Digital microsystems design

Marius Marcu

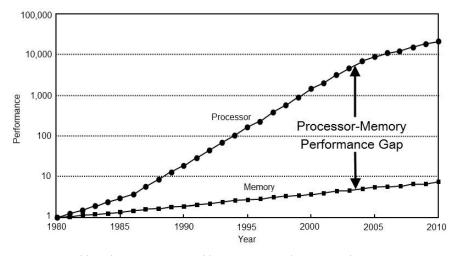
Objectives

- Specific objectives
 - Memory subsystem and how it is connected to a microprocessor

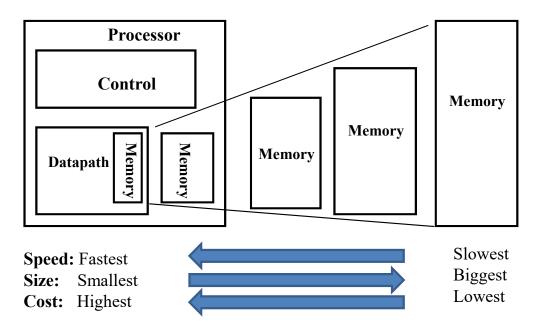
Outline

- Memory subsystem
 - Memory types
 - Memory connection
 - Data transfer diagrams
 - Memory organization

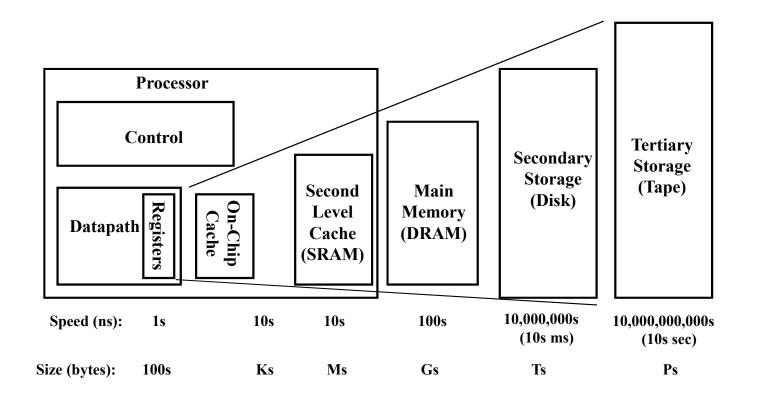
- Memory subsystem of a microprocessor system is a bottleneck
- Multi-core further increase the performance gap stalls on shared memory access



- One of the main components of a computer system, that stores programs and data
- Memory hierarchy



- The Principle of Locality:
 - Programs access a relatively small portion of the address space at any instant of time.
 - Temporal Locality (Locality in Time):
 - Keep most recently accessed data items closer to the processor
 - Spatial Locality (Locality in Space):
 - Move blocks consists of contiguous words closer to the processor
 - Why is that useful?



Memory types

- ROM Read Only Memories
- RAM Random Access Memories
 - SRAM Static RAM
 - DRAM Dynamic RAM
- Random access time is the same for all locations

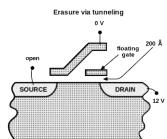
Memory types

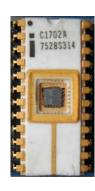
ROM

- The content is not lost when the device's power is turned off
- It contains the initialization and start-up (boot-up) binary program
- Specific methods to write its content, slower than reading
- Non-volatile memory

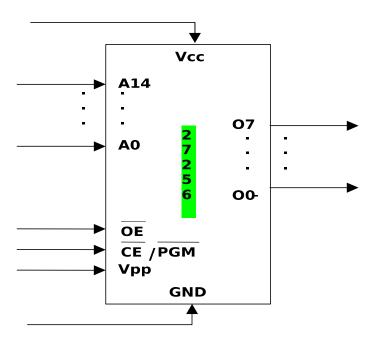
Memory types

- ROM memories
 - Read Only Memories
 - PROM Programmable ROM
 - EPROM Erasable PROM
 - EEPROM Electrical Erasable PROM
 - NVRAM (Non-Volatile RAM)
 - Flash memory floating gate transistor
 - NOR
 - NAND





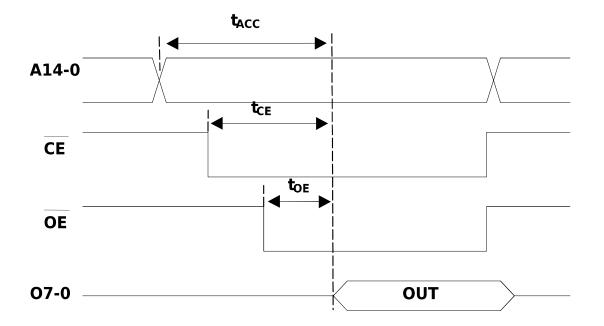
- Example
 - Memory chip EPROM 27256
 - 32 KB = 256Kbits
 - Access time (read)
 - 90-200 ns



- Example: Memory chip EPROM 27256
 - Memory content and addresses
 - Address space: 0x0000 0x7FFF
 - Each address points to a location (2^15 locations)
 - Each location stores a number of bits (data width) 8 bits

0x0000	Location 0
0x0001	Location 1
0x0002	Location 2
0x0003	
0x7FFF	Location 2^15-1

- Example: Memory chip EPROM 27256
 - Read access cycle



- Example: EPROM 27C2048
 - One-time programmable
 - 128 K x 16 bits (=256KB=2048Kbits)
 - Access time: 55 150 ns
 - Programming time: 16 s

PIN DESIGNATIONS

A0-A16 = Address Inputs

CE# (E#) = Chip Enable Input

DQ0-DQ15 = Data Input/Outputs

OE# (G#) = Output Enable Input

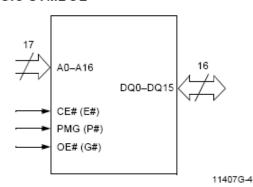
PGM# (P#) = Program Enable Input

V_{CC} = V_{CC} Supply Voltage

V_{PP} = Program Voltage Input

 V_{SS} = Ground

LOGIC SYMBOL



14

- EPROM 27C2048
 - Byte/word transfers
 - A-1 LSb address bit in BYTE mode

BYTE MODE(BYTE = GND)

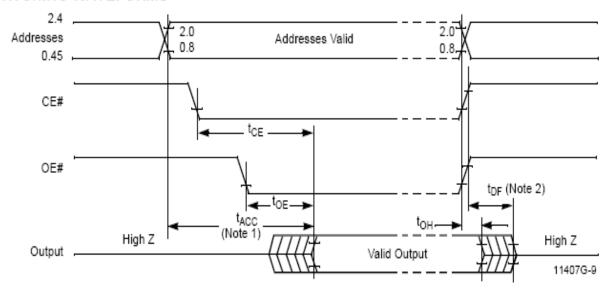
CE	ŌĒ	Q15/A-1	MODE	Q0-Q7	SUPPLY CURRENT
Н	Х	X	Non selected	High Z	Standby(ICC2)
L	Н	Х	Non selected	High Z	Operating(ICC1)
L	L	A-1 input	Selected	DOUT	Operating(ICC1)

WORD MODE(BYTE = VCC)

CE	ŌĒ	Q15/A-1	MODE	Q0-Q14	SUPPLY CURRENT
Н	Χ	High Z	Non selected	High Z	Standby(ICC2)
L	Н	High Z	Non selected	High Z	Operating(ICC1)
L	L	DOUT	Selected	DOUT	Operating(ICC1)

