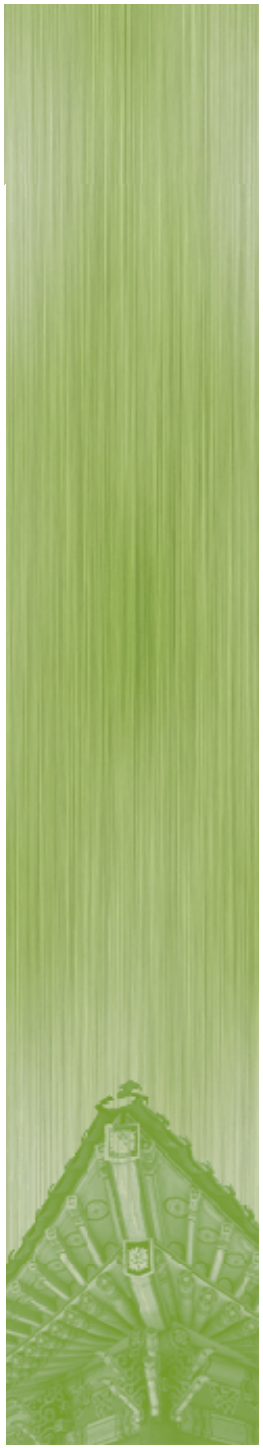


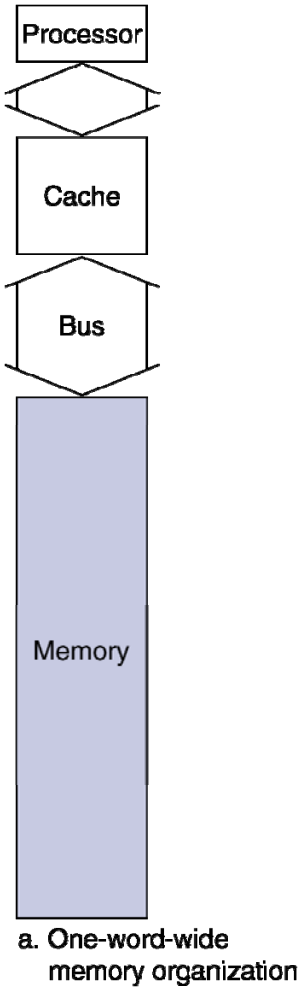
DRAMs

Jin-Soo Kim (jinsookim@skku.edu)
Computer Systems Laboratory
Sungkyunkwan University
http://csl.skku.edu

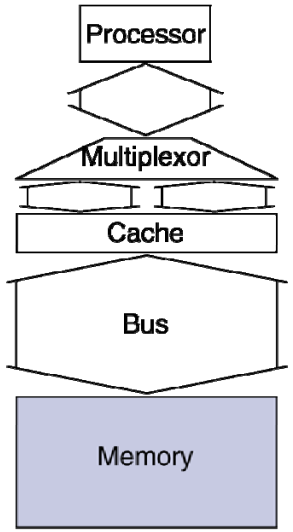
Main Memory & Caches

- **Use DRAMs for main memory**
  - Fixed width (e.g., 1 word)
  - Connected by fixed-width clocked bus
    - Bus clock is typically slower than CPU clock
- **Example cache block read**
  - 1 bus cycle for address transfer
  - 15 bus cycles per DRAM access
  - 1 bus cycle per data transfer
- **For 4-word block, 1-word-wide DRAM**
  - Miss penalty = 1 + 4 x 15 + 4 x 1 = 65 bus cycles
  - Bandwidth = 16 bytes / 65 cycles = 0.25 B/cycle

