Machine-Learning Like a Physicist

Michele Ceriotti
https://cosmo.epfl.ch

AI-4-MATER
September 2018, ICMPE

http://cosmo.epfl.ch

Follow @COSMO_EPFL

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

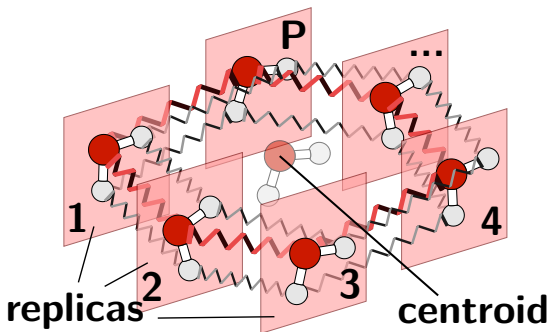# Bringing Fluctuations into Materials Modelling

- Quantum nuclei with path integral methods
- Anharmonic free energies in solids
- Activated events and phase transitions
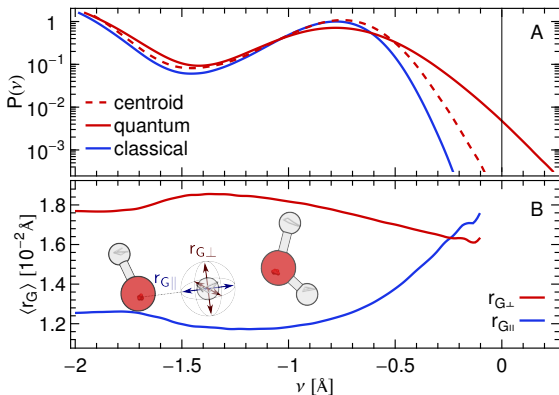
Li, Walker, Michaelides, PNAS (2011); **MC** et al., PNAS (2013); **MC** et al., Chem. Rev. (2016)
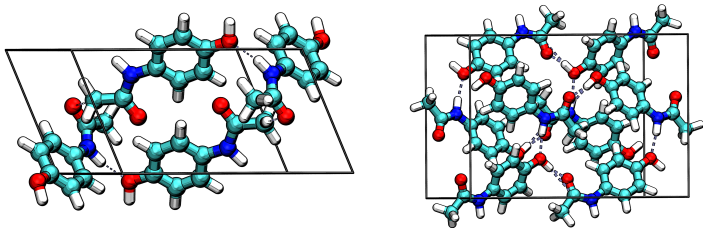
Bringing Fluctuations into Materials Modelling

- Quantum nuclei with path integral methods
- Anharmonic free energies in solids
- Activated events and phase transitions

Markland & **MC**, Nat. Rev. Chem. (2018); http://ipi-code.org

Bringing Fluctuations into Materials Modelling

- Quantum nuclei with path integral methods
- Anharmonic free energies in solids
- Activated events and phase transitions

Binding Energy (meV/mol)	fI	fII	$\Delta\Delta$
Lattice Energy (PBE+D3)	-1492	-1489	-3
Lattice Energy (PBE0+D3)	-1271	-1271	0
Classical Harmonic	-1500	-1487	-13
Quantum Anharmonic	**-1152**	**-1107**	**-45**

Bringing Fluctuations into Materials Modelling

- Quantum nuclei with path integral methods
- Anharmonic free energies in solids
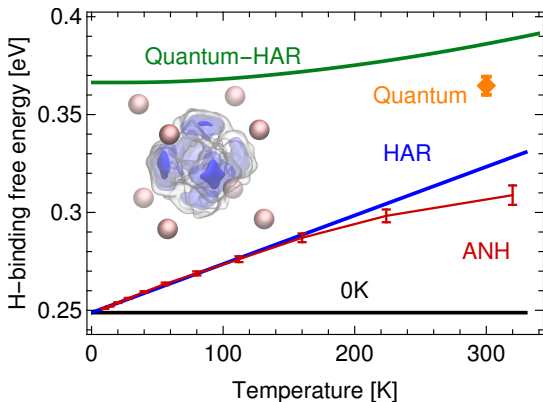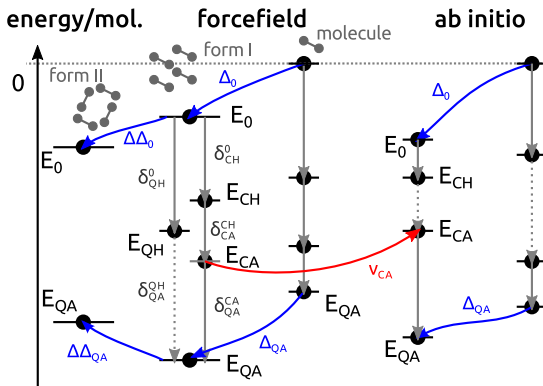- Activated events and phase transitions

Cheng, Paxton, **MC**, Phys. Rev. Lett. (2018)

# Bringing Fluctuations into Materials Modelling
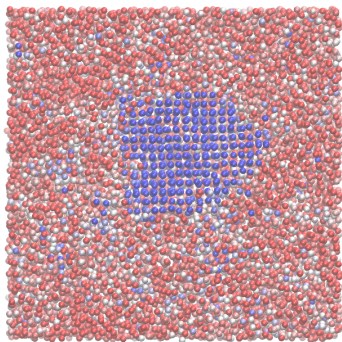
- Quantum nuclei with path integral methods
- Anharmonic free energies in solids
- Activated events and phase transitions

Rossi, Gasparotto & **MC**, Phys. Rev. Lett. (2016)

Bringing Fluctuations into Materials Modelling

- Quantum nuclei with path integral methods
- Anharmonic free energies in solids
- Activated events and phase transitions
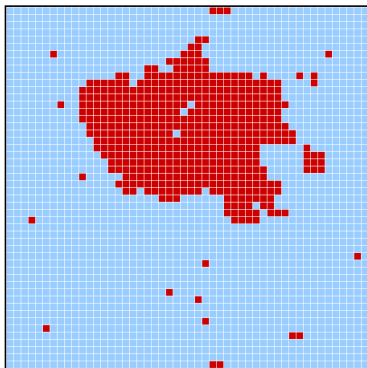
Cheng & **MC**, J. Chem. Phys. (2017)

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
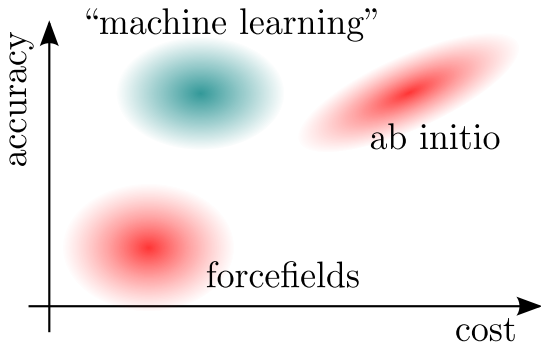- Use machine learning to get around these limitations
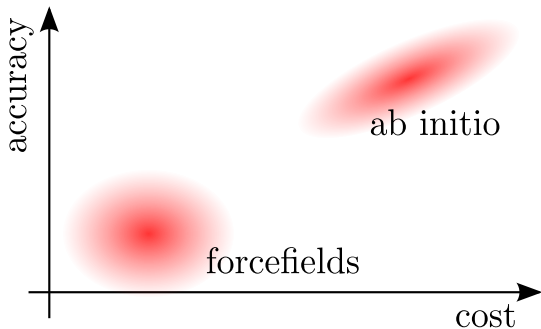
Why Machine Learning?

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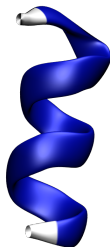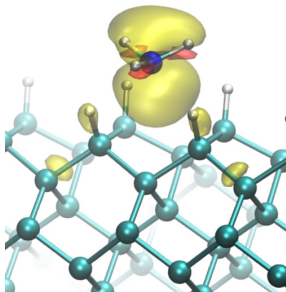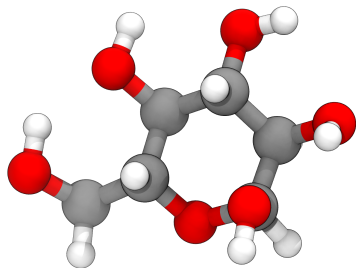
$$\hat{\mathcal{H}}\Psi = E\Psi \quad E(\mathbf{q}) = \sum_{ij} v(r_{ij}) + \dots, \quad E(\mathbf{q}) = ML(\mathbf{q} | \{\mathbf{q}_i, V_i\})$$

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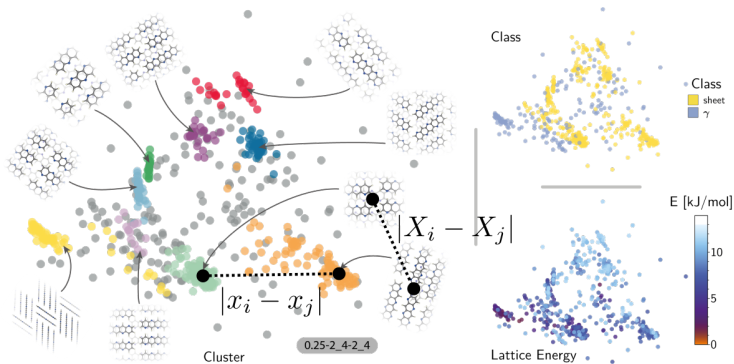
$$\{X_i\} \Rightarrow \{x_i\} \qquad \{x_i\} = \operatorname{argmin} \sum_{ij} \left[s\left(|X_i - X_j|\right) - s\left(|x_i - x_j|\right) \right]^2$$