NOTE: X = H or L

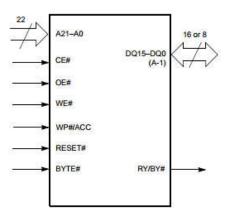
- EPROM 27C2048
 - Read memory access diagram switching waveforms



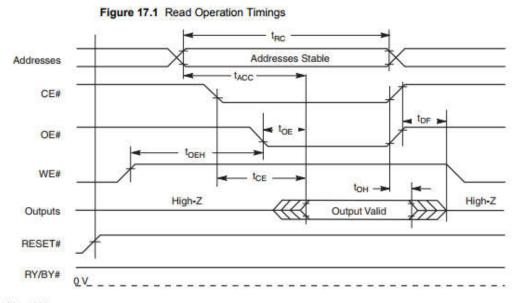
Notes:

- OE# may be delayed up to t_{ACC} t_{OE} after the falling edge of the addresses without impact on t_{ACC}.
- 2. t_{DF} is specified from OE# or CE#, whichever occurs first.

- S29GL064S
 - 64 Mbit (8MB/4Mx16bits)
 - Byte/word operation
 - 70ns access time



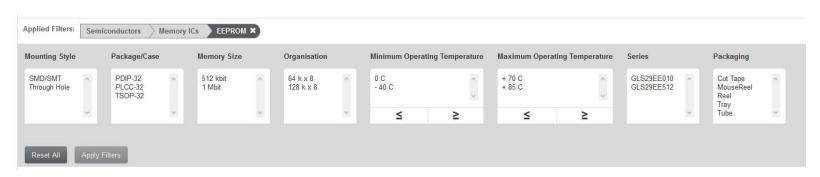
• S29GL064S



Matan

- 1. Word Configuration: Toggle A0, A1, and A2.
- 2. Byte Configuration: Toggle A-1, A0, A2, and A3.

- Flash memories
 - Data access interface
 - Serial interface
 - Parallel interface
 - Manufacturing technology
 - NAND
 - NOR
 - Package



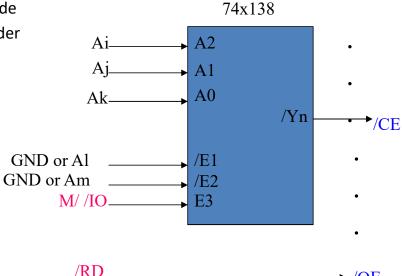
NOR

- Random access to any memory location
- Short read times
- Slower write times
- Allow code execution
- Used to store BIOS, firmware

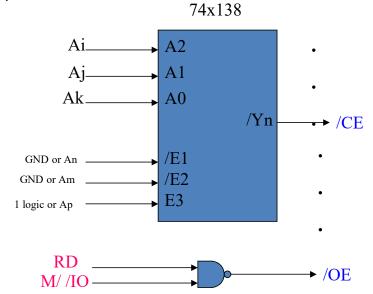
NAND

- Faster writes
- Slower reads
- Cheaper
- Smaller area per bit
- Block transfers
- Higher endurance (number of write cycles)
- Used as storage device
- I/O access

- EPROM commands signal generation (27C2048):
 - uP address lines A17-A0 are connected to memory address lines
 - A17-1 (uP) -> A16-0 (memory)
 - A0 -> A-1 (only for BYTE transfers)
 - Most significant address lines are connected to memory decoder inputs
 - Changing lines to the selection inputs of the decode
 - · Not changing lines to the enabling inputs of decoder
 - /OE, /CE



- EPROM commands signal generation:
 - uP address lines A17-A0 are connected to memory address lines
 - A17-1 (uP) -> A16-0 (memory)
 - A0 -> A-1 (only for BYTE transfers)
 - /OE, /CE



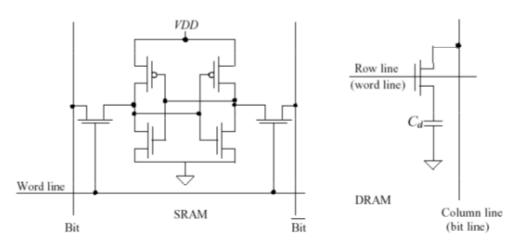
SRAM

- Low density, high power, expensive, but fast
- Static content will last as long the memory is powered

DRAM

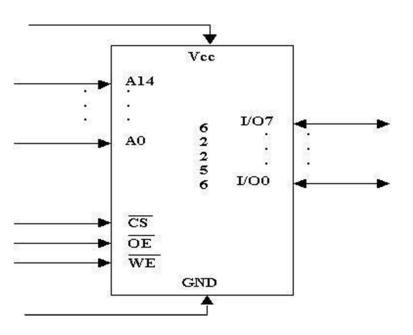
- High density, low power, cheap, but slow
- Dynamic need refresh regularly to keep memory content

• SRAM vs. DRAM

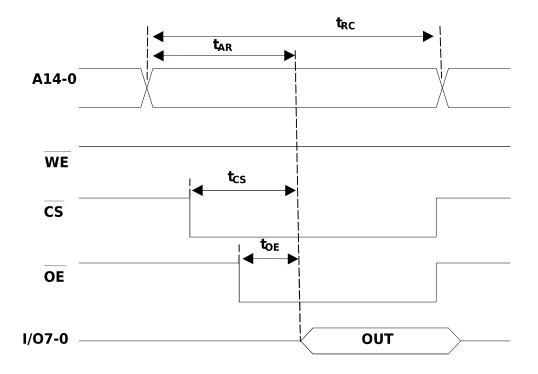


	Tran. per bit	Access time	Persist?	Sensitive?	Cost	Applications
SRAM	6	1X	Yes	No	100x	cache memories
DRAM	1	10X	No	Yes	1X	Main memories, frame buffers

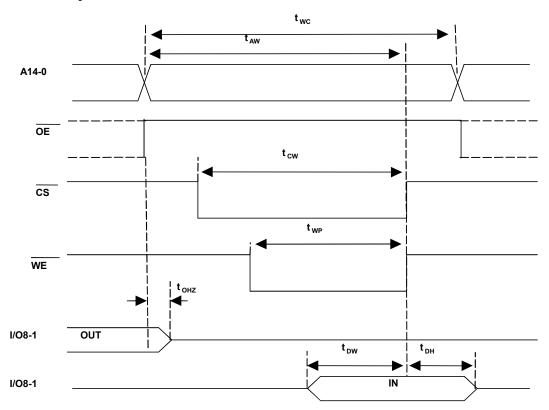
- SRAM memory chip 62256:
 - 32 KB (256 Kbits)
 - Access time: 45 84 ns;



• SRAM read access cycle



• SRAM write access cycle



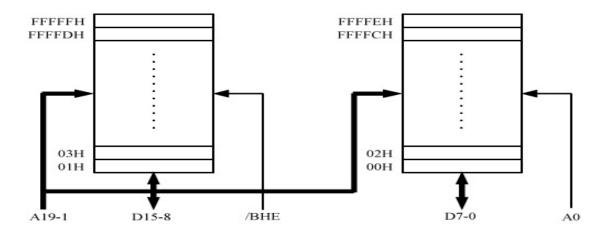
Data transfers

- 16 bits processors have to implement both:
 - 16 bits transfers
 - 8 bits transfers
- 8086: /BHE and A0

