MOTIVATION

Speculative Precharge and Activate
Use Prefetching techniques for speculation
OVERVIEW

Motivation
Scheduler
Memory Requirements
Experimental Results
Questions
OVERVIEW

Motivation

Scheduler

Memory Requirements

Experimental Results

Questions
SCHEDULER

Three components:

- BASE SCHEDULER
- CONSTANT STRIDE DETECTING OPEN/CLOSE PAGE PREDICTOR
- GLOBAL HISTORY- BASED CLOSE PAGE PREDICTOR

COMMAND ISSUING PRIORITY
SCHEDULER

Three components:

- BASE SCHEDULER
- CONSTANT STRIDE DETECTING OPEN/CLOSE PAGE PREDICTOR
- GLOBAL HISTORY- BASED CLOSE PAGE PREDICTOR

COMMAND ISSUING PRIORITY
BASE SCHEDULER

Read and Write requests served in bursts

• BUS switching delay

WRITE DRAIN MODE:
WRITE QUEUE LENGTH > 40
Until
WRITE QUEUE LENGTH < 20

READ MODE:
WRITE DRAIN MODE

FR-FCFS (write queue)

IF NONE ISSUABLE

PRE or ACT cmd from read queue

NO BANK CONFLICT WITH PENDING WRITE REQUESTS
READ MODE

FR-FCFS (read queue)

IF NONE ISSUABLE

PRE from write queue

NO BANK CONFLICT WITH PENDING READ REQUESTS
ISUSED DATA STRUCTURE

VALUES:

0:
BANK NOT TARGETTED BY ANY REQUEST

1:
BANK TARGETTED BY ATLEAST ONE READ REQUEST

2:
BANK TARGETTED BY ATLEAST ONE WRITE REQUEST

3:
BANK TARGETTED BY ATLEAST ONE WRITE & ONE READ REQUEST
SCHEDULER

Three components:

- BASE SCHEDULER
- CONSTANT STRIDE DETECTING OPEN/CLOSE PAGE PREDICTOR
- GLOBAL HISTORY-BASED CLOSE PAGE PREDICTOR

COMMAND ISSUING PRIORITY
CONSTANT STRIDE
DETECTING OPEN/CLOSE
PAGE PREDICTOR

Instruction PC % 1024

Stride Table

<table>
<thead>
<tr>
<th>instruction pc%1024</th>
<th>last stride</th>
<th>prev address</th>
<th>detected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 bytes</td>
<td>8 bytes</td>
<td>1 bit</td>
</tr>
</tbody>
</table>
CONSTANT STRIDE
DETECTING OPEN/CLOSE
PAGE PREDICTOR

Address, a

| a + s |
| a + 2s |
| a + 3s |
| a + 4s |
| ... |
| a + ds |

If different channel

Optimal Depth : 7

ToBeIssued

PRE

OR

ACT
SCHEDULER

Three components:

- BASE SCHEDULER
- CONSTANT STRIDE DETECTING OPEN/CLOSE PAGE PREDICTOR
- GLOBAL HISTORY-BASED CLOSE PAGE PREDICTOR

COMMAND ISSUING PRIORITY
GLOBAL HISTORY-BASED CLOSE PAGE PREDICTOR

<table>
<thead>
<tr>
<th>Thread id</th>
<th>Instruction PC XOR Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 bits</td>
<td>8 bits</td>
</tr>
</tbody>
</table>

Index Table: 1024 entries
GHB: 512 entries

GHB:

A
I
R
C
G
T
F
A
I
R
C
G
T
F
A

DEPTH

BREADTH

head ptr

FIFO

link

Index Table:

A
B
C
D

Link
GLOBAL HISTORY- BASED CLOSE PAGE PREDICTOR

Optimal Configuration:

- Breadth : 3
- Depth : 20

Address, a
OVERVIEW

Motivation
Scheduler

Memory Requirements
Experimental Results
Questions
## MEMORY REQUIREMENTS

### Memory Requirements for Scheduler

<table>
<thead>
<tr>
<th>Entity</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isused</td>
<td>$16 \times 16 \times 32 \times 2B = 16\text{KB}$</td>
</tr>
<tr>
<td>Stride Table</td>
<td>$1024 \times (12B + 1b) = (12\text{K} + 128)\text{B}$</td>
</tr>
<tr>
<td>ToBeIssued</td>
<td>$16 \times 20B = 320B$</td>
</tr>
<tr>
<td>GHB</td>
<td>$512 \times 32B = 16\text{KB}$</td>
</tr>
<tr>
<td>Index Table</td>
<td>$1024 \times 8B = 8\text{KB}$</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>$(52\text{K} + 448)\text{B}$</td>
</tr>
</tbody>
</table>
OVERVIEW

