

ZHENG (EDDY) ZHANG

Department of Computer Science
Rutgers, The State University of New Jersey
110 Frelinghuysen Road
Piscataway, NJ 08854-8019

Voice: (848) 445-8310
Fax: (732) 445-0537
Email: eddy.zhengzhang@cs.rutgers.edu
www.cs.rutgers.edu/~zz124

RESEARCH INTERESTS

Many-Core Programming System, GPU Computing, Shared Cache Locality Enhancement, Information Flow and Security, Graph Processing, Energy Efficiency, Performance Evaluation, Markov Chains and Stochastic Simulation

Employment

Assistant Professor - Computer Science Department Rutgers, The State University of New Jersey	9/2012 - Now
Research Assistant - Computer Science Department	8/2009 - 7/2012
Teaching Assistant - Math Department and Computer Science Department The College of William and Mary	8/2005 - 7/2009
Research Intern - eXtreme Computing Group (XCG) Microsoft Research (Redmond)	4/2011 - 8/2011
Web and Database Applications Programmer Intern Virginia Institute of Marine Science	5/2006 - 8/2006

AWARDS

Google Faculty Research Award	2014
Google Anita Borg Scholarship	2011-2012
Stephen K. Park Graduate Research Award at the College of William and Mary	2011
NVIDIA Graduate Fellowship Finalist	2011
PPoPP'10 Best Paper Award - <i>The 15th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)</i>	2010
QEST'08 Best Student Paper Award - <i>The 5th International Conference on the Quantitative Evaluation of SysTems (QEST)</i>	2008

EDUCATION

The College of William and Mary, Williamsburg, VA Ph.D. in Computer Science M.S. in Computer Science Specialized in <i>Computational Operations Research</i>	Aug. 2012
Shanghai Jiaotong University, Shanghai, China B.S. in Electronic Engineering	Jul. 2004

TEACHING

CS 515 Programming Languages and Compilers I (Graduate Course)	Fall 2012
CS 516 Programming Languages and Compilers II (Graduate Course)	Spring 2015/2017
CS 671 Program Compilation and Optimization in Exa-scale Computing Era (Graduate Course)	Fall 2013
CS 415 Compilers (Undergraduate Course)	Spring 2013/2014/2016
CS 314 Principles of Programming Languages (Undergraduate Course)	Fall 2014/2016

