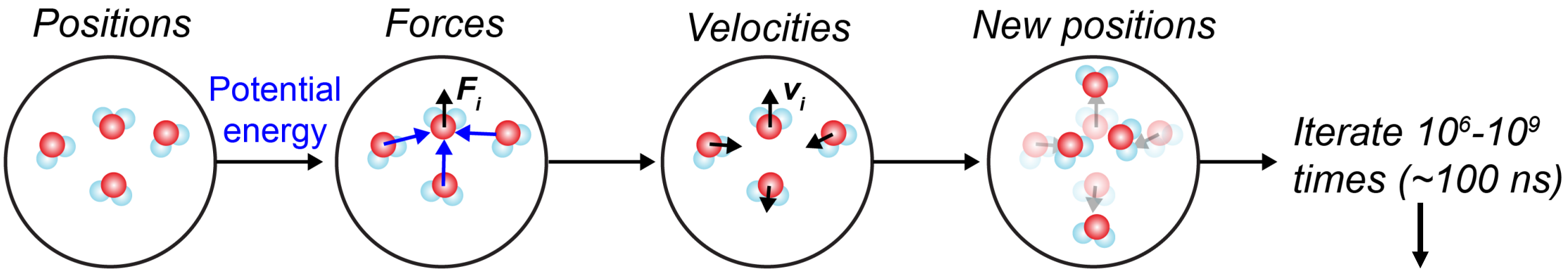
Developing a High Throughput Computational Workflow for Rationalizing Molecular Crystallization

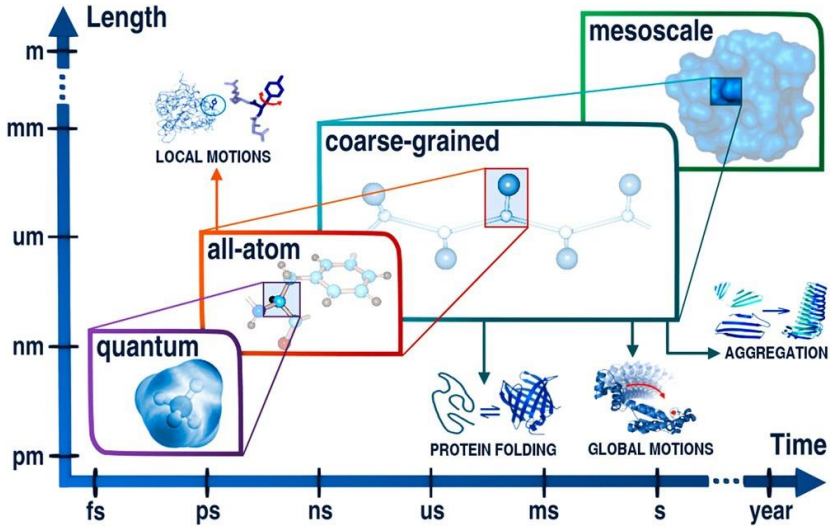
Jordan Loeffler[*1], Christian Jorgensen[*1], Rose Cersonsky[1, 2, 3]

1. Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI
2. Department of Materials Science and Engineering, University of Wisconsin-Madison, Madison, WI
3. Data Science Institute, University of Wisconsin-Madison, Madison, WI

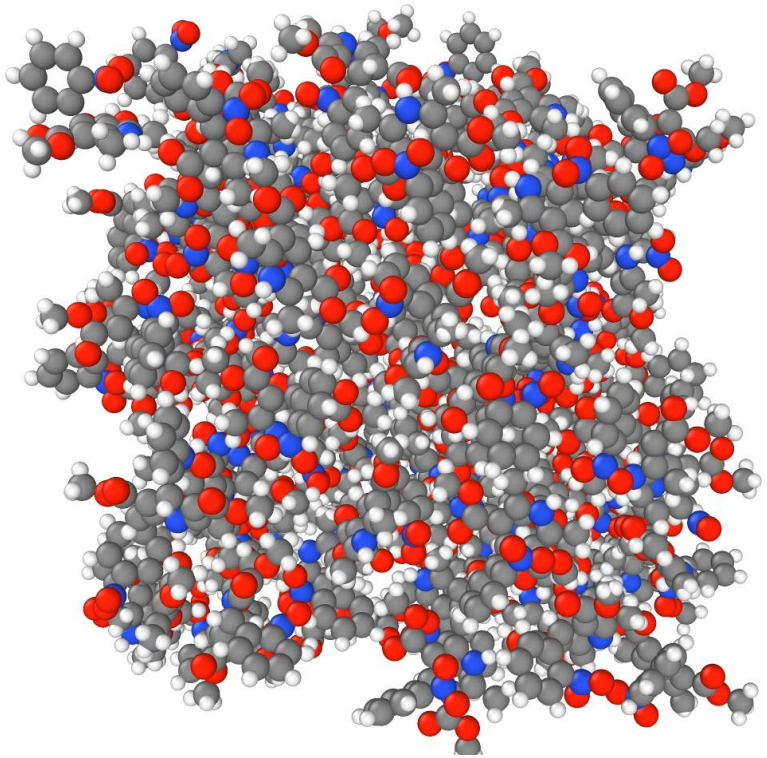
What is Molecular Dynamics?



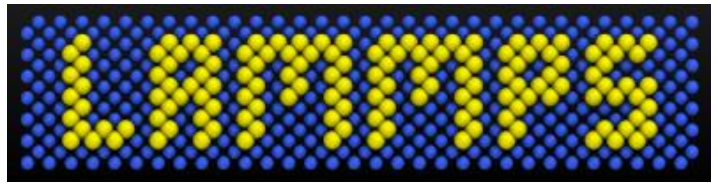
Molecular dynamics simulations apply **Newtonian mechanics** to **capture molecular motion**



There are many molecular simulations with various resolution capabilities, molecular dynamics is just one of them



Molecular Dynamics Simulation Engines Offer GPU Compatibility



- GPU package
  - CPU computes neighbor list building, bonding forces, and electrostatics
  - GPU handles pairwise force calculations
- KOKKOS Package
  - Write physics kernel once, and run entirely on GPU
- CUDA, AMD, SYCL, OpenCL, OpenMP supported
- Best for materials research, polymers

- Hybrid CPU/GPU offloading
  - Non-bonded forces on GPU and electrostatics and bonded forces on CPU
- CUDA, SYCL, OpenCL supported
- Best for standard high-throughput biomolecular simulations (e.g., proteins, lipids)

- Forces, integration, and constraints all calculated and updated on GPU
- Eliminates CPU-to-GPU data transfer bottlenecks
- CUDA, OpenCL supported
- Best for custom Python-scripted workflows and integration of machine learning

Research by Cersonsky Lab Group

Molecular Representations

Arthur Lin

Colloidal Behavior in Concentrated Electrolyte Solutions

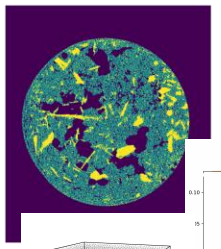
Hwigwang Lim

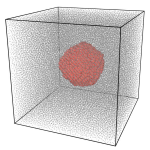
Uncovering Dynamics Underlying Glassy Systems

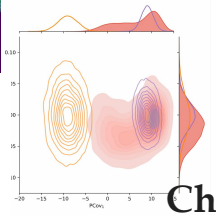
Ethan Deutsch

Charles Carroll

Enhanced Crystal Structure Prediction







Christian Jorgensen

Modes of Drug Delivery

Solid
Ex. tablets, capsules, powders

Patient preferred non-invasive, self-administration
Challenging design and formulation

Liquid
Ex. Intravenous injection, subcutaneous injection, topical creams

While injection most effective delivery method, not patient preferred

Gas
Ex. inhalers, aerosols

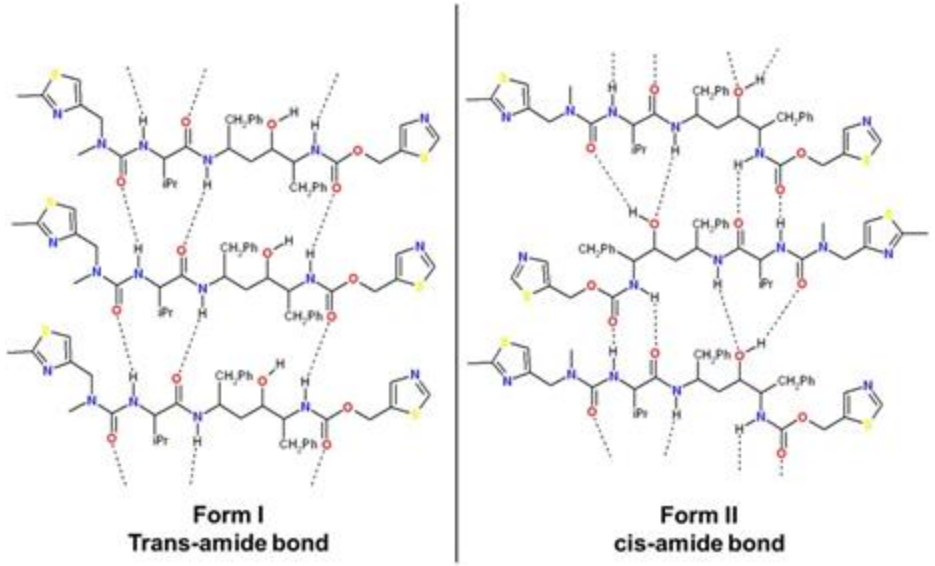
Specific use-cases, non-generalizable route of administration

The Challenge of Polymorphism

- Polymorphism – the ability for one molecule to adopt many crystalline forms
  - Significant variation in material properties (e.g., solubility, bioavailability stability)

- Organic molecules are prone to polymorphism (~37%)
  - Due to flexibility, weak intermolecular interactions, and H-bonding networks

- Each polymorph must be characterized differently in it's formulation testing
  - If another crystal structure is present than previously, it must be tested)



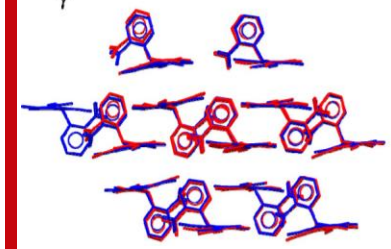
The "Ritonavir crisis" led to $250M reformulation effort by Abbott Laboratories (now AbbVie)

Iuzzolino, L., et al., Nature Communications in Chemistry (2025), 8, 404

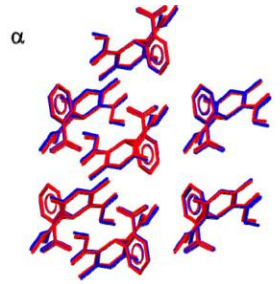
Crystal Structure Prediction Current Limitations

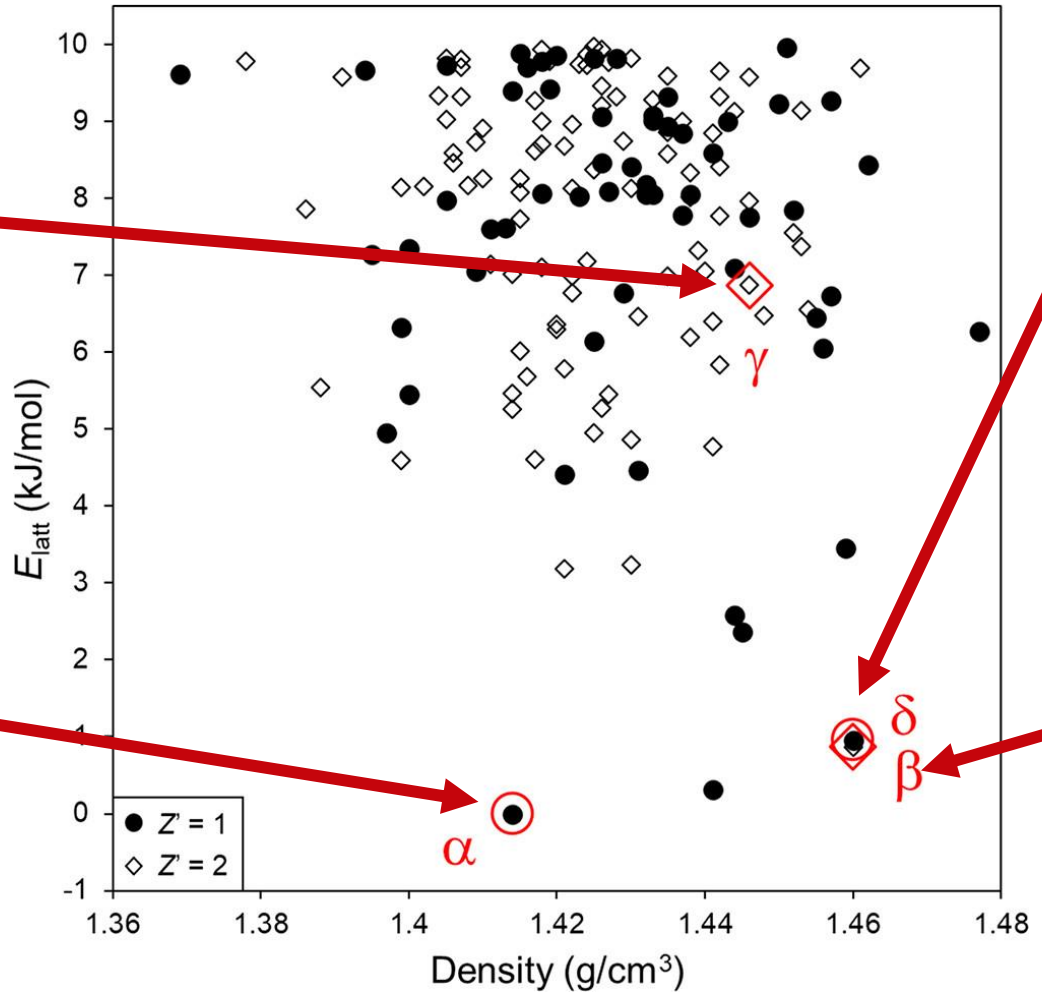
$$\Delta G(T,P) = \underbrace{\Delta U_{latt}}_{\text{lattice energy}} + \underbrace{\Delta F_{vib}(T)}_{\text{vibrational free energy}} - \underbrace{T\Delta S_{conf}(T)}_{\text{conformational entropy}} + P\Delta V$$

