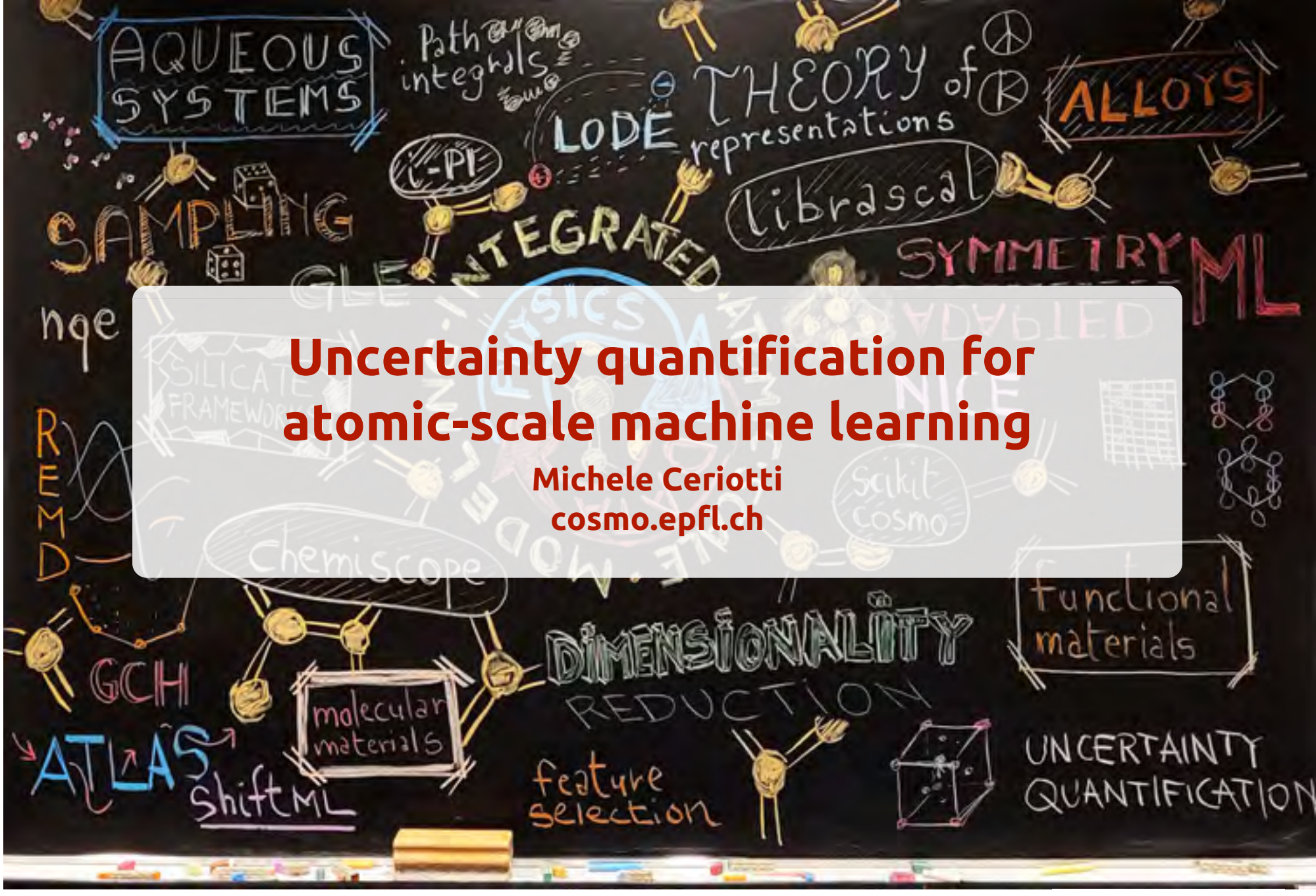


Uncertainty quantification for atomic-scale machine learning

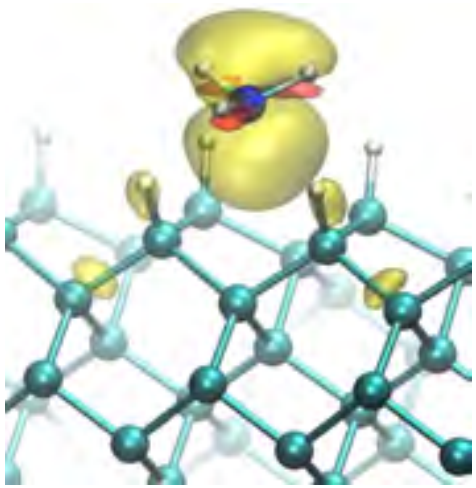
Michele Ceriotti
cosmo.epfl.ch

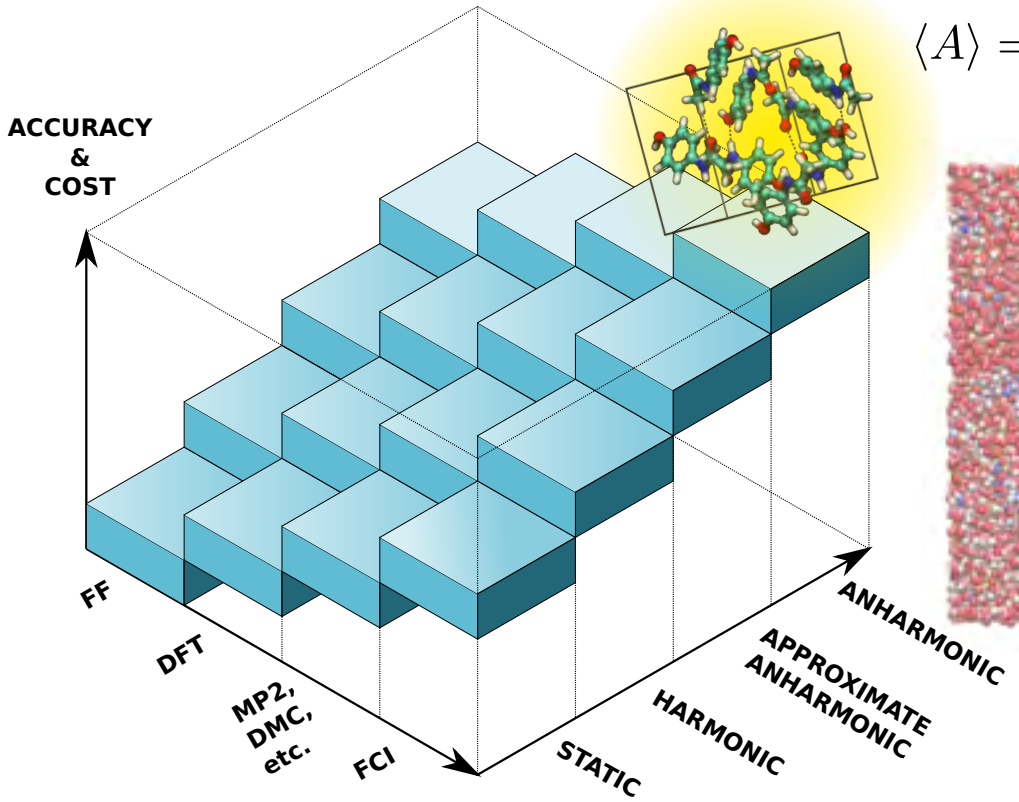
Predictive accuracy for real materials

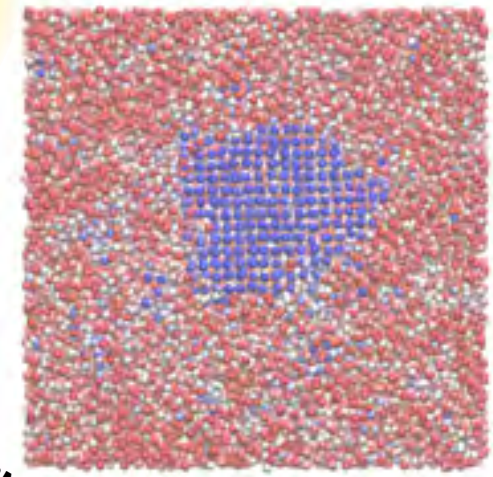
- First-principles calculations promise quantitatively accurate simulations that make no use of experimental data
- Emergent physics from first principles: still a tremendous challenge
- Machine learning to the rescue

$$\hat{H}(\mathbf{q})|\Psi\rangle = V(\mathbf{q})|\Psi\rangle$$

$$\langle A\rangle = \int \mathrm{d}\mathbf{q} e^{-\beta V(\mathbf{q})} A(\mathbf{q})$$







ACCURACY OF ENERGETICS

ACCURACY OF SAMPLING

Grabowski et al., PRB (2009); Kapil, Engel, Rossi, **MC**, JCTC (2019)

Predictive accuracy for real materials

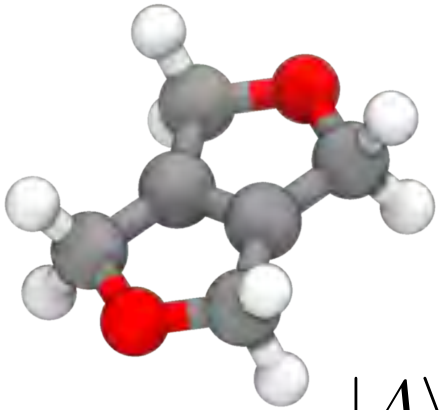
- First-principles calculations promise quantitatively accurate simulations that make no use of experimental data
- Emergent physics from first principles: still a tremendous challenge
- Machine learning to the rescue

FIRST-PRINCIPLES QUANTUM MODELS

$A \equiv \{a_i, \mathbf{r}_i\}$

$\hat{H}\,|\Psi\rangle = V\,|\Psi\rangle$

$V(\{a_i, \mathbf{r}_i\})$



$\langle\Psi|\,\hat{Y}\,|\Psi\rangle$

$|A\rangle$

$\tilde{V}(\,|A\rangle)$

DATA-DRIVEN SURROGATE MODELS

μ

$|A; \lambda\mu\rangle$

$\tilde{y}^{\mu}_{\lambda}(\,|A\rangle)$

H

ρ

Musil et al., Chem. Rev. (2021)

Uncertainty and errors

We don’t talk about errors

UQ for atomistic simulations

- Errors from model approximations:
  - 50 shades of DFT (and implementations!)
  - Classical vs quantum nuclei, Born-Oppenheimer vs non-adiabatic
- Statistical convergence of averages, accelerated sampling

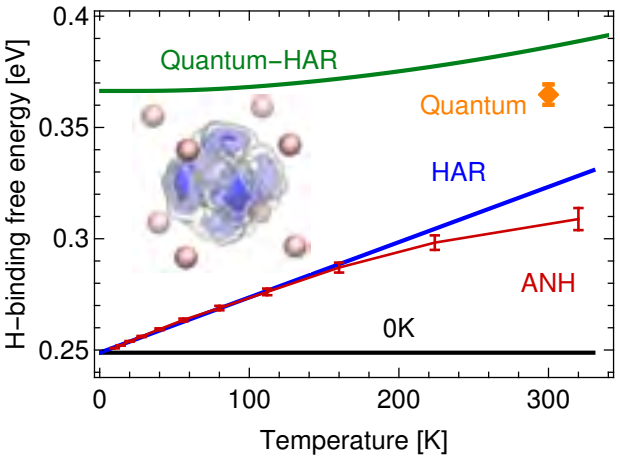
Scorecard

	0.3	0.3	0.6	1.0	0.9	0.3	1.5	0.6	0.9	0.4	1.0	0.4	0.4	6.3	13.5	1.1	2.1	0.7	1.4
0.3		0.1	0.5	0.9	0.8	0.2	1.5	0.6	0.8	0.4	1.0	0.5	0.3	6.3	13.4	1.1	2.1	0.7	1.4
0.3	0.1		0.5	0.9	0.8	0.2	1.5	0.6	0.8	0.4	0.9	0.5	0.3	6.3	13.4	1.1	2.1	0.7	1.4
0.6	0.5	0.5		0.8	0.6	0.4	1.5	0.6	0.8	0.6	1.0	0.7	0.5	6.3	13.2	1.0	1.9	0.6	1.3
1.0	0.9	0.9	0.8		0.9	0.9	1.8	0.9	1.3	1.0	1.4	1.0	0.9	6.4	13.0	1.2	1.8	1.0	1.6
0.9	0.8	0.8	0.6	0.9		0.8	1.7	0.7	1.1	0.8	1.3	1.0	0.8	6.5	13.2	1.1	1.8	0.8	1.5
0.3	0.2	0.2	0.4	0.9	0.8		1.5	0.5	0.8	0.3	1.0	0.5	0.3	6.2	13.4	1.0	2.0	0.6	1.4

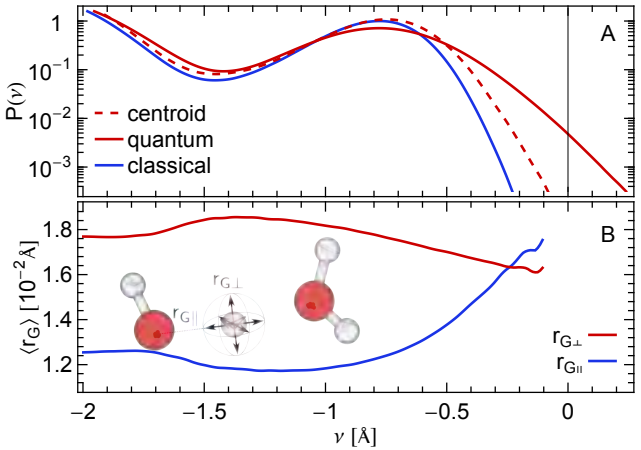
Lejaeghere et al., Science (2016)

UQ for atomistic simulations

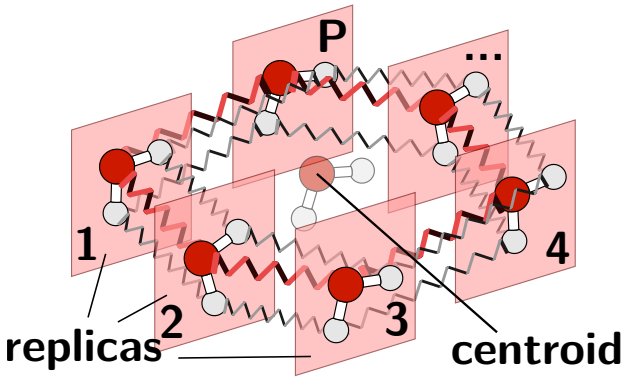
- Errors from model approximations:
  - 50 shades of DFT (and implementations!)
  - Classical vs quantum nuclei, Born-Oppenheimer vs non-adiabatic
- Statistical convergence of averages, accelerated sampling



Cheng, Paxton, **MC**, Phys. Rev. Lett. (2018)



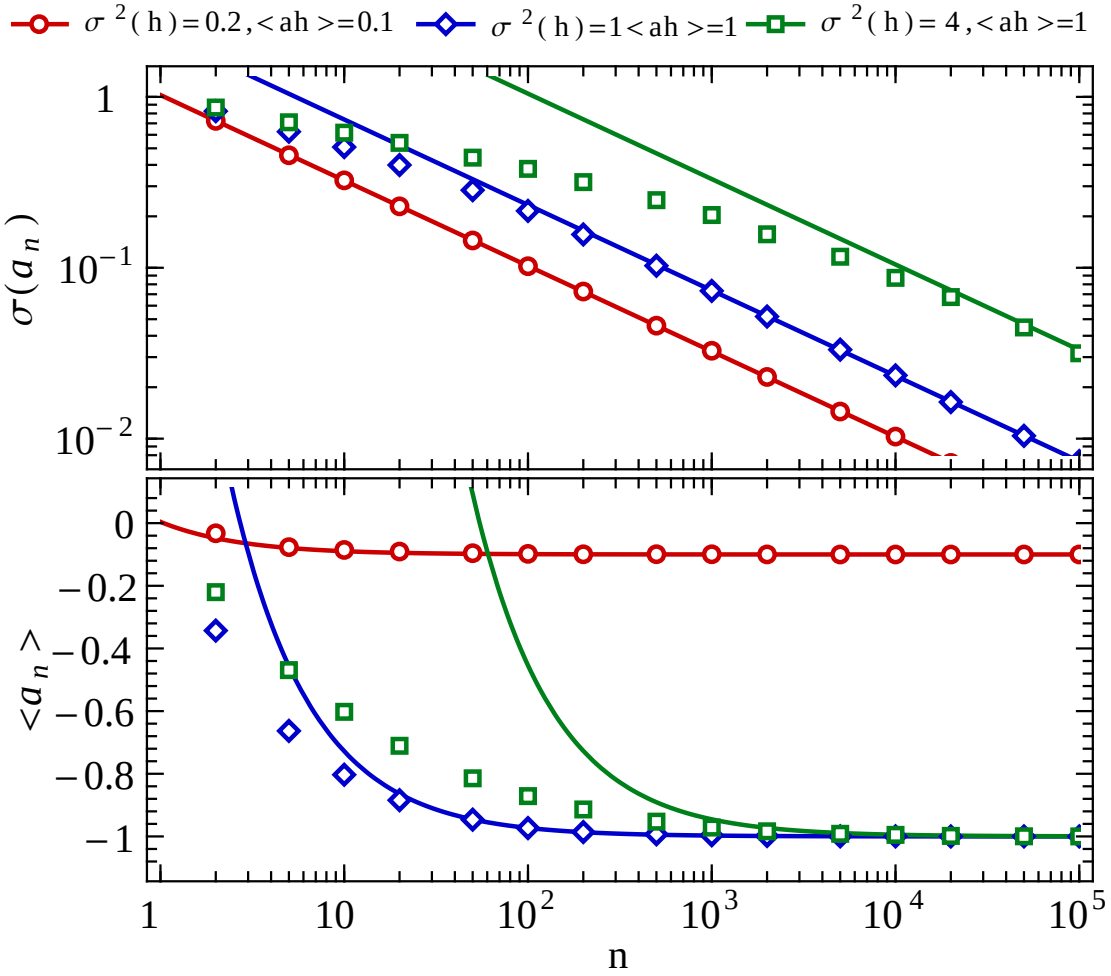
MC et al., Chem. Rev. (2016)



Markland & **MC**, Nat. Rev. Chem. (2018)

UQ for atomistic simulations

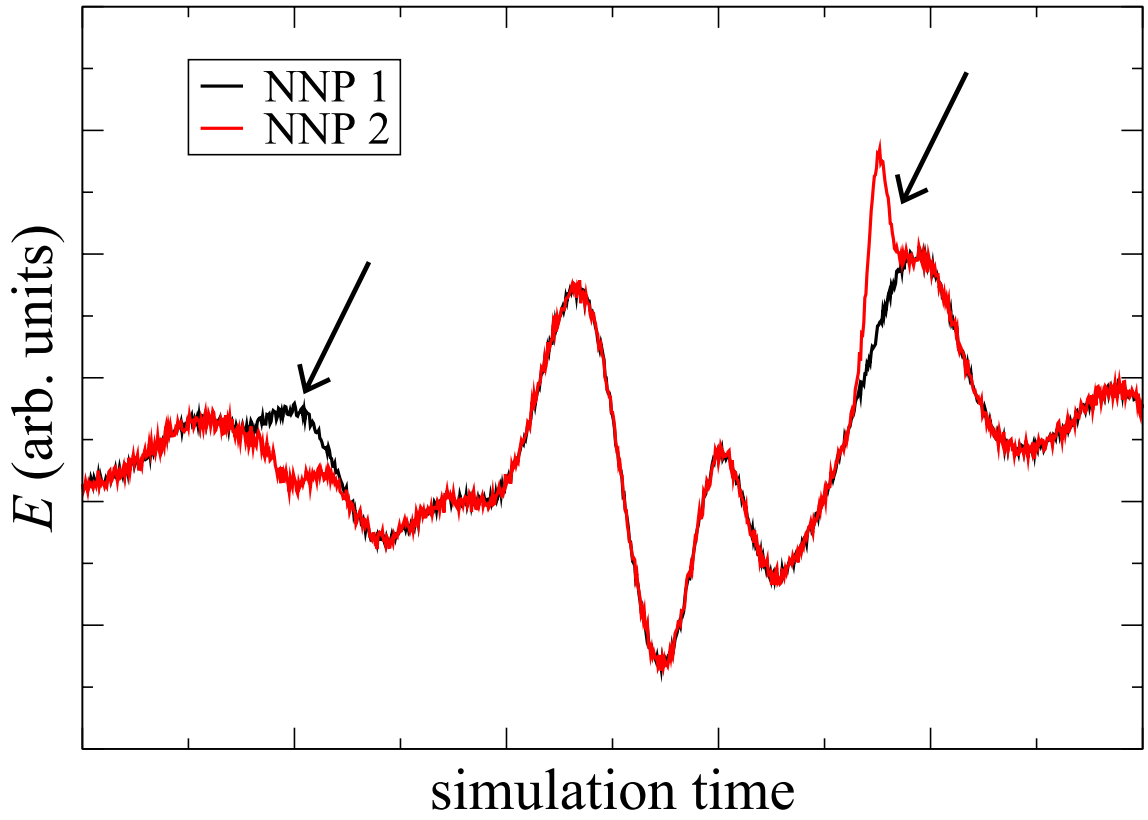
- Errors from model approximations:
  - 50 shades of DFT (and implementations!)
  - Classical vs quantum nuclei, Born-Oppenheimer vs non-adiabatic
- Statistical convergence of averages, accelerated sampling



MC, Brain, Riordan, Manolopoulos Proc. Royal Soc. A (2012)

UQ for ML: the classics

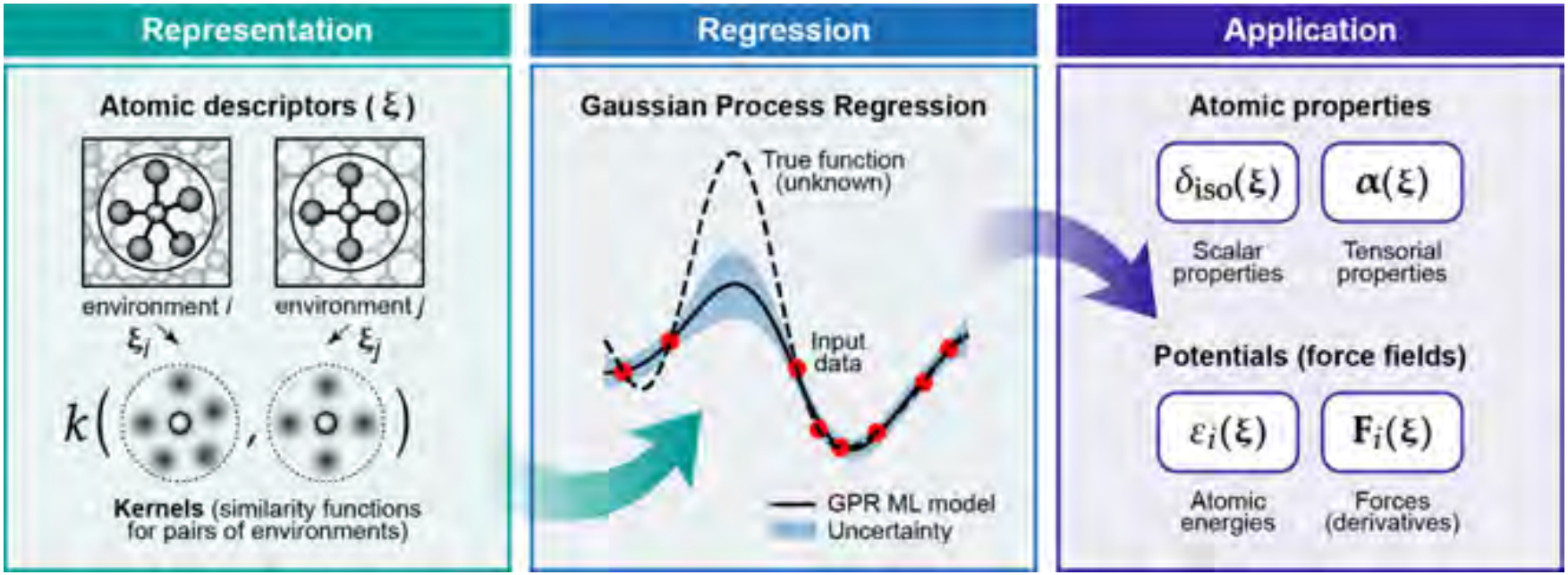
- NN sanity check comparing the outputs of multiple fits
- Gaussian process regression provides built-in uncertainty measure
- Active learning driven by error estimates



Behler, J. Phys. Cond. Mat (2014)

UQ for ML: the classics

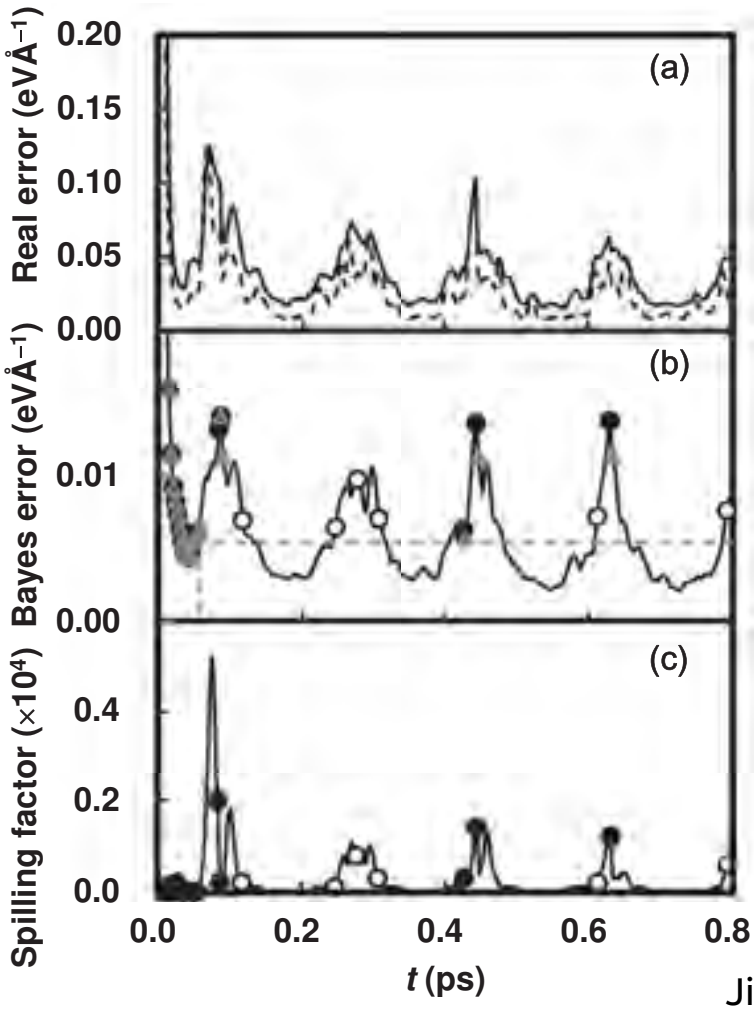
- NN sanity check comparing the outputs of multiple fits
- Gaussian process regression provides built-in uncertainty measure
- Active learning driven by error estimates



Deringer et al. Chem. Rev. (2021)

UQ for ML: the classics

- NN sanity check comparing the outputs of multiple fits
- Gaussian process regression provides built-in uncertainty measure
- Active learning driven by error estimates

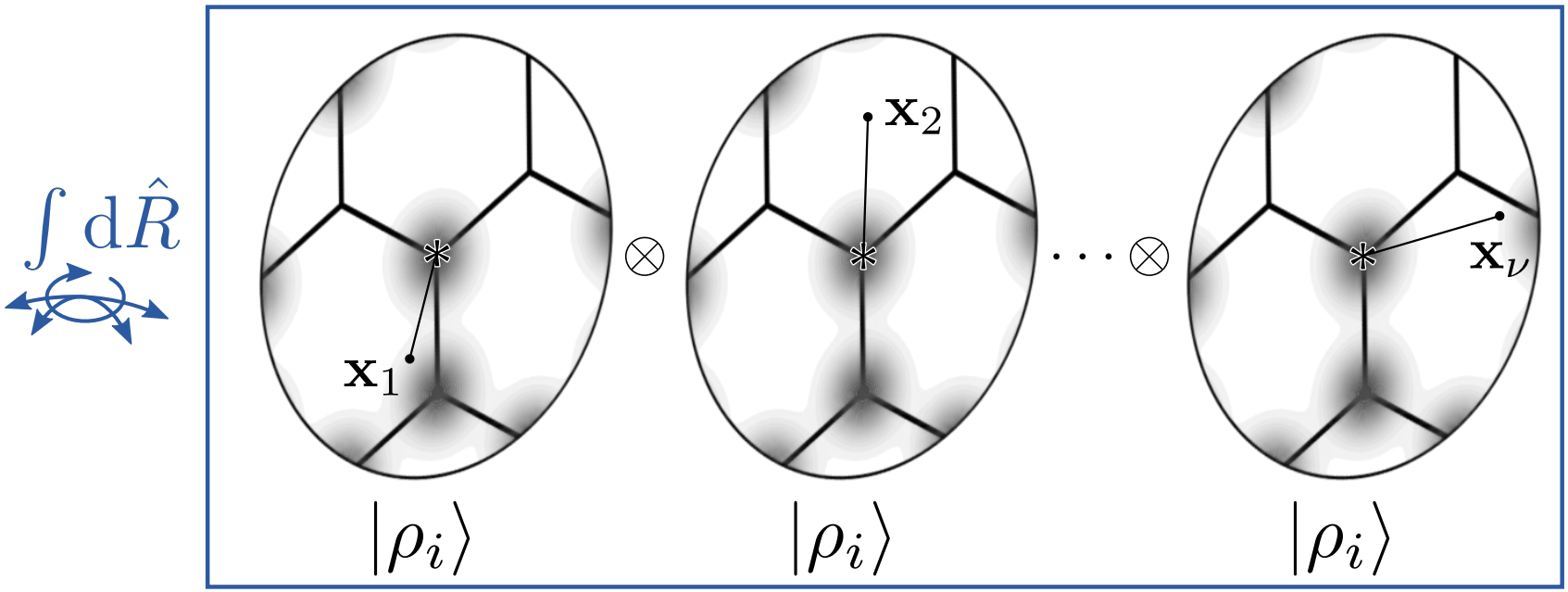


Jinnouchi et al., Phys. Rev. B (2019)

What we do, what we need

A universal feature construction

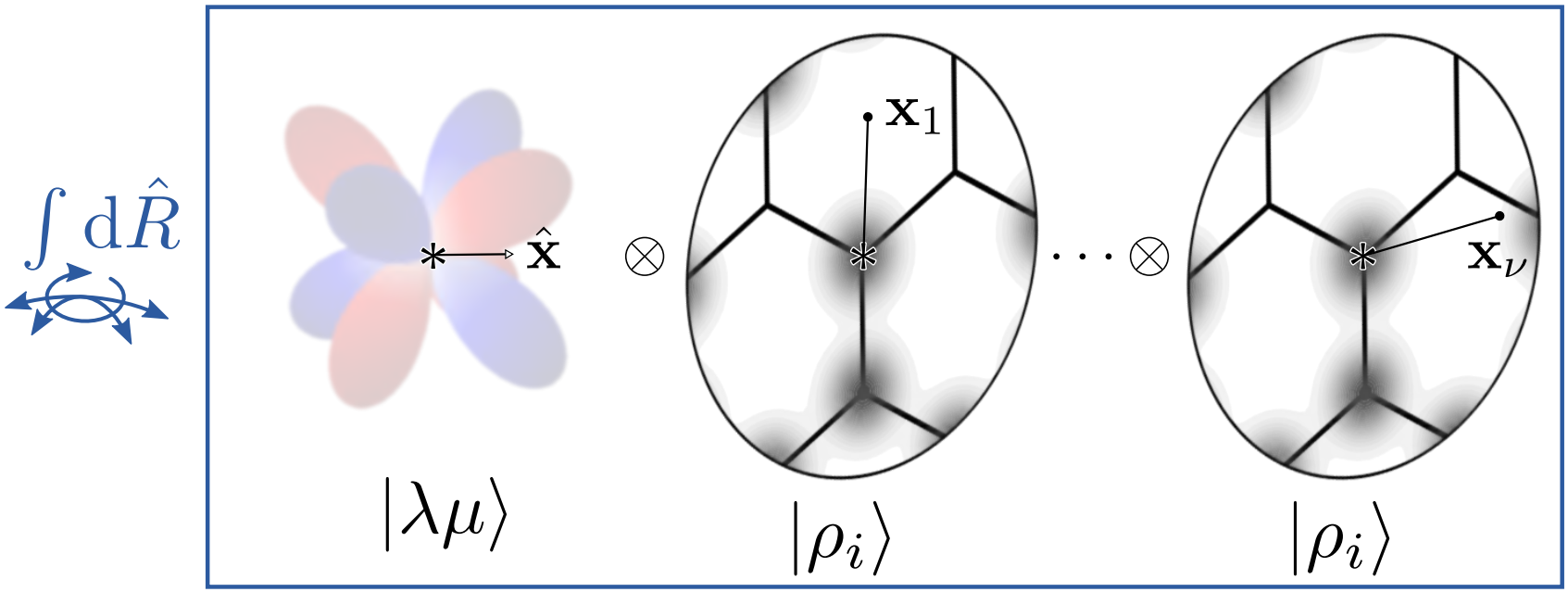
- Most frameworks can be expressed in terms of *n*-body correlations of atom positions. Only difference - the choice of basis
- Extension to a fully equivariant framework (NICE)
- ... to features to describe long-range interactions (LODE)
- ... and to message-passing, *N*-center features (MP-ACDC)



Introductory review: Musil et al., Chem. Rev. (2021)

A universal feature construction

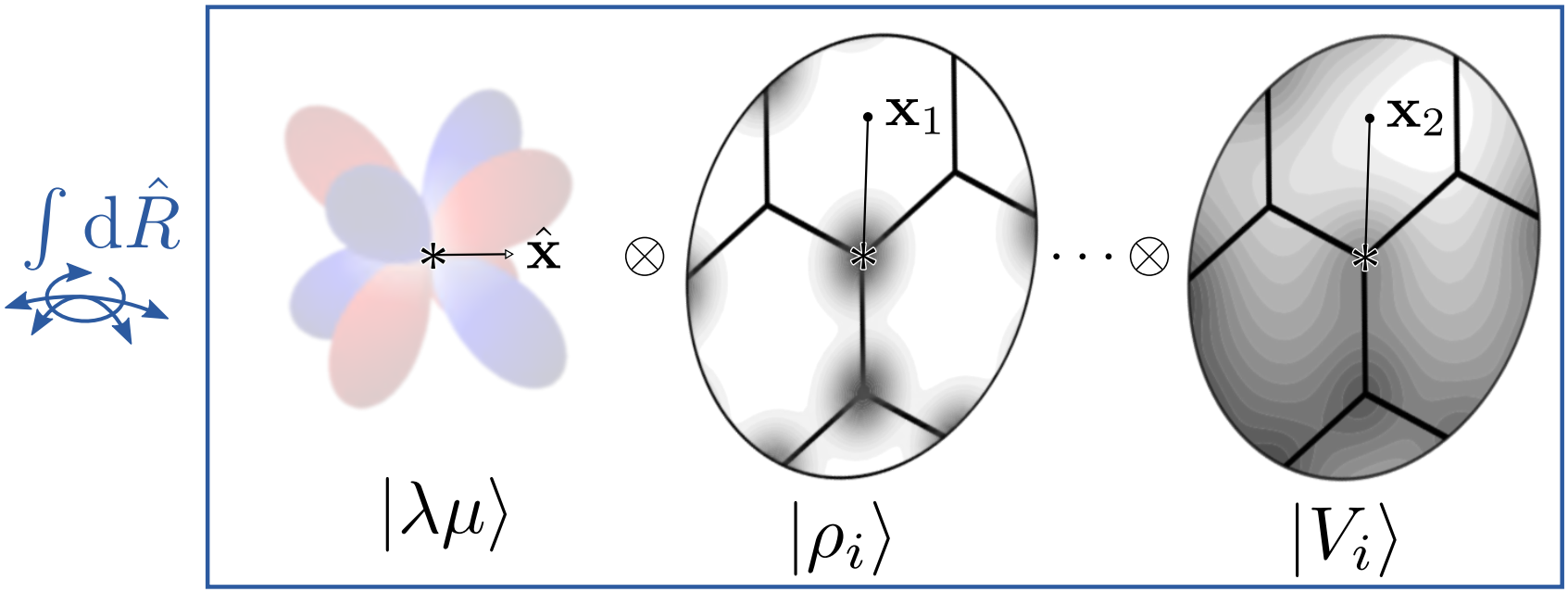
- Most frameworks can be expressed in terms of *n*-body correlations of atom positions. Only difference - the choice of basis
- Extension to a fully equivariant framework (NICE)
- ... to features to describe long-range interactions (LODE)
- ... and to message-passing, *N*-center features (MP-ACDC)



Introductory review: Musil et al., Chem. Rev. (2021)

A universal feature construction

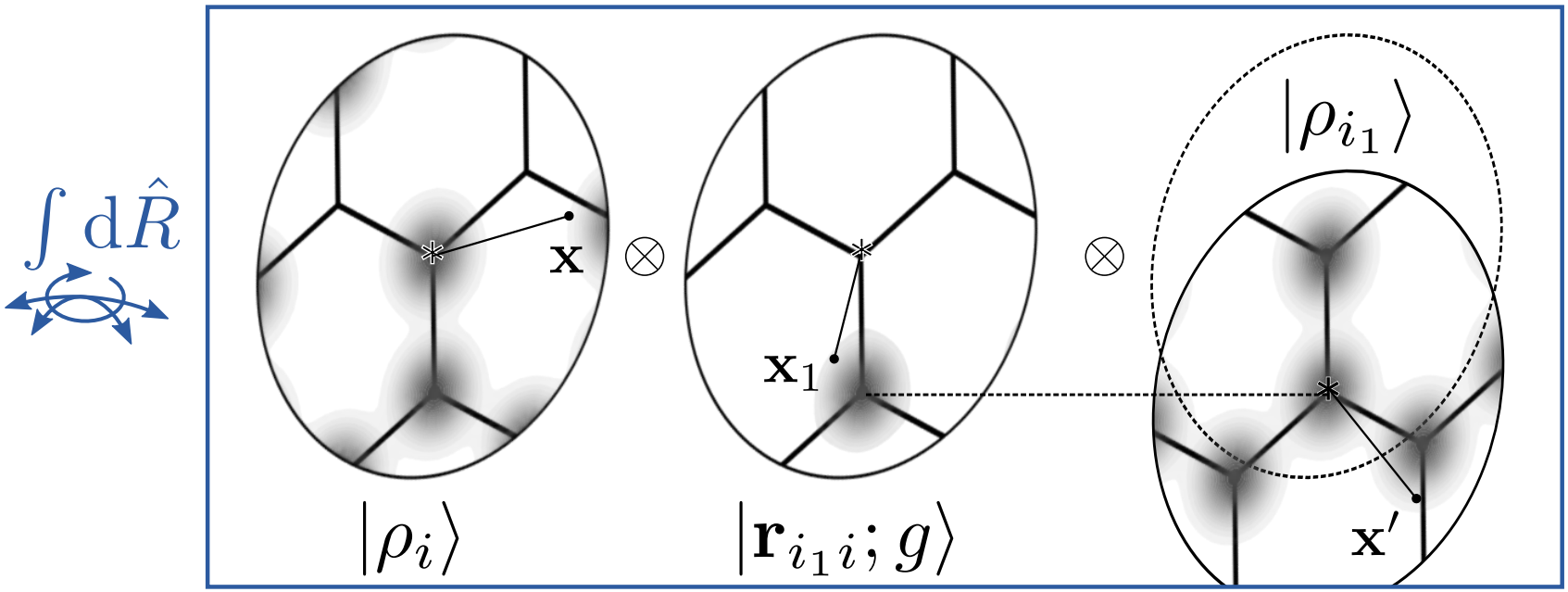
- Most frameworks can be expressed in terms of *n*-body correlations of atom positions. Only difference - the choice of basis
- Extension to a fully equivariant framework (NICE)
- ... to features to describe long-range interactions (LODE)
- ... and to message-passing, *N*-center features (MP-ACDC)



Introductory review: Musil et al., Chem. Rev. (2021)

A universal feature construction

- Most frameworks can be expressed in terms of *n*-body correlations of atom positions. Only difference - the choice of basis
- Extension to a fully equivariant framework (NICE)
- ... to features to describe long-range interactions (LODE)
- ... and to message-passing, *N*-center features (MP-ACDC)

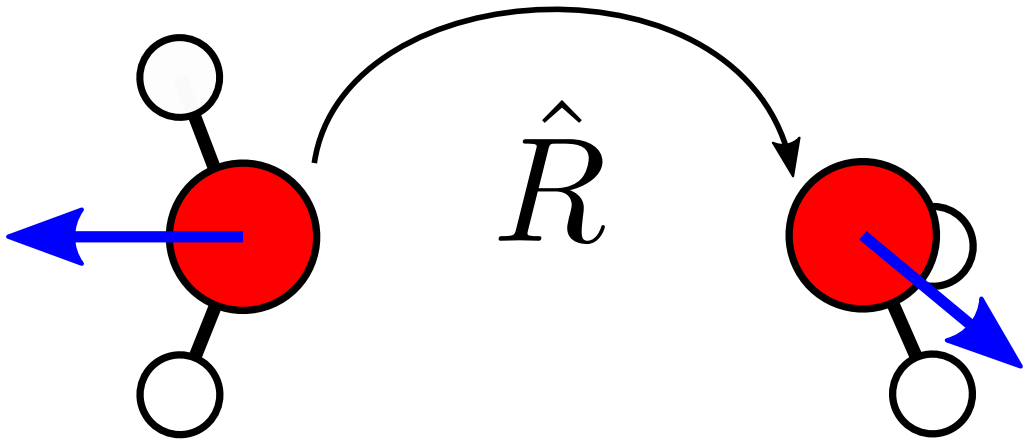


Latest & greatest: Nigam, Pozdnyakov, Fraux, **MC**, JCP (2022)

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$d_\alpha\left(\hat{R}A_i\right) = \sum_q \langle d|q\rangle \, \langle q|\hat{R}A; \overline{\rho_i^{\otimes\nu}}; \alpha\rangle$$

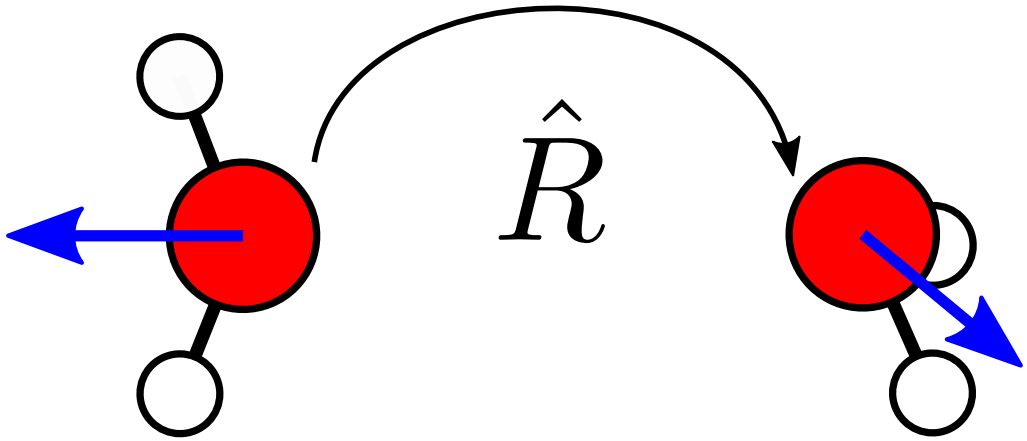


Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$d_\alpha\left(\hat{R}A_i\right) = \sum_q \langle d|q\rangle \sum_{\alpha'} R_{\alpha\alpha'} \langle q|A;\overline{\rho_i^{\otimes\nu}};\alpha'\rangle = \sum_{\alpha'} R_{\alpha\alpha'}\, d_{\alpha'}\,(A_i)$$

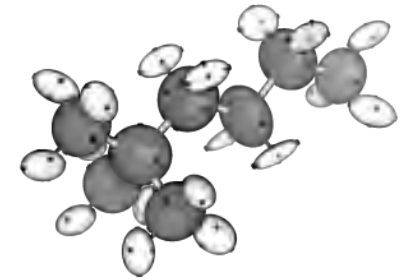


Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

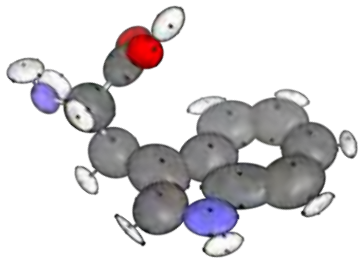
Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$y^{\lambda}_{\mu}\left(\hat{R}A_i\right) = \sum_{q}\langle d|q\rangle \sum_{\mu'} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right)\,\langle q|A;\overline{\rho_i^{\otimes\nu}};\lambda\mu\rangle$$

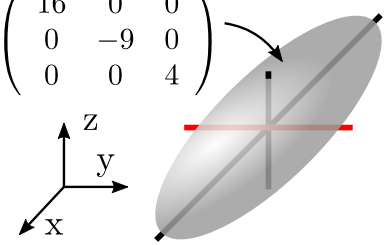


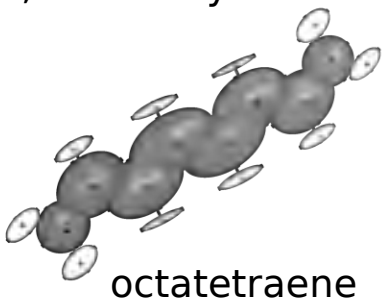
2,2-dimethylhexane



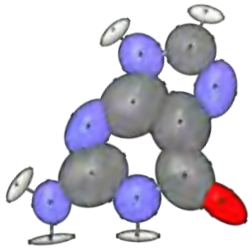
tryptophan

$$\boldsymbol{\alpha} \approx \begin{pmatrix} 16 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

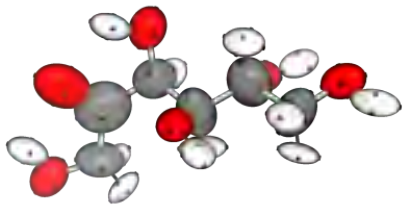




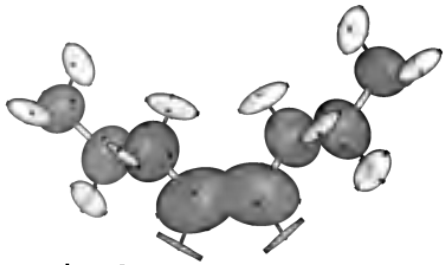
octatetraene



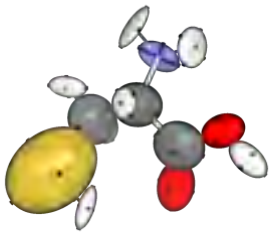
guanine



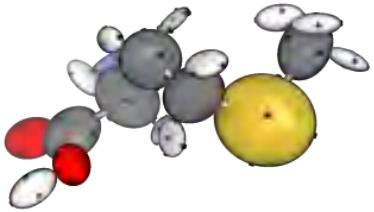
fructose



cis-4-octene



cysteine



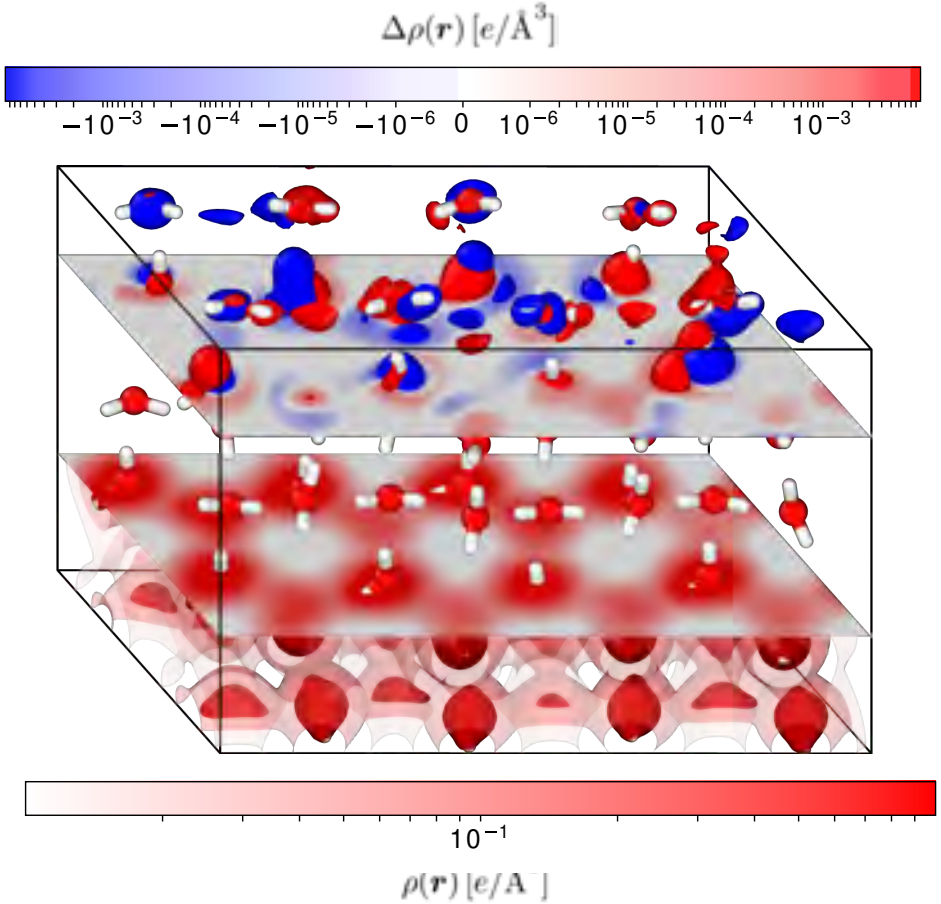
methionine

Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, PNAS (2019);

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$y^{\lambda}_{\mu}\left(\hat{R}A_i\right) = \sum_{q}\langle d|q\rangle \sum_{\mu'} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right)\,\langle q|A;\overline{\rho_i^{\otimes\nu};\lambda\mu}\rangle$$

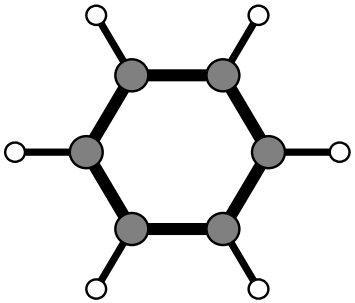


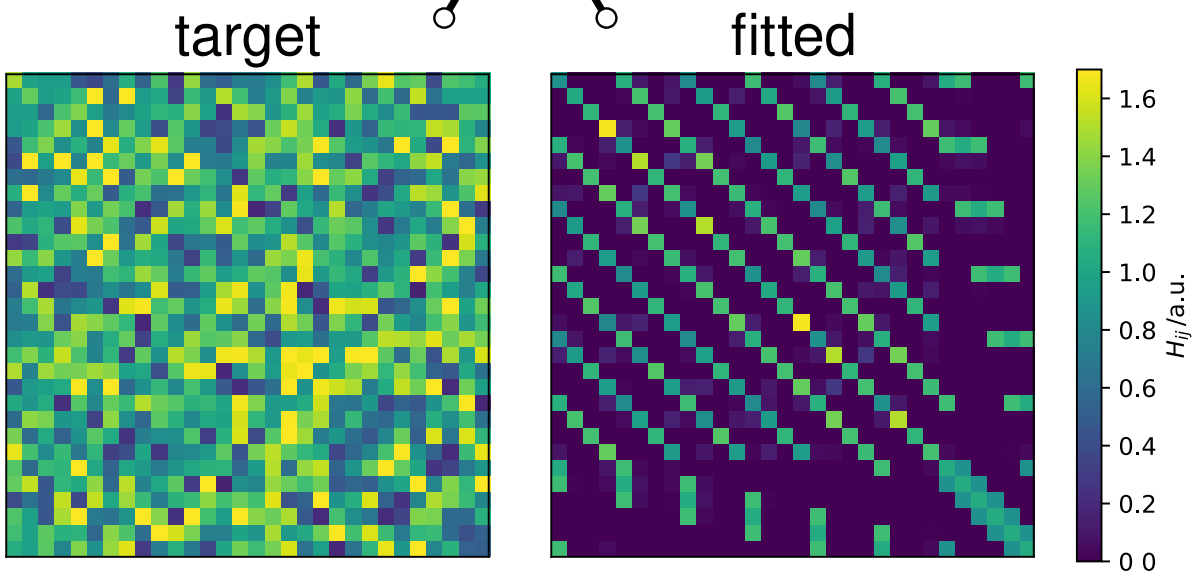
Lewis, Grisafi, **MC**, Rossi, JCTC (2021)

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$y^{\lambda}_{\mu}\left(\hat{R}A_i\right) = \sum_{q}\langle d|q\rangle \sum_{\mu'} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right)\,\langle q|A;\overline{\rho_i^{\otimes\nu}};\lambda\mu\rangle$$



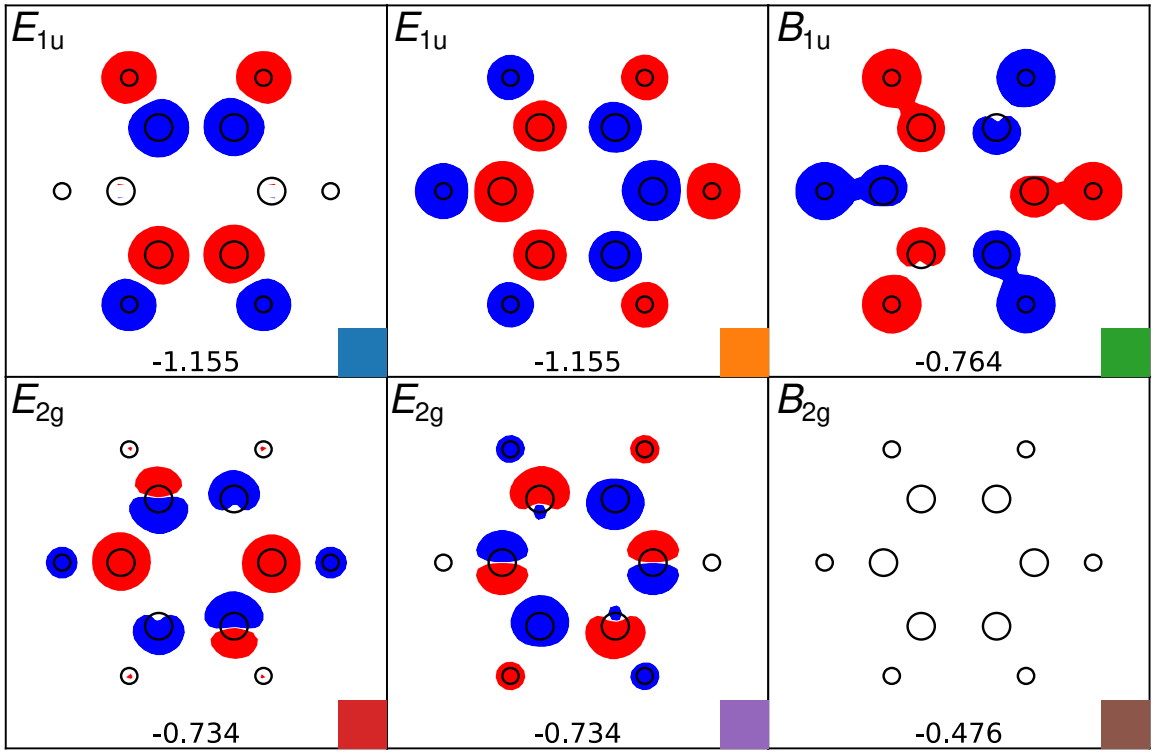


Nigam, Willatt, **MC**, JCP (2022)

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

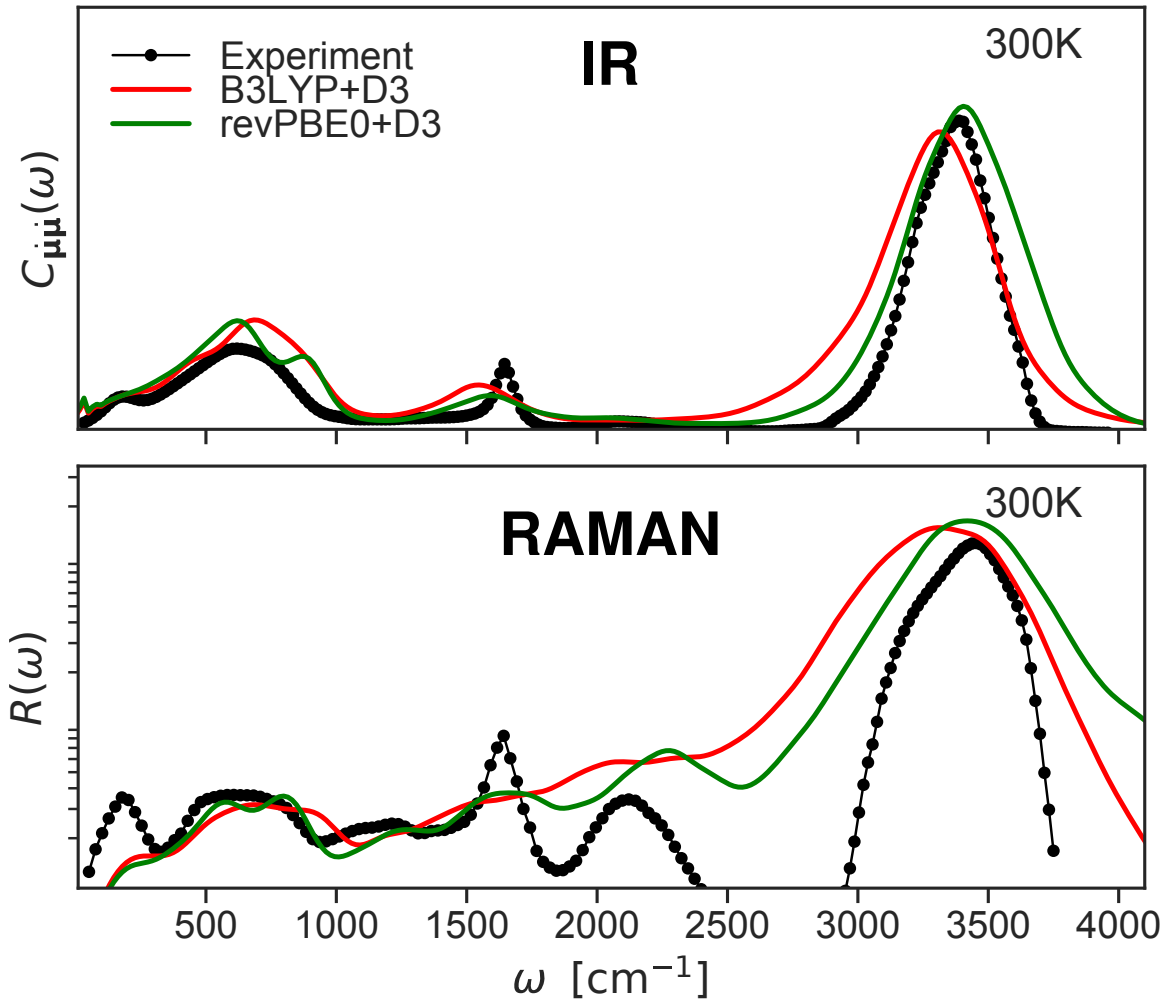
$$y^{\lambda}_{\mu}\left(\hat{R}A_i\right) = \sum_{q}\langle d|q\rangle \sum_{\mu'} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right)\,\langle q|A;\overline{\rho_i^{\otimes\nu}};\lambda\mu\rangle$$



Nigam, Willatt, **MC**, JCP (2022)

Integrated ML, beyond potentials

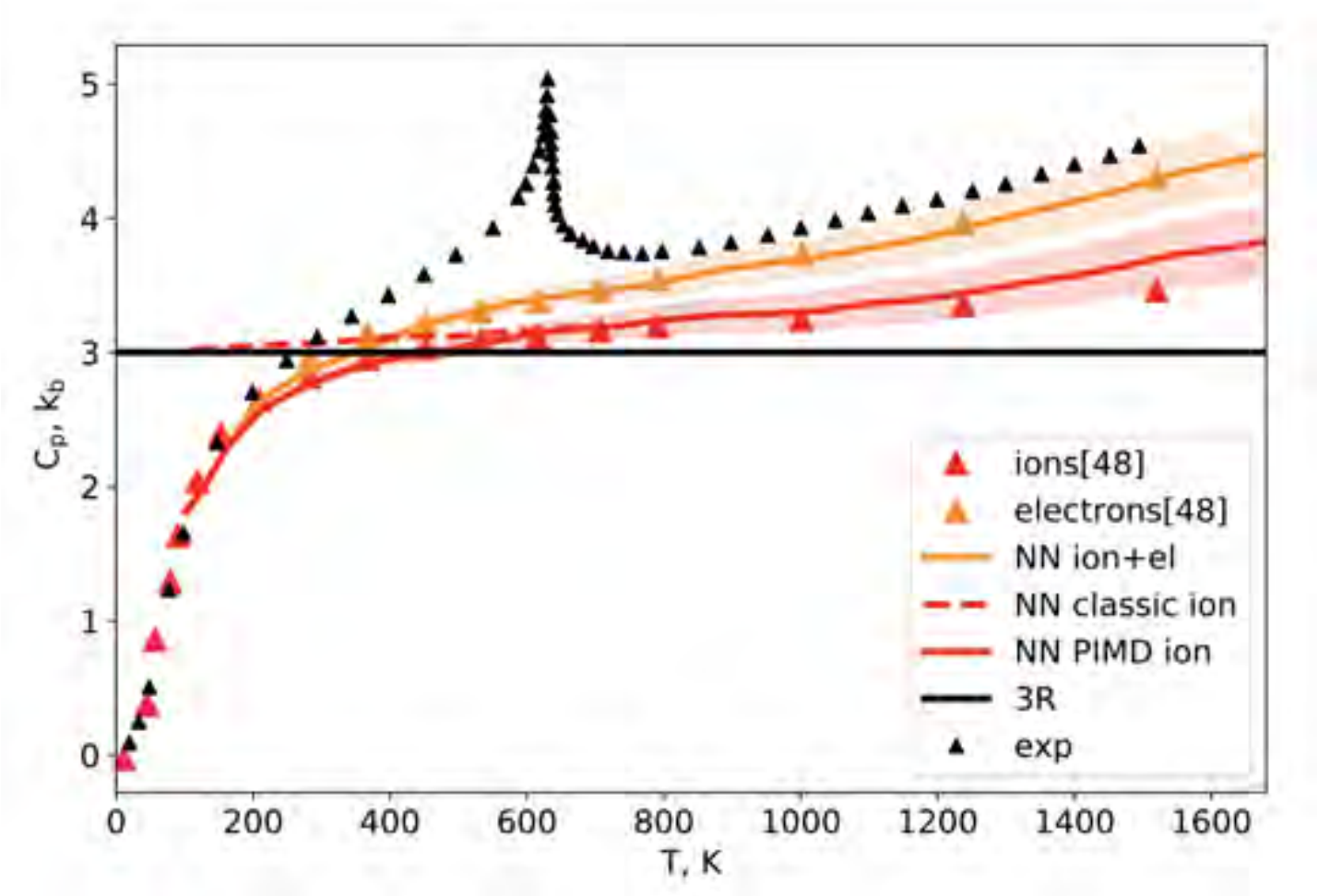
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



Kapil, Wilkins, Lan, **MC**, JCP (2020)

Integrated ML, beyond potentials

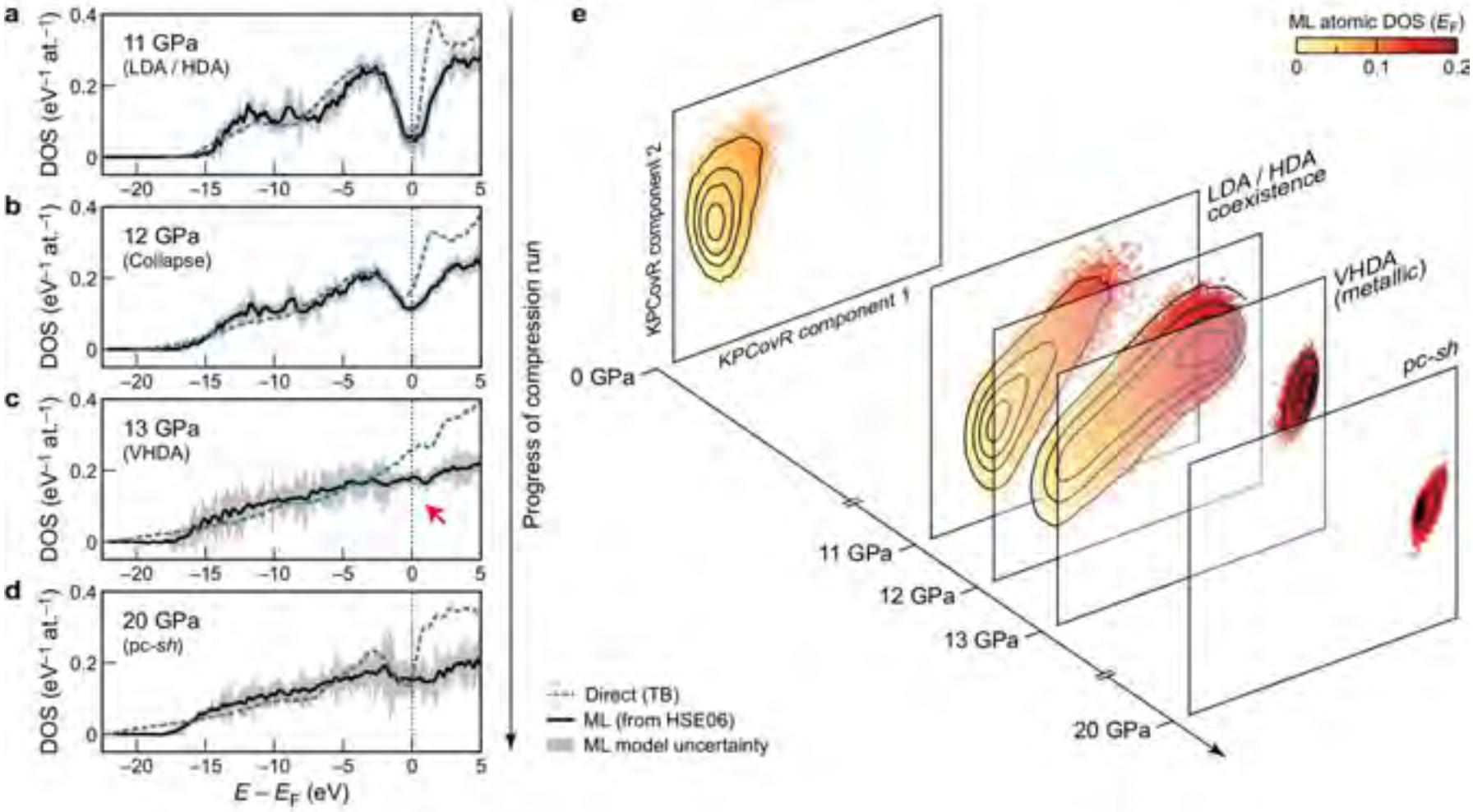
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



N. Lopanitsyna, C. Ben Mahmoud, **MC**, Phys. Rev. Mater. (2021)

Integrated ML, beyond potentials

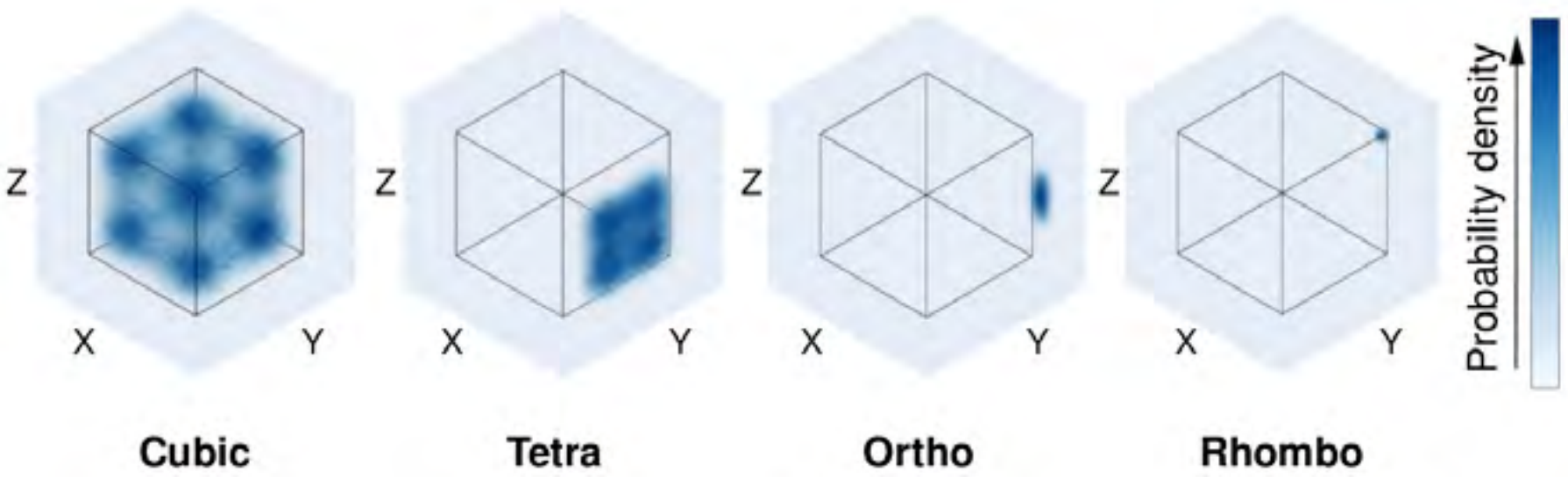
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



V. Deringer et al., Nature (2021)

Integrated ML, beyond potentials

- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



Gigli et al., NPJ Comp. Mat. in press (2022)

Uncertainty quantification made easy

Calibrated committee models

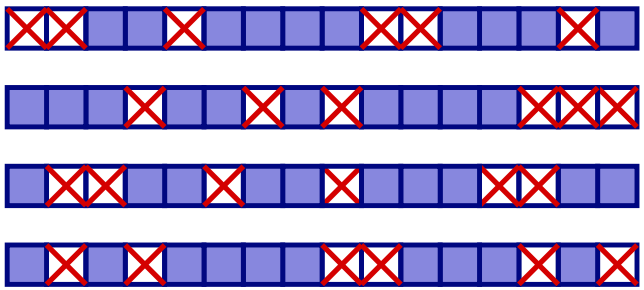
- Ensemble of N_{RS} models, trained on subsets of the train set

$$\bar{y}(A) = \frac{1}{N_{RS}} \sum_j \tilde{y}_j(A), \qquad \sigma^2(A) = \frac{1}{N_{RS}-1} \sum_i \left(\tilde{y}_j(A) - \bar{y}(A)\right)^2$$

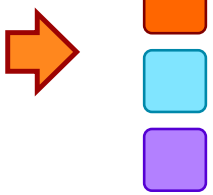
- Verify accuracy by the distribution of errors $P\left(|\bar{y}(A) - y_A(A)| \;\middle|\; \sigma(A)\right)$
- Adjust the spread by maximum likelihood $\tilde{y}_j \leftarrow \bar{y} + \alpha\left(\tilde{y}_j - \bar{y}\right)$
- Calibrated model can be used for easy uncertainty propgation

1. train set $\{(A, y_A)\}$

2. subselection



3. committee of models



$\tilde{y}_j(A)$

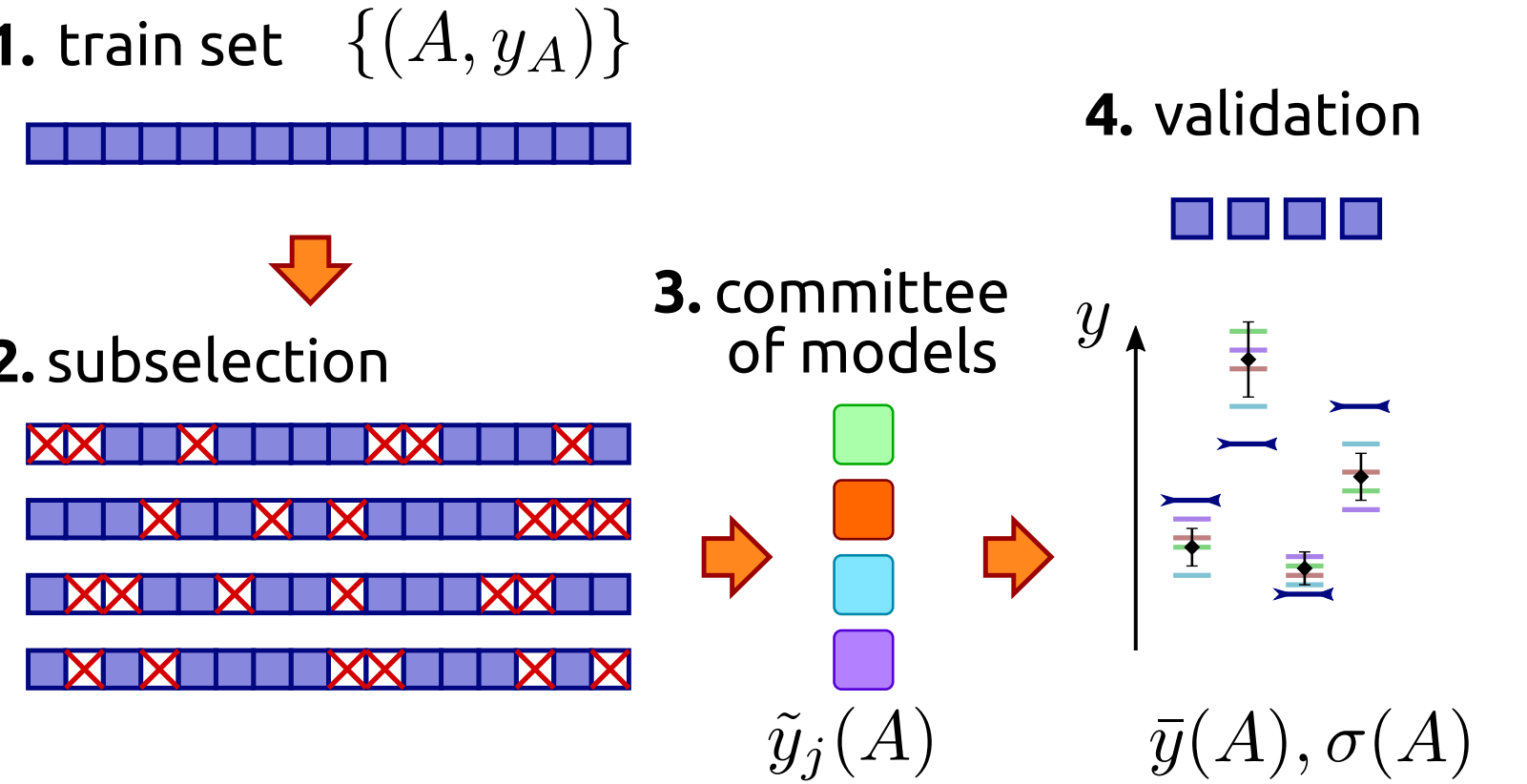
Musil et al., JCTC (2019)

Calibrated committee models

- Ensemble of N_{RS} models, trained on subsets of the train set

$$\bar{y}(A) = \frac{1}{N_{RS}} \sum_j \tilde{y}_j(A), \qquad \sigma^2(A) = \frac{1}{N_{RS}-1} \sum_i \left(\tilde{y}_j(A) - \bar{y}(A)\right)^2$$

- Verify accuracy by the distribution of errors $P\left(|\bar{y}(A) - y_A(A)| \;\middle|\; \sigma(A)\right)$
- Adjust the spread by maximum likelihood $\tilde{y}_j \leftarrow \bar{y} + \alpha\left(\tilde{y}_j - \bar{y}\right)$
- Calibrated model can be used for easy uncertainty propgation



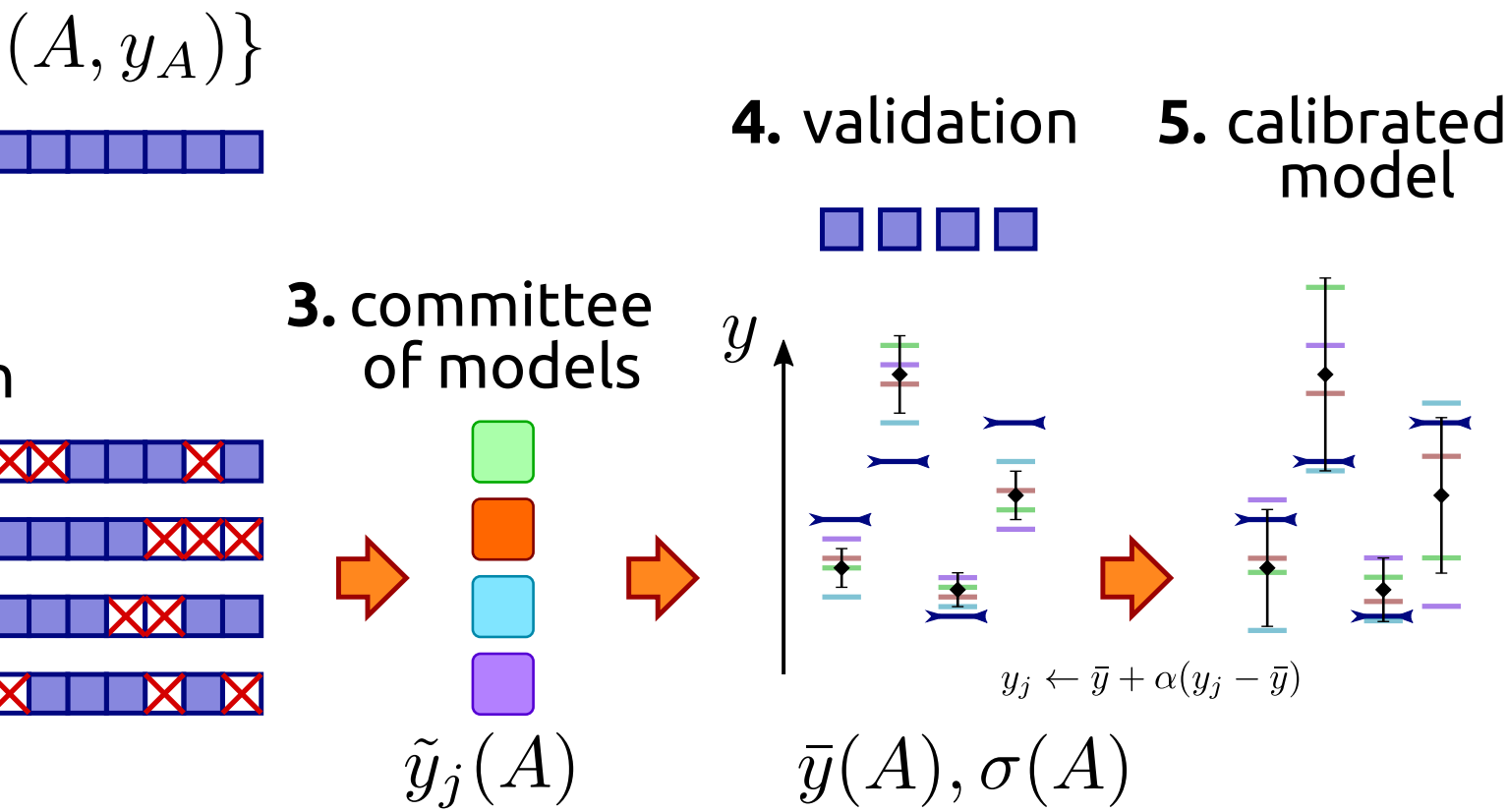
Musil et al., JCTC (2019)

Calibrated committee models

- Ensemble of N_{RS} models, trained on subsets of the train set

$$\bar{y}(A) = \frac{1}{N_{RS}} \sum_j \tilde{y}_j(A), \qquad \sigma^2(A) = \frac{1}{N_{RS}-1} \sum_i \left(\tilde{y}_j(A) - \bar{y}(A)\right)^2$$

- Verify accuracy by the distribution of errors $P\left(|\bar{y}(A) - y_A(A)| \;\middle|\; \sigma(A)\right)$
- Adjust the spread by maximum likelihood $\tilde{y}_j \leftarrow \bar{y} + \alpha\left(\tilde{y}_j - \bar{y}\right)$
- Calibrated model can be used for easy uncertainty propgation



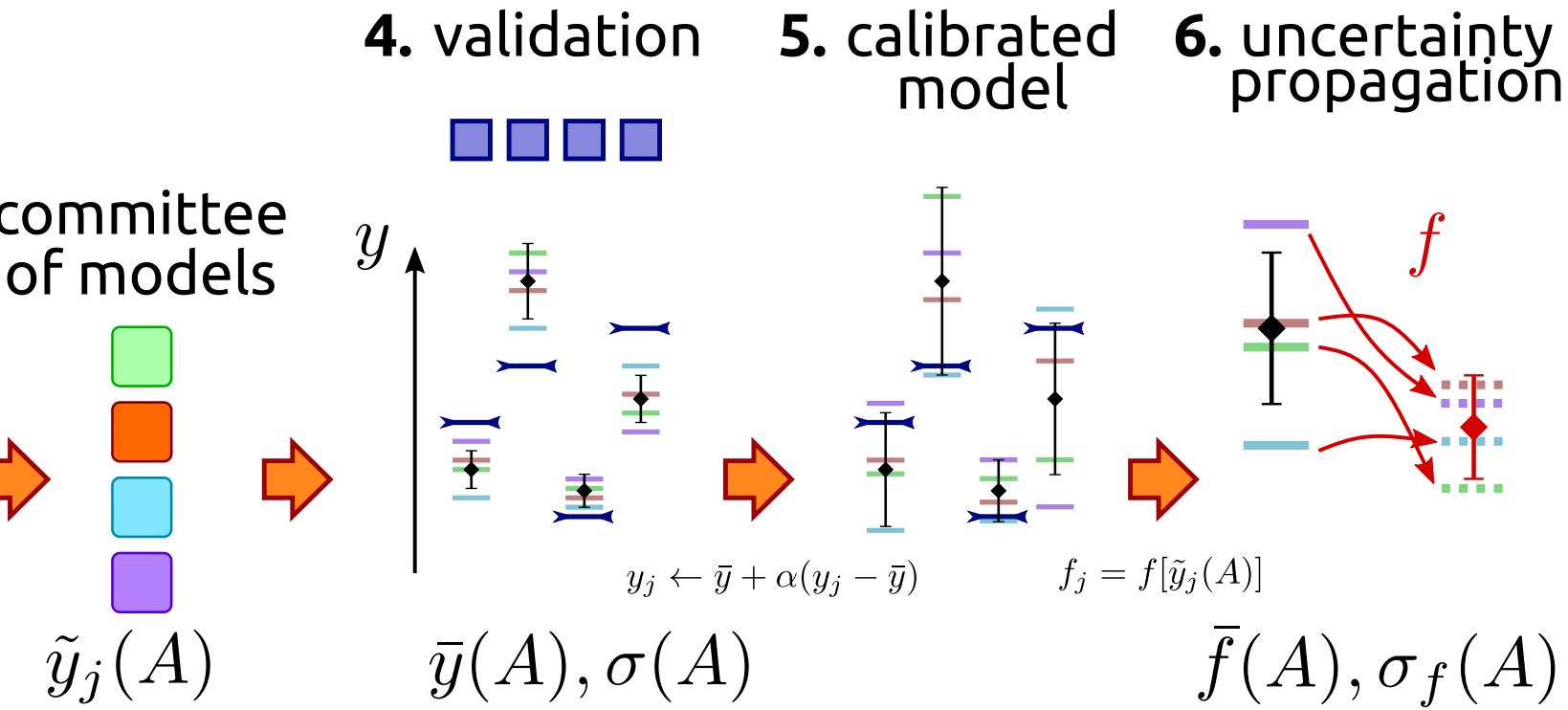
Musil et al., JCTC (2019)

Calibrated committee models

- Ensemble of N_{RS} models, trained on subsets of the train set

$$\bar{y}(A) = \frac{1}{N_{RS}} \sum_j \tilde{y}_j(A), \qquad \sigma^2(A) = \frac{1}{N_{RS}-1} \sum_i \left(\tilde{y}_j(A) - \bar{y}(A)\right)^2$$

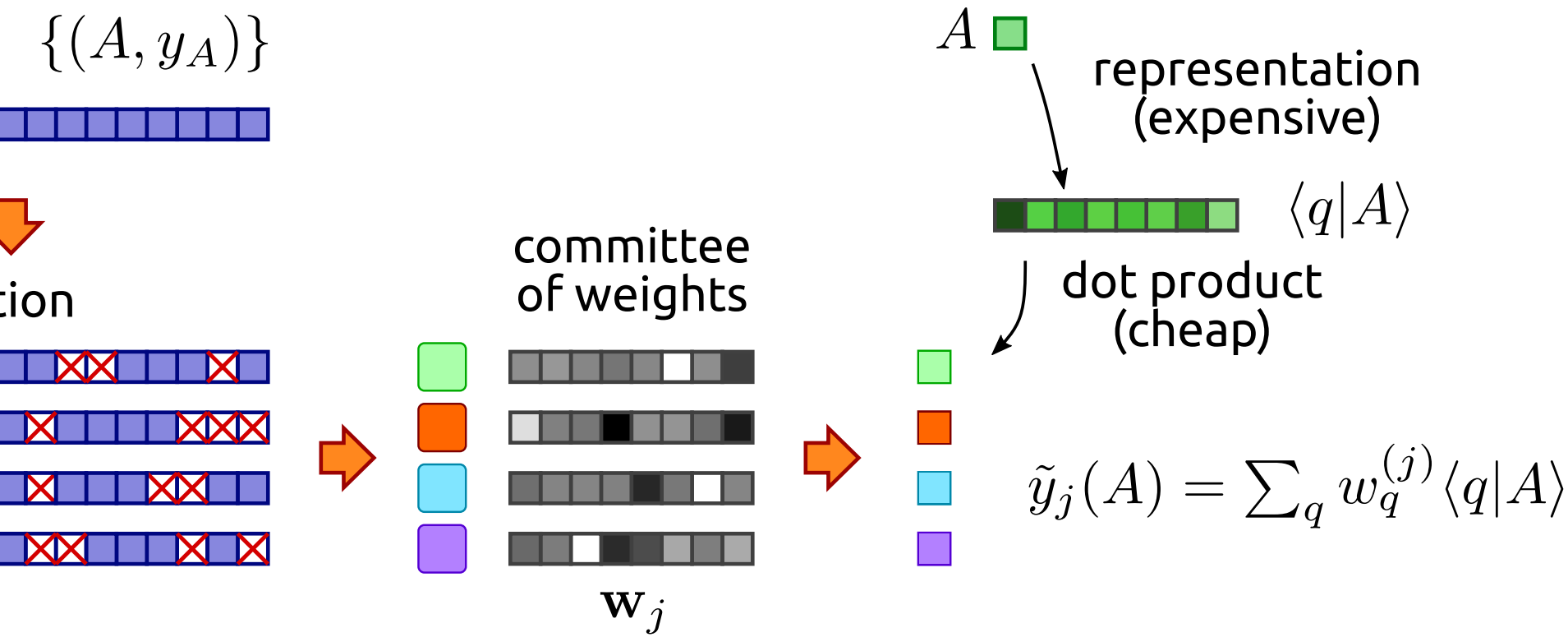
- Verify accuracy by the distribution of errors $P\left(|\bar{y}(A) - y_A(A)| \;\; |\sigma(A)\right)$
- Adjust the spread by maximum likelihood $\tilde{y}_j \leftarrow \bar{y} + \alpha\left(\tilde{y}_j - \bar{y}\right)$
- Calibrated model can be used for easy uncertainty propgation



Musil et al., JCTC (2019)

Uncertainty estimation on a shoestring

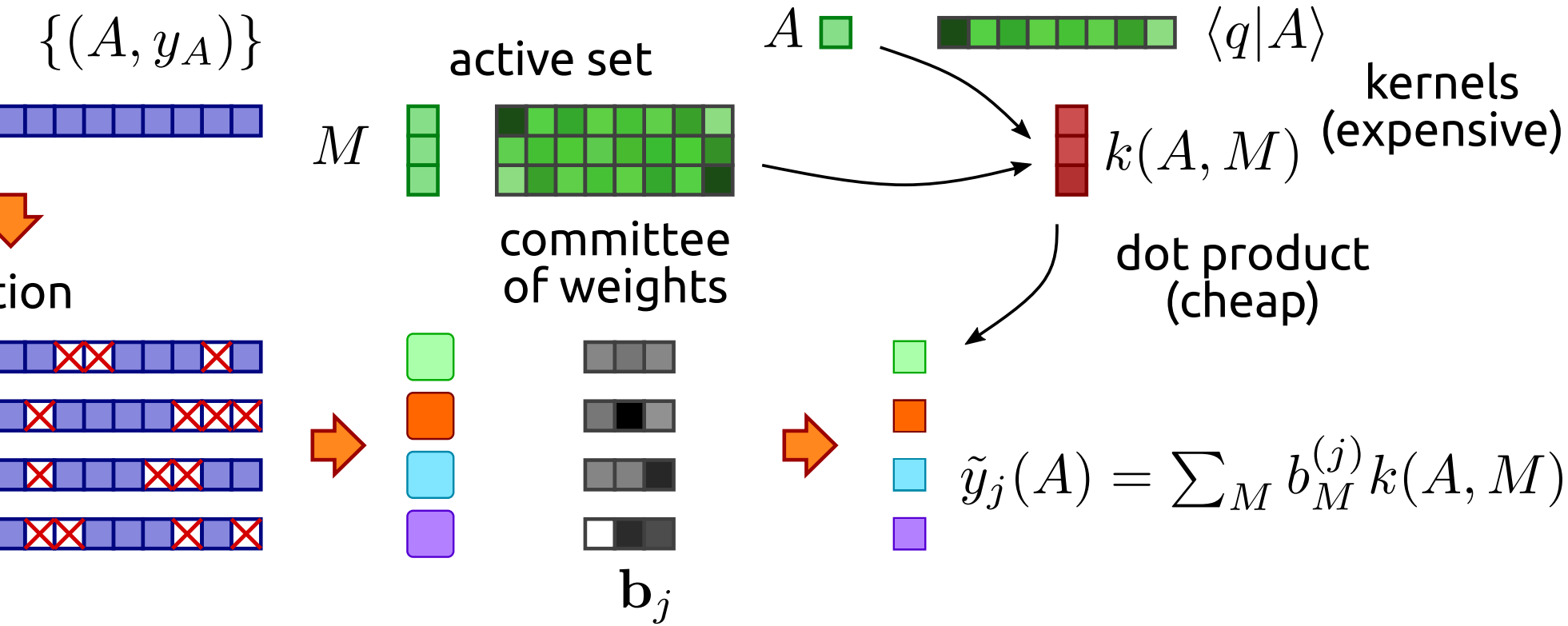
- Most expensive step is usually features/kernels evaluation. Generating committee by resampling allows inexpensive error estimation
- Same idea applies to linear & (sparse) kernel models
- Extension to NNs by only re-training output layer



Musil et al., JCTC (2019)

Uncertainty estimation on a shoestring

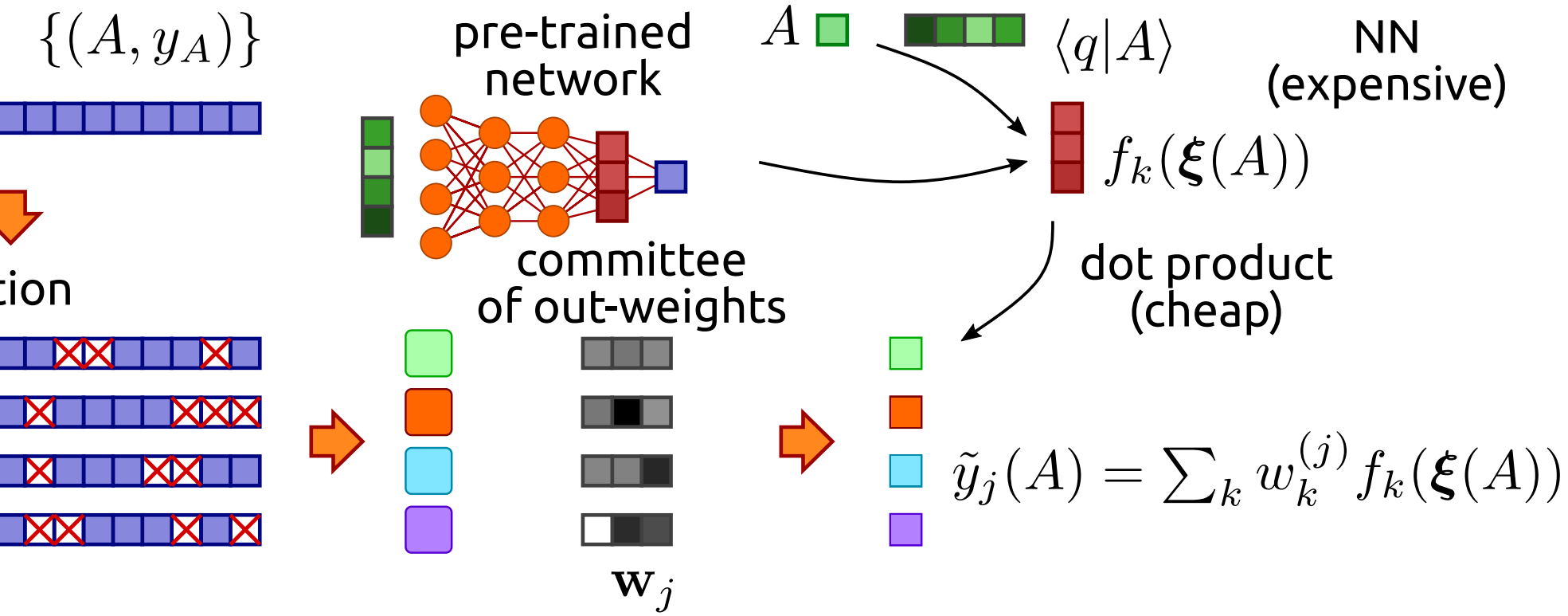
- Most expensive step is usually features/kernels evaluation. Generating committee by resampling allows inexpensive error estimation
- Same idea applies to linear & (sparse) kernel models
- Extension to NNs by only re-training output layer



Musil et al., JCTC (2019)

Uncertainty estimation on a shoestring

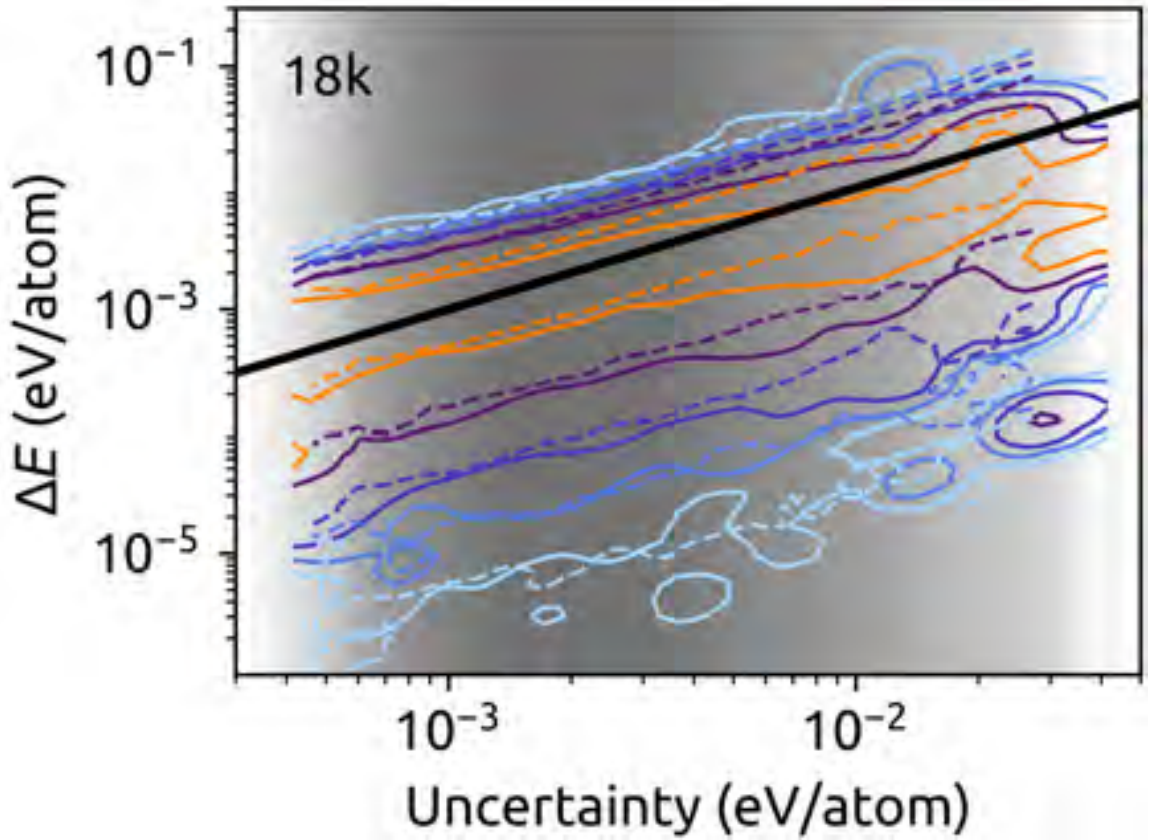
- Most expensive step is usually features/kernels evaluation. Generating committee by resampling allows inexpensive error estimation
- Same idea applies to linear & (sparse) kernel models
- Extension to NNs by only re-training output layer



Musil et al., JCTC (2019)

Uncertainty estimation in action

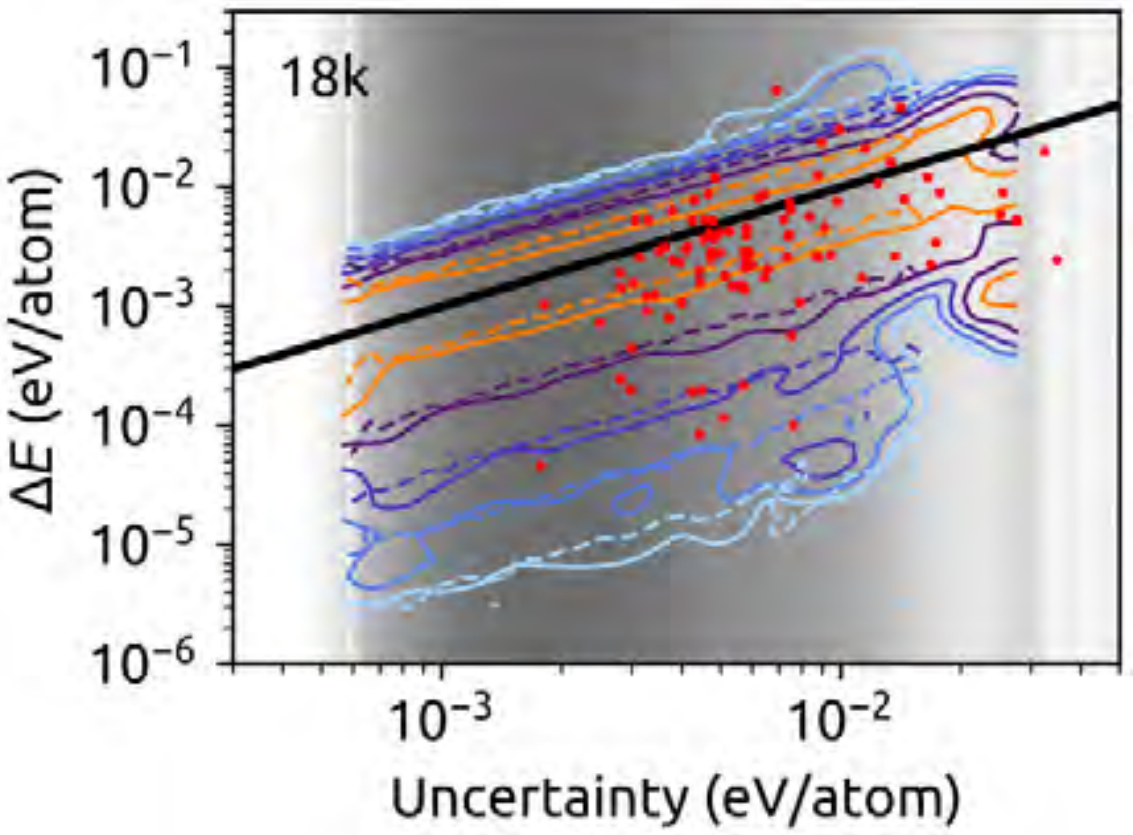
- Error vs uncertainty for the atomization energies of QM9
  - Uncertainty $\neq$ error! Spread of predictions is what matters
  - "Rejected" QM9 structures are highly uncertain
- Raman spectra with uncertainty propagation



Musil, Willatt, **MC** JCTC (2019)

Uncertainty estimation in action

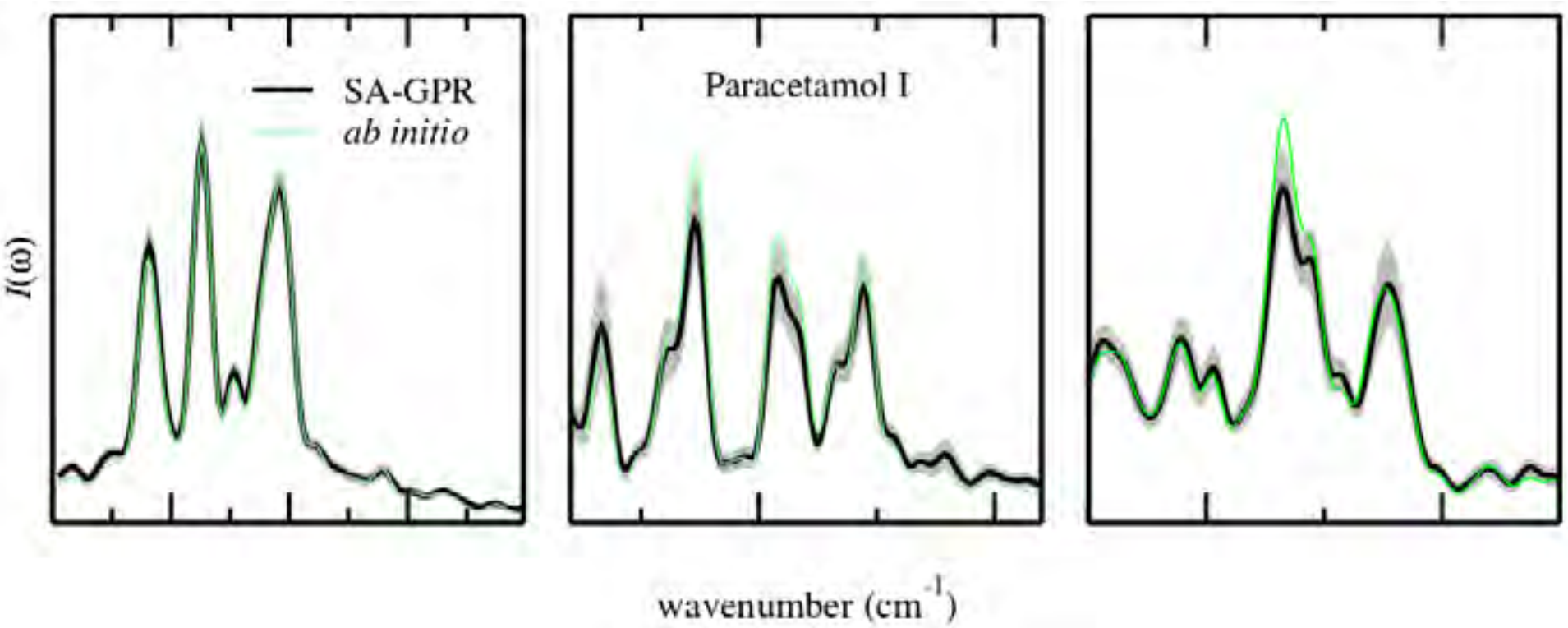
- Error vs uncertainty for the atomization energies of QM9
  - Uncertainty $\neq$ error! Spread of predictions is what matters
  - "Rejected" QM9 structures are highly uncertain
- Raman spectra with uncertainty propagation



Musil, Willatt, **MC** JCTC (2019)

Uncertainty estimation in action

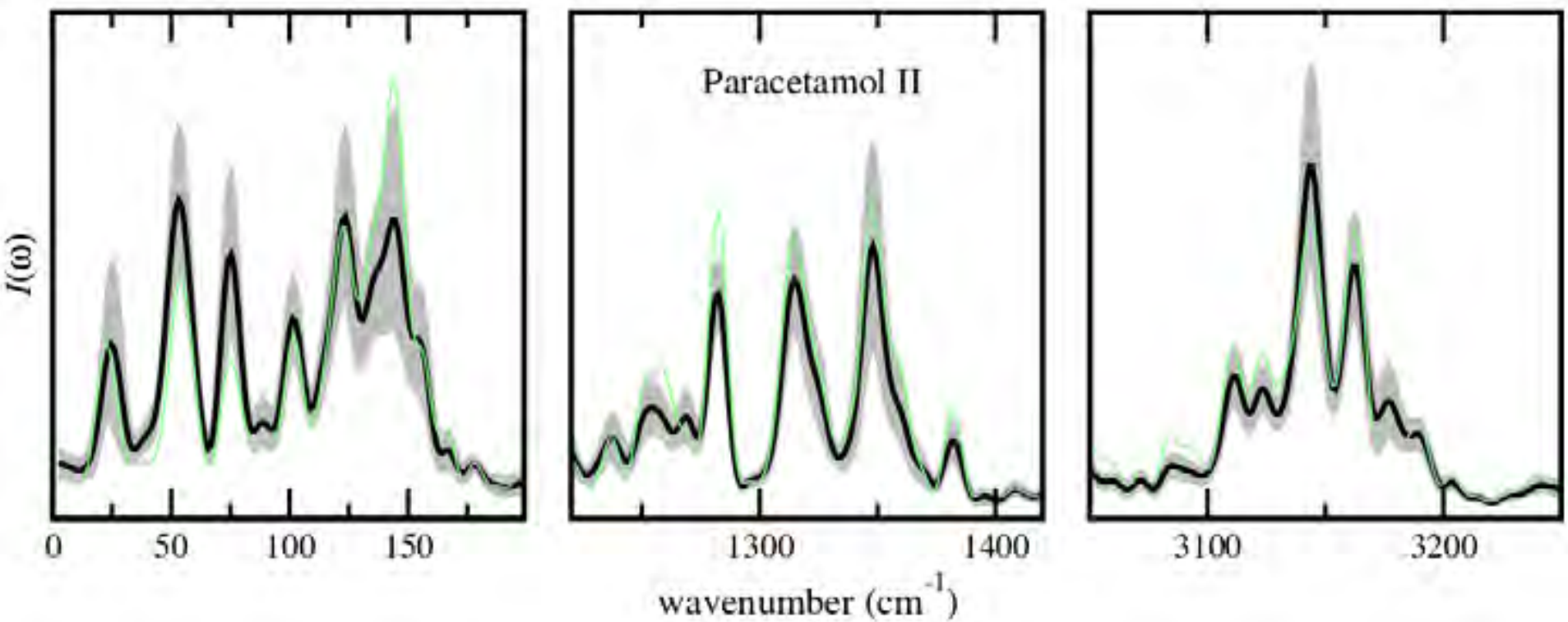
- Error vs uncertainty for the atomization energies of QM9
  - Uncertainty $\neq$ error! Spread of predictions is what matters
  - "Rejected" QM9 structures are highly uncertain
- Raman spectra with uncertainty propagation



N. Raimbault, A. Grisafi, **MC**, M. Rossi, New J. Phys. (2019)

Uncertainty estimation in action

- Error vs uncertainty for the atomization energies of QM9
  - Uncertainty $\neq$ error! Spread of predictions is what matters
  - "Rejected" QM9 structures are highly uncertain
- Raman spectra with uncertainty propagation

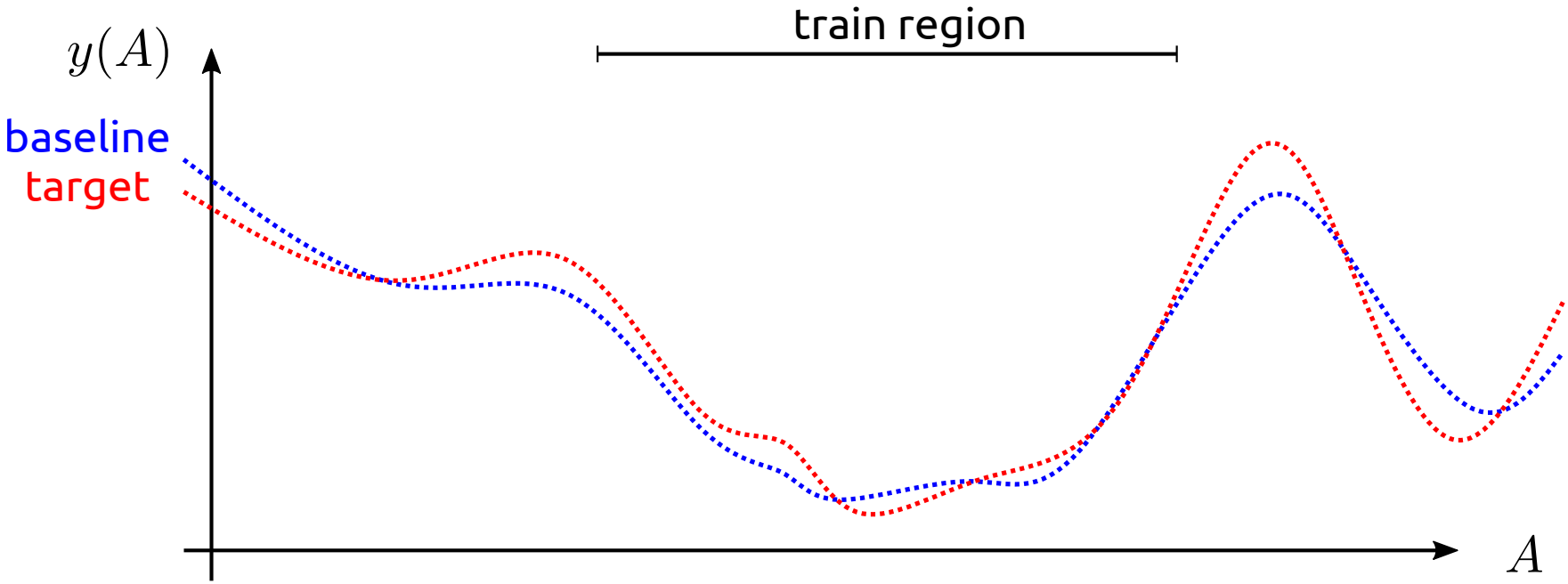


N. Raimbault, A. Grisafi, **MC**, M. Rossi, New J. Phys. (2019)

Weighted baseline models

- ML potential as a correction to a (semi)empirical baseline
- Use baseline error (constant σ_b) and uncertainty (structure-dependent $\sigma(A)$) to get a weighted-baseline model that avoids instability

$$V(A) = V_b(A) + \bar{V}_\delta(A)$$

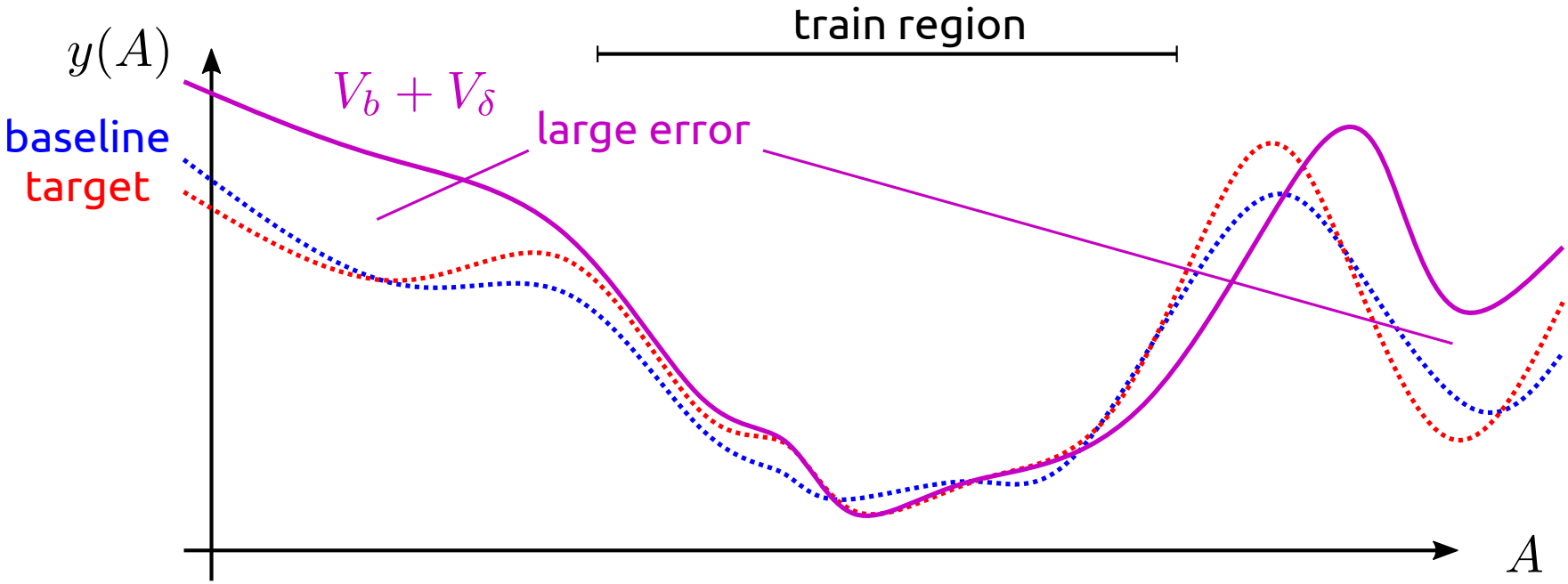


Imbalzano et al. JCP (2021)

Weighted baseline models

- ML potential as a correction to a (semi)empirical baseline
- Use baseline error (constant σ_b) and uncertainty (structure-dependent $\sigma(A)$) to get a weighted-baseline model that avoids instability

$$V(A) = V_b(A) + \bar{V}_\delta(A)$$

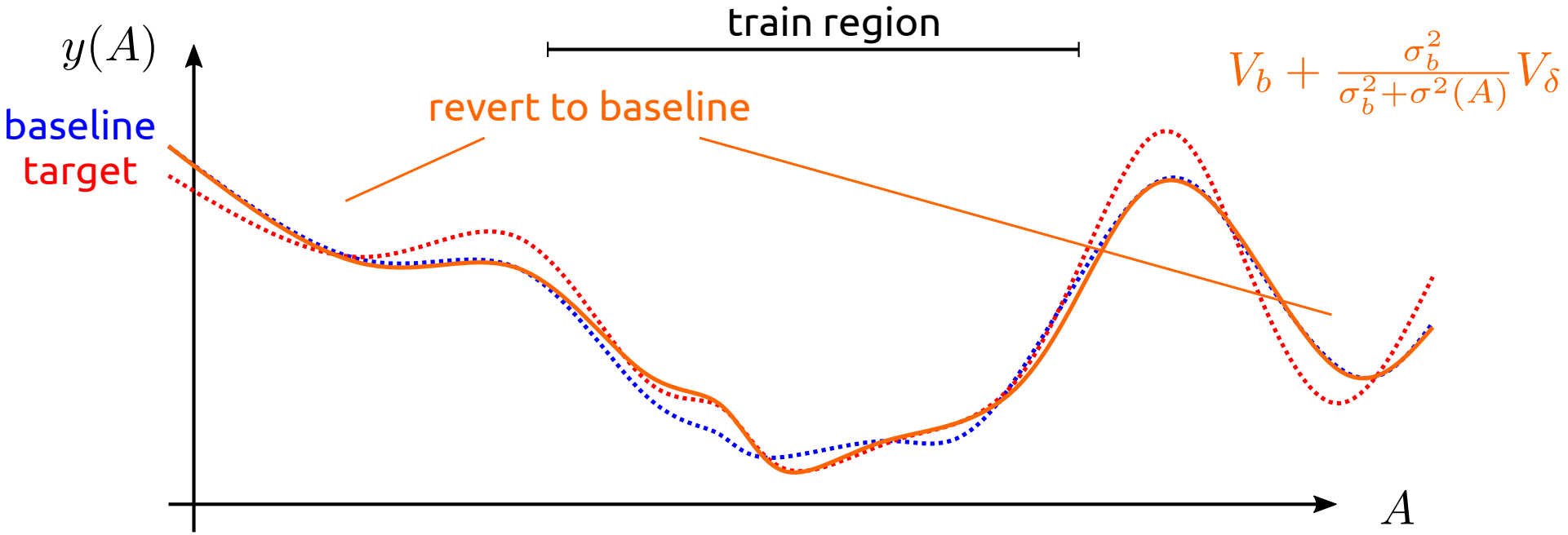


Imbalzano et al. JCP (2021)

Weighted baseline models

- ML potential as a correction to a (semi)empirical baseline
- Use baseline error (constant σ_b) and uncertainty (structure-dependent $\sigma(A)$) to get a weighted-baseline model that avoids instability

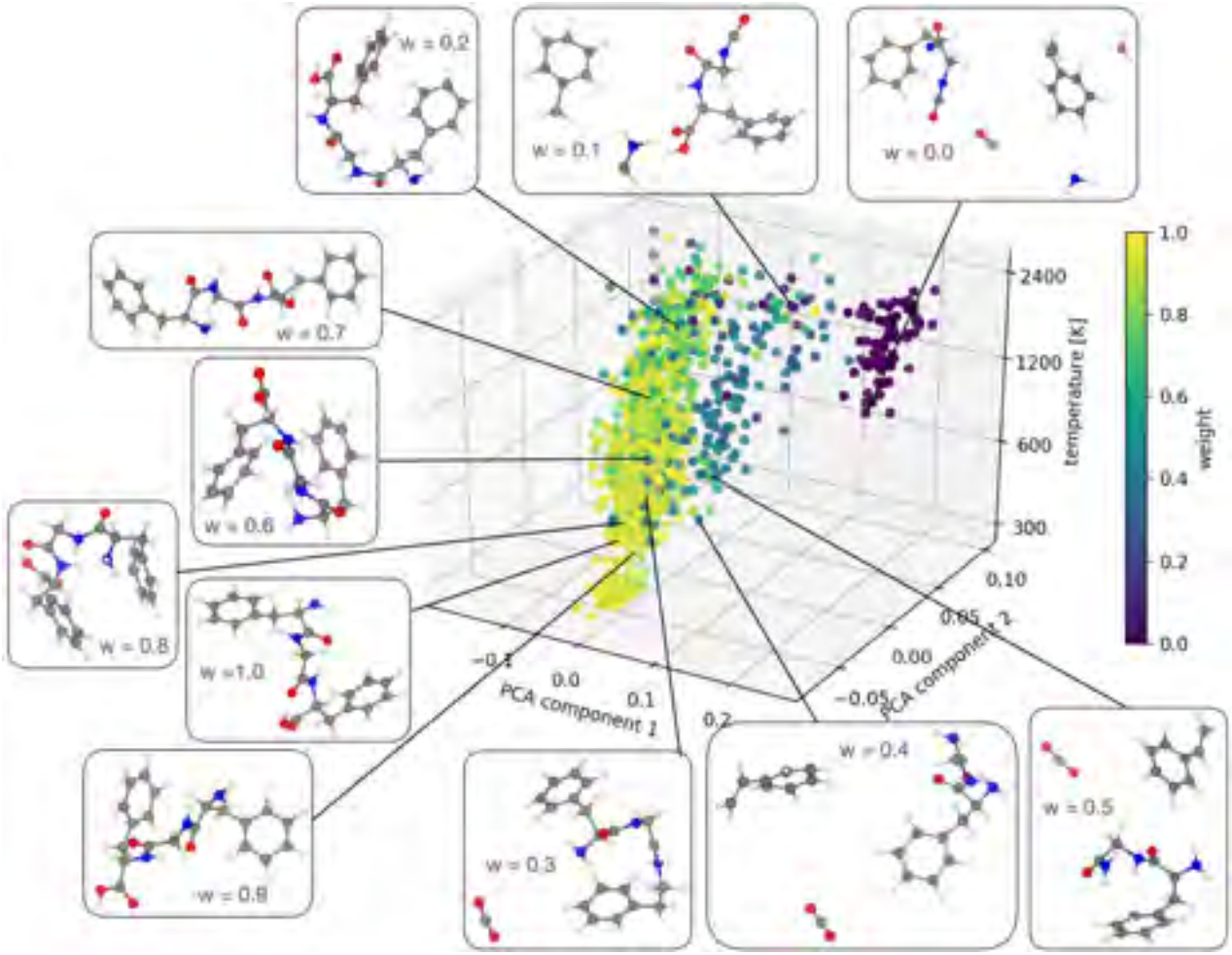
$$U(A) = \left[\frac{1}{\sigma_b^2} + \frac{1}{\sigma^2(A)}\right]^{-1} \left[\frac{1}{\sigma_b^2} V_b(A) + \frac{1}{\sigma^2(A)} \left[V_b(A) + \bar{V}_\delta(A)\right]\right]$$



Imbalzano et al. JCP (2021)

Weighted baseline models

- ML potential as a correction to a (semi)empirical baseline
- Use baseline error (constant σ_b) and uncertainty (structure-dependent $\sigma\,(A)$) to get a weighted-baseline model that avoids instability



Imbalzano et al. JCP (2021)

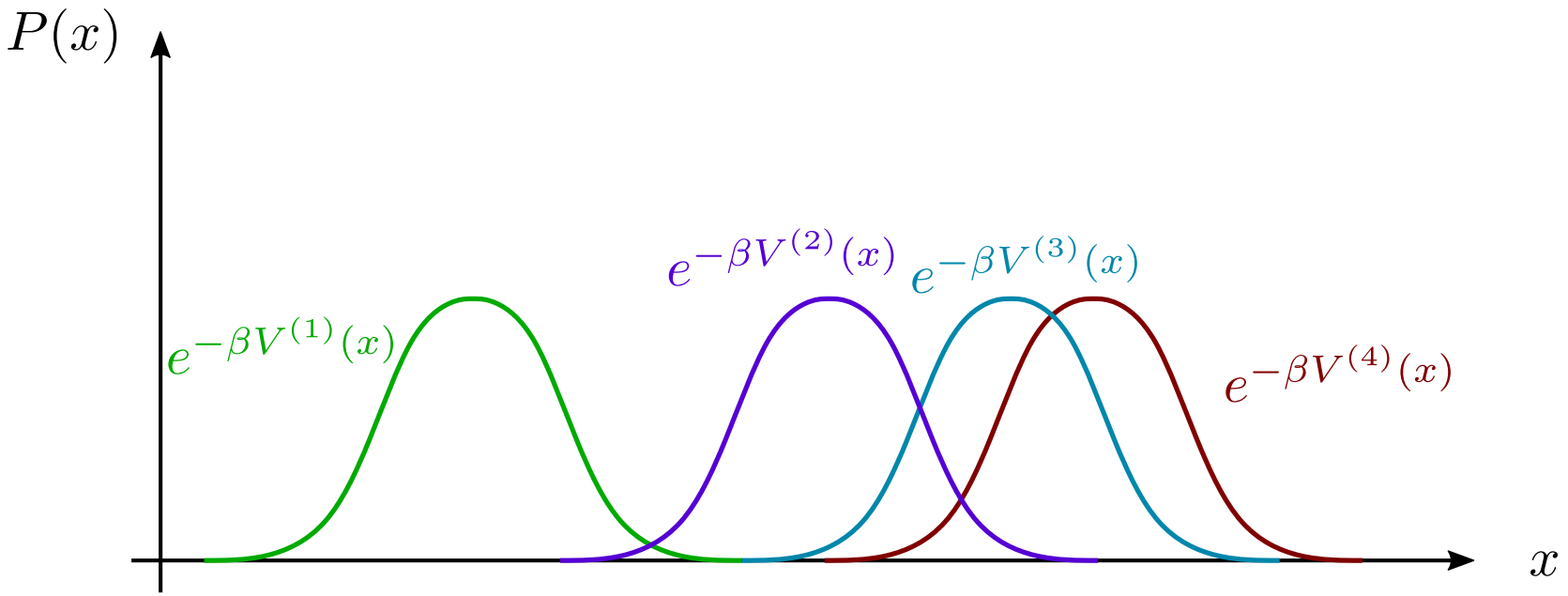
Uncertain sampling

- Errors in ML-based thermodynamic averages combine effects on the observable σ^2_a and those from sampling σ^2_{aV}
- A committee of predictions can be obtained from a single trajectory!

$$\langle a \rangle_{V^{(i)}} = \left\langle a\, e^{\beta(\bar{V}-V^{(i)})} \right\rangle_{\bar{V}}$$

- Statistically stable estimates with a Cumulant Expansion Approximation

$$\langle a \rangle_{V^{(i)},\mathrm{CEA}} \approx \langle a \rangle_{\bar{V}} - \beta \left[\left\langle a \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} - \langle a \rangle_{\bar{V}} \left\langle \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} \right]$$



Imbalzano et al. JCP (2021)

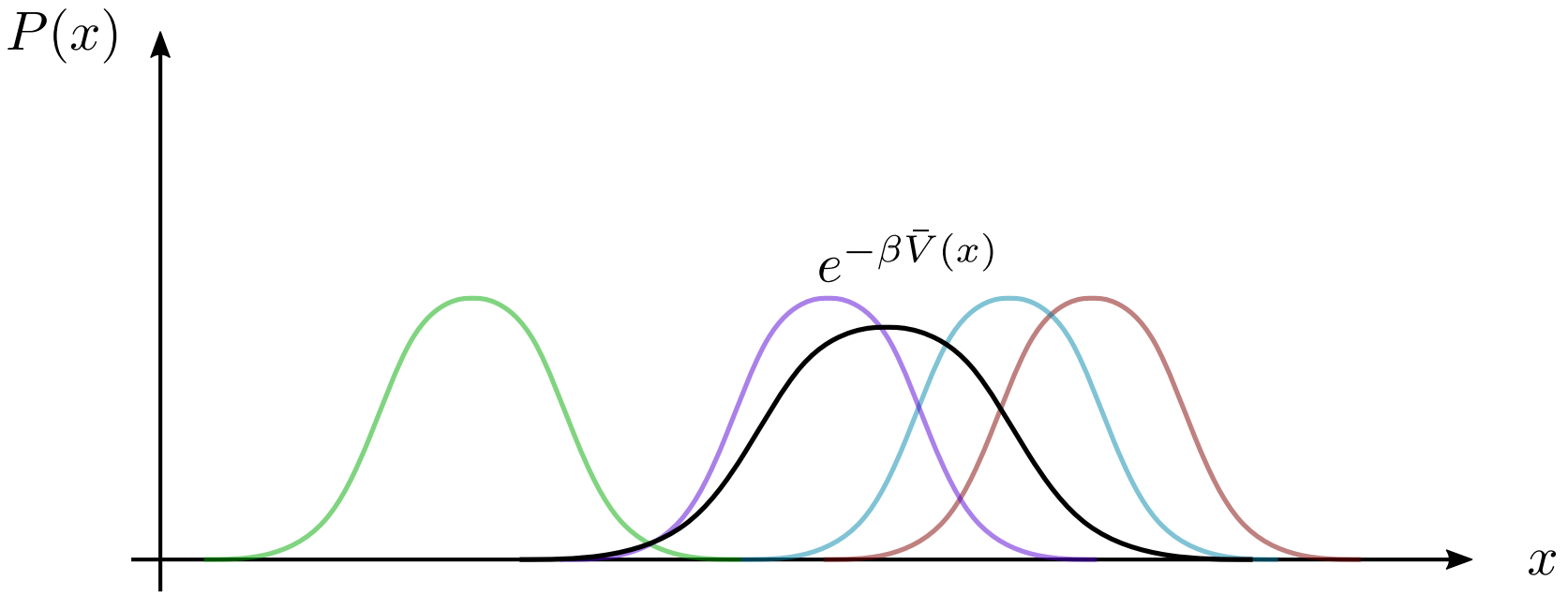
Uncertain sampling

- Errors in ML-based thermodynamic averages combine effects on the observable σ_a^2 and those from sampling σ_{aV}^2
- A committee of predictions can be obtained from a single trajectory!

$$\langle a \rangle_{V^{(i)}} = \left\langle a\, e^{\beta(\bar{V}-V^{(i)})} \right\rangle_{\bar{V}}$$

- Statistically stable estimates with a Cumulant Expansion Approximation

$$\langle a \rangle_{V^{(i)},\mathrm{CEA}} \approx \langle a \rangle_{\bar{V}} - \beta \left[\left\langle a \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} - \langle a \rangle_{\bar{V}} \left\langle \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} \right]$$



Imbalzano et al. JCP (2021)

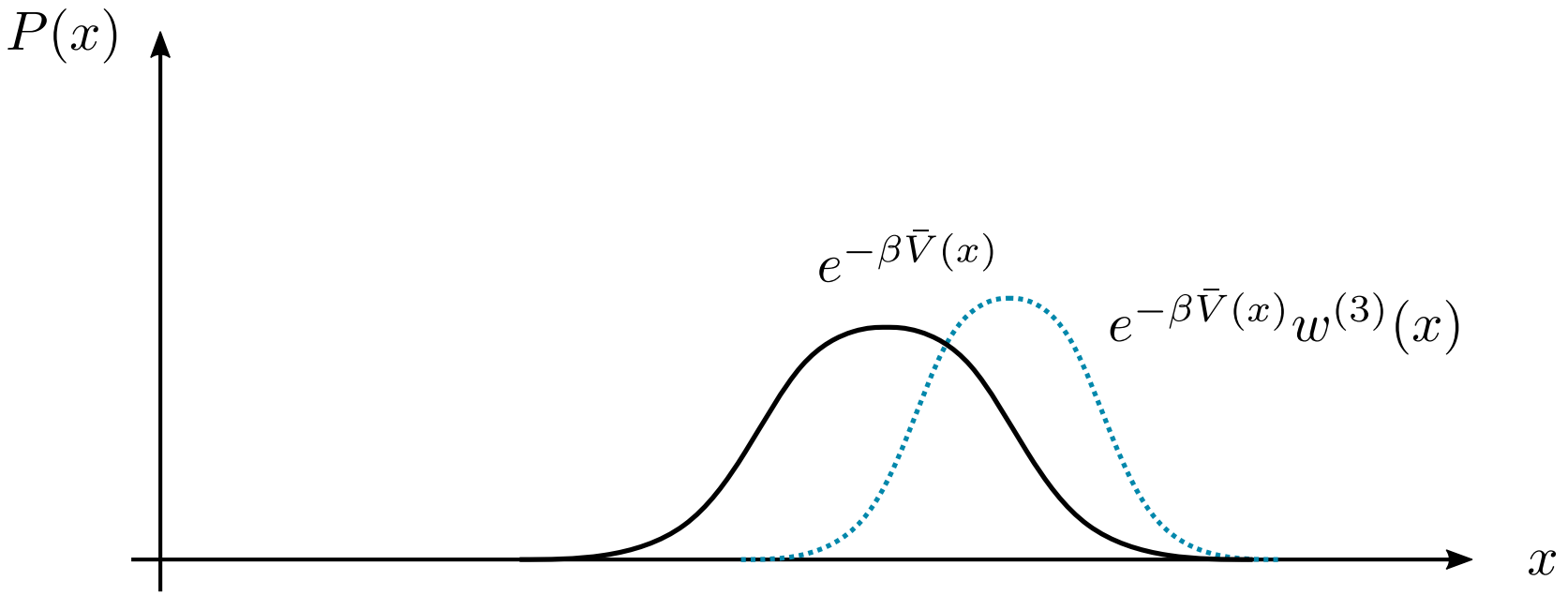
Uncertain sampling

- Errors in ML-based thermodynamic averages combine effects on the observable σ_a^2 and those from sampling σ_{aV}^2
- A committee of predictions can be obtained from a single trajectory!

$$\langle a \rangle_{V^{(i)}} = \left\langle a\, e^{\beta\left(\bar{V}-V^{(i)}\right)} \right\rangle_{\bar{V}}$$

- Statistically stable estimates with a Cumulant Expansion Approximation

$$\langle a \rangle_{V^{(i)},\mathrm{CEA}} \approx \langle a \rangle_{\bar{V}} - \beta \left[\left\langle a \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} - \langle a \rangle_{\bar{V}} \left\langle \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} \right]$$



Imbalzano et al. JCP (2021)

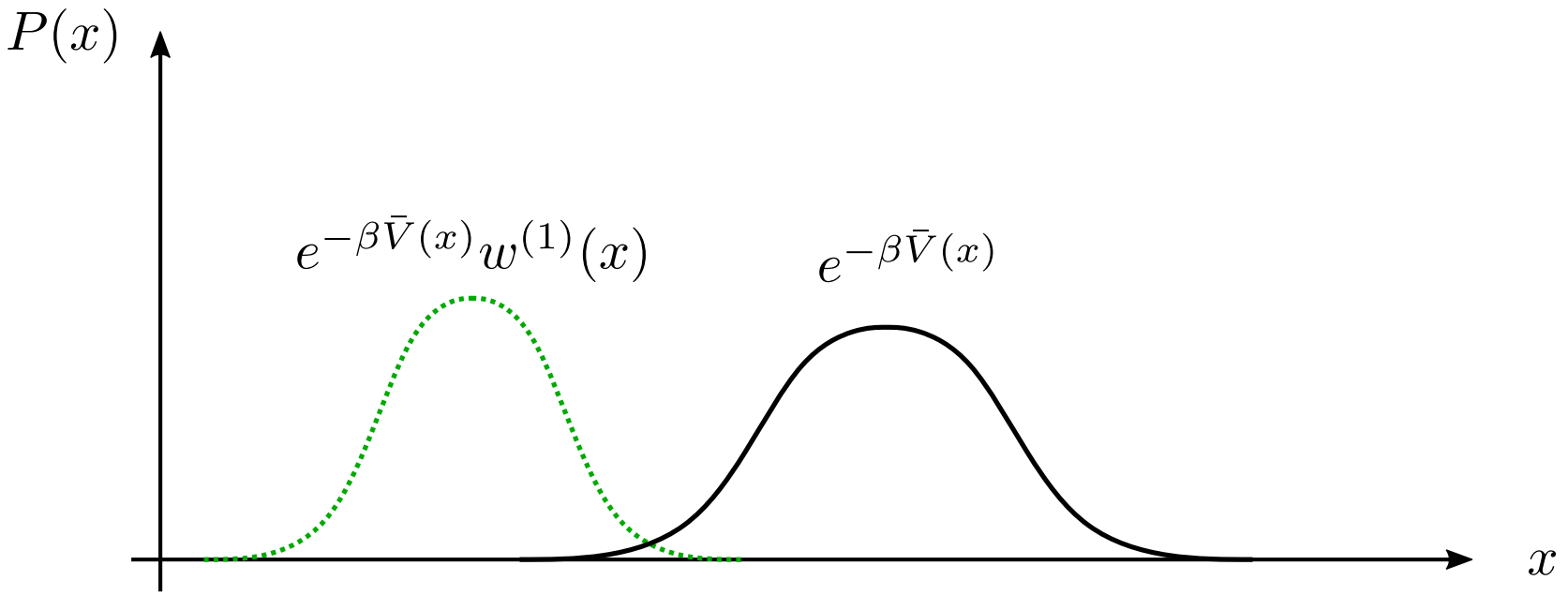
Uncertain sampling

- Errors in ML-based thermodynamic averages combine effects on the observable σ^2_a and those from sampling σ^2_{aV}
- A committee of predictions can be obtained from a single trajectory!

$$\langle a \rangle_{V^{(i)}} = \left\langle a\, e^{\beta\left(\bar{V}-V^{(i)}\right)} \right\rangle_{\bar{V}}$$

- Statistically stable estimates with a Cumulant Expansion Approximation

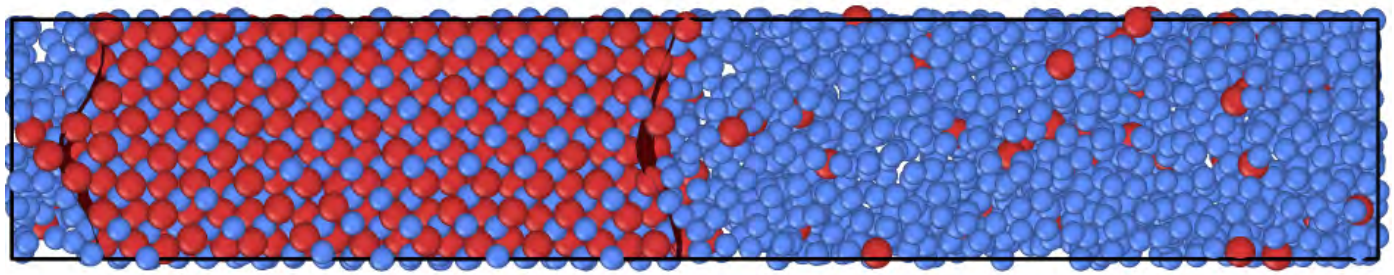
$$\langle a \rangle_{V^{(i)},\mathrm{CEA}} \approx \langle a \rangle_{\bar{V}} - \beta \left[\left\langle a \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} - \langle a \rangle_{\bar{V}} \left\langle \left(V^{(i)} - \bar{V} \right) \right\rangle_{\bar{V}} \right]$$



MC et al. PRSA (2012)

UQ for interface pinning

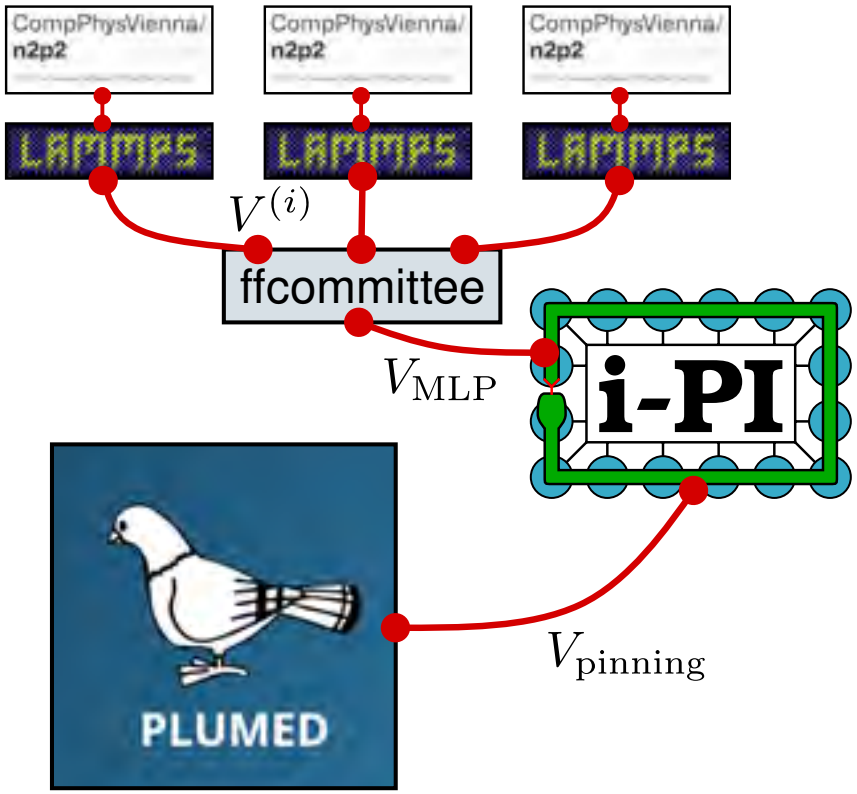
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

UQ for interface pinning

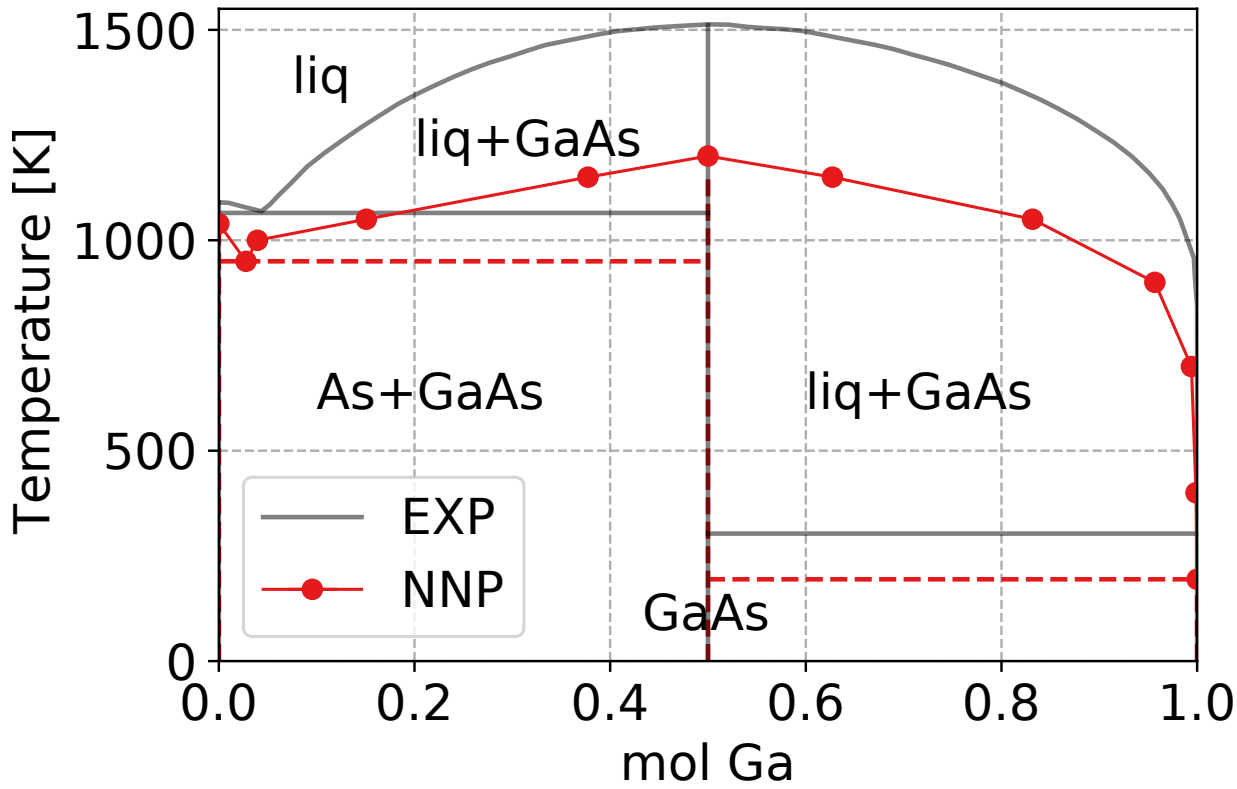
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

UQ for interface pinning

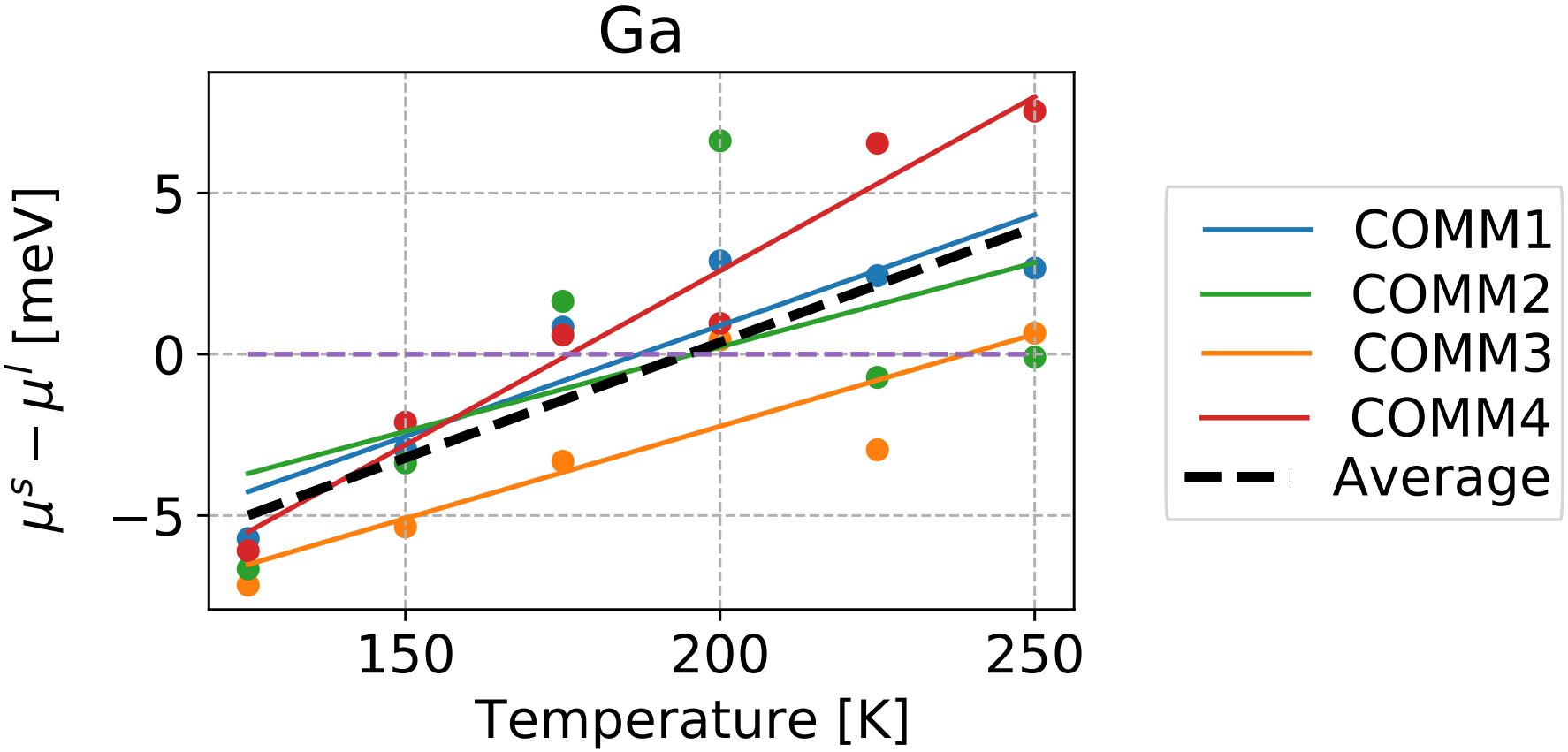
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

UQ for interface pinning

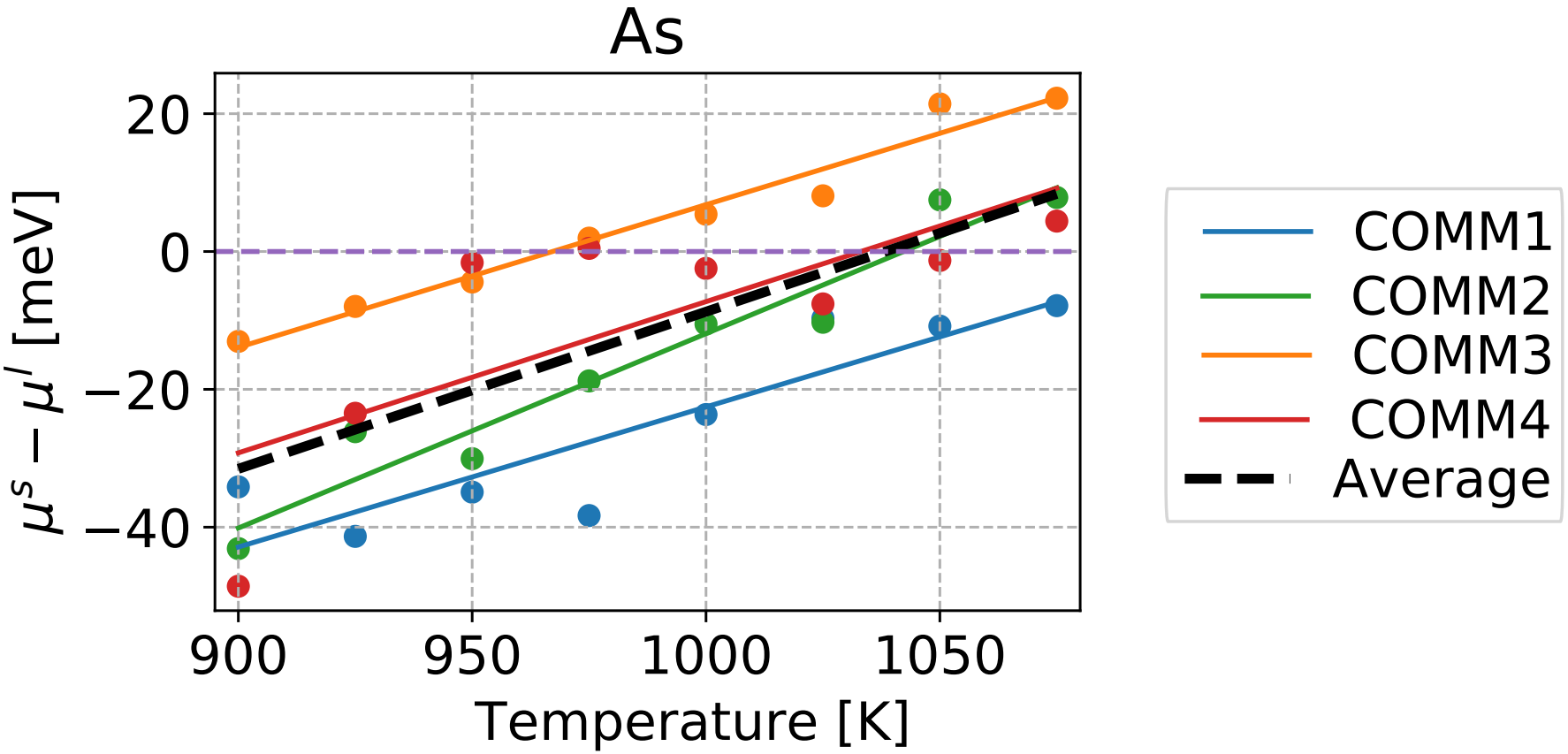
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

UQ for interface pinning

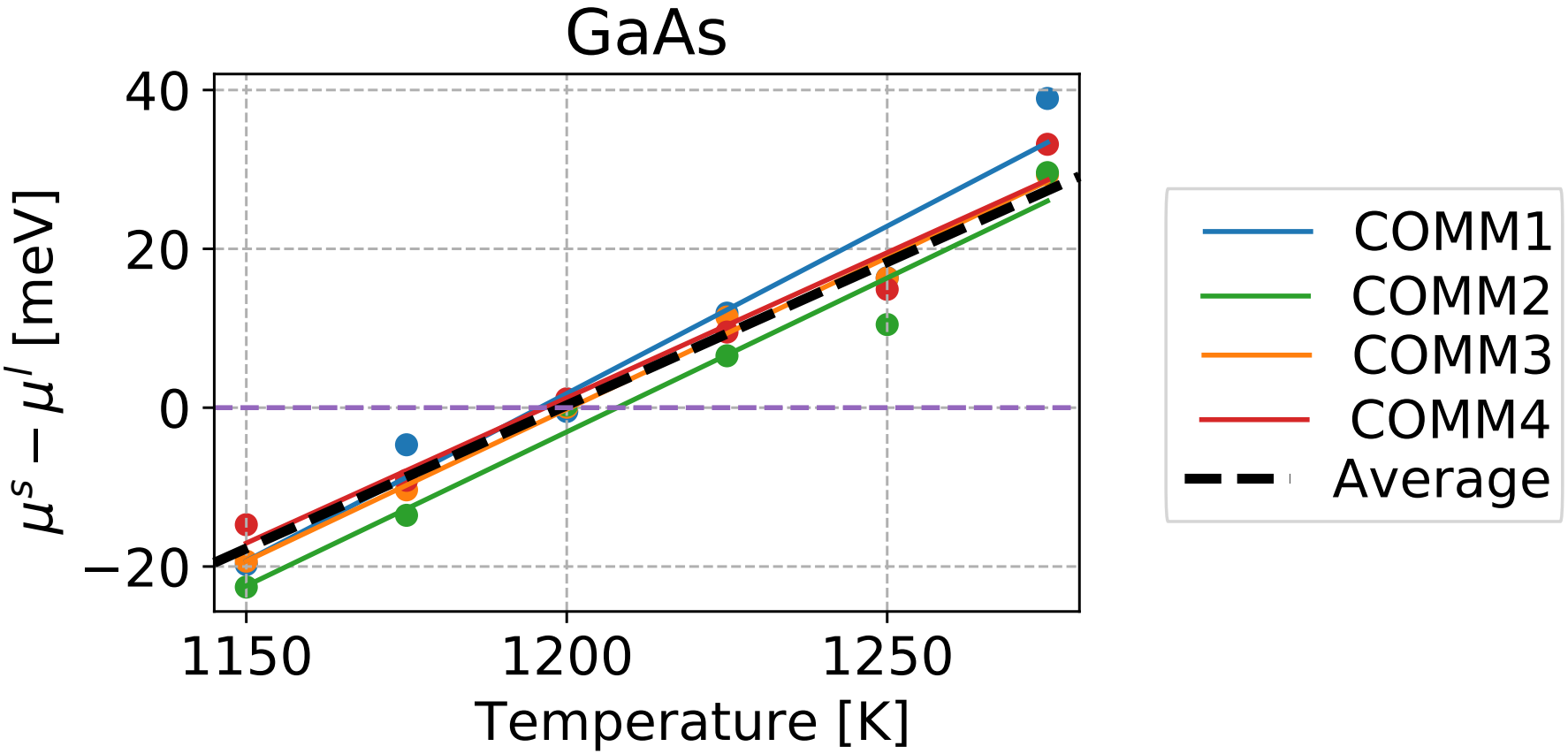
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

UQ for interface pinning

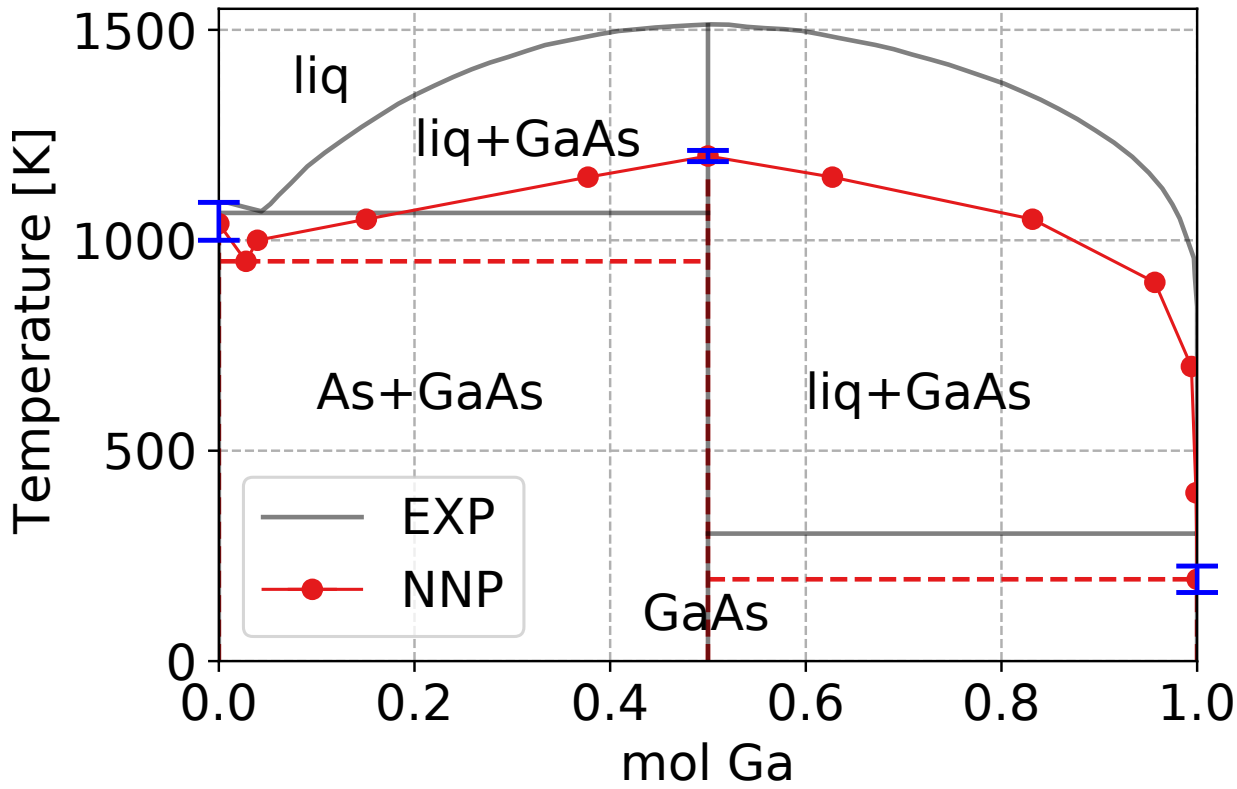
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

UQ for interface pinning

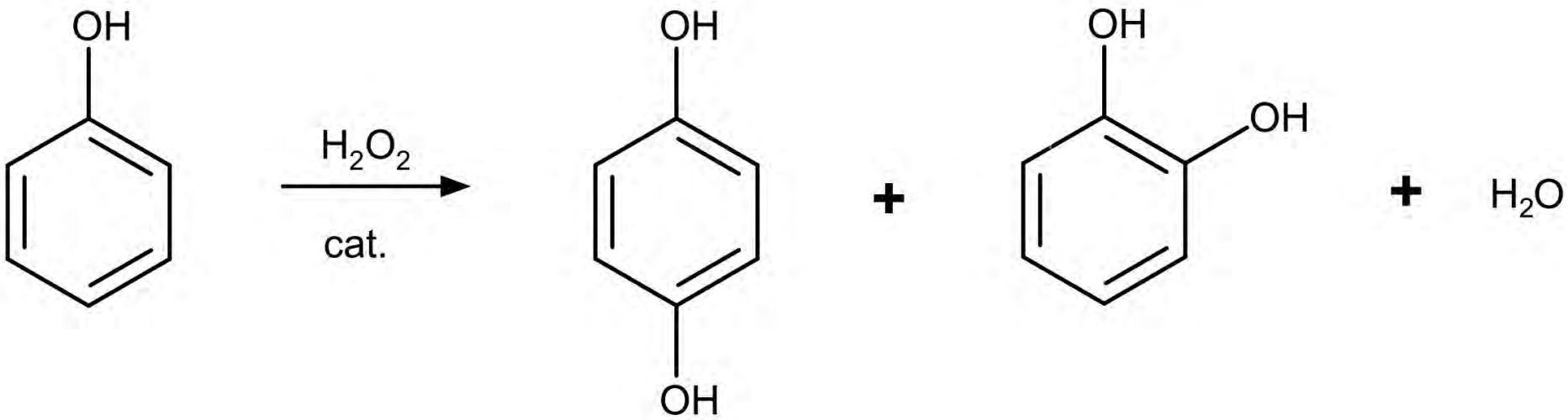
- Compute the phase diagram of Ga_xAs_{1-x}: interface-pinning simulations for a 2-component system
- DFT-accurate ML potential, i-PI+LAMMPS+PLUMED setup
- Estimate uncertainty in melting points



Imbalzano & **MC**, Phys. Rev. Materials (2021); `ipi-code.org`

Regio-selective catalysis

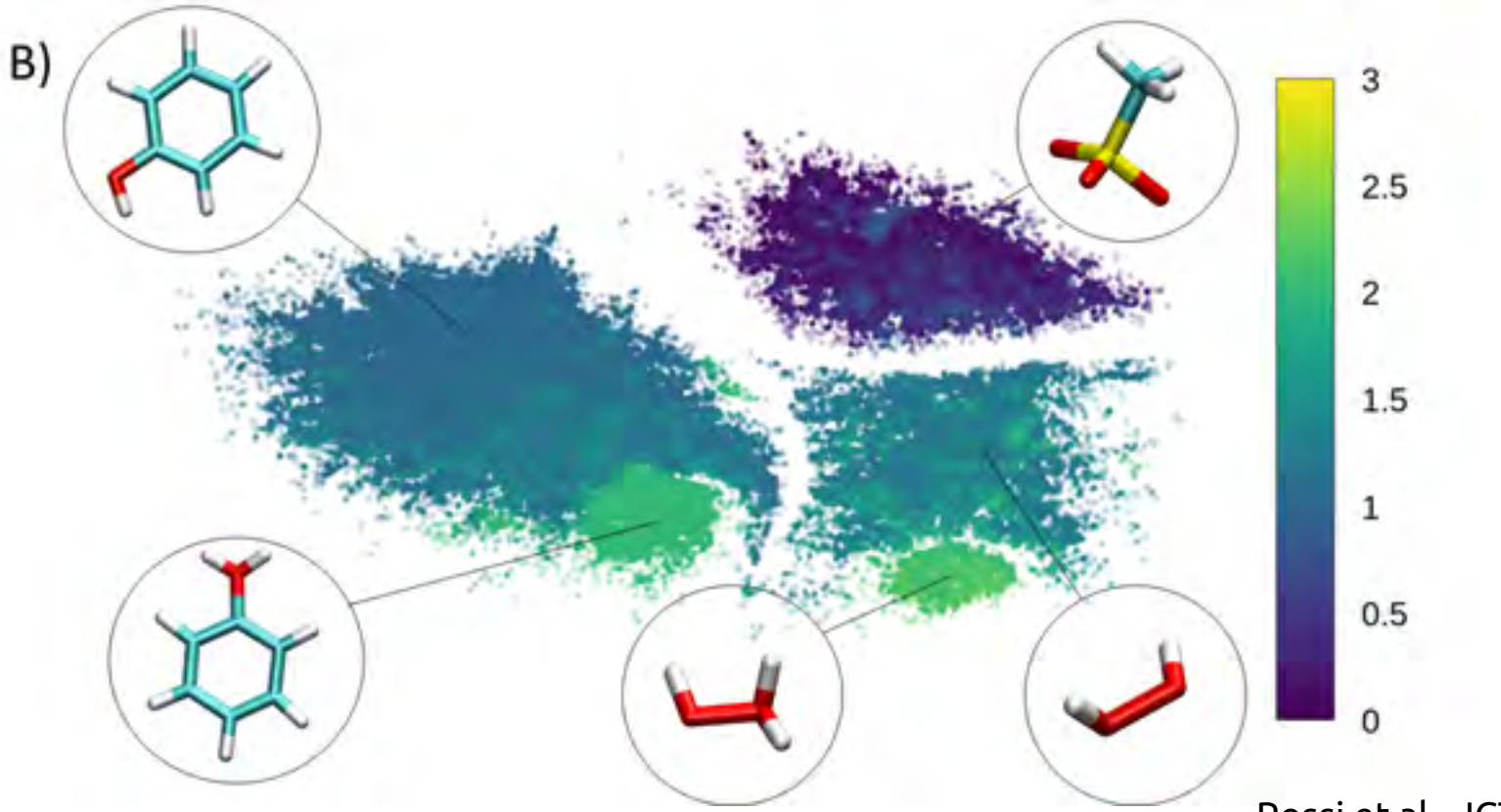
- Acid-catalyzed oxidation of phenol yields catechol and hydroquinone. Regioselectivity is poorly understood
- Very complex solution: phenol, H_2O_2, methanesulphonic acid. Explicit simulations require hybrid DFT accuracy



Rossi et al., JCTC (2020)

Regio-selective catalysis

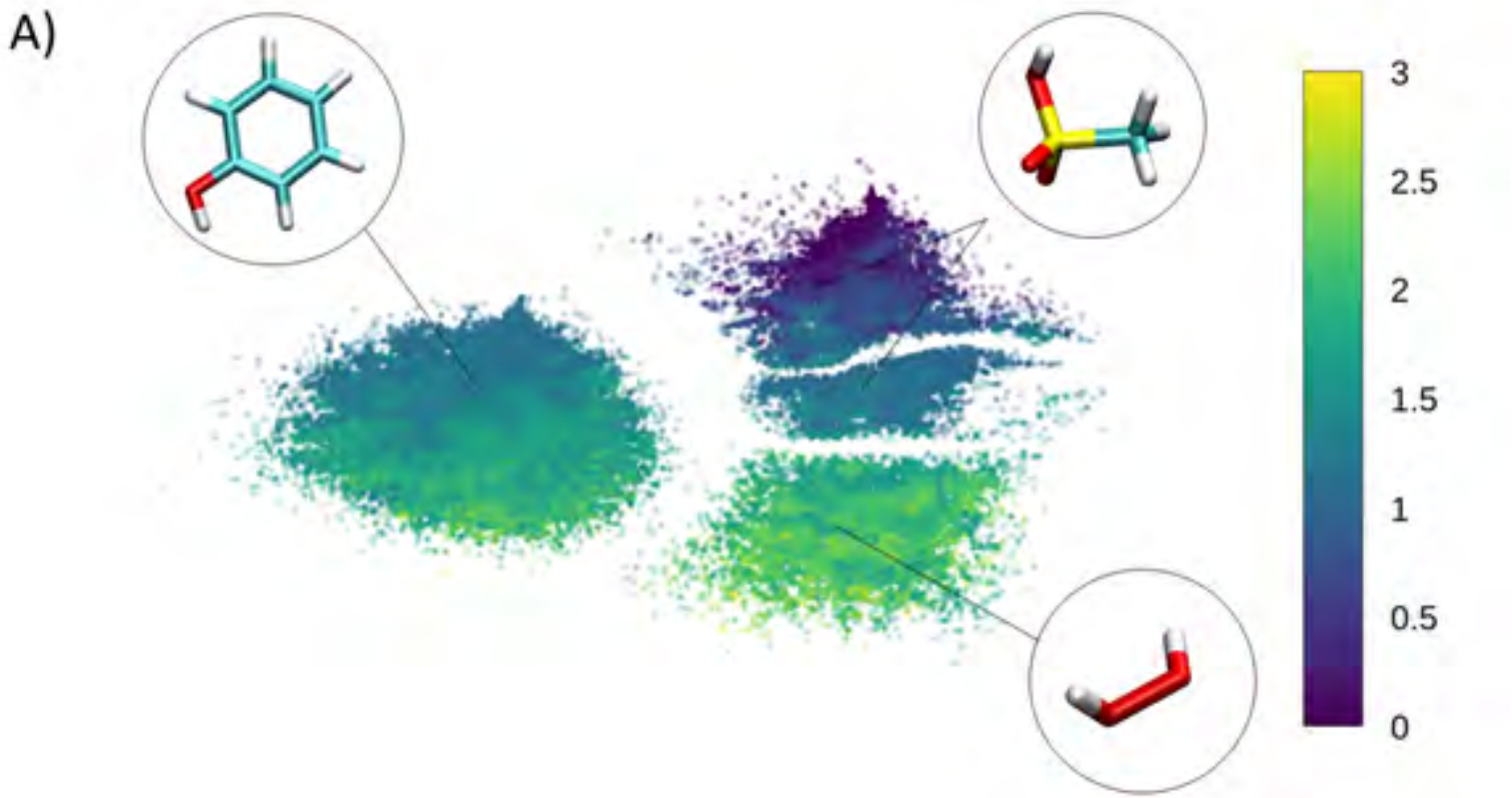
- Acid-catalyzed oxidation of phenol yields catechol and hydroquinone. Regioselectivity is poorly understood
- Very complex solution: phenol, H_2O_2, methanesulphonic acid. Explicit simulations require hybrid DFT accuracy



Rossi et al., JCTC (2020)

Regio-selective catalysis

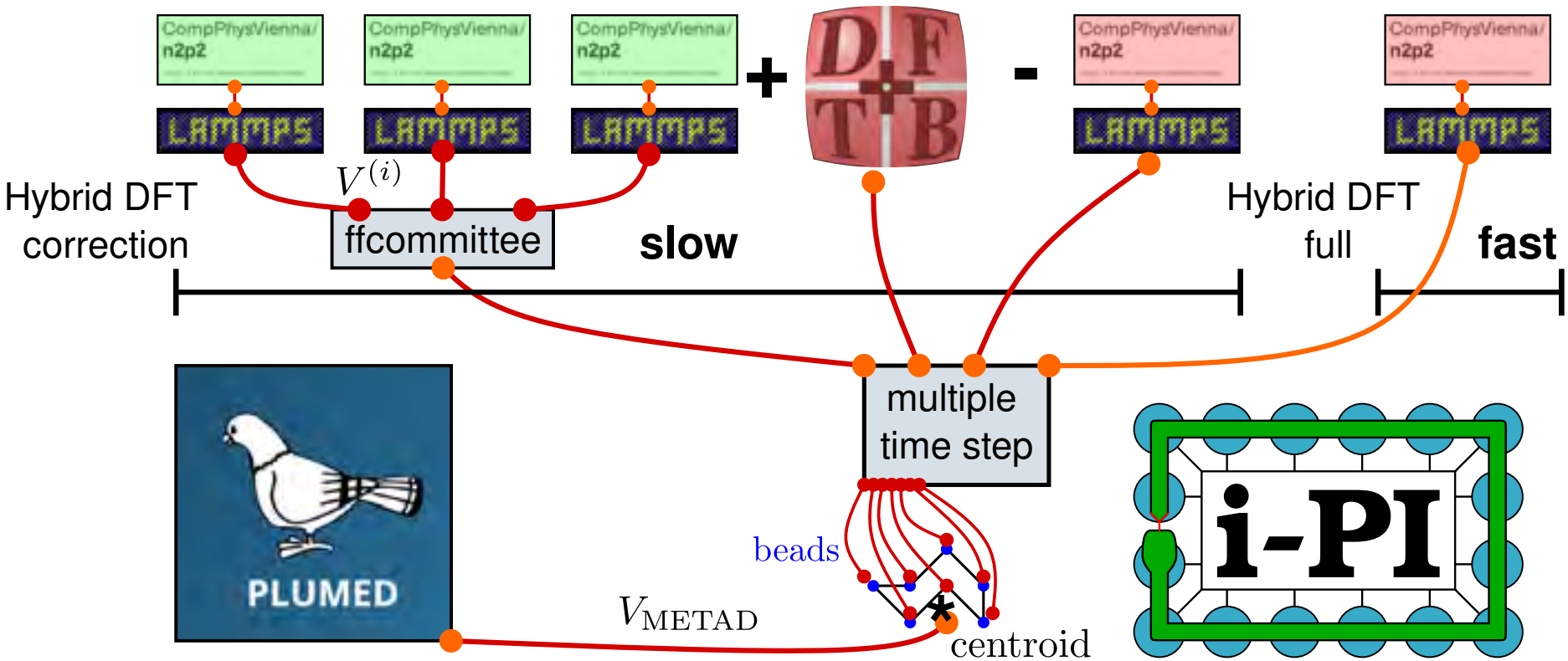
- Acid-catalyzed oxidation of phenol yields catechol and hydroquinone. Regioselectivity is poorly understood
- Very complex solution: phenol, H_2O_2, methanesulphonic acid. Explicit simulations require hybrid DFT accuracy



Rossi et al., JCTC (2020)

UQ for acid-base equilibria

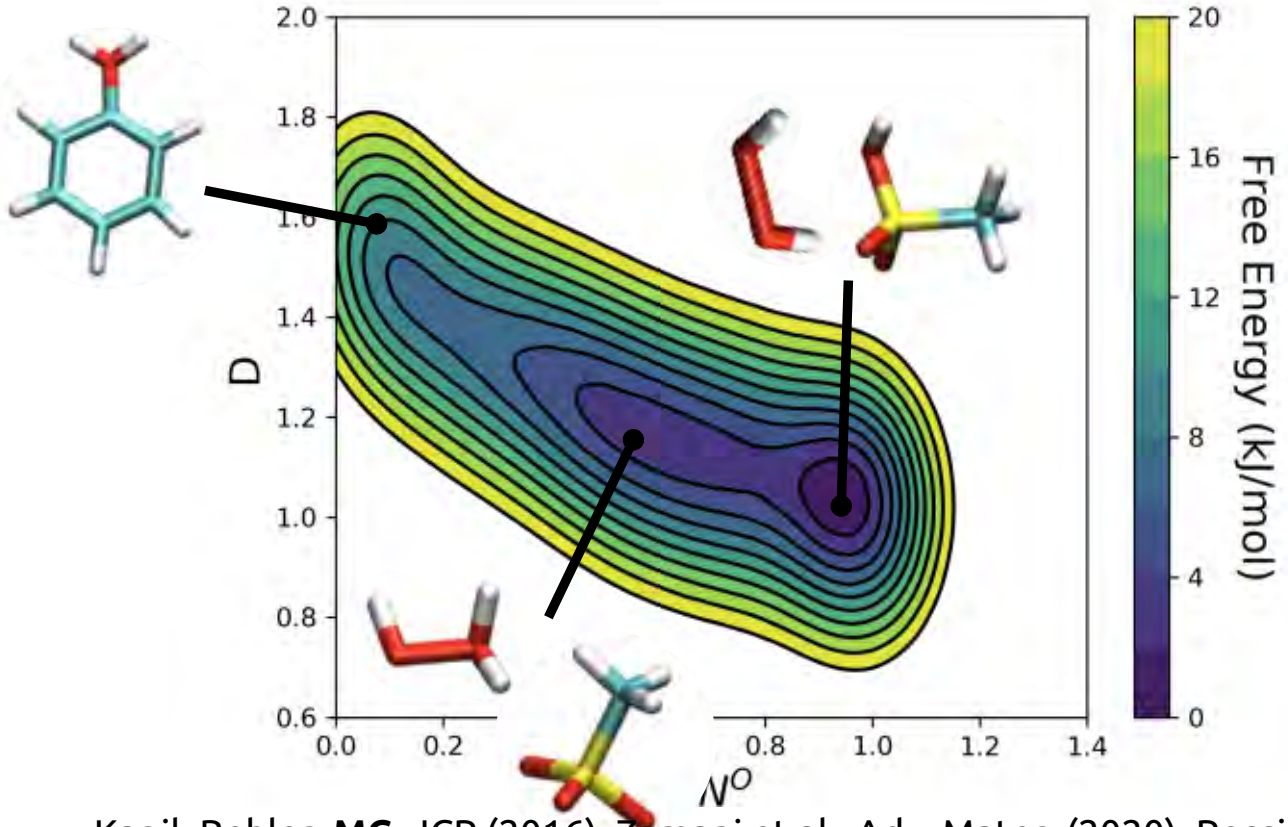
- Energetics described with a DFTB baseline and a MLP correction. Accelerated by multiple time stepping
- PIGLET thermostatting & ring-polymer contraction for the quantum sampling, Plumed-driven metadynamics applied to the centroid
- Metadynamics sampling of the dissociation of CH_3SO_2OH.
- **UQ for the free-energy profile!**



Kapil, Behler, **MC**, JCP (2016); Zamani et al., Adv. Mater. (2020); Rossi et al., JCTC (2020)

UQ for acid-base equilibria

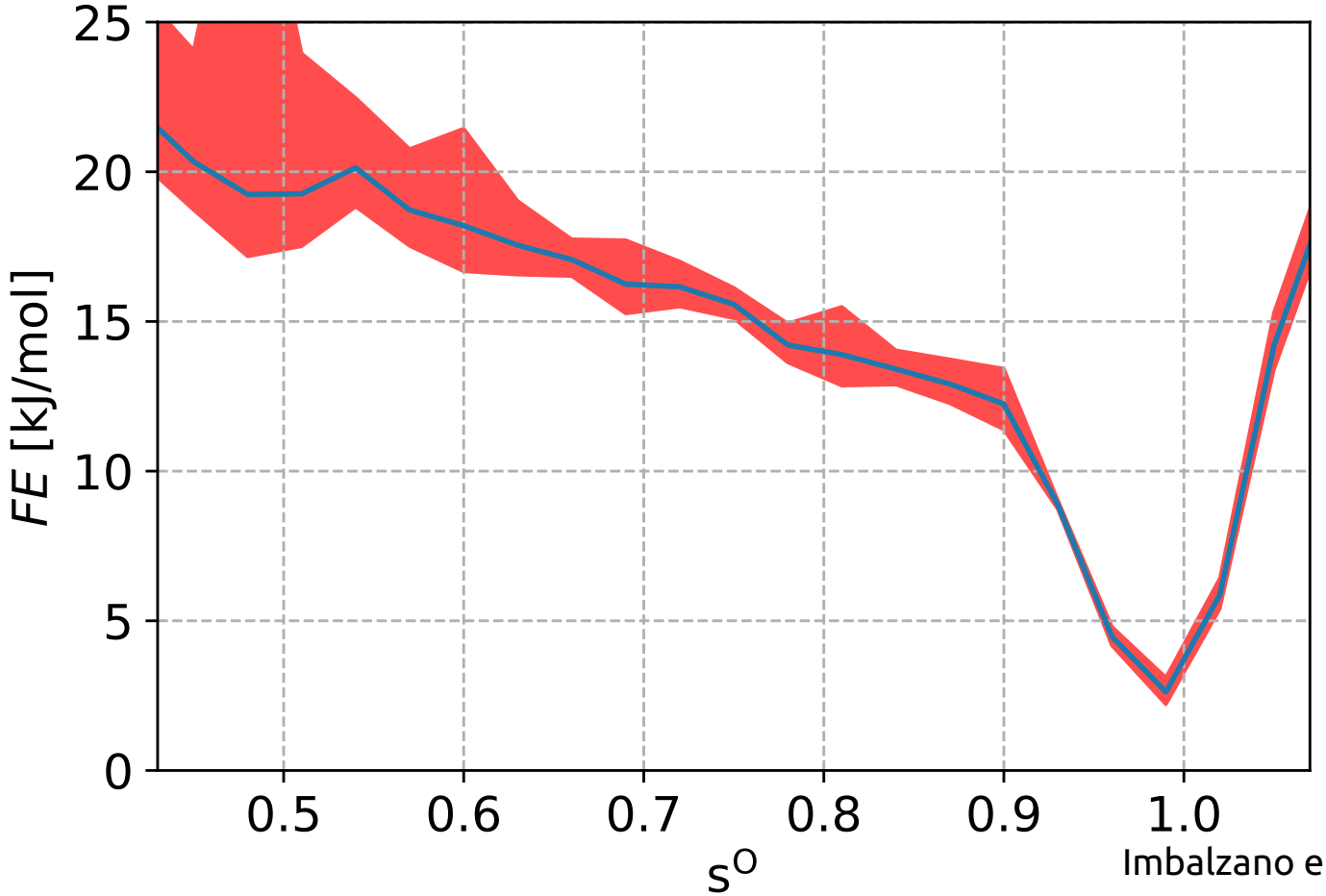
- Energetics described with a DFTB baseline and a MLP correction. Accelerated by multiple time stepping
- PIGLET thermostatting & ring-polymer contraction for the quantum sampling, Plumed-driven metadynamics applied to the centroid
- Metadynamics sampling of the dissociation of CH_3SO_2OH.
- UQ for the free-energy profile!



Kapil, Behler, **MC**, JCP (2016); Zamani et al., Adv. Mater. (2020); Rossi et al., JCTC (2020)

UQ for acid-base equilibria

- Energetics described with a DFTB baseline and a MLP correction. Accelerated by multiple time stepping
- PIGLET thermostatting & ring-polymer contraction for the quantum sampling, Plumed-driven metadynamics applied to the centroid
- Metadynamics sampling of the dissociation of CH_3SO_2OH.
- **UQ for the free-energy profile!**

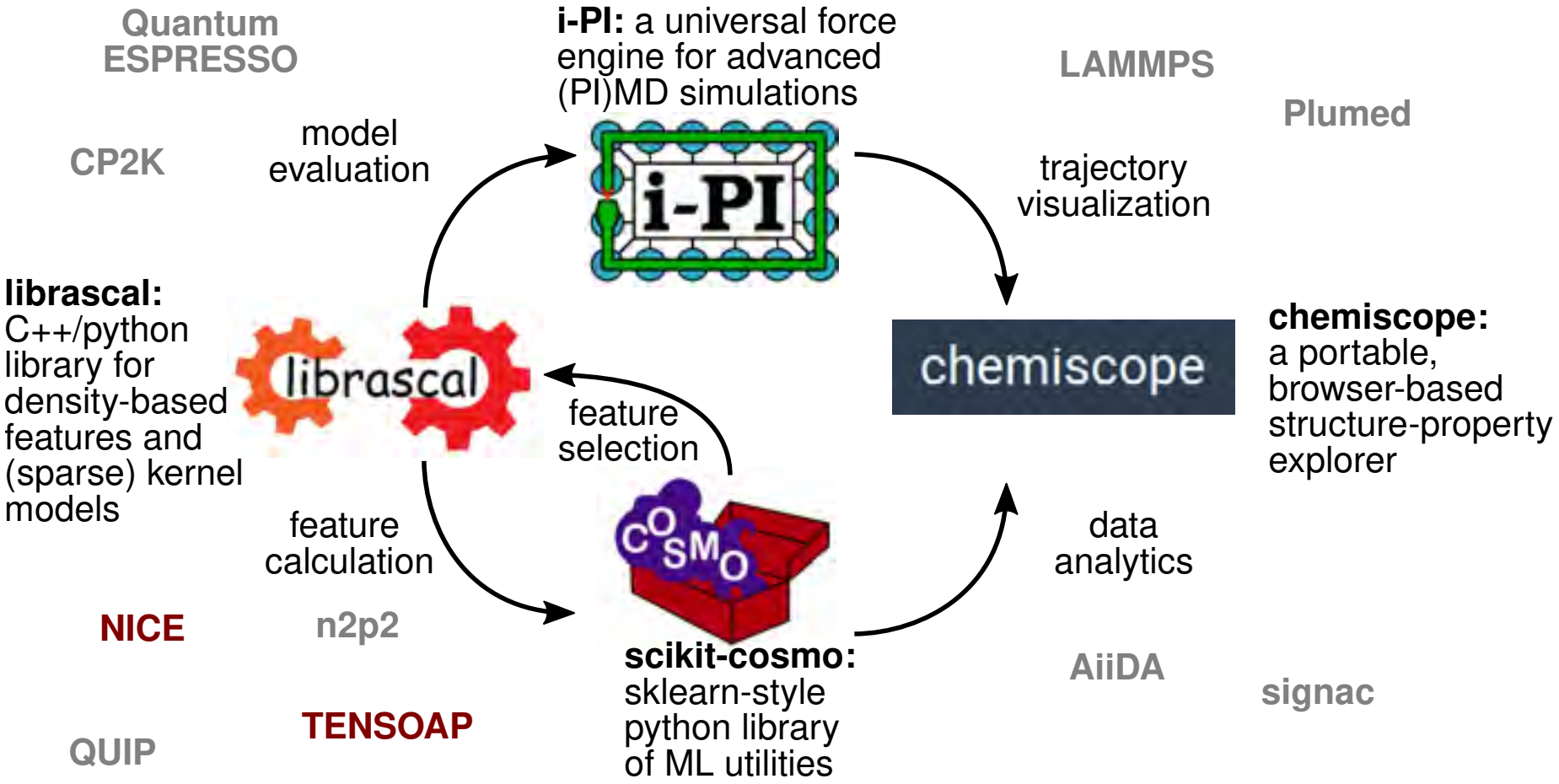


Imbalzano et al., JCP (2021)

Outlook

A software stack for atomistic machine learning

- Integrating ML and atomistic simulations: from representations to models to advanced MD
- Interoperability and data sharing with the rest of the ecosystem

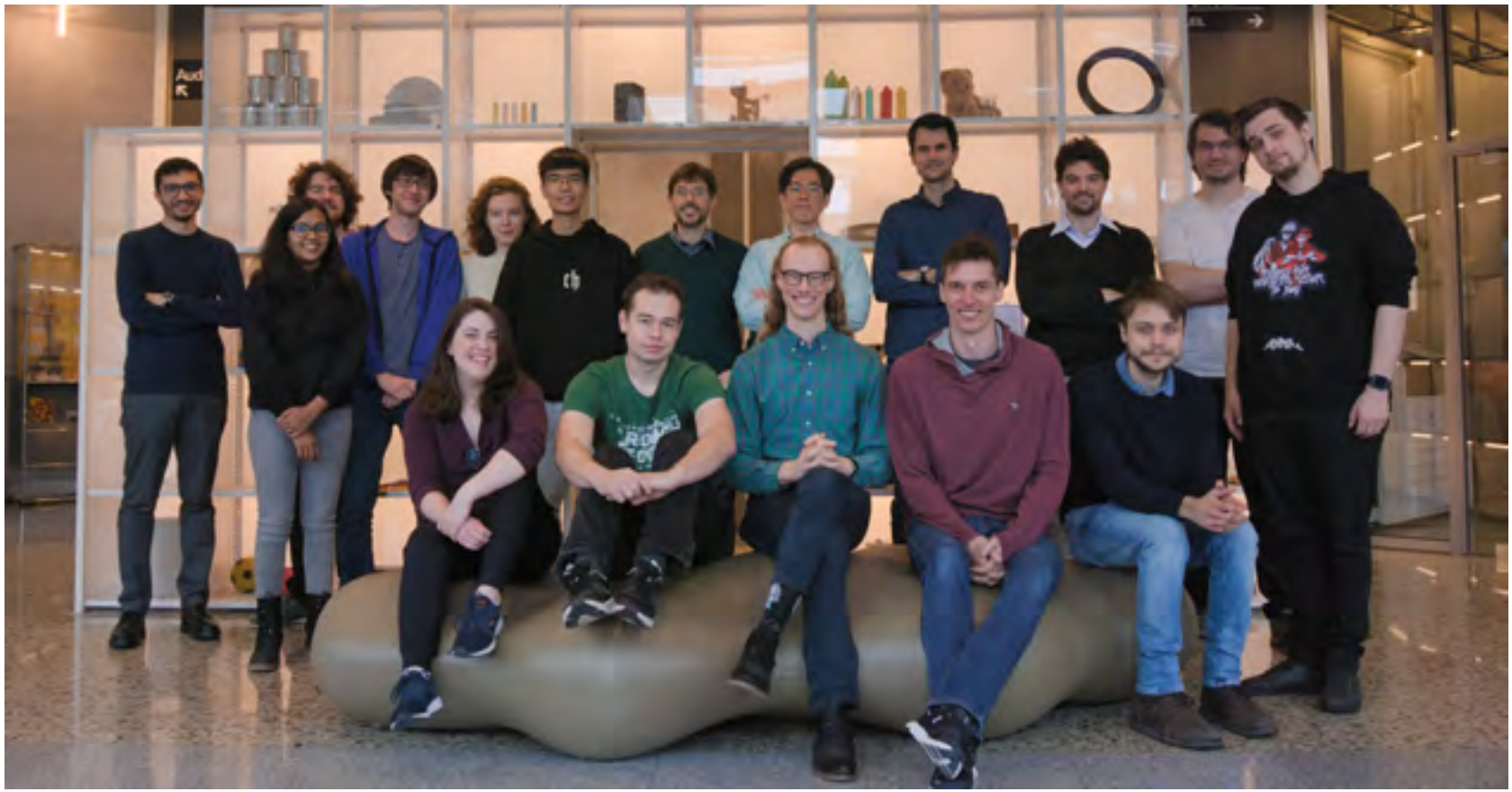


https://github.com/lab-cosmo/

Computational science & modeling @ EPFL

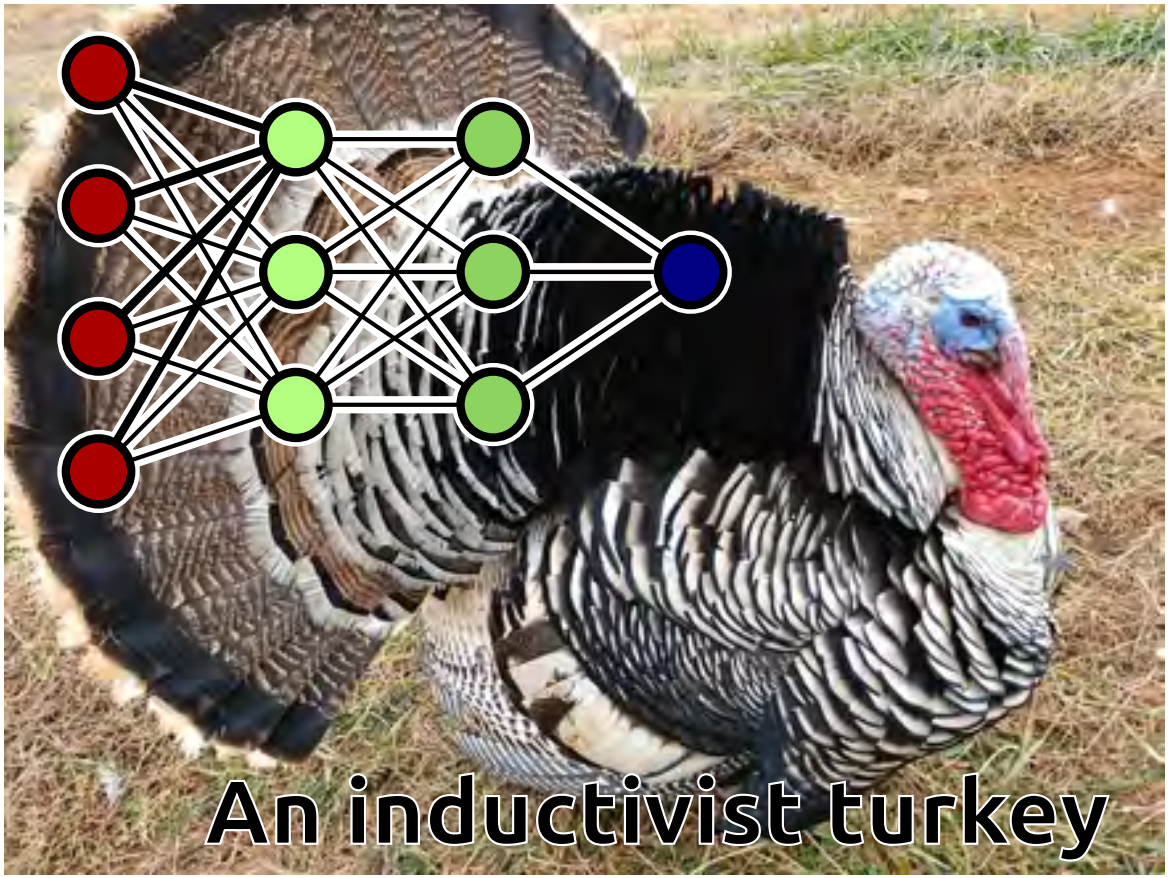
cosmo.epfl.ch

Follow @lab_COSMO



Building trust and understanding for ML

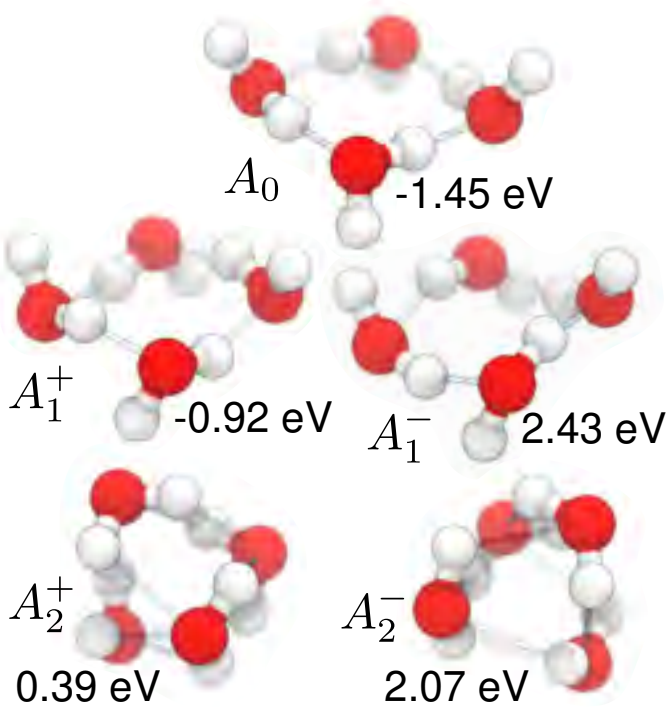
- ML is a purely inductive approach: dangerous in extrapolative contexts
- Physics-based constraints, and understanding of mathematical underpinnings, provide useful inductive biases
- Uncertainty quantification: easy, cheap, and universally applicable

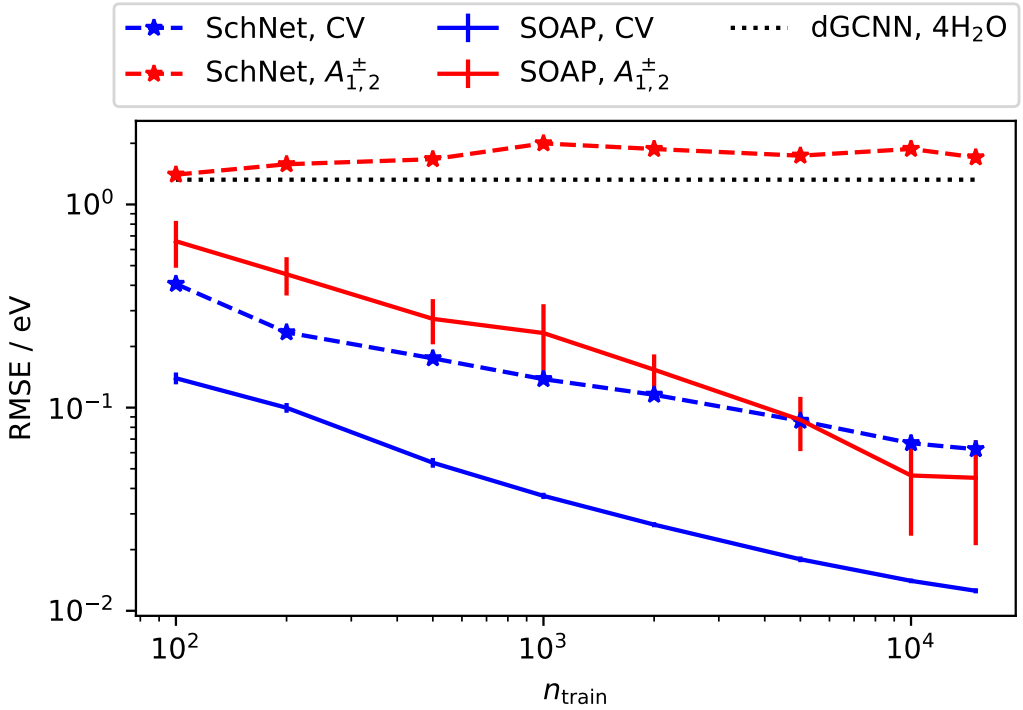


B. Russel, *The problems of philosophy*, Chapter VI

Building trust and understanding for ML

- ML is a purely inductive approach: dangerous in extrapolative contexts
- Physics-based constraints, and understanding of mathematical underpinnings, provide useful inductive biases
- Uncertainty quantification: easy, cheap, and universally applicable

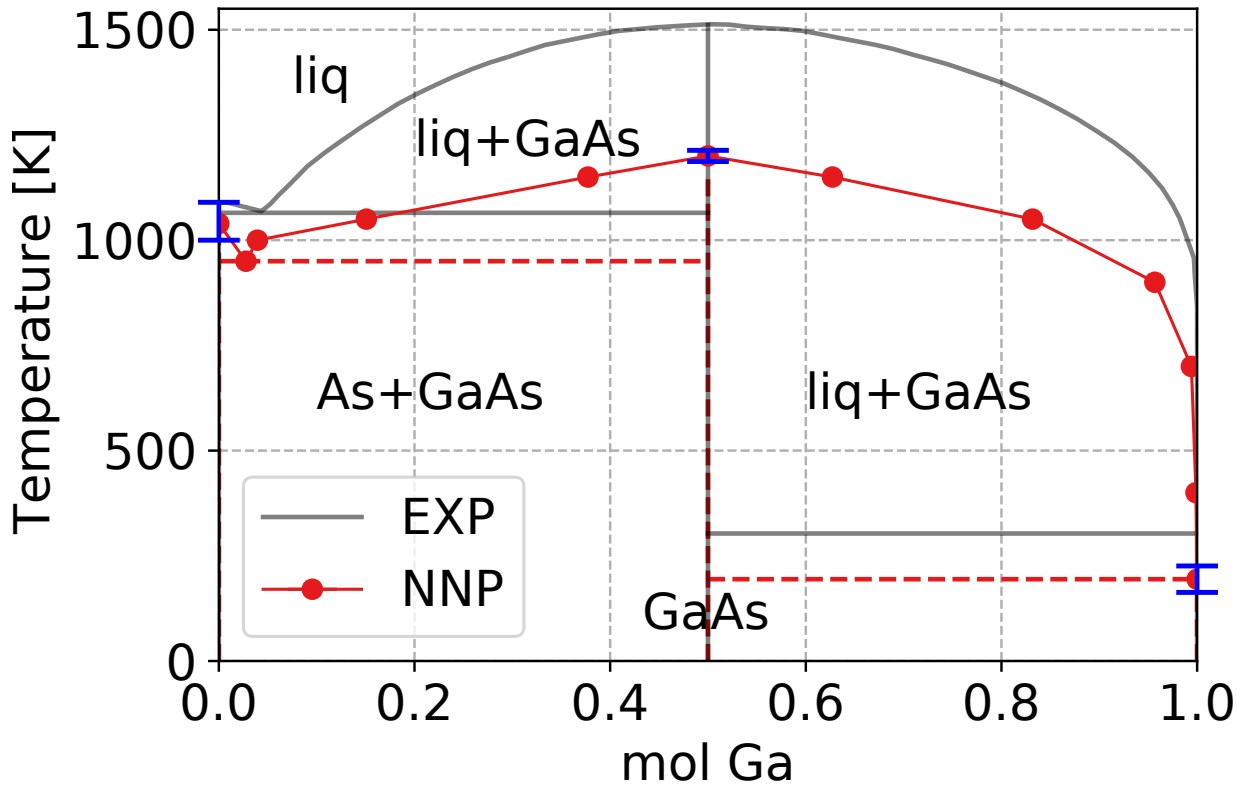




Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

Building trust and understanding for ML

- ML is a purely inductive approach: dangerous in extrapolative contexts
- Physics-based constraints, and understanding of mathematical underpinnings, provide useful inductive biases
- Uncertainty quantification: easy, cheap, and universally applicable



Imbalzano et al., JCP (2021); Imbalzano & **MC**, Phys. Rev. Materials (2021)

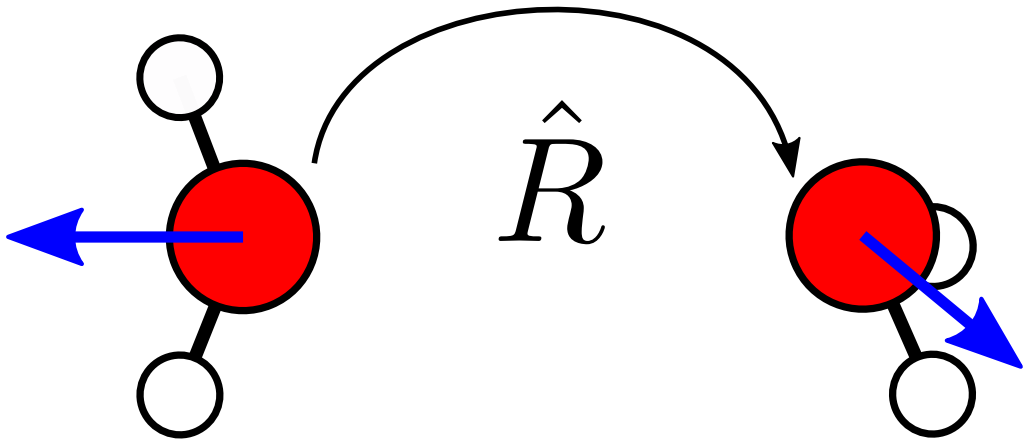
Backup slides

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$d_{\underset{\uparrow}{\alpha}}(A_i) = \sum_{q} \langle d|q\rangle \, \langle q|A; \overline{\rho_i^{\otimes\nu}; \underset{\uparrow}{\alpha}}\rangle$$

$$d_{\alpha}\left(\hat{R}A_i\right) = \sum_{q} \langle d|q\rangle \, \langle q|\hat{R}A; \overline{\rho_i^{\otimes\nu}; \alpha}\rangle$$



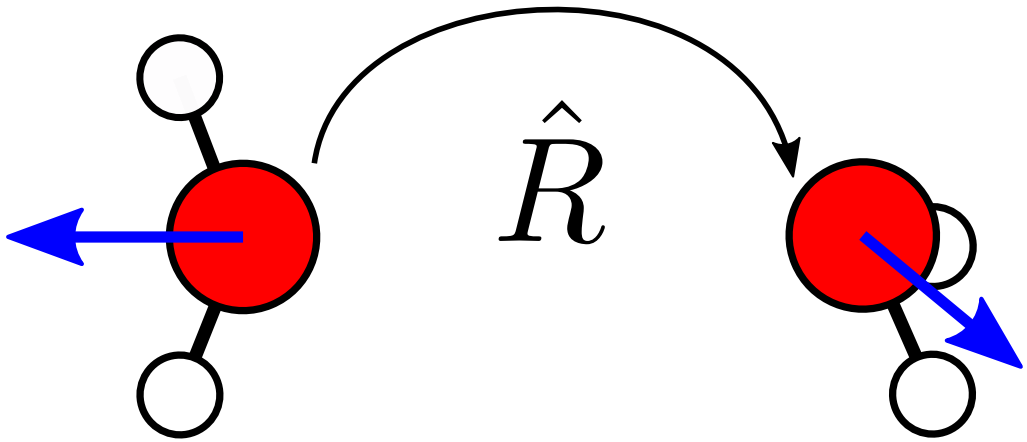
Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018); Veit et al., JCP (2020)

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$d_{\underset{\uparrow}{\alpha}}(A_i) = \sum_{q} \langle d|q\rangle \, \langle q|A; \overline{\rho_i^{\otimes\nu}; \underset{\uparrow}{\alpha}}\rangle$$

$$d_\alpha\left(\hat{R}A_i\right) = \sum_{q} \langle d|q\rangle \sum_{\alpha'} R_{\alpha\alpha'} \langle q|A; \overline{\rho_i^{\otimes\nu}; \alpha'}\rangle = \sum_{\alpha'} R_{\alpha\alpha'} d_{\alpha'}(A_i)$$



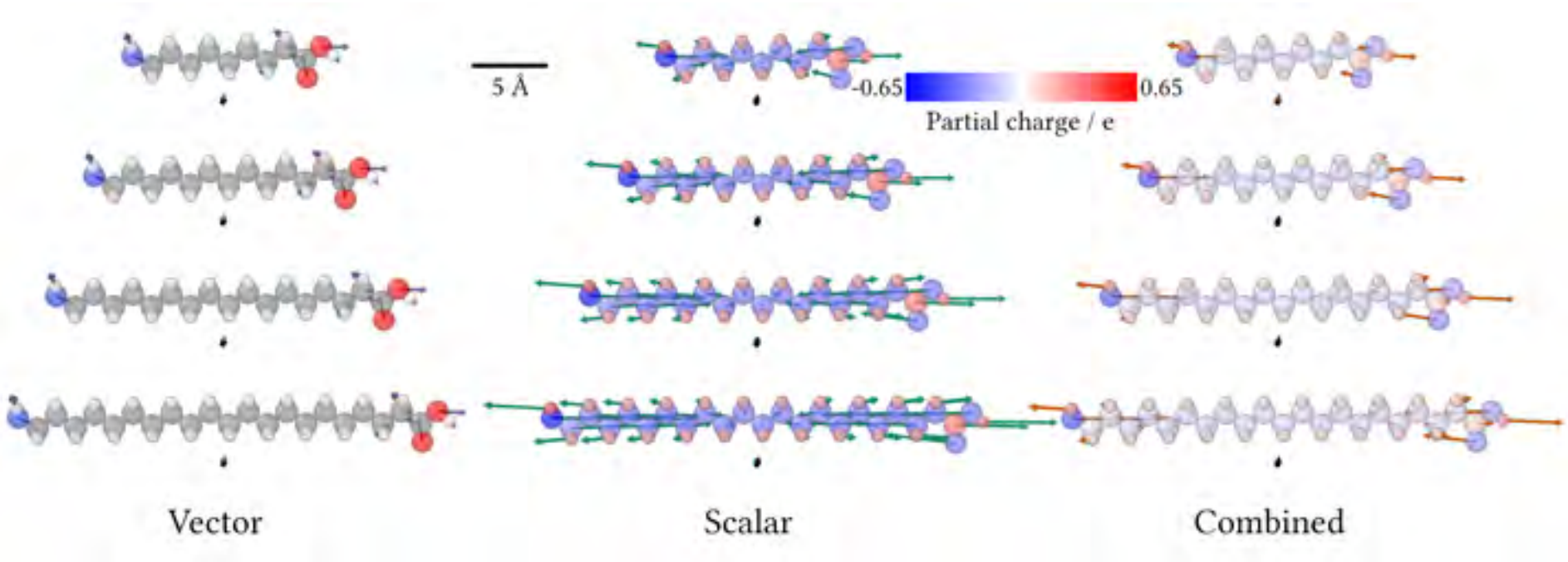
Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018); Veit et al., JCP (2020)

Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$d_{\underset{\uparrow}{\alpha}}(A_i) = \sum_{q} \langle d|q\rangle \, \langle q|A; \overline{\rho_i^{\otimes\nu}; \underset{\uparrow}{\alpha}}\rangle$$

$$d_\alpha\left(\hat{R}A_i\right) = \sum_{q} \langle d|q\rangle \sum_{\alpha'} R_{\alpha\alpha'} \langle q|A; \overline{\rho_i^{\otimes\nu}; \alpha'}\rangle = \sum_{\alpha'} R_{\alpha\alpha'} d_{\alpha'}(A_i)$$



Veit, Wilkins, Yang, DiStasio, **MC**, JCP (2020)

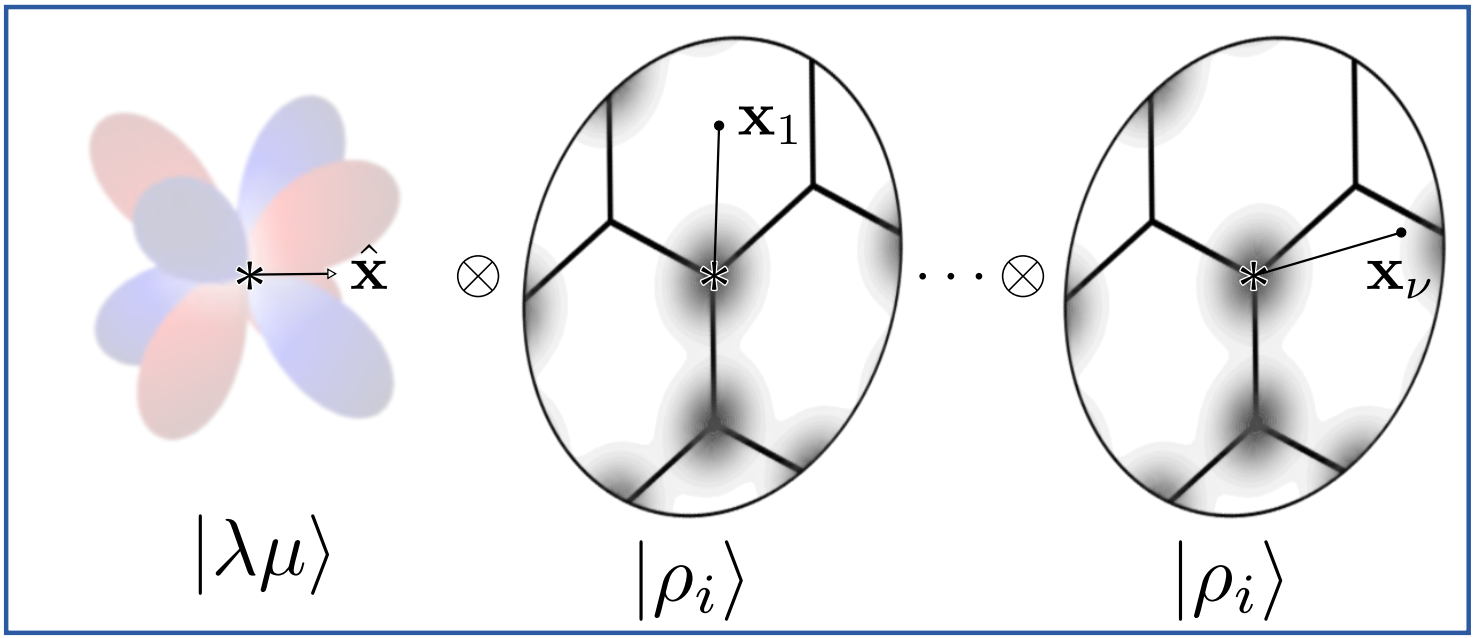
Machine-learning for tensors

- Want to learn vectors or general tensors?
  Need features that are *equivariant* to rotations

$$d_{\underset{\uparrow}{\alpha}}(A_i) = \sum_{q} \langle d|q\rangle \, \langle q|A; \overline{\rho_i^{\otimes\nu}; \underset{\uparrow}{\alpha}}\rangle$$

$$y_\mu^\lambda\left(\hat{R}A_i\right) = \sum_{q} \langle d|q\rangle \sum_{\mu'} D^\lambda_{\mu\mu'}\left(\hat{R}\right) \langle q|A; \overline{\rho_i^{\otimes\nu}; \lambda\mu}\rangle$$

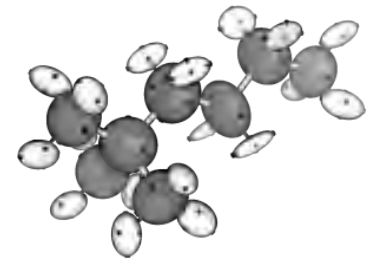
$\int \mathrm{d}\hat{R}$



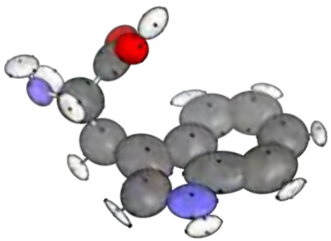
Grisafi, Wilkins, Csányi, & **MC**, PRL (2018); Willatt, Musil, & **MC**, JCP (2019)

Molecular polarizabilities at the CCSD level

- Symmetry-adapted tensorial model of polarizabilities
- Training on high-end CCSD calculations of small molecules
- Extrapolate to larger molecules (tested up to aciclovir $C_8H_{11}N_5O_3$)
- Better-than-DFT accuracy - try it on **alphaml.org**

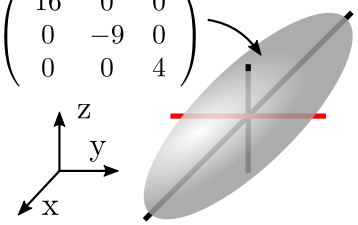


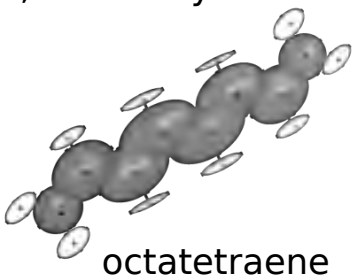
2,2-dimethylhexane



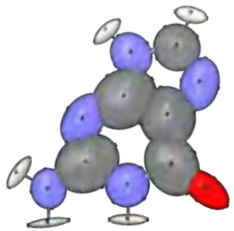
tryptophan

$$\boldsymbol{\alpha} \approx \begin{pmatrix} 16 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

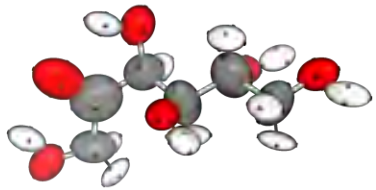




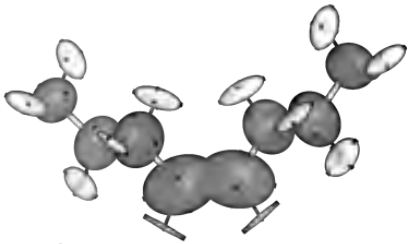
octatetraene



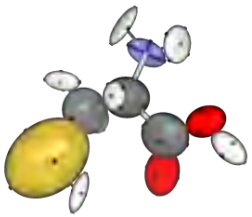
guanine



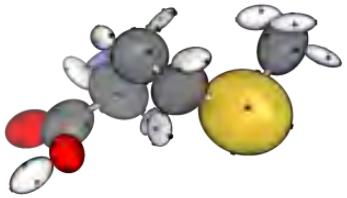
fructose



cis-4-octene



cysteine



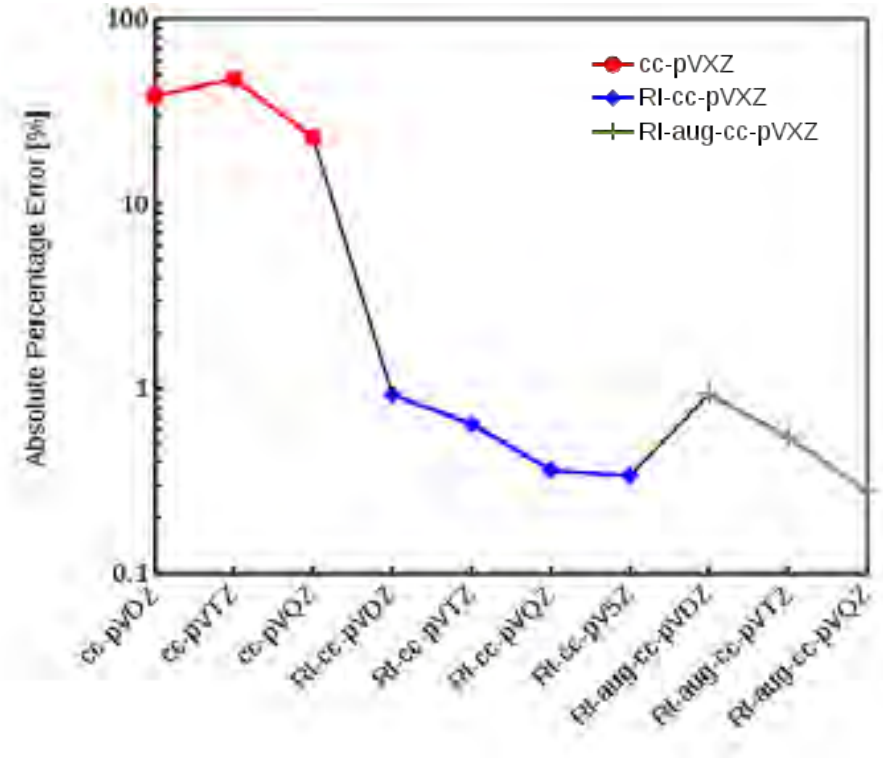
methionine

Method	RMSE
CCSD/ML	0.304
CCSD/DFT	0.573

Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, PNAS (2019);

Transferable model of the electron density

- Expand the charge density on atom-centered basis $\phi_k \equiv R_n Y_l^m$
- Learning of coefficients with symmetry-adapted kernels
- Highly transferable: train on small molecules, predict on polypeptides
- Condensed phase implementation. Scaling up complexity!



cc-pVQZ

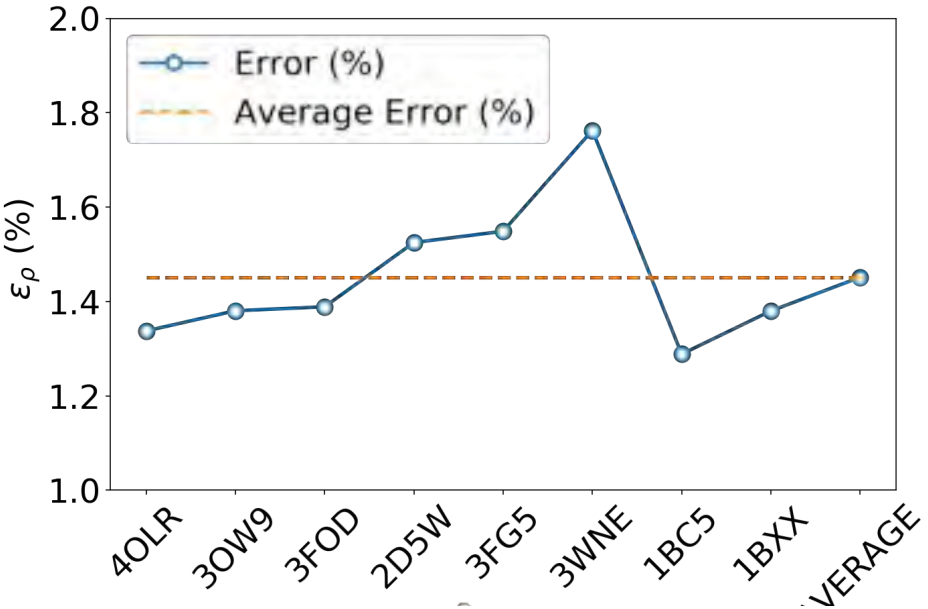
RI-cc-pVQZ

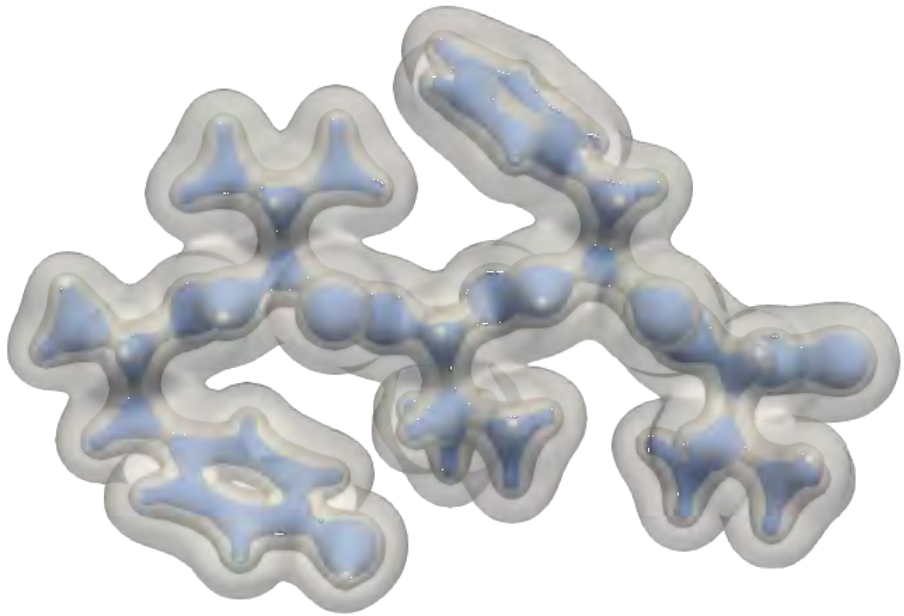
Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, ACS Central Science (2019);

Meyer, Grisafi, Fabrizio, **MC**, Corminboeuf, Chem. Sci., (2019)

Transferable model of the electron density

- Expand the charge density on atom-centered basis $\phi_k \equiv R_n Y_l^m$
- Learning of coefficients with symmetry-adapted kernels
- Highly transferable: train on small molecules, predict on polypeptides
- Condensed phase implementation. Scaling up complexity!



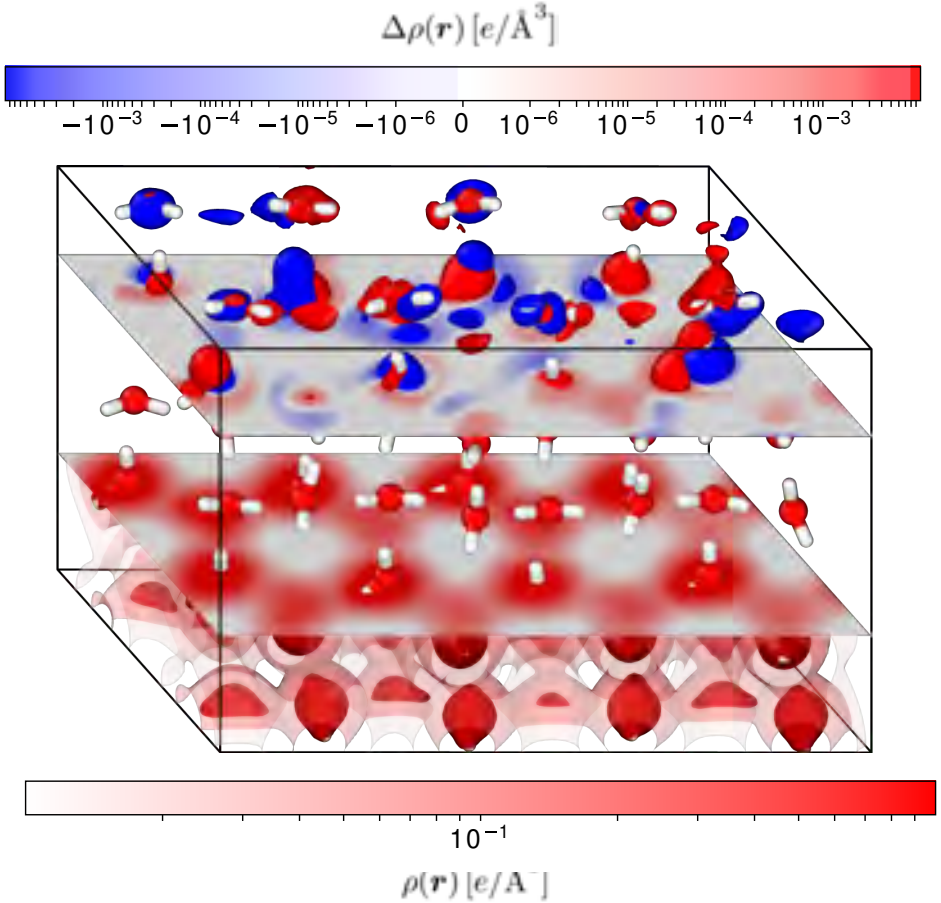


Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, ACS Central Science (2019);

Meyer, Grisafi, Fabrizio, **MC**, Corminboeuf, Chem. Sci., (2019)

Transferable model of the electron density

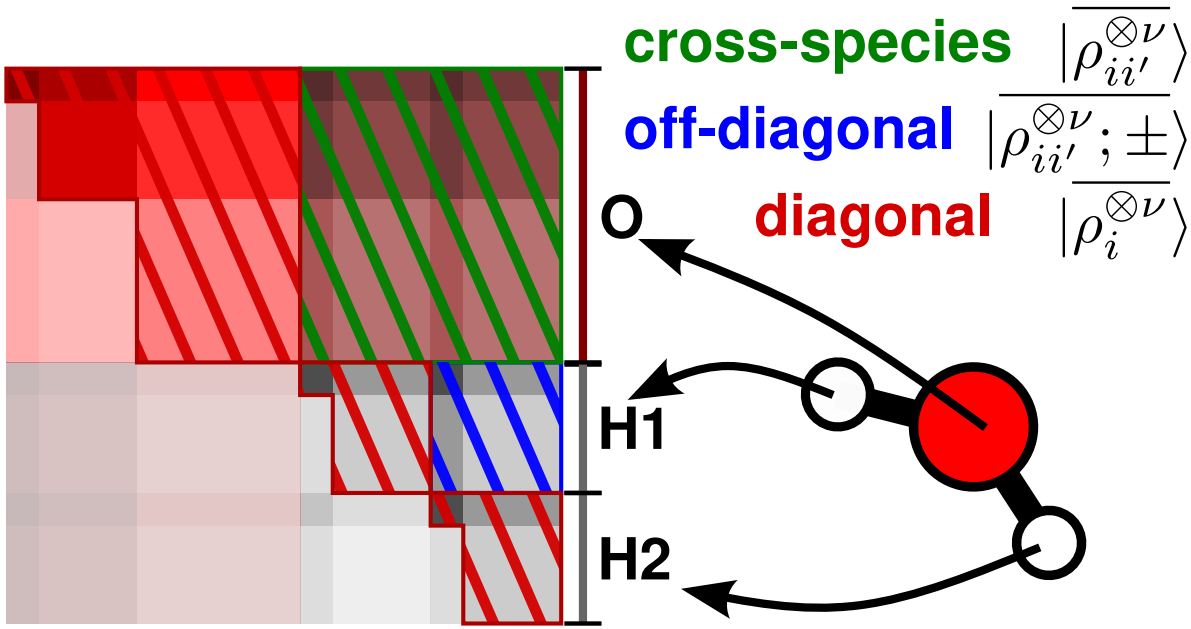
- Expand the charge density on atom-centered basis $\phi_k \equiv R_n Y_l^m$
- Learning of coefficients with symmetry-adapted kernels
- Highly transferable: train on small molecules, predict on polypeptides
- Condensed phase implementation. Scaling up complexity!



Lewis, Grisafi, **MC**, Rossi, JCTC (2021)

Learning molecular Hamiltonians

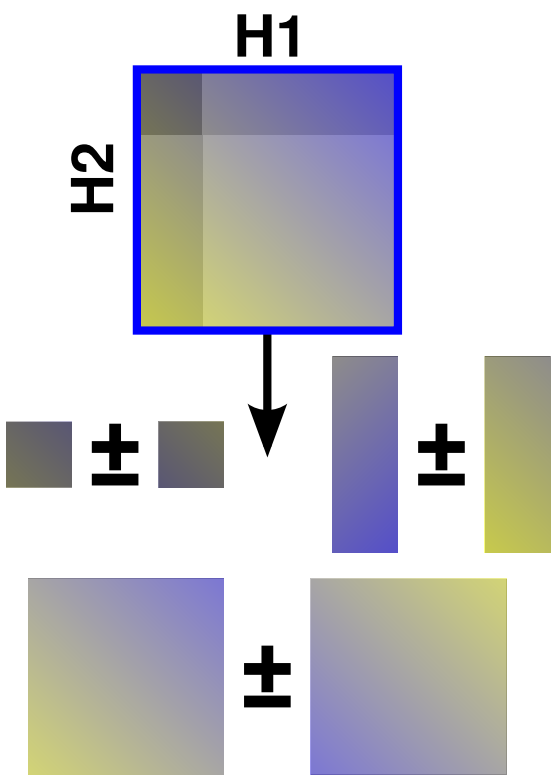
- Molecular Hamiltonian in an atomic orbital basis
- Decompositon into irreducible symmetric blocks
- Learn with a fully equivariant model
- Consistent with molecular orbital theory by design

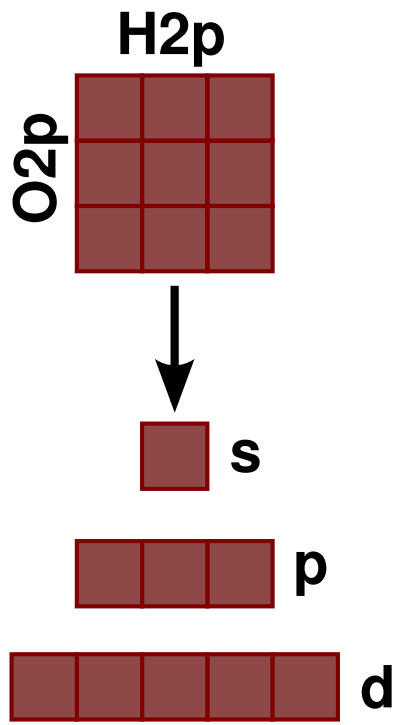


Nigam, Willatt, **MC**, JCP (2022)

Learning molecular Hamiltonians

- Molecular Hamiltonian in an atomic orbital basis
- Decompositon into irreducible symmetric blocks
- Learn with a fully equivariant model
- Consistent with molecular orbital theory by design

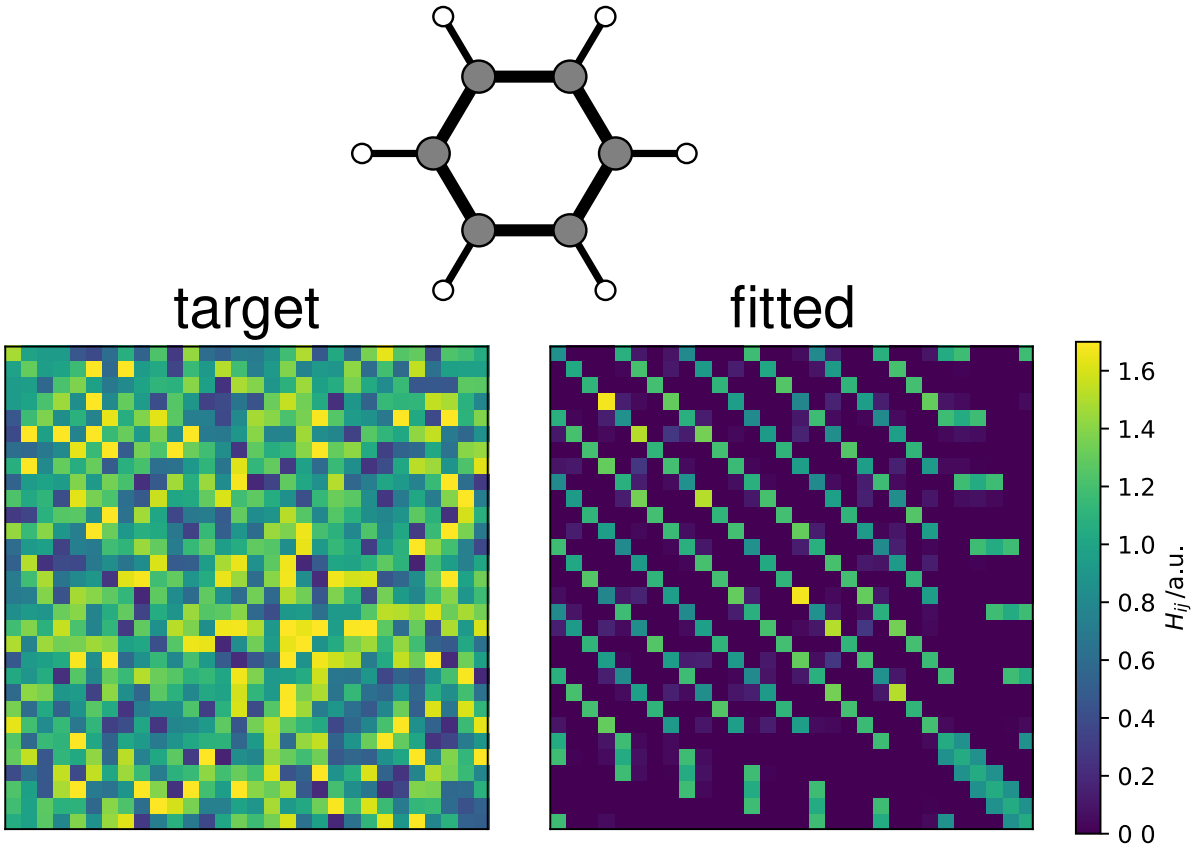




Nigam, Willatt, **MC**, JCP (2022)

Learning molecular Hamiltonians

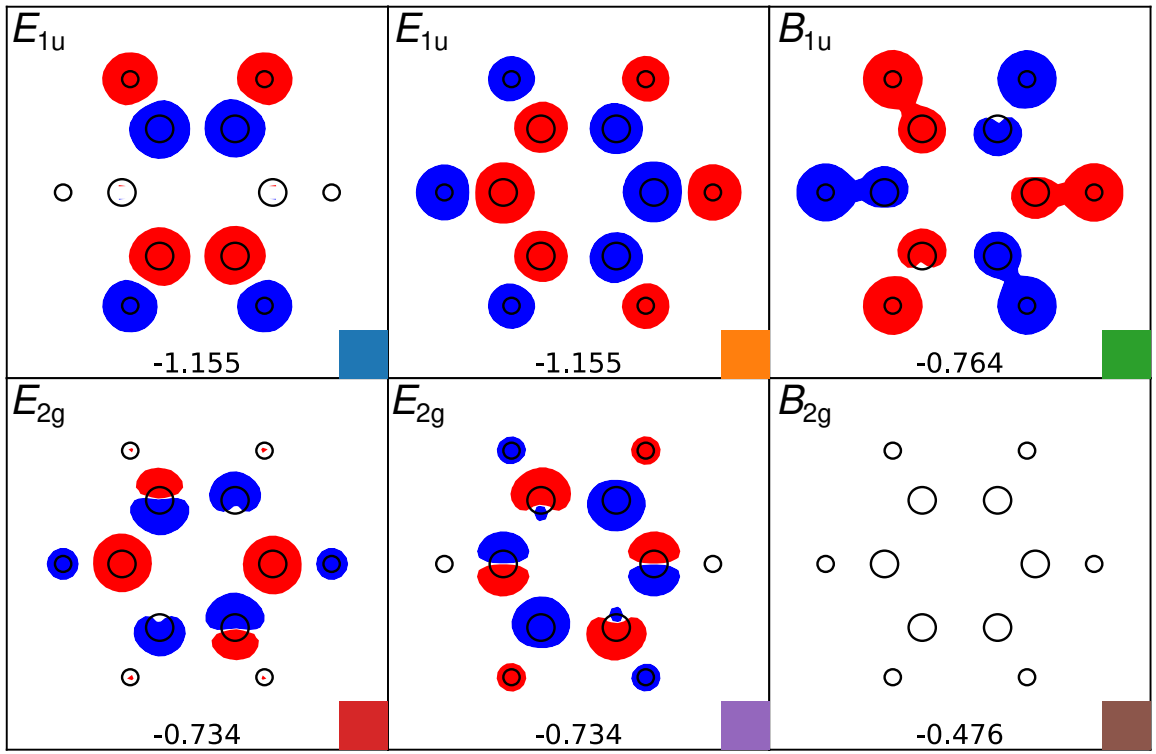
- Molecular Hamiltonian in an atomic orbital basis
- Decompositon into irreducible symmetric blocks
- Learn with a fully equivariant model
- Consistent with molecular orbital theory by design



Nigam, Willatt, **MC**, JCP (2022)

Learning molecular Hamiltonians

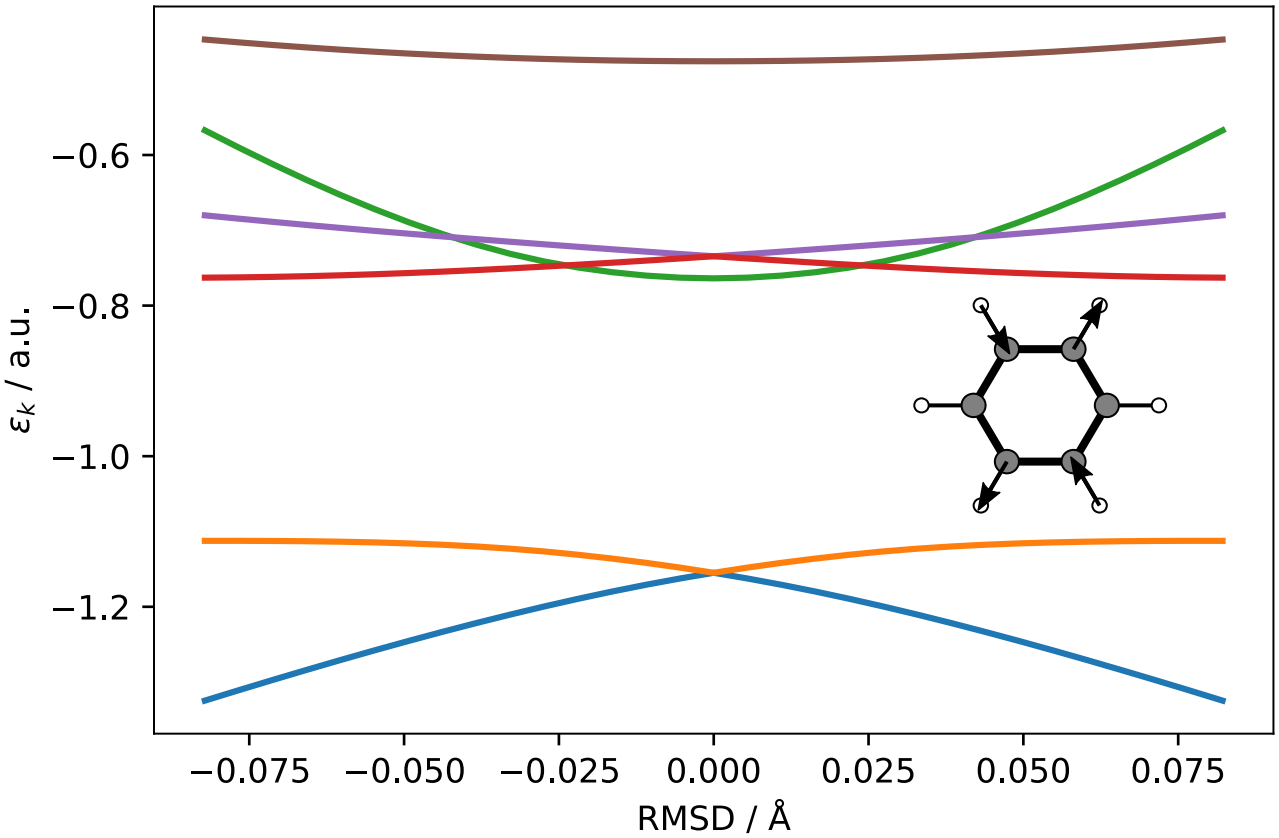
- Molecular Hamiltonian in an atomic orbital basis
- Decompositon into irreducible symmetric blocks
- Learn with a fully equivariant model
- Consistent with molecular orbital theory by design



Nigam, Willatt, **MC**, JCP (2022)

Learning molecular Hamiltonians

- Molecular Hamiltonian in an atomic orbital basis
- Decompositon into irreducible symmetric blocks
- Learn with a fully equivariant model
- Consistent with molecular orbital theory by design

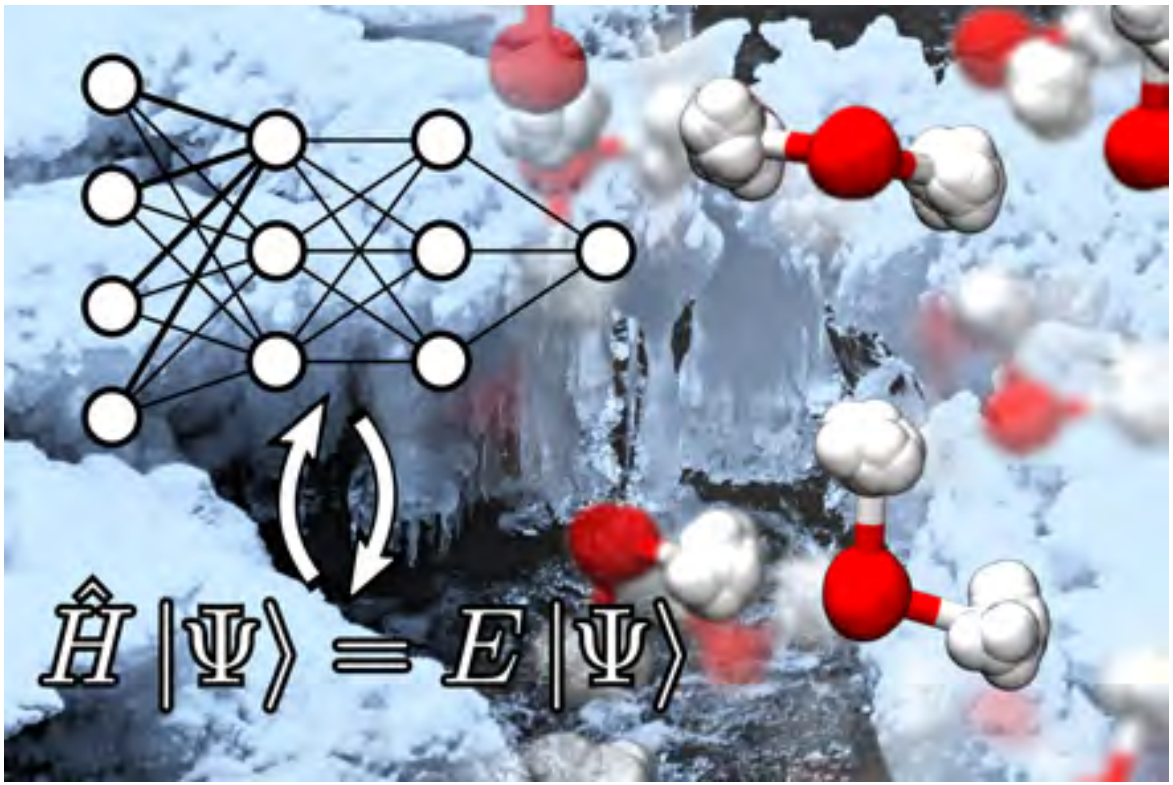


Nigam, Willatt, **MC**, JCP (2022)

ML, ML everywhere
(and all the errors did shrink)

Ab initio thermodynamics of water

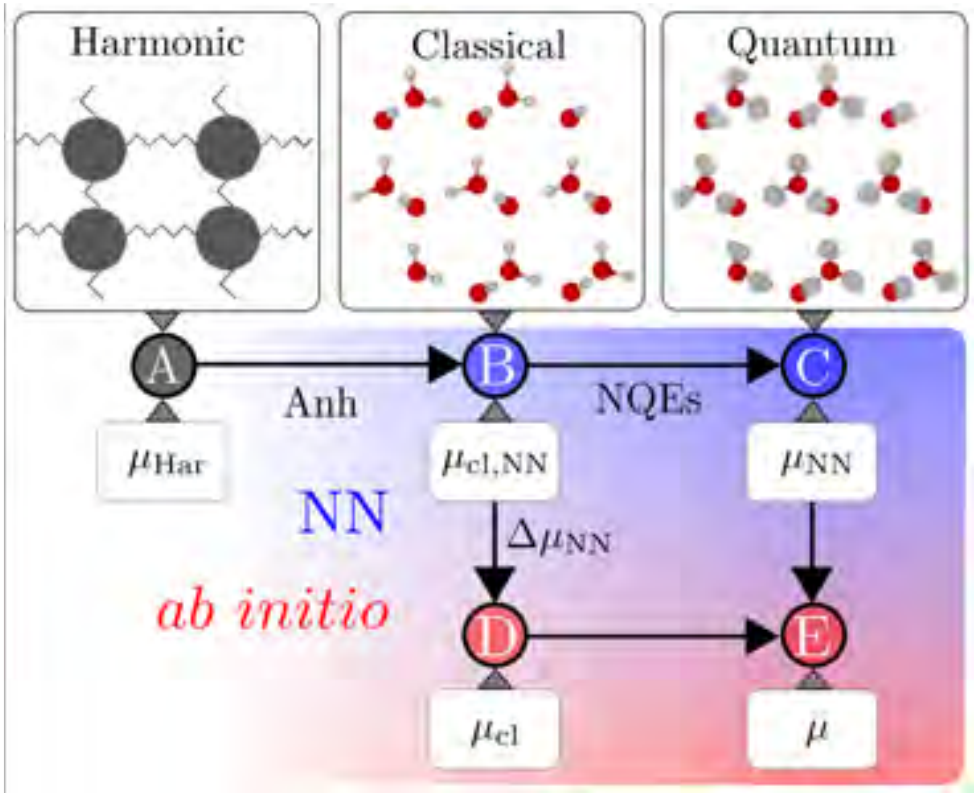
- Thermodynamics of water liquid/ice-Ih/ice-Ic.
- Hybrid DFT+D3 level, with quantum nuclei
- ML, for sampling, promote to full-DFT with free-energy perturbation
- Melting point of water within ~5K. $\Delta G_{Ih/Ic}$ within 0.1 meV/molecule



Cheng, Engel, Behler, Dellago, **MC**, PNAS (2019)

Ab initio thermodynamics of water

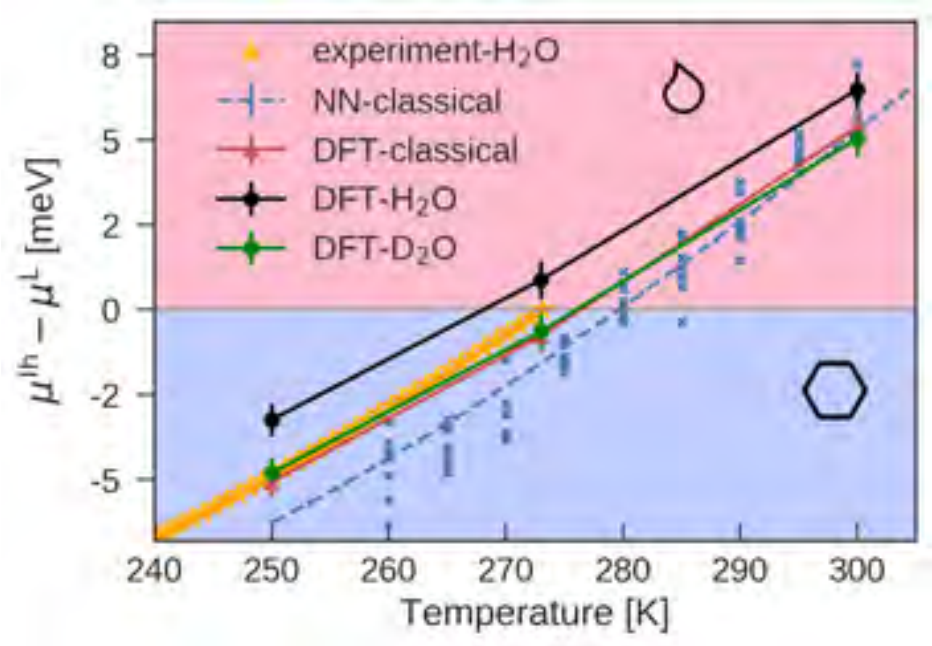
- Thermodynamics of water liquid/ice-Ih/ice-Ic.
- Hybrid DFT+D3 level, with quantum nuclei
- ML, for sampling, promote to full-DFT with free-energy perturbation
- Melting point of water within ~5K. $\Delta G_{Ih/Ic}$ within 0.1 meV/molecule

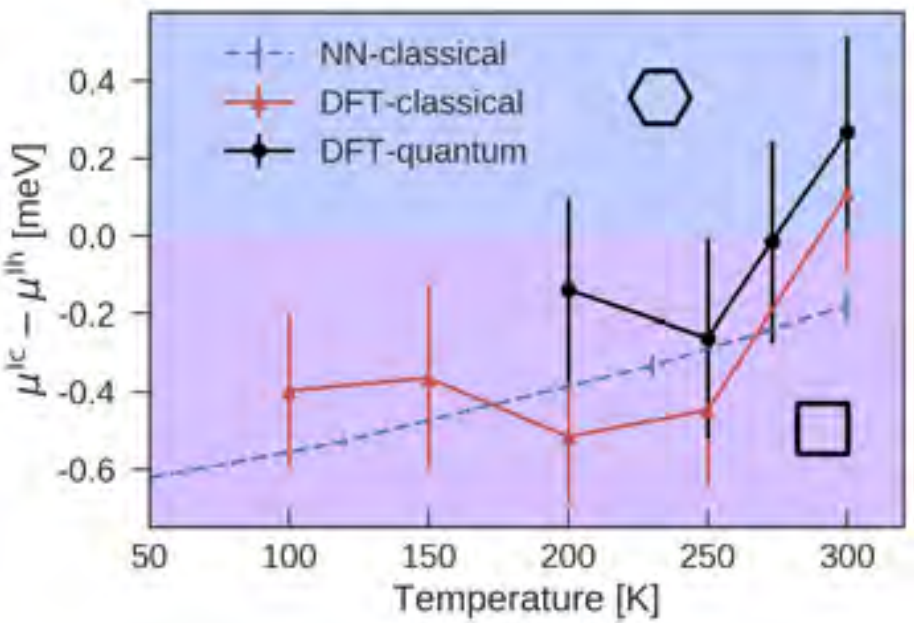


Cheng, Engel, Behler, Dellago, **MC**, PNAS (2019)

Ab initio thermodynamics of water

- Thermodynamics of water liquid/ice-Ih/ice-Ic.
- Hybrid DFT+D3 level, with quantum nuclei
- ML, for sampling, promote to full-DFT with free-energy perturbation
- Melting point of water within ~5K. $\Delta G_{Ih/Ic}$ within 0.1 meV/molecule

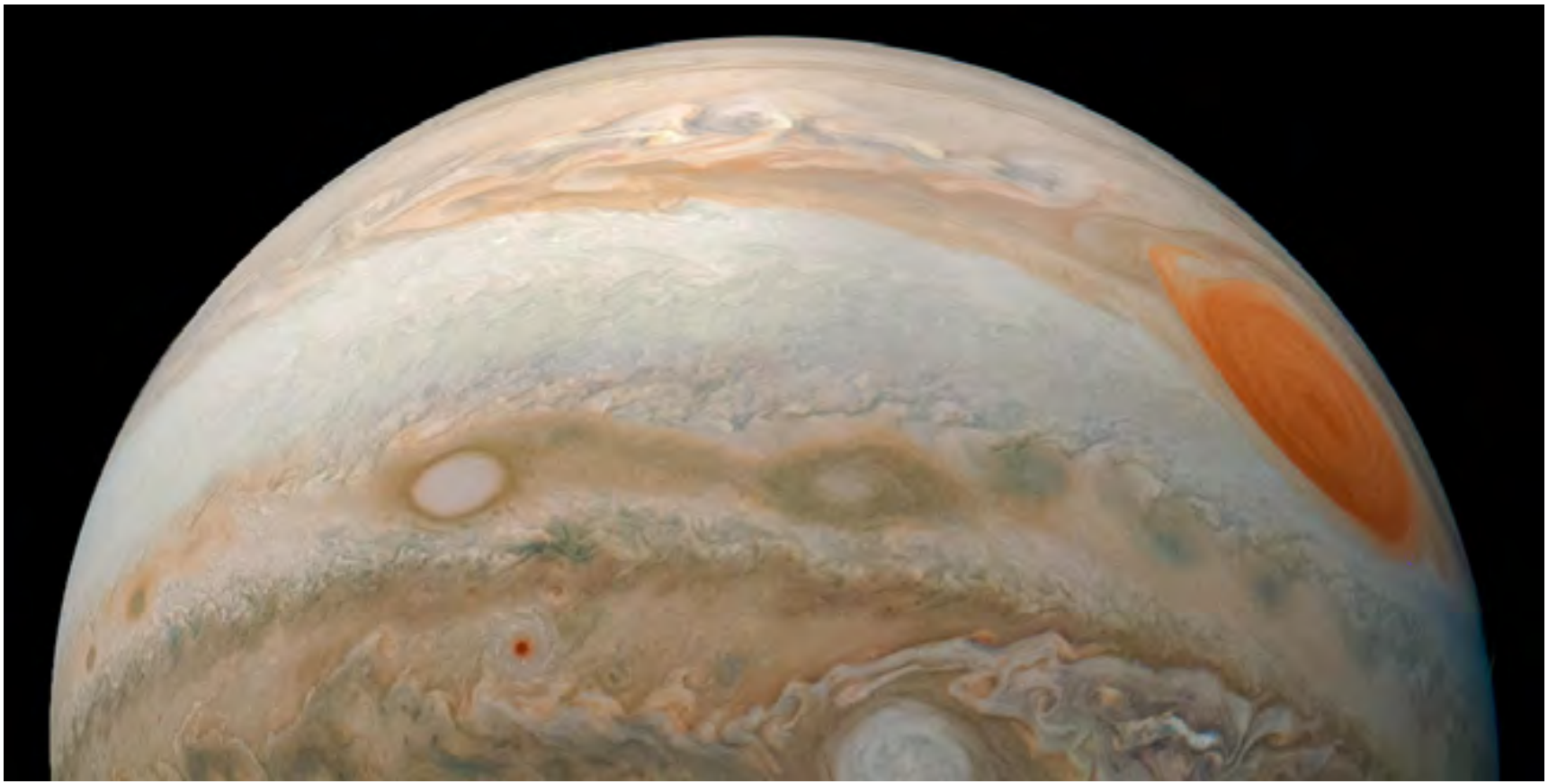




Cheng, Engel, Behler, Dellago, **MC**, PNAS (2019)

Liquid-liquid phase transition in dense H

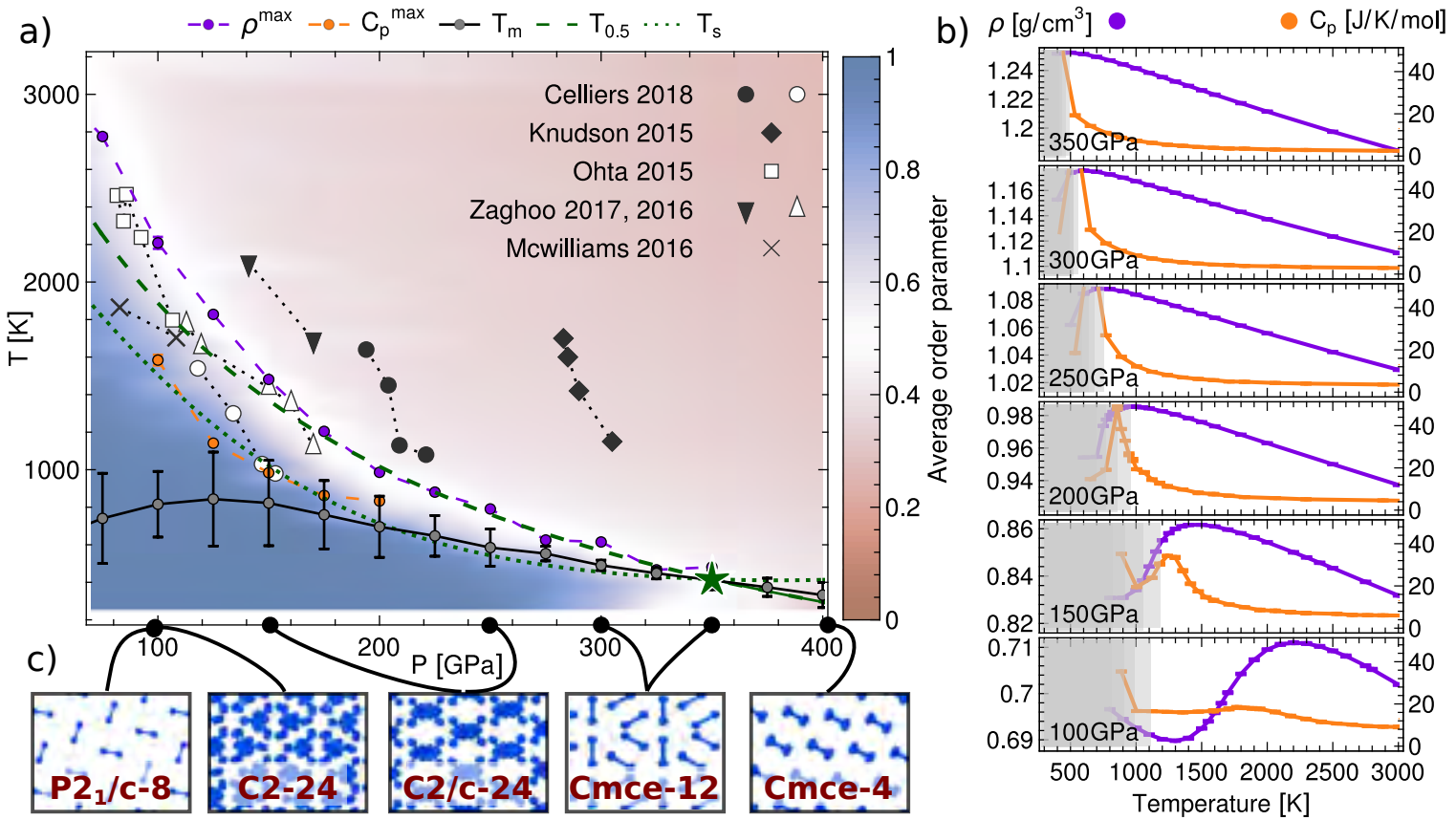
- Atomic/molecular fluid for H at giant-planets conditions
- Difficult to characterize the transition experimentally
- First-order transition predicted by ab initio MD: finite-size effect?
- ML potential and polyamorphic solution model finds critical point to lie on the solid-liquid phase line



Cheng, Mazzola, Pickard, **MC**, Nature (2020)

Liquid-liquid phase transition in dense H

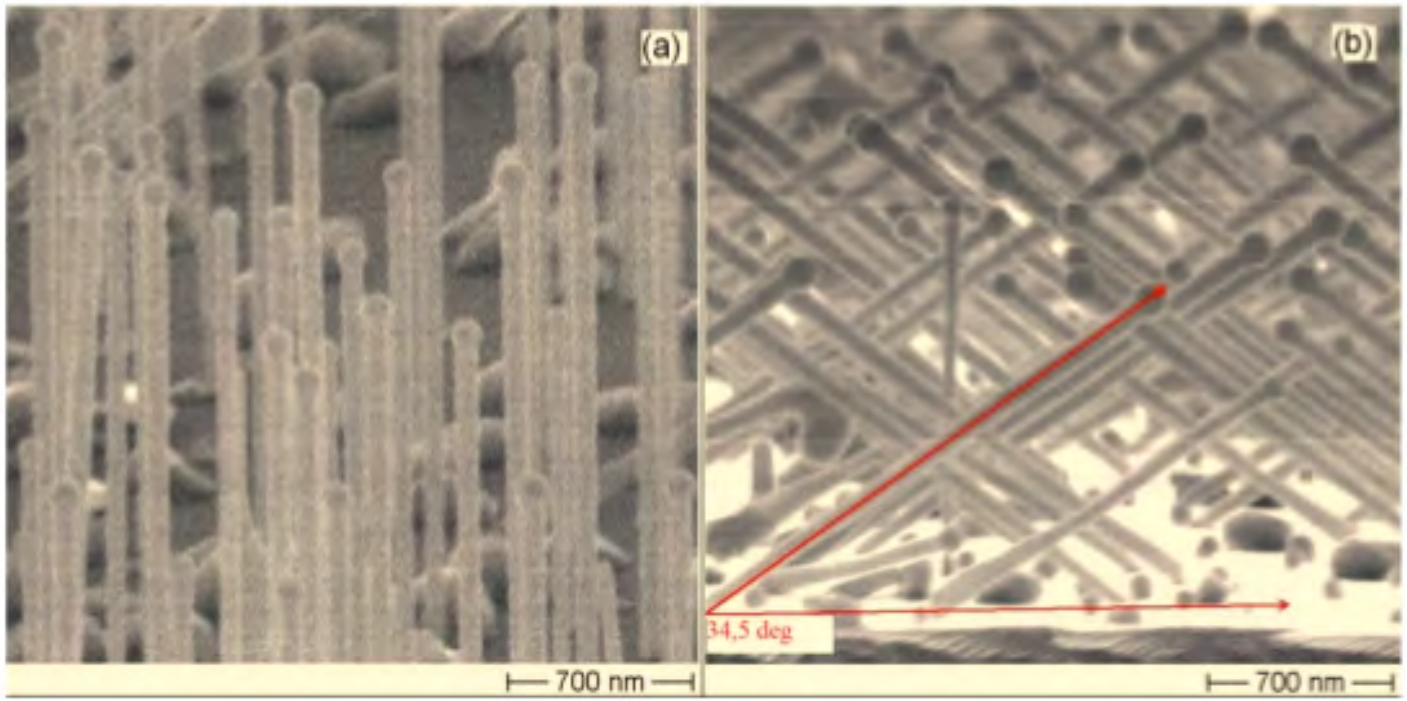
- Atomic/molecular fluid for H at giant-planets conditions
- Difficult to characterize the transition experimentally
- First-order transition predicted by ab initio MD: finite-size effect?
- ML potential and polyamorphic solution model finds critical point to lie on the solid-liquid phase line



Cheng, Mazzola, Pickard, **MC**, Nature (2020)

Synthesis of GaAs nanowires

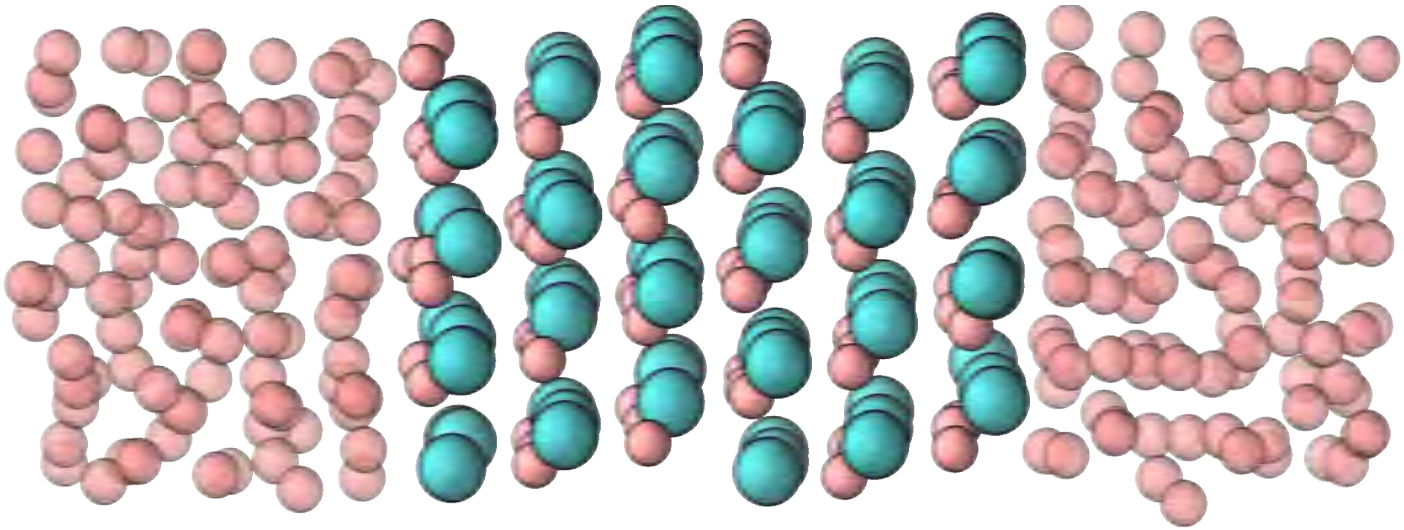
- Synthesis of GaAs nanostructures: solid-liquid-gas process, Ga_l/GaAs interface based on droplets of molten Ga
- Problem: controlling the zinc-blende/wurtzite polymorphism
- Role of surface polarity and liquid structure on defect rate
- MD simulation with DFT/NN to study liquid ordering at the interfaces
- More to come: transferable model of the Ga_xAs_{1-x} binary



Mahdi Zamani et al., Adv. Materials (2020)

Synthesis of GaAs nanowires

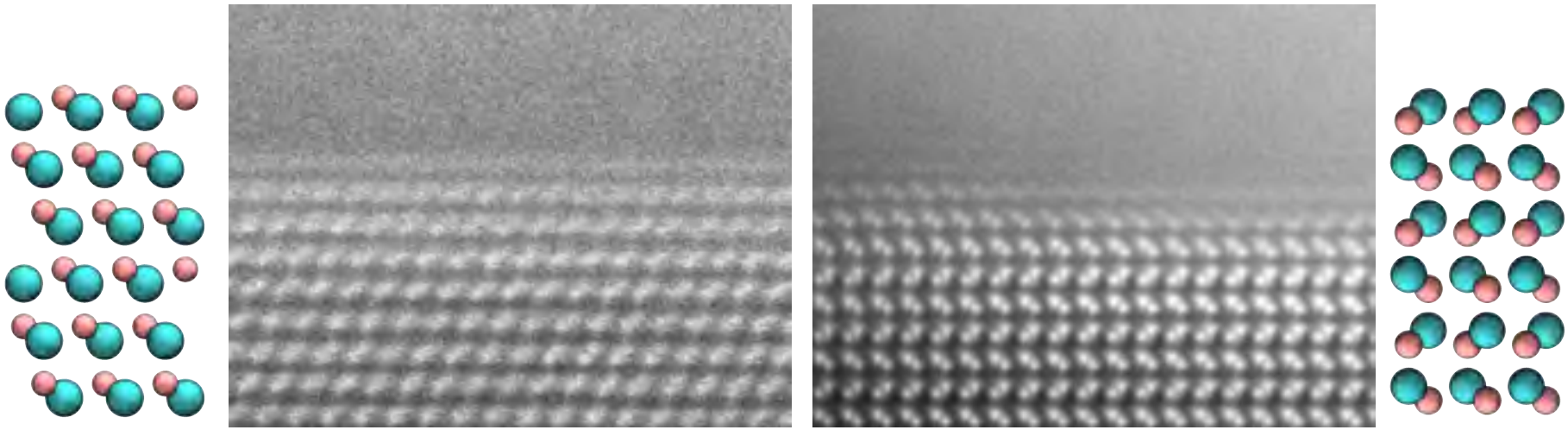
- Synthesis of GaAs nanostructures: solid-liquid-gas process, Ga_l/GaAs interface based on droplets of molten Ga
- Problem: controlling the zinc-blende/wurtzite polymorphism
- Role of surface polarity and liquid structure on defect rate
- MD simulation with DFT/NN to study liquid ordering at the interfaces
- More to come: transferable model of the Ga_xAs_{1-x} binary



Mahdi Zamani et al., Adv. Materials (2020)

Synthesis of GaAs nanowires

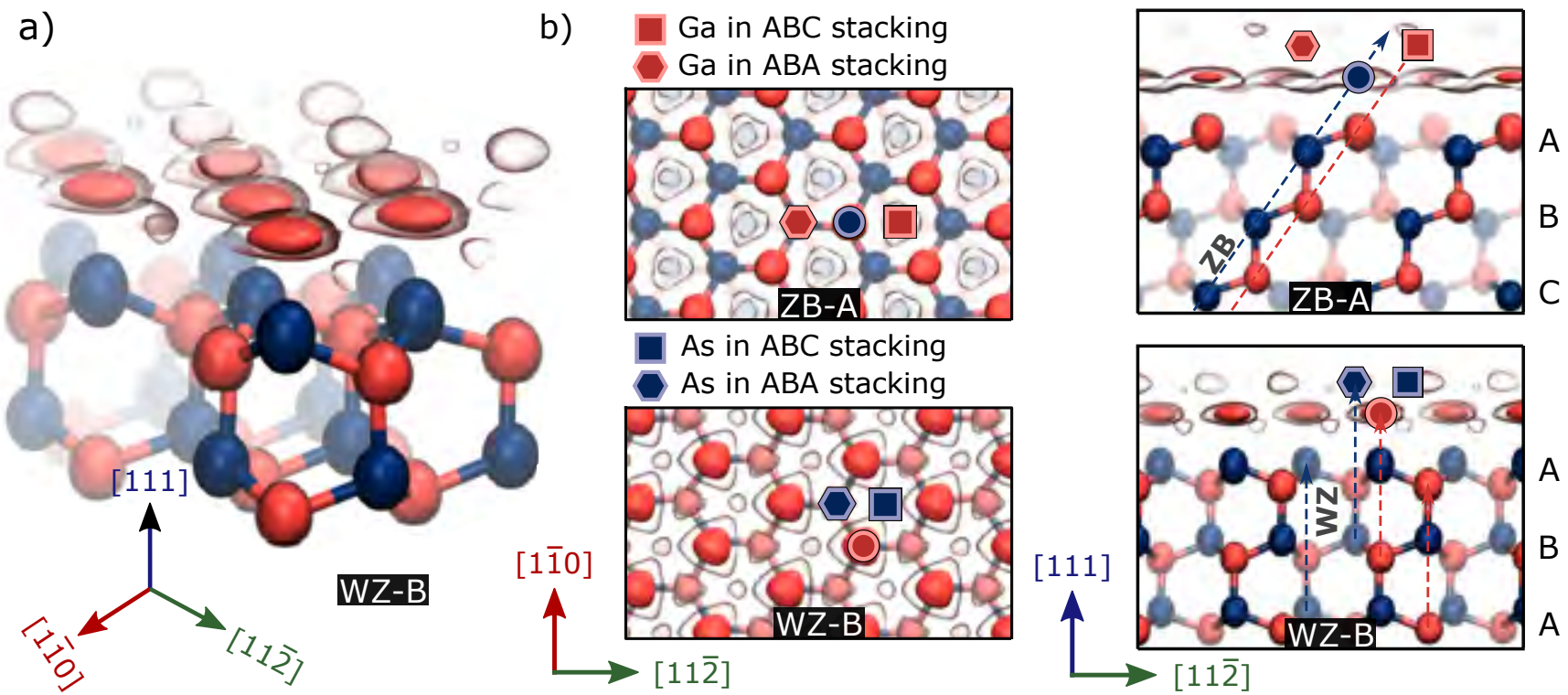
- Synthesis of GaAs nanostructures: solid-liquid-gas process, Ga_l/GaAs interface based on droplets of molten Ga
- Problem: controlling the zinc-blende/wurtzite polymorphism
- Role of surface polarity and liquid structure on defect rate
- MD simulation with DFT/NN to study liquid ordering at the interfaces
- More to come: transferable model of the Ga_xAs_{1-x} binary



Mahdi Zamani et al., Adv. Materials (2020)

Synthesis of GaAs nanowires

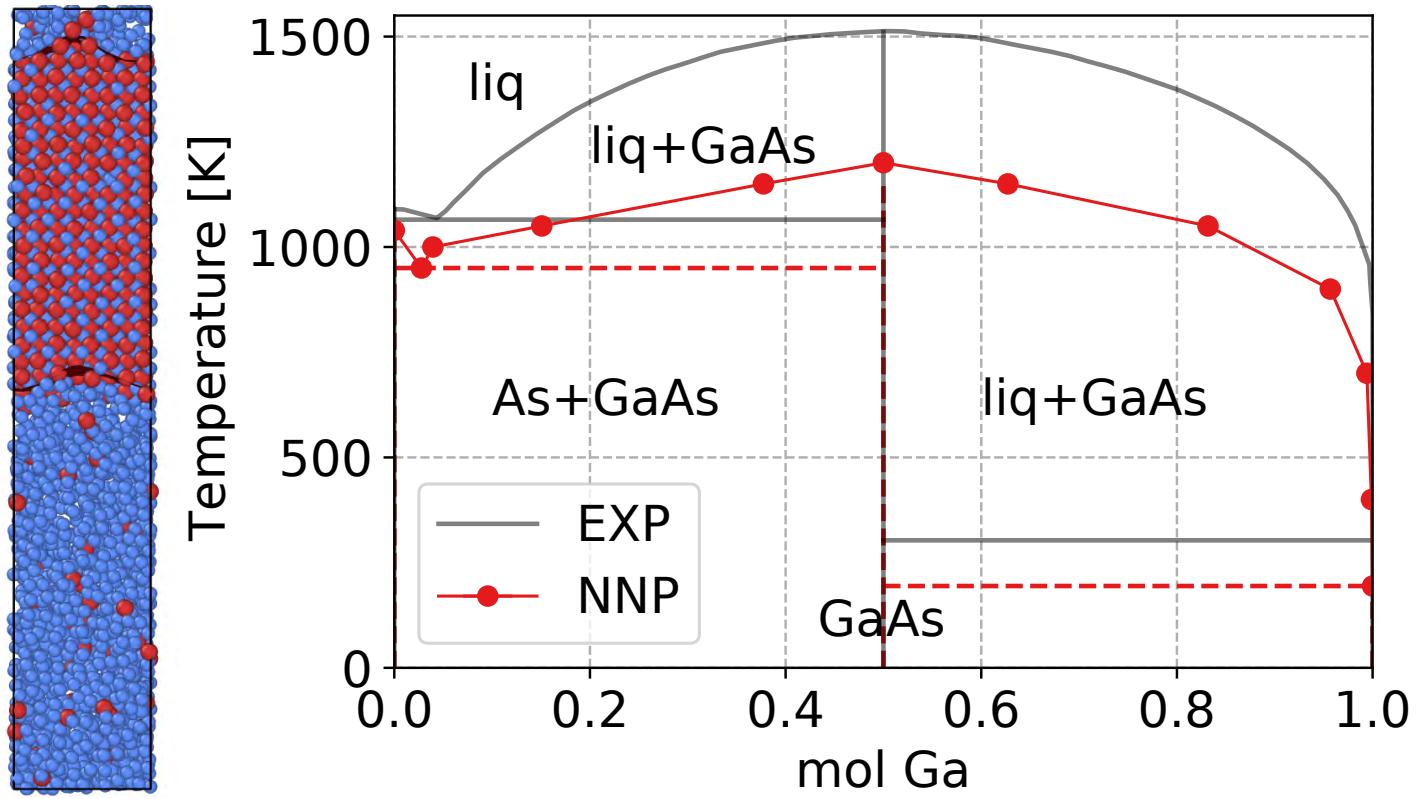
- Synthesis of GaAs nanostructures: solid-liquid-gas process, Ga_l/GaAs interface based on droplets of molten Ga
- Problem: controlling the zinc-blende/wurtzite polymorphism
- Role of surface polarity and liquid structure on defect rate
- MD simulation with DFT/NN to study liquid ordering at the interfaces
- More to come: transferable model of the Ga_xAs_{1-x} binary



Mahdi Zamani et al., Adv. Materials (2020)

Synthesis of GaAs nanowires

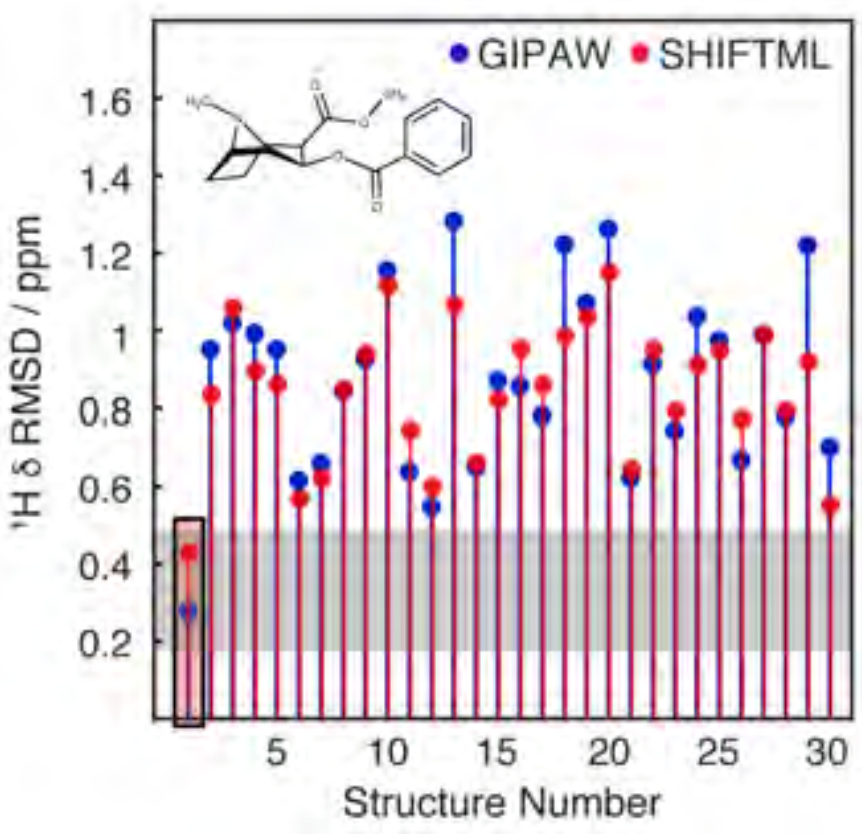
- Synthesis of GaAs nanostructures: solid-liquid-gas process, Ga_l/GaAs interface based on droplets of molten Ga
- Problem: controlling the zinc-blende/wurtzite polymorphism
- Role of surface polarity and liquid structure on defect rate
- MD simulation with DFT/NN to study liquid ordering at the interfaces
- More to come: transferable model of the Ga_xAs_{1-x} binary

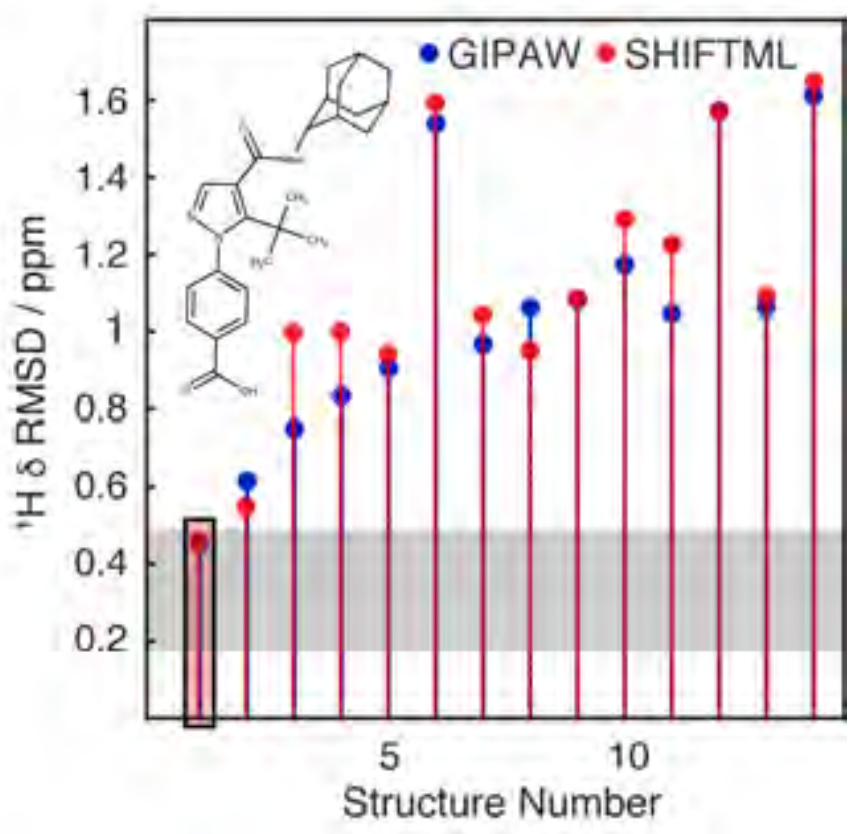


Imbalzano, **MC**, Phys. Rev. Mat. (2021)

ML-powered NMR crystallograpy

- Solid-state NMR relies on GIPAW-DFT to determine crystal structure
- ML model trained on 2000 CSD structures. DFT accuracy of chemical shielding predictions (RMSE H: 0.5, C: 5, N: 13, O: 18 ppm)
- Precise enough to do structure determination - try it on **shiftml.org**

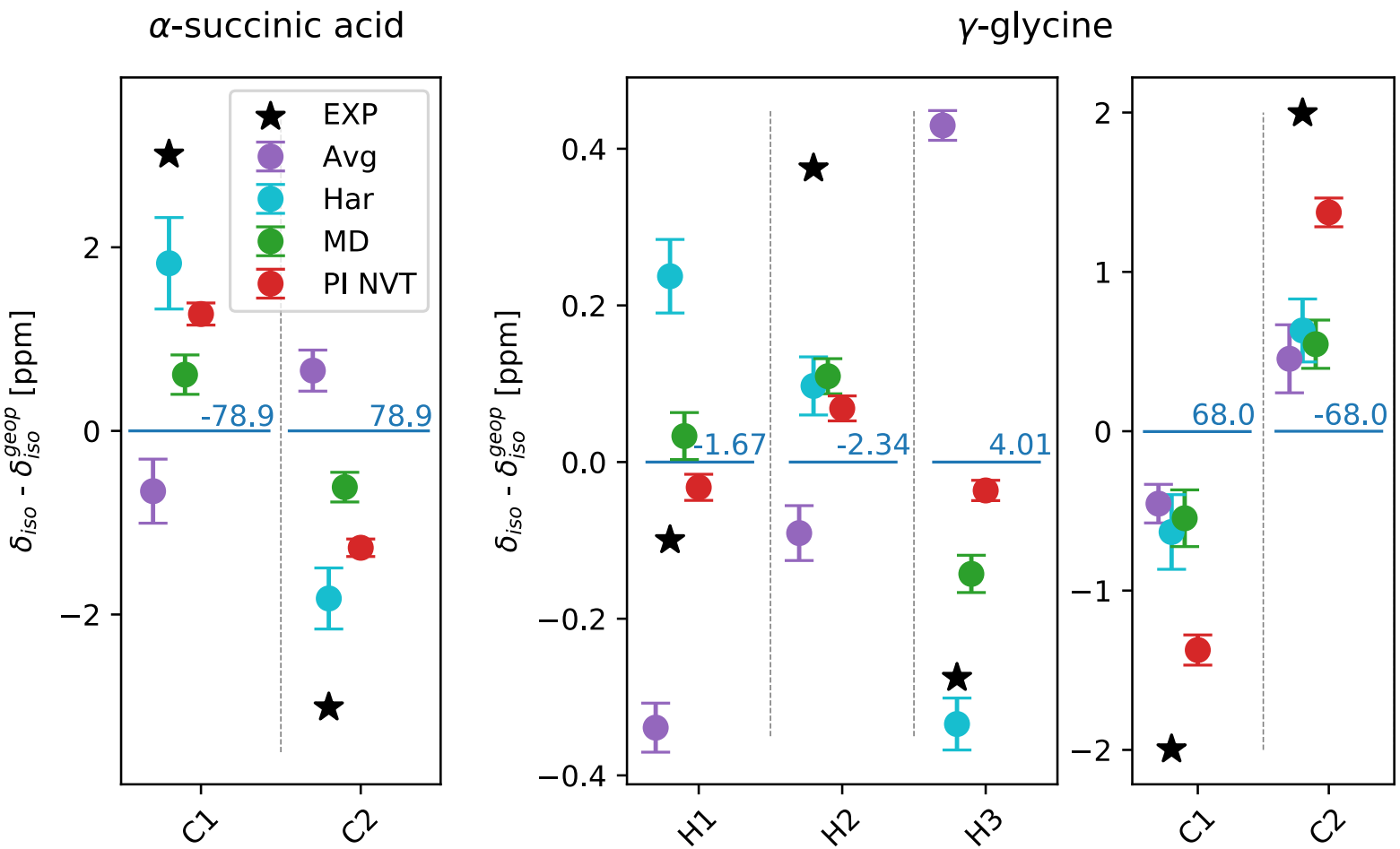




Paruzzo, Hofstetter, Musil, De, **MC**, Emsley, Nature Comm. (2018) *[data: CSD-500]*

Quantum nuclei in NMR experiments

- Combining a ML potential and NMR shielding model → full ab initio modeling of NMR at constant temperature
- Thermal and quantum fluctuations affect the NMR *shifts*: comparable to typical errors of reference DFT calculations

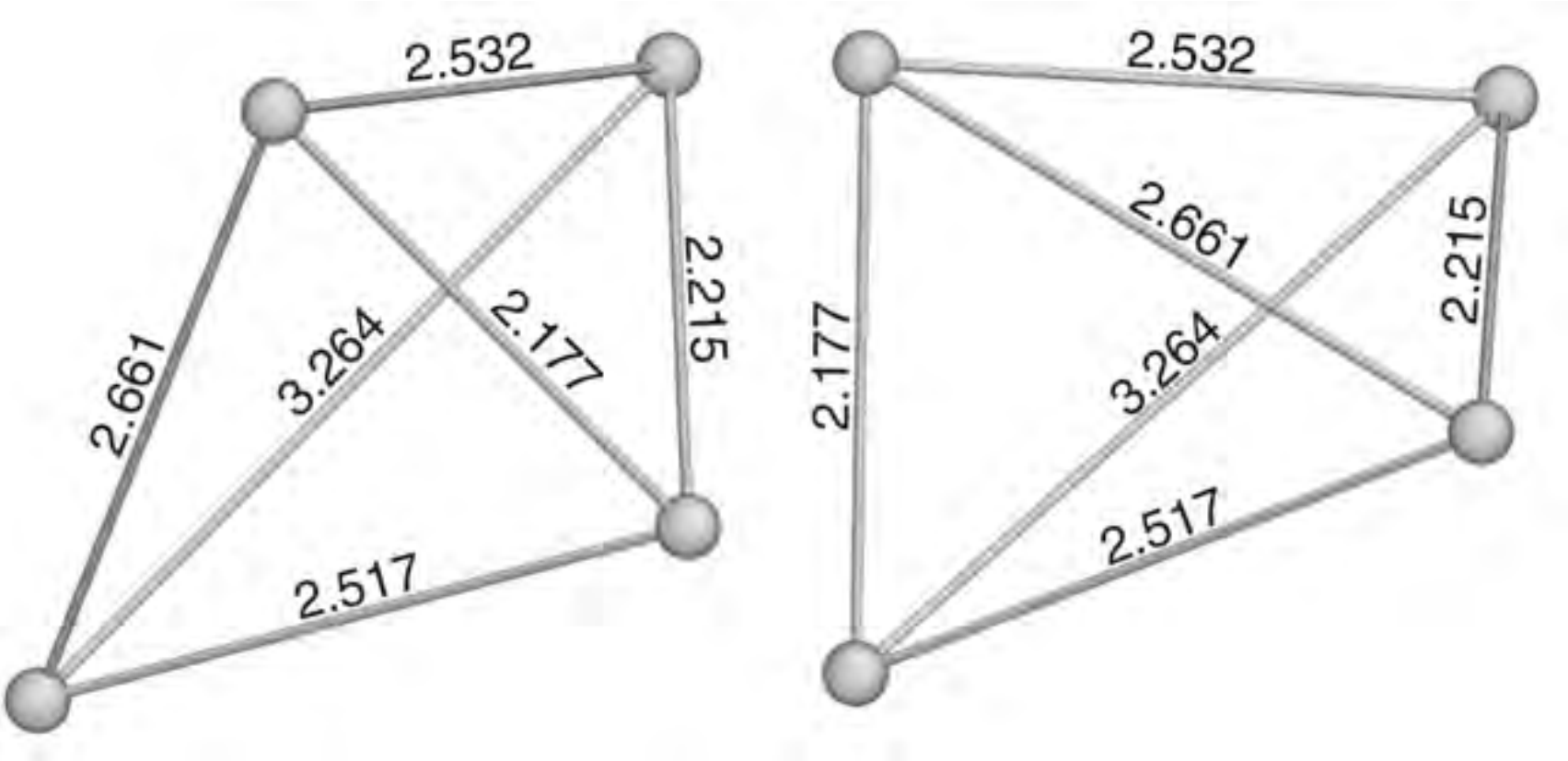


Engel, Kapil, **MC** J. Phys. Chem. Lett. (2021)

Consolidating the mathematical foundations

Are representations complete?

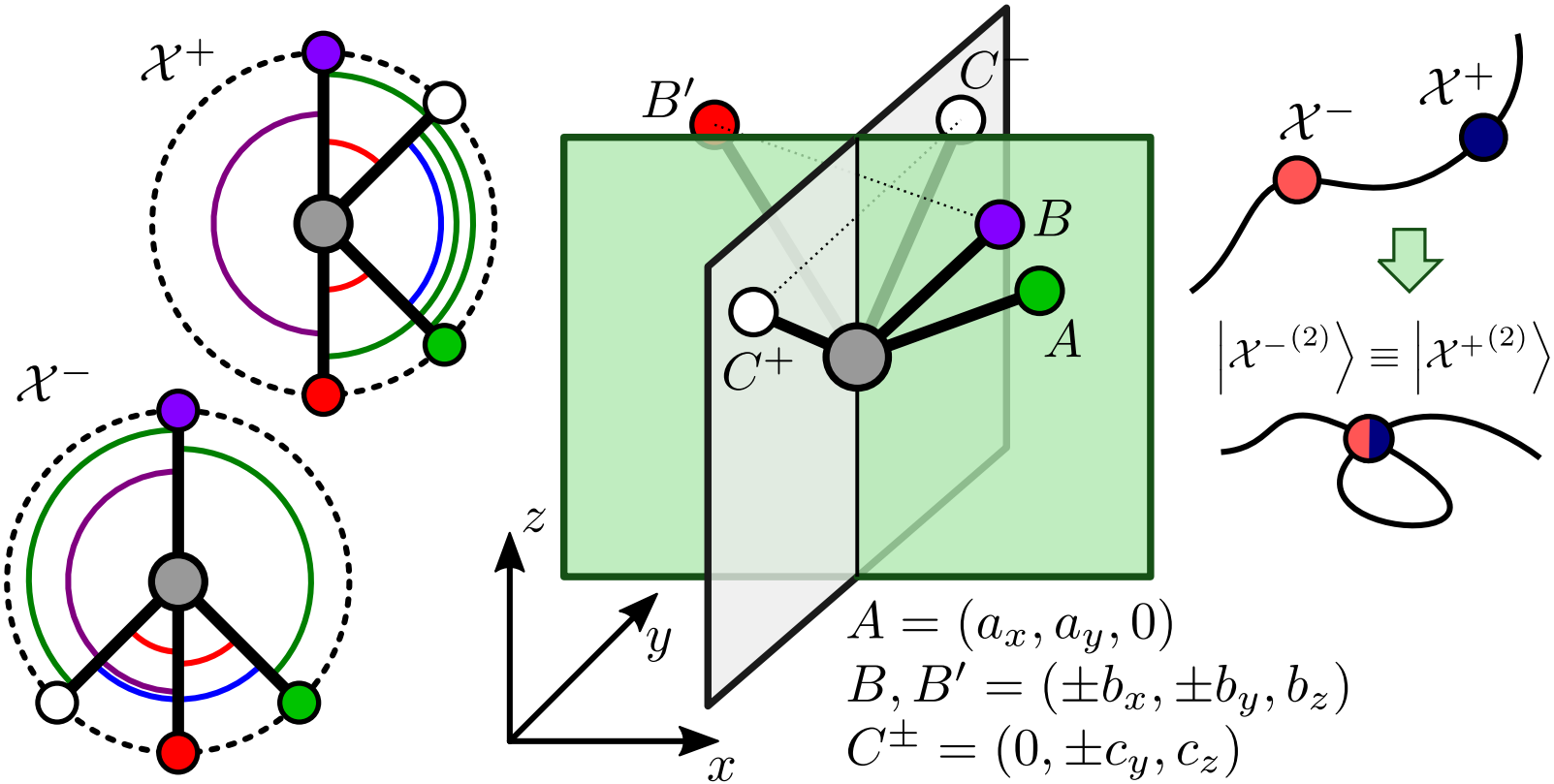
- More fundamental: are representations complete (injective)?
- Well-known: 2-body correlations (distances) are ambiguous
- Surprise: neither are 3 (angles) and 4 (dihedrals) body features!
- Limits the asymptotic accuracy of models



Boutin, Kemper, Ann. Adv. Math. (2004); Figure from Bartók, Kondor, Csányi, PRB (2013)

Are representations complete?

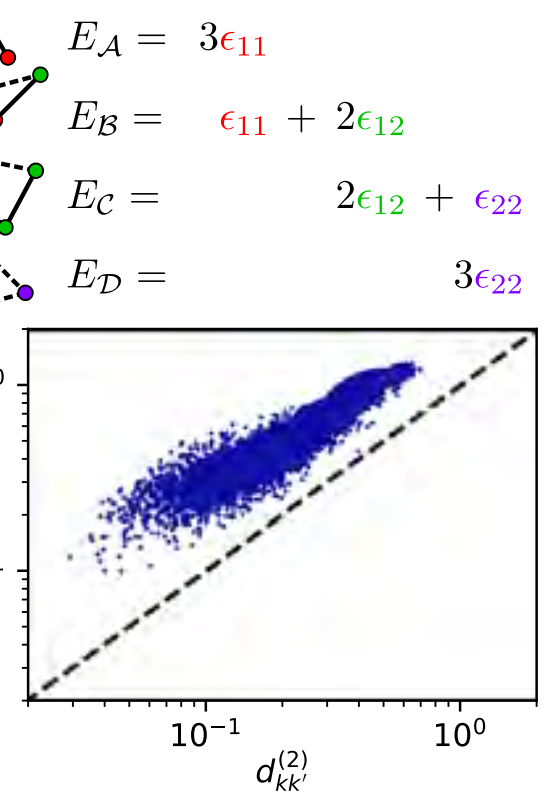
- More fundamental: are representations complete (injective)?
- Well-known: 2-body correlations (distances) are ambiguous
- Surprise: neither are 3 (angles) and 4 (dihedrals) body features!
- Limits the asymptotic accuracy of models

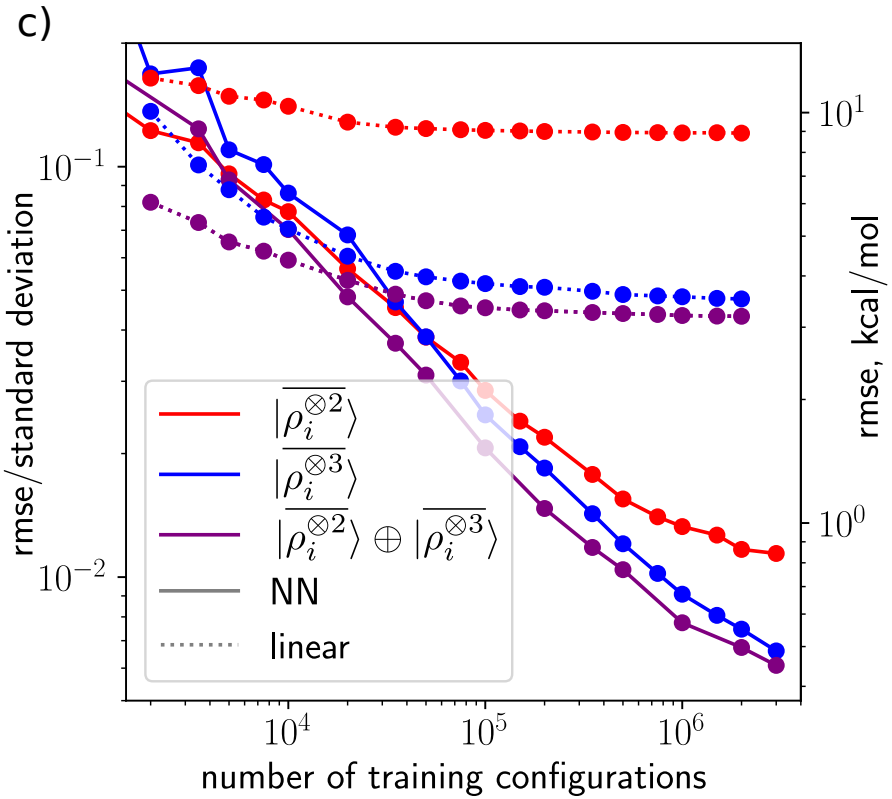


Pozdniakov, Willatt, Bartók, Ortner, Csányi, **MC** PRL (2020)

Are representations complete?

- More fundamental: are representations complete (injective)?
- Well-known: 2-body correlations (distances) are ambiguous
- Surprise: neither are 3 (angles) and 4 (dihedrals) body features!
- Limits the asymptotic accuracy of models

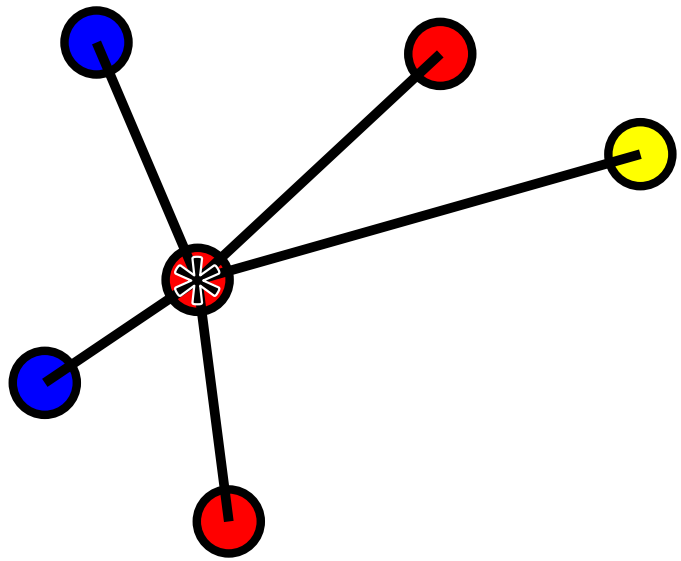




Pozdniakov, Willatt, Bartók, Ortner, Csányi, **MC** PRL (2020)

How about graph convolution?

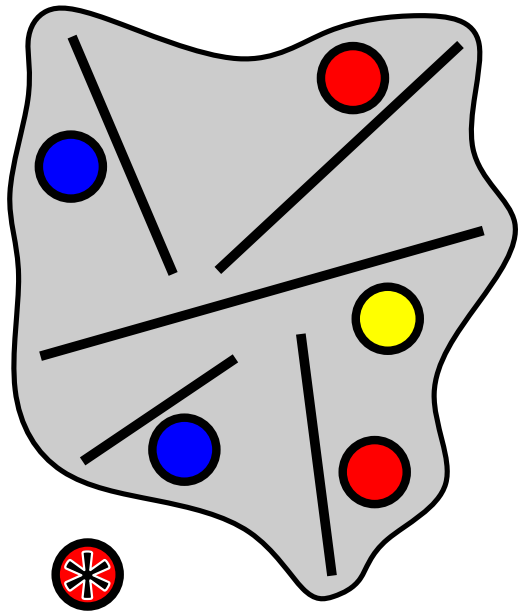
- Atoms: nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is augmented with information on its neighbors and their distance: $h(A_i) = (a_i, \{(a_j, r_{ij})\})$
- The multiset of neighbors and edges is hashed, and used as a label to describe the nodes. The process can be iterated



SchNET: Schütt et al., JCP (2018)

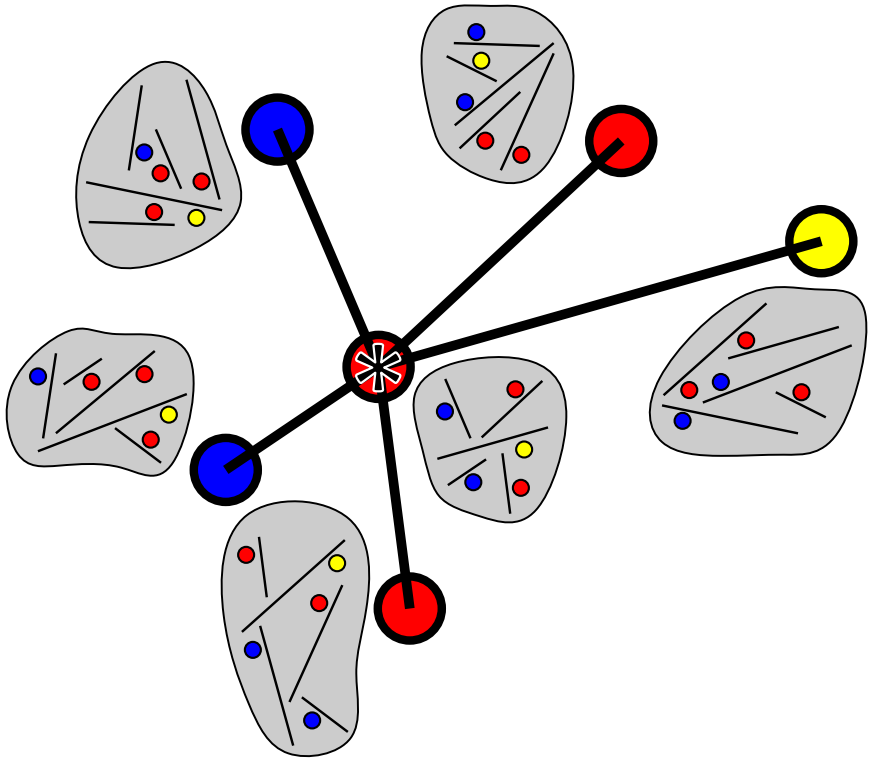
How about graph convolution?

- Atoms: nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is augmented with information on its neighbors and their distance: $h(A_i) = \left(a_i, \left\{(a_j, r_{ij})\right\}\right)$
- The multiset of neighbors and edges is hashed, and used as a label to describe the nodes. The process can be iterated



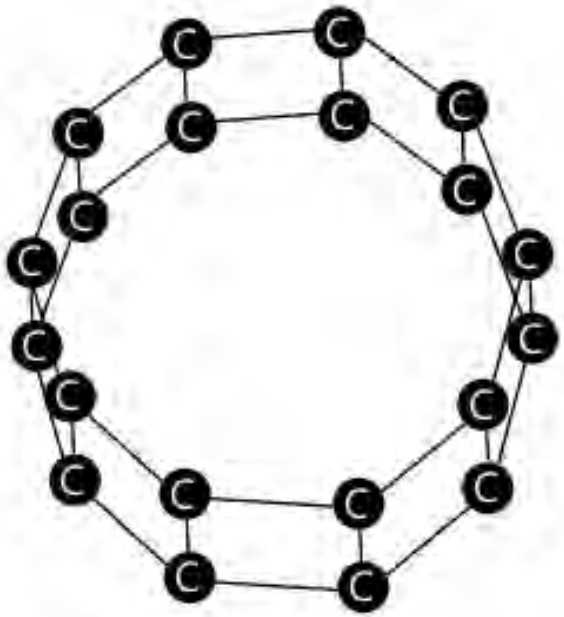
How about graph convolution?

- Atoms: nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is augmented with information on its neighbors and their distance: $h(A_i) = \left(a_i, \left\{(a_j, r_{ij})\right\}\right)$
- The multiset of neighbors and edges is hashed, and used as a label to describe the nodes. The process can be iterated

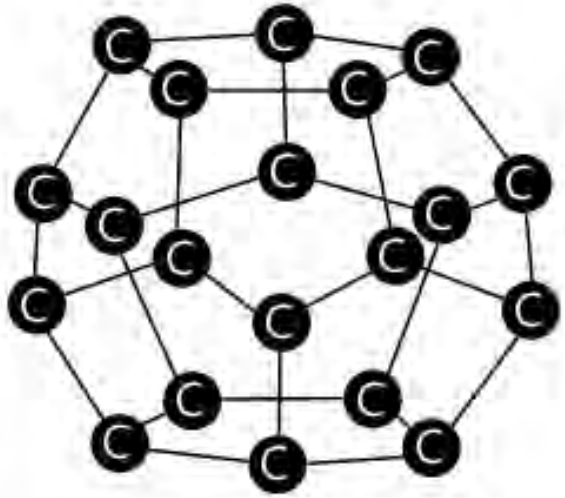


Graph convolution, pros and cons

- Bad news: there are known *discrete* graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things look good for molecular graphs (*fully-connected*, distance-decorated 3D point clouds);
- Distance-GNN resolve all known ACDC counterexamples



(a) Decaprismane.

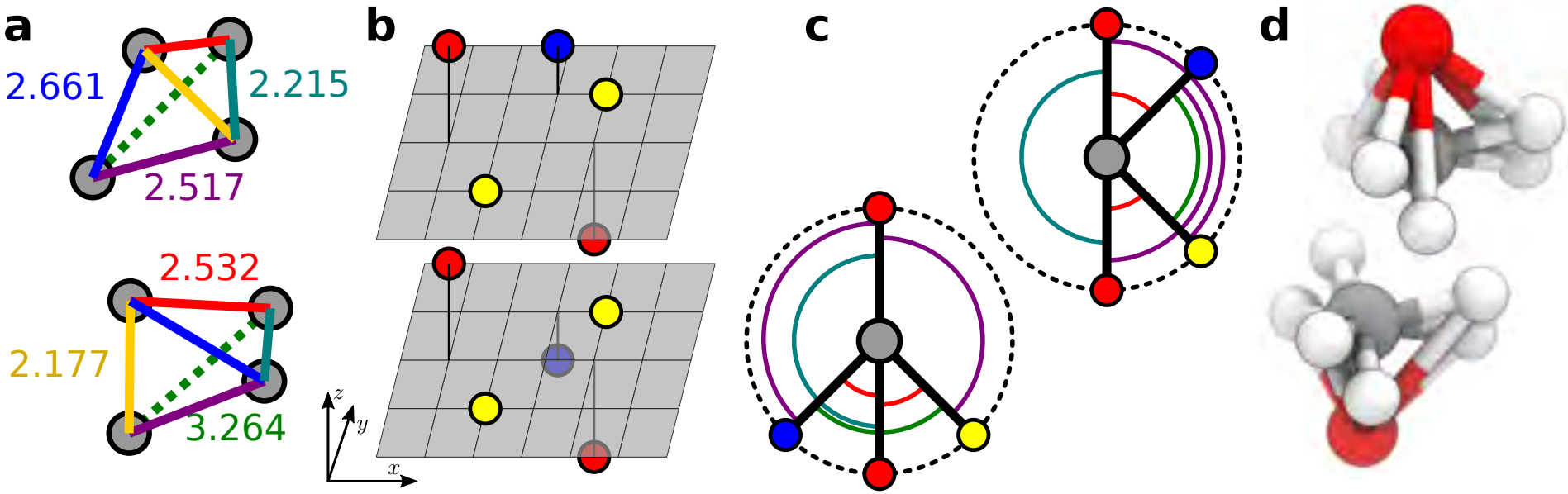


(b) Dodecahedrane.

Sato, arxiv:2003.04078

Graph convolution, pros and cons

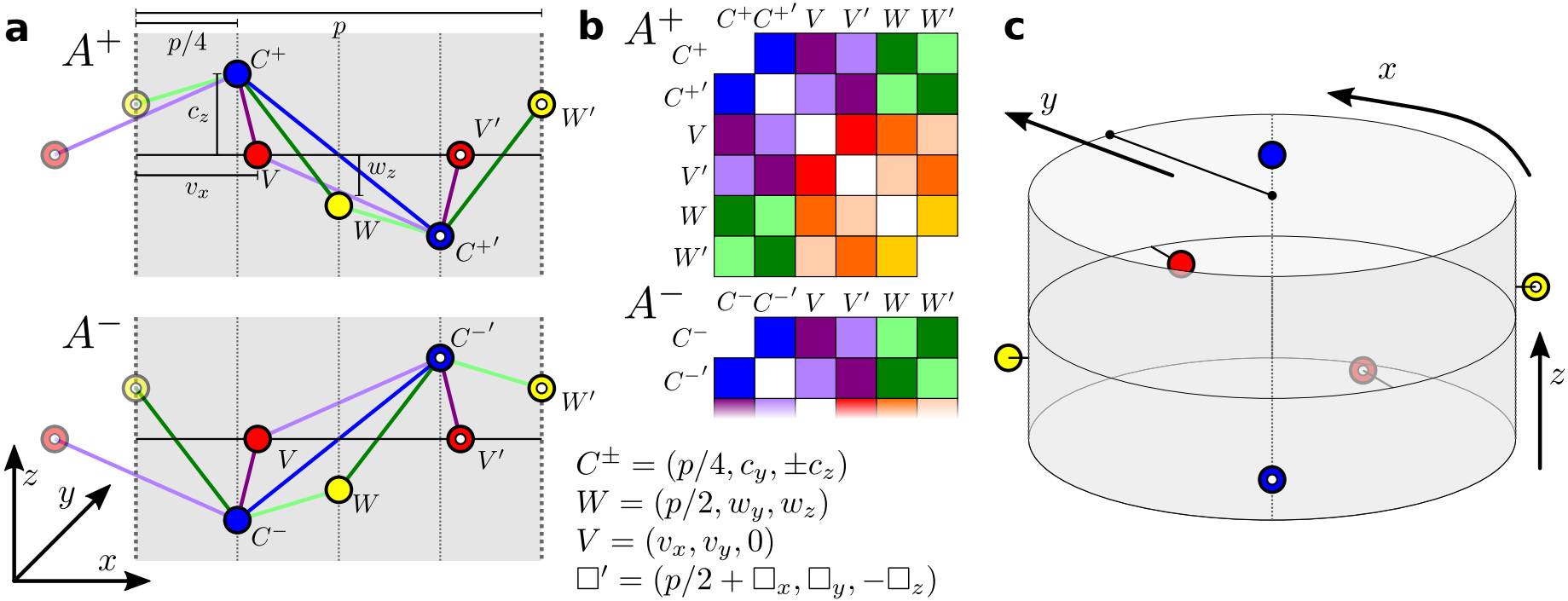
- Bad news: there are known *discrete* graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things look good for molecular graphs (*fully-connected*, distance-decorated 3D point clouds);
- Distance-GNN resolve all known ACDC counterexamples



Bartók et al. PRB (2013); Pozdnyakov et al. PRL (2020)

A counterexample for distance-based GNN

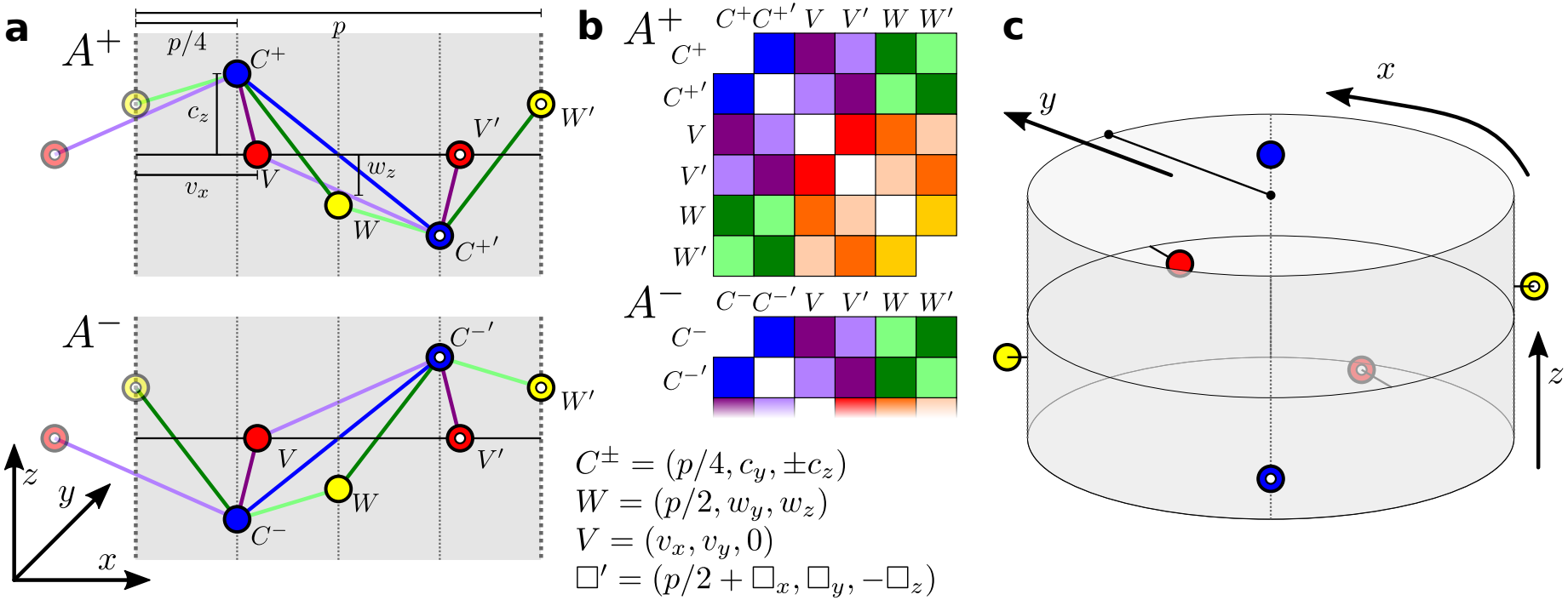
- A family of 3D point clouds with degenerate pairs for GNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and simple models based on $|\overline{\rho_i^{\otimes 2}}\rangle$) are immune



Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

A counterexample for distance-based GNN

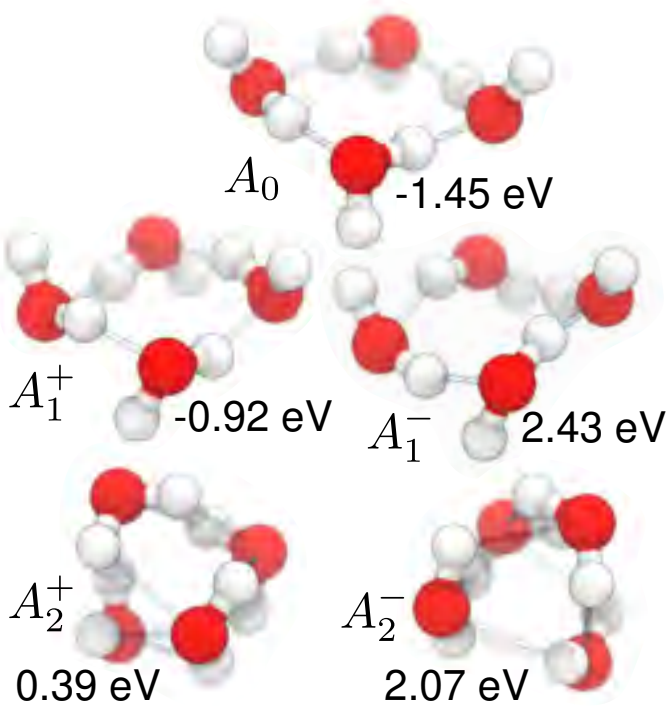
- A family of 3D point clouds with degenerate pairs for GNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and simple models based on $|\overline{\rho_i^{\otimes 2}}\rangle$) are immune

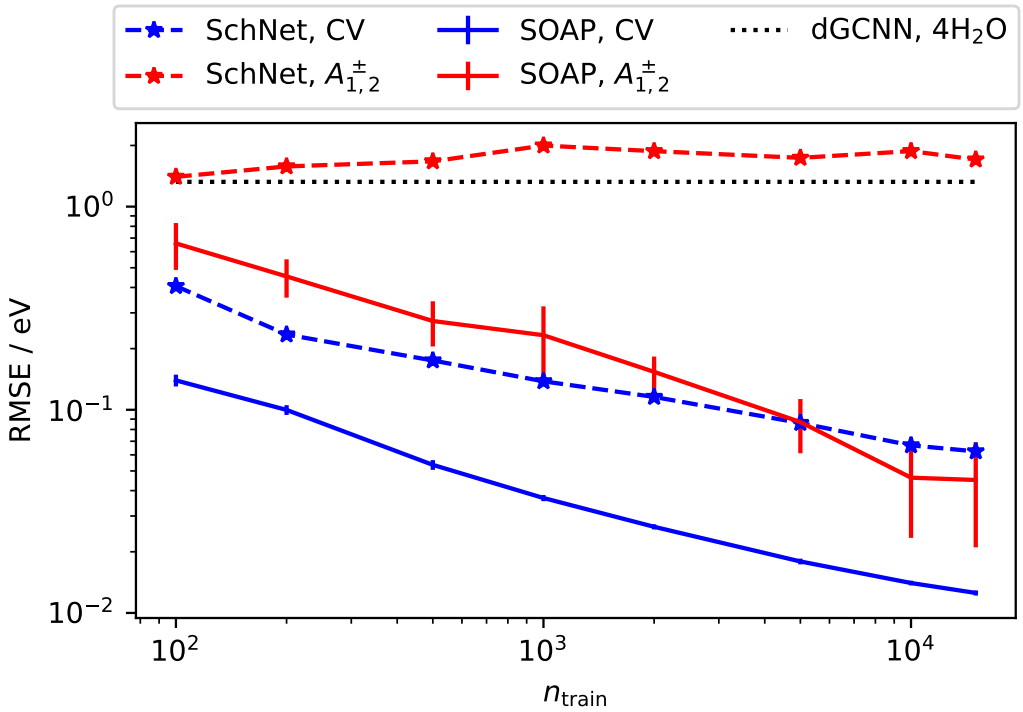


Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

A counterexample for distance-based GNN

- A family of 3D point clouds with degenerate pairs for GNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and simple models based on $|\overline{\rho_i^{\otimes 2}}\rangle$) are immune

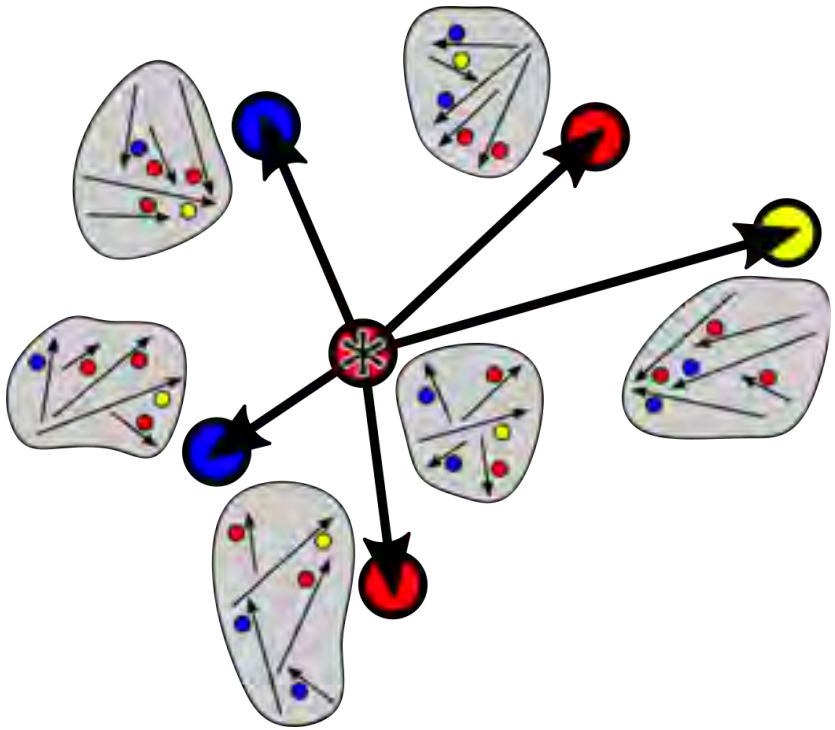




Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

A counterexample for distance-based GNN

- A family of 3D point clouds with degenerate pairs for GNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and simple models based on $|\overline{\rho_i^{\otimes 2}}\rangle$) are immune

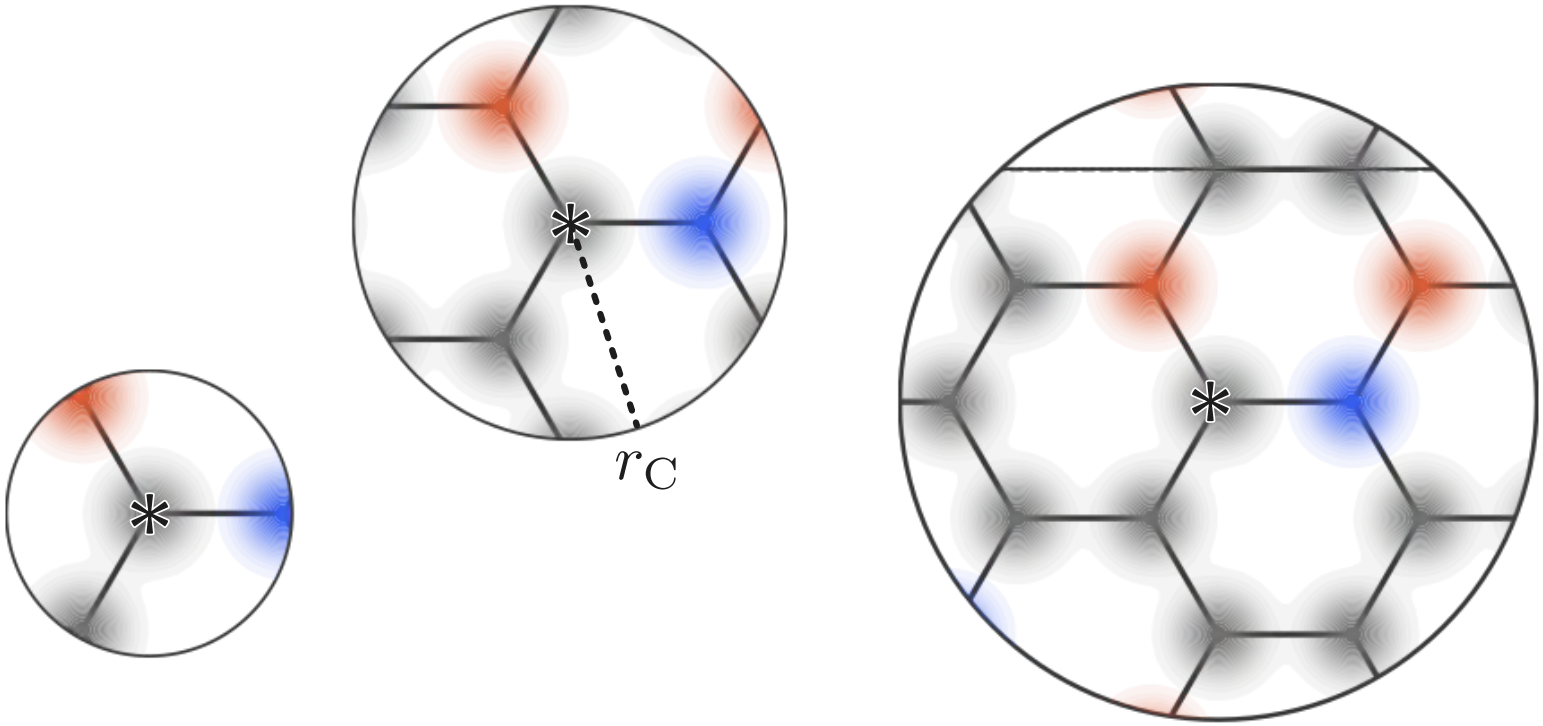


Nigam, Fraux, **MC**, arXiv:2202.01566 (2022)

The charged elephant in the other room

Understanding the range of interactions

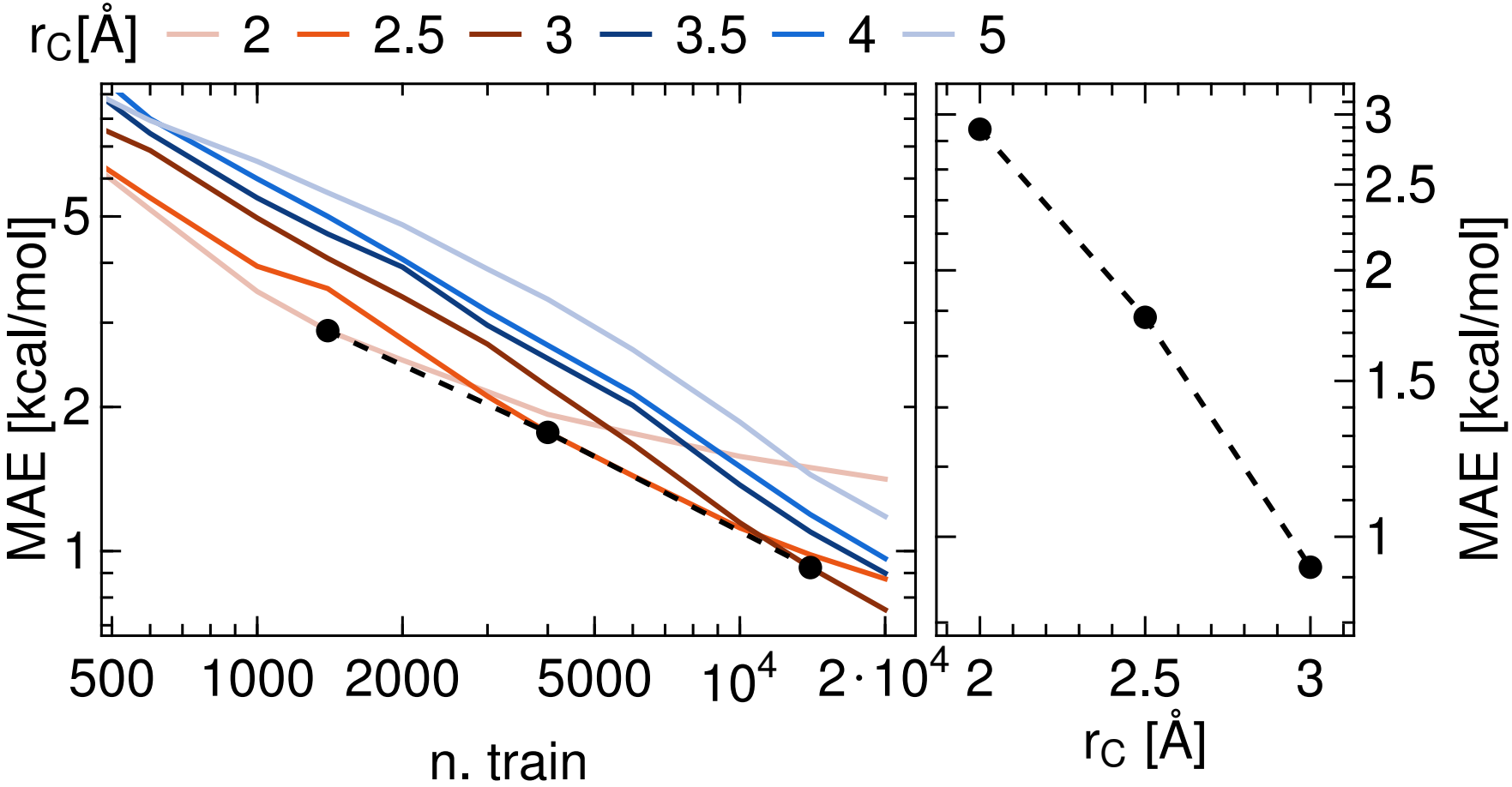
- Representations are built for different cutoff radii
- Dimensionality/accuracy tradeoff: a measure of the range of interactions
- Multi-scale kernels $K(A,B) = \sum_i w_i K_i(A,B)$ yield the best of all worlds



Bartók, De, Poelking, Kermode, Bernstein, Csányi, **MC**, Science Advances (2017) *[data: QM9, von Lilienfeld&C]*

Understanding the range of interactions

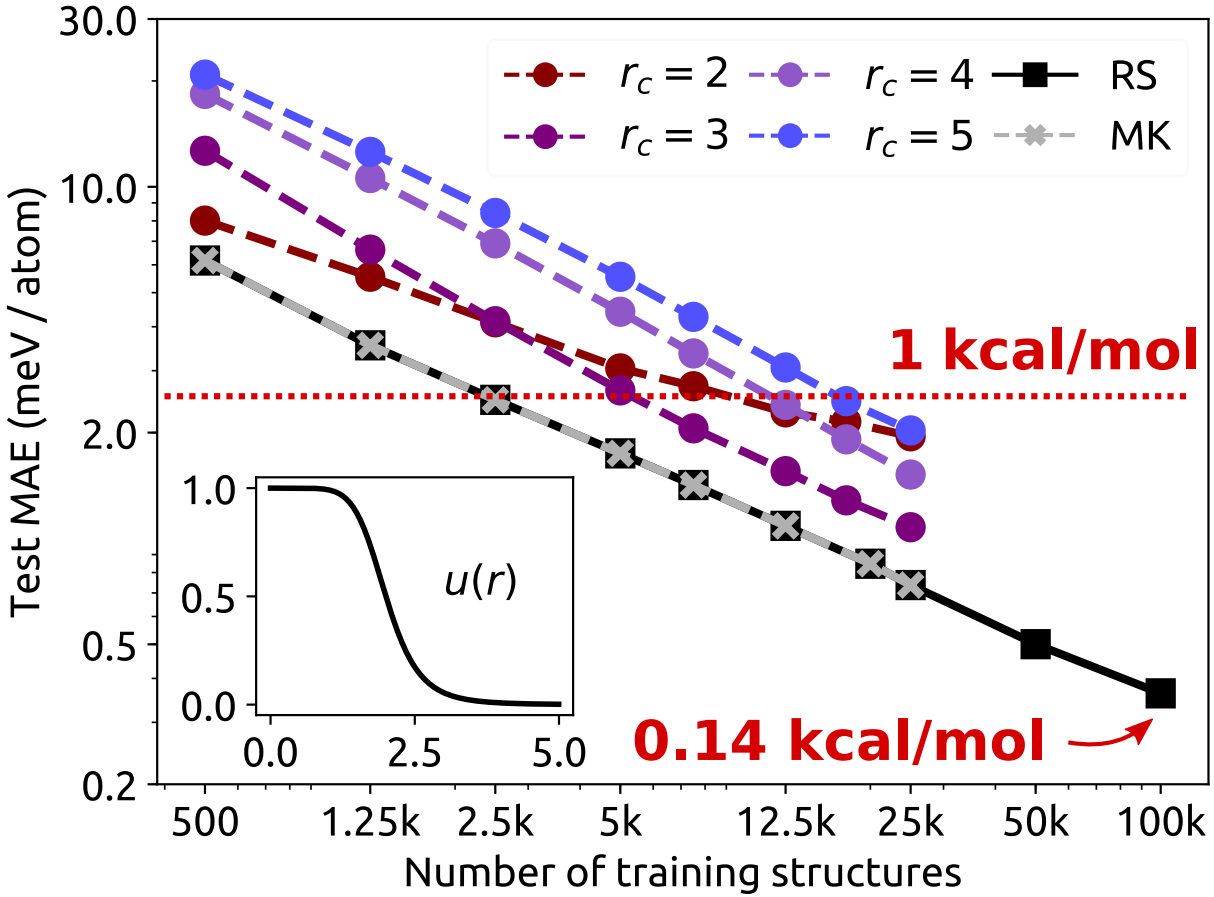
- Representations are built for different cutoff radii
- Dimensionality/accuracy tradeoff: a measure of the range of interactions
- Multi-scale kernels $K(A,B) = \sum_i w_i K_i(A,B)$ yield the best of all worlds



Willatt, Musil, **MC**, PCCP (2018)

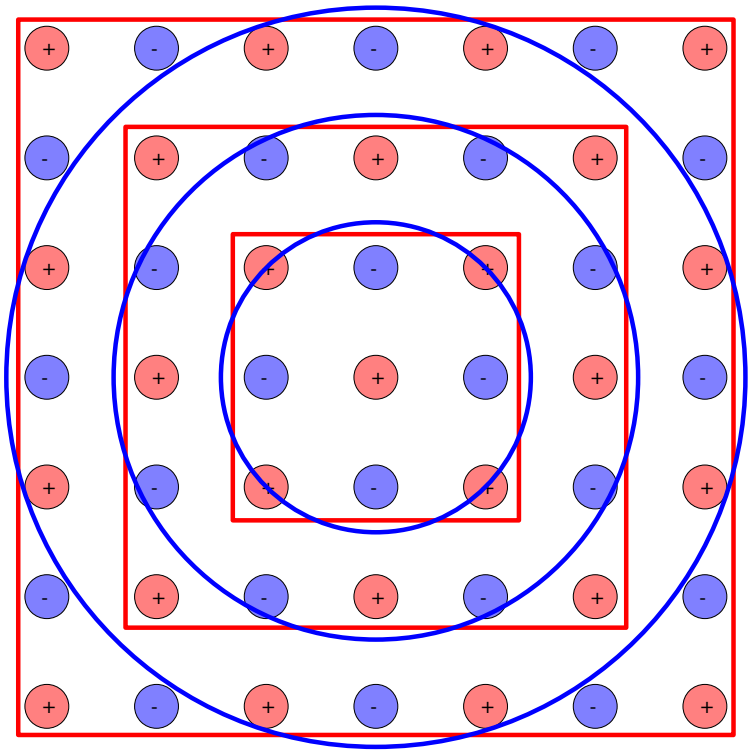
Understanding the range of interactions

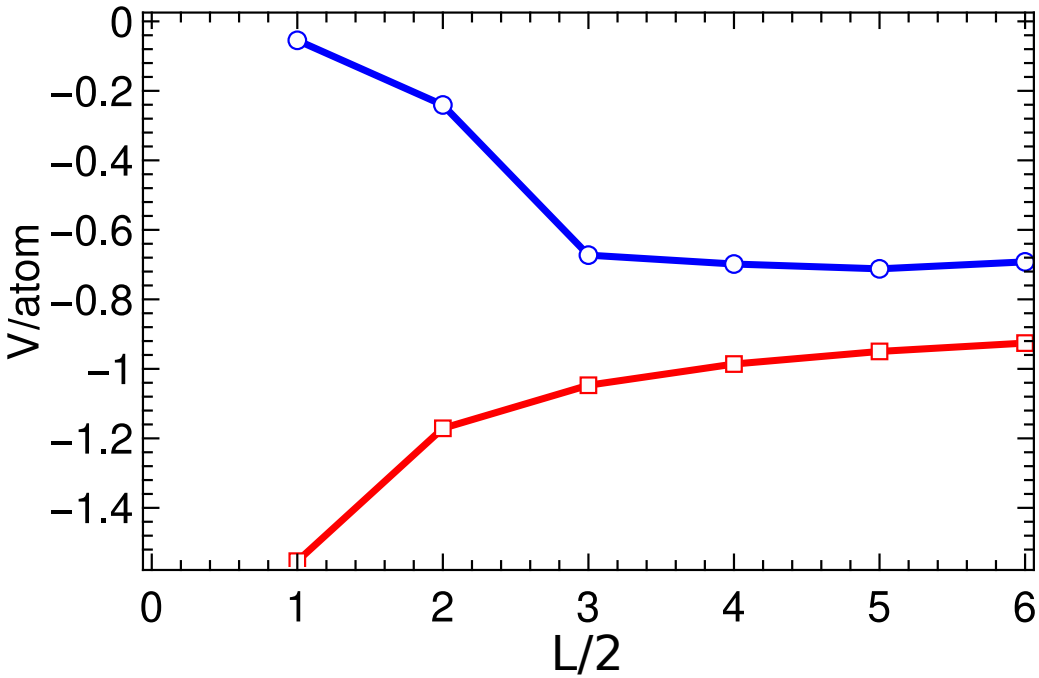
- Representations are built for different cutoff radii
- Dimensionality/accuracy tradeoff: a measure of the range of interactions
- Multi-scale kernels $K(A,B) = \sum_i w_i K_i(A,B)$ yield the best of all worlds



The problem with electrostatics

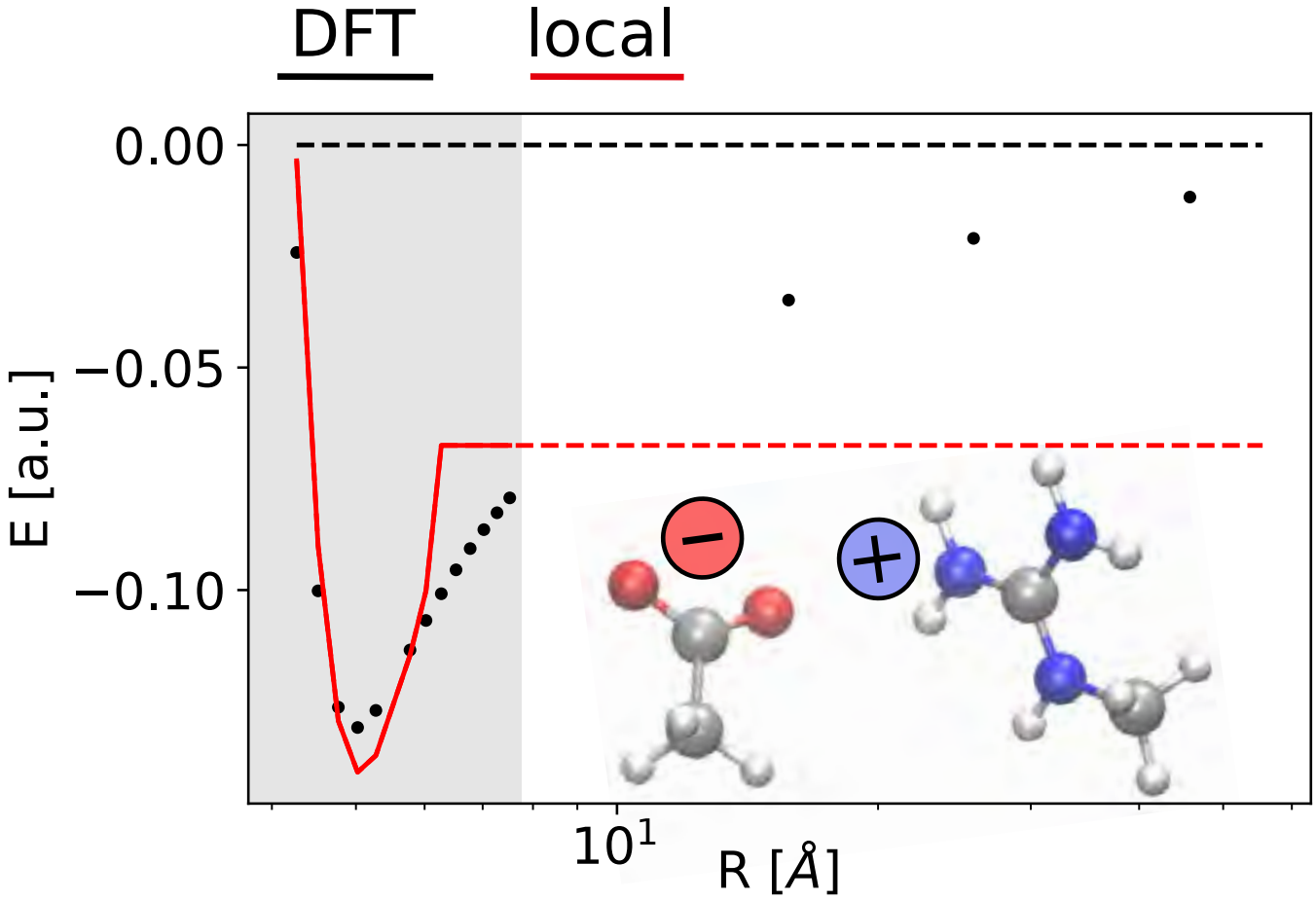
- $1/r$ decay $\rightarrow$ pathological convergence of with interaction cutoff
- Capturing true long-range effects with local models is hopeless





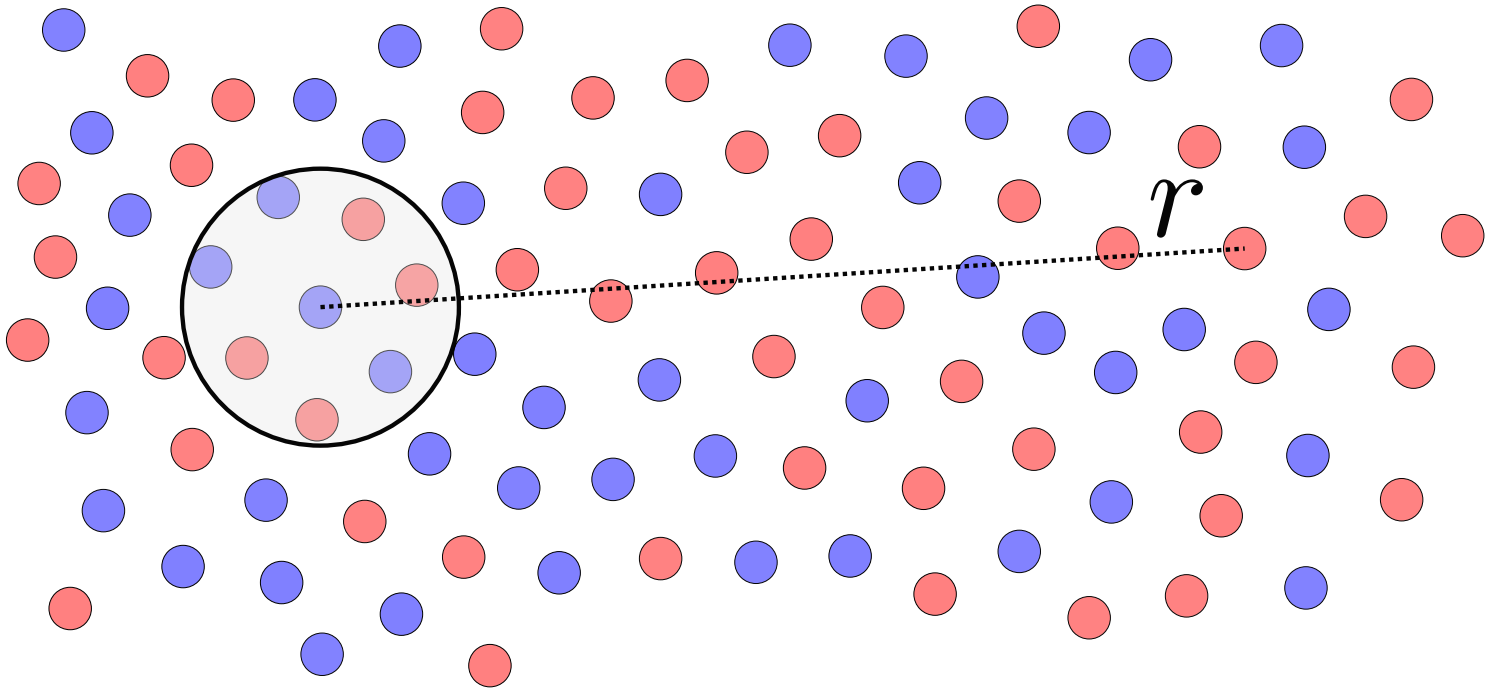
The problem with electrostatics

- $1/r$ decay $\rightarrow$ pathological convergence of with interaction cutoff
- Capturing true long-range effects with local models is hopeless



Long-distance equivariant representation

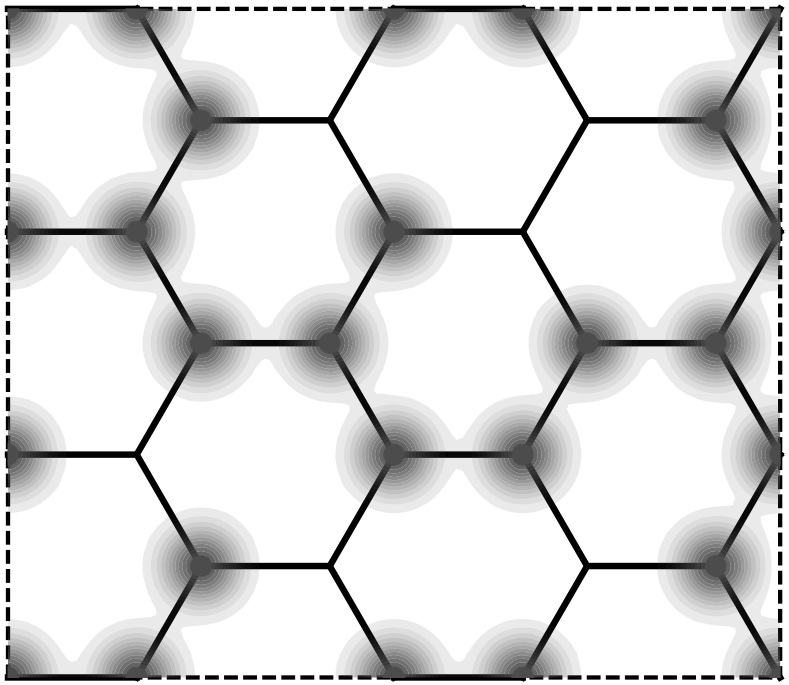
- Idea: *local* representation that reflects long-range *asymptotics*
  1. Atom-density potential $\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$
  2. Usual gig: symmetrize, decompose locally, learn!
- Efficient evaluation in reciprocal space



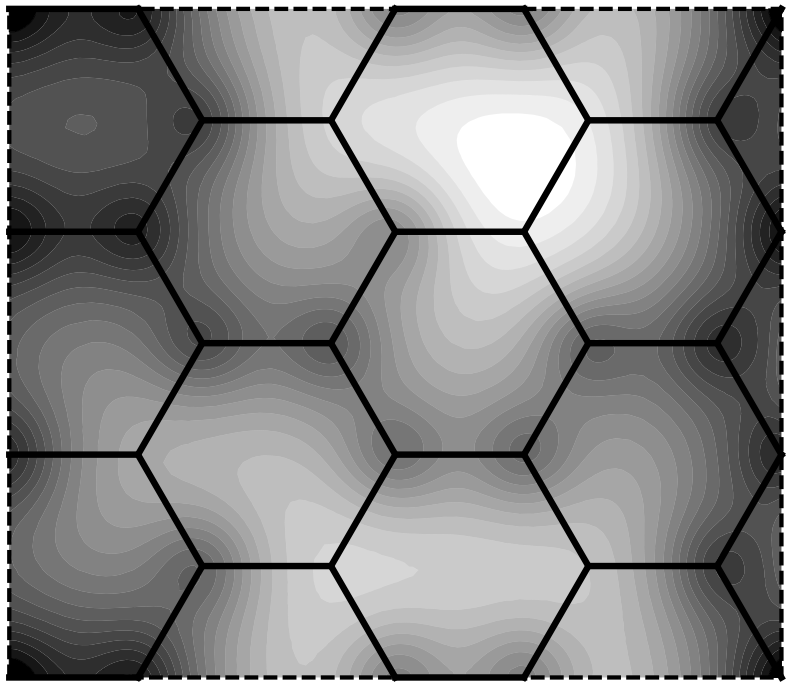
Grisafi, **MC**, JCP (2019)

Long-distance equivariant representation

- Idea: *local* representation that reflects long-range *asymptotics*
  1. Atom-density potential $\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$
  2. Usual gig: symmetrize, decompose locally, learn!
- Efficient evaluation in reciprocal space



$$\langle a\mathbf{r}|\rho\rangle = \sum_i \delta_{aa_i} g(\mathbf{r} - \mathbf{r}_i)$$

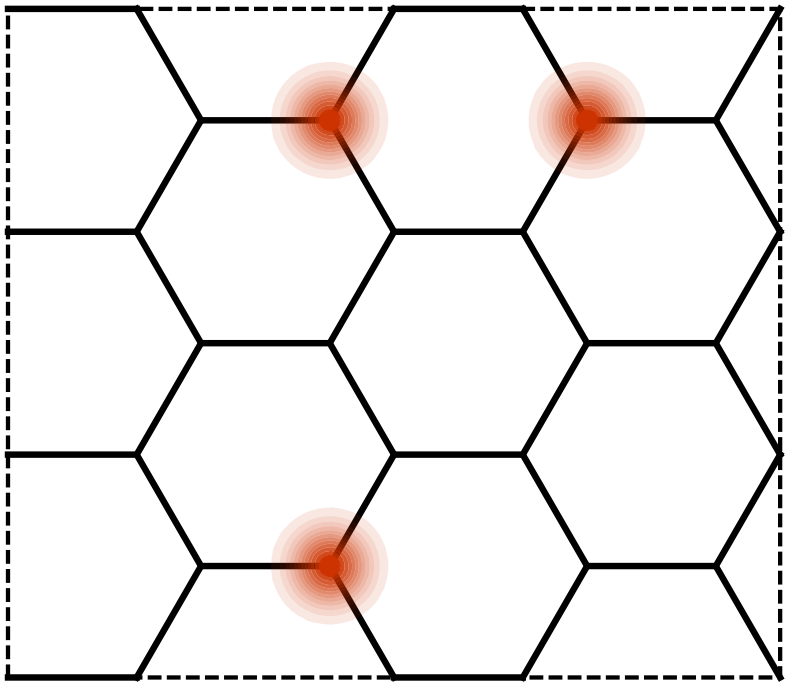


$$\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$$

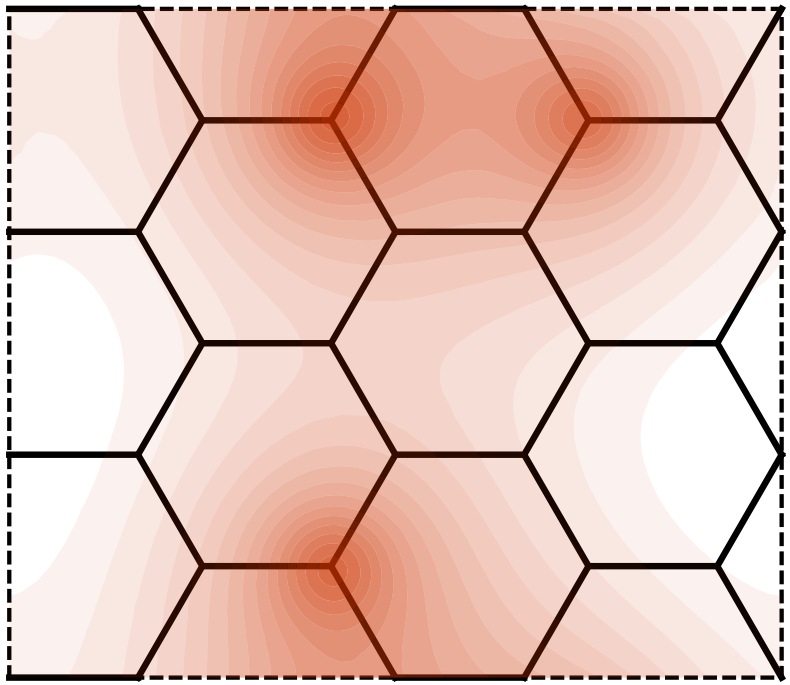
Grisafi, **MC**, JCP (2019)

Long-distance equivariant representation

- Idea: *local* representation that reflects long-range *asymptotics*
  1. Atom-density potential $\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$
  2. Usual gig: symmetrize, decompose locally, learn!
- Efficient evaluation in reciprocal space



$$\langle a\mathbf{r}|\rho\rangle = \sum_i \delta_{aa_i} g(\mathbf{r} - \mathbf{r}_i)$$

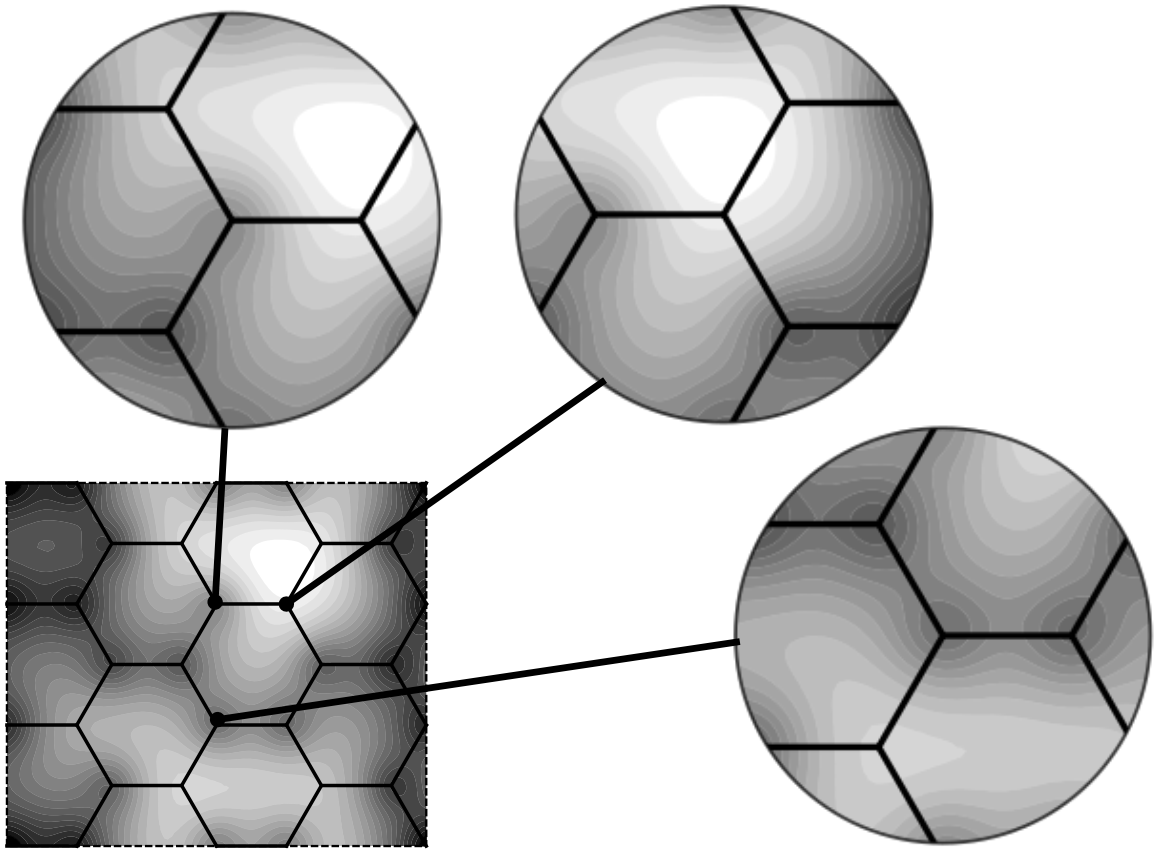


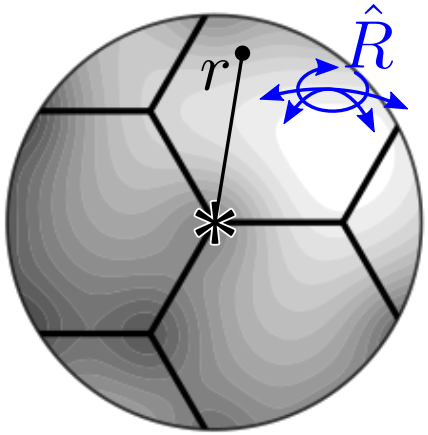
$$\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$$

Grisafi, **MC**, JCP (2019)

Long-distance equivariant representation

- Idea: *local* representation that reflects long-range *asymptotics*
  1. Atom-density potential $\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$
  2. Usual gig: symmetrize, decompose locally, learn!
- Efficient evaluation in reciprocal space

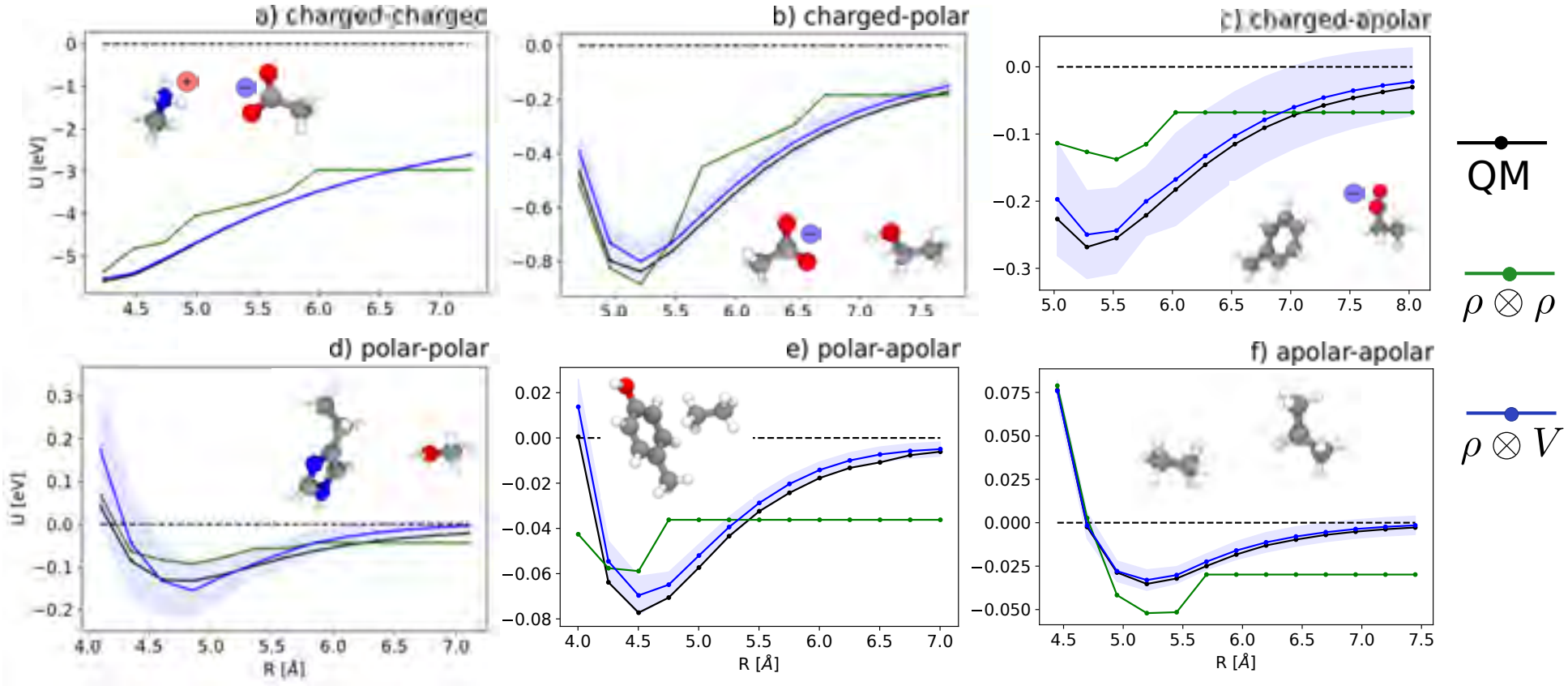




Grisafi, **MC**, JCP (2019)

Binding of charged molecules, and beyond

- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- "Multi-scale" LODE features $|\overline{\rho_i \otimes V_i}\rangle$ map to multipole electrostatics but enable learning all sorts of long-range physics

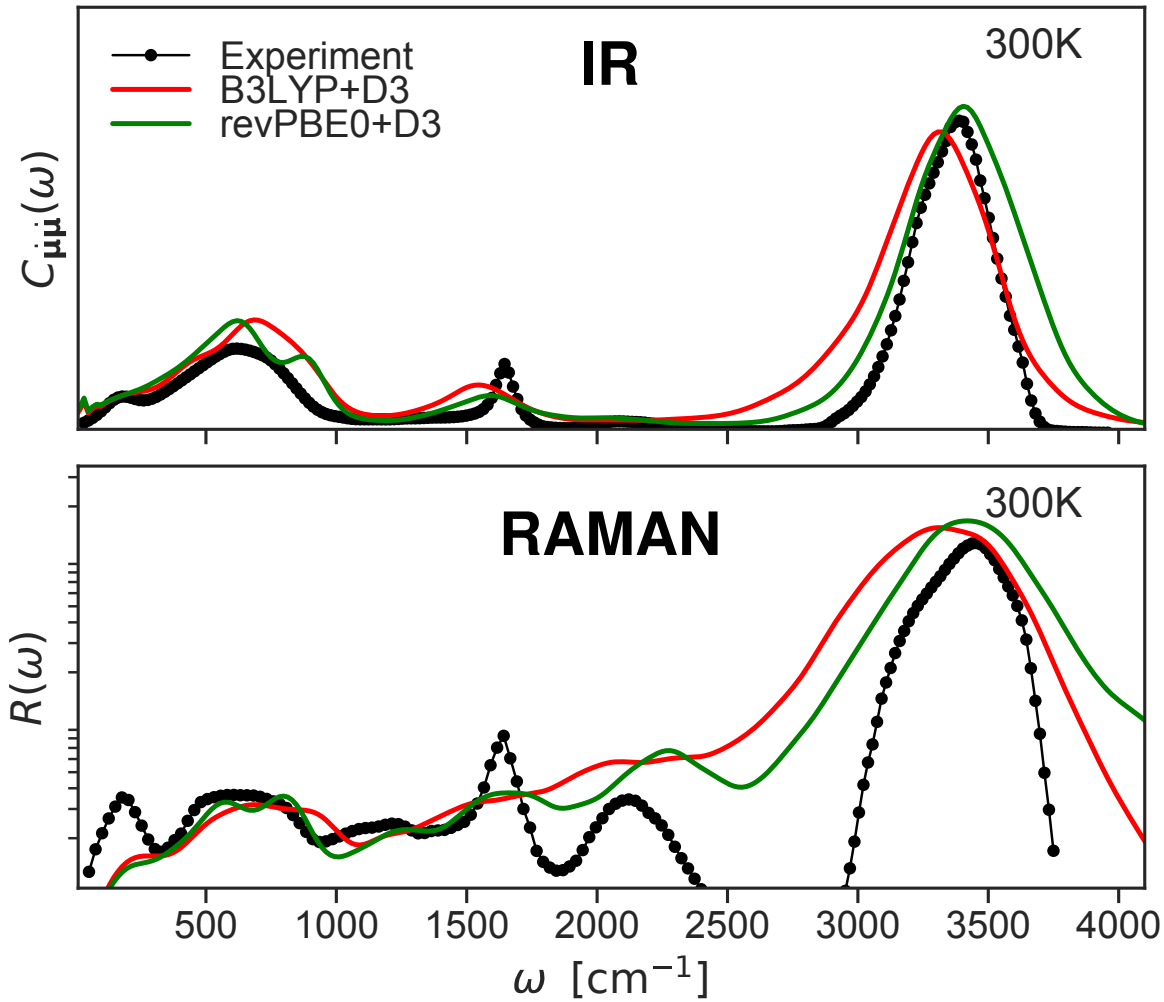


Grisafi, **MC**, JCP (2019); Grisafi, Nigam, **MC**, Chem. Sci. (2021)

Integrated ML models beyond size and time limits

Structural and functional properties, combined

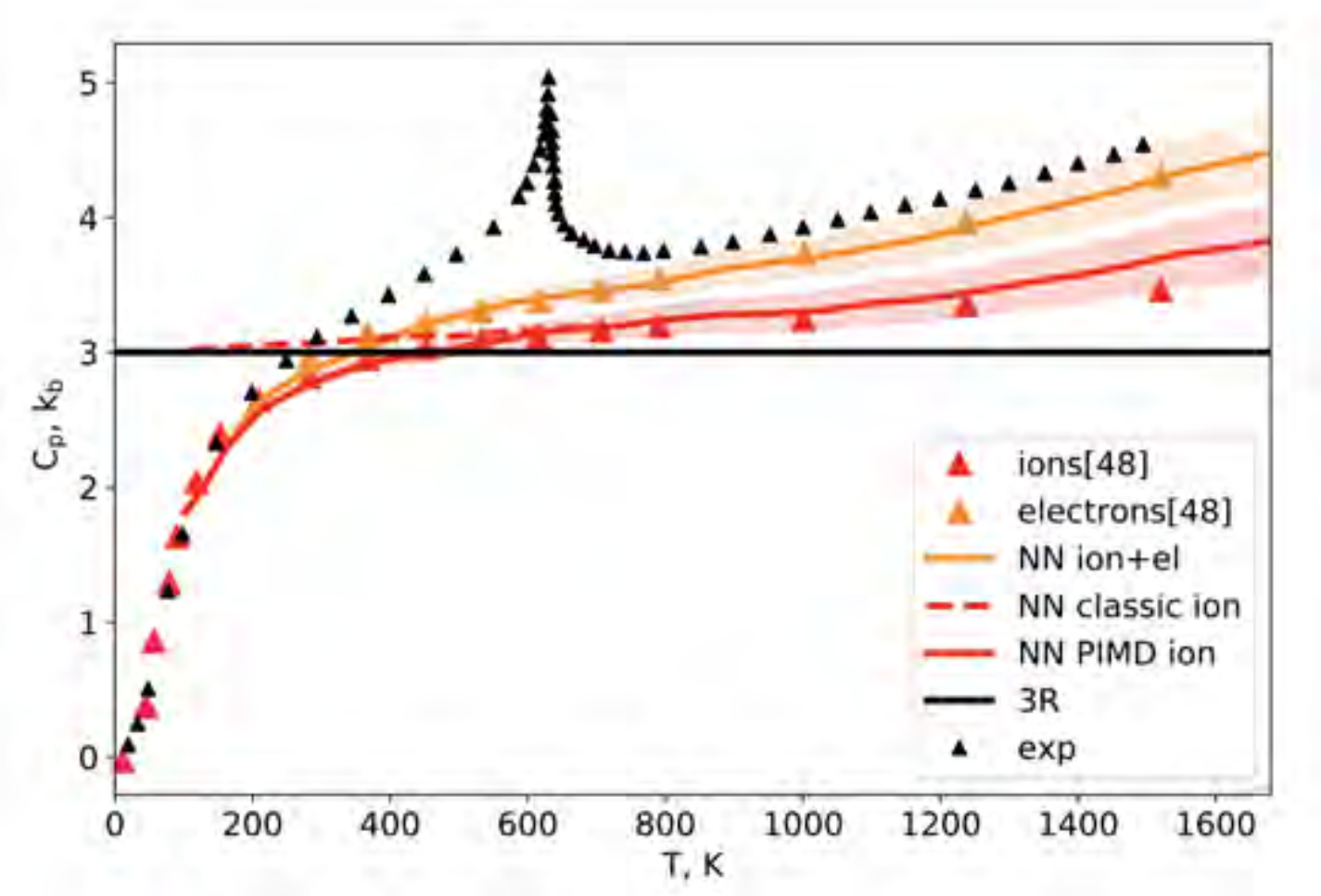
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



Kapil, Wilkins, Lan, **MC**, JCP (2020)

Structural and functional properties, combined

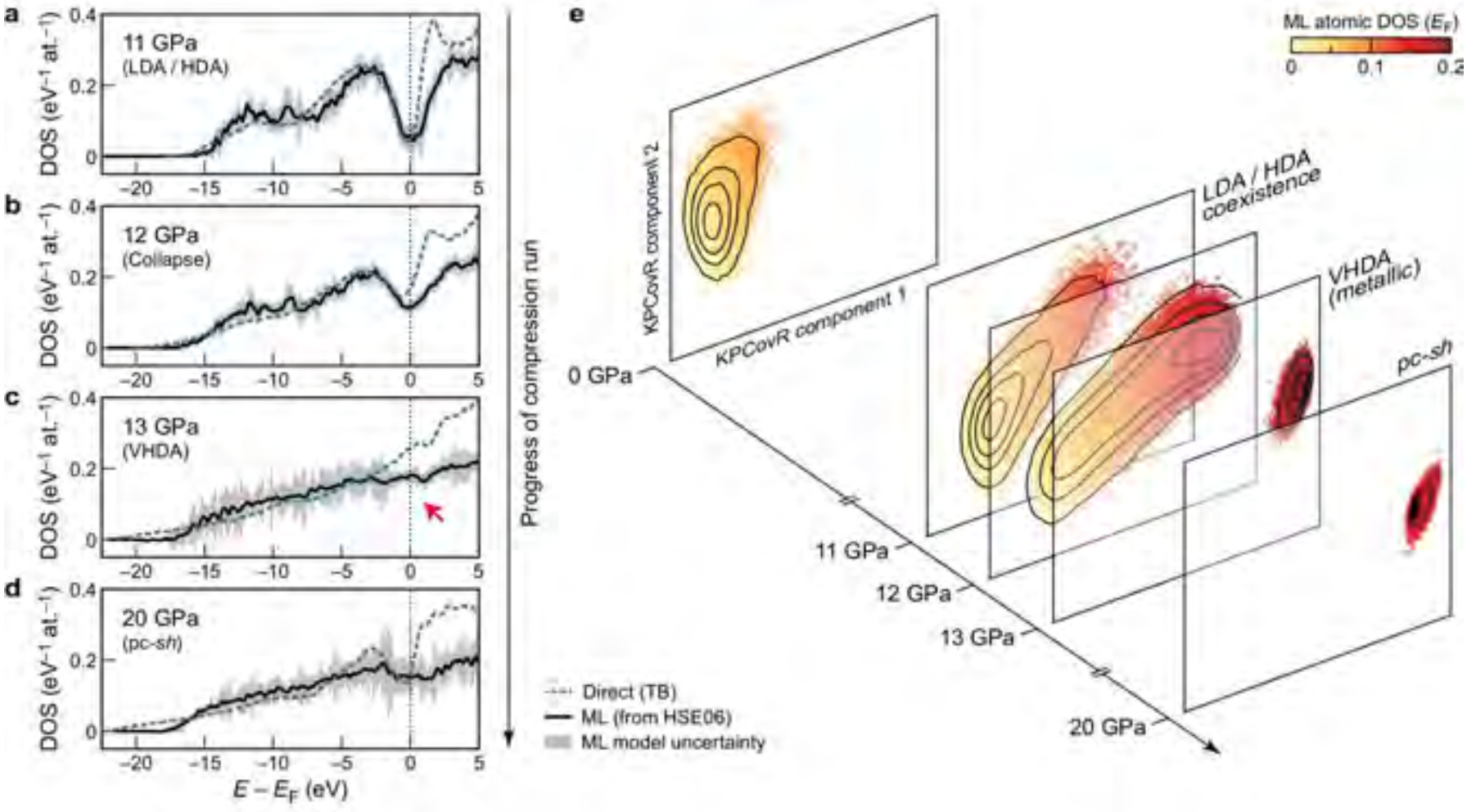
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



N. Lopanitsyna, C. Ben Mahmoud, **MC**, Phys. Rev. Mater. (2021)

Structural and functional properties, combined

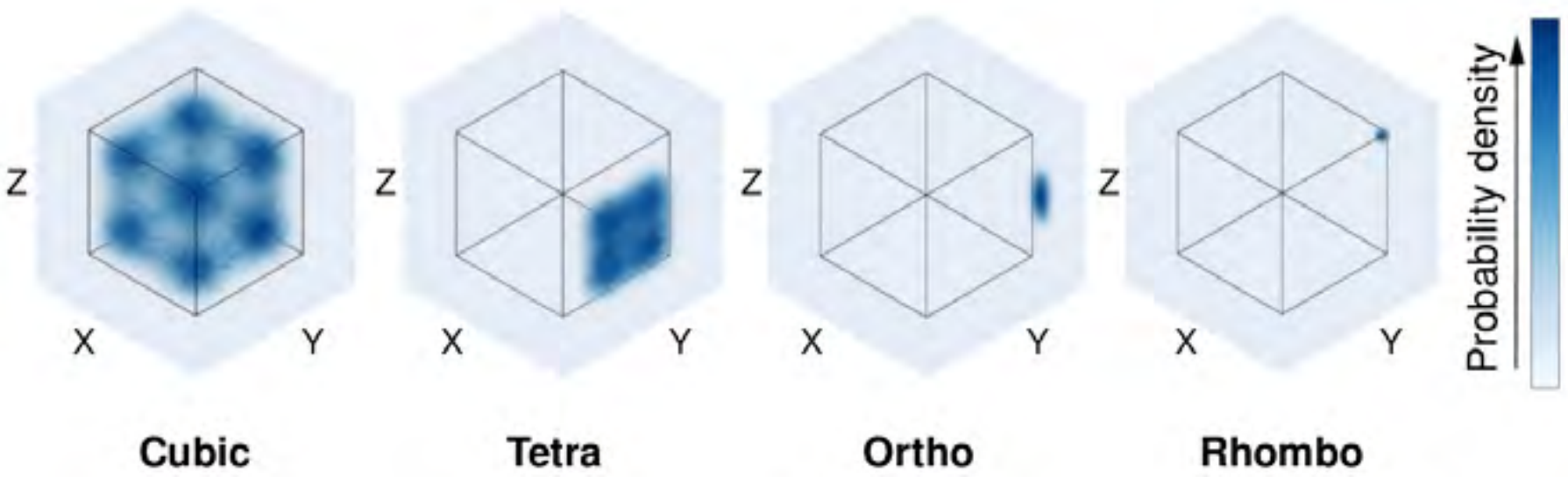
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



V. Deringer et al., Nature (2021)

Structural and functional properties, combined

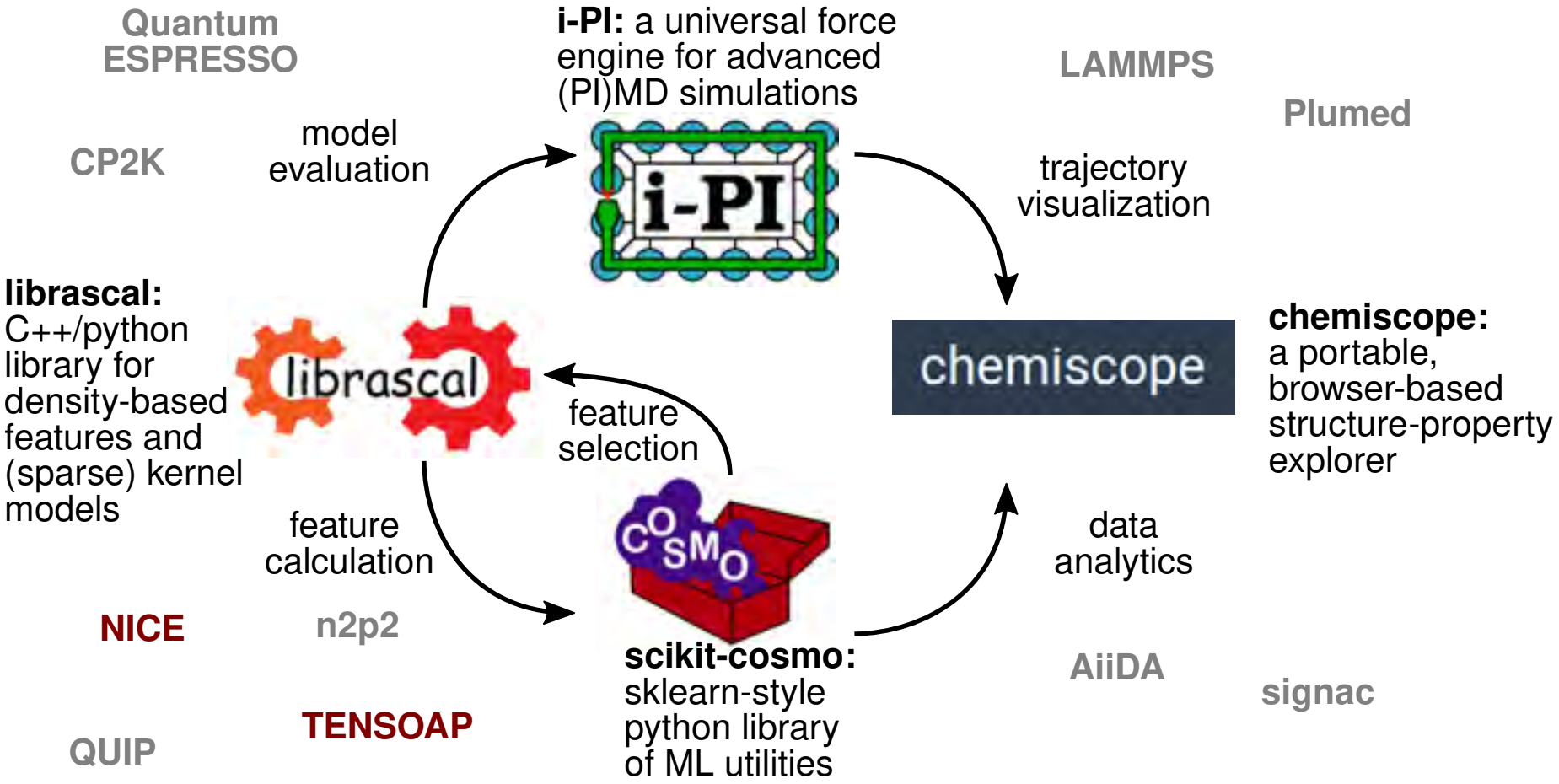
- Predicting *any* property accessible to quantum calculations
- Realistic time and size scales, with first-principles accuracy *and* mapping of structural and functional properties



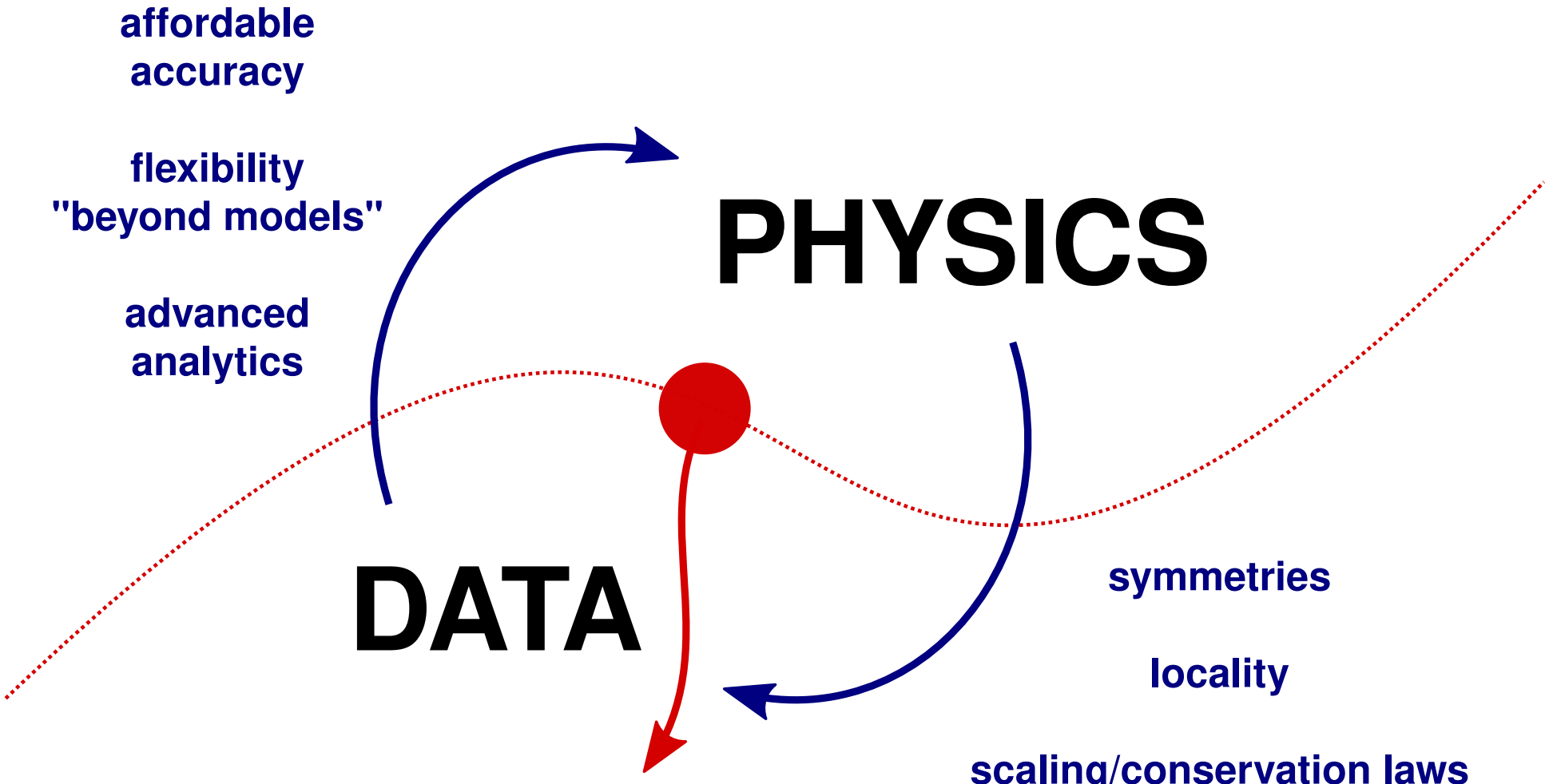
Gigli et al., arxiv: 2111.05129

A software stack for atomistic machine learning

- Integrating ML and atomistic simulations: from representations to models to advanced MD
- Interoperability and data sharing with the rest of the ecosystem



https://github.com/lab-cosmo/



Slides → tinyurl.com/ceriotti-2022-caltech
Review → Musil et al. ChemRev (2021)
Code: → github.com/lab-cosmo

