Crossing the Simulation-Experiment Gap: Style Translation for Enhanced Atomic Structure Discovery from AFM Images

Jie Huang[1], Niko Oinonen[1,2], Fabio Priante[3], Lauri Kurki[1], and Adam S. Foster[1,4]

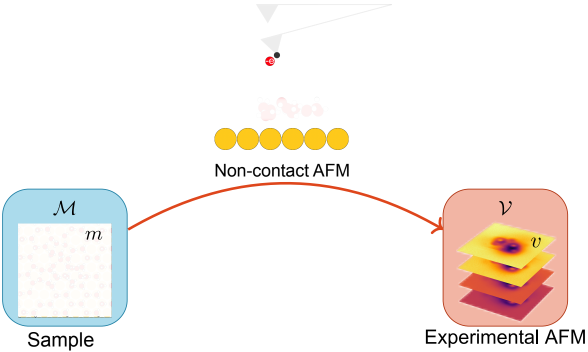
1. Surfaces and Interfaces at the Nanoscale (SIN), School of Science, Aalto University, Finland
2. Nanolayers Research Computing Ltd., United Kingdom
3. Department of Chemistry and Materials Science, Aalto University, Finland
4. WPI Nano Life Science Institute, Kanazawa University, Japan

Surfaces and Interfaces at the Nanoscale (SIN)

Aalto University School of Science

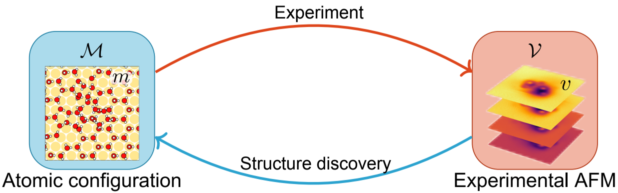
Structure discovery and its challenges

Why is atomic structure prediction so hard?



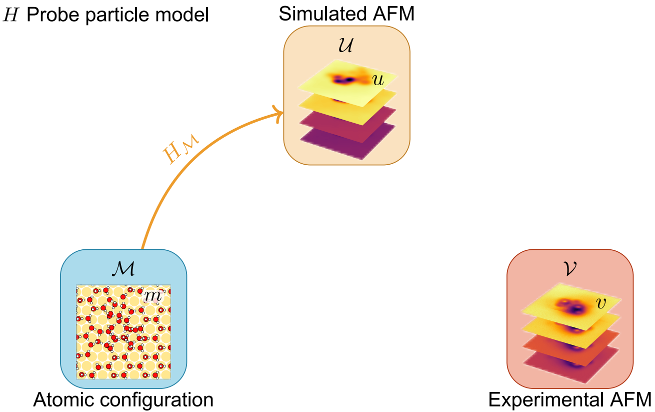
Atomic force microscopy (AFM) allows us to see atoms.

Why is atomic structure prediction so hard?



However, we don't fully understand what they show us.

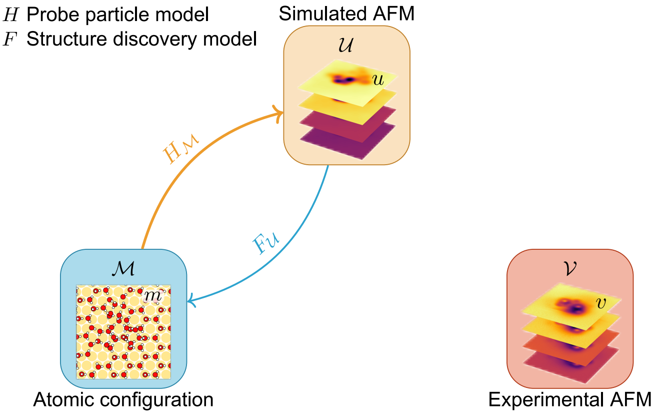
Why is atomic structure prediction so hard?



Probe particle model (PPM) [1,2,3] $H_{\mathcal{M}}$ is used to simulate AFM images.

