Seminar in Computer Architecture

Lecture 1: Intro & Logistics

Dr. Mohammad Sadrosadati
mohammad.sadrosadati@safari.ethz.ch

ETH Zürich
Spring 2024
22 February 2024
Brief Self Introduction

Onur Mutlu

- Full Professor @ ETH Zurich ITET (INFK), since Sept 2015
- Strecker Professor @ Carnegie Mellon University ECE (CS), 2009-2016, 2016-...
- Started the Comp Arch Research Group @ Microsoft Research, 2006-2009
- Worked @ Google, VMware, Microsoft Research, Intel, AMD
- PhD in Computer Engineering from University of Texas at Austin in 2006
- BS in Computer Engineering & Psychology from University of Michigan in 2000
- https://people.inf.ethz.ch/omutlu/  omutlu@gmail.com

Research and Teaching in:

- **Computer architecture, systems, hardware security, bioinformatics**
- Memory and storage systems
- Robust & dependable hardware systems: security, safety, predictability, reliability
- Hardware/software cooperation
- New computing paradigms; architectures with emerging technologies/devices
- Architectures for bioinformatics, genomics, health, medicine, AI/ML
- ...

2
Current Research Mission

Computer architecture, HW/SW, systems, bioinformatics, security

Build fundamentally better architectures
Four Key Current Directions

- Fundamentally Secure/Reliable/Safe Architectures

- Fundamentally Energy-Efficient Architectures
  - Memory-centric (Data-centric) Architectures

- Fundamentally Low-Latency and Predictable Architectures

- Architectures for AI/ML, Genomics, Medicine, Health
The Transformation Hierarchy

Computer Architecture (expanded view)

- Problem
- Algorithm
- Program/Language
- System Software
- SW/HW Interface
- Micro-architecture
- Logic
- Devices
- Electrons

Computer Architecture (narrow view)
Computing System

Leiserson+, "There’s plenty of room at the Top: What will drive computer performance after Moore’s law?", Science, 2020

Richard Feynman, "There's Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics", a lecture given at Caltech, 1959.

Image source: https://science.sciencemag.org/content/368/6495/eaam9744
Software & Hardware Optimizations

### Multiplying Two 4096-by-4096 Matrices

```python
for i in xrange(4096):
    for j in xrange(4096):
        for k in xrange(4096):
            C[i][j] += A[i][k] * B[k][j]
```

<table>
<thead>
<tr>
<th>Implementation</th>
<th>Running time (s)</th>
<th>Absolute speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python</td>
<td>25,552.48</td>
<td>1x</td>
</tr>
<tr>
<td>Java</td>
<td>2,372.68</td>
<td>11x</td>
</tr>
<tr>
<td>C</td>
<td>542.67</td>
<td>47x</td>
</tr>
<tr>
<td>Parallel loops</td>
<td>69.80</td>
<td>366x</td>
</tr>
<tr>
<td>Parallel divide and conquer</td>
<td>3.80</td>
<td>6,727x</td>
</tr>
<tr>
<td>plus vectorization</td>
<td>1.10</td>
<td>23,224x</td>
</tr>
<tr>
<td>plus AVX intrinsics</td>
<td>0.41</td>
<td>62,806x</td>
</tr>
</tbody>
</table>

Leiserson+, "There’s plenty of room at the Top: What will drive computer performance after Moore’s law?", Science, 2020
Axiom

To achieve the highest energy efficiency and performance:

**we must take the expanded view**

of Computer Architecture

---

Co-design across the hierarchy:
Algorithms to devices

Specialize as much as possible within the design goals
Current Research Mission & Major Topics

**Build fundamentally better architectures**

- **Data-centric arch. for low energy & high perf.**
  - Proc. in Mem/DRAM, NVM, unified mem/storage
- **Low-latency & predictable architectures**
  - Low-latency, low-energy yet low-cost memory
  - QoS-aware and predictable memory systems
- **Fundamentally secure/reliable/safe arch.**
  - Tolerating all bit flips; patchable HW; secure mem
- **Architectures for ML/AI/Genomics/Graph/Med**
  - Algorithm/arch./logic co-design; full heterogeneity
- **Data-driven and data-aware architectures**
  - ML/AI-driven architectural controllers and design
  - Expressive memory and expressive systems
Onur Mutlu’s SAFARI Research Group

Computer architecture, HW/SW, systems, bioinformatics, security, memory

Think BIG, Aim HIGH!
SAFARI Newsletter June 2023 Edition

- https://safari.ethz.ch/safari-newsletter-june-2023/
https://safari.ethz.ch/safari-newsletter-december-2021/
Dear SAFARI friends,

Happy New Year! We are excited to share our group highlights with you in this second edition of the SAFARI newsletter (You can find the first edition from April 2020 here). 2020 has
Dear SAFARI friends,

