



Luxembourg National  
Research Fund



Deutsche  
Forschungsgemeinschaft



# Learning from Machine Learning

Alexandre Tkatchenko

*Physics and Materials Science (PhyMS), University of Luxembourg*

[alexandre.tkatchenko@uni.lu](mailto:alexandre.tkatchenko@uni.lu)



Cargese, August 22, 2018

BERLIN BIG  
DATA CENTER

UNIVERSITÉ DU  
LUXEMBOURG

# Machine Learning in a nutshell

Typical scenario: learning from data

- given data set **X** and labels **Y** (generated by some joint probability distribution  $p(x,y)$ )

- **LEARN/INFERENCE** underlying **unknown** mapping

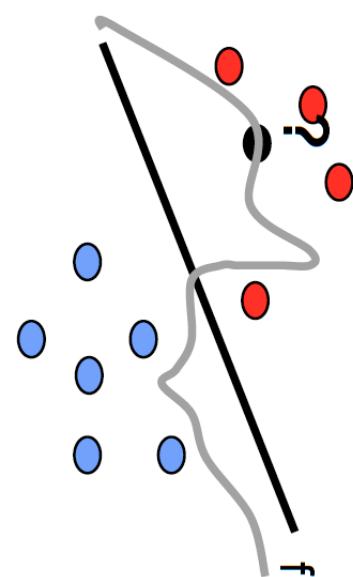
$$Y = f(X)$$

**fit**

Example: ~~understand~~ chemical compound space, distinguish brain states ...

BUT: how to do this optimally with good performance on **unseen** data?

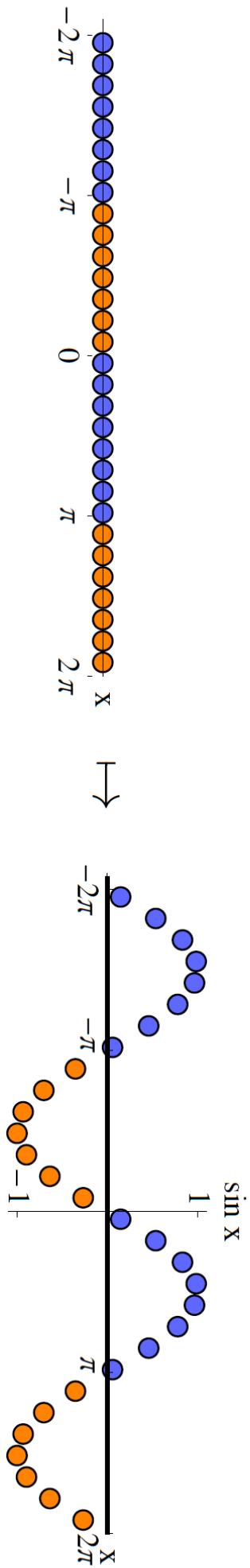
Most popular techniques: **kernel methods** and (deep) **neural networks**



# Kernel Learning

Idea:

- Transform samples into higher-dimensional space
- *Implicitly* compute inner products there
- Rewrite linear algorithm to use only inner products



Input space  $\mathcal{X}$        $\xrightarrow{\phi}$       Feature space  $\mathcal{H}$

$$k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}, \quad k(x, z) = \langle \phi(x), \phi(z) \rangle$$

# Kernel Ridge Regression

- Regularized form of ordinary regression
- Regularization prevents over-fitting by penalizing large coefficients
- Use of kernels for non-linearity

Solution has form

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x})$$

Coefficients  $\alpha$  are obtained by solving

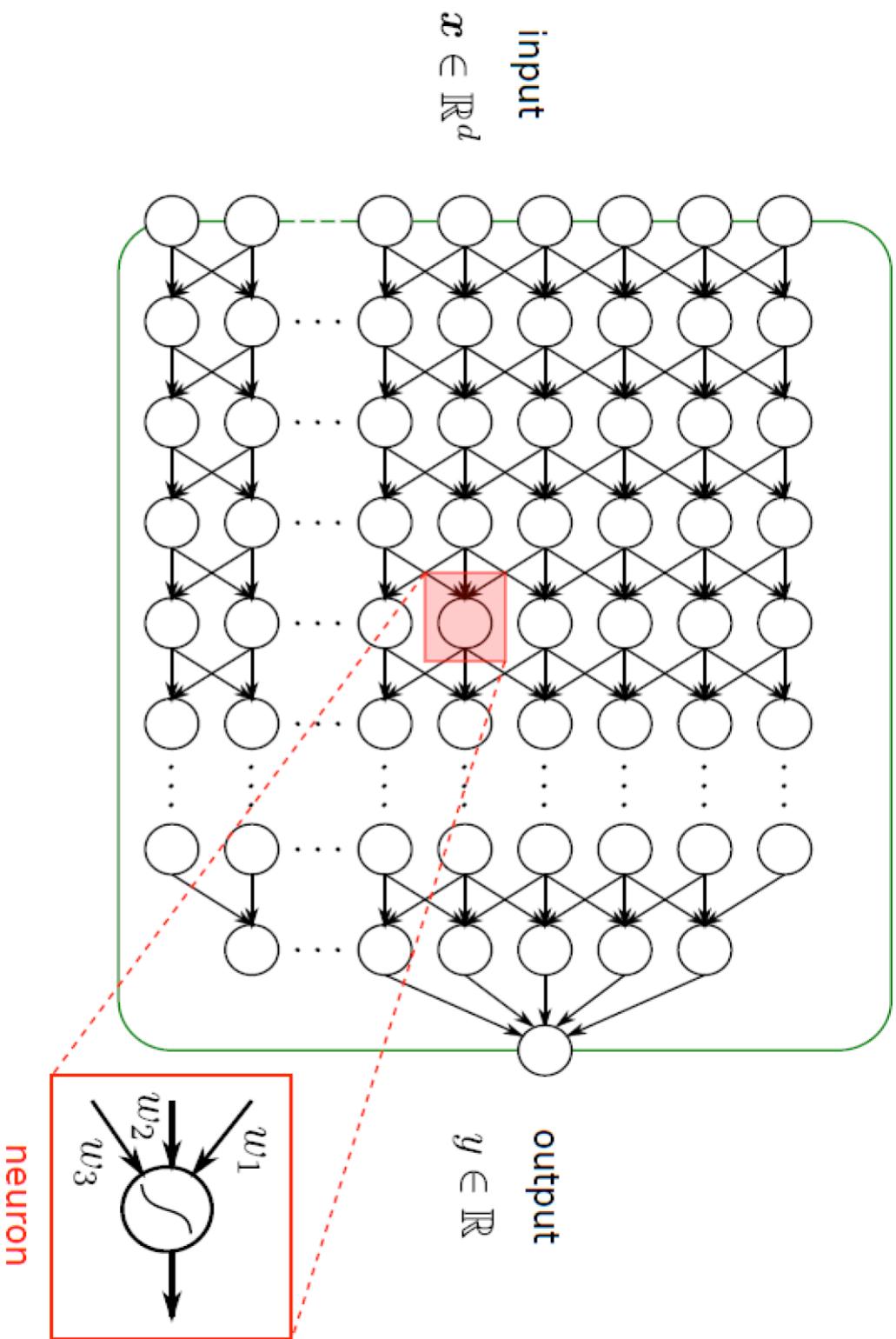
$$\sum_{i=1}^n (f(\mathbf{x}_i) - y_i)^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha},$$

which has solution

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}.$$

# Neural Networks

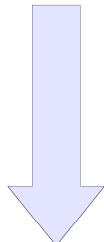
deep neural network



- ▼ Neuron applies a nonlinear function to its input.
- ▼ Examples of functions: hyperbolic tangent, rectification.

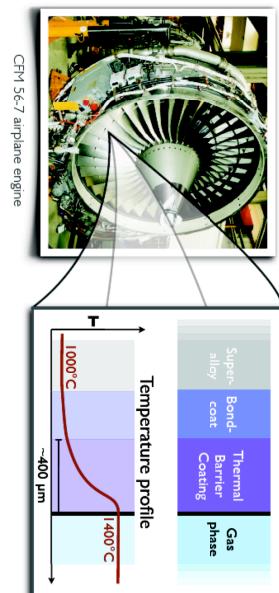
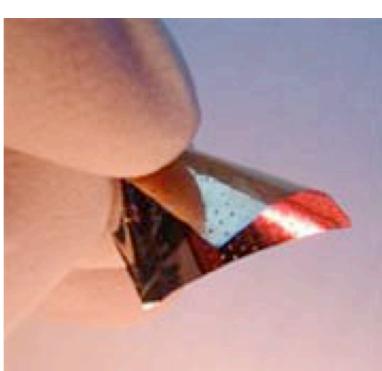
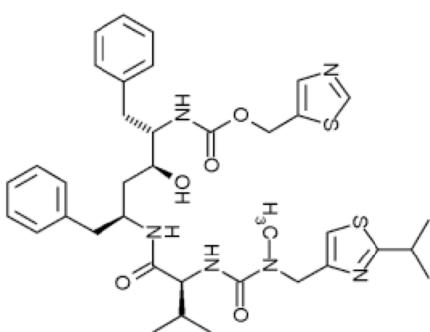
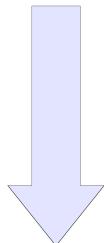
# Physics and Chemistry (and Biology?)

