Topic 21:
Memory Technology

COS / ELE 375

Computer Architecture and Organization

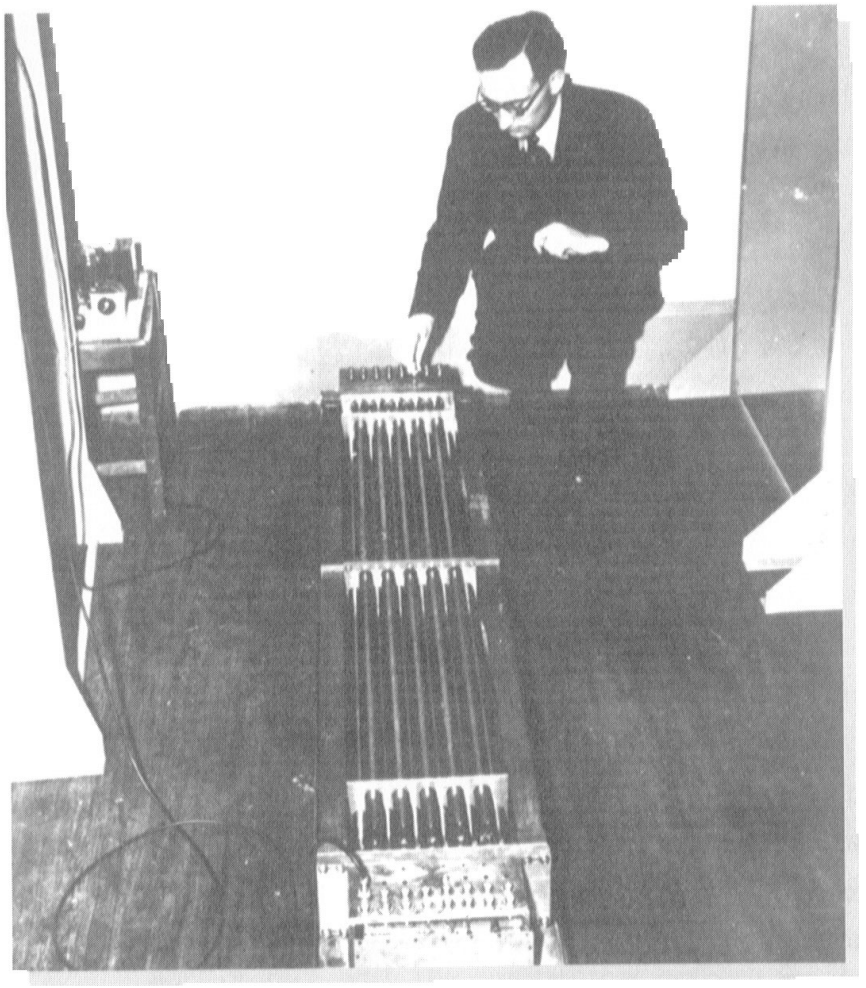
Princeton University
Fall 2015

Prof. David August

Old Stuff Revisited

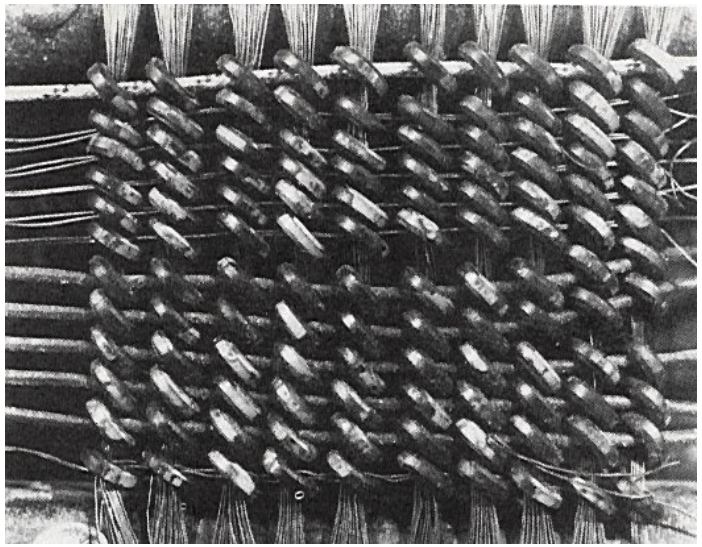
Mercury Delay Line Memory

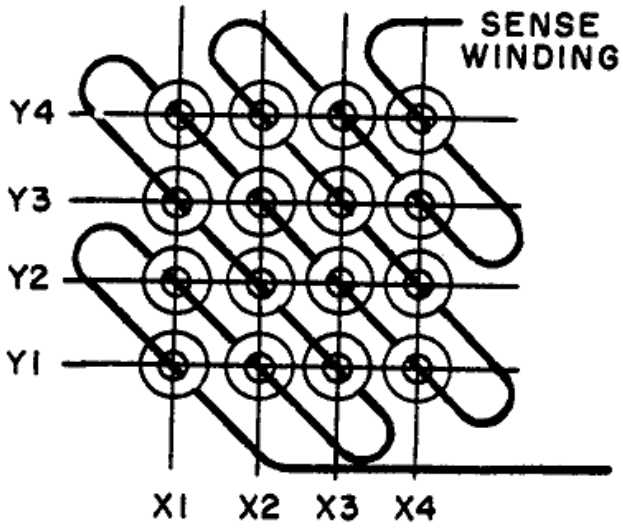
Maurice Wilkes, in 1947, with first mercury tank memories built for EDSAC.



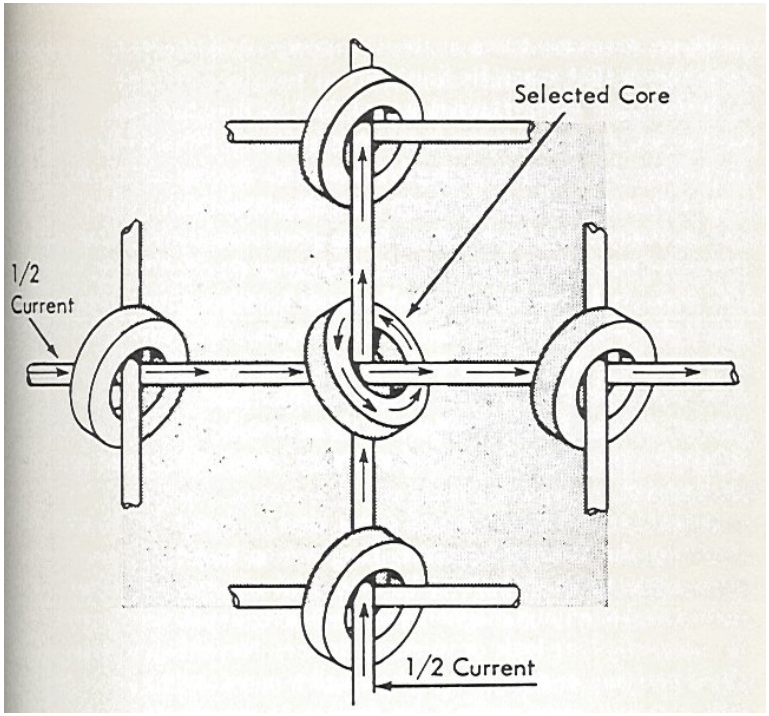
Core Memory

16K!

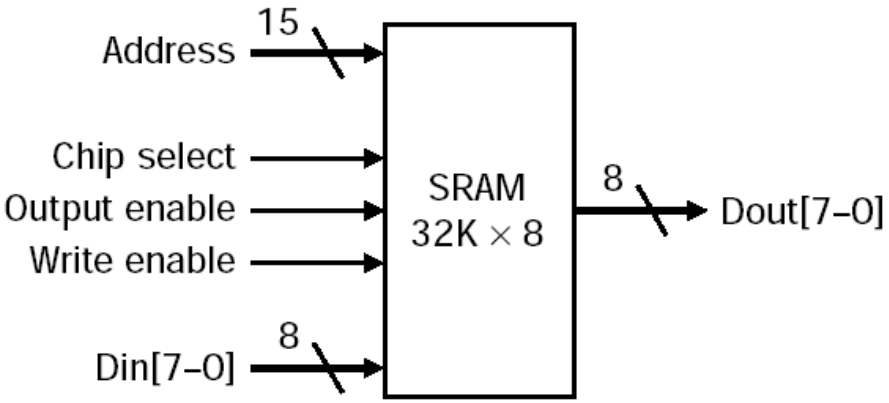


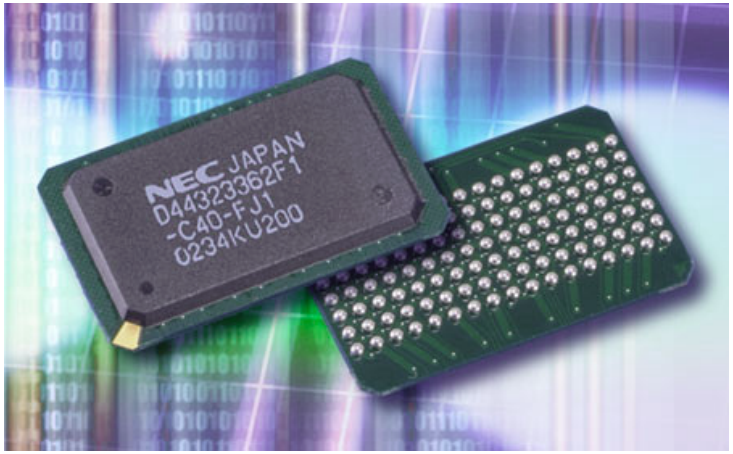


- Theory of operation
- Threaded by hand!
- The Lifesaver connection
- Refresh



Static RAM (SRAM)



