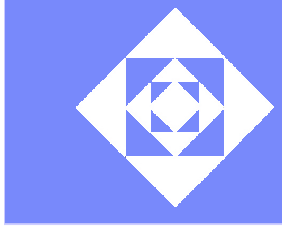


IBM

Compiler and Architecture Seminar, November 11th, 2003

Subsetting SPEC When Measuring Results: Research vs. Industry

Daniel Citron

citron@il.ibm.com

IBM Labs in Haifa

Production History

ISCA 2003 Panel, June 9th, 2003, San Diego, CA :

Subsetting SPEC when measuring results: valid or manipulative?

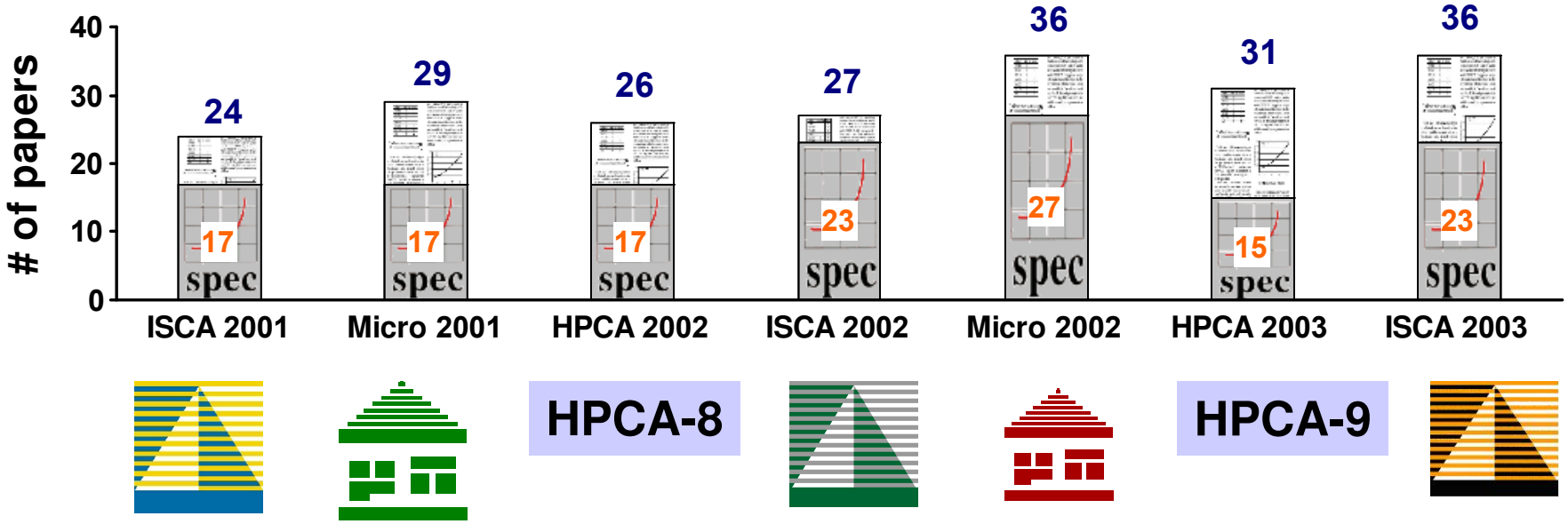
Mis ulation: Partial and Misleading Use of SPEC CPU 2000 in Computer Architecture Conferences

IEEE micro IEEE Micro, July/August 2003 (Vol. 23, No. 4)

