Discussion 9

March 5th

Outline

- gem5 assignment 5
  - Introduction to the gem5 simulator
  - Look at the assignment 5 components
- Quiz 9

Introduction to the gem5 simulator

What is gem5?

Michigan m5 + Wisconsin GEMS = gem5

"The gem5 simulator is a modular platform for computer-system architecture research, encompassing system-level architecture as well as processor microarchitecture."

Lowe-Power et al. **The gem5 Simulator: Version 20.0+**. ArXiv Preprint ArXiv:2007.03152, 2021. https://doi.org/10.48550/arXiv.2007.03152

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. 2011. **The gem5 simulator**. *SIGARCH Comput. Archit. News* 39, 2 (August 2011), 1-7. DOI=http://dx.doi.org/10.1145/2024716.2024718

Created at Michigan by Steve Reinhardt and his students, principally Nate Binkert.

"A tool for simulating systems"

Two Views of M5

1. A framework for event-driven simulation
   - Events, objects, statistics, configuration
2. A collection of predefined object models
   - CPUs, caches, busses, devices, etc.

- This tutorial focuses on #2
- You may find #1 useful even if #2 is not

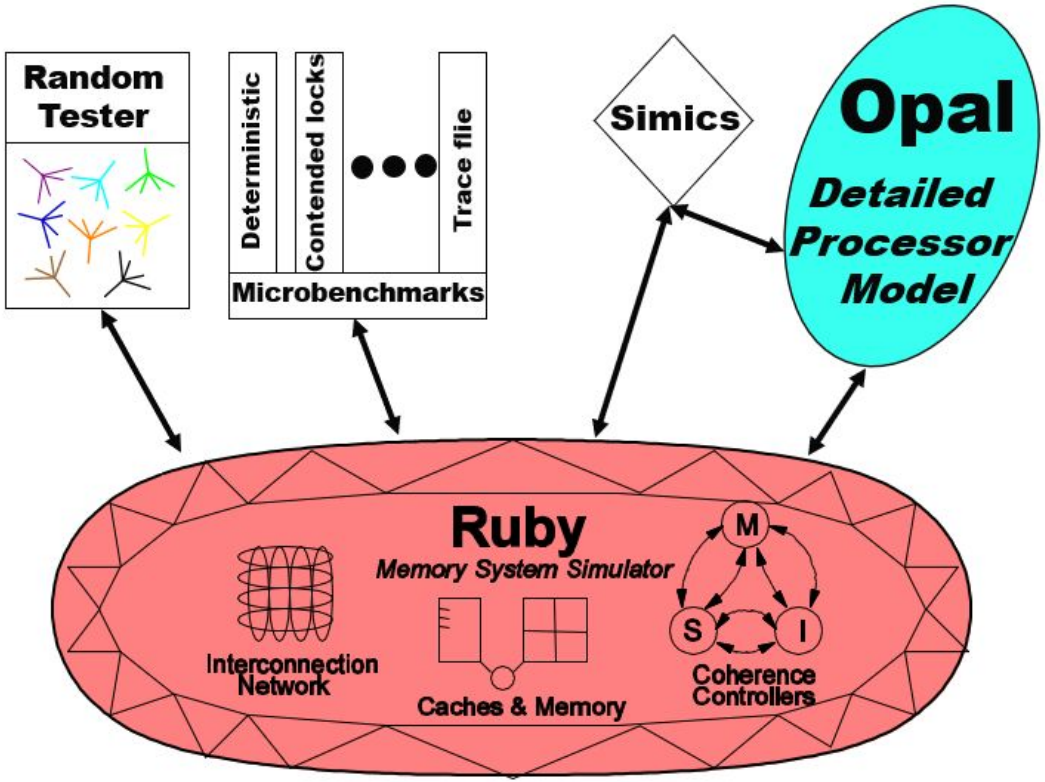
Created at Michigan by students of Steve Reinhardt, principally Nate Binkert.