PUBLICATIONS

Refereed Conference Publications

- A.B. Hayes, L. Li, M. Hedayati, J. He, E.Z. Zhang, K. Shen, “GPU Taint Tracking”, The 2016 USENIX Annual Technical Conference (USENIX ATC 2017), Santa Clara, CA, July 2017. Acceptance Rate: 21.2% (60/283).
- L. Li, R. Geda, A. B. Hayes, P. Chaudhari, E.Z. Zhang, M. Szegedy, “A Simple Yet Effective Balanced Graph Edge Partition Model for Parallel Computing”, The 2017 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 2017), Urbana-Champaign, Illinois, June 2017. Acceptance Rate: 25.9% (56/216).
- P. Li, X. Hu, D. Chen, J. Brock, H. Luo, E.Z. Zhang, C. Ding, “LD: Low-Overhead GPU Race Detection without Access Monitoring”. ACM Transaction on Architecture and Code Optimization (TACO), 2017, Invited for Presentation at the 12th HiPEAC Conference.
- A.B. Hayes, L. Li, D. Chavarria, S.L. Song, E.Z. Zhang, “ORION: A Framework for GPU Occupancy Tuning”, The 17th ACM/IFIP/USENIX Middleware Conference (Middleware 2016) , Trento, Italy, December, 2016. Acceptance Rate: 19.6% (21/107).
- L. Li, A.B. Hayes, S. Song, E.Z. Zhang, “Tag-Split Cache for Efficient GPGPU Cache Utilization”, The 30th ACM International Conference on Supercomputing (ICS 2016), Istanbul, Turkey, June, 2016. Acceptance Rate: 24.1% (43/178).
- D. Tao, S. Leon Song, S. Krishnamoorthy, P. Wu, E. Z. Zhang, Z. Chen, D. Kerbyson, “New-Sum: A Novel Online ABFT Scheme For General Iterative Methods ”, The 25th ACM Symposium on High-Performance Parallel and Distributed Computing (HPDC 2016), Kyoto, Japan, June, 2016. Acceptance Rate: 15.5% (20/129).
- A. Li, S. Song, A. Kumar, E.Z. Zhang, D. Chavarria, H. corporaal, “Critical Point Based Register-Concurrency Autotuning For GPUs”. The 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE 2016), Dresden, Germany, March, 2016. Acceptance Rate: 24%.
- G. Haldeman, I. Rodero, M. Parashar, S. Ramos, E. Z. Zhang, U. Kremer, “Exploring Energy-Performance-Quality Tradeoffs for Scientific Workflows With In-situ Data Analyses”, International Conference on Energy-Aware High Performance Computing (EnA-HPC 2014), Dresden, Germany, September, 2014.
- I. Egielski, J. Huang, E. Z. Zhang, “Massive Atomics for Massive Parallelism on GPUs”, ACM SIGPLAN International Symposium on Memory Management (ISMM 2014), Edinburgh, Scotland, UK, June, 2014. Acceptance Rate: 50% (11/22).
- A. Hayes, E. Z. Zhang, “Unified On-chip Memory Allocation for SIMT Architecture”, ACM International Conference on Supercomputing (ICS 2014), Munich, Germany, June, 2014. Acceptance Rate: 20.9% (34/162).
- B. Wu, Z. Zhao, E. Z. Zhang, Y. Jiang, X. Shen, “Complexity Analysis and Algorithm Design for Reorganizing Data to Minimize Non-Coalesced GPU Memory Accesses”, 18th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP’13), Shenzhen, China, February 2013. Acceptance Rate: 17.8% (26/146).
- B. Wu, E. Z. Zhang, X. Shen, “Enhancing Data Locality for Dynamic Simulations through Asynchronous Data Transformations and Adaptive Control”, The 20th International Conference on Parallel Architectures and Compilation Techniques (PACT’11), Galveston Island, Texas, USA, October 2011. Acceptance Rate: 16% (36/121).
- Z. Guo, E. Z. Zhang, X. Shen, “Correctly Treating Synchronizations in Compiling Fine-Grained SPMD-Threaded Programs for CPU”, The 20th International Conference on Parallel Architectures and Compilation Techniques (PACT’11), Galveston Island, Texas, USA, October 2011. Acceptance Rate: 16% (36/121).
- K. Tian, E. Z. Zhang, X. Shen, “A Step Towards Transparent Integration of Input-Consciousness into Dynamic Program Optimizations”, ACM International Conference on Systems, Programming, Languages and Applications (SPLASH 2011), Portland, Oregon, USA, October 2011.

- E. Z. Zhang, Y. Jiang, Z. Guo, K. Tian, X. Shen, “On-the-Fly Elimination of Dynamic Irregularities for GPU Computing”, International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2011), Newport Beach, California, March 2011. Acceptance Rate: 21% (32/152).
- K. Tian, Y. Jiang, E. Z. Zhang, X. Shen, “An Input-Centric Paradigm for Program Dynamic Optimizations”, ACM International Conference on Systems, Programming, Languages and Applications (OOPSLA 2010), Reno, Nevada, October 2010. Acceptance Rate: 27% (45/164).
- E. Z. Zhang, Y. Jiang, Z. Guo, X. Shen, “Streamlining GPU Applications On the Fly”, ACM International Conference on Supercomputing (ICS 2010), Tsukuba, Japan, June 2010. Acceptance Rate: 17.8% (32/180).
- Y. Jiang, E. Z. Zhang, K. Tian, X. Shen, “Is Reuse Distance Applicable to Data Locality Analysis on Chip Multiprocessors?”, International Conference on Compiler Construction (CC 2010), Paphos, Cyprus, March 2010. Acceptance Rate: 28% (15/56).
- Y. Jiang, E.Z. Zhang, K. Tian, F. Mao, M. Gethers, X. Shen, Y. Gao, “Exploiting Statistical Correlations for Proactive Prediction of Program Behaviors”, ACM/IEEE International Symposium on Code Generation and Optimization (CGO 2010), Toronto April, 2010.
- E.Z. Zhang, Y. Jiang, X. Shen, “Does Cache Sharing on Modern CMP Matter to the Performance of Contemporary Multithreaded Programs?”, The 15th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP 2010), Bangalore, India, January 2010. Acceptance Rate: 16.7%(29/173). (Best Paper Award)
- F. Mao, E.Z. Zhang, X. Shen. “Influence of Program Inputs on the Selection of Garbage Collectors”, The 2009 ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE 2009), Washington, DC, March 2009
- Y. Liu, E.Z. Zhang, X. Shen. “A Cross-Input Adaptive Framework for GPU Programs Optimization”. The 2009 IEEE International Parallel & Distributed Processing Symposium (IPDPS 2009), Rome, Italy, May 2009. Acceptance Rate: 23% (100/440).
- G. Casale, E.Z. Zhang, E. Smirni, “KPC-Toolbox: Simple Yet Effective Trace Fitting Using Markovian Arrival Processes”, Proceedings of the 5th International Conference on the Quantitative Evaluation of Systems (QEST 2008), pp. 83-92, St. Malo, France, IEEE Press, September 2008. (Best Student Paper Award)
- E.Z. Zhang, G. Casale, E. Smirni, “KPC-Toolbox: Best Recipes Toward Automation of Workload Fitting”, Special Issue on the SIGMETRICS Demo Competition 2008. ACM SIGMETRICS Performance Evaluation Review 36(2):134-136, ACM Press, September 2008.

