Roger Allan Pearce

Research Interests

Graph Algorithms, Parallel Algorithms, External Memory Algorithms, Data-Intensive Computing, Machine Learning, Optimization, Computational Science, High-Performance Computing (HPC)

Education

2013   **Ph.D., Computer Science**, *Texas A&M University*, College Station, Texas, USA.
Research Focus: Parallel & External Memory Graph Algorithms
Adviser: Dr. Nancy M. Amato
Dissertation: "Scalable Parallel Algorithms for Massive Scale-free Graphs"

2004   **B.S., Computer Engineering**, *Texas A&M University*, College Station, Texas, USA.
Senior Honors Thesis: "Extracting Optimal Paths from Roadmaps for Motion Planning"

Professional Experience

2013 - present   **Center for Applied Scientific Computing**, *Lawrence Livermore National Laboratory*.

new! ○ PI of internally funded R&D project investigating persistent memory technologies for exploratory data analytics (EDA) on HPC (FY21-FY23 $1.5M):
- Investigating interactive exploratory labeled graph pattern matching [5]

new! ○ LLNL-PI of multi-institution NSF funded (via UCSF) project investigating graph analytics for system biology knowledge graph SPOKE (FY21-FY22 $290K).

○ PI of internally funded R&D project investigating graph algorithms on CORAL2 arch (FY19 $150K).
- Demonstrated largest Graph500 to date; 70-trillion edges with distributed out-of-core alg.
- Demonstrated largest GraphChallenge to date; awarded *champion* [8].

○ Co-PI of internally funded R&D project investigating graph algorithms on HPC (FY17-FY19 $2M); algorithms investigated include:
- Labeled Graph Pattern Matching [1, 18, 23, 20]
- GraphChallenge Triangle Counting [22] & K-Truss [16]
- Exact vertex Eccentricity [13]
- Distributed-GPU Breadth-First Search (BFS) [15]
- Streaming on-line graph algorithms [11, 25]

○ Selected systems & architecture research contributions include:
- Distributed communication optimizations for irregular algorithms on HPC [10]
- System software for Non-Volatile Memory (NVM & NVRAM) [24, 27].
- Disaggregated memory for HPC [3]
- Persistent Memory (PM) C++ allocator: Metall [7]

○ Supervised 3 post-docs, 15 Ph.D., and 5 undergraduate student interns.
○ Served on 2 Ph.D. Committees as an external member.
○ Served on the Lawrence Fellowship review committee as computation representative (2016-Present)

2009 - 2013   **Graduate Fellowship**, *Lawrence Livermore National Lab*,   Adviser: Dr. Maya Gokhale.

○ Awarded Lawrence Scholar Program (LSP) graduate fellowship.
○ Developed asynchronous graph traversal algorithms for massive real-world graphs: HavoqGT.
○ Applied asynchronous techniques to External Memory using NAND Flash devices [35].
○ Developed scalable techniques for processing massive scale-free graphs in distributed memory [29, 30].

2004 - 2009	**Graduate Research Assistant**, *Parasol Lab*, Texas A&M, Adviser: Dr. Nancy Amato.

2004 - 2009   **Graduate Research Assistant**, *Parasol Lab*, Texas A&M,   Adviser: Dr. Nancy Amato.
- Applied machine learning techniques to robotic motion planning problems [39, 42, 43].
- Developed techniques to analyze the evolution of sampling based motion planning algs. [38, 40, 41].
- Improved motion planning sampling techniques using graph structural information [36].

2007, 2008   **Summer Intern**, *Lawrence Livermore National Laboratory*,   Adviser: Dr. Maya Gokhale.
- Member of *Storage Intensive Super Computing (SISC)* group [37]
  - Developed an I/O profiling tool using dynamic interposition libraries.
  - Researched techniques for implementing external memory graph algorithms.
  - Experimented with distributed file systems and *Hadoop MapReduce*.

2001 – 2004   **Undergrad Research Assistant**, *Parasol Lab*, Texas A&M Univ., Adviser: Dr. Nancy Amato.
- Developed autonomous navigation and localization algorithms for mobile robots [46, 45].
- Researched path optimization algorithms for robotics using roadmaps [44].

2003   **Software Developer**, *CAPSHER Technology, Inc.*, Bryan, TX, USA.

