

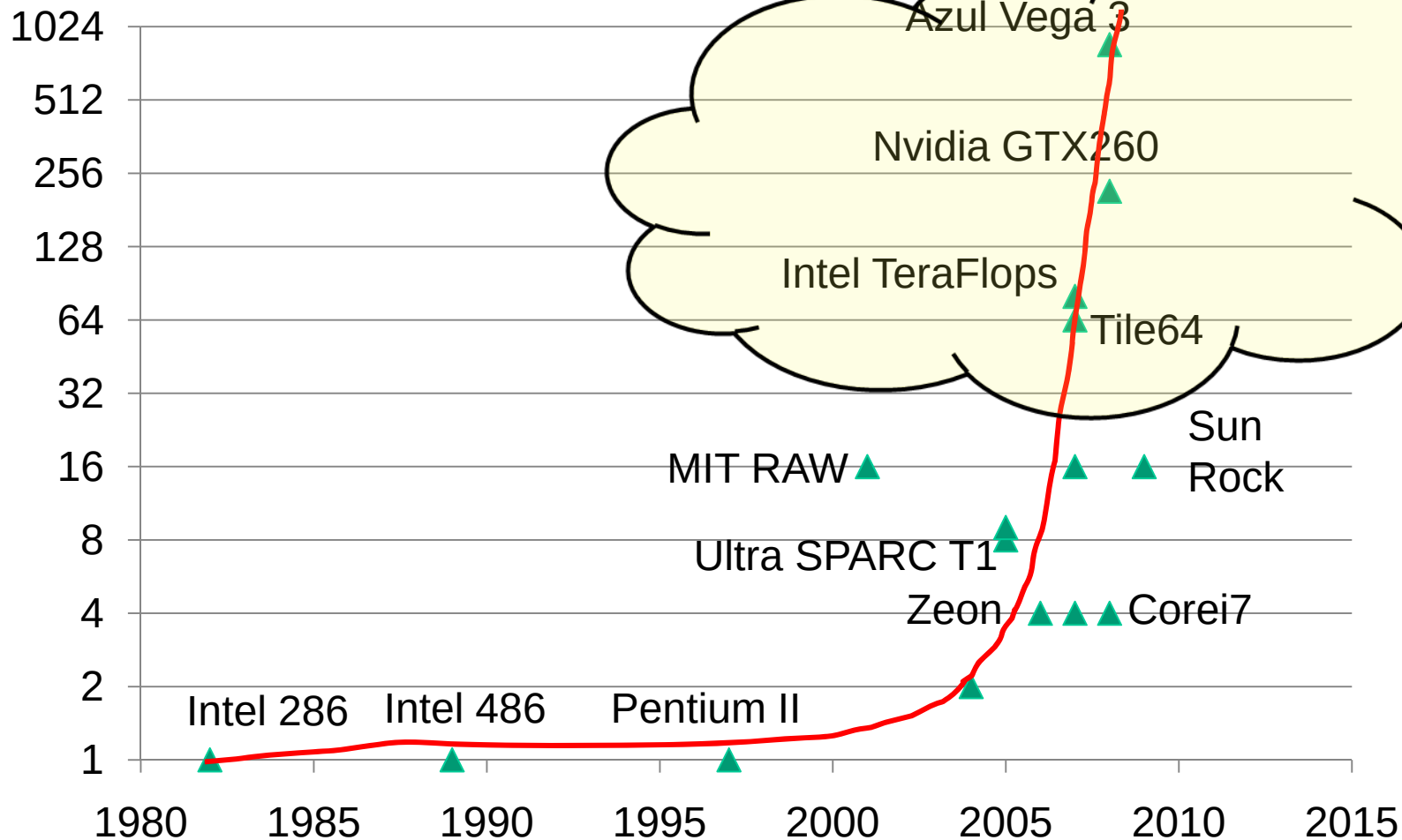
Multi-Execution: Multicore Caching for Data-Similar Executions

**Susmit Biswas, Diana Franklin, Alan Savage, Ryan
Dixon, Timothy Sherwood, Frederic T Chong**

susmit@cs.ucsb.edu

Manycore Era

Evolution of Manycore



Manycore Era

How to use so many cores?

