Xuhao Chen — *Curriculum vitae*

Contact Information

Mailing Address: 32 Vassar Street, 32-G836, Cambridge, MA, 02139
Webpage: http://people.csail.mit.edu/xchen ✉ cxh@mit.edu 512-9887388

Research Interest

Specialized AI & Big-data systems, for data mining and machine learning on graphs, spanning parallel algorithms, runtime system, domain-specific compiler and hardware architecture.

Academic Positions

Research Scientist — **Sep. 2020 - Present**
Computer Science & Artificial Intelligence Laboratory (CSAIL),
Massachusetts Institute of Technology (MIT)
Supervisor: Professor Arvind
Research Area: Efficient AI Systems and 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 & Graph Algorithms

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

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: Massively Parallel Computer Architecture

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

Peer-Reviewed Publications

[1] Xuhao Chen, Tianhao Huang, Shuotao Xu, Thomas Bourgeat, Chanwoo Chung, Arvind, "FlexMiner: A Pattern-Aware Accelerator for Graph Pattern Mining", International Symposium on Computer Architecture (**ISCA**), 2021 [pdf]

[2] Xuhao Chen, Roshan Dathathri, Gurbinder Gill, Loc Hoang, Keshav Pingali, "Sandslash: A Two-Level Framework for Efficient Graph Pattern Mining", International Conference on Supercomputing (**ICS**), 2021 [pdf]

[3] Loc Hoang, Xuhao Chen, Hochan Lee, Roshan Dathathri, Gurbinder Gill, Keshav Pingali, "Efficient Distribution for Deep Learning on Large Graphs", Workshop on Graph Neural Networks and Systems (**GNNSys**), 2021 [pdf]

[4] 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 [pdf]

[5] 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 [pdf]

[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 [pdf]

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

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

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

[17] 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

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

Research Experience

Hardware Accelerator for Graph Pattern Mining **Sep. 2020 - Present**

With Arvind

Design and implemented a pattern-aware hardware accelerator, FlexMiner [1], for graph pattern mining (GPM). FlexMiner offers 11× speedup over the state-of-the-art CPU baseline

with 10× less logic area than CPU, while retaining high programming productivity.

Programming Framework for Graph Pattern Mining **Jan. 2019 - Aug. 2020**
With Keshav Pingali
Design and implemented graph pattern mining (GPM) frameworks, Pangolin [4] and Sandslash [2]. Pangolin hides away parallel programming complexity, and is 100× ∼ 1000× faster than prior systems. Sandslash provides a novel two-level API that further simplifies programming, and enables optimizations that deliver 8× speedup over Pangolin on CPU.

Parallel Graph Algorithms on GPU **Jan. 2015 - Dec. 2018**
With Zhiying Wang
Design and implemented graph algorithms [6,7,8] on GPU, e.g., vertex coloring, strongly connected components, sparse neural networks. It pushes the frontier of GPU graph processing.

Cache Architecture for Irregular Algorithms on GPU **Sep. 2009 - Dec. 2014**
With Wen-Mei Hwu and *Zhiying Wang*
Designed and implemented efficient cache architectures [5,9,10,11,12] for GPU. An *adaptive* cache management policy [5] is proposed for GPU, which effectively combines warp throttling and cache bypassing, to mitigate the memory divergence and cache contention on GPU.

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

6.886: Algorithm Engineering (graduate course) **Spring 2021, MIT**
- With Professor Julian Shun
- Guest lecture on Pangolin [Slides]

CS 380C: Advanced Topics in Compilers (graduate course) **Fall 2019, UT Austin**
- With Professor Keshav Pingali
- Teaching Assistant to setup a course project and mentor students on the project

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

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

Mentoring Experience

1. Tianhao Huang (third year PhD with Prof. Arvind).
   Closely collaborate on project: Hardware Accelerator for Graph Pattern Mining
2. Loc Hoang (second year PhD with Prof. Keshav Pingali).
   Closely collaborate on project: Distributed Framework for Graph Neural Networks
3. Siyu Zhang & Bozhi You (first year PhD with Prof. Keshav Pingali).
   Mentoring course project: Parallel and Distributed k-clique Listing on Large Graphs

4. Jie Lv (master student with Prof. Wen-Mei Hwu).
   Closely collaborate on project: Adaptive Cache Management for GPU
5. Zhihao Xie (undergraduate student with Prof. Wen-Mei Hwu).
   Mentoring course project: Characterization of Irregular Algorithms on GPU

Grants

- Samsung Memory Solutions Lab. (MSL) UR Collaboration Program (2022), PI: Arvind
  "In-storage Acceleration of Massive Graph Neural Networks"
  Co-led the project and wrote successful grant proposal
- MIT-IBM Watson AI Lab Research Funding (2021), PI: Arvind
  "Storage-Centric Infrastructure for Graph-Based Deep Learning on Massive Datasets"
  Co-led the project and wrote successful grant proposal
- NSF of China Grant No.61502514 (2015), PI: Xuhao Chen
  "Memory Hierarchy for Energy-efficient Heterogeneous Processors"
  Wrote successful grant proposal, led and completed the project

Talks

1. Invited Talk at the CS Seminar in UMass Amherst, November 5, 2021
2. Invited Talk at the MiDAS Seminar in Boston University, October 29, 2021
3. Invited Talk at the COSI Seminar in Brandeis University, October 22, 2021
4. Software and Hardware Systems for Emerging Graph Algorithms [Video]
   MIT Fast Code Seminar, April 28, 2021
5. Pangolin: An Efficient and Flexible Graph Mining System on CPU and GPU [Video]
   International Conference on Very Large Databases (VLDB), Tokyo, Japan, Sep. 2020
6. High Performance Detection of SCCs in Sparse Graphs on GPUs
   PMAM Workshop in conjunction with PPoPP-22, Austin, TX, Feb. 2017
7. Adaptive Cache Management for Energy-efficient GPU Computing
   International Symposium on Microarchitecture (MICRO), Cambridge, UK, Dec. 2014
8. 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 IEEE Transactions on Knowledge and Data Engineering
- Invited reviewer for ACM Transactions on Architecture and Code Optimization
- 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

Arvind
Johnson Professor
CSAIL at MIT
Postdoc Supervisor
arvind@csail.mit.edu

Keshav Pingali
Professor
CS at UT Austin
Postdoc Supervisor
pingali@cs.utexas.edu

Wen-Mei Hwu
Professor and Sanders-AMD Chair
ECE at UIUC
Supervisor
w-hwu@illinois.edu

Zhiying Wang
Professor
CS at NUDT
PhD Thesis Advisor
zywang@nudt.edu.cn