FireSim/Chipyard/Gemmini Tutorial: Intro

Seah Kim

UC Berkeley

seah@berkeley.edu

Berkeley Architecture Research

GEMMINI

CHIPYARD

Presenters / Organizers

Seah Kim

Abraham Gonzalez

Jerry Zhao

Joonho Whangbo

Vikram Jain

Sagar Karandikar

Krste Asanović

Sophia Shao

Borivoje Nikolić

A *Golden Age* in Comp Arch Tools

- No more traditional scaling…
- An architect's dream: everyone wants custom microarchitectures and HW/SW co-designed systems
- Also, a golden age to have *direct impact* as researchers
  - Exploding open-source hardware environment
  - An open-ISA that can run software we care about

https://cacm.acm.org/magazines/2019/2/234352-a-new-golden-age-for-computer-architecture/fulltext

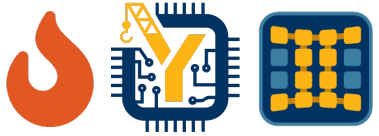
A *Dark Age* in Comp Arch Tools

- What do we need to do good architecture research?
  - Need tools that let us evaluate designs on a variety of metrics:
    - Functionality
    - Performance
    - Power
    - Area
    - Frequency
  - Especially in small teams (grad students, startups), these tools need to be *agile*
  - Historically, without good open IP, had to build abstract arch/uarch simulators out of necessity
    - But now, we have much better IP and software compatibility, so what's stopping us?

A *Dark Age* in Comp Arch Tools

- Designed to be operated by hundreds of engineers
- Not, 10s of engineers or 1s-10s of grad students
- Three hard questions:
  - How do I easily generate a realistic, well-optimized, programmable ML accelerator for full-system exploration?
  - Where do I get a collection of well-tested hardware IP + complex software stacks that run on it?
  - How do I quickly obtain performance measurements for a novel HW/SW system?

Three hard questions, answered!



- How do I easily generate a realistic, well-optimized, programmable ML accelerator for full-system exploration?

- Where do I get a collection of well-tested hardware IP + complex software stacks that run on it?

- How do I quickly obtain performance measurements for a novel HW/SW system?

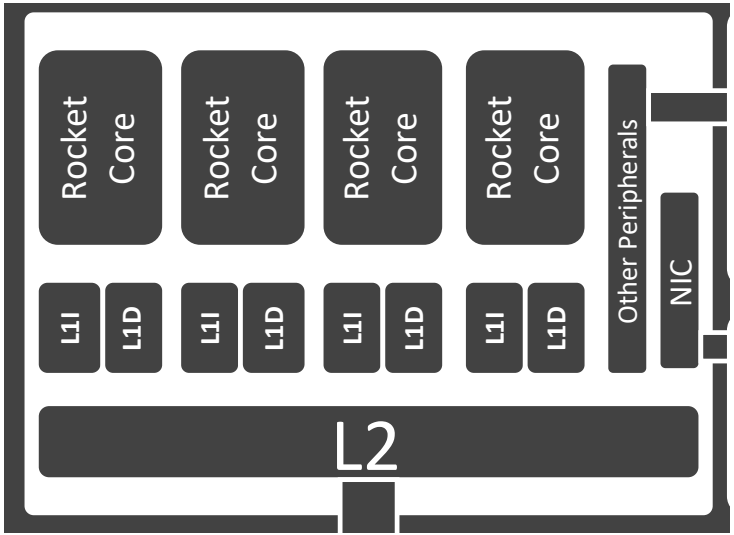
Three hard questions, answered!

+

Measure Functionality, Performance, Power, Area, Frequency *for real HW/SW systems*, quickly and easily, with small teams of engineers

What kinds of designs can I work with?

