CSE 560M Computer Systems Architecture I

Assignment 3

The goal of this assignment is to choose an interesting topic in computer architecture and explore it further. Research in the field is constantly evolving, and we want to learn something new and exciting about computer architecture and share it with the rest of the class.

This assignment is divided into three parts, which will be discussed next.

Part (a) – due Friday, Nov. 1, 2024

Look through the course topics and find one that piques your interest, something you'd like to dive deeper into. Let's explore further! Browse or search for a paper published in the last five years (2020-2024) in one of the ISCA, ASPLOS, or HPCA conferences that interests you. (In Part (b), you'll read the paper and post an executive summary on Piazza, but we'll cover that below.)

For convenience, we have provided links to the conference papers and a list of topics covered in each conference below in the Conference Paper Links and Conference Topics sections.

Make a Piazza post in the **hw3** folder with your paper selection. The post should **not** be anonymous so we can give credit, and it should be posted to the Entire Class. (After all, we'd like to learn what our peers are interested in, too!)

For full credit, you must giver the following information:

1. Which paper did you choose? (At minimum, include the title, year published, conference, and the URL to the paper.)
2. Why did you choose this paper? What drew your interest to it?

Part (b) – due Friday, Nov. 22, 2024

Read the paper and consider the key ideas the author(s) are communicating. Write an executive summary of those ideas and what you found interesting. The summary should be longer than an abstract but focused on the key ideas. We're not focusing on length, but rather the quality of it, and making it digestible for your peers. We want to learn from each other!

Make a new Piazza post in the **hw3** folder with your executive summary. Be sure to reference your original post by typing @ followed by the number of your previous post. (This links the two posts together.)

Again, this post should **not** be anonymous and should be posted to the Entire Class.

For full credit, your post must:

1. Include an executive summary of the paper that is longer than an abstract but focused on the key ideas.
2. Reference your original post that identifies the paper.
3. Be posted to the Entire Class.

Part (c) – due Tuesday, Dec. 3, 2024

There will be a peer review component for this assignment. Read at least three posts from your peers and respond to them in Piazza as part of a follow-up discussion. Again, these posts should **not** be anonymous.

Conference Paper Links

1. 2024 ACM/IEEE 51st International Symposium on Computer Architecture (ISCA)
   – https://ieeexplore.ieee.org/xpl/conhome/10609566/proceeding
2. 2023 ACM/IEEE 50th International Symposium on Computer Architecture (ISCA)
   – https://dl.acm.org/doi/proceedings/10.1145/3579371
3. 2022 ACM/IEEE 49th International Symposium on Computer Architecture (ISCA)
   – https://dl.acm.org/doi/proceedings/10.1145/3470496
4. 2021 ACM/IEEE 48th International Symposium on Computer Architecture (ISCA)
   – https://ieeexplore.ieee.org/xpl/conhome/9499716/proceeding
5. 2020 ACM/IEEE 47th International Symposium on Computer Architecture (ISCA)
   – https://ieeexplore.ieee.org/xpl/conhome/9136582/proceeding
6. ASPLOS '24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1
   – https://dl.acm.org/doi/proceedings/10.1145/3617232
7. ASPLOS '24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2
   – https://dl.acm.org/doi/proceedings/10.1145/3620665
8. ASPLOS '24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3
   – https://dl.acm.org/doi/proceedings/10.1145/3620666
9. ASPLOS '23: Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1
   – https://dl.acm.org/doi/proceedings/10.1145/3567955
10. ASPLOS '23: Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2
    – https://dl.acm.org/doi/proceedings/10.1145/3575693
11. ASPLOS '23: Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3
    – https://dl.acm.org/doi/proceedings/10.1145/3582016
12. ASPLOS '22: Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems
    – https://dl.acm.org/doi/proceedings/10.1145/3503222
13. ASPLOS '21: Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems
    – https://dl.acm.org/doi/proceedings/10.1145/3445814