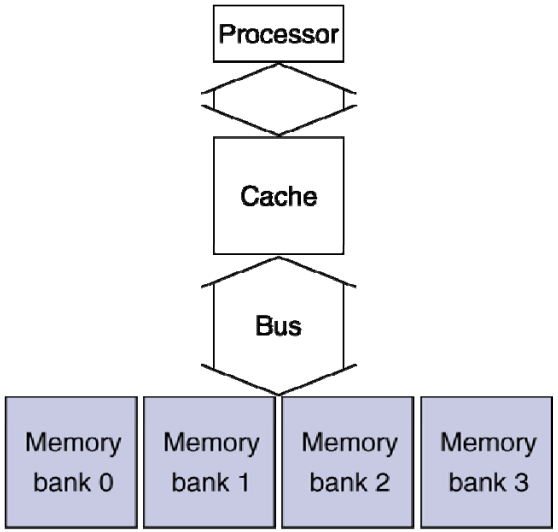
Increasing Memory Bandwidth



a. One-word-wide memory organization



b. Wider memory organization



c. Interleaved memory organization

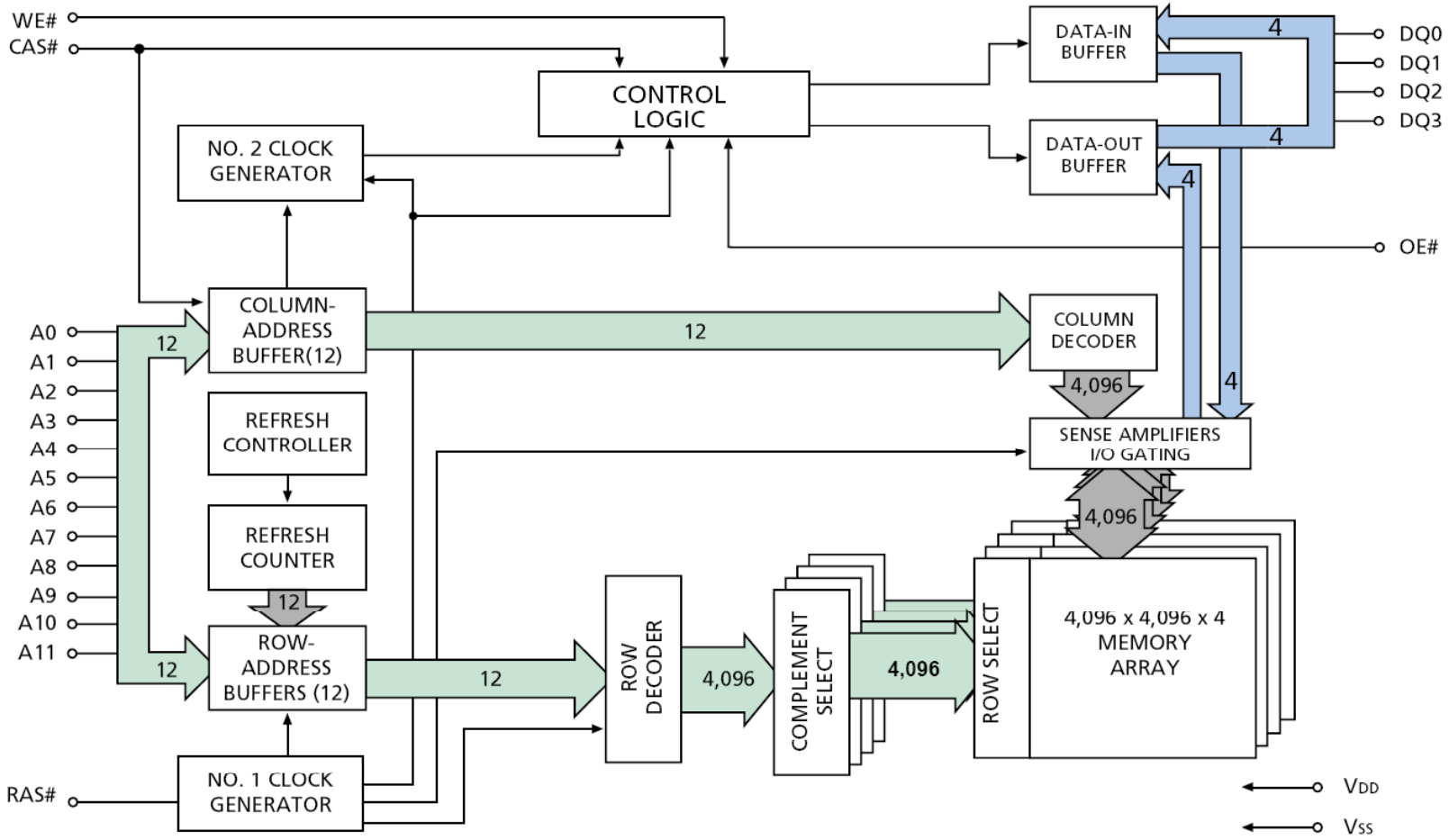
- **4-word wide memory**
  - Miss penalty = 1 + 15 + 1 = 17 bus cycles
  - Bandwidth = 16 bytes / 17 cycles = 0.94 B/cycle
- **4-bank interleaved memory**
  - Miss penalty = 1 + 15 + 4 x 1 = 20 bus cycles
  - Bandwidth = 16 bytes / 20 cycles = 0.8 B/cycle

Advanced DRAM Organization

- **Bits in a DRAM are organized as a rectangular array**
  - DRAM accesses an entire row
  - Burst mode: supply successive words from a row with reduced latency
- **Double data rate (DDR) DRAM**
  - Transfer on rising and falling clock edges
- **Quad data rate (QDR) DRAM**
  - Separate DDR inputs and outputs

DRAM Organization (1)

- Micron MT4LC16M4T8 (16M x 4bit)