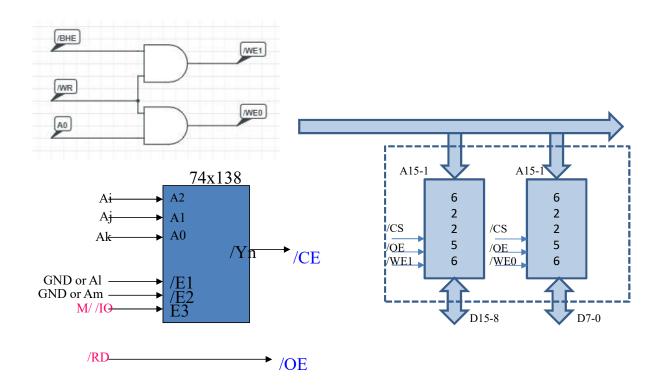
BHE	A ₀	Characteristics		
0	0	Whole word		
0	1	Upper byte from/to odd address		
1	0	Lower byte from/to even address		
1	1	None		

Data transfers

- 16 bits transfers even address (A0 = 0 , /BHE = 0)
- 8 bits transfers
 - Odd address (A0 = 1, /BHE = 0)
 - Even address (A0 = 0, /BHE = 1)

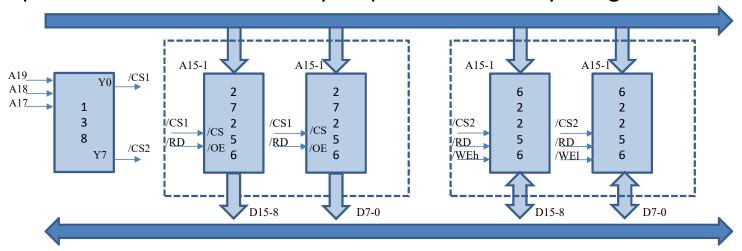


• SRAM setup



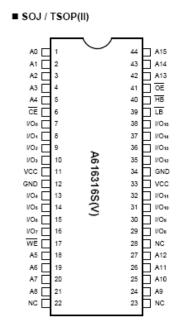
Data transfers

- 8 bits memory connection (example)
 - Group 2 circuits into one memory block for 16 bits bus
 - Provide access at byte and word level
- Example determine the memory map for the memory design bellow

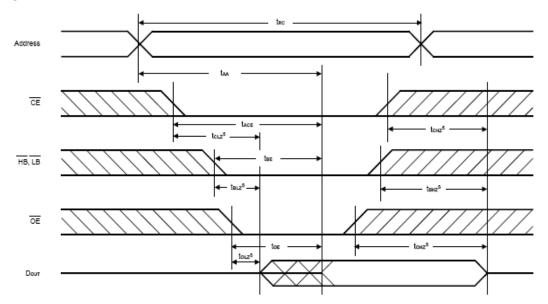


- SRAM memory chip A616316
 - 64 K x 16bits,
 - LB,HB control signals
 - High speed memory, access time: 12/15 ns (max),
 - TTL inputs and outputs, 3-states outputs,
 - Current consumption:
 - Normal operation: 170 mA (12 ns), 165 mA (15 ns),
 - Stand-by: 25 mA (TTL), 8 mA (CMOS).

- SRAM memory chip A616316
 - A15-0
 - **IO15-0**
 - LB,HB
 - WE, CE, OE



• Read cycle 2(1, 2, 3)

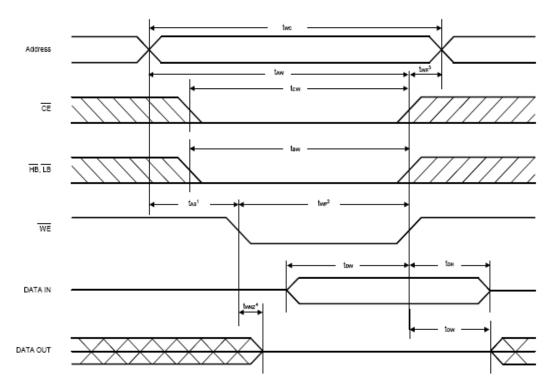


Notes: 1. WE is high for Read Cycle.

- Device is continuously enabled \(\overline{CE} = V_{IL}, \overline{HB} = V_{IL} \) and, or \(\overline{LB} = V_{IL}.\)
 Address valid prior to or coincident with \(\overline{CE} \) and \((\overline{HB} \) and, or \(\overline{LB} \)) transition low.
- 4. OE = VIL.
- 5. Transition is measured ±500mV from steady state. This parameter is sampled and not 100% tested.

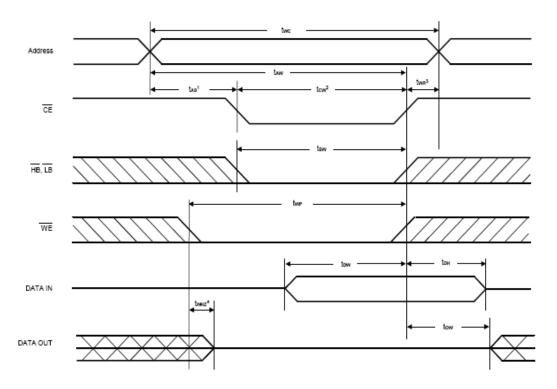
• Write cycle

Write Cycle 1 (Write Enable Controlled)

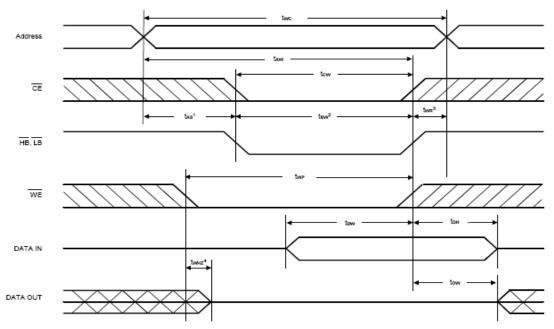


• Write cycle

Write Cycle 2 (Chip Enable Controlled)

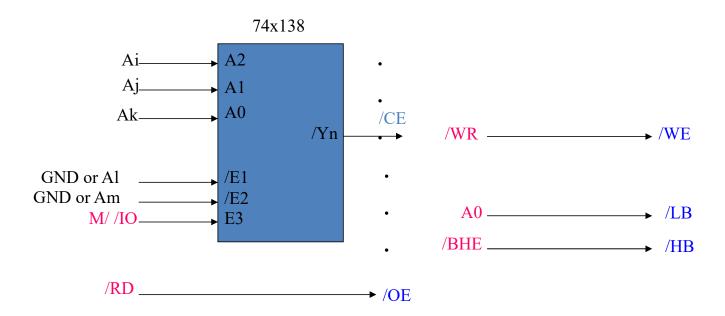


Write Cycle 3 (Byte Enable Controlled)

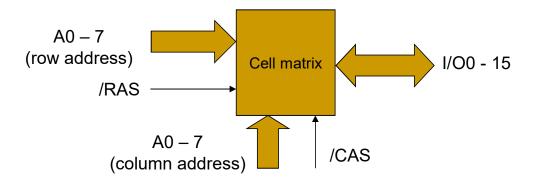


- Notes: 1. tas is measured from the address valid to the beginning of Write.
 - 2. A Write occurs during the overlap (twp, tsw) of a low \overline{CE} , \overline{WE} and (\overline{HB} and, or \overline{LB}).
 - 3. twn is measured from the earliest of $\overline{\text{CE}}$ or $\overline{\text{WE}}$ or $(\overline{\text{HB}}$ and, or $\overline{\text{LB}})$ going high to the end of the Write cycle.
 - 4. OE level is high or low.
 - 5. Transition is measured ±500mV from steady state. This parameter is sampled and not 100% tested.