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, $K(\mathcal{A}, \mathcal{B}) \sim \langle \mathcal{A} | \mathcal{B} \rangle$.
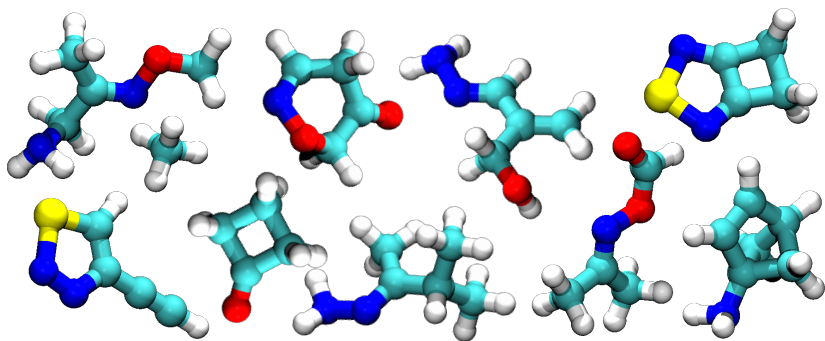
A Universal Surrogate Quantum Model

- 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, $K(\mathcal{A}, \mathcal{B}) \sim \langle \mathcal{A} | \mathcal{B} \rangle$.

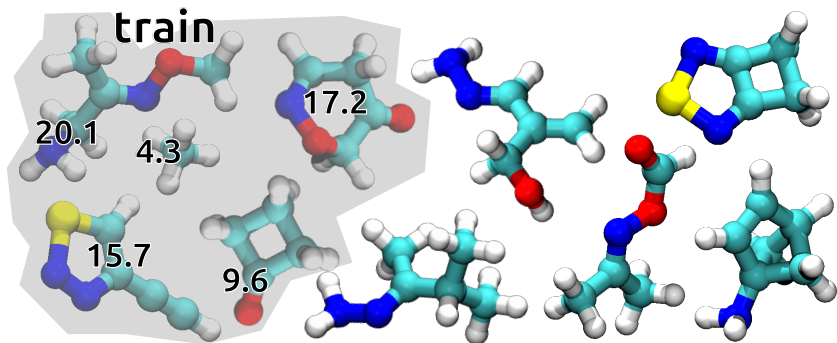$$E(\mathcal{A}_j) = \sum_i w_i K(\mathcal{A}_j, \mathcal{A}_i)$$

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, $K(\mathcal{A}, \mathcal{B}) \sim \langle \mathcal{A} | \mathcal{B} \rangle$.

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

Machine-Learning Like a Physicist
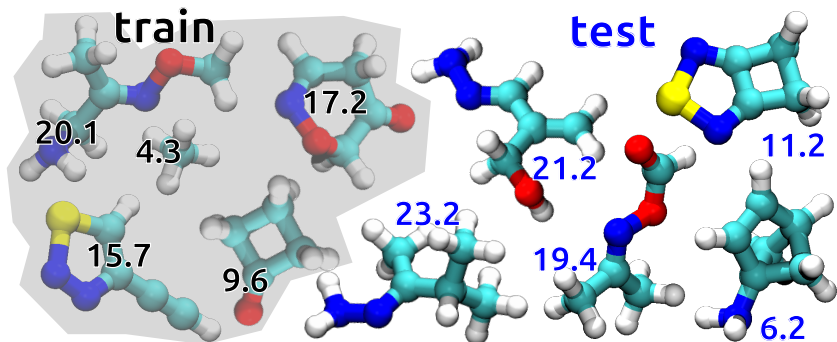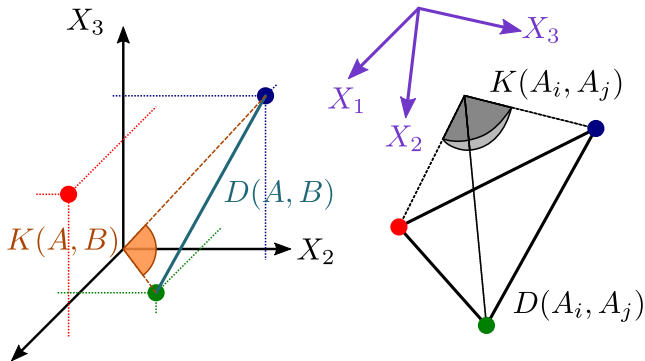
A Universal Surrogate Quantum Model

- 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, $K(\mathcal{A}, \mathcal{B}) \sim \langle \mathcal{A} | \mathcal{B} \rangle$.

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
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition
- Rotational average yields $(\nu + 1)$-body correlation functions $\left|\mathcal{X}^{(\nu)}\right\rangle_{\hat{R}}$

Willatt, Musil, **MC**, https://arxiv.org/pdf/1807.00408
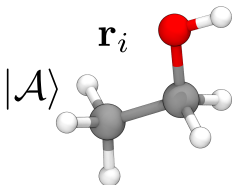
Symmetry-Adapted Atom-Density Representations

- 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
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition
- Rotational average yields $(\nu + 1)$-body correlation functions $\left|\mathcal{X}^{(\nu)}\right\rangle_{\hat{R}}$

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

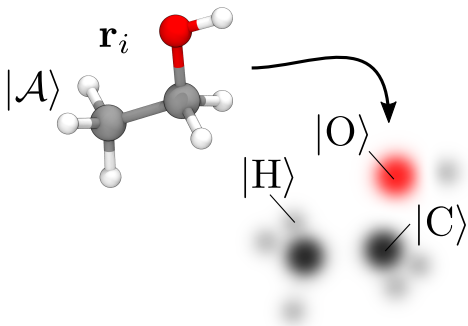Willatt, Musil, **MC**, https://arxiv.org/pdf/1807.00408

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
- $\left|\mathcal{A}^{(\nu)}\right\rangle_{\hat{T}}$ leads naturally to atom-centered decomposition
- Rotational average yields $(\nu + 1)$-body correlation functions $\left|\mathcal{X}^{(\nu)}\right\rangle_{\hat{R}}$

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

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
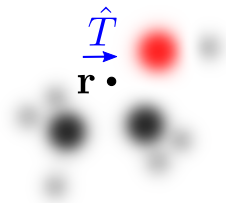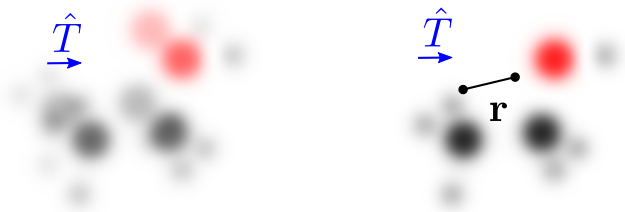- Rotational average yields $(\nu + 1)$-body correlation functions $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$

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

Willatt, Musil, **MC**, 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
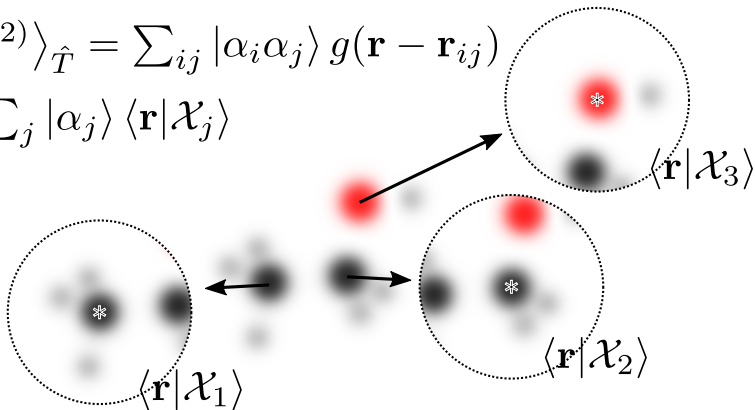- Rotational average yields $(\nu + 1)$-body correlation functions $|\mathcal{X}^{(\nu)}\rangle_{\hat{R}}$

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

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

$\langle \mathbf{r} | \mathcal{X}_3 \rangle$

$\langle \mathbf{r} | \mathcal{X}_1 \rangle$

$\langle \mathbf{r} | \mathcal{X}_2 \rangle$

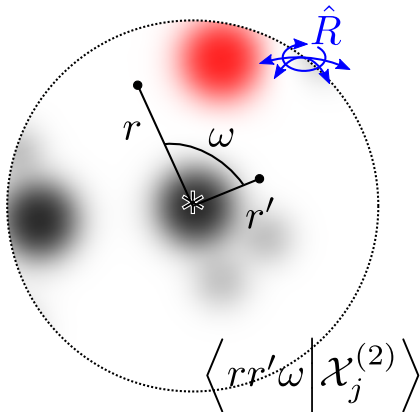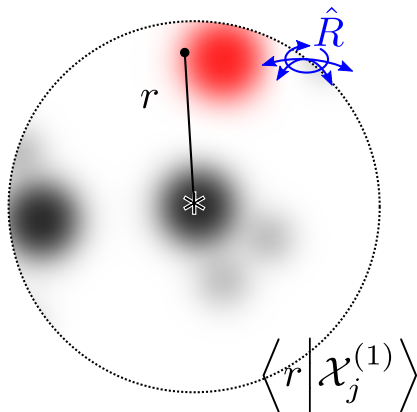# Symmetry-Adapted Atom-Density Representations

