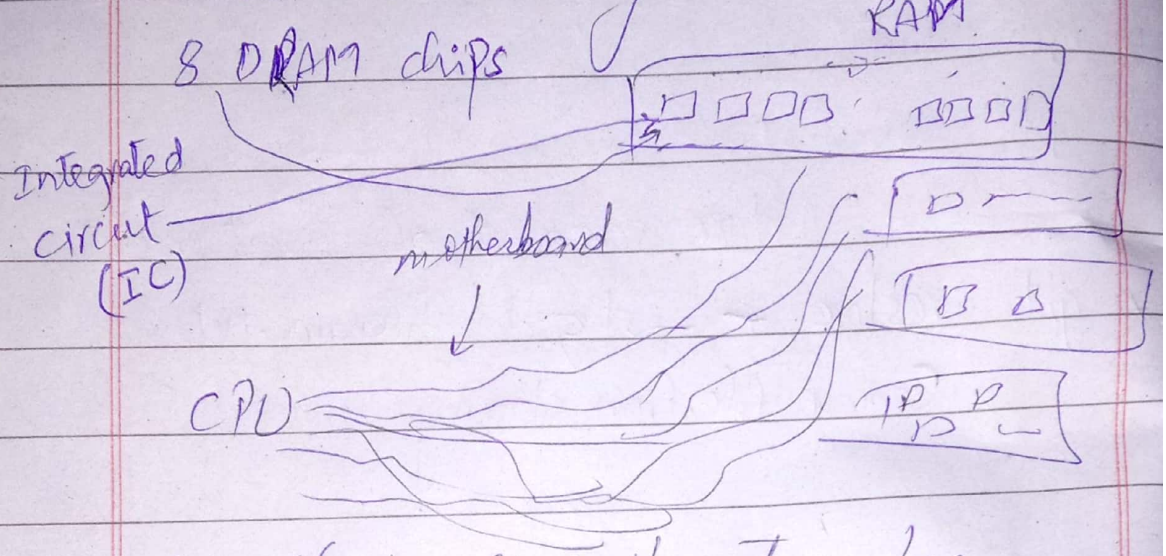
How Computer memory Works

Day: How does comput memory work Date:

- DRAM is also called (DIMM)
  Dual inline memory module
  8 DRAM chips



~~two~~ if 4 DRAM then two has different channel and other two diff.

⇒ CPU

* In CPU memory controller manage the ~~comptr~~ communication with DRAM
* There is a one other section that control the communication with M2/SATA HDD.

⇒ Memory channel DDR5
memory is devided into two channels

A A B

Day: Date:

→ A & B transfer 32 bits at a time using 32 data wires.

→ 21 additional wires to memory that carries address where to read & write data

→ 7 control signal wires, command are relayed. These 7 signal are used to send and receive all the bits

Power (electricity) supplied by the motherboard and handled by the chips that are placed in center.

→ DRAM microchips (8 bit)
there are four layers ~~between~~ two at center interconnect the lower (ball Grid array) to upper (Die).

→ Die
It has 8 Bank groups
one Bank group consists of 4 Bank.
so total Bank 8x4 = 32 Banks
~~Each~~ ⇒ Bank
Each Bank has a memory

Day: Date: Fall

array (65536 ~~memory cells~~ rows) vertically falls / length
(8192 ~~cells~~ columns) accross / horizontally

There is a circtry outsides the Banks which consists of thousand of wires for communication.

⇒ How to access ~~info~~ / data in banks
31 Bit address ∪

⇒ first 3 bits to identify Bank group
~~4 2 1~~ ✓
1 1 1 $2^3 = 8$
8 combination ~~as group also~~ 8

⇒ next two Bits to select Bank
$2^2 = 4$ as Per group
there are ~~four~~ 4 Banks.

⇒ Next 16 Bits to identify the row of data for accessing like
$2^{16} = 65536$

⇒ 8192 cells group of 8 columns in one Bank. why in 8 columns Because it can only read and write 8 bits

Day: Date:

at a time.

first 5 bits Bank address
Next 16 bits Row address
Next 10 bits column address

⇒ Cell / 1T1C DRAM memory cell.
One memory cell

Two Parts

→ Capacitor: to store 1 bit of data either 1 or 0 in the form of electrical charges or electrons. Its shape is like deep trench. access

→ Transistor: It allows for reading and writing the data. It controls the flow of electrons. like emplify which represent (1) or to zero.

