Symmetry Matters

Learning Scalars and Tensors in Materials and Molecules

David M. Wilkins
http://cosmo.epfl.ch

MaX 2018, Trieste

COSMO
COMPUTATIONAL SCIENCE AND MODELLING

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

http://cosmo.epfl.ch

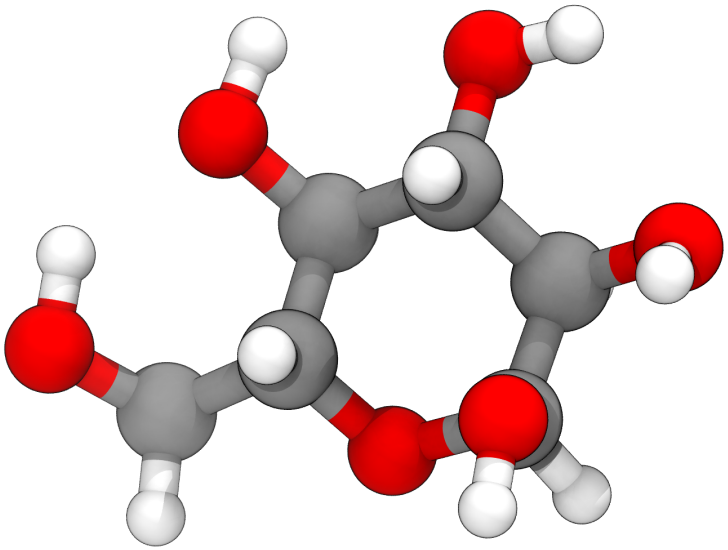
Follow @COSMO_EPFL

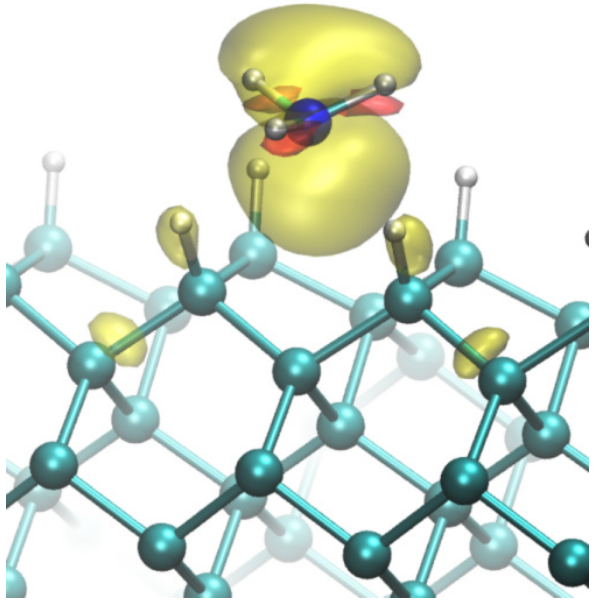
S. De, F. Musil, M. Willatt
Michele Ceriotti, Andrea Grisafi

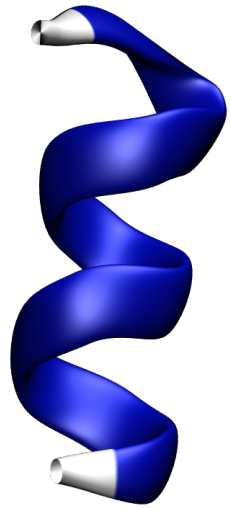
G. Csányi, A. Bartók, C. Poelking,
J. Kermode, N. Bernstein

A Universal Predictor of Atomic-Scale Properties

- The Schrödinger Equation allows – in principle! – prediction of any property for any kind of molecule or material
- Prohibitive computational cost
- A proliferation of ad-hoc electronic-structure methods and empirical potentials tuned to specific problems



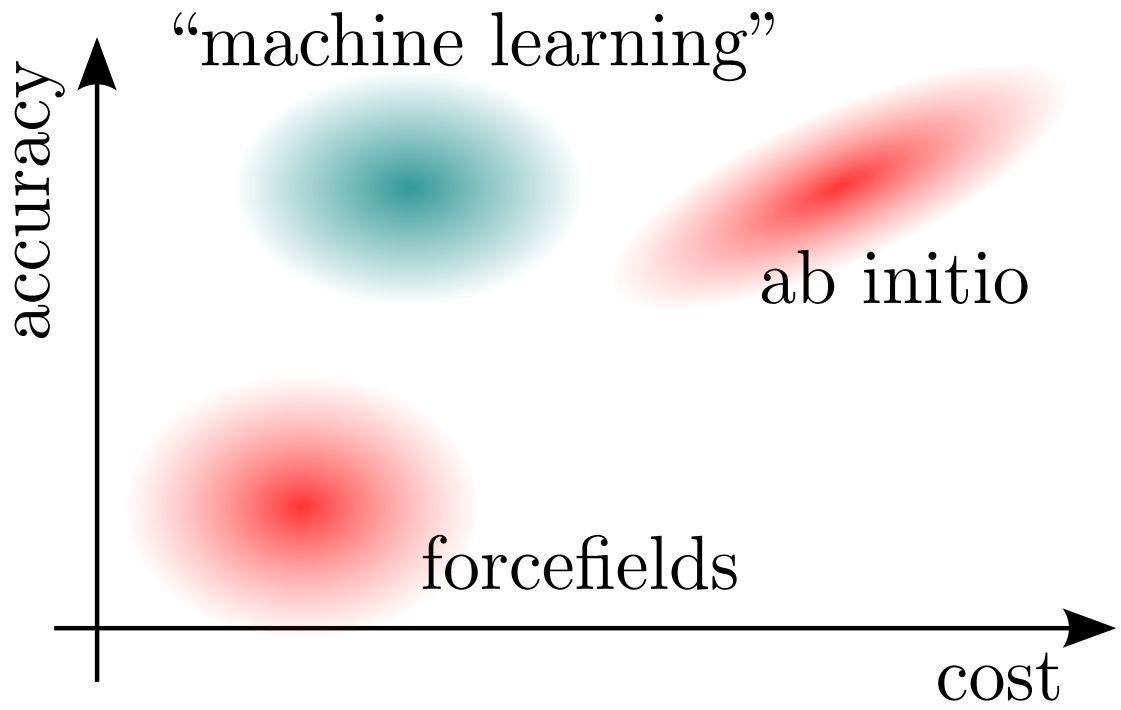




$$\hat{\mathcal{H}}\Psi = \mathrm{E}\Psi$$

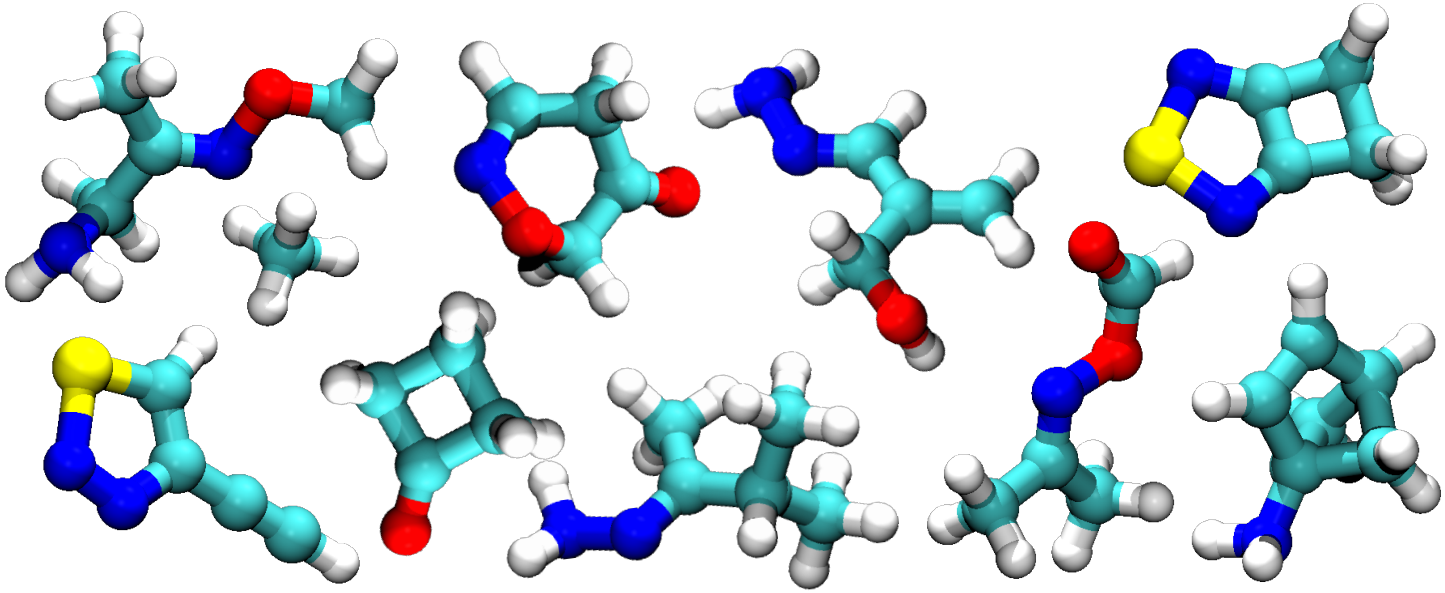
A Universal Predictor of Atomic-Scale Properties

- The Schrödinger Equation allows – in principle! – prediction of any property for any kind of molecule or material
- Prohibitive computational cost
- A proliferation of ad-hoc electronic-structure methods and empirical potentials tuned to specific problems



