SAND2022-2243C

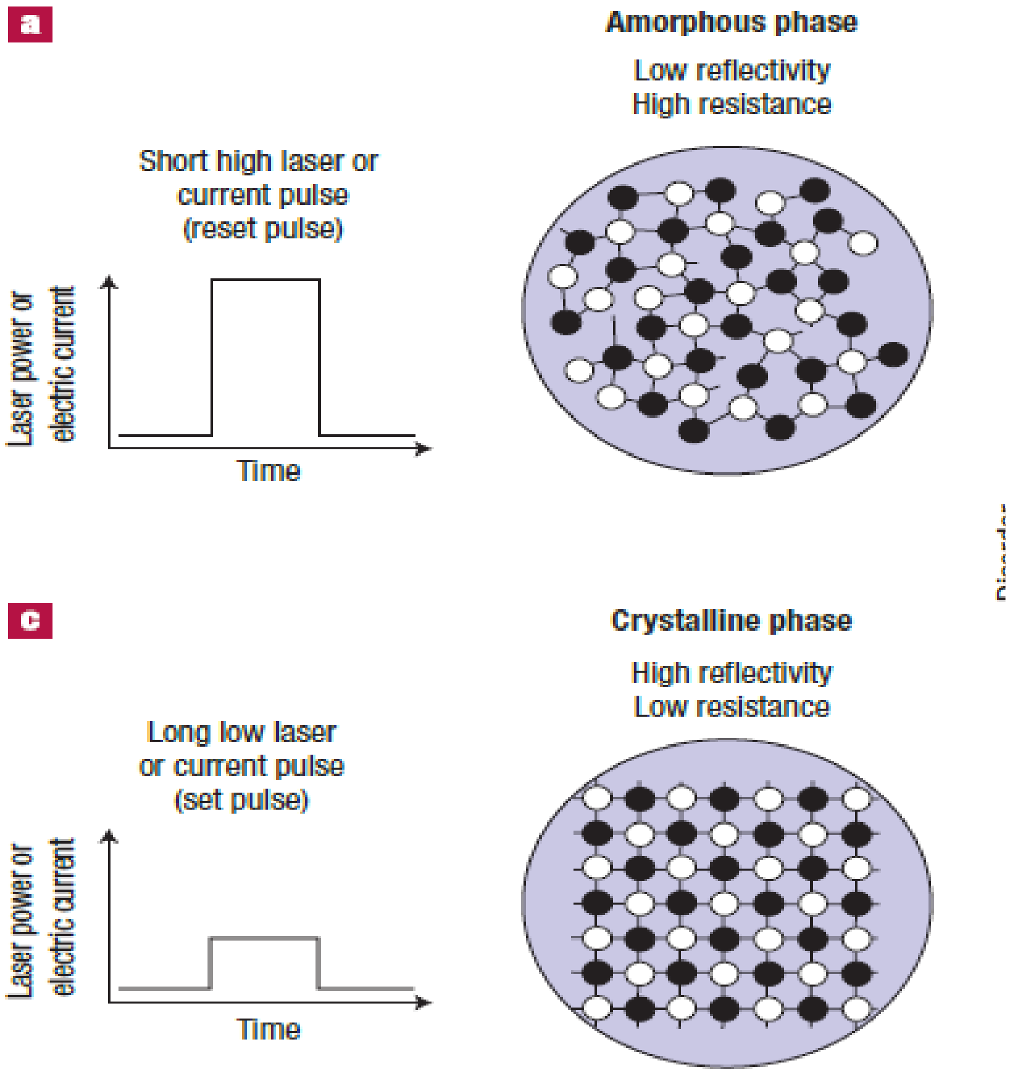
Practical database enrichment strategies with iterative learning: Neural Network Potentials for Phase Change Memory

Zachary McClure*, Robert Appleton, David Adams, Alejandro Strachan
zmcclure@purdue.edu

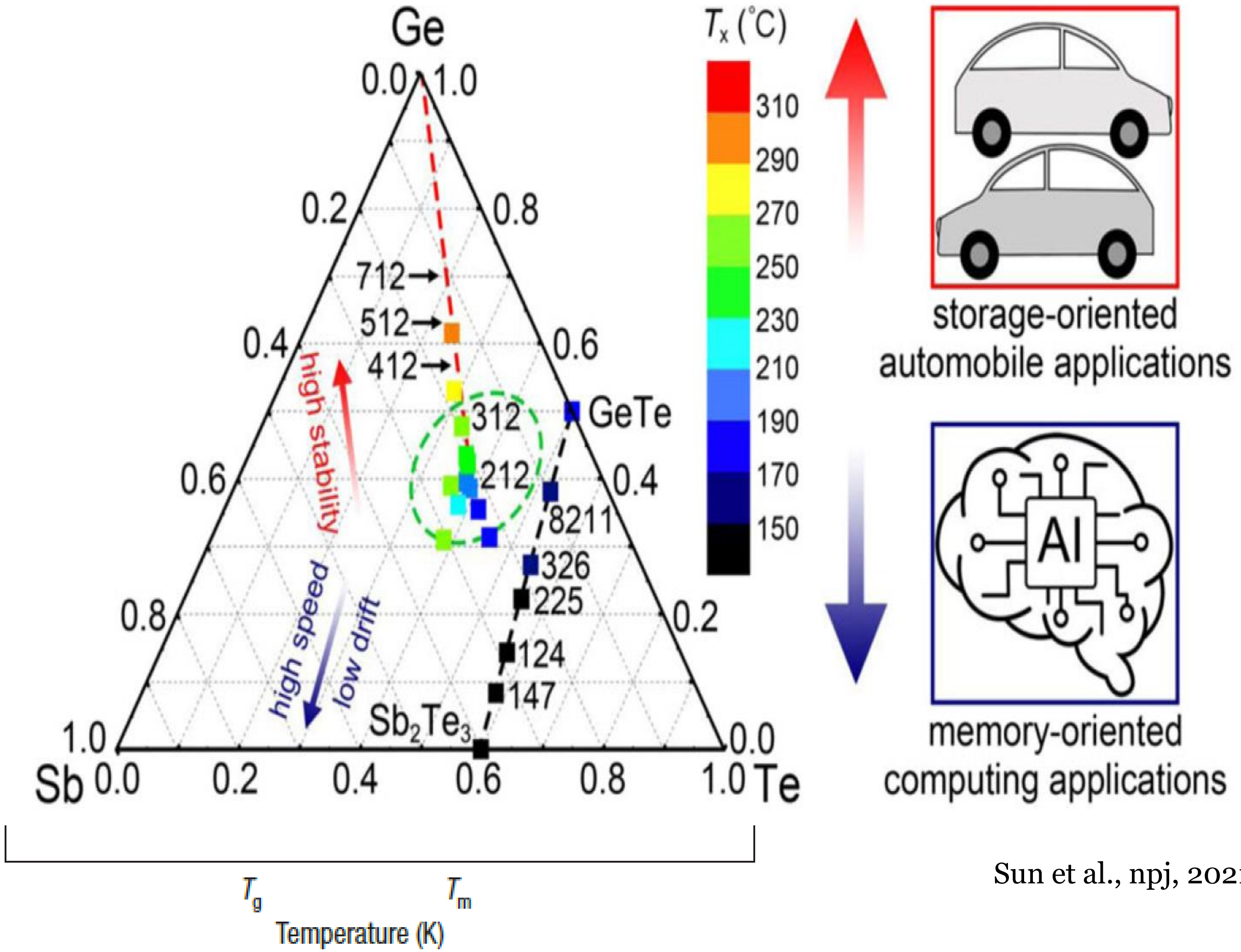
Project Supported by:
https://www.sandia.gov/

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Phase Change Memory: Atomistic precision for macroscale applications



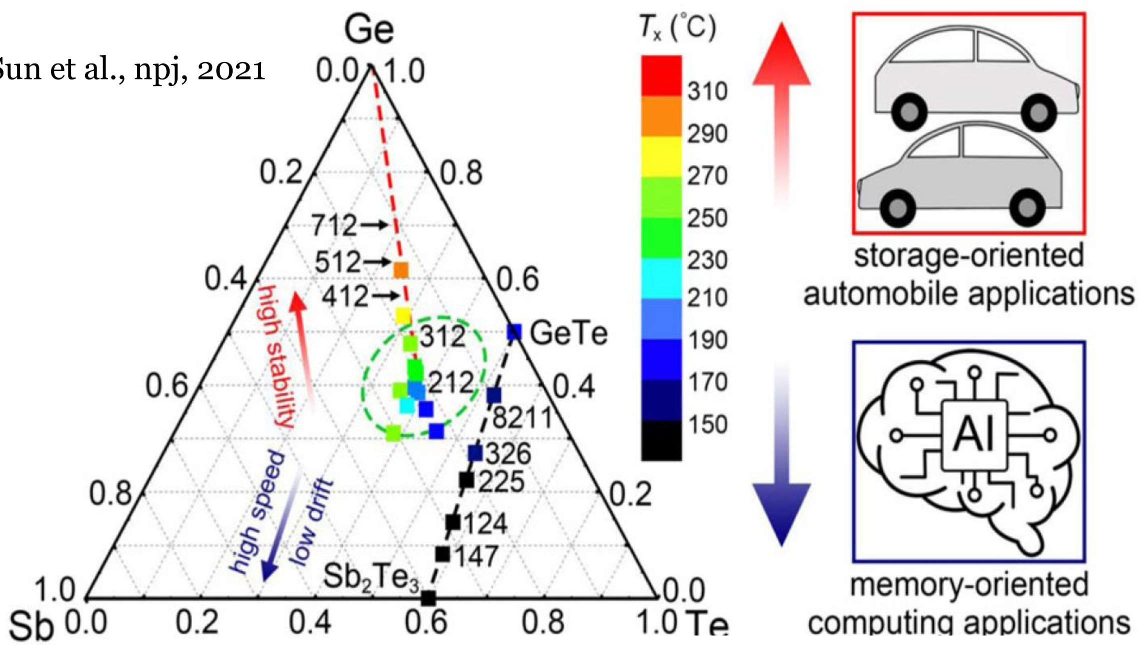
Wuttig and Yamada, Nat Mat, 2009



Sun et al., npj, 2021

Stabilizing the amorphous phase

Sun et al., npj, 2021



Han et al., RSC Advances, 2021

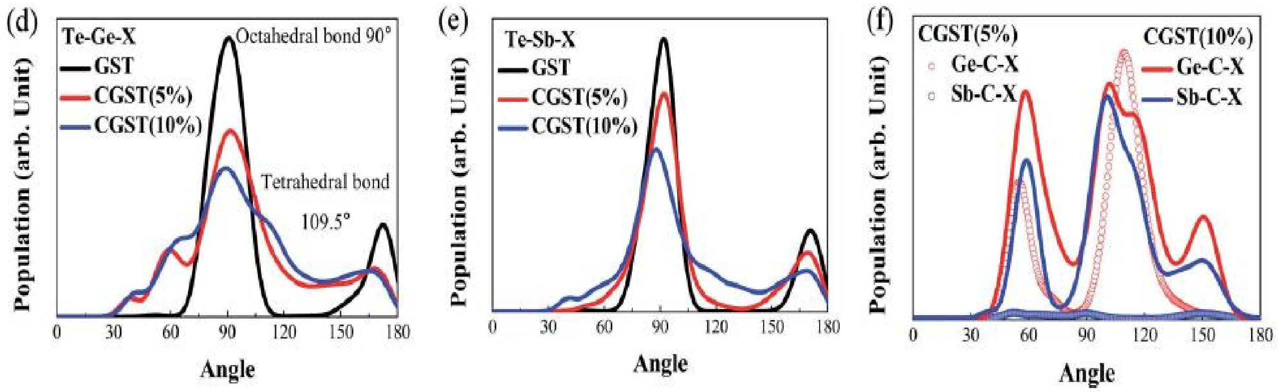
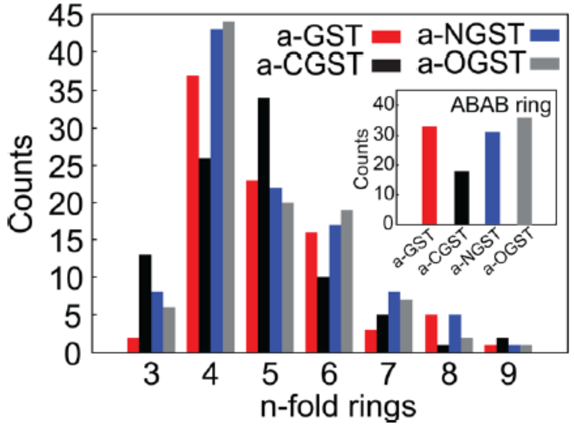


TABLE I. The coordination numbers, bond lengths, the first peak position of ADF around Ge atoms in amorphous GST structures. Ge(I) atoms are bonded to dopants, while Ge(II) atoms are not.

	Coordination numbers		Bond lengths (Å)		Averaged angle around Ge atoms (°)	
	Ge(I)	Ge(II)	Ge(I)-Te	Ge(II)-Te	-Ge(I)-	-Ge(II)-
a-GST	3.67		2.78		93	
a-CGST	4.00	3.67	2.67	2.74	106	98
a-NGST	4.00	4.00	2.73	2.80	102	95
a-OGST	3.74	3.90	2.84	2.81	98	93



Cho et al., APL, 2011

FIG. 2. (Color online) Ring statistics for amorphous GST structures counted per supercell. The inset figure shows the numbers of ABAB-type squared rings.

What atomistic process can we model so far?

<u>*Ab initio* methods:</u>

- Phase change
- Density of states
- Dopant stability
- Recrystallization
- Limited to <u>*constant volume*</u>

**limited to small system sizes

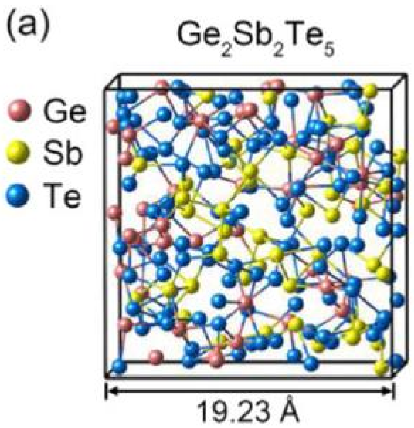
npj | Computational Materials

www.nature.com/npjcompumats

ARTICLE OPEN

Ab initio molecular dynamics and mater
embedded phase-change memory

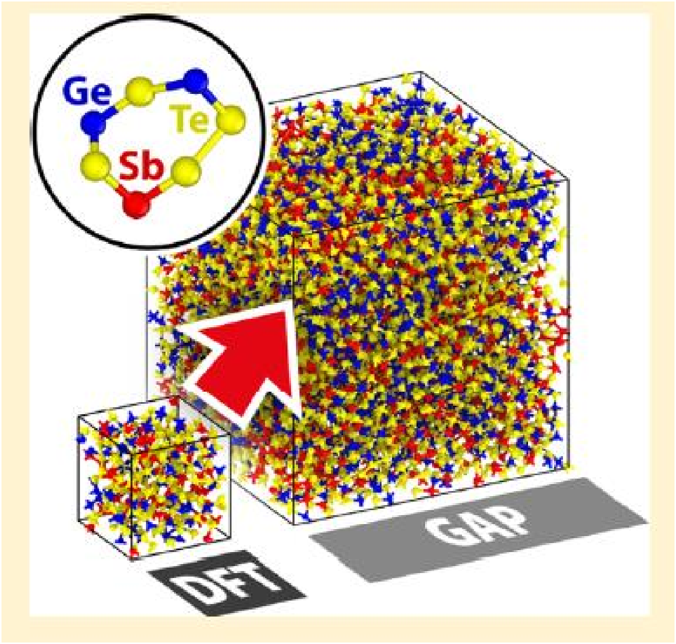
Liang Sun[1,8], Yu-Xing Zhou[2,3,7,8], Xu-Dong Wang[2,3], Yu-Han Chen[2,3], Volker L. Deringer