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



Density-functional theory, perturbation theory,  
coupled cluster, configuration interaction, ...

$$E_v[n] = T_s[n] + \int v(\mathbf{r})n(\mathbf{r})d^3\mathbf{r} + E^{\text{Hartree}}[n] + E^{\text{xc}}[n]$$

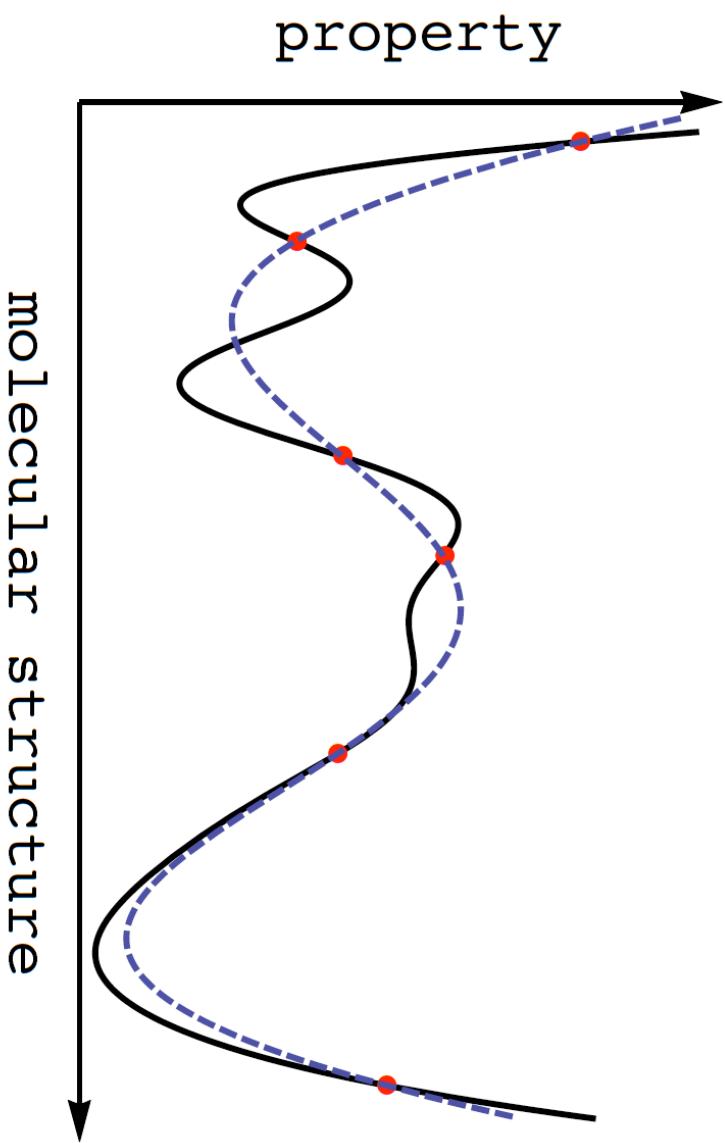


# Quantum Mechanics / ML models

Exploit redundancy in a series of QM calculations

- QM/ML = quantum mechanics + machine learning
- Interpolate between QM calculations using ML
- Smoothness assumption (regularization)

- reference calculations
  - QM
  - ML



# Big Data for Molecules and Materials

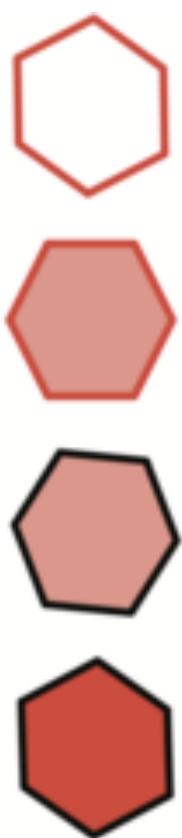


e-cam2020.eu

[nomad-coe.eu](http://nomad-coe.eu)

MARVEL

[max-centre.eu](http://max-centre.eu)

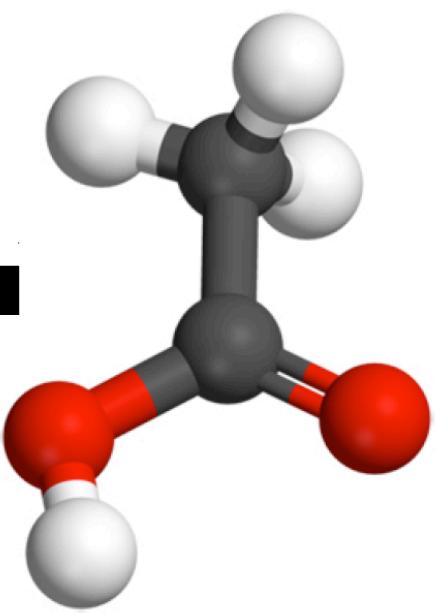


NATIONAL CENTRE OF COMPETENCE IN RESEARCH

[ncer-marvel.ch](http://ncer-marvel.ch)

~~H2020-EU~~  
H2020-EU  
platform technology ecosystem education  
strategic computing  
excellence framework  
development  
researchers  
strategy  
fast pilot  
pilot projects  
new  
center  
workable  
dissemination  
modelling  
Molinari  
ETP4HPC  
1000x  
drop  
PSIK  
navigation  
research  
Modena  
European  
Excellence  
Material  
Material  
material  
challenge  
funded  
Update  
e-infrastructure  
H2020-EU  
Mar  
MARVEL

# Quantum physics/chemistry today



$$\begin{array}{l} \text{DFT} \\ \text{MP2} \\ \text{CCSD(T)} \\ \dots \end{array} \hat{\mathcal{H}}(R_1, Z_1, \dots, R_N, Z_N) \tilde{\Psi} = E \tilde{\Psi}$$



Properties: Energy, polarizability, HOMO, LUMO, ...  
Dynamics: Thermal properties, spectroscopy, ...

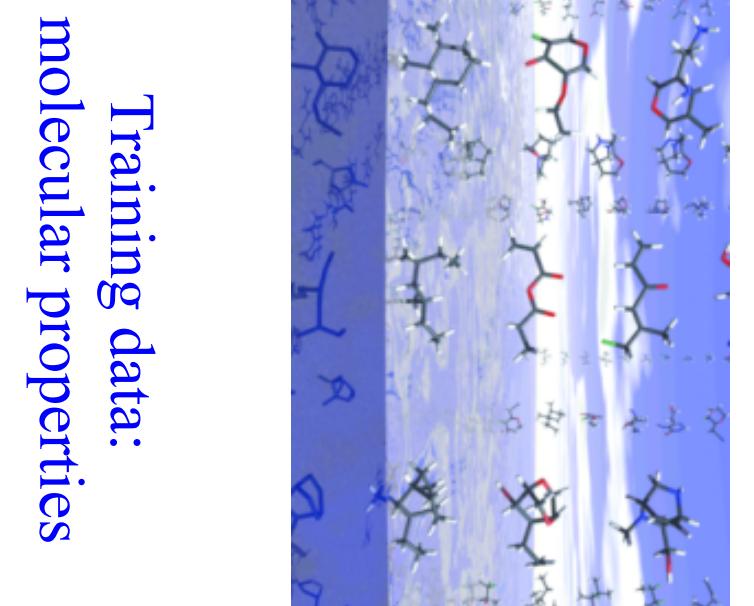
# Quantum physics/chemistry tomorrow?

ML Insights:

- Structure of chemical space

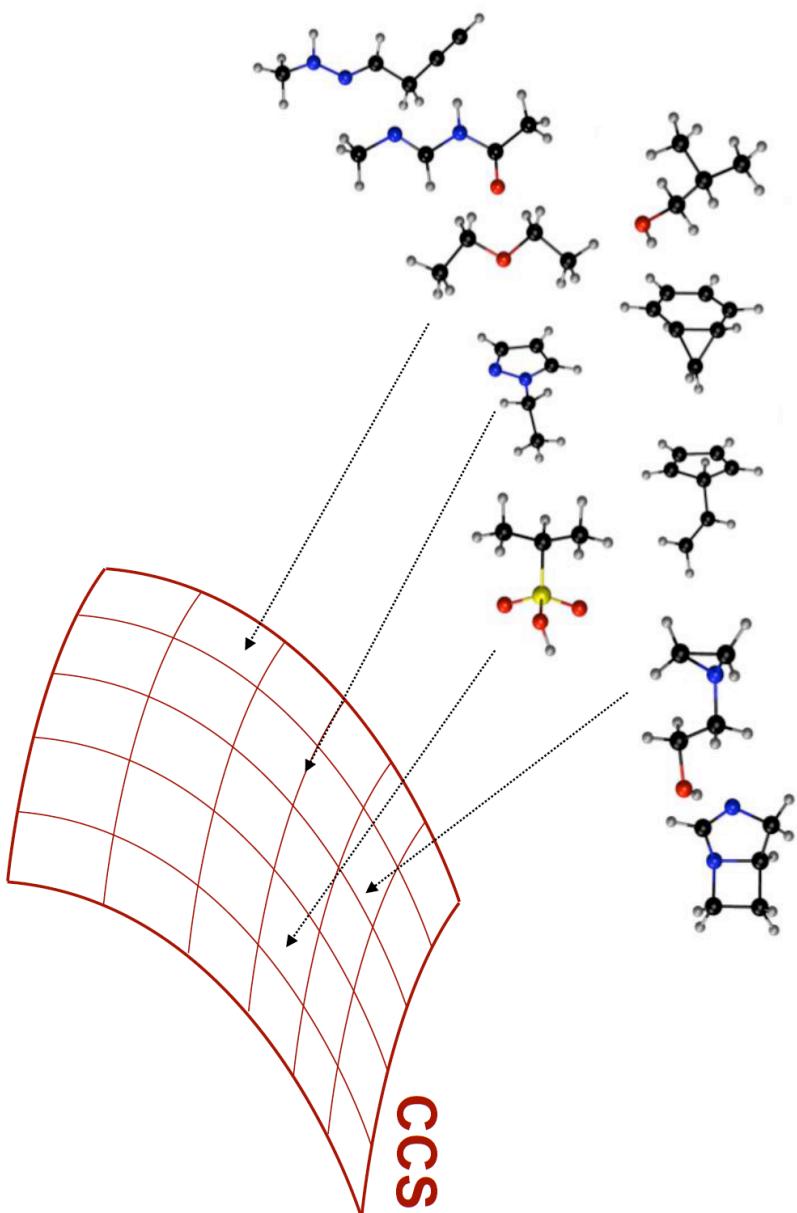
- Reactivity trends,

- aromaticity,  
“new” chemistry



- Training data:  
molecular properties
- Molecular design through multi-property optimization
  - ...

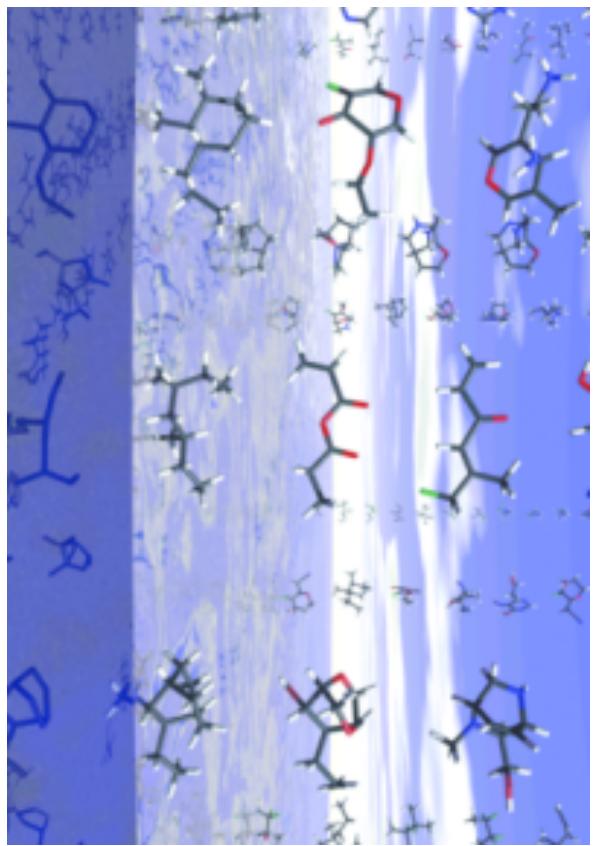
# Molecular big data



$\{R_i, Z_i\}$  maps to  $\{P_1, P_2, P_3, P_4, \dots\}$

- Graph theory:  
combinatorial explosion
- At least  $10^{60}$  small drug  
candidate molecules
- Finding needles in a  
haystack

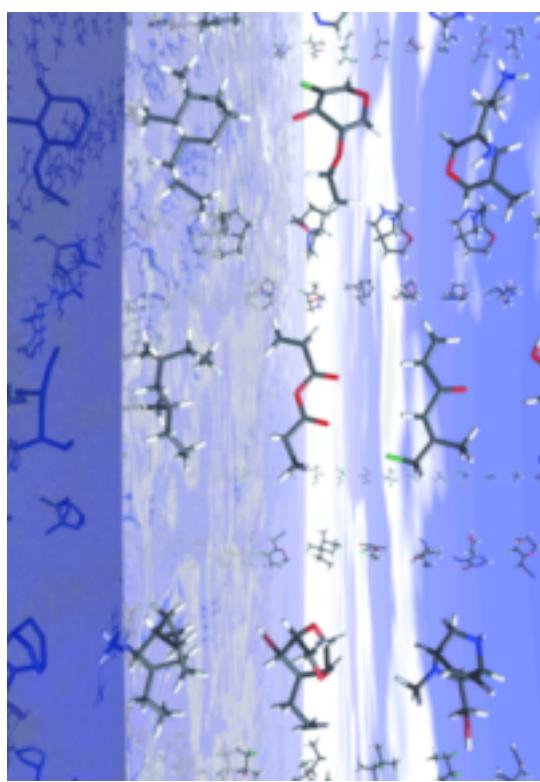
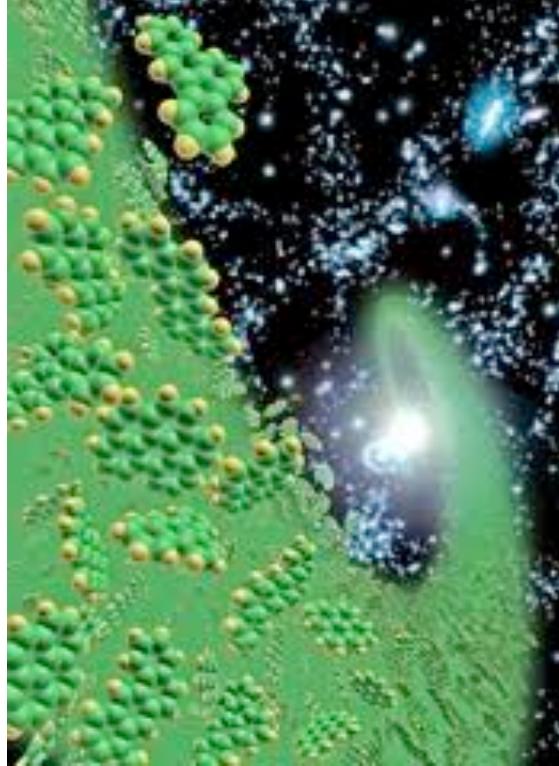
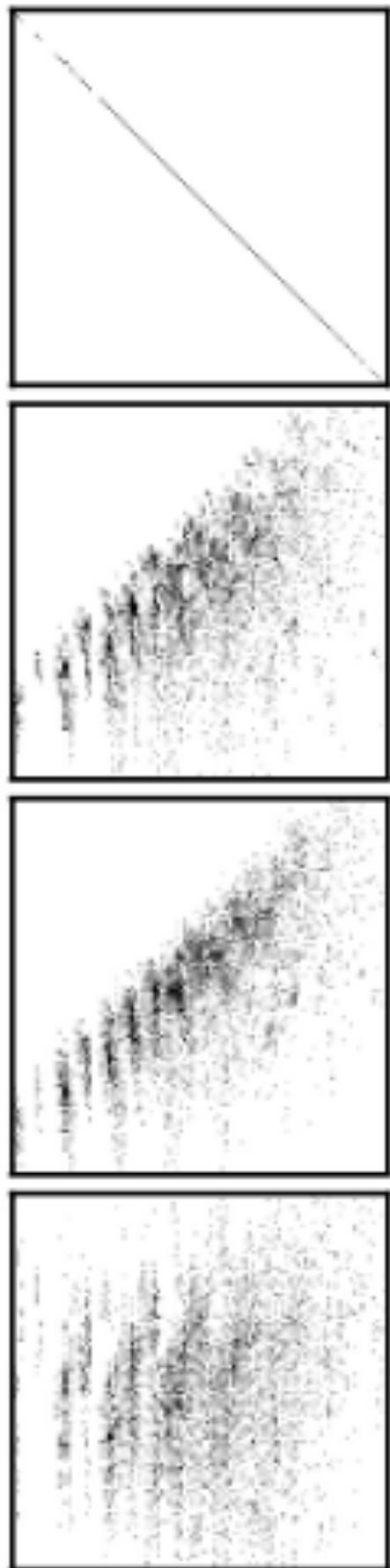
# Machine learning for molecular big data



- **Descriptor:** what's a good representation of a molecule?
  - **Metric:** how to define distance between two molecules?
  - **Data selection:** Which molecules to use for training?
  - **Properties:** which set of properties uniquely defines a molecule?
- $\{R_i, Z_i\}$  maps to  $\{P_1, P_2, P_3, P_4, \dots\}$

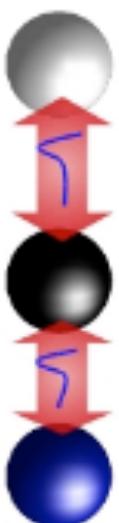
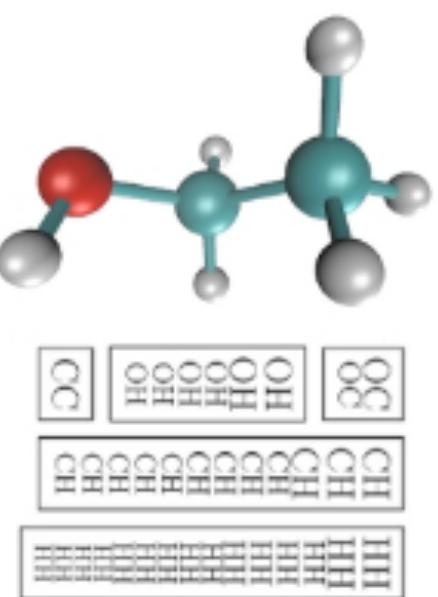
# Chemical Compound Space: Freedom of design

$E(P)$        $\alpha(P)$        $\alpha(S)$       HOMO(G)



G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Mueller, A. von Lilienfeld, *New J. Phys.* 15, 095003 (2013).

