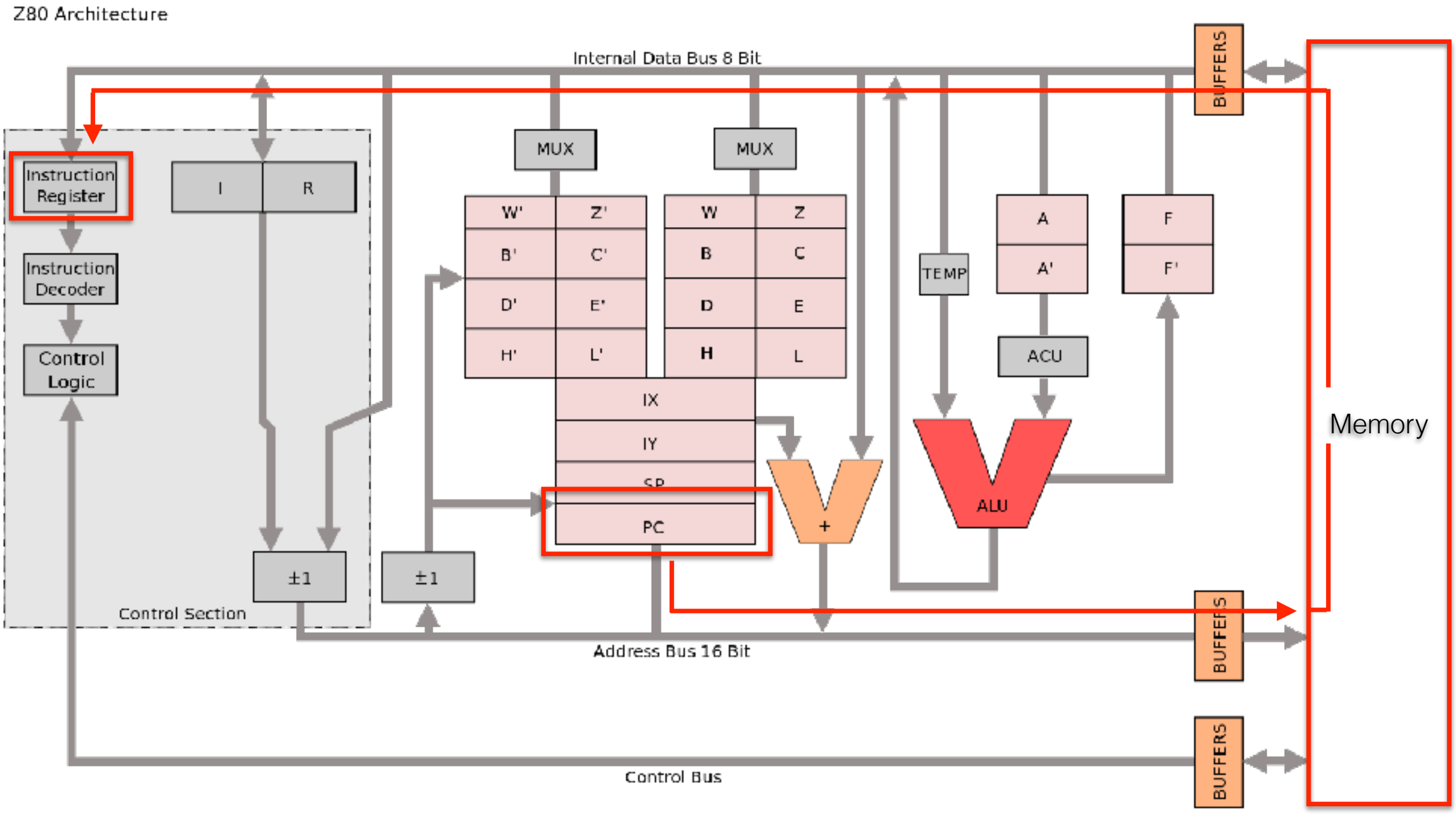
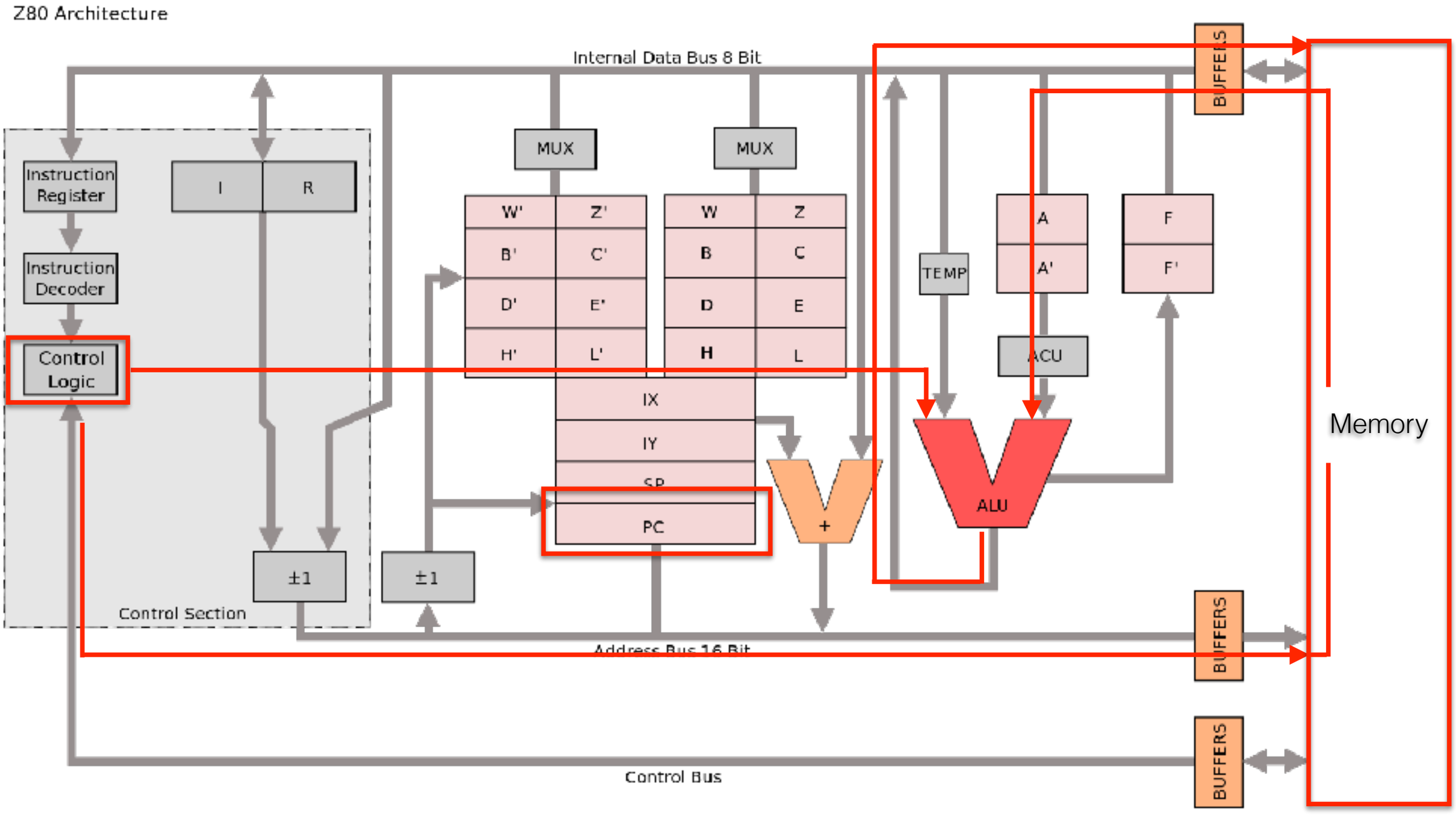
Memory interface