- SRAM command signals generation
 - Address lines A16-A1 connected to memory address pins
 - Most significant address lines are connected to memory decode selection inputs



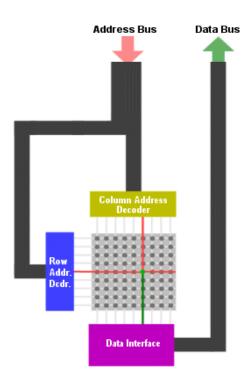
- Dynamic RAM memory
- Each bit is stored by a MOS transistor (capacitor)
- Multiplexed addresses
 - Row address
 - Column address
- Matrix structure



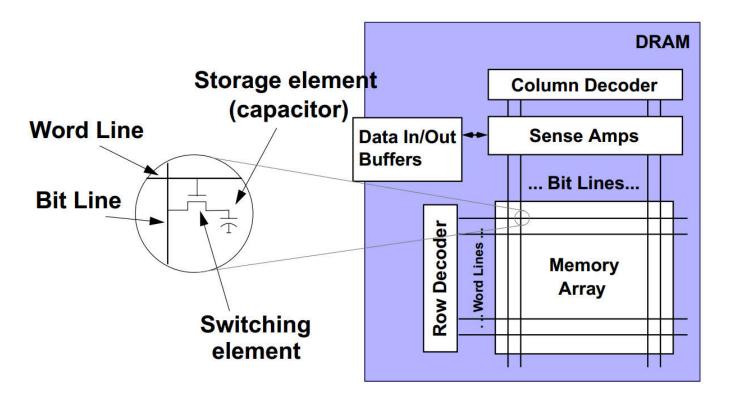
- DRAM memory
 - /RAS row address select signal
 - /CAS column address select signal
 - Memory refresh
 - Refresh cycle 4 ms
 - Row level refresh row read
 - Every MOS transistor needs to be refreshed in order to keep the logical value
 - Fast access block cycles
 - Consecutive locations of the same row

- DRAM access
 - A square grid the address splits in half
 - Half bits for row, other half for column
- Multiplex address pins
 - Read row & column address on two edges
 - Saves space, money
- Typically there are more columns than rows
 - Better row buffer hit rate
 - Less time spent refreshing (just a row read)

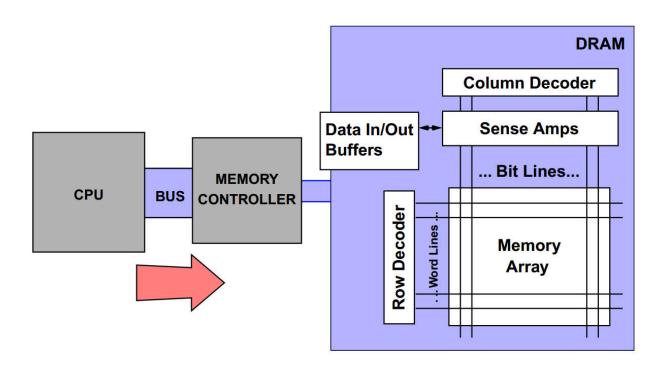
- Multiplexed address is latched on successive clock cycle
 - 1 bit / memory location



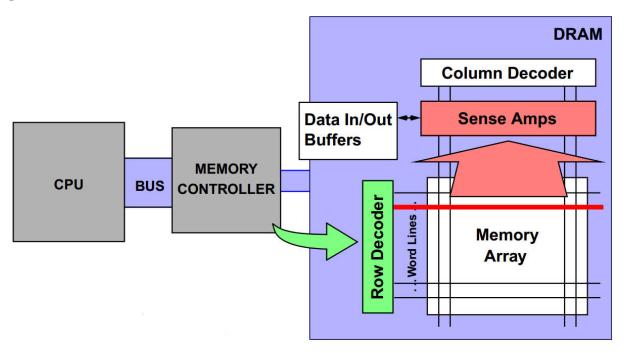
DRAM components



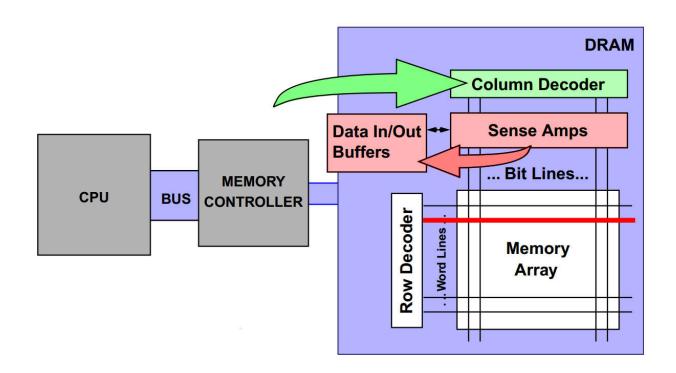
Read request



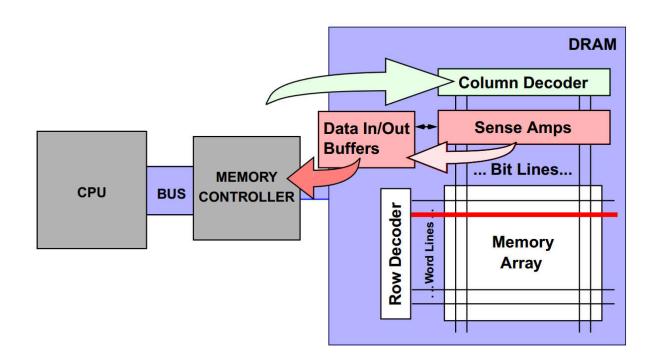
- Read command
 - Row selection (Row address & RAS signal)
 - Load row buffer



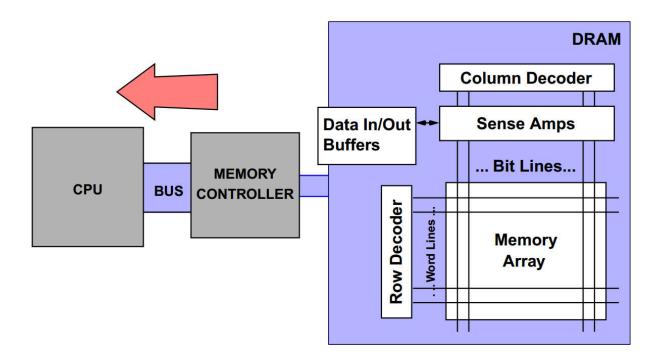
- Read command
 - Column selection (column address & CAS signal)



