Learning and Correcting Static and Dynamic Properties of Materials in Atomistic Simulations

Michele Ceriotti
http://cosmo.epfl.ch

IPAM Workshop, November 2017, L.A.

COSMO
COMPUTATIONAL SCIENCE AND MODELLING

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

http://cosmo.epfl.ch

Follow @COSMO_EPFL

Sandip De, Felix Musil,
Andrea Grisafi, David Wilkins
Venkat Kapil, M. Rossi

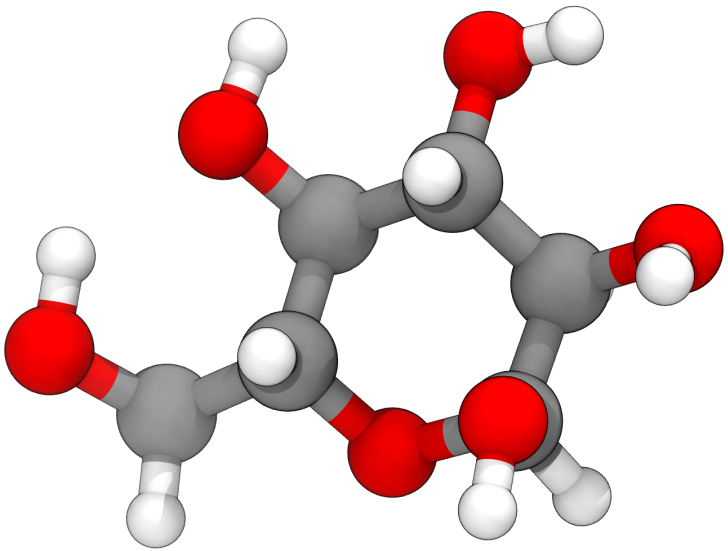
G. Csányi, A. Bartók, C. Poelking,
J. Kermode, N. Bernstein

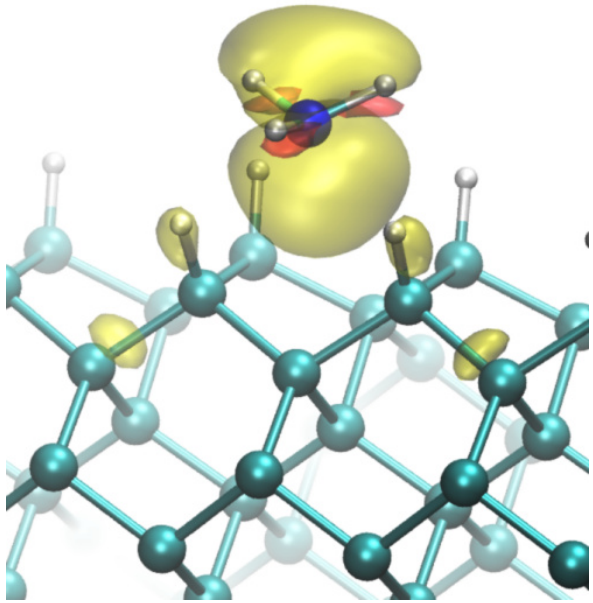
Hiring post-docs with machine-learning background/interests

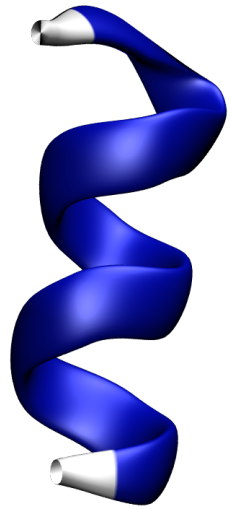
Machine-Learning Static Properties of Materials and Molecules

A Universal Predictor of Atomic-Scale Properties

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



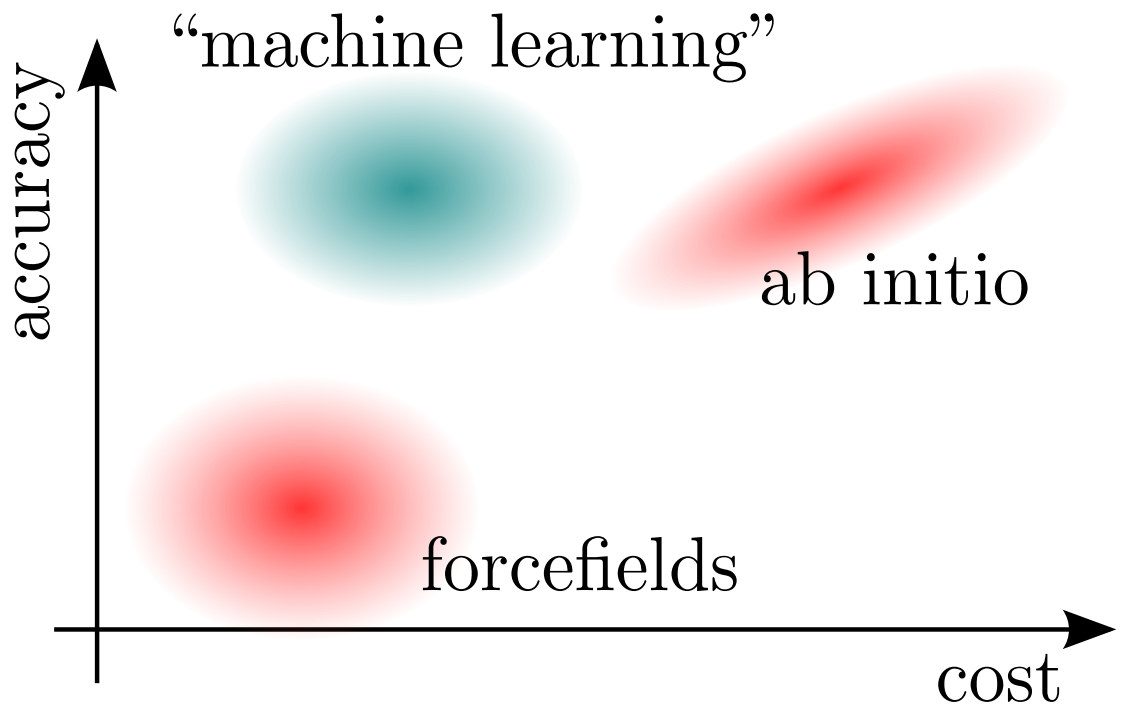




$$\hat{\mathcal{H}}\Psi = E\Psi$$

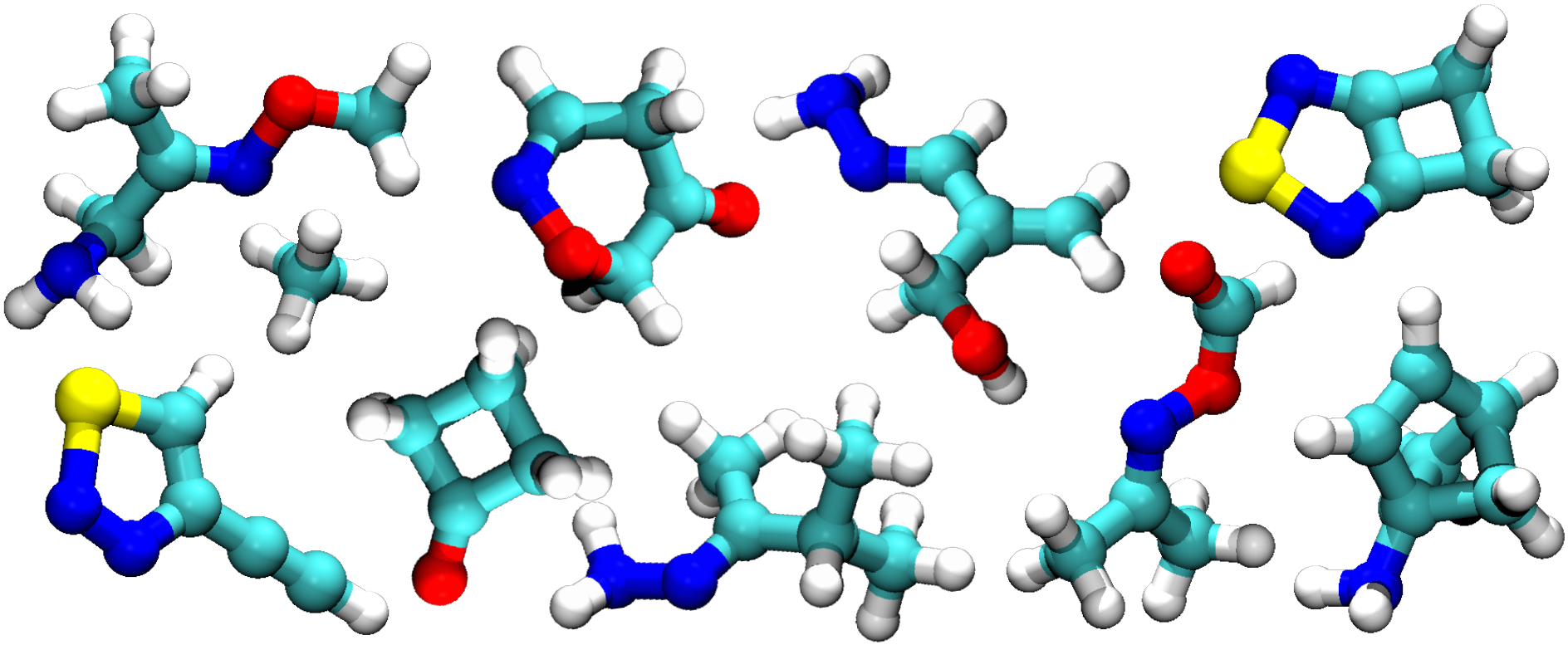
A Universal Predictor of Atomic-Scale Properties

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



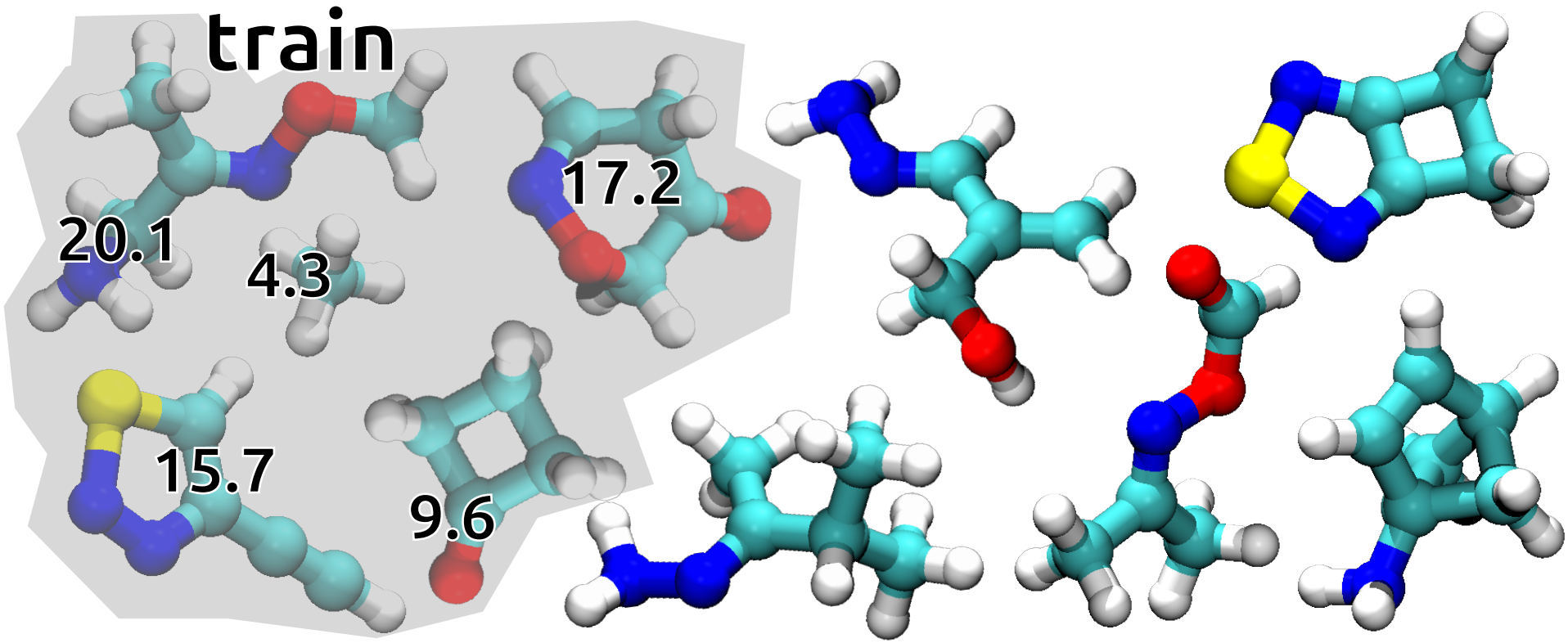
Machine-Learning as a Universal Interpolator

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



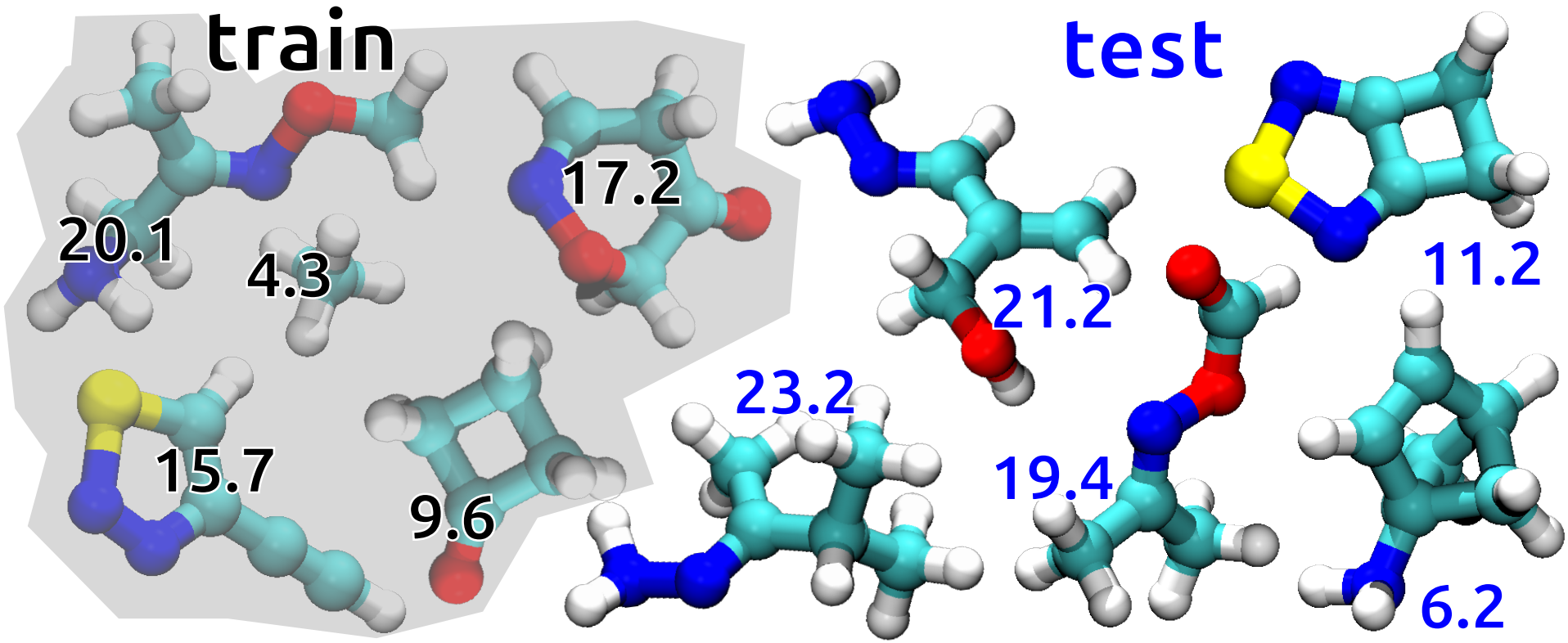
Machine-Learning as a Universal Interpolator

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



Machine-Learning as a Universal Interpolator

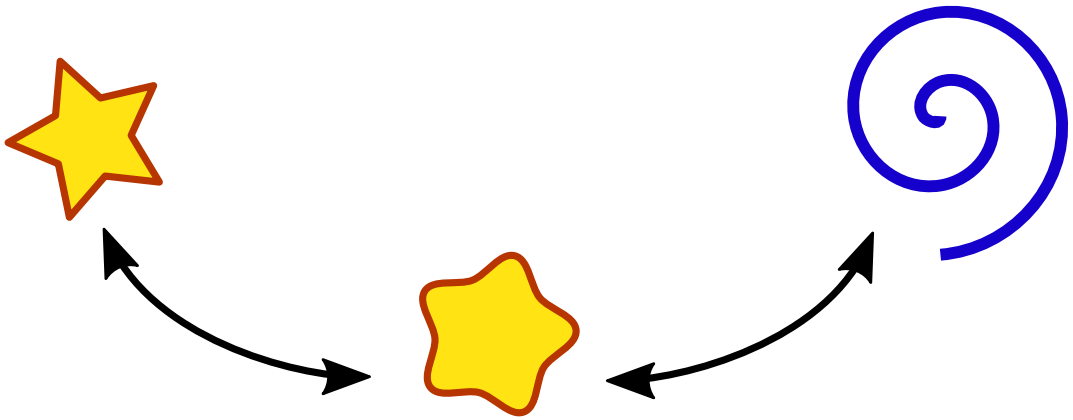
- Machine-learning can be regarded as a sophisticated interpolation between a few known values of the properties
- Can it be made as accurate and general as the Schrödinger equation?