Donghwa Shin

Instruction in bus and memory

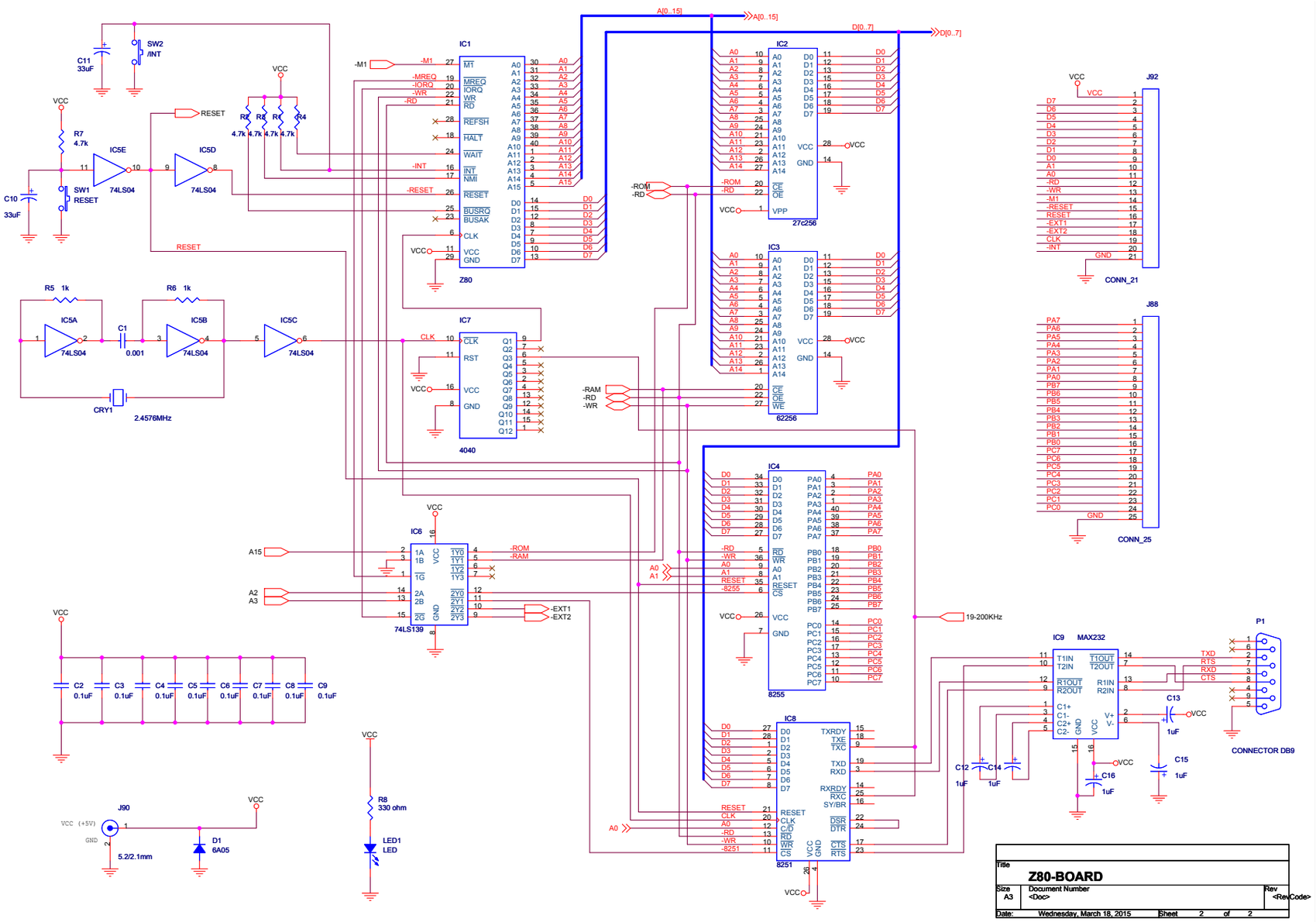


Instruction in bus and memory



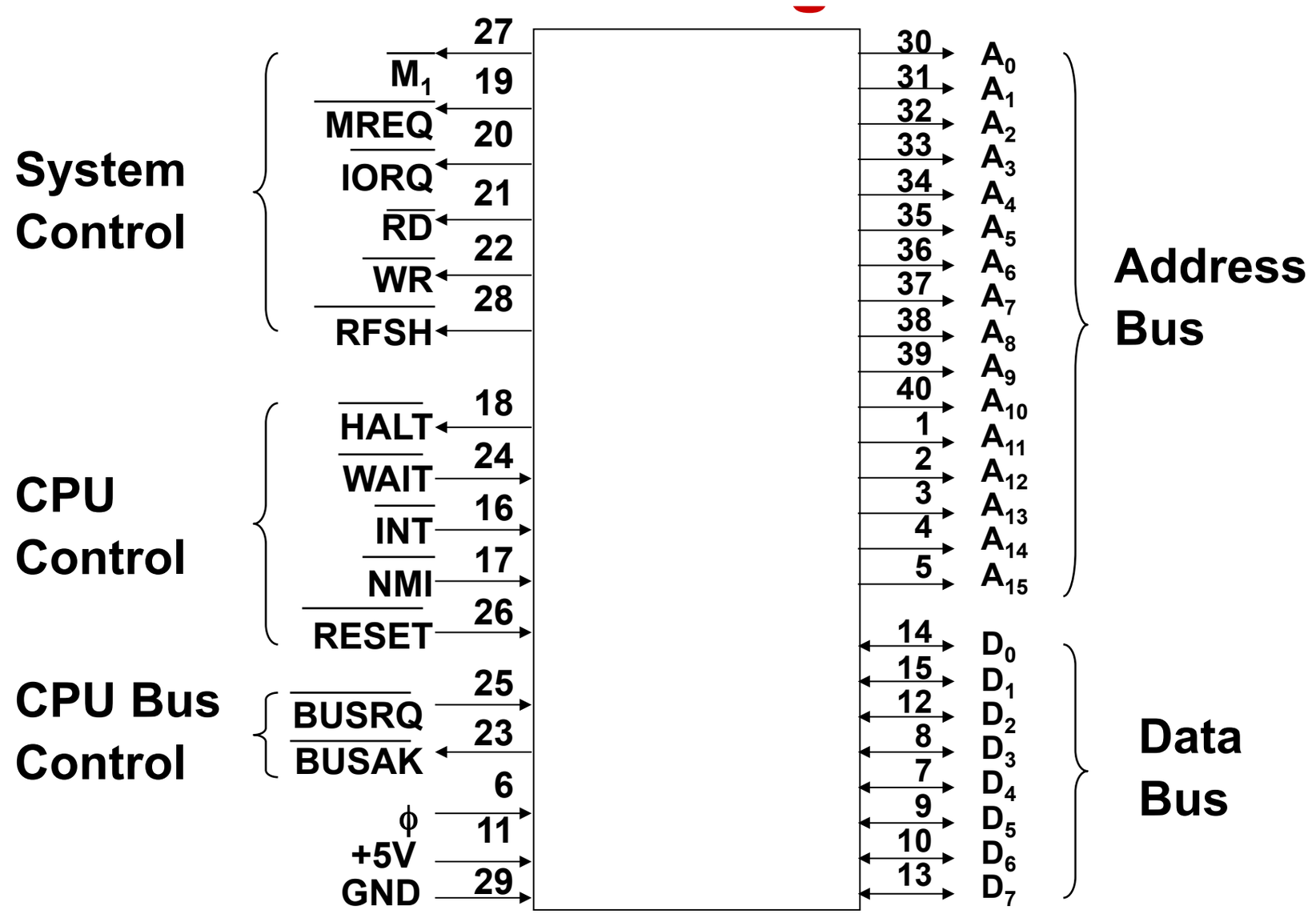
Processor and memory

- Memory devices in Lab board
- SRAM
- EEPROM



Processor and memory

- Memory interface signals in microprocessor
- Address bus
- Data bus
- Control signals

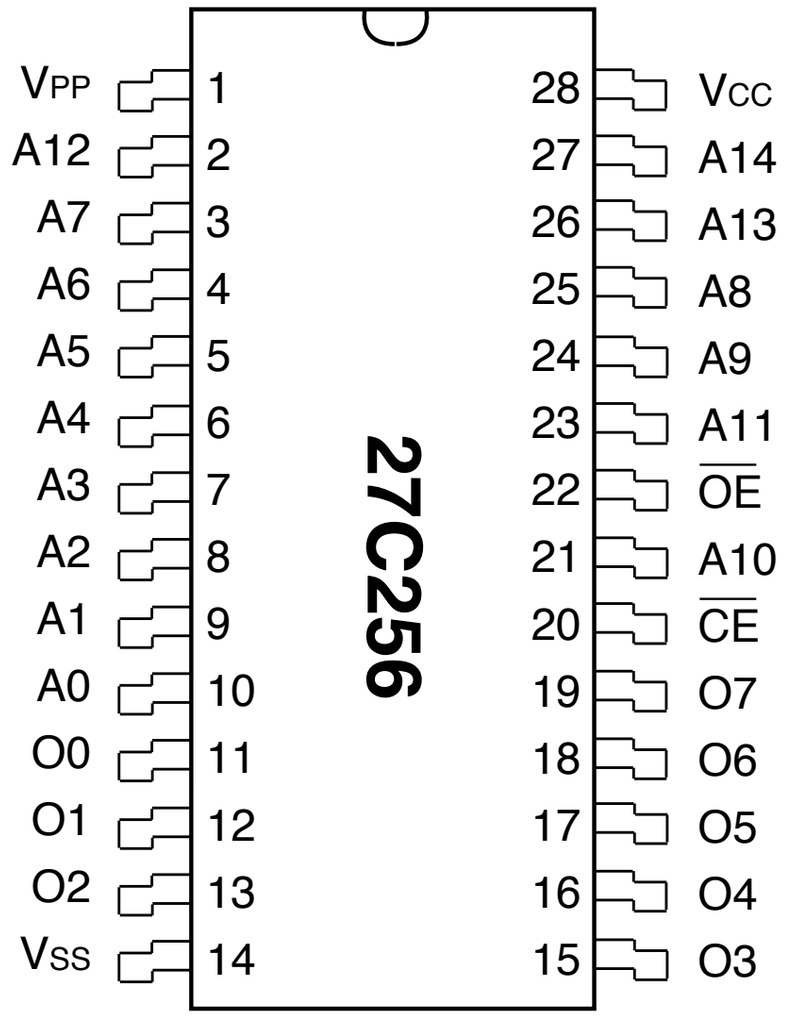


Memory Devices

- Read only memory (Non-volatile)
    - ROM: Programmed at factory
    - PROM and EPROM: Programmable and Erasable ROM
    - Flash memory
- Random access memory (Volatile): SRAM, DRAM
- Emerging memory technologies
    - Phase-change memory (PCM) , STT-RAM, MRAM

