Machine learning
More than potentials

Michele Ceriotti
https://cosmo.epfl.ch

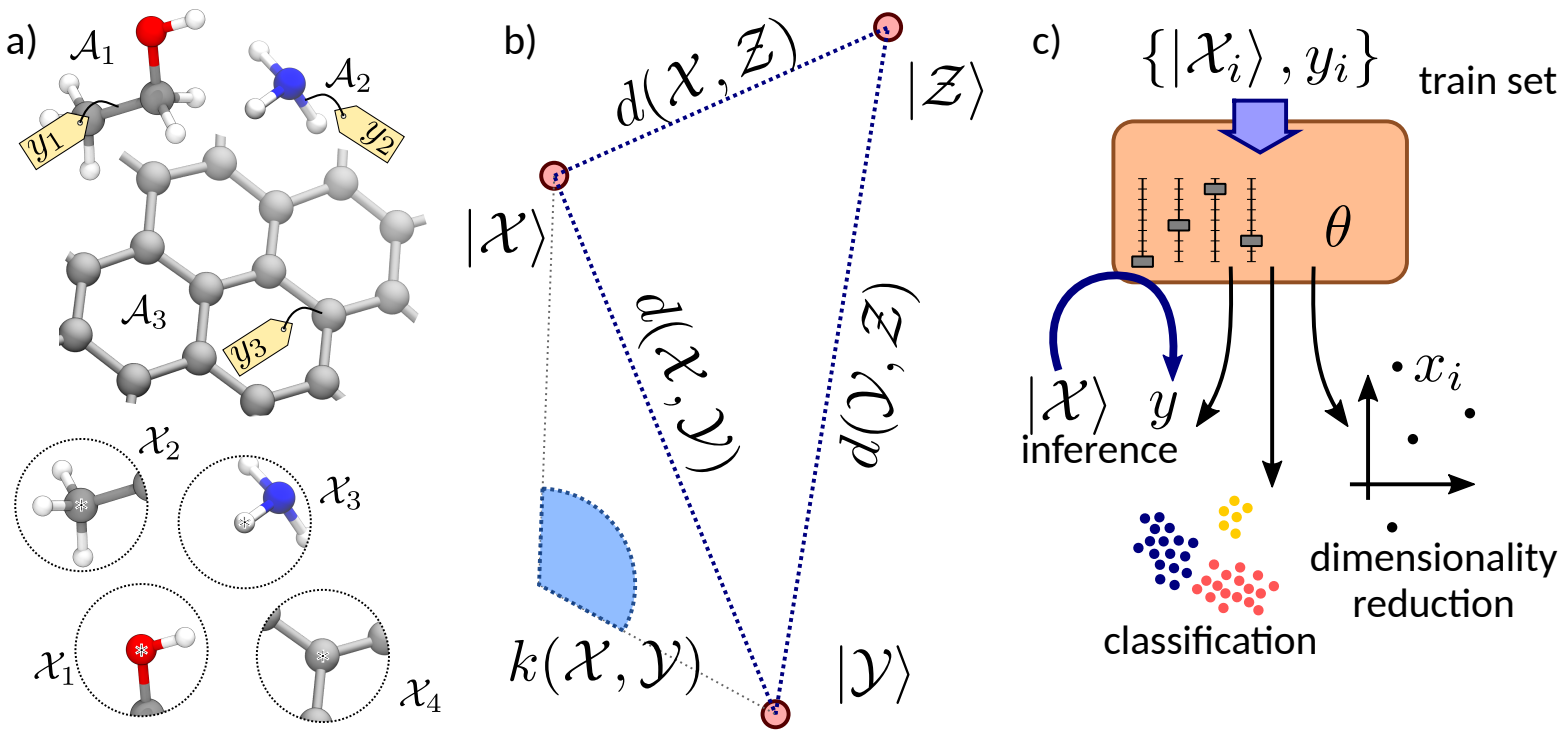
ML4MS, Helsinki, May 2019

COSMO
COMPUTATIONAL SCIENCE AND MODELLING

EPFL

Machine-learning in a nutshell

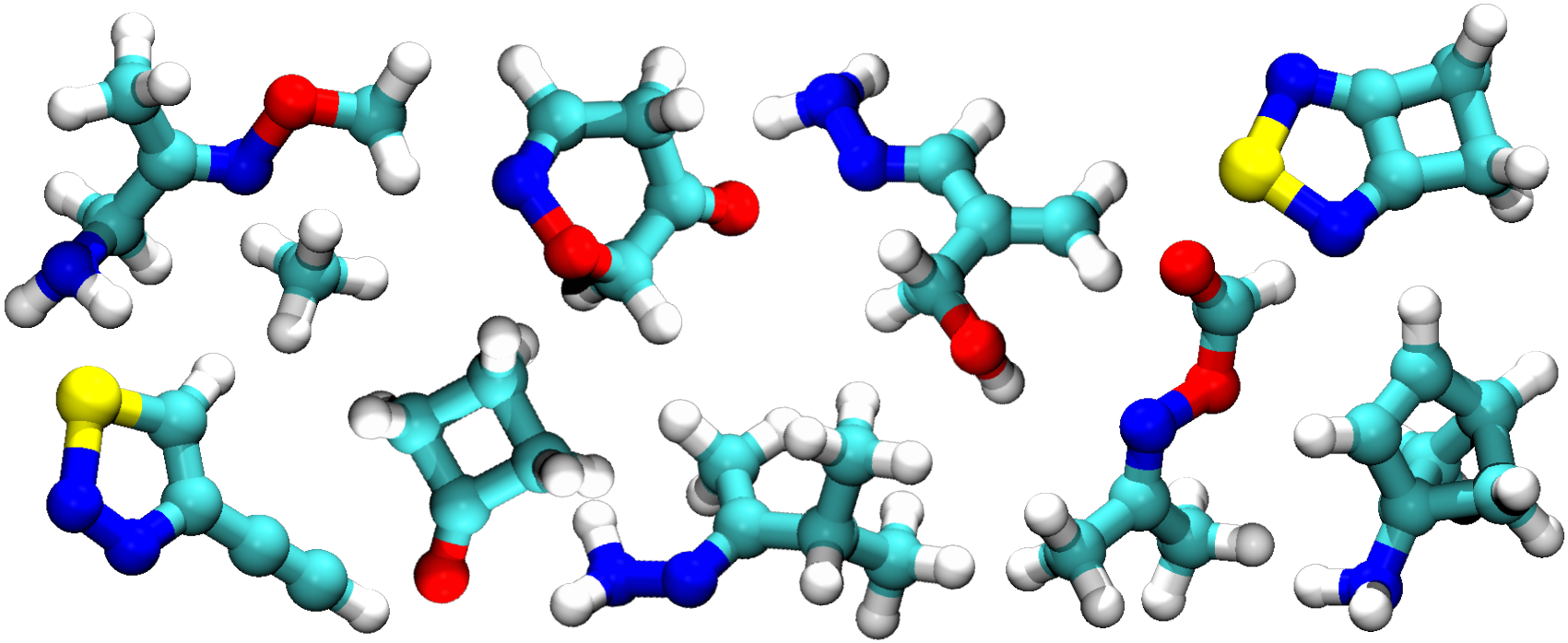
- Chemical structures/environments need to be cast in a complete but concise mathematical representation
- The input/label pairs are fed to a learning scheme, tuned by hyperparameters θ, that can then be used to perform different tasks on new data



Machine Learning of Atomic-Scale Properties Based on Physical Principles,
M Ceriotti, MJ Willatt, G Csányi, Handbook of Materials Modeling: Methods: Theory and Modeling 2018

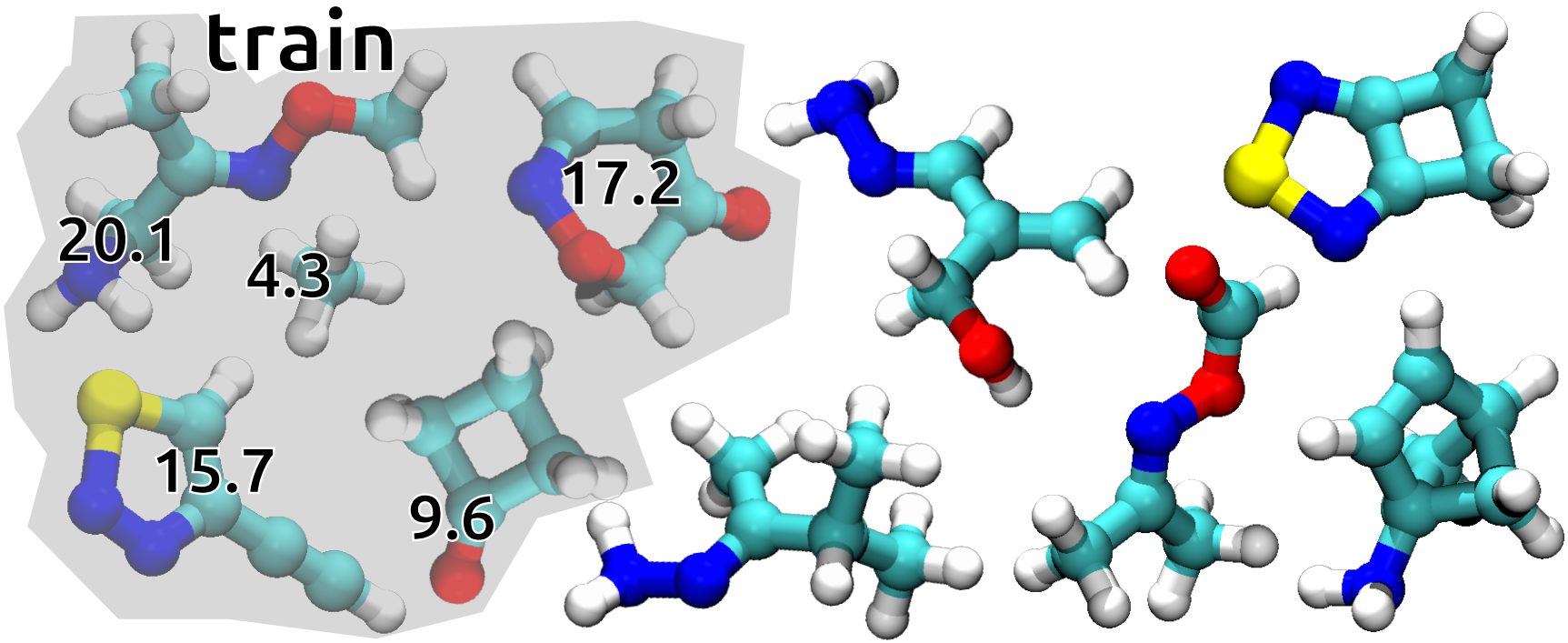
Supervised learning with kernels

- Predict the properties of a bunch of molecules based on few reference calculations
- Use kernel functions as basis - turns the problem into a linear fit onto a training set
- Make predictions on new molecules, benchmark accuracy on a test set



Supervised learning with kernels

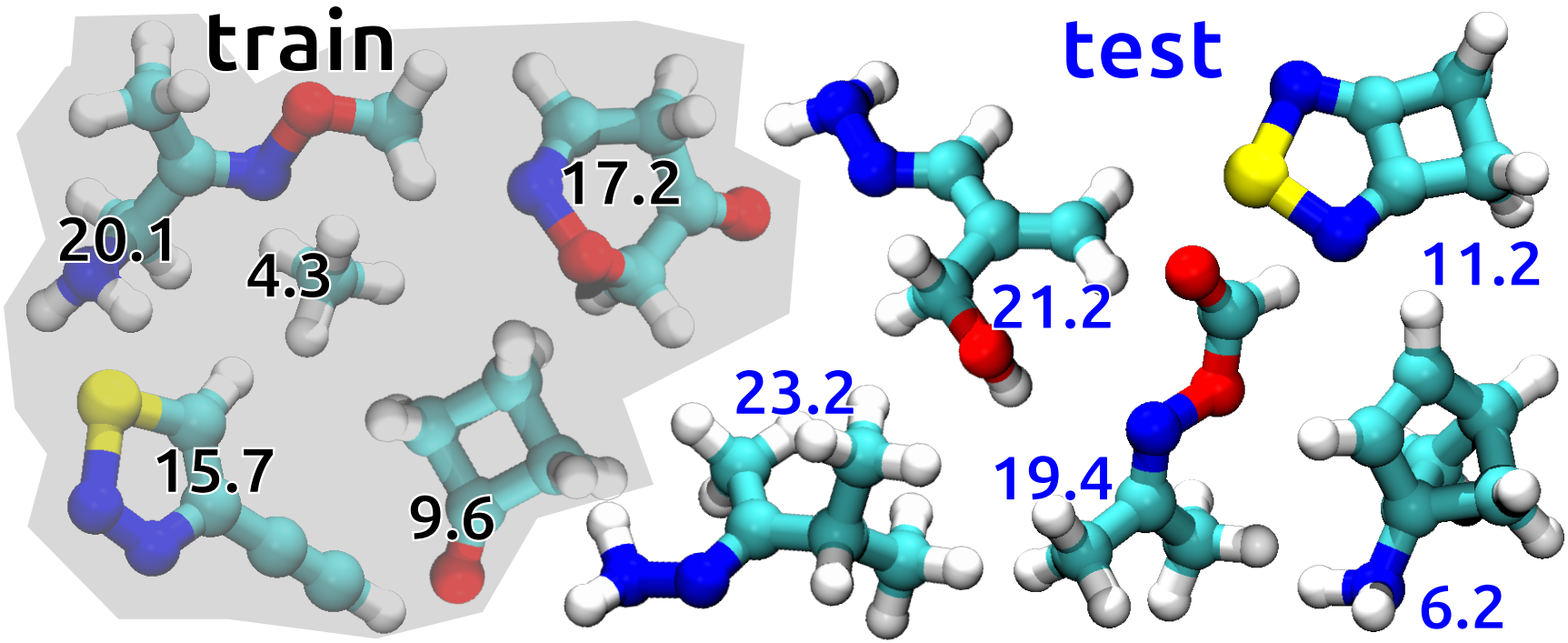
- Predict the properties of a bunch of molecules based on few reference calculations
- Use kernel functions as basis - turns the problem into a linear fit onto a training set
- Make predictions on new molecules, benchmark accuracy on a test set



$$E(\mathcal{A}_j) = \sum_i w_i K(\mathcal{A}_j, \mathcal{A}_i)$$

Supervised learning with kernels

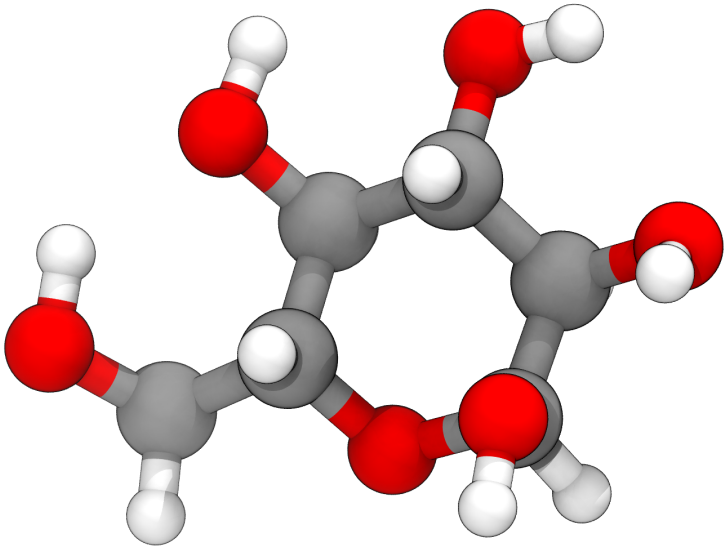
- Predict the properties of a bunch of molecules based on few reference calculations
- Use kernel functions as basis - turns the problem into a linear fit onto a training set
- Make predictions on new molecules, benchmark accuracy on a test set

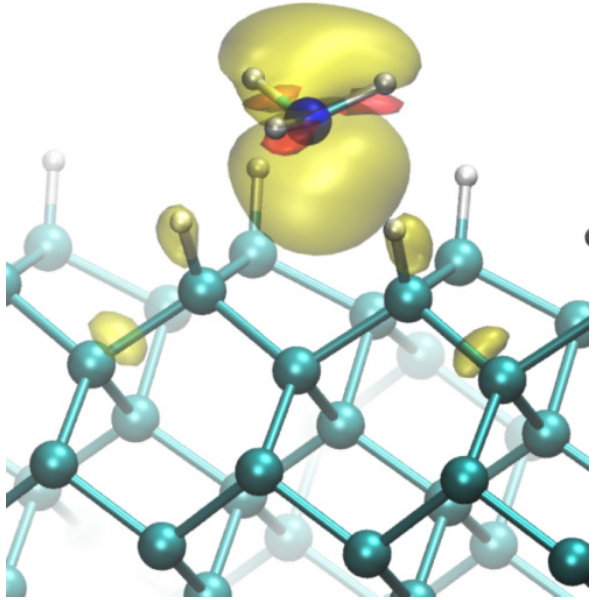


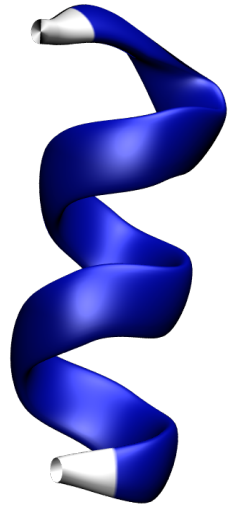
$$E(\mathcal{A}) = \sum_i w_i K(\mathcal{A}, \mathcal{A}_i)$$

Machine learning with a physical mindset

- General applicability: suitable for all systems and all types of properties
- Well-principled: incorporates structure and symmetries of physical laws
- Not only a fancy interpolator: use ML to gain insights and understanding



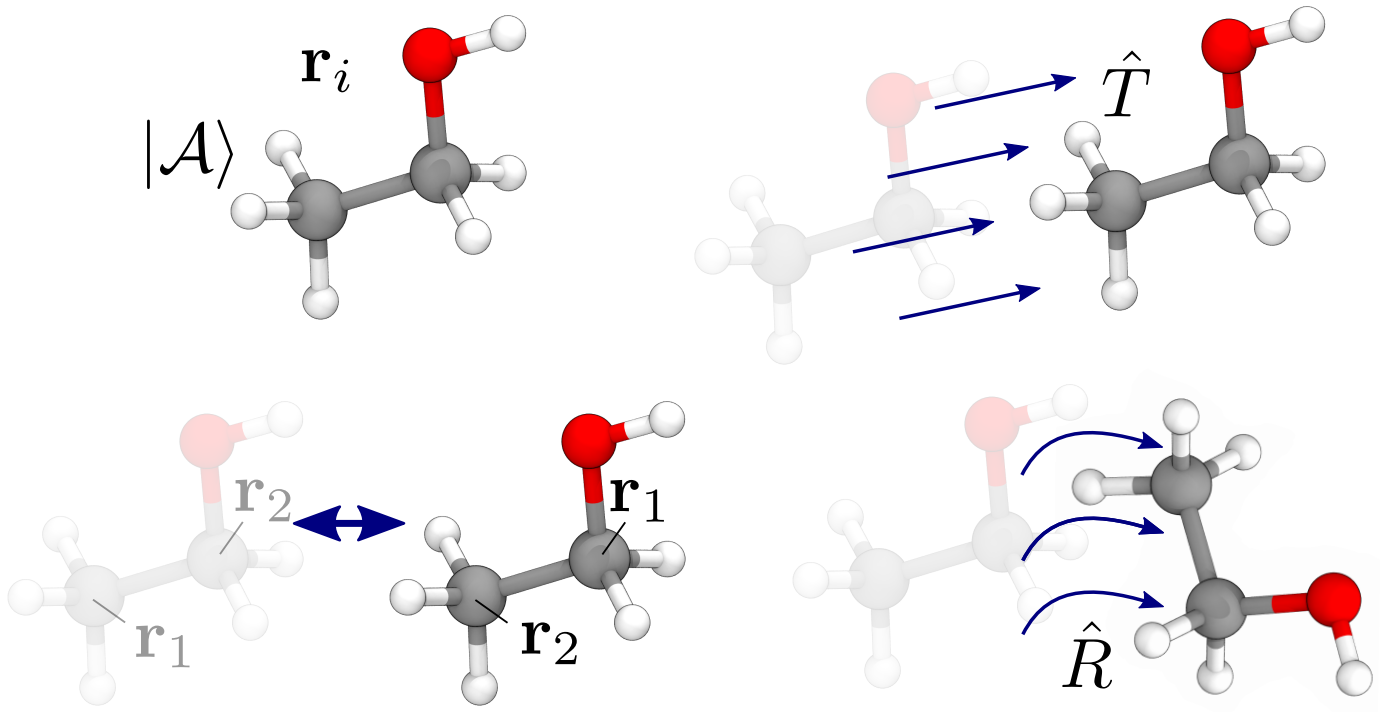




$$\hat{\mathcal{H}}\Psi = E\Psi \qquad E(\mathbf{q}) = \sum_{ij} v(r_{ij}) + \ldots, \qquad E(\mathbf{q}) = ML(\mathbf{q}|\{\mathbf{q}_i, V_i\})$$

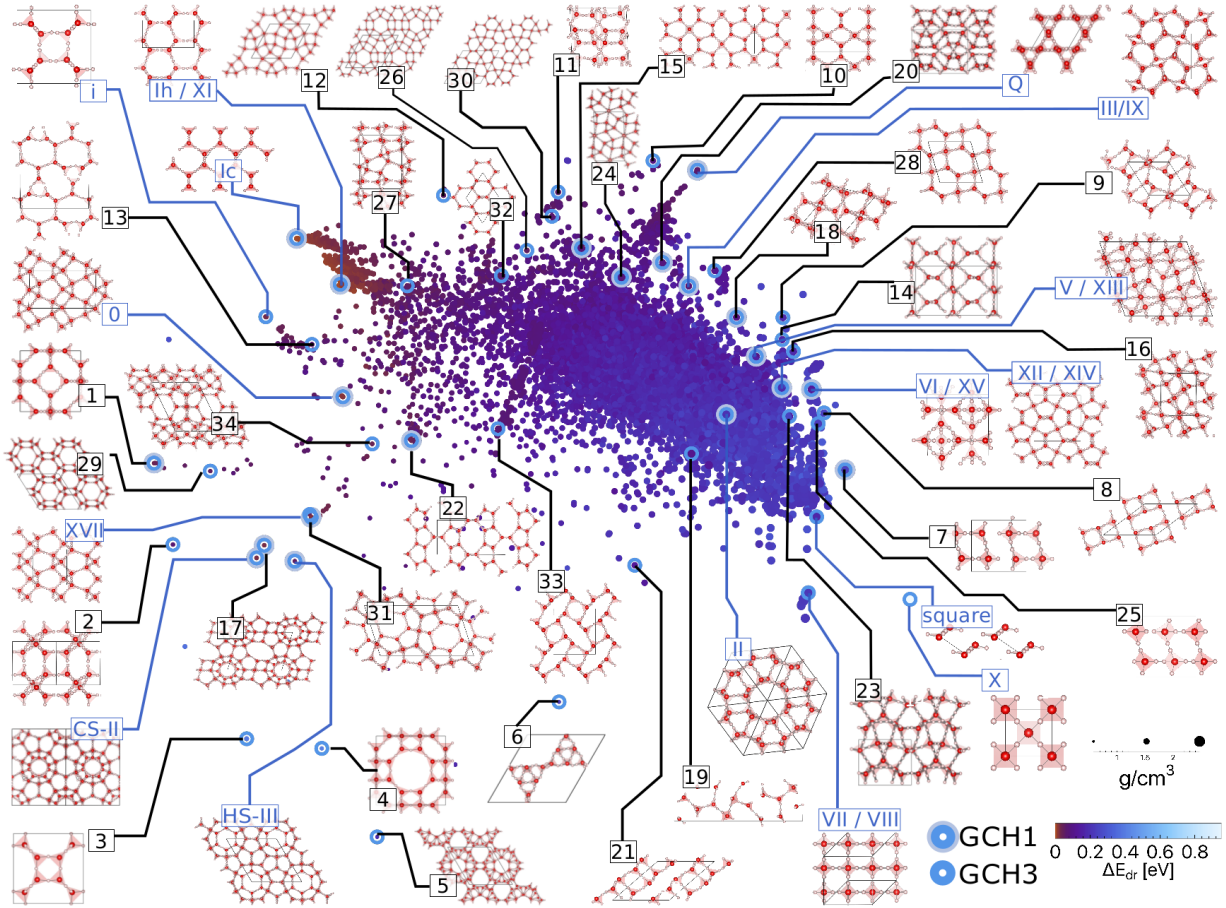
Machine learning with a physical mindset

- General applicability: suitable for all systems and all types of properties
- Well-principled: incorporates structure and symmetries of physical laws
- Not only a fancy interpolator: use ML to gain insights and understanding



Machine learning with a physical mindset

- General applicability: suitable for all systems and all types of properties
- Well-principled: incorporates structure and symmetries of physical laws
- Not only a fancy interpolator: use ML to gain insights and understanding

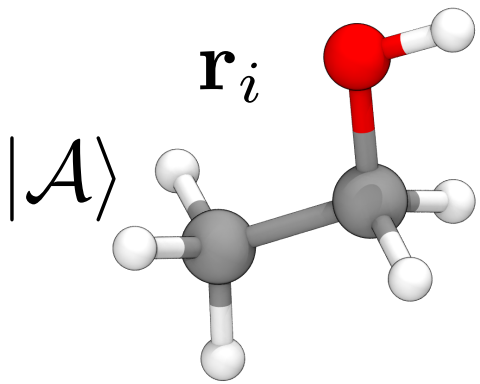


MC, Tribello, Parrinello, PNAS (2011); Engel et al, Nat. Comm. (2018); Anelli et al., PRM (2018); http://interactive.sketchmap.org