The Use and Abuse of SPEC: An ISCA Panel

Revenge of the SPEC

Research: Computer Architecture Conferences

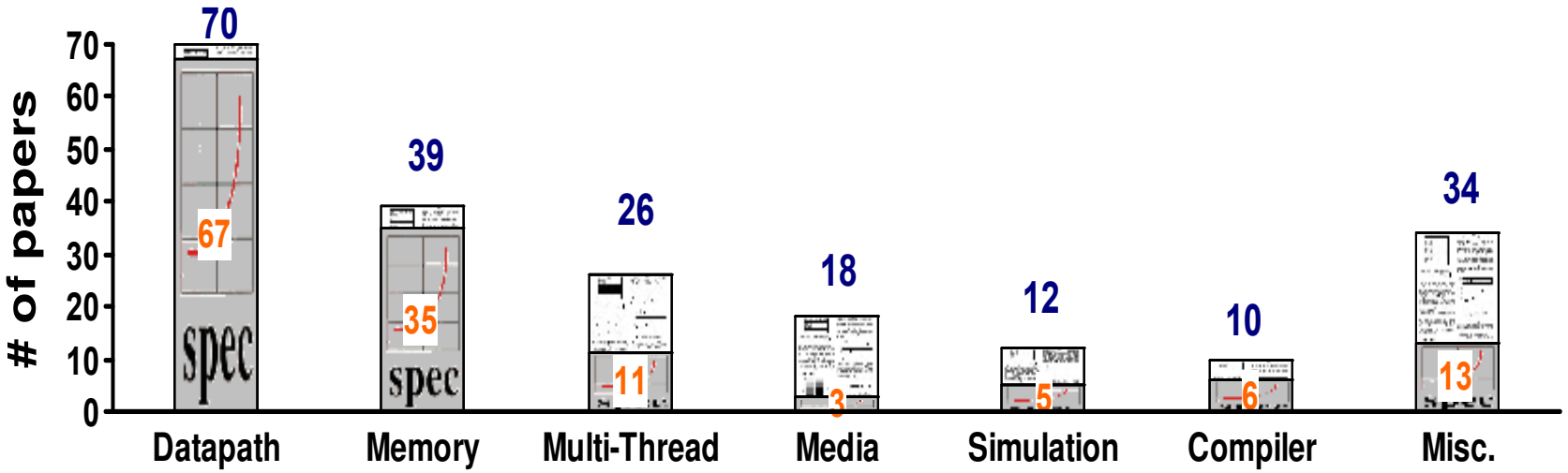


- Number of papers published: **209**
- Papers that used a version of SPEC: **140 (66%)**
- Earliest conference deadline: **December 2000**
- SPEC CPU2000 announced: **December 1999**