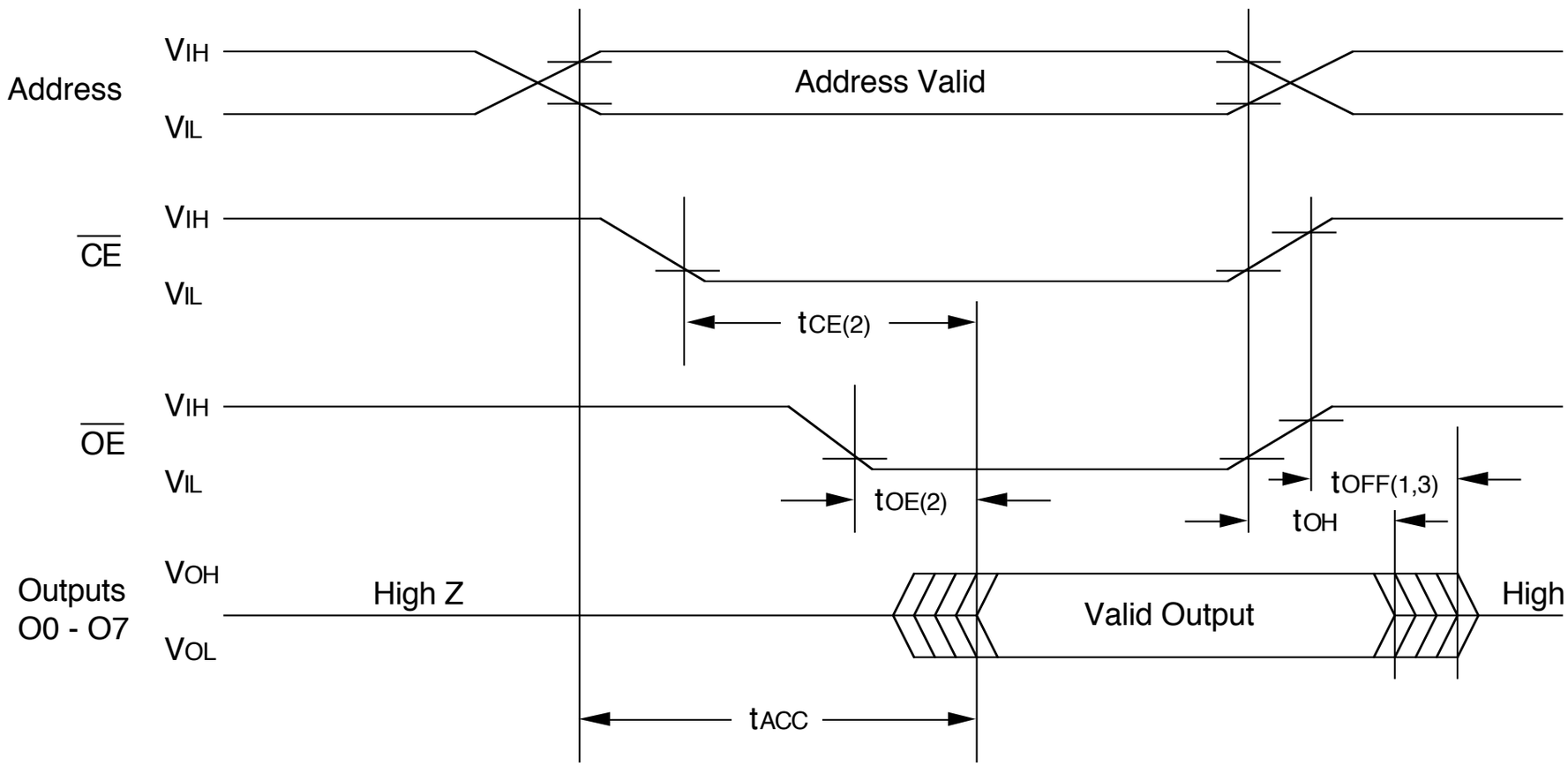
Non-volatile memory: EPROM

- Erasable programmable read only memory (EPROM)
  - Non-volatile: storage for code
  - Array of floating-gate transistors individually by higher voltages
  - Once programmed, an EPROM can be erased by exposing it to strong ultraviolet light source
  - 90 ns access time



Non-volatile memory: EPROM

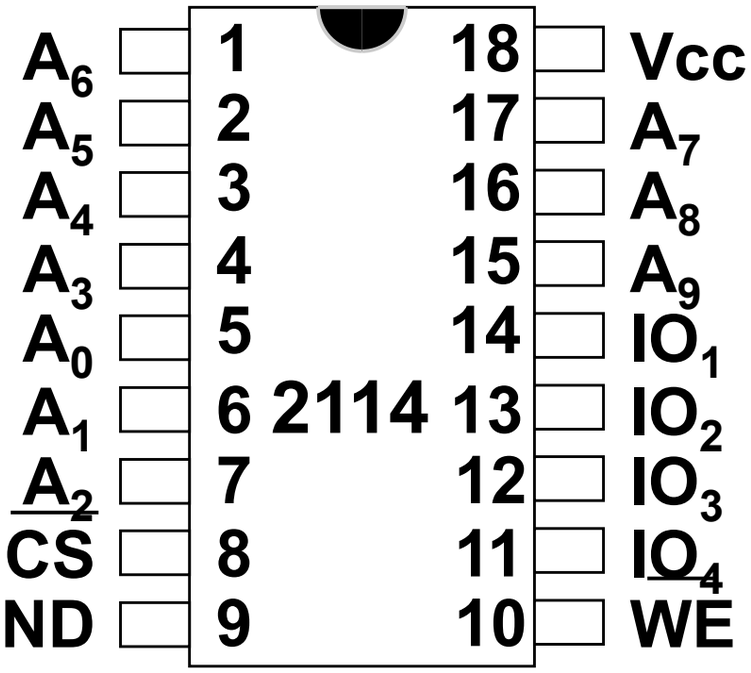
- Read operation
  - Set address address bus
  - Chip enable
  - Output enable
  - Output on data bus
- Can be regarded as the simplest way of memory access



Notes: (1) t_{OFF} is specified for $\overline{OE}$ or $\overline{CE}$, whichever occurs first
(2) $\overline{OE}$ may be delayed up to $t_{CE} - t_{OE}$ after the falling edge of $\overline{CE}$ without impact on t_{CE}
(3) This parameter is sampled and is not 100% tested.

Volatile memory: SRAM

- A typical SRAM with
  - 4 x 1Kbyte memory arrays
    - 4 bit data x 10 bit address
  - Chip select (=Chip enable) and write enable for control signals
- Q) What is the relation between the memory size and data/address bus width?

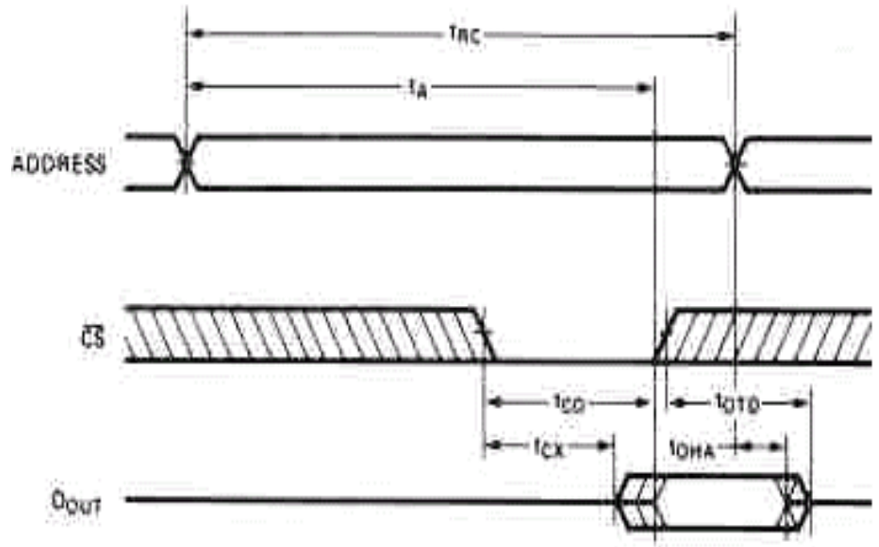


Pin Names

$A_0 - A_9$	Address Inputs
$\overline{CS}$	Chip Select
$\overline{WE}$	Write Enable
$I/O_1 - I/O_4$	Data Input/Output