- 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 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 → SOAP power spectrum and kernel
- Other strategies to combine local kernels (entropy-regularized match)
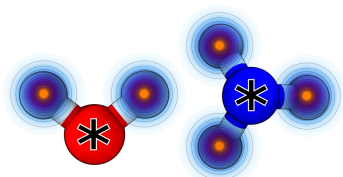
$$\langle \mathbf{r} | \mathcal{X}_j \rangle = \psi(\mathbf{r}) = \sum_i g(\mathbf{r} - \mathbf{r}_{ij})$$

$$k(\mathcal{X}, \mathcal{X}') \equiv \langle \mathcal{X} | \mathcal{X}' \rangle \sim \int \psi(\mathbf{r}) \psi'(\mathbf{r}) \mathrm{d}\mathbf{r}$$
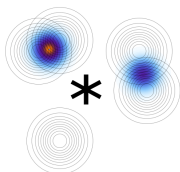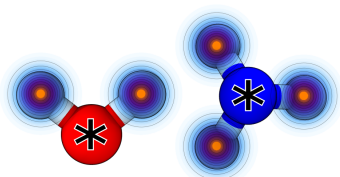
Symmetry Adapted Representations & SOAP Kernel

- 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
- Other strategies to combine local kernels (entropy-regularized match)

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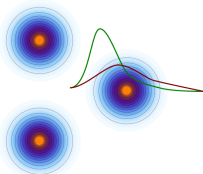
$$\langle nlm | \mathcal{X}_j \rangle = \int d\mathbf{x}\, \psi(\mathbf{r}) R_n(r) Y_m^l(\hat{\mathbf{r}})$$

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 → SOAP power spectrum and kernel
- Other strategies to combine local kernels (entropy-regularized match)

$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim$$

$$\int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r})\psi'(\hat{R}\mathbf{r})\mathrm{d}\mathbf{r} \right|^{\nu}$$

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
- Other strategies to combine local kernels (entropy-regularized match)
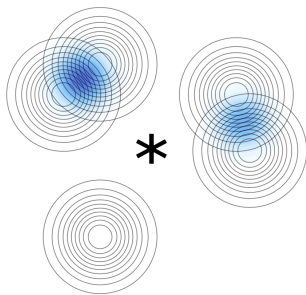

$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim$$

$$\int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r})\psi'(\hat{R}\mathbf{r})\mathrm{d}\mathbf{r} \right|^{\nu}$$
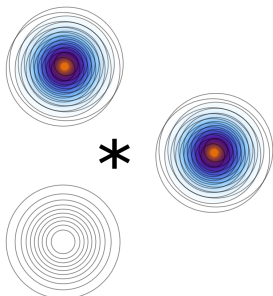
Symmetry Adapted Representations & SOAP Kernel

- 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
- Other strategies to combine local kernels (entropy-regularized match)

$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim$$

$$\int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r})\psi'(\hat{R}\mathbf{r})\mathrm{d}\mathbf{r} \right|^{\nu}$$
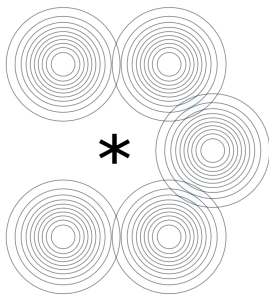
Symmetry Adapted Representations & SOAP Kernel

- 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
- Other strategies to combine local kernels (entropy-regularized match)

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

Bartók, Kondor, Csányi, PRB (2013)

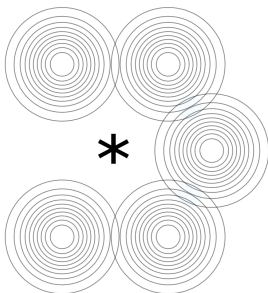# Symmetry Adapted Representations & SOAP Kernel

- 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
- Other strategies to combine local kernels (entropy-regularized match)

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

$$P_{ij} = \frac{1}{N_A N_B}$$

De, Bartók, Csányi, **MC**, PCCP (2016); M. Cuturi, NIPS (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 → SOAP power spectrum and kernel
- Other strategies to combine local kernels (entropy-regularized match)

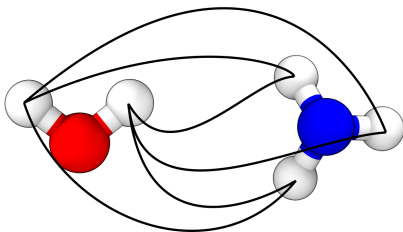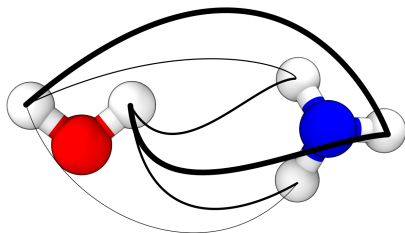$$K(A, B) = \sum_{i \in A, j \in B} P_{ij} k(\mathcal{X}_i, \mathcal{X}_j)$$

$$\hat{K}^\gamma(A, B) \propto$$
$$\max_{\mathbf{P} \in \mathcal{U}} \sum_{ij} P_{ji}(C_{ij}^{A,B} - \gamma \ln P_{ji})$$

De, Bartók, Csányi, **MC**, PCCP (2016); M. Cuturi, NIPS (2013);

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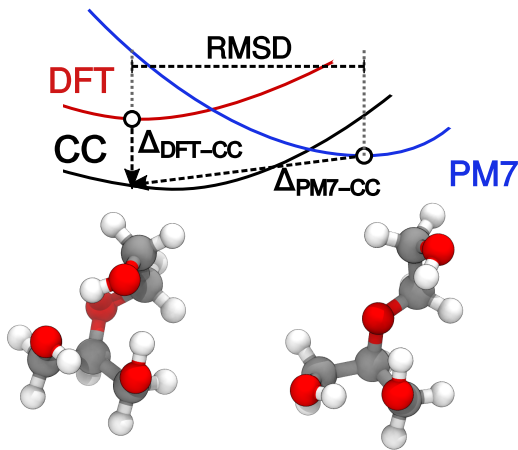
Ramakrishnan et al., Scientific Data (2014); Ramakrishnan et al., JCTC (2015)

100k Molecules with Coupled-Clusters

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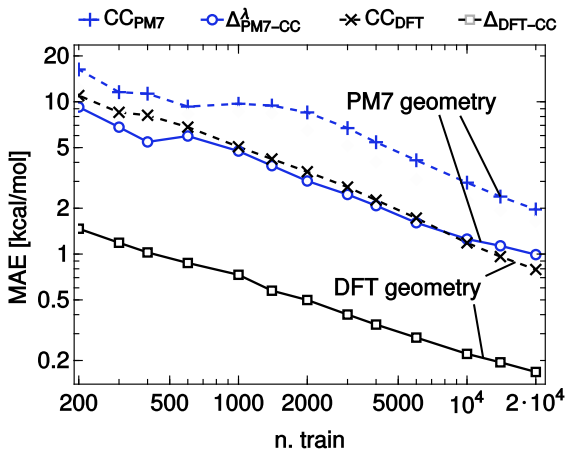- 1kcal/mol error for predicting CCSD(T) based on PM7 geometries; 0.18kcal/mol error for predicting CCSD(T) based on DFT geometries!

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

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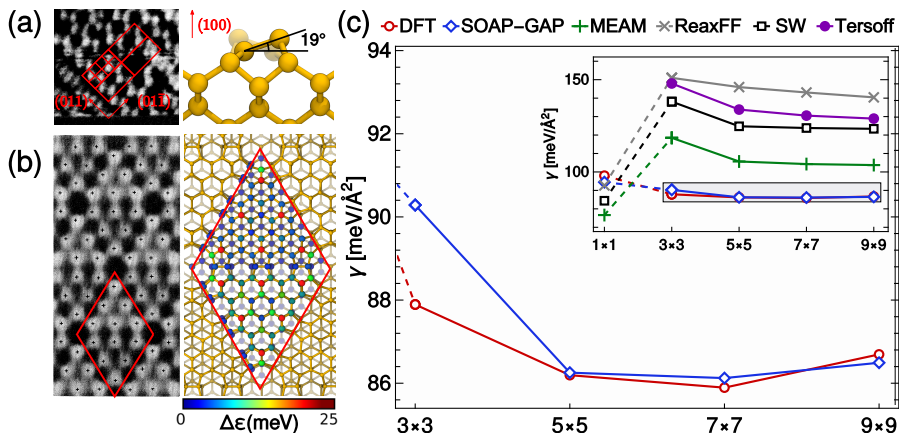
Bartok, De, Kermode, Bernstein, Csanyi, *MC*, Science Advances (2017)

