

# **SNS COLLEGE OF TECHNOLOGY**



Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

# DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

19ECT312 – EMBEDDED SYSTEM DESIGN

III YEAR/ VI SEMESTER

UNIT 1 – INTRODUCTION TO EMBEDDED SYSTEMS

TOPIC 1.7 –EMBEDDED MEMORY





- •fMemory Hierarchy fCache Memory –
- Different types of Cache Mappings –
- Cache Impact on System

Performance fDynamic Memory -

Different types of Dynamic RAMs f

**Memory Management Unit** 





Memory Hierarchy Objective is to use inexpensive, fast memory

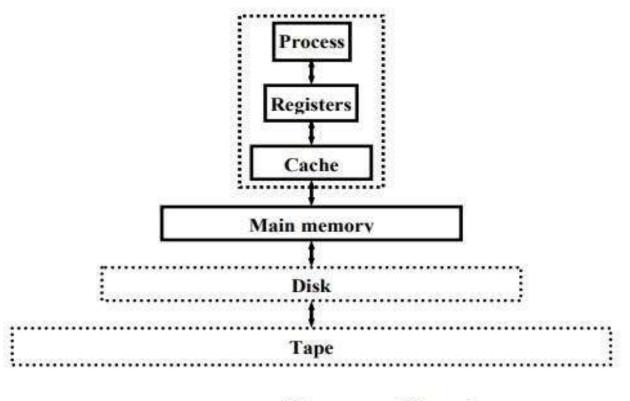
• Main memory f

Large, inexpensive, slow memory stores entire program and data

• Cache f

Small, expensive, fast memory stores copy of likely accessed parts of larger

memory fCan be multiple levels of cache



The memory Hierarchy





#### Cache

- Usually designed with SRAM f
  - faster but more expensive than DRAM
- Usually on same chip as processor f
  - space limited, so much smaller than off-chip main memory f Faster access (1 cycle vs. several cycles for main memory)
- Cache operation f

Request for main memory access (read or write) f First, check cache for copy f cache hit

- copy is in cache, quick access fcache miss - copy not in cache, read address and possibly its neighbors into cache

• Several cache design choices f cache mapping, replacement policies, and write techniques





## **Cache Mapping**

- Necessary as there are far fewer number of available cache addresses than the memory
- Are address' contents in cache?
- •Cache mapping used to assign main memory address to cache address and determine hit or miss
- Three basic techniques: f
  - 1. Direct mapping
  - 2. fFully associative mapping f
  - 3. Set-associative mapping
- •Caches partitioned into indivisible blocks or lines of adjacent memory addresses f usually 4 or 8 addresses per line





## **Direct Mapping**

• Main memory address divided into 2 fields f

Index which contains

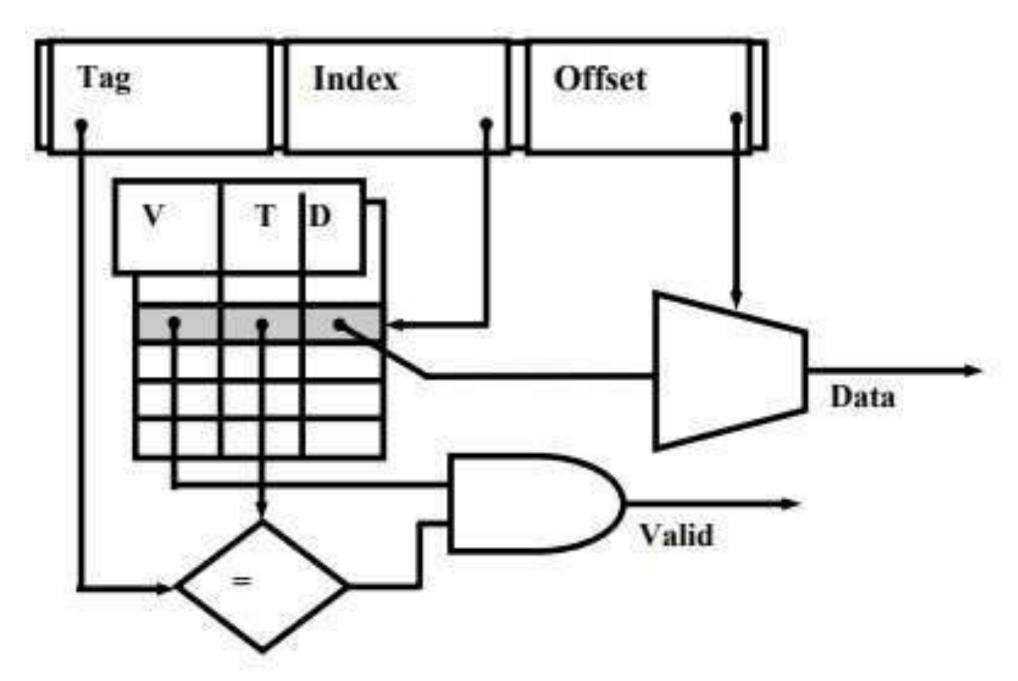
- cache address
- -number of bits determined by cache size f