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

Measuring distances between materials

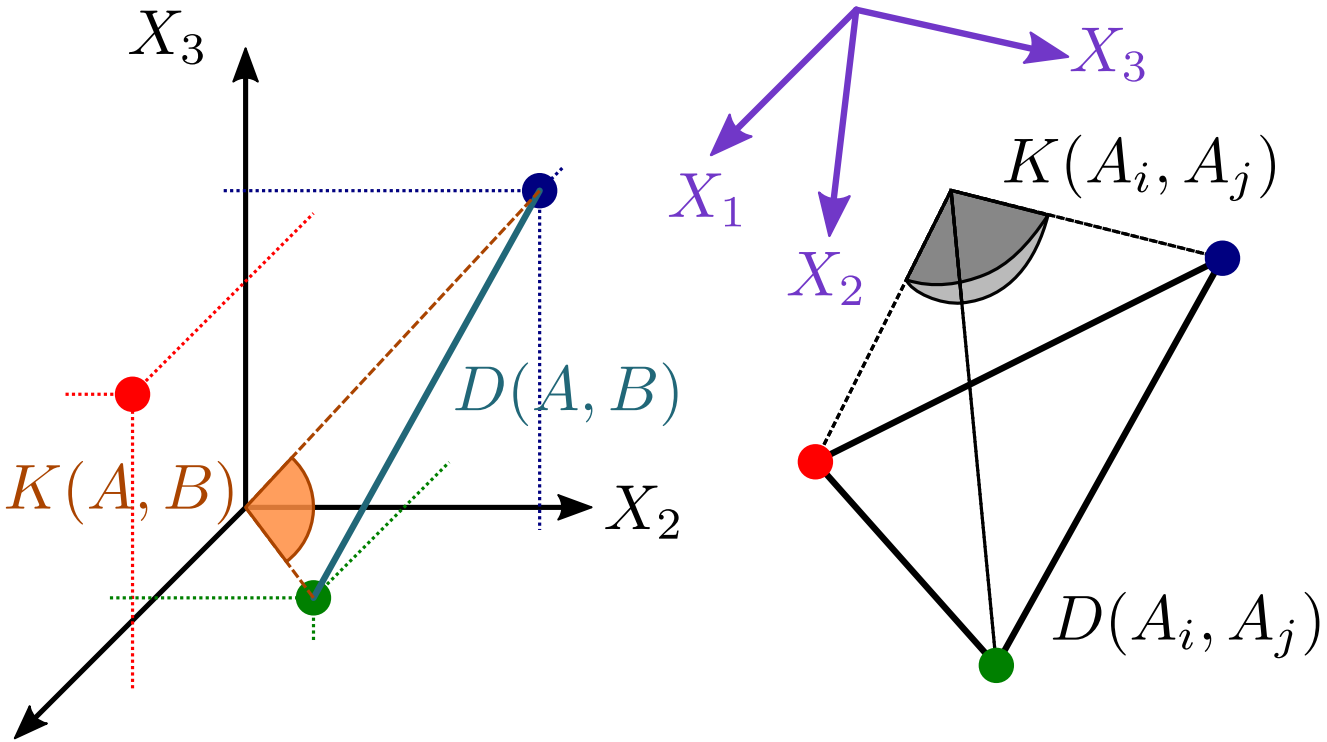
- The crucial ingredient in machine-learning is a method to compare the items whose properties should be predicted
- A distance $D(A, B)$ or a kernel function $K(A, B)$ can be used to assess the (dis)-similarity between items in a set
- Under reasonable assumptions one can always convert a distance $D(A,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



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

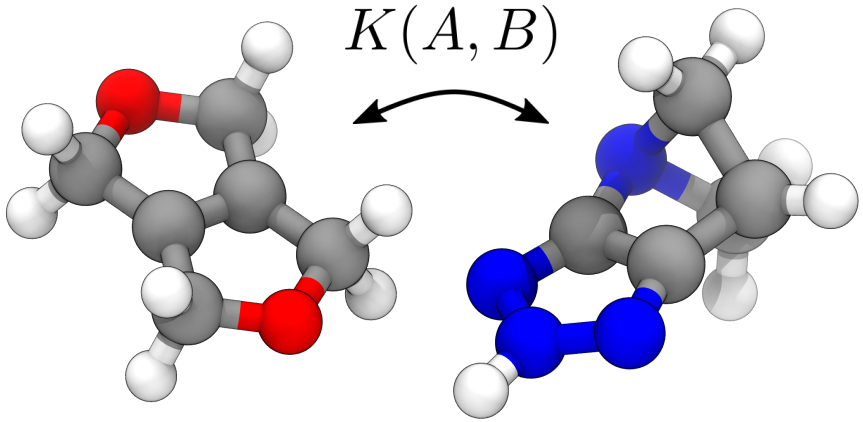
Measuring distances between materials

- 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 and vice versa



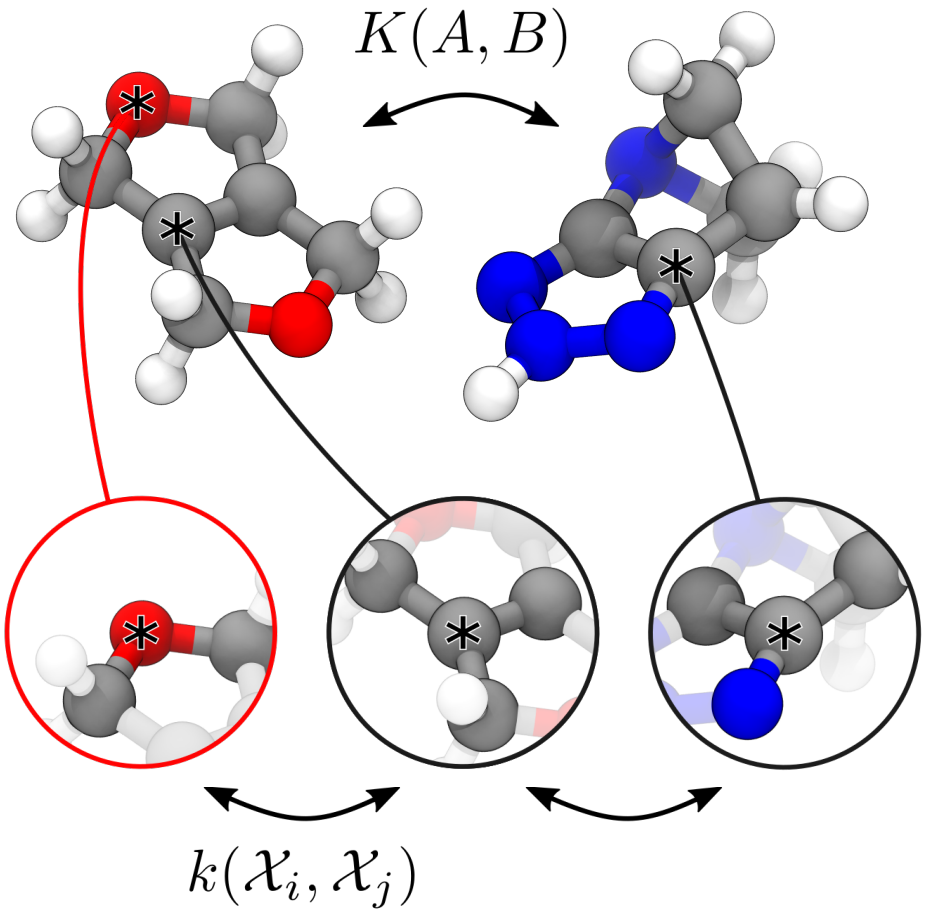
A General-Purpose Similarity Kernel

- How to compare two atomic *structures*?
- Start from a comparison of *local environments*!
- We use SOAP kernels that are smooth, and invariant to translations, rotations and permutations of identical atoms



A General-Purpose Similarity Kernel

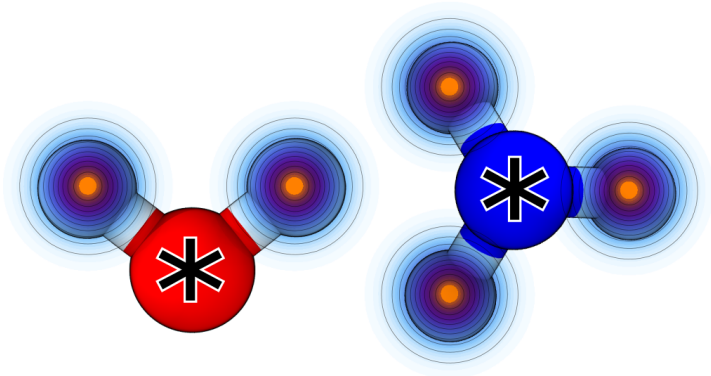
- How to compare two atomic *structures*?
- Start from a comparison of *local environments*!
- We use SOAP kernels that are smooth, and invariant to translations, rotations and permutations of identical atoms



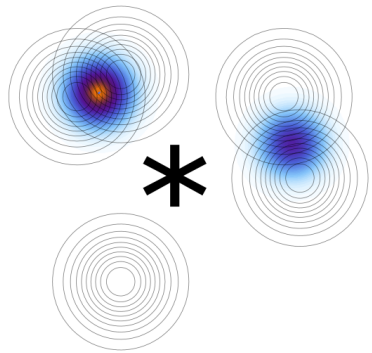
Prodan, Kohn, PNAS (2005)

A General-Purpose Similarity Kernel

- How to compare two atomic *structures*?
- Start from a comparison of *local environments*!
- We use SOAP kernels that are smooth, and invariant to translations, rotations and permutations of identical atoms



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



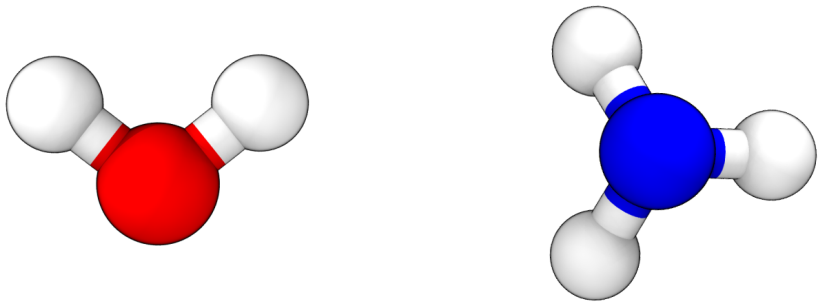
$$k(\mathcal{X}, \mathcal{X}') \sim \int \rho(\mathbf{x})\rho'(\mathbf{x})$$

Bartók, Kondor, Csányi, PRB (2013)

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

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

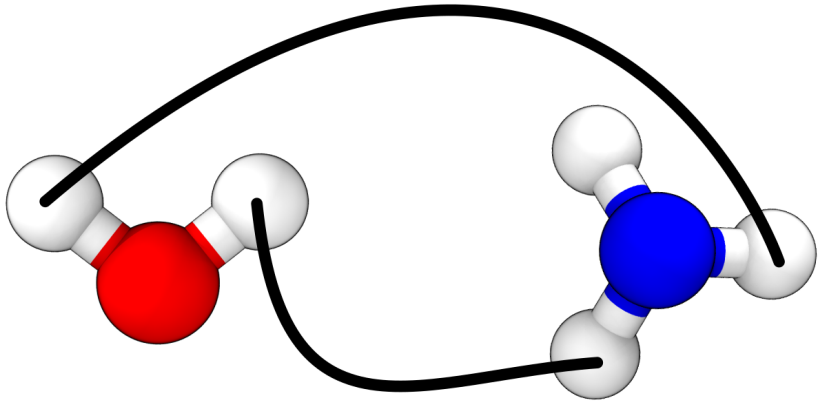


De, Bartók, Csányi, Ceriotti, PCCP (2016); M. Cuturi, NIPS (2013);

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

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



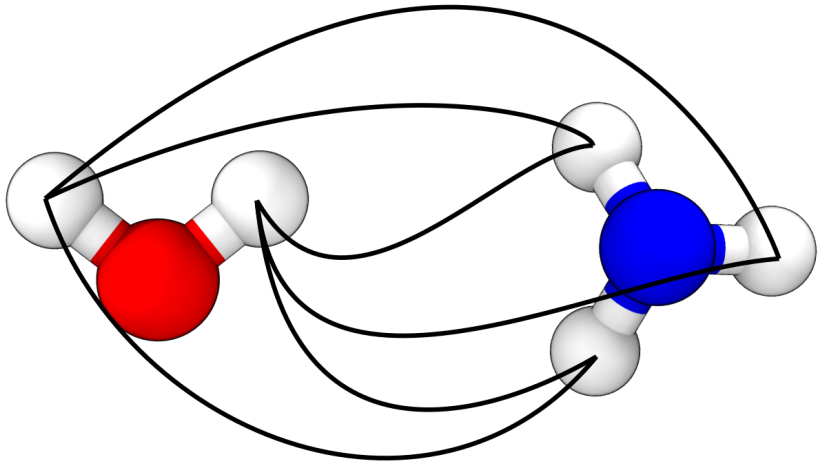
$$\hat{K}(A,B) \propto \max_{\mathbf{P}\in\mathcal{U}} \sum_{ij} P_{ji} C_{ij}^{A,B}$$

De, Bartók, Csányi, Ceriotti, 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

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



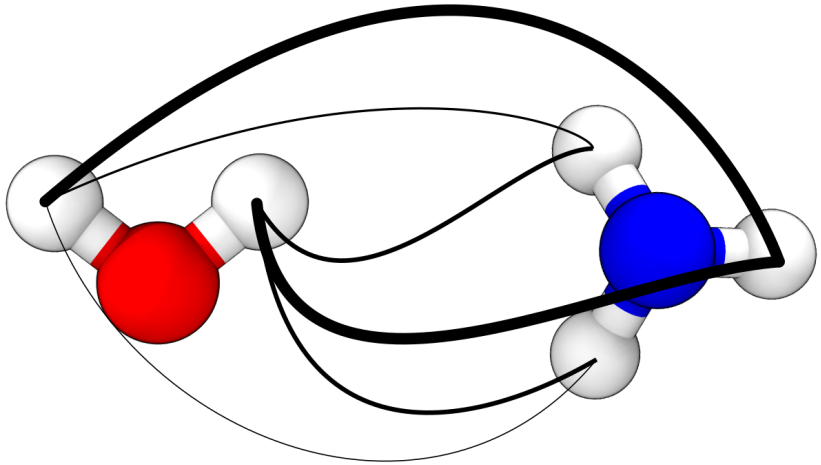
$$P_{ij} = \frac{1}{N_A N_B}$$

De, Bartók, Csányi, Ceriotti, 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

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



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

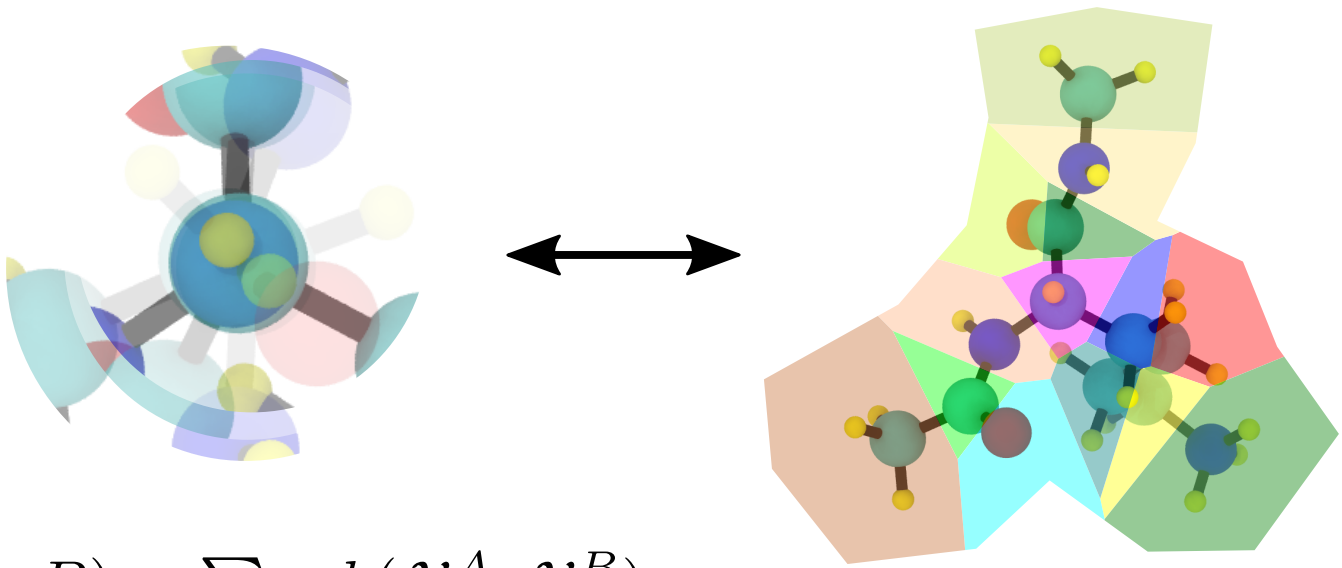
De, Bartók, Csányi, Ceriotti, PCCP (2016); M. Cuturi, NIPS (2013);

Additive Property Models & Beyond

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

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

- Entropy-regularized match provides a natural strategy to go beyond additive models



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

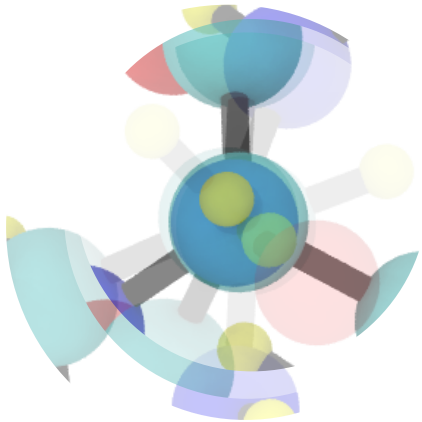
$$E(A) = \sum_i \epsilon(\mathcal{X}_i^A)$$

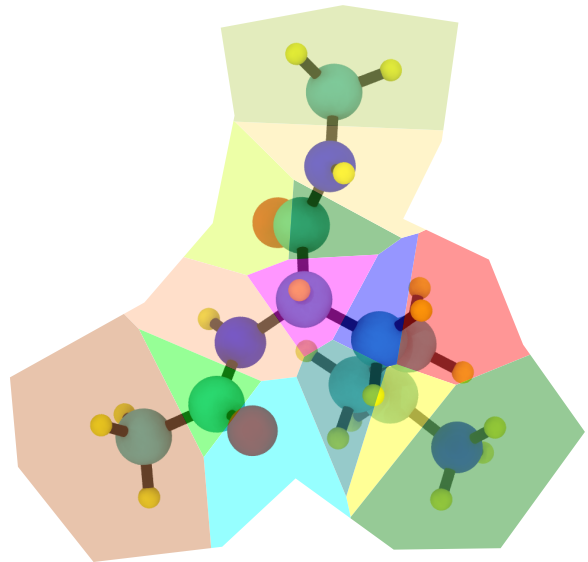
Additive Property Models & Beyond

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

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

- Entropy-regularized match provides a natural strategy to go beyond additive models