Selected Synergistic Activities

2020 - present	Postdoctoral mentor for Dr. Trevor Steil (previously at University of Minnesota)
2020 - present	Postdoctoral mentor for Dr. Tahsin Reza (previously at University of British Columbia)
2017 - 2020	Postdoctoral mentor for Dr. Keita Iwabuchi (previously at Tokyo Tech)
2019	Presented overview of HPC Graph research at *Graph Fest*, National Security Agency (NSA)
2019	Served on the Ph.D. Defense Committe for Dr. Benjamin Priest, advised by Prof. George Cybenko at Dartmouth
2019	Served on the Ph.D. Defense Committe for Dr. Tahsin Reza, advised by Prof. Matei Ripeanu at the University of British Columbia (UBC)

Honors and Awards

2020	Global Security/Z Program team Silver Award, Lawrence Livermore National Laboratory
2017, '18, '19	IEEE HPEC GraphChallenge Champion

- *"One Quadrillion Triangles Queried on One Million Processors"* [8]
- *"K-truss decomposition for Scale-Free Graphs at Scale in Distributed Memory"* [16]
- *"Triangle counting for scale-free graphs at scale in distributed memory"* [22]

2017, 2018	Outstanding Mentor Award, Lawrence Livermore National Laboratory

- For mentorship of graduate and undergraduate students.

2017	Best Paper Award, IEEE Int. Workshop on High-Performance Big Data Computing [24]

- *"Performance evaluation of scale-free graph algorithms in low latency Non-Volatile Memory"*

2011	Computation Directorate Noteworthy Achievement Award, Lawrence Livermore National Lab
2009	LSP Fellow (Lawrence Scholar Program), Lawrence Livermore National Laboratory
2006	GAANN Fellow (Graduate Assistance in Areas of National Need), Texas A&M University
2003	University Undergraduate Research Fellow, Texas A&M University

- First place for senior honors thesis in category of engineering and physics

1999	Member of Eagle Scout Association

Publications – Google Scholar Profile

Peer-reviewed Publications. * denotes students or postdocs I supervised.

[1] Tahsin Reza*, Hassan Halawa, Matei Ripeanu, Geoffrey Sanders, and **Roger Pearce**. Scalable pattern matching in metadata graphs via constraint checking. *ACM Transactions on Parallel Computing (TOPC)*, 8(1):1–45, 2021.

[2] Benjamin Brock, Aydin Buluç, Timothy G Mattson, Scott McMillan, José E Moreira, **Roger Pearce**, Oguz Selvitopi, and Trevor Steil*. Considerations for a distributed graphblas api. In *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 215–218. IEEE, 2020.

[3] Ivy Peng, **Roger Pearce**, and Maya Gokhale. On the memory underutilization: Exploring disaggregated memory on hpc systems. In *2020 IEEE 32nd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*, pages 183–190. IEEE, 2020.

[4] Tahsin Reza*, Matei Ripeanu, Geoffrey Sanders, and **Roger Pearce**. Approximate pattern matching in massive graphs with precision and recall guarantees. In *ACM SIGMOD/PODS International Conference on Management of Data*, 2020.

[5] Tahsin Reza*, Matei Ripeanu, Geoffrey Sanders, and **Roger Pearce**. Towards interactive pattern search in massive graphs. In *Proceedings of the 3rd Joint International Workshop on Graph Data Management Experiences; Systems (GRADES) and Network Data Analytics (NDA)*, GRADES-NDA'20, New York, NY, USA, 2020. Association for Computing Machinery.

[6] Trevor Steil*, Scott McMillan, Geoffrey Sanders, **Roger Pearce**, and Benjamin Priest. Kronecker graph generation with ground truth for 4-cycles and dense structure in bipartite graphs. In *2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 237–246. IEEE, 2020.

[7] Keita Iwabuchi*, Lance Lebanoff*, Maya Gokhale, and **Roger Pearce**. Metall: A persistent memory allocator enabling graph processing. In *2019 IEEE/ACM 9th Workshop on Irregular Applications: Architectures and Algorithms (IA3)*, pages 39–44. IEEE, 2019.

[8] **Roger Pearce**, Trevor Steil*, Benjamin Priest*, and Geoffrey Sanders. One quadrillion triangles queried on one million processors. In *2019 IEEE High Performance extreme Computing Conference (HPEC)*. IEEE, 2019. **(Selected Champion)**.