# Predicting Molecular Properties: Descriptors From “Dressed Atoms” to Bag-of-Bonds

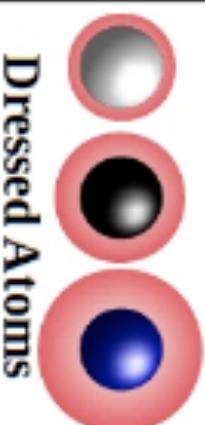


2-Body Potentials

Bag-of-Bonds



Sum over Bonds



Dressed Atoms

ACCURACY →

COMPLEXITY ↓

K. Hansen, F. Biegler,  
R. Ramakrishnan, W. Pronobis,  
O. A. von Lilienfeld,  
K.-R. Mueller, and A. Tkatchenko,  
*J. Phys. Chem. Lett.* 6, 2326 (2015).

# Predicting Molecular Properties: GDB-7 dataset

model	MAE [kcal/mol]
dressed atoms	15.1
sum-overbonds	9.9
Lennard-Jones potential	8.7
polynomial pot. ( $n = 6$ )	5.6
polynomial pot. ( $n = 10$ )	3.9
polynomial pot. ( $n = 18$ )	3.0
Bag of Bonds ( $p = 2$ , Gaussian)	4.5
Bag of Bonds ( $p = 1$ , Laplacian)	1.5
Coulomb matrix ( $p = 2$ , Gaussian) <sup>17</sup>	10.0
Coulomb matrix ( $p = 1$ , Laplacian) <sup>16</sup>	4.3

K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Mueller, and A. Tkatchenko, *J. Phys. Chem. Lett.* 6, 2326 (2015).

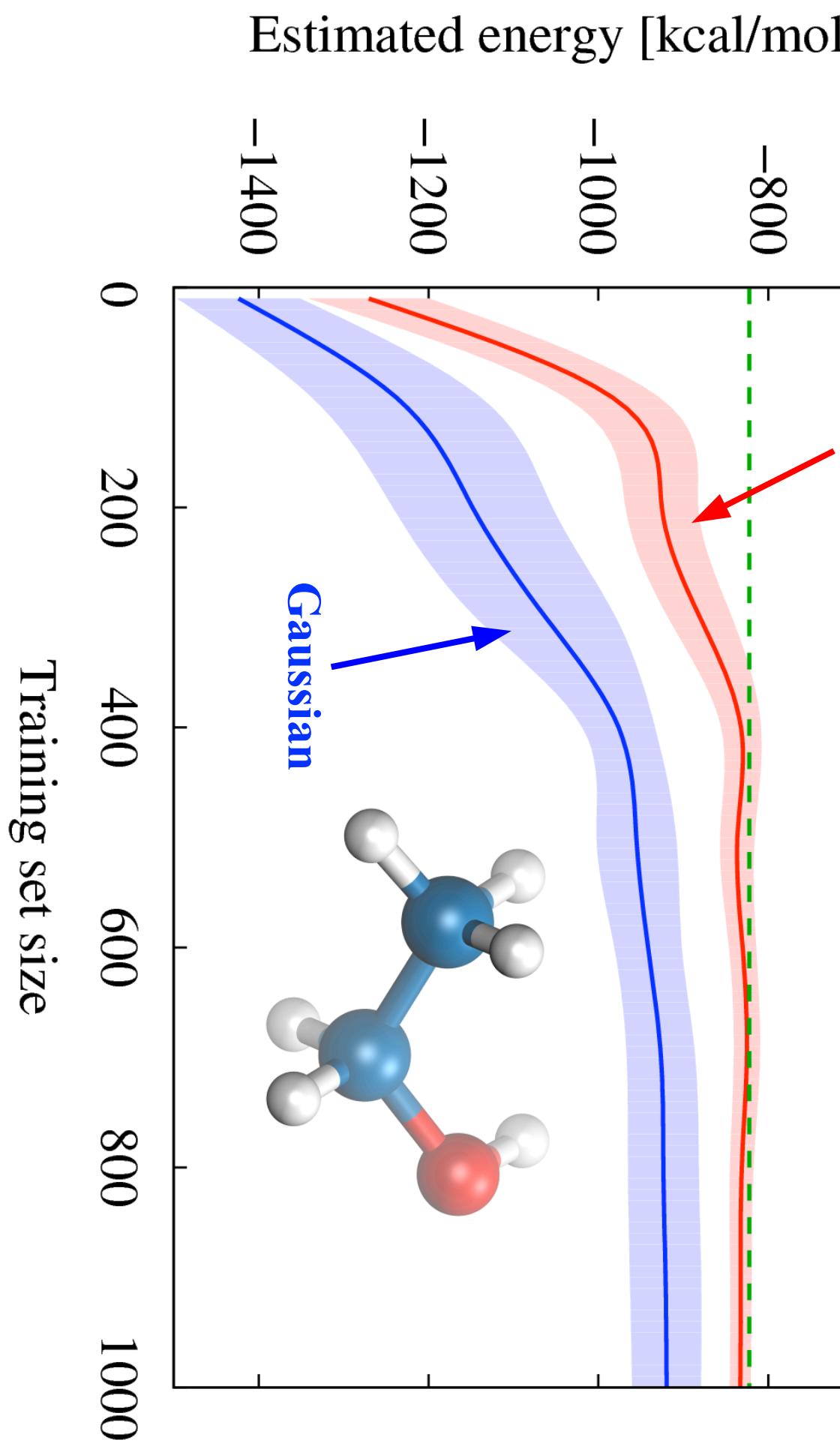
# Predicting Molecular Properties: GDB-7 dataset

model	MAE [kcal/mol]
dressed atoms	15.1
sum-overbonds	9.9
Lennard-Jones potential	8.7
polynomial pot. ( $n = 6$ )	5.6
polynomial pot. ( $n = 10$ )	3.9
polynomial pot. ( $n = 18$ )	3.0
Bag of Bonds ( $p = 2$ , Gaussian)	4.5
Bag of Bonds ( $p = 1$ , Laplacian)	1.5
Coulomb matrix ( $p = 2$ , Gaussian) <sup>17</sup>	10.0
Coulomb matrix ( $p = 1$ , Laplacian) <sup>16</sup>	4.3

