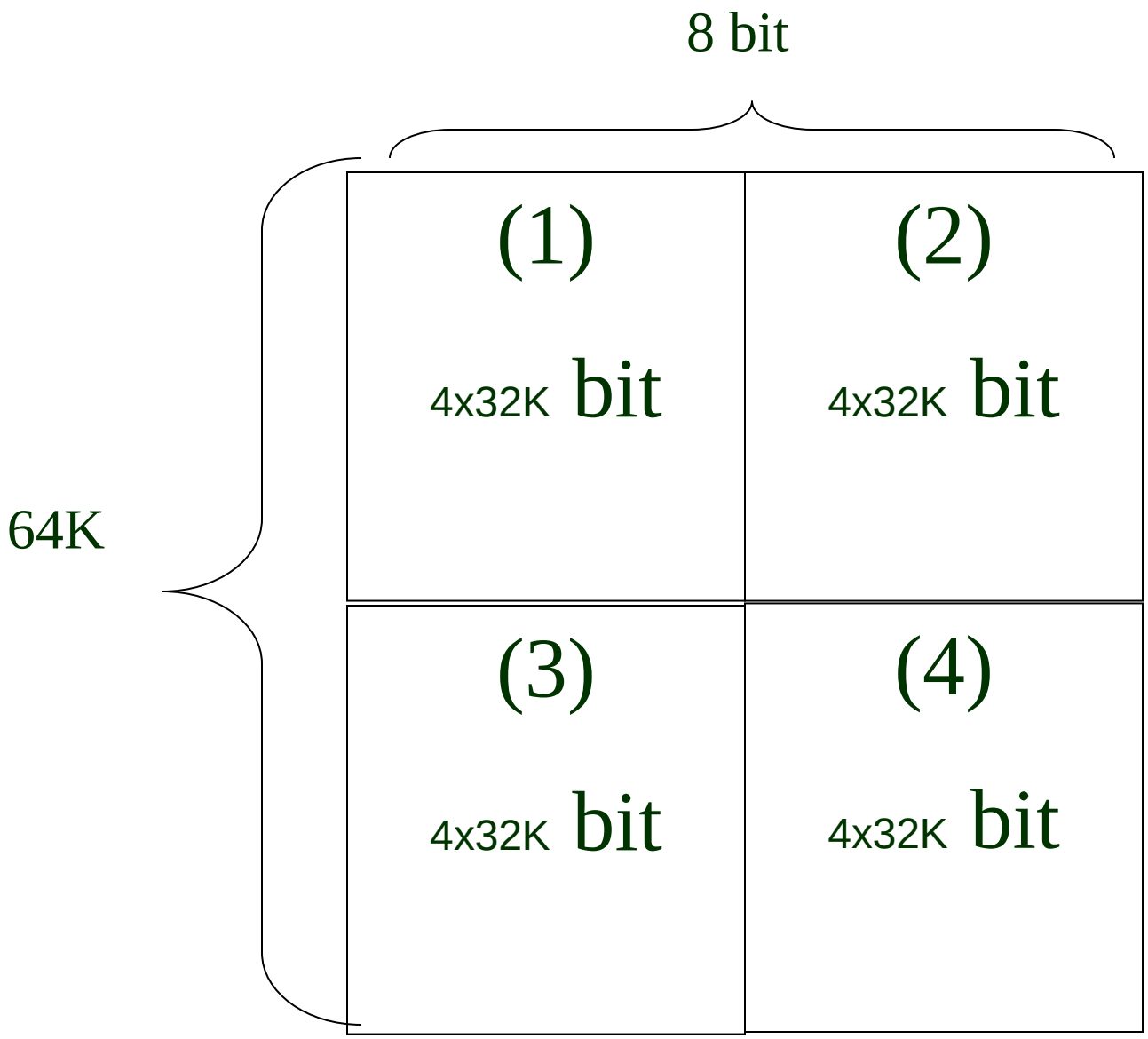
Construct an 8x8K memory using units of an 4x8K memories


D0
A0 ~ A12
4x8K bit
D3
D4
4x8K bit
D7


8 bit
8K
4x8K bit
4x8K bit

Address & Word expansion. Construct an 8x64K memory using 4 units of an 4x32K memories




8 bit
64K
(1)
4x32K bit
(2)
4x32K bit
(3)
4x32K bit
(4)
4x32K bit

Construct an 8x32k memory using 4x16k memories

Construct an 8x16K memory using 4x4k memories

Construct 2x64K memory using 2x16K units

Construct an 4x8K memory using units of an 1x8K memories

Construct an 4x16K memory using 1x4k memories

R A M (Random Access Memory)

- Because the bits of data can be individually accessed, retrieved, and modified at random, this type of main memory used in computers is called random access memory (RAM).
- RAM is a volatile form of memory, which means that it must have power in order to retain data. When the power is turned off, data in RAM is lost.
- Contrast this to other storage media such as disks, tapes, and CDs that retain data even without power.

