

Sustainable gem5 Simulations

Mohammad Alian

Assistant Professor

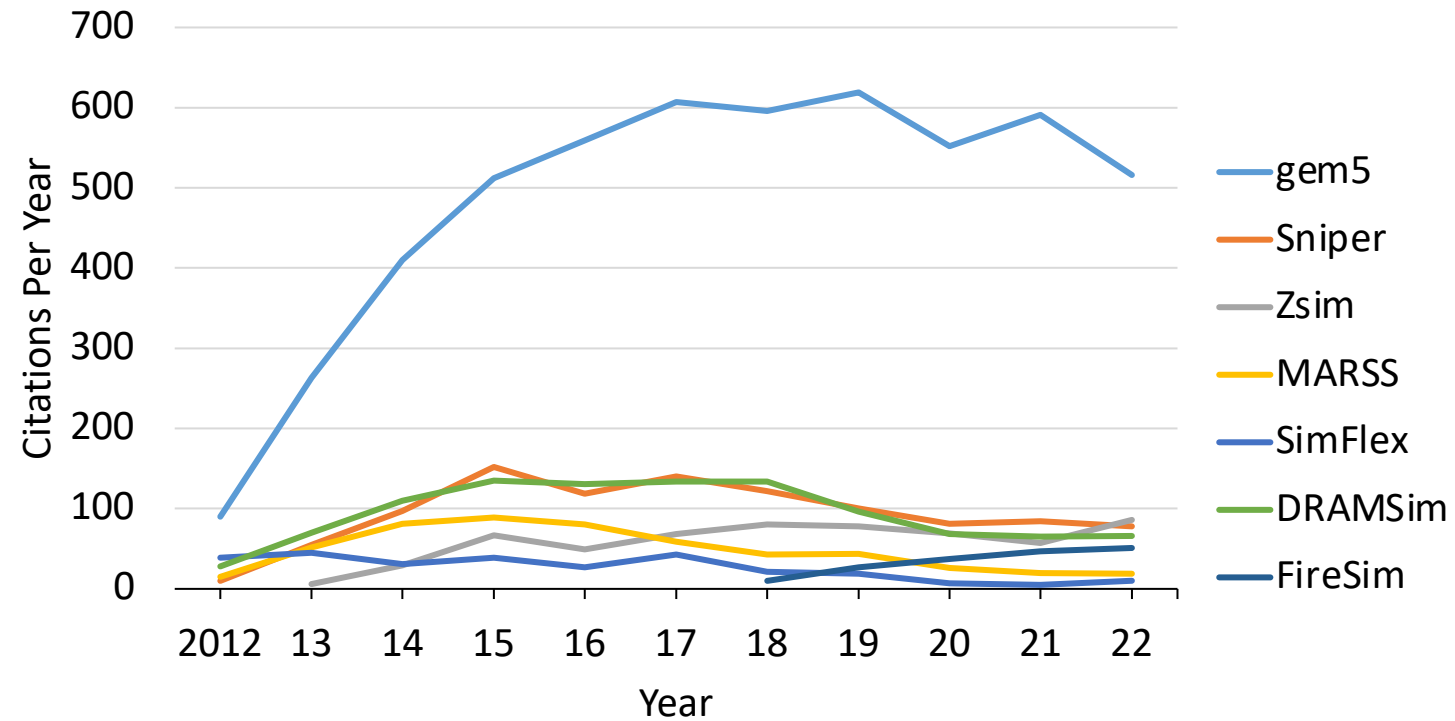
Architecture Research Group

Electrical Engineering and Computer Science



gem5 vs. All

- Industry
- Academia
 - ✓ Research
 - ✓ Education



Simulation is Wide Spread and Slow!

- Real system (single core): 1000 ~ 3000 MIPS
- gem5: 100 ~ 1000 KIPS
- 1000~30000 X slowdown

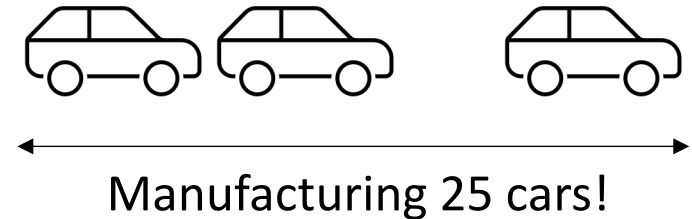
- I.e., 1 second of a single core simulation takes 17 min to 8.3 hours
- gem5 is single-threaded, so multi-core simulation time increases linearly

