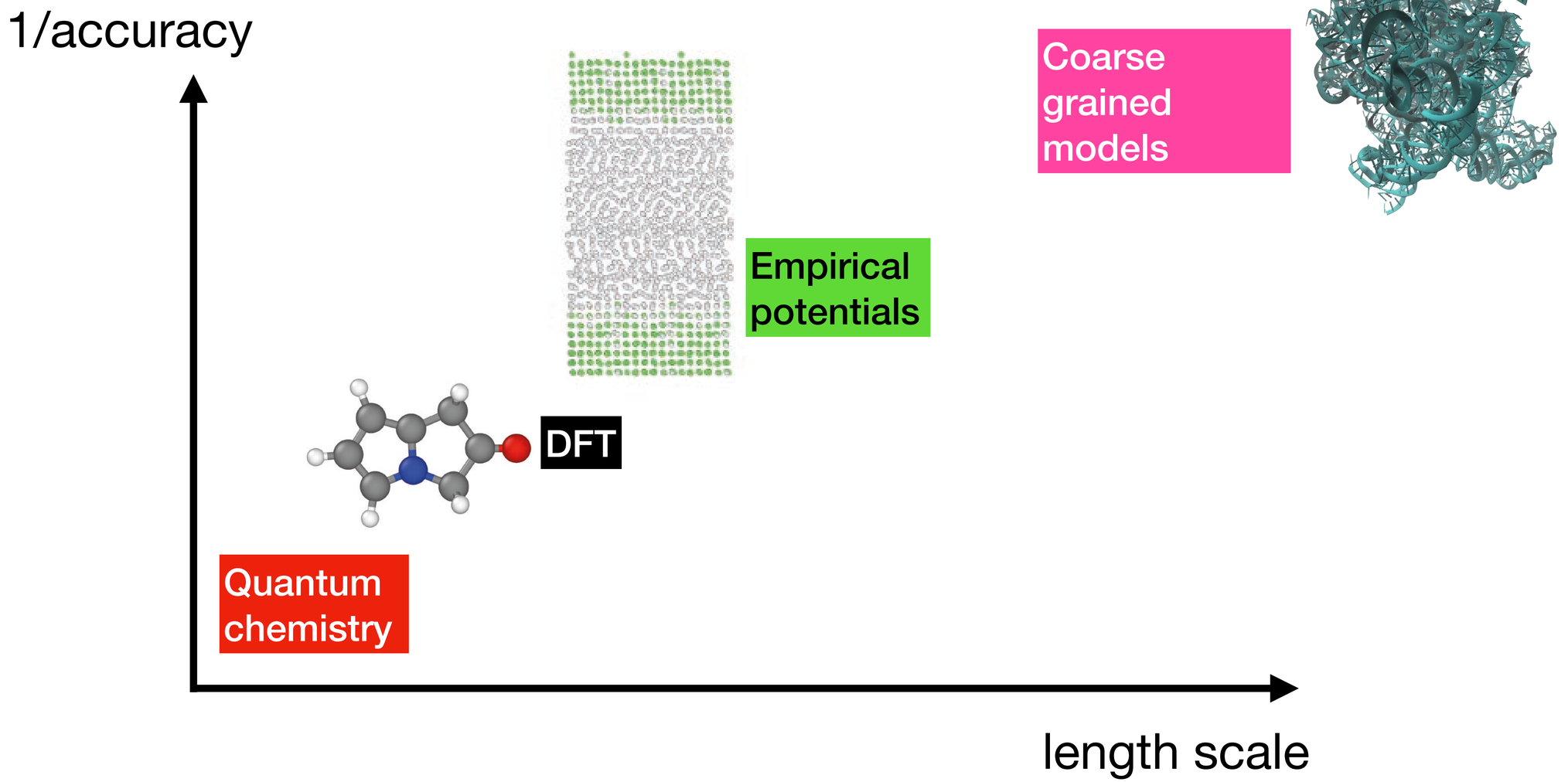
Data-driven approaches to atomistic modeling beyond interatomic potentials

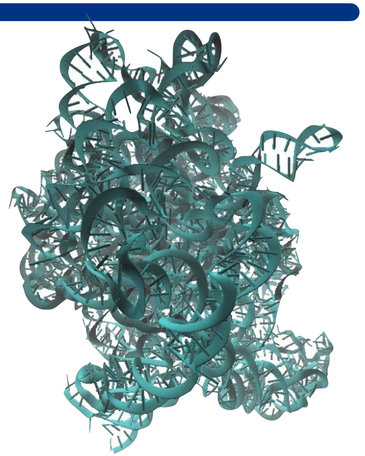
Chiheb Ben Mahmoud

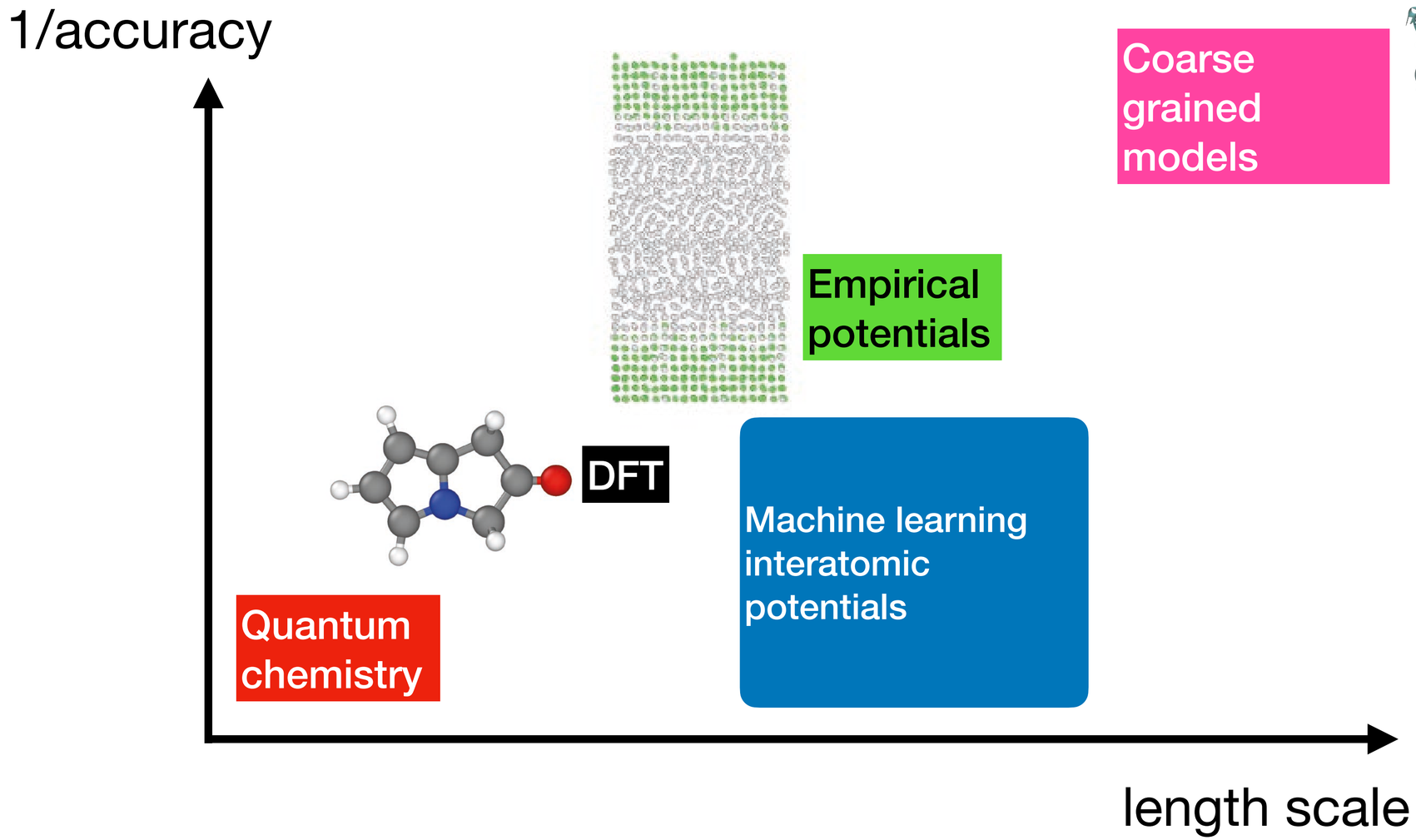
NHR4CES Community Workshop 2025 — Online —07.05.2025

QM results at a fraction of the cost



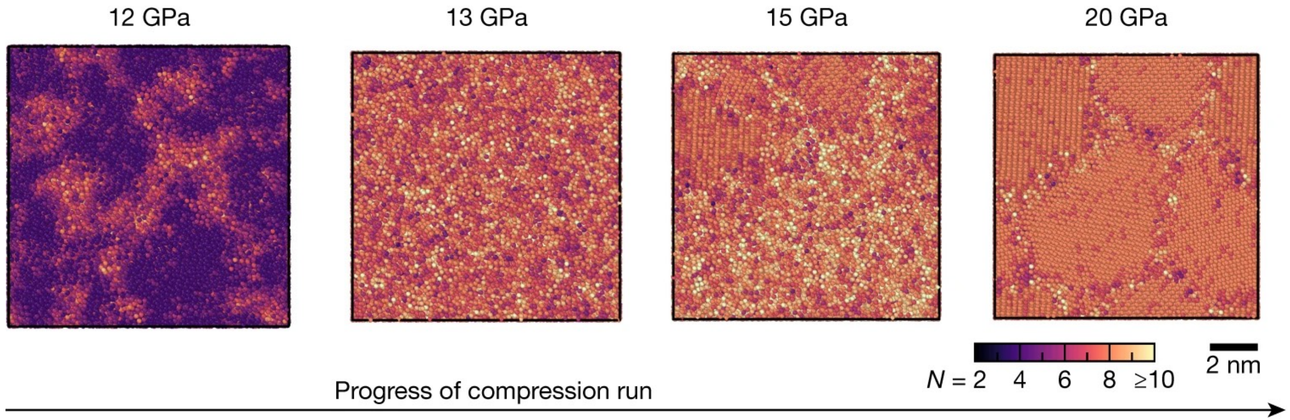
QM results at a fraction of the cost





QM results at a fraction of the cost

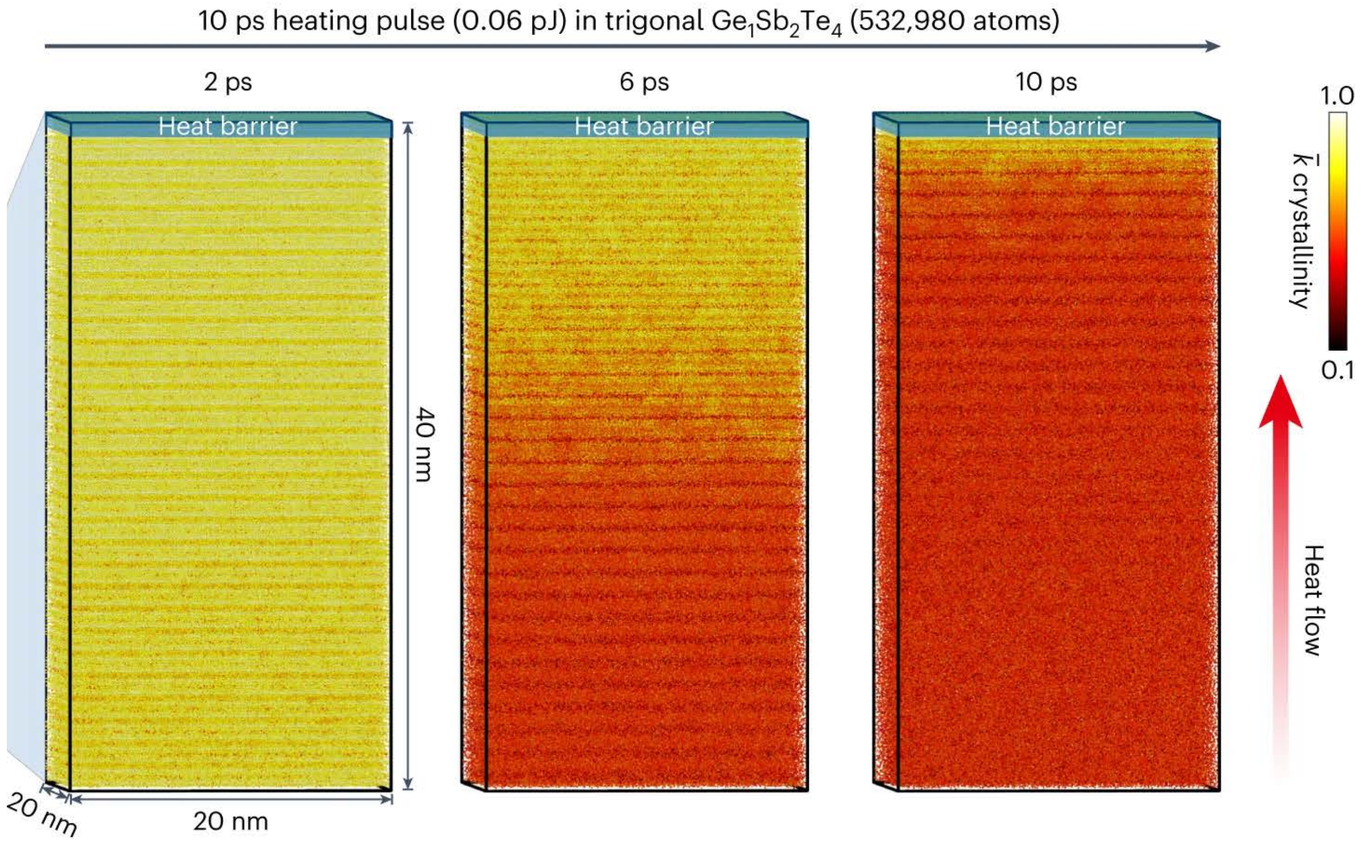
Amorphous Si compression -
100,000 atoms



Deringer et al, 2021

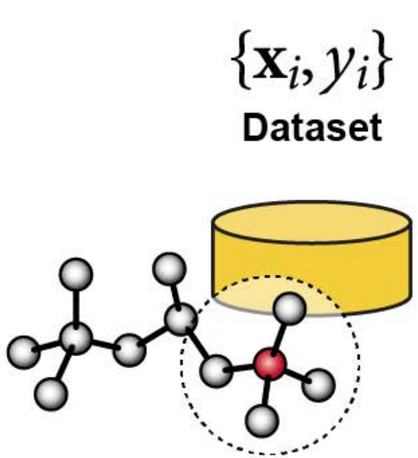
QM results at a fraction of the cost

Heat flow in phase change materials - ~500,000 atoms



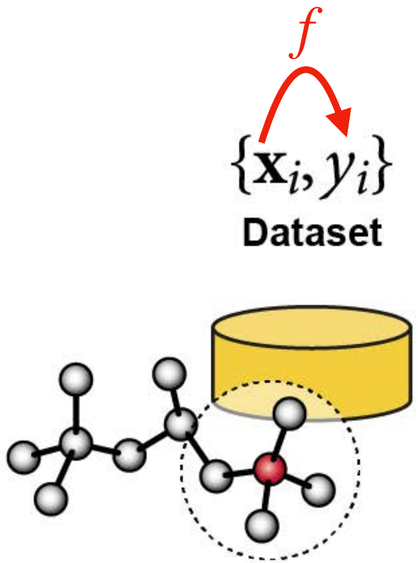
Zhou et al, 2023

Important atomistic ML ingredients



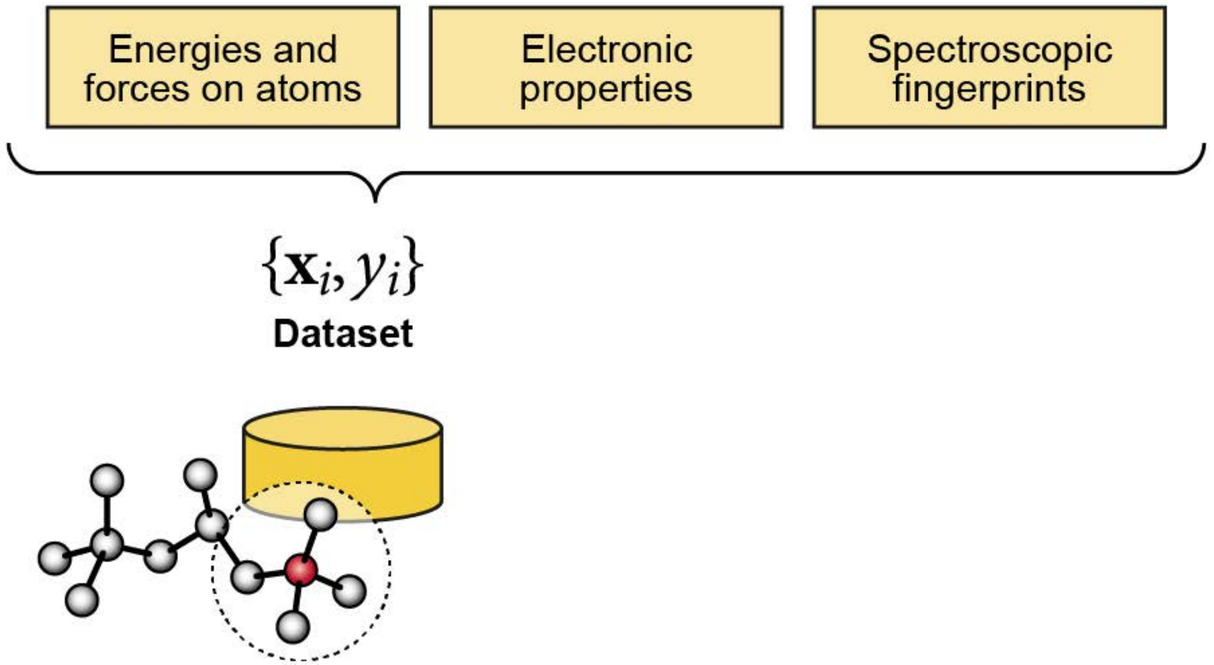
CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Important atomistic ML ingredients



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

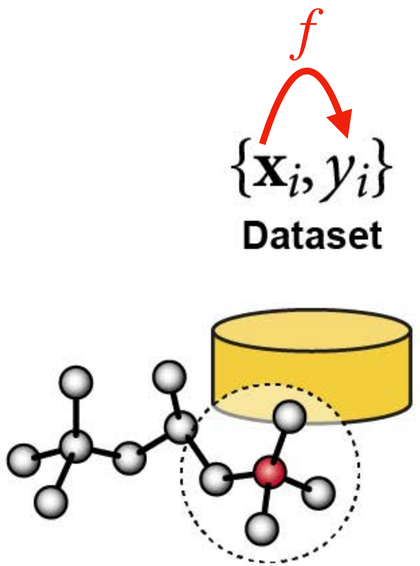
Important atomistic ML ingredients



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Important atomistic ML ingredients

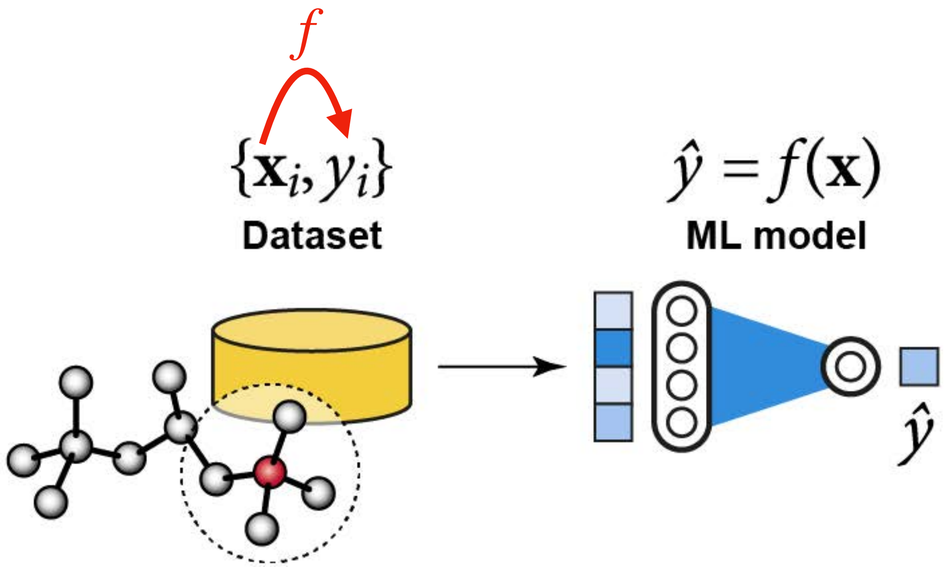
Empirical model Quantum mecahnics



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Important atomistic ML ingredients

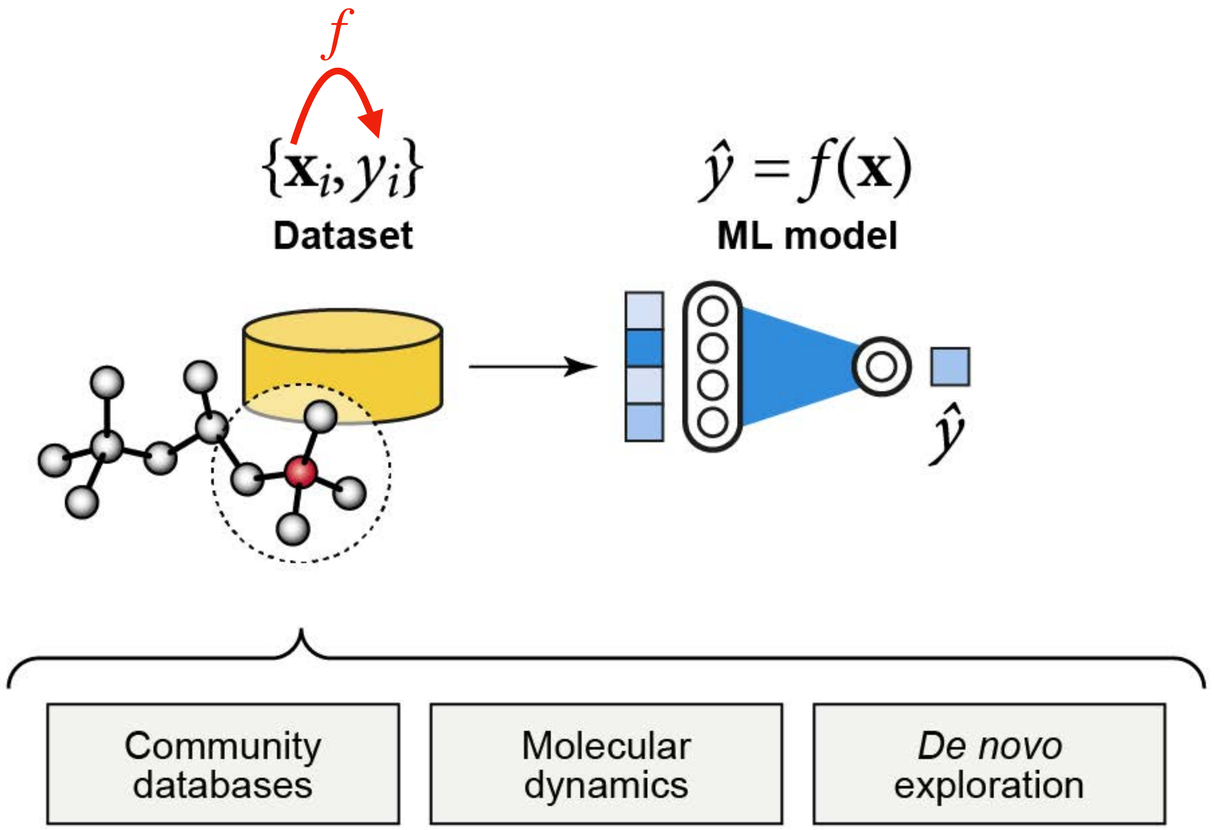
Empirical model Quantum mecahnics



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Important atomistic ML ingredients

Empirical model Quantum mecahnics



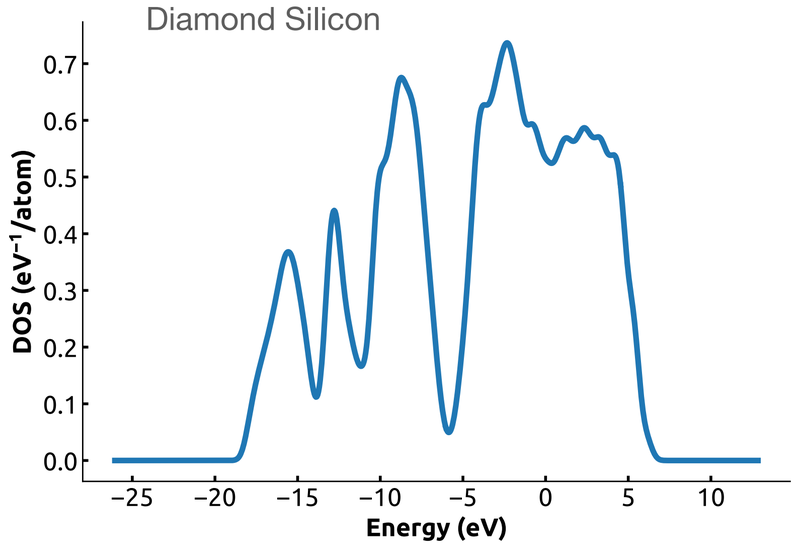
CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

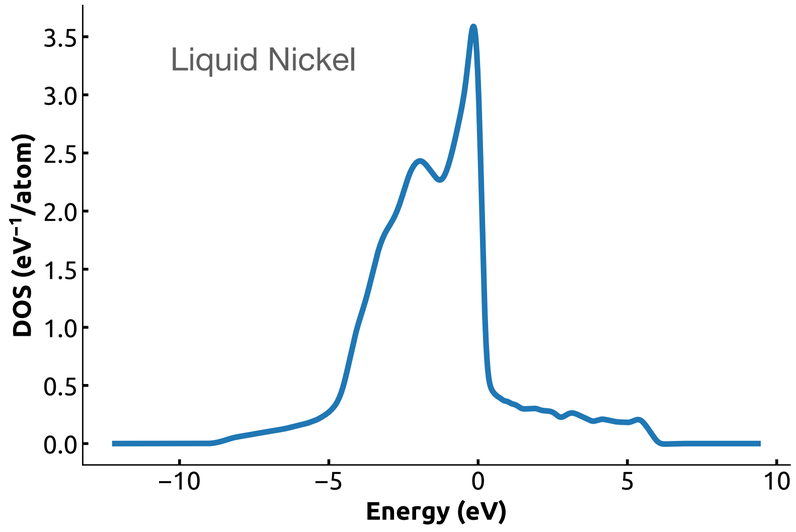
Property 1:

electronic density of states

Property 1: Electronic density of states

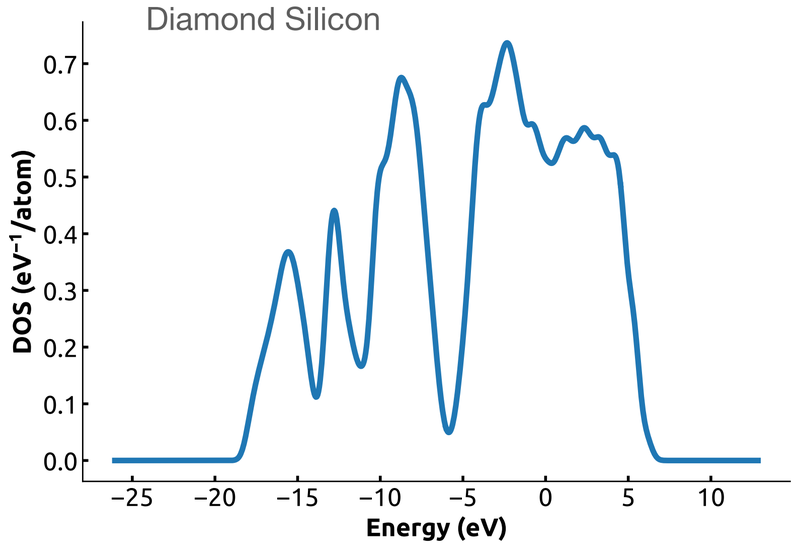
- The electronic density of states (DOS) is useful for:

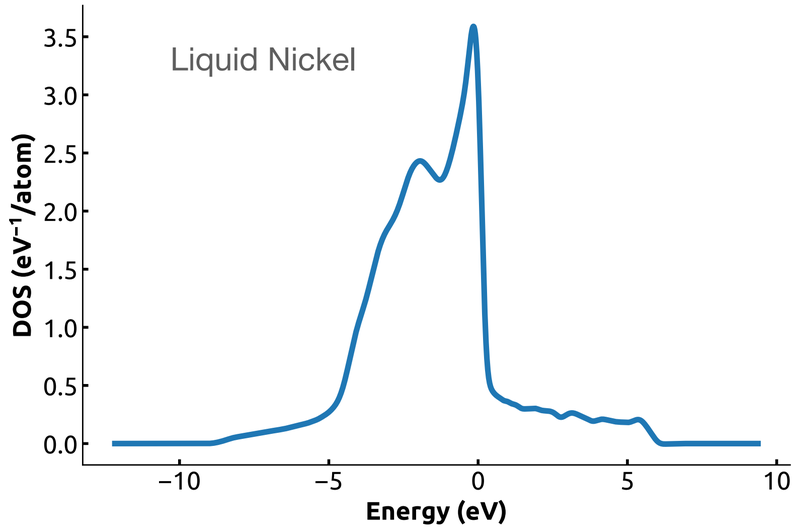




Property 1: Electronic density of states

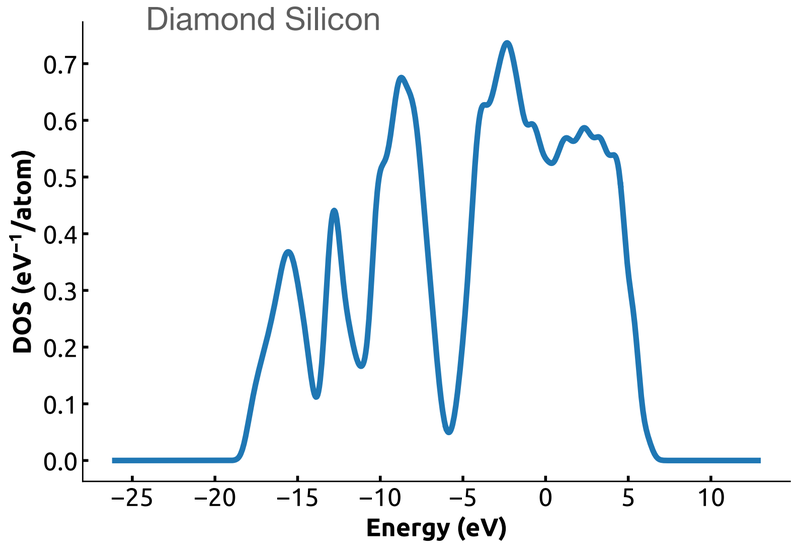
- The electronic density of states (DOS) is useful for:
  - **optical properties**: band gap and joint density of states

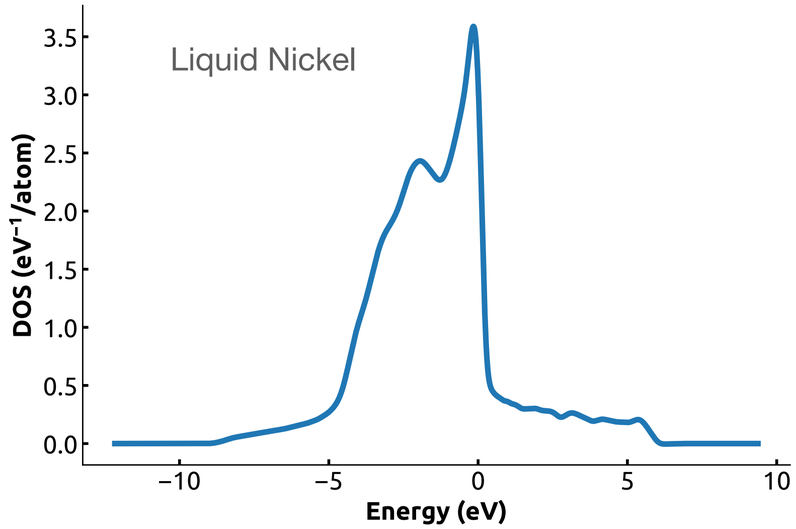




Property 1: Electronic density of states

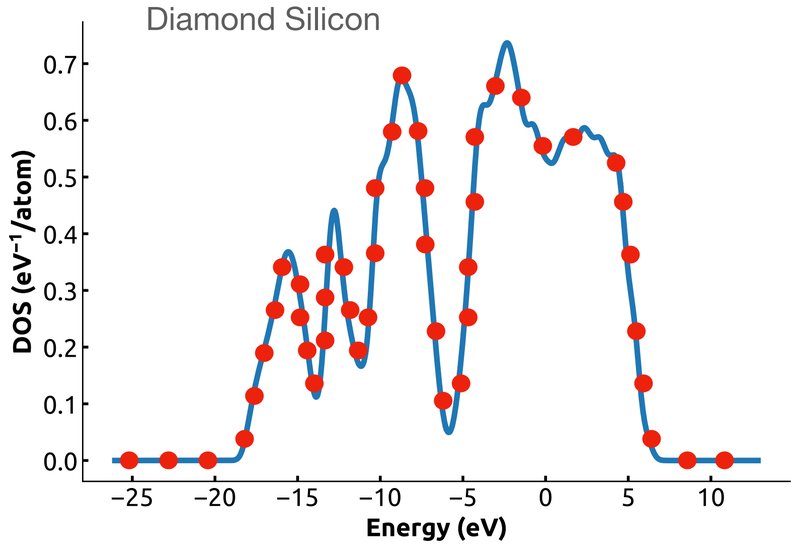
- The electronic density of states (DOS) is useful for:
  - **optical properties**: band gap and joint density of states
  - **thermal properties**: heat capacity of metals

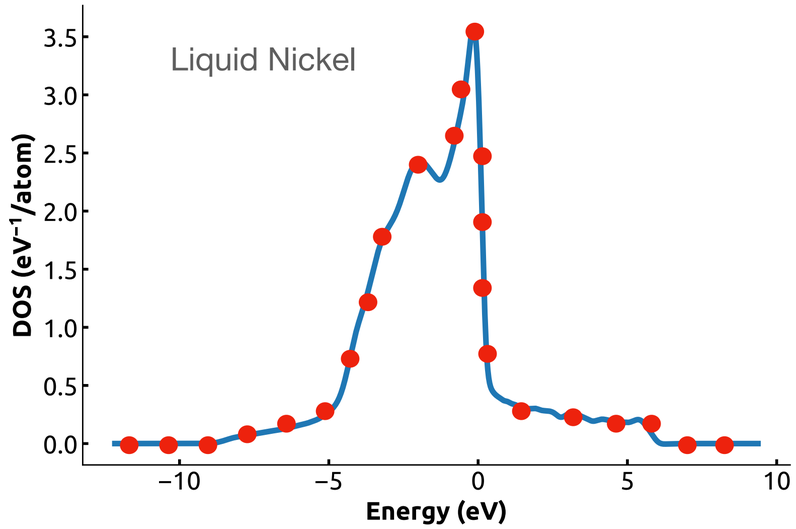




Property 1: Electronic density of states

- The electronic density of states (DOS) is useful for:
  - **optical properties**: band gap and joint density of states
  - **thermal properties**: heat capacity of metals





Objectives

- Build a **transferable ML model** for the DOS
- Identify the **challenges** in modelling and interpreting the DOS
- Apply the ML DOS models to study :
  - the **feature/property** relations
  - thermal properties of materials in physics-based approximations

Atom-centred model for the DOS

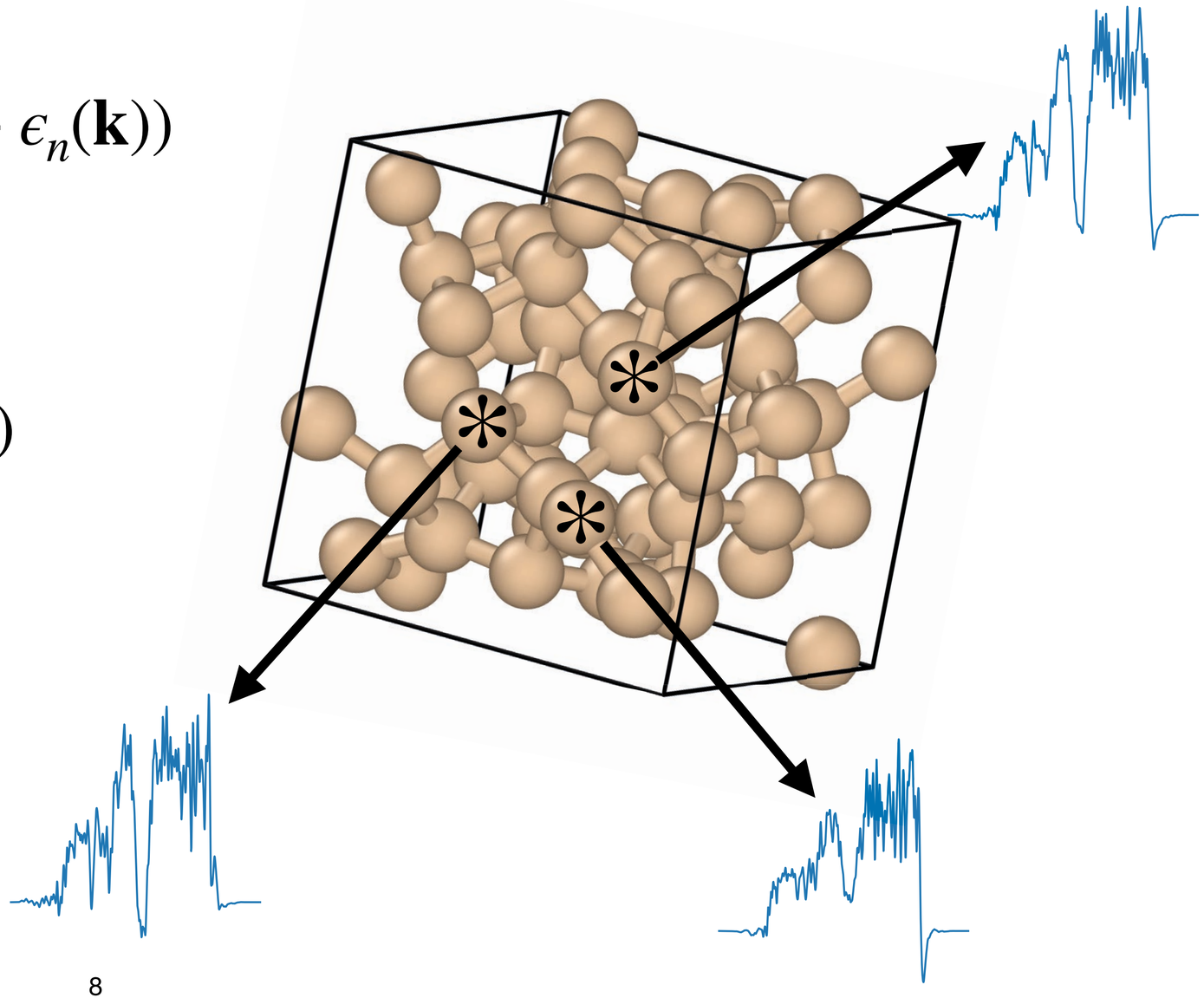
- $\mathrm{DOS}(\mathscr{A}, E) = \frac{2}{N_k} \sum_{n}^{bands} \sum_{\mathbf{k}} \tilde{\delta}(E - \epsilon_n(\mathbf{k}))$

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

Atom-centred model for the DOS

- $\mathrm{DOS}(\mathscr{A}, E) = \dfrac{2}{N_k} \sum_{n}^{bands} \sum_{\mathbf{k}} \tilde{\delta}(E - \epsilon_n(\mathbf{k}))$

- $\mathrm{DOS}(\mathscr{A}, \epsilon) = \sum_{\mathscr{A}_j \in \mathscr{A}} \mathrm{LDOS}(\mathscr{A}_j, \epsilon)$



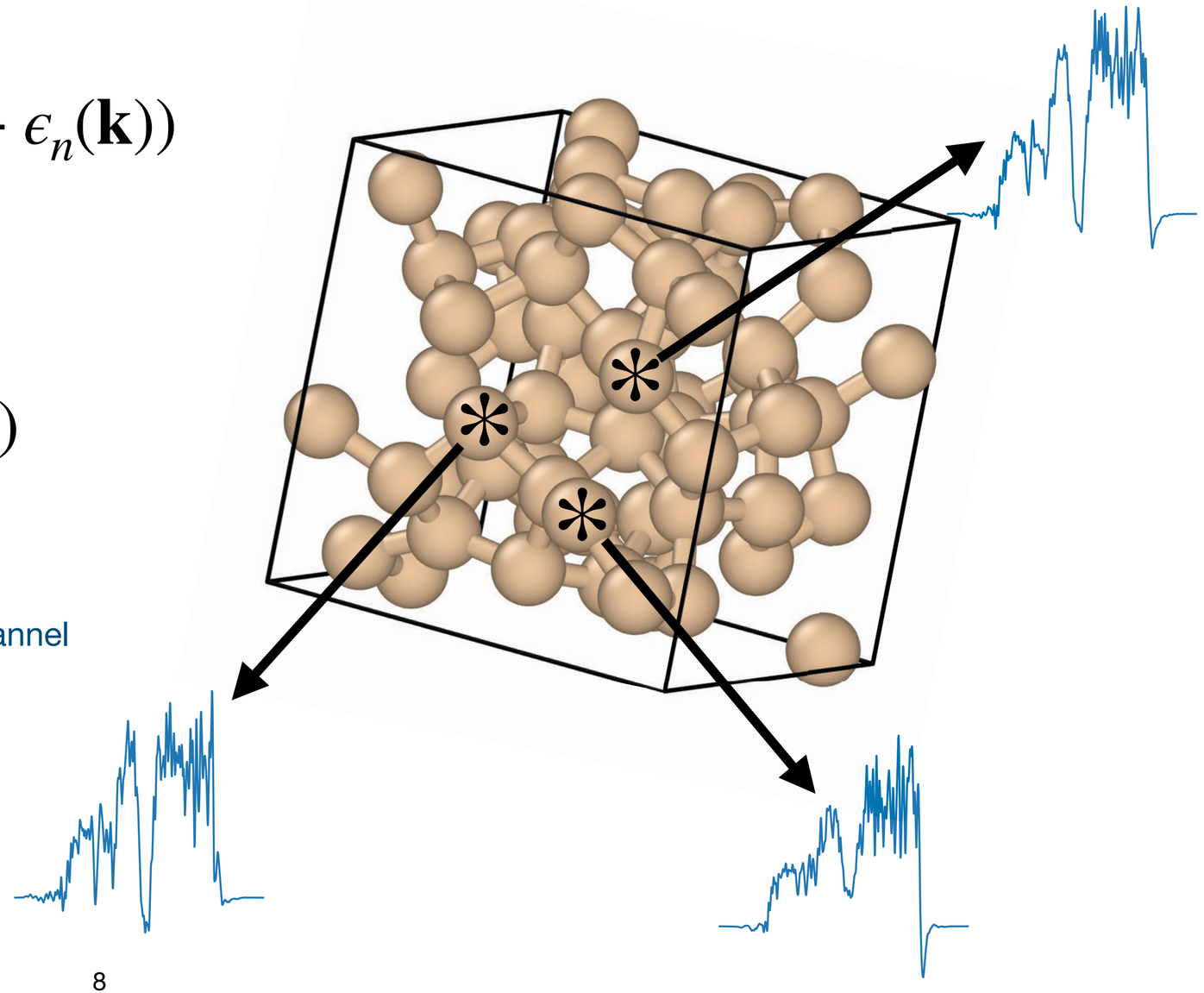
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

Atom-centred model for the DOS

- $\mathrm{DOS}(\mathscr{A}, E) = \frac{2}{N_k} \sum_{n}^{bands} \sum_{\mathbf{k}} \tilde{\delta}(E - \epsilon_n(\mathbf{k}))$

- $\mathrm{DOS}(\mathscr{A}, \epsilon) = \sum_{\mathscr{A}_j \in \mathscr{A}} \mathrm{LDOS}(\mathscr{A}_j, \epsilon)$

Build a machine learning model for each energy channel



CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

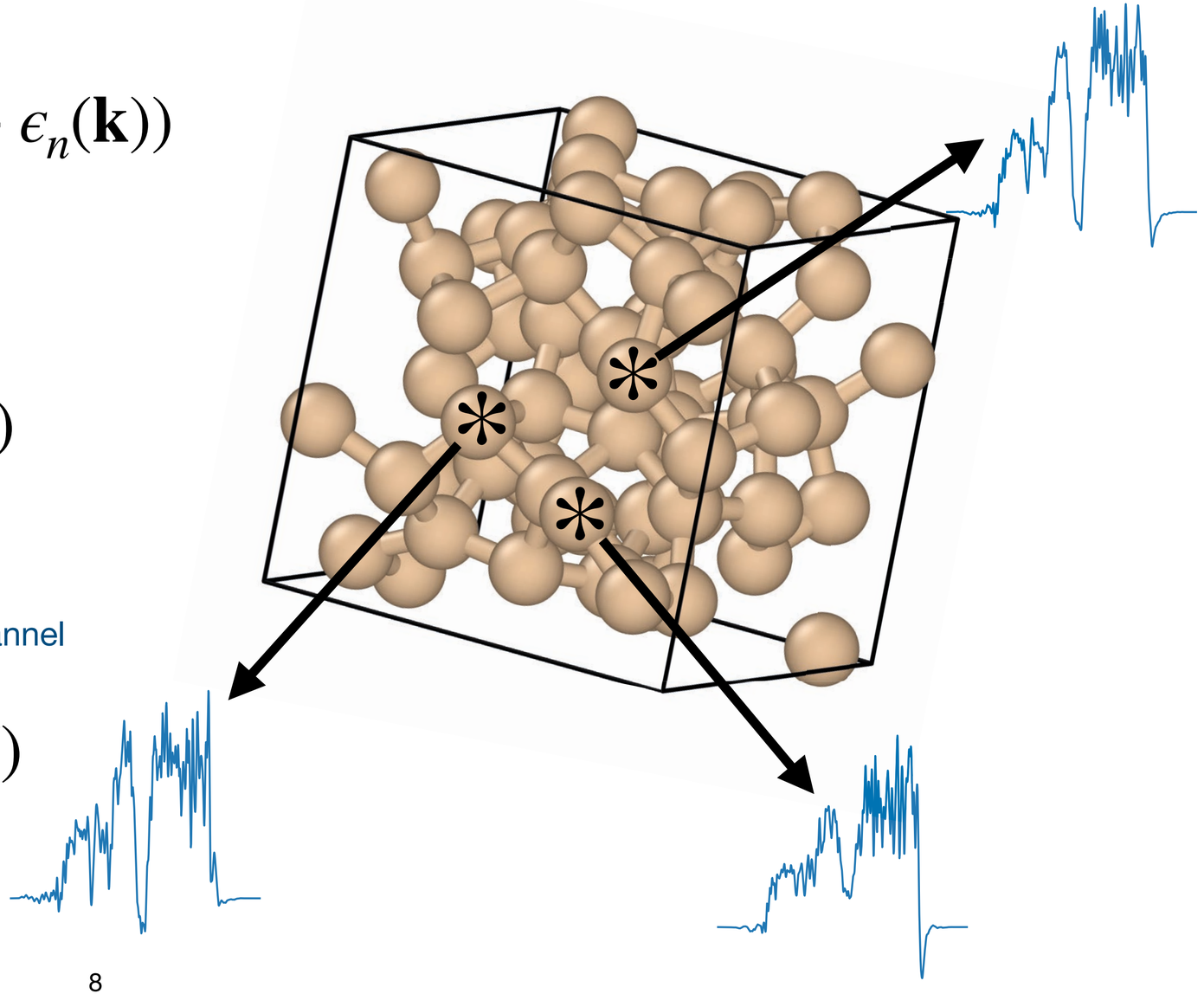
Atom-centred model for the DOS

- $\mathrm{DOS}(\mathscr{A}, E) = \frac{2}{N_k} \sum_{n}^{bands} \sum_{\mathbf{k}} \tilde{\delta}(E - \epsilon_n(\mathbf{k}))$

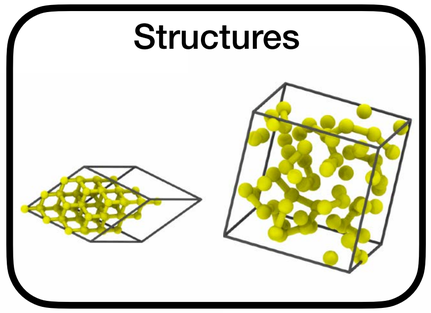
- $\mathrm{DOS}(\mathscr{A}, \epsilon) = \sum_{\mathscr{A}_j \in \mathscr{A}} \mathrm{LDOS}(\mathscr{A}_j, \epsilon)$

Build a machine learning model for each energy channel

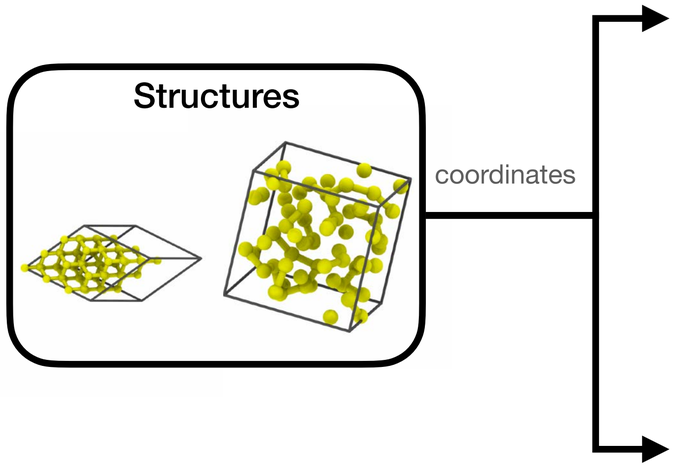
- $\mathrm{DOS}(\mathscr{A}, \epsilon) = \sum_{\mathscr{A}_j \in \mathscr{A}} \mathbf{x_j}(\epsilon) k(\mathscr{A}_j, M)$



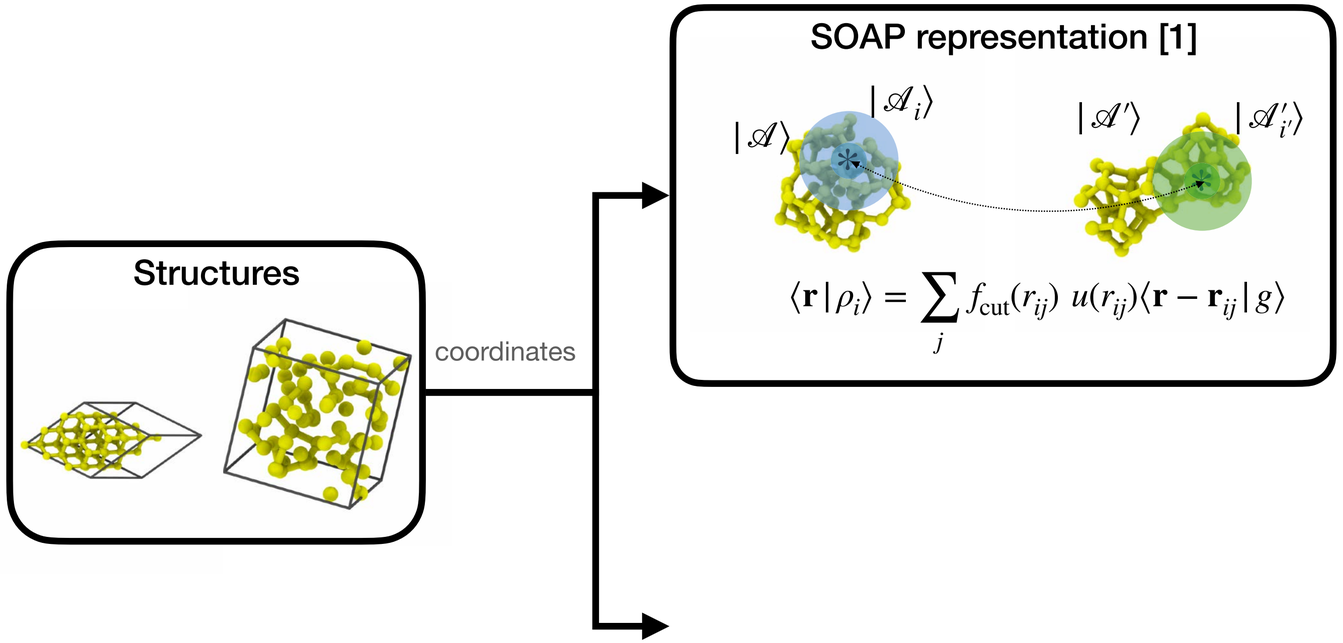
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter.
Phys. Rev. B 102, 235130

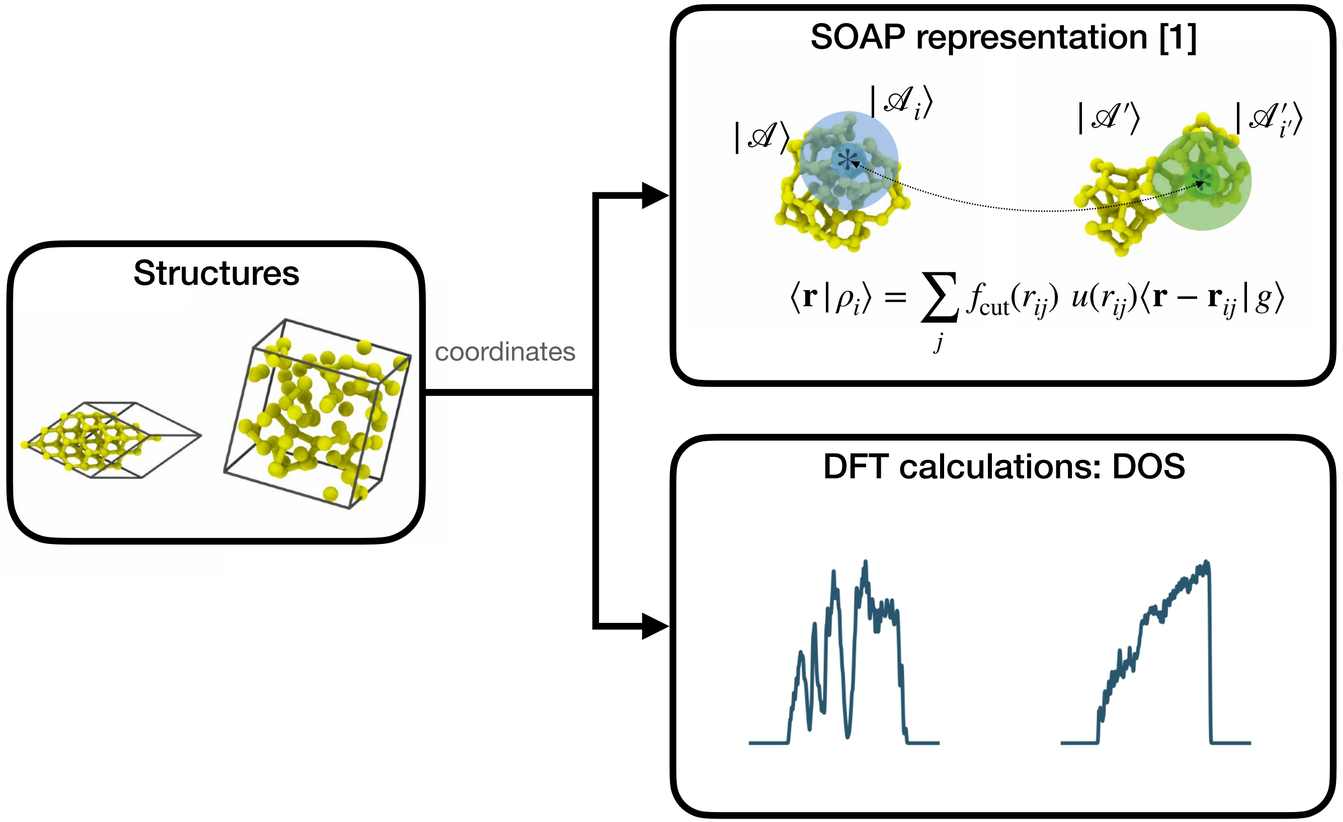


CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter.
Phys. Rev. B 102, 235130



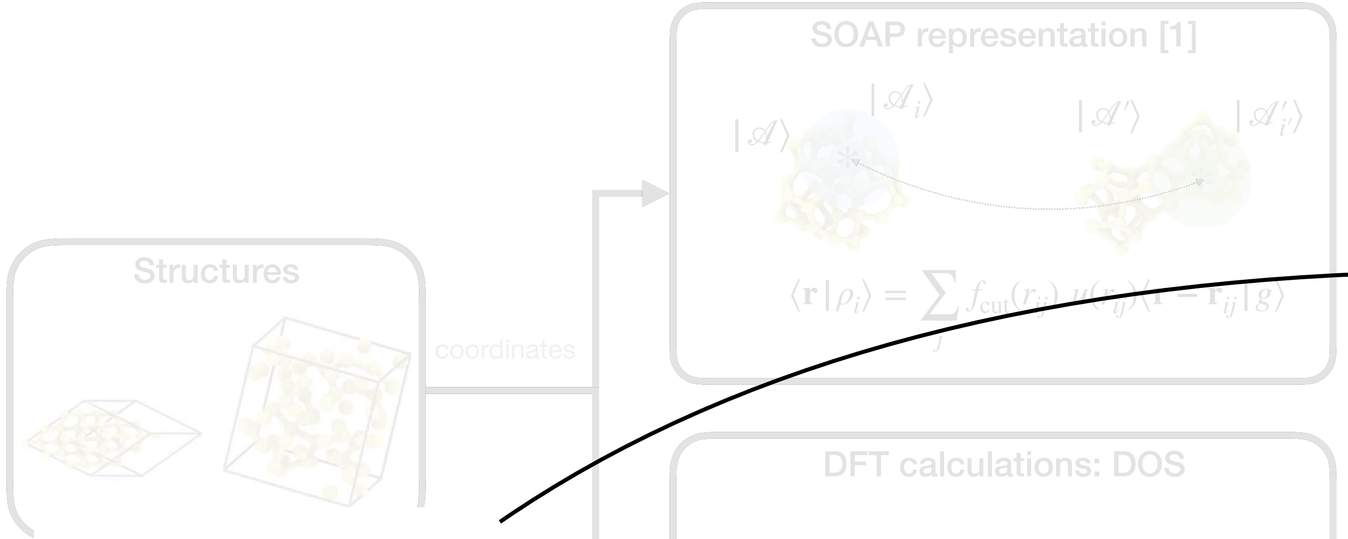
[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

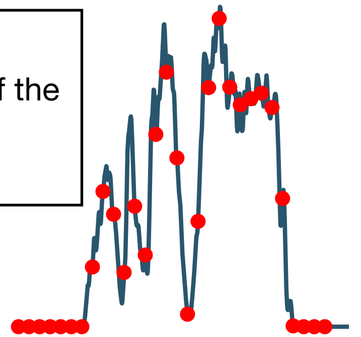


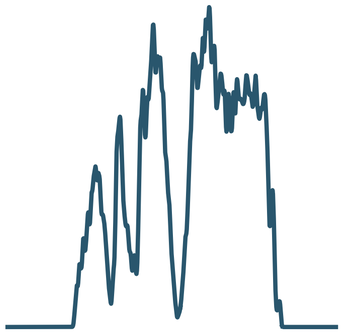
[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



Discretisation of the DOS curve



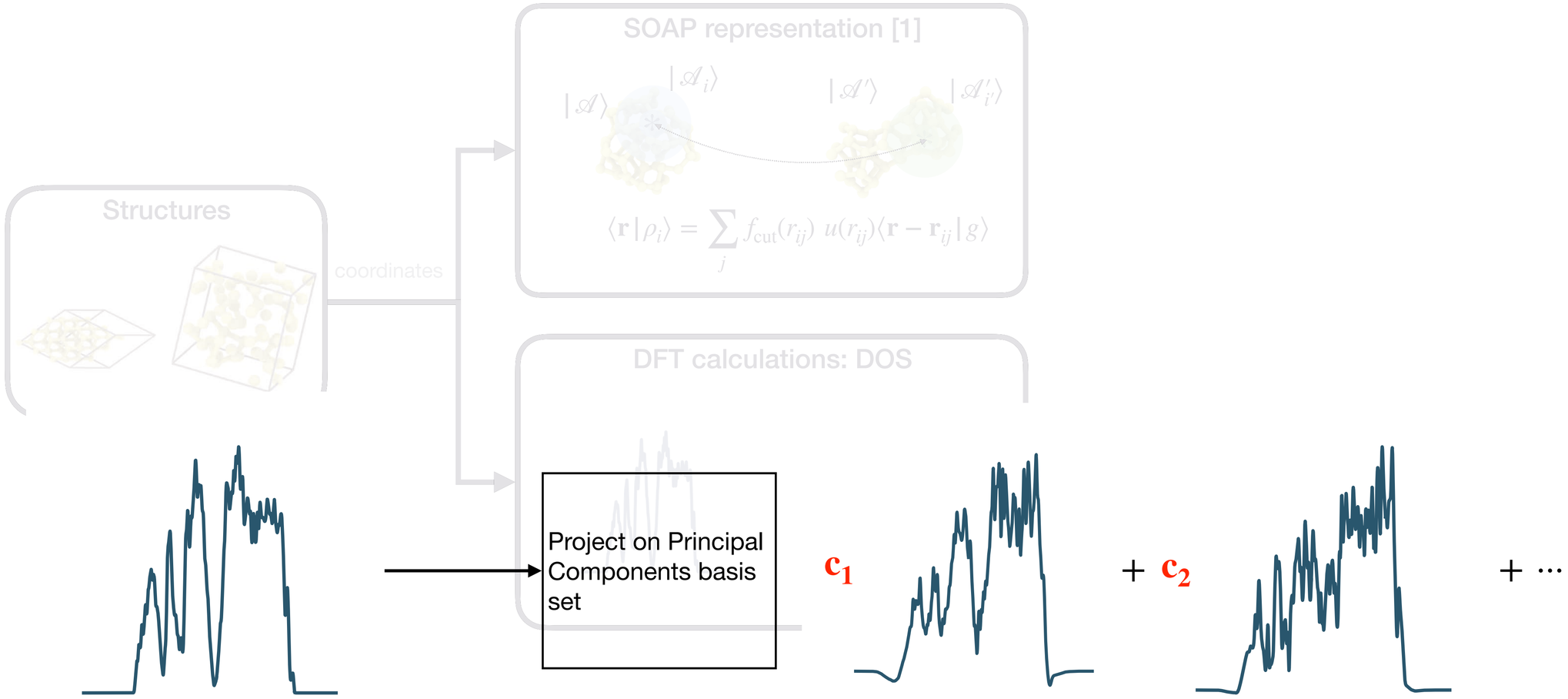


DOS

- We build a ML model for every point
- Non continuous prediction

[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



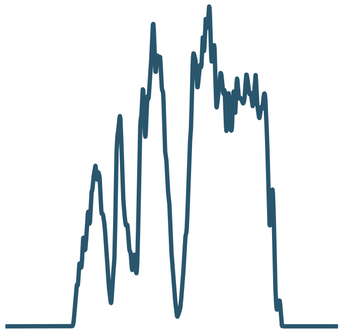
DOS

- We build a ML model for every coefficient
- Continuous prediction
- Extra smoothening

[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

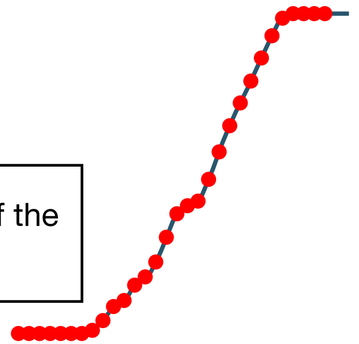
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

- DOS is a histogram
- Compute its cumulative distribution function (CDF)
- We build a model for every point
- Better sensitivity to the shift of the peaks



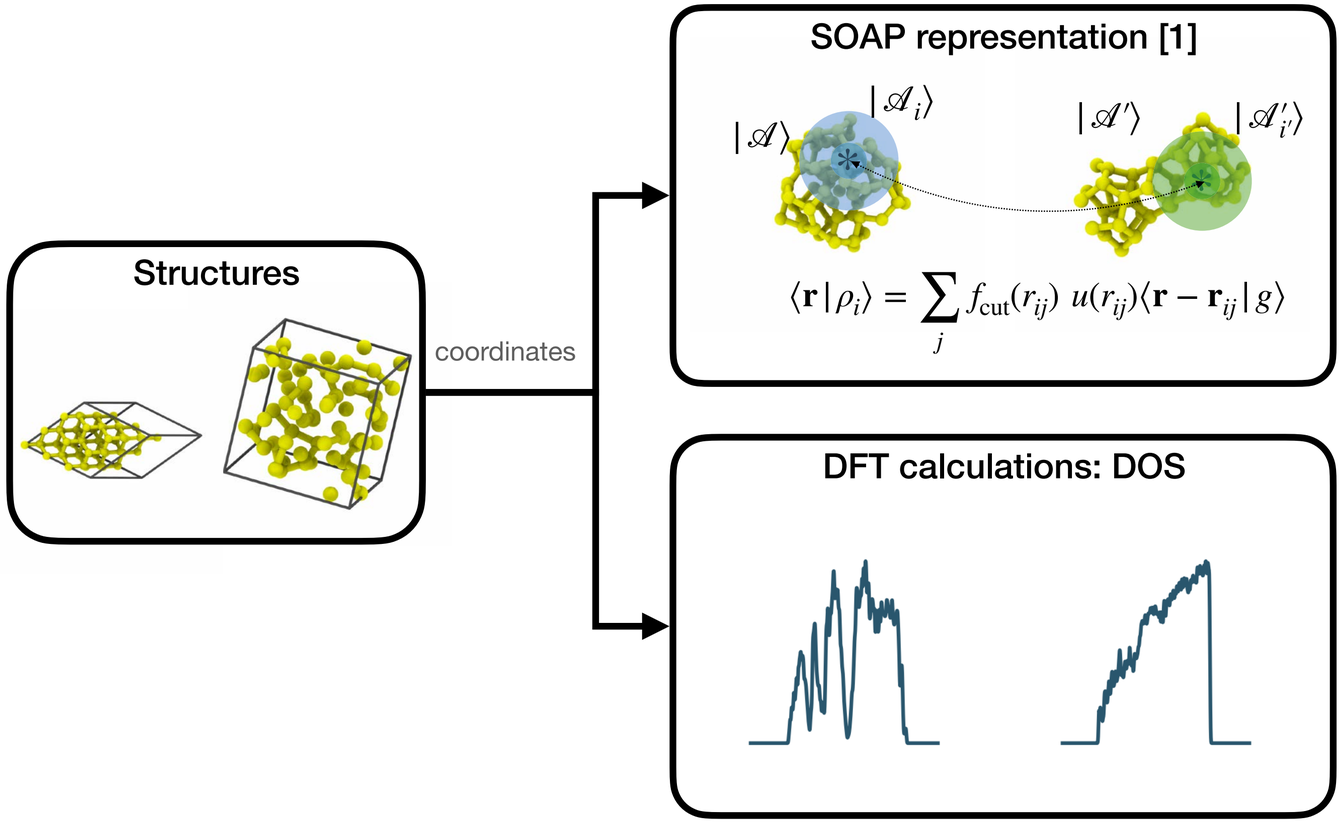
DOS

Discretization of the CDF



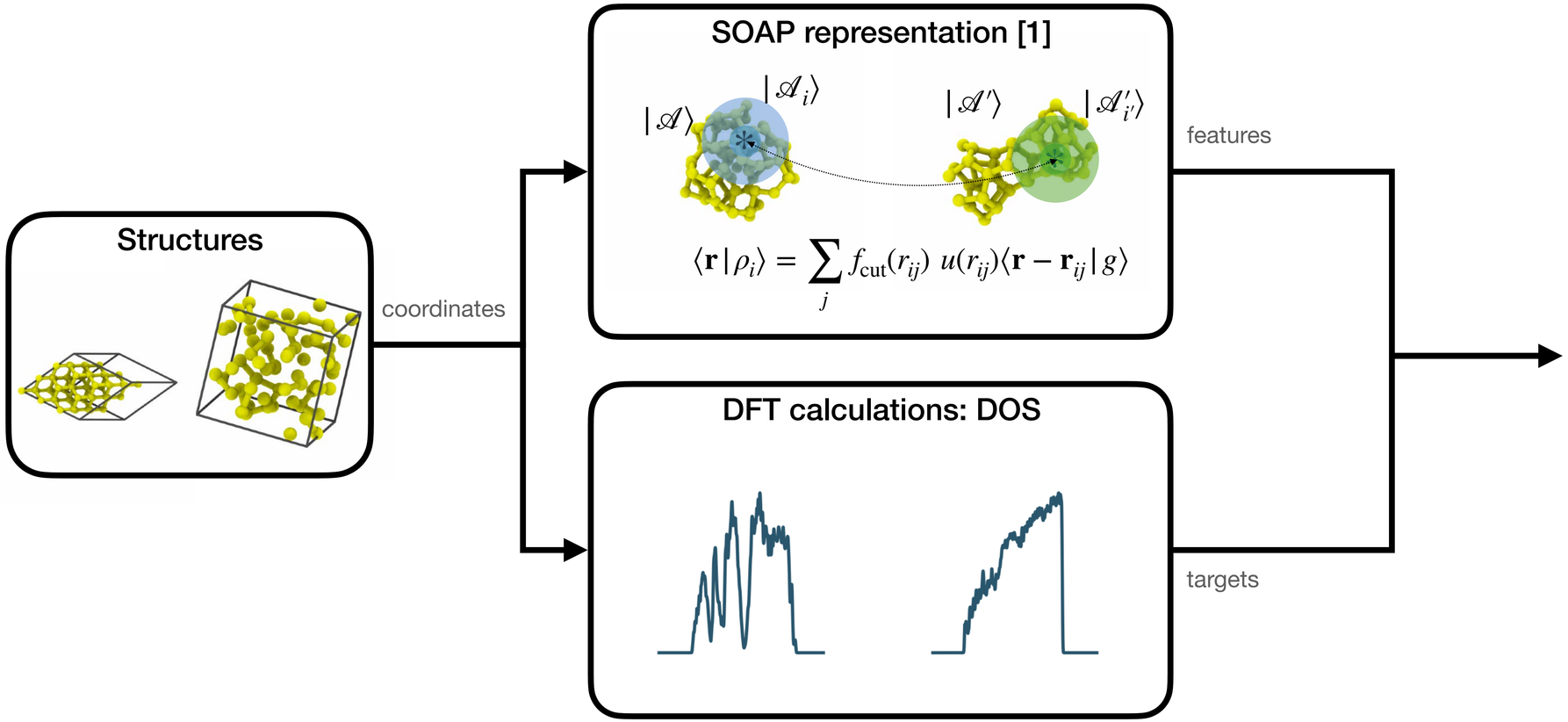
[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter.
Phys. Rev. B 102, 235130



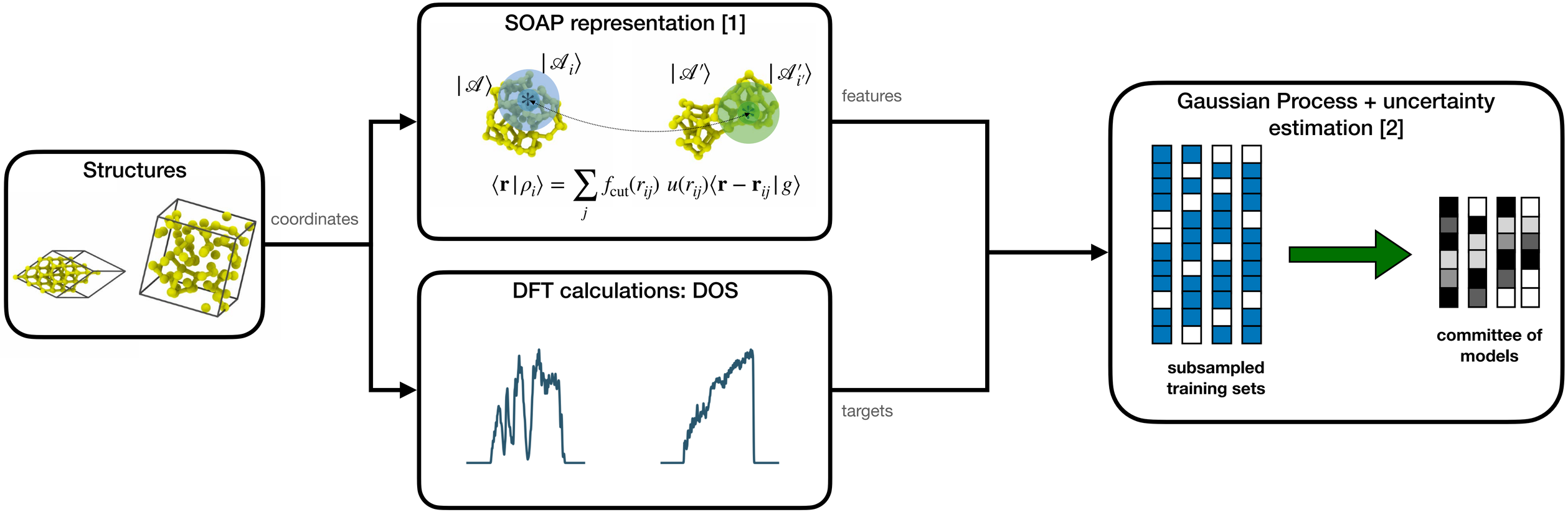
[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

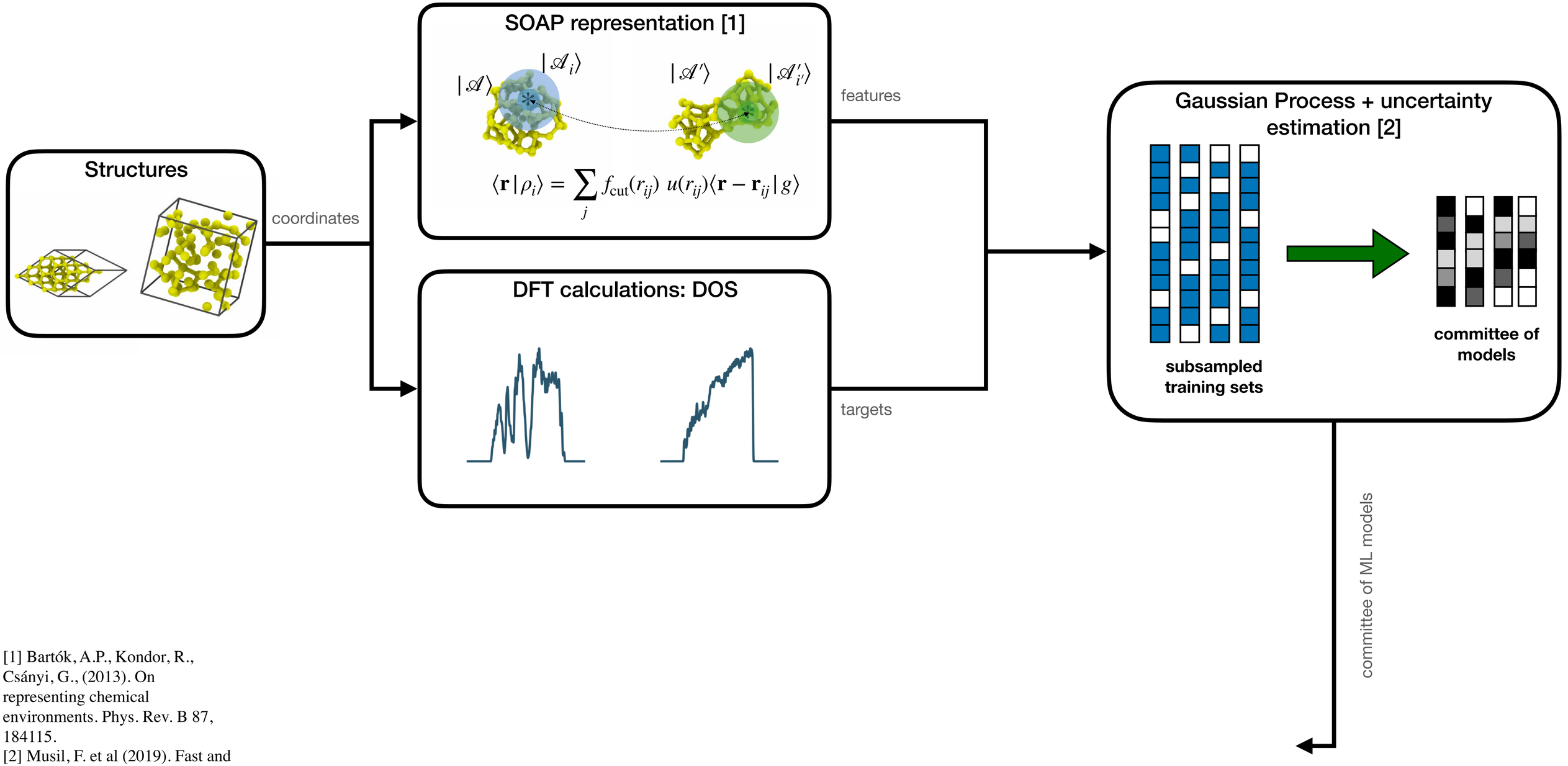
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

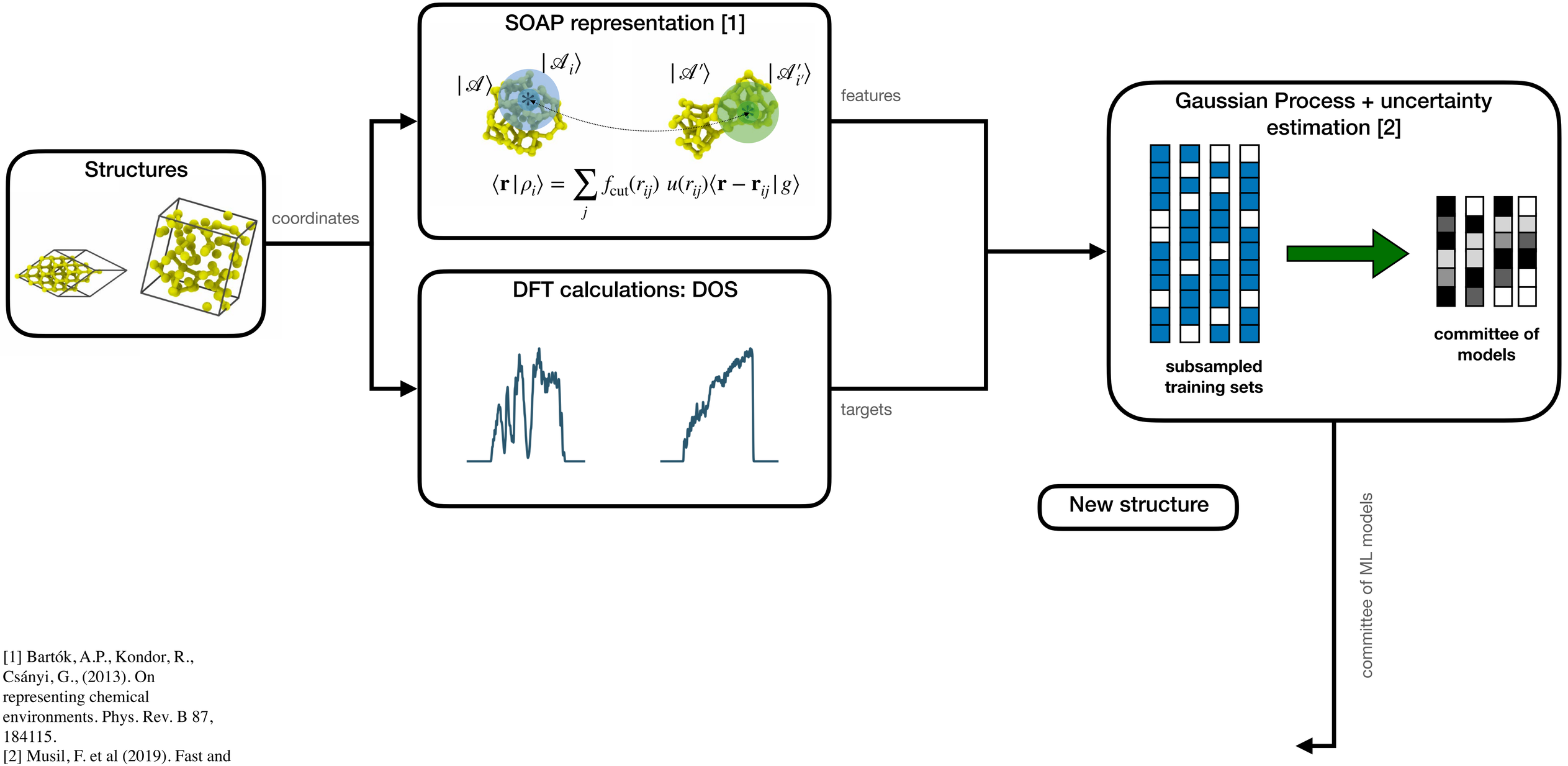
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

