

Chao Chen

Amazon Science

✉ chao.chen@gatech.edu

🌐 <http://cchen435.com>

📞 +1(806) 620-8598

RESEARCH INTERESTS

My research interest lies in the intersection of compiler and computing system. I am particularly interested in applying compiler techniques to improve systems' dynamic properties, e.g., resilience and scheduling.

EDUCATION

Georgia Institute of Technology, Atlanta, GA

Ph.D. in Computer Science

2020/12

Thesis: "Compiler-Assisted Resilience Framework for Recovery from Transient Faults"

Advisors: Santosh Pande and Greg Eisenhauer

Hunan University, Changsha, China

Master in Computer Science

2011/06

Advisor: Cheng Xu

Hunan University, Changsha, China

BS in Computer Science

2008/06

Rank 6 / 203, postgraduate recommendation

EXPERIENCE

Software Development Engineer II

Amazon Science, Santa Clara, CA

April. 2021 – . Present

- Deep Learning Compiler

Software Development Engineer II

Amazon, Seattle, WA

Nov. 2020 – April. 2021

- Distributed Storage System

Research Intern

VMWare CTO Office, Boston, MA

May. 2016 – Aug. 2016

- **Topic:** Photon platform for container-based high performance computing. Designed and implemented a framework to spawn/create virtual clusters based on the container technology.
- **Manager:** Josh Simons, **Mentor:** Na Zhang

Research Assistant

Georgia Institute of Technology, Atlanta, GA

Sep. 2014 – May. 2017

Research Intern

New Mexico Consortium/Los Alamos National Lab, Los Alamos, NM

May. 2013 – Dec. 2013

- **Topic:** Active burst-buffer for data-intensive computing. Designed and implemented a framework to explore compute power in I/O path of Supercomputers to mitigate I/O overheads.
- **Manager/Mentor:** Michael Lang

RESEARCH PROJECTS

GPU Sharing and Scheduling

Sep. 2019 – .

GPUs are gaining increasing popularity in data-centers due to resource requirements by various emerging workloads, e.g. machine learning and graph analysis. A modern GPU consists of dozens of streaming multi-processors and suffers from under-utilization when used in data centers for emerging workloads, and yet there is no strong support for multi-tasking. In this project, I am going to explore the potential of sharing a single GPU among different workloads by designing a new scheduler and leveraging compiler techniques.

Compiler-assisted Scheduling

Oct. 2018 – .

Modern software executes on multi-core systems that share resources such as caches and main memory. In such a co-execution environment, the performance of modern software is critically affected by the resource conflicts that occur from this sharing. In this project, we are to explore a synergistic approach that combines compiler generated dynamic application attributes with runtime aggregation of such information to undertake smart scheduling decisions, thereby improving performance. Particularly, we leverage the compiler to insert “beacons” in the application at strategic program points to periodically produce and update details of anticipated resource requirements, and design a throughput-oriented scheduler to consume this information to improve system performance.

Resilience for large-scale HPC systems

May. 2015 – .

Transient faults are becoming a big threat to modern scientific applications running on large-scale supercomputers. They could lead to incorrect outputs (SDCs), or crash the execution of applications. In this project, we explore lightweight and efficient mechanisms to detect SDCs and repair the failures online based on program analysis.

Active Burst-Buffer

May. 2013 – Dec. 2013

This project investigates the computing resources of burst-buffer nodes on modern HPC systems (Supercomputers) for data analysis or visualization to mitigate the I/O overheads.

Active Storage

Sep. 2011 – May. 2014

The I/O on HPC systems is increasingly becoming the performance bottleneck. In this project, we explore the potential opportunities of decoupling and offloading data-intensive computing kernels to storage nodes. We develop advanced active storage solutions to improve I/O performance for data-intensive applications, with emphasis on data dependence and resource contention problems for active storage.

RESEARCH FUNDING

National Science Foundation (China), “Study on storage system optimization based on logical and physical I/O information”, RMB 230,000 (Co-PI, 01/14 - 12/16).

PUBLICATIONS

Preprints

- **Compiler-Guided Throughput Scheduling for Many-core Machines**
Girish Mururu, *Chao Chen*, Chris Porter and Santosh Pande
arXiv:2103.06647.