Volatile memory: SRAM

READ CYCLE ③



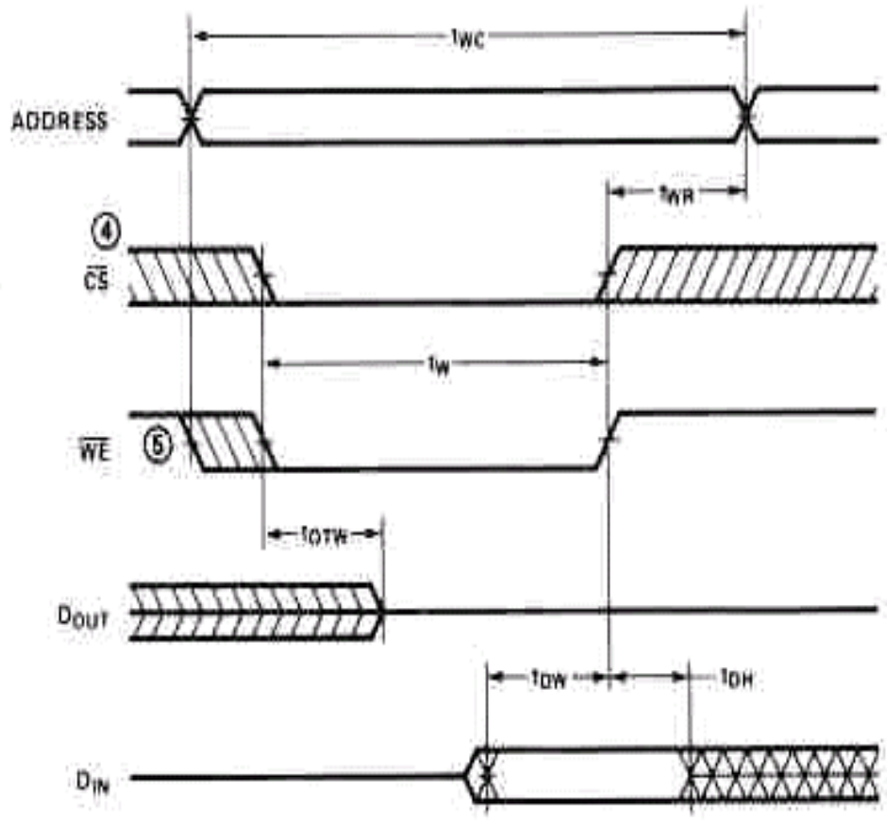
NOTES:

③ $\overline{WE}$ is high for a Read Cycle.

④ If the $\overline{CS}$ low transition occurs simultaneously with the $\overline{WE}$ low transition, the output buffers remain in a high impedance state.

⑤ $\overline{WE}$ must be high during all address transitions.

WRITE CYCLE



- Read: Address -> Chip select -> data output
- Write: Address -> Chip select & Write enable -> data in (rising edge)