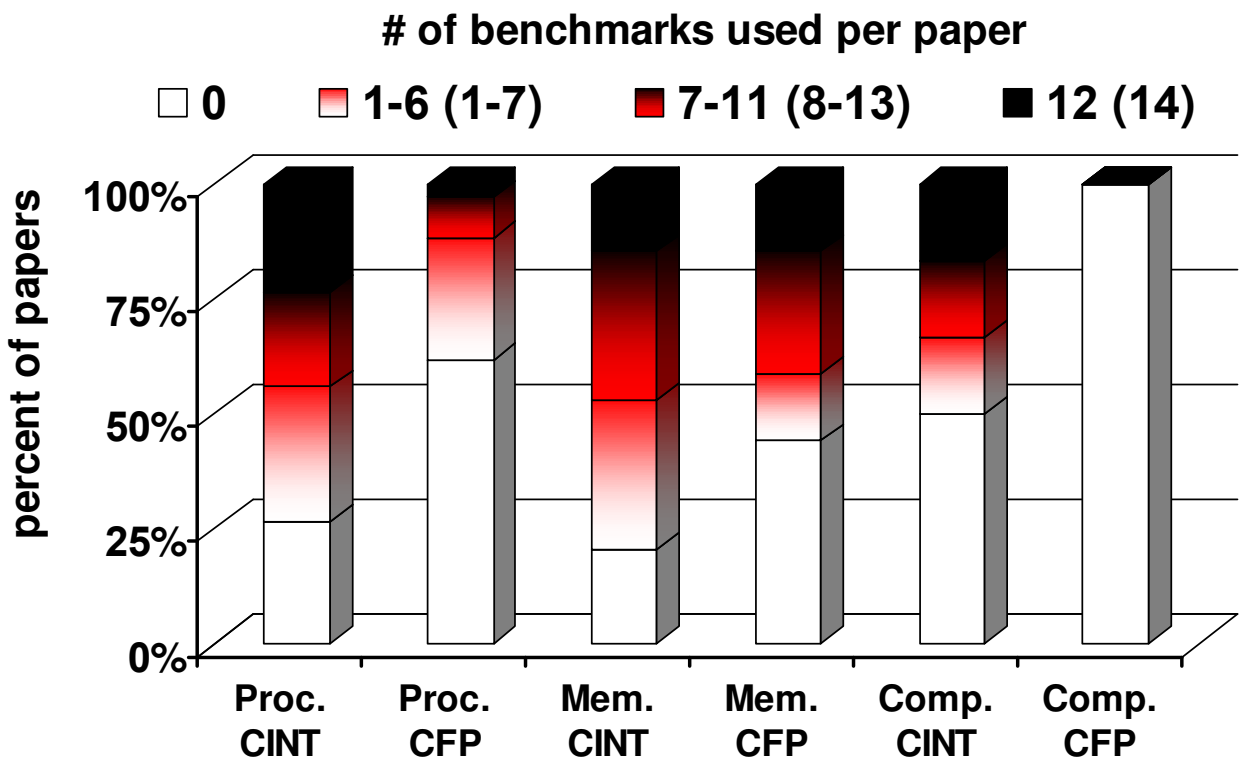
Industry: SPEC CPU2000

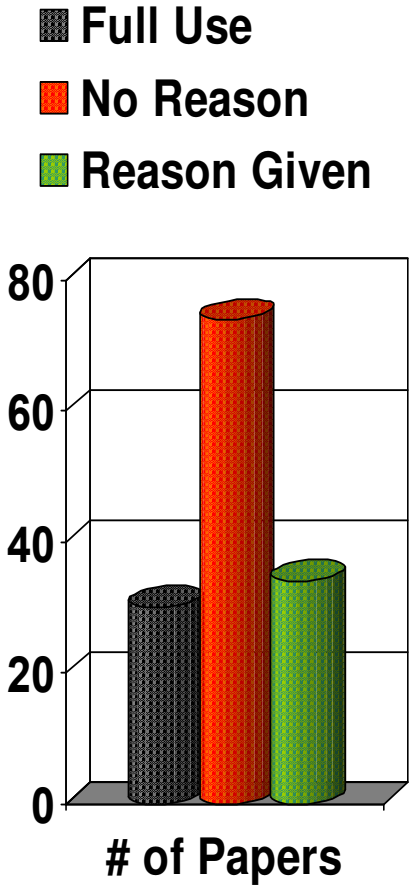
Quotes from www.spec.org

- "SPEC CPU2000 is the next-generation industry-standardized CPU-intensive benchmark suite."
- "These benchmarks measure the performance of the processor, memory and compiler on the tested system."