There are two main forms of RAM:

- **DRAM**
- **SRAM**

SRAM *(Static RAM)*

- SRAM are built with flip-flop circuits that allow electricity to flow through one side or the other depending on which one of two transistors is activated.
- This "flow through" design is faster than the "stored charge" design of RAM, but it consumes more power.

- SRAM is also volatile (must have power to retain data) but SRAM does not need to be constantly refreshed like DRAM.
- SRAM is about five times faster, twice as expensive, and twice as big as DRAM.
- Because of its lower cost and smaller size, DRAM is preferred for use in computer main memory, while SRAM is used primarily for **cache memory** because of its speed.

DRAM *(Dynamic RAM)*

- DRAM is the most common type of computer memory.
- DRAM chips usually forms the core of a computer's main memory peripherals.
- The system uses this memory to temporarily store programs, data, and processed information that moves to and from the processor, video card, and other
- It is called "dynamic" RAM because it must be refreshed, or reenergized, hundreds of times each second in order to retain data in its memory cells

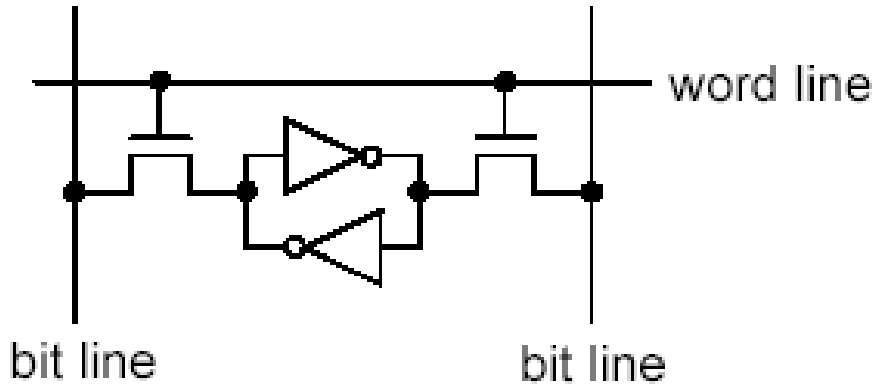
It has to be refreshed because its memory cells are designed around tiny capacitors that store electrical charges.

These capacitors work like very tiny batteries and will gradually lose their stored charges if they are not re-energized.

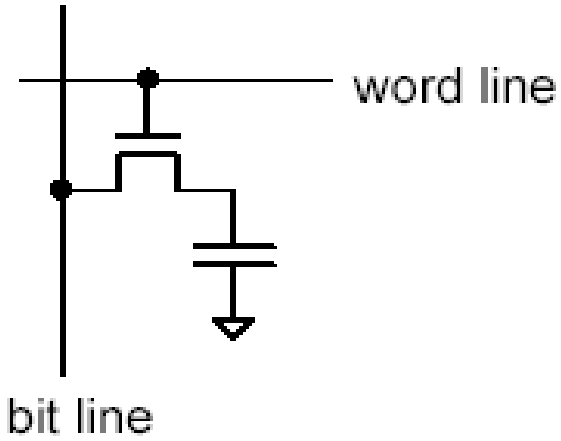
Different types of DRAM:
SDRAM (Synchronous DRAM)
DDR-SDRAM(Double Data Rate SDRAM)

Volatile Memory Comparison

- SRAM Cell



- DRAM Cell



Larger cell, lower density, higher cost/bit

No refresh required

Simple read faster access

Standard IC process natural for integration with logic

Smaller cell, higher density, lower cost/bit

Needs periodic refresh, and refresh after read

Complex read longer access time

Special IC process difficult to integrate with logic circuits

Cache Memory

Cache memory stores copies of chunks of frequently accessed RAM which is slower memory running at system board speed.