A Dirac notation for ML

features index

$$\langle Q | A \rangle$$

representation target & nature

$$\langle n_1 l_1; \dots n_\nu l_\nu k_\nu | A; \rho_i^{\otimes\nu}; \sigma; \lambda\mu \rangle$$

radial indices

angular channels

structure

correlation order

parity

field center

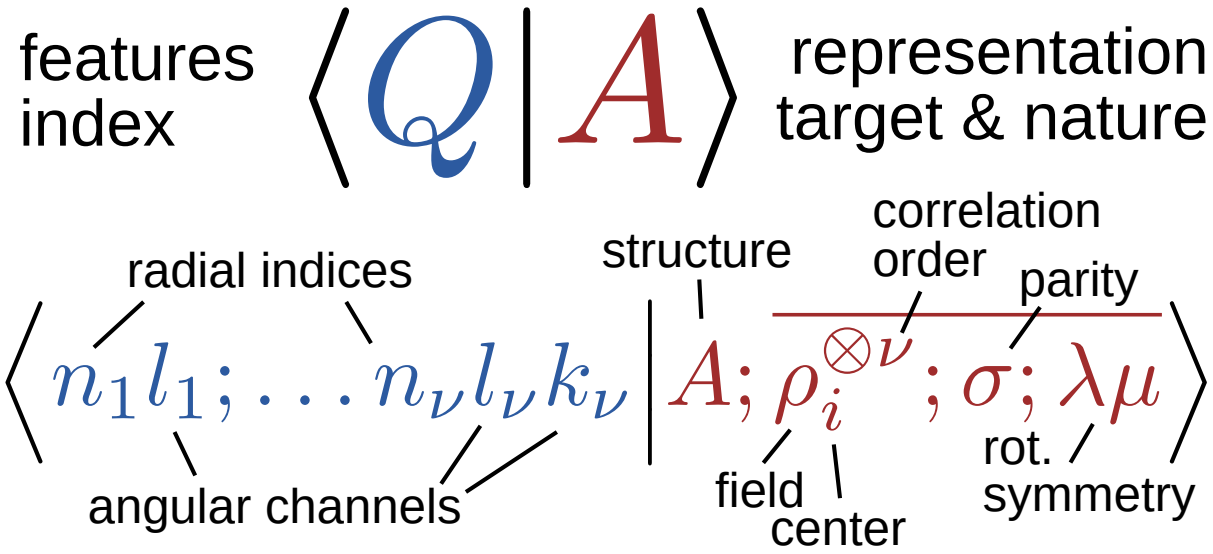
rot. symmetry

- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index Q
- Bra-ket notation $\langle Q|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally a change of basis, the construction of a kernel, or a linear model

$$\langle Y|A\rangle = \int \mathrm{d}Q \, \langle Y|Q\rangle \, \langle Q|A\rangle$$

Willatt, Musil, **MC**, JCP (2019); `https://tinyurl.com/dirac-rep`

A Dirac notation for ML



- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index Q
- Bra-ket notation $\langle Q|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally a change of basis, the construction of a kernel, or a linear model

$$k(A,A') = \langle A|A'\rangle \approx \int \mathrm{d}Q\, \langle A|Q\rangle\, \langle Q|A'\rangle$$

Willatt, Musil, **MC**, JCP (2019); `https://tinyurl.com/dirac-rep`

A Dirac notation for ML

features index $\langle Q | A \rangle$ representation target & nature

$\langle n_1 l_1; \ldots n_\nu l_\nu k_\nu | A; \rho_i^{\otimes\nu}; \sigma; \lambda\mu \rangle$

radial indices; angular channels; structure; correlation order; parity; field center; rot. symmetry

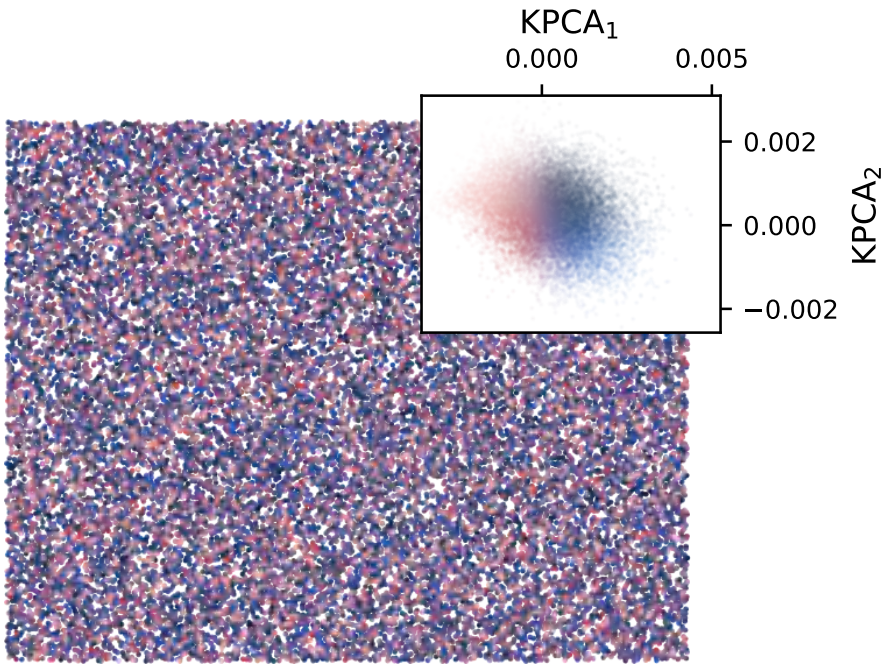
- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index Q
- Bra-ket notation $\langle Q|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally a change of basis, the construction of a kernel, or a linear model

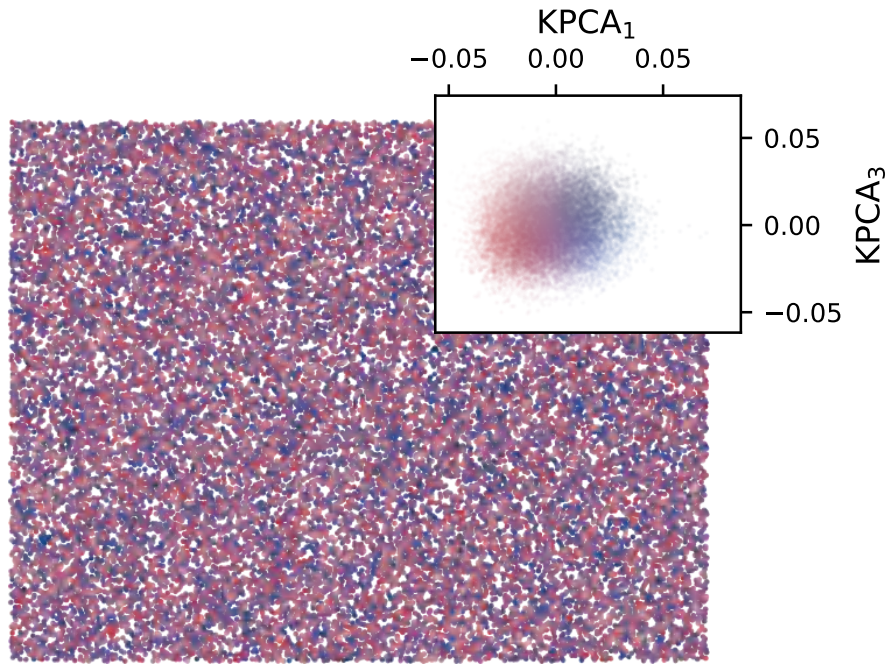
$$E(A) = \langle E|A\rangle \approx \int \mathrm{d}Q \, \langle E|Q\rangle \, \langle Q|A\rangle$$

Willatt, Musil, **MC**, JCP (2019); `https://tinyurl.com/dirac-rep`

What you ask is what you get

- Understanding what goes into a representation is key to achieve meaningful results from automated data analytics
- Example: you don't *always* want to have rotational invariance

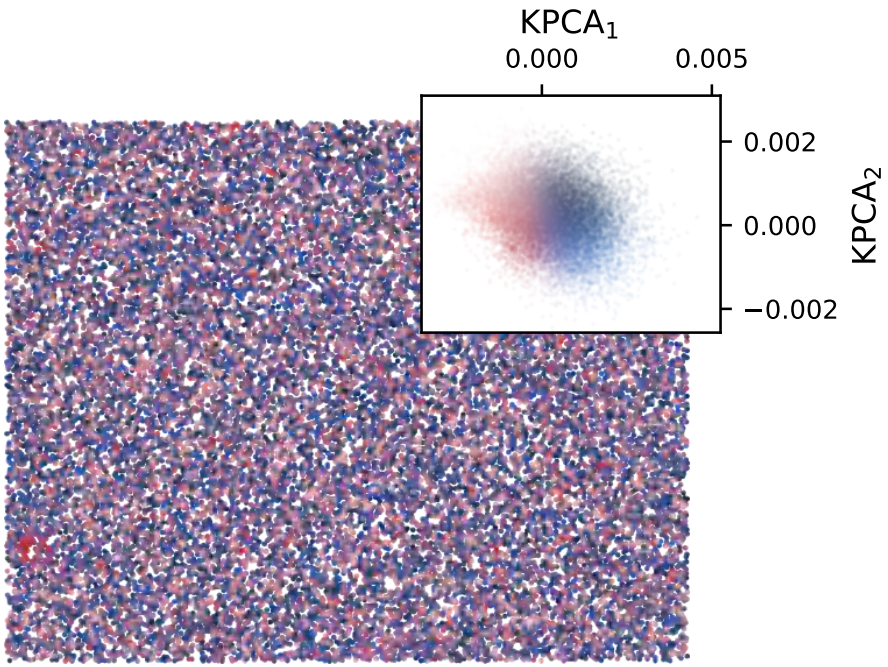


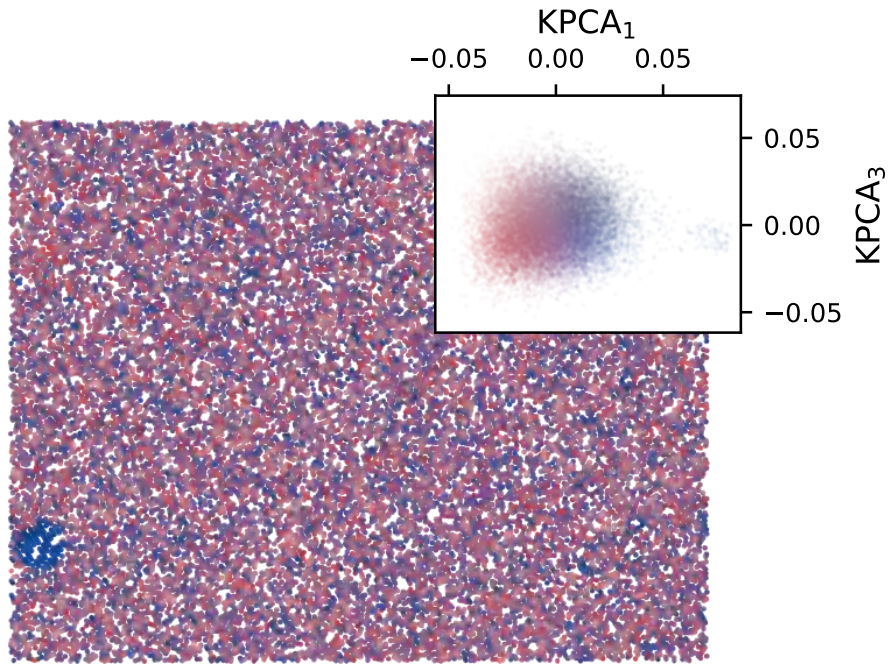


data: Shibuta, Sakane, Takaki, Ohno, Acta Mat. (2016)

What you ask is what you get

- Understanding what goes into a representation is key to achieve meaningful results from automated data analytics
- Example: you don't *always* want to have rotational invariance

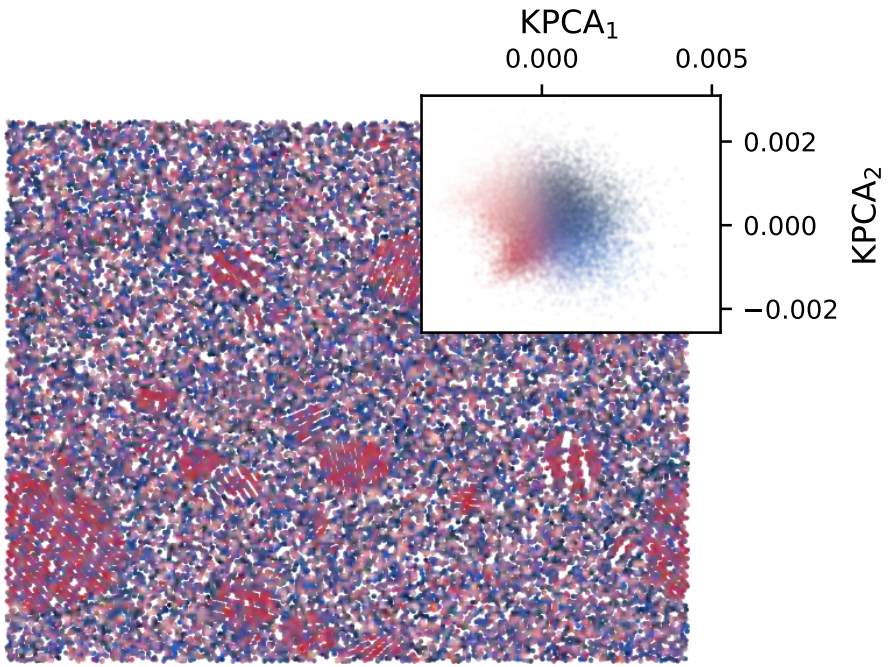


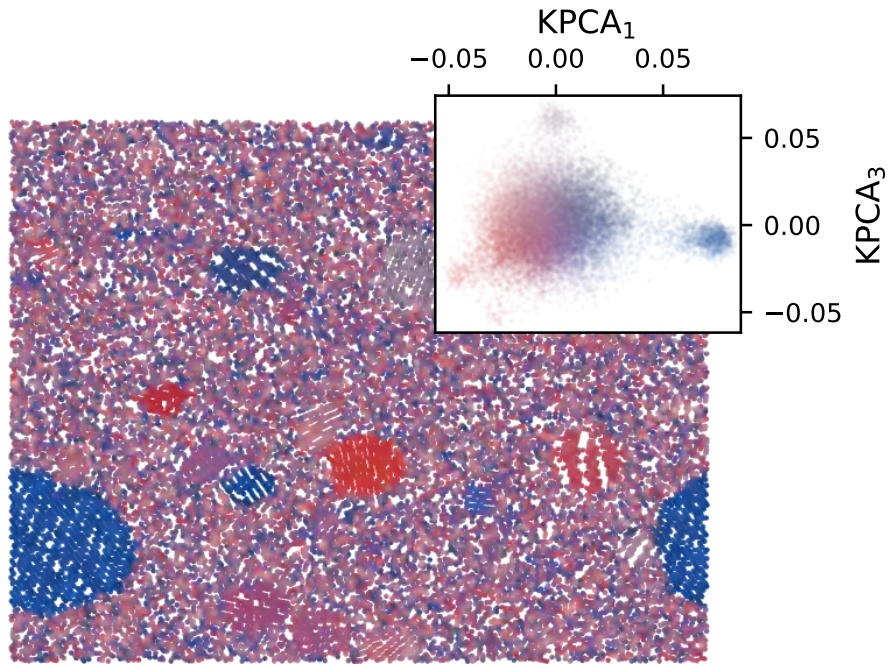


data: Shibuta, Sakane, Takaki, Ohno, Acta Mat. (2016)

What you ask is what you get

- Understanding what goes into a representation is key to achieve meaningful results from automated data analytics
- Example: you don't *always* want to have rotational invariance

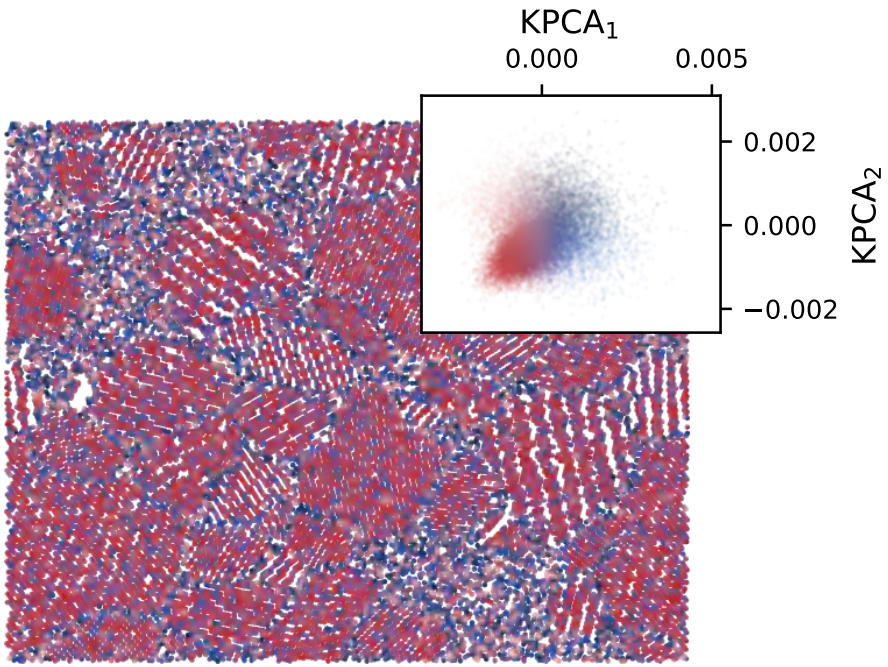


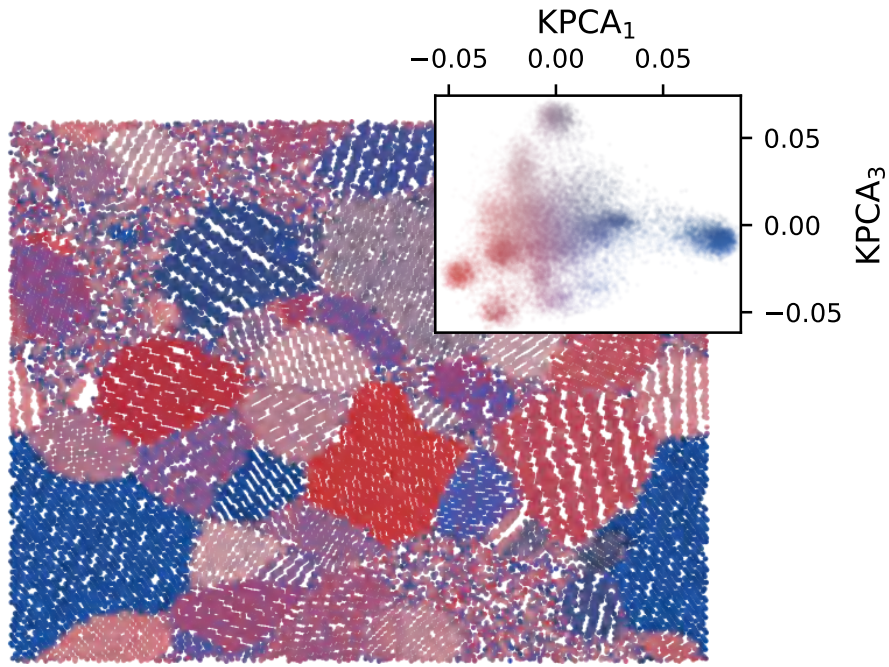


data: Shibuta, Sakane, Takaki, Ohno, Acta Mat. (2016)

What you ask is what you get

- Understanding what goes into a representation is key to achieve meaningful results from automated data analytics
- Example: you don't *always* want to have rotational invariance

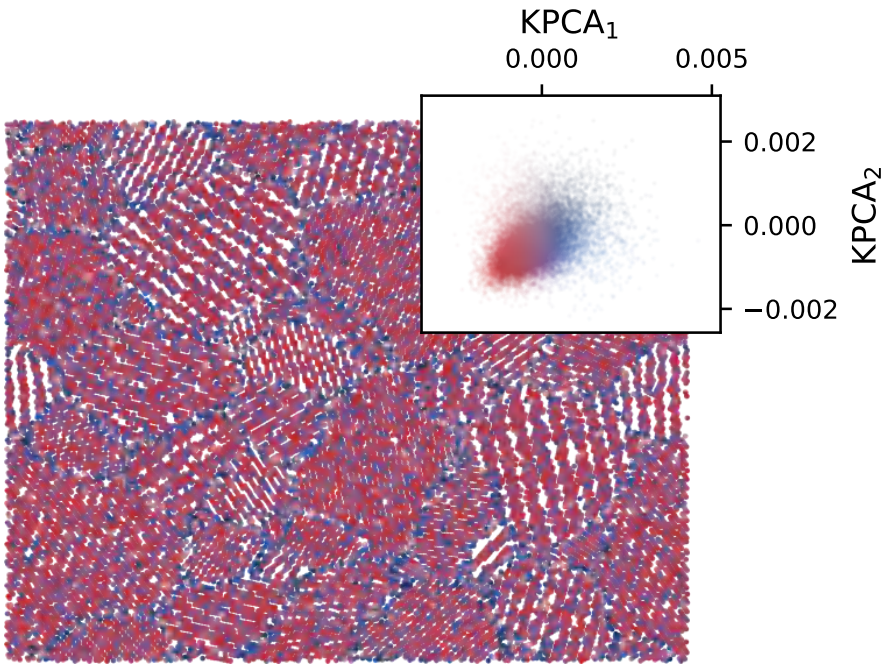


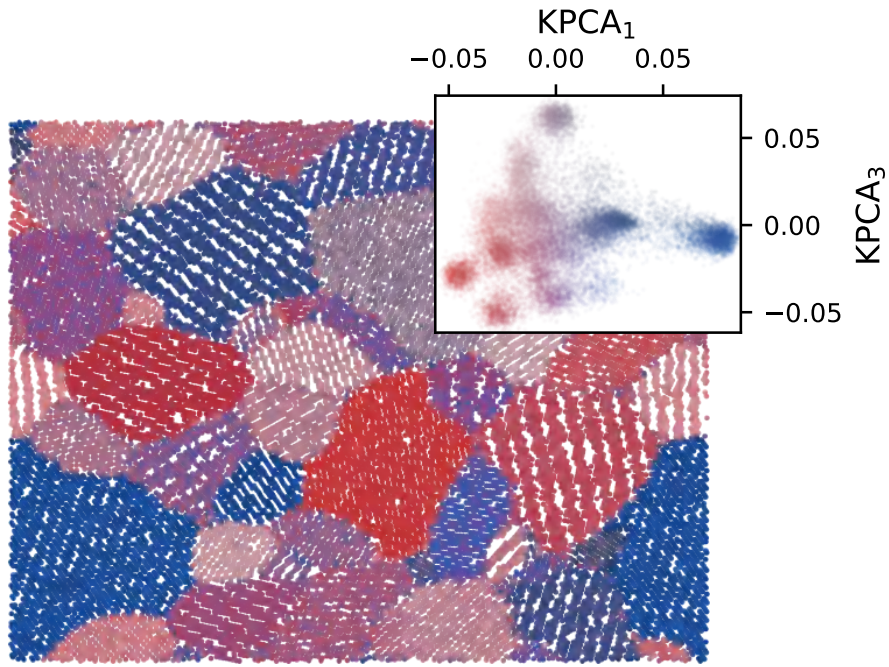


data: Shibuta, Sakane, Takaki, Ohno, Acta Mat. (2016)

What you ask is what you get

- Understanding what goes into a representation is key to achieve meaningful results from automated data analytics
- Example: you don't *always* want to have rotational invariance



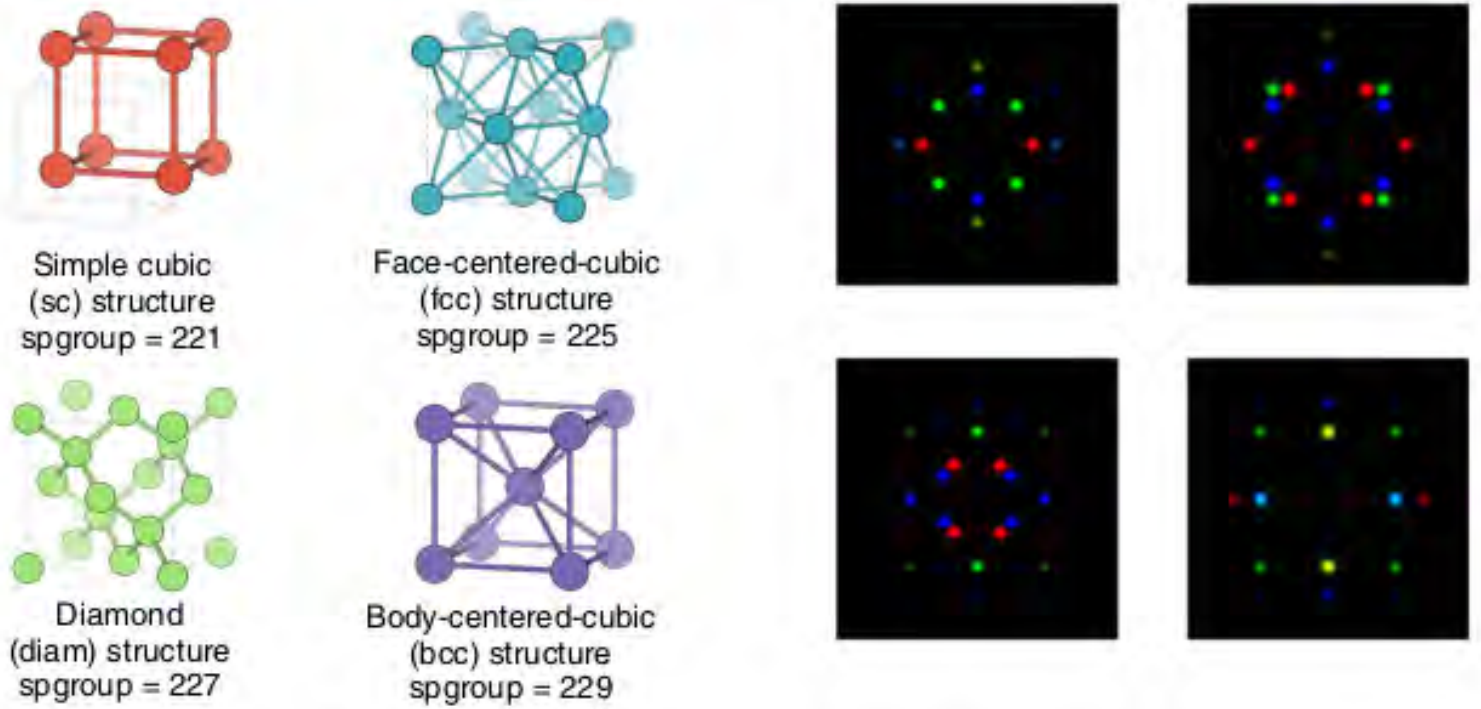


data: Shibuta, Sakane, Takaki, Ohno, Acta Mat. (2016)

Variations on a theme

- Most of the existing density-based representations and kernels emerge as special cases of this framework
  - Basis set choice - e.g. plane waves basis for $|\overline{\rho_i^{\otimes 2}}\rangle$ (Ziletti et al. N.Comm 2018)
  - Projection on symmetry functions (Behler-Parrinello, DeepMD)

$$\langle \mathbf{k}|A;\overline{\rho^{\otimes 2}}\rangle = \sum_{ij\in A} e^{i\mathbf{k}\cdot\mathbf{r}_{ij}}$$

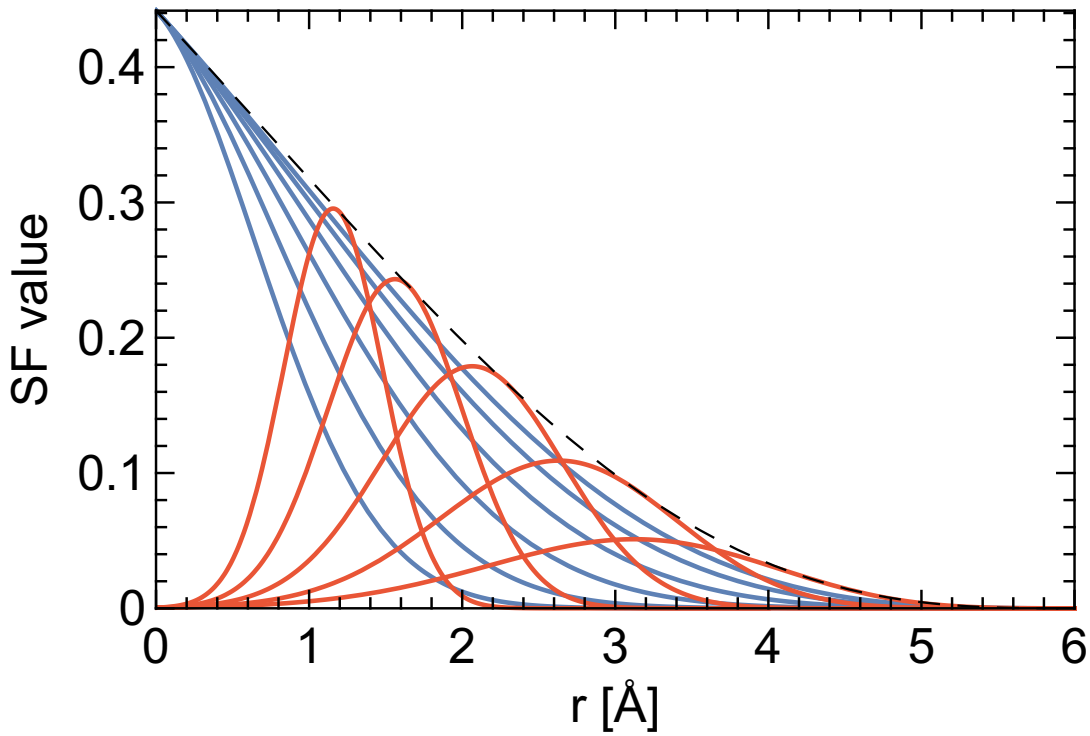


Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

Variations on a theme

- Most of the existing density-based representations and kernels emerge as special cases of this framework
  - Basis set choice - e.g. plane waves basis for $|\overline{\rho_i^{\otimes 2}}\rangle$ (Ziletti et al. N.Comm 2018)
  - Projection on symmetry functions (Behler-Parrinello, DeepMD)

$$\langle abG_2|\overline{\rho_i^{\otimes 1}}\rangle = \delta_{aa_i} \int \mathrm{d}r\, G_2(r)\, \langle br|\overline{\rho_i^{\otimes 1}}; g \to \delta\rangle$$

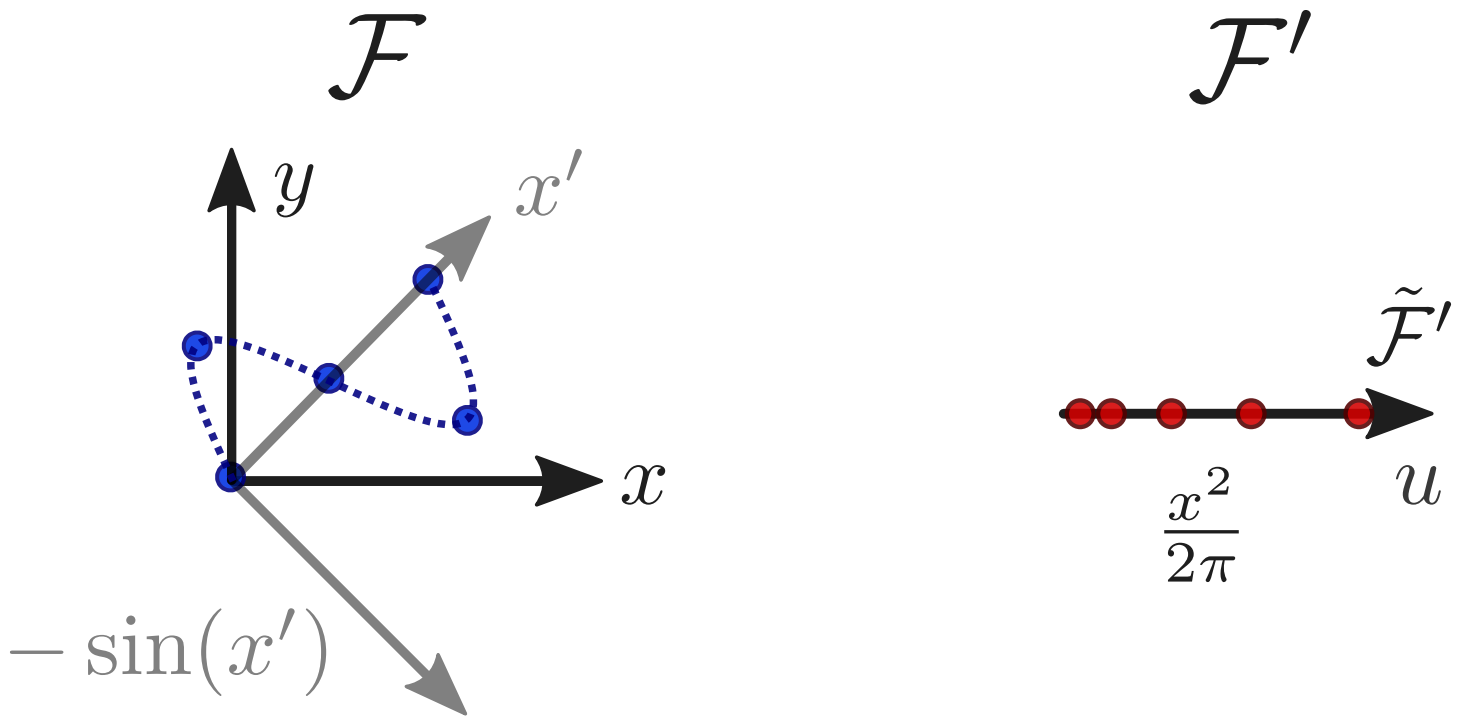


Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

Measuring feature spaces

- Quantitative comparison of relative information content of different features, metrics & kernels
- Feature space Reconstruction Error (**FRE**): linearly-embeddable mutual information

$$\mathrm{GFRE}(\mathcal{F} \to \mathcal{F}') = \min_{\mathbf{P} \in \mathbb{R}^{n_{\mathcal{F}} \times n_{\mathcal{F}'}}} \|\mathbf{X}_{\mathcal{F}'} - \mathbf{X}_{\mathcal{F}}\mathbf{P}\|$$

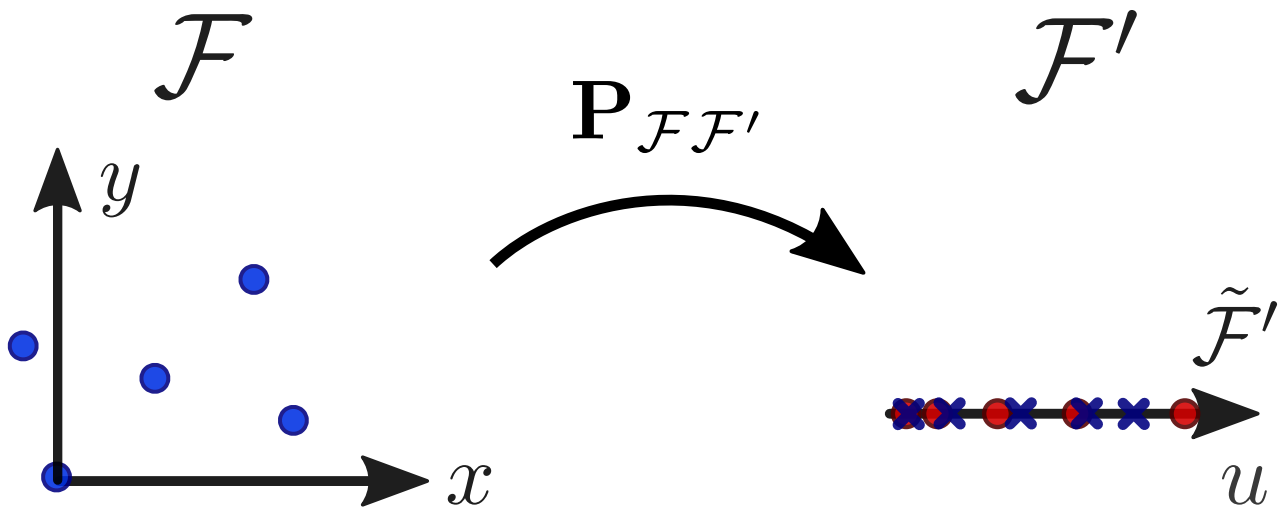


Goscinski, Fraux, **MC**, MLST (2021)

Measuring feature spaces

- Quantitative comparison of relative information content of different features, metrics & kernels
- Feature space Reconstruction Error (**FRE**): linearly-embeddable mutual information

$$\text{GFRE}(\mathcal{F} \to \mathcal{F}') = \min_{\mathbf{P} \in \mathbb{R}^{n_{\mathcal{F}} \times n_{\mathcal{F}'}}} \|\mathbf{X}_{\mathcal{F}'} - \mathbf{X}_{\mathcal{F}}\mathbf{P}\|$$



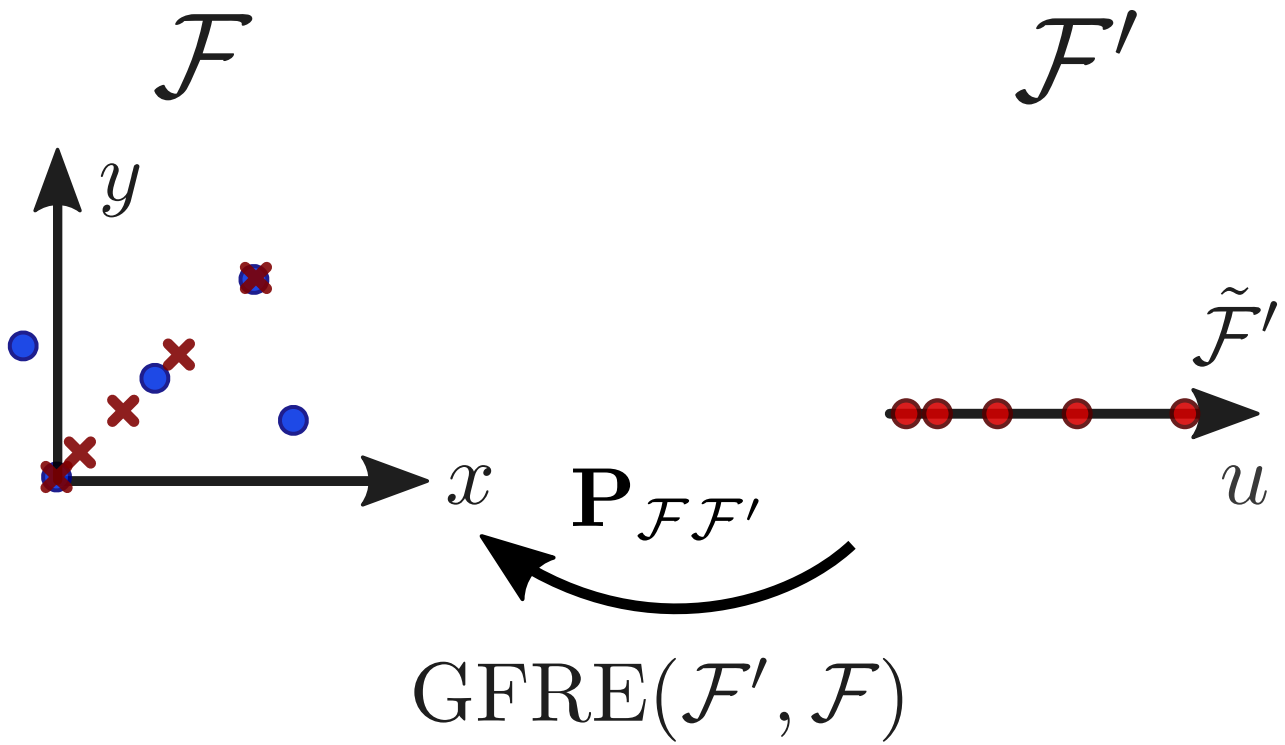
$$\text{GFRE}(\mathcal{F}, \mathcal{F}')$$

Goscinski, Fraux, **MC**, MLST (2021)

Measuring feature spaces

- Quantitative comparison of relative information content of different features, metrics & kernels
- Feature space Reconstruction Error (**FRE**): linearly-embeddable mutual information

$$\mathrm{GFRE}(\mathcal{F} \to \mathcal{F}') = \min_{\mathbf{P} \in \mathbb{R}^{n_{\mathcal{F}} \times n_{\mathcal{F}'}}} \|\mathbf{X}_{\mathcal{F}'} - \mathbf{X}_{\mathcal{F}}\mathbf{P}\|$$

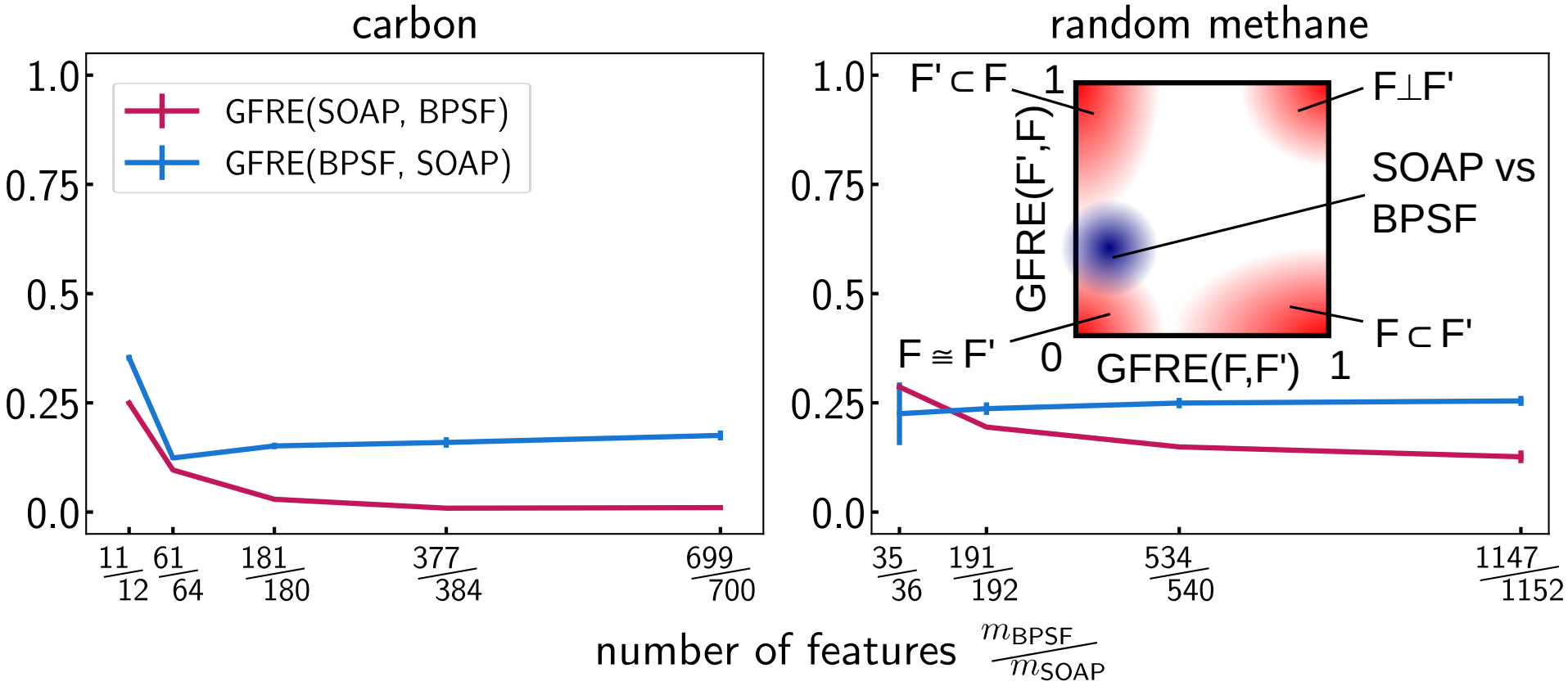


Goscinski, Fraux, **MC**, MLST (2021)

Measuring feature spaces

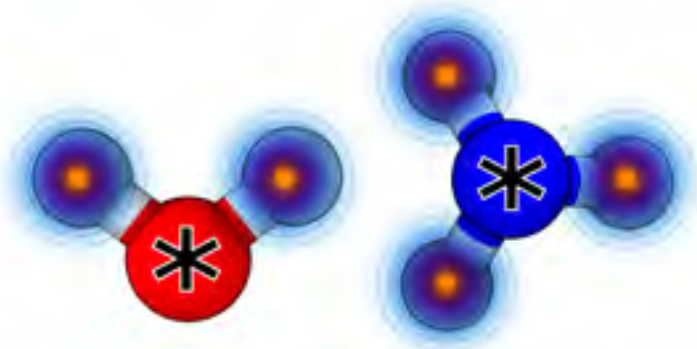
- Quantitative comparison of relative information content of different features, metrics & kernels
- Feature space Reconstruction Error (**FRE**): linearly-embeddable mutual information

$$\mathrm{GFRE}(\mathcal{F} \to \mathcal{F}') = \min_{\mathbf{P} \in \mathbb{R}^{n_{\mathcal{F}} \times n_{\mathcal{F}'}}} \|\mathbf{X}_{\mathcal{F}'} - \mathbf{X}_{\mathcal{F}}\mathbf{P}\|$$

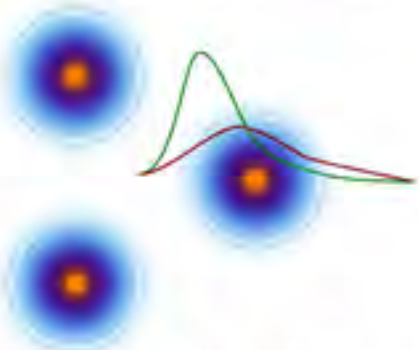


Density expansion and SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product $\rightarrow$ SOAP power spectrum!
- Easily generalized to higher body order.
  δ-distribution limit $\rightarrow$ atomic cluster expansion



$$\langle \mathbf{r}|\rho_i\rangle = \sum_i g(\mathbf{r}-\mathbf{r}_{ij})$$

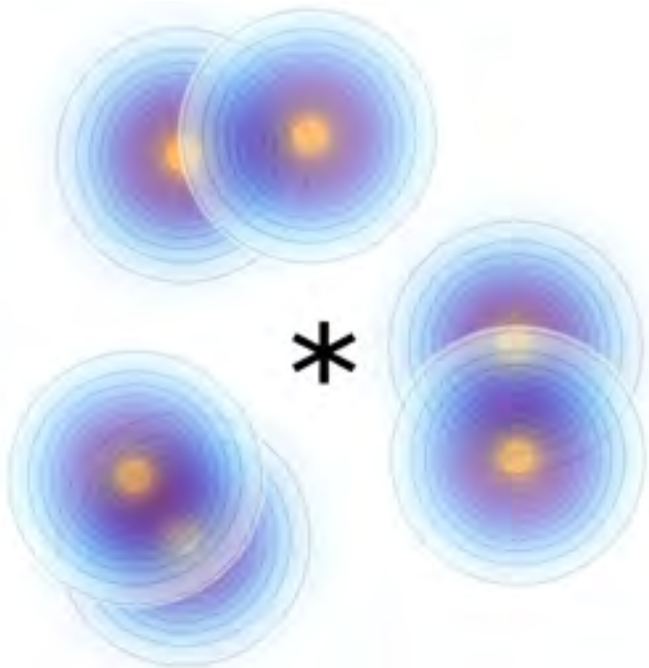


$$\langle nlm|\rho_i\rangle = \int \mathrm{d}\mathbf{r}\, \langle \mathbf{r}|\rho_i\rangle R_n(r) Y_m^l(\hat{\mathbf{r}})$$

Bartók, Kondor, Csányi, PRB (2013); Willatt, Musil, **MC**, JCP (2019); Drautz, PRB (2019)

Density expansion and SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product $\rightarrow$ SOAP power spectrum!
- Easily generalized to higher body order.
  δ-distribution limit $\rightarrow$ atomic cluster expansion

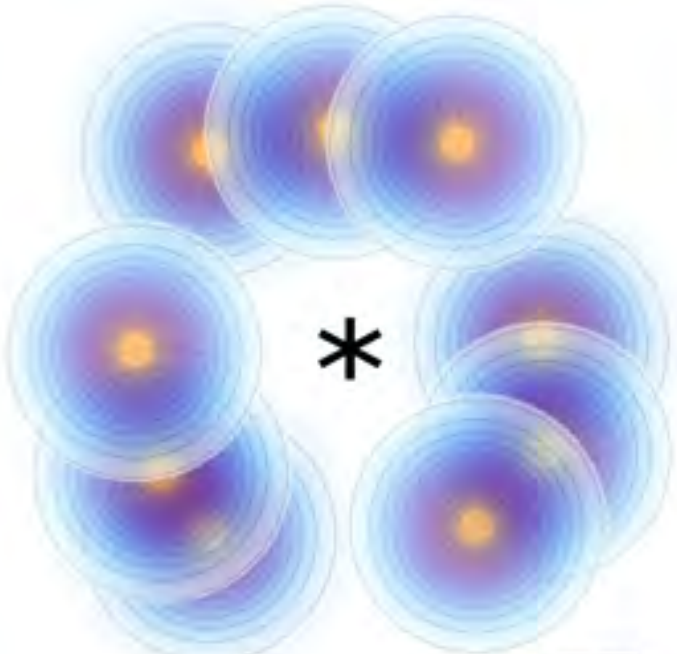


$$\langle nn'l|\overline{\rho_i^{\otimes 2}}\rangle = \sum_m \langle nlm|\rho_i\rangle^\star \langle n'lm|\rho_i\rangle$$
$$p_{nn'l} = \sum_m c^\star_{nlm} c_{n'lm}$$

Bartók, Kondor, Csányi, PRB (2013); Willatt, Musil, **MC**, JCP (2019); Drautz, PRB (2019)

Density expansion and SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product $\rightarrow$ SOAP power spectrum!
- Easily generalized to higher body order.
  δ-distribution limit $\rightarrow$ atomic cluster expansion

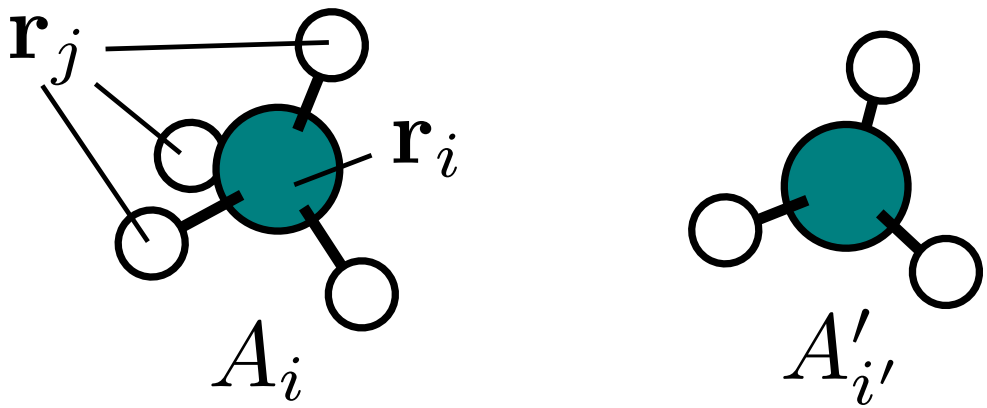


$$\langle n_1 l_1 m_1; n_2 l_2 m_2; \dots n_\nu l_\nu m_\nu | \overline{\rho_i^{\otimes \nu}} \rangle =$$
$$\int \mathrm{d}\hat{R} \langle n_1 l_1 m_1 | \hat{R} | \rho_i \rangle \cdots \langle n_\nu l_\nu m_\nu | \hat{R} | \rho_i \rangle =$$
$$\sum_{m'_1 \dots m'_\nu} \langle n_1 l_1 m'_1 | \rho_i \rangle \cdots \langle n_\nu l_\nu m'_\nu | \rho_i \rangle$$
$$\times \int \mathrm{d}\hat{R}\, D^{l_1}_{m_1 m'_1}(\hat{R}) \cdots D^{l_\nu}_{m_\nu m'_\nu}(\hat{R})$$

Bartók, Kondor, Csányi, PRB (2013); Willatt, Musil, **MC**, JCP (2019); Drautz, PRB (2019)

Two-neighbors descriptors

- Construction of a three-body ($\nu = 2$) invariant atomic descriptor
  1. Define relative position of neighbors (*translation-invariant*)
  2. Positions are transformed in a neighbor density (*permutation invariant*)
  3. Symmetrize over rotations a tensor product of the neighbor densities
  4. This is equivalent to a function of two distances and one angle
  5. $g \to \delta$ limit $\Rightarrow$ list of 2-neighbors tuples $(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$
  6. Linear model $\Rightarrow$ 3-body potential!

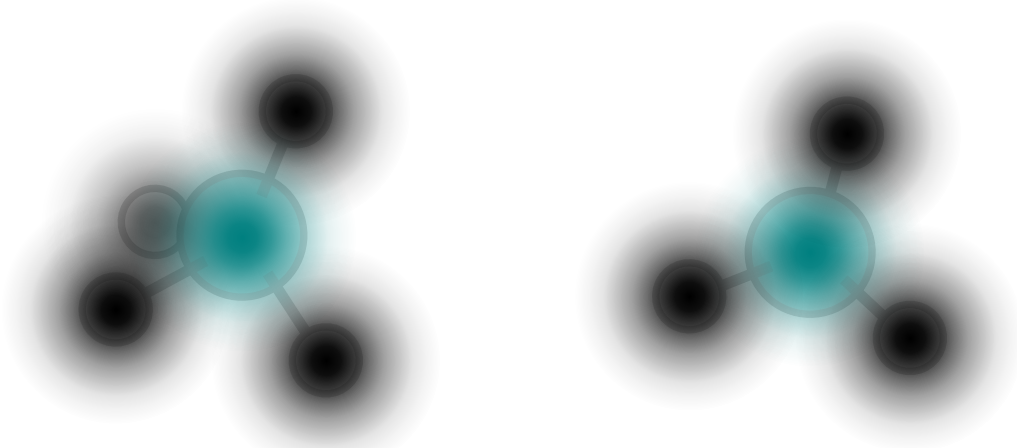


$$\{\mathbf{r}_{ji} = \mathbf{r}_j - \mathbf{r}_i\} \leftrightarrow A_i$$

Bartók, Kondor, Csányi, PRB (2013)

Two-neighbors descriptors

- Construction of a three-body ($\nu = 2$) invariant atomic descriptor
  1. Define relative position of neighbors (*translation-invariant*)
  2. Positions are transformed in a neighbor density (*permutation invariant*)
  3. Symmetrize over rotations a tensor product of the neighbor densities
  4. This is equivalent to a function of two distances and one angle
  5. $g \to \delta$ limit $\Rightarrow$ list of 2-neighbors tuples $(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$
  6. Linear model $\Rightarrow$ 3-body potential!

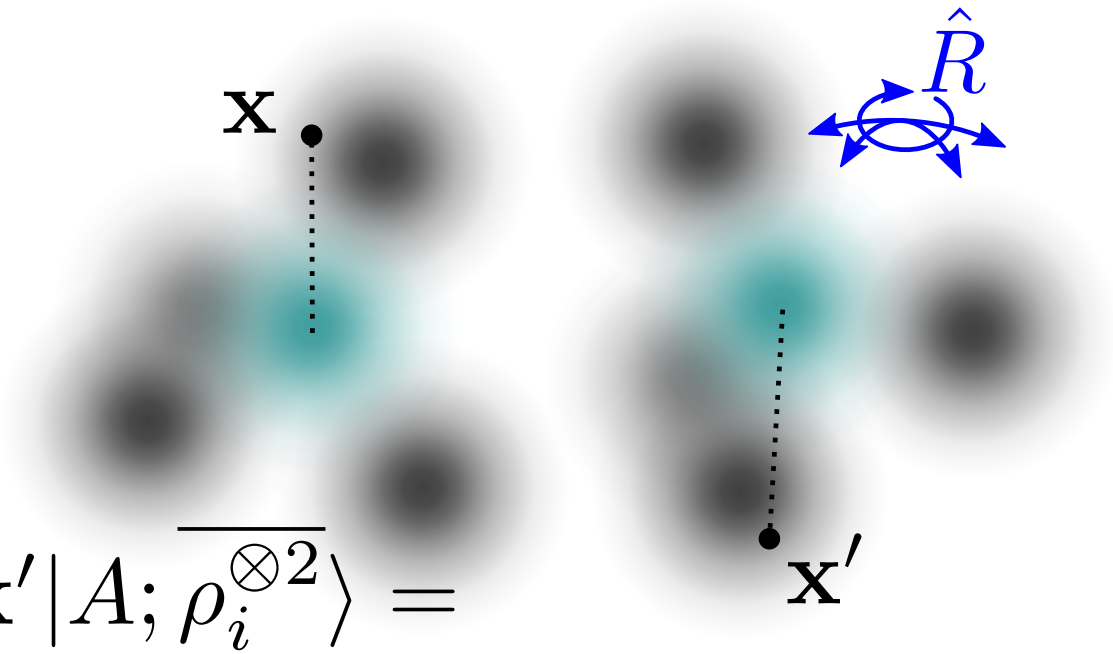


$$\langle a\mathbf{x}|\rho_i\rangle = \sum_{j\in A_i} \delta_{aa_j} \langle \mathbf{x}|\mathbf{r}_{ji}; g\rangle$$
$$\langle \mathbf{x}|\mathbf{r}_{ji}; g\rangle \equiv g(\mathbf{x} - \mathbf{r}_{ji})$$

Bartók, Kondor, Csányi, PRB (2013)

Two-neighbors descriptors

- Construction of a three-body ($\nu = 2$) invariant atomic descriptor
  1. Define relative position of neighbors (*translation-invariant*)
  2. Positions are transformed in a neighbor density (*permutation invariant*)
  3. Symmetrize over rotations a tensor product of the neighbor densities
  4. This is equivalent to a function of two distances and one angle
  5. $g \to \delta$ limit $\Rightarrow$ list of 2-neighbors tuples $(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$
  6. Linear model $\Rightarrow$ 3-body potential!

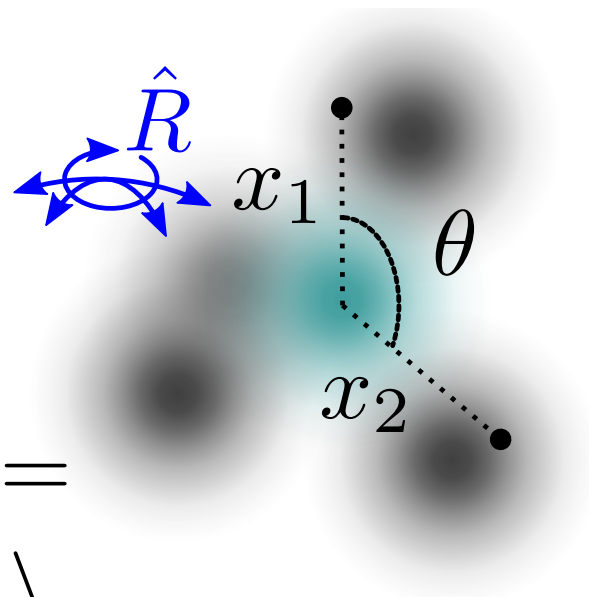


$$\langle \mathbf{x}; \mathbf{x}' | A; \overline{\rho_i^{\otimes 2}} \rangle = \int \mathrm{d}\hat{R} \langle \mathbf{x} | \hat{R}A; \rho_i \rangle \langle \mathbf{x}' | \hat{R}A; \rho_i \rangle$$

Willatt, Musil, **MC**, JCP (2019)

Two-neighbors descriptors

- Construction of a three-body ($\nu = 2$) invariant atomic descriptor
  1. Define relative position of neighbors (*translation-invariant*)
  2. Positions are transformed in a neighbor density (*permutation invariant*)
  3. Symmetrize over rotations a tensor product of the neighbor densities
  4. This is equivalent to a function of two distances and one angle
  5. $g \to \delta$ limit $\Rightarrow$ list of 2-neighbors tuples $(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$
  6. Linear model $\Rightarrow$ 3-body potential!



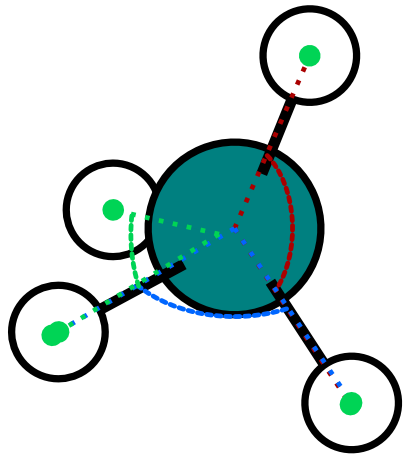
$$\langle x_1; x_2; \theta | A; \overline{\rho_i^{\otimes 2}} \rangle =$$
$$\int \mathrm{d}\hat{R} \langle x_1 \hat{R} \hat{\mathbf{e}}_z | A; \rho_i \rangle$$
$$\langle x_2 \hat{R} (\hat{\mathbf{e}}_z \cos\theta + \hat{\mathbf{e}}_x \sin\theta) | A; \rho_i \rangle$$

Willatt, Musil, **MC**, JCP (2019)

Two-neighbors descriptors

- Construction of a three-body ($\nu = 2$) invariant atomic descriptor
  1. Define relative position of neighbors (*translation-invariant*)
  2. Positions are transformed in a neighbor density (*permutation invariant*)
  3. Symmetrize over rotations a tensor product of the neighbor densities
  4. This is equivalent to a function of two distances and one angle
  5. $g \to \delta$ limit $\Rightarrow$ list of 2-neighbors tuples $(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$
  6. Linear model $\Rightarrow$ 3-body potential!

