Christopher R. Collins

chris@crcollins.com
https://github.com/crcollins
https://crcollins.com

Education

Carnegie Mellon University — **Pittsburgh, PA**
Ph.D. in Theoretical/Computational Chemistry — *2014 – 2019 Expected*
Using Machine Learning and Molecular Similarity to Predict Chemical Properties

Carnegie Mellon University — **Pittsburgh, PA**
M.S. in Chemistry — *2019*

Carnegie Mellon University — **Pittsburgh, PA**
M.S. in Machine Learning — *2018*

University of North Georgia — **Dahlonega, GA**
B.S. in Chemistry with Chemical Physics Concentration — *2010 – 2014*

Research

Carnegie Mellon University — **Pittsburgh, PA**
Graduate Student; Advisor: David Yaron — *2014 – Present*

- Developed parallel and distributed code for optimizing INDO parameters with SGD (50x speedup on 60 core cluster)
- Achieved state-of-the-art performance predicting polymer properties with deep learning with multi-task regression
- Invented new molecular descriptors that reduced previous prediction errors by 50% while reducing $O(n^2)$ storage costs to $O(1)$
- Implemented ranking heuristic for global minimum energy conformation search using kernel ridge regression
- Developed method for predicting atomic forces using by differentiating LSTM-based energy models
- Examined LASSO regression model for feature selection in Density Functional Theory Model
- Managed two groups of ML masters students on big data term projects (one active learning and the other multilabel/ranking classification)
- Built semi-supervised ML pipeline to map functional brain regions from fMRI data
- Built interpretable models for analyzing experimental data involving iridium catalysts

University of North Georgia — **Dahlonega, GA**
Undergraduate Researcher; Advisor: Aimée Tomlinson — *2011 – 2014*

- Researched Benzobisazole Cruciform structures using DFT for use as efficient organic solar cells
- Built set of tools to automate research process (reduced hours of work to minutes)
- Formulated SVM model for predicting Benzobisazole properties (~300,000x faster with comparable accuracy to similar methods)

Carnegie Mellon University — **Pittsburgh, PA**
Undergraduate Researcher; Advisor: David Yaron — *Summer 2013*

- Researched decomposition of Thiaphospholes using quantum computational methods
- Wrote MATLAB code to automate running Gaussian/AMPAC calculations and parsing results
- Wrote code to calculate molecular overlaps and generate molecular orbital diagrams
- Created routines to draw structures with molecular orbital overlaps indicated on each atom

Work Experience

Google, Inc. **Mountain View, CA**
Software Engineer *August 2019–*

Google, Inc. **Los Angeles, CA**
Software Engineering Intern; Host: Laura Book *Summer 2018*

- Wrote design doc to plan, scope, and compare options for a new ML pipeline for the Keyword Suggestion Service (KSS)
- Implemented and deployed 4 ML pipelines built on TFX and other tooling
- Built C++ client interface to ML models
- Instructed other teams on TFX best practices

Google, Inc. **Pittsburgh, PA**
Software Engineering Intern; Host: John Tantalo *Summer 2017*

- Implemented and deployed library for path critical data transfer format (all traffic in Google Shopping)
- Developed code generation code with multiple target languages (Python, Java, C++) for data format API
- Developed design documents for further migration to library

Tripwire, Inc. **Alpharetta, GA**
Engineering Intern *2012 – 2014*

- Rewrote entire test suite for Benchmark product using Selenium and PyTest (Python)
- Aided in migration of critical IP360 product and wrote automated test suite for API
- Maintained and expanded installation automation for CCM product
- Wrote automated script to test for CSRF and XSS vulnerabilities in IP360 web application

Technical Skills

Proficient Languages: Python 2/3, C, C++ , MATLAB, Octave, Javascript, Bash, Java, Mathematica, LaTeX, HTML5, SQL

Familiar Languages: Go, CUDA C, TI-BASIC, CSS, Maple, Haskell, Custom ASM, x86 ASM, Ruby

Software: *nix (Linux, Centos, FreeBSD, AIX, Solaris), Vim, Gaussian/AMPAC, Git, SVN, SSH, rsync, coreutils, Make, Logisim, Vagrant, PostgreSQL, Varnish, Hadoop, MapReduce, Spark, AWS, nginx, Apache, Torque, ab, LabVIEW, Pig/GuineaPig

Libraries: NumPy, SciPy, Matplotlib, Scikit-Learn, Pandas, Django, CVXPY, CVX, Keras, PyBrain, Caffe, Protobuf, Jinja2, TensorFlow, Tensorflow Extended (TFX), Theano, Selenium, Paramiko, PyTest, Bootstrap, jQuery, jQueryUI, Swig, SDL, OpenMP, OpenMPI, SCIP, Alamo