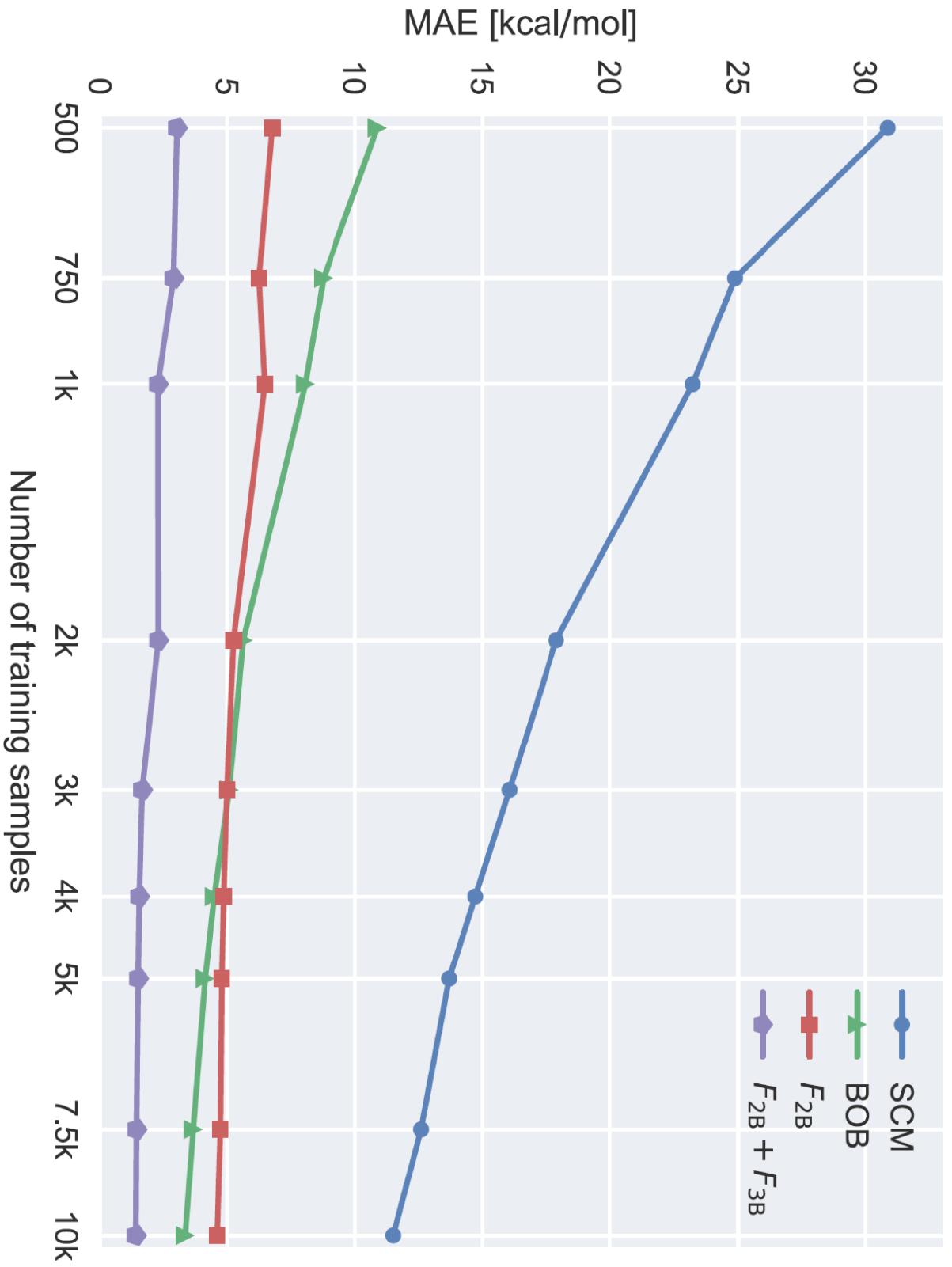
**2+3body many-body expansion**

**0.8**

# *Bag-of-Bonds* (BoB): Non-Locality in Chemical Space

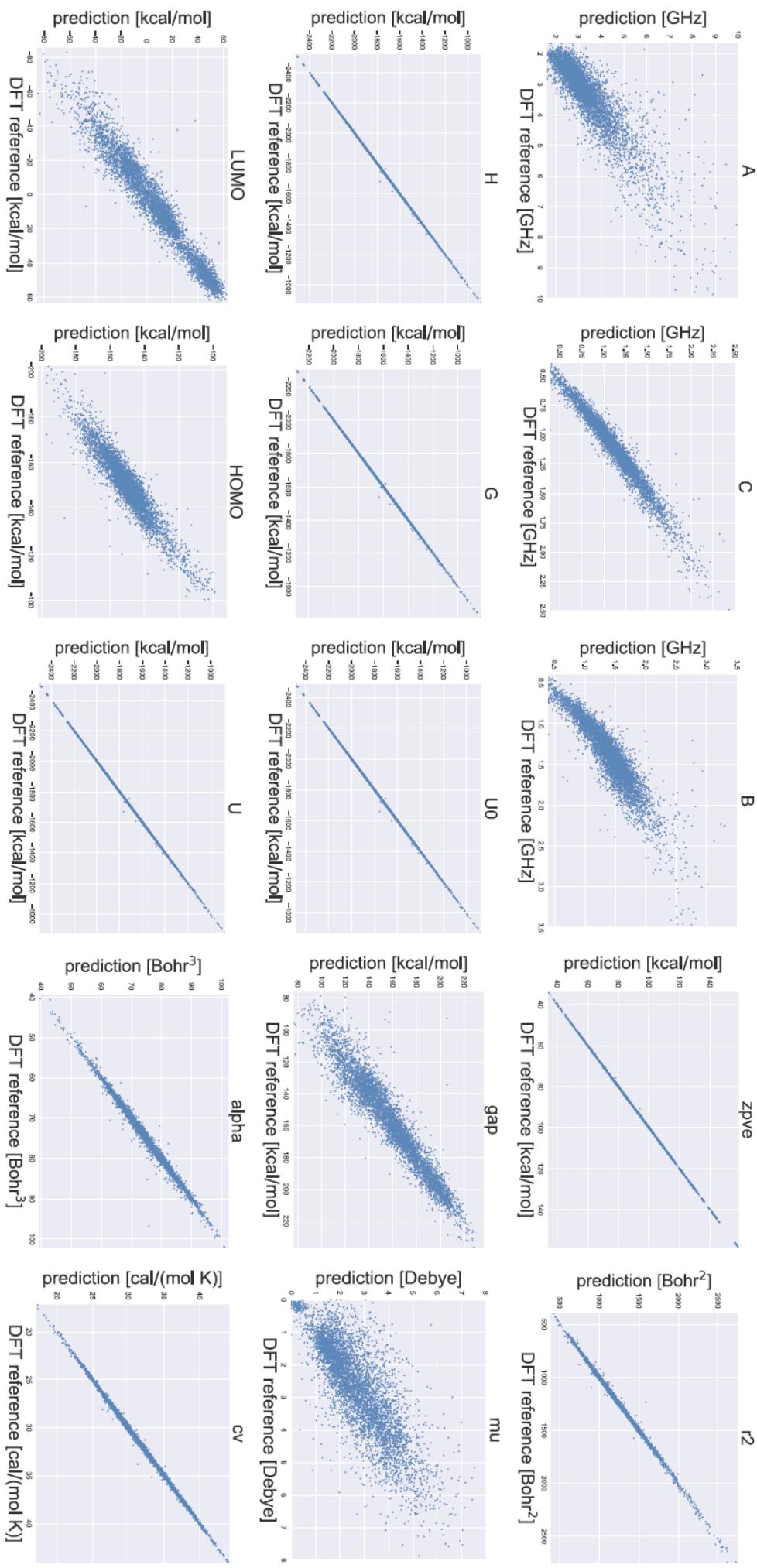


# QM9 dataset: Evolution from Coulomb Matrix to Many-Body Representation



# QM9 dataset: Extensive and Intensive Properties

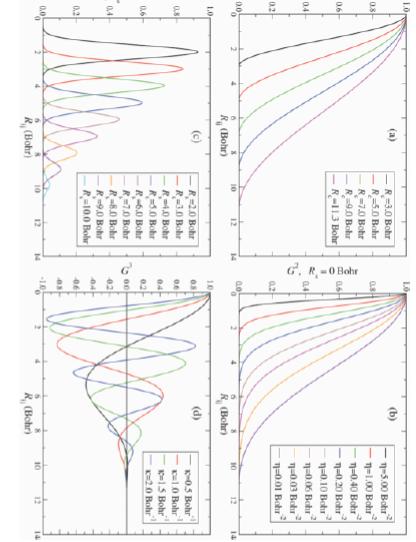
W. Pronobis, A. Tkatchenko, and K.-R. Mueller, *J. Chem. Theory Comput.* (2018).



# Zoo of Descriptors for Molecules and Solids

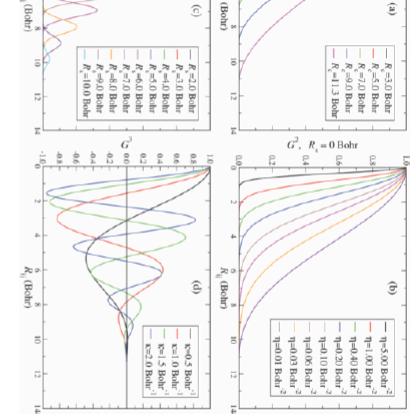


$$M_{ij} = \begin{cases} 0.5 Z_i^{2.4} & \text{for } i = j \\ \frac{Z_i Z_j}{d_{ij}} & \text{for } i \neq j \end{cases}$$



Coulomb matrix  
(Rupp et al. 2012)

Bag of bonds  
(Hansen et al. 2015)



O	C	C	H	H	H	H	H	H	H	O-pag
O	o	oc	oc	OH	OH	OH	OH	OH	OH	0
C	oc	c	cc	CH	CH	CH	CH	CH	CH	C-bag
C	oc	c	cc	CH	CH	CH	CH	CH	CH	0
H	OH	CH	CH	H	HH	HH	HH	HH	HH	H-bag
H	OH	CH	CH	H	HH	HH	HH	HH	HH	0
H	OH	CH	CH	H	HH	HH	HH	HH	HH	OC-bag
H	OH	CH	CH	H	HH	HH	HH	HH	HH	OH-bag
H	OH	CH	CH	H	HH	HH	HH	HH	HH	CC-bag
H	OH	CH	CH	HH	HH	HH	HH	HH	HH	0
H	OH	CH	CH	HH	HH	HH	HH	HH	HH	CH-bag
H	OH	CH	CH	HH	HH	HH	HH	HH	HH	0
CC	CH	CC-bag								
CC	CH	0								
H	CH	0								
H	CH	HH-bag								
H	CH	0								

Atom-centered  
symmetry functions  
(Behler et al. 2007)

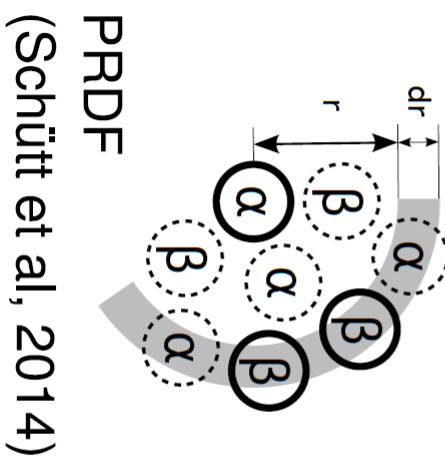
$$\{Z_i, \mathbf{R}_i\}$$

$$\{Z_i, d_{ij}\}$$

$$k(\rho, \rho') = \int d\hat{R} \left| \rho(\mathbf{r}) \rho'(\hat{R}\mathbf{r}) \right|^n$$

$$x_{ij} = \begin{cases} 0.5 Z_i^{2.4} & \text{if } i = j \\ \frac{Z_i Z_j}{\phi(\mathbf{r}_i, \mathbf{r}_j)} & \text{if } i \neq j \end{cases}$$