Gamma Polymorph

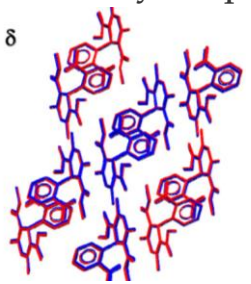


Alpha Polymorph

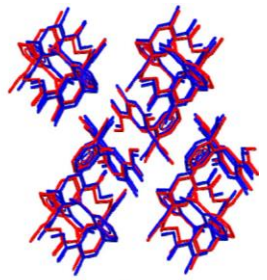




Delta Polymorph

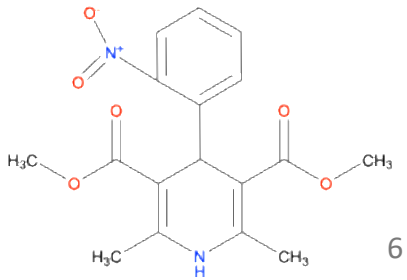


Beta Polymorph



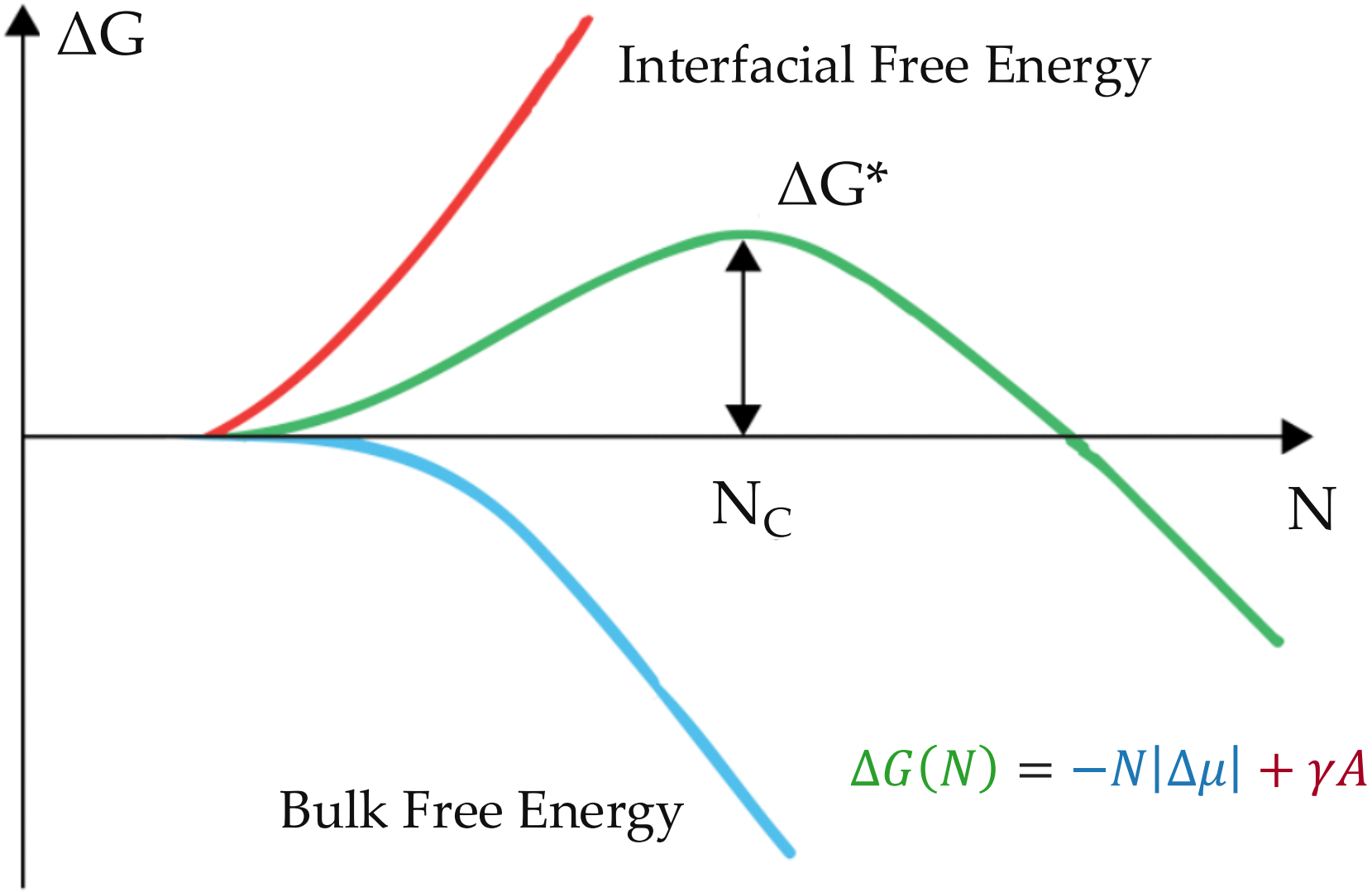
- Current crystal structure prediction (CSP) methods rely on thermodynamic stability at 0K
- These predictions omit kinetic and entropic information
- CSP methods often fail to reproduce experimentally observed behavior

Nifedipine



Gui, Y., et al., Journal American Chemical Society (2022), 22, 2, 1365-1370

Leveraging Classical Nucleation Theory: Interfacial Free Energy



$$\Delta G(N) = -N|\Delta\mu| + \gamma A$$

- **Classical Nucleation Theory:** crystallization is the competition between the bulk free energy gained and the interfacial free energy cost of forming a crystal-liquid interface
- **Hypothesis:** We can leverage differences in interfacial free energies to enhance CSP and reproduce experimental behavior
- Direct calculation of interfacial free energies is non-trivial

Gui, Y., et al., Journal American Chemical Society (2022), 22, 2, 1365-1370

Leveraging Classical Nucleation Theory: Interfacial Free Energy

- **Classical Nucleation Theory:** crystallization is the competition

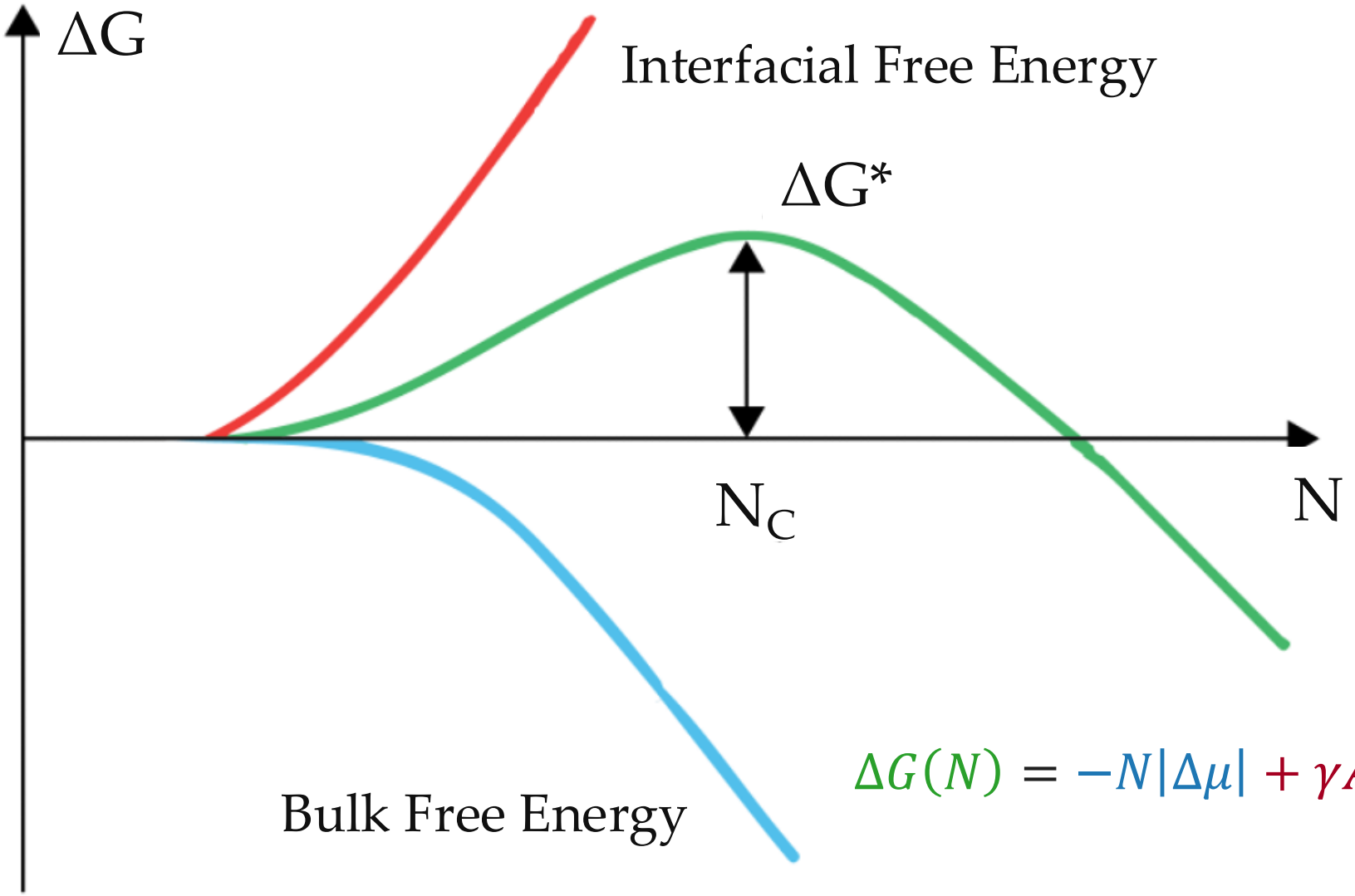
What simulation techniques can we leverage to determine interfacial free energy, and properly rank polymorphs based on kinetic factors that govern experimental behavior?

- Direct calculation of interfacial free energies is non-trivial

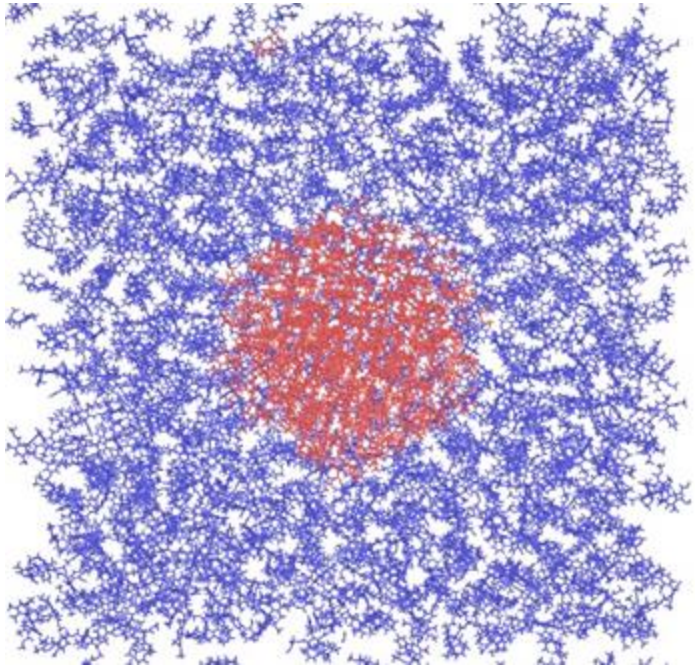
$$\Delta G(N) = -N|\Delta\mu| + \gamma A$$

Gui, Y., et al., Journal American Chemical Society (2022), 22, 2, 1365-1370

Calculating Interfacial Free Energy Using Seeding Method



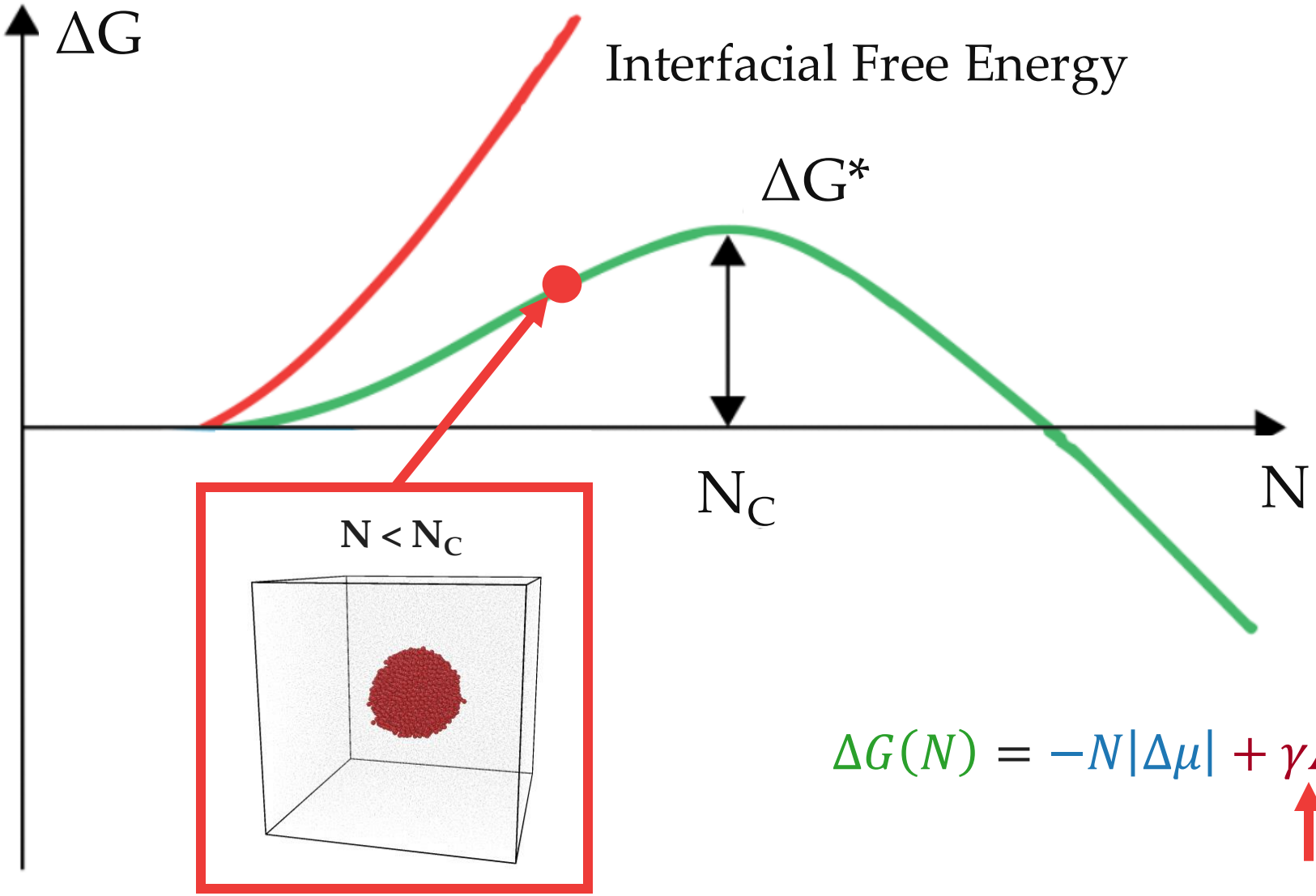
$$\Delta G(N) = -N|\Delta\mu| + \gamma A$$