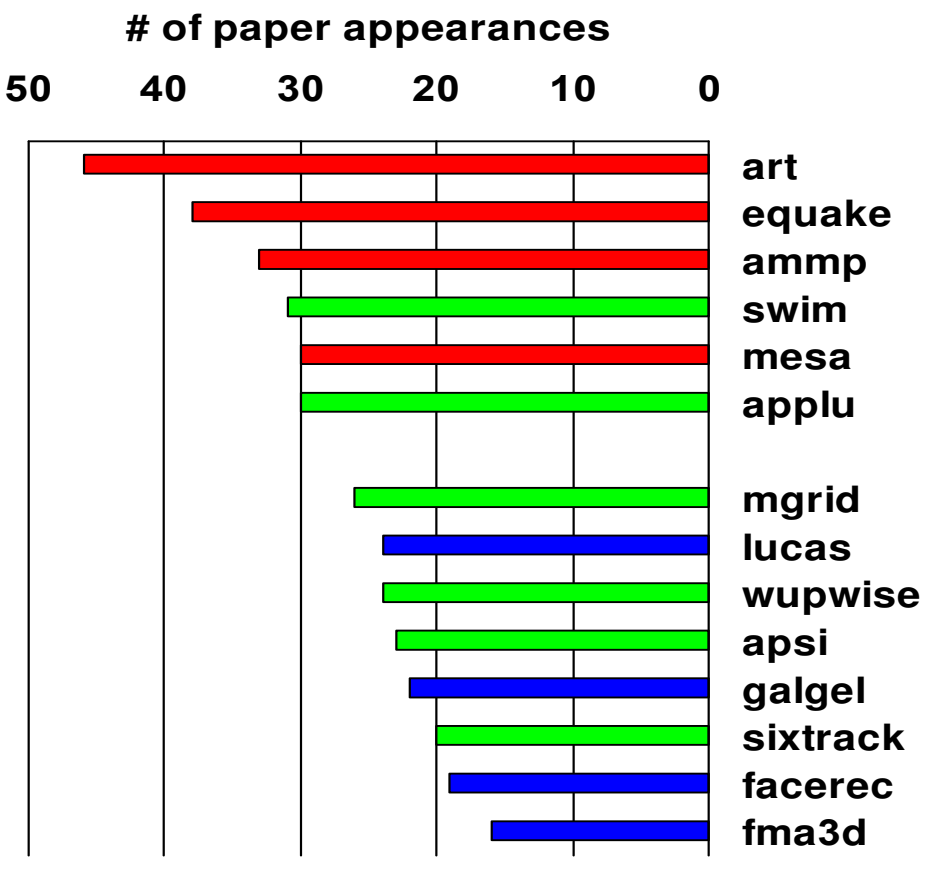
Research: Partial Use of CPU2000

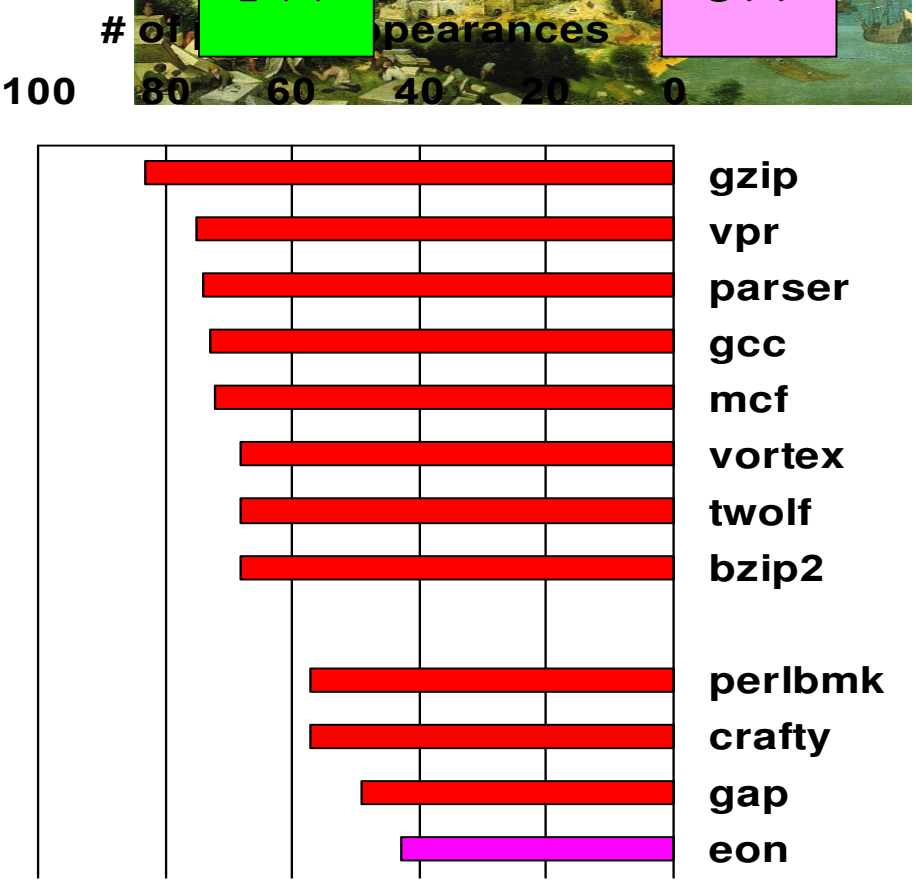




Authors omit benchmarks, suites, and explanations!

