The cause of memory limitation in computer architecture today

Mohammad Saeed Rajabi
Mohammad mahdi nemati
Amir Ali Roshanzadeh

Static RAM vs. dynamic RAM

STATIC RAM	DYNAMIC RAM
Does not require refreshing	Must be continuously refreshed
Requires multiple transistors to store one bit	Requires one transistor and one capacitor to store one bit
Delivers faster access times	Delivers slower performance
Consumes less power, especially in idle	Consumes more power
Takes up more space	Requires less space
Can hold only a small amount of data	Can hold much more data
Costs more than DRAM	Costs less than SRAM
Typically used for a processor's cache	Typically used for a computer's main memory

Type of DRAM

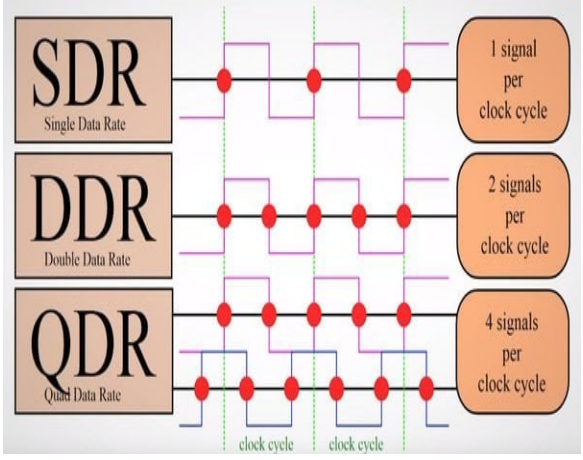
- **SDRAM (Synchronous Dynamic RAM)**
- **RDRAM (Rambus DRAM)**
- **LPDDR (Low Power DDR)**
- **DDR ECC (Error-Correcting Code) RAM**
- **RDIMM (Registered DIMM)**
- **UDIMM (Unbuffered DIMM)**
- **SO-DIMM (Small Outline DIMM)**
- **DIMM (Dual In-line Memory Module)**
- **NVRAM (Non-Volatile RAM)**
- **HBM (High Bandwidth Memory)**
- **XDR RAM (Extreme Data Rate RAM)**

Type of SDRAM

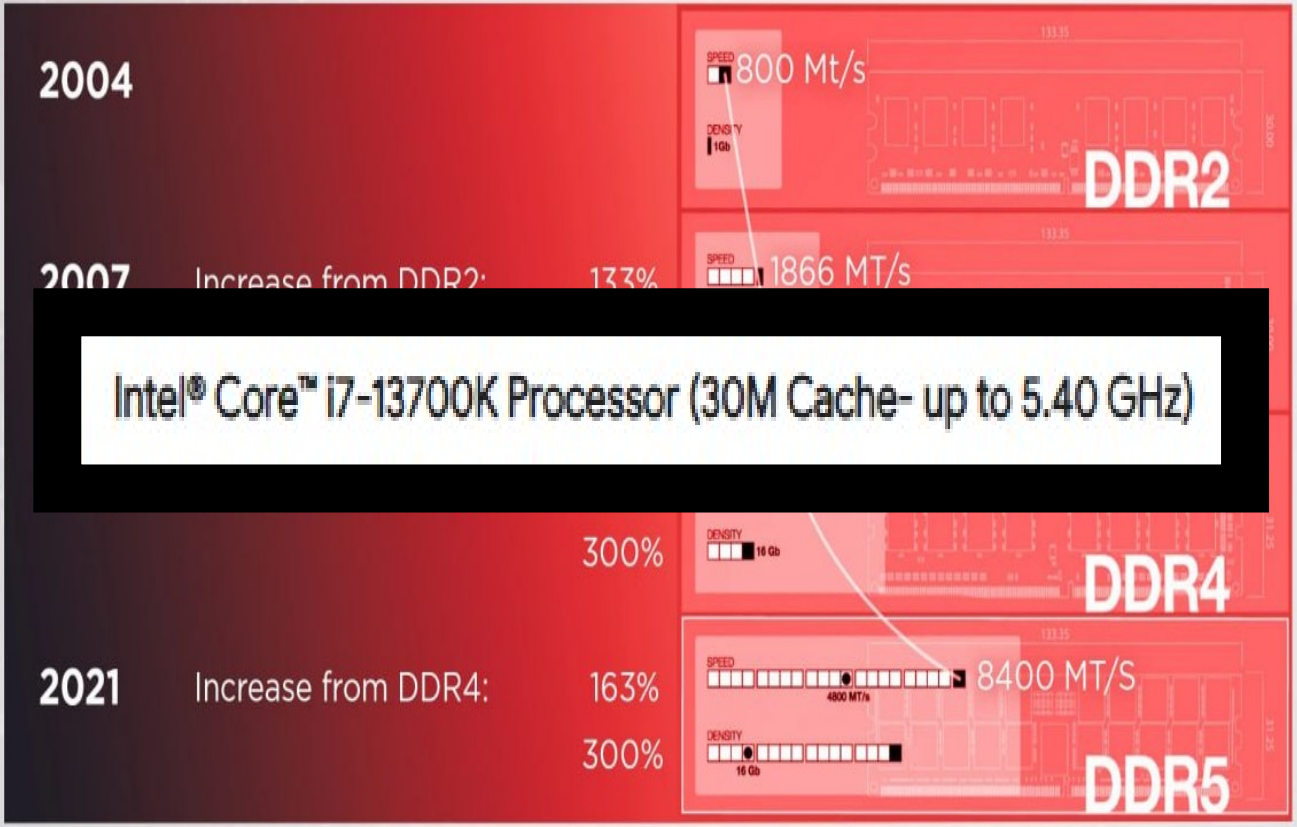
SDR (Single Data Rate) SDRAM: SDR SDRAM is a version of Synchronous Dynamic RAM (SDRAM) in which data is transferred only once in each clock cycle. In other words, only one data transaction occurs per clock cycle. This type of DRAM had limited data transfer speeds and is generally considered an older generation of SDRAM.

