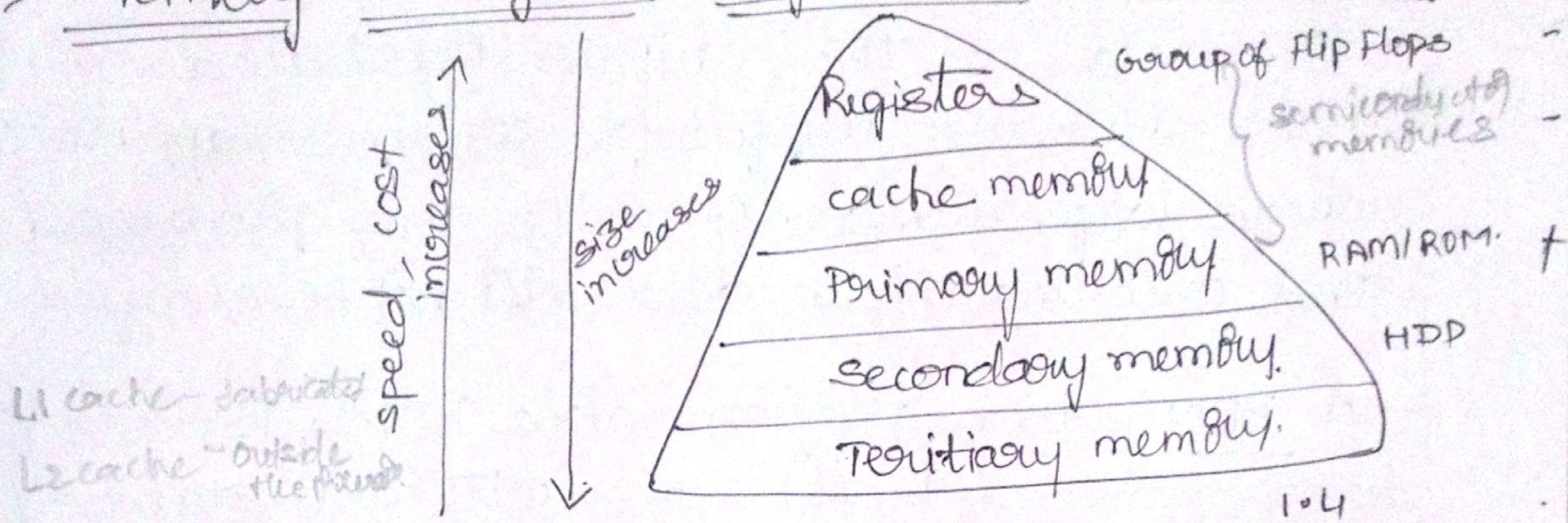
⇒ Memory Subsystem Organization & Interfacing :-



L1 cache - fabricated
L2 cache - outside the processor

Types of memory :-

- There are two types of memory chips:
  - ROM (Read Only Memory)
  - RAM (Random Access Memory)
- ROM chips are programmed with data by an external programming units before they are added to computer system.
- ROM chip always retains its data even when power to chip is turned off.
- RAM, also called Read/Write memory. RAM chips lose their data once power is shut off.
- Many computers include both RAM & ROM.

⇒ ROM chips :-

- There are several types of ROM chips, which are differentiated by how & how often they can be programmed.
  - masked ROM
  - PROM - Programmable ROM
  - EPROM - erasable PROM
  - EEPROM - E^2PROM - Electrically erasable PROM
  - flash EEPROM

- A masked ROM, or simply ROM is preprogrammed with data as the chip is fabricated. These chips are useful for consumer appliances, where large number of units are produced and once installed, data will not be changed.

- A PROM is a programmable ROM. PROM can be programmed by the user with any standard PROM programmer. Rather than being hard-wired, the PROM has a series of internal connections similar to fuses. Programming data into the PROM essentially blows the proper fuses so that each word of memory stores the correct value.