[9] Ivy Peng, Marty McFadden, Eric Green, Keita Iwabuchi*, Kai Wu, Dong Li, **Roger Pearce**, and Maya Gokhale. Umap: Enabling application-driven optimizations for page management. In *2019 IEEE/ACM Workshop on Memory Centric High Performance Computing (MCHPC)*, pages 71–78. IEEE, 2019.

[10] Benjamin Priest*, Trevor Steil*, Geffrey Sanders, and **Roger Pearce**. You've got mail: Building missing asynchronous communication primitives. In *GrAPL 2019: Workshop on Graphs, Architectures, Programming, and Learning*, 2019.

[11] Scott Sallinen*, **Roger Pearce**, and Matei Ripeanu. Incremental graph processing for on-line analytics. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2019.

[12] Trevor Steil*, Benjamin Priest*, Geoffrey Sanders, **Roger Pearce**, Timothy La Fond, and Keita Iwabuchi. Distributed kronecker graph generation with ground truth of many graph properties. In

2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pages 251–260. IEEE, 2019.

[13] Keita Iwabuchi*, Geoffrey Sanders, Keith Henderson, and **Roger Pearce**. Computing exact vertex eccentricity on massive-scale distributed graphs. In *2018 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 257–267. IEEE, 2018.

[14] Jeremy Kepner, Siddharth Samsi, William Arcand, David Bestor, Bill Bergeron, Tim Davis, Vijay Gadepally, Michael Houle, Matthew Hubbell, Hayden Jananthan, Michael Jones, Anna Klein, Peter Michaleas, **Roger Pearce**, Lauren Milechin, Julie Mullen, Andrew Prout, Antonio Rosa, Geoff Sanders, Charles Yee, and Albert Reuther. Design, generation, and validation of extreme scale power-law graphs. In *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 279–286. IEEE, 2018.

[15] Yuechao Pan*, **Roger Pearce**, and John D Owens. Scalable breadth-first search on a gpu cluster. In *2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 1090–1101. IEEE, 2018.

[16] **Roger Pearce** and Geoffrey Sanders. K-truss decomposition for scale-free graphs at scale in distributed memory. In *2018 IEEE High Performance extreme Computing Conference (HPEC)*, pages 1–6. IEEE, 2018. **(Selected Champion)**.

[17] Benjamin W Priest*, **Roger Pearce**, and Geoffrey Sanders. Estimating edge-local triangle count heavy hitters in edge-linear time and almost-vertex-linear space. In *2018 IEEE High Performance extreme Computing Conference (HPEC)*, pages 1–7. IEEE, 2018.

[18] Tahsin Reza*, Matei Ripeanu, Nicolas Tripoul, Geoffrey Sanders, and **Roger Pearce**. Prunejuice: pruning trillion-edge graphs to a precise pattern-matching solution. In *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 265–281. IEEE, 2018.

[19] Geoffrey Sanders, **Roger Pearce**, Timothy La Fond, and Jeremy Kepner. On large-scale graph generation with validation of diverse triangle statistics at edges and vertices. In *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 287–296. IEEE, 2018.

[20] Nicolas Tripoul, Hassan Halawa, Tahsin Reza*, Geoffrey Sanders, **Roger Pearce**, and Matei Ripeanu. There are trillions of little forks in the road. choose wisely! In *2018 IEEE/ACM 8th Workshop on Irregular Applications: Architectures and Algorithms (IA3)*, pages 20–27. IEEE, 2018.

[21] Sam Adé Jacobs, Nikoli Dryden, **Roger Pearce**, and Brian Van Essen. Towards scalable parallel training of deep neural networks. In *Proceedings of the Machine Learning on HPC Environments*, MLHPC'17, pages 5:1–5:9, New York, NY, USA, 2017. ACM.

[22] **Roger Pearce**. Triangle counting for scale-free graphs at scale in distributed memory. In *2017 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–4. IEEE, 2017. **(Selected Champion)**.

[23] Tahsin Reza*, Christine Klymko, Matei Ripeanu, Geoffrey Sanders, and **Roger Pearce**. Towards practical and robust labeled pattern matching in trillion-edge graphs. In *2017 IEEE International Conference on Cluster Computing (CLUSTER)*, pages 1–12. IEEE, 2017. **(Nominated for best paper)**.

[24] Manu Shantharam, Keita Iwabuchi, Pietro Cicotti, Laura Carrington, Maya Gokhale, and **Roger Pearce**. Performance evaluation of scale-free graph algorithms in low latency non-volatile memory. In *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 1021–1028. IEEE, 2017. **(Best Paper Award)**.