A transferable ML model for materials and molecules

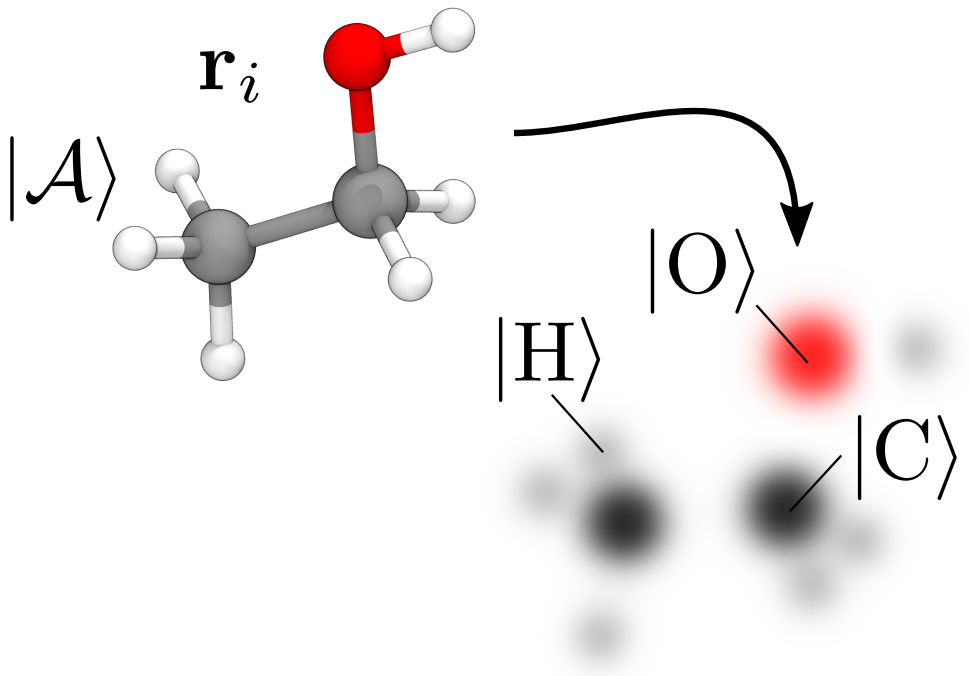
Symmetry-adapted atom-density representations

- Structural representation based on a decorated atom-density vector $|\mathcal{A}\rangle$
- Write in position representation as a sum of atom-centered Gaussians
- Use abstract kets $|\alpha\rangle$ to encode the nature of the atoms
- Permutation-invariant ☺, but not translation-invariant ☹



Symmetry-adapted atom-density representations

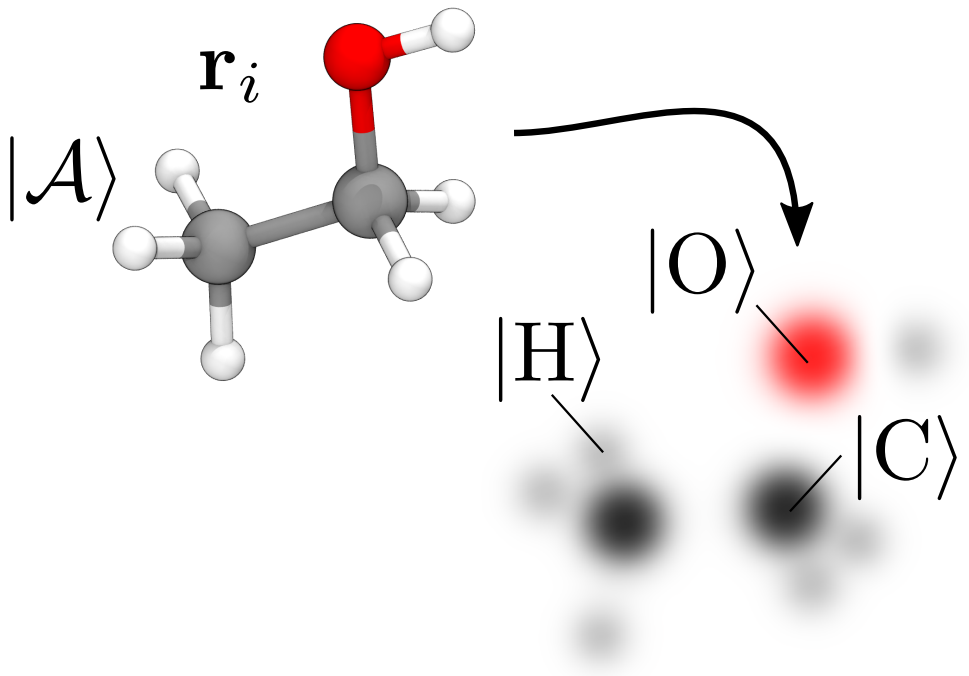
- Structural representation based on a decorated atom-density vector $|\mathcal{A}\rangle$
- Write in position representation as a sum of atom-centered Gaussians
- Use abstract kets $|\alpha\rangle$ to encode the nature of the atoms
- Permutation-invariant ☺, but not translation-invariant ☹



$$\langle \mathbf{r} | \mathcal{A} \rangle = \sum_i g(\mathbf{r} - \mathbf{r}_i) \, |\alpha_i\rangle$$

Symmetry-adapted atom-density representations

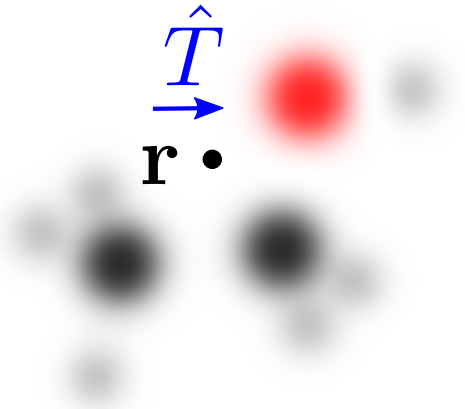
- Structural representation based on a decorated atom-density vector $|\mathcal{A}\rangle$
- Write in position representation as a sum of atom-centered Gaussians
- Use abstract kets $|\alpha\rangle$ to encode the nature of the atoms
- Permutation-invariant ☺, but not translation-invariant ☹



$$\langle \mathbf{r} | \mathcal{A} \rangle = \sum_i g(\mathbf{r} - \mathbf{r}_i) \, |\alpha_i\rangle$$

Translational invariance

- Translational symmetry can be recovered by integration over $\hat{T}$ group
- Integration leads to severe information loss
- Symmetrize tensor products to reduce information loss
- A convolution of Gaussians is a Gaussian . . .
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition



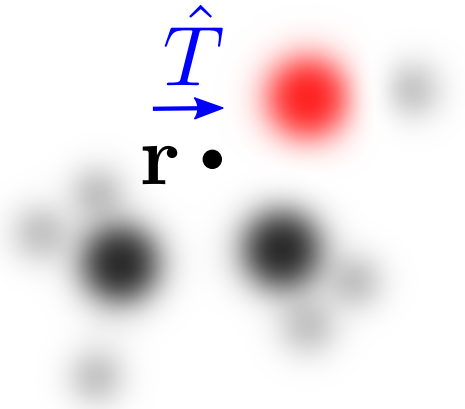
$$\int \mathrm{d}\hat{T} \left\langle \mathbf{r} \middle| \hat{T} \middle| \mathcal{A} \right\rangle =$$

$$\sum_i \int \mathrm{d}\mathbf{t}\, g(\mathbf{r} + \mathbf{t} - \mathbf{r}_i) \left|\alpha_i\right\rangle = \sum_\alpha N_\alpha \left|\alpha\right\rangle$$

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

Translational invariance

- Translational symmetry can be recovered by integration over $\hat{T}$ group
- Integration leads to severe information loss
- Symmetrize tensor products to reduce information loss
- A convolution of Gaussians is a Gaussian . . .
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition



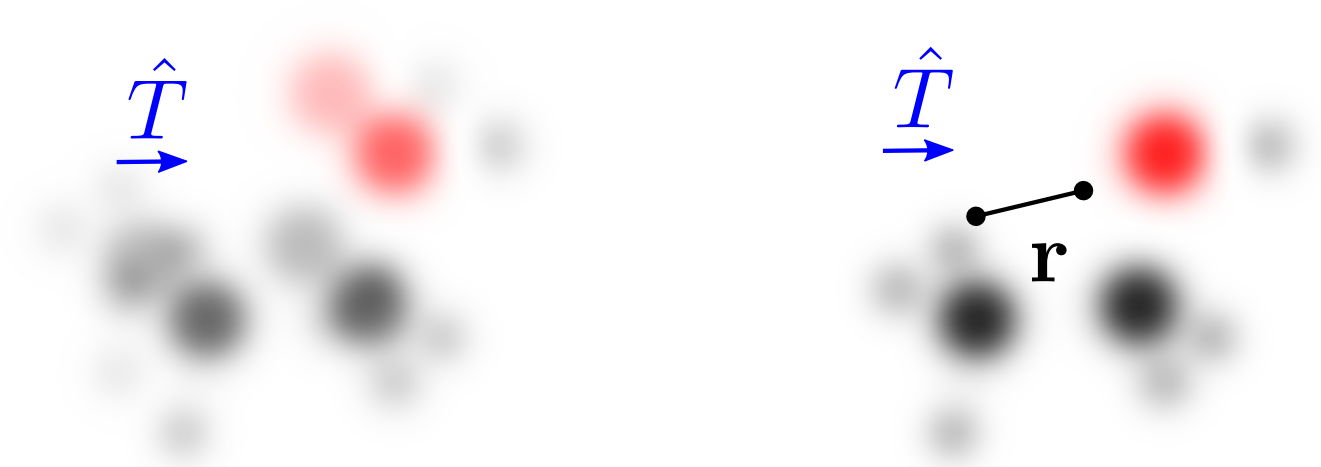
$$\int \mathrm{d}\hat{T} \left\langle \mathbf{r} \middle| \hat{T} \middle| \mathcal{A} \right\rangle =$$

$$\sum_i \int \mathrm{d}\mathbf{t}\, g(\mathbf{r} + \mathbf{t} - \mathbf{r}_i) \left|\alpha_i\right\rangle = \sum_\alpha N_\alpha \left|\alpha\right\rangle$$

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

Translational invariance

- Translational symmetry can be recovered by integration over $\hat{T}$ group
- Integration leads to severe information loss
- Symmetrize tensor products to reduce information loss
- A convolution of Gaussians is a Gaussian . . .
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition



$$\int \mathrm{d}\hat{T}\ \langle \mathbf{r}|\,\hat{T}\,|\mathcal{A}\rangle\,\langle \mathbf{r}'|\,\hat{T}\,|\mathcal{A}\rangle = \int \mathrm{d}\mathbf{r}'\,\langle \mathbf{r}'|\mathcal{A}\rangle\,\langle \mathbf{r}'+\mathbf{r}|\mathcal{A}\rangle$$

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

Translational invariance

- Translational symmetry can be recovered by integration over $\hat{T}$ group
- Integration leads to severe information loss
- Symmetrize tensor products to reduce information loss
- A convolution of Gaussians is a Gaussian . . .
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition

$$\int \mathrm{d}\mathbf{t}\, g\left(\mathbf{t}+\mathbf{r}-\mathbf{r}_i\right) g\left(\mathbf{t}+\mathbf{r}'-\mathbf{r}_j\right) =$$

$$\int \mathrm{d}\mathbf{t}\, g\left(\mathbf{t}\right) g\left(\mathbf{t}+\mathbf{r}'-\mathbf{r}_j-\mathbf{r}+\mathbf{r}_i\right) =$$

$$\tilde{g}\left(\left(\mathbf{r}'-\mathbf{r}\right)-\left(\mathbf{r}_j-\mathbf{r}_i\right)\right)$$

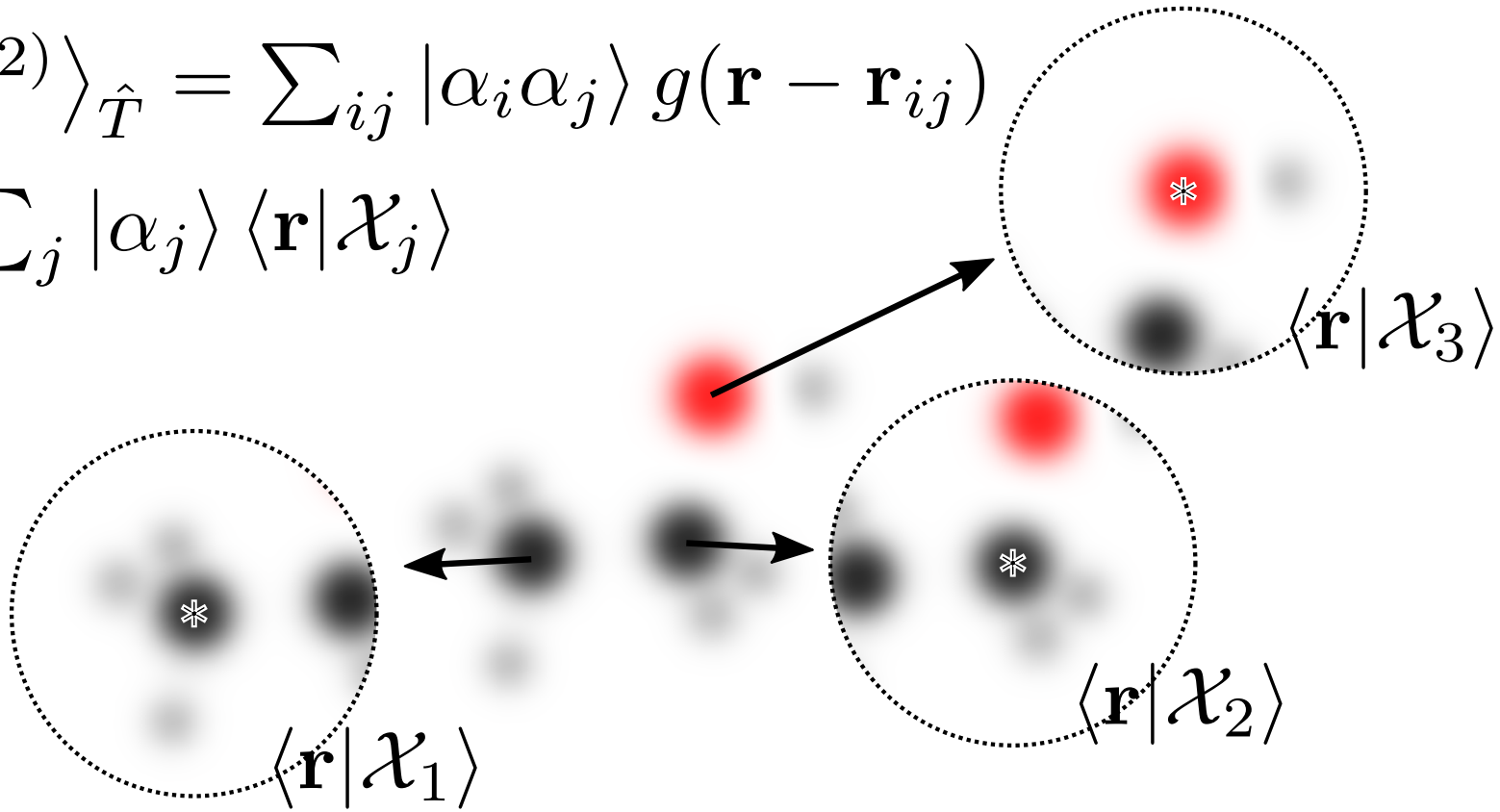
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

Translational invariance

- Translational symmetry can be recovered by integration over $\hat{T}$ group
- Integration leads to severe information loss
- Symmetrize tensor products to reduce information loss
- A convolution of Gaussians is a Gaussian . . .
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition

$$\left\langle \mathbf{r}\middle|\mathcal{A}^{(2)}\right\rangle_{\hat{T}} = \sum_{ij} \left|\alpha_i\alpha_j\right\rangle g(\mathbf{r}-\mathbf{r}_{ij})$$

$$= \sum_j \left|\alpha_j\right\rangle \left\langle \mathbf{r}\middle|\mathcal{X}_j\right\rangle$$



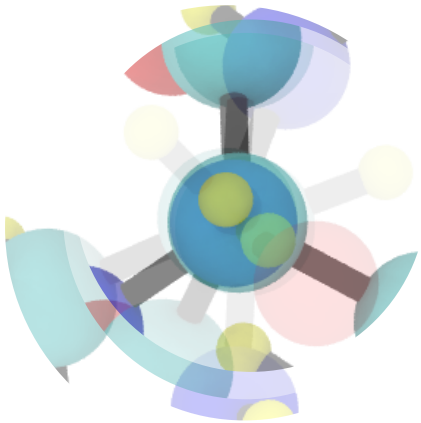
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

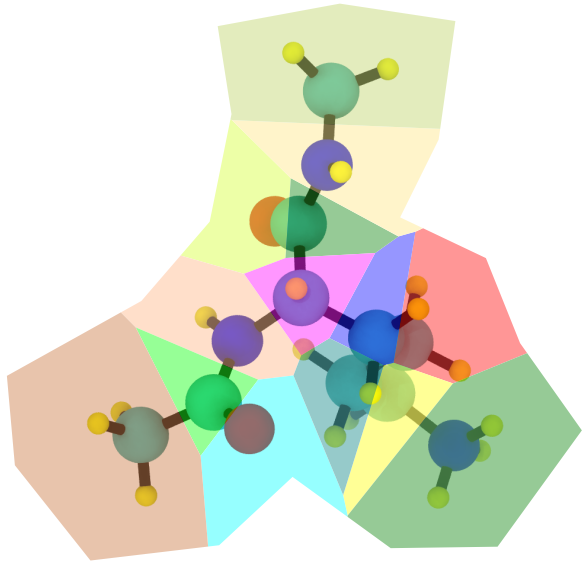
Additive Property Models & Beyond

- Crucial observation: learning with an average kernel is equivalent to learning an atom-centered *additive* energy model

$$E(A) = \sum_i w_i K(A, A_i) \qquad \epsilon(\mathcal{X}) = \sum_i w_i k(\mathcal{X}, \mathcal{X}_i)$$
$$K(A, B) = \sum_{i \in A, j \in B} k(\mathcal{X}_i, \mathcal{X}_j) \iff E(A) = \sum_{i \in A} \epsilon(\mathcal{X}_i)$$

- Entropy-regularized match provides a natural way to go beyond additive models, retaining a local environment expansion





$$K(A, B) = \sum_{i,j} k(\mathcal{X}_i^A, \mathcal{X}_j^B)$$

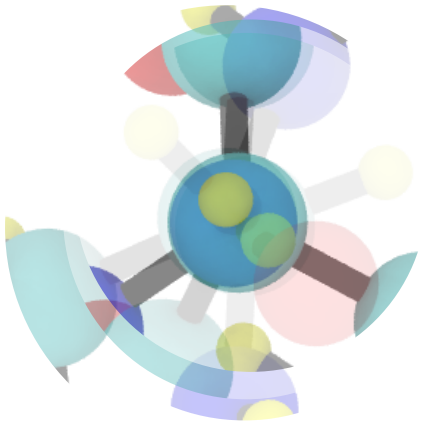
$$E(A) = \sum_i \epsilon(\mathcal{X}_i^A)$$

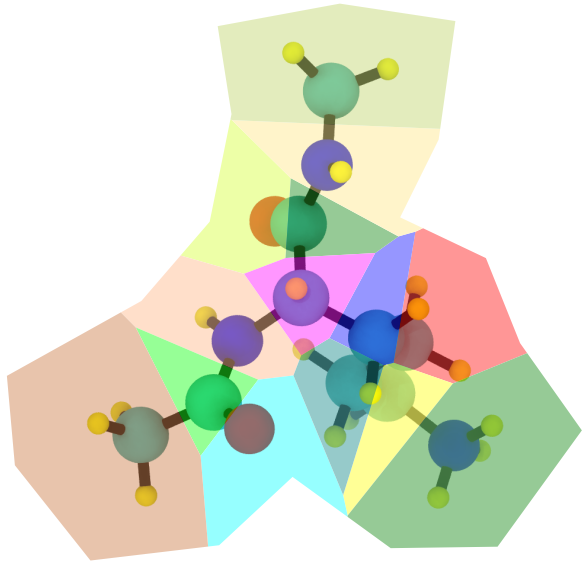
Additive Property Models & Beyond

- Crucial observation: learning with an average kernel is equivalent to learning an atom-centered *additive* energy model

$$\begin{array}{c} E(A) = \sum_i w_i K(A, A_i) \\ K(A,B) = \sum_{i \in A, j \in B} k(\mathcal{X}_i, \mathcal{X}_j) \end{array} \iff \begin{array}{c} \epsilon(\mathcal{X}) = \sum_i w_i k(\mathcal{X}, \mathcal{X}_i) \\ E(A) = \sum_{i \in A} \epsilon(\mathcal{X}_i) \end{array}$$

- Entropy-regularized match provides a natural way to go beyond additive models, retaining a local environment expansion





$$K(A,B) = \sum_{i,j} k(\mathcal{X}_i^A, \mathcal{X}_j^B)$$

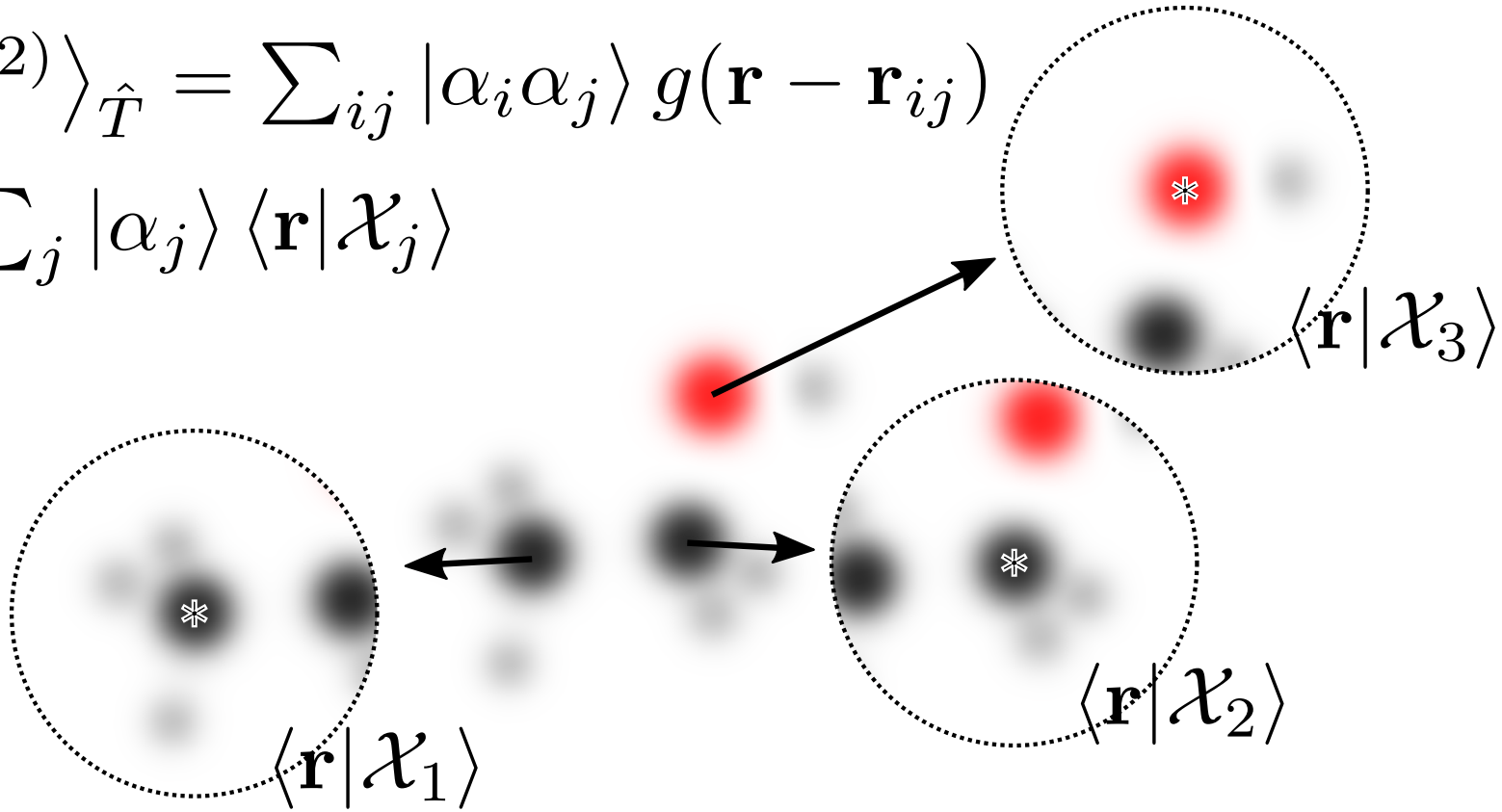
$$E(A) = \sum_i \epsilon(\mathcal{X}_i^A)$$

De, Bartók, Csányi, **MC**, PCCP (2016)

Rotational invariance

- How about rotations? Not invariant, $\left\langle \hat{R}\mathbf{r}\middle|\mathcal{X}\right\rangle_{\hat{T}} \neq \langle \mathbf{r}|\mathcal{X}\rangle_{\hat{T}}$
- Integration over rotation yields $\left|\mathcal{X}^{(1)}\right\rangle$ (atom pair distribution!)
- Again, we can use tensor products to retain information.
  $\int \mathrm{d}\hat{R}\hat{R}\,|\mathcal{X}\rangle \otimes |\mathcal{X}\rangle = \left|\mathcal{X}^{(2)}\right\rangle$ incorporates 3-body correlations