"A tool for simulating systems"

n by students of Mark Hill and David Wood.

Detailed memory system

GEMS From 50,000 Feet



Why simulation

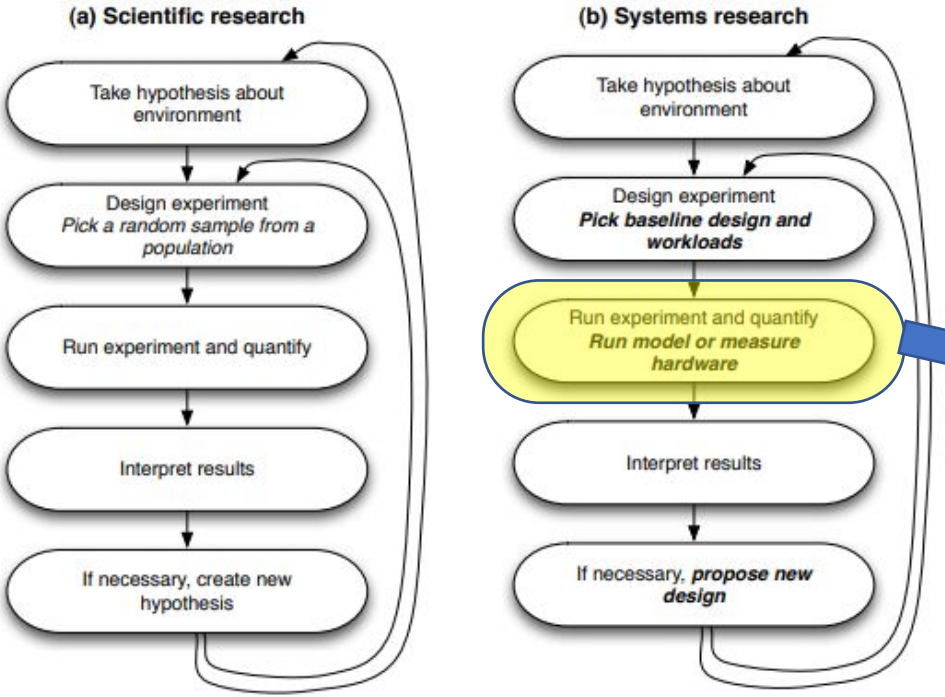
Need a tool to evaluate systems that don't exist (yet)
- Performance, power, energy, etc.

Very costly to actually make the hardware

Computer systems are complex with many interdependent parts
- Not easy to be accurate without the full system

Simulation can be parameterized
- Design-space exploration
- Sensitivity analysis