- Insert a spherical crystalline seed in a liquid melt and observe the nucleation behavior

Espinosa, J. R., et al., Journal of Chemical Physics (2016), 144, 3, 034501

Calculating Interfacial Free Energy Using Seeding Method

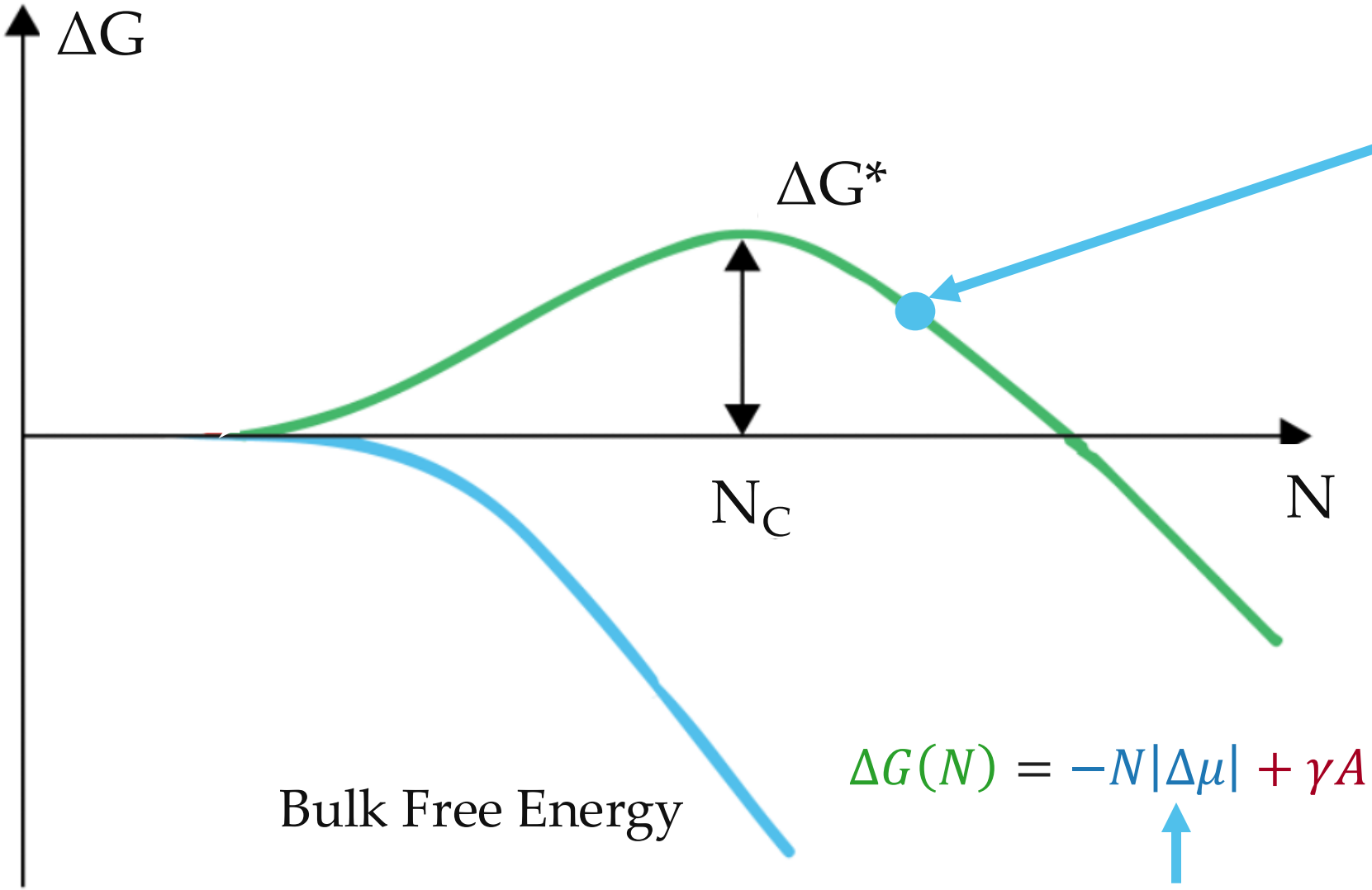


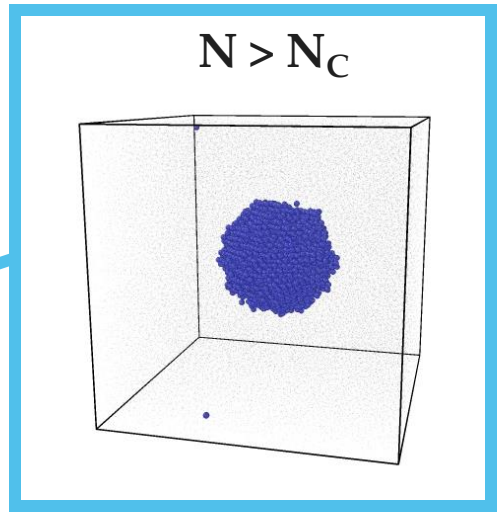
$$\Delta G(N) = -N|\Delta\mu| + \gamma A$$

- When $N < N_C$ the **interfacial free energy** term dominates and the seed **shrinks**

Espinosa, J. R., et al., Journal of Chemical Physics (2016), 144, 3, 034501

Calculating Interfacial Free Energy Using Seeding Method



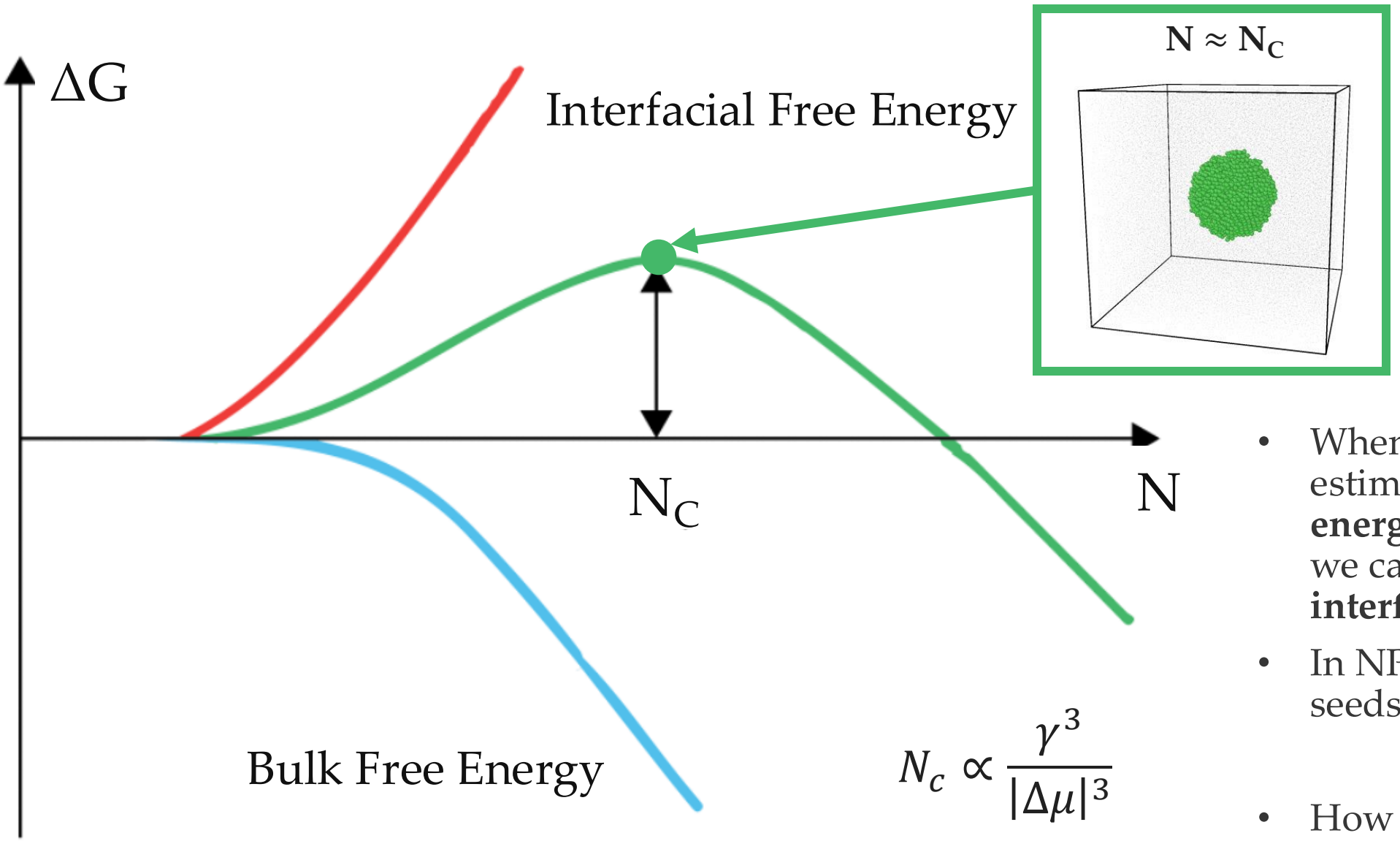


$$\Delta G(N) = -N|\Delta\mu| + \gamma A$$

- When **N > N_C** the **bulk free energy** term dominates and the seed **grows**

Espinosa, J. R., et al., Journal of Chemical Physics (2016), 144, 3, 034501

Calculating Interfacial Free Energy Using Seeding Method



- When $\mathbf{N \approx N_C}$ we estimate this as the **free energy barrier** where we can estimate the **interfacial free energy**
- In NPT, where 50% of seeds at this size melt
- How do we quantify crystallinity?

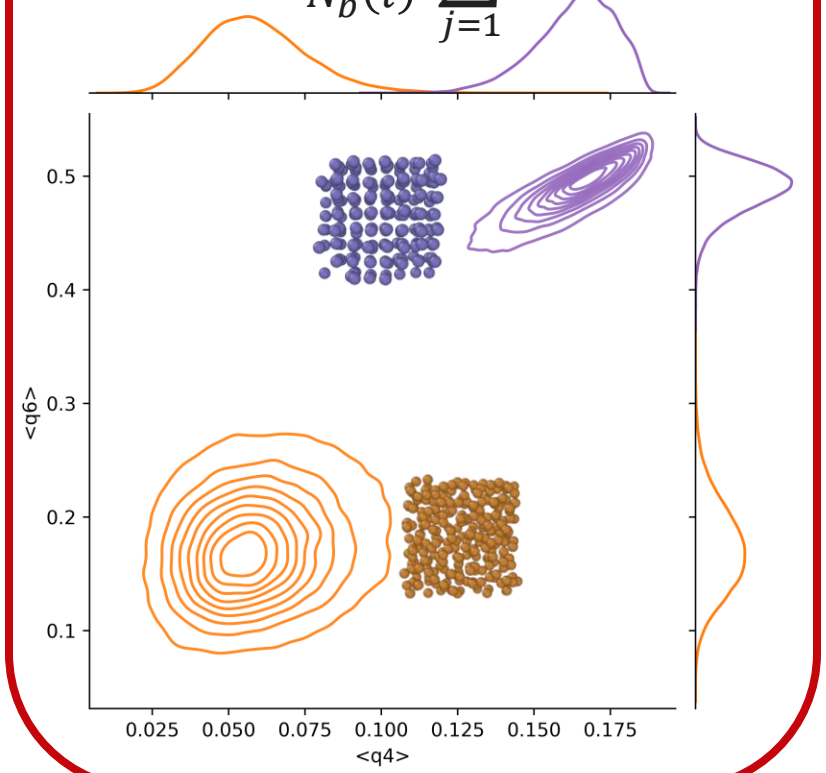
Espinosa, J. R., et al., Journal of Chemical Physics (2016), 144, 3, 034501

Order Parameters to Distinguish and Quantify Crystalline Seed

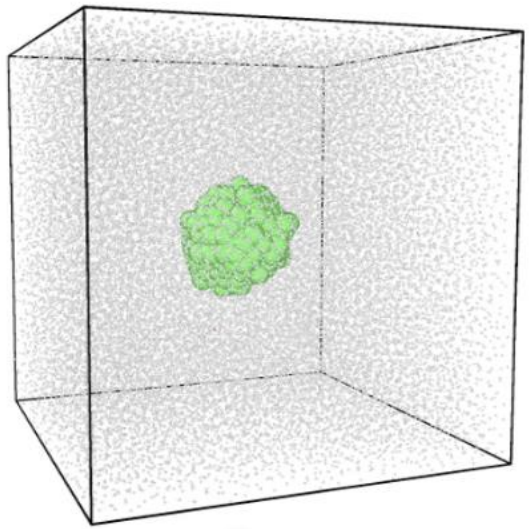
- Existing order parameters are useful for atomic systems, but are not designed for use on molecular systems with complex symmetries
- **Goal:** Design a new order parameter that can distinguish crystalline from liquid environments, allowing us to track molecular crystal growth and determine a critical nucleus size

Steinhardt Order Parameters

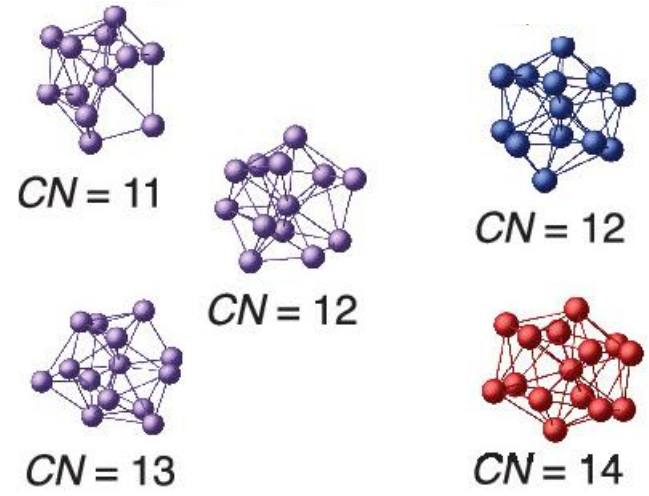
$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\mathbf{r}_{ij}).$$



Polyhedral Template Matching



Local Density/Coordination No.