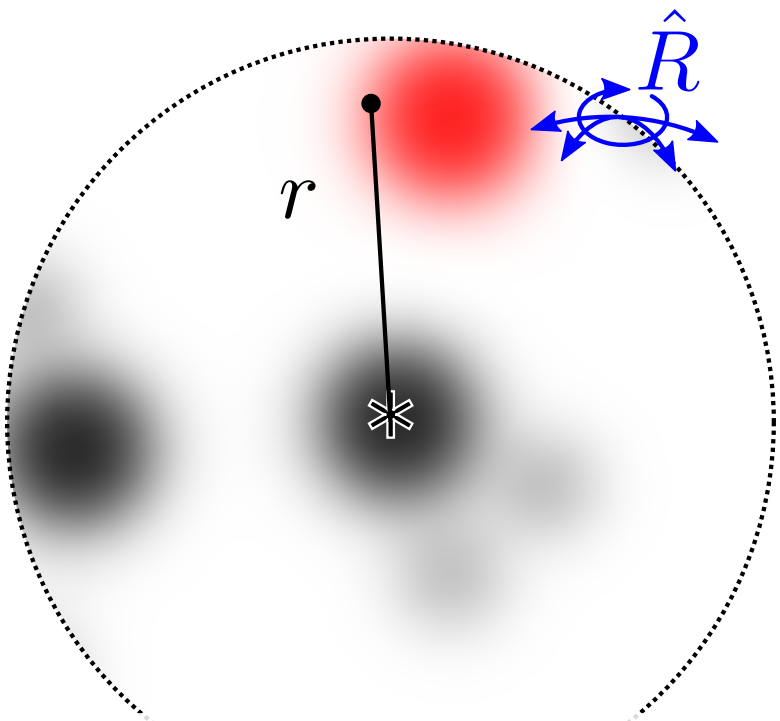
$$\left\langle \mathbf{r}\middle|\mathcal{A}^{(2)}\right\rangle_{\hat{T}} = \sum_{ij} |\alpha_i\alpha_j\rangle\, g(\mathbf{r}-\mathbf{r}_{ij})$$
$$= \sum_j |\alpha_j\rangle\, \langle \mathbf{r}|\mathcal{X}_j\rangle$$



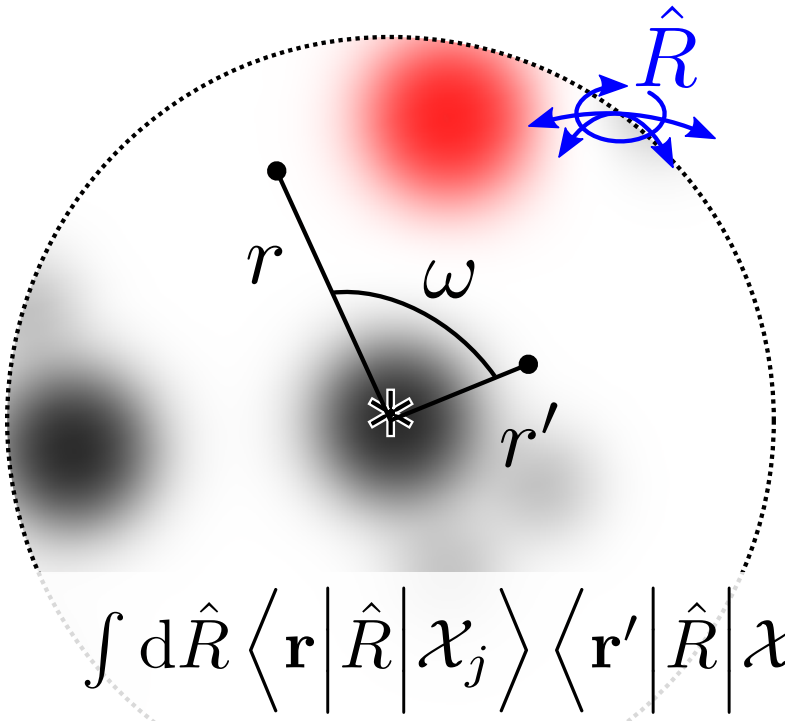
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

Rotational invariance

- How about rotations? Not invariant, $\left\langle \hat{R}\mathbf{r}\middle|\mathcal{X}\right\rangle_{\hat{T}} \neq \langle \mathbf{r}|\mathcal{X}\rangle_{\hat{T}}$
- Integration over rotation yields $\left|\mathcal{X}^{(1)}\right\rangle$ (atom pair distribution!)
- Again, we can use tensor products to retain information.
  $\int \mathrm{d}\hat{R}\hat{R}\,|\mathcal{X}\rangle \otimes |\mathcal{X}\rangle = \left|\mathcal{X}^{(2)}\right\rangle$ incorporates 3-body correlations



$$\int \mathrm{d}\hat{R}\left\langle \mathbf{r}\middle|\hat{R}\middle|\mathcal{X}_j\right\rangle \rightarrow \left\langle r\middle|\mathcal{X}_j^{(1)}\right\rangle$$

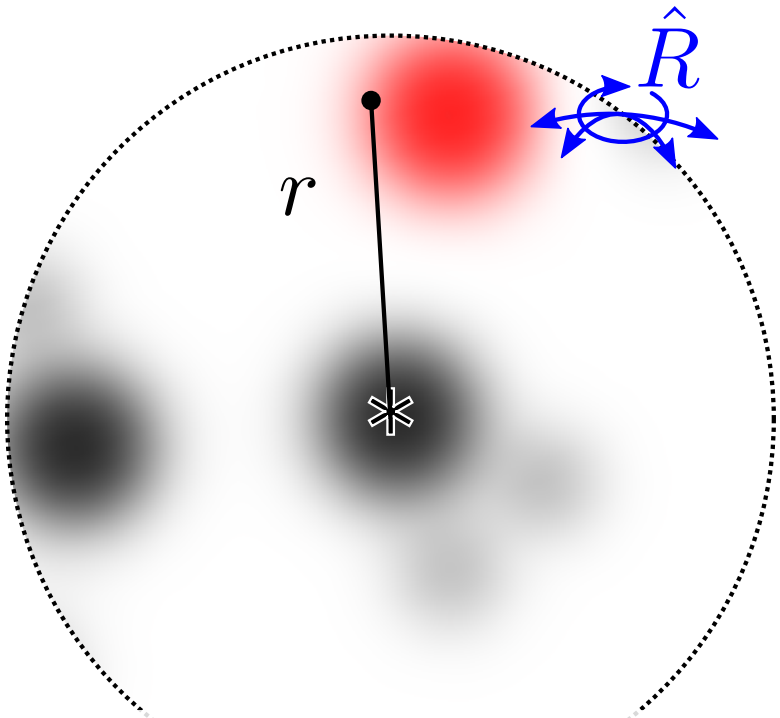


$$\int \mathrm{d}\hat{R}\left\langle \mathbf{r}\middle|\hat{R}\middle|\mathcal{X}_j\right\rangle \left\langle \mathbf{r}'\middle|\hat{R}\middle|\mathcal{X}_j\right\rangle \rightarrow \left\langle rr'\omega\middle|\mathcal{X}_j^{(2)}\right\rangle$$

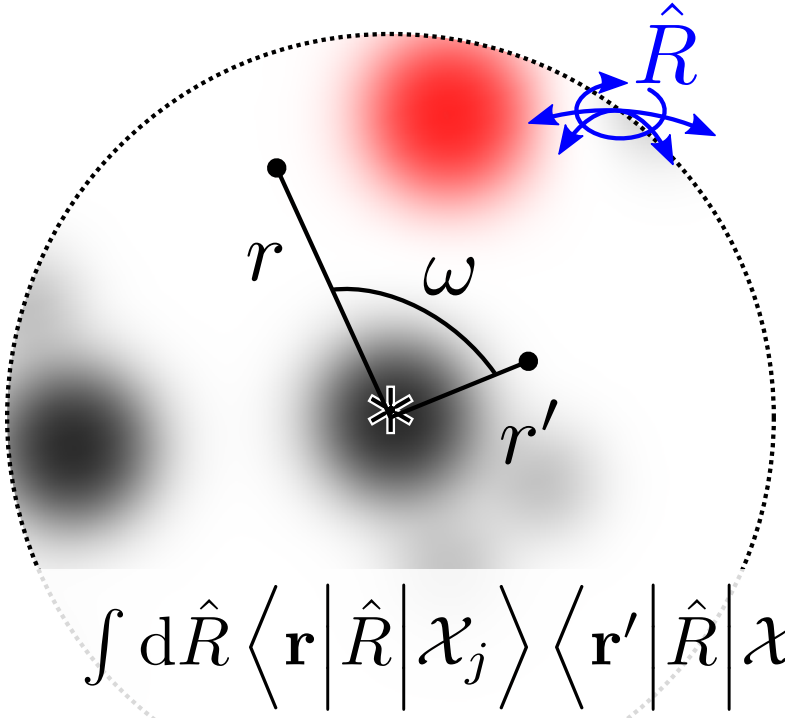
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

Rotational invariance

- How about rotations? Not invariant, $\left\langle \hat{R}\mathbf{r}\middle|\mathcal{X}\right\rangle_{\hat{T}} \neq \langle \mathbf{r}|\mathcal{X}\rangle_{\hat{T}}$
- Integration over rotation yields $\left|\mathcal{X}^{(1)}\right\rangle$ (atom pair distribution!)
- Again, we can use tensor products to retain information.
  $\int \mathrm{d}\hat{R}\hat{R}\,|\mathcal{X}\rangle \otimes |\mathcal{X}\rangle = \left|\mathcal{X}^{(2)}\right\rangle$ incorporates 3-body correlations



$$\int \mathrm{d}\hat{R}\left\langle \mathbf{r}\middle|\hat{R}\middle|\mathcal{X}_j\right\rangle \rightarrow \left\langle r\middle|\mathcal{X}_j^{(1)}\right\rangle$$

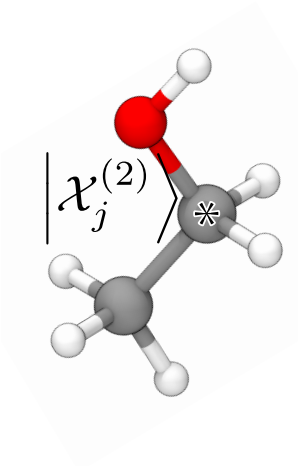


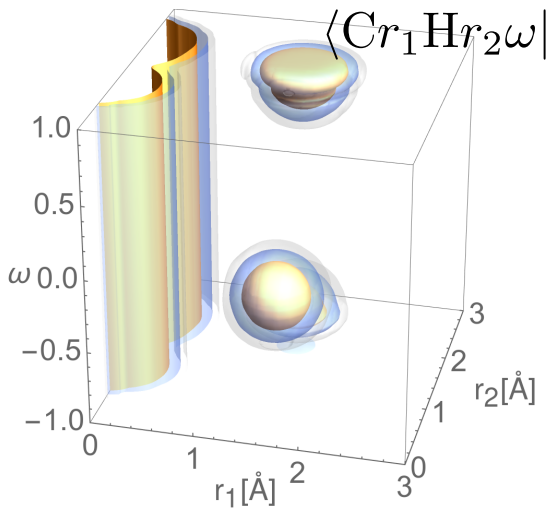
$$\int \mathrm{d}\hat{R}\left\langle \mathbf{r}\middle|\hat{R}\middle|\mathcal{X}_j\right\rangle \left\langle \mathbf{r}'\middle|\hat{R}\middle|\mathcal{X}_j\right\rangle \rightarrow \left\langle rr'\omega\middle|\mathcal{X}_j^{(2)}\right\rangle$$

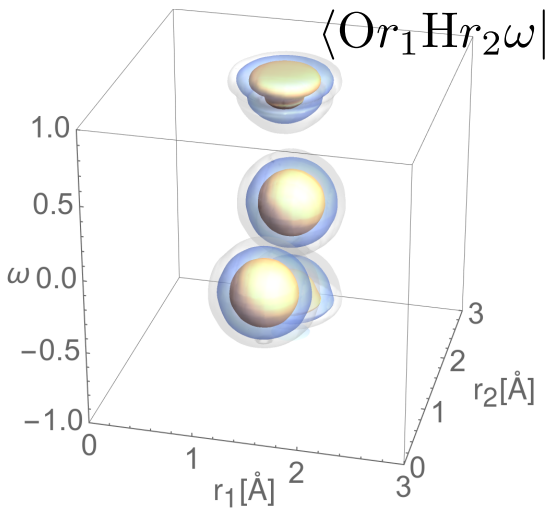
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408

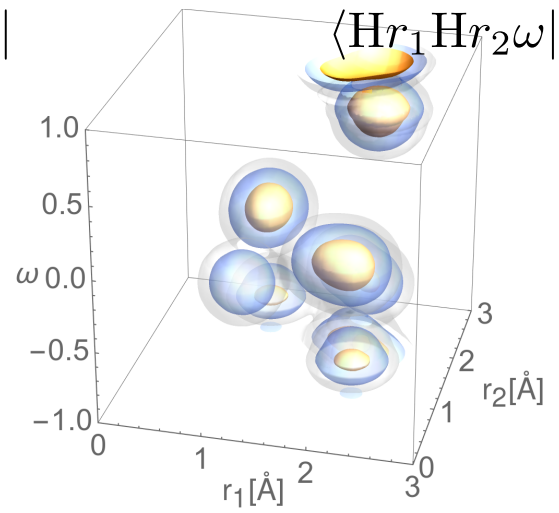
Rotational invariance

- How about rotations? Not invariant, $\left\langle \hat{R}\mathbf{r}\middle|\mathcal{X}\right\rangle_{\hat{T}} \neq \langle \mathbf{r}|\mathcal{X}\rangle_{\hat{T}}$
- Integration over rotation yields $\left|\mathcal{X}^{(1)}\right\rangle$ (atom pair distribution!)
- Again, we can use tensor products to retain information.
  $\int \mathrm{d}\hat{R}\hat{R}\,|\mathcal{X}\rangle \otimes |\mathcal{X}\rangle = \left|\mathcal{X}^{(2)}\right\rangle$ incorporates 3-body correlations









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 $\left|\mathcal{A}^{(2)}\right\rangle_{\hat{T}}$)
  - Projection on symmetry functions (Behler-Parrinello)

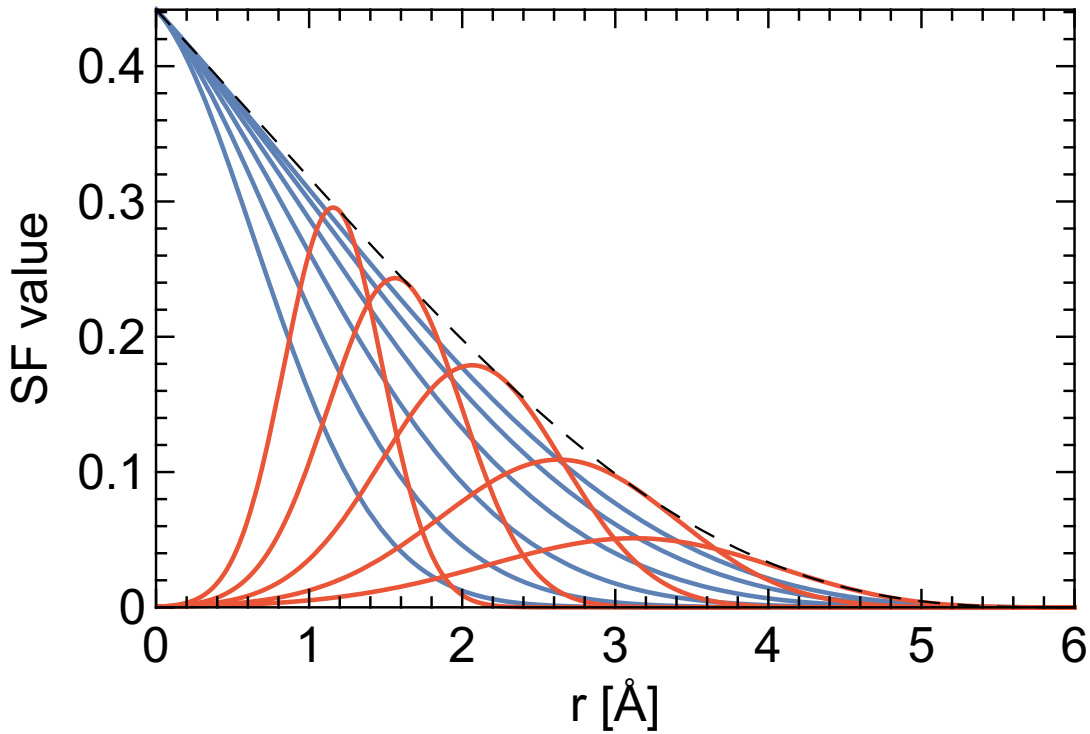
$$\left\langle \mathbf{k}\middle|\mathcal{A}^{(2)}\right\rangle_{\hat{T}} = \sum_{ij} \left|\alpha_i\alpha_j\right\rangle e^{i\mathbf{k}\cdot\mathbf{r}_{ij}}$$

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

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 $\left|\mathcal{A}^{(2)}\right\rangle_{\hat{T}}$)
  - Projection on symmetry functions (Behler-Parrinello)

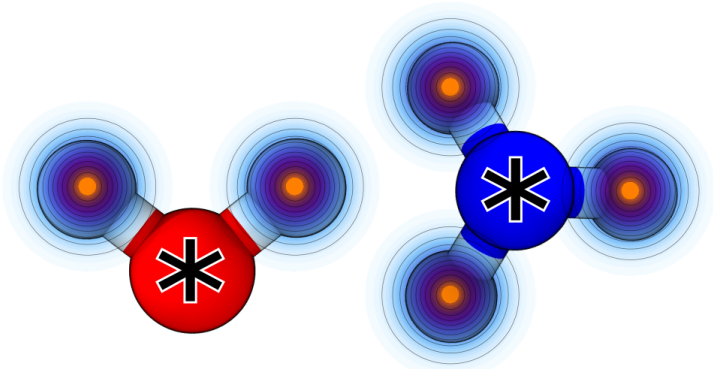
$$\left\langle \alpha\beta G_2 \middle| \mathcal{X}_j \right\rangle = \left\langle \alpha \middle| \alpha_j \right\rangle \int \mathrm{d}r G_2(r) \left\langle \beta r \middle| \mathcal{X}_j^{(1)} \right\rangle_{\hat{R}, g\to\delta}$$



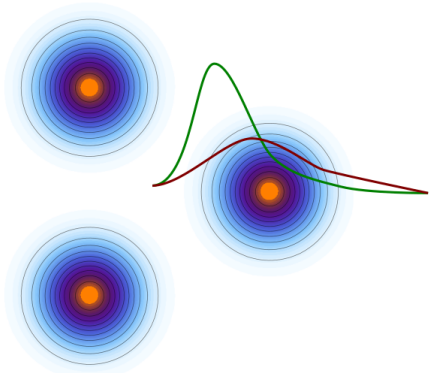
Willatt, Musil, **MC**, JCP (2019), https://arxiv.org/pdf/1807.00408;
R. Drautz, PRB (2019)

The connection with SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product can be written in a breeze
  $\rightarrow$ SOAP power spectrum!
- Alternative construction, but fully equivalent to kernel formulation



$$\langle \mathbf{r}|\mathcal{X}_j\rangle = \psi(\mathbf{r}) = \sum_i g(\mathbf{r}-\mathbf{r}_{ij})$$

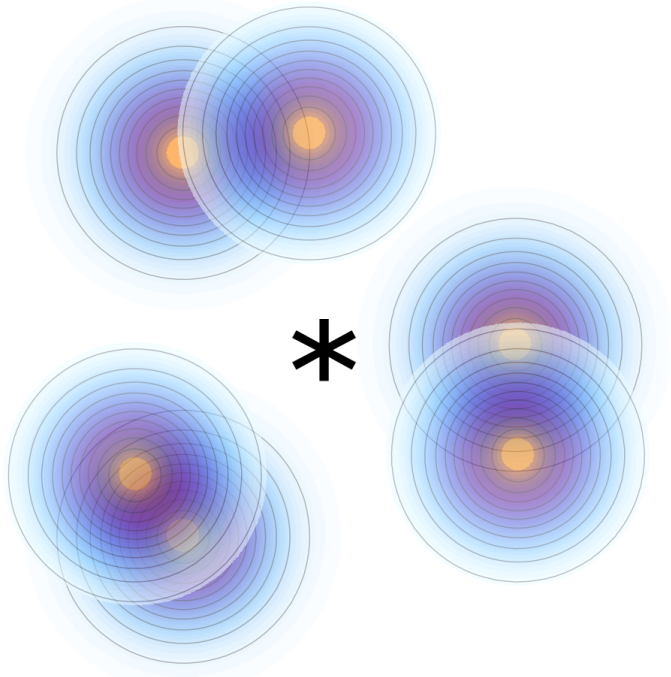


$$\langle nlm|\mathcal{X}_j\rangle = \int \mathrm{d}\mathbf{x}\, \psi(\mathbf{r}) R_n(r) Y_m^l(\hat{\mathbf{r}})$$

Bartók, Kondor, Csányi, PRB (2013)

The connection with SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product can be written in a breeze
  $\rightarrow$ SOAP power spectrum!
- Alternative construction, but fully equivalent to kernel formulation

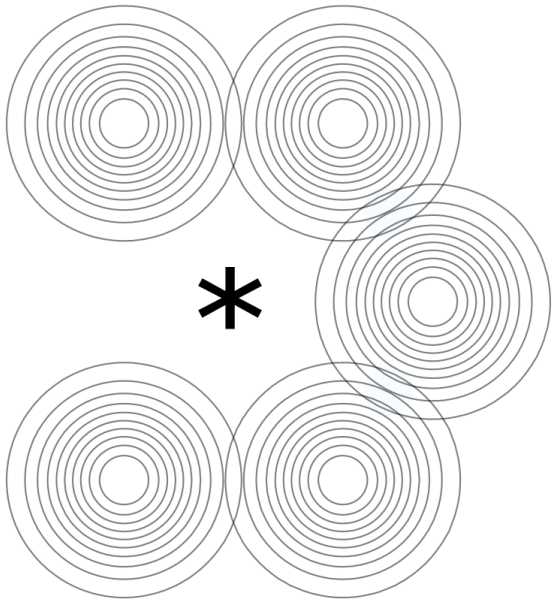


$$\left\langle nn'l\middle|\mathcal{X}^{(2)}\right\rangle_{\hat{R}} = \sum_m \langle nlm|\mathcal{X}\rangle \langle n'lm|\mathcal{X}\rangle$$

Bartók, Kondor, Csányi, PRB (2013)

The connection with SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product can be written in a breeze
  $\rightarrow$ SOAP power spectrum!
- Alternative construction, but fully equivalent to kernel formulation

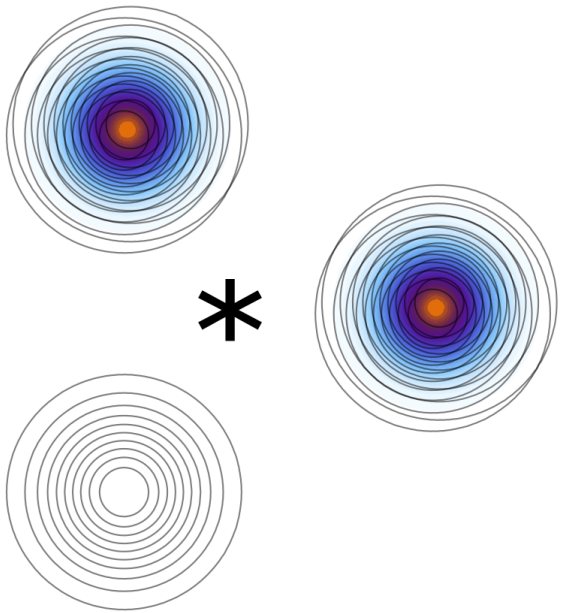


$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim \int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r}) \psi'(\hat{R}\mathbf{r}) \mathrm{d}\mathbf{r} \right|^{\nu}$$

Bartók, Kondor, Csányi, PRB (2013)

The connection with SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product can be written in a breeze
  $\rightarrow$ SOAP power spectrum!
- Alternative construction, but fully equivalent to kernel formulation

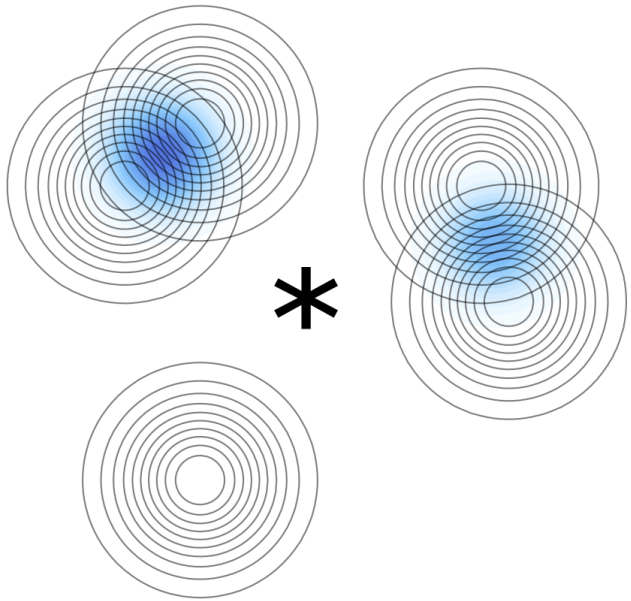


$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim \int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r}) \psi'(\hat{R}\mathbf{r}) \mathrm{d}\mathbf{r} \right|^{\nu}$$

Bartók, Kondor, Csányi, PRB (2013)

The connection with SOAP

- What if we use radial functions and spherical harmonics?
- Symmetrized tensor product can be written in a breeze
  $\rightarrow$ SOAP power spectrum!
- Alternative construction, but fully equivalent to kernel formulation



$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim \int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r}) \psi'(\hat{R}\mathbf{r}) \mathrm{d}\mathbf{r} \right|^{\nu}$$