2019 and the first three months of 2020 have been very positive eventful times for SAFARI.
SAFARI PhD and Post-Doc Alumni

https://safari.ethz.ch/safari-alumni/

- Hasan Hassan (ETH Zurich), S&P 2020 Best Paper Award, 2020 Pwnie Award, IEEE Micro Top Picks HM 2020
- Christina Giannoula (National Technical University of Athens),
- Minesh Patel (ETH Zurich), MICRO 2020 and DSN 2020 Best Paper Awards; ISCA Hall of Fame 2021
- Damla Senol Cali (Bionano Genomics), SRC TECHCON 2019 Best Student Presentation Award, RECOMB-Seq 2018 Best Poster Award
- Nastaran Hajinazar (ETH Zurich)
- Gagandeep Singh (ETH Zurich), FPL 2020 Best Paper Award Finalist
- Amirali Boroumand (Stanford Univ → Google), SRC TECHCON 2018 Best Student Presentation Award
- Jeremie Kim (ETH Zurich), EDAA Outstanding Dissertation Award 2020; IEEE Micro Top Picks 2019; ISCA/MICRO HoF 2021
- Nandita Vijaykumar (Univ. of Toronto, Assistant Professor), ISCA Hall of Fame 2021
- Kevin Hsieh (Microsoft Research, Senior Researcher)
- Justin Meza (Facebook), HiPEAC 2015 Best Student Presentation Award; ICCD 2012 Best Paper Award
- Mohammed Alser (ETH Zurich), IEEE Turkey Best PhD Thesis Award 2018
- Yixin Luo (Google), HPCA 2015 Best Paper Session
- Kevin Chang (Facebook), SRC TECHCON 2016 Best Student Presentation Award
- Rachata Ausavarungnirun (KMUNTB, Assistant Professor), NOCS 2015 and NOCS 2012 Best Paper Award Finalist
- Gennady Pekhimenko (Univ. of Toronto, Assistant Professor), ISCA Hall of Fame 2021; ASPLOS 2015 SRC Winner
- Vivek Seshadri (Microsoft Research)
- Donghyuk Lee (NVIDIA Research, Senior Researcher), HPCA Hall of Fame 2018
- Yoongu Kim (Software Robotics → Google), TCAD’19 Top Pick Award; IEEE Micro Top Picks’10; HPCA’10 Best Paper Session
- Lavanya Subramanian (Intel Labs → Facebook)
- Samira Khan (Univ. of Virginia, Assistant Professor), HPCA 2014 Best Paper Session
- Saugata Ghose (Univ. of Illinois, Assistant Professor), DFRWS-EU 2017 Best Paper Award
- Jawad Haj-Yahya (Huawei Research Zurich, Principal Researcher)
Onur Mutlu,
"SAFARI Research Group: Introduction & Research"
Invited Talk at the ETH Future Computing Laboratory
Huawei Day, Virtual, 19 October 2021.
[Slides (pptx) (pdf)]
[Talk Video (15 minutes)]
Short Video on SAFARI Research Group

SAFARI Research Group

Introduction & Research

Onur Mutlu
omutlu@gmail.com
https://people.inf.ethz.ch/omutlu
19 October 2021
EFCL Huawei Day

SAFARI Research Group: Introduction & Research -- ETH Future Computing Laboratory Talk - Onur Mutlu

529 views • Premiered Jan 15, 2022

https://www.youtube.com/watch?v=mSr1QQmYuX0
Our Major Courses & Lectures

- **Computer Architecture & Digital Design Course**
  - Spring 2023 Livestream Edition:
    [https://www.youtube.com/playlist?list=PL5Q2soXY2Zi-EImKxYYY1SZuGiOAOBKaf](https://www.youtube.com/playlist?list=PL5Q2soXY2Zi-EImKxYYY1SZuGiOAOBKaf)

- **Advanced Computer Architecture Course**
  - Fall 2023 Livestream Edition:
    [https://www.youtube.com/playlist?list=PL5Q2soXY2Zi9UFWwfRtSjpKX_IRF0lPSe](https://www.youtube.com/playlist?list=PL5Q2soXY2Zi9UFWwfRtSjpKX_IRF0lPSe)

- **Seminar in Computer Architecture**
  - Fall 2023 Livestream Edition:
    - [https://www.youtube.com/playlist?list=PL5Q2soXY2Zi9w66HWHS7UQ0zK4qOl.IMU](https://www.youtube.com/playlist?list=PL5Q2soXY2Zi9w66HWHS7UQ0zK4qOl.IMU)

[SAFARI](https://www.youtube.com/onurmutlulectures)
Accessing All Our Courses

Spring 2024:
- Digital Design and Computer Architecture
- Seminar in Computer Architecture
- SAFARI Project & Seminars courses

Fall 2023:
- Computer Architecture
- Seminar in Computer Architecture
- SAFARI Project & Seminars courses

Spring 2023:
- Digital Design and Computer Architecture
- Seminar in Computer Architecture
- SAFARI Project & Seminars courses

Fall 2022:
- Computer Architecture
- Seminar in Computer Architecture
- SAFARI Project & Seminars courses

Spring 2022:
- Digital Design and Computer Architecture
- Seminar in Computer Architecture
- SAFARI Project & Seminars courses

Fall 2021:
- Computer Architecture
- Seminar in Computer Architecture
- SAFARI Project & Seminars courses

https://safari.ethz.ch/courses
Hands-On Project Courses

- https://safari.ethz.ch/projects_and_seminars/doku.php

SAFARI Project & Seminars Courses (Spring 2024)

Welcome to the wiki for Project and Seminar courses SAFARI offers.