Peer-reviewed Publications

- **CASE: A Compiler-Assisted SchEduling Framework for Multi-GPU Systems**
Chao Chen, Chris Porter and Santosh Pande
accepted by 27th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP'22).
- **Near-zero Downtime Recovery from Transient-error-induced Crashes**
(*Early Access Available*)
Chao Chen, Greg Eisenhauer and Santosh Pande
In *IEEE Transactions on Parallel and Distributed Systems (TPDS)*.
Volume: 33, Issue: 4, April, 2022.
- **CARE: Compiler-assisted Recovery from Soft Failures**
Best Student Paper Finalist
Chao Chen, Greg Eisenhauer, Santosh Pande and Qiang Guan
In *International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*.
Denver, CO, Nov, 2019.
- **LADR: Low-cost Application-level Detector for Reducing Silent Output Corruptions**
Chao Chen, Greg Eisenhauer, Matthew Wolf and Santosh Pande
In *ACM International Symposium on High-Performance Parallel and Distributed Computing (HPDC)*.
Pages 156-167, Tempe, Arizona, Jun, 2018.
- **Active Burst-Buffer: In-Transit Processing Integrated into Hierarchical Storage**
Best Paper Award
Chao Chen, Michael Lang, Latchesar Ionkov and Yong Chen
In *11th IEEE International Conference on Networking, Architecture, and Storage (NAS)*.
Pages 1-10, Long Beach, CA, Aug, 2016.
- **Rethinking High Performance Computing System Architecture for Scientific Big Data Applications**
Best Paper Award
Yong Chen, *Chao Chen*, Yanlong Yin, Xianhe Sun, Rajeev Thakur and William Gropp
In *14th IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA)*.
Pages 1605-1612, Tianjin, China, Aug, 2016.
- **Multilevel Active Storage for Big Data Applications in High Performance Computing**
(*short paper*)
Chao Chen, Michael Lang and Yong Chen
In *The 2013 IEEE International Conference on Big Data (BigData)*.
Pages 169-174, Santa Clara, CA, Oct, 2013.
- **A Decoupled Execution Paradigm for Data-Intensive High-End Computing**
Yong Chen, *Chao Chen*, Xian-He Sun, William D. Gropp, and Rajeev Thakur
In *International Conference on Cluster Computing (Cluster)*.
Pages 200-208, Beijing, China, Sep, 2012.
- **DOSAS: Mitigating the Resource Contention in Active Storage Systems**
Chao Chen, Yong Chen and Philip C. Roth
In *International Conference on Cluster Computing (Cluster)*.
Page 164-172, Beijing, China, Sep, 2012.

- **Dynamic Active Storage for High Performance I/O**
Chao Chen and Yong Chen
In *41st International Conference on Parallel Processing (ICPP)*.
Pages 379-388, Pittsburgh, PA, Sep, 2012.

TEACHING

- Teaching Assistant for CS3210 Operating System Design** *Fall 2016*
Grading Projects, Holding Office Hours, and Giving Guest Lectures
- Teaching Assistant for CS6290 High Performance Computer Architecture** *Summer 2017*
Grading Homework and Projects, as well as Holding Office Hours
- Teaching Assistant for CS4240 Compilers and Interpreters** *Fall 2017*
Grading Homework and Projects, as well as Holding Office Hours
- Teaching Assistant for CS8803-O08 Compiler: Theory and Practice** *Spring 2018 ~ Now*
Grading Homework, Designing and Grading Projects, and Holding Office Hours

SERVICE

- External Reviewer of PACT'18, ASPLOS'17, IEEE BigData'13, CCGrid'13, NAS'12, ISPA'11.
Volunteer of SC'13.
Program Committee: IPDPS 2022.

AWARDS

- Best Student Paper Nominee** from **International conference on high performance computing, networking, storage and analysis (SC)**, 2019.
- Best Paper Award** from **11st International Conference on Networking, Architecture and Storage**, 2016.
- Best Paper Award** from **14th International Symposium on Parallel and Distributed Processing with Applications**, 2016.
- Student Travel Grant from **OSDI**, 2014.
- Student Travel Grant from **FAST**, 2014.
- Champion** of The Future Challenge: Intelligent Vehicles and Beyond Contest (**Team**). By **National Science Foundation (China)**, 2009.
- First Prize** of **National Undergraduate Electronic Design Contest-Embedded System Design Invitational Contest (Team)**. By **Intel, Ministry of Education (China)**, and **Ministry of Industry and Information Technology (China)**, 2008.
- First-class scholarships**. By **Hunan University**, 2005, 2007, 2008.
- Second-class scholarship**. By **Hunan University**, 2006.

REFERENCES

Santosh Pande (Advisor)
Professor
Georgia Institute of Technology
✉ gatechpande@gmail.com

Greg Eisenhauer (Advisor)
Senior Research Scientist
Georgia Institute of Technology
✉ geisenhauer@gmail.com

Richard Vuduc
Professor
Georgia Institute of Technology
✉ richie@cc.gatech.edu