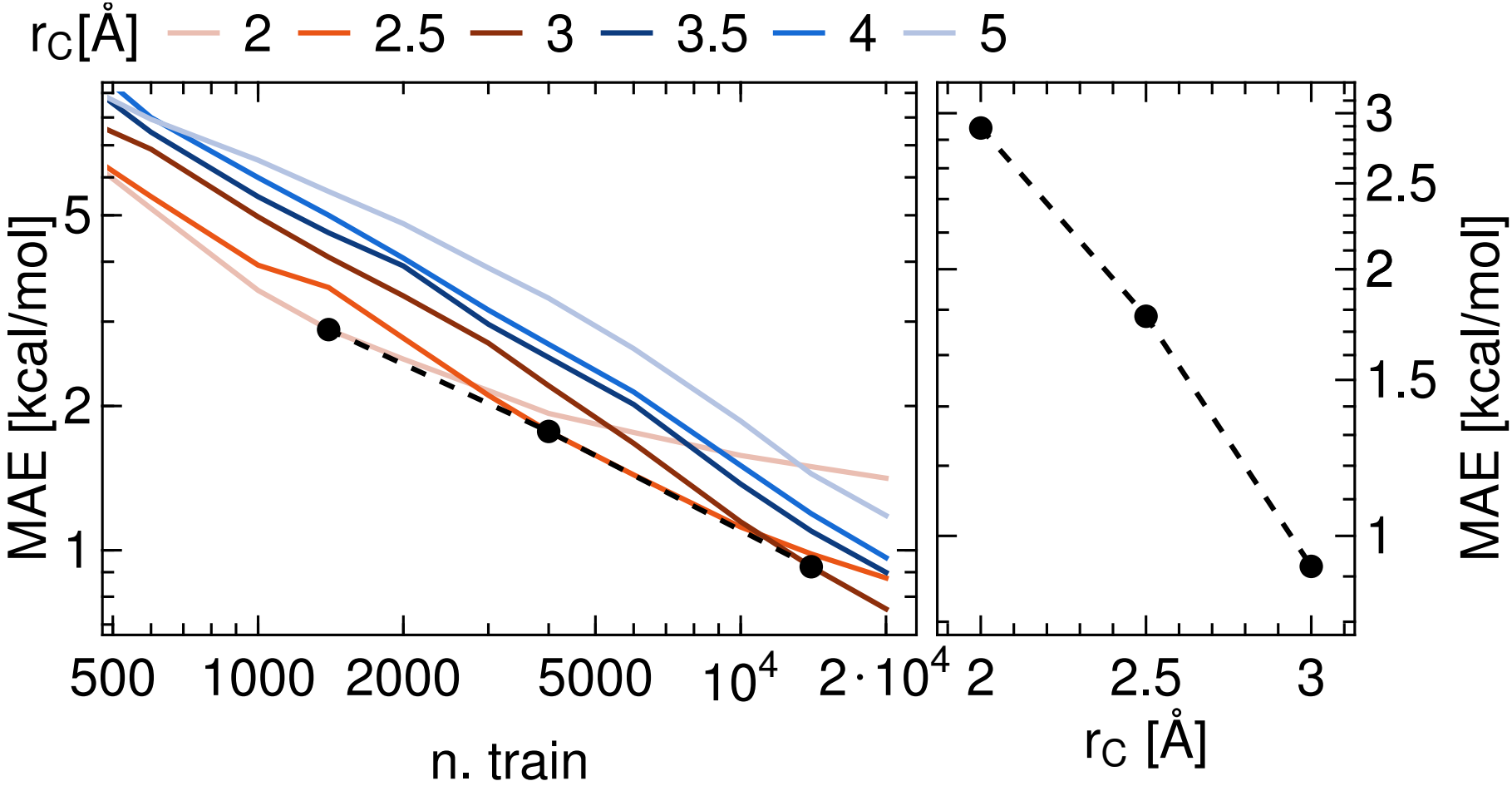
Bartók, Kondor, Csányi, PRB (2013)

Optimizing representations

and what we learn in the process

Understanding the range of interactions

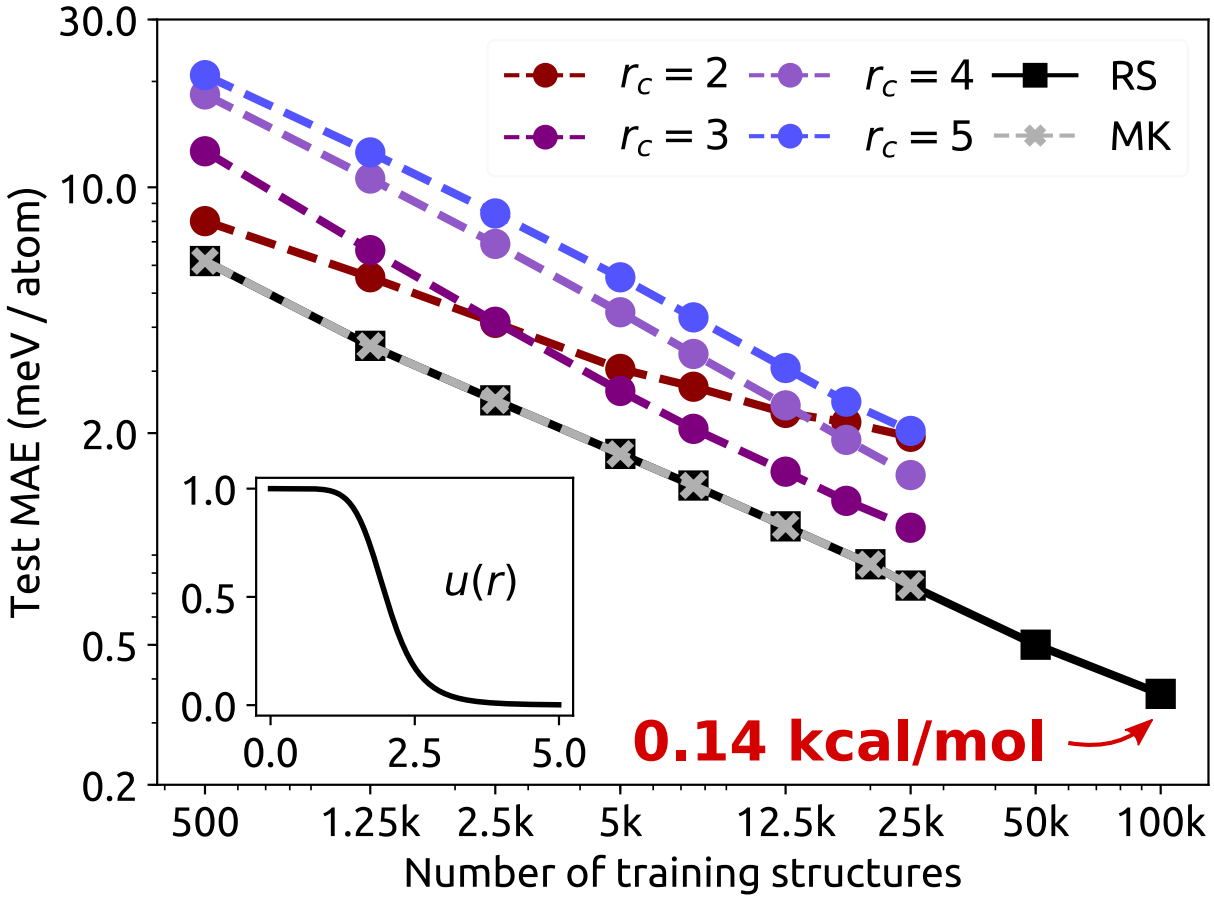
- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel $K(A,B) = \sum_i w_i K_i(A,B)$ yields the best of all worlds. Same results can be achieved by optimized radial scaling of $\langle \mathbf{r}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



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

Understanding the range of interactions

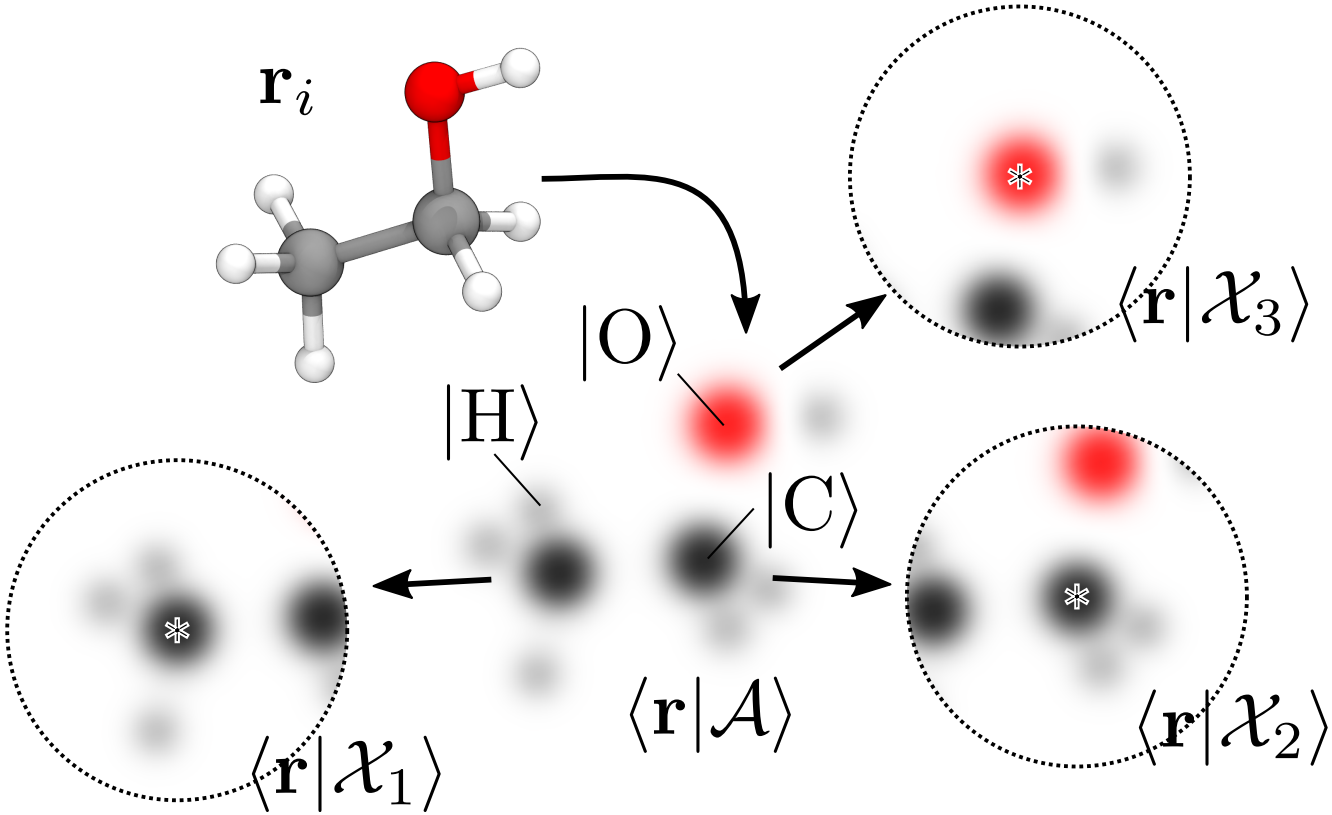
- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel $K(A,B) = \sum_i w_i K_i(A,B)$ yields the best of all worlds. Same results can be achieved by optimized radial scaling of $\langle \mathbf{r}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



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

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

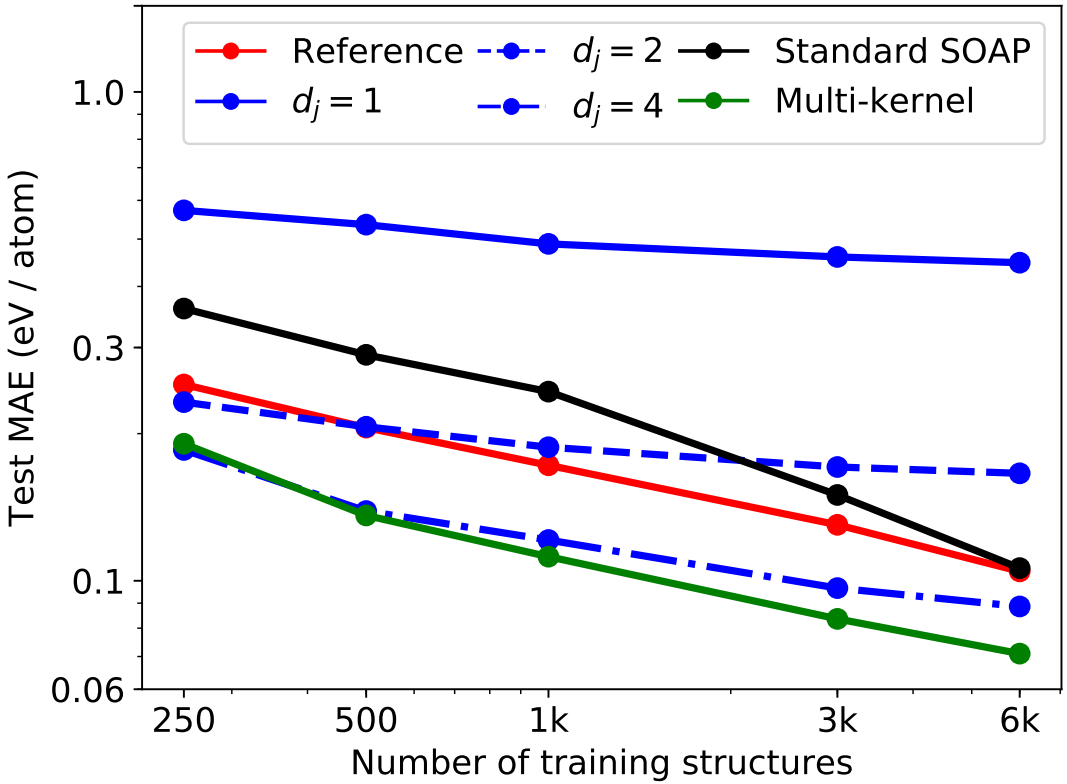
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective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

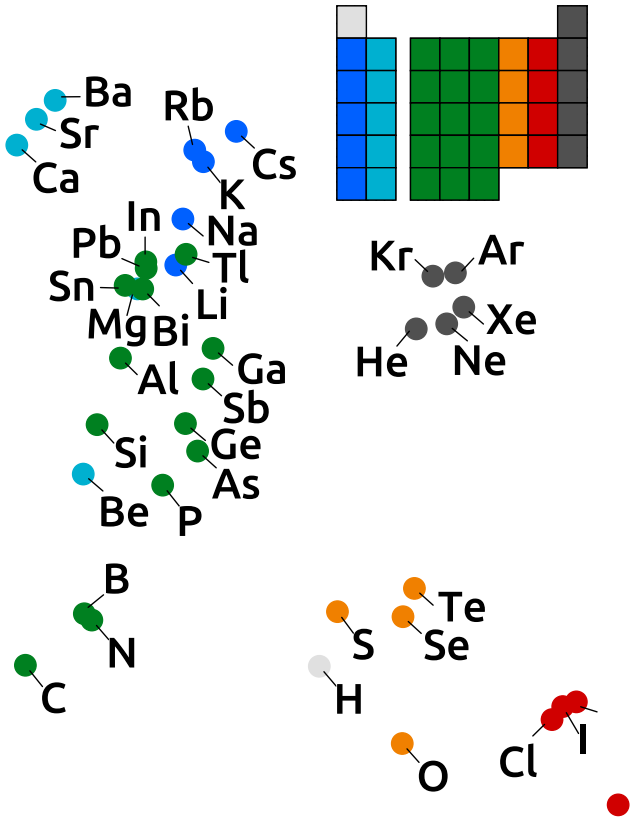
A data-driven periodic table of the elements

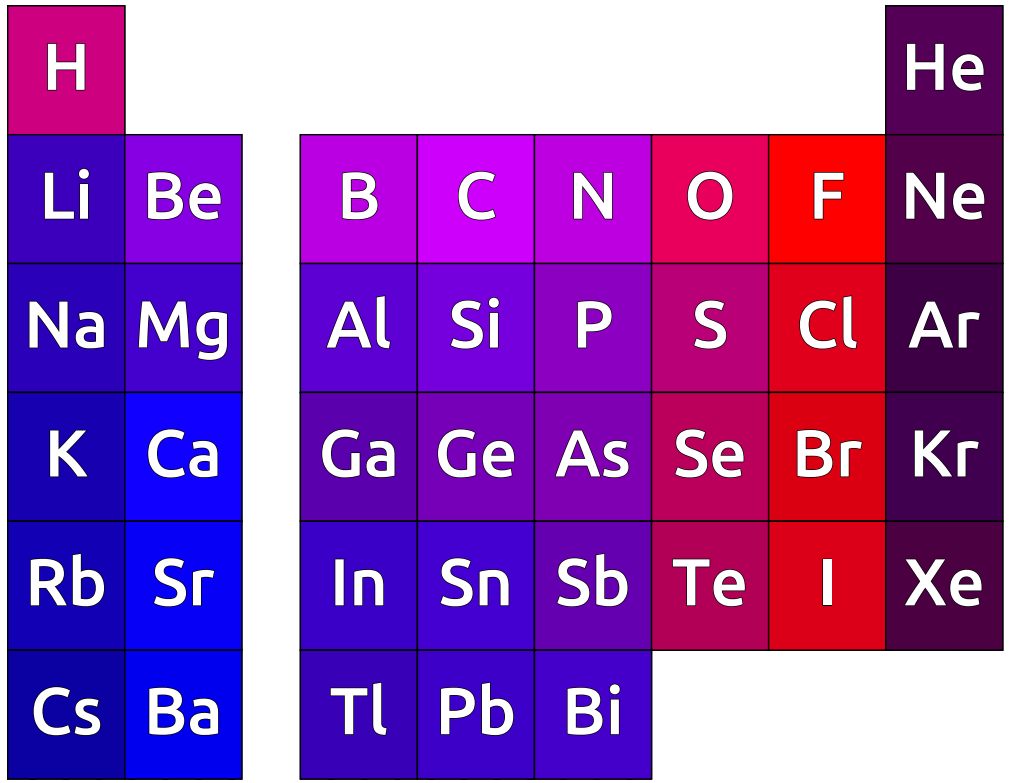
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|\mathrm{H}\rangle$, $|\mathrm{O}\rangle$, . . .
- Expand each ket in a finite basis, $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$. Optimize coefficients
- Dramatic reduction of the descriptor space, more effective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



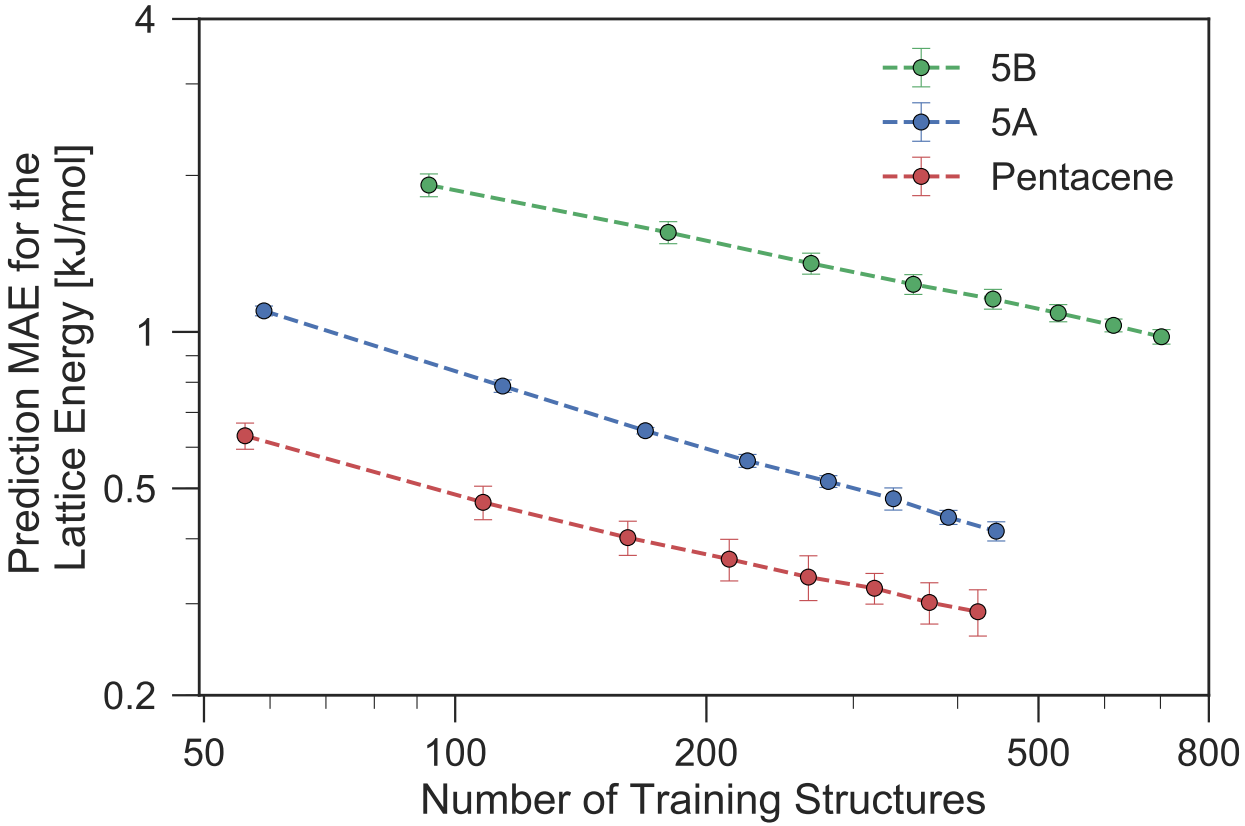


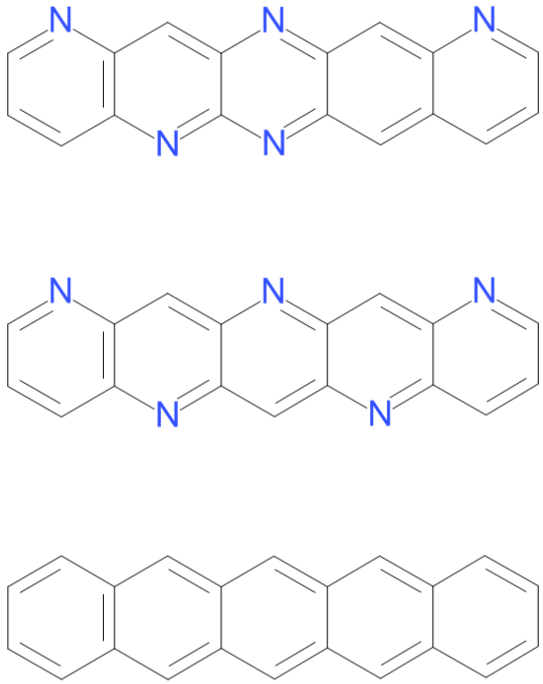
Willatt, Musil, **MC**, PCCP (2018); *[data: Elpasolites, von Lilienfeld&C]*

Examples & Applications

Accurate predictions for molecular crystals

- Substituted pentacenes - model systems for molecular electronics
- Easily achieve sub-kcal/mol accuracy, with REMatch-SOAP kernels

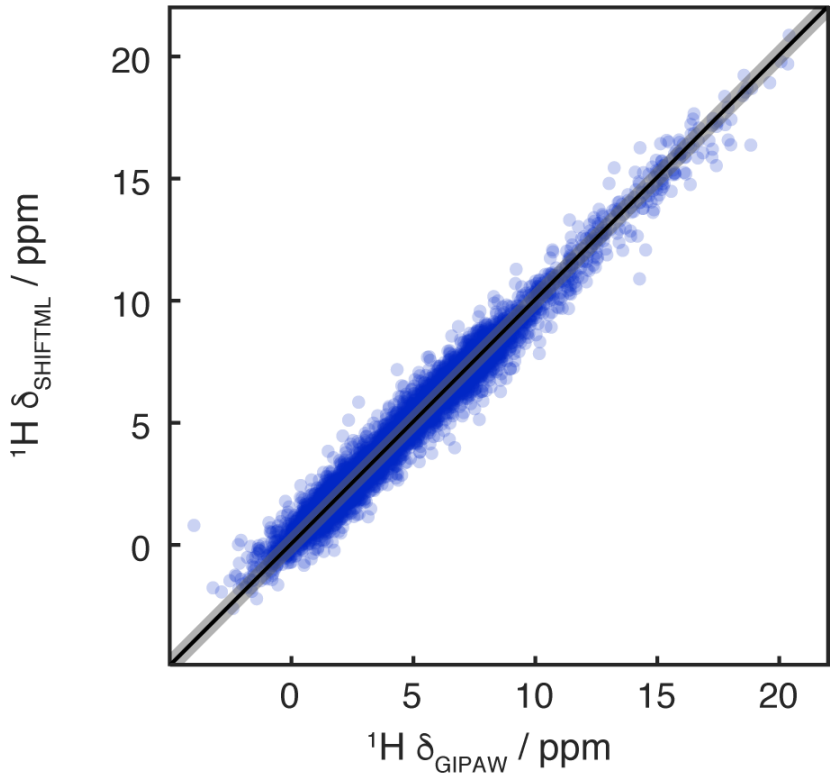


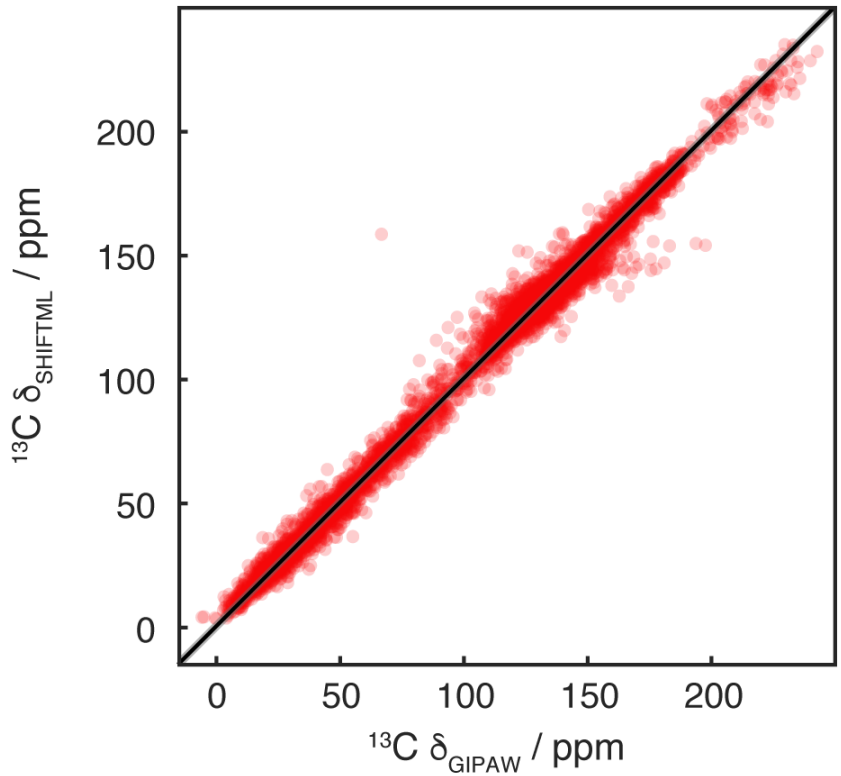


Musil, De, Yang, Campbell, Day, **MC**, Chemical Science (2018)

More than interatomic potentials

- Solid-state NMR relies on GIPAW-DFT to determine crystal structure of molecular materials
- Train a ML model on 2000 CSD structures, predict chemical shieldings with DFT accuracy (RMSE H: 0.5, C: 5, N: 13, 0: 18 ppm)
- Accurate enough to do structure determination!