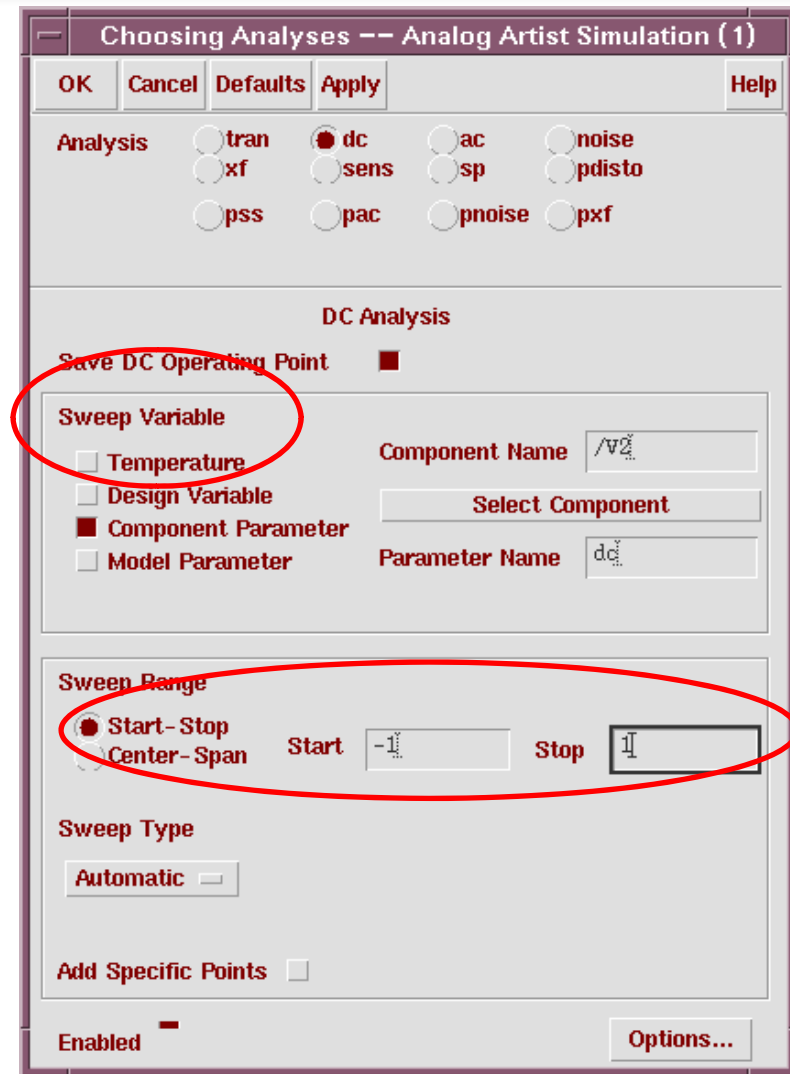
Transactional memory [Herlihy-ISCA93]

Thread level speculation [Steffan-ISCA00]

...

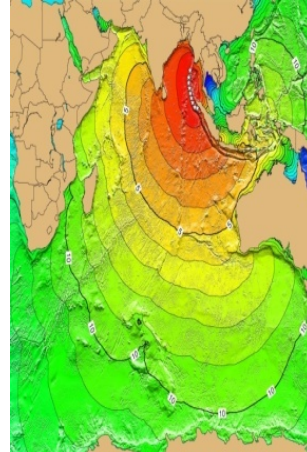
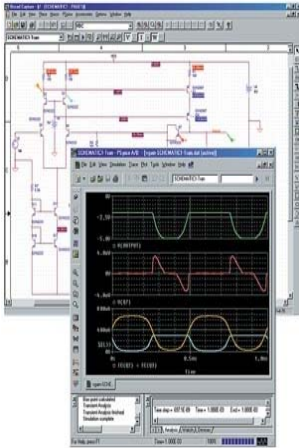
How to use so many cores?

- “Multi-execution”
- Run multiple instances
 - Serial code
 - Input variation

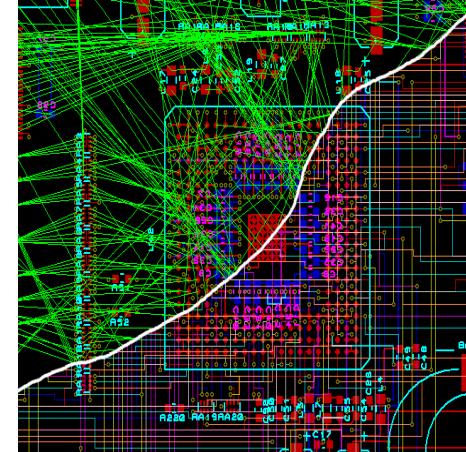


Sic3a Analog Circuit Simulator

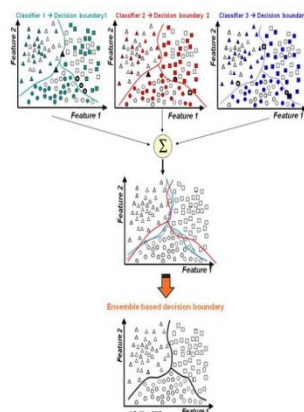
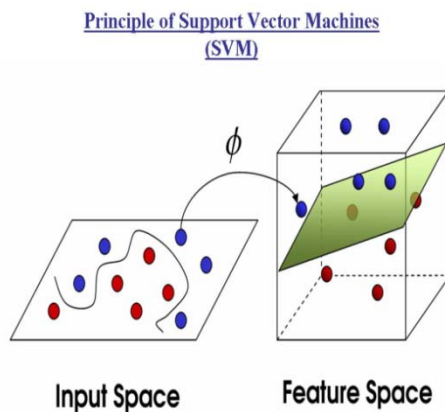
Multi-Execution: Is it common?



