Performance Analysis of the Compaq ES40--An Overview

- Paper evaluates Compaq's ES40 system, based on the Alpha 21264
- Only concern is performance: no power or cost issues
- Compares the ES40 and systems from other vendors in terms of speed, throughput, and sustained memory bandwidth
- Primary focus of paper is on comparison of ES40 against previous-generation AlphaServer 4100 to explore the improvements in the new system
  - Larger L1 cache
  - Tournament branch predictor
  - Out-of-order execution
  - Larger number of execution pipelines
  - More advanced memory system

Methodology

- On-line transaction processing and technical/scientific workloads
  - TPC-C: transaction processing benchmark
    - Measures throughput (transactions per minute)
    - Exercises processor power, memory interface, and I/O
  - SPEC95 benchmark suite
    - SPECint95/SPECfp95 measure speed; SPECint_rate95/SPECfp_rate95 measure throughput
    - Exercises processor, memory hierarchy, and compiler performance (to be included in the suite, can not stress I/O or include networking or graphics)
    - Measures geometric mean of normalized execution times, which can present a skewed perspective
  - Linpack: solves linear equations and least-squares problems
    - Paper uses benchmark once to get a MFLOPS rating
  - STREAM: synthetic benchmark program that includes four vector kernels
    - Measures sustainable memory bandwidth (in MB/s)
- ProfileMe and DCPI tools used to profile data
- Paper does not specify compiler used

I-Cache, Branch Prediction, and IPC Performance

- Instruction cache performance
  - 21164 has 8 KB direct-mapped L1 cache, while 21264 has 64 KB two-way associative L1 cache
  - Most cache misses in 21264 appear to be compulsory misses for the given benchmarks; difficult to tell how the cache would handle larger programs
- Branch prediction performance
  - 21164 uses a simple two-bit branch predictor; 21264 uses a tournament branch predictor that tracks local and global history
  - 21264 typically has improved branch prediction over the 21164, but it does not do as well with transaction processing branch prediction
- Comparison of Instructions Per Cycle
  - Out-of-order execution, up to 80 in-flight instructions, additional functional units, and other advances lead to IPC improvements in the 21264 over the 21164
  - IPC comparison could be skewed: IPC in 21264 measures number of retired instructions, which favors the 21164, but the ES40 results were obtained with code that targeted 21264, which favors that system

Instruction Retire and Memory System Performance

- Non-retired instructions
  - A large percentage of issued instructions in the 21264 are killed due to speculation
  - Little performance down-side due to speculative execution
- Memory bandwidth and latency
  - STREAM Copy benchmark shows big improvement in memory bandwidth in the 21264 over the 21164, as well as better scaling, due to several factors
    - Larger, more associative L1 cache
    - Crossbar interconnect
    - New WH64 prefetch
  - 21264 has impressive advantage (4x) over 21164 for “independent-load” memory latency, but only 40% improvement for “dependent-load” memory latency.
- Memory replay traps
  - 21264 has significantly fewer memory traps than 21164, due to more highly-parallel memory system
  - Further reduction in number of memory traps can be obtained through compiler modification

Conclusions

- The 21264-based Compaq ES40 has many architectural features that improve performance over the 21164-based AlphaServer 4100
- Authors claim that the ES40 has performance advantage over other vendor machines based on results of SPEC95 and STREAM benchmarks
- It is acknowledged in the paper that further study is required using a greater variety of applications--SPEC95 results are generally impressive, but TPC results show that there is room for improvement
- Paper does a good job of examining the architectural features of the ES40 and their effect on performance
- Other areas of interest--power consumption, area, cost