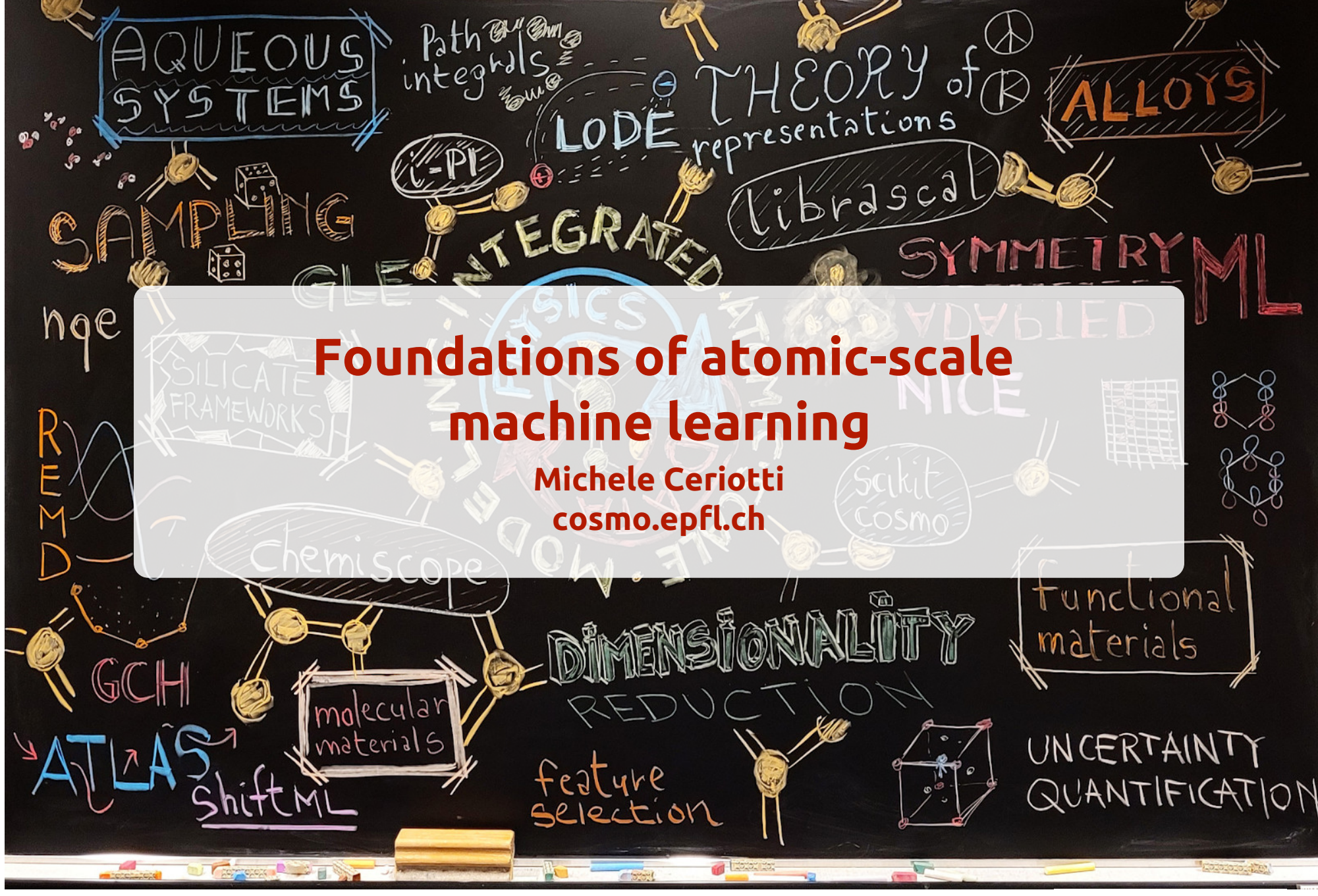


Foundations of atomic-scale machine learning

Michele Ceriotti
cosmo.epfl.ch

Predictive materials modeling with atomic-scale simulations

Simple models, complex physics

- Simple models, with minimal number of parameters fitted by comparison with experimental quantities
- Aim: capture the essence of atomic-scale interactions, and understand emergent phenomena (phase transitions, equations of state...)

$$V(\{\mathbf{r}\}) \sim \underbrace{\sum_{ij} \frac{z_i z_j}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\text{electrostatics}} + \underbrace{\sum_{\text{bonds}} k_i (\mathbf{r}_i - \mathbf{r}'_i)^2}_{\text{bonded terms}} - \underbrace{\sum_{ij} \frac{A_{ij}}{|\mathbf{r}_i - \mathbf{r}_j|^6}}_{\text{dispersion}} + \dots$$

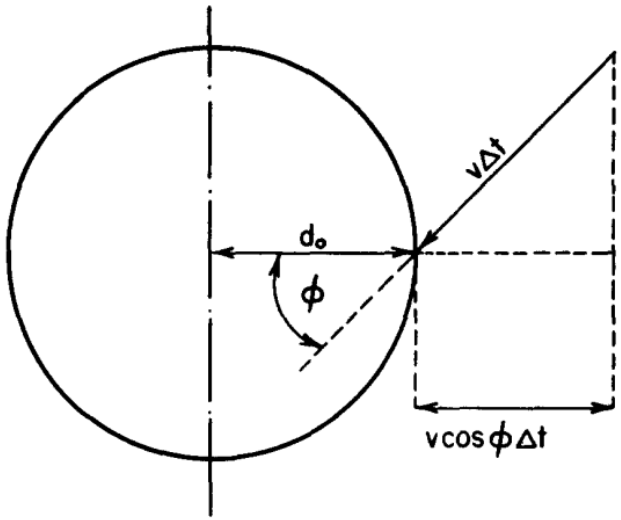
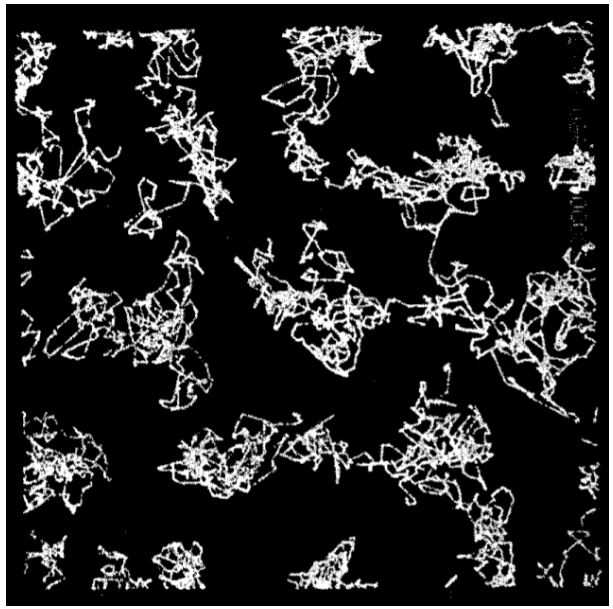
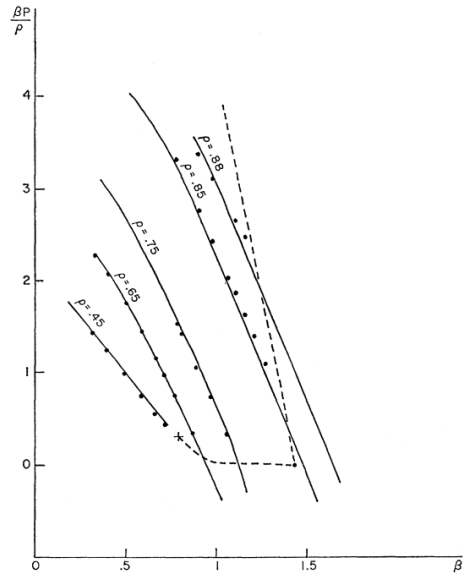


FIG. 1. Collisions of rigid spheres.





Metropolis et al., JCP (1953); Alder & Wainwright, JCP (1959); Verlet, Phys. Rev. (1969)

First-principles calculations

- Practical approaches to evaluate the electronic structure
- Quantitatively accurate simulations, without fit to experiments
- High computational effort

$$\{-\tfrac{1}{2}\nabla^2+\varphi(\mathbf{r})+\mu_c(\mathbf{r})\}\psi_i(\mathbf{r})$$

$$-\int \frac{n_1(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\psi_i(\mathbf{r}')\,d\mathbf{r}'=\epsilon_i\psi_i(\mathbf{r})\,, \qquad (2.22)$$

where

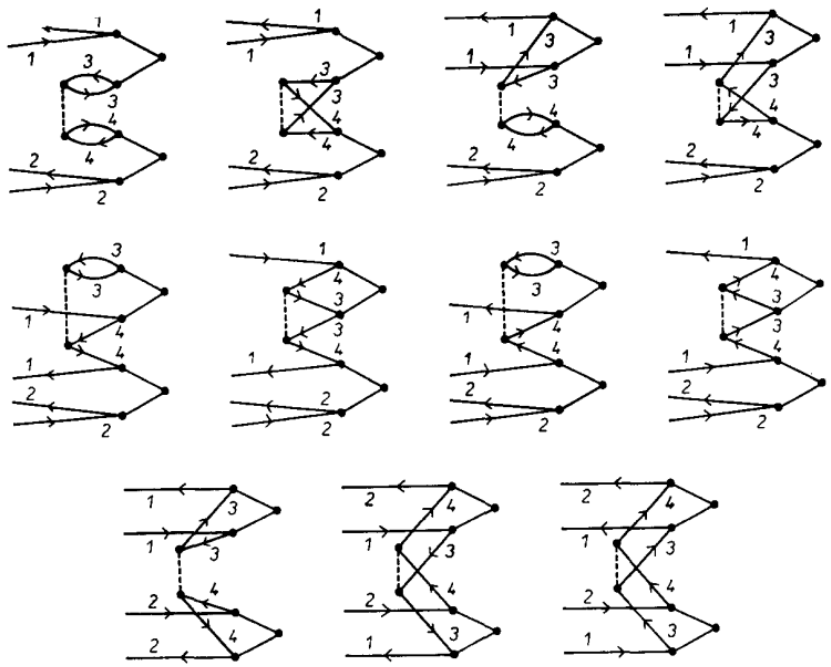
$$\mu_c=d(n\epsilon_c)/dn\,, \qquad (2.23)$$

$$n_1(\mathbf{r},\mathbf{r}')=\sum_{j=1}^{N}\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')\,, \qquad (2.24)$$

$$\mu\ddot{\psi}_i(\mathbf{r},t)=-\delta E/\delta\psi_i^*(\mathbf{r},t)+\sum_k\Lambda_{ik}\psi_k(\mathbf{r},t), \qquad (5a)$$

$$M_I\ddot{R}_I=-\nabla_{R_I}E, \qquad (5b)$$

$$\mu_\nu\ddot{\alpha}_\nu=-(\partial E/\partial\alpha_\nu), \qquad (5c)$$

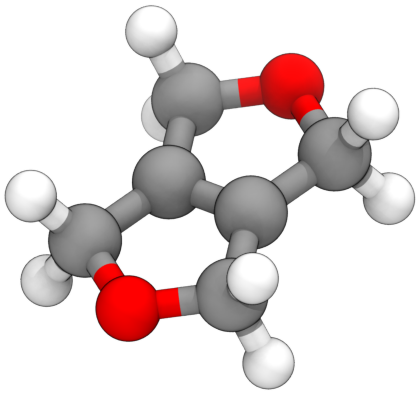


Kohn & Sham, Phys. Rev. (1965); Cížek, JCP (1966); Car & Parrinello, PRL (1985)

Surrogate models for quantum chemistry

- Electronic-structure calculations predict accurately molecular properties
- Machine-learning models provide inexpensive approximations
- There is more to life than energy and forces!

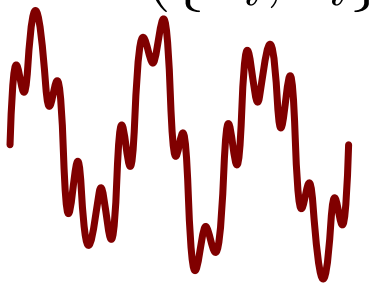
$A \equiv \{a_i, \mathbf{r}_i\}$



FIRST-PRINCIPLES QUANTUM MODELS

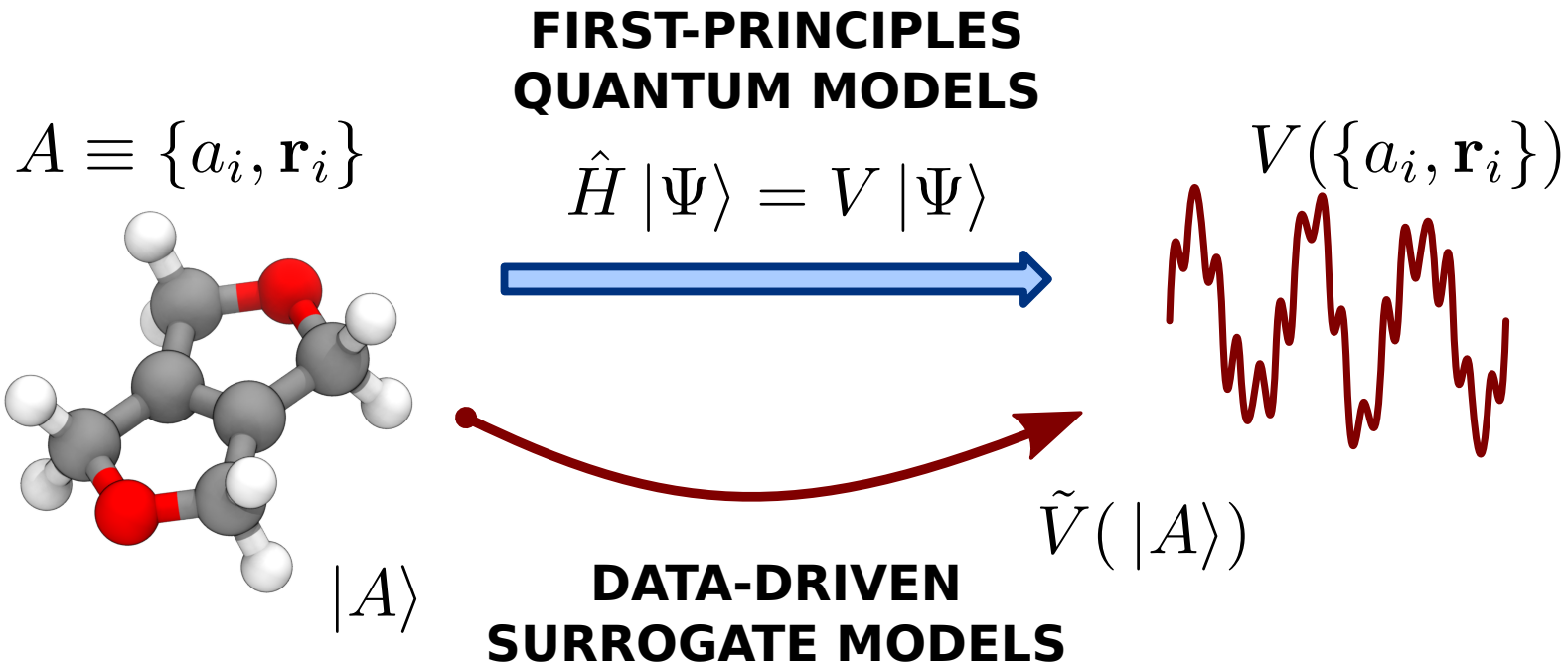
$\hat{H}\,|\Psi\rangle = V\,|\Psi\rangle$

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



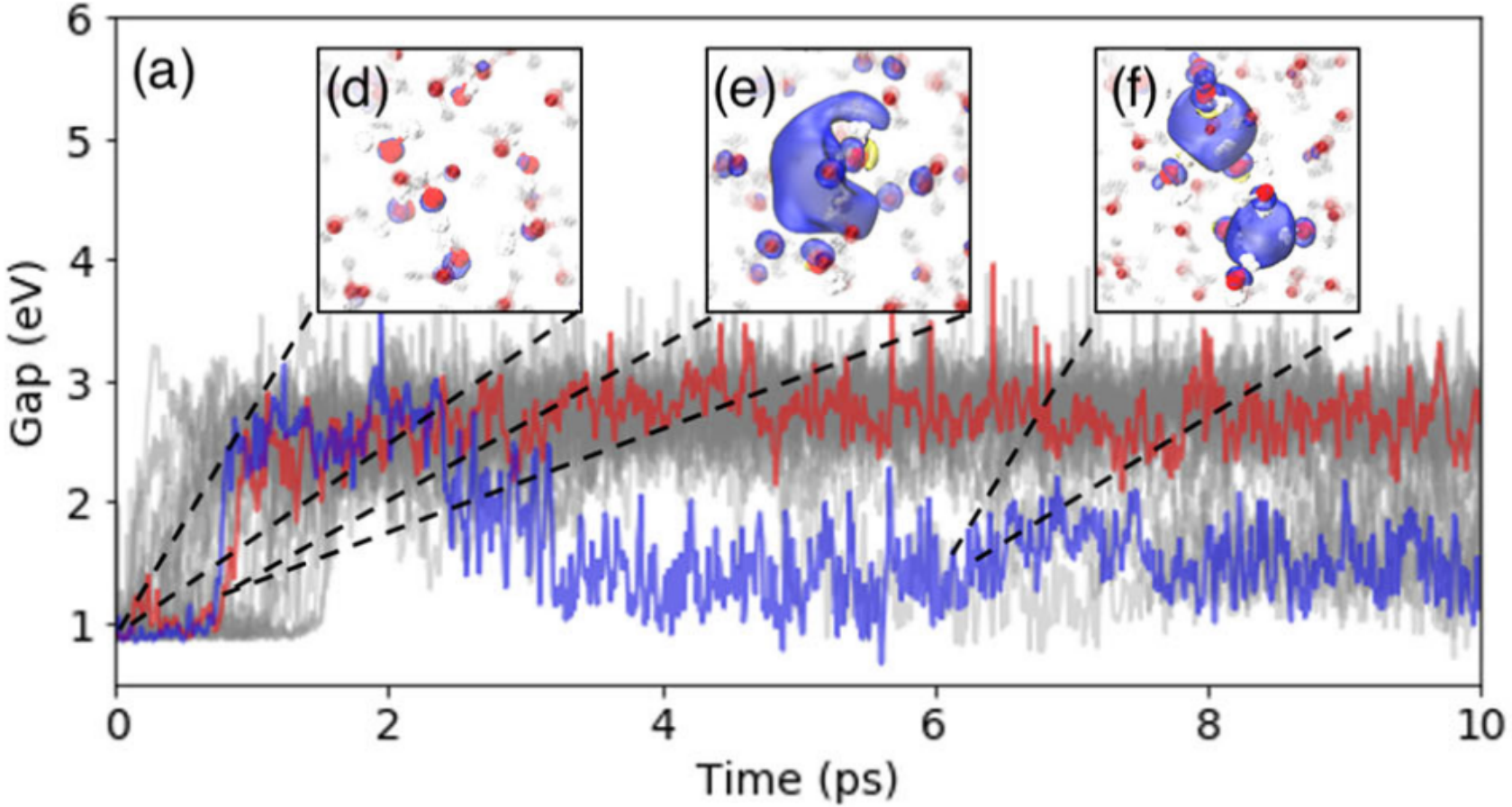
Surrogate models for quantum chemistry

- Electronic-structure calculations predict accurately molecular properties
- Machine-learning models provide inexpensive approximations
- There is more to life than energy and forces!



Surrogate models for quantum chemistry

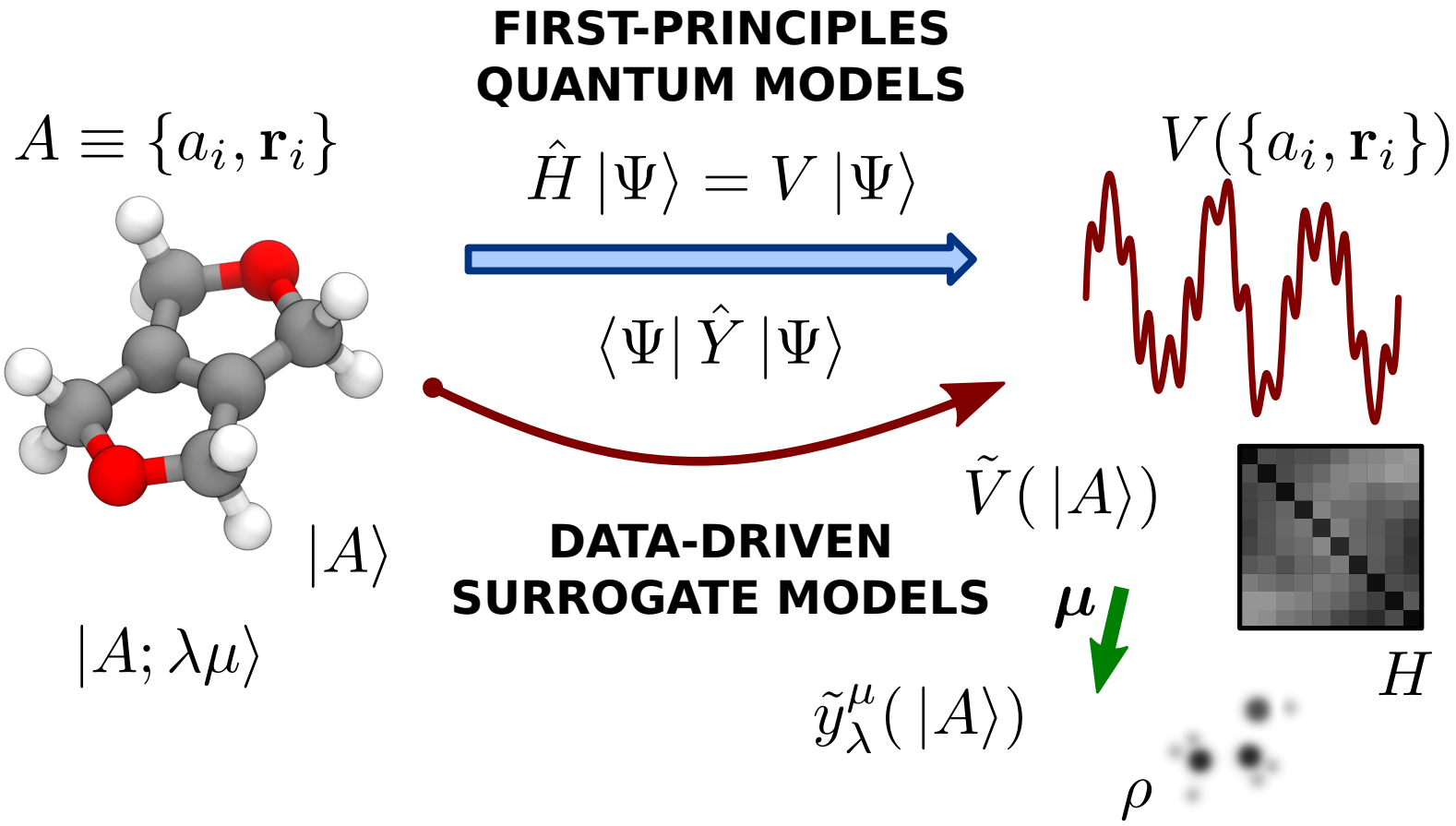
- Electronic-structure calculations predict accurately molecular properties
- Machine-learning models provide inexpensive approximations
- There is more to life than energy and forces!