[25] Scott Sallinen*, Keita Iwabuchi*, Suraj Poudel*, Maya Gokhale, Matei Ripeanu, and **Roger Pearce**. Graph colouring as a challenge problem for dynamic graph processing on distributed systems. In *SC'16: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 347–358. IEEE, 2016.

[26] Julia Bernd, Damian Borth, Carmen Carrano, Jaeyoung Choi, Benjamin Elizalde, Gerald Friedland, Luke Gottlieb, Karl Ni, **Roger Pearce**, Doug Poland, et al. Kickstarting the commons: The yfcc100m and the yli corpora. In *Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions*, pages 1–6. ACM, 2015.

[27] Brian Van Essen, Henry Hsieh, Sasha Ames, **Roger Pearce**, and Maya Gokhale. Di-mmap—a scalable memory-map runtime for out-of-core data-intensive applications. *Cluster Computing*, 18(1):15–28, 2015.

[28] Brian Van Essen, Hyojin Kim, **Roger Pearce**, Kofi Boakye, and Barry Chen. Lbann: Livermore big artificial neural network hpc toolkit. In *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments*, page 5. ACM, 2015.

[29] **Roger Pearce**, Maya Gokhale, and Nancy M Amato. Faster parallel traversal of scale free graphs at extreme scale with vertex delegates. In *SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 549–559. IEEE, 2014.

[30] **Roger Pearce**, Maya Gokhale, and Nancy M. Amato. Scaling techniques for massive scale-free graphs in distributed (external) memory. In *IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, 2013.

[31] Brian Van Essen, Henry Hsieh, Sasha Ames, **Roger Pearce**, and Maya Gokhale. DI-MMAP – a scalable memory-map runtime for out-of-core data-intensive applications. *Cluster Computing*, pages 1–14, 2013.

[32] Brian Van Essen, **Roger Pearce**, Sasha Ames, and Maya Gokhale. On the role of NVRAM in data-intensive architectures: an evaluation. In *IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, Shanghai, China, May 2012.

[33] Andy Yoo, Allison H. Baker, **Roger Pearce**, and Van Emden Henson. A scalable eigensolver for large scale-free graphs using 2D partitioning. In *Supercomputing 2011 (SC'11)*, Seattle, WA, November 2011.

[34] Brian Van Essen, **Roger Pearce**, Sasha Ames, and Maya Gokhale. On the role of NVRAM in data intensive HPC architectures. In *Workshop on Emerging Supercomputing Technologies*, Tucson, AZ, May 2011.

[35] **Roger Pearce**, Maya Gokhale, and Nancy M. Amato. Multithreaded asynchronous graph traversal for in-memory and semi-external memory. In *Supercomputing 2010 (SC'10)*, New Orleans, LA, November 2010.

[36] **Roger Pearce**, Marco Morales, and Nancy M. Amato. Structural improvement filtering strategy for PRM. In *Proceedings of Robotics: Science and Systems (RSS)*, Zurich, Switzerland, June 2008.

[37] Maya Gokhale, Jonathan Cohen, Andy Yoo, W. Marcus Miller, Arpith Jacob, Craig Ulmer, and **Roger Pearce**. Hardware technologies for high-performance data-intensive computing. In *IEEE Computer*, pages 60–68, April 2008. (Cover Feature).

[38] Marco A. Morales A., **Roger Pearce**, and Nancy M. Amato. Analysis of the evolution of c-space models built through incremental exploration. In *Proceedings of IEEE International Conference on Robotics Automation (ICRA)*, Rome, Italy, April 2007.

[39] Samuel Rodriguez, Shawna Thomas, **Roger Pearce**, and Nancy M. Amato. RESAMPL: A region-sensitive adaptive motion planner. In *Proceedings of International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, New York City, NY, July 2006.

[40] Dawen Xie, Marco A. Morales A., **Roger Pearce**, Shawna Thomas, Jyh-Ming Lien, and Nancy M. Amato. Incremental map generation (IMG). In *Proceedings of International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, New York City, NY, July 2006.

[41] Marco A. Morales A., **Roger Pearce**, and Nancy M. Amato. Metrics for analyzing the evolution of c-space models. In *Proceedings of IEEE International Conference on Robotics Automation (ICRA)*, Orlando, FL, May 2006.