<u>ML MD Potential:</u>

Gaussian Approximation Potential – DFT accuracy with scaling to ~7K atoms

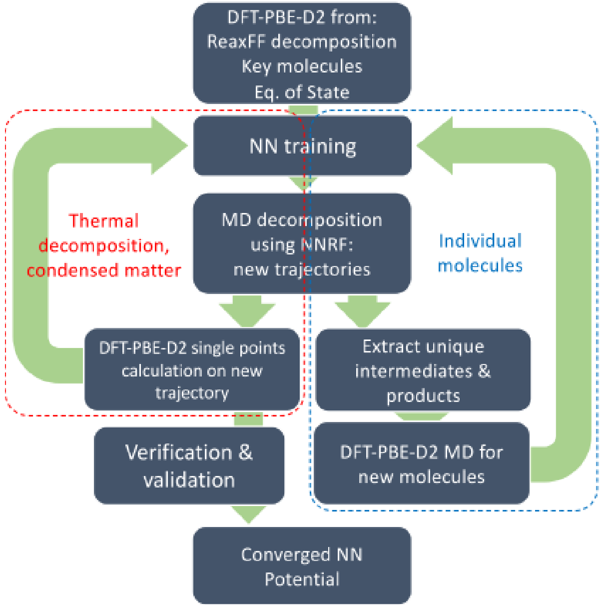
NPT possible for phase change

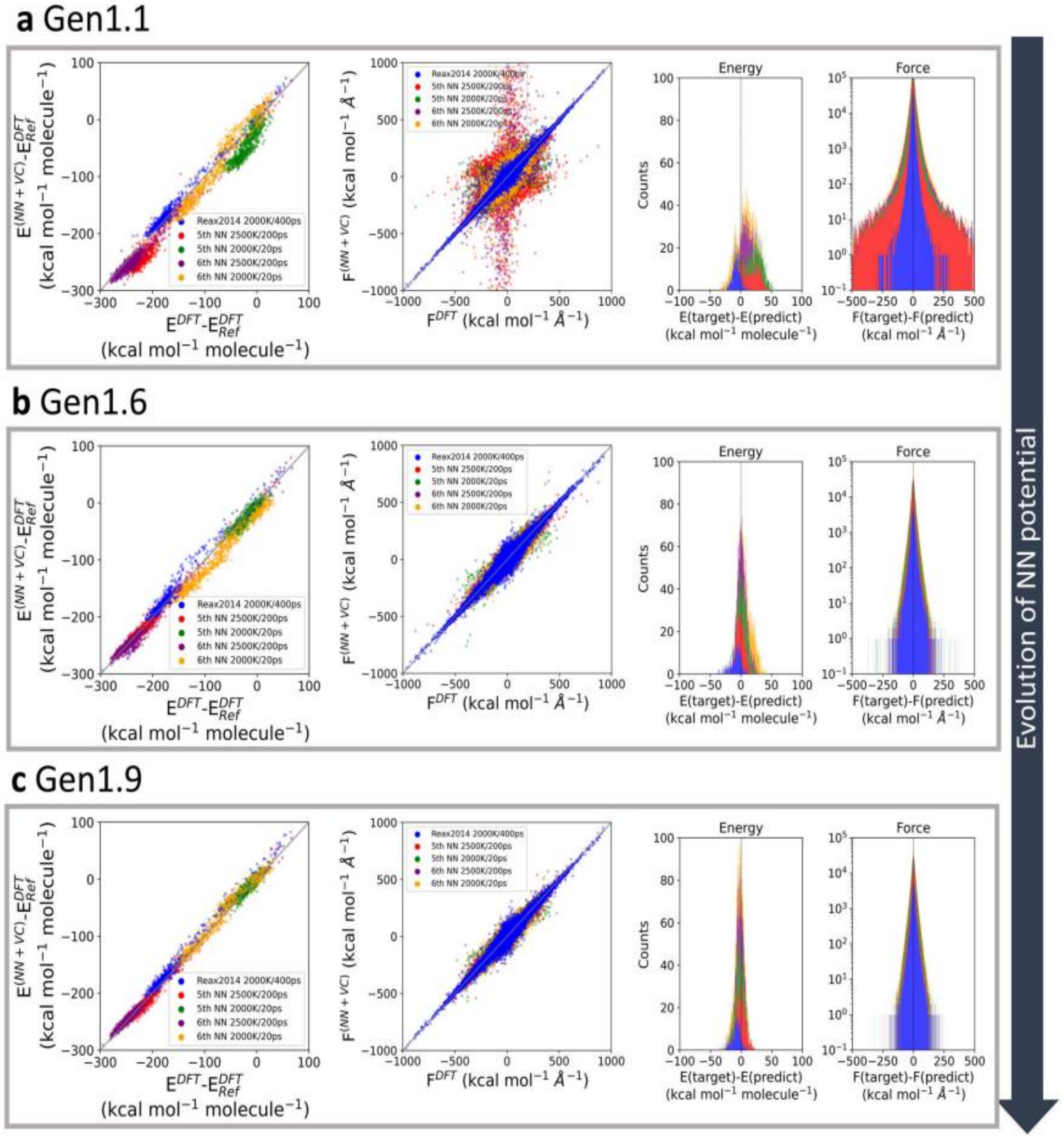


Mocanu et al., JPC B, 2018

Developing machine learning potentials

High density neural net
(HDNNI
- High variability du
hyperparam
- Requires higher vc
- High sensitivity
initialization fc





initialization

HDNNP+Iteration:

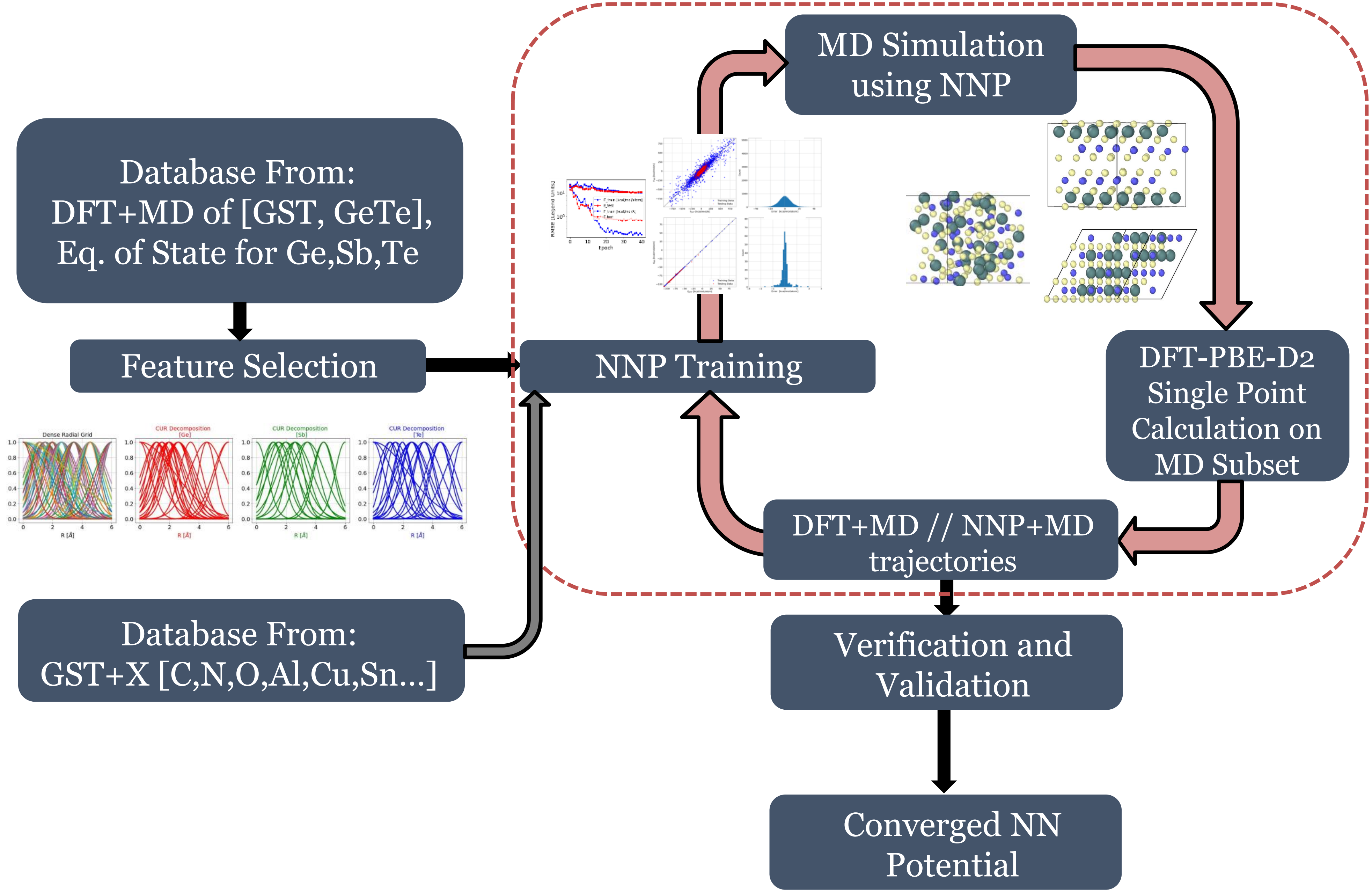
- Enrich dataset with NNMD trajectories
  - Explore edge cases

npj | Computational Materials
www.nature.com/npjcompumats
ARTICLE OPEN
Check for updates

Neural network reactive force field for C, H, N, and O systems

Pilsun Yoo[1], Michael Sakano[1], Saaketh Desai[1], Md Mahbubul Islam[1,2], Peilin Liao[1] and Alejandro Strachan[1]

NNP Iterative Workflow

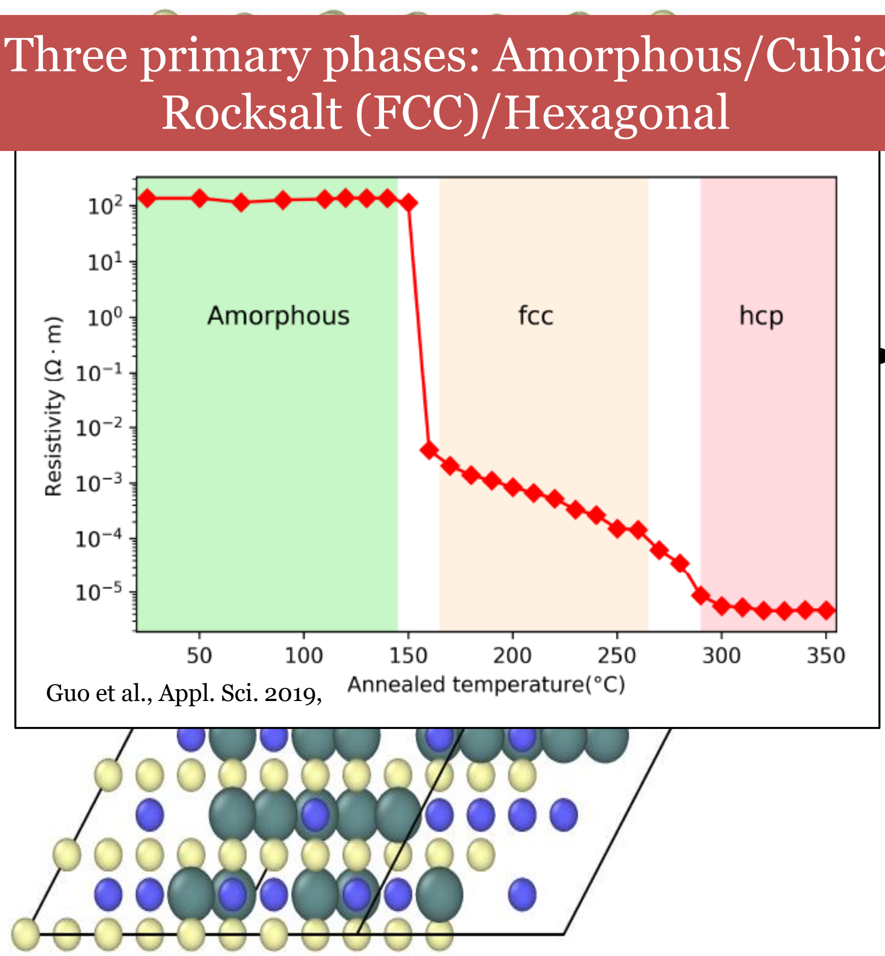


DFT Simulations – GGA – PBE+D2

Robert Appleton, Zachary McClure, Alejandro Strachan, David Adams

Hexagonal

Three primary phases: Amorphous/Cubic Rocksalt (FCC)/Hexagonal



Guo et al., Appl. Sci. 2019,

T_m = 900 K

$T_{x\text{-}tal}$ = 423 K

Ab initio MD – NVT

Analyze and break down free energy contributions

Xtal:

- T=[400,600,800]K
- Gather trajectories for each temp

Amorph:

- Melt at 2000K
- Quench and repeat temps from Xtal

GST+C:

- Create cubic structure with desired composition
- Melt & quench [see work by Robert Appleton]

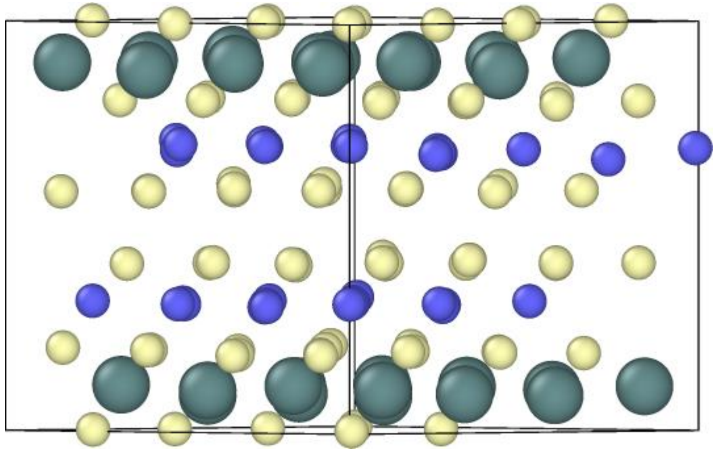
Run Params:

K-point grid: 1x1x1

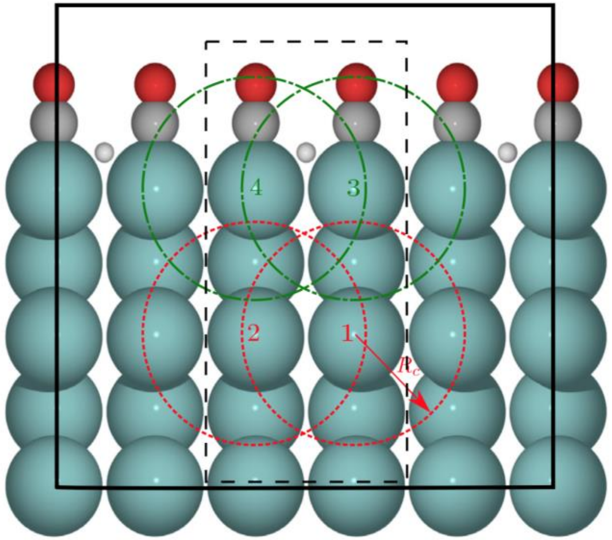
Supercell: 144 atoms

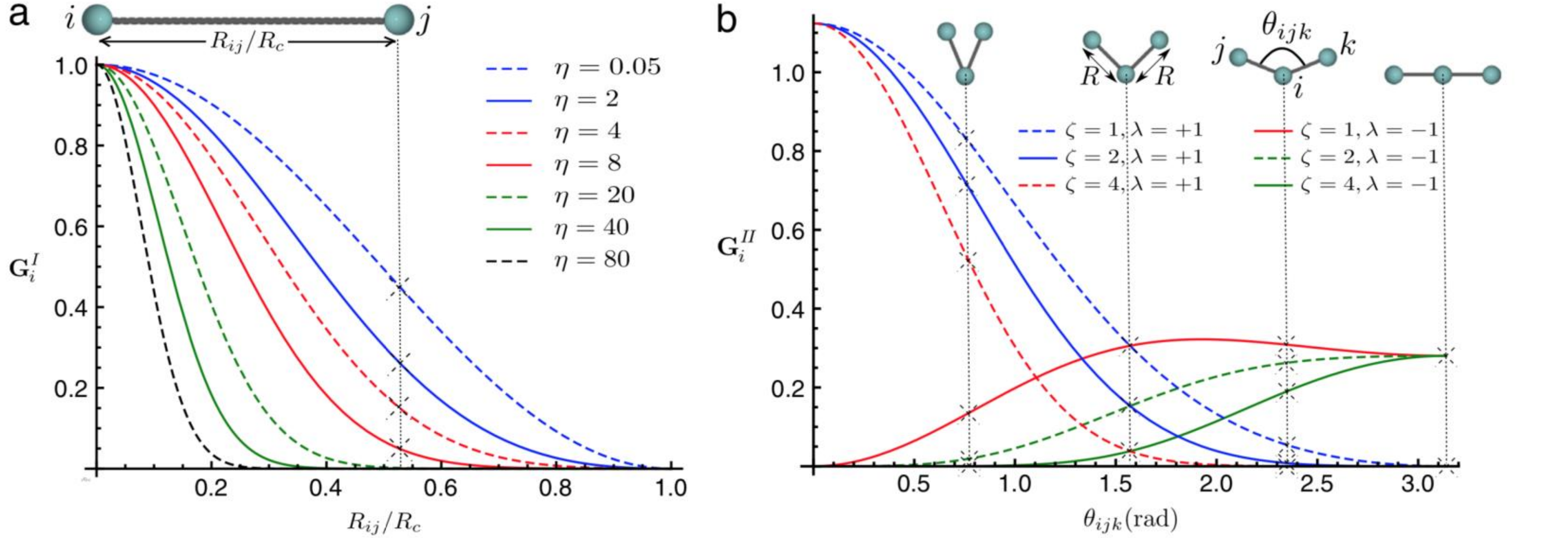
KE cutoff: 300 eV

Using neural network potentials



$$E_{\text{short}} = \sum_{N_{i=1}}^{N_{\text{elem}}} \sum_{j=1}^{N^{i}_{\text{atom}}} E^{i}_{j}$$



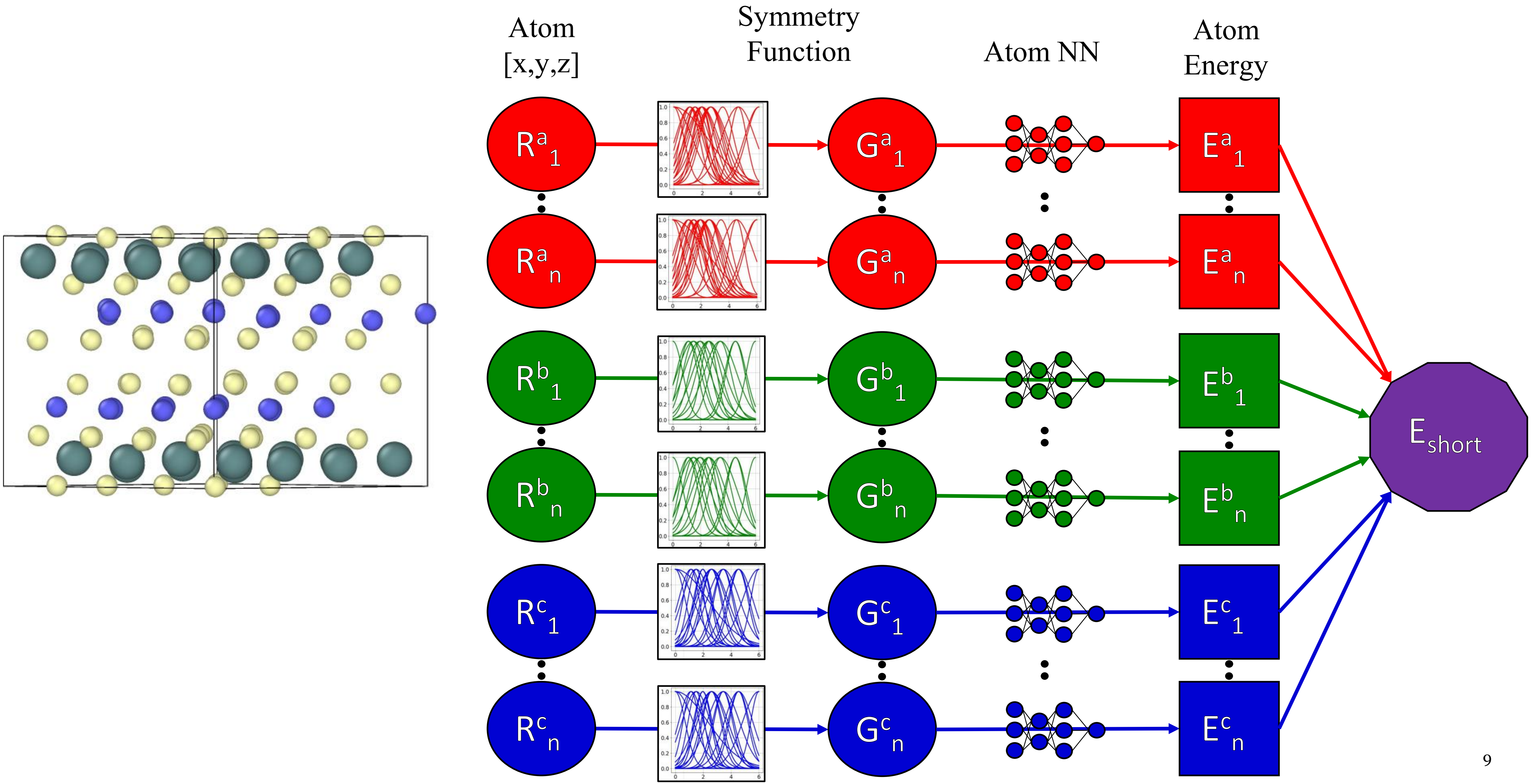


$$G_i^{w,rad} = \sum_{j \neq i}^{N_{\text{atom}} \in R_c} g(Z_j) \mathrm{e}^{-\eta (R_{ij} - R_s)^2} f_c(R_{ij})$$

$$G_i^{w,ang} = 2^{1-\zeta} \sum_{\substack{j,k \neq i \\ j<k}} h(Z_j, Z_k)(1 + \lambda \cos \theta_{ijk})^{\zeta} \mathrm{e}^{-\eta[(R_{ij}-R_s)^2 + (R_{ik}-R_s)^2 + (R_{jk}-R_s^2)]} \cdot f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk})$$