14. ASPLOS '20: Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems
– https://dl.acm.org/doi/proceedings/10.1145/3373376

15. 2024 IEEE International Symp. on High-Performance Computer Architecture (HPCA)
– https://ieeexplore.ieee.org/xpl/conhome/10476359/proceeding

16. 2023 IEEE International Symp. on High-Performance Computer Architecture (HPCA)
– https://ieeexplore.ieee.org/xpl/conhome/10070856/proceeding

17. 2022 IEEE International Symp. on High-Performance Computer Architecture (HPCA)
– https://ieeexplore.ieee.org/xpl/conhome/9773179/proceeding

18. 2021 IEEE International Symp. on High-Performance Computer Architecture (HPCA)
– https://ieeexplore.ieee.org/xpl/conhome/9406784/proceeding

19. 2020 IEEE International Symp. on High Performance Computer Architecture (HPCA)
– https://ieeexplore.ieee.org/xpl/conhome/9044284/proceeding

Conference Topics

1. 2024 ACM/IEEE 51st International Symposium on Computer Architecture (ISCA) – `https://ieeexplore.ieee.org/xpl/conhome/10609566/proceeding`
    - Accelerators for Emerging Workloads I
    - Accelerators for Emerging Workloads II
    - Cloud Technologies
    - Compilers and Programming Models
    - Emerging Technologies
    - Industry Session
    - Machine Learning Accelerators I
    - Machine Learning Accelerators II
    - Memory Systems
    - Microarchitecture
    - NDP Technologies
    - Networking
    - PIM Accelerators
    - Parallel Architectures
    - Prefetching
    - Quantum Computing
    - Security
    - Tools and Analysis
2. 2023 ACM/IEEE 50th International Symposium on Computer Architecture (ISCA) – `https://dl.acm.org/doi/proceedings/10.1145/3579371`
    - CPU Microarchitecture
    - Cloud
    - Domain Specific Accelerators
    - Emerging - Edge Sensor
    - Emerging - Robotics
    - Emerging - Vision/Graphics/AR-VR
    - GPU
    - ML Accelerators and Scheduling
    - ML Systems
    - Memory Systems

- Quantum
- SSD
- Security

3. 2022 ACM/IEEE 49th International Symposium on Computer Architecture (ISCA) – `https://dl.acm.org/doi/proceedings/10.1145/3470496`
    - Datacenters and Sustainability
    - Embedded Systems and HW Synthesis
    - Graph Applications
    - Learning
    - Microarchitecture and Parallelism
    - Novel Architectures
    - Persistent Memory
    - Processing in Memory
    - Quantum
    - Security
4. 2021 ACM/IEEE 48th International Symposium on Computer Architecture (ISCA) – `https://ieeexplore.ieee.org/xpl/conhome/9499716/proceeding`
    - Accelerators
    - Compilers
    - DRAM/IO/Network
    - Data Center
    - Graph Processing
    - Low Temperature/Low Energy Computing
    - Machine Learning
    - Memory
    - Microarchitecture
    - Network Storage and Acceleration
    - Processing in/near Memory
    - Quantum/Photonics
    - Reliability and Security
    - Security
    - Sparse Processing
5. 2020 ACM/IEEE 47th International Symposium on Computer Architecture (ISCA) – `https://ieeexplore.ieee.org/xpl/conhome/9136582/proceeding`

- Accelerator-Based/Application-Specific Archs.
- Architectural Support for Machine Learning
- Coherence, Consistency, and Memory
- DRAM / Quantum Computing
- GPUs / Memory
- Microarchitecture
- Near-Data Processing / Processing-in-Memory
- Non-Traditional Computing / Graph Processing
- Non-Volatile / Persistent Memory and Storage
- Paralellism/IoT and Mobile
- Performance Evaluation / Virtualization
- Security

6. ASPLOS '24: Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems
   – `https://dl.acm.org/doi/proceedings/10.1145/3617232`
   – `https://dl.acm.org/doi/proceedings/10.1145/3620665`
   – `https://dl.acm.org/doi/proceedings/10.1145/3620666`