Research: The Tower of Babel

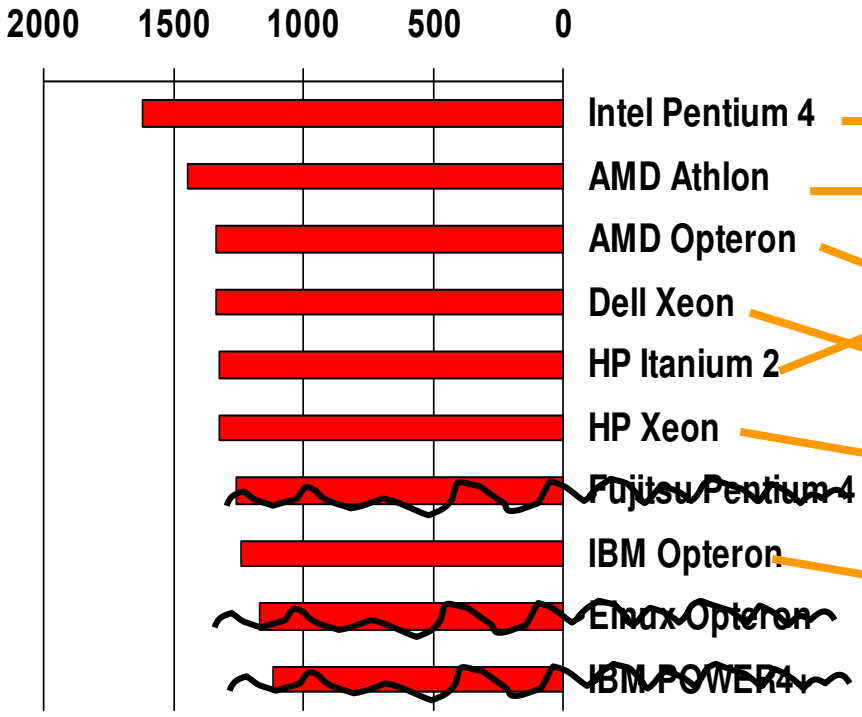




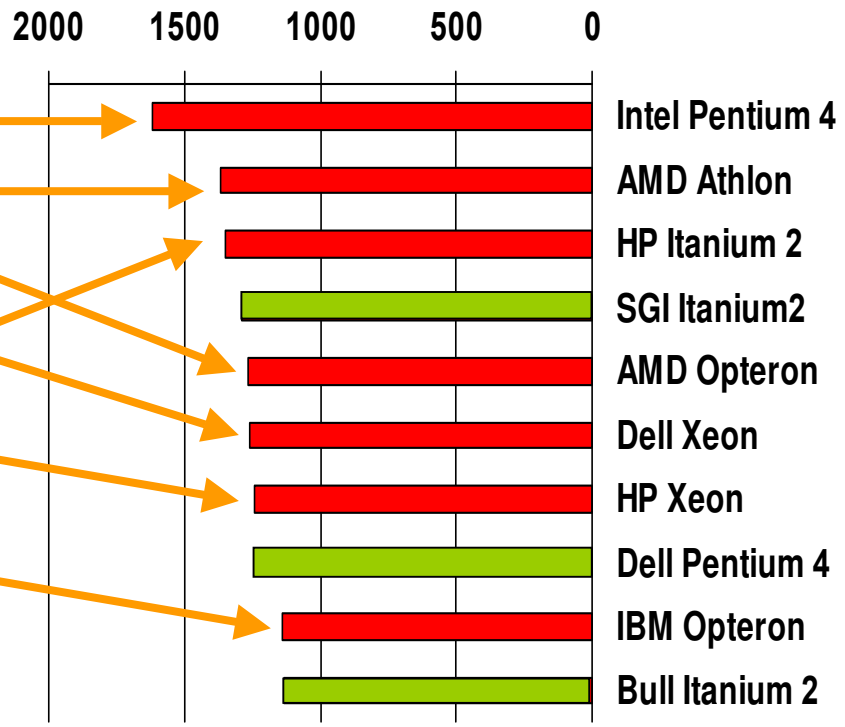
Average of 8 (CINT) and 6 (CFP) Benchmarks per Paper!

