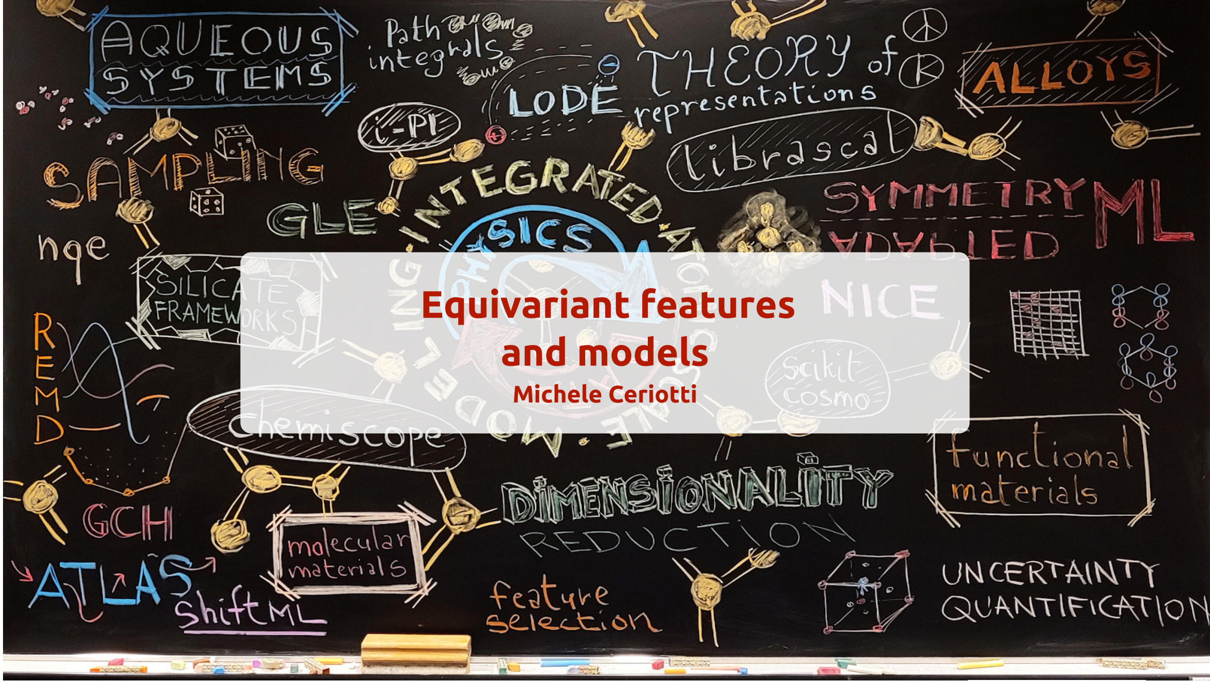
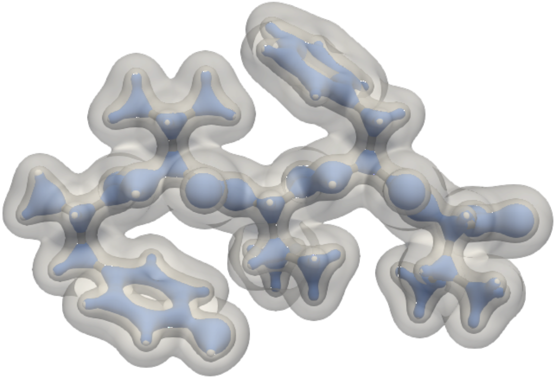


Equivariant features and models

Michele Ceriotti

Outline

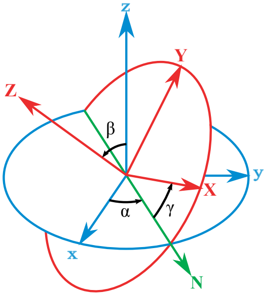
- Rotations and the rotation group
- Rotational equivariance for linear and kernel models
- The design space of equivariant representations
- Permutational equivariance: multi-center descriptors and message passing

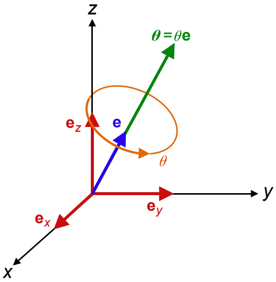


Rotations and equivariance

3D rotations and the rotation group

- Rotations are isometries of 3D space that preserve the origin and distances
- Rotations form a *group* $SO(3) = \left\{\hat{R}\right\}$: the composition of two rotations yields another rotation, each rotation has an inverse operation that is also a rotation and there is an identity operation
- Rotations are not commutative, so in general $\hat{R}_1\hat{R}_2 \neq \hat{R}_2\hat{R}_1$
- Rotations can be parameterized by a set of continuous parameters, e.g. the orientation of an axis and an angle of rotation around it, or a set of *Euler angles*. A hellscape of alternative conventions.

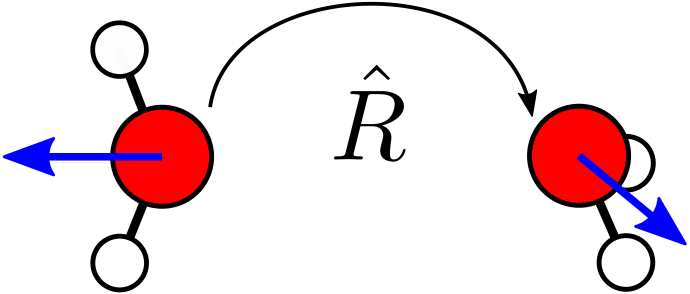




How Cartesian functions transform under rotations

- 3D vectors $\mathbf{u} = (x, y, z)$ transform under the action of a rotation $\hat{R}$ as the application of a rotation matrix, $\hat{R}\mathbf{u} = \mathbf{R}\left(\hat{R}\right)\mathbf{u}$
- For general functions of the coordinates, things are more complicated. For instance, x^2 does not transform in a simple way
- It is possible to identify groups of polynomials that transform into each other in a prescribed way. For instance,

$$\mathbf{U} = \begin{pmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{pmatrix} \rightarrow \hat{R}\mathbf{U} = \mathbf{R}\mathbf{U}\mathbf{R}^T$$

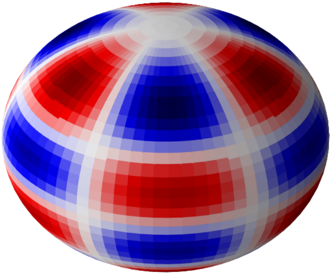
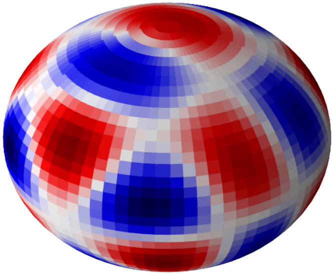


Spherical harmonics and irreducible representations

- Spherical harmonics are sets of polynomials in $(x/r, y/r, z/r)$ that transform in the most concise way possible under rotations.
- Spherical harmonics Y_l^m are labeled by two indices $l \geq 0$ and $-l \leq m \leq l$ (so there are $2l+1$ Y_l^m that are mixed by a rotation). These $(2l+1)$-sized vectors cannot be further disentangled: they are said to form *irreducible representations*
- Products of spherical harmonics can be re-orthogonalized by *Clebsch-Gordan products*
- Rotations are enacted by a generalization of **R**, so-called *Wigner D matrices*,

$$Y_l^m\left(\mathbf{R}\left(\hat{R}\right)\hat{\mathbf{r}}\right) = \sum_{m'} D^l_{mm'}\left(\hat{R}\right) Y_l^{m'}$$

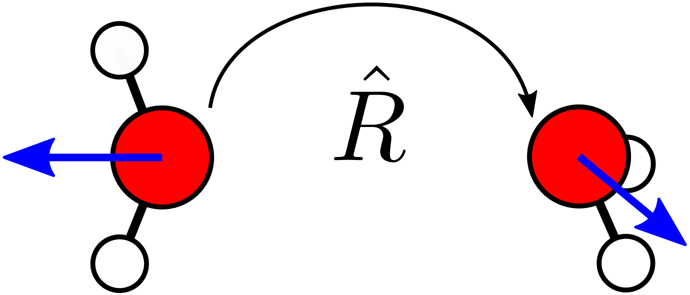
- Another hellscape of arbitrary conventions: complex vs real, different normalizations

Equivariance for your favourite model

- Want to learn vectors or general tensors? They should transform with the molecular geometry
- Equivariance constraints the structure of the model
- Equivariance ensures symmetry-adapted behavior

$$y_\alpha\left(\hat{R}A_i\right) = \sum_{\alpha'} R_{\alpha\alpha'} y_{\alpha'}\left(A_i\right)$$

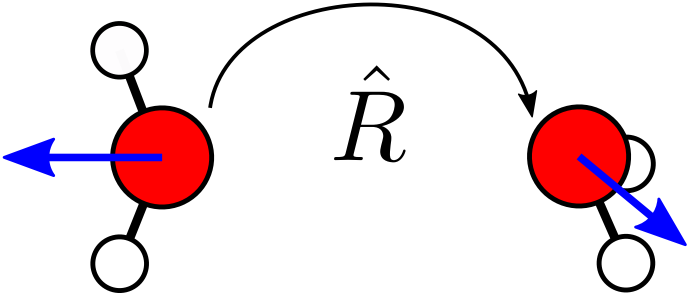


T. Cohen, M. Welling, ICML (2016); Glielmo et al. PRB (2017); Grisafi et al., PRL (2018)

Equivariance for your favourite model

- Want to learn vectors or general tensors? They should transform with the molecular geometry
- Equivariance constraints the structure of the model
- Equivariance ensures symmetry-adapted behavior

$$y_\alpha(A_i) = \sum_q w_q^\alpha \xi_q^\alpha(A_i) \equiv \sum_q \langle y;\alpha|q\rangle\langle q|A_i;\alpha\rangle$$

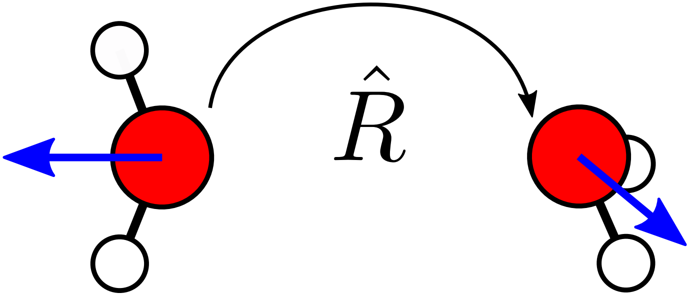


T. Cohen, M. Welling, ICML (2016); Glielmo et al. PRB (2017); Grisafi et al., PRL (2018)

Equivariance for your favourite model

- Want to learn vectors or general tensors? They should transform with the molecular geometry
- Equivariance constraints the structure of the model
- Equivariance ensures symmetry-adapted behavior

$$y_\alpha\left(\hat{R}A_i\right) = \sum_q w_q^\alpha \xi_q^\alpha\left(\hat{R}A_i\right) \equiv \sum_q \langle y;\alpha|q\rangle\,\langle q|\hat{R}A_i;\alpha\rangle$$

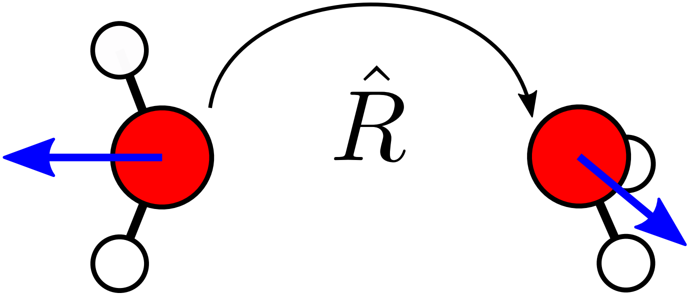


T. Cohen, M. Welling, ICML (2016); Glielmo et al. PRB (2017); Grisafi et al., PRL (2018)

Equivariance for your favourite model

- Want to learn vectors or general tensors? They should transform with the molecular geometry
- Equivariance constraints the structure of the model
- Equivariance ensures symmetry-adapted behavior

$$y_\alpha\left(\hat{R}A_i\right) = \sum_q \langle y|q\rangle \langle q|\hat{R}A_i;\alpha\rangle = \sum_q \langle y|q\rangle \sum_{\alpha'} R_{\alpha\alpha'} \langle q|A_i;\alpha'\rangle = \sum_{\alpha'} R_{\alpha\alpha'} y_{\alpha'}(A_i)$$

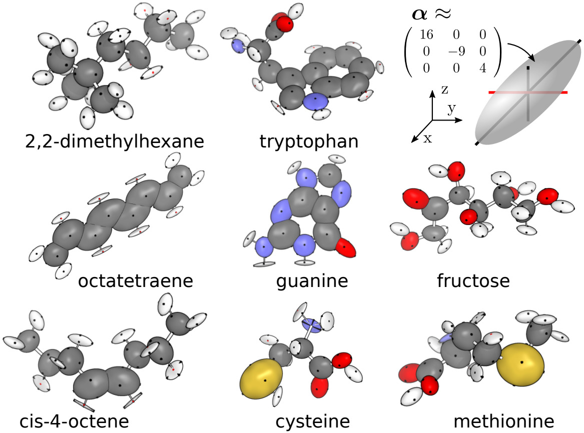


T. Cohen, M. Welling, ICML (2016); Glielmo et al. PRB (2017); Grisafi et al., PRL (2018)

Equivariance for your favourite model

- Want to learn vectors or general tensors? They should transform with the molecular geometry
- Equivariance constraints the structure of the model
- Equivariance ensures symmetry-adapted behavior

$$y^{\lambda}_{\mu}\left(\hat{R}A_i\right) = \sum_{q} \langle y;\lambda|q\rangle \sum_{\mu'} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right) \langle q|A_i;\lambda\mu'\rangle$$

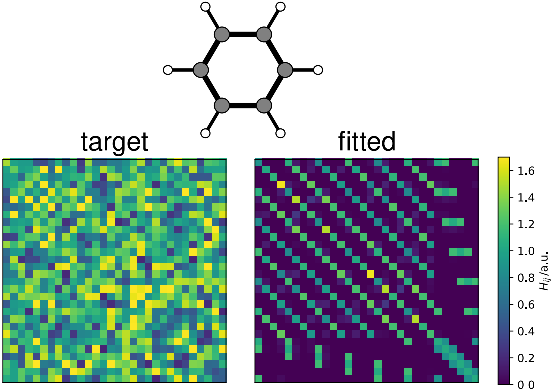


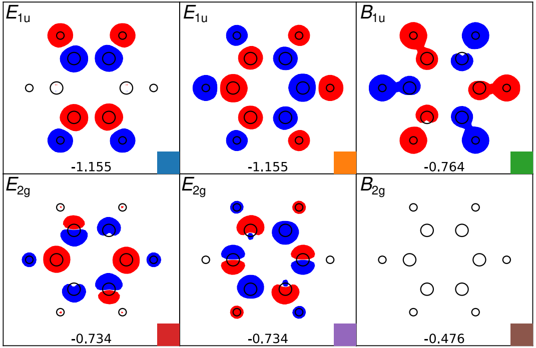
T. Cohen, M. Welling, ICML (2016); Glielmo et al. PRB (2017); Grisafi et al., PRL (2018)

Equivariance for your favourite model

- Want to learn vectors or general tensors? They should transform with the molecular geometry
- Equivariance constraints the structure of the model
- Equivariance ensures symmetry-adapted behavior

$$y^{\lambda}_{\mu}\left(\hat{R}A_i\right) = \sum_{q} \langle y;\lambda|q\rangle \sum_{\mu'} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right) \langle q|A_i;\lambda\mu'\rangle$$



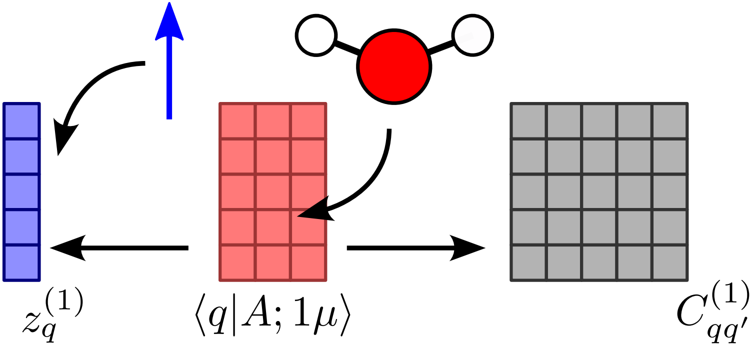


Nigam, Willatt, MC, JCP (2022)

Equivariant ridge regression

- Recall: ridge regression weights involve computing the covariance $\mathbf{C} = \mathbf{X}^T\mathbf{X}$ of the features
- "Virtual" rotational augmentation shows that the covariance must be block-diagonal
- Ridge regression weights $w_q^{(\lambda)}$ are decoupled and μ-independent

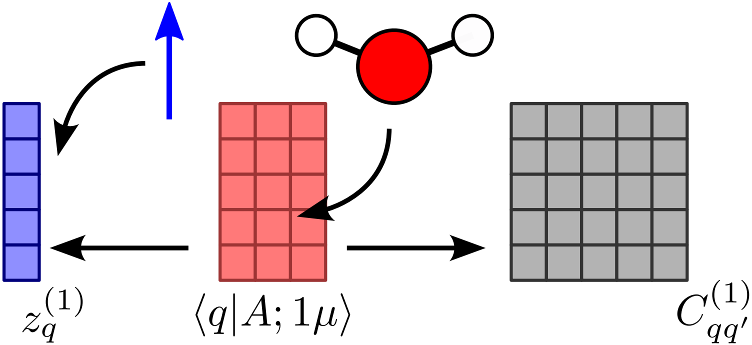
$$\mathbf{P}_{XY} = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$



Equivariant ridge regression

- Recall: ridge regression weights involve computing the covariance $\mathbf{C} = \mathbf{X}^T\mathbf{X}$ of the features
- "Virtual" rotational augmentation shows that the covariance must be block-diagonal
- Ridge regression weights $w_q^{(\lambda)}$ are decoupled and μ-independent

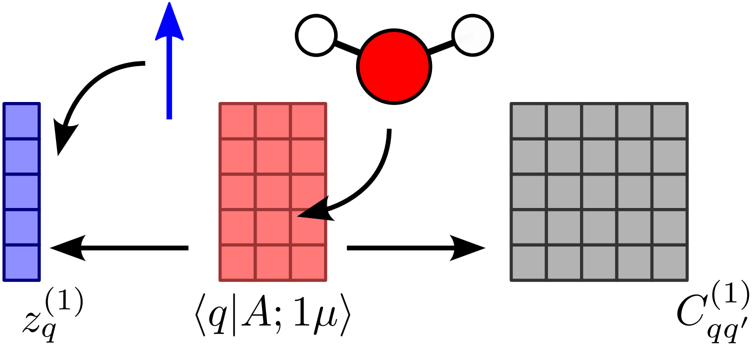
$$C_{q\lambda\mu,q'\lambda'\mu'} = \frac{1}{N}\sum_A \langle q|\hat{R}A;\lambda\mu\rangle\,\langle q'|\hat{R}A;\lambda'\mu'\rangle$$



Equivariant ridge regression

- Recall: ridge regression weights involve computing the covariance $\mathbf{C} = \mathbf{X}^T\mathbf{X}$ of the features
- "Virtual" rotational augmentation shows that the covariance must be block-diagonal
- Ridge regression weights $w_q^{(\lambda)}$ are decoupled and μ-independent

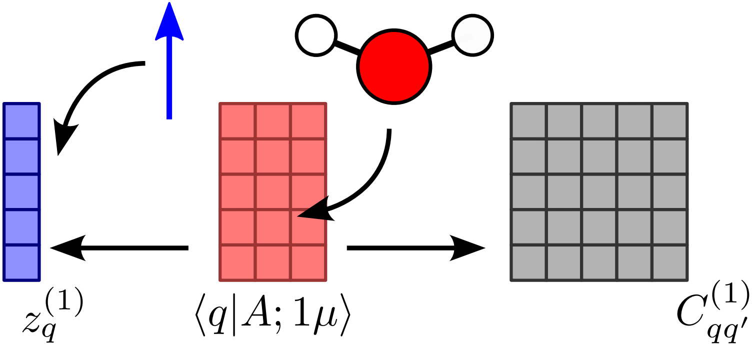
$$\int \mathrm{d}\hat{R}\, \langle q|\hat{R}A;\lambda\mu\rangle\ \langle q'|\hat{R}A;\lambda'\mu'\rangle = \sum_{mm'} \langle q|A;\lambda m\rangle\ \langle q'|A;\lambda' m'\rangle \int \mathrm{d}\hat{R} D^{\lambda}_{m\mu}(\hat{R}) D^{\lambda'}_{m'\mu'}(\hat{R}) \sim \delta_{\lambda\lambda'}\delta_{\mu\mu'}$$



Equivariant ridge regression

- Recall: ridge regression weights involve computing the covariance $\mathbf{C} = \mathbf{X}^T\mathbf{X}$ of the features
- "Virtual" rotational augmentation shows that the covariance must be block-diagonal
- Ridge regression weights $w_q^{(\lambda)}$ are decoupled and μ-independent

$$C_{qq'}^{(\lambda)} = \frac{1}{N}\sum_{A\mu} \langle q|\hat{R}A;\lambda\mu\rangle\ \langle q'|\hat{R}A;\lambda\mu\rangle \quad z_q^{(\lambda)} = \frac{1}{N}\sum_{A\mu} \langle q|\hat{R}A;\lambda\mu\rangle\, y_\lambda^\mu(A)) \quad \mathbf{w}^{(\lambda)} = \left[\mathbf{C}^{(\lambda)} + \sigma\mathbf{I}\right]^{-1}\mathbf{z}^{(\lambda)}$$



Equivariant kernels

- Equivariant kernel models require tensorial kernels that transform like Wigner D matrices
- Kernels can be computed as scalar products of equivariant descriptors
- Non-linear kernels: combination with scalar kernels

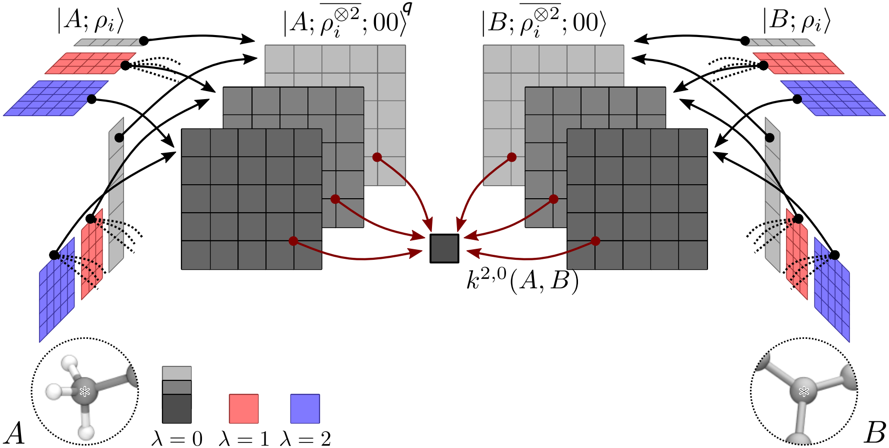
$$y^{\mu}_{\lambda}(A) = \sum_{M\mu'} b_{M\mu'} k^{\lambda}_{\mu\mu'}(A, M) \qquad k^{\lambda}_{\mu\mu'}(A, A') = \int \mathrm{d}\hat{R} D^{\lambda}_{\mu\mu'}\left(\hat{R}\right) k(A, A')$$

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

Equivariant kernels

- Equivariant kernel models require tensorial kernels that transform like Wigner D matrices
- Kernels can be computed as scalar products of equivariant descriptors
- Non-linear kernels: combination with scalar kernels

$$k^{\lambda}_{\mu\mu'}(A, A') = \sum_q \langle q|A; \lambda\mu\rangle \langle q|A'; \lambda\mu'\rangle$$

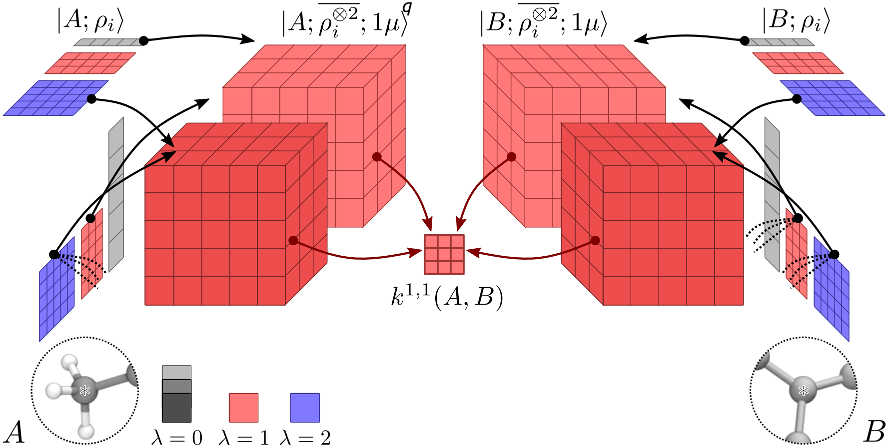


Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

Equivariant kernels

- Equivariant kernel models require tensorial kernels that transform like Wigner *D* matrices
- Kernels can be computed as scalar products of equivariant descriptors
- Non-linear kernels: combination with scalar kernels

$$k^{\lambda}_{\mu\mu'}(A, A') = \sum_q \langle q|A;\lambda\mu\rangle \langle q|A';\lambda\mu'\rangle$$

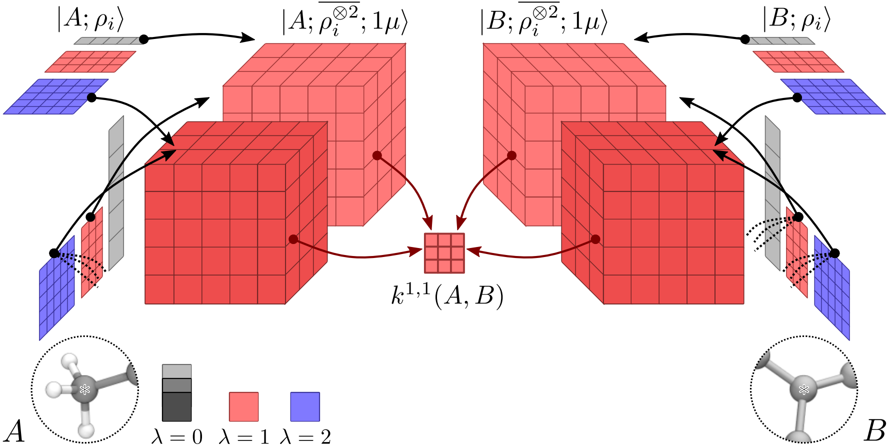


Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

Equivariant kernels

- Equivariant kernel models require tensorial kernels that transform like Wigner *D* matrices
- Kernels can be computed as scalar products of equivariant descriptors
- Non-linear kernels: combination with scalar kernels

$$\tilde{k}^{\lambda}_{\mu\mu'}(A, A') \equiv k^{\lambda}_{\mu\mu'}(A, A')\, k(A, A')^{\zeta}$$



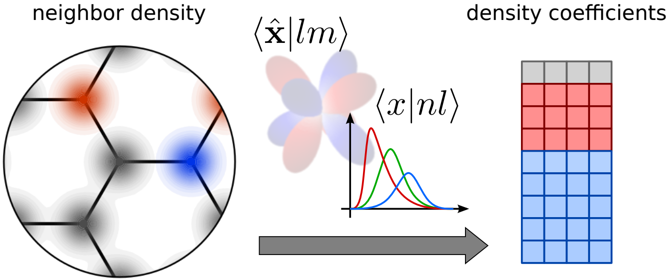
Glielmo, Sollich, De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

The design space of equivariant representations

Ingredients of an equivariant model

- Density expansion coefficients are (irreducible) $O(3)$ equivariant
- **Clebsch-Gordan products**: combine $O(3)$ equivariants $\to$ $O(3)$ *irreducible* equivariants
- Special case: combination with a arbitrary non-linear *scalar* functions
- Recursive construction generates a *complete linear basis* to expand functions of ν neighbors (MTP, ACE, NICE). Extension to message-passing schemes
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

$$x_n^{(\lambda\mu)}(A_i) \equiv \langle n|A_i; \lambda\mu\rangle \equiv \sum_{j\in A_i} R_{n\lambda}(r_{ij})\, Y_\lambda^\mu(\hat{\mathbf{r}}_{ij})$$

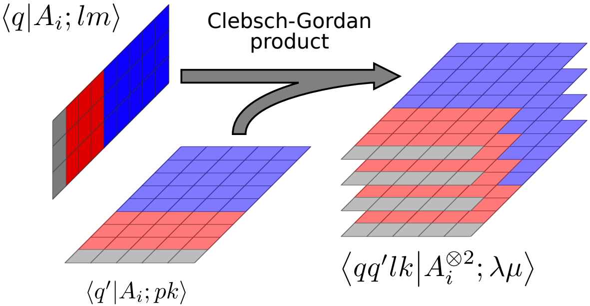


Anderson et al., NIPS (2019); Nigam, Pozdnyakov, MC, JCP (2020); `https://github.com/cosmo-epfl/nice`

Ingredients of an equivariant model

- Density expansion coefficients are (irreducible) $O(3)$ equivariant
- **Clebsch-Gordan products**: combine $O(3)$ equivariants $\to$ $O(3)$ *irreducible* equivariants
- Special case: combination with a arbitrary non-linear *scalar* functions
- Recursive construction generates a *complete linear basis* to expand functions of ν neighbors (MTP, ACE, NICE). Extension to message-passing schemes
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

$$\langle qq'lk|A_i^{\otimes 2};\lambda\mu\rangle = \sum_{mp} \langle q|A_i;lm\rangle\,\langle q'|A_{'i};kp\rangle\,\langle lm;kp|\lambda\mu\rangle$$

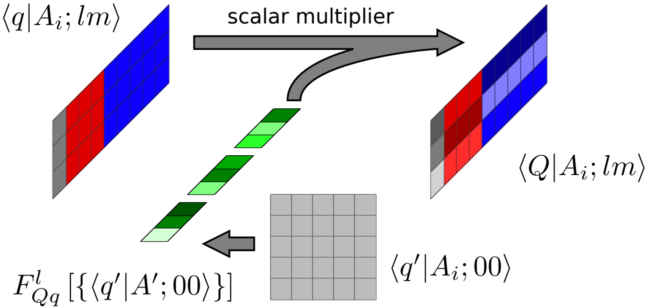


Anderson et al., NIPS (2019); Nigam, Pozdnyakov, MC, JCP (2020); `https://github.com/cosmo-epfl/nice`

Ingredients of an equivariant model

- Density expansion coefficients are (irreducible) $O(3)$ equivariant
- **Clebsch-Gordan products**: combine $O(3)$ equivariants $\rightarrow$ $O(3)$ *irreducible* equivariants
- Special case: combination with a arbitrary non-linear *scalar* functions
- Recursive construction generates a *complete linear basis* to expand functions of ν neighbors (MTP, ACE, NICE). Extension to message-passing schemes
- Linear models built on $\left|\overline{\rho_i^{\otimes\nu}}; g \rightarrow \delta\right\rangle$ yield $(\nu+1)$-body potential expansion

$$\langle Q|AA';\lambda\mu\rangle = \sum_q \langle q|A;\lambda\mu\rangle\, F^{\lambda}_{Qq}\left[\{\langle q'|A';00\rangle\}\right]$$

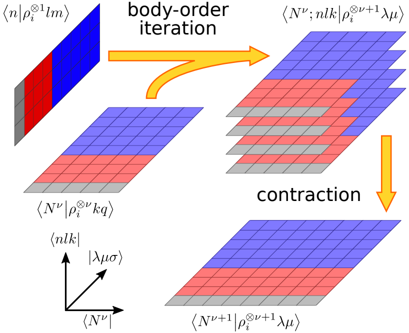


Anderson et al., NIPS (2019); Nigam, Pozdnyakov, MC, JCP (2020); https://github.com/cosmo-epfl/nice

Ingredients of an equivariant model

- Density expansion coefficients are (irreducible) $O(3)$ equivariant
- **Clebsch-Gordan products**: combine $O(3)$ equivariants $\to$ $O(3)$ *irreducible* equivariants
- Special case: combination with a arbitrary non-linear *scalar* functions
- Recursive construction generates a *complete linear basis* to expand functions of ν neighbors (MTP, ACE, NICE). Extension to message-passing schemes
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \to \delta\rangle$ yield $(\nu+1)$-body potential expansion

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

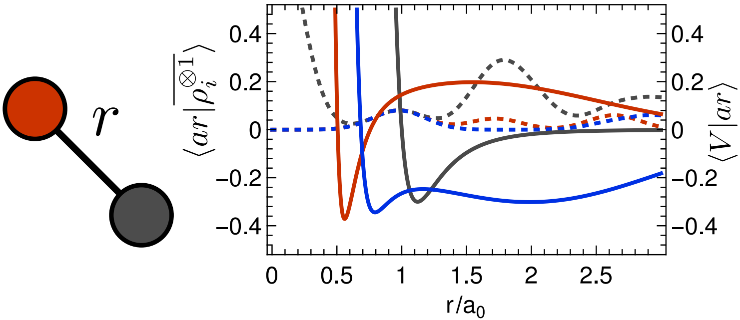


Anderson et al., NIPS (2019); Nigam, Pozdnyakov, MC, JCP (2020); `https://github.com/cosmo-epfl/nice`

Ingredients of an equivariant model

- Density expansion coefficients are (irreducible) $O(3)$ equivariant
- **Clebsch-Gordan products**: combine $O(3)$ equivariants $\rightarrow$ $O(3)$ *irreducible* equivariants
- Special case: combination with a arbitrary non-linear *scalar* functions
- Recursive construction generates a *complete linear basis* to expand functions of ν neighbors (MTP, ACE, NICE). Extension to message-passing schemes
- Linear models built on $|\overline{\rho_i^{\otimes\nu}}; g \rightarrow \delta\rangle$ yield $(\nu+1)$-body potential expansion

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

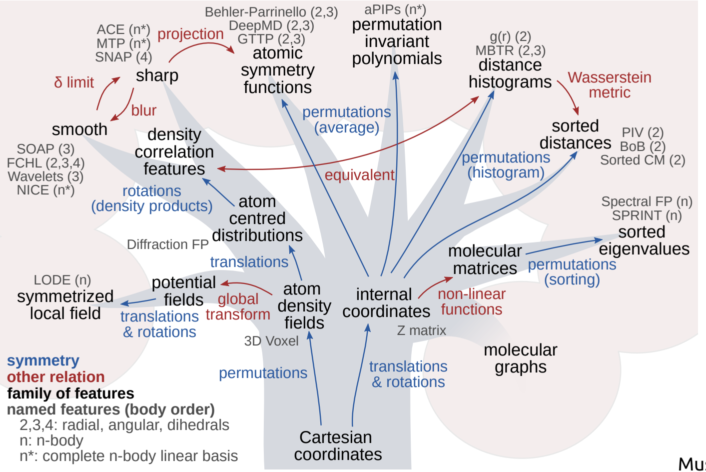


$$V(A_i) = \int \mathrm{d}x \langle V|ax\rangle\langle ax|\overline{\rho_i^{\otimes 1}}\rangle \approx \sum_j V_a(r_{ij})$$

Shapeev, MMS (2016); Willatt, Musil, MC, JCP (2019); Drautz, PRB (2019); Glielmo, Zeni, De Vita, PRB (2018)

What do equivariant models compute?

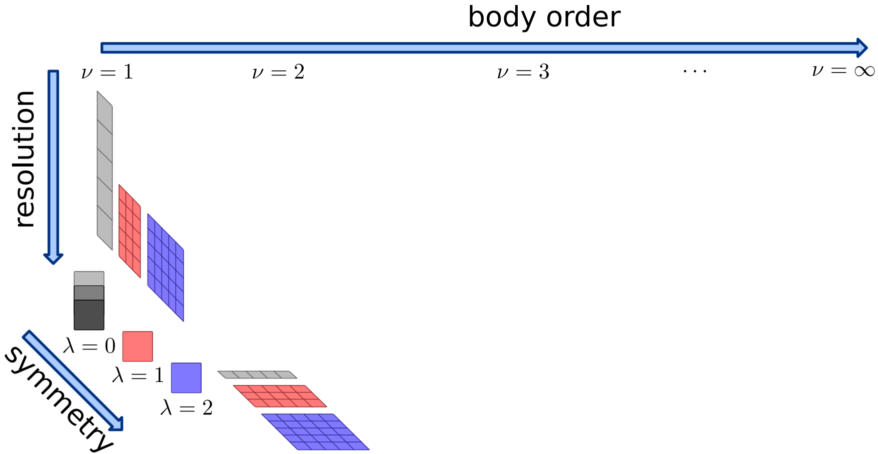
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\rightarrow$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



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

What do equivariant models compute?

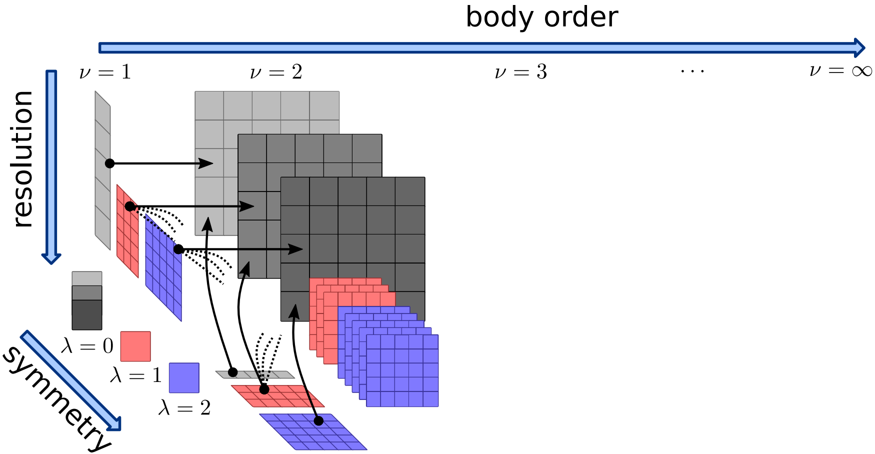
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\rightarrow$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

What do equivariant models compute?

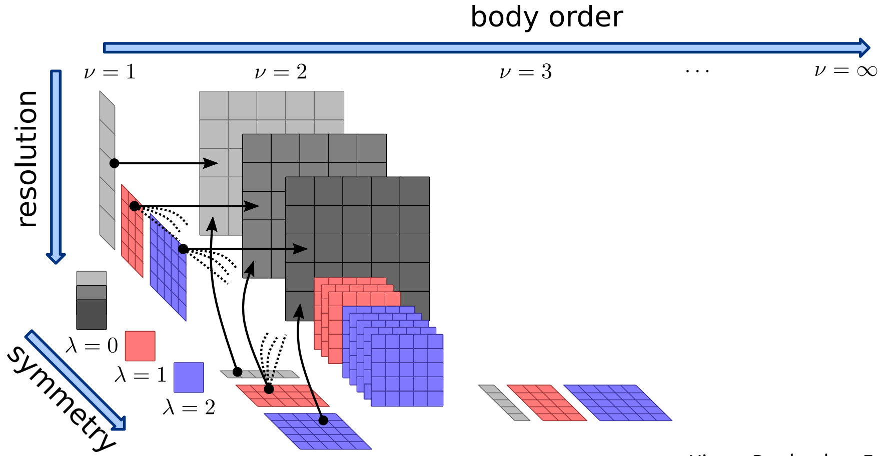
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\rightarrow$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

What do equivariant models compute?

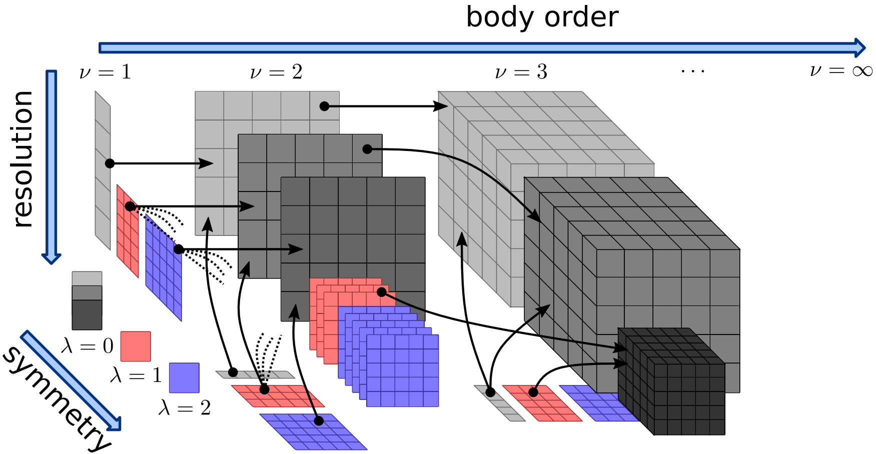
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\to$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

What do equivariant models compute?

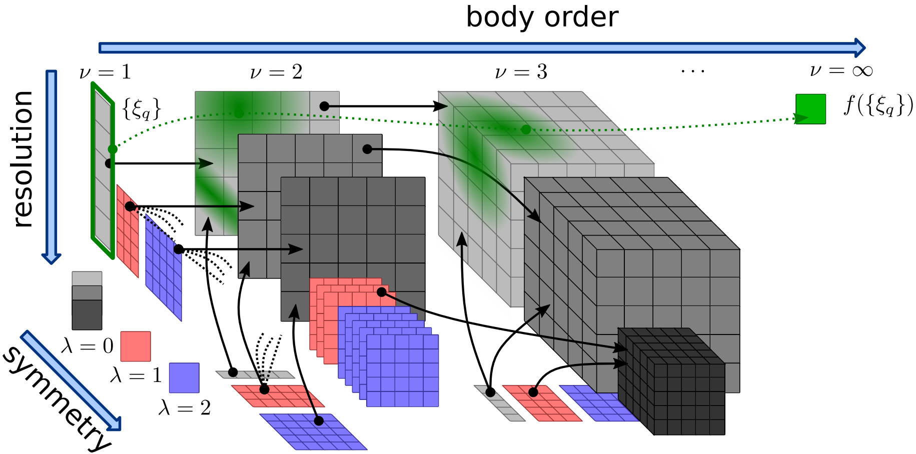
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\rightarrow$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

What do equivariant models compute?

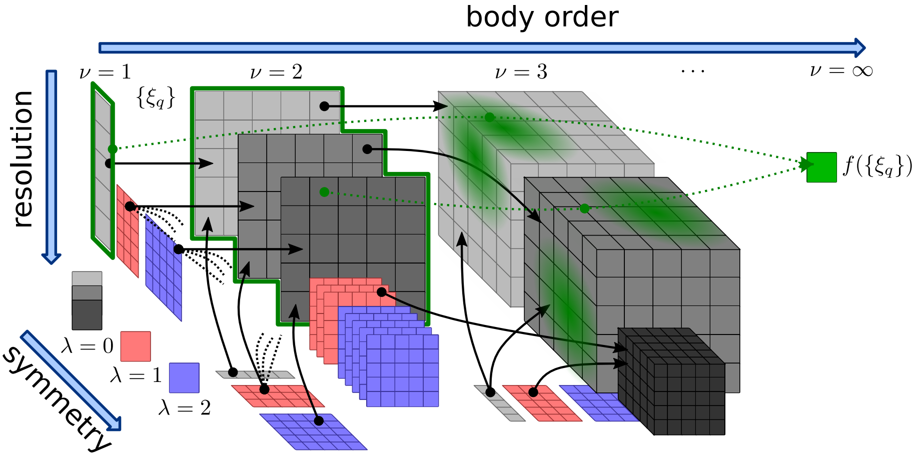
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\rightarrow$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

What do equivariant models compute?

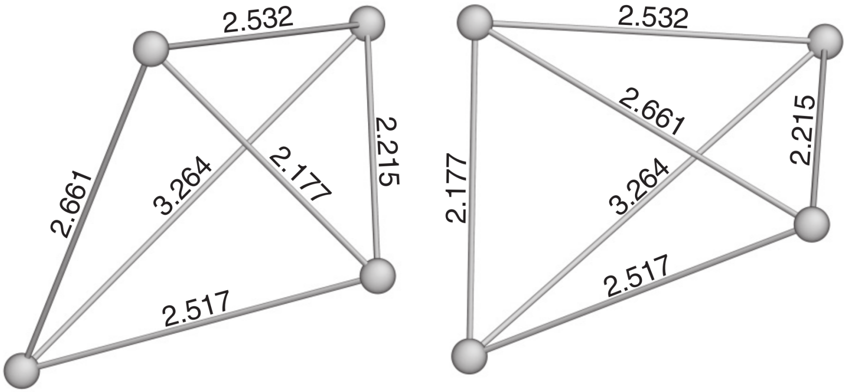
- Models differ in the contraction strategies, but the complete basis, $\nu = \infty$ limit of all equivariant models is the same
- *Symmetry ties your hands*: polynomial CG iterations $\rightarrow$ only systematic way to increase body order while keeping irreducible equivariance
- Important (and poorly understood) role played by non-linearities: infinite-body-order descriptors. Complete?



Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

Problem 1: Low-order invariants are incomplete

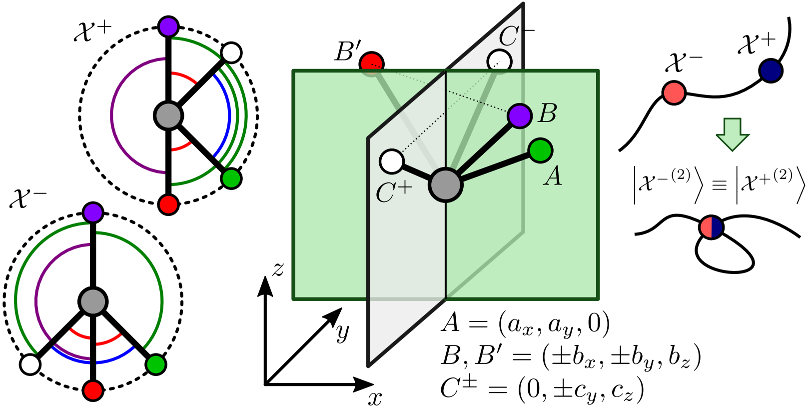
- It is well-known that 2-body correlations are ambiguous: one can build tetrahedra with same pair distances that are different
- Surprise: neither are 3 (and 4!!) body feature!
- Degeneracy "spills over" to neighboring structures: numerical instability



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

Problem 1: Low-order invariants are incomplete

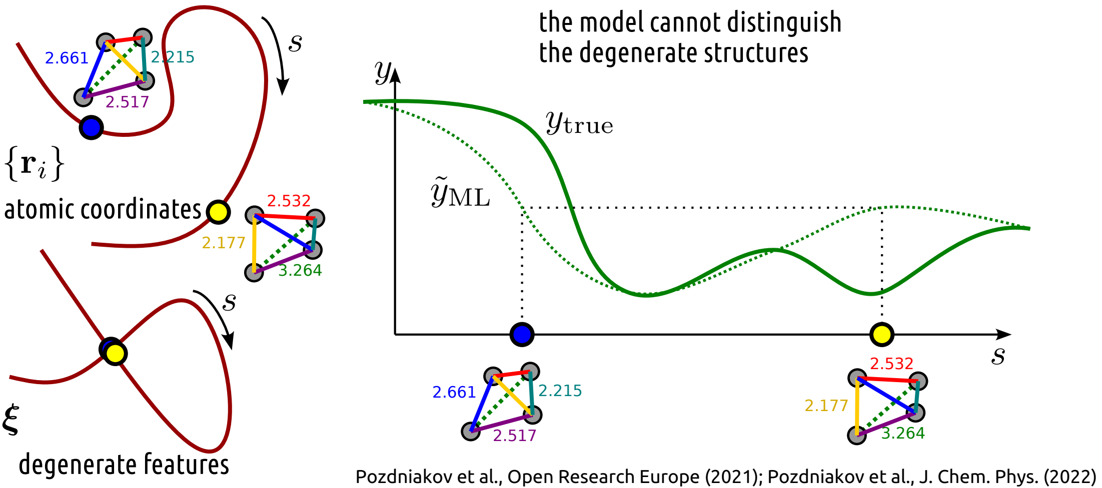
- It is well-known that 2-body correlations are ambiguous: one can build tetrahedra with same pair distances that are different
- Surprise: neither are 3 (and 4!!) body feature!
- Degeneracy "spills over" to neighboring structures: numerical instability



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

Problem 1: Low-order invariants are incomplete

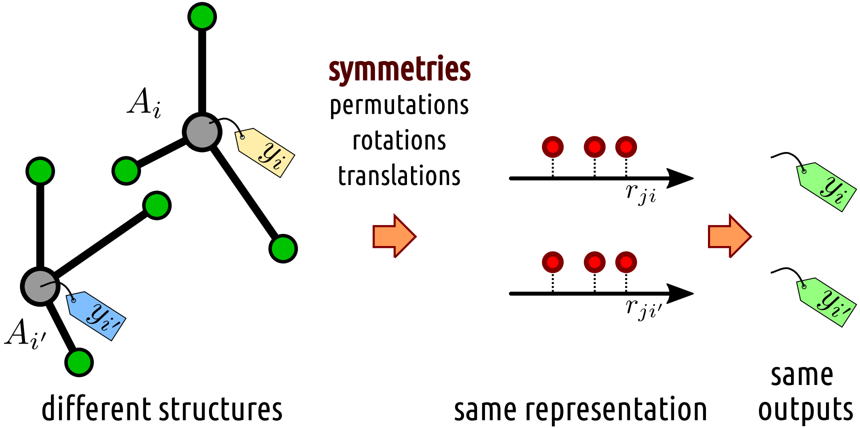
- It is well-known that 2-body correlations are ambiguous: one can build tetrahedra with same pair distances that are different
- Surprise: neither are 3 (and 4!!) body feature!
- Degeneracy "spills over" to neighboring structures: numerical instability



Pozdniakov et al., Open Research Europe (2021); Pozdniakov et al., J. Chem. Phys. (2022)

Solution: Deep-set 3-center representations

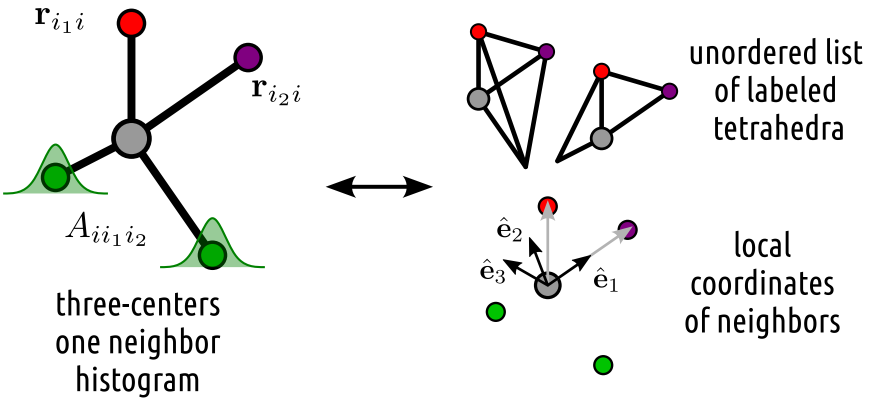
- Atom-centered equivariants are symmetrized over rotations *and* neighbor permutations
- Change the order of symmetrization to obtain a universal approximator
  - 3-center descriptors provide rotationally-invariant features in a local reference frame
  - symmetrize over 2 neighbor indices using a deep-set encoder



Nigam, Pozdnyakov, MC, arXiv:2302.14770

Solution: Deep-set 3-center representations

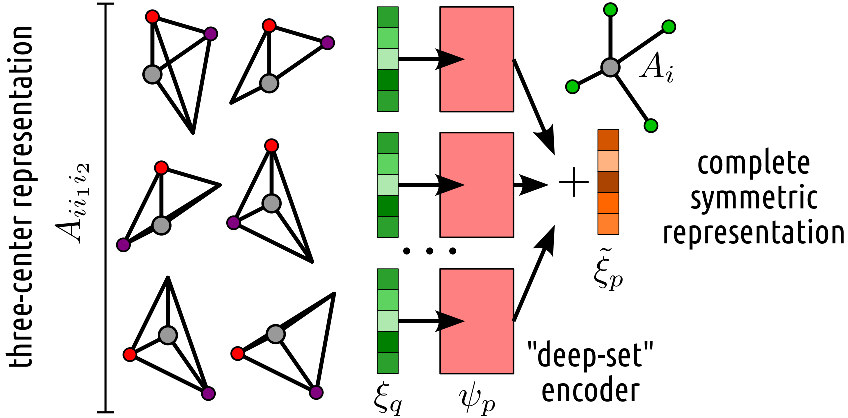
- Atom-centered equivariants are symmetrized over rotations *and* neighbor permutations
- Change the order of symmetrization to obtain a universal approximator
  - 3-center descriptors provide rotationally-invariant features in a local reference frame
  - symmetrize over 2 neighbor indices using a deep-set encoder



Nigam, Pozdnyakov, MC, arXiv:2302.14770

Solution: Deep-set 3-center representations

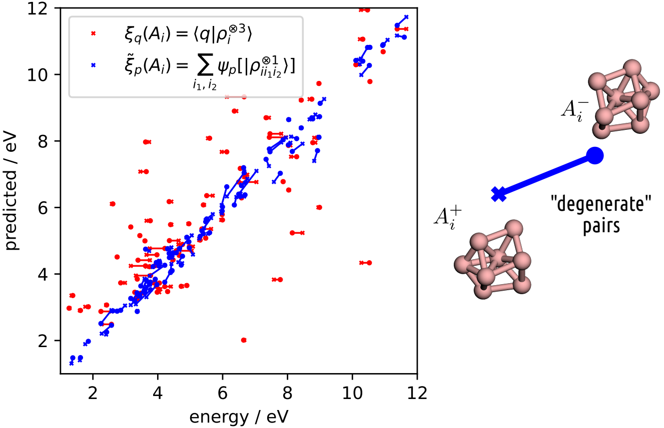
- Atom-centered equivariants are symmetrized over rotations *and* neighbor permutations
- Change the order of symmetrization to obtain a universal approximator
  - 3-center descriptors provide rotationally-invariant features in a local reference frame
  - symmetrize over 2 neighbor indices using a deep-set encoder



Nigam, Pozdnyakov, MC, arXiv:2302.14770

Solution: Deep-set 3-center representations

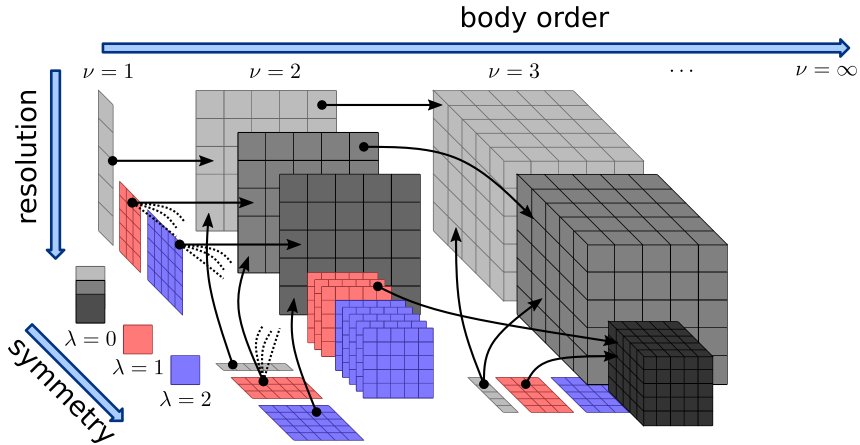
- Atom-centered equivariants are symmetrized over rotations *and* neighbor permutations
- Change the order of symmetrization to obtain a universal approximator
  - 3-center descriptors provide rotationally-invariant features in a local reference frame
  - symmetrize over 2 neighbor indices using a deep-set encoder



Nigam, Pozdnyakov, MC, arXiv:2302.14770

Problem 2: Truncation of the discrete basis

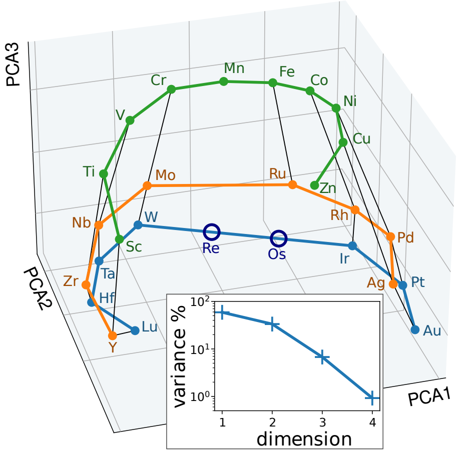
- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels

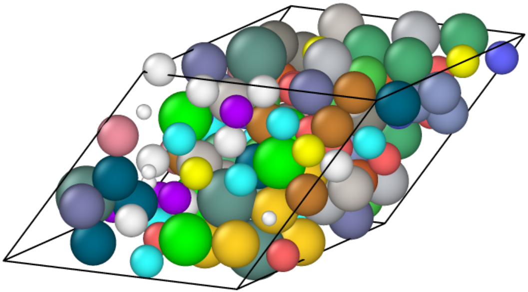


Nigam, Pozdnyakov, Fraux, MC, JCP (2022)

Problem 2: Truncation of the discrete basis

- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}\left(M^\nu\right)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels

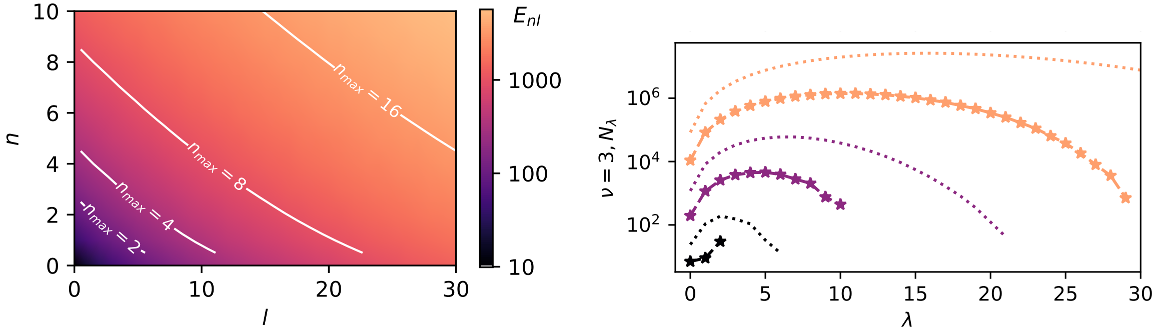




Willatt et al., PCCP (2018); Lopanitsyna et al. PRM (2023); Darby et al., PRL (2023);
Nigam, Pozdnyakov, MC, J. Chem. Phys. (2020); Dusson et al., J. Comp. Phys. (2022);
Bigi, Huguenin-Dumittan, MC, Manolopoulos, J. Chem. Phys. (2022)

Problem 2: Truncation of the discrete basis

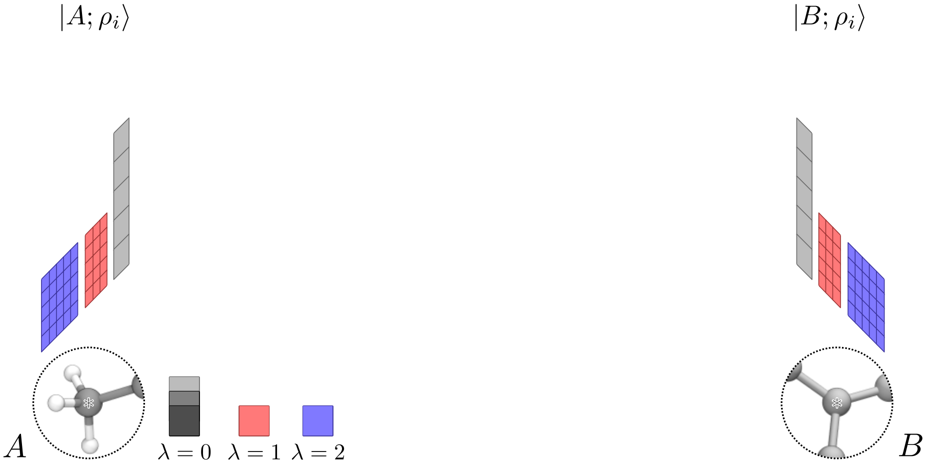
- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels



Willatt et al., PCCP (2018); Lopanitsyna et al. PRM (2023); Darby et al., PRL (2023); Nigam, Pozdnyakov, MC, J. Chem. Phys. (2020); Dusson et al., J. Comp. Phys. (2022); Bigi, Huguenin-Dumittan, MC, Manolopoulos, J. Chem. Phys. (2022)

Problem 2: Truncation of the discrete basis

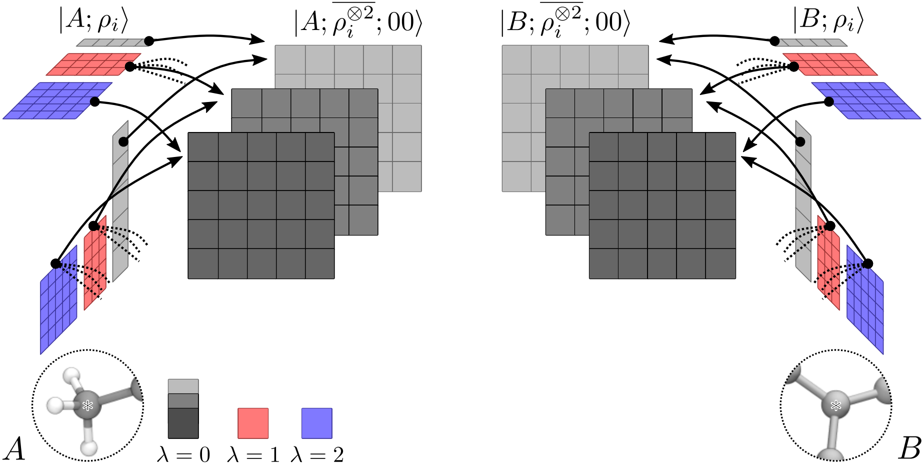
- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels



Bigi, Pozdnyakov, MC, JCP (2024)

Problem 2: Truncation of the discrete basis

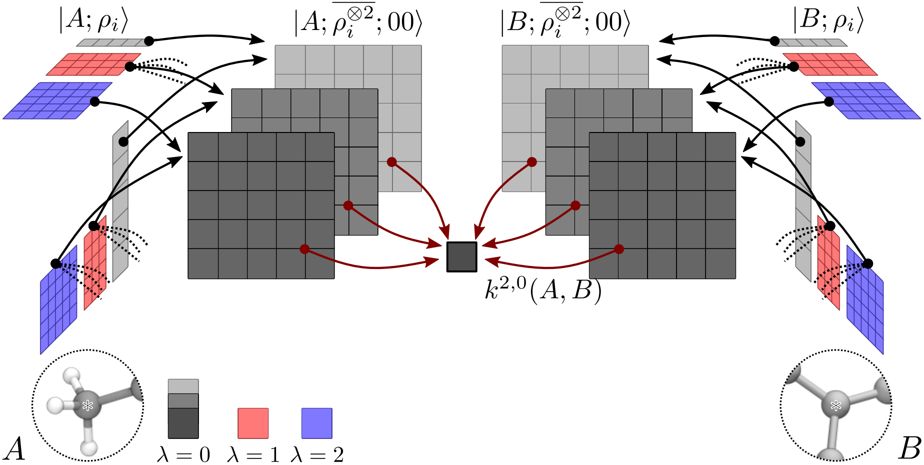
- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels



Bigi, Pozdnyakov, MC, JCP (2024)

Problem 2: Truncation of the discrete basis

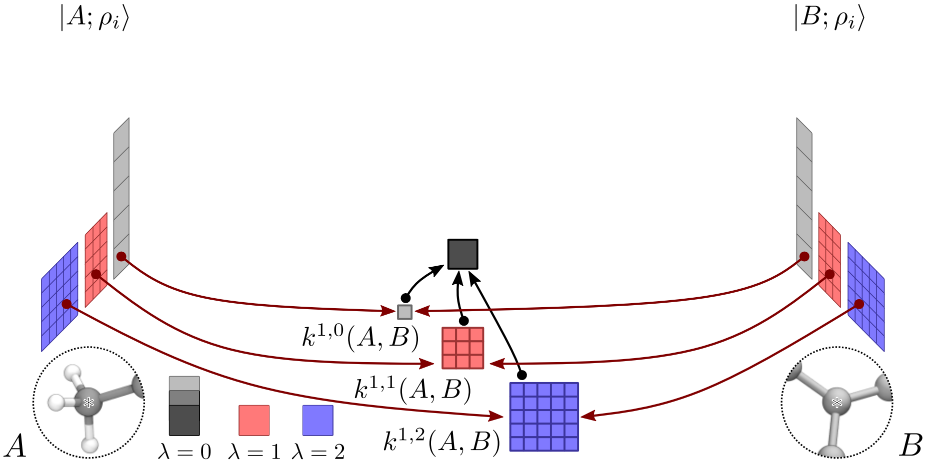
- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels



Bigi, Pozdnyakov, MC, JCP (2024)

Problem 2: Truncation of the discrete basis

- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels



Bigi, Pozdnyakov, MC, JCP (2024)

Problem 2: Truncation of the discrete basis

- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}\left(M^\nu\right)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels

$$k^{(\nu_1+\nu_2),\lambda}_{\mu\mu'}(A,A') = \sum_{\substack{l_1\, m_1\, m_1' \\ l_2\, m_2\, m_2'}} \langle l_1\, m_1; l_2\, m_2|\lambda\mu\rangle\, k^{\nu_1,l_1}_{m_1\, m_1'}(A,A')\, k^{\nu_2,l_2}_{m_2\, m_2'}(A,A')\, \langle l_1\, m_1'; l_2\, m_2'|\lambda\mu'\rangle\,,$$

Bigi, Pozdnyakov, MC, JCP (2024)

Problem 2: Truncation of the discrete basis

- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels

MAE on QM9 (100k training points, no forces)

Model	U_0	U	H	G	Avg.
NoisyNodes	7.3	7.6	7.4	8.3	7.65
SphereNet	6.3	6.4	6.3	7.8	6.70
DimeNet++	6.3	6.3	6.5	7.6	6.68
ET	6.2	6.4	6.2	7.6	6.60
PaiNN	5.9	5.8	6.0	7.4	6.28
MACE	5.2 (0.2)	**4.1**	4.7	**5.5**	4.88
Allegro	4.7 (0.2)	4.4	4.4	5.7	4.80
TensorNet	**4.3** (0.3)	4.3 (0.1)	4.3 (0.2)	6.0 (0.1)	4.72
Wigner Kernels	**4.3** (0.1)	4.2 (0.2)	**4.2** (0.2)	6.0 (0.1)	**4.68**

Bigi, Pozdnyakov, MC, JCP (2024)

Problem 2: Truncation of the discrete basis

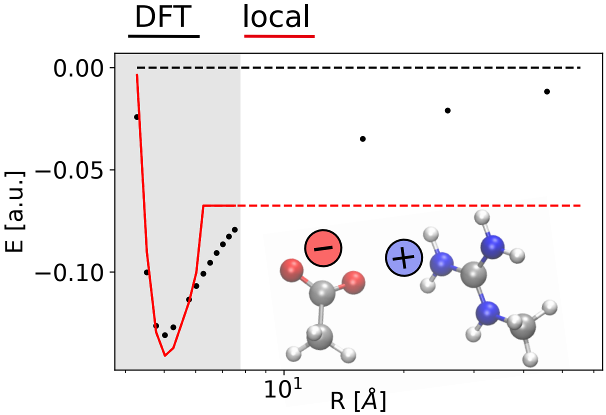
- If $\langle anlm|\rho_i\rangle$ contains M coefficients, $\langle q|\overline{\rho_i^{\otimes\nu}}\rangle$ contains $\mathcal{O}(M^\nu)$
- Many solutions to mitigate: alchemical embedding; tensor sketch; iterative contraction (NICE); total degree truncation (ACE); optimal-smoothness truncation (LE basis);
- Body-ordered potentials **without a basis**: Wigner kernels

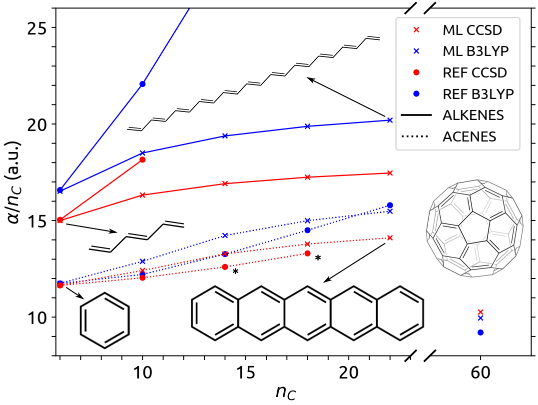
Molecule		LE-ACE	NequIP	MACE	WK
Aspirin	F	59.1	52.0	**43.9**	50.2
Azobenzene	F	27.5	20.0	**17.7**	25.6
Benzene	F	1.44	2.9	2.7	**1.31**
Ethanol	F	32.0	40.3	32.6	**30.8**
Malonaldehyde	F	50.9	52.5	**43.3**	43.8
Naphthalene	F	13.9	10.0	**9.2**	12.5
Paracetamol	F	45.1	39.7	**31.5**	37.2
Salicylic acid	F	36.7	35.0	**28.4**	31.9
Toluene	F	18.4	15.1	**12.1**	16.4
Uracil	F	30.7	40.1	**25.9**	27.8
Average latency		-	-	92 ms	56 ms

Bigi, Pozdnyakov, MC, JCP (2024)

Problem 3: Finite range of interactions

- Finite cutoff for the density expansion → impossible to learn long-range physics
- Many solutions to mitigate: baselining, point-charge/charge-equilibration models, message passing (more in next lecture...)
- Long-distance equivariants: reciprocal-space formulation to model non-bonded interactions

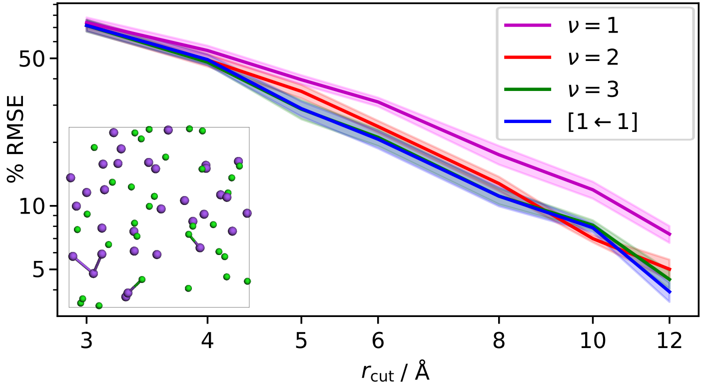




Grisafi, MC, JCP (2019); Wilkins et al. PNAS (2019)

Problem 3: Finite range of interactions

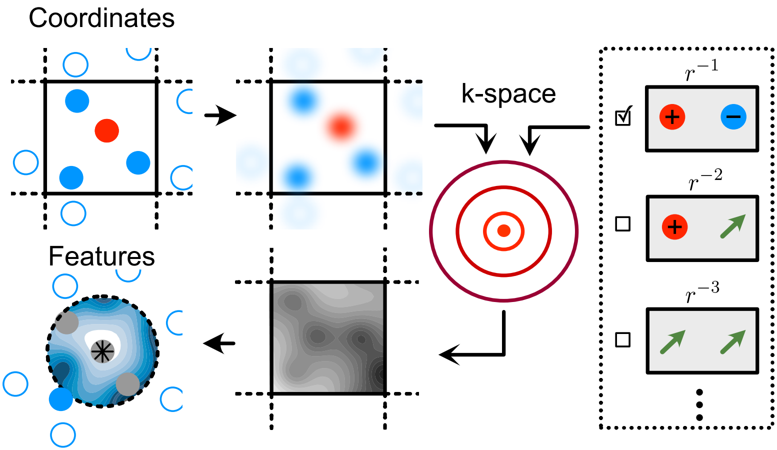
- Finite cutoff for the density expansion $\rightarrow$ impossible to learn long-range physics
- Many solutions to mitigate: baselining, point-charge/charge-equilibration models, message passing (more in next lecture...)
- Long-distance equivariants: reciprocal-space formulation to model non-bonded interactions



Nigam et al., JCP (2022)

Problem 3: Finite range of interactions

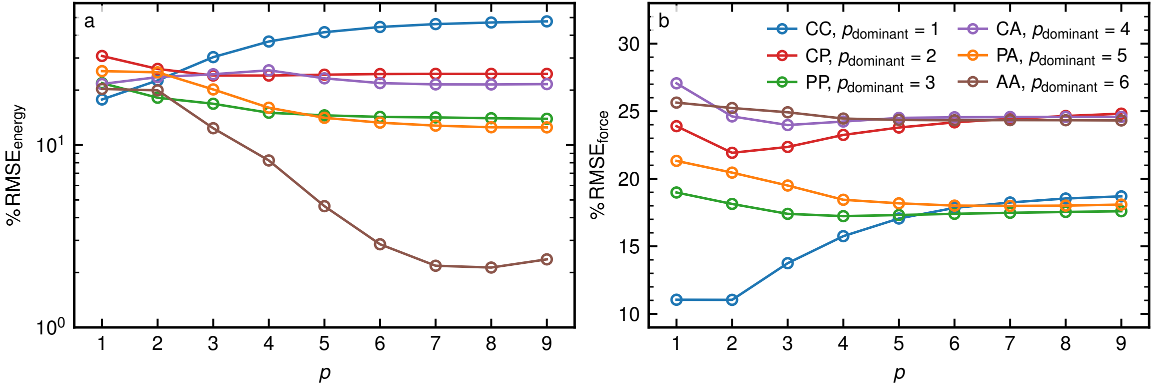
- Finite cutoff for the density expansion → impossible to learn long-range physics
- Many solutions to mitigate: baselining, point-charge/charge-equilibration models, message passing (more in next lecture...)
- Long-distance equivariants: reciprocal-space formulation to model non-bonded interactions



Huguenin-Dumittan, Loche, MC, JCPL (2023)

Problem 3: Finite range of interactions

- Finite cutoff for the density expansion → impossible to learn long-range physics
- Many solutions to mitigate: baselining, point-charge/charge-equilibration models, message passing (more in next lecture...)
- Long-distance equivariants: reciprocal-space formulation to model non-bonded interactions



Huguenin-Dumittan, Loche, MC, JCPL (2023)

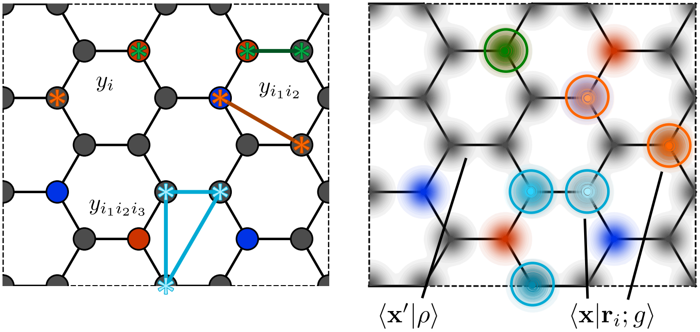
Beyond atom centering

Permutation equivariance with N-center features

- Problem: there are properties that are associated with *pairs* or *N-uples* of atoms

$$y_{ii_2\ldots i_N} = \int \mathrm{d}Q \, \langle y|Q\rangle \, \langle Q|\overline{\rho_{ii_2\ldots i_N}^{\otimes\nu}; \lambda\mu}\rangle$$

- Models rely on N-center *features*, that can be built as symmetrized correlations with tagged *centers*



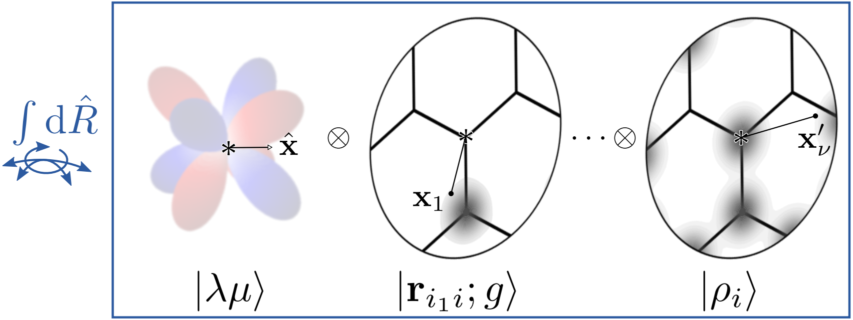
Nigam, Willatt, **MC**, JCP (2022)

Permutation equivariance with N-center features

- Problem: there are properties that are associated with *pairs* or *N-uples* of atoms

$$y_{ii_2\ldots i_N} = \int \mathrm{d}Q \, \langle y|Q\rangle \, \langle Q|\overline{\rho^{\otimes\nu}_{ii_2\ldots i_N}; \lambda\mu}\rangle$$

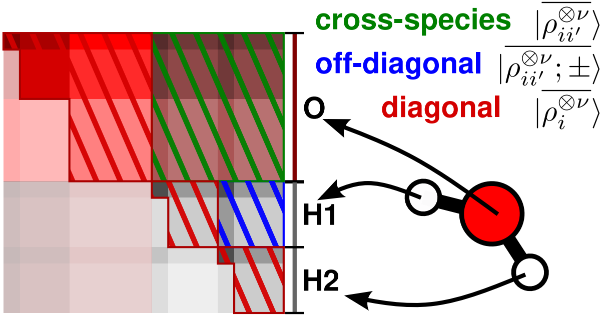
- Models rely on N-center *features*, that can be built as symmetrized correlations with tagged *centers*



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

Hamiltonian learning

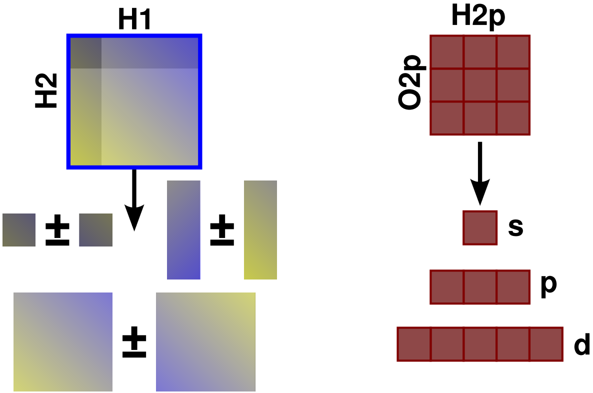
- In an atomic orbital basis the Hamiltonian of a molecule can be decomposed into irreducible symmetric blocks
- These can be learned with a fully equivariant model, that incorporates automatically molecular orbital theory results for symmetric molecules



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

Hamiltonian learning

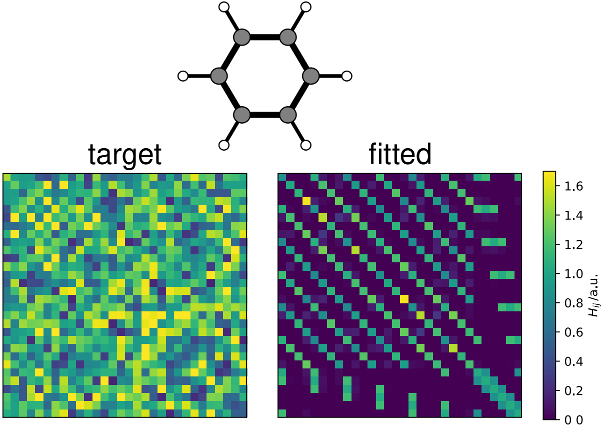
- In an atomic orbital basis the Hamiltonian of a molecule can be decomposed into irreducible symmetric blocks
- These can be learned with a fully equivariant model, that incorporates automatically molecular orbital theory results for symmetric molecules



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

Hamiltonian learning

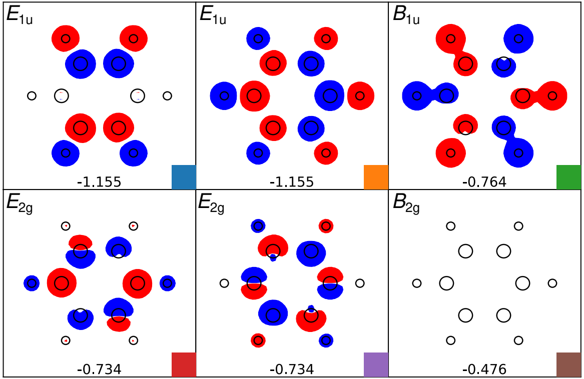
- In an atomic orbital basis the Hamiltonian of a molecule can be decomposed into irreducible symmetric blocks
- These can be learned with a fully equivariant model, that incorporates automatically molecular orbital theory results for symmetric molecules



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

Hamiltonian learning

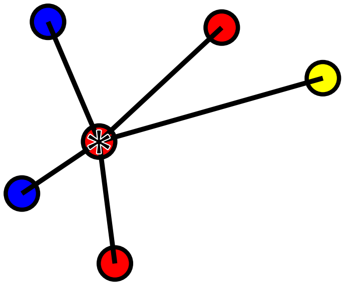
- In an atomic orbital basis the Hamiltonian of a molecule can be decomposed into irreducible symmetric blocks
- These can be learned with a fully equivariant model, that incorporates automatically molecular orbital theory results for symmetric molecules



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

Basic distance graph convolution

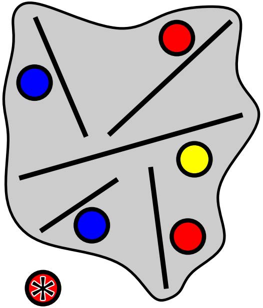
- Atoms are nodes in a fully-connected network. Edges are decorated by (functions of) interatomic distances r_{ij}
- Each node is decorated by the nature of its neighbors and their distance $h(A_i) = (a_i, \{(a_j, r_{ij})\})$
- The multiset of neighbors and edges is hashed, and used as a label to describe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



SchNET: Schütt et al., JCP (2018)

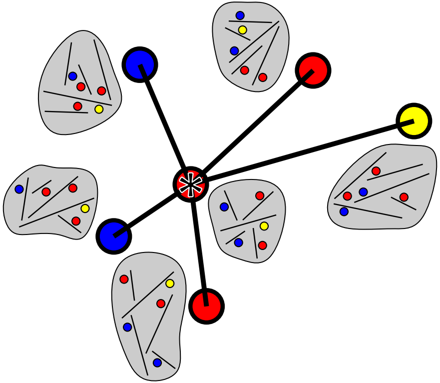
Basic distance graph convolution

- 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 $h\left(A_i\right) = \left(a_{i,} \left\{\left(a_j, r_{ij}\right)\right\}\right)$
- The multiset of neighbors and edges is hashed, and used as a label to describe the nodes. The process can be iterated



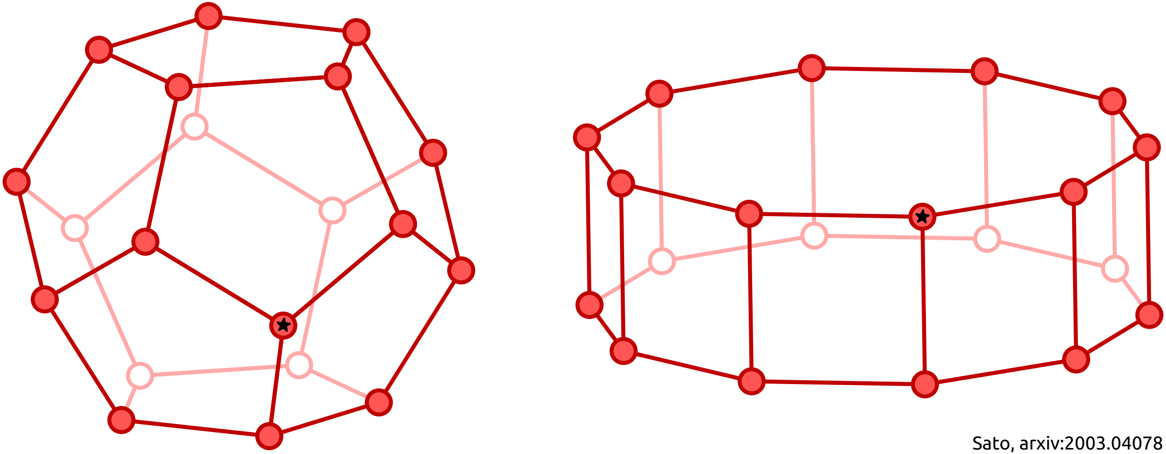
Basic distance graph convolution

- 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 $h\left(A_i\right) = \left(a_i, \left\{\left(a_j, r_{ij}\right)\right\}\right)$
- 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

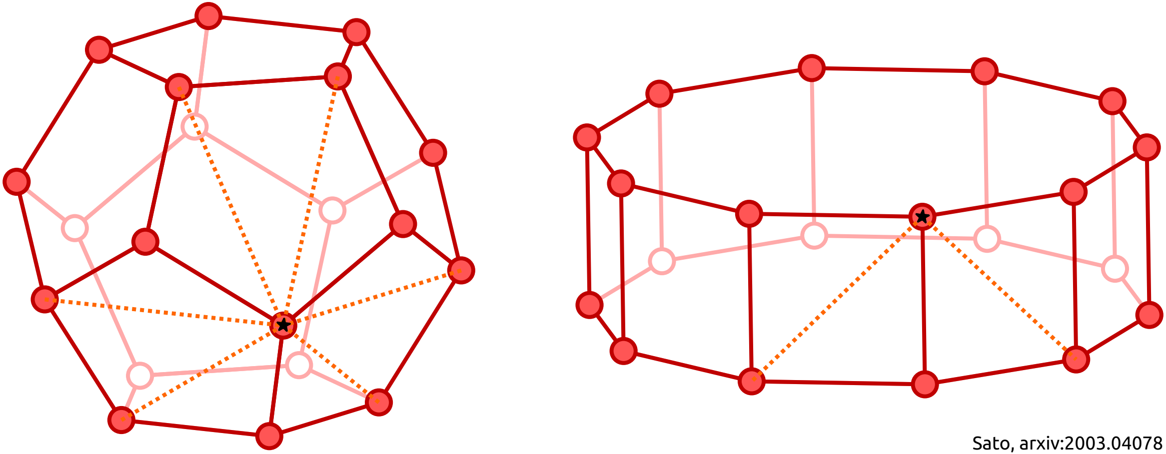
- Bad news: there are known discrete graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things seem to be fine for molecular graphs (*fully-connected*, distance-decorated 3D point clouds); GC resolves all known counterexamples of atom-centered representations



Sato, arxiv:2003.04078

Graph convolution, pros and cons

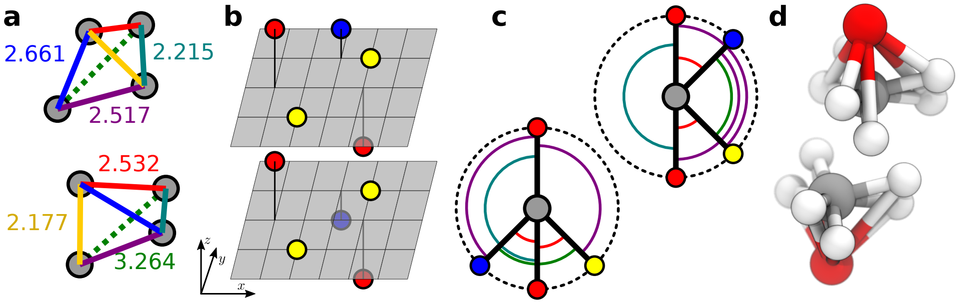
- Bad news: there are known discrete graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things seem to be fine for molecular graphs (*fully-connected*, distance-decorated 3D point clouds); GC resolves all known counterexamples of atom-centered representations



Sato, arxiv:2003.04078

Graph convolution, pros and cons

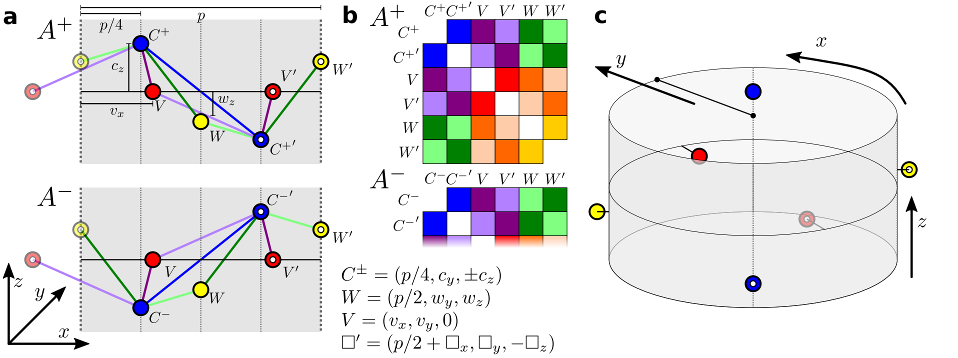
- Bad news: there are known discrete graphs that cannot be distinguished by this procedure (W-L test)
- Good news: things seem to be fine for molecular graphs (*fully-connected*, distance-decorated 3D point clouds); GC resolves all known counterexamples of atom-centered representations



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

A counterexample for distance-based CNN

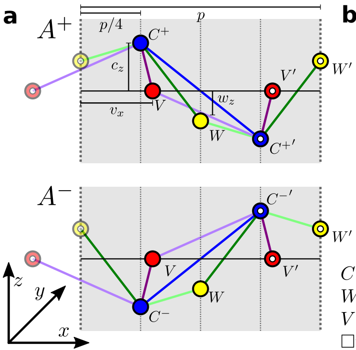
- A family of 3D point clouds with degenerate pairs for GCNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and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

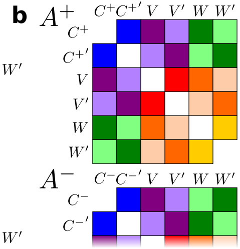


Pozdnyakov, **MC**, MLST (2022)

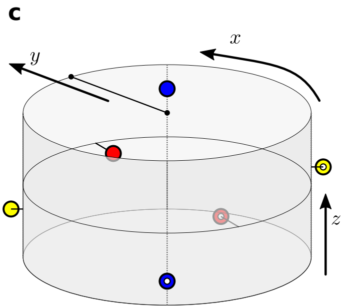
A counterexample for distance-based CNN

- 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





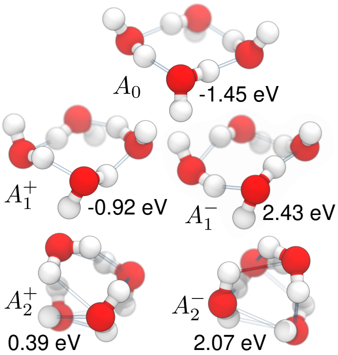
$$C^{\pm} = (p/4, c_y, \pm c_z)$$
$$W = (p/2, w_y, w_z)$$
$$V = (v_x, v_y, 0)$$
$$\Box' = (p/2 + \Box_x, \Box_y, -\Box_z)$$

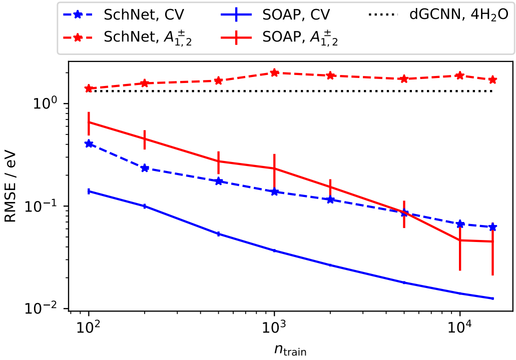


Pozdnyakov, **MC**, MLST (2022)

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

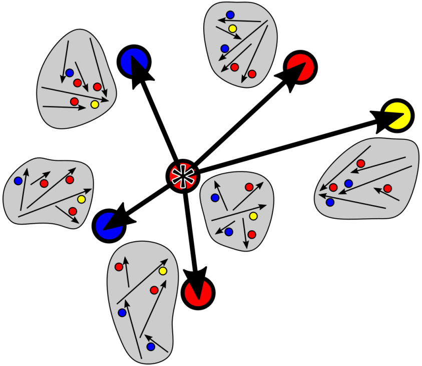




Pozdnyakov, **MC**, MLST (2022)

A counterexample for distance-based CNN

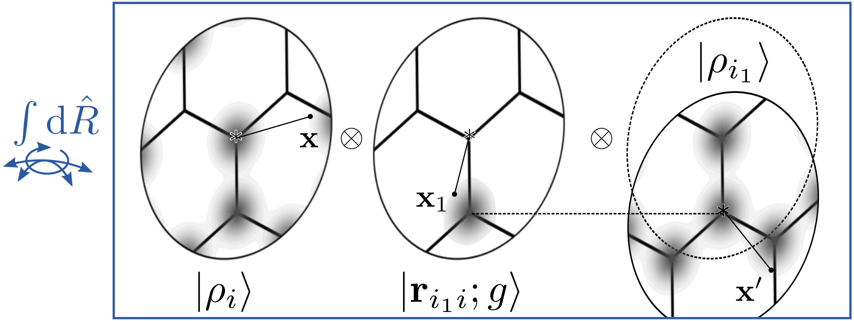
- A family of 3D point clouds with degenerate pairs for GCNN. Key idea: the distance matrix is identical, except for a swap
- Can be folded to give finite 3D structures
- Hard limit to the accuracy for plausible molecular geometries
- Modern architectures that use angular/directional information (and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



Pozdnyakov, **MC**, MLST (2022)

Unified theory of equivariant models

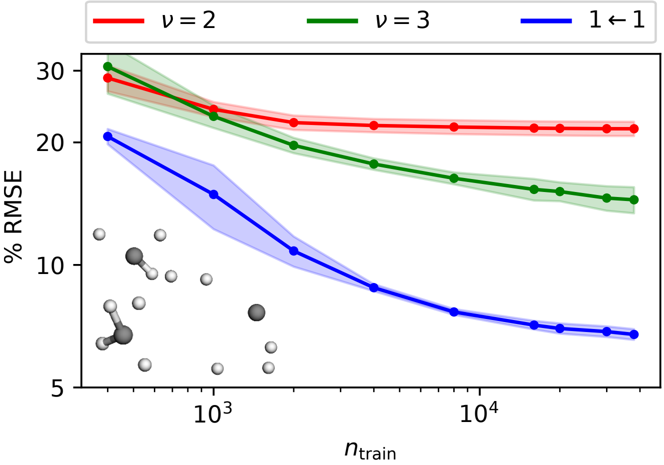
- The construction of N-centers correlations can include features centered on multiple atoms, and message-passing-like contractions
  $|\rho_i^{\otimes[\nu\leftarrow\nu_1]}\rangle = \sum_{i_1} |\rho_i^{\otimes\nu}\rangle \otimes |\mathbf{r}_{i_1 i}\rangle \otimes |\rho_{i_1}^{\otimes\nu_1}\rangle$
- Symmetry-adapted versions can be obtained with CG iterations
  $\langle q_1 l_1; q_2 l_2 | \lambda\mu\rangle = \sum_{m_1 m_2} \langle q_1 | l_1 m_1\rangle \langle q_2 | l_2 m_2\rangle \langle l_1 m_1; l_2 m_2 | \lambda\mu\rangle$
- 1-to-1 mapping with several equivariant message-passing frameworks



Nigam, Pozdnyakov, Fraux, **MC**, JCP (2022)

Resolution and range

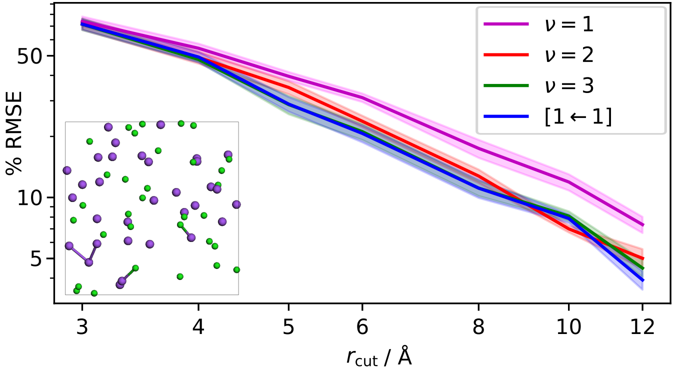
- Empirical tests of the role of MP constructs
- Much better discretization convergence for body-ordered expansions
- . . . but very little impact on long-range interactions



Nigam, Pozdnyakov, Fraux, **MC**, JCP (2022)

Resolution and range

- Empirical tests of the role of MP constructs
- Much better discretization convergence for body-ordered expansions
- . . . but very little impact on long-range interactions



Nigam, Pozdnyakov, Fraux, **MC**, JCP (2022)

Wrapping up

Summary

- ML modeling of tensorial properties require more complicated symmetry constraints
- Symmetry restricts the design space of ML models
- Several problems shared by many frameworks, and shared solutions
- Multi-center models and message-passing provide a framework that can also be applied to predict directly electronic-structure quantities

M. Ceriotti, "Beyond potentials: Integrated machine learning models for materials," MRS Bulletin 47(10), 1045-1053 (2022).
A. Grisafi, A. M. Lewis, M. Rossi, and M. Ceriotti, "Electronic-Structure Properties from Atom-Centered Predictions of the Electron Density," J. Chem. Theory Comput. 19(14), 4451-4460 (2023).
J. Nigam, M. J. Willatt, and M. Ceriotti, "Equivariant representations for molecular Hamiltonians and N-center atomic-scale properties," J. Chem. Phys. 156(1), 014115 (2022).