Xuhao Chen *Curriculum vitae*

CONTACT INFORMATION
Massachusetts Institute of Technology (MIT), 32 Vassar St, 32-G885, Cambridge, MA, 02139
Webpage: https://chenxuhao.github.io ✉ cxh@mit.edu ☎ 512-9887388

RESEARCH INTEREST
Parallel computer system and architecture, with an emphasis on domain-specific acceleration of emerging parallel graph algorithms (pattern mining and machine learning on graphs).

EDUCATION

Ph.D. in Computer Science **Sep. 2009 - Dec. 2014**
National University of Defense Technology (NUDT), China
Advisor: Professor Zhiying Wang
Thesis: Cache Management for Manycore Accelerators

B.S. in Computer Science **Sep. 2005 - Jun. 2009**
National University of Defense Technology (NUDT), China
Rank: 1/144

ACADEMIC POSITIONS

Postdoctoral Research Associate **Sep. 2020 - Present**
Computer Science & Artificial Intelligence Laboratory (CSAIL),
Massachusetts Institute of Technology (MIT)
Supervisor: Professor Arvind
Research Area: Parallel Computer Architecture

Research Fellow **Jan. 2019 - Aug. 2020**
Institute for Computational Engineering and Sciences,
University of Texas at Austin
Supervisor: Professor Keshav Pingali
Research Area: Parallel Computing

Assistant Research Scientist **Jan. 2015 - May. 2018**
Department of Computer Science,
National University of Defense Technology (NUDT), China
Research Area: Computer Architecture

Visiting Student **Oct. 2012 - Oct. 2014**
Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign (UIUC)
Advisor: Professor Wen-Mei Hwu
Research Area: Computer Architecture

PEER-REVIEWED PUBLICATIONS

[1] Xuhao Chen, Roshan Dathathri, Gurbinder Gill, Keshav Pingali, "Pangolin: An Efficient and Flexible Graph Pattern Mining System on CPU and GPU", International Conference on Very Large Databases (**VLDB**), 13(8): 1190-1205, 2020

[2] Xuhao Chen, Cheng Chen, Jie Shen, Jianbin Fang, Tao Tang, Canqun Yang, Zhiying Wang, "Orchestrating Parallel Detection of Strongly Connected Components on GPUs", Parallel Computing (**ParCo**), Volume 78, Pages 101-114, 2018

[3] Xuhao Chen, Pingfan Li, Jianbin Fang, Tao Tang, Zhiying Wang, Canqun Yang, "Efficient and High-quality Sparse Graph Coloring on the GPU", Concurrency and Computation: Practice and Experience (**CPE**), Volume 29, Issue 10, 2017

[4] Xuhao Chen, Li-Wen Chang, Christopher I. Rodrigues, Jie Lv, Zhiying Wang, Wen-Mei W. Hwu. "Adaptive Cache Management for Energy-efficient GPU Computing", In the 47th International Symposium on Microarchitecture (**MICRO**), 2014

[5] Xuhao Chen, Shengzhao Wu, Li-Wen Chang, Wei-Sheng Huang, Carl Pearson, Zhiying Wang, Wen-Mei W. Hwu. "Adaptive Cache Bypass and Insertion for Many-core Accelerators", In Proceeding of the MES Workshop in conjunction with ISCA-41, 2014

[6] Loc Hoang*, Vishwesh Jatala*, Xuhao Chen, Udit Agarwal, Roshan Dathathri, Grubinder Gill, Keshav Pingali, "DistTC: High Performance Distributed Triangle Counting", IEEE High Performance Extreme Computing Conference (**HPEC**), 2019

[7] Hang Zhang, Xuhao Chen, Nong Xiao, Fang Liu, "Optimizing STT-RAM Based Register File Energy Consumption on GPGPU with Delta Compression", In Proceeding of the 53rd Design Automation Conference (**DAC**), 2016

[8] Zhen Xu, Xuhao Chen, Jie Shen, Yang Zhang, Cheng Chen, Canqun Yang, "GARDENIA: A Domain-specific Benchmark Suite for Next-generation Accelerators", ACM Journal on Emerging Technologies in Computing Systems (**JETC**), 15(1): 9, 2019

[9] Hang Zhang, Xuhao Chen, Nong Xiao, Fang Liu, "Red-Shield: Shielding Read Disturbance for STT-RAM Based Register files on GPUs", In Proceeding of the 26th Great Lakes Symposium on VLSI (**GLSVLSI**), 2016

