Physics-Based Machine Learning for Materials and Molecules

Michele Ceriotti
https://cosmo.epfl.ch

CECAM Discussion Meeting
December 2018, CERMICS
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.

![Diagram showing the tradeoff between accuracy and cost for different methods: ab initio and forcefields.](https://cosmo.epfl.ch)
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

\[ \hat{H} \psi = E \psi \]

\[ E (q) = \sum_{ij} v (r_{ij}) + \ldots, \quad E (q) = ML (q \mid \{q_i, V_i\}) \]
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

\[ |A\rangle \]

\[ \mathbf{r}_i \]

\[ \hat{T} \]

\[ \mathbf{r}_1 \leftrightarrow \mathbf{r}_2 \]

\[ \hat{R} \]
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

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

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

$$E(A_j) = \sum_i w_i K(A_j, A_i)$$
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, \( K(A, B) \sim \langle A|B \rangle \).

\[
E(A) = \sum_i w_i K(A, A_i)
\]
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, $K(A, B) \sim \langle A|B \rangle$. 

![Diagram showing machine-learning in a quantum context.](https://cosmo.epfl.ch)
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 $|\mathcal{X}^{(\nu)}\rangle \hat{R}$

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

Symmetry-adapted atom-density representations

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

\[
\int d\hat{T} \left< r | \hat{T} | A \right> = \sum_i \int dt \, g(r + t - r_i) |\alpha_i\rangle = \sum_{\alpha} N_{\alpha} |\alpha\rangle
\]

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
- $|\mathcal{A}^{(\nu)}\rangle$ leads naturally to atom-centered decomposition
- Rotational average yields $(\nu + 1)$-body correlation functions $|\mathcal{X}^{(\nu)}\rangle$

\[
\int d\hat{T} \langle r | \hat{T} | \mathcal{A}\rangle \langle r' | \hat{T} | \mathcal{A}\rangle = \int dr' \langle r' | \mathcal{A}\rangle \langle r' + r | \mathcal{A}\rangle
\]

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 $|\chi^{(\nu)}\rangle_{\hat{R}}$

\[
\langle \mathbf{r} \mid \mathcal{A}^{(2)} \rangle_{\hat{T}} = \sum_{ij} |\alpha_i \alpha_j\rangle g(\mathbf{r} - \mathbf{r}_{ij}) \\
= \sum_j |\alpha_j\rangle \langle \mathbf{r} \mid \chi_j \rangle
\]

Symmetry-adapted atom-density representations

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

\[
\int d\hat{R} \langle r | \hat{R} | \chi_j \rangle \rightarrow \langle r | \chi_j^{(1)} \rangle
\]

\[
\int d\hat{R} \langle r | \hat{R} | \chi_j \rangle \langle r' | \hat{R} | \chi_j \rangle \rightarrow \langle rr' \omega | \chi_j^{(2)} \rangle
\]
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.

\[ \langle r | \mathcal{X}_j \rangle = \psi(r) = \sum_i g(r - r_{ij}) \]

\[ k(\mathcal{X}, \mathcal{X}') \equiv \langle \mathcal{X} | \mathcal{X}' \rangle \sim \int \psi(r) \psi'(r) dr \]
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.

\[
\langle r | \mathcal{X}_j \rangle = \psi(r) = \sum_i g(r - r_{ij})
\]

\[
\langle nlm | \mathcal{X}_j \rangle = \int dx \: \psi(r) R_n(r) Y^l_m(\hat{r})
\]

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.

\[
\langle nn'l | \chi^{(2)} \rangle_R = \sum_m \langle nlm | \chi \rangle \langle n'lm | \chi \rangle
\]

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

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

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

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

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.

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

Bartók, Kondor, Csányi, PRB (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
- 1 kcal/mol error for predicting CCSD(T) based on PM7 geometries;
- 0.18 kcal/mol error for predicting CCSD(T) based on DFT geometries!

Ramakrishnan et al., Scientific Data (2014); Ramakrishnan et al., JCTC (2015)
100k molecules with coupled-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
- 1 kcal/mol error for predicting CCSD(T) based on PM7 geometries;
- 0.18 kcal/mol error for predicting CCSD(T) based on DFT geometries!