Courses we offer:

- FPGA-based Exploration of DRAM and RowHammer: 227-0085-35L
- Accelerating Genome Analysis with FPGAs, GPUs, and New Execution Paradigms: 227-0085-33L
- Genomic Sequencing on Mobile Devices: 227-0085-36L
- Understanding and Designing Modern NAND Flash-Based SSDs (Solid-State Drives): 227-0085-44L

SAFARI Projects & Seminars Courses (Spring 2024)
SAFARI Live Seminars

SAFARI Live Seminars in Computer Architecture
Dr. Juan Gómez Luna, ETH Zurich
Understanding a Modern Processing-in-Memory Architecture: Benchmarking and Experimental Characterization

SAFARI Live Seminars in Computer Architecture
Dr. Andrew Walker, Schiltron Corporation & Nexgen Power Systems
An Addiction to Low Cost Per Memory Bit – How to Recognize It and What to Do About It

SAFARI Live Seminars in Computer Architecture
Gernady Pekshemen, University of Toronto
Efficient DNN Training at Scale: from Algorithms to Hardware

SAFARI Live Seminars in Computer Architecture
Jawad Haj-Yahya, Huawei Research Center Zurich
Power Management Mechanisms in Modern Microprocessors and Their Security Implications

Overview of a Modern SoC Architecture
- 3 domains in modern thermally constrained mobile SoC (CPU, Memory, IO)
- Several voltage sources exist, and some of them are shared between domains
- IO controllers and engines, IO interconnect, memory controller, and DDR4 typically each has an independent clock

SAFARI Live Seminars in Computer Architecture
Ataberk Olgun, TOBB & ETH Zurich
QUAC-PINQ: High-Throughput True Random Number Generation Using Quadruple Row Activation in Connects DRAM Chips

SAFARI Live Seminars in Computer Architecture
Minaw Patel, ETH Zurich
Enabling Effective Error Mitigation in Memory Chips That Use On-Die ECCs

SAFARI Live Seminars in Computer Architecture
Christina Giannoule, National Technical University of Athens
Efficient Synchronization Support for Near-Data-Processing Architectures

SAFARI Live Seminars in Computer Architecture
Jawad Haj-Yahya, Huawei Research Center Zurich

https://safari.ethz.ch/safari-seminar-series/
Open Source Artifacts: SAFARI GitHub

SAFARI Research Group at ETH Zurich and Carnegie Mellon University
Site for source code and tools distribution from SAFARI Research Group at ETH Zurich and Carnegie Mellon University.

 ETH Zurich and Carnegie Mellon ...
 https://safari.ethz.ch/
 omutlu@gmail.com

Pinned

**ramulator**
A Fast and Extensible DRAM Simulator, with built-in support for modeling many different DRAM technologies including DDRx, LPDDRx, GDDRx, WIOx, HBMx, and various academic proposals. Described in the...

**prim-benchmarks**
PrIM (Processing-In-Memory benchmarks) is the first benchmark suite for a real-world processing-in-memory (PIM) architecture. PrIM is developed to evaluate, analyze, and characterize the first publ...

**DAMOV**
DAMOV is a benchmark suite and a methodical framework targeting the study of data movement bottlenecks in modern applications. It is intended to study new architectures, such as near-data processin...

Repositories

Find a repository...

**Pythia**
A Customizable Hardware Prefetching Framework Using Online Reinforcement Learning.

**BurstLink**

https://github.com/CMU-SAFARI/
Papers, Courses, Talks, Videos, Artifacts

- All are available at
  
  [https://people.inf.ethz.ch/omutlu/projects.htm](https://people.inf.ethz.ch/omutlu/projects.htm)

  [https://www.youtube.com/onurmutlulectures](https://www.youtube.com/onurmutlulectures)

  [https://github.com/CMU-SAFARI/](https://github.com/CMU-SAFARI/)
Some Basic Principles We Follow
Principle: Teaching and Research

Teaching drives Research  
Research drives Teaching
Principle: Insight and Ideas

Focus on Insight
Encourage New Ideas
Principle: Learning and Scholarship

Focus on learning and scholarship
Principle: Environment of Freedom

Create an environment that values free & critical exploration, openness, collaboration, hard work, creativity.
Principle: Learning and Scholarship

