ECE 571 – Advanced Microprocessor-Based Design Lecture 24

Vince Weaver

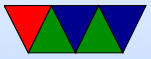
https://web.eece.maine.edu/~vweaver

vincent.weaver@maine.edu

4 November 2024

Announcements

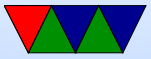
- HW8 will be posted soon
- Don't forget project topic e-mail due Friday (go over common project topics)
- Midterm not graded yet



Advances in Memory Technology

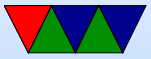
In general the actual bit array stays same, only interface changes up.
A historical progression:

- Clocked
- Asynchronous
- Fast page mode (FPM) – row can remain active across multiple CAS.
- Extended Data Out (EDO) – like FPM, but buffer "caches" a whole page of output if the CAS value



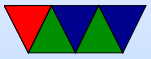
the same.

- Burst Extended Data Out (BEDO) – has a counter and automatically will return consecutive values from a page



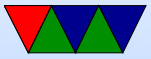
Synchronous DRAM (SDRAM)

- The big change was to SDRAM. Most modern RAM is some variant on SDRAM
- Drives internal circuitry from clock rather than outside RAS and CAS lines.
- Previously the commands could happen at any time. Less skew.
- Prefetch architecture? Reads multiple words from row and bursts them back without separate column requests
- SDRAM not necessarily faster than previous types but



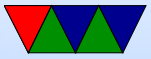
interleaving and bursts improve bandwidth