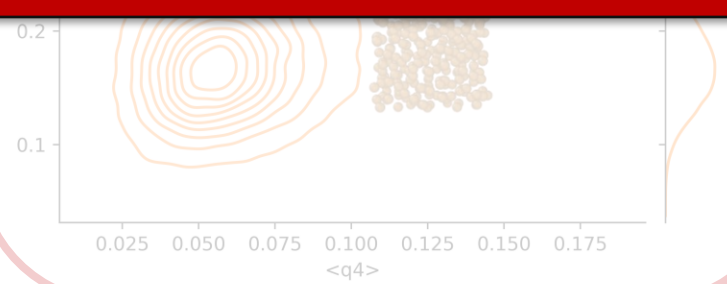
Espinosa, J. R., et al., Journal of Chemical Physics (2016), 144, 3, 034501

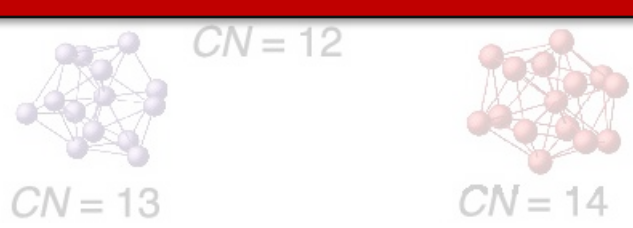
Order Parameters to Distinguish and Quantify Crystalline Seed

- Existing order parameters are useful for atomic systems, but are not designed for use on molecular systems with complex symmetries
- **Goal:** Design a new order parameter that can distinguish crystalline from liquid environments, allowing us to track molecular crystal growth and determine critical nucleus size

Steinhardt Order Parameters

$$\bar{q}_l(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} \bar{Y}_{lm}(\vec{r}_{ij})$$





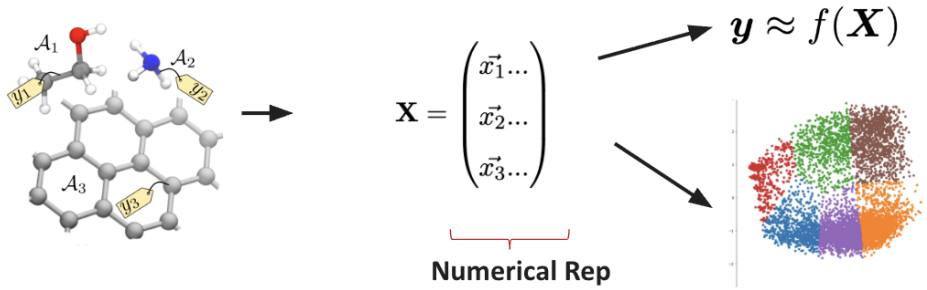
Can we develop robust order parameters that capture the contributions of complex molecular symmetries to distinguish and quantify crystallinity?

Espinosa, J. R., et al., Journal of Chemical Physics (2016), 144, 3, 034501

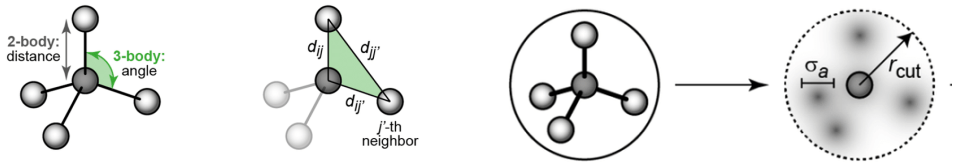
Tools to Create an Automatable Machine Learning Model

Smooth Overlap of Atomic Positions (SOAP)

- To distinguish phases, we need a highly expressive representation



- 3-Body SOAP Descriptors

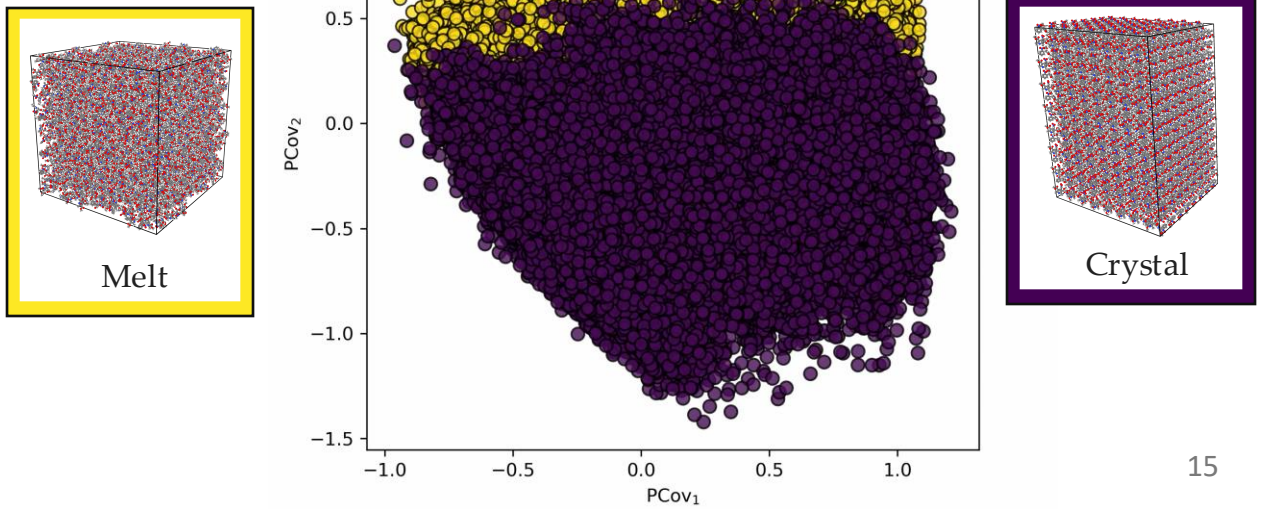


- Want a representation is invariant to translations, rotations and permutations
- Encodes all 2-body distances, 3-body angles, and atom types

Principal Covariates Classification (PCovC)

- To understand the descriptor space, we need a low-dimensional data representation
- Supervised dimensionality reduction for classification

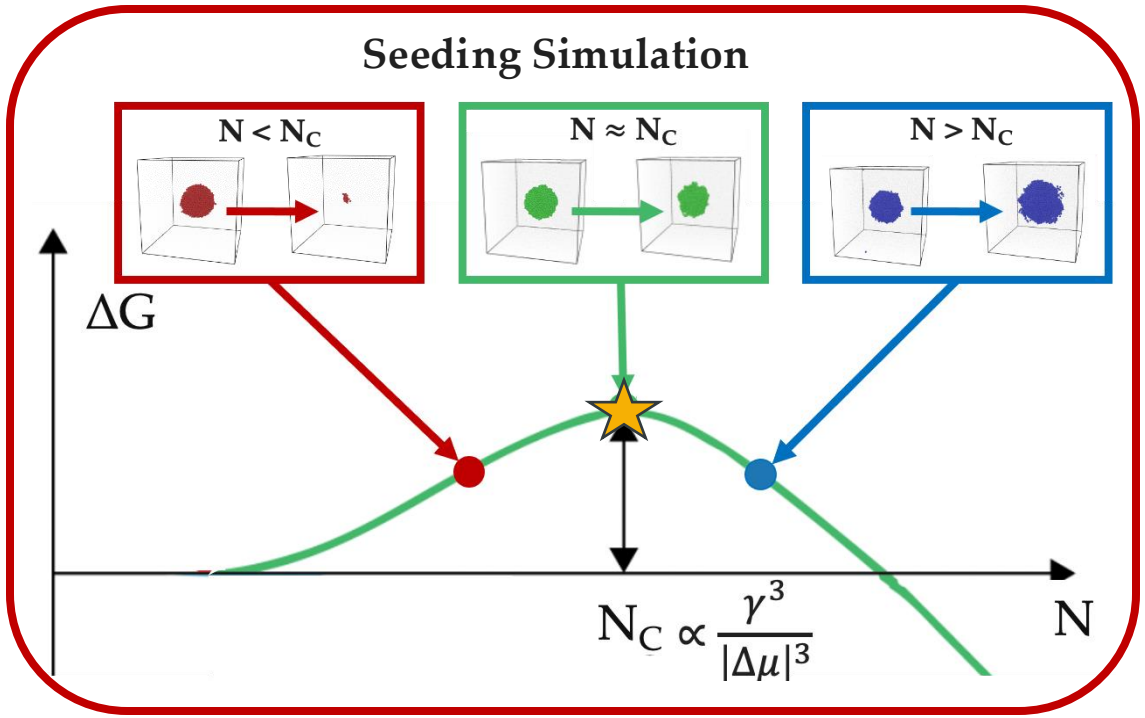
$$loss = \alpha[\text{loss in reconstructing features } (X)] + (1-\alpha)[\text{loss in prediction } (y)]$$



Deringer, V. L., & Barók, A. P, et al., Chemical Reviews (2021), 121, 16

Jorgensen, C. A., et al., ML Sci. and Tech. (2026), 7, 2, 023008

High Throughput Molecular Dynamics Seeding Simulations

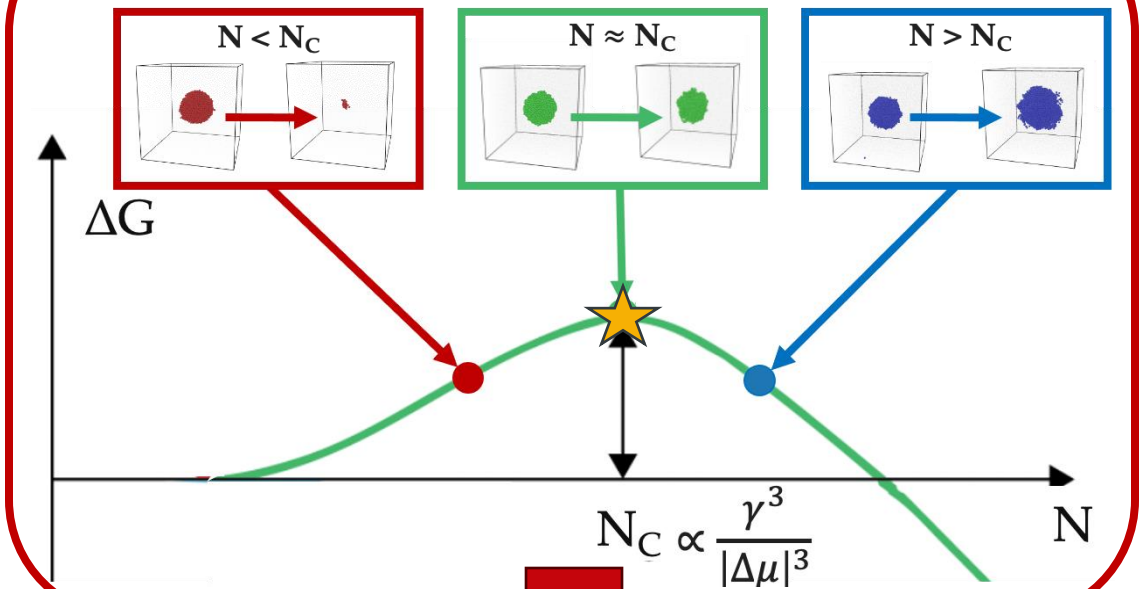


OpenMM

GPU Compatible
Simulation fully written in python

Integrating Machine Learning Techniques for In-Line Simulation Analysis

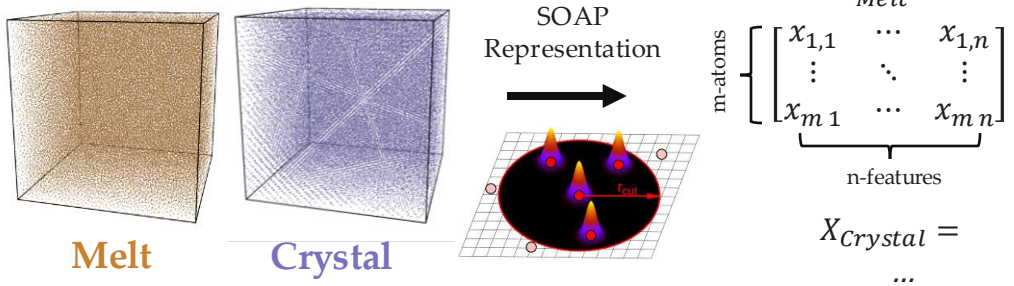
Seeding Simulation



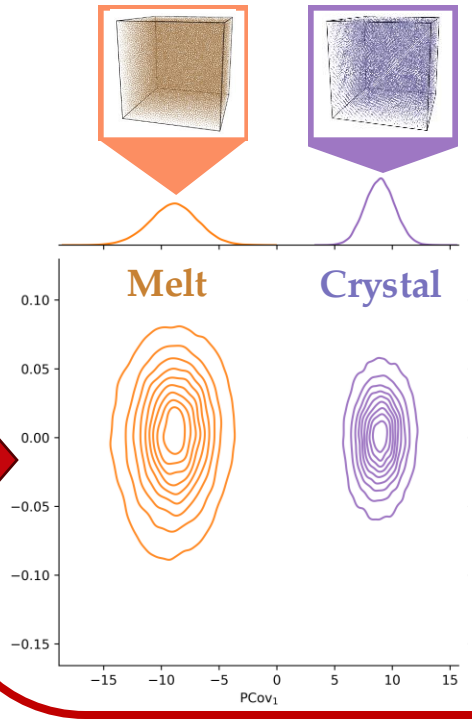
scikit matter

Soon to be GPU compatible!