If capacitor charged up with electron it means 1. and if no charge Present and it is zero ~~charge~~ volt then

Day: Date:

it is a binary zero.

Types of capacitors / evaluation.

(1) MOS capacitor (2) stacked

(3) Trench (4) substrate plate trench capa-

(5) 3-D stacked capacitor.

~~wordline~~ Bitline

~~There is a gate which has on~~ which if wordline act then the gate will open and through channel the capacitor charged up. and vice versa

⇒ How to read stored value in capacitor.

we can read the value in the capacitor by measuring the ~~volt~~ amount of charge in capacitor.

⇒ ~~token word~~ horizontal ⟺

Wordline

Bitline ↕ vertical

Applying voltage on the wordline turn on the transistors. which leads to the full charge of capacitor

Day: Date:

Now electron can flow through the channel and connect to bitline.

→ when the wordline is off transistor also turned off as well and capacitor isolated from bitline saving the data/charge that were previously written.

→ Electron Leakage → Refresh memory

The transistor is encredibly small so the electrons leakage happened slowly through chanel, So we need to capacitor needs to be refereshed to recharge the leaked electrons.

To resolve this a issue we use refreshing memory cells technique.

~~cover later.~~ → Solution 2

Each colums is connected to a sense amplifier. when a 32 bit address come then first 5 bits select the specific Bank and next 16 bit select the row

Day: _____ Date: _____

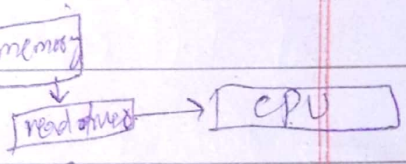
in the Bank, all the wordline accept the selected one turned off. Now in that one active row which contain 8192 columns is active. Some capacitors are charged (1) some not (0). So in the charged capacitors the electron started to leak in the bitline. However the bitline is connected to the sense amplifier so it sense the leakage of electrone so it amplify the electrone and make it full. Also in the capacitor that store (0) zero electron the electron started to charge in the capacitor through bitline channel. So sense amplifier then reduce the electron from that bitline to zero.

Day: Date:

Now the row is selected from 65536 lines/rows but what about columns we only want to access 8 bit but one row contain 8192 bits so how?

⇒ Read Memory

The last 10 bits from the 31 bit addresses are used to identify which colums are need to be accessed. These 10 bits select the 8 bits from columns and then send to Read drivers through wires and then send to CPU



⇒ Write Memory (31 bit address)

First Five bits select the Bank

Next 16 Bit select the row.

next 10 bit select the specific columns and connect to the Write Driver. At this time the bits that should be

Day: Date:

written in memory was already sent to the write driver by CPU which is in the DRAM. the write drivers are much more stronger than sense amplifier. So it charge/overwrite the capacitor if (1) needed otherwise if not charge if (0) needed.

At the time of row selection all the capacitor charged to 0.5 volt. which is the half of full signal voltage. It is called Precharge.

⇒ Read & write happen concurrently to the all four IC's with shared address & command wires But separate data wires

Voltages for 1 in different DDR

DDR	2.5 V	DDR4	1.2 V
DDR2	1.8 V	DDR5	1.1 V
DDR3	1.5 V		

precharge is half for each.

Day: Date:

DDR5 precharged voltages 0.55V
It store 1.4V because of
leaked electron/voltages.

2^3 8 Bank group 2^5 32 Bank group
2^{13} 8,192 Bitline 2^{16} 65536 wordline

⇒ Refreshing the rows because of volt leakage.

Sense amplifies open the first row and then refresh the whole row like 1 to full charge 0 to zero charge.
This happen row by row.

⇒ Refreshing one raw takes 15 nano sec time.
⇒ 3 miliseconds to refresh 65536 rows
1 whole Bank.

Refresh action occures after every 64.00 miliseconds for each Bank.

Each IC on Memory handles

Day: _____ Date: _____

4800 million Requests/second

In one second the whole IC's Banks refreshes 16 times.