Research vs. Industry: Top 10 Systems (CINT)

Top Industry SPECint ratios (12/12)



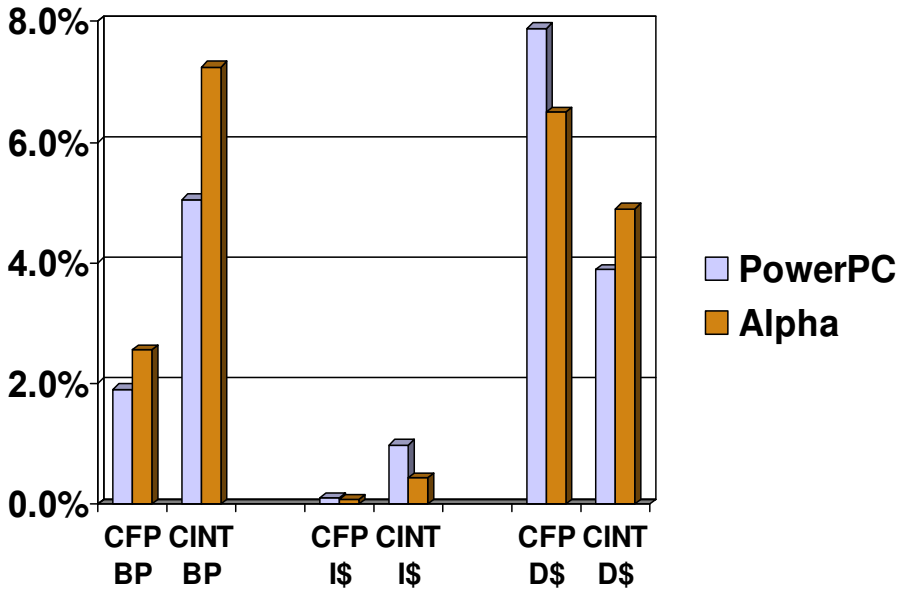
Top Research SPECint ratios (8/12)

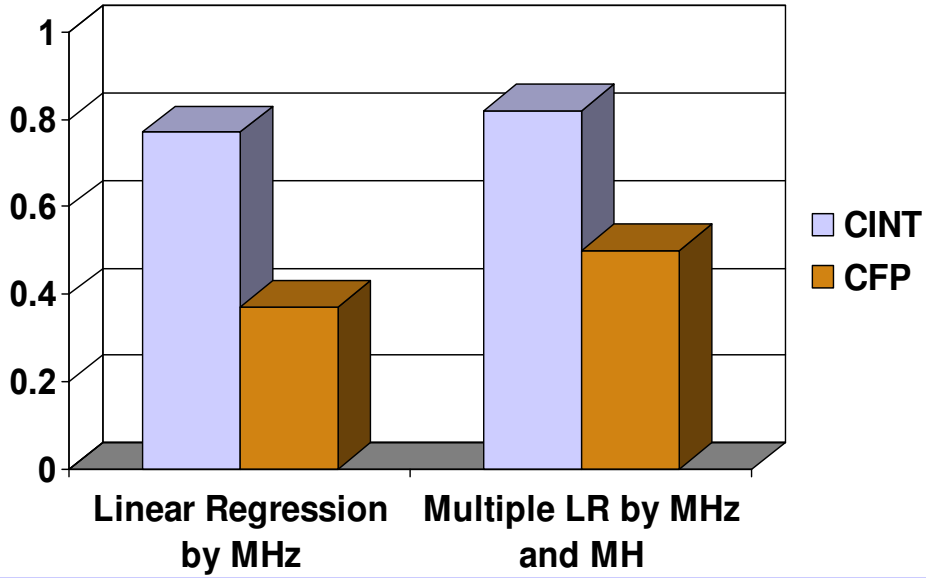


Research vs. Industry: CFP2000

- 105 papers used CINT2000, 63 used CFP2000 (60%)
- CFP2000 considered more regular and predictable
- CFP2000 has higher Dcache miss rate
- Only 35% of Memory-Hierarchy papers use a majority of CFP2000

