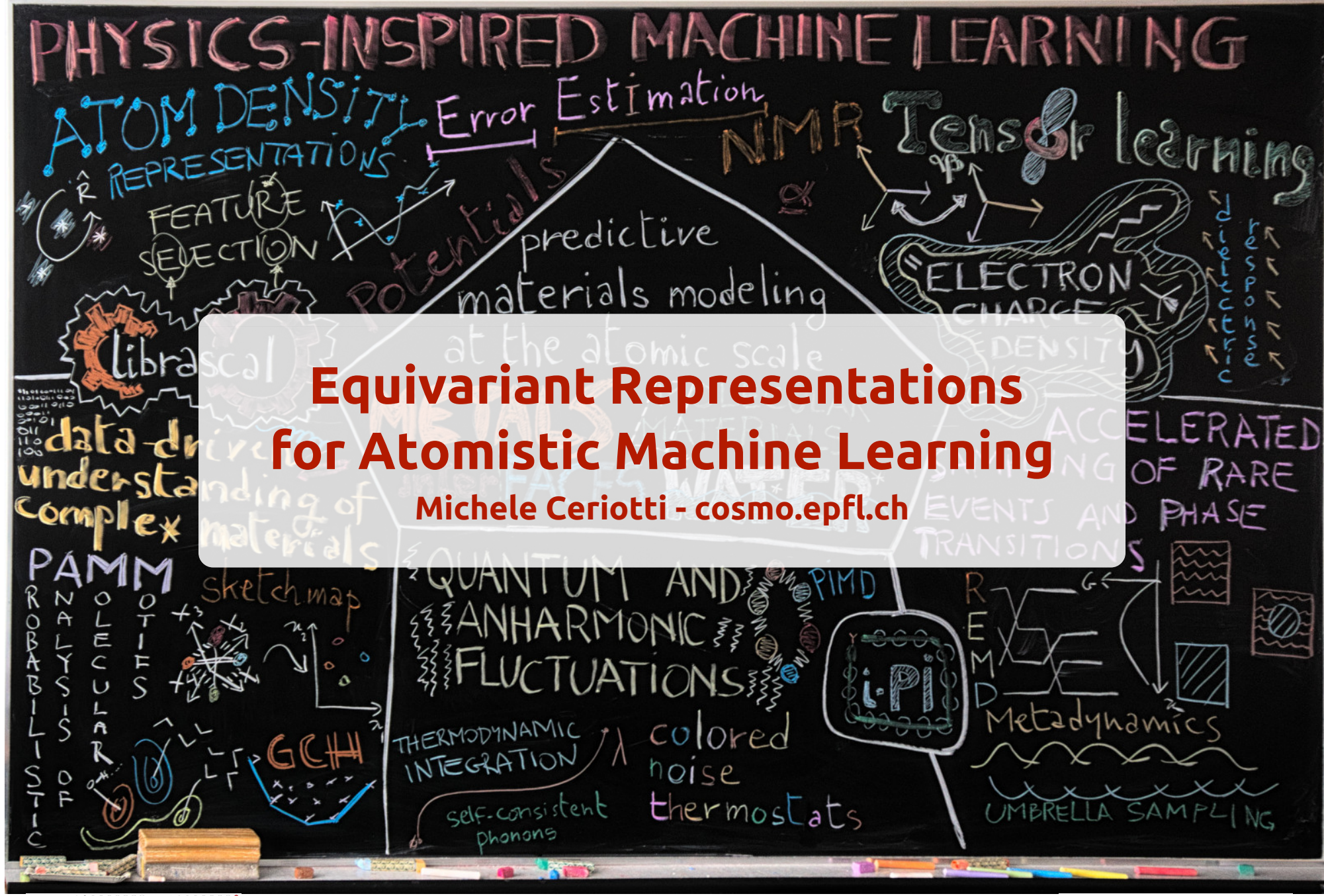


Equivariant Representations for Atomistic Machine Learning

Michele Ceriotti - cosmo.epfl.ch

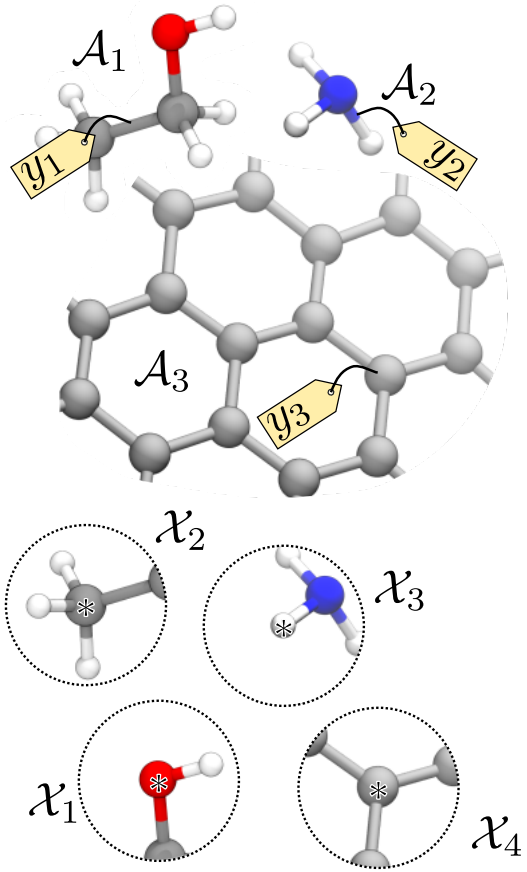
Workshop on Molecular Dynamics and its Applications to Biological Systems, Sept. 2020

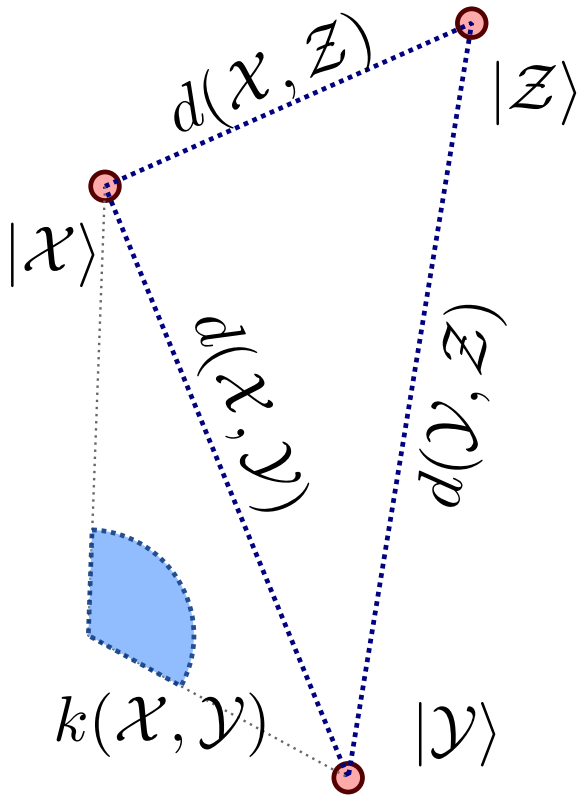
COSMO COMPUTATIONAL SCIENCE AND MODELLING

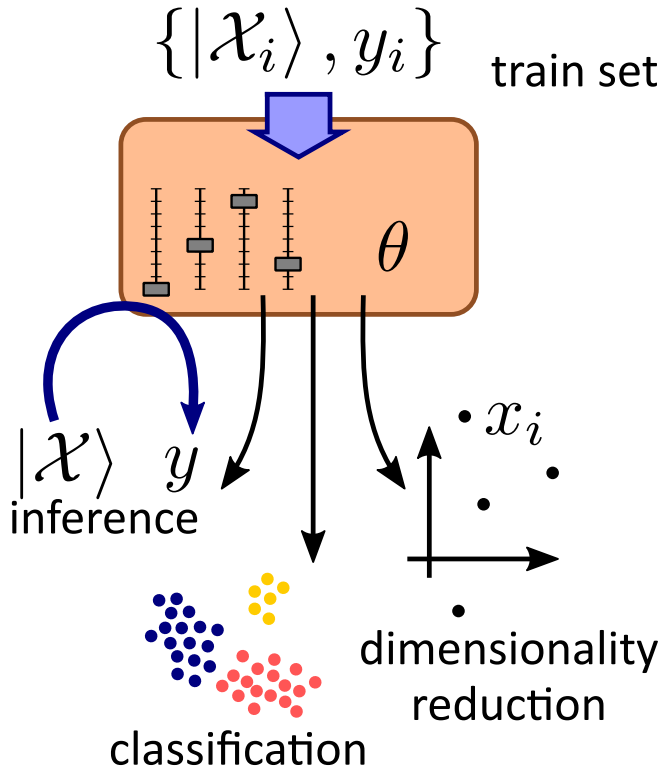
EPFL

The problem of representation

- Mapping an atomic structure to a mathematical representation suitable to ML is the first and perhaps most important step for atomistic machine learning

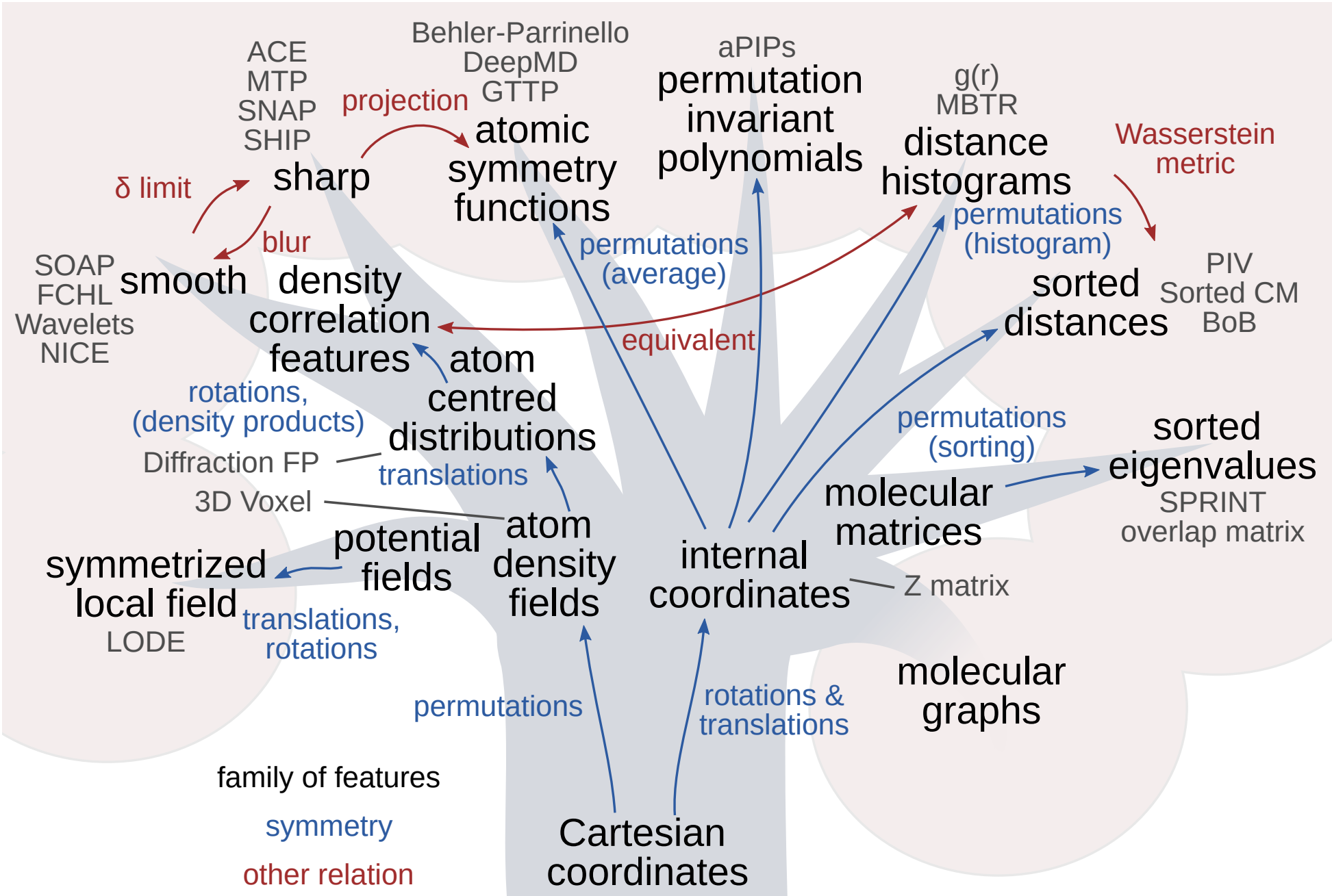






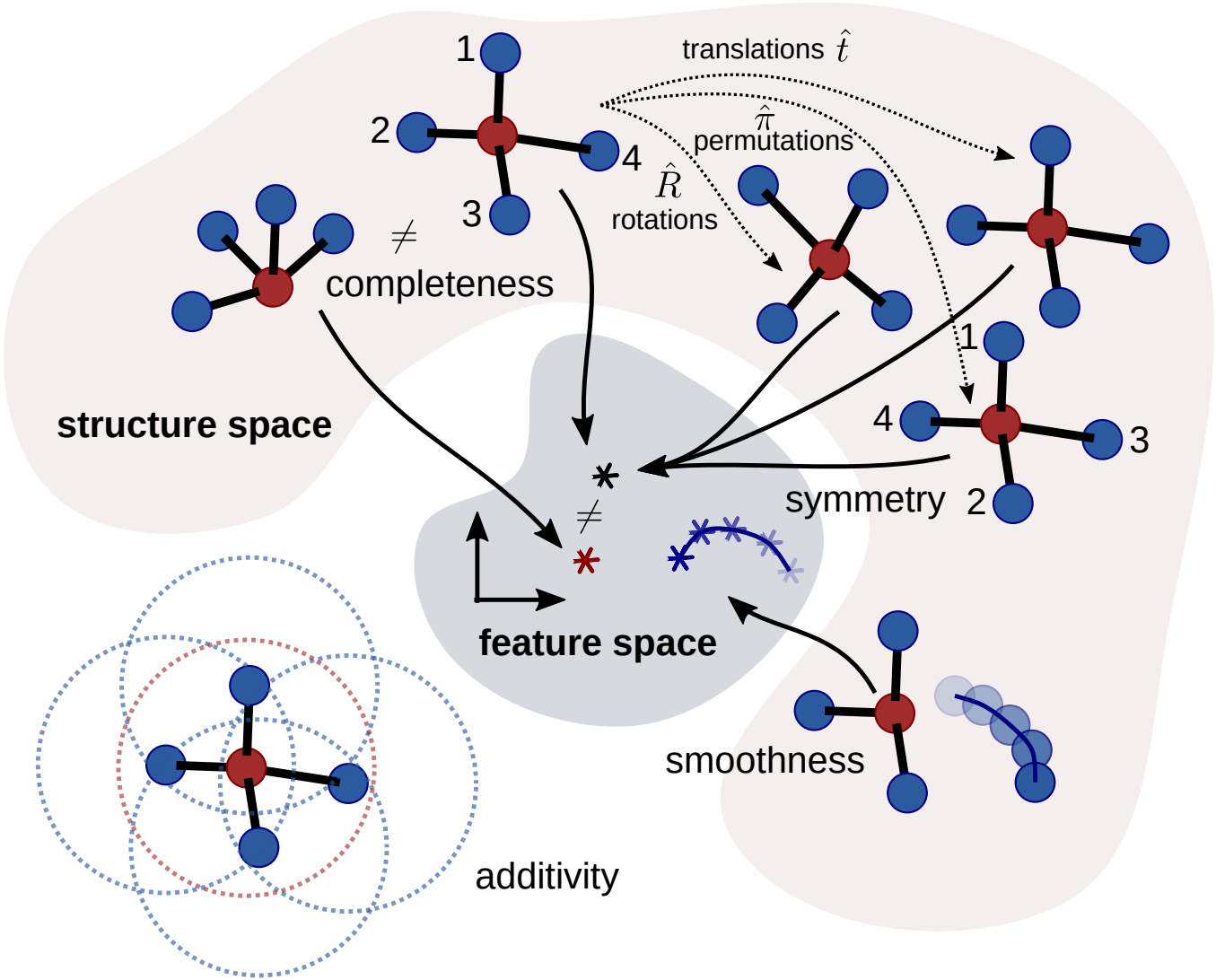
MC, *Unsupervised machine learning in atomistic simulations, between predictions and understanding*, JCP (2019)