- An EPROM is an erasable PROM. EPROM can be programmed like a PROM, but its contents can be erased and the chip reprogrammed.
Programming an EPROM is similar to charging capacitors. The charged and uncharged capacitors cause each word of memory to store the correct value. The chip is erased by being placed under ultraviolet light causes the capacitors to leak their charge, thus resetting the chip's contents.

- EEPROM, sometimes denoted E^2PROM, is an electrically erasable PROM. It works like an EPROM, except its contents are erased and reprogrammed electrically, rather than by using ultraviolet light.

- In EPROM, it must be entirely erased and then reprogrammed, but in E^2PROM it is possible to modify individual locations of the EEPROM while leaving other locations unchanged.
- One common use of EEPROM is basic input/output system or BIOS of personal computer.

→ A Special type of EEPROM called a flash EEPROM is electrically erasable in blocks of data, rather than individual locations.

- The external configuration of all ROM types is almost same.
A chip with 2^n words, each having m bits, has n address i/ps, A_{n-1} to A_0 and m data o/ps D_{m-1} to D_0. It also have chip enable input (CE) and an output enable (OE).
Both must receive high (logic 1) signals if the chip is to output data;

⇒ RAM chips

RAM chips are differentiated by how they mairitain their data.
Dynamic RAM /DRAM
static RAM / SRAM.

- DRAM chips are like leaky capacitors. Initially data is stored in the DRAM chip, charging its memory cells to their maximum values. The charge slowly leaks out and would eventually go too low to represent valid data. (to be able to retrieve correct data)

(15 times/sec)

Before this happens refresh circuitry reads the contents of the DRAM and rewrites the data to its original locations, thus restoring the memory cells to their maximum charges.

- DRAM is used to construct the RAM in personal computer.

Access time 10-30 ns - SRAM
50 ns - DRAM

- SRAM is more like a register. Once data is written to SRAM, its contents stay valid; it does not have to be refreshed.
- SRAM is faster than DRAM, but it is also much more expensive.
- The cache memory in personal computer is constructed from SRAM.
- External configuration of both types of RAM is identical.

Each $2^n \times m$ chip has n address inputs and m bidirectional data pins, which are similar to those of the ROM chip except that the data pins can also input data under normal operating conditions.

RAM chips have a chip enable (CE or CE').

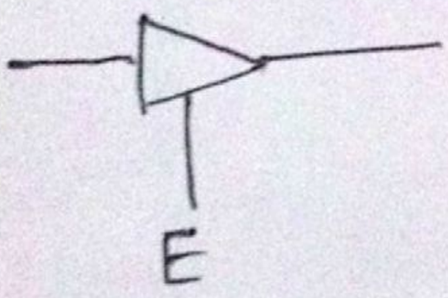
In place of output enable RAM chips may have either one read enable input (RD or RD') and one write enable input (WR or WR'), or one combined signal such as R/W'.

The R/W' signal would be set to 1 for a memory read operation or 0 for a memory write.

⇒ Internal chip organization:-

- The internal organizations of ROM and RAM chips are similar.
- To illustrate the simplest organization, a linear organization consider an 8 X 2 ROM chip. [fig 1.5]
- chip has 3 address i/p & 2 data outputs and 16 bits of internal storage arranged as eight 2-bit locations.
- The three address bits are decoded to select one of the eight locations, but only if chip enable(CE) is active. If CE=0, the decoder is disabled and no location is selected. The tristate buffers for that location is selected, allowing data to pass to the output buffers.

If both CE & OE are set to 1, these buffers are enabled and the data is output from the chip; otherwise the outputs are tristated.