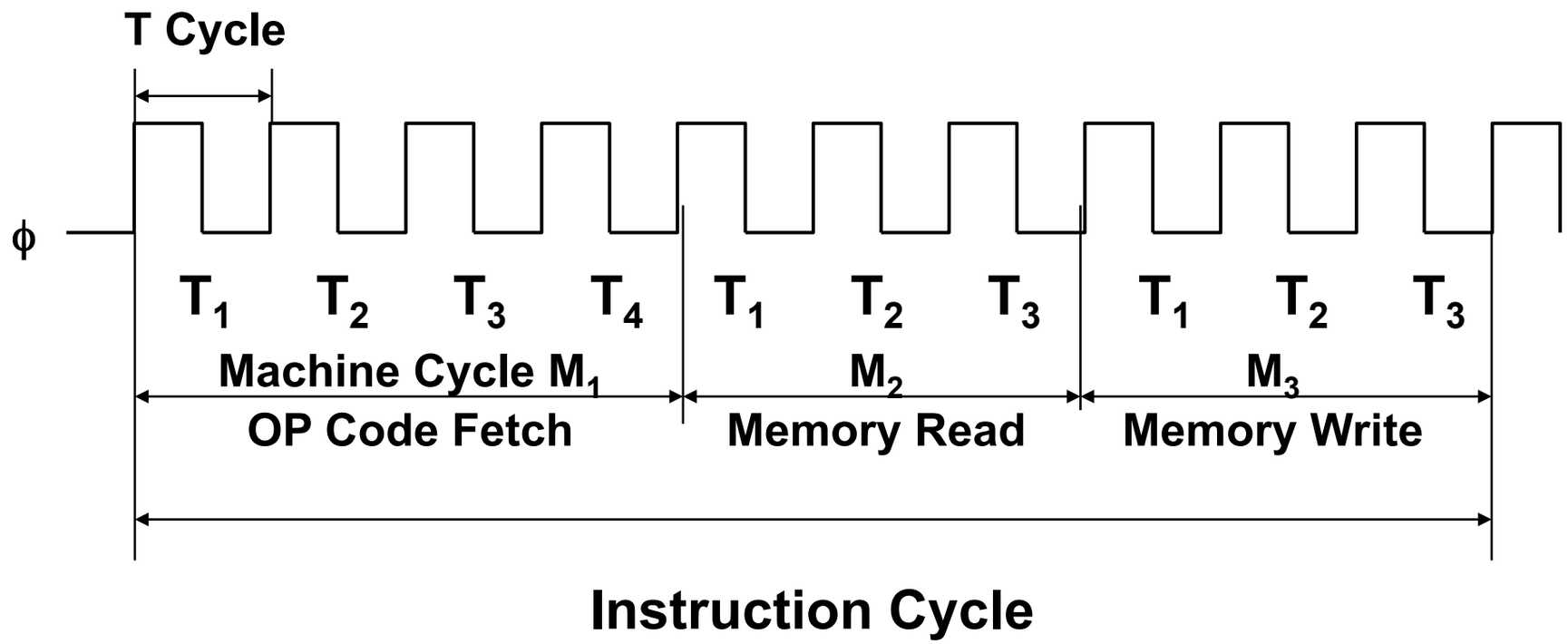
Z80 instruction cycle

- M1: Opcode fetch - M2: memory read - M3: memory write

- Why? Let's think about the case of moving data from memory to memory

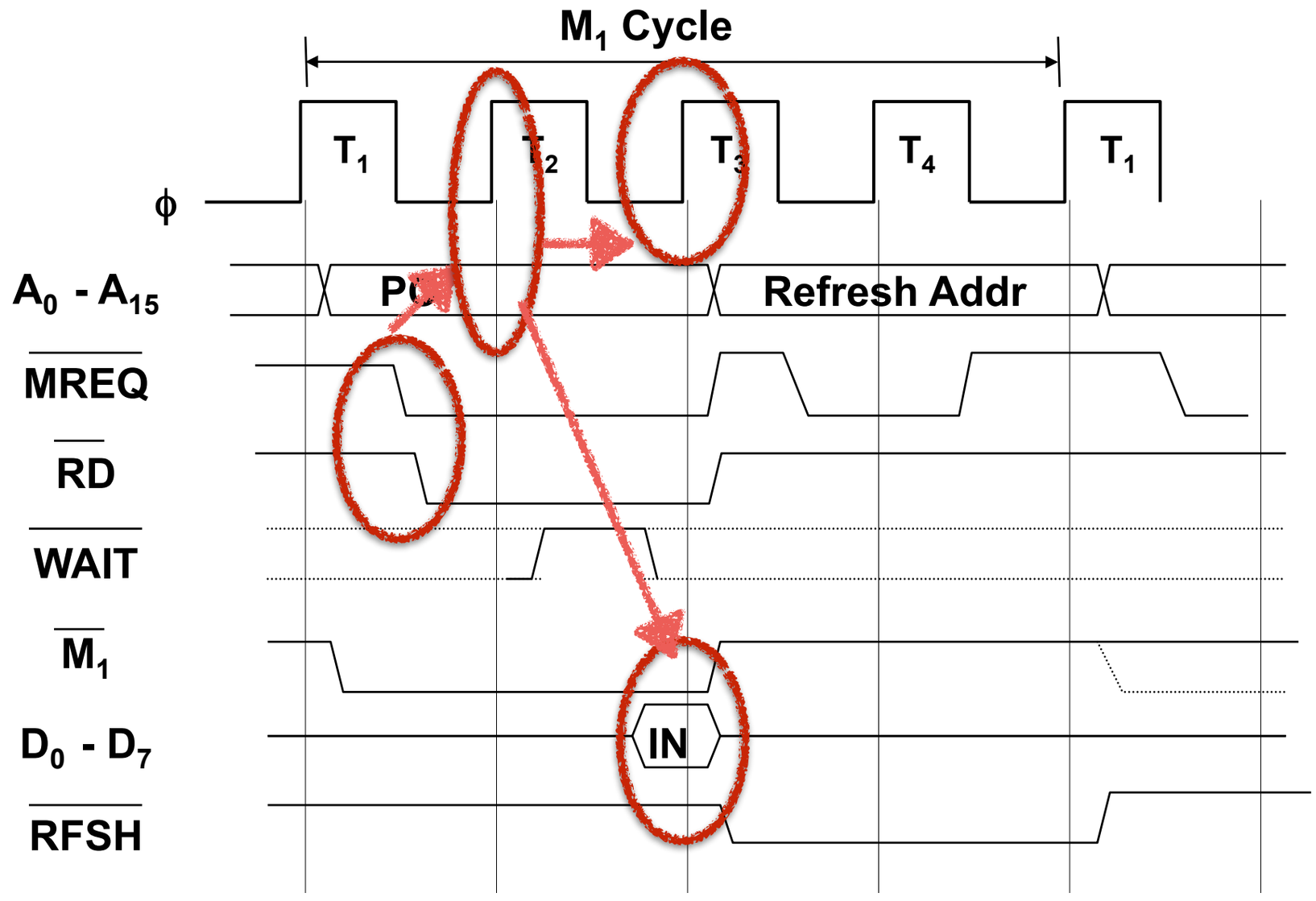
- LD r (HL)
  ->M1 + M2

- LD (HL) r
  ->M1 + M3



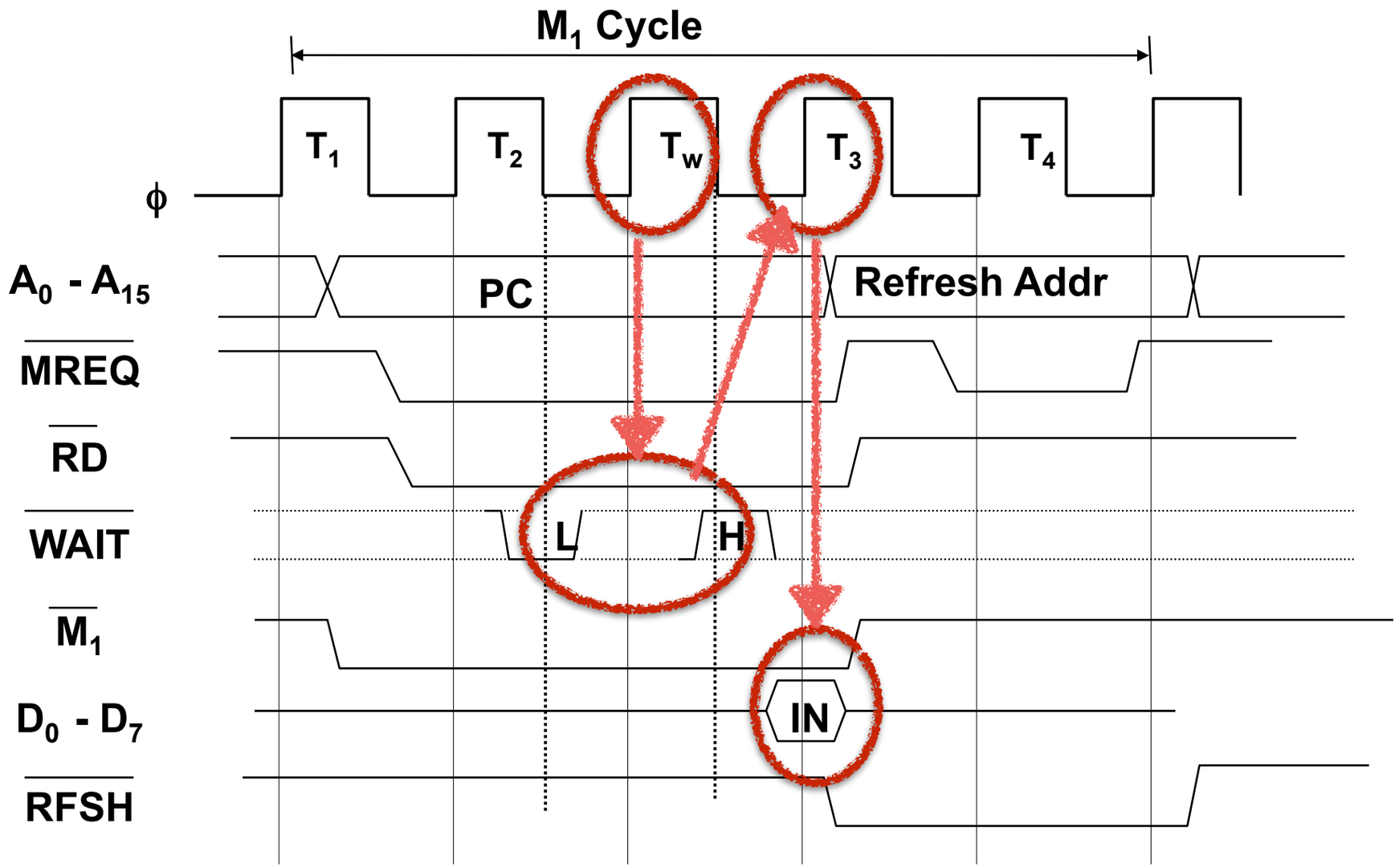
Z80 instruction cycle and memory access

- M1: Opcode fetch
  - read code memory addressed by PC (program counter)
  - Address, control signals and data are synchronized by the clock signal
  - Data will be the next clock rising edge after the address is given



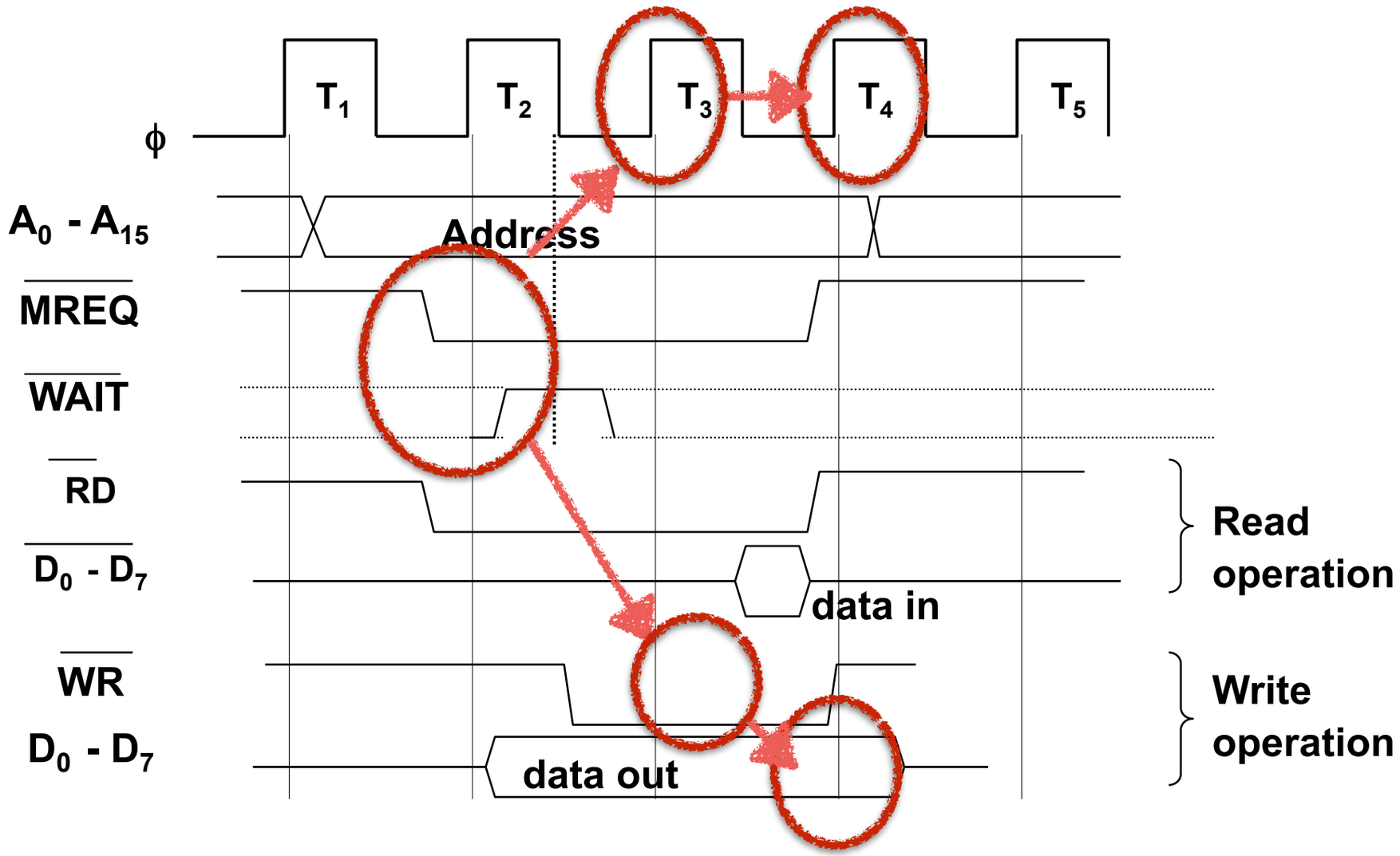
Z80 instruction cycle and memory access

- Delayed memory access
  - If the data cannot be delivered after the address,
  - Then, wait signal will hold the access cycle