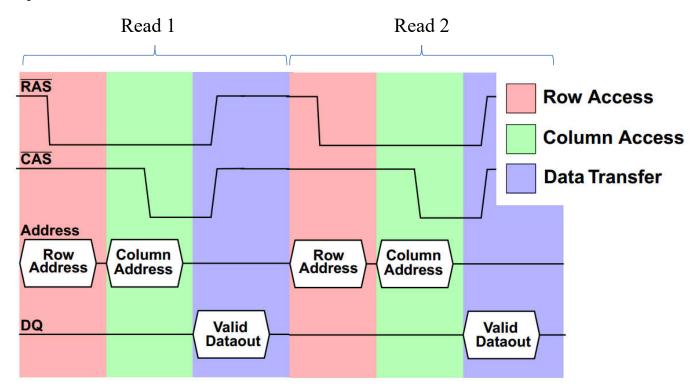
- Data transfer
 - From row buffer



- Bus transfer
 - FSB



Read cycles for conventional DRAM



- DRAM memory chip IC41C1665
 - 64K x 16 bits
 - 8 address pins
 - Multiplexed gives 16 bits addresses
 - Read and write cycles on 8 or 16 bits
 - Two CAS signals, one selection signal per byte

A0-A7	Address Inputs			
I/O0-I/O15	Data Inputs/Outputs			
WE	Write Enable			
ŌĒ	Output Enable			
RAS	Row Address Strobe			
UCAS	Upper Column Address Strobe			
LCAS	Lower Column Address Strobe			
Vcc	Power			
GND	Ground			
NC	No Connection			

• DRAM memory chip IC41C1665

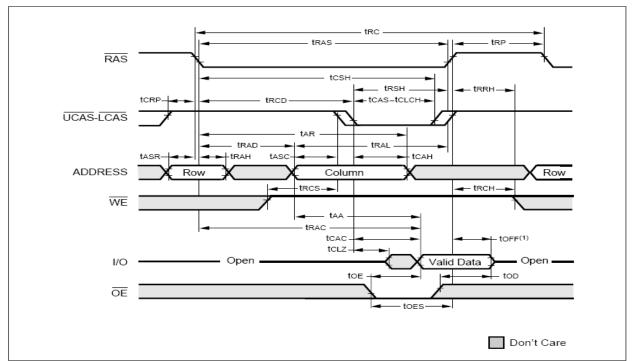
Function	RAS	LCAS	UCAS	WE	ŌΕ	Address tr/tc	I/O
Standby	Н	Н	Н	Χ	Χ	X	High-Z
Read: Word	L	L	L	Н	L	ROW/COL	Dout
Read: Lower Byte	L	L	Н	Н	L	ROW/COL	Lower Byte, Dout Upper Byte, High-Z
Read: Upper Byte	L	Н	L	Н	L	ROW/COL	Lower Byte, High-Z Upper Byte, Dout
Write: Word (Early Write)	L	L	L	L	Х	ROW/COL	Din
Write: Lower Byte (Early Write)) L	L	Н	L	Х	ROW/COL	Lower Byte, DIN Upper Byte, High-Z
Write: Upper Byte (Early Write)) L	Н	L	L	Х	ROW/COL	Lower Byte, High-Z Upper Byte, DIN
Read-Write ^(1,2)	L	L	L	H→L	L→H	ROW/COL	Dout, DIN
Hidden Refresh ²⁾	Read L→H→L	L	L	Н	L	ROW/COL	Dout
	Write $L \rightarrow H \rightarrow L$	L	L	L	X	ROW/COL	Din
RAS-Only Refresh	L	Н	Н	X	Х	ROW/NA	High-Z
CBR Refresh ⁽³⁾	H→L	L	L	Χ	Χ	Х	High-Z

Notes:

- These WRITE cycles may also be BYTE WRITE cycles (either LCAS or UCAS active).
 These READ cycles may also be BYTE READ cycles (either LCAS or UCAS active).
 At least one of the two CAS signals must be active (LCAS or UCAS).

Read cycle

READ CYCLE



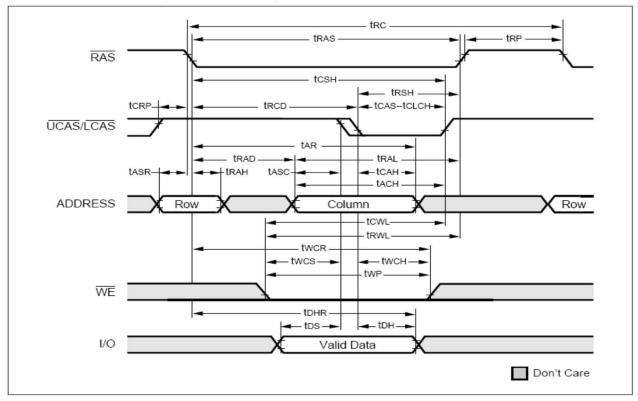
Note:

toff is referenced from rising edge of RAS or CAS, whichever occurs last.

- Read cycle
 - Set row address /RAS
 - Set column address CAS (/LCAS, /UCAS)
 - Activate /OE
 - Column address has to be kept valid for a specific amount of time
 - Data (8 or 18 bits) is valid after specific access times

Write cycle

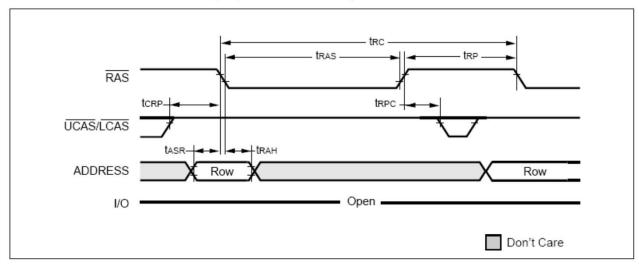
EARLY WRITE CYCLE (OE = DON'T CARE)



- Write cycle
 - Set row address /RAS
 - Activate /WE
 - Set column address CAS (/LCAS, /UCAS)
 - Column address has to be kept valid for a specific amount of time
 - Data has to be kept valid for a specific time

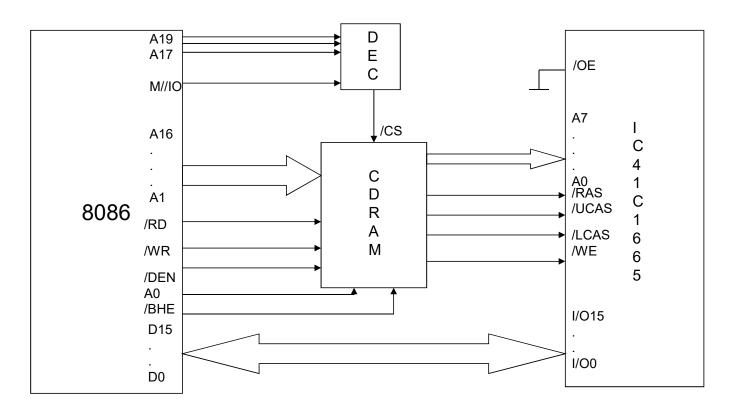
- Refresh cycle
 - Row refresh

RAS-ONLY REFRESH CYCLE (OE, WE = DON'T CARE)



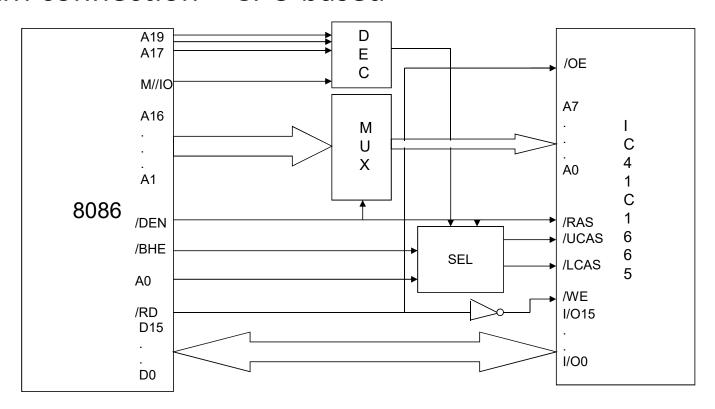
- DRAM refresh cycle
 - Read and write accesses do refresh the logical addressed locations
 - Every location has to be refreshed at 4 ms
 - 256 refresh cycles at 4 ms in order to refresh all memory locations
 - Two refresh modes
 - /RAS-only A7-0 are set by external components and RAS is activated
 - CBR (/CAS before /RAS) /RAS after /CAS, addresses are generated internally