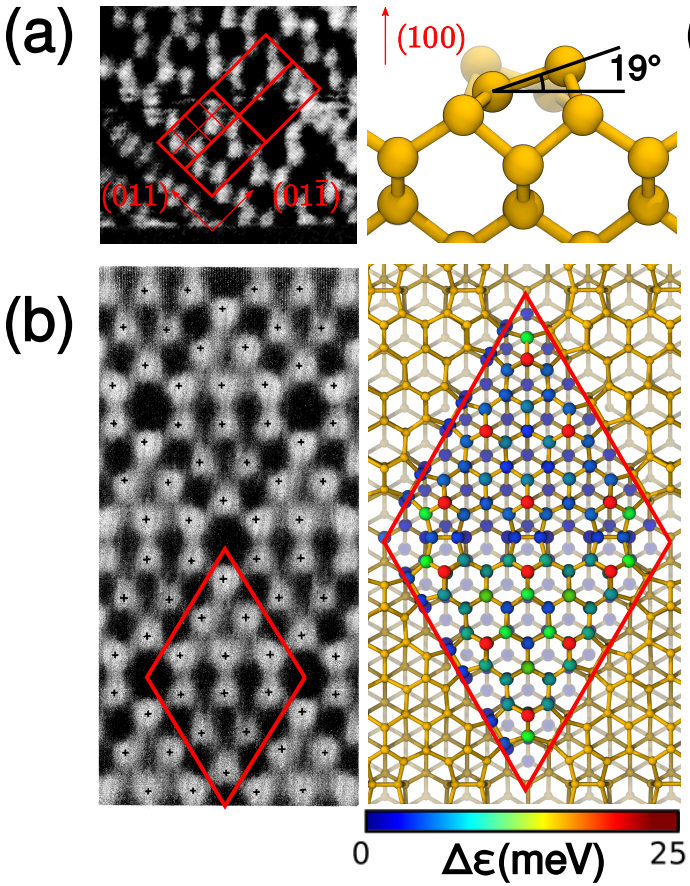


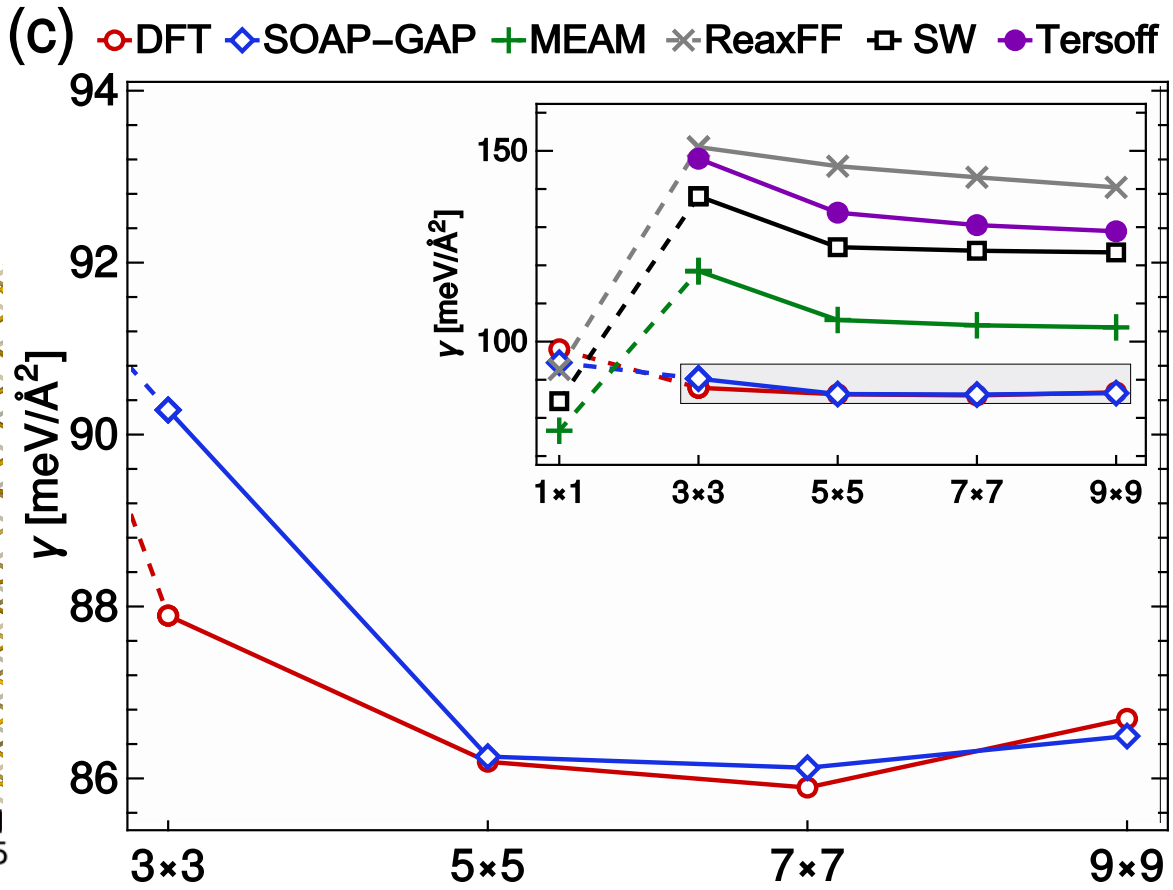
$$K(A,B) = \sum_{i,j} k(\mathcal{X}_i^A, \mathcal{X}_j^B)$$

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

Silicon Surfaces - Complexity in a Simple Material

- 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

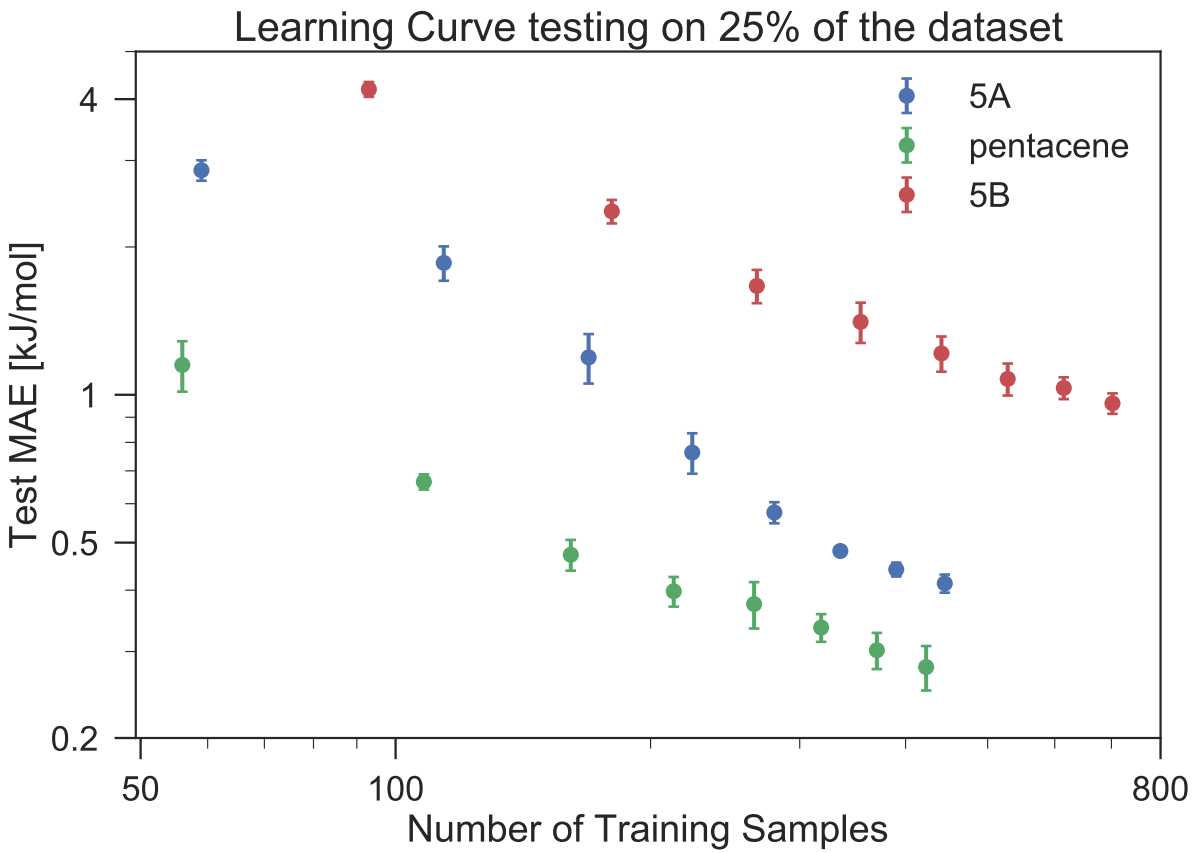


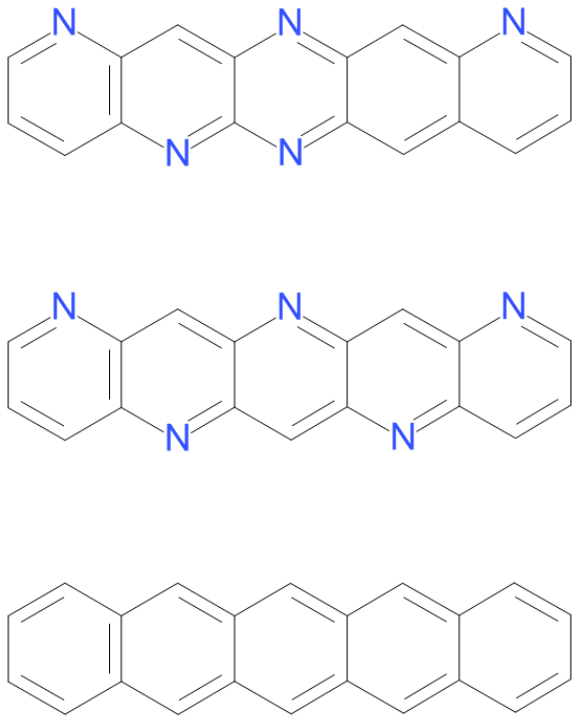


Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Accurate Predictions for Molecular Crystals

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

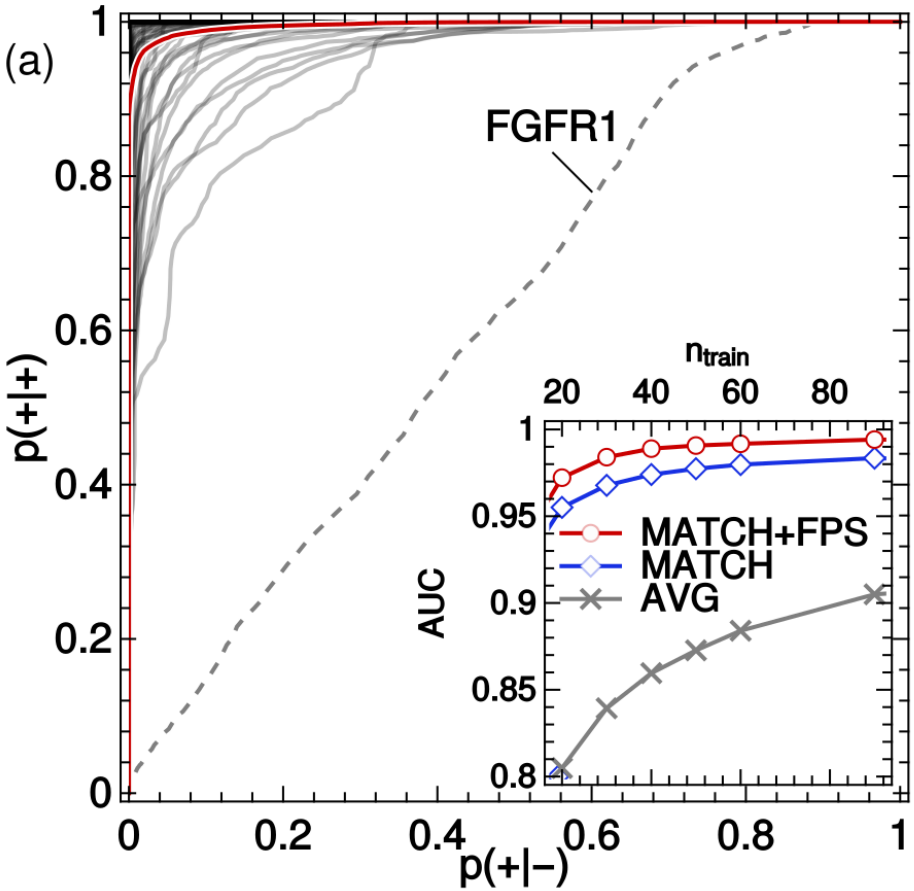


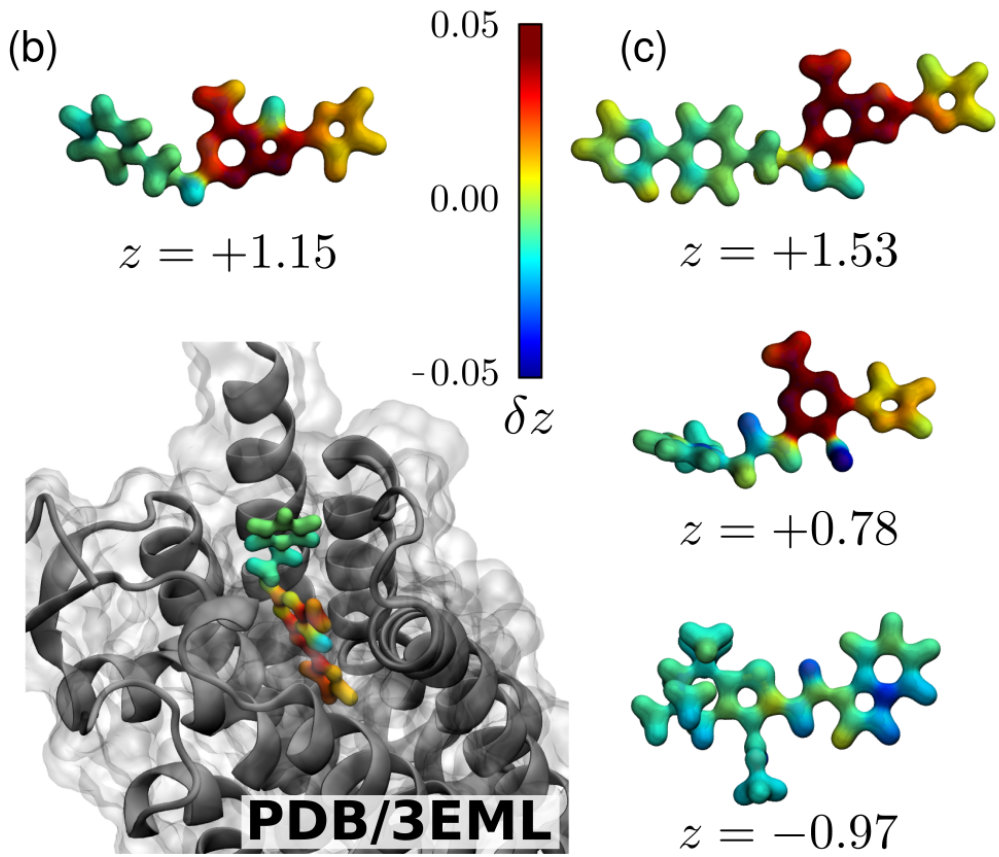


Data from G. Day and J. Yang

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

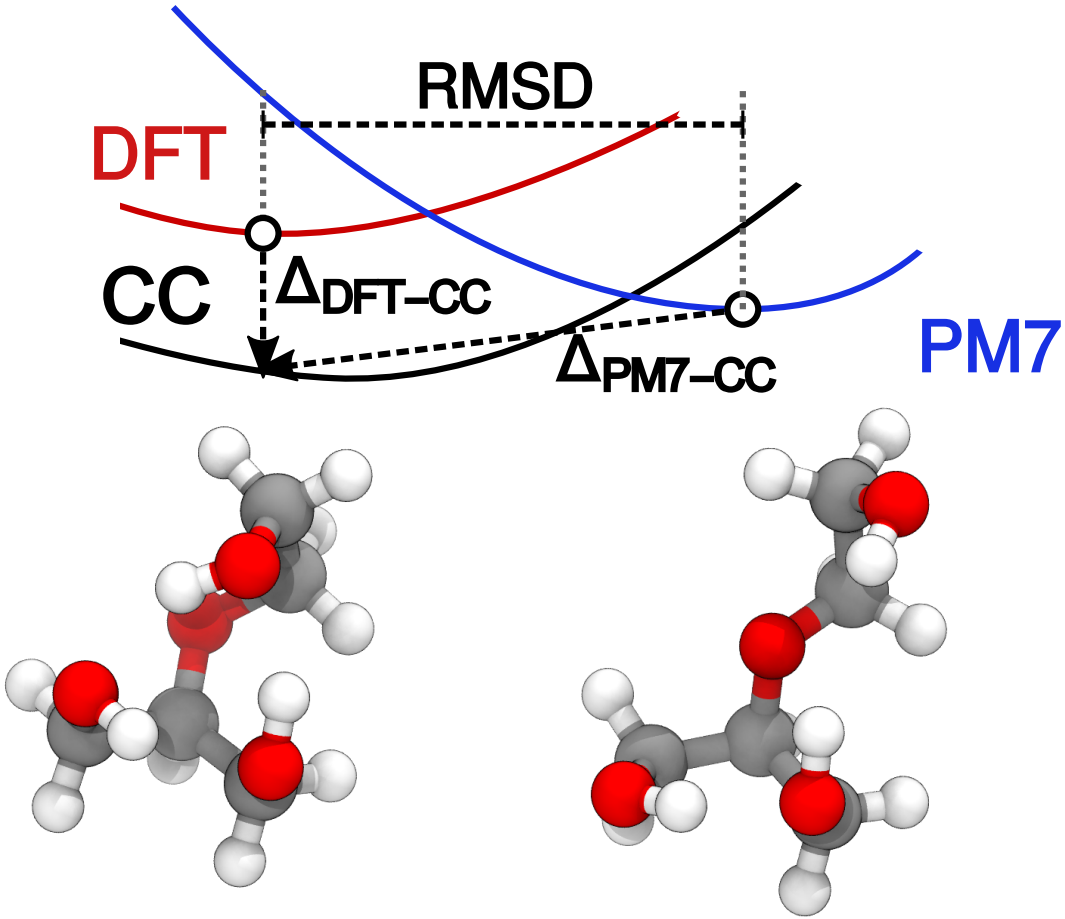




Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

100k Molecules with Coupled-Clusters

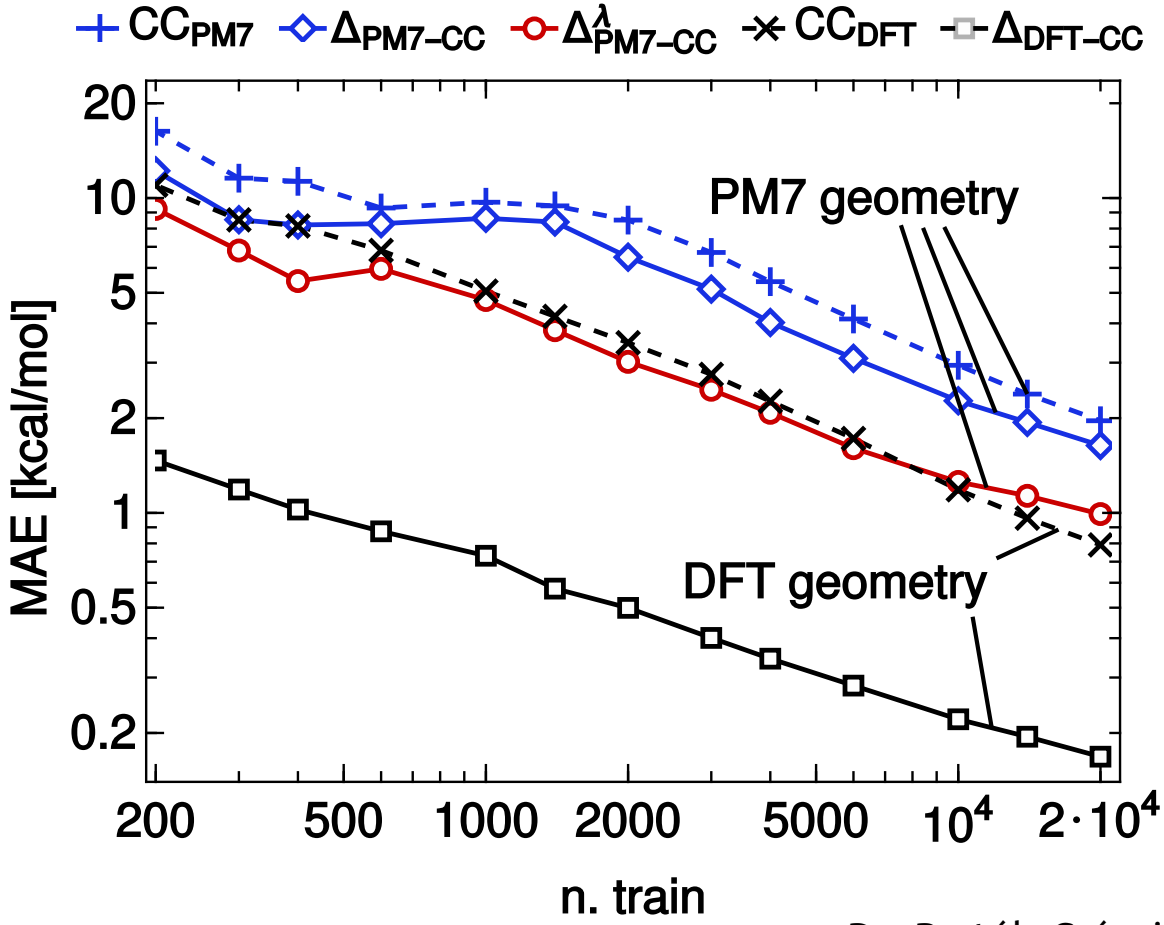
- CCSD(T) Energetics on the GDB9 - 114k *useful* predictions based on 20k training calculations
- 1kcal/mol error for predicting CCSD(T) based on PM7 geometries; 0.18kcal/mol error for predicting CCSD(T) based on DFT geometries!



Ramakrishnan et al. Scientific Data (2014)

100k Molecules with Coupled-Clusters

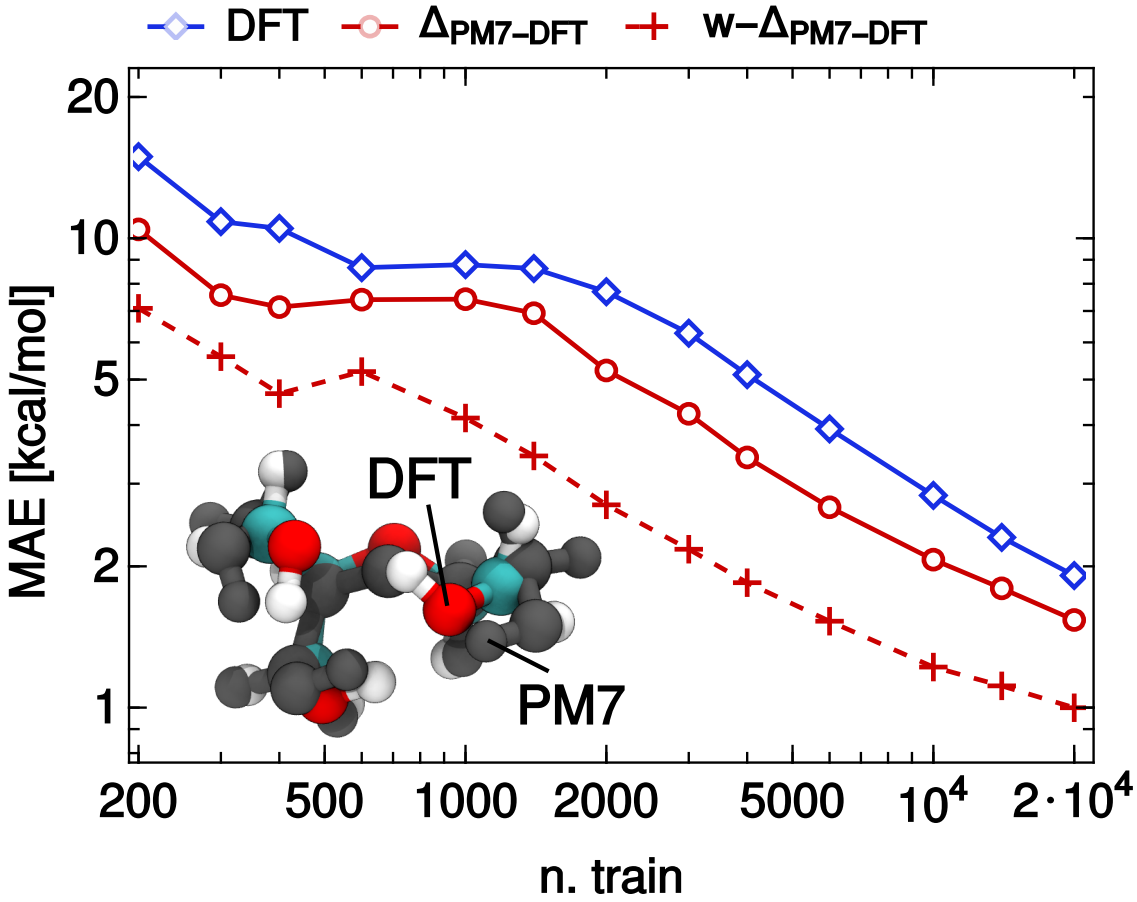
- CCSD(T) Energetics on the GDB9 - 114k *useful* predictions based on 20k training calculations
- 1kcal/mol error for predicting CCSD(T) based on PM7 geometries; 0.18kcal/mol error for predicting CCSD(T) based on DFT geometries!



De, Bartók, Csányi, Ceriotti, PCCP (2016);
Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Useful Semi-empirical Predictions

- In some cases PM7 optimization fails catastrophically and gives a different compound than DFT. These structures *hurt* predictions
- Solution: weight the structures in the train phase by using a diagonal regularization $\propto e^{-|x_{DFT}-x_{PM7}|^2/\lambda^2}$



Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

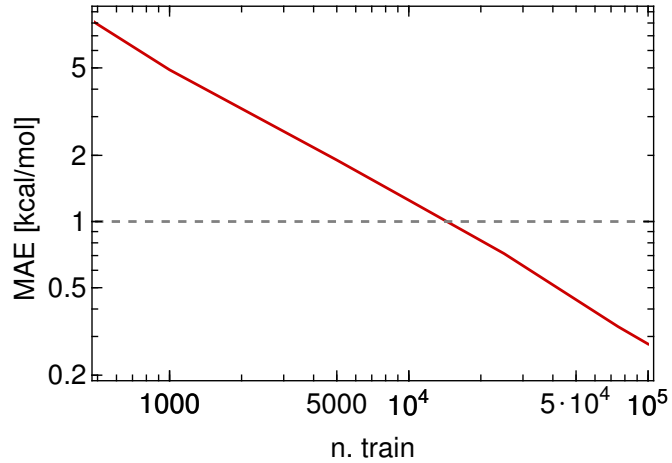
Let's Talk About Benchmarks

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

GDB9, MAE on atomization energies (eV)

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

90% TRAIN, arxiv:1702.0553



KRR/SOAP: 0.012 eV (1kJ/mol)

80% TRAIN, arxiv:1706.00179

Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Let's Talk About Benchmarks

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

QM7B, MAE on atomization energies (kcal/mol)

		E / kcal mol^{-1}	
Repr.	Kernel	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, arxiv:1706.00179

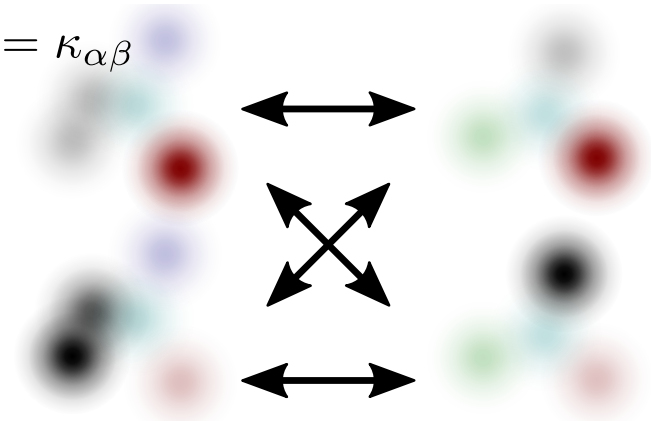
Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Let's Talk About Benchmarks

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

QM7B, MAE on atomization energies (kcal/mol)

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



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

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

KRR/SOAP: 0.40 kcal/mol

ALCHEMY: 0.33 kcal/mol

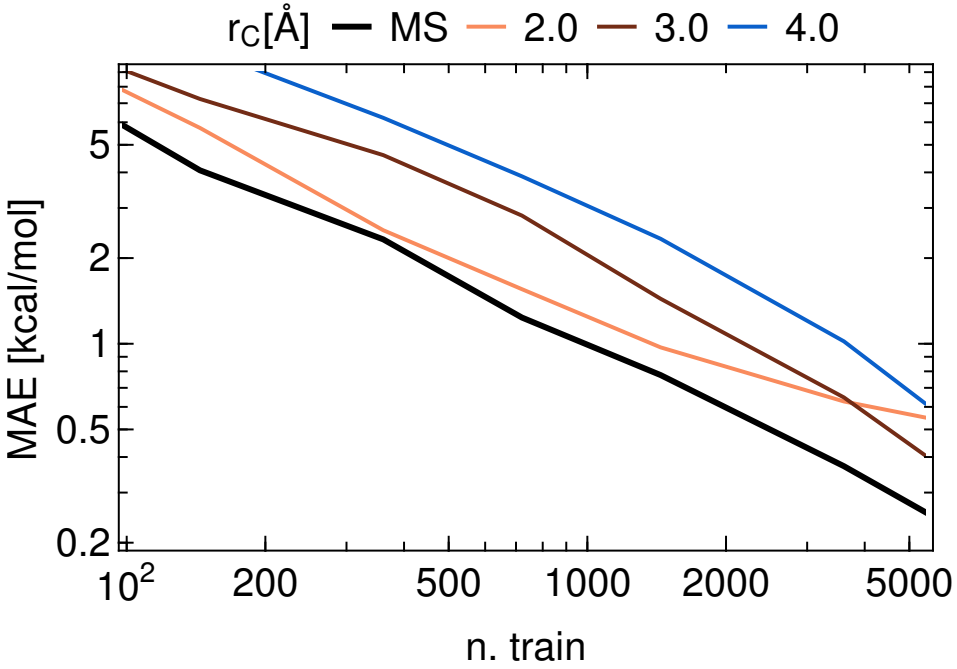
75% TRAIN, arxiv:1706.00179

Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Let's Talk About Benchmarks

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

QM7B, MAE on atomization energies (kcal/mol)



KRR/SOAP: 0.40 kcal/mol

MULTISCALE: 0.26 kcal/mol

75% TRAIN, arxiv:1706.00179

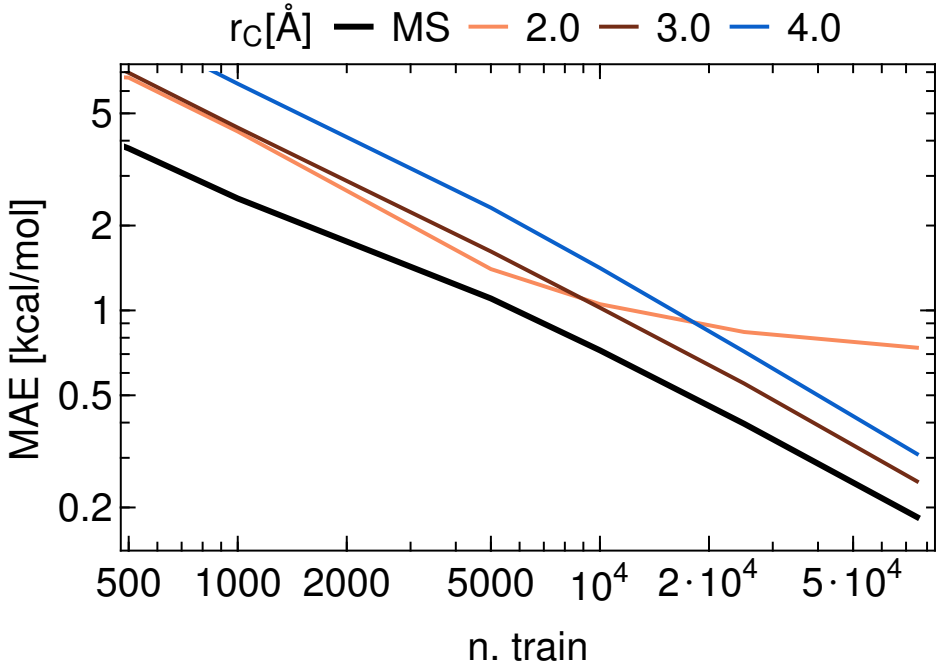
$$K(A,B) = \sum_i w_i K_i(A,B)$$

Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Let's Talk About Benchmarks

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

Multiscale for GDB9 atomization energies (kcal/mol)



KRR/SOAP: 0.27 kcal/mol

MULTISCALE: 0.18 kcal/mol

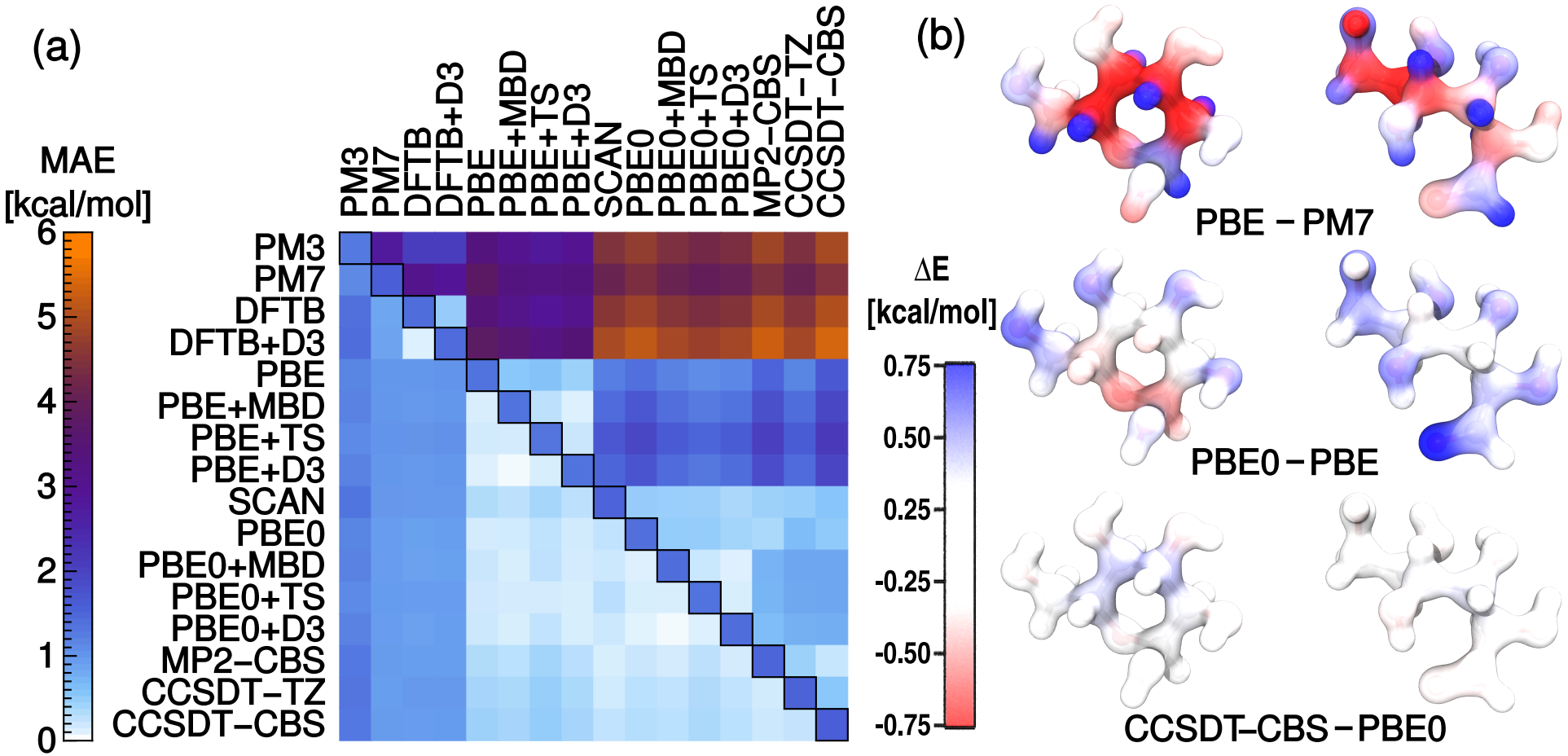
80% TRAIN, arxiv:1706.00179

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

Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Understanding Errors in Quantum Chemistry

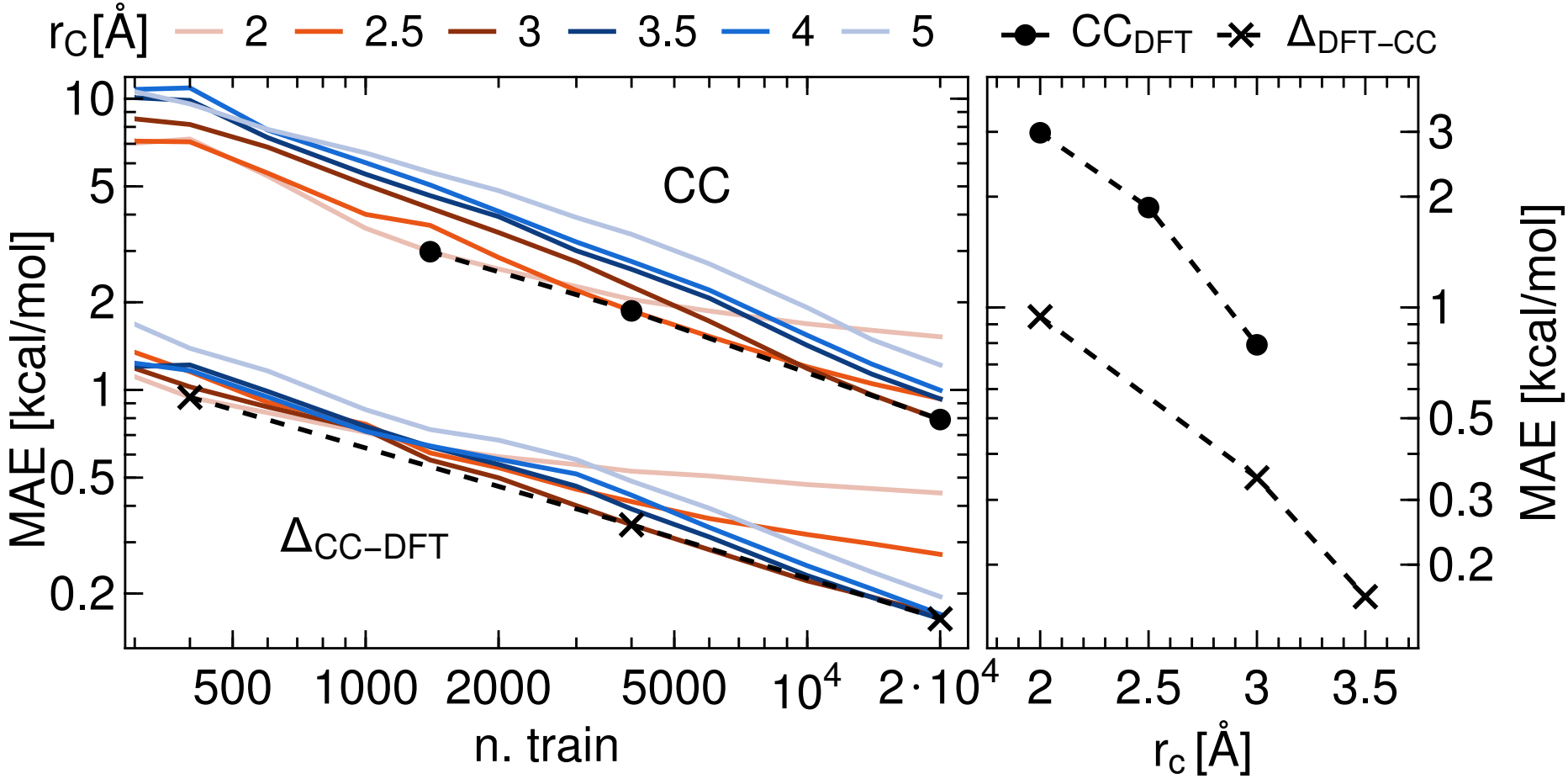
- Learning differences between electronic structure methods is simpler
- Atom-centered energetics give insight into the impact of approximations



Von Lilienfeld & C, JCTC 2015; M. Marianski et al., J. Chem. Theory Comput. 12, 6157 (2016); Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

Assessing the Range of Chemical Interactions

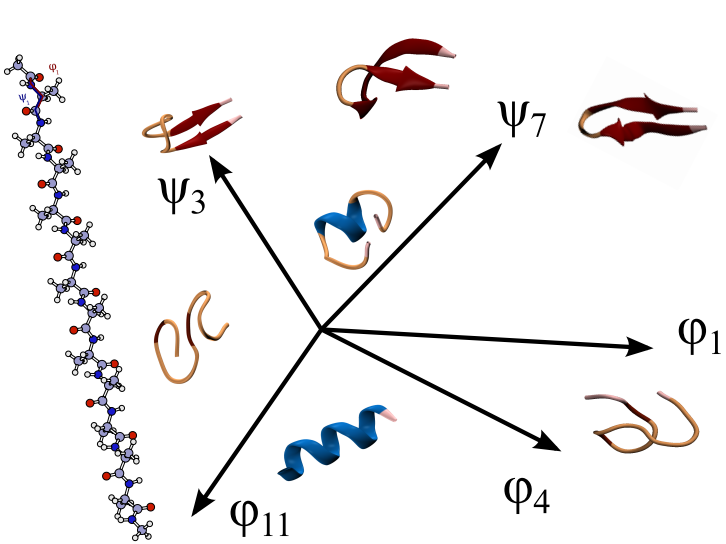
- Different SOAP ranges saturate at different levels of accuracy
- We can use the information to understand "how much" energy information is available within a given cutoff

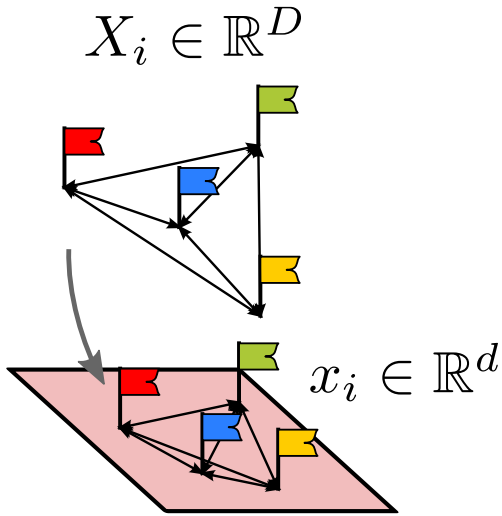


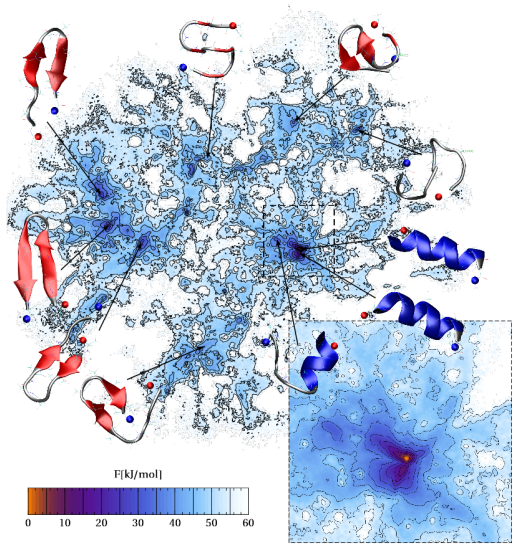
Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179

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







High-dimensional descriptors of a complex system

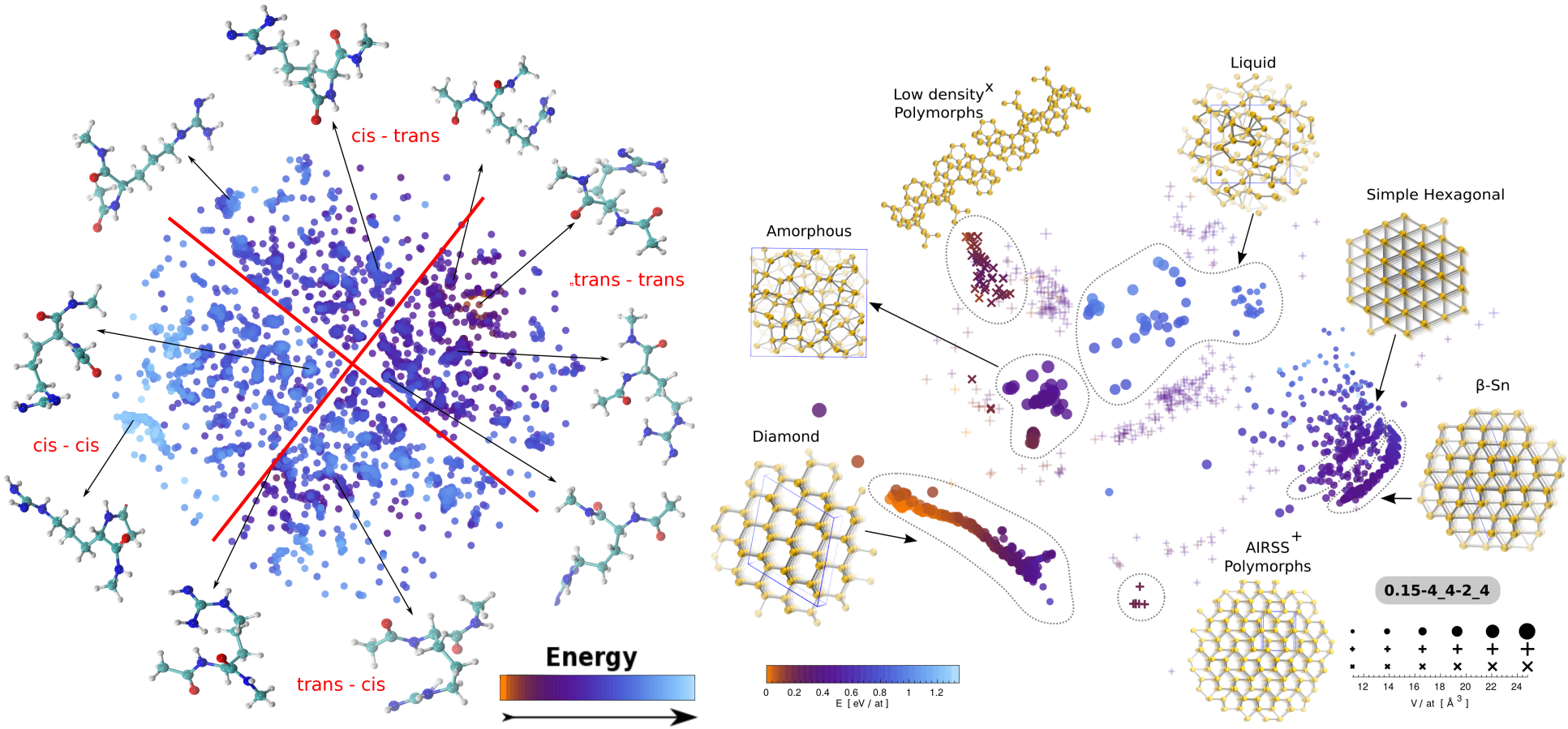
Non-linear dimensionality reduction (**sketch-map**)

Intuitive maps of structural stability

Ceriotti, Tribello, Parrinello, PNAS (2011); **interactive.sketchmap.org**

A Map to Navigate Materials & Molecules

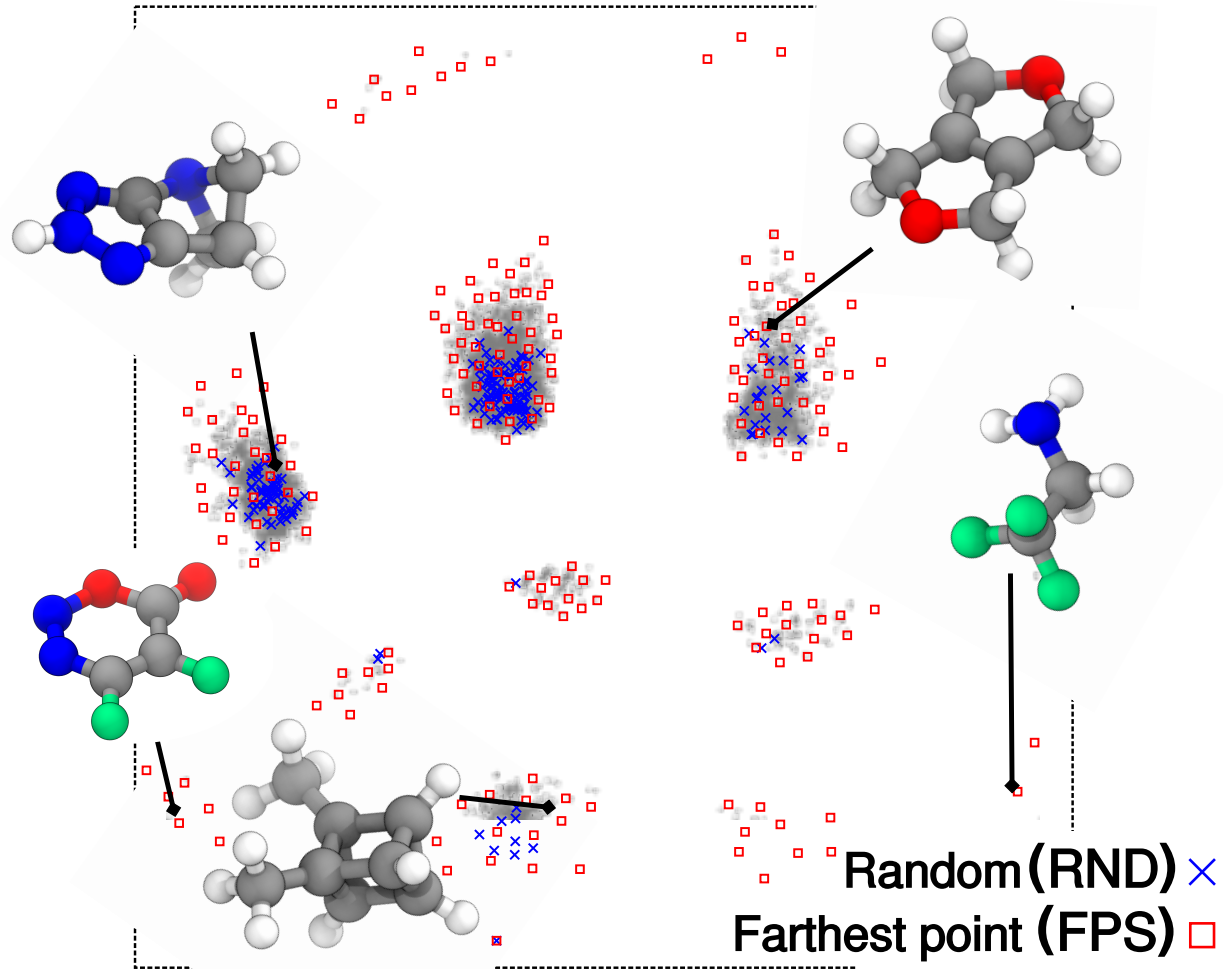
- Kernel-induced distances can be also used as the basis of clustering and dimensionality-reduction techniques
- Generate insightful representations of the (free)-energy landscape of complex systems



Ceriotti, Tribello, Parrinello, PNAS (2011); **interactive.sketchmap.org**

Train Set Selection & Active Learning

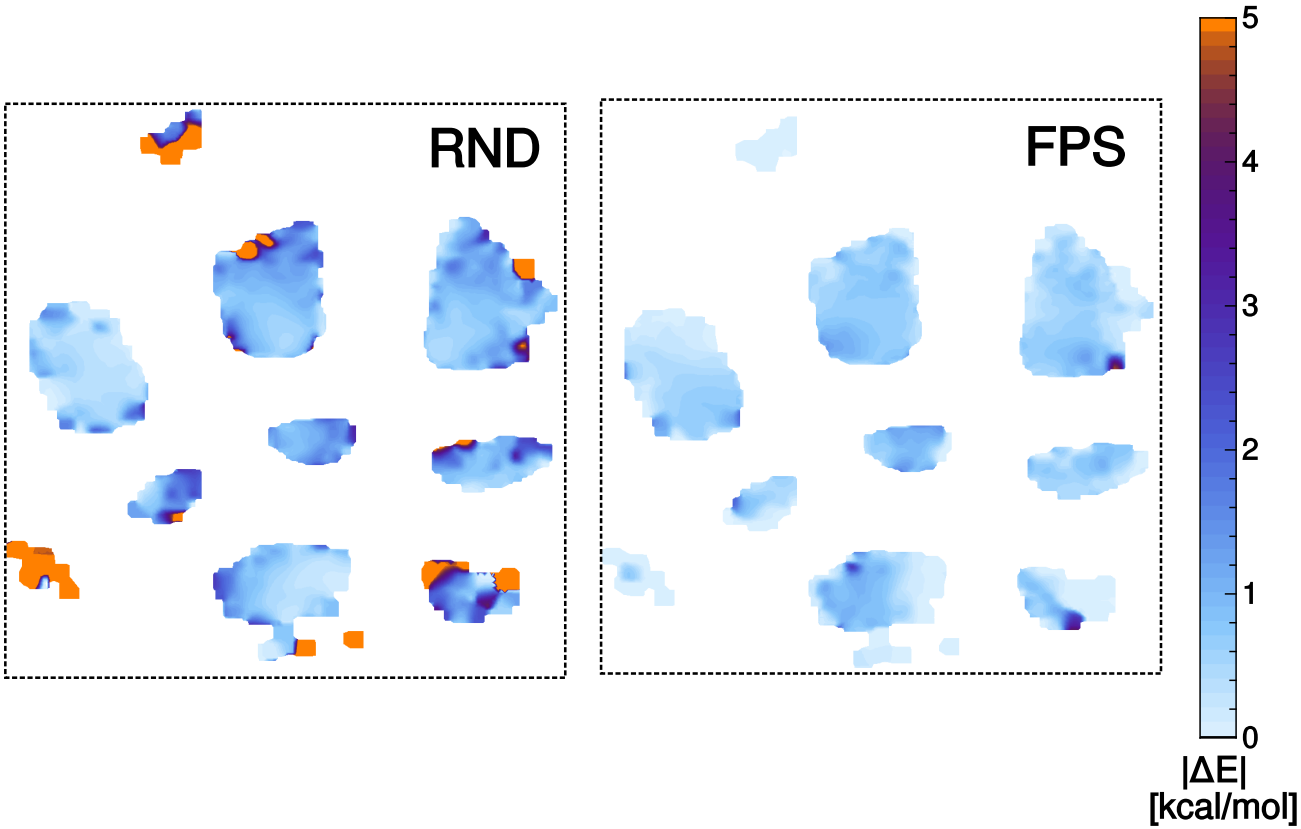
- The train set should cover uniformly the relevant space
- FPS is a simple, constructive strategy to optimize the training set, opening doors to active learning



Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179
Browning et al., JPCL 2017; Ceriotti, Tribello, Parrinello, PNAS (2011); http://sketchmap.org

Train Set Selection & Active Learning

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



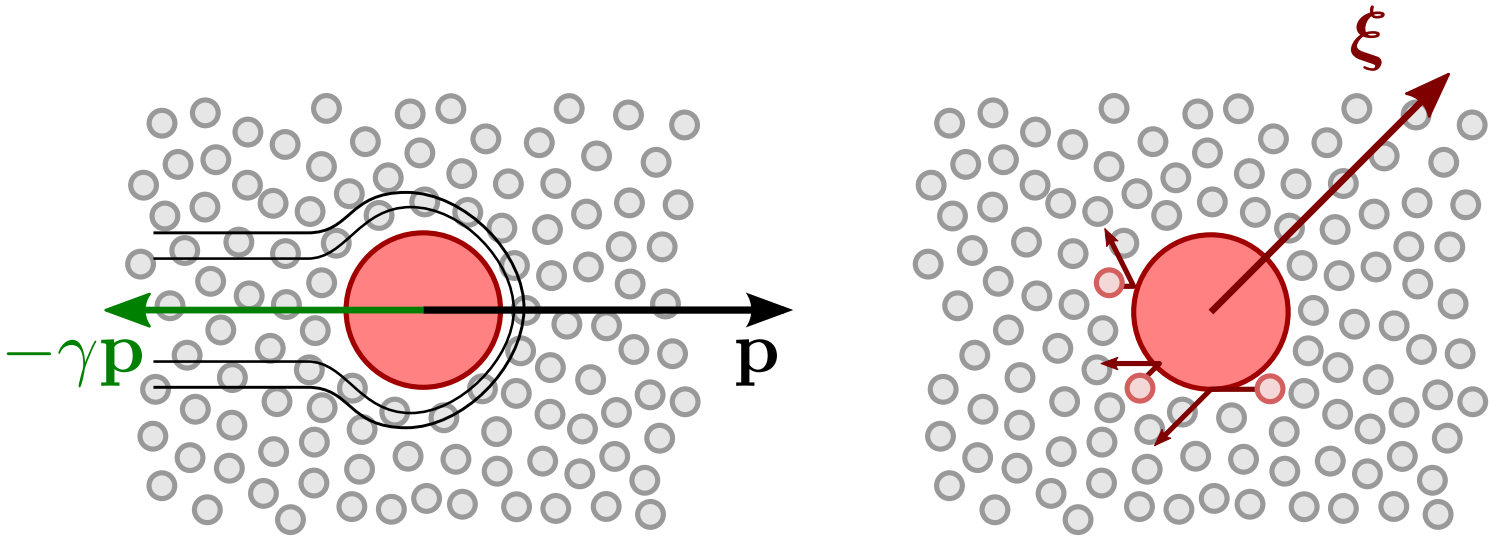
Bartok, De, Kermode, Bernstein, Csanyi, Ceriotti, arxiv:1706.00179
Browning et al., JPCL 2017; Ceriotti, Tribello, Parrinello, PNAS (2011); http://sketchmap.org

How About Dynamics?

Colored-noise Langevin dynamics

- The Langevin equation is Markovian: $\dot{p}(t)$ only depends on $p(t)$
- Can we formulate a non-Markovian generalized Langevin equation?

