

# Physics-Based Machine Learning for Materials and Molecules

Michele Ceriotti  
<https://cosmo.epfl.ch>

CECAM Discussion Meeting  
December 2018, CERMICS



<http://cosmo.epfl.ch>

Follow @COSMO\_EPFL



MARVEL



NATIONAL CENTRE OF COMPETENCE IN RESEARCH



CCMX

Competence Centre for  
Materials Science and Technology



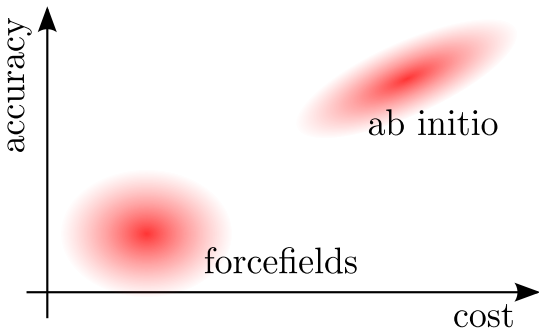
FNSNF

**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

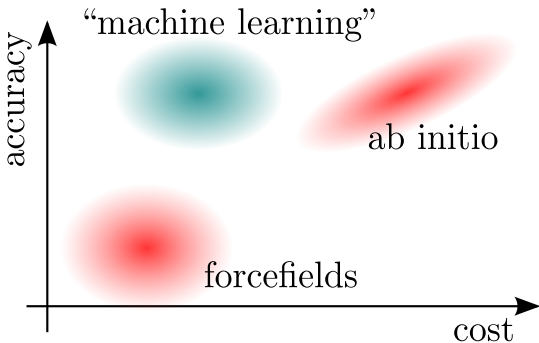
# Why machine learning?

- Statistical sampling requires large size and long time simulations, but also an accurate evaluation of electronic energy and properties
- Traditionally a tradeoff between cost, accuracy and transferability
- Use machine learning to get around these limitations



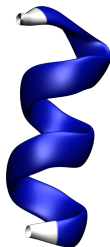
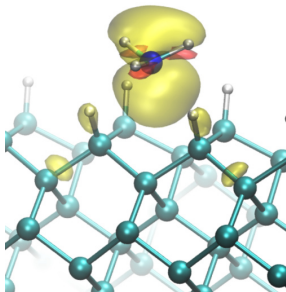
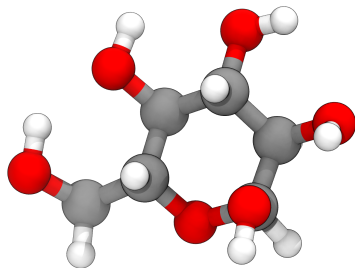
# Why machine learning?

- Statistical sampling requires large size and long time simulations, but also an accurate evaluation of electronic energy and properties
- Traditionally a tradeoff between cost, accuracy and transferability
- Use machine learning to get around these limitations



# My machine learning wishlist

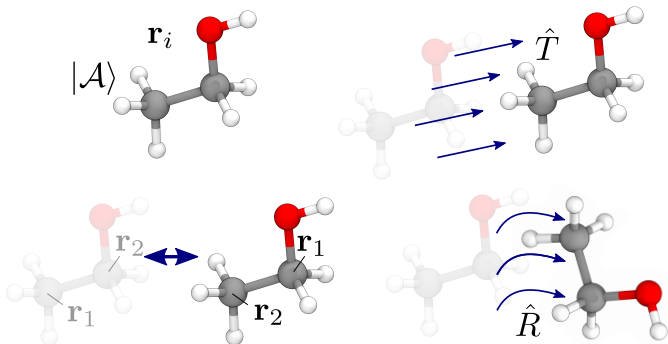
- General applicability: suitable for all systems and all types of properties
- Well-principled: incorporates structure and symmetries of physical laws
- Not only a fancy interpolator: use ML to gain insights and understanding



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

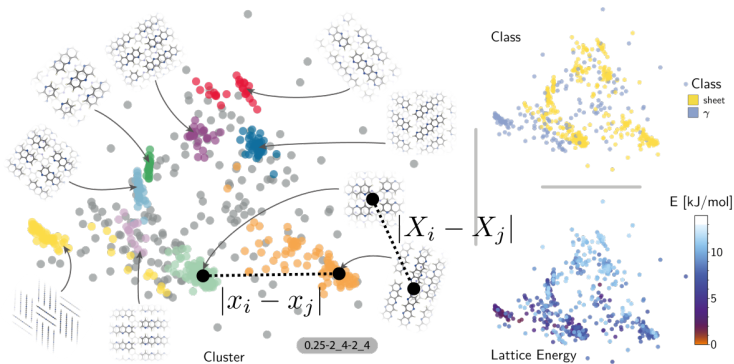
# My machine learning wishlist

- 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



# My machine learning wishlist

- 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



$$\{X_i\} \Rightarrow \{x_i\} \quad \{x_i\} = \operatorname{argmin} \sum_{ij} [s(|X_i - X_j|) - s(|x_i - x_j|)]^2$$

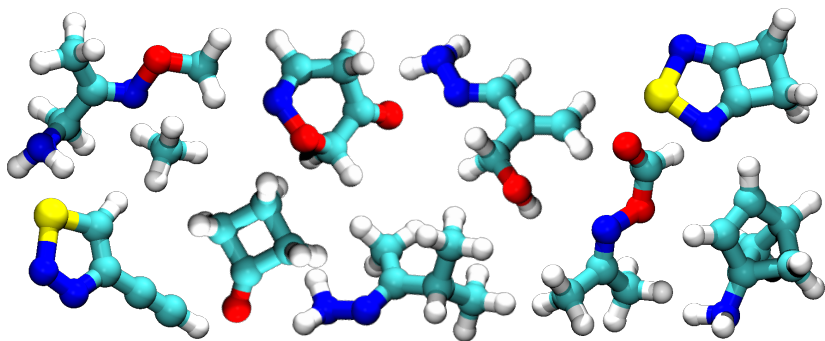
MC, Tribello, Parrinello, PNAS (2011); Musil, [...], MC, Chem. Sci. (2018); <http://interactive.sketchmap.org>

# **A transferable ML model for materials and molecules**



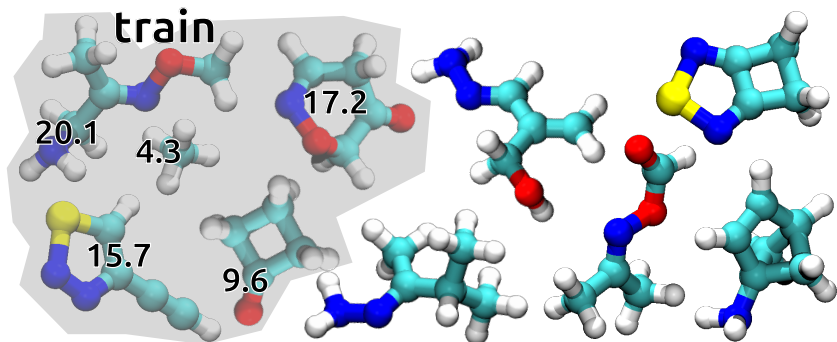
# A universal surrogate quantum model

- 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?
- Kernels are the main ingredient. Think of them as scalar products between structures,  $K(A, B) \sim \langle A|B \rangle$ .



# A universal surrogate quantum model

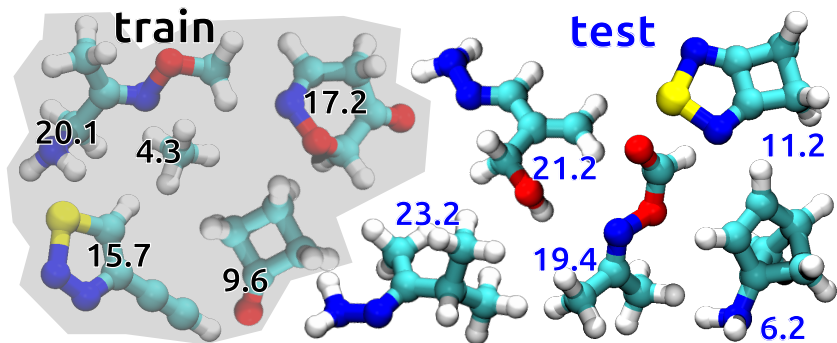
- Machine-learning can be regarded as a sophisticated interpolation between a few known values of the properties
- Can it be made as accurate and general as the Schrödinger equation?
- Kernels are the main ingredient. Think of them as scalar products between structures,  $K(\mathcal{A}, \mathcal{B}) \sim \langle \mathcal{A} | \mathcal{B} \rangle$ .