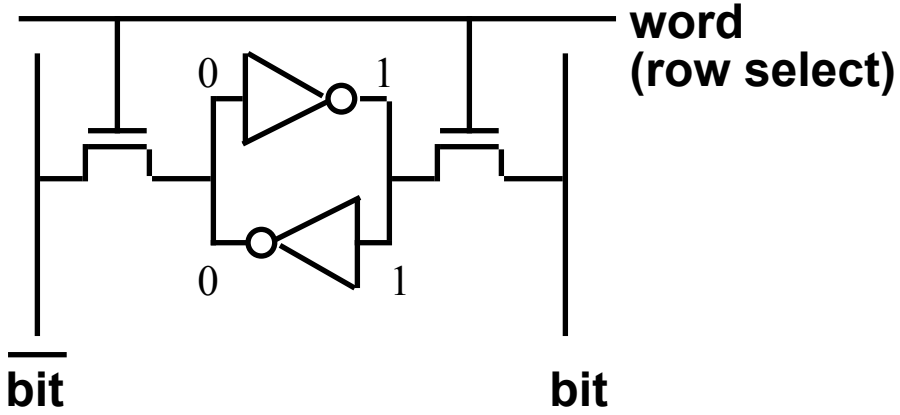


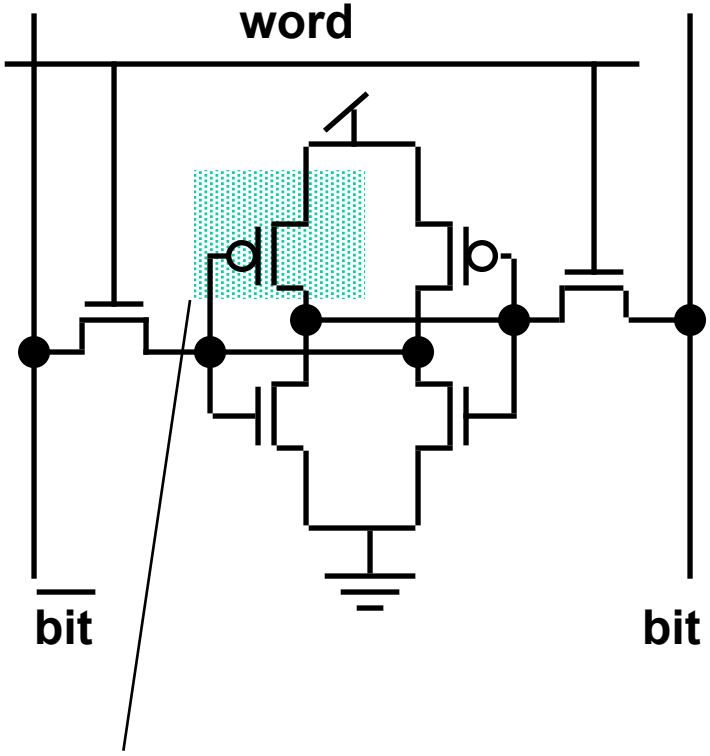
- SRAM - Fast, but not the most dense (better than core!)
- Chip select can be viewed as another address line
- Din and Dout are often combined to save pins
  - Need output enable (OE_L - enable low)
  - Need write enable (WE_L - enable low)
- Don't assert both write enable and output enable
  - Result is unknown.
  - This is bad.
  - Don't do it!!!

The Transistor Makes It Possible!

Static RAM (SRAM) Cell

6-Transistor SRAM Cell





Replaced with pullup resistor to save area

Write:

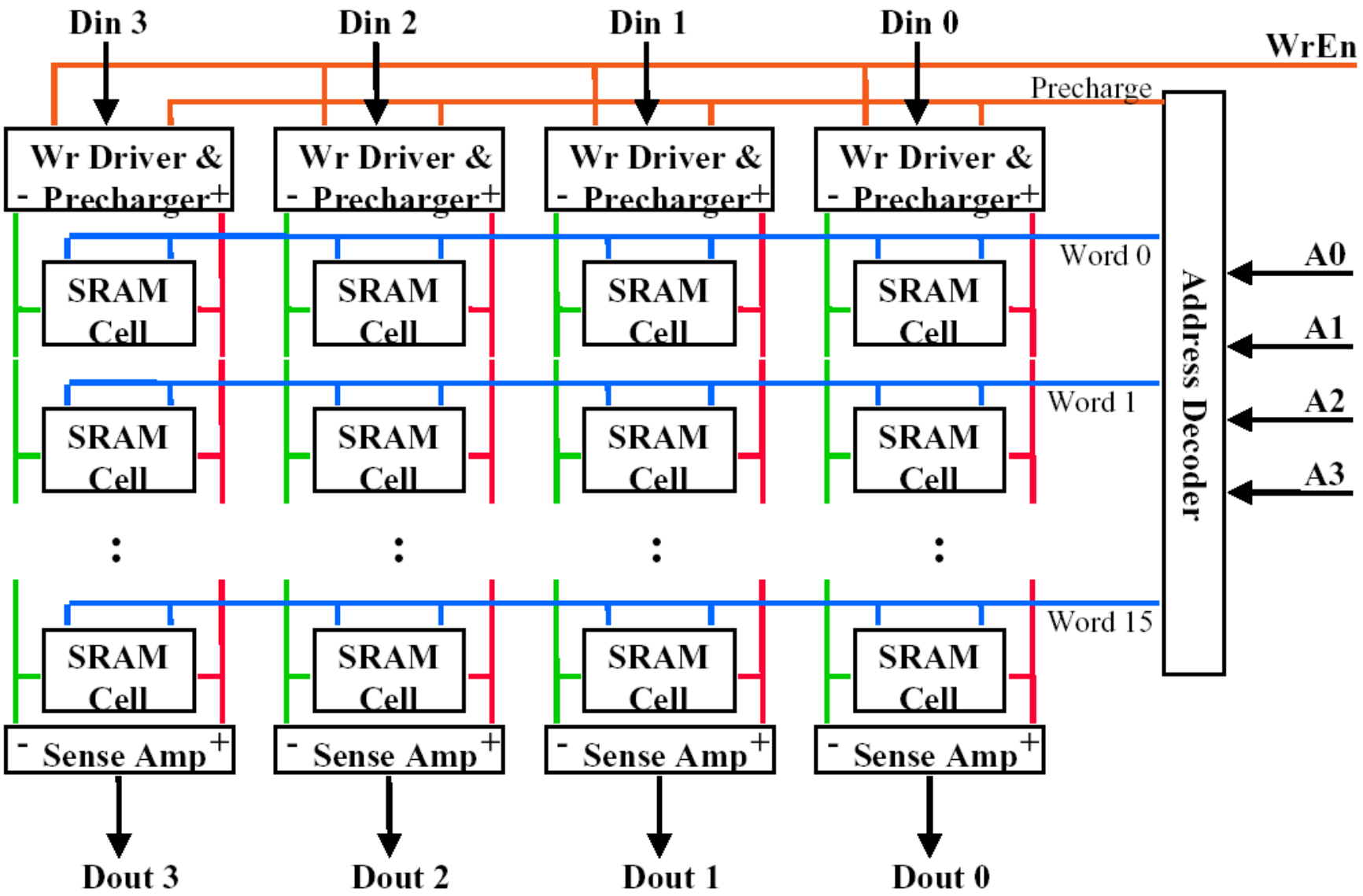
1. Drive bit lines (bit=1, bitbar=0)
2. Select row

Read:

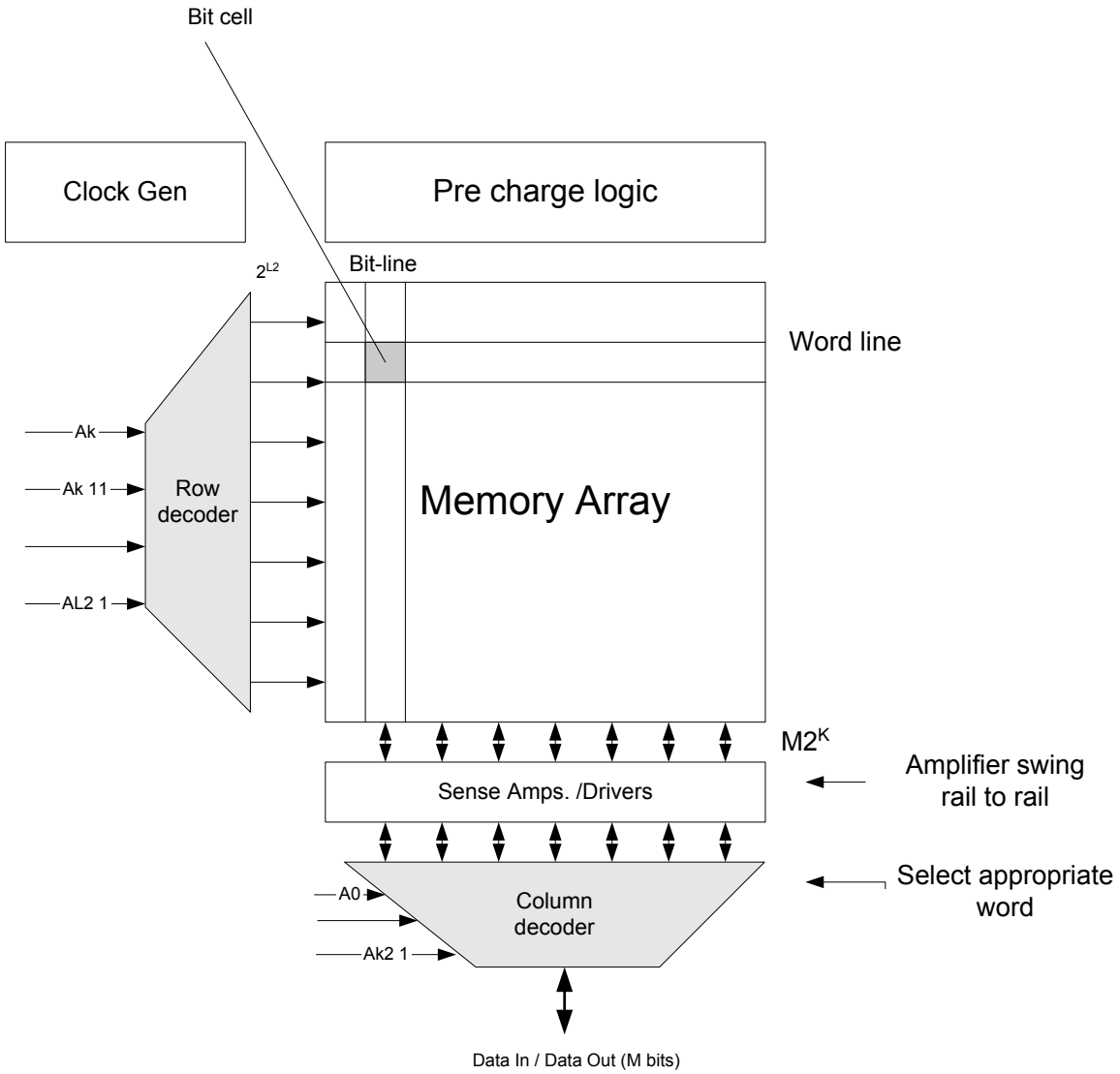
1. Precharge bit and bitbar to Vdd
2. Select row
3. Cell pulls one line low
4. Sense amp on column detects difference between bit and bitbar

Why is it Static?

Typical SRAM Organization

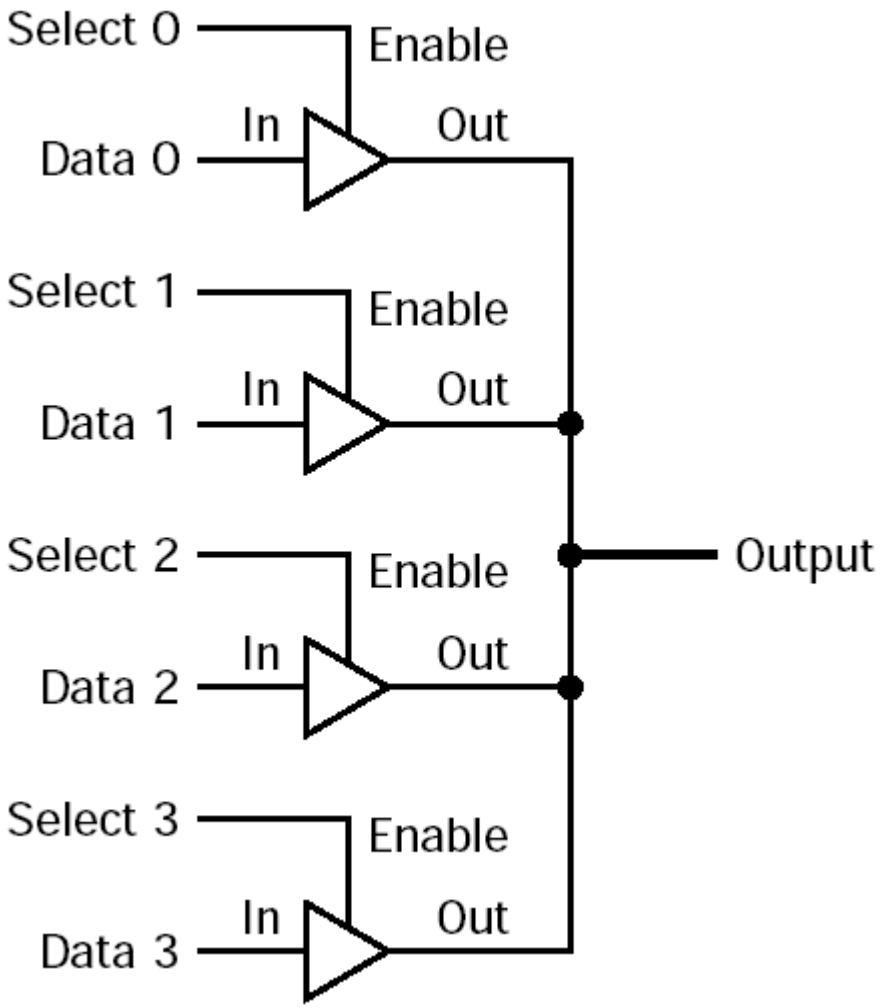


Typical SRAM Organization

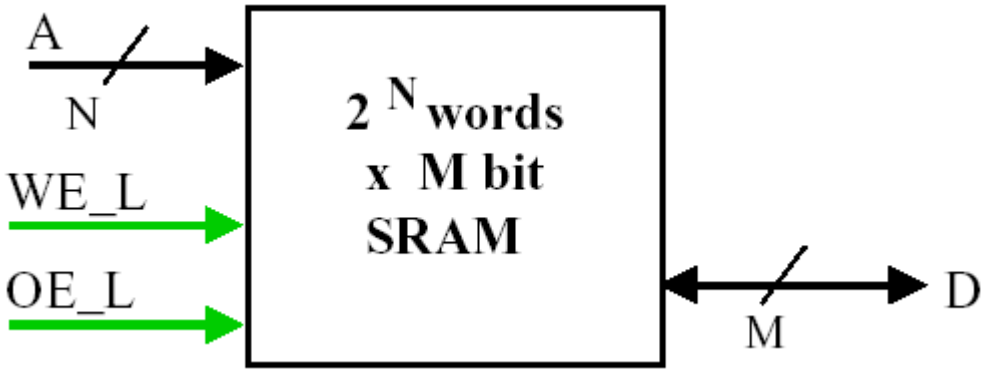


Three-State Buffers

Avoid HUGE MUX using three-state buffers

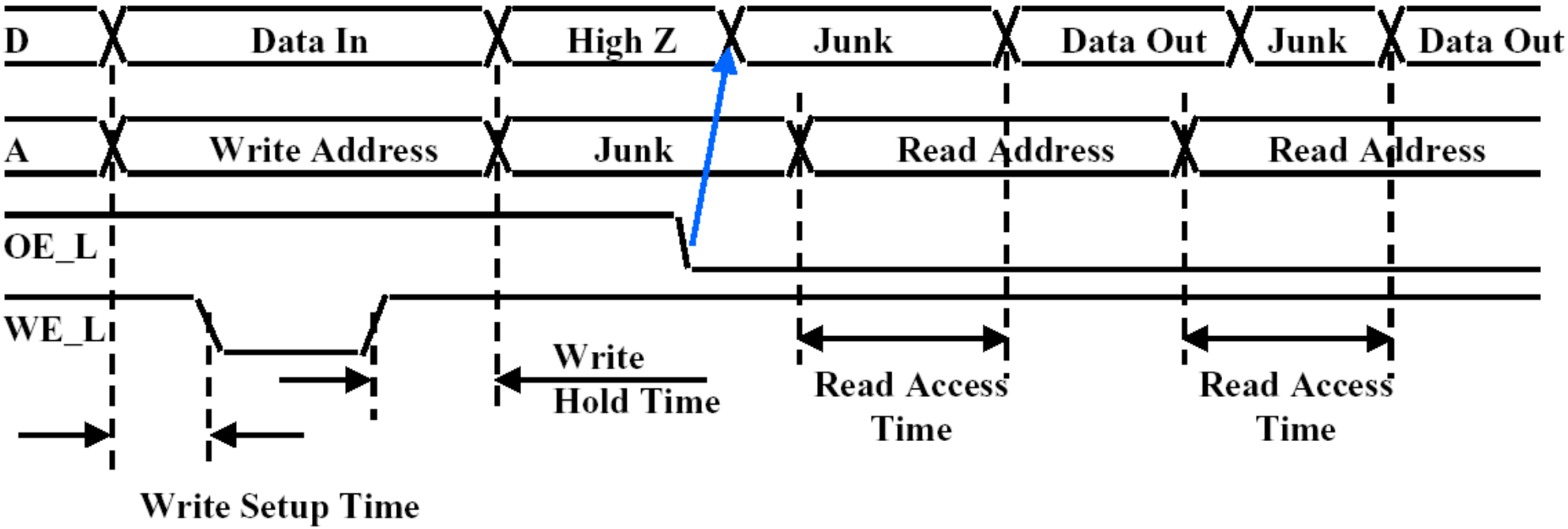


Typical SRAM Timing



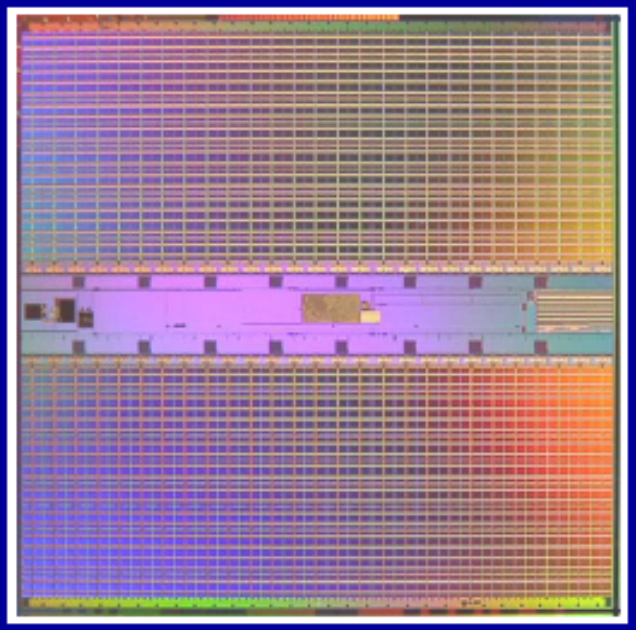
Write Timing:

Read Timing:



70 Mbit SRAM

- Fully functional 70 Mbit SRAM chips have been made
- >0.5 billion transistors
- 0.57 μm^2 cell size
- Uses all process features needed for 65 nm logic products

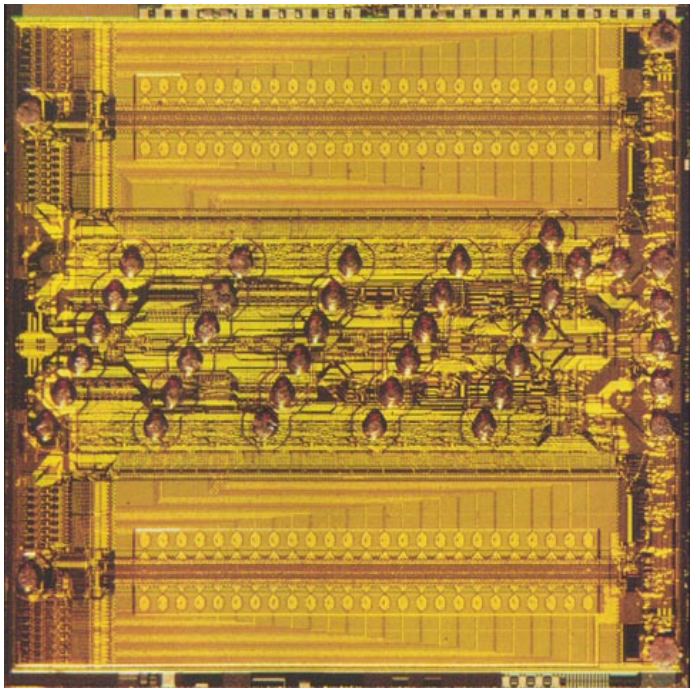


110 mm^2 chip size

Mark Bohr: Intel 04

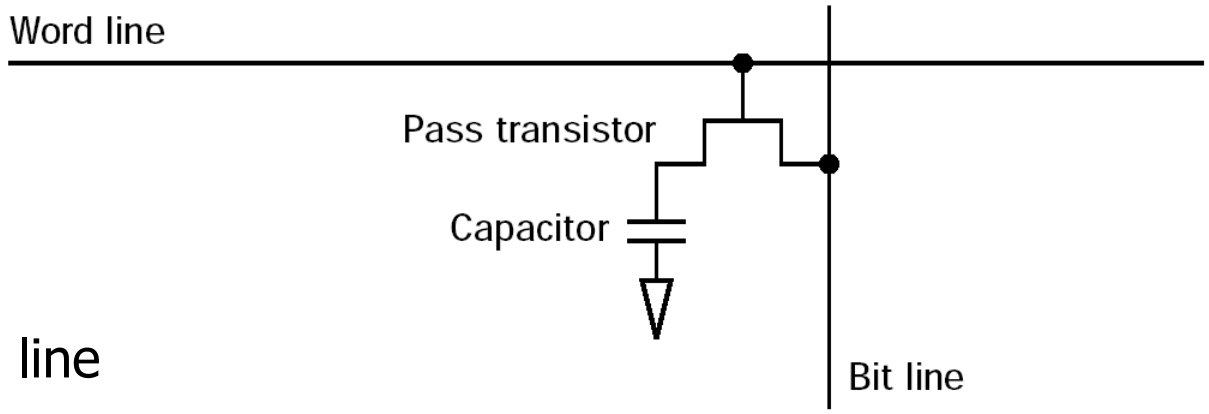
Dynamic RAM (DRAM)

- Slower, cheaper, more dense than SRAM
- Dynamic?



Dynamic RAM Cell

DRAM



Write:

1. Drive bit line
2. Select row/word line

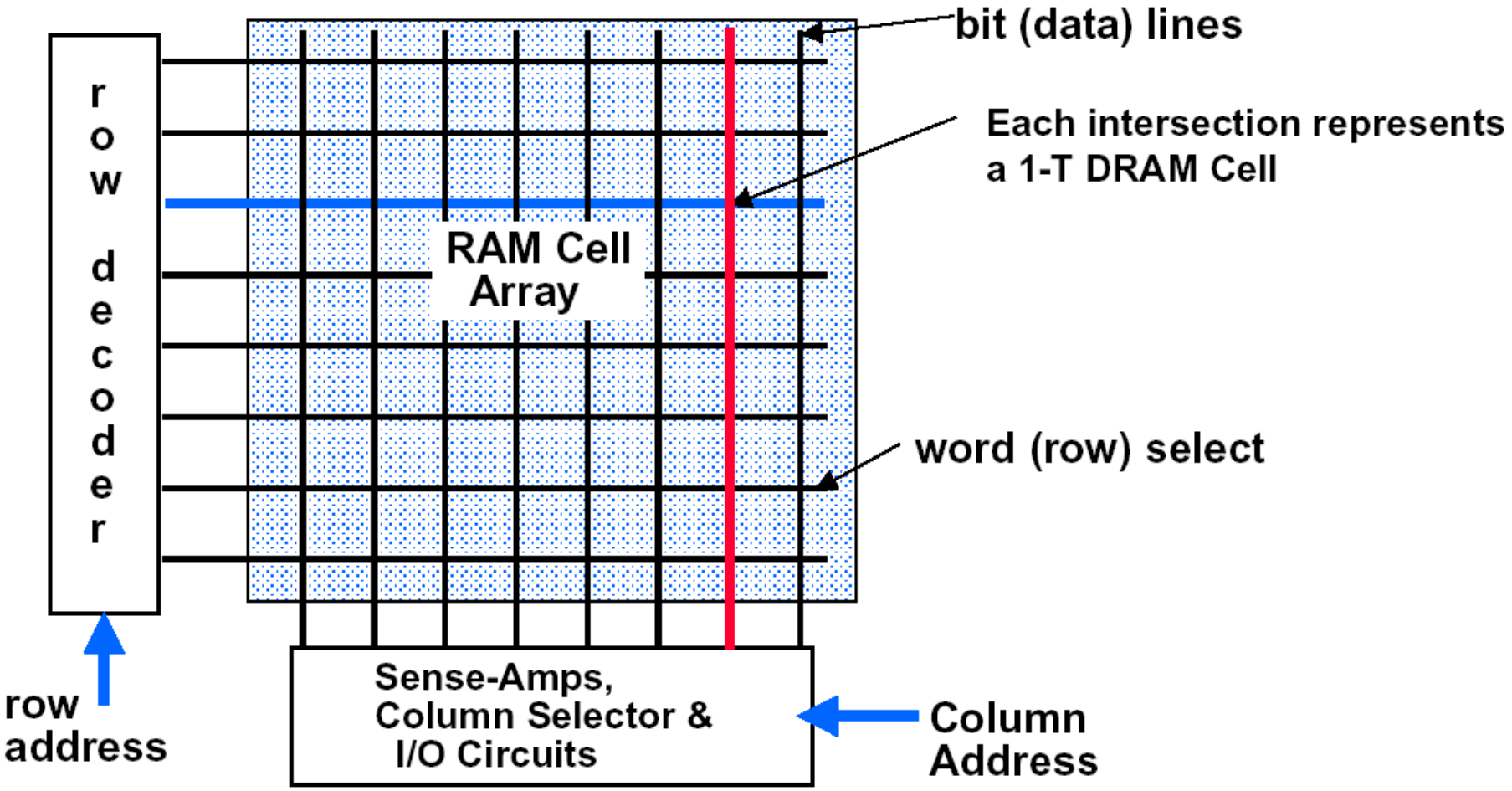
Read:

1. Precharge bit line to Vdd
2. Select row/word line
3. Cell and bit line share charge
4. Sense (sense amp can detect changes of ~10-100k electrons)
5. Write: restore the value