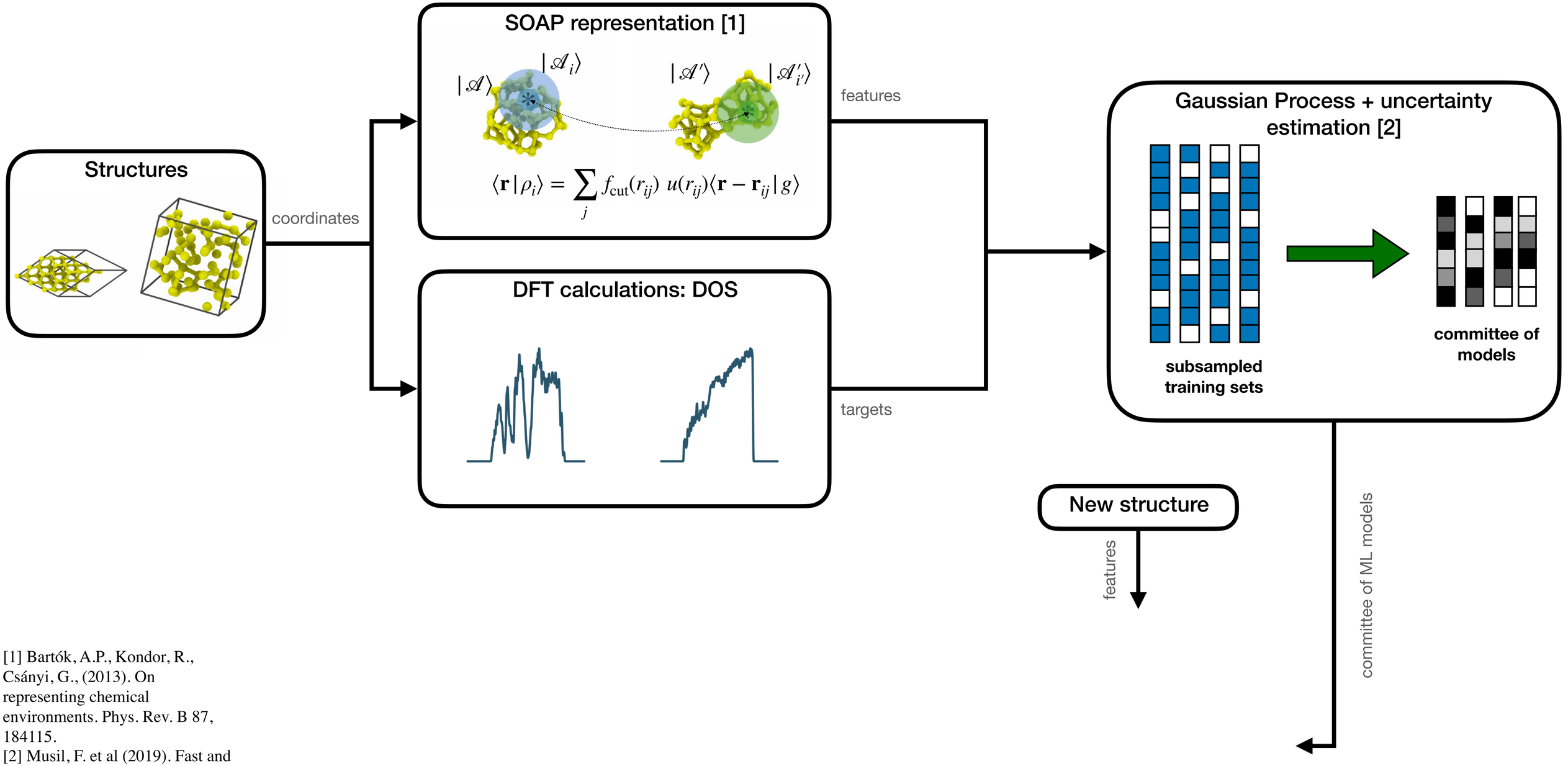
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

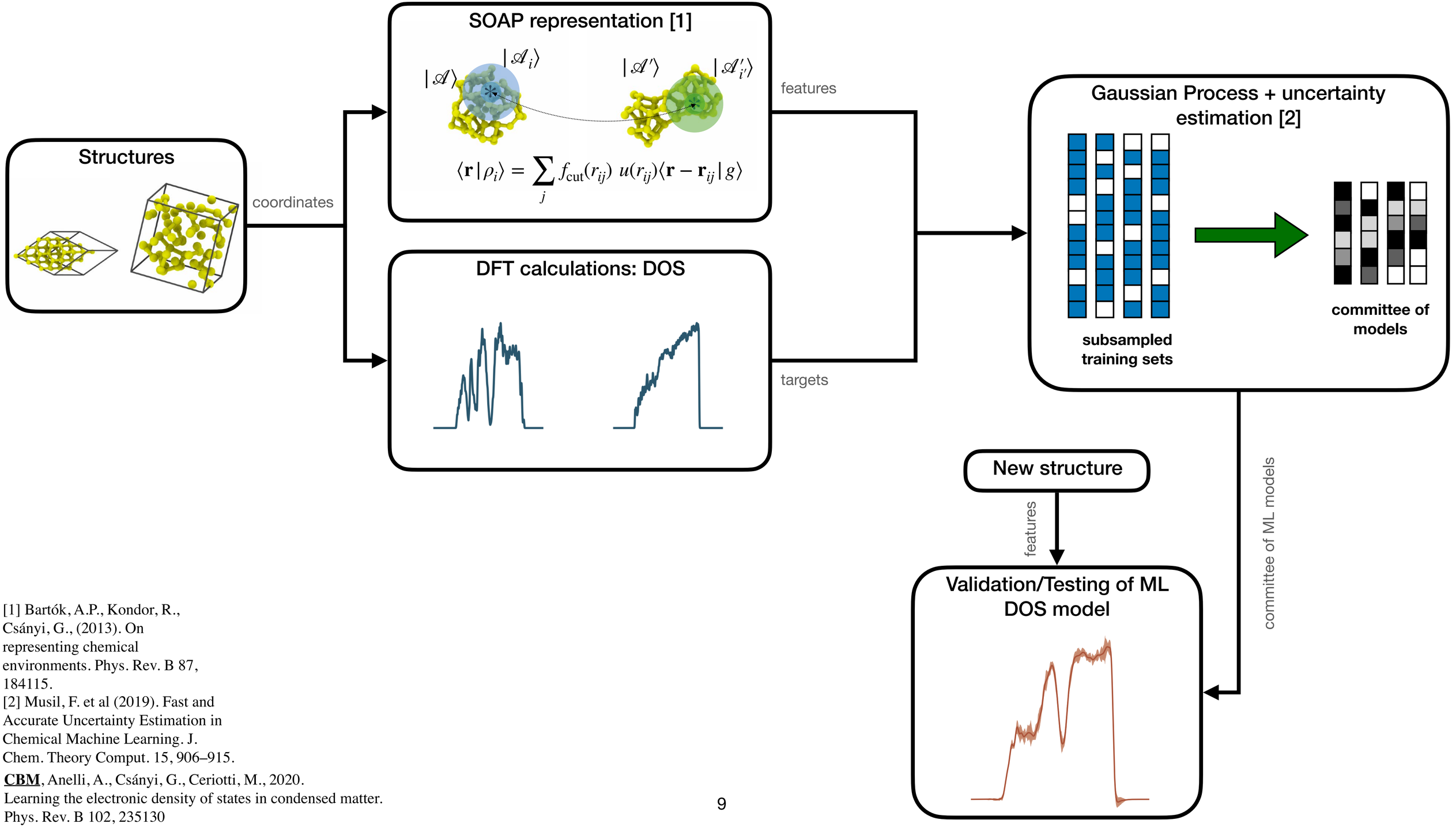
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

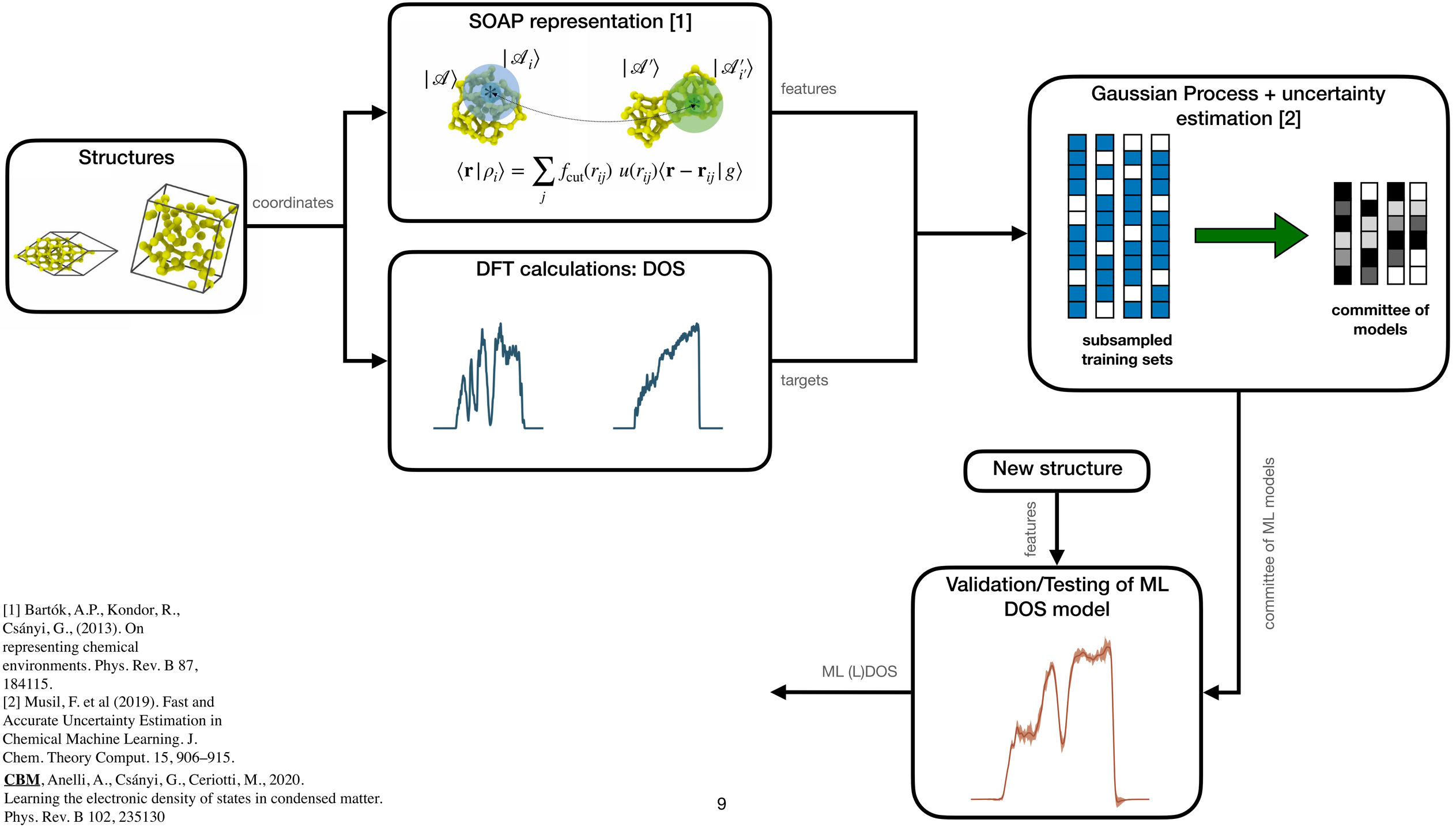
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130



[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

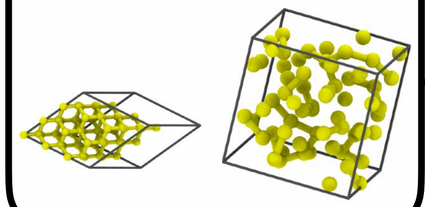


[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

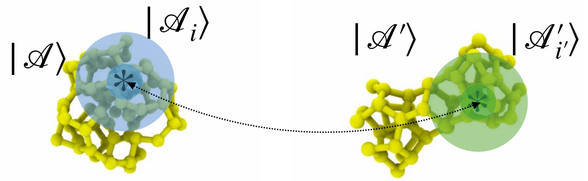
CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

Structures



coordinates

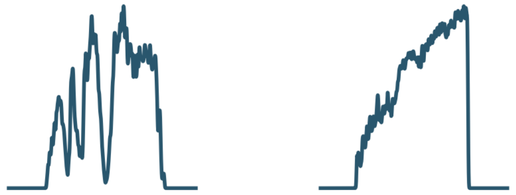
SOAP representation [1]



$$\langle \mathbf{r} \,|\, \rho_i \rangle = \sum_j f_{\text{cut}}(r_{ij})\; u(r_{ij}) \langle \mathbf{r} - \mathbf{r}_{ij} \,|\, g \rangle$$

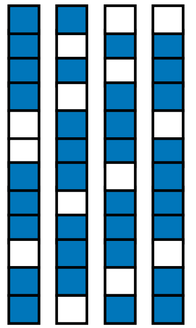
features

DFT calculations: DOS

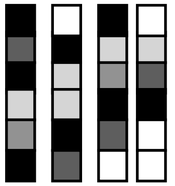


targets

Gaussian Process + uncertainty estimation [2]



subsampled training sets



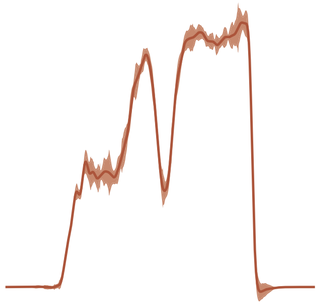
committee of models

committee of ML models

New structure

features

Validation/Testing of ML DOS model



ML (L)DOS

Postprocessing

1. structure/property correlations
2. band energy and heat capacity
3. finite-temperature potentials

[1] Bartók, A.P., Kondor, R., Csányi, G., (2013). On representing chemical environments. Phys. Rev. B 87, 184115.

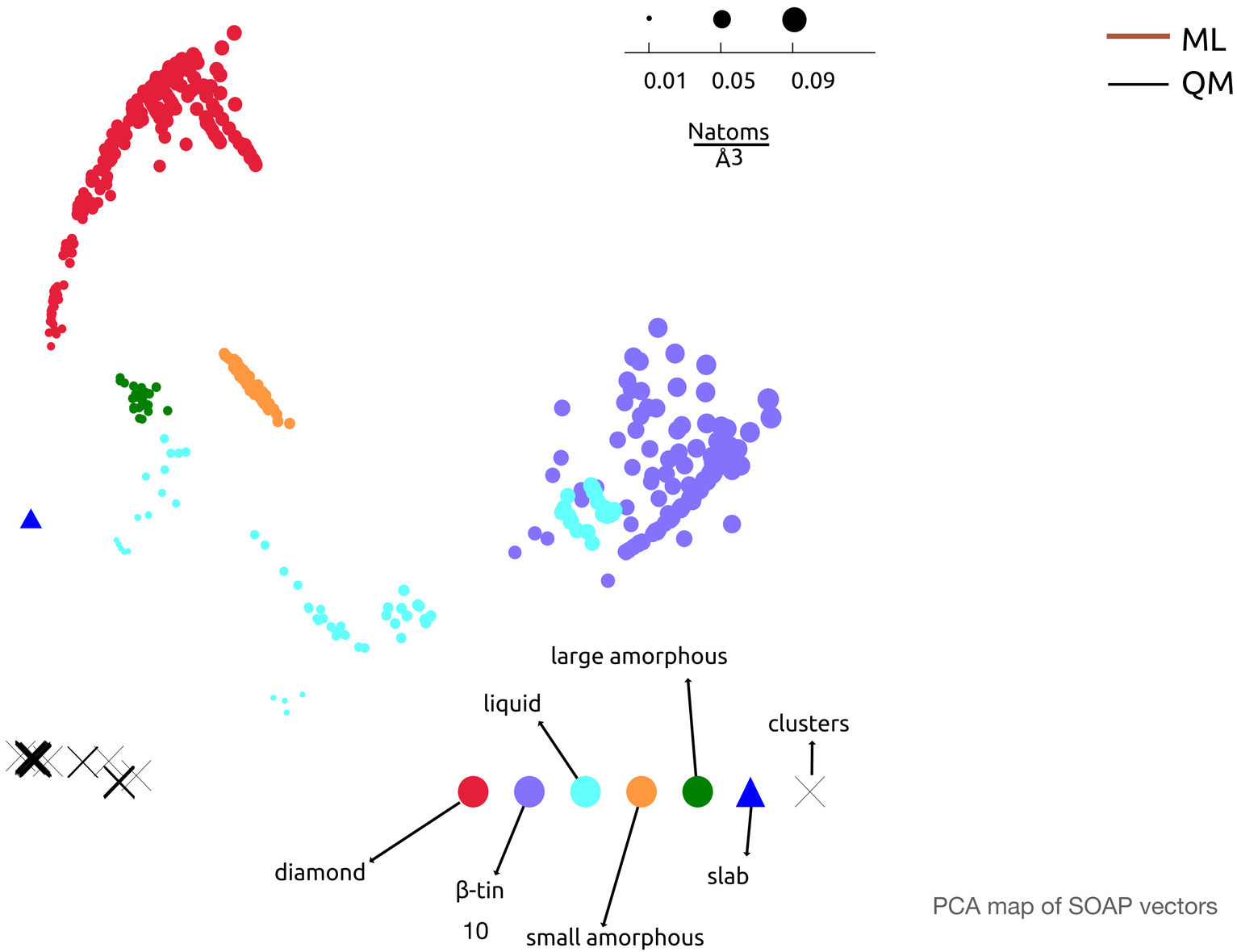
[2] Musil, F. et al (2019). Fast and Accurate Uncertainty Estimation in Chemical Machine Learning. J. Chem. Theory Comput. 15, 906–915.

CBM, Anelli, A., Csányi, G., Ceriotti, M., 2020. Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130

Benchmark on diverse silicon data set

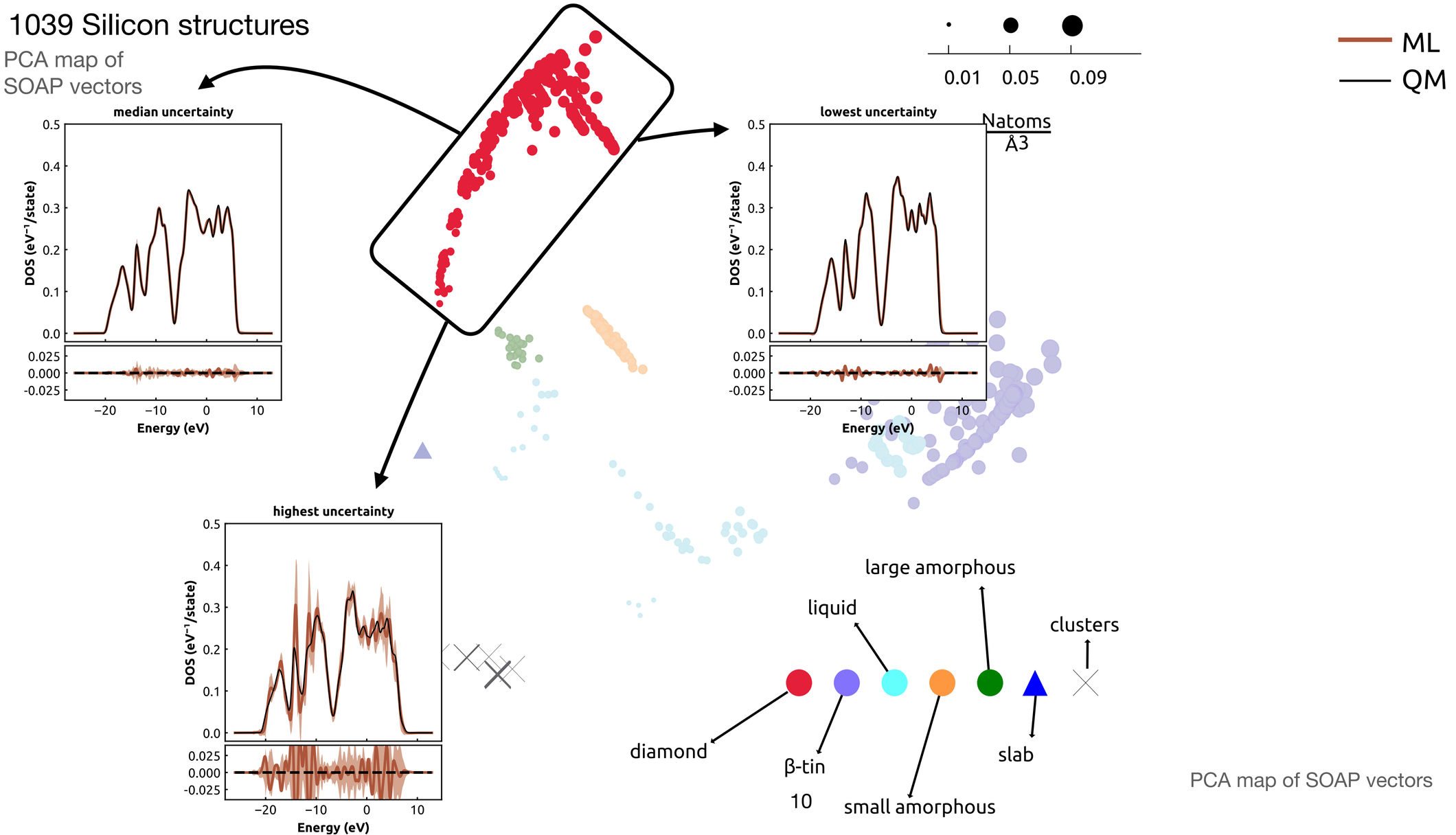
1039 Silicon structures

PCA map of
SOAP vectors



PCA map of SOAP vectors

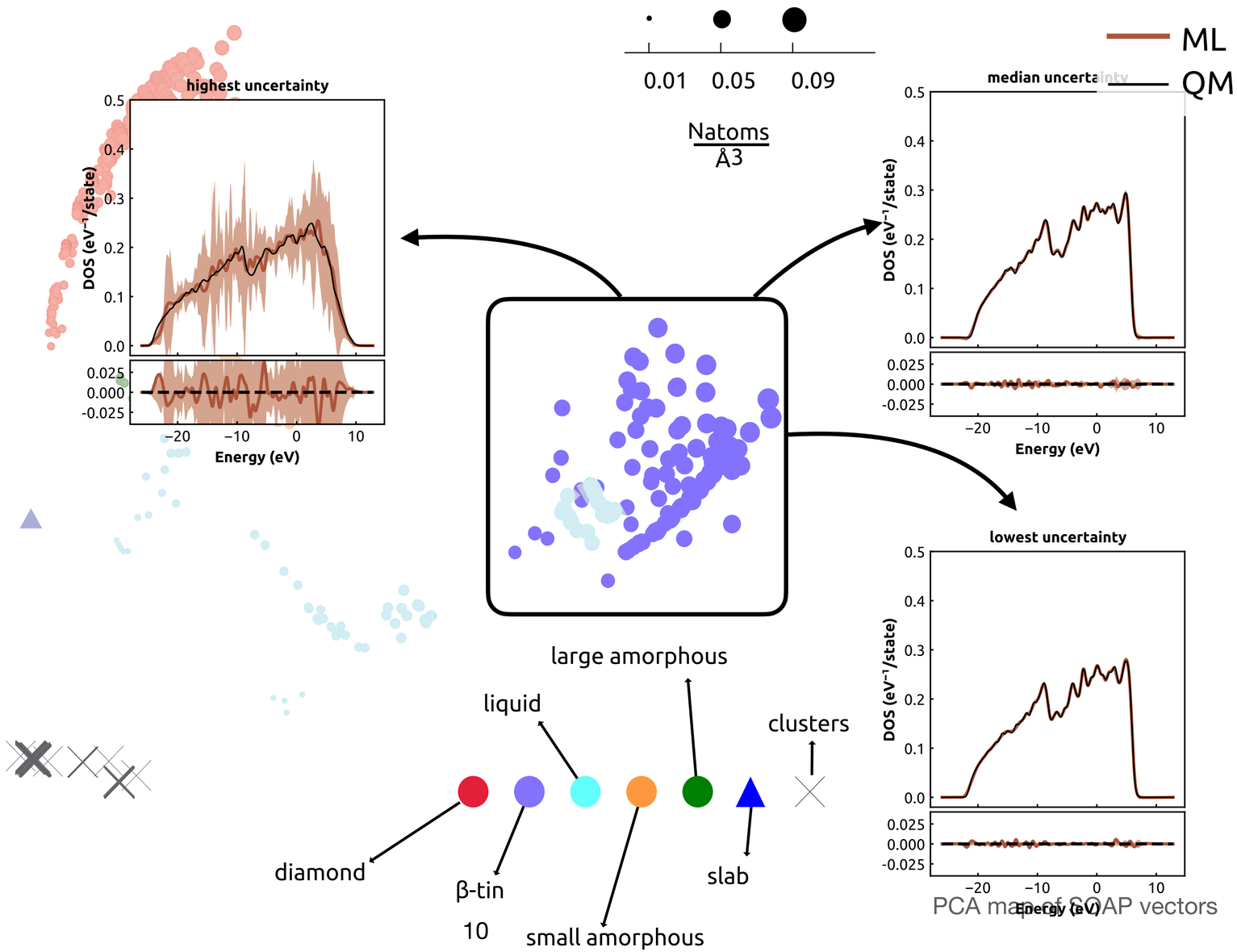
Benchmark on diverse silicon data set



Benchmark on diverse silicon data set

1039 Silicon structures

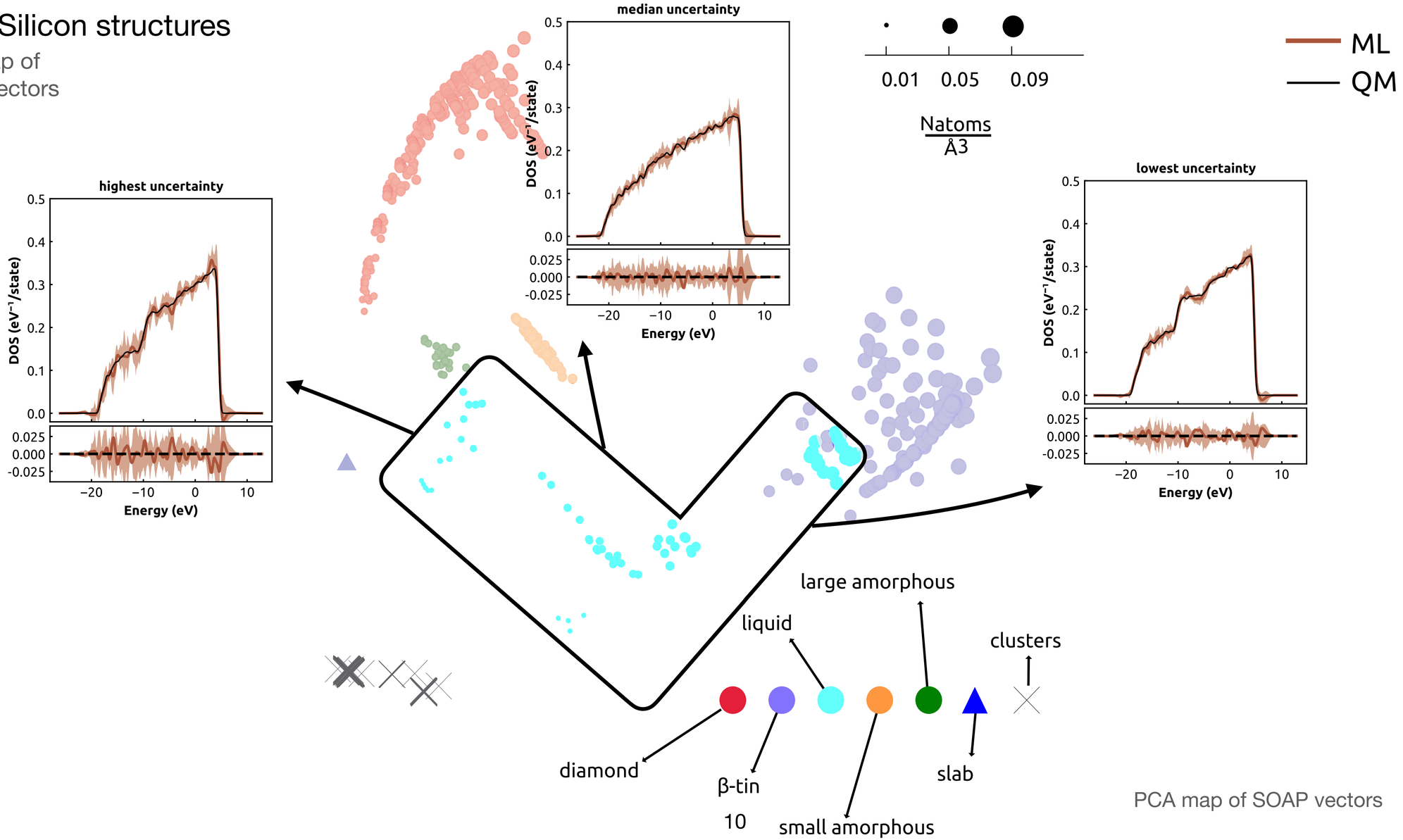
PCA map of
SOAP vectors



Benchmark on diverse silicon data set

1039 Silicon structures

PCA map of
SOAP vectors

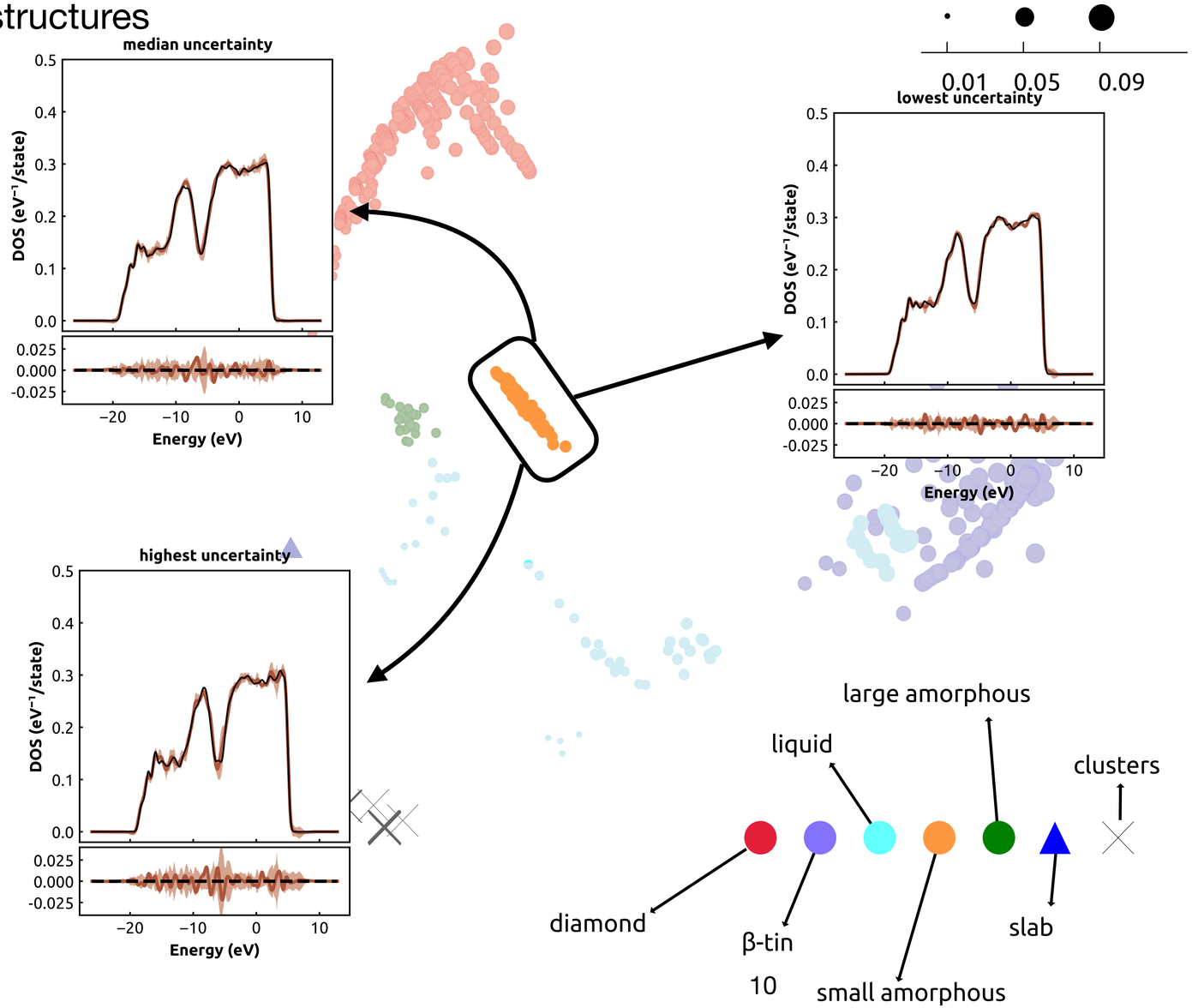


PCA map of SOAP vectors

Benchmark on diverse silicon data set

1039 Silicon structures

PCA map of SOAP vectors

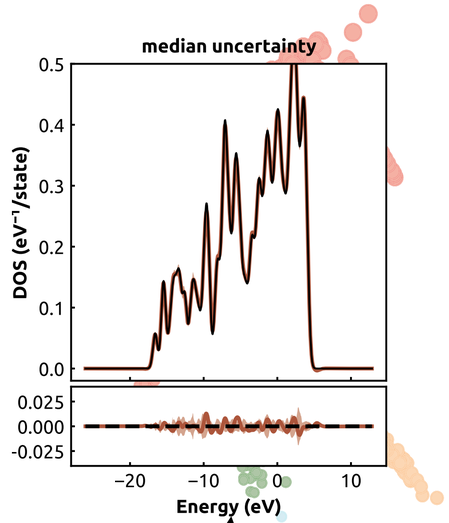


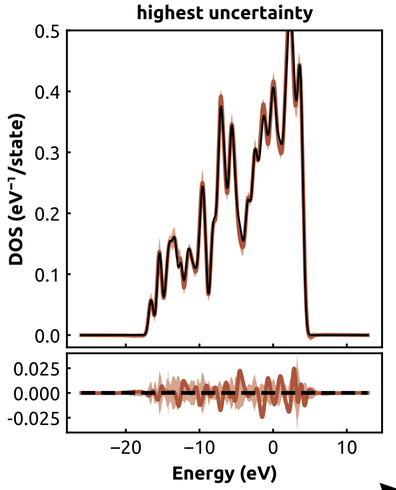
PCA map of SOAP vectors

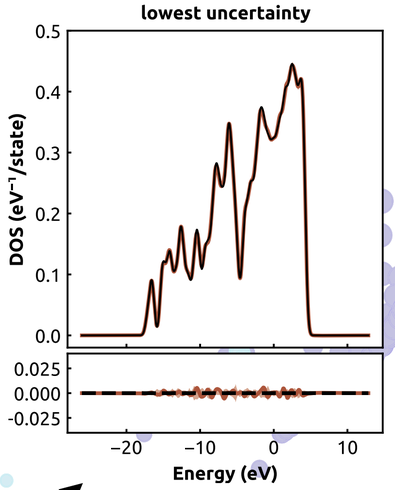
Benchmark on diverse silicon data set

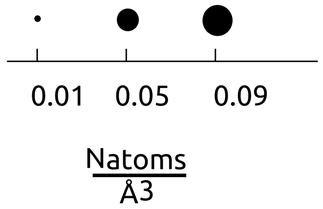
1039 Silicon structures

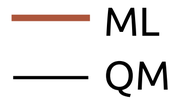
PCA map of
SOAP vectors

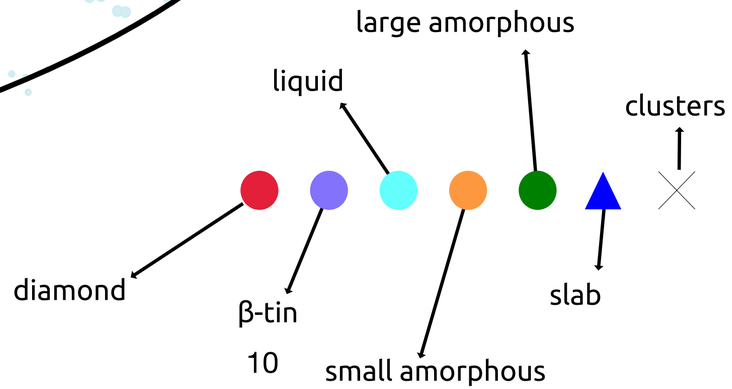










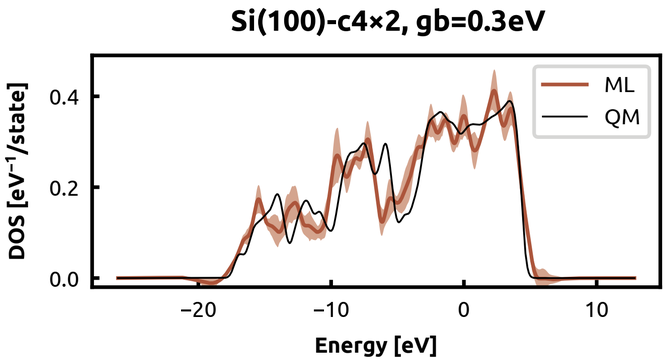


PCA map of SOAP vectors

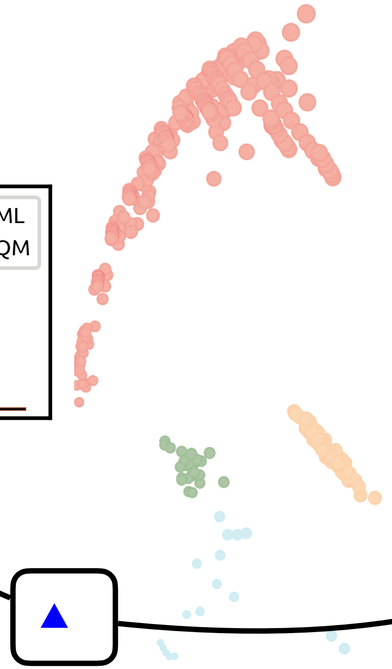
Benchmark on diverse silicon data set

1039 Silicon structures

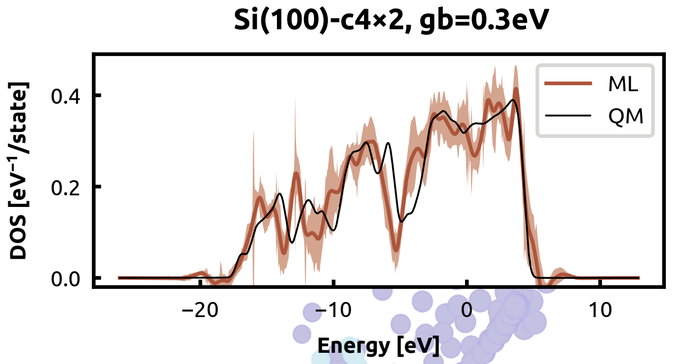
PCA map of
SOAP vectors



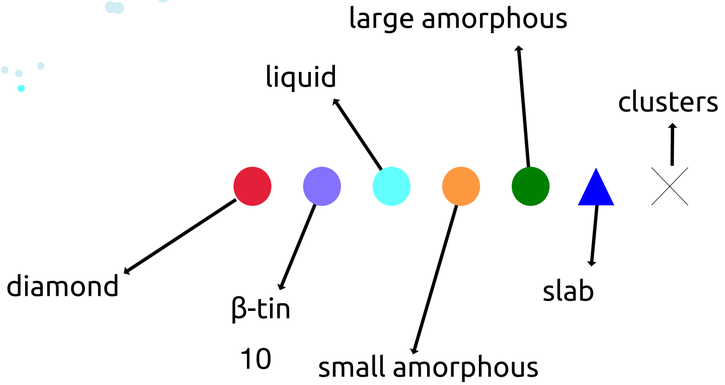
from PW



ML
QM



from CDF



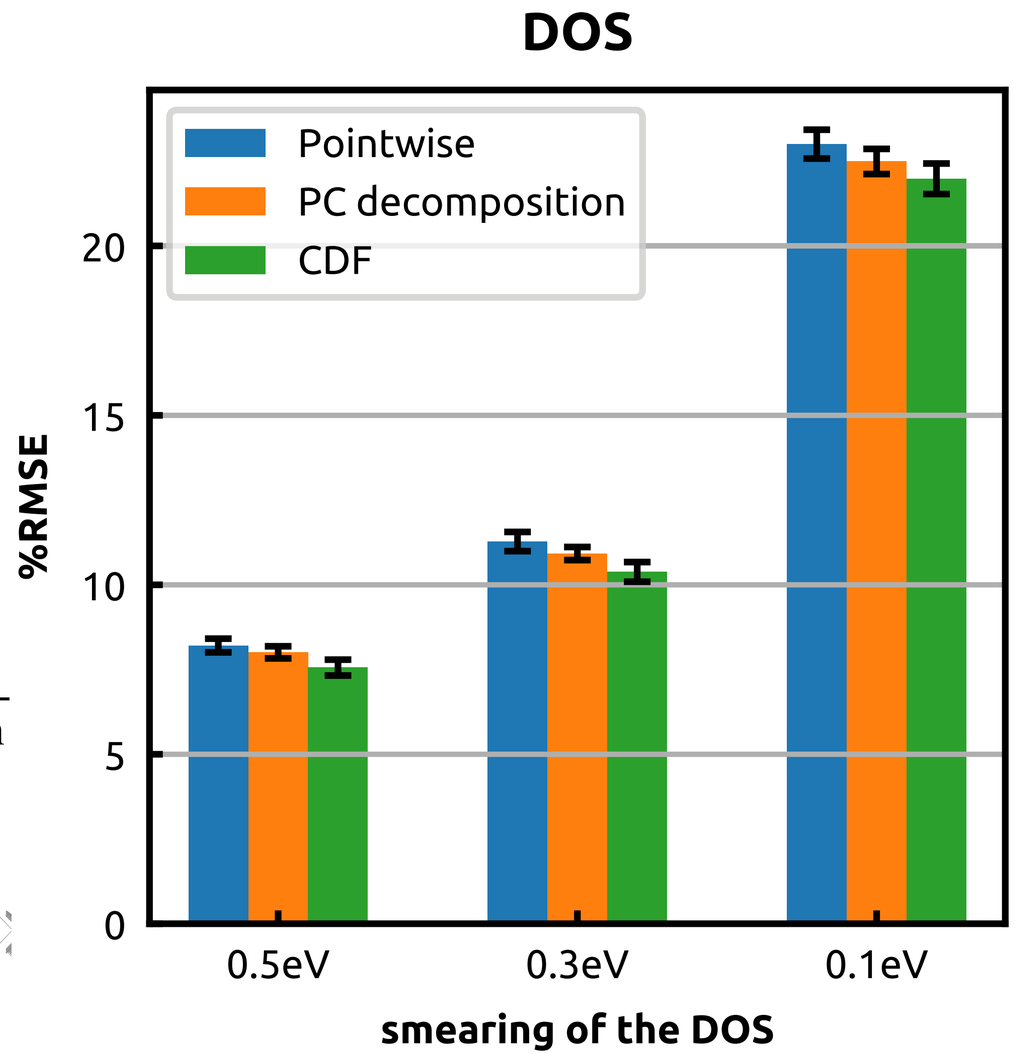
PCA map of SOAP vectors

Benchmark on diverse silicon data set

1039 Silicon structures

PCA map of
SOAP vectors

$$\%\,\mathrm{RMSE} = 100 \times \frac{\mathrm{RMSE}}{\text{standard deviation}}$$

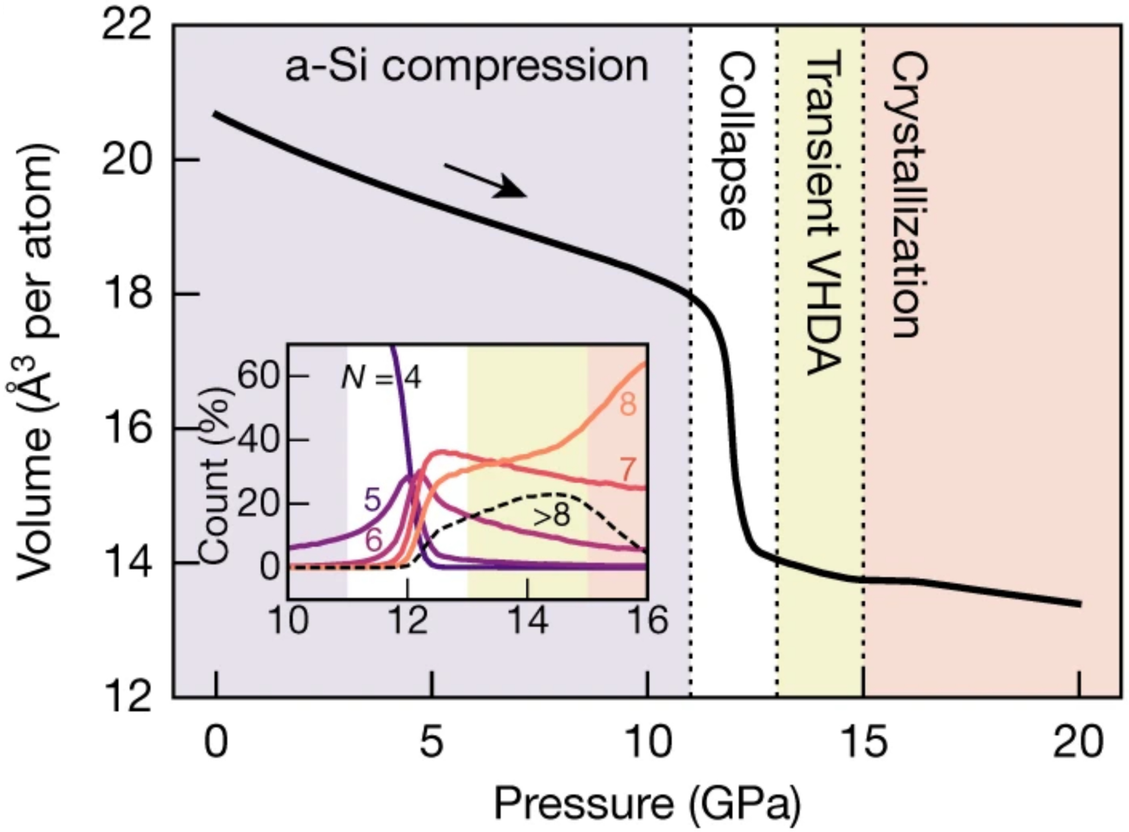


ML
QM

diamond
β-tin
slab
small amorphous

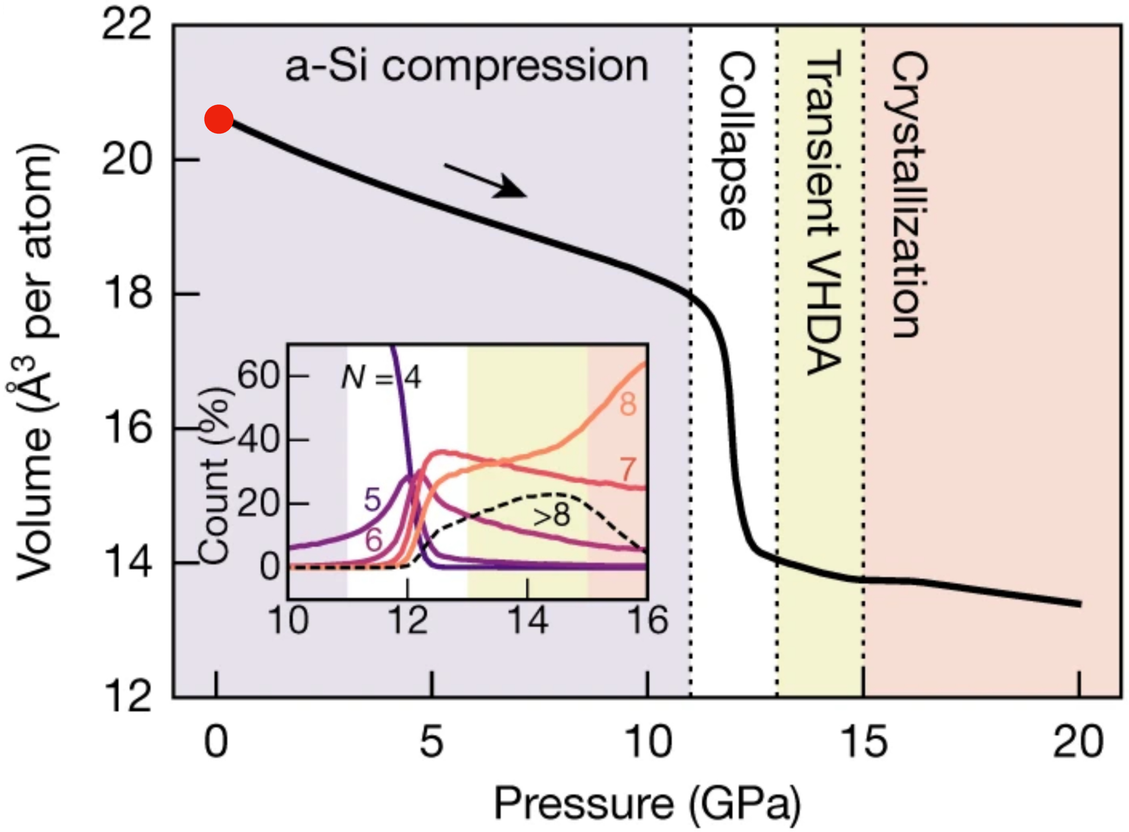
PCA map of SOAP vectors

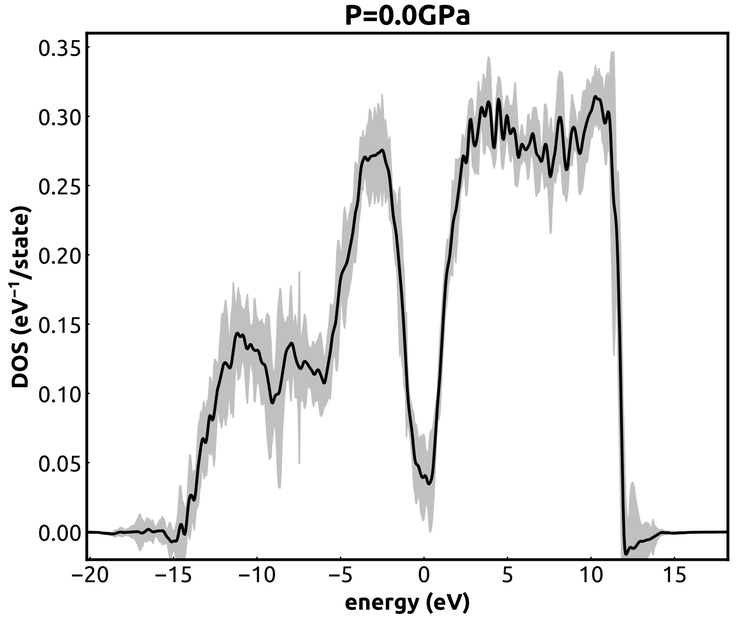
Combining structural and electronic fingerprints