Computer systems research/engineering

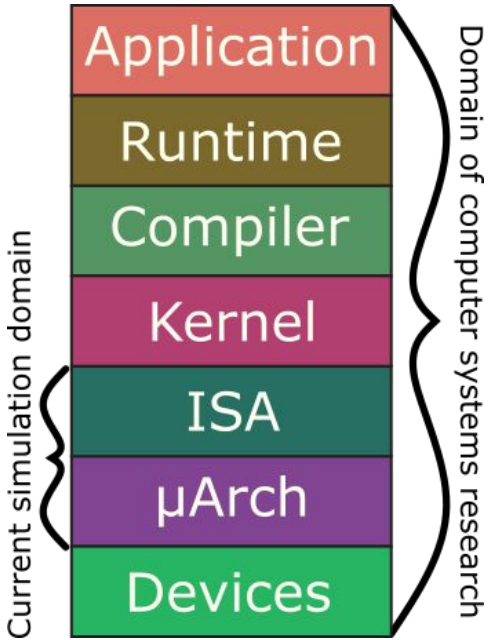


From Computer Architecture Performance Evaluation Methods by Lieven Eeckhout

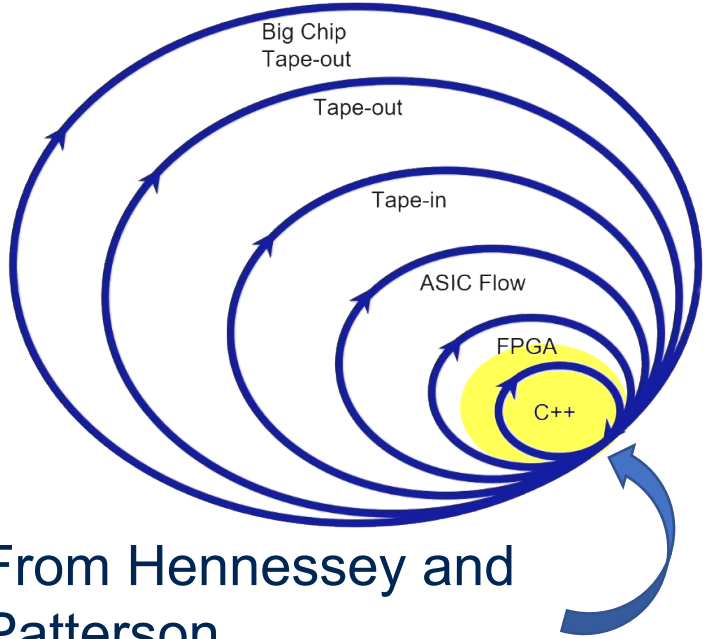
Computer architecture simulation!