[42] Marco A. Morales A., Lydia Tapia, **Roger Pearce**, Samuel Rodriguez, and Nancy M. Amato. C-space subdivision and integration in feature-sensitive motion planning. In *Proceedings of IEEE Inter- national Conference on Robotics Automation (ICRA)*, Barcelona, Spain, May 2005.

[43] Marco Morales, Lydia Tapia, **Roger Pearce**, Samuel Rodriguez, and Nancy M. Amato. A machine learning approach for feature-sensitive motion planning. In *Proceedings of International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, Utrecht/Zeist, The Netherlands, July 2004.

[44] Jinsuck Kim, **Roger Pearce**, and Nancy M. Amato. Extracting optimal paths from roadmaps for motion planning. In *Proceedings of IEEE International Conference on Robotics Automation (ICRA)*, Taipei, Taiwan, September 2003.

[45] Jinsuck Kim, **Roger Pearce**, and Nancy M. Amato. Feature-based localization using scannable visibility sectors. In *Proceedings of IEEE International Conference on Robotics Automation (ICRA)*, Taipei, Taiwan, September 2003.

[46] Jinsuck Kim, **Roger Pearce**, and Nancy M. Amato. Robust geometric-based localization in indoor environments using sonar sensors. In *Proceedings of IEEE/RSJ International Conference on intelligent Robots and Systems (IROS)*, October 2002.

Collaborators and Co-Editors

Co-Authors Nancy Amato (UIUC), Sasha Ames (LLNL), Allison H Baker (NCAR), Kofi Boakye (Flickr at Yahoo), Laura Carrington (SDSC), Barry Chen (LLNL), Pietro Cicotti (SDSC), Nikoli Dryden (ETH Zurich), Maya Gokhale (LLNL), Eric Green (LLNL), Hassan Halawa (UBC), Keith Henderson (LLNL), Henry Hsieh (unknown), Keita Iwabuchi (LLNL), Sam Adé Jacobs (LLNL), Jeremy Kepner (MIT-LL), Jinsuck Kim (IMVU), Christine Klymko (LLNL), Timothy La Fond (LLNL), Dong Li (UC Merced), G Scott Lloyd (LLNL), Satoshi Matsuoka (Tokyo Tech), Marty McFadden (LLNL), Marco Morales (ITAM), Karl Ni (Google), John Owens (UC Davis), Yuechao Pan (Google), Ivy Peng (LLNL), Suraj Poudel (Facebook), Benjamin Priest (LLNL), Tahsin Reza (UBC), Matei Ripeanu (UBC), Samuel Rodriguez (Texas Wesleyan), Scott Sallinen (UBC), Geoff Sanders (LLNL), Manu Shantharam (SDSC), Trevor Steil (UMN), Lydia Tapia (UNM), Shawna Thomas (TAMU), Nicolas Tripoul (UBC), Craig Ulmer (SNL/CA), Brian Van Essen (LLNL), Kai Wu (UC Merced), Andy Yoo (LLNL)

Ph.D. Advisor Nancy Amato (UIUC)

Ph.D. Committee Marvin Adams (TAMU), Nancy Amato (UIUC), Yoonsuck Choe (TAMU), Maya Gokhale (LLNL), Lawrence Rauchwerger (UIUC)

Fellowship Advisor Maya Gokhale (LLNL)

Postdoc Advisees Keita Iwabuchi (LLNL), Tahsin Reza (LLNL), Trevor Steil(LLNL)

Ph.D. Committees Benjamin Priest (advised by Prof. George Cybenko at Dartmouth), Tahsin Reza (advised by Prof. Matei Ripeanu at UBC)

Supervised Interns Allyson Cauble-Chantrenne (EP Analytics), Yuxin Chen (UC Davis), Jocelyn Cheng (IBM), Ryan Dozier (UCF), Steven Feldman (Google), Keita Iwabuchi (LLNL), Griffin Kimura (UC Davis), Vasant Kurvari (TAMU), Lance Lebanof (Leidos), Brian Page (IU), Yuechao Pan (Google), Jae Sang Park(TAMU), Suraj Poudel (Facebook), Benjamin Priest (Dartmouth), Tahsin Reza (UBC), Scott Sallinen (UBC), Trevor Steil (UMN), Ancy Tom (UMN), Kaushik Velusamy (UMBC), Elizabeth Wang (UIUC), Karim Youssef (VT), Jiyuan Zhang (CMU)