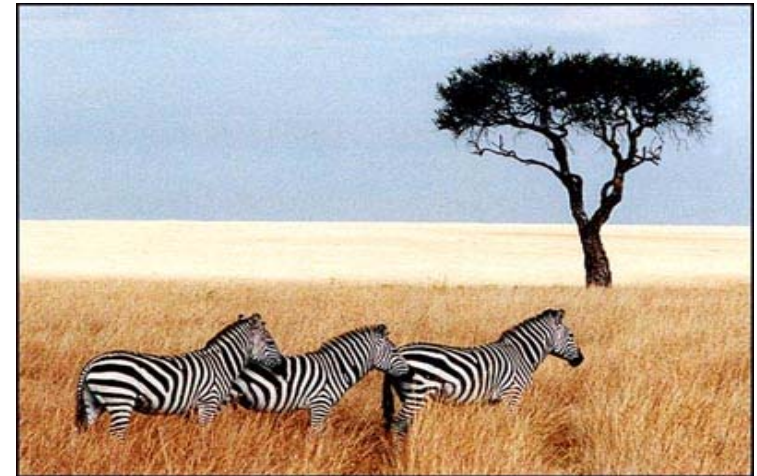
Simulation



CAD

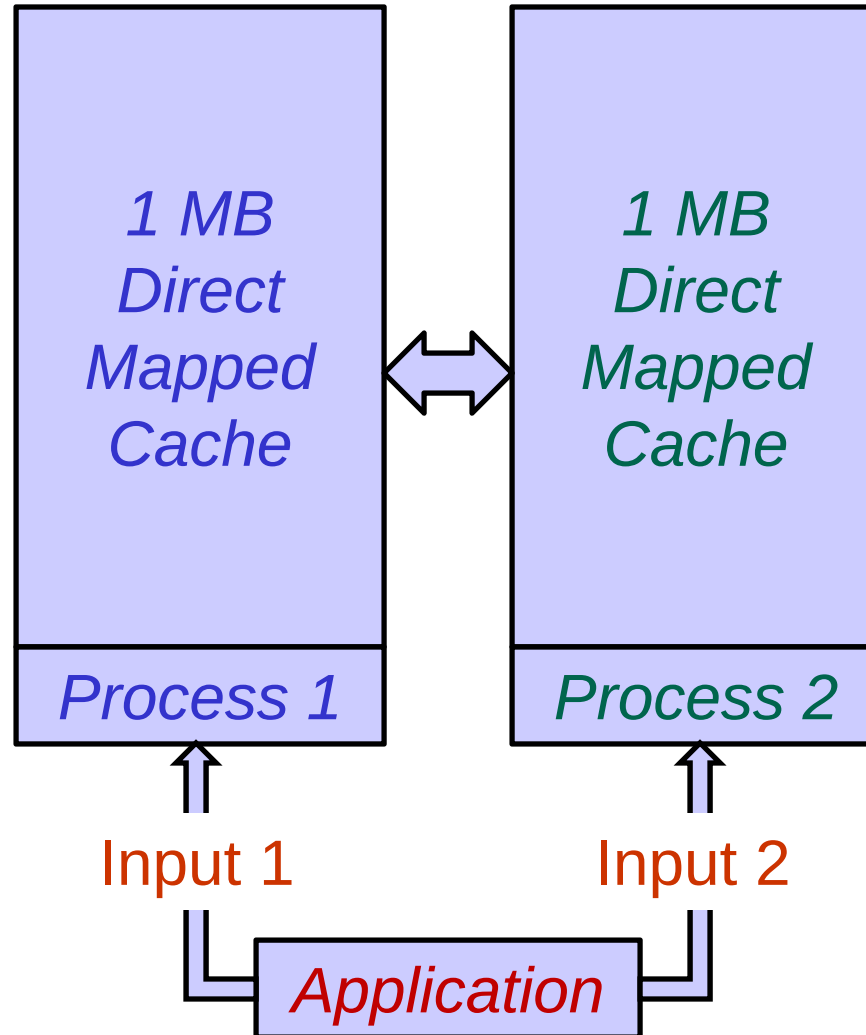


Machine Learning

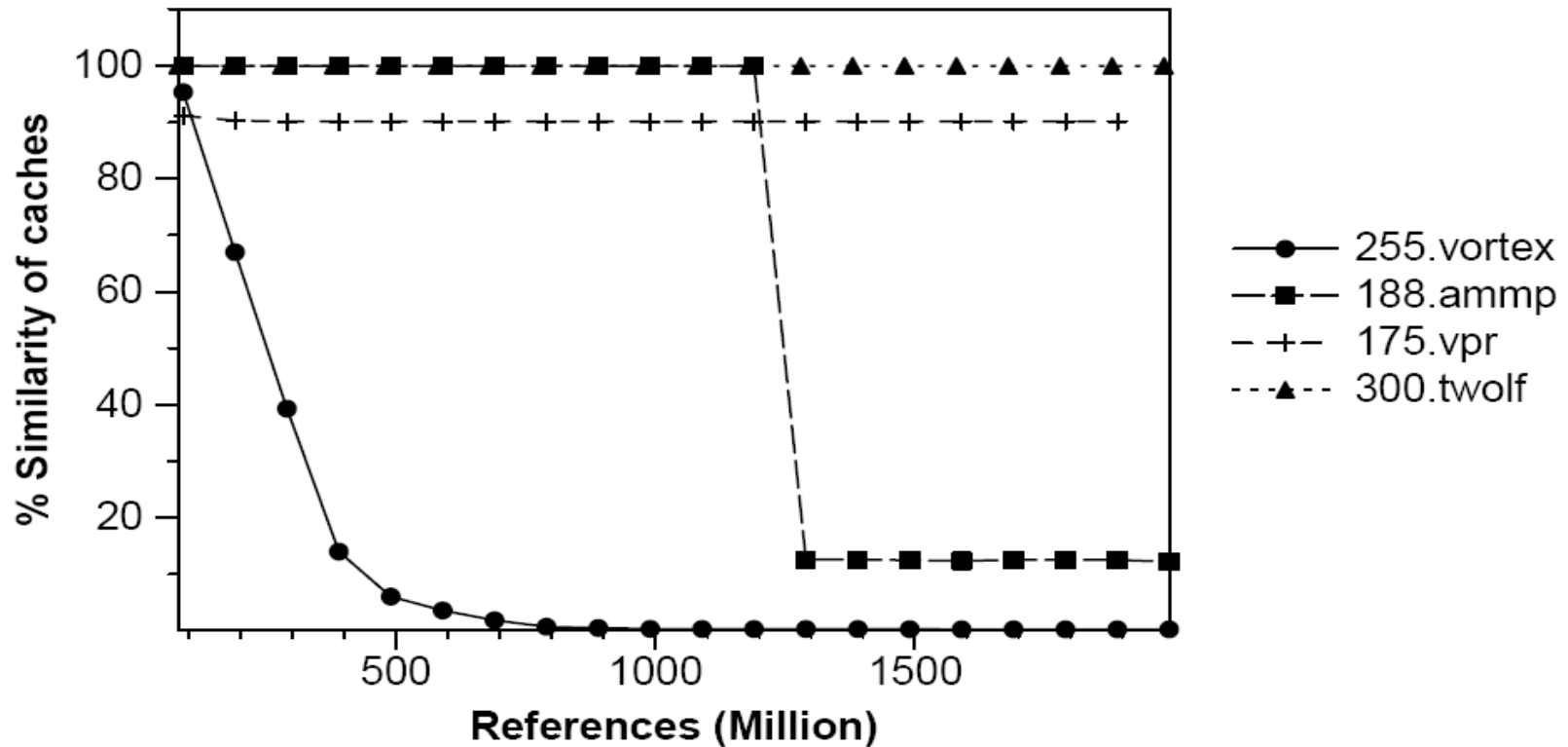


Data Hiding

Multi-Execution Data Similarity