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

# A universal surrogate quantum model

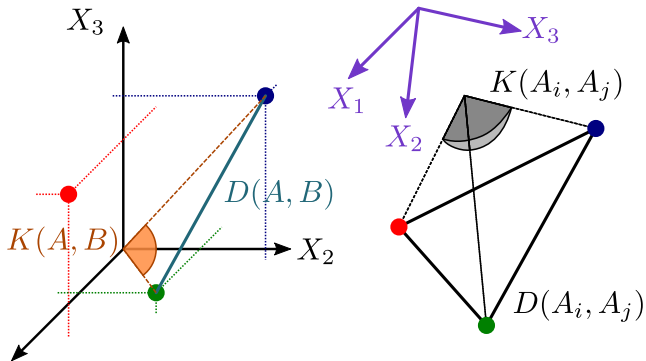
- Machine-learning can be regarded as a sophisticated interpolation between a few known values of the properties
- Can it be made as accurate and general as the Schrödinger equation?
- Kernels are the main ingredient. Think of them as scalar products between structures,  $K(\mathcal{A}, \mathcal{B}) \sim \langle \mathcal{A} | \mathcal{B} \rangle$ .



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

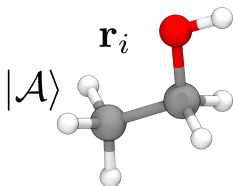
# A universal surrogate quantum model

- 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?
- Kernels are the main ingredient. Think of them as scalar products between structures,  $K(A, B) \sim \langle A|B \rangle$ .



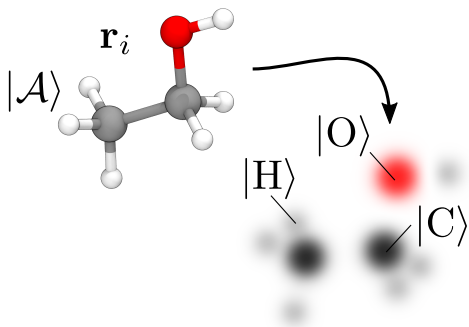
# Symmetry-adapted atom-density representations

- Structural representation based on a decorated atom-density vector  $|\mathcal{A}\rangle$
- Physical symmetries are recovered by integration over group
- Use tensor products to reduce information loss
- $|\mathcal{A}^{(\nu)}\rangle_{\hat{T}}$  leads naturally to atom-centered decomposition
- Rotational average yields  $(\nu + 1)$ -body correlation functions  $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$



# Symmetry-adapted atom-density representations

- Structural representation based on a decorated atom-density vector  $|\mathcal{A}\rangle$
- Physical symmetries are recovered by integration over group
- Use tensor products to reduce information loss
- $|\mathcal{A}^{(\nu)}\rangle_{\hat{T}}$  leads naturally to atom-centered decomposition
- Rotational average yields  $(\nu + 1)$ -body correlation functions  $|\chi^{(\nu)}\rangle_{\hat{R}}$

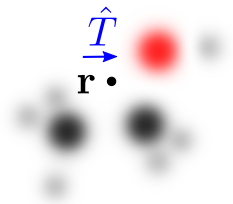


$$\langle \mathbf{r} | \mathcal{A} \rangle = \sum_i g(\mathbf{r} - \mathbf{r}_i) |\alpha_i\rangle$$

Willatt, Musil, **MC**, <https://arxiv.org/pdf/1807.00408>

# Symmetry-adapted atom-density representations

- Structural representation based on a decorated atom-density vector  $|\mathcal{A}\rangle$
- Physical symmetries are recovered by integration over group
- Use tensor products to reduce information loss
- $|\mathcal{A}^{(\nu)}\rangle_{\hat{T}}$  leads naturally to atom-centered decomposition
- Rotational average yields  $(\nu + 1)$ -body correlation functions  $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$

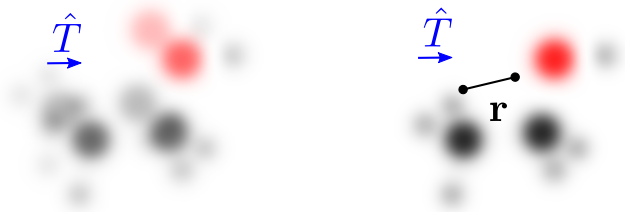


$$\int d\hat{T} \langle \mathbf{r} | \hat{T} | \mathcal{A} \rangle = \sum_i \int d\mathbf{t} g(\mathbf{r} + \mathbf{t} - \mathbf{r}_i) |\alpha_i\rangle = \sum_{\alpha} N_{\alpha} |\alpha\rangle$$

Willatt, Musil, **MC**, <https://arxiv.org/pdf/1807.00408>

# Symmetry-adapted atom-density representations

- Structural representation based on a decorated atom-density vector  $|\mathcal{A}\rangle$
- Physical symmetries are recovered by integration over group
- Use tensor products to reduce information loss
- $|\mathcal{A}^{(\nu)}\rangle_{\hat{T}}$  leads naturally to atom-centered decomposition
- Rotational average yields  $(\nu + 1)$ -body correlation functions  $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$



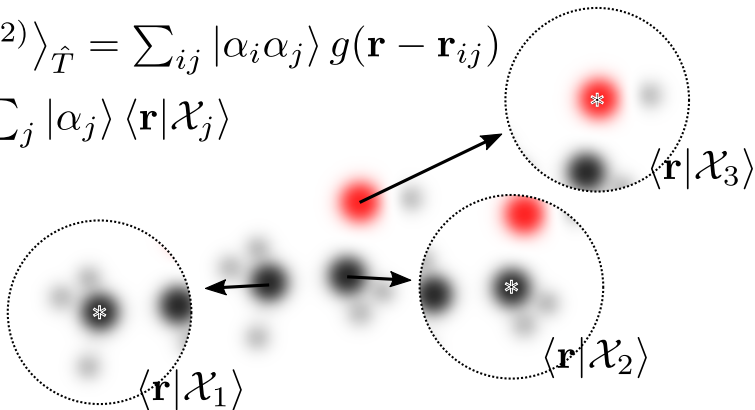
$$\int d\hat{T} \langle \mathbf{r} | \hat{T} | \mathcal{A} \rangle \langle \mathbf{r}' | \hat{T} | \mathcal{A} \rangle = \int d\mathbf{r}' \langle \mathbf{r}' | \mathcal{A} \rangle \langle \mathbf{r}' + \mathbf{r} | \mathcal{A} \rangle$$



# Symmetry-adapted atom-density representations

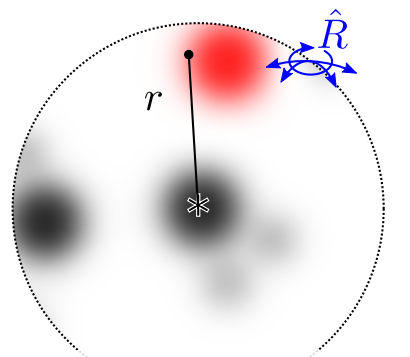
- Structural representation based on a decorated atom-density vector  $|\mathcal{A}\rangle$
- Physical symmetries are recovered by integration over group
- Use tensor products to reduce information loss
- $|\mathcal{A}^{(\nu)}\rangle_{\hat{T}}$  leads naturally to atom-centered decomposition
- Rotational average yields  $(\nu + 1)$ -body correlation functions  $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$

$$\begin{aligned}\langle \mathbf{r} | \mathcal{A}^{(2)} \rangle_{\hat{T}} &= \sum_{ij} |\alpha_i \alpha_j\rangle g(\mathbf{r} - \mathbf{r}_{ij}) \\ &= \sum_j |\alpha_j\rangle \langle \mathbf{r} | \mathcal{X}_j \rangle\end{aligned}$$