#### -Tag

- -- compared with tag stored in cache at address indicated by index
- - if tags match, check valid bit
  - -• Valid bit f
    - indicates whether data in slot has been loaded from memory
  - -• Offset
    - -fused to find particular word in cache line







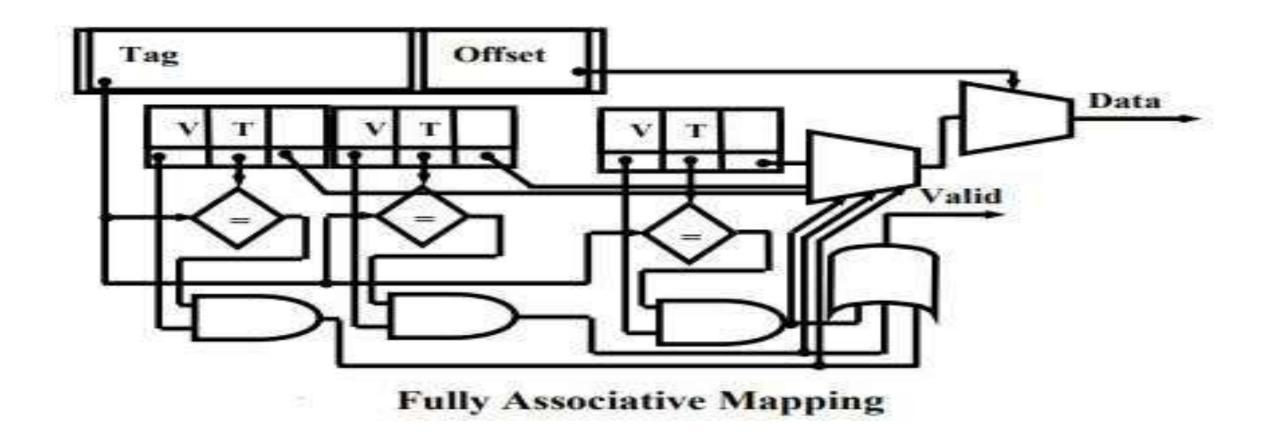
Direct Mapping





#### **Fully Associative Mapping**

- Complete main memory address stored in each cache address
- All addresses stored in cache simultaneously compared with desired address
- Valid bit and offset same as direct mapping

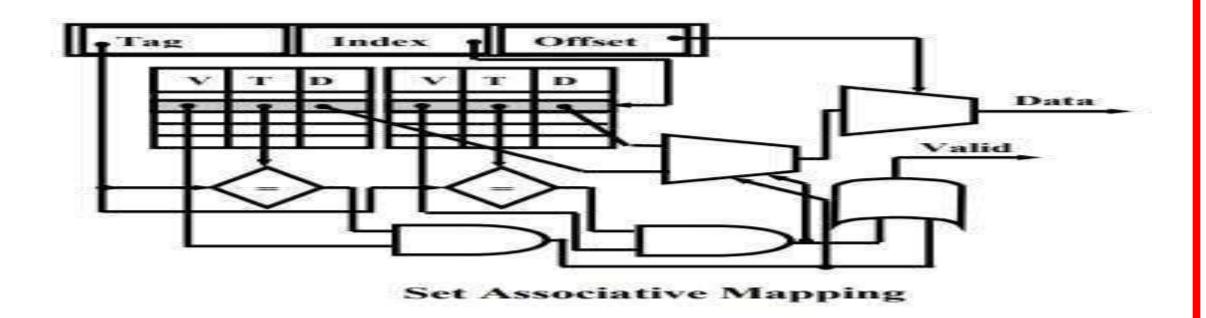






# **Set-Associative Mapping**

- Compromise between direct mapping and fully associative mapping
- Index same as in direct mapping
- •But, each cache address contains content and tags of 2 or more memory address locations
- Tags of that set simultaneously compared as in fully associative mapping
- Cache with set size N called N-way set-associative f
   2-way, 4-way, 8-way are common