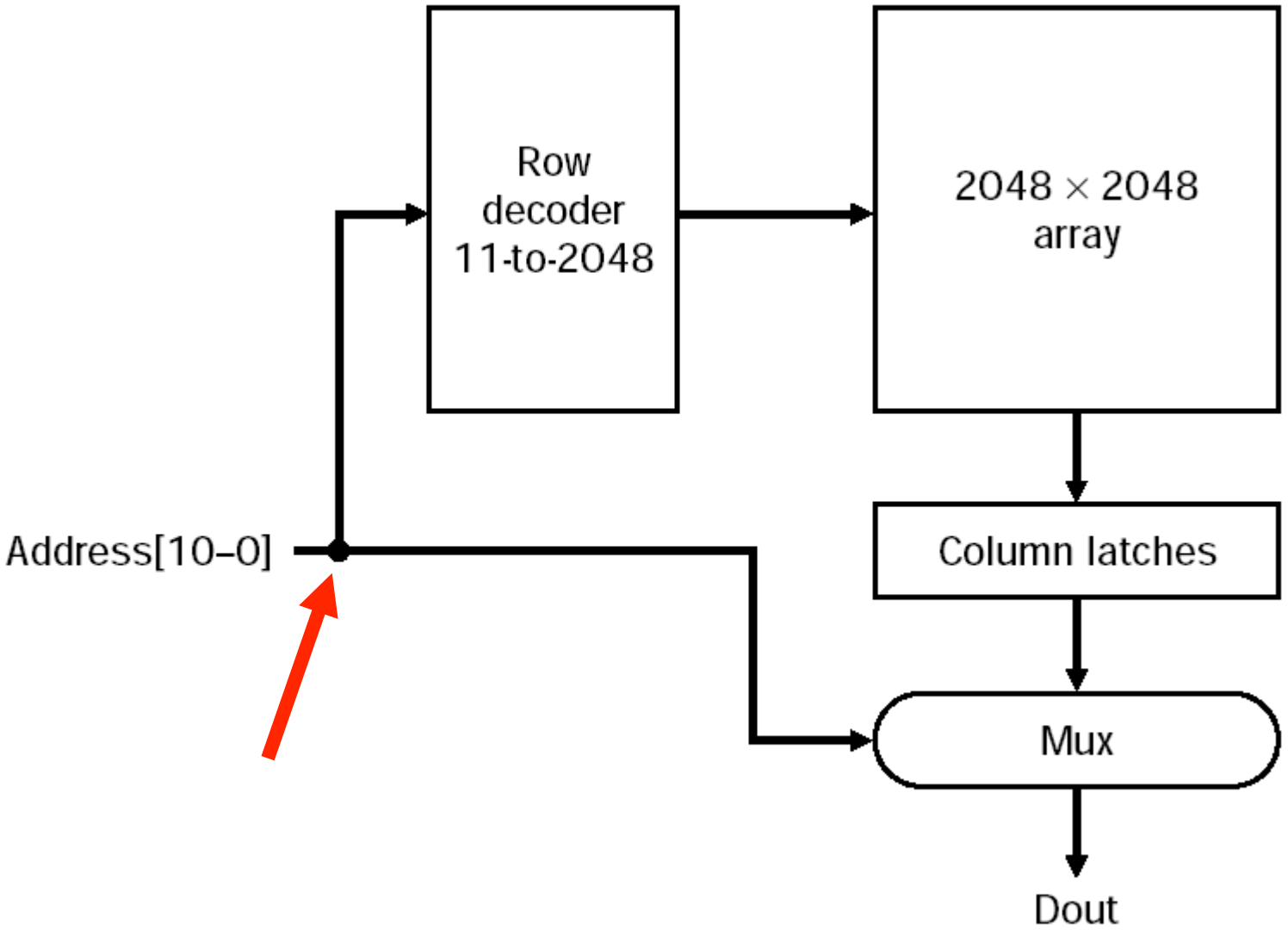
Refresh (capacitor leaks):

1. Just do a dummy read to every cell.

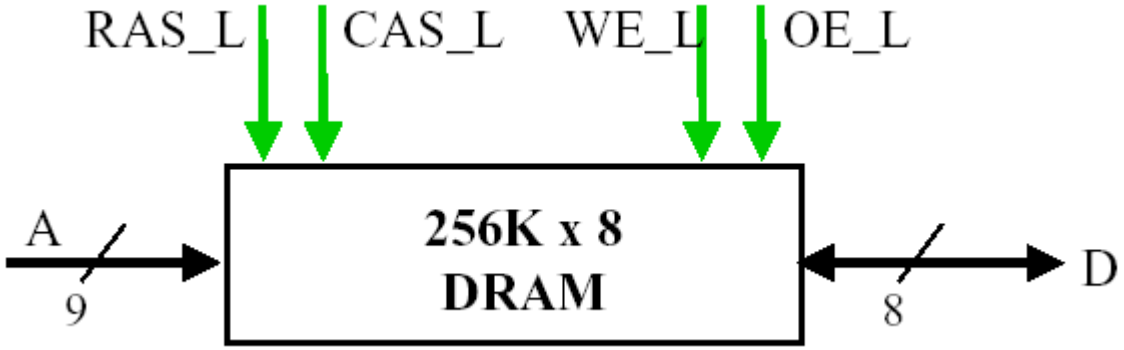
Classical DRAM Organization (Square)