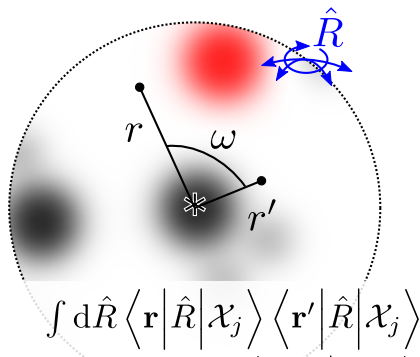
# Symmetry-adapted atom-density representations

- Structural representation based on a decorated atom-density vector  $|\mathcal{A}\rangle$
- Physical symmetries are recovered by integration over group
- Use tensor products to reduce information loss
- $|\mathcal{A}^{(\nu)}\rangle_{\hat{T}}$  leads naturally to atom-centered decomposition
- Rotational average yields  $(\nu + 1)$ -body correlation functions  $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$



The diagram shows a central atom (marked with an asterisk) and a red density cloud. A vector  $r$  points from the atom to the center of the cloud. A blue rotation operator  $\hat{R}$  is shown with arrows indicating rotation around the atom.

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

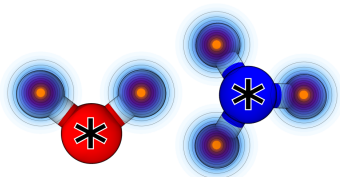


The diagram shows two atoms (marked with asterisks) and two red density clouds. Vectors  $r$  and  $r'$  point from the atoms to the centers of their respective clouds. An angle  $\omega$  is shown between the two vectors. A blue rotation operator  $\hat{R}$  is shown with arrows indicating rotation around the atoms.

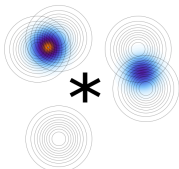
$$\int d\hat{R} \langle \mathbf{r} | \hat{R} | \mathcal{X}_j \rangle \langle \mathbf{r}' | \hat{R} | \mathcal{X}_j \rangle \rightarrow \langle rr' \omega | \mathcal{X}_j^{(2)} \rangle$$

# Symmetry adapted representations & SOAP kernel

- Most of the existing density-based representations and kernels emerge as special cases of this framework
- Not necessary to use position basis. Radial functions and spherical harmonics  $\rightarrow$  SOAP power spectrum and kernel



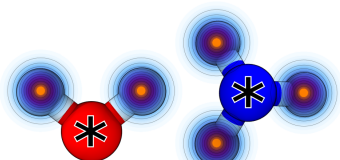
$$\langle \mathbf{r} | \mathcal{X}_j \rangle = \psi(\mathbf{r}) = \sum_i g(\mathbf{r} - \mathbf{r}_{ij})$$



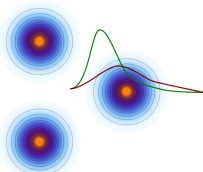
$$k(\mathcal{X}, \mathcal{X}') \equiv \langle \mathcal{X} | \mathcal{X}' \rangle \sim \int \psi(\mathbf{r}) \psi'(\mathbf{r}) d\mathbf{r}$$

# Symmetry adapted representations & SOAP kernel

- Most of the existing density-based representations and kernels emerge as special cases of this framework
- Not necessary to use position basis. Radial functions and spherical harmonics  $\rightarrow$  SOAP power spectrum and kernel



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

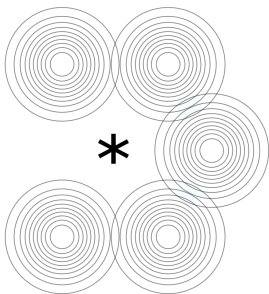


$$\langle nlm | \mathcal{X}_j \rangle = \int d\mathbf{x} \psi(\mathbf{r}) R_n(r) Y_m^l(\hat{\mathbf{r}})$$

Bartók, Kondor, Csányi, PRB (2013)

# Symmetry adapted representations & SOAP kernel

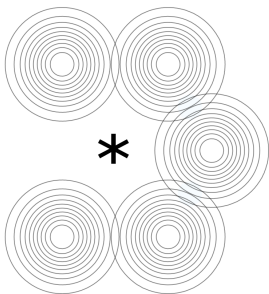
- Most of the existing density-based representations and kernels emerge as special cases of this framework
- Not necessary to use position basis. Radial functions and spherical harmonics  $\rightarrow$  SOAP power spectrum and kernel



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

# Symmetry adapted representations & SOAP kernel

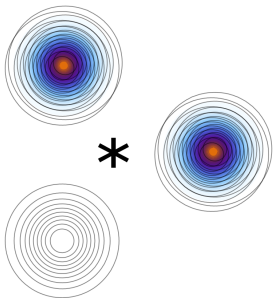
- Most of the existing density-based representations and kernels emerge as special cases of this framework
- Not necessary to use position basis. Radial functions and spherical harmonics  $\rightarrow$  SOAP power spectrum and kernel



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

# Symmetry adapted representations & SOAP kernel

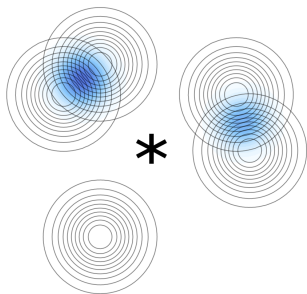
- Most of the existing density-based representations and kernels emerge as special cases of this framework
- Not necessary to use position basis. Radial functions and spherical harmonics  $\rightarrow$  SOAP power spectrum and kernel



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

# Symmetry adapted representations & SOAP kernel

- Most of the existing density-based representations and kernels emerge as special cases of this framework
- Not necessary to use position basis. Radial functions and spherical harmonics  $\rightarrow$  SOAP power spectrum and kernel



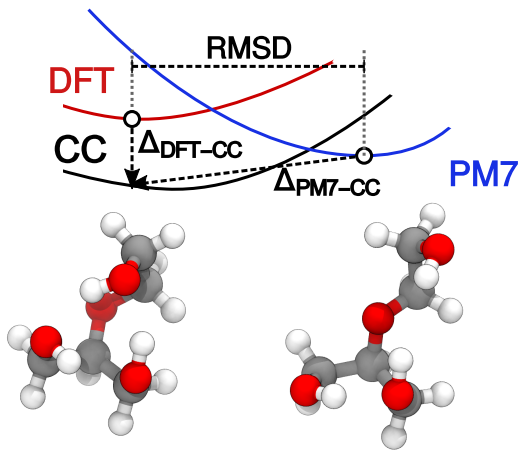
$$\langle \chi^{(\nu)} | \chi'^{(\nu)} \rangle_{\hat{R}} \sim \int d\hat{R} \left| \int \psi(\mathbf{r}) \psi'(\hat{R}\mathbf{r}) d\mathbf{r} \right|^\nu$$



**How well does this work?**

# 100k molecules with coupled-clusters

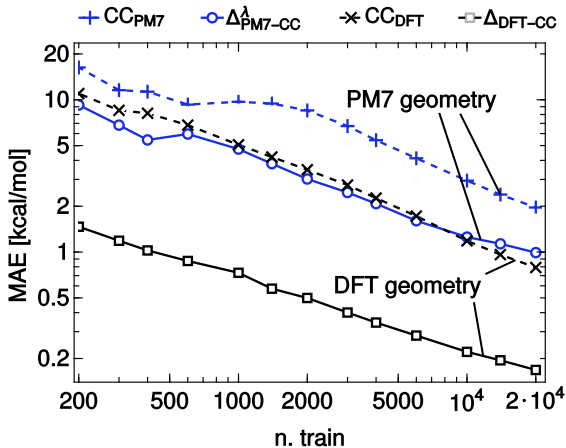
- CCSD(T) Energetics on the QM9 - 114k *useful* predictions based on 20k training calculations
- 1kcal/mol error for predicting CCSD(T) based on PM7 geometries;  
0.18kcal/mol error for predicting CCSD(T) based on DFT geometries!