The quality of your work defines your impact
Principle: Good Mindset, Goals & Focus

You can make a good impact on the world
Suggestion: Principle: Passion

Follow Your Passion
(Do not get derailed by naysayers)
Principle: Build Infrastructure

Build Infrastructure to Enable Your Passion
Principle: Work Hard

Work Hard to Enable Your Passion
Suggestion: Principle: Resilience & Focus

Be Resilient & Focused

Make It Happen
You can make a good impact on the world
Richard Hamming

``You and Your Research''

Transcription of the
Bell Communications Research Colloquium Seminar
7 March 1986

If you really want to be a first-class scientist you need to know yourself, your weaknesses, your strengths, and your bad faults, like my egotism. How can you convert a fault to an asset? How can you convert a situation where you haven't got enough manpower to move into a direction when that's exactly what you need to do? I say again that I have seen, as I studied the history, the successful scientist changed the viewpoint and what was a defect became an asset.

In summary, I claim that some of the reasons why so many people who have greatness within their grasp don't succeed are: they don't work on important problems, they don't become emotionally involved, they don't try and change what is difficult to some other situation which is easily done but is still important, and they keep giving themselves alibis why they don't. They keep saying that it is a matter of luck. I've told you how easy it is; furthermore I've told you how to reform. Therefore, go forth and become great scientists!

The Role of This Course
Computer Architecture

- is the **science** and **art** of designing **computing platforms** (hardware, interface, system SW, and programming model)

- to achieve a set of **design goals**
  - E.g., highest performance on earth on workloads X, Y, Z
  - E.g., longest battery life at a form factor that fits in your pocket with cost < $$$ CHF
  - E.g., best average performance across all known workloads at the best performance/cost ratio
  - ...

- Designing a supercomputer is different from designing a smartphone → But, many fundamental principles are similar
Different Platforms, Different Goals

Source: http://www.sia-online.org (semiconductor industry association)
Different Platforms, Different Goals

iPhone 14 Pro

VS

iPhone 11 Pro

https://www.phonearena.com/reviews/iPhone-14-pro-max-vs-iPhone-13-pro-max_id5373
Different Platforms, Different Goals

Source: https://iq.intel.com/5-awesome-uses-for-drone-technology/
Different Platforms, Different Goals
Different Platforms, Different Goals
Different Platforms, Different Goals
Different Platforms, Different Goals

- ML accelerator: 260 mm$^2$, 6 billion transistors, 600 GFLOPS GPU, 12 ARM 2.2 GHz CPUs.
- Two redundant chips for better safety.

https://youtu.be/Ucp0TTmvqOE?t=4236
Different Platforms, Different Goals

Why Study Computer Architecture?

- Enable better systems: make computers faster, cheaper, smaller, more reliable, ...
  - By exploiting advances and changes in underlying technology/circuits

- Enable new applications
  - Life-like 3D visualization 20 years ago? Virtual reality?
  - Self-driving cars?
  - Personalized genomics? Personalized medicine?

- Enable better solutions to problems
  - Software innovation is built on trends and changes in computer architecture
    - > 50% performance improvement per year has enabled this innovation

- Understand why computers work the way they do
Today is a very exciting time to study computer architecture.

Industry is in a large paradigm shift (to novel architectures) – many different potential system designs possible.

Many difficult problems motivating and caused by the shift:
- Huge hunger for data and new data-intensive applications
- Power/energy/thermal constraints
- Complexity of design
- Difficulties in technology scaling
- Memory bottleneck
- Reliability problems
- Programmability problems
- Security and privacy issues

No clear, definitive answers to these problems.
Seminar in Computer Architecture

- We will cover **fundamental** and **cutting-edge** research papers in computer architecture

- Multiple components that are aimed at improving students’
  - technical skills in computer architecture
  - critical thinking and analysis
  - technical presentation of concepts and papers
    - in both spoken and written forms
  - familiarity with key research directions
Key Goal

(Learn how to) rigorously analyze, present, discuss papers and ideas in computer architecture
Steps to Achieve the Key Goal

- **Steps for the Presenter**
  - Read
  - Absorb, read more (other related works)
  - Critically analyze; think; synthesize
  - Prepare a clear and rigorous talk
  - Present
  - Answer questions
  - Analyze and synthesize (in meeting, after, and at course end)

- **Steps for the Participants**
  - Discuss
  - Ask questions
  - Analyze and synthesize (in meeting, after, and at course end)
Topics of Papers and Discussion

- hardware security;
- architectural acceleration mechanisms for key applications like machine learning, graph processing, and bioinformatics;
- memory systems;
- interconnects;
- processing inside memory;
- various fundamental and emerging ideas/paradigms in computer architecture;
- hardware/software co-design and cooperation;
- fault tolerance;
- energy efficiency;
- heterogeneous and parallel systems;
- new execution models, etc.
Recap: Some Goals of This Course