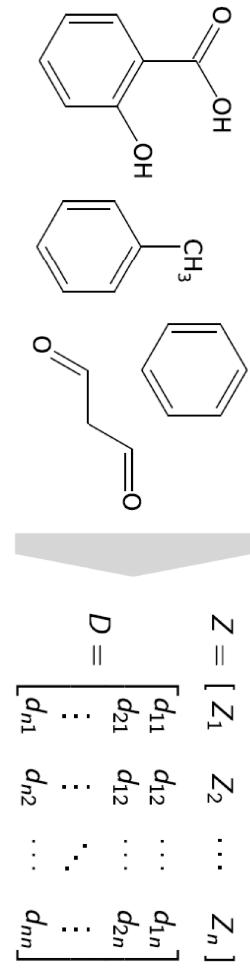
Sine matrix  
(Faber et al. 2015)



# Learning the Representation: Deep Tensor Neural Networks (DTNN)

# Deep Tensor Neural Networks (DTNN)

**Input:** Atomic numbers and interatomic distances



**Embedding of based on atom types**

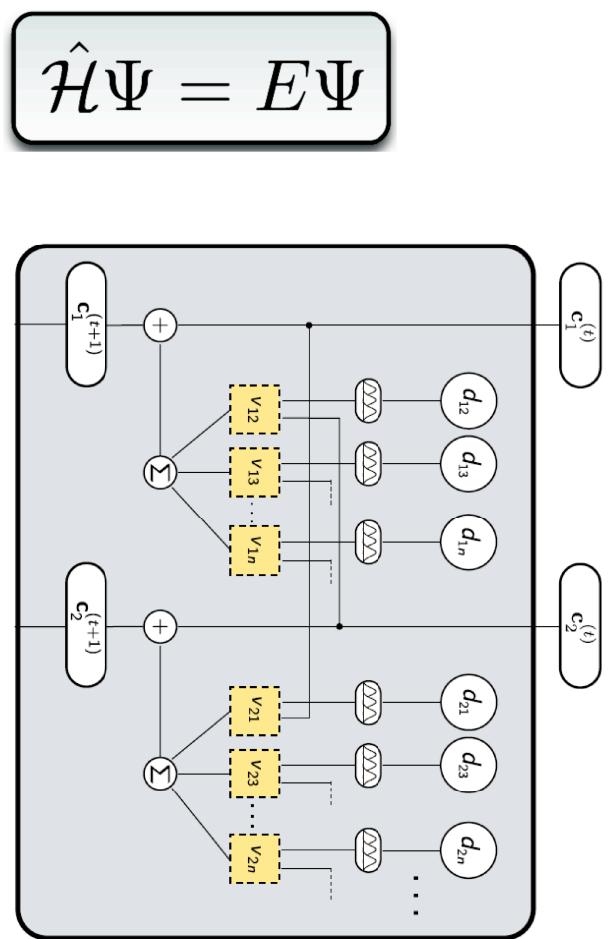
$$\mathbf{x}_i^{(0)} = \mathbf{x}_{Z_i} \in \mathbb{R}^d$$

**Add interaction with environment using  $t = 1 \dots T$**   
**sequential refinements  $\mathbf{v}_i^{(t)}$**

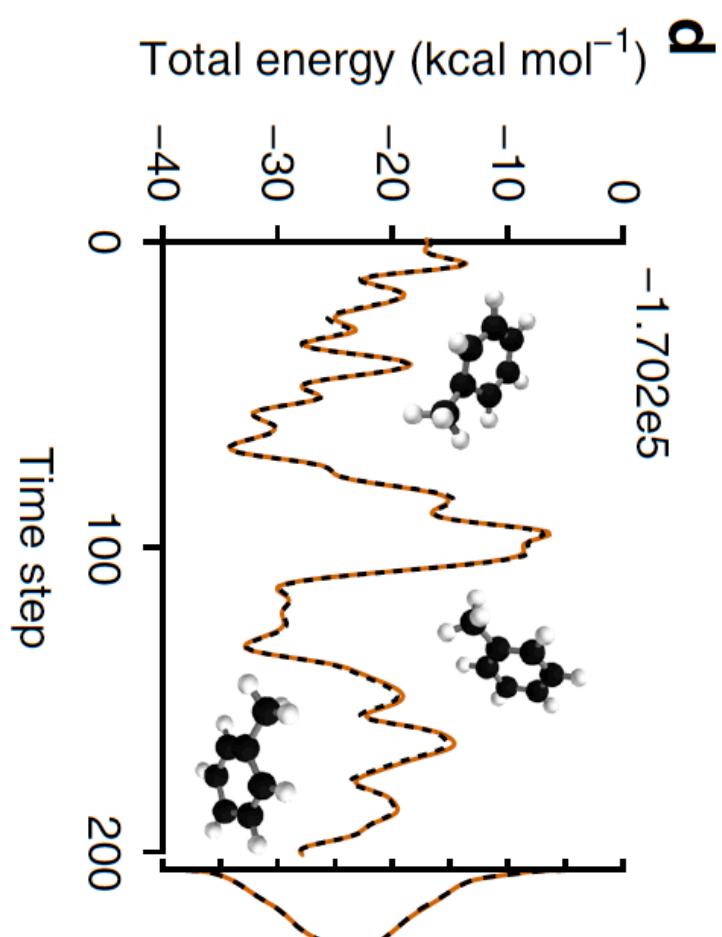
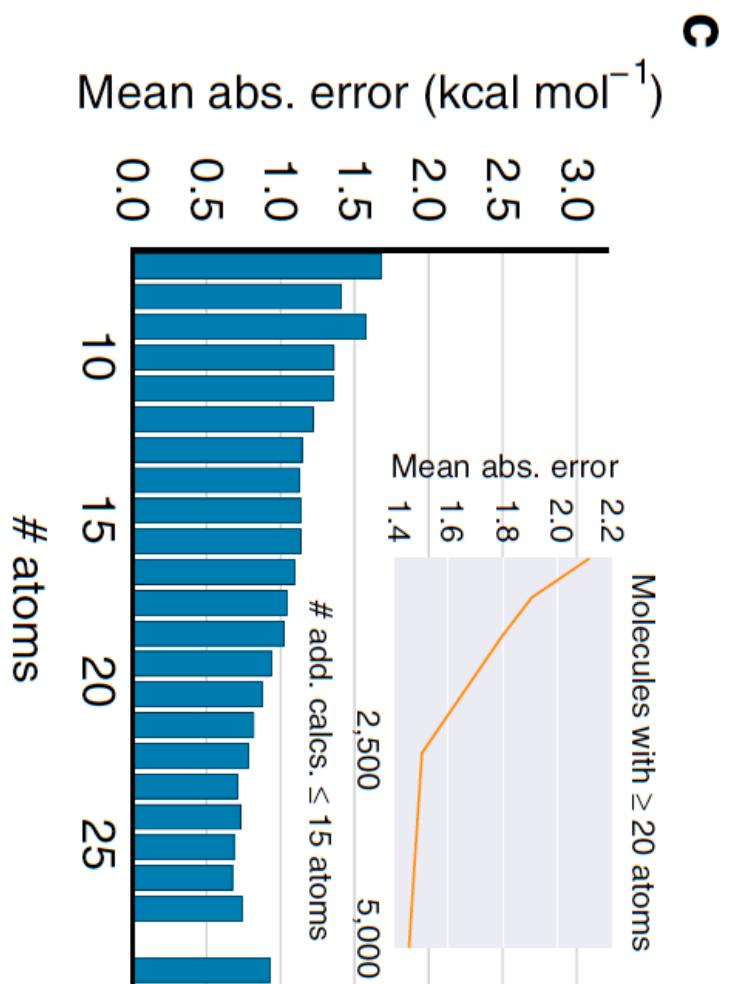
$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \mathbf{v}_i^{(t)} \left( \mathbf{x}_1^{(t)}, \dots, \mathbf{x}_{n_{\text{atoms}}}^{(t)}, d_{i1}, \dots, d_{in_{\text{atoms}}} \right)$$

