Gopinath Chennupati

3000 Trinity Dr, Apt 30, Los Alamos, NM, 87544
cgnath.dr@gmail.com
http://cgnath.gitlab.io
(+1) 505-570-0914

Education and Qualifications

2012 – 2015 **Ph.D.**, *Computer Science*, University of Limerick, Ireland
Thesis: Grammatical Evolution + Multi-Cores = Automatic Parallel Programming!

2010 – 2011 **M.Sc.**, *Web Technologies*, National College of Ireland, Ireland
Thesis: Adaptation of Web Services Based on User Preferences & User Evaluation

2006 – 2010 **B.Tech.**, *Information Technology*, Jawaharlal Nehru Technological University, India.

Professional Experience

Jan 2021 – Current **Applied Scientist**, Amazon Alexa, Sunnyvale, CA.

- Federated Learning, Automatic Speech Recognition (ASR), Language/speech understanding, semi-supervised learning

May 2019 – Dec 2020 **Scientist**, Los Alamos National Laboratory, Los Alamos, NM.

- Non-negative matrix/tensor factorization, MPI, decomposed 10 TB data
- NLP, Deep Learning models, uncertainty quantification
- Probabilistic models for **performance prediction** of scientific codes

July 2016 – Apr, 2019 **Post-doc**, Information Sciences Group (CCS-3), Los Alamos National Laboratory, Los Alamos, NM.

- Simulated half-an-exascale (16.7M node) super computer on 4096 ranks
- Deep learning, NLP, protein 3D structure prediction, graph theory, ensemble learning

Jan – Nov 2016 **Data Scientist**, Mobacar, Killarney, Ireland.

- World's first retail intelligent ground transportation, Decision Neural Analytics (DNA)
- Serves 650 suppliers in 30,000 locations with a response time of 0.1 sec

Sep 2012 – Dec 2015 **Teaching Assistant**, Univeristy of Limerick, Limerick, Ireland.

- Tutored Object Oriented Programming (Java) for 130+ students each year

Aug 2011 – Jan 2012 **Production Engineer**, Lionbridge, Dublin, Ireland.

- Implemented Xbox web sites in HTML, translated and tested in 40+ locales
- Reduced the broken links by a factor of 11% over the previous release

PhD Dissertation Summary

- The key contribution is the synthesis of parallel programs (directly compliable source code) on multi-core architectures (OpenMP parallelization and Context Free Grammars) using a machine learning [C25, C26, C27, C29].
- The parallel programs were deployed on an Intel Xeon E7-4820 and optimized the speed-up of the synthesized parallel programs by a factor of 25.21% limiting the excessive thread on an with 32 cores [C22, C24].
- Performance of the synthesized parallel programs competes with the human written parallel programs. These programs adapt to the underlying hardware architectures when run on varying number of cores [C21, C23].

Awards and Honors

2015	**Humies Silver Award** ACM Special Interest Group for Genetic and Evolutionary Computation (SIGEVO), 17^{th} Genetic and Evolutionary Computation Conference (GECCO), 2015, Madrid, Spain
2015	**Evostar Student Accommodation Bursary** 18^{th} EuroGP, Copenhagen, Denmark
2014	**Best Graduate Student Contribution Award** 16^{th} GECCO, 2014, Vancouver, Canada
2014, 2015	**ACM Student Travel Grants** To attend GECCO – 2014 (Canada) and GECCO – 2015 (Spain)
2012 – 2015	**PhD Scholarship** Funded by Science Foundation of Ireland (SFI) during the years 2012 – 2015

Software Implementations

DnMFk Distributed Nonnegative Matrix Factorization [`https://github.com/lanl/DnMFk`]

- MPI, large scale unsupervised learning, decomposed 10TB matrices, clustering, silhouettes

PPT Performance Prediction Toolkit [`https://github.com/lanl/PPT`]

- Developed hardware models for Broadwell, Haswell and AMD processors
- Scalable simulation of a super computer (16.7M ranks) on a super computer (4096 ranks)
- Thousands of lines of Python scripts, and scientific codes in C/C++

NinjaDock Protein-Protein 3D Structure Prediction

- Protein surface structure extraction to identify graph features for optimal docking poses
- Python scripts for processing docking results, datasets and ML models
- Deploy machine learning models for optimal docking poses out of millions of results

GAPP Grammatical Automatic Parallel Programming [in C/C++ (10000+ lines)]

- Optimized the speed-up of the synthesized parallel programs by a factor of 25.21% limiting the excessive thread on an Intel Xeon E7-4820 with 32 cores
- Synthesized *architecture adaptive parallel programs* on multi-cores

RPM Run Prediction Model

- Meta-learning in Java (2500+ lines), improved accuracy by 9%, optimized runtime by 14%

Programming Skills

Experienced	Python, C/C++, Java, MPI, OpenMP, PyTorch
Good	TensorFlow, Scikit-Learn, Keras, Julia

Professional Activities

Professional	Member Association for Computing Machinery (ACM), IEEE High Performance Extreme Computing Conference (HPEC), Program Committee member of International Conference on Simulation and Modeling Methodologies Technologies and Applications (SIMULTECH), Program Committee member of Neural Information and Processing Systems (NeurIPS) 2020, Co-lead Applied Machine Learning Summer Research Fellowship at Los Alamos National Laboratory (LANL), Advisory committee member LANL Research Library 2019–2020
Referee Conferences	International Conference on Learning Representations (ICLR) 2021, NeurIPS 2021 and 2020, ICML 2021, SIMULTECH 2021 – 2019, ACM International Symposium on Memory Systems (MEMSYS) 2020 and 2018, ACM International Conference on Parallel Processing (ICPP) 2018, IEEE International Performance Computing and Communications Conference (IPCCC) 2018, Session chair of the *Interconnects* at Supercomputing Conference (SC) 2017
Referee Journals	IEEE Transaction on Parallel and Distributed Systems (TPDS), ACM Transaction on Modeling and Computer Simulation (TOMACS), IEEE Access, Journal of Supercomputing, Elsevier Parallel Computing, IEEE Journal on Selected Areas in Communications (JSAC) series on Machine Learning for Communications and Networks, Elsevier Journal of Parallel and Distributed Computing (JPDC)

Graduate Student Mentoring

Name		Thesis Title	Publications
Raviteja Vangara	Post-doc, 2020	Distributed nonnegative matrix factorizations	[J3]
Bhargava Kalla	M.S., 2017	Probabilistic monte carlo framework for branch prediction	[C19, C20]
Yehia Arafa	Ph.D., (Ongoing)	Performance prediction models on GPUs	[C13, J6]
Atanu Barai	Ph.D., (Ongoing)	Performance prediction models on CPUs	[P1]
Nasrin Akhter	Ph.D. (2020)	Decoy selection strategies using energy landscapes	[C12, J1, J5]

Publications

Ph.D. Thesis

D1. **G Chennupati** (2015). "Grammatical evolution+ multi-cores= automatic parallel programming!" PhD thesis. University of Limerick.

Book Chapters

B1. **G Chennupati**, R Azad, C Ryan, S Eidenbenz, and N Santhi (2018). "Synthesis of Parallel Programs on Multi-Cores". In: *Handbook on Grammatical Evolution*. Springer, pp.289–315.

Journal Articles

J1. N Akhter, **G Chennupati**, H Djidjev, and A Shehu (2020). Decoy Selection for Protein Structure Prediction Via Extreme Gradient Boosting and Ranking. *BMC Bioinformatics* **21**(1), x–x. DOI: 10.1186/s12859-020-3523-9.

J2. H Carrillo-Cabada, E Skau, **G Chennupati**, B Alexandrov, and H Djidjev (2020). An Out of Memory tSVD for Big-Data Factorization. *IEEE Access* **8**, 107749–107759.

J3. **G Chennupati**, R Vangara, E Skau, H Djidjev, and B Alexandrov (2020). Distributed non-negative matrix factorization with determination of the number of latent features. *The Journal of Supercomputing* **76**(9), 7458–7488. DOI: 10.1007/s11227-020-03181-6.

J4. P Haridas, **G Chennupati**, N Santhi, P Romero, and SJ Eidenbenz (2020). Code Characterization With Graph Convolutions and Capsule Networks. *IEEE Access* **8**, 136307–136315. DOI: 10.1109/ACCESS.2020.3011909.

J5. N Akhter, **G Chennupati**, KL Kabir, H Djidjev, and A Shehu (2019). Unsupervised and Supervised Learning over the Energy Landscape for Protein Decoy Selection. *Biomolecules* **9**(10).

J6. Y Arafa, A Badawy, **G Chennupati**, N Santhi, and S Eidenbenz (Jan. 1, 2019). PPT-GPU: Scalable GPU Performance Modeling. *IEEE Computer Architecture Letters* **18**(1), 55–58. DOI: 10.1109/LCA.2019.2904497.

In Conference Proceedings

C1. Y Arafa, AA Badawy, **G Chennupati**, A Barai, N Santhi, and SJ Eidenbenz (2020). Fast, Accurate, and Scalable Memory Modeling of GPGPUs using Reuse Profiles. In: *In Proceedings of The 34th ACM International Conference on Supercomputing (ICS)*. ICS '20. ACM, pp.31:1–31:12.

C2. Y Arafa, A ElWazir, A ElKanishy, Y Aly, A Elsayed, AH Badawy, **G Chennupati**, S Eidenbenz, and N Santhi (2020). Verified Instruction-Level Energy Consumption Measurement for NVIDIA GPUs. In: *Proceedings of the 17th ACM International Conference on Computing Frontiers*. CF '20. ACM, pp.60–70.

C3. A Barai, **G Chennupati**, N Santhi, AA Badawy, Y Arafa, and SJ Eidenbenz (2020). PPT-SASMM: Scalable Analytical Shared Memory Model. In: *Press of the International Symposium on Memory Systems, MEMSYS*. ACM.

C4. M Bhattarai, **G Chennupati**, E Skau, R Vangara, H Djidjev, and BS Alexandrov (2020). Distributed Non-Negative Tensor Train Decomposition. In: *Press IEEE High Performance Extreme Computing Conference, HPEC*.

C5. N Akhter, R Vangara, **G Chennupati**, BS Alexandrov, H Djidjev, and A Shehu (2019). Non-Negative Matrix Factorization for Selection of Near-Native Protein Tertiary Structures. In: *International Conference on Bioinformatics and Biomedicine BIBM*. IEEE, pp.70–73.

C6. Y Arafa, AA Badawy, **G Chennupati**, N Santhi, and SJ Eidenbenz (2019). Low Overhead Instruction Latency Characterization for NVIDIA GPGPUs. In: *High Performance Extreme Computing Conference, HPEC*. IEEE, pp.1–8.

C7. Y Arafa, AA Badawy, **G Chennupati**, N Santhi, and SJ Eidenbenz (2019). POSTER: GPUs Pipeline Latency Analysis. In: *International Conference on Application-specific Systems, Architectures and Processors, ASAP*. IEEE, pp.139.

C8. Y Arafa, **G Chennupati**, A Barai, AA Badawy, N Santhi, and SJ Eidenbenz (2019). GPUs Cache Performance Estimation using Reuse Distance Analysis. In: *International Performance Computing and Communications Conference, IPCCC*. IEEE, pp.1–8.

C9. **G Chennupati**, N Santhi, and S Eidenbenz (2019). Scalable Performance Prediction of Codes with Memory Hierarchy and Pipelines. In: *SIGSIM Conference on Principles of Advanced Discrete Simulation, SIGSIM-PADS*. ACM, pp.13–24.

C10. S Thulasidasan, T Bhattacharya, JA Bilmes, **G Chennupati**, and J Mohd-Yusof (2019). Combating Label Noise in Deep Learning using Abstention. In: *Proceedings of the 36th International Conference on Machine Learning, ICML*. PMLR, pp.6234–6243.

C11. S Thulasidasan, **G Chennupati**, JA Bilmes, T Bhattacharya, and S Michalak (2019). On Mixup Training: Improved Calibration and Predictive Uncertainty for Deep Neural Networks. In: *Neural Information Processing Systems, NeurIPS*, pp.13888–13899.

C12. N Akhter, **G Chennupati**, H Djidjev, and A Shehu (2018). Improved Decoy Selection via Machine Learning and Ranking. In: *International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*. IEEE, pp.1–1.

C13. Y Arafa, AHA Badawy, **G Chennupati**, N Santhi, and S Eidenbenz (2018). PPT-GPU: Performance Prediction Toolkit for GPUs: Identifying the impact of caches. In: *International Symposium on Memory Systems (MEMSYS)*. ACM, pp.301–302.

C14. **G Chennupati**, S Eidenbenz, A Long, O Tkachenko, J Zerr, and J Liu (2018). IMCSIM: Parameterized performance prediction for implicit monte carlo codes. In: *Winter Simulation Conference (WSC)*. IEEE, pp.491–502.

C15. MA Obaida, J Liu, **G Chennupati**, N Santhi, and S Eidenbenz (2018). Parallel Application Performance Prediction Using Analysis Based Models and HPC Simulations. In: *SIGSIM Conference on Principles of Advanced Discrete Simulation*. ACM, pp.49–59.

C16. **G Chennupati**, N Santhi, R Bird, S Thulasidasan, AHA Badawy, S Misra, and S Eidenbenz (2017). A scalable analytical memory model for cpu performance prediction. In: *International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems*. Springer, pp.114–135.

C17. **G Chennupati**, N Santhi, S Eidenbenz, and S Thulasidasan (2017). AMM: scalable memory reuse model to predict the performance of physics codes. In: *IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, pp.649–650.

C18. **G Chennupati**, N Santhi, S Eidenbenz, and S Thulasidasan (2017). An analytical memory hierarchy model for performance prediction. In: *Winter Simulation Conference (WSC)*. IEEE, pp.908–919.

C19. B Kalla, N Santhi, AHA Badawy, **G Chennupati**, and S Eidenbenz (2017). A probabilistic monte carlo framework for branch prediction. In: *IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, pp.651–652.

C20. B Kalla, N Santhi, AHA Badawy, **G Chennupati**, and S Eidenbenz (2017). Probabilistic Monte Carlo simulations for static branch prediction. In: *IEEE International Performance Computing and Communications Conference (IPCCC)*. IEEE, pp.1–4.

C21. **G Chennupati**, RMA Azad, and C Ryan (2016). Automatic lock-free parallel programming on multi-core processors. In: *IEEE Congress on Evolutionary Computation (CEC)*. IEEE, pp.4143–4150.

C22. **G Chennupati**, R Azad, and C Ryan (2015). On the automatic generation of efficient parallel iterative sorting algorithms. In: *Genetic and evolutionary computation conference companion*. ACM, pp.1369–1370.

C23. **G Chennupati**, R Azad, and C Ryan (2015). Performance optimization of multi-core grammatical evolution generated parallel recursive programs. In: *Genetic and evolutionary computation conference*. ACM, pp.1007–1014.

C24. **G Chennupati**, R Azad, and C Ryan (2015). Synthesis of parallel iterative sorts with multi-core grammatical evolution. In: *Genetic and evolutionary computation conference companion*. ACM, pp.1059–1066.

C25. **G Chennupati**, RMA Azad, and C Ryan (2015). Automatic evolution of parallel recursive programs. In: *European Conference on Genetic Programming*. Springer, pp.167–178.

C26. **G Chennupati**, RMA Azad, and C Ryan (2015). Automatic evolution of parallel sorting programs on multi-cores. In: *European Conference on the Applications of Evolutionary Computation*. Springer, pp.706–717.

C27. **G Chennupati**, RMA Azad, and C Ryan (2014). Multi-core GE: automatic evolution of CPU based multi-core parallel programs. In: *Genetic and evolutionary computation conference companion*. ACM, pp.1041–1044.

C28. **G Chennupati**, RA Azad, and C Ryan (2014). Predict the performance of GE with an ACO based machine learning algorithm. In: *Genetic and evolutionary computation conference companion*. ACM, pp.1353–1360.

C29. **G Chennupati**, J Fitzgerald, and C Ryan (2014). On The Efficiency of Multi-core Grammatical Evolution (MCGE) Evolving Multi-Core Pallel Programs. In: *Sixth World Congress on Nature and Biologically Inspired Computing*. IEEE, pp.238–243.

C30. **G Chennupati**, C Ryan, and RMA Azad (2014). Predict the success or failure of an evolutionary algorithm run. In: *Genetic and evolutionary computation conference companion*. ACM, pp.131–132.

C31. **G Chennupati**, C Ryan, and RMA Azad (2013). An Empirical Analysis Through the Time Complexity of GE Problems. In: *19th International Conference on Soft Computing, MENDEL 2013*, pp.37–44.

Technical Reports

T1. PJ Coles, S Eidenbenz, S Pakin, A Adedoyin, J Ambrosiano, P Anisimov, W Casper, **G Chennupati**, C Coffrin, H Djidjev, et al. (2018). Quantum Algorithm Implementations for Beginners. *arXiv preprint arXiv:1804.03719*.

T2. **G Chennupati** (2014). Eant-miner: an ensemble ant-miner to improve the ACO classification. *arXiv preprint arXiv:1409.2710*.

Under Preparation

P1. A Barai, AHA Badawy, **G Chennupati**, N Santhi, and S Eidenbenz (2020). Performance Prediction of loop parallel codes on multi-cores using shared memory models. *Computer Architecture Letters*. (Expected submission: 2020).

P2. **G Chennupati** and H Djidjev (2020). Extended Parallel Seed-based Approach with Machine Learning-based Scoring Function and Graph Models for Protein Structure Prediction. *Journal of Biomedical Informatics*. (Expected submission:: 2020).

P3. **G Chennupati**, K Ganguly, M Benjamin, S Thulasidasan, and T Bhattacharya (2020). Multi-task Learning to Extract Tumor Genomic Markers from Breast Cancer Pathology Reports and Uncertainty Quantification. *Journal of American Medical Informatics Association (JAMIA)*. (Expected submission: 2020).

P4. MA Obaida, J Liu, **G Chennupati**, N Santhi, and S Eidenbenz (2020). PyPaast: Parallel Application Performance Prediction Using Memory Models and HPC Simulations. *Transactions on Modeling and Computer Simulation (TOMACS)*. (Expected submission: 2020).

References

Available upon request.