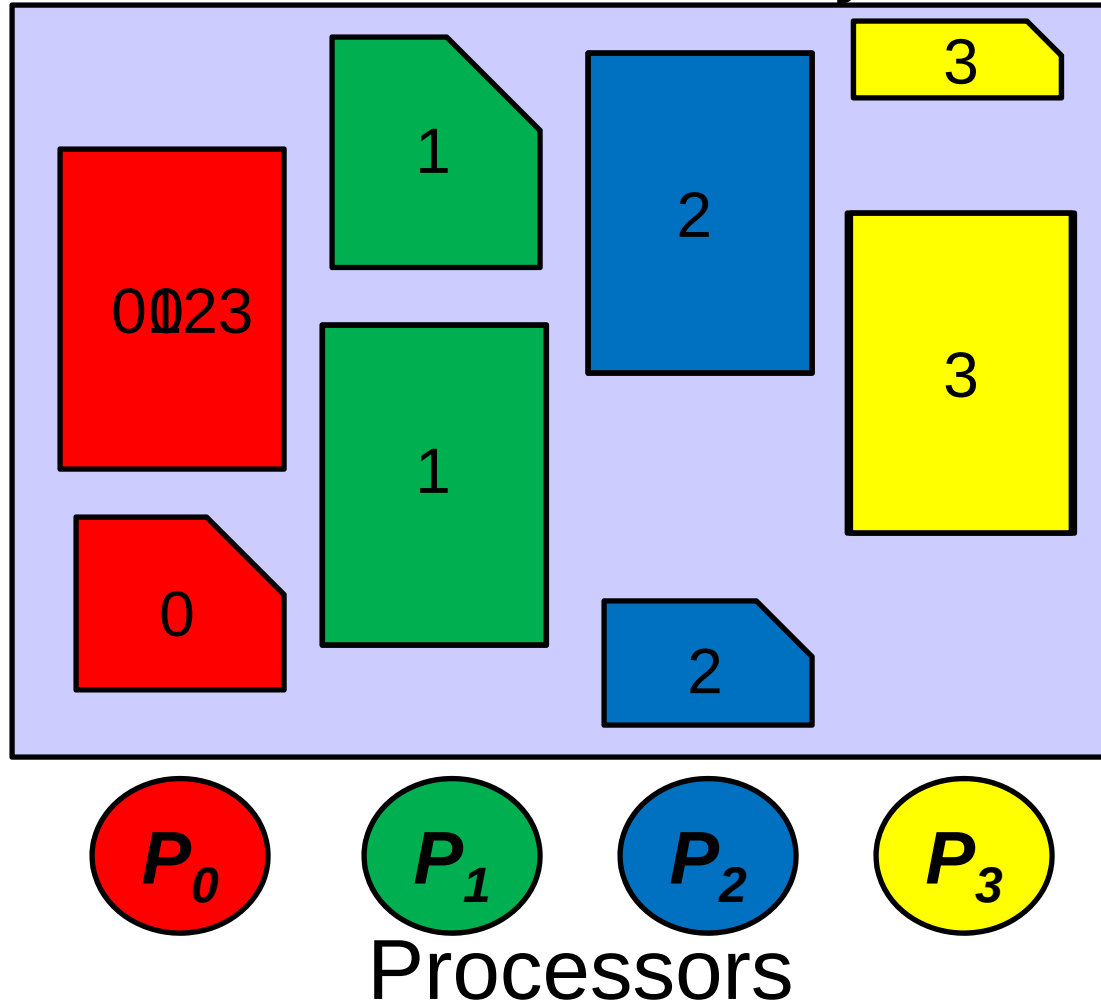
Multi-Execution Data Similarity



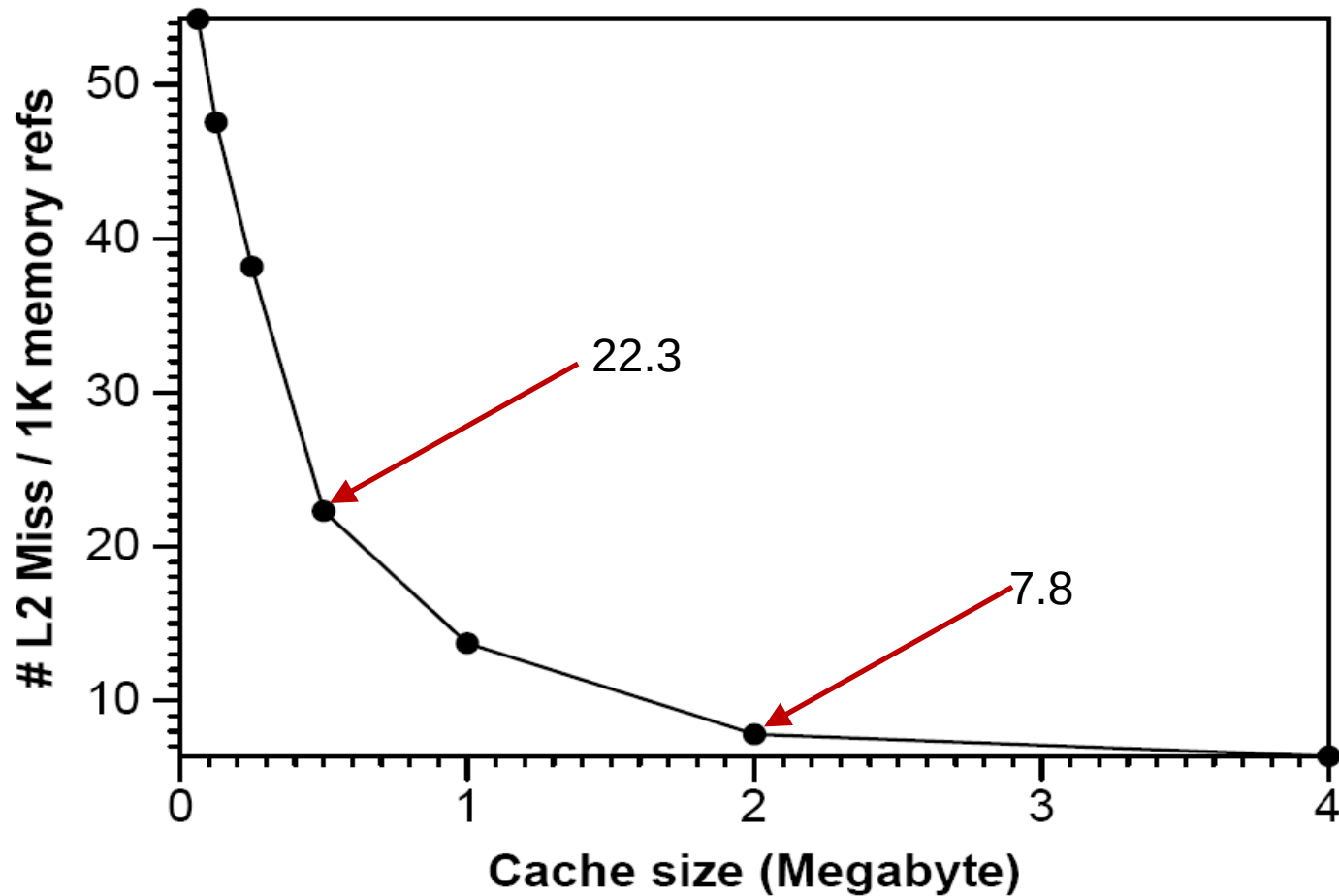
1 MB Direct Mapped Cache

Data Block Merging

Shared Memory

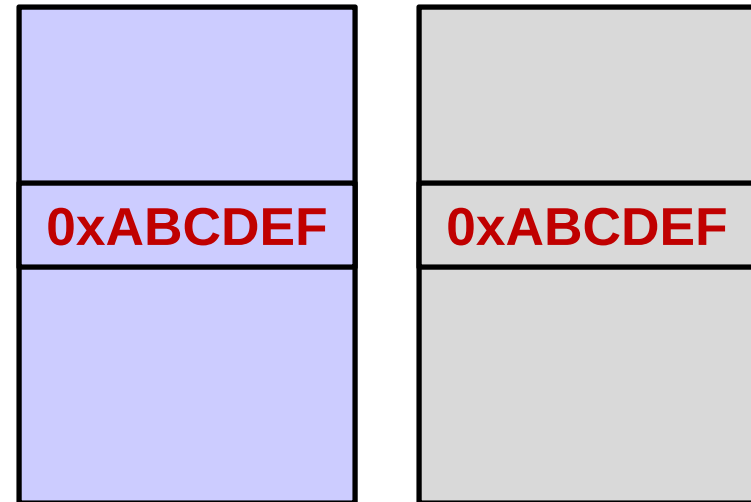


