Understanding Crystallization with Ab Initio Machine Learning Simulation

Pablo Piaggi
CIC nanoGUNE and Ikerbasque
San Sebastian, Spain

CIC nanoGUNE
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

ikerbasque
Basque Foundation for Science

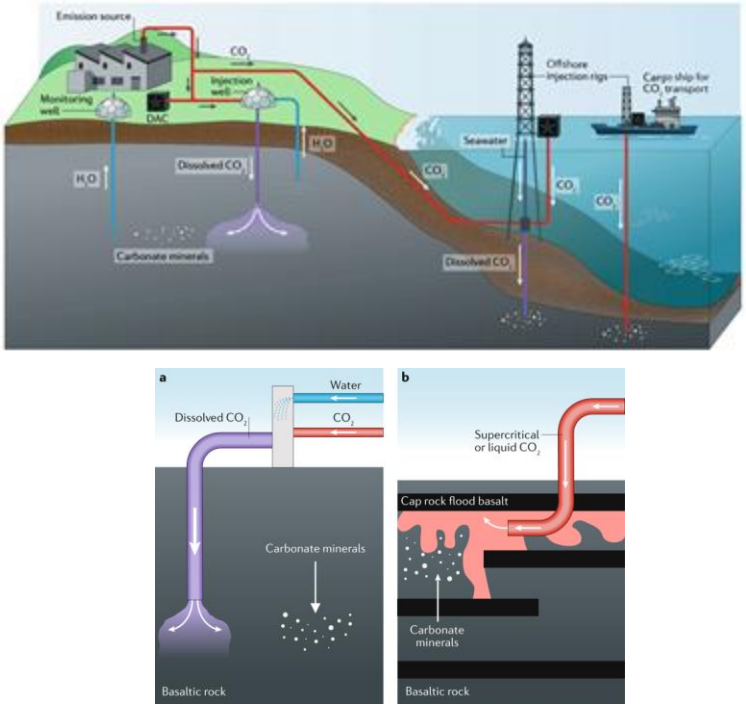
Conference on Frontiers in Atomistic Simulations
ICTP, Trieste, Italy

September 8, 2025

Tackling challenging and useful problems

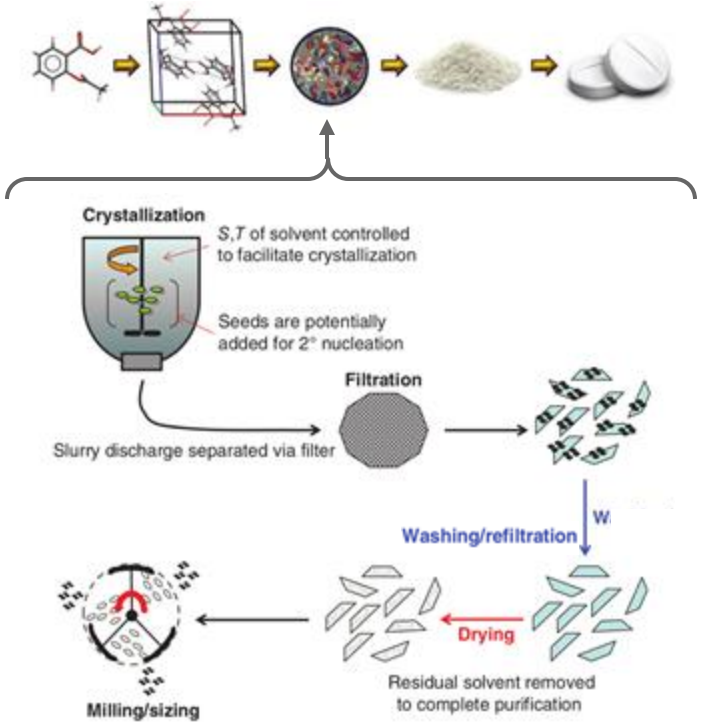
The case of crystallization

Climate change mitigation via mineral carbonation



Snæbjörnsdóttir et al, *Nature Reviews Earth & Environment* 1, 90 (2020)

Manufacturing of pharmaceutical materials



Li and Mattei, Pharmaceutical Crystals (2018)

Rare, but beautiful fluctuations

Classical Nucleation Theory (CNT)

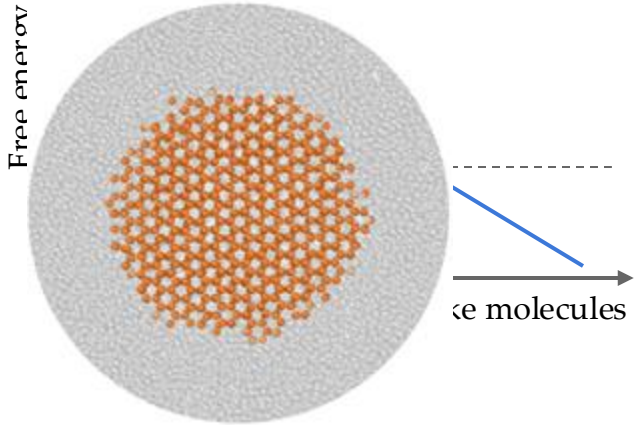
$$\Delta G = -\Delta\mu\, n + C\gamma n^{2/3}$$

total free energy | bulk | interface

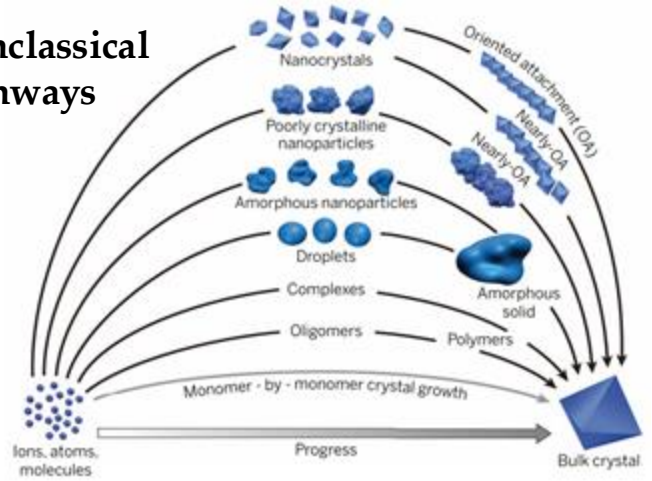
$\Delta\mu$: supersaturation

γ : interfacial free energy

n : # of solid-like molecules

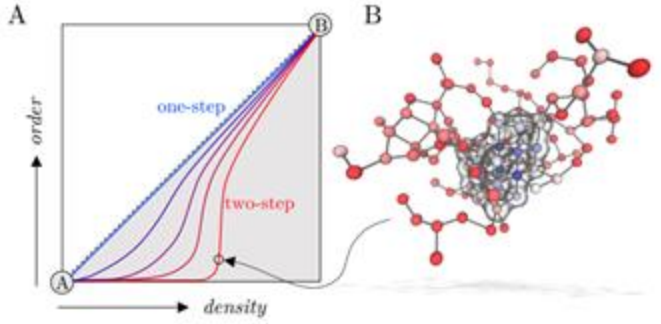


Nonclassical pathways



Crystallization via attachment of complex particle assemblies

Two-step nucleation mechanisms



J. De Yoreo et al., *Science* 349, 6247 (2015)

Finney and Salvalaglio, *Faraday Discussions* (2021)

