Saurabh Sivakumar

saurabh.sivakumar97@gmail.com | LinkedIn | Github | Website | Google Scholar | +1 412 636 6213

Summary: Computational chemist with 6+ years of experience developing predictive models to uncover atomistic structure–property relationships in catalytic and materials systems. Designs scalable simulation workflows and machine learning tools to accelerate materials design and enable data-driven decision making, partnering with cross-functional experimental teams to support product development.

Education

University of California — Davis, CA
Ph.D. *in Chemical Engineering (Advisor: Ambarish Kulkarni)* — ***Sep 2021 – June 2026***

Carnegie Mellon University — Pittsburgh, PA
M.S *in Chemical Engineering (Advisor: Zachary Ulissi)* — ***Aug 2019 – Dec 2020***

National Institute of Technology — Tiruchirappalli, India
B.Tech *in Chemical Engineering, Minor in Economics* — ***Jul 2015 – May 2019***

Skills & Awards

Languages: Python, Bash, R, C++, SQL, LaTeX

Python Packages: ASE, RDKit, Scipy, Pandas, Seaborn, Pymatgen, Pynta, RMG, VASP-kit, MDanalysis, OpenMM, CuPy, Numba, Parsl, Dask, Deap

Machine Learning: PyTorch, scikit-learn, PyG, WandB, LangGraph

Technologies/Frameworks: Git, Plotly, Jupyter, OpenAcc, AI-assisted development (Claude Code, Copilot, Antigravity)

Simulation Tools: VASP, ORCA, COMSOL, QuantumEspresso, LAMMPS, PLUMED, Gaussian, Zacros, PySCF, Cantera, MATLAB, OVITO, VMD, ASPEN Plus, ANSYS

Awards: NREL Kestrel compute grant, NSF ACCESS Discover compute grants, Multiple NERSC compute grants; Total CPU/GPU hours: 6 Million

Experience

Graduate Researcher — Sep 2021 – May 2026
University of California — *Davis, CA*

Machine Learning for Catalysis and Materials Design

- Led ML-enabled simulation efforts for DOE- and industry-funded projects on catalyst deactivation, electrochemical separations, and surface reaction chemistry, while mentoring 7 students in computational methods and data-driven materials modeling.
- Engineered machine-learned interatomic potentials (MLIPs) to model adsorption, diffusion, and near-surface gas phase reactions, achieving up to 1000× speedups over DFT and enabling mechanistic insight to inform catalyst and process design decisions.
- Developed automated frameworks for transition state discovery and kinetic modeling by integrating graph- and descriptor-based ML models (MACE, DeepMD-Kit, FairChem framework, ORB, NEP, NequIP) with open-source simulation frameworks.
- Collaborated with experimental teams to model diverse catalysis and materials systems (surfaces, nanoparticles and zeolites), linking atomistic simulations with multiscale studies through microkinetic modeling, Raman spectra prediction, kinetic Monte Carlo, metadynamics, and other enhanced sampling techniques.
- Co-developed and deployed active learning pipelines for model training and data generation.

Data-Driven Screening of Transition Metal (TM) Complexes for Oxygen Separation

- Designed and deployed high-throughput DFT and ML workflows to screen transition metal complexes for O_2 binding, performing over 6M+ atomistic calculations and identifying several candidate materials for electrochemical oxygen separation.
- Leveraged structure–property relationships and thermodynamic analysis to design and optimize oxygen capture candidates, providing guidance to experimental teams.
- Exploring diffusion-based generative models and LLM-based agents to autonomously design transition-metal complexes for O_2 capture.

- Benchmarked DFT methods against high-accuracy quantum chemical calculations (CCSD(T)) and built scalable multi-fidelity workflows with ML surrogates, reducing property prediction cost and turnaround time by orders of magnitude.

Research Assistant Jan 2020 – Jul 2021
Carnegie Mellon University *Pittsburgh, PA*

Development of ML-Enhanced Frameworks for Atomistic Simulations

- Worked on surface segregation using deep reinforcement learning, focusing on dataset preparation and Bayesian hyperparameter optimization.
- Contributed to the development of open-source tools and sampling strategies to build an active learning framework leveraging physics-informed Behler-Parrinello Neural Networks to accelerate Nudged elastic band calculations and molecular dynamics. Code
- Built and validated software for global optimization of nanoclusters using a modified genetic algorithm enhanced with neural networks and Gaussian processes. Code

Research Intern May 2018 – Aug 2018
Singapore University of Technology and Design *Singapore*

- Researched lightweight photovoltaic modules by designing and evaluating solar PV systems using polycarbonate substitutes, under the guidance of *Arief S Budiman*.

Selected Publications

1. Toward an Ab-Initio Description of Adsorbate Surface Dynamics
   Saurabh Sivakumar *and Ambarish Kulkarni* [Link]
2. Computational Insights into O_2 Capture by Cobalt Salen and its Derivative
   Saurabh Sivakumar *and Ambarish Kulkarni* [Link]
3. An Automated Pynta-based Curriculum for ML-Accelerated Calculation of Transition States
   Trevor Price, ***Saurabh Sivakumar****, Matthew S. Johnson, Judit Zador and Ambarish Kulkarni* [Link]
4. Cluster-MLP: An Active-Learning Genetic-Algorithm Framework for Accelerated Discovery of Global-Minimum Configurations of Pure and Alloyed Nanoclusters
   Rajesh K. Raju, ***Saurabh Sivakumar****, Xiaoxiao Wang and Zachary W. Ulissi* [Link]
5. Enabling robust offline active learning for machine learning potentials using simple physics-based priors
   Muhammed Shuaibi, ***Saurabh Sivakumar****, Rui Qi Chen and Zachary W. Ulissi* [Link]

Selected Projects

1. Extended the All-atom **Diffusion Transformer (ADiT)** with energy-aware fine-tuning, generating chemically valid, diverse 3D conformers on QM9 with state-of-the-art accuracy while reducing computational cost versus traditional ab initio sampling. Code
2. Implemented and extended a **multi-modal transformer** (DOSTransformer) for predicting density of states in crystalline materials by adapting and enhancing existing open-source code, integrating structural and energy-level embeddings, prompt tuning, dynamic positional encoding, and benchmarking against GNN, MLP, and E(3)NN baselines. Code
3. Applied supervised and unsupervised **ML techniques (NNs, k-NN, decision trees, logistic regression, PCA, t-SNE, XGBoost)** to classification and time-series analysis, achieving ≥96% accuracy on the UC Irvine Musk dataset and uncovering trends in COVID-19 data from the New York Times.Code

Selected Conferences & Posters/Talks

MLIPs and Microkinetic Modeling for Probing Near Surface Adsorbate Kinetics Feb 2026
GRC AI for Energy, Materials and Chemistry *Galveston, TX*

ML Accelerated Design of Transition Metal Oxygen Capture Agents Dec 2025
MRS Fall Meeting 2025 *Boston, MA*

Optimizing Chebyshev Interaction Model with Parallel Programming Oct 2024
NERSC Annual User Group Meeting *Berkeley, CA*

An Active Learning Framework for Accelerating Saddle Point Searches Nov 2020
Presented at AIChE Annual Meeting 2020 *(Virtual)*