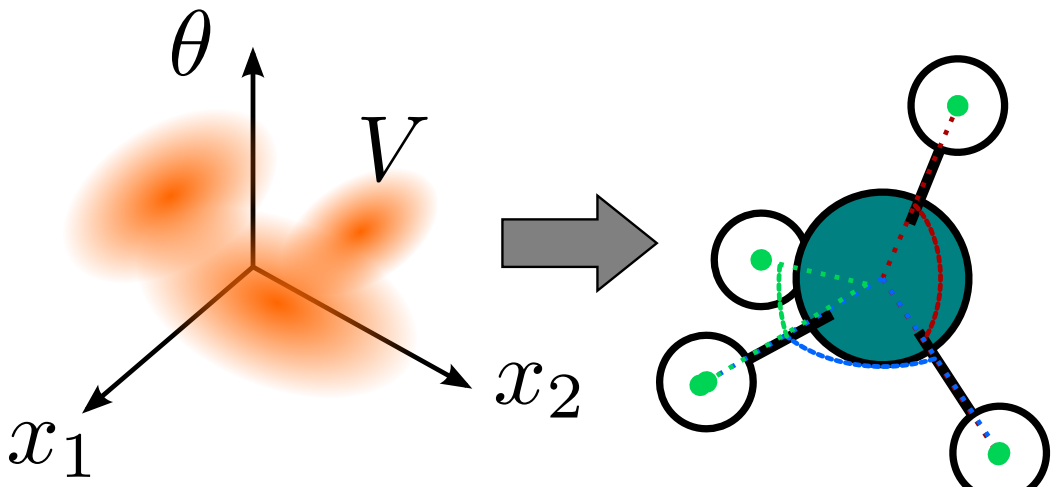
$$\langle x_1; x_2; \theta | A; \overline{\delta_i^{\otimes 2}} \rangle = \sum_{j_1 j_2} \delta(x_1 - r_{j_1 i}) \delta(x_2 - r_{j_2 i}) \delta(\cos\theta - \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$$



Willatt, Musil, **MC**, JCP (2019)

Two-neighbors descriptors

- Construction of a three-body ($\nu = 2$) invariant atomic descriptor
  1. Define relative position of neighbors (*translation-invariant*)
  2. Positions are transformed in a neighbor density (*permutation invariant*)
  3. Symmetrize over rotations a tensor product of the neighbor densities
  4. This is equivalent to a function of two distances and one angle
  5. $g \to \delta$ limit $\Rightarrow$ list of 2-neighbors tuples $(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$
  6. Linear model $\Rightarrow$ 3-body potential!

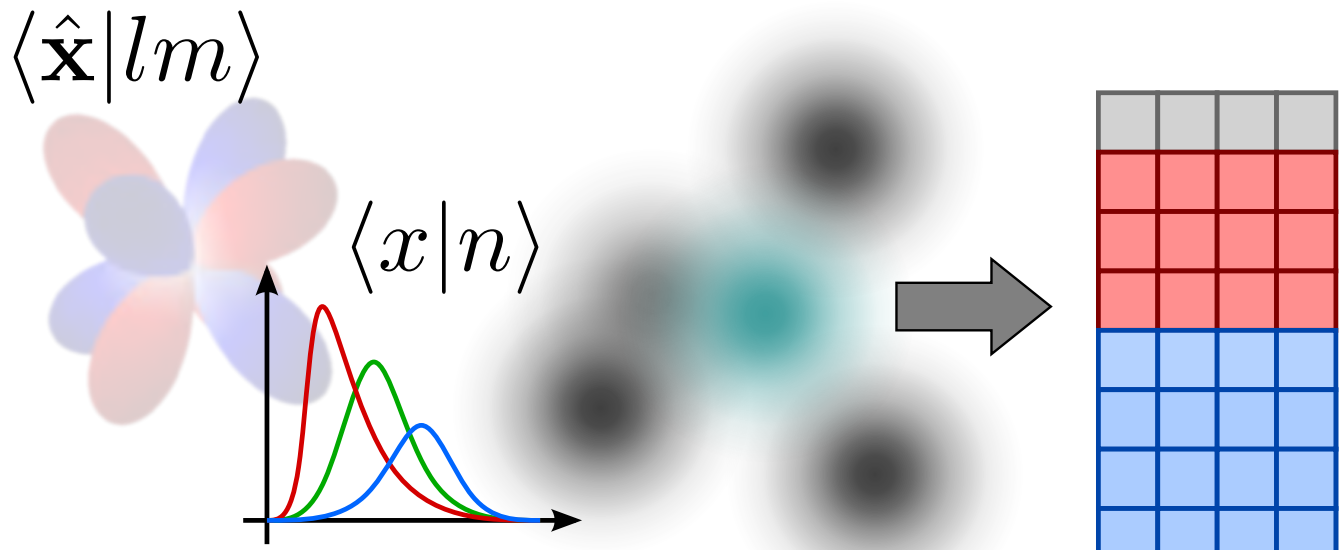


$$\int \langle V|x_1;x_2;\theta\rangle\langle x_1;x_2;\theta|A;\overline{\delta_i^{\otimes 2}}\rangle = \sum_{j_1 j_2} V\left(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i}\cdot\hat{\mathbf{r}}_{j_2 i}\right)$$

Willatt, Musil, **MC**, JCP (2019)

Density trick in an $\langle nlm|$ basis

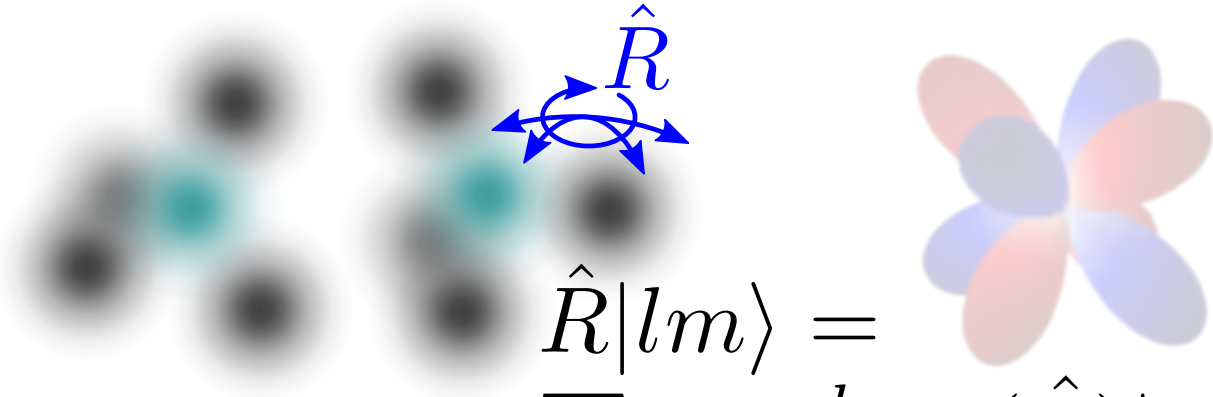
- The symmetrized correlations can be computed in closed form using a discrete basis
  - The neighbor density can be expanded on a basis of radial functions $\langle x|n\rangle \equiv R_n(x)$ and spherical harmonics $\langle \hat{\mathbf{x}}|lm\rangle \equiv Y_l^m(\hat{\mathbf{x}})$
  - Spherical harmonics transform linearly under rotations based on Wigner rotation matrices $\mathbf{D}^l\left(\hat{R}\right)$
  - Orthogonality of Wigner matrices yields the SOAP powerspectrum



$$\langle nlm|A;\rho_i\rangle = \int \mathrm{d}\mathbf{x}\langle n|x\rangle\langle lm|\hat{\mathbf{x}}\rangle\langle \mathbf{x}|A;\rho_i\rangle$$

Density trick in an $\langle nlm|$ basis

- The symmetrized correlations can be computed in closed form using a discrete basis
  - The neighbor density can be expanded on a basis of radial functions $\langle x|n\rangle \equiv R_n(x)$ and spherical harmonics $\langle \hat{\mathbf{x}}|lm\rangle \equiv Y_l^m(\hat{\mathbf{x}})$
  - Spherical harmonics transform linearly under rotations based on Wigner rotation matrices $\mathbf{D}^l\left(\hat{R}\right)$
  - Orthogonality of Wigner matrices yields the SOAP powerspectrum



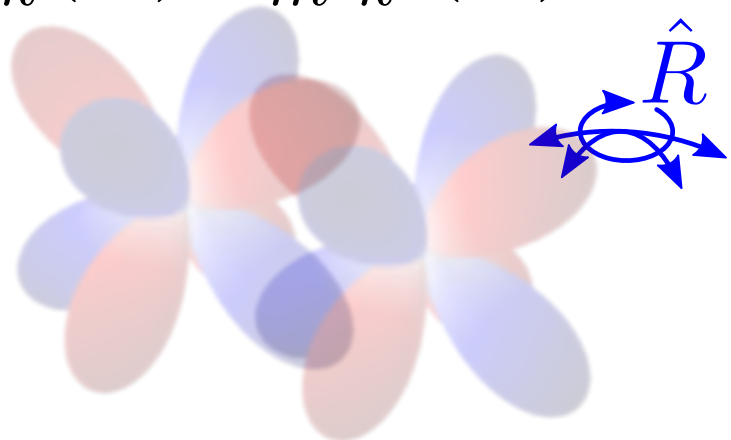
$$\hat{R}|lm\rangle = \sum_{m'} D^l_{mm'}(\hat{R})|lm'\rangle$$

$$\langle nlm; n'l'm'|A; \overline{\rho_i^{\otimes 2}}\rangle = \int \mathrm{d}\hat{R}\langle nlm|\hat{R}A; \rho_i\rangle\langle n'l'm'|\hat{R}A; \rho_i\rangle$$

Density trick in an $\langle nlm|$ basis

- The symmetrized correlations can be computed in closed form using a discrete basis
  - The neighbor density can be expanded on a basis of radial functions $\langle x|n\rangle \equiv R_n(x)$ and spherical harmonics $\langle \hat{\mathbf{x}}|lm\rangle \equiv Y_l^m(\hat{\mathbf{x}})$
  - Spherical harmonics transform linearly under rotations based on Wigner rotation matrices $\mathbf{D}^l\left(\hat{R}\right)$
  - Orthogonality of Wigner matrices yields the SOAP powerspectrum

$$\int \mathrm{d}\hat{R} \sum_{kk'} D^l_{mk}(\hat{R}) D^{l'}_{m'k'}(\hat{R}) \propto \delta_{ll'}\delta_{mm'}\delta_{kk'}$$



$$\langle nn'l|A;\overline{\rho_i^{\otimes 2}}\rangle = \sum_m \langle nlm|A;\rho_i\rangle\langle n'lm|A;\rho_i\rangle$$

A hierarchy of equivariant features

- A generalization of the definition yields N-body features that transform like angular momenta

$$\langle X | \overline{\rho_i^{\otimes \nu}; \sigma; \lambda\mu} \rangle$$

- Recursive construction based on sums of angular momenta and an expansion of the atom density

$$\langle n_1 l_1 k_1 | \overline{\rho_i^{\otimes 1}; \lambda\mu} \rangle \equiv \langle n_1 \lambda\, (-\mu) | \rho_i \rangle\, \delta_{l_1 \lambda} \delta_{k_1 \lambda} \delta_{\sigma 1} \equiv \langle n_1 | \overline{\rho_i^{\otimes 1}; \lambda\mu} \rangle$$

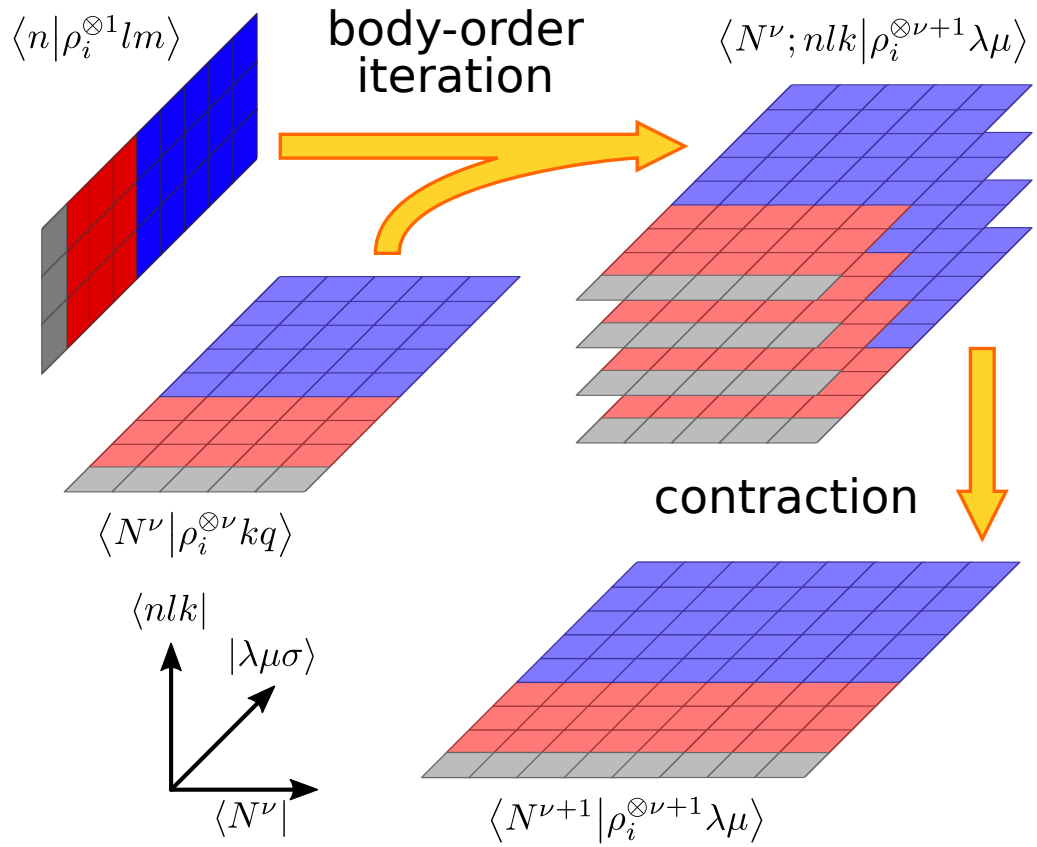
$$\langle \ldots; n_\nu l_\nu k_\nu; nlk | \overline{\rho_i^{\otimes(\nu+1)}; \sigma; \lambda\mu} \rangle = \delta_{\sigma((-1)^{l+k+\lambda} s)} c_{k\lambda} \times \sum_{qm} \langle lm; kq | \lambda\mu \rangle \, \langle n | \overline{\rho_i^{\otimes 1}; lm} \rangle \, \langle \ldots; n_\nu l_\nu k_\nu | \overline{\rho_i^{\otimes \nu}; s; kq} \rangle$$

- Can be used to compute efficiently *invariant* features $|\overline{\rho_i^{\otimes \nu}; 0; 00}\rangle$

Nigam, Pozdnyakov, **MC**, JCP (2020)

NICE features for ML

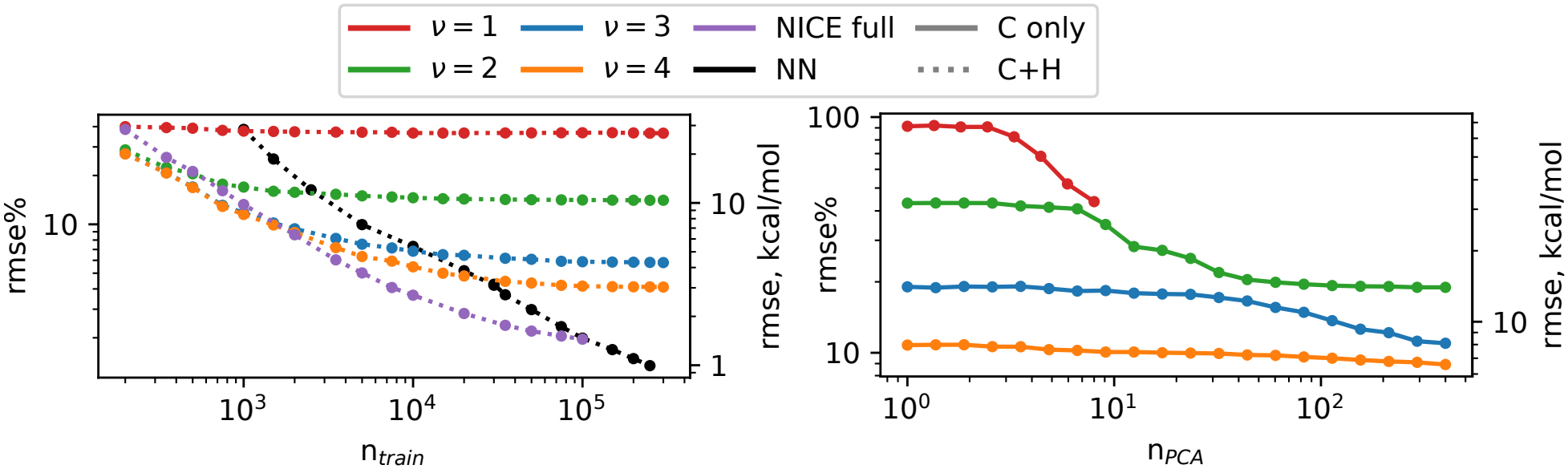
- Problem: number of features grows exponentially with ν
- Solution: an *N*-body iterative contraction of equivariants (NICE) framework
  - After each body order increase, the most relevant features are selected and used for the next iteration
  - Systematic convergence with ν and contraction truncation



Nigam, Pozdnyakov, **MC**, JCP (2020)

NICE features for ML

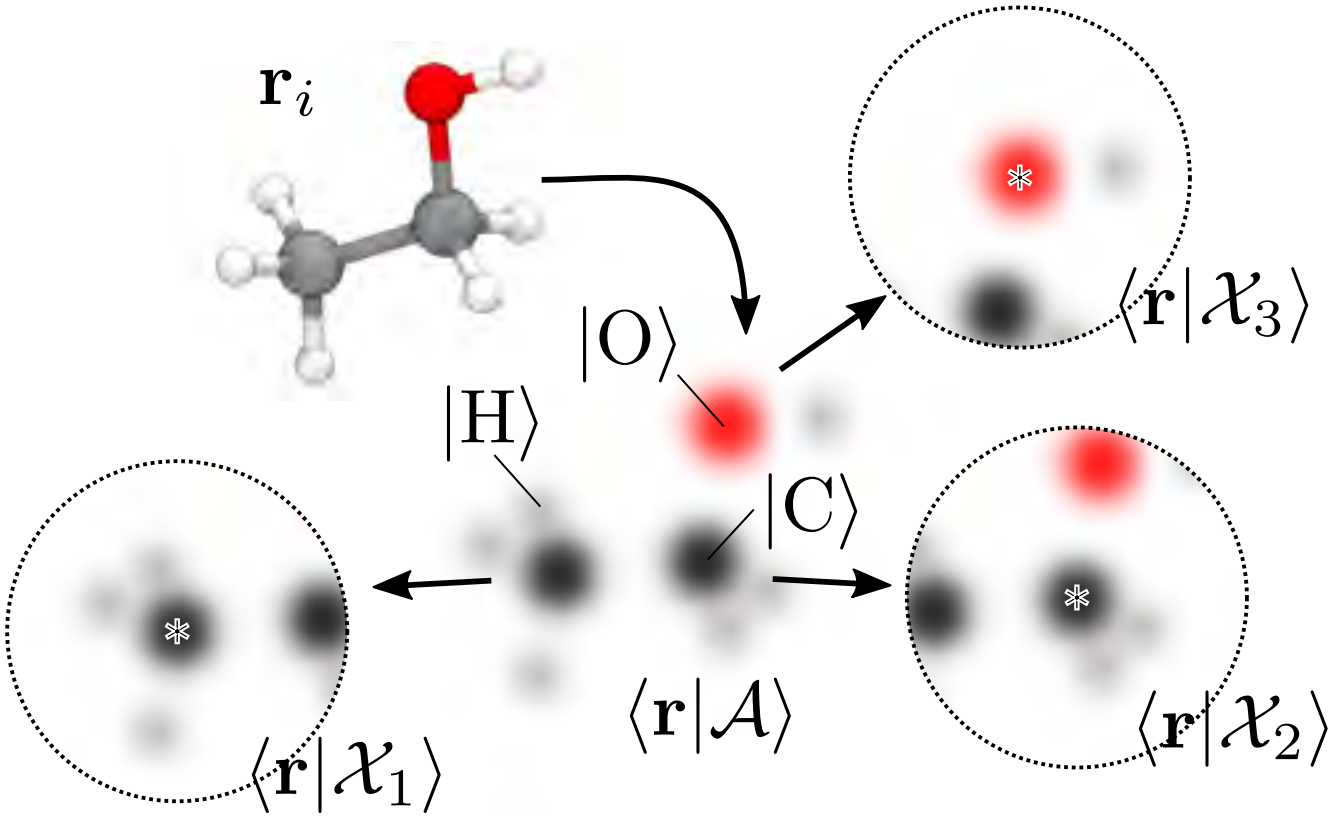
- Problem: number of features grows exponentially with ν
- Solution: an *N*-body iterative contraction of equivariants (NICE) framework
  - After each body order increase, the most relevant features are selected and used for the next iteration
  - Systematic convergence with ν and contraction truncation



Nigam, Pozdnyakov, **MC**, JCP (2020)

Machine-learning the periodic table

- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective learning . . .
- . . . and as by-product get a data-driven version of the periodic table!



Machine-learning the periodic table

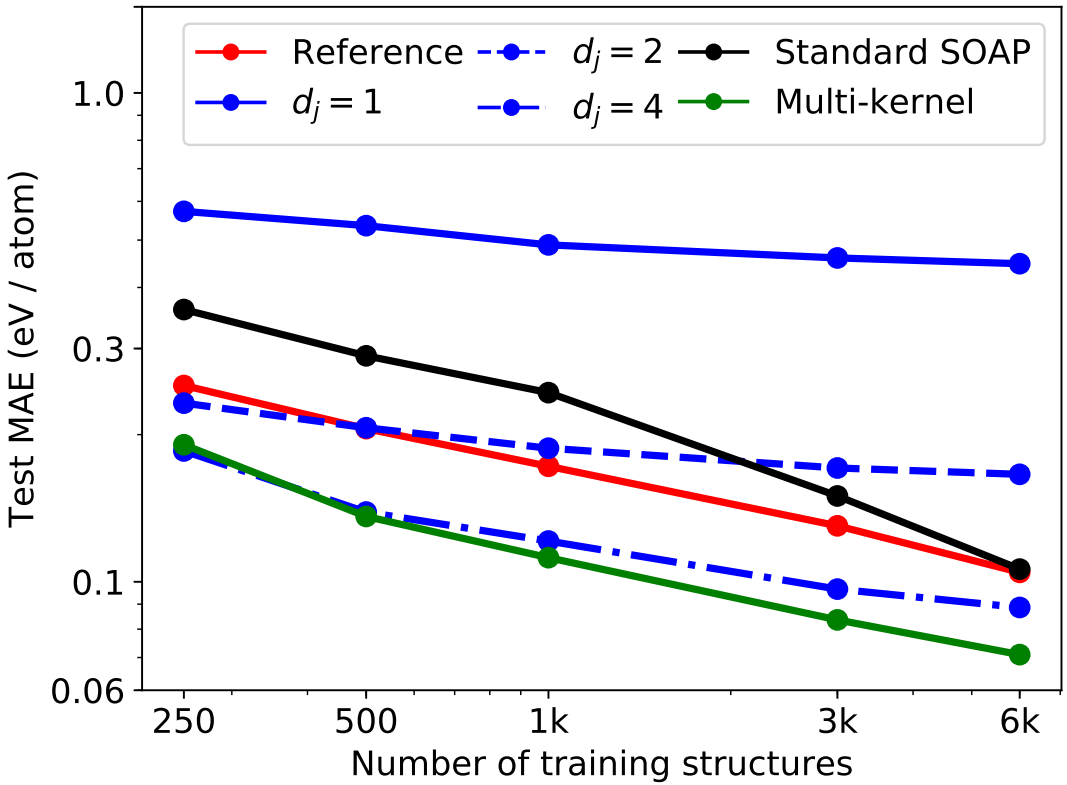
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective learning . . .
- . . . and as by-product get a data-driven version of the periodic table!

$$|\mathrm{H}\rangle = 0.5\,|🔥\rangle + 0.1\,|🪨\rangle + 0.2\,|💧\rangle$$
$$|\mathrm{C}\rangle = 0.2\,|🔥\rangle + 0.8\,|🪨\rangle + 0.3\,|💧\rangle$$
$$|\mathrm{O}\rangle = 0.1\,|🔥\rangle + 0.1\,|🪨\rangle + 0.6\,|💧\rangle$$

Empedocles et al. (ca 360BC). Metaphor courtesy of Albert Bartók

Machine-learning the periodic table

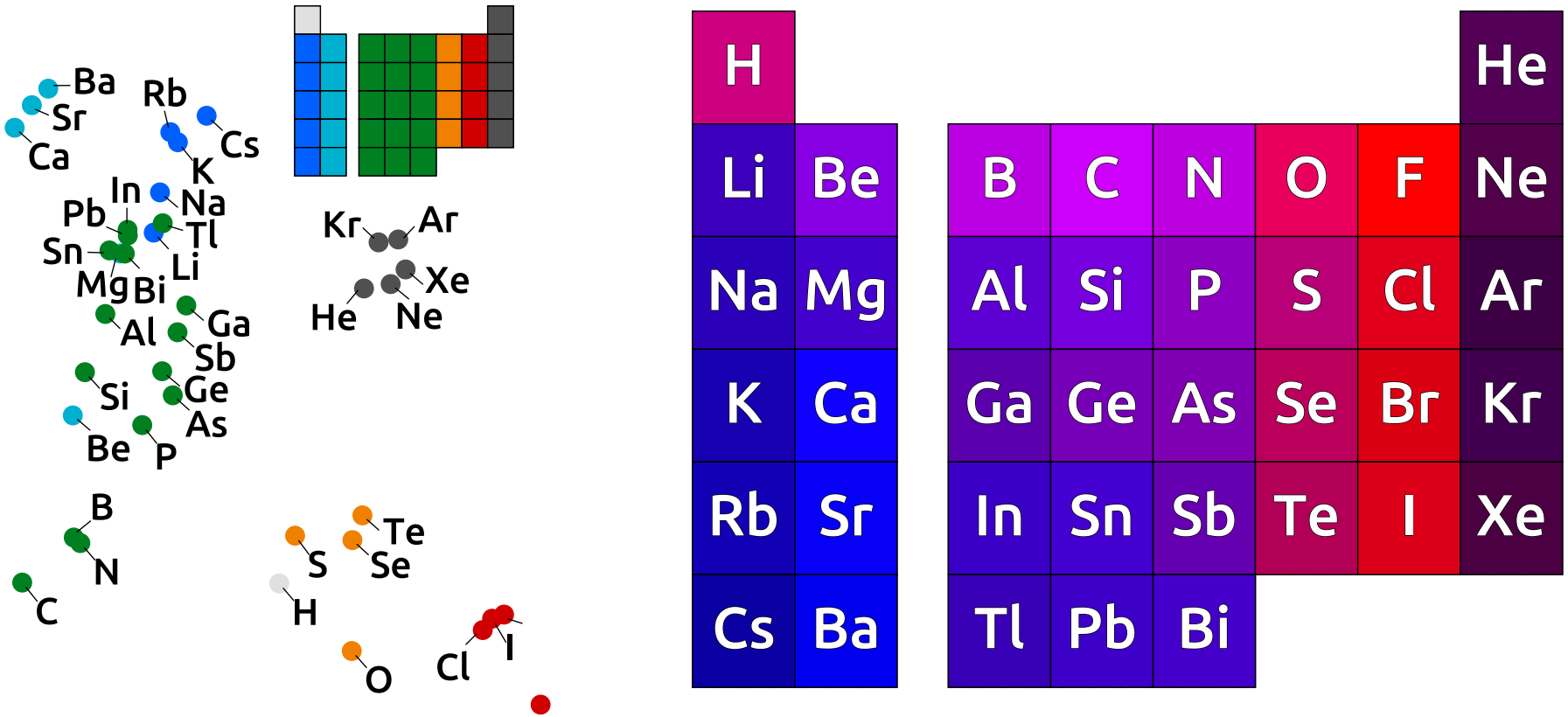
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective learning . . .
- . . . and as by-product get a data-driven version of the periodic table!



Elpasolite dataset. Reference curve (red) from Faber et al. JCP (2018)

Machine-learning the periodic table

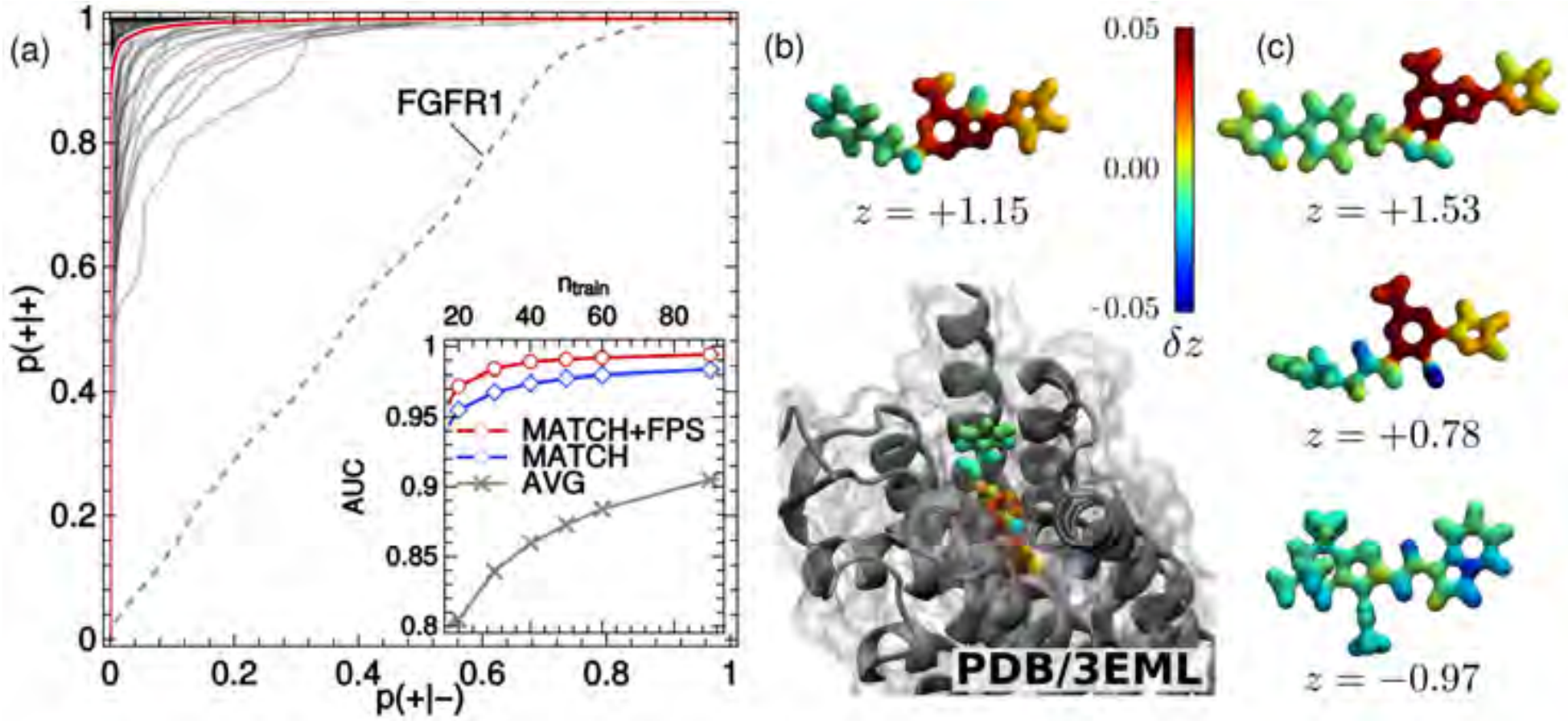
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective learning . . .
- . . . and as by-product get a data-driven version of the periodic table!



Willatt, Musil, Ceriotti, PCCP (2018)

Recognizing active protein ligands

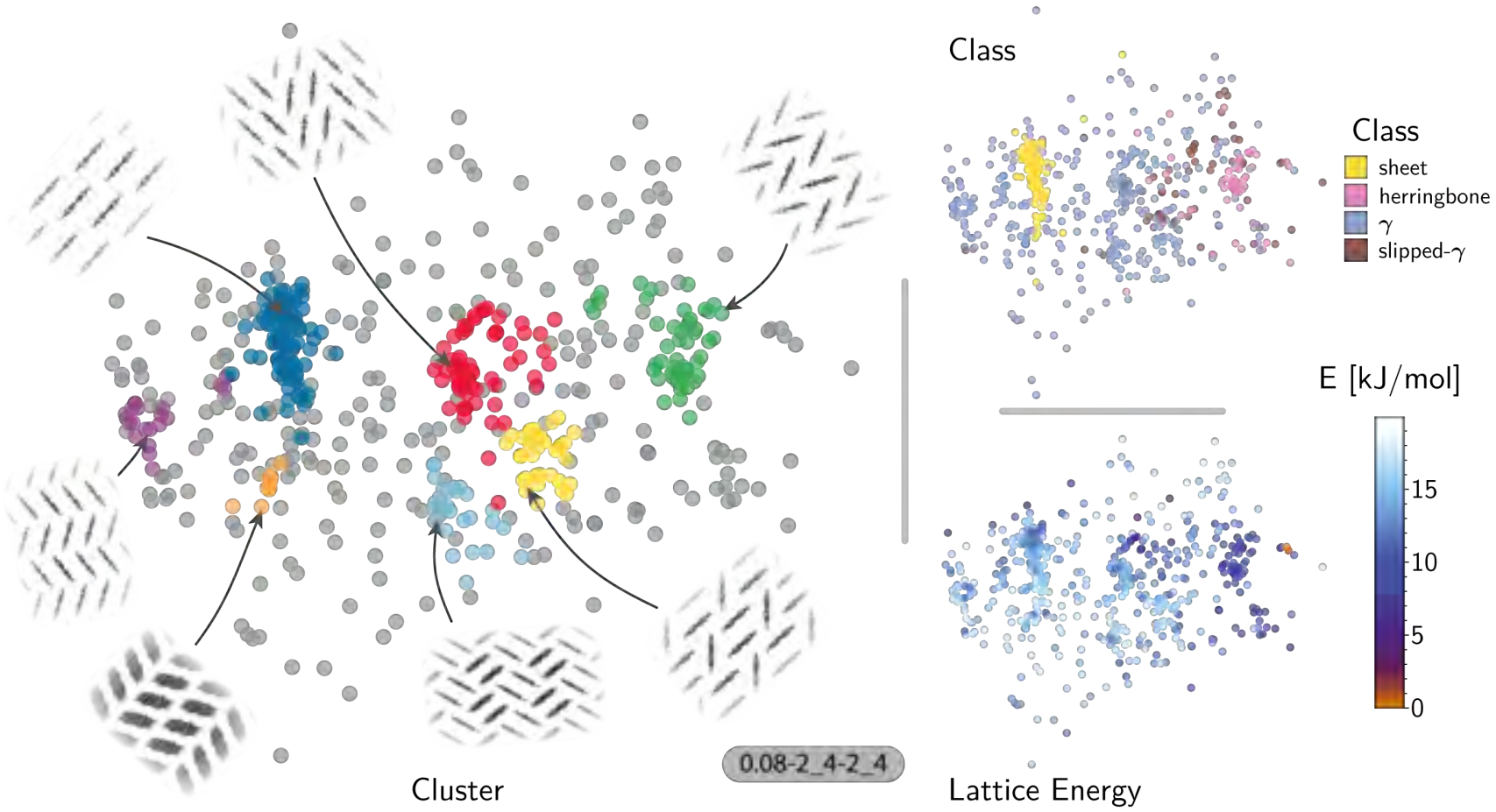
- A SOAP-REMatch-based KSVM classifies active and inactive ligands with 99% accuracy; non-additive model is crucial!
- Sensitivity analysis help identify the active “warhead” and could guide drug design and optimization



Bartok, De, Poelking, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017) *[data: DUD-E, Shoichet]*

Structure-property landscapes

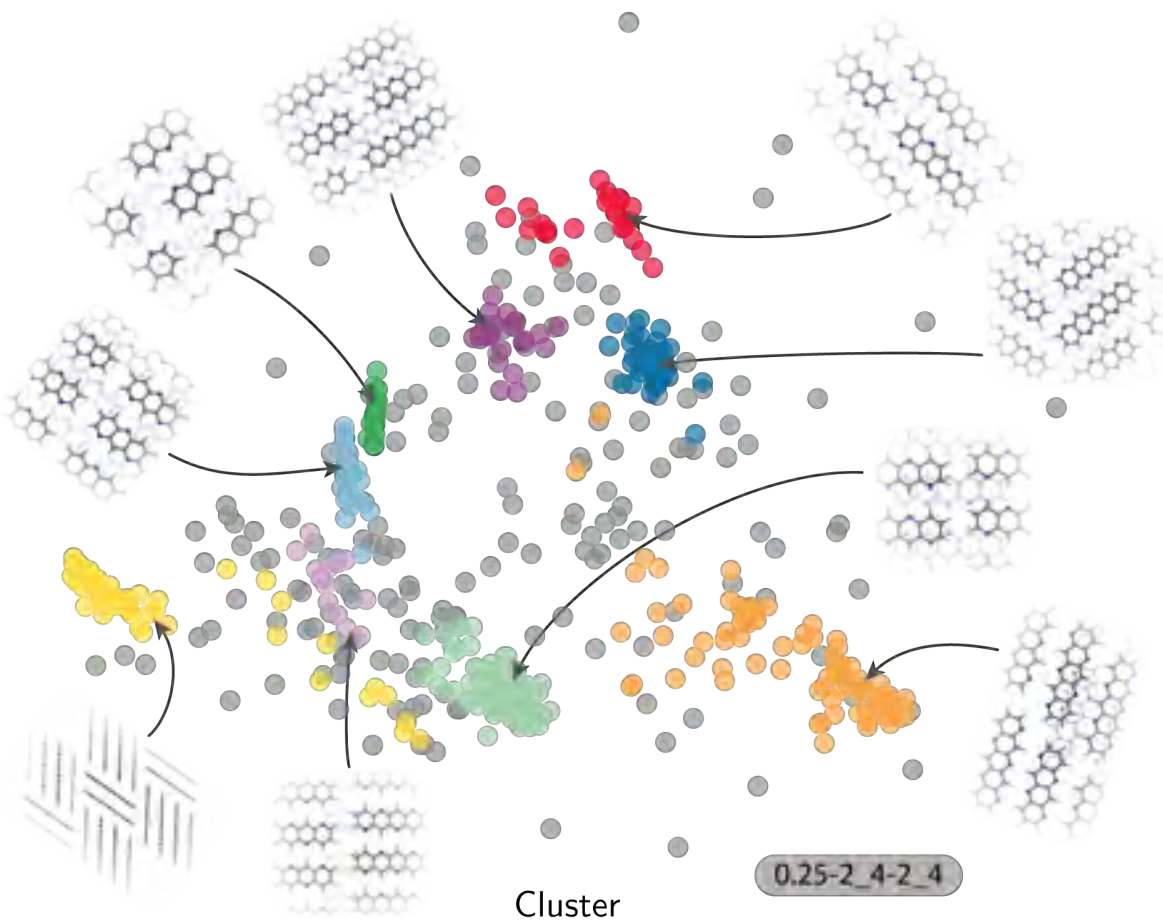
- Clustering/sketch-maps based on REMatch-SOAP correlate well with qualitative classification of packing motifs, and with properties (ex.: azapentacene structure-energy-property landscape maps)

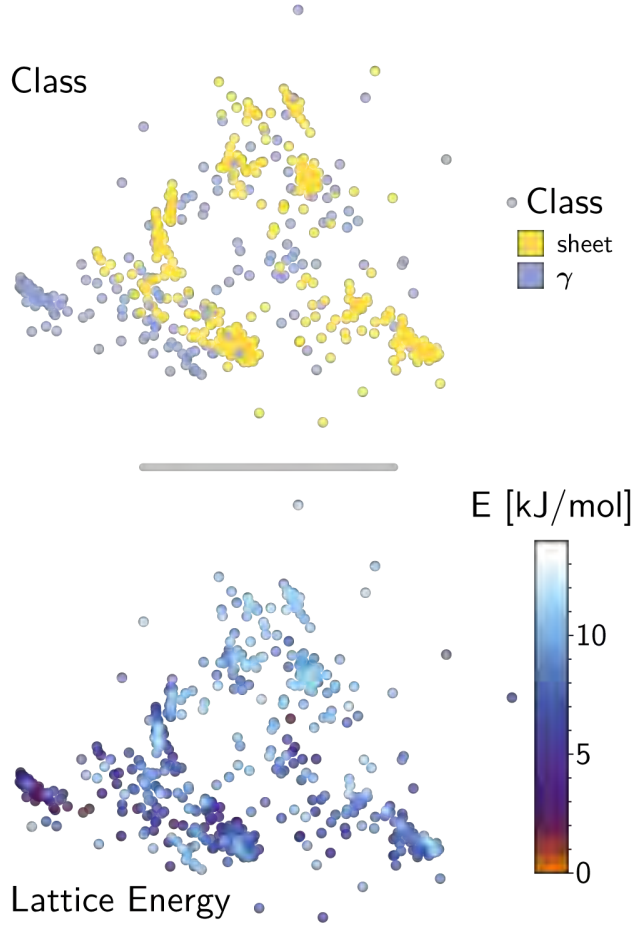


Musil, De, Yang, Campbell, Day, **MC**, Chemical Science (2018);`http://interactive.sketchmap.org`

Structure-property landscapes

- Clustering/sketch-maps based on REMatch-SOAP correlate well with qualitative classification of packing motifs, and with properties (ex.: azapentacene structure-energy-property landscape maps)

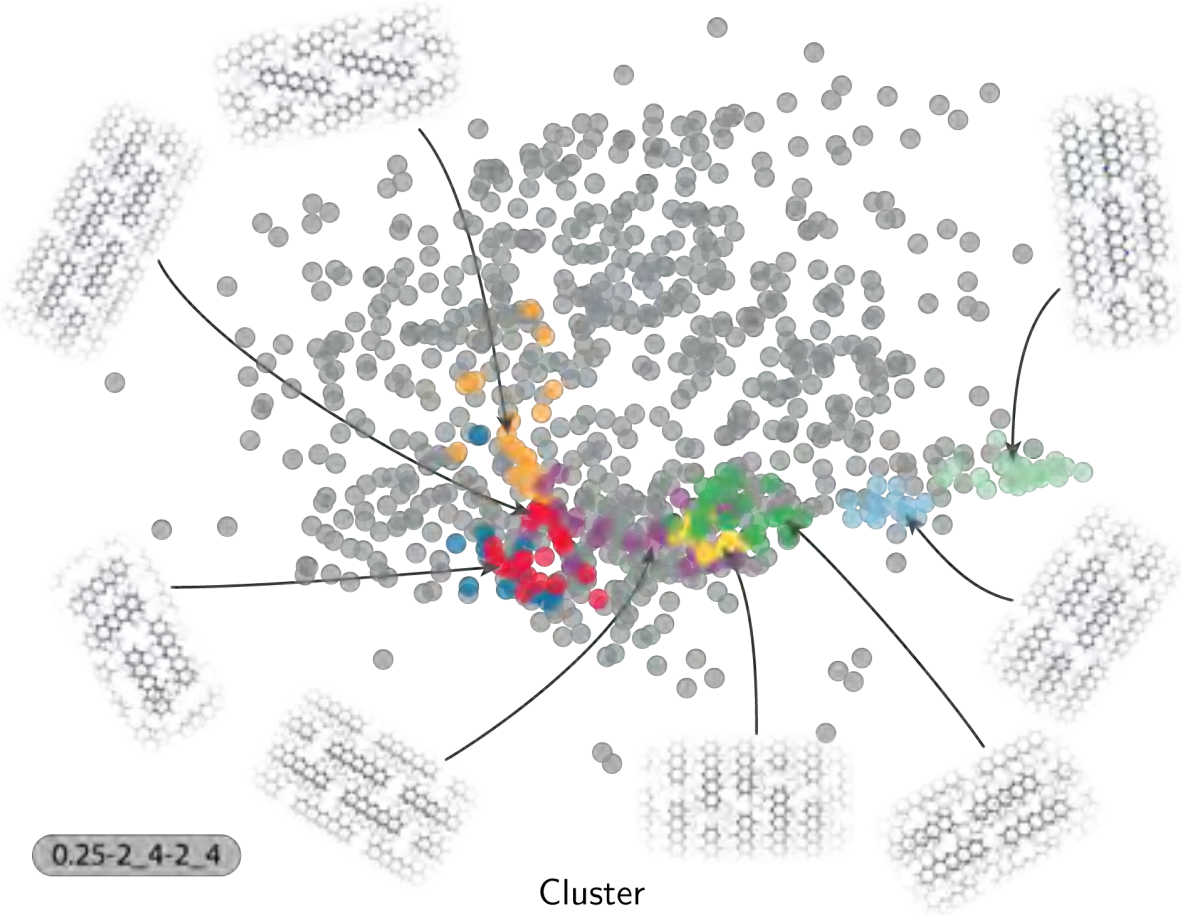


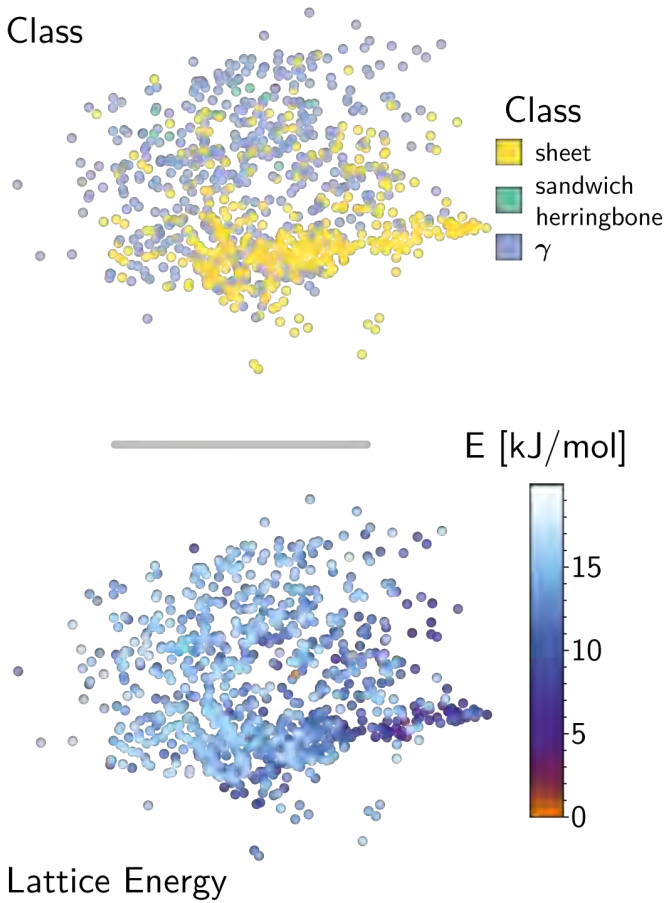


Musil, De, Yang, Campbell, Day, **MC**, Chemical Science (2018);`http://interactive.sketchmap.org`

Structure-property landscapes

- Clustering/sketch-maps based on REMatch-SOAP correlate well with qualitative classification of packing motifs, and with properties (ex.: azapentacene structure-energy-property landscape maps)





Musil, De, Yang, Campbell, Day, **MC**, Chemical Science (2018);`http://interactive.sketchmap.org`

Principal Covariates Regression

- Very simple idea to combine PCA and latent-space LR to find a dimensionality reduction that preserves variance and predicts well

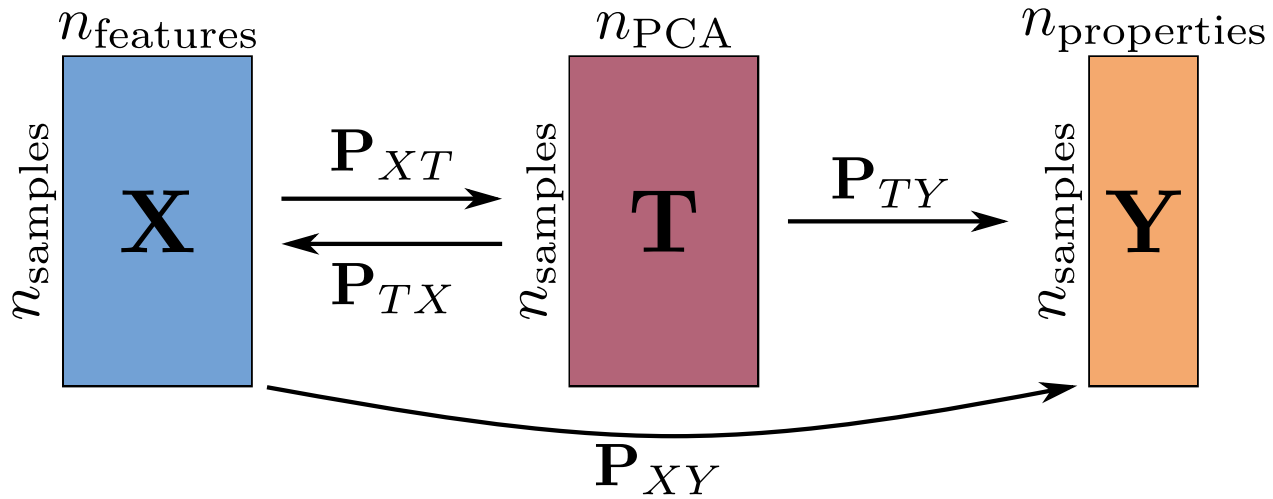
$$\ell = \alpha \left\| \mathbf{X} - \mathbf{X}\mathbf{P}_{XT}\mathbf{P}_{TX} \right\|^2 + (1-\alpha) \left\| \mathbf{Y} - \mathbf{X}\mathbf{P}_{XT}\mathbf{P}_{TY} \right\|^2$$

- Solution can be found working in sample space (looking for the eigenvectors of a modified Gram matrix)

$$\tilde{\mathbf{K}} = \alpha \mathbf{X}\mathbf{X}^T + (1-\alpha)\, \mathbf{X}\mathbf{P}_{XY}\mathbf{P}_{XY}^T\mathbf{X}^T$$

- ... or in feature space by diagonalizing a modified covariance

$$\tilde{\mathbf{C}} = \alpha \mathbf{X}^T\mathbf{X} + (1-\alpha)\, \left(\mathbf{X}^T\mathbf{X}\right)^{-1/2} \mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X} \left(\mathbf{X}^T\mathbf{X}\right)^{-1/2}$$

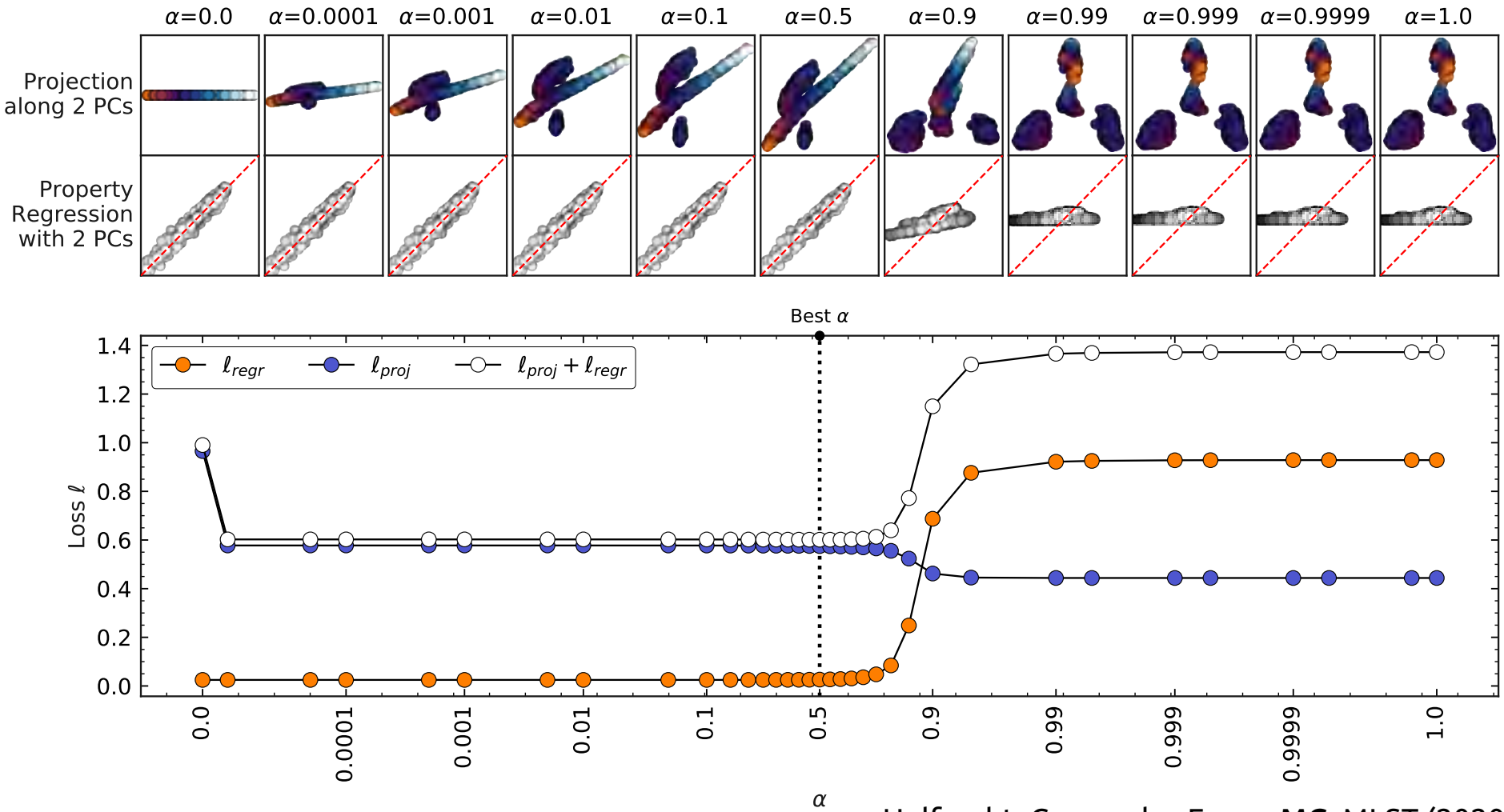


S. de Jong and HAL Kiers, Scandinavian Symposium on Chemometrics (1992)

Kernel PCovR

- Kernel versions of PCovR can be obtained with a modified kernel $\tilde{\mathbf{K}} = \alpha\mathbf{K} + (1-\alpha)\,\hat{\mathbf{Y}}\hat{\mathbf{Y}}^T$, diagonalizing it and finding the projector

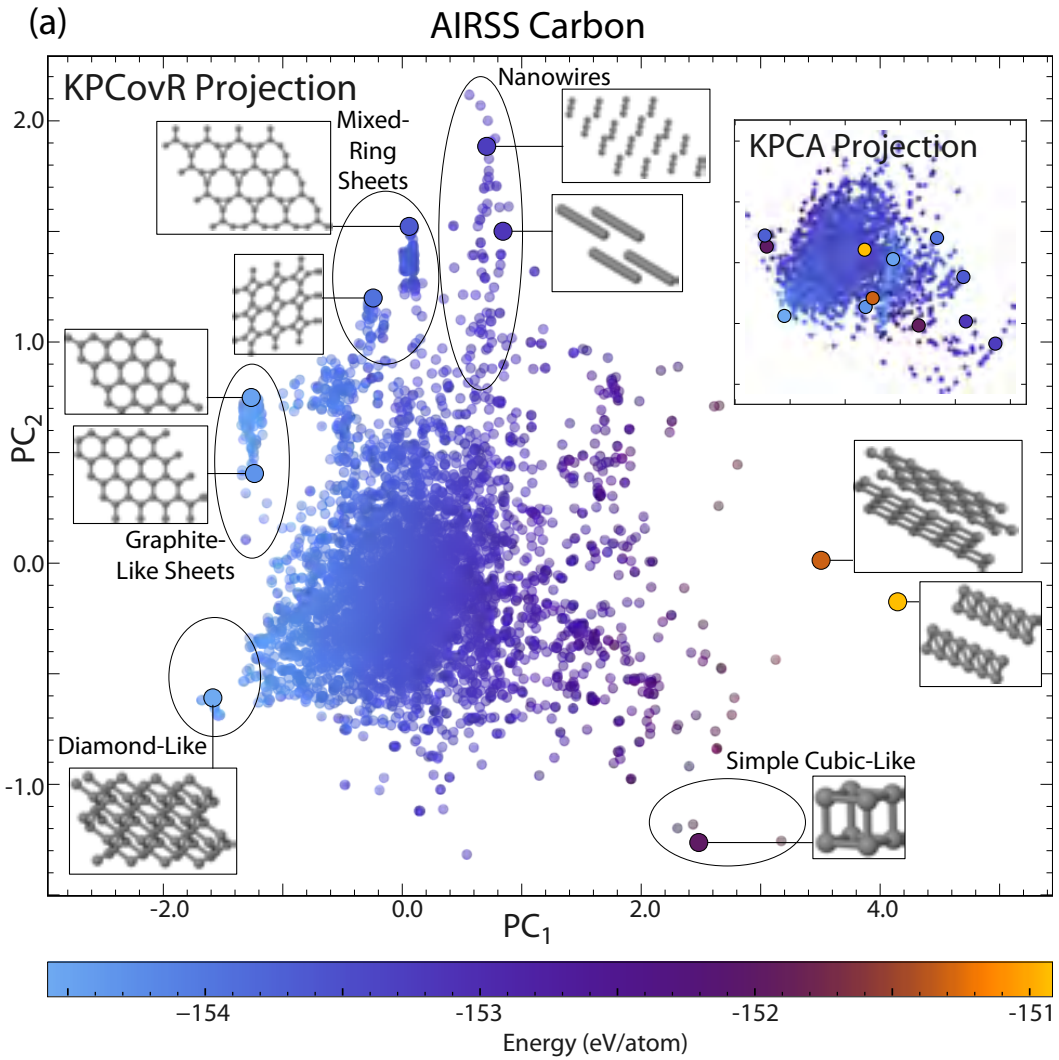
$$\mathbf{P}_{\mathbf{K}T} = \left(\alpha\mathbf{I} + (1-\alpha)\left(\mathbf{K} + \lambda\mathbf{I}\right)^{-1}\mathbf{Y}\hat{\mathbf{Y}}\right)\mathbf{U}_{\tilde{\mathbf{K}}}\Lambda_{\tilde{\mathbf{K}}}^{1/2}$$



Helfrecht, Cersonsky, Fraux, **MC**, MLST (2020)

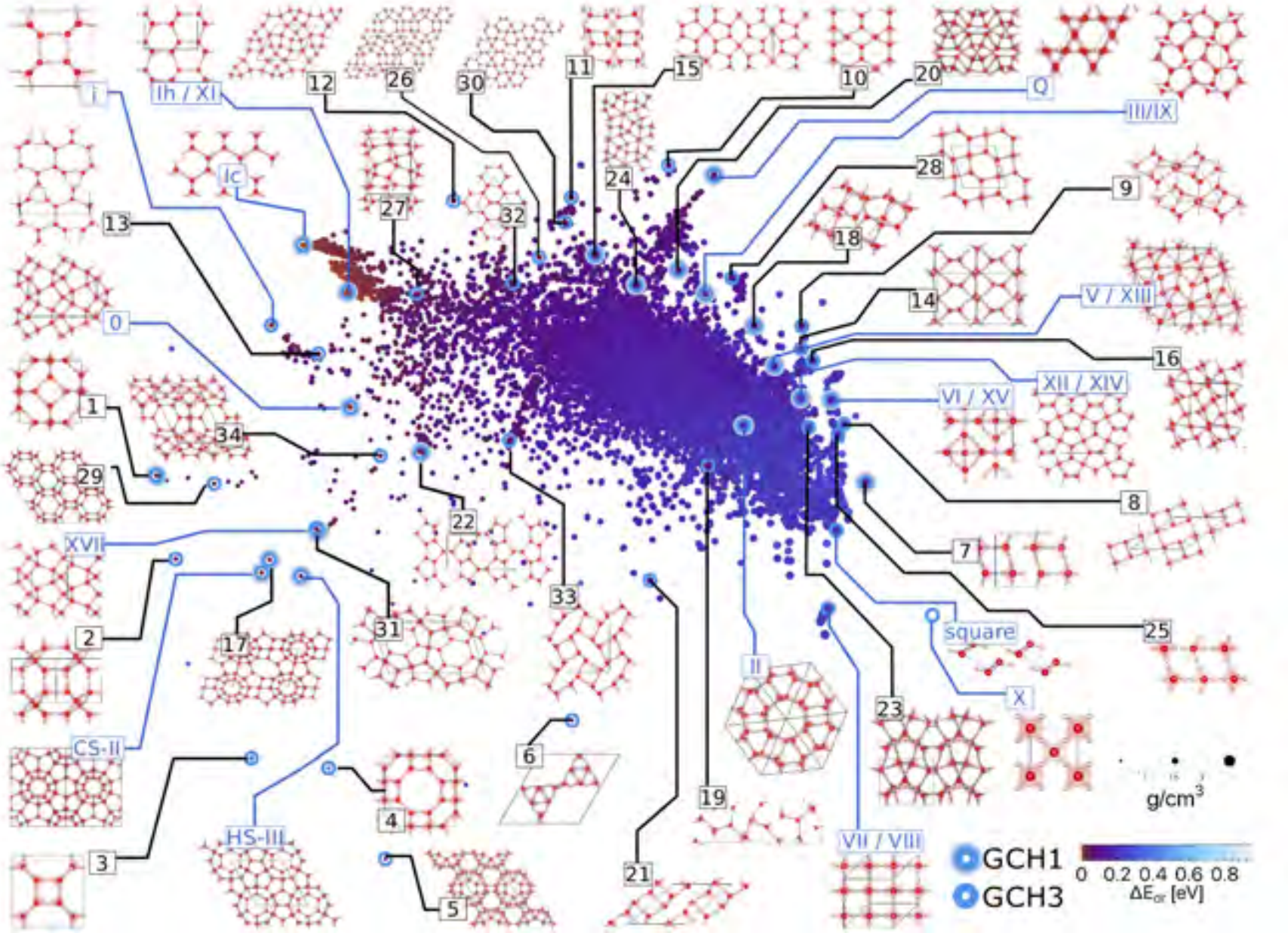
Where unsupervised meets supervised

- Using KPCovR to reveal structure-property relations in databases of materials structures



Helfrecht, Cersonsky, Fraux, **MC**, MLST (2020)

A Generalized Convex Hull Construction



Anelli, Engel, Pickard & **MC**, PRM (2019); Engel, Anelli, **MC**, Pickard & Needs, Nature Comm. (2018)