Accurate Predictions for Molecular Crystals

- Substituted pentacenes - model systems for molecular electronics
- Easily achieve sub-kcal/mol accuracy, with REMatch-SOAP kernels
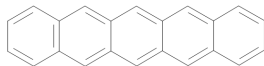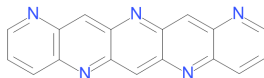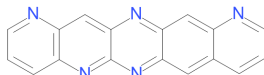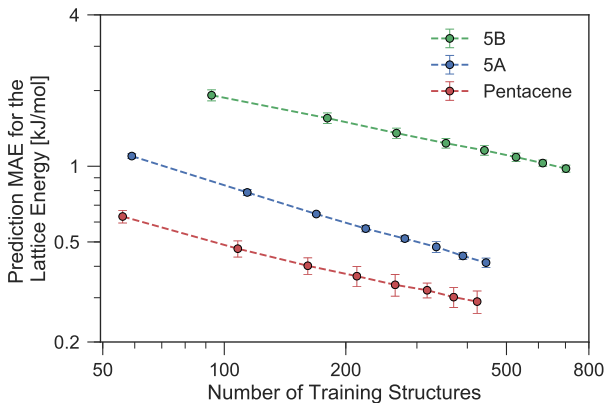
Musil, De, Yang, Campbell, Day, **MC**, Chemical Science (2018)

Recognizing Active Ligands for Receptor Proteins

- 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

(a) FGFR1

n_{train}
20 40 60 80

AUC

MATCH+FPS
MATCH
AVG

(b) $z = +1.15$

(c) $z = +1.53$

$z = +0.78$

$z = -0.97$

PDB/3EML
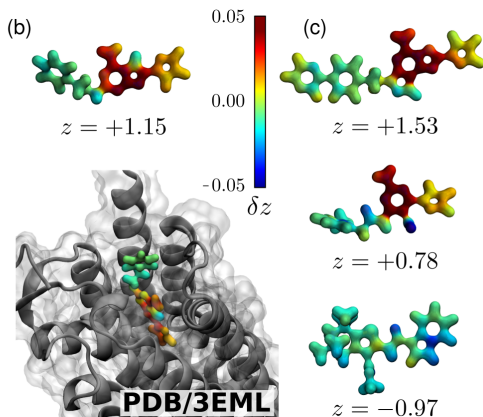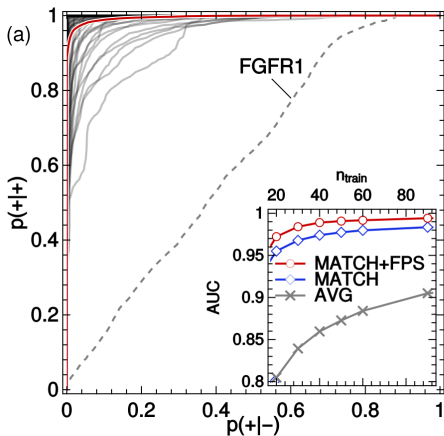
Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

Thorough Sampling of Compound Space

- 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

Random (RND) ×
Farthest point (FPS) □
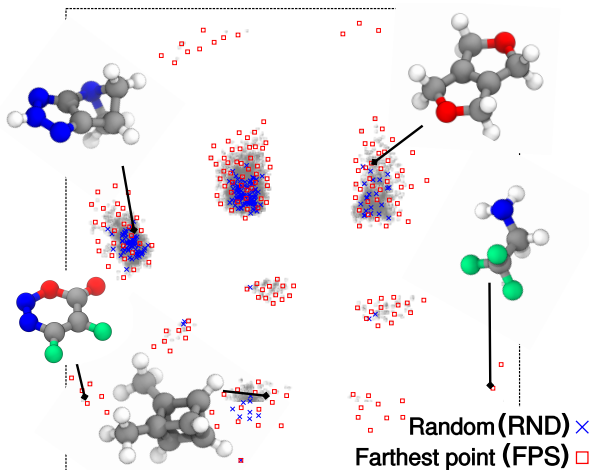
Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

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

Thorough Sampling of Compound Space

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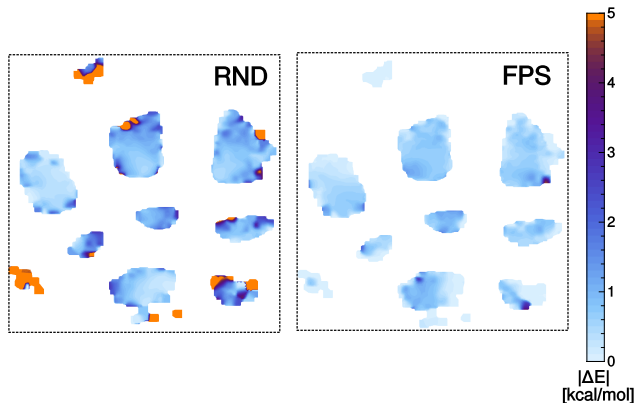
MC, Tribello, Parrinello, PNAS (2011); http://sketchmap.org

Machine-Learning Like a Physicist

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 $P(|y(\mathcal{X}) - y_{\text{ref}}(\mathcal{X})|| \sigma(\mathcal{X}))$
- Use maximum-likelihood to calibrate the uncertainty $\sigma(\mathcal{X}) \rightarrow \alpha\sigma(\mathcal{X})^{\gamma-1}$
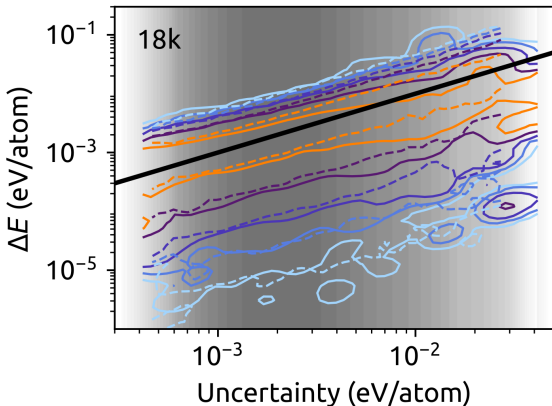
Musil, Willatt, **MC** arxiv.org/abs/1809.07653

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!

w/Emsley, Paruzzo, Hofstetter, 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!

w/Emsley, Paruzzo, Hofstetter, 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!
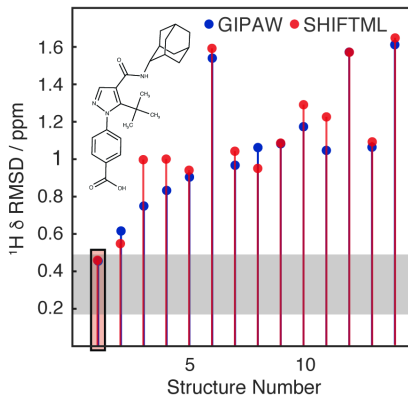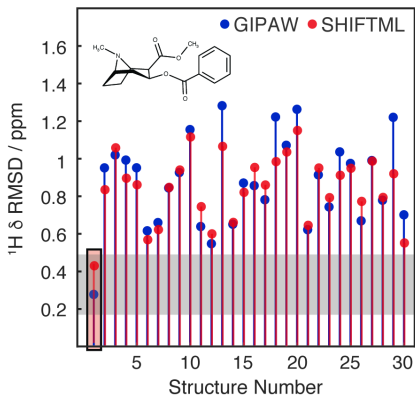
w/Emsley, Paruzzo, Hofstetter, http://shiftml.org

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 ~ 5000 train structures
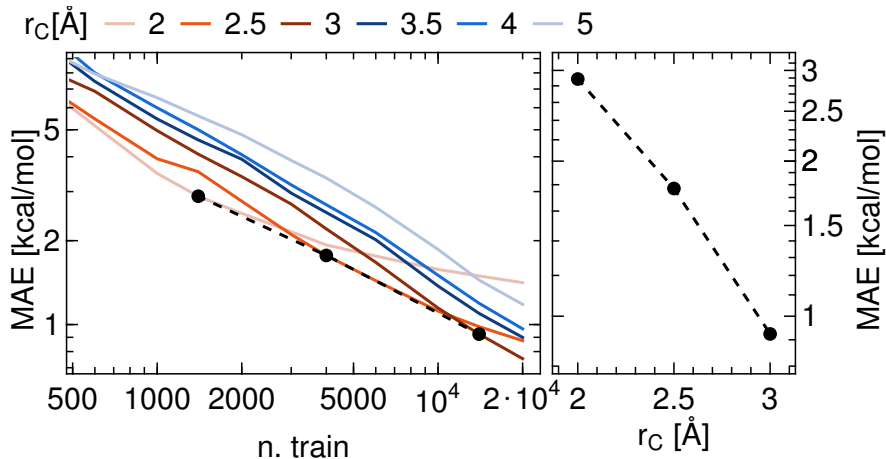
Bartók, De, Kermode, Bernstein, Csányi, **MC**, Science Advances (2017)