   - Accelerated Applications
   - Accelerators
   - Architecture Support for ML
   - Attacks and Mitigations
   - Binary Analysis
   - Bug Finding and Testing
   - Caching and Prefetching
   - Case Studies and Experiences
   - Compiler and Optimization Techniques
   - Cryptography and Privacy
   - Dynamic Analysis and Instrumentation
   - Emerging Non-Traditional Technologies
   - FPGAs and Reconfigurable Hardware
   - Graph Neural Networks
   - High Performance Systems
   - IoT and Embedded
   - ML Cluster Scheduling

- ML Inference Systems
- ML Quantization and Memory Optimizations
- ML Sparsity and Dynamic Shapes
- ML Systems and Optimizations
- ML Training Optimization
- Memory Optimizations
- Memory: Address Translation and Tiering
- Memory: Allocation and Management
- More Processing-In-Memory
- Optimization of Tensor Programs
- Optimizing ML Communication
- Power and Energy
- Processing-In-Memory (PIM) for ML
- Program and Configuration Syn
- Quantum Architecture
- Quantum Software
- SSDs
- Scheduling
- Security
- Serverless Computing 1
- Serverless Computing 2
- Side Channels
- Static Analysis and Verification
- Storage Optimization in Software
- Synthesis for Architectures
- Trusted Computing
- Variational Quantum Computing

7. ASPLOS '23: Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems
   – `https://dl.acm.org/doi/proceedings/10.1145/3567955`
   – `https://dl.acm.org/doi/proceedings/10.1145/3575693`
   – `https://dl.acm.org/doi/proceedings/10.1145/3582016`

   - Accelerators
   - Clouds/Data Centers

- Compiler Techniques and Optimization
- Debugging
- Deep Learning Systems
- Design Tools
- Disaggregated Memory
- Distributed Systems
- GPUs
- Graphs
- Hardware Security
- IoT/Embedded/Mobile
- Machine Learning
- Memory Mgmt./Near Data Processing
- Networking
- OS/Virtualization
- Persistence
- Quantum
- Shared Memory/Mem Consistency
- Software Security and Privacy
- Storage
- Sustainability
- Systems for ML
- Tensor Computation

8. ASPLOS '22: Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems
   – `https://dl.acm.org/doi/proceedings/10.1145/3503222`

   - Accelerating Emerging Applications
   - Accelerators
   - Address and Memory
   - Bugs
   - Data Center and Cloud Services
   - GPU and Data Analytics
   - Hardware Security
   - Non-traditional Computing and Reconfigurable Hardware
   - Operating Systems

- Privacy and Software Security
- Quantum Computing
- Serverless Computing
- Smart Networking
- Systems for Machine Learning

9. ASPLOS '21: Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems
   – `https://dl.acm.org/doi/proceedings/10.1145/3445814`

   - Better Hardware through Compilers
   - Beyond Neural Nets
   - Beyond the Pixels
   - Data Driven Optimization
   - Flow
   - Languages and Systems I
   - Languages and Systems II
   - Mapping and Management of Quantum and Cloud
   - Memory Systems
   - Microservices
   - Neural Net Optimization
   - Packet Up
   - Pages and Machine Architecture
   - Persistence I
   - Persistence II
   - Quantum Abstractions
   - Races and Concurrency
   - Robots, Optimization, and Robo-Optimization
   - Security I
   - Security II
   - Solid State Drives
   - Supporting Hardware Parallelism
   - Systems Software
   - Tools and Frameworks
   - Towards Improved Throughputs

10. ASPLOS '20: Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems
    – `https://dl.acm.org/doi/proceedings/10.1145/3373376`