- SDRAM – 3.3V

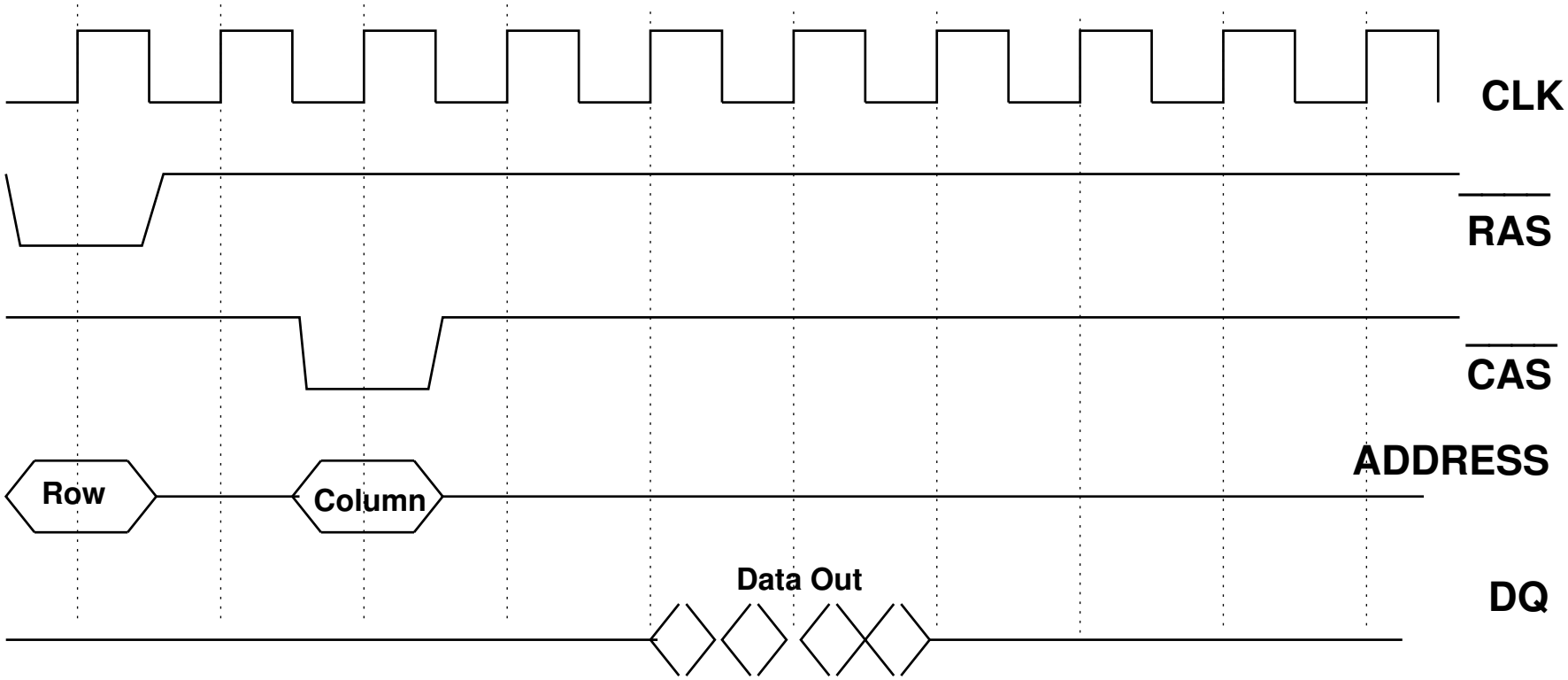


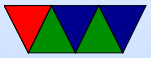
DDR

- year(?)
- transfer and both rising and falling edge of clock
- 2.5V (lower voltage is lower power)
- Adds DLL (delay locked loop) to keep clocks in sync (but burns power)



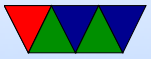
DDR Timing Diagram





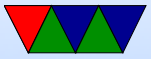
DDR2

- 2003
- runs internal bus half the speed of data bus
- 4 data transfers per external clock
- memory clock rate * 2 (for bus clock multiplier) * 2 (for dual rate) * 64 (number of bits transferred) / 8 (number of bits/byte) so at 100MHz, gives transfer rate of 3200MB/s.
- not pin compatible with DDR3.
- 1.8 or 2.5V



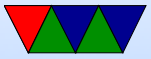
DDR3

- 2007
- internal doubles again
- Up to 6400MB/s, up to 16GB DIMMs.
- 1.5V or 1.35V
- 400...1067MHz.
- DDR3-800 ... DDR3-2133

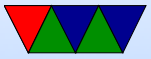


DDR4

- 2014
- Higher density, faster speed, lower voltage than DDR3
- 1.2V, 2.5V auxiliary wordline boost
- 16 internal banks, up to 8 ranks per DIMM
- 1600 ... 3200 MT/s
- DDR4-1600 (PC4-12800) (12.8GB/s) to DDR4-3200 (PC-25600) (25.6GB/s)
- Up to 64GB DIMMs
- parity on command/address busses, crc on data busses.

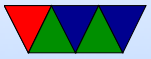


- Data bus inversion (if more power/noise causes by sending lots of 0s you can set this bit and then send things as 1s instead)
- Pins closer together (0.85mm vs 1.0mm)
- 288 pins vs 240 pins, new package. Slightly curved edge connector so not forcing all pins at once when pushing in
- Example: DDR4-2400R, memory clock 300MHz, I/O bus clock 1200MHz, Data rate 2400MT/S, PC4-2400 19200MB/s (8B or 64 bits per transaction) CAS latency around 13ns



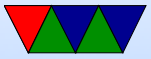
DDR5

- 2020 (Alder Lake / Raptor Lake optional, Sapphire Rapids required)
- DDR5-4000 (PC5-32000) to DDR5-8800 (PC5-70400), up to 8GT/s
- Reduce power and Doubled bandwidth over DDR4
- 1.1V, on board voltage regulation (5V or 12V from board?)
- Up to 512GB DIMM
- On-chip ECC to avoid errors on chip, but not required



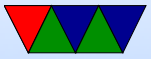
to traditional report-to-CPU if something is wrong ECC

- Decision Feedback Equalizer (DFE)?
- LRDIMM (12V buffered) and UDIMM (5V)
- 288 pins like DDR4 but different layout
- C0-C2 column bits removed to speed protocol, only get multiples of 8
- Built in temp sensor?