Ab initio Atomistic Machine Learning

Training set

$$\hat{\mathcal{H}}_e(\mathbf{R})\,|\Psi_0\rangle = E_0(\mathbf{R})\,|\Psi_0\rangle$$

$$O(\mathbf{R}) = \langle\Psi_0|\hat{O}(\mathbf{R})|\Psi_0\rangle$$

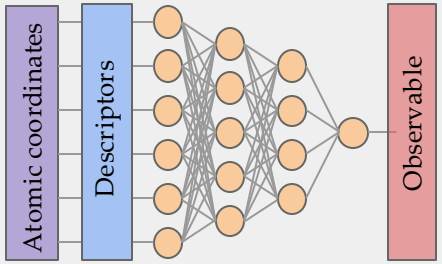
Compute observables from quantum mechanical electronic structure calculations

Typically at the density-functional theory (DFT) level

Multiple observables:

- Potential energy (potential energy surface)
- Forces
- Dipoles and Wannier centers
- Charges
- Magnetic moment

Model



$$O(\mathbf{R};\theta) = \sum_{i=1}^{N} O_i(\{\mathbf{r}_{ij} \in \mathcal{N}_i\};\theta)$$

Define a model for the observables

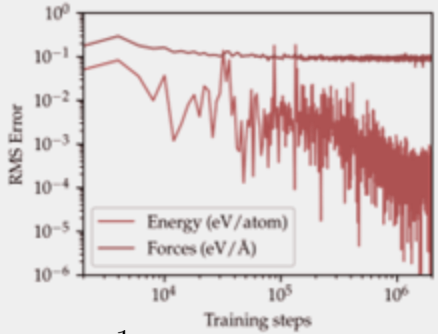
Model based on neural networks

Descriptors preserve symmetries and are key tools for analysis

Short-range with cutoffs around 6 Å

Global or local, and can be scalars, vectors, or tensors. Equivariant or invariant in O(3)

Training process



$$\mathcal{L}(\theta) = \frac{1}{N_{\mathcal{B}}} \sum_{i\in\mathcal{B}} |O^i(\theta) - O^i_{\mathrm{DFT}}|^2$$

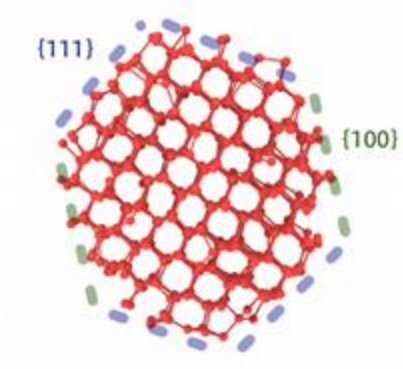
Train model by minimizing a loss function

Mean square error loss function leads to a normal distribution of errors

Minimization with algorithm akin to steepest-gradient descent

Behler and Parrinello, *Phys. Rev. Lett. 98*, 14 (2007)

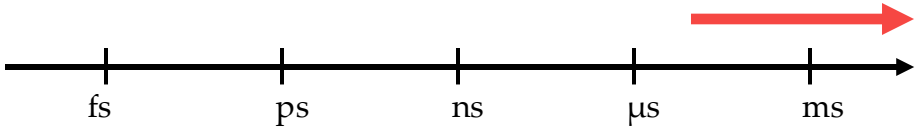
Enhanced sampling and the rise of OPES



Crystal nucleation

Rare-event techniques

- Seeding technique
- Umbrella sampling
- Metadynamics
- Path sampling



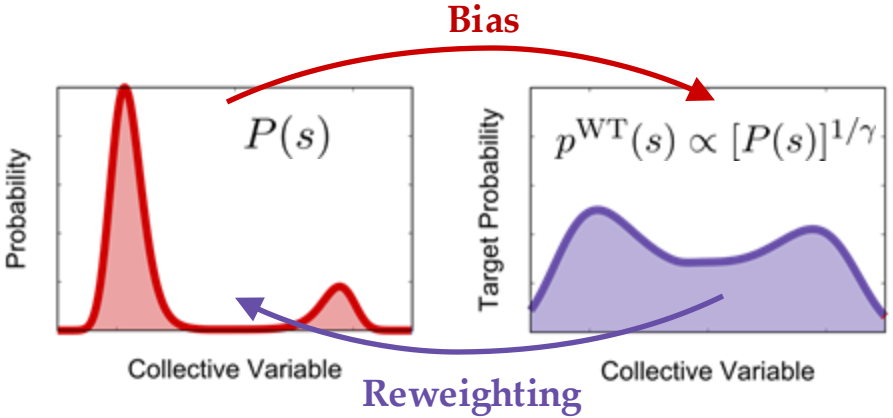
Direct ab initio MD

MD with machine learning models

MD with empirical models

Laio and Parrinello, *Proc. Natl. Acad. Sci.* 99 (2002)
Valsson, Tiwary, and Parrinello, *Annu. Rev. Phys. Chem. 67* (2016)

OPES (On the fly Probability Enhanced Sampling)



$$V(s) = (1 - 1/\gamma)\frac{1}{\beta}\log P(s)$$

Iterative reweighted KDE to determine P(s)

OPES advantages:

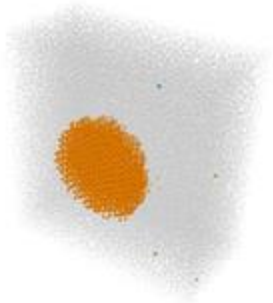
- Very easy to use
- Fast convergence
- Stationary bias → easy reweighting
- Kernel compression to avoid slowdown

PLUMED
PLUgin for MolEcular Dynamics
www.plumed.org

Invernizzi and Parrinello, *J. Phys. Chem. Lett.*, 11, 2731 (2020)
Invernizzi, **Piaggi**, and Parrinello, *Phys. Rev. X* 10, 041034 (2020)