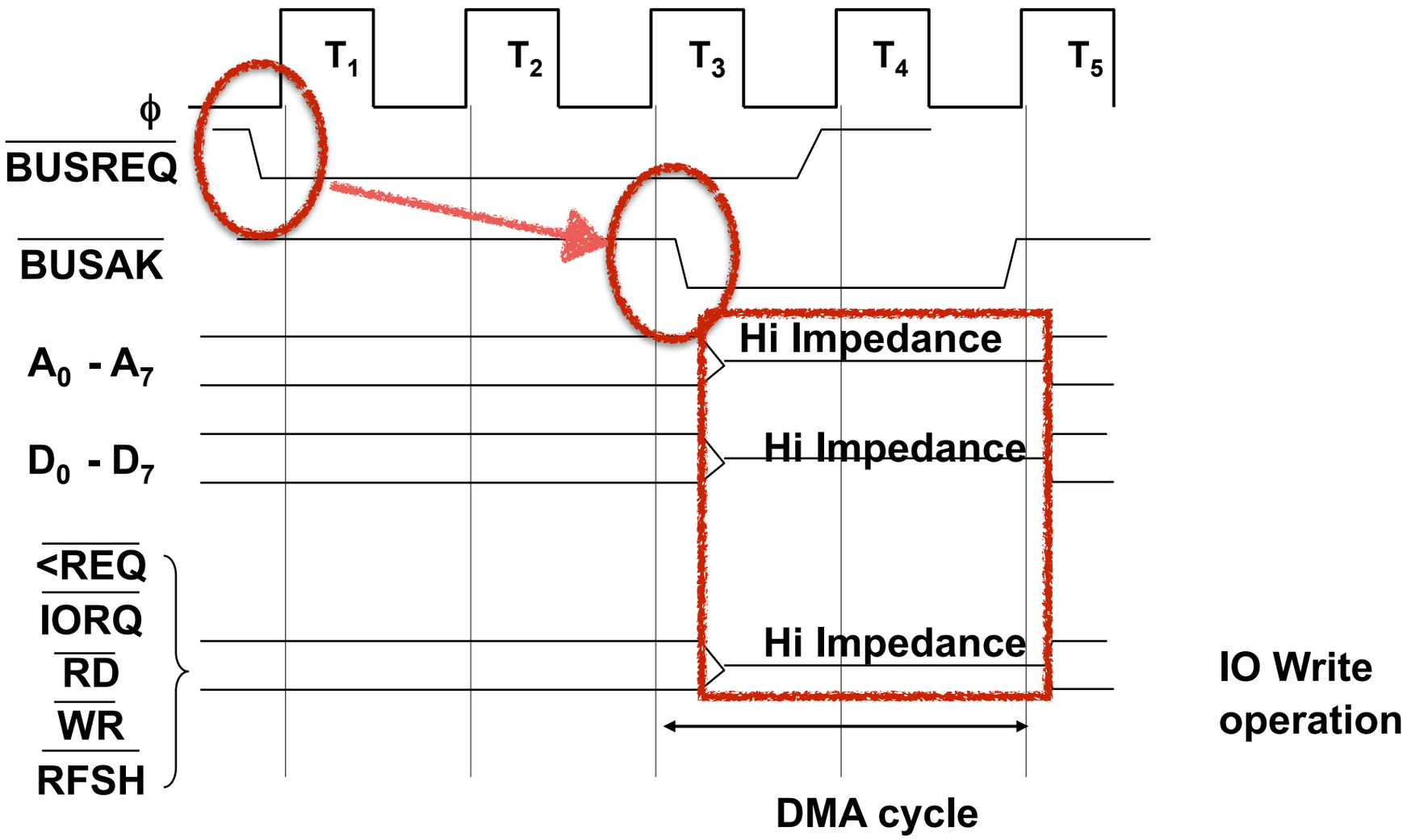
Z80 instruction cycle and memory access

- Write operation
  - Select the chip
  - Set address on the address bus
  - Set output data on the bus in the next clock rising edge



Bus arbitration

- Bust request from the other master device
- Response by the ACK signal, and then
- Change the bus master by setting the output signals to high-Z state

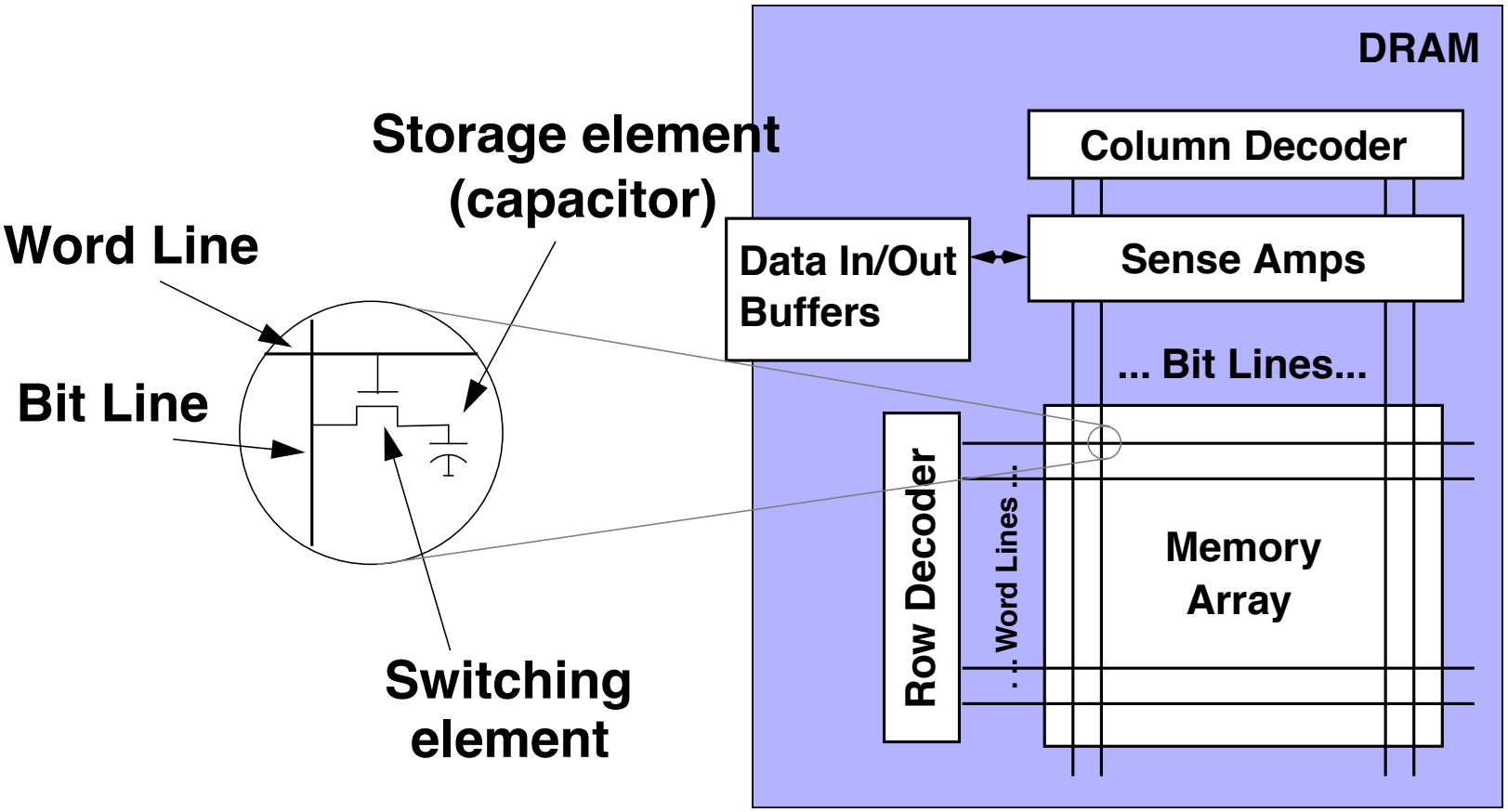


Different types of memory

	SRAM	DRAM	Flash (NAND)	HDD
Reciprocal density *(F^2)*	140	6-12	1-4	2/3
Energy per bit *(pJ)*	0.0005	0.005	0.00002 (a)	$5x10^3$- 10^4 (b)
Read time *(ns)*	0.1-0.3	10	100 000	$5\text{-}8x10^6$
Write time *(ns)*	0.1-0.3	10	100 000	$5\text{-}8x10^6$
Retention	as long as V applied	<<second	years	years
Endurance *(cycles)*	> 10^{16}	> 10^{16}	10^4	10^4 (c)

