

Lixue Cheng

[Email](#) [Homepage](#) [GitHub](#) [Google Scholar](#)

EDUCATION

California Institute of Technology

Pasadena, CA

Ph.D. in Chemistry, with Prof. Thomas F. Miller III.

June 2022

Thesis: Accurate and transferable molecular-orbital-based machine learning for molecular modeling [[Link](#)]

University of Wisconsin-Madison

Madison, WI

Bachelor of Sciences in Honors, GPA: 3.933/4.0

May 2016

Majors: Chemistry (Honor), Mathematics (Honor), Biochemistry, Molecular Biology.

Minor (Certificate): Computer Sciences

RESEARCH INTERESTS

Apply computational and mathematical tools to improve the chemistry and biology study strategies.

(1) AI for quantum chemistry

- Deep Quantum Monte Carlo (DeepQMC, Microsoft Research)
- Machine Learning Force Fields (MLFF, Microsoft Research)
- Molecular-orbital-based machine learning (MOB-ML, Miller group, [Theme twitter thread](#))
- TransOrb (In preparation, with Goddard group)

(2) Quantum computing algorithms

- Quantum computing for quantum chemistry (Independent)
- Machine Learning for variational quantum algorithms (Independent)

(3) AI for biology

- Bayesian optimization for experimental design (ODBO, Tencent Quantum Lab, [Repo](#))

RESEARCH AND TEACHING EXPERIENCES

Microsoft Research AI4Science Lab (Berlin)

June 2023–Present

Visiting Researcher

AI for quantum chemistry: DeepQMC and MLFF

- Apply Transformer and other deep learning techniques to develop deep quantum Monte Carlo (DeepQMC) and machine learning force field (MLFF)
- Collect and generate highly accurate quantum simulation data as input data for deep learning model

LLMs for Science: ability of GPT-4 in chemistry

- Assess the ability of GPT-4 in various scientific tasks, such as understanding theoretical chemistry, generating useful simulation inputs, and designing reasonable research plan, with different prompt engineering techniques.
- Participate in the preparation of foundation model for molecular sciences.

Tencent Quantum Lab, Tencent

August 2022–May 2023

Research scientist

Machine learning assisted quantum computing algorithm design for quantum chemistry

- Develop AI tools for quantum computing, quantum chemistry and biology
- Develop, implement and test novel Bayesian learning algorithm to optimize quantum computing algorithms running on real quantum devices.
- Design shallow and efficient quantum circuits for VQE to achieve chemical accuracy for molecular systems.

Miller Group, California Institute of Technology

August 2017–Mar 2022

PhD student

MOB-ML: Molecular Orbital-Based Machine Learning

- Use the molecular orbital representations to predict the molecular energies via ML tools to give high accuracy as wave function theory method but a low computational cost as density functional theory method.
- Perform electronic structure calculations and quantum simulations for different molecular systems.
- Develop different supervised and unsupervised learning methodologies to model different molecular properties.
- Develop a ML module in open-source quantum chemistry software (*entos*) and build a MOB-ML database.

TranOrb: Transferable Electronic Structure Learning via General Orbital Information (migrated to Goddard group)

INSPIRE: Computational Parameterization of Nucleic Acid Secondary Structure Models

- Combine ML tools with molecular dynamics (MD) to derive new thermodynamics and kinetics nucleic acid models with secondary structure information.
- Develop and revise ML algorithms that can improve the efficiency of the strategy.
- Perform MD simulations for different nucleic acid reaction systems as a source of data for ML model.

Record Group, University of Wisconsin-Madison

January 2013–July 2017

Undergraduate researcher

Interpreting and predicting Hofmeister salt ion, polyol, PEG and sugar effects on biopolymer processes using solubility assay and solute-partitioning model

- Design and lead projects using solubility experiments to measure different solutes-biomolecule interactions.
- Recruit and mentor other undergraduate students to collect solute–nucleobase thermodynamics data.
- Develop mathematical models and analyze the data to interpret functional group chemical interactions.
- Determine interaction constants between different bases and different solutes and apply them to predict and interpret solute-biopolymer interaction and drug solubility data.

Math studies supervised by Qin Li, University of Wisconsin-Madison

August 2015–May 2016

Undergraduate researcher

Randomized Singular Value Decomposition algorithm for low rank matrices & its numerical applications

- Reprove and improve the theorems in the literature as a mathematical exercise.
- Develop MATLAB codes to implement the algorithm studied in the literature and perform numerical tests.
- Compare and analyze the numerical test results and explain the mathematical origins in different cases.