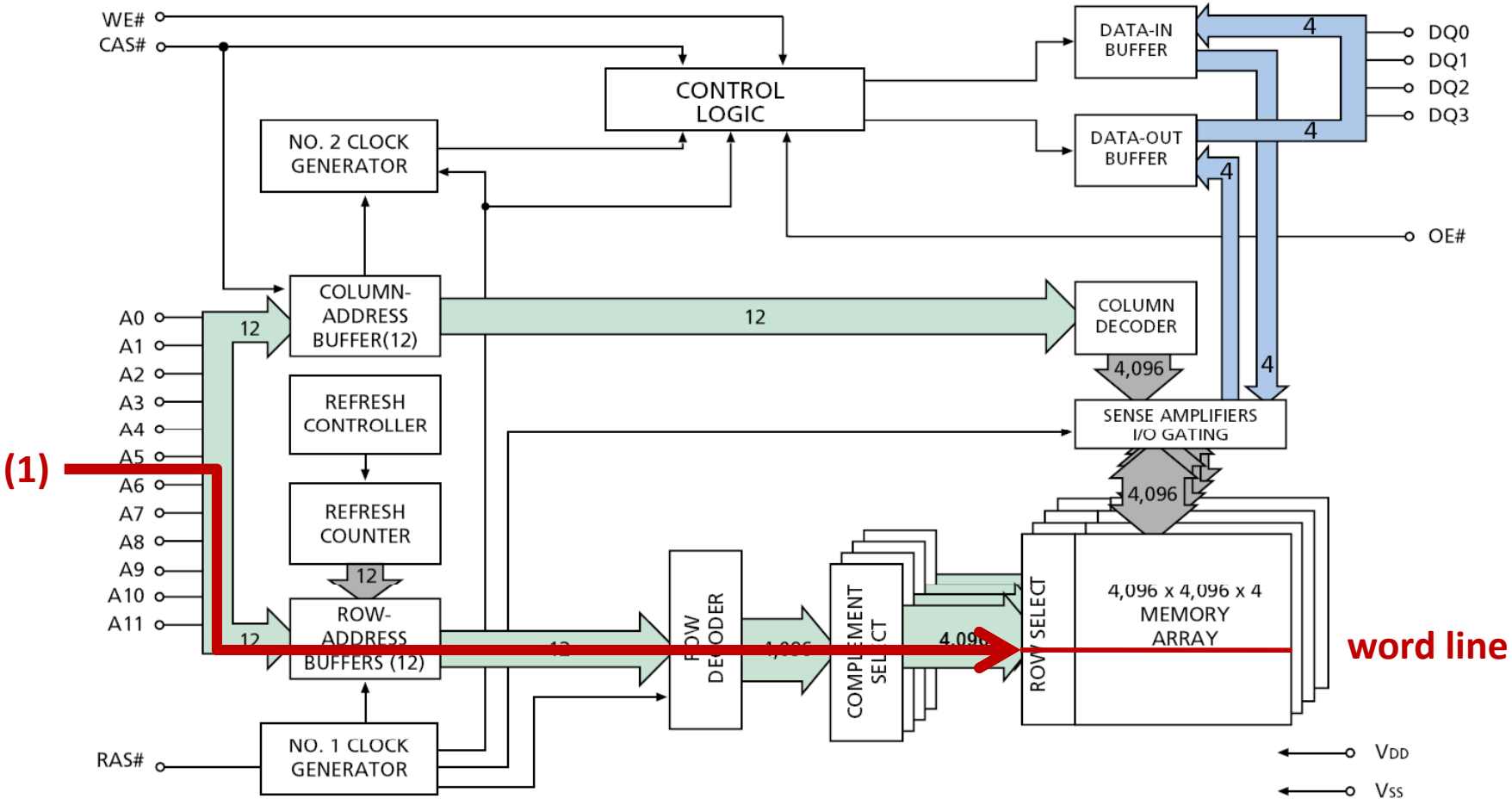


DRAM Organization (2)

- **DRAM configuration**
  - Asynchronous: no clock
  - Large capacity: 1 – 4Gb
    - Arranged as 2D matrix
    - Minimizes wire length
    - Maximizes refresh efficiency
  - Narrow data interface: 1 – 16 bits (x1, x4, x8, x16)
    - Cheap packages → few bus pins
    - Pins are expensive
  - Narrow address interface:
    - Multiplexed address lines: row and column address
    - Signaled by RAS# and CAS# respectively

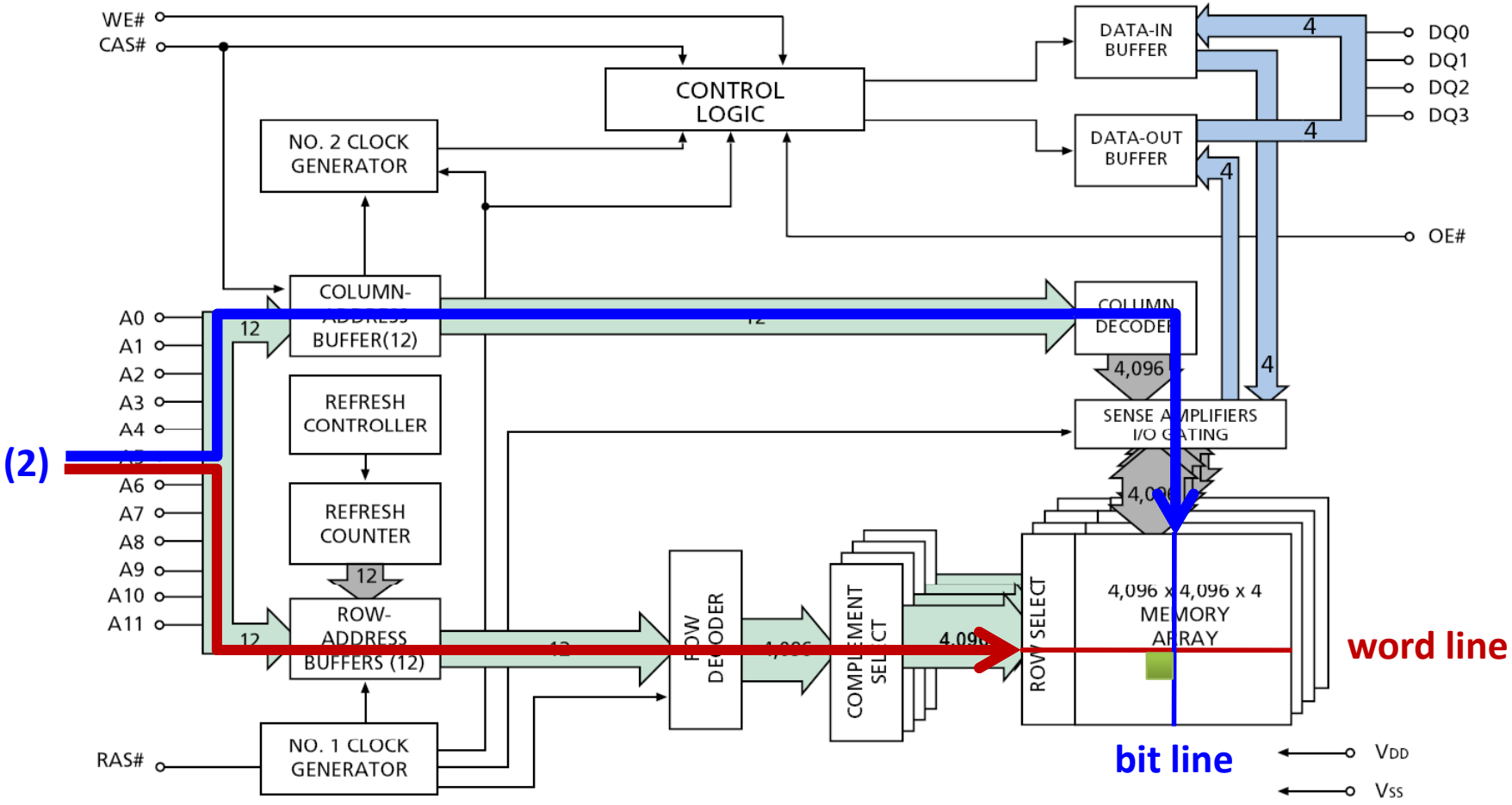
DRAM Operation (1)