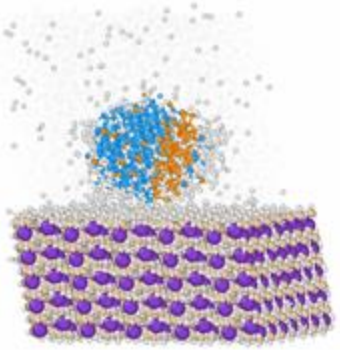
A few examples

Homogeneous ice nucleation



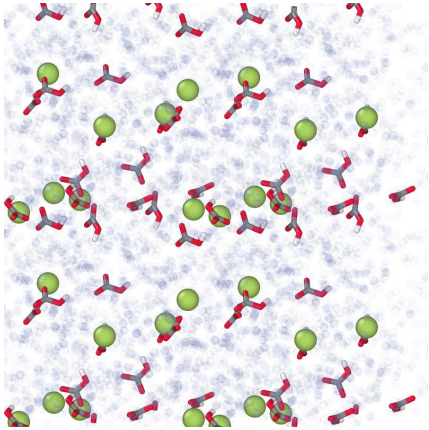
Piaggi, Weis, Panagiotopoulos, Debenedetti, and Car, *PNAS* 119, 33 (2022)

Heterogeneous ice nucleation on feldspar



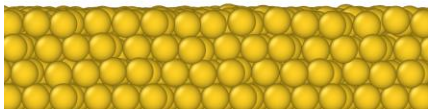
Piaggi, Selloni, Panagiotopoulos, Car, and Debenedetti, *Faraday Discussions* (2024)

Calcium carbonate precipitation



Piaggi, Gale, and Raiteri, *PNAS* (2025)

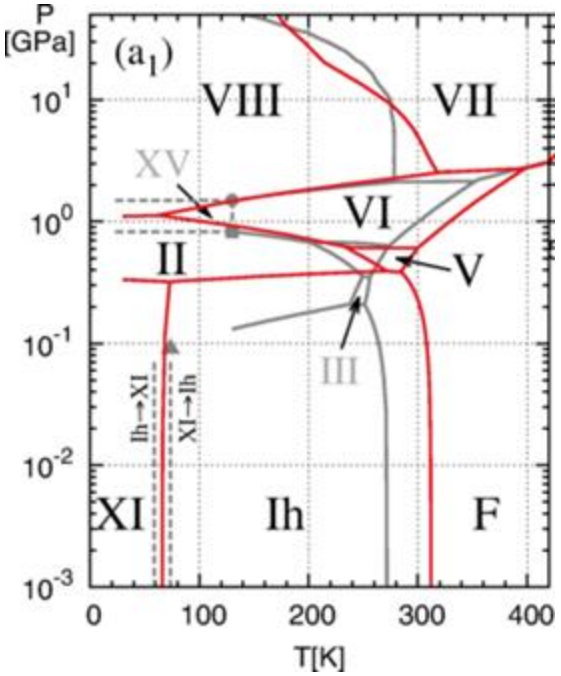
Growth of $CoBr_2$ on Au(111)



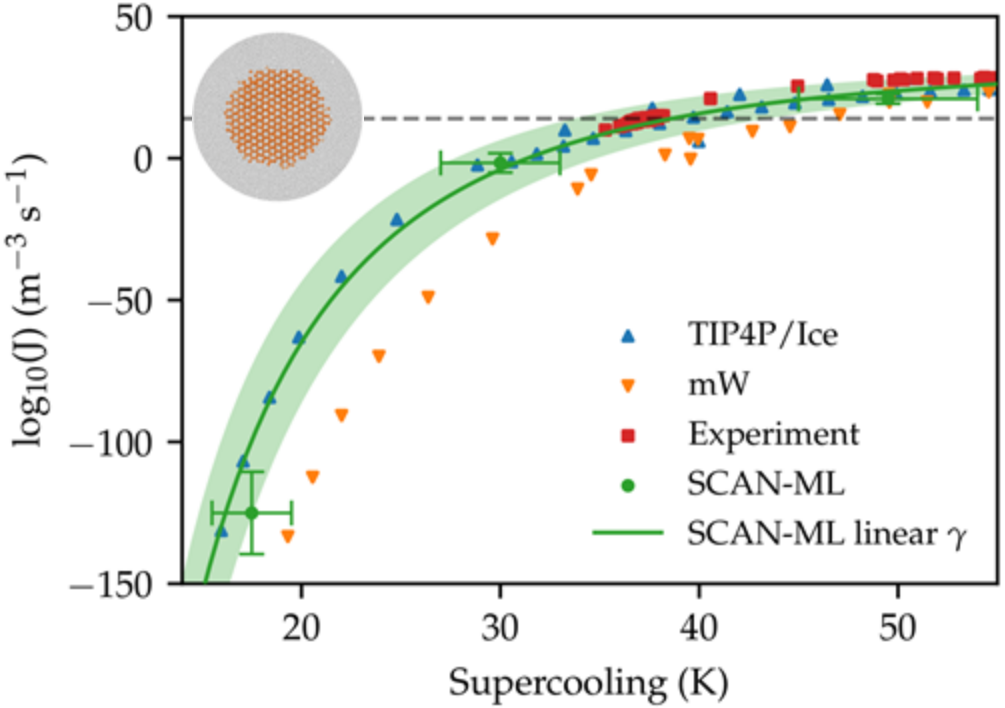
Kerschbaumer, **Piaggi**, Rogero, et al., *submitted* (2025)

Ab initio water and ice (SCAN DFT functional)

Water/ice phase diagram from machine-learning driven MD



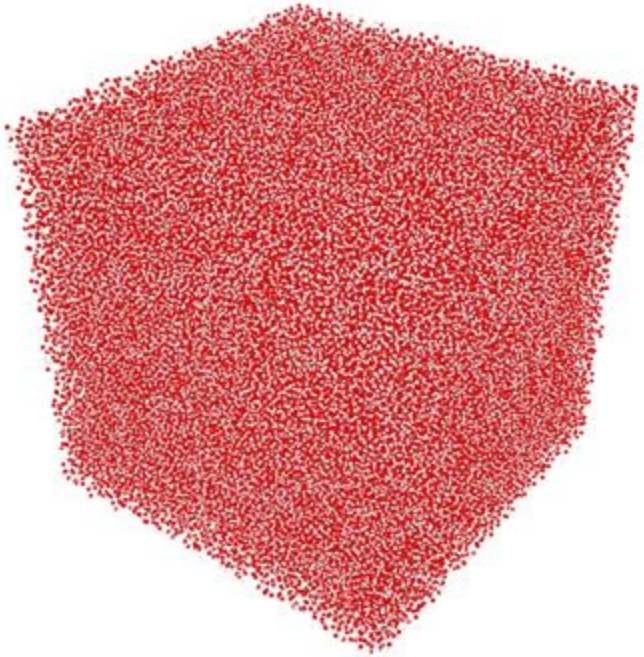
Ice Ih nucleation rates from supercooled liquid water