Effect of Cache Capacity on Miss Rate

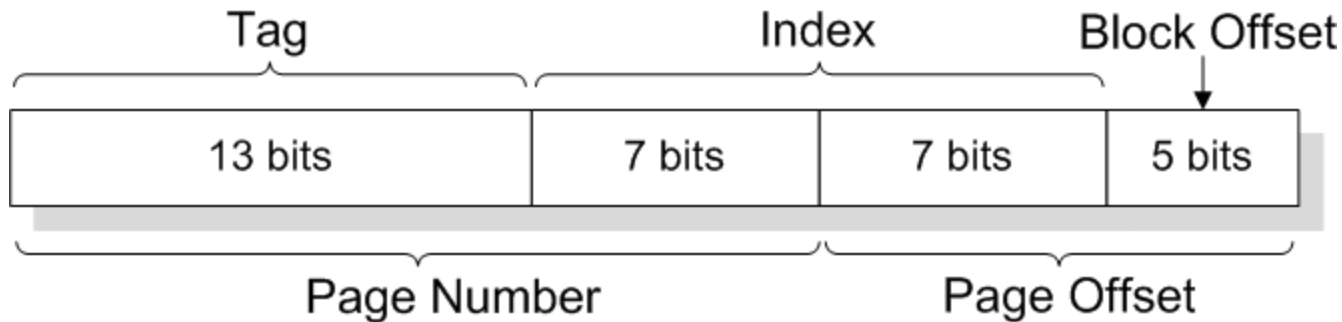


Challenges

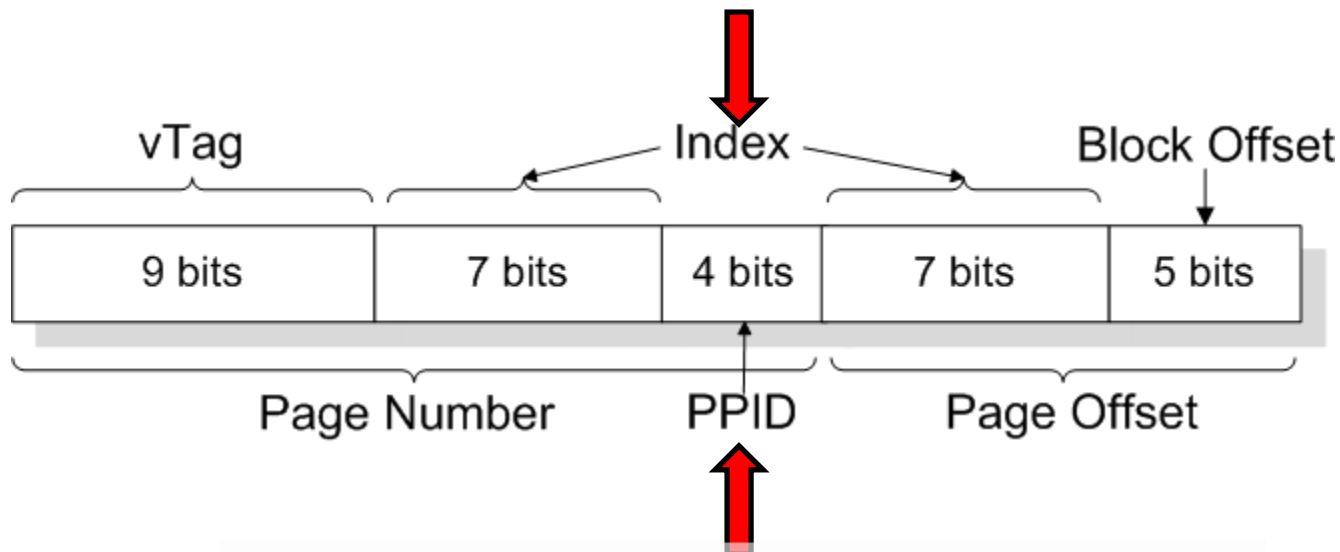
- Fast searching of identical blocks
 - Similarity at same virtual address
 - Limit within set
- Track Merged blocks