4Mx1 DRAM Organization

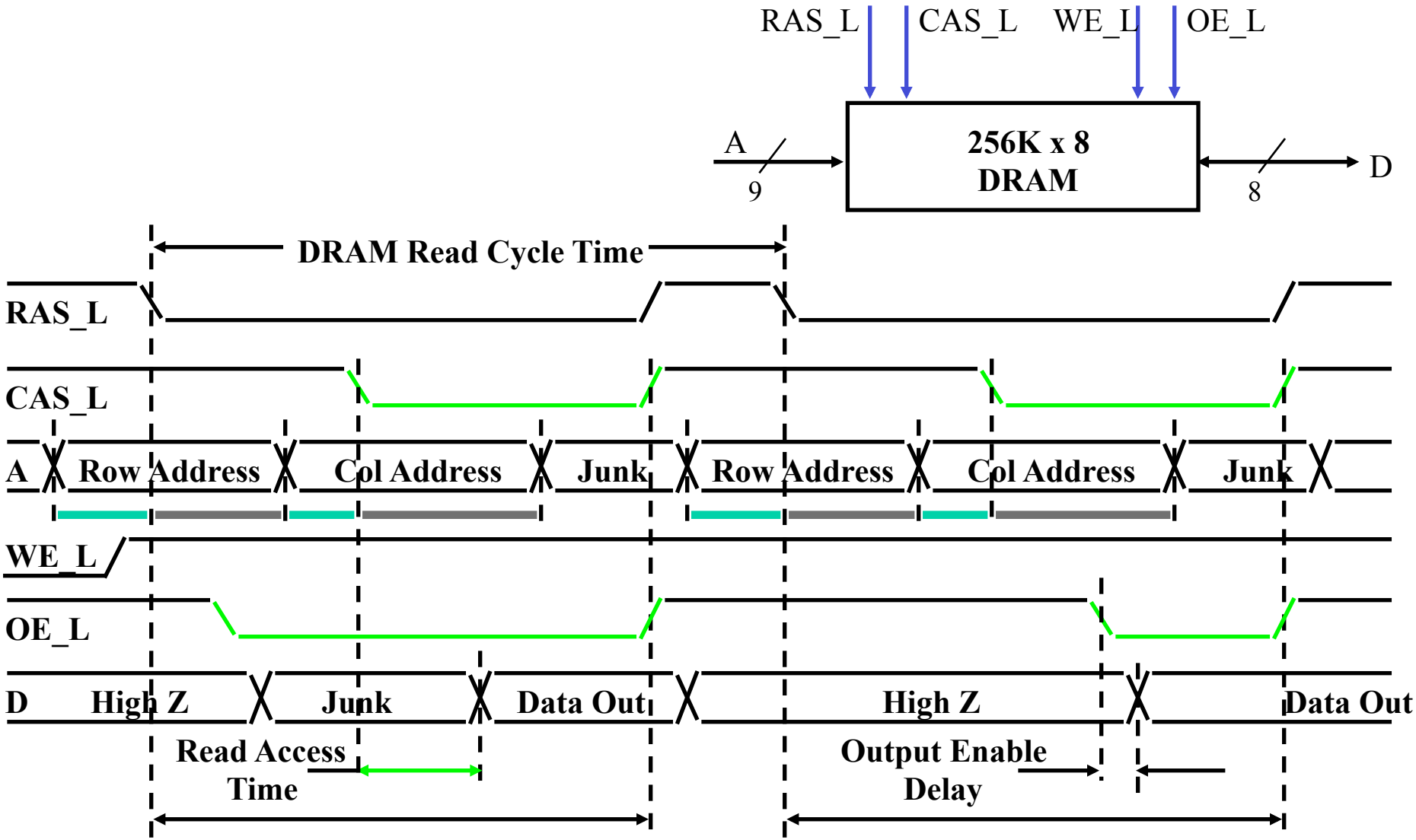


Logic Diagram of a Typical DRAM



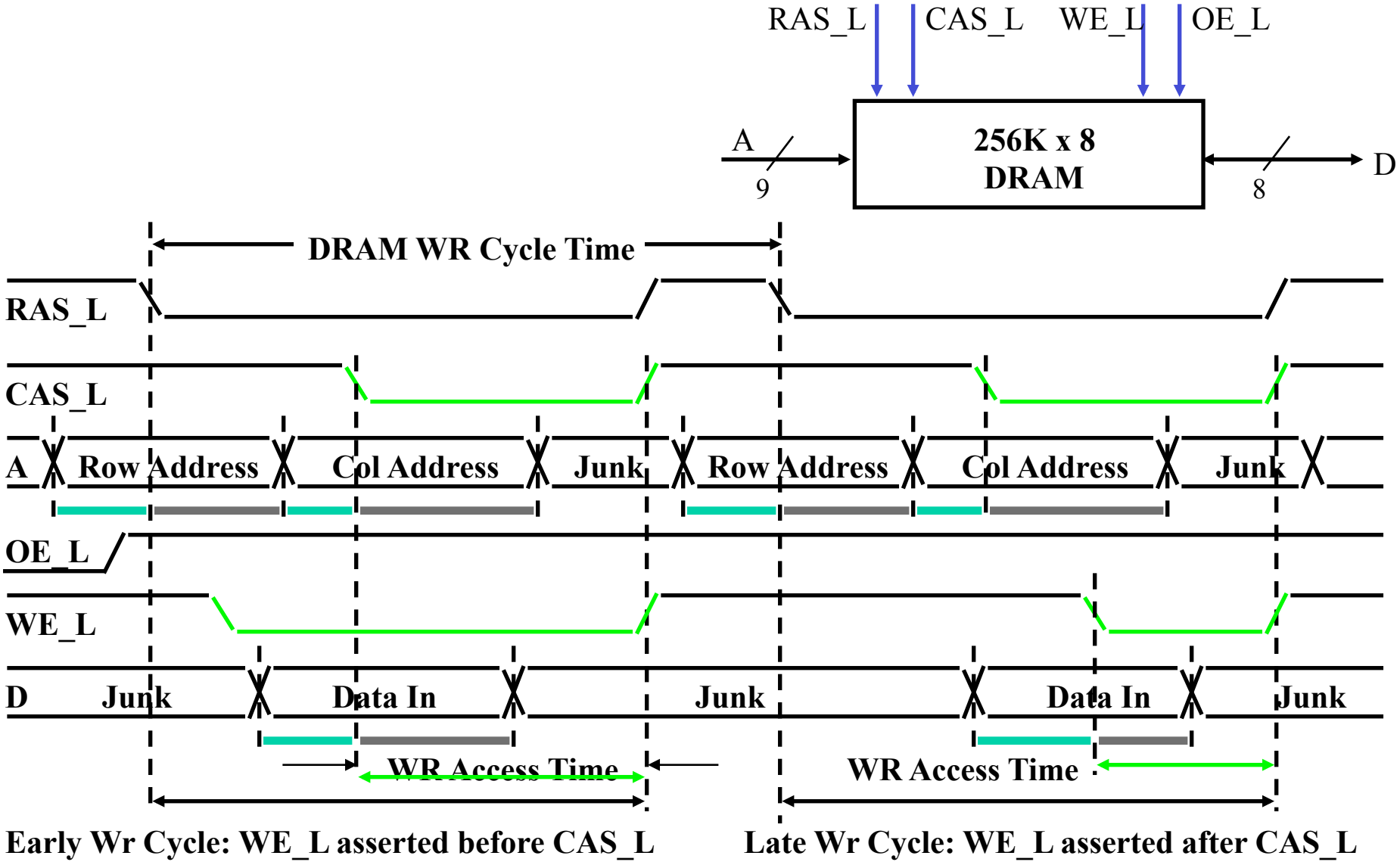
- Row and column addresses share the same pins (A)
  - RAS_L goes low: Pins A are latched in as row address
  - CAS_L goes low: Pins A are latched in as column address
  - RAS/CAS edge-sensitive
- Din and Dout share the same pins (D)
- Control Signals (RAS_L, CAS_L, WE_L, OE_L) typically active low

DRAM Read Timing



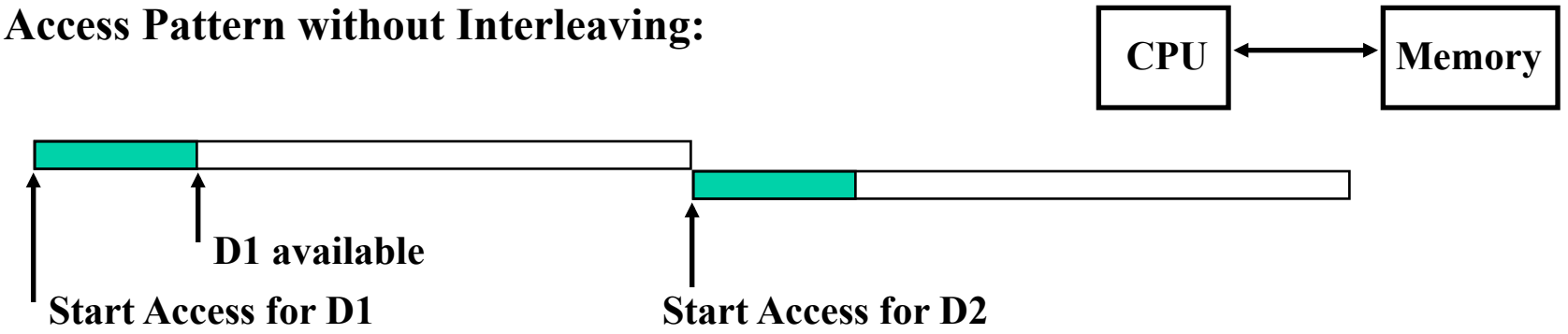
Early Read Cycle: OE_L asserted before CAS_L **Late Read Cycle: OE_L asserted after CAS_L**

