LA-UR-21-28037
Approved for public release; distribution is unlimited.

Title: Machine Architecture Impact on Application Performance

Author(s): Przybylski, Nicklaus River

Intended for: HPC Intern Showcase, 2021-07-22 (Online, New Mexico, United States)

Issued: 2021-08-11

Machine Architecture Impact on Application Performance

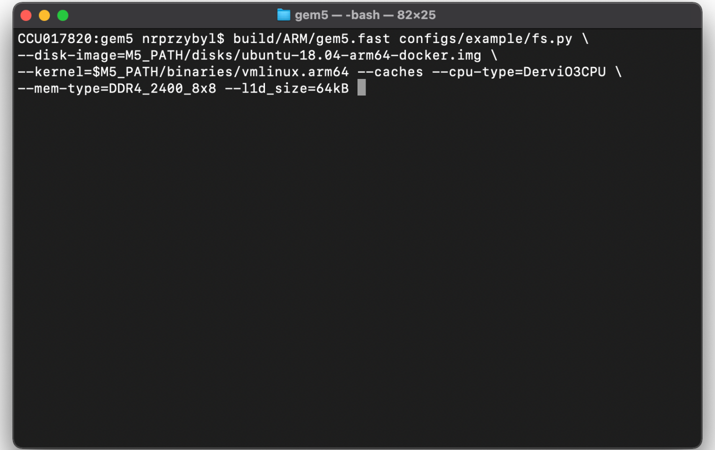
Nicklaus Przybylski

8/11/21

Agenda

The gem5 Simulator

- Simulation platform
- Models
  - CPU, cache, memory, etc.
- Two use cases
  - Full system mode
  - Syscall emulation mode

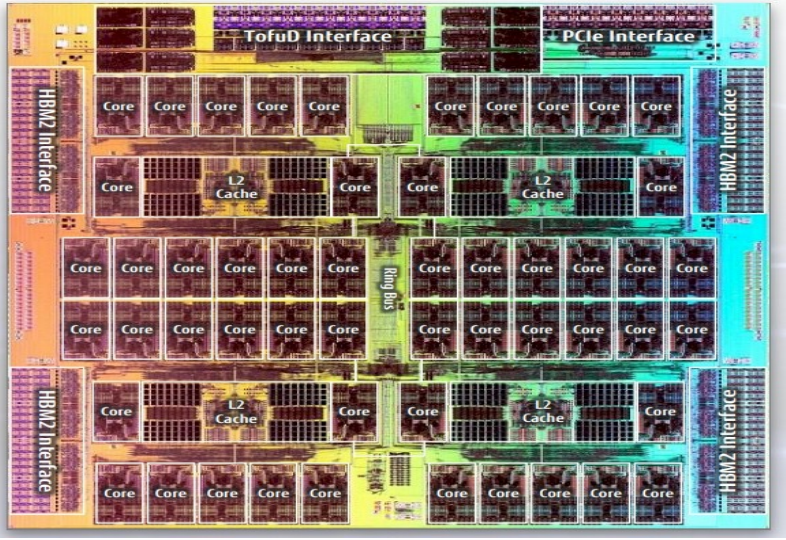


Bottom: gem5 full system example

Los Alamos NATIONAL LABORATORY

Motivation

- Prior experience using gem5
  - X86
    - Used in most desktops, servers, supercomputers
    - Sacrifices power efficiency for performance
- Joined Next Generation Platforms (NGP) team
  - ARM
    - Used in smaller electronic devices
    - Sacrifices performance for power efficiency