[1] Hapala, P. et al., *Phys. Rev. B*, 2014, 90(8), 085421. DOI: 10.1103/physrevb.90.085421.
[2] Hapala, P. et al., *Phys. Rev. Lett.*, 2014, 113(22), 226101. DOI: 10.1103/physrevlett.113.226101
[3] Oinonen, N. et al., *Comput. Phys. Commun.*, 2024, 305, 109341. DOI: 10.1016/j.cpc.2024.109341

Why is atomic structure prediction so hard?

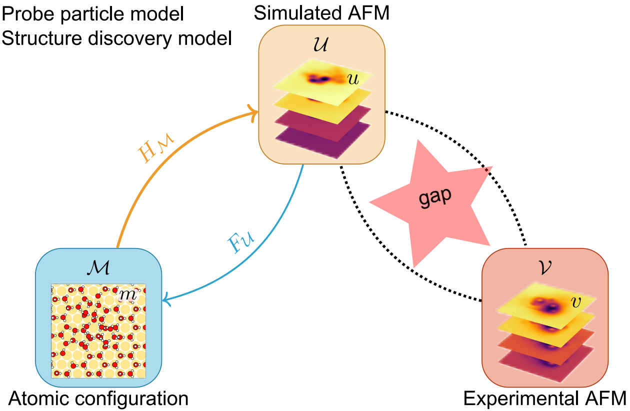


Structure discovery models [1,2,3] learn the inverse map from the simulated AFM images to the atomic structure.

[1] Oinonen, N. et al., *MRS Bulletin*, 2022, 47(9), 895-905. DOI: 10.1557/s43577-022-00324-3.
[2] Kurki, L. et al., *ACS Nano*, 2024, 18(17), 11130-11138. DOI: 10.1021/acsnano.3c12654.
[3] Priante, D. et al., *ACS Nano*, 2024, 18(1), 1234-1245. DOI: 10.1021/acsnano.3c10958

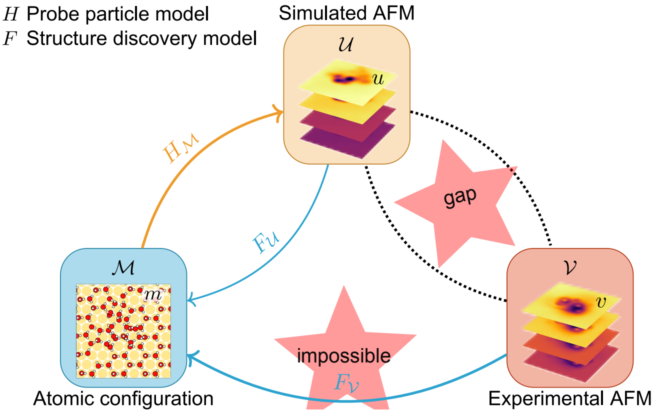
Why is atomic structure prediction so hard?

H Probe particle model
F Structure discovery model



Challenge 1: The simulated AFM images can't capture all the features of the experimental AFM images.

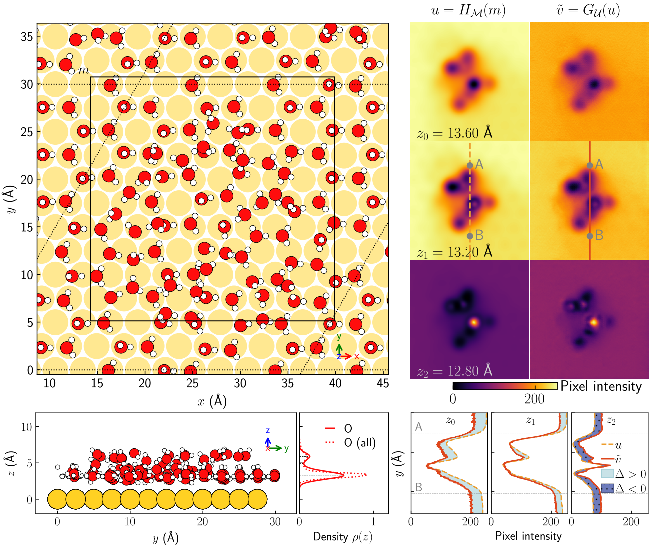
Why is atomic structure prediction so hard?