Zhang, Wang, Car, and E, *Phys. Rev. Lett.* 126, 236001 (2021)

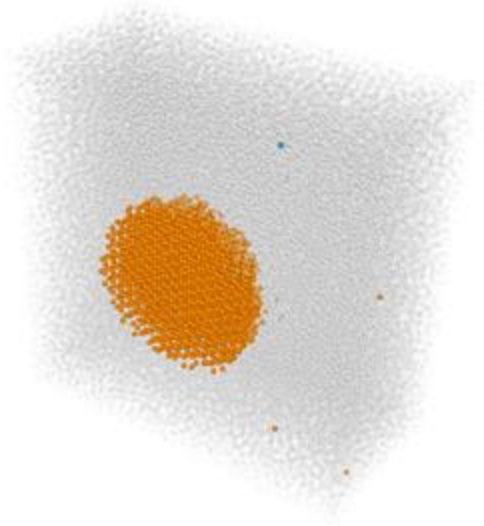
Piaggi, Weis, Panagiotopoulos, Debenedetti, and Car, *PNAS* 119, 33 (2022)

Simulations require high-performance computing capabilities



: Oxygen

: Hydrogen



: Oxygen with **liquid**-like environment

: Oxygen with **ice Ih**-like environment

: Oxygen with **ice Ic**-like environment

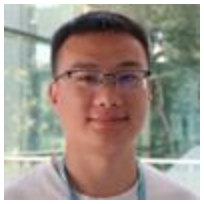
We used DeePMD and the **SUMMIT supercomputer** to simulate ~300,000 atoms with ab initio accuracy on **600 NVIDIA V100 GPUs**

TensorFlow

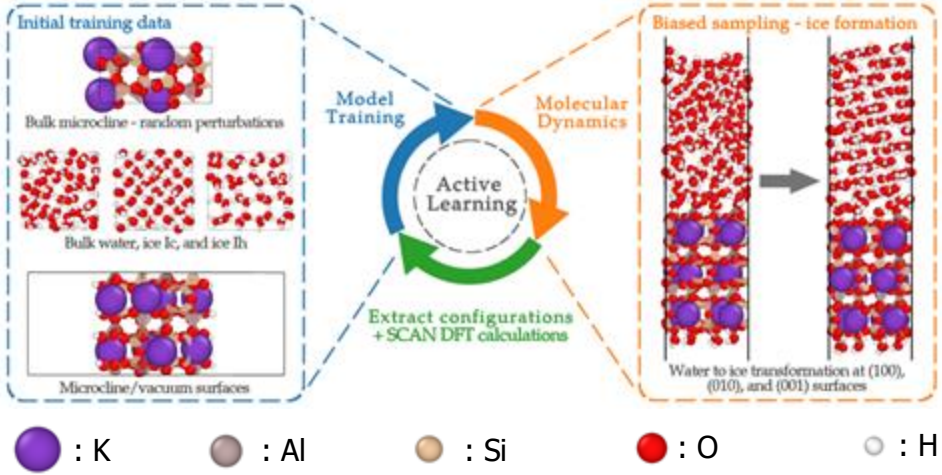
DeePMD-kit

Piaggi, Weis, Panagiotopoulos, Debenedetti, and Car, *Proc. Natl. Acad. Sci.* 119, 33 (2022)

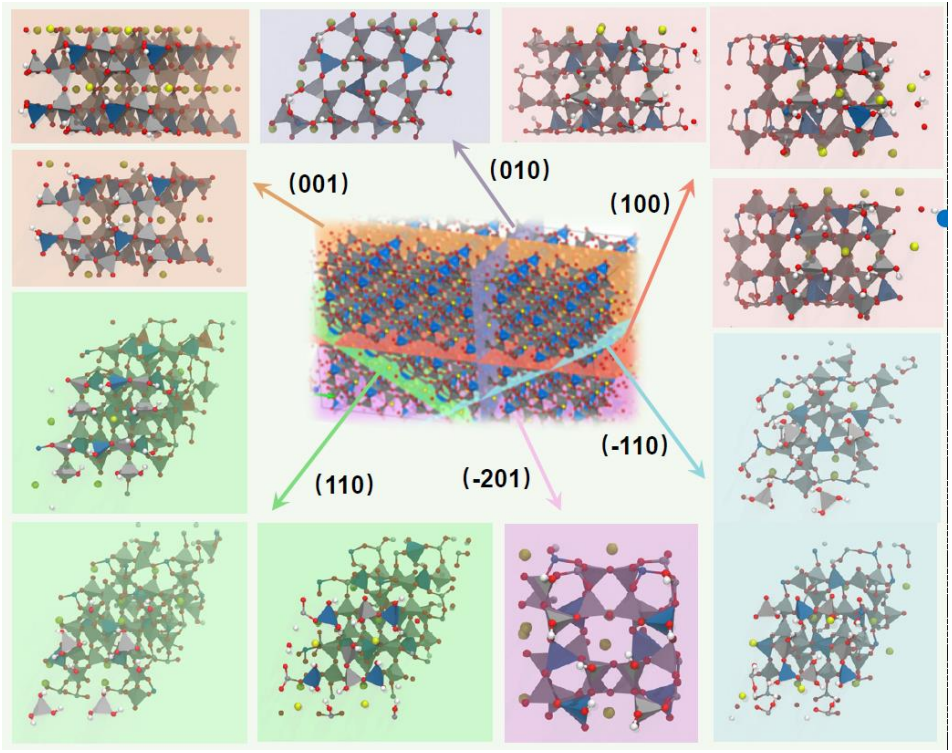
Ab initio machine learning simulations of ice nucleation on feldspar



Wanqi Zhou
Postdoctoral researcher @
nanoGUNE



Trained using biased sampling and active learning to capture the formation of ice at surfaces.



