Reconciling Repeatable Timing with Pipelining and Memory Hierarchy

Stephen A. Edwards, Sungjun Kim (Columbia), Edward A. Lee (UC Berkeley), Hiren D. Patel (UC Berkeley Waterloo), and Martin Schoeberl (TU Vienna)

October 15, 2009
1963 Chrysler Turbine Ghia

Ran on any fuel: gas, diesel, scotch, Chanel #5, tequila
The Disadvantage of Turbine Cars: Acceleration

*When* can be just as important as *what*.
Great Except for Hazards

Forwarding Can Reduce the Need to Stall...

...But It Does Not Solve Everything...

LD R1, 45(r2)  
DADD R5, R1, R7  
BE R5, R3, R0  
ST R5, 48(R2)

Unpipelined  

The Dream  

The Reality  

Memory Hazard  
Data Hazard  
Branch Hazard
Motorola Coldfire pipeline from Ferdinand et al., Reliable and precise WCET determination for a real-life processor, EMSOFT 2001
Our Solution: Thread-Interleaved Pipelines

An old idea from the 1960s

T1: FDEMW FDEMW
T2: FDEMW FDEMW
T3: FDEMW FDEMW
T4: FDEMW FDEMW
T5: FDEMW FDEMW

But what about memory?

Memory Hierarchy is an Old Idea

Ideally one would desire an indefinitely large memory capacity such that any particular ... word ... would be immediately available. We ... recognize the possibility of constructing a hierarchy of memories, each of which has greater capacity than the preceding but which is less quickly accessible.

Memory System Design 101

DRAM Circuit Operation

- Bit Cells
- Sense Amplifier
- Equalization Circuit
**DRAM Circuit Operation**

- **Precharge**
  - Set bit lines to $V_{cc}/2$

![Diagram showing DRAM circuit operation with precharge and set bit lines to $V_{cc}/2$]
DRAM Circuit Operation

- **Precharge**
  Set bit lines to $V_{cc}/2$

- **Access**
  Select a row; perturb bit lines
DRAM Circuit Operation

- **Precharge**
  Set bit lines to $V_{cc}/2$

- **Access**
  Select a row; perturb bit lines

- **Sense**
  Enable sense amplifier
  Drive bit lines to rails
DRAM Circuit Operation

**Precharge**
Set bit lines to $V_{cc}/2$

**Access**
Select a row; perturb bit lines

**Sense**
Enable sense amplifier
Drive bit lines to rails

**Restore**
Disable sense amplifier
DRAM Circuit Operation

- **Precharge**
  - Set bit lines to $V_{cc}/2$

- **Access**
  - Select a row; perturb bit lines

- **Sense**
  - Enable sense amplifier
  - Drive bit lines to rails

- **Restore**
  - Disable sense amplifier

\[ \begin{array}{c}
V_{cc} \\
V_{cc}/2 \\
0 \\
\end{array} \]
Modern DRAMs Have Banks

DDR2: 4–8 pipelined banks
DDR3: 8+ pipelined banks
Banks Enable Pipelining

B1: ASRPAASRPAASRP
B2: ASRPAASRPAASRP
B3: ASRPAASRPAASRP
B4: ASRPAASRPAASRP

Not to scale!
Our Solution: Pipeline-Synchronized Memory

- One per
- Tune processor frequency to memory
Pipeline-Synchronized Memory

T1: FDEMFDEMFDEM
B1: AASSRRPPPAASSRRPP
T2: FDEMFDEMFDEM
B2: AASSRRPPPAASSRRPP
T3: FDEMFDEMFDEM
B3: AASSRRPPPAASSRRPP
T4: FDEMFDEMFDEM
B3: AASSRRPPPAASSRRPP
Pipeline-Synchronized Refinements

- Add scratchpad memories
- Use round-robin scheduling for sharing banks
- Add ranks and DIMMs for more parallelism
- Get easier-to-pipeline memory
Conclusions

- PRET goal: predictable, temporally isolated threads
- Thread-interleaved pipelines avoid hazards
- Pipelined DRAMs allow elegant sharing of resources
- Pipeline-synchronized memory hierarchy
- Working on an FPGA prototype