DRAM Write Timing

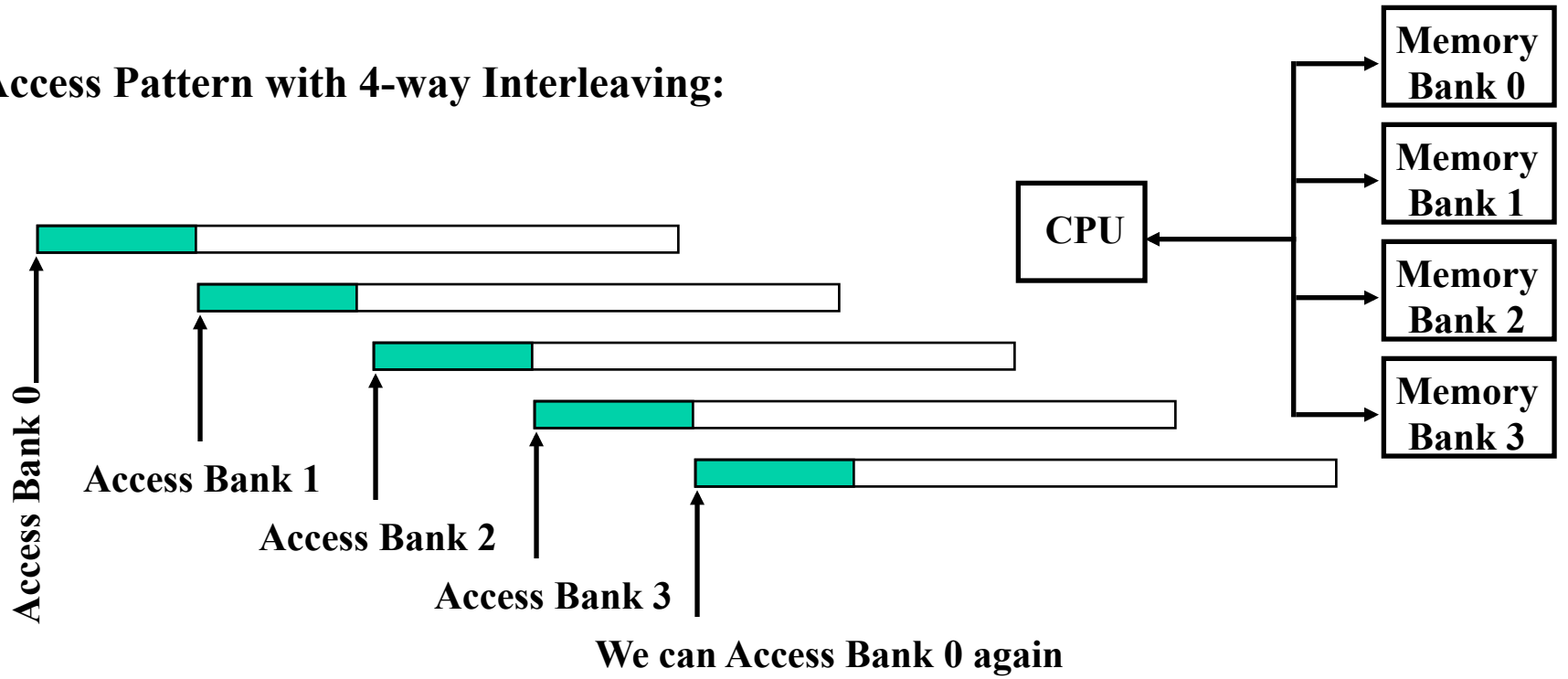


Increasing Bandwidth - Interleaving

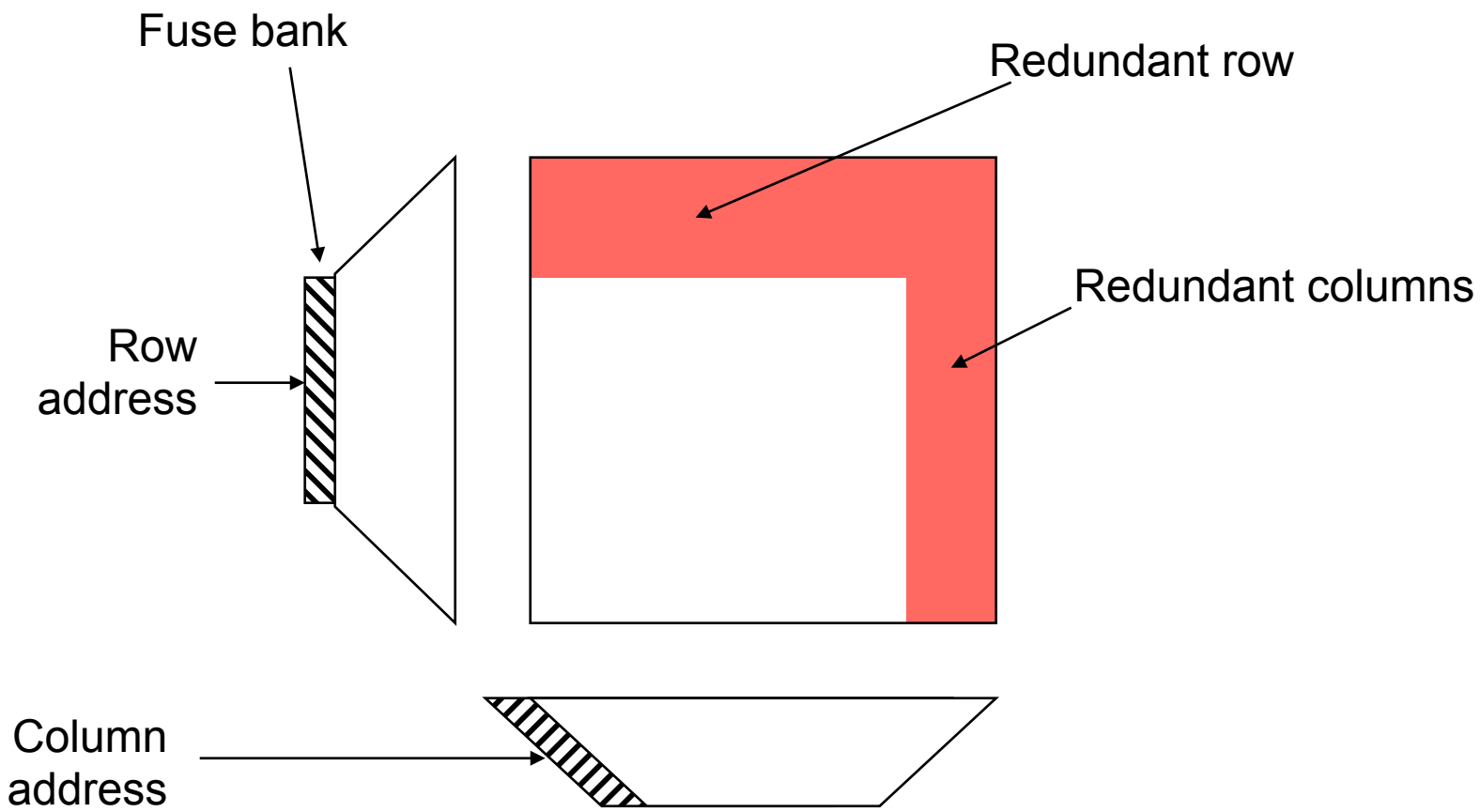
Access Pattern without Interleaving:



Access Pattern with 4-way Interleaving:

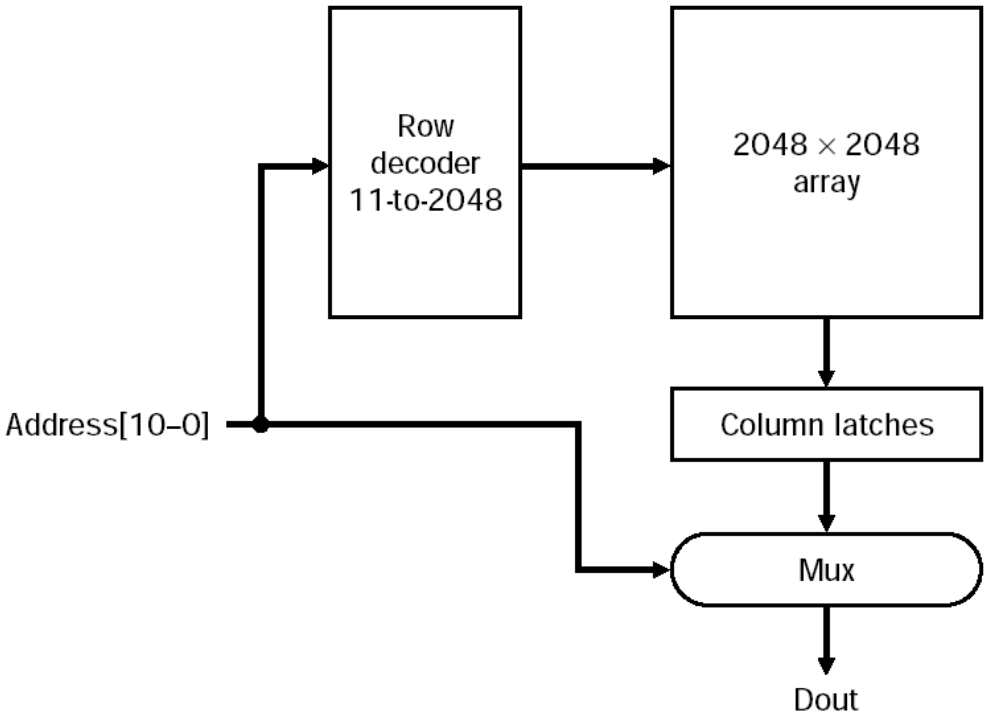


Improving Yield with Redundancy

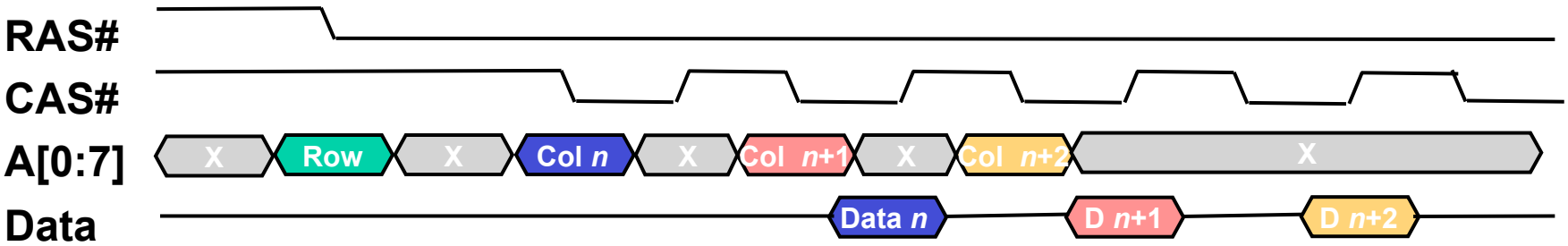


"New" DRAMs

- For decades, DRAM Interface was stable (RAS, CAS, etc.)
- Only in past decade has it begun to evolve again
  - Especially in systems with few DRAM chips
  - Bandwidth/Througput
  - Ease of design
- Several Enhancements:
  - Page Mode
  - EDO RAM
  - Burst DRAM
  - Synchronous DRAM
  - Rambus DRAM

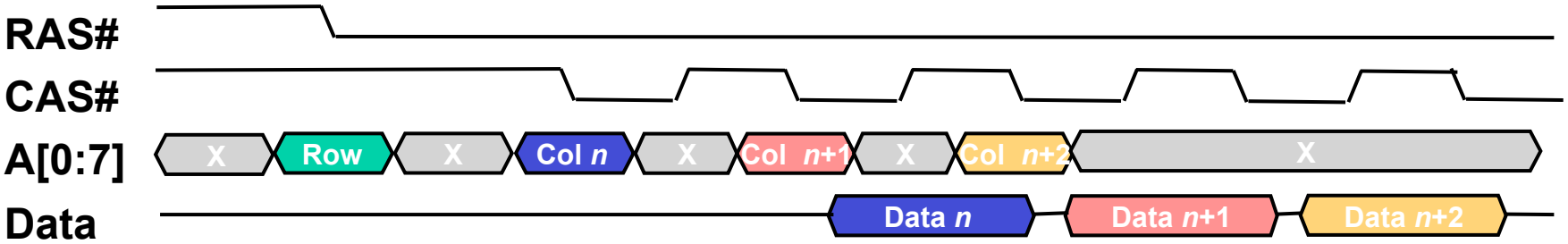


Paged Mode DRAM



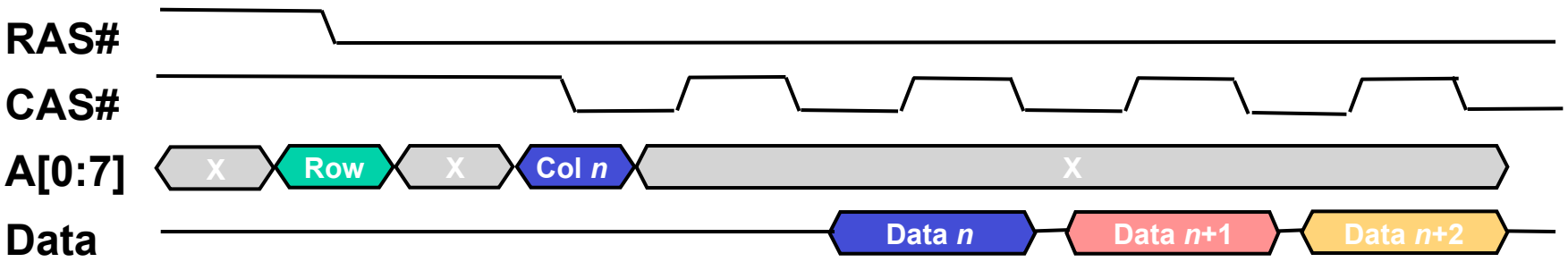
- Multiple accesses to different columns from same row
- Saves RAS and RAS to CAS delay