- Read operation (1)



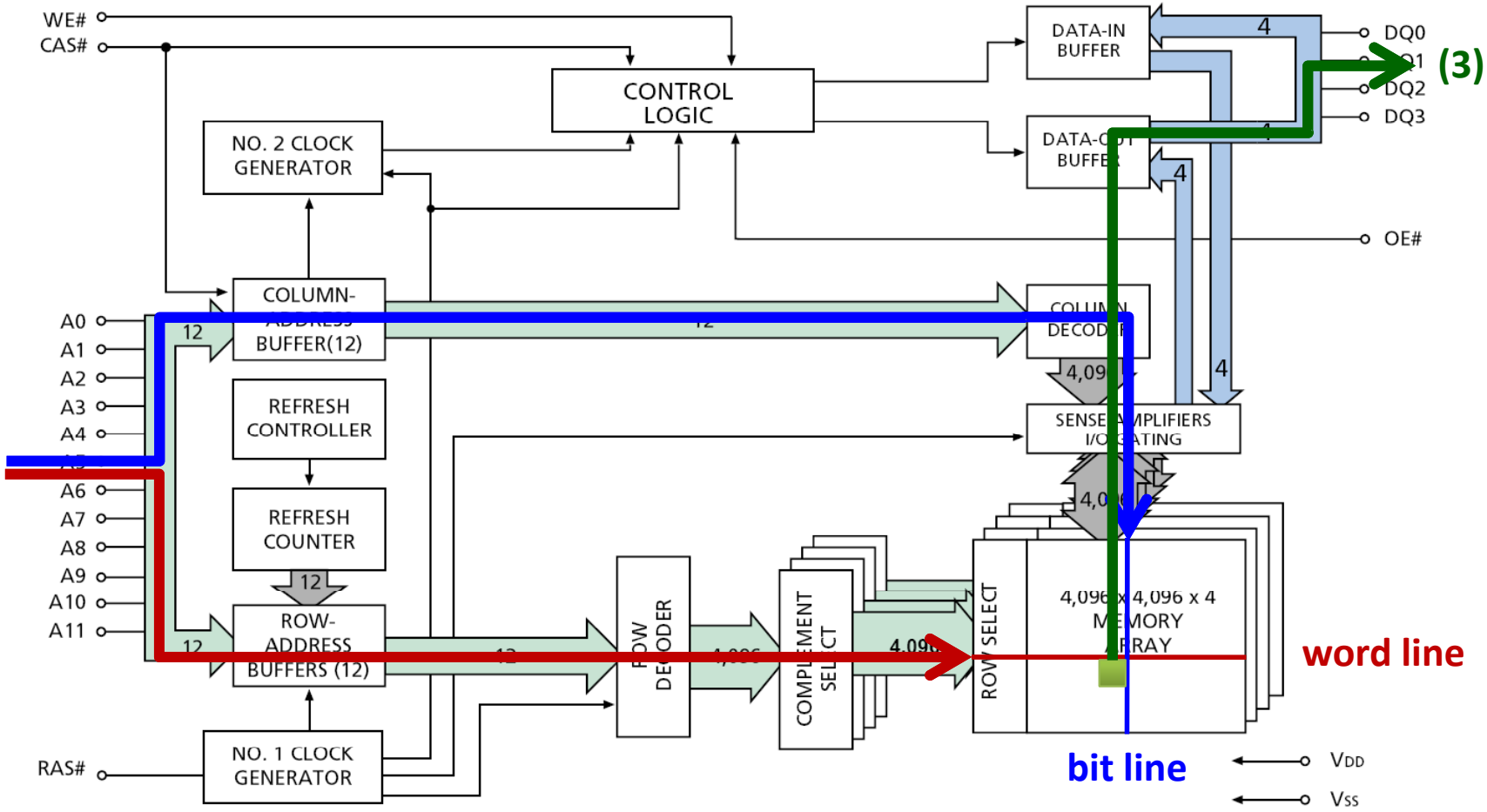
DRAM Operation (2)