Khorshidi et al., CPC, 2016
Behler, JCP 2011
Gastegger et al., JCP, 2018

Using neural network potentials



Hyperparameters of features

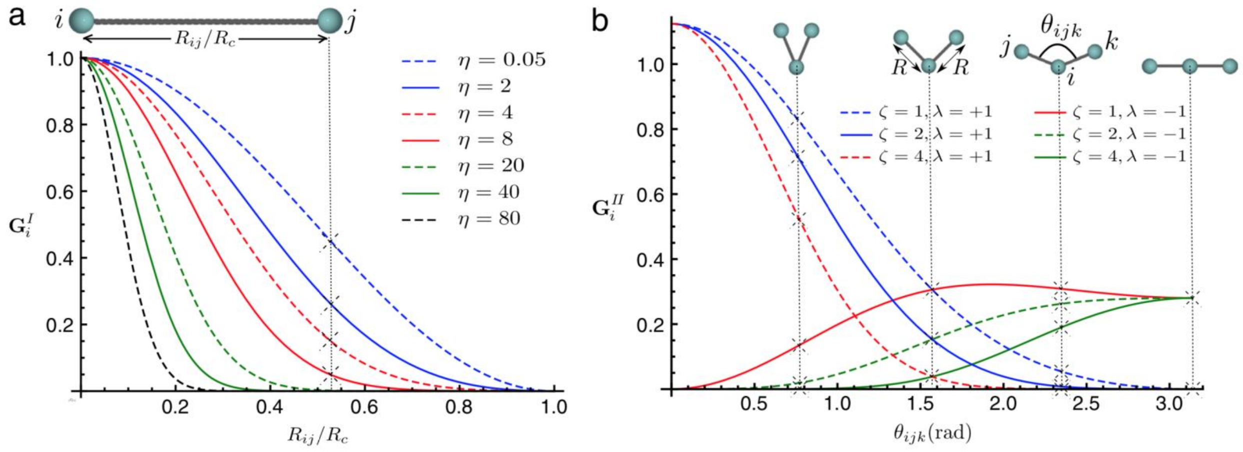
A simple gaussian function with a cutoff radius R_c, a width tuner η, and a shift parameter R_s

$$G_i^{w,rad} = \sum_{j \neq i}^{N_{\text{atom}} \in R_c} g(Z_j) \mathrm{e}^{-\eta (R_{ij} - R_s)^2} f_c(R_{ij})$$

A complex gaussian function with a cutoff radius R_c, a width tuner η, and a shift parameter R_s, now with a cosine function flipped by λ±1, and band width tuned by ζ

$$G_i^{w,ang} = 2^{1-\zeta} \sum_{\substack{j,k \neq i \\ j<k}} h(Z_j, Z_k)(1 + \lambda \cos \theta_{ijk})^{\zeta} \mathrm{e}^{-\eta[(R_{ij}-R_s)^2 + (R_{ik}-R_s)^2 + (R_{jk}-R_s^2)]}$$

$$\cdot f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk})$$



<u>How many features are we optimizing?</u>

3 elements

G^{rad}

R_s = 3x

η = 6x

G^{ang}

λ = 2x

ζ = 1x

R_s = 3x

η = 6x

54 possible features / element given initial grid

Initial Selection Database:

Robert Appleton, Zachary McClure,
Alejandro Strachan, David Adams

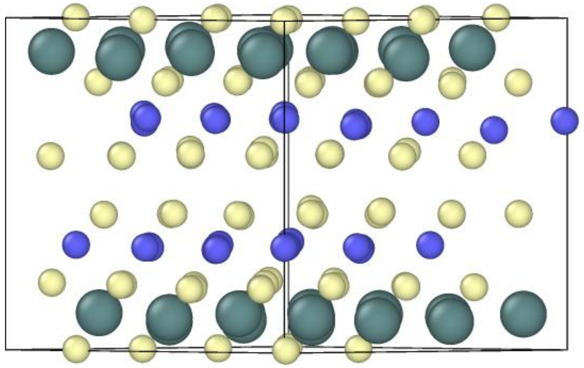
Hex – GST – 225
- @ [300,400,600,800] K

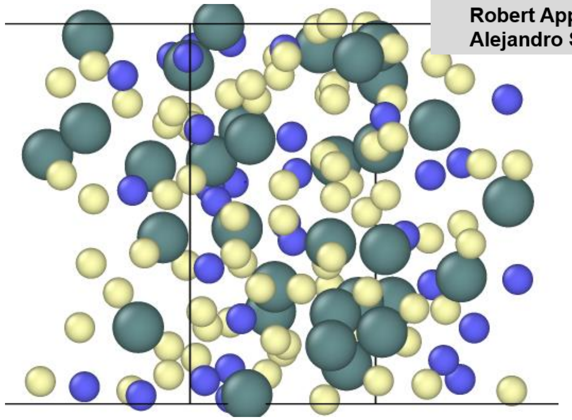
Cub – GST – 225
- @ [300,400,600,800] K

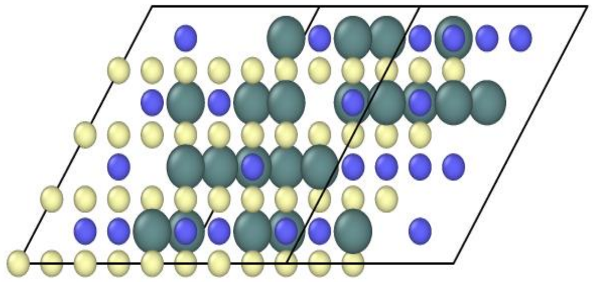
Amorph – GST – 225
- @ [300,400,600,800] K

Liq - GST – 225
- @ [1100K, 2000] K

Liq & Crystalline GeTe
- @ varied temperatures





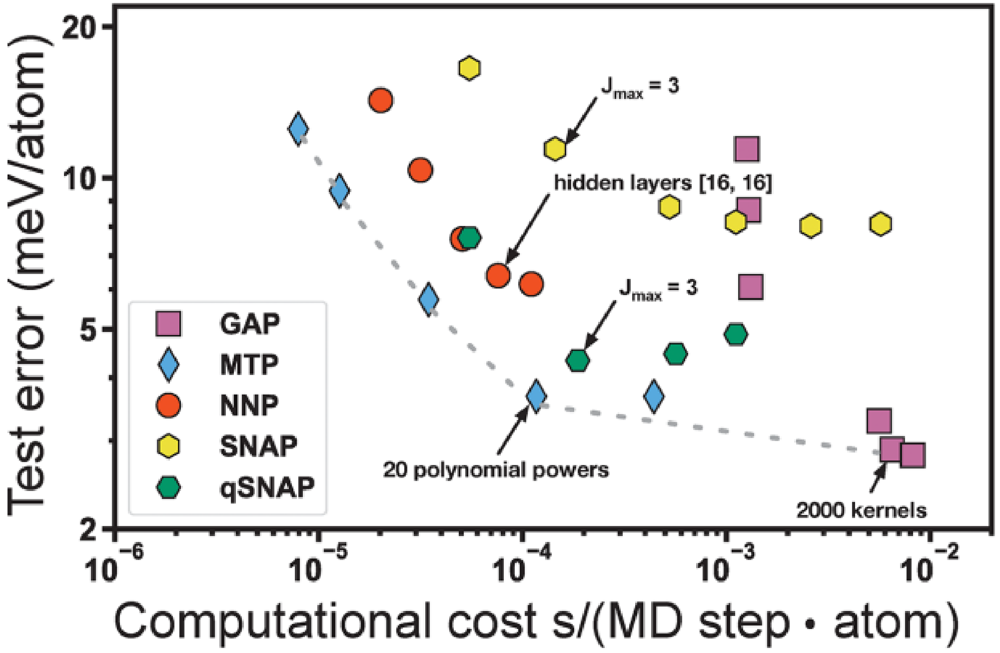


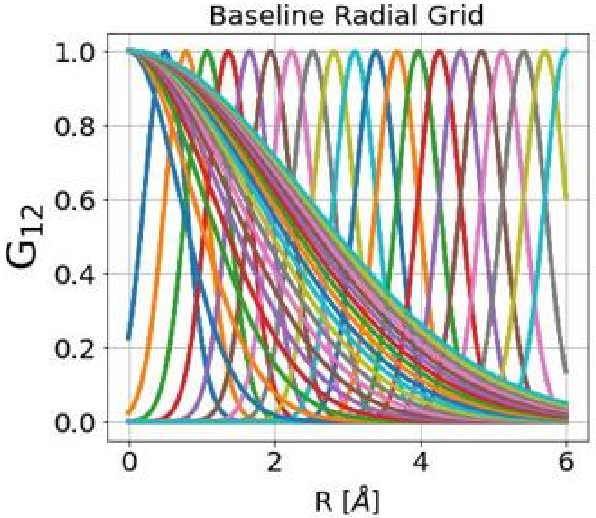
- Selected 1% of full DFT+MD database for initial training and calibration
- 90/10 train/test split for initial verification
- 2 Hidden Layers: [50/50]

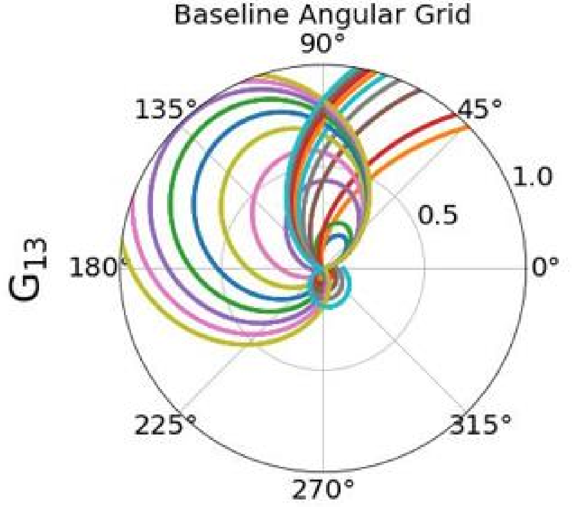
	Eq. of State [Ge,Sb,Te] S.P. DFT	GeTe MD+DFT	Hex. $Ge_2Sb_2Te_5$ MD+DFT	Cub. $Ge_2Sb_2Te_5$ MD+DFT	Amorph. $Ge_2Sb_2Te_5$ MD+DFT	Liq. $Ge_2Sb_2Te_5$ MD+DFT
Gen1.1	150	828	100	100	114	40

Limitation 1

Limitations to overcome:
1. **Feature selection**
2. Architecture initialization







3 elements

G^{rad}

R_s = 20x constant η
η = 18x constant R_s

G^{ang}

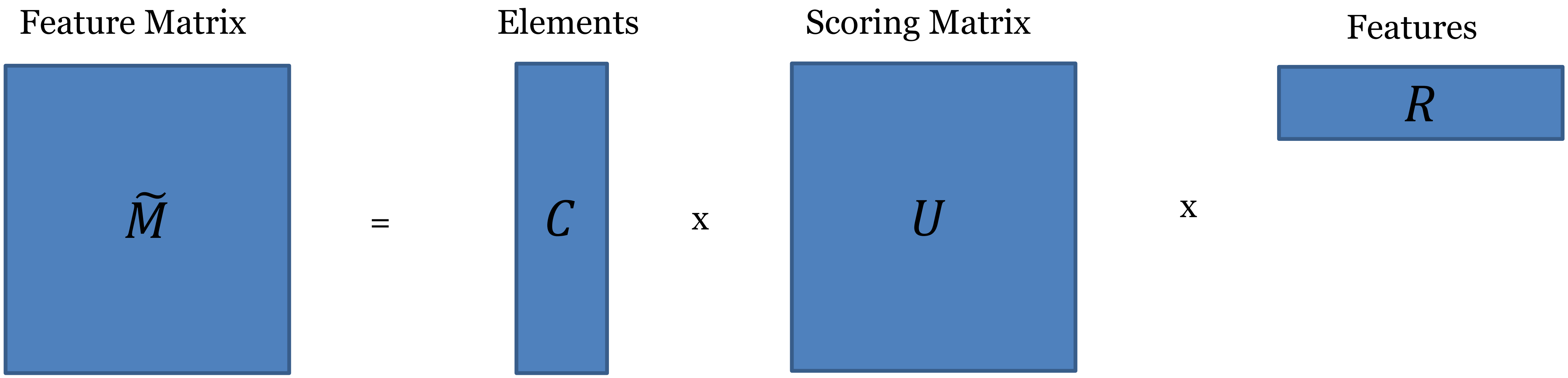
λ = 2x
ζ = 1x
R_s = 3x constant η
η = 3x constant R_s

Which symmetry functions best describe my environment?

What is my trade off for description vs. computational weight?

Solution 1: CUR Decomposition

$\tilde{M} \approx CUR$



$$\pi_c = \Sigma_{j=1}^{k} (v_c^{(j)})^2.$$

Scaling functions from symmetry functions used as weighting parameters in U

Highest expressivity, lowest overlap, most unique fingerprints retained

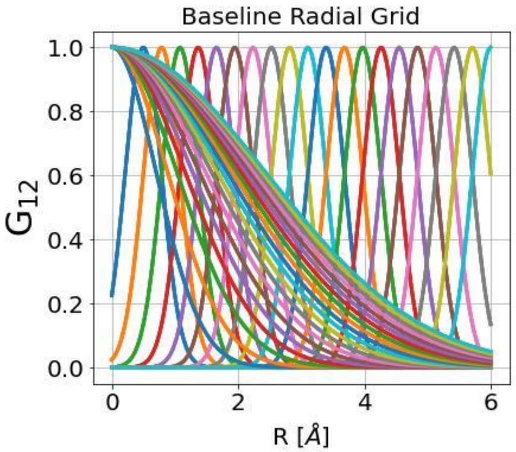
Retain the original matrix

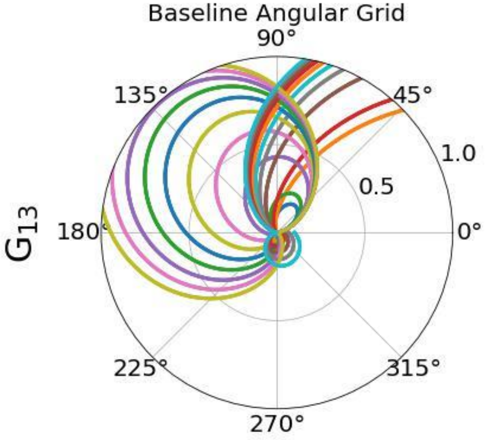
Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials

Cite as: J. Chem. Phys. **148**, 241730 (2018); https://doi.org/10.1063/1.5024611
Submitted: 02 February 2018 • Accepted: 10 April 2018 • Published Online: 30 April 2018

Giulio Imbalzano, Andrea Anelli, Daniele Giofré, et al.