- **Teach/enable/empower you to:**
  - Think critically
  - Think broadly
  - Learn how to understand, analyze, and present papers and ideas
  - Get familiar with key first steps in research
  - Get familiar with key research directions
The Virtuous Cycle of Scientific Progress

- Read/critique papers; understand problems
- Question, Brainstorm
- Do great research and publish (educate others and more)
- Collaborate, work hard
- Develop new, out-of-the-box ideas

(educate others and more)
Course Info and Logistics
Onur Mutlu

- Full Professor @ ETH Zurich ITET (INFK), since Sept 2015
- Strecker Professor @ Carnegie Mellon University ECE (CS), 2009-2016, 2016-...
- Started the Comp Arch Research Group @ Microsoft Research, 2006-2009
- Worked @ Google, VMware, Microsoft Research, Intel, AMD
- PhD in Computer Engineering from University of Texas at Austin in 2006
- BS in Computer Engineering & Psychology from University of Michigan in 2000
- https://people.inf.ethz.ch/omutlu/  omutlu@gmail.com

Research and Teaching in:
- Computer architecture, systems, hardware security, bioinformatics
- Memory and storage systems
- Robust & dependable hardware systems: security, safety, predictability, reliability
- Hardware/software cooperation
- New computing paradigms; architectures with emerging technologies/devices
- Architectures for bioinformatics, genomics, health, medicine, AI/ML
- ...
Lecturers

Mohammad Sadrosadati
Senior Researcher and Lecturer
Heterogeneous computing | Processing-In-Memory | Memory Systems | Interconnection Networks

Yu (Lenny) Liang
Senior Researcher
Mobile Storage Systems | Hardware/Software Cooperation | Storage for Genomics | Processing-In-Memory
Course Info: Who Are We?

- **Teaching Assistants/Mentors**
  - Dr. Haiyu Mao,
  - Ataberk Olgun,
  - Abdullah Giray Yaglikci,
  - Can Firtina,
  - Geraldo Francisco De Oliveira Junior,
  - Rahul Bera,
  - Konstantinos Kanellopoulos,
  - Nika Mansouri Ghiasi,
  - Nisa Bostanci,
  - Rakesh Nadig,
  - Joël Lindegger,
  - Ismail Emir Yüksel
  - Haocong Luo,
  - Banu Cavlak,
  - Zulal Bingol

- Get to know them and their research as they will be your mentors
  https://safari.ethz.ch/group-members/
Course Requirements and Expectations

- Attendance required for all meetings (preferably in-person)
- Each student presents one paper
  - Prepare for presentation with engagement from the mentors
  - Full presentation + questions + discussion

- Non-presenters participate during the meeting
  - Ask questions, contribute thoughts/ideas
  - Better if you read/skim the paper beforehand

- Non-presenters take an online short quiz after each session
  - 3 MCQs for each presentation (1 Full day to submit)

- Everyone comments on papers in the online review system
  - After presentation

- Write synthesis report at the end of semester
  - (sample synthesis report online)
Course Website

- https://safari.ethz.ch/architecture_seminar/spring2024

- All course materials to be posted

- Plus other useful information for the course

- Check frequently for announcements and due dates
Moodle


- Check frequently for:
  - Announcements and due dates
  - Quizzes
  - Paper discussion
  - Assignment submissions
Homework 0: Student Information

- Due on March 7
- [https://safari.ethz.ch/architecture_seminar/spring2024/doku.php?id=homework](https://safari.ethz.ch/architecture_seminar/spring2024/doku.php?id=homework)
- Information about yourself
- All future grading is predicated on homework 0
- If it is not submitted on time, we cannot schedule you for a presentation.
Homework 1: Analyzing a Paper

- Due on March 7
- [https://safari.ethz.ch/architecture_seminar/spring2024/doku.php?id=homework](https://safari.ethz.ch/architecture_seminar/spring2024/doku.php?id=homework)

- Review and write a critical analysis of a paper

- It is mandatory to submit the review
Paper Review Preferences

- Due TBD

- Check the website and Moodle for instructions

- If it is not submitted on time, we cannot schedule you for a presentation.
Synthesis Report

- Due TBD (Usually 1 month after the last lecture)

- Answering questions based on everything you have read, seen, listened to, and experienced in the course during the entire semester.