**Prediction via atom-wise contributions:**

$$\hat{E} = \sum_{i=1}^{n_{\text{atoms}}} f_{\text{out}}(\mathbf{x}_i^{(T)})$$

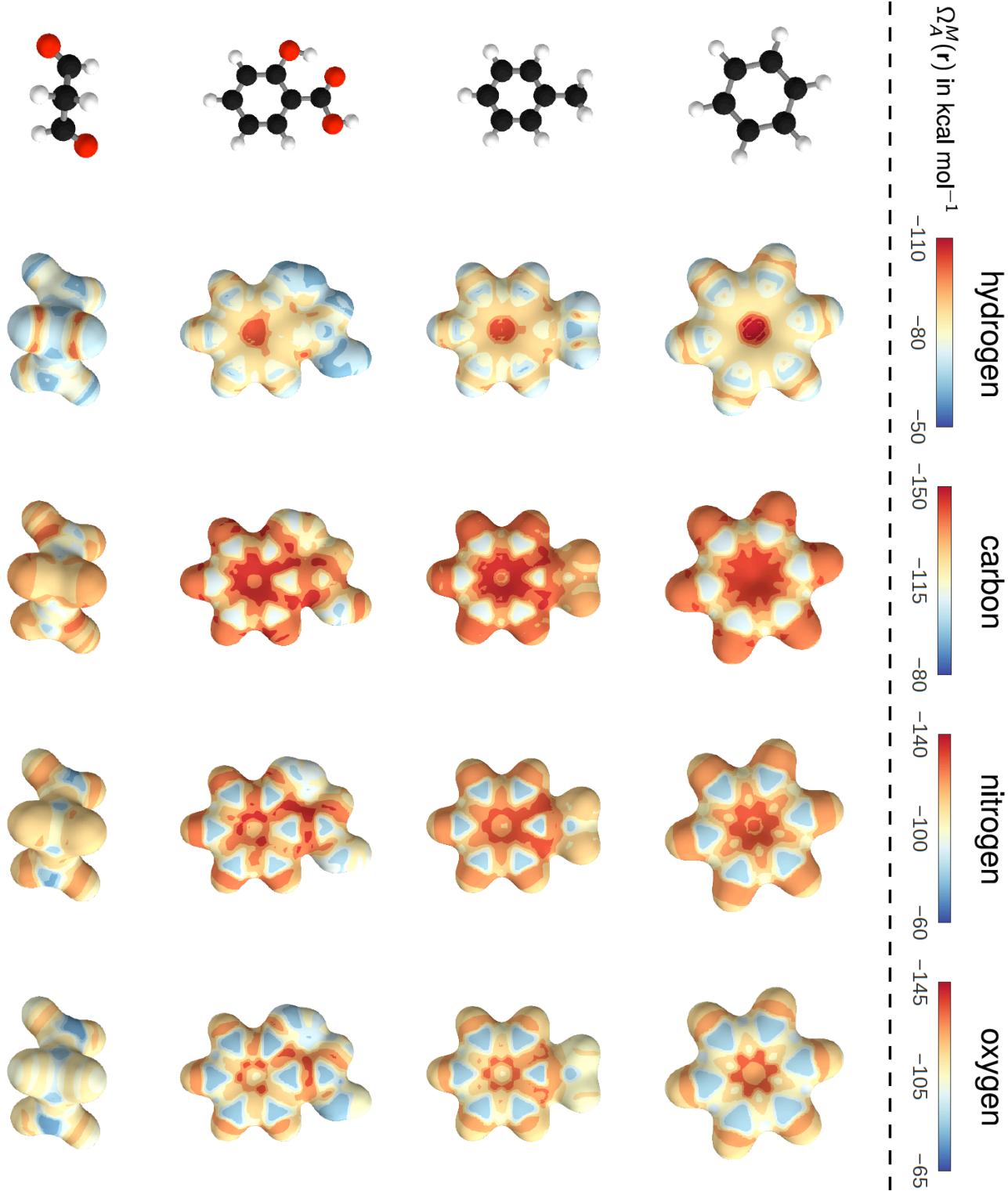


# Molecular DTNN: Performance on GDB-9 and MD



K. T. Schuett, F. Arbabzadah, S. Chmiela, K.-R. Mueller, and A. Tkatchenko,  
*Nature Commun.* 8, 13890 (2017).

# Molecular DTNN: What Did it Learn ?



# Quantum Chemical Insights: Aromaticity

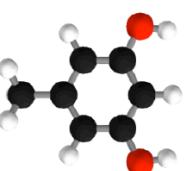
# 1 - 10



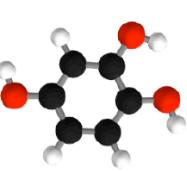
-859.9



-858.3



-857.8



-857.4



-857.4

$E_{\text{ring}}$  in kcal mol $^{-1}$

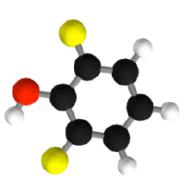
-856.9

-856.8

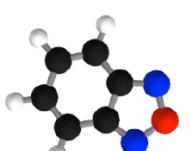
-856.8

-856.6

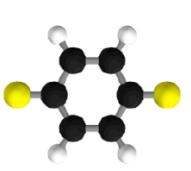
# 281 - 290



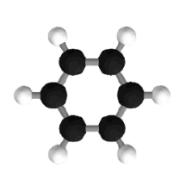
-857.3



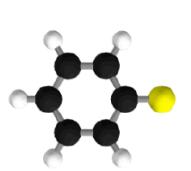
-856.9



-856.8



-856.8



-856.6

$E_{\text{ring}}$  in kcal mol $^{-1}$

-841.7

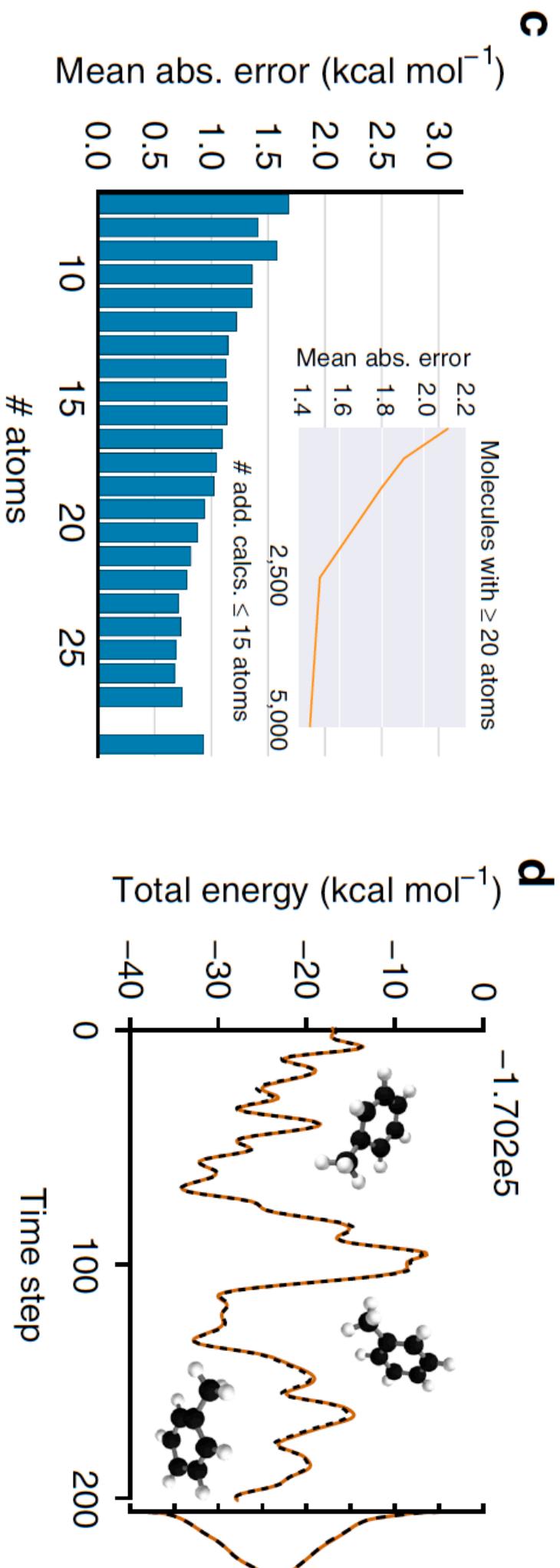
-841.7

-841.4

-841.2

-841.1

# Learning Full Chemical Space with DTNN?

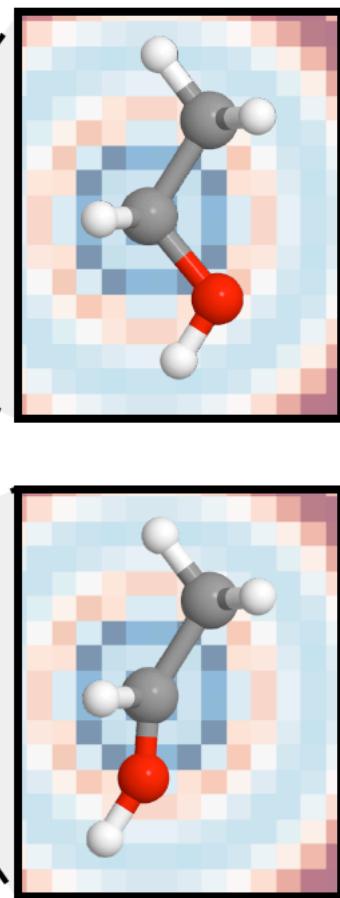


Accurately representing **BOTH** compositional and conformational degrees of freedom is difficult.