DDR (Double Data Rate) SDRAM: DDR SDRAM has a double data transfer rate compared to the previous generation SDR SDRAM. In this type, data is transferred on both the rising and falling edges of the clock, so two data transfers occur in each clock cycle. In other words, DDR means that data is transferred twice in each clock cycle.

QDR (Quad Data Rate) SRAM: QDR SDRAM is an advanced generation beyond DDR SDRAM, where data is transferred four times in each clock cycle. This means that four data transactions occur in each clock cycle. QDR SDRAM is typically used in systems that require high speed and processing of large data



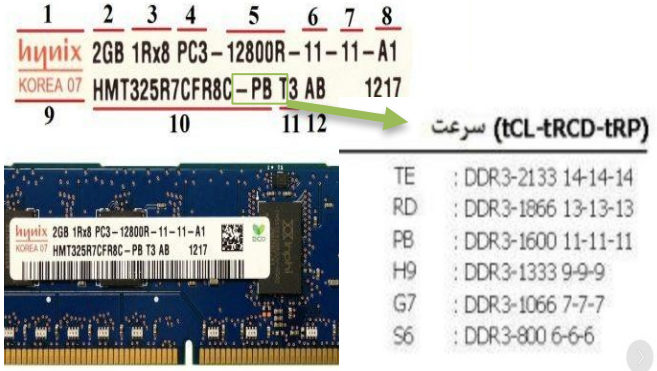
Generational Leaps in DDR RAM Generations