    - ACID — Trippy!
    - Accelerators — Holding hands!
    - Automata — Hopping around the states.
    - Datacenter/cloud power/performance — Managing the beast.
    - Dynamic compilation — Who moved my cheese?
    - Edge/intermittent computing support — Life is too short!
    - Enclaves and memory security — Who
    - Evaluation techniques — Accelerating accelerator adoption.
    - Exotic architectures — Keep architecture weird!
    - Frameworks for deep learning — Layering the ML cake.
    - Huge memories and distributed databases — Now I remember!
    - Memory behavior — Where did I put it?
    - Memory management — I forgot.
    - Mobile/intermittent applications — Off and on again?
    - Persistence and correctness — Or... persistent correctness?
    - Persistent data structures — Keep all cats in mind!
    - Privacy and security in machine learning — In ML we trust???
    - Quantum computing — Who says you can't
    - SIMT — So many instructions, multiple tricks!
    - Security with little performance loss — Fast and furious!
    - Smart peripherals — Outside the box.
    - Speculation and consistency — Brain teasers.
    - Speculation and security — Life after meltdown.
    - Storage — Cache is the answer, what is the question?
    - Streaming computational models — In the flow!
    - Tensor computation and data
    - Virtualized acceleration — Don't keep it real!
    - Virtualized environments — Take the blue pill.

11. 2024 IEEE International Symp. on High-Performance Computer Architecture (HPCA)
    – `https://ieeexplore.ieee.org/xpl/conhome/10476359/proceeding`

    - Accelerator

- Accelerator (non-DNN)
- Accelerators & Memory Systems
- Best of CAL
- Cache & Memory Systems
- Computational SSD
- Datacenter & Networks
- Distributed DNN & Training
- Emerging Technology
- GNN
- GPU
- Industrial Track
- IoT/Edge
- Microarchitecture
- Near Memory Processing
- Reconfigurable Architecture & FPGA
- Rowhammer
- SSD
- Security
- Security-Side Channel
- Side-Channel & Microarchitecture

12. 2023 IEEE International Symp. on High-Performance Computer Architecture (HPCA) – `https://ieeexplore.ieee.org/xpl/conhome/10070856/proceeding`

    - Accelerators
    - Applications
    - Caching and Memory Management
    - Cloud and Edge Computing
    - Datacenters and HPC
    - Encryption and SGX
    - GPUs
    - Microarchitecture and Memory Systems
    - NICs and Networks
    - NVRAM and Hybrid Memory
    - Neural Networks and Accelerators
    - PIMs and Persistent Memory

- Quantum and FPGAs
- Reliability
- Security

13. 2022 IEEE International Symp. on High-Performance Computer Architecture (HPCA) – `https://ieeexplore.ieee.org/xpl/conhome/9773179/proceeding`
    - Accelerators
    - At Scale
    - Cache Hierarchy
    - Memory
    - Network on Chip
    - Non-Volatile Memory
    - Quantum
    - Security
    - Simulation
    - Storage, Scheduling, Interfaces
    - Synthesis
    - Traditional Architecture
14. 2021 IEEE International Symp. on High-Performance Computer Architecture (HPCA) – `https://ieeexplore.ieee.org/xpl/conhome/9406784/proceeding`
    - Accelerators for Machine Learning
    - Cache Design
    - Emerging Technologies and Applications
    - Hardware Accelerators Beyond Machine Learning
    - High Throughput Architectures
    - Memory and Storage Architectures
    - Network on Chip
    - Power Efficiency and Resiliency
    - Quantum Computing
    - Security Architectures
    - Security Attacks
    - Storage Systems
    - Systems for Machine Learning
15. 2020 IEEE International Symp. on High Performance Computer Architecture (HPCA) – `https://ieeexplore.ieee.org/xpl/conhome/9044284/proceeding`

- Accelerators and DSA
- Cloud
- Fault Tolerance and Security
- GPUs
- Machine Learning Acceleration
- Memory and Memory Hierarchy and Cloud
- Microarchitecture
- NoC
- Reliability and Fault Tolerance
- Security and NoC