Extended Data Output RAM



- A data output latch enables to parallel next column address with current column data

Burst DRAM

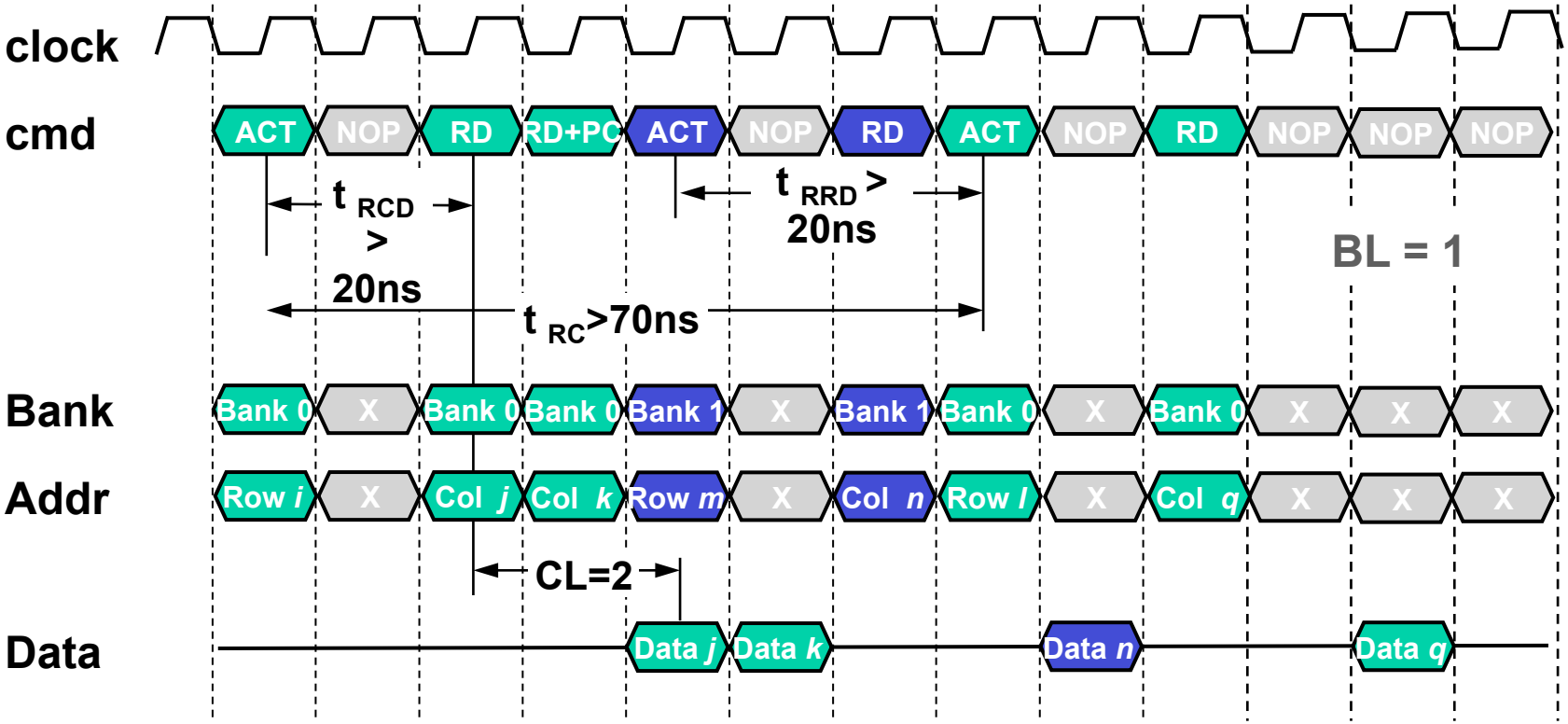


- Generate a consecutive address by itself

Synchronous DRAM – SDRAM

- All signals are referenced to an external clock (100MHz-200MHz)
  - Makes timing more precise with other system devices
- Multiple Banks
  - Multiple pages open simultaneously (one per bank)
- Command driven functionality instead of signal driven
  - ACTIVE: selects both the bank and the row to be activated
    - ACTIVE to a new bank can be issued while accessing current bank
  - READ/WRITE: select column
- Read and write accesses to the SDRAM are burst oriented
  - Successive column locations accessed in the given row
  - Burst length is programmable: 1, 2, 4, 8, and full-page
    - Full-page burst may end with a BURST TERMINATE to get arbitrary burst lengths
- A user programmable Mode Register
  - CAS latency, burst length, burst type
- Auto pre-charge: may close row at last read/write in burst
- Auto refresh: internal counters generate refresh address

SDRAM Timing

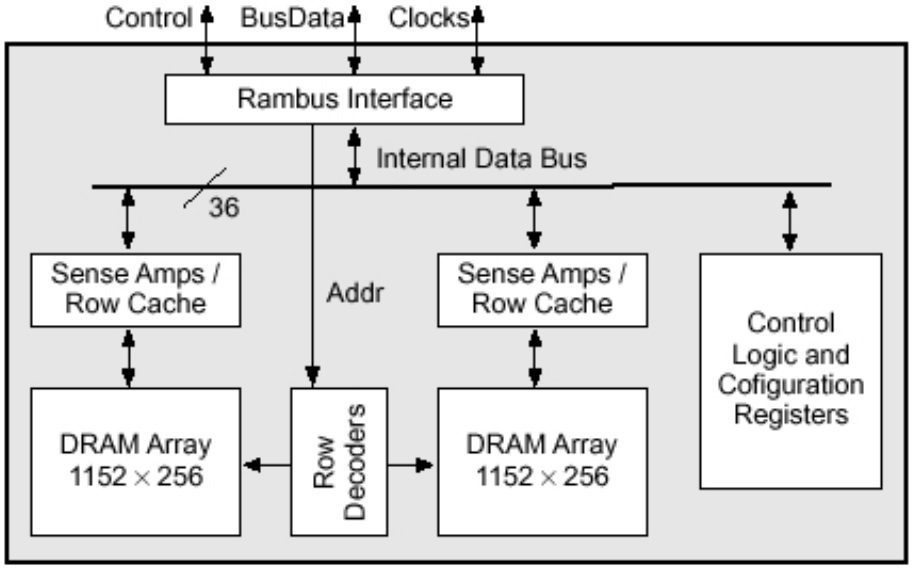


- t_{RCD}: ACTIVE to READ/WRITE gap = $\lceil t_{RCD}(MIN) / \text{clock period} \rceil$
- t_{RC}: successive ACTIVE to a different row in the same bank
- t_{RRD}: successive ACTIVE commands to different banks

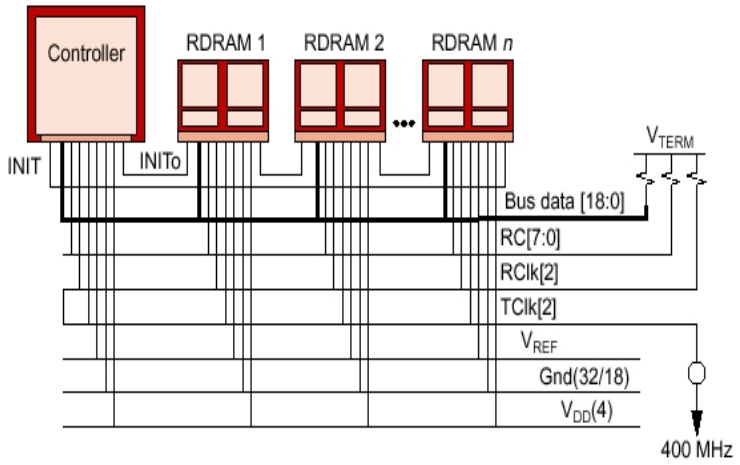
Rambus (R)DRAM

- Develop by Rambus Inc. and Intel
- Based on a narrow bus (16bits) runs at high speed 400Mhz
- Pipeline operation
- Multi array
- Data transfer on both edges

RDRAM System



RAMBUS Bank



RDRAM Memory System

Summary

DRAM → slow, cheap, dense

- Good for BIG main memory
- Must be refreshed

SRAM → fast, expensive, not very dense

- Good choice for fast memory like caches!
- Holds state while power applied

Memory hierarchy to get the best of both!