I/P	E	O/P
X	0	Z
0	1	0
1	1	1

A_2	A_1	A_0	CE	DE	D_1	D_0
0	0	1	1	1	1	1
0	1	1	1	0	NO o/p / high impedance.	
1	1	1	1	1	1	0

Assume 2 bit locations

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

fig 1.5

INTERNAL LINEAR ORGANIZATION OF AN 8 X 2 ROM CHIP

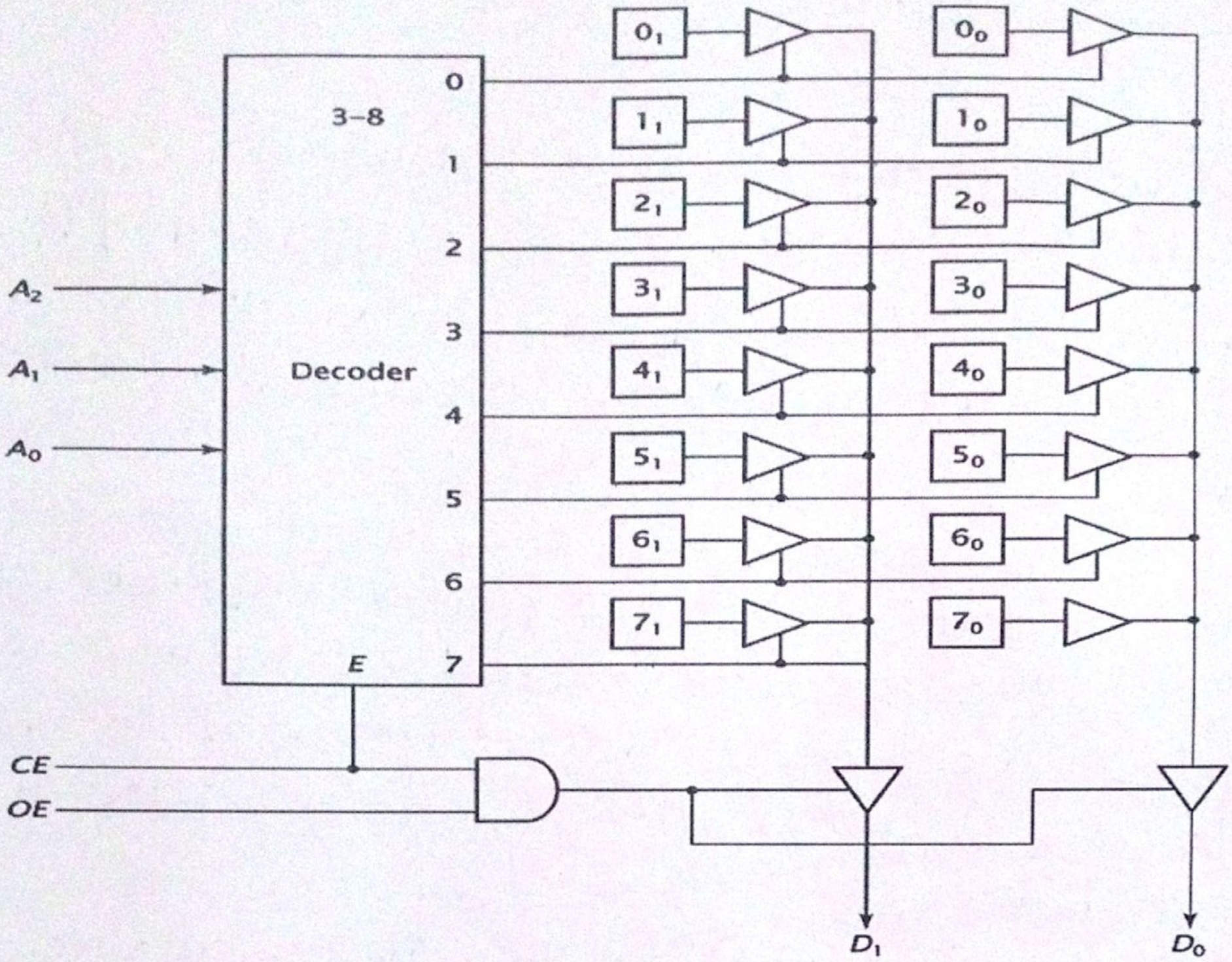
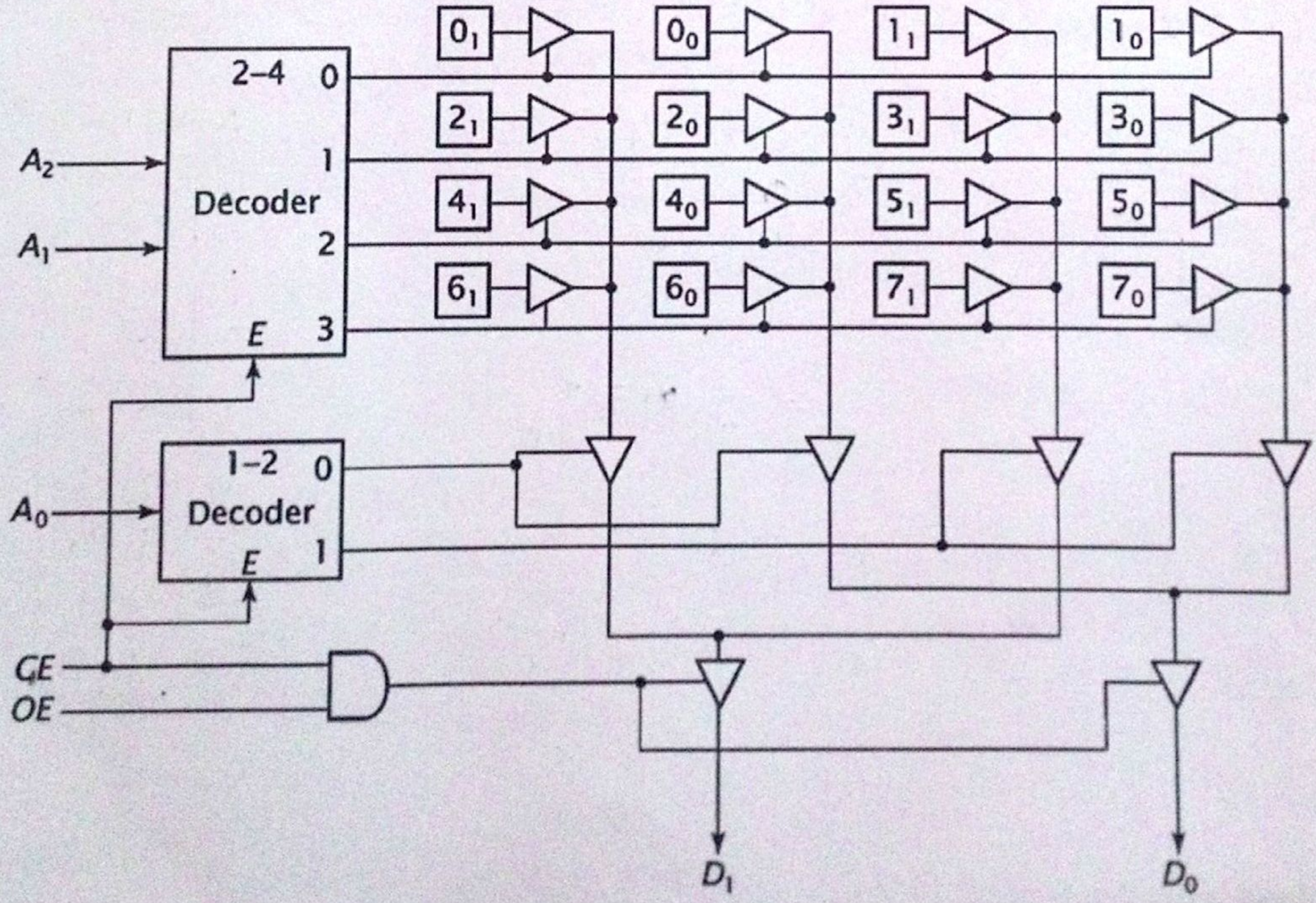