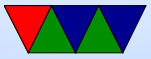
DDR6

- Soon (2024/2025)

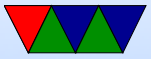


GDDR – Graphics Card RAM

- Despite similar name, not related to same DDR version usually optimized for high-bandwidth
- GDDR2 – graphics card, but actually halfway between DDR and DDR2 technology wise
- GDDR3 – like DDR2 with a few other features. lower voltage, higher frequency, signal to clear bits
- GDDR4 – based on DDR3, replaced quickly by GDDR5
- GDDR5 – based on DDR3. addresses sent at double rate

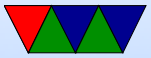


- GDDR5X – 2016
- GDDR6 – quad data rate
- GDDR6X

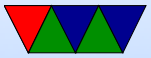


LPDDR

- Despite similar name, not related to same DDR version
- LPDDR usually soldered to motherboard. Low voltage, short traces
- LP-DDR
- LP-DDR2 – low power states, 1.2V, different bus
- LP-DDR3 – higher data rate
- LP-DDR4 – change from 10-bit DDR to 6-bit SDR bus
- LP-DDR4X – I/o voltage 0.6V
- LP-DDR5 – (2019) 6.4Gbit/s/pin, differential clocks

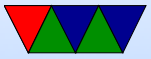


- LPCAMM2 from Micron is LPDDR5X-9600?

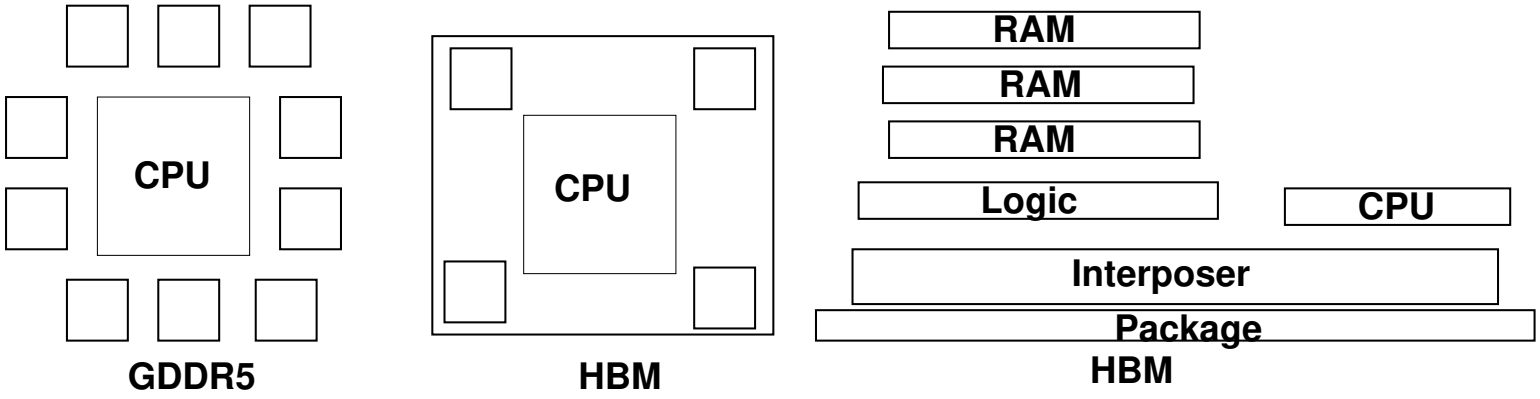


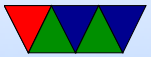
DDRL

- DDR3L - low voltage, 1.35V (not same as LPDDR3)
  DDR3U (ultra-low voltage) 1.25V
- DDR4L – does not exist?



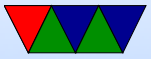
HBM RAM


CPU
GDDR5
CPU
HBM
RAM
RAM
RAM
Logic
CPU
Interposer
Package
HBM



HBM/HBM2 RAM

- HBM
  - High bandwidth memory
  - 3d-stacked RAM, stacked right on top of CPU
  - Silicon through VIA
  - Higher bandwidth, two 128-bit channels per die
  - 4096 bit wide bus compared to GDDR5 where you might have 32-bit channel times 16 chips for 512 bit
- HBM2
  - Eight dies per stack, up to 2GT/s



- HBM3/HBM4
  - Specified, doesn't exist yet? HPC?
- In newer GPUs, AMD and NVIDIA. HBM2 in new Nvidia Pascal Tesla P100