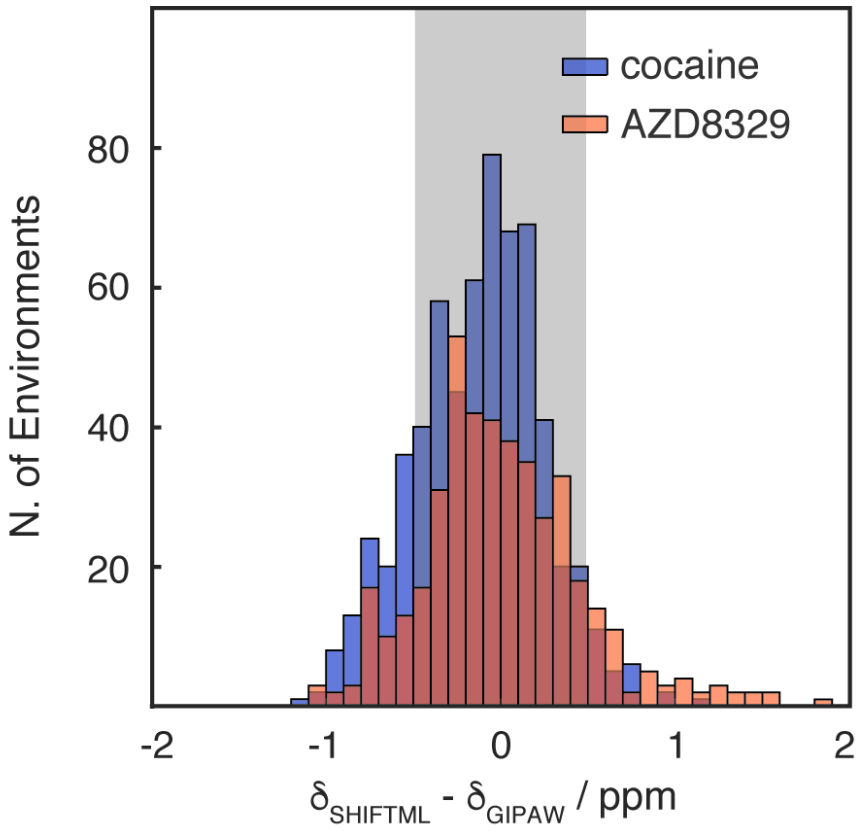


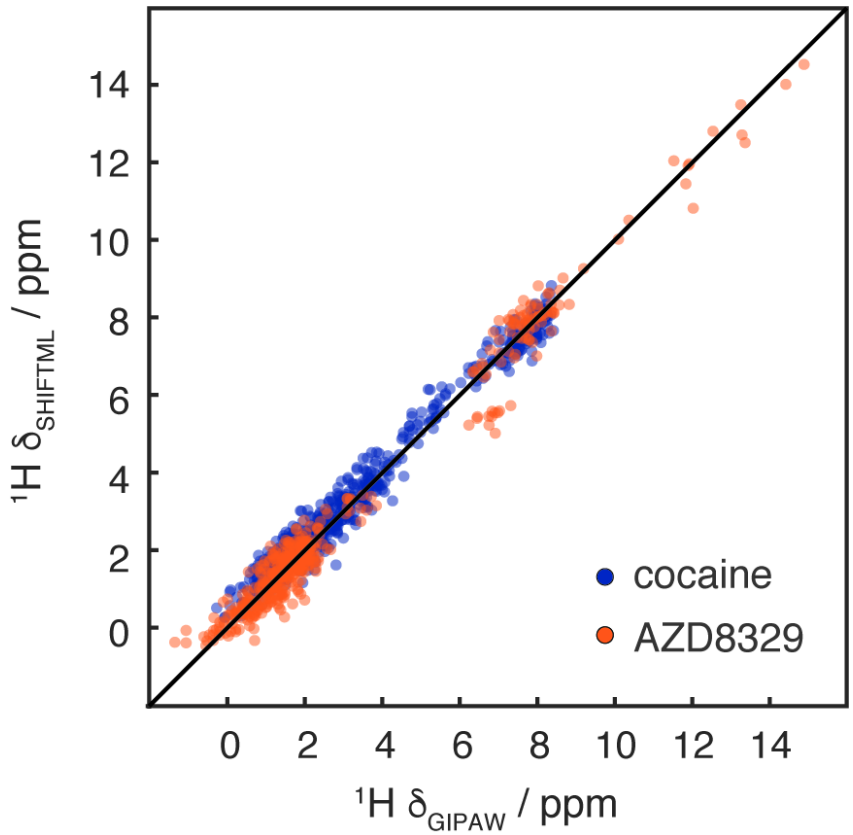


Paruzzo, Hofstetter, Musil, De, **MC**, Emsley, Nature Comm. (2018); http://shiftml.org

More than interatomic potentials

- Solid-state NMR relies on GIPAW-DFT to determine crystal structure of molecular materials
- Train a ML model on 2000 CSD structures, predict chemical shieldings with DFT accuracy (RMSE H: 0.5, C: 5, N: 13, 0: 18 ppm)
- Accurate enough to do structure determination!

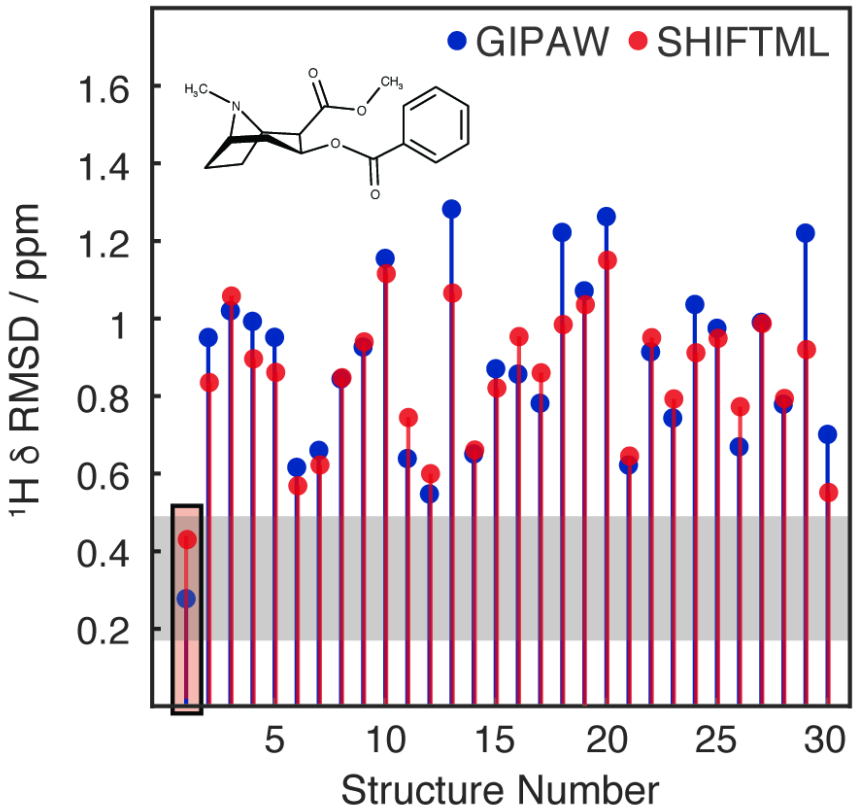


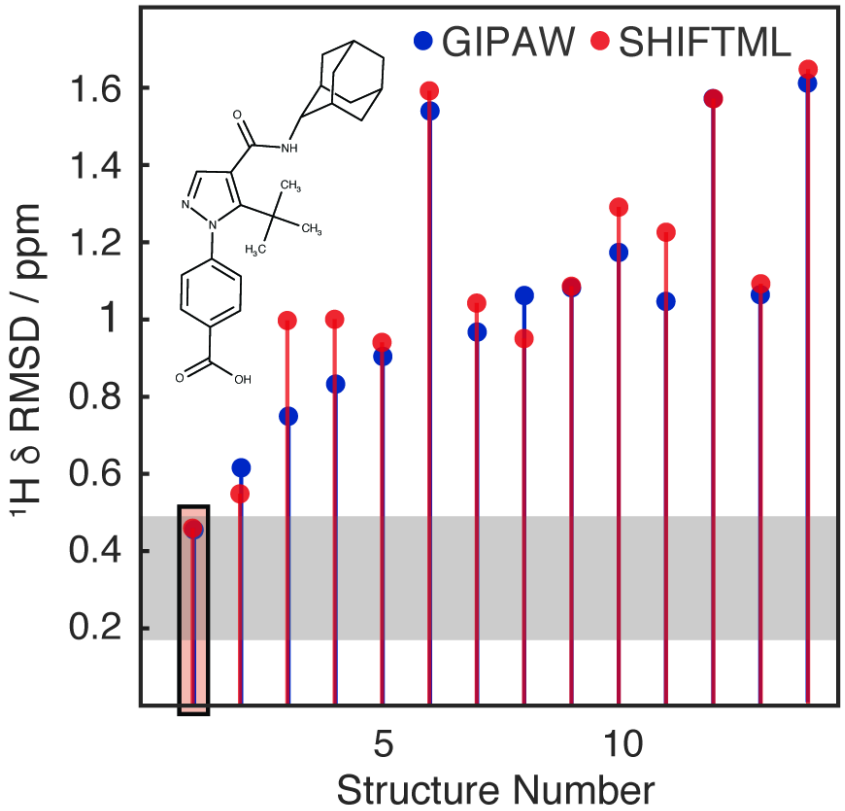


Paruzzo, Hofstetter, Musil, De, **MC**, Emsley, Nature Comm. (2018); http://shiftml.org

More than interatomic potentials

- Solid-state NMR relies on GIPAW-DFT to determine crystal structure of molecular materials
- Train a ML model on 2000 CSD structures, predict chemical shieldings with DFT accuracy (RMSE H: 0.5, C: 5, N: 13, 0: 18 ppm)
- Accurate enough to do structure determination!

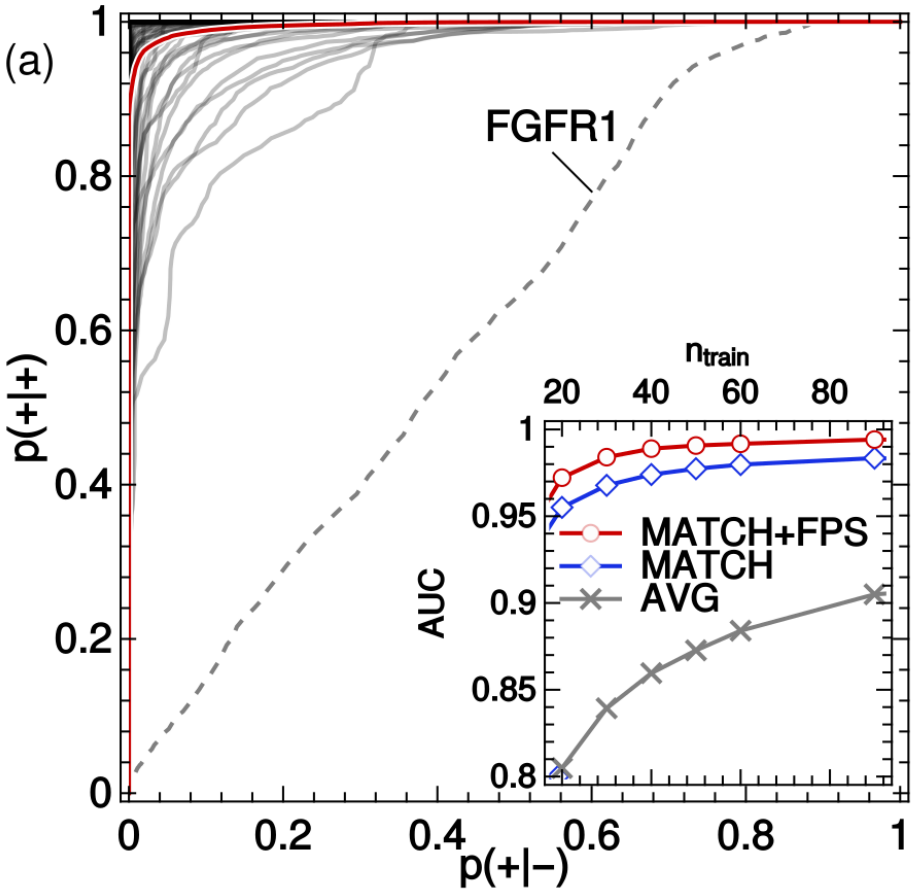


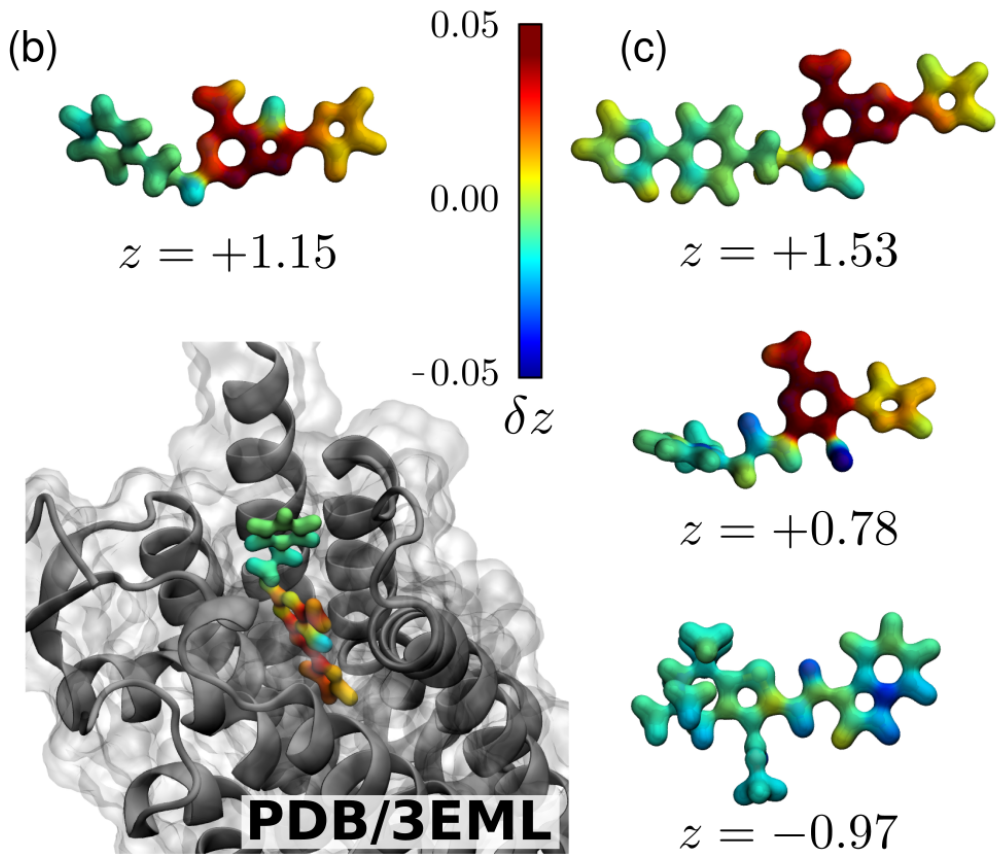


Paruzzo, Hofstetter, Musil, De, **MC**, Emsley, Nature Comm. (2018); http://shiftml.org

Recognizing active protein ligands

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





Bartok, De, Poelking, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

Questions?

Tensorial properties and beyond

Machine-learning for tensors

- In a Gaussian Process framework, the kernel represents correlations between properties. This must be reflected in how it transforms under symmetry operations applied to the inputs

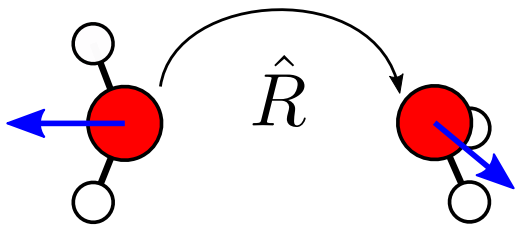
$$k\left(\mathcal{X},\mathcal{X}'\right) \leftrightarrow \left\langle y\left(\mathcal{X}\right); y\left(\mathcal{X}'\right)\right\rangle, \text{ so } k\left(\hat{S}\mathcal{X},\hat{S}'\mathcal{X}'\right) \leftrightarrow \left\langle y\left(\hat{S}\mathcal{X}\right); y\left(\hat{S}'\mathcal{X}'\right)\right\rangle$$

- Properties that are *invariant* under $\hat{S}$ must be learned with a kernel that should be insensitive to the operation

$$k\left(\hat{S}\mathcal{X},\hat{S}'\mathcal{X}'\right) = k\left(\mathcal{X},\mathcal{X}'\right)$$

- How about machine-learning tensorial properties **T**? The kernel should be *covariant* to rigid rotations - need a symmetry-adapted framework

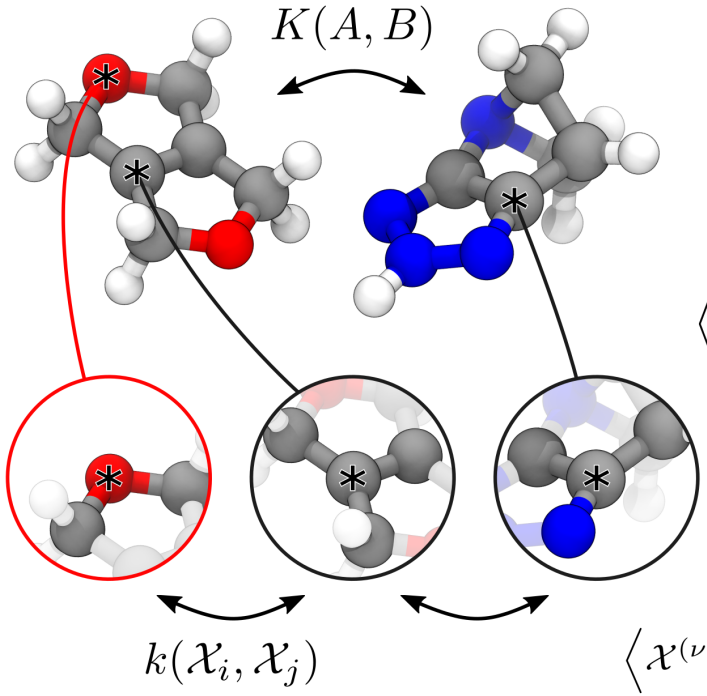
$$k_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right) \leftrightarrow \left\langle T_\mu\left(\mathcal{X}\right); T_\nu\left(\mathcal{X}'\right)\right\rangle \rightarrow k_{\mu\nu}\left(\hat{R}\mathcal{X},\hat{R}'\mathcal{X}'\right) = R_{\mu\mu'}k_{\mu'\nu'}\left(\mathcal{X},\mathcal{X}'\right)R'_{\nu\nu'}$$

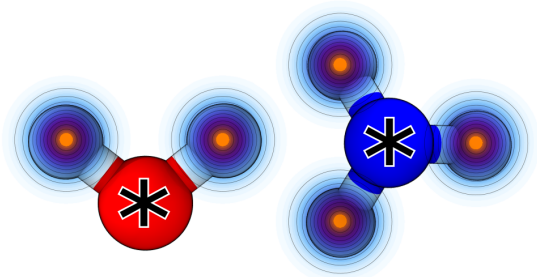


Glielmo, Sollich, & De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

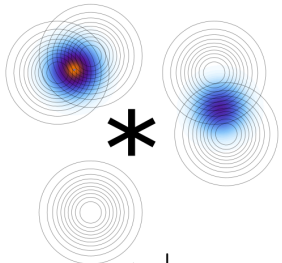
$\lambda-$SOAP: a $SO\,(3)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\,(3)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities





$$\langle \mathbf{r}|\mathcal{X}_j\rangle = \psi(\mathbf{r}) = \sum_i g(\mathbf{r}-\mathbf{r}_{ij})$$



$$\left\langle \mathcal{X}^{(\nu)}\middle|\mathcal{X}'^{(\nu)}\right\rangle_{\hat{R}} \sim \int \mathrm{d}\hat{R}\left|\int \psi(\mathbf{r})\psi'(\hat{R}\mathbf{r})\mathrm{d}\mathbf{r}\right|^{\nu}$$

$$k\left(\mathcal{X},\mathcal{X}'\right) = \int \mathrm{d}\hat{R}\,\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right), \qquad \kappa\left(\mathcal{X},\mathcal{X}'\right) = \left|\int \psi_{\mathcal{X}}\left(\boldsymbol{x}\right)\psi_{\mathcal{X}'}\left(\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x}\right|^{2}$$

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

$\lambda-$SOAP: a $SO\,(3)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\,(3)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

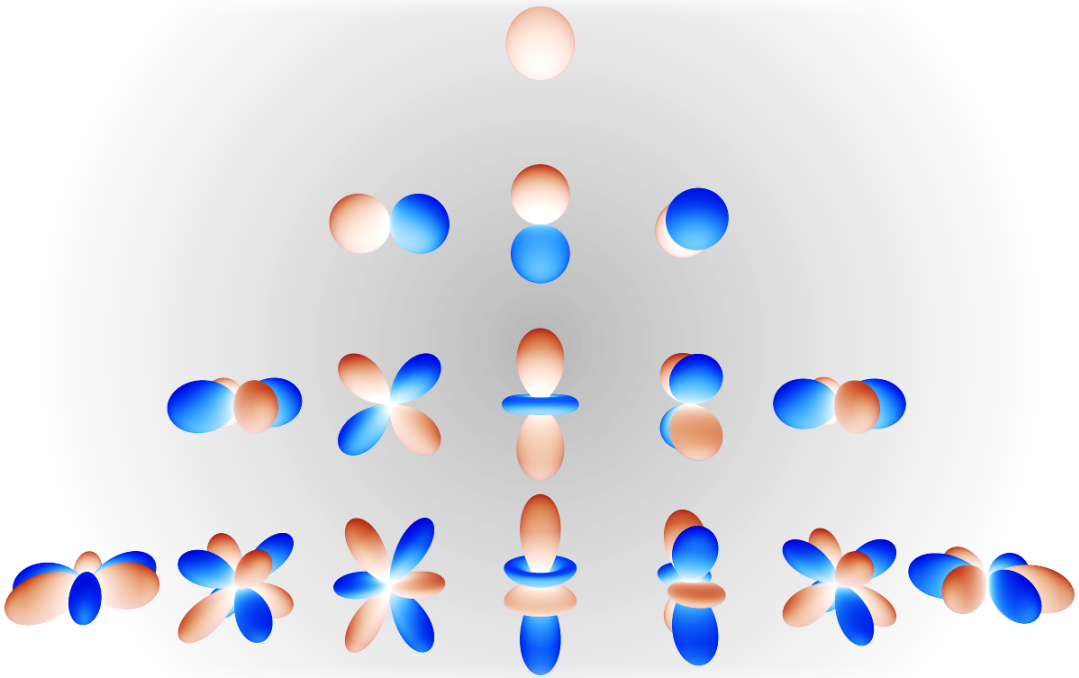


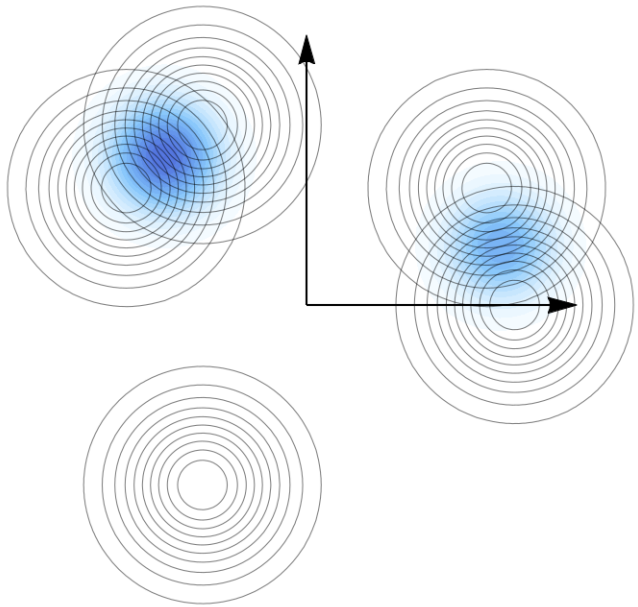
image fro Wikipedia

$$T^{\lambda}_{\mu}\left(\hat{R}\,(\mathcal{X})\right) = D^{\lambda}_{\mu\mu'}\left(\hat{R}\right) T^{\lambda}_{\mu'}\,(\mathcal{X})$$

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

$\lambda-$SOAP: a $SO\,(3)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\,(3)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

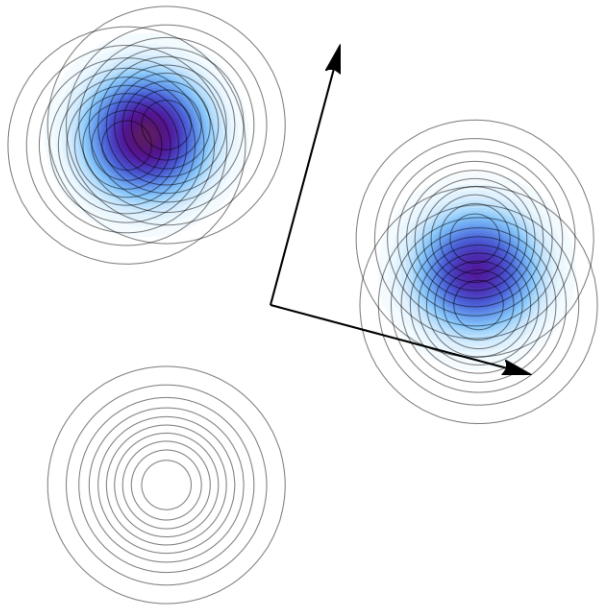


$$k^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{R}\, D^{\lambda}_{\mu\nu}\left(\hat{R}\right)\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right)$$