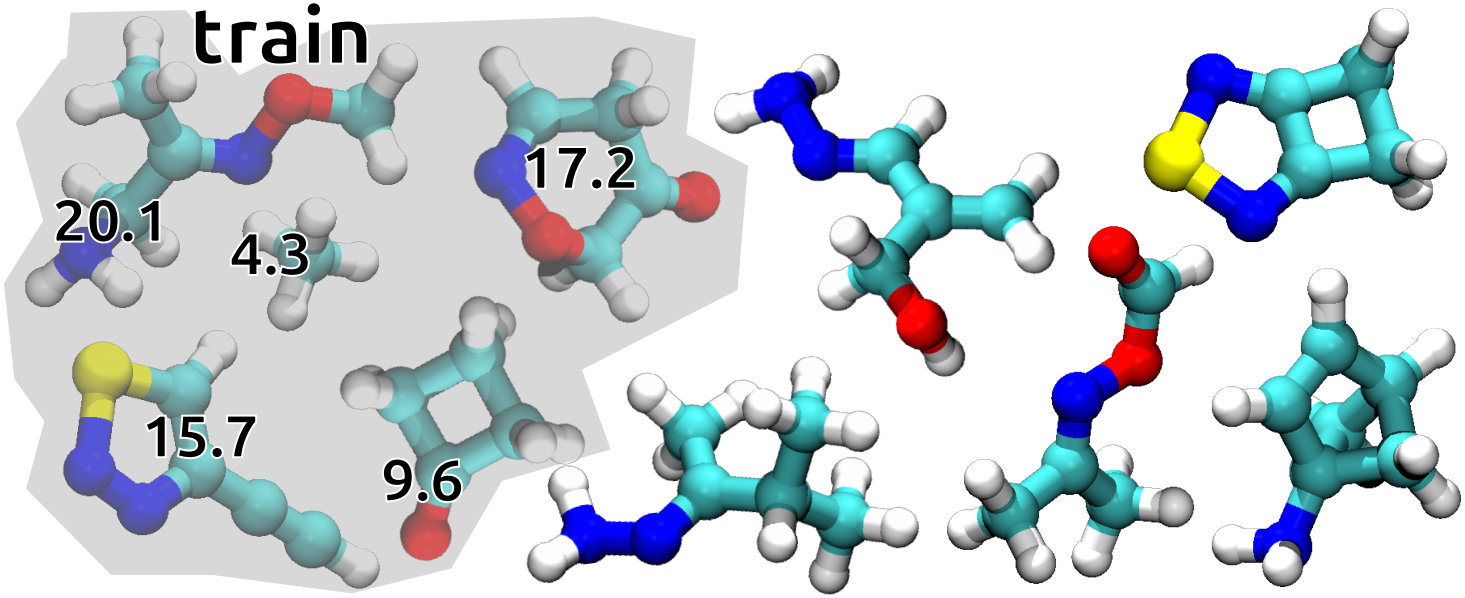
Machine-Learning as a Universal Interpolator

- Machine-learning can be regarded as a sophisticated interpolation between a few known values of the properties
- Can it be made as accurate and general as the Schrödinger equation?



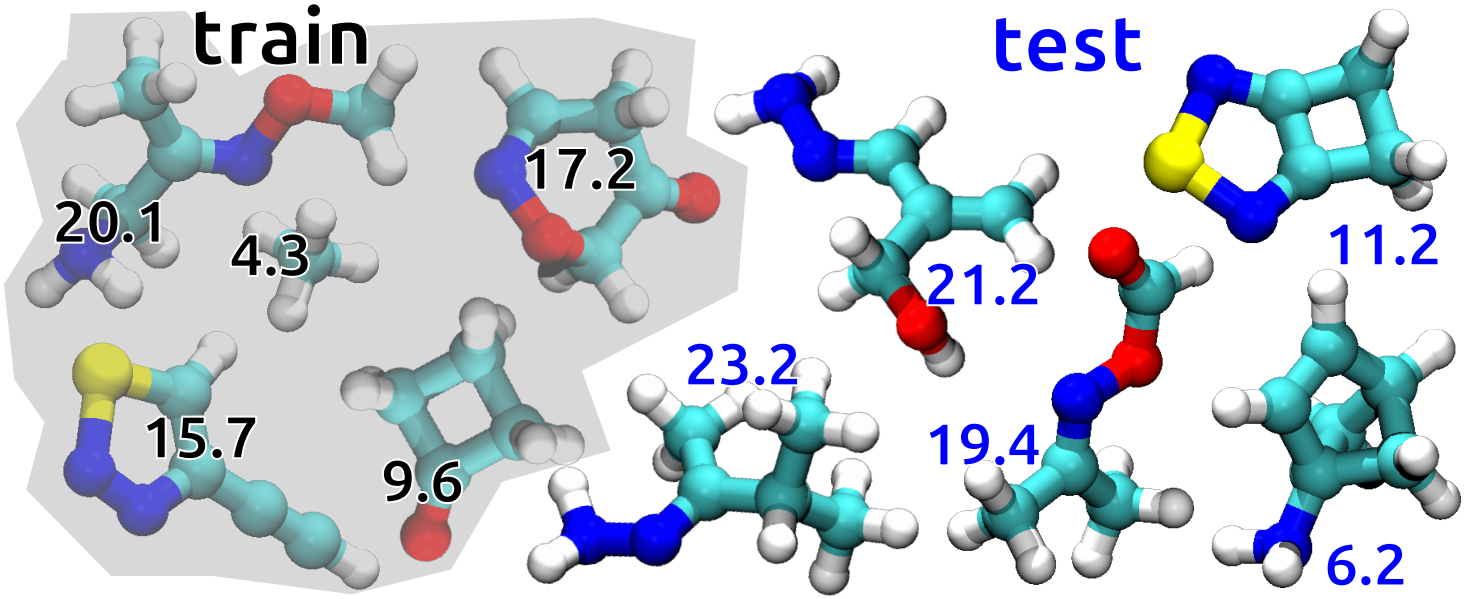
Machine-Learning as a Universal Interpolator

- Machine-learning can be regarded as a sophisticated interpolation between a few known values of the properties
- Can it be made as accurate and general as the Schrödinger equation?



Machine-Learning as a Universal Interpolator

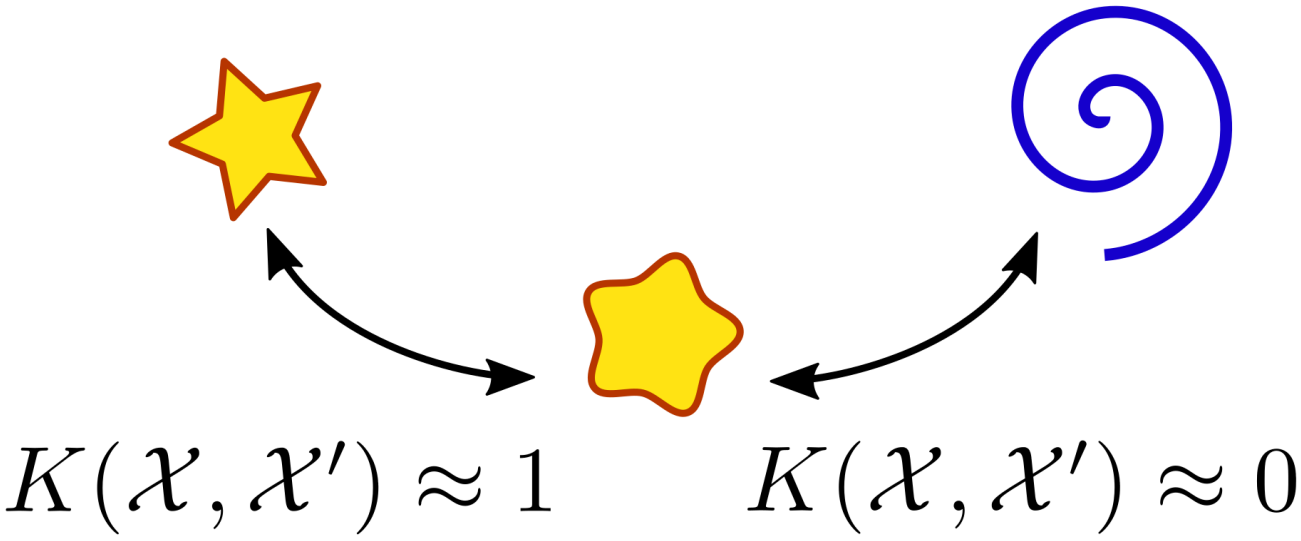
- Machine-learning can be regarded as a sophisticated interpolation between a few known values of the properties
- Can it be made as accurate and general as the Schrödinger equation?



$$\mathrm{E}\left(\mathcal{A}\right) = \sum_{\mathrm{i}} \mathrm{w_i} \mathrm{K}\left(\mathcal{A}, \mathcal{A}_\mathrm{i}\right)$$

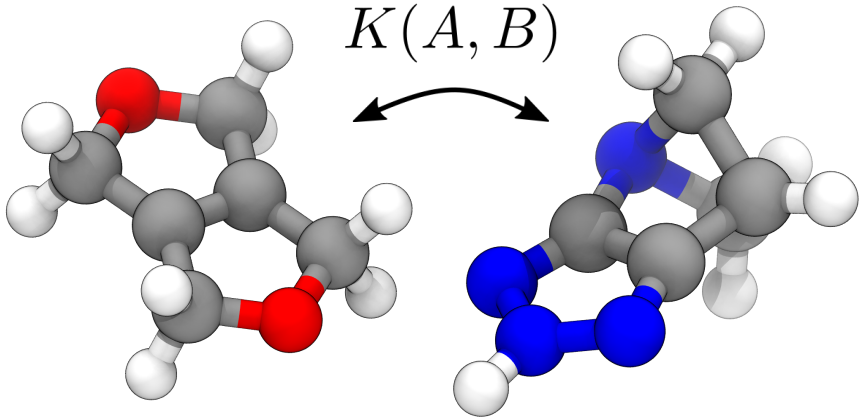
Measuring distances between materials

- The crucial ingredient in machine-learning is a method to compare the items whose properties should be predicted
- A kernel function K(A, B) can be used to assess the (dis)-similarity between items in a set



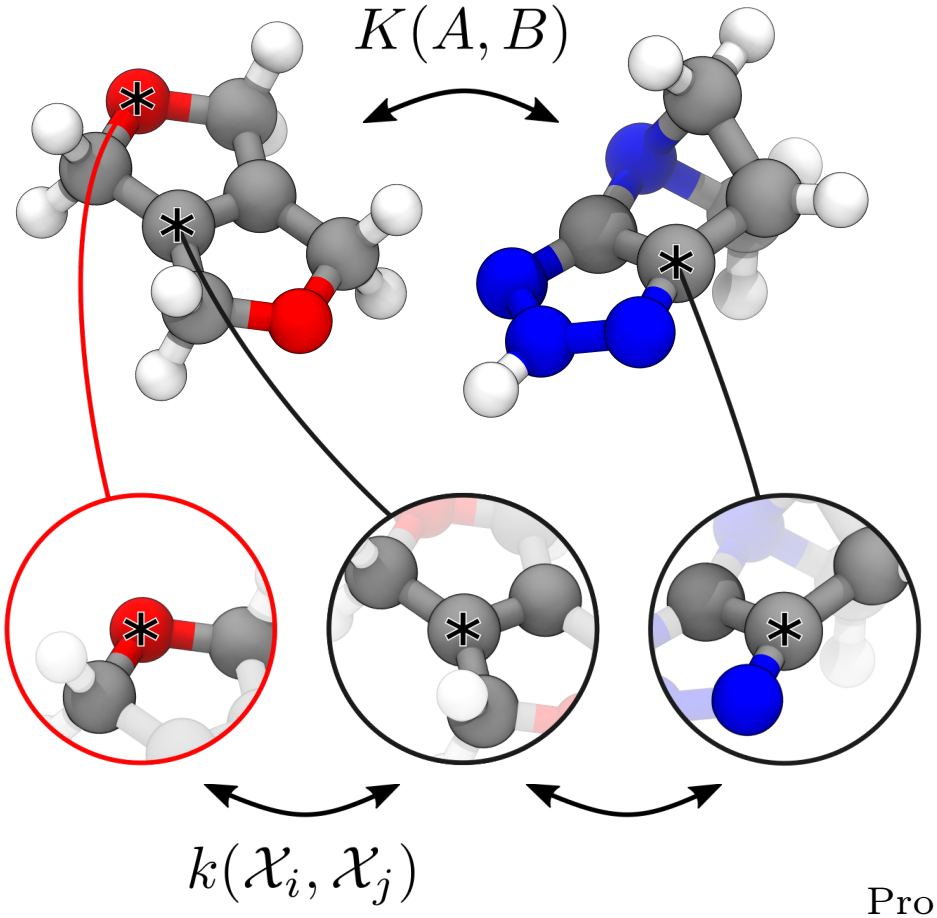
A General-Purpose Similarity Kernel

- How to compare two atomic structures?
- Start from a comparison of local environments!
- We use SOAP (smooth overlap of atomic positions) kernels – smooth, invariant to translations, rotations and permutations of identical atoms.