Grader/TA for Chem/Biochem 565/665 Biophysical Chemistry

University of Wisconsin-Madison

2015 Fall & 2016 Fall

Grader/TA for Math 341 Linear Algebra (Honor) and Math 561 Differential Geometry

University of Wisconsin-Madison

2016 Spring

VOLUNTEERING, TEACHING & COMMUNITY EXPERIENCES

I'm an active volunteer & organizer for "[AI for science](#)" research community to bridge scientists from different backgrounds, a practitioner for "AI for science" education to cultivate the next generation talents, and a part-time open-source software developer to break the knowledge barrier in research.

AI for Science community & workshops

May 2023–Present

Core member & Organizer

- Organize the AI for science workshops in top conferences
- Start the AI for Edu and AI for science paper collection initiatives
- Participate in the proposal and review report writing

AI for Science education for high school students

Jan 2021–Present

Instructor

- Introduce cutting-edge AI for science contents to high school students all over the world
[\[Link\]](#)[\[An example course page\]](#)

AI for Science paper collection (over 40 stars) [\[GitHub\]](#)

June 2024–Present

Organizer & Core maintainer

MathTranslate (over 1000 stars) [\[GitHub\]](#)[\[YouTube\]](#)[\[Bilibili\]](#)

April 2023–Present

Core developer

- A core developer for this free open-source software to provide high quality translation of scientific papers with heavy math symbols.
- Maintain social media & user discussion groups.

Department of Mathematics, University of Wisconsin–Madison

Administrative Office and Front Desk Student Help

August 2015–May 2016

Sonoco Products Co., 455 Science Drive, Madison, WI

Research Project Assistant

September 2014–January 2015

- Measure geometries and properties of different materials and perform mathematical modeling of products.
- Analyzing data using various software, for example, Excel, Auto CAD, FEA.
- Develop test devices and search the literature to support our own research projects.

Department of Chemistry and Mathematics, University of Wisconsin–Madison

Private Tutor

August 2014–May 2017

College Library, University of Wisconsin–Madison

Circulation Student Assistant

August 2013–May 2015

Greater University Tutoring Service (GUTS), University of Wisconsin–Madison

Academic Tutor (Chemistry and Mathematics)

September 2012–August 2013

HONORS AND AWARDS

- 2017 Caltech General Fellowship, Division of Chemistry and Chemical Engineering, Caltech
 - 2016 Mary Ellen Rudin Foundation Scholarship Award, Department of Mathematics, UW-Madison
 - 2015 David H. Durra Scholarship, College of Letters & Sciences, UW-Madison
 - 2015 Eugene and Patricia Kreger Herscher Undergraduate Scholarship for the 2015–16 Academic Year, Department of Chemistry, UW-Madison
 - 2015 Walter W. & Young-Ja C. Toy Summer Research Fellowship, Department of Chemistry, UW-Madison
 - 2014 Biochemistry Undergraduate Summer Research Scholarship, Department of Biochemistry, UW-Madison
 - 2014 Gerald W. and Tui G. Hedstrom Scholarship, College of Letters & Science, UW-Madison
 - 2014 Martha Gunhild Weeks Undergraduate Scholarship for the 2014-15 Academic Year, Department of Chemistry, UW-Madison
 - 2014 The Honor Society of Phi Kappa Phi Member, University of Wisconsin-Madison Chapter
 - 2014 Trewartha Senior Thesis Grant, Letters & Science Honors Program, UW-Madison
 - 2014 Wisconsin Hilldale Undergraduate/Faculty Research Fellowship, UW-Madison
 - 2013 John & Elizabeth Moore Awards for Excellence, Department of Chemistry, UW-Madison: Excellence for Chem109 Advanced General Chemistry, 1/350
-

PATENTS, PUBLICATIONS AND THESES

Post-Caltech work

- **Cheng, L.***; Shen, J.*; Helie, J.; Sun, L. Robust Long-range Interaction Modelling in Molecular Graphs via Equivariant Ewald-based Message Passing (EEMP), *In preparation*. (*co-first author)
- Jacobson, G.; **Cheng, L.**; Sun, J.; Bhethanabotla, V.; McCoy, A. B., *In preparation*.
- **Cheng, L.**; Szabó, P.B.; Schätzle, Z.; Kooi, D.; Köhler, J.; Noé, F.; Gori-Giorgi, P.; Foster, A. Highly Accurate Real-space Electron Densities with Neural Networks, arXiv:2409.01306. [[Link](#)]
- Sun, J.; **Cheng, L.**; Zhang, S.X. Stabilizer ground state: Theory, algorithms and applications. arXiv: 2403.08441 (2024). [[Link](#)][[Repo](#)]
- **Cheng, L.***; Chen, Y.Q.*; Zhang, S.X.; Zhang, S. Quantum approximate optimization via learning-based adaptive optimization. *Commun. Phys.*, **2024** (*co-first author) [[Link](#)] [[Repo](#)]
- Sun, J.; **Cheng, L.**; Li, W. Towards chemical accuracy with shallow quantum circuits: A Clifford-based Hamiltonian engineering approach. *J. Chem. Theory Comput.*, **2024**. [[Link](#)][[Repo](#)]
- MR AI4Science*, MA Quantum. The Impact of Large Language Models on Scientific Discovery: a Preliminary Study using GPT-4. arXiv:2311.07361 (2023). [[Link](#)] (***Major contributor to Chap 4**)

