

# Chapter 08: The Memory System

## Lesson 11: **Memory Banking**

# Objective

- Learn parallel access and memory bank concepts

# Parallel Access

# Improving Performance by Banking or Interleaving

- Banking— placing chips in number of rows and columns at a given level of a memory hierarchy
- Interleaving— for facilitating the simultaneous accesses to the memory system

# Parallel memory system — a banked memory system

- A banked memory system, the data is divided, or *interleaved*, across the memories so that each memory contains only a fraction of the data
- Typically, some of the bits of the address are used to select which memory bank a given datum resides in

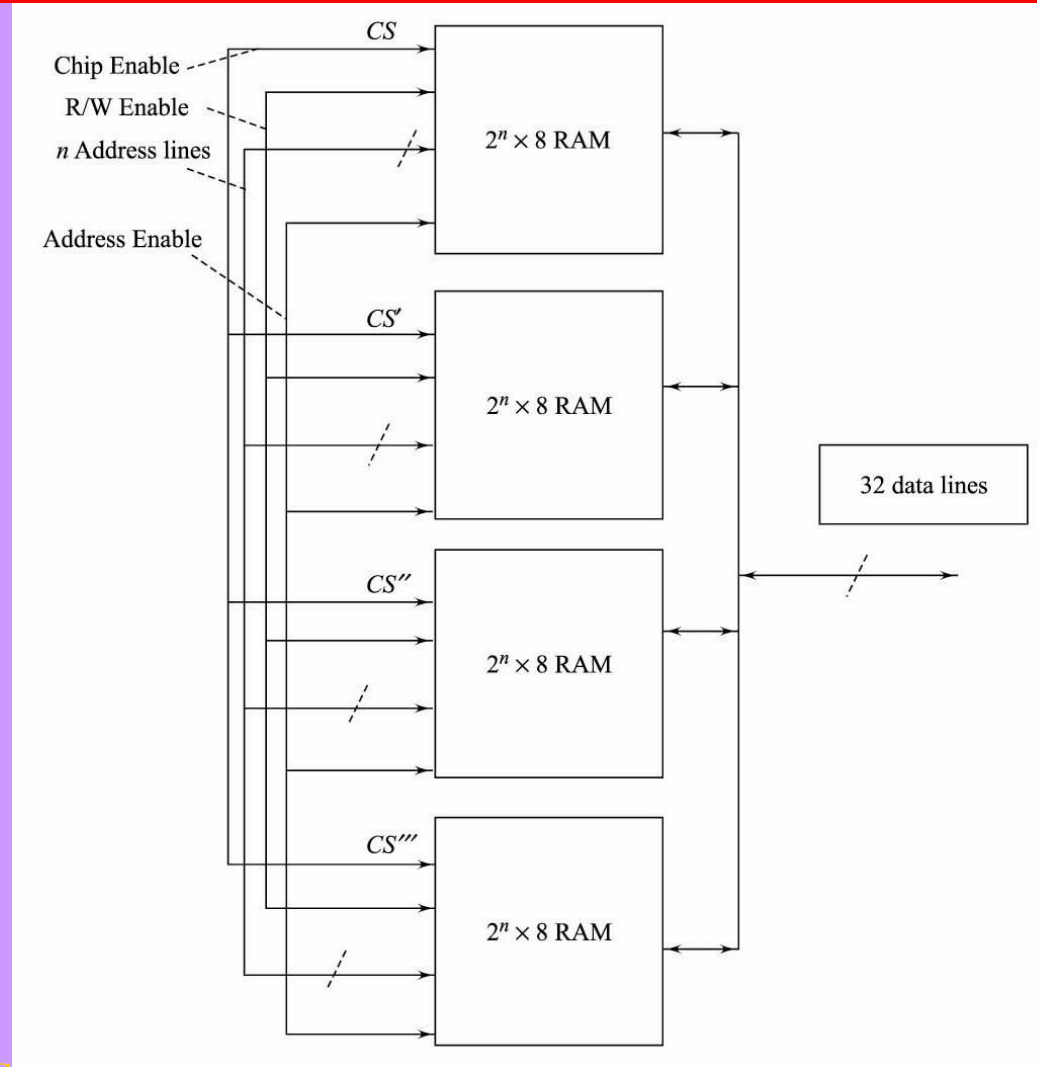
# 32-Word or 64-bit or 128 bit Word Accesses

- A word or the data bus is of 32 or 64 or 128 bits access in a memory design
- Improves the performance of memory systems
- Support to multiple memory references in parallel