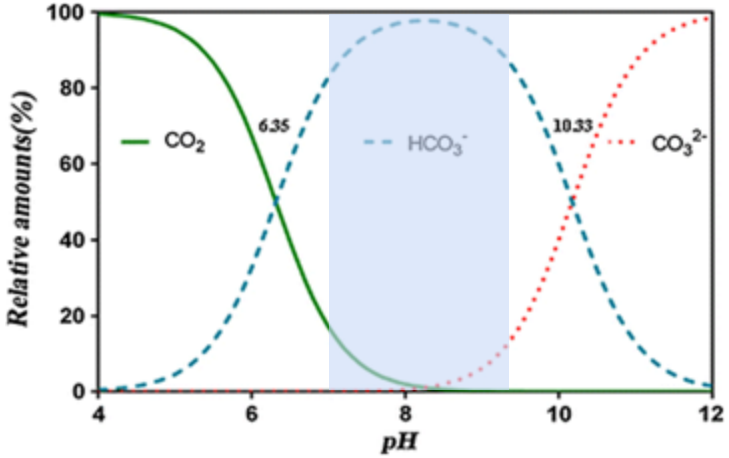
Thorough analysis of ice nucleation activity for the multiple felspar crystallographic surfaces

Piaggi, Selloni, Panagiotopoulos, Car, and Debenedetti, *Faraday Discussions* 249, 98 (2024)

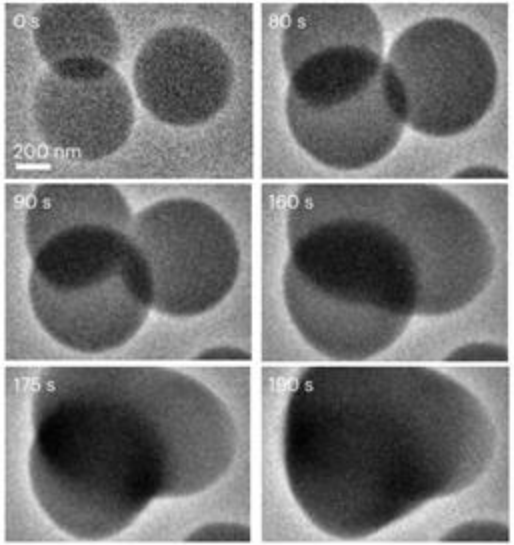
Crystallization of $CaCO_3$ at near neutral pH

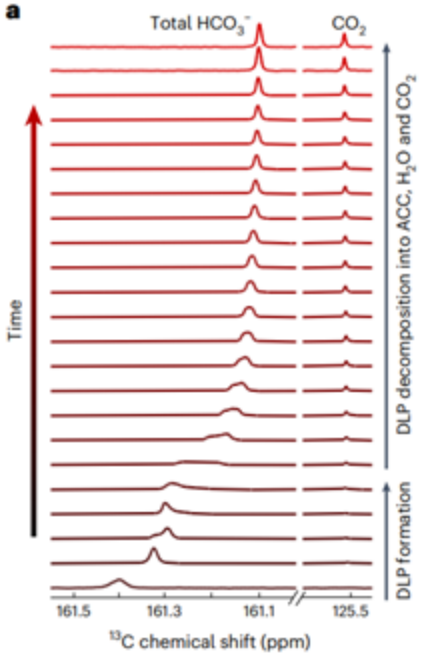
Eslam Ibrahim
Visiting PhD student @ nanoGUNE (collaboration with Ralf Drautz)

CO_2 / HCO_3^- / CO_3^{2-} speciation vs pH



Mechanism for the bicarbonate to carbonate transformation is mot fully understood



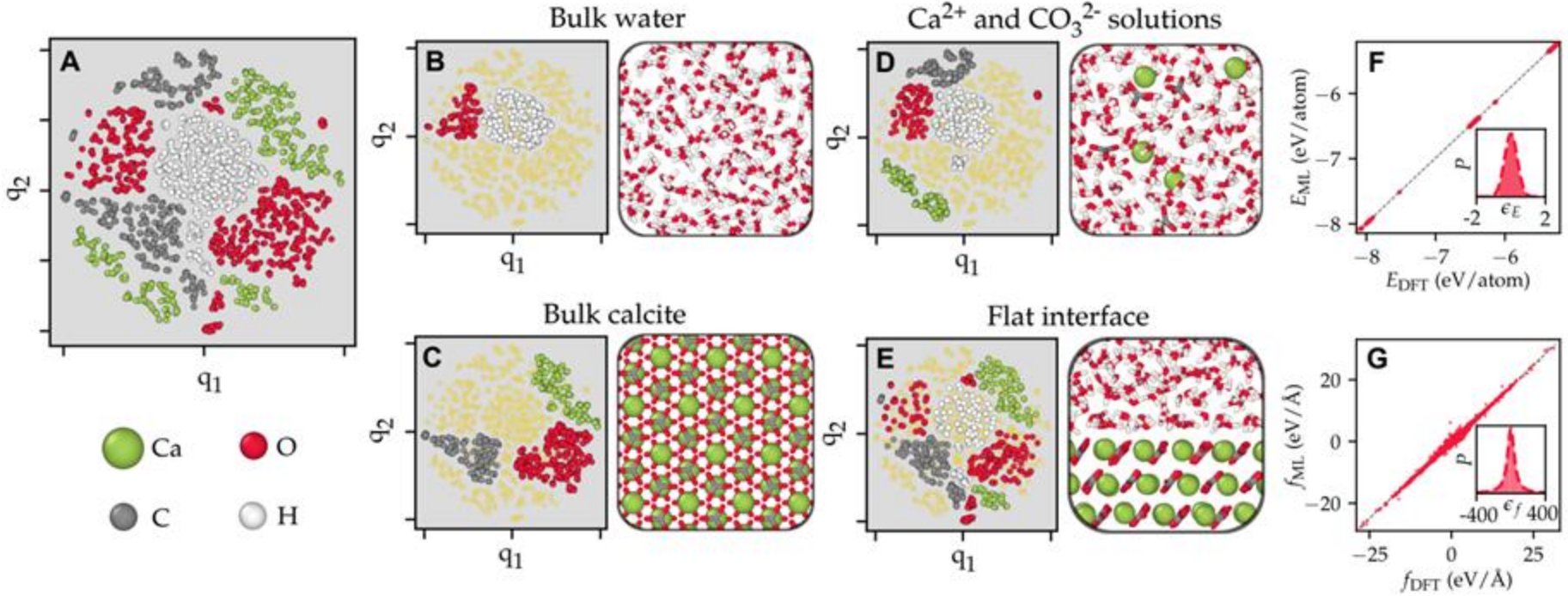


Reactivity is key for this process and **ML potentials are fully reactive**

Jin, De Yoreo, et al, *Nat. Mater.* (2024)