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 ~ 5000 train structures
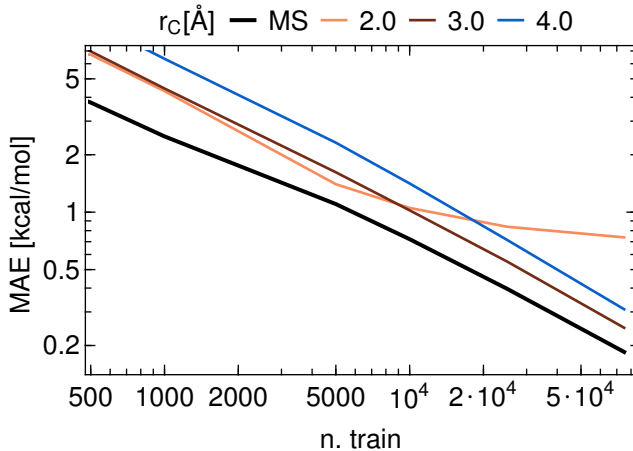
Bartók, De, Kermode, Bernstein, Csányi, **MC**, Science Advances (2017)

A Data-Driven Periodic Table of the Elements

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

$\mathbf{r}_i$

$|O\rangle$

$\langle \mathbf{r} | \mathcal{X}_3 \rangle$

$|H\rangle$

$|C\rangle$

$\langle \mathbf{r} | \mathcal{A} \rangle$

$\langle \mathbf{r} | \mathcal{X}_1 \rangle$

$\langle \mathbf{r} | \mathcal{X}_2 \rangle$

A Data-Driven Periodic Table of the Elements

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

$$|H\rangle = 0.5\,|🔥\rangle + 0.1\,|🪨\rangle + 0.2\,|💧\rangle$$
$$|C\rangle = 0.2\,|🔥\rangle + 0.8\,|🪨\rangle + 0.3\,|💧\rangle$$
$$|O\rangle = 0.1\,|🔥\rangle + 0.1\,|🪨\rangle + 0.6\,|💧\rangle$$

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 $|H\rangle$, $|O\rangle$, . . .
- 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!
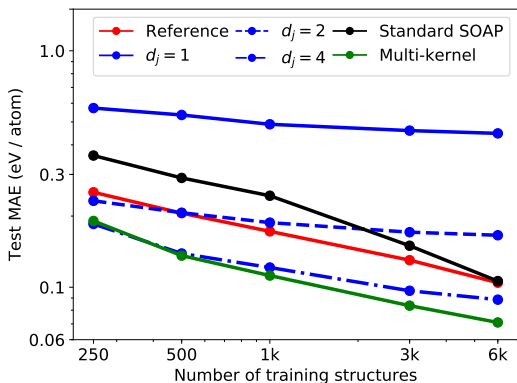
Elpasolite dataset. Reference curve (red) from Faber et al. JCP (2018)

A Data-Driven Periodic Table of the Elements

- How to learn with multiple species? Decorate atomic Gaussian with elemental kets $|H\rangle$, $|O\rangle$, . . .
- 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 . . .
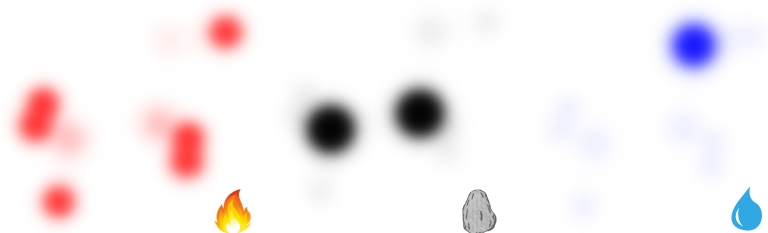- . . . and as by-product get a data-driven version of the periodic table!

Willatt, Musil, **MC**, https://arxiv.org/abs/1807.00236

Machine-Learning Like a Physicist
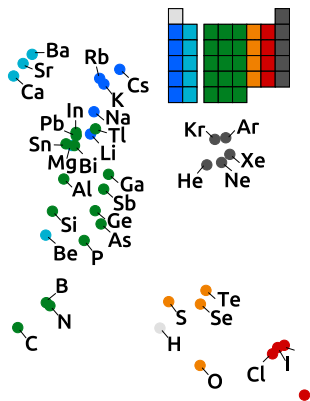
Tensorial properties and beyond

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

$$k_{\mu\nu}\left(\mathcal{X}, \mathcal{X}'\right) \leftrightarrow \left\langle T_\mu\left(\mathcal{X}\right); T_\nu\left(\mathcal{X}'\right)\right\rangle \rightarrow k_{\mu\nu}\left(\hat{R}\mathcal{X}, \hat{R}'\mathcal{X}'\right) = R_{\mu\mu'}\, k_{\mu'\nu'}\left(\mathcal{X}, \mathcal{X}'\right) R'_{\nu\nu'}$$
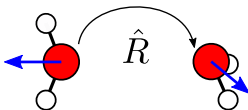
Glielmo, Sollich, & De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)

λ−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

$$K(A,B)$$

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

$$k(\mathcal{X}_i, \mathcal{X}_j)$$

$$\left\langle \mathcal{X}^{(\nu)} \middle| \mathcal{X}'^{(\nu)} \right\rangle_{\hat{R}} \sim \int \mathrm{d}\hat{R} \left| \int \psi(\mathbf{r}) \psi'(\hat{R}\mathbf{r}) \mathrm{d}\mathbf{r} \right|^\nu$$
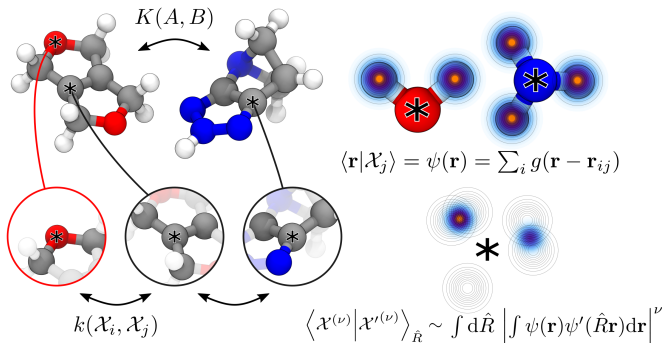
$$k(\mathcal{X}, \mathcal{X}') = \int \mathrm{d}\hat{R}\, \kappa\left(\mathcal{X}, \hat{R}\mathcal{X}'\right), \qquad \kappa(\mathcal{X}, \mathcal{X}') = \left| \int \psi_{\mathcal{X}}(\boldsymbol{x}) \psi_{\mathcal{X}'}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right|^2$$

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

image fro Wikipedia

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

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

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

$\lambda-$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 $SO(3)$
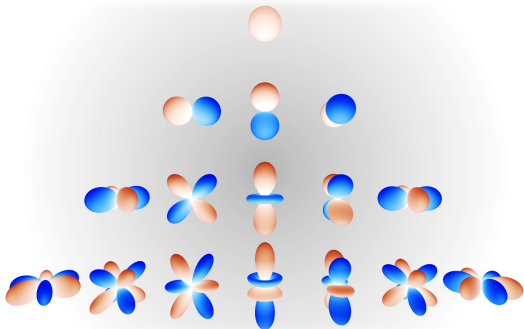- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

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

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
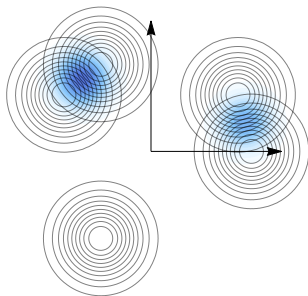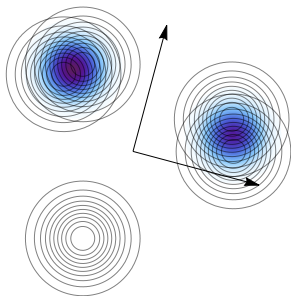

$$k_{\mu\nu}^\lambda (\mathcal{X}, \mathcal{X}') = \int \mathrm{d}\hat{R} D_{\mu\nu}^\lambda \left(\hat{R}\right) \kappa \left(\mathcal{X}, \hat{R}\mathcal{X}'\right)$$

Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)
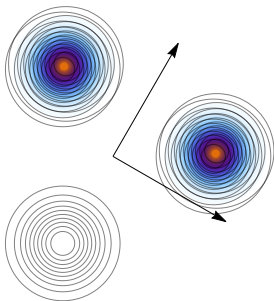
$\lambda-$SOAP: A $SO(3)$ Compliant Kernel

- 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

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

Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)
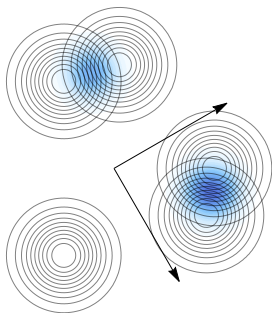
$\lambda-$SOAP: A $SO(3)$ Compliant Kernel

- 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

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

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

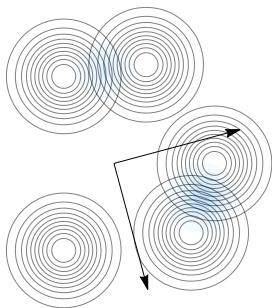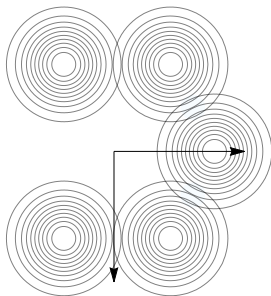# $\lambda-$SOAP: A $SO(3)$ Compliant Kernel

- 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

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

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

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 - and that gives meaningful partitioning in molecular contributions
- Works great for bulk systems (liquid & ice) after fixing non-additive terms

a) H_2O b) $(H_2O)_2$ c) $H_5O_2^+$

Legend: μ, α^0, α^2, β^1, β^3

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

Grisafi, Wilkins, Csányi, & **MC**, PRL (2018)
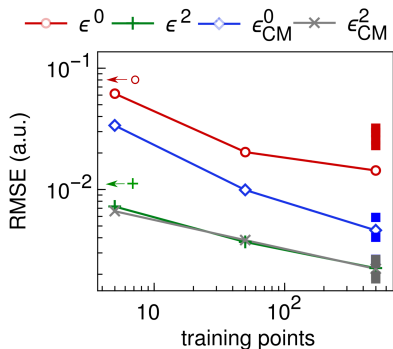
Learning the Dielectric Response of Water

- 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 - and that gives meaningful partitioning in molecular contributions
- Works great for bulk systems (liquid & ice) after fixing non-additive terms

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

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.