A General-Purpose Similarity Kernel

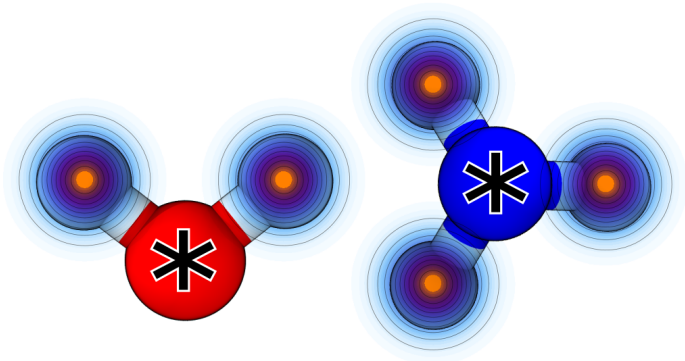
- How to compare two atomic structures?
- Start from a comparison of local environments!
- We use SOAP (smooth overlap of atomic positions) kernels – smooth, invariant to translations, rotations and permutations of identical atoms.



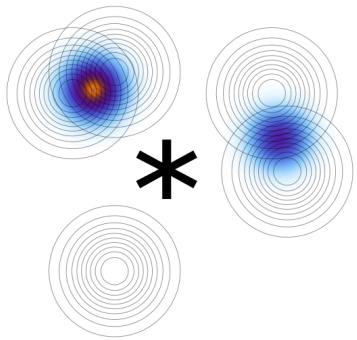
Prodan, Kohn, PNAS (2005)

A General-Purpose Similarity Kernel

- How to compare two atomic structures?
- Start from a comparison of local environments!
- We use SOAP (smooth overlap of atomic positions) kernels – smooth, invariant to translations, rotations and permutations of identical atoms.



$$\rho(\mathbf{x}) = \sum_i g(\mathbf{x} - \mathbf{x}_i)$$



$$k(\mathcal{X}, \mathcal{X}') \sim \int \rho(\mathbf{x})\rho'(\mathbf{x})$$

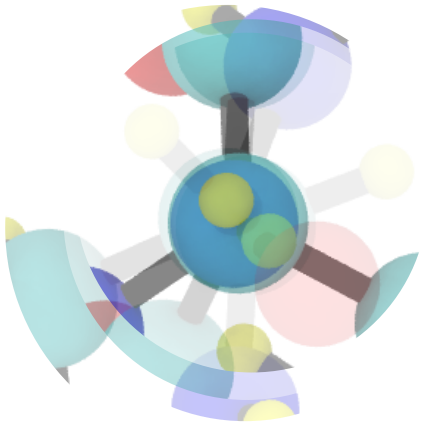
Bartók, Kondor, Csányi, PRB (2013)

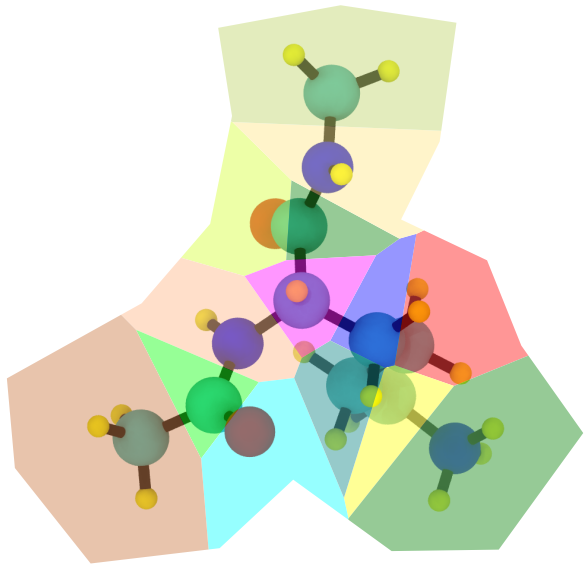
Additive Property Models & Beyond

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

$$\begin{array}{r} \mathrm{E}\left(\mathrm{A}\right) = \sum_{\mathrm{i}} \mathrm{W}_{\mathrm{i}} \mathrm{K}\left(\mathrm{A}, \mathrm{A}_{\mathrm{i}}\right) \\ \mathrm{K}\left(\mathrm{A}, \mathrm{B}\right) = \sum_{\mathrm{i}\in\mathrm{A},\mathrm{j}\in\mathrm{B}} \mathrm{k}\left(\mathcal{X}_{\mathrm{i}}, \mathcal{X}_{\mathrm{j}}\right) \end{array} \quad \Longleftrightarrow \quad \begin{array}{r} \epsilon\left(\mathcal{X}\right) = \sum_{\mathrm{i}} \mathrm{w}_{\mathrm{i}} \mathrm{k}\left(\mathcal{X}, \mathcal{X}_{\mathrm{i}}\right) \\ \mathrm{E}\left(\mathrm{A}\right) = \sum_{\mathrm{i}\in\mathrm{A}} \epsilon\left(\mathcal{X}_{\mathrm{i}}\right) \end{array}$$

- Entropy-regularized matching provides a natural way to go beyond additive models



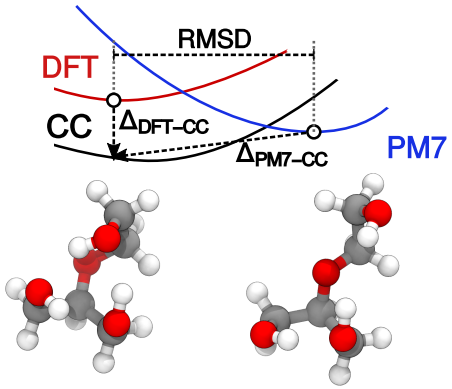


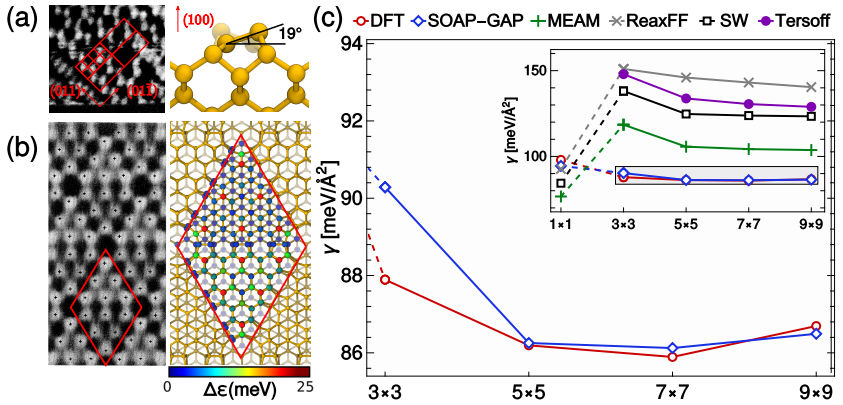
$$K(A,B) = \sum_{i,j} k(\mathcal{X}_i^A, \mathcal{X}_j^B)$$

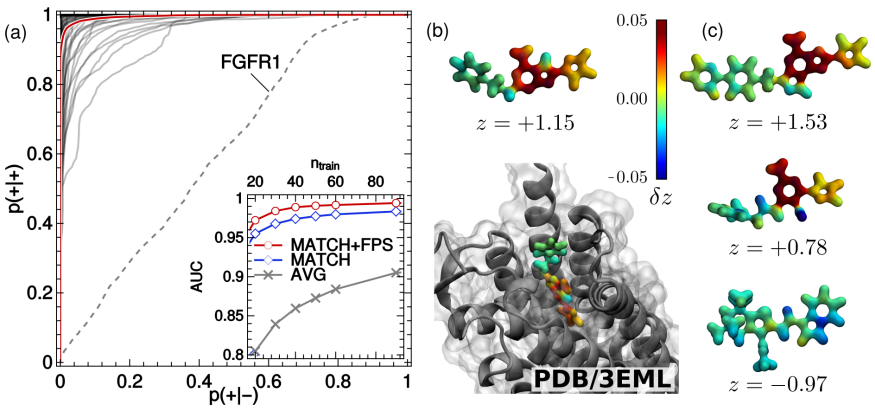
$$E(A) = \sum_i \epsilon(\mathcal{X}_i^A)$$

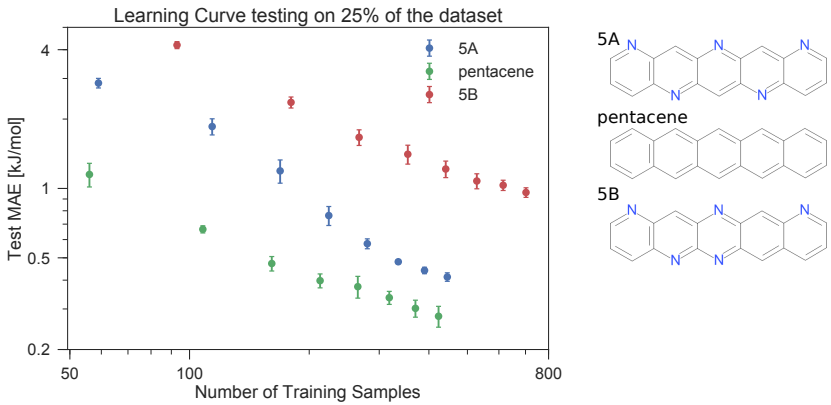
Accurate Prediction of Scalar Properties