[10] Pingfan Li, Xuhao Chen, Jie Shen, Jianbin Fang, Tao Tang, Canqun Yang, "High Performance Detection of Strongly Connected Components in Sparse Graphs on GPUs", In the Proceedings of the PMAM Workshop in conjunction with PPoPP-22, 2017

[11] Hang Zhang, Xuhao Chen, Nong Xiao, Lei Wang, Fang Liu, Wei Chen, Zhiguang Chen, "Shielding STT-RAM Based Register files on GPUs Against Read Disturbance", ACM Journal on Emerging Technologies in Computing Systems (**JETC**), 13(2): 27, 2016

[12] Pingfan Li, Xuhao Chen, Zhe Quan, Jianbin Fang, Huayou Su, Tao Tang, Canqun Yang, "High Performance Parallel Graph Coloring on GPGPUs", IPDPS Workshop, 2016

[13] Jing Chen, Jianbin Fang, Weifeng Liu, Tao Tang, Xuhao Chen, Canqun Yang, "Efficient and Portable ALS Matrix Factorization for Recommender Systems", IPDPS Workshop, 2017

[14] Jianbin Fang, Peng Zhang, Zhaokui Li, Tao Tang, Xuhao Chen, Cheng Chen, Canqun Yang, "Evaluating Multiple Streams on Heterogeneous Platforms", Parallel Processing Letters, Volume 26, Issue 4, 2016

[15] Canqun Yang, Cheng Chen, Tao Tang, Xuhao Chen, Jianbin Fang, Jingling Xue, "An Energy-Efficient Implementation of LU Factorization on Heterogeneous Systems". ICPADS, 2016

[16] Zhaokui Li, Jianbin Fang, Tao Tang, Xuhao Chen, Cheng Chen, Canqun Yang, "Evaluating the Performance Impact of Multiple Streams on the MIC-Based Heterogeneous Platform". IPDPS Workshop, 2016

Under Submission

[17] Xuhao Chen*, Tianhao Huang*, Shuotao Xu, Thomas Bourgeat, Arvind, "Software/Hardware Co-designed System for Efficient Graph Pattern Mining", Tianhao and I co-lead

[18] Xuhao Chen*, Loc Hoang*, Roshan Dathathri, Gurbinder Gill, Keshav Pingali, "DeepGalois: An Efficient Framework for Deep Learning on Large Graphs", Loc and I co-lead

[19] Xuhao Chen, Roshan Dathathri, Gurbinder Gill, Loc, Hoang, Keshav Pingali, "Sandslash: A Two-Level Framework for Efficient Graph Pattern Mining"

[20] Xuhao Chen, Zhen Xu, Jie Shen, Hao Zhu, "GraphCage: Cache Aware Graph Processing on GPUs", CoRR, https://arxiv.org/abs/1904.02241

[21] Xuhao Chen, Zhen Xu, Jie Shen, Hao Zhu, "Escort: Efficient Sparse Convolutional Neural Networks Inference on GPUs", CoRR, https://arxiv.org/abs/1802.10280

Research Experience

Programming Framework for Graph Pattern Mining **2019-2020**

With Keshav Pingali

Design and implemented Pangolin [1], an efficient and flexible graph pattern mining (GPM) framework targeting shared-memory CPUs and GPUs. It is the first GPM system that supports GPU mining. Pangolin provides a novel programming interface which allows users to express application-specific optimizations. It also employs novel architecture oriented optimizations, particularly for the GPU architecture. These innovations makes Pangolin orders-of-magnitude faster than previous GPM systems.

Parallel Graph Algorithms on GPU **2015-2018**

With Zhiying Wang

Design and implemented various parallel graph algorithms [2,3,6], frameworks [20] and benchmarks [8] on the GPU, for diverse graph problems ranging from graph coloring and and strongly connected components to sparse neural networks. I proposed techniques to overcome the challenges of insufficient parallelism, indirect memory access pattern and load imbalance, which leads to substantial speedups over previous parallel CPU and GPU solutions.

Cache Architecture for Irregular Algorithms on GPU **2011-2014**

With Wen-Mei Hwu and *Zhiying Wang*

Designed and implemented efficient cache architectures [4,5,7,9] for irregular applications on GPU. Irregular algorithms, e.g., graph algorithms, have indirect memory accesses that cause memory divergence on GPU. The massive amount of diverged memory requests causes severe cache contention and resource congestion. I proposed an adaptive cache management scheme [4] specifically for the GPU architecture, which effectively combines the techniques of warp throttling and cache bypassing, and achieves significant performance improvement.