Challenge 2: It's almost impossible to train a model directly on the experimental AFM images.

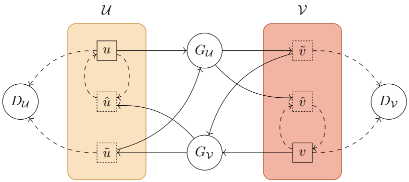
Crossing the gap

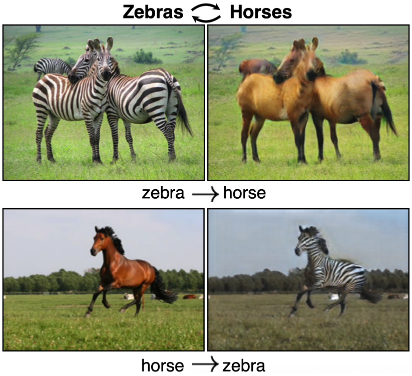
Simulated AFM must look real to train useful ML models



We add experimental features automatically to simulated AFM images.

CycleGAN enables unpaired style translation between image domains

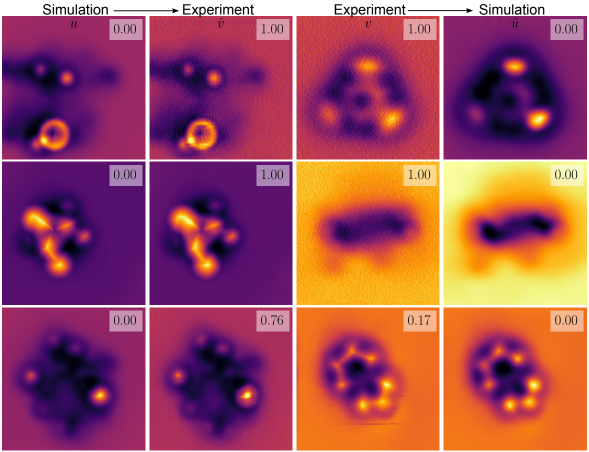




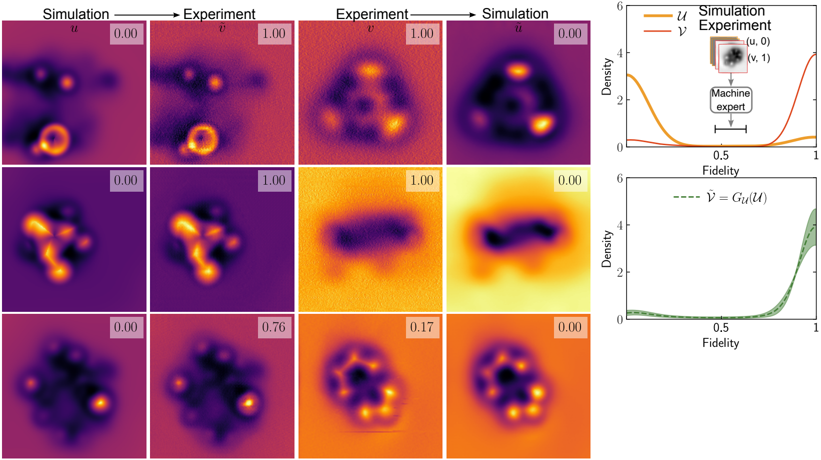
- CycleGAN[a] learns to map between two domains — without paired training data.
- It enables realistic style translation from simulated to experimental AFM images.