Solution 1: CUR Decomposition



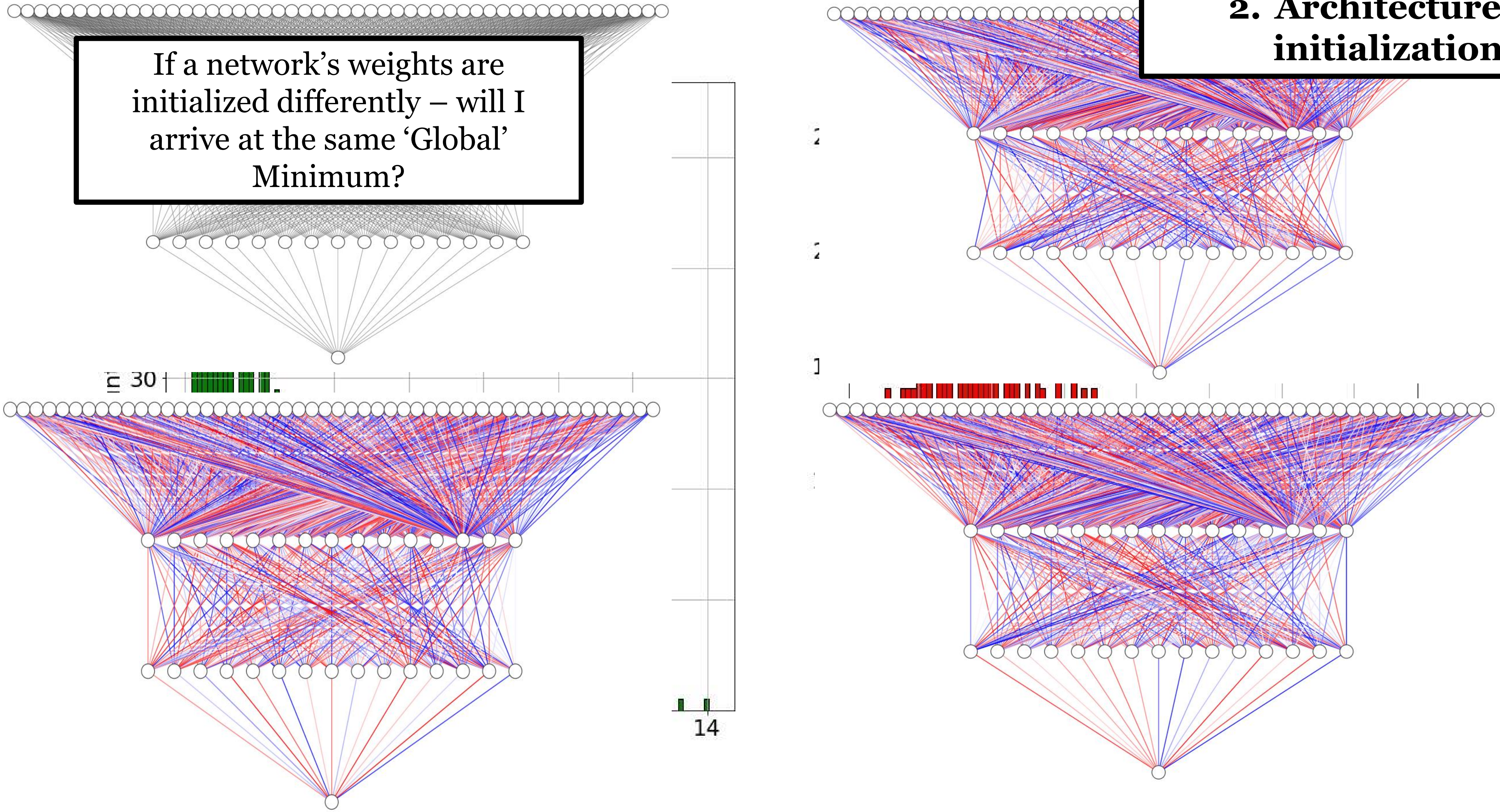


Sampling 1000 networks

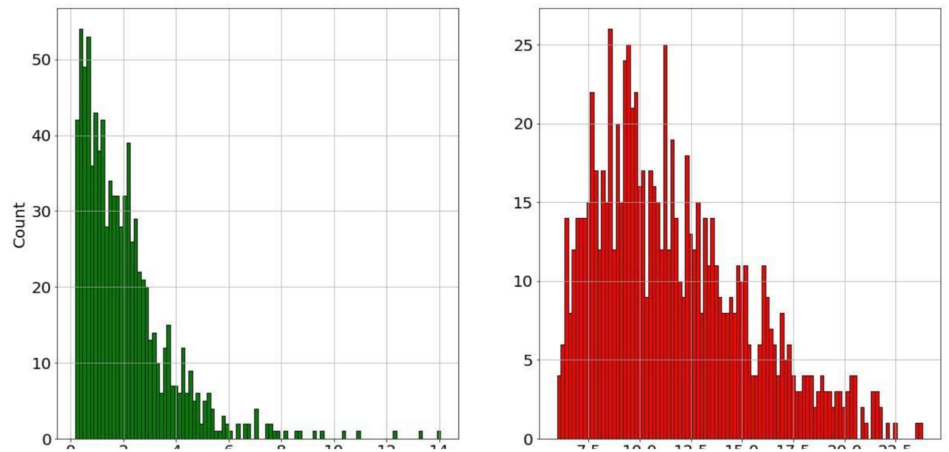
Limitations to overcome:

1. Feature selection
2. **Architecture initialization**

If a network's weights are initialized differently – will I arrive at the same 'Global' Minimum?



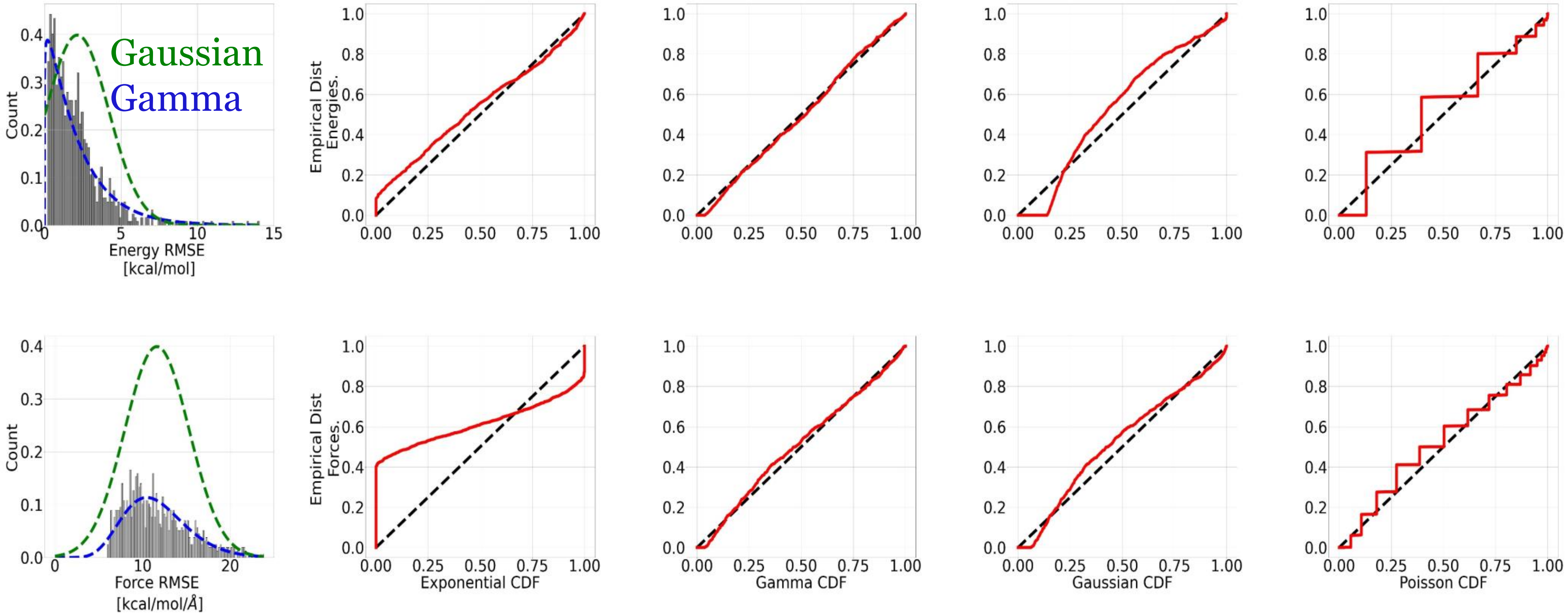
Characterizing the probability distribution



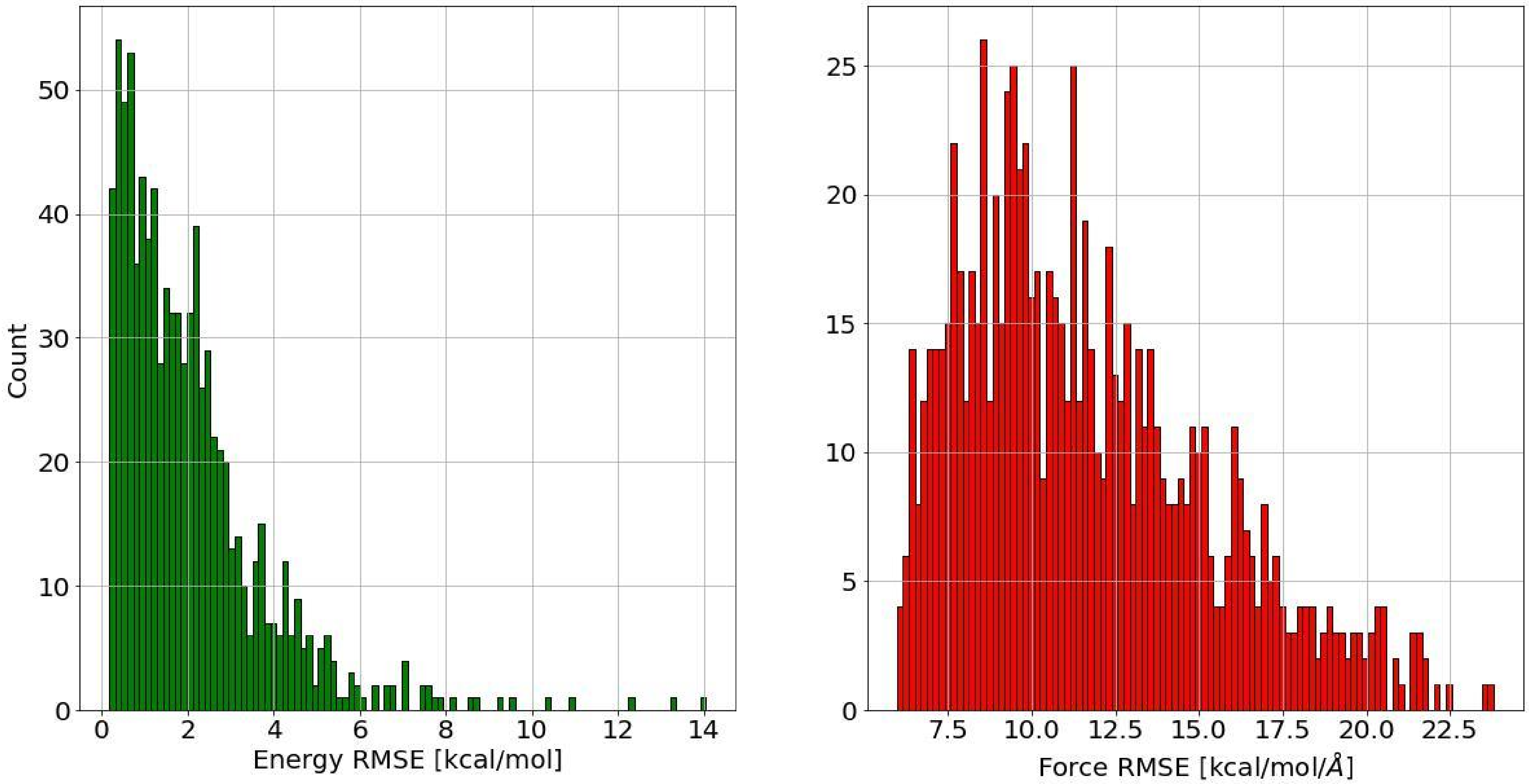
Distribution Type Hypothesis:

- Exponential
- Gamma
- Gaussian
- Poisson

Test with Cumulative Distribution Function:
$F_x(x) = P(X \leq x)$



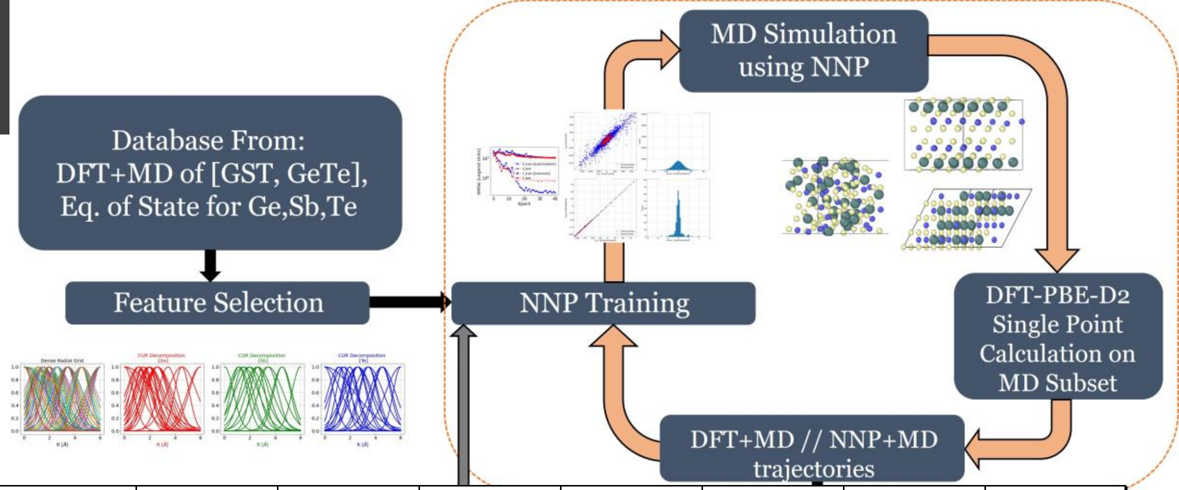
The Rule of 5



If I pick 5 different initializers, the probability of me selecting a value below the mean is XX
[the math needs to be done before the conference]

Generational Training Roadmap

- For each NNP+MD run create isotherm
- 144 atom configurations mapped from DFT structures
- 1 ns simulation time || sample 10 trajectories



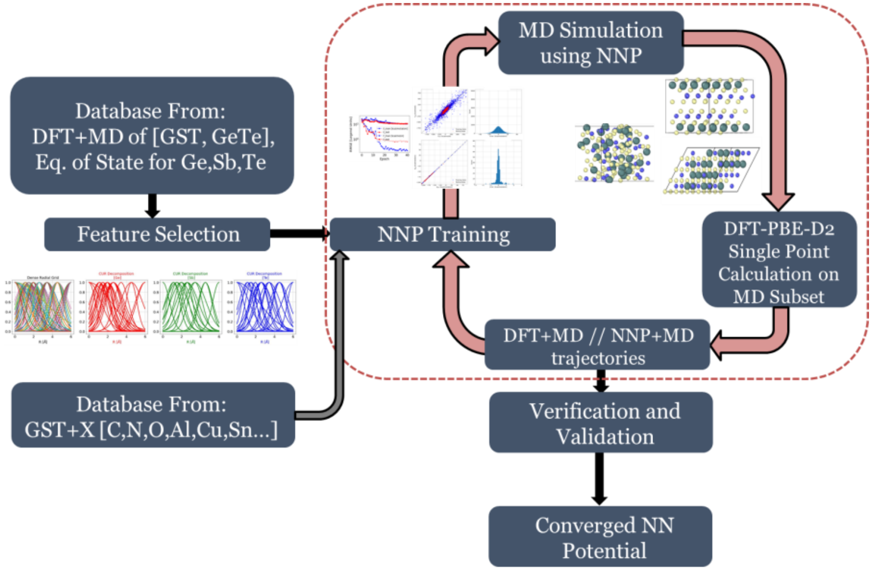
DFT+MD Database	Gen 1.1	Gen1.2	Gen1.3	Gen1.4	Gen1.5	Gen1.6	Gen1.7	Gen1.8	Gen1.9	Gen1.10	Gen1.11	Gen1.12	Gen1.13	Gen1.14	Gen1.15
Ge/Sb/Te															
GeTe															
h-GST															
c-GST															
a-GST[6.2]								300K						2000K	
a-GST[6.11]															
a-GST[5.88]															
l-GST															

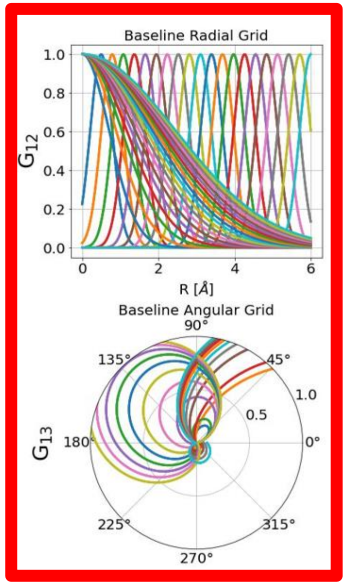
Generational Training Roadmap

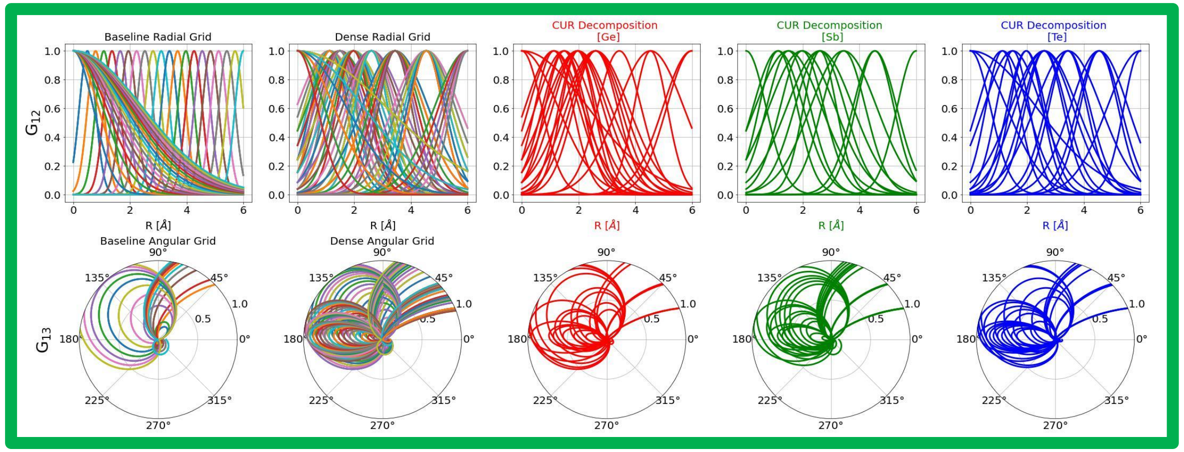
<u>Family 1</u>: Grid of features, use one network per iteration

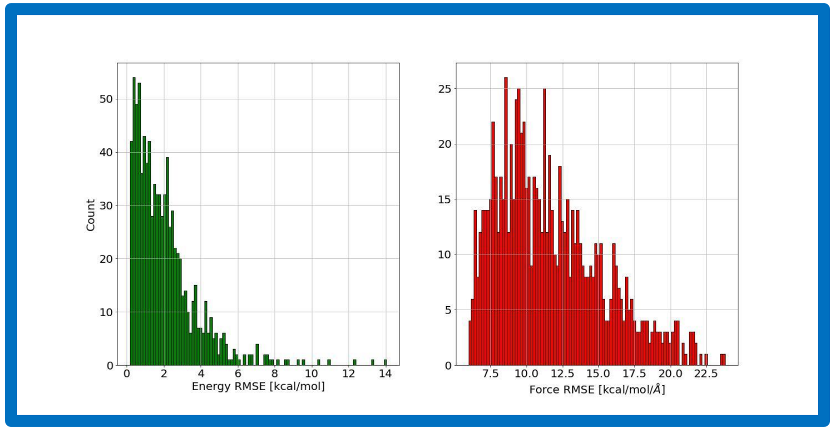
<u>Family 2</u>: CUR selected features, use one network per iteration

