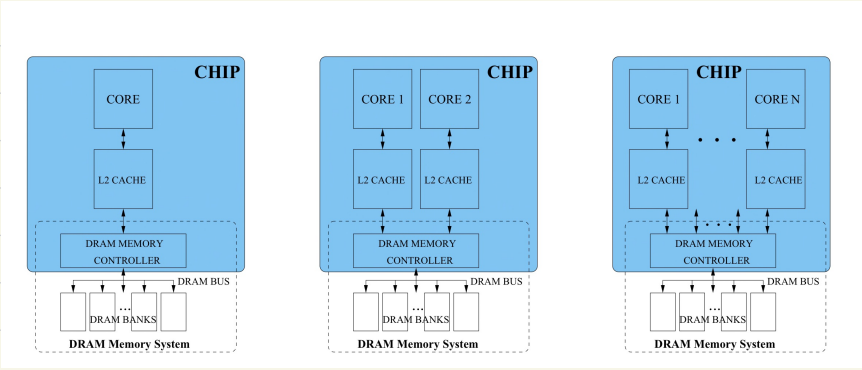
DRAM Bank

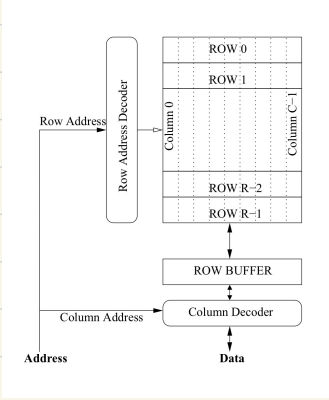
假設為 - Multicore, DRAM shared memory 系統

多个 core 共同存取 DRAM

∴ DRAM 設計為分割成多个 DRAM Bank, 可平行處理来自 core 的 request



DRAM Bank 為資料真正存放的地方, 各个 Bank 具各自的 row buffer, 以加速 2-D 資料存取



∴ 存取 2 維資料時, 可分為: (1) row-buffer hit

(2) row-buffer miss

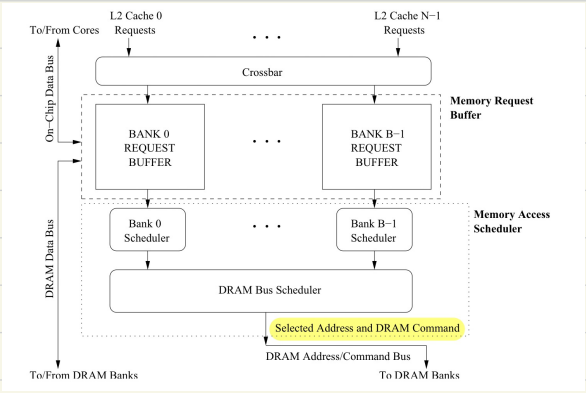
若為 miss, 會有 miss penalty, 增加 access time

∴ 具高 row-buffer locality 之 request 較快

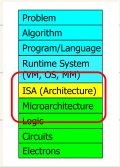
sequential access 之 program access time 較低

DRAM controller 含有各个 bank 之 bank request buffer 和 bank request scheduler 和 DRAM bus scheduler
來排程 memory bank request 之處理順序

為了增加 DRAM bandwidth, bank request scheduler 會以具高 row-buffer locality 之 request 具 high priority
但在 multicore 系統中, 可能會造成具 high locality of row-buffer 之 thread 持續具 high priority
使其它 thread 無法得到所需資源而執行時間上升, 而 slowdown



Solution:

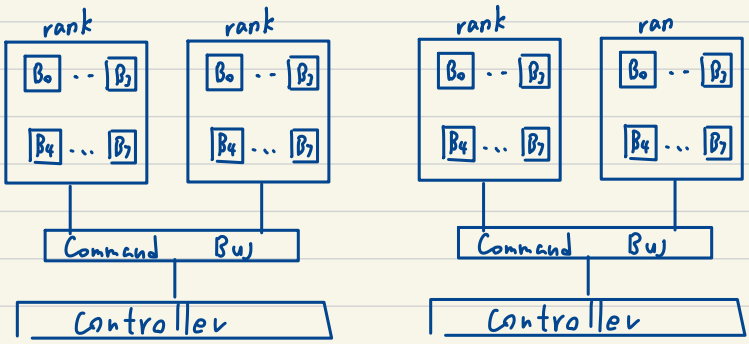


從 DRAM controller 著手, 給來自不同 core 的 request 依 core 編号
當有單一 core starvation 時, 利用 Aging 處理
逐步調高該 core 之該 request 之 priority

8. (15% total) A memory system is configured as:
   - There are two DRAM channels.
   - Each channel has two ranks of DRAM chips.
   - Each memory channel is controlled by a separate memory controller.
   - Each rank of DRAM contains eight banks.
   - Each bank contains 32K rows.
   - Each row in one bank is 8KB.

   Assume the minimum retention time among all DRAM rows in the system is 64 ms. In order to ensure that no data is lost, every DRAM row is refreshed once per 64 ms. Every DRAM row refresh is initiated by a command from memory controller which occupies the command bus on the associated memory channel. Consider a 1.024 second span of time.
   Define the *utilization* (of a resource such as a bus or a memory bank) as the fraction of total time for which a resource is occupied by a refresh command.
   For each calculation in this question, you may leave your answer in simplified form in terms of powers of 2 and powers of 10.

8.1 (4%) How long does each refresh command occupy the command bus (in ns) such that across all memory channels, the command bus utilization due to refreshes is 8.192%? (Hint $8.192 = 2^{13}/1000$)
8.2 (4%) How long does each refresh command occupy the DRAM banks (in ns) such that across all the banks, the banks utilization due to refreshes is 8.192%?
8.3 (3%) What data bus utilization, across all memory channels, is directly caused by DRAM refreshes?
8.4 (4%) How many refreshes are performed by the memory controllers during the 1.024 second period in total across both memory channels combined?



- 一个 Bank 有 32K rows ⇒ $32K \times 64\ ms = 2^{21}\ ms$

Refresh step: