

I work at the intersection of deep learning and high-performance computing. I build systems for training deep neural networks fast and at scale, design more efficient neural networks, and apply deep learning to scientific applications. My research focuses on identifying current or emerging computational bottlenecks for training deep neural networks and developing approaches to optimally utilize resources. I also work collaboratively with domain scientists to apply deep learning to science, leveraging new achievements in scale and efficiency, with a focus on weather & climate.

## EDUCATION

- 2014–2019 **PhD Computer Science** **University of Illinois at Urbana-Champaign**  
Committee: [Marc Snir](#), William Gropp, Wen-mei Hwu, Alexander Schwing, Brian Van Essen  
Thesis: *Large-Scale Training of Deep Neural Networks*
- 2010–2014 **BS Computer Science** **University of Illinois at Urbana-Champaign**  
Minor in Mathematics, James Scholar

## AWARDS & HONORS

- 2023 **R&D 100 Award** **Lawrence Livermore National Laboratory**  
Received as part of the Cancer Distributed Learning Environment (CANDLE) team.
- 2019 **ETH Postdoctoral Fellow** **ETH Zürich**  
Competitive fellowship for “young researchers who have already demonstrated scientific excellence”.
- 2019 **Kenichi Miura Award** **University of Illinois at Urbana-Champaign**  
“Honors a graduate student for excellence in high-performance computing”.
- 2018 **State Farm Doctoral Scholar** **University of Illinois at Urbana-Champaign**  
Fellowship “. . . to support outstanding continuing doctoral students”.
- 2013 **Student of the Year** **National Center for Supercomputing Applications**  
Awarded once per year to the most outstanding student working at NCSA.
- 2011 **“Defining the Future” 1<sup>st</sup> Place** **UIUC Engineering Open House**  
Top award in the theme category for Tacchi, an interactive multi-touch table with gesture recognition.

## PROFESSIONAL EXPERIENCE

- 2023–present **Computer Scientist** **Lawrence Livermore National Laboratory**  
Leading research on exascale high-performance & scalable deep learning, and scientific machine learning.
- 2019–2022 **ETH Postdoctoral Fellow** **ETH Zürich**  
Research on systems for deep learning, efficient neural networks, and deep learning for weather & climate in Torsten Hoefer’s Scalable Parallel Computing Laboratory. Mentored ten bachelors and ten masters students.
- 2014–2019 **Research Assistant** **University of Illinois at Urbana-Champaign**  
Research on algorithms and systems for scalable training of deep neural networks, and runtime systems for future exascale supercomputers in Marc Snir’s High-Performance Computing group.
- Summer 2016, 2017, 2018 **Computation Intern** **Lawrence Livermore National Laboratory**  
Research on scalable deep learning on supercomputers. Led development of the Aluminum communication library and contributed to the LBANN toolkit for training deep neural networks.
- Summer 2015 **Summer Intern** **Raytheon Centers of Innovation**  
Developed tools for cloud-based analysis of OS images and large-scale indexing/search of unstructured data.
- 2012–2014 **SPIN Fellow** **National Center for Supercomputing Applications**  
Member of the inaugural Students Pushing Innovation (SPIN) program. Subsequently co-mentored a student in the program. Research and development for parallel debugging tools on supercomputers.
- Summer 2012 **High Energy Density Physics Intern** **Lawrence Livermore National Laboratory**  
Developed PGDB, a parallel debugger for MPI applications at scale.

## PUBLICATIONS

\*: *Equal contribution.* †: *Bachelors or masters student at time of writing.*

- WACV 2024 Piotr Teterwak, Soren Nelson, **Nikoli Dryden**, Dina Bashkirova, Kate Saenko, and Bryan Plummer. “Learning to Compose SuperWeights for Neural Parameter Allocation Search”. Winter Conference on Applications of Computer Vision, 2024.
- arXiv 2023 Julia Bazińska†, Andrei Ivanov, Tal Ben-Nun, **Nikoli Dryden**, Maciej Besta, Siyuan Shen, and Torsten Hoefer. “Cached Operator Reordering: A Unified View for Fast GNN Training”. arXiv, 2023. [X](#)

- arXiv 2023 Andrei Ivanov, **Nikoli Dryden**, Tal Ben-Nun, Saleh Ashkboos, and Torsten Hoefler. “STen: Productive and Efficient Sparsity in PyTorch”. arXiv, 2023. [X](#)
- LOG 2022 Maciej Besta, Patrick Iff, Florian Scheidl<sup>†</sup>, Kazuki Osawa, **Nikoli Dryden**, Michał Podstawski, Tiancheng Chen<sup>†</sup>, and Torsten Hoefler. “Neural Graph Databases”. Learning on Graphs Conference, 2022. [X](#) [B](#)
- NeurIPS 2022 **Nikoli Dryden** and Torsten Hoefler. “Spatial Mixture-of-Experts”. Advances in Neural Information Processing Systems, 2022. [X](#) [B](#)
- NeurIPS 2022 Saleh Ashkboos, Langwen Huang, **Nikoli Dryden**, Tal Ben-Nun, Peter Dueben, Lukas Gianinazzi, Luca Kummer<sup>†</sup>, and Torsten Hoefler. “ENS-10: A Dataset for Post-Processing Ensemble Weather Forecasts”. Advances in Neural Information Processing Systems, 2022. [X](#) [B](#)
- SIGKDD 2022 Maciej Besta, Raphael Grob<sup>†</sup>, Cesare Miglioli, Nicola Bernoldi<sup>†</sup>, Grzegorz Kwasniewski, Gabriel Gjini<sup>†</sup>, Raghavendra Kanakagiri, Saleh Ashkboos, Lukas Gianinazzi, **Nikoli Dryden**, and Torsten Hoefler. “Motif Prediction with Graph Neural Networks”. ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022. [X](#) [B](#)
- ICS 2022 Oliver Rausch<sup>†</sup>, Tal Ben-Nun, **Nikoli Dryden**, Andrei Ivanov, Shigang Li, and Torsten Hoefler. “A Data-Centric Optimization Framework for Machine Learning”. ACM International Conference on Supercomputing, 2022. [X](#) [B](#)
- ICLR 2022 Bryan Plummer\*, **Nikoli Dryden\***, Julius Frost<sup>†</sup>, Torsten Hoefler, and Kate Saenko. “Neural Parameter Allocation Search”. International Conference on Learning Representations, 2022. [X](#) [B](#)
- arXiv 2021 Lukas Gianinazzi, Maximilian Fries<sup>†</sup>, **Nikoli Dryden**, Tal Ben-Nun, and Torsten Hoefler. “Learning Combinatorial Node Labeling Algorithms”. arXiv, 2021. [X](#)
- JMLR 2021 Torsten Hoefler, Dan Alistarh, Tal Ben-Nun, **Nikoli Dryden**, and Alexandra Peste. “Sparsity in Deep Learning: Pruning and Growth for Efficient Inference and Training in Neural Networks”. Journal of Machine Learning Research, 2021. [B](#)
- SC 2021 **Nikoli Dryden**, Roman Böhringer<sup>†</sup>, Tal Ben-Nun, and Torsten Hoefler. “Clairvoyant Prefetching for Distributed Machine Learning I/O”. Supercomputing, 2021. [X](#) [B](#)
- MLSys 2021 Andrei Ivanov\*, **Nikoli Dryden\***, Tal Ben-Nun, Shigang Li, and Torsten Hoefler. “Data Movement Is All You Need: A Case Study on Optimizing Transformers”. Conference on Machine Learning and Systems, 2021. **Oustanding paper** [X](#) [B](#)
- RSTA 2021 Peter Grönquist<sup>†</sup>, Chengyuan Yao<sup>†</sup>, Tal Ben-Nun, **Nikoli Dryden**, Peter Dueben, Shigang Li, and Torsten Hoefler. “Deep Learning for Post-Processing Ensemble Weather Forecasts”. Philosophical Transactions of the Royal Society A, 2021. [X](#) [B](#)
- TPDS 2020 Yosuke Oyama, Naoya Maruyama, **Nikoli Dryden**, Erin McCarthy, Peter Harrington, Jan Balewski, Satoshi Matsuoka, Peter Nugent, and Brian Van Essen. “The Case for Strong Scaling in Deep Learning: Training Large 3D CNNs with Hybrid Parallelism”. IEEE Transactions on Parallel and Distributed Systems, 2020. [X](#) [B](#)
- TPDS 2020 Shigang Li, Tal Ben-Nun, Giorgi Nadiradze, Salvatore Di Girolamo, **Nikoli Dryden**, Dan Alistarh, and Torsten Hoefler. “Breaking (Global) Barriers in Parallel Stochastic Optimization with Wait-Avoiding Group Averaging”. IEEE Transactions on Parallel and Distributed Systems, 2020. [X](#) [B](#)
- ML4PS 2019 Peter Grönquist<sup>†</sup>, Tal Ben-Nun, **Nikoli Dryden**, Peter Dueben<sup>†</sup>, Luca Lavarini, Shigang Li, and Torsten Hoefler. “Predicting Weather Uncertainty with Deep Convnets”. Machine Learning and the Physical Sciences Workshop @ NeurIPS, 2019. [X](#)
- SC 2019 **Nikoli Dryden**, Naoya Maruyama, Tim Moon, Tom Benson, Marc Snir, and Brian Van Essen. “Channel and Filter Parallelism for Large-Scale CNN Training”. Supercomputing, 2019. [B](#)
- IPDPS 2019 **Nikoli Dryden\***, Naoya Maruyama\*, Tom Benson, Tim Moon, Marc Snir, and Brian Van Essen. “Improving Strong-Scaling of CNN Training by Exploiting Finer-Grained Parallelism”. IEEE International Parallel and Distributed Processing Symposium, 2019. [X](#) [B](#)
- MLHPC 2018 **Nikoli Dryden**, Naoya Maruyama, Tim Moon, Tom Benson, Andy Yoo, Marc Snir, and Brian Van Essen. “Aluminum: An Asynchronous, GPU-aware Communication Library Optimized for Large-Scale Training of Deep Neural Networks on HPC Systems”. Workshop on Machine Learning and High Performance Computing, 2018. [B](#)
- Cluster 2018 Chen Wang, **Nikoli Dryden**, Franck Cappello, and Marc Snir. “Neural Network Based Silent Error Detector”. IEEE Cluster Conference, 2018. **Best paper** [B](#)
- PLDI 2018 Roshan Dathathri, Gurbinder Gill, Loc Hoang, Hoang-Vu Dang, Alex Brooks, **Nikoli Dryden**, Andrew Lenharth, Marc Snir, and Keshav Pingali. “Gluon: A Communication Optimizing Framework for Distributed Heterogeneous Graph Analytics”. ACM SIGPLAN Conference on Programming Language Design and Implementation, 2018. [B](#)

- IPDPS 2018 Hoang-Vu Dang, Roshan Dathathri, Gurbinder Gill, Alex Brooks, **Nikoli Dryden**, Andrew Lenharth, Loc Hoang, Keshav Pingali, and Marc Snir. “A Lightweight Communication Runtime for Distributed Graph Analytics”. IEEE International Parallel and Distributed Processing Symposium, 2018. [📄](#)
- MLHPC 2017 Sam Adé Jacobs, **Nikoli Dryden**, Roger Pearce, and Brian Van Essen. “Towards Scalable Parallel Training of Deep Neural Networks”. Workshop on Machine Learning and High Performance Computing, 2017. [📄](#)
- MLHPC 2016 **Nikoli Dryden**, Tim Moon, Sam Adé Jacobs, and Brian Van Essen. “Communication Quantization for Data-parallel Training of Deep Neural Networks”. Workshop on Machine Learning and High Performance Computing, 2016. [📄](#)
- ESPM2 2015 Alex Brooks, Hoang-Vu Dang, **Nikoli Dryden**, and Marc Snir. “PPL: An abstract runtime system for hybrid parallel programming”. ACM/IEEE International Workshop on Extreme Scale Programming Models and Middleware, 2015. [📄](#)
- XSEDE 2014 **Nikoli Dryden**. “PGDB: A Debugger for MPI Applications”. Conference on Extreme Science and Engineering Discovery Environment, 2014. [📄](#)

## POSTERS

- LOG 2022 Maciej Besta, Patrick Iff, Florian Scheidl, Kazuki Osawa, **Nikoli Dryden**, Michał Podstawski, Tiancheng Chen, and Torsten Hoefler. “Neural Graph Databases”. Learning on Graphs Conference, 2022.
- NeurIPS 2022 **Nikoli Dryden** and Torsten Hoefler. “Spatial Mixture-of-Experts”. NeurIPS 2022.
- NeurIPS 2022 Saleh Ashkboos, Langwen Huang, **Nikoli Dryden**, Tal Ben-Nun, Peter Dueben, Lukas Gianinazzi, Luca Kummer, and Torsten Hoefler. “ENS-10: A Dataset for Post-Processing Ensemble Weather Forecasts”. NeurIPS 2022.
- ML4PS 2019 Peter Grönquist, Tal Ben-Nun, **Nikoli Dryden**, Peter Dueben, Luca Lavarini, Shigang Li, and Torsten Hoefler. “Predicting Weather Uncertainty with Deep ConvNets”. Machine Learning and the Physical Sciences Workshop @ NeurIPS, 2019.
- GTC 2019 Naoya Maruyama, **Nikoli Dryden**, Tom Benson, Tim Moon, Brian Van Essen, and Marc Snir. “Spatial Parallel Convolution for Scalable Training of Large Neural Networks”. GPU Technology Conference, 2019.
- SysML 2018 **Nikoli Dryden**, Naoya Maruyama, Tom Benson, Tim Moon, Marc Snir, and Brian Van Essen. “Scalable CNN Training on Large-Scale HPC Systems”. Workshop on Systems for ML and Open Source Software @ NeurIPS, 2018.
- LLNL DSI 2018 Naoya Maruyama, **Nikoli Dryden**, Tim Moon, Brian Van Essen, and Marc Snir. “Generalized Distributed-Memory Convolutional Neural Networks for Large-Scale Parallel Systems”. LLNL Data Science Institute Workshop, 2018.
- LLNL DSI 2018 Sam Adé Jacobs, **Nikoli Dryden**, Tim Moon, Brian Van Essen, Stewart He, and Jonathan Allen. “Scaling Deep Learning for Cancer Drug Discovery on HPC Systems”. LLNL Data Science Institute Workshop, 2018.
- Salishan 2017 **Nikoli Dryden**, Brian Van Essen, and Marc Snir. “Gradient Quantization for Data-Parallel DNN Training”. Salishan Conference on High Speed Computing, 2017.

## TALKS

- LLNL 2022 **Optimizing Systems for Deep Learning and Deep Learning for Systems**  
Invited talk at Lawrence Livermore National Laboratory
- PASC 2022 **Deep Learning for Weather Prediction and Ensemble Post-Processing**  
Invited mini-symposium talk
- LLNL 2021 **Motif Prediction with Graph Neural Networks**  
Invited talk to the graph motif group at Lawrence Livermore National Laboratory
- Huawei 2021 **Clairvoyant Prefetching for High Performance Distributed Machine Learning I/O**  
Invited talk at the Huawei Cloud InnovWave Conference
- ISC 2021 **Clairvoyant Prefetching for High Performance Distributed Machine Learning I/O**  
Invited talk at the ISC 2021 ML Systems and Tools session
- ECMWF 2020 **Deep Learning for Post-Processing Ensemble Weather Forecasts**  
At the ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction

- ESPT 2015 **Large-Scale Debugging with Graphs**  
Contributed talk
- NCSA 2013 **Towards an Improved PGDB**  
Invited talk at the National Center for Supercomputing Applications
- ANL 2013 **A GDB-based Debugger for MPI Applications**  
Invited talk at Argonne National Laboratory
- NCSA 2013 **A Parallelized GDB-based Debugger**  
Invited talk at the National Center for Supercomputing Applications
- LLNL 2012 **PGDB: A Parallel GDB**  
End-of-summer internship talk at Lawrence Livermore National Laboratory

## TEACHING

- Fall 2019–2022 **Design of Parallel and High Performance Computing** ETH Zürich  
Guest lectures on deep learning, ≈50–100 students.
- Spring 2020 **Parallel Programming** ETH Zürich  
Led exercise section, ≈30 students.
- Fall 2017 **Parallel Programming** University of Illinois at Urbana-Champaign  
Teaching assistant for CS420; managed assignments, exams, grading, several guest lectures, ≈100 students.

## ADVISING & MENTORING

Advised and mentored ten bachelors and ten masters students at ETH Zürich, plus interns at Lawrence Livermore National Laboratory. Published ten papers with bachelors or masters students as co-authors.

### ETH Zürich Bachelors Students

- 2020 Christoph Amevor, *Partitioned: A Method for Block-Sparsification of Fully Connected Layers in Neural Networks*
- 2020 Roman Böhringer, *Clairvoyant Prefetching for Machine Learning I/O*
- 2020 Neville Walo, *cuCOSMA: Near Optimal Matrix-Matrix Multiplication in CUDA C++*
- 2020 Tobia Clagliùna, *Exploring Reduced Data Exchange in Domain and Pipeline Parallel CNNs*
- 2021 Jinfan Chen, *Large Scale Distributed Deep Learning via Fine-grained Tensor Partition*
- 2021 Anton Schäfer, *Machine Learning for Graph Partitioning*
- 2021 Maximilian Fries, *A General Reinforcement Learning Solver for Combinatorial Optimization Problems on Graphs*
- 2021 Julian Schenkel, *Improving the Generalization of Recurrent Neural Networks on Compiler Dataflow Tasks*
- 2021 Stefan Scholbe, *Extending DaCe to Support the Scalable Vector Extension*
- 2022 Simon Jacob, *Data-Centric GPU Matrix-Matrix Multiplication: Implementing cuCOSMA in DaCe*

### ETH Zürich Masters Students

- 2020 Andreas Zingg, *Quantized Allreduces for Distributed Deep Learning Training*
- 2020 Peter Tatkowski, *Using (Sparse) Transformers for Weather Prediction*
- 2021 Ali Nasser (at KAUST, remotely advised, with David Keyes), *An Empirical Study of the Distributed Ellipsoidal Trust Region Method for Large Batch Training*
- 2021 Lukas Ernst, *Structured Attention Transformers on Weather Prediction*
- 2021 Bernhard Walser, *On the Similarities between Vision Transformers and Residual Networks for Weather Prediction*
- 2021 George Mtui, *Pretraining Residual Networks and Transformers for Weather Prediction with PDE Generated Data*
- 2021 Siméone de Fremond de la Merveillère, *Exploiting Structured Sparsity in Transformers through Data Centric Optimizations*
- 2022 Cliff Hodel (MSc semester project), *Scheduling N-dimensional Reductions on GPUs*
- 2022 Julia Bazińska (MSc semester project), *Accelerating Graph Neural Networks using DaCeML*
- 2023 Julia Bazińska, *To Cache or not to Cache: Rethinking Computation Schemes in Graph Neural Networks*
- 2023 Akanksha Baranwal, *Optimizing CUDA Convnets*

### Lawrence Livermore National Laboratory Interns

- 2023 Senhe Hao, *Scalable Deep Learning Benchmarks for Emerging HPC Systems*

## SERVICE

- 2024 **Cluster, Cloud and Internet Computing (CCGrid)**  
Diversity & inclusion co-chair
- 2024 **IEEE International Parallel and Distributed Processing Symposium (IPDPS)**  
Program committee member for algorithms and computational science
- 2024 **International Conference on Learning Representations (ICLR)**  
Reviewer
- 2024 **SIAM Conference on Parallel Processing for Scientific Computing (SIAM-PP)**  
Program committee member
- 2020–2024 **Supercomputing**  
Program committee member for machine learning and high-performance computing; session chair
- 2022–2024 **ISC High Performance**  
Program committee member for machine learning, AI, & quantum computing; session chair
- 2022–2023 **International Conference on Machine Learning (ICML)**  
Reviewer
- 2022–2023 **Neural Information Processing Systems (NeurIPS)**  
Reviewer
- 2022–2023 **Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks track**  
Reviewer
- 2023 **Cluster, Cloud and Internet Computing (CCGrid)**  
Program committee member for ML for systems and systems for ML
- 2020–present **IEEE Transactions on Parallel and Distributed Systems (TPDS)**  
Reviewer
- 2023–present **The International Journal of High Performance Computing Applications (IJHPCA)**  
Reviewer
- 2023–present **IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)**  
Reviewer
- 2021 **IEEE TPDS Special Section on Parallel and Distributed Computing Techniques for AI, ML, and DL**  
Program committee member
- 2019–2020 **Deep Learning on Supercomputers (@ Supercomputing)**  
Program committee member
- 2020 **Workshop on Benchmarking Machine Learning Workloads on Emerging Hardware (@ Supercomputing)**  
Program committee member
- 2019 **Principles and Practice of Parallel Programming (PPoPP)**  
Artifact evaluation committee
- 2018 **Workshop on Machine Learning in HPC Environments (@ Supercomputing)**  
Program committee member

## NOTABLE PROJECTS

- 2019–present **DaCeML** [[github.com/spcl/daceml](https://github.com/spcl/daceml) & [github.com/spcl/substation](https://github.com/spcl/substation)]  
Co-led research and development of a deep learning compiler based on data movement analysis. As part of this we created the fastest single-GPU implementation for training BERT.
- 2019–2022 **Deep Weather / MAELSTROM** [[github.com/spcl/deep-weather](https://github.com/spcl/deep-weather) & [github.com/spcl/ens10](https://github.com/spcl/ens10)]  
Co-led research in deep learning for weather and ETH's role in the EuroHPC MAELSTROM project. Set new state-of-the-arts for neural ensemble post-processing and weather prediction, and released the ENS-10 dataset.
- 2017–present **Aluminum** [[github.com/LLNL/Aluminum](https://github.com/LLNL/Aluminum)]  
A high-performance GPU communication library. Lead developer.
- 2016–present **LBANN** [[github.com/LLNL/LBANN](https://github.com/LLNL/LBANN)]  
Research and development “across the stack” for the Livermore Big Artificial Neural Network toolkit. Major contributions to communication infra, GPU acceleration, model-parallelism, and its Python interface.