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

- Machine-learn directly the full density (non-orthogonal basis is tricky!)
- Highly transferable: learn on C4, predict on C8
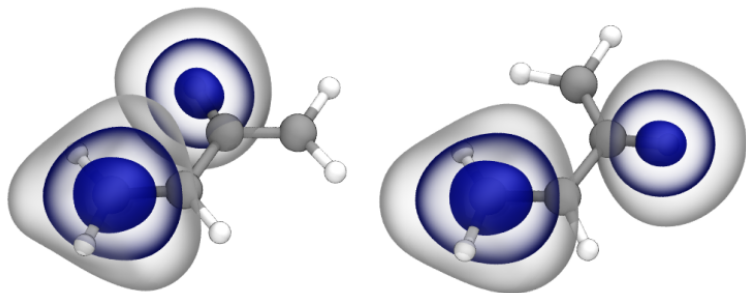
Marzari, Vanderbilt, PRB 1997

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.

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

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

QM	ML	ML − QM

$\varepsilon_\rho = 1.41\%$

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

Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, arxiv.org/abs/1809.05349
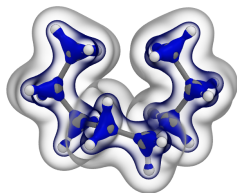
A Transferable Model of the Electron Density

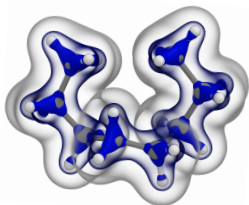- Write the density in atom-centered terms. Use a $\phi_k \equiv R_n Y_m^l$ expansion.

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

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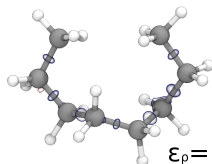- Highly transferable: learn on C4, predict on C8
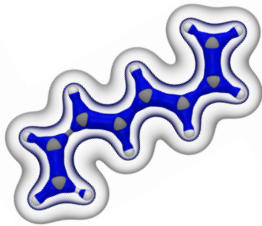
QM   ML   ML − QM

$\varepsilon_\rho = 1.81\%$

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

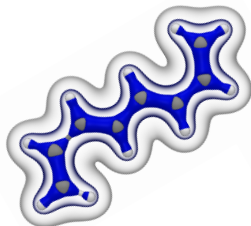Grisafi, Wilkins, Meyer, Fabrizio, Corminboeuf, **MC**, arxiv.org/abs/1809.05349

My Wishlist - Revisited

- General applicability: suitable for all systems and all types of properties
  - "Nearsightedness" of electronic matter ↔ **local** environment decomposition
  - Excellent performance on benchmark DBs, accurate & cheap **error estimate**
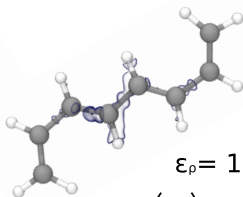 - Predict CCSD from PM7, potentials for solids, 99% prediction of drug activity, silicon & molecular crystals, NMR shieldings in solids
  - Huge potential of a **SA-GPR framework to learn tensors** - electric multipoles and response, but also densities, Hamiltonians, . . .
- Not only a fancy interpolator: 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
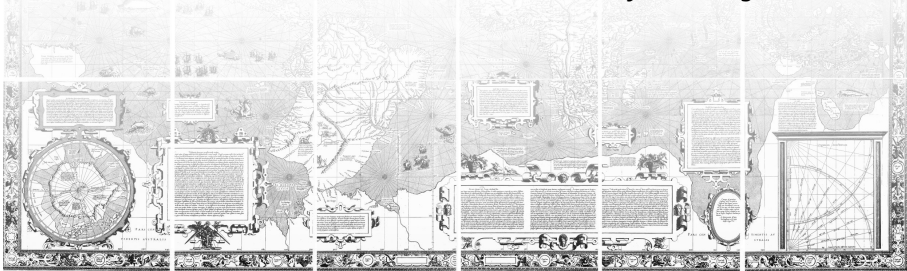
(Development) code available on http://cosmo-epfl.github.io & http://sketchmap.org/

A Map to Navigate Materials & Molecules

- Kernel-induced distances can be also used as the basis of clustering and dimensionality-reduction techniques
  - Generate insightful representations of the (free)-energy landscape of complex systems

$$X_i \in \mathbb{R}^D$$

$$x_i \in \mathbb{R}^d$$

High-dimensional descriptors of a complex system

Non-linear dimensionality reduction (**sketch-map**)

Intuitive maps of structural stability

http://sketchmap.org
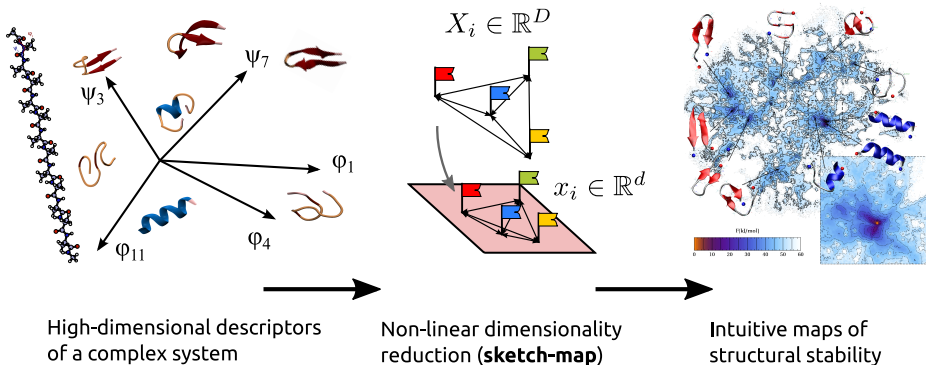
A Map to Navigate Materials & Molecules

- Kernel-induced distances can be also used as the basis of clustering and dimensionality-reduction techniques
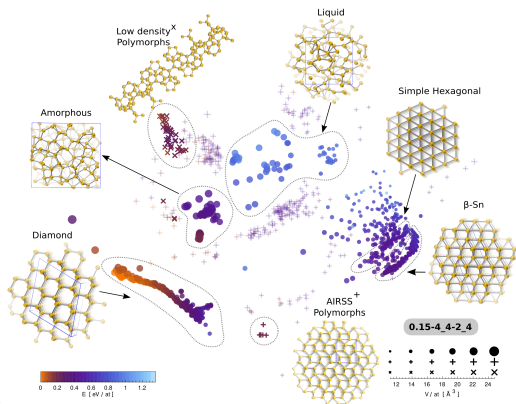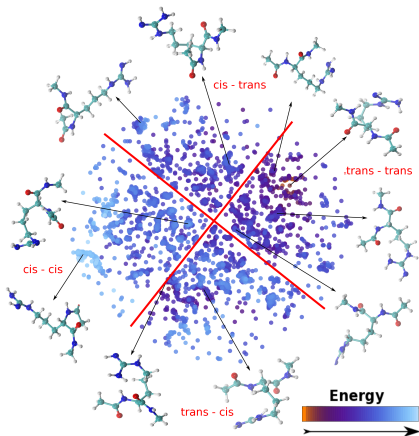 - Generate insightful representations of the (free)-energy landscape of complex systems

MC, Tribello, Parrinello, PNAS (2011).

http://sketchmap.org

Machine-Learning Like a Physicist

Measuring Distances Between Materials

- The crucial ingredient in machine-learning is a method to compare the items whose properties should be predicted
- A distance $D(A, B)$ or a kernel function $K(A, B)$ can be used to assess the (dis)-similarity between items in a set
- Under reasonable assumptions one can always convert a distance $D(A, B)$ to a kernel, or to fingerprints and vice versa

$$K(\mathcal{X}, \mathcal{X}') \approx 1 \qquad K(\mathcal{X}, \mathcal{X}') \approx 0$$
$$D(\mathcal{X}, \mathcal{X}') \approx 0 \qquad D(\mathcal{X}, \mathcal{X}') \gg 0$$

Measuring Distances Between Materials

- The crucial ingredient in machine-learning is a method to compare the items whose properties should be predicted
- A distance $D(A, B)$ or a kernel function $K(A, B)$ can be used to assess the (dis)-similarity between items in a set
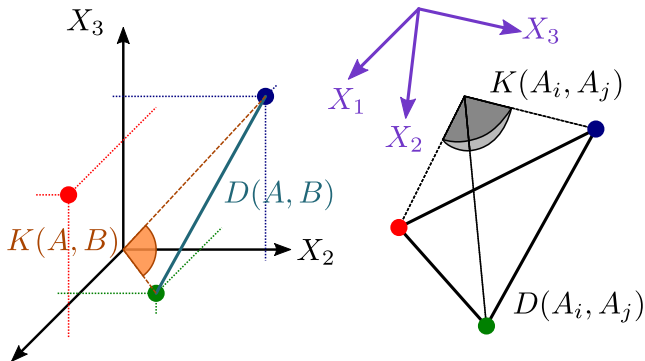- Under reasonable assumptions one can always convert a distance $D(A, B)$ to a kernel, or to fingerprints and vice versa

Michele Ceriotti https://cosmo.epfl.ch   *Machine-Learning Like a Physicist*

How accurate can we get (and what we learn)?