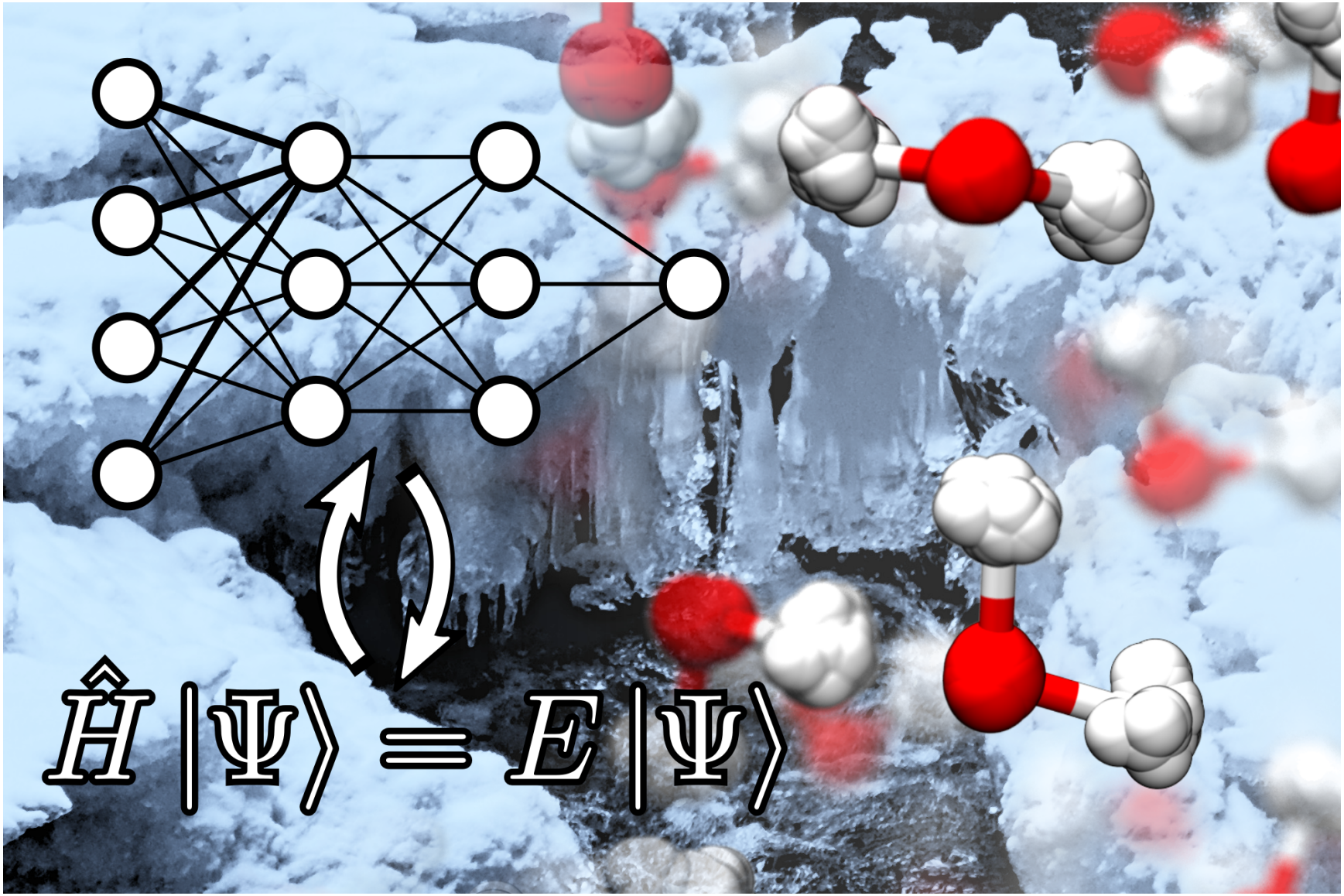
Lan et al., Nat. Comm. (2021)

Surrogate models for quantum chemistry

- Electronic-structure calculations predict accurately molecular properties
- Machine-learning models provide inexpensive approximations
- There is more to life than energy and forces!

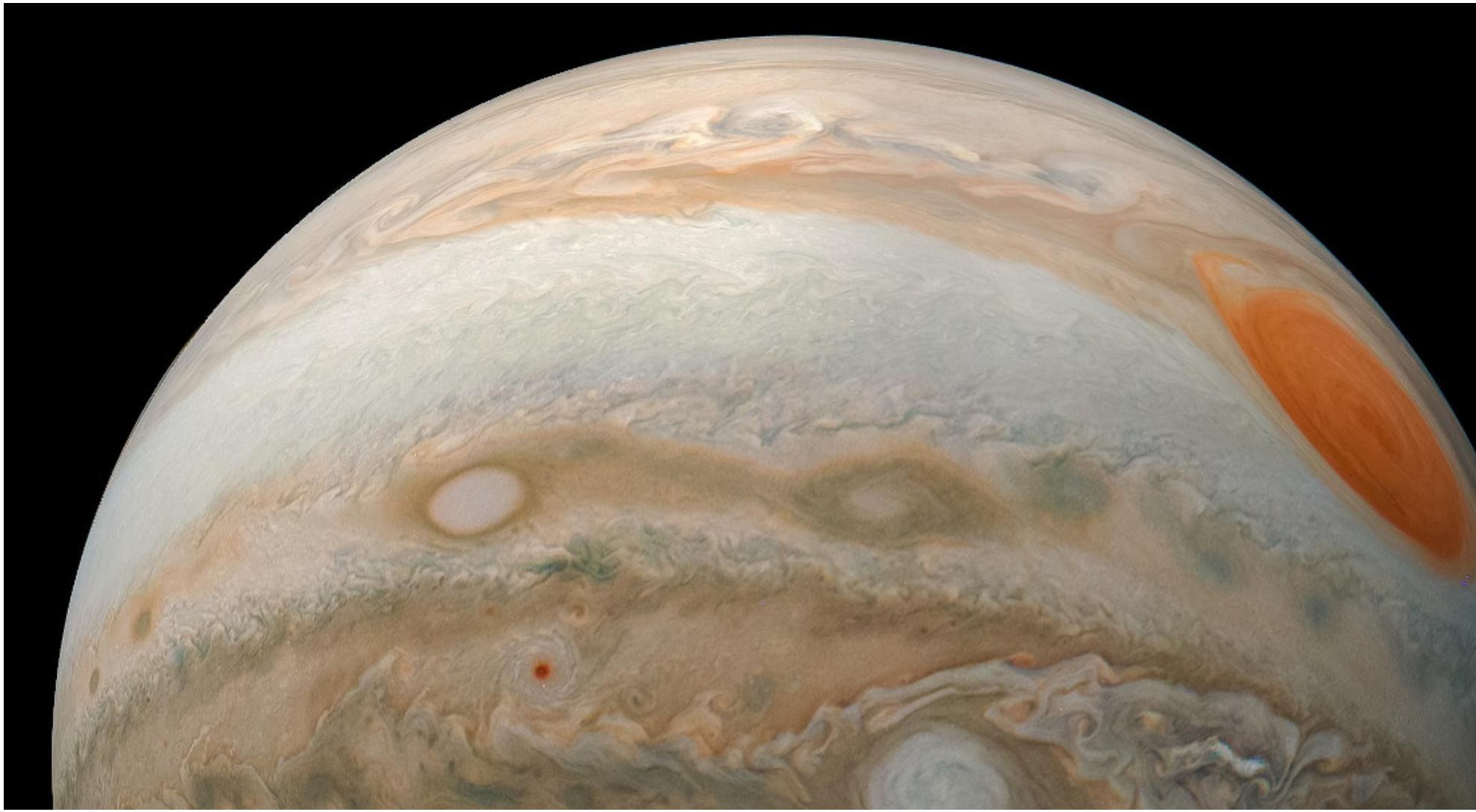


Atomistic machine learning & the glossy press



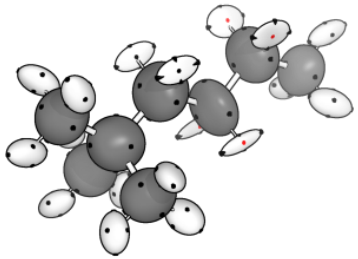
Cheng, Engel, Behler, Dellago, **MC**, PNAS (2019)

Atomistic machine learning & the glossy press

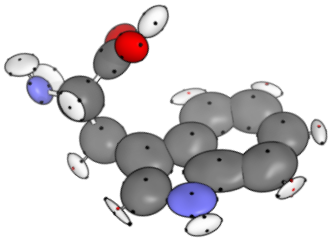


Cheng, Mazzola, Pickard, **MC**, Nature (2020)

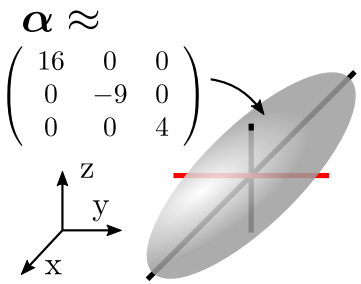
Atomistic machine learning & the glossy press

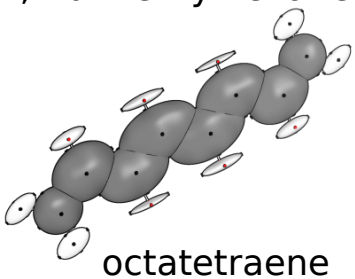


2,2-dimethylhexane

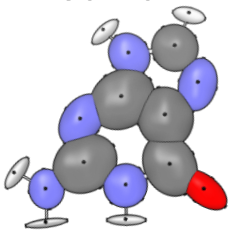


tryptophan

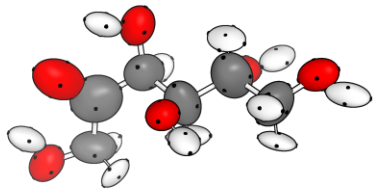




octatetraene

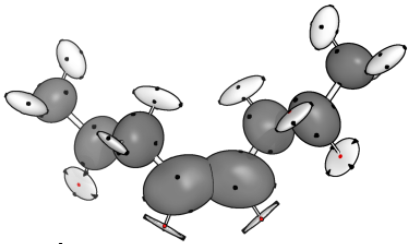


guanine

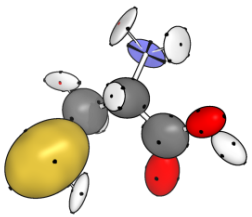


fructose

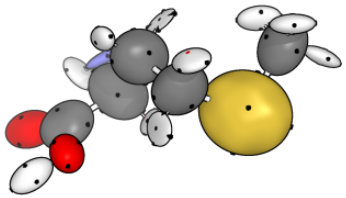
Method	RMSE
CCSD/ML	0.304
CCSD/DFT	0.573



cis-4-octene



cysteine



methionine

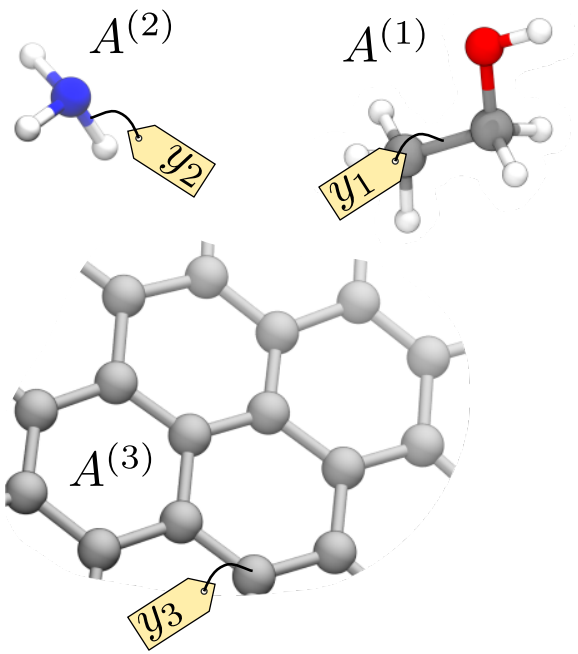
Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, PNAS (2019);

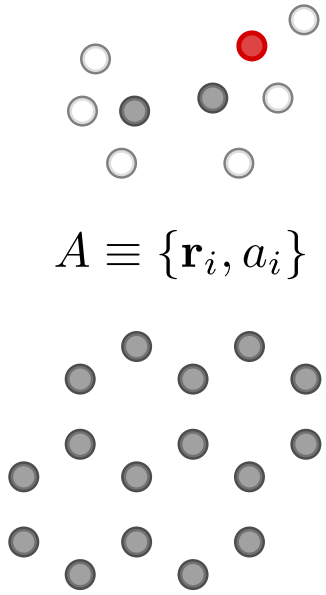
Atomistic machine learning & the glossy press

V. Deringer et al., Nature (2021)

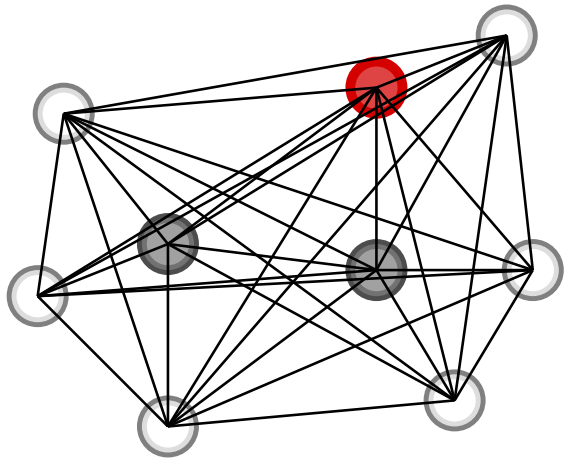
Molecules as atom clouds

- Can we develop a rigorous theory of representations for atomic-scale ML?
- From a mathematical perspective, molecules and materials are treated as point clouds, decorated by the chemical nature of the atoms
- An alternative view is that each molecule is a fully-connected graph, with atoms as nodes and *separating vectors* as edges





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



$$A \equiv \{a_i, \mathbf{r}_{ji}\}$$

A unified theory of ML representations

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

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

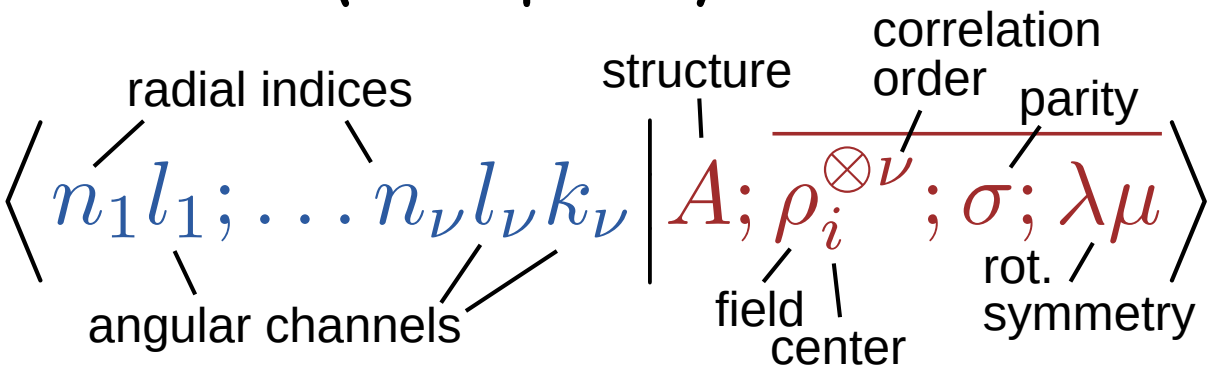
Willatt, Musil, **MC**, JCP (2019), arxiv:1807.00408; `https://tinyurl.com/dirac-rep`

A Dirac notation for ML

features index

$$\langle Q | A \rangle$$

representation target & nature

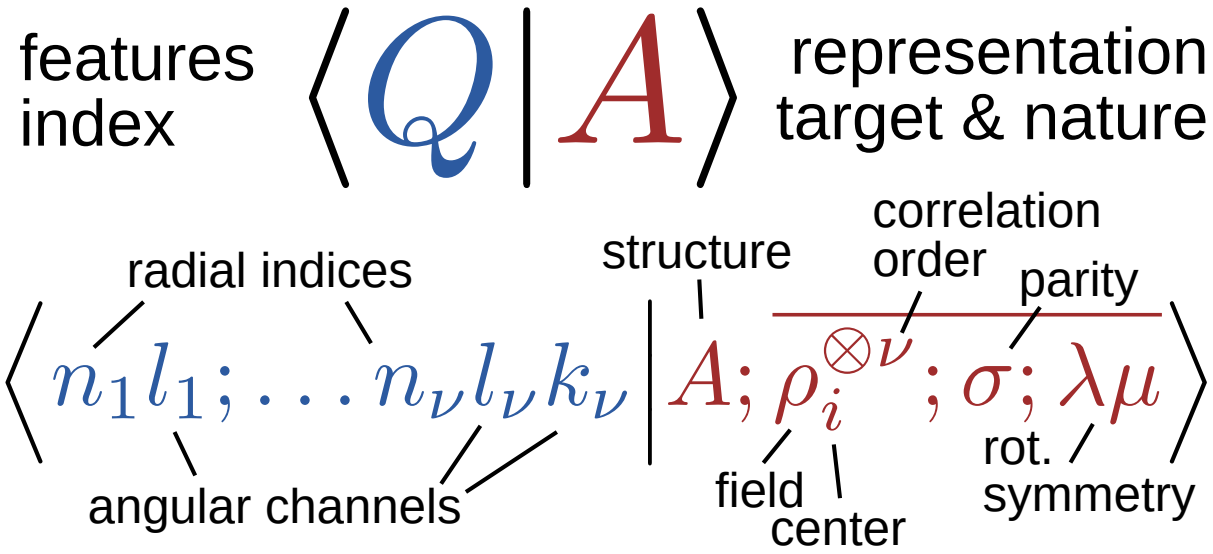


- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index Q
- Bra-ket notation $\langle Q|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally 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), arxiv:1807.00408; `https://tinyurl.com/dirac-rep`

A Dirac notation for ML

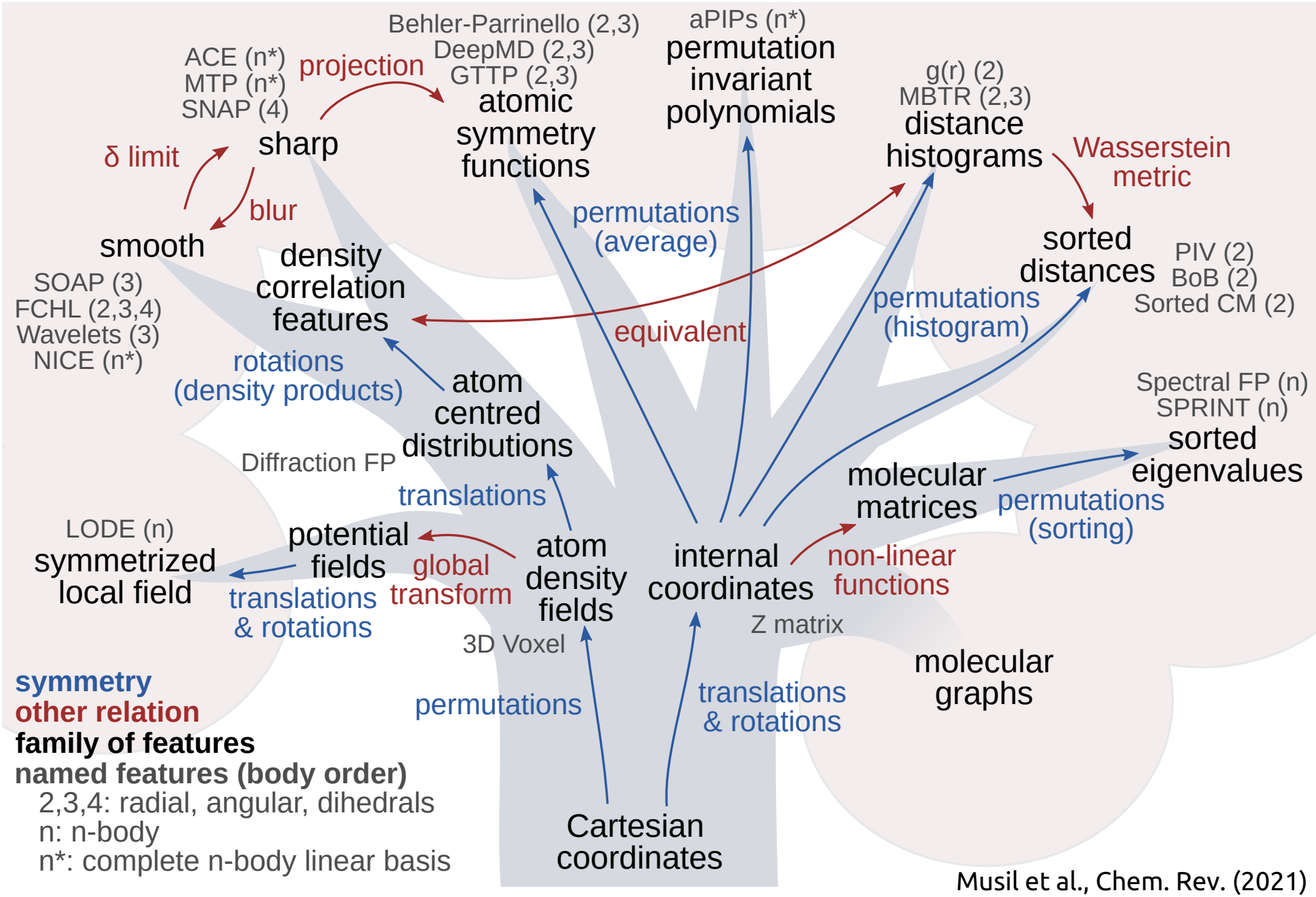


- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index Q
- Bra-ket notation $\langle Q|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally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

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

Willatt, Musil, **MC**, JCP (2019), arxiv:1807.00408; `https://tinyurl.com/dirac-rep`

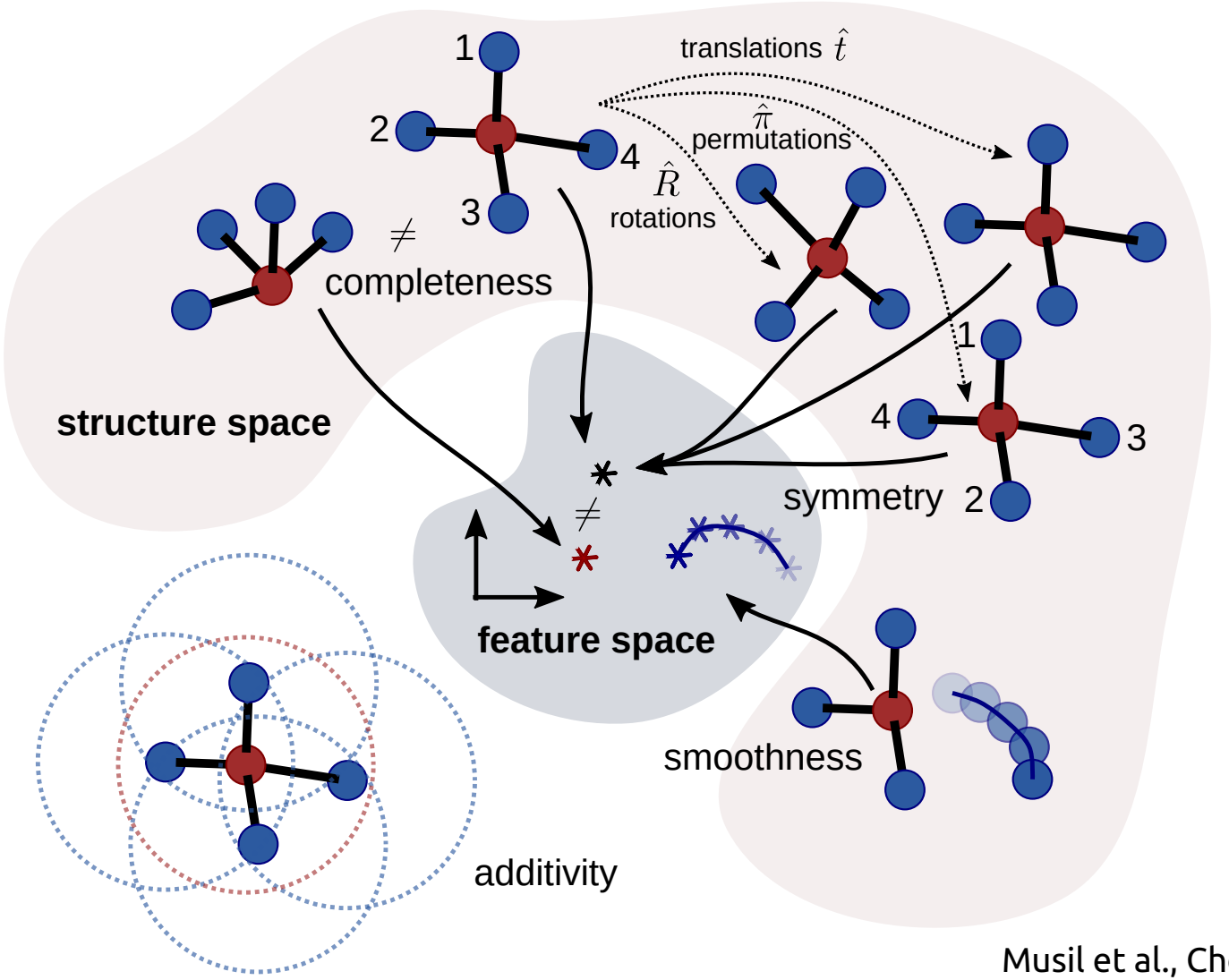
A phylogenetic tree of ML representations



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

What do we want from a representation?

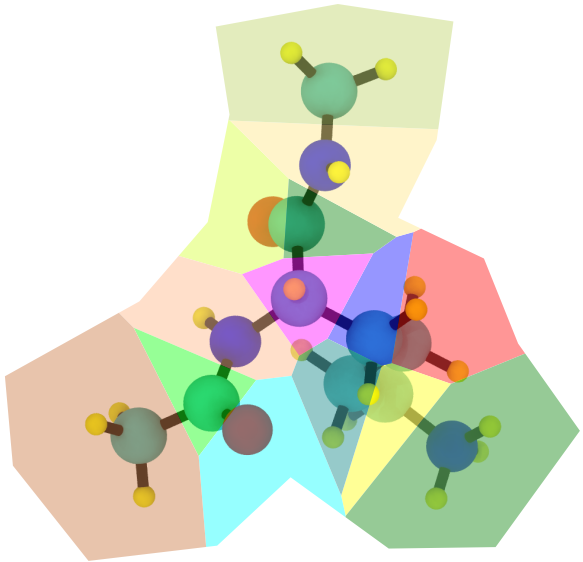
- Structure representations should: 1. be complete (injective); 2. reflect basic physical symmetries; 3. be smooth, regular; 4. exploit additivity
- Cartesian coordinates fulfill only 1 and 3

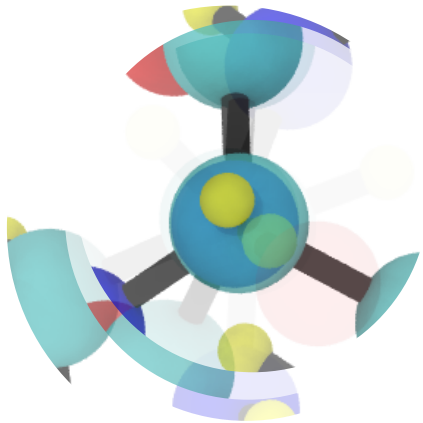


Musil et al., Chem. Rev. (2021)

Additivity, and locality

- A representation of a structure in terms of a sum over atom-centered terms implies (for a linear model or an average kernel) an additive form of the property
- Additivity and locality lead to transferable models with a *divide and conquer* structure





$$V(A) = \sum_{i \in A} V(A_i)$$

$$|A\rangle = \sum_i |A_i\rangle$$

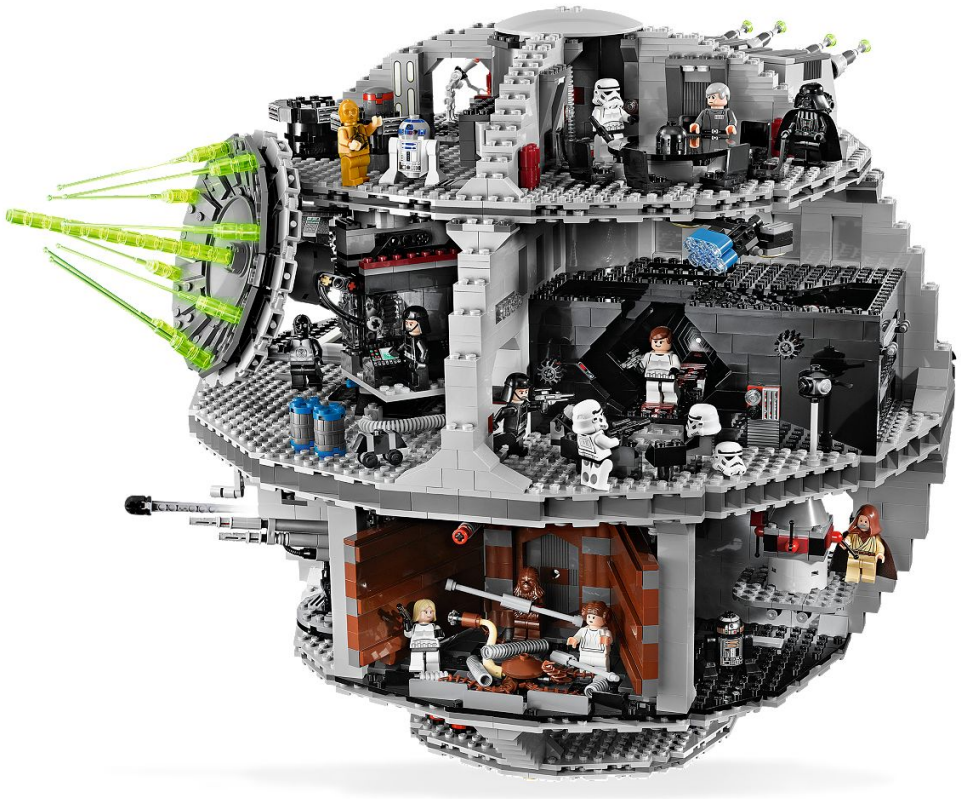
$$K(A,B) = \sum_{i,j} k(A_i, B_j)$$

Additivity, and locality

- A representation of a structure in terms of a sum over atom-centered terms implies (for a linear model or an average kernel) an additive form of the property
- Additivity and locality lead to transferable models with a *divide and conquer* structure

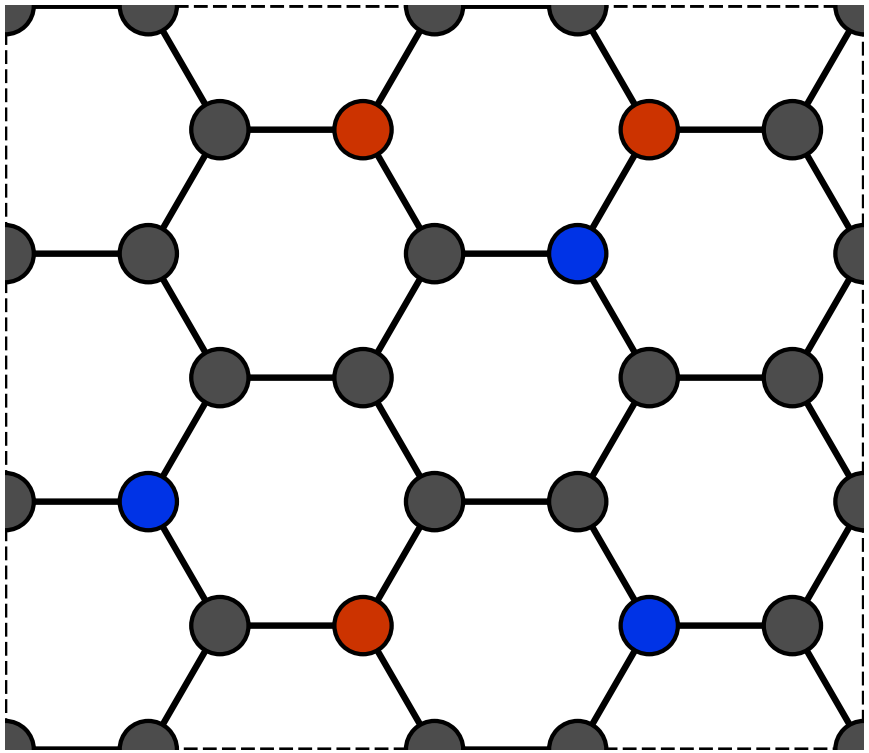
Additivity, and locality

- A representation of a structure in terms of a sum over atom-centered terms implies (for a linear model or an average kernel) an additive form of the property
- Additivity and locality lead to transferable models with a *divide and conquer* structure



Symmetrized field construction

- Start from a non-symmetric representation (Cartesian coordinates)
- Define a decorated atom-density $|\rho\rangle$ (permutation invariant)
- Translational average *of a tensor product* $|\rho\rangle \otimes |\rho\rangle$ yields atom-centred (and $\hat{t}$ invariant) $|\rho_i\rangle$



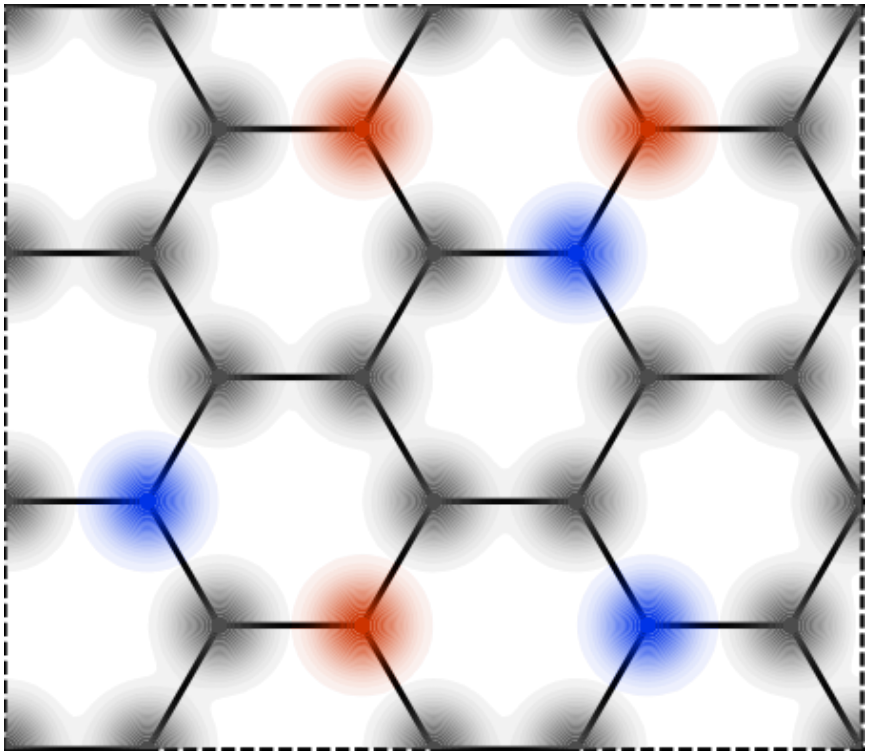
$A \equiv$

```
C 0.00 0.00 0.00
C 0.00 1.00 0.00
B 1.00 2.00 0.00
.......
```

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

Symmetrized field construction

- Start from a non-symmetric representation (Cartesian coordinates)
- Define a decorated atom-density $|\rho\rangle$ (permutation invariant)
- Translational average *of a tensor product* $|\rho\rangle \otimes |\rho\rangle$ yields atom-centred (and $\hat{t}$ invariant) $|\rho_i\rangle$



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

$|\mathrm{C}\rangle$ ■

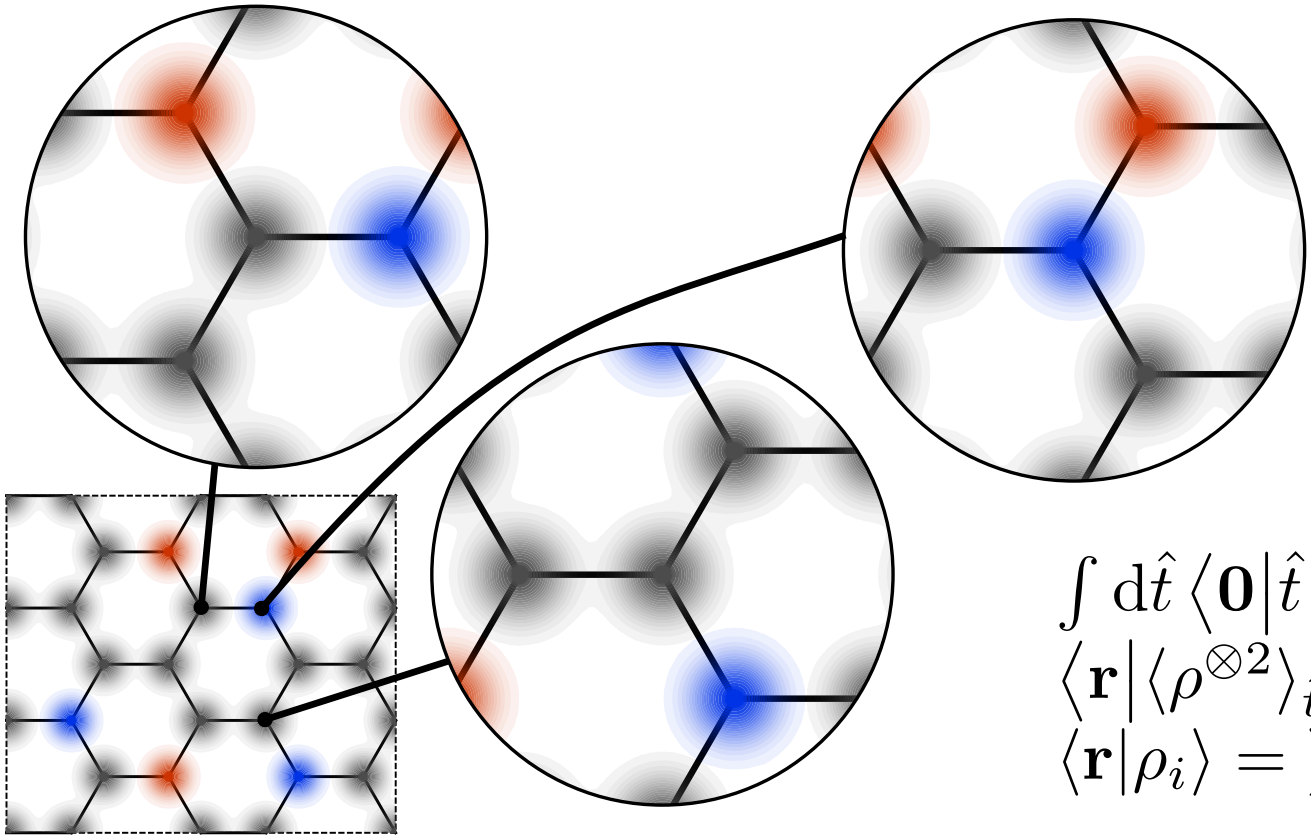
$|\mathrm{N}\rangle$ ■

$|\mathrm{B}\rangle$ ■

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

Symmetrized field construction

- Start from a non-symmetric representation (Cartesian coordinates)
- Define a decorated atom-density $|\rho\rangle$ (permutation invariant)
- Translational average *of a tensor product* $|\rho\rangle \otimes |\rho\rangle$ yields atom-centred (and $\hat{t}$ invariant) $|\rho_i\rangle$



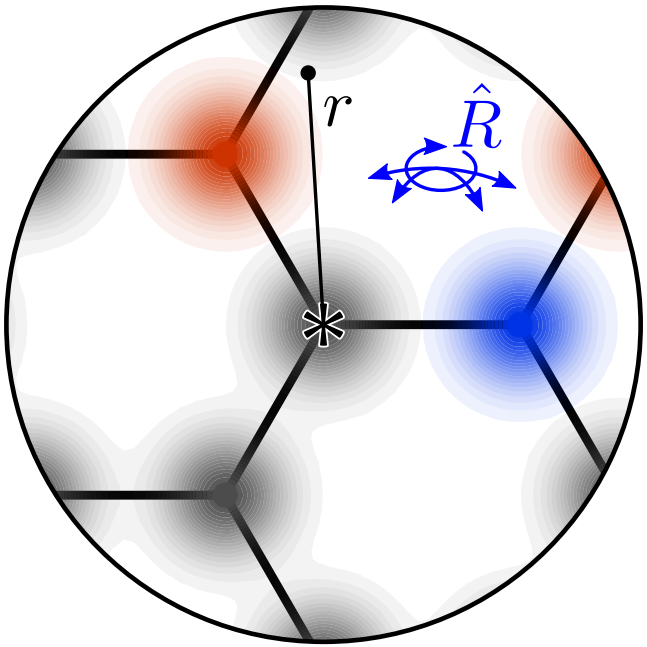
$$\int \mathrm{d}\hat{t} \, \langle \mathbf{0} | \hat{t} | \rho \rangle \, \langle \mathbf{r} | \hat{t} | \rho \rangle =$$
$$\langle \mathbf{r} | \langle \rho^{\otimes 2} \rangle_{\hat{t}} \rangle = \sum_i \langle \mathbf{r} | \rho_i \rangle$$
$$\langle \mathbf{r} | \rho_i \rangle = \sum_j g(\mathbf{r} - \mathbf{r}_{ij})$$

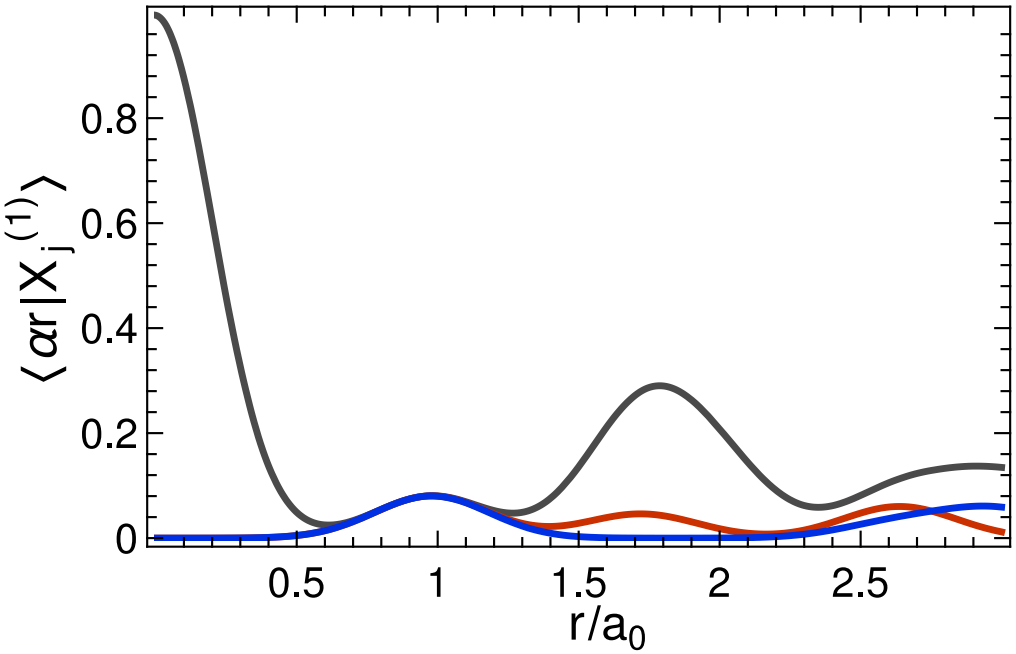
Willatt, Musil, **MC**, JCP (2019)

A universal feature construction

- Rotationally-averaged representations are essentially the same *n*-body correlations that are used in statistical theories of liquids
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

$$V(A_i) = \sum_{ij} V^{(2)}(r_{ij}) + \sum_{ij} V^{(3)}(r_{ij}, r_{ik}, \omega_{ijk}) \dots$$





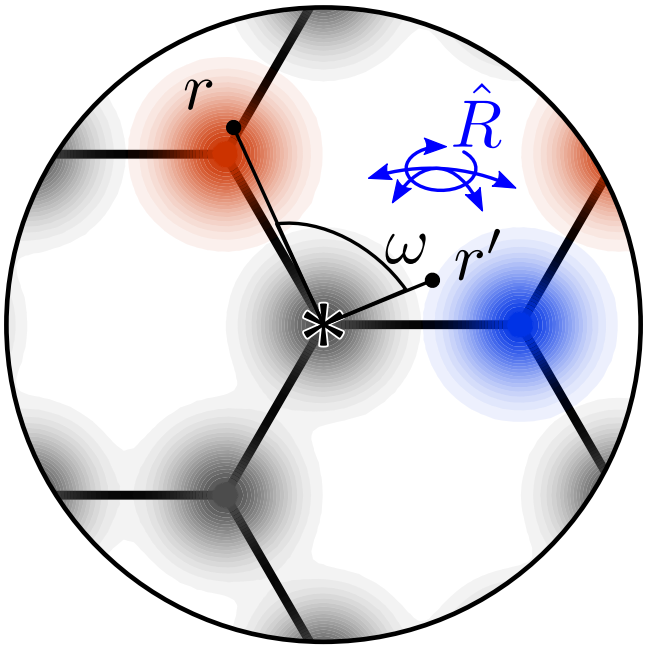
$$\langle ar|\overline{\rho_i^{\otimes 1}}\rangle = \int \mathrm{d}\hat{R} \langle ar\hat{\mathbf{r}}|\hat{R}|\rho_i\rangle$$

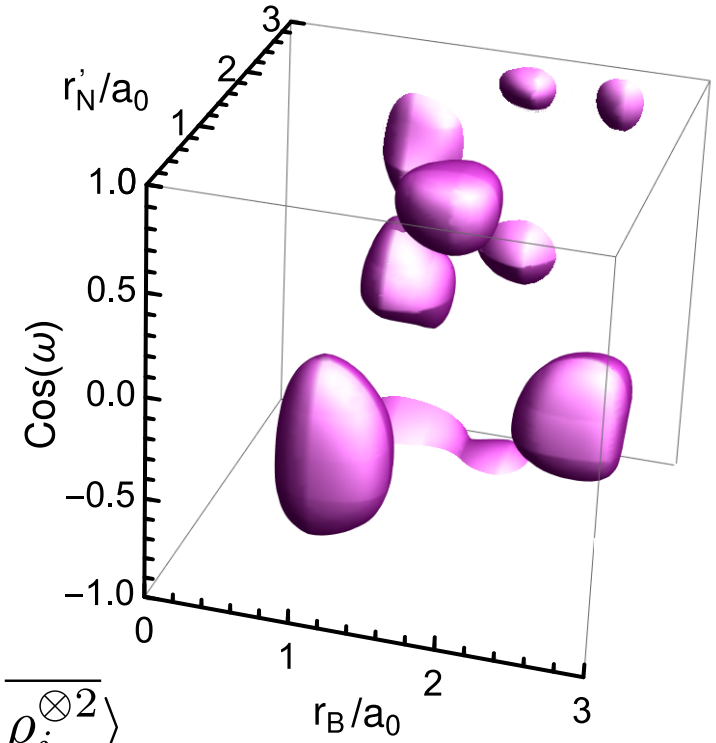
Willatt, Musil, **MC**, JCP (2019); Bartók, Kondor, Csányi PRB 2013

A universal feature construction

- Rotationally-averaged representations are essentially the same *n*-body correlations that are used in statistical theories of liquids
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

$$V(A_i) = \sum_{ij} V^{(2)}(r_{ij}) + \sum_{ij} V^{(3)}(r_{ij}, r_{ik}, \omega_{ijk}) \ldots$$





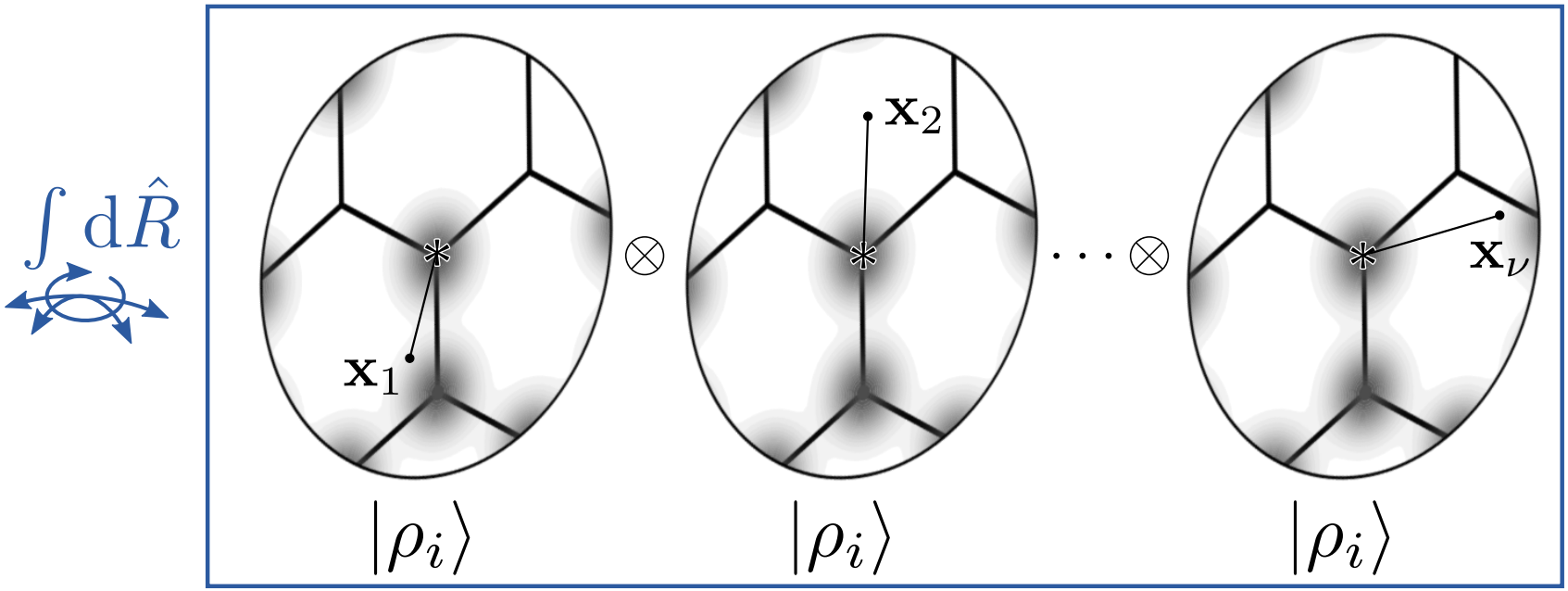
$$\langle a_1 r_1; a_2 r_2; \omega | \overline{\rho_i^{\otimes 2}} \rangle = \int \mathrm{d}\hat{R} \langle a_1 r_1 \hat{\mathbf{r}} | \hat{R} | \rho_i \rangle \langle a_2 r_2 \hat{\mathbf{r}}'(\omega) | \hat{R} | \rho_i \rangle$$