fig 1.6

INTERNAL TWO-DIMENSIONAL ORGANIZATION OF AN 8 X 2 ROM CHIP



⇒ Internal chip organization:-

- As the number of locations increases, the size of the address decoder needed in a linear organization becomes prohibitively large.
- To solve this problem, the memory chip can be designed using multiple dimensions of decoding.
- Consider the two-dimensional organization of the same 8×2 ROM chip (fig 1.6)
- This configuration has four rows with four bits per row; each row holds two data values.
- The two high-order address bits of the two locations in each row are the same. These two high order address bit select one of the four rows, and the low-order address bit selects one of the two locations in the row.

A_2	A_1	A_0
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

when $A_0 = 0$ then the locations 0, 2, 4, 6 are selected

when $A_0 = 1$; 1, 3, 5, 7 locations are selected.

⇒ Memory sub system Configuration:-

- Two or more chips can be combined to create memory with more bits per location.
- This is done by connecting the corresponding address and control signals of the chips, and connecting their data pins to different bits of the data bus.
- For ex, two 8X2 chips can be combined to create an 8X4 memory.

Fig 1.7

- Both chips receive the same three address inputs from the bus, as well as the same chip enable and output enable signals.
- The data pins of the first chip are connected to bits 3 and 2 of the data bus, and those of the other chip are connected to bits 1 and 0.
- When the CPU reads data, it places the address on the address bus.
- Both chips read in address bits A_2, A_1, A_0 and perform their internal decoding..
- If the CE and OE signals are activated, the chips output their data onto the four bits of the data bus.
- Since the address and enable signals are the same for both chips, they act as a single 8X4 chip.