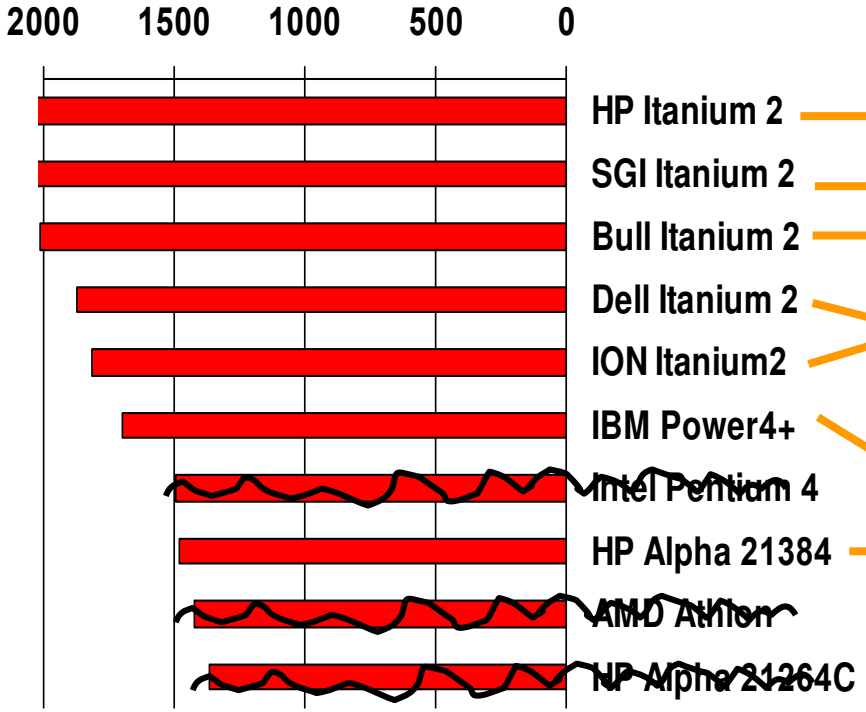
- 496 CINT2000 and 488 CFP2000 results published by SPEC
- Linear Regression model of SPECratio by the CPU clock rate
- Multiple Linear Regression model of SPECratio by speed and Memory-Hierarchy



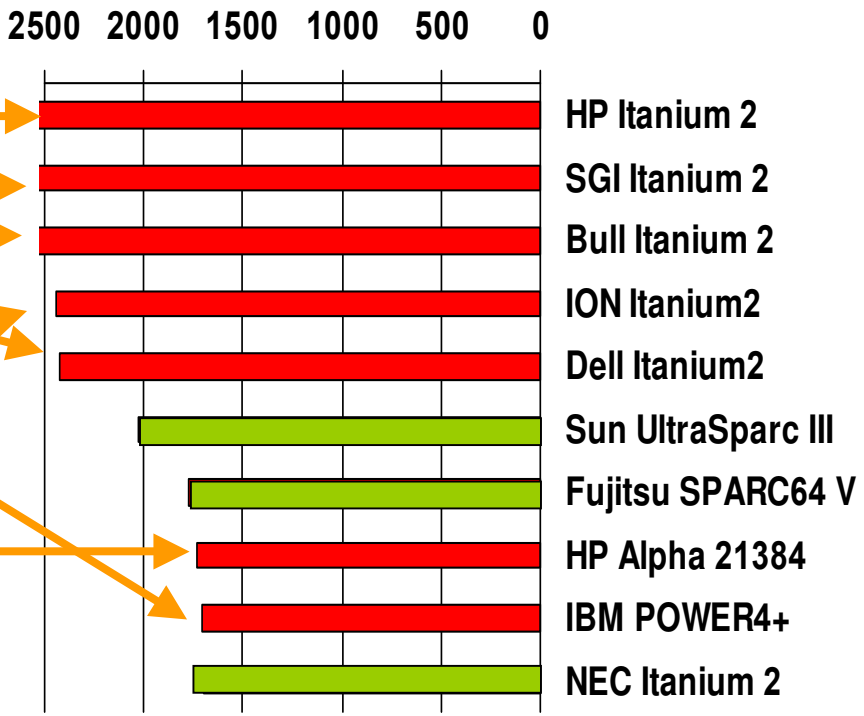


Research vs. Industry: Top 10 Systems (CFP)