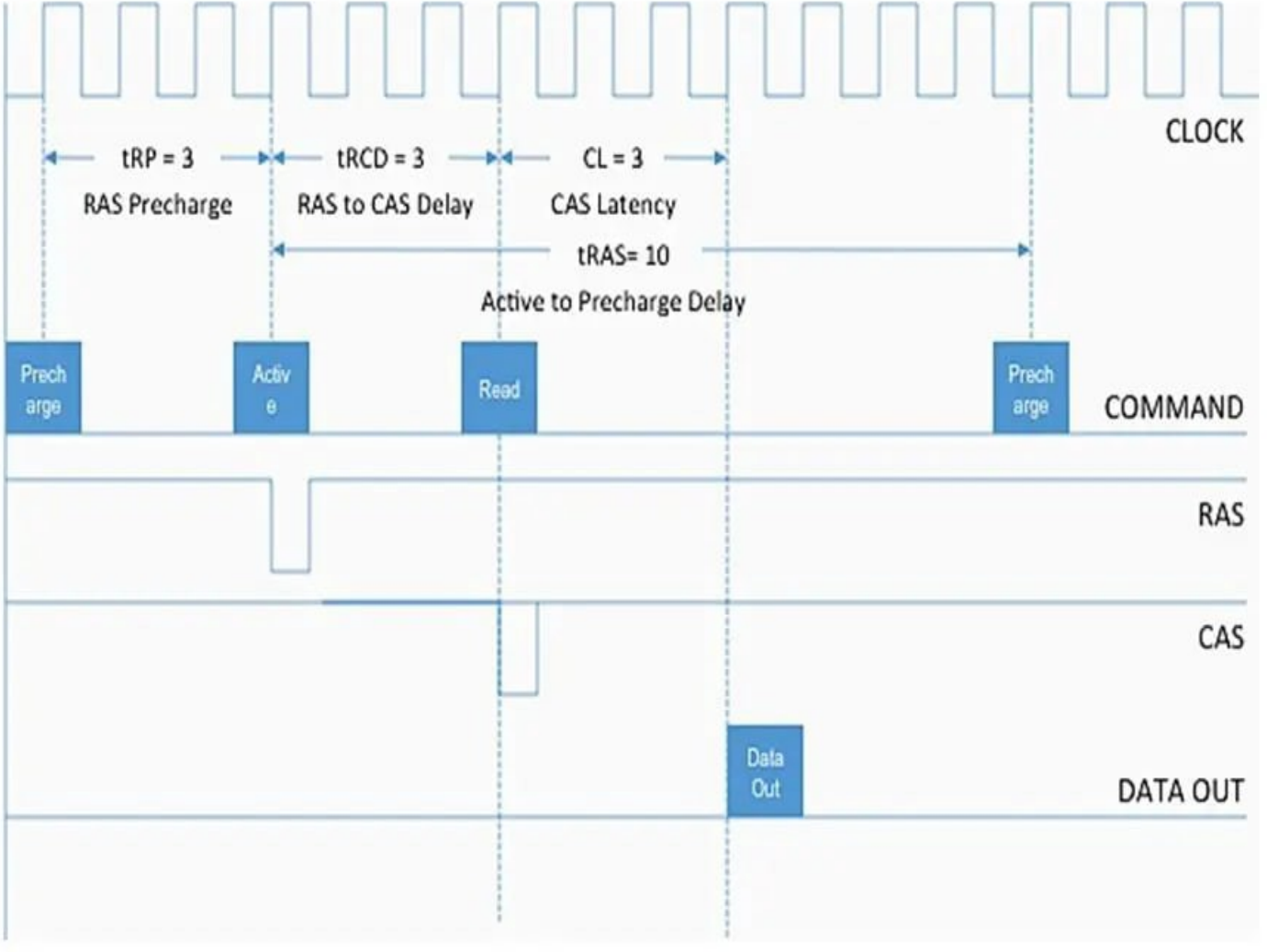


CGDIRECTOR.COM

RAM Timing

- RAM timing means the time or clock pulses that it takes for a data to sit in the RAM memory or to be called.
- Each RAM module has parameters such as "CAS OR CL" , "TRCD“ , "TRP“ , "TRAS" and "CMD."is that each of them represents a range of the working cycle of RAM




CLOCK
tRP = 3
RAS Precharge
tRCD = 3
RAS to CAS Delay
CL = 3
CAS Latency
tRAS= 10
Active to Precharge Delay
Prech arge
Activ e
Read
Prech arge
COMMAND
RAS
CAS
Data Out
DATA OUT

Factors determining RAM efficiency

- ***memory Bandwidth***

refers to the rate at which data can be transferred from one location to another over a communication channel.
** The best measure for large continuous data.*

- ***memory Latency***

refers to the delay or the time it takes for data to be accessed or transferred from one location to another.
**The best measure for small random data.*

Memory Bandwidth

To calculate the bandwidth, you can use the following formula:

Bandwidth (in GB/s) = (Data Rate * Bus Width) / 8

For example, let's consider DDR5-4800 with a bus width of 128 bits:

Bandwidth = (4800 MT/s * 128 bits) / 8

= 76800 MB/s

= 76.8 GB/s

Memory Bandwidth

In reality:

Effective Memory Bandwidth			Effective Memory Bandwidth per CPU Core						
Memory Standard	System Bandswidth (GB/s)	Core Count	4.0	6.0	8.0	10.0	12.0	14.0	16.0
DDR5-5600	69.21	GB/s/Core	17.3	11.5	8.7	6.9	5.8	4.9	4.3
DDR5-5200	66.12		16.5	11.0	8.3	6.6	5.5	4.7	4.1
DDR5-4800	62.74		15.7	10.5	7.8	6.3	5.2	4.5	3.9
DDR5-4400	58.81		14.7	9.8	7.4	5.9	4.9	4.2	3.7
DDR5-4000	54.65		13.7	9.1	6.8	4.4	4.6	3.9	3.4
DDR5-3600	50.26		12.6	8.4	6.3	5.0	4.2	3,6	3.1
DDR5-3200	45.62		11.4	7.6	5.7	4.6	3.8	3.3	2.9
DDR5-3200	33.57		8.4	5.6	4.2	3.4	2.8	2.4	2.1

Memory Latency

True memory latency (ns) = (2000/RAM Speed)(ns) x CAS latency

Therefore,
true memory latency of DDR4-3200 CL22 = 13.75 ns and
true memory latency of DDR5-4800 CL40 = 16.67 ns

In reality:

Memory Specification	System Latency[1]
DDR5-4800 CL40	92.8 ns
DDR4-3200 CL22	90.0 ns