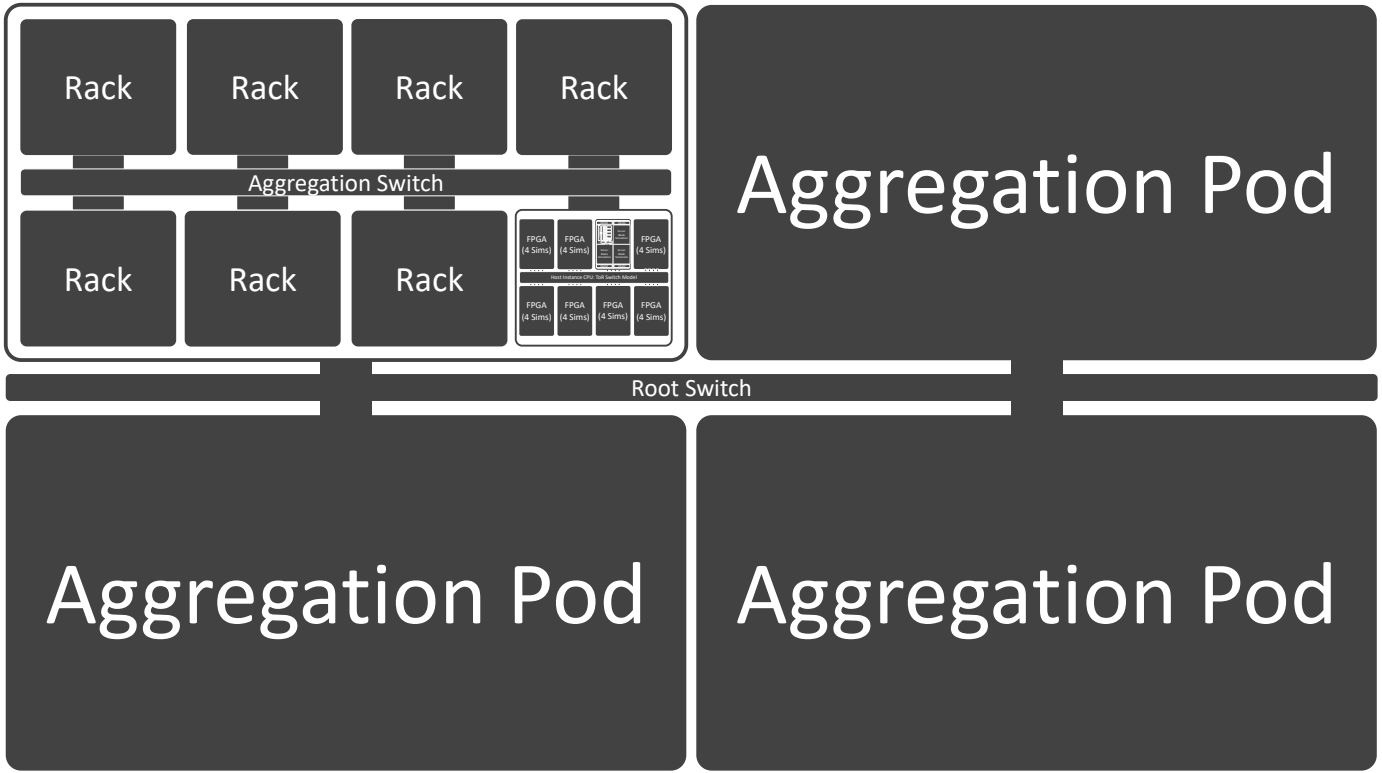
- RISC-V Cores:
  - Rocket Chip In-Order core, industry proven
  - SonicBOOM Out-of-Order Superscalar core
  - CVA6
  - Ibex
- Accelerators
  - **Gemmini (Berkeley DNN Accelerator)**
  - sha3 accelerator
  - NVDLA (NVIDIA Deep Learning Accelerator)
  - Hwacha Vector Accelerator
  - FFT Generator
  - Many more
- Peripherals/other IP
  - L2 Cache, UART, Disk, Ethernet NIC, etc.
- FPGA-Simulation Models
  - Large LLCs, large DDR3 memory systems