⇒ Why need this much faster memory

eg: for video games

each moment and changing environment shadow and many more features need fast and on-time calculation in the memory.

⇒ Raw HIT/Page Hit

If there is data that is already loaded on the memory and we are also reading and writing that data (same address) it means the row is open and then we can directly read or write the data skipping some steps like. For opening a row we need to

Day:

Times ~~Ref~~: Perform actions in clock cycle

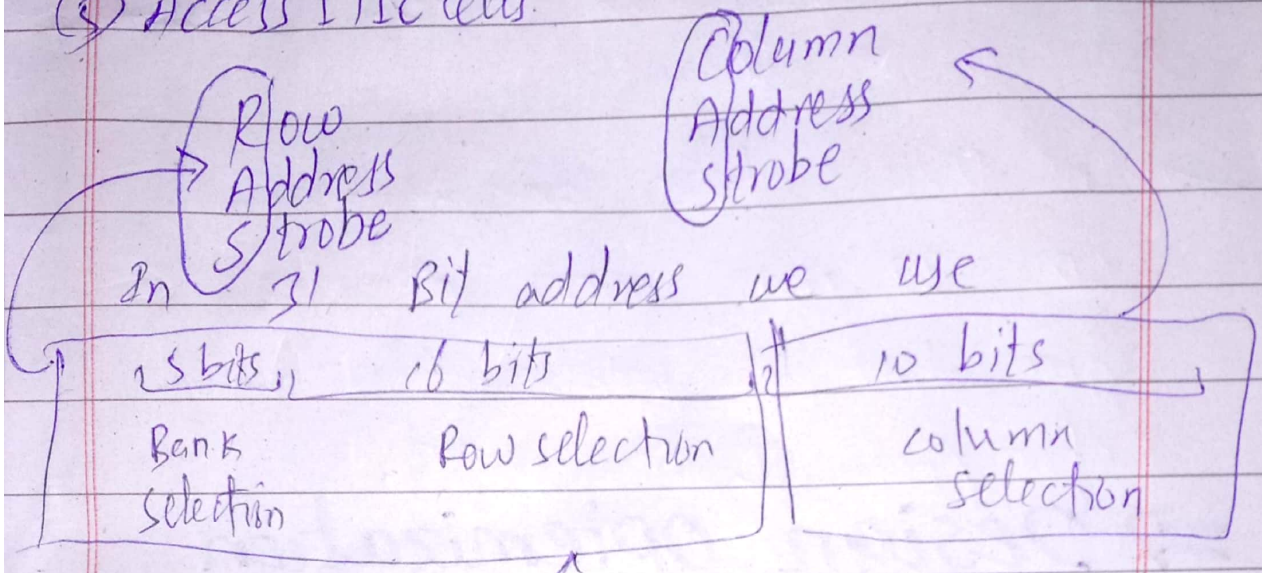
(1) Row close

(2) Precharge Bitlines → 39 Precharge tRP

(3) Row open → 39 RAS to CAS Delay tRCD

(4) Column address → 40 CAS latency.

(5) Access ITIC cells.



If a row is already selected we can skip this step and directly use 10 bits to select columns.

It reduce the amount of time.

If the same row is requested that was already open for read/~~rea~~write we called page hit/Row hit.

⇒ If another row requested for read and write accept the open one. then it is called Row miss.

Day: Date:

⇒ Row Thrashing

when a program request runs/jumps from one row to another over and over. It becomes inefficient in terms of both energy and times.

⇒ In DDR5 there are 32 banks so 32 rows opens at the same time to increase the liklehood of Row hit in each IC.

⇒ Design optemization

⇒ Burst Buffer

Read and write driver each has 8 wires that connects to memory.

In Burst Buffer now we place Burst Buffer between Read ~~and~~ write driver and memory. Now memory connected to each Burst Buffer is with 128 wires.

Day: Date:

For column selection 10 bits divided into, 6 bits, for column multiplexer and 4 bits for Burst Buffer.

⇒ Burst Buffer for reading.
It consist of 128 bits
16 columns $2^4 = 16$
8 Rows
Burst Buffer one column ~~for~~ which has 8 bits connected to Read driver through 4 bits.

☆ Same for writing.