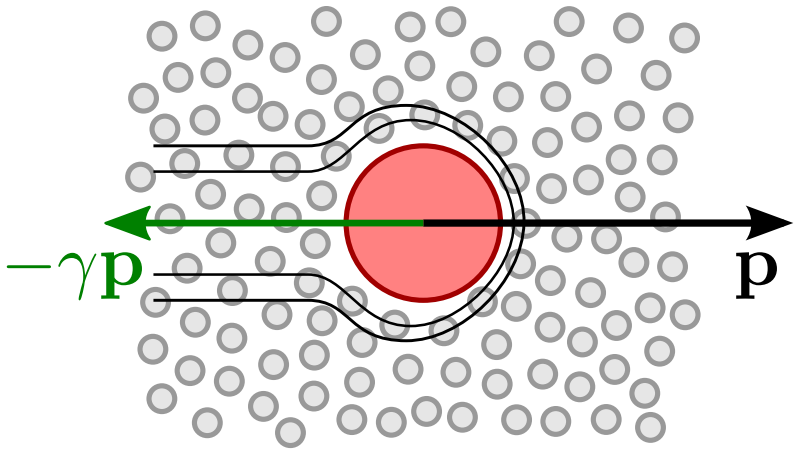
$$\dot{p}(t) = -\gamma p(t) + \sqrt{2m\gamma/\beta}\xi(t), \ \langle \xi(t)\xi(0)\rangle = \delta(t)$$

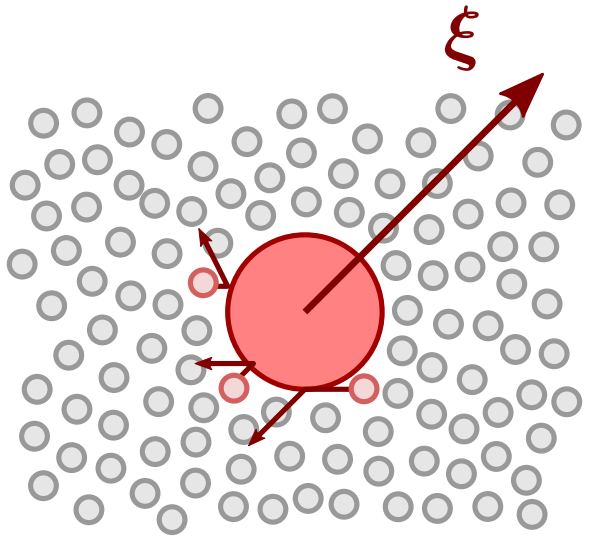


Colored-noise Langevin dynamics

- The Langevin equation is Markovian: $\dot{p}(t)$ only depends on $p(t)$
- Can we formulate a non-Markovian generalized Langevin equation?

$$\dot{p}(t) = -\int_0^\infty K(s)\, p(t-s)\, \mathrm{d}s + \sqrt{2m/\beta}\zeta(t), \quad \langle \zeta(t)\zeta(0)\rangle = K(t)$$

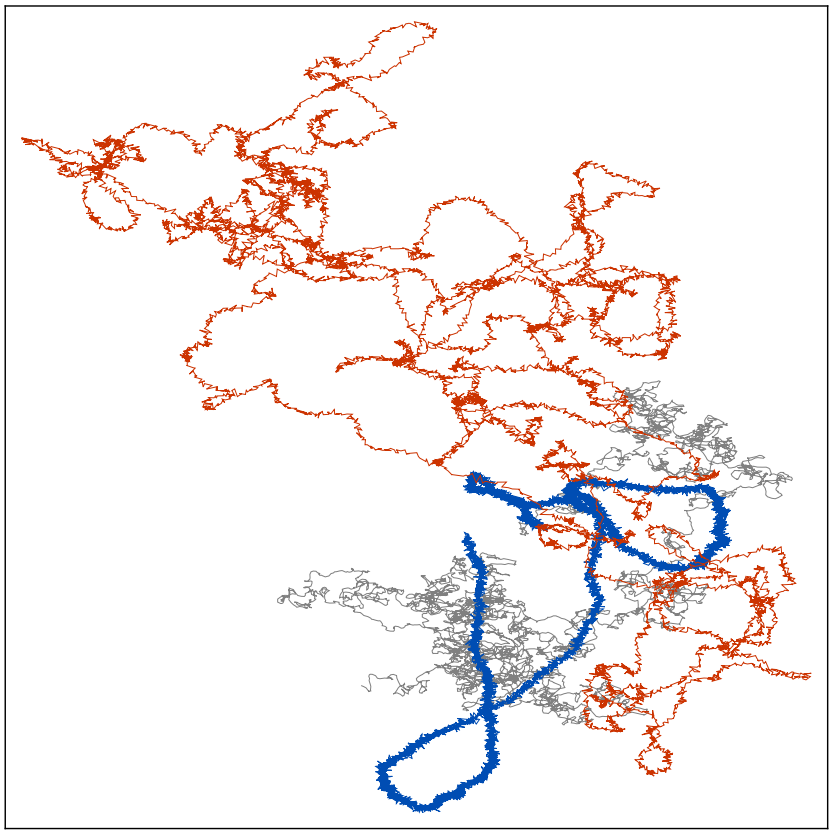




Colored-noise Langevin dynamics

- The Langevin equation is Markovian: $\dot{p}(t)$ only depends on $p(t)$
- Can we formulate a non-Markovian generalized Langevin equation?

$$\dot{p}(t) = -\int_0^\infty K(s)\, p(t-s)\, \mathrm{d}s + \sqrt{2m/\beta}\zeta(t), \quad \langle \zeta(t)\zeta(0)\rangle = K(t)$$



Details of the noise **dramatically change dynamics!**

The trajectories of a particle subject to white and colored-noise LE with the same diffusion coefficient yield very different short-times behavior.

Colored-noise Langevin dynamics

- The Langevin equation is Markovian: $\dot{p}(t)$ only depends on $p(t)$
- Can we formulate a non-Markovian generalized Langevin equation?

$$\dot{p}(t) = -\int_0^\infty K(s)\, p(t-s)\, \mathrm{d}s + \sqrt{2m/\beta}\zeta(t), \ \langle \zeta(t)\zeta(0)\rangle = K(t)$$

- Non-Markovian generalized Langevin equations are used in many contexts as a physical model
  - Can we exploit the flexibility to manipulate the sampling properties of a molecular dynamics trajectory?
  - Can we make it as general and "user-friendly" as possible?

A GLE framework for molecular dynamics

Non-Markovian GLE

$$\begin{aligned} \dot{q}(t) &= p(t)/m \\ \dot{p}(t) &= -V'(q) - \int_0^\infty K(t')\, p(t-t')\, \mathrm{d}t' + \sqrt{2mT}\,\zeta(t) \\ \langle \zeta(t)\,\zeta(0)\rangle &= TK(t) \end{aligned}$$

- We want to use a non-Markovian GLE, but it is inconvenient
- A large class of non-Markovian dynamics can be **mapped onto a Markovian dynamics** in an extended phase space
- The Markovian GLE corresponds to a (possibly complex) exponential memory kernel $K(t) = a_{pp}\delta(t) - \mathbf{a}_p^T e^{-\mathbf{A}t}\bar{\mathbf{a}}_p$
- Except for the non-linear potential, this stochastic differential equation is an **Ornstein-Uhlenbeck process** $\dot{\mathbf{u}} = -\mathbf{A}\mathbf{u} + \mathbf{B}\xi$ which can be solved analytically.

http://gle4md.org; Ceriotti, Bussi, Parrinello, Phys. Rev. Lett. 102&103 (2009)

A GLE framework for molecular dynamics

Markovian GLE

$$\begin{aligned} \dot{q}(t) &= p(t)/m \\ \begin{pmatrix} \dot{p} \\ \dot{\mathbf{s}} \end{pmatrix} &= \begin{pmatrix} -V'(q) \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} a_{pp} & \mathbf{a}_p^T \\ \bar{\mathbf{a}}_p & \mathbf{A} \end{pmatrix} \begin{pmatrix} p \\ \mathbf{s} \end{pmatrix} + \mathbf{B}\xi \\ \mathbf{B}\mathbf{B}^T &= T\left(\mathbf{A} + \mathbf{A}^T\right) \end{aligned}$$

- We want to use a non-Markovian GLE, but it is inconvenient
- A large class of non-Markovian dynamics can be **mapped onto a Markovian dynamics** in an extended phase space
- The Markovian GLE corresponds to a (possibly complex) exponential memory kernel $K(t) = a_{pp}\delta(t) - \mathbf{a}_p^T e^{-\mathbf{A}t}\bar{\mathbf{a}}_p$
- Except for the non-linear potential, this stochastic differential equation is an **Ornstein-Uhlenbeck process** $\dot{\mathbf{u}} = -\mathbf{A}\mathbf{u} + \mathbf{B}\xi$ which can be solved analytically.

http://gle4md.org; Ceriotti, Bussi, Parrinello, Phys. Rev. Lett. 102&103 (2009)

A GLE framework for molecular dynamics

Markovian GLE

$$\begin{aligned} \dot{q}(t) &= p(t)/m \\ \begin{pmatrix} \dot{p} \\ \dot{\mathbf{s}} \end{pmatrix} &= \begin{pmatrix} -V'(q) \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} a_{pp} & \mathbf{a}_p^T \\ \bar{\mathbf{a}}_p & \mathbf{A} \end{pmatrix} \begin{pmatrix} p \\ \mathbf{s} \end{pmatrix} + \mathbf{B}\xi \\ \mathbf{B}\mathbf{B}^T &= T\left(\mathbf{A} + \mathbf{A}^T\right) \end{aligned}$$

- We want to use a non-Markovian GLE, but it is inconvenient
- A large class of non-Markovian dynamics can be **mapped onto a Markovian dynamics** in an extended phase space
- The Markovian GLE corresponds to a (possibly complex) exponential memory kernel $K(t) = a_{pp}\delta(t) - \mathbf{a}_p^T e^{-\mathbf{A}t}\bar{\mathbf{a}}_p$
- Except for the non-linear potential, this stochastic differential equation is an **Ornstein-Uhlenbeck process** $\dot{\mathbf{u}} = -\mathbf{A}\mathbf{u} + \mathbf{B}\xi$ which can be solved analytically.

http://gle4md.org; Ceriotti, Bussi, Parrinello, Phys. Rev. Lett. 102&103 (2009)

A GLE framework for molecular dynamics

Markovian GLE

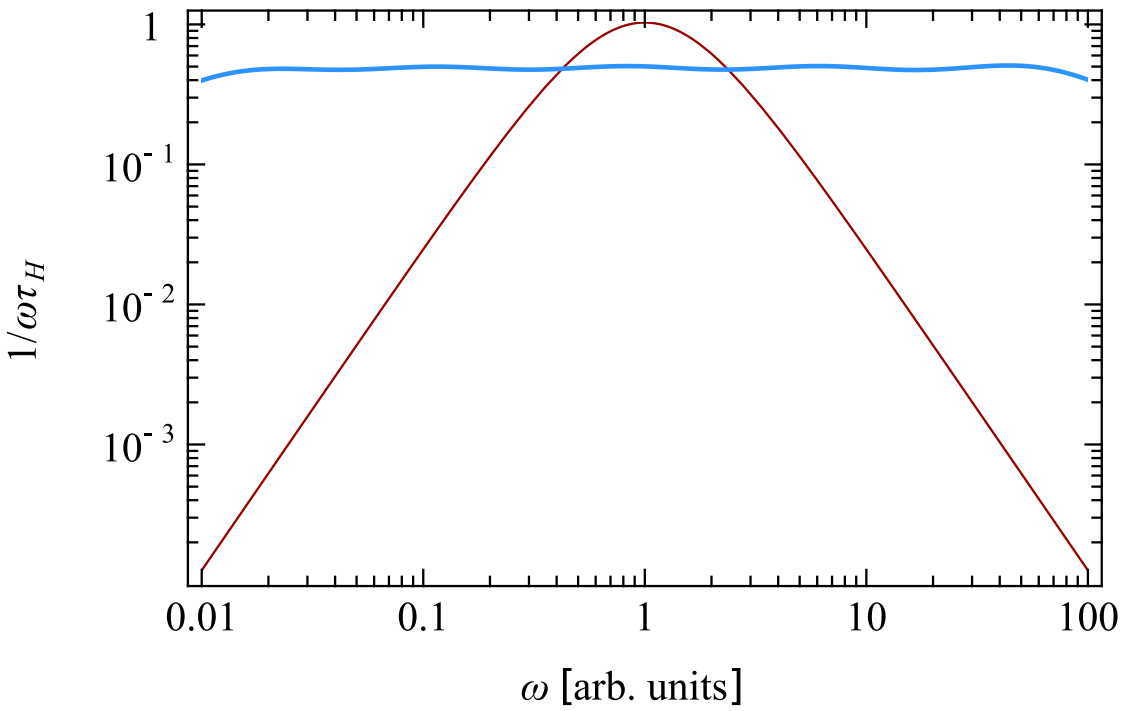
$$\begin{aligned} \dot{q}(t) &= p(t)/m \\ \begin{pmatrix} \dot{p} \\ \dot{\mathbf{s}} \end{pmatrix} &= \begin{pmatrix} -V'(q) \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} a_{pp} & \mathbf{a}_p^T \\ \bar{\mathbf{a}}_p & \mathbf{A} \end{pmatrix} \begin{pmatrix} p \\ \mathbf{s} \end{pmatrix} + \mathbf{B}\xi \\ \mathbf{B}\mathbf{B}^T &= T\left(\mathbf{A} + \mathbf{A}^T\right) \end{aligned}$$

- We want to use a non-Markovian GLE, but it is inconvenient
- A large class of non-Markovian dynamics can be **mapped onto a Markovian dynamics** in an extended phase space
- The Markovian GLE corresponds to a (possibly complex) exponential memory kernel $K(t) = a_{pp}\delta(t) - \mathbf{a}_p^T e^{-\mathbf{A}t}\bar{\mathbf{a}}_p$
- Except for the non-linear potential, this stochastic differential equation is an **Ornstein-Uhlenbeck process** $\dot{\mathbf{u}} = -\mathbf{A}\mathbf{u} + \mathbf{B}\xi$ which can be solved analytically.

http://gle4md.org; Ceriotti, Bussi, Parrinello, Phys. Rev. Lett. 102&103 (2009)

A Thermostat in All Colors

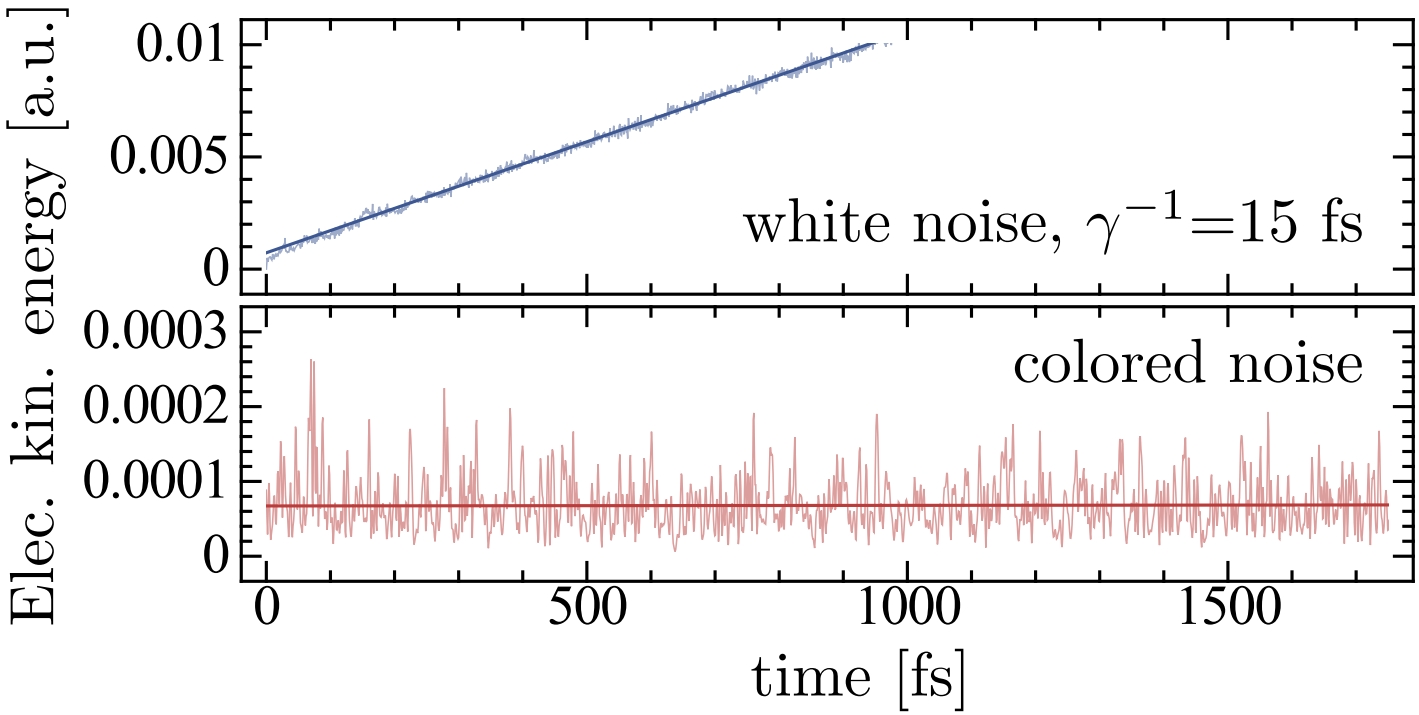
- GLEs give exquisite control over canonical sampling
- Efficient sampling for forcefield and ab initio molecular dynamics
- Controlling resonances in multiple-time step methods
- Frequency-dependent thermalization
- Accelerate modelling of nuclear quantum effects



Ceriotti, Bussi, Parrinello, PRL 020601 (2009)

A Thermostat in All Colors

- GLEs give exquisite control over canonical sampling
- Efficient sampling for forcefield and ab initio molecular dynamics
- Controlling resonances in multiple-time step methods
- Frequency-dependent thermalization
- Accelerate modelling of nuclear quantum effects



Ceriotti, Bussi, Parrinello, PRL 020601 (2009)

A Thermostat in All Colors

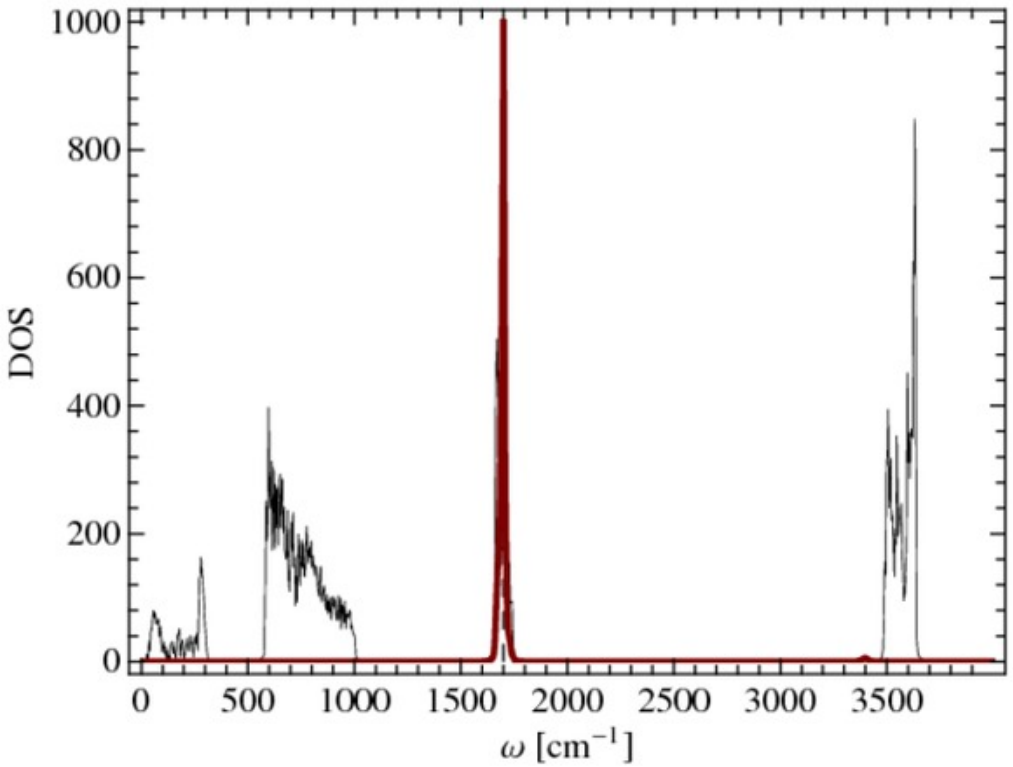
- GLEs give exquisite control over canonical sampling
- Efficient sampling for forcefield and ab initio molecular dynamics
- Controlling resonances in multiple-time step methods
- Frequency-dependent thermalization
- Accelerate modelling of nuclear quantum effects

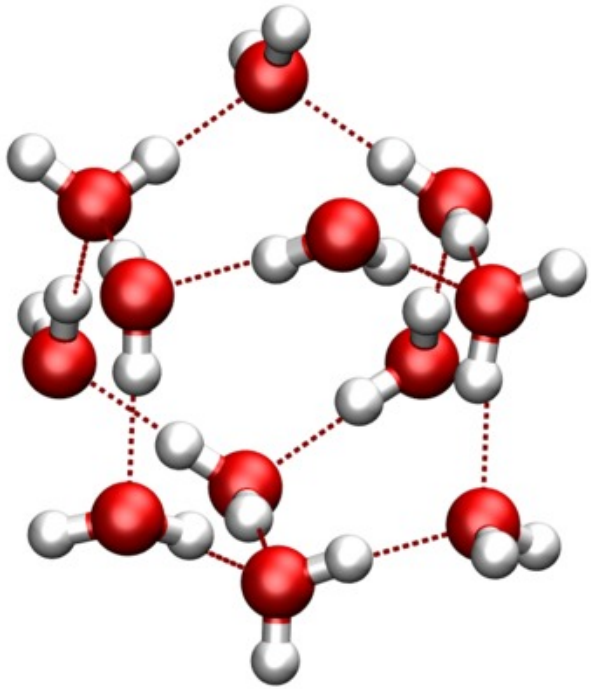
	$\langle T \rangle / K$	$\langle V_b \rangle$ /kcal/mol	$D/\text{Å}^2/\text{ps}$
REF	300.0	0.5227	0.25
White noise, 12 fs	301.7	0.5278	0.095
White noise, 20 fs	302.3	0.5246	0.019
GLE, 12 fs	301.5	0.5245	0.24
GLE, 20 fs	302.0	0.5236	0.20

Morrone, Markland, Ceriotti & Berne, JCP (2011).