# Cache-Replacement Policy

- Technique for choosing which block to replace f
  when fully associative cache is full f
  when set-associative cache's line is full
- Direct mapped cache has no choice
- Random f
   replace block chosen at random
- LRU: **least-recently used** f replace block not accessed for longest time
- FIFO: first-in-first-out f
   push block onto queue when accessed f
   choose block to replace by popping queue





#### Cache Write Techniques

- When written, data cache must update main memory
- Write-through f

write to main memory whenever cache is written to f
easiest to implement
fprocessor must wait for slower main memory write f
potential for unnecessary writes

#### Write-back f

main memory only written when "dirty" block replaced f
extra dirty bit for each block set when cache block written to f
reduces number of slow main memory writes





## Cache Impact on System Performance

- •Most important parameters in terms of performance: fTotal size of cache
  - total number of data bytes cache can hold
  - -tag, valid and other house keeping bits not included in total f
- -Degree of associatively f
- -Data block size





Larger caches achieve lower miss rates but higher access cost

- e.g.,
  - 2 Kbyte cache: miss rate = 15%, hit cost = 2 cycles, miss cost = 20 cycles
    - avg. cost of memory access
    - = (0.85 \* 2) + (0.15 \* 20) = 4.7 cycles
  - 4 Kbyte cache: miss rate = 6.5%, hit cost = 3 cycles, miss cost will not change
    - avg. cost of memory access = (0.935 \* 3) + (0.065 \* 20) = 4.105
       cycles (improvement)
  - 8 Kbyte cache: miss rate = 5.565%, hit cost = 4 cycles, miss cost will not change
    - avg. cost of memory access = (0.94435 \* 4) + (0.05565 \* 20) = 4.8904 cycles

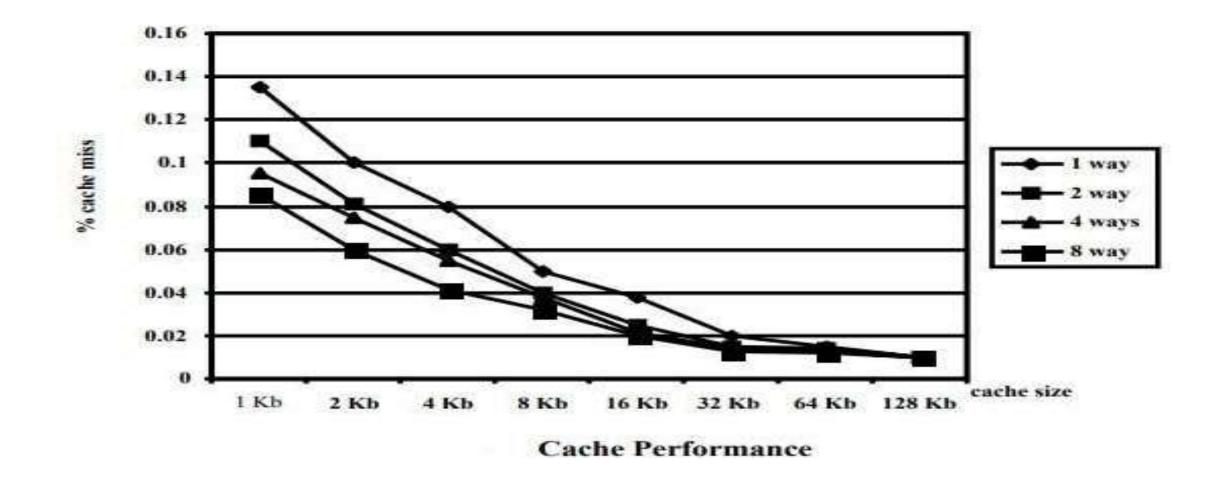




#### Cache Performance Trade-Offs

• Improving cache hit rate without increasing size f
Increase line size f

Change set-associativity







#### Advanced RAM

DRAMs commonly used as main memory in processor based embedded systems f high capacity, low cost

## Many variations of DRAMs proposed f

need to keep pace with processor speeds

- **1 FPM DRAM**: fast page mode DRAM f
- 2. EDO DRAM: extended data out DRAM f
- 3. SDRAM/ESDRAM: synchronous and enhanced synchronous DRAM f
- 4. RDRAM: rambus DRAM