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

A universal feature construction

- Rotationally-averaged representations are essentially the same *n*-body correlations that are used in statistical theories of liquids
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

$$V(A_i) = \sum_{ij} V^{(2)}(r_{ij}) + \sum_{ij} V^{(3)}(r_{ij}, r_{ik}, \omega_{ijk}) \dots$$

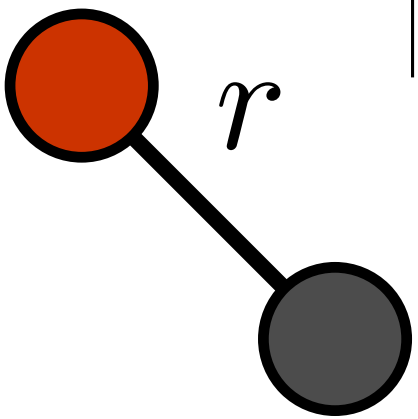


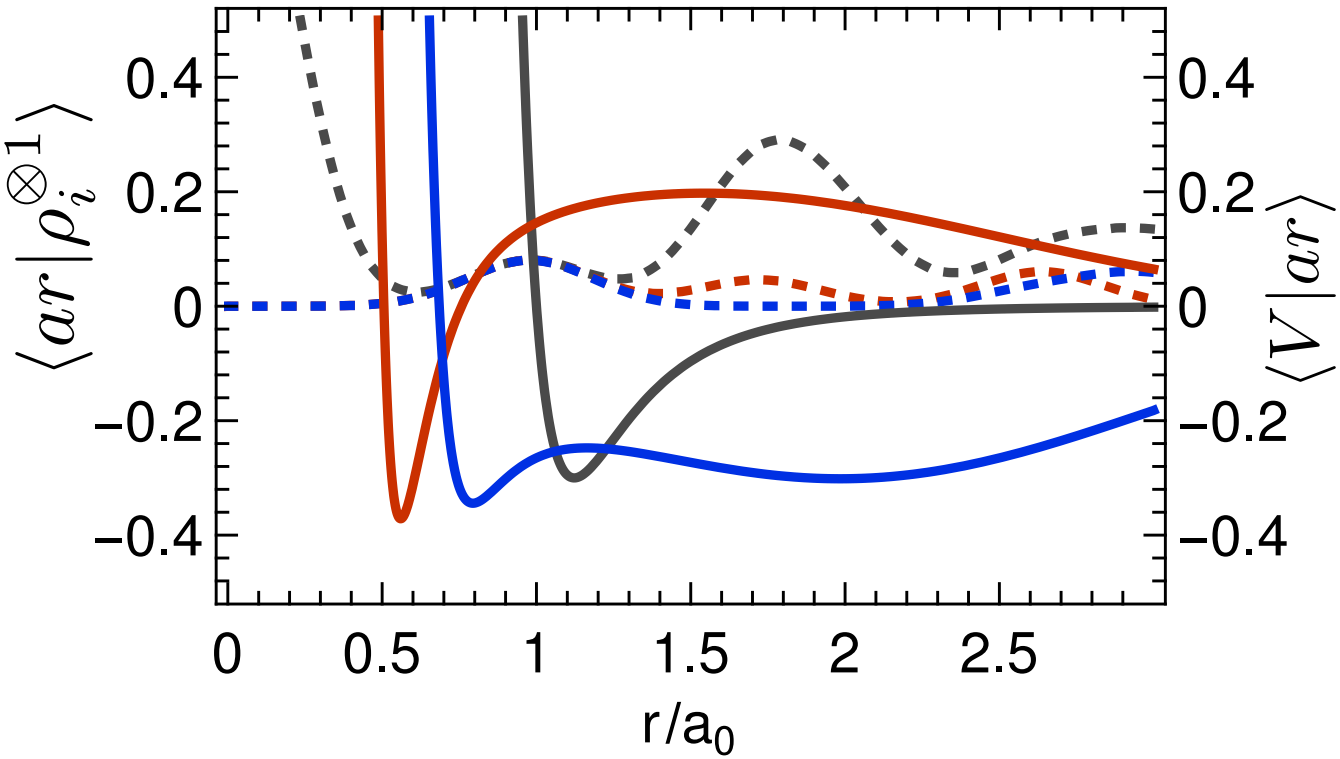
Willatt, Musil, **MC**, JCP (2019); Bartók, Kondor, Csányi PRB 2013

A universal feature construction

- Rotationally-averaged representations are essentially the same *n*-body correlations that are used in statistical theories of liquids
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

$$V(A_i) = \sum_{ij} V^{(2)}\left(r_{ij}\right) + \sum_{ij} V^{(3)}\left(r_{ij}, r_{ik}, \omega_{ijk}\right) \ldots$$





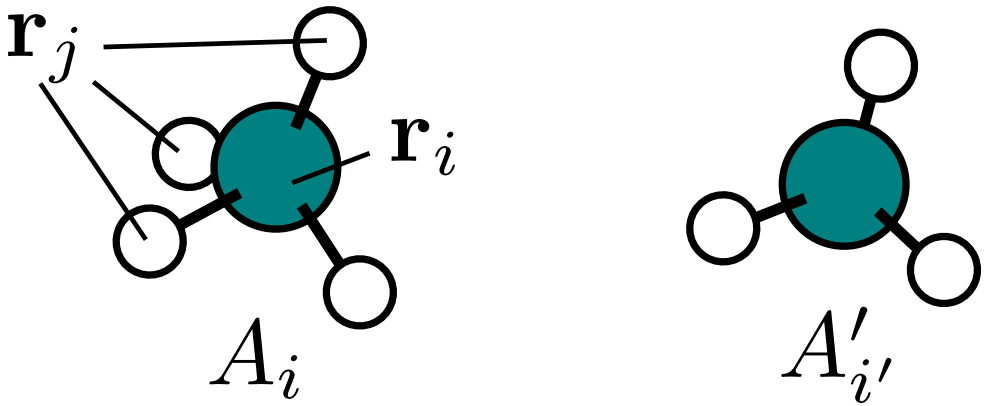
$$V(A_i) = \int \mathrm{d}r \langle V|ar\rangle\langle ar|\overline{\rho_i^{\otimes 1}}\rangle \approx \sum_j V_a(r_{ij})$$

Willatt, Musil, **MC**, JCP (2019); Drautz, PRB (2019); Glielmo, Zeni, De Vita, PRB (2018)

Smooth overlap of atomic positions: a worked example

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!

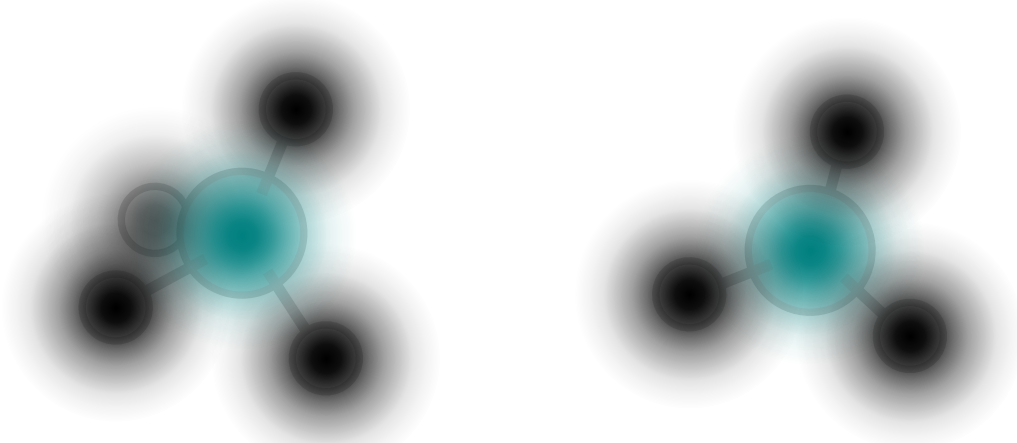


$$\{\mathbf{r}_{ji} = \mathbf{r}_j - \mathbf{r}_i\} \leftrightarrow A_i$$

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!

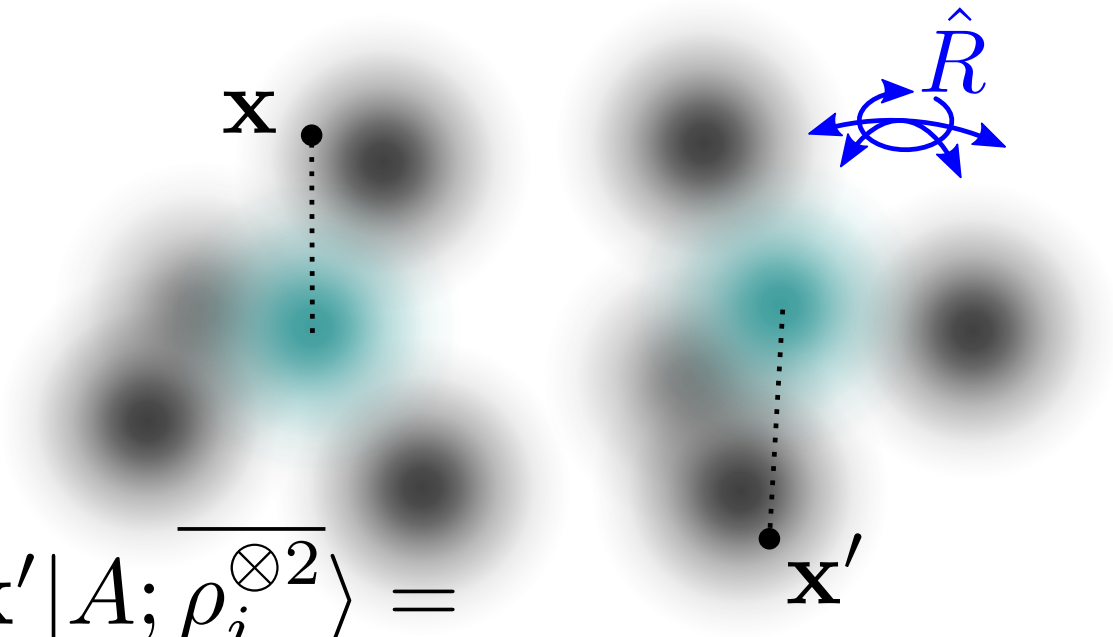


$$\langle a\mathbf{x}|\rho_i\rangle = \sum_{j\in A_i} \delta_{aa_j} \langle \mathbf{x}|\mathbf{r}_{ji}; g\rangle$$
$$\langle \mathbf{x}|\mathbf{r}_{ji}; g\rangle \equiv g(\mathbf{x} - \mathbf{r}_{ji})$$

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!

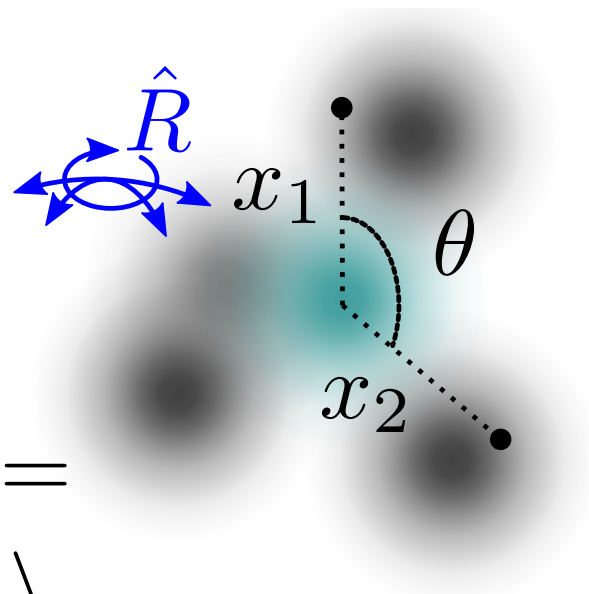


$$\langle \mathbf{x}; \mathbf{x}' | A; \overline{\rho_i^{\otimes 2}} \rangle = \int \mathrm{d}\hat{R} \langle \mathbf{x} | \hat{R}A; \rho_i \rangle \langle \mathbf{x}' | \hat{R}A; \rho_i \rangle$$

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!



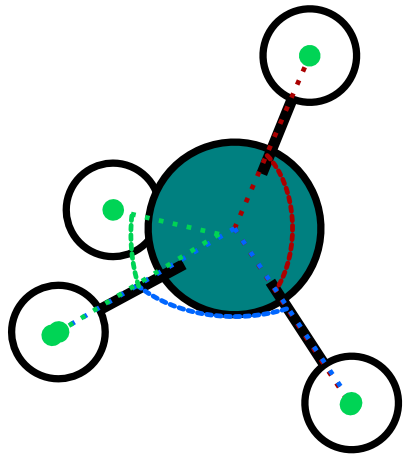
$$\langle x_1; x_2; \theta | A; \overline{\rho_i^{\otimes 2}} \rangle =$$
$$\int \mathrm{d}\hat{R} \langle x_1 \hat{R} \hat{\mathbf{e}}_z | A; \rho_i \rangle$$
$$\langle x_2 \hat{R} (\hat{\mathbf{e}}_z \cos\theta + \hat{\mathbf{e}}_x \sin\theta) | A; \rho_i \rangle$$

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!

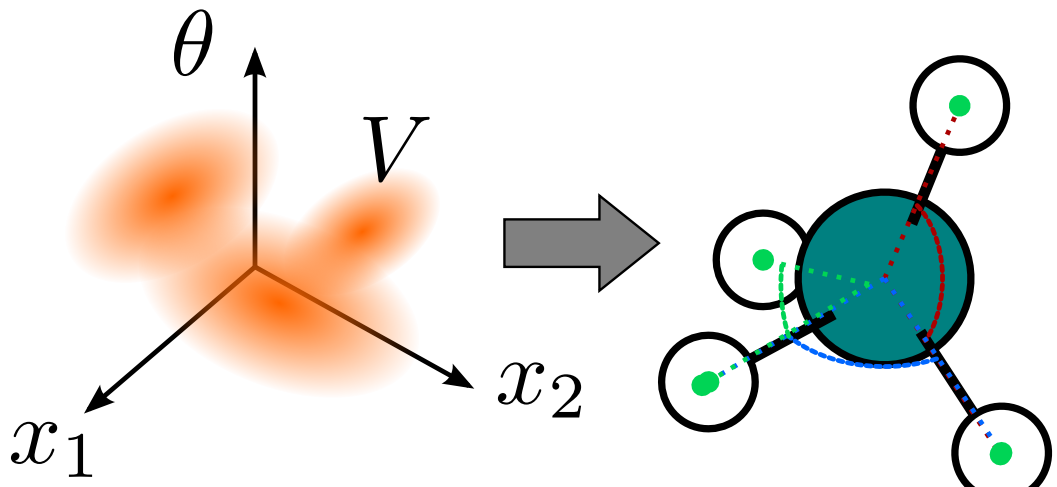
$$\langle x_1; x_2; \theta | A; \overline{\delta_i^{\otimes 2}} \rangle = \sum_{j_1 j_2} \delta(x_1 - r_{j_1 i}) \delta(x_2 - r_{j_2 i}) \delta(\cos\theta - \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i})$$



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!

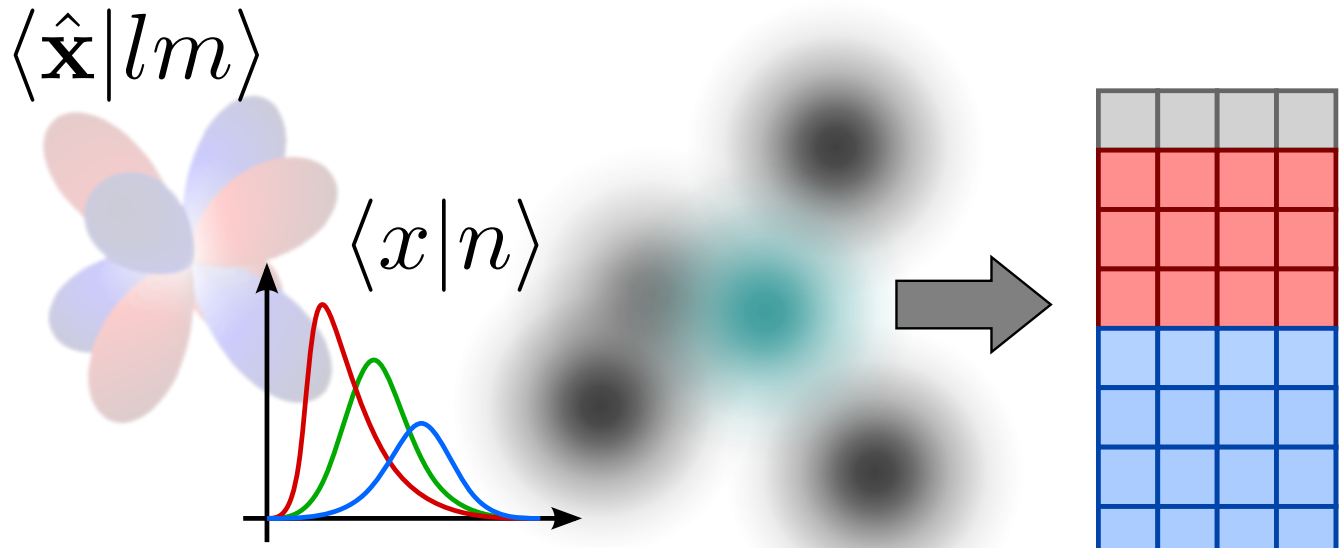


$$\int \langle V | x_1; x_2; \theta \rangle \langle x_1; x_2; \theta | A; \overline{\delta_i^{\otimes 2}} \rangle = \sum_{j_1 j_2} V\left(r_{j_1 i}, r_{j_2 i}, \hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i}\right)$$

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

Density trick in an $\langle nlm|$ basis

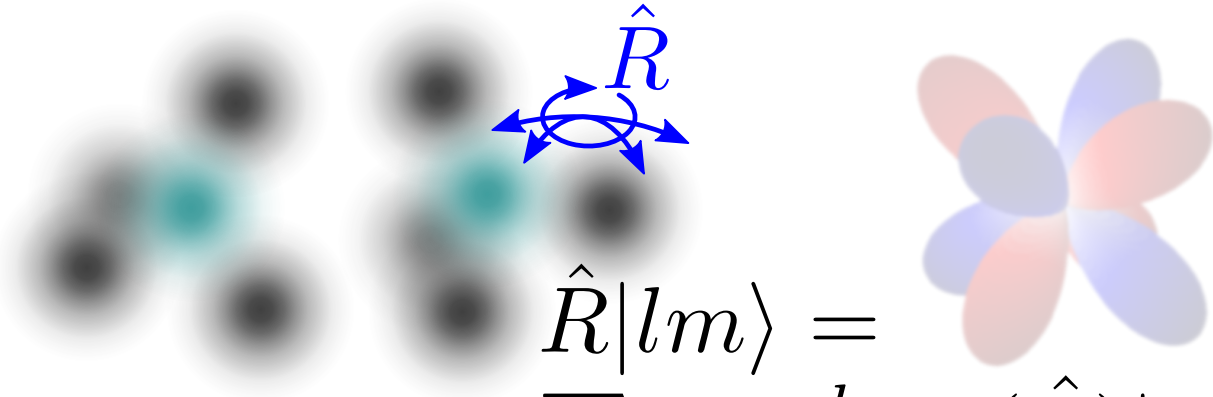
- The symmetrized correlations can be computed in closed form using a discrete basis
  - The neighbor density can be expanded on a basis of radial functions $\langle x|n\rangle \equiv R_n(x)$ and spherical harmonics $\langle \hat{\mathbf{x}}|lm\rangle \equiv Y_l^m(\hat{\mathbf{x}})$
  - Spherical harmonics 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



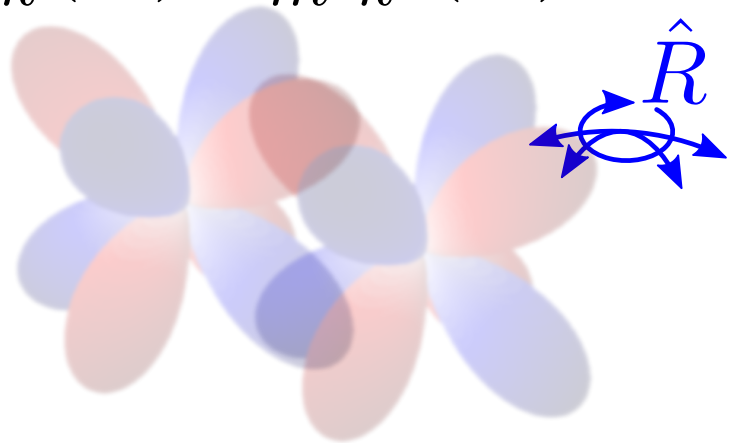
$$\hat{R}|lm\rangle = \sum_{m'} D^l_{mm'}(\hat{R})|lm'\rangle$$

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

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

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



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

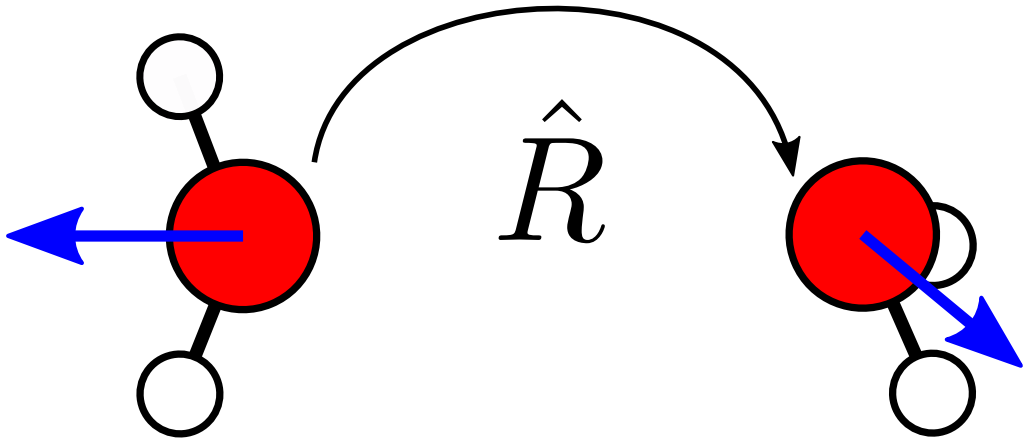
There are more things in heaven and earth, Horatio, than those transforming like a scalar

Machine-learning for tensors

- What if we want to learn vectors or general tensors? We need features that are *equivariant* to the tensor under rotations.

$$d_\alpha\left(A_i\right) = \int \mathrm{d}Q \, \langle d|Q\rangle \, \langle Q|A; \overline{\rho_i^{\otimes\nu}; \alpha}\rangle$$