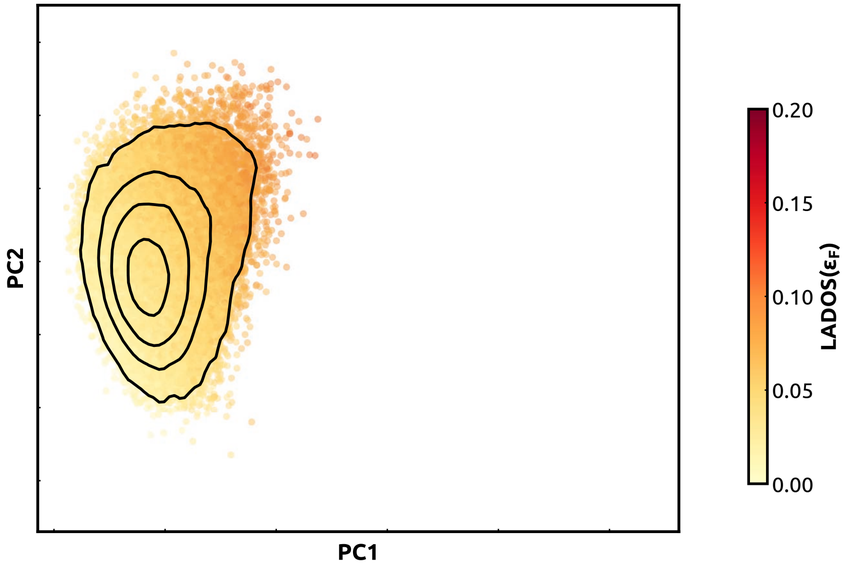


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

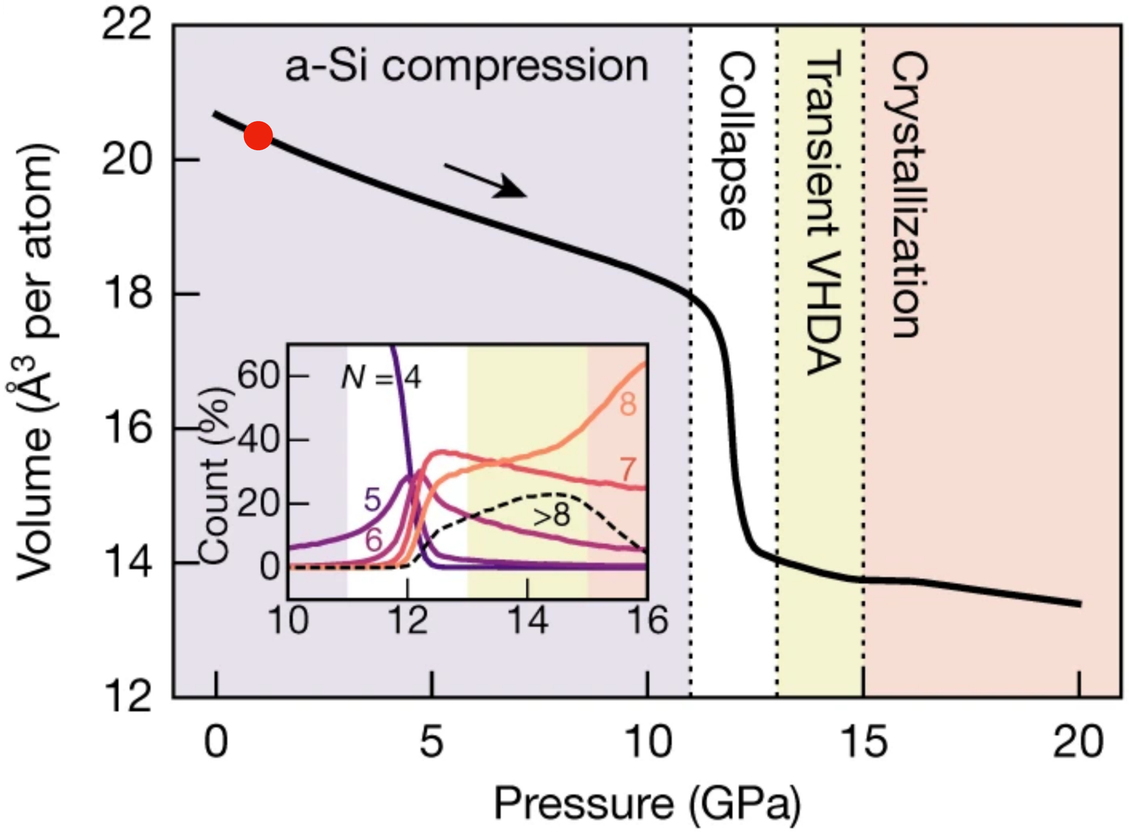


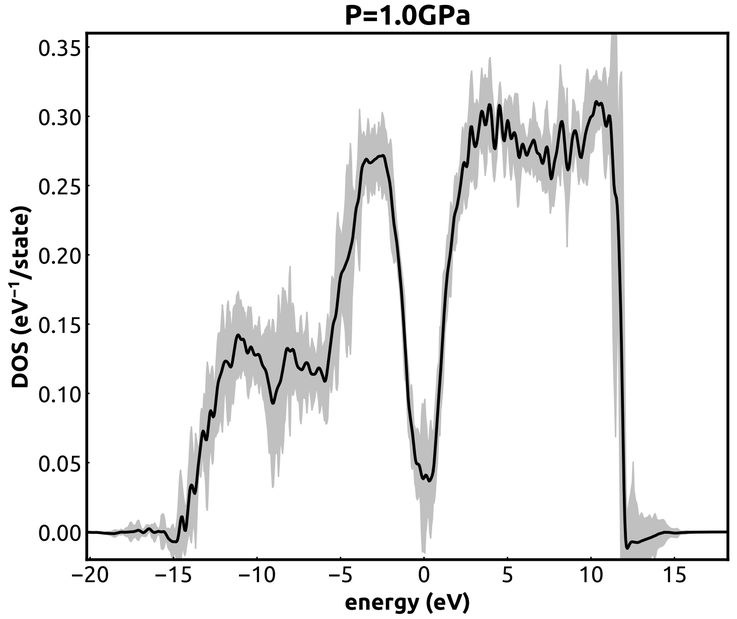


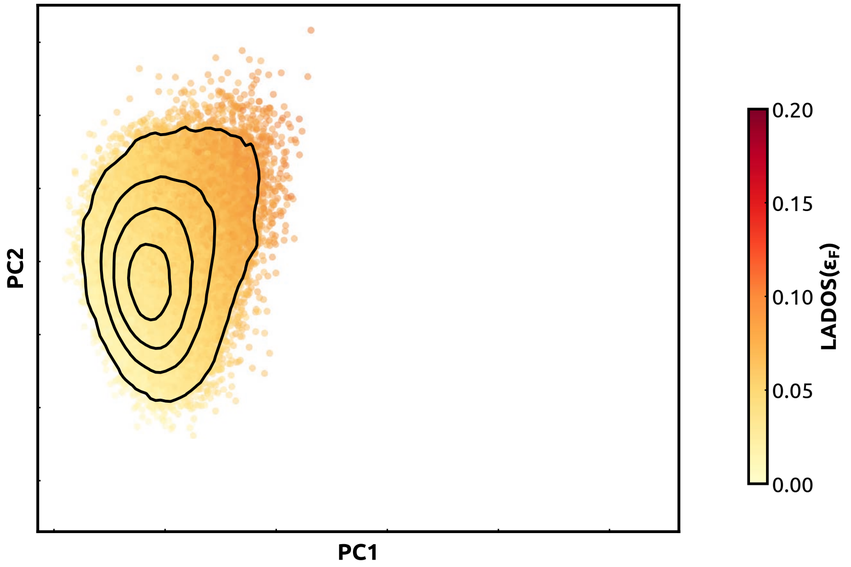


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

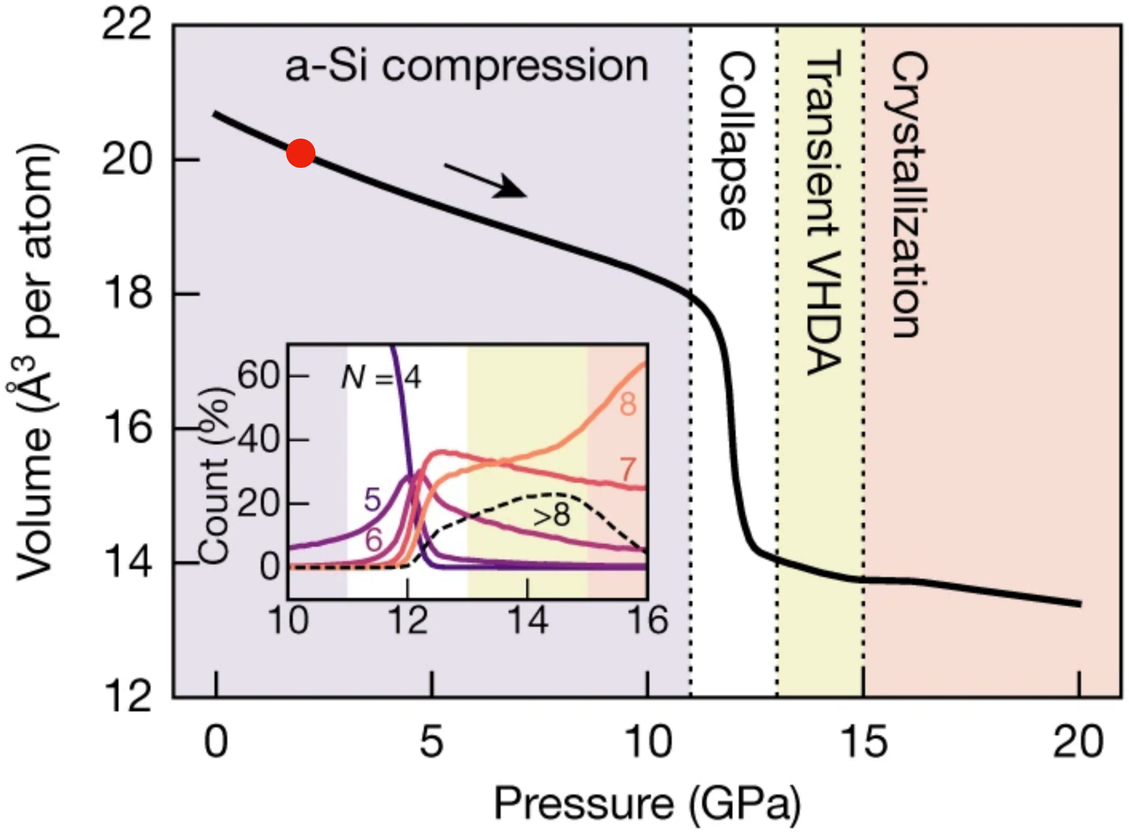


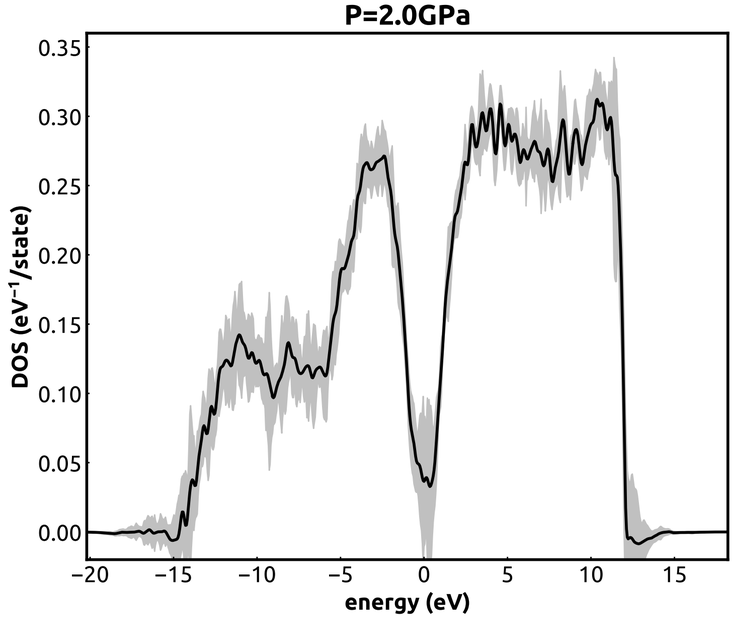


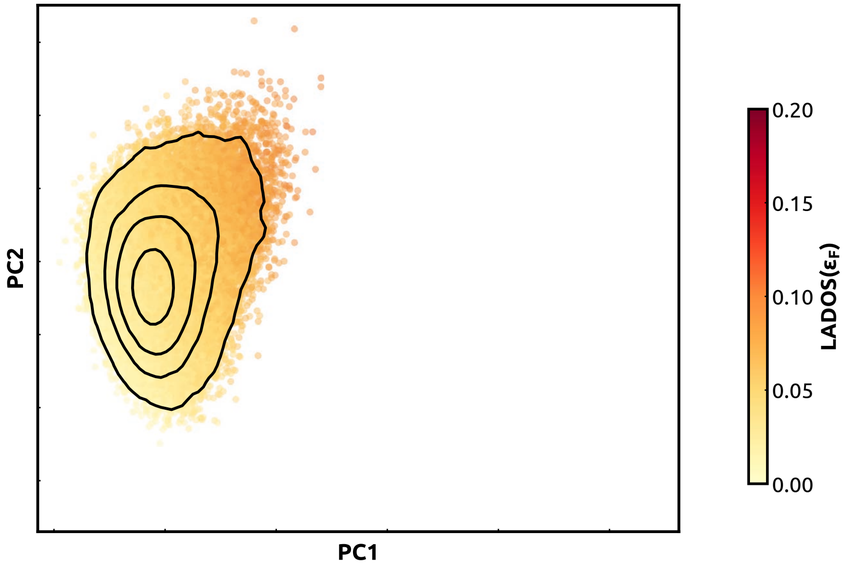


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

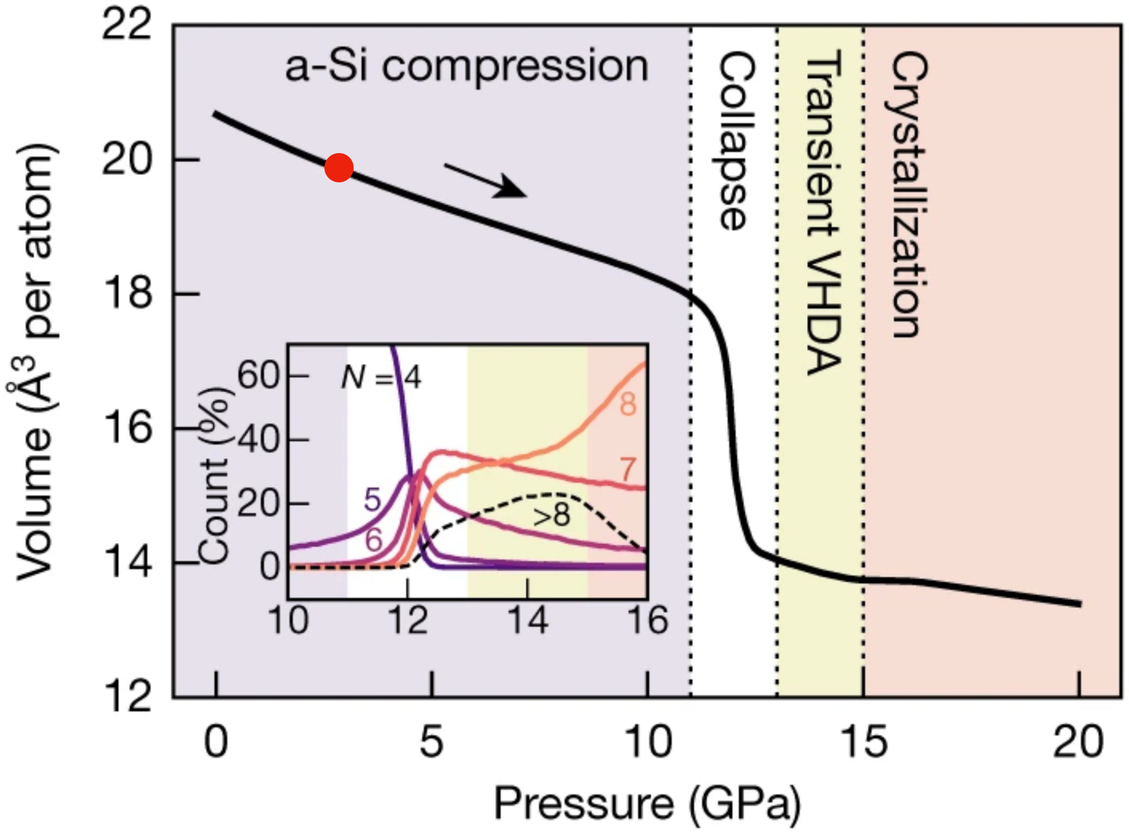


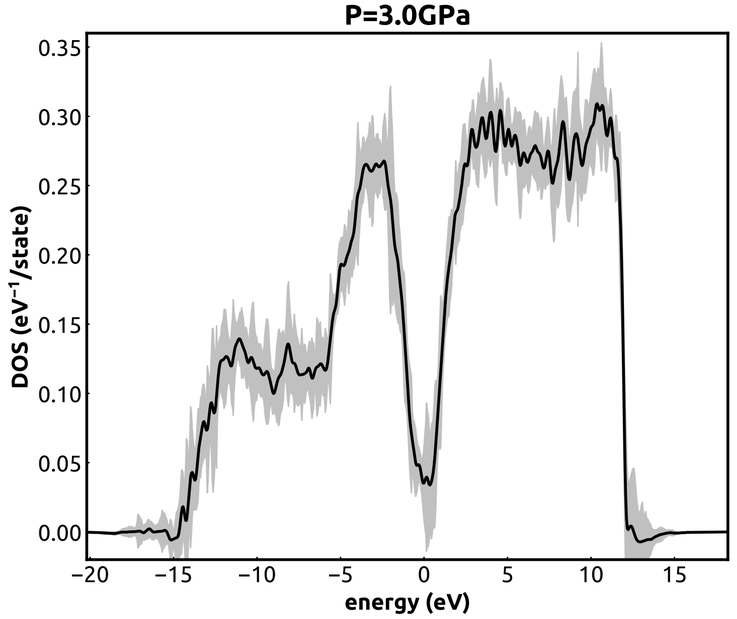


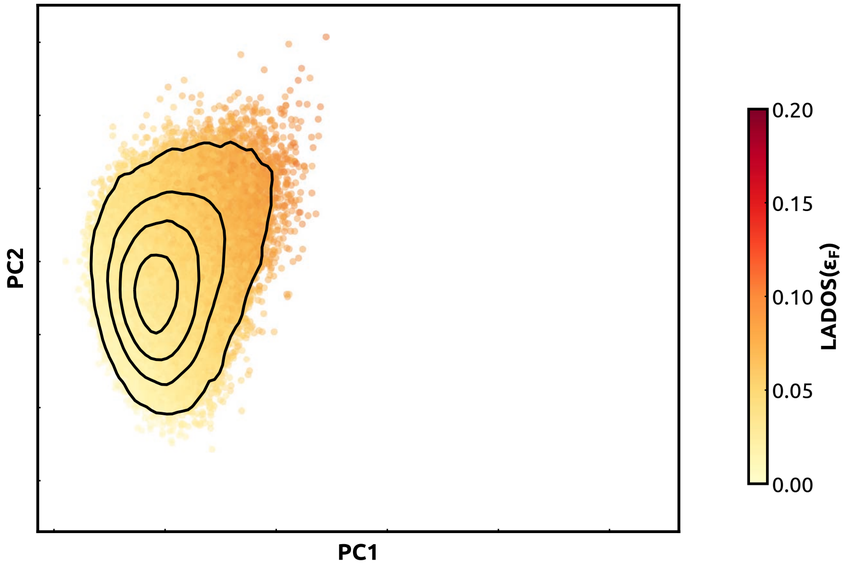


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

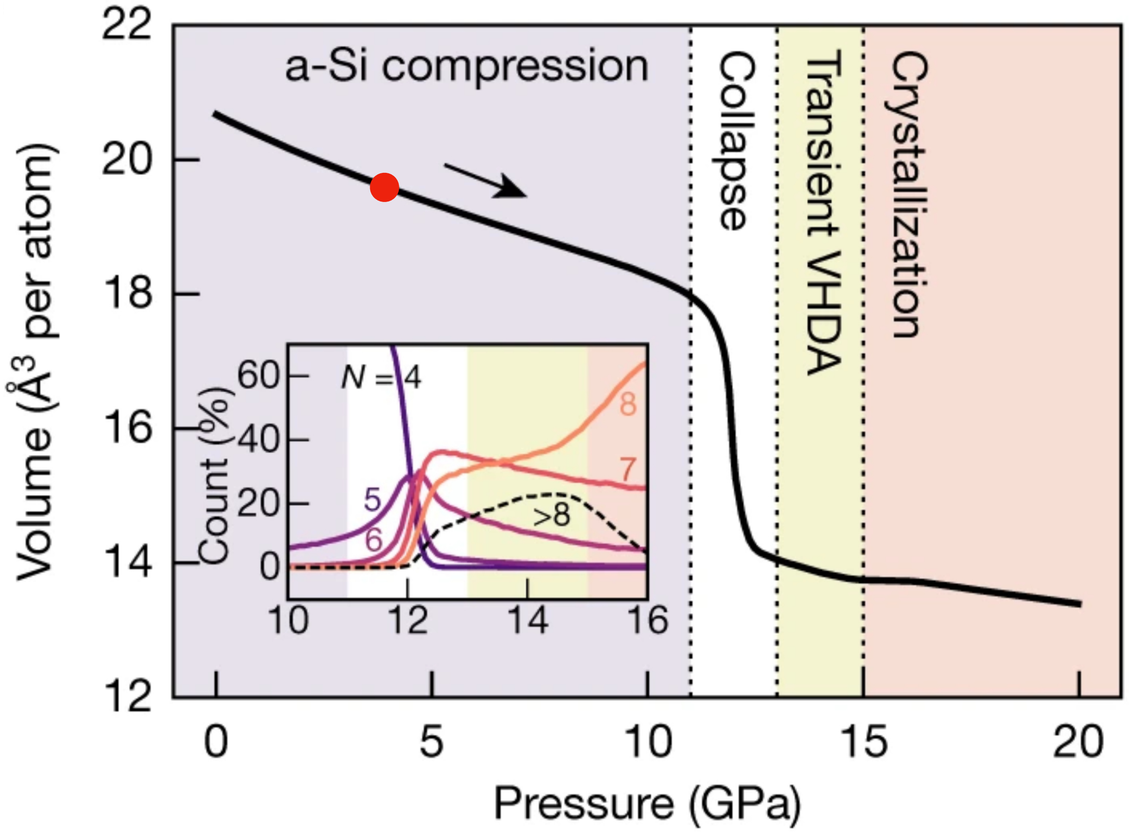


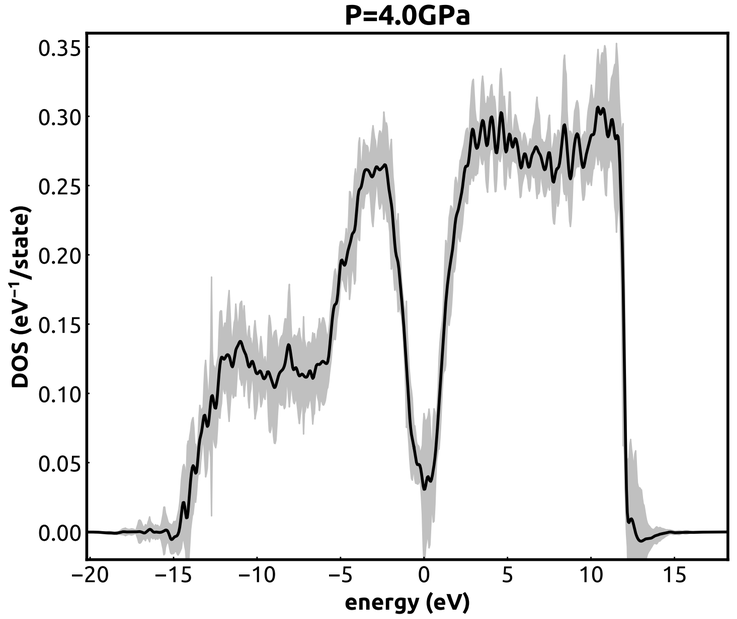


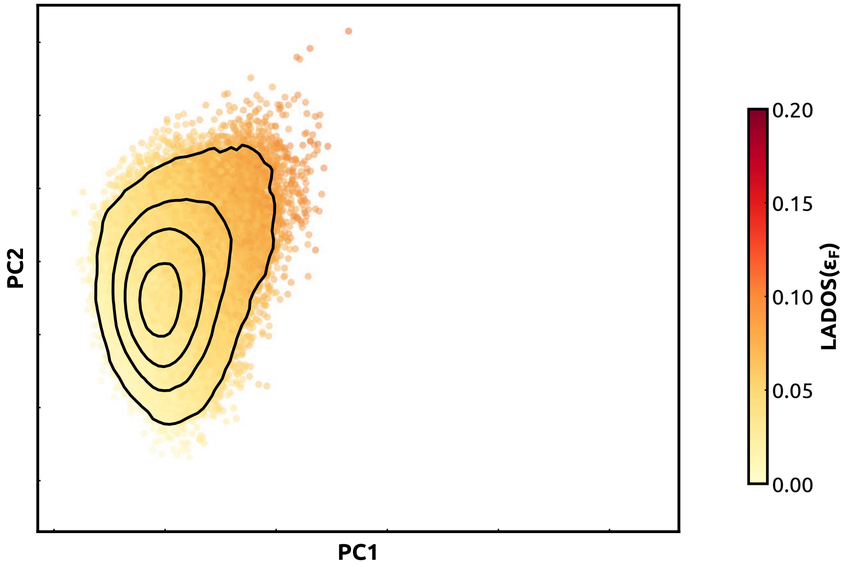


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

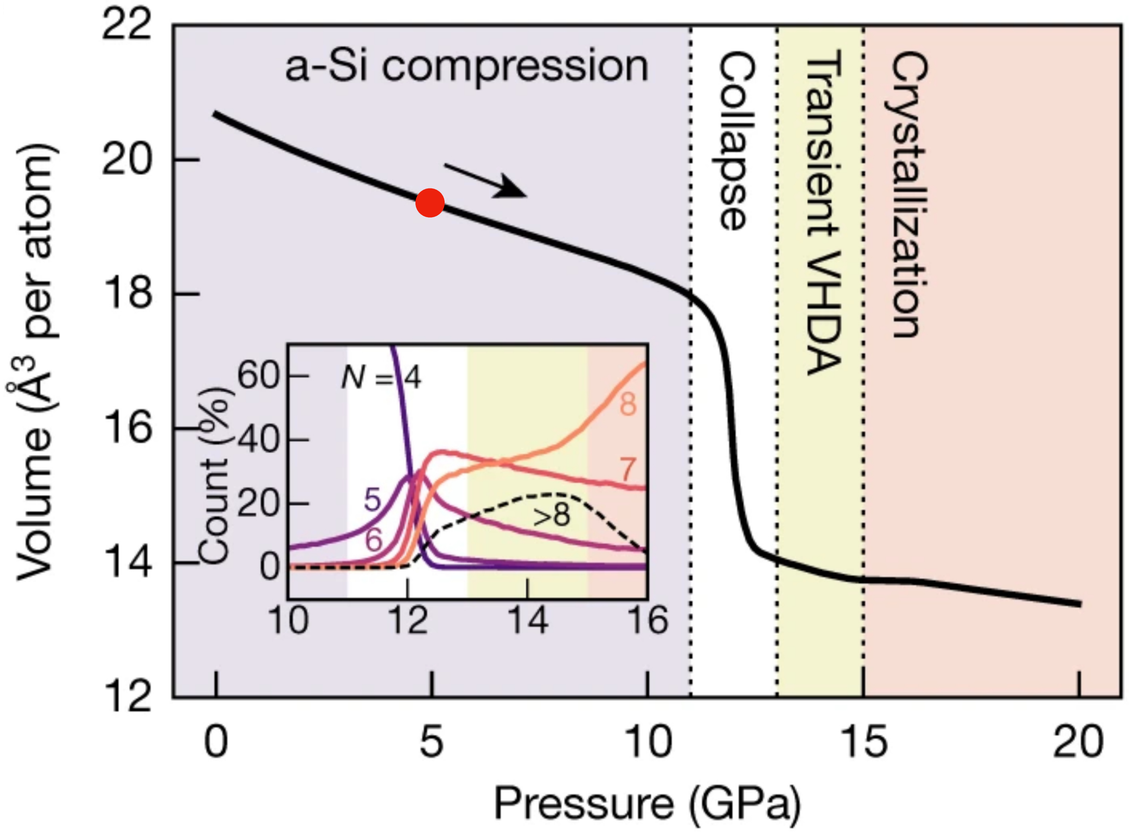


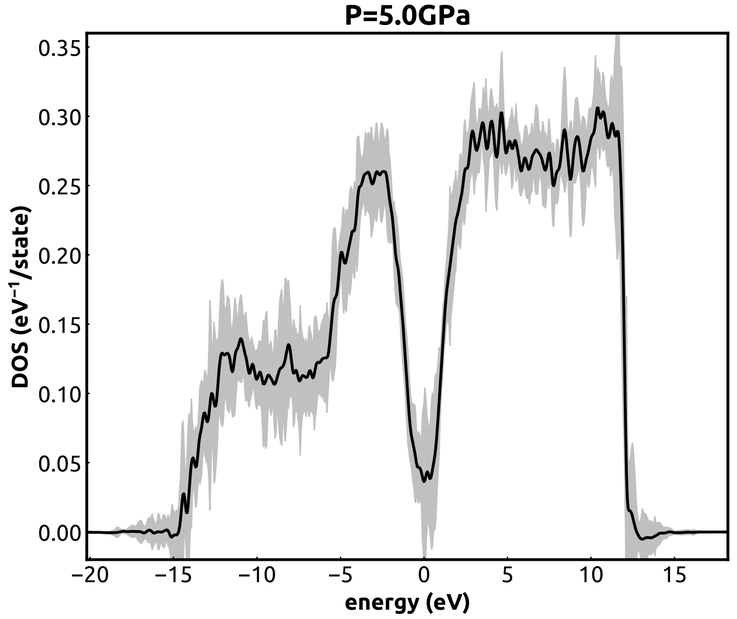


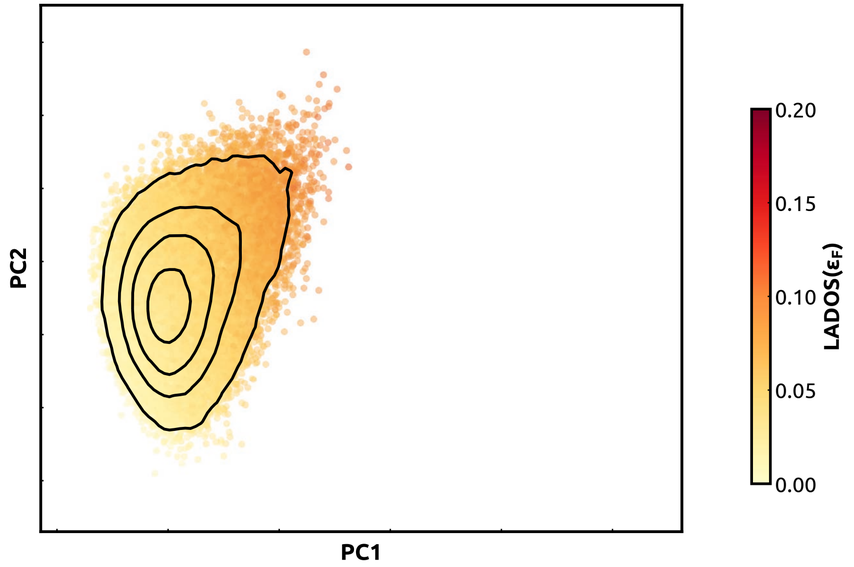


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

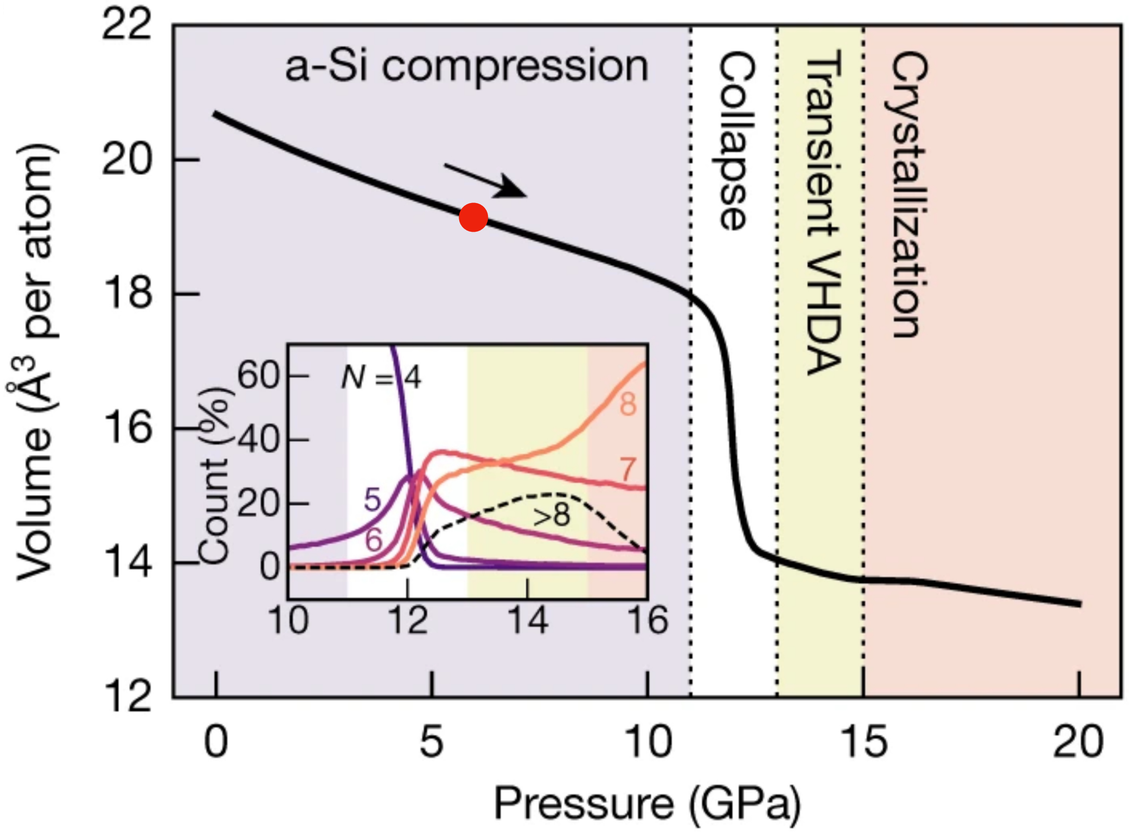


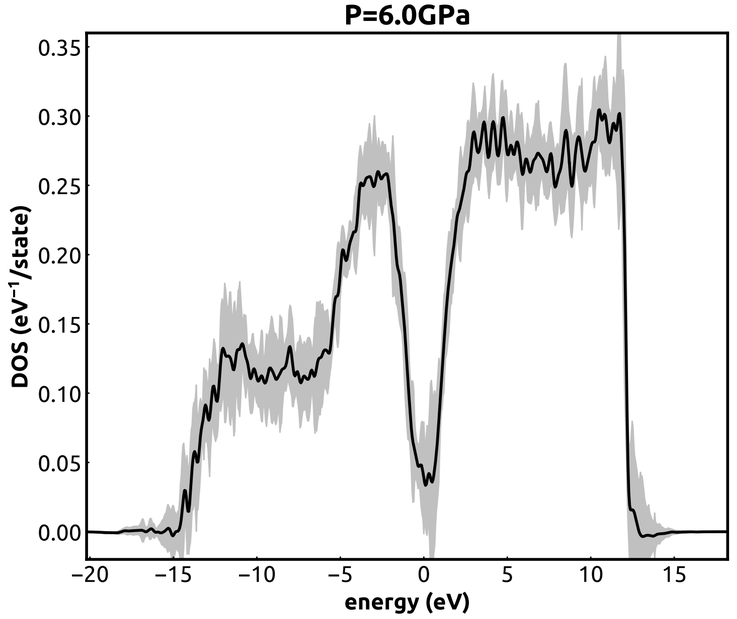


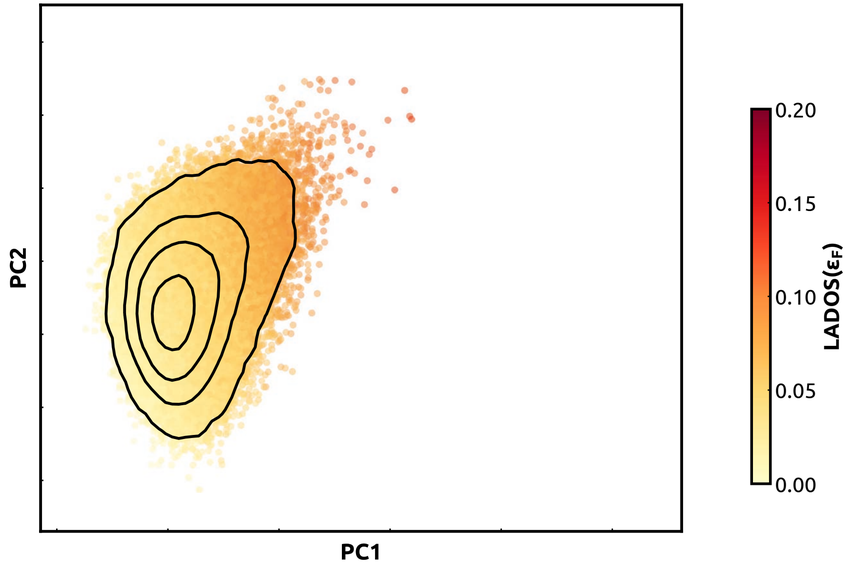


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

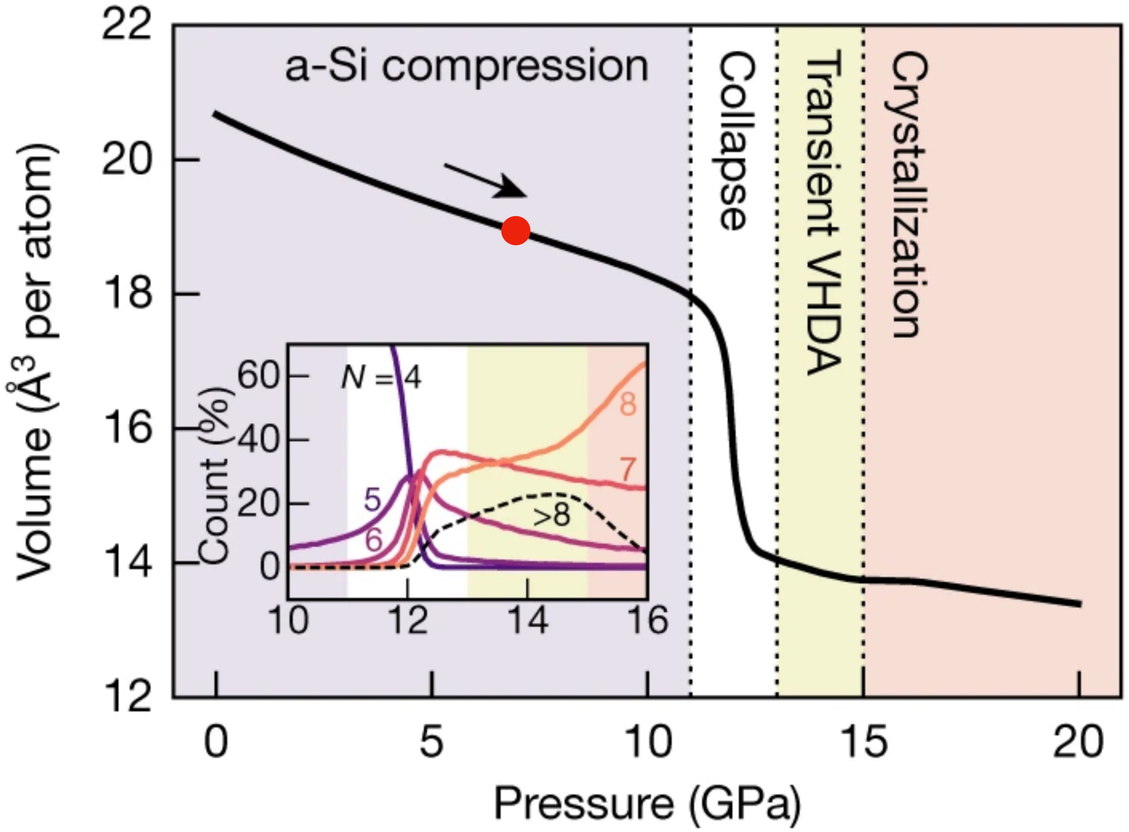


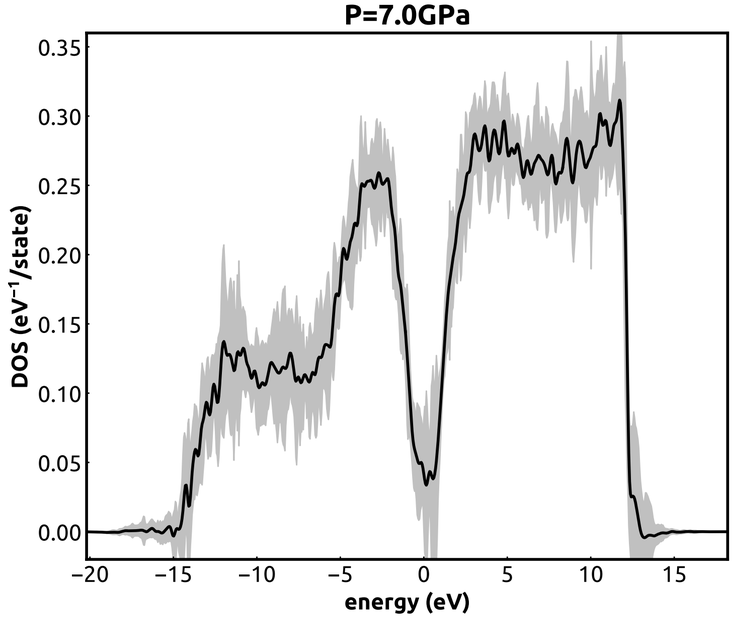


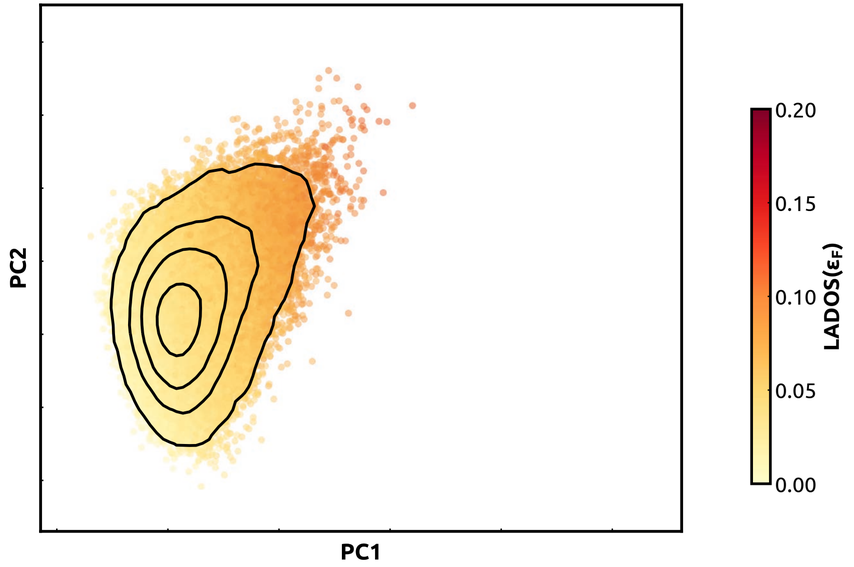


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

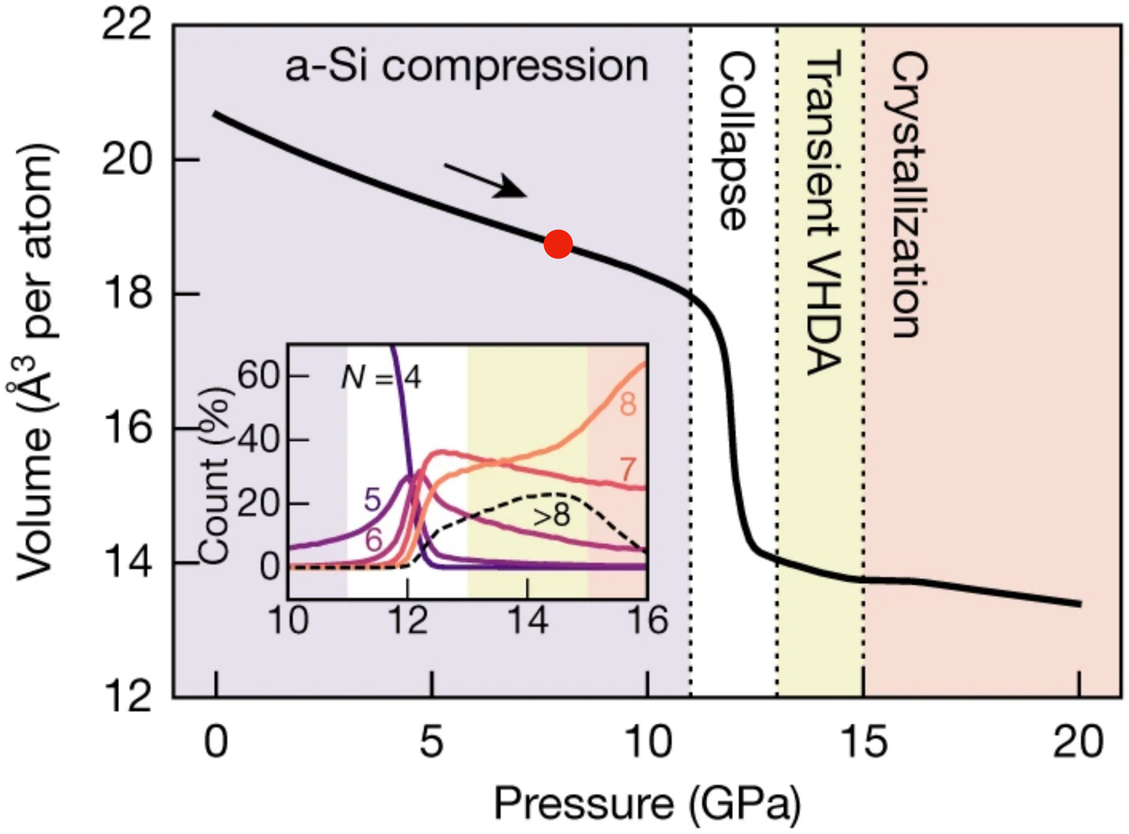


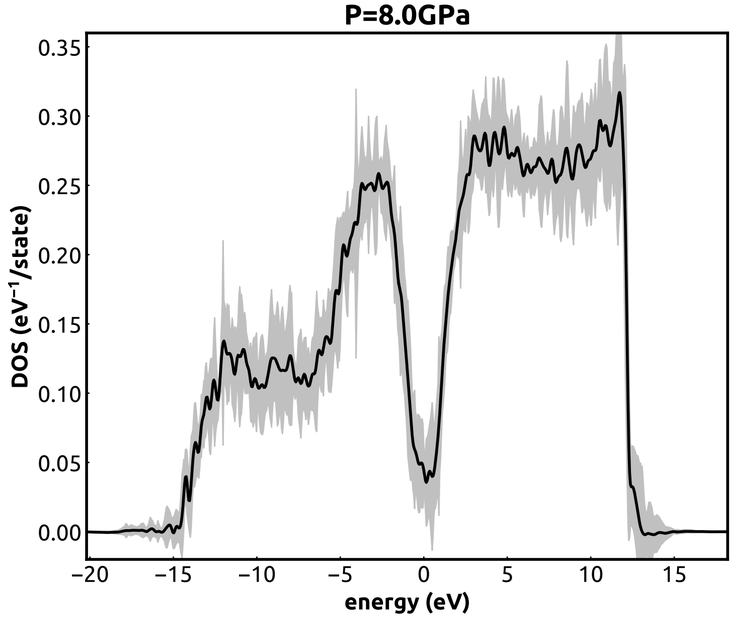


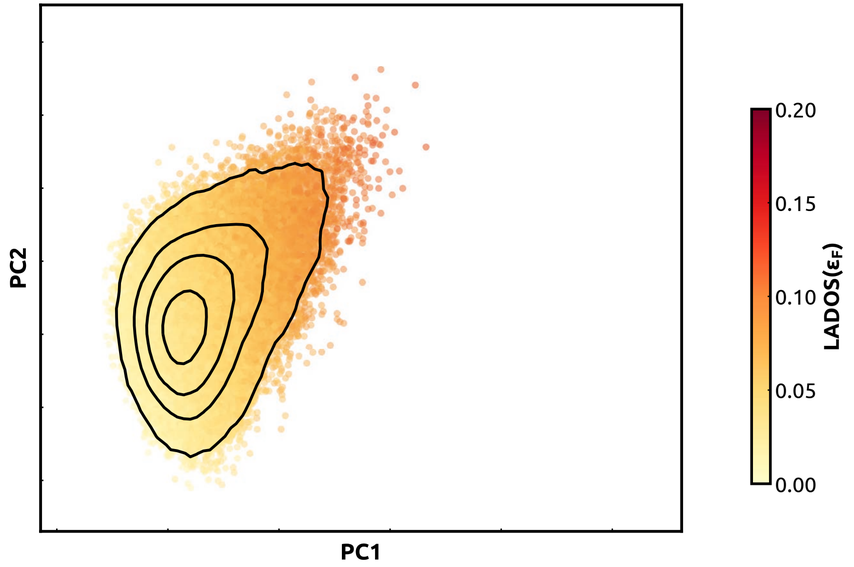


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

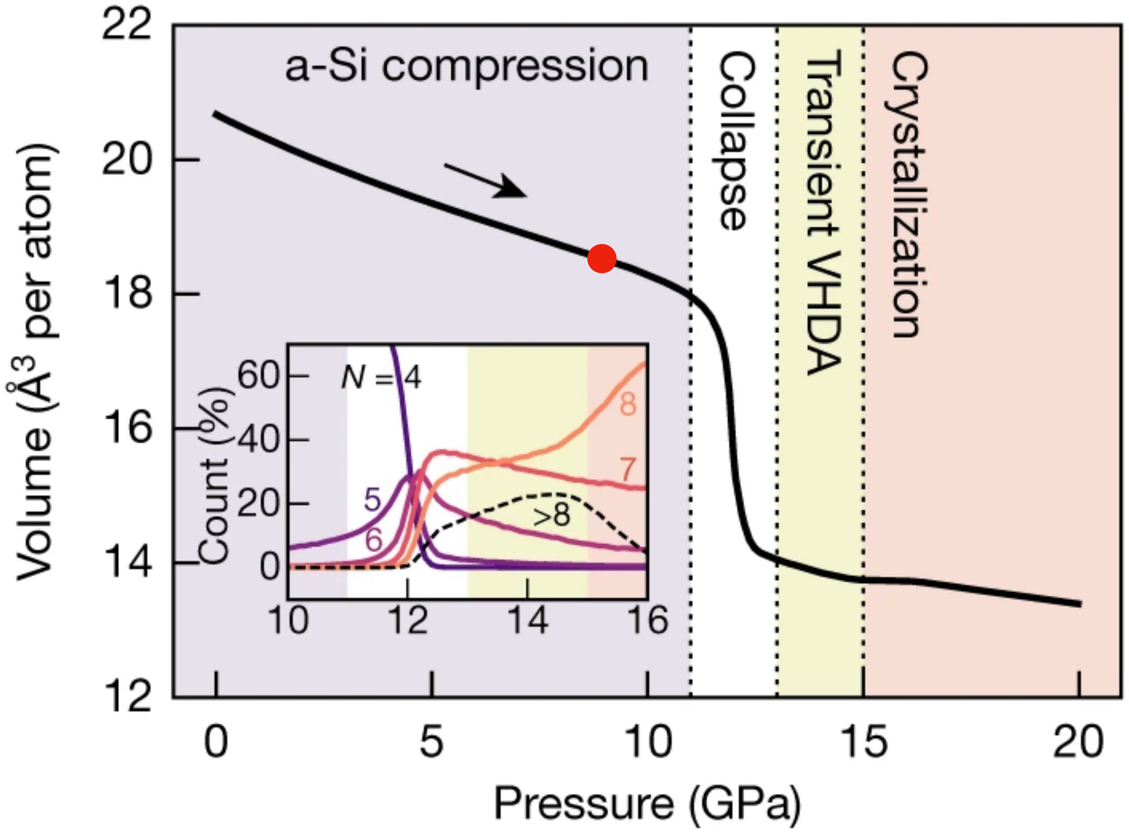


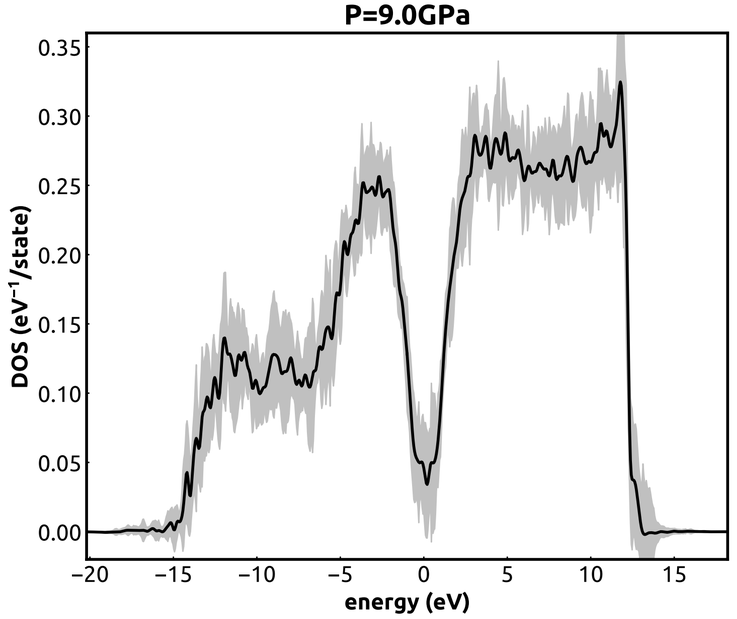


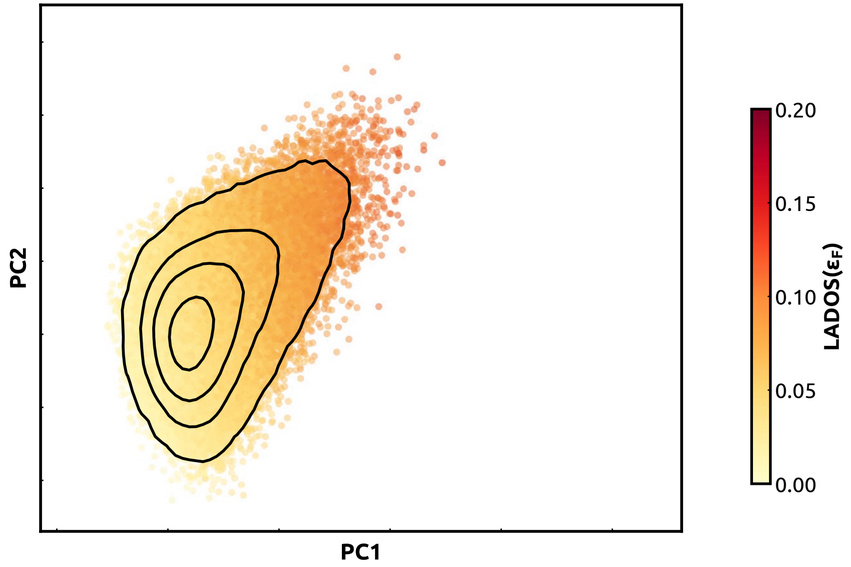


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

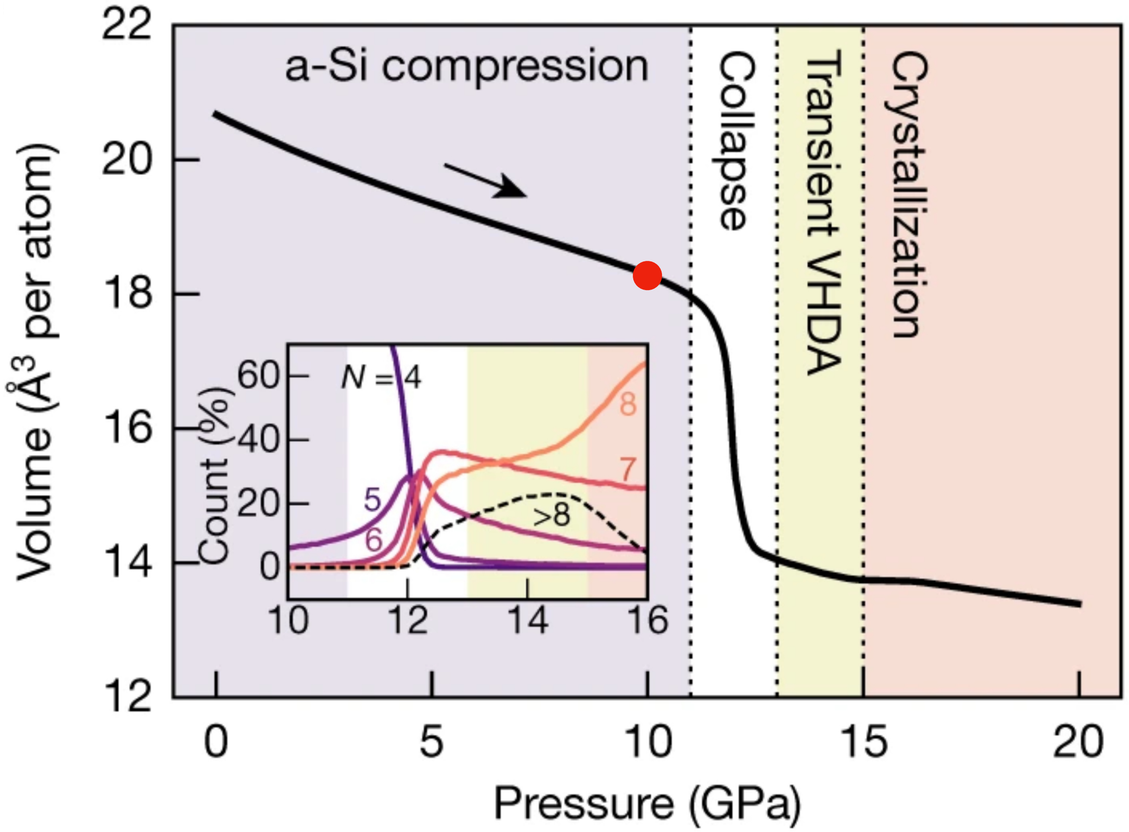


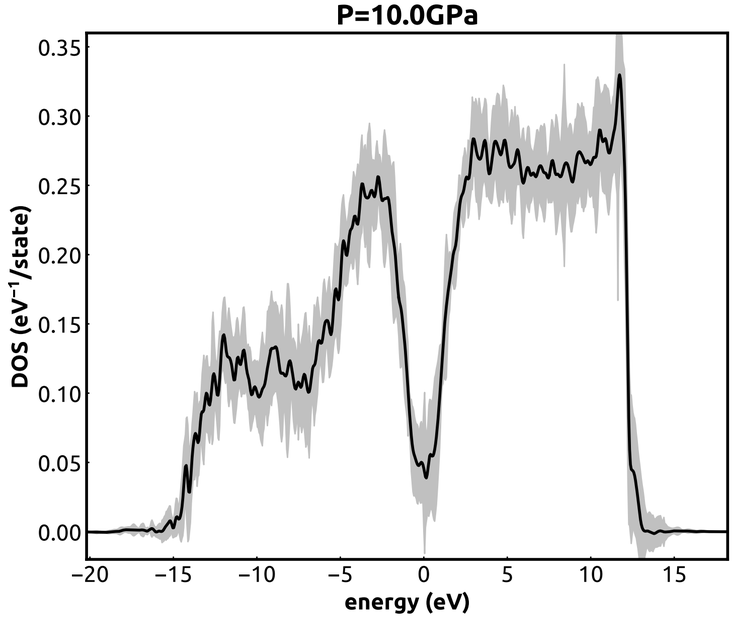


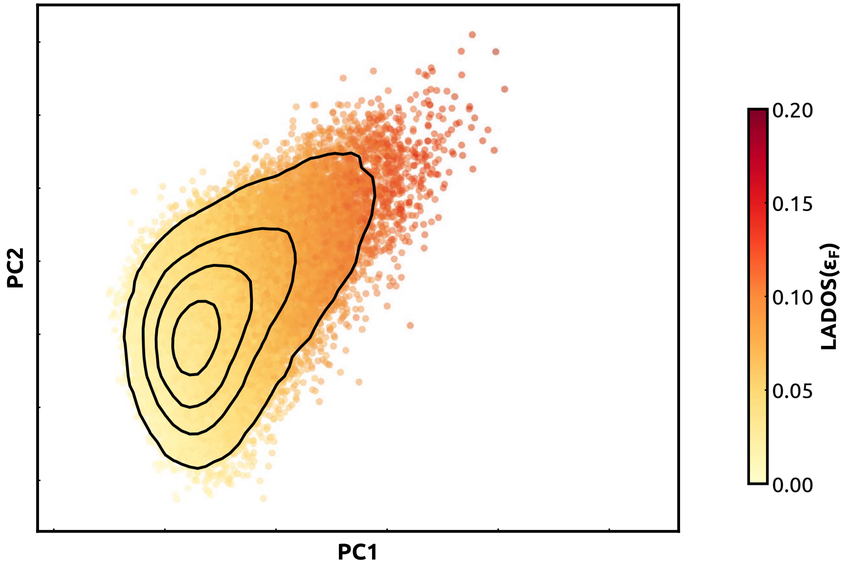


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

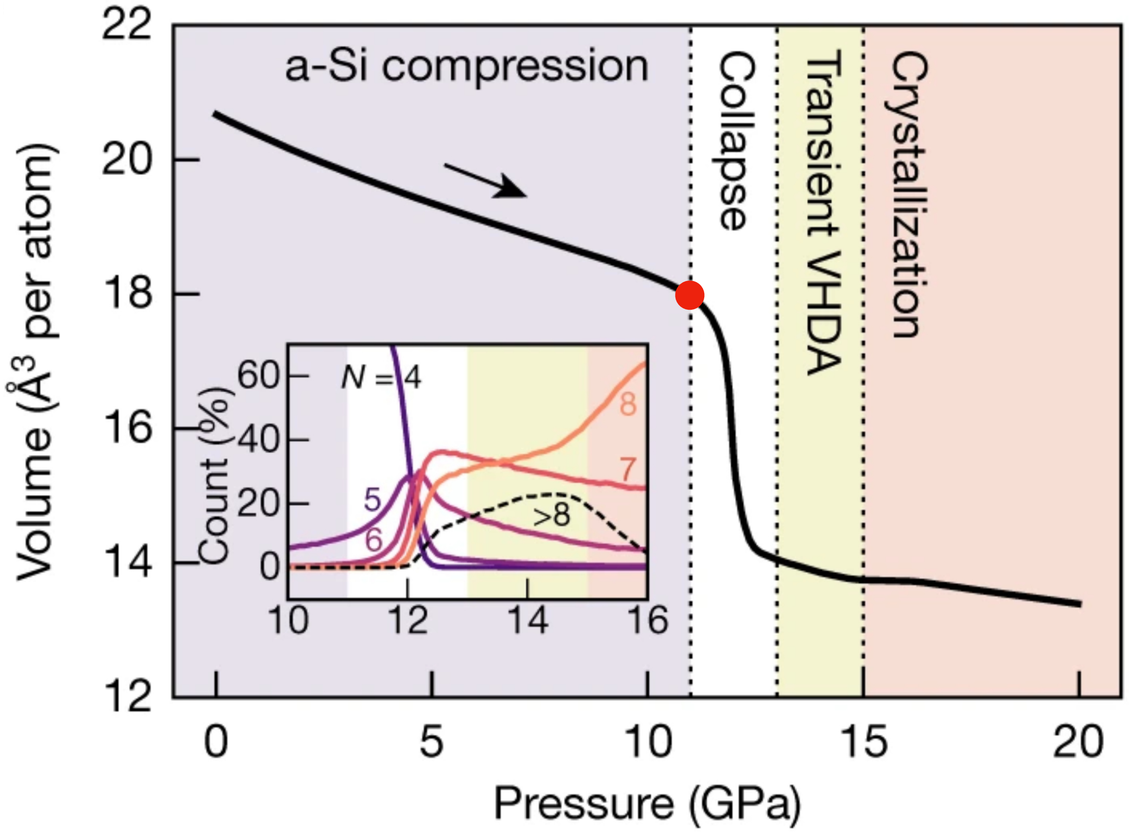


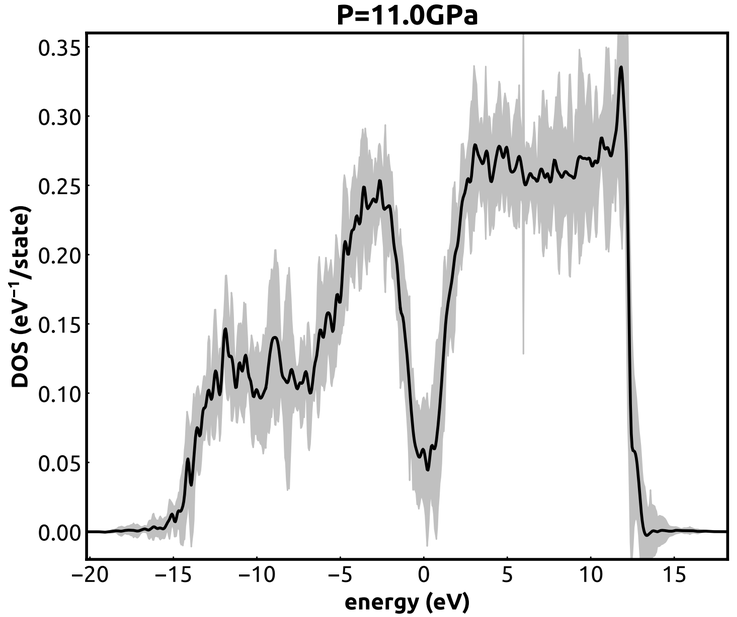


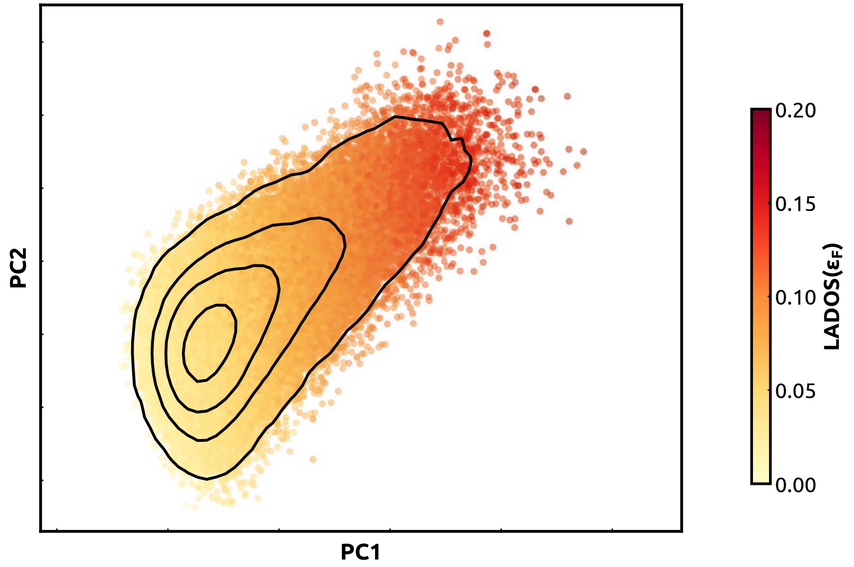


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

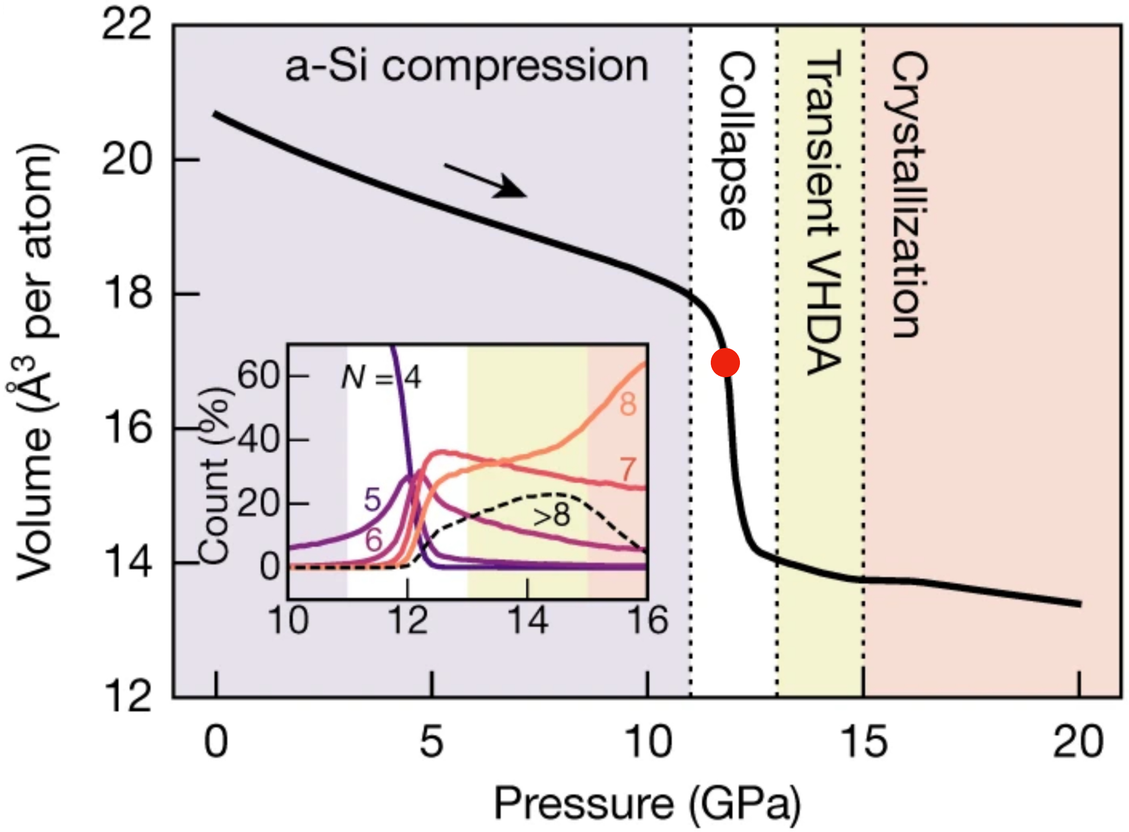


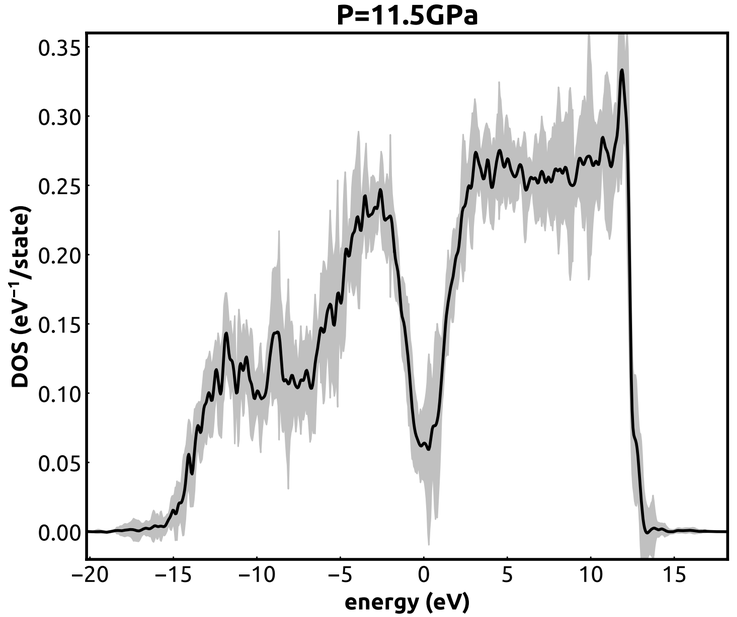


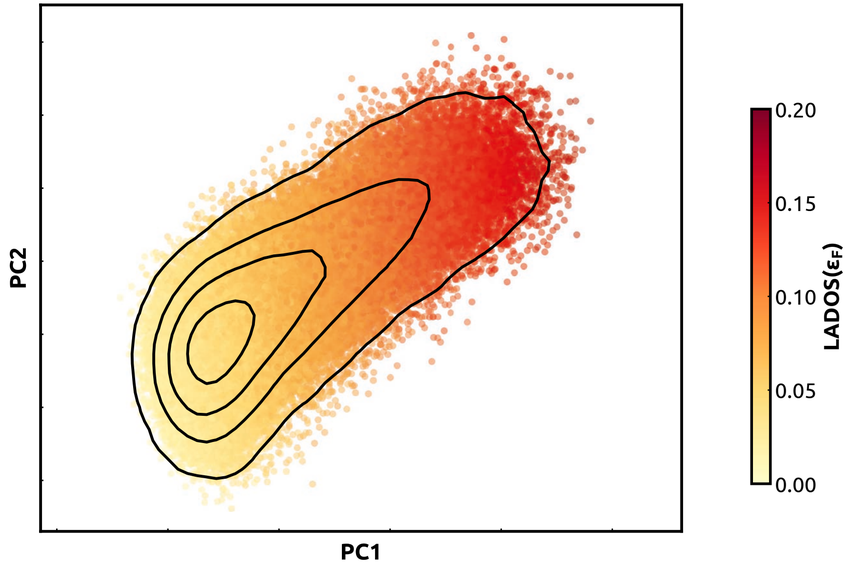


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

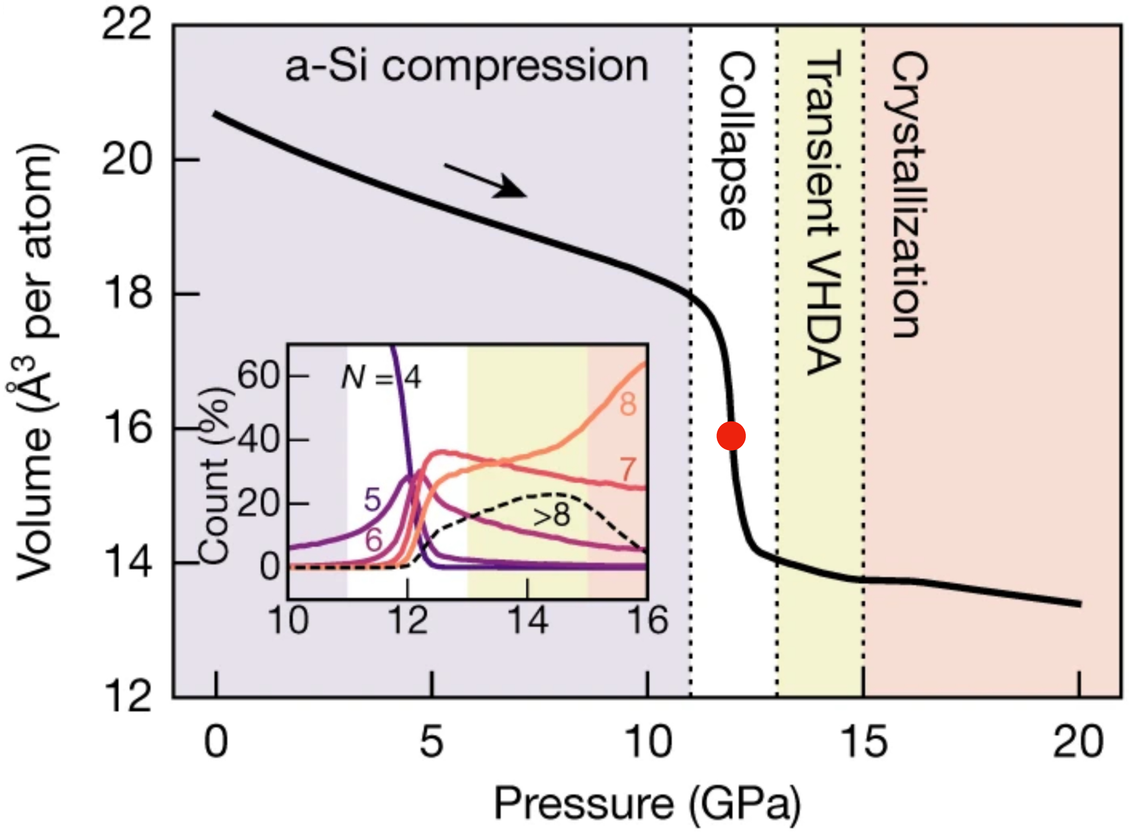


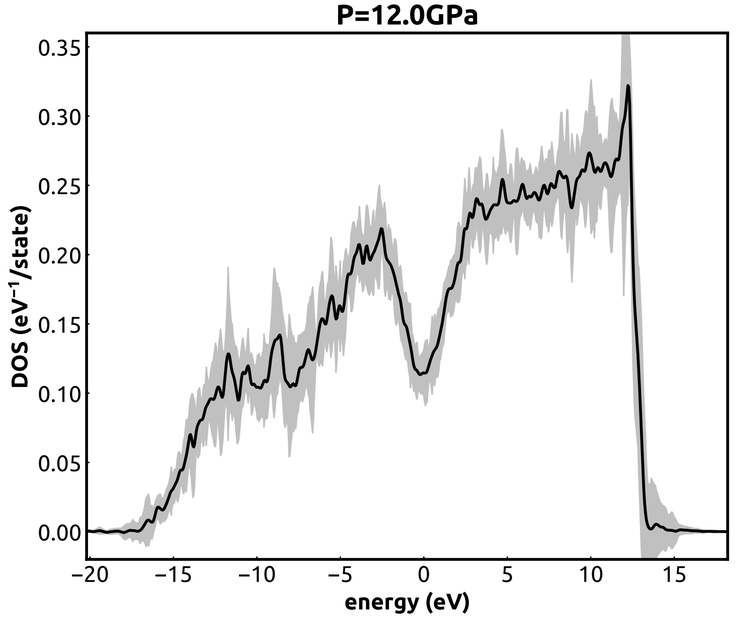


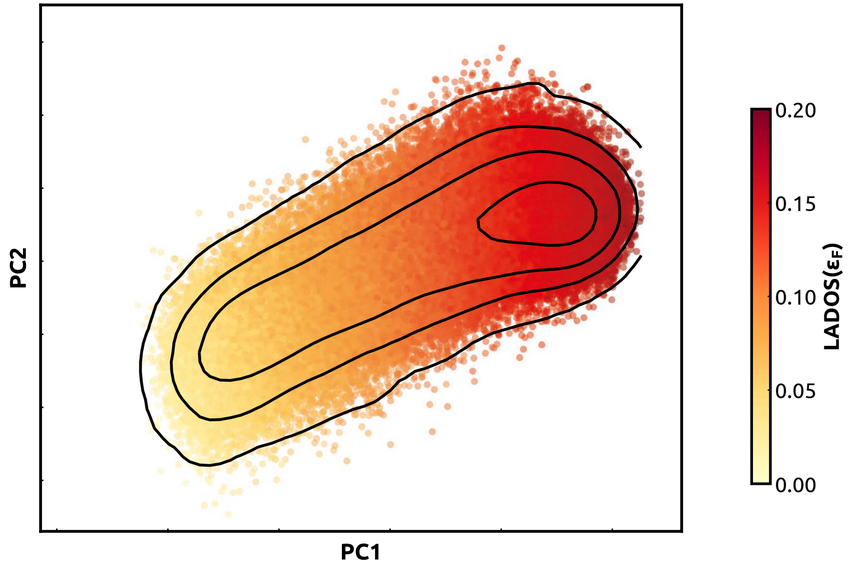


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

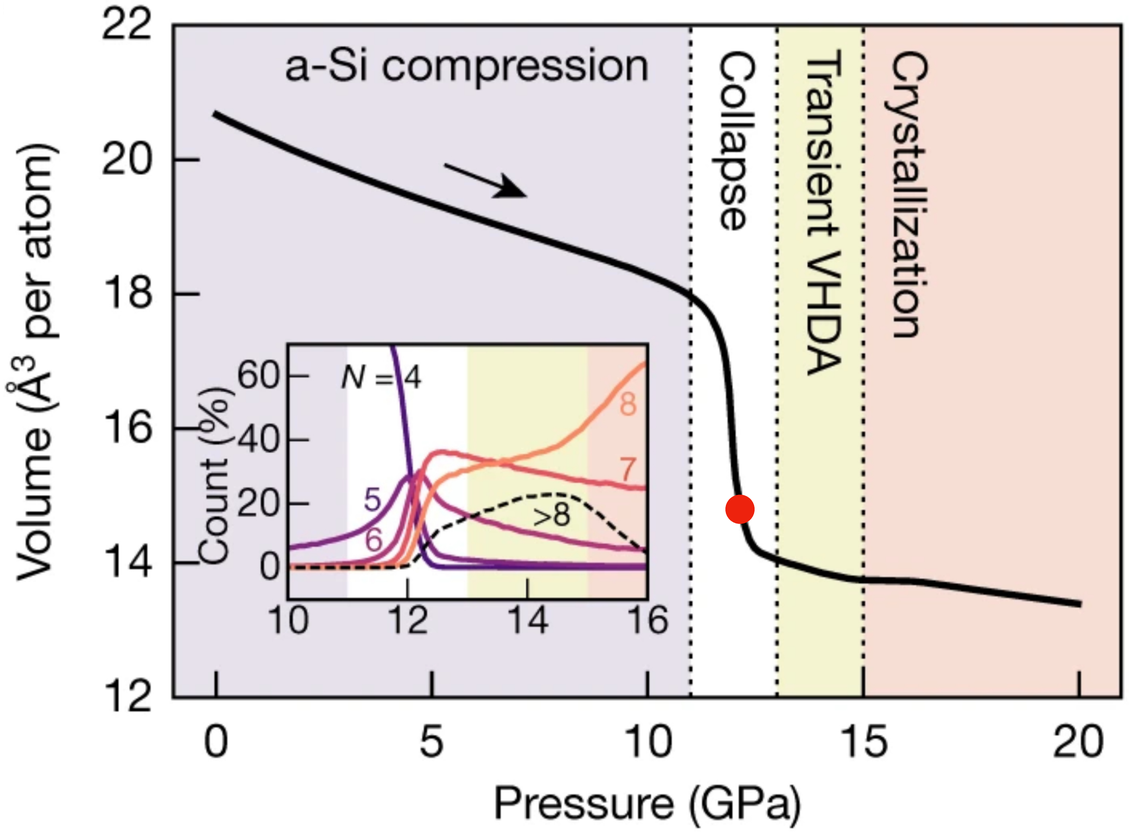


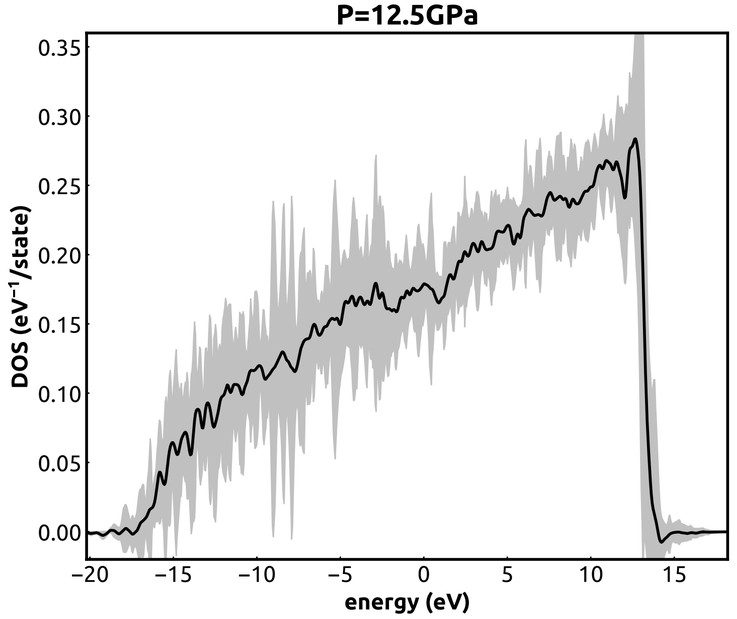


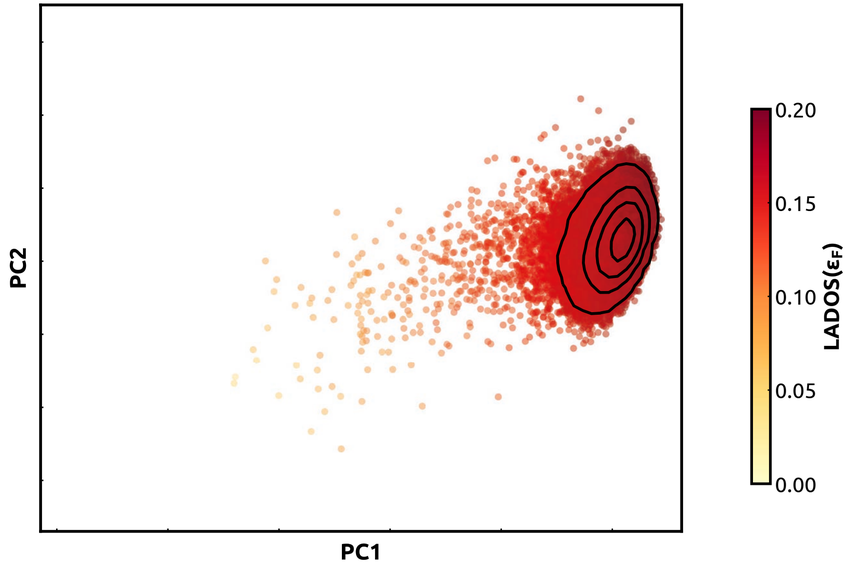


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

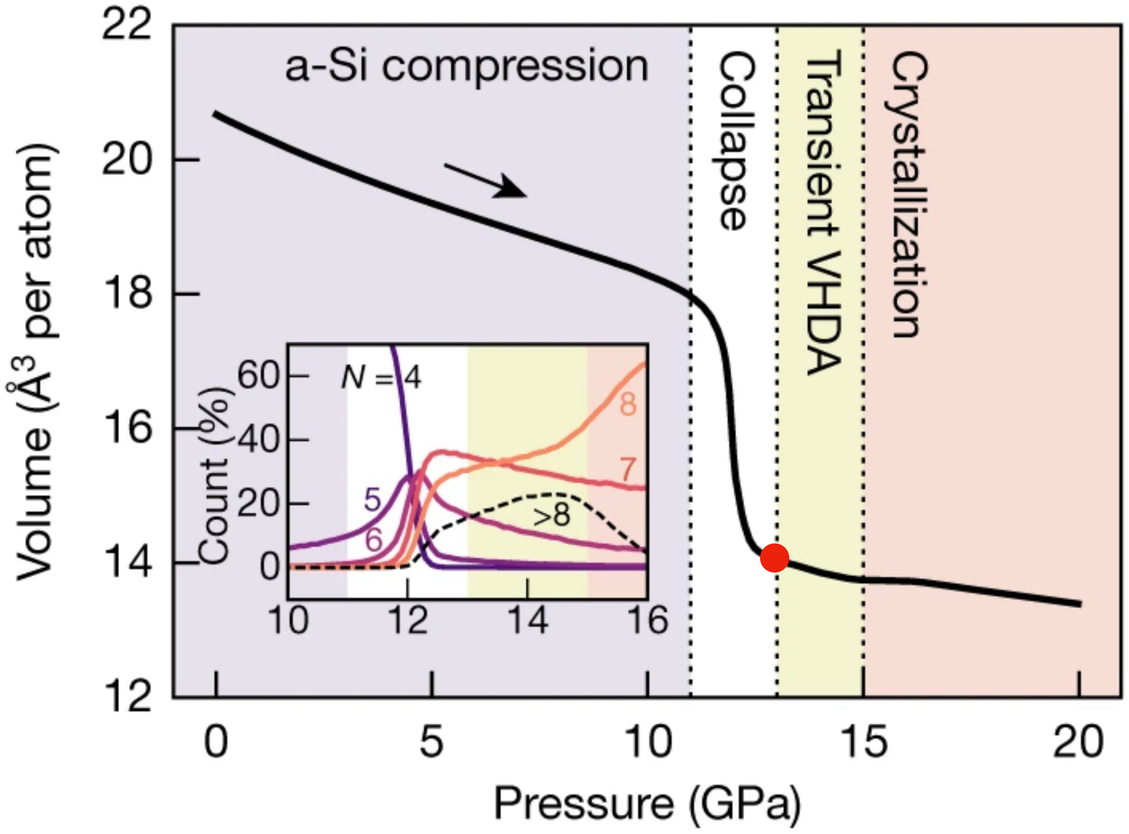


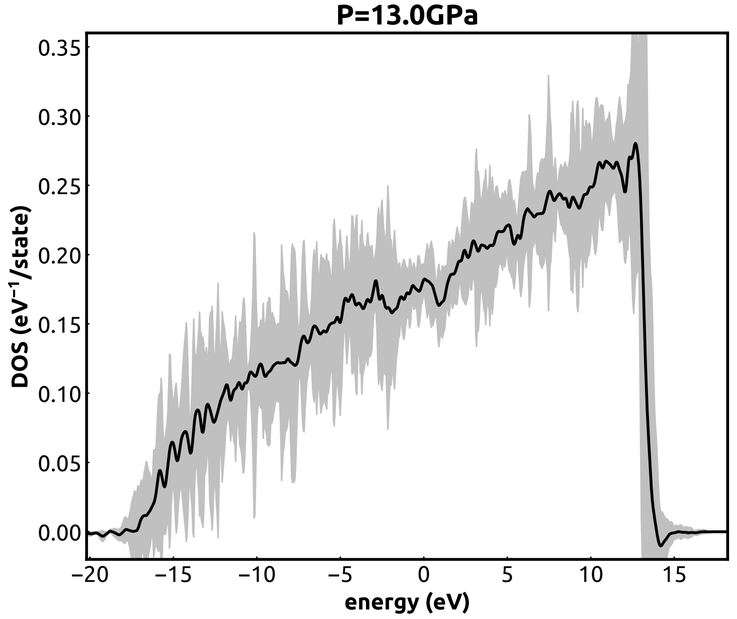


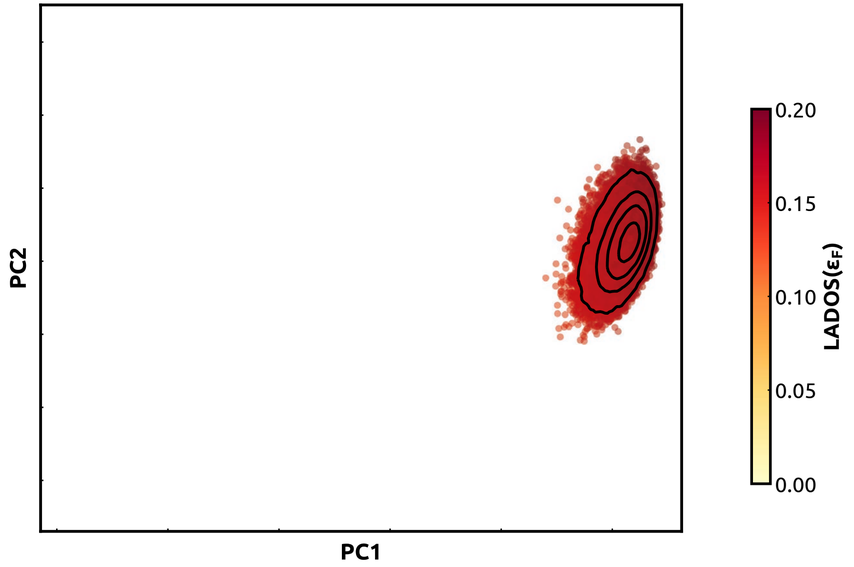


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

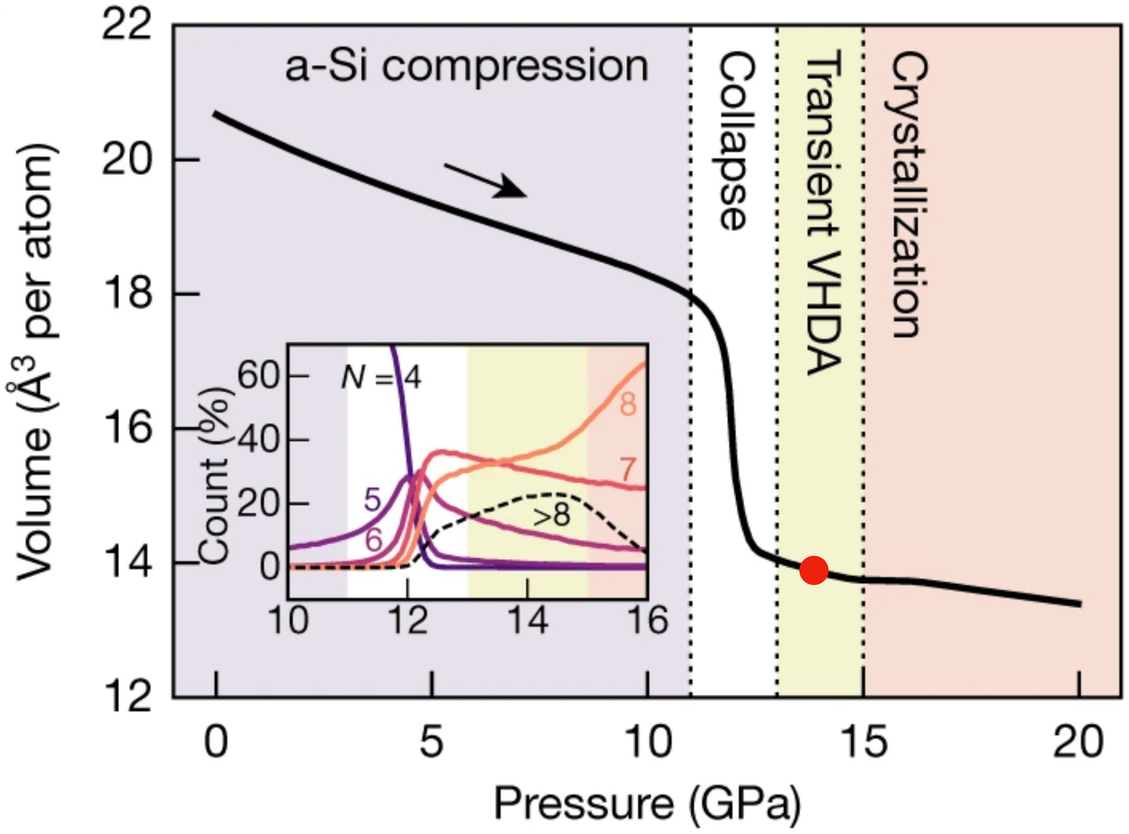


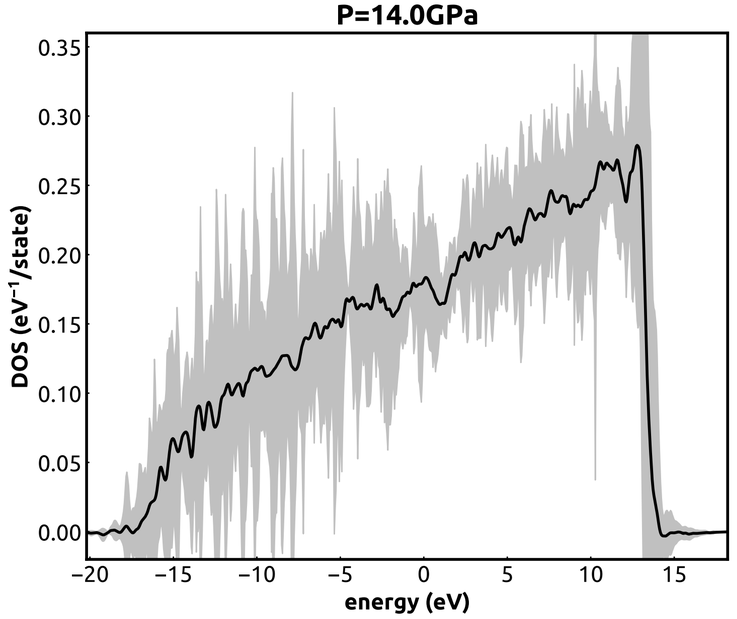


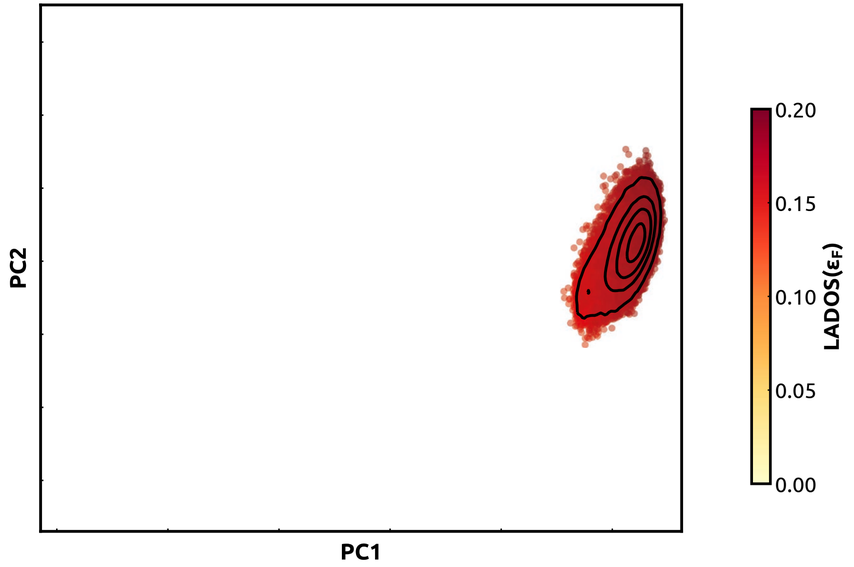


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

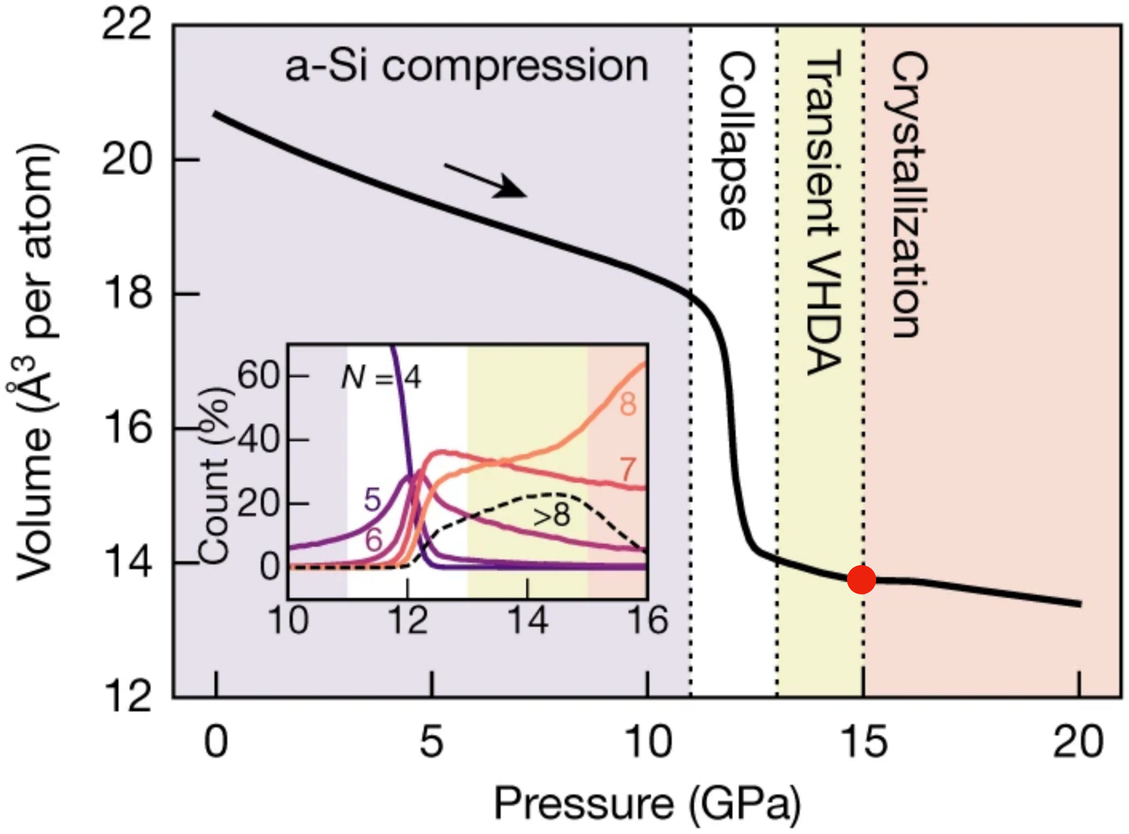


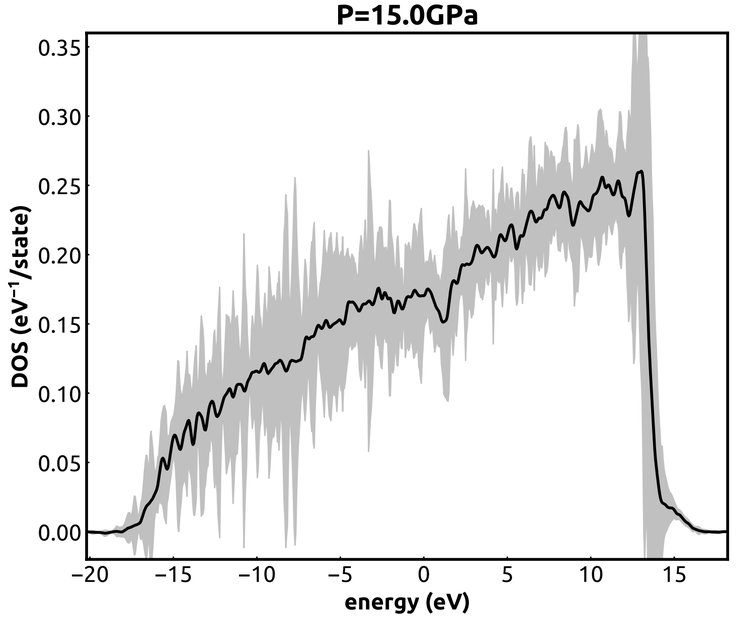


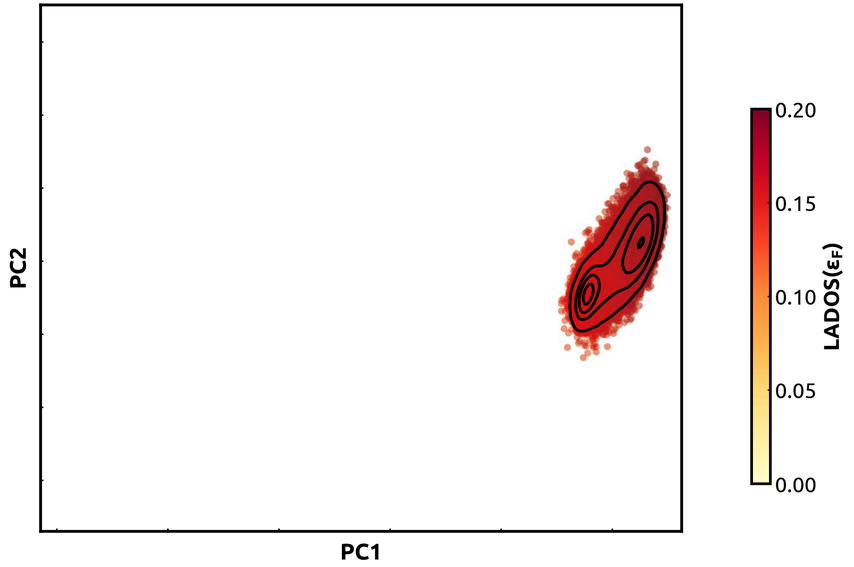


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

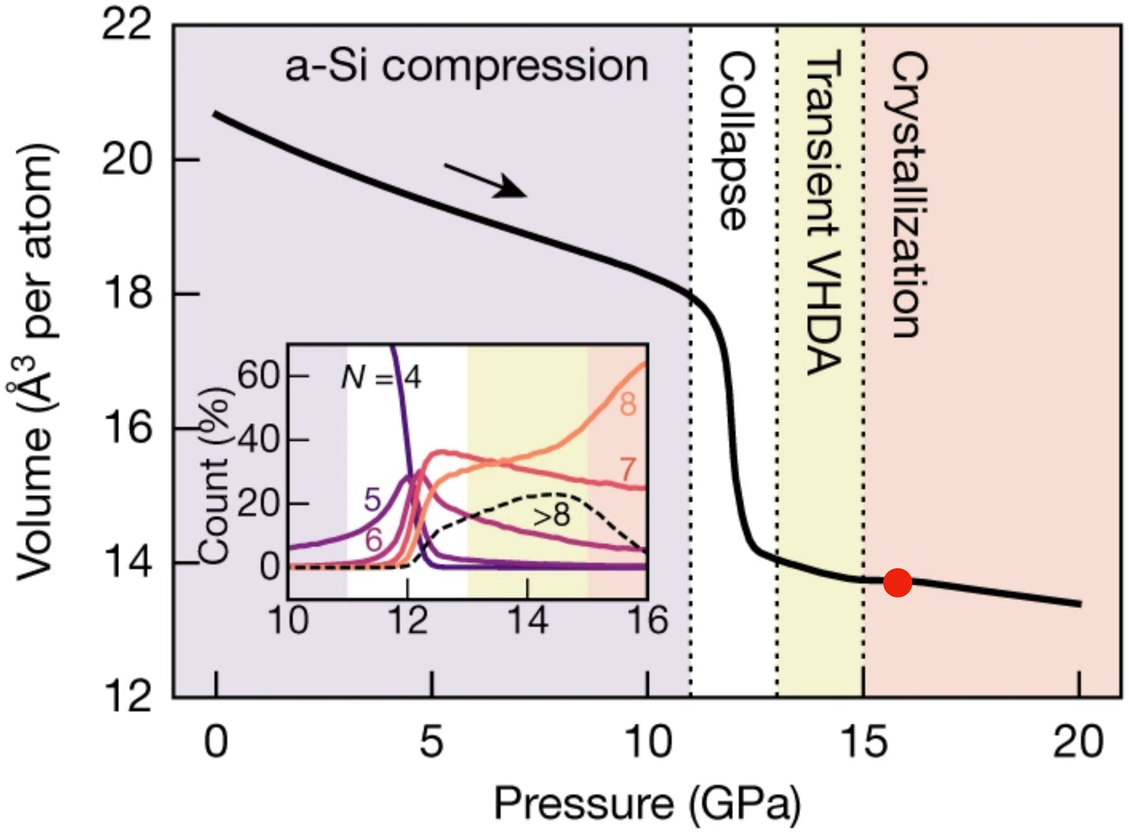


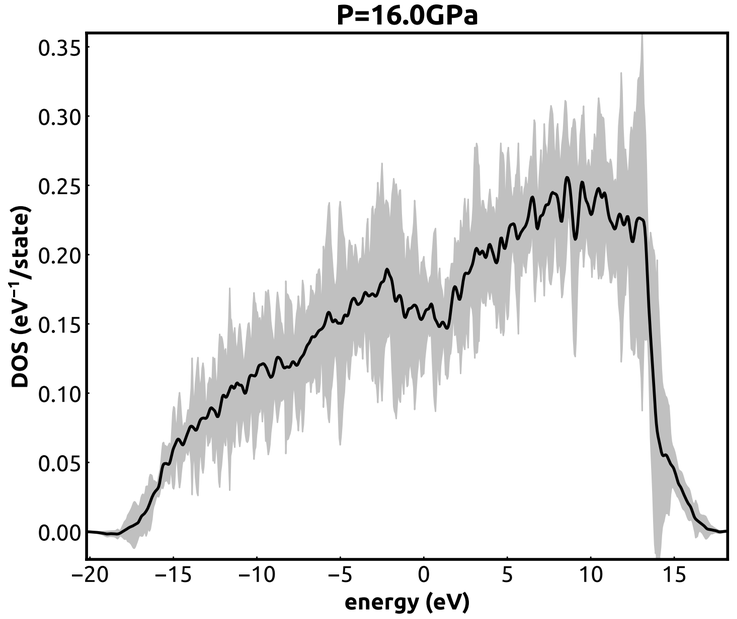


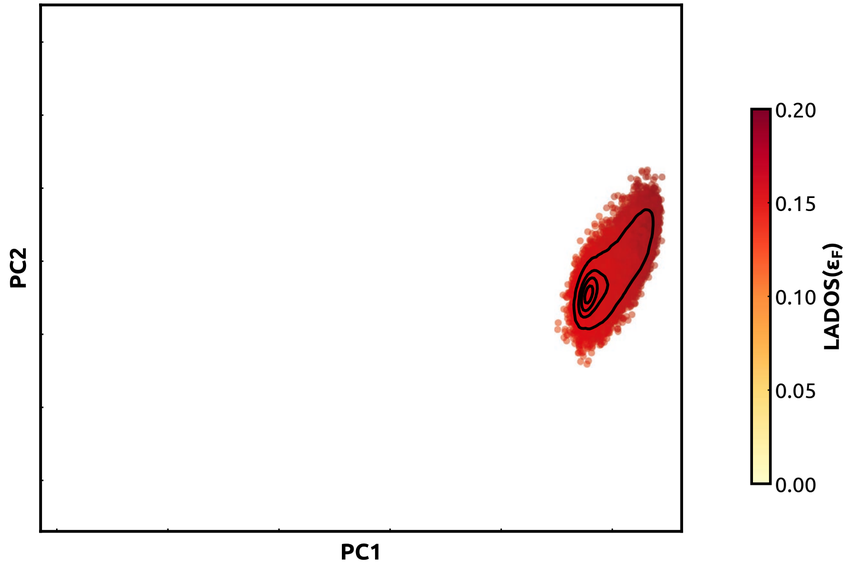


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

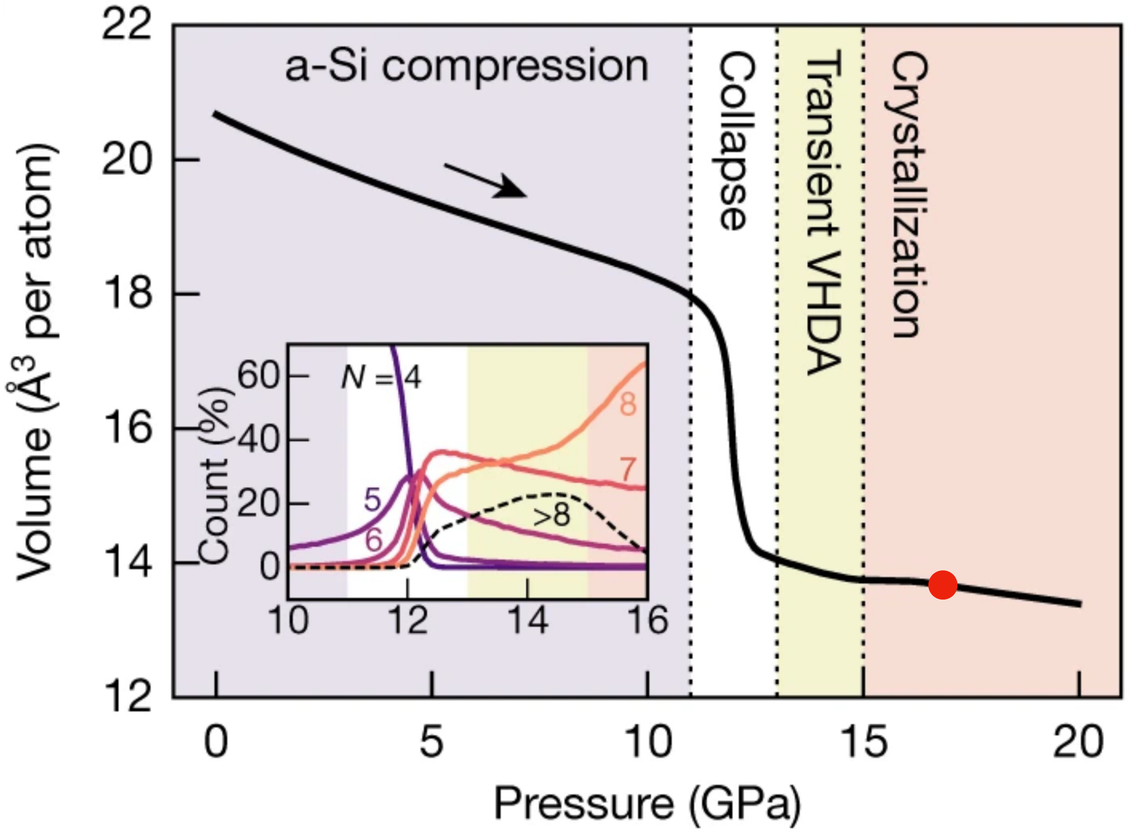


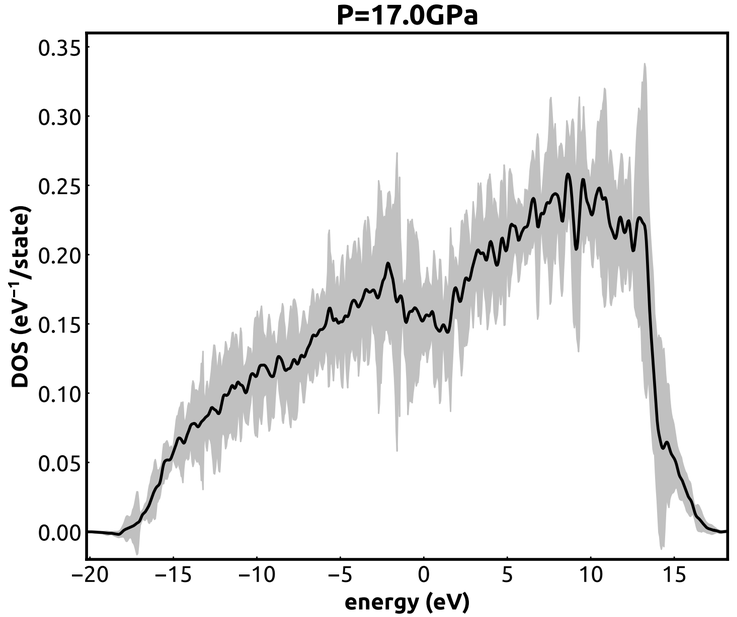


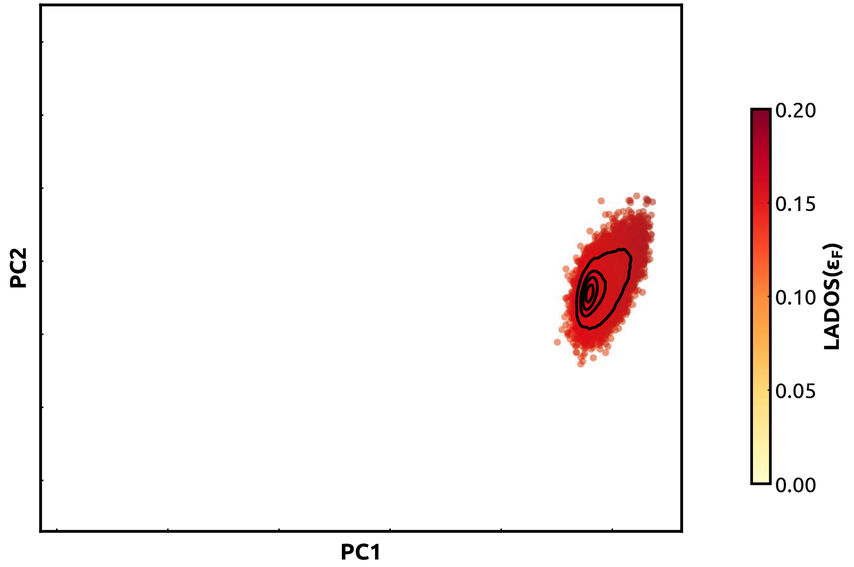


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

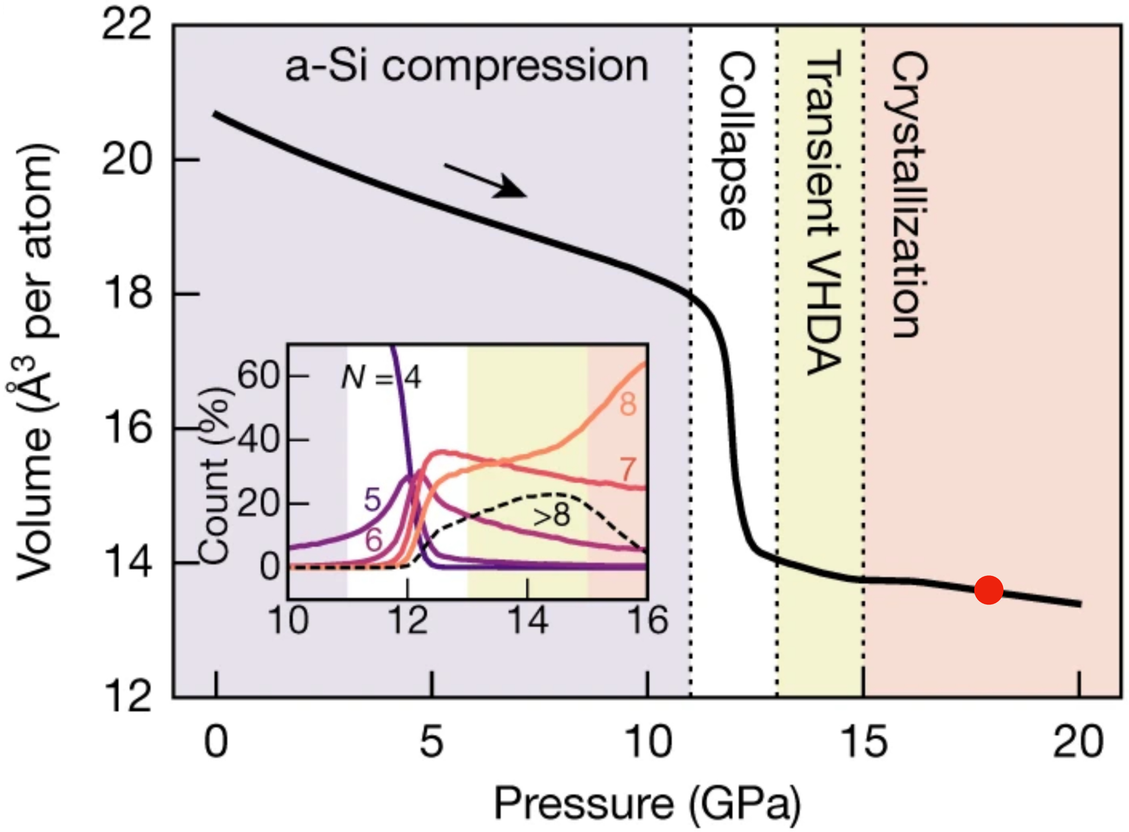


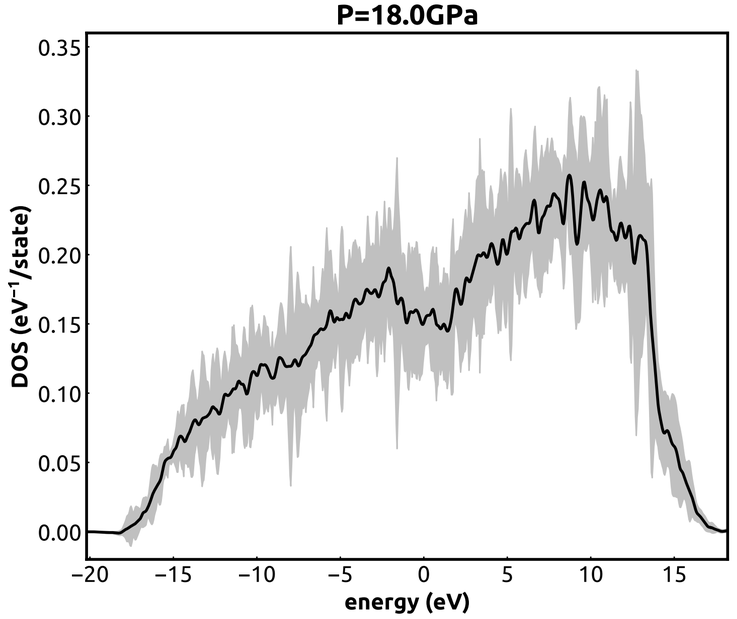


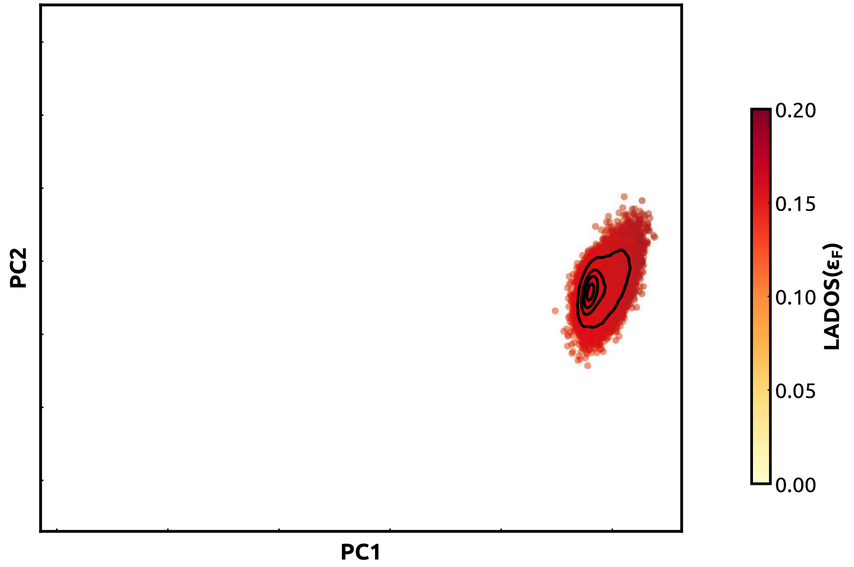


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

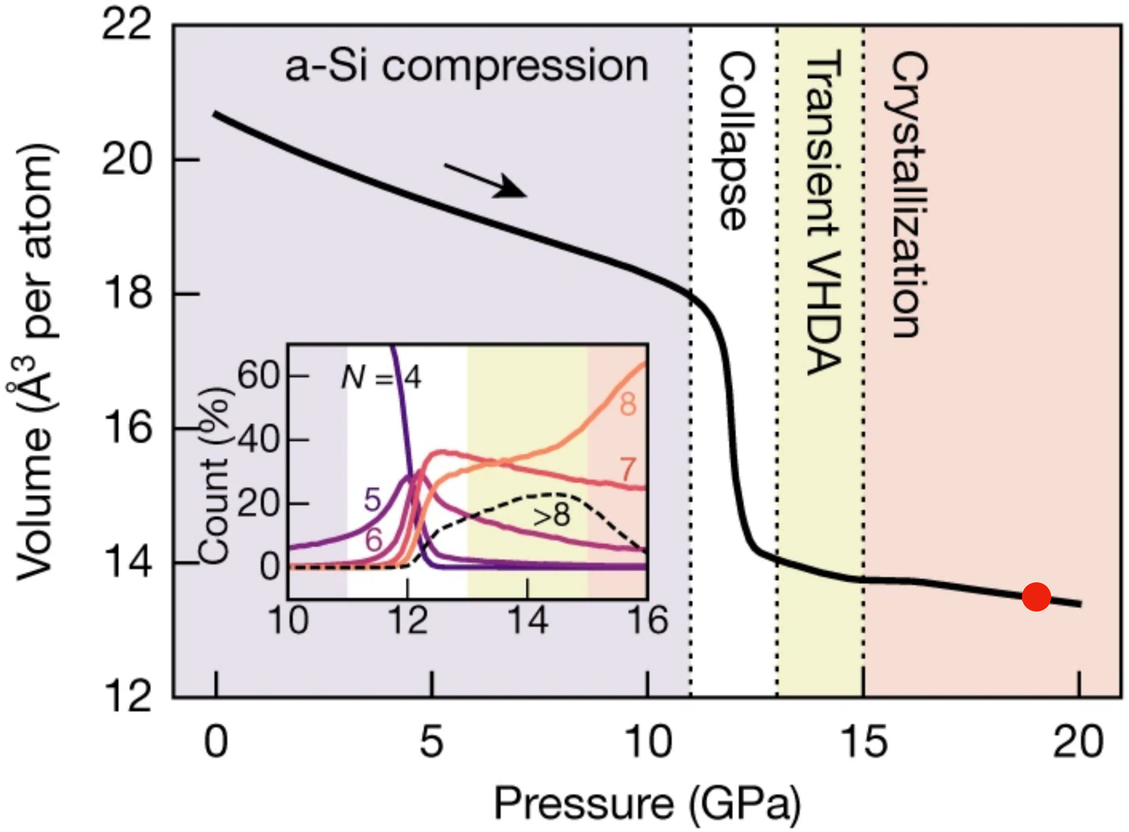


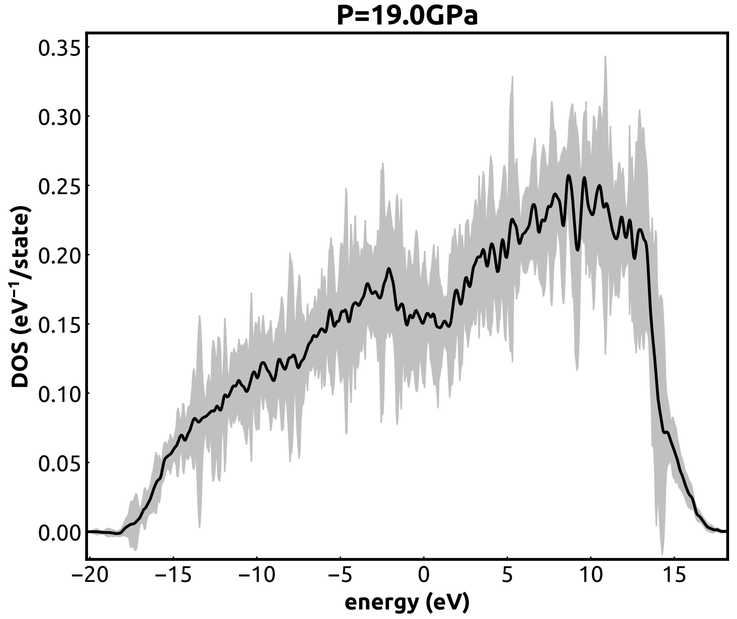


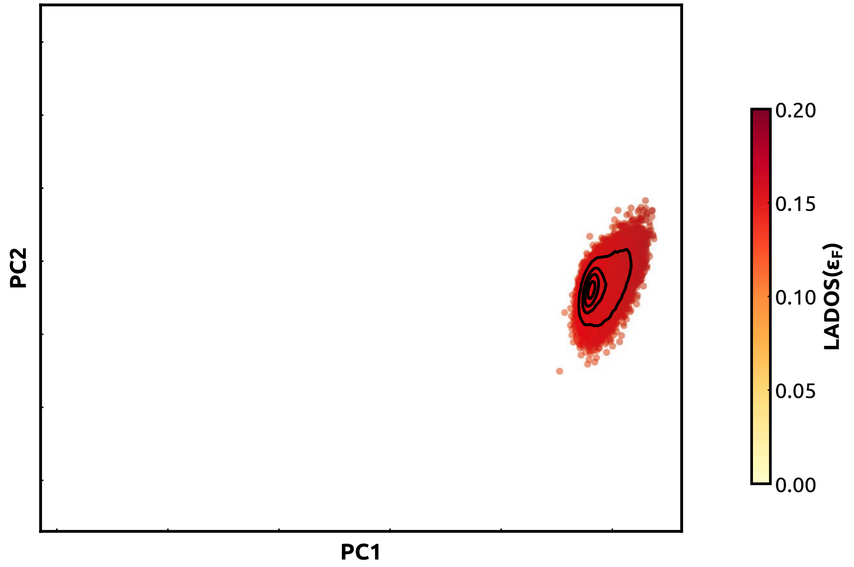


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

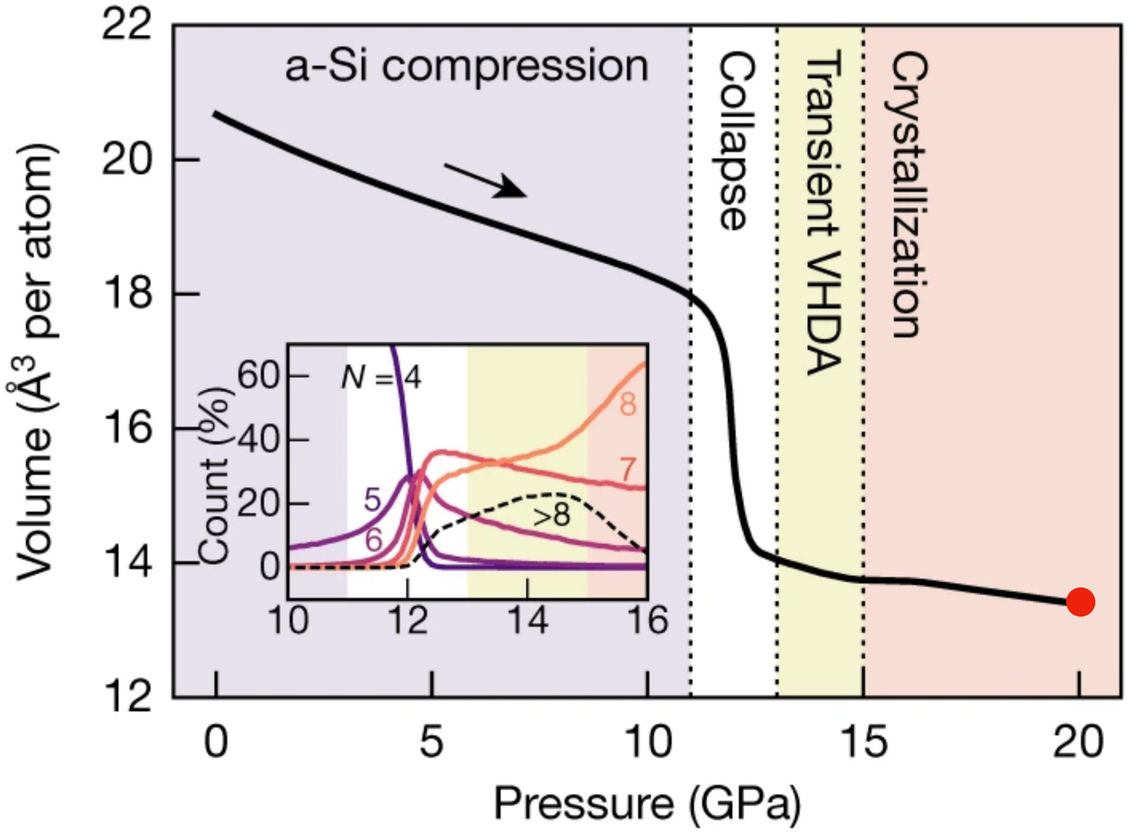


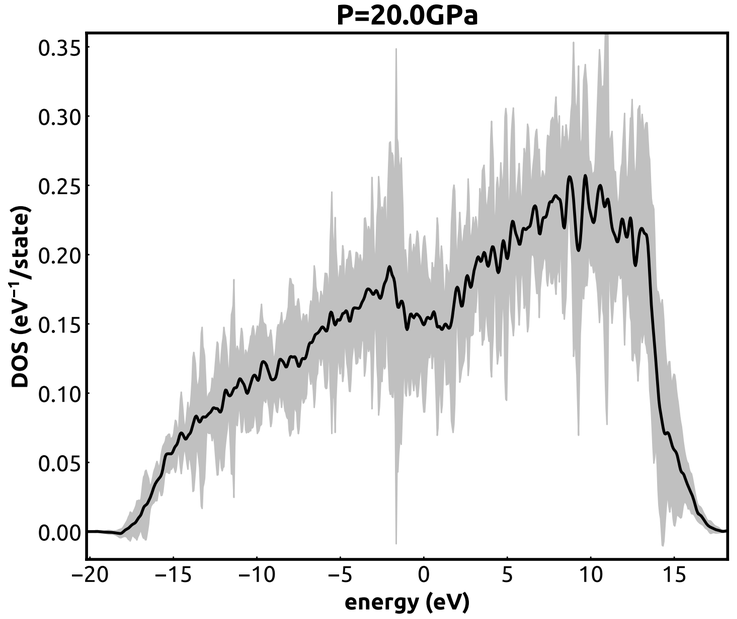


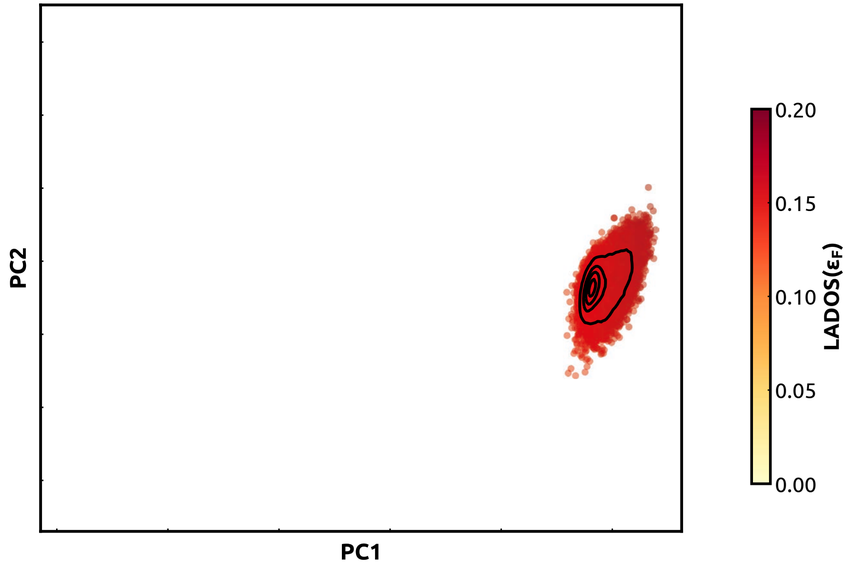


Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

Combining structural and electronic fingerprints

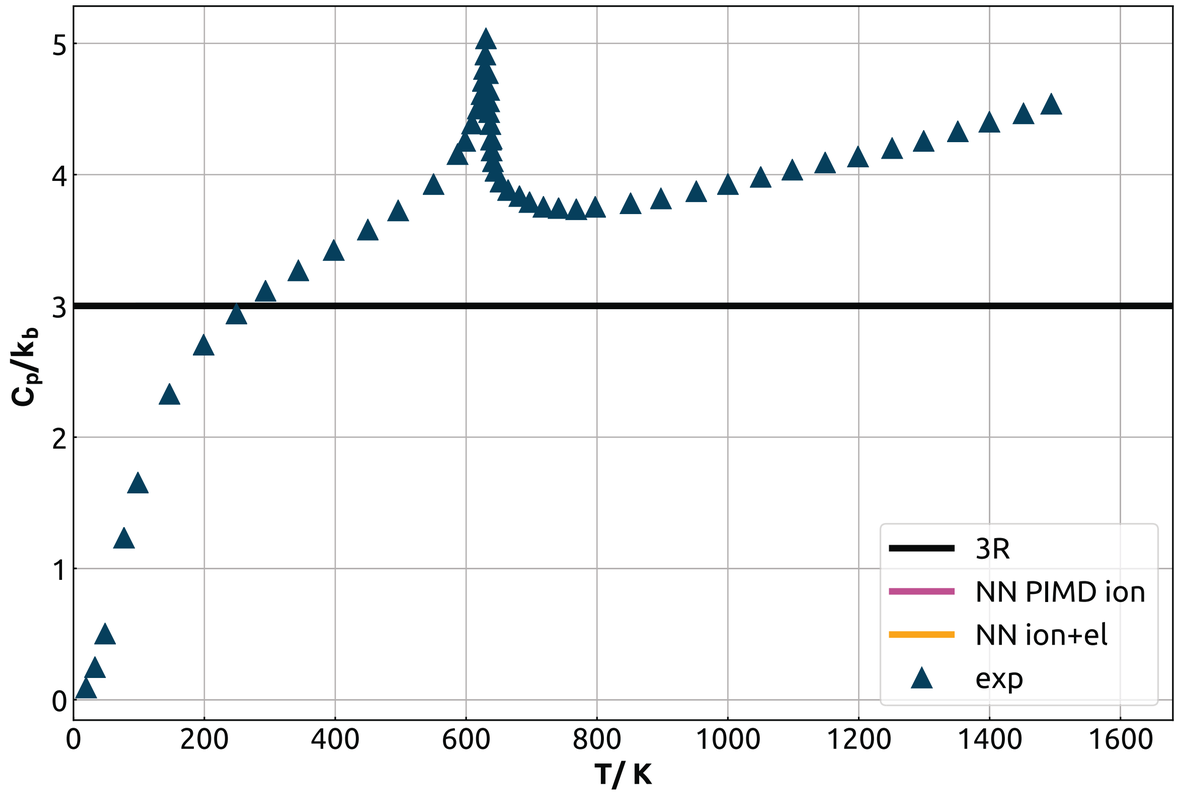






Deringer, V.L., Bernstein, N., Csányi, G., **CBM,** Ceriotti, M., Wilson, M., Drabold, D.A., Elliott, S.R., 2021. Origins of structural and electronic transitions in disordered silicon. Nature 589, 59–64.

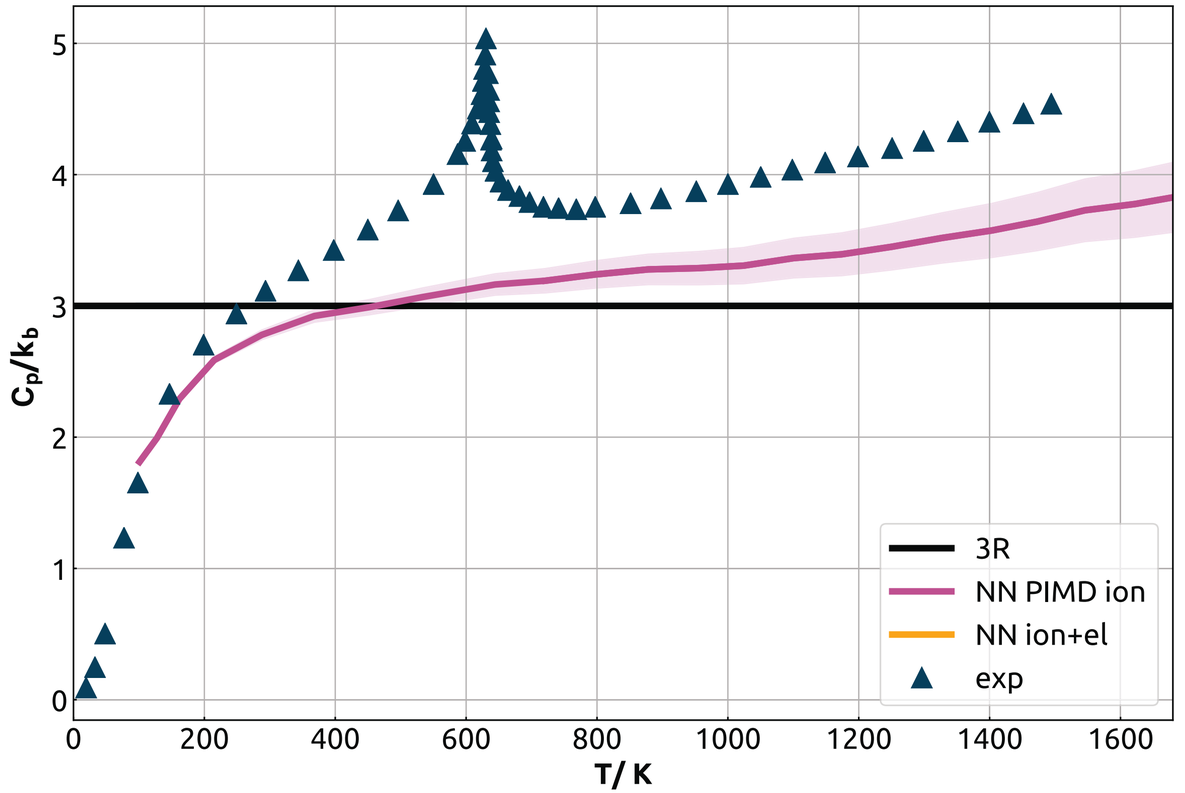
Thermodynamic properties from ground-state



Constant pressure heat capacity of nickel

Lopanitsyna, N., **CBM**, Ceriotti, M., 2021. Finite-temperature materials modeling from the quantum nuclei to the hot electron regime. Phys. Rev. Materials 5, 043802

Thermodynamic properties from ground-state



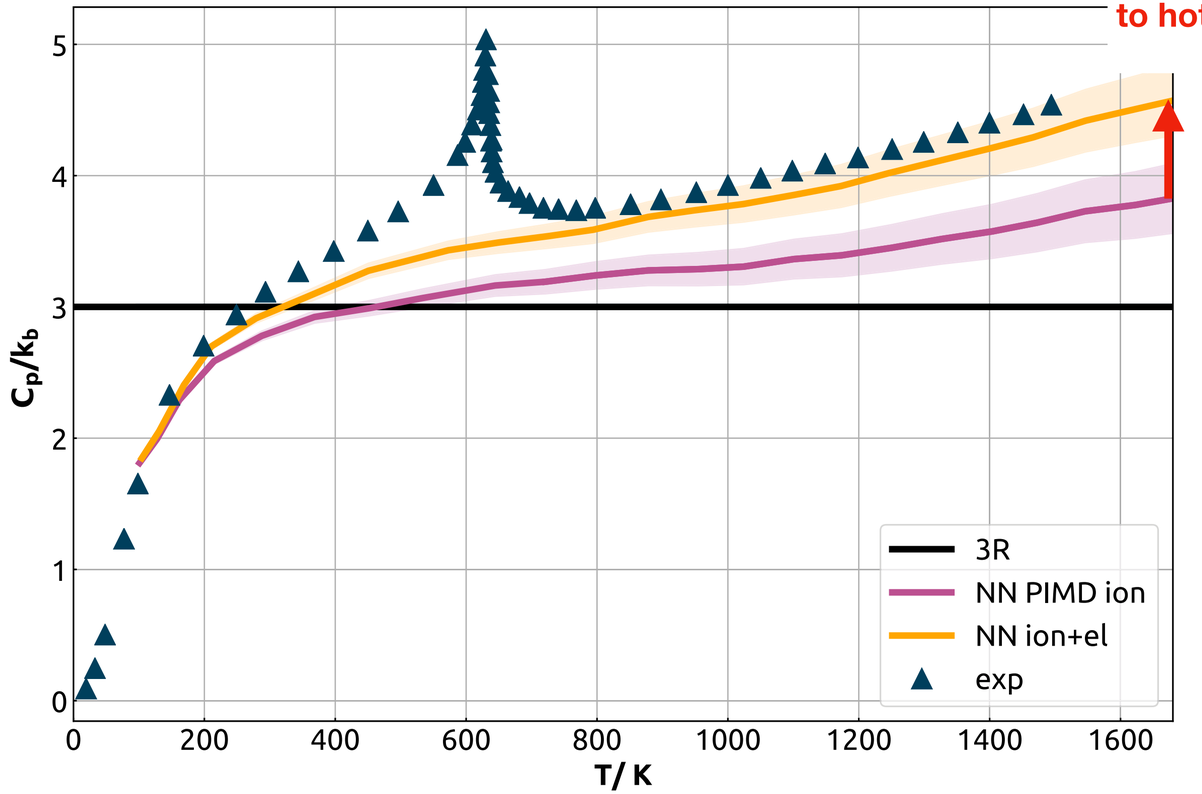
Constant pressure heat capacity of nickel

Lopanitsyna, N., **CBM**, Ceriotti, M., 2021. Finite-temperature materials modeling from the quantum nuclei to the hot electron regime. Phys. Rev. Materials 5, 043802

Thermodynamic properties from ground-state

$$U_{\mathrm{DOS}}^{\mathrm{el}}(T) = \int_{-\infty}^{\infty} \varepsilon \mathrm{DOS}(\varepsilon) \left[f_{\mathrm{FD}}(\varepsilon - \mu(T), T) - f_{\mathrm{FD}}(\varepsilon - \varepsilon_F, T = 0) \right] \mathrm{d}\varepsilon$$

$$C_P^{\mathrm{el}}(T) = \frac{\partial \langle U_{\mathrm{DOS}}^{\mathrm{el}}(T) \rangle}{\partial T}$$

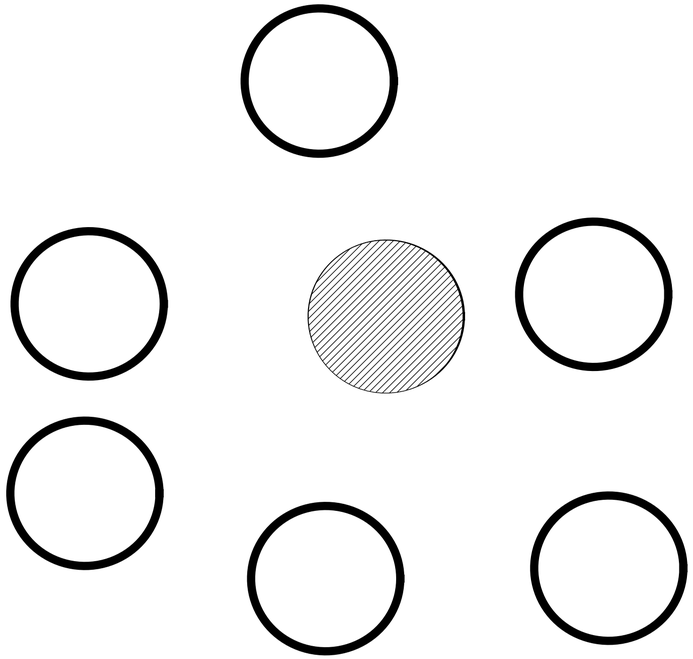


Constant pressure heat capacity of nickel

Lopanitsyna, N., **CBM**, Ceriotti, M., 2021. Finite-temperature materials modeling from the quantum nuclei to the hot electron regime. Phys. Rev. Materials 5, 043802

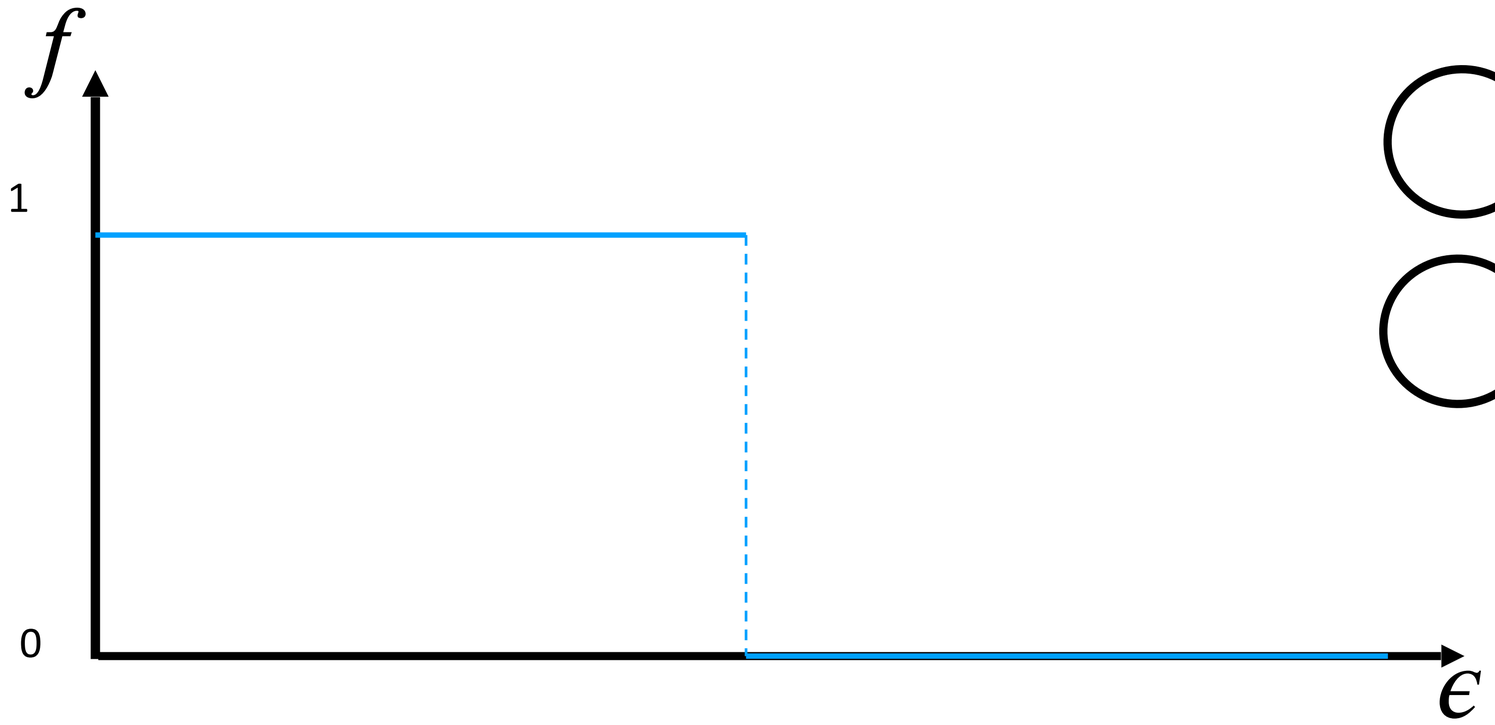
The finite-temperature problem

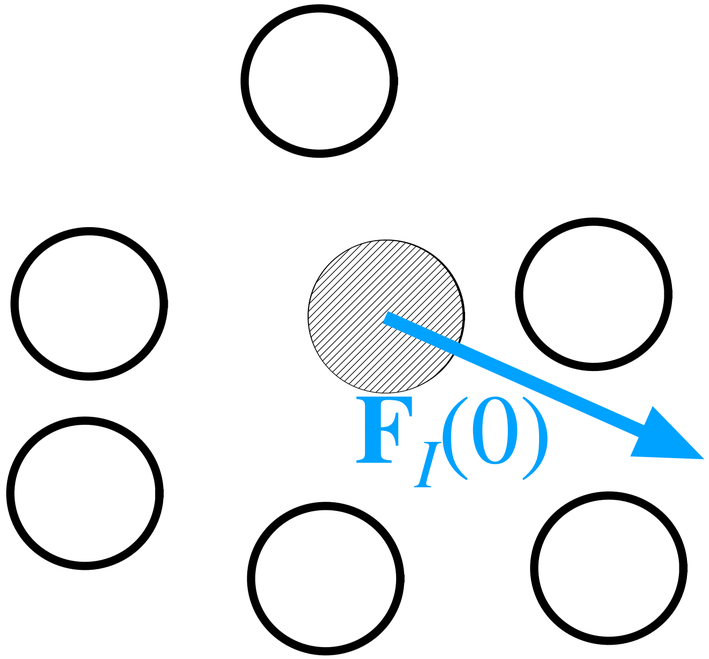
$$\mathbf{F}_I = \mathbf{F}_I\left(\mathbf{R}_1, \ldots, \mathbf{R}_N; T^{el}\right)$$



The finite-temperature problem

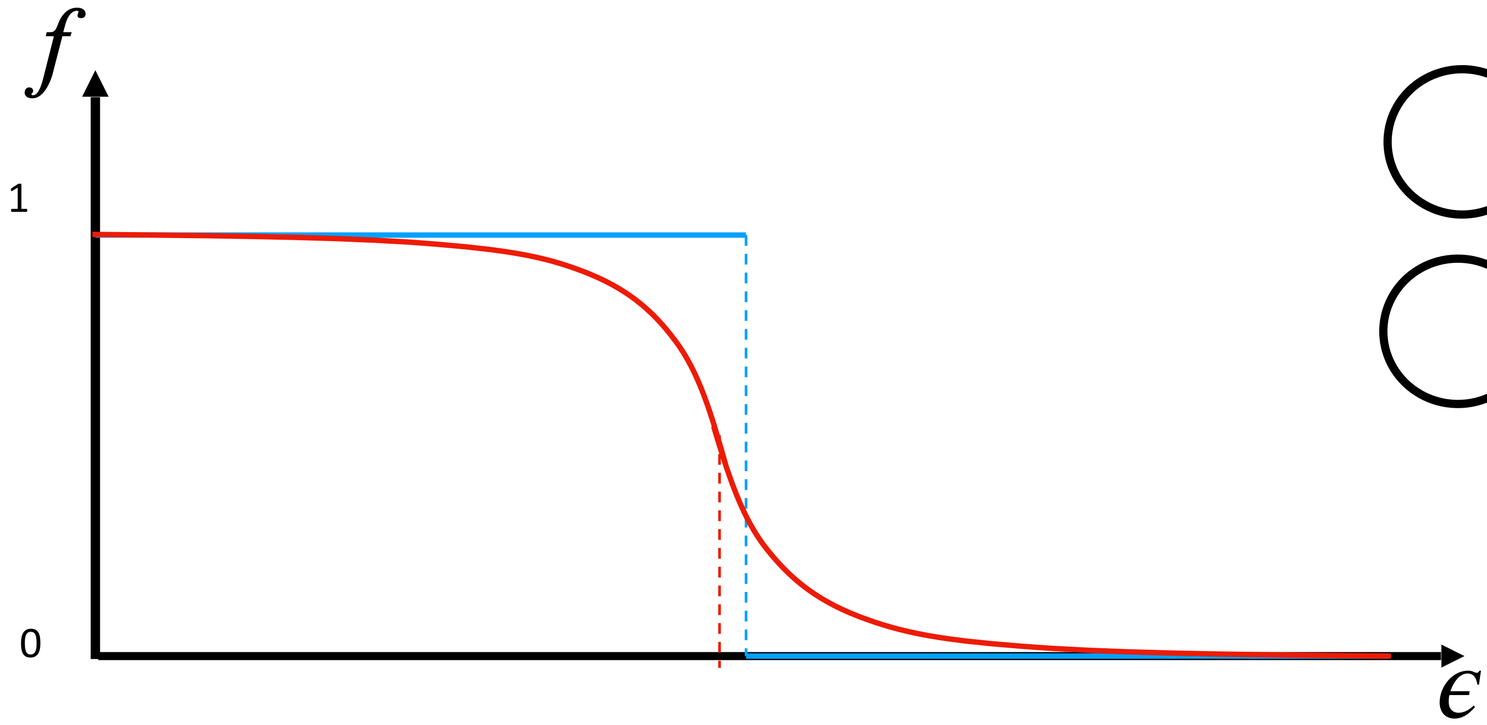
$$\mathbf{F}_I = \mathbf{F}_I\left(\mathbf{R}_1, \ldots, \mathbf{R}_N; T^{el}\right)$$

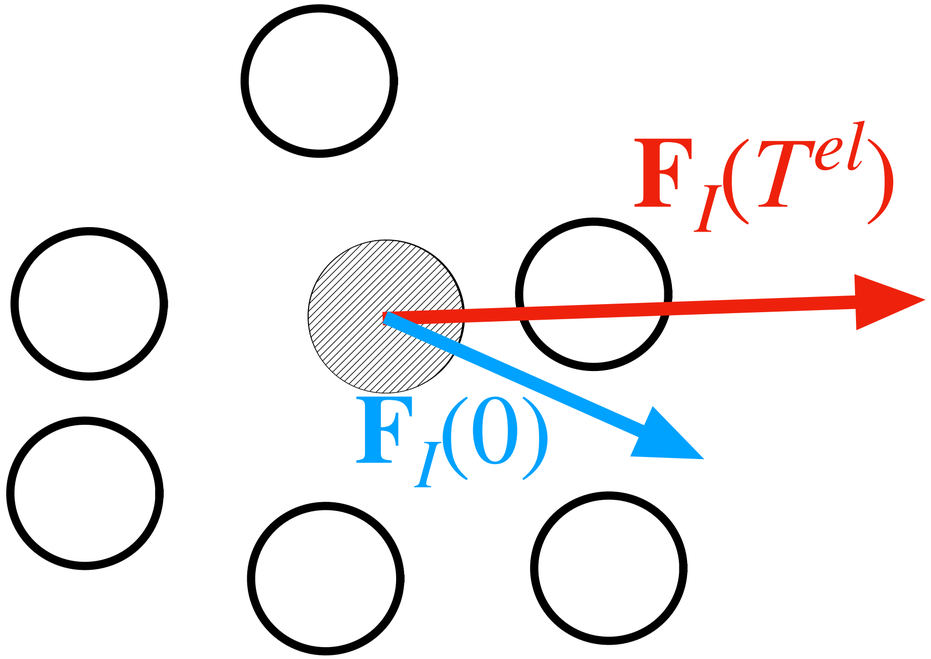




The finite-temperature problem

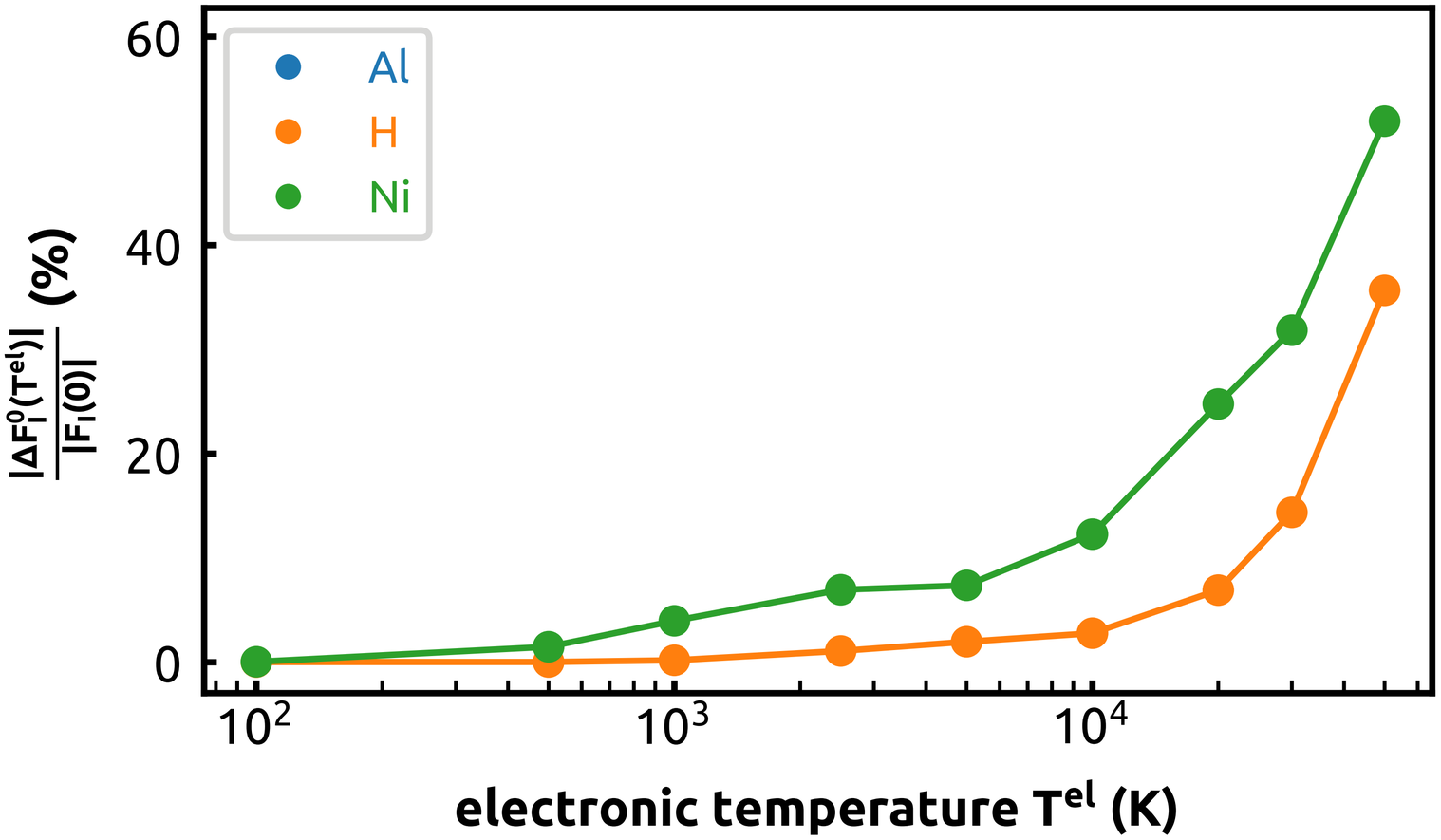
$$\mathbf{F}_I = \mathbf{F}_I\left(\mathbf{R}_1, \ldots, \mathbf{R}_N; T^{el}\right)$$





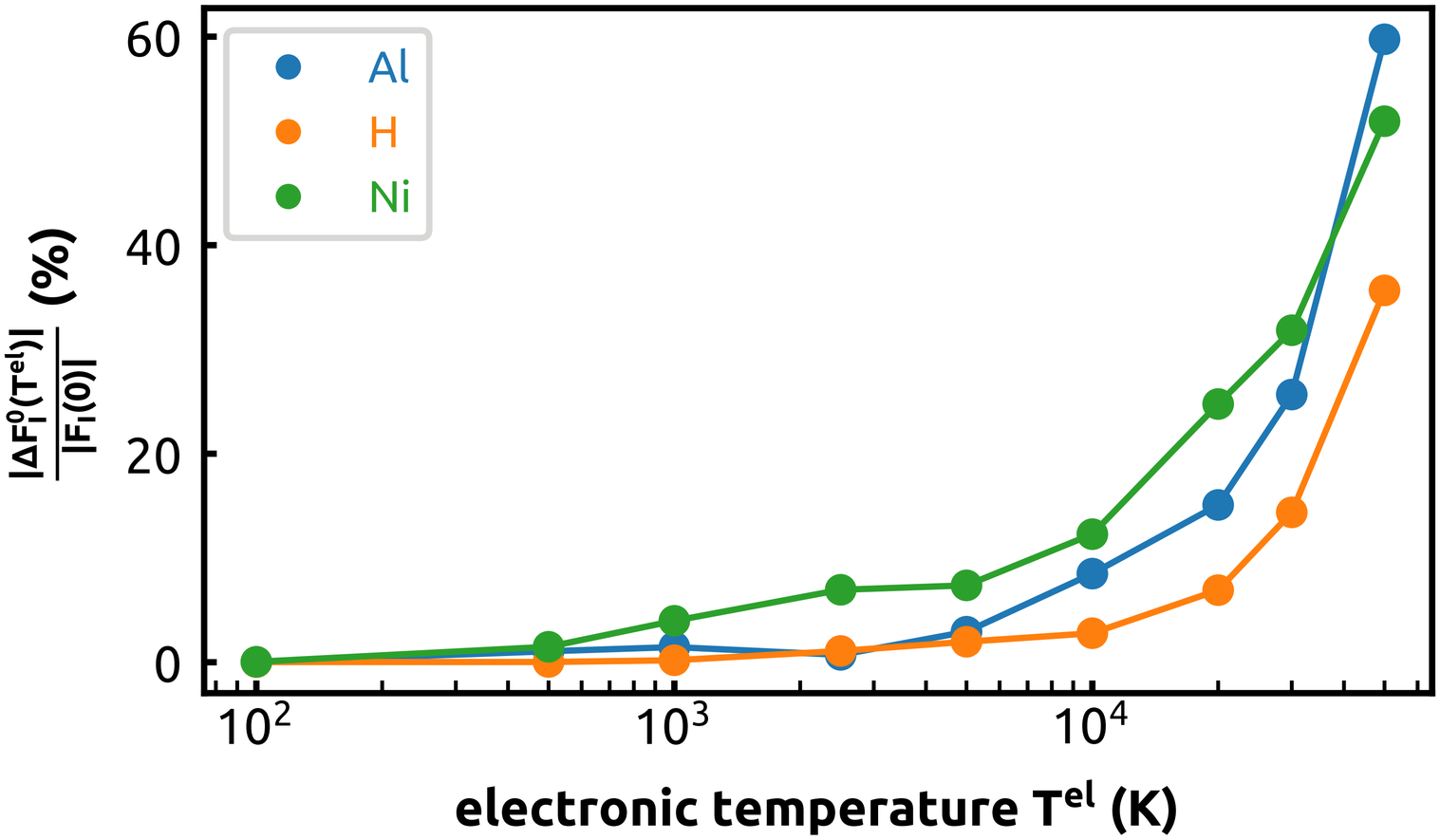
The finite-temperature problem

- High errors with respect to T=0 forces

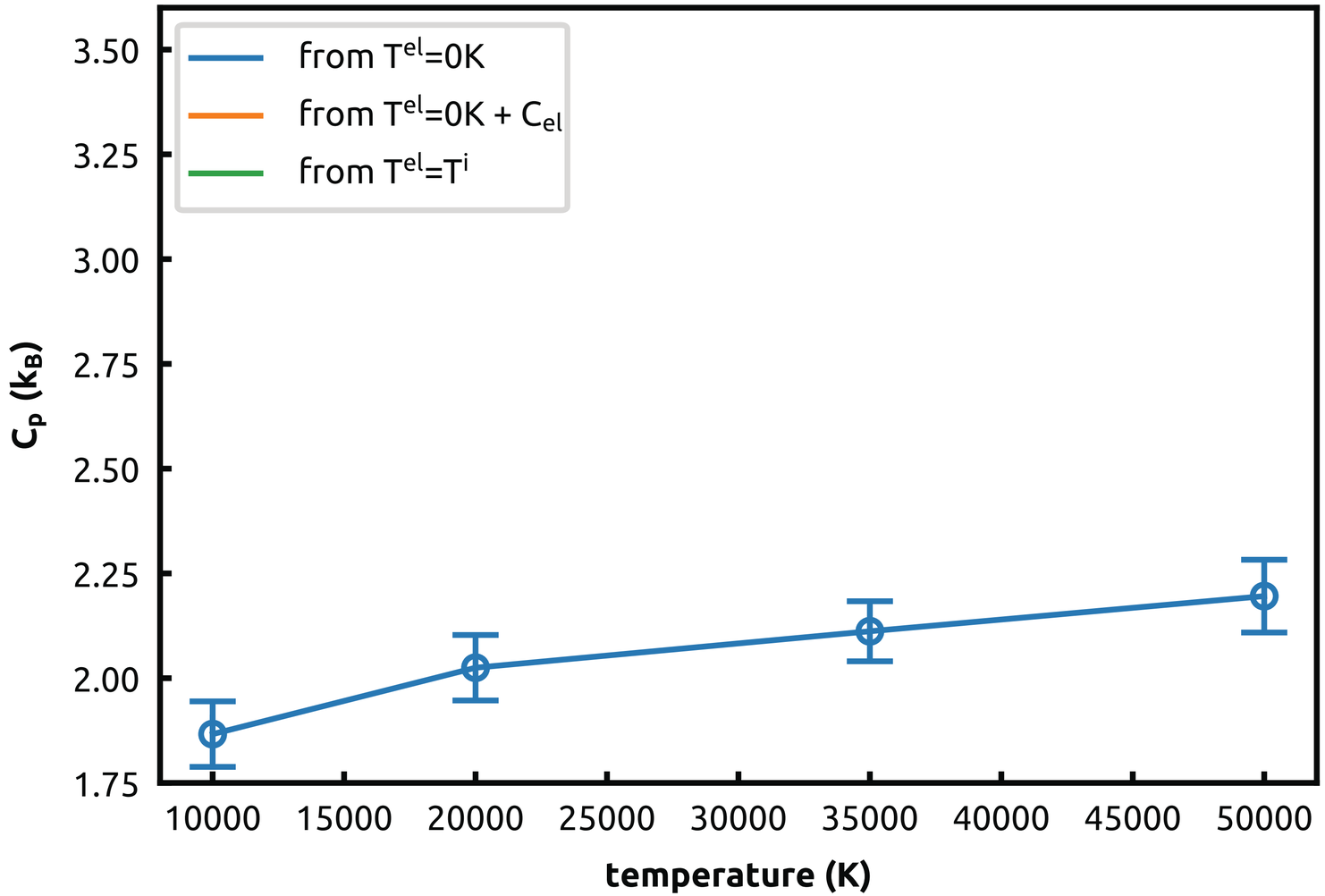


The finite-temperature problem

- High errors with respect to T=0 forces

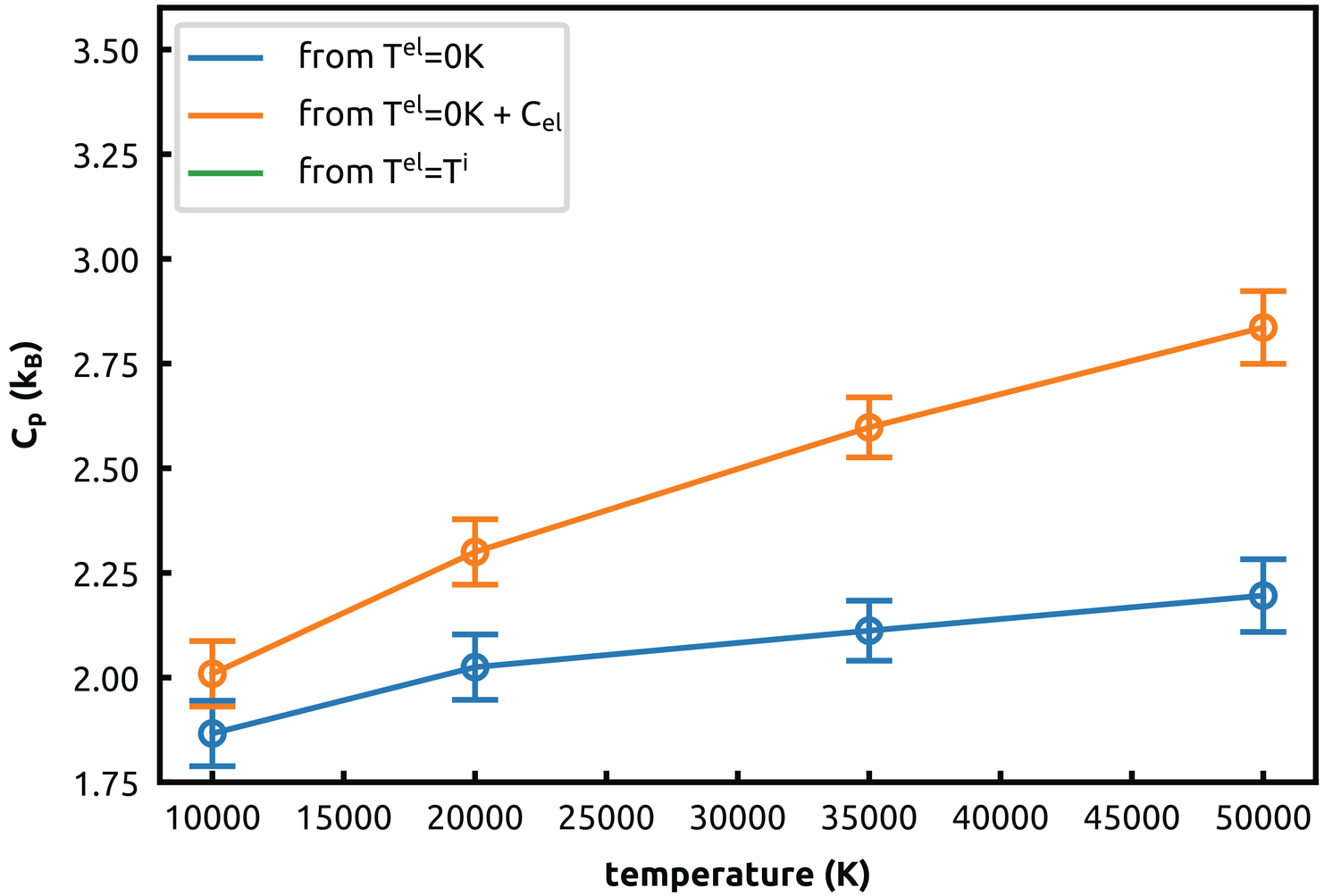


The finite-temperature problem



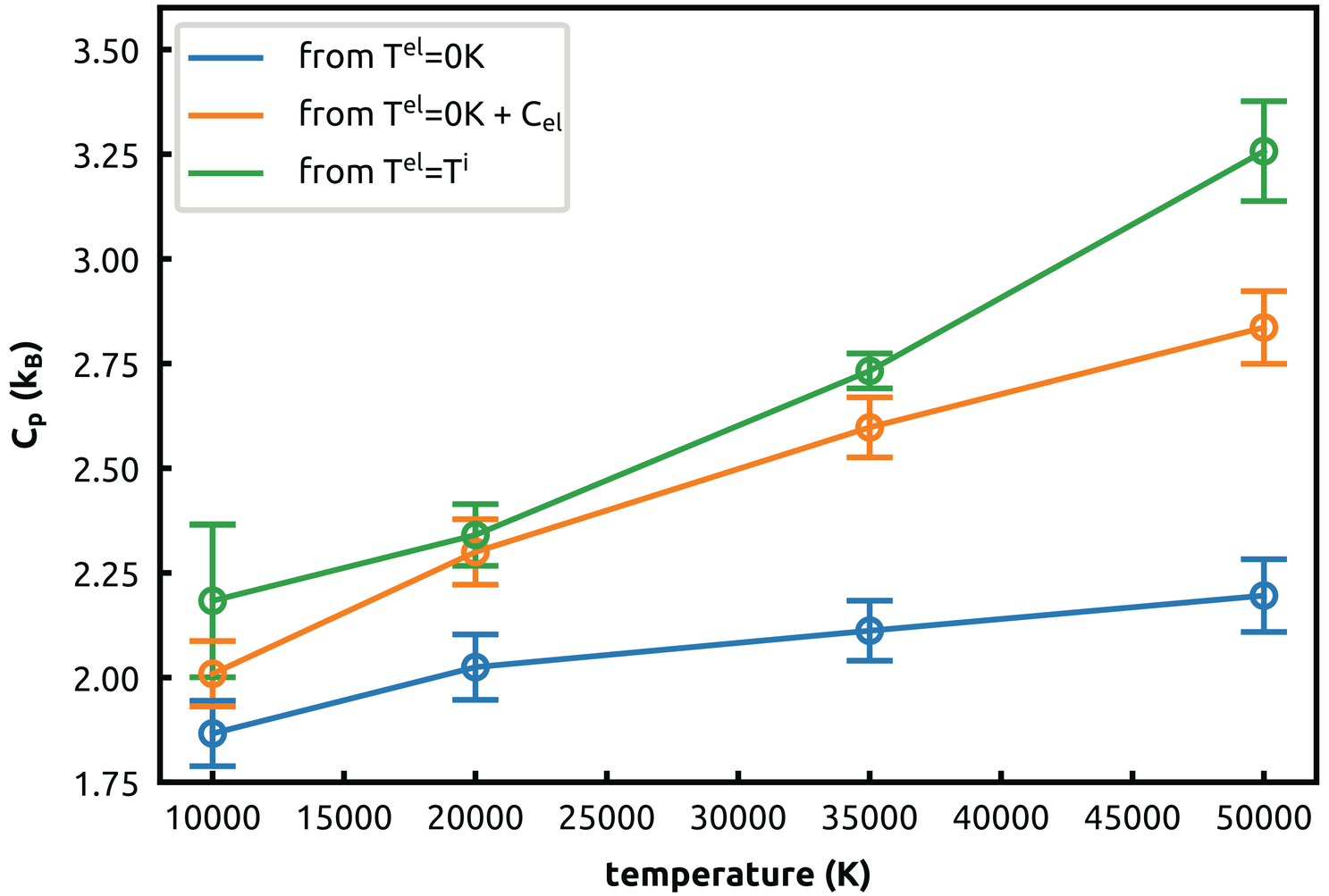
Constant pressure heat capacity of hydrogen at 400GPa from DFT simulations

The finite-temperature problem



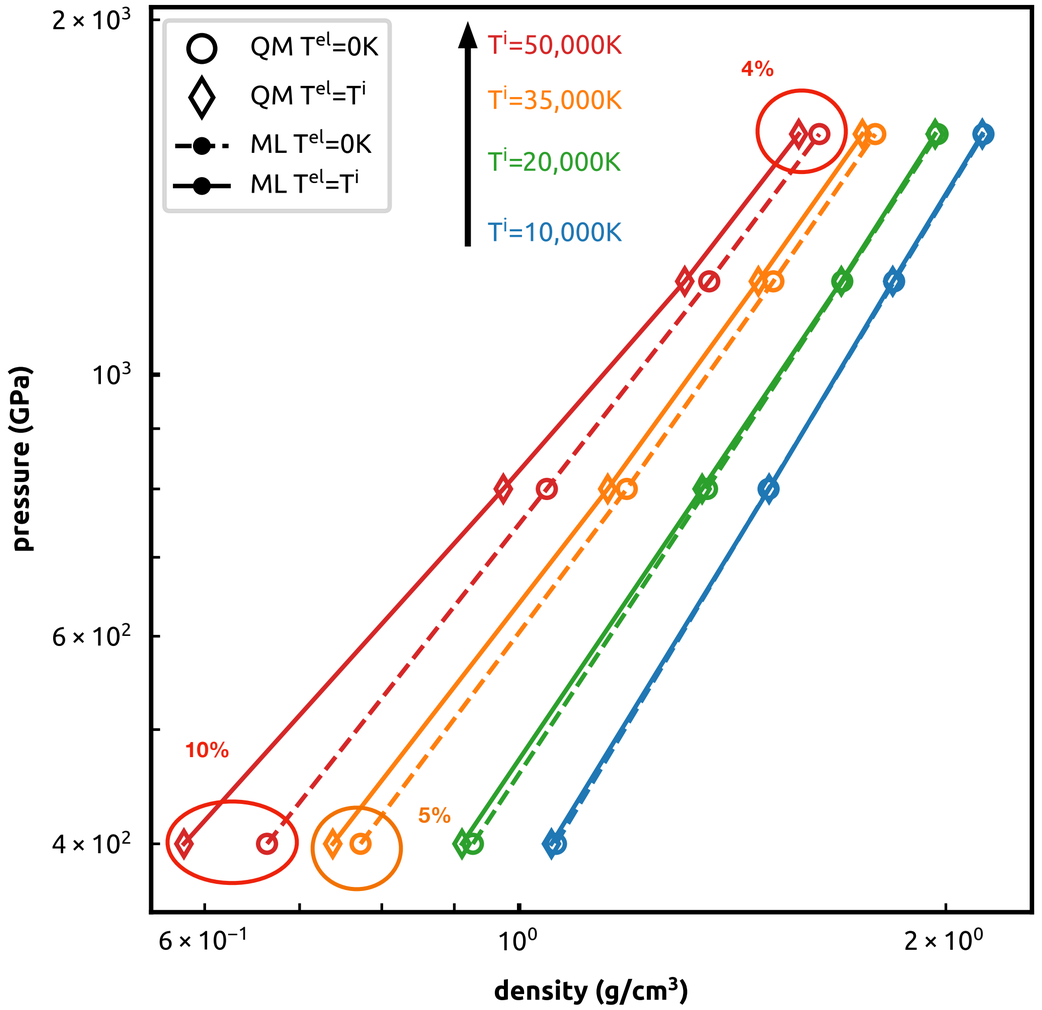
Constant pressure heat capacity of hydrogen at 400GPa from DFT simulations

The finite-temperature problem



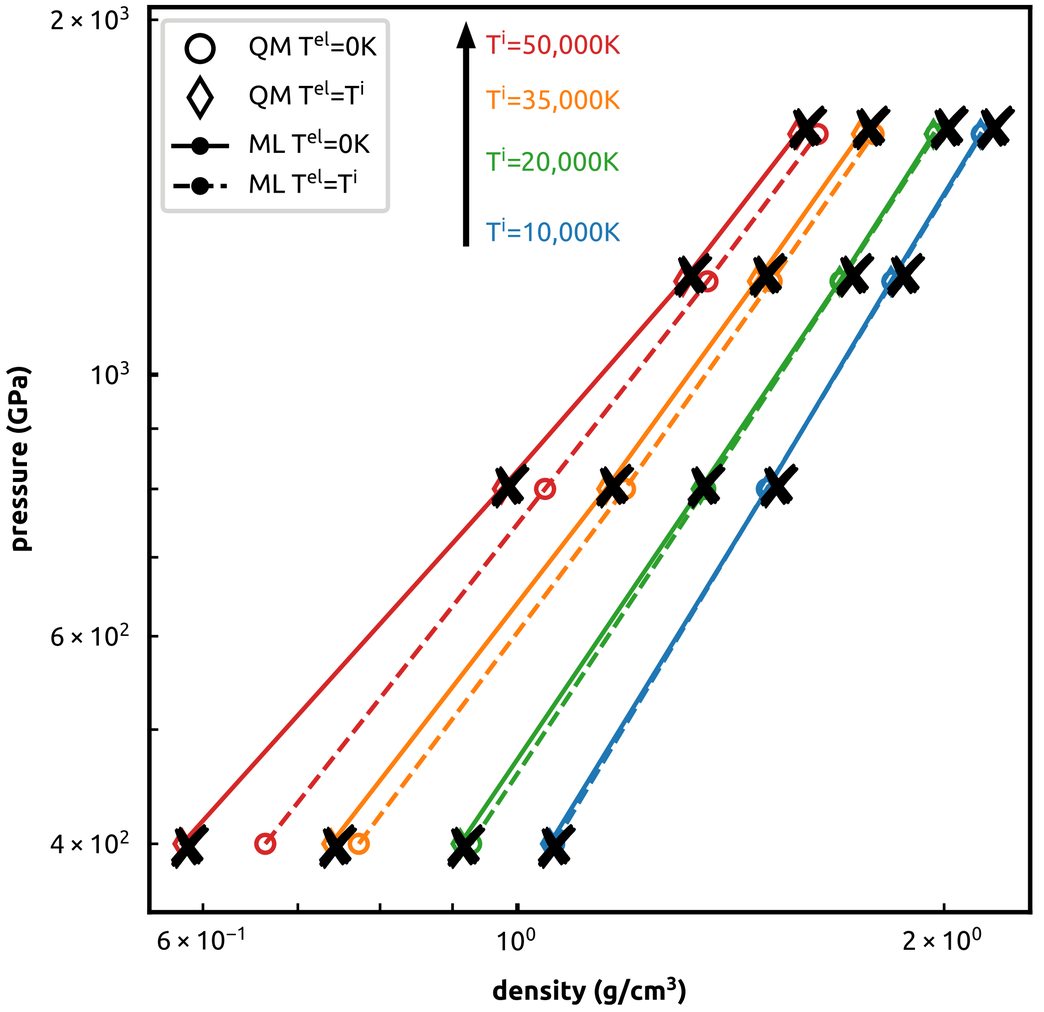
Constant pressure heat capacity of hydrogen at 400GPa from DFT simulations

The finite-temperature problem

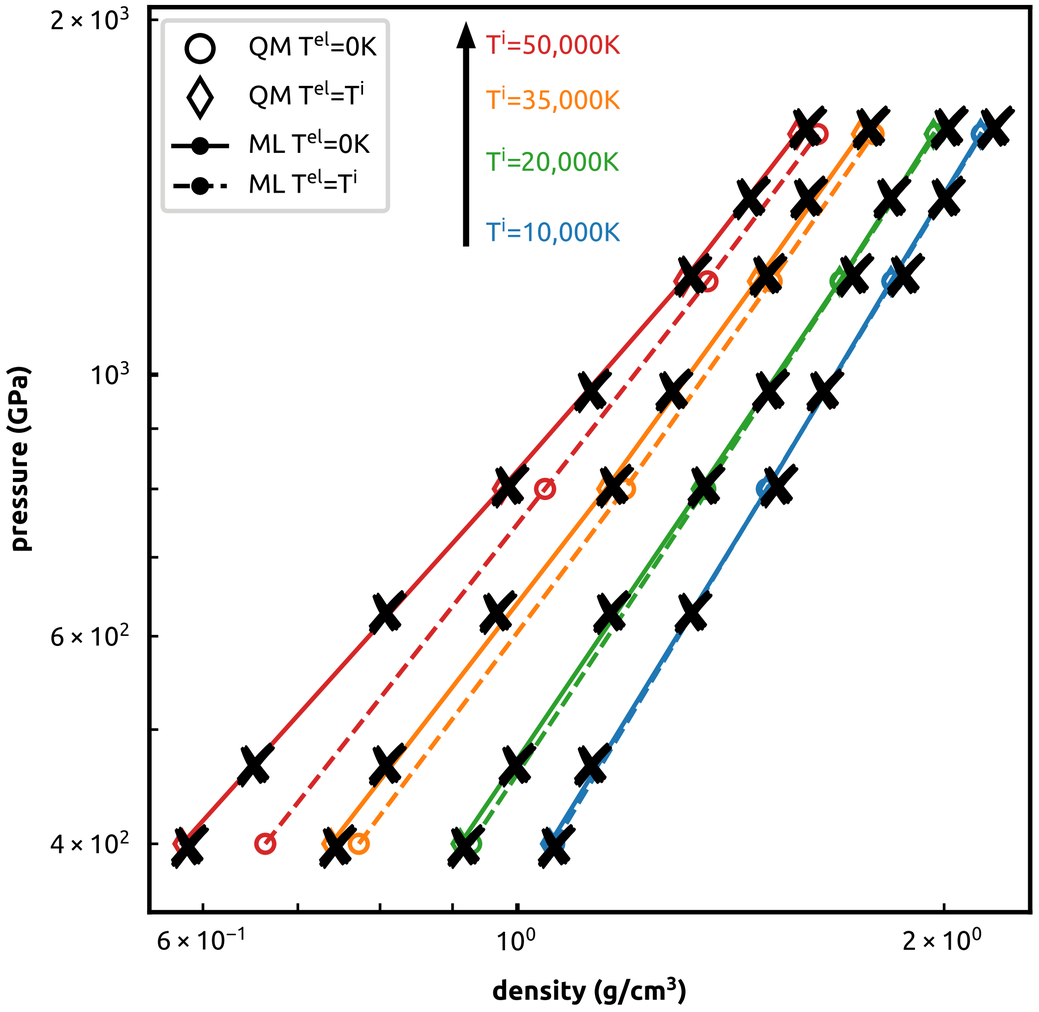


Phase diagram of hydrogen computed with DFT

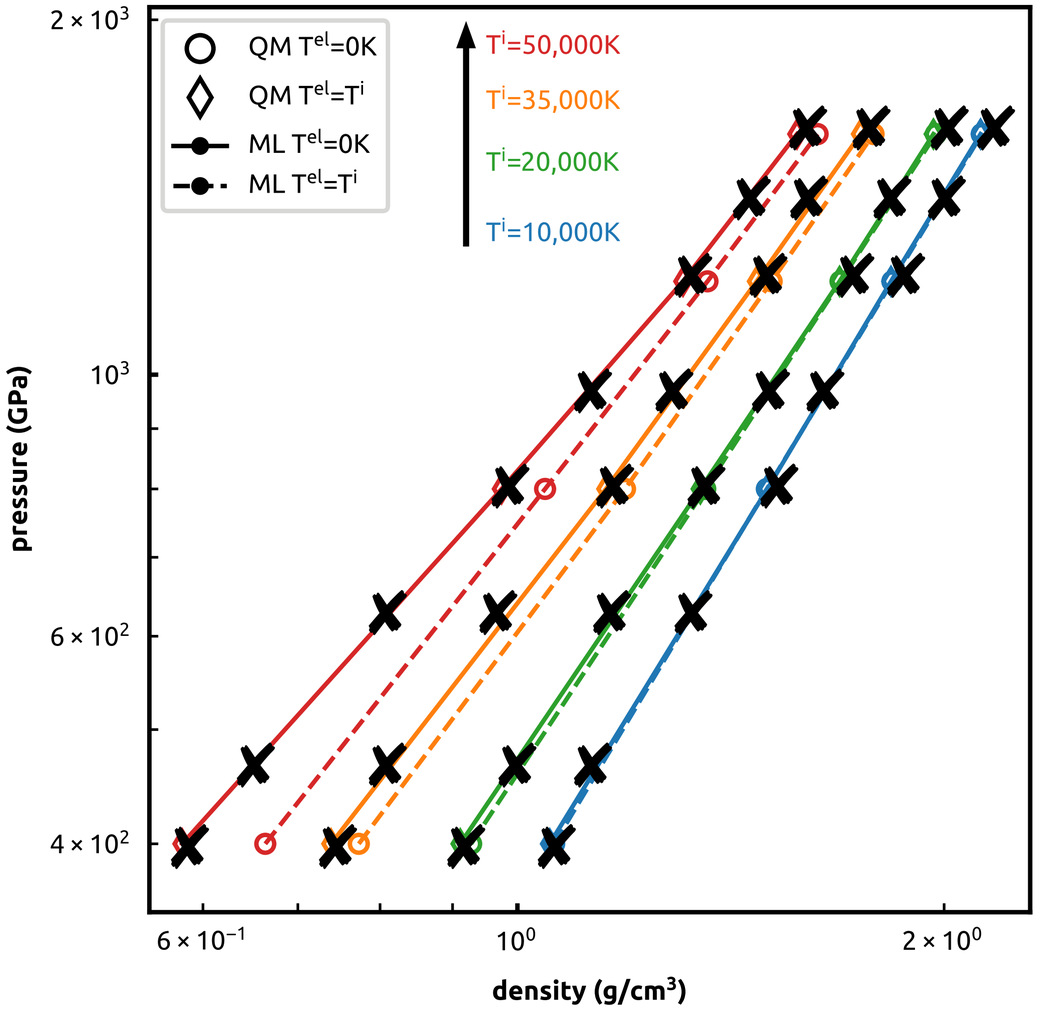
Naive solution



Naive solution

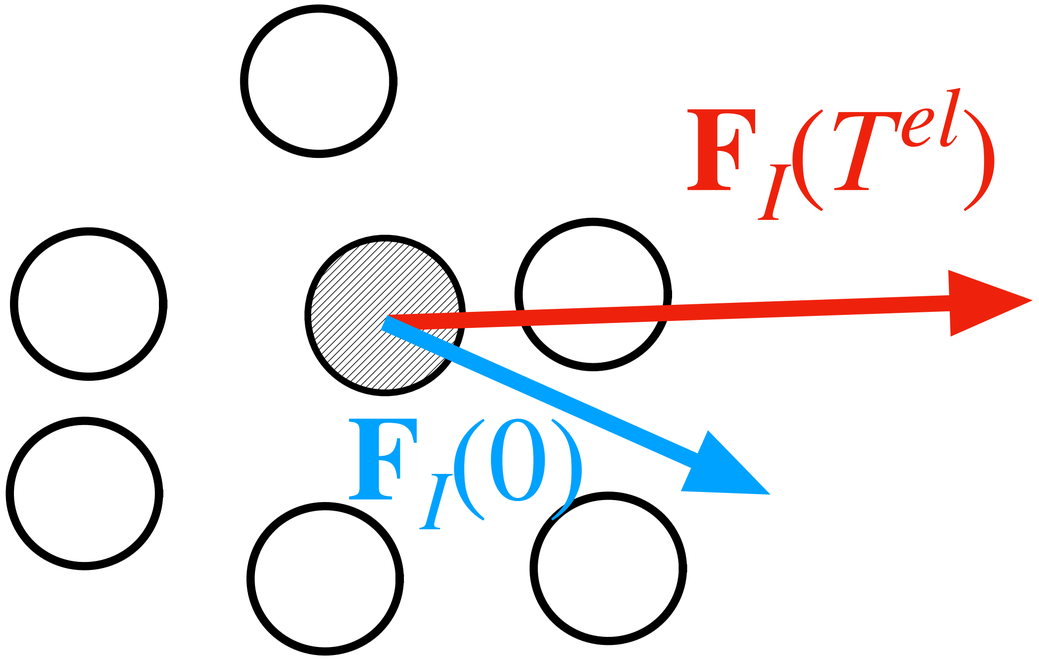


Naive solution



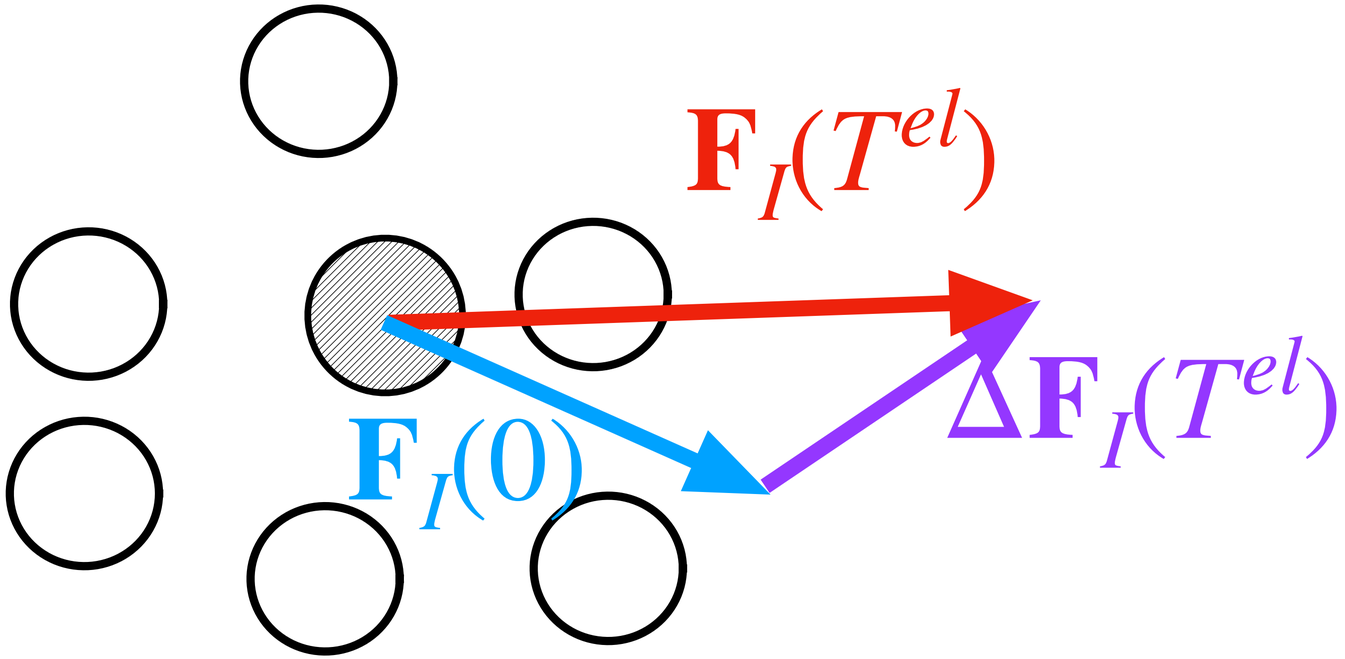
New ML potential for every temperature (4 in total)

