Jason Liang

Email: yhljason49 -at- gmail.com
New York, NY
yhliang.com

Scientific software researcher building Python workflows and validated datasets from close to 1,000 quantum-chemistry calculations; developed periodic MP2 that reduced formal scaling from $O(N^5)$ to $O(N^3)$ ($\sim$90% lower cost).

Education

- **Columbia University** | *Ph.D. Candidate, Chemical Physics* | *Gu Family & Dean's Fellowship ($100k)* *Expected Aug. 2026*
- **University of California, Berkeley** | *B.S. Chemistry, Honors (3.86/4.0, Top 7%); Dean's List* *May 2021*

Research and Software Experience

- **Columbia Center for Computational Electrochemistry** | *Graduate Researcher* *Aug. 2021 - Present*
  - Built object-oriented Python workflows on HPC clusters that automated, organized, and validated close to 1,000 DFT, CCSD, MP2, and quantum-embedding calculations across molecular and materials systems
  - Built validation and provenance tooling for a battery-modeling dataset; traced a 2×-inflated diffusion barrier to a symmetry-broken reference whose error cancelled on small clusters but not at scale, and rebuilt the analysis from a corrected data source
  - Developed domain-based local MP2 for periodic systems in PySCF, reducing first-principles scaling from $O(N^5)$ to $O(N^3)$ ($\sim$90% lower cost); wrote Cython kernels calling BLAS `dgemm` directly to accelerate the residual equations
  - Independently developed a machine-learning project on features from self-written local-correlation code; trained Gaussian-process, scikit-learn, and PyTorch neural-network models (including a physics-constrained architecture), evaluating against pre-registered criteria and held-out transfer sets
- **Martin Head-Gordon Group, UC Berkeley** | *Undergraduate Researcher* *Apr. 2020 - May 2021*
  - Benchmarked 47 computational chemistry methods across 132 molecular systems; created a systematic decision framework for method selection, published in a peer-reviewed journal
- **Sumitomo Electric Industries** | *Materials Science Intern* *May - Aug. 2018*
  - Performed root-cause failure analysis (electron microscopy, spectroscopy) of materials defects in products shipped to Toyota and Toshiba; presented findings to senior R&D leadership in a Japanese-language work environment

Open-Source & Selected Projects

- **PySCF-Based Method Development** | *Research software* *2021 - Present*
  - Built 4 modules on the PySCF framework: periodic double hybrid (kdh), periodic local MP2 (lpno), DLPNO-RPA, and DLPNO-MP2; packaged and installable
  - Wrote regression and validation suites for these modules (several hundred tests), checking method output against published literature values, provenance metadata, and analytic limits such as exact energy gradients
- **Published-Research Code & Data** | *Reproducibility* *2021 - Present*
  - Released the geometries, basis sets, data, and analysis pipelines behind two publications: x23_mp2 (*JPCL* 2023) and research_su (*JCP* 2021)
- **pymatgen (Materials Project)** | *Developer, TEM diffraction calculator* *2019 - 2020*
  - Contributed the TEM diffraction calculator module to pymatgen, the Materials Project's open-source materials-analysis library (object-oriented Python); listed project developer (pymatgen.org/team)

Publications

Three first-author peer-reviewed publications in *J. Chem. Theory Comput.*, *J. Phys. Chem. Lett.*, and *J. Chem. Phys.* (2021 to 2025); manuscripts in preparation on periodic local MP2 and lithium-ion battery modeling.

Technical Skills

Programming:	Python (5+ years; object-oriented programming; NumPy, SciPy, scikit-learn, GPy, PyTorch), Cython, Bash, Git/GitHub, pytest
Data/ML:	Scientific data pipelines, metadata and provenance, validation diagnostics, regression testing, feature engineering, Gaussian-process regression, neural networks
Computing:	HPC clusters, automated batch workflows across $\sim$1,000 calculations
Software:	PySCF, Quantum ESPRESSO, ORCA, pymatgen, LaTeX
Methods:	First-principles/DFT (finite and periodic), MP2, random-phase approximation, coupled cluster, local correlation, quantum embedding
Talks:	PySCF Developer Conference; academic symposia
Languages:	English (Native), Mandarin (Native), German (Elementary)