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



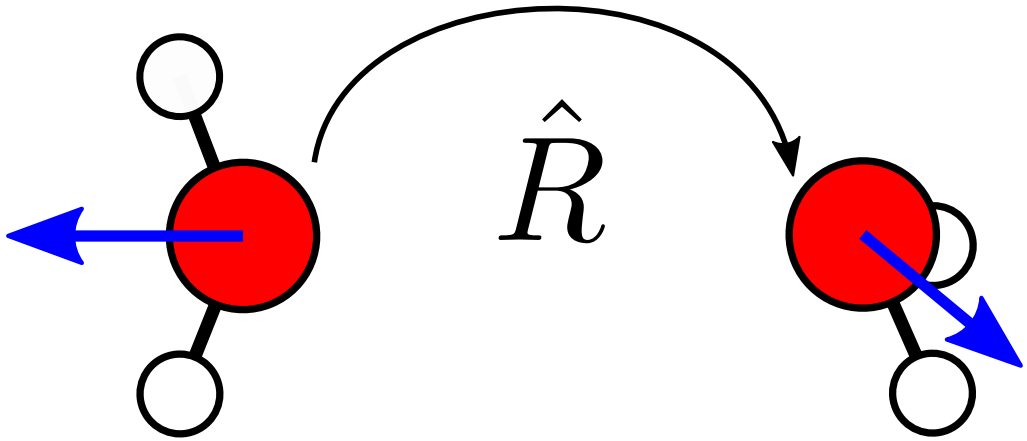
Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018); Veit et al., JCP (2020)

Machine-learning for tensors

- What if we want to learn vectors or general tensors? We need features that are *equivariant* to the tensor under rotations.

$$d_\alpha\left(A_i\right) = \int \mathrm{d}Q \, \langle d|Q\rangle \, \langle Q|A; \overline{\rho_i^{\otimes\nu}; \alpha}\rangle$$

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



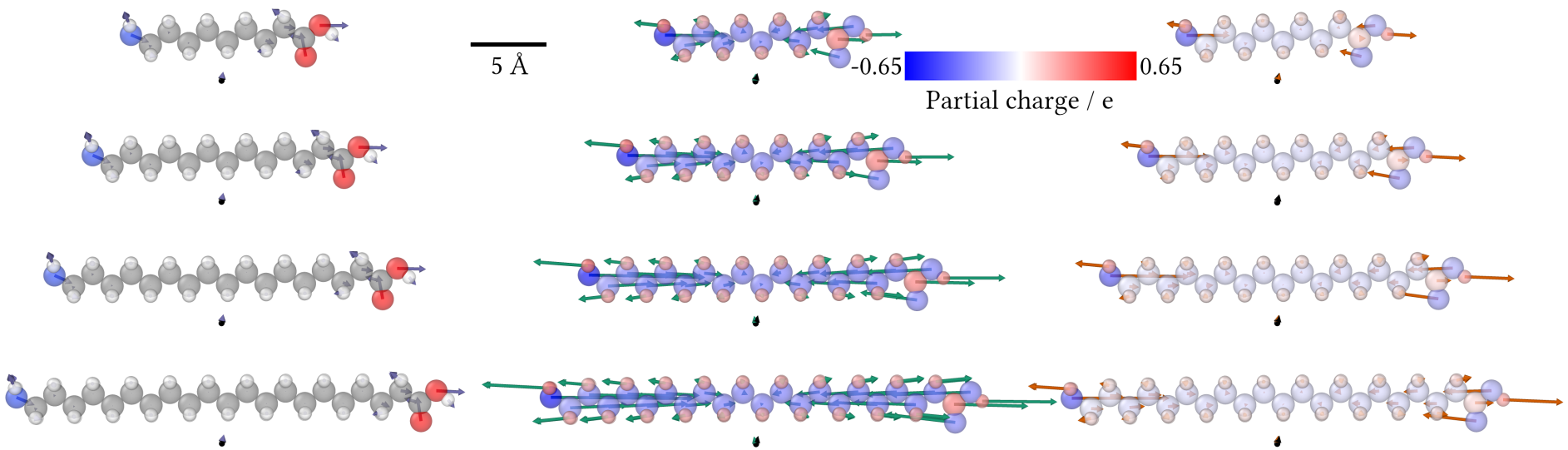
Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018); Veit et al., JCP (2020)

Machine-learning for tensors

- What if we want to learn vectors or general tensors? We need features that are *equivariant* to the tensor under rotations.

$$d_\alpha\left(A_i\right) = \int \mathrm{d}Q \, \langle d|Q\rangle \, \langle Q|A; \overline{\rho_i^{\otimes\nu}; \alpha}\rangle$$

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



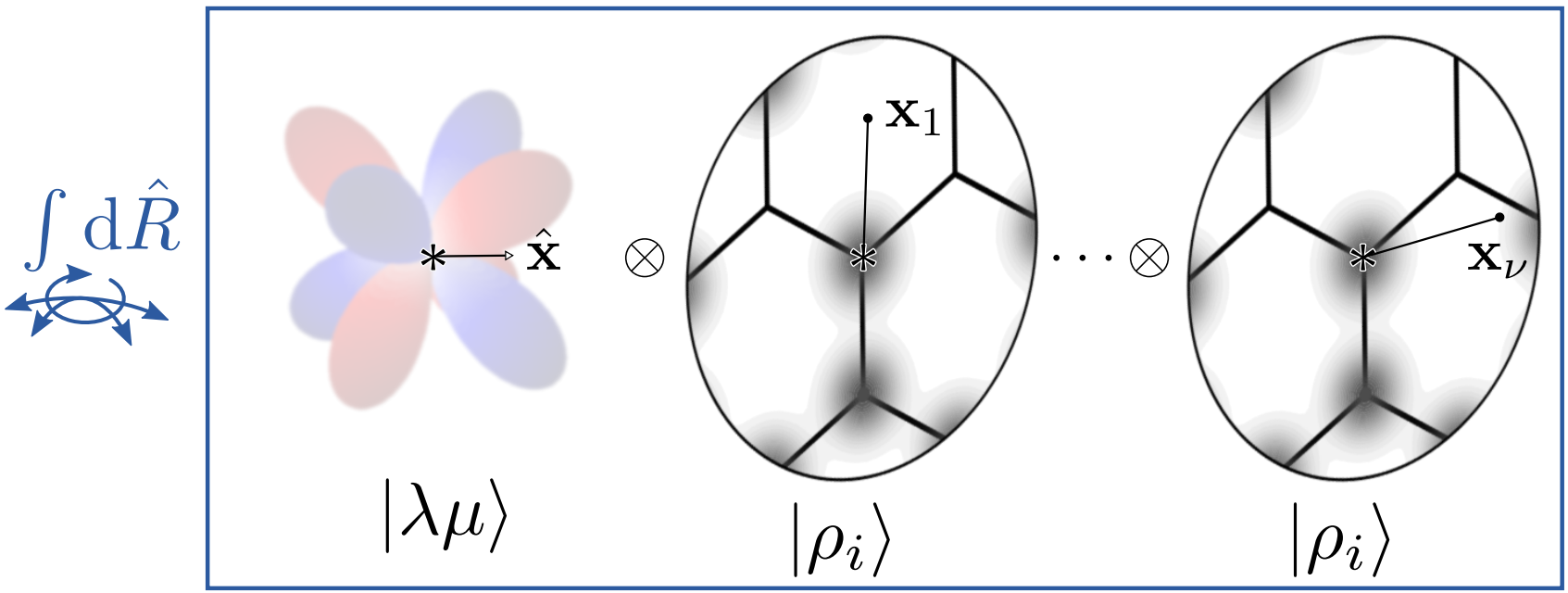
Veit, Wilkins, Yang, DiStasio, **MC**, JCP (2020)

Machine-learning for tensors

- What if we want to learn vectors or general tensors? We need features that are *equivariant* to the tensor under rotations.

$$d_\alpha\left(A_i\right) = \int \mathrm{d}Q \, \langle d|Q\rangle \, \langle Q|A; \overline{\rho_i^{\otimes\nu}}; \alpha\rangle$$

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



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

A hierarchy of equivariant features

- Equivariant *N*-body features transform like angular momenta

$$\langle Q|\hat{R}A;\overline{\rho_i^{\otimes\nu};\lambda\mu}\rangle \sim \sum_{\mu'} D^{\lambda}_{\mu\mu'}(R)\ \langle Q|A;\overline{\rho_i^{\otimes\nu};\lambda\mu'}\rangle$$

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

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

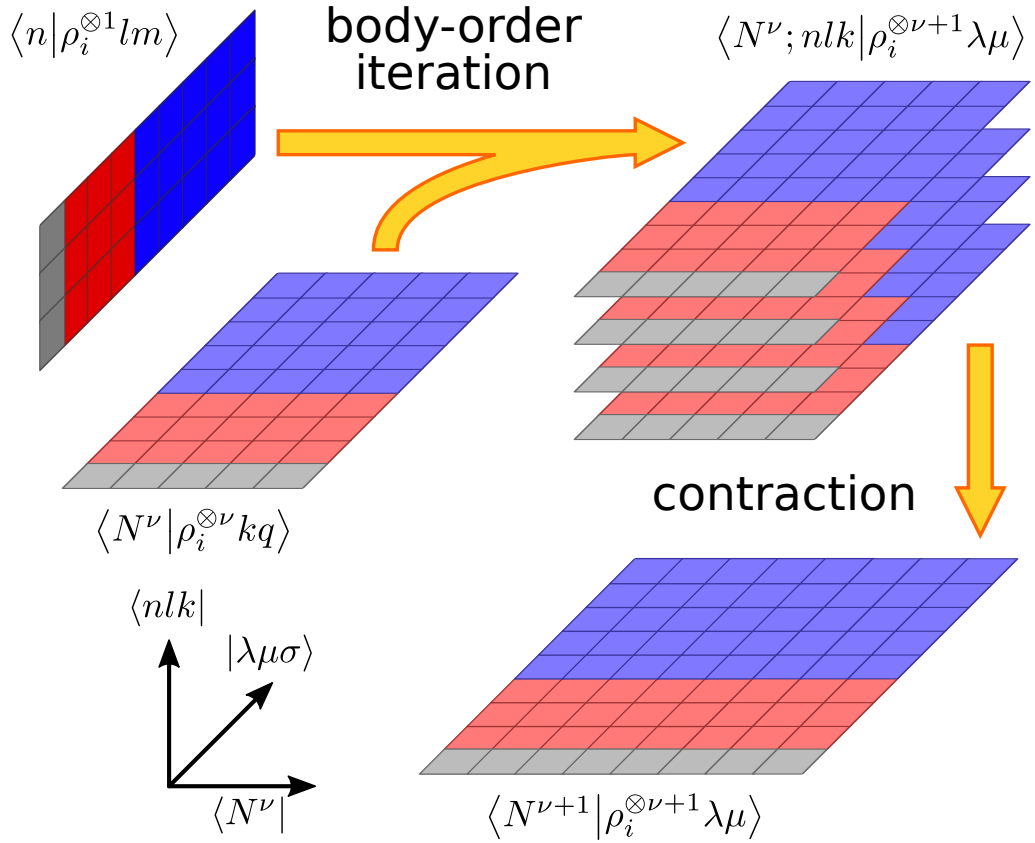
$$\langle \ldots;n_\nu l_\nu k_\nu;nlk|\overline{\rho_i^{\otimes(\nu+1)};\lambda\mu}\rangle = \sum_{qm}\langle n|\overline{\rho_i^{\otimes 1};lm}\rangle\ \langle \ldots;n_\nu l_\nu k_\nu|\overline{\rho_i^{\otimes\nu};kq}\rangle\ \langle lm;kq|\lambda\mu\rangle$$

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

Nigam, Pozdnyakov, **MC**, JCP (2020); `https://github.com/cosmo-epfl/nice`

NICE features for ML

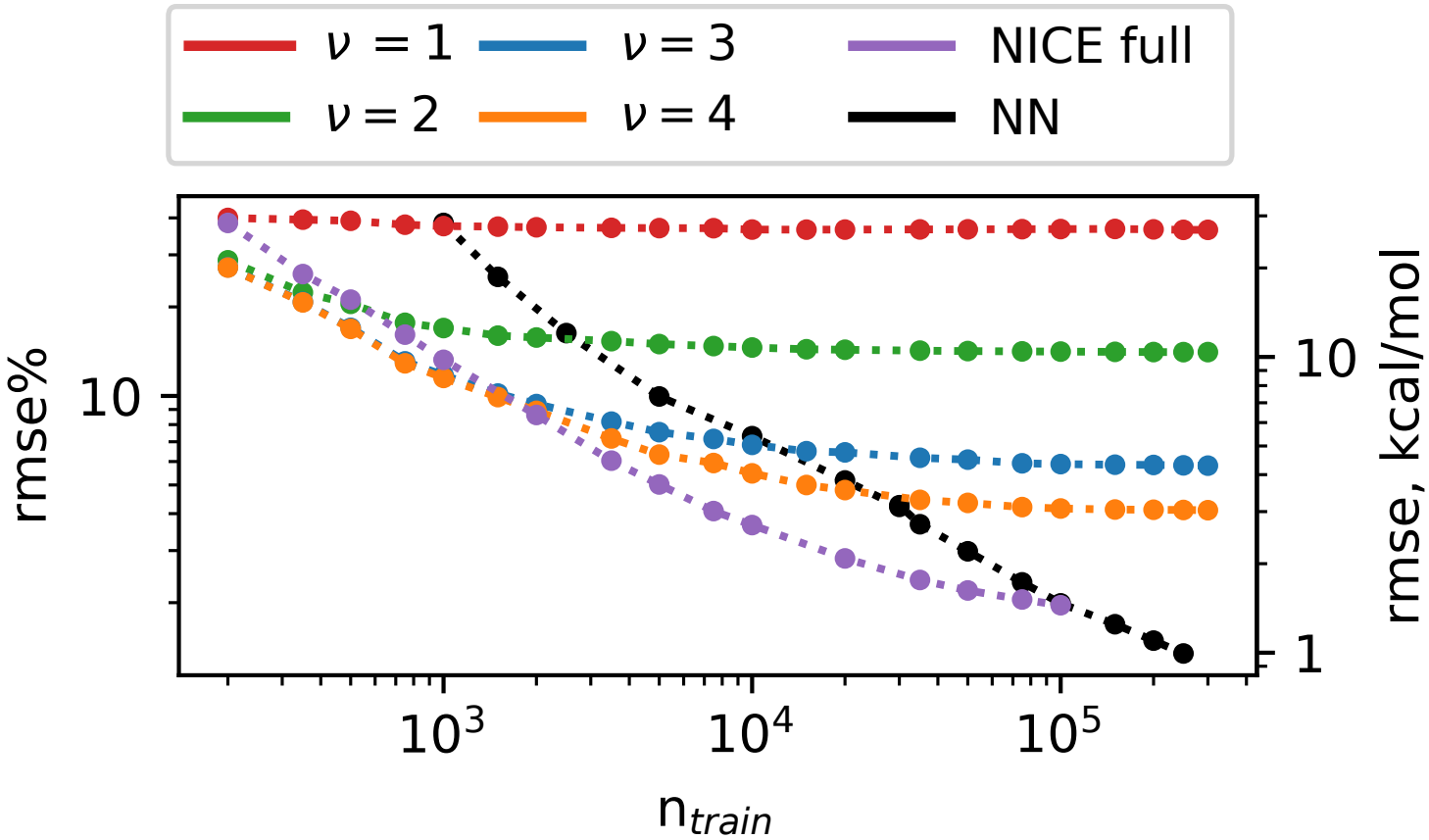
- Problem: number of features grows exponentially with ν
- Solution: *N*-body iterative contraction of equivariants (NICE)
  - After each body order increase, the most relevant features are selected and used for the next iteration
  - Systematic convergence with ν and contraction truncation



Nigam, Pozdnyakov, **MC**, JCP (2020); `https://github.com/cosmo-epfl/nice`

NICE features for ML

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



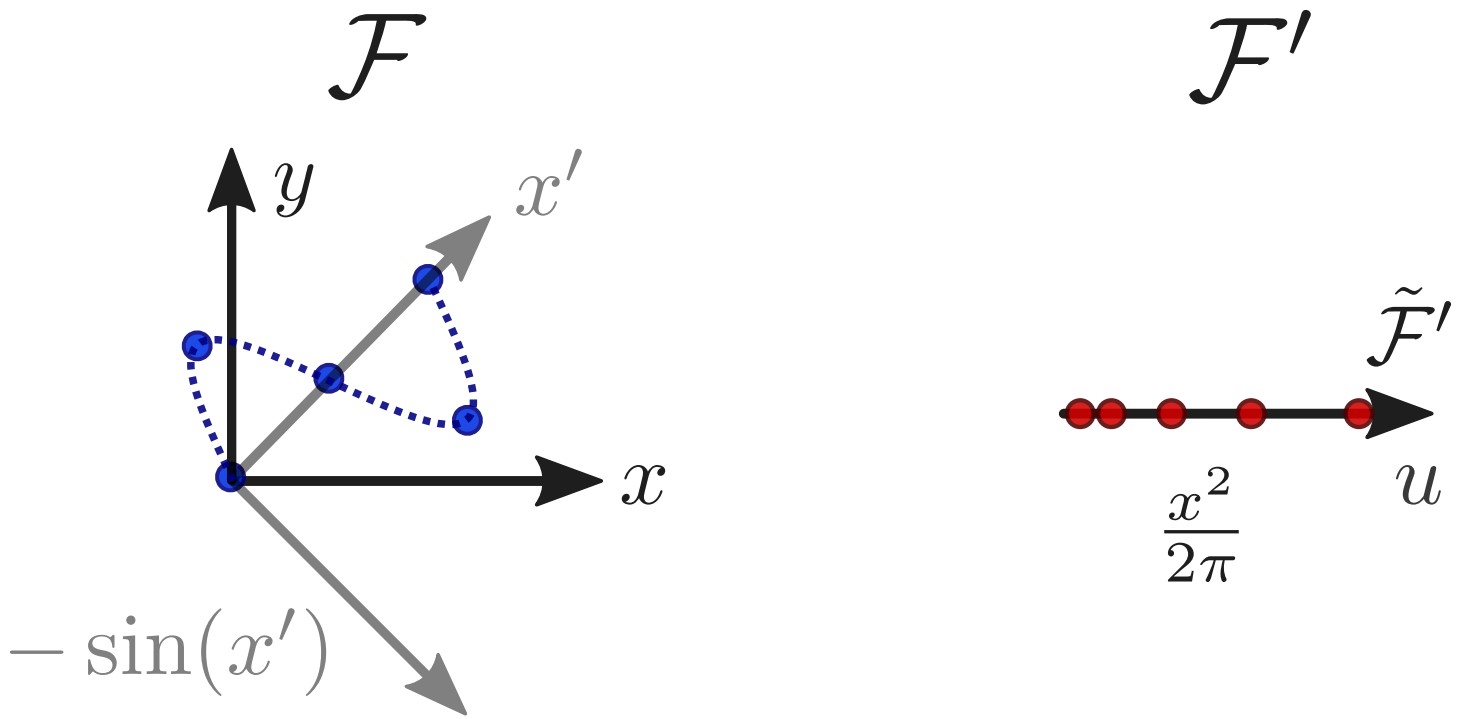
Nigam, Pozdnyakov, **MC**, JCP (2020); `https://github.com/cosmo-epfl/nice`

Comparing representations

Measuring feature spaces

- General problem: how do we compare information content of different choices of features? How to compare metrics and kernels?
- Feature space Reconstruction Error (**FRE**) measures the linearly-embeddable mutual information. Locally-linear and kernelized extensions also available

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

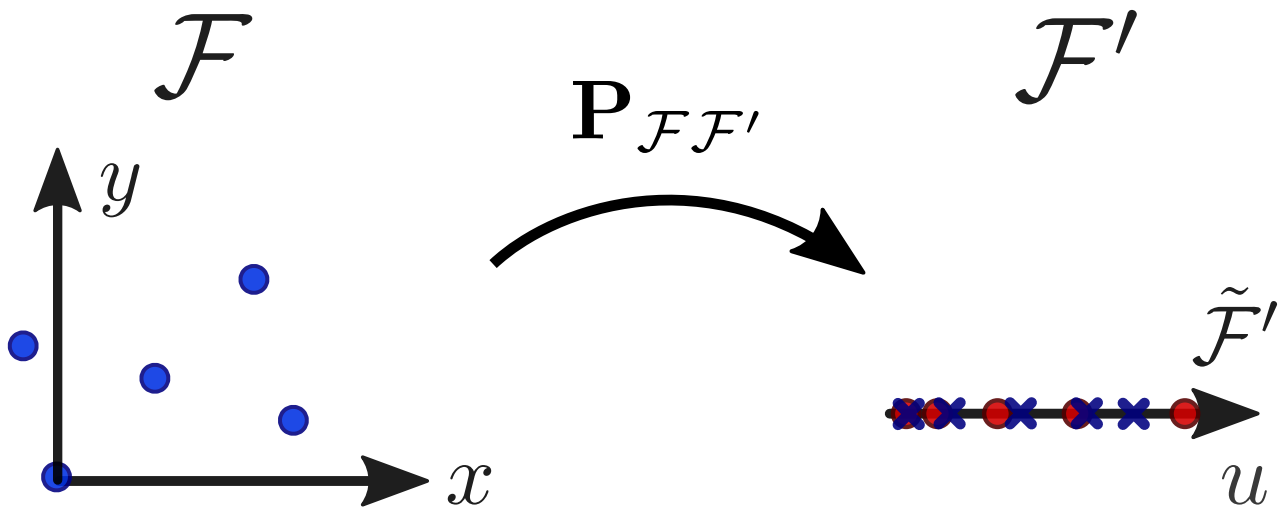


Goscinski, Fraux, **MC**, MLST (2021)

Measuring feature spaces

- General problem: how do we compare information content of different choices of features? How to compare metrics and kernels?
- Feature space Reconstruction Error (**FRE**) measures the linearly-embeddable mutual information. Locally-linear and kernelized extensions also available

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



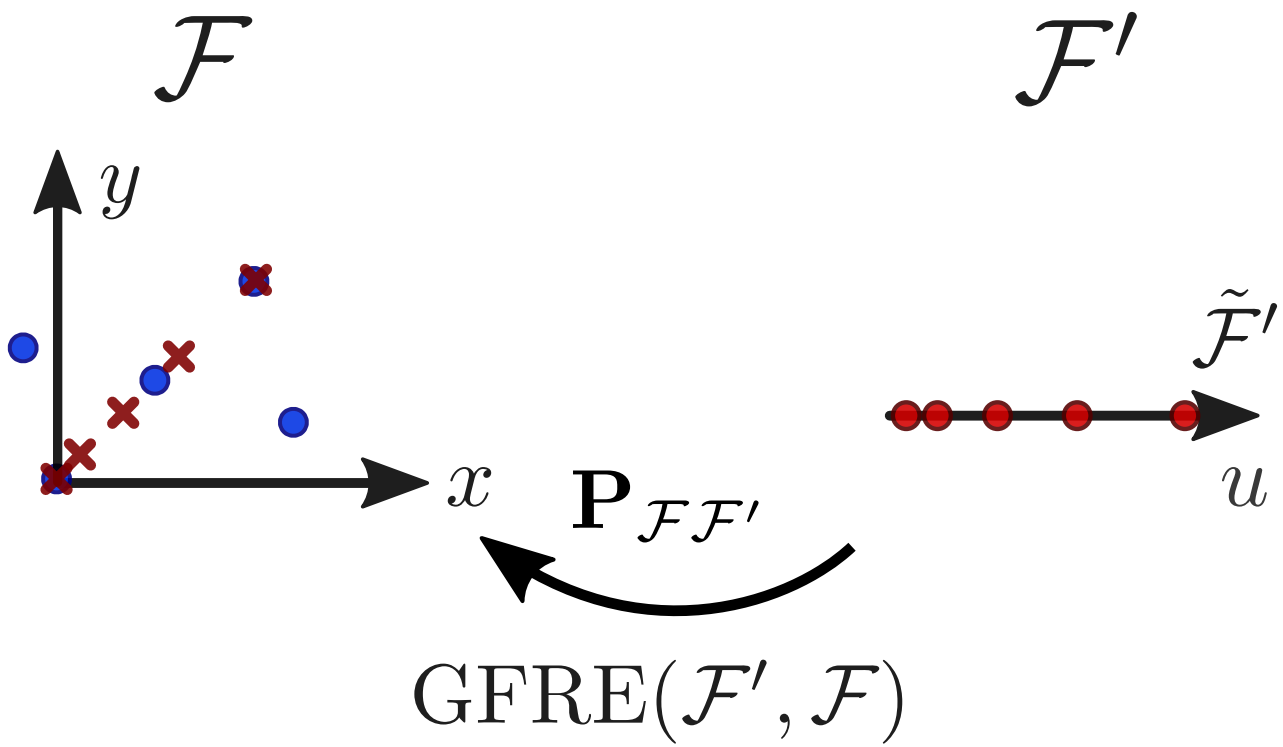
$$\mathrm{GFRE}(\mathcal{F}, \mathcal{F}')$$

