Shared History Instruction Fetch for Lean-Core Server Processors

Cansu Kaynak
EPFL

Boris Grot
University of Edinburgh

Babak Falsafi
EPFL

Instruction Fetch Bottleneck in Server Workloads

Intuition: Instruction access sequences repeat

Sources of discontinuities:
• Branches
• Function calls
• Library calls

Problem: Prohibitively large history storage

Integrating Instruction History into Modern Servers

A Few Fat Cores

high parallelism
Low power

Many Lean Cores

Cores run same application
⇒ Same history
90% of L1-I accesses to common sequences

Result Highlights

Performance
• Almost as good as per-core history!
• 18% speedup (avg)

Storage
• 14X reduction (for 16-core server)

Goal
• Eliminate history replication

Design
• One core records,
all cores replay history

Outcome
• Per-core history performance,
at cost of one history