 - Bit per processor in tag
- Unmerge
 - Exclusive policy
 - unmerge on L1 miss



Modification in Addressing



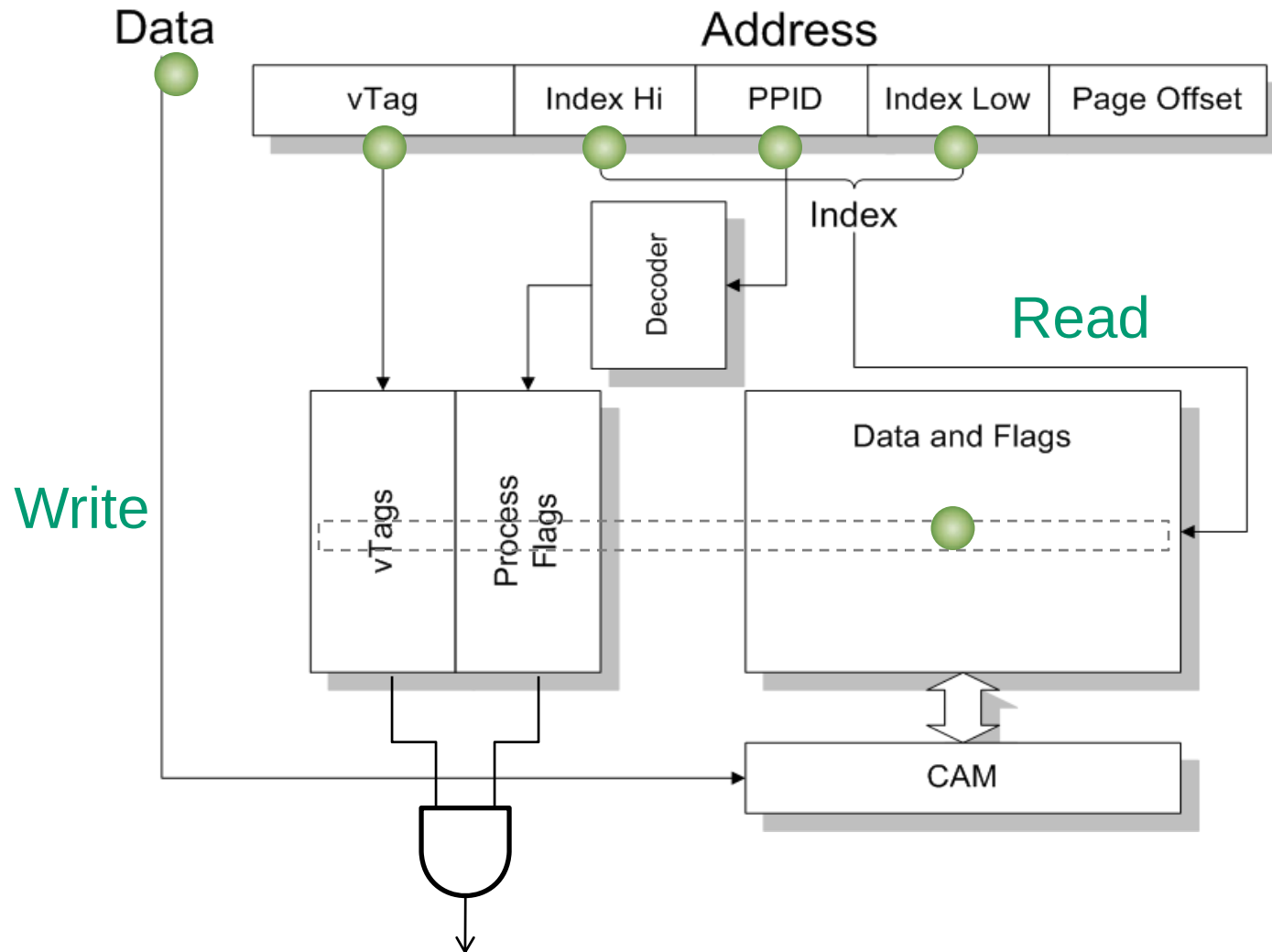
- 32 bit addr
- 4MB, 8W
- 14 bits index
- 4KB page
- 16 Proc