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

# 100k molecules with coupled-clusters

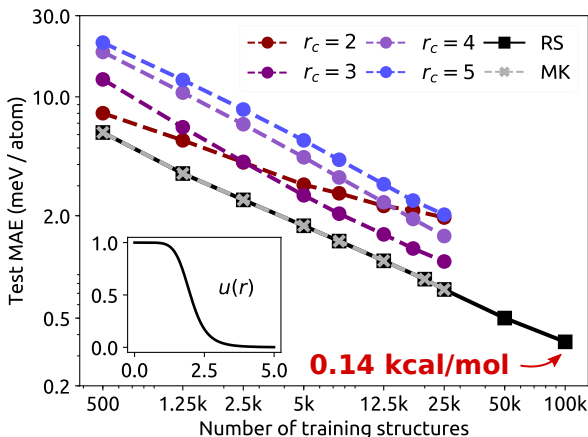
- CCSD(T) Energetics on the QM9 - 114k *useful* predictions based on 20k training calculations
- 1kcal/mol error for predicting CCSD(T) based on PM7 geometries;  
0.18kcal/mol error for predicting CCSD(T) based on DFT geometries!



De, Bartók, Csányi, **MC**, PCCP (2016);  
Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

# 100k molecules with coupled-clusters

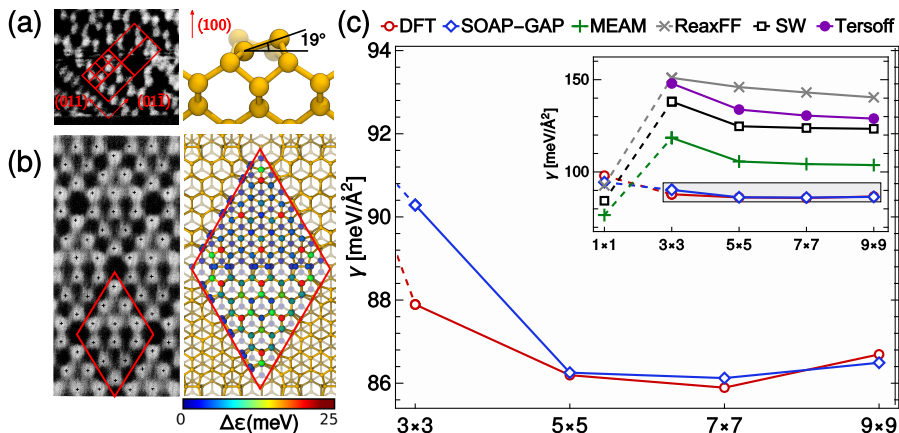
- CCSD(T) Energetics on the QM9 - 114k *useful* predictions based on 20k training calculations
- 1kcal/mol error for predicting CCSD(T) based on PM7 geometries;  
0.18kcal/mol error for predicting CCSD(T) based on DFT geometries!



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

# Silicon surfaces - complexity in a simple material

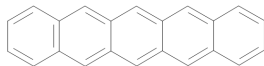
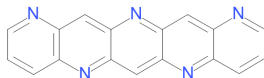
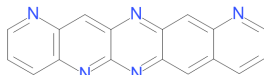
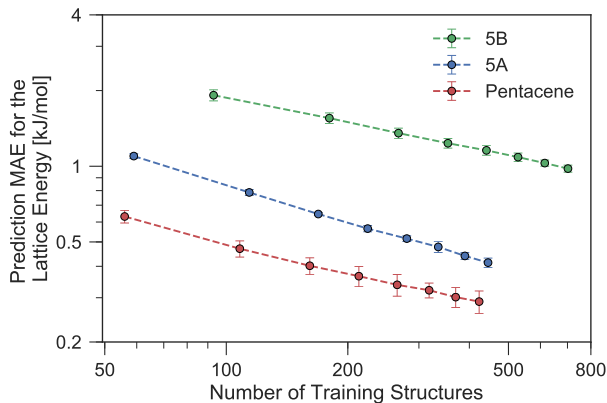
- More than just molecules: a SOAP-GAP model for Si can capture the dimer tilt in Si(100)-2x1, and the delicate energy balance that determines the stability of the Si(111) 7x7 DAS reconstruction



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

# Accurate predictions for molecular crystals

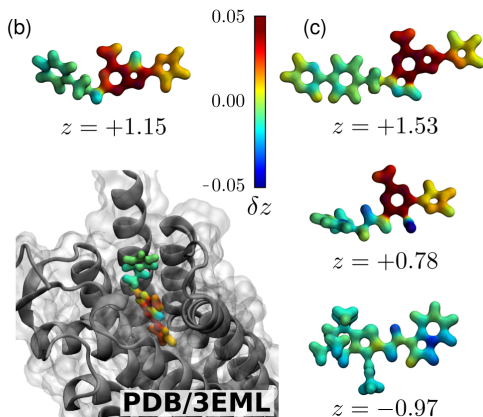
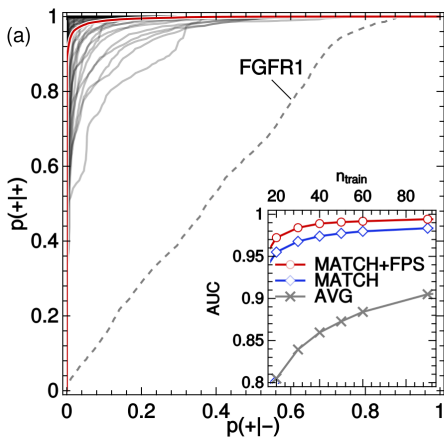
- Substituted pentacenes - model systems for molecular electronics
- Easily achieve sub-kcal/mol accuracy, with REMatch-SOAP kernels



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

# Recognizing active protein ligands

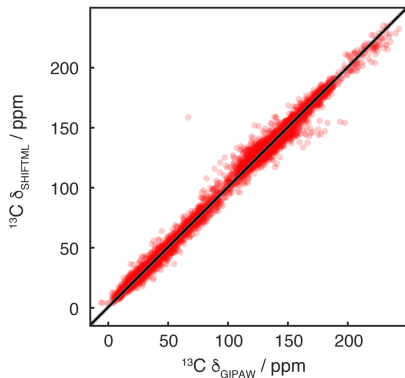
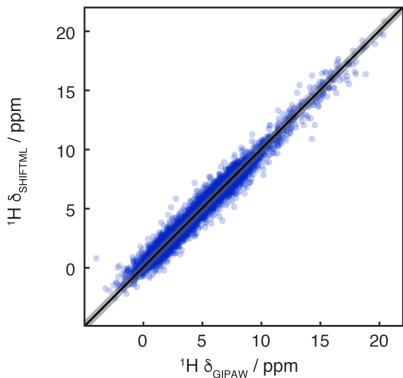
- A SOAP-REMatch-based KSVM classifies active and inactive ligands with 99% accuracy; non-additive model is crucial!
- Sensitivity analysis help identify the active “warhead” and could guide drug design and optimization



Bartok, De, Kermodé, Bernstein, Csanyi, MC, Science Advances (2017)

# 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, O: 18 ppm)
- Accurate enough to do structure determination!

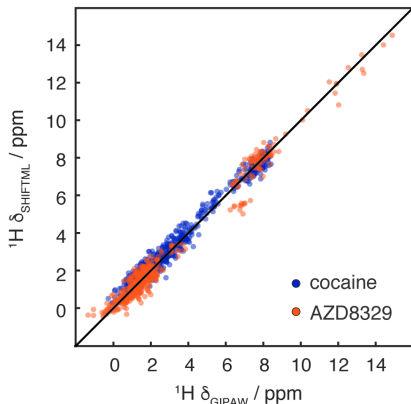
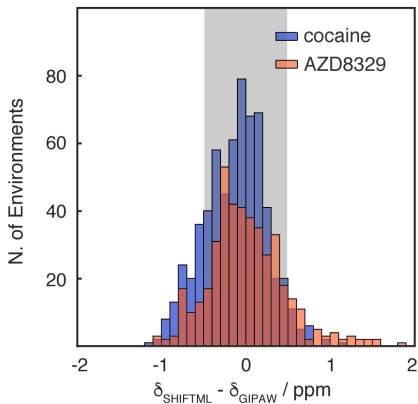