- What happens if we increase the train set fraction?
  - Can we improve the accuracy by tuning the kernel?

QM9, MAE on atomization energies (eV)

	CM	0.128
	BOB	0.0667
	BAML	0.0519
KRR	ECFP4	4.25
	HDAD	**0.0251**
	HD	0.0644
	MARAD	0.0529

90% TRAIN, arxiv:1702.0553

KRR/SOAP: 0.012 eV (1kJ/mol)
75% TRAIN, Science Advances (2017)

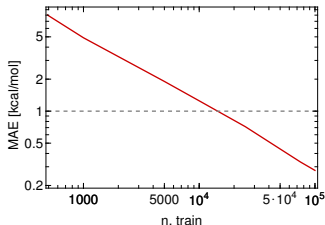Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

How accurate can we get (and what we learn)?

- What happens if we increase the train set fraction?
  - Can we improve the accuracy by tuning the kernel?

QM7B, MAE on atomization energies (eV)

Repr.	Kernel	E / kcal mol^{-1}	
		RMSE	MAE
CM	Laplacian	5.48	3.54
BoB	Laplacian	3.32	1.95
BAML [17]	Laplacian	2.54	1.15
~~SOAP [35]~~	~~REMatch~~	~~1.61~~	~~0.92~~
MBTR	Linear	1.81	0.82
MBTR	Gaussian	**0.94**	**0.60**

75% TRAIN, arxiv:1704.06439

KRR/SOAP: 0.40 kcal/mol
75% TRAIN, Science Advances (2017)

Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)
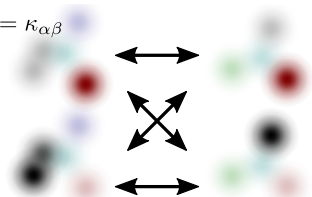
How accurate can we get (and what we learn)?

- What happens if we increase the train set fraction?
  - Can we improve the accuracy by tuning the kernel?

QM7B, MAE on atomization energies (eV)

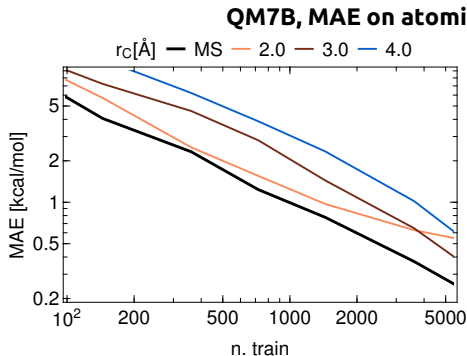

$$\langle \alpha | \beta \rangle = \kappa_{\alpha\beta}$$

$$\int \mathrm{d}\hat{R} \left| \sum_{\alpha\beta} \kappa_{\alpha\beta} \int \rho_\alpha(\mathbf{x}) \rho'_\beta(\hat{R}\mathbf{x}) \right|^2$$

$$\kappa_{\alpha\beta} = e^{-\left(E_\alpha - E_\beta\right)^2 / 2\lambda^2}$$

KRR/SOAP: 0.40 kcal/mol
ALCHEMY: 0.33 kcal/mol
75% TRAIN, Science Advances (2017)

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

How accurate can we get (and what we learn)?

- What happens if we increase the train set fraction?
  - Can we improve the accuracy by tuning the kernel?

QM7B, MAE on atomization energies (eV)

r_C[Å] — MS — 2.0 — 3.0 — 4.0

KRR/SOAP: 0.40 kcal/mol
MULTISCALE: 0.26 kcal/mol
75% TRAIN, Science Advances (2017)

$$K(A, B) = \sum_i w_i K_i(A, B)$$

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

How accurate can we get (and what we learn)?

- What happens if we increase the train set fraction?
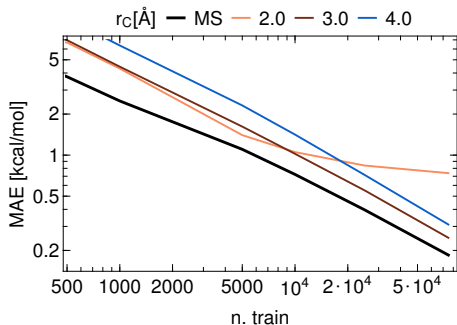 - Can we improve the accuracy by tuning the kernel?

QM9, MAE on atomization energies (eV)

$r_C[\text{Å}]$ — MS — 2.0 — 3.0 — 4.0

MAE [kcal/mol] vs n. train

KRR/SOAP: 0.3 kcal/mol
MULTISCALE: 0.18 kcal/mol
75% TRAIN, Science Advances (2017)

$$K(A, B) = \sum_i w_i K_i(A, B)$$

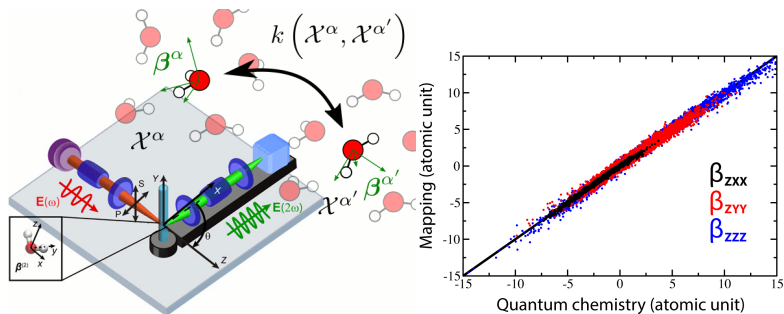Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

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

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

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

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

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

Understanding Errors in Quantum Calculations

- Learning differences between electronic structure methods is simpler
- Atom-centered energetics give insight into the impact of approximations

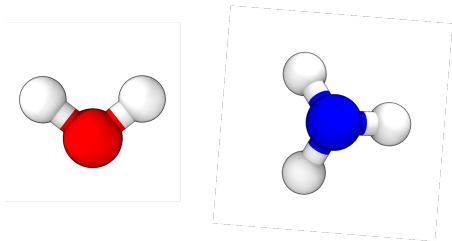ΔE [kcal/mol]

PBE − PM7

PBE0 − PBE
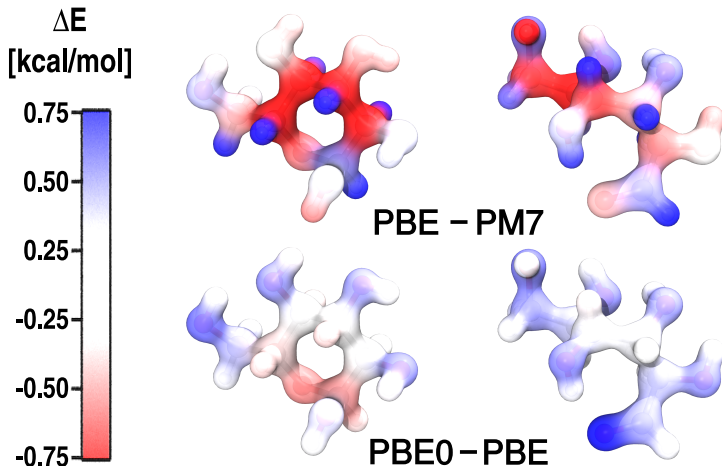
Ramakrishnan et al., JCTC (2015); M. Marianski et al., J. Chem. Theory Comput. 12, 6157 (2016); Bartok, De, Kermode, Bernstein, Csanyi, **MC**, Science Advances (2017)

Structures as Combinations of Local Environments

- One can write *structural* kernels as a combination of *local* kernels
  - Entropy-regularized Wasserstein distance interpolates between "best-match" and "average" constructions
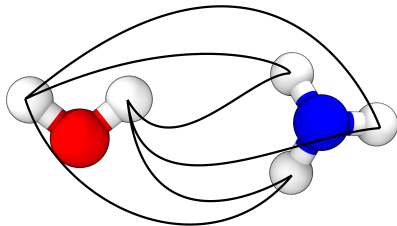
$$K(A, B) = \sum_{i \in A, j \in B} P_{ij} k(\mathcal{X}_i, \mathcal{X}_j)$$

De, Bartók, Csányi, **MC**, PCCP (2016); M. Cuturi, NIPS (2013);

Structures as Combinations of Local Environments

- One can write *structural* kernels as a combination of *local* kernels
  - Entropy-regularized Wasserstein distance interpolates between "best-match" and "average" constructions