Side Projects

- **MolML** (`https://github.com/crcollins/molml`)
  - Library to extract features from molecules
    - Extracts features at the atom, molecule, fragment, kernel, or crystal level (~20-30 descriptors)
    - Parallel feature generation
    - Automated transformation saving/loading/packaging
  - Full automated testing suite (100% code coverage)
- **Chemtools/Chemtools-Webapp** (`https://gauss.crcollins.com`)
  - Implemented a set of tools for use in Benzobisazole research
    - Gaussian log output parser to collect molecular properties
    - Machine learning model to predict polymer properties from oligomers
    - Molecule graph detector/classifier
    - Benzobisazole structure generator and machine learning predictor for optoelectronic properties

        * Supercomputer job submitter and curator
    - Full automated testing suite (over 95% code coverage)
    - Python/Django Application with a Bootstrap front end
    - Vagrant VM build system

- **pyOS**
    - Implementation of a *nix like operating system using Python
    - Created shell similar to Bash Shell
    - Reimplemented standard Unix programs (cp, mv, rm, ls, etc)
    - Included permissions, pipes, multiple users, and a virtual file system

- **Dragonfly** (`https://github.com/dragonfly/dragonfly/`)
    - Library to do scalable Bayesian optimization
    - Made pip installable
    - Restructured test code to be usable
    - Added continuous integration support for multiple Python versions and platforms

- **Other**
    - 8 bit CPU (Logisim)
    - Mandelbrot/Buddhabrot Generator (Python, Javascript, C/C++ , CUDA C, TI-BASIC)
    - Various Quadtree Visualizations (Python, C/C++ , Javascript)
    - Cellular Automata Based Electronic Simulation (Javascript)
    - Object Relation Management Library (Python/SQL)
    - University Course Scraper (Python)
    - Cryptography Library (Python, Logisim)
    - Automated Peer Review System Web Application (Python/Django)
    - Virtual Cluster and Server Architecture Vagrant Builds
    - Resident-Hospital Matching Program (Python, SCIP)
    - Malloc (C)
    - Terminal Text Editor (C)
    - LC-3 Virtual Machine (C)

Courses

- (UNG) Linear Algebra
- (UNG) Intro to MATLAB
- (UNG) Intro to Cryptography
- (UNG) Differential Equations
- (UNG) Computational Methods in Physics
- (Udacity) Intro to Parallel Programming
- (CMU) Convex Optimization
- (CMU) Machine Learning
- (CMU) Statistical Machine Learning
- (CMU) Graduate Artificial Intelligence
- (CMU) Deep Learning
- (CMU) Machine Learning with Large Datasets
- (CMU) Deep Reinforcement Learning and Control
- (CMU) Data Analysis
- (CMU) Intermediate Statistics
- (CMU) Quantum Chemistry
- (CMU) Special Topics in Computational Quantum Chemistry
- (CMU) Computational Chemistry

Publications

- Kandasamy, K.; Vysyaraju, K. R.; Neiswanger, W.; Paria, B.; **Collins, C. R.** ; Schneider, J.; Póczos, B.; Xing, E. Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly. *Submitted JMLR.* **2019**. `https://arxiv.org/abs/1903.06694`

- Li, H.; **Collins, C. R.** ; Tanha, M.; Gordon, G. J.; Yaron, D. J. A Density Functional Tight Binding Layer for Deep Learning of Chemical Hamiltonians. *J. Chem. Theory Comput.* **2018**. `https://arxiv.org/abs/1808.04526`

- Li, H.; **Collins, C. R.** ; Ribelli, T. G.; Matyjaszewski, K.; Gordon, G. J.; Kowalewski, T.; Yaron, D. J. Tuning the Molecular Weight Distribution from Atom Transfer Radical Polymerization Using Deep Reinforcement Learning. *Mol. Syst. Des. Eng.* **2018**. `https://arxiv.org/abs/1712.04516`
- **Collins, C. R.** ; Gordon, G. J.; von Lilienfeld, O. A.; Yaron, D. J. Constant Size Descriptors for Accurate Machine Learning Models of Molecular Properties. *J. Chem. Phys.* **2018**. `https://arxiv.org/abs/1701.06649`.
- **Collins, C. R.** ; Yaron, D. J. Developing Coarse-Grained Site Models for Excited Electronic States of Conjugated Polymers. *SPIE* **2015**.
- Qiu, Y.; Worch, J. C.; Chirdon, D. N.; Kaur, A.; Maurer, A. B.; Amsterdam, S.; **Collins, C. R.** ; Pintauer, T.; Yaron, D.; Bernhard, S.; Noonan, K. J. T. Tuning Thiophene with Phosphorus: Synthesis and Electronic Properties of Benzobisthiaphospholes. *Chem. Eur. J.* **2014**.
- Tlach, B. C.; Tomlinson, A. L.; Morgan, K. D.; **Collins, C. R.** ; Jeffries-EL, M. Evaluation of the Impact of Extended Conjugation on the Optoelectronic Properties Benzo[1,2-d:4,5-d']bisoxazole Polymers. *Aust. J. Chem.* **2014**.