Sustainability of Simulations

- Sustainability == environmental impact
 - ✓ Operational: energy to run the simulation
 - ✓ Embodied: energy to purchase servers, server rooms, cooling, etc.
- Assume that each paper that uses gem5
 - ✓ Run 1000 experiments
 - ✓ Simulating 10 sec
 - 50 hours per experiment
- Simulation on a Dell server with 20 cores (250 Watts fully loaded)
- Operational energy consumption: 625kWh per paper
- Operational CO2 emission: 232 kgs of CO2 per paper

Emission of gem5 Simulations

- Assumption: 600 projects use gem5 per year
- 139 tones of CO2 per year



- 2x more when considering embodied carbon of simulation servers
 - ✓ More when considering underutilized periods
- Much more if considering:
 - ✓ Cold room, rack, IT service, building, cables, ...

Many Efforts on Reducing Simulation Time

- Throughput simulation
 - ✓ Run many simulations at the same time
- Parallelizing
 - ✓ dist-gem5 [ISPASS'2018]
 - ✓ Multi-threading
- Sampling
 - ✓ Fast forwarding (checkpointing), warmup, then detailed simulation

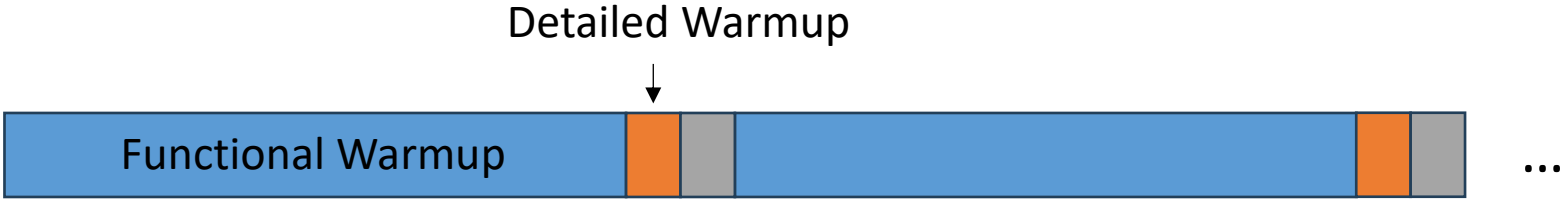
Do they improve sustainability?

Sustainability?

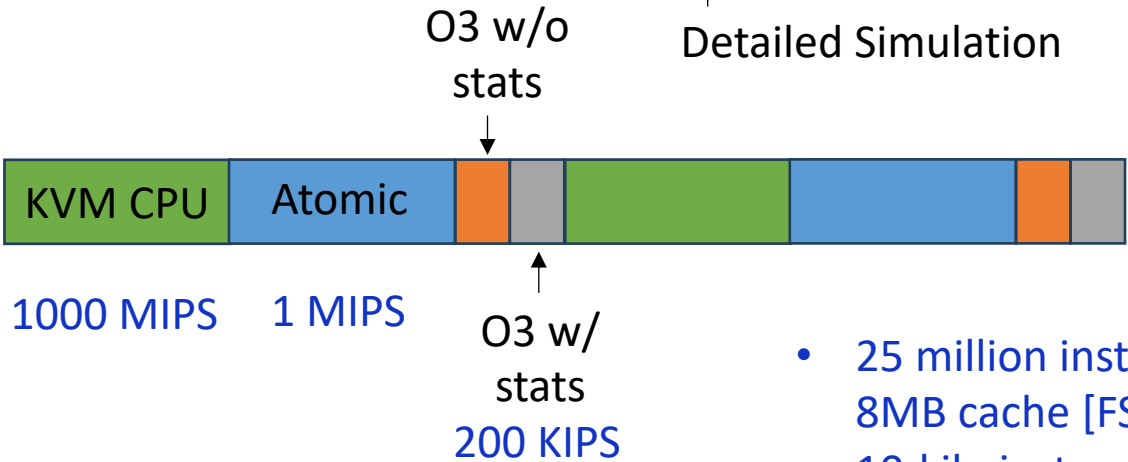
- Throughput simulations can hurt sustainability!
 - ✓ Many redundant and careless simulations
- Parallelizing simulation improves speed, not sustainability
 - ✓ Useful work + synchronization
- Sampling can improve overall simulation time, BUT
 - ✓ Requires improving the speed of the functional warmup phase