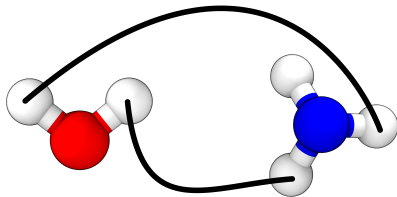$$K(A, B) = \sum_{i \in A, j \in B} P_{ij} k(\mathcal{X}_i, \mathcal{X}_j)$$

$$P_{ij} = \frac{1}{N_A N_B}$$

De, Bartók, Csányi, **MC**, PCCP (2016); M. Cuturi, NIPS (2013);

Structures as Combinations of Local Environments

- One can write *structural* kernels as a combination of *local* kernels
  - Entropy-regularized Wasserstein distance interpolates between "best-match" and "average" constructions

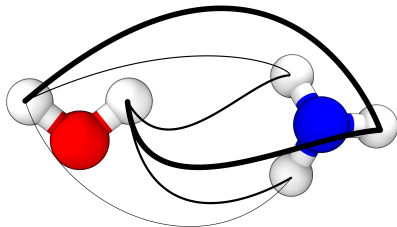$$K(A, B) = \sum_{i \in A, j \in B} P_{ij} k(\mathcal{X}_i, \mathcal{X}_j)$$

$$\hat{K}(A, B) \propto$$
$$\max_{\mathbf{P} \in \mathcal{U}} \sum_{ij} P_{ji} C_{ij}^{A,B}$$

De, Bartók, Csányi, **MC**, PCCP (2016); M. Cuturi, NIPS (2013);

Structures as Combinations of Local Environments

- One can write *structural* kernels as a combination of *local* kernels
  - Entropy-regularized Wasserstein distance interpolates between "best-match" and "average" constructions

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

$$\hat{K}^\gamma(A, B) \propto$$
$$\max_{\mathbf{P} \in \mathcal{U}} \sum_{ij} P_{ji}(C_{ij}^{A,B} - \gamma \ln P_{ji})$$

De, Bartók, Csányi, **MC**, PCCP (2016); M. Cuturi, NIPS (2013);

Additive Property Models & Beyond

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

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

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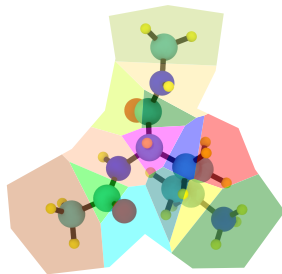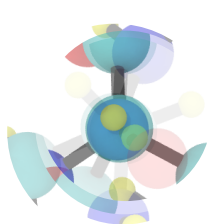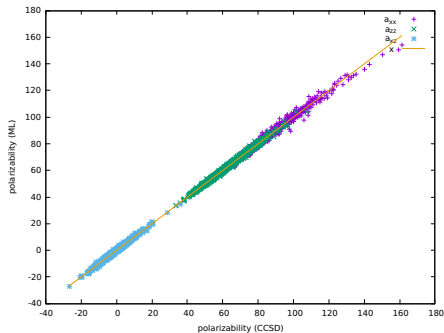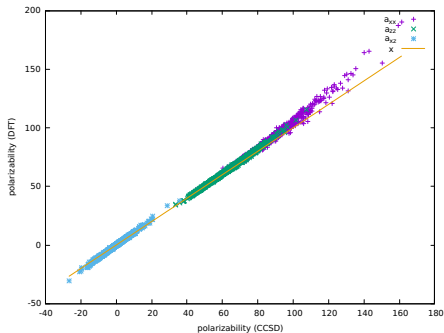

$$K(A, B) = \sum_{i,j} k(\mathcal{X}_i^A, \mathcal{X}_j^B)$$

$$E(A) = \sum_i \epsilon(\mathcal{X}_i^A)$$
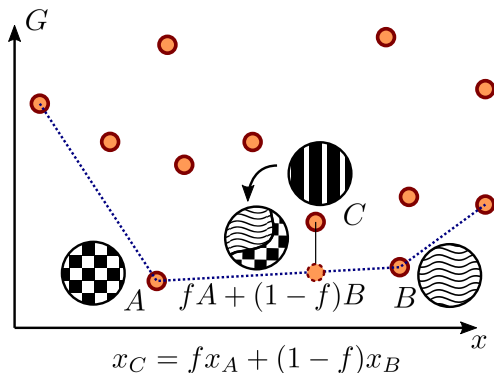
Predicting the Full Polarizability of Molecules

- Benchmarking polarizability learning on the QM7b dataset. DFT and high-end coupled-cluster references (Rob DiStasio@Cornell)
- WIP (*<2000 training structures, no optimization!) - we can predict α with better accuracy than DFT.

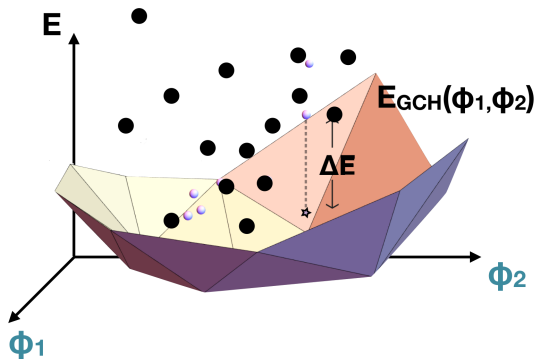| $\langle \cdot \rangle, \sigma(\cdot)$ [a.u.]	α_{xx}	α_{yy}	α_{zz}	α_{xy}	α_{xz}	α_{yz}
DFT vs CCSD	2.6, 2.6	2.0, 2.1	0.9, 0.9	0.6, 1.3	0.0, 0.6	0.1, 0.6
SA-GPR* vs CCSD	0.0, 1.5	0.0, 1.4	0.0, 0.9	0.0, 1.0	0.0, 0.7	0.0, 0.6
ΔSA-GPR*	0.0, 0.7	0.0, 0.6	0.0, 0.3	0.0, 0.4	0.0, 0.3	0.0, 0.2

A Generalized Convex Hull Construction

- The convex hull construction determines phases that are stable for given thermodynamic boundary conditions (volume, composition, ...)
- We use a kernel principal component analysis to assign abstract descriptors to each phase and build a generalized convex hull. Probabilistic construction, with uncertainty quantification

$$x_C = f x_A + (1-f) x_B$$

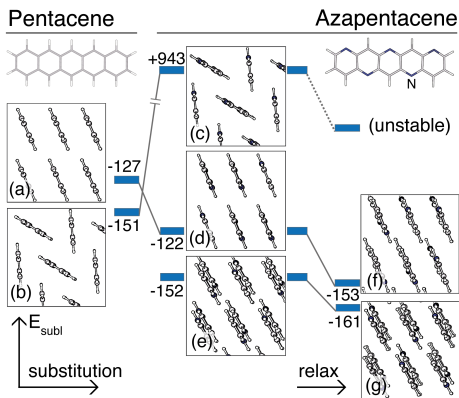Anelli, Engel, Pickard & **MC** arxiv.org/1803.01932

A Generalized Convex Hull Construction

- The convex hull construction determines phases that are stable for given thermodynamic boundary conditions (volume, composition, …)
- We use a kernel principal component analysis to assign abstract descriptors to each phase and build a generalized convex hull. Probabilistic construction, with uncertainty quantification

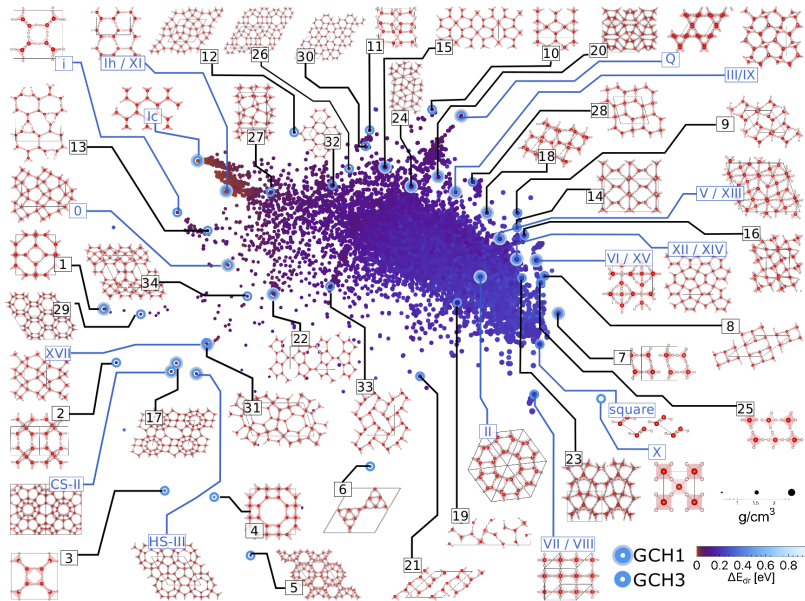Anelli, Engel, Pickard & **MC** arxiv.org/1803.01932

A Generalized Convex Hull Construction

- The convex hull construction determines phases that are stable for given thermodynamic boundary conditions (volume, composition, ...)
- We use a kernel principal component analysis to assign abstract descriptors to each phase and build a generalized convex hull. Probabilistic construction, with uncertainty quantification

Anelli, Engel, Pickard & **MC** arxiv.org/1803.01932

A Generalized Convex Hull Construction

Engel, Anelli, **MC**, Pickard & Needs, Nature Comm. (2018)