Journal

- Lingda Li, Robel Geda, Ari B. Hayes, Yanhao Chen, Pranav Chaudhari, Eddy Z. Zhang, and Mario Szegedy. “A Simple Yet Effective Balanced Edge Partition Model for Parallel Computing”. Proceedings of the ACM Measurement Analysis Computing Systems, Vol. 1, Issue. 1, Article 14 (June 2017), 21 pages. DOI: <https://doi.org/10.1145/3084451>
- Pengcheng Li, Xiaoyu Hu, Dong Chen, Jacob Brock, Hao Luo, Eddy Z. Zhang, and Chen Ding. 2017. “LD: Low-Overhead GPU Race Detection Without Access Monitoring”. ACM Transactions on Architecture and Code Optimization, Volume 14, Issue 1, Article 9 (March 2017), 25 pages. DOI: <https://doi.org/10.1145/3046678>
- X. Shen, Y. Liu, E. Z. Zhang, P. Bhamidipati, “An Infrastructure for Tackling Input-Sensitivity of GPU Program Optimizations”, International Journal of Parallel Programming 41(6): 855-869, 2013.
- E. Z. Zhang, Y. Jiang, X. Shen, “The Significance of the Influence of Cache Sharing in Modern CMPs on the Performance of Contemporary Multithreaded Programs”, IEEE Transaction on Parallel and Distributed Systems, vol. 23, no. 2, pp. 367-374, DOI: 10.1109/TPDS.2011.130, February 2012.
- G. Casale, E. Z. Zhang, E. Smirni, “KPC-Toolbox: Best recipes for automatic trace fitting using Markovian Arrival Processes”, Elsevier Performance Evaluation, vol. 67, Issue 9, 873-896, September 2010.
- G. Casale, E. Z. Zhang, E. Smirni, “Trace data characterization and fitting for Markov modeling”, Elsevier Performance Evaluation, vol. 67, pp. 61-79, February 2010.

- X. Shen, F. Mao, K. Tian, E. Z. Zhang, “The Study and Handling of Program Inputs in the Selection of Garbage Collectors”, ACM SIGOPS Operating Systems Review, July 2009.

Refereed Workshop Publications

- J.D. Catarata, S. Corbett, H. Stern, M. Szegedy, T. Vyskocil, Z. Zhang (Eddy Z. Zhang), “The Moser-Tardos Resample algorithm: Where is the limit? (an experimental inquiry)”, Proceedings of the Nineteenth Workshop on Algorithm Engineering and Experiments (ALENEX 2017), Barcelona, Spain, 2017.
- Dmitry Mikushin, Nikolay Likhogrud, Eddy Z. Zhang, Christopher Bergströ, “KernelGen - the design and implementation of a next generation compiler platform for accelerating numerical models on GPUs”, The 28th IEEE International Parallel & Distributed Processing Symposium: Programming Models, Languages and Compilers Workshop for Manycore and Heterogeneous Architectures (PLC2014).
- E.Z. Zhang, H. Li, and X. Shen, “A Study Towards Optimal Data Layout for GPU Computing”, Proceedings of the ACM SIGPLAN Workshop on Memory Systems Performance and Correctness (MSPC 2012), June, 2012, in conjunction with PLDI’12.
- G. Casale, E.Z. Zhang, E. Smirni, “Characterization and Synthesis of Markovian Workload Models”, Proceedings of the GlobeCom 2007 Workshop on Future Service-Oriented Internet, 1-5, Washington DC, IEEE Press, November 2007.
- G. Casale, E.Z. Zhang, E. Smirni, “Characterization of Moments and Autocorrelation in MAPs”, Special issue of the MAMA 2007 workshop. ACM SIGMETRICS Performance Evaluation Review 35(2):27-29, ACM Press, September 2007.