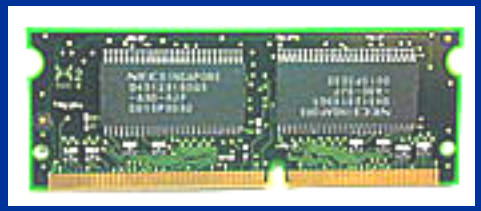
When the processor needs data, it will check in high-speed cache first to see if the data is there (fastest option). If not, then the processor will retrieve the data from slower main memory (RAM, slowest option).

Cache memory can be divided into two parts and two speeds.

- Level 1 (L1) cache is the fastest as it resides on the CPU chip itself and hence runs at CPU speed.

- Level 2 (L2) cache memory is the next fastest. It resides off-chip but close by for easy access. L2 cache is a small block of high-speed memory located between the CPU and main memory. Although, more and more CPUs are including L2 caches into their architectures

Physical View

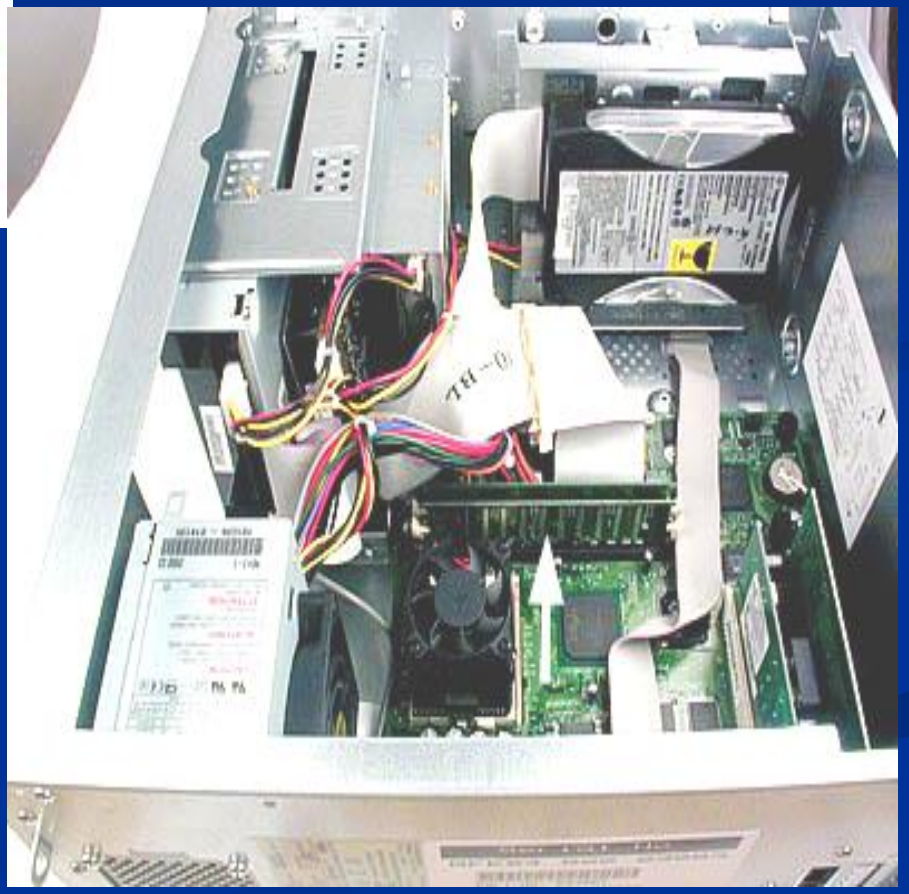


From the top:

SIMM, DIMM and SODIMM

memory modules

RAM installed in a PC



R O M (Read Only Memory)

- **ROM**
- **PROM**
- **EPROM**
- **EEPROM**
- **Flash memory**

Each type has unique characteristics, but there are two things in common. Data stored in these chips is

- **nonvolatile**
- either **unchangeable** or requires a special operation to change

ROM

- Similar to RAM, ROM chips contain a grid of columns and rows.
- ROM uses a **diode** to connect the lines if the value is **1**.
- If the value is **0**, then the lines are not connected at all.
- A diode normally allows current to flow in only one direction. If a diode is present at that cell, the charge will be conducted through to the ground, and, under the binary system, the cell will be read as being "on" (a value of 1)

This is a factory programmed ROM, Diodes are being inserted at the required cross points which cannot be reprogrammed.