Ab initio machine learning simulation of calcium carbonate

Training set and accuracy of SCAN-DFT model

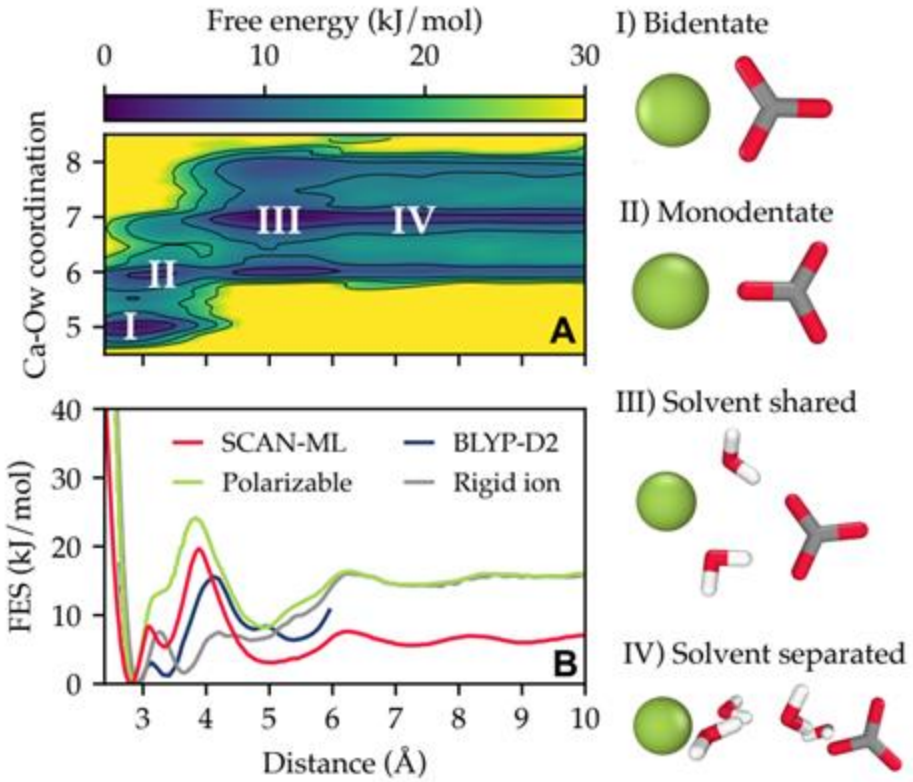


t-SNE dimensionality reduction using DeePMD descriptors

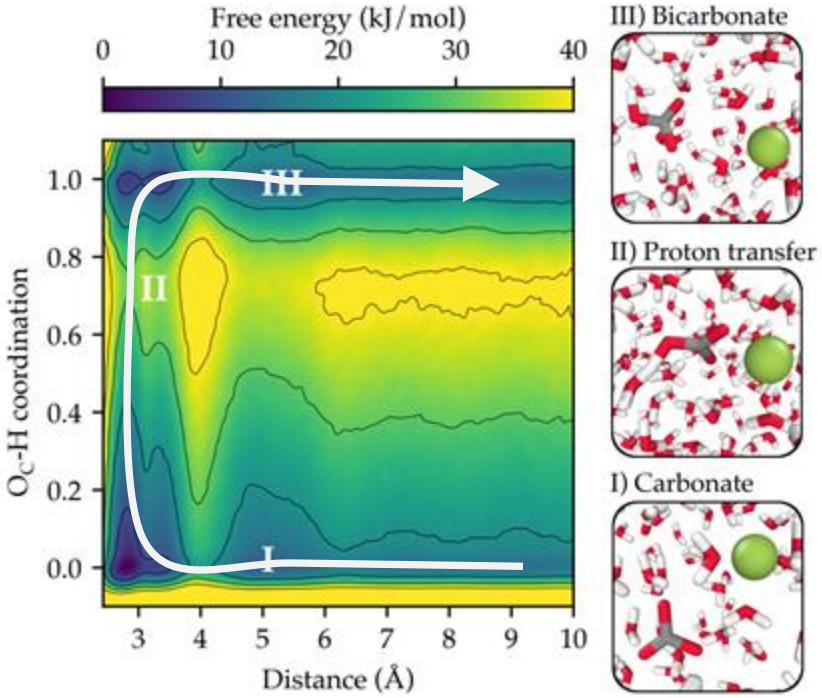
Piaggi, Gale, and Raiteri, *PNAS* (2025)