gem5

gem5's goals

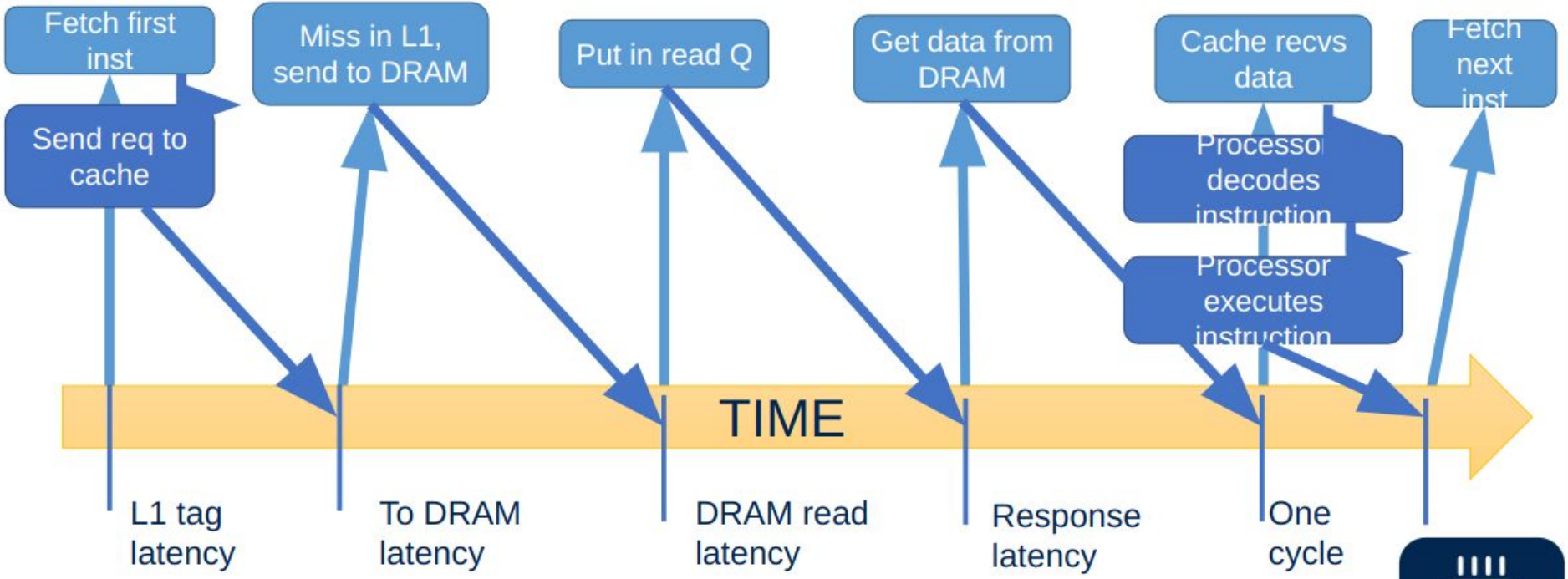


Agile Hardware Dev. Methodology



From Hennessey and Patterson
Turing Lecture

Discrete event simulation example



gem5

Let’s look at the gem5 assignment 5’s setup

Quiz 9

Question 1 — 2 pts

There is a program in which 72% can be parallelized.

When you run this program on 32 cores, what is the speedup compared to the serial version?

$$\frac{1}{1-P+\frac{P}{\#C}} = \frac{1}{28\% + \frac{72\%}{32}} = 3.3X$$

Question 2 — 2 pts

There is a program in which 62% can be parallelized.

When you run this program on 16 cores, what is the speedup compared to the serial version?

$$\frac{1}{38\% + \frac{62\%}{16}} = 2.4X$$

Question 3 — 2 pts

There is a program in which 82% can be parallelized.

What is the *maximum possible* speedup you can achieve through parallelization?

$$\frac{1}{18\% + \frac{82\%}{\infty}} = \frac{1}{18\%} = 5.6X$$

Question 4

4 pts

t & C are constant.

Assume you have an application where 75.0% is parallelizable.

You have three systems, a one core system, a two core system, and a four core system. Fill in the table below with the relative performance, power, and energy for each of the systems.

Cores	Capacitance	Voltage	Frequency	Time	Power	Energy 2ct
1	1*c	1V	2 GHz	1*t	2*c	
2	1.8*c	.9V	1.5 GHz	[Select] 0.83t	[Select]	
4	2.5*c	.7V	1 GHz	[Select]	[Select]	

Which system is the most *energy efficient* (uses the least energy)?

[Select]

$t_1 = I \times CPI_{old} \times \frac{1}{f_1}$

$t_2 = I \times (25\% + \frac{75\%}{2}) \times CPI_{old} \times \frac{1}{f_2}$.

$t_2 = t_1 \times (25\% + \frac{75\%}{2}) \times \frac{f_1}{f_2} = 1 \times t \times (25\% + \frac{75\%}{2}) \times \frac{2GHz}{1.5GHz}$

$= 0.83t$.

Question 4 — 4 pts

t & C are constant.

Assume you have an application where 75.0% is parallelizable.

You have three systems, a one core system, a two core system, and a four core system. Fill in the table below with the relative performance, power, and energy for each of the systems.

Cores	Capacitance	Voltage	Frequency	Time	Power	Energy 2ct
1	1*c	1V	2 GHz	1*t	2*c	
2	1.8*c	.9V	1.5 GHz	[Select] 0.833t	[Select]	
4	2.5*c	.7V	1 GHz	[Select] 0.88t.	[Select]	

Which system is the most *energy efficient* (uses the least energy)?

[Select]

$$t_4 = t_1 \times \left(25\% + \frac{75\%}{4}\right) \times \frac{2\,GHz}{1\,GHz}$$

$$= 0.88t.$$

Question 4

4 pts

t & C are constant.

Assume you have an application where 75.0% is parallelizable.

You have three systems, a one core system, a two core system, and a four core system. Fill in the table below with the relative performance, power, and energy for each of the systems.

$\frac{CV^2f}{2}$

Cores	Capacitance	Voltage	Frequency	Time	Power	Energy 2ct
1	1*c	1V	2 GHz	1*t	2*c	
2	1.8*c	.9V	1.5 GHz	[Select] 0.833t.	[Select] 2.19c.	
4	2.5*c	.7V	1 GHz	[Select] 0.88t.	[Select] 1.22c.	

Which system is the most *energy efficient* (uses the least energy)?