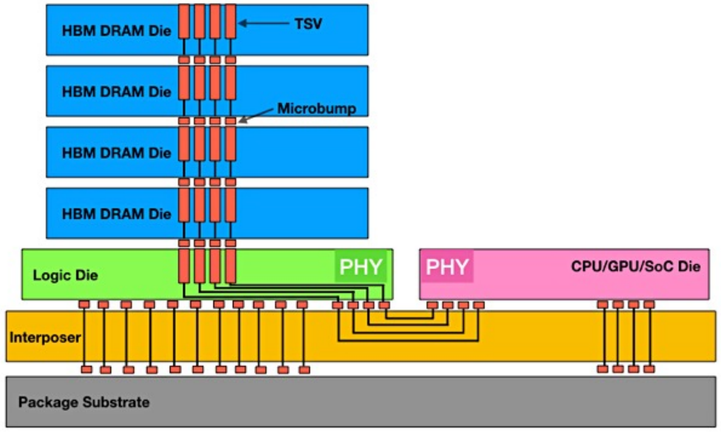
Top: Fugaku supercomputer

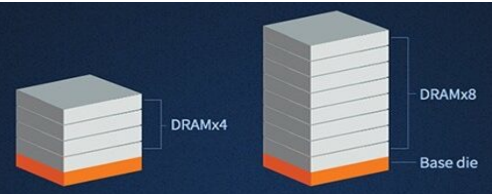
Bottom: Fujitsu A64FX topology

Los Alamos NATIONAL LABORATORY

Edits to gem5

- HBM2 (High Bandwidth Memory)
- Private L2, shared L3
- Minor changes to checkpointing



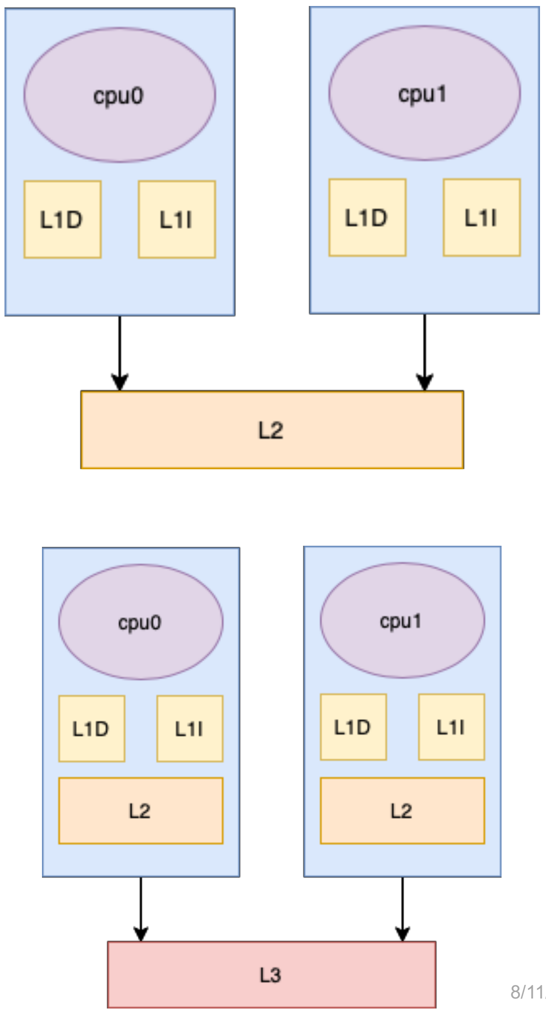


	HBM1	HBM2	HBM2E
Die Density (Gb)	2Gb	8Gb	16Gb
Stack Height	4Hi	4Hi / 8Hi	4Hi / 8Hi
Capacity (GB)	1GB	4GB / 8GB	8GB / 16GB
Bandwidth(GB/s)	128GB/s (1Gbps)	307GB/s (2.4Gbps)	410GB/s (3.2Gbps)

Top: HBM layout
Bottom: HBM1/2/2E comparison

Edits to gem5

- Memory
  - HBM2 (High Bandwidth Memory)
- Cache
  - Private L2, shared L3
- Minor changes to checkpointing



Top: default gem5 cache hierarchy

Bottom: updated gem5 cache hierarchy

Los Alamos NATIONAL LABORATORY

Parameter sweep experiments

- Pick applications to run through gem5
  - SNAP, PENNANT
- Run gem5 many times
  - Different combinations of inputs
  - SLURM
- Store
  - Inputs to gem5
    - Cache size, cache associativity, memory type, etc.
  - Output produced by the simulator
    - Simulated execution time, cache miss rate, memory latency, etc.