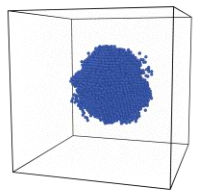
SOAP Calculation on Bulk Crystal & Melt



Project Seeding Simulation Frames into the Latent Space

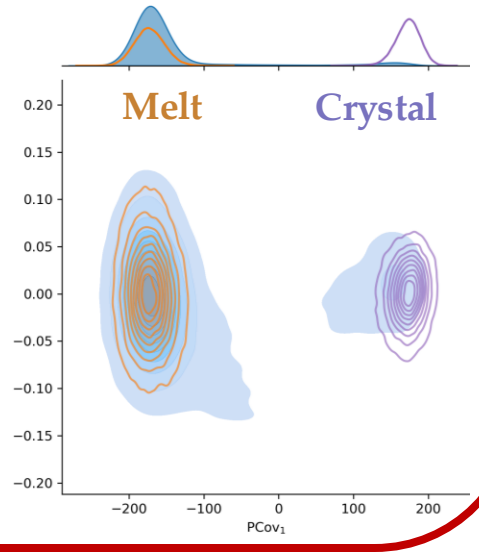


SOAP Representation & Projection



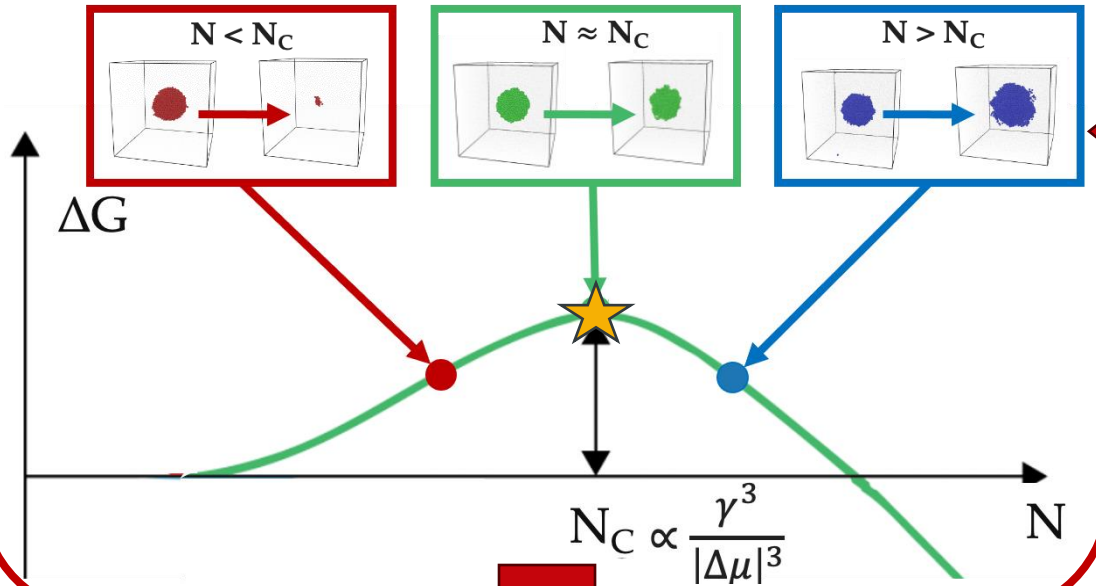
Is Seed Stable?

$N \approx N_C$



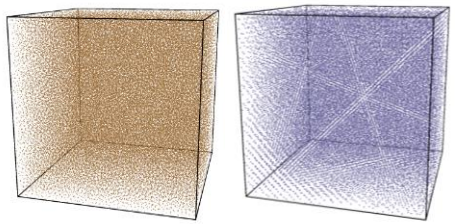
Integrating the Seeding Method with Low-Dimensional Order Parameters to Determine Critical Seed Radius

Seeding Simulation

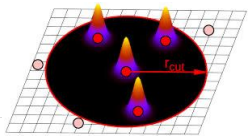


New State Point: T, N

SOAP Calculation on Bulk Crystal & Melt



SOAP Representation



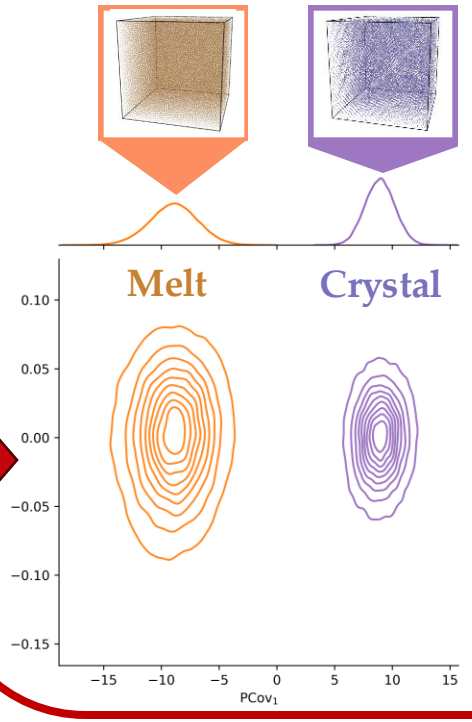
$$X_{Melt} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m\,1} & \cdots & x_{m\,n} \end{bmatrix}$$

m-atoms

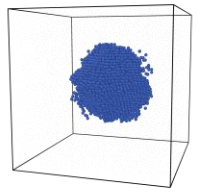
n-features

$$X_{Crystal} = \ldots$$

Project Seeding Simulation Frames into the Latent Space

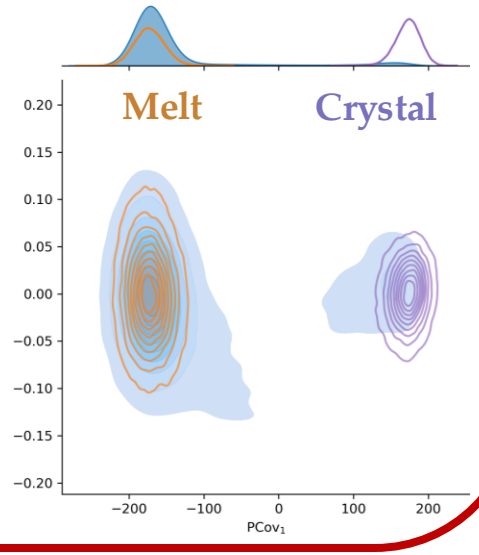


SOAP Representation & Projection



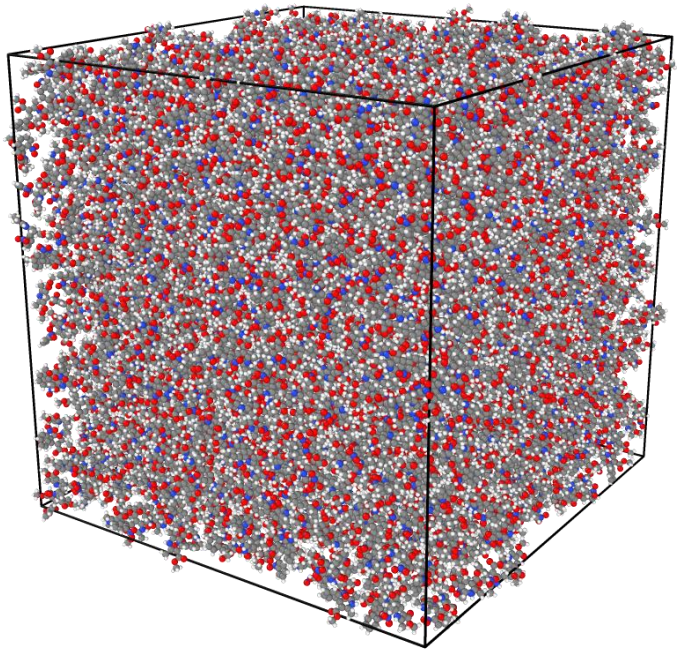
Is Seed Stable?

$N \approx N_C$

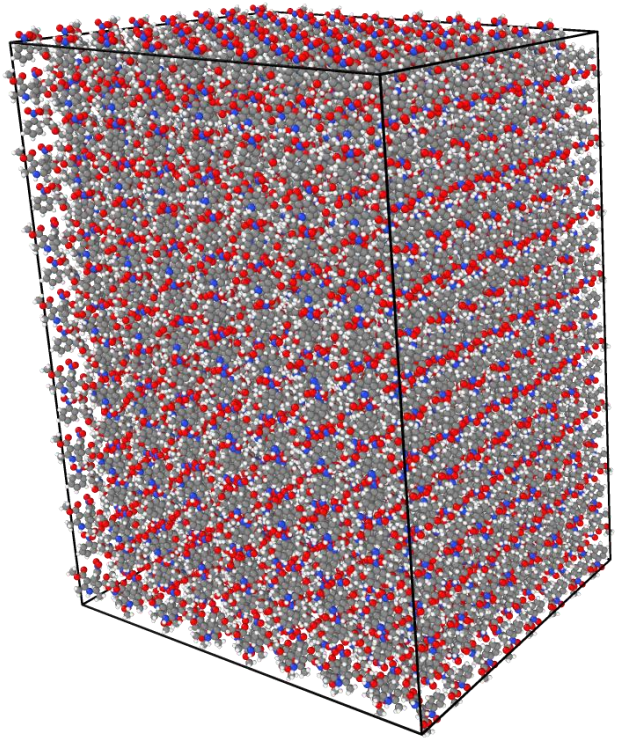


Nifedipine: SOAP-PCovC Principal Covariate Axes

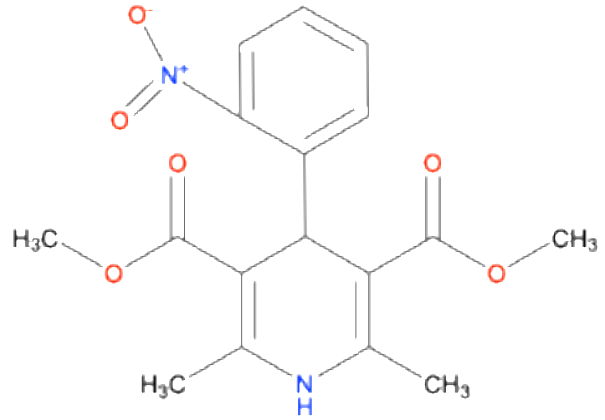
Liquid Melt



Nifedipine α-Polymorph

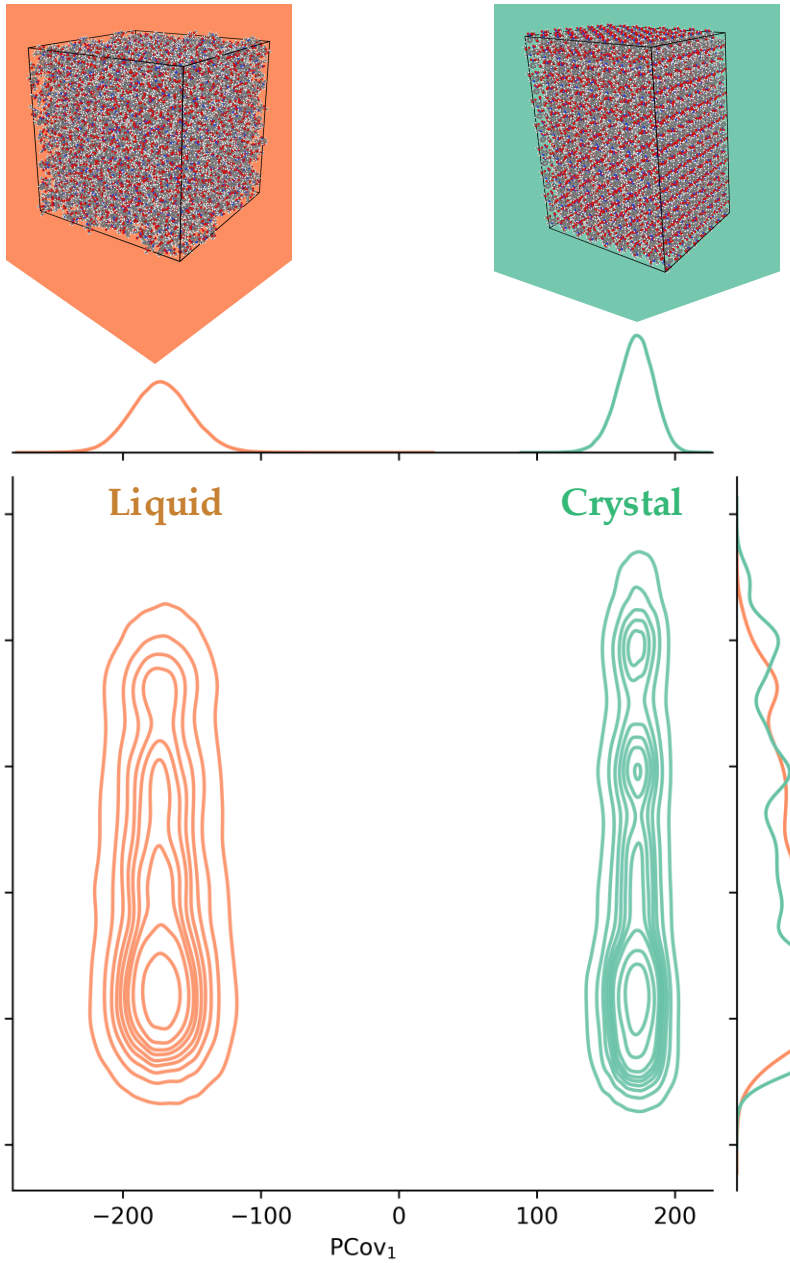


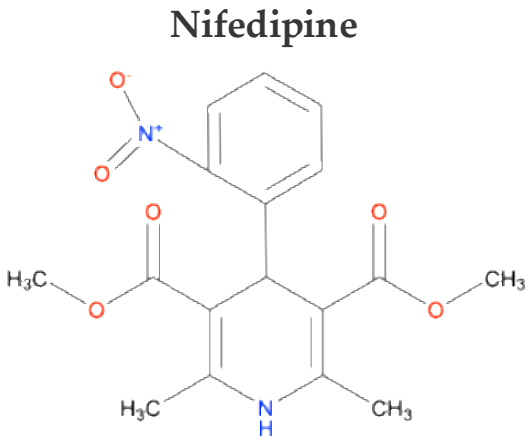
Nifedipine Structure



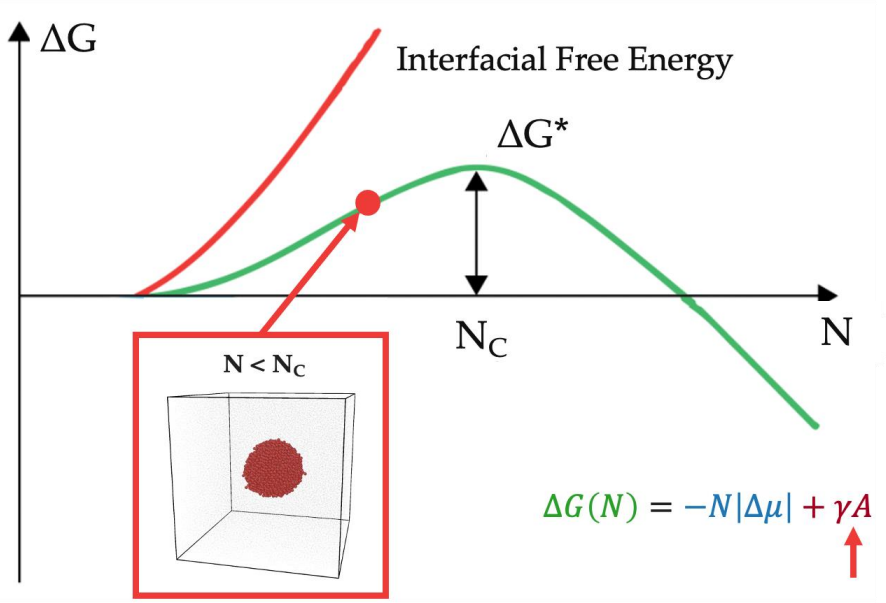
To train the PCovC model, we input all atom SOAP descriptors of the equilibrated solid and crystal