Paruzzo, Hofstetter, Musil, De, **MC**, Emsley, Nature Comm. (2018); <http://shiftml.org>



# More than interatomic potentials

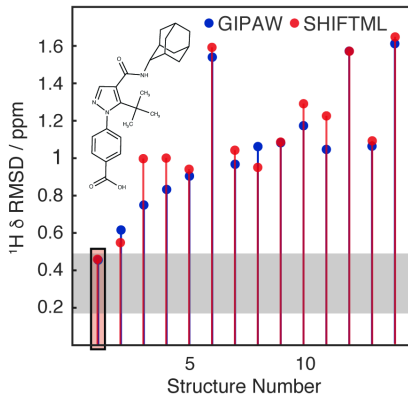
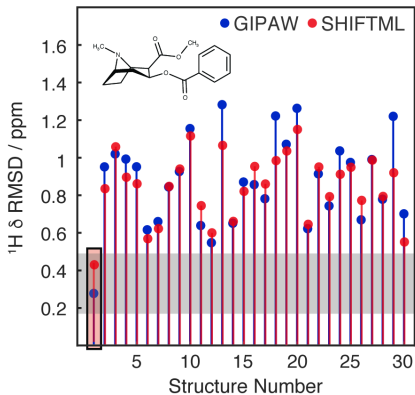
- Solid-state NMR relies on GIPAW-DFT to determine crystal structure of molecular materials
- Train a ML model on 2000 CSD structures, predict chemical shieldings with DFT accuracy (RMSE H: 0.5, C: 5, N: 13, O: 18 ppm)
- Accurate enough to do structure determination!



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, O: 18 ppm)
- Accurate enough to do structure determination!

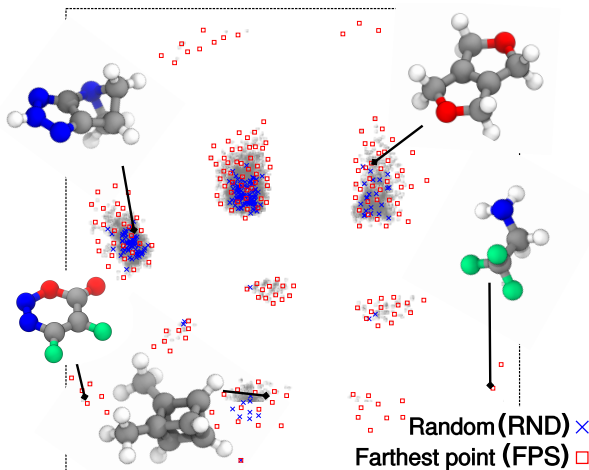


Paruzzo, Hofstetter, Musil, De, **MC**, Emsley, Nature Comm. (2018); <http://shiftml.org>

# **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

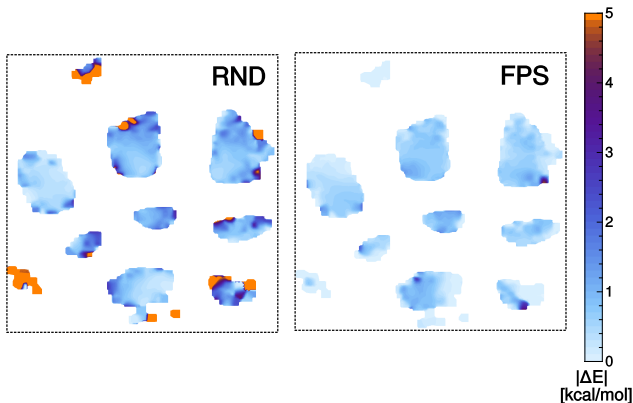


Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

**MC**, Tribello, Parrinello, PNAS (2011); <http://sketchmap.org>  
Physics-Based Machine Learning for Materials and Molecules

# 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

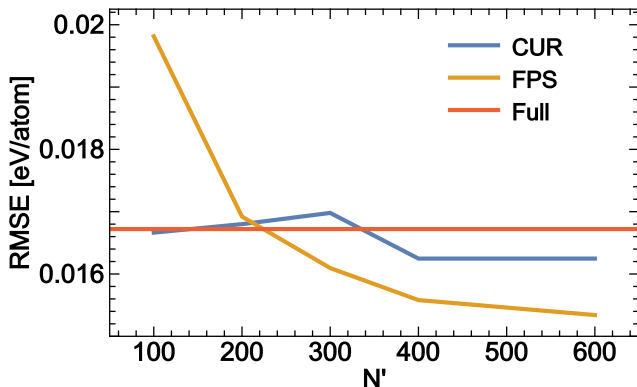


Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

**MC**, Tribello, Parrinello, PNAS (2011); <http://sketchmap.org>  
Physics-Based Machine Learning for Materials and Molecules

# 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



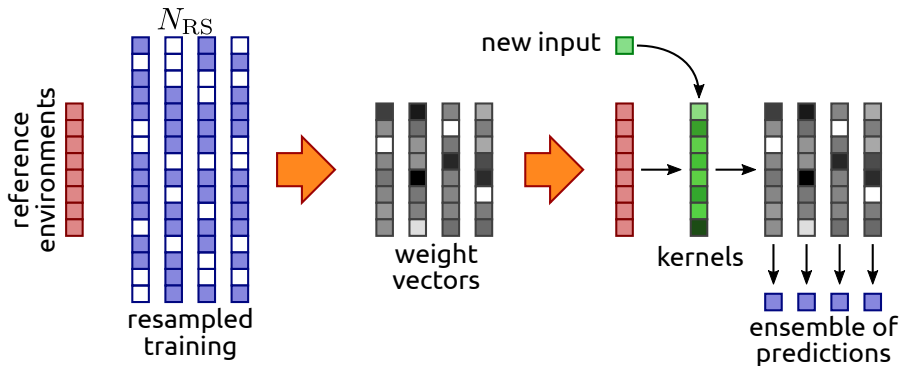
Imbalzano, Anelli, Giofr , Klees, Behler, **MC**, JCP (2018)

# An accurate & inexpensive error estimation

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

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

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

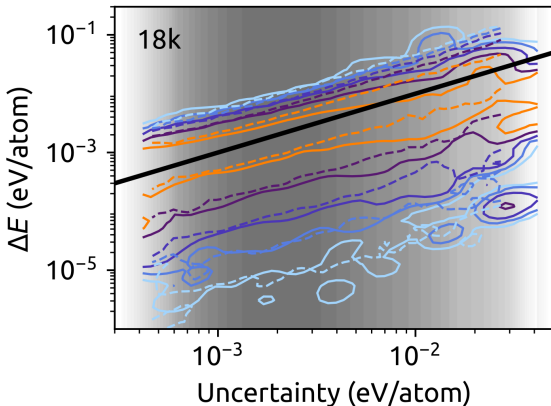


# An accurate & inexpensive error estimation

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

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

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



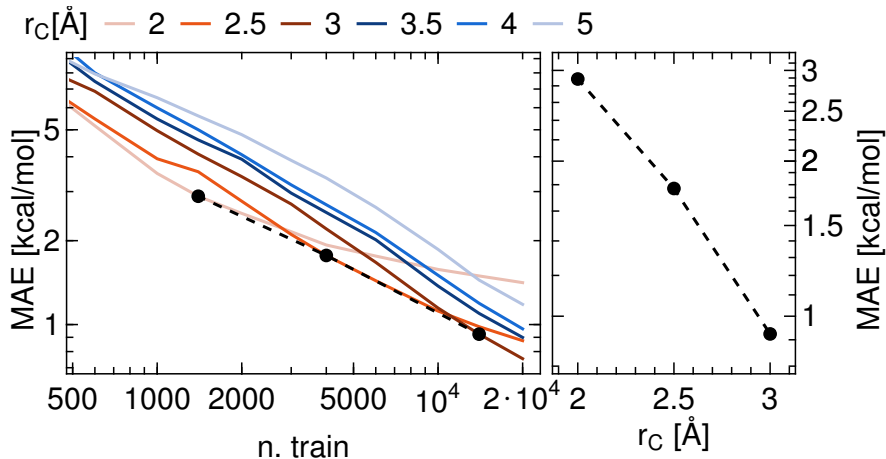
Musil, Willatt, **MC** [arxiv.org/abs/1809.07653](https://arxiv.org/abs/1809.07653)



