Avoiding memory banks Conflicts

- Suppose that we have 128 banks, and we will store 512x512 array.
- All the elements of a row will be mapped to the same bank (conflicts if we access a row.
- Usually, the number of banks is a power of 2, in this case
- Bank number = address MOD number of banks
- Address within a bank =Address/Number of banks
- This is a trivial calculation if the number of banks is a power of 2.
- If the number of memory banks is a prime number, that will decrease conflicts, but division and MOD will be very expensive

Copyright © 2012, Elsevier Inc. All rights reserved.

Avoiding memory Banks Conflicts

- MOD can be calculated very efficiently if the prime number is 1 less than a power of 2.
- Division still a problem
- But if we change the mapping such that
- Address in a bank = address MOD number of words in a bank.
- Since the number of words in a bank is usually a power of 2, that will lead to a very efficient implementation.
- Consider the following example, the first case is the usual 4 banks, then 3 banks with sequential interleaving and modulo interleaving and notice the conflict free access to rows and columns of a 4 by 4 matrix

Copyright © 2012, Elsevier Inc. All rights reserved.

Example

Add in a bank					SE	Q		M	O	D
	0	1	2	3	0	1	2	0	1	2
0	0	1	2	3	0	1	2	0	16	8
1	4	5	6	7	3	4	5	9	1	17
2	8	9	10	11	6	7	8	18	10	2
3	12	13	14	15	9	10	11	3	19	11
4	16	17	18	19	12	13	14	12	4	20
5	20	21	22	23	15	16	17	21	13	5
6	24	25	26	27	18	19	20	6	22	14
7	28	29	30	31	21	22	23	15	7	23

Copyright © 2012, Elsevier Inc. All rights reserved.

Memory Optimizations

- DDR:
  - DDR2
    - Lower power (2.5 V -> 1.8 V)
    - Higher clock rates (266 MHz, 333 MHz, 400 MHz)
  - DDR3
    - 1.5 V
    - 800 MHz
  - DDR4
    - 1-1.2 V
    - 1600 MHz
- GDDR5 is graphics memory based on DDR3

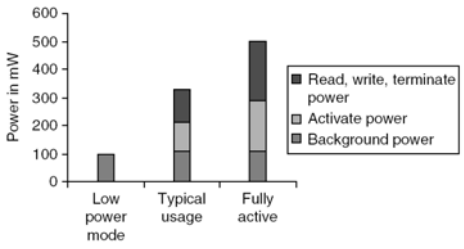
Copyright © 2012, Elsevier Inc. All rights reserved.

Memory Optimizations

- Graphics memory:
  - Achieve 2-5 X bandwidth per DRAM vs. DDR3
    - Wider interfaces (32 vs. 16 bit)
    - Higher clock rate
      - Possible because they are attached via soldering instead of socketted DIMM modules
- Reducing power in SDRAMs:
  - Lower voltage
  - Low power mode (ignores clock, continues to refresh)

Copyright © 2012, Elsevier Inc. All rights reserved.

Memory Power Consumption



Copyright © 2012, Elsevier Inc. All rights reserved.

Flash Memory

- Type of EEPROM
- Must be erased (in blocks) before being overwritten
- Non volatile
- Limited number of write cycles
- Cheaper than SDRAM, more expensive than disk
- Slower than SRAM, faster than disk

Copyright © 2012, Elsevier Inc. All rights reserved.

Memory Dependability

- Memory is susceptible to cosmic rays
- *Soft errors*: dynamic errors
  - Detected and fixed by error correcting codes (ECC)
- *Hard errors*: permanent errors
  - Use sparse rows to replace defective rows
- Chipkill: a RAID-like error recovery technique

Copyright © 2012, Elsevier Inc. All rights reserved.

Virtual Memory

- Protection via virtual memory
  - Keeps processes in their own memory space
- Role of architecture:
  - Provide user mode and supervisor mode
  - Protect certain aspects of CPU state
  - Provide mechanisms for switching between user mode and supervisor mode
  - Provide mechanisms to limit memory accesses
  - Provide TLB to translate addresses

Copyright © 2012, Elsevier Inc. All rights reserved.

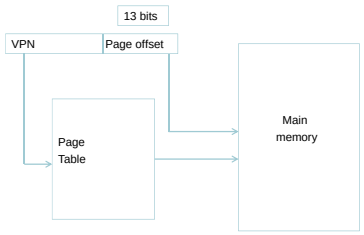
Virtual Memory

- Virtual memory references are generated by the compiler
- Physical memory is shared between many processes.
- Physical memory may be smaller than virtual memory.
- Need some mechanism to translate between virtual and physical memory.
- Need also a protection scheme to allow processes to reference only memory that belongs to them.

Copyright © 2012, Elsevier Inc. All rights reserved.

Virtual Memory

- Page table is used to translate virtual memory to physical memory



Copyright © 2012, Elsevier Inc. All rights reserved.

TLB

- Every memory reference takes 2 memory accesses.
- TLB is used to improve performance
- TLB is a small cache to store part of the page table

Copyright © 2012, Elsevier Inc. All rights reserved.