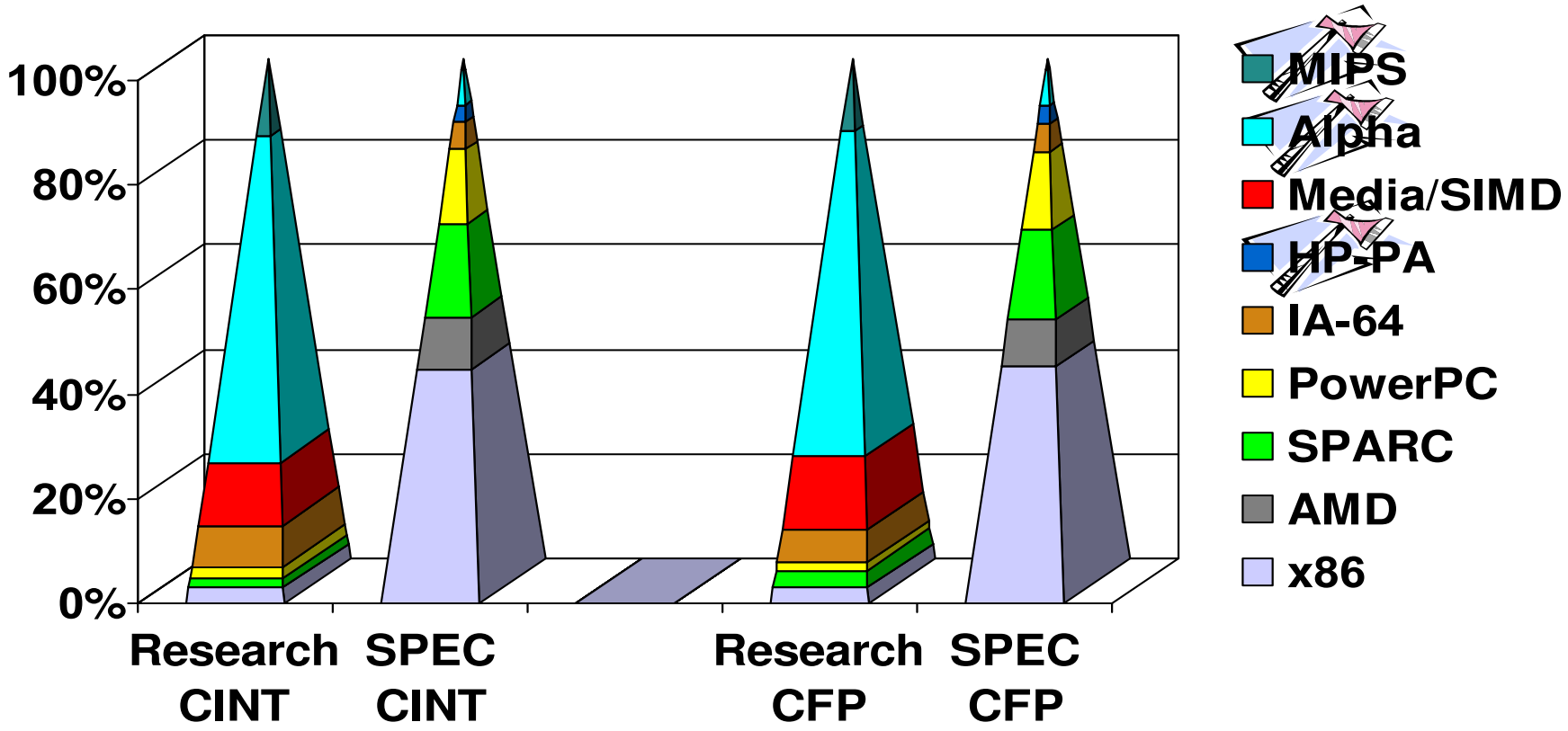
Top Industry SPECfp ratios (14/14)



Top Research SPECfp ratios (6/14)



Research vs. Industry: Processor Architecture



Use Architectures of the Future, not the Past!

SimpleScalar – A Two-Edged Sword

Research

- C[illegible]t of base standards for incremental research:
  - [illegible]arks: CP[illegible]4, Splash-3, MediaBench-2
  - Input sets: Fixed num[illegible]nstructions, MinneS[illegible]piler flags
  - Processor attributes: latencies, c[illegible]e sizes, bran[illegible]tion, …
  - [illegible]lation tools: SimpleScalar, Tim[illegible]ET, EXPRESS, …
  - Av[illegible] methods: Geometric, harmo[illegible]arithmetic, just pick one
- Innovate and disregard th[illegible]t above

Industry

SP[illegible] upport research with simulation [illegible]nput sets

Vendors: Get simulation tools out to the field