Sketch of the idea



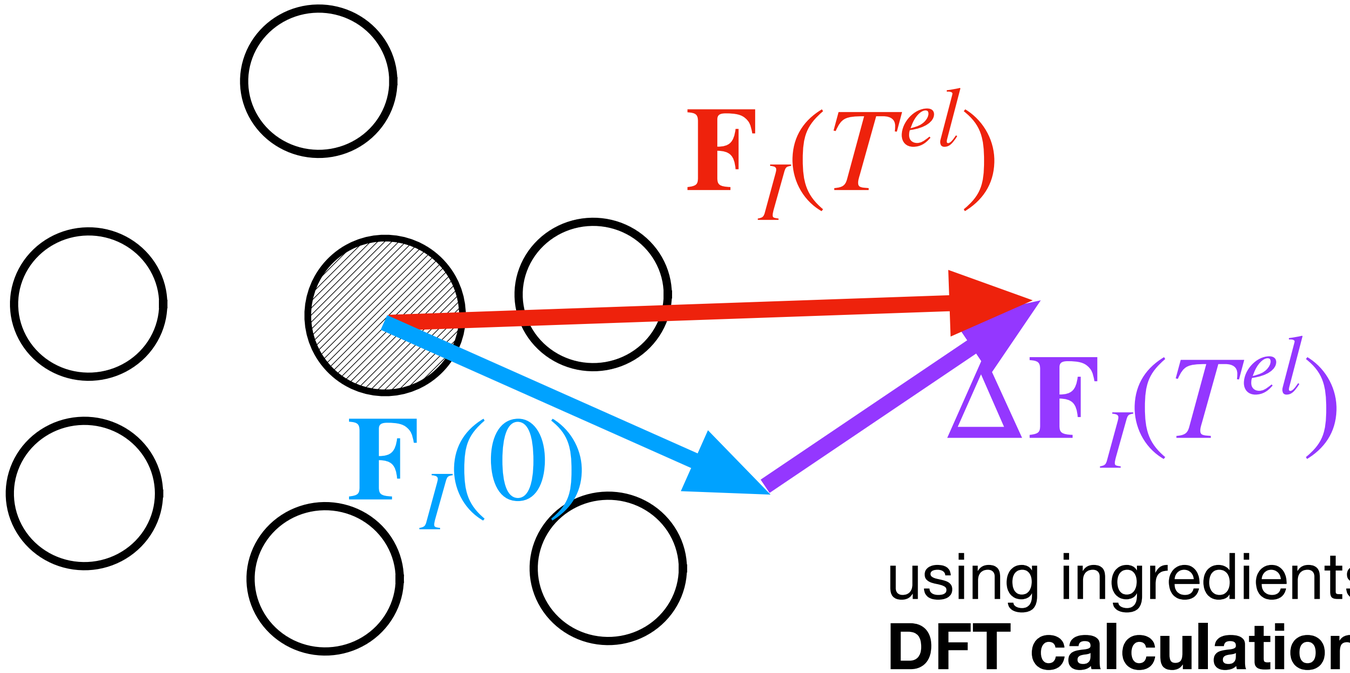
CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Sketch of the idea



CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Sketch of the idea



using ingredients from **the same DFT calculation** and extrapolating to **ANY** electronic temperature

CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Physics-motivated framework

Physics-motivated framework

- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$

[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

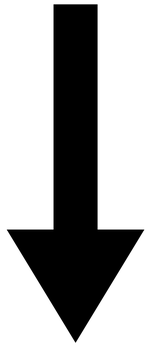
Physics-motivated framework

- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(0) + \Delta E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$

[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

Physics-motivated framework

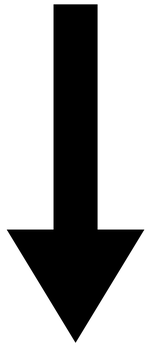
- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(0) + \Delta E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$



[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

Physics-motivated framework

- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(0) + \Delta E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$

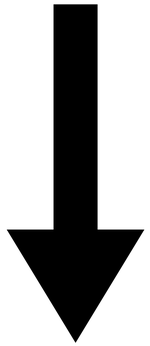


- Approximation: $A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) + \mathcal{O}(f^{2})$

[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

Physics-motivated framework

- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(0) + \Delta E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$



- Approximation: $A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) + \mathcal{O}(f^2)$

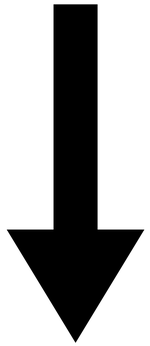
$$\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) = \int \mathrm{d}\epsilon \ \mathrm{DOS}^{0}(\epsilon) \ \epsilon \ [f(\mu, T^{\mathrm{el}}) - f(\mu_0, T = 0)]$$

$$S^{0}(T^{\mathrm{el}}) = -k_B \int \mathrm{d}\epsilon \ \mathrm{DOS}^{0}(\epsilon) \ \left[f \log(f) - (1-f) \log(1-f) \right]$$

[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

Physics-motivated framework

- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(0) + \Delta E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$



- Approximation: $A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) + \mathcal{O}(f^2)$

$$\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) = \int \mathrm{d}\epsilon \ \mathrm{DOS}^{0}(\epsilon) \ \epsilon \ [f(\mu, T^{\mathrm{el}}) - f(\mu_0, T = 0)]$$

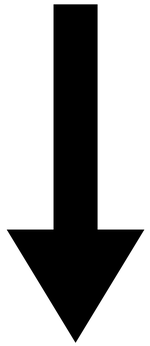
$$S^{0}(T^{\mathrm{el}}) = -k_B \int \mathrm{d}\epsilon \ \mathrm{DOS}^{0}(\epsilon) \ \left[f \log(f) - (1-f) \log(1-f) \right]$$

The temperature T^{el} is an external parameter

[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

Physics-motivated framework

- Ensemble DFT[1, 2]: $A(T^{\mathrm{el}}) = E(0) + \Delta E(T^{\mathrm{el}}) - T^{\mathrm{el}} S_{KS}(T^{\mathrm{el}})$



Fixed DOS approximation

- Approximation: $A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) + \mathcal{O}(f^2)$

$$\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) = \int \mathrm{d}\epsilon \ \mathrm{DOS}^{0}(\epsilon) \ \epsilon \ [f(\mu, T^{\mathrm{el}}) - f(\mu_0, T = 0)]$$

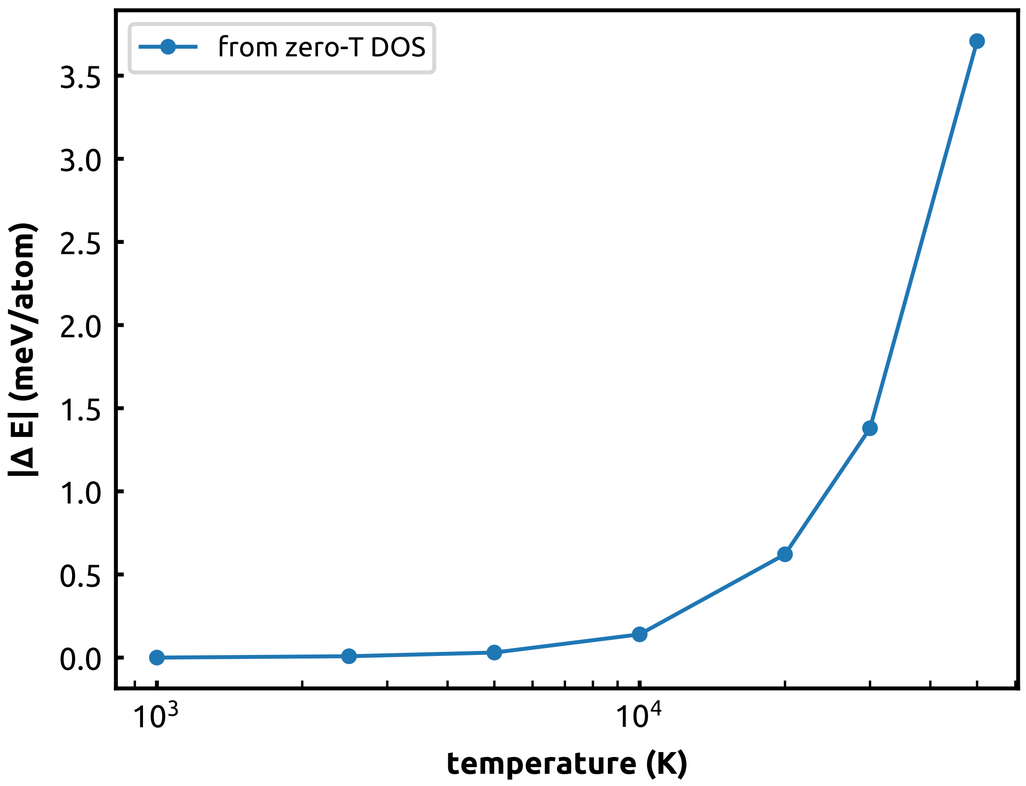
$$S^{0}(T^{\mathrm{el}}) = -k_B \int \mathrm{d}\epsilon \ \mathrm{DOS}^{0}(\epsilon) \ \left[f \log(f) - (1-f) \log(1-f) \right]$$

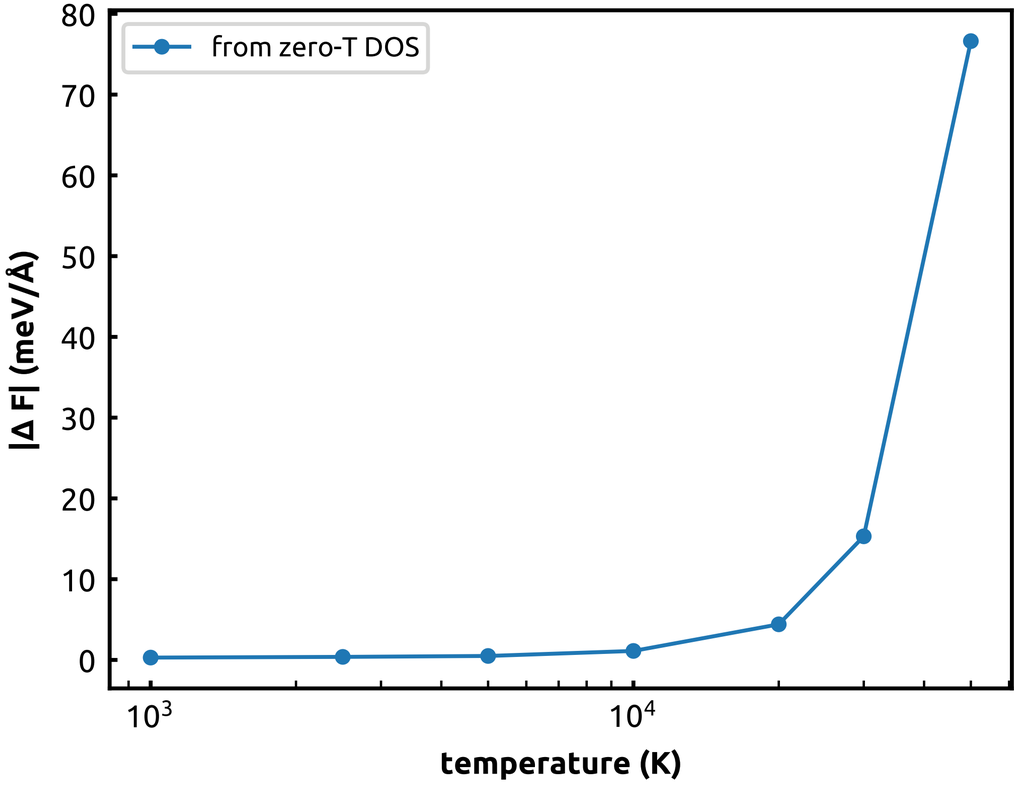
The temperature T^{el} is an external parameter

[1]: Mermin, 1965
[2]: Marzari and Vanderbilt, 1997

Systematic errors of the approximation

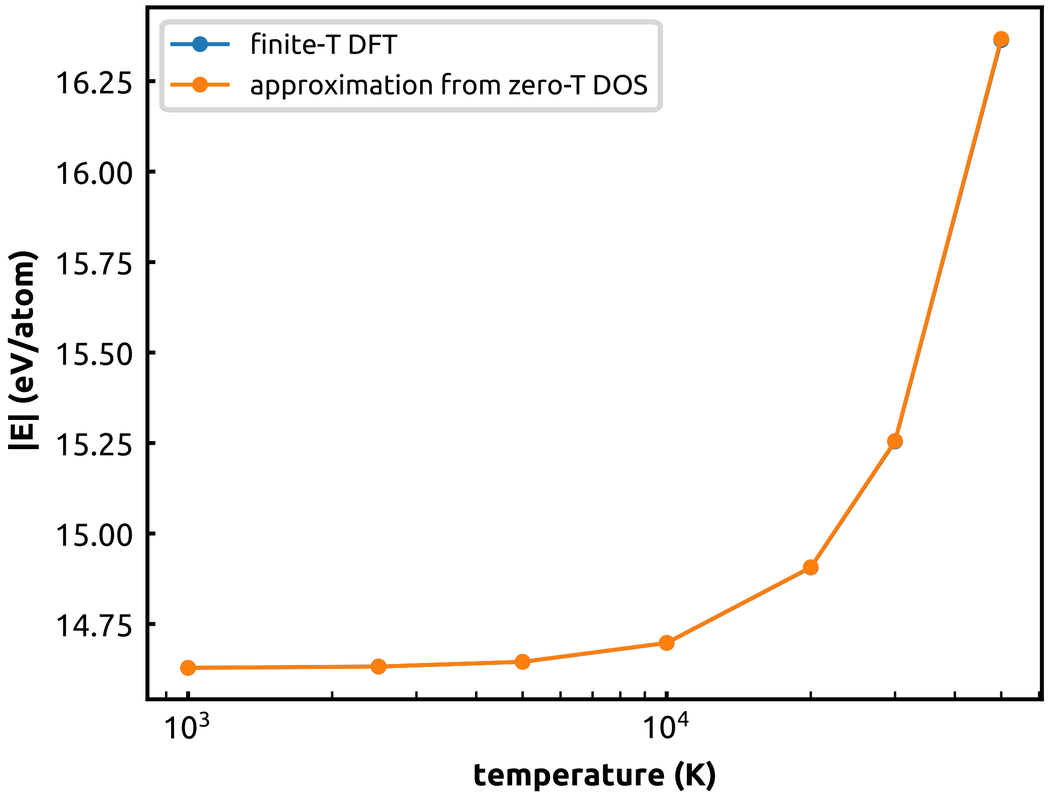
Approximation error in the total free energy and a single force component for a liquid hydrogen structure

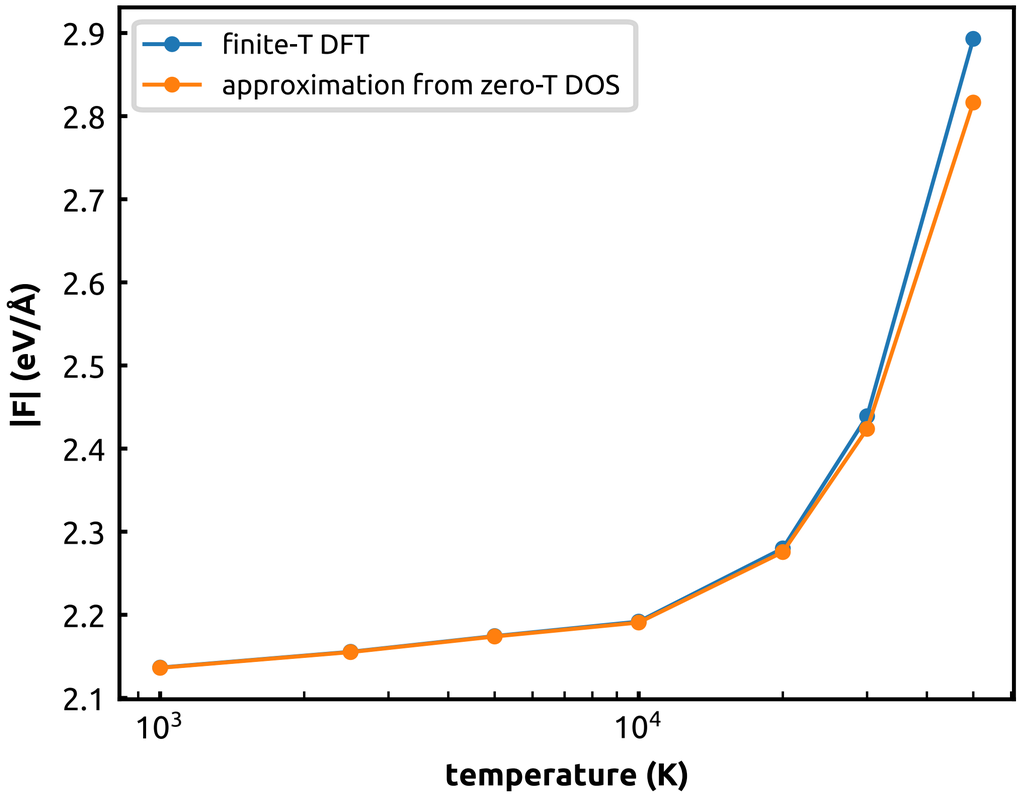




Systematic errors of the approximation

Approximation error in the total free energy and a single force component for a liquid hydrogen structure





Application in ML workflows

energy $A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}})$

CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Application in ML workflows

energy $$A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}})$$

force $$\mathbf{F}_I(T^{\mathrm{el}}) \approx \mathbf{F}_I(0) - \nabla_I \left(\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) \right)$$

stress $$\nabla_\eta A(T^{\mathrm{el}}) \approx \nabla_\eta E(0) + \nabla_\eta \left(\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) \right)$$

CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Application in ML workflows

ANY ML interatomic potentials (GAP, DeepMD, BPNN, (M/Gr)ACE,…)

energy $$A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}})$$

force $$\mathbf{F}_I(T^{\mathrm{el}}) \approx \mathbf{F}_I(0) - \nabla_I \left(\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) \right)$$

stress $$\nabla_\eta A(T^{\mathrm{el}}) \approx \nabla_\eta E(0) + \nabla_\eta \left(\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) \right)$$

CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Application in ML workflows

ANY ML interatomic potentials (GAP, DeepMD, BPNN, (M/Gr)ACE,…)

ML DOS models

energy $A(T^{\mathrm{el}}) \approx E(0) + \Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}})$

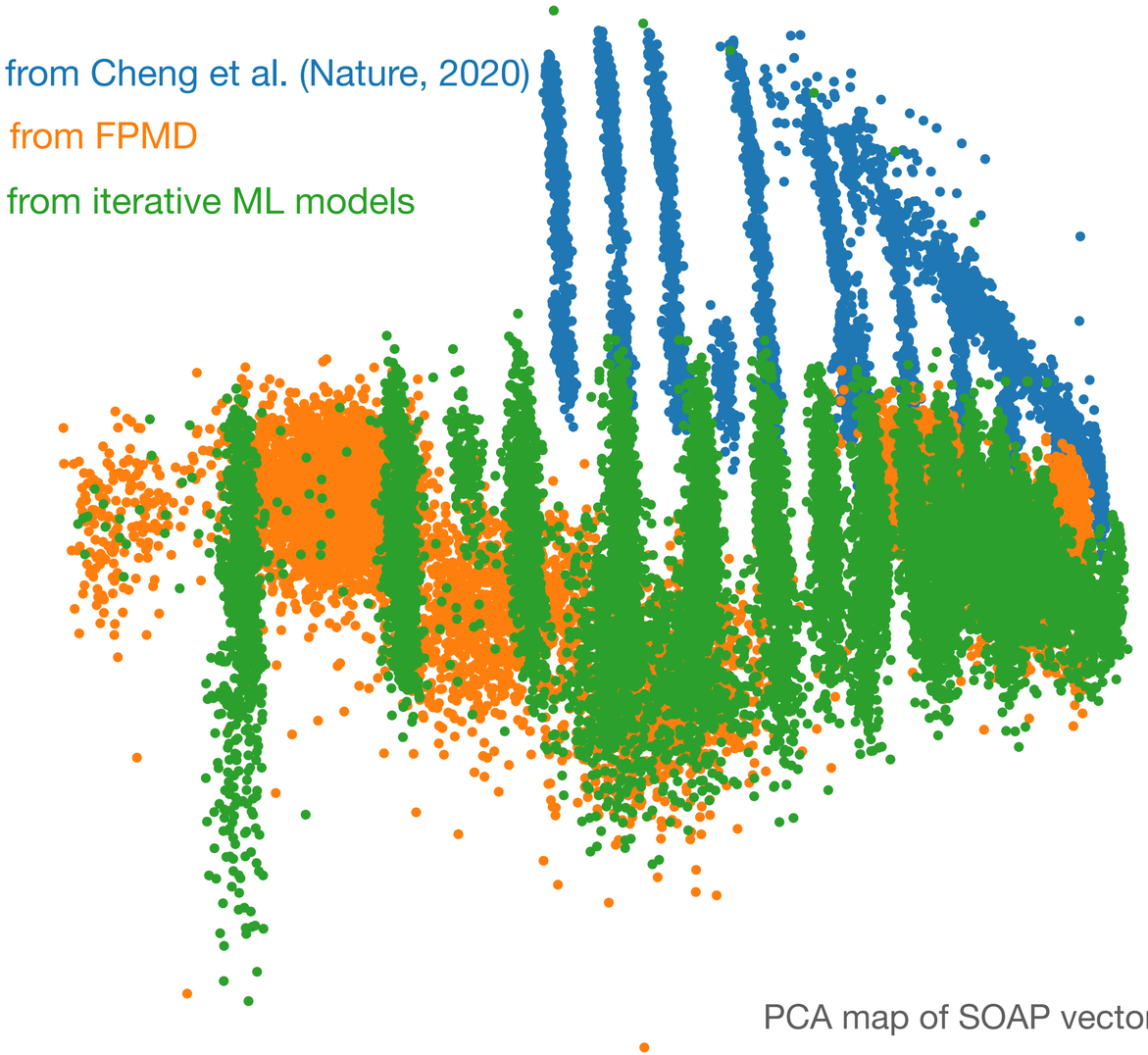
force $\mathbf{F}_I(T^{\mathrm{el}}) \approx \mathbf{F}_I(0) - \nabla_I \left(\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) \right)$

stress $\nabla_\eta A(T^{\mathrm{el}}) \approx \nabla_\eta E(0) + \nabla_\eta \left(\Delta E^{0}_{\mathrm{band}}(T^{\mathrm{el}}) - T^{\mathrm{el}} S^{0}(T^{\mathrm{el}}) \right)$

CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Application to phase diagram of hydrogen

- training set size: ~28,500 structures
- 128 atoms per structure
- DFT calculations on the training set targeting the ground state **ONLY**

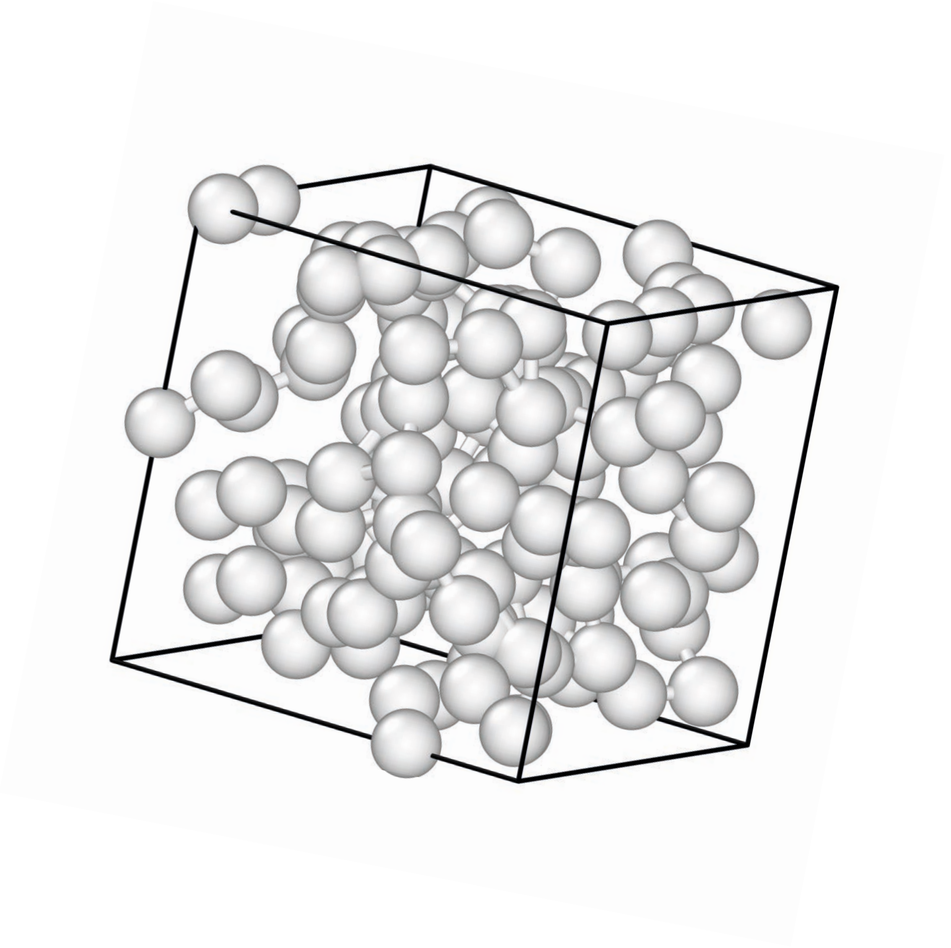


PCA map of SOAP vectors

CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Cheng et al., 2020

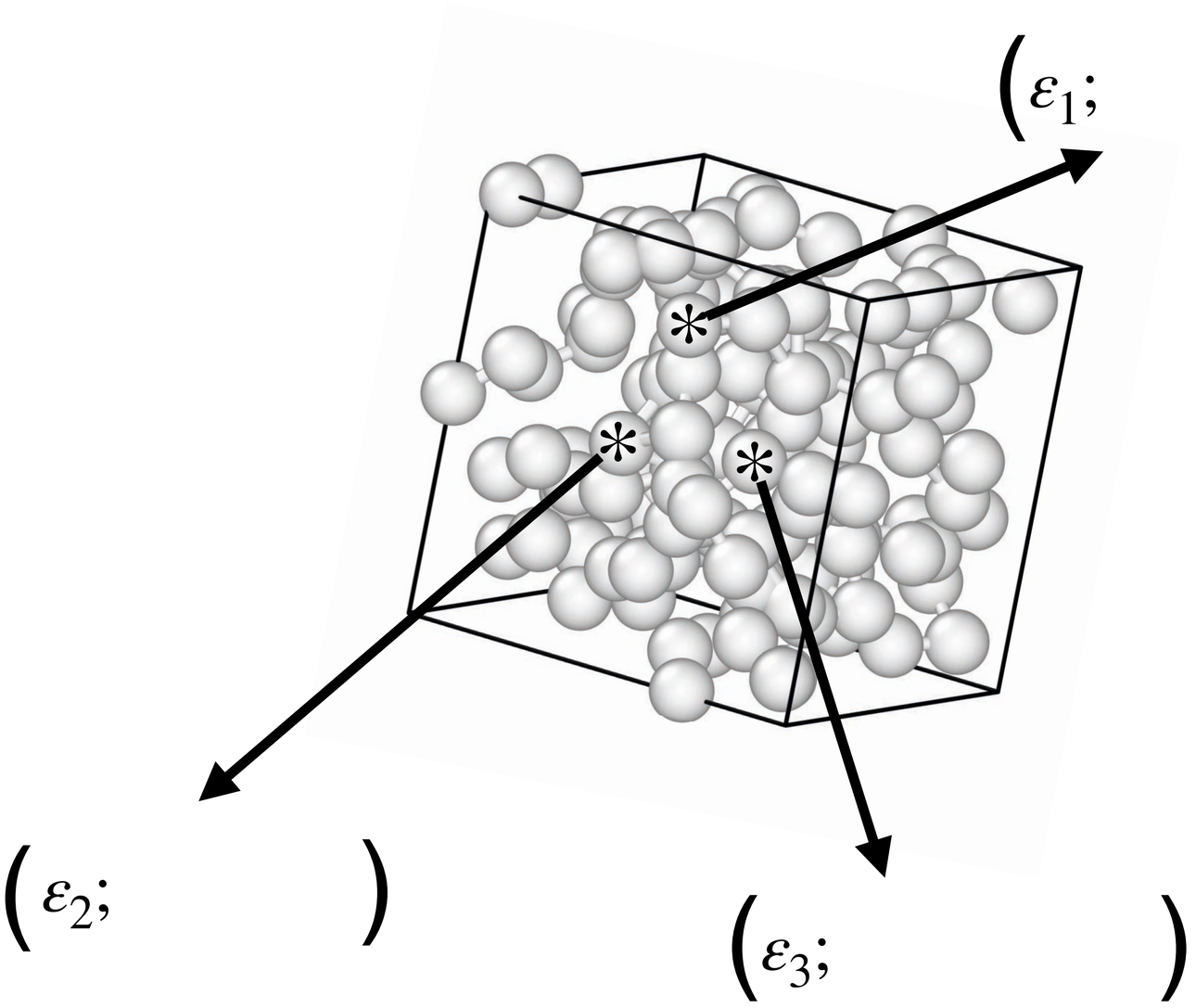
ML models for the energy contributions



ML models for the energy contributions

SOAP based model

- $E(0) = E_{2B} + E_{MB}$



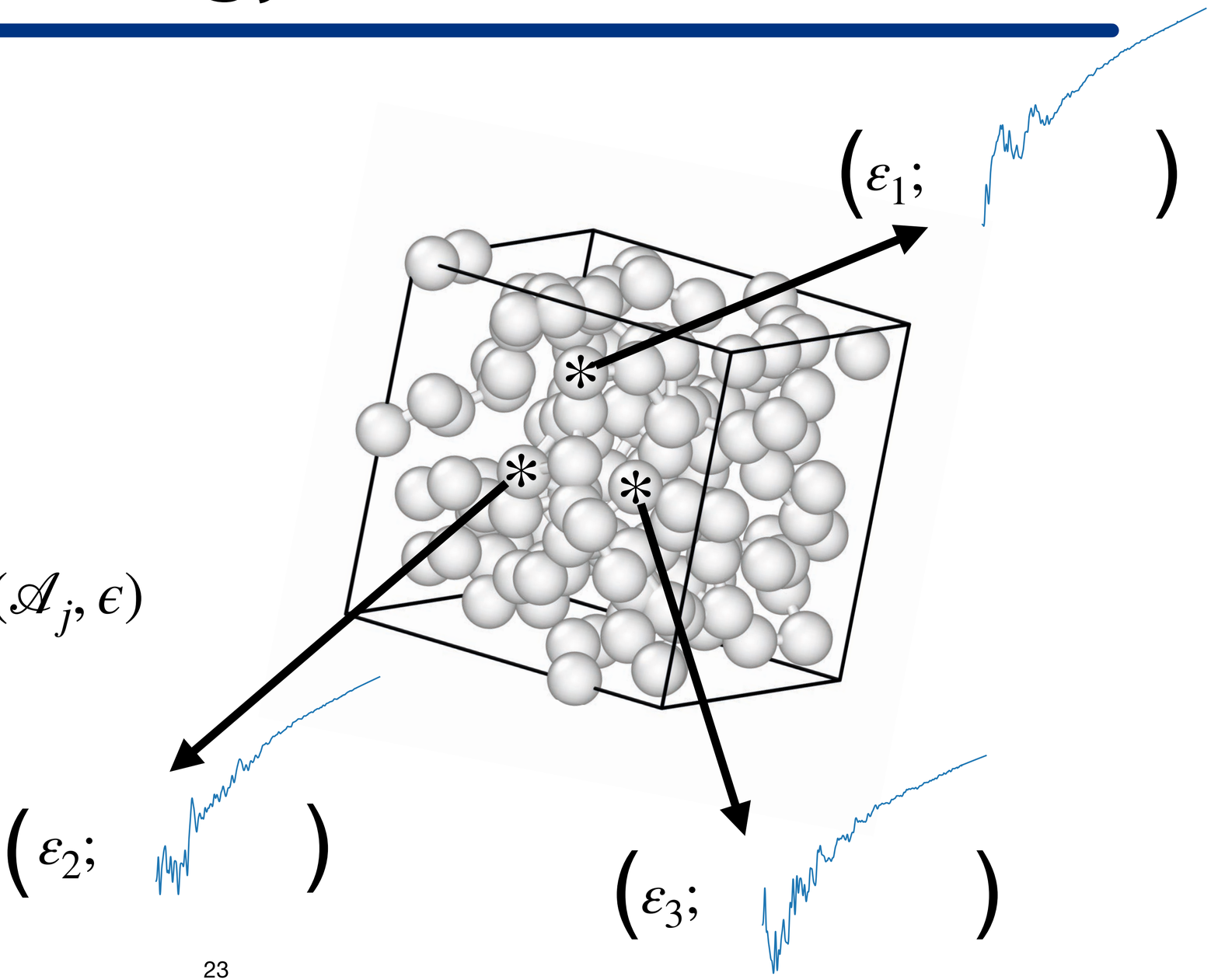
Bartók et al., 2010

ML models for the energy contributions

SOAP based model

- $E(0) = E_{2B} + E_{MB}$

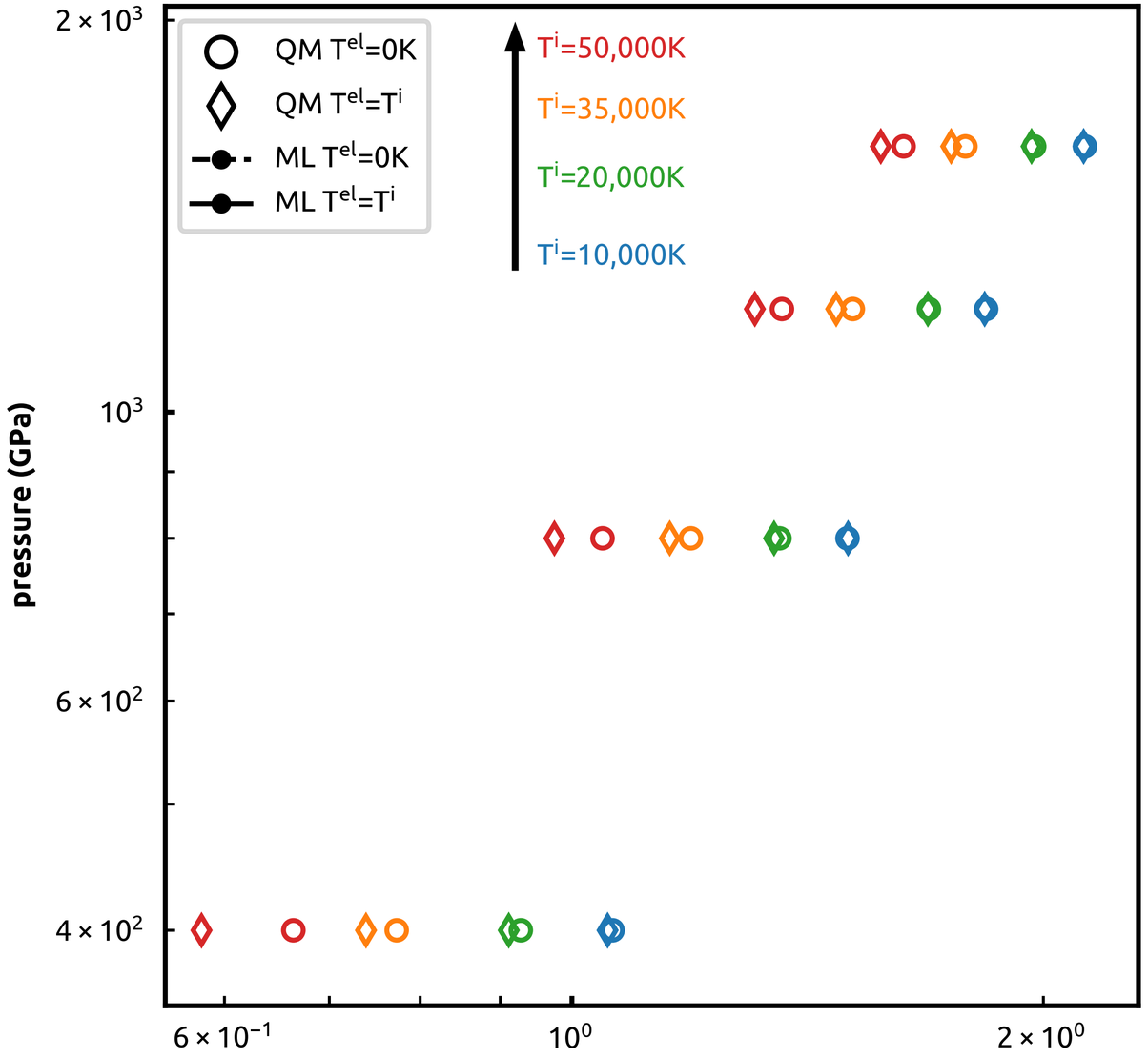
- $\mathrm{DOS}(\mathscr{A}, \epsilon) = \sum_{\mathscr{A}_j \in \mathscr{A}} \mathrm{LDOS}(\mathscr{A}_j, \epsilon)$



Bartók et al., 2010

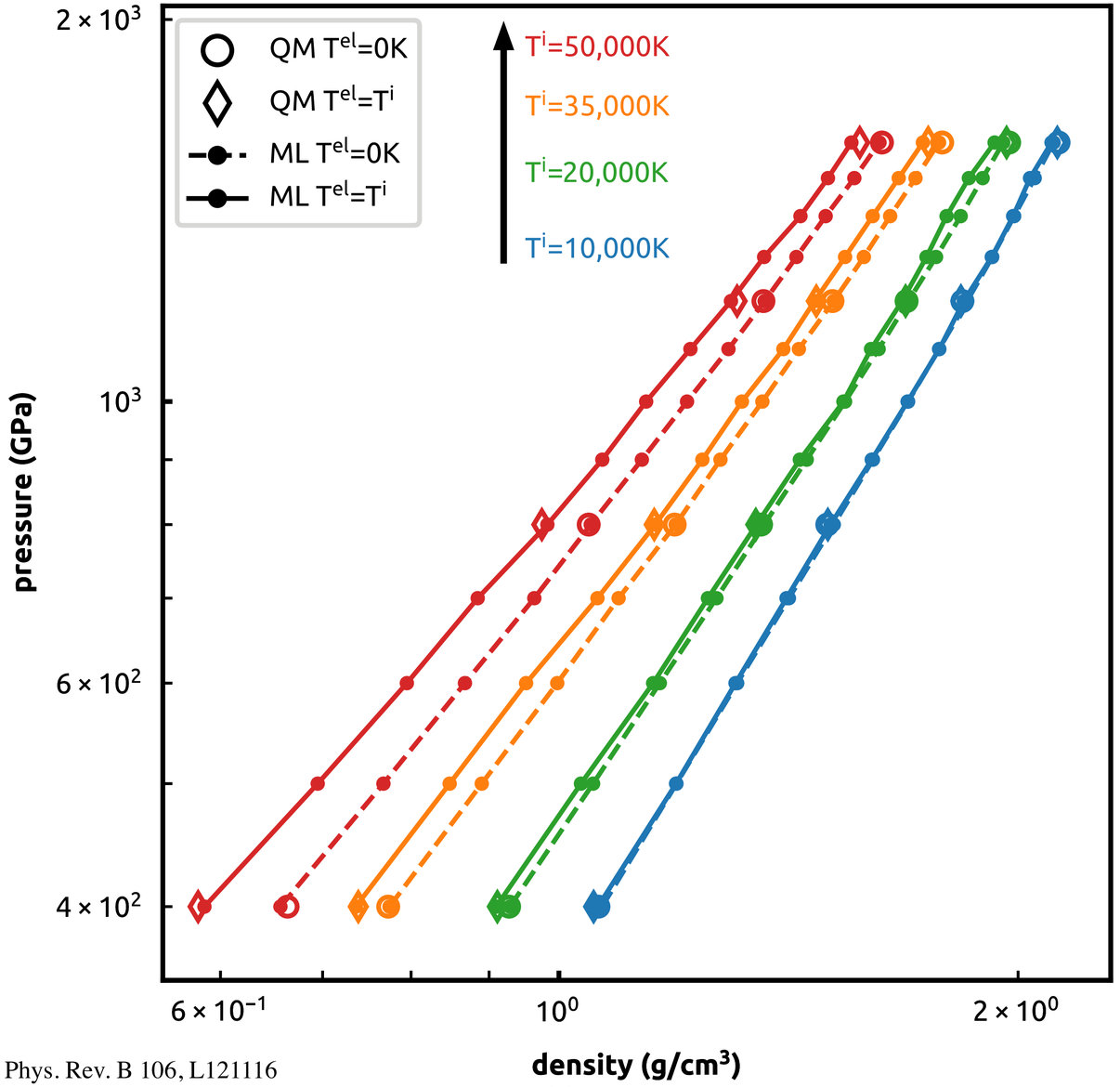
CBM., Anelli, A., Csányi, G., Ceriotti, M., 2020.
Learning the electronic density of states in condensed matter. Phys. Rev. B 102, 235130.

Hydrogen EOS



CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Hydrogen EOS



CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Heat capacity of high pressure hydrogen

- Using finite differences:

$$C_p = \frac{\langle H(T + \Delta T)\rangle_{Np\ T+\Delta T} - \langle H(T)\rangle_{Np\ T}}{\Delta T}$$

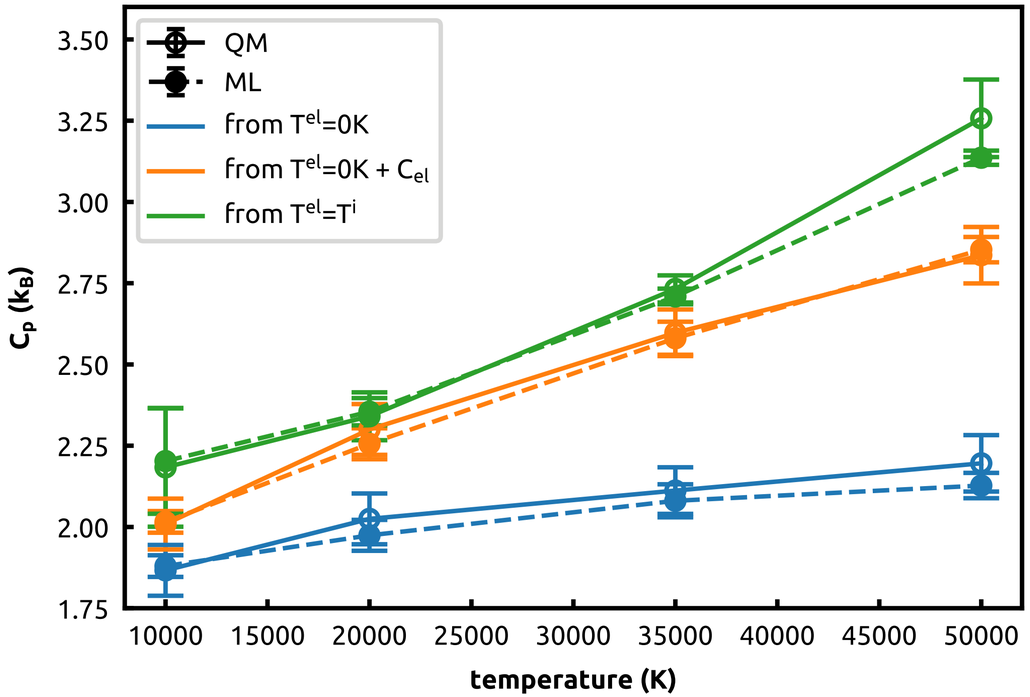
CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Heat capacity of high pressure hydrogen

- Using finite differences:

$$C_p = \frac{\langle H(T+\Delta T)\rangle_{Np\ T+\Delta T} - \langle H(T)\rangle_{Np\ T}}{\Delta T}$$

Constant pressure heat capacity of hydrogen at 400GPa



CBM, Grasselli, F., Ceriotti, M., 2022.
Predicting hot-electron free energies from ground-state data. Phys. Rev. B 106, L121116

Takehome message

Takehome message

- Atom-centred DOS models are versatile and transferable

Takehome message

- Atom-centred DOS models are versatile and transferable
- ML DOS models can be used in physics-inspired modelling of materials:

Takehome message

- Atom-centred DOS models are versatile and transferable
- ML DOS models can be used in physics-inspired modelling of materials:
  - characterise local electronic fingerprints

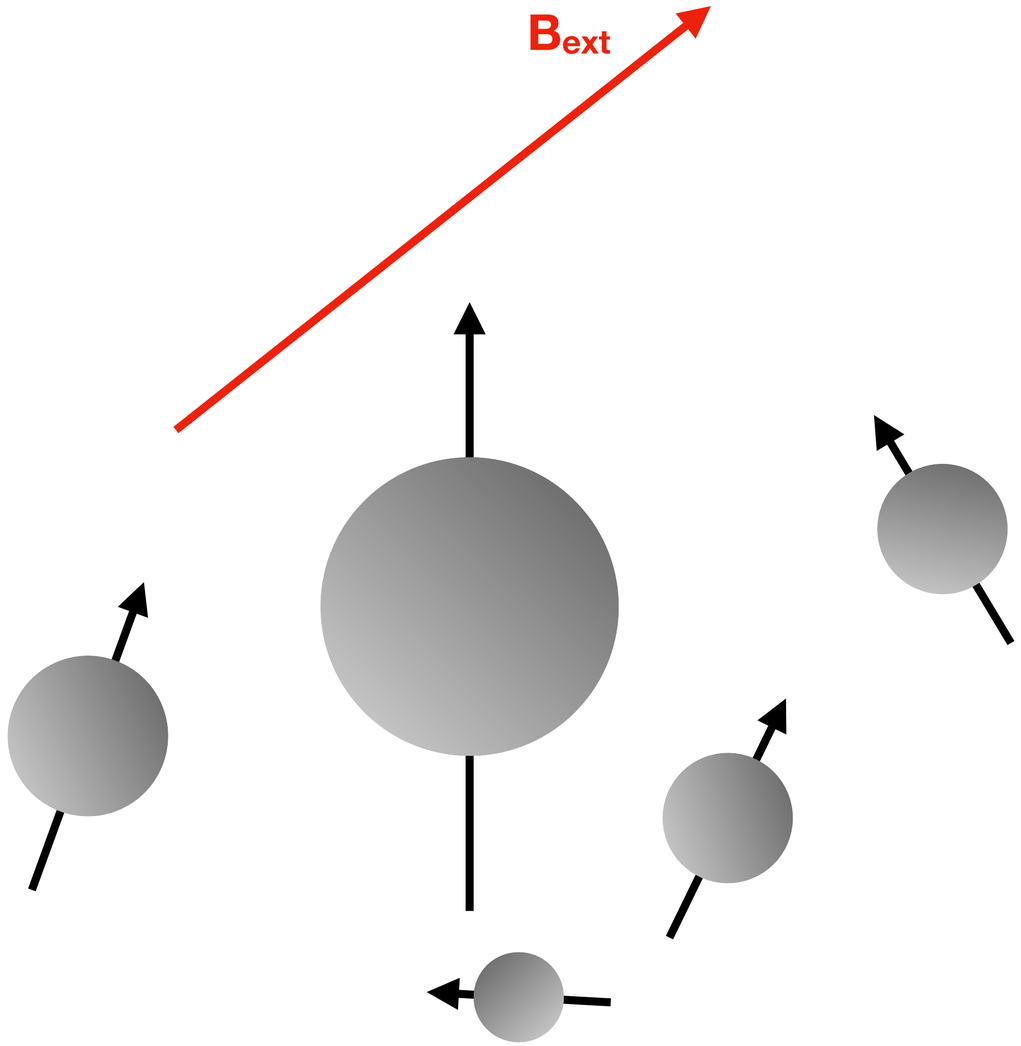
Takehome message

- Atom-centred DOS models are versatile and transferable
- ML DOS models can be used in physics-inspired modelling of materials:
  - characterise local electronic fingerprints
  - incorporate electron-finite-temperature effects

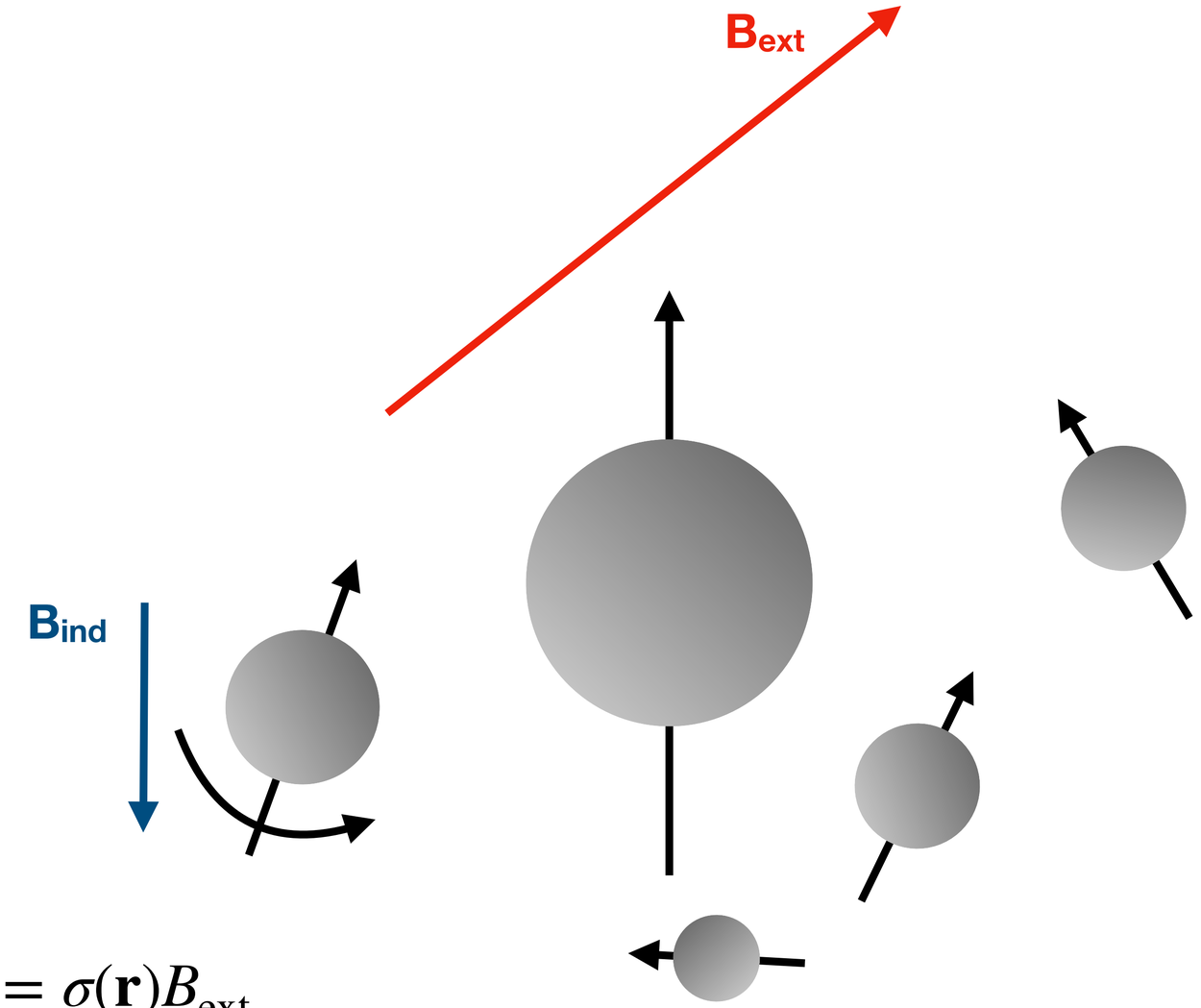
Property 2:

solid-state nuclear magnetic resonance parameters

(Quick) Introduction to NMR

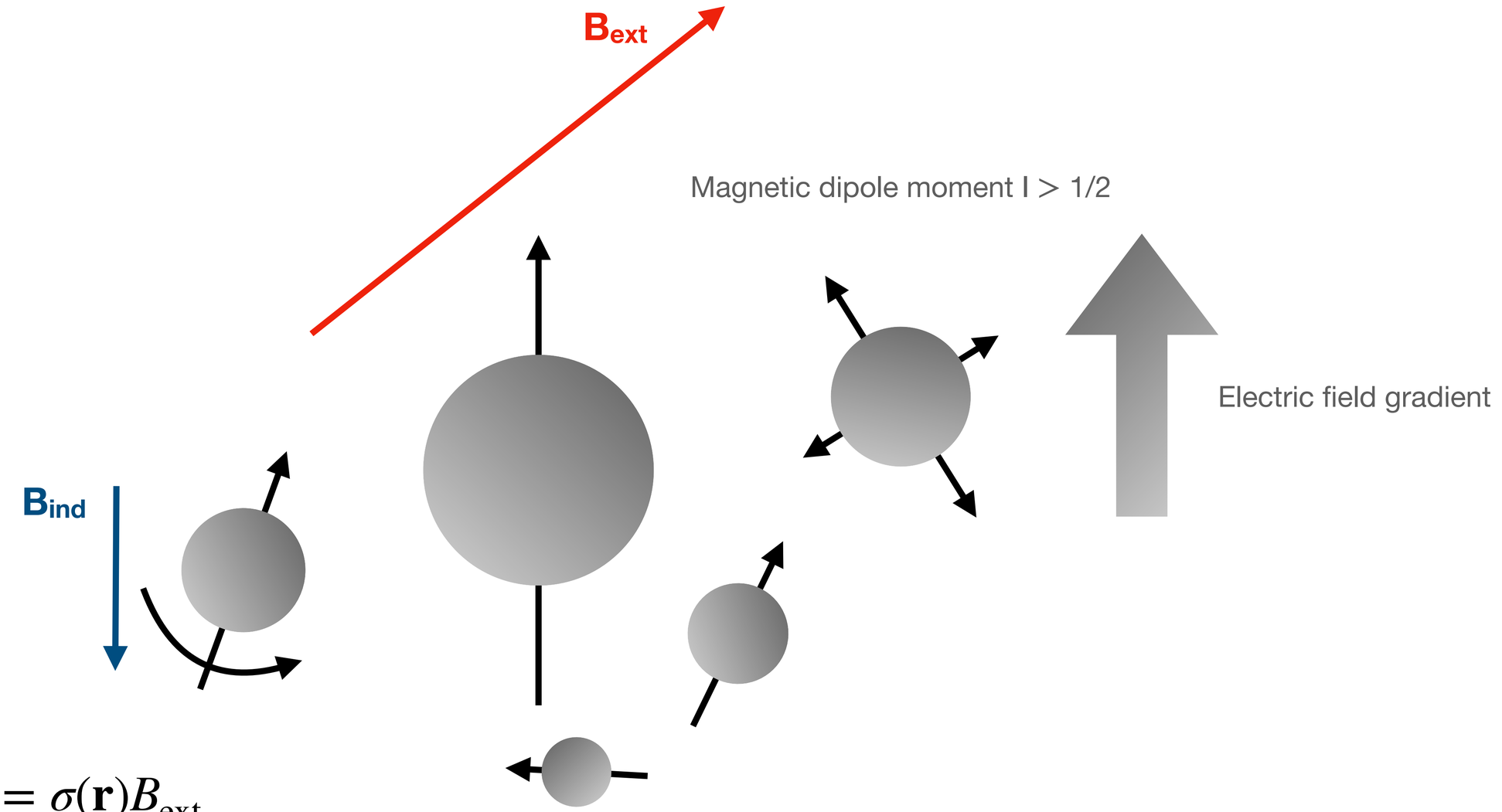


(Quick) Introduction to NMR



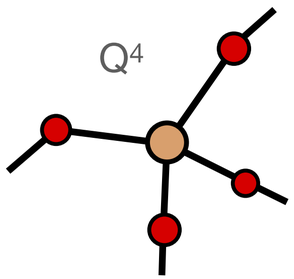
$$B_{\text{ind}}(\mathbf{r}) = \sigma(\mathbf{r})B_{\text{ext}}$$

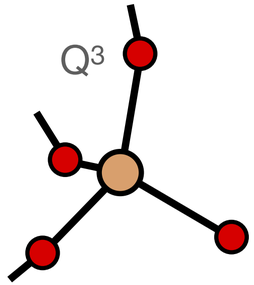
(Quick) Introduction to NMR

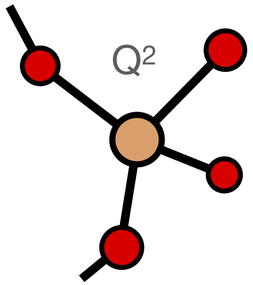


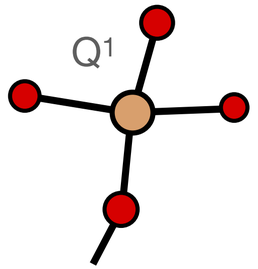
$$B_{\text{ind}}(\mathbf{r}) = \sigma(\mathbf{r})B_{\text{ext}}$$

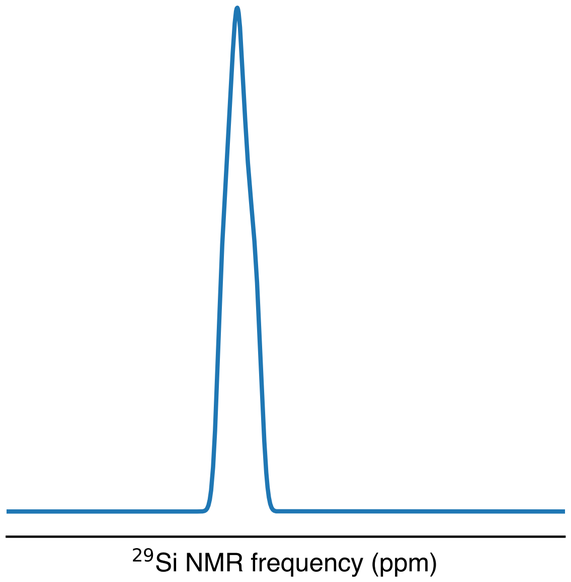
NMR parameters as local fingerprints

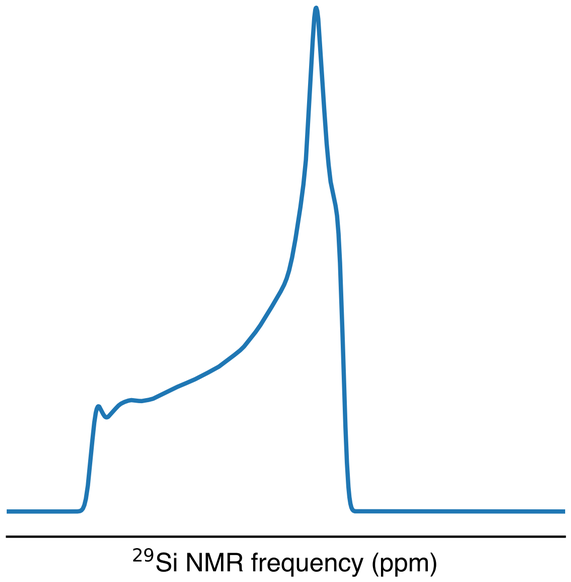


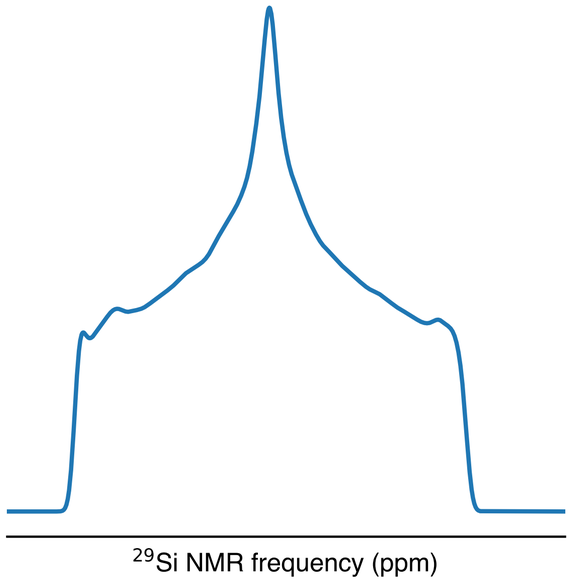


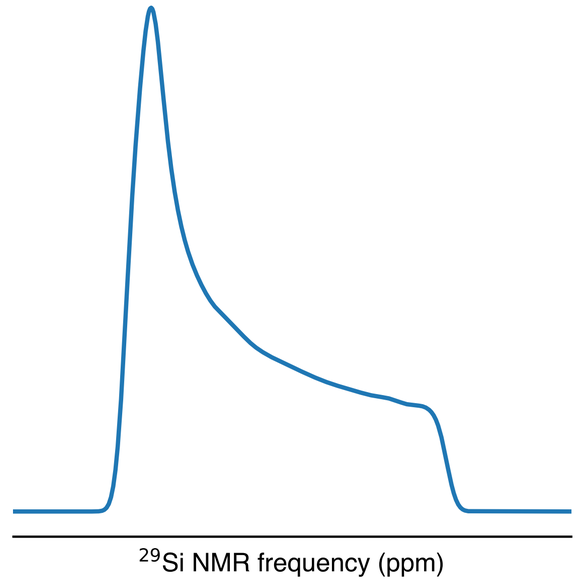








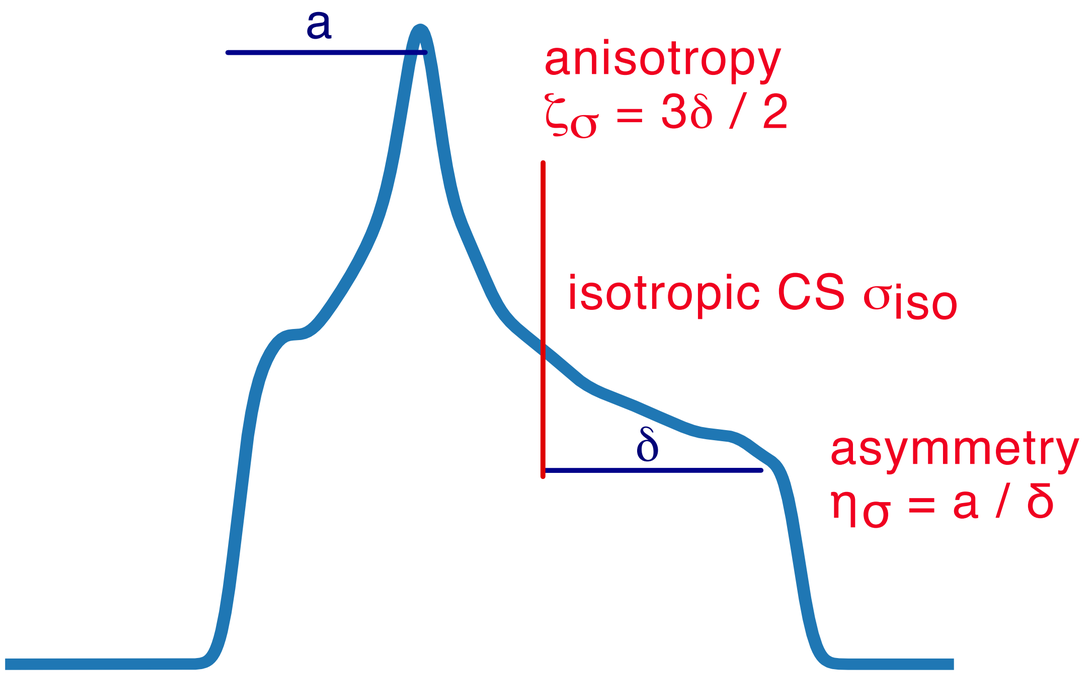




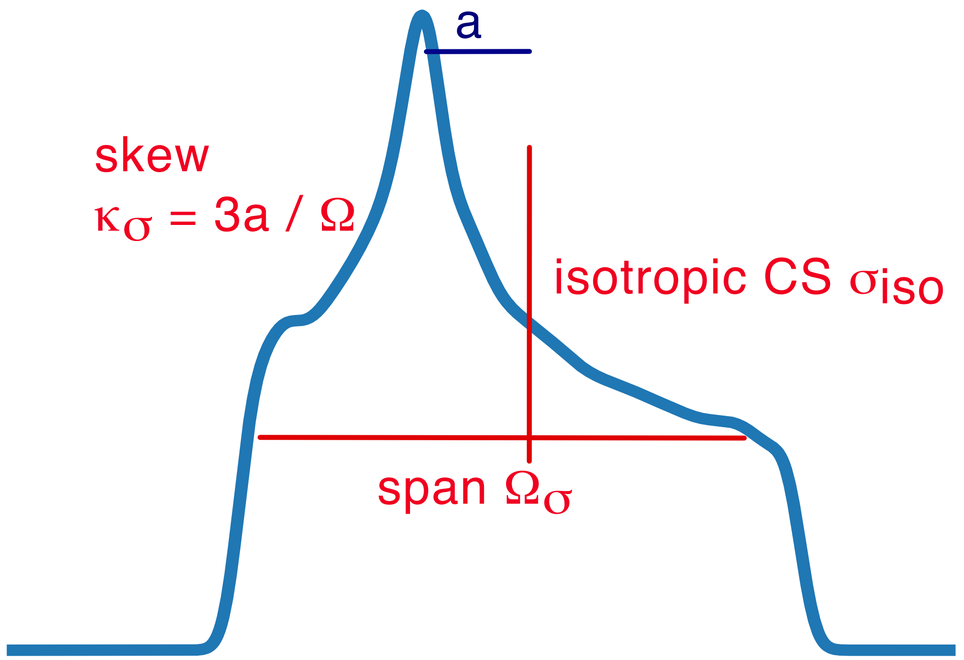
Objectives

- Build a **transferable ML model** for the NMR parameters
- Identify the **challenges** in modelling NMR parameters as tensors
- Apply the ML NMR models to study :
  - the **feature/property** relations
  - dynamics of solids

Conventions for NMR tensors



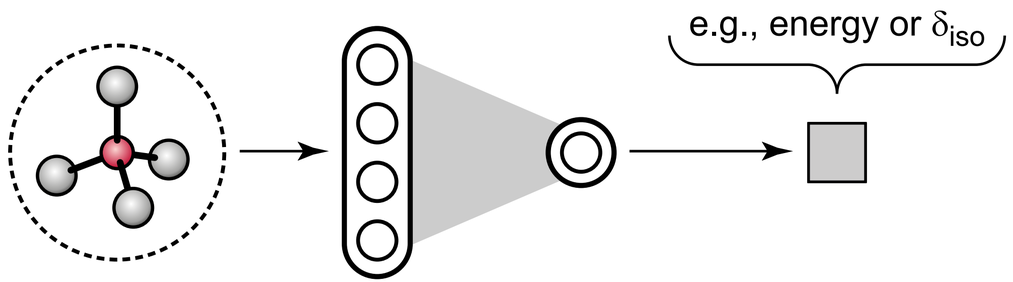
Haeberlen convention



Maryland convention

ML architecture

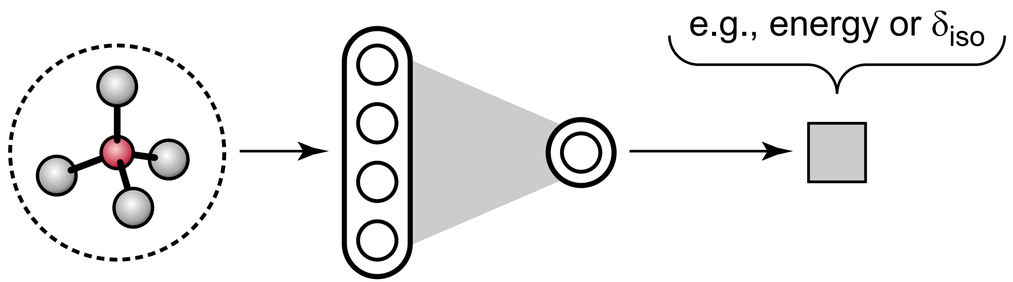
a Scalar atomistic ML



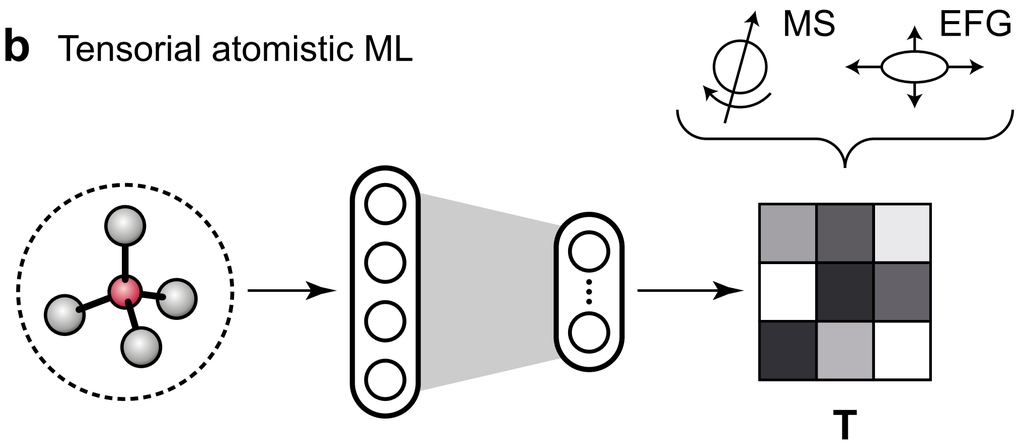
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

ML architecture

a Scalar atomistic ML



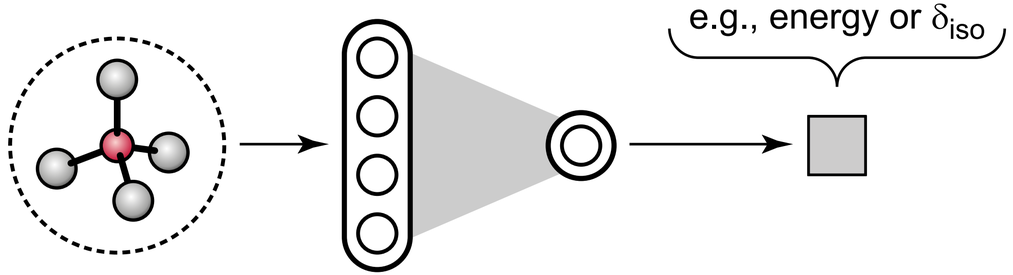
b Tensorial atomistic ML



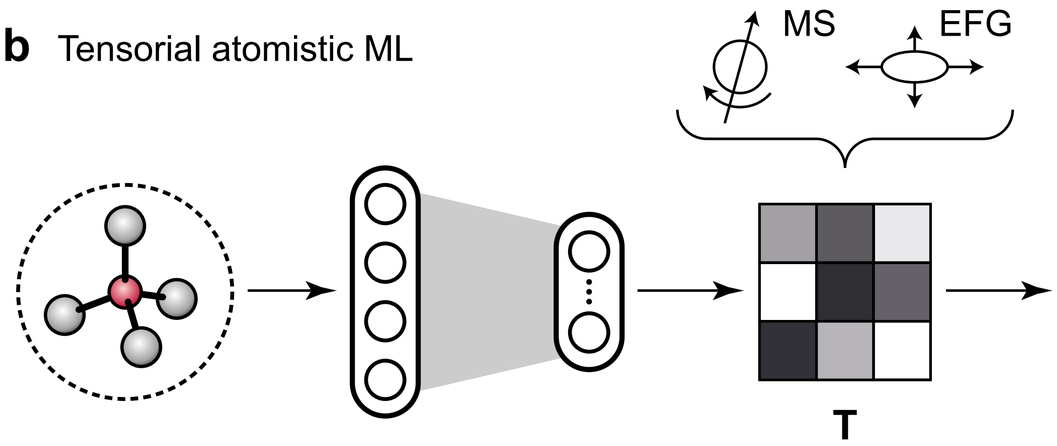
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

ML architecture

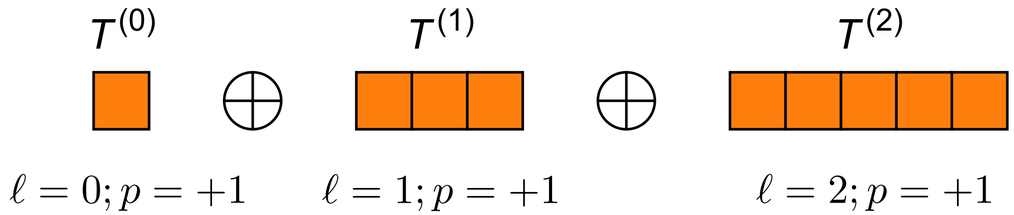
a Scalar atomistic ML



b Tensorial atomistic ML



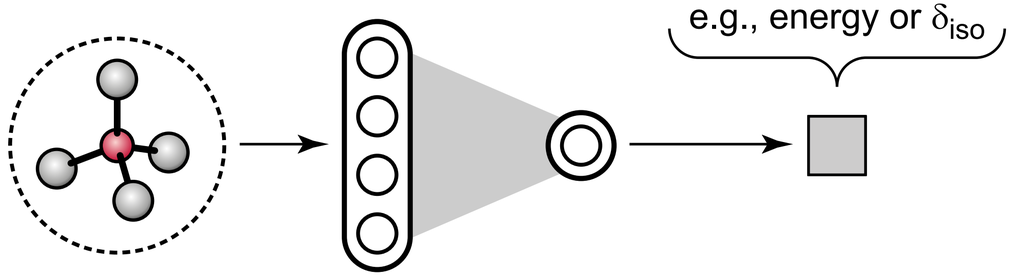
c **Option 1**: Irreducible spherical decomposition (ISD)



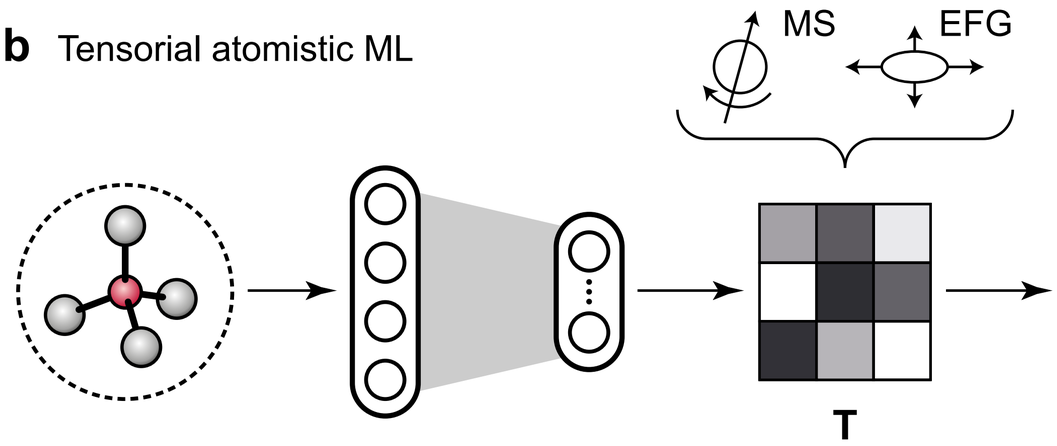
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

ML architecture

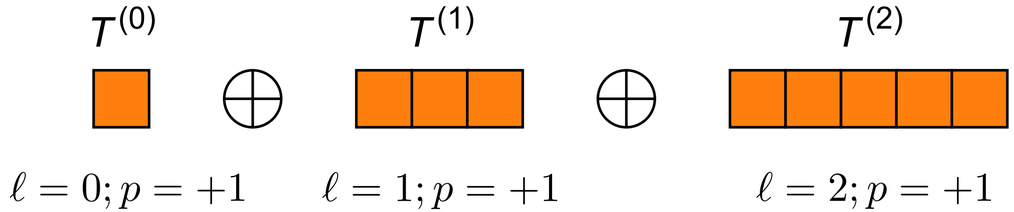
a Scalar atomistic ML



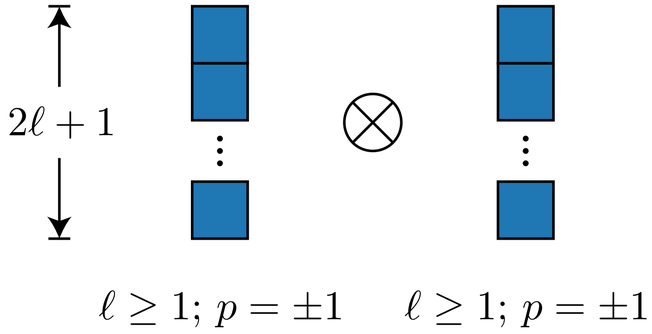
b Tensorial atomistic ML



c **Option 1**: Irreducible spherical decomposition (ISD)



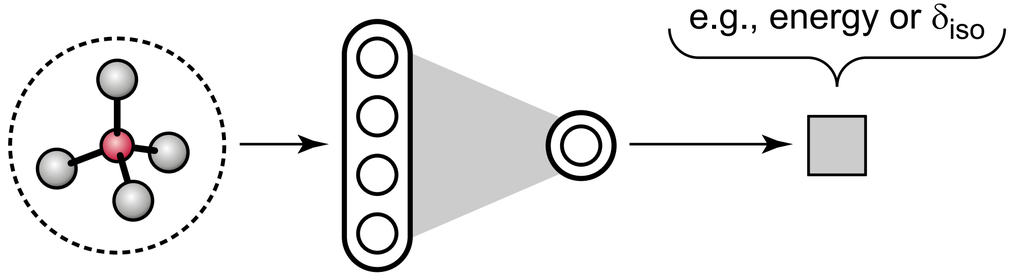
d **Option 2:** Tensor product (TP)



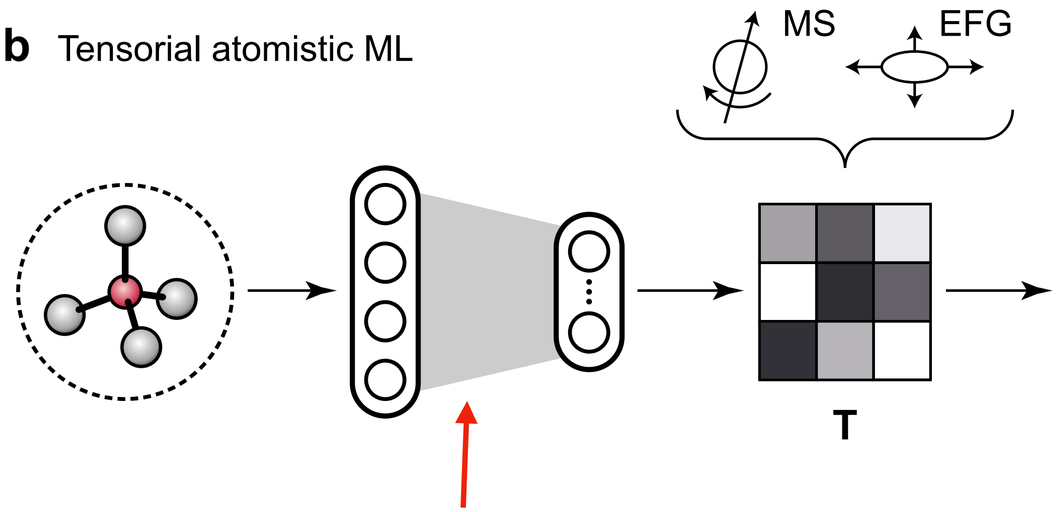
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

ML architecture

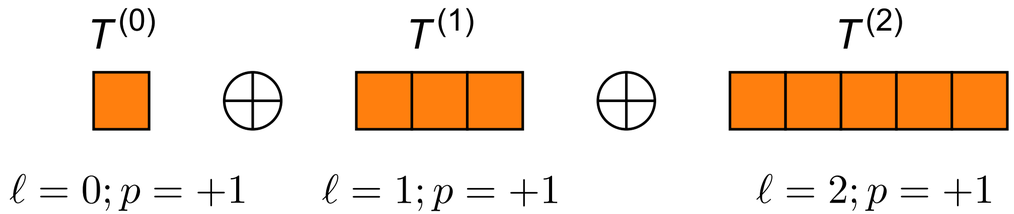
a Scalar atomistic ML



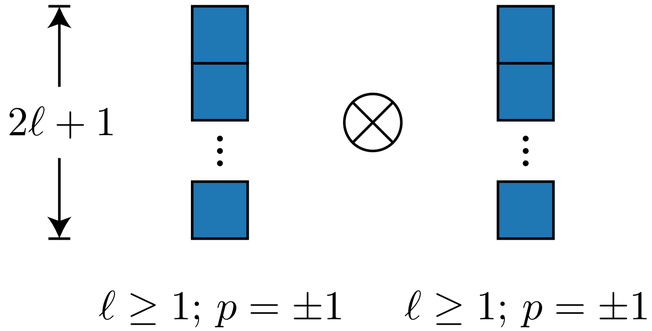
b Tensorial atomistic ML



c **Option 1**: Irreducible spherical decomposition (ISD)

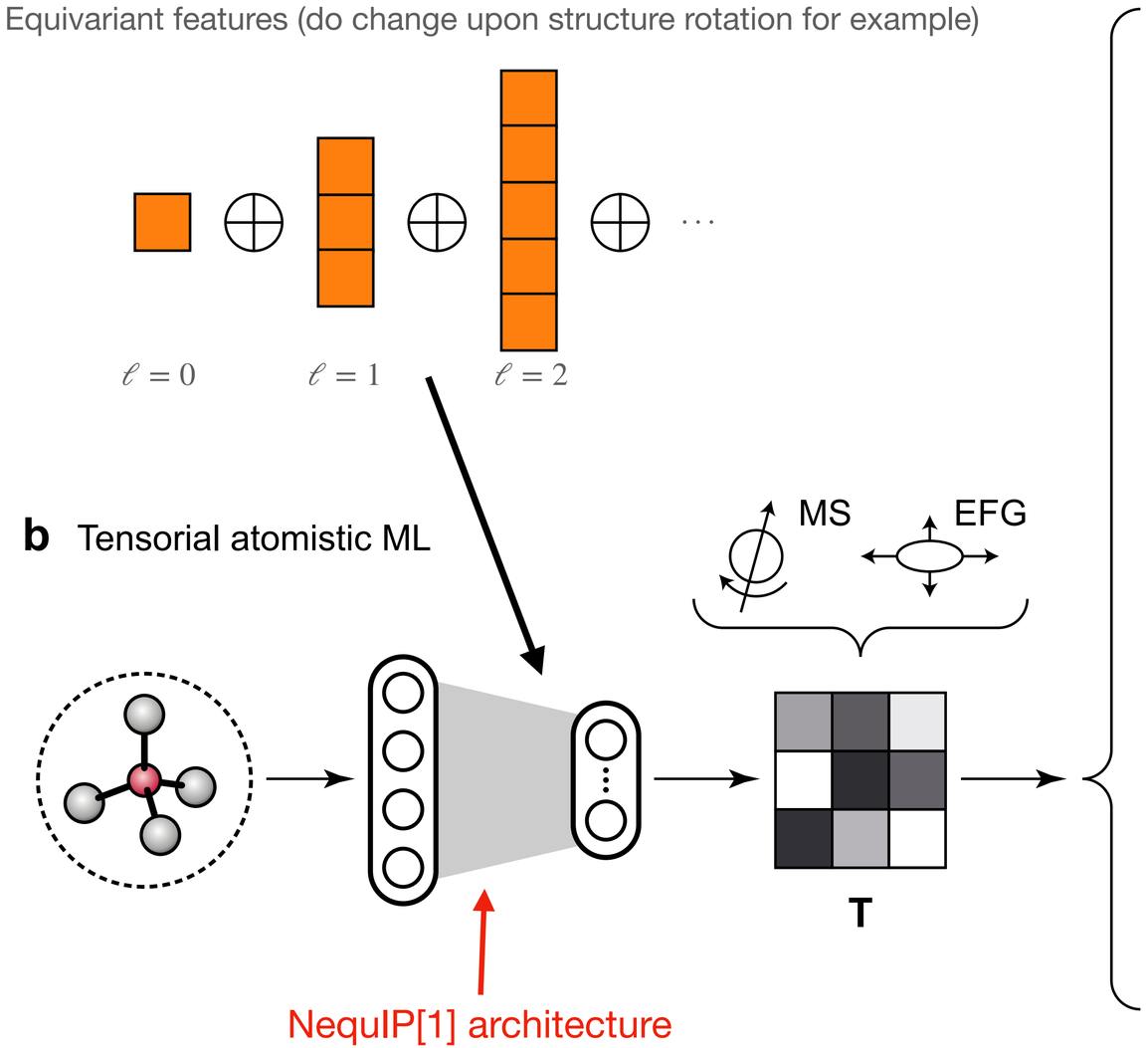


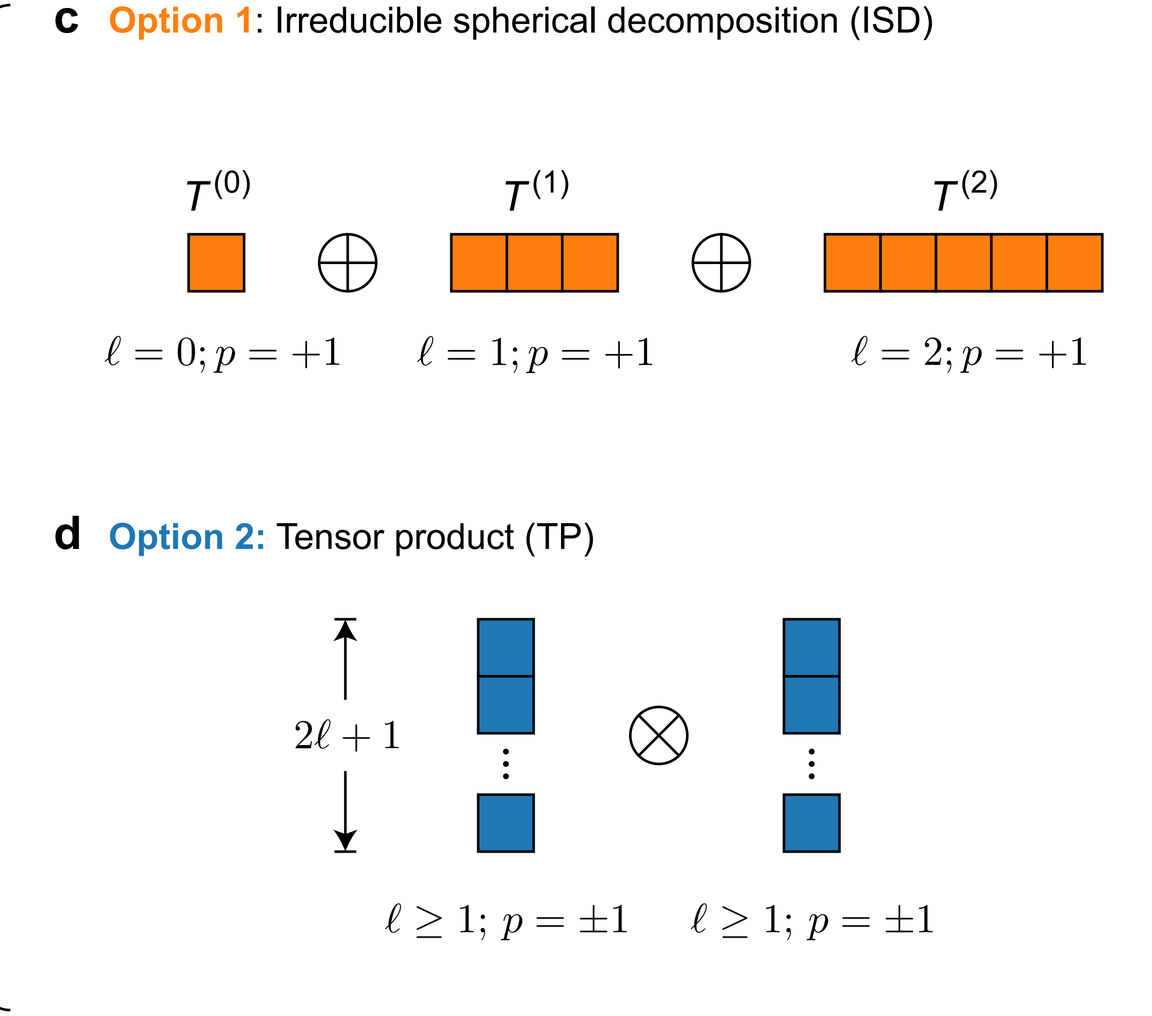
d **Option 2:** Tensor product (TP)



CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063
[1] Batzner et al., 2022

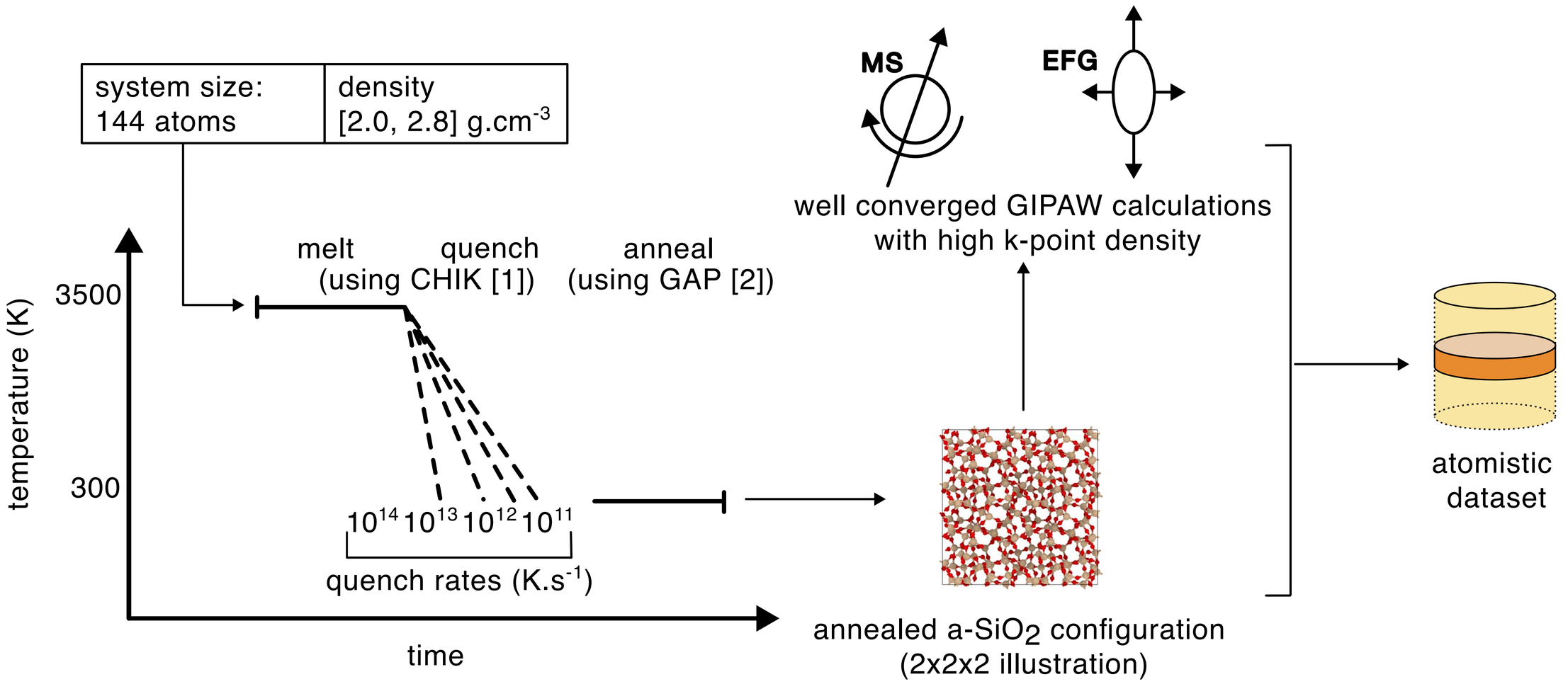
ML architecture





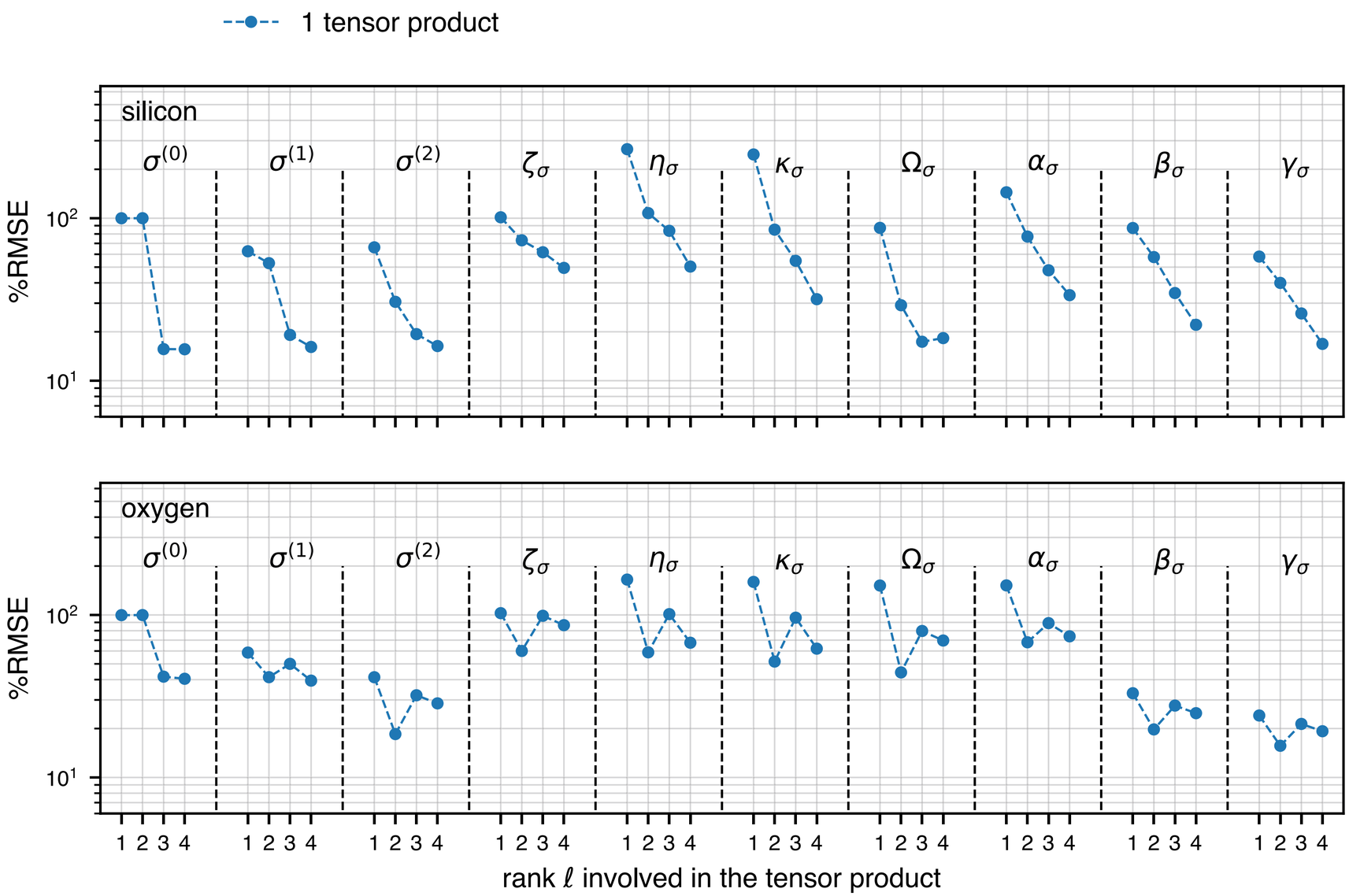
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063
[1] Batzner et al., 2022

a-SiO2 dataset generation protocol



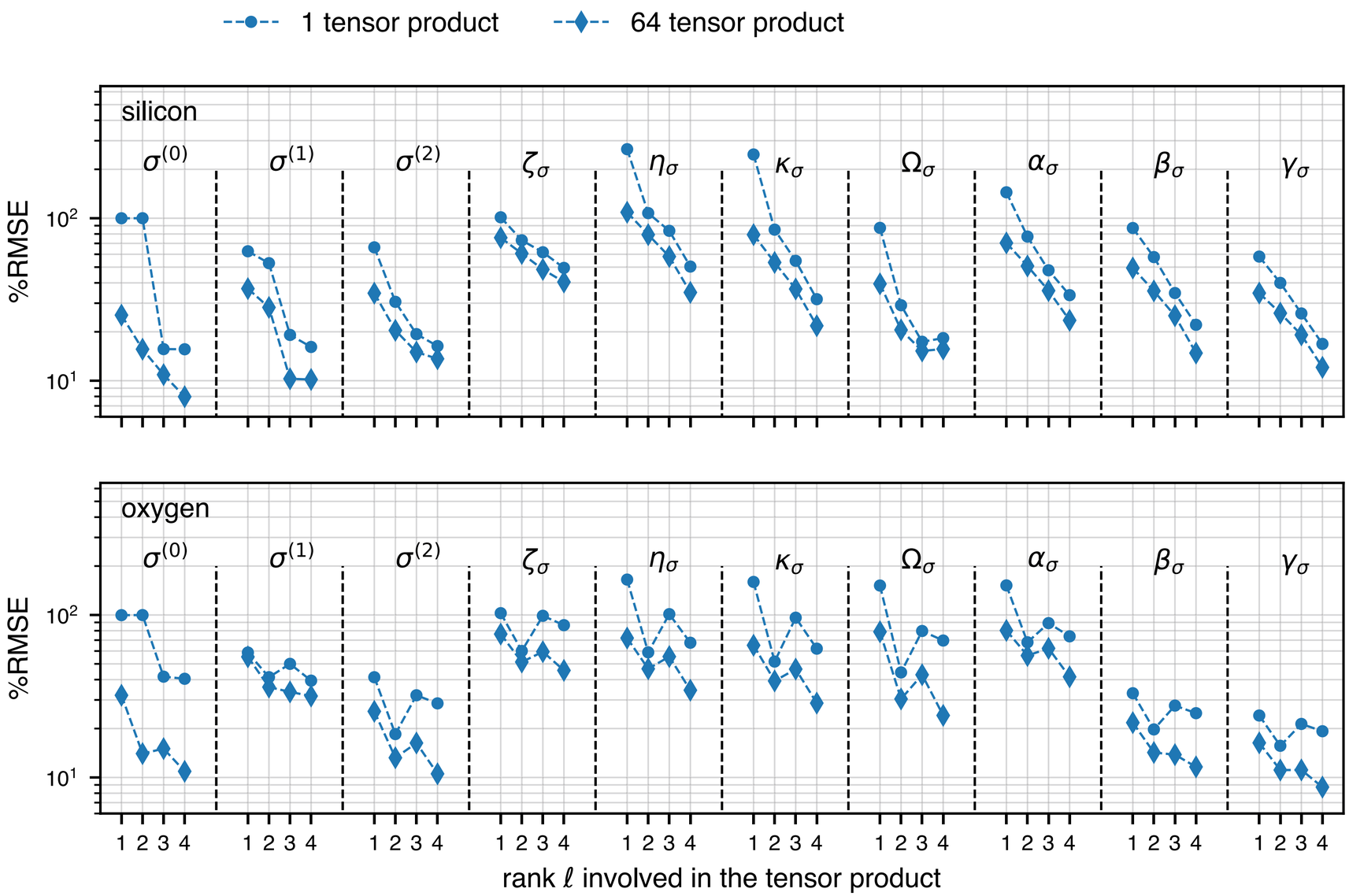
[1] Carré et al., 2008
[2] Bartók et al., 2010

NMR parameters as SH tensor products



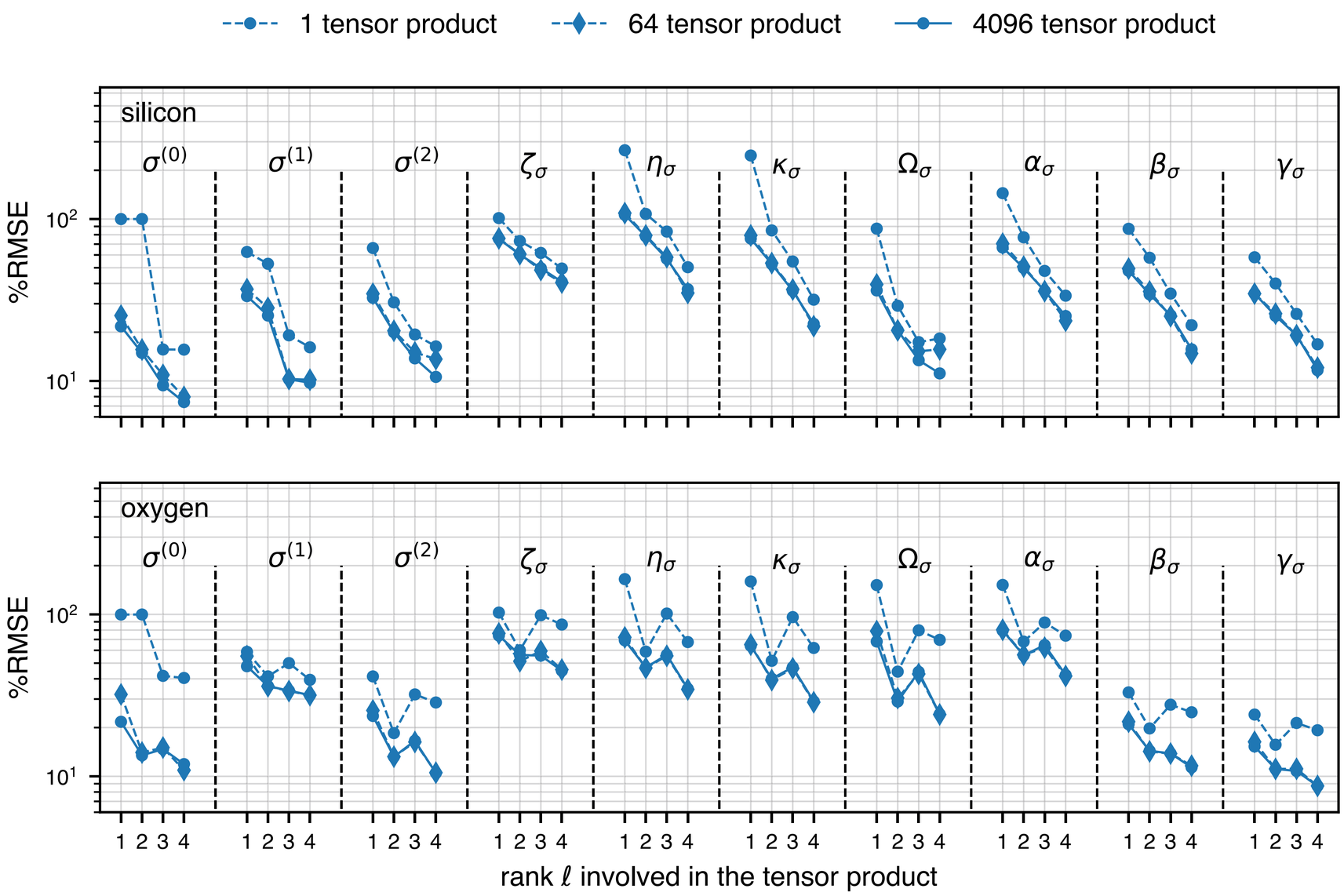
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024. Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition. arXiv: 2412.15063

NMR parameters as SH tensor products



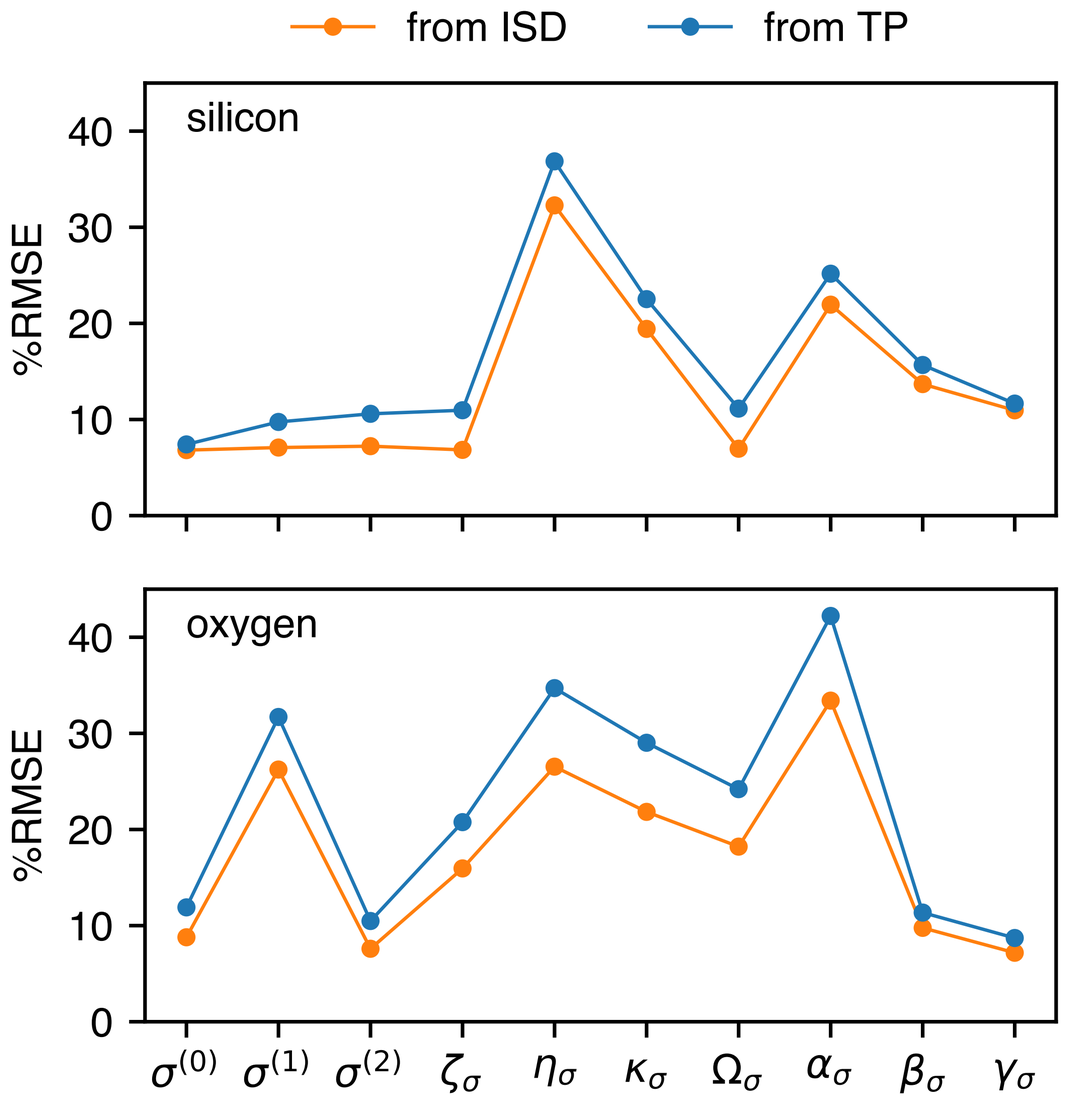
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024. Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition. arXiv: 2412.15063

NMR parameters as SH tensor products



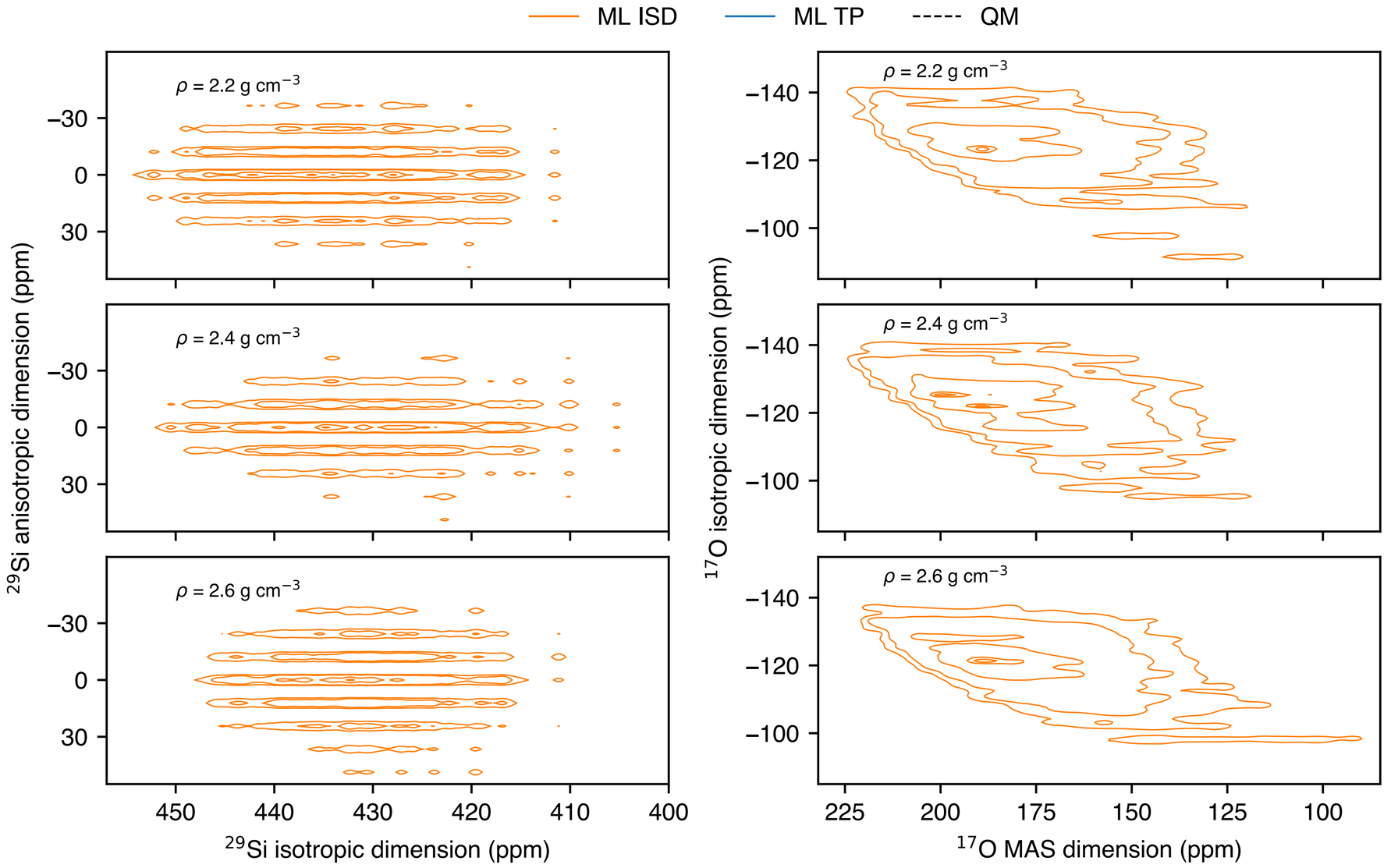
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024. Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition. arXiv: 2412.15063

Two (nearly) equivalent representations

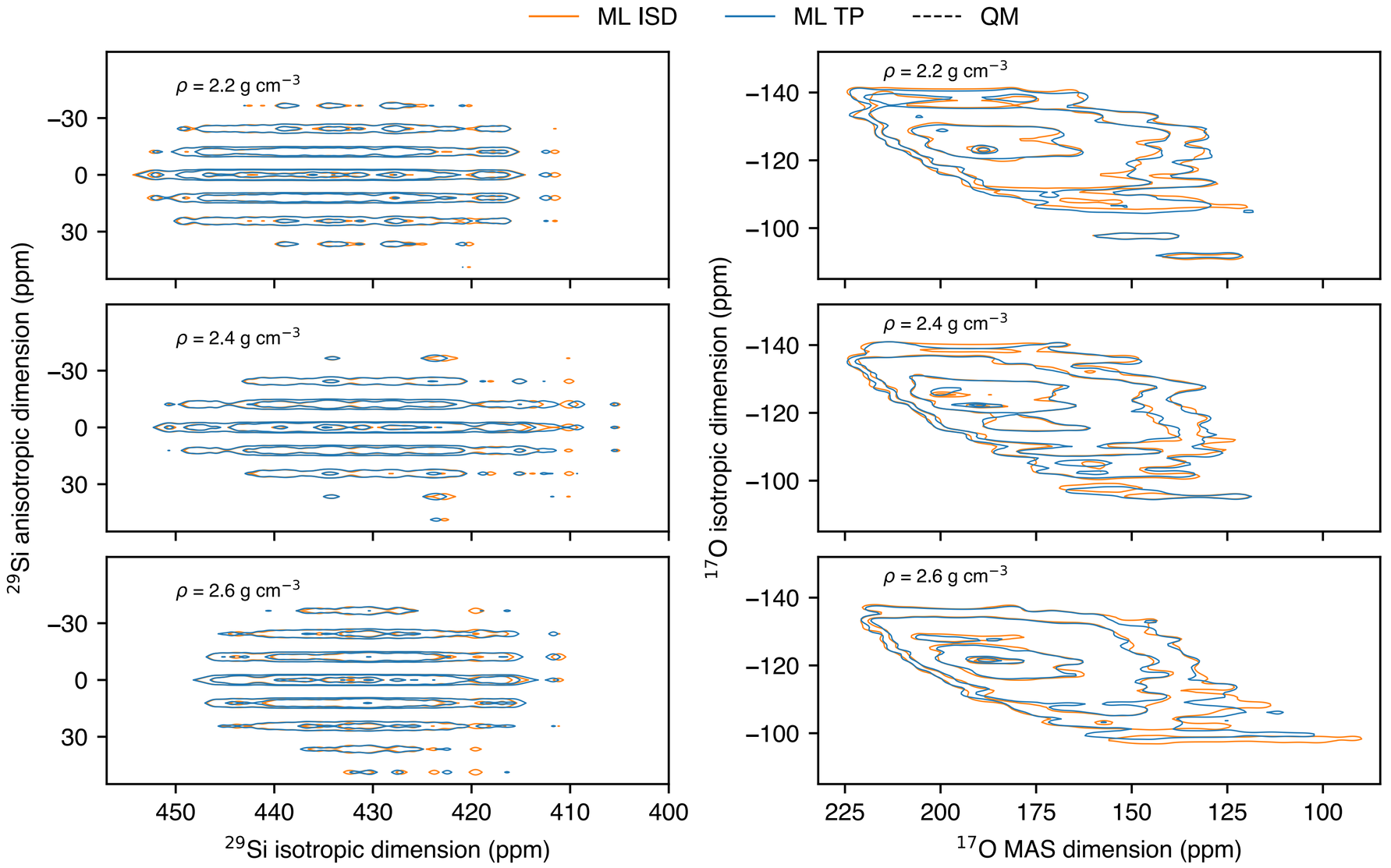


CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024. Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition. arXiv: 2412.15063

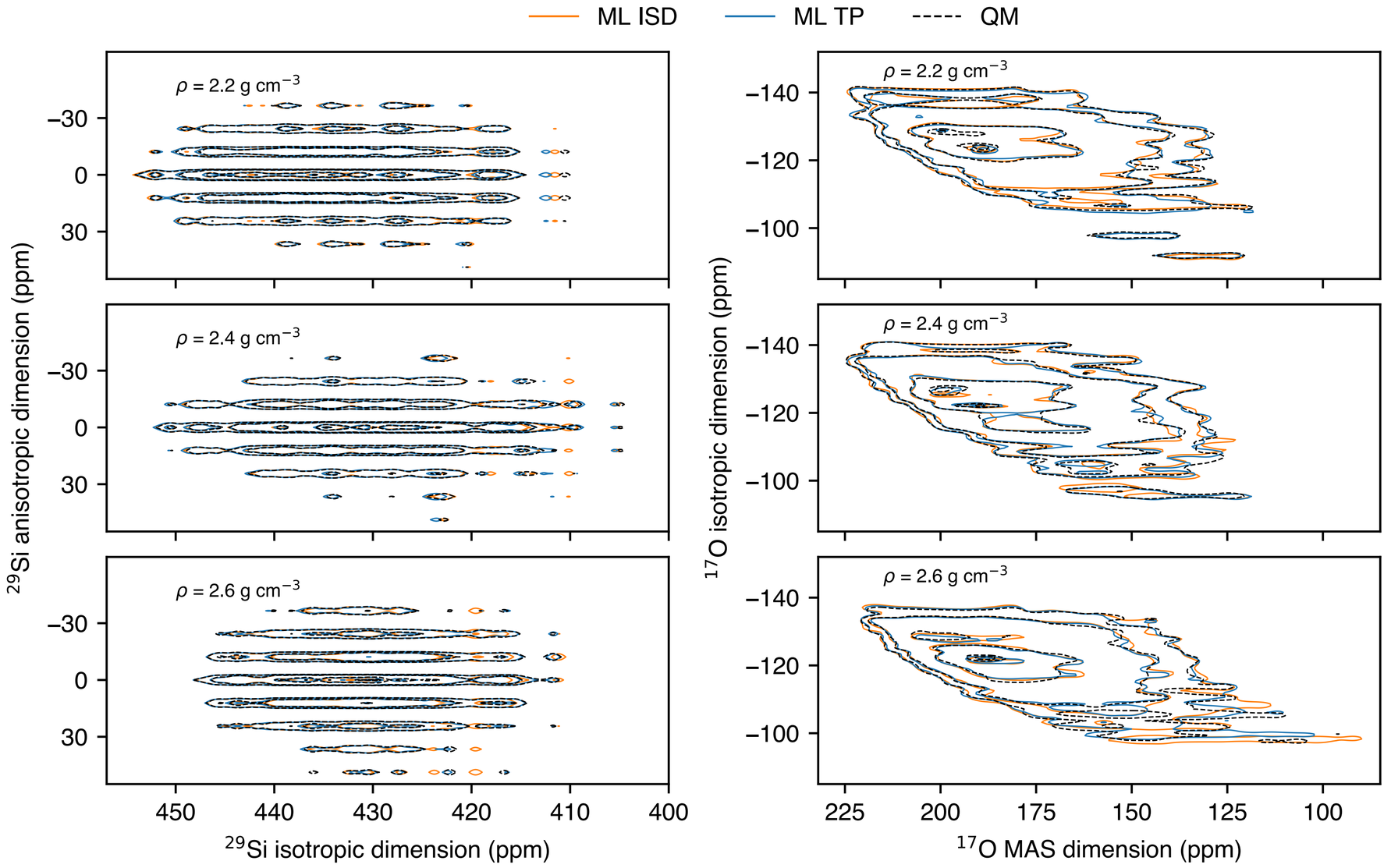
2D NMR spectra of a-SiO_2 models



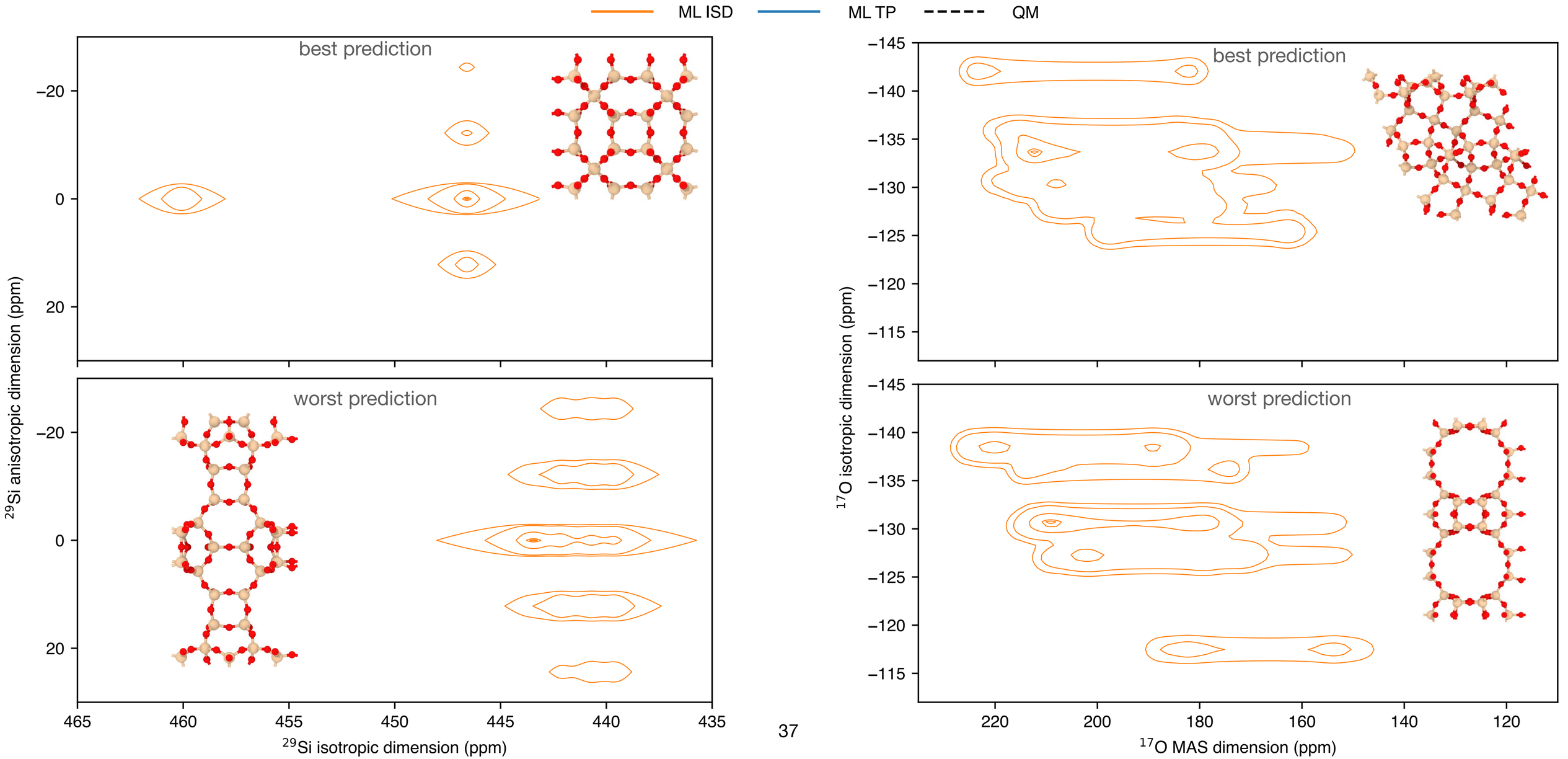
2D NMR spectra of a-SiO_2 models



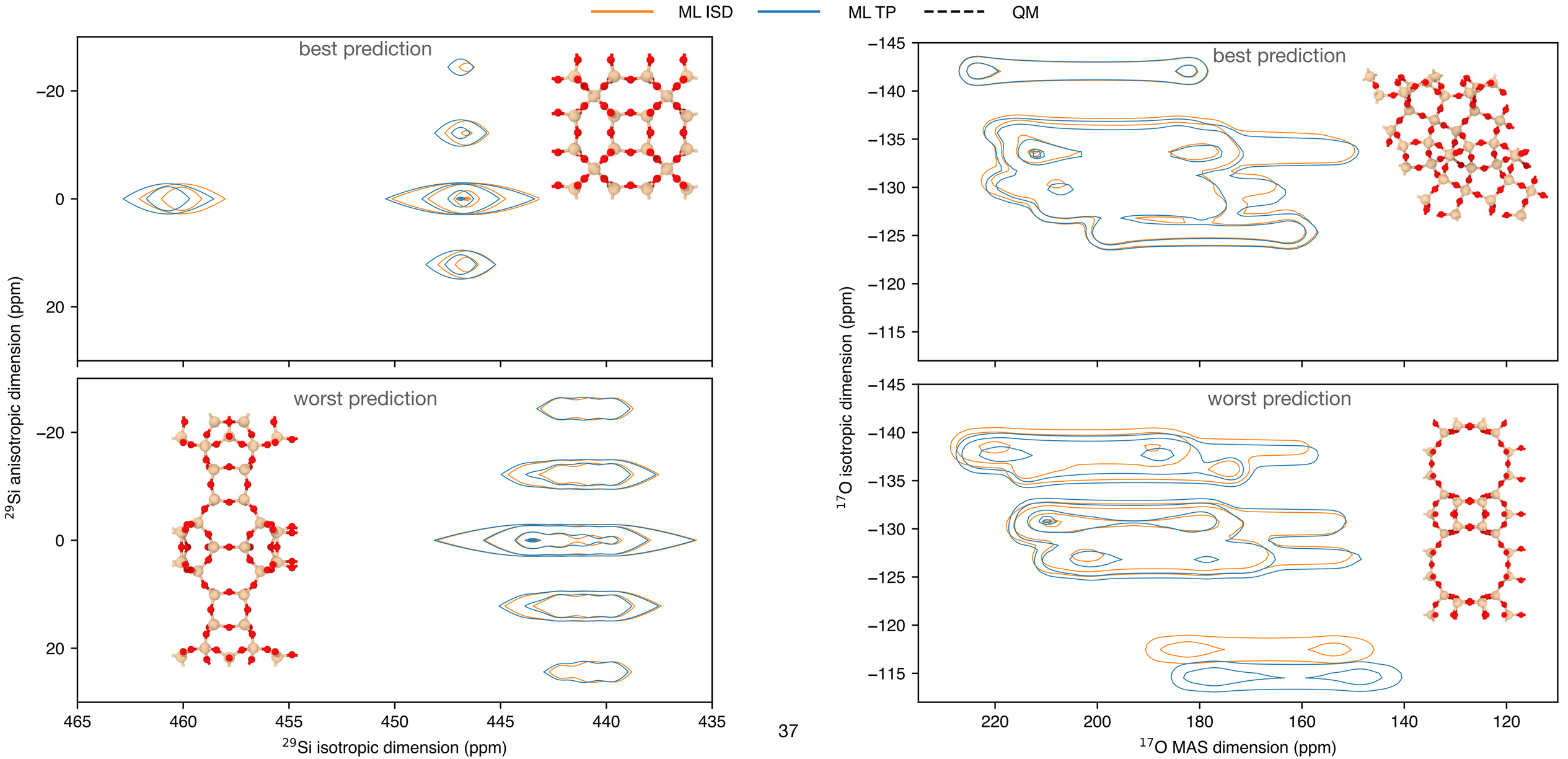
2D NMR spectra of a-SiO_2 models



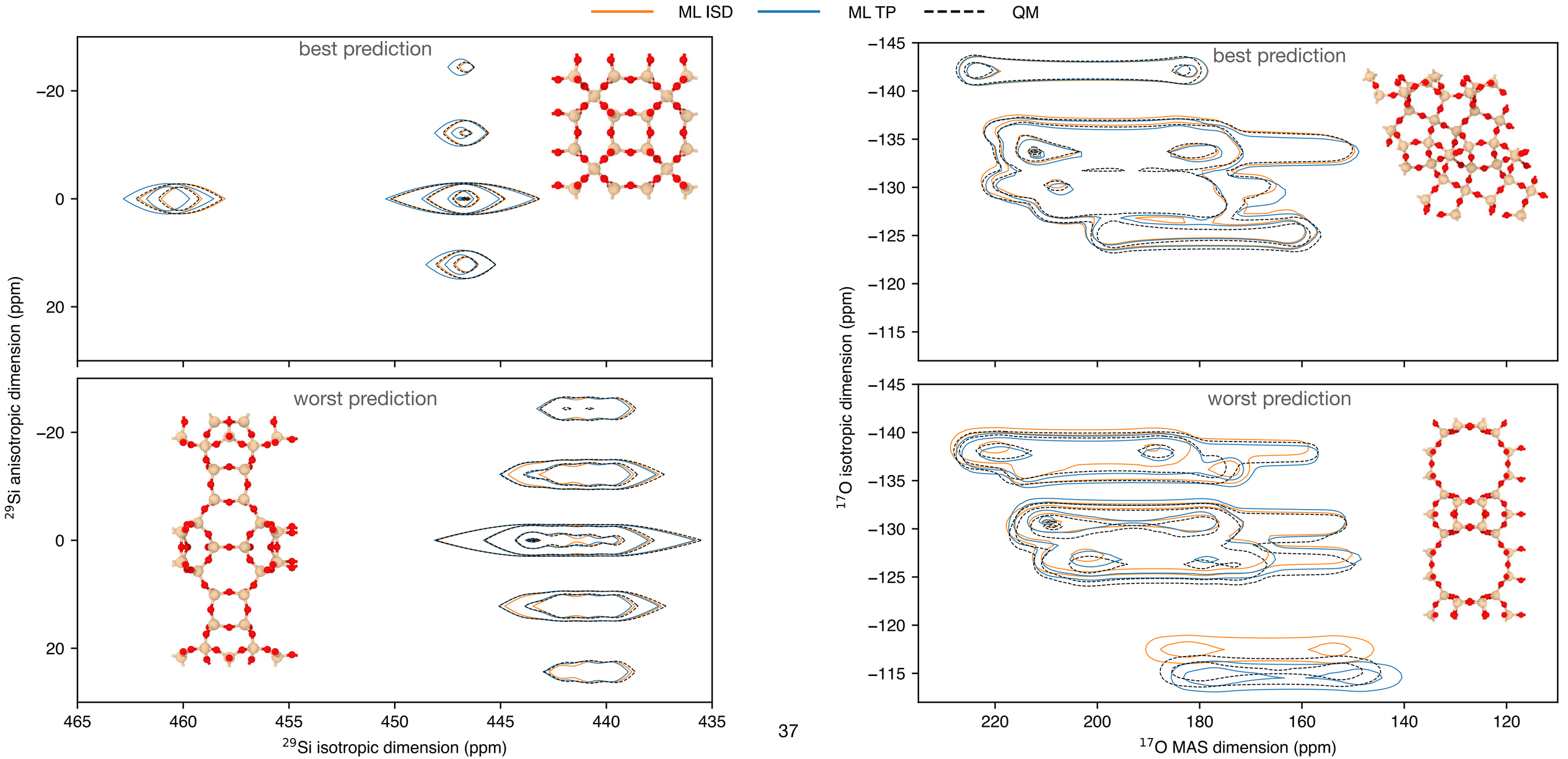
2D NMR spectra for hypothetical zeolites



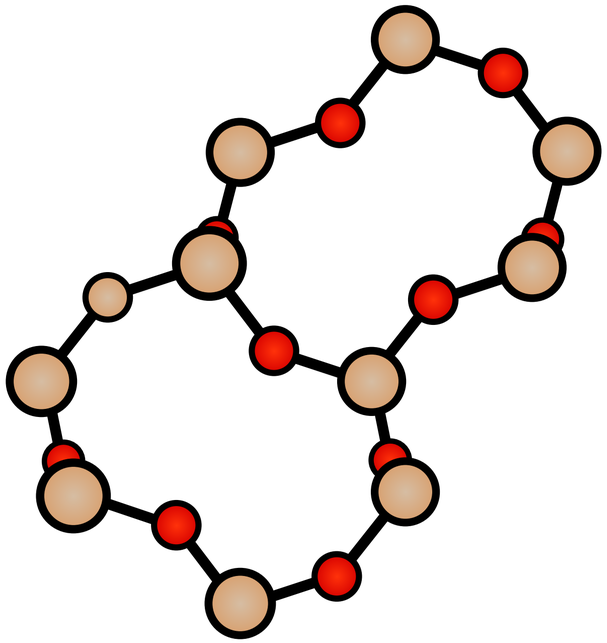
2D NMR spectra for hypothetical zeolites



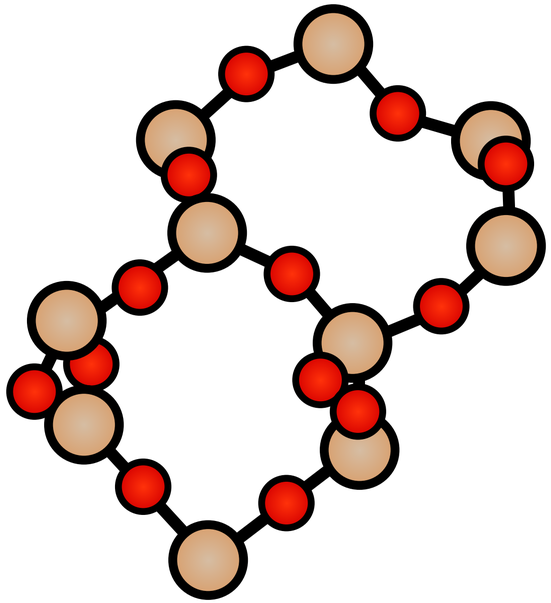
2D NMR spectra for hypothetical zeolites



Dynamics of α—β inversion in cristobalite

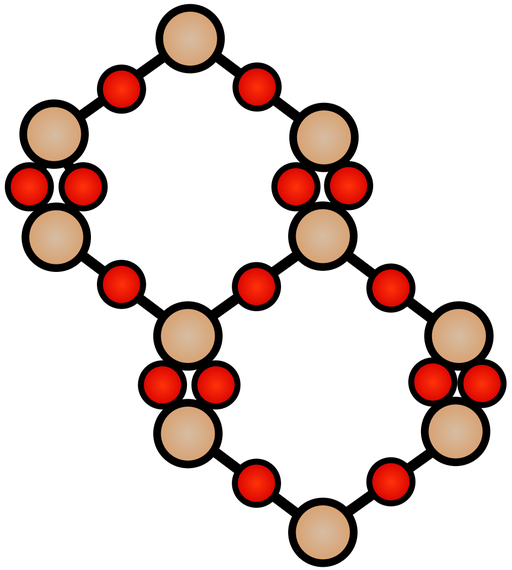


α environments



50% β environments

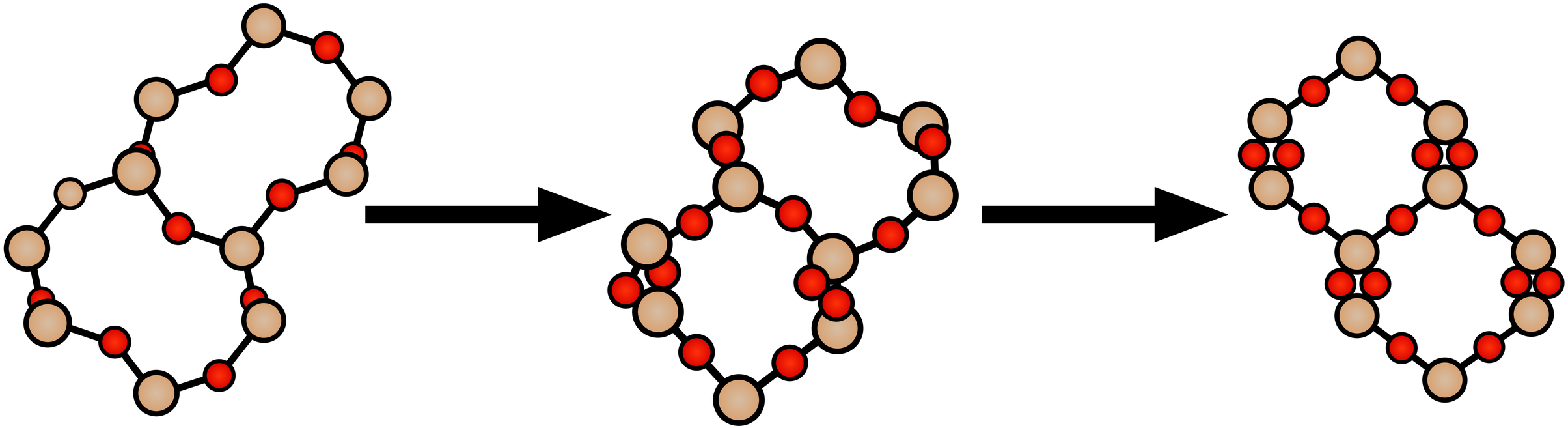
50% α environments



β environments

Erhard et al, 2024

Dynamics of α—β inversion in cristobalite



α environments

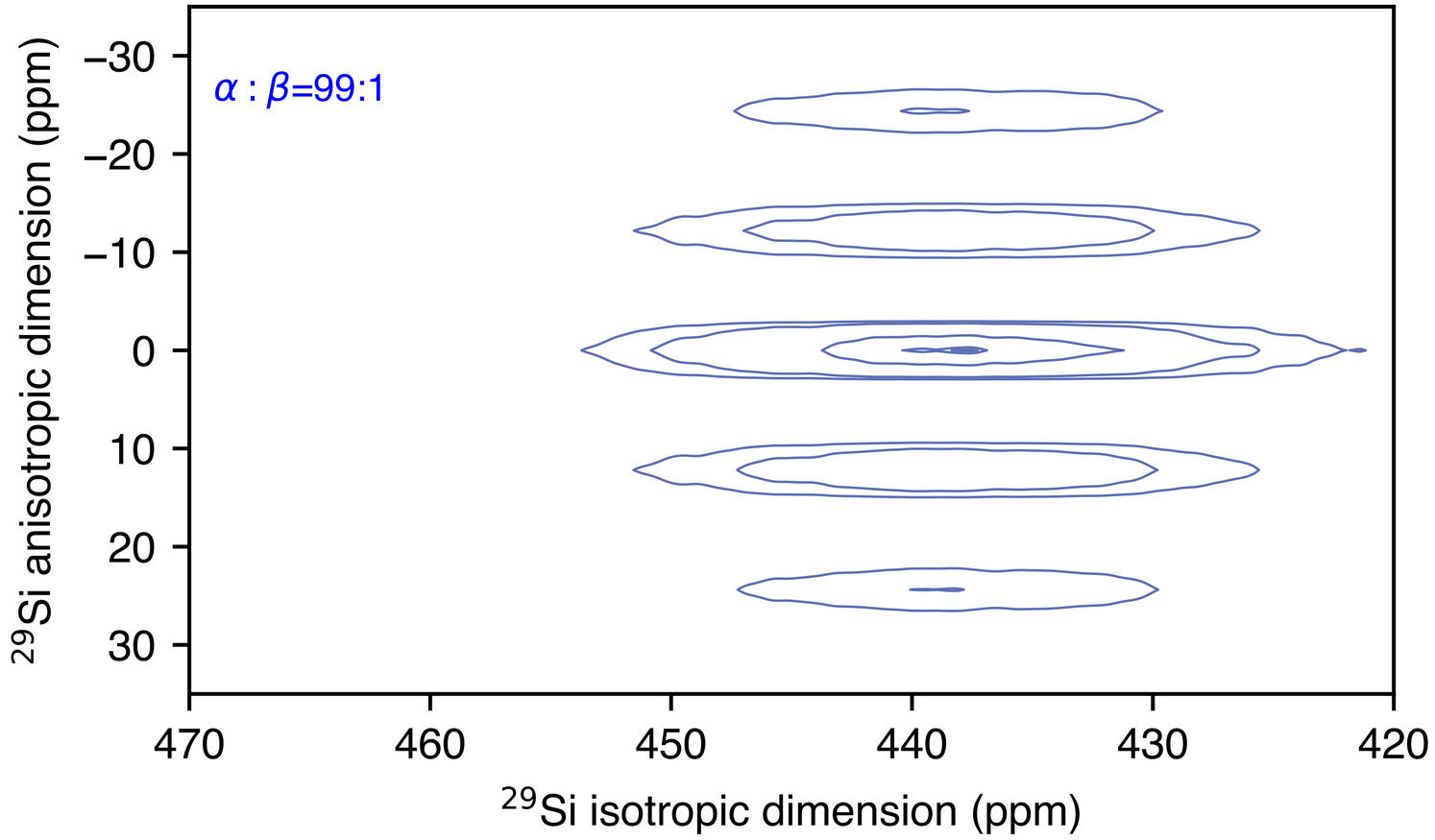
50% β environments

50% α environments

β environments

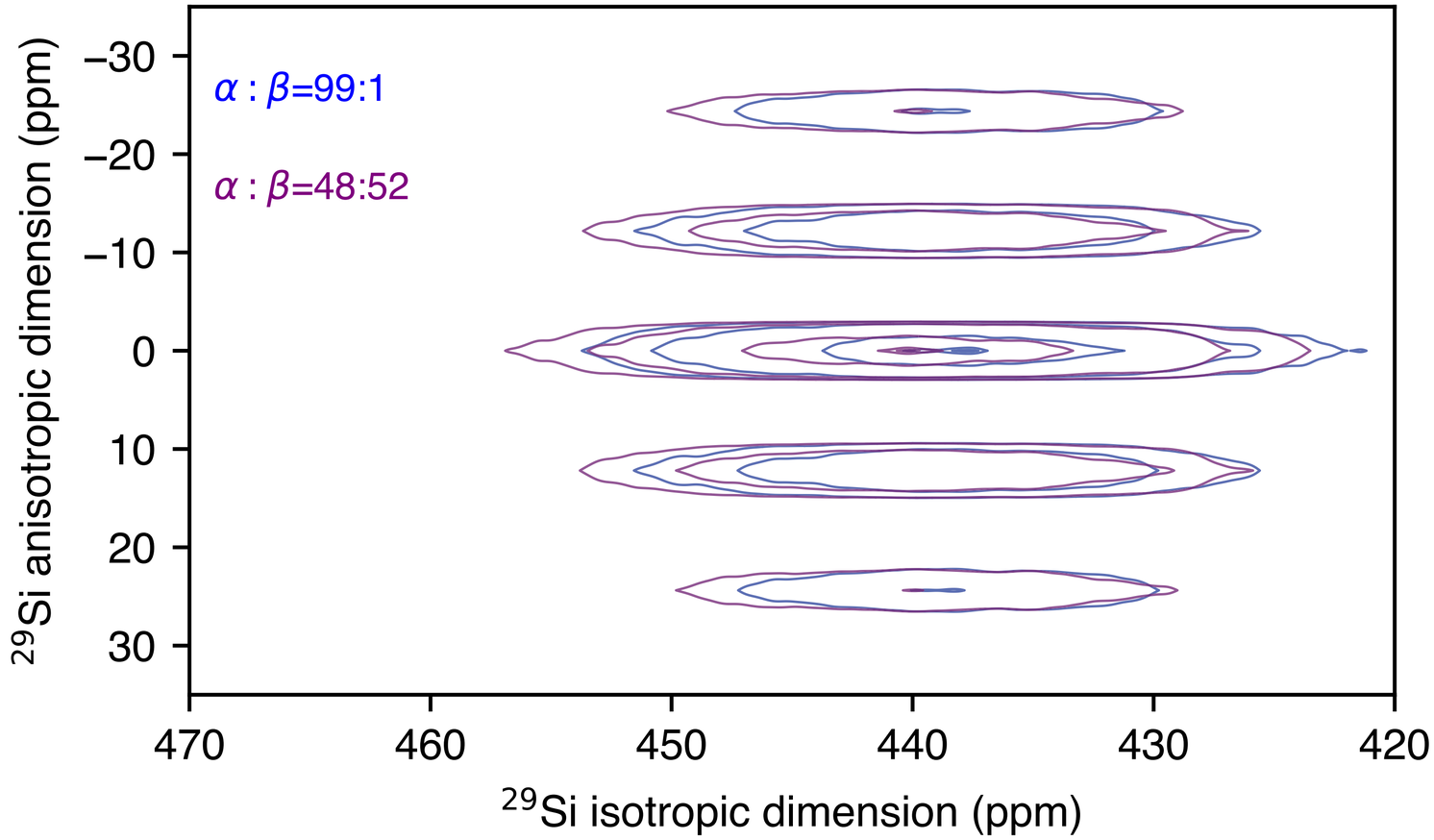
Erhard et al, 2024

Dynamics of α—β inversion in cristobalite



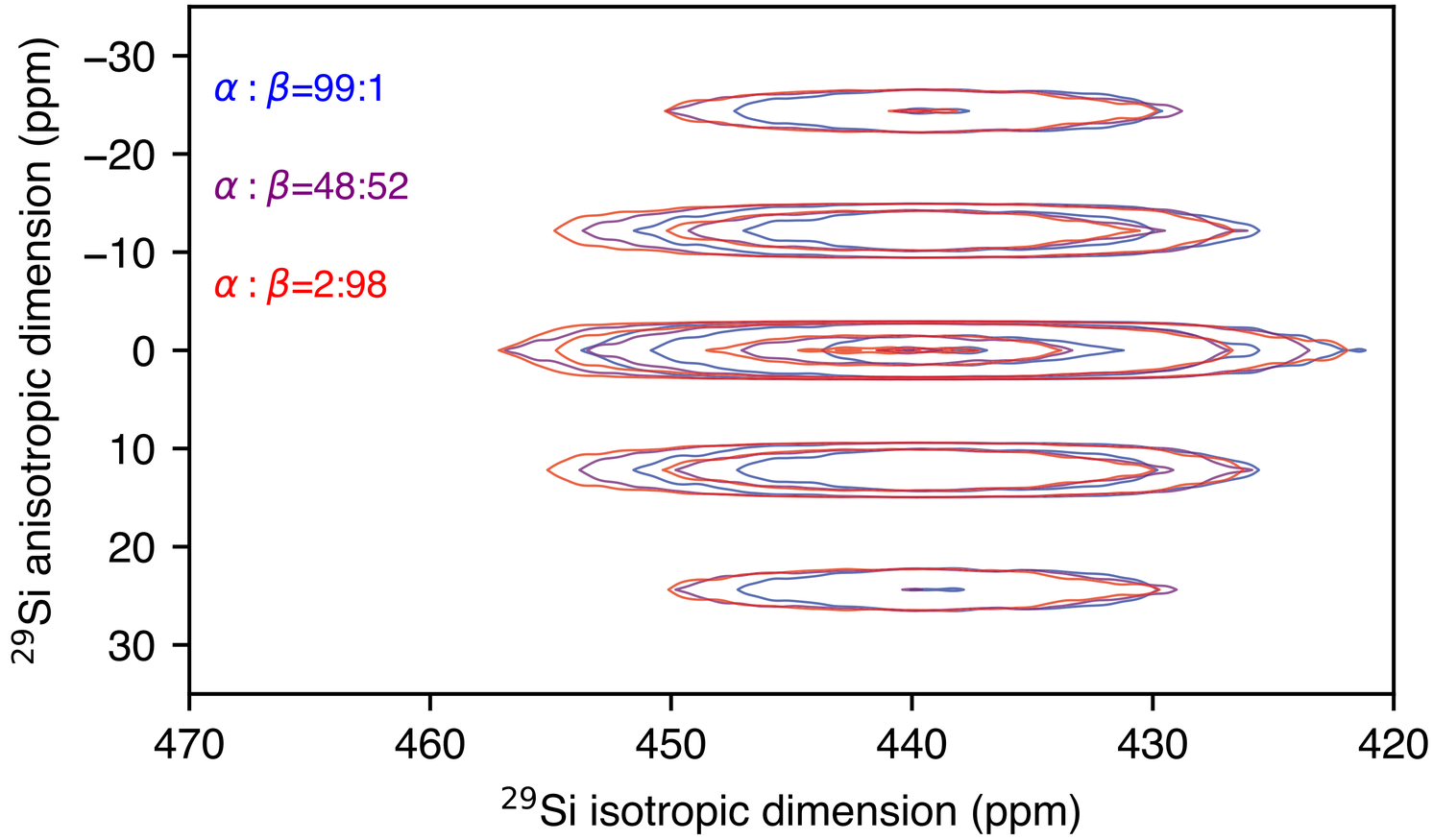
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

Dynamics of α—β inversion in cristobalite



CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

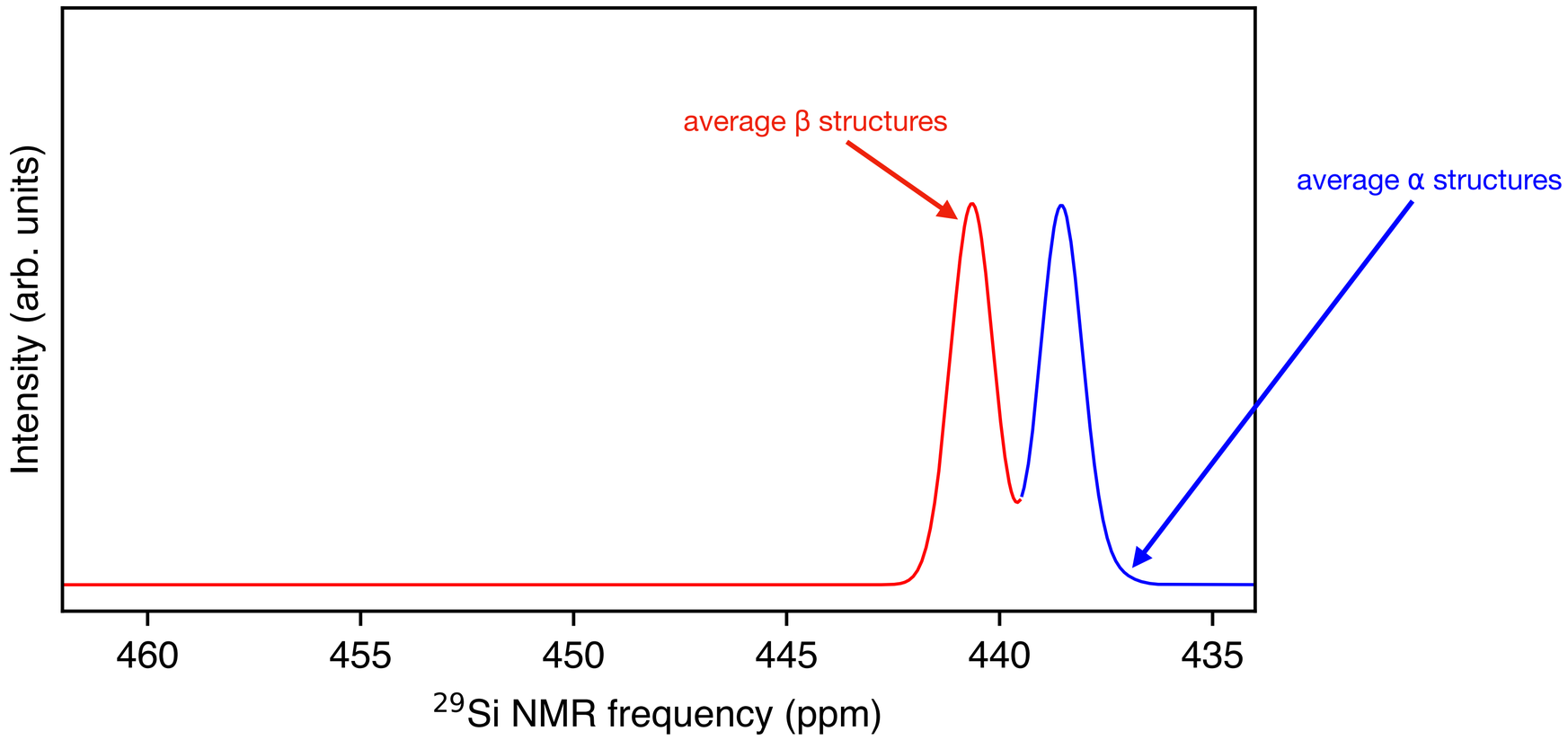
Dynamics of α—β inversion in cristobalite



CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

Dynamics of α—β inversion in cristobalite

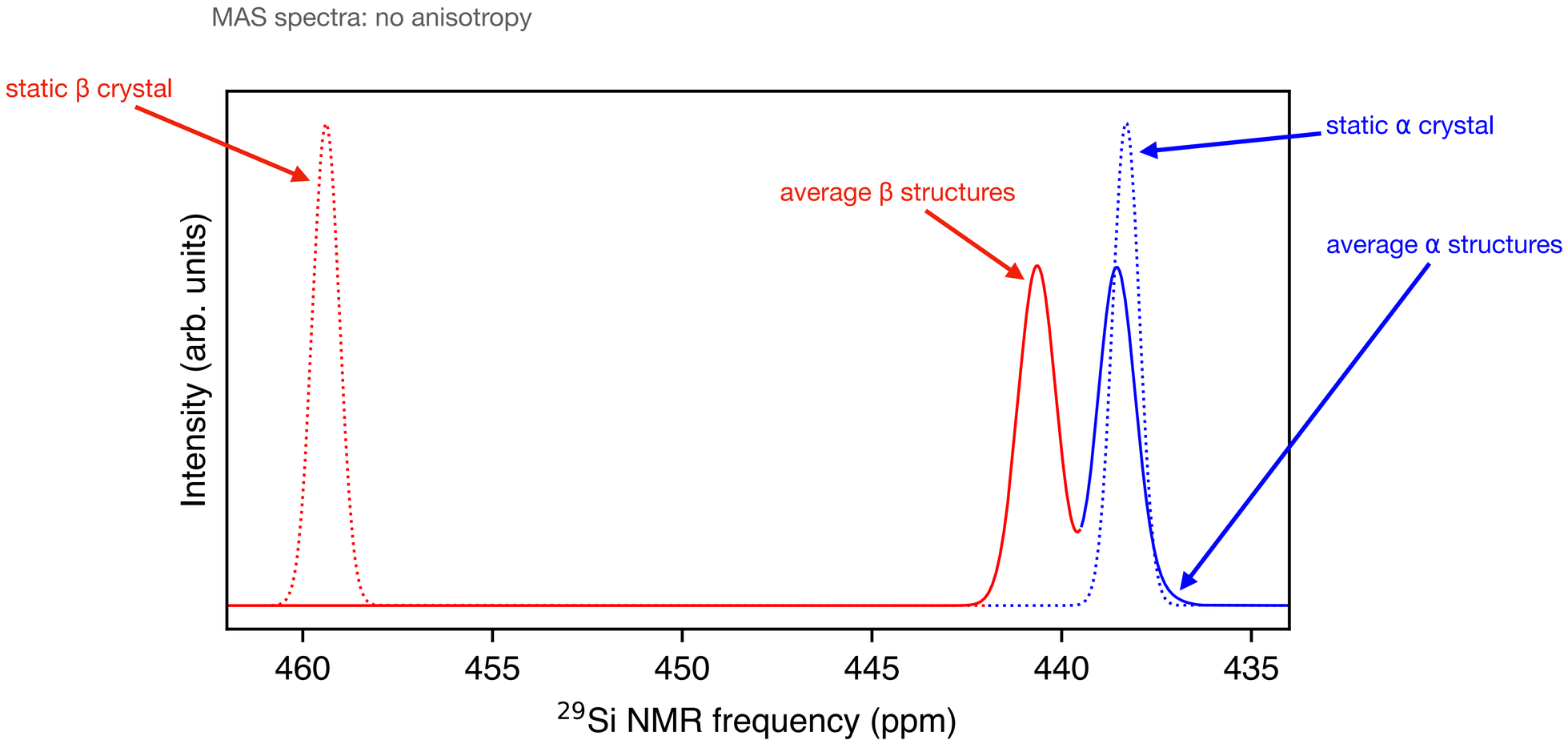
MAS spectra: no anisotropy



CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

Spearing, Farnan, and Stebbins, 1992

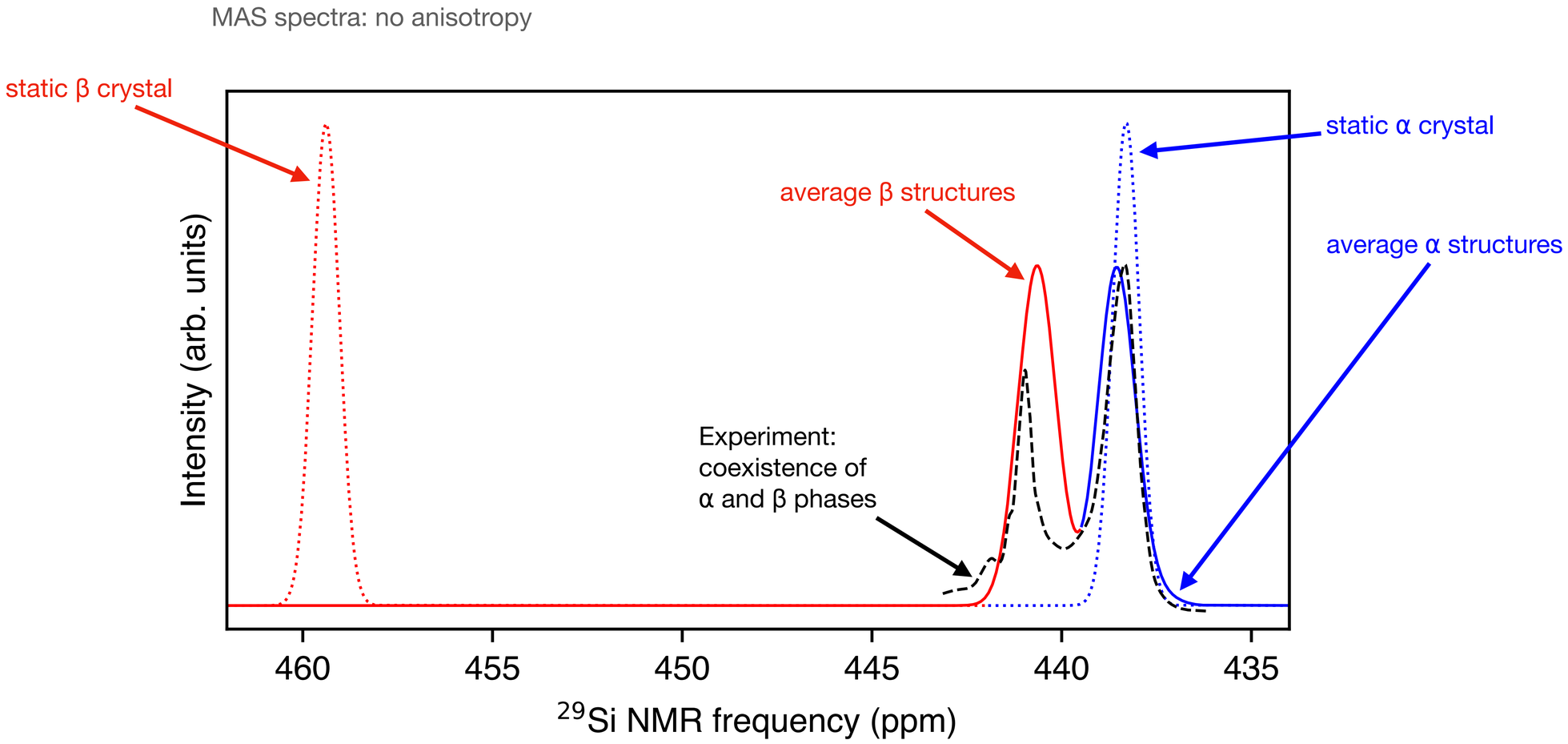
Dynamics of α—β inversion in cristobalite



CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

Spearing, Farnan, and Stebbins, 1992

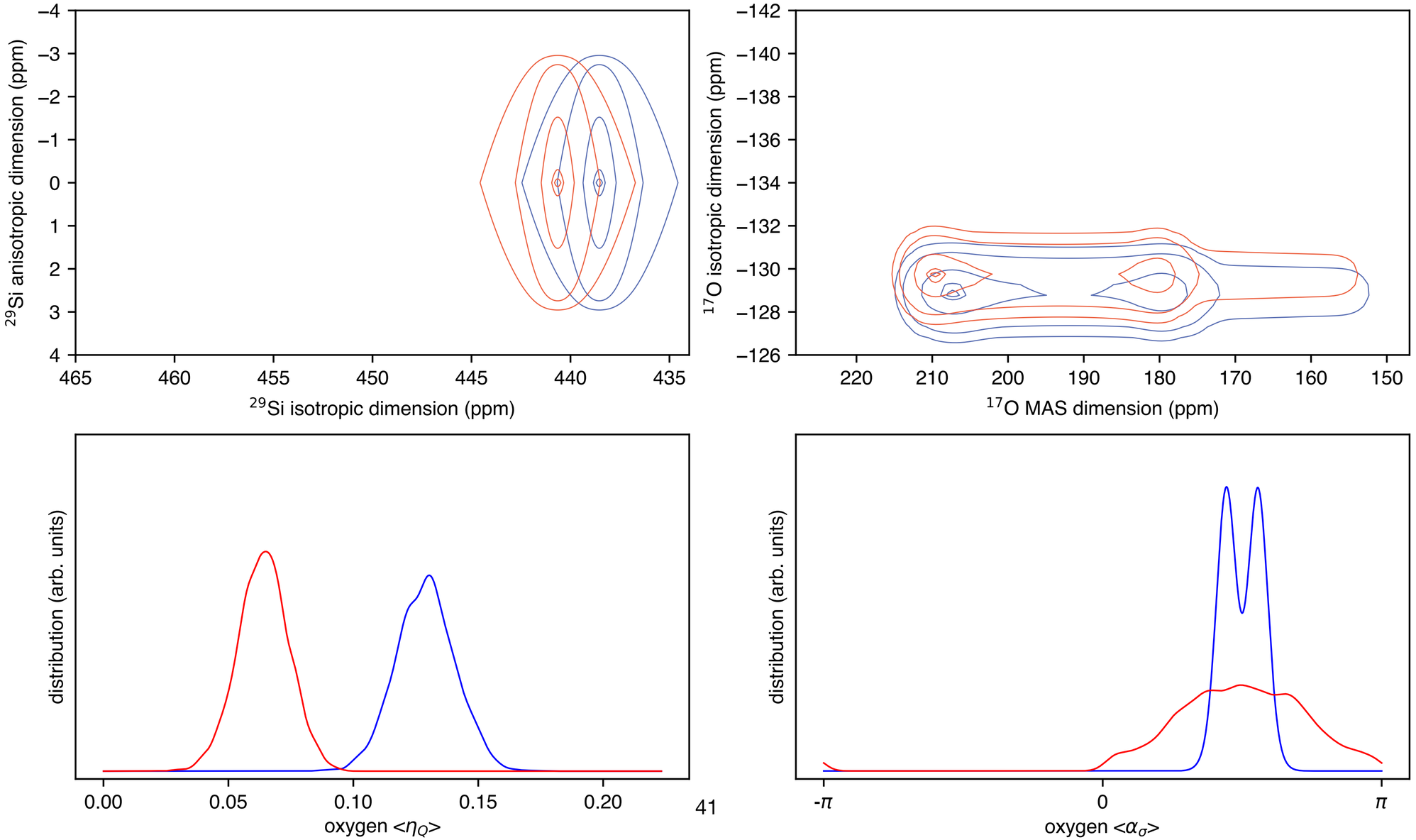
Dynamics of α—β inversion in cristobalite



CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024.
Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition.
arXiv: 2412.15063

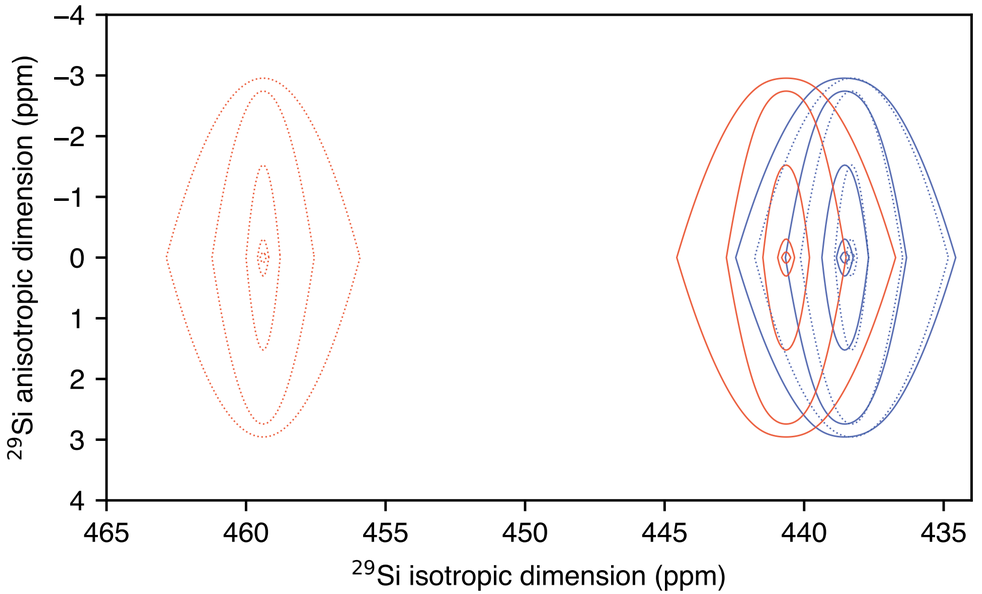
Spearing, Farnan, and Stebbins, 1992

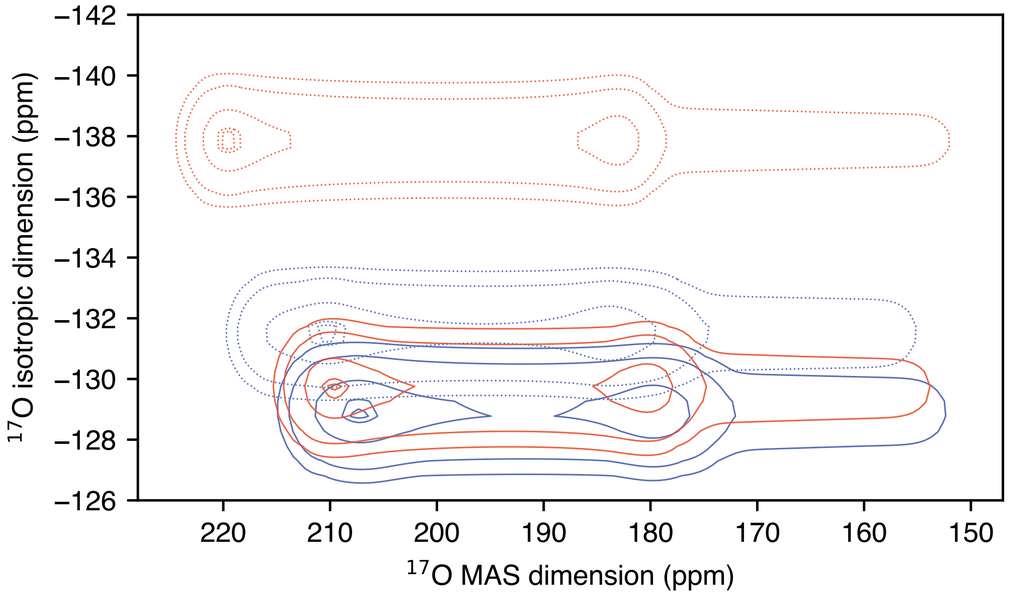
Dynamics of α—β inversion in cristobalite

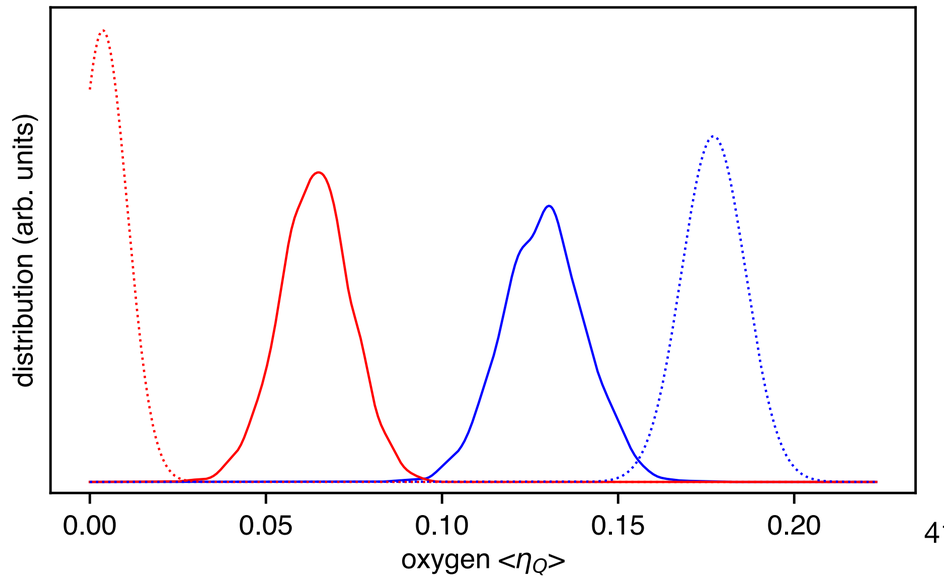


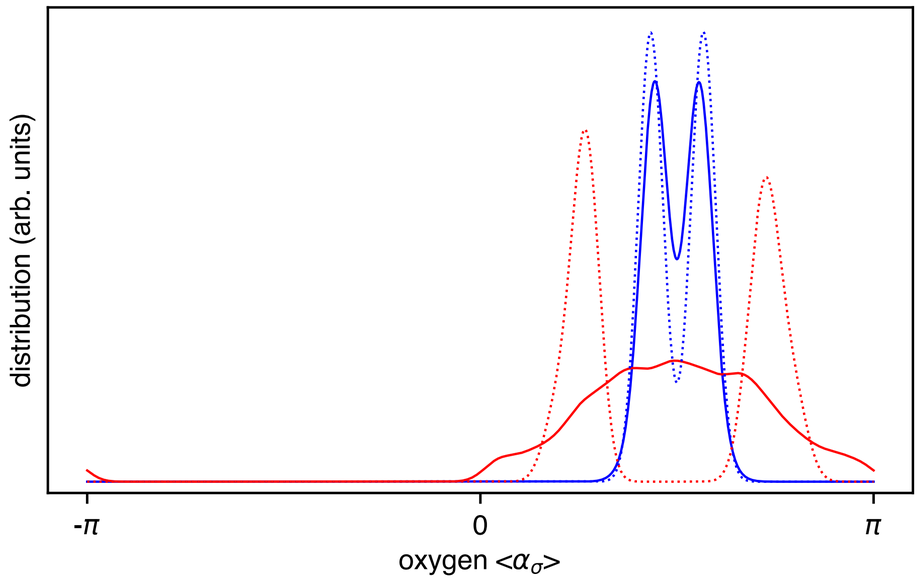
CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024. Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition. arXiv: 2412.15063

Dynamics of α—β inversion in cristobalite









CBM, Rosset, L.A.M., Yates, J.R., Deringer, V.L., 2024. Graph-Neural-Network predictions of solid-state NMR parameters from spherical tensor decomposition. arXiv: 2412.15063

Takehome message

Takehome message

- Atom-centered atomistic ML formulation extends naturally to NMR parameters

Takehome message

- Atom-centered atomistic ML formulation extends naturally to NMR parameters
- NMR models provide an atomistic interpretation of experimental observables for solids and amorphous phases

Takehome message

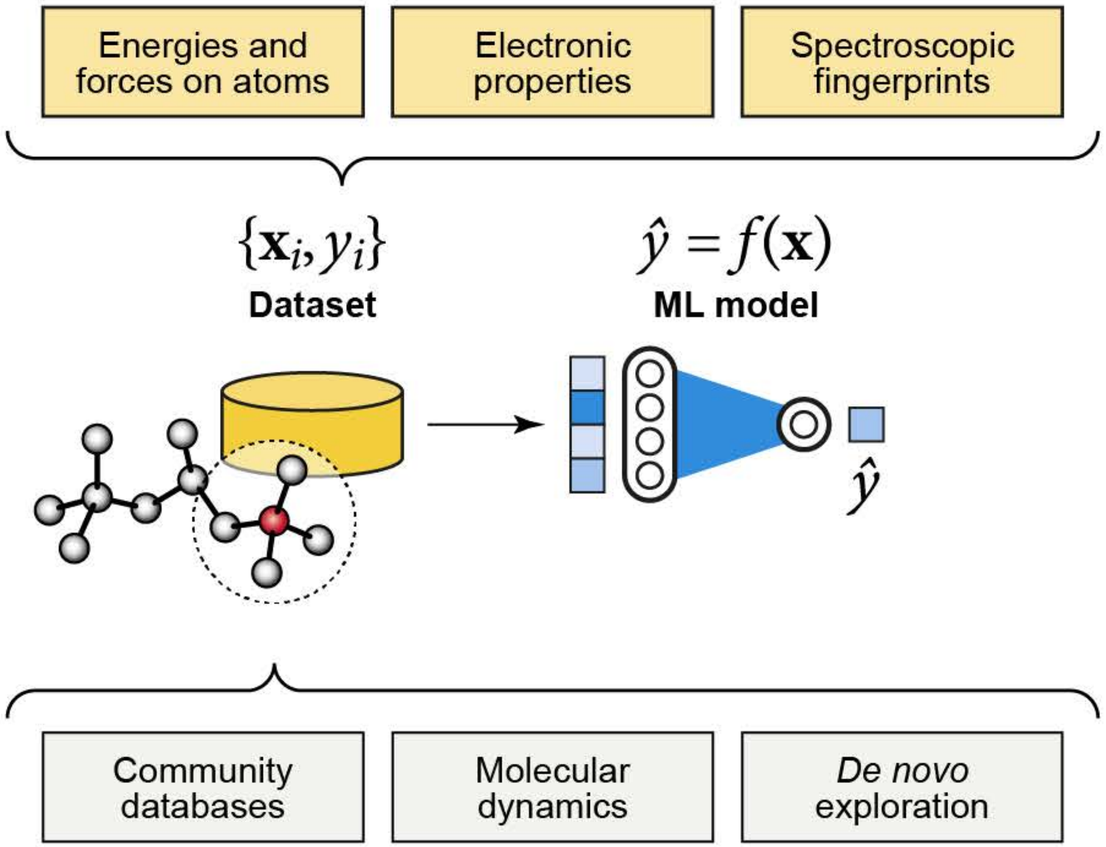
- Atom-centered atomistic ML formulation extends naturally to NMR parameters
- NMR models provide an atomistic interpretation of experimental observables for solids and amorphous phases
- Flexible architectures open the door for multi-property learning

Property 3:

potential energy surface

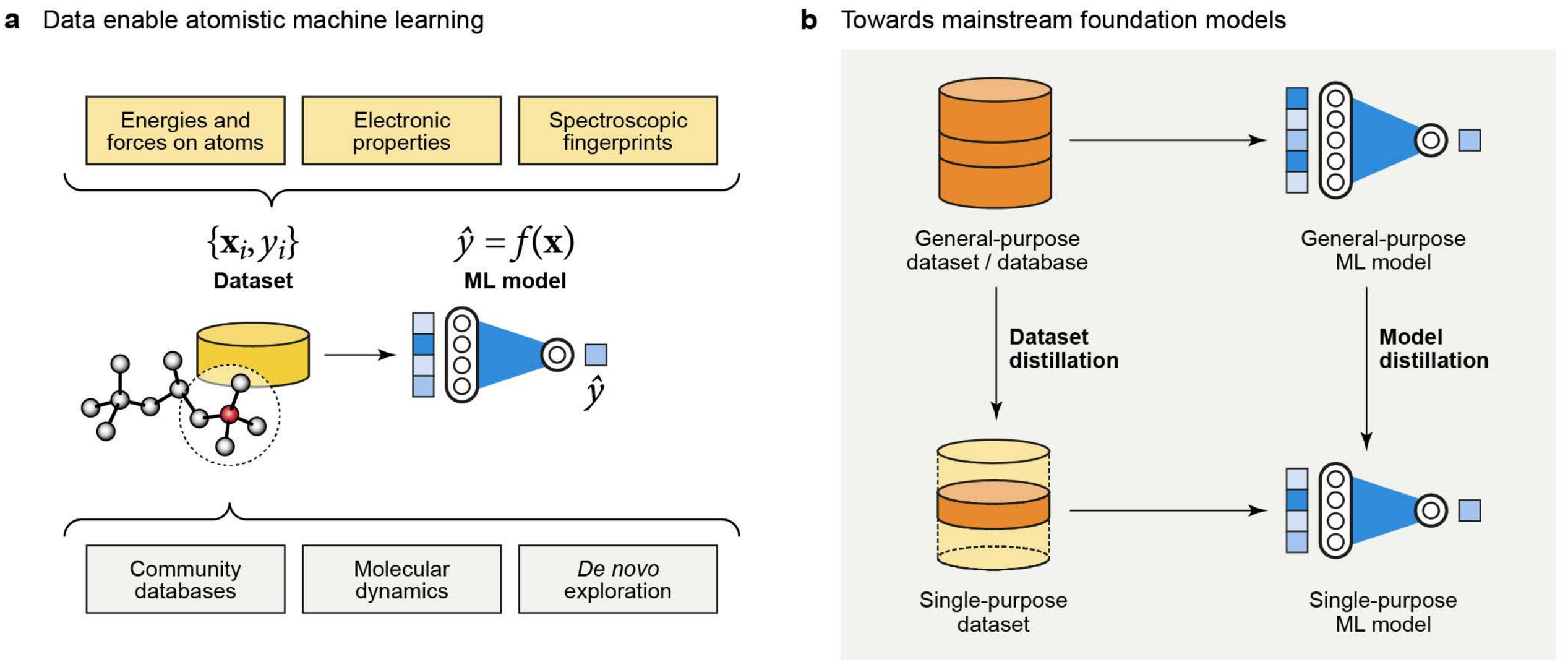
Towards general-purpose models/datasets

a Data enable atomistic machine learning



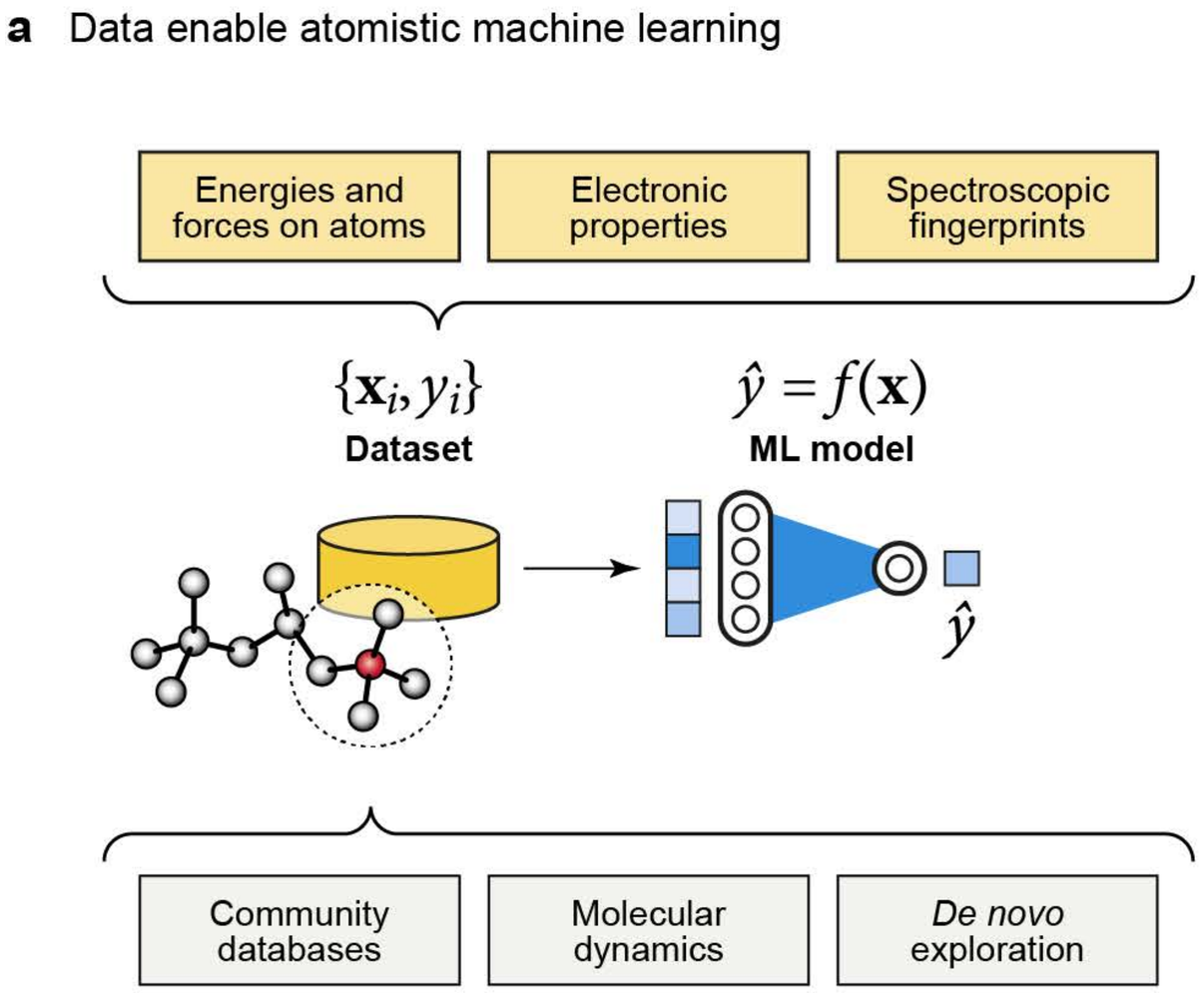
CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

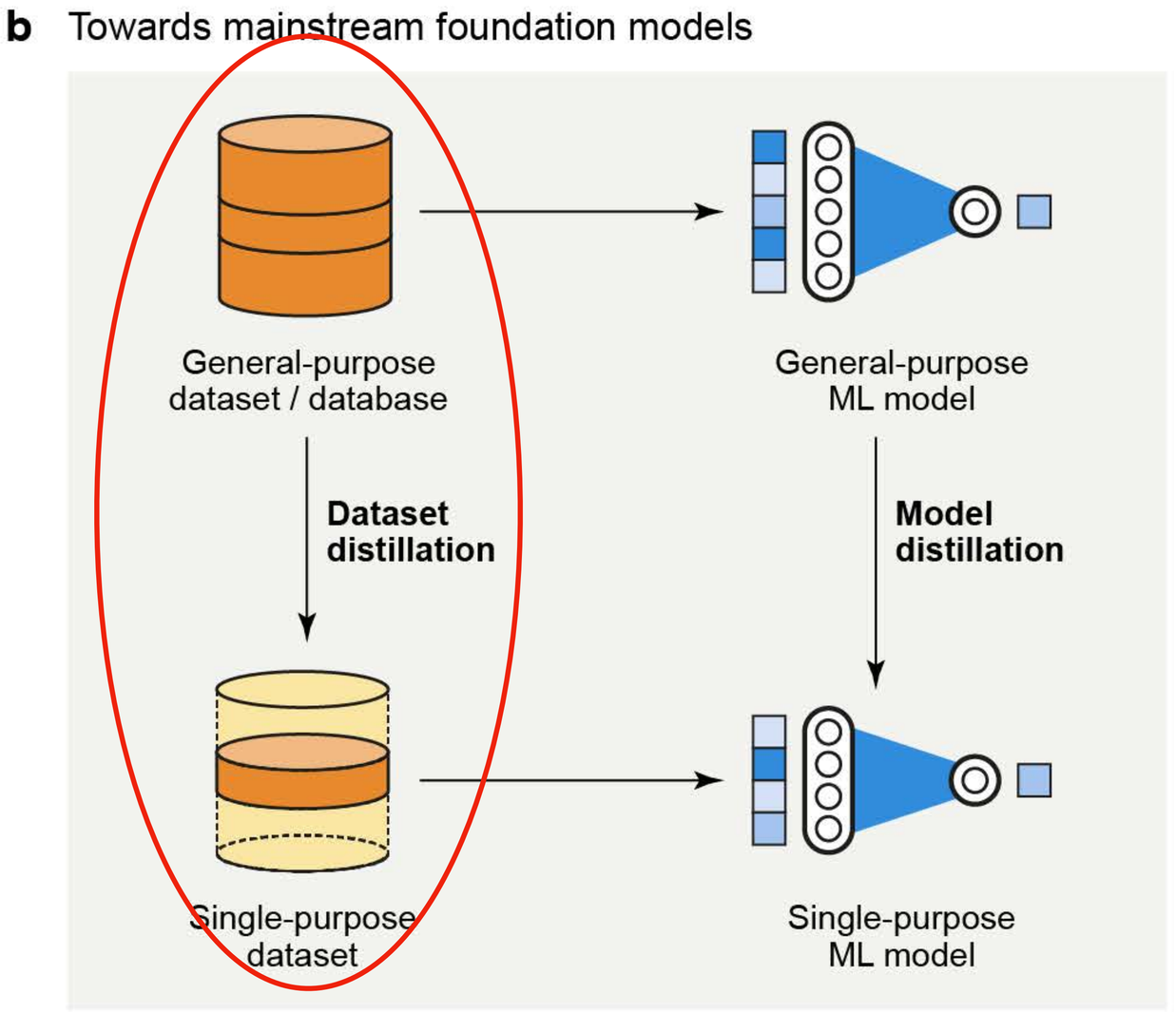
Towards general-purpose models/datasets



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

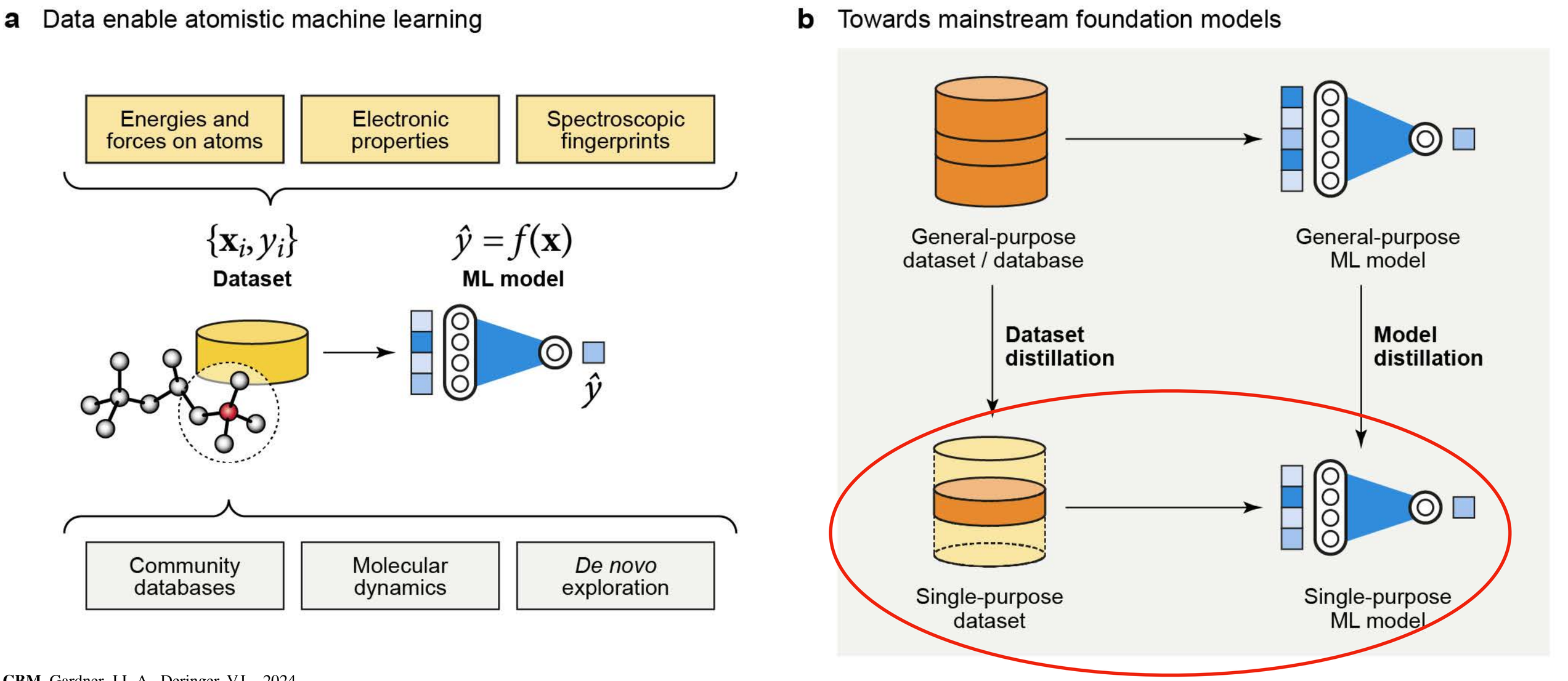
Towards general-purpose models/datasets





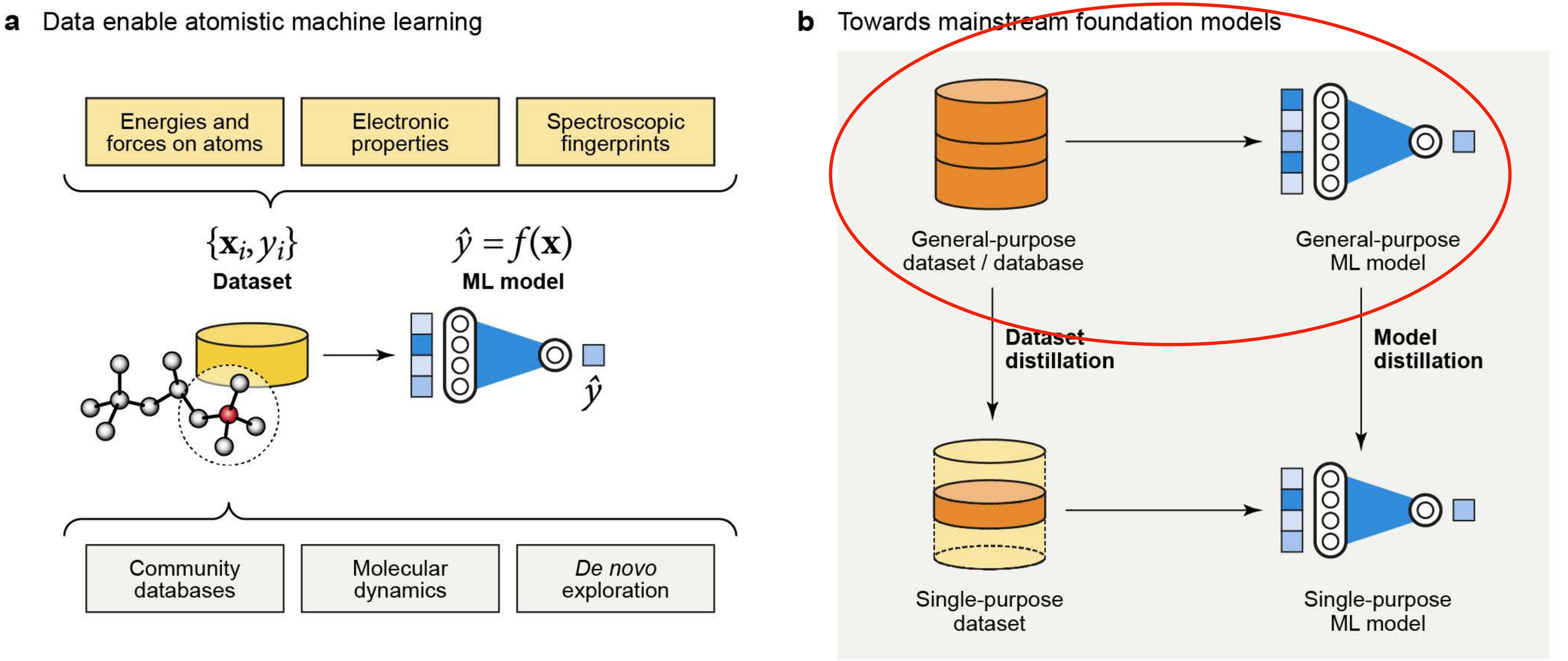
CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Towards general-purpose models/datasets



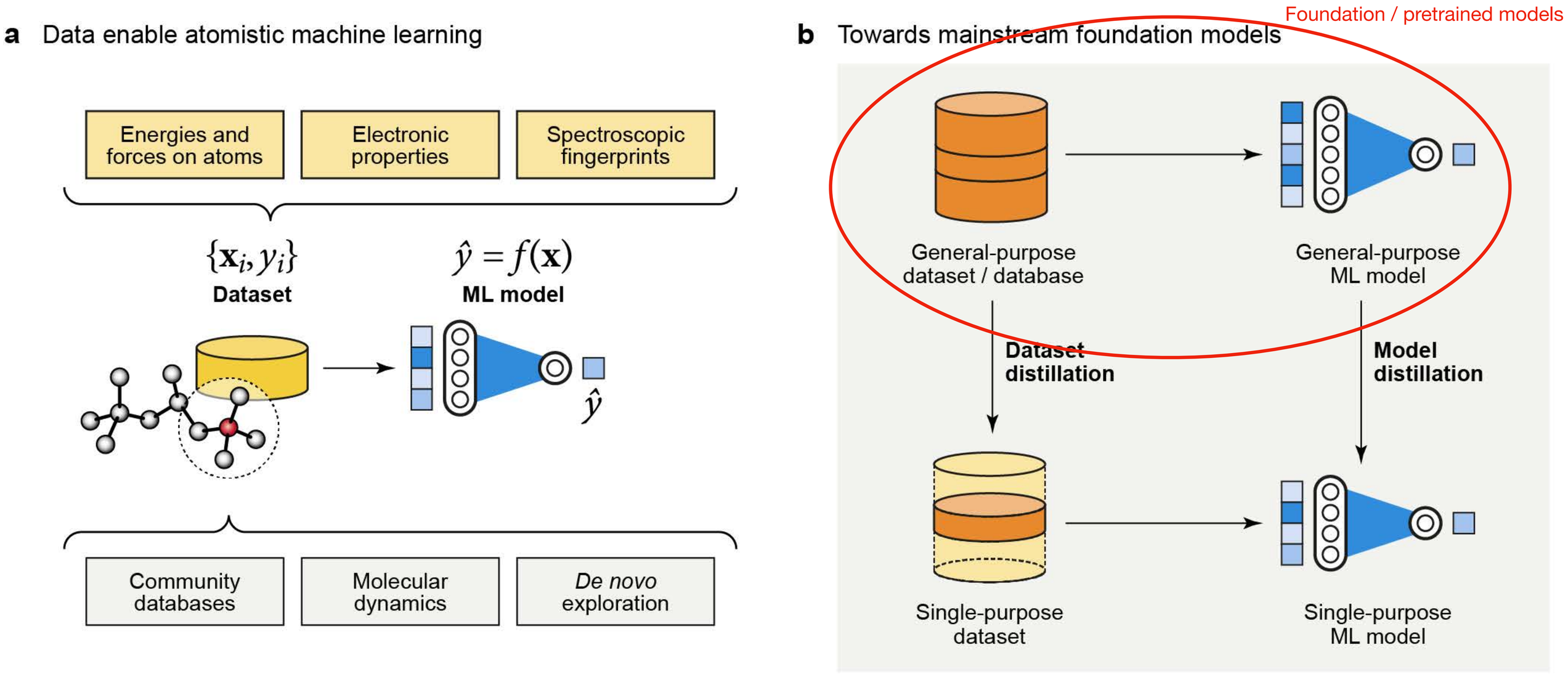
CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Towards general-purpose models/datasets



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Towards general-purpose models/datasets



CBM, Gardner, J.L.A., Deringer, V.L., 2024.
Data as the next challenge in atomistic machine learning. Nat Comput Sci 4, 384-387

Example of a pretrained graph neural networks

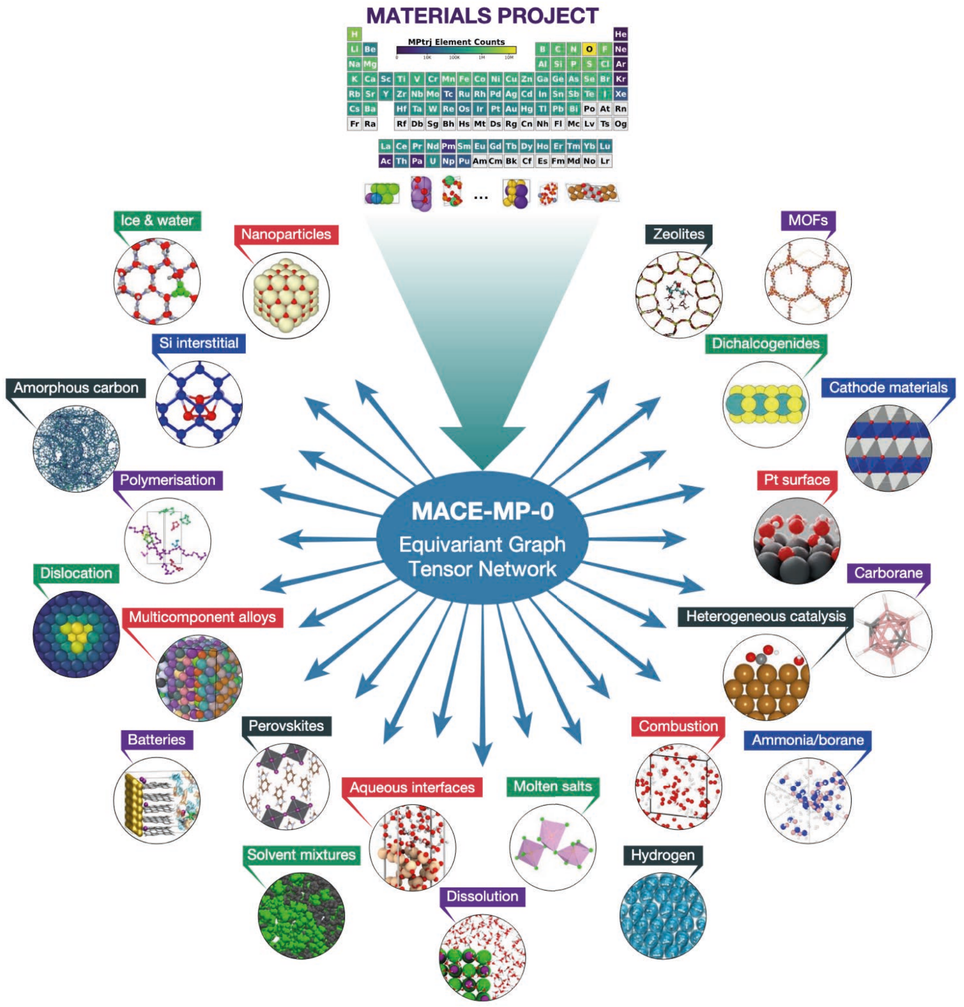


Figure 1: **A foundation model for materials modelling.** Trained only on Materials Project data (*19*) which consists primarily of inorganic crystals and is skewed heavily towards oxides, MACE-MP-0 is capable of molecular dynamics simulation across a wide variety of chemistries in the solid, liquid and gaseous phases.

Batatia et al., 2023

Example of a pretrained graph neural networks

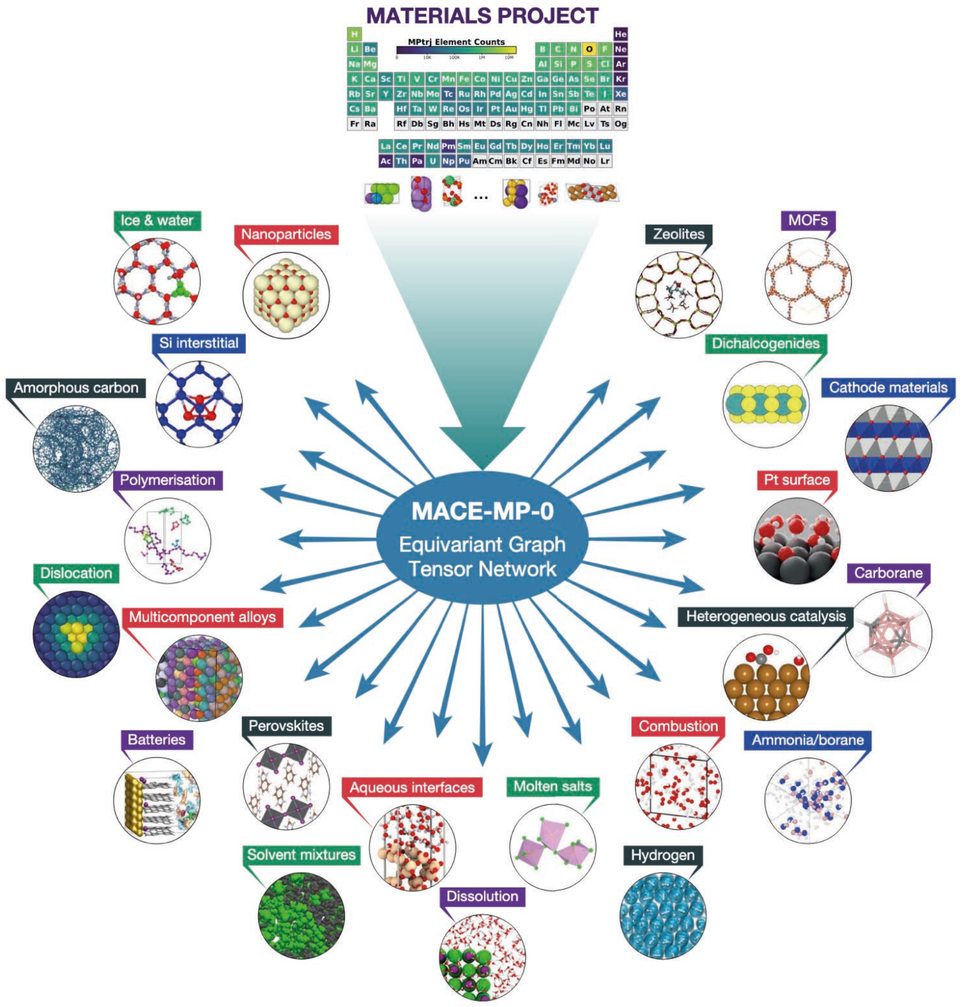


Figure 1: **A foundation model for materials modelling.** Trained only on Materials Project data (*19*) which consists primarily of inorganic crystals and is skewed heavily towards oxides, MACE-MP-0 is capable of molecular dynamics simulation across a wide variety of chemistries in the solid, liquid and gaseous phases.

Batatia et al., 2023

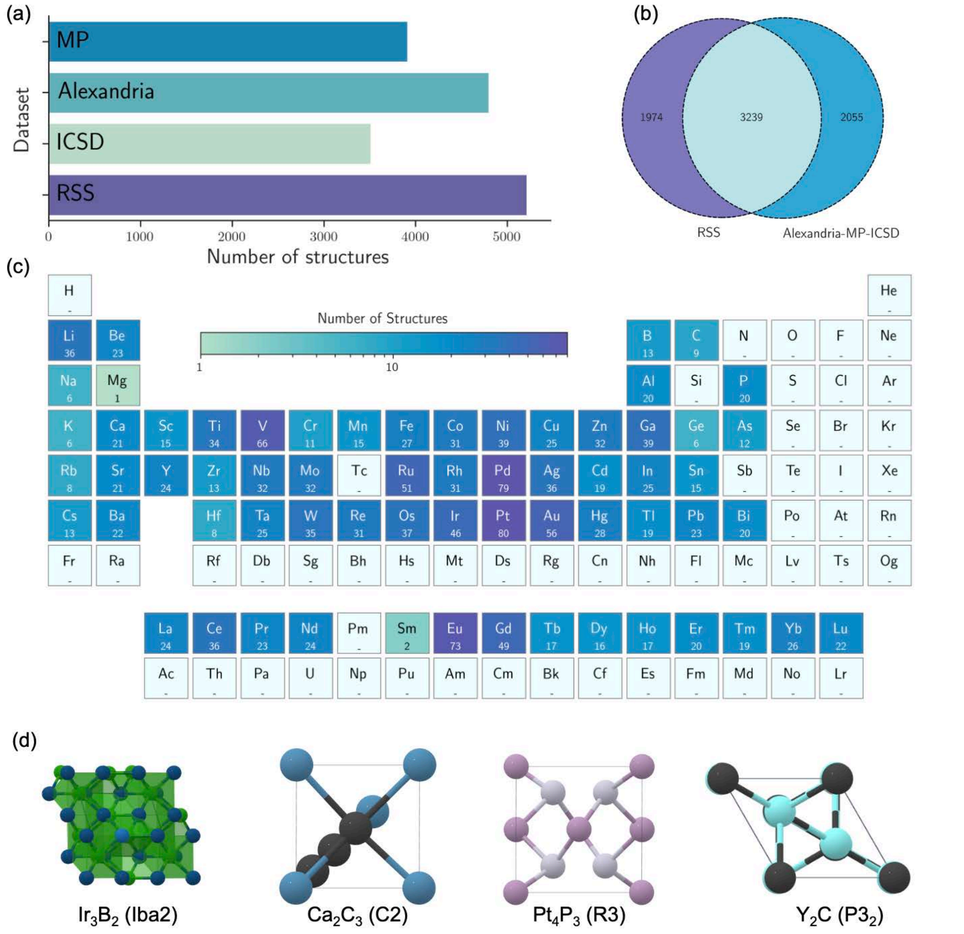


Fig. 3: **MatterSim as a zero-shot emulator empowering materials discovery. (a)** and **(b)** are the contribution of each dataset to the combined convex hull formed by `Alexandria-MP-ICSD` dataset (see text) and RSS-generated materials; **(c)** Elementwise appearance distribution[37] of the 852 RSS-generated materials found be to on the combined convex hull formed by the `Alexandria-MP-ICSD` and RSS-generated materials. The materials containing H, Si, N, Sb, O, S, Se, Te, F, Cl, Br, I are removed due to potential issue with how anion corrections are implemented in Materials Project when applied to hypothetical materials[38]. **(d)** exhibits examples of materials found to be lower than the `Alexandria-MP-ICSD` hull, with the corresponding space group in the parentheses.

Yang et al., 2024

Example of a pretrained graph neural networks

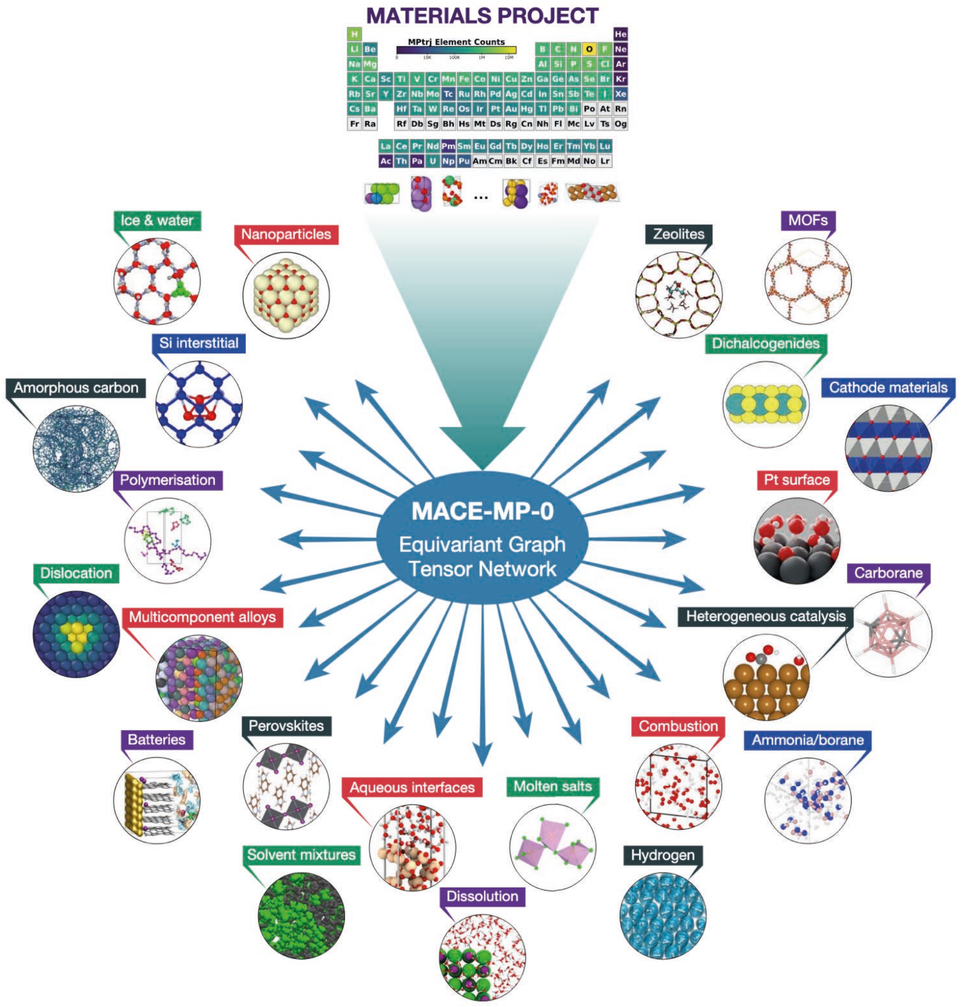


Figure 1: **A foundation model for materials modelling.** Trained only on Materials Project data (*19*) which consists primarily of inorganic crystals and is skewed heavily towards oxides, MACE-MP-0 is capable of molecular dynamics simulation across a wide variety of chemistries in the solid, liquid and gaseous phases.

Batatia et al., 2023

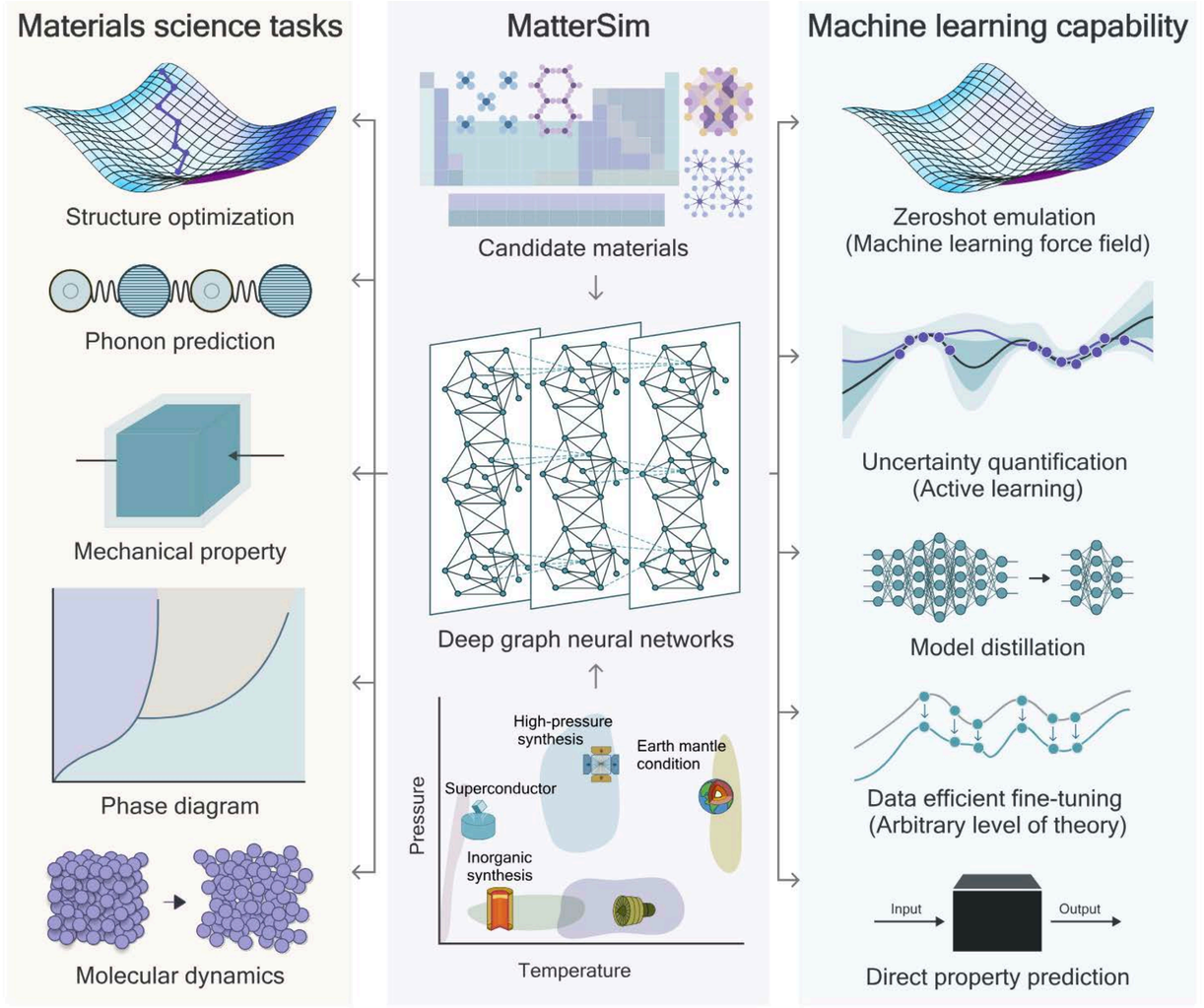
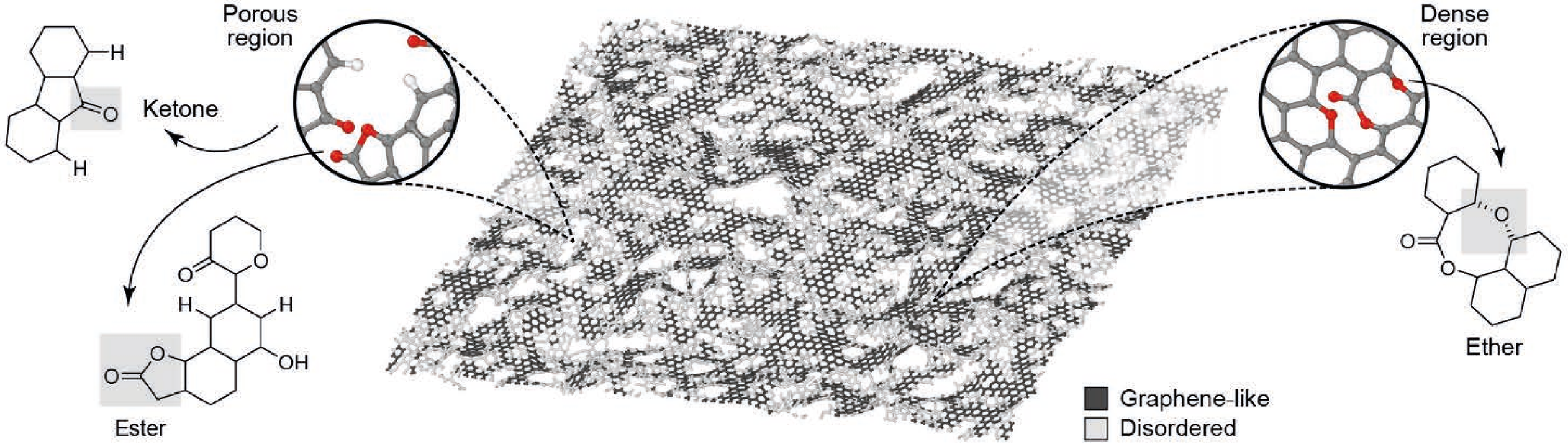


Fig. 1: MatterSim is a deep learning atomistic model for predicting materials properties with high predictive accuracy across chemical elements, temperatures and pressures, enabling a wide range of applicability and functionality.

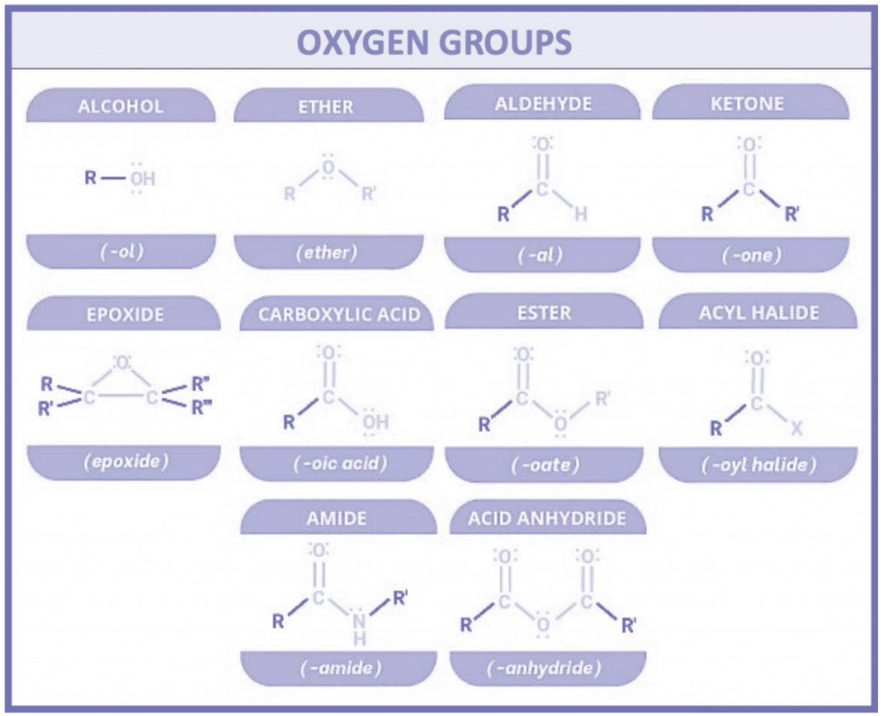
Yang et al., 2024

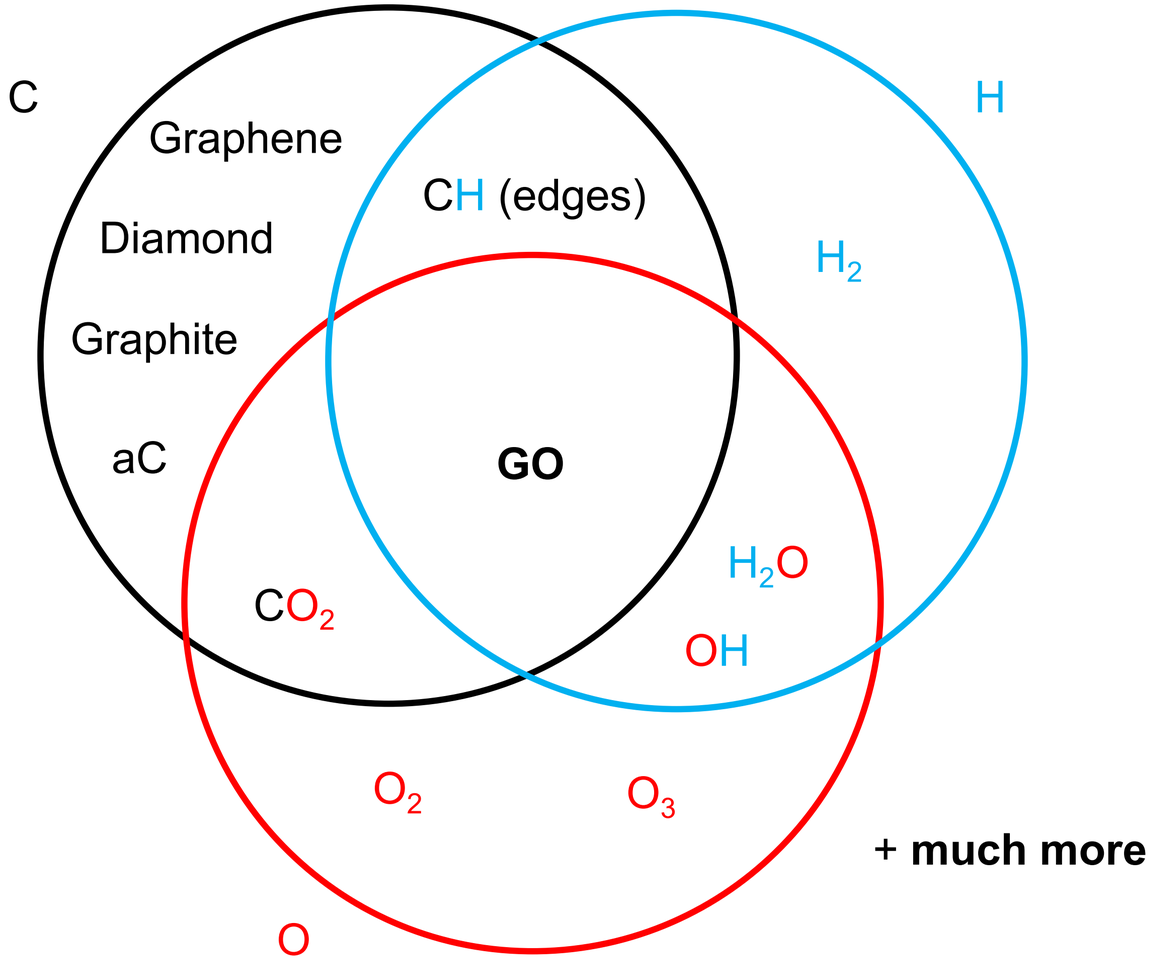
Graphene oxide as a basis for transferable MLIPs



El-Machachi et al., 2024

The structure of reduced GO





Courtesy of Z. El-Machachi

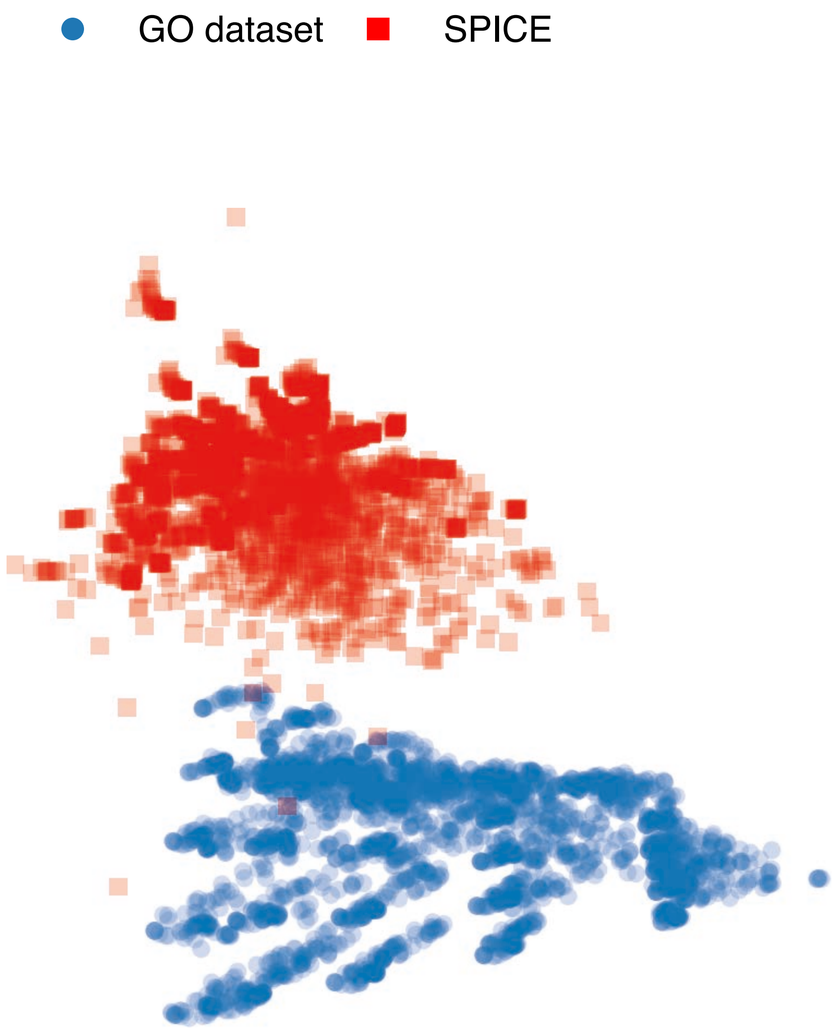
Objectives

- Not to train an MLIP from scratch (yes, it is painful!)
- Assess the out-of-domain performance of an MLIP trained on (some) chemistry: **GO-MACE-23**
- Quantify the role of data and architecture in out-of-domain performance
- Identify requirements for general-purpose MLIP targeting (organic) chemistry

GO vs molecules from GO-MACE-23 perspective

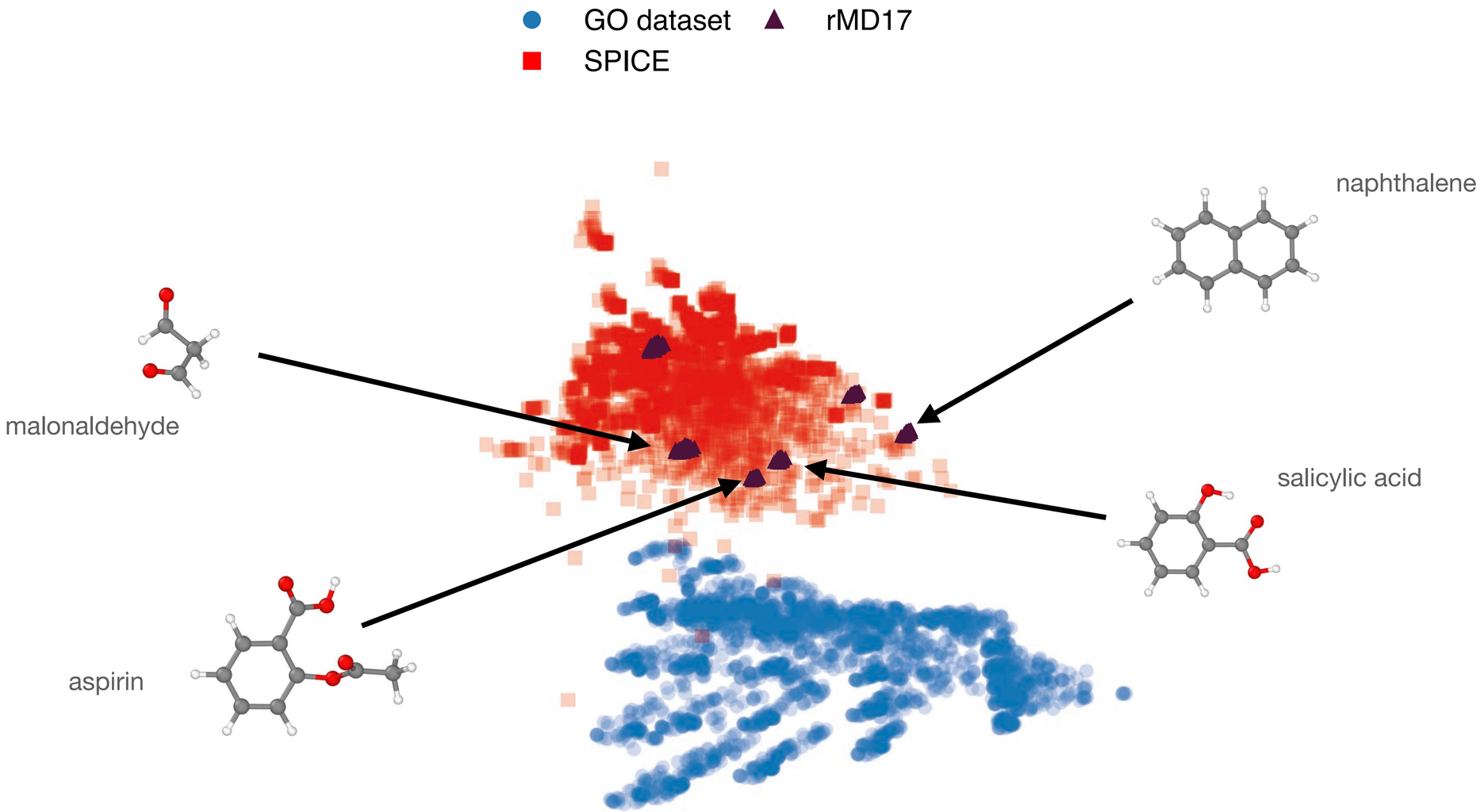
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

GO vs molecules from GO-MACE-23 perspective



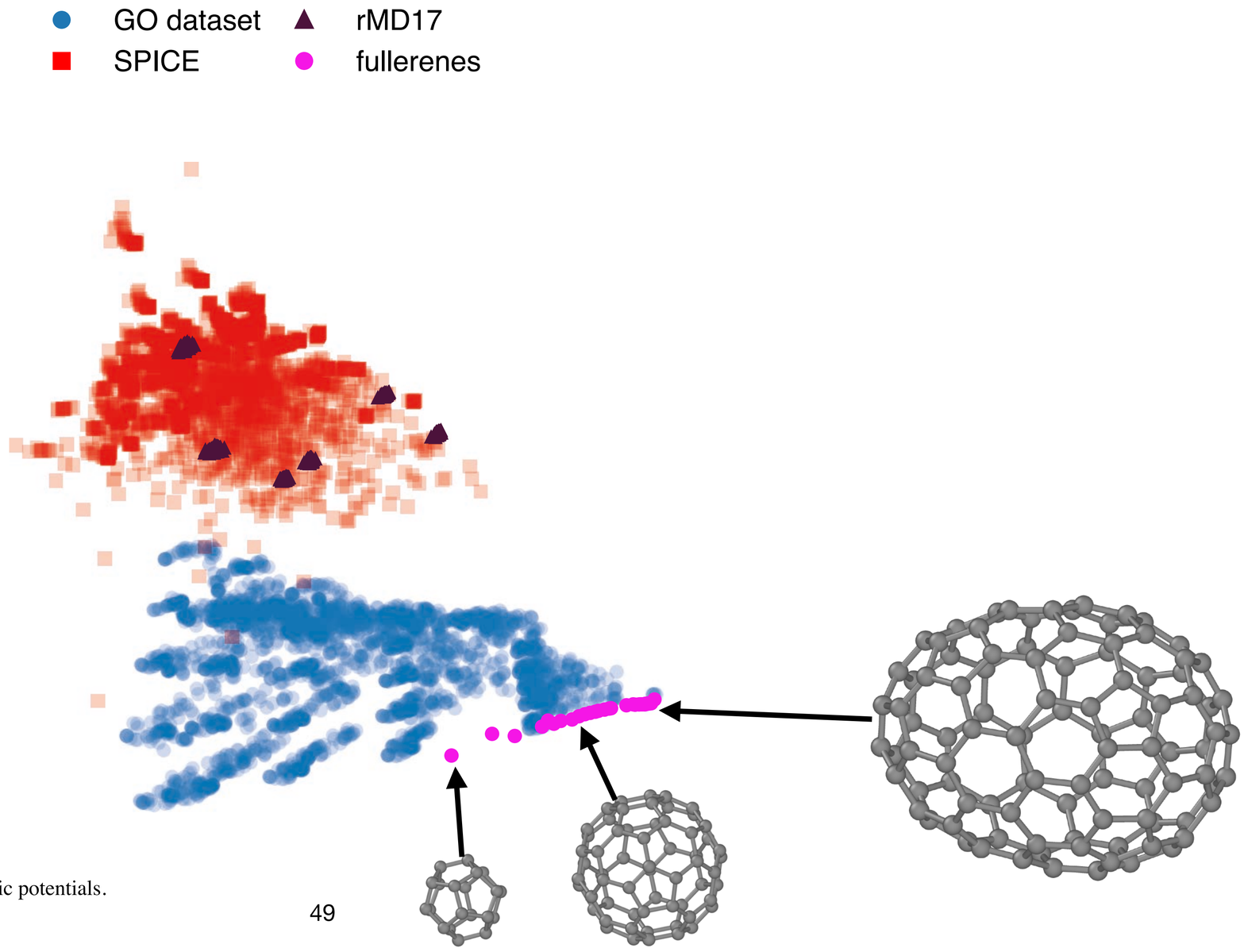
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

GO vs molecules from GO-MACE-23 perspective



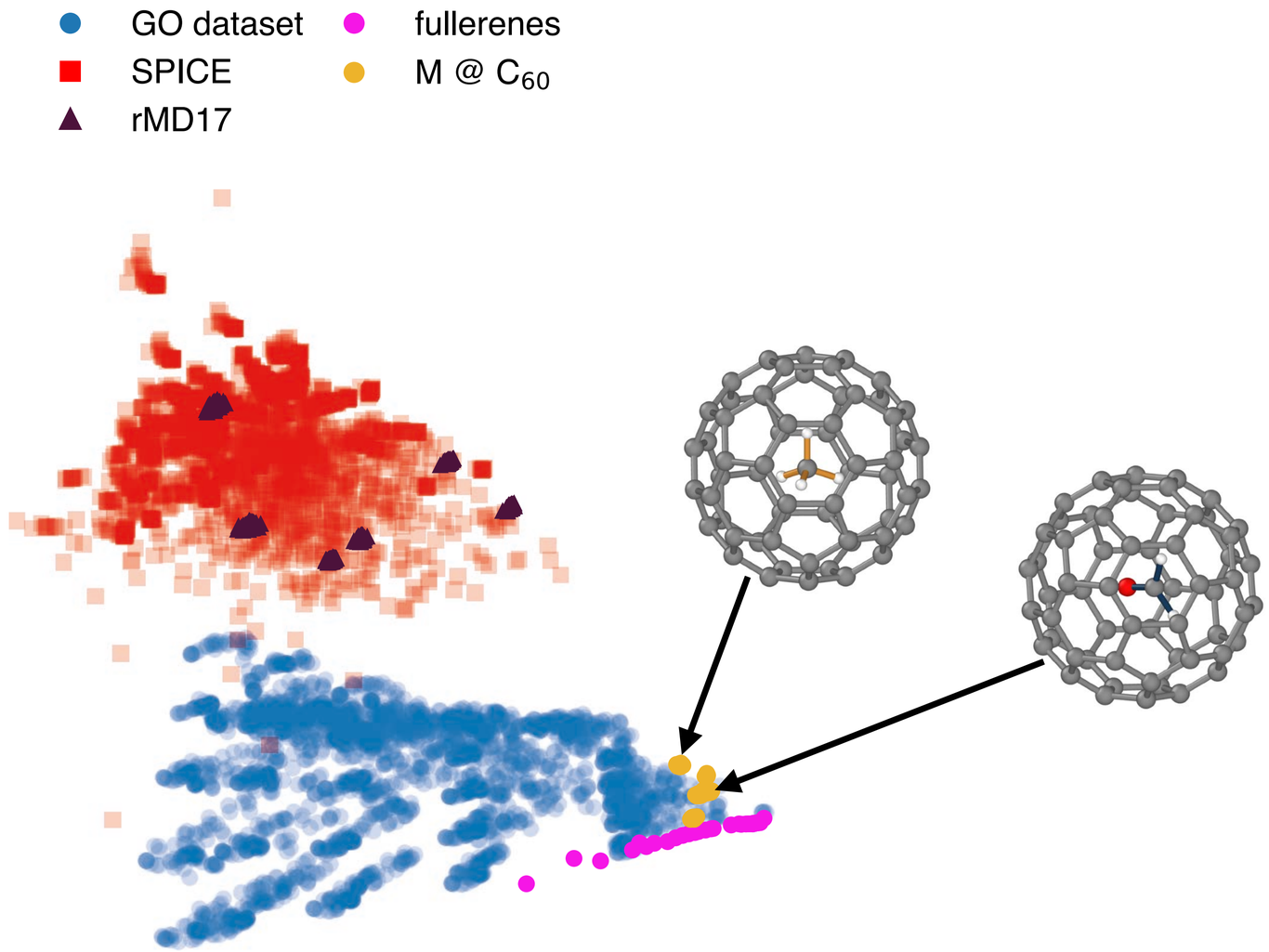
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

GO vs molecules from GO-MACE-23 perspective



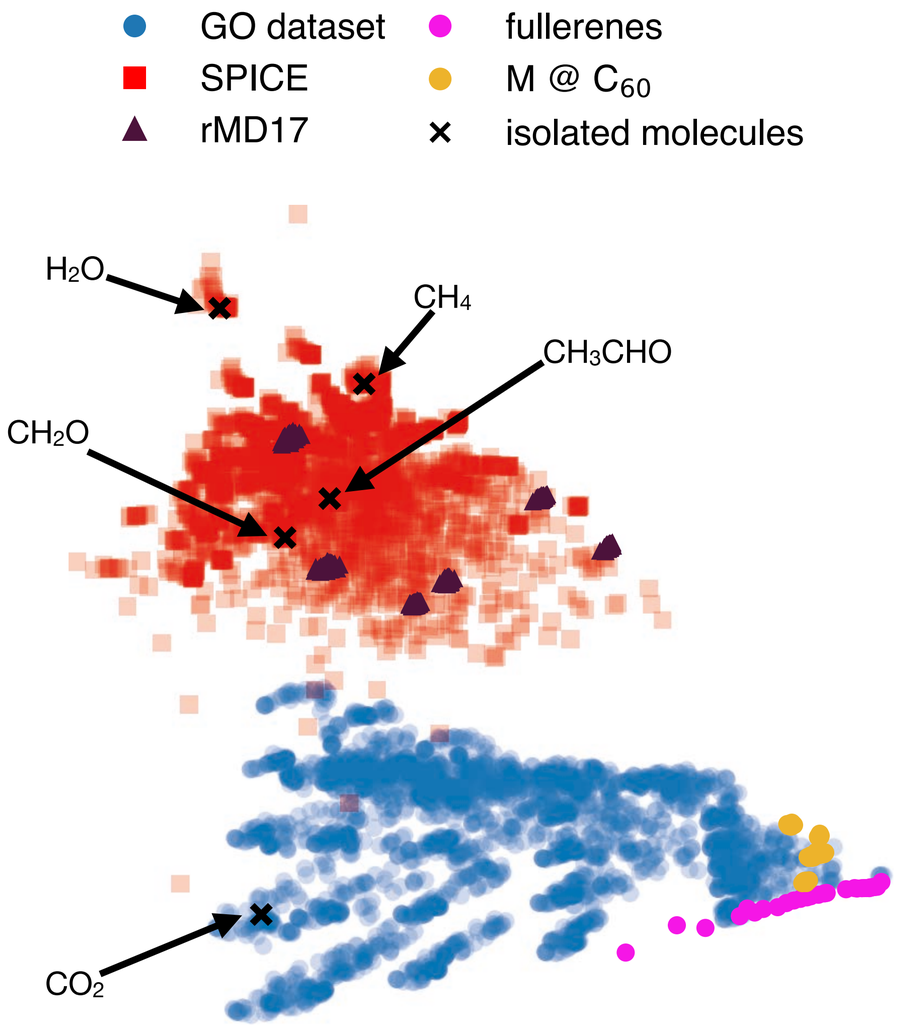
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

GO vs molecules from GO-MACE-23 perspective



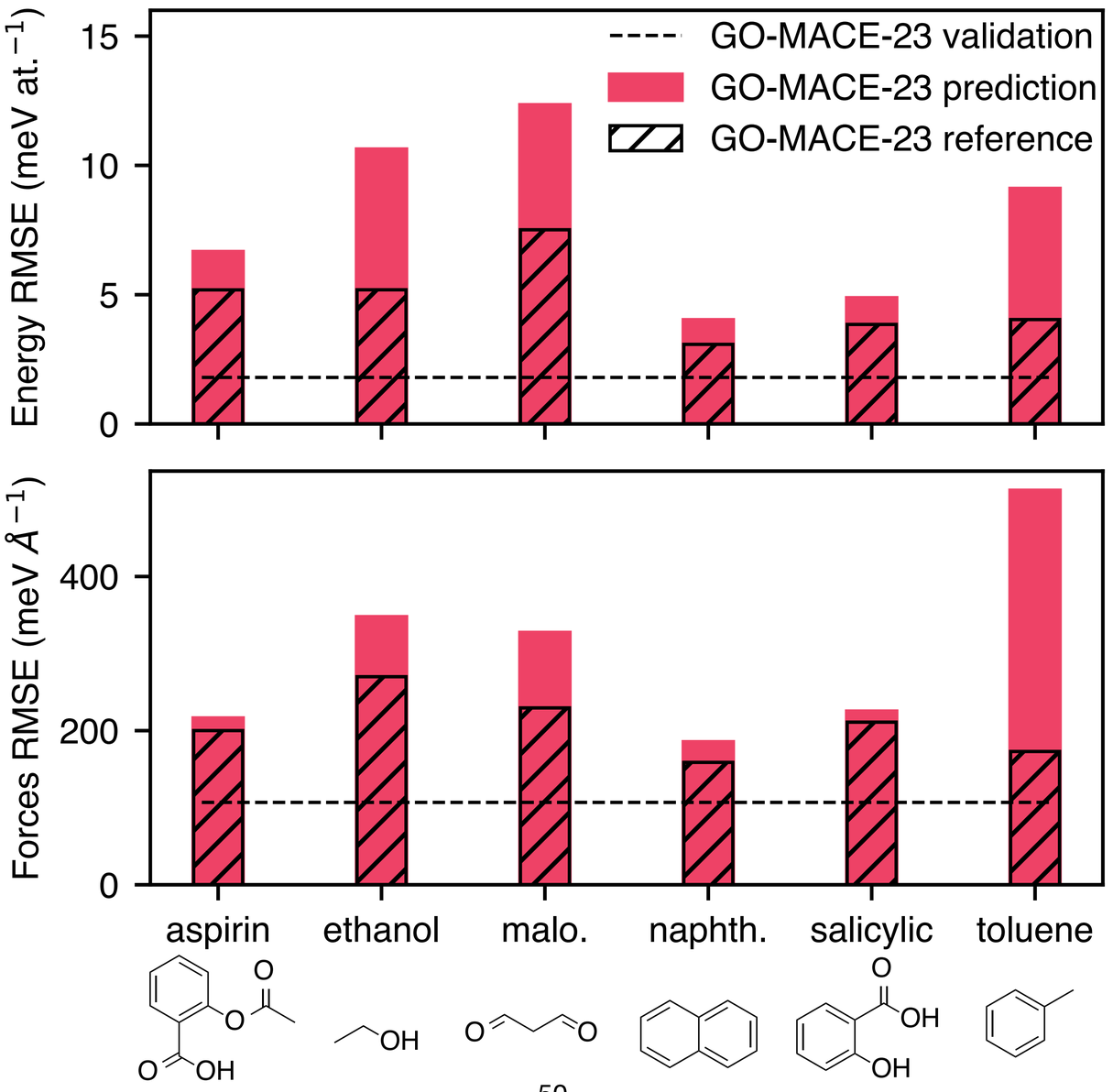
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

GO vs molecules from GO-MACE-23 perspective



CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

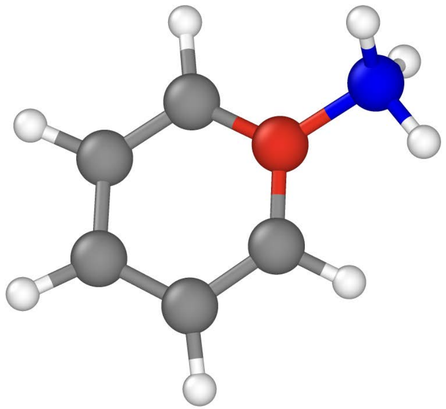
Performance on the rMD17 dataset



CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

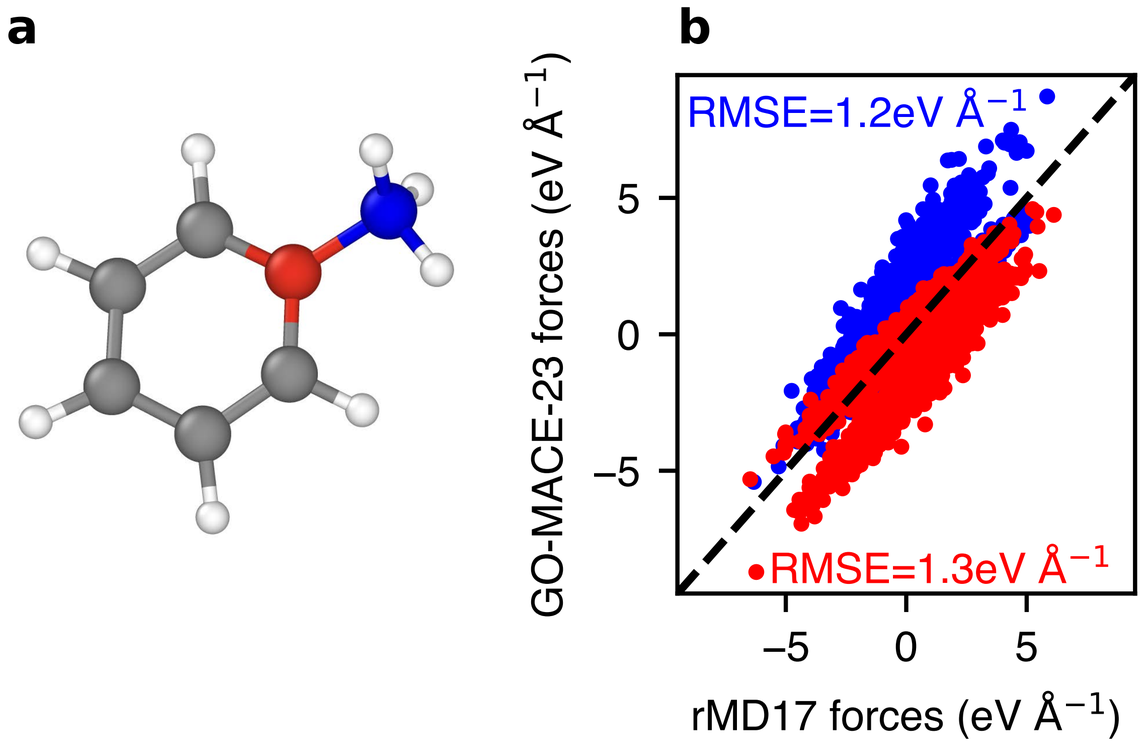
Why the toluene is unlearnable?

a



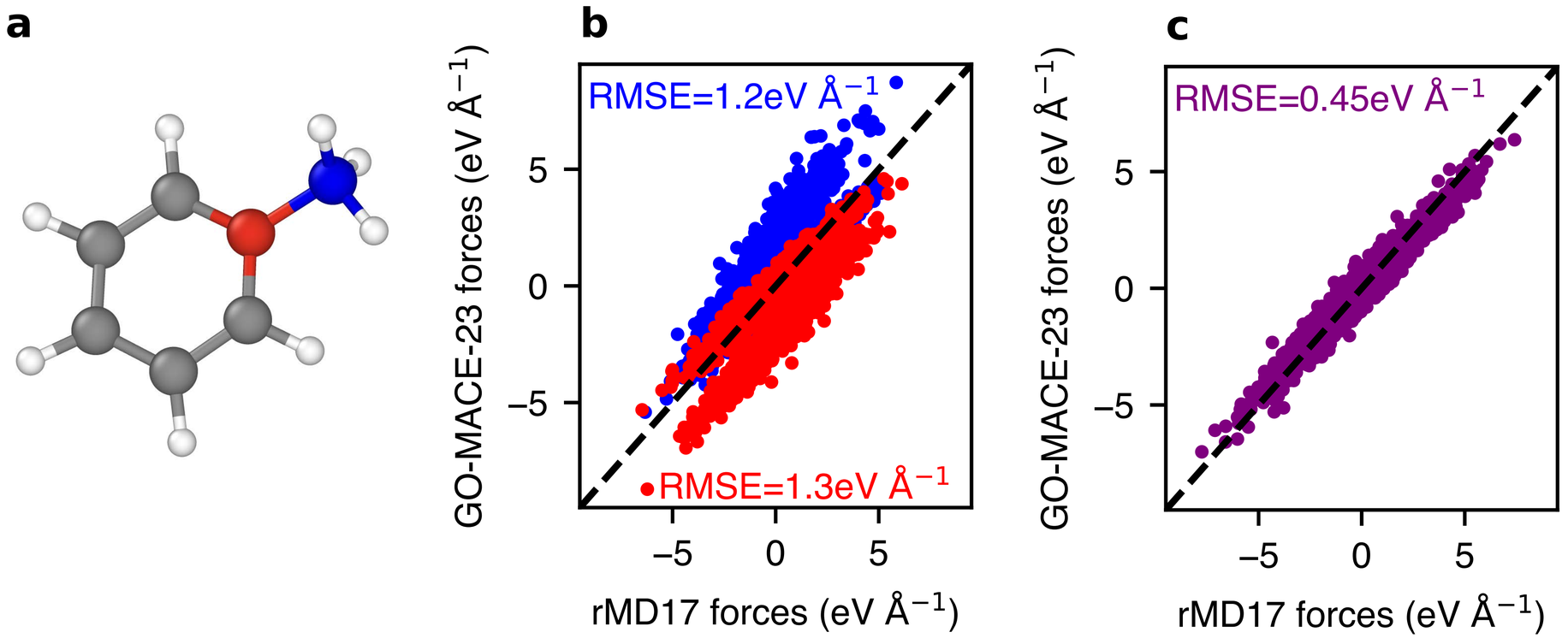
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025.
Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials.
arXiv: 2502.21317

Why the toluene is unlearnable?