# Insights from machine learning

# Understanding the range of interactions

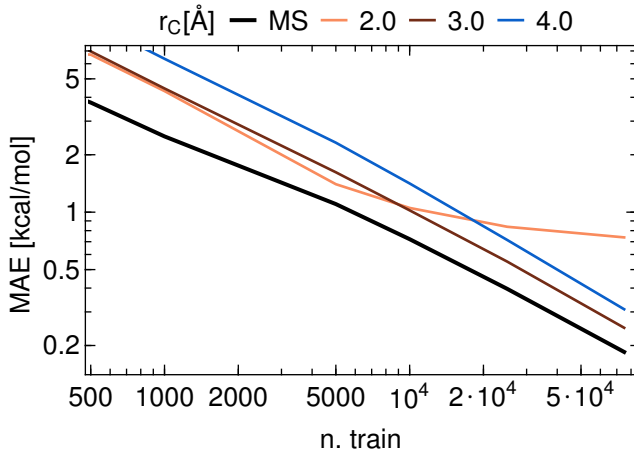
- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel  $K(A, B) = \sum_i w_i K_i(A, B)$  yields the best of all worlds - chemical accuracy on QM9 with  $\sim 5000$  train structures



Bartók, De, Kermode, Bernstein, Csányi, **MC**, Science Advances (2017)

# Understanding the range of interactions

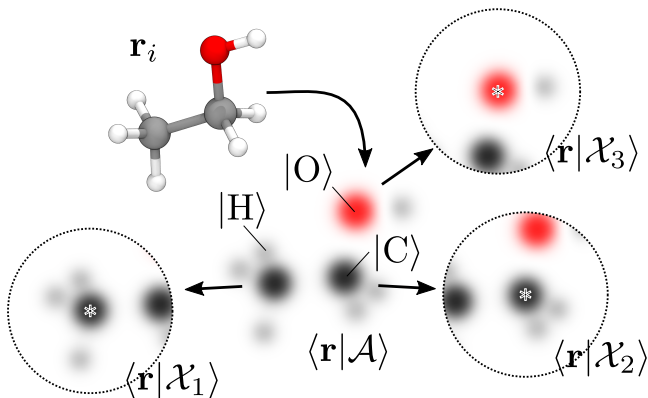
- Environment kernels can be built for different cutoff radii
- Dimensionality/accuracy tradeoff, a measure of the range of interactions
- A multi-scale kernel  $K(A, B) = \sum_i w_i K_i(A, B)$  yields the best of all worlds - chemical accuracy on QM9 with  $\sim 5000$  train structures



Bartók, De, Kermode, Bernstein, Csányi, **MC**, Science Advances (2017)

# A data-driven periodic table of the elements

- How to learn with multiple species? Decorate atomic Gaussian with elemental kets  $|H\rangle, |O\rangle, \dots$
- 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!



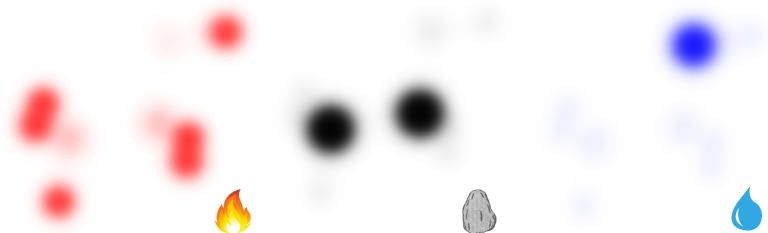
# A data-driven periodic table of the elements

- How to learn with multiple species? Decorate atomic Gaussian with elemental kets  $|H\rangle, |O\rangle, \dots$
- 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!

$$|H\rangle = 0.5 \left| \text{🔥} \right\rangle + 0.1 \left| \text{🪨} \right\rangle + 0.2 \left| \text{💧} \right\rangle$$

$$|C\rangle = 0.2 \left| \text{🔥} \right\rangle + 0.8 \left| \text{🪨} \right\rangle + 0.3 \left| \text{💧} \right\rangle$$

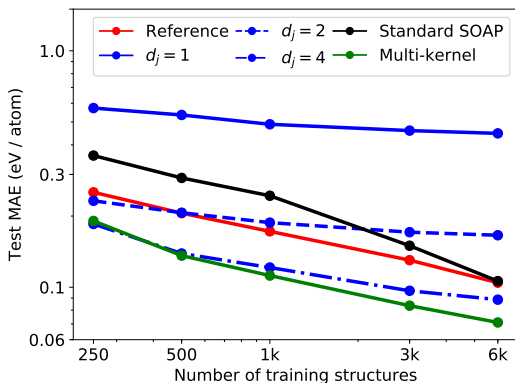
$$|O\rangle = 0.1 \left| \text{🔥} \right\rangle + 0.1 \left| \text{🪨} \right\rangle + 0.6 \left| \text{💧} \right\rangle$$



Empedocles et al. (ca 360BC). Metaphor courtesy of Albert Bartók

# A data-driven periodic table of the elements

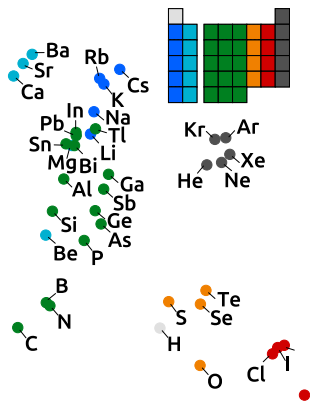
- How to learn with multiple species? Decorate atomic Gaussian with elemental kets  $|H\rangle, |O\rangle, \dots$
- Expand each ket in a finite basis,  $|\alpha\rangle = \sum_J u_{\alpha J} |J\rangle$ . Optimize coefficients
- Dramatic reduction of the descriptor space, more effective learning . . .
- . . . and as by-product get a data-driven version of the periodic table!



Elpasolite dataset. Reference curve (red) from Faber et al. JCP (2018)

# A data-driven periodic table of the elements

- How to learn with multiple species? Decorate atomic Gaussian with elemental kets  $|H\rangle, |O\rangle, \dots$
- 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!



H						He	
Li	Be	B	C	N	O	F	Ne
Na	Mg	Al	Si	P	S	Cl	Ar
K	Ca	Ga	Ge	As	Se	Br	Kr
Rb	Sr	In	Sn	Sb	Te	I	Xe
Cs	Ba	Tl	Pb	Bi			

# 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

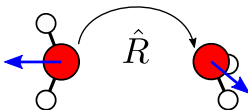
$$k(\mathcal{X}, \mathcal{X}') \leftrightarrow \langle \mathbf{y}(\mathcal{X}); \mathbf{y}(\mathcal{X}') \rangle, \text{ so } k(\hat{S}\mathcal{X}, \hat{S}'\mathcal{X}') \leftrightarrow \langle \mathbf{y}(\hat{S}\mathcal{X}); \mathbf{y}(\hat{S}'\mathcal{X}') \rangle$$

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

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

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

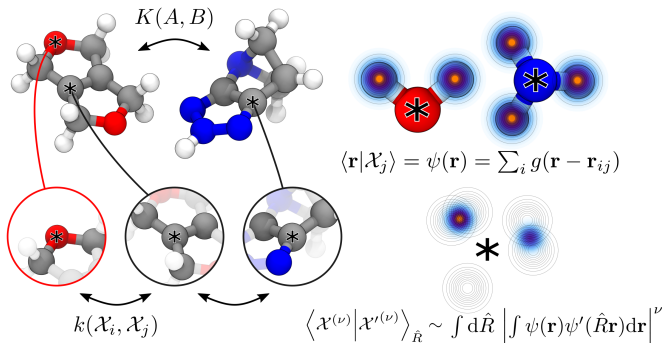
$$k_{\mu\nu}(\mathcal{X}, \mathcal{X}') \leftrightarrow \langle T_{\mu}(\mathcal{X}); T_{\nu}(\mathcal{X}') \rangle \rightarrow k_{\mu\nu}(\hat{R}\mathcal{X}, \hat{R}'\mathcal{X}') = R_{\mu\mu'} k_{\mu'\nu'}(\mathcal{X}, \mathcal{X}') R'_{\nu\nu'}$$



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  $T^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities



$$k(\mathcal{X}, \mathcal{X}') = \int d\hat{R} \kappa(\mathcal{X}, \hat{R}\mathcal{X}'),$$

$$\kappa(\mathcal{X}, \mathcal{X}') = \left| \int \psi_{\mathcal{X}}(\mathbf{x}) \psi_{\mathcal{X}'}(\mathbf{x}) d\mathbf{x} \right|^2$$

# $\lambda$ -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  $T^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities

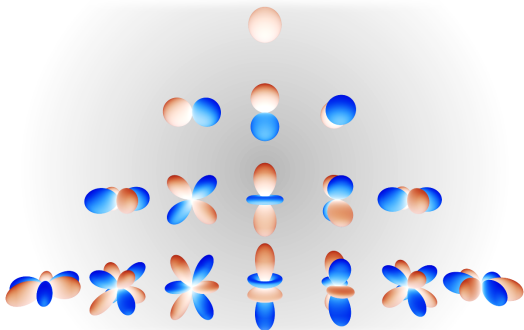


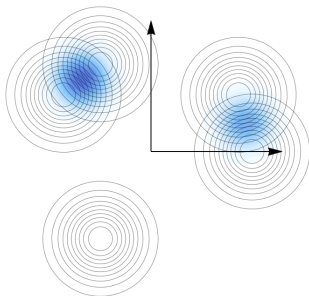
image fro Wikipedia

$$T_\mu^\lambda \left( \hat{R}(\mathcal{X}) \right) = D_{\mu\mu'}^\lambda \left( \hat{R} \right) T_{\mu'}^\lambda (\mathcal{X})$$

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

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

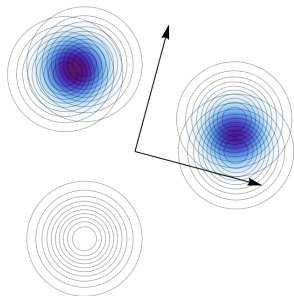
- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components  $\mathbf{T}^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities



$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} D_{\mu\nu}^\lambda(\hat{R}) \kappa(\mathcal{X}, \hat{R}\mathcal{X}')$$

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

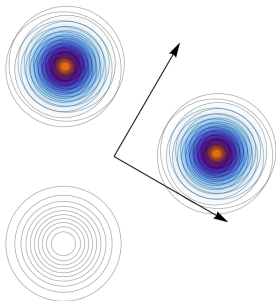
- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components  $\mathbf{T}^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities



$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} D_{\mu\nu}^\lambda(\hat{R}) \kappa(\mathcal{X}, \hat{R}\mathcal{X}')$$

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

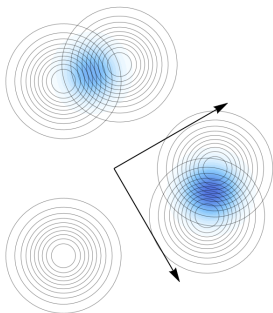
- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components  $\mathbf{T}^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities



$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} D_{\mu\nu}^\lambda(\hat{R}) \kappa(\mathcal{X}, \hat{R}\mathcal{X}')$$

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

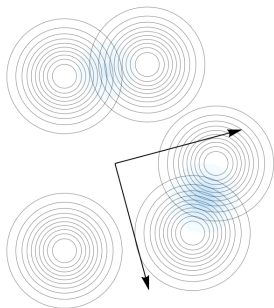
- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components  $\mathbf{T}^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities



$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} D_{\mu\nu}^\lambda(\hat{R}) \kappa(\mathcal{X}, \hat{R}\mathcal{X}')$$

# $\lambda$ -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  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities

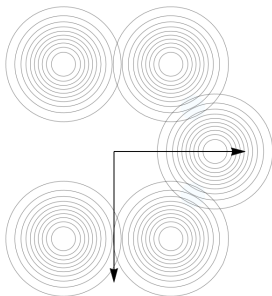


$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} D_{\mu\nu}^\lambda(\hat{R}) \kappa(\mathcal{X}, \hat{R}\mathcal{X}')$$



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

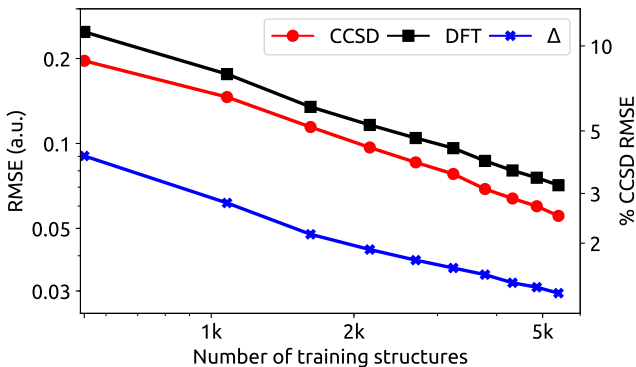
- Recall the definition of SOAP, based on the atom-density overlap
- Each tensor can be decomposed into irreducible spherical components  $\mathbf{T}^\lambda$ , corresponding to the representations of  $SO(3)$
- A hierarchy of  $\lambda$ -SOAP kernels can be defined to learn tensorial quantities



$$k_{\mu\nu}^\lambda(\mathcal{X}, \mathcal{X}') = \int d\hat{R} D_{\mu\nu}^\lambda(\hat{R}) \kappa(\mathcal{X}, \hat{R}\mathcal{X}')$$

# Molecular polarizabilities at the CCSD level

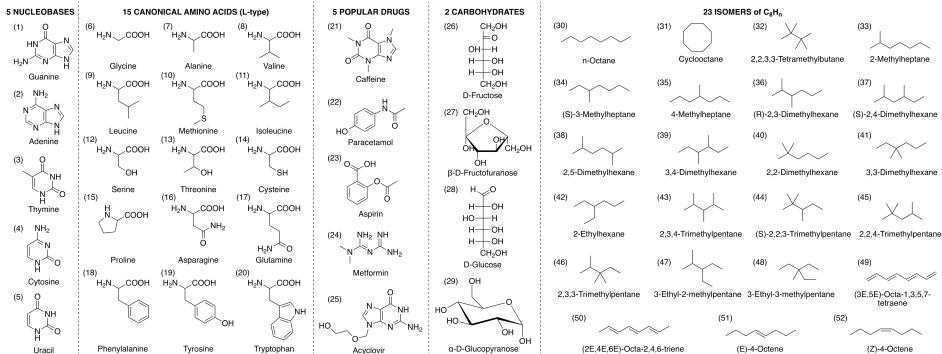
- DFT is not very accurate for the dielectric response. Train a ML model on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir  $C_8H_{11}N_5O_3$ ) with better-than-DFT accuracy
- Atom-centered environment decomposition of  $\alpha$  and the DFT error



Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, [arxiv.org/abs/1809.05349](https://arxiv.org/abs/1809.05349)

# Molecular polarizabilities at the CCSD level

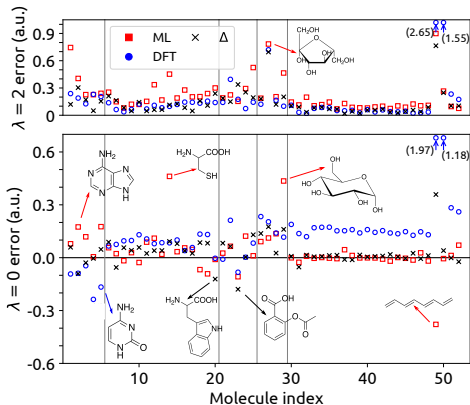
- DFT is not very accurate for the dielectric response. Train a ML model on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir  $C_8H_{11}N_5O_3$ ) with better-than-DFT accuracy
- Atom-centered environment decomposition of  $\alpha$  and the DFT error