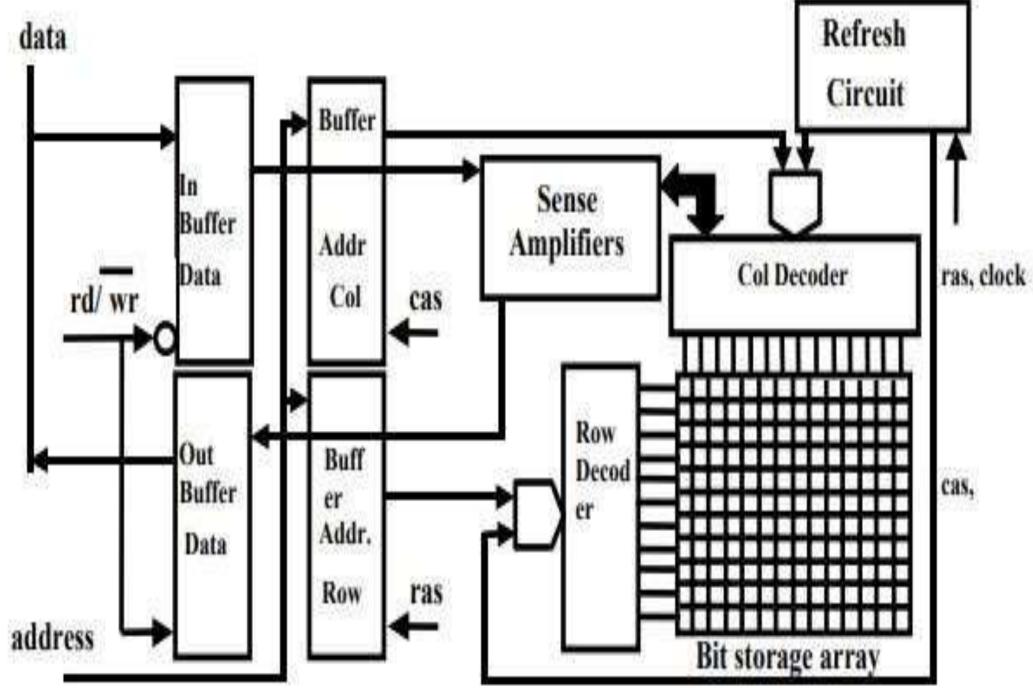


#### **Basic DRAM**

- Address bus multiplexed between row and column components
- •Row and column addresses are latched in, sequentially, by strobing ras (row address strobe) and cas (column address strobe) signals, respectively
- Refresh circuitry can be external or internal to DRAM device f
  - •strobes consecutive memory address periodically causing memory content to be refreshed f
  - •Refresh circuitry disabled during read or write operation







The Basic Dynamic RAM Structure





## Fast Page Mode DRAM (FPM DRAM)

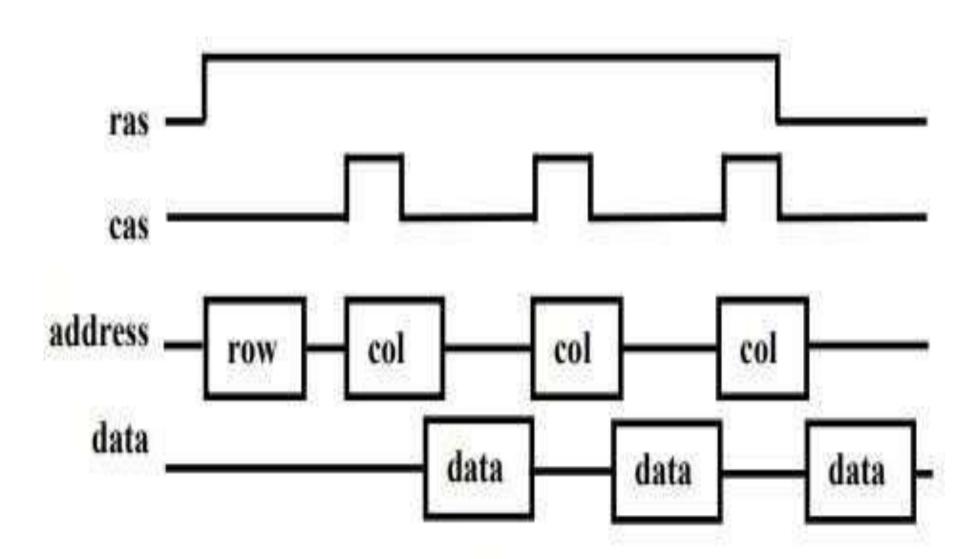
- Each row of memory bit array is viewed as a page
- Page contains multiple words
- Individual words addressed by column address
- Timing diagram: f

row (page) address sent

f3 words read consecutively by sending column address for each Extra cycle eliminated on each read/write of words from same