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

Measuring feature spaces

- General problem: how do we compare information content of different choices of features? How to compare metrics and kernels?
- Feature space Reconstruction Error (**FRE**) measures the linearly-embeddable mutual information. Locally-linear and kernelized extensions also available

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

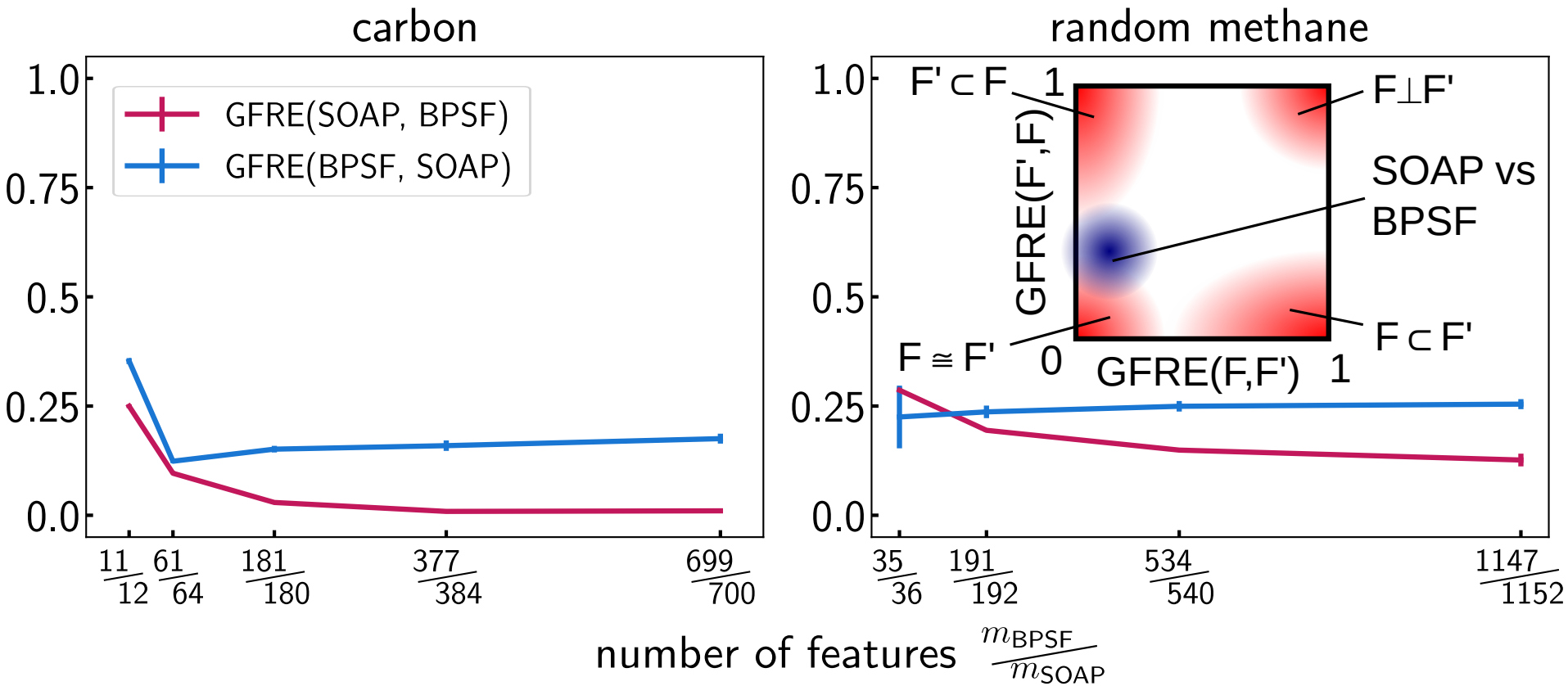


Goscinski, Fraux, **MC**, MLST (2021)

Measuring feature spaces

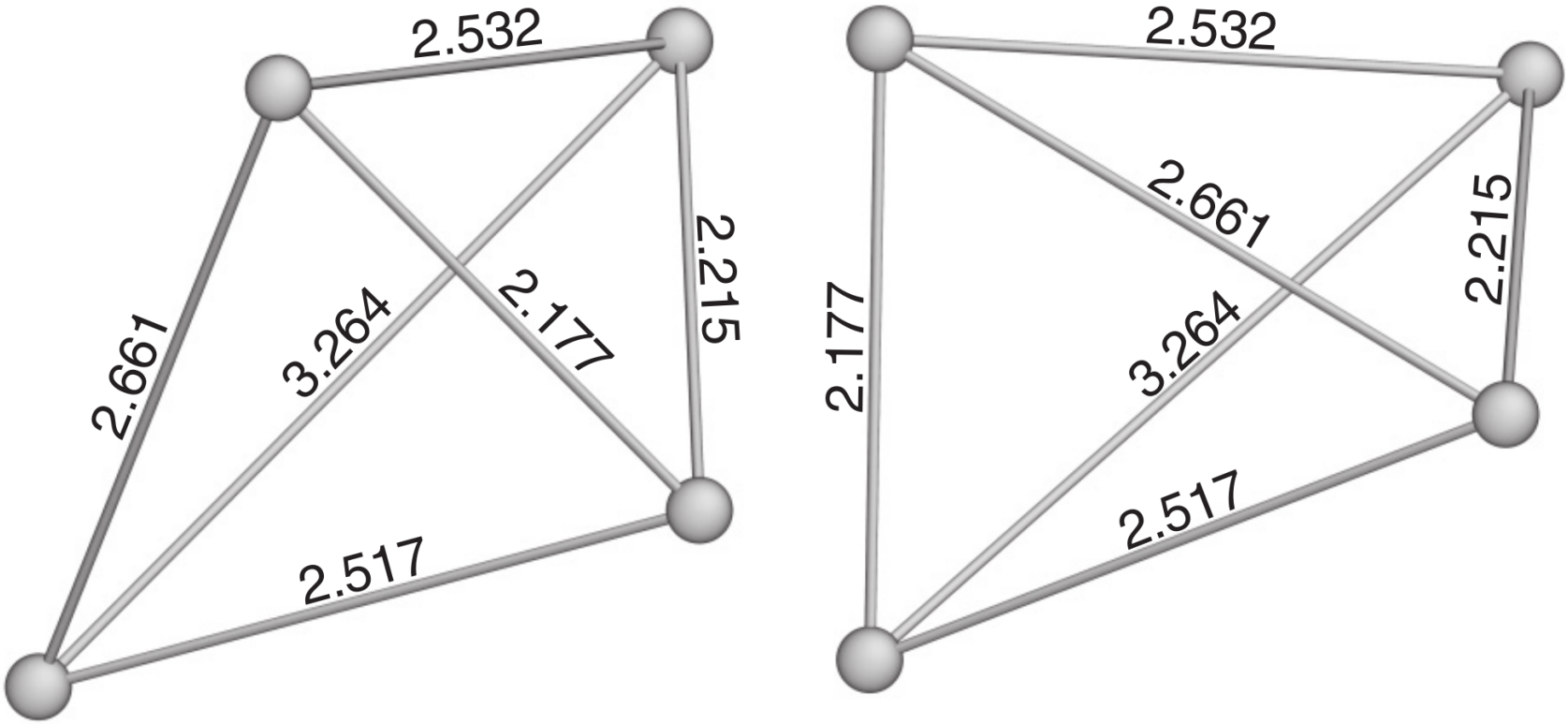
- General problem: how do we compare information content of different choices of features? How to compare metrics and kernels?
- Feature space Reconstruction Error (**FRE**) measures the linearly-embeddable mutual information. Locally-linear and kernelized extensions also available

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



Are these representations complete?

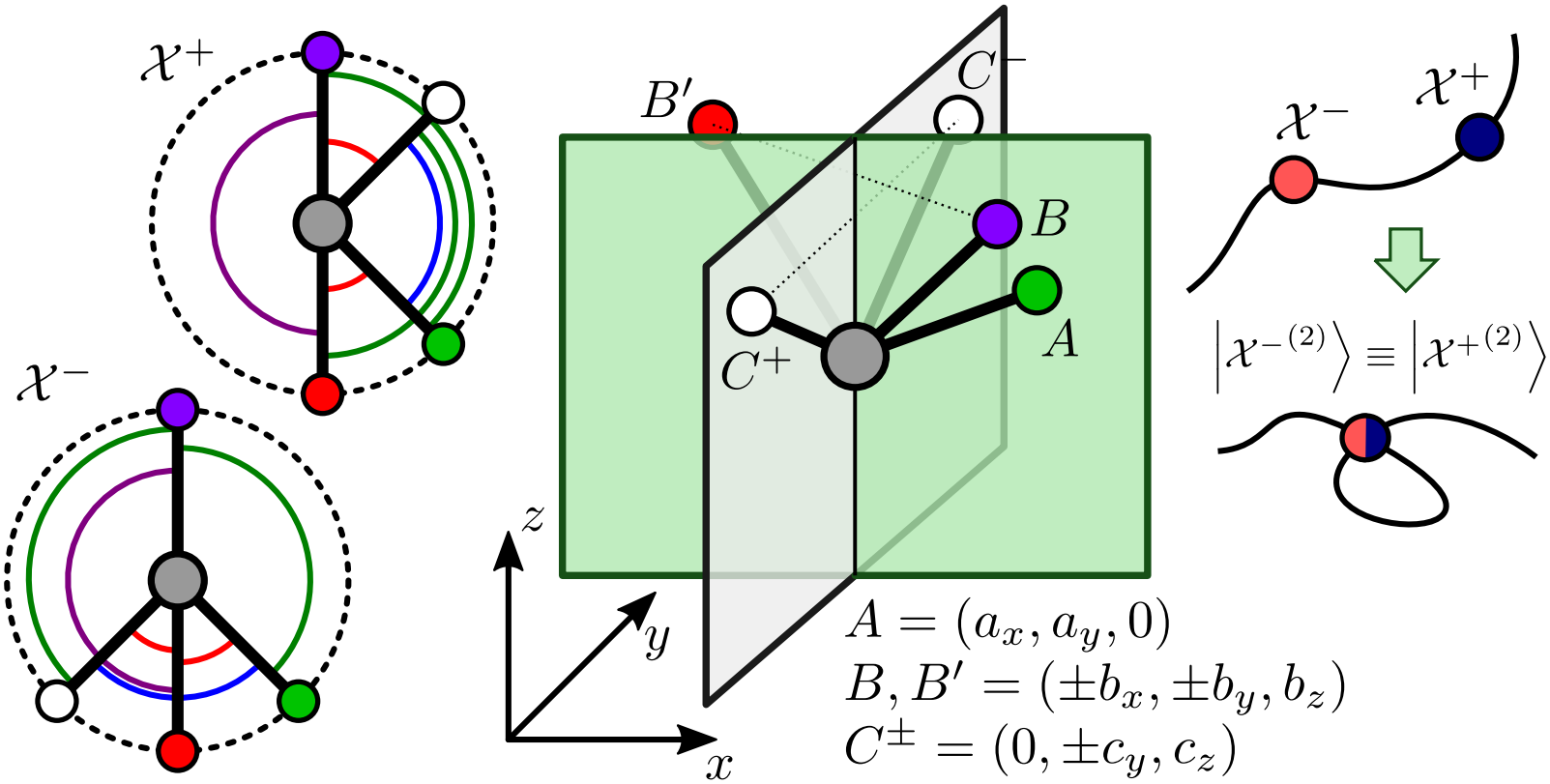
- It is well-known that 2-body correlations are ambiguous: one can build tetrahedra with same pair distances that are different
- Surprise: neither are 3 (and 4!!) body feature! Problem gets important as model accuracy is increased



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

Are these representations complete?

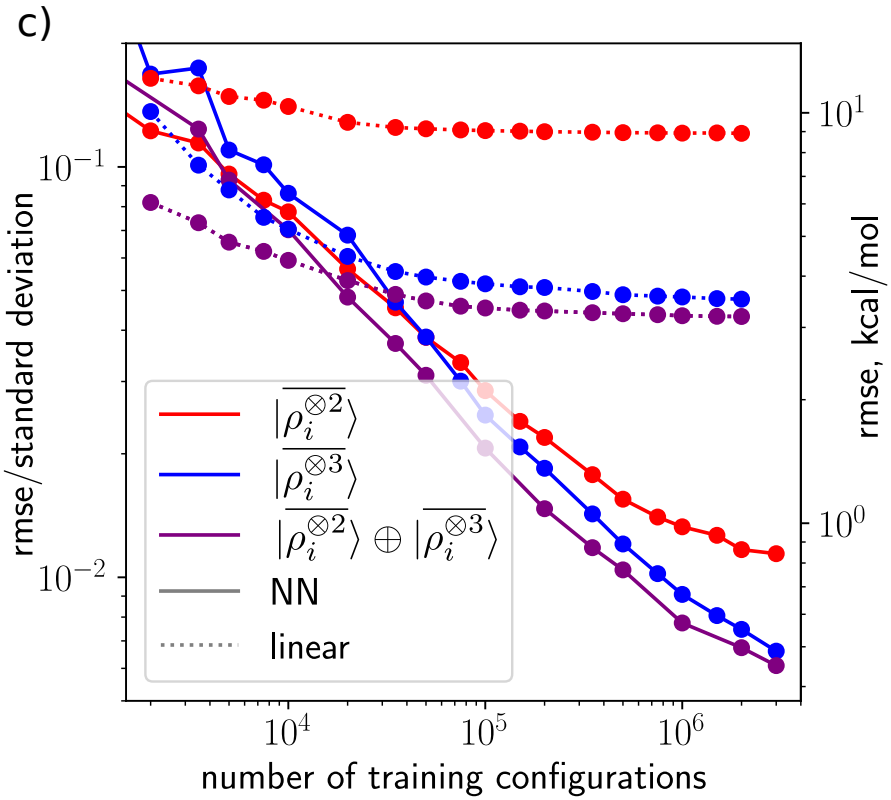
- It is well-known that 2-body correlations are ambiguous: one can build tetrahedra with same pair distances that are different
- Surprise: neither are 3 (and 4!!) body feature! Problem gets important as model accuracy is increased



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

Are these representations complete?

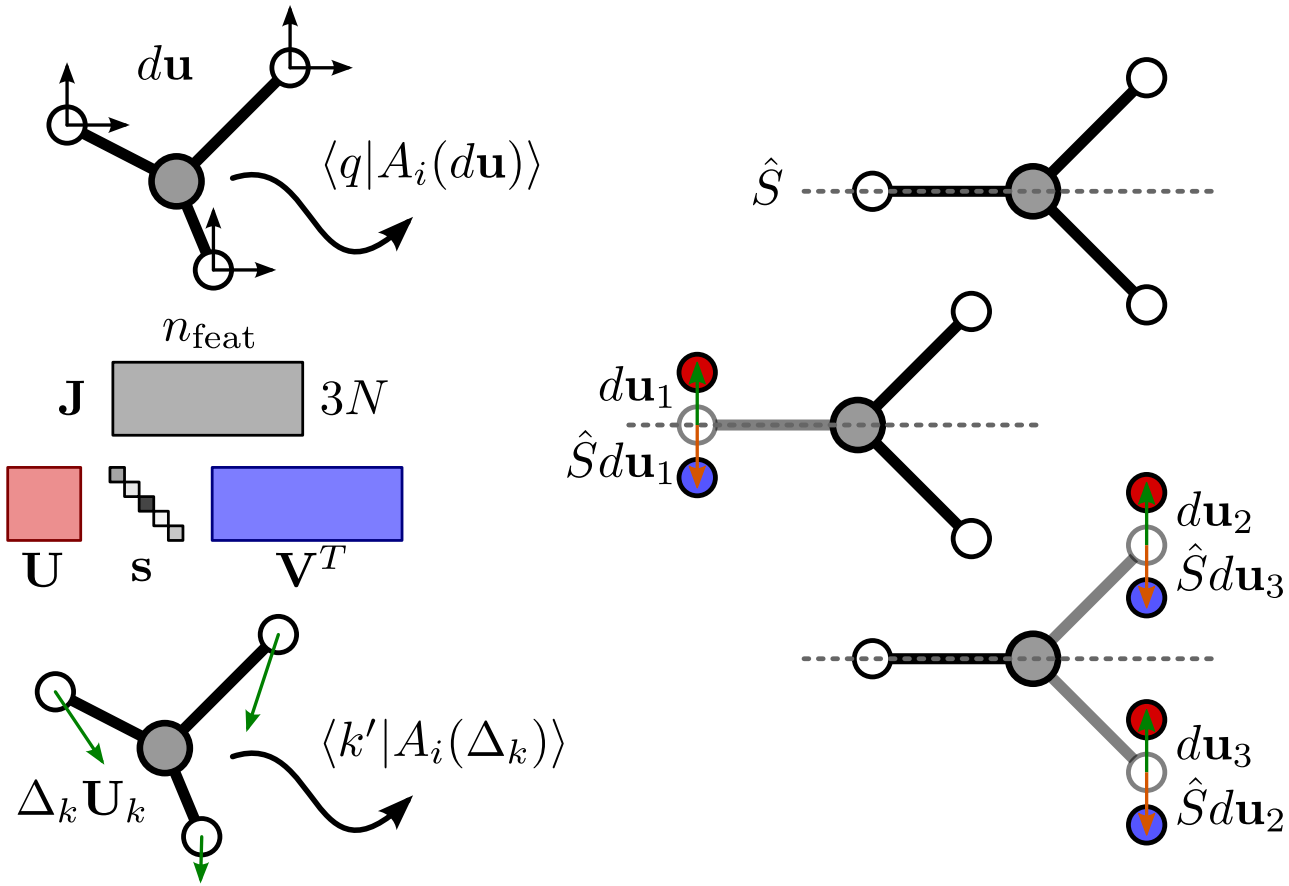
- It is well-known that 2-body correlations are ambiguous: one can build tetrahedra with same pair distances that are different
- Surprise: neither are 3 (and 4!!) body feature! Problem gets important as model accuracy is increased



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

Symmetry and sensitivity

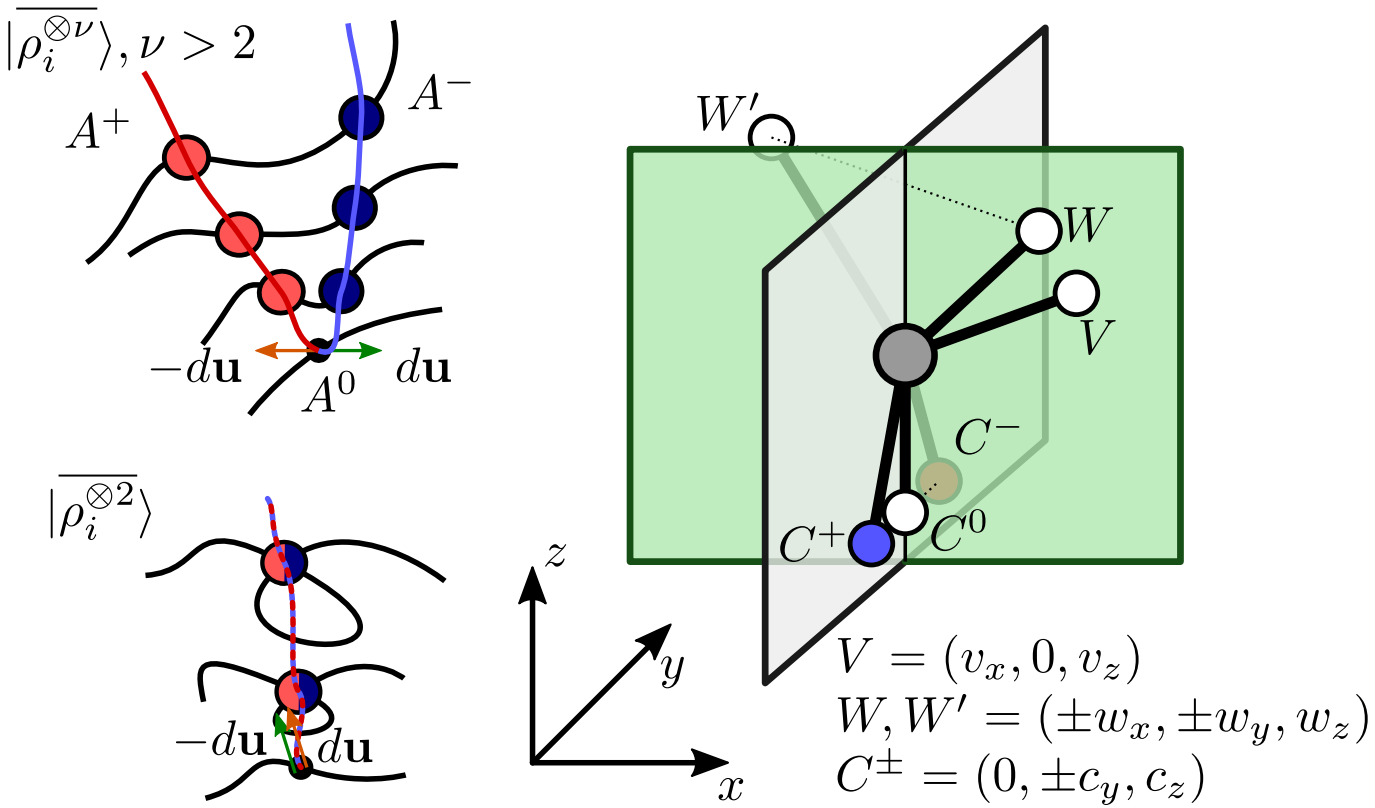
- *Local* characterization of features based on the SVD of the Jacobian
- Symmetric features must have low-rank **J** environments, connected with discrete symmetries
- Interesting twist: degenerate manifolds crossing → *spurious* singularities



Parsaeifard et al., MLST (2021); Pozdnyakov et al., Open Research Europe (2021)

Symmetry and sensitivity

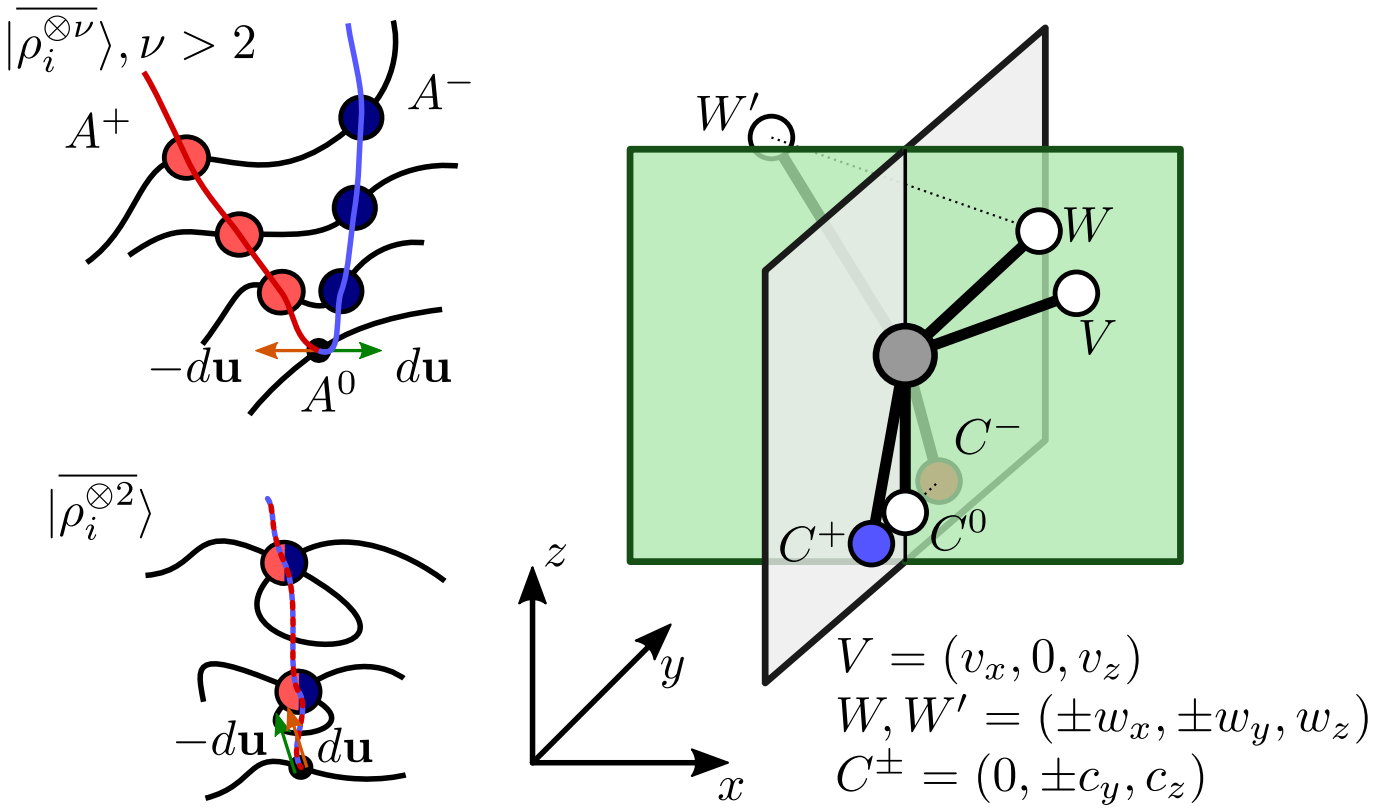
- *Local* characterization of features based on the SVD of the Jacobian
- Symmetric features must have low-rank **J** environments, connected with discrete symmetries
- Interesting twist: degenerate manifolds crossing → *spurious* singularities