Improved ion association thermodynamics and reactivity

Ion pair association via OPES simulation



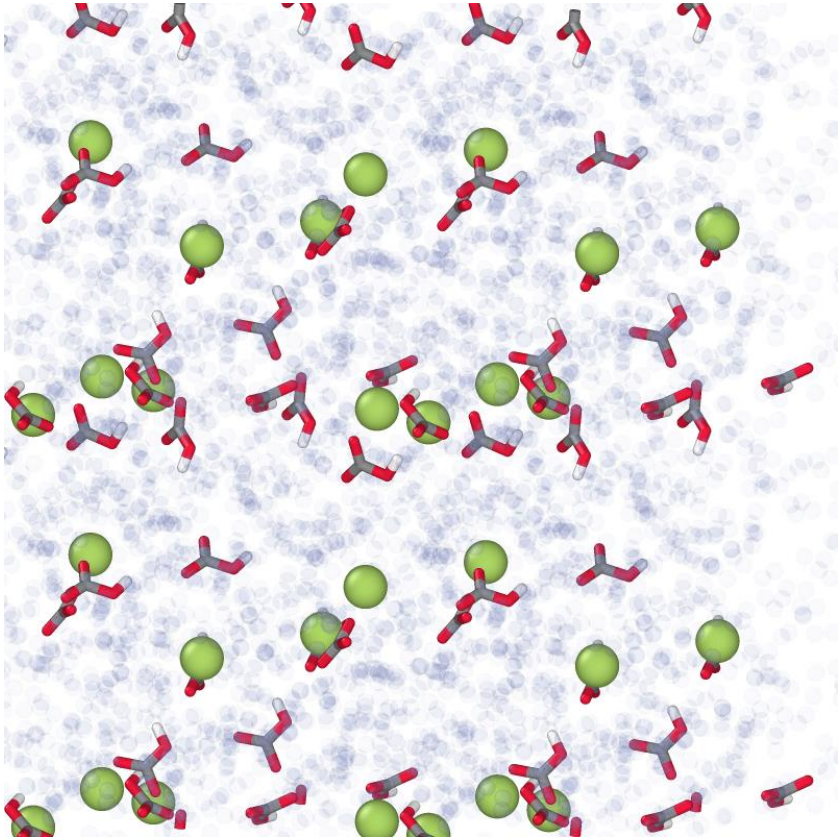
Carbonate to bicarbonate transformation

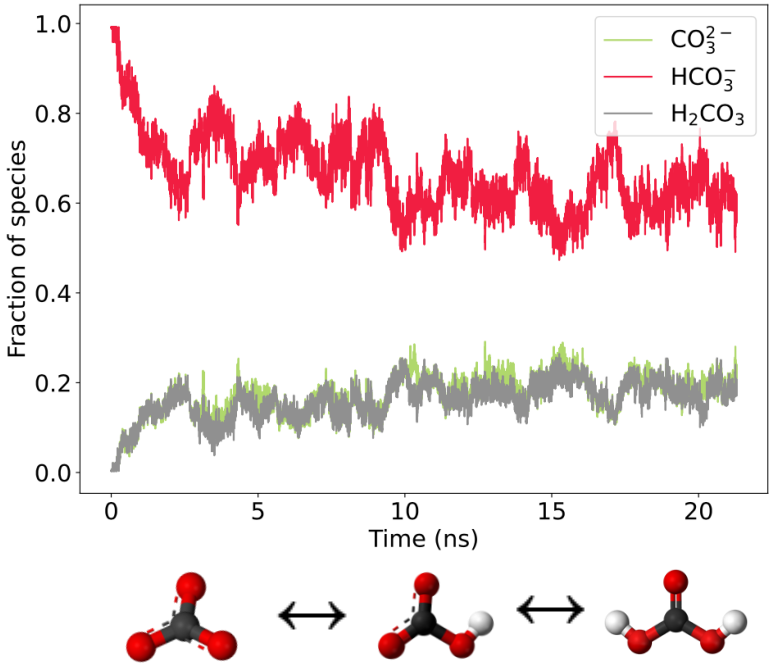


50-ns-long simulations with 5000 atoms

Piaggi, Gale, and Raiteri, *PNAS* (2025)

$CaCO_3$ precipitation with unconstrained reactivity





- Species interconvert via proton transfer
- Mean (max) error in forces 20 meV/Å (150 meV/Å)
- First time this process is simulated with full reactivity

Piaggi, Gale, and Raiteri, *PNAS* (2025)

Ab initio machine learning electrostatics

Wannier centers

$$\hat{\mathcal{H}}_e(\mathbf{R})\,|\Psi_0\rangle = E_0(\mathbf{R})\,|\Psi_0\rangle$$

Bloch's theorem (delocalized eigenstates)

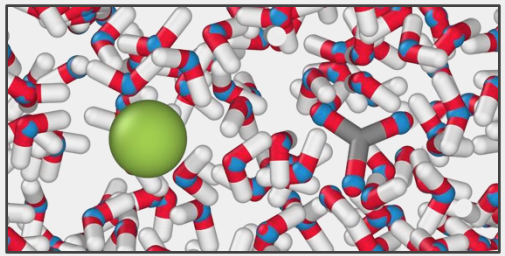
$$|\Psi_{nk}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{nk}\rangle$$

Wannierization (localized eigenstates)

$$|w_{nR}\rangle = \frac{\Omega}{(2\pi)^2}\int d\mathbf{k}\, e^{i\mathbf{k}\cdot\mathbf{r}}|\Psi_{nk}\rangle$$

Wannier centers

$$\mathbf{r}^W_{n0} = \langle w_{n0}|\hat{\mathbf{r}}|w_{n0}\rangle$$



Ca O C H

Wannier centroids

Marzari and Vanderbilt., *Phys. Rev. B 56*, 20 (1997)

Electrostatics and ML model

Electrostatic energy

$$E_{\text{elec}}(\mathbf{R},\mathbf{R}^{\mathbf{W}}) = \sum_{i,j}\frac{Z_iZ_je^2}{4\pi\epsilon_0|\mathbf{r}_i-\mathbf{r}_j|} + \sum_{i,j}\frac{-2Z_ie^2}{4\pi\epsilon_0|\mathbf{r}_i-\mathbf{r}^W_j|} + \sum_{i,j}\frac{4e^2}{4\pi\epsilon_0|\mathbf{r}^W_i-\mathbf{r}^W_j|}$$

Charges are constant, but the Wannier centers can move!
→ polarization and charge conservation
Large systems cannot be tackled directly

Train an ML model to predict positions of Wannier centers

$$\mathbf{W}_i(\{\mathbf{r}_{ij}\in\mathcal{N}_i\})$$

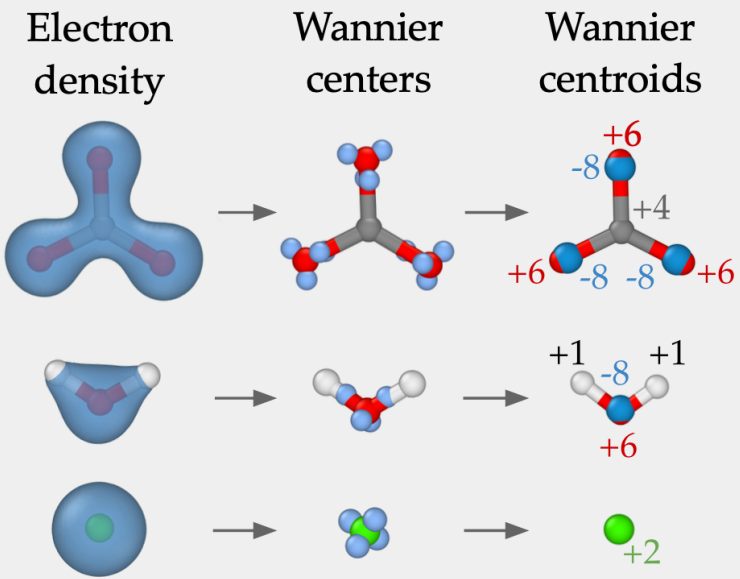
Using a loss function

$$\mathcal{L}(\theta) = \frac{1}{N_{\mathcal{B}}}\sum_{i\in\mathcal{B}}|\mathbf{W}^i(\theta)-\mathbf{r}^i_{\text{W,DFT}}|^2$$

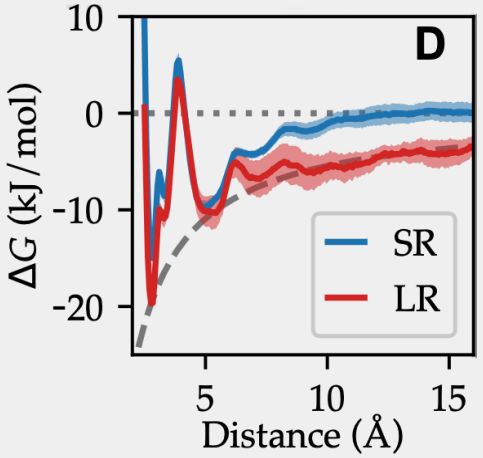
Zhang et al., *J. Chem. Phys.* 156, 12 (2022)

Are long-range interactions relevant for $CaCO_3$?