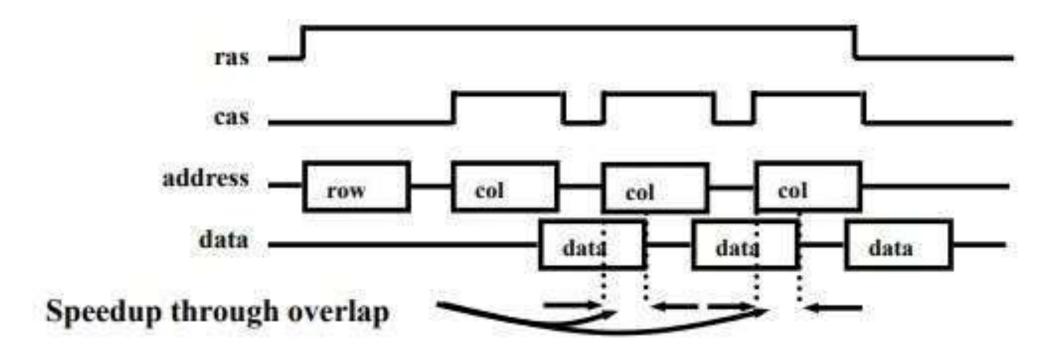
The timing diagram in FPM DRAM





## Extended data out DRAM (EDO DRAM)

- Improvement of FPM DRAM
- Extra latch before output buffer f allows strobing of cas before data read operation completed
- Reduces read/write latency by additional cycle



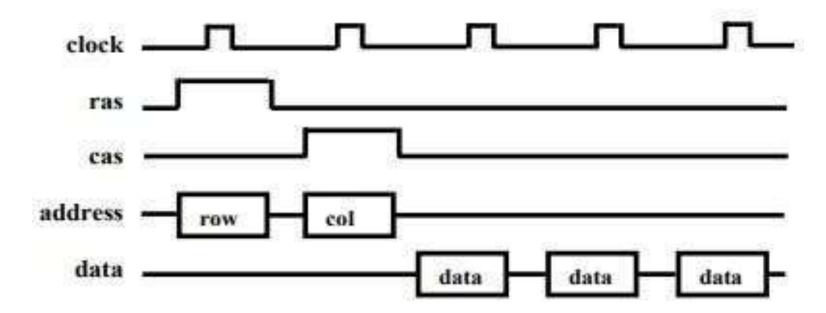
The timing diagram in EDORAM





## (S)ynchronous and Enhanced Synchronous (ES) DRAM

- SDRAM latches data on active edge of clock
- Eliminates time to detect ras/cas and rd/wr signals
- •A counter is initialized to column address then incremented on active edge of clock to access consecutive memory locations
- ESDRAM improves SDRAM f
   added buffers enable overlapping of column addressing f
   faster clocking and lower read/write latency possible



The timing diagram in SDRAM

## Rambus DRAM (RDRAM)



- Data is latched on both rising and falling edge of clock
- •Broken into 4 banks each with own row decoder fcan have 4 pages open at a time
- Capable of very high throughput

## **DRAM Integration Problem**

- SRAM easily integrated on same chip as processor
- DRAM more difficult f

Different chip making process between DRAM and conventional logic f

## Goal of conventional logic (IC) designers: -

 $\label{eq:minimize} minimize parasitic capacitance \ to \ reduce \ signal \ propagation \ delays \ and \\ power \ consumption \ f$ 

## •Goal of DRAM designers: -

create capacitor cells to retain stored information fIntegration processes beginning to appear





## Memory Management Unit (MMU)

- Duties of MMU f
  - •Handles DRAM refresh, bus interface and arbitration f
  - •Takes care of memory sharing among multiple processors
  - •fTranslates logic memory addresses from processor to physical memory addresses of DRAM
- Modern CPUs often come with MMU built-in
- Single-purpose processors can be used





# **SUMMARY & THANK YOU**