INVITED TALKS

- “*GPU Taint Tracking*”
The USENIX Annual Technical Conference (USENIX ATC’17), Santa Clara, CA, July 2017.
- “*A Simple Yet Effective Balanced Edge Partition Model for Parallel Computing*”
University of Rochester, Rochester, NY, May 2017.
- “*ORION: A Framework for GPU Occupancy Autotuning and Resource Allocation*”
Rice University, Houston, TX, May 2017.
- “*ORION: A Framework for GPU Occupancy Autotuning and Resource Allocation*”
Huawei Research, Santa Clara, CA, Aug 2016.
- “*ORION: A Framework for GPU Occupancy Autotuning and Resource Allocation*”
University of California, Irvine. Irvine, CA, Jan 2016.
- “*Unified On-chip Memory Allocation for SIMT Architecture*”
The 28th ACM International Conference on Supercomputing (ICS’14), Munich, Germany, June 2014.
- “*Orchestrating On-chip Memory Allocation for SIMT Architecture*”
South China University of Technology, Guangzhou, China, Jan 2014.

SERVICES

Technical Conference Committee Memberships

- *ACM SIGPLAN International Symposium on Memory Management (ISMM)*
ISMM Steering Committee Member 2016 - Present
- *ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*
Program Committee Member 2014, 2016, 2017
- *ACM International Conference on Supercomputing (ICS)*
Program Committee Member 2016, 2018

- *USENIX Annual Technical Conference (ATC)*
Program Committee Member 2018
- *ACM SIGPLAN conference on Systems, Programming, Languages and Applications: Software for Humanity (SPLASH/OOPSLA)*
External Review Committee Member 2017
- *ACM SIGPLAN conference on Programming Language Design and Implementation (PLDI)*
Program Committee Member 2015
- *IEEE/ACM International Symposium on Code Generation and Optimization (CGO)*
Program Committee Member 2015
- *ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP)*
Program Committee Member 2014, 2018
- *ACM SIGPLAN International Symposium on Memory Management (ISMM)*
Program Committee Member 2015
- *IEEE International Conference on High Performance Computing and Communications (HPCC)*
Program Committee Member 2013
- *IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*
Program Committee Member 2015

Technical Conference Committee Chairmanships

- *ACM International Symposium on Memory Management (ISMM)*
Program Committee Chair 2016
- *ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis*
Program Committee Co-Chair for the Area of Visualization, Storage and Analytics 2015

Technical Reviewer for Journals

ACM Transaction on Computers (TC), ACM Transactions on Architecture and Code Optimization (TACO), ACM Transactions on Programming Languages and Systems (TOPLAS), ACM Transactions on Parallel Computing (TOPC), Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), ACM Transactions on Parallel Computing (TOPC), IEEE Computer Architecture Letters (CAL)

GRANTS

- *NSF-CCF-1628401 (Sole-PI)*
Title: Cache Management for Data Parallel Architectures”
2016 - 2019 Amount: \$300,000.00
- *NSF-CCF-1421505 (Sole-PI)*
Title: Optimizing Compiler and Runtime for Concurrency-Oriented Execution Model
2014 - 2017 Amount: \$378,141.00
- *Google Faculty Research Award*
Title: Programming GPUs with Maximal Concurrent Efficiency
2014 - Now Amount: \$45,053.00
- *Rutgers Research Council Grant (Sole-PI)*
Title: Concurrency Oriented Compiler and Runtime Optimization
2014 - 2015 Amount: \$2,000.00
- *Rutgers Aresty Undergraduate Research Grant (Sole-PI)*
Title: A Compiler and Runtime Framework for Exascale Computing
2014 - 2015 Amount: \$2,000.00
- *Rutgers Student Computation Fee (SCF) Grant (Co-PI)*
2014 - 2015 Amount: \$30,000.00