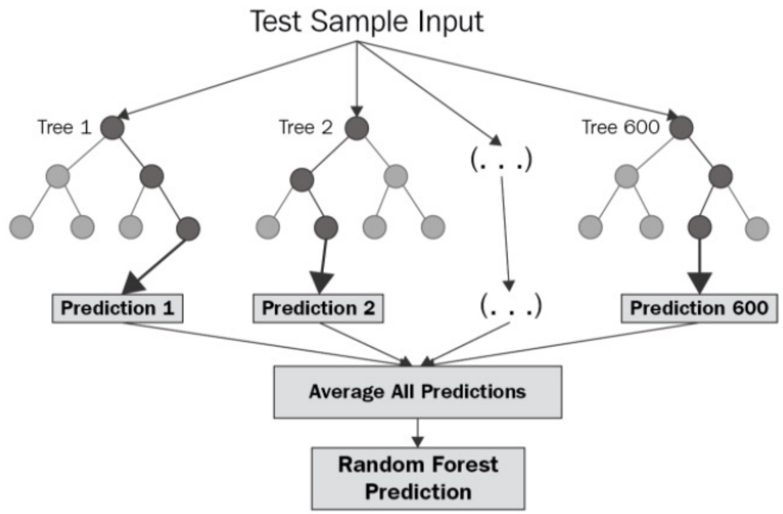
Sample data

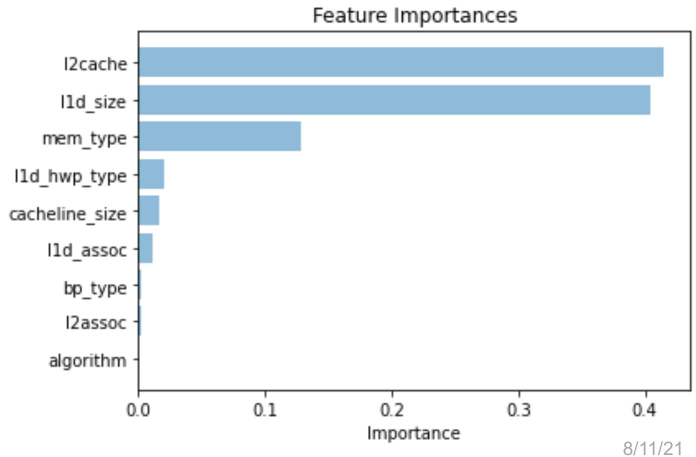
l1d_size	l1d_assoc	l2cache	l2assoc	mem_type	l1d_hwp_type	cacheline_size	algorithm	bp_type	sim_seconds
2	8	OFF	8	LPDDR5_6400_1x16_8B_BL32	BOPPrefetcher	128	quick	TournamentBP	0.0369
8	1	ON	16	DDR4_2400_4x16	BOPPrefetcher	64	quick	TournamentBP	0.028114
512	1	ON	4	DDR4_2400_4x16	AMPMPrefetcher	32	mm	BiModeBP	0.028122
8	1	ON	1	LPDDR5_6400_1x16_BG_BL32	AMPMPrefetcher	32	mm	LocalBP	0.061662
512	1	ON	4	HBM2_2000_4H_1x128	SignaturePathPrefetcher	32	mm	TournamentBP	0.028095
64	1	OFF	8	DDR4_2400_4x16	BOPPrefetcher	64	snap	BiModeBP	0.853243
16	4	OFF	8	DDR4_2400_16x4	BOPPrefetcher	128	pen	TournamentBP	0.121244
32	4	OFF	16	DDR4_2400_16x4	None	32	snap	MultiperspectivePerceptron64KB	0.8911690000000001
256	16	OFF	2	DDR4_2400_16x4	BOPPrefetcher	128	pen	LocalBP	0.094301
128	4	ON	4	DDR4_2400_16x4	TaggedPrefetcher	64	snap	TournamentBP	0.5381239999999999
256	4	OFF	8	LPDDR5_6400_1x16_BG_BL16	SignaturePathPrefetcher	32	snap	BiModeBP	1.0997370000000002
128	8	ON	1	DDR3_1600_8x8	SignaturePathPrefetcher	32	insert	LocalBP	0.613696
32	16	OFF	8	DDR4_2400_4x16	AMPMPrefetcher	128	insert	MultiperspectivePerceptron64KB	0.650169
4	8	ON	4	LPDDR5_6400_1x16_BG_BL32	TaggedPrefetcher	128	snap	TournamentBP	0.664966
8	1	ON	8	DDR3_1600_8x8	None	64	mm	LocalBP	0.05778099999999999

A sample of data collected after a randomized sweep of gem5 parameters

ML – Random Forests

- Random Forest
  - Ensemble of decision trees
  - Each tree uses different features to predict outcome
- Use gem5 inputs to predict output
- Feature importance