A phylogenetic tree of ML representations



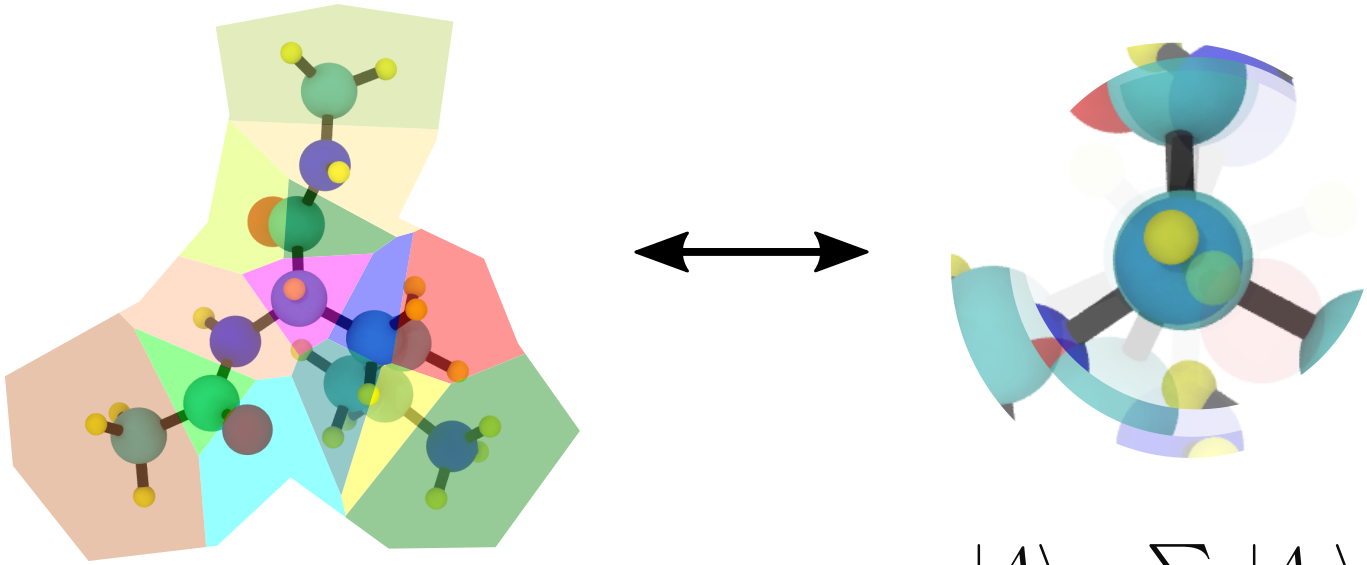
What we want from a representation

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



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



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

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

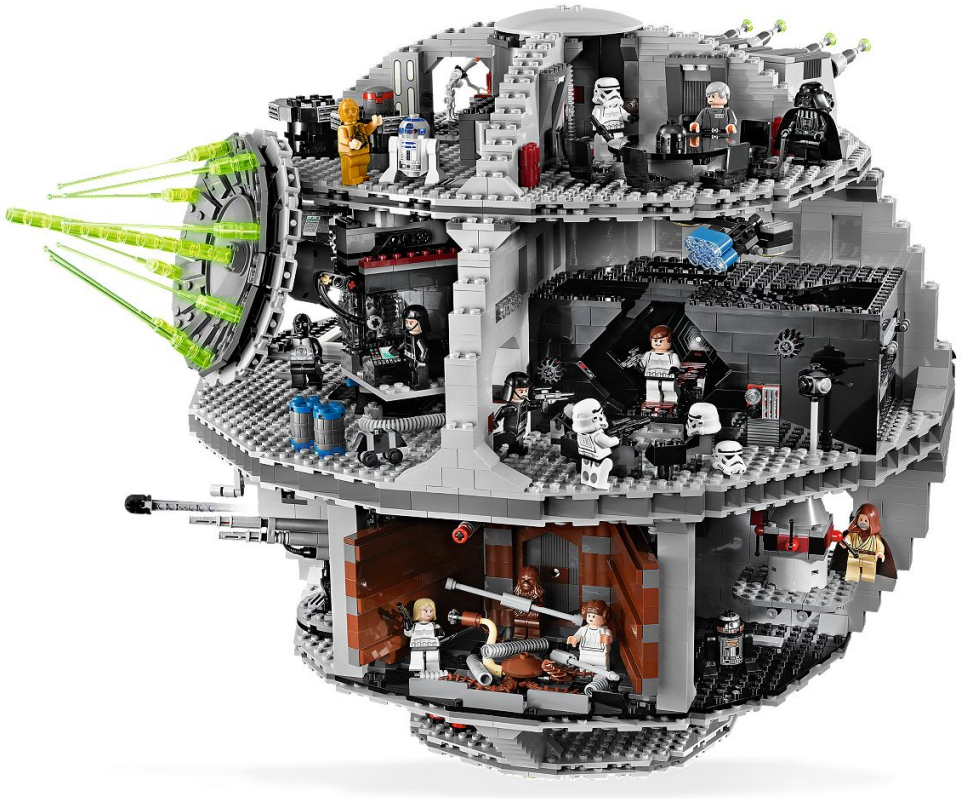
$$K(A,B) = \sum_{i,j} k(A_i, B_j)$$

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

- 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



A Dirac notation for representations

features index $\langle X | A \rangle$ 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; field center; correlation order; parity; rot. symmetry

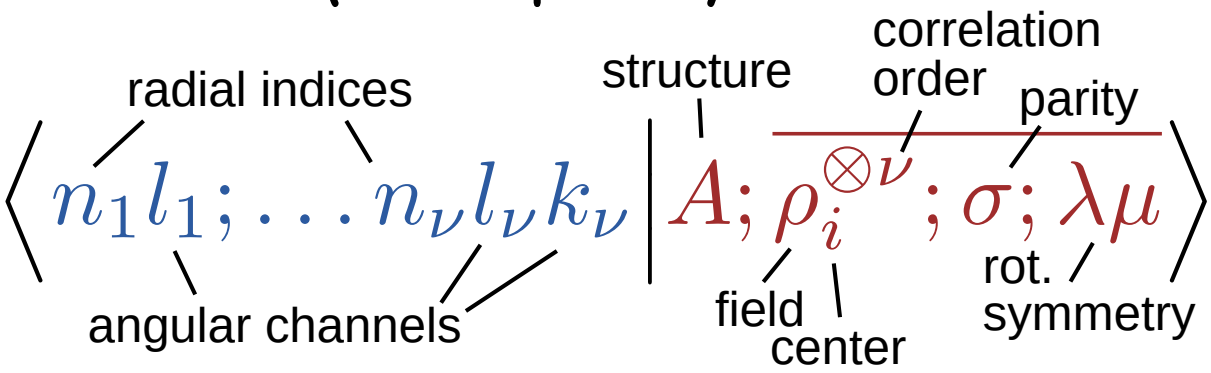
- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index X
- Bra-ket notation $\langle X|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally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

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

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

A Dirac notation for representations

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



- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index X
- Bra-ket notation $\langle X|A;\text{rep.}\rangle$ 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

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

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

A Dirac notation for representations

features index $\langle X | A \rangle$ 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

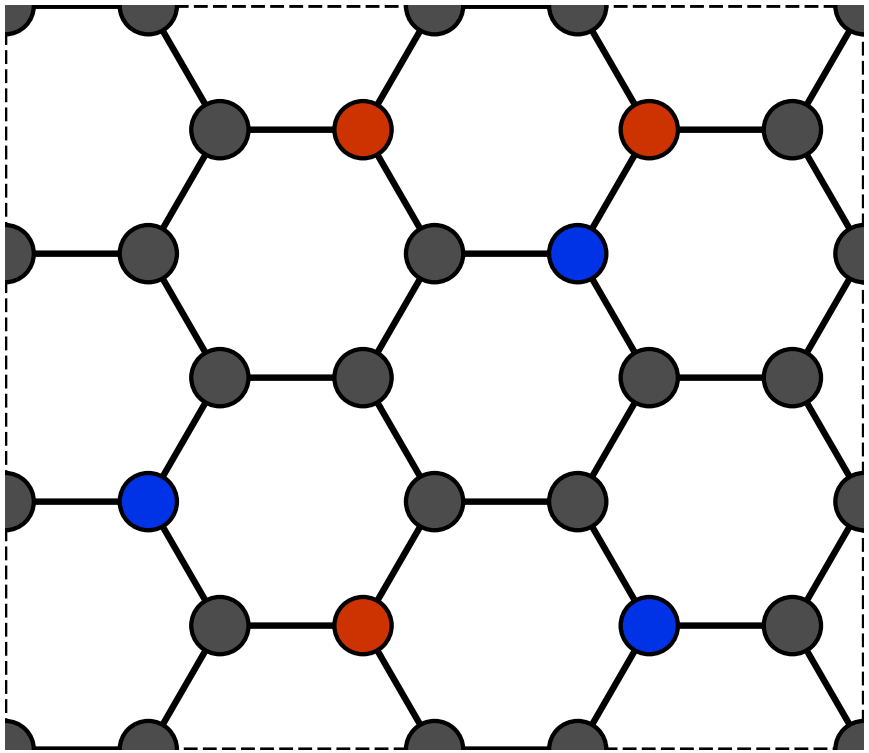
- A representation maps a structure A (or one environment A_i) to a vector discretized by a feature index X
- Bra-ket notation $\langle X|A;\text{rep.}\rangle$ indicates in an abstract way this mapping, leaving plenty of room to express the details of a representation
- Dirac-like notation reflects naturally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