- It is mandatory to submit the review
How to Deliver a Good Talk
Anatomy of a Good Paper Review (Talk)

- **0: Title, Authors, Venue**
- **1: Summary**
  - What is the problem the paper is trying to solve?
  - What are the key ideas of the paper? Key insights?
  - What are the key mechanisms? What is the implementation?
  - What are the key results? Key conclusions?
- **2: Strengths** (most important ones)
  - Does the paper solve the problem well? Is it well written? ...
- **3: Weaknesses** (most important ones)
  - This is where you should think critically. Every paper/idea has a weakness. This does not mean the paper is necessarily bad. It means there is room for improvement and future research can accomplish this.
- **4: Thoughts/Ideas**: Can you do better? Present your ideas.
- **5: Takeaways**: What you learned/enjoyed/disliked? Why?
- **6: Discussion starters and questions**.
- Review should be short and concise (20 minutes or < one page)
Executive Summary

Motivation:

- Repeatedly toggling a DRAM row’s wordline voltage causes bit flips in nearby rows
- This vulnerability, RowHammer, worsens in denser DRAM chips
- Understanding RowHammer enables designing effective and efficient solutions

Problem: No study demonstrates how wordline voltage (VPP) affects RowHammer

Goal: Experimentally understand how VPP affects RowHammer and DRAM operation

Experimental study: 272 DRAM chips from three major DRAM manufacturers

VPP’s effect on RowHammer: Six observations show that with reduced VPP,

- Bit error rate caused by a RowHammer attack reduces by 15.2% (66.9% max)
- A row needs to be activated 7.4% more times (85.8% max) to induce the first bit flip

VPP’s effect on DRAM operation: Nine observations show that with reduced VPP,

- 208 out of 272 tested DRAM chips reliably operate using nominal timing parameters
- Erroneous DRAM chips can reliably operate with
- A longer row activation latency, i.e., 24ns/15ns for 48/16 chips,
- Single-error-correcting codes or 2x the refresh rate only for 16.4% of rows

Conclusion: Reducing wordline voltage can reduce RowHammer vulnerability without significantly affecting reliable DRAM operation

SAFARI
Executive Summary

**Problem:** DRAM on-die ECC *complicates* third-party reliability studies
- Proprietary design *obfuscates* raw bit errors in an *unpredictable* way
- *Interferes* with (1) design, (2) test & validation, and (3) characterization

**Goal:** understand *exactly how* on-die ECC obfuscates errors

**Contributions:**
1. **BEER:** new testing methodology that determines a DRAM chip’s *unique* on-die ECC function (i.e., its parity-check matrix)
   - Exploits *ECC-function-specific* uncorrectable error patterns
   - Requires no hardware support, inside knowledge, or metadata access
2. **BEEP:** new error profiling methodology that infers the *raw bit error* locations of error-prone cells from the *observable uncorrectable errors*

**BEER Evaluations:**
- Apply BEER to 80 real LPDDR4 chips from 3 major DRAM manufacturers
- Show correctness in simulation for 115,300 codes (4-247b ECC words)

We hope BEER and BEEP enable valuable studies in the future

SAFARI
Suggested Paper Discussion Format

- Problem & Goal
- Key Ideas/solution
- Novelty
- Mechanisms & Implementation
- Major Results
- Takeaways/Conclusions

- Strengths
- Weaknesses
- Alternatives
- New ideas/problems
- Brainstorming and Discussion

~15-20 minute Summary

~10 min Critique plus
~20 min Discussion

SAFARI
More Advice on Paper Review/Talk

- When doing the paper reviews and analyses, be very critical

- Always think about better ways of solving the problem or related problems
  - Question the problem as well
  - Read background papers (both past and future)

- This is how things progress in science and engineering (or anywhere), and how you can make big leaps
  - By critical analysis

- A few sample text reviews provided online
Try to Avoid Rat Hole Discussions

Performance Analysis Rat Holes

Workload  Metrics  Configuration Details

Source: https://www.cse.wustl.edu/~jain/iucee/ftp/k_10adp.pdf
Try to Avoid Rat Hole Discussions

Aside: A Recommended Book

Even if the performance analysis is correctly done and presented, it may not be enough to persuade your audience—the decision makers—to follow your recommendations. The list shown in Box 10.2 is a compilation of reasons for rejection heard at various performance analysis presentations. You can use the list by presenting it immediately and pointing out that the reason for rejection is not new and that the analysis deserves more consideration. Also, the list is helpful in getting the competing proposals rejected!

There is no clear end of an analysis. Any analysis can be rejected simply on the grounds that the problem needs more analysis. This is the first reason listed in Box 10.2. The second most common reason for rejection of an analysis and for endless debate is the workload. Since workloads are always based on the past measurements, their applicability to the current or future environment can always be questioned. Actually workload is one of the four areas of discussion that lead a performance presentation into an endless debate. These “rat holes” and their relative sizes in terms of time consumed are shown in Figure 10.26. Presenting this cartoon at the beginning of a presentation helps to avoid these areas.