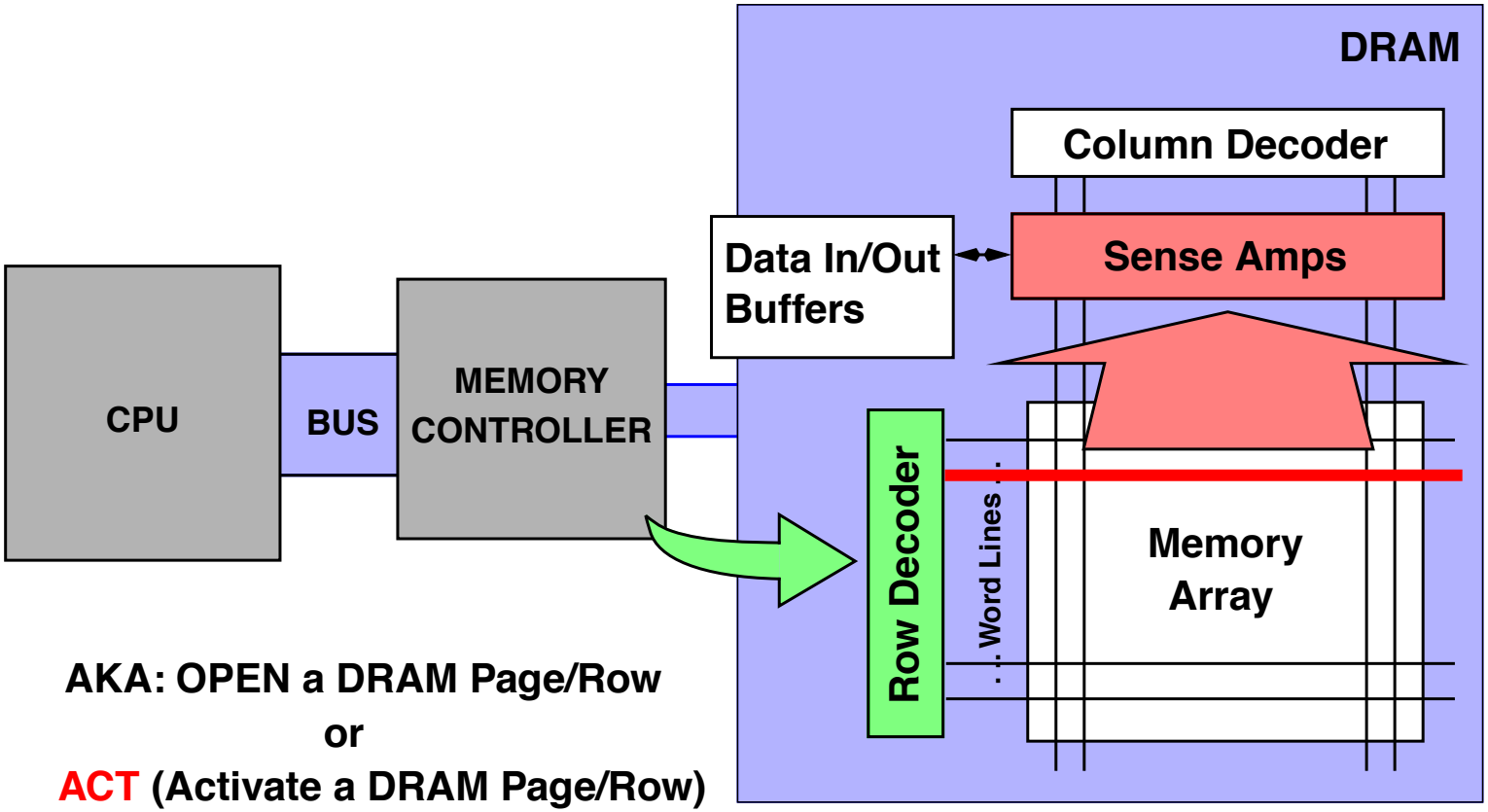
DRAM

- Principle and structure



DRAM access

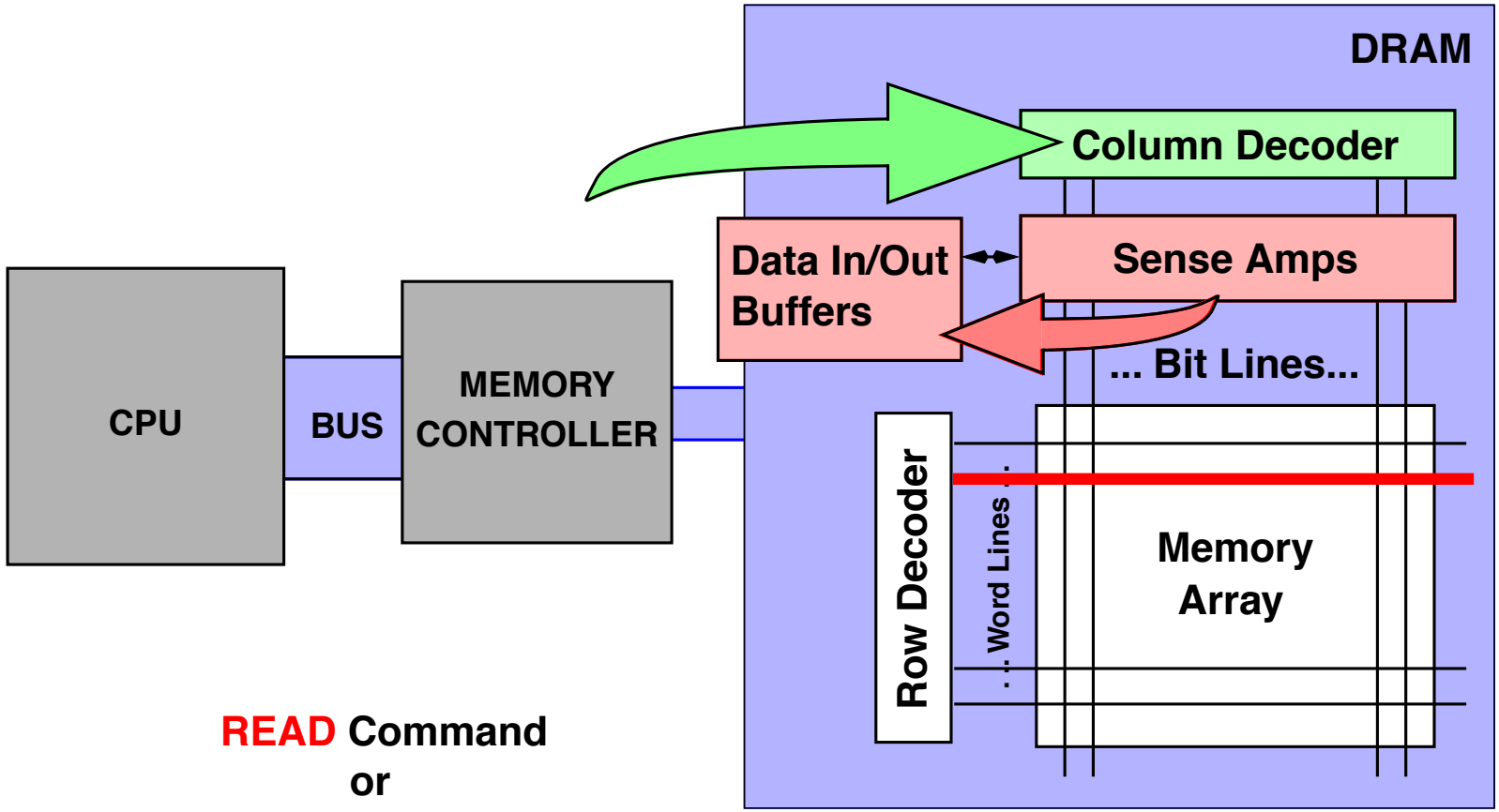
- Bus transaction - Precharge - Row access



AKA: OPEN a DRAM Page/Row
or
ACT (Activate a DRAM Page/Row)
or
RAS (Row Address Strobe)