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.

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

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!

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!

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!

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


MC, Tribello, Parrinello, PNAS (2011); http://sketchmap.org
Train set optimization to reduce errors

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

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

![Graph showing RMSE vs. N']

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 \( P(|y(\mathcal{X}) - y_{\text{ref}}(\mathcal{X})| \mid \sigma(\mathcal{X})) \)

- 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
An accurate & inexpensive error estimation

- Generate an ensemble of GPR models, and use distribution of predictions
  \[ y(\mathbf{x}) = \frac{1}{N_{RS}} \sum_i y^{(i)}(\mathbf{x}), \quad \sigma^2(\mathbf{x}) = \frac{1}{N_{RS} - 1} \sum_i \left( y^{(i)}(\mathbf{x}) - y(\mathbf{x}) \right)^2 \]

- Verify accuracy by the distribution of errors
  \[ P(|y(\mathbf{x}) - y_{\text{ref}}(\mathbf{x})| | \sigma(\mathbf{x})) \]

- Use maximum-likelihood to calibrate the uncertainty
  \[ \sigma(\mathbf{x}) \rightarrow \alpha \sigma(\mathbf{x})^{\gamma^{-1}} \]

Musil, Willatt, MC 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

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

A data-driven periodic table of the elements

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

$|H\rangle = 0.5 |\text{火烧}\rangle + 0.1 |\text{石头}\rangle + 0.2 |\text{水}\rangle$

$|C\rangle = 0.2 |\text{火烧}\rangle + 0.8 |\text{石头}\rangle + 0.3 |\text{水}\rangle$

$|O\rangle = 0.1 |\text{火烧}\rangle + 0.1 |\text{石头}\rangle + 0.6 |\text{水}\rangle$

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**, PCCP (2018)
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

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

- Properties that are \textit{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 \textit{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)
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:

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

$$\kappa (\mathcal{X}, \mathcal{X}') = \left| \int \psi_\mathcal{X} (\mathbf{x}) \psi_\mathcal{X}' (\mathbf{x}) d\mathbf{x} \right|^2$$
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 $T^\lambda$, corresponding to the representations of SO(3)
- A hierarchy of λ-SOAP kernels can be defined to learn tensorial quantities

\[
T^\lambda_{\mu} \left( \hat{R} (\chi') \right) = D^\lambda_{\mu \mu'} \left( \hat{R} \right) T^\lambda_{\mu'} (\chi')
\]

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

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

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

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

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

$$k^\lambda_{\mu\nu} (\vec{x}, \vec{x}') = \int d\hat{R} \, D^\lambda_{\mu\nu} (\hat{R}) \, \kappa (\vec{x}, \hat{R}\vec{x}')$$

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

$$k^\lambda_{\mu \nu} (\chi, \chi') = \int d\hat{R} \, D^\lambda_{\mu \nu} (\hat{R}) \kappa (\chi, \hat{R} \chi')$$
\( \lambda \)-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

\[
k_{\mu\nu}^\lambda (x, x') = \int d\hat{R}D_{\mu\nu}^\lambda (\hat{R}) \kappa (x, \hat{R}x')
\]

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

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

Grisafi, Wilkins, Csányi, & MC, PRL (2018)
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 $\text{C}_8\text{H}_{11}\text{N}_5\text{O}_3$) with better-than-DFT accuracy
- Atom-centered environment decomposition of $\alpha$ and the DFT error