<u>Family 3</u>: CUR selected features, use five networks per iteration

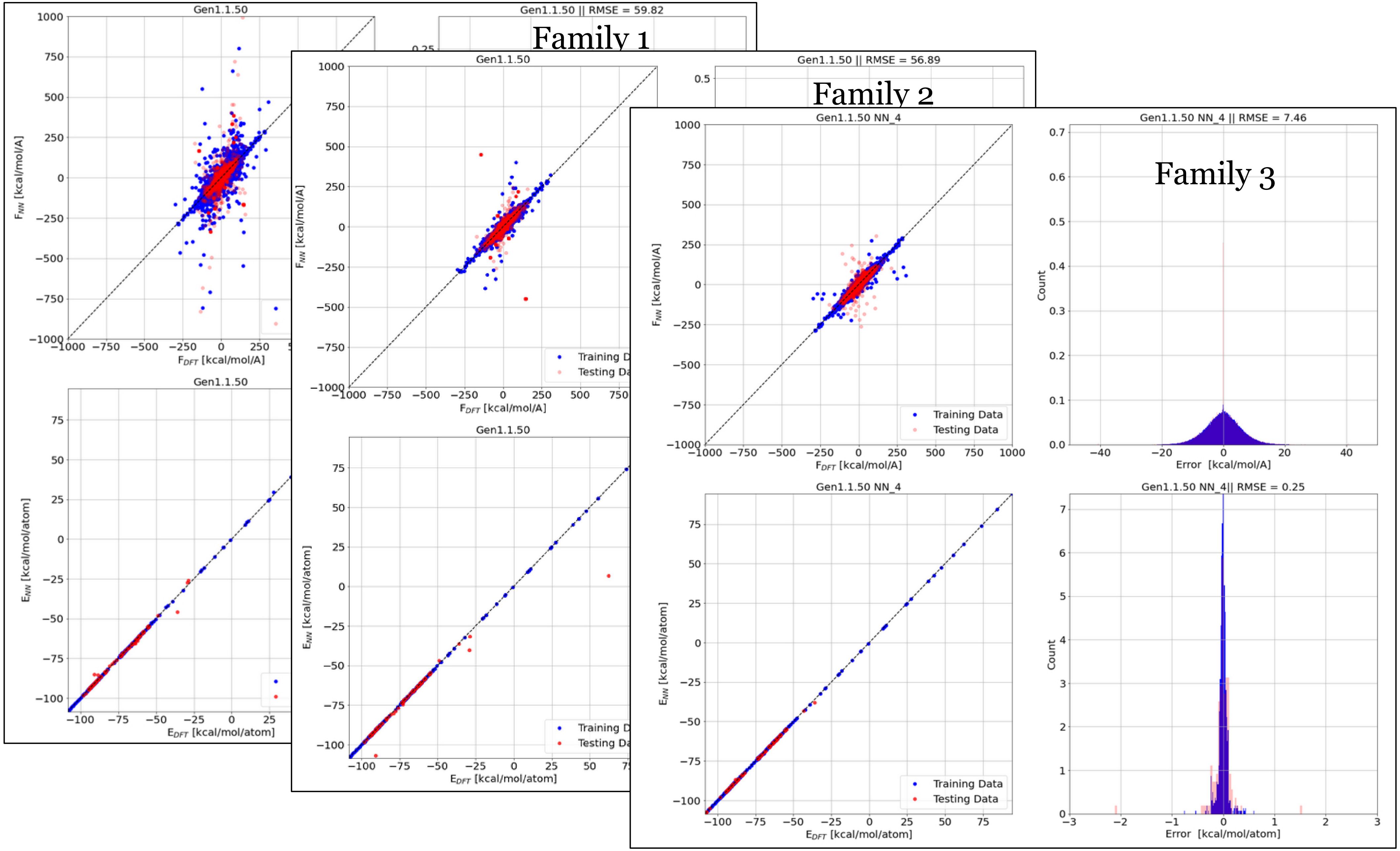




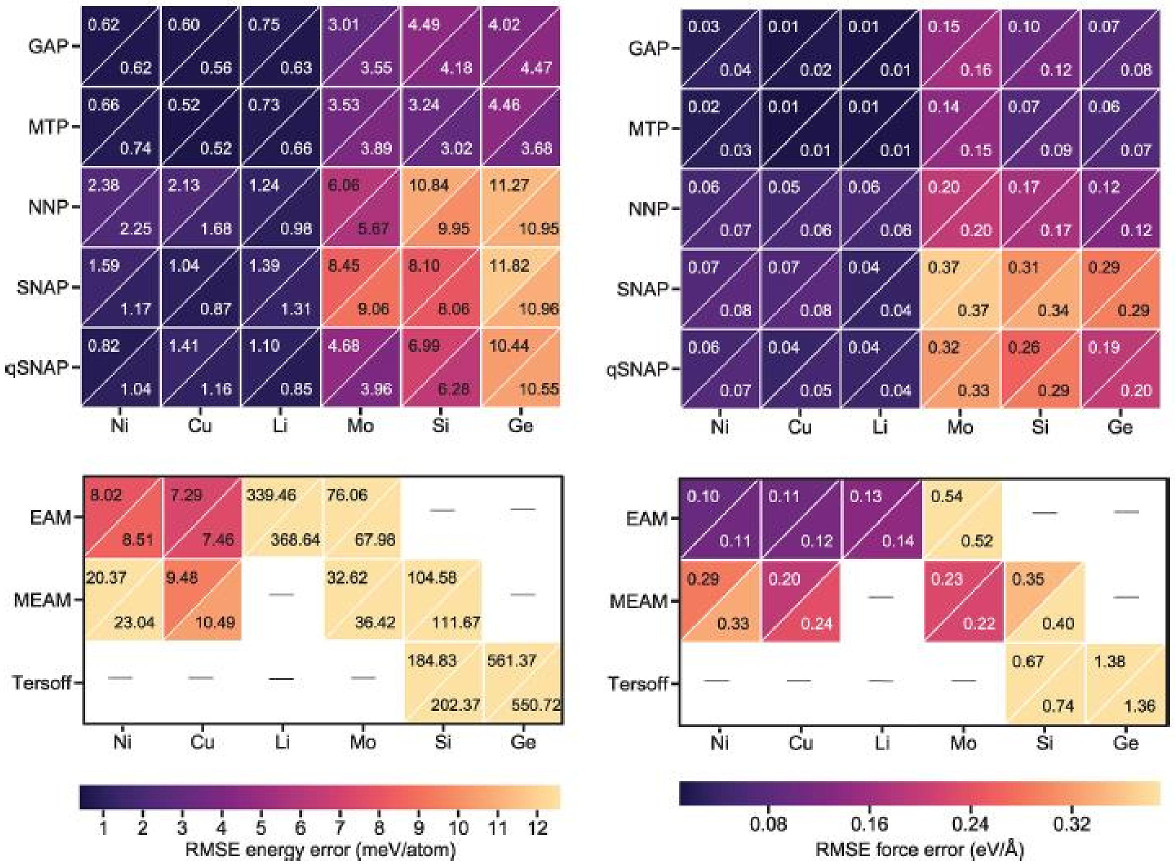


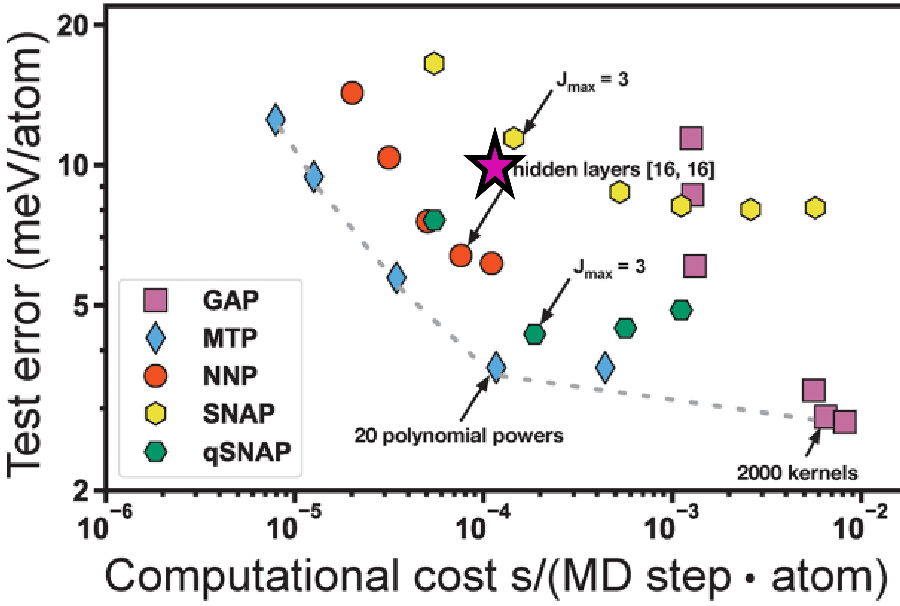


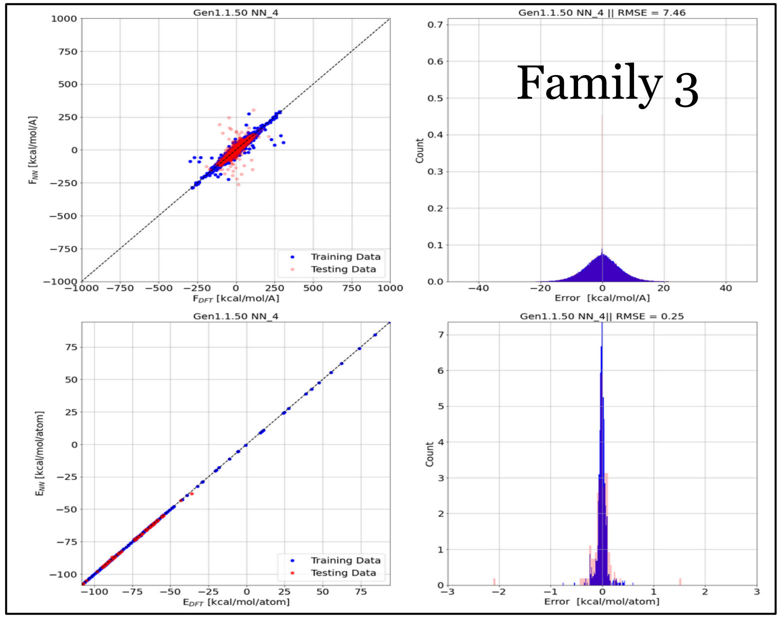
Baseline Training



How do we compare?