- Read operation (2)



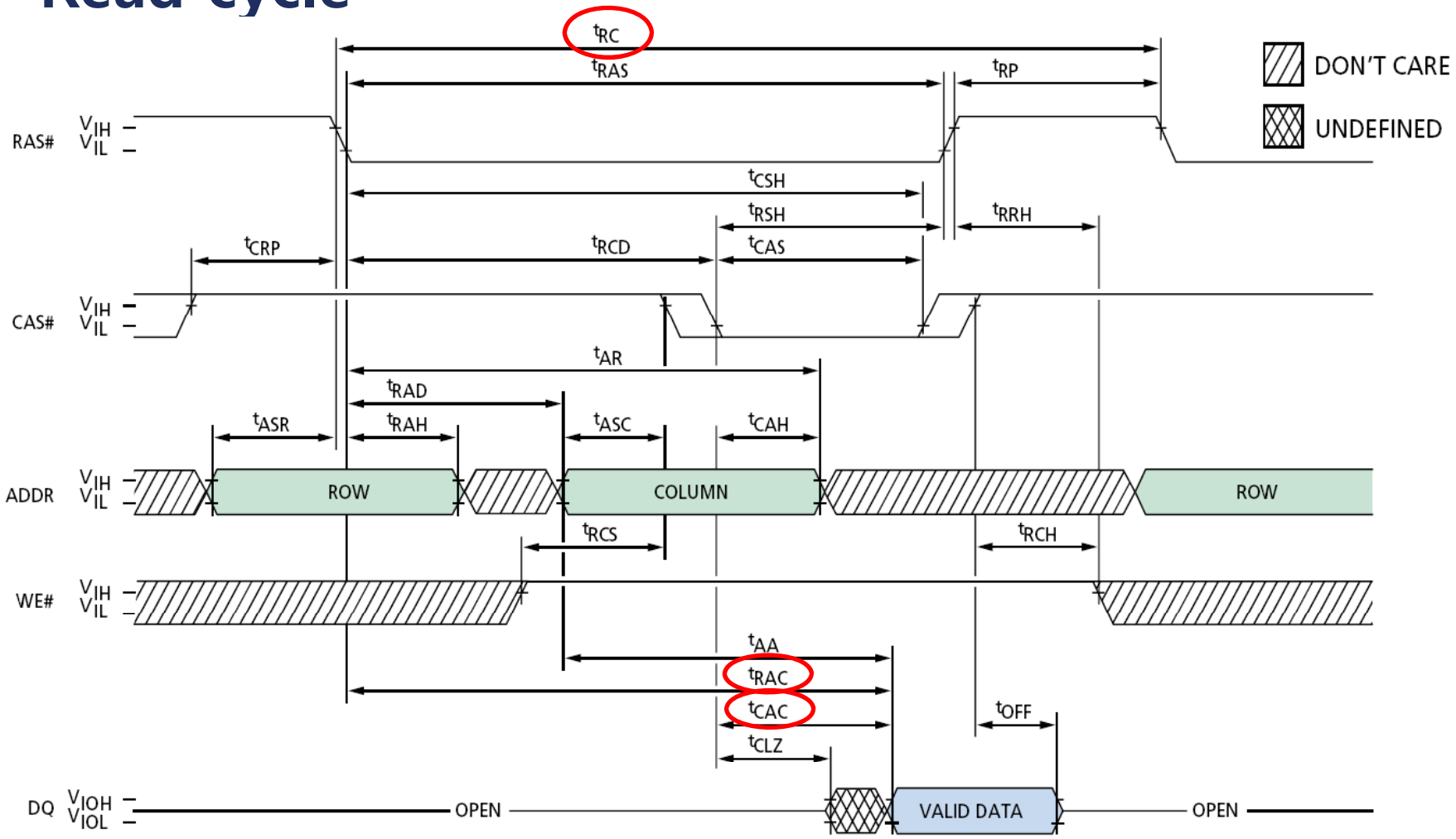
DRAM Operation (3)

- Read operation (3)



DRAM Operation (4)