AN 8 X 4 MEMORY SUB SYSTEM CONSTRUCTED FROM TWO 8 X 2 ROM CHIPS

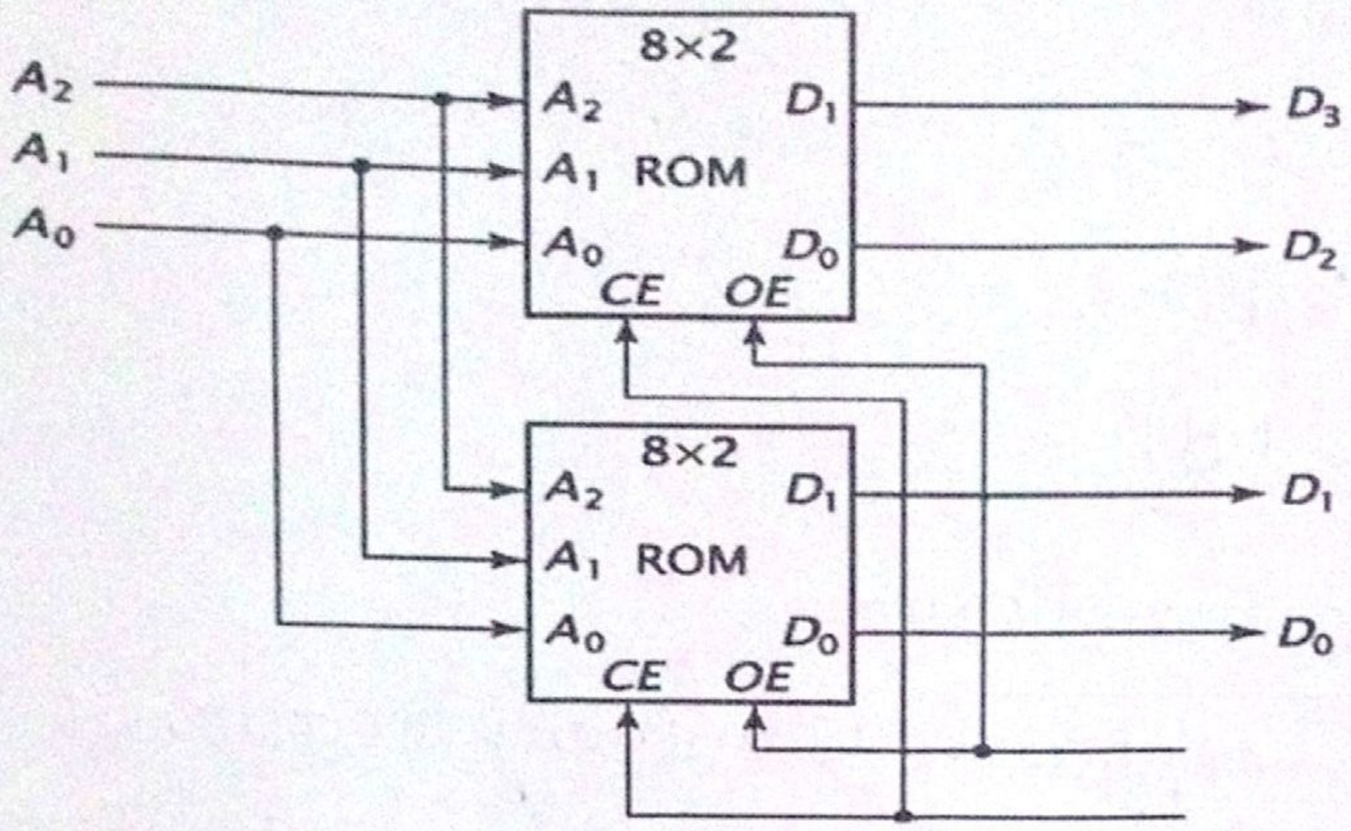
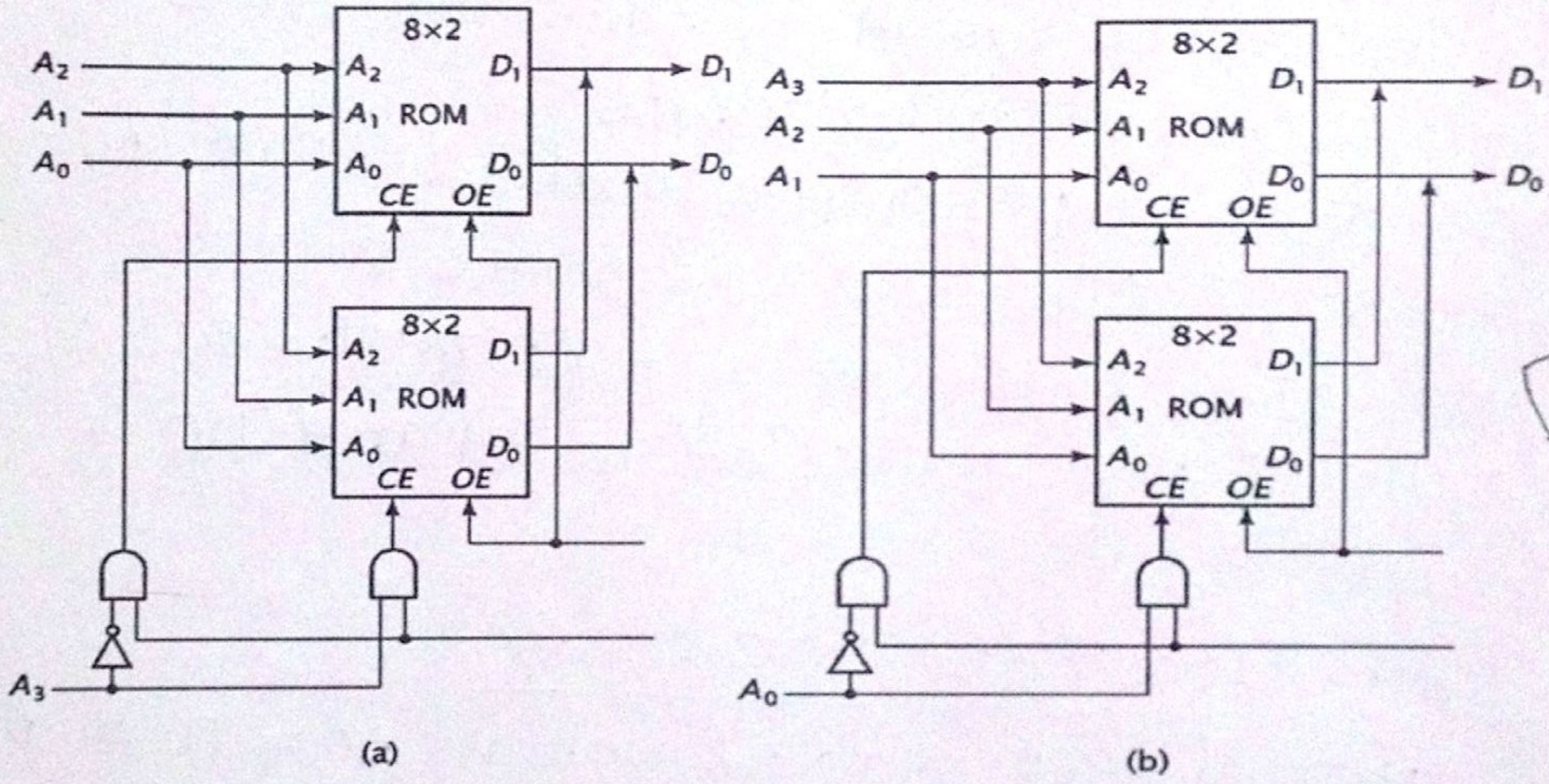


Fig 1.7

A 16 X 2 MEMORY SUBSYSTEM CONSTRUCTED FROM TWO 8X2 ROM CHIPS WITH

A)HIGH-ORDER INTERLEAVING B) LOW-ORDER INTERLEAVING



(a) (b)

Fig 1.8

AN 8X4 MEMORY SUB SYSTEM CONSTRUCTED FROM TWO 8X 2 ROM CHIPS WITH CONTROL SIGNALS

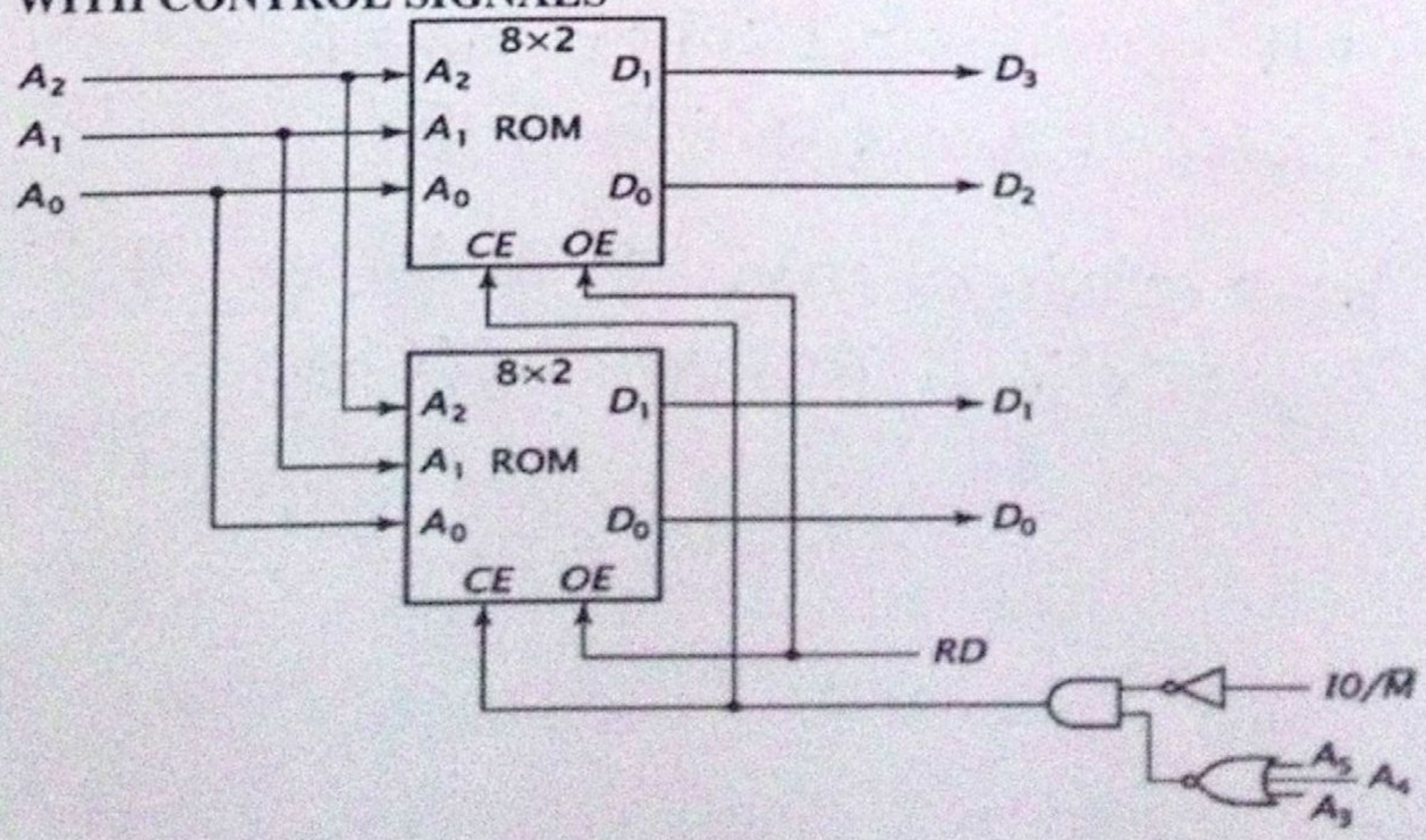


Fig 1.9

⇒ Instead of creating wider words, chips can be combined to create more words.

- Two 8×2 chips could be used to configure a 16×2 memory subsystem.

Fig 1.8

Fig 1.8 a - High-order Interleaving

- The upper chip is configured as memory locations 0 to 7 (0000 to 0111) and lower chip locations are 8 to 15 (1000 to 1111). The upper chip always has A3=0 and lower chip has A3=1. This difference is used to select one of the two chips.
- when A3=0 upper chip is enabled & lower chip-disabled
- when A3=1 upper chip - disabled & lower chip - enabled.
- Since both chips corresponds to the same data bits, both are connected to D1 and D0 of the data bus.

Fig 1.8 b - Low-Order Interleaving

- The upper chip is enabled when A0 = 0 & by addresses XXX0, in this case 0, 2, 4, 6, 8, 10, 12 & 14.
- The lower chip is enabled when A0=1, which is true for addresses 1, 3, 5, 7, 9, 11, 13 & 15.

- Low-order interleaving can offer some speed advantages for pipelined memory access.

→ Input logic for CE & OE :-

- The CPU generally outputs a control signal called RD & RD', which it sets active when it wants to read data from memory. This signal is sufficient to drive OE.

- The CE signal makes use of the unused address bits. Assume that the 8×4 memory is used in a system with a 6-bit address bus. Assume this chip corresponds to locations 0 to 7 (00 0000 to 00 0111).
Address bits A_2, A_1 and A_0 select a location within the memory chips; bits A_5, A_4, A_3 must be 000 for the chips to be active.

Fig 1.9

The fig includes the IO/$\overline{M}$ signal that is supplied by processors that use isolated I/O.

→ Multibyte Data Organization :-

There are two commonly used organizations for multibyte data: big endian
little endian

- In big endian format, the most significant byte of a value is stored in location X, the following byte in location X+1, and so on
- In little endian, the order is reversed. The least significant byte is stored in location X, the next byte in location X+1 and so on.

Eg:- For ex, the hexadecimal value 01020304H organized in big endian & little endian formats as follows.

Memory Address	Data (in Hex)
100	01
101	02
102	03
103	04

Big endian

Memory Address	Data (in Hex)
100	04
101	03
102	02
103	01

Little endian.

- The main problem comes in transferring datee between compilers with different organisations.
- One other issue for multibyte words is alignment.
~~Modern computers~~ modern microprocessors can read in more than one byte of data at a time.
Alignment simply means storing multibyte values in locations such that they begin at a location that also begins a multibyte read block.