**Figure 10.26** Four issues in performance presentations that commonly lead to endless discussion.

Box 10.2 Reasons for Not Accepting the Results of an Analysis

1. This needs more analysis.
2. You need a better understanding of the workload.
3. It improves performance only for long I/O’s, packets, jobs, and files, and most of the I/O’s, packets, jobs, and files are short.
4. It improves performance only for short I/O’s, packets, jobs, and files, but who cares for the performance of short I/O’s, packets, jobs, and files; it’s the long ones that impact the system.
5. It needs too much memory/CPU/bandwidth and memory/CPU/bandwidth isn’t free.
6. It only saves us memory/CPU/bandwidth and memory/CPU/bandwidth is cheap.
7. There is no point in making the networks (similarly, CPUs/disks/…) faster; our CPUs/disks (any component other than the one being discussed) aren’t fast enough to use them.
8. It improves the performance by a factor of $x$, but it doesn’t really matter at the user level because everything else is so slow.
9. It is going to increase the complexity and cost.
10. Let us keep it simple stupid (and your idea is not stupid).
11. It is not simple. (Simplicity is in the eyes of the beholder.)
12. It requires too much state.
13. Nobody has ever done that before. (You have a new idea.)
14. It is not going to raise the price of our stock by even an eighth. (Nothing ever does, except rumors.)
15. This will violate the IEEE, ANSI, CCITT, or ISO standard.
16. It may violate some future standard.
17. The standard says nothing about this and so it must not be important.
18. Our competitors don’t do it. If it was a good idea, they would have done it.
19. Our competition does it this way and you don’t make money by copying others.
20. It will introduce randomness into the system and make debugging difficult.
21. It is too deterministic; it may lead the system into a cycle.
22. It’s not interoperable.
23. This impacts hardware.
24. That’s beyond today’s technology.
25. It is not self-stabilizing.
26. Why change—it’s working OK.
More Advice on Talks

- Kayvon Fatahalian, “Tips for Giving Clear Talks”
  - [http://graphics.stanford.edu/~kayvonf/misc/cleartalktips.pdf](http://graphics.stanford.edu/~kayvonf/misc/cleartalktips.pdf)
  - Many useful and simple principles here

  "Every sentence matters"

  "The audience prefers not to think" (about things you can just tell them)

  "Surprises are bad": say why before what
  (indicate why you are saying something before you say it)

  Explain every figure, graph, or equation

  When improving the talk, the audience is always right
How to Participate
How to Make the Best Out of This?

- Come prepared → Read and critically evaluate the paper
- Think new ideas
- Bring discussion points and questions; read other papers
- Be critical
- Brainstorm – be open to new ideas
- Pay attention and discuss+contribute
- Participate online before and after each meeting
Guided Talk Preparation
Preparing a Talk

1. Check your presentation date
2. Study your paper(s)
3. Create draft presentation
4. Meet advisor, get feedback
Preparing a Talk: Start Early

- Preparing a good presentation takes time
- Start early!

1. Check your presentation date
2. Study your paper(s)
3. Create draft presentation
4. Meet advisor, get feedback
Preparing a Talk: Study Paper

- **Check your presentation date**
- **Study your paper(s)**
- **Create draft presentation**
- **Meet advisor, get feedback**

- **3 ‘C’s of reading**
  - *Carefully*: look up terms, possibly read cited papers
  - *Critically*: find limitations, flaws
  - *Creatively*: think of improvements

- **Try examples by hand**
- **Try tools if available**

- **Consult with mentors if questions**
Preparing a Talk: Create Draft

- Explain the motivation for the work
- Clearly present the technical solution and results
  - Include a demo if appropriate
- Outline limitations or improvements
- Focus on the key concepts
  - Do not present all of the details

1. Check your presentation date
2. Study your paper(s)
3. Create draft presentation
4. Meet advisor, get feedback
Preparing a Talk: Get Feedback

- Prepare for the meeting
  - Schedule early
  - Send slides in advance
  - Write down questions

- Make sure you address feedback
  - Take notes

- Meetings are mandatory!
  - At least one week before the talk
  - Two meetings
Grading and Feedback
Grading Rubric

- Quality of your presentation (60%)
  - How well did you understand the material?
  - How well did you present it?
  - How well did you answer the questions?
  - Be prepared to explain technical terms
  - We will take into account the difficulty of the paper and the time you had to prepare.

- Quality of the final synthesis paper (30%)
  - How well did you understand some of the papers presented during the seminar?

- Attendance & Quizzes (10%)

- Participation (during class and online) (BONUS 10%)
  - Did you ask good questions?
  - Did you participate and contribute to the discussion?
Feedback

- We will try to (briefly) discuss strengths/weaknesses of your talk in class
  - Let us know upfront if you would prefer **not** to

- You can arrange a meeting with your TA to get feedback
Expected Schedule
Schedule

- We will meet once a week, with two presentations per session
  - Next meeting is next week
  - Your presentations start on March 14
  - 2-3 presentations per session
  - Each presentation: 50 minutes including questions & discussion

- Paper assignment
  - Will be done online
  - Study the list of papers
  - Check your email and be responsive
Homework 0: Student Information

- Due on March 7
- https://safari.ethz.ch/architecture_seminar/spring2024/doku.php?id=homework

- Information about yourself

- All future grading is predicated on homework 0

- If it is not submitted on time, we cannot schedule you for a presentation.
Homework 1: Analyzing a Paper

- Due on March 7
- https://safari.ethz.ch/architecture_seminar/spring2024/doku.php?id=homework

- Review and write a critical analysis of a paper

- It is mandatory to submit the review
Paper Review Preferences

- Due TBD

- Check the website and Moodle for instructions

- If it is not submitted on time, we cannot schedule you for a presentation.
Synthesis Report

- Due TBD (Usually 1 month after the last lecture)

- Answering questions based on everything you have read, seen, listened to, and experienced in the course during the entire semester.