- Read cycle



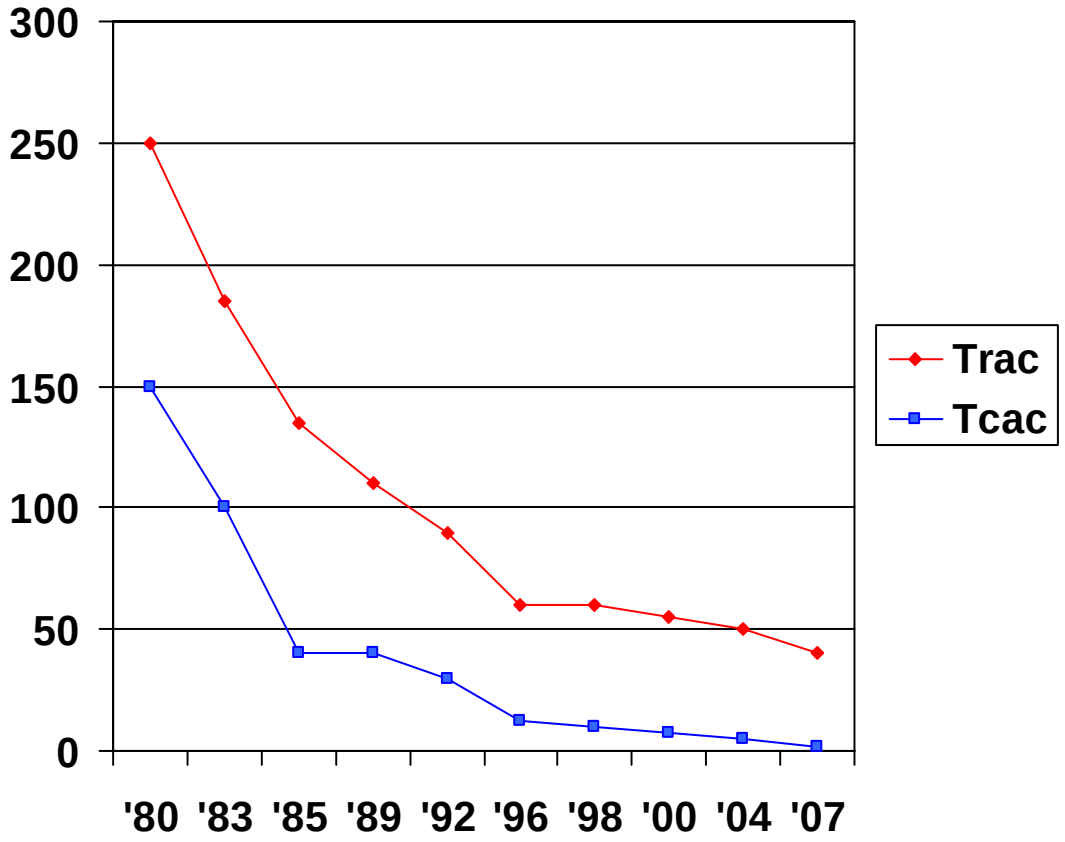
DRAM Operation (5)

- **Timing parameters**
  - t_{RC}: minimum time from the start of one row access to the start of the next (cycle time)
  - t_{RAC}: minimum time from RAS# line falling to the valid data output (access time)
    - Used to be quoted as the nominal speed of a DRAM chip
  - t_{CAC}: minimum time from CAS# line falling to valid data output

Model	t_{RC}	t_{RAC}	t_{CAC}
MT4LC16M4T8-5	90 ns	50 ns	13 ns
MT4LC16M4T8-6	110 ns	60 ns	15 ns

DRAM Generations

Year	Capacity	$/GB
1980	64Kbit	$1500000
1983	256Kbit	$500000
1985	1Mbit	$200000
1989	4Mbit	$50000
1992	16Mbit	$15000
1996	64Mbit	$10000
1998	128Mbit	$4000
2000	256Mbit	$1000
2004	512Mbit	$250
2007	1Gbit	$50



Evolution of DRAM

- **Asynchronous DRAM**
  - Fast Page Mode (FPM DRAM) 1987
  - Extended Data Out (EDO DRAM) 1993

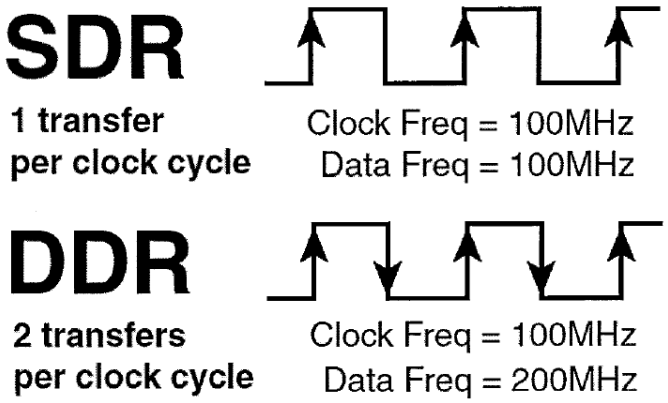
- **Synchronous DRAM**
  - Synchronous (SDRAM) 1996
  - Rambus (RDRAM) 1999
  - Double Data Rate (DDR SDRAM) 2000
  - Double Data Rate 2 (DDR2 SDRAM) 2003
  - Double Data Rate 3 (DDR3 SDRAM) 2007

SDRAM

- **(Single Data Rate) Synchronous DRAM**
  - Add a clock signal to DRAM interface
    - Eliminates time to detect RAS#/CAS# and WR# signals
    - "Command" inputs sampled on positive clock edge
    - CPU knows when data will be ready; can do something else
  - Multiple, semi-independent banks on each device
    - SDRAM command scheme allows overlapped bank operations
    - One bank is activated and accessed, while the other is precharged
    - More efficient use of pin bandwidth via interleaving
  - PC66, PC100, PC133 (66-133 MHz)