# Parallel access to 32-Word or 64-bit or 128 bit Word Accesses

- 4 or 8 or 16 addresses accessed by 2 or 3 or 4 address lines simultaneously

# Parallel Access to four sets of RAMs





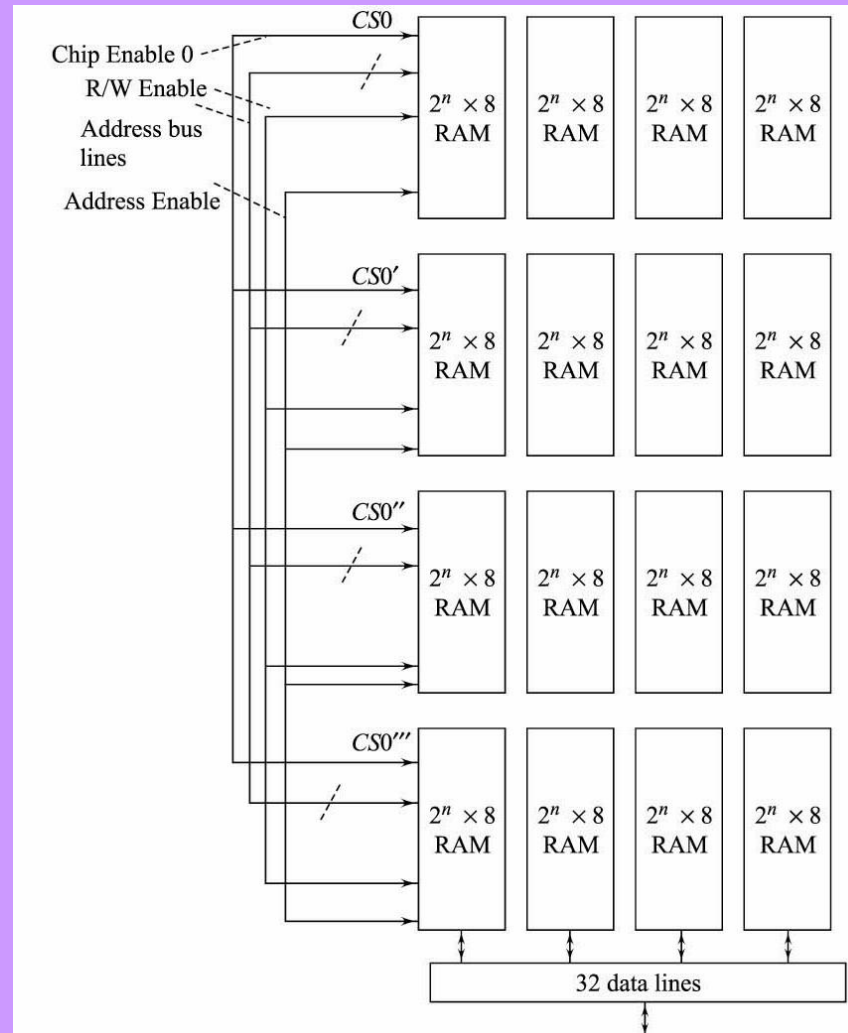
# Parallel access to 32-Word or 64-bit or 128 bit Word Accesses from multiple banks

- 1 or 4 or 8 addresses accessed by 1 or 2 or 3 address lines simultaneously vertically placed banks
- 4 horizontally placed chips in 4 banks accessed simultaneously

# Banking

- Two low-order address bits are 0b00 might be placed in the leftmost bank
- Bytes whose two low-order address bits are 0b01 might go in the next bank
- So on

# Parallel access to 32-Word bit Word Accesses from 4 banks



# Why actual bandwidth < Peak

# Why actual bandwidth < Peak

- No requests going to the memory system at times
- Conflicts for memory banks at times
- Other factors

# Example

- Assume— a memory system is constructed out of devices with a latency of 10 ns and no precharge delay
- Each bank transfers 4 bytes per access
- Compute how many banks does the memory system need to provide a peak bandwidth of at least  $1.5 \times 10^{10}$  bytes/s

# Solution

- Problem is of finding how many operations the memory system needs to be able to handle in parallel to achieve the  $1.5 \times 10^{10}$  bytes/s bandwidth
- First find the bandwidth of each memory bank
- Then divide the total bandwidth by that to get the number of banks required
- At 10 ns per access and 4 bytes transferred per access, a bank has a bandwidth =  $(1/10 \text{ ns}) \times 4 = 4 \times 10^8 \text{ Byte/s}$

## ... Solution

- Dividing  $1.5 \times 10^{10}$  by that value gives 38 as the required number of banks
- Number of banks must be rounded up to the next integer



# Summary

# We learnt

- Parallel access by common chip select signals
- Memory bank concept for the parallel access

# End of Lesson 11 on **Memory Banking**