Parsaeifard et al. Pozdnyakov et al., Open Research Europe (2021)

Symmetry and sensitivity

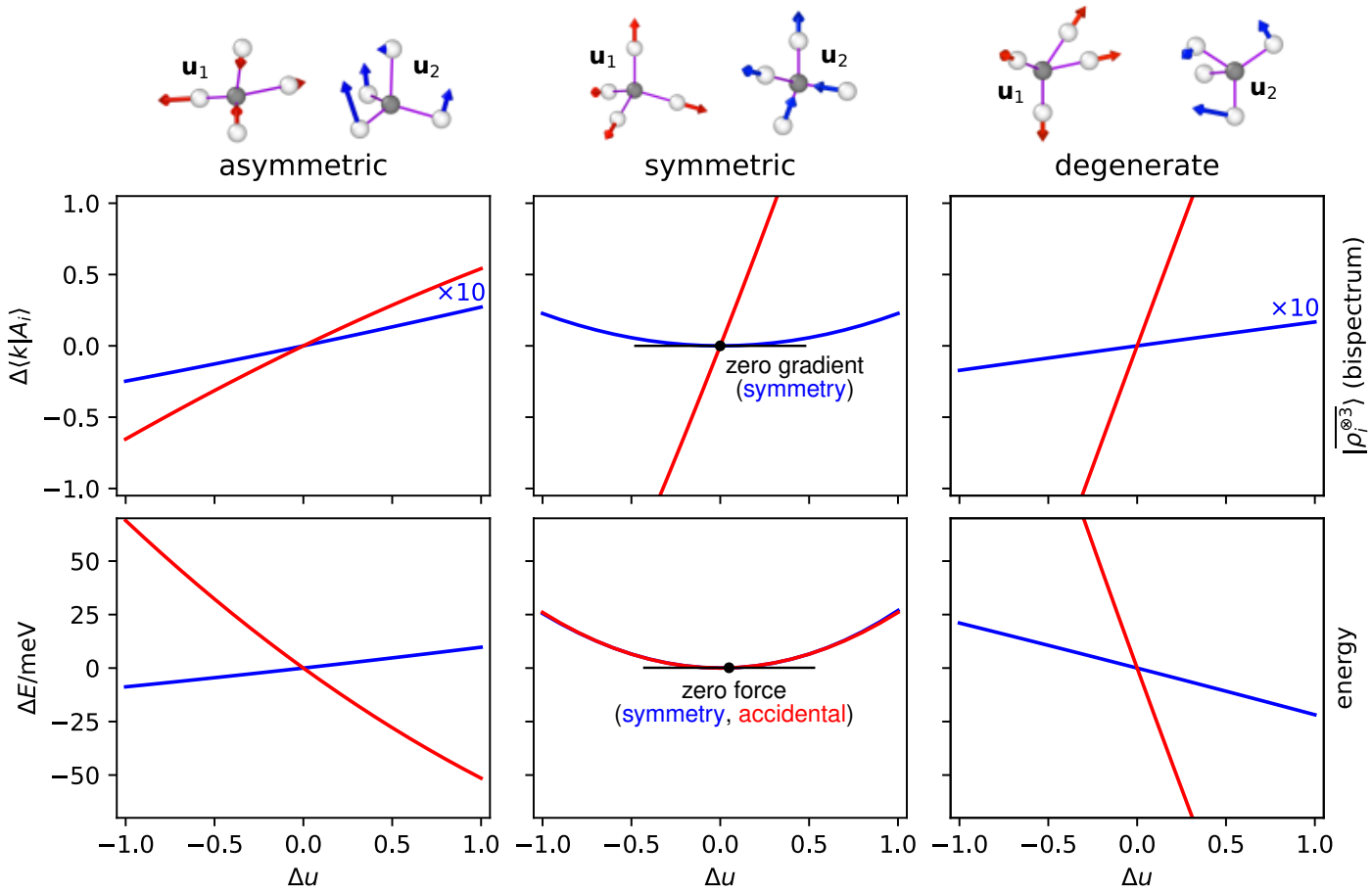
- *Local* characterization of features based on the SVD of the Jacobian
- Symmetric features must have low-rank **J** environments, connected with discrete symmetries
- Interesting twist: degenerate manifolds crossing → *spurious* singularities



Parsaeifard et al. Pozdnyakov et al., Open Research Europe (2021)

Symmetry and models

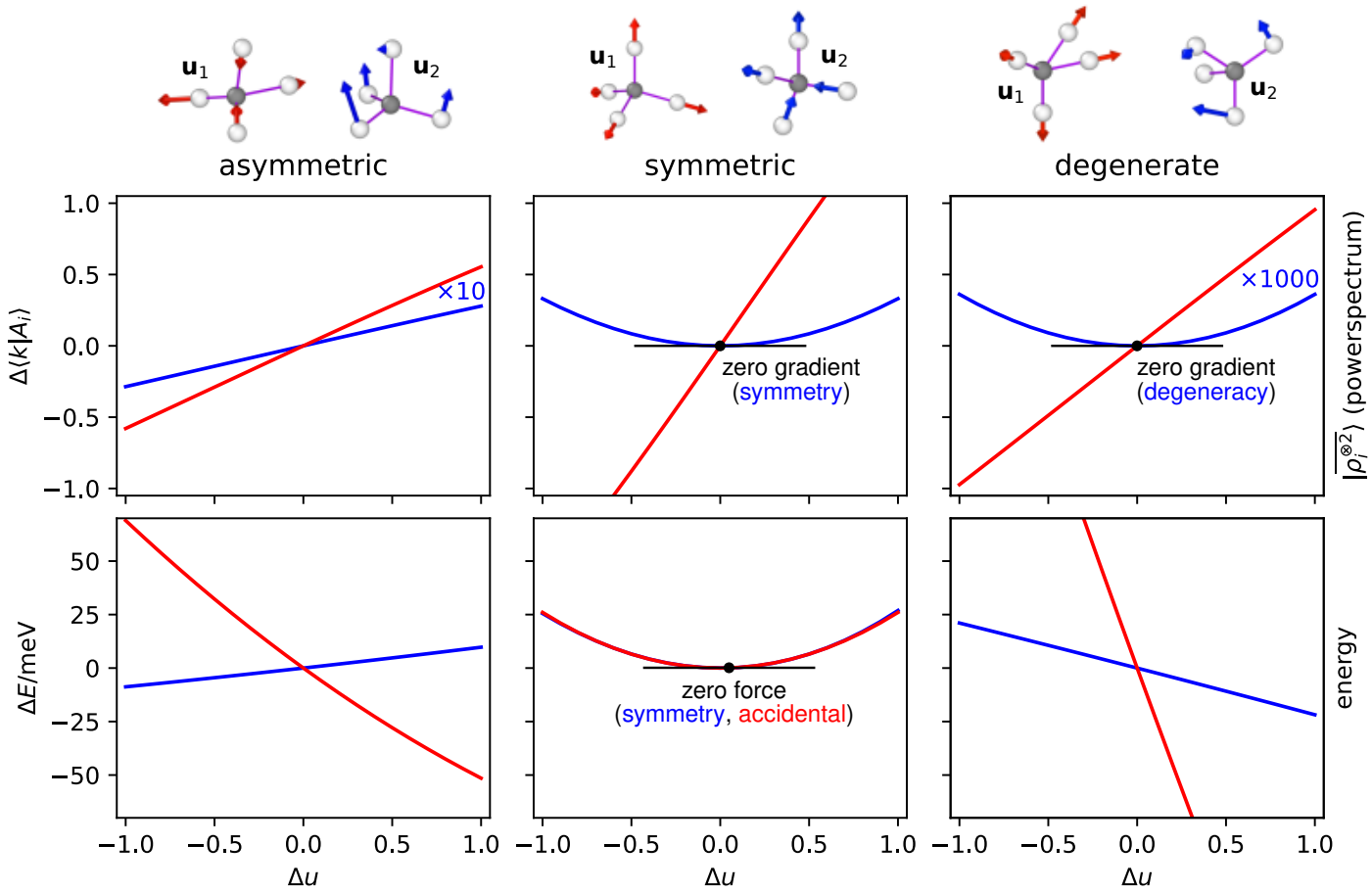
- Zeros in the Jacobian singular value spectrum enforce the correct extremal behavior for symmetric structures
- ... but for degenerate structures: artificial zero-force points!
- Deep consequences for equivariant models



Pozdnyakov et al., Open Research Europe (2021)

Symmetry and models

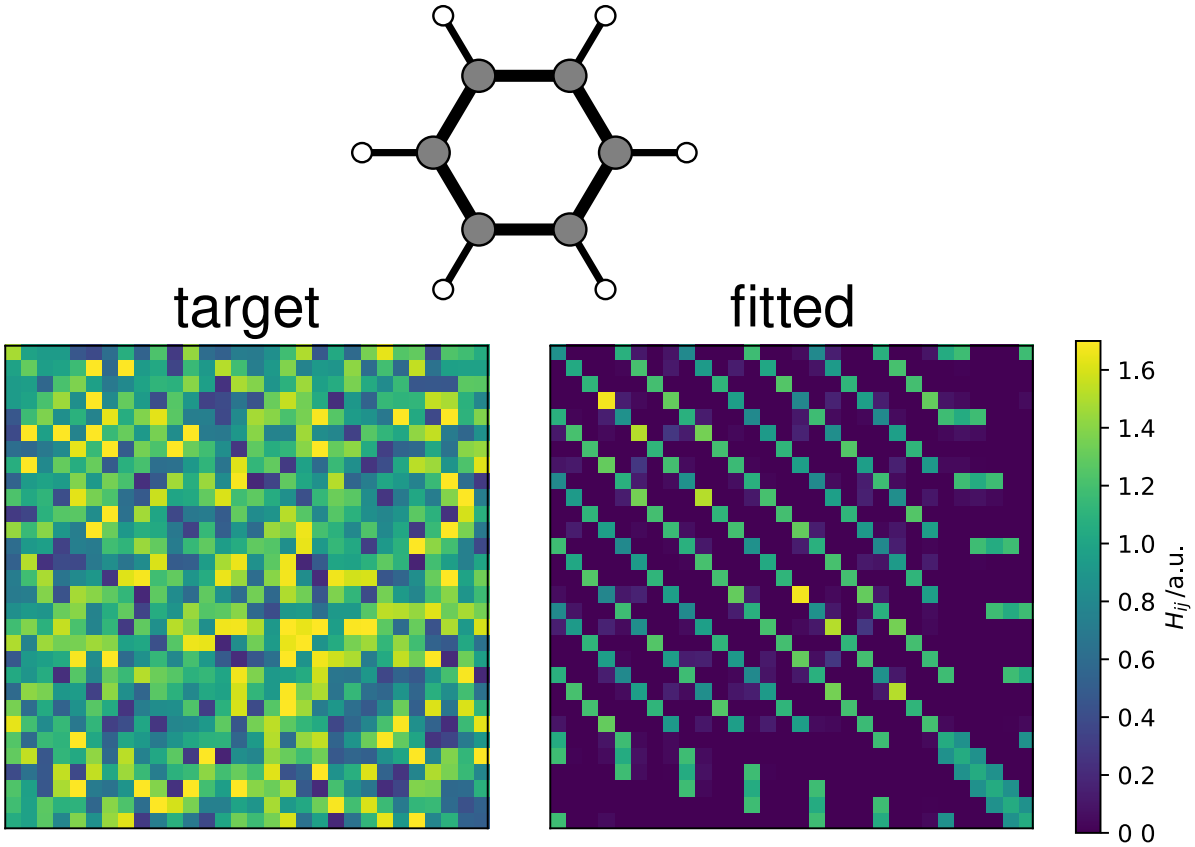
- Zeros in the Jacobian singular value spectrum enforce the correct extremal behavior for symmetric structures
- ... but for degenerate structures: artificial zero-force points!
- Deep consequences for equivariant models



Pozdnyakov et al., Open Research Europe (2021)

Symmetry and models

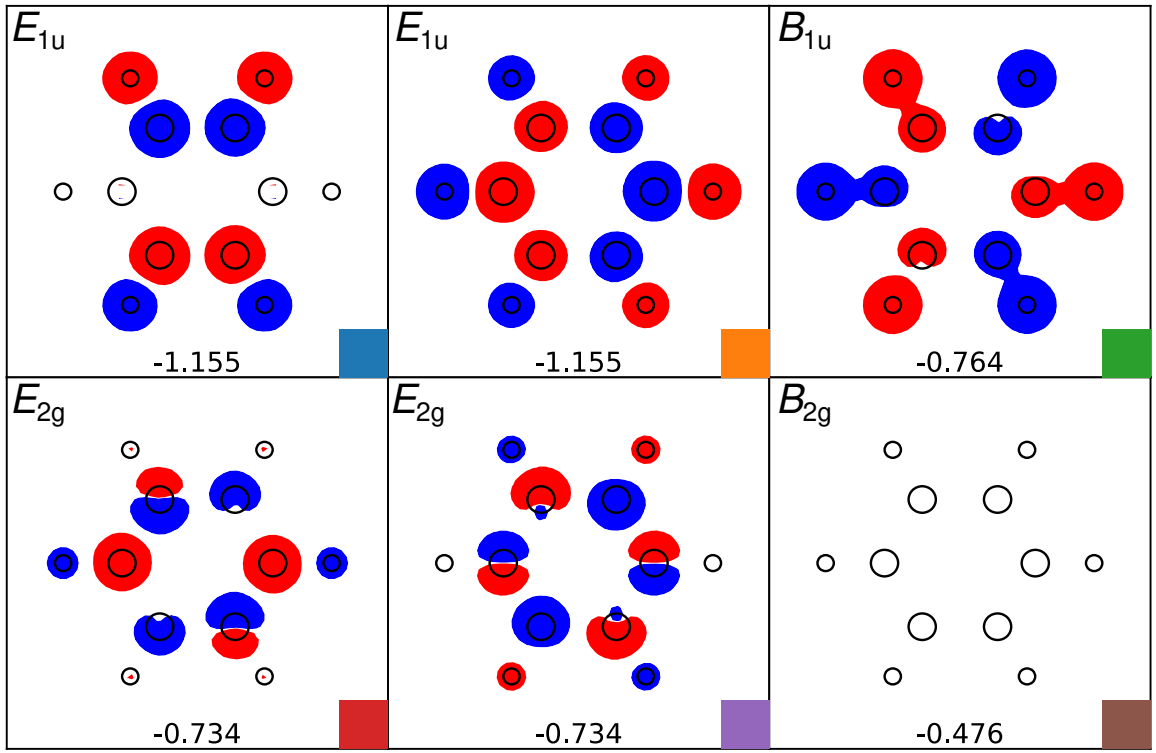
- Zeros in the Jacobian singular value spectrum enforce the correct extremal behavior for symmetric structures
- ... but for degenerate structures: artificial zero-force points!
- Deep consequences for equivariant models



Nigam, Willatt, **MC**, JCP (2021)

Symmetry and models

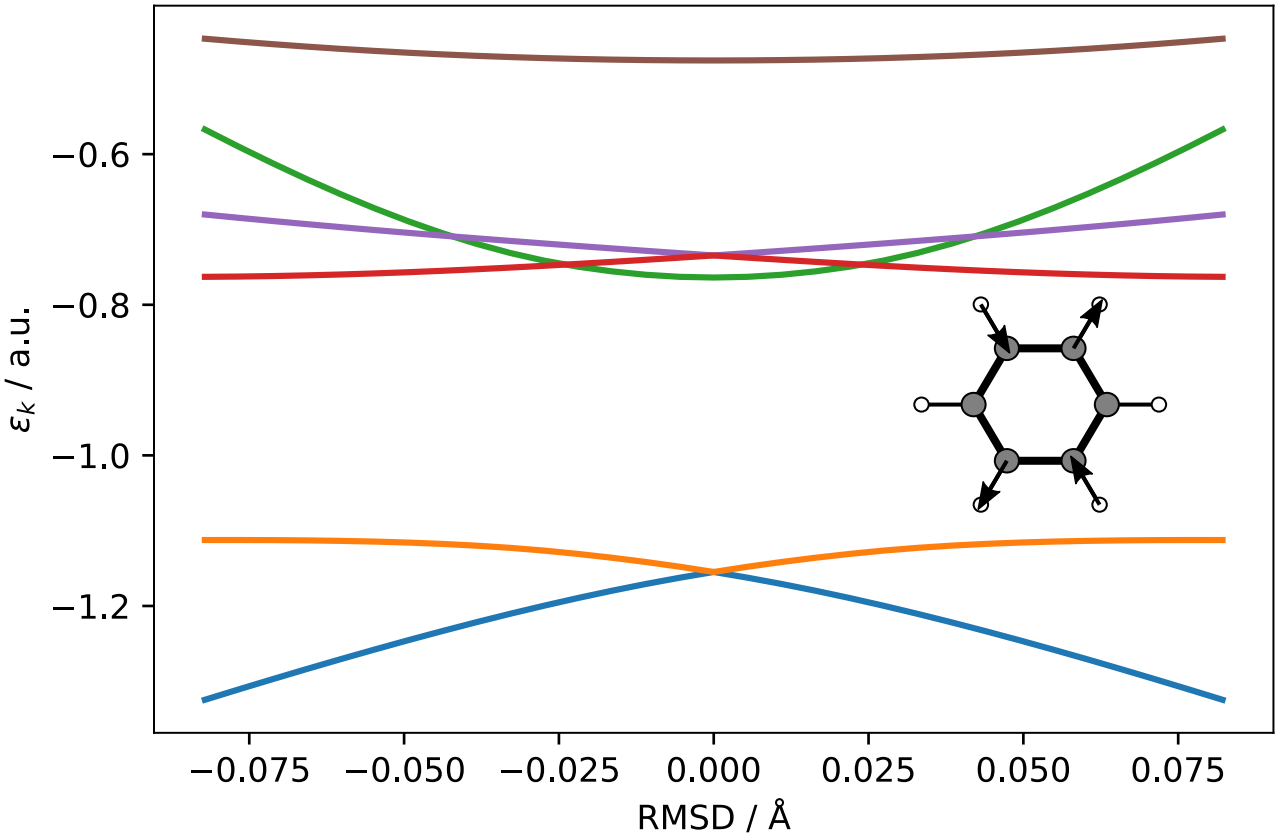
- Zeros in the Jacobian singular value spectrum enforce the correct extremal behavior for symmetric structures
- ... but for degenerate structures: artificial zero-force points!
- Deep consequences for equivariant models



Nigam, Willatt, **MC**, JCP (2021)

Symmetry and models

- Zeros in the Jacobian singular value spectrum enforce the correct extremal behavior for symmetric structures
- ... but for degenerate structures: artificial zero-force points!
- Deep consequences for equivariant models

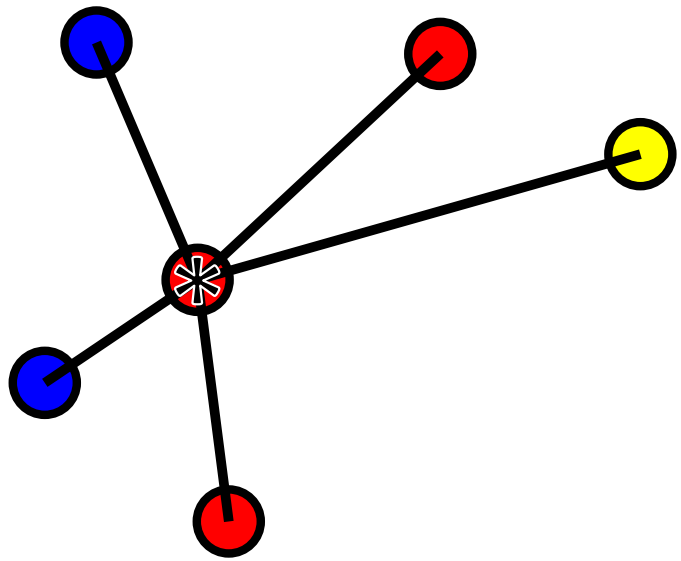


Nigam, Willatt, **MC**, JCP (2021)

How about graph convolution schemes?