- Li, W.; Allcock, J.; **Cheng, L.**; Zhang, S.X.; Chen, Y.Q.; Mailoa, J.P.; Zhang, S. TenCirChem: An efficient quantum computational chemistry package for the NISQ era. *J. Chem. Theory Comput.*, **2023**. [[Link](#)][[Repo](#)]
- **Cheng, L.**; Yang, Z.; Liao, B.; Hsieh, C.; Zhang, S. ODBO: Bayesian Optimization with prescreening for directed protein evolution. arXiv:2205.09548 (2022). [[Link](#)][[Repo](#)]

Caltech work

- Patent: Miller III, T.F.; Welborn, M.; **Cheng, L.**; Husch, T.; Song, J.; Kovachiki, N.; Burov, D.; Teh, Y.S.; Anandkumar, A.; Ding, F.; Lee, S.J.R.; Qiao, Z.; Lale, A.S. Systems and methods for determining molecular structures with molecular-orbital-based features. U.S. Patent 16817489, 2020 [[Link](#)]
- **Cheng, L.**; Sun, J.; Deustua, J. E.; Bhethanabotla, V.C.; Miller III, T.F. Molecular-orbital-based machine learning for open-shell and multi-reference systems with kernel addition Gaussian process regression. *J. Chem. Phys.*, **2022**. [[Link](#)]
- Sun, J.; **Cheng, L.**; Miller III, T. F. Molecular dipole moment learning via rotationally equivariant Gaussian process regression with derivatives in molecular-orbital-based machine learning. *J. Chem. Phys.* **2022**. [[Link](#)]
- **Cheng, L.**; Sun, J.; Miller III, T.F. Accurate molecular-orbital-based machine learning energies via unsupervised clustering of chemical space. *J. Chem. Theory Comput.*, **2022**. [[Link](#)]
- Sun, J.; **Cheng, L.**; Miller III, T.F. Molecular energy learning using alternative blackbox matrix-matrix multiplication algorithm for exact Gaussian process. arXiv: 2109.09817 (2022). Appears in NeurIPS 2021 AI for Science Workshop [[Link](#)].
- Lu, F.*; **Cheng, L.***; DiRisio, R. J.*; Finney, J. M.; Boyer, M. A.; Sun, J.; Lee, S. J. R.; Deustua, J. E.; Miller III, T. F.; McCoy, A. B. Fast near *ab initio* potential energy surfaces using machine learning. *J. Phys. Chem. A*, **2022**. (*co-first author) [[Link](#)]
- Husch, T.; Sun, J.; **Cheng, L.**; Lee, S. J. R.; Miller III, T.F. Improved accuracy and transferability of molecular- orbital-based machine learning: Organics, transition-metal complexes, non-covalent interactions, and transition states. *J. Chem. Phys.*, **2021**. [[Link](#)]
- **Cheng, L.**; Kovachki, N.; Welborn, M.; Miller III, T. F. Regression clustering for improved accuracy and training costs with molecular-orbital-based machine learning. *J. Chem. Theory Comput.*, **2019**. [[Link](#)]
- **Cheng, L.**; Welborn, M.; Miller III, T. F. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules. *J. Chem. Phys.*, **2019**. [[Link](#)]
- Welborn, M.; **Cheng, L.**; Miller III, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **2018**. [[Link](#)] (Highlighted with commentary in [C&EN](#) and [Caltech News](#))

UW-Madison work

- Knowles, D. B.; Shkel, I. A.; Phan, N. M.; Sternke, M.; Lingeman, E.; Cheng, X.; **Cheng, L.**; O'Connor, K.; Record, M. T. Chemical interactions of polyethylene glycols (PEGs) and glycerol with protein functional groups: Applications to effects of PEG and glycerol on protein processes. *Biochemistry* **2015**, 54 (22), 3528–3542. [[Link](#)]
- Undergraduate Senior Honor Thesis for Chemistry with Prof. M. Thomas Record: *Interactions of polyols with aromatics and nucleobases: Large opposing contributions from interactions of nonpolar and polar groups*
- Undergraduate Senior Honor Thesis for Mathematics with Prof. Qin Li: *Randomized singular value decomposition (RSVD) algorithm for low rank matrices and its numerical applications*