[a] Zhu, J.-Y. et al., 2020, arXiv:1703.10593

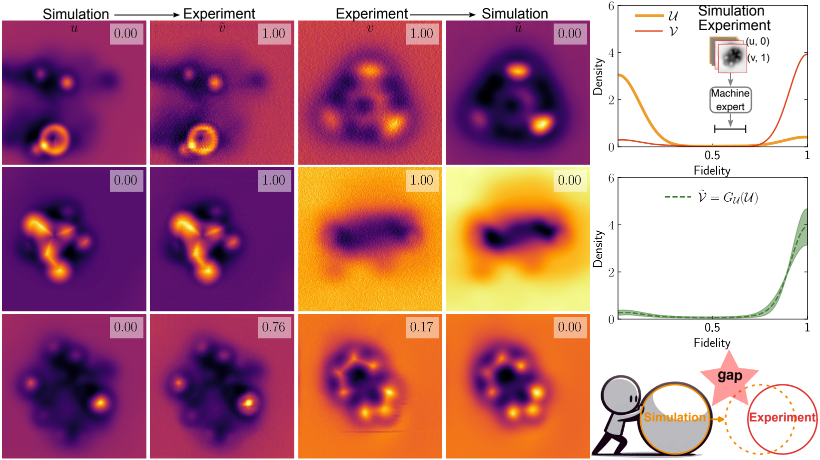
Our model learns realistic image features from experimental data



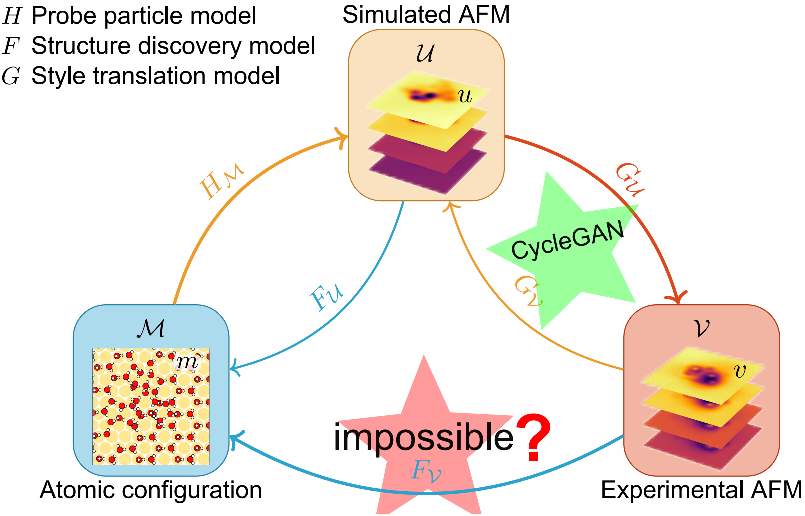
Our model learns realistic image features from experimental data