- SOAP kernels with additive environment kernels allow for high-accuracy predictions of molecular and material properties





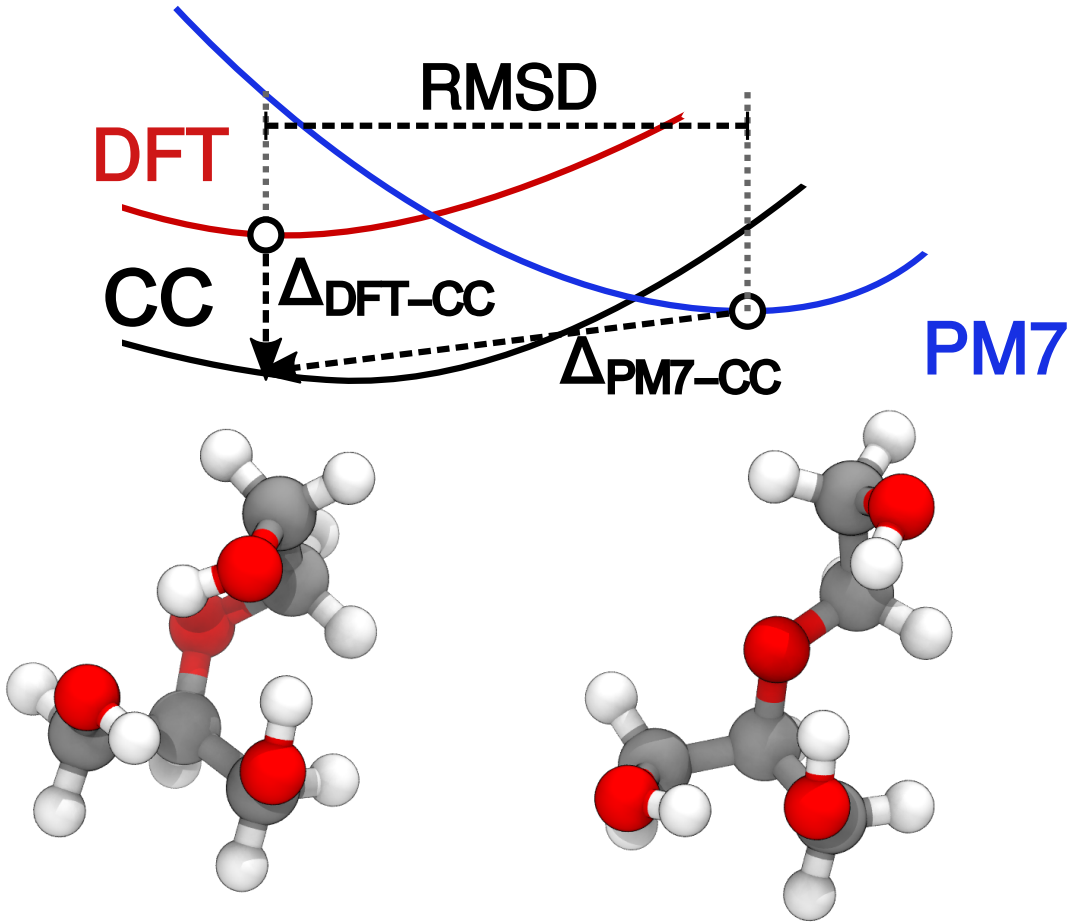




Bartok, De, Kermode, Bernstein, Csányi, Ceriotti, Sci. Adv. (2017);
pentacene data from G. Day and J. Yang

100k Molecules with Coupled-Cluster

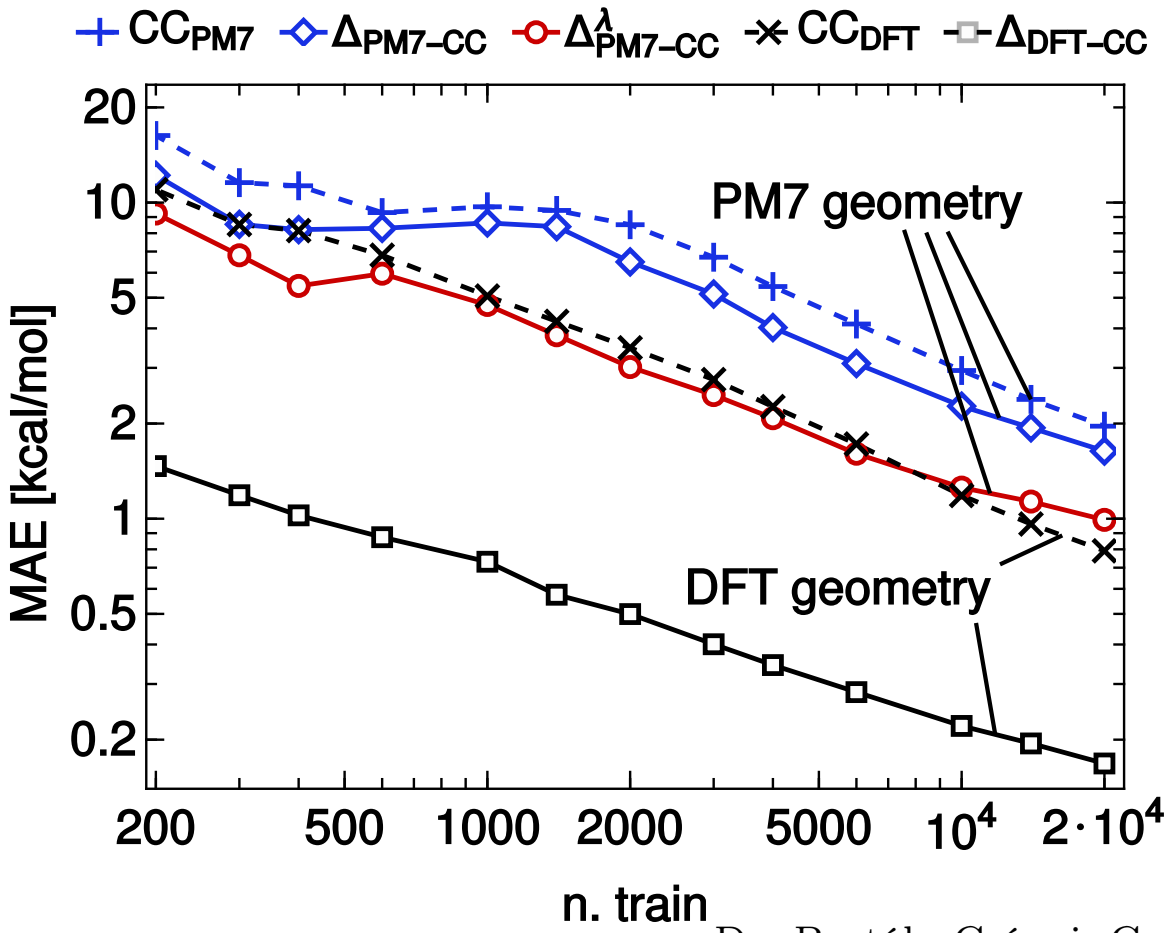
- CCSD(T) Energetics on the GDB9 database of small molecules - 114k useful predictions based on 20k training calculations
- 1 kcal/mol error for predicting CCSD(T) based on PM7 geometries; 0.18 kcal/mol error for predicting CCSD(T) based on DFT geometries!



Ramakrishnan et al., Scientific Data (2014); Ramakrishnan et al., JCTC (2015)

100k Molecules with Coupled-Cluster

- CCSD(T) Energetics on the GDB9 database of small molecules - 114k useful predictions based on 20k training calculations
- 1 kcal/mol error for predicting CCSD(T) based on PM7 geometries; 0.18 kcal/mol error for predicting CCSD(T) based on DFT geometries!



De, Bartók, Csányi, Ceriotti, PCCP (2016);
Bartok, De, Kermode, Bernstein, Csányi, Ceriotti, Sci. Adv. (2017)

Symmetries in Machine-Learning

- 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:

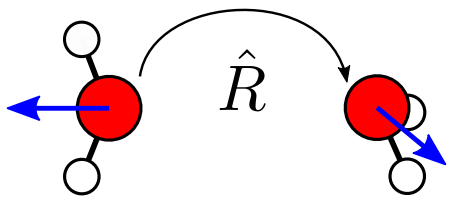
$$\mathrm{k}\left(\mathcal{X},\mathcal{X}'\right) \leftrightarrow \left\langle \mathrm{y}\left(\mathcal{X}\right);\mathrm{y}\left(\mathcal{X}'\right)\right\rangle, \text{ so } \mathrm{k}\left(\hat{\mathrm{S}}\mathcal{X},\hat{\mathrm{S}}'\mathcal{X}'\right) \leftrightarrow \left\langle \mathrm{y}\left(\hat{\mathrm{S}}\mathcal{X}\right);\mathrm{y}\left(\hat{\mathr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

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

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

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

$$\mathrm{k}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right) \leftrightarrow \left\langle \mathrm{T}_{\mu}\left(\mathcal{X}\right);\mathrm{T}_{\nu}\left(\mathcal{X}'\right)\right\rangle \rightarrow \mathrm{k}_{\mu\nu}\left(\hat{\mathrm{R}}\mathcal{X},\hat{\mathrm{R}}'\mathcal{X}'\right) = \mathrm{R}_{\mu\mu'}\mathrm{k}_{\mu'\nu'}\left(\mathcal{X},\mathcal{X}'\right)\mathrm{R}'_{\nu\nu'}$$



Glielmo, Sollich, De Vita, PRB (2017);
Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

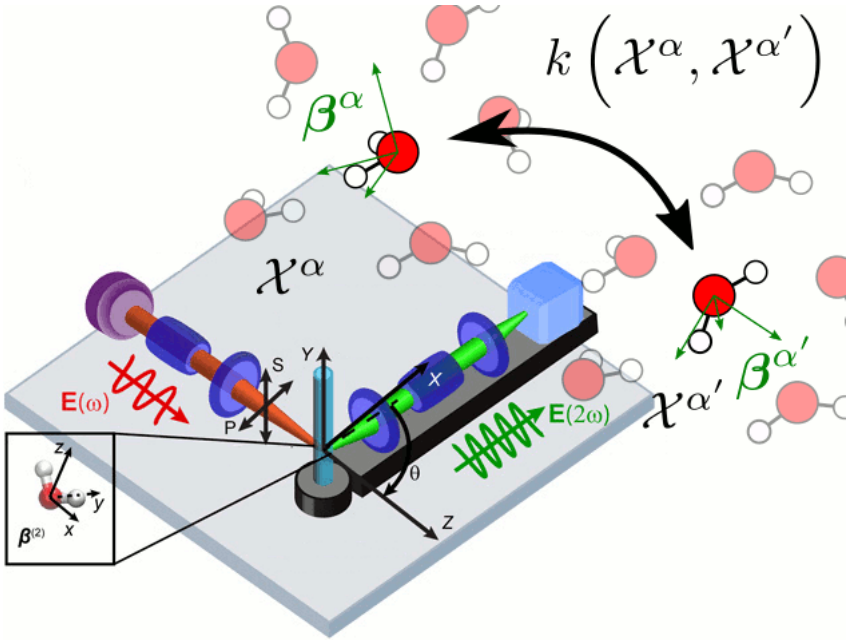
A Simple (but Limited) Solution

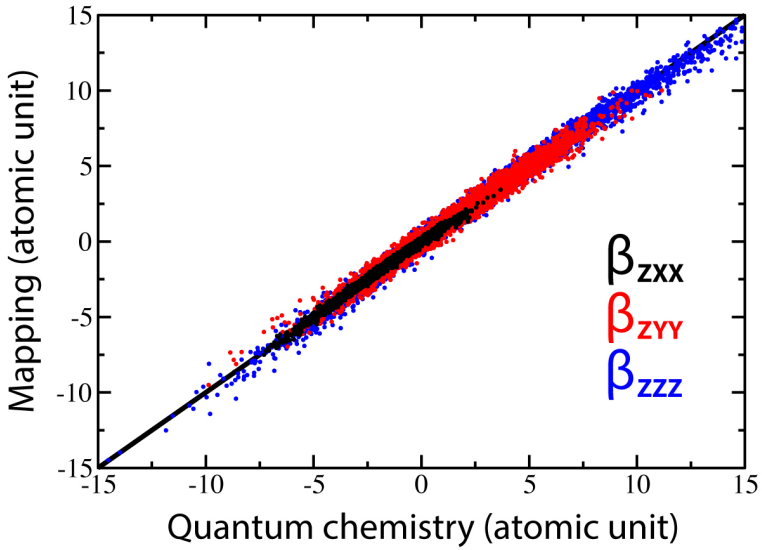
- For rigid molecules, one can convert the tensor to a reference frame and learn individual components using an invariant kernel

$$\mathrm{k}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right) \equiv \mathrm{R}\left(\mathcal{X}\right)_{\mu \mathrm{j}}\, \mathrm{k}\left(\mathcal{X},\mathcal{X}'\right) \mathrm{R}\left(\mathcal{X}'\right)_{\nu \mathrm{j}}\,,$$

$$\mathrm{k}\left(\mathcal{X},\mathcal{X}'\right) = \tilde{\mathrm{k}}\left(\mathrm{R}\left(\mathcal{X}\right)\mathcal{X}, \mathrm{R}\left(\mathcal{X}'\right)\mathcal{X}'\right)$$

- Learning of second-harmonic response of water solutions (SHS experiments)

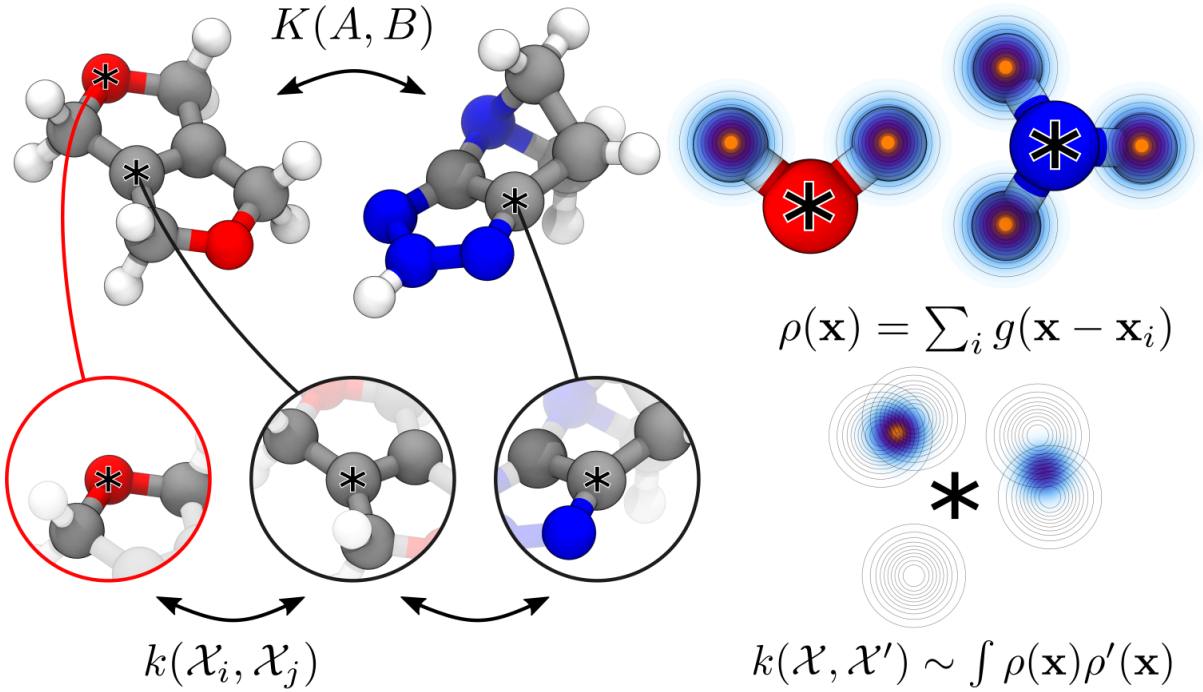




Bereau, Andrienko, von Lilienfeld, JCTC (2015);
Liang, Tocci, Wilkins, Grisafi, Roke, Ceriotti, PRB (2017)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities



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

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

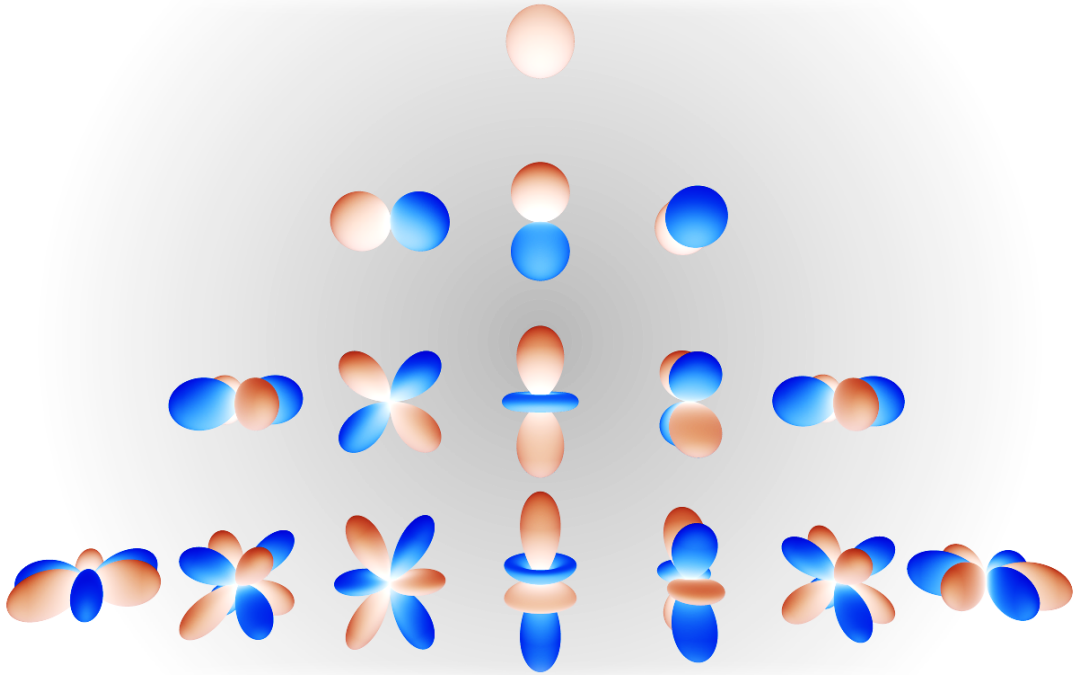


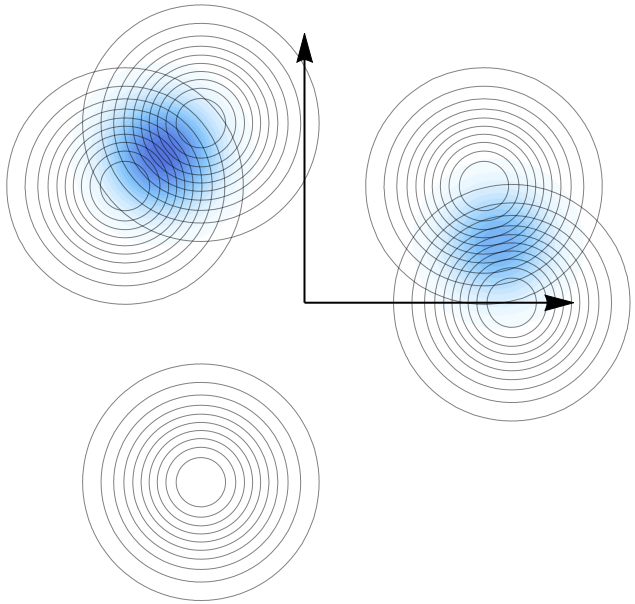
image from: Wikipedia

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

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

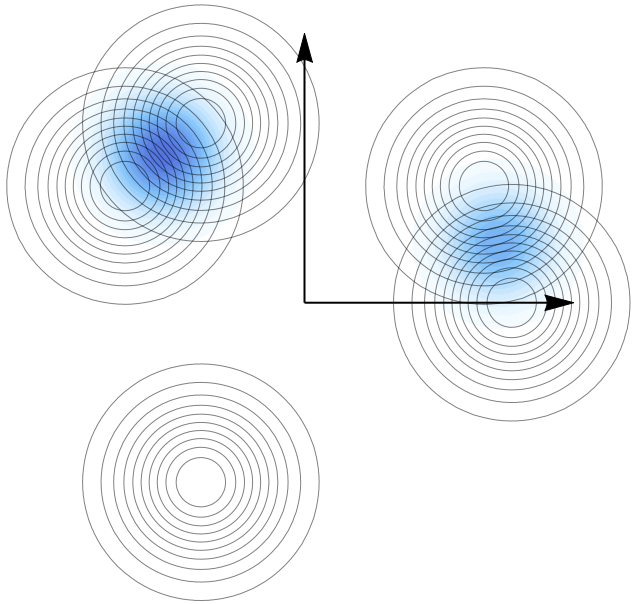


$$\mathrm{k}^{0}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

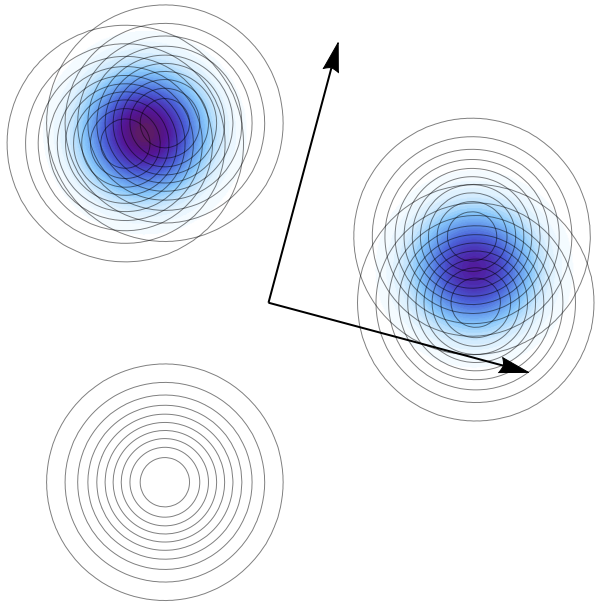


$$\mathrm{k}^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\mathrm{D}^{\lambda}_{\mu\nu}\left(\hat{\mathrm{R}}\right)\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