- 4-bit PPID
- 9-bit vtag
- coloring

Lower $\log_2(p)$ -bits of page number,
 p =number of processors

Mergeable Cache Design



Extra 5.28% power, 4.21% area of 4MB cache

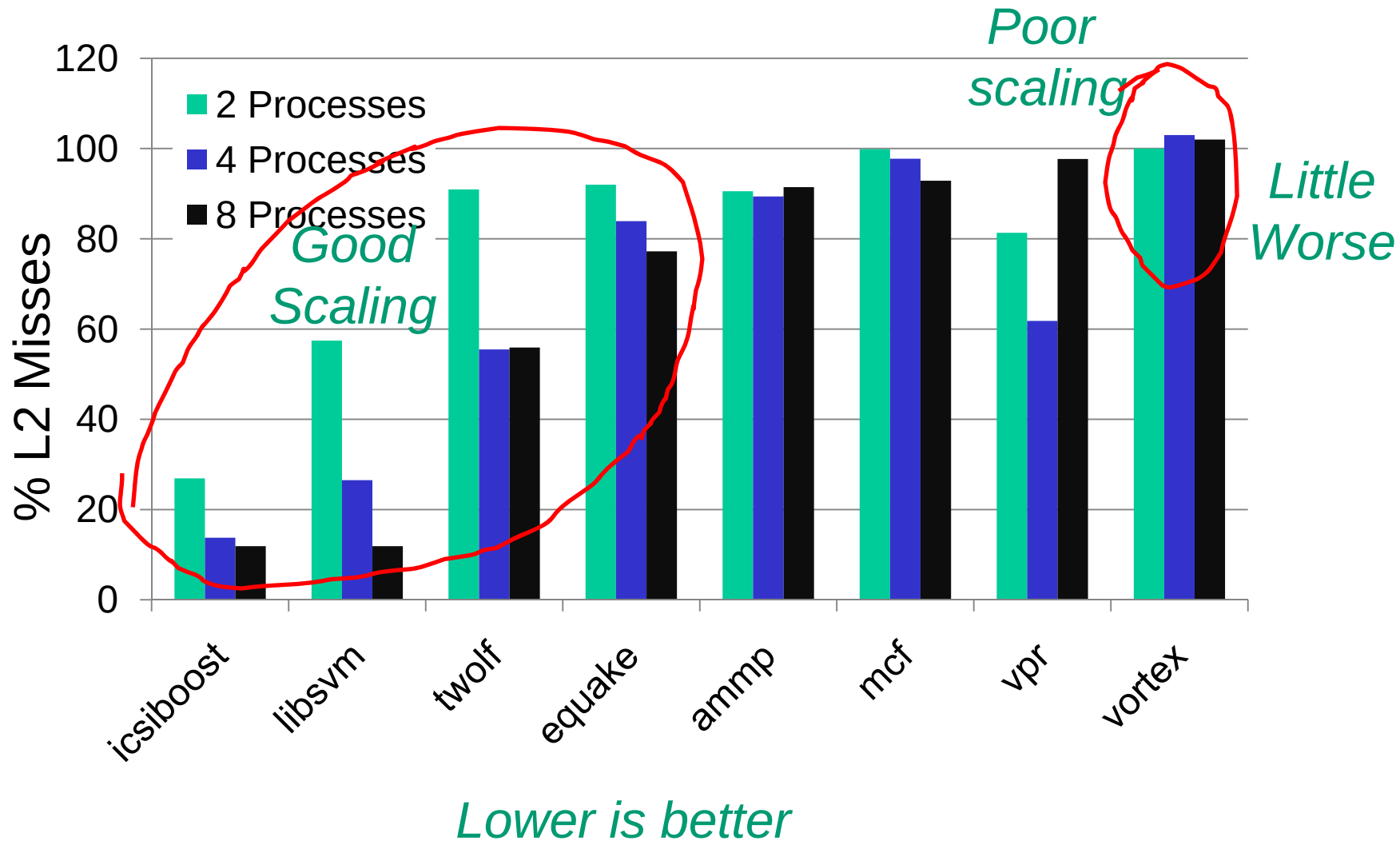
Simulation Framework

- Polyscalar OOO-multiprocessor simulator
- 2 – 8 instances of applications
- L1 cache: 32KB-I+32KB-D direct mapped
- L1 latency: 1 cycle
- Shared L2 cache: 4MB-8way-32B lines
- L2 latency: 6 cycles
- DRAM latency 200 cycles