Basic distance graph convolution

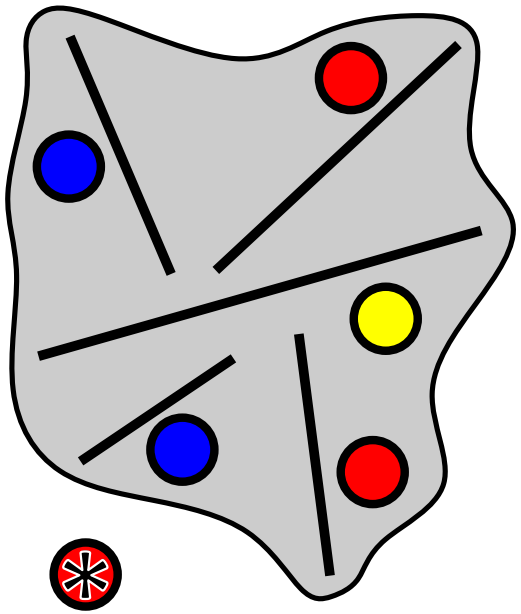
- Atoms are nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is decorated by the nature of its neighbors and their distance $h(A_i) = \left(a_i, \left\{(a_j, r_{ij})\right\}\right)$
- The multiset of neighbors and edges is hashed, and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



SchNET: Schütt et al., JCP (2018)

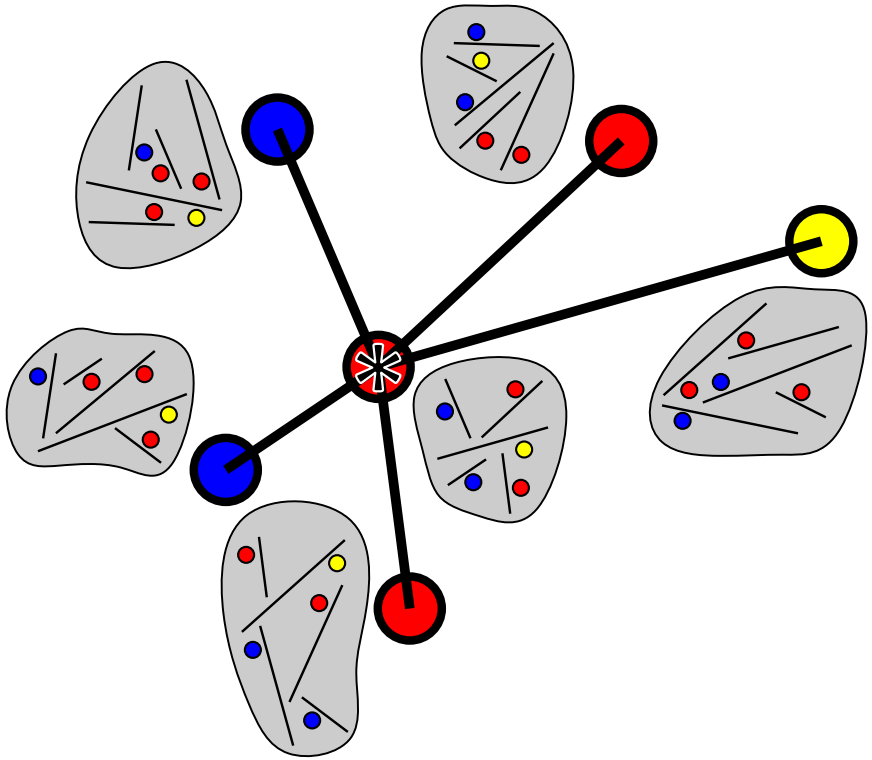
Basic distance graph convolution

- Atoms are nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is decorated by the nature of 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



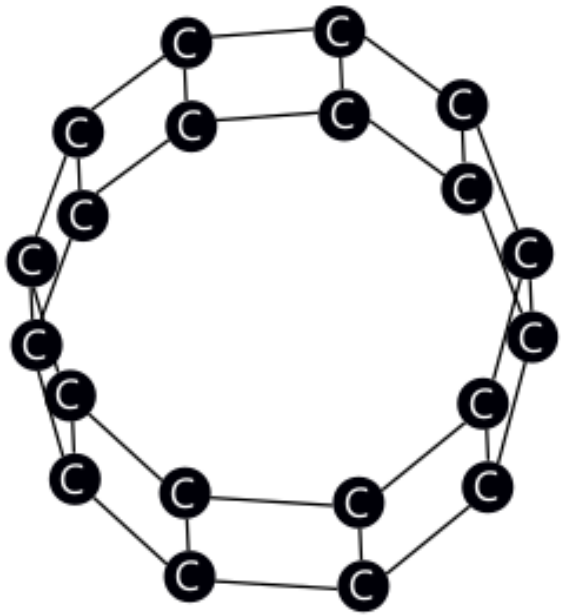
Basic distance graph convolution

- Atoms are nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is decorated by the nature of 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

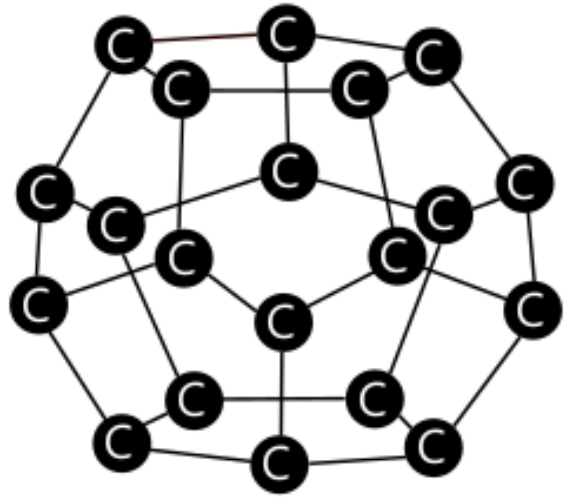


Graph convolution, pros and cons

- Bad news: there are known discrete graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things seem to be fine for molecular graphs (*fully-connected*, distance-decorated 3D point clouds); GC resolves all known counterexamples of atom-centered representations



(a) Decaprismane.

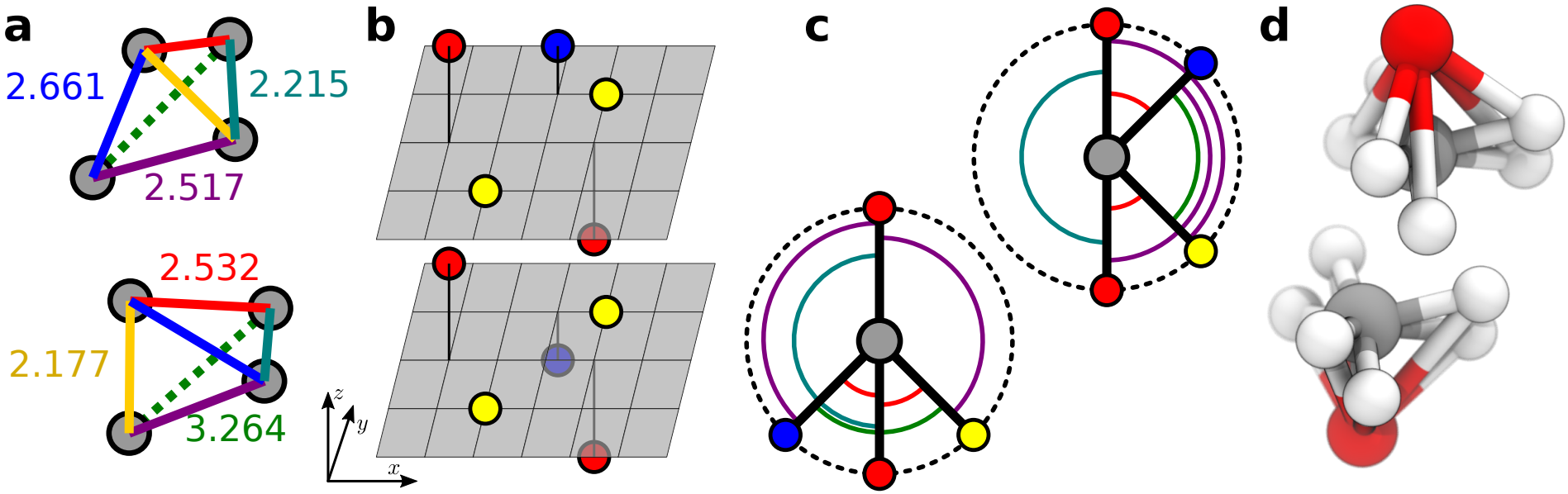


(b) Dodecahedrane.

Sato, arxiv:2003.04078

Graph convolution, pros and cons

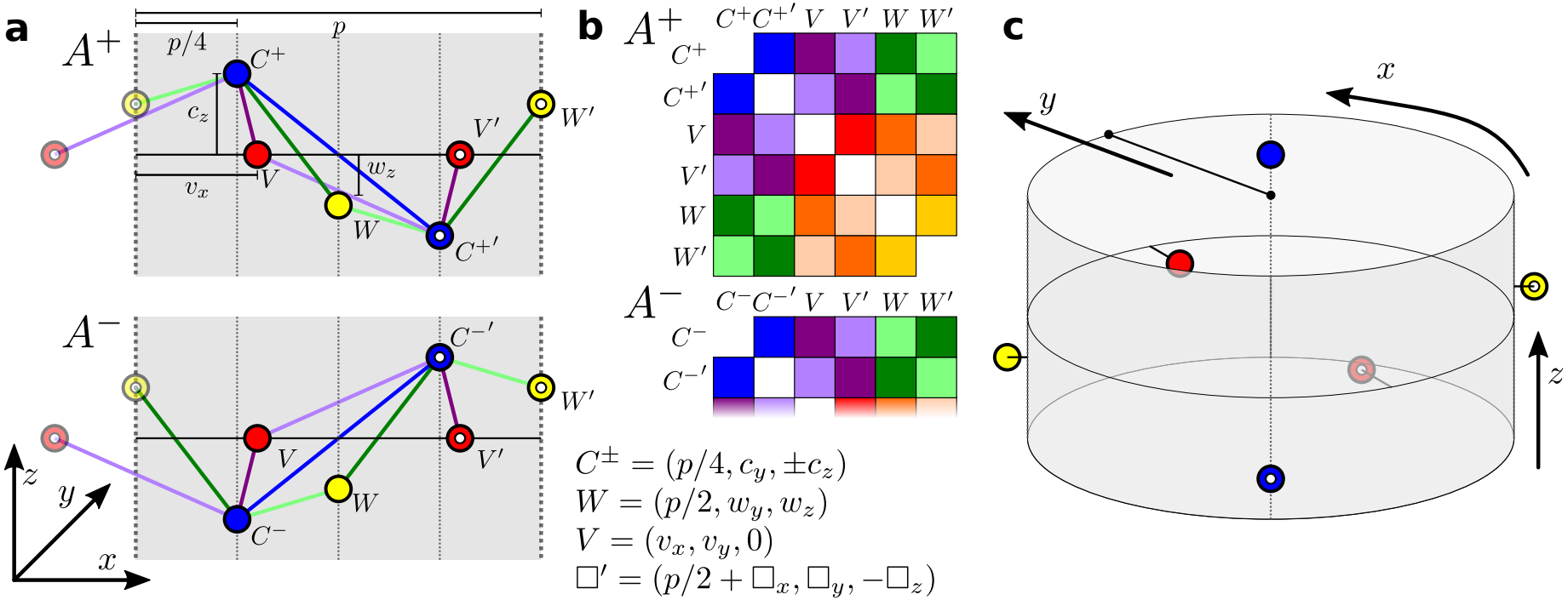
- Bad news: there are known discrete graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things seem to be fine for molecular graphs (*fully-connected*, distance-decorated 3D point clouds); GC resolves all known counterexamples of atom-centered representations



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

A counterexample for distance-based CNN

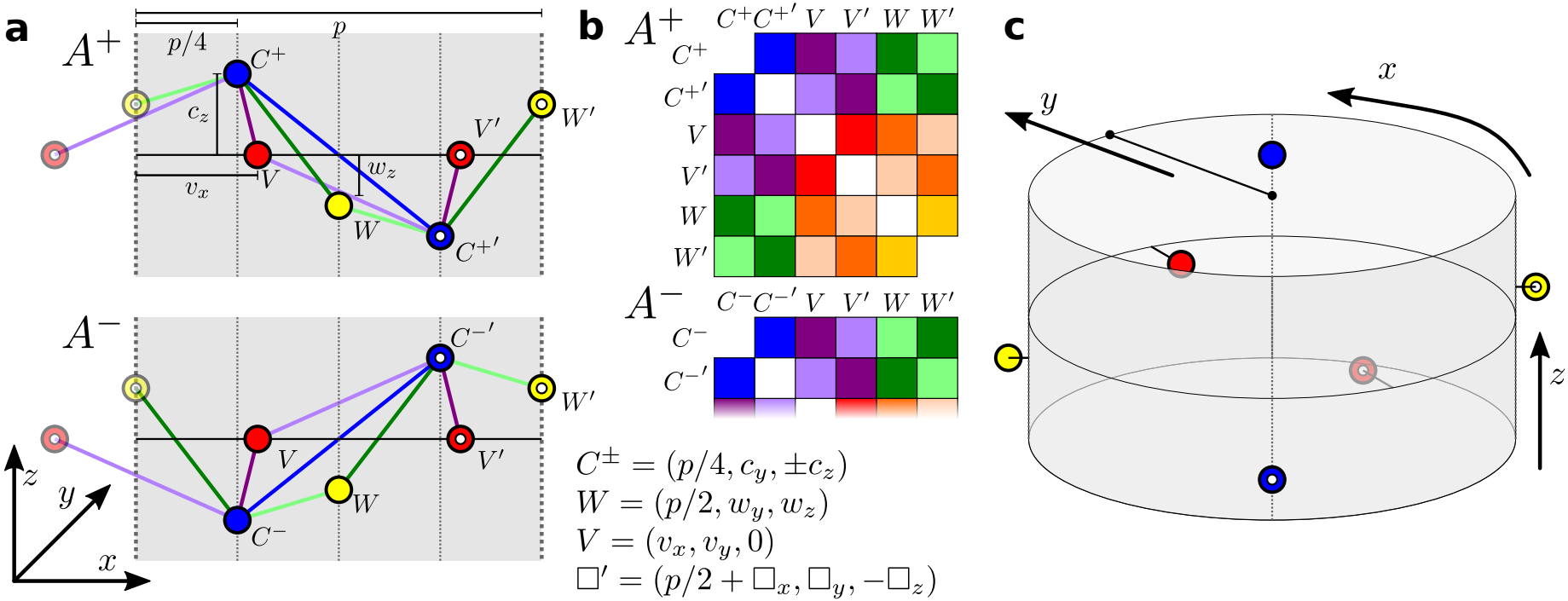
- A family of 3D point clouds with degenerate pairs for GCNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and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



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

A counterexample for distance-based CNN

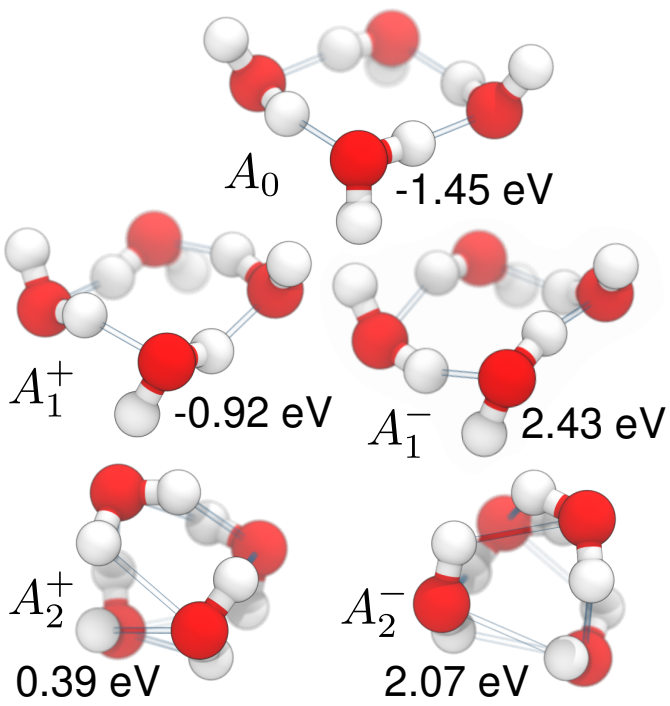
- A family of 3D point clouds with degenerate pairs for GCNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and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

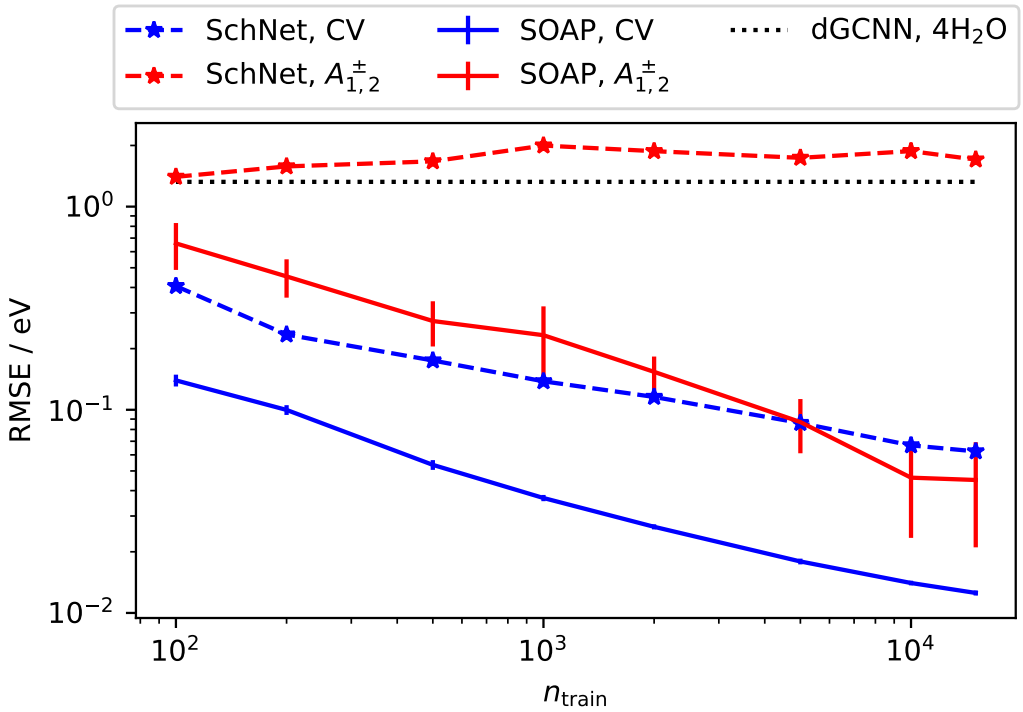


Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

A counterexample for distance-based CNN

- A family of 3D point clouds with degenerate pairs for GCNN. 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

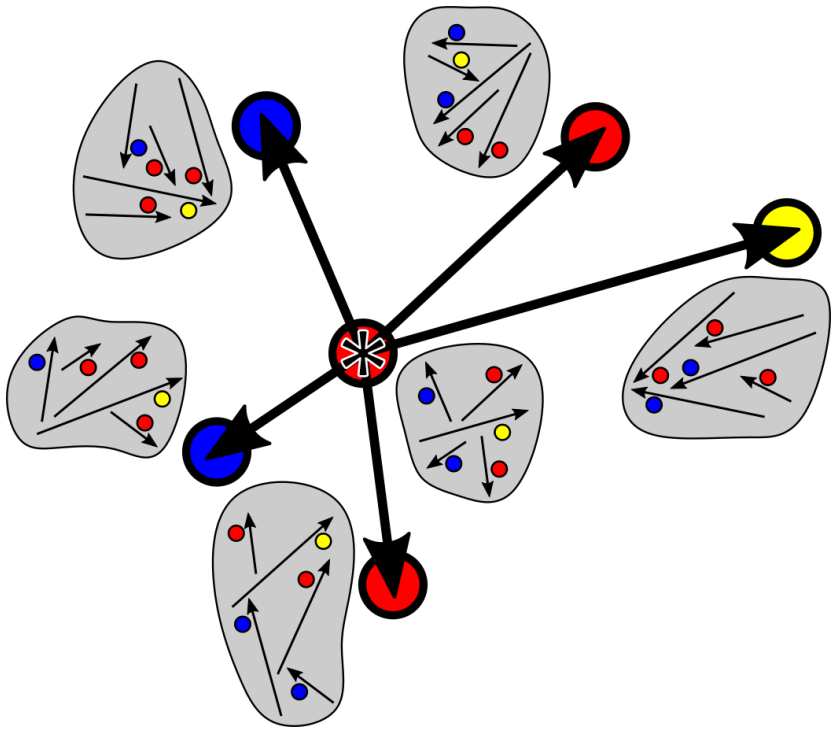




Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

A counterexample for distance-based CNN

- A family of 3D point clouds with degenerate pairs for GCNN. 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

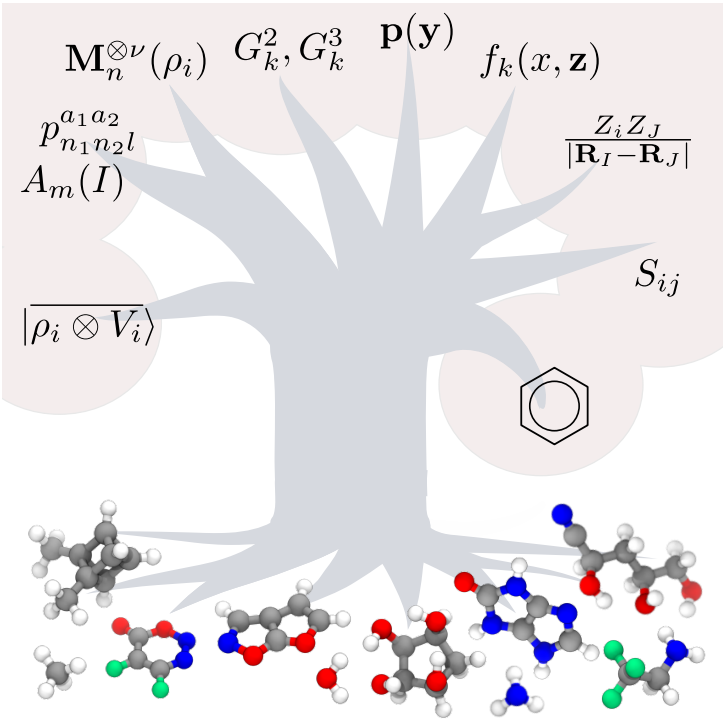


Pozdnyakov, **MC**, arXiv:2201.07136 (2022)

Outlook

Towards a rigorous theory of atomistic ML

- Atomistic simulations have a long tradition of fitting potentials, giving a framework to understand the implication of representations and models to machine learn atomic-scale properties
- Locality, symmetry and equivariance are key elements to incorporate domain priors. *Non-locality: another interesting story!*
- A different perspective on descriptive power and transferability. *Relevant for general geometric ML?*
- A systematic theory of atom-centered representations. *Extend to message-passing architectures?*



Computational science & modeling @ EPFL

cosmo.epfl.ch

Follow @lab_COSMO

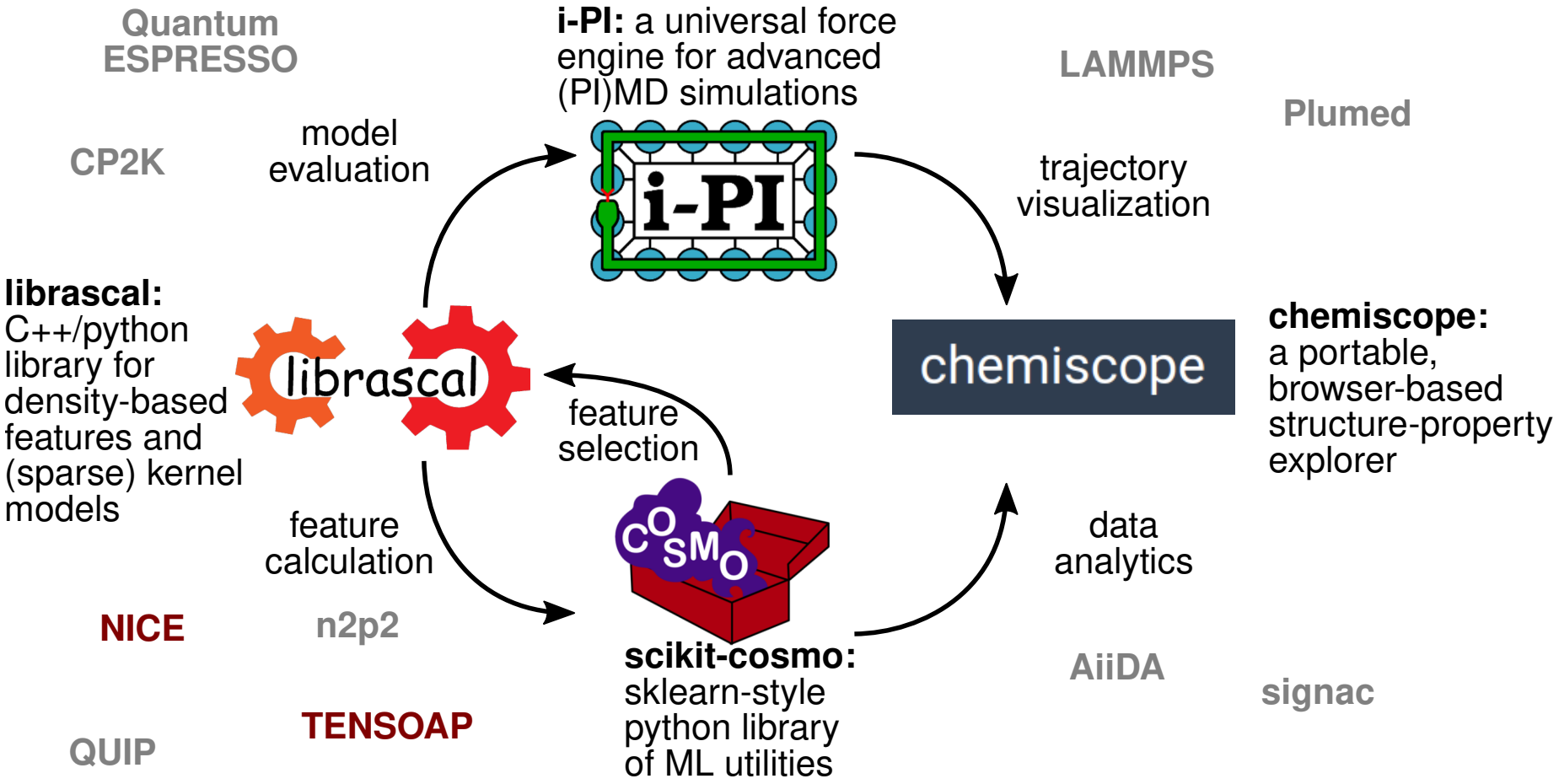
Computational science & modeling @ EPFL

cosmo.epfl.ch

Follow @lab_COSMO

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/riken-epfl-2022**
Review → **Musil et al. ChemRev (2020)**