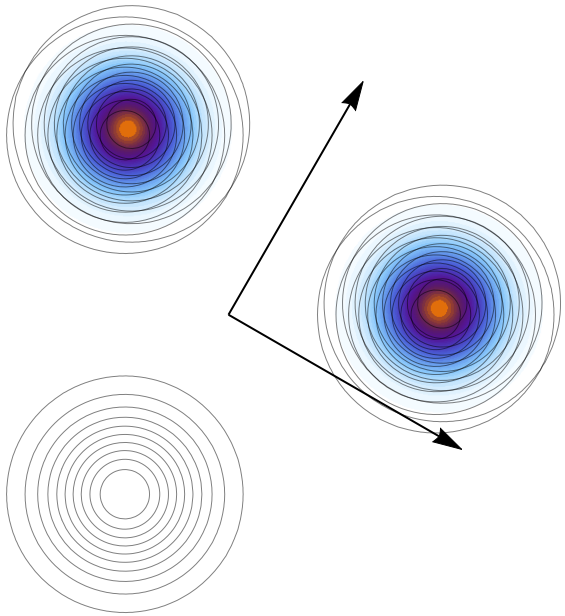


$$\mathrm{k}^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\mathrm{D}^{\lambda}_{\mu\nu}\left(\hat{\mathrm{R}}\right)\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

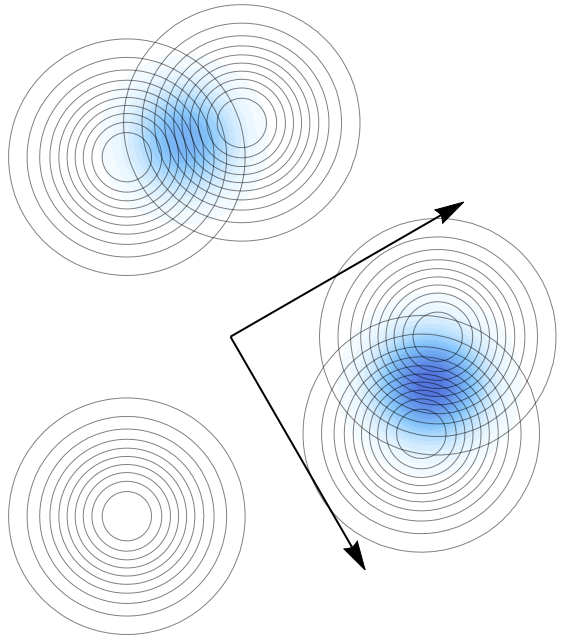


$$\mathrm{k}^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\mathrm{D}^{\lambda}_{\mu\nu}\left(\hat{\mathrm{R}}\right)\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

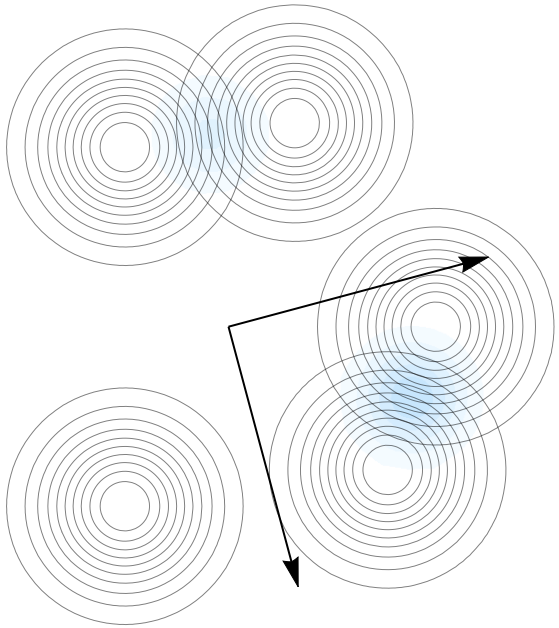


$$\mathrm{k}^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\mathrm{D}^{\lambda}_{\mu\nu}\left(\hat{\mathrm{R}}\right)\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

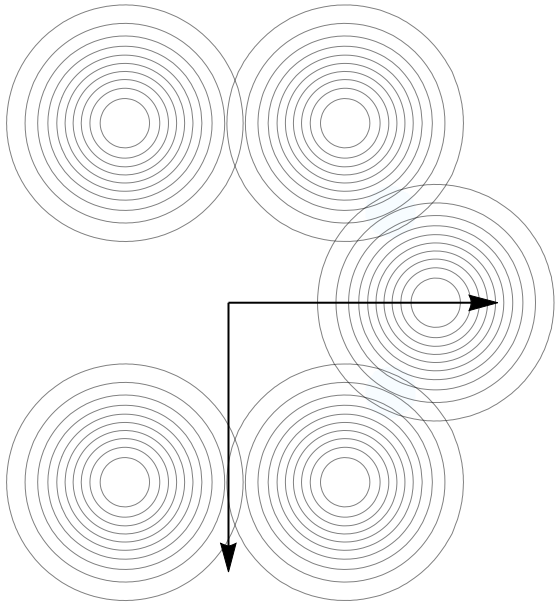


$$\mathrm{k}^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\mathrm{D}^{\lambda}_{\mu\nu}\left(\hat{\mathrm{R}}\right)\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

$\lambda-$SOAP Kernel

- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components T^{λ}
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

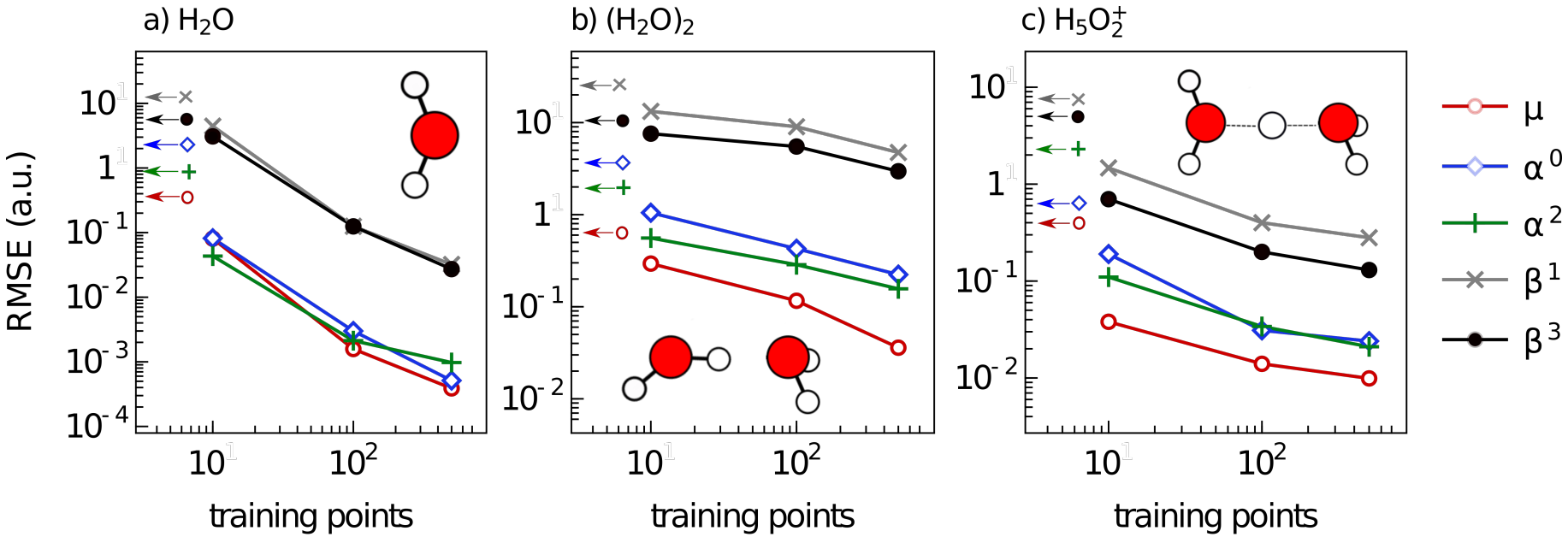


$$\mathrm{k}^{\lambda}_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\int \mathrm{d}\hat{\mathrm{R}}\,\mathrm{D}^{\lambda}_{\mu\nu}\left(\hat{\mathrm{R}}\right)\kappa\left(\mathcal{X},\hat{\mathrm{R}}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

Machine-Learning the Dielectric Response of Water

- 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 - which gives meaningful partitioning into molecular contributions
- Works well for bulk systems (liquid & ice) after fixing non-additive terms