• DRAM connection – controller based



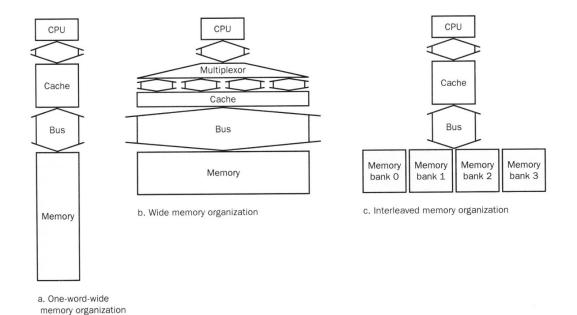
- DRAM controller
 - Multiplex row and column addresses
 - Generates row address for refresh cycles
 - Generates /RAS, /LCAS and /UCAS out of processor's signals (A0, /BHE)
 - Generates /OE and /WE
- Advantages
 - Simplified system design
 - Refresh is executed by the controller
- Disadvantages
 - Controller design
 - Controller software

DRAM connection – CPU based

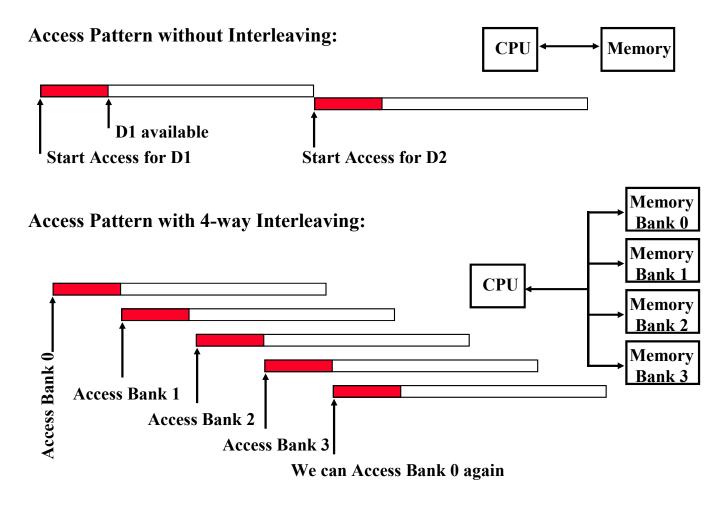


- DRAM connection CPU based
 - MUX multiplexes row and column address lines out of CPU address bus
 - SEL generates CAS signals out of AO, /BHE and /RAS (delayed)
- Advantages
 - Does not require specialized controller
- Disadvantages
 - More complex system design
 - Refresh by the processor

- Memory organization
 - Modern processors have internal caches
 - One cache line is read/written



62



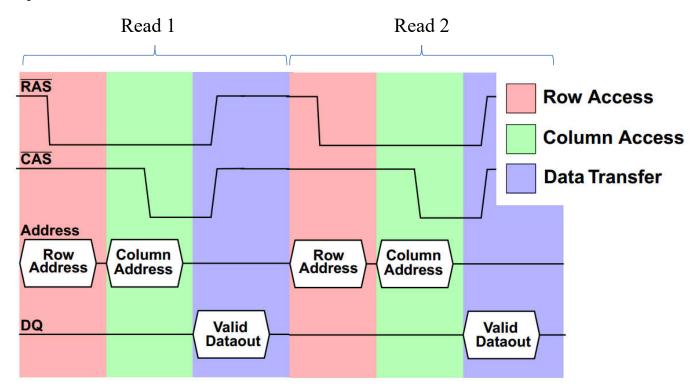
DRAM evolution

- DRAM
 - Access [RAS,CAS], [RAS,CAS],...
- FPM DRAM
 - Fast page mode DRAM
 - Access contents of row with [RAS, CAS, CAS, CAS, CAS] instead of [(RAS,CAS), (RAS,CAS), (RAS,CAS)].
 - Burst access
- EDO DRAM
 - Extended data out DRAM
 - Shorter CAS signals
- BEDO DRAM
 - Burst EDO
- SDRAM
 - Synchronous DRAM
 - Driven with rising clock edge instead of asynchronous control signals

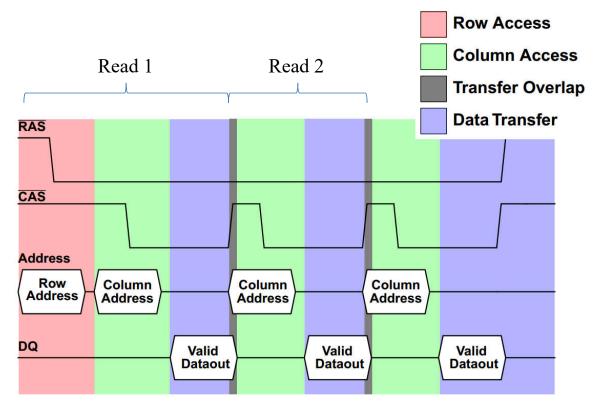
DRAM evolution

- DDR SDRAM (2000)
 - Double data-rate synchronous DRAM
 - Enhancement of SDRAM that uses both clock edges as control signals
- DDR2 SDRAM (2003)
- DDR3 SDRAM (2007)
- DDR4 SDRAM (2014) Intel Haswell