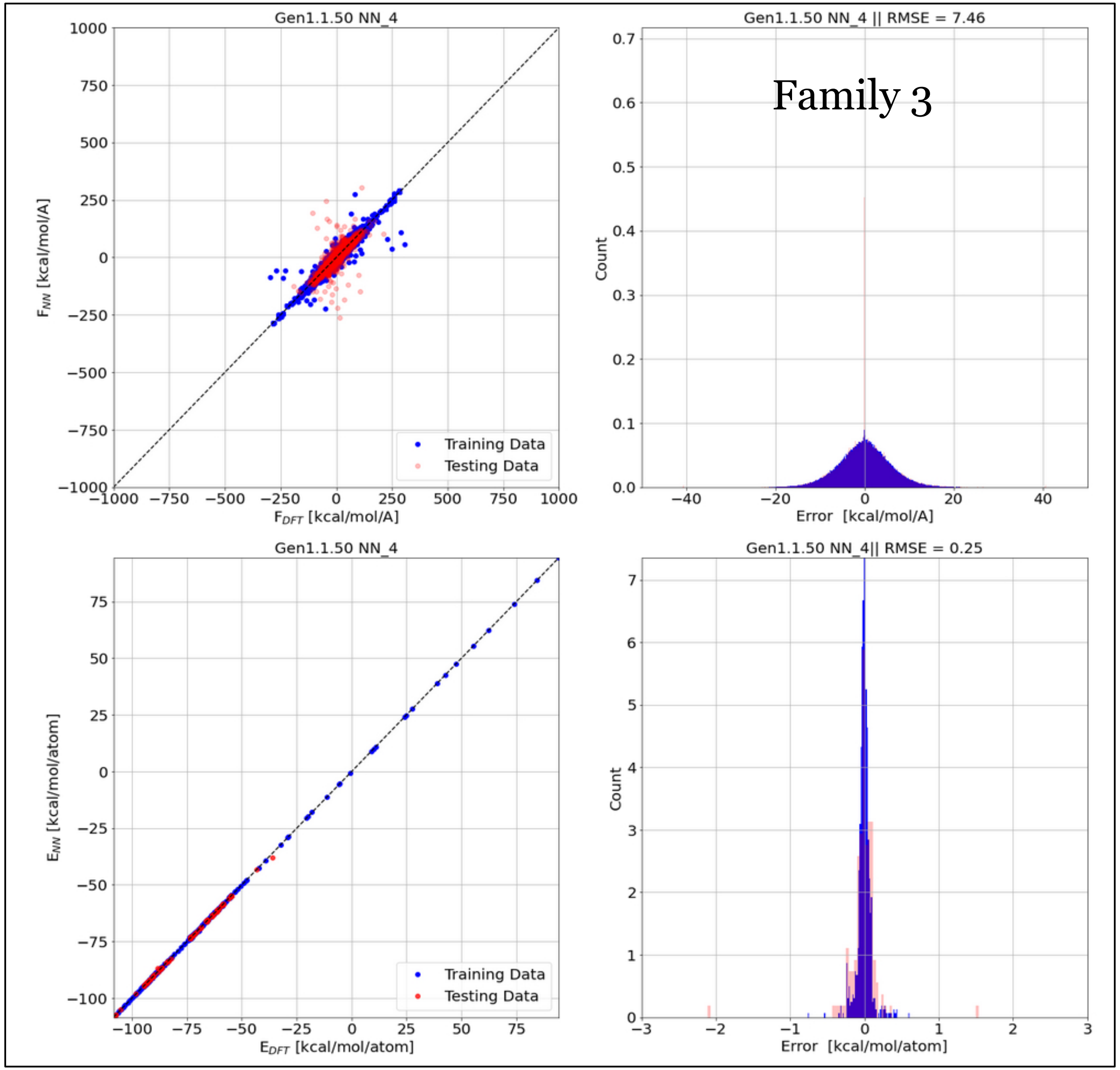




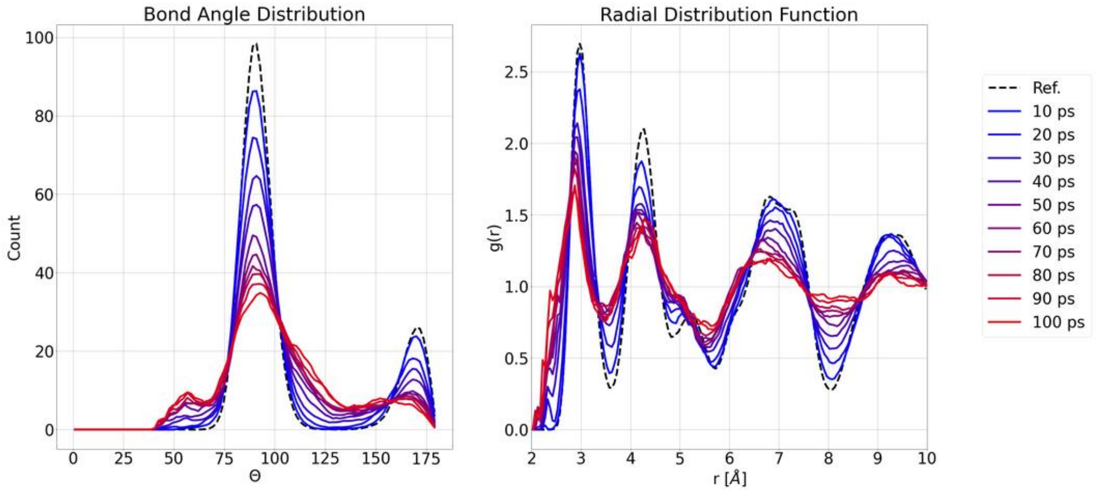


<u>NNP GeSbTe</u>
Force RMSE: 0.32 eV/Å
Energy RMSE: 10.87 meV/atom

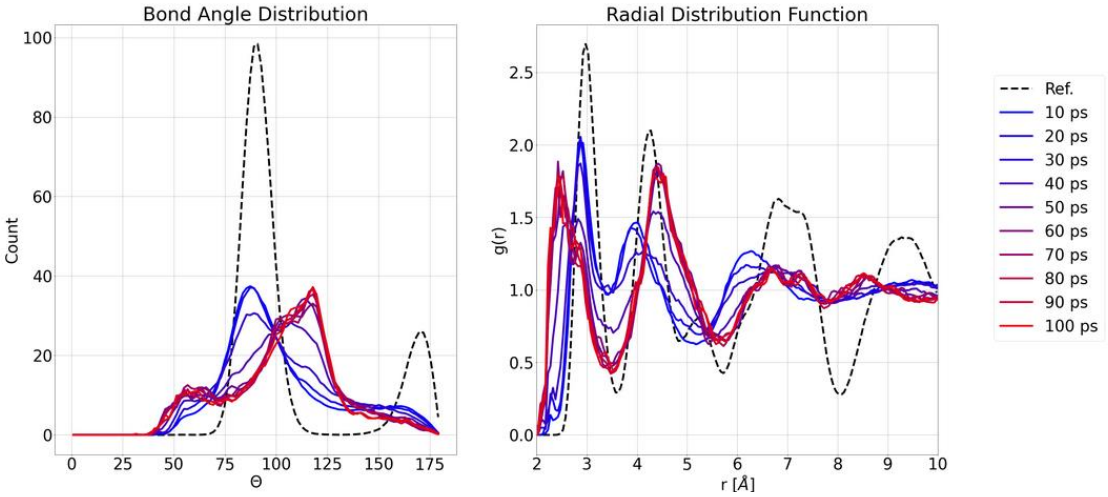
Training Gen1.1 – Molecular Dynamics



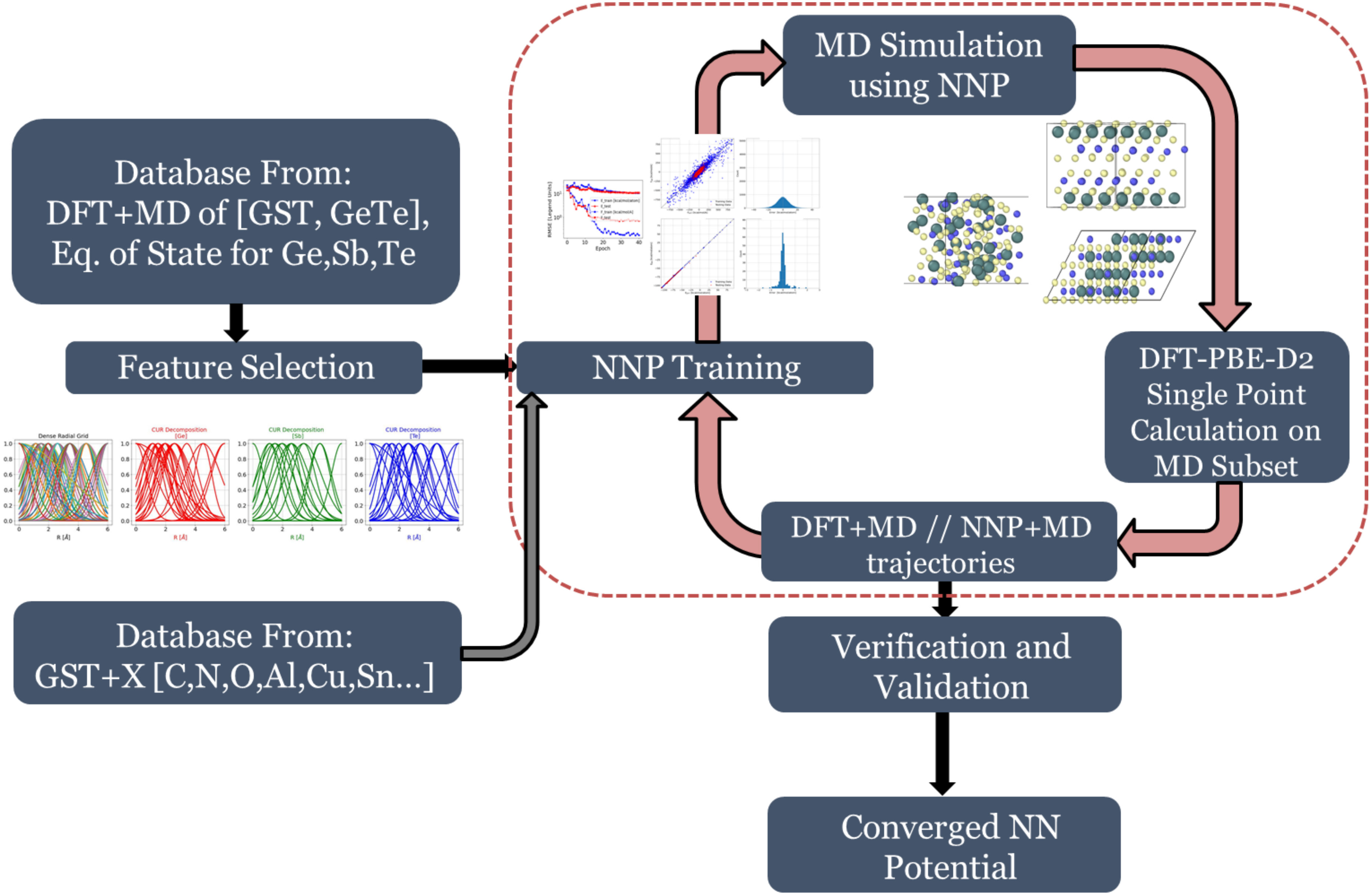
Stabilizing the Cubic Phase [X]



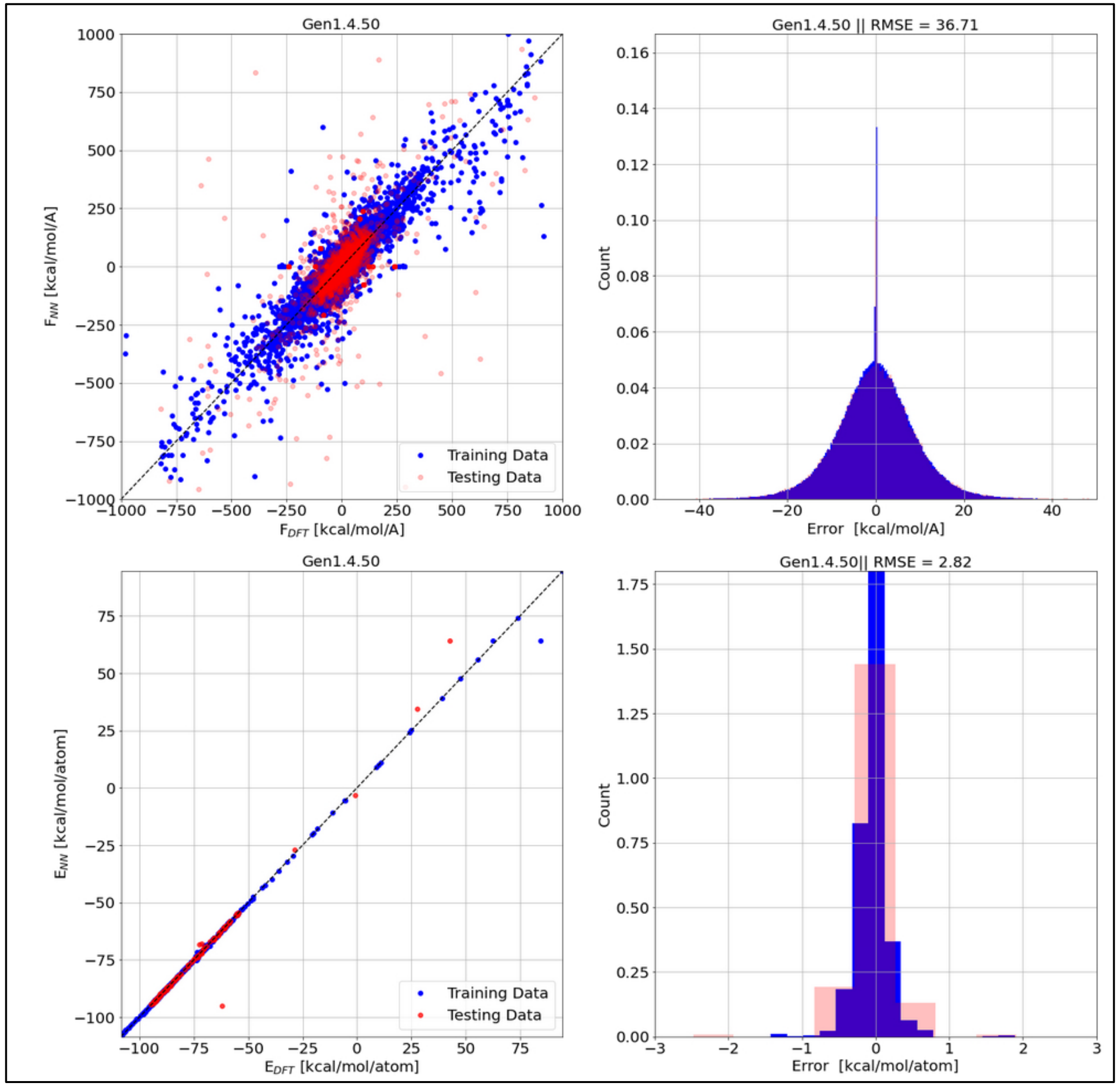
Amorphous to Xtal at 600K [X]



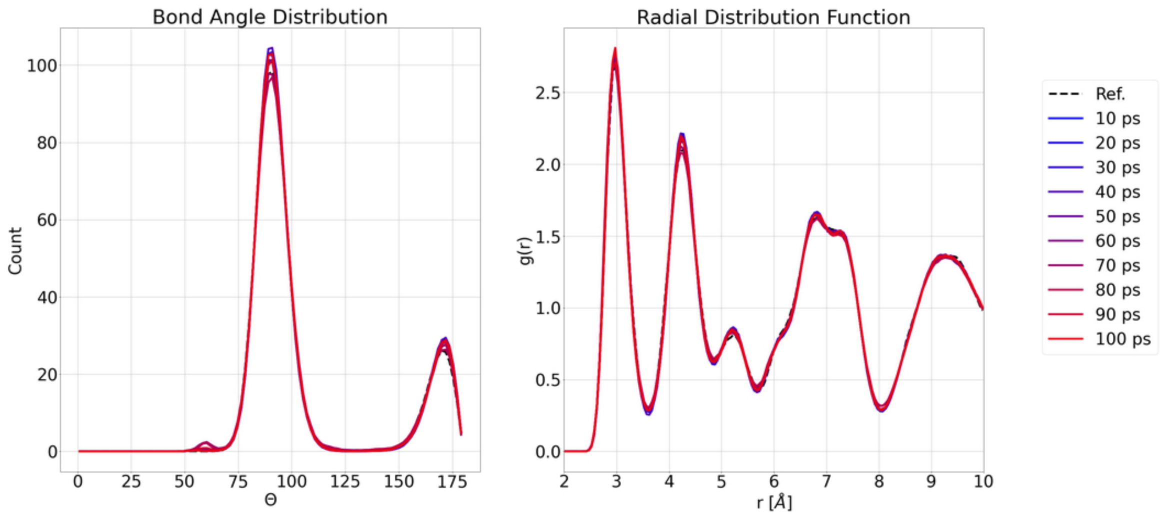
Begin the loop!



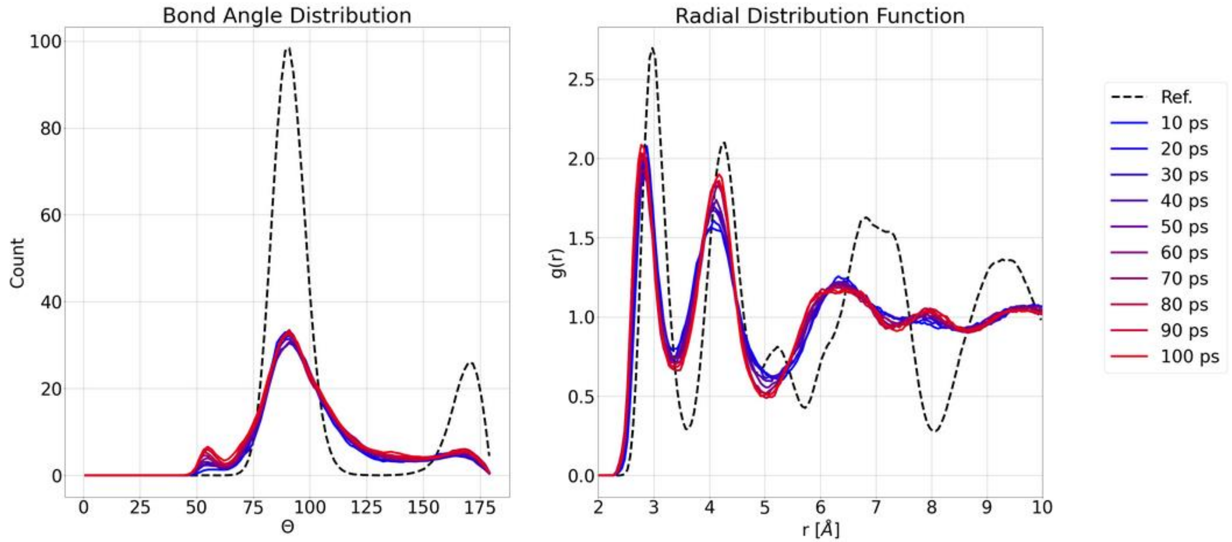
Training Gen1.4



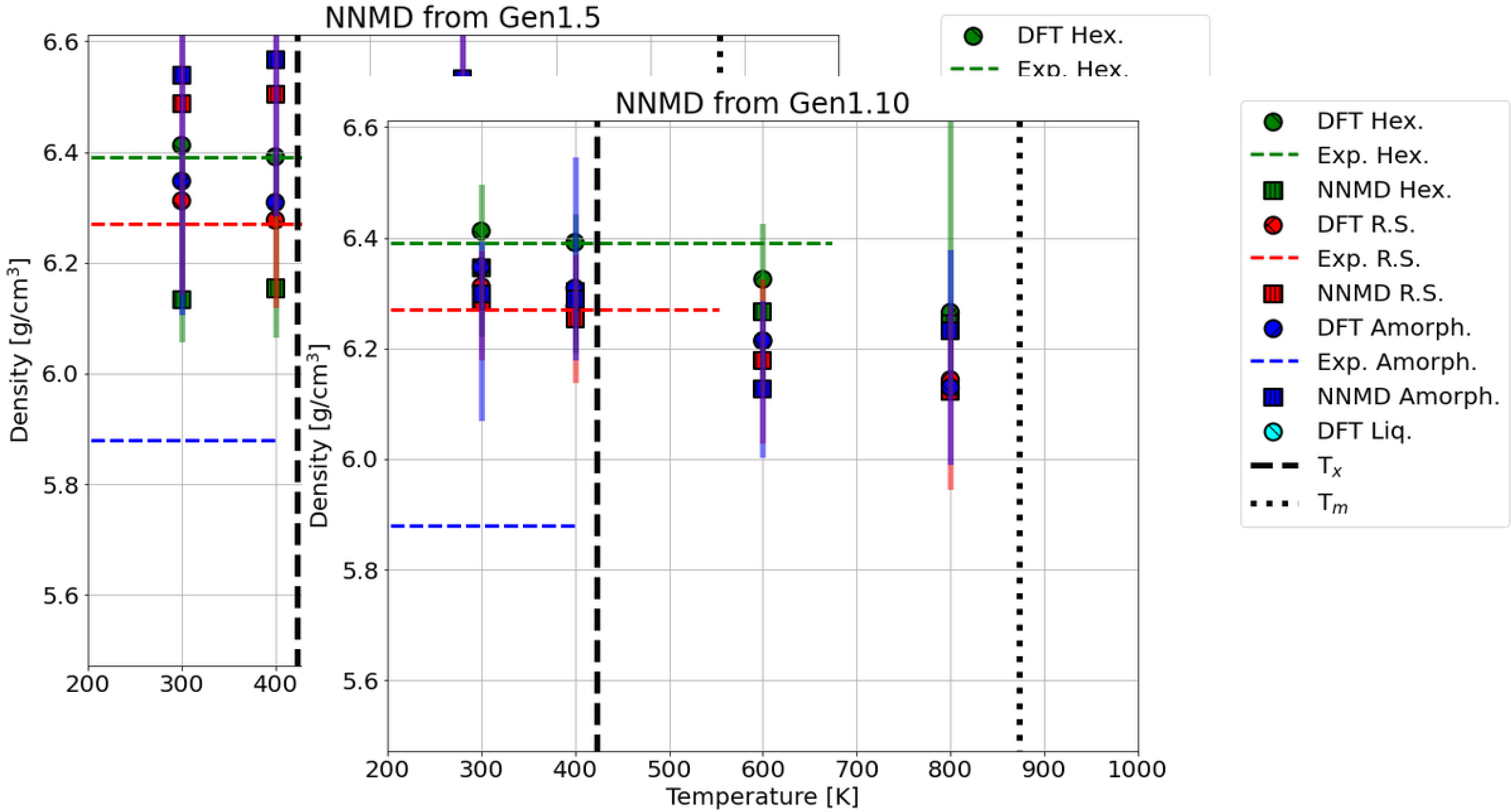
Stabilizing the Cubic Phase [√]



Amorphous to Xtal at 600K [X]