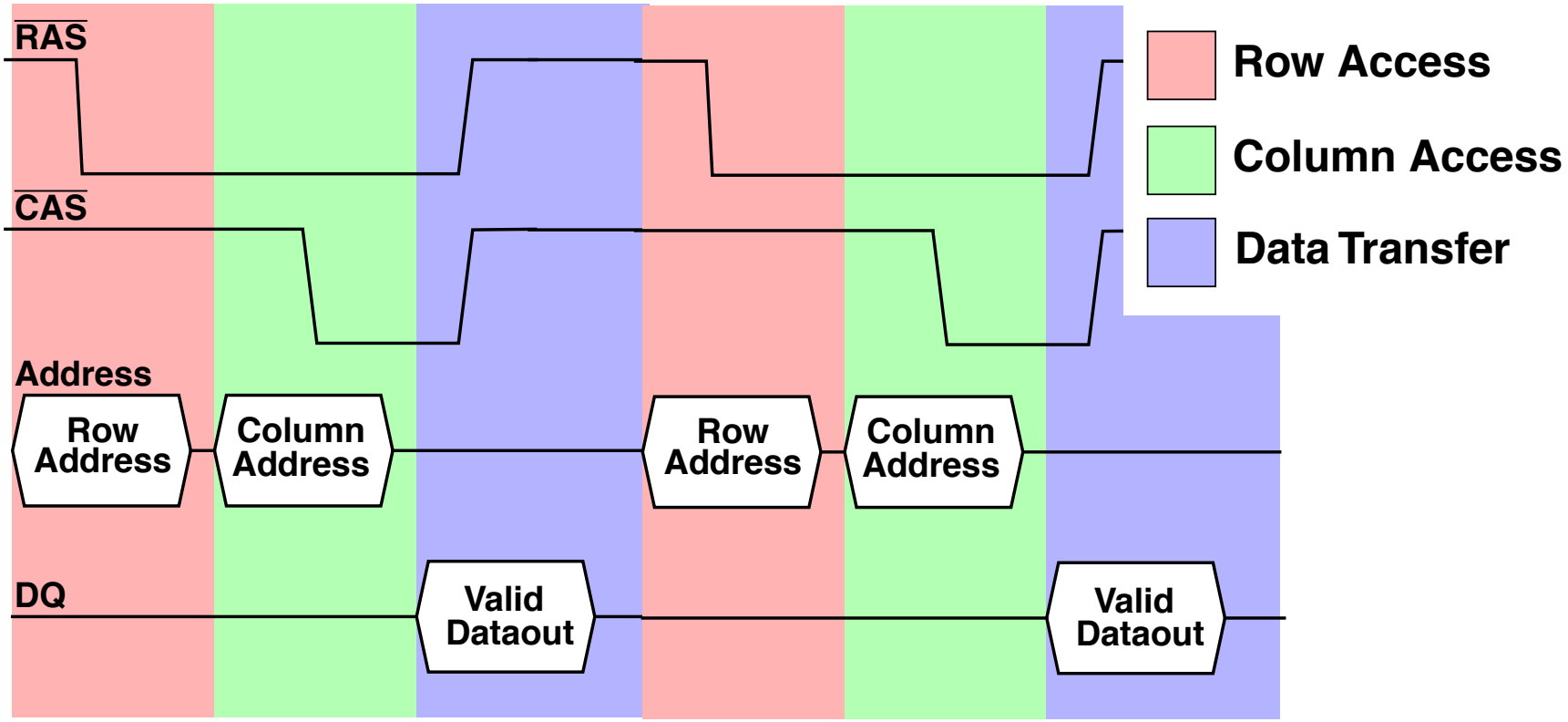
DRAM access

- Column access - Bus tramsacation



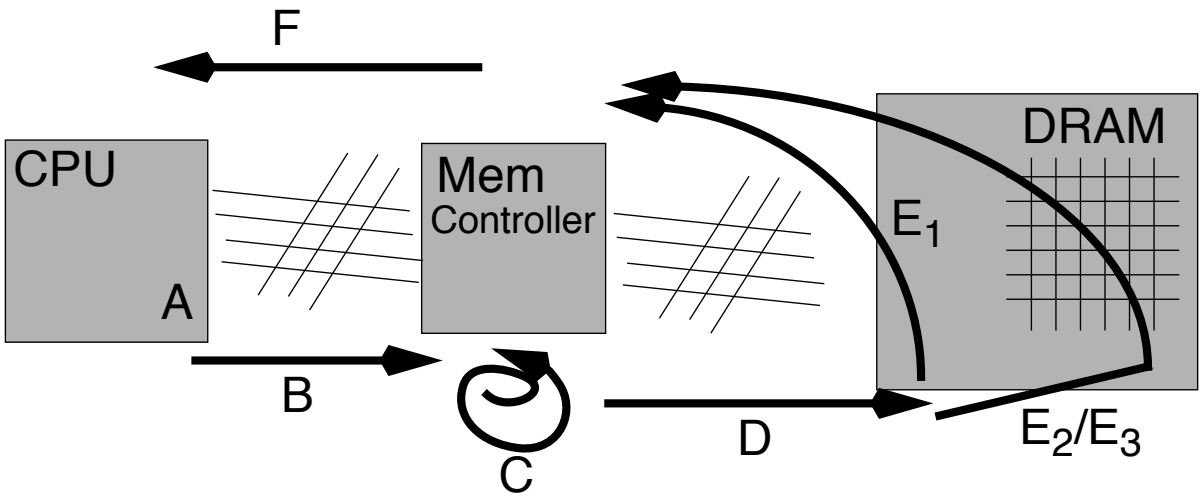
DRAM access

- RAS-CAS timing



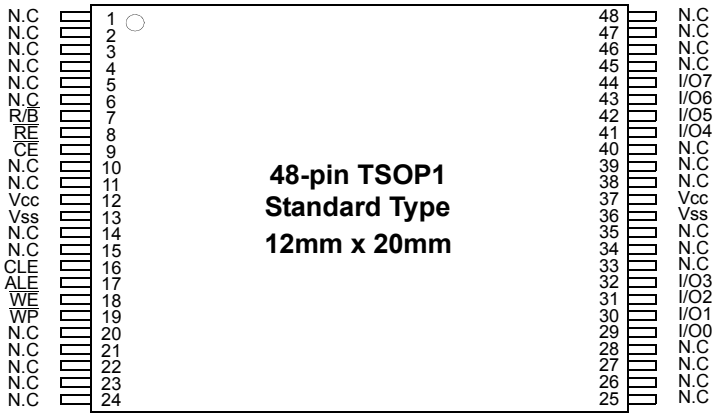
DRAM latency

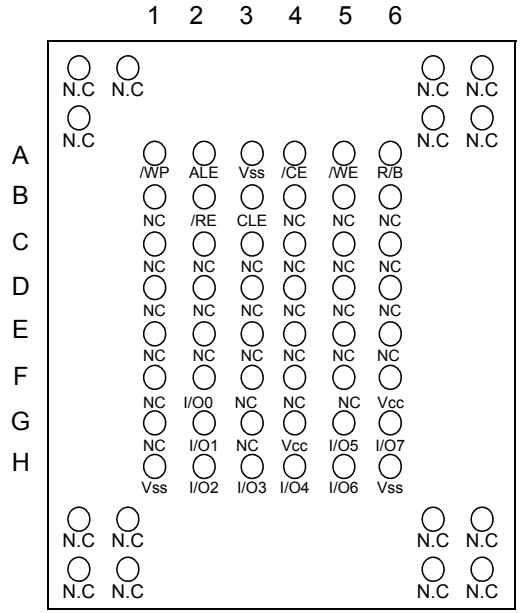
- A: Transaction request may be delayed in Queue B: Transaction request sent to Memory Controller C: Transaction converted to Command Sequences
- D: Command/s Sent to DRAM E1: Requires only a CAS or
- E2: Requires RAS + CAS or
- E3: Requires PRE + RAS + CAS
- F: Transaction sent back to CPU
- “DRAM Latency” = A + B + C + D + E + F



Flash

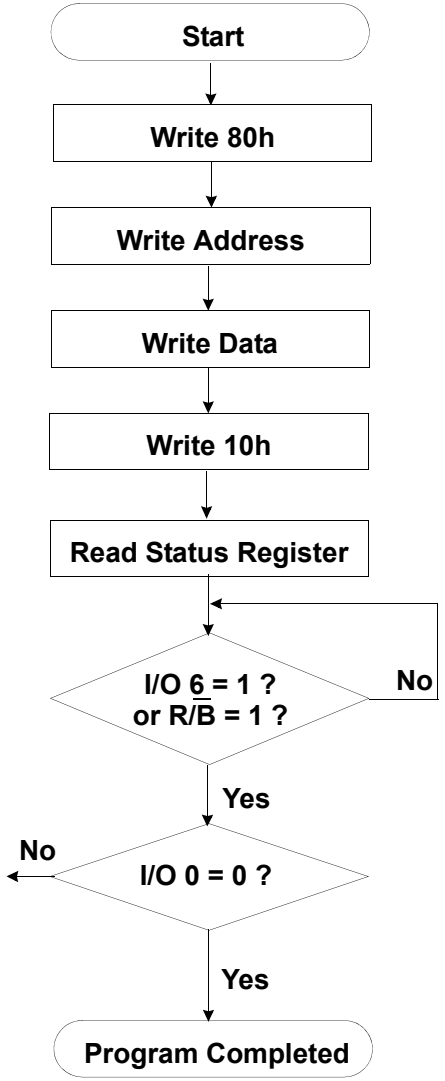
- K9F1G08U0D: 1Gbit NAND flash from Samsung
- Organization
  - Memory Cell Array : (128M + 4M) x 8bit
  - Page Size : (2K + 64)Byte
  - Erase block: (128K + 4K)Byte
- Performance
  - Random Read : 40μs(Max.)
  - Serial Access : 25ns(Min.)
  - Page Program time : 250μs(Typ.)
  - Block Erase Time : 2ms(Typ.)





Flash interface

- Erase and program -> Read
- Sequence of command to control register and command buffer
  - Send ignition command
  - Fill the data to the buffer
  - Send program start command
  - Check the program status flag
  - Finish and ready for the next command



Memory hierarchy

- Memory hierarchy: a concept used to discuss performance issues in computer architectural design, algorithm predictions, and lower level programming constructs involving locality of reference
- The memory hierarchy in computer storage separates each of its levels based on response time. Since response time, complexity, and capacity are related, the levels may also be distinguished by their performance and controlling technologies.
- Major storage levels:
    - Internal – Processor registers and cache
    - Main – the system RAM and controller cards
    - On-line mass storage – Secondary storage
    - Off-line bulk storage – Tertiary and Off-line storage

Memory hierarchy

- Designing for high performance requires considering the size and capabilities of each component
- In general, as part of a hierarchy of memories, each member is typically smaller and faster than the next highest member of the hierarchy
- To limit waiting by higher levels, a lower level will respond by filling a buffer and then signaling to activate the transfer
- In short, avoid to access the slow devices!
  - Shouting in datacenter (https://www.youtube.com/watch?v=tDacjrSCeq4)

Computer Memory Hierarchy