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

$\lambda-$SOAP: a $SO\left(3\right)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\left(3\right)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

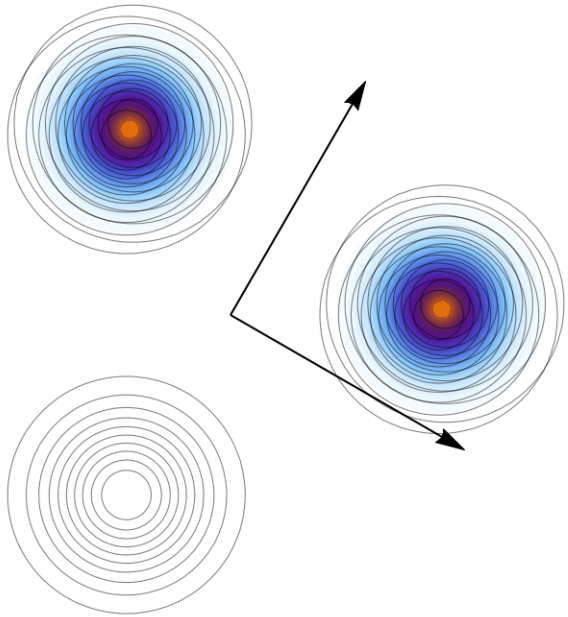


$$k^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{R} D^{\lambda}_{\mu\nu}\left(\hat{R}\right)\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right)$$

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

$\lambda-$SOAP: a $SO\,(3)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\,(3)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

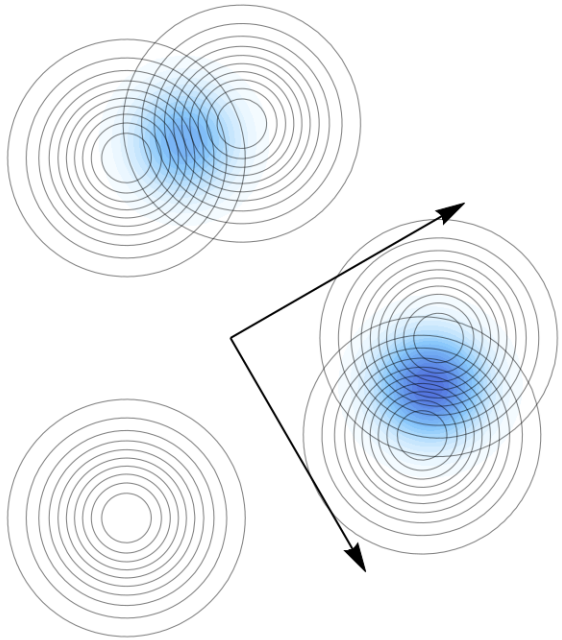


$$k^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{R}\, D^{\lambda}_{\mu\nu}\left(\hat{R}\right)\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right)$$

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

$\lambda-$SOAP: a $SO\,(3)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\,(3)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

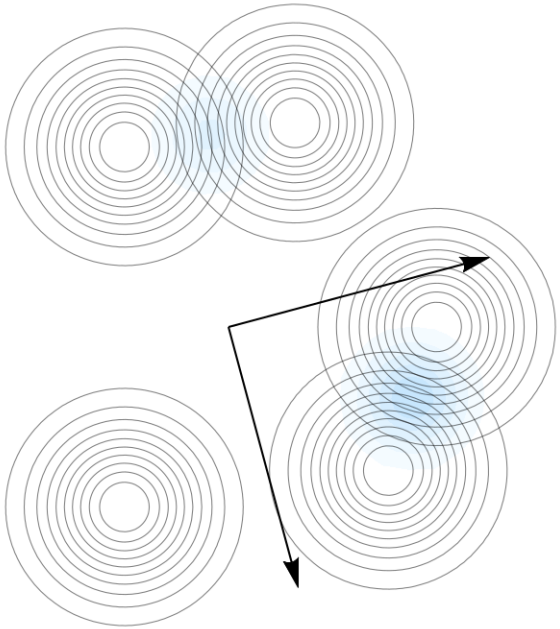


$$k^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{R}\, D^{\lambda}_{\mu\nu}\left(\hat{R}\right)\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right)$$

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

$\lambda-$SOAP: a $SO\left(3\right)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\left(3\right)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

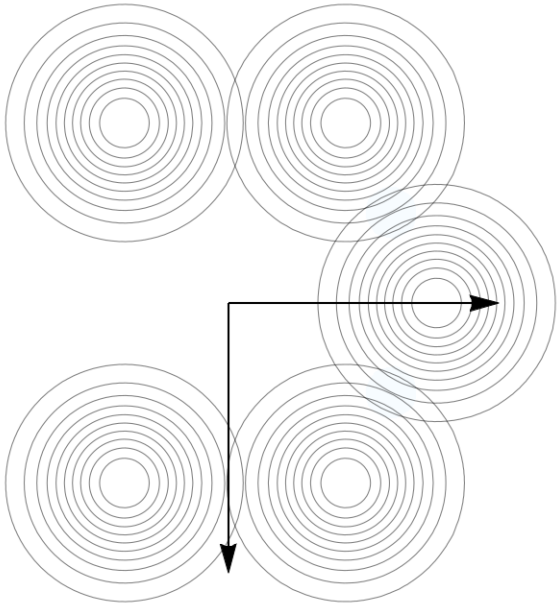


$$k^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{R} D^{\lambda}_{\mu\nu}\left(\hat{R}\right)\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right)$$

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

$\lambda-$SOAP: a $SO\left(3\right)$ compliant kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components $\mathbf{T}^{\lambda}$, corresponding to the representations of $SO\left(3\right)$
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities



$$k^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{R}\, D^{\lambda}_{\mu\nu}\left(\hat{R}\right)\kappa\left(\mathcal{X},\hat{R}\mathcal{X}'\right)$$

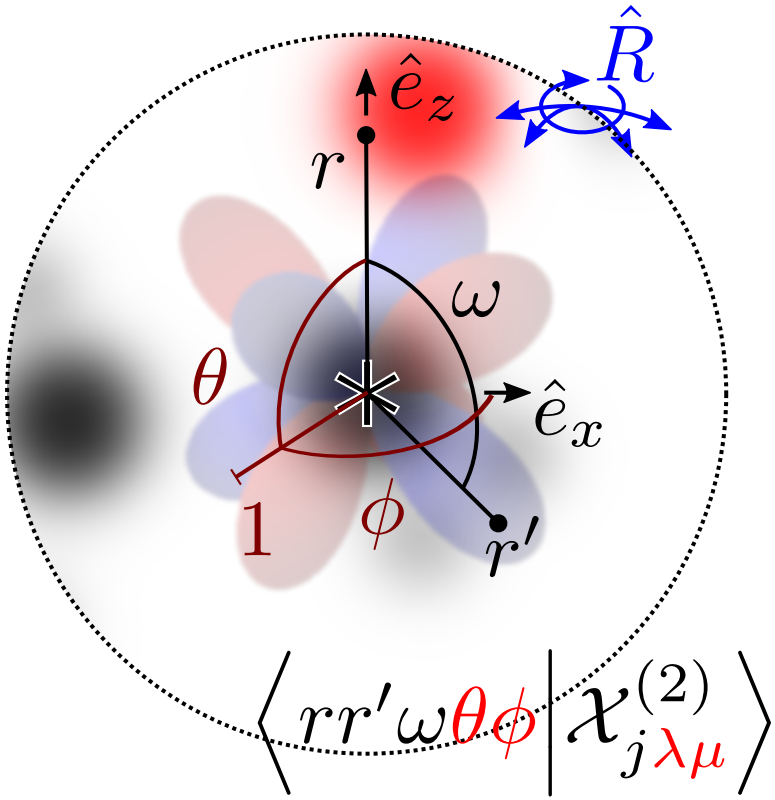
Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

$\lambda-$SOAP as a descriptor

- For the mathematically-inclined: we can see this as an extension of the density-representation framework

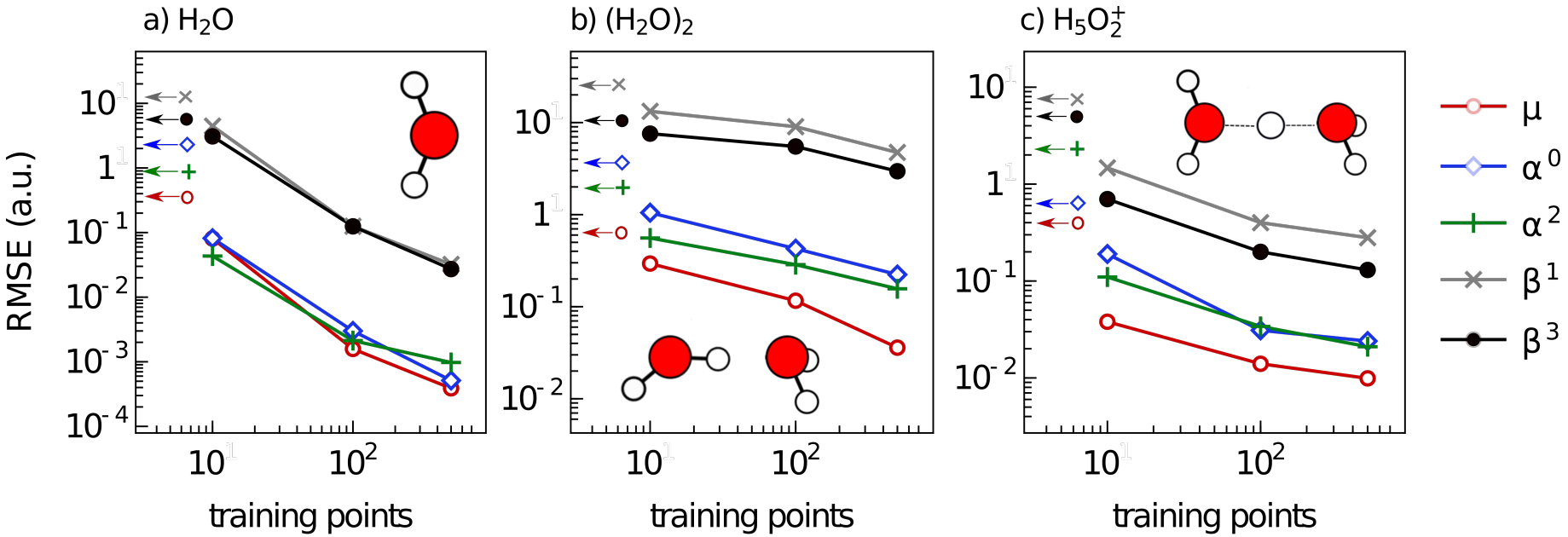
$$\int \mathrm{d}\hat{R} \langle \mathbf{r}| \hat{R} \left|\mathcal{X}_j\right\rangle \langle \mathbf{r}'| \hat{R} \left|\mathcal{X}_j\right\rangle \langle \mathbf{r}''| \hat{R} |\lambda\mu\rangle \rightarrow \left\langle rr'\omega\,\theta\phi \middle| \mathcal{X}^{(2)}_{j\lambda\mu} \right\rangle$$

- Easier to compute by expanding the density in $R_n(r)\, Y^l_m(\hat{r})$, leading explicit power-spectrum-like representation $\left\langle nn'll' \middle| \mathcal{X}^{(2)}_{j\lambda\mu} \right\rangle$



Machine-Learning the Dielectric Responses

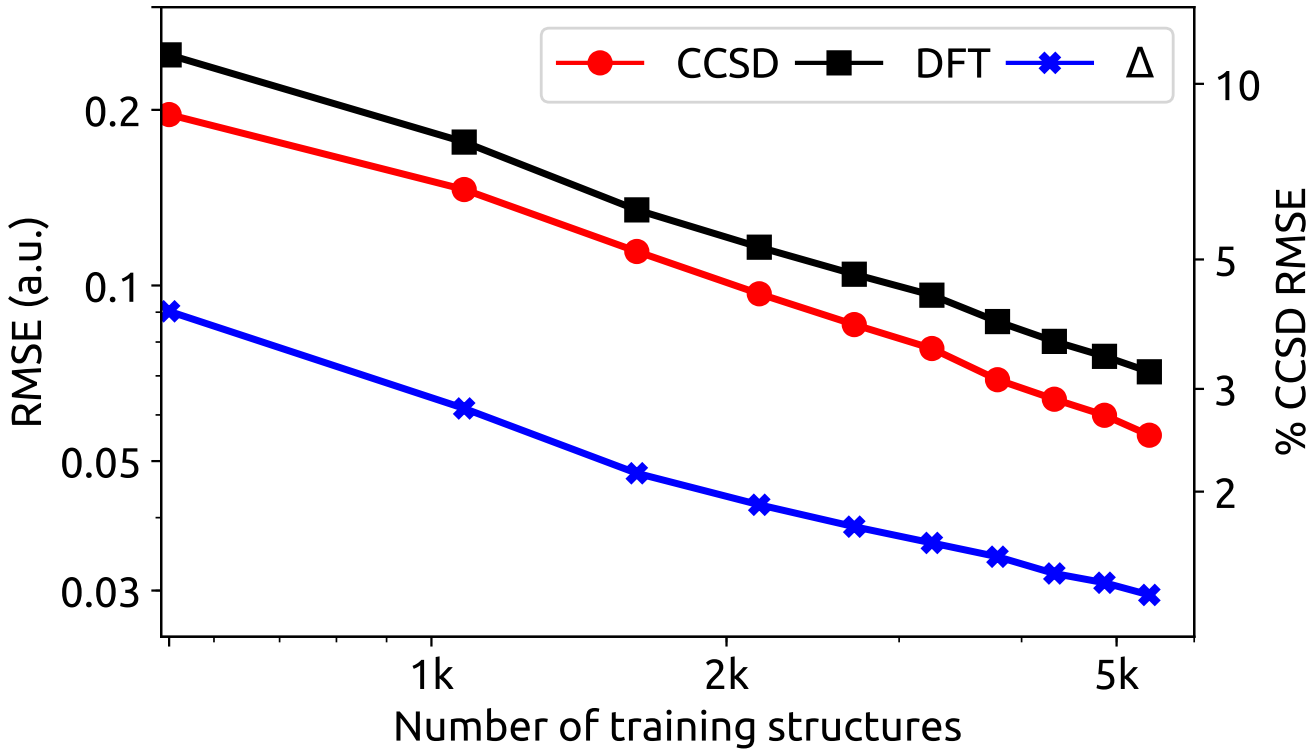
- A demonstration of the SA-GPR framework, and the λ-SOAP kernel - learning the dielectric response of water oligomers
- The kernels for multi-atomic systems can be built with an additive ansatz - and that gives meaningful partitioning in atomic/molecular contributions



$$k_{\mu\nu}\left(\mathcal{A},\mathcal{B}\right)=\frac{1}{N_{\mathcal{A}}N_{\mathcal{B}}}\sum_{ij}k_{\mu\nu}\left(\mathcal{X}_i^{\mathcal{A}},\mathcal{X}_j^{\mathcal{B}}\right)$$

Molecular polarizabilities at the CCSD level

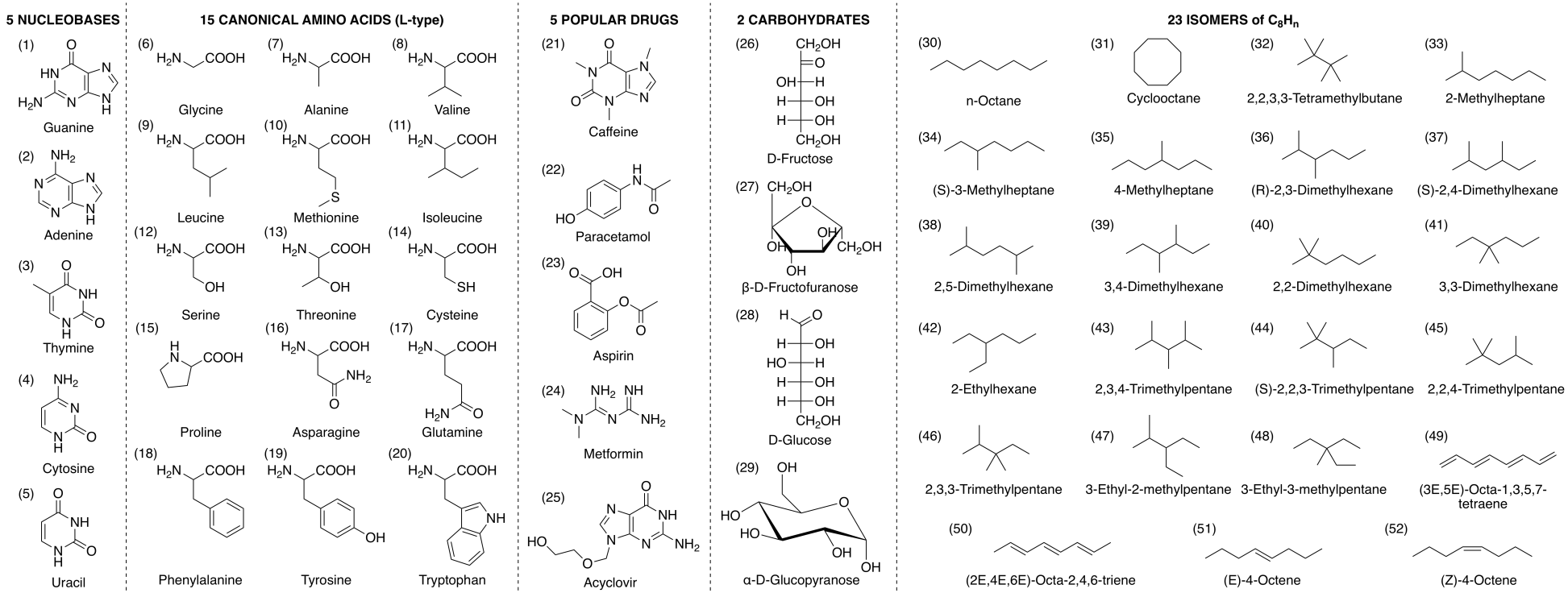
- DFT is not very accurate for the dielectric response. Train a ML model (AlphaML) on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir $C_8H_{11}N_5O_3$) with better-than-DFT accuracy
- Atom-centered environment decomposition of α and the DFT error



Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, arxiv.org/abs/1809.05349; `http://alphaml.org`

Molecular polarizabilities at the CCSD level

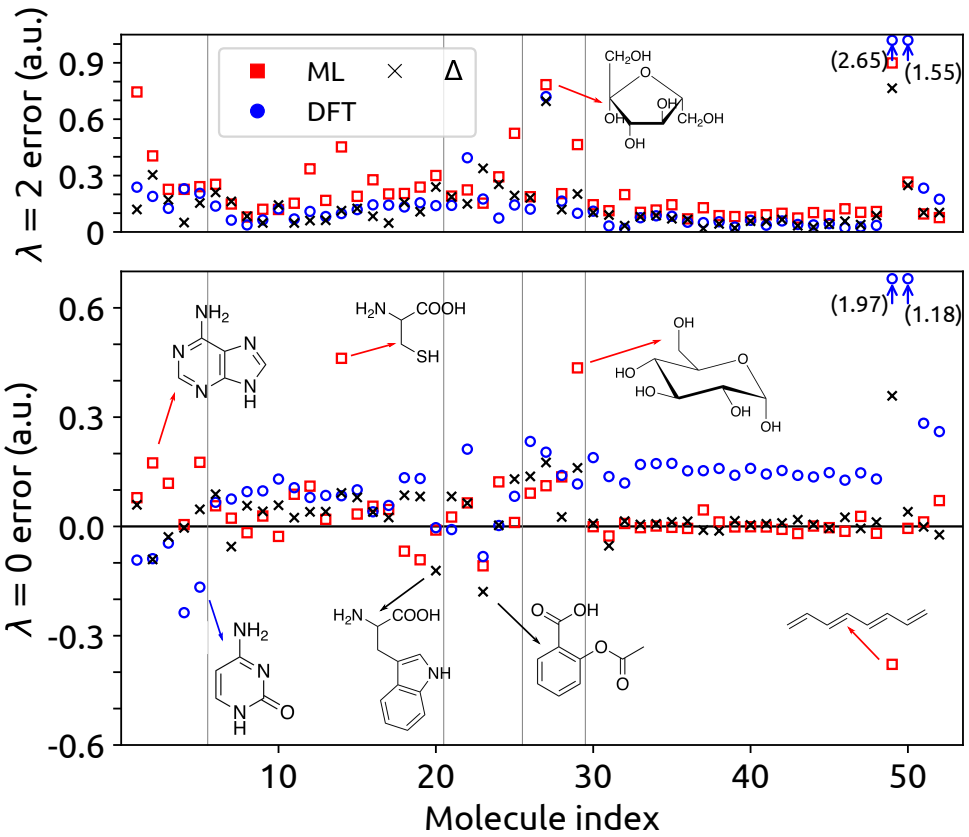
- DFT is not very accurate for the dielectric response. Train a ML model (AlphaML) on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir $C_8H_{11}N_5O_3$) with better-than-DFT accuracy
- Atom-centered environment decomposition of α and the DFT error



Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, arxiv.org/abs/1809.05349; `http://alphaml.org`

Molecular polarizabilities at the CCSD level

- DFT is not very accurate for the dielectric response. Train a ML model (AlphaML) on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir $C_8H_{11}N_5O_3$) with better-than-DFT accuracy
- Atom-centered environment decomposition of α and the DFT error

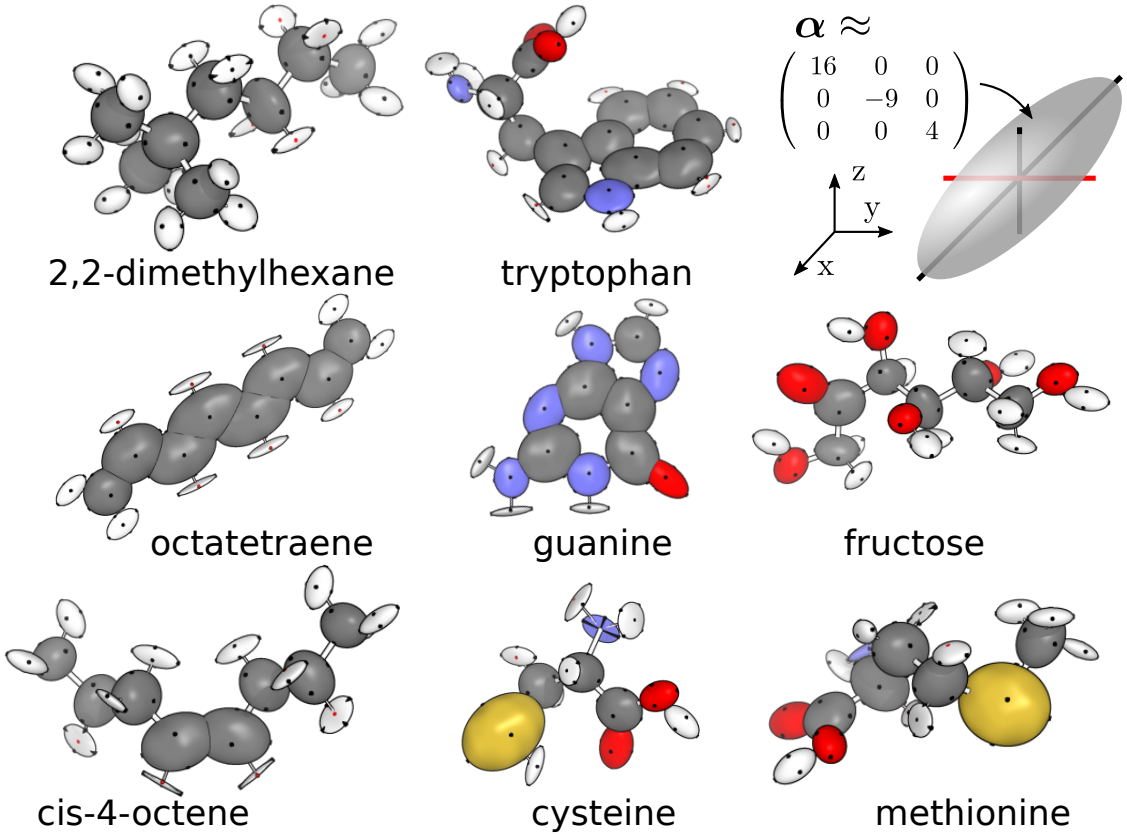


Method	RMSE
CCSD/DFT	0.573
CCSD/ML	0.304
DFT/ML	0.403
Δ(CCSD-DFT)/ML	0.212

Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, arxiv.org/abs/1809.05349; `http://alphaml.org`

Molecular polarizabilities at the CCSD level

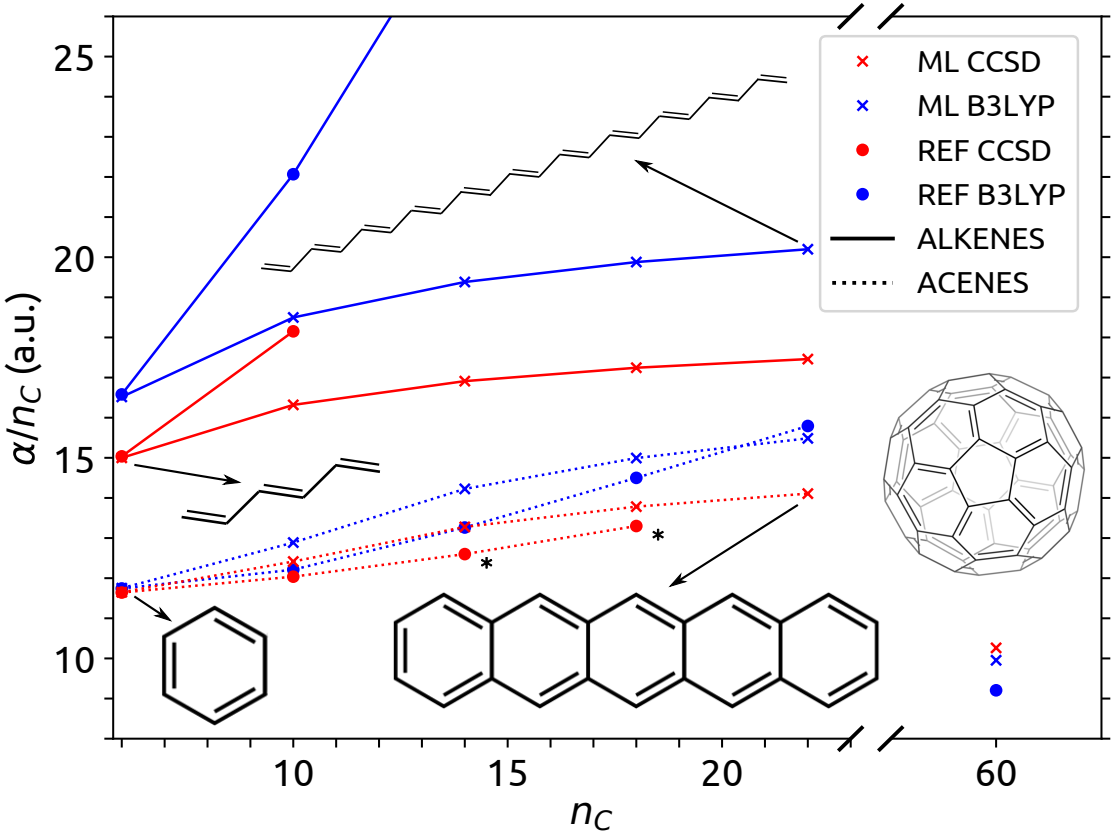
- DFT is not very accurate for the dielectric response. Train a ML model (AlphaML) on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir $C_8H_{11}N_5O_3$) with better-than-DFT accuracy
- Atom-centered environment decomposition of α and the DFT error



Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, arxiv.org/abs/1809.05349; `http://alphaml.org`

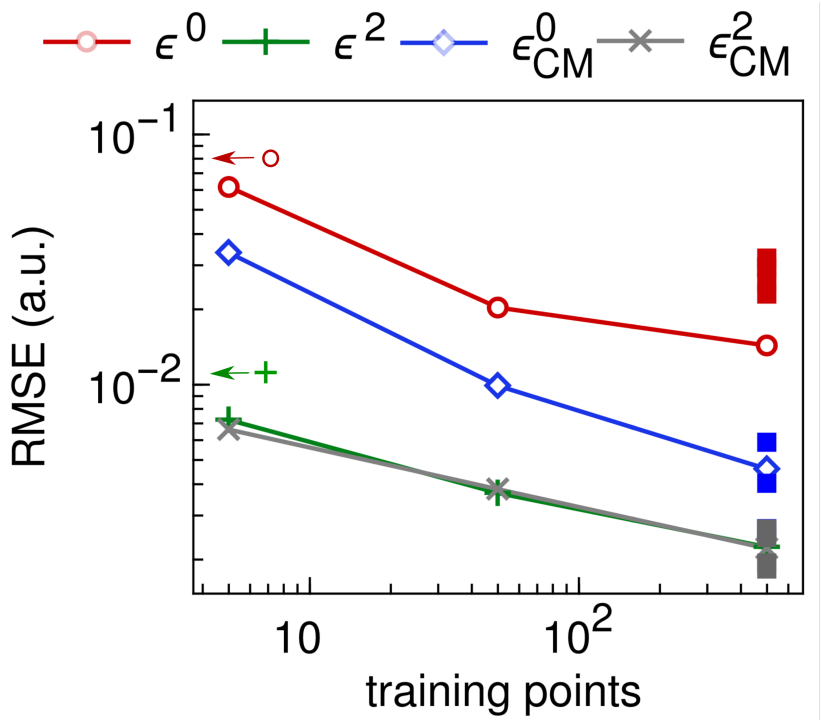
Locality: Curse or Blessing?