Presentations and Posters

Building a Scalable Machine Learning Pipeline for fMRI Data *CMU 10-7185 Data Analysis Term Project*	**Pittsburgh, PA** *2018*
Using Active Learning in Quantum Chemistry to Reduce Experimental Costs *CMU 10-805 Machine Learning with Large Datasets Term Project*	**Pittsburgh, PA** *2017*
A Data-driven Approach for Selecting Optimal Quantum Chemical Methods *CMU 10-805 Machine Learning with Large Datasets Term Project*	**Pittsburgh, PA** *2017*
Using Generative Adversarial Networks to Estimate Uncertainty in Quantum Chemical Methods by Exploiting Similarity in Chemical Compound Space *CMU Original Proposal*	**Pittsburgh, PA** *2017*
Dueling Recurrent Network for Partially Observable Markov Decision Process *CMU 10-703 Deep Reinforcement Learning and Control Term Project*	**Pittsburgh, PA** *2017*
Accelerating the (Augmented) Roothaan-Hall Method in Solving the Density Functional Theory Problem *CMU 10-725 Convex Optimization Term Project*	**Pittsburgh, PA** *2016*
Constant Size Molecular Descriptors For Use With Machine Learning *Midwest Theoretical Chemistry Conference*	**Pittsburgh, PA** *2016*
Constant Size Molecular Descriptors For Use With Machine Learning *CECAM Chemical Space Workshop*	**Zürich, Switzerland** *2016*
Using Data to Accelerate Quantum Chemical Calculations by Getting Better at Guessing *CMU 15-780 Graduate Artificial Intelligence Term Project*	**Pittsburgh, PA** *2016*
Using Machine Learning and Molecular Similarity to Predict Chemical Properties *CMU Progress Report*	**Pittsburgh, PA** *2016*
Using Machine Learning and Molecular Similarity to Predict Chemical Properties *CMU Chemistry Department Poster Session*	**Pittsburgh, PA** *2016*
Predicting Chemical Properties Using Machine Learning Methods *CMU Chemistry Department Retreat Poster Session*	**Farmington, PA** *2015*

The Influence of Chemical Representations on the Efficiency of Molecular Screening — **Pittsburgh, PA**
CMU Graduate Seminar — *2015*

A Mission to Multivariate Adaptive Regression Splines (MARS) and Its Other-worldly Neighbors — **Pittsburgh, PA**
CMU 10-702 Statistical Machine Learning Term Project — *2015*

Predicting Chemical Properties Using Machine Learning Methods — **Pittsburgh, PA**
CMU 10-701 Machine Learning Term Project — *2014*

Application of Machine Learning to Predict the Optoelectronic Properties of Benzobisazoles — **Dahlonega, GA**
UNG Senior Seminar — *2014*

Machine Learning in Chemical Compound Space — **Dahlonega, GA**
UNG Junior Seminar — *2013*

Donor-Acceptor Behavior of Benzobisazole Cruciformic Polymers — **Atlanta, GA**
American Chemical Society Southeastern Regional Meeting — *2013*

The Impact of Conjugation Length on Benzobisazole Cruciforms — **Raleigh, NC**
American Chemical Society Southeastern Regional Meeting — *2012*

  - Runner-up Undergraduate Poster in Organic Chemistry

Teaching

- Teaching Assistant
  - 09-221 Laboratory I: Introduction to Chemical Analysis
  - 09-214 Physical Chemistry
  - 09-101 Introduction to Experimental Chemistry (2 times)
  - 09-103 Atoms, Molecules, and Chemical Change
  - 09-231 Mathematical Methods for Chemists
  - 09-106 Modern Chemistry II (2 times)
  - 09-107 Honors Chemistry
- Weekly Lecture Series in Theory Suite (Organizer)
  - Intro to Python
  - Intro to Bash Scripting (2 lectures)
  - Intro to Object Oriented Programming
  - Intro to NumPy (2 lectures)

Awards

- State Level Technology Literacy Challenge Winner
- SERMACS Runner-up Undergraduate Poster in Organic Chemistry
- 2nd Place Spam Classifier in CMU Machine Learning out of 100.
- CMU Chemistry Department TA Award

Service

- CMU Graduate Student Assembly Representative
- Mellon College of Science Graduate Student Advisory Committee
- Department of Chemistry Social Committee
- Department of Chemistry Student Ambassador
- MellonFit Treasurer and Authorized Signer
- CMU President's Advisory Committee