Complete Single-SoC RISC-V System

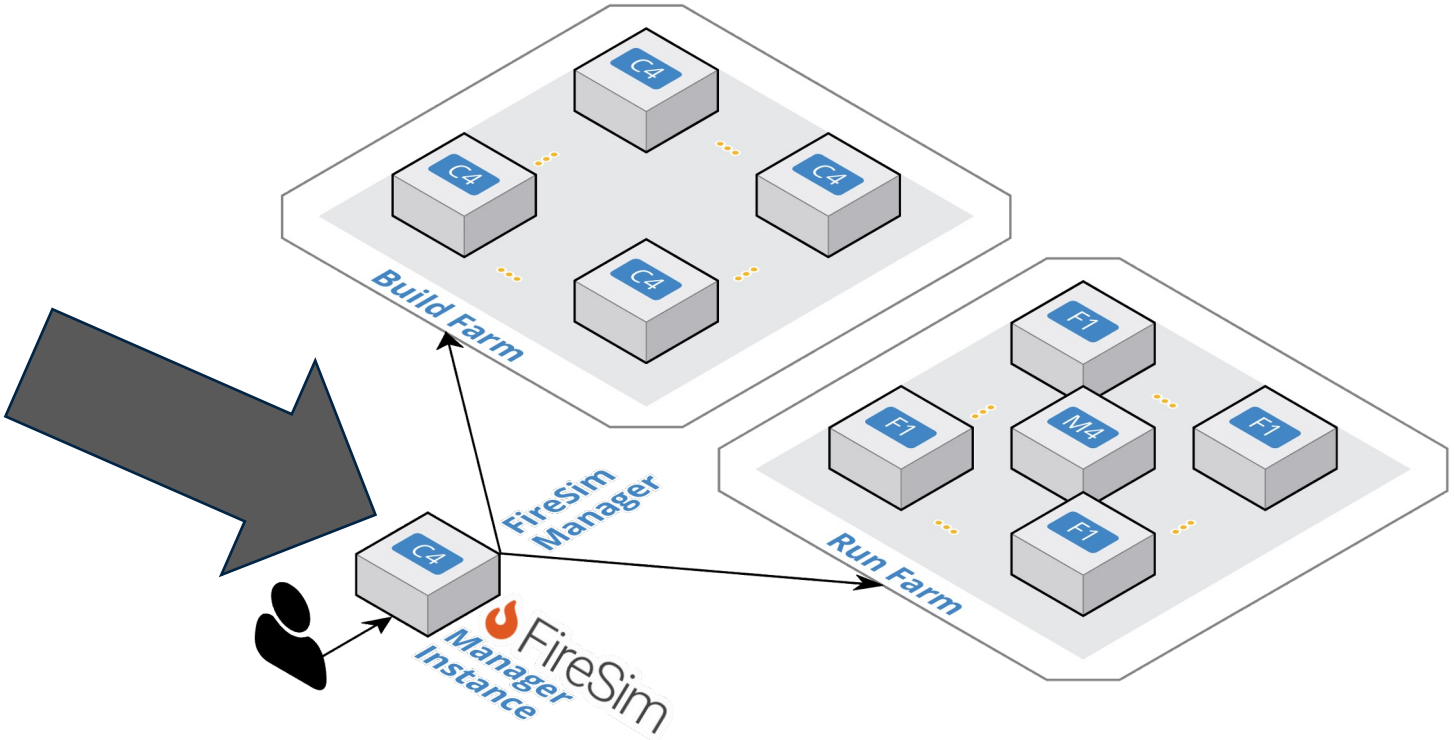
What kinds of designs can I work with?

Ethernet-Networked 1024 Node (4096 Core) System on 256 Cloud FPGAs



Today's Logistics

You are here



Today's Agenda

8:25am: Introduction/Overview, Amazon EC2 Instance Setup, Logistics - Seah

8:40am: Chipyard Basics – Jerry

9:00am: Building Custom RISC-V SoCs in Chipyard – Jerry

9:30am: Hammer VLSI Flow, Chips, and Chiplet-yard - Vikram

10:00am: Coffee break

10:30am: Gemmini Introduction: Architecture and Programming – Seah

11:10am: AuRORA Introduction – Seah

11:30am: Gemmini SoC Integration with AuRORA - Seah

12:00pm – 1:00pm: Lunch

Today's Agenda

1:00pm: FireSim Introduction – Abe

1:20pm: FireSim Update and FireAxe - Joonho

1:50pm: Building Hardware Designs in FireSim – Joonho

2:20pm: Running FireAxe - Joonho

3:00pm: Coffee break

3:30pm: Building Software ML Workloads with FireMarshal - Abe

4:00pm: Running a FireSim Simulation: End-to-End ML Workload – Vikram

4:30pm: Debugging and Profiling FireSim-Simulated Designs – Abe

4:55pm: Conclusion – Abe

5:00pm: End of Tutorial