ENEE 759H, Spring 2005
Memory Systems: Architecture and Performance Analysis

Introduction

Credit where credit is due:
Slides contain original artwork (© Jacob, Wang 2005)
Stuff that will be covered in this class

- DRAM Device Architecture
- Memory System Organization
- System Controller
- DRAM Access Protocol
- Performance Analysis
- Reliability (Error detection/correction)
- Signalling (Data transport/reception)
- New System Architecture (FB-DIMM)
- Non Volatile RAM (Flash)
- A lot of other stuff...
DRAM Device Architecture I

- Stacked capacitor
- Access transistor
- Wordline
- Bitline
- Poly 1
- Poly 2
- Poly 3
- Metal 1 bitline
- Sense amp array
- I/O gating
- Row select
- DRAM Array
- Wordlines @ V_{ref}
- Sense amp
- Bitlines @ V_{ref}
- Contact
Memory System Organization

DMC

Chip select 0
Chip select 1

Address and command

Data bus 16
Data bus 16
Data bus 16
Data bus 16

Rank ID = 1
Bank ID = 1
Row ID = 0x0B1D
Column ID = 0x187

Channel ?
Rank ?
Bank ?
Row ?
Column ?

Memory System
Heavy demand placed on memory system

Heavier still in SMP/SMT/CMP system

System Controller == System traffic cop
DRAM Access Protocol

\( n \neq m \)

\( i \neq j \)

\( \text{addr & cmd} \)

\( \text{Rank n} \)

\( \text{Rank m} \)

\( \text{data bus} \)

\( \text{decode} \)

\( \text{decode} \)

\( \text{cmd & addr} \)

\( \text{bank "i" of rank "m"} \)

\( \text{bank "j" of rank "n"} \)

\( \text{bank "i" open} \)

\( \text{bank "j" open} \)

\( \text{I/O gating} \)

\( \text{data burst} \)

\( \text{sync} \)

\( \text{data burst} \)

\( \text{t}_{\text{CAS}} \)

\( \text{t}_{\text{Burst}} \)

\( \text{t}_{\text{DQS}} \)

\( \text{t}_{\text{Burst}} \)

\( \text{t}_{\text{Burst}} + \text{t}_{\text{DQS}} \)

\( \text{time} \)
**Memory Request Overview**

**Part A: Searching on-chip for data**
- virtual to physical address translation (DTLB access) \([A_1]\)
- L1 D-Cache access. If miss then proceed to \([A_2]\)
- L2 Cache access. If miss then send to BIU \([A_3]\)

**Part B: Going off-chip for data**
- physical to memory address translation \([B_1]\)
- read data buffer \([B_2]\)
- memory request scheduling \([B_3]\)
- DRAM dev. obtains data and returns to controller \([B_6, B_7]\)
- system controller returns data to CPU \([B_8]\)

**System controller**
- BIU (Bus Interface Unit)
- arbitration \([B_2]\)
- request sent to system controller \([B_5]\)
- memory address translation \([B_3]\)
- memory request scheduling \([B_4]\)
- memory addr. Setup (RAS/CAS) \([B_5]\)

**Stages of instruction execution**
- Fetch
- Decode
- Exec
- Mem
- WB

**Progression of a Memory Read Transaction Request Through Memory System**

**Steps not required for some processor/system controllers, protocol dependant.**
“Memory Latency”

A: Transaction request may be delayed in Queue
B: Transaction request sent to Memory Controller
C: Transaction converted to Command Sequences (may be queued)
D: Command/s Sent to DRAM
E₁: Requires only a **CAS** or
E₂: Requires **RAS** + **CAS** or
E₃: Requires **PRE** + **RAS** + **CAS**
F: Transaction sent back to CPU

“DRAM Latency” = A + B + C + D + E + F
Performance Analysis

- tRC = 60ns, burst of eight, 8B wide channel

- 2 ranks of 8 banks perform same as 1 rank 16 banks if tRTRS = 0.
- tRTRS minutely impacts performance even with only 1 rank of memory due to R/W turnarounds.
- 2 ranks of 8 banks perform worse than 1 rank 16 banks if tRTRS = 3 clocks (6 beats).
- 2 ranks of 8 banks with separate tFAW limits on each rank does not alleviate tFAW impact much.