3 Benefits
16 sets of 8 bits per microchip = 128 bits
Total 8 microchip so = 128 x 8 = 1024 bits

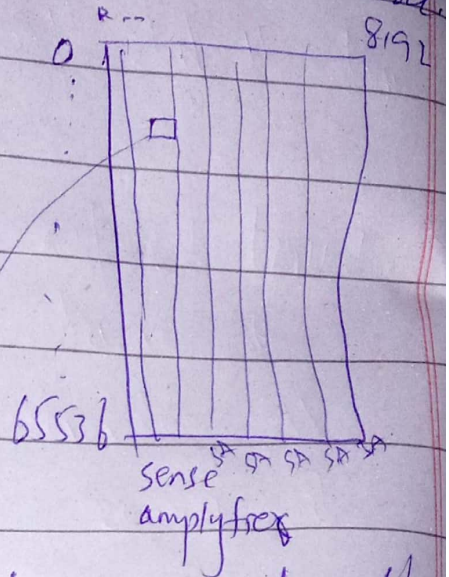
⇒ Another optimization
each IC has 65536 rows and 8192 columns
So the ~~the~~ the bitlines are very long

SA = sense amplyfier

Day: _____ Date: _____

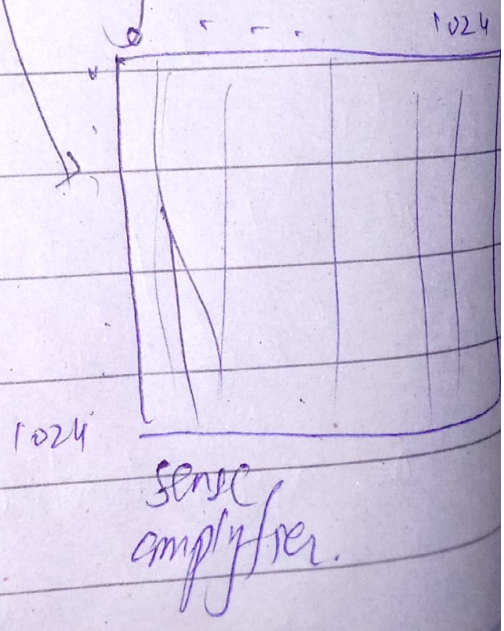
and also charging the capacitor for sense amplyfier is hard.

It is a massive structure.



So, we devide and grouped the block of 1024 x 1024 bits and each block has little sub sense amplyfier

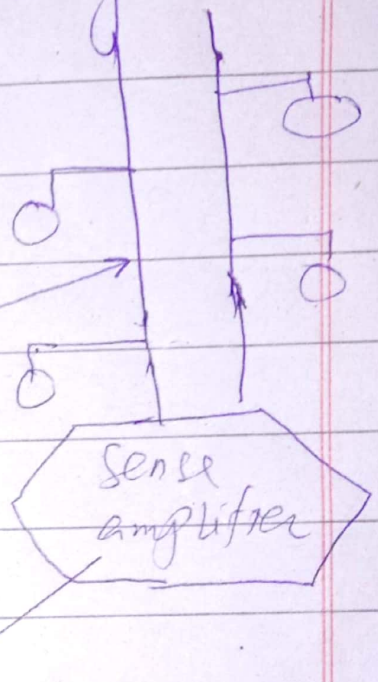
So, for charging the capacitor becam fast because of having SA near to capacitors.



Day: Another Date:

- Sense Amplifies connected to each bit line. or column.

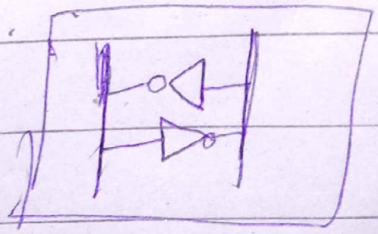
Now we connected two bitlines to single column but alternating the rows of memory cells



Now if one row is open ~~we can e~~ then half of the bitline active. Other half are passive(off) also vice versa.

⇒ How above works?

↳ Inside the sense amplifier there is a Cross Coupled inverter. It turned on one bit line and off the other bitline.



Sense Amplifier.

Day:_____ Date:_____

→ Benefits

1. Precharge
2. Noise Immunity
3. Parasitic Capacitance.