Nifedipine: SOAP-PCovC Principal Covariate Axes





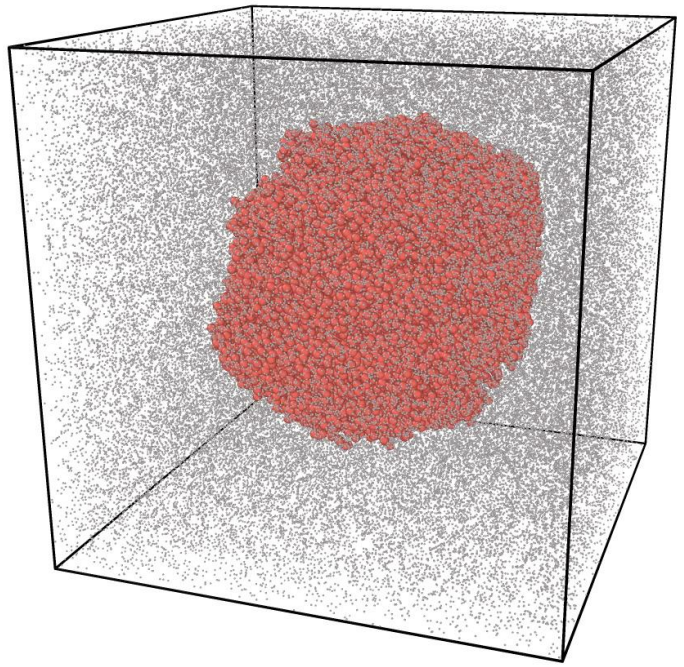
- In model training, we can distinguish crystalline from liquid using the first principal covariate ($PCov_1$)
- We calculate SOAP descriptor with all atoms within an 8Å cutoff radius



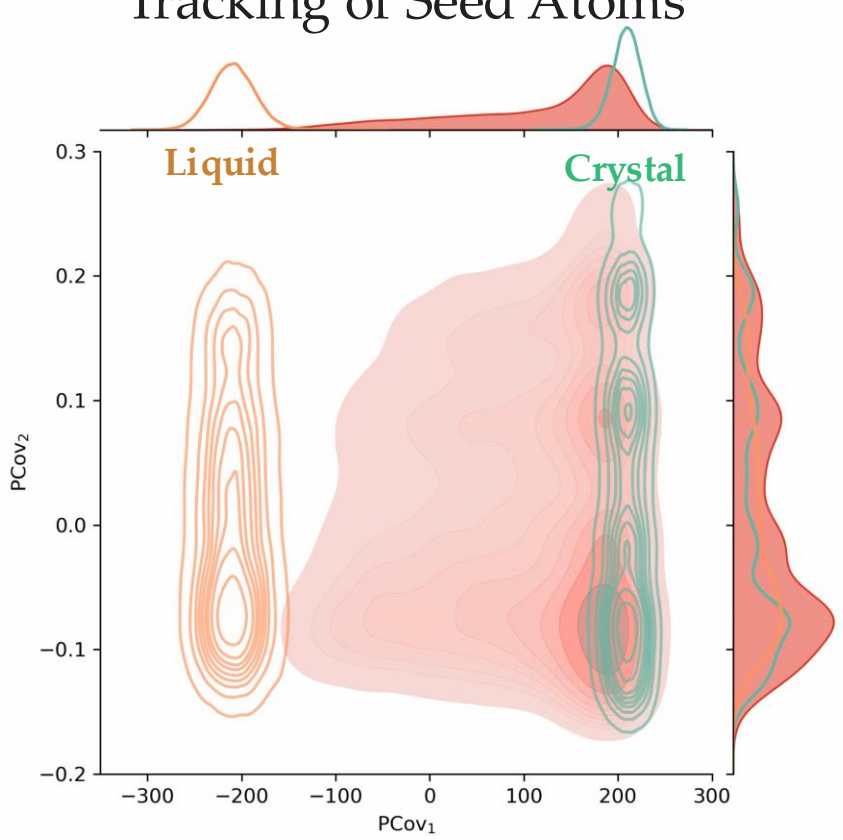
- When $N < N_C$ the **interfacial free energy** term dominates and the seed **shrinks**

Nifedipine: Tracking Crystalline Seed as Seeding Simulation Progresses using SOAP-PCovC Order Parameter

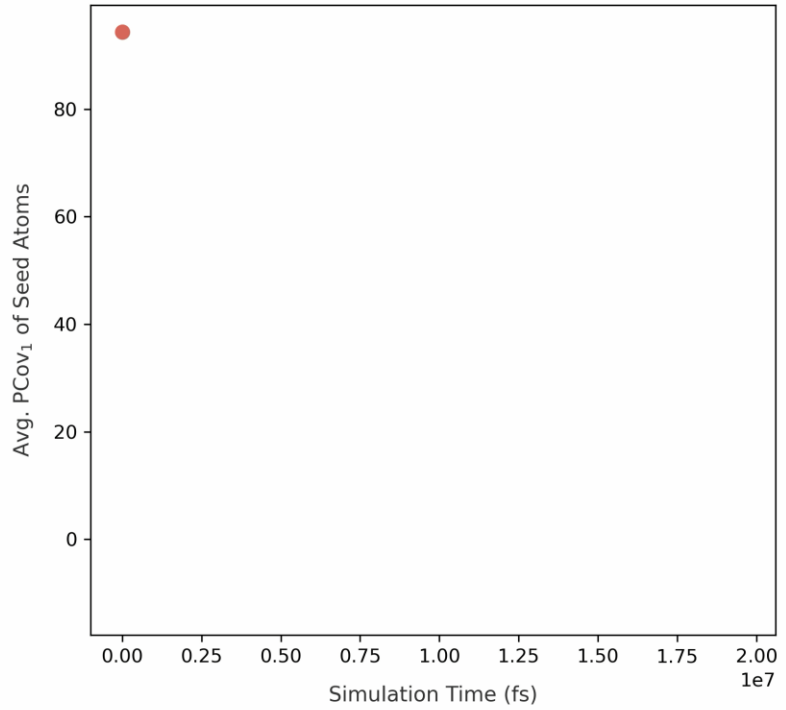
Nifedipine Seeding Simulation Cell



SOAP-PCovC Latent Space Tracking of Seed Atoms

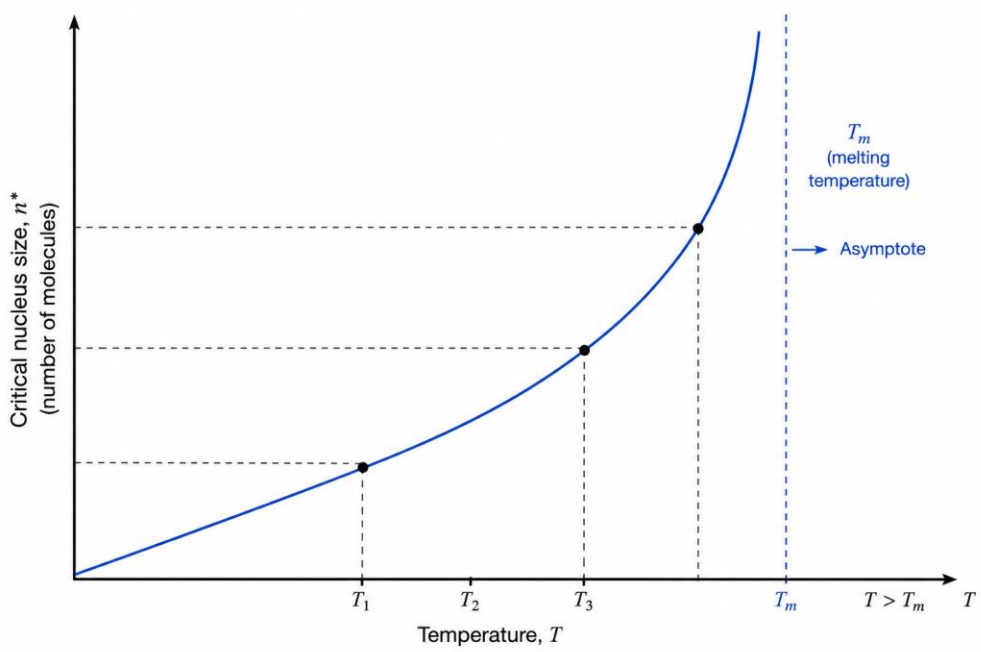


$PCov_1$ Distribution Tracking of Seed Atoms



- Using a modified workflow, we can track molecular crystal (nifedipine) seed melting using the first principal covariate

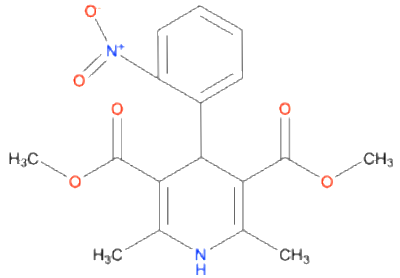
Understanding the Design Space Necessary to Determine Critical Nuclei, and Rank Polymorph Stability



We want to find points along this critical nucleus size vs. temperature curve for each crystal form so we can identify differences in stability

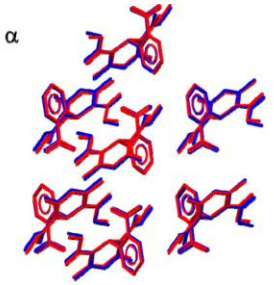
Molecule (CCDC_ID)

Nifedipine

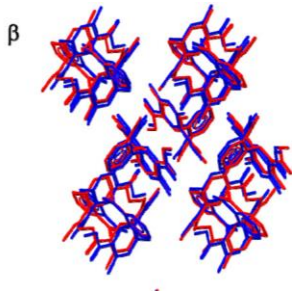


Crystal Form (Sub_No)

Alpha Polymorph



Beta Polymorph



Seed Size (N_C)

Small Seeds + Large Seeds must be tested

e.g., 25 molecules – 500 molecules

Temperature (T)

Low T : >Effective Melting T (T_m)

e.g., 350K – 550K, testing 10K intervals

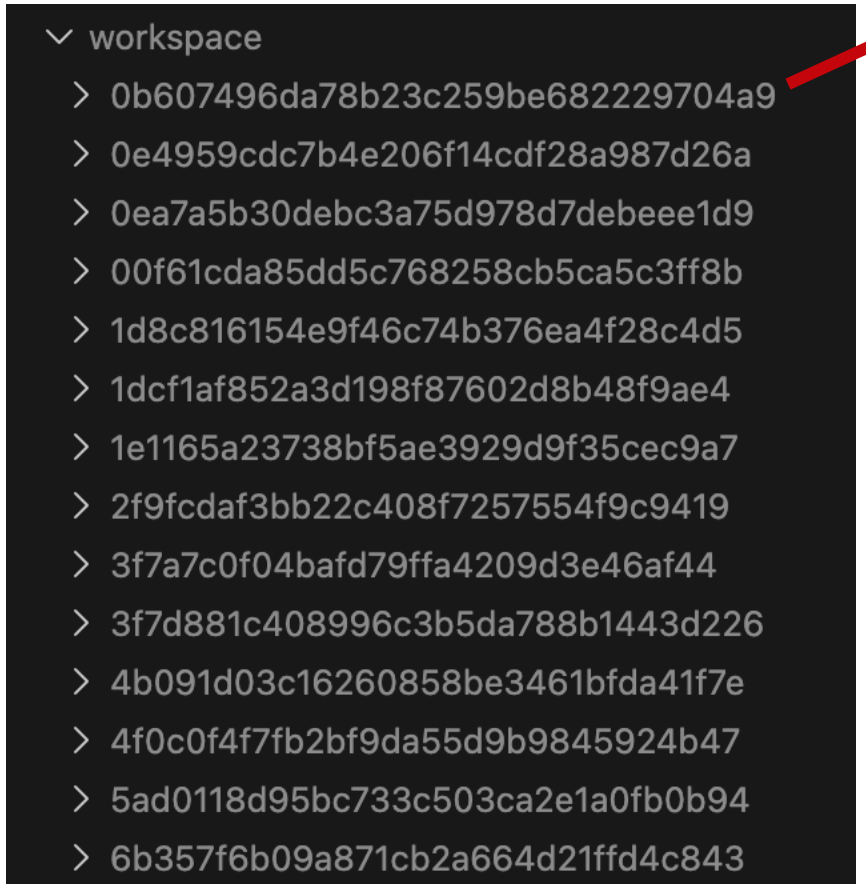
Minimum of 320 jobs (without replicates) with this proposed setup!

