Changwen Xu

412-999-1188 | changwex@umich.edu | linkedin.com/in/changwenxu | github.com/ChangwenXu98

Education

University of Michigan — Ann Arbor, MI
Doctor of Philosophy in Mechanical Engineering — *May. 2027 (expected)*

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Computational Materials Science and Engineering, GPA: 4.0/4.0 — *Dec. 2022*

South China University of Technology — Guangzhou, China
Bachelor of Engineering in Materials Science and Engineering, GPA: 3.96/4.0 — *Jun. 2021*

Publication

1. Huang, H., Magar, R., **Xu, C.**, Barati Farimani, A., Materials Informatics Transformer: A Language Model for Interpretable Materials Properties Prediction, *arXiv preprint arXiv:2308.16259*, 2023 (under review by *Nature Machine Intelligence*)
2. Wang, Y., **Xu, C.**, Li, Z., Barati Farimani, A., Denoise Pretraining on Nonequilibrium Molecules for Accurate and Transferable Neural Potentials, *J. Chem. Theory Comput.*, 2023, 19, 15, 5077–5087
3. **Xu, C.**, Wang, Y., Barati Farimani, A., TransPolymer: a Transformer-based language model for polymer property predictions, *npj Computational Materials*, 2023, 9, 64
4. Wang, L. (#), **Xu, C.** (#), Zhang, W., Zhang, Q., Zhao, M., Zeng, C., Jiang, Q., Gu, C., Ma, Y., Electrocleavage Synthesis of Solution-Processed, Imine-Linked, Crystalline Covalent Organic Framework Thin Films, *J. Am. Chem. Soc.*, 2022, 144, 20, 8961–8968. (# These authors contributed equally.)

Research Experience

Machine Learning Summer Internship — Remote
Redesign Science — *Jun. 2023 – Aug. 2023*

Improving ML Collective Variables with Energy-based Path Construction | *Python, MD*

- Constructed **Minimum Energy Path** for better **Machine-Learned Collective Variables** for sampling rare events by metadynamics
- Trained time-lagged autoencoder to learn CVs that capture the metastability in low dimensions
- Conducted **interpolations** in CV space between start and end state and minimized the energy along the path for better **transition path**
- Ran **MetaD** simulation using the ML-CVs and compared the results to those without energy-aware paths

Graduate Student Researcher — Pittsburgh, PA
Carnegie Mellon University — *Oct. 2021 – May. 2022, Aug. 2022 – Dec. 2022*

Graduate Research Assistant — Pittsburgh, PA
Carnegie Mellon University — *May. 2022 – Aug. 2022, Jan. 2023 – Apr. 2023*

Materials Informatics Transformer for Interpretable Property Predictions | *Python, Language Model*

- Proposed **MatInFormer**, a **Transformer**-based language model for interpretable material property predictions
- Introduced a novel **pretraining** approach which combined Masked Language Modeling and **Lattice Parameter Prediction** for learning the grammar of crystallography
- **Finetuned** the model on Matbench and MOFs data, achieving comparable performance to structure-based or structure-agnostic models on most chosen datasets
- Demonstrated the **flexibility** of the framework by manipulating the input informatics tokens and analyzed the **interpretability** of prediction results by attention visualization

Denoise Pre-training for Accurate and Transferable Neural Potentials | *Python, PyTorch, Neural Potential*

- Proposed **denoise** pretraining on **non-equilibrium** molecular conformations to achieve more accurate and transferable GNN potential predictions
- Finetuned the model pretrained on small molecules to improve performance on diverse molecular systems with remarkable **transferability**

- Demonstrated **data efficiency** of pretrained GNNs for large and complex molecular systems with limited training data

Transformer-based Language Model for Polymer Property Predictions | *Python, PyTorch, Language Model*

- Proposed **TransPolymer**, a **Transformer**-based model for accurate and efficient polymer property predictions
- **Pretrained** with **Masked Language Modeling** on a large unlabeled dataset of polymer sequences for learning expressive representations
- **Finetuned** the pretrained model in different downstream prediction tasks, achieving the state-of-the-art (**SOTA**) results and surpassing baseline models by large margins in most cases
- Conducted various ablation studies on pretraining, frozen transformer encoder, and data augmentation, and identified influential parts of polymers on prediction results by **attention visualization**
- Highlighted the model as a promising computational tool for understanding chemistry in a data science view

Undergraduate Student Researcher Guangzhou, China

South China University of Technology *Sep. 2018 – Jun. 2021*

Electrocleavage Synthesis of Solution-Processed Covalent Organic Framework (COF) Thin Films

- Developed an unprecedented electrocleavage synthesis strategy to produce **imine-linked COF films** directly on electrodes from electrolyte solutions at room temperature
- Carried out **cathodic exfoliation** of COF powders into nanosheets by electrochemical reduction and protonation and **anodic oxidation** to reproduce COF structures
- Characterized the COF films, COF solutions as well as precipitated COF nanosheets to verify the method

Synthesis and Solution Processing of Triazine-based Crosslinked Conjugated Polymers (TCPs)

- Reported the method of **protonating** the polymeric skeletons to enhance the solubility of TCPs under electrostatic repulsions to achieve solution processing
- Synthesized triazine-based n-type crosslinked conjugated polymers with **perylene diimide** units and achieved **solution-processing** of crosslinked conjugated polymers for the fabrication of polymeric films
- Developed the method of electrocleavage synthesis of triazine-based COFs from previous research and compared the results of the two methods.

Projects

AI Ramanujan: Discovery of Formula Equivalence (CMU) | *PyTorch, Language Model* Oct. 2022 – Dec. 2022

- Train **Transformer** encoders to evaluate whether two mathematical formulas map to the same value
- Construct a **tokenizer trainer** for adaptation to formulas in Mathematica format and design a **translator** to convert formulas in Maple format into Mathematica format
- Adapt the encoders with **cross attention** to encode correlations between two embeddings and train the model by **contrastive learning**

JPX Tokyo Stock Exchange Prediction (CMU) | *Python, PyTorch, sklearn, EDA* May. 2022 – Jul. 2022

- Trained **LSTM** model for JPX stock prediction and compared the model against real future returns for evaluation
- Applied exploratory data analysis on JPX stock data, conducted data cleaning, and carried out **feature engineering** to construct financial descriptors
- Achieved the highest rank of top 15% in Kaggle competition "JPX Tokyo Stock Exchange Prediction"

Talks And Presentations

Poster Presentation at Molecular Machine Learning Conference Jun. 2024

- Presented the poster on *CLOUD: A Scientific Foundation Model for Crystal Property Prediction*
- The poster was chosen as the spotlight at the conference

Poster Presentation at MICDE Scientific Foundation Model Conference Apr. 2024

- Presented the poster on *CLOUD: A Scientific Foundation Model for Crystal Property Prediction*

Invited Speaker at "Math + Machine Learning + X" Seminar at Brown University Nov. 2022

- Gave a talk on *TransPolymer: a Transformer-based Language Model for Polymer Property Predictions*

Professional Service

Reviewer: NeurIPS'23-24, ICLR'24, ICML'24