Our model learns realistic image features from experimental data

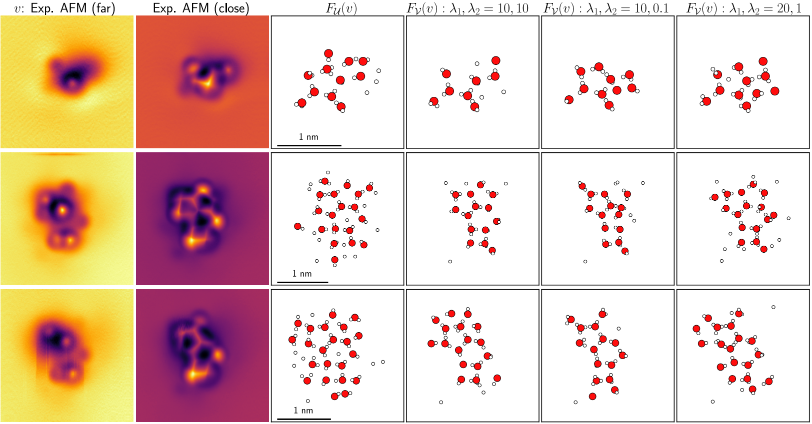


Obtain the training data in experimental-like AFM images

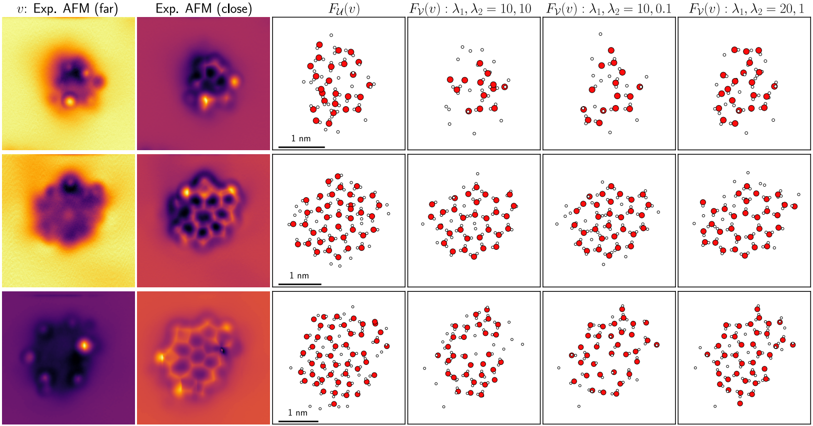


Train models on style-translated data

Atomic structure prediction: with and without style translation

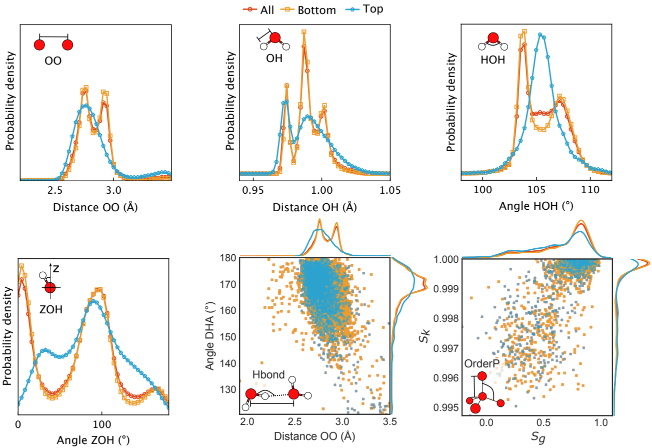


Atomic structure prediction: with and without style translation



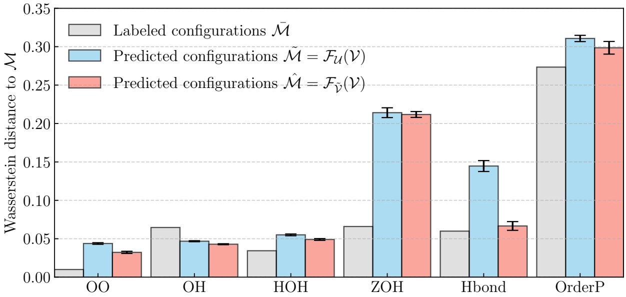
If we don't know the answers, how can we tell if the predictions are good?

Theoretical local structral distributions obtained from simulations



The good predictions of of atomic structures shoould be consistant with the theoretical local structural distributions.

Performance evaluation based on the theoretical structural distributions



We use the Wasserstein distance between the prediction and the theoretical distributions as a metric to evaluate the model performance.

Summary and outlook

Summary and outlook

We want to make structure discovery from AFM images more reliable expecially in real experimental data. We believe machine learning can help us to achieve this goal. But only if it learns from the right data.

- Simulated AFM $\rightarrow$ style-translated to look real
- Realistic AFM $\rightarrow$ better ML training $\rightarrow$ better predictions
- It makes AFM image interpretation more reliable
- A step closer to **automated atomic-level discovery**
- Bayesian optimization and diffusion model-based methods can be used

Acknowledgments

Colleagues at the SIN group in Aalto University, Finland.

- PPM: https://github.com/Probe-Particle/ppafm
- Prof. Ying Jiang's lab, Peking University, China.
- World Premier International Research Center Initiative (WPI), MEXT, Japan.
- The Academy of Finland (Projects 347319 and 346824).
- Computational resources provided by the Aalto Science-IT Project and CSC.
- Organizers of the ML4MD 2025 conference.