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

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

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) $|\rho_i\rangle$



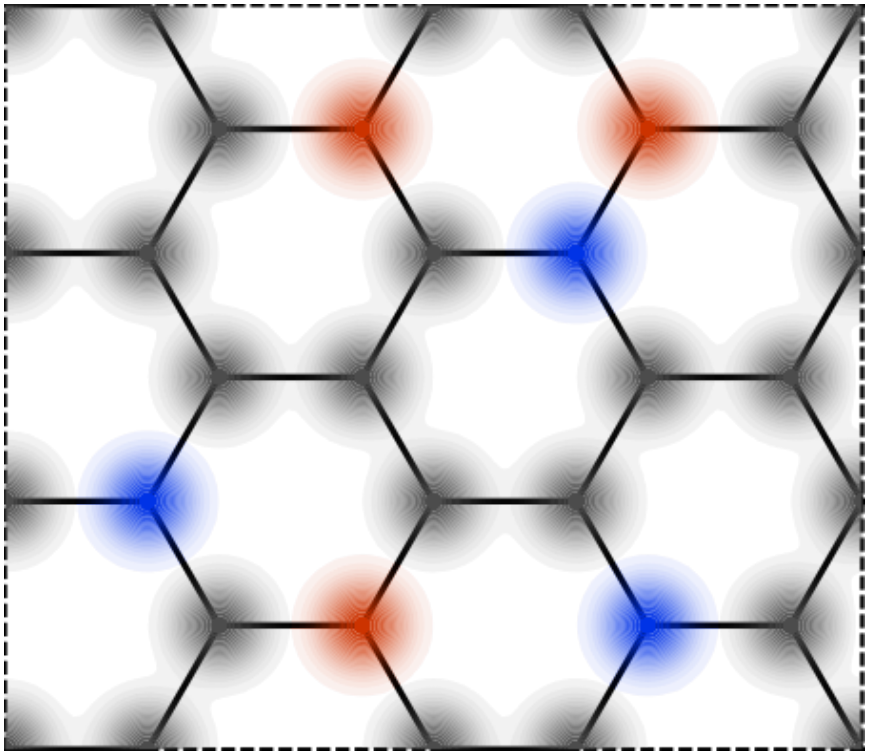
$\mathcal{A} \rightarrow$

```
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) $|\rho_i\rangle$



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

$|\mathrm{C}\rangle$ ■

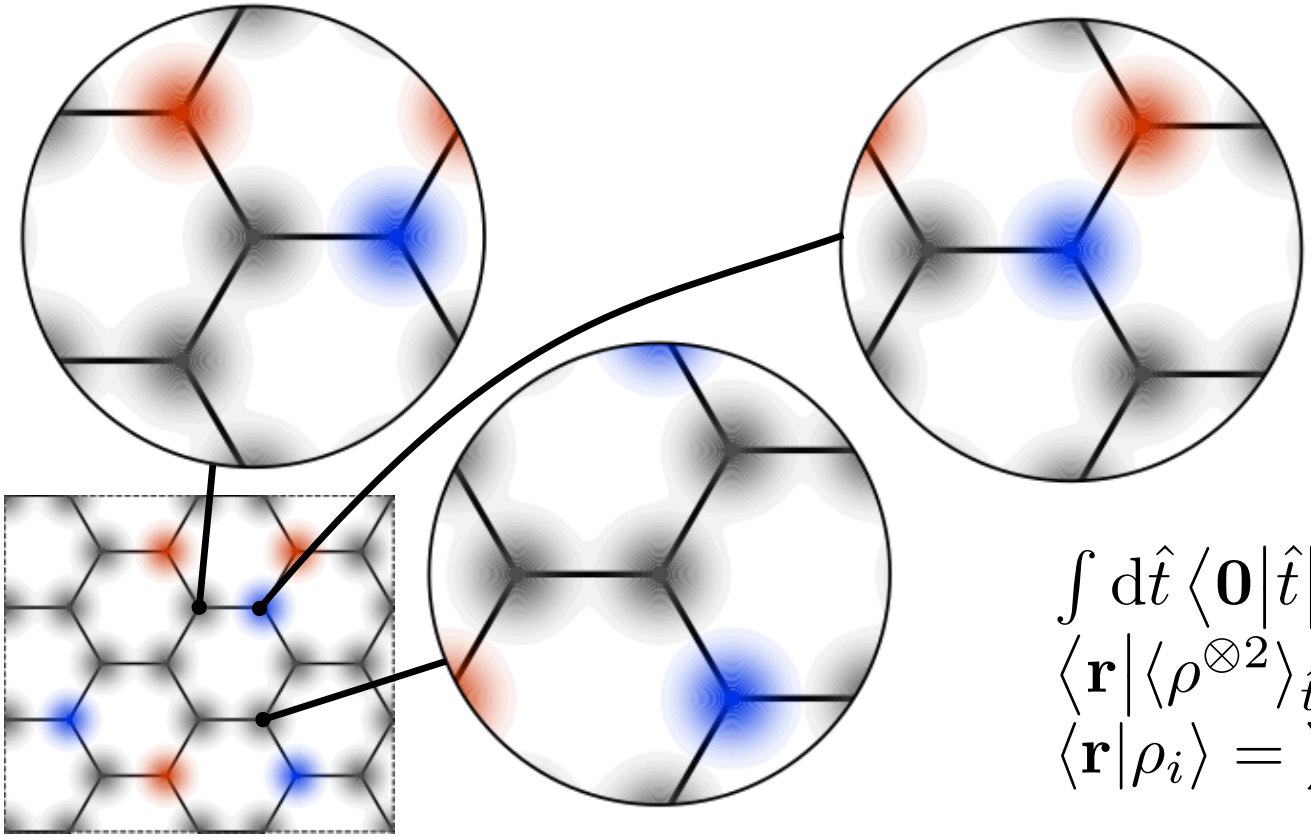
$|\mathrm{N}\rangle$ ■

$|\mathrm{B}\rangle$ ■

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) $|\rho_i\rangle$



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

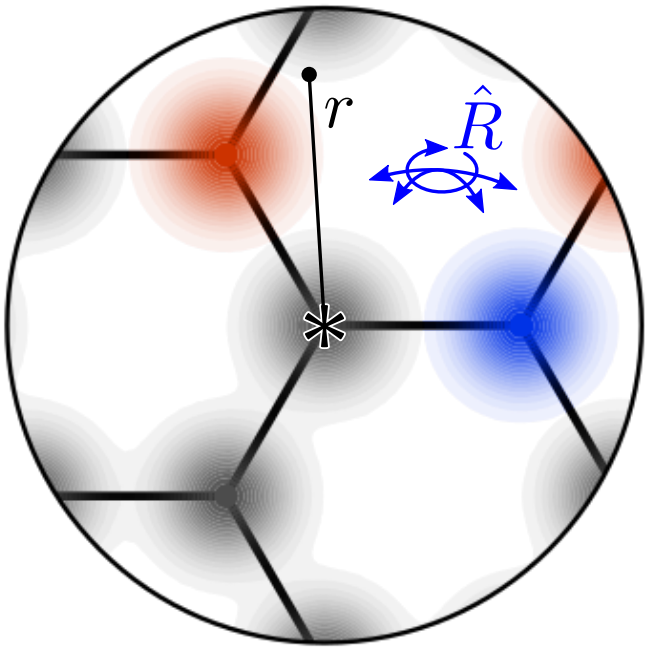
Willatt, Musil, **MC**, JCP (2019)

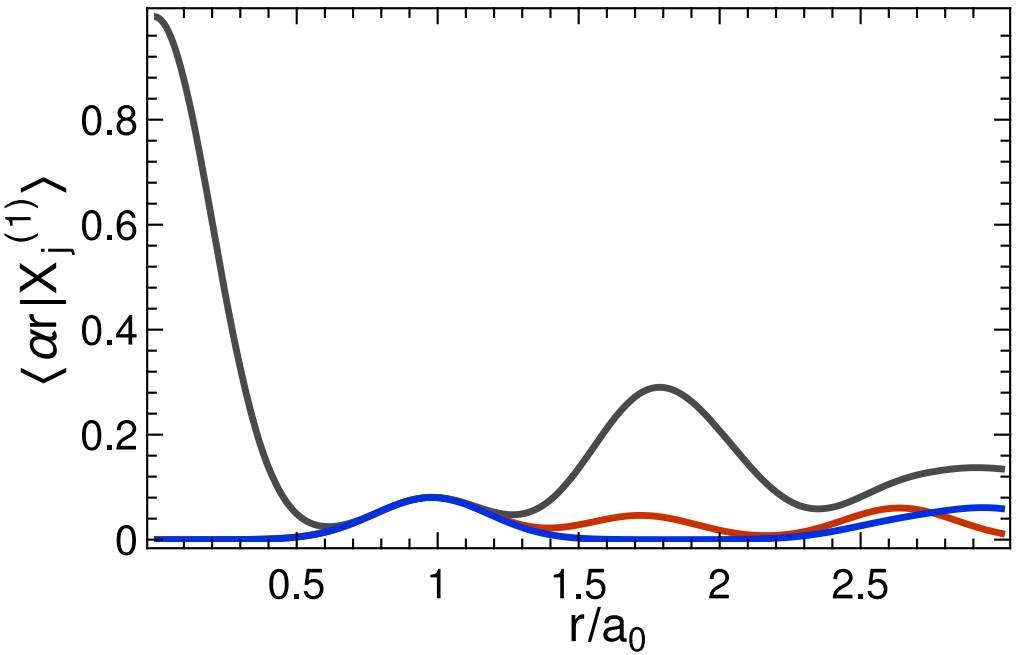
A universal feature construction

- 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

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

- Basically any atom-centred feature can be seen as a projection of $|\overline{\rho_i^{\otimes\nu}}\rangle$





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

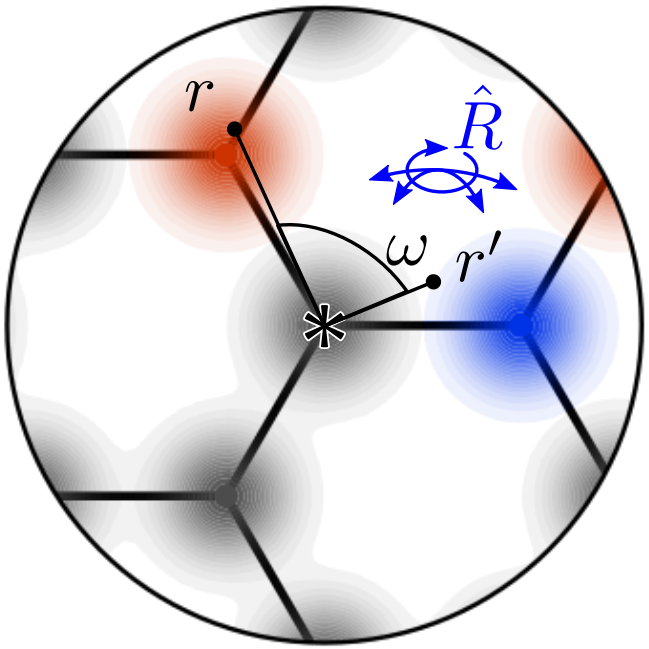
Willatt, Musil, **MC**, JCP (2019); Bartók, Kondor, Csányi PRB 2013

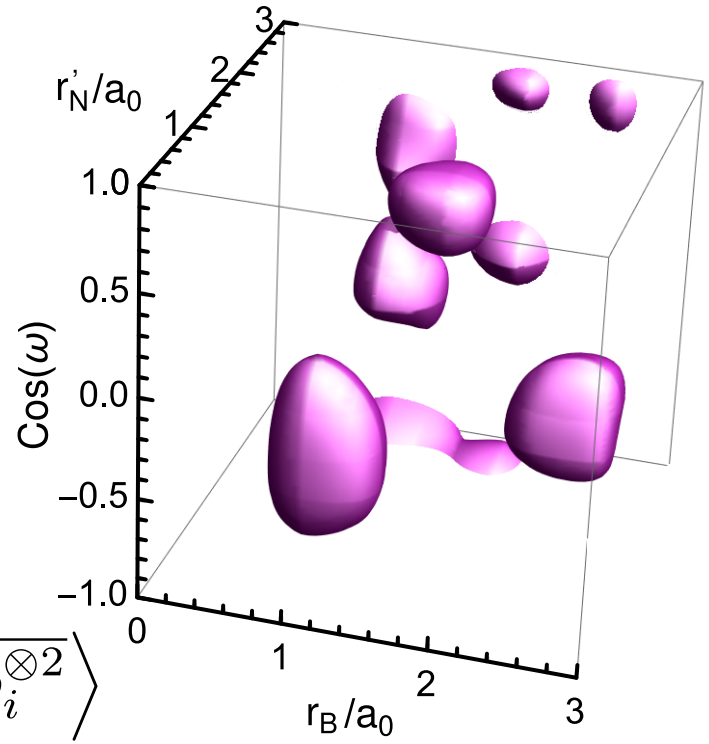
A universal feature construction

- 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

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

- Basically any atom-centred feature can be seen as a projection of $|\overline{\rho_i^{\otimes\nu}}\rangle$





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

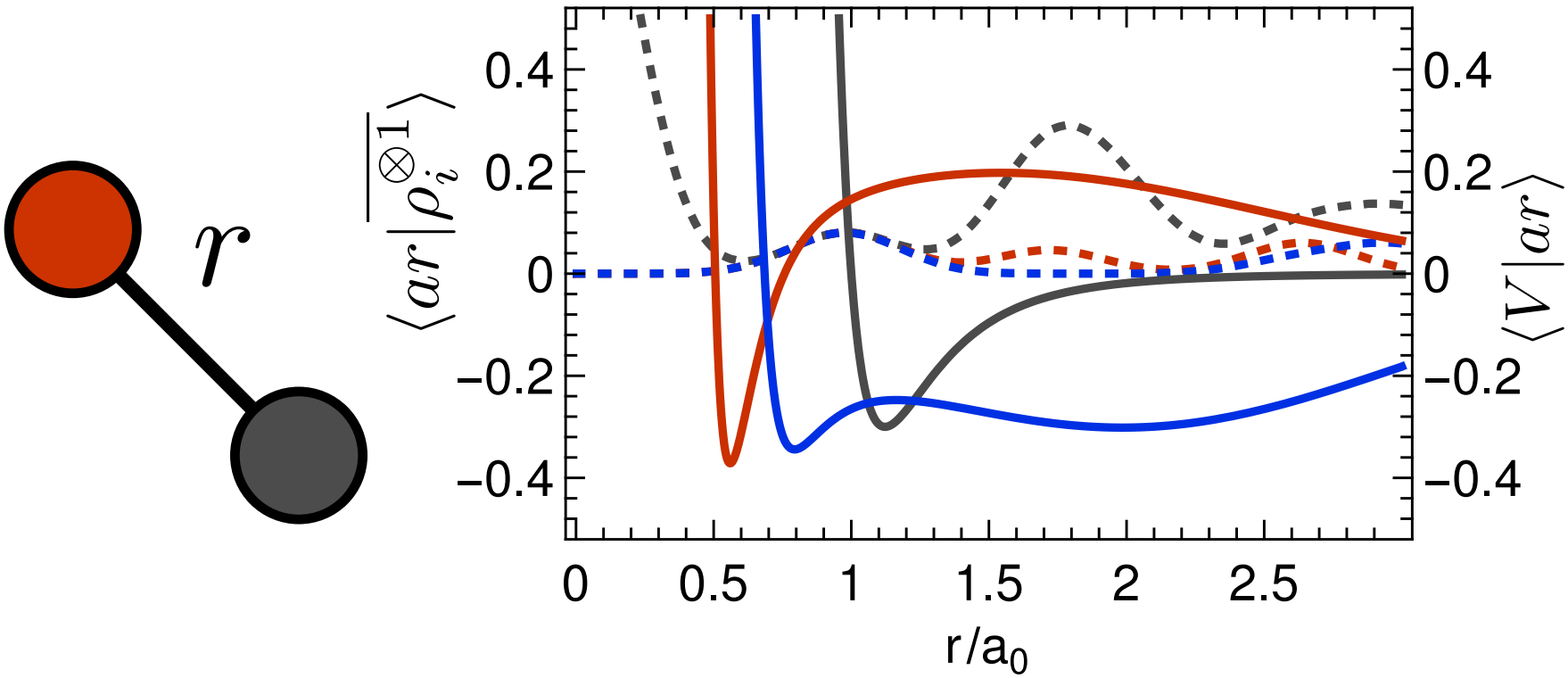
Willatt, Musil, **MC**, JCP (2019); Drautz, PRB (2019); Glielmo, Zeni, De Vita, PRB (2018)

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

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

- Basically any atom-centred feature can be seen as a projection of $|\overline{\rho_i^{\otimes\nu}}\rangle$



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

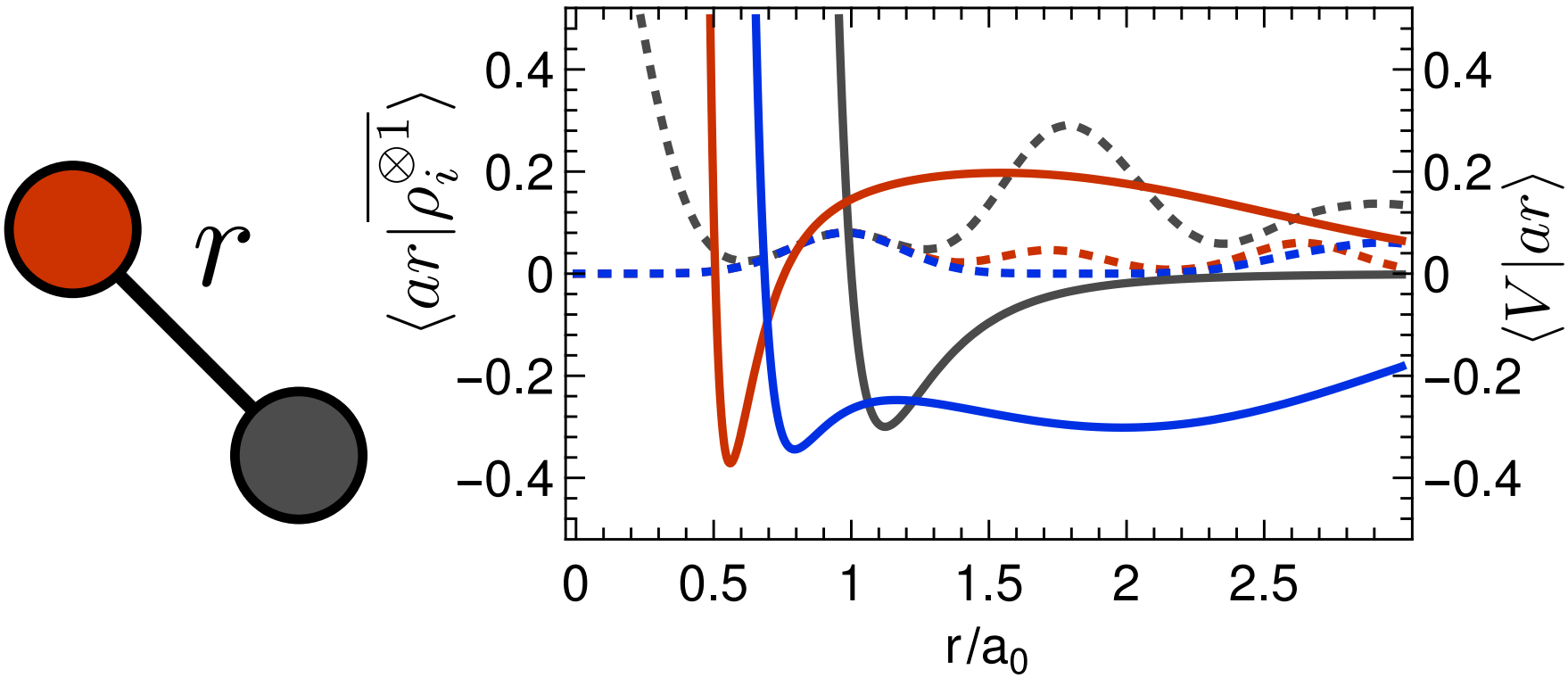
Willatt, Musil, **MC**, JCP (2019); Drautz, PRB (2019); Glielmo, Zeni, De Vita, PRB (2018)

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

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

- Basically any atom-centred feature can be seen as a projection of $|\overline{\rho_i^{\otimes\nu}}\rangle$



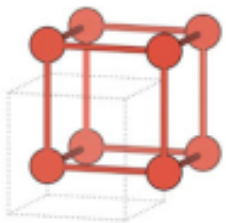
$$V(A_i) = \int \mathrm{d}r \langle V|ar\rangle\langle ar|\overline{\rho_i^{\otimes 1}}\rangle \approx \sum_j V_a(r_{ij})$$

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

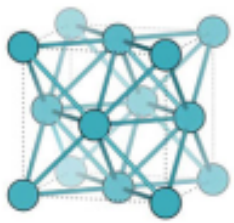
Variations on a theme

- 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)

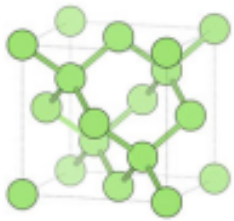
$$\langle \mathbf{k}|A;\overline{\rho^{\otimes 2}}\rangle = \sum_{ij\in A} e^{i\mathbf{k}\cdot\mathbf{r}_{ij}}$$