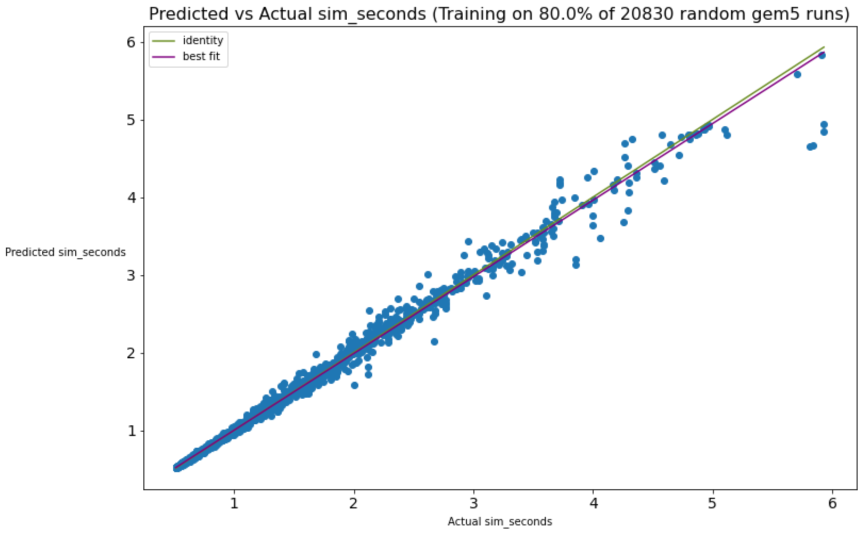


Top: Random Forest example

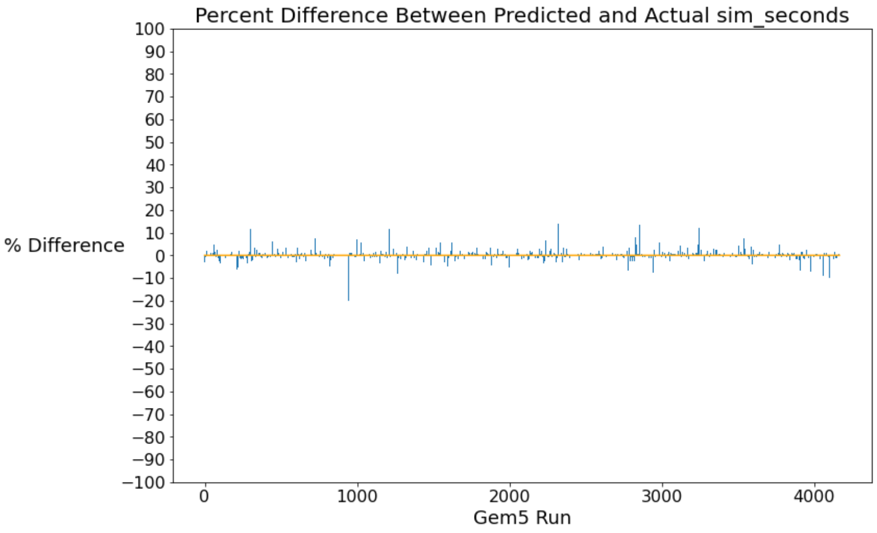
Bottom: Feature Importance for Random Forest model trained to predict simulation time on gem5 instances running SNAP

Los Alamos NATIONAL LABORATORY

Predictions



Scatterplot showing the correlation between predicted and actual values



Bar chart showing the percent difference between predicted and actual values for each prediction

Additional Efforts

- Hyperparameter tuning
  - Optimize size of train set
    - How little data can we train on and still get acceptable accuracy?
    - What exactly is acceptable accuracy?
- Compare gem5 output to real life performance
  - Work in progress

Questions?

- Mentors: Steve Poole, Terry Grové, Jeff Inman, Kevin Sheridan, Reid Rivenburgh

Sources

- The gem5 Simulator. Nathan Binkert, Bradford Beckmann, Gabriel Black, Steven K. Reinhardt, Ali Saidi, Arkaprava Basu, Joel Hestness, Derek R. Hower, Tushar Krishna, Somayeh Sardashti, Rathijit Sen, Korey Sewell, Muhammad Shoaib, Nilay Vaish, Mark D. Hill, and David A. Wood. May 2011, ACM SIGARCH Computer Architecture News.
- "Snap." *ECP Proxy Applications*, 25 Apr. 2019, proxyapps.exascaleproject.org/app/snap/.
- "Pennant." *ECP Proxy Applications*, 25 Apr. 2019, proxyapps.exascaleproject.org/app/pennant/.