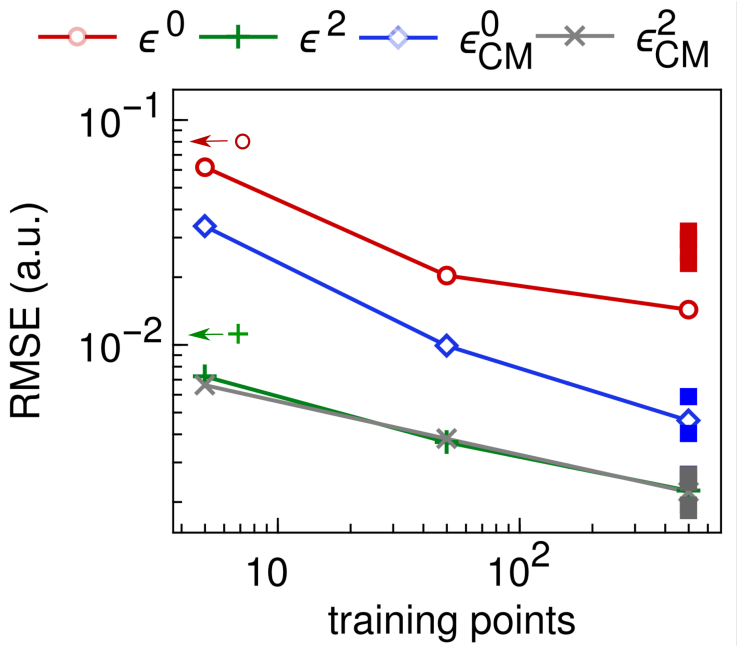


$$\mathrm{K}_{\mu\nu}\left(\mathrm{A},\mathrm{B}\right) = \frac{1}{\mathrm{N_A N_B}} \sum_{\mathrm{ij}} \mathrm{k}_{\mu\nu}\left(\mathcal{X}_{\mathrm{i}}^{\mathrm{A}}, \mathcal{X}_{\mathrm{j}}^{\mathrm{B}}\right)$$

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

Machine-Learning the Dielectric Response of Water

- 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 - which gives meaningful partitioning into molecular contributions
- Works well for bulk systems (liquid & ice) after fixing non-additive terms



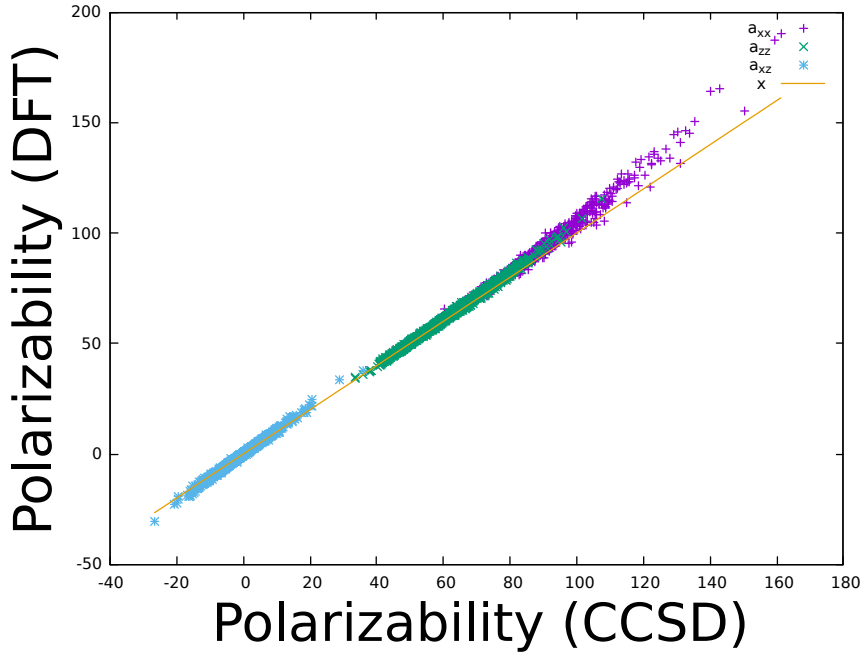
Clausius-Mossotti: $\boldsymbol{\alpha} = (\boldsymbol{\varepsilon} - 1)(\boldsymbol{\varepsilon} + 2)^{-1}\mathrm{V}$

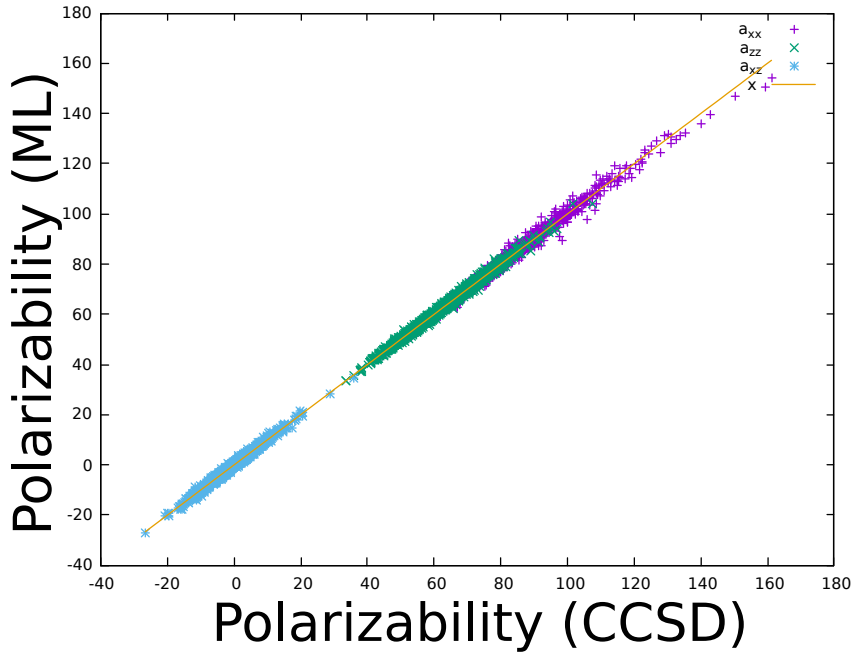
Learning a localized property gives much better results!

Grisafi, Wilkins, Csányi, Ceriotti, PRL (2018)

Predicting the Full Polarizability of Molecules

- Benchmarking polarizability learning on the QM7b dataset. DFT and high-end coupled-cluster references (Rob DiStasio@Cornell)
- Preliminary tests (1400 training structures) - we can predict $\boldsymbol{\alpha}$ with better accuracy than DFT.



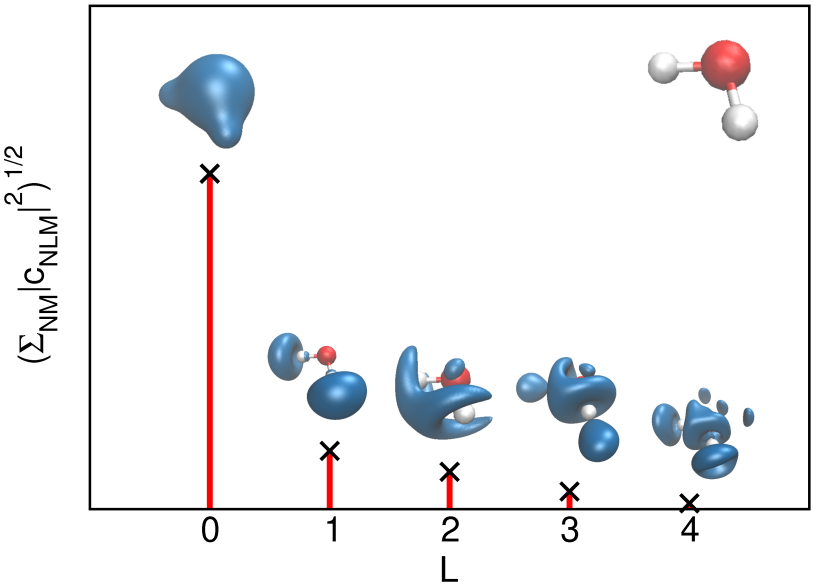


$\langle\cdot\rangle, \sigma(\cdot)$ [a.u.]	α_{xx}	α_{yy}	α_{zz}	α_{xy}	α_{xz}	α_{yz}
DFT vs CCSD	2.6, 2.6	2.0, 2.1	0.9, 0.9	0.6, 1.3	0.0, 0.6	0.1, 0.6
SA-GPR vs CCSD	0.0, 1.5	0.0, 1.4	0.0, 0.9	0.0, 1.0	0.0, 0.7	0.0, 0.6
ΔSA-GPR	0.0, 0.7	0.0, 0.6	0.0, 0.3	0.0, 0.4	0.0, 0.3	0.0, 0.2

Learning Charge Densities

- Charge density gives access to a wide variety of properties
- Decomposing the density into (localized) components that transform as spherical harmonics means we can learn them with SA-GPR:

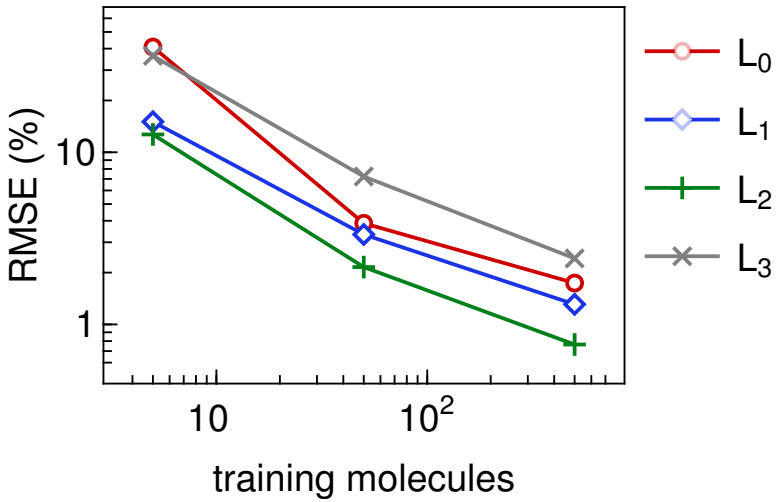
$$\rho(\mathrm{r}) = \sum_{\mathrm{i}} \sum_{\mathrm{nlm}} \mathrm{c}^{(\mathrm{i})}_{\mathrm{nlm}} \mathrm{g}_{\mathrm{n}}(|\mathrm{r} - \mathrm{r}_{\mathrm{i}}|) \mathrm{Y}^{\mathrm{m}}_{\mathrm{l}}(\Theta)$$