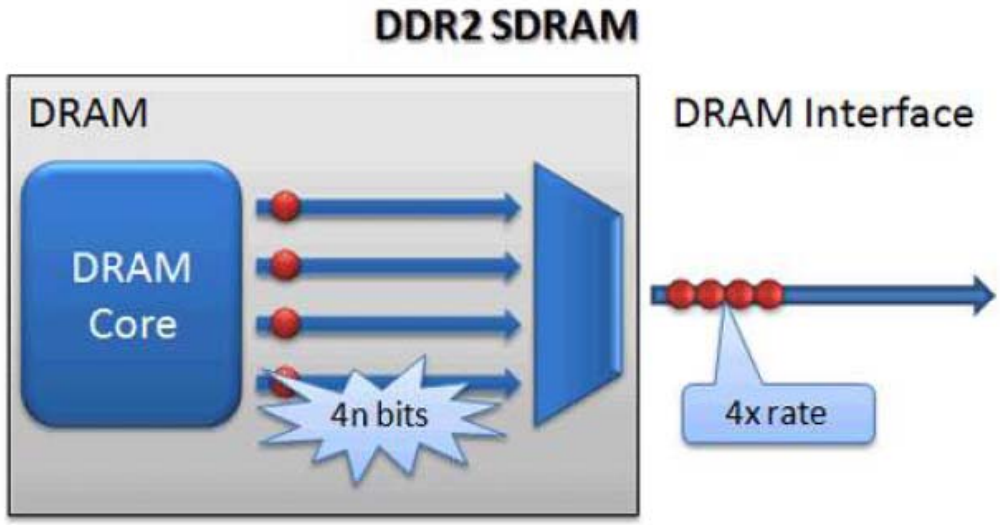
DDR/DDR2/DDR3 (1)

- **Double Data Rate (DDR) SDRAM**
  - SDRAM can only send data once per clock
  - DDR SDRAM transfers data on the rising and falling edges of the clock cycle
  - Doubles the transfer rate without increasing the frequency of the front side bus
  - Commands still sent only with positive clock edge
  - Operating voltage: 2.5 V
  - DDRx improves bandwidth, not latency

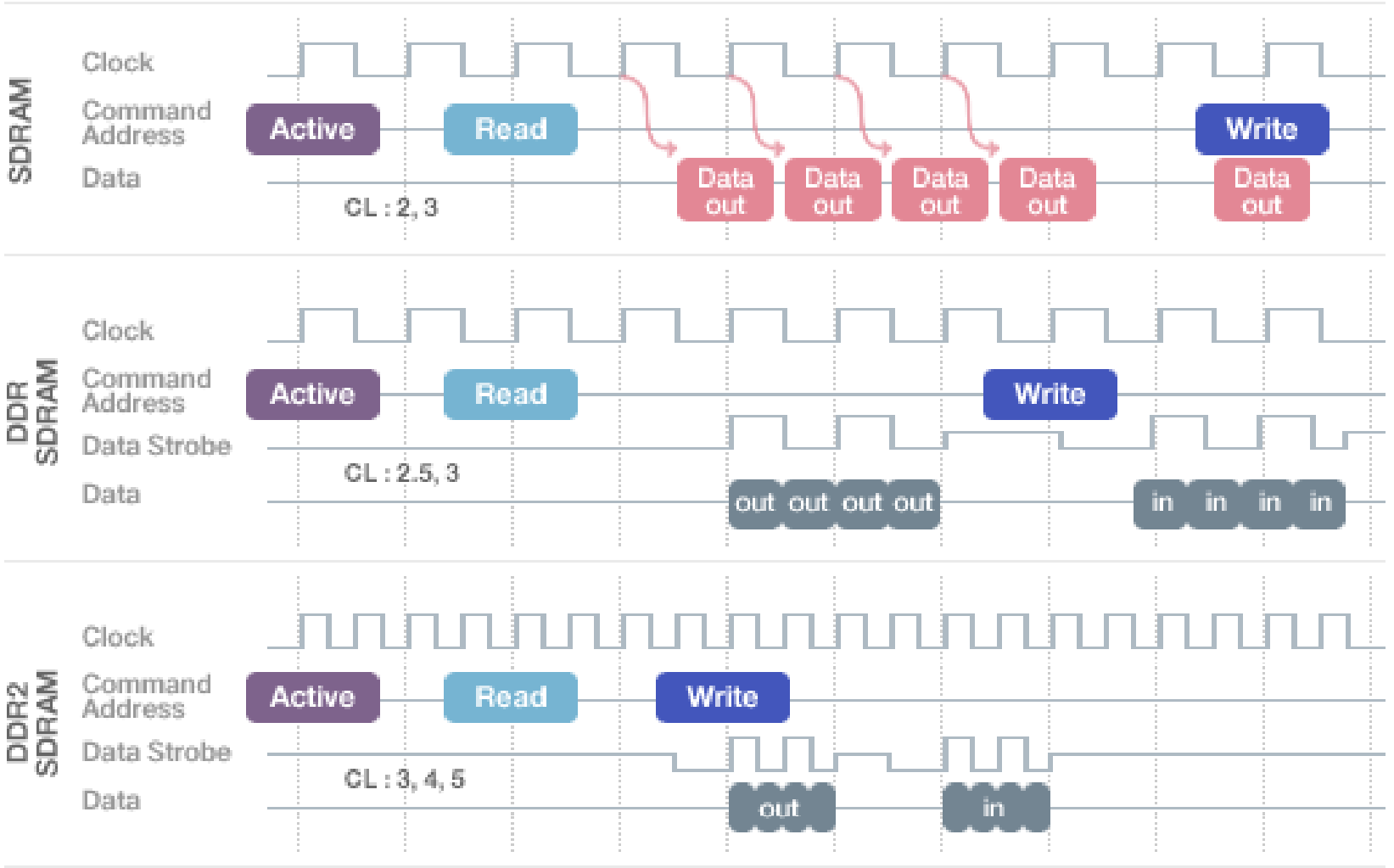


DDR/DDR2/DDR3 (2)