More jobs required to resolve true critical nucleus size at a given T

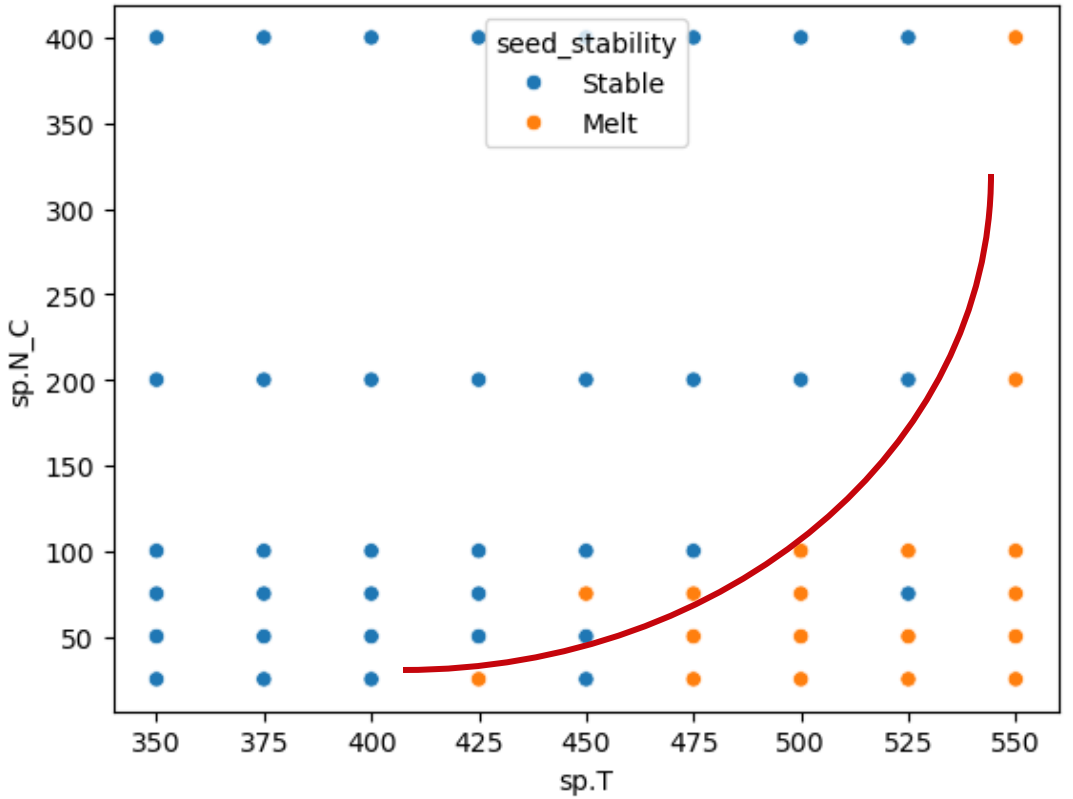
Generating an Initial Dataset Using Signac Package to Submit Jobs

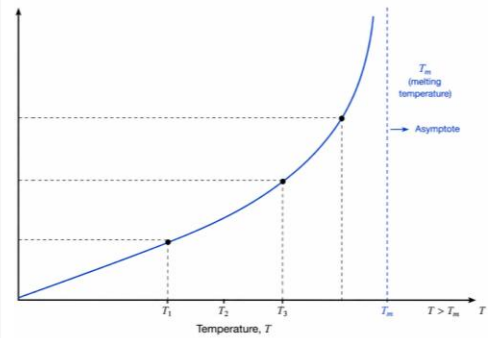
Example job state-point

{"CCDC_ID": "BICCIZ", "sub_No": "02", "N_T": 2500, "N_C": 500, "T": 530, "n_replicates": 1, "forcefield_string": "GAFF-RESP", "ensemble": "NVT", "dt": 1, "total_timesteps": 10000000}

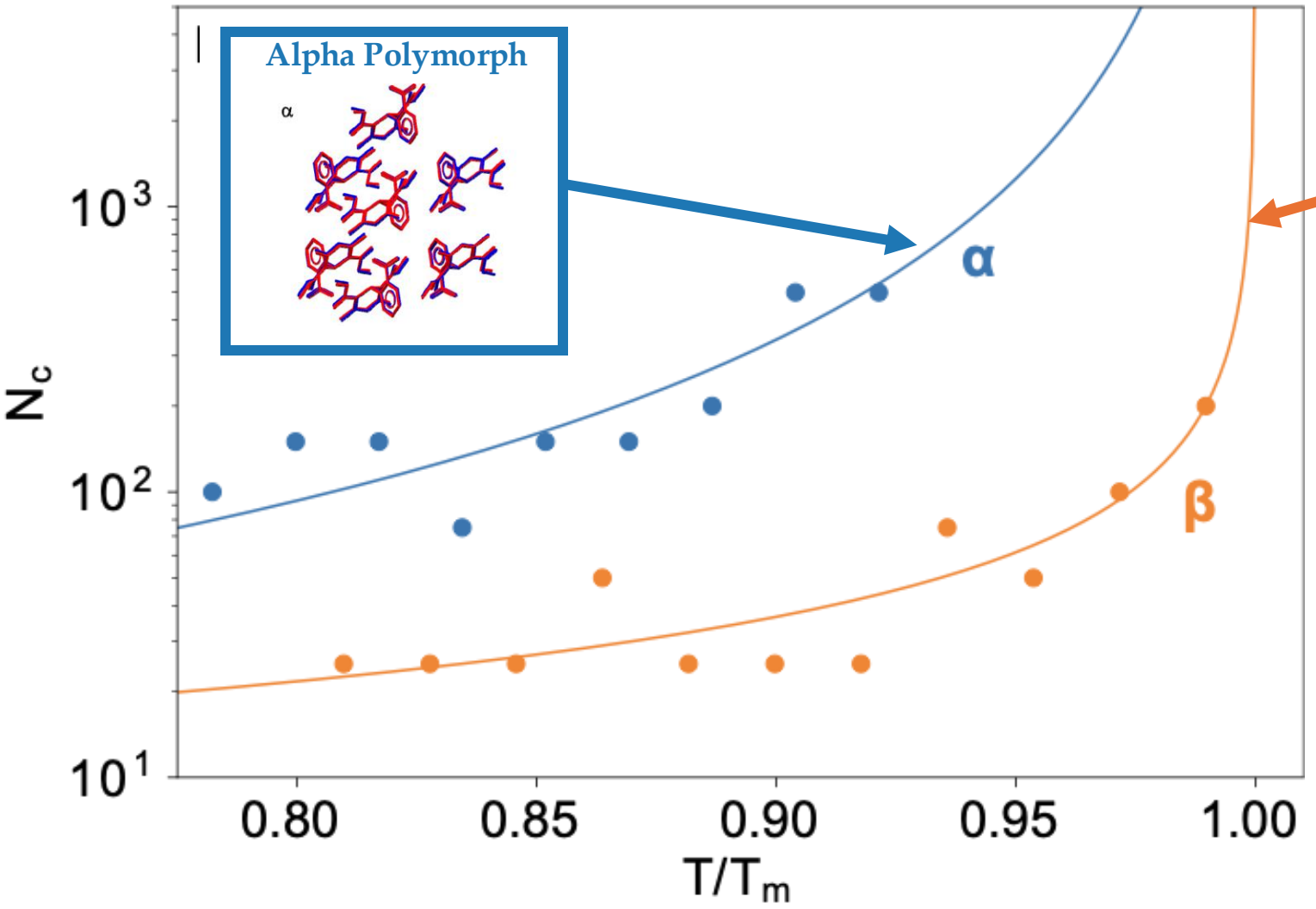


Alpha Polymorph





Nifedipine: Mapping Critical Seed Nucleus Size



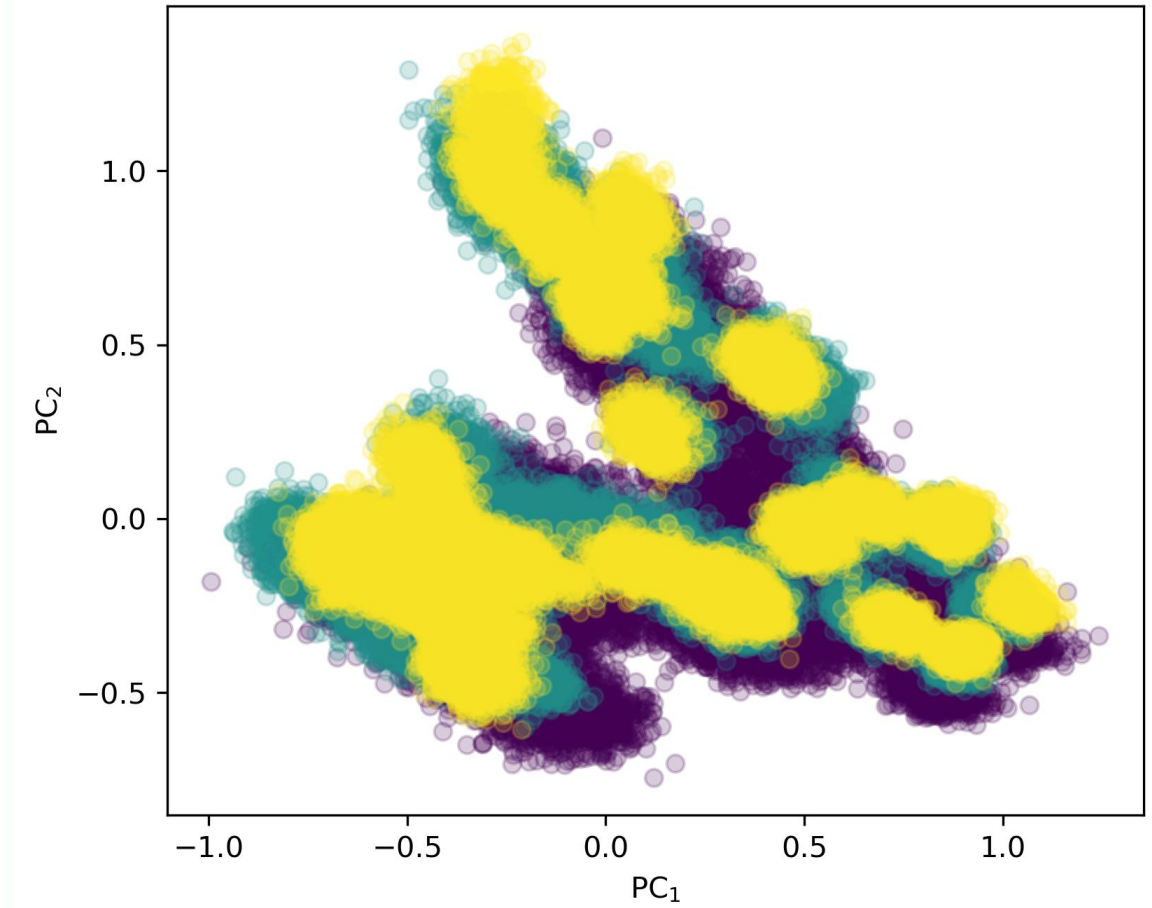
- Preliminary results show seed stability differences between α and β forms
- In experiment β forms from the supercooled melt
- α has the lowest lattice energy, and would be the predicted structure using traditional CSP methods

Gunn, E, et at., Crystal Growth and Design (2012), 12, 4, 2037-2043

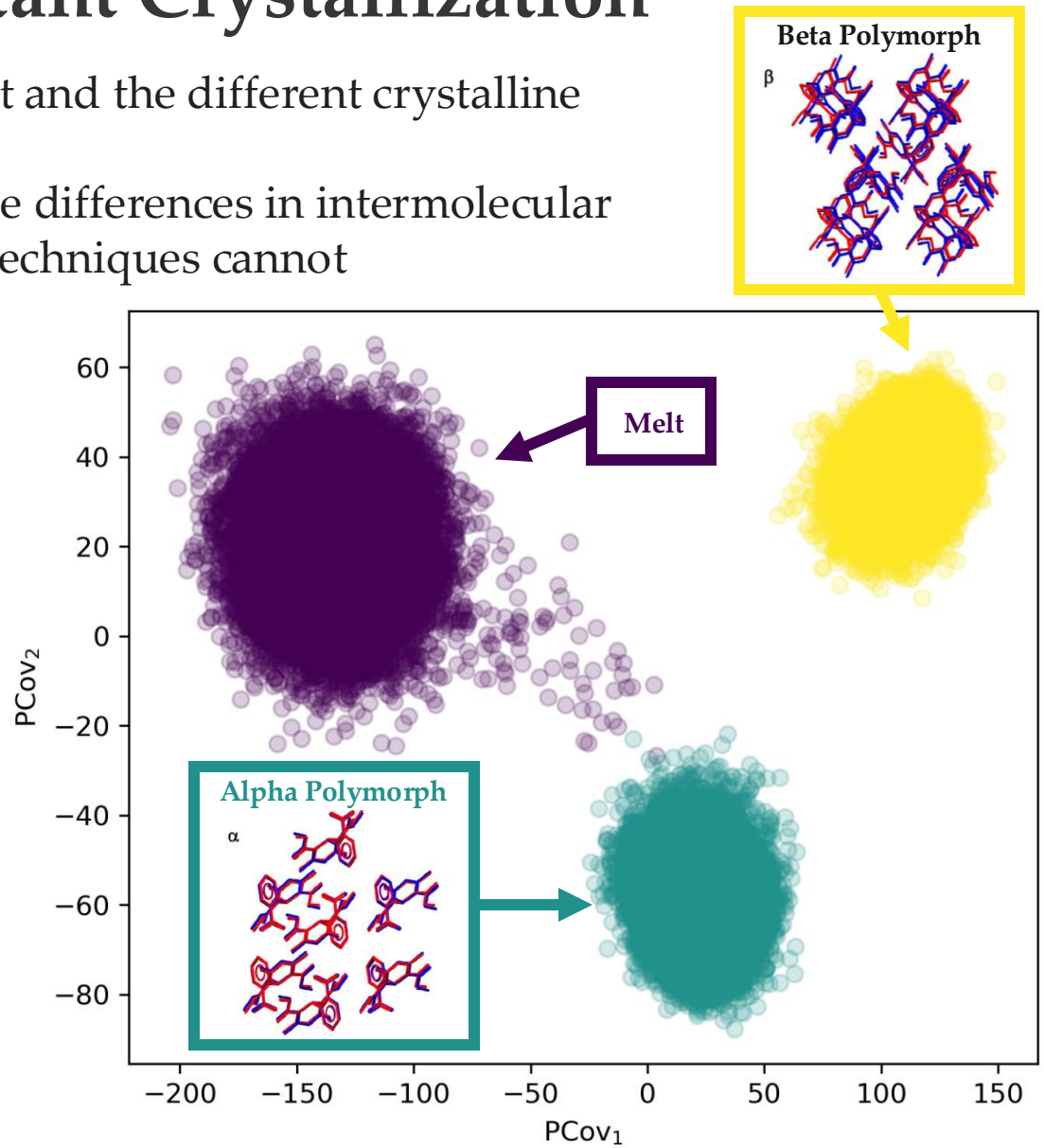
Gui, Y., et al., Crystal Growth and Design (2022), 22, 2, 1365-1370

Future Work: Tracking Concomitant Crystallization

- We can train our PCovC on the SOAPs of the melt and the different crystalline structures of nifedipine
- SOAP-PCovC order parameter can distinguish the differences in intermolecular interactions that other dimensionality reduction techniques cannot



SOAP-PCA



SOAP-PCovC

Gui, Y., et al., Journal American Chemical Society (2022), 22, 2, 1365-1370

Acknowledgments

Censonsky Lab of Multiscale Matter

Censonsky Group

- **Prof. Rose Censonsky**
- Charles Carroll
- Max Depperschmidt
- Ethan Deutsch
- **Christian Jorgensen**
- Hwigwang Lim
- Arthur Lin
- Caleb Youngwerth
- James Yeom