- It is mandatory to submit the review
How to Approach This Course?
“Formative Experience”
“Reading and analyzing papers will help us a lot into the future”
How to Approach This Course

“High investment, high return”
“Guidance from 2 top researchers in the field”
How to Approach This Course

“I would definitely recommend this course”
“How to Approach This Course

“I really love Computer Architecture”
How to Approach This Course

Learning experience
Long-term tradeoff analysis
Critical thinking & decision making
Synergistic Activities
An Interview on Computing Futures

https://www.youtube.com/watch?v=8ffSEKZhmvo
Short Video on SAFARI Research Group


SAFARI Research Group

Introduction & Research

Onur Mutlu
omutlu@gmail.com
https://people.inf.ethz.ch/omutlu
19 October 2021
EFCL Huawei Day

SAFARI
ETH zürich
Carnegie Mellon

https://www.youtube.com/watch?v=mSr1QQmYuX0
Research & Teaching: Some Overview Talks

https://www.youtube.com/onurmutlulectures

- Future Computing Architectures
  - [https://www.youtube.com/watch?v=kgiZISOcGFM&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=1](https://www.youtube.com/watch?v=kgiZISOcGFM&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=1)

- Enabling In-Memory Computation
  - [https://www.youtube.com/watch?v=njX_14584Jw&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=16](https://www.youtube.com/watch?v=njX_14584Jw&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=16)

- Accelerating Genome Analysis
  - [https://www.youtube.com/watch?v=r7sn41lH-4A&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=41](https://www.youtube.com/watch?v=r7sn41lH-4A&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=41)

- Rethinking Memory System Design
  - [https://www.youtube.com/watch?v=F7xZLNMIY1E&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=3](https://www.youtube.com/watch?v=F7xZLNMIY1E&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=3)

- Intelligent Architectures for Intelligent Machines
  - [https://www.youtube.com/watch?v=c6_LgzuNdkw&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=25](https://www.youtube.com/watch?v=c6_LgzuNdkw&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=25)

- The Story of RowHammer
  - [https://www.youtube.com/watch?v=sgd7PHQQ1AI&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=39](https://www.youtube.com/watch?v=sgd7PHQQ1AI&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=39)
An Interview on Research and Education

- **Computing Research and Education (@ ISCA 2019)**
  - https://www.youtube.com/watch?v=8ffSEKZhmvvo&list=PL5Q2soXY2Zi_4oP9LdL3cc8G6NIjD2Ydz

- **Maurice Wilkes Award Speech (10 minutes)**
  - https://www.youtube.com/watch?v=tcQ3zZ3JpuA&list=PL5Q2soXY2Zi8D_5MGV6EnXEJHnV2YFBJl&index=15
More Thoughts and Suggestions

- Onur Mutlu,
  "Some Reflections (on DRAM)"
  Award Speech for ACM SIGARCH Maurice Wilkes Award, at the ISCA Awards Ceremony, Phoenix, AZ, USA, 25 June 2019.
  [Slides (pptx) (pdf)]
  [Video of Award Acceptance Speech (Youtube; 10 minutes) (Youku; 13 minutes)]
  [Video of Interview after Award Acceptance (Youtube; 1 hour 6 minutes) (Youku; 1 hour 6 minutes)]
  [News Article on "ACM SIGARCH Maurice Wilkes Award goes to Prof. Onur Mutlu"]

- Onur Mutlu,
  "How to Build an Impactful Research Group"
  57th Design Automation Conference Early Career Workshop (DAC), Virtual, 19 July 2020.
  [Slides (pptx) (pdf)]
More Thoughts and Suggestions (II)

- Onur Mutlu,
  "Computer Architecture: Why Is It So Important and Exciting Today?"
  Invited Lecture at Izmir Institute of Technology (IYTE), Virtual, 16 October 2020.
  [Slides (pptx) (pdf)]
  [Talk Video (2 hours 12 minutes)]

- Onur Mutlu,
  "Applying to Graduate School & Doing Impactful Research"
  Invited Panel Talk at the 3rd Undergraduate Mentoring Workshop, held with the 48th International Symposium on Computer Architecture (ISCA), Virtual, 18 June 2021.
  [Slides (pptx) (pdf)]
  [Talk Video (50 minutes)]
Seminar in Computer Architecture
Lecture 1: Intro & Logistics

Dr. Mohammad Sadrosadati
mohammad.sadrosadati@safari.ethz.ch

ETH Zürich
Spring 2024
22 February 2024