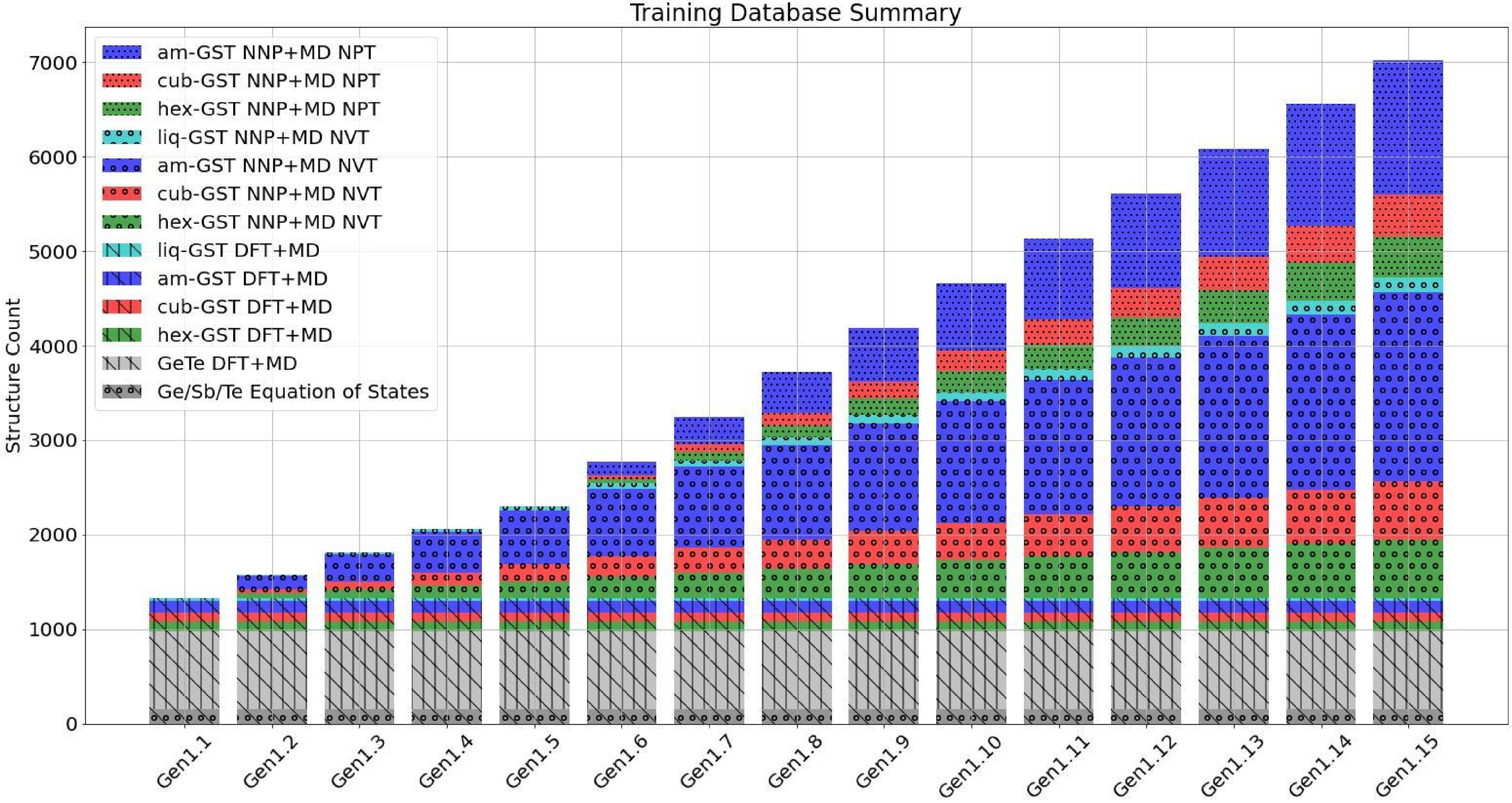


Switching to NPT and Evaluating Stability

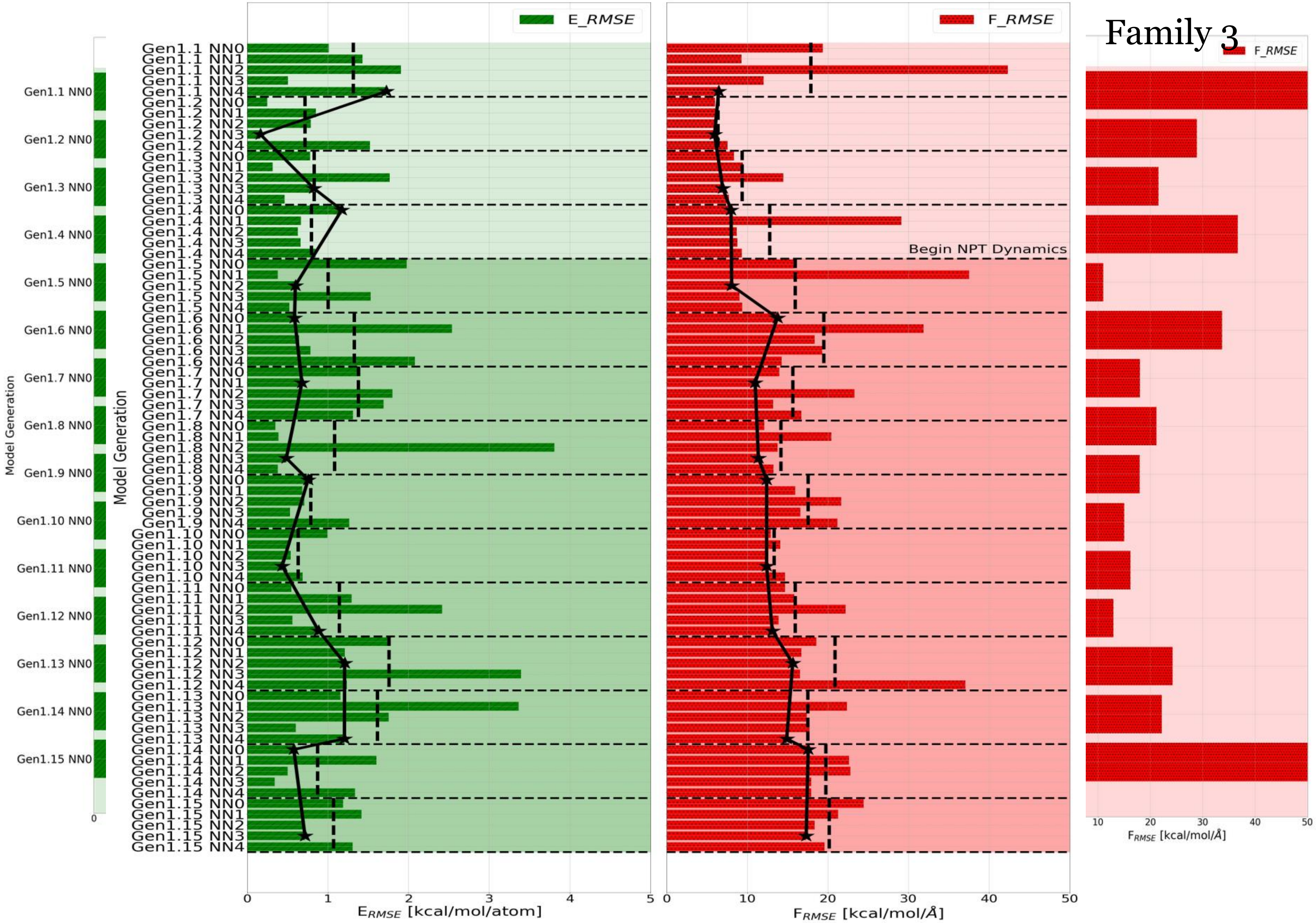


Database Composition

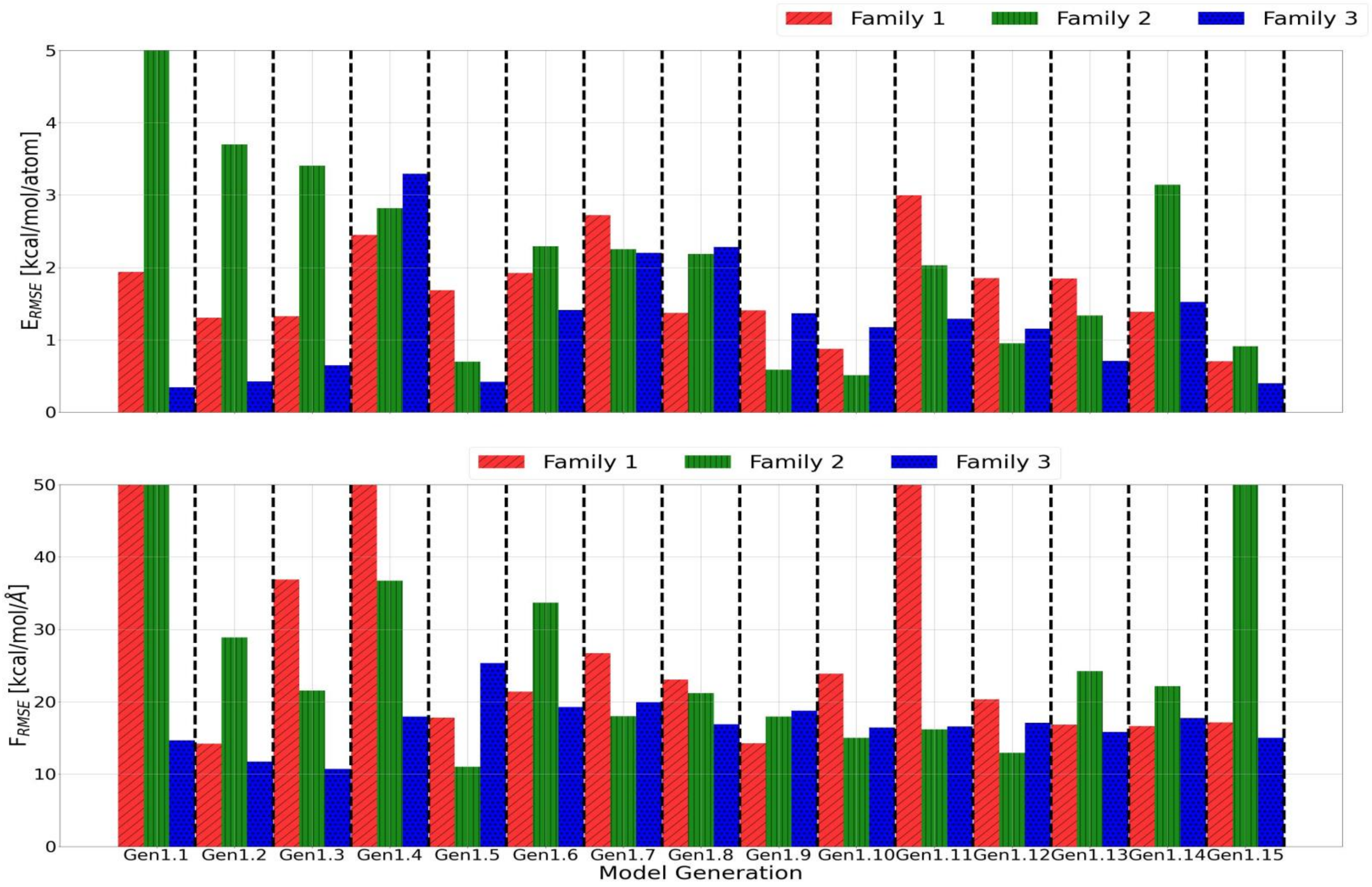
	Eq. of State [Ge,Sb,Te] S.P. DFT	GeTe MD+DFT	Hex. $Ge_2Sb_2Te_5$ MD+DFT	Cub. $Ge_2Sb_2Te_5$ MD+DFT	Amorph. $Ge_2Sb_2Te_5$ MD+DFT	Liq. $Ge_2Sb_2Te_5$ MD+DFT	Hex. $Ge_2Sb_2Te_5$ NVT	Cub. $Ge_2Sb_2Te_5$ NVT	Amorph. $Ge_2Sb_2Te_5$ NVT	Liq. $Ge_2Sb_2Te_5$ NVT	Hex. $Ge_2Sb_2Te_5$ NPT	Cub. $Ge_2Sb_2Te_5$ NPT	Amorph. $Ge_2Sb_2Te_5$ NPT
Gen1.1	150	828	100	100	114	40	0	0	0	0	0	0	0
Gen1.2	150	828	100	100	114	40	44	44	143	11	0	0	0
Gen1.3	150	828	100	100	114	40	88	88	286	22	0	0	0
										33	0	0	0
										44	0	0	0
										55	44	44	143
										66	88	88	284
										77	132	132	427
										88	176	176	570
										99	220	220	713
										110	264	264	856
										121	308	308	999
										132	352	352	1142
										143	396	396	1285
										154	440	440	1428



Comparing Three Families



Comparing Three Families



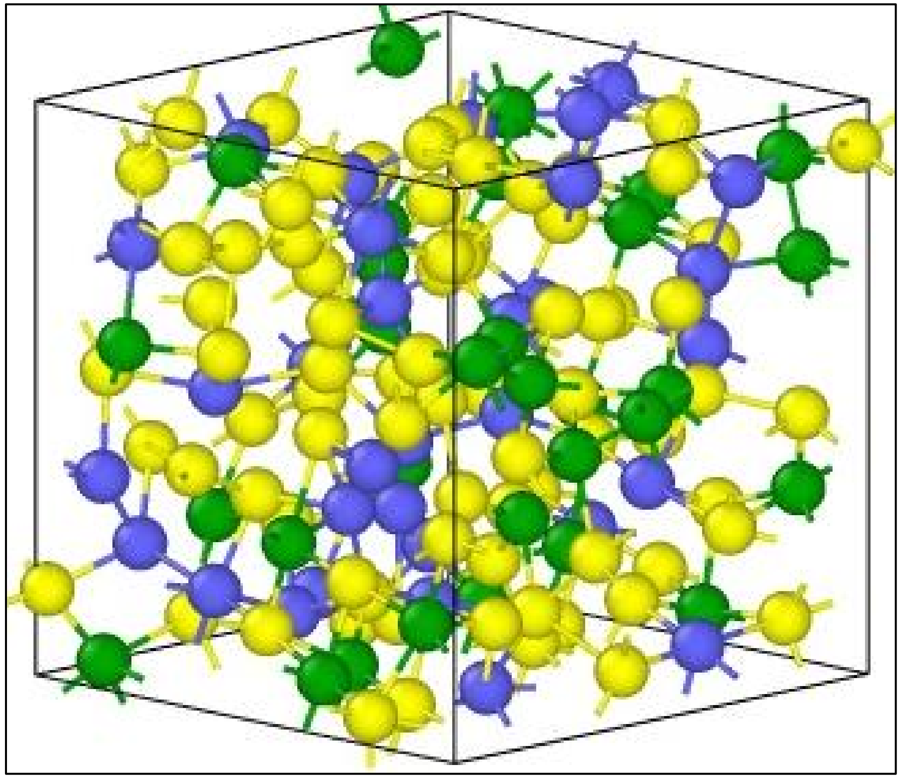
- CUR selected features [Family 2/3] outperform grid of features [Family 1]

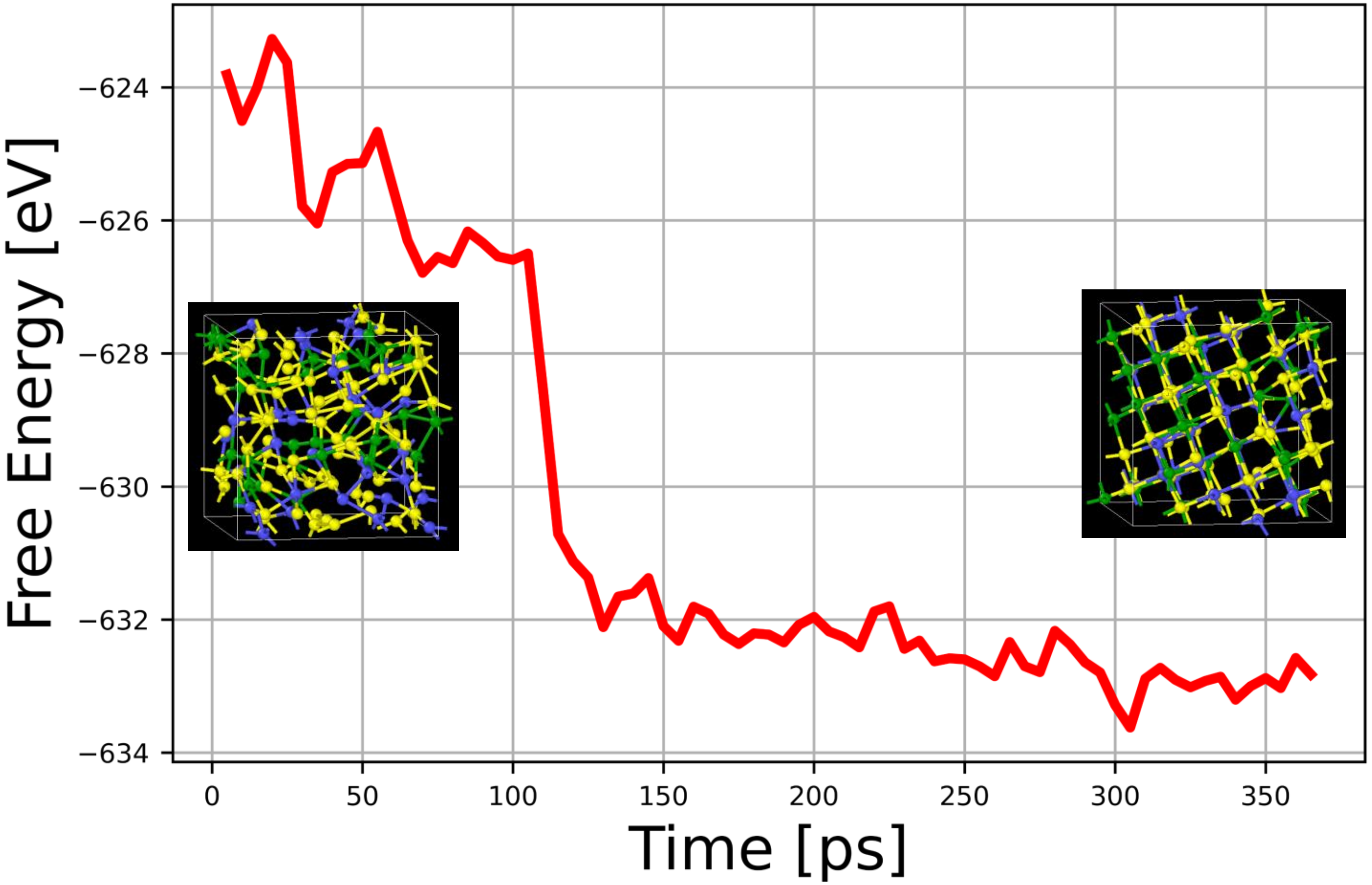
- [Family 3] allows for well controlled dynamics and database enrichment