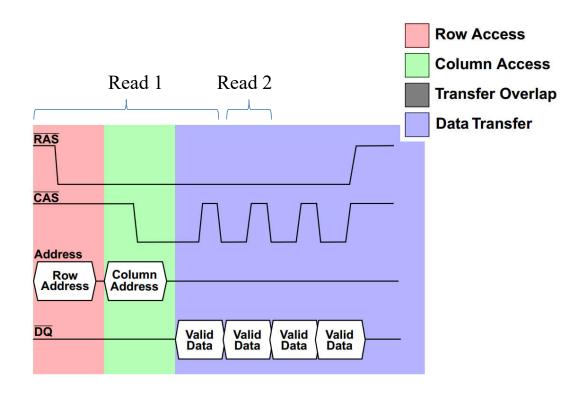
Read cycles for conventional DRAM



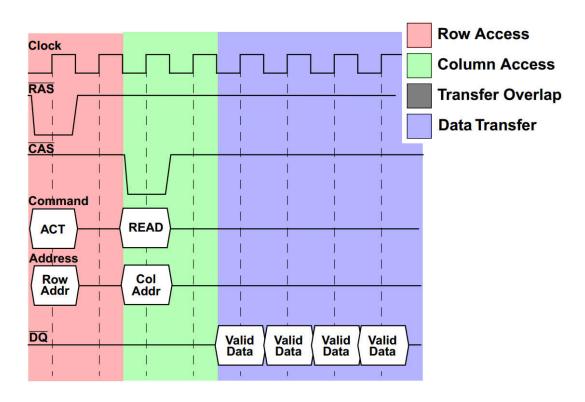
Read cycles for FPM



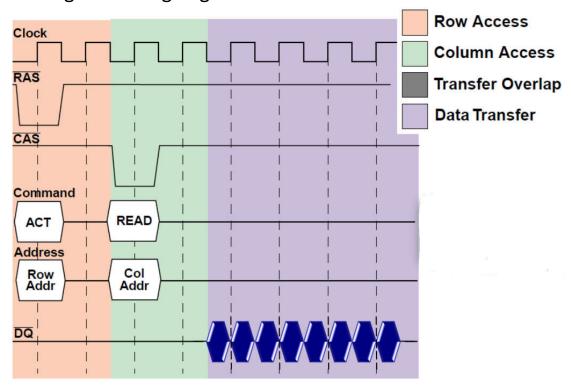
Read cycles for Burst EDO



Read cycles for SDRAM

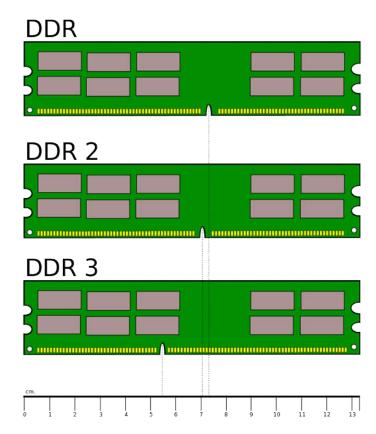


- Read cycles for DDR
 - Transfer data on both rising and falling edge of the clock



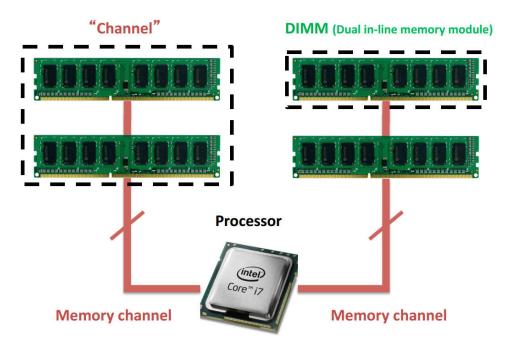
- Memory models are not backward or forward compatible
 - Different signals, protocol/timings, voltages
- DDR3 two times faster and consumes 30% less power than DDR2
 - DDR3 transfer data at a rate of 800–2133 MTps using both rising and falling edges of a 400–1066 MHz clock
 - DDR2 400–1066 MTps using a 200–533 MHz clock
 - DDR 200–400 MTps using a 100–200 MHz clock
- DDR4 is two times faster than DDR3

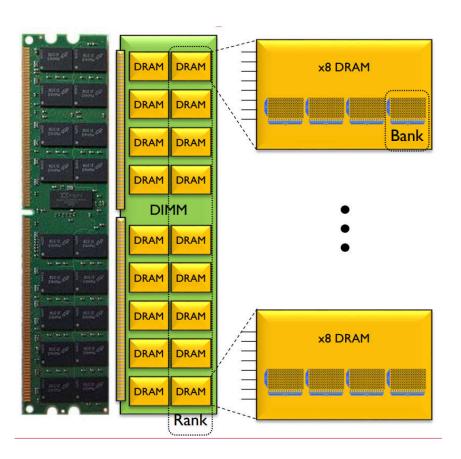
- Main memory interface types
 - RIMM Rambus Inline Memory Module
 - SIMM single inline memory module
 - DIMM Dual Inline Memory Module



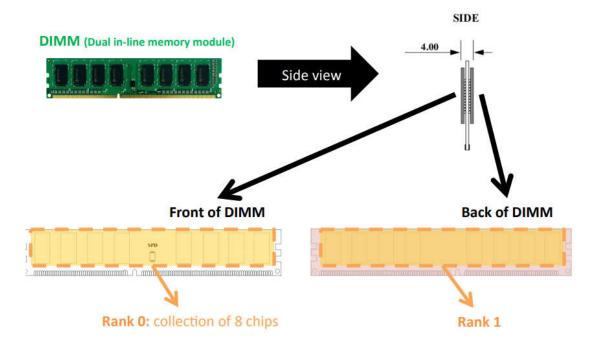
- Structures
 - DDR Channels
 - Single-channel, dual-channel, quad-channels
 - DDR Ranks
 - 1, 2, 4
 - DRAM Chips
 - 16, 8, 9
 - DRAM Banks
 - 1, 2, 4, 8
 - DRAM Rows
 - DRAM Columns
 - DRAM data bits

- DRAM channels
 - Separate paths/channels to independent memories
 - Symultaneous access from the controller

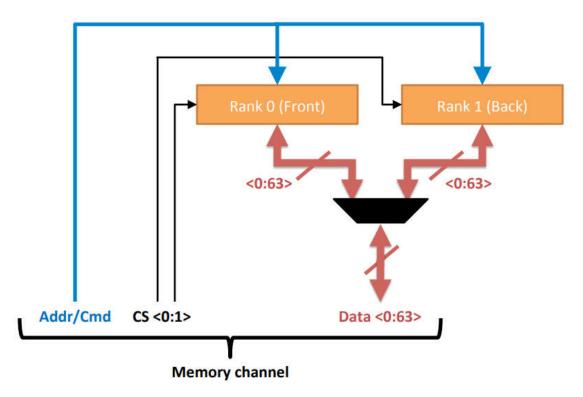




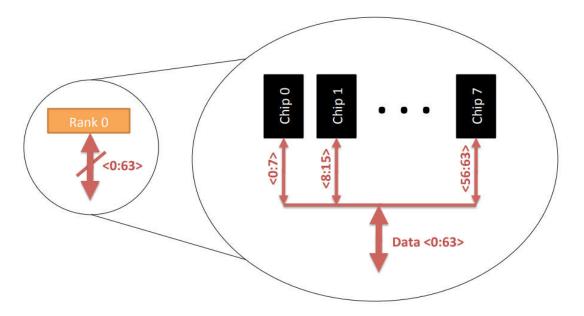
- Memory ranks
 - Each rank has dedicated CS signal (a rank is a mem block)
 - 1, 2, 4 ranks / DIMM



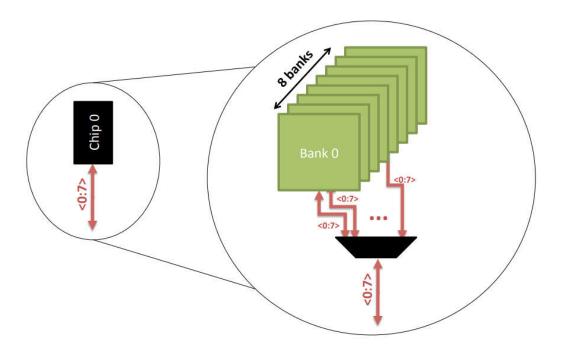
Memory ranks



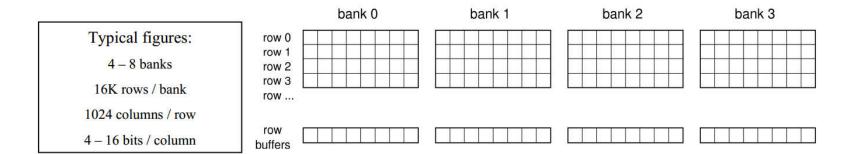
- Memory chips per rank
 - A rank has 64 bits data bus width
 - Number of circuits per rank