Wannier centers and centroids



Ion pair association with long-range interactions



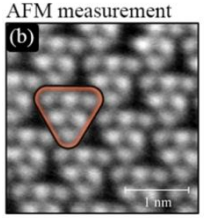
$$\Delta G = -\frac{1}{\beta} \ln\left(C_0 \int_0^{r^*} dr\, 4\pi r^2 e^{-\beta W(r)} \right)$$

Long-range interactions are important to capture the *ab initio* thermodynamics of this system

Piaggi, Gale, and Raiteri, *PNAS* (2025)

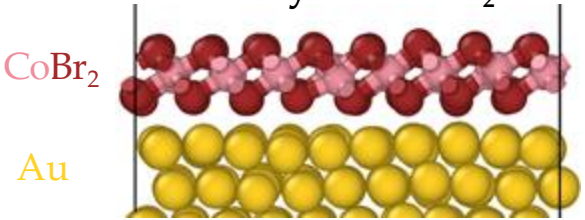
Growth of magnetic 2D materials

Samuel Kerschbaumer
PhD student @ CFM
(collaboration with Celia Rogero / Max Ilyn)



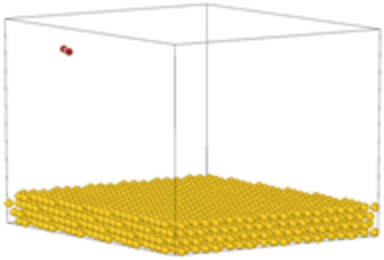
There is a precursor phase (PP) that precedes the formation of the monolayer (ML). There seems to be a triangular motif in the PP.

Van der Waals layer of $CoBr_2$ on Au(111)

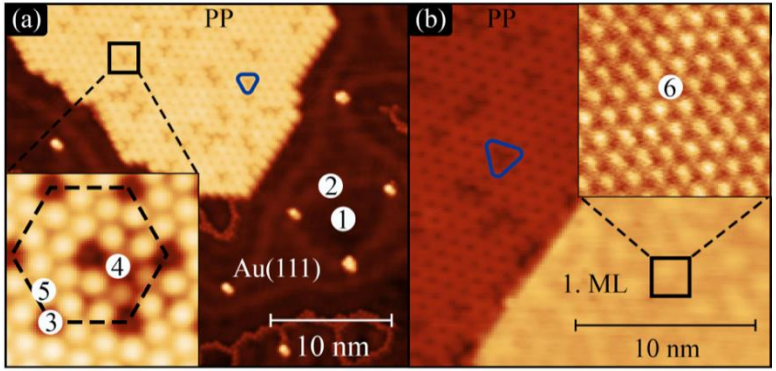


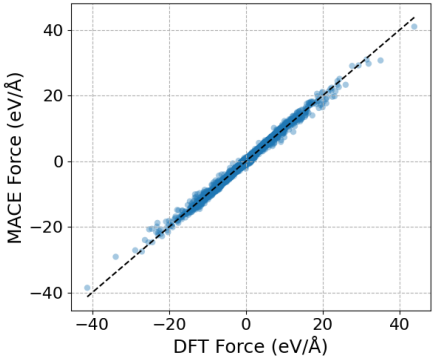
Was thought to form directly during vapor depositions

Simulate the **reactive** deposition of molecules using the **MACE-MP** foundation model



STM images of growth process





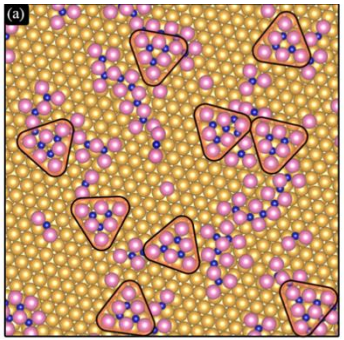
Validation with spin-polarized DFT calculations

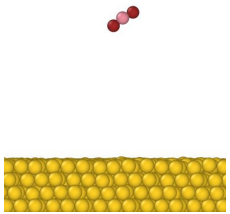
- MAE = 103 meV/Å
- RMSE = 152 meV/Å
- R^2 = 0.9743

Kerschbaumer, **Piaggi**, Rogero, et al., *submitted* (2025)

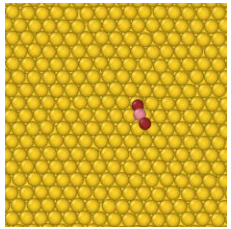
Growth of magnetic 2D materials

MD simulations identify the triangular motif

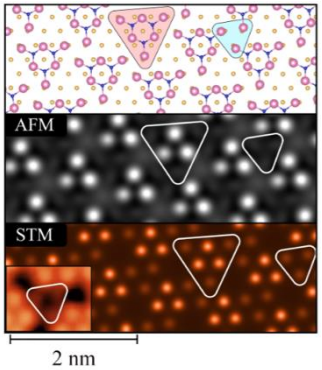




Side View

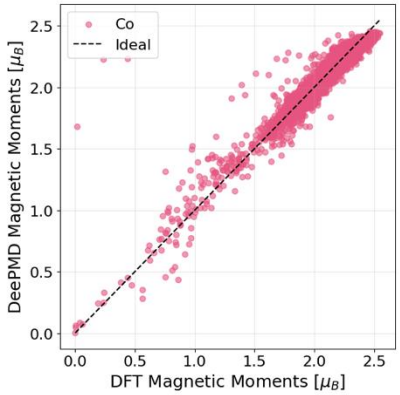


Top View



The PP structure is finally confirmed with direct DFT calculations, bolstering the evidence for a **two-step growth mechanism** in this system

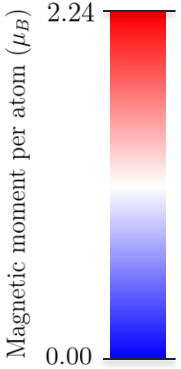
Machine learning model for the magnetic moment



- MAE = 0.04 μ_B
- RMSE = 0.06 μ_B
- R^2 = 0.93

Co

Br or Au



Kerschbaumer, **Piaggi**, Rogero, et al., *submitted* (2025)

Collaborators and funding sources

Team members @ nanoGUNE

- Wanqi Zhou
- Pablo Peña
- Eslam Ibrahim (Visiting from Bochum, Germany)
- Sergio Estrada

Collaborators @ Curtin University

- Paolo Raiteri
- Julian Gale

Collaborators @ CFM

- Samuel Kerschbaumer
- Max Ilyn
- Celia Rogero

Funding

Swiss National Science Foundation

ikerbasque Basque Foundation for Science

Funded by the European Union

U.S. DEPARTMENT OF ENERGY | Office of Science

Supercomputing resources

RES RED ESPAÑOLA DE SUPERCOMPUTACIÓN

DIPC Donostia International Physics Center