Motivation
Scheduler
Memory Requirements

Experimental Results

Questions
## EXPERIMENTAL RESULTS

### CONFIGS USED

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1channel.cfg</th>
<th>4channel.cfg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor clock speed</td>
<td>3.2 GHz</td>
<td>3.2 GHz</td>
</tr>
<tr>
<td>Processor ROB size</td>
<td>128</td>
<td>160</td>
</tr>
<tr>
<td>Processor retire width</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Processor fetch width</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Processor pipeline depth</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Memory bus speed</td>
<td>800 MHz (plus DDR)</td>
<td>800 MHz (plus DDR)</td>
</tr>
<tr>
<td>DDR3 Memory channels</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Ranks per channel</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Banks per rank</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Rows per bank</td>
<td>32768 × NUMCORES</td>
<td>32768 × NUMCORES</td>
</tr>
<tr>
<td>Columns (cache lines) per row</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>Cache line size</td>
<td>64 B</td>
<td>64 B</td>
</tr>
<tr>
<td>Address bits (function of above params)</td>
<td>32 + log(NUMCORES)</td>
<td>34 + log(NUMCORES)</td>
</tr>
<tr>
<td>Write queue capacity</td>
<td>64</td>
<td>96</td>
</tr>
<tr>
<td>Write queue bypass latency</td>
<td>10 cpu cycles</td>
<td>10 cpu cycles</td>
</tr>
</tbody>
</table>
## WORKLOADS

<table>
<thead>
<tr>
<th>No.</th>
<th>Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MTc</td>
</tr>
<tr>
<td>2</td>
<td>bl-bl-fr-fr</td>
</tr>
<tr>
<td>3</td>
<td>c1-c1</td>
</tr>
<tr>
<td>4</td>
<td>c1-c1-c2-c2</td>
</tr>
<tr>
<td>5</td>
<td>c2</td>
</tr>
<tr>
<td>6</td>
<td>fa-fa-fe-fe</td>
</tr>
<tr>
<td>7</td>
<td>fl-sw-c2-c2</td>
</tr>
<tr>
<td>8</td>
<td>st-st-st-st</td>
</tr>
<tr>
<td>9</td>
<td>fl-fl-sw-sw-c2-c2-fe-fe</td>
</tr>
<tr>
<td>10</td>
<td>fl-fl-sw-sw-c2-c2-fe-fe-</td>
</tr>
<tr>
<td></td>
<td>bl-bl-fr-fr-c1-c1-st-st</td>
</tr>
</tbody>
</table>

### BENCHMARK TRACES

<table>
<thead>
<tr>
<th>Trace</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARSECs blackscholes</td>
<td>bl</td>
</tr>
<tr>
<td>PARSECs facesim</td>
<td>fa</td>
</tr>
<tr>
<td>PARSECs ferret</td>
<td>fe</td>
</tr>
<tr>
<td>PARSECs fluidanimate</td>
<td>fl</td>
</tr>
<tr>
<td>PARSECs freqmne</td>
<td>fr</td>
</tr>
<tr>
<td>PARSECs stream</td>
<td>st</td>
</tr>
<tr>
<td>PARSECs swaptions</td>
<td>sw</td>
</tr>
<tr>
<td>Server-class transaction processing workload -1</td>
<td>c1</td>
</tr>
<tr>
<td>Server-class transaction processing workload -2</td>
<td>c2</td>
</tr>
<tr>
<td>PARSECs canneal (4 threads)</td>
<td>MTc</td>
</tr>
</tbody>
</table>
**METRIC USED**

Speculation Accuracy

\[
Speculation\ Accuracy = \frac{Accurate\ Speculation}{Total\ Speculation} \times 100
\]

**PRE**

Next Request to bank
Is a different row than That closed

**ACT**

Next Request to bank
Is a ROW BUFFER HIT
METRIC USED

Speculation Coverage

\[ \text{Speculation Coverage} = \frac{\text{Accurate Speculation}}{\text{Total ACTs and PREs}} \times 100 \]
METRIC USED

% Decrease in Execution Time (compared to FR-FCFS)

\[ \% \text{Decrease in Exec. Time} = \frac{E_{FR-FCFS} - E}{E_{FR-FCFS}} \times 100 \]
Average: 3.97%
THANKS

Prof. Mikko Lipasti
QUESTIONS