Wilkins, Grisafi, Yang, Lao, DiStasio, MC, arxiv.org/abs/1809.05349

# Molecular polarizabilities at the CCSD level

- DFT is not very accurate for the dielectric response. Train a ML model 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  $\alpha$  and the DFT error

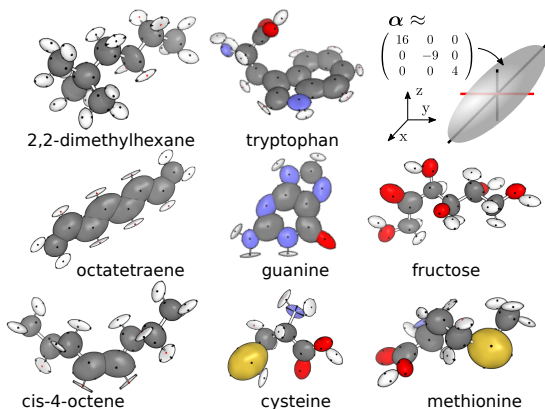


Method	RMSE
CCSD/DFT	0.573
CCSD/ML	0.304
DFT/ML	0.403
$\Delta(\text{CCSD-DFT})/\text{ML}$	0.212

Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, [arxiv.org/abs/1809.05349](https://arxiv.org/abs/1809.05349)

# Molecular polarizabilities at the CCSD level

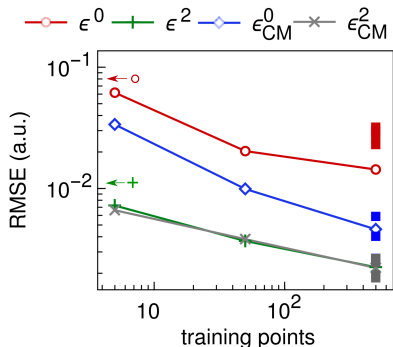
- DFT is not very accurate for the dielectric response. Train a ML model on the QM7 dataset with CCSD accuracy
- The model can extrapolate to much large compounds (up to aciclovir  $C_8H_{11}N_5O_3$ ) with better-than-DFT accuracy
- Atom-centered environment decomposition of  $\alpha$  and the DFT error



Wilkins, Grisafi, Yang, Lao, DiStasio, **MC**, [arxiv.org/abs/1809.05349](https://arxiv.org/abs/1809.05349)

# Learning the dielectric response of water

- The SA-GPR framework, and the  $\lambda$ -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



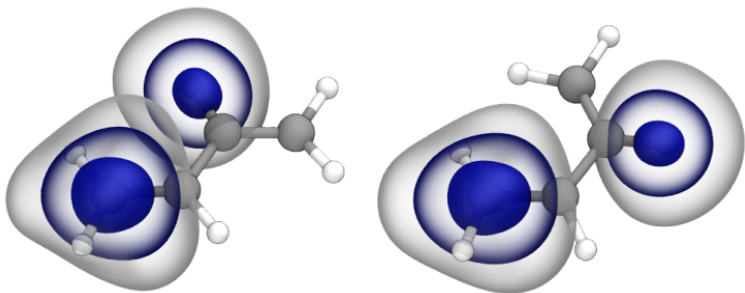
$$\text{Clausius-Mossotti: } \alpha = (\epsilon - 1)(\epsilon + 2)^{-1} V$$

# 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.

$$\mathcal{F}(\rho) = \int d\mathbf{r} \left| \sum_{ik} c_{ik} \phi_k(\mathbf{r} - \mathbf{r}_i) \right|^2 + \eta |\mathbf{c}|^2, \quad c_{inlm} = \sum_{jm'} x_{jnml} k_{mm'}^l(x_i, x_j)$$

- Machine-learn directly the full density (non-orthogonal basis is tricky!)
- Highly transferable: learn on C4, predict on C8



Marzari, Vanderbilt, PRB 1997

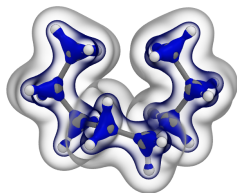
# A transferable model of the electron density

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

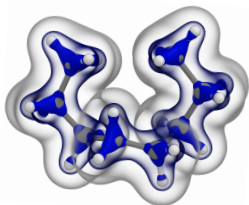
$$\mathcal{F}(\rho) = \int d\mathbf{r} \left| \sum_{ik} c_{ik} \phi_k(\mathbf{r} - \mathbf{r}_i) \right|^2 + \eta |\mathbf{c}|^2, \quad c_{inlm} = \sum_{jm'} x_{jnml} k_{mm'}^l(x_i, x_j)$$

- Machine-learn directly the full density (non-orthogonal basis is tricky!)
- Highly transferable: learn on C4, predict on C8

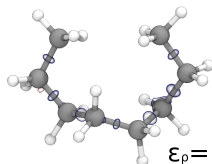
QM



ML



ML - QM



$$\epsilon_\rho = 1.41\%$$

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

Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, [arxiv.org/abs/1809.05349](https://arxiv.org/abs/1809.05349)



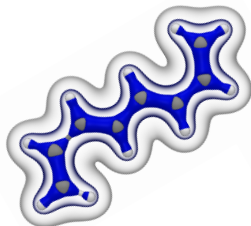
# A transferable model of the electron density

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

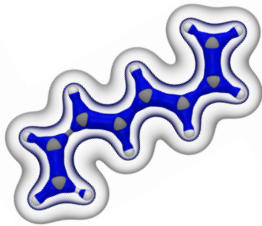
$$\mathcal{F}(\rho) = \int d\mathbf{r} \left| \sum_{ik} c_{ik} \phi_k(\mathbf{r} - \mathbf{r}_i) \right|^2 + \eta |\mathbf{c}|^2, \quad c_{inlm} = \sum_{jm'} x_{jnml} k_{mm'}^l(x_i, x_j)$$

- Machine-learn directly the full density (non-orthogonal basis is tricky!)
- Highly transferable: learn on C4, predict on C8

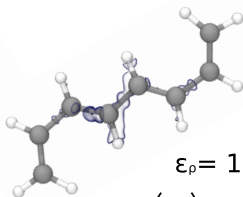
QM



ML



ML - QM

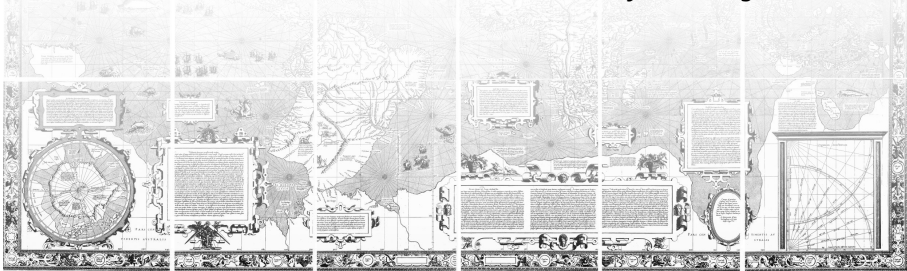


$$\varepsilon_\rho = 1.81\% \\ \langle \varepsilon_\rho \rangle = 1.83\%$$

Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, [arxiv.org/abs/1809.05349](https://arxiv.org/abs/1809.05349)

# My wishlist - revisited

- General applicability: suitable for all systems and all types of properties
  - “Nearsightedness” of electronic matter  $\leftrightarrow$  **local** environment decomposition
  - Excellent performance on benchmark DBs, accurate & cheap **error estimate**
  - Predict CCSD from PM7, potentials for solids, 99% prediction of drug activity, silicon & molecular crystals, NMR shieldings in solids
  - Huge potential of a **SA-GPR framework to learn tensors** - electric multipoles and response, but also densities, Hamiltonians, . . .
- Not only a fancy interpolator: use ML to gain **insights** and understanding
  - Structure-energy-property maps based on the kernel distance
  - Understand the nature of chemical interactions by dissecting the ML model



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