Selected Honors and Awards

- Graph Challenge 2019 Student Innovation Awards 2019
- China Computer Federation (CCF) Distinguished PhD Dissertation Nominee 2015
- Ci Yun-Gui Scholarship for Graduate, NUDT (top 1%) 2010
- **Meritorious Winner**, Mathematical Contest In Modeling (MCM), COMAP 2009
- Distinguished Graduate, NUDT (top 1%), 2009
- First rank, Scholarship of Excellent Achievements, NUDT (top 3%) 2009
- Ci Yun-Gui Scholarship for Undergraduate, NUDT (top 1%) 2008
- **First-rank Prize**, China Undergraduate Mathematical Contest in Modeling 2007
- First rank, Scholarship of Excellent Achievements, NUDT (top 3%) 2007

Teaching Experience

Computer Architecture (undergraduate course) **Fall 2008 NUDT**

- With Professor Zhiying Wang
- Teaching Assistant to redesign the labs and the final project for a 5-stage pipelined in-order processor, and mentor students on course projects

Design and Analysis of Algorithms (undergraduate course) **Fall 2010 NUDT**

- With Professor Jianping Yin
- Teaching Assistant to mentor students on labs and final projects and help with scoring

CS 380C: Advanced Topics in Compilers (graduate course) **Fall 2019, UT Austin**

- With Professor Keshav Pingali
- Teaching Assistant to setup a course project on distributed graph pattern mining algorithms and mentor students on the project

Mentoring Experience

1. Tianhao Huang (second year PhD with Prof. Arvind).
   Project: Algorithm-aware Hardware Accelerator for Graph Pattern Mining

2. Loc Hoang (second year PhD with Prof. Keshav Pingali).
   Project: Programming Framework for Graph Neural Networks on Distributed System
3. Siyu Zhang (first year PhD with Prof. Keshav Pingali).
   Project: Parallel and Distributed k-clique Listing on Large Graphs
4. Pingfan Li (master student with Prof. Zhiying Wang).
   Project: Parallel Graph Coloring and SCC Detection on GPU
5. Ping Li (undergraduate student with Prof. Zhiying Wang).
   Project: Benchmarking Multicore CPU and GPU using PARSEC, Parboil and Rodinia

Grants

- NSF of China Grant No.61502514 (2015), PI: **Xuhao Chen**
  "Memory Hierarchy for Energy-efficient Heterogeneous Processors"
  Wrote successful grant proposal and led the project
- NSF of China Grant No.61272144 (2012), PI: Zhiying Wang
  "Energy-efficient Asynchronous Manycore Processor"
  Co-wrote successful grant proposal and presented proposed research at kickoff meeting

Talks

1. Pangolin: An Efficient and Flexible Graph Mining System on CPU and GPU
   International Conference on Very Large Databases (VLDB), Tokyo, Japan, Sep. 2020
2. High Performance Detection of SCCs in Sparse Graphs on GPUs
   PMAM Workshop in conjunction with PPoPP-22, Austin, TX, Feb. 2017
3. Adaptive Cache Management for Energy-efficient GPU Computing
   International Symposium on Microarchitecture (MICRO), Cambridge, UK, Dec. 2014
4. Adaptive Cache Bypass and Insertion for Many-core Accelerators
   MES Workshop in conjunction with ISCA-41, Minneapolis, MN, June 2014

Academic Service

- Invited reviewer for ACM Transactions on Modeling and Performance Evaluation of Computing Systems
- Invited reviewer for Microprocessors and Microsystems: Embedded Hardware Design
- Invited reviewer for Journal of Supercomputing

References (alphabetical)

Arvind
Johnson Professor
Computer Science and Artificial Intelligence Laboratory at MIT
arvind@csail.mit.edu

Keshav Pingali
Professor,
Department of Computer Science at UT Austin
pingali@cs.utexas.edu

Wen-Mei Hwu
Sanders III AMD Endowed Chair, Professor
Department of Electrical and Computer Engineering at UIUC
w-hwu@illinois.edu

Zhiying Wang
Professor
Department of Computer Science at NUDT
zywang@nudt.edu.cn