Van Lehn Group

- Prof. Reid Van Lehn
- Dr. Juriti Rajbangshi
- Ali Altamimi
- Srija Chakraborty
- Avery Greene
- Yuhui Huang
- Ugo Ikegwu
- Changsu Kim
- Jung Min Lee
- Yicheng Li
- ByungUk Park
- Letian Wang

CENTER FOR HIGH THROUGHPUT COMPUTING

WARF Wisconsin Alumni Research Foundation

ACCESS


WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Aside: GPU simulations can be set up by modifying the pixi.toml and apptainer image definition file.

pixi.toml

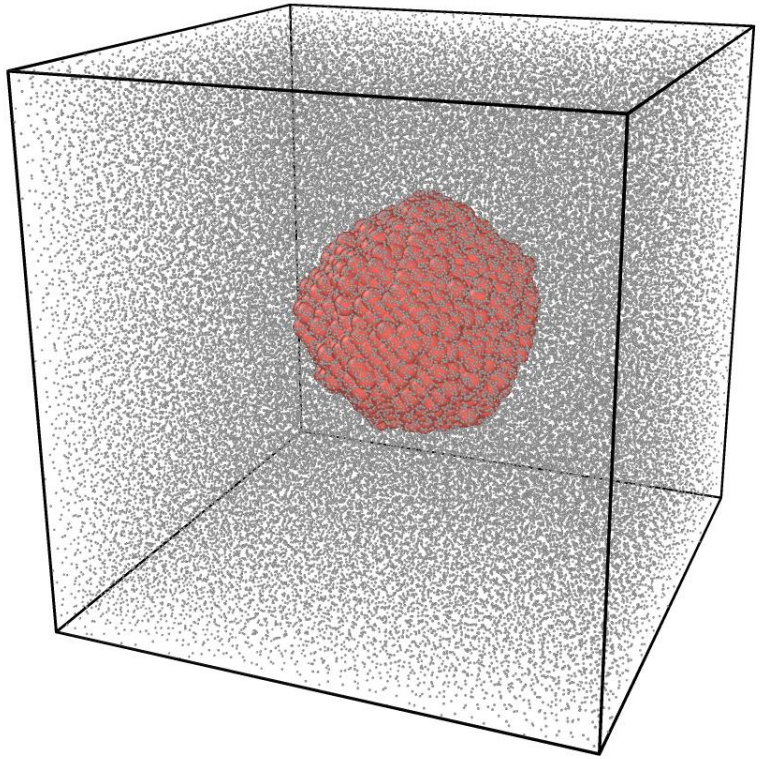
You, 25 seconds ago | 2 authors (You and one other)

```
[workspace]
channels = ["https://conda.ovito.org", "conda-forge"]
name = "Seeding"
platforms = ["linux-64", "osx-arm64"]
version = "0.1.0"

[system-requirements]
cuda = "12"

[dependencies]

openmm = "*"
openmm-ml = "*"
mdtraj = ">=1.10.3,<2"
openff-toolkit = ">=0.16.10,<0.17"
nnpops = ">=0.6,<0.7"
openmmforcefields = ">=0.14.2,<0.15"
packmol = ">=21.1.2,<22"
openff-interchange = ">=0.4.2,<0.5"
nglview = ">=3.1.4,<4"
ase = ">=3.27.0,<4"
openmm-setup = ">=1.6,<2"
mdanalysis = ">=2.9.0,<3"
signac = ">=2.3.0,<3"
signac-flow = ">=0.29.0,<0.30"
python = ">=3.10.19,<3.11"
cuda-version = "12.1.*"          You, 25 seconds ago • Uncom
```

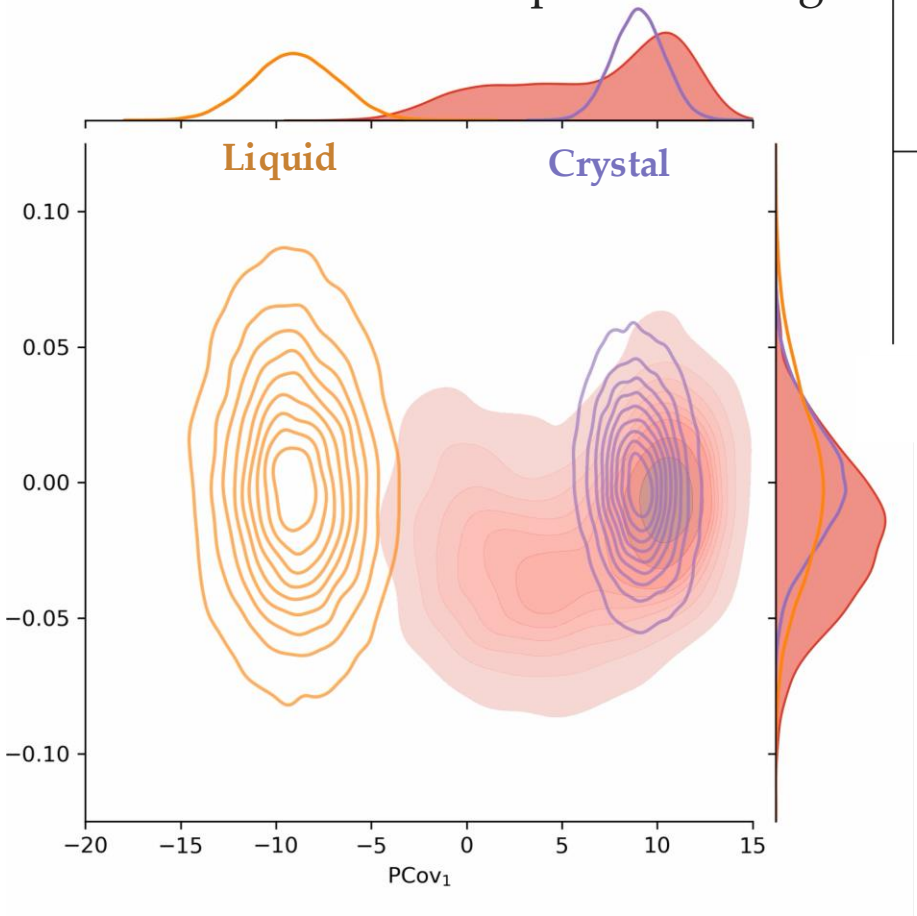
```
%arguments
    CUDA_VERSION=12.0
    ENVIRONMENT=default

%files
./pixi.toml /app/
./pixi.lock /app/

%post
#!/bin/bash
export CONDA_OVERRIDE_CUDA={{ CUDA_VERSION }}
cd /app/
pixi info
pixi install --locked --environment {{ ENVIRONMENT }}
echo "#!/bin/bash" > /app/entrypoint.sh && \
pixi shell-hook --environment {{ ENVIRONMENT }} -s bash >> /app/entrypoint.sh && \
echo 'exec "$@"' >> /app/entrypoint.sh

Bootstrap: docker
From: ghcr.io/prefix-dev/pixi:noble
Stage: final

%arguments
    ENVIRONMENT=default

%files from build
/app/.pixi/envs/{{ ENVIRONMENT }} /app/.pixi/envs/{{ ENVIRONMENT }}
/app/pixi.toml /app/pixi.toml
/app/pixi.lock /app/pixi.lock
```

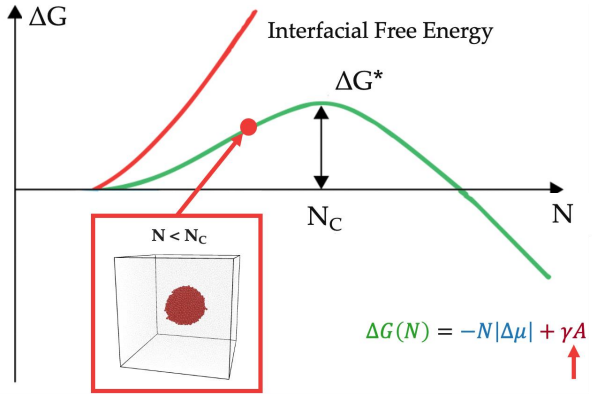
Lennard-Jones: Tracking Crystalline Seed as Seeding Simulation Progresses using SOAP-PCovC Order Parameter

Lennard Jones Seeding Simulation Cell



SOAP-PCovC Latent Space Tracking



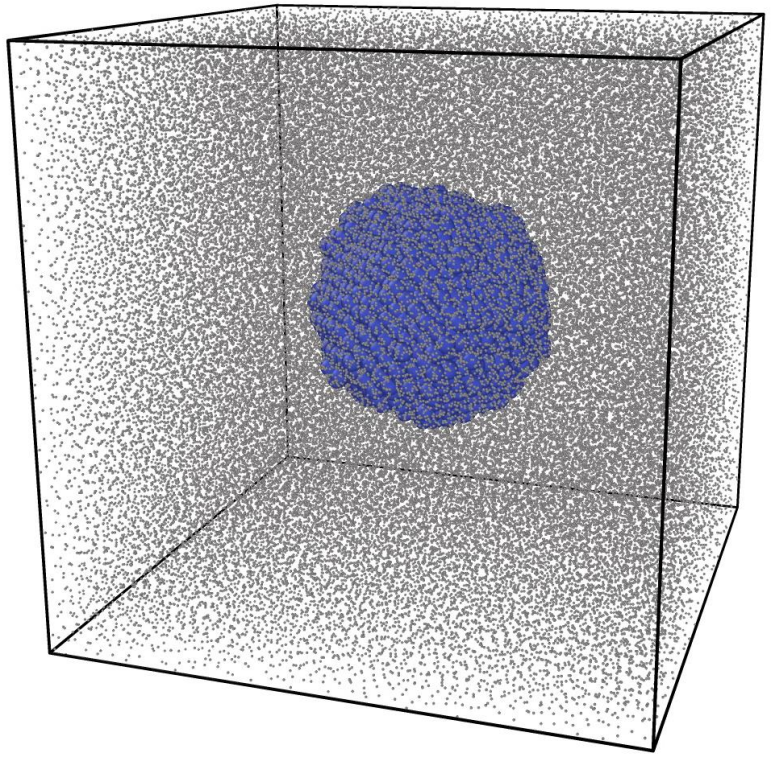


- When $N < N_C$ the **interfacial free energy** term dominates and the seed **shrinks**

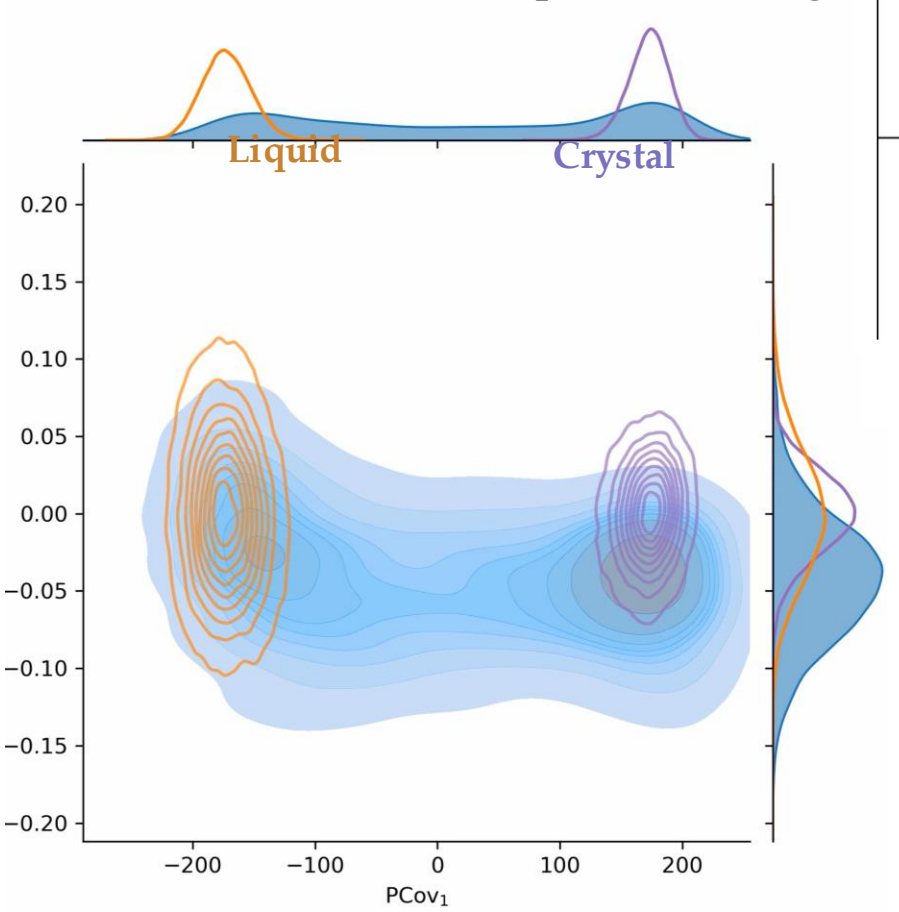
Using seeding simulations and SOAP-PCovC order parameters, we have established a workflow for automated crystal detection

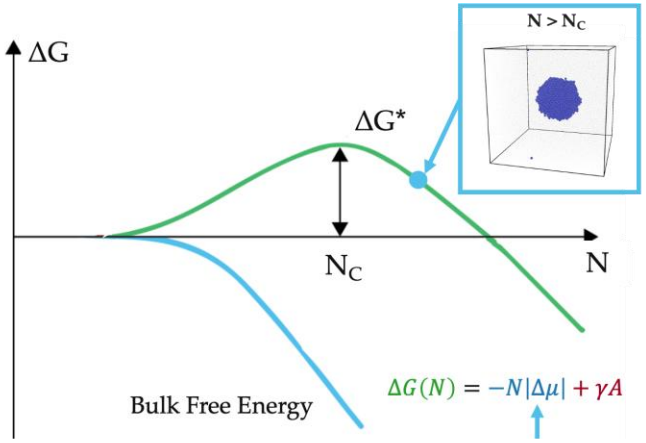
Lennard-Jones: Tracking Crystalline Seed as Seeding Simulation Progresses using SOAP-PCovC Order Parameter

Lennard Jones Seeding Simulation Cell



SOAP-PCovC Latent Space Tracking





- When **N > N_C** the **bulk free energy** term dominates and the seed **grows**

Using seeding simulations and SOAP-PCovC order parameters, we have established a workflow for automated crystal detection