CS 152 Computer Architecture and Engineering

Lecture 18: Multithreading

Krste Asanovic
Electrical Engineering and Computer Sciences
University of California, Berkeley

http://www.eecs.berkeley.edu/~krste
http://inst.cs.berkeley.edu/~cs152

Last Time: Vector Computers

- Vectors provide efficient execution of data-parallel loop codes
- Vector ISA provides compact encoding of machine parallelism
- ISAs scale to more lanes without changing binary code
- Vector registers provide fast temporary storage to reduce memory bandwidth demands, & simplify dependence checking between vector instructions
- Scatter/gather, masking, compress/expand operations increase set of vectorizable loops
- Requires extensive compiler analysis (or programmer annotation) to be certain that loops can be vectorized
- Full "long" vector support still only in supercomputers (NEC SX9, Cray X1E); microprocessors have limited "short" vector operations
  - Intel x86 MMX/SSE/AVX
  - IBM/Motorola PowerPC VMX/Altivec

Multithreading

- Difficult to continue to extract instruction-level parallelism (ILP) or data-level parallelism (DLP) from a single sequential thread of control
- Many workloads can make use of thread-level parallelism (TLP)
  - TLP from multiprogramming (run independent sequential jobs)
  - TLP from multithreaded applications (run one job faster using parallel threads)
- Multithreading uses TLP to improve utilization of a single processor