- Extended conjugated systems are a challenge for both ML and electronic-structure methods (diverging polarizability, need multi-reference methods)
- AlphaML is local and additive, so it saturates to the core of the largest structure in training. Ends up being much less insane than DFT, but clearly points at the challenge of non-local physics.



Learning the dielectric response of water

- The SA-GPR framework, and the λ-SOAP kernel, works as well for bulk systems
- The dielectric constant involves non-additive effects. ML improves dramatically by learning a proxy that is approximately additive



Clausius-Mossotti: $\alpha = (\varepsilon - 1)(\varepsilon + 2)^{-1} V$

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

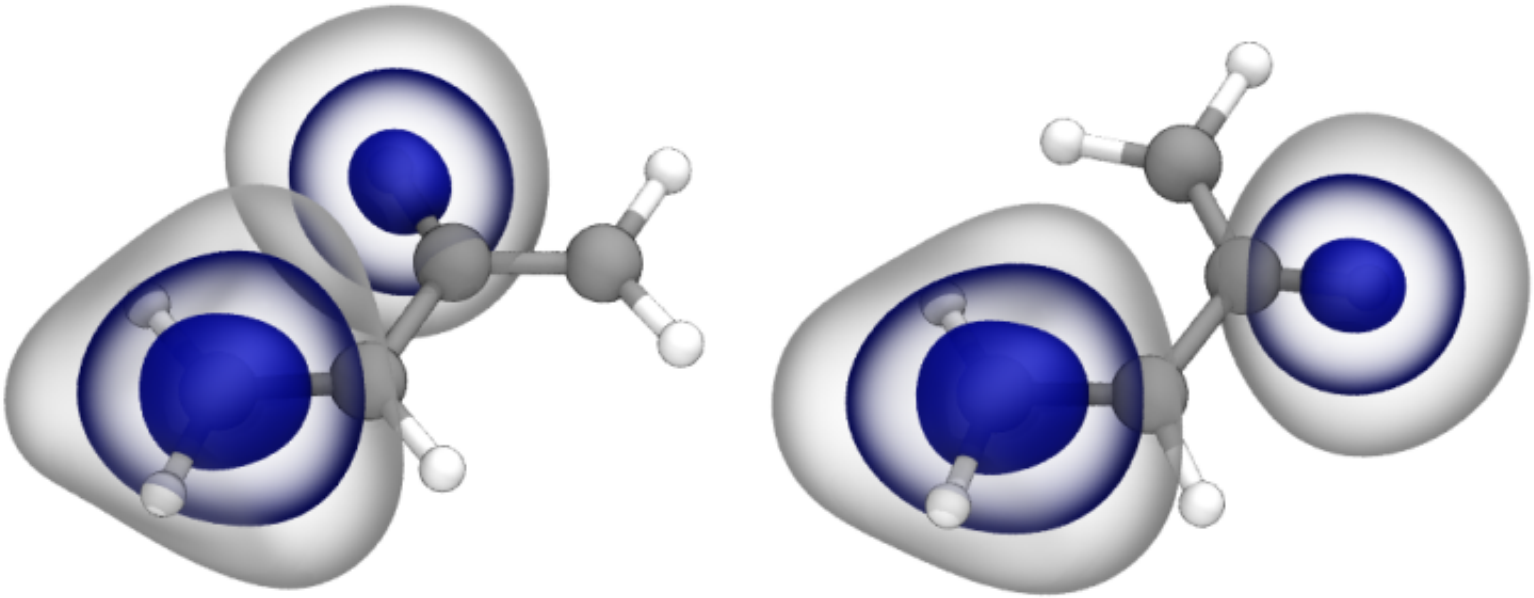
A transferable model of the electron density

- Write the density in atom-centered terms. Use a $\phi_k \equiv R_n Y_m^l$ expansion.

$$\mathcal{F}(\rho) = \int d\mathbf{r} \left| \rho(\mathbf{r}) - \sum_{ik} c_{ik} \phi_k (\mathbf{r} - \mathbf{r}_i) \right|^2 + \eta |\mathbf{x}|^2, \quad c_{inlm} = \sum_{jm'} x_{jnlm} k^l_{mm'} (\mathcal{X}_i, \mathcal{X}_j)$$

- Machine-learn directly the full density
  - Avoid the non-uniqueness of atoms-in-molecules decompositions
  - Tricky due to non-orthogonality: $\mathbf{x}$ coefficients of different orbitals and atoms are coupled by

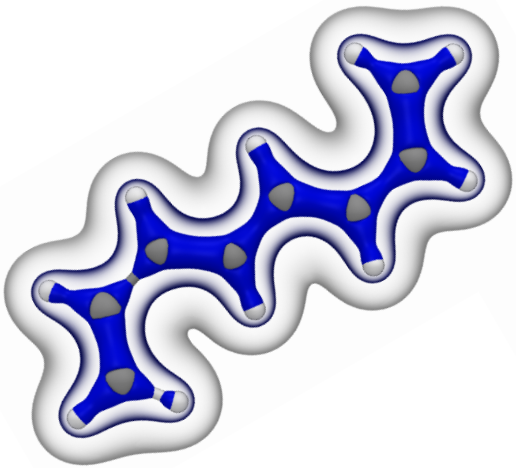
$$\left\langle \phi_k (\mathbf{r} - \mathbf{r}_i) \, \phi_{k'} \left(\mathbf{r} - \mathbf{r}'_i \right) \right\rangle$$



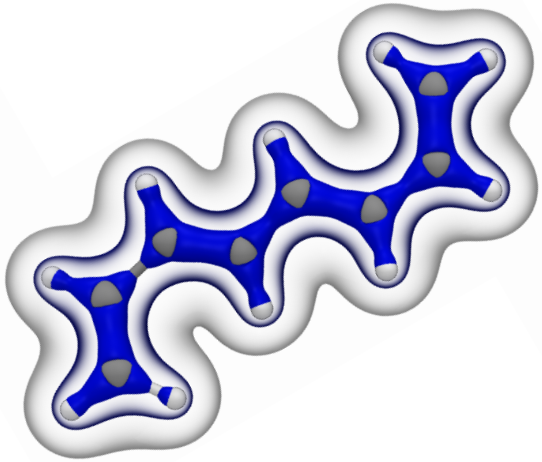
A transferable model of the electron density

- Very efficient learning, but limited by the basis set accuracy
- Extremely transferable: learn on C4 molecules, predict on C8
- Needs more work on optimizing the basis set
- Somewhat disappointing accuracy on energetics. Better to learn directly?

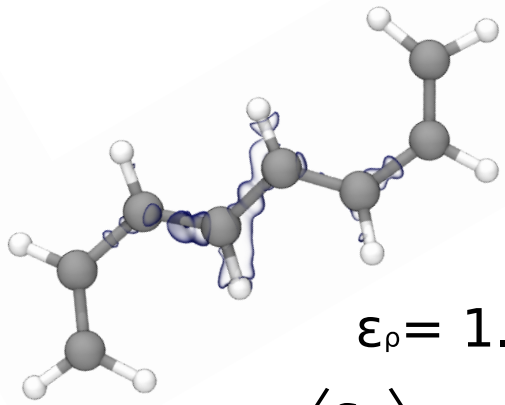
QM



ML



ML − QM



$\varepsilon_\rho = 1.81\%$

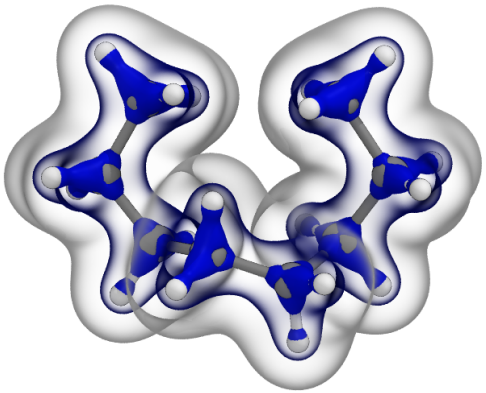
$\langle \varepsilon_\rho \rangle = 1.83\%$

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

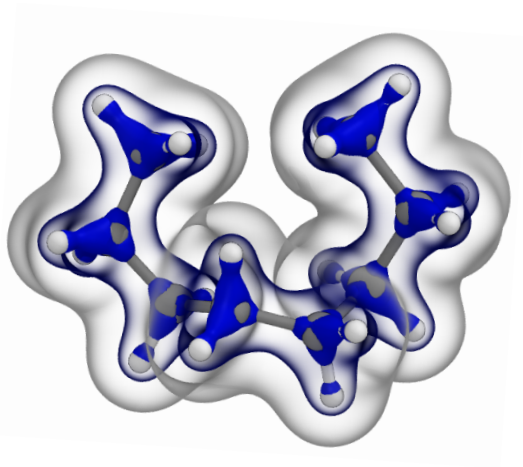
A transferable model of the electron density

- Very efficient learning, but limited by the basis set accuracy
- Extremely transferable: learn on C4 molecules, predict on C8
- Needs more work on optimizing the basis set
- Somewhat disappointing accuracy on energetics. Better to learn directly?

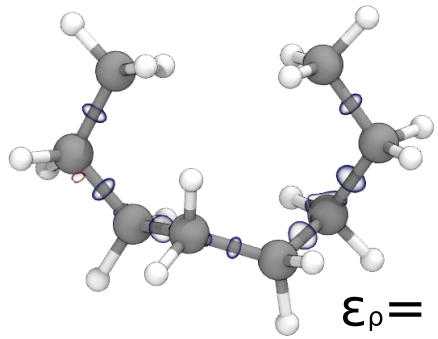
QM



ML



ML − QM



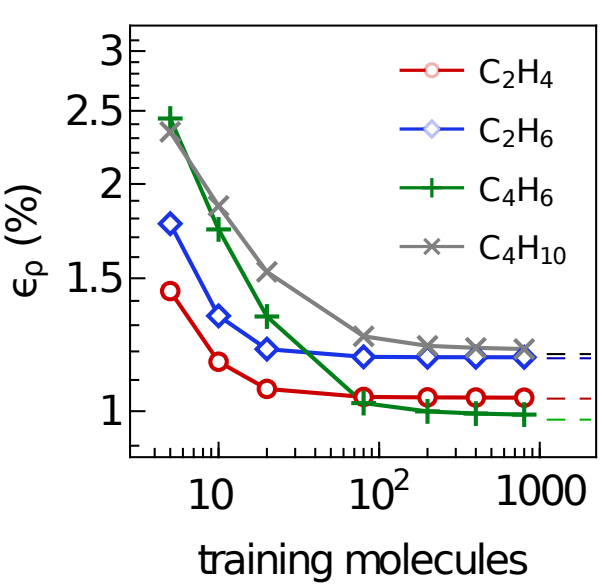
$\varepsilon_\rho = 1.41\%$

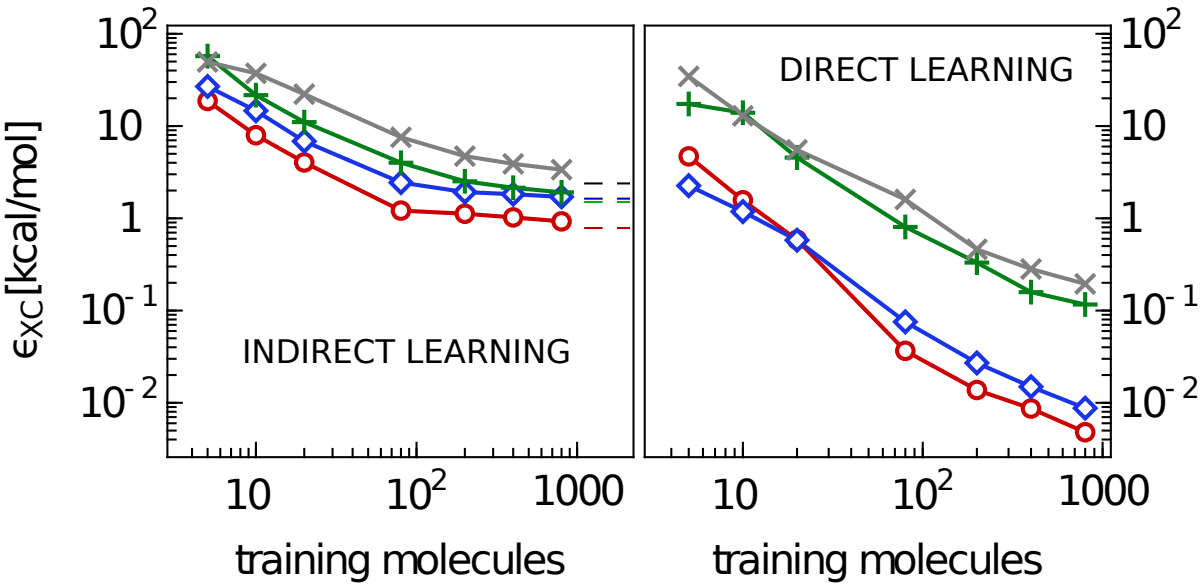
$\langle \varepsilon_\rho \rangle = 1.40\%$

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

A transferable model of the electron density

- Very efficient learning, but limited by the basis set accuracy
- Extremely transferable: learn on C4 molecules, predict on C8
- Needs more work on optimizing the basis set
- Somewhat disappointing accuracy on energetics. Better to learn directly?



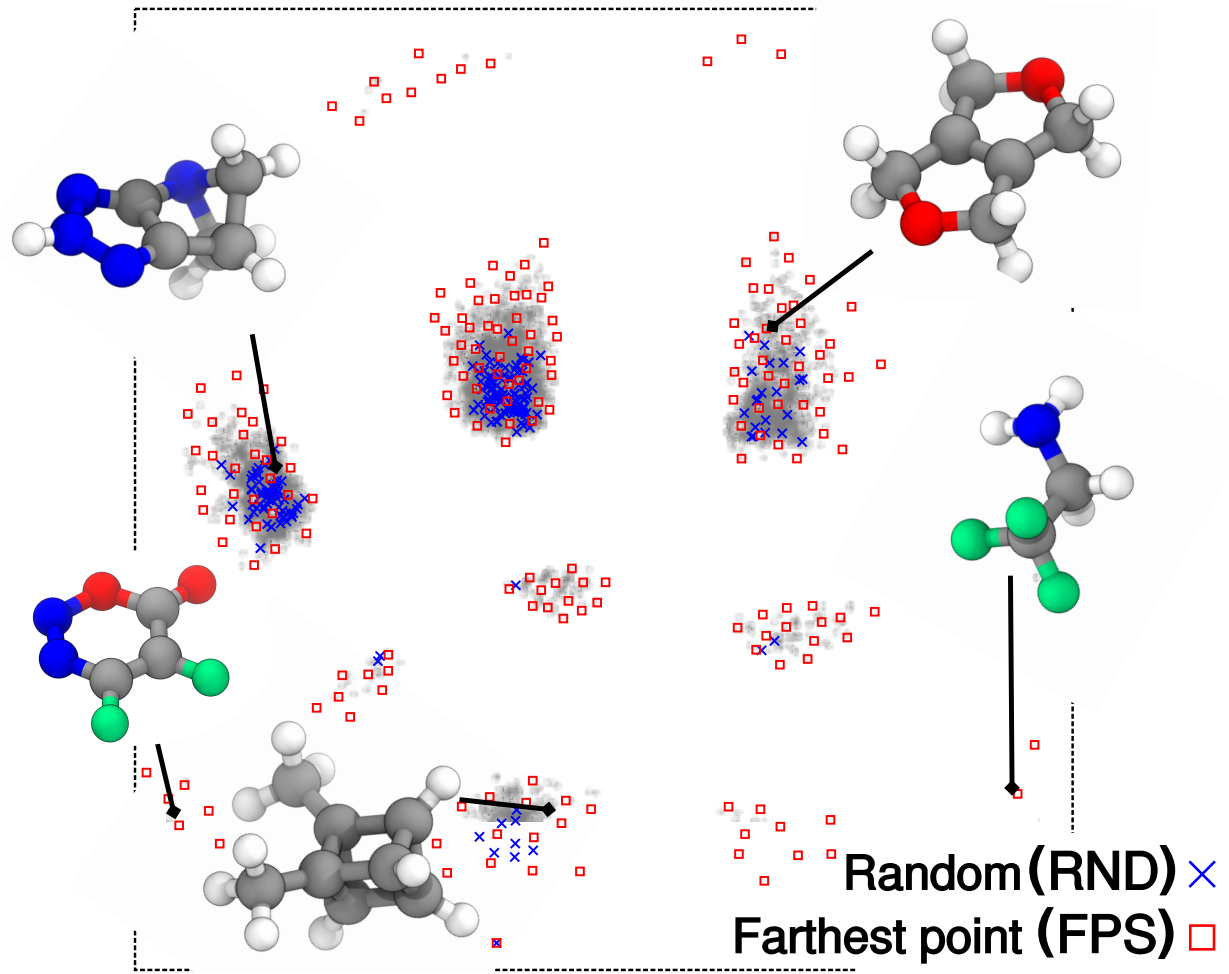


Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, ACS Central Science (2019)

Accuracy, efficiency and errors

Train set optimization to reduce errors

- The train set should cover uniformly the relevant space
  - Farthest point sampling is a simple, constructive strategy to optimize the training set, opening doors to active learning

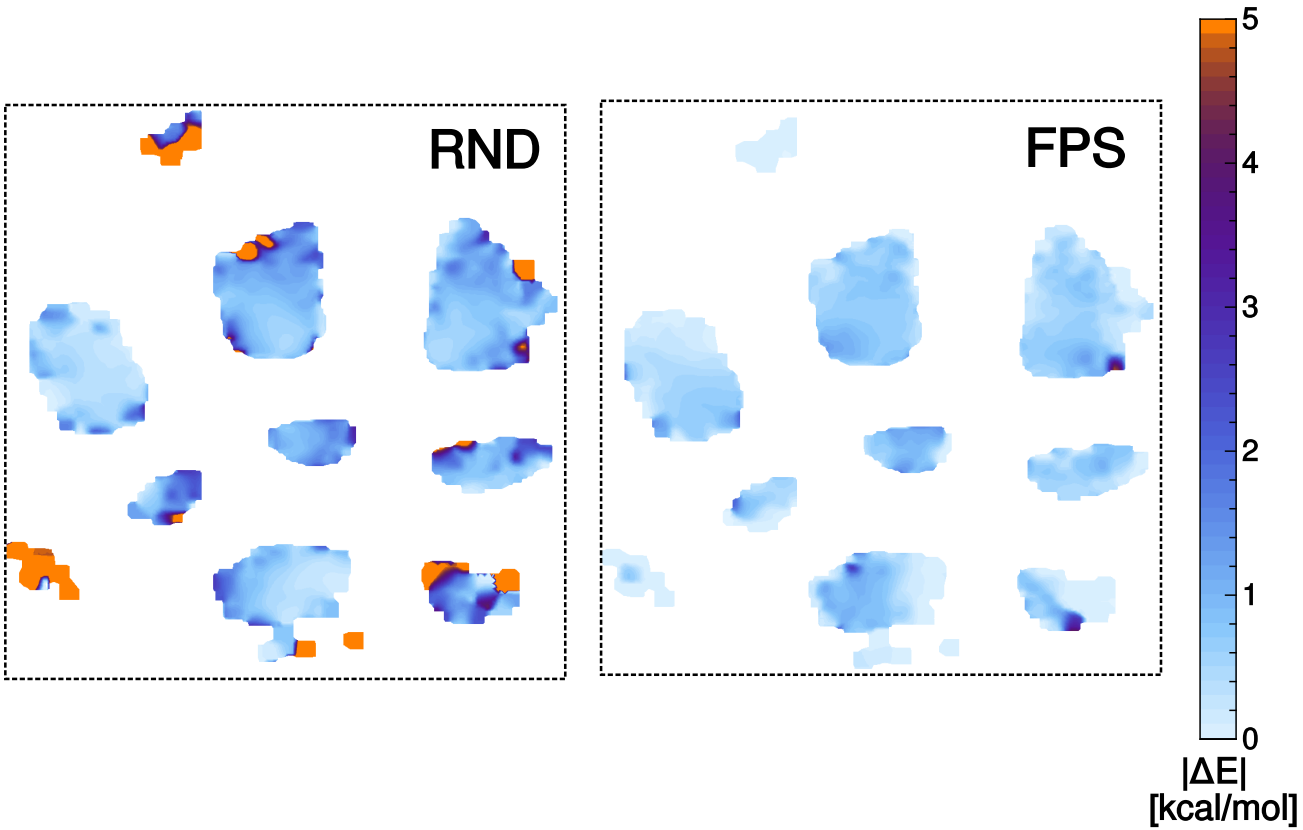


Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

MC, Tribello, Parrinello, PNAS (2011); http://sketchmap.org

Train set optimization to reduce errors

- The train set should cover uniformly the relevant space
  - Farthest point sampling is a simple, constructive strategy to optimize the training set, opening doors to active learning



Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

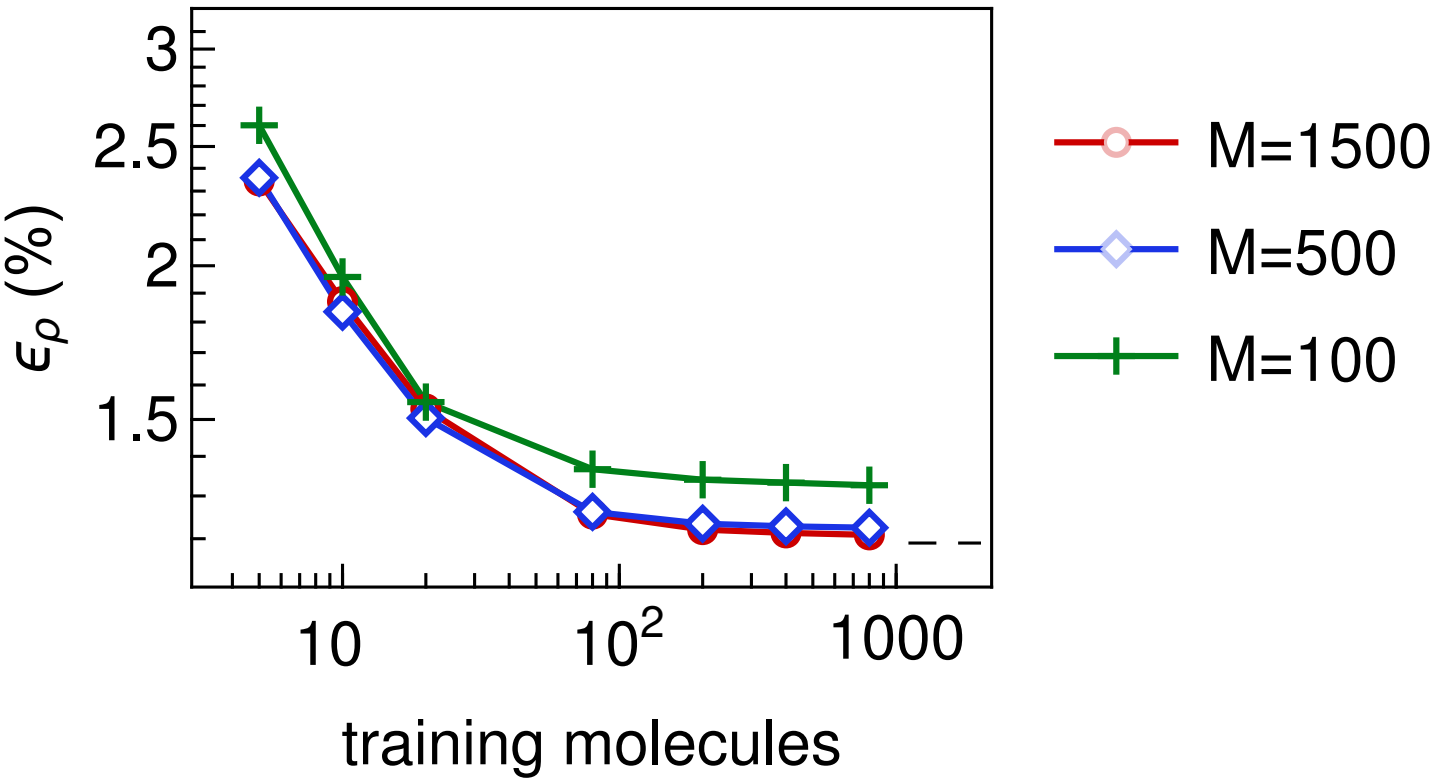
MC, Tribello, Parrinello, PNAS (2011); http://sketchmap.org

Sparse kernel training

- "But the cost of kernel methods grow with train set size!" . . . does it?
- Sparse kernel learning is actually easy and effective ($M \ll N$)

$$y(\mathcal{X}) = \sum_M x_M k(\mathcal{X}, \mathcal{X}_M), \qquad L^2 = \sum_N |y_N - y(\mathcal{X}_N)|^2 + \lambda |\mathbf{x}|^2$$

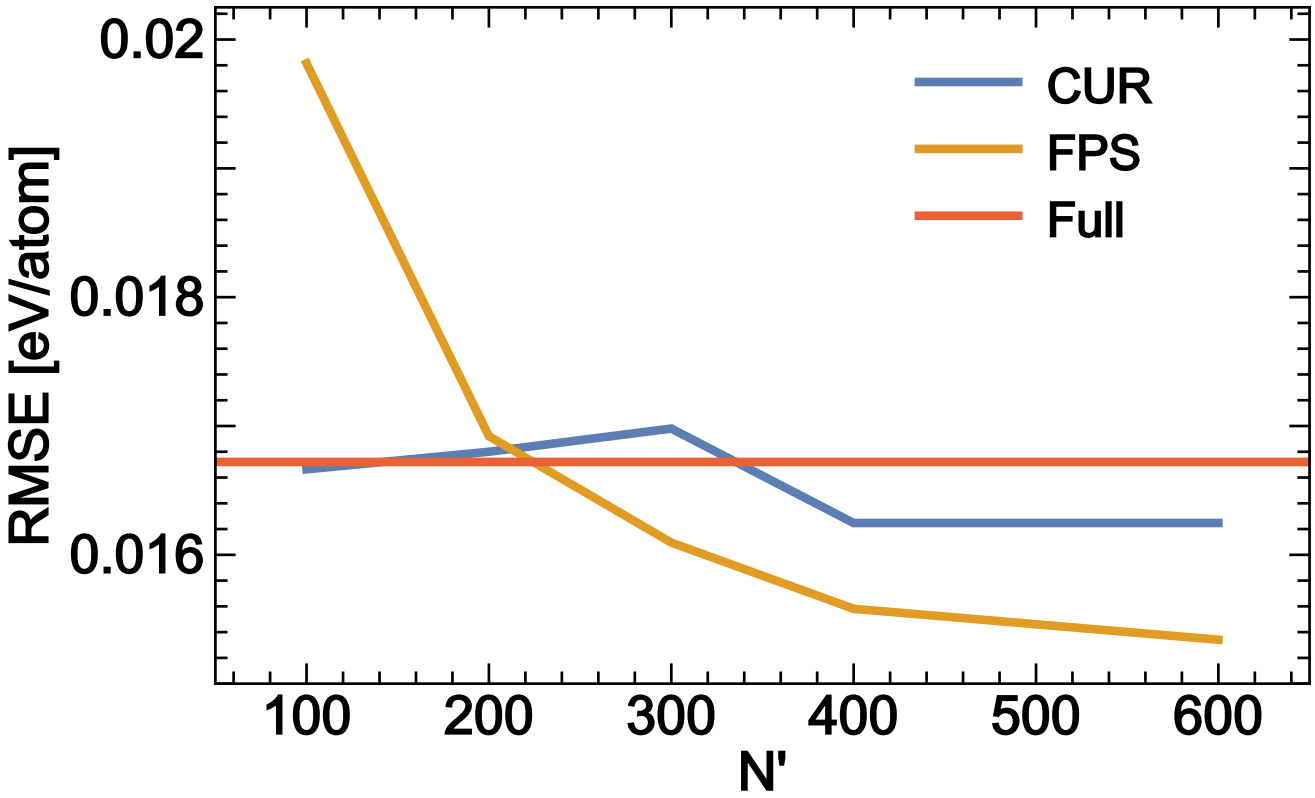
- Learning charge density: we can keep 100 environments out of 10000



"Machine Learning of Atomic-Scale Properties Based on Physical Principles", in *Handbook of Materials Modeling* (2018);

Sparse representation for data efficiency

- Symmetry-functions are hard to choose
- Systematic expansions à la SOAP are huge and expensive
- Solution: automatic feature selection based on CUR or FPS idea applied to representation space



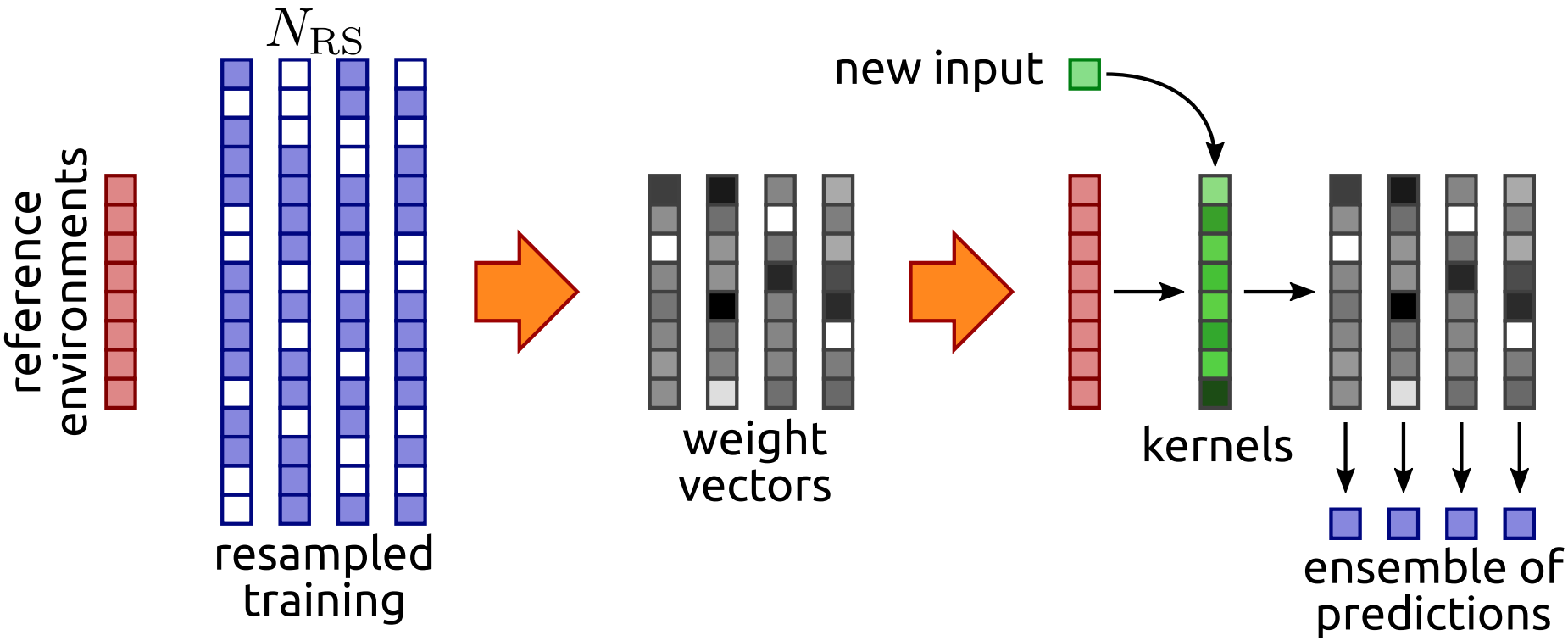
Imbalzano, Anelli, Giofré, Klees, Behler, **MC**, JCP (2018)

An accurate & inexpensive error estimation

- Generate an ensemble of GPR models, and use distribution of predictions

$$y(\mathcal{X}) = \frac{1}{N_{RS}} \sum_i y^{(i)}(\mathcal{X}), \qquad \sigma^2(\mathcal{X}) = \frac{1}{N_{RS}-1} \sum_i \left(y^{(i)}(\mathcal{X}) - y(\mathcal{X}) \right)^2$$

  - Verify accuracy by the distribution of errors $P\left(|y(\mathcal{X}) - y_{\text{ref}}(\mathcal{X})| \;\; |\sigma(\mathcal{X})\right)$
  - Use maximum-likelihood to calibrate the uncertainty $\sigma(\mathcal{X}) \rightarrow \alpha\sigma(\mathcal{X})^{\gamma-1}$



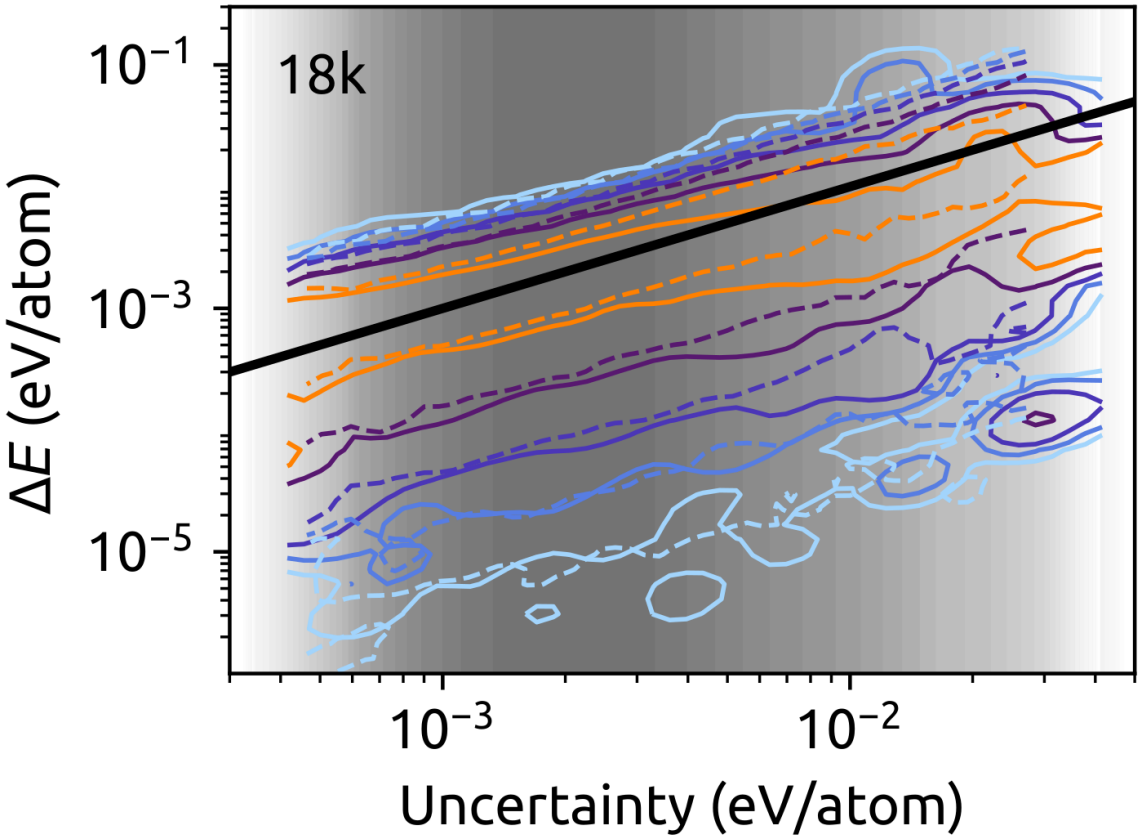
Musil, Willatt, **MC** JCTC (2019)

An accurate & inexpensive error estimation

- Generate an ensemble of GPR models, and use distribution of predictions

$$y(\mathcal{X}) = \frac{1}{N_{RS}} \sum_i y^{(i)}(\mathcal{X}), \qquad \sigma^2(\mathcal{X}) = \frac{1}{N_{RS}-1} \sum_i \left(y^{(i)}(\mathcal{X}) - y(\mathcal{X}) \right)^2$$

  - Verify accuracy by the distribution of errors $P\left(|y(\mathcal{X}) - y_{\text{ref}}(\mathcal{X})| \;\middle|\; \sigma(\mathcal{X})\right)$
  - Use maximum-likelihood to calibrate the uncertainty $\sigma(\mathcal{X}) \to \alpha\sigma(\mathcal{X})^{\gamma-1}$



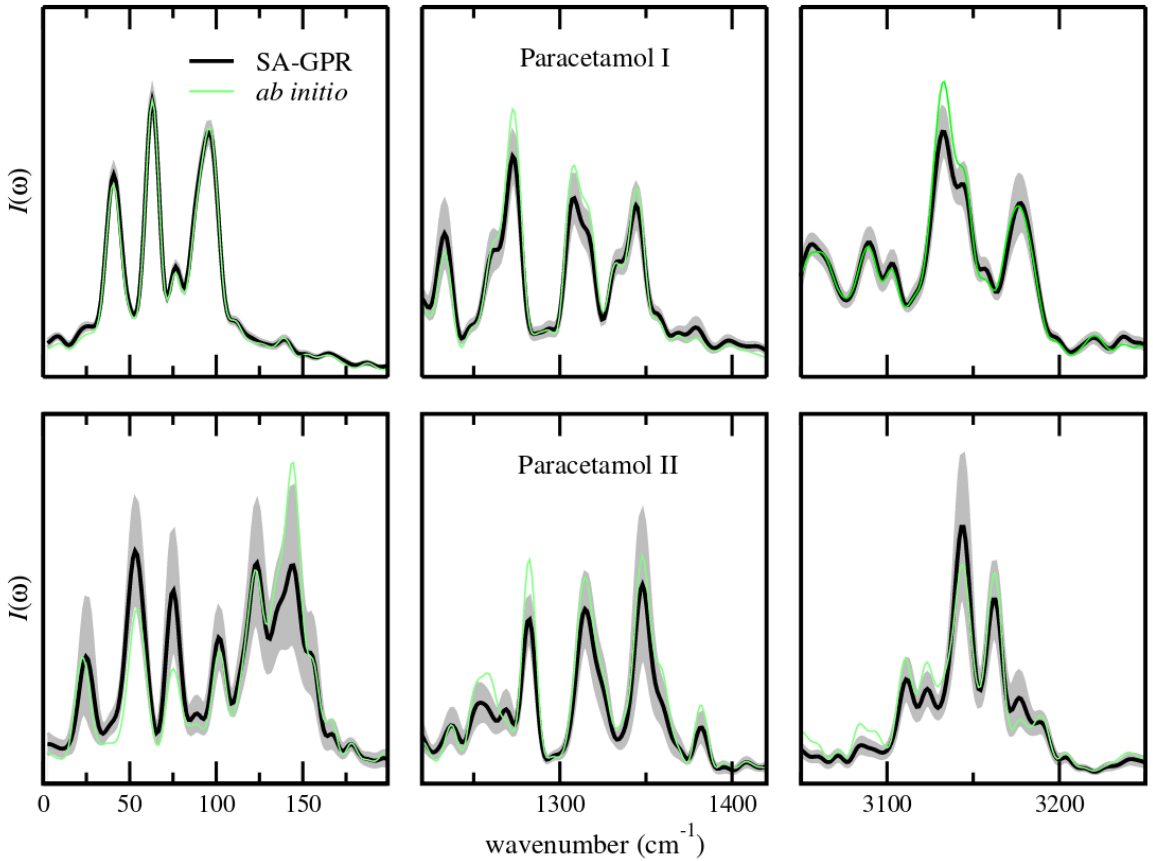
Musil, Willatt, **MC** JCTC (2019)

An accurate & inexpensive error estimation

- Generate an ensemble of GPR models, and use distribution of predictions

$$y(\mathcal{X}) = \frac{1}{N_{RS}} \sum_i y^{(i)}(\mathcal{X}), \qquad \sigma^2(\mathcal{X}) = \frac{1}{N_{RS}-1} \sum_i \left(y^{(i)}(\mathcal{X}) - y(\mathcal{X}) \right)^2$$

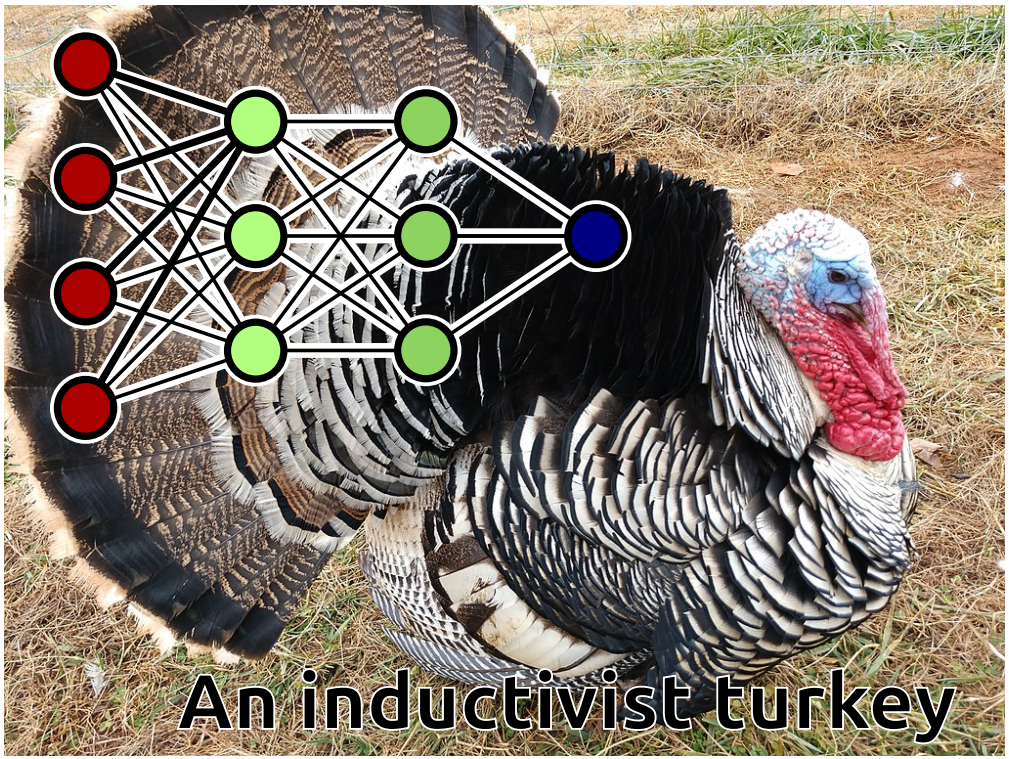
  - Verify accuracy by the distribution of errors $P\left(|y(\mathcal{X}) - y_{\text{ref}}(\mathcal{X})| \;\middle|\; \sigma(\mathcal{X})\right)$
  - Use maximum-likelihood to calibrate the uncertainty $\sigma(\mathcal{X}) \to \alpha\sigma(\mathcal{X})^{\gamma-1}$



N. Raimbault, A. Grisafi, **MC**, M. Rossi, work in progress

My machine learning wishlist

- General applicability: suitable for all systems and all types of properties
- Well-principled, mathematically robust and physically inspired
  - Symmetries of representations and target quantities
  - Locality, additivity, smoothness, conservation laws. . .
- Not only a fancy interpolator: use ML to gain insights and understanding



(Development) code available on http://cosmo-epfl.github.io

My machine learning wishlist

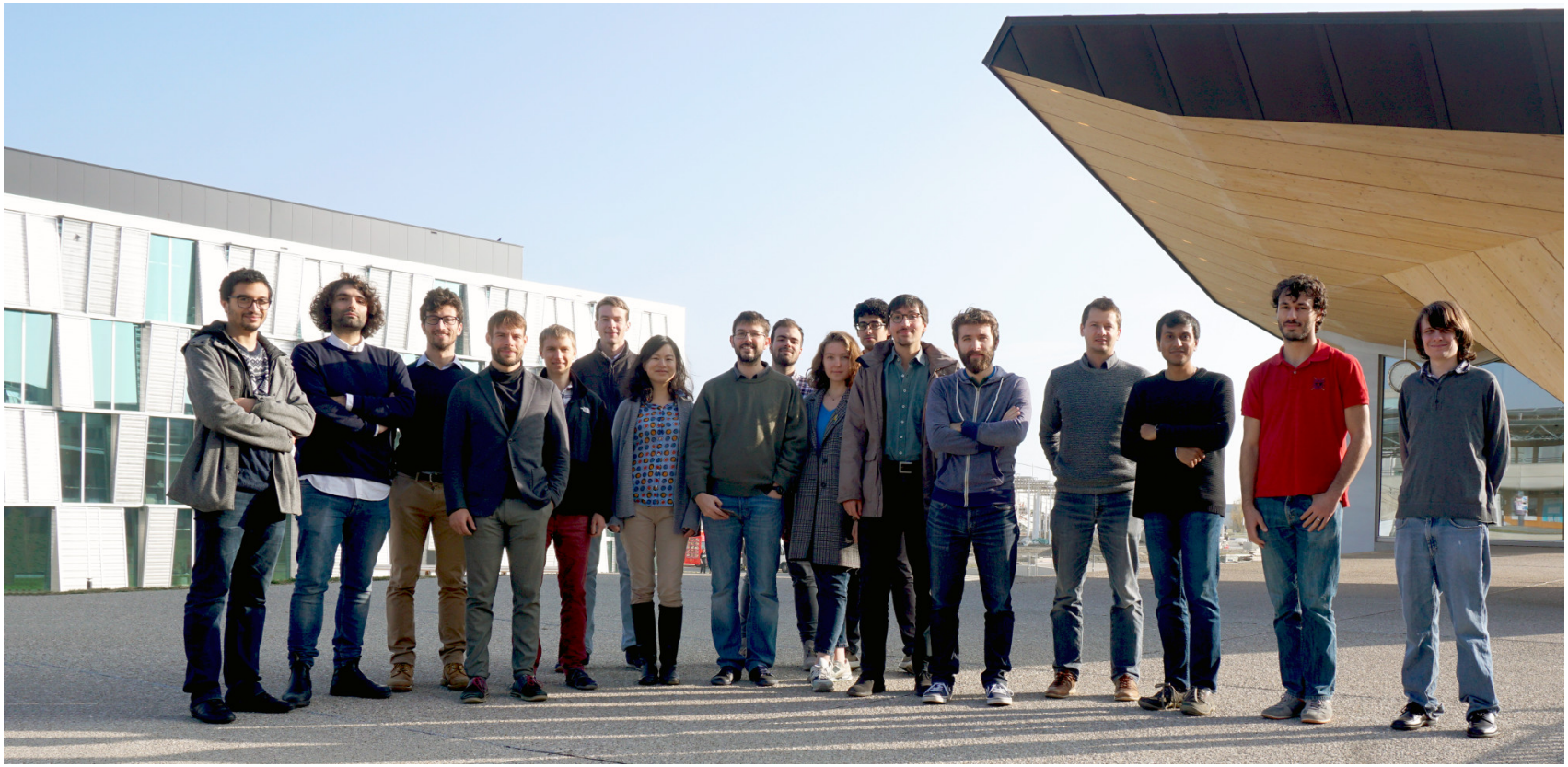
- General applicability: suitable for all systems and all types of properties
- Well-principled, mathematically robust and physically inspired
  - Symmetries of representations and target quantities
  - Locality, additivity, smoothness, conservation laws. . .
- Not only a fancy interpolator: use ML to gain insights and understanding

Deep connections between most representations . . . . . Willatt et al. arxiv:1807.00408
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)
Molecular polarizability . . . . . Wilkins et al. PNAS (2019)
Electron density . . . . . Grisafi et al., ACS Central Science (2019c)
Applications from water to biomolecules . . . . . .
Bartók et al. Science Adv. (2017); Musil et al., Chem. Sci. (2018);
Paruzzo et al. Nat.Comm. (2018); Cheng et al., PNAS (2019);
Deringer et al. PRB (2017); Dragoni et al. PRM (2017)

(Development) code available on http://cosmo-epfl.github.io

http://cosmo.epfl.ch

Follow @COSMO_EPFL



erc

European Commission

MARVEL

NATIONAL CENTRE OF COMPETENCE IN RESEARCH

CCMX Competence Centre for Materials Science and Technology

FNSNF

S.De, F.Musil, M.Willatt
A.Grisafi, D.Wilkins,
A. Anelli, E. Engel, G. Imbalzano

G.Csányi, A.Bartók, C.Poelking,
J.Kermode, N.Bernstein
F.Paruzzo, A.Hofstetter, L.Emsley
C.Corminboeuf, J.Behler, A.Paxton