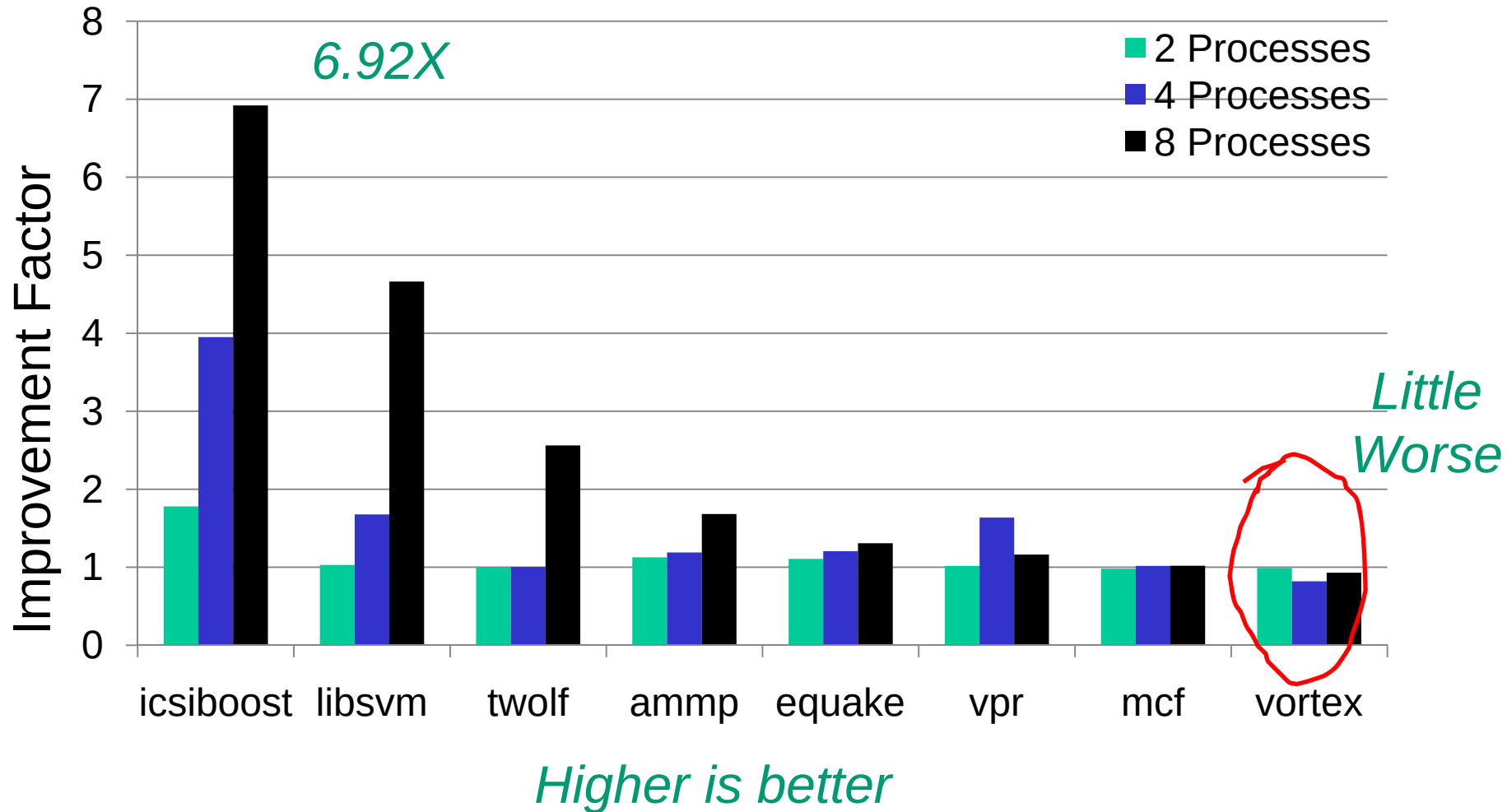
Benchmarks

Benchmark	Description	Input Modification	Run Length
175.vpr	FPGA Place and Route	routing-channel-width	3.3 B
181.mcf	Combinatorial Optimization	reduced trips	6.99 B
183.equake	Seismic Wave Propagation	epicenter and intensity	4.36 B
188.ammp	Chemistry	simulation parameters	6.13 B
255.vortex	Database	random insert, lookup	5.85 B
300.twolf	Place and Route	intercell gaps	4.13 B
libsvm	Machine learning	C and γ parameter	5.67 B
icsiboost	Ensemble learning	distribution of sample	2.30 B

Results: Off-chip Accesses



Improvement Over No Merging



Conclusion

- “*Multi-execution*” domain
- Data similarity exists
- Mergeable cache
 - improves performance
 - 6.92X max, 2.5X in average

Future Work

- Dynamic hybrid mergeable & conventional cache
- Wider application set
 - More applications
 - Libraries
- Coarse-grain software implementations

Thanks!



<http://www.cs.ucsb.edu/~arch>

susmit@cs.ucsb.edu