Statistical Sampling

SMARTS [ISCA'03]



SMARTS in gem5



- 25 million instructions to warm up an 8MB cache [FSA-IISWC'15]
- 10-kilo instructions for detailed sim

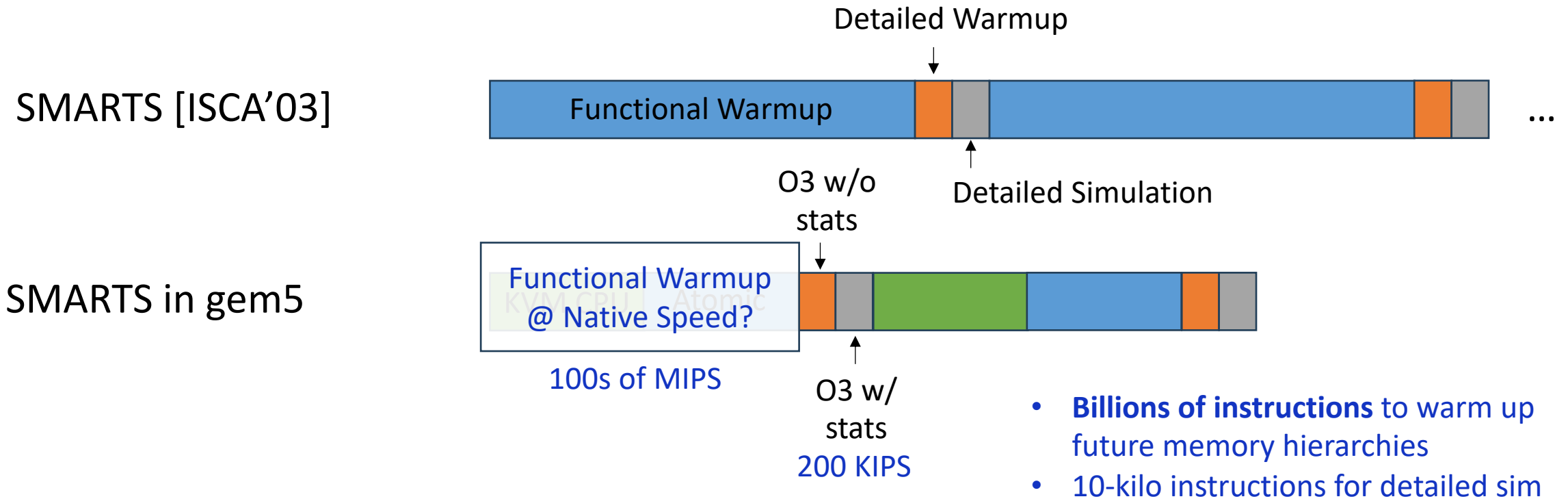
Statistical Sampling for Future Architectures

- Memcached w/ 10GB dataset
- 60 Billion instructions to warm up the cache
- If we do it with 1MIPS:
 - ✓ 60,000 seconds -> 16 Hours just for one sampling unit
- 800 samples to eliminate biased results
 - ✓ 533 days worth of Atomic simulation!
 - ✓ You can parallelize this, but it is not sustainable!

1GB cache
(e.g., HBM)

DDR

Sustainable Statistical Sampling



We need to run functional warmup phase at native speed!

Final Thoughts

- Optimize simulation servers for running simulations!

✓ *“Profiling gem5 Simulation” ISPASS’23*

Platform	CPU	TDP	Simulation Rate (KIPS)	KIPS/Watt (Fully loaded)
Dell Xeon	Xeon Gold 6242	150W	642	85
MacStudio	M1	60W	1252	417
Ampere Altra Max	Neoverse N1	105W	732	557

- On-prem simulations servers are highly underutilized

✓ Disaggregate/share resources or go cloud!

✓ *NSF CCRI seed funding for building accelerated, cloud-based simulation infrastructure*

Conclusion

- Demand for simulation is growing
- Sustainability-aware simulations
 - ✓ Proper sampling
 - ✓ Shared infrastructure
 - ✓ Sustainable infrastructure

Thank You!

For more information visit KU Architecture Research Group at
<https://arg.ku.edu>

Acknowledgment

This work was supported in part by National Science Foundation (CNS-2213807) grant

Electrical Engineering and Computer Science