Maximum Sustainable Bandwidth: GB/s

Datarate - Mbits/sec

- 1 rank 16 banks
- 2 rank 8 banks

Performance gap between 1 rank of 16 banks and 2 ranks of 8 banks closes with higher data rates. For systems limited by tFAW, gap closes faster.
Reliability

light charged particle (α, p, e, etc.)
neutron
neutron
recoil nucleus

Si

\[ R = \{0\,1\,1\,0\,1\,0\,0\,1\,1\,1\,1\,0\} \quad \text{One bit error. Can we detect and correct?} \]

R0001 = R0011 + R1011 + R1001 + R1011 = 1 + 1 + 1 + 1 = 1
R0010 = R0011 + R0110 + R0111 + R1011 = 1 + 0 + 1 + 1 = 0
R0100 = R0101 + R0110 + R0111 + R1100 = 1 + 0 + 1 + 0 = 0
R1000 = R1001 + R1010 + R1011 + R1100 = 1 + 1 + 1 + 0 = 0

XOR old check bits against new check bits

<table>
<thead>
<tr>
<th>R0001</th>
<th>R0010</th>
<th>R0010</th>
<th>R0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \oplus \]

ECC Syndrome

Syndrome \(!=\) 0000 \quad \text{Bit position 11 is rotten}
Signalling

The Digital Fantasy

Pretend that the world looks like this

But...
DRAM Interface: Signals

FCRAM side

Controller side

VDDQ(Pad)

DQS (Pin)

DQ0-15 (Pin)

VSSQ(Pad)

skew=158psec

skew=102psec

*Toshiba Presentation, Denali MemCon 2002

*UNIVERSITY OF MARYLAND
Interface: Signal Propagation

Ideal Transmission Line

\[ \sim 0.66c = 20 \text{ cm/ns} \]

PC Board + Module Connectors + Varying Electrical Loads

= Rather non-Ideal Transmission Line
Interface: Clocking Issues

Figure 1: Sliding Time

Figure 2: H Tree?

What Kind of Clocking System?
Path Length Differential

High Frequency AND Wide Parallel Busses are Difficult to Implement
# Future Trends I

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi Generation (nm)</td>
<td>90</td>
<td>65</td>
<td>45</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>CPU MHz</td>
<td>3990</td>
<td>6740</td>
<td>12000</td>
<td>19000</td>
<td>29000</td>
</tr>
<tr>
<td>MLogicTransistors/cm^2</td>
<td>77.2</td>
<td>154.3</td>
<td>309</td>
<td>617</td>
<td>1235</td>
</tr>
<tr>
<td>High Perf chip pin count</td>
<td>2263</td>
<td>3012</td>
<td>4009</td>
<td>5335</td>
<td>7100</td>
</tr>
<tr>
<td>High Performance chip cost (cents/pin)</td>
<td>1.88</td>
<td>1.61</td>
<td>1.68</td>
<td>1.44</td>
<td>1.22</td>
</tr>
<tr>
<td>Memory pin cost (cents/pin)</td>
<td>0.34 -</td>
<td>0.27 -</td>
<td>0.22 -</td>
<td>0.19 -</td>
<td>0.19 -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.39</td>
<td>0.84</td>
<td>0.34</td>
<td>0.39</td>
</tr>
<tr>
<td>Memory pin count</td>
<td>48-160</td>
<td>48-160</td>
<td>62-208</td>
<td>81-270</td>
<td>105-351</td>
</tr>
</tbody>
</table>
Future Trends II

- **Direct Connect**
  - Custom DRAM: Highest Bandwidth + Low Latency
- **Indirect Connection**
  - Highest Bandwidth
  - Highest Latency
- **Inexpensive DRAM**
  - Lowest Bandwidth + Highest Latency
Research Areas: Topology

Unidirectional Topology:

- Write Packets sent on Command Bus
- Pins used for Command/Address/Data
- Further Increase of Logic on DRAM chips
Memory Commands?

Instead of A[ ] = 0; Do “write 0”

Why do A[ ] = B[ ] in CPU?

Move Data inside of DRAM or between DRAMs.

Why do STREAMadd in CPU?

A[ ] = B[ ] + C[ ]

Active Pages *(Chong et. al. ISCA ‘98)
- Grading
- Projects
- “Textbook”