- Memory banks per chip
 - Support for interleaving

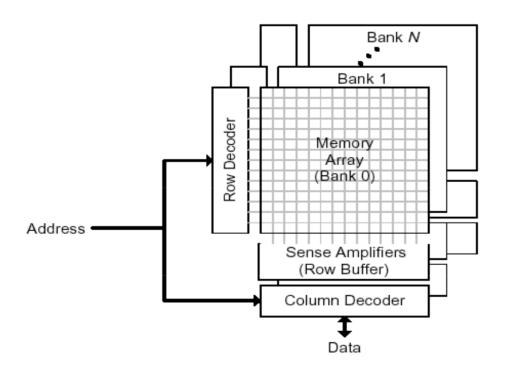


Memory banks



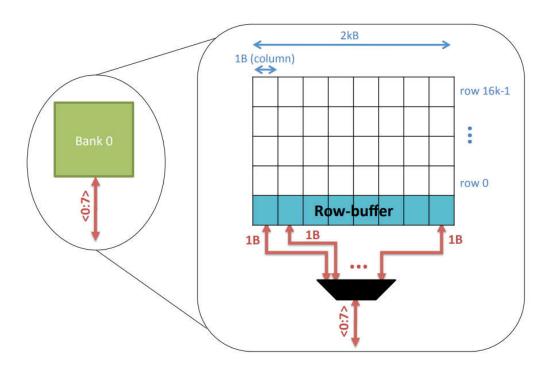
DRAM memories

• 3-D DRAM



S. Rixner et al. Memory Access Scheduling. ISCA 2000.

Memory arrays per bank



Main memory

DDR3

- 2GB 2Rx8 PC3-8500U
- 2GB 1Rx4 PC3-10600R

Standard name	Memory clock (MHz)	Cycle time (ns)	I/O bus clock (MHz)	Data rate (MT/s)	Module name	Peak transfer rate (MB/s)	Timings (CL-tRCD-tRP)	CAS latency (ns)
DDR3-800D DDR3-800E	100	10	400	800	PC3-6400	6400	5-5-5 6-6-6	12½ 15
DDR3-1066E DDR3-1066F DDR3-1066G	1331/2	71/2	5331/3	1066%	PC3-8500	85331/4	6-6-6 7-7-7 8-8-8	11½ 13½ 15
DDR3-1333F* DDR3-1333G DDR3-1333H DDR3-1333J*	166%	6	666%	13331⁄₃	PC3-10600	10666%	7-7-7 8-8-8 9-9-9 10-10-10	10½ 12 13½ 15
DDR3-1600G* DDR3-1600H DDR3-1600J DDR3-1600K	200	5	800	1600	PC3-12800	12800	8-8-8 9-9-9 10-10-10 11-11-11	10 11½ 12½ 13¾
DDR3-1866J* DDR3-1866K DDR3-1866L DDR3-1866M*	2331/4	42/7	9331⁄2	1866%	PC3-14900	149331/3	10-10-10 11-11-11 12-12-12 13-13-13	10 ⁵ / ₇ 11 ¹¹ / ₁₄ 12 ⁶ / ₇ 13 ¹³ / ₁₄
DDR3-2133K* DDR3-2133L DDR3-2133M DDR3-2133N*	266%	3¾	1066%	21331⁄3	PC3-17000	17066%	11-11-11 12-12-12 13-13-13 14-14-14	10 ⁵ / ₁₆ 11 ¹ / ₄ 12 ³ / ₁₆ 13 ¹ / ₈

^{*} optiona

CL - Clock cycles between sending a column address to the memory and the beginning of the data in response

tRCD - Clock cycles between row activate and reads/writes

Main memory

DDR4

Standard name	Memory clock (MHz)	I/O bus clock (MHz)	Data rate (MT/s)	Module name	Peak trans- fer rate (MB/s)	Timings CL-tRCD-tRP	CAS latency (ns)
DDR4-1600J*						10-10-10	12.5
DDR4-1600K	200	800	1600	PC4-12800	12800	11-11-11	13.75
DDR4-1600L						12-12-12	15
DDR4-1866L*						12-12-12	12.857
DDR4-1866M	233.33	933.33	1866.67	PC4-14900	14933.33	13-13-13	13.929
DDR4-1866N						14-14-14	15
DDR4-2133N*						14-14-14	13.125
DDR4-2133P	266.67	1066.67	2133.33	PC4-17000	17066.67	15-15-15	14.063
DDR4-2133R						16-16-16	15
DDR4-2400P*						15-15-15	12.5
DDR4-2400R	200	1200	2400	PC4-19200	19200	16-16-16	13.32
DDR4-2400T	300	1200	2400	PC4-19200	19200	17-17-17	14.16
DDR4-2400U						18-18-18	15
DDR4-2666T						17-17-17	12.75
DDR4-2666U	325	1333	2666	PC4-21333	21333	18-18-18	13.50
DDR4-2666V						19-19-19	14.25
DDR4-2666W						20-20-20	15
DDR4-2933V						19-19-19	12.96
DDR4-2933W	366.6	1466.5	2933	PC4-23466	23466	20-20-20	13.64
DDR4-2933Y						21-21-21	14.32
DDR4-2933AA						22-22-22	15
DDR4-3200W						20-20-20	12.50
DDR4-3200AA	400	1600	3200	PC4-25600	25600	22-22-22	13.75
DDR4-3200AC						24-24-24	15

Main memory

- Chip size = 2^14*2^10*8banks*8bits = 2^4*2^20*8*8bits = 128 MB
- Rank size = 128 MB * 8 chips/rank = 1024 MB = 1GB
- DDR size = 2 ranks * 1GB = 2GB

Raw Card Version	DIMM Capacity	DIMM Organization	SDRAM Density	SDRAM Organization	Number of SDRAMs	Number of Physical Ranks	Number of Banks in SDRAM	Number of Address Bits Row/Column
A	512MB	64 Meg x 64	512 Megabit	64 Meg x 8	8	1	8	13/10
	1GB	128 Meg x 64	1 Gigabit	128 Meg x 8	8	1	8	14/10
	2GB	256 Meg x 64	2 Gigabit	256 Meg x 8	8	1	8	15/10
	4GB	512 Meg x 64	4 Gigabit	512 Meg x 8	8	1	8	16/10
	8GB	1 Gig x 64	8 Gigabit	1 Gig x 8	8	1	8	16/11
В	1GB	128 Meg x 64	512 Megabit	64 Meg x 8	16	2	8	13/10
	2GB	256 Meg x 64	1 Gigabit	128 Meg x 8	16	2	8	14/10
	4GB	512 Meg x 64	2 Gigabit	256 Meg x 8	16	2	8	15/10
	8GB	1 Gig x 64	4 Gigabit	512 Meg x 8	16	2	8	16/10
	16GB	2 Gig x 64	8 Gigabit	1 Gig x 8	16	2	8	16/11
C ₁	256MB	32 Meg x 64	512 Megabit	32 Meg x 16	4	1	8	12/10
	512MB	64 Meg x 64	1 Gigabit	64 Meg x 16	4	1	8	13/10
	1GB	128 Meg x 64	2 Gigabit	128 Meg x 16	4	1	8	14/10
	2GB	256 Meg x 64	4 Gigabit	256 Meg x 16	4	1	8	15/10