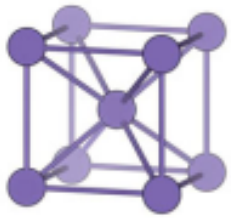
Simple cubic
(sc) structure
spgroup = 221



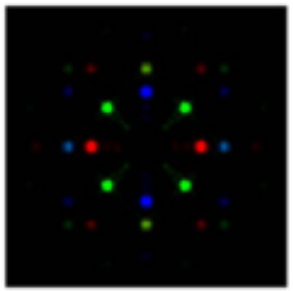
Face-centered-cubic
(fcc) structure
spgroup = 225

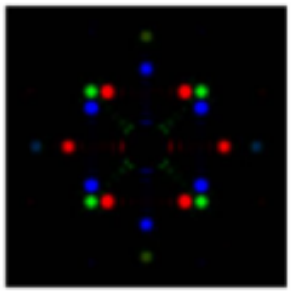


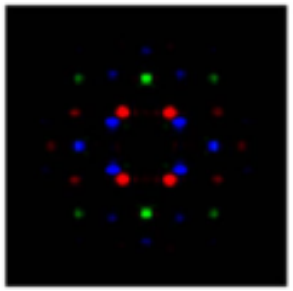
Diamond
(diam) structure
spgroup = 227

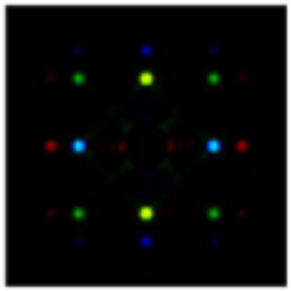


Body-centered-cubic
(bcc) structure
spgroup = 229







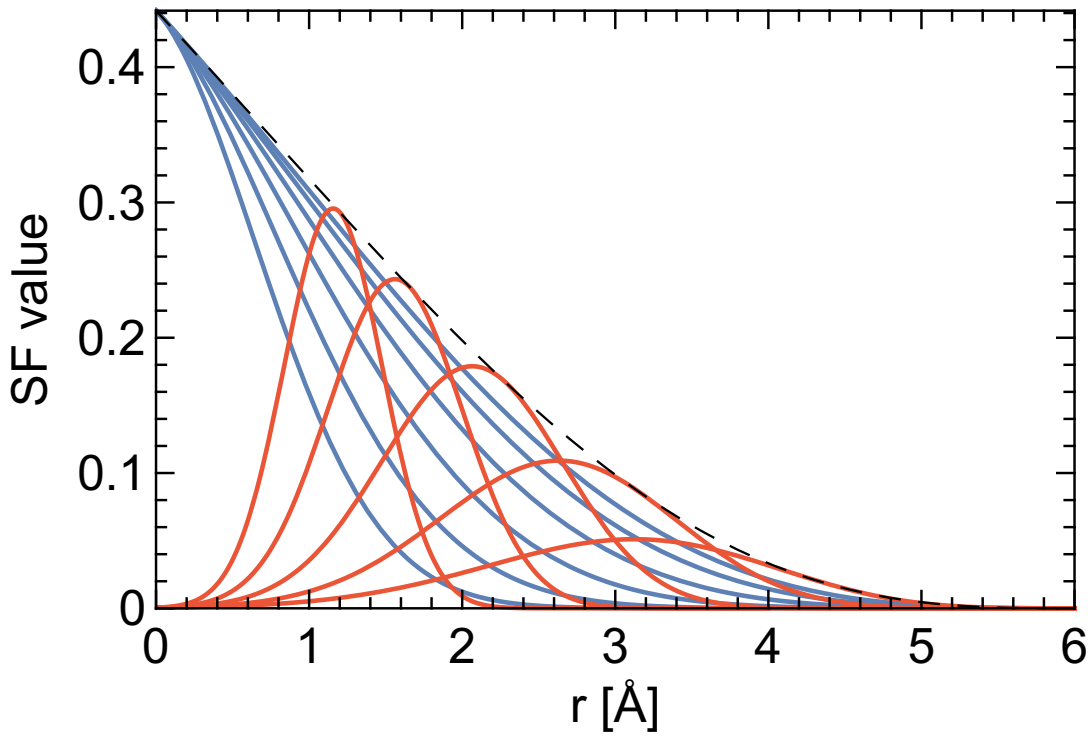


Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

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)

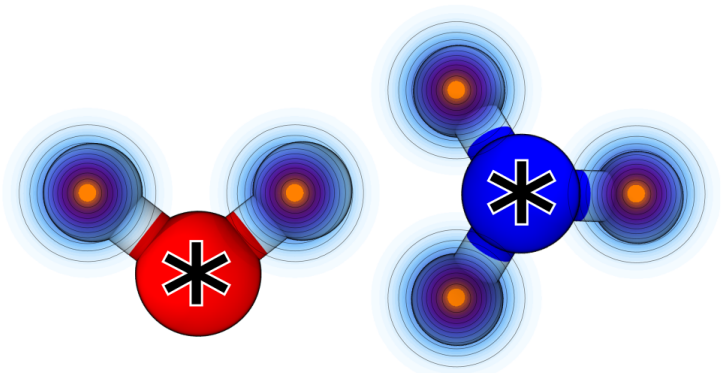
$$\langle abG_2|\overline{\rho_i^{\otimes 1}}\rangle = \delta_{aa_i} \int \mathrm{d}r\, G_2(r)\, \langle br|\overline{\rho_i^{\otimes 1}}; g \to \delta\rangle$$



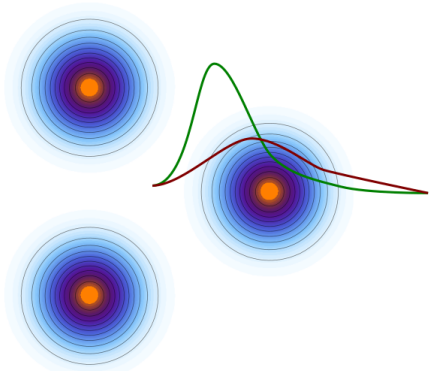
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

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



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

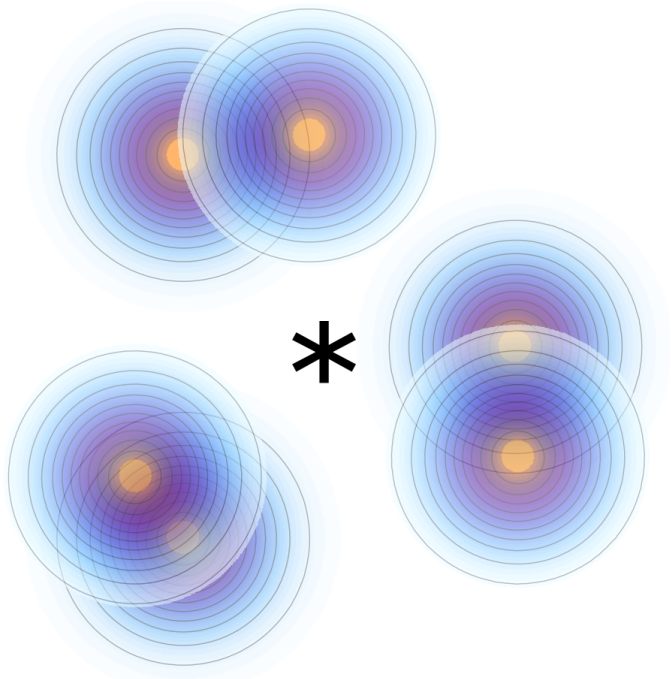


$$\langle nlm|\rho_i\rangle = \int \mathrm{d}\mathbf{r}\, \langle \mathbf{r}|\rho_i\rangle R_n(r) Y_m^l(\hat{\mathbf{r}})$$

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

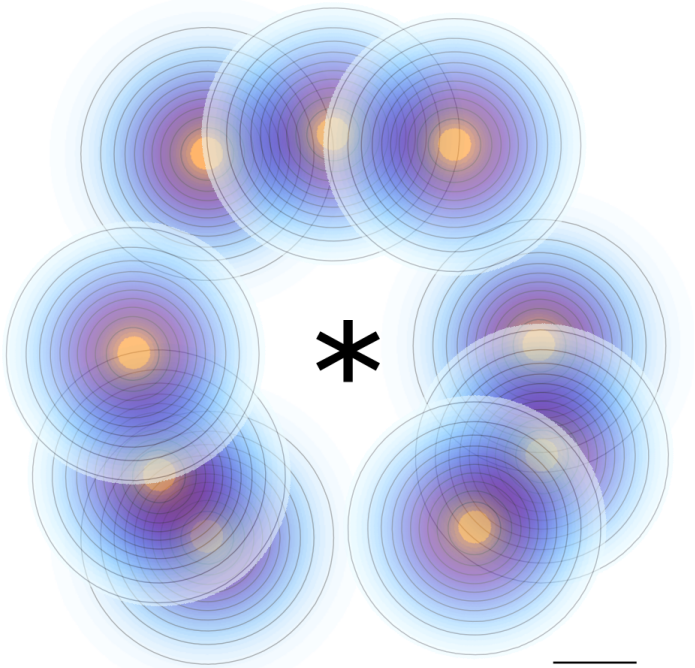


$$\langle nn'l|\overline{\rho_i^{\otimes 2}}\rangle = \sum_m \langle nlm|\rho_i\rangle^\star \langle n'lm|\rho_i\rangle$$
$$p_{nn'l} = \sum_m c^\star_{nlm} c_{n'lm}$$

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)

Are these representations complete?

- It is well-known that 2-body correlations are ambiguous: can build tetrahedra with same pair distances that are different
- One can also build examples of pairs of environments that have the same 3B and 4B correlations. Problem becomes important as model accuracy is increased

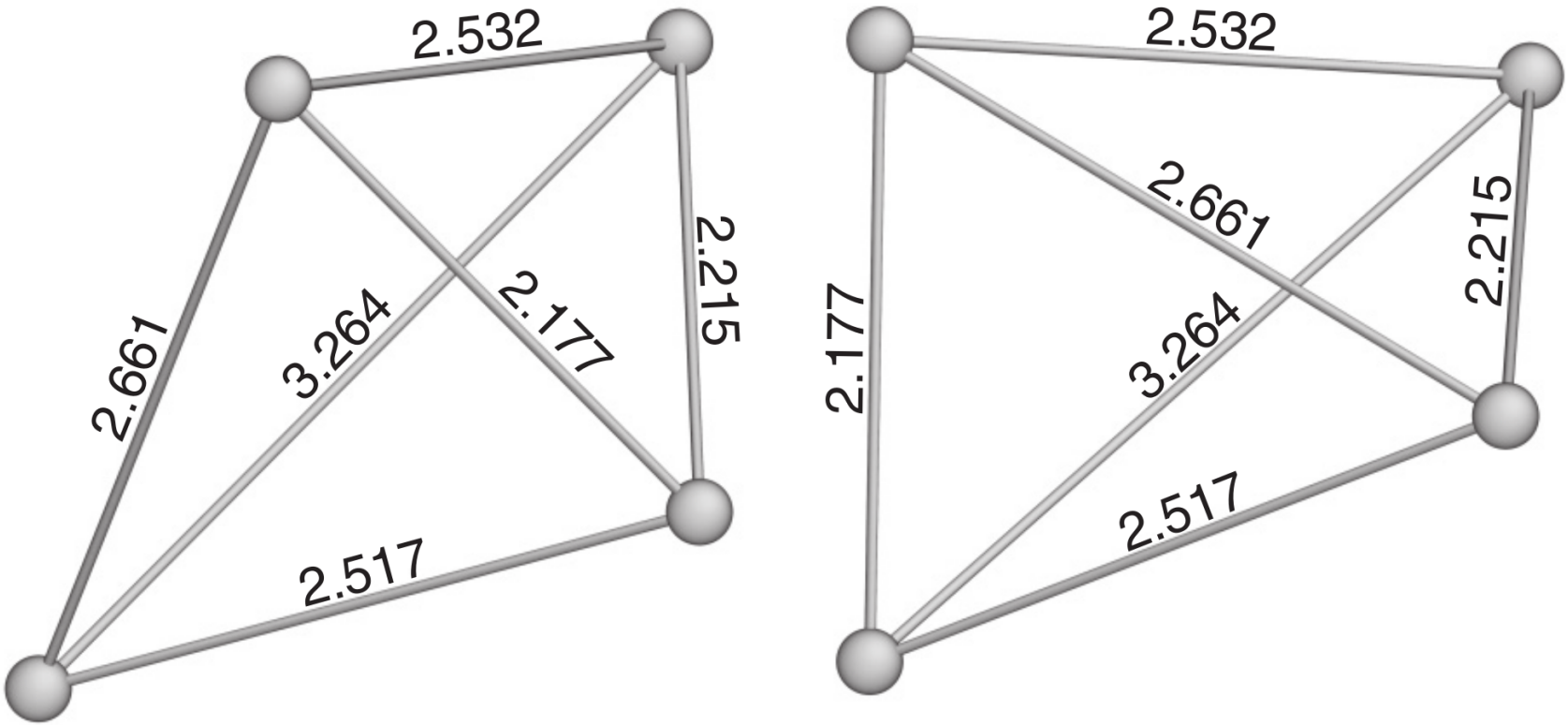
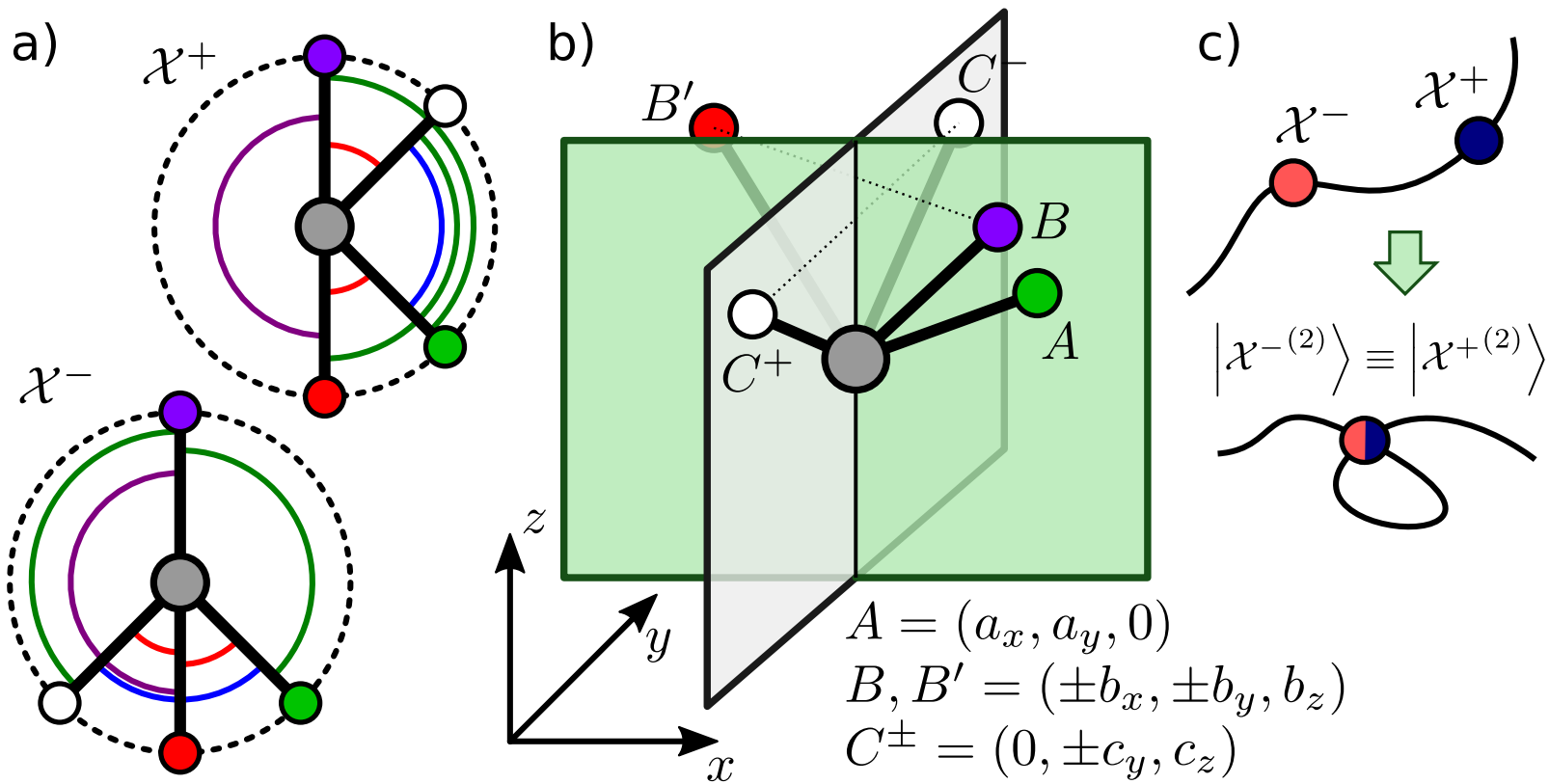


Figure from Bartók, Kondor, Csányi, PRB (2013)

Are these representations complete?

- It is well-known that 2-body correlations are ambiguous: can build tetrahedra with same pair distances that are different
- One can also build examples of pairs of environments that have the same 3B and 4B correlations. Problem becomes important as model accuracy is increased



Pozdniakov, Willatt, Bartók, Ortner, Csányi, **MC**, arxiv:2001.11696

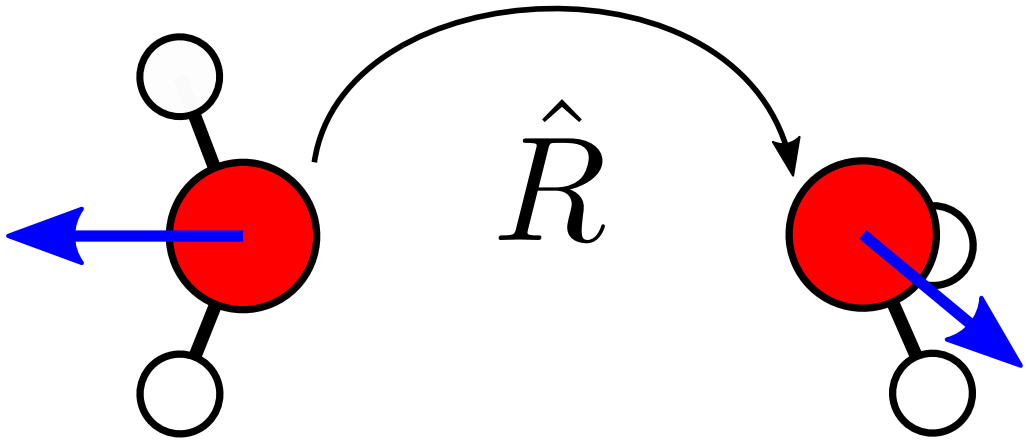
There are more things in heaven and earth, Horatio, than those transforming like a scalar

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.

$$\epsilon_\mu^\lambda(A_i) = \int \mathrm{d}X \, \langle \epsilon | X \rangle \, \langle X | A; \overline{\rho_i^{\otimes\nu}}; \lambda\mu \rangle$$

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



Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