- Creating the original template for a ROM chip is often a laborious process full of trial and error.
- But the benefits of ROM chips outweigh the drawbacks. Once the template is completed, the actual chips can cost as little as a few cents each.

- They use very little power, are extremely reliable and, in the case of most small electronic devices, contain all the necessary programming to control the device.

- A great example is the small chip in the singing toy. This chip, about the size of your fingernail, contains the 30-second song clips in ROM and the control codes to synchronize the motors to the music.

PROM- Programmable ROM

- PROM chips have a grid of columns and rows just as ordinary ROMs.
- The difference is that every intersection of a column and row in a PROM chip has a fuse connecting them.
- A charge sent through a row will pass through the fuse in a cell to a grounded column indicating a value of 1. Since all the cells have a fuse, the initial (blank) state of a PROM chip is all 1s.

- To change the value of a cell to 0, you use a programmer to send a specific amount of current to the cell.
- The higher voltage breaks the connection between the column and row by burning out the fuse.
- This process is known as burning the PROM.

- Creating ROM chips is time-consuming and very expensive in small quantities.
- For this reason, mainly, developers created a type of ROM known as programmable read-only memory (PROM).
- Blank PROM chips can be bought inexpensively and coded by anyone with a special tool called a programmer.

- PROMs can only be programmed once. They are more fragile than ROMs. A jolt of static electricity can easily cause fuses in the PROM to burn out, changing essential bits from 1 to 0.
- But blank PROMs are inexpensive and are great for prototyping the data for a ROM before committing to the costly ROM fabrication process.

EPROM

- Working with ROMs and PROMs can be a wasteful business even though they are inexpensive.
- **Erasable programmable read-only memory** (EPROM) addresses this issue.
- EPROM chips can be rewritten many times. Erasing an EPROM requires a special tool that emits a certain frequency of ultraviolet (UV) light.
- EPROMs are configured using an EPROM programmer that provides voltage at specified levels depending on the type of EPROM used.

- An EPROM eraser is not selective, it will erase the entire EPROM.
- Each EPROM chip has a quartz window on top of it for UV rays to enter in to the chip for erasing.

EEPROM

Though EPROMs are a big step up from PROMs in terms of reusability, they still require dedicated equipment and a labor-intensive process to remove and reinstall them each time a change is necessary.

Also, changes cannot be made incrementally to an EPROM; the whole chip must be erased.

Electrically erasable programmable read-only memory (EEPROM) chips remove the biggest drawbacks of EPROMs.

In EEPROMs:

- The chip does not have to removed to be rewritten.
- The entire chip does not have to be completely erased to change a specific portion of it.
- Changing the contents does not require additional dedicated equipment.

- Instead of using UV light, you can return the electrons in the cells of an EEPROM to normal with the localized application of an electric field to each cell.
- This erases the targeted cells of the EEPROM, which can then be rewritten.
- EEPROMs are changed 1 byte at a time, which makes them versatile but slow.
- In fact, EEPROM chips are too slow to use in many products that make quick changes to the data stored on the chip.

Manufacturers responded to this limitation with Flash memory, a type of EEPROM that uses in-circuit wiring to erase by applying an electrical field to the entire chip or to predetermined sections of the chip called blocks.

Flash memory works much faster than traditional EEPROMs because it writes data in chunks, usually 512 bytes in size, instead of 1 byte at a time.

FLASH MEMORY

- Few examples of Flash memory:
  - Your computer's BIOS chip
  - CompactFlash (most often found in digital cameras)
  - SmartMedia (most often found in digital cameras)
  - Memory Stick (most often found in digital cameras)
  - PCMCIA Type I and Type II memory cards (used as solid-state disks in laptops)
  - Memory cards for video game consoles

- Operation is same as in EEPROM but Flash memory works much faster than traditional EEPROMs because instead of erasing one byte at a time, it erases a block or the entire chip, and then rewrites it.
- Computer's BIOS chip is the most common form of Flash memory, removable solid-state storage devices are becoming increasingly popular.
- Smart Media and Compact Flash cards are both well-known, especially as "electronic film" for digital cameras. Other removable Flash memory