


CHRISTOPHER R. COLLINS

chris@crcollins.com

 <https://github.com/crcollins>

<https://crcollins.com>

EDUCATION

Carnegie Mellon University

Ph.D. in Theoretical/Computational Chemistry

Thesis Topic: Using Machine Learning and Molecular Similarity to Predict Chemical Properties

Pittsburgh, PA

2014 – 2019 Expected

Carnegie Mellon University

M.S. in Machine Learning

Pittsburgh, PA

2019 Expected

University of North Georgia

B.S. in Chemistry with Chemical Physics Concentration

Dahlonega, GA

2010 – 2014

RESEARCH

Carnegie Mellon University

Graduate Student; Advisor: David Yaron

Pittsburgh, PA

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)

University of North Georgia

Undergraduate Researcher; Advisor: Aimée Tomlinson

Dahlonega, GA

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 (orders of magnitude faster than comparable methods for the same accuracy)

Carnegie Mellon University

Undergraduate Researcher; Advisor: David Yaron

Pittsburgh, PA

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.

Software Engineering Intern; Host: Laura Book

Los Angeles, CA

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.

Software Engineering Intern; Host: John Tantalo

Pittsburgh, PA

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.

Engineering Intern

Alpharetta, GA

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, C, C++ , MATLAB, Octave, Javascript, Bash, Java, Mathematica, L^AT_EX, 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

SIDE PROJECTS

- **MolML** (<https://github.com/crcollins/molml>)
 - Library to extract features from molecules
 - * Extracts features at the atom, molecule, 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

- **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)

COURSES

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

PUBLICATIONS

- 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

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 <i>CMU 10-805 Machine Learning with Large Datasets Term Project</i>	Pittsburgh, PA 2017
Using Generative Adversarial Networks to Estimate Uncertainty in Quantum Chemical Methods by Exploiting Similarity in Chemical Compound Space <i>CMU Original Proposal</i>	Pittsburgh, PA 2017
Dueling Recurrent Network for Partially Observable Markov Decision Process <i>CMU 10-703 Deep Reinforcement Learning and Control Term Project</i>	Pittsburgh, PA 2017
Accelerating the (Augmented) Roothaan-Hall Method in Solving the Density Functional Theory Problem <i>CMU 10-725 Convex Optimization Term Project</i>	Pittsburgh, PA 2016
Constant Size Molecular Descriptors For Use With Machine Learning <i>Midwest Theoretical Chemistry Conference</i>	Pittsburgh, PA 2016
Constant Size Molecular Descriptors For Use With Machine Learning <i>CECAM Chemical Space Workshop</i>	Zürich, Switzerland 2016
Using Data to Accelerate Quantum Chemical Calculations by Getting Better at Guessing <i>CMU 15-780 Graduate Artificial Intelligence Term Project</i>	Pittsburgh, PA 2016
Using Machine Learning and Molecular Similarity to Predict Chemical Properties <i>CMU Progress Report</i>	Pittsburgh, PA 2016
Using Machine Learning and Molecular Similarity to Predict Chemical Properties <i>CMU Chemistry Department Poster Session</i>	Pittsburgh, PA 2016
Predicting Chemical Properties Using Machine Learning Methods <i>CMU Chemistry Department Retreat Poster Session</i>	Farmington, PA 2015
The Influence of Chemical Representations on the Efficiency of Molecular Screening <i>CMU Graduate Seminar</i>	Pittsburgh, PA 2015
A Mission to Multivariate Adaptive Regression Splines (MARS) and Its Other-worldly Neighbors <i>CMU 10-702 Statistical Machine Learning Term Project</i>	Pittsburgh, PA 2015
Predicting Chemical Properties Using Machine Learning Methods <i>CMU 10-701 Machine Learning Term Project</i>	Pittsburgh, PA 2014
Application of Machine Learning to Predict the Optoelectronic Properties of Benzobisazoles <i>UNG Senior Seminar</i>	Dahlonega, GA 2014
Machine Learning in Chemical Compound Space <i>UNG Junior Seminar</i>	Dahlonega, GA 2013
Donor-Acceptor Behavior of Benzobisazole Cruciformic Polymers <i>American Chemical Society Southeastern Regional Meeting</i>	Atlanta, GA 2013
The Impact of Conjugation Length on Benzobisazole Cruciforms <i>American Chemical Society Southeastern Regional Meeting</i>	Raleigh, NC 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