A Thermostat in All Colors

- GLEs give exquisite control over canonical sampling
- Efficient sampling for forcefield and ab initio molecular dynamics
- Controlling resonances in multiple-time step methods
- Frequency-dependent thermalization
- Accelerate modelling of nuclear quantum effects

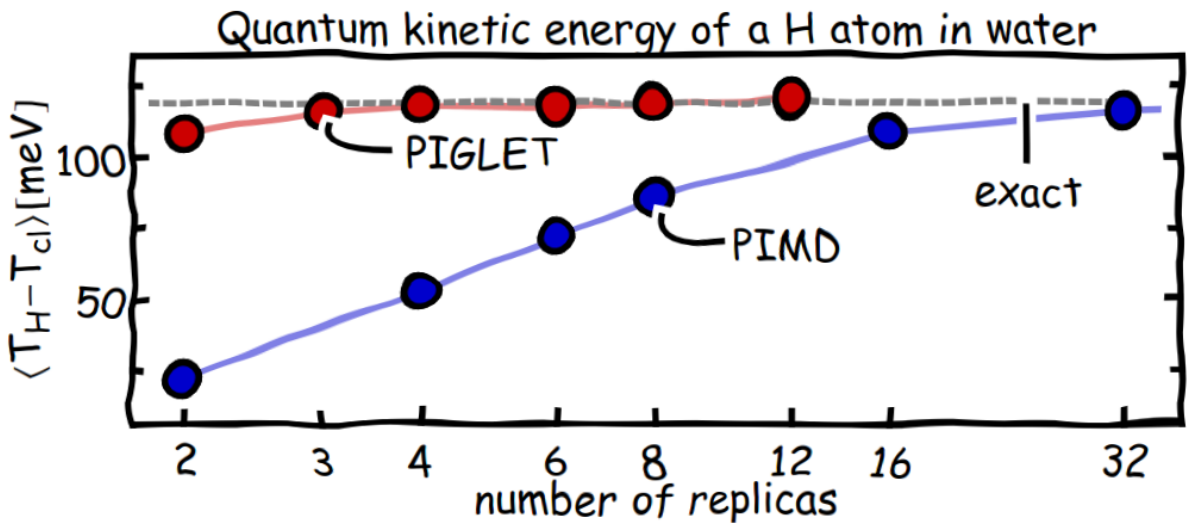


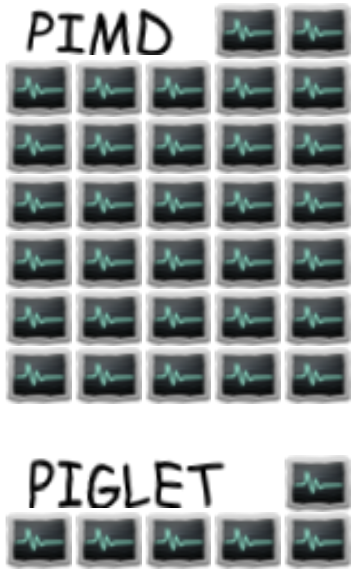


Dettori, Ceriotti, Melis, Hunger, Colombo, Donadio, JCTC (2017).

A Thermostat in All Colors

- GLEs give exquisite control over canonical sampling
- Efficient sampling for forcefield and ab initio molecular dynamics
- Controlling resonances in multiple-time step methods
- Frequency-dependent thermalization
- Accelerate modelling of nuclear quantum effects

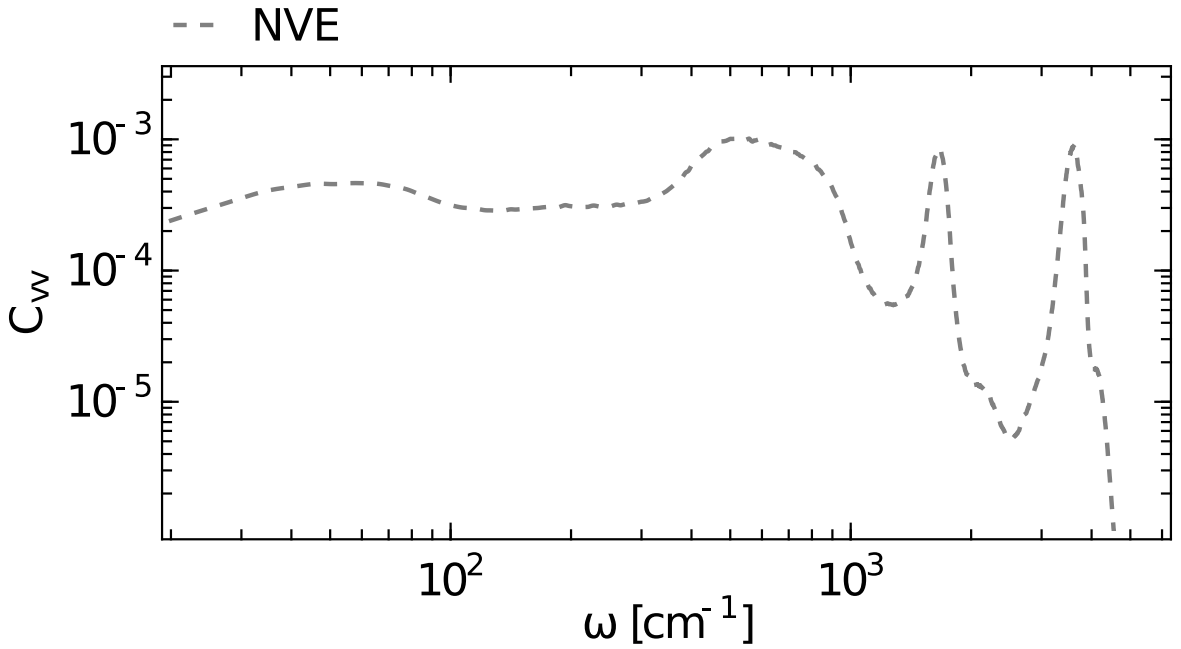




Ceriotti, Manolopoulos, PRL (2012)

Controlling and Correcting Dynamics

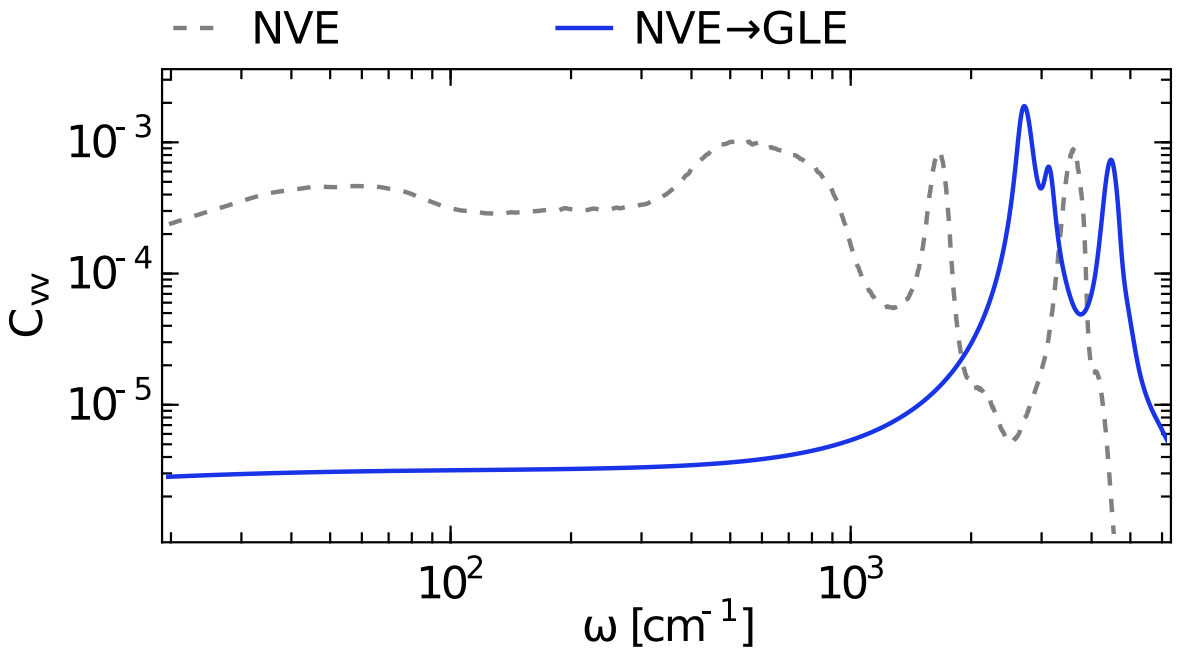
- Can we control (& correct) the disturbance introduced by (Generalized) Langevin dynamics on top of microcanonical MD?
- We can predict the spectrum of a GLE-thermostatted HO of frequency ω_0, and write the perturbed power spectrum as a convolution
  $c_{vv}^{\mathrm{GLE}}(\omega) = \int \mathrm{d}\omega_0 \mathcal{C}(\omega, \omega_0)\, c_{vv}^{\mathrm{NVE}}(\omega_0)$
- It is also possible to deconvolute a (G)LE spectrum: great potential to extract dynamical properties from thermostatted MD



M. Rossi, V. Kapil, Ceriotti, JCP (2018)

Controlling and Correcting Dynamics

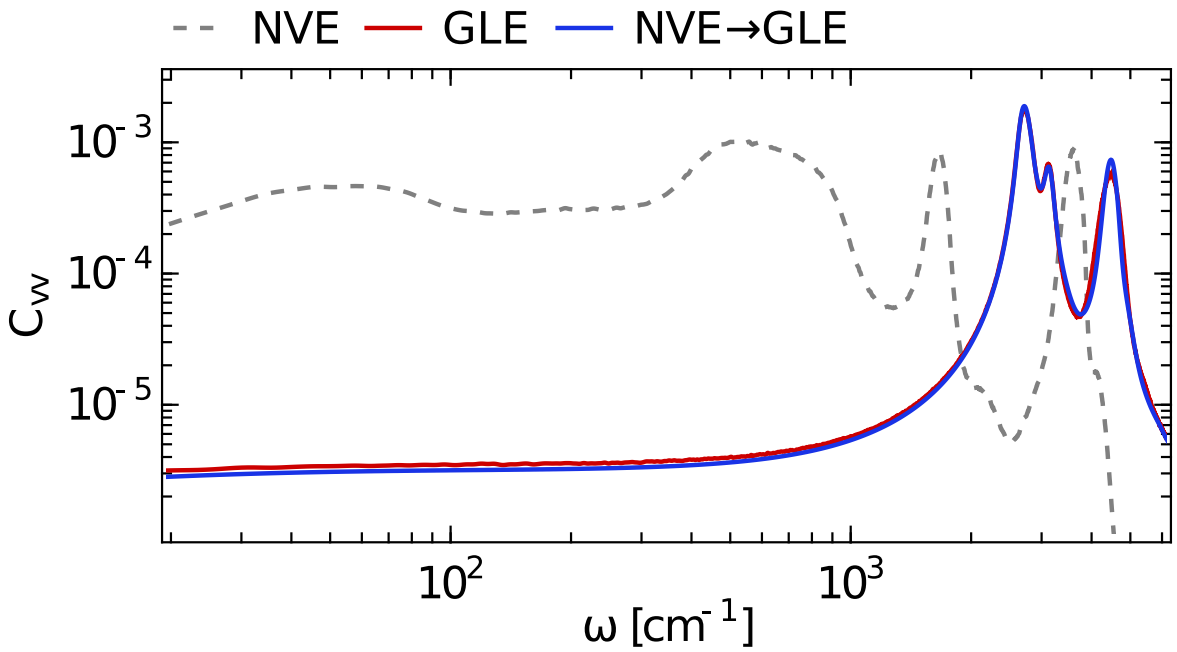
- Can we control (& correct) the disturbance introduced by (Generalized) Langevin dynamics on top of microcanonical MD?
- We can predict the spectrum of a GLE-thermostatted HO of frequency ω_0, and write the perturbed power spectrum as a convolution
  $c_{vv}^{\mathrm{GLE}}(\omega) = \int \mathrm{d}\omega_0 \mathcal{C}(\omega, \omega_0)\, c_{vv}^{\mathrm{NVE}}(\omega_0)$
- It is also possible to deconvolute a (G)LE spectrum: great potential to extract dynamical properties from thermostatted MD



M. Rossi, V. Kapil, Ceriotti, JCP (2018)

Controlling and Correcting Dynamics

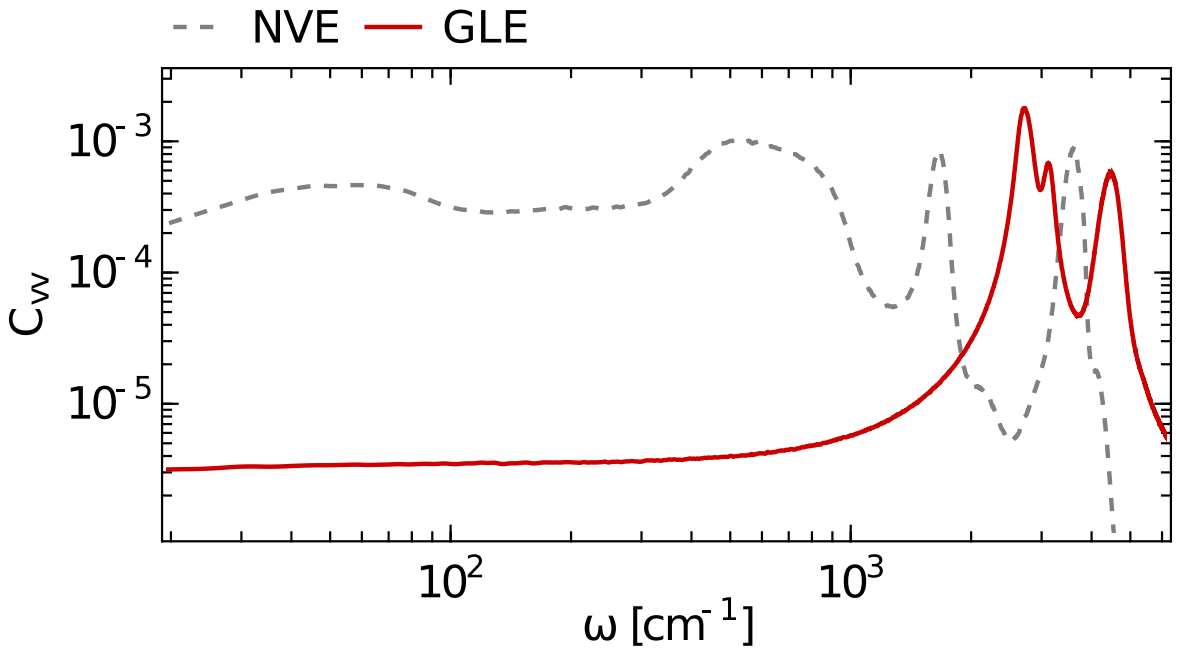
- Can we control (& correct) the disturbance introduced by (Generalized) Langevin dynamics on top of microcanonical MD?
- We can predict the spectrum of a GLE-thermostatted HO of frequency ω_0, and write the perturbed power spectrum as a convolution
  $c_{vv}^{\mathrm{GLE}}(\omega) = \int \mathrm{d}\omega_0 \mathcal{C}(\omega, \omega_0)\, c_{vv}^{\mathrm{NVE}}(\omega_0)$
- It is also possible to deconvolute a (G)LE spectrum: great potential to extract dynamical properties from thermostatted MD



M. Rossi, V. Kapil, Ceriotti, JCP (2018)

Controlling and Correcting Dynamics

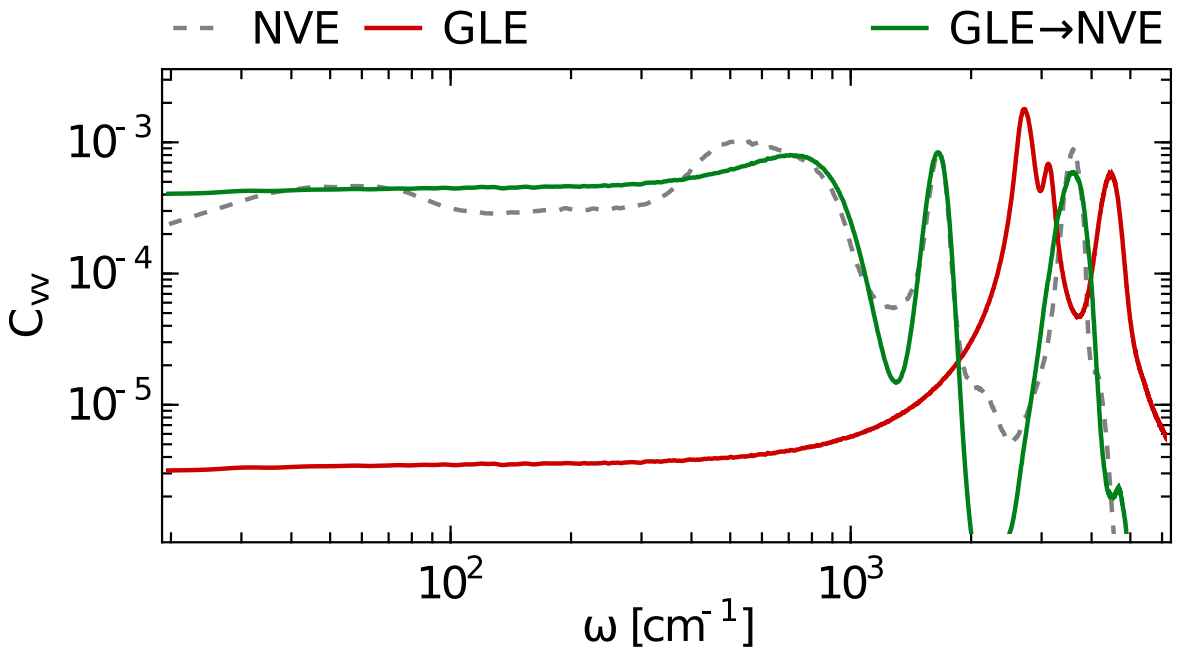
- Can we control (& correct) the disturbance introduced by (Generalized) Langevin dynamics on top of microcanonical MD?
- We can predict the spectrum of a GLE-thermostatted HO of frequency ω_0, and write the perturbed power spectrum as a convolution
  $c_{vv}^{\mathrm{GLE}}(\omega) = \int \mathrm{d}\omega_0 \mathcal{C}(\omega, \omega_0)\, c_{vv}^{\mathrm{NVE}}(\omega_0)$
- It is also possible to deconvolute a (G)LE spectrum: great potential to extract dynamical properties from thermostatted MD