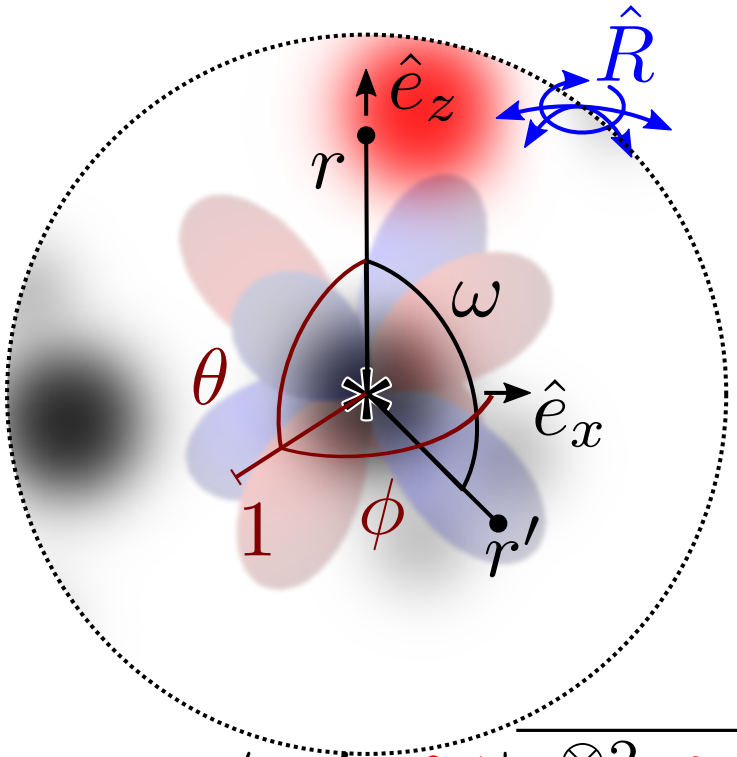
Symmetrized-field equivariants

- Include a $|\lambda\mu\rangle$ in the Haar integral to obtain $SO(3)$ equivariants

$$\int \mathrm{d}\hat{R}\, \langle \mathbf{r}|\, \hat{R}\, |\rho_i\rangle\, \langle \mathbf{r}'|\, \hat{R}\, |\rho_i\rangle\, \langle \mathbf{r}''|\, \hat{R}\, |\lambda\mu\rangle \rightarrow \langle rr'\omega\,\theta\phi|\overline{\rho_i^{\otimes 2}; \lambda\mu}\rangle$$

- Easier to compute by expanding the density in $R_n(r)\, Y_m^l(\hat{r})$: explicit power-spectrum-like representation

$$\langle n_1 l_1; n_2 l_2|\overline{\rho_i^{\otimes 2}; \lambda\mu}\rangle = \sum_m \langle n_1 l_1 m|\rho_i\rangle\, \langle n_2 l_2(\mu - m)|\rho_i\rangle\, \langle l_1 m; l_2(\mu - m)|\lambda\mu\rangle$$



$$\langle rr'\omega\theta\phi|\overline{\rho_i^{\otimes 2}; \lambda\mu}\rangle$$

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

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

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

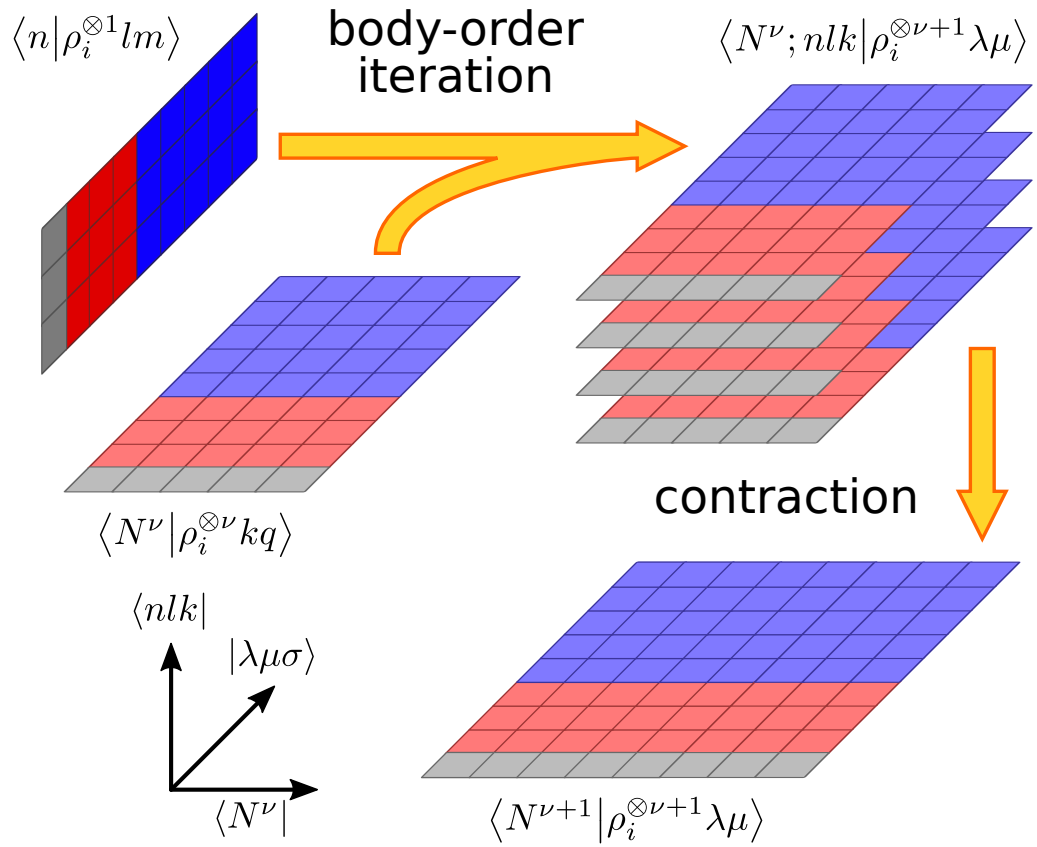
$$\langle \ldots; n_\nu l_\nu k_\nu; nlk | \overline{\rho_i^{\otimes(\nu+1)}; \sigma; \lambda\mu} \rangle = \delta_{\sigma((-1)^{l+k+\lambda} s)} c_{k\lambda} \times$$
$$\sum_{qm} \langle lm; kq | \lambda\mu \rangle < n || \overline{\rho_i^{\otimes 1}; lm} > \langle \ldots; n_\nu l_\nu k_\nu | \overline{\rho_i^{\otimes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

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

Nigam, Pozdnyakov, **MC**, https://arxiv.org/pdf/2007.03407 (2020)

NICE features for ML

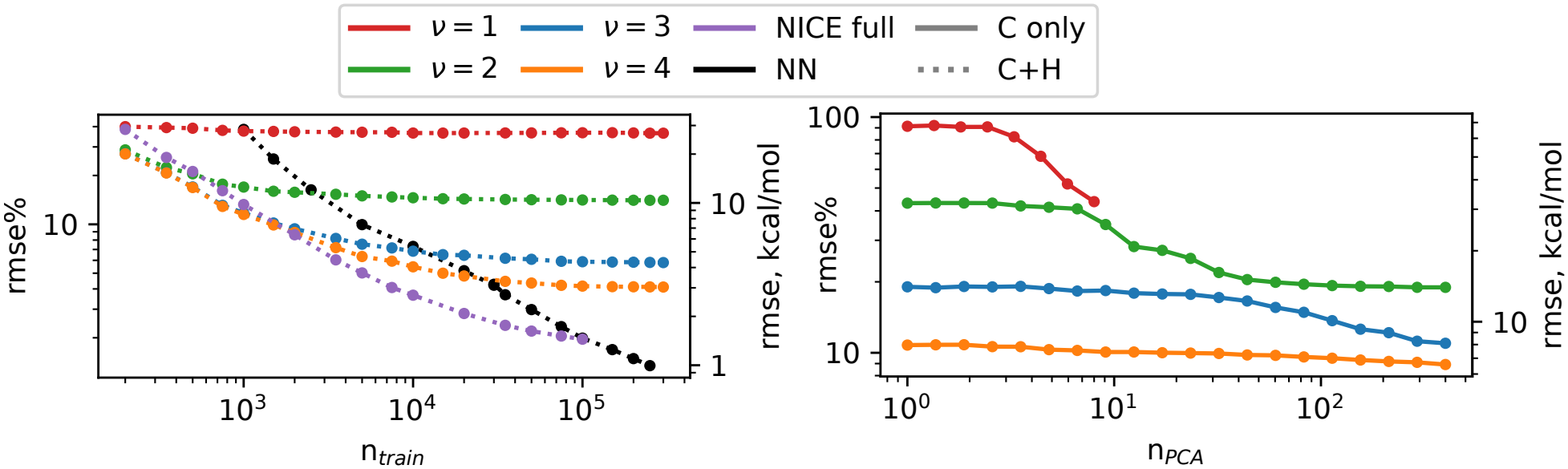
- Problem: number of features grows exponentially with ν
- Solution: an *N*-body iterative contraction of equivariants (NICE) framework
  - After each body order increase, the most relevant features are selected and used for the next iteration
  - Systematic convergence with ν and contraction truncation



Nigam, Pozdnyakov, **MC**, https://arxiv.org/pdf/2007.03407 (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

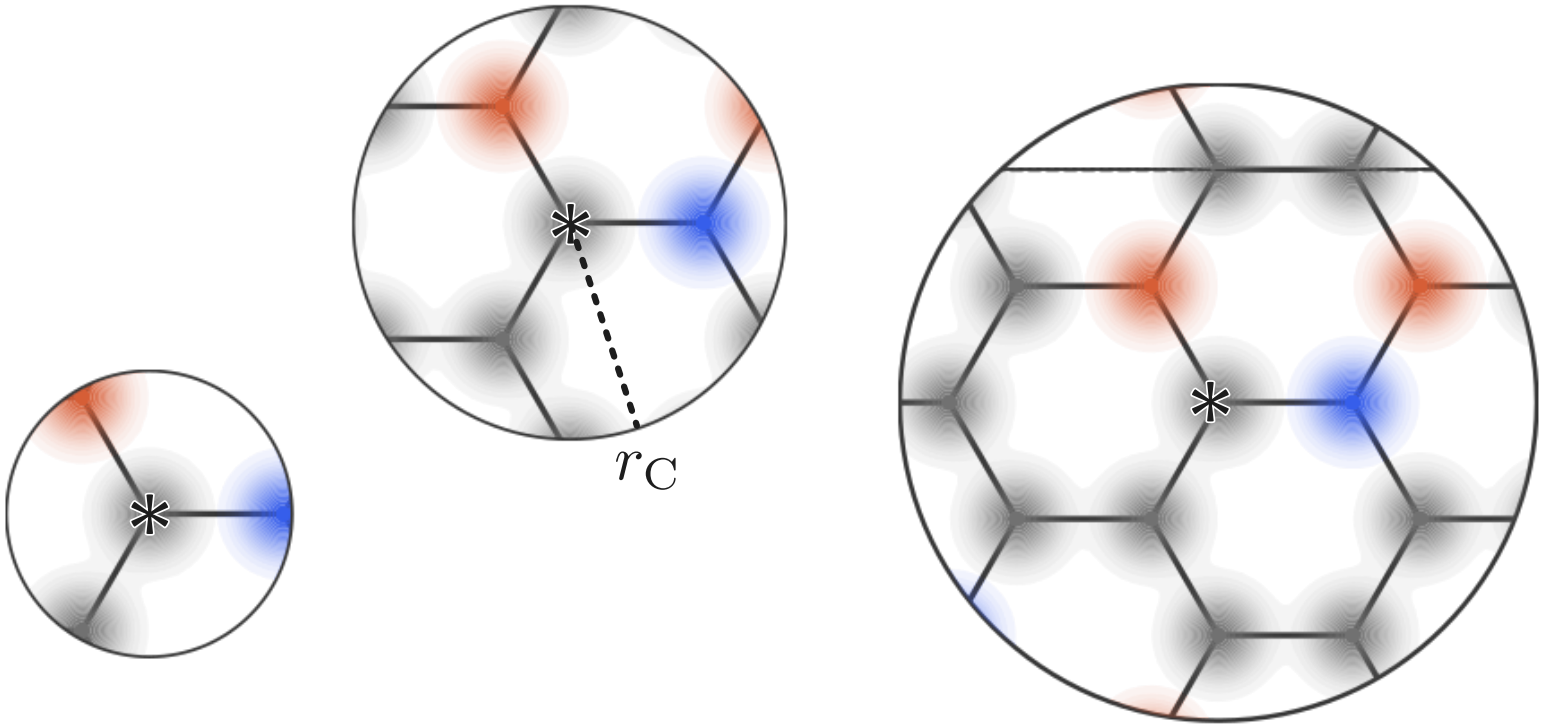


Nigam, Pozdnyakov, **MC**, https://arxiv.org/pdf/2007.03407 (2020)

The charged elephant in the other room

Understanding the range of interactions

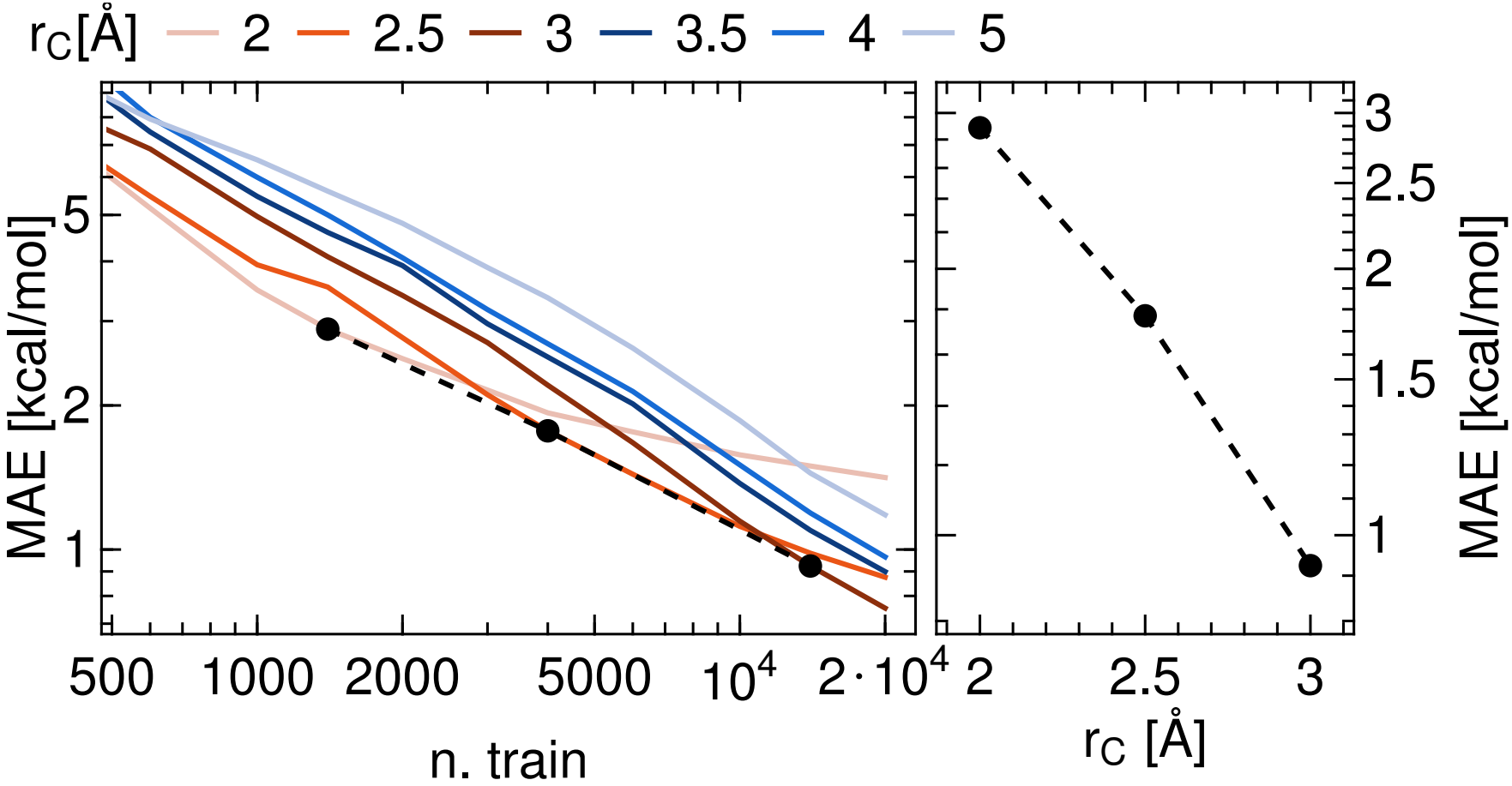
- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel $K(A,B) = \sum_i w_i K_i(A,B)$ yields the best of all worlds. Same results can be achieved by optimized radial scaling of $\langle \mathbf{r} | \overline{\rho_i^{\otimes \nu}} \rangle$



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

Understanding the range of interactions

- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel $K(A,B) = \sum_i w_i K_i(A,B)$ yields the best of all worlds. Same results can be achieved by optimized radial scaling of $\langle \mathbf{r} | \overline{\rho_i^{\otimes \nu}} \rangle$

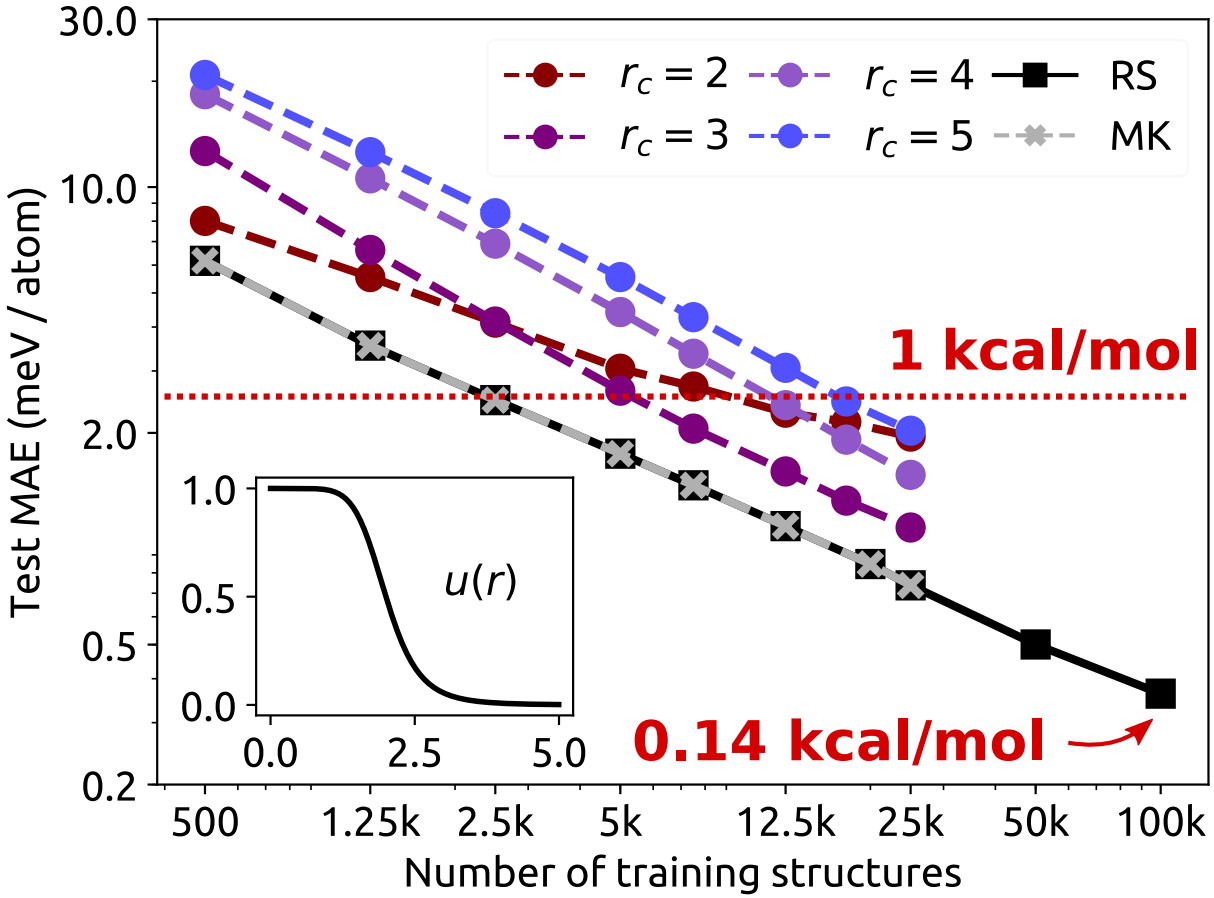


Willatt, Musil, **MC**, PCCP (2018)

Understanding the range of interactions

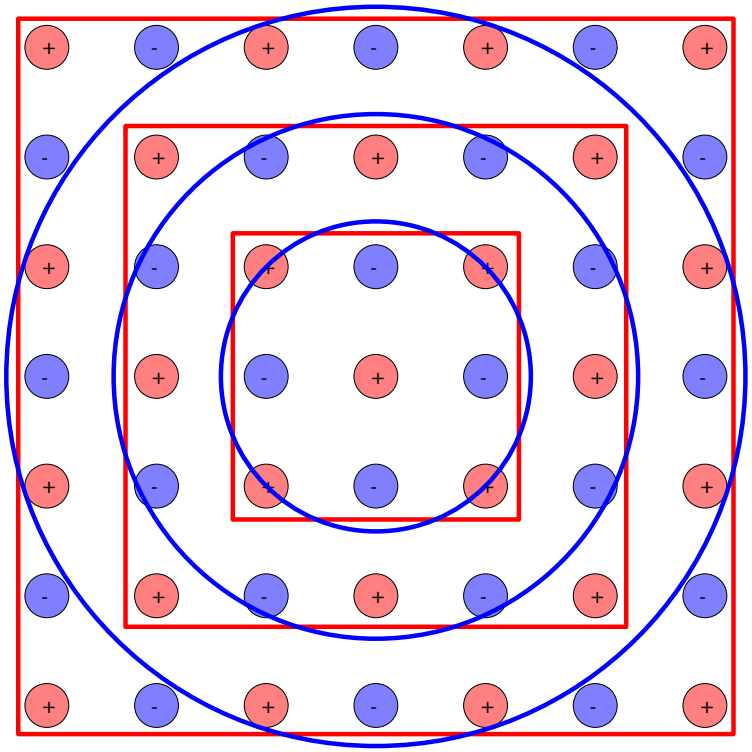
- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel $K(A,B) = \sum_i w_i K_i(A,B)$ yields the best of all worlds.

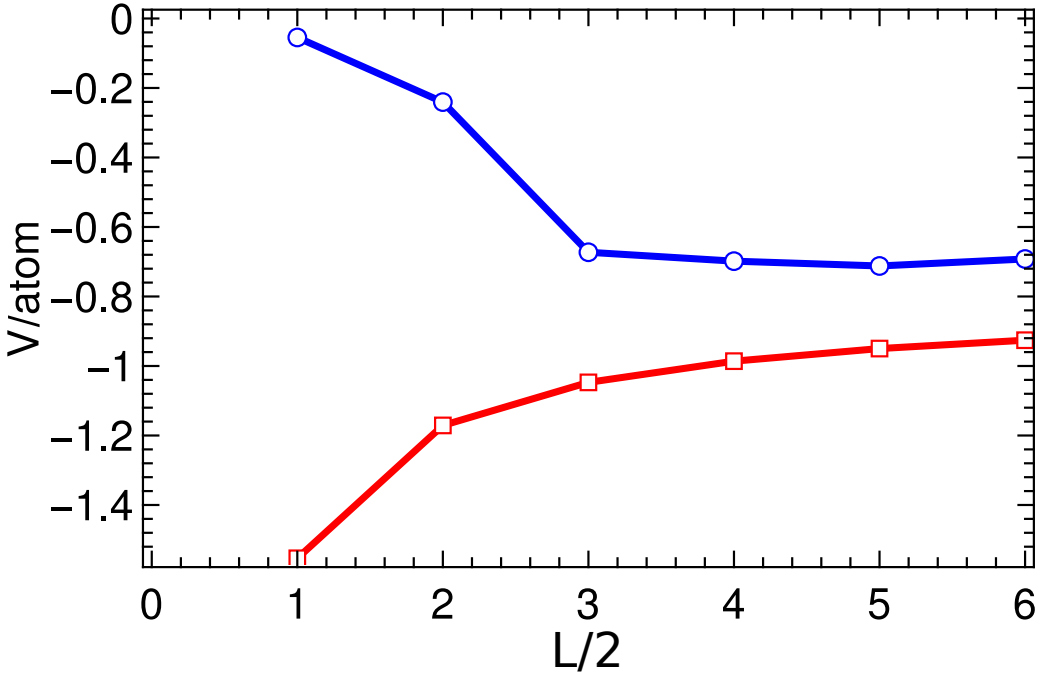
  Same results can be achieved by optimized radial scaling of $\langle \mathbf{r} | \overline{\rho_i^{\otimes \nu}} \rangle$



The problem with electrostatics

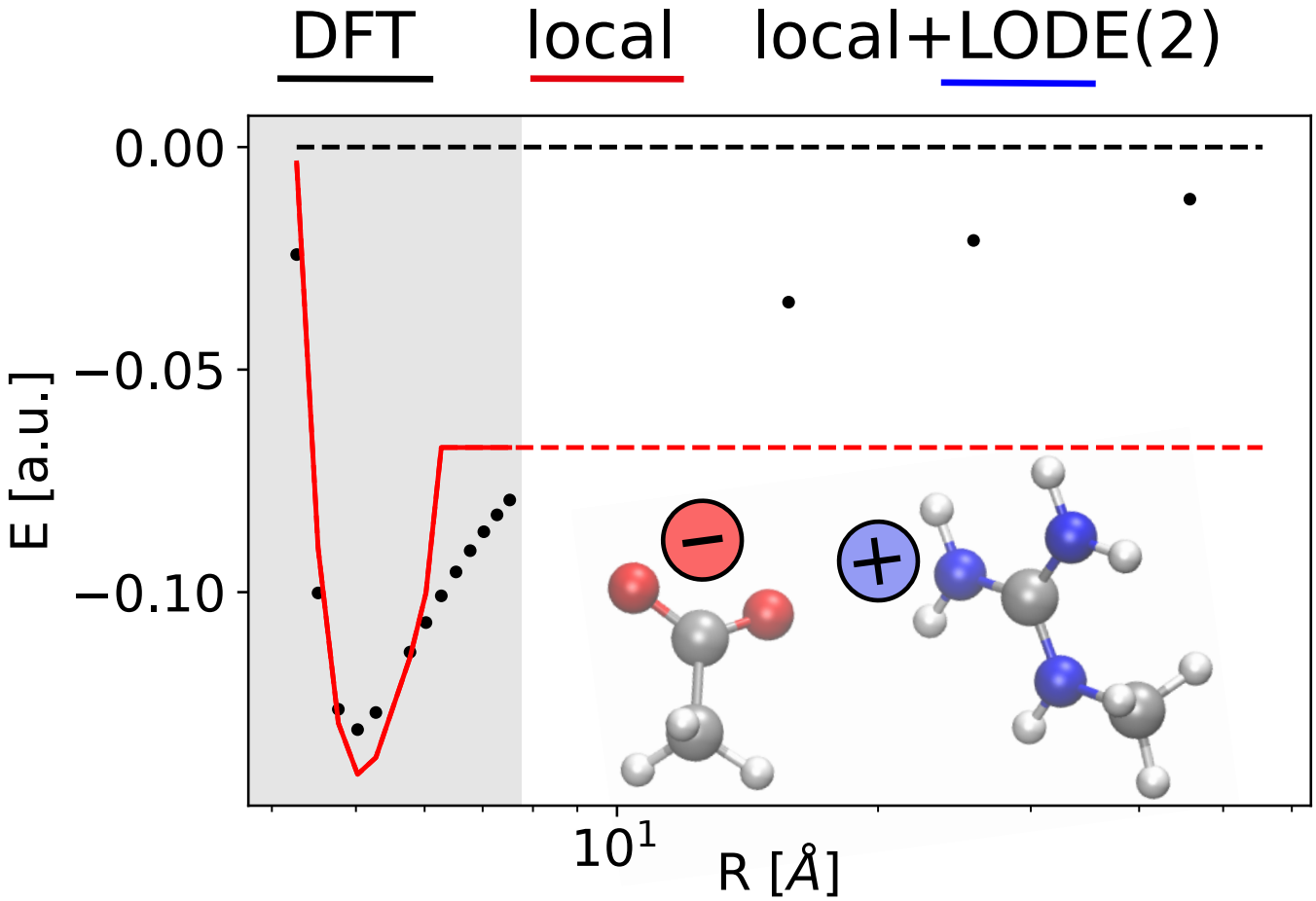
- Electrostatic interactions decay as $1/r$, leading to very slow convergence of properties with interaction cutoff
- Local ML models are hopeless to capture long-range effects, e.g. binding curves of charged fragments





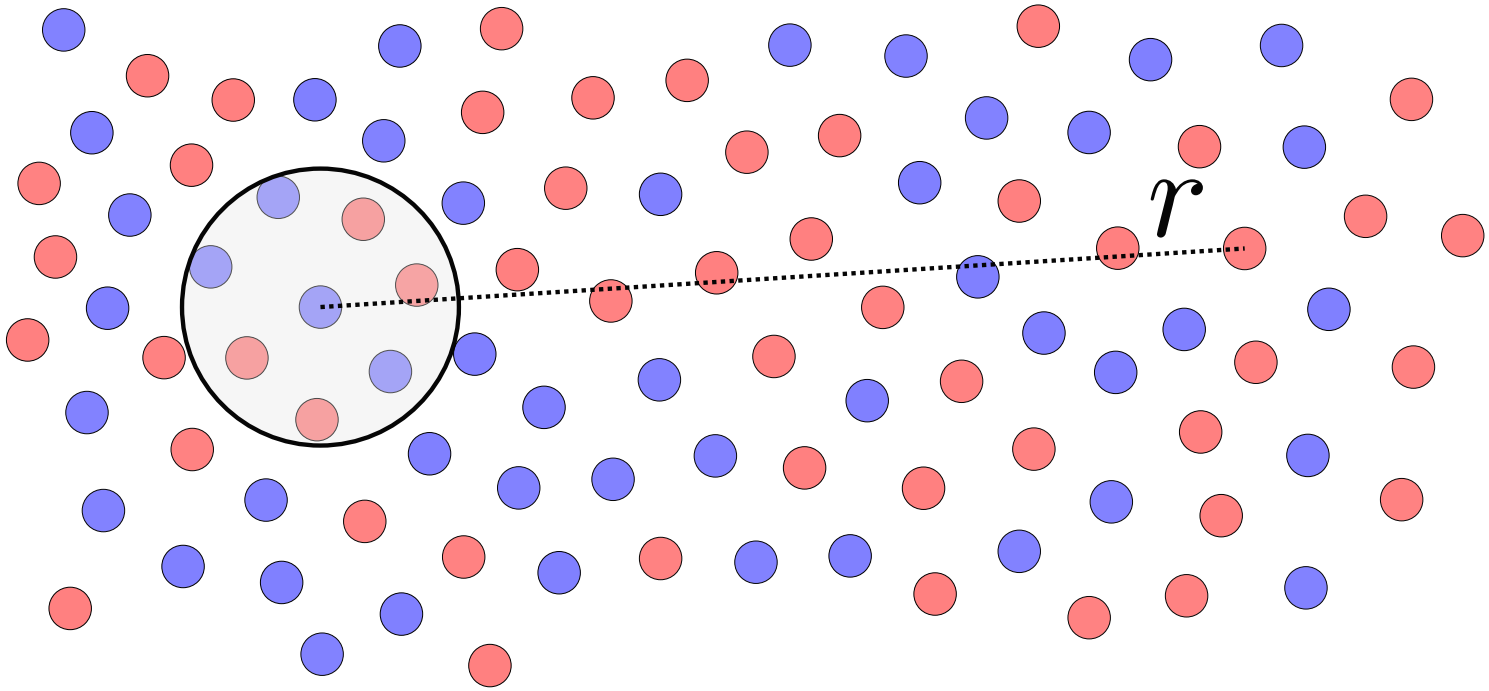
The problem with electrostatics

- Electrostatic interactions decay as $1/r$, leading to very slow convergence of properties with interaction cutoff
- Local ML models are hopeless to capture long-range effects, e.g. binding curves of charged fragments



Long-distance equivariant representation

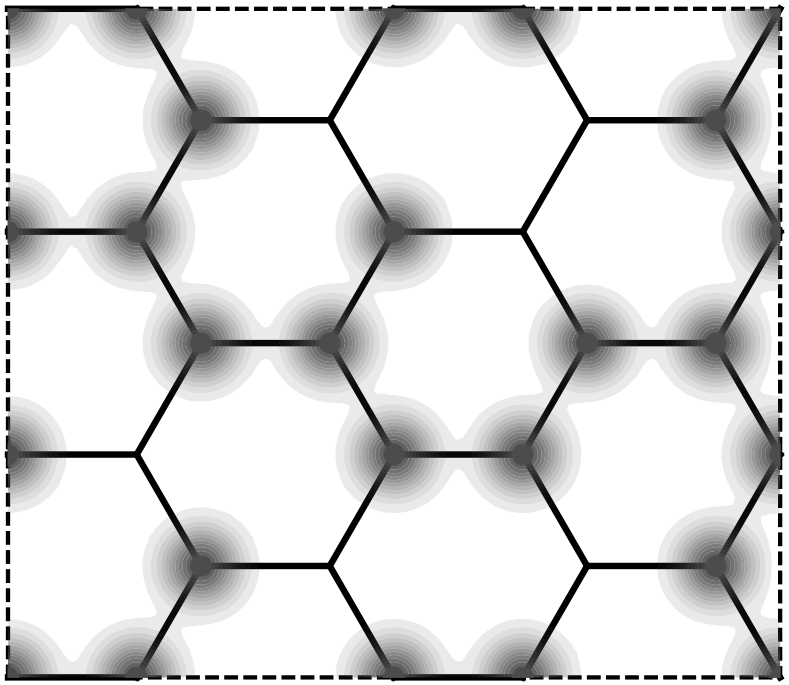
- Idea: get a local representation that reflects long-range correlations, with proper asymptotics
  1. Define an atom-density potential $\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$
  2. Do the usual gig: symmetrize, decompose locally, learn!
- Can be computed efficiently in reciprocal space



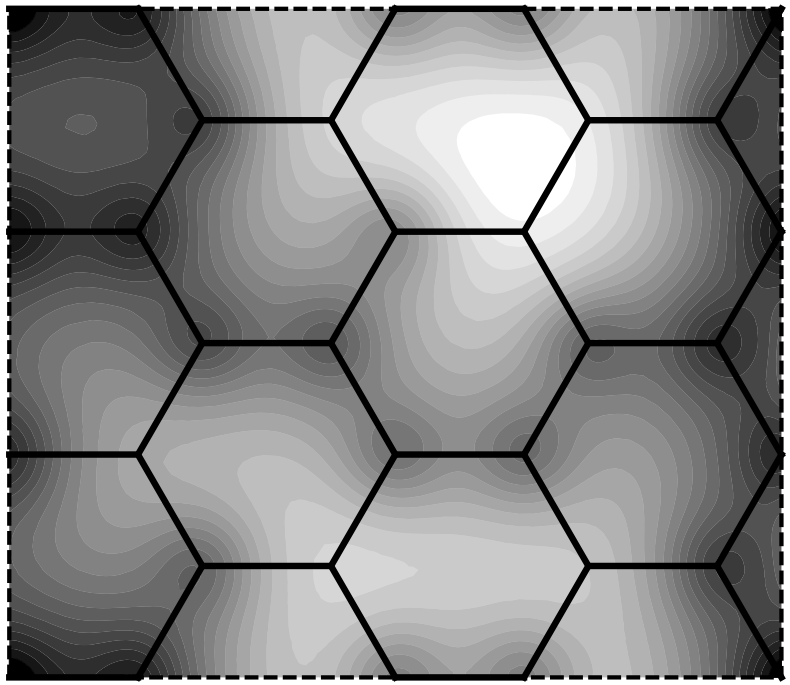
Grisafi, **MC**, JCP (2019)

Long-distance equivariant representation

- Idea: get a local representation that reflects long-range correlations, with proper asymptotics
  1. Define an 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. Do the usual gig: symmetrize, decompose locally, learn!
- Can be computed efficiently in reciprocal space



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

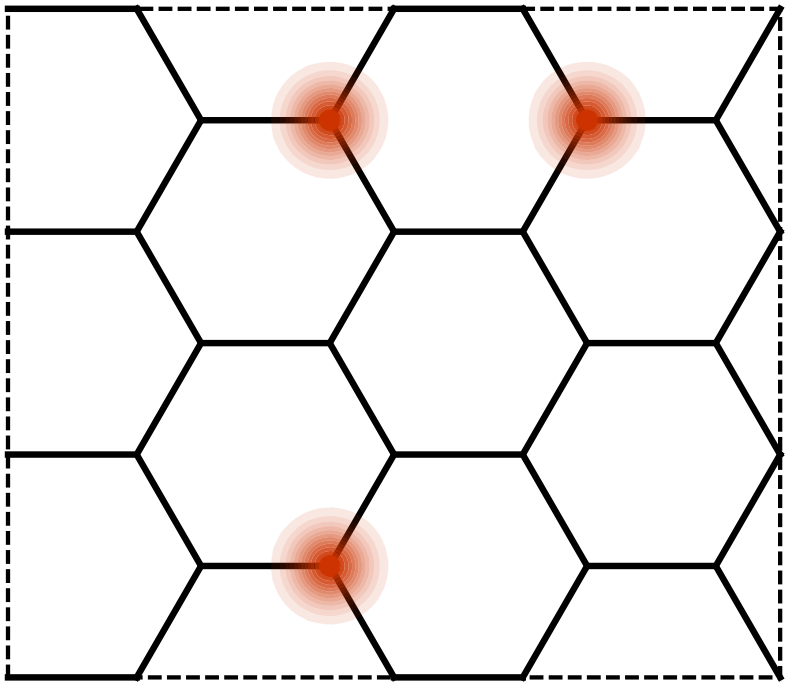


$$\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$$

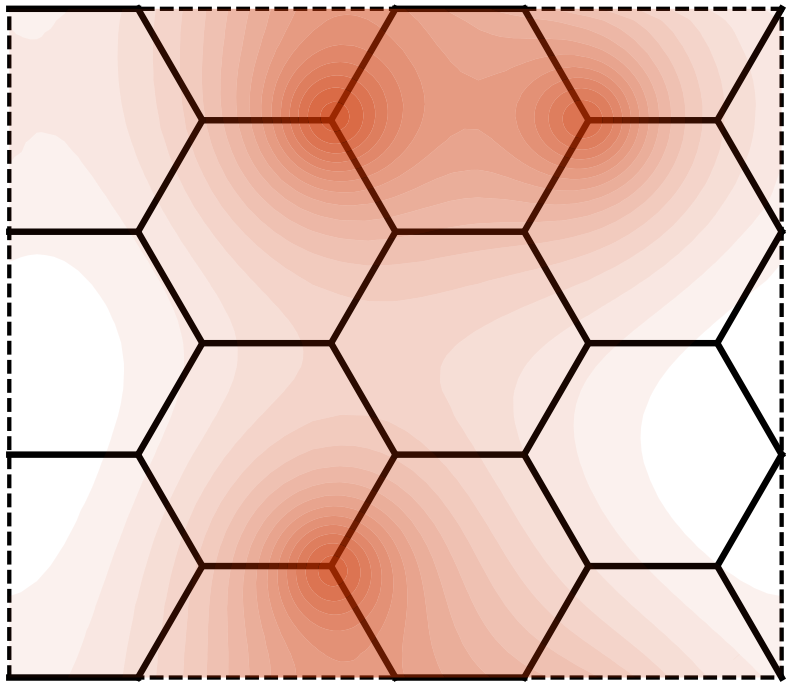
Grisafi, **MC**, JCP (2019)

Long-distance equivariant representation

- Idea: get a local representation that reflects long-range correlations, with proper asymptotics
  1. Define an 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. Do the usual gig: symmetrize, decompose locally, learn!
- Can be computed efficiently in reciprocal space



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

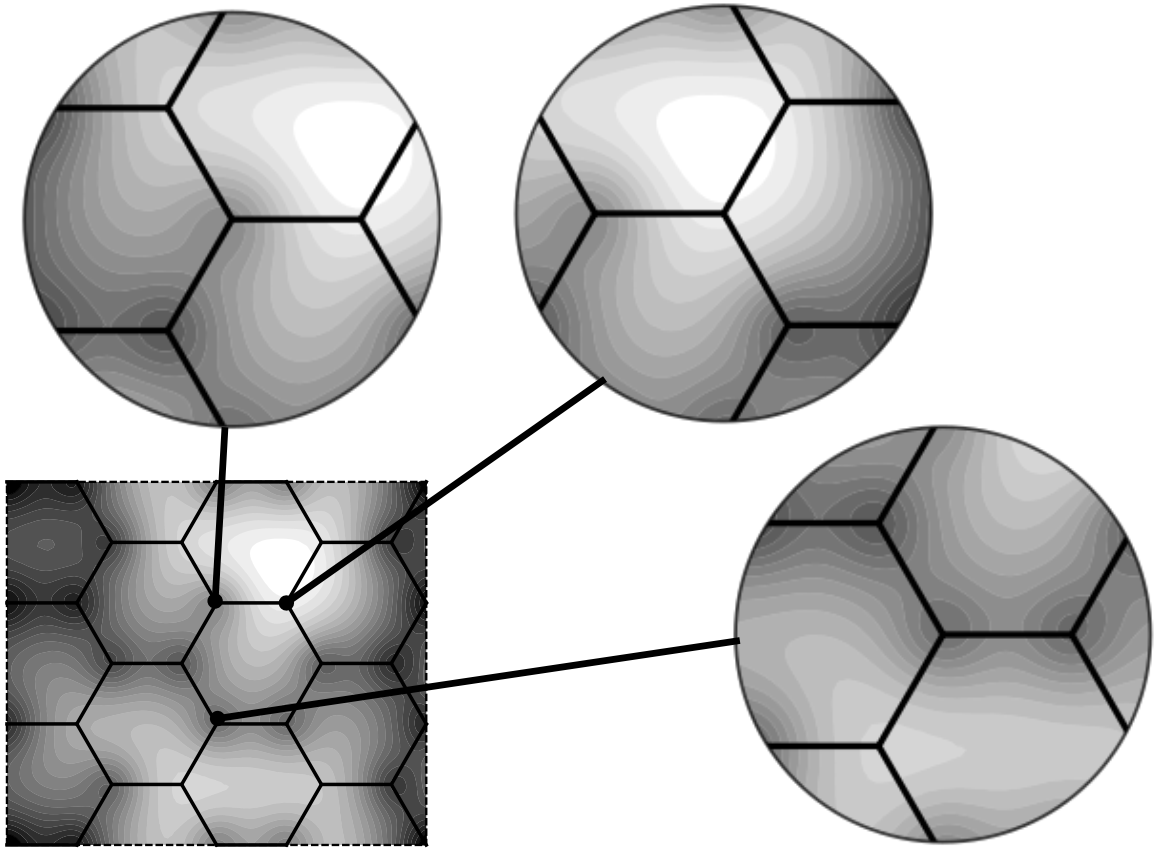


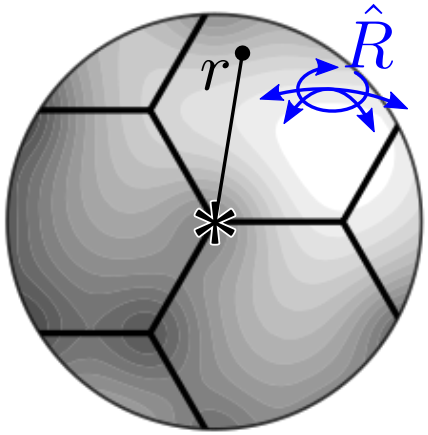
$$\langle a\mathbf{r}|V\rangle = \int \langle a\mathbf{r}'|\rho\rangle \,/\, |\mathbf{r}' - \mathbf{r}|\, \mathrm{d}\mathbf{r}'$$

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

Long-distance equivariant representation

- Idea: get a local representation that reflects long-range correlations, with proper asymptotics
  1. Define an 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. Do the usual gig: symmetrize, decompose locally, learn!
- Can be computed efficiently in reciprocal space

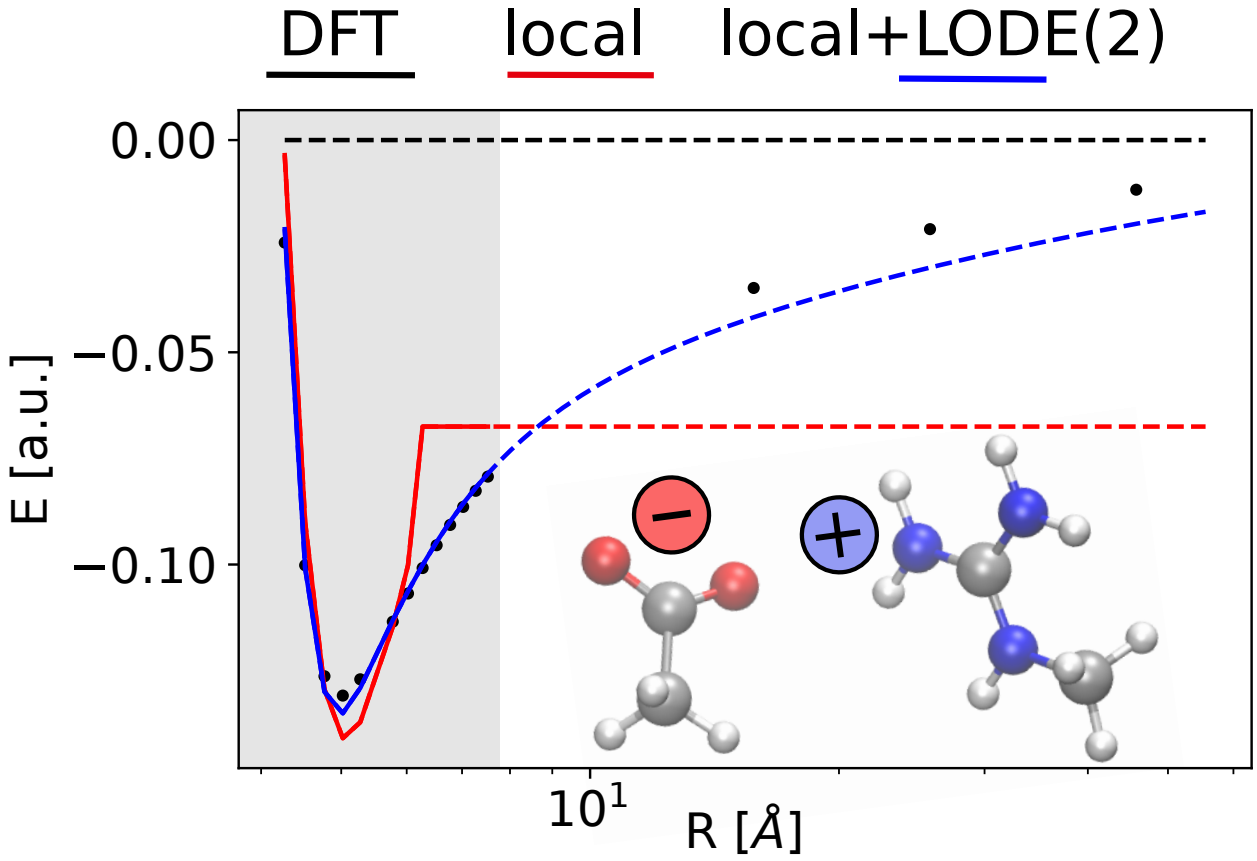




Grisafi, **MC**, JCP (2019)

Predicting binding curves for charged molecules

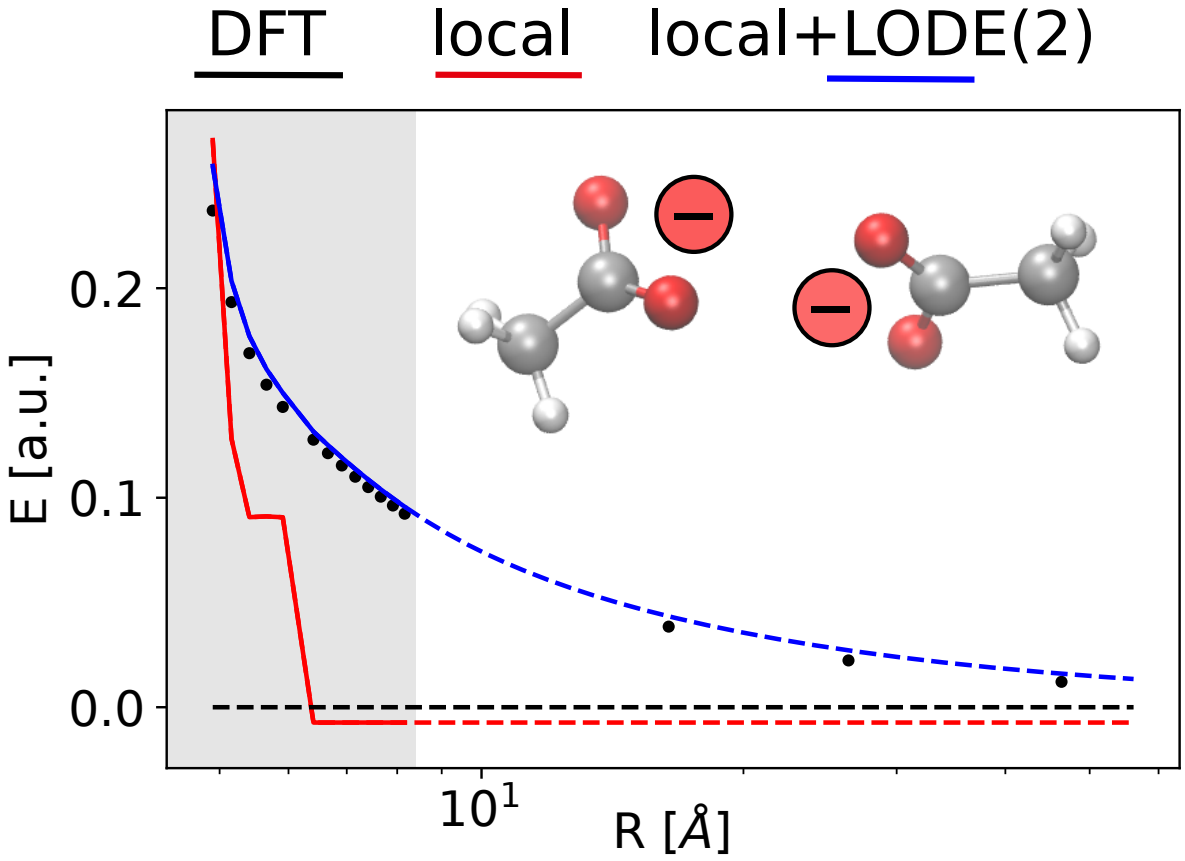
- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- Train on ~600 dimers, separations <8Å; test on ~60 dimers, up to $>$ 50Å
- Local ML alone fails, but SOAP+LODE combination extrapolates greatly for both monopole-monopole and monopole-dipole interactions



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

Predicting binding curves for charged molecules

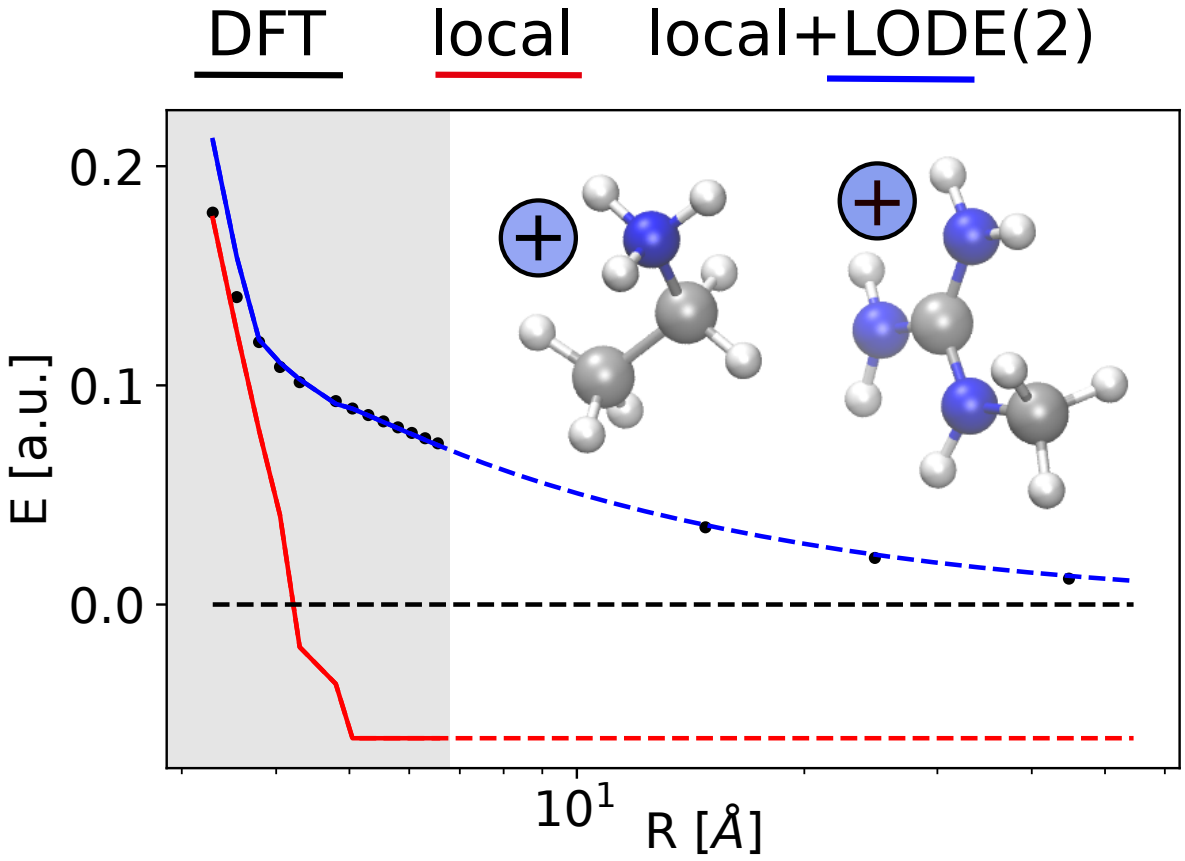
- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- Train on ~600 dimers, separations <8Å; test on ~60 dimers, up to $>$ 50Å
- Local ML alone fails, but SOAP+LODE combination extrapolates greatly for both monopole-monopole and monopole-dipole interactions



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

Predicting binding curves for charged molecules

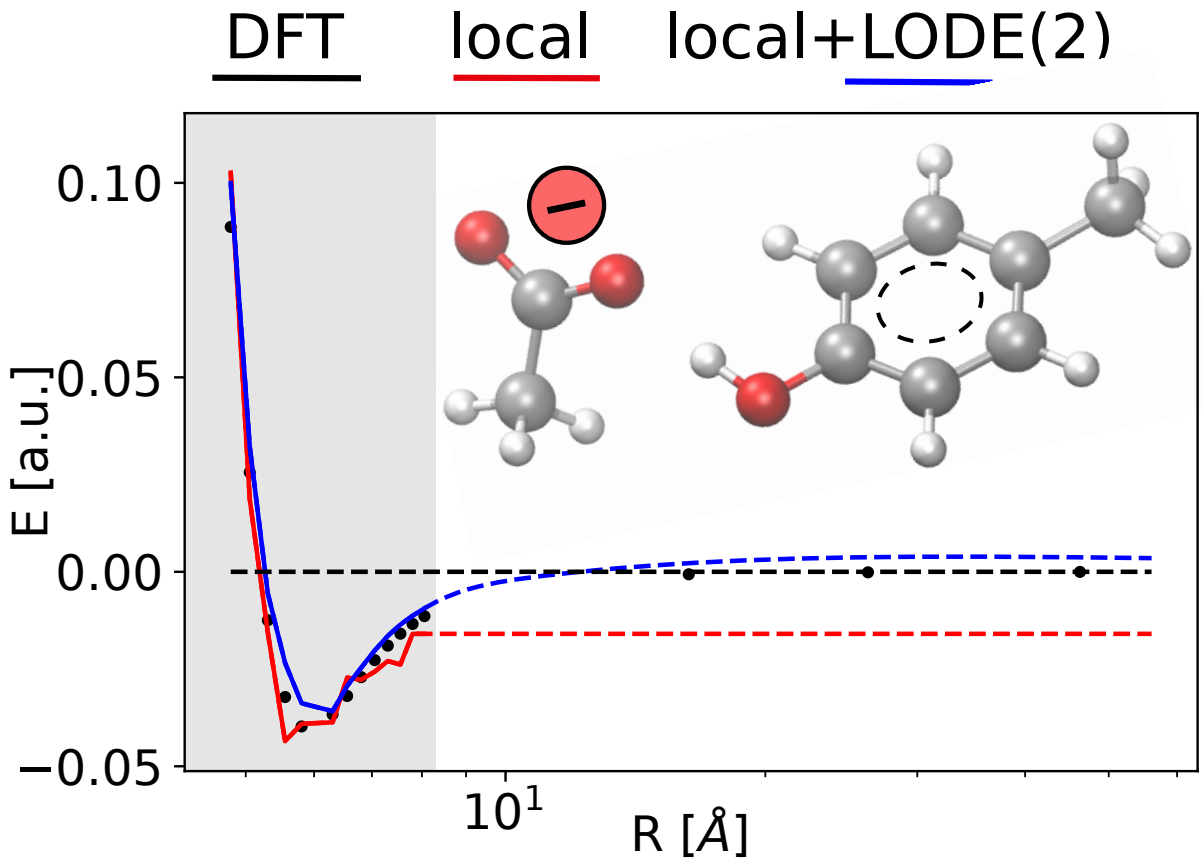
- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- Train on ~600 dimers, separations <8Å; test on ~60 dimers, up to $>$ 50Å
- Local ML alone fails, but SOAP+LODE combination extrapolates greatly for both monopole-monopole and monopole-dipole interactions



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

Predicting binding curves for charged molecules

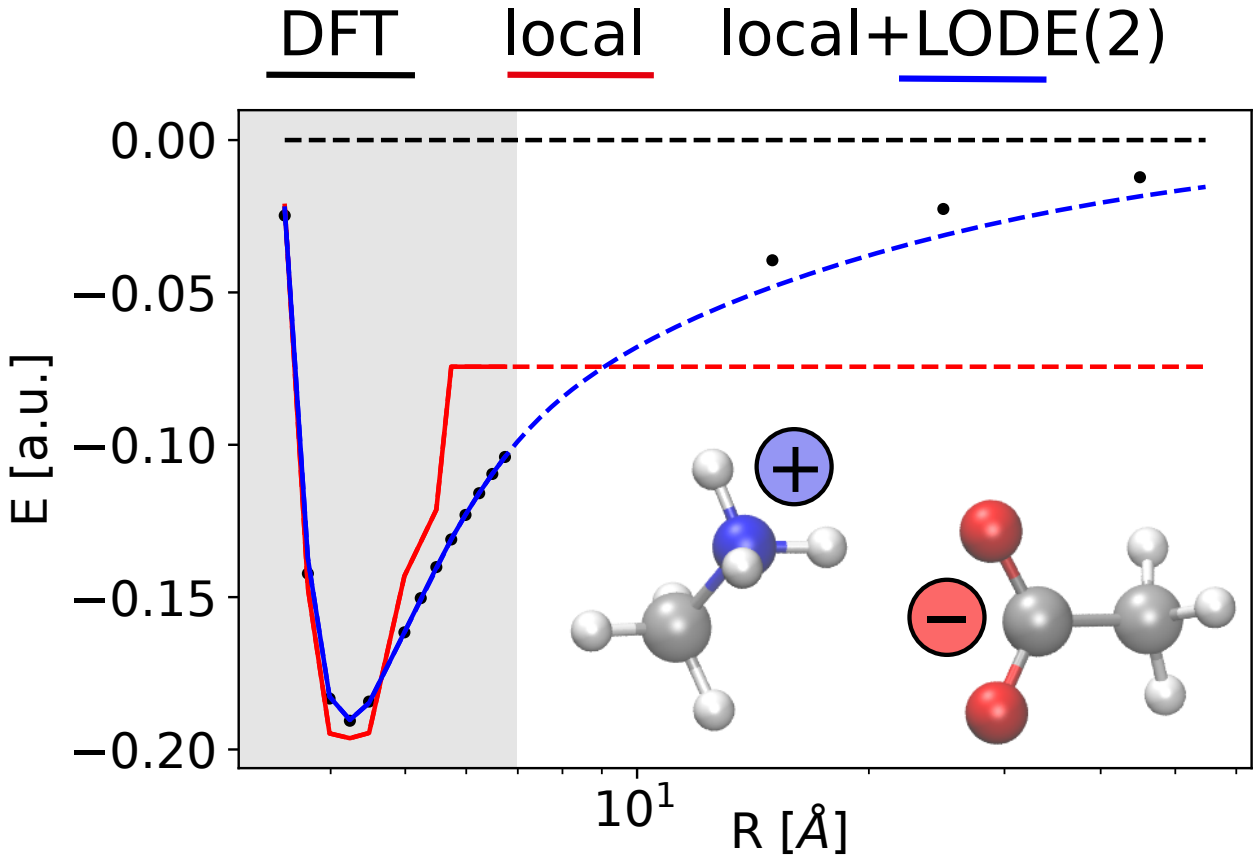
- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- Train on ~600 dimers, separations <8Å; test on ~60 dimers, up to $>$ 50Å
- Local ML alone fails, but SOAP+LODE combination extrapolates greatly for both monopole-monopole and monopole-dipole interactions



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

Predicting binding curves for charged molecules

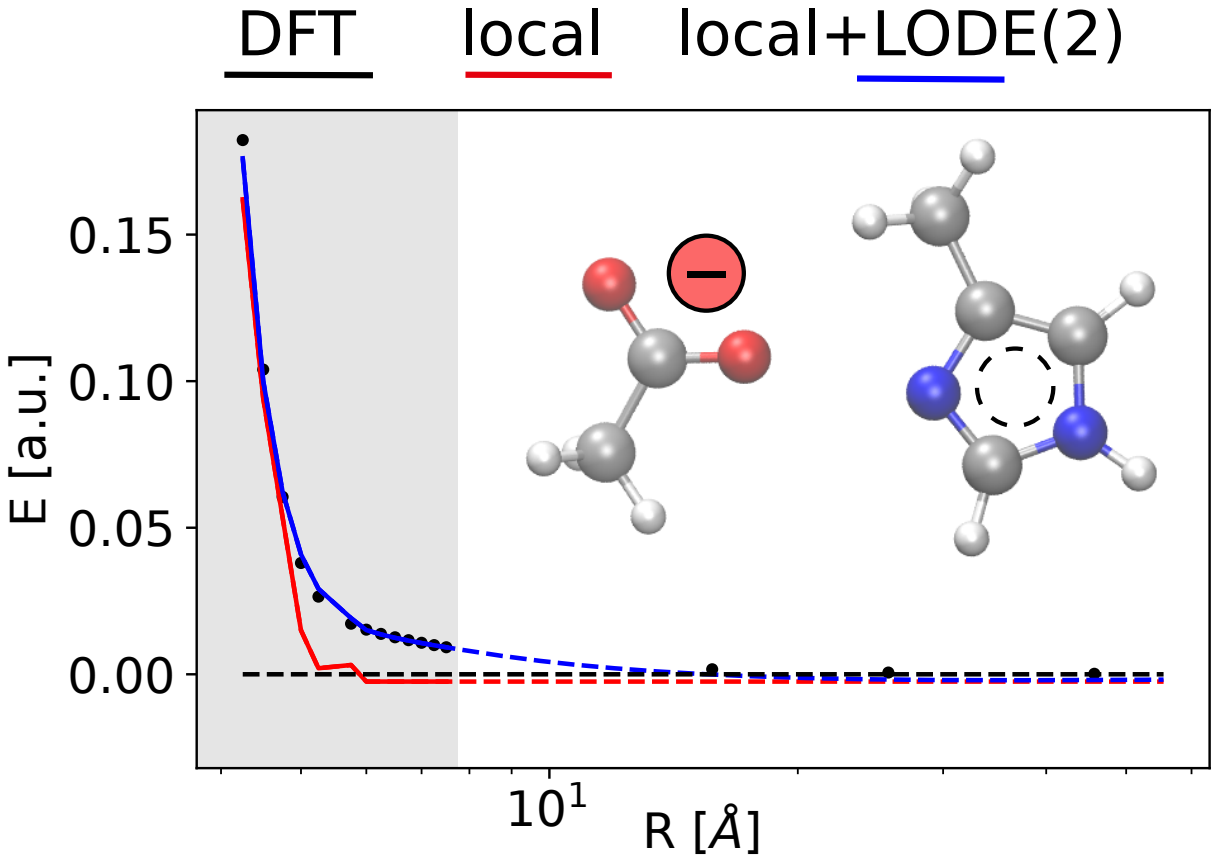
- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- Train on ~600 dimers, separations <8Å; test on ~60 dimers, up to $>$ 50Å
- Local ML alone fails, but SOAP+LODE combination extrapolates greatly for both monopole-monopole and monopole-dipole interactions



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

Predicting binding curves for charged molecules

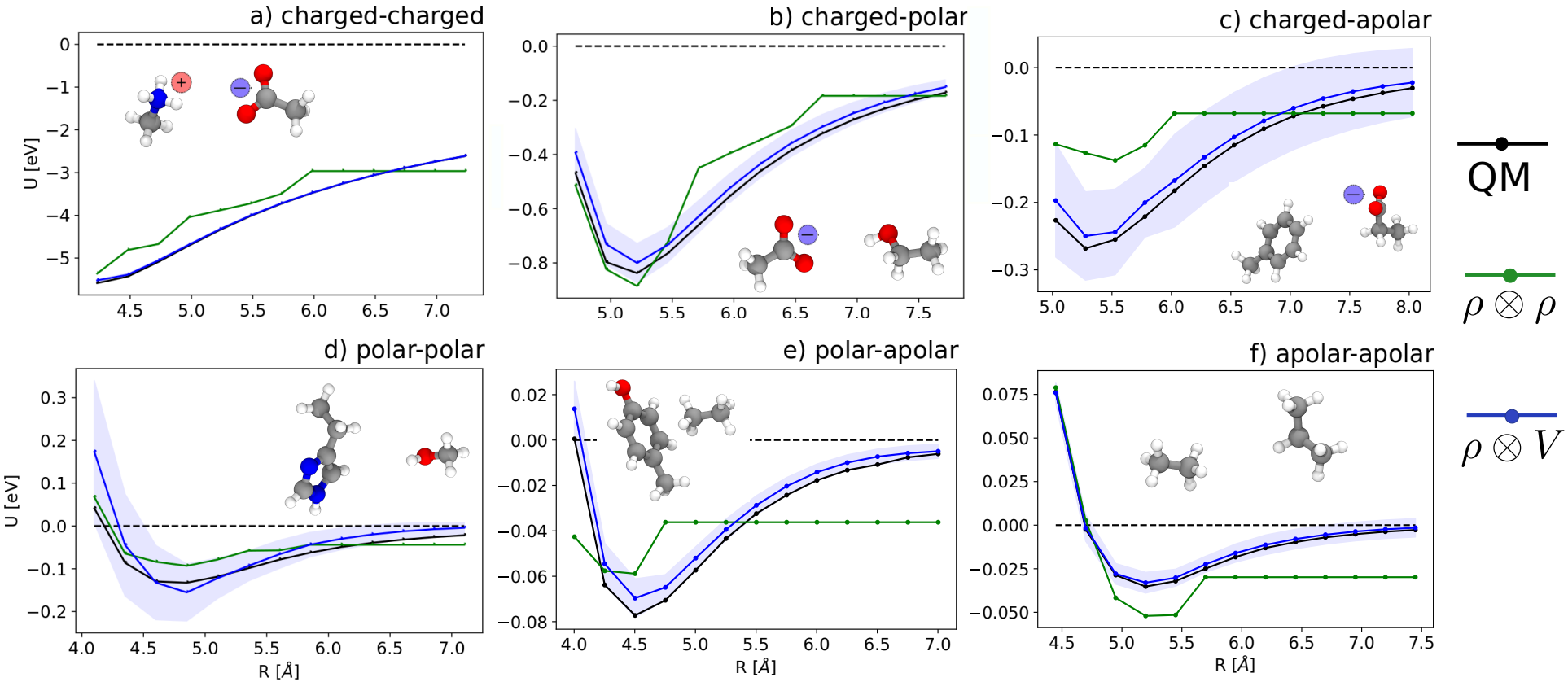
- A challenging test: rigid-molecule binding curves of charged dimers from the BioFragmentsDB
- Train on ~600 dimers, separations <8Å; test on ~60 dimers, up to $>$ 50Å
- Local ML alone fails, but SOAP+LODE combination extrapolates greatly for both monopole-monopole and monopole-dipole interactions



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

... and beyond

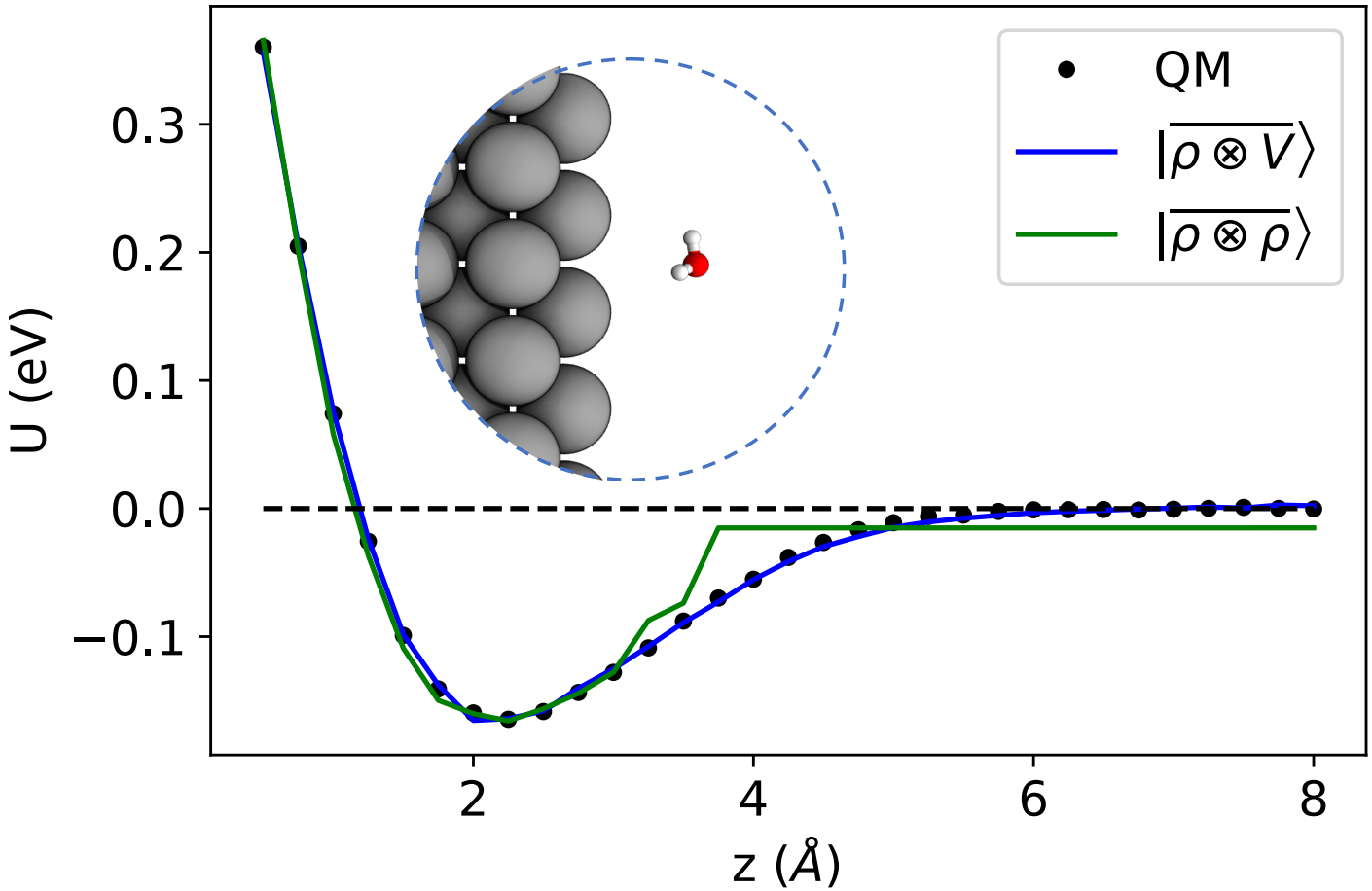
- "Multi-scale" LODE features $|\overline{\rho_i \otimes V_i}\rangle$ map to multipole electrostatics but enable learning all sorts of long-range physics



Grisafi, Nigam, **MC**, arXiv:2008.12122 (2020)

... and beyond

- "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, Nigam, **MC**, arXiv:2008.12122 (2020)

Conclusions & outlook

- Representations play a central role in any data-driven application
  - Symmetries of representations and target quantities are key
  - Locality, additivity, smoothness, conservation laws. . .
  - Incorporating long-range interactions in a physics-inspired way
- Very useful to keep the treatment abstract, and to understand whether different representations are *substantially* different, or just a matter of practical implementation

Conclusions & outlook

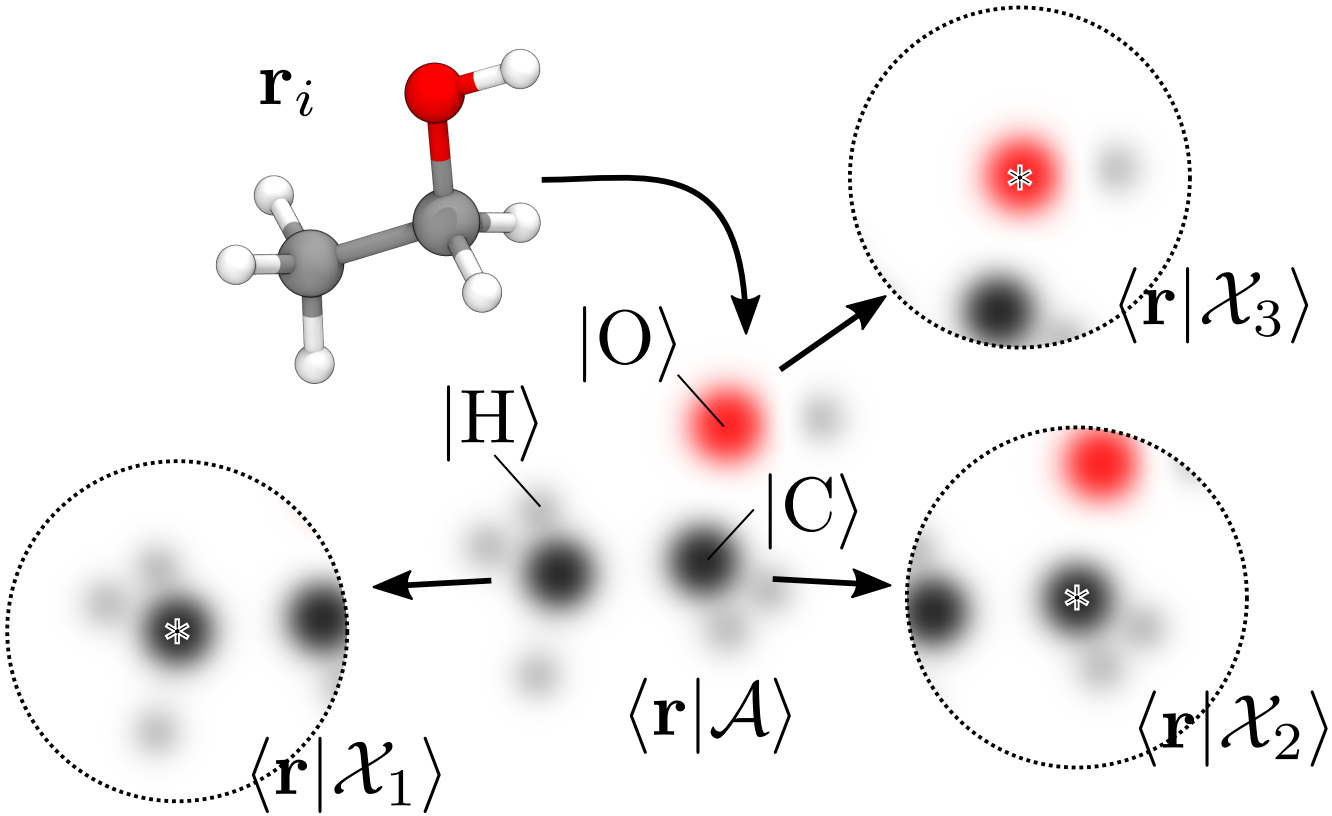
- Representations play a central role in any data-driven application
  - Symmetries of representations and target quantities are key
  - Locality, additivity, smoothness, conservation laws. . .
  - Incorporating long-range interactions in a physics-inspired way
- Very useful to keep the treatment abstract, and to understand whether different representations are *substantially* different, or just a matter of practical implementation

Deep connections between most representations . . . . . Willatt et al. JCP (2019)
Strategies to reduce the computational cost . . . . . Imbalzano et al. J. Chem. Phys. (2018)
Feature optimization: efficiency and insight . . . . . Willatt et al. PCCP (2018)
Fast and accurate error estimation . . . . . Musil et al. JCTC (2019)
Symmetry-adapted regression for tensors: . . . . . Grisafi et al., Phys. Rev. Lett. (2018)
Completeness of representations . . . . . Podznyakov et al. arXiv:2001.11696
NICE features . . . . . Nigam et al., arXiv:2007.03407
Comparing features . . . . . Goscinski et al., arXiv:2009.02741
Multi-scale equivariants . . . . . Grisafi et al., arXiv:2008.12122

https://tinyurl.com/ceriotti-sissa-ictp-2020

A data-driven periodic table of the elements

- 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!



A data-driven periodic table of the elements

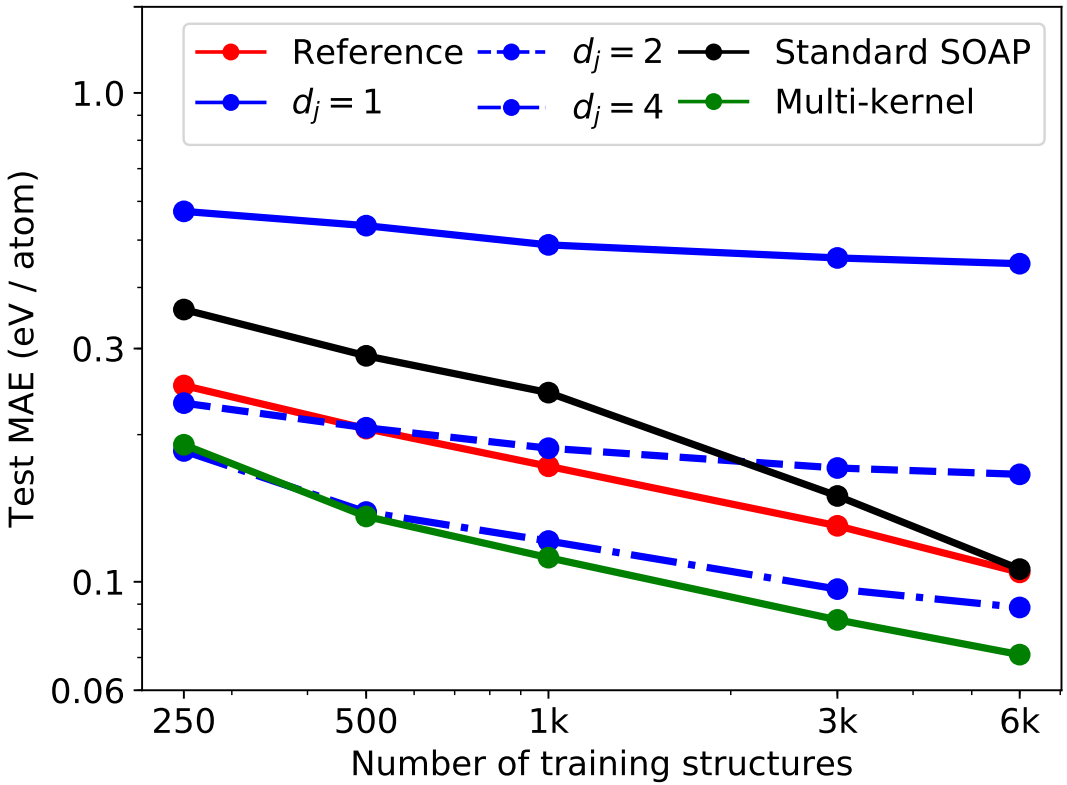
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective 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

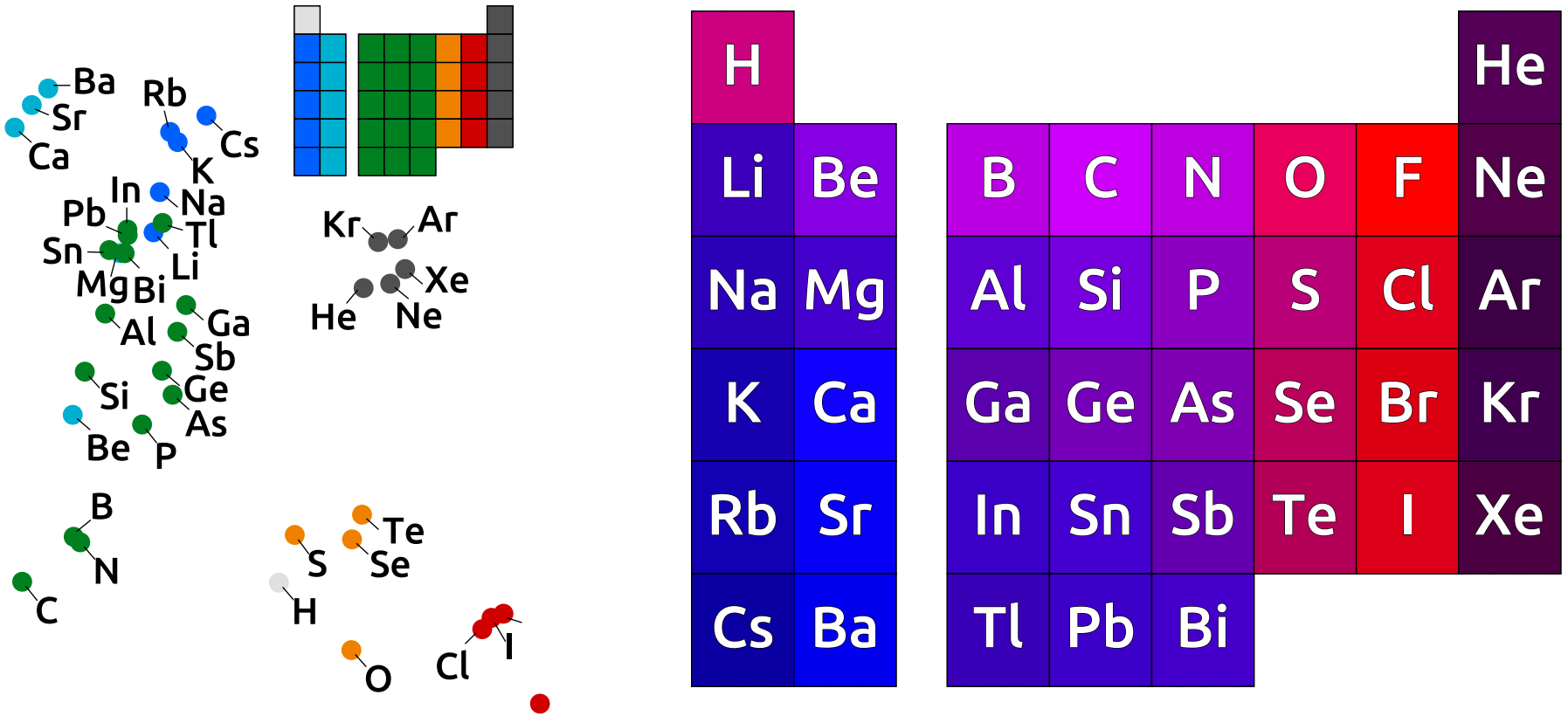
A data-driven periodic table of the elements

- 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!



A data-driven periodic table of the elements

- 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, **MC**, PCCP (2018); *[data: Elpasolites, von Lilienfeld&C]*