Machine Learning for Thread-Level Speculation on Multi-core Architectures

Nikolas Ioannou

December 02, 2008

School of informatics

Chip-Multiprocessors

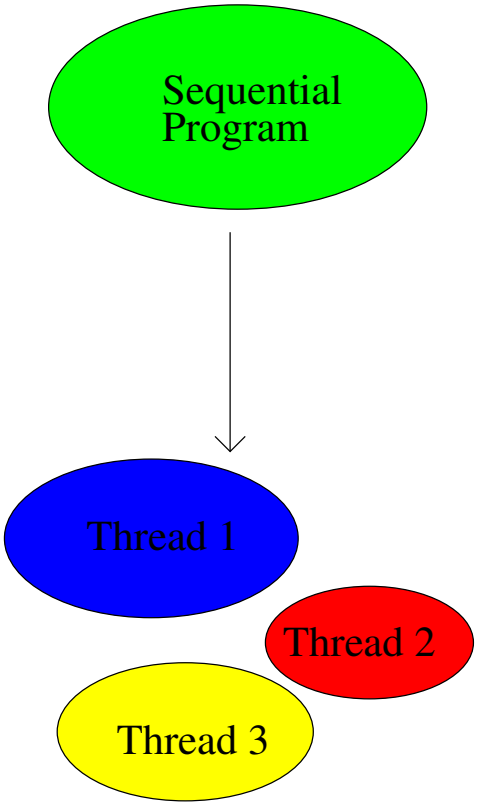
- Irreversible turn towards multi-cores
- Currently 2,4,8 and 16 cores per chip. Intel claims hundreds of cores in the near future

However:

- Imposes a major challenge for programmers
- Huge amount of *sequential* legacy code that does not benefit from multi-cores
- Unfortunately, auto-parallelizing compilers still do not work very well

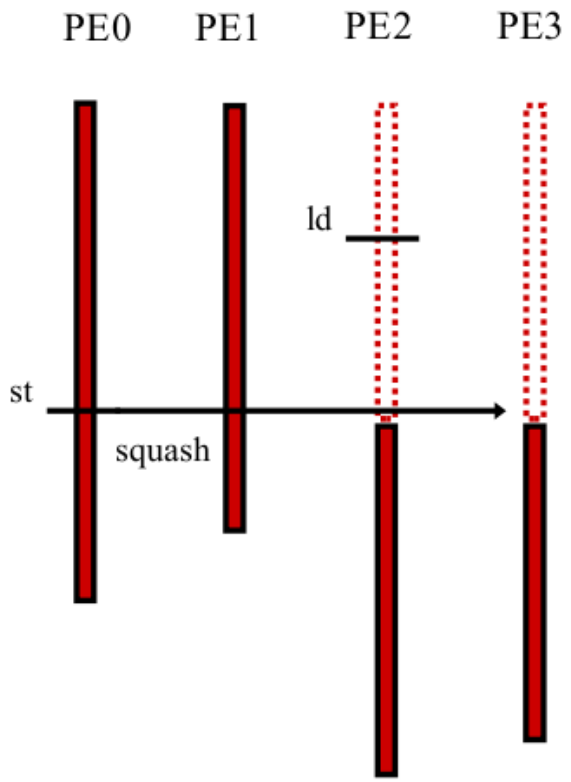
Thread-Level Speculation

- An approach to automatically parallelize sequential applications
- Assumes that two sections of a software application are parallelizable and tries to execute them concurrently
- If data conflicts occur, the underlying hardware enforces correct execution by re-executing the offending threads



Inefficiencies of Thread-Level Speculation

- Squash & restart: re-executing the threads
- Inter-thread communication: waiting for value from predecessor thread
- Load imbalance: processor waiting for thread to become non-speculative to commit



Machine Learning Techniques

Despite many years of research TLS is still inefficient. The effort of this project is to formulate the TLS inefficiencies as ML problems.

- Squash Prediction
  - Find the correlation between features of threads and their likelihood of being squashed
  - A feature selection problem in ML terms
- Inter-thread communication
  - Predict the value of a variable accessed by one thread while being computed by another
  - Learning/Prediction overhead must not negate the benefits of ML

Machine Learning Techniques

- Load Imbalance
  - Load imbalance can be minimized by *correctly* splitting the sequential program to tasks
  - Compiler has to be smart enough to find the optimal splitting based on static features

Current Work

- Develop fully working TLS compiler research prototypes (e.g. extending GCC), integrating ML predictors
- Investigate on which ML techniques can be beneficial for TLS systems
- Develop architectural models of CMP systems that support TLS and implement the ML predictors

Thank you!