M. Rossi, V. Kapil, Ceriotti, JCP (2018)

Controlling and Correcting Dynamics

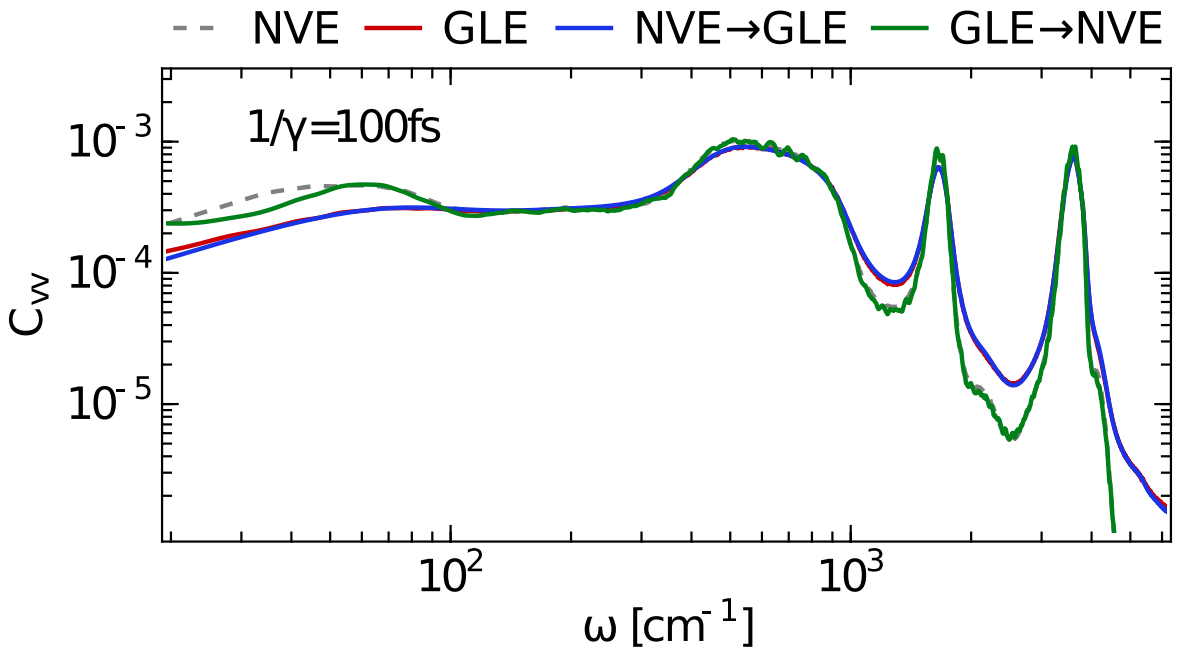
- Can we control (& correct) the disturbance introduced by (Generalized) Langevin dynamics on top of microcanonical MD?
- We can predict the spectrum of a GLE-thermostatted HO of frequency ω_0, and write the perturbed power spectrum as a convolution
  $c_{vv}^{\mathrm{GLE}}(\omega) = \int \mathrm{d}\omega_0 \mathcal{C}(\omega, \omega_0)\, c_{vv}^{\mathrm{NVE}}(\omega_0)$
- It is also possible to deconvolute a (G)LE spectrum: great potential to extract dynamical properties from thermostatted MD



M. Rossi, V. Kapil, Ceriotti, JCP (2018)

Controlling and Correcting Dynamics

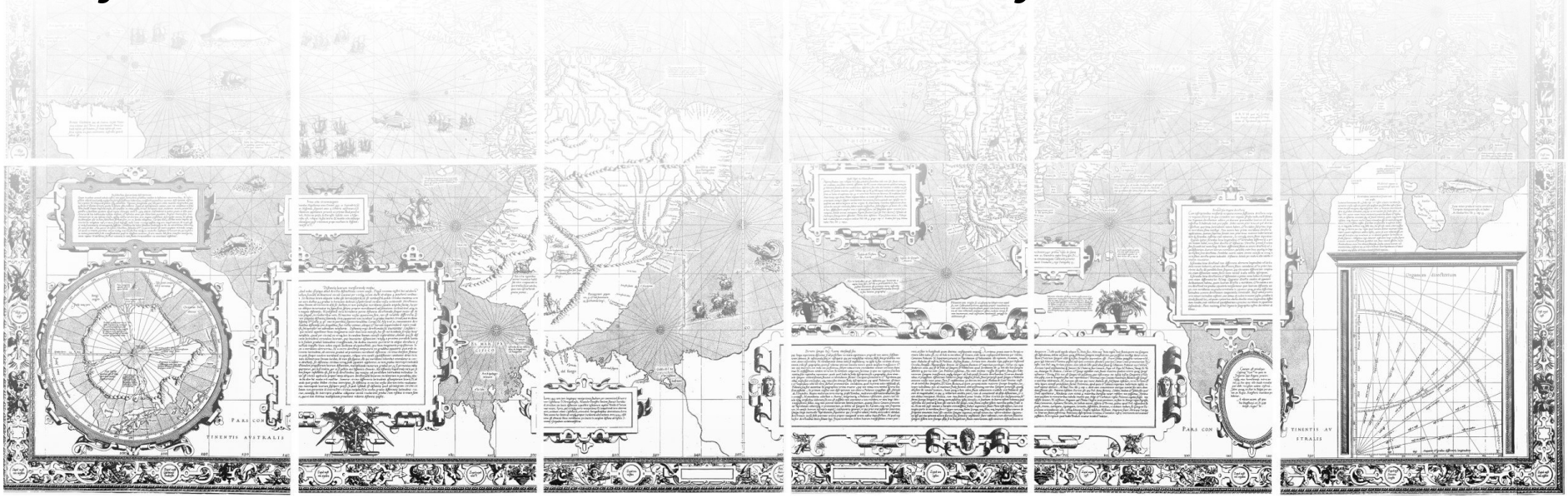
- Can we control (& correct) the disturbance introduced by (Generalized) Langevin dynamics on top of microcanonical MD?
- We can predict the spectrum of a GLE-thermostatted HO of frequency ω_0, and write the perturbed power spectrum as a convolution
  $c_{vv}^{\mathrm{GLE}}(\omega) = \int \mathrm{d}\omega_0 \mathcal{C}(\omega, \omega_0)\, c_{vv}^{\mathrm{NVE}}(\omega_0)$
- It is also possible to deconvolute a (G)LE spectrum: great potential to extract dynamical properties from thermostatted MD



M. Rossi, V. Kapil, Ceriotti, JCP (2018)

Outlook

- Building structural kernels from local-environments SOAP fingerprints
  - "Nearsightedness" of electronic matter: excellent accuracy on benchmark DBs - **~1kJ/mol** for 80%GDB9 and 75%QM7b(multi-scale)
  - Beyond additive models using entropy-regularized kernels
  - Potentials for solids, 99% accuracy in predicting drug activity, silicon & molecular crystals, "useful" predictions: PM7$\rightarrow$CC
- Ingredients for effective learning: sound mathematical foundation, train set sparsification, cross-species learning & multi-scale kernels.
- Colored noise Langevin equations as a mean to manipulate sampling and dynamics in a controllable and correctable way



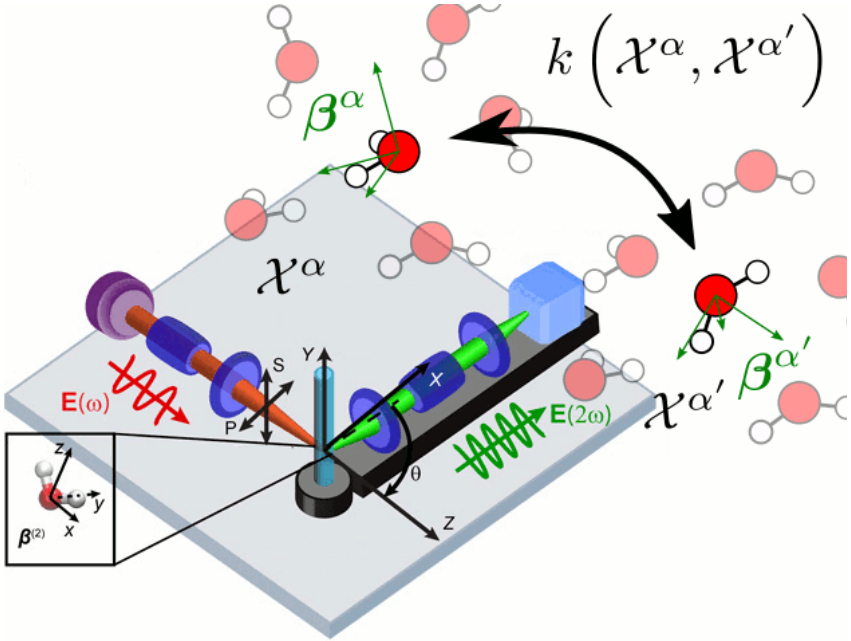
(Development) code available on http://cosmo-epfl.github.io & http://sketchmap.org/

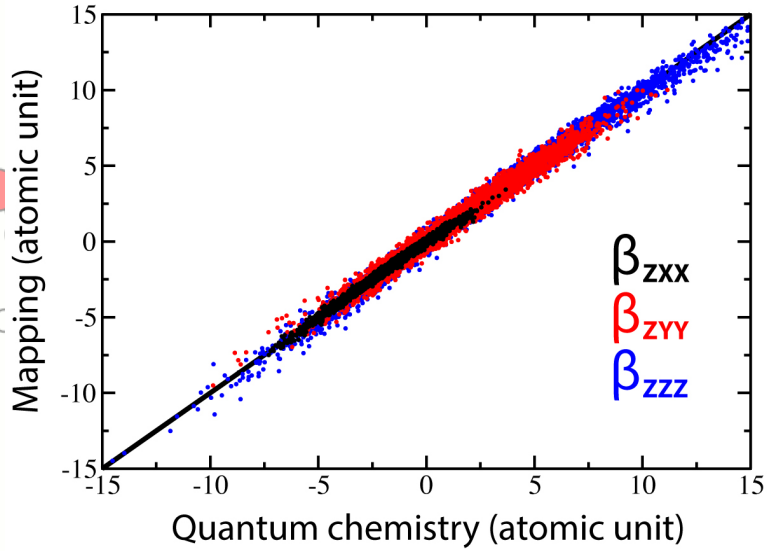
Symmetry-Adapted Gaussian Process Regression

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

$$k_{\mu\nu}\left(\mathcal{X},\mathcal{X}'\right)=\left\langle T_{\mu}\left(\mathcal{X}\right);T_{\nu}\left(\mathcal{X}'\right)\right\rangle \rightarrow k_{\mu\nu}\left(S\mathcal{X},S'\mathcal{X}'\right)=S_{\mu\mu'}k_{\mu'\nu'}\left(\mathcal{X},\mathcal{X}'\right)S_{\nu\nu'}$$

- For rigid molecules, one can convert in the local frame and learn individual components
- A more general strategy for fluxional, flexible environments: λ-SOAPs





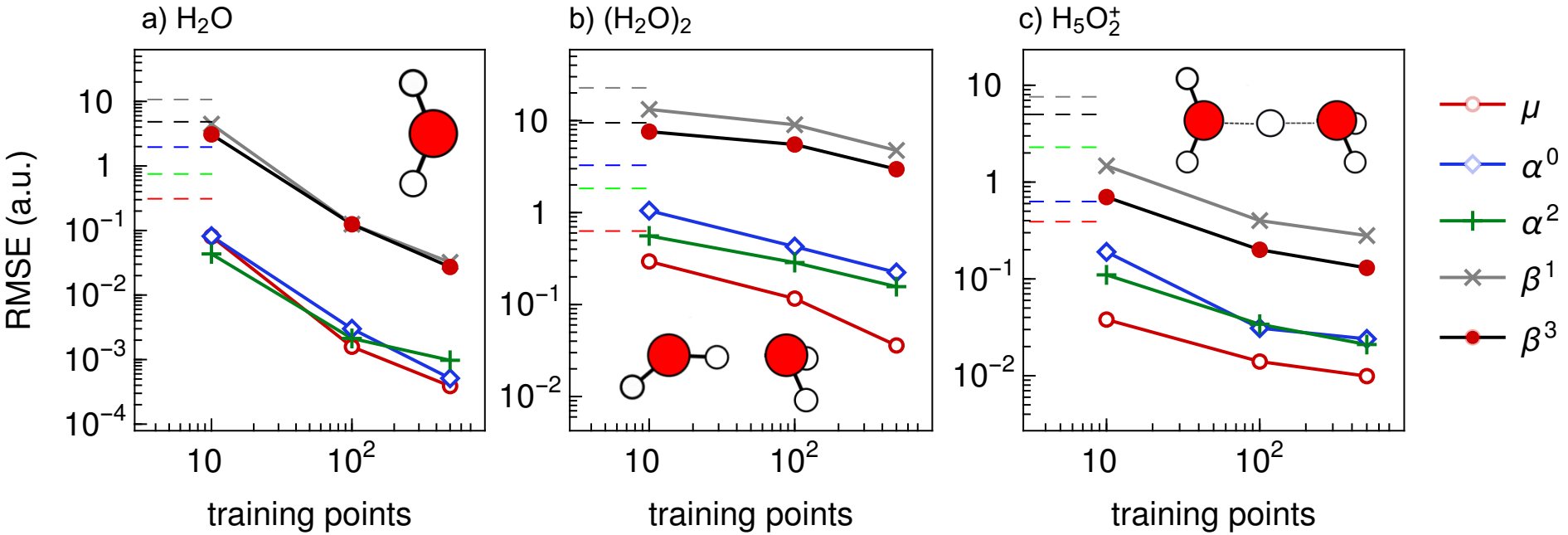
Liang, Tocci, Wilkins, Grisafi, Roke, & **MC**, PRB (2017);
Glielmo, Sollich, & De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, arXiv:1709.06757.

Symmetry-Adapted Gaussian Process Regression

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

$$k_{\mu\nu}(\mathcal{X},\mathcal{X}') = \langle T_\mu(\mathcal{X}); T_\nu(\mathcal{X}')\rangle \rightarrow k_{\mu\nu}(S\mathcal{X},S'\mathcal{X}') = S_{\mu\mu'}k_{\mu'\nu'}(\mathcal{X},\mathcal{X}')S_{\nu\nu'}$$

- For rigid molecules, one can convert in the local frame and learn individual components
- A more general strategy for fluxional, flexible environments: λ-SOAPs



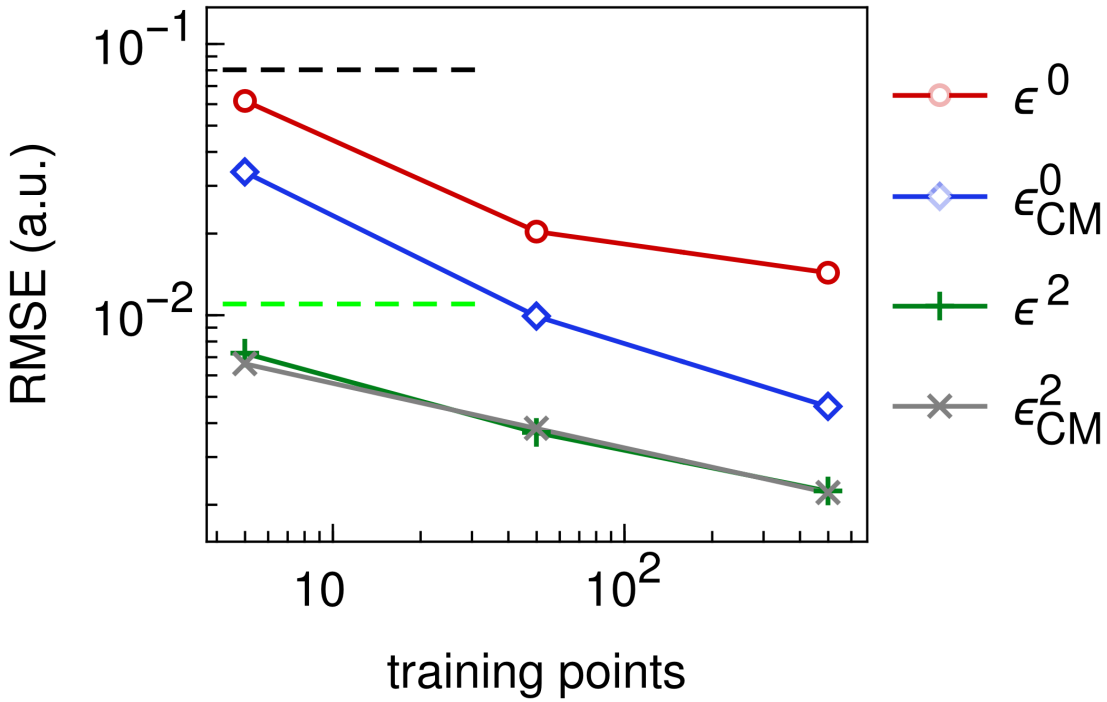
Liang, Tocci, Wilkins, Grisafi, Roke, & **MC**, PRB (2017);
Glielmo, Sollich, & De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, arXiv:1709.06757.

Symmetry-Adapted Gaussian Process Regression

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

$$k_{\mu\nu}(\mathcal{X},\mathcal{X}') = \langle T_\mu(\mathcal{X}); T_\nu(\mathcal{X}')\rangle \rightarrow k_{\mu\nu}(S\mathcal{X},S'\mathcal{X}') = S_{\mu\mu'}k_{\mu'\nu'}(\mathcal{X},\mathcal{X}')S_{\nu\nu'}$$

- For rigid molecules, one can convert in the local frame and learn individual components
- A more general strategy for fluxional, flexible environments: λ-SOAPs



Liang, Tocci, Wilkins, Grisafi, Roke, & **MC**, PRB (2017);
Glielmo, Sollich, & De Vita, PRB (2017); Grisafi, Wilkins, Csányi, & **MC**, arXiv:1709.06757.