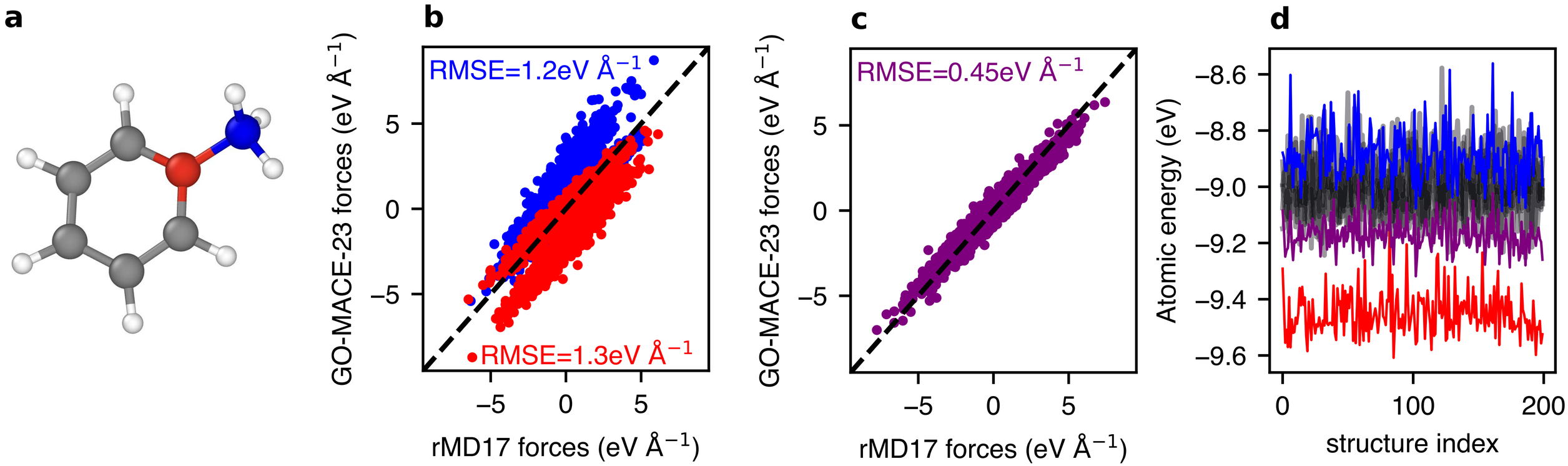
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

Why the toluene is unlearnable?



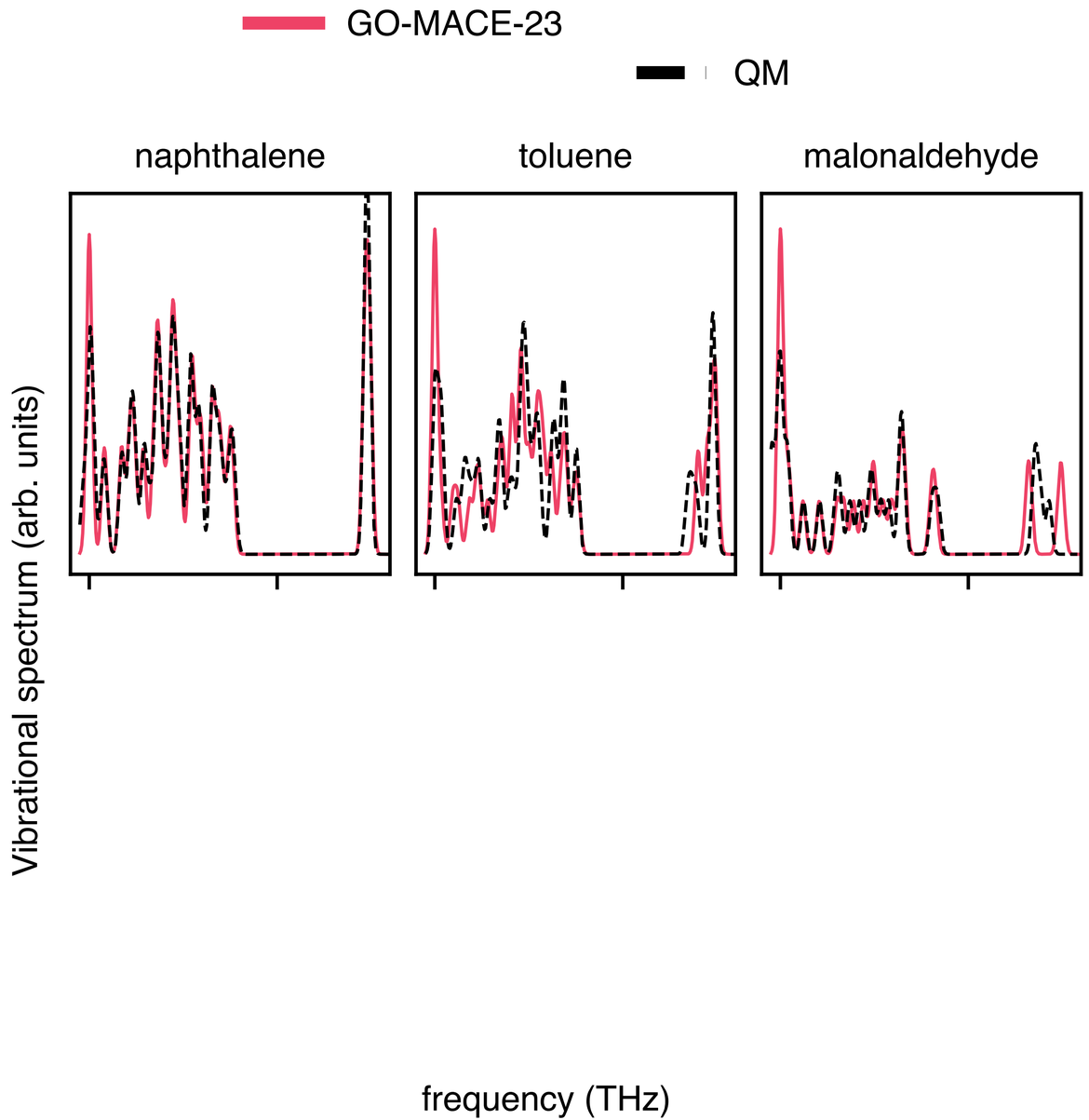
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

Why the toluene is unlearnable?

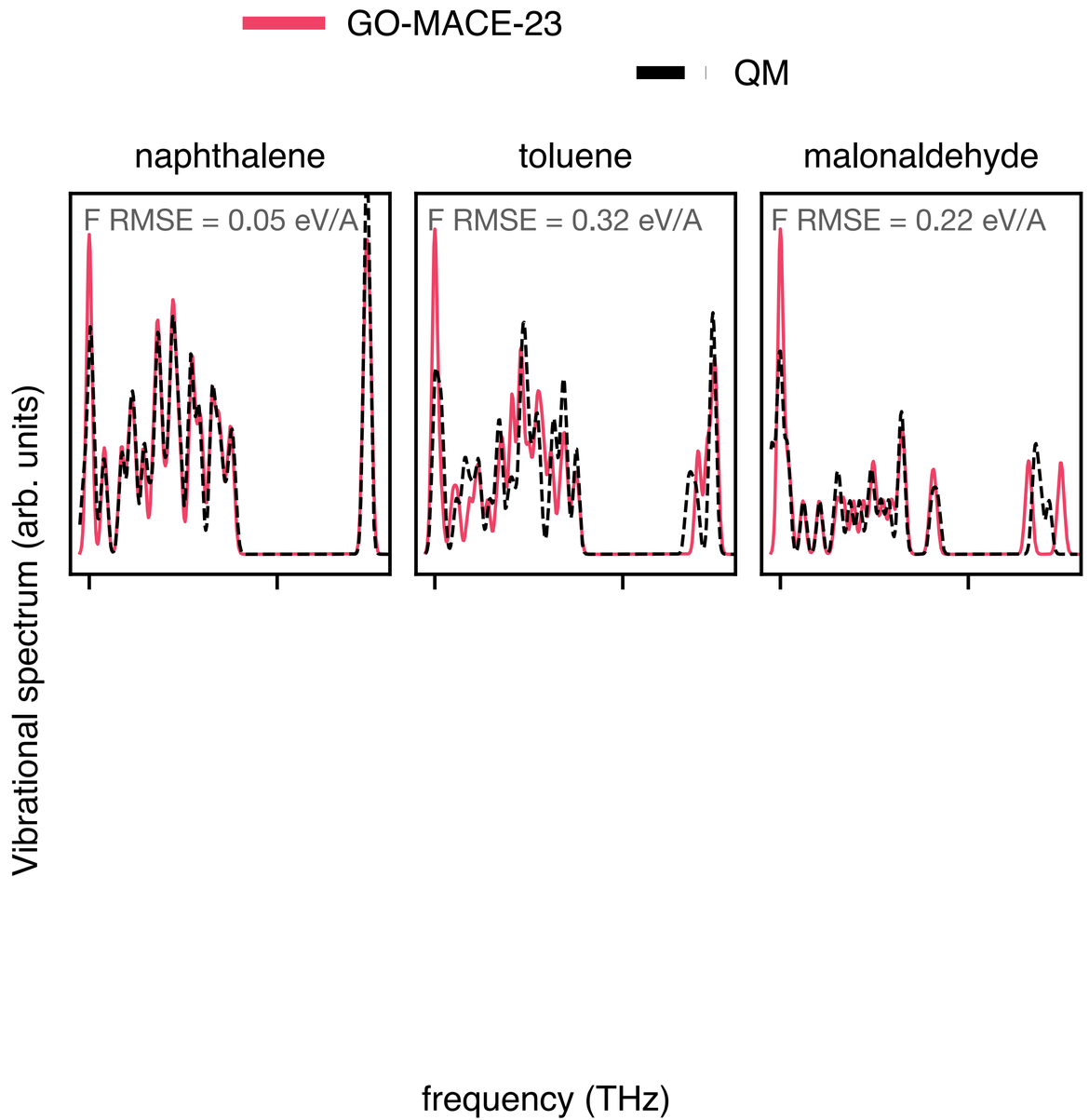


CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

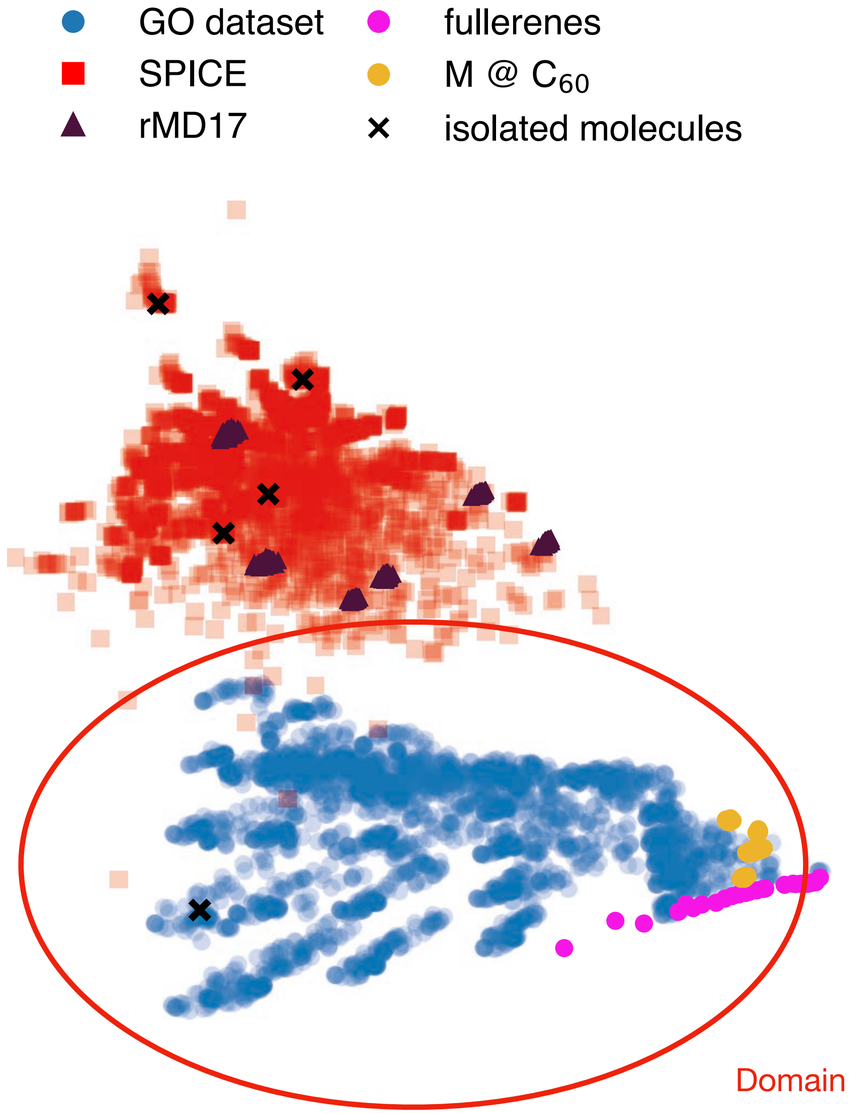
Vibrational spectra of molecules

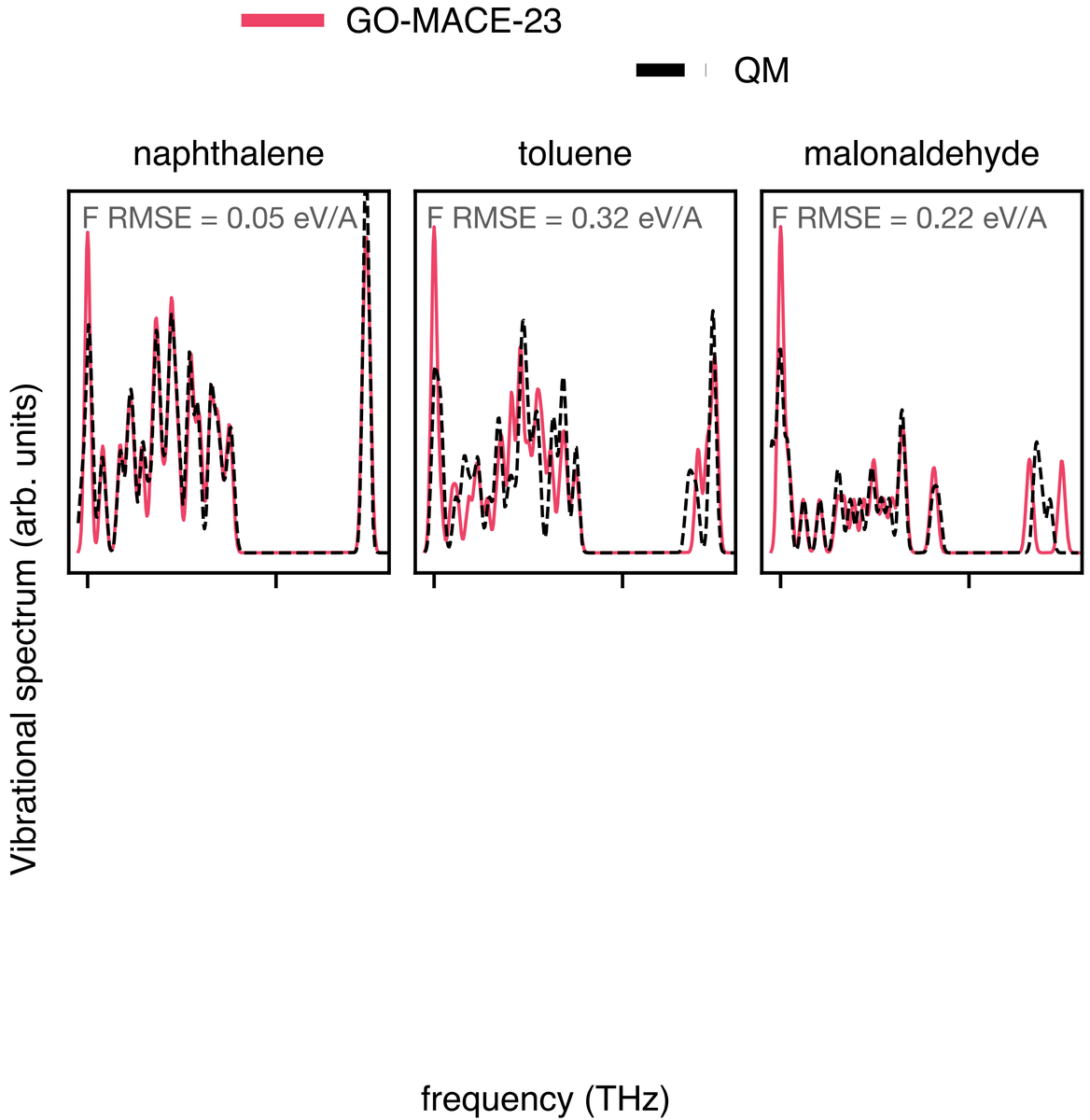


Vibrational spectra of molecules

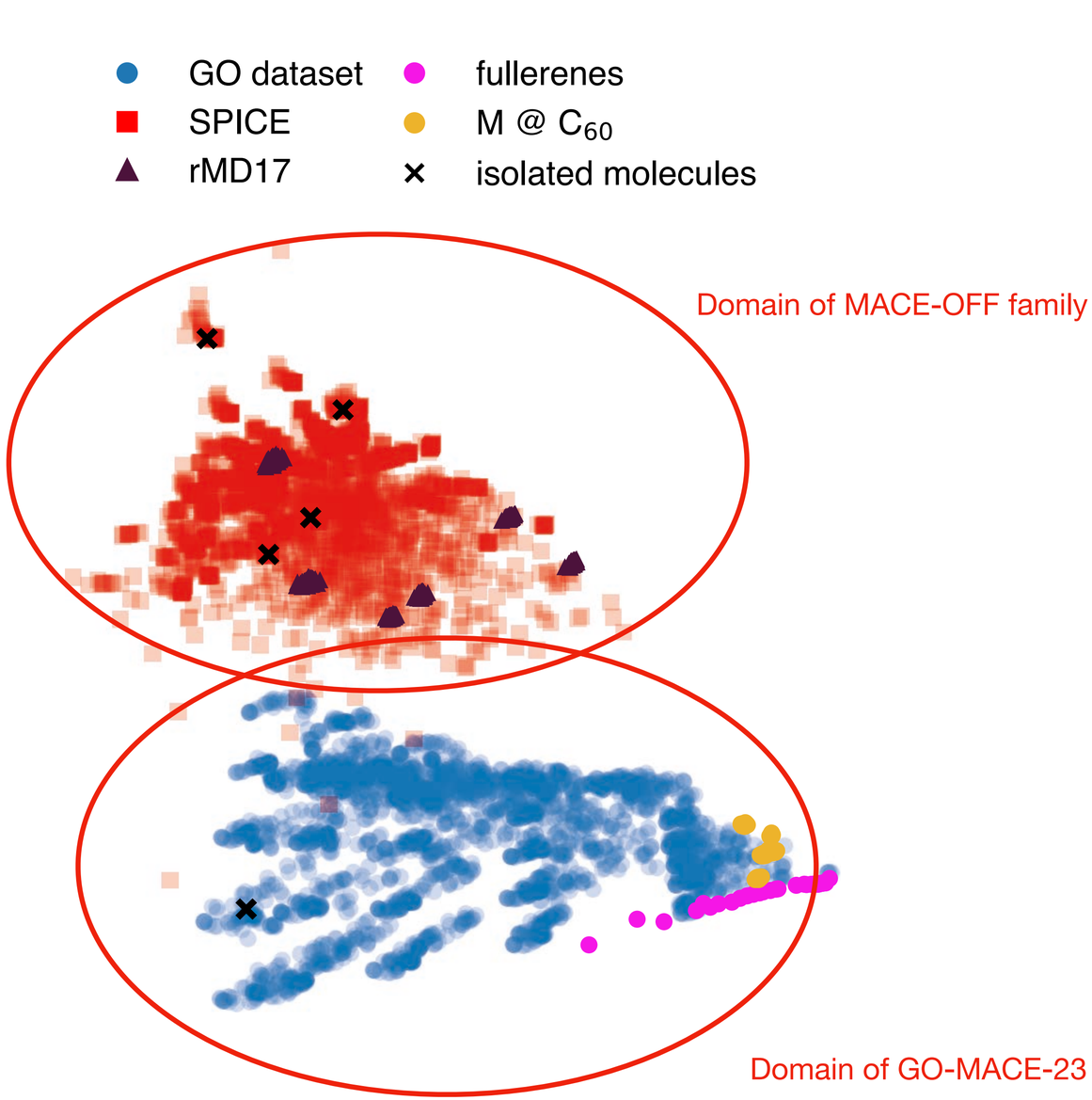


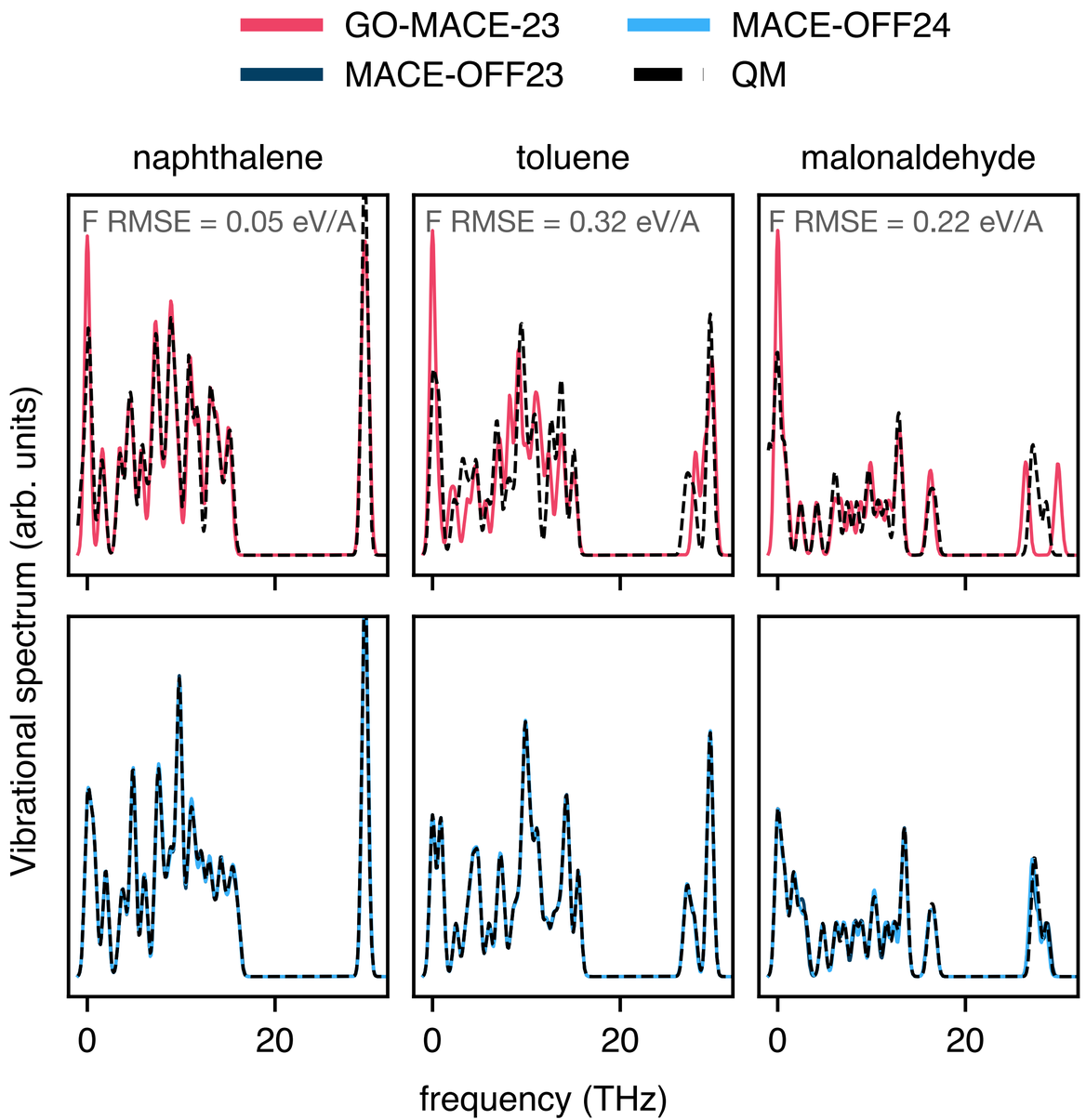
Vibrational spectra of molecules



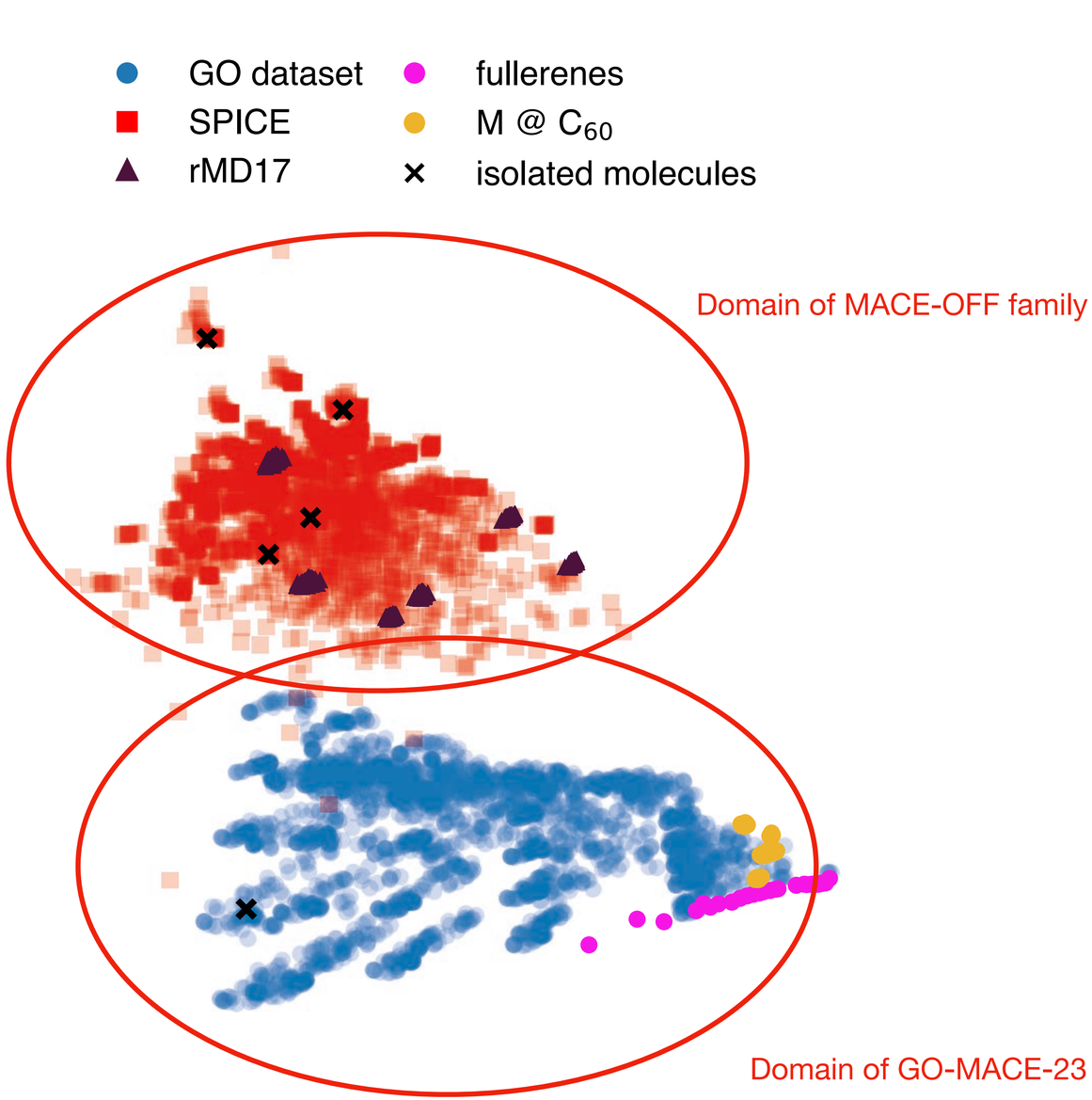


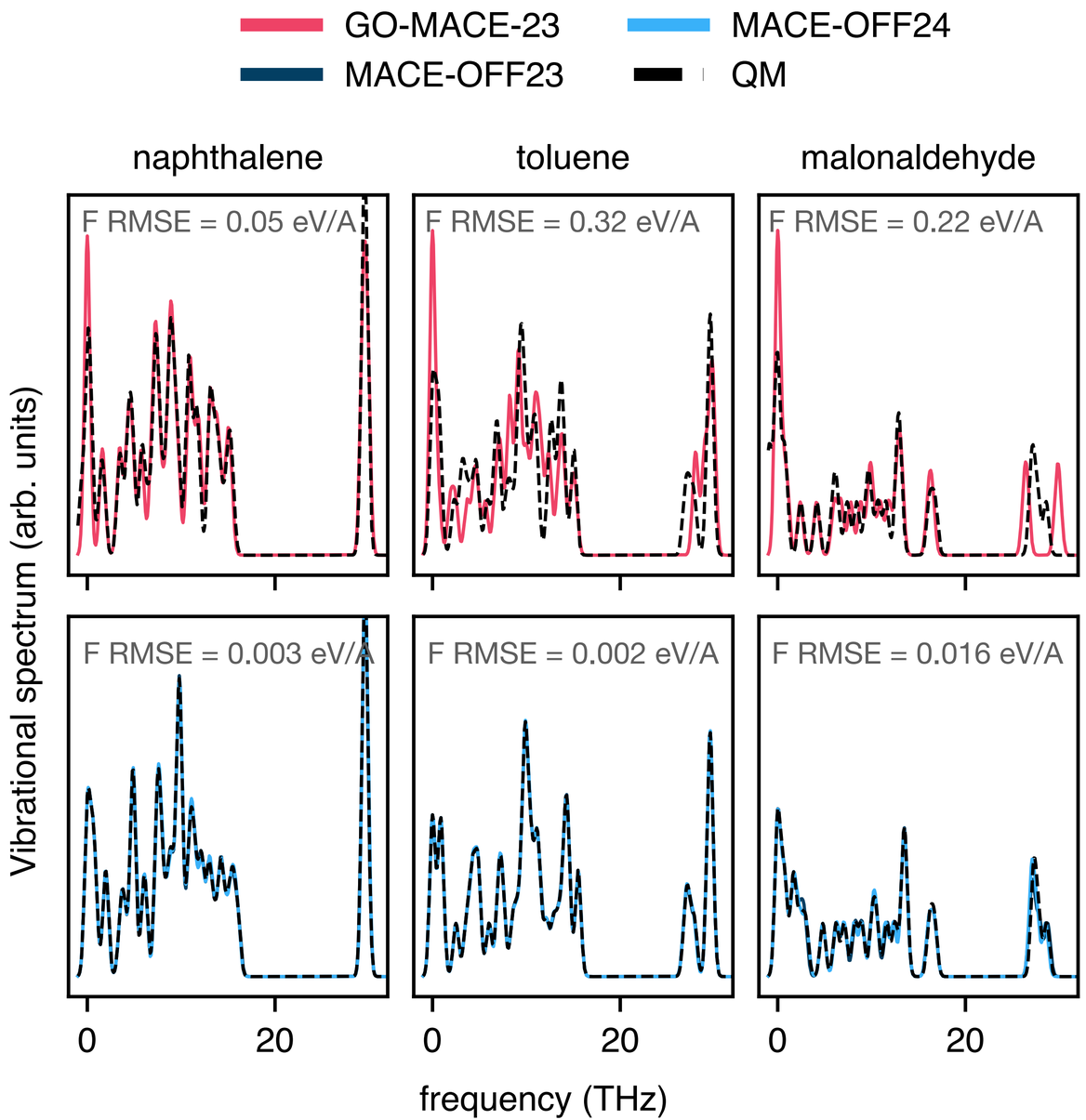
Vibrational spectra of molecules



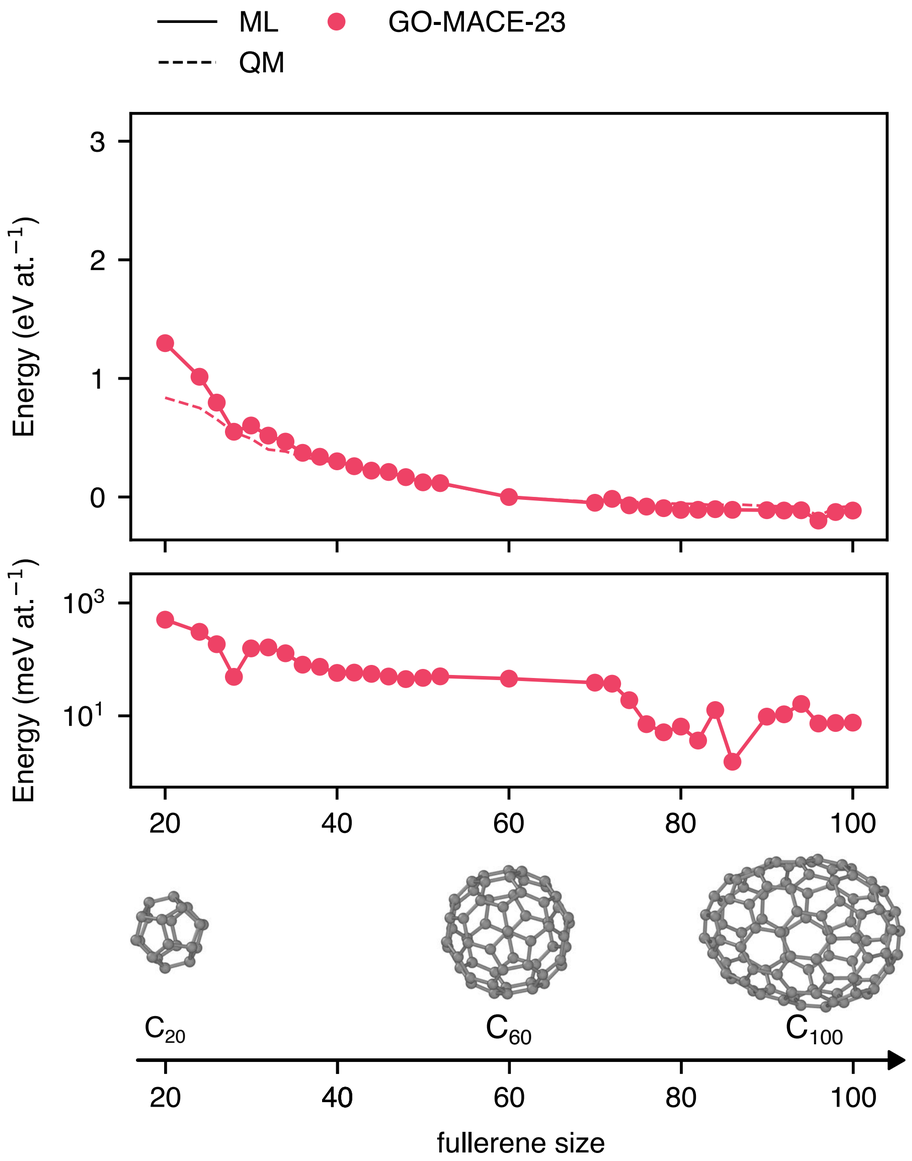


Vibrational spectra of molecules



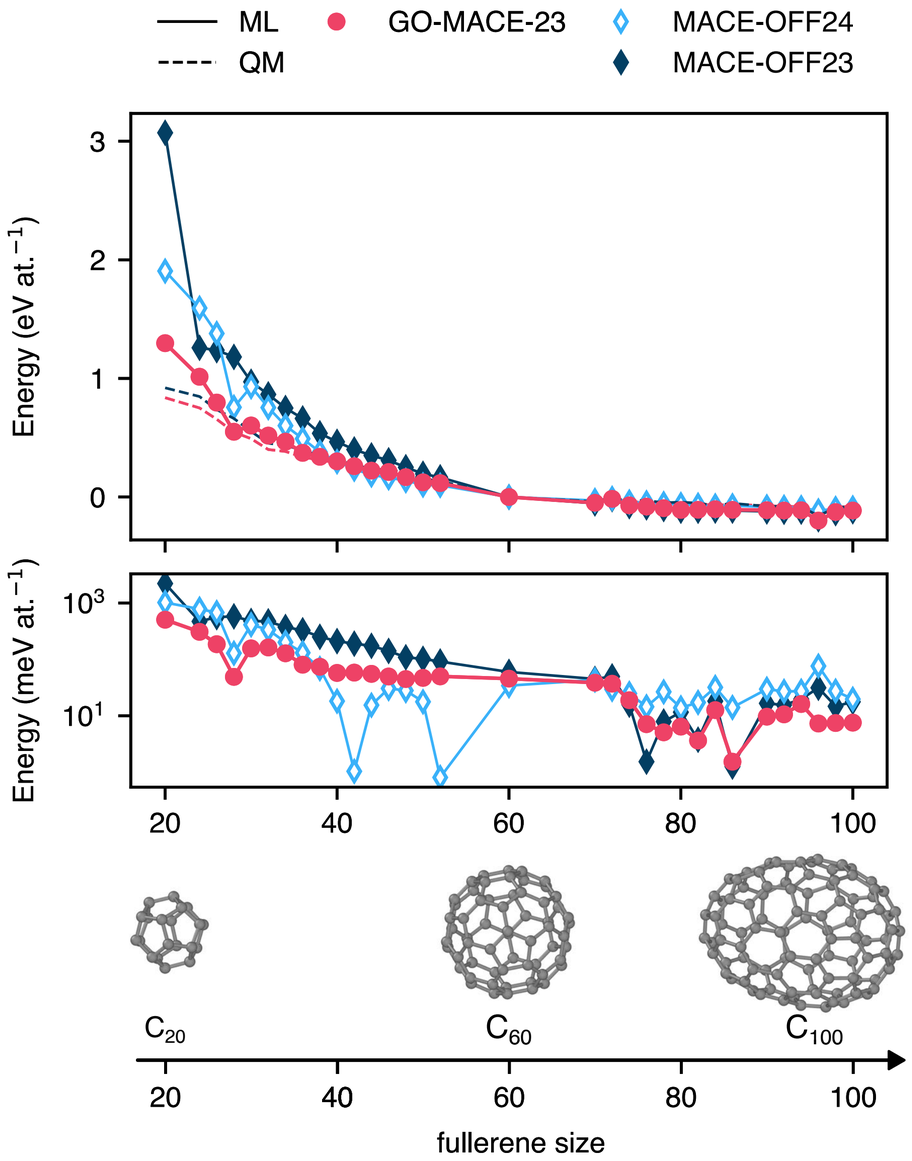


Fullerenes



CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

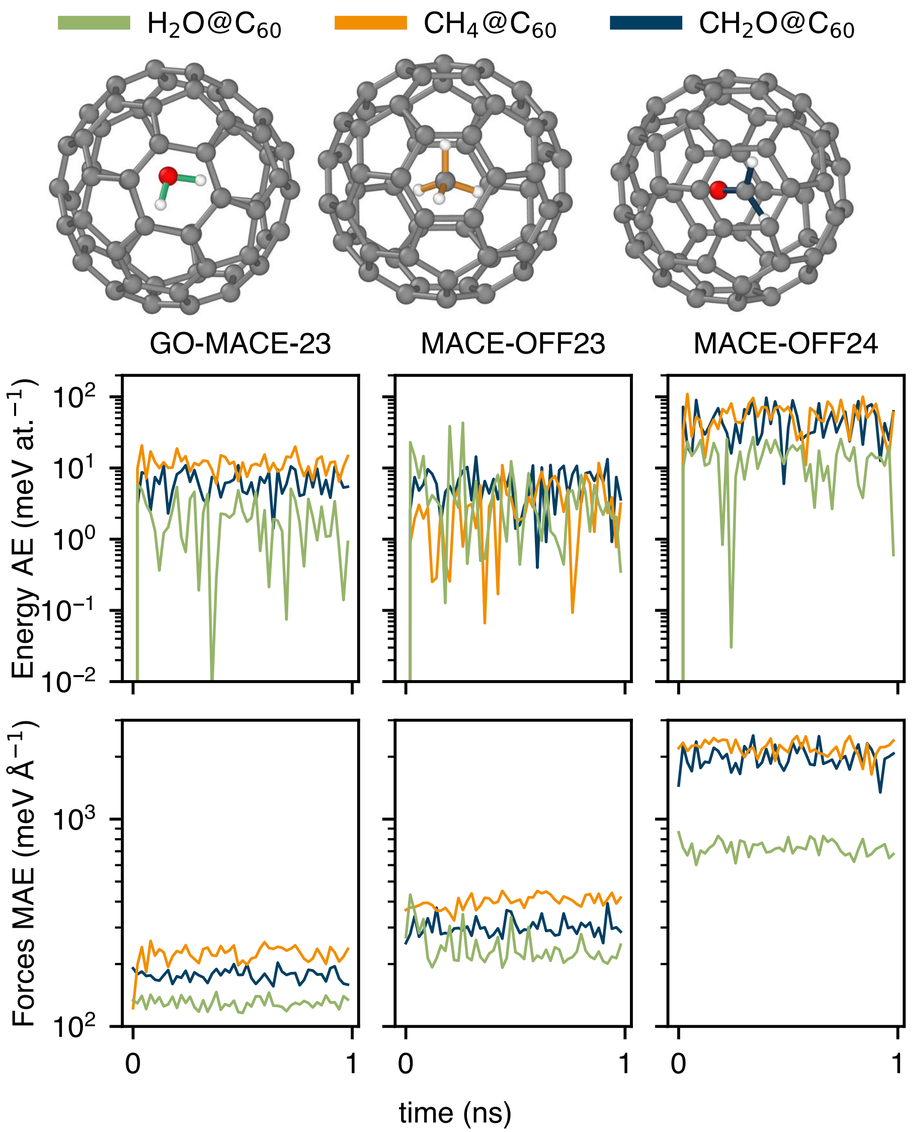
Fullerenes



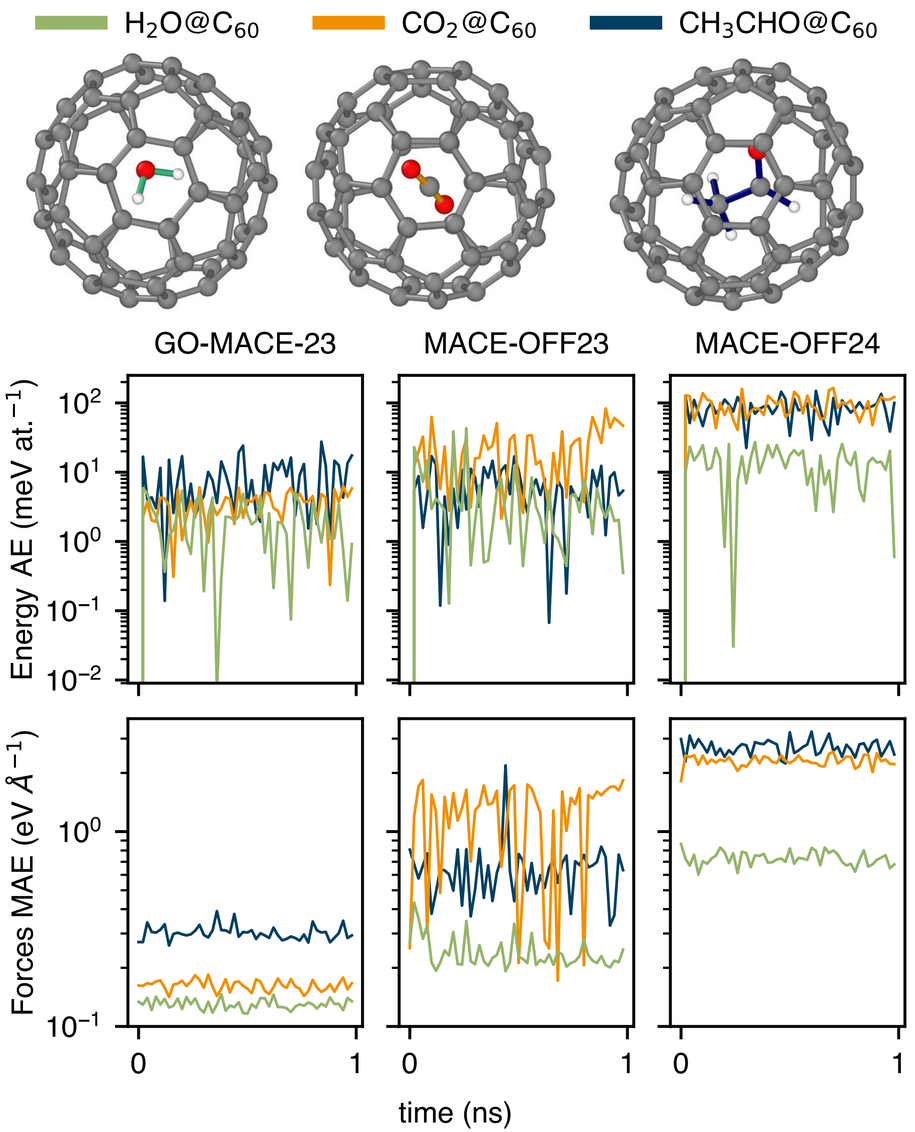
CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

Encapsulated reactions

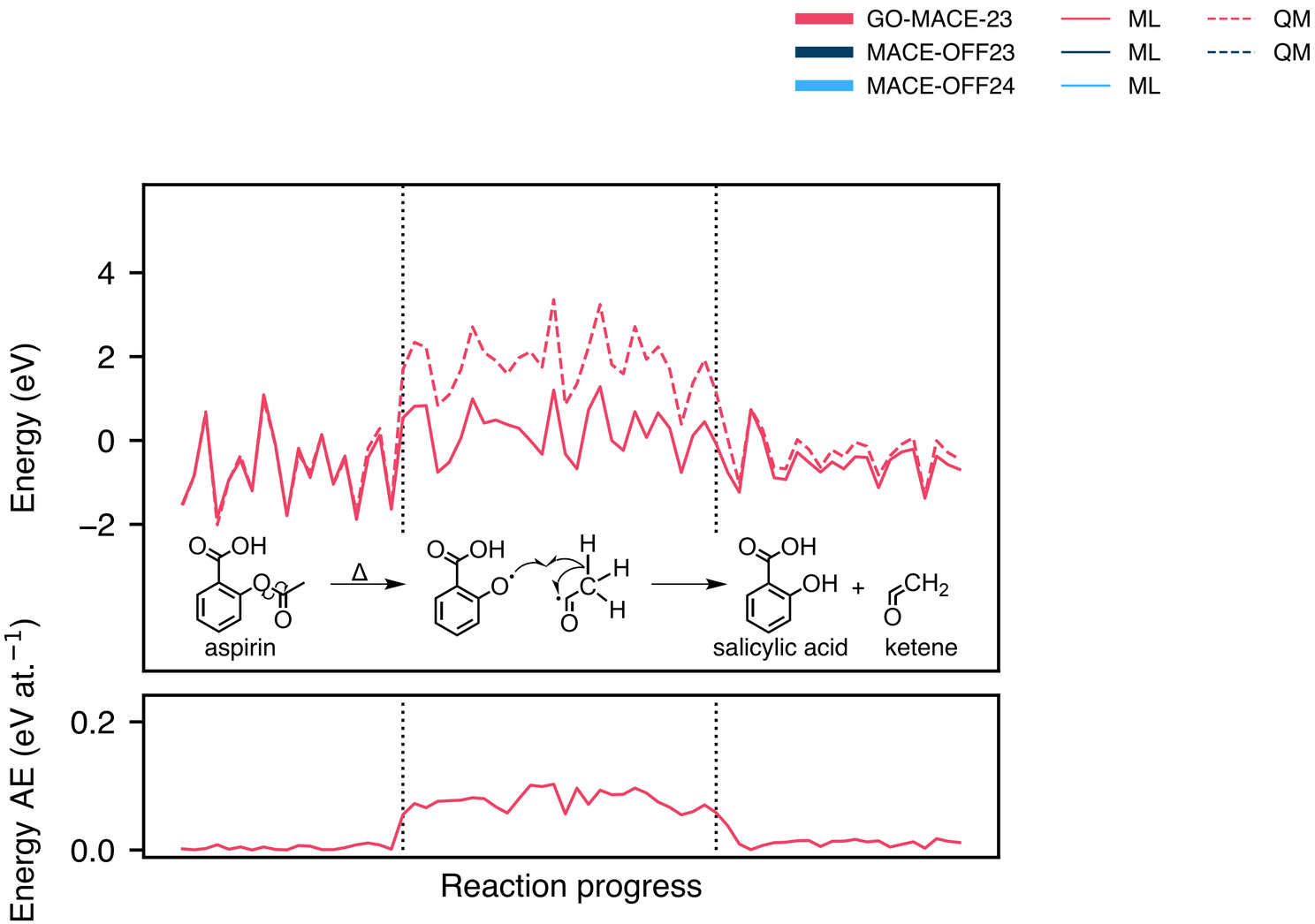
Reported in the literature



Yet to be reported in the literature

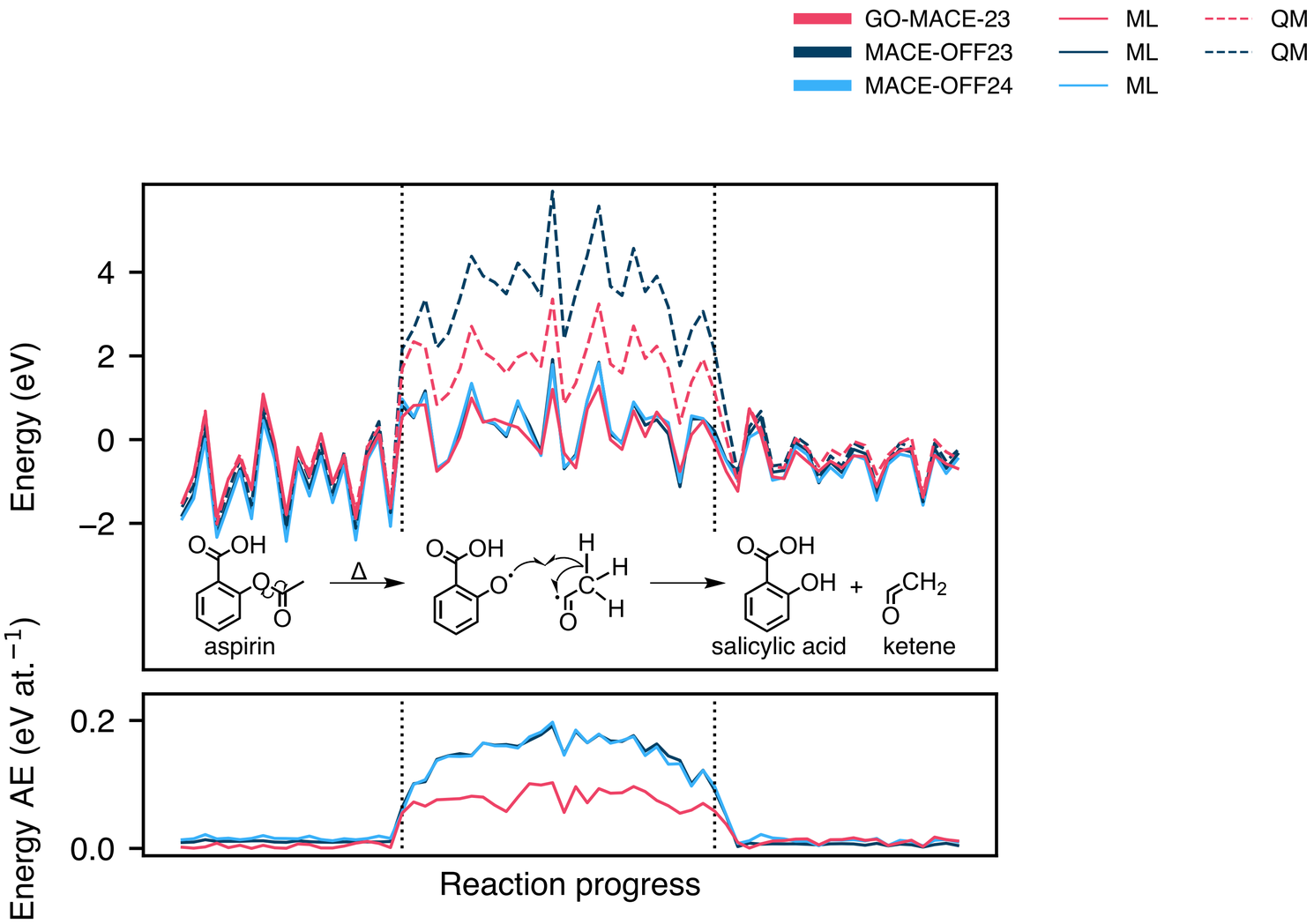


Transferability to reaction pathways



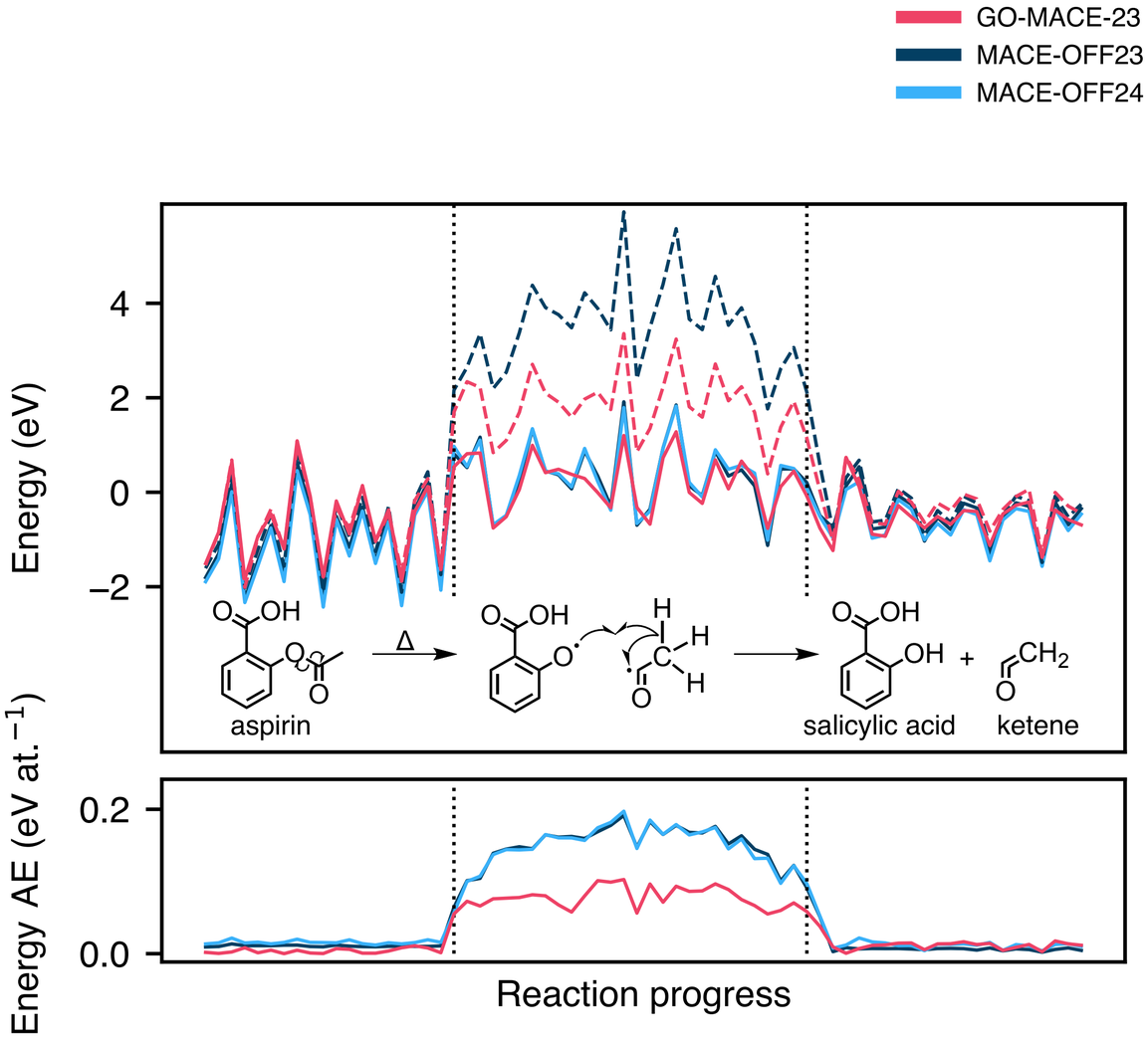
D. A. Nightingale, US Pat., 1604472, 1926

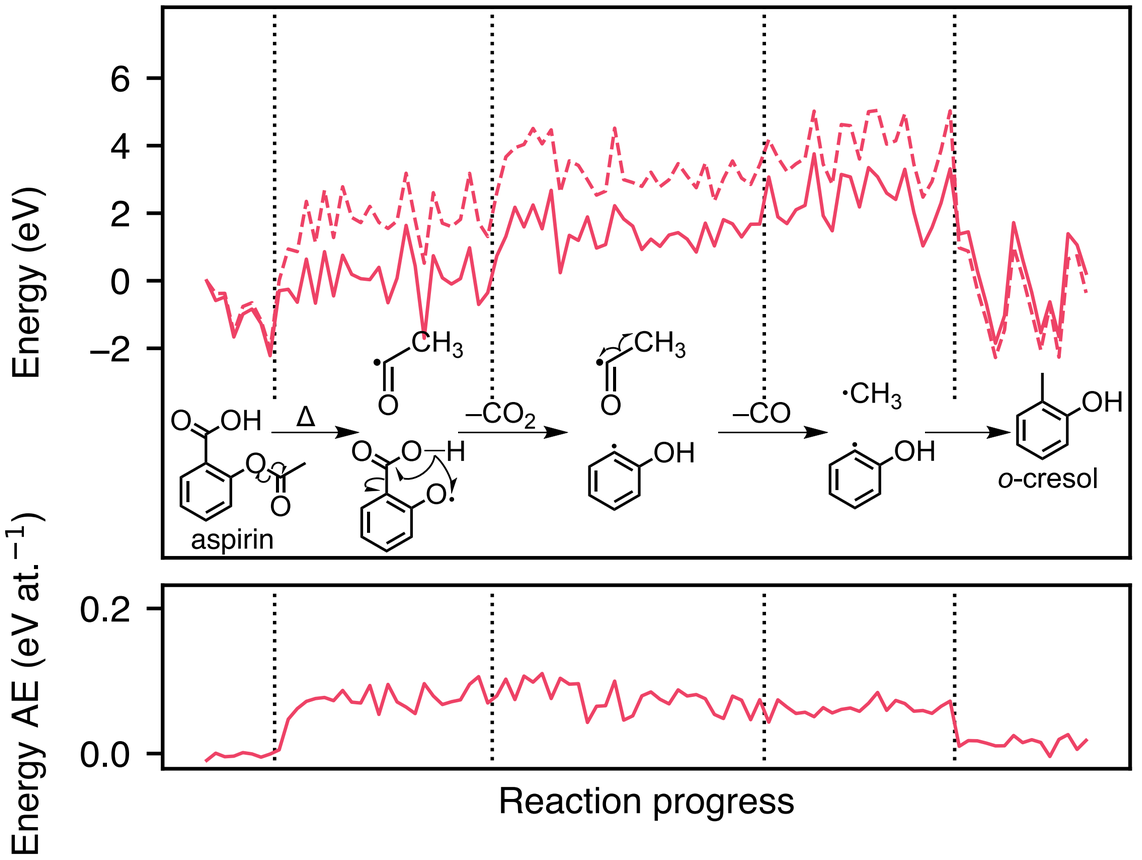
Transferability to reaction pathways



D. A. Nightingale, US Pat., 1604472, 1926

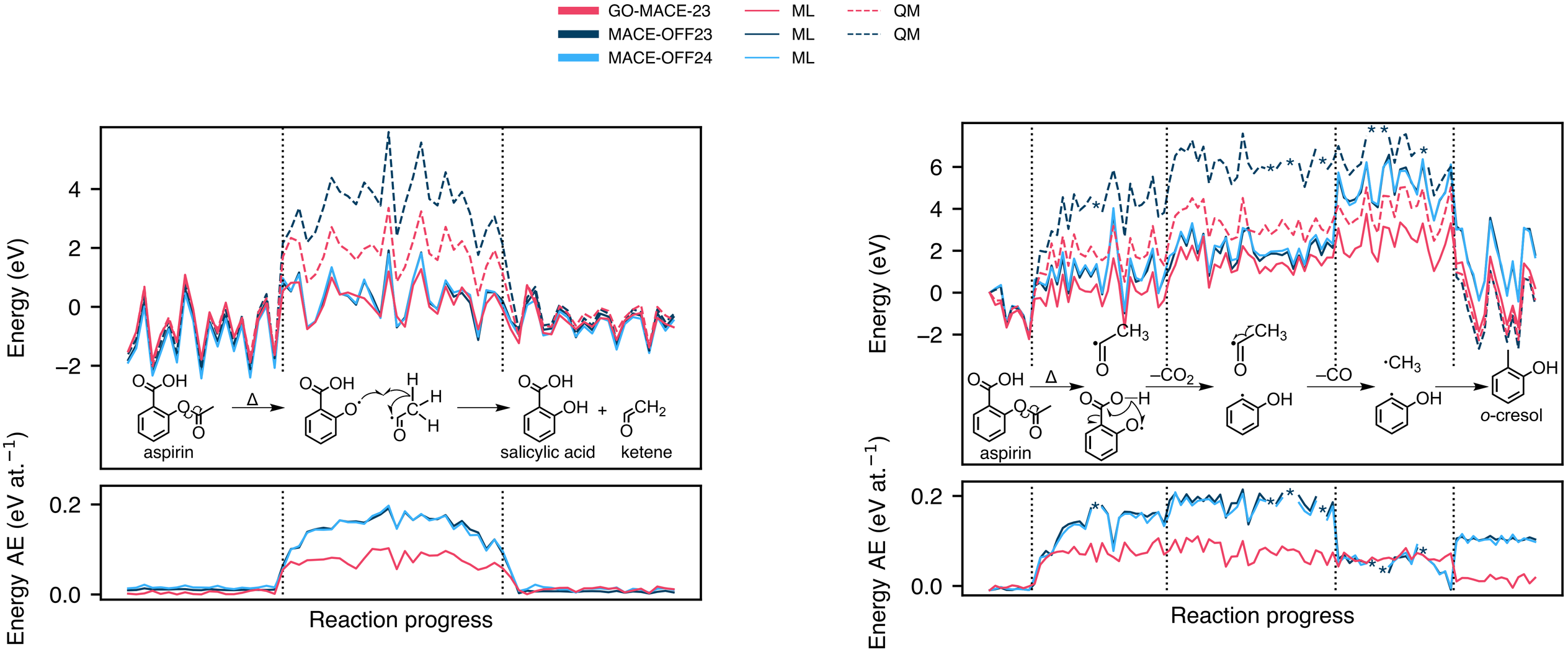
Transferability to reaction pathways





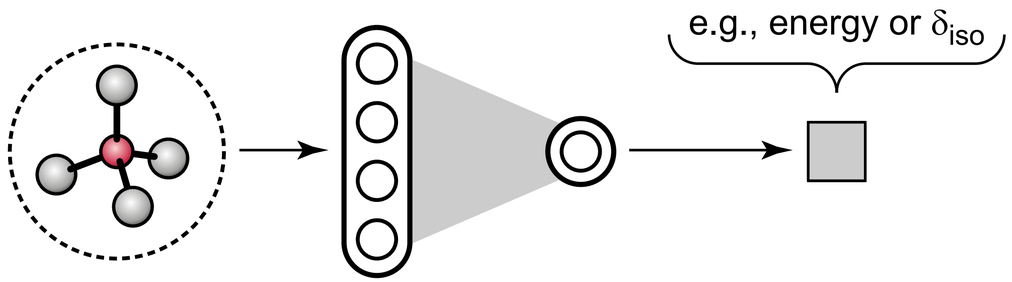
D. A. Nightingale, US Pat., 1604472, 1926

Transferability to reaction pathways

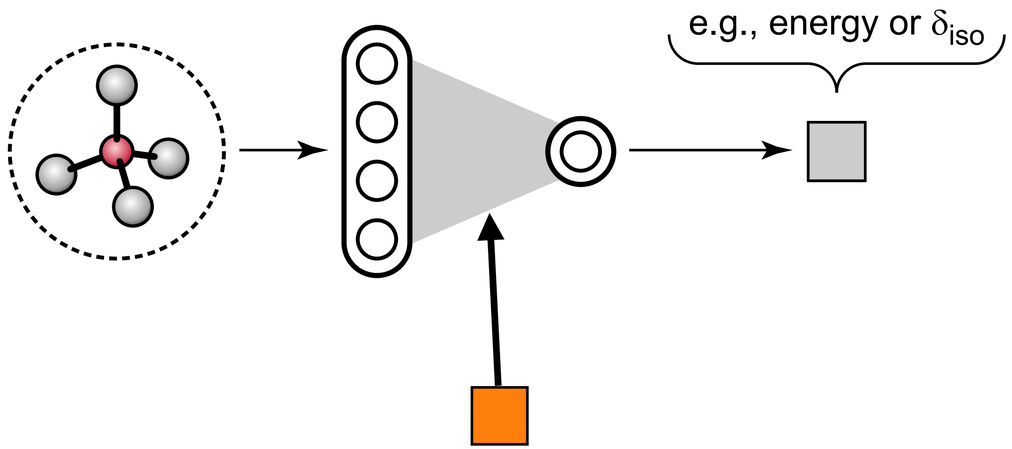


D. A. Nightingale, US Pat., 1604472, 1926

Does model architecture matter?

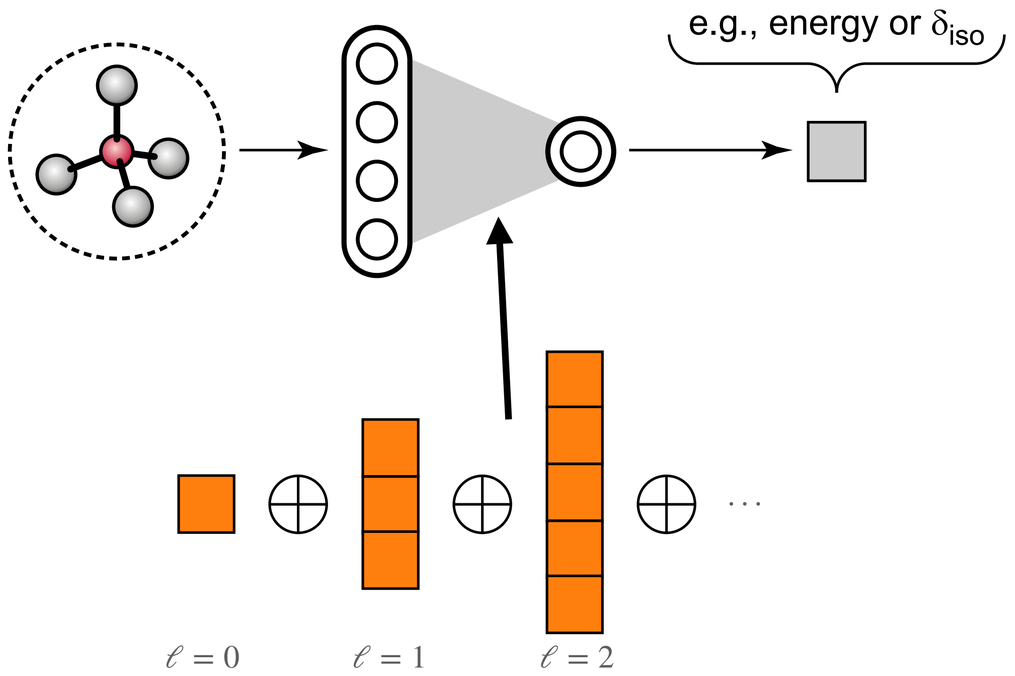


Does model architecture matter?



Invariant features (do not change upon structure rotation for example)

Does model architecture matter?



Equivariant features (do change upon structure rotation for example)

Does model architecture matter?

	Energy RMSE (meV at.$^{-1}$)			Force RMSE (eV Å^{-1})		
`max L`	0	1	2	0	1	2
aspirin	6.2	6.6	**4.9**	0.25	**0.22**	0.28
ethanol	12.3	**10.6**	12.2	0.49	**0.35**	0.48
malonaldehyde	**7.7**	12.3	9.2	0.28	0.33	**0.25**
naphthalene	**3.3**	4.0	3.6	0.18	0.18	**0.17**
salicylic acid	5.3	**4.9**	6.8	**0.22**	**0.22**	0.26
toluene	**5.6**	9.1	6.9	0.32	0.51	**0.25**

CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

Other MLIPs trained on GO

- MLIPs trained with **GraphPES** : a universal interface to train and fine-tune grpah-neural network potentials
- GraphPES accessible on: **https://github.com/jla-gardner/graph-pes**
- **pip install graph-pes**

	Energy RMSE (meV at.$^{-1}$)				
	GO-MACE-23	SchNet	TensorNet	NequIP	PaiNN
aspirin	6.6	22.4	6.6	**5.7**	11.3
ethanol	**10.6**	33.4	17.4	17.2	27.6
malonaldehyde	12.3	38.5	10.8	**8.8**	17.2
naphthalene	4.0	9.9	5.0	**3.9**	5.8
salicylic	4.9	19.7	5.6	**3.9**	7.4
toluene	9.1	16.8	**8.7**	24.0	14.0
	Force RMSE (eV Å^{-1})				
	GO-MACE-23	SchNet	TensorNet	NequIP	PaiNN
aspirin	**0.22**	0.86	0.38	0.31	0.57
ethanol	**0.35**	1.13	0.61	0.47	1.01
malonaldehyde	**0.33**	0.98	0.34	**0.33**	0.38
naphthalene	**0.18**	0.54	0.21	0.21	0.30
salicylic	0.22	0.42	0.24	**0.19**	0.25
toluene	0.51	0.59	**0.28**	0.38	0.32

CBM, El-Machachi, Z., Gierczak, K.A., Gardner, J.L.A., Deringer, V.L., 2025. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials. arXiv: 2502.21317

Takehome message

- Systematic assessment of out-of-domain performance is essential for MLIPs
- Structural variability is as important as chemical diversity when training general-purpose MLIPs for chemical reactions

Conclusions

Conclusions

- ML tools are essential in atomistic modelling

Conclusions

- ML tools are essential in atomistic modelling
- Data is central to atomistic ML

Conclusions

- ML tools are essential in atomistic modelling
- Data is central to atomistic ML
- The emergence of pretrained/foundation models will close the gap between material and molecular modelling

Conclusions

- ML tools are essential in atomistic modelling
- Data is central to atomistic ML
- The emergence of pretrained/foundation models will close the gap between material and molecular modelling
- Universal models describing structural, electronic, and spectroscopic fingerprints within physics-based approximations can enable more accurate and predictive materials modelling and design

Acknowledgements

Lab COSMO, EPFL:

- Prof. Michele Ceriotti
- Dr Andrea Anelli
- Dr Nataliya Lopanitsyna
- Dr Federico Grasselli

University of Cambridge:

- Prof. Gábor Csányi

Deringer Group, University of Oxford:

- Prof. Volker Deringer
- Louise Rosset
- John Gardner
- Zakariya El-Machachi
- Krystian Gierczak

University of Oxford:

- Prof. Jonathan Yates

Swiss National Science Foundation

UKRI Engineering and Physical Sciences Research Council