<table>
<thead>
<tr>
<th>5 NUCLEOBASES</th>
<th>15 CANONICAL AMINO ACIDS (L-type)</th>
<th>5 POPULAR DRUGS</th>
<th>2 CARBOHYDRATES</th>
<th>23 ISOMERS of $\text{C}_9\text{H}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Guanine</td>
<td>(6) $\text{H}_2\text{N}-\text{COOH}$</td>
<td>(21) $\text{N}^+\text{O}$</td>
<td>(26) CH$_2$OH</td>
<td>(30) n-Octane</td>
</tr>
<tr>
<td>(2) Adenine</td>
<td>(7) $\text{H}_2\text{N}-\text{COOH}$</td>
<td>(22) Paracetamol</td>
<td>(27) CH$_3$OH</td>
<td>(31) Cyclooctane</td>
</tr>
<tr>
<td>(3) Thymine</td>
<td>(8) $\text{H}_2\text{N}-\text{COOH}$</td>
<td>(23) Aspirin</td>
<td>(28) H$_2$O</td>
<td>(32) 2,2,3,3-Tetramethylbutane</td>
</tr>
<tr>
<td>(4) NH$_2$</td>
<td>(9) $\text{H}_2\text{N}-\text{COOH}$</td>
<td>(24) Metformin</td>
<td>(29) $\alpha$-D-Glucopyranose</td>
<td>(33) 2-Methylheptane</td>
</tr>
<tr>
<td>(5) O</td>
<td>(10) $\text{H}_2\text{N}-\text{COOH}$</td>
<td>(25) Acyclovir</td>
<td></td>
<td>(34) -</td>
</tr>
<tr>
<td></td>
<td>(11) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(35) -</td>
</tr>
<tr>
<td></td>
<td>(12) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(36) -</td>
</tr>
<tr>
<td></td>
<td>(13) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(37) -</td>
</tr>
<tr>
<td></td>
<td>(14) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(38) -</td>
</tr>
<tr>
<td></td>
<td>(15) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(39) -</td>
</tr>
<tr>
<td></td>
<td>(16) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(40) -</td>
</tr>
<tr>
<td></td>
<td>(17) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(41) -</td>
</tr>
<tr>
<td></td>
<td>(18) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(42) -</td>
</tr>
<tr>
<td></td>
<td>(19) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(43) -</td>
</tr>
<tr>
<td></td>
<td>(20) $\text{H}_2\text{N}-\text{COOH}$</td>
<td></td>
<td></td>
<td>(44) -</td>
</tr>
<tr>
<td></td>
<td>(21) $\text{N}^+\text{O}$</td>
<td></td>
<td></td>
<td>(45) -</td>
</tr>
<tr>
<td></td>
<td>(22) Paracetamol</td>
<td></td>
<td></td>
<td>(46) -</td>
</tr>
<tr>
<td></td>
<td>(23) Aspirin</td>
<td></td>
<td></td>
<td>(47) -</td>
</tr>
<tr>
<td></td>
<td>(24) Metformin</td>
<td></td>
<td></td>
<td>(48) -</td>
</tr>
<tr>
<td></td>
<td>(25) Acyclovir</td>
<td></td>
<td></td>
<td>(49) -</td>
</tr>
<tr>
<td></td>
<td>(26) CH$_2$OH</td>
<td></td>
<td></td>
<td>(50) -</td>
</tr>
<tr>
<td></td>
<td>(27) CH$_3$OH</td>
<td></td>
<td></td>
<td>(51) -</td>
</tr>
<tr>
<td></td>
<td>(28) H$_2$O</td>
<td></td>
<td></td>
<td>(52) -</td>
</tr>
</tbody>
</table>

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

\[
\begin{aligned}
\text{Method} & \quad \text{RMSE} \\
\text{CCSD/DFT} & \quad 0.573 \\
\text{CCSD/ML} & \quad 0.304 \\
\text{DFT/ML} & \quad 0.403 \\
\Delta(\text{CCSD-DFT})/\text{ML} & \quad 0.212
\end{aligned}
\]

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 \( \text{C}_8\text{H}_{11}\text{N}_5\text{O}_3 \)) 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
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.

\[
\alpha = (\epsilon - 1)(\epsilon + 2)^{-1} \nu
\]

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

Grisafi, Wilkins, Csányi, & MC, PRL (2018)
A transferable model of the electron density

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

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

- 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.

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

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

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.

$$\mathcal{F}(\rho) = \int dr \left| \sum_{ik} c_{ik} \phi_k (r - r_i) \right|^2 + \eta |c|^2, \quad c_{inlm} = \sum_{jm'} x_{jnlm} k_{mm'}^l (X_i, X_j)$$

Machine-learn directly the full density (non-orthogonal basis is tricky!)

Highly transferable: learn on C4, predict on C8

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/