[Select]

CV^2f

$Power_2 = 1.8c \times (.9)^2 \times 1.5GHz. = 2.19c.$

$Power_4 = 2.5c \times (.7)^2 \times 1 = 1.22c.$

Question 4

4 pts

t & C are constant.

Assume you have an application where 75.0% is parallelizable.

You have three systems, a one core system, a two core system, and a four core system. Fill in the table below with the relative performance, power, and energy for each of the systems.

Cores	Capacitance	Voltage	Frequency	Time	Power	Energy 2ct
1	1*c	1V	2 GHz	1*t	2*c	
2	1.8*c	.9V	1.5 GHz	[Select] 0.833t.	[Select] 2.19C.	1.82ct.
4	2.5*c	.7V	1 GHz	[Select] 0.88t.	[Select] 1.22C.	1.07ct.

Which system is the most *energy efficient* (uses the least energy)?

[Select] 4 core.

$Energy_2 = 0.833t \times 2.19c = 1.82ct.$

$Energy_4 = 0.88t \times 1.22c = 1.07ct.$

Question 5

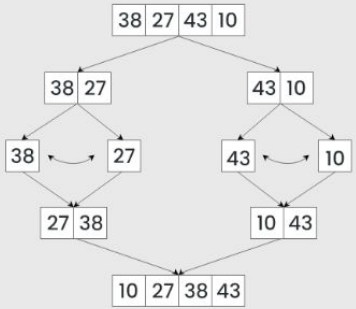
2 pts

Which of the following represent data-level parallelism?

- [x] Applying a filter to an image (e.g., making every pixel darker)
- [x] `for (int i=0; i<size; i++) result[i] = a*x[i] + y[i]`
- [ ] Parallel version of mergesort
- [ ] A webserver serving requests from many clients

Merge Sort

Algorithm



Question 6

2 pts

When the application has [Select] parallelism, a SIMD architecture is probably more efficient than a MIMD architecture.

Data-level.

Question 7

2 pts

Implementing communication via message passing is required when which of the following is true?

- [ ] The workload uses multiple cores on a single machine
- [x] The workload has a large parallel portion
- [ ] The workload will only execute on a single machine at a time
- [x] The workload requires more memory than will fit on a single machine

Question 8

2 pts

Which of the following are examples of **message passing** architectures?

- [ ] Most smartphones
- [ ] The AMD Epyc system discussed at the end of the memory section
- [ ] A GPU
- [x] Super computers (e.g., Summit at ORNL and Sierra and LLNL)
- [ ] Most desktop computers
- [x] Data centers

shared mem

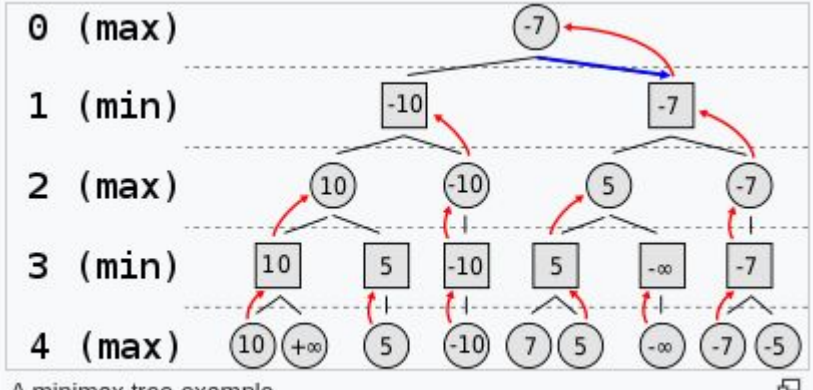
Question 9

2 pts

Read the wikipedia pages on the minimax algorithm. This algorithm can be used for playing games and maximizes the minimum value of an opponents' move.

Which parallel programming technique do you think would be most appropriate to parallelize minimax?

- ✓ Task-parallism / MIMD possibly via recursion
- Data-level parallelism / SIMD by using array operations



A minimax tree example