Validation with DFT

Robert Appleton, Zachary McClure, Alejandro Strachan, David Adams

DFT+MD Simulation @ 600K || NVT ensemble

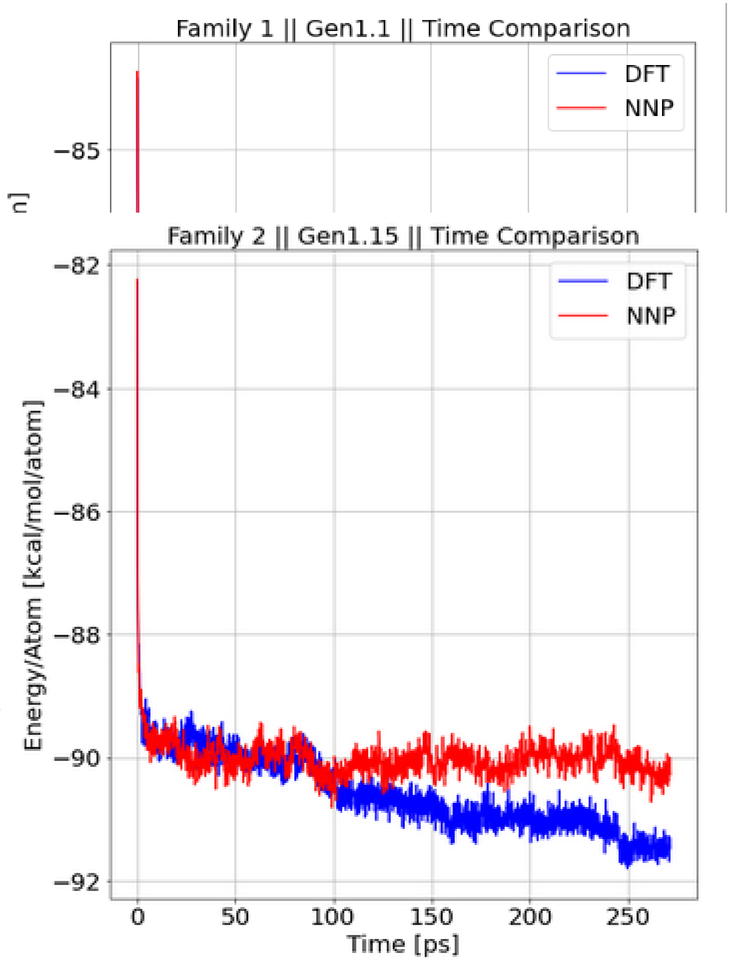




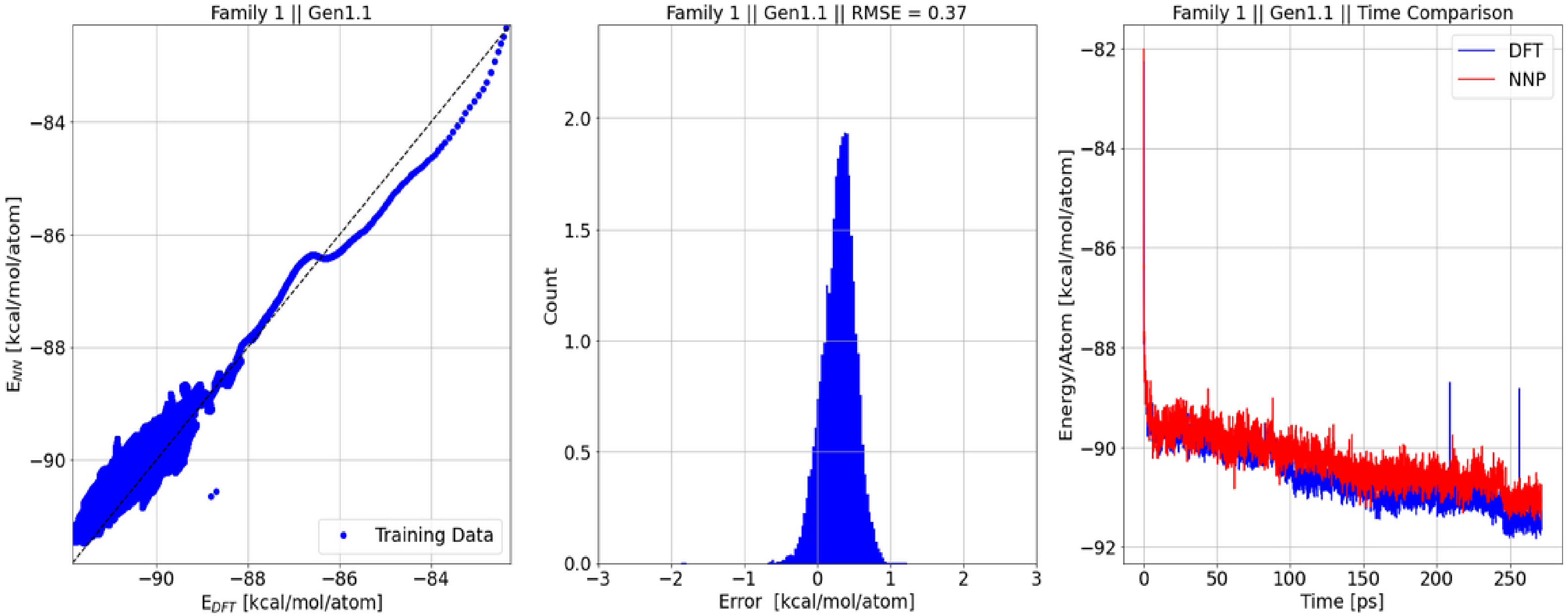
Crystallization via molecular dynamics

2 Flavors of Validation

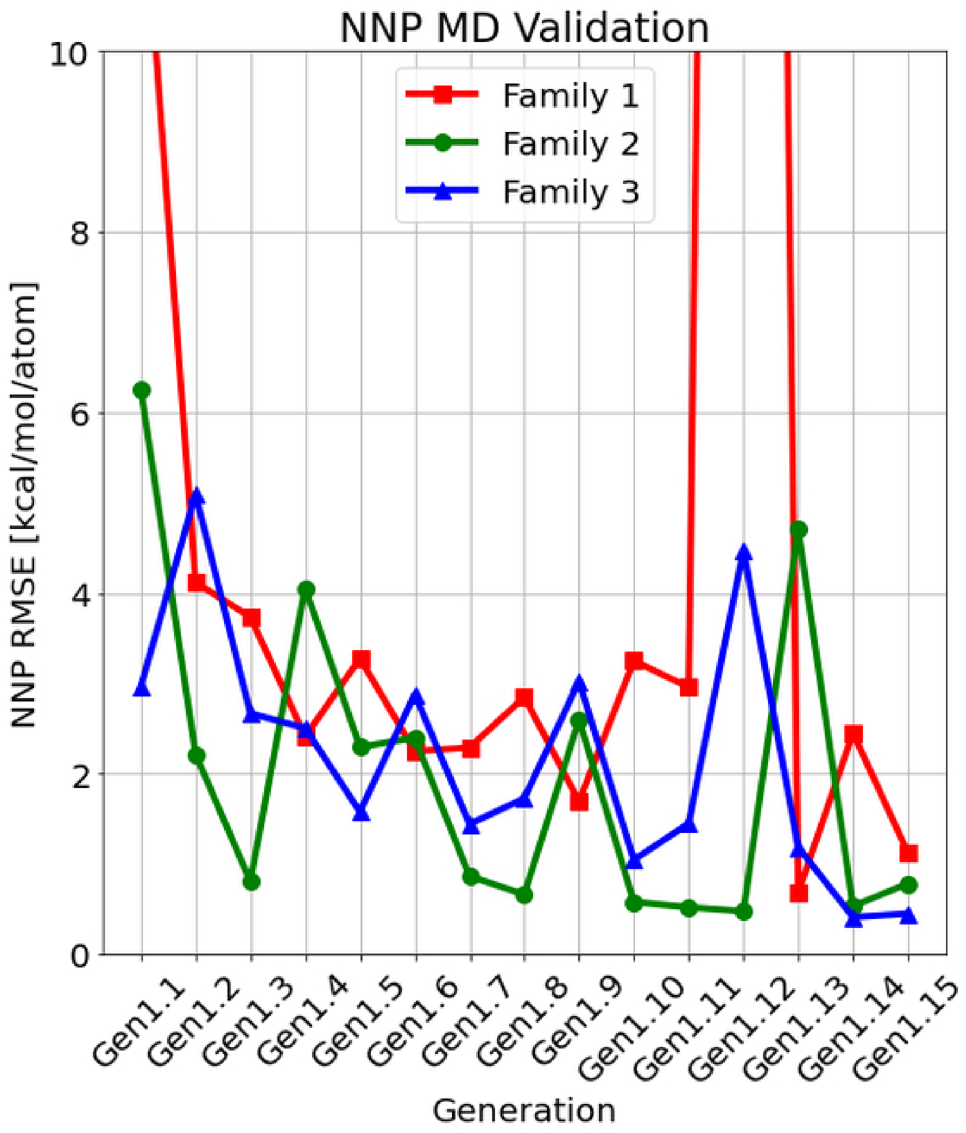
1.) How well can my trained weights reproduce an DFT+MD simulation of amorphous recrystallization with NNP+MD?

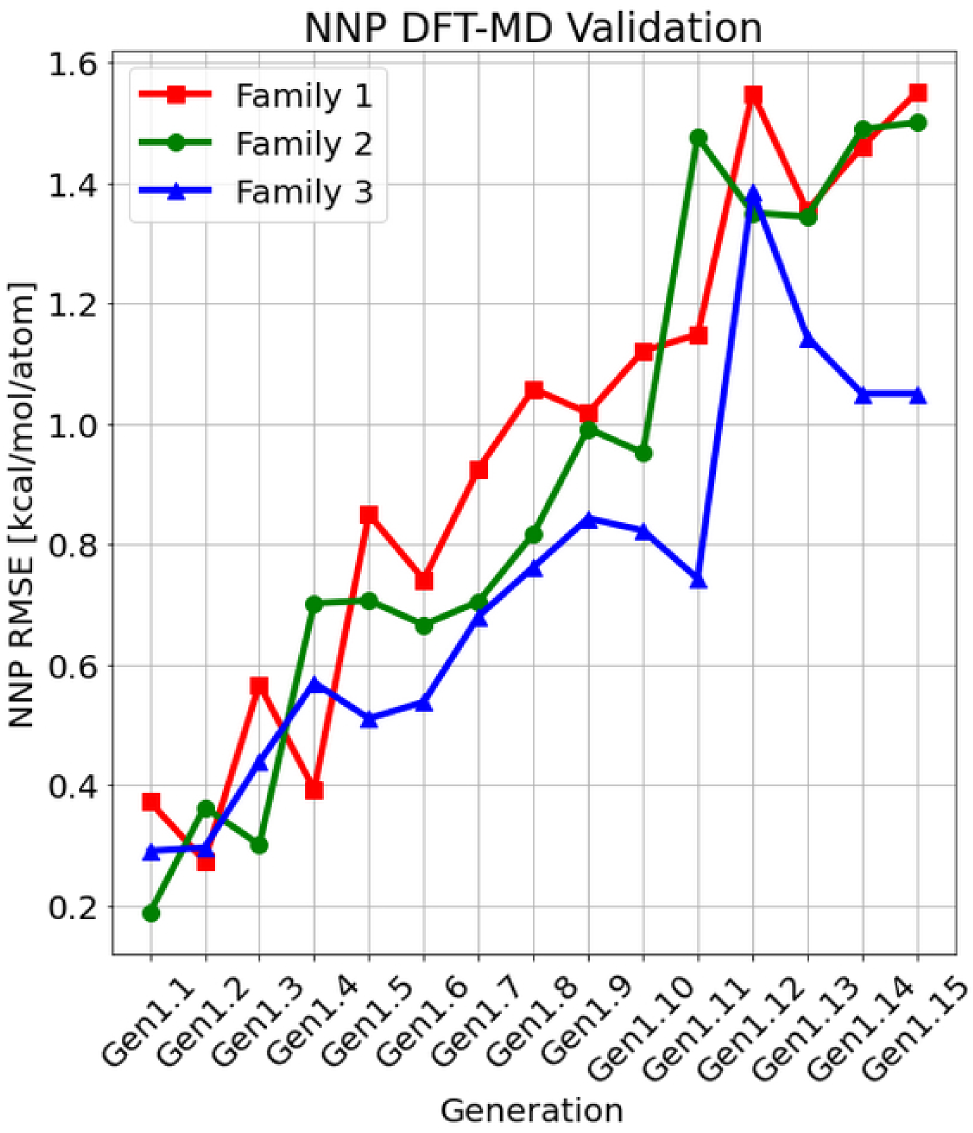


2.) How well can my weights reproduce energies of the DFT+MD trajectory



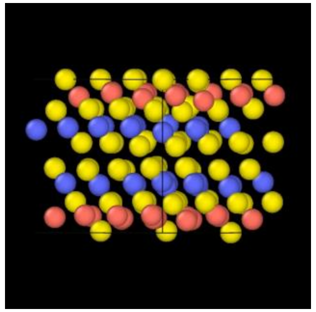
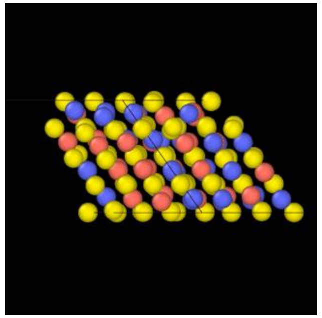
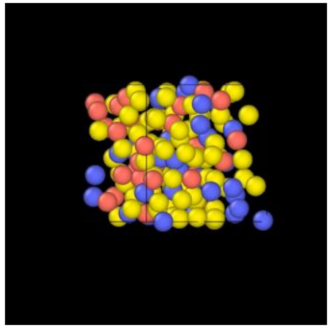
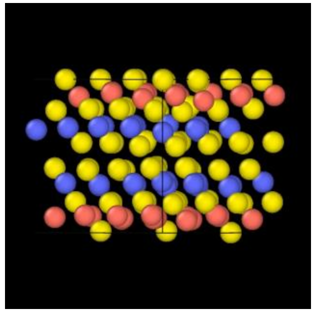
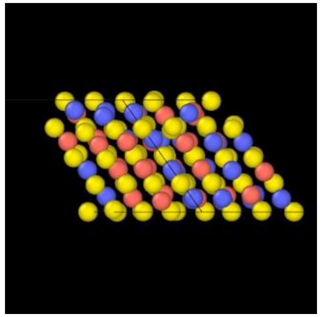
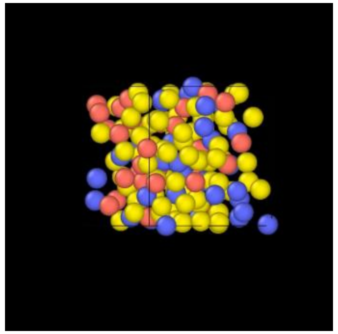
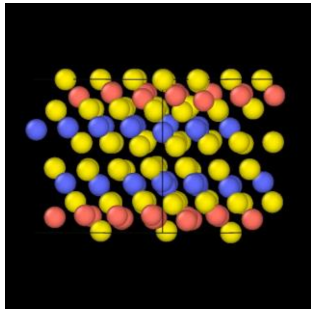
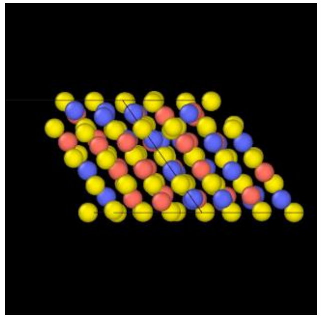
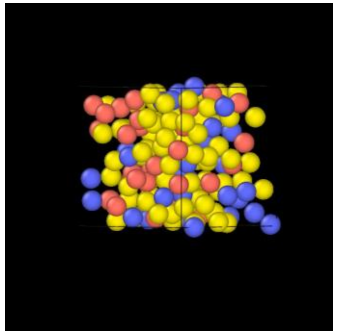
Generation Validation





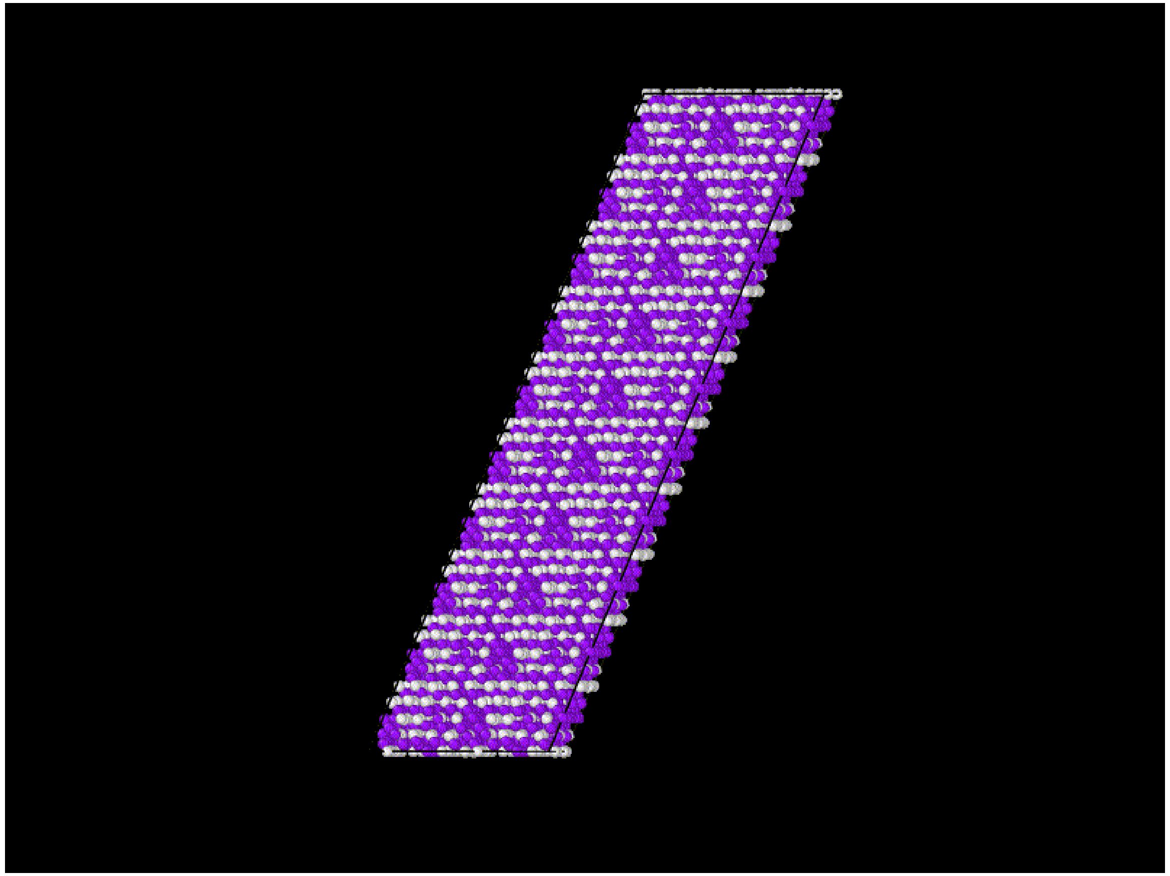
Three Families – Three Phases

Ge Sb Te

	Hexagonal	Cubic	Amorphous
1	 √	 √	 X
2	 √	 √	 √
3	 √	 √	 √

Current State:

~7000 atoms – recrystallizing anneal 600K – 2 ns run



Successes:

- Stability of all three phases in NVT/NPT
- Recrystallization of amorphous into cubic phase shown
- Good agreement with DFT structural representations for xtal/amorphous

Limitations:

- Large scale recrystallization needs a heterogeneous seed
- Liquid temperatures above 1200K become unstable in NPT

Soon to come:
GST+C iterative training

Acknowledgements

Strachan Research Group – Purdue University

Project Colleague – Robert Appleton
PI Alejandro Strachan – Purdue University
Co PI David Adams – Sandia National Laboratory

In part, work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US Department of Energy's National Nuclear Security Administration under Contract No. DE-NA-0003525.

Sandia National Laboratories