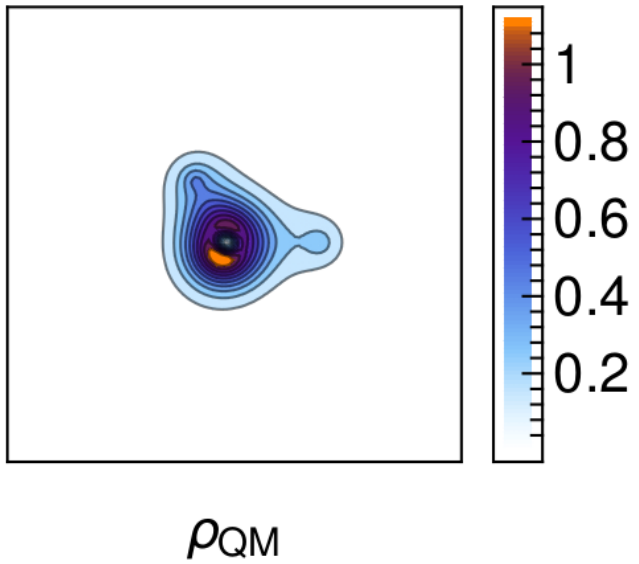


Learning Charge Densities

- Charge density gives access to a wide variety of properties
- Decomposing the density into (localized) components that transform as spherical harmonics means we can learn them with SA-GPR:

$$\rho(\mathrm{r}) = \sum_{\mathrm{i}} \sum_{\mathrm{nlm}} \mathrm{c}^{(\mathrm{i})}_{\mathrm{nlm}} \mathrm{g}_{\mathrm{n}}(|\mathrm{r} - \mathrm{r}_{\mathrm{i}}|) \mathrm{Y}^{\mathrm{m}}_{\mathrm{l}}(\Theta)$$

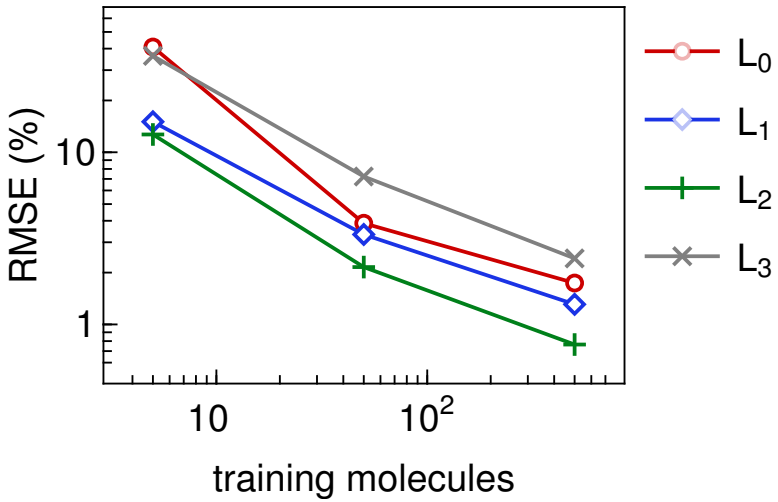


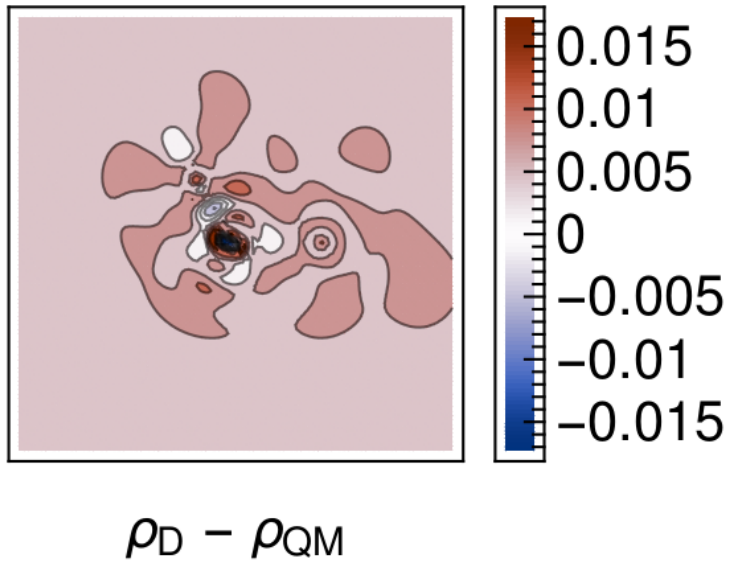


Learning Charge Densities

- Charge density gives access to a wide variety of properties
- Decomposing the density into (localized) components that transform as spherical harmonics means we can learn them with SA-GPR:

$$\rho(\mathrm{r}) = \sum_{\mathrm{i}} \sum_{\mathrm{nlm}} \mathrm{c}^{(\mathrm{i})}_{\mathrm{nlm}} \mathrm{g}_{\mathrm{n}}(|\mathrm{r} - \mathrm{r}_{\mathrm{i}}|) \mathrm{Y}^{\mathrm{m}}_{\mathrm{l}}(\Theta)$$

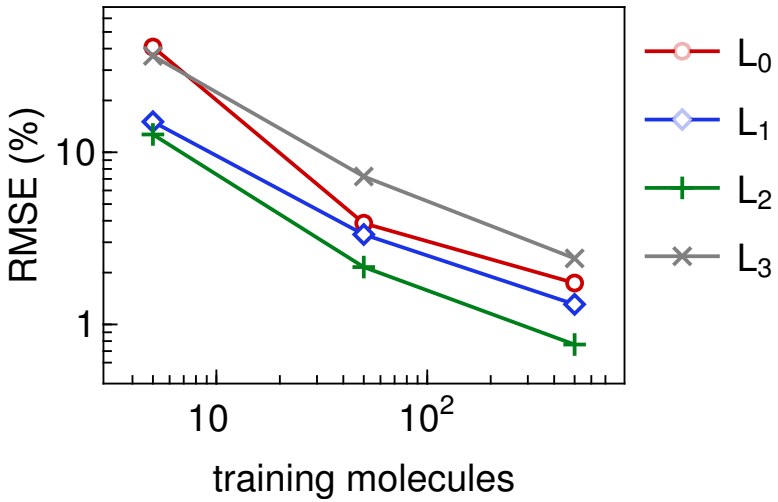


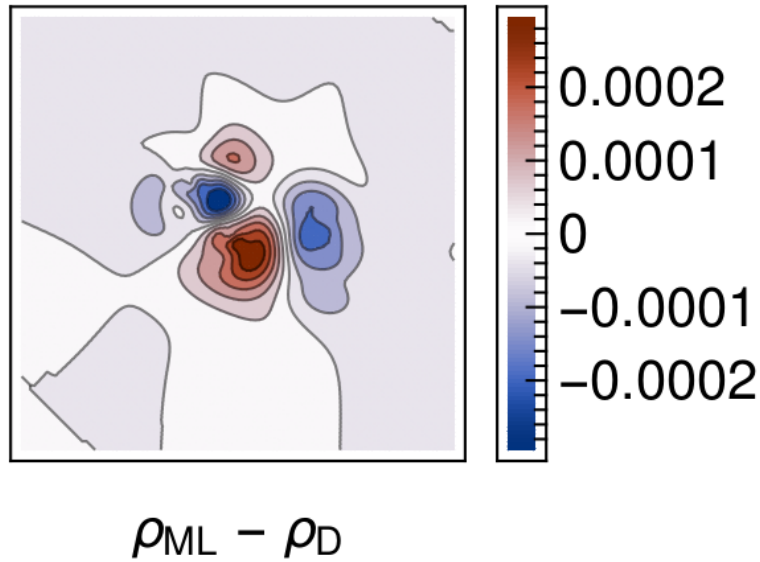


Learning Charge Densities

- Charge density gives access to a wide variety of properties
- Decomposing the density into (localized) components that transform as spherical harmonics means we can learn them with SA-GPR:

$$\rho(\mathrm{r}) = \sum_{\mathrm{i}} \sum_{\mathrm{nlm}} \mathrm{c}^{(\mathrm{i})}_{\mathrm{nlm}} \mathrm{g}_{\mathrm{n}}(|\mathrm{r} - \mathrm{r}_{\mathrm{i}}|) \mathrm{Y}^{\mathrm{m}}_{\mathrm{l}}(\Theta)$$

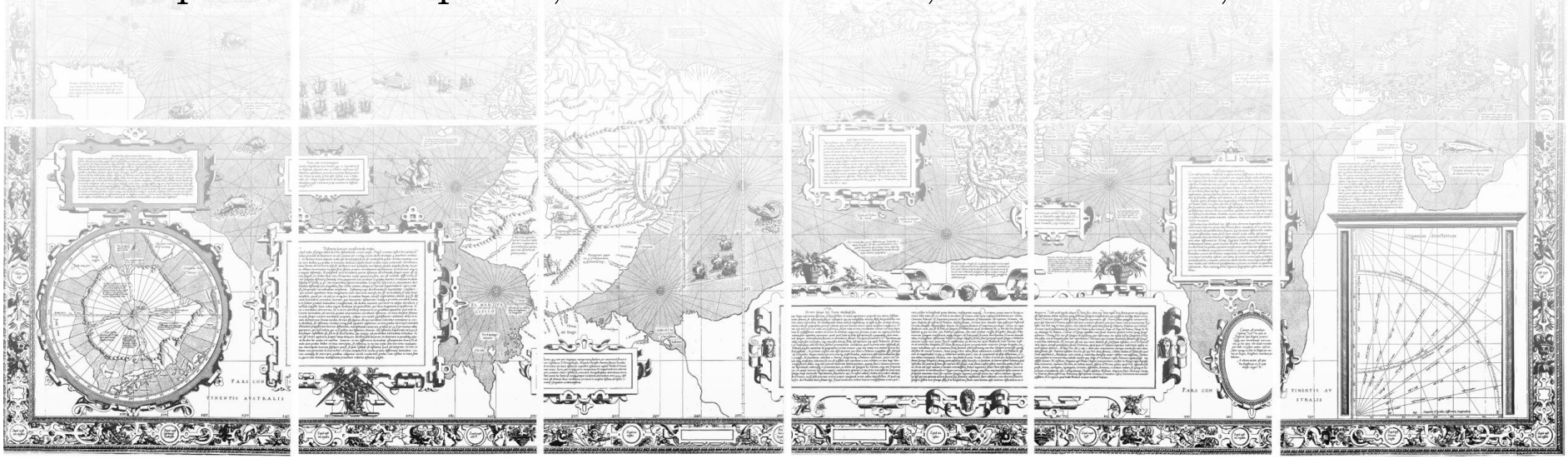




$\sim 1\%$ error for ab initio density with 500 training points

Outlook

- Building structural kernels from local-environments SOAP fingerprints
  - "Nearsightedness" of electronic matter, beyond additive models using entropy-regularized kernels
  - Excellent perfomance on benchmark DBs: ∼1 kJ/mol for 80%GDB9 and 75%QM7b(multi-scale)
  - Predictions all the way PM7→CC, potentials for solids, silicon & molecular crystals
  - Ingredients for effective learning: sound mathematical foundation, cross-species learning & multi-scale kernels, training set sparsification
- Huge potential of a SA-GPR framework to learn tensors - electric multipoles and response, but also densities, Hamiltonians, . . .



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