- **DDR2 SDRAM**
  - Read data four times over a clock cycle
  - Lowers power by dropping to 1.8V
  - Offers higher clock rates: up to 533 MHz
  - 256Mb – 4Gb (4 or 8 banks)

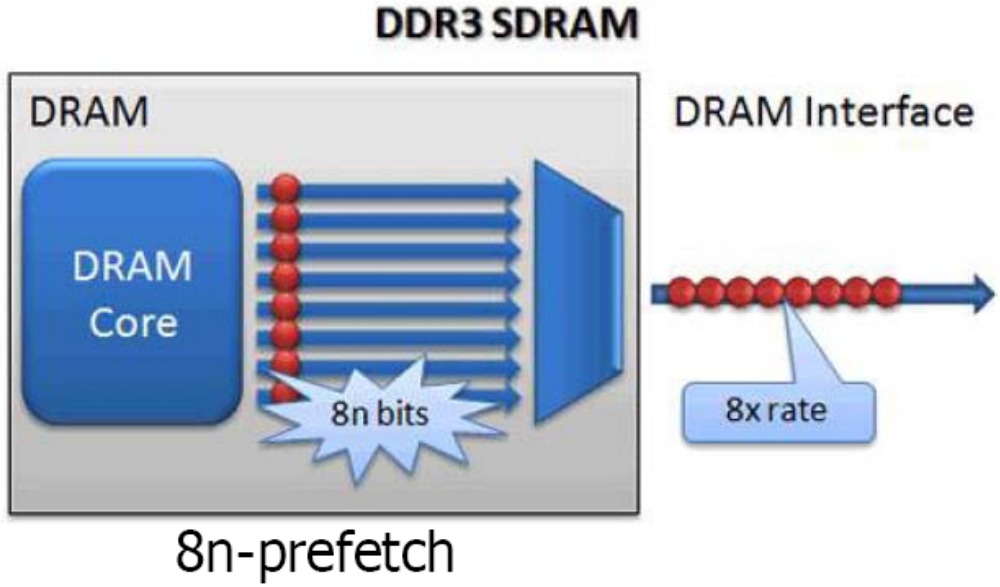


DDR/DDR2/DDR3 (3)



DDR/DDR2/DDR3 (4)

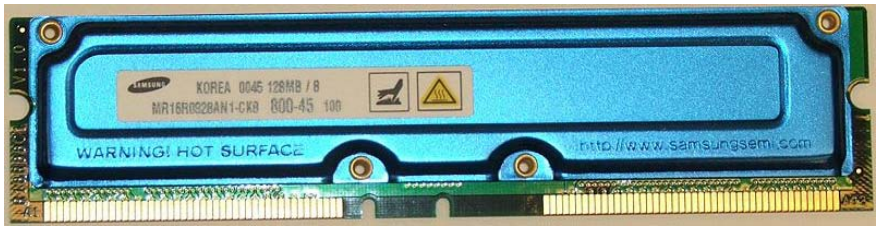
- **DDR3 SDRAM**
  - Read data eight times over a clock cycle
  - Operating at 1.5V (~30% reduction in power)
  - Offers higher clock rates: up to 800 MHz
  - 512Mb – 8Gb (8 banks)



DDR/DDR2/DDR3 (5)

Chip type	Memory clock	I/O bus clock	Data transfers/s	MB/s /DIMM	DIMM
DDR200	100 MHz	100 MHz	200 M	1600	PC1600
DDR266	133 MHz	133 MHz	266 M	2133	PC2100
DDR333	166 MHz	166 MHz	333 M	2667	PC2700
DDR400	200 MHz	200 MHz	400 M	3200	PC3200
DDR2-400	100 MHz	200 MHz	400 M	3200	PC2-3200
DDR2-533	133 MHz	266 MHz	533 M	4266	PC2-4200
DDR2-667	166 MHz	333 MHz	667 M	5333	PC2-5300
DDR2-800	200 MHz	400 MHz	800 M	6400	PC2-6400
DDR2-1066	266 MHz	533 MHz	1066 M	8533	PC2-8500
DDR3-800	100 MHz	400 MHz	800 M	6400	PC3-6400
DDR3-1066	133 MHz	533 MHz	1066 M	8533	PC3-8500
DDR3-1333	166 MHz	667 MHz	1333 M	10667	PC3-10600
DDR3-1600	200 MHz	800 MHz	1600 M	12800	PC3-12800

DRAM Modules (1)

	**SIMM (Single In-line Memory Module)** • 30-pin • 72-pin
	**DIMM (Dual In-line Memory Module)** • 168-pin • 184-pin • 240-pin
	**RIMM (Rambus In-line Memory Module)** • 184-pin
	**SO-DIMM (Small Outline Dual In-line Memory Module)** • 72-pin, 100-pin • 144-pin, 200-pin, 204-pin

DRAM Modules (2)

Modules	Pins	Data bits	Capacity	Technology
SIMM	30	8	256KB – 16MB	FPM, EDO
	72	32	1MB – 256MB	EDO
DIMM	168	64	32MB – 1GB	SDR
	184	64	128MB – 2GB	DDR
	240	64	256MB – 4GB	DDR2, DDR3
SODIMM	72	32	8MB – 32MB	FPM, EDO
	100	32	32MB – 128MB	FPM, EDO
	144	64	32MB – 256MB	FPM, EDO, SDR, DDR2
	200	64	128MB – 4GB	DDR, DDR2
	204	64	1GB – 4GB	DDR3
RIMM	184	16	64MB – 512MB	RDRAM