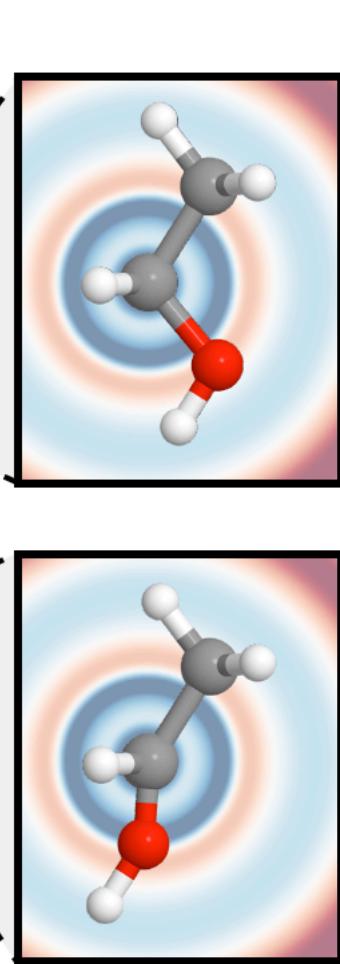
For C<sub>7</sub>O<sub>2</sub>H<sub>10</sub> isomer and MD data, the error grows to **1.7 kcal/mol**.

# From DTNN to SchNet architecture

Discrete filter



Continuous filter

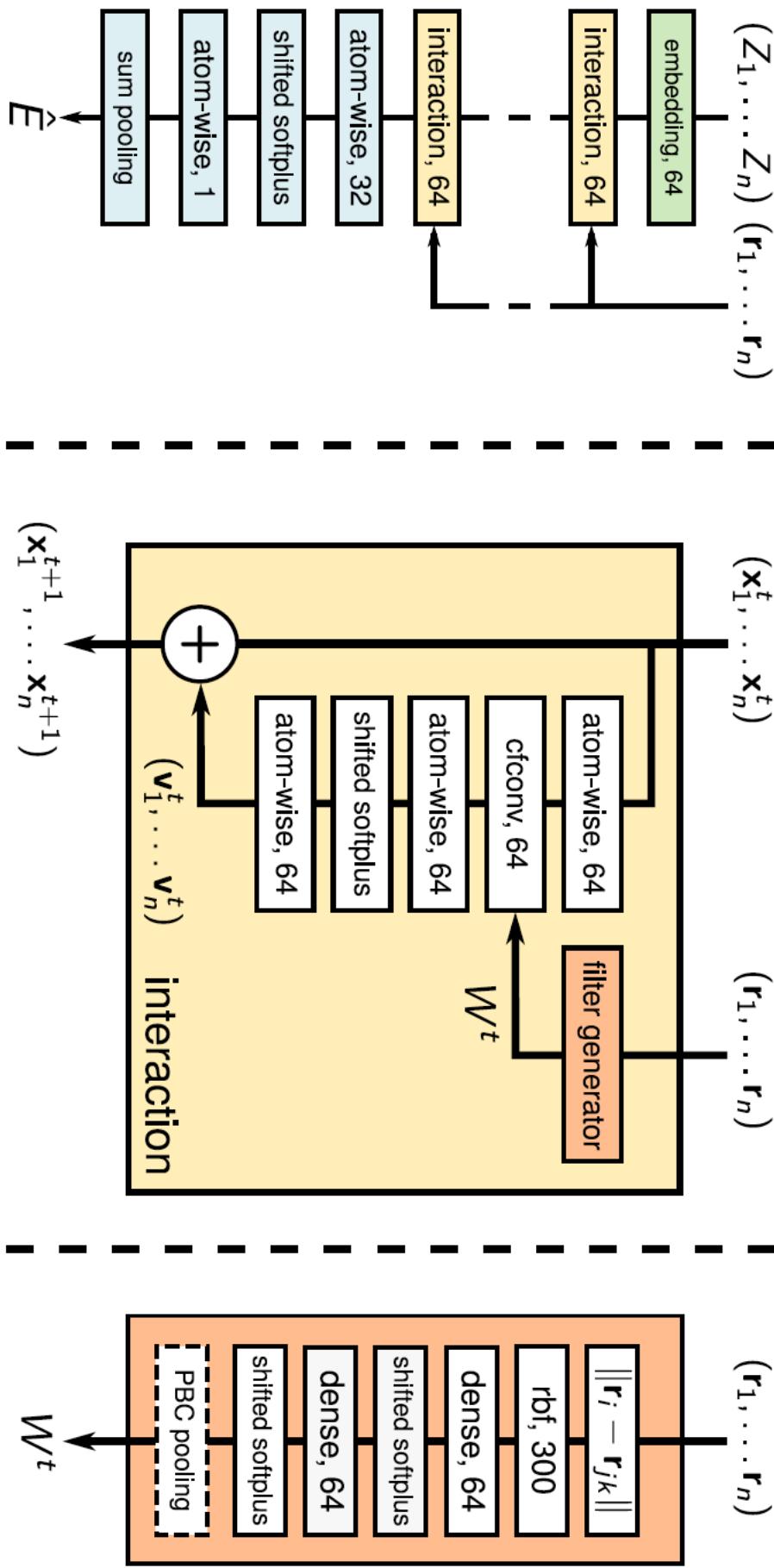


$$\mathbf{v}_i^{(t)} = \sum_{j=1}^{N_{\text{atom}}} \mathbf{x}_j^{(t)} \circ \underbrace{W_{[d_{ij}]}^{(t)}}_{\text{parameter tensor}}$$

$$\mathbf{v}_i^{(t)} = \sum_{j=1}^{N_{\text{atom}}} \mathbf{x}_j^{(t)} \circ \underbrace{W^{(t)}(d_{ij})}_{\text{neural network}}$$

K.T. Schuett, P.J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, K.-R. Mueller (2017). *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*. NIPS.

# From DTNN to SchNet architecture

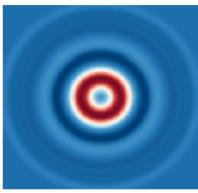


K.T. Schuett, P.J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, K.-R. Mueller (2017). *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*. NIPS.

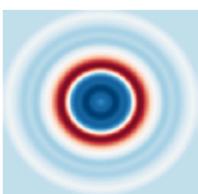
# SchNet architecture: Application to Materials

$$\mathbf{x}_i^{l+1} = \sum_{j=0}^{n_{\text{atoms}}} \mathbf{x}_j^l \circ \underbrace{\left( \sum_{b=0}^{n_{\text{cells}}} \tilde{W}^l(\mathbf{r}_{jb} - \mathbf{r}_{ia}) \right)}_{\text{periodic filter}}$$

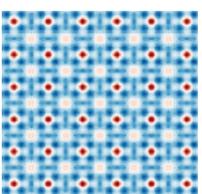
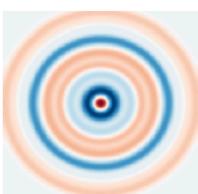
No PBC



Diamond



Graphite

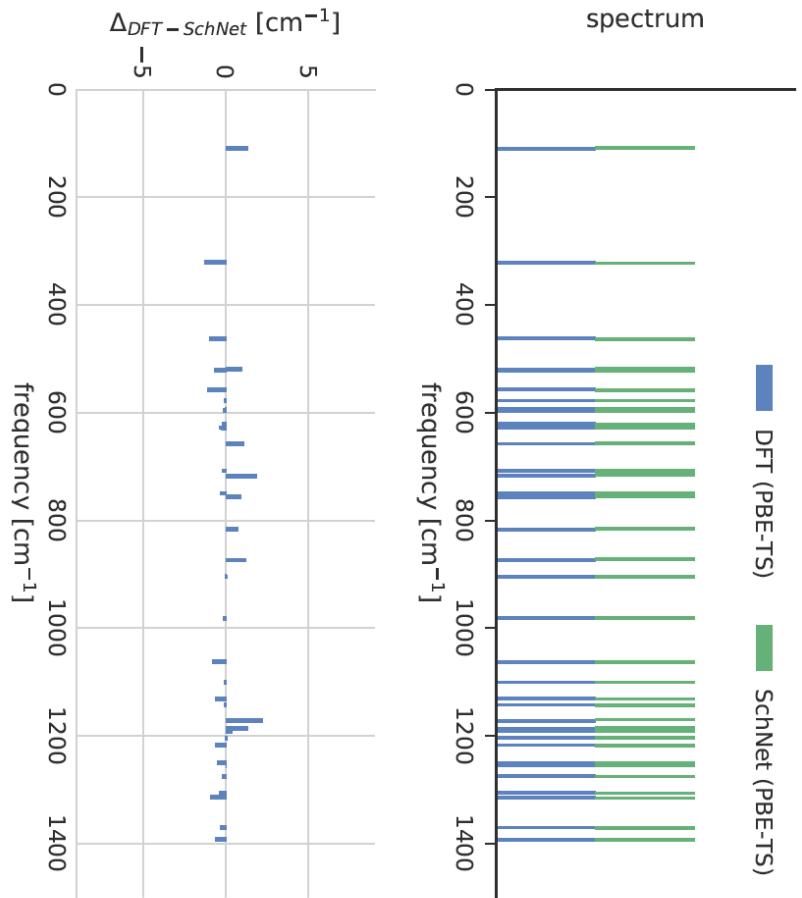


Model	$N = 3,000$	$N = 60,000$
ext. Coulomb matrix <sup>[1]</sup>	0.640	—
Ewald sum matrix <sup>[1]</sup>	0.490	—
sine matrix <sup>[1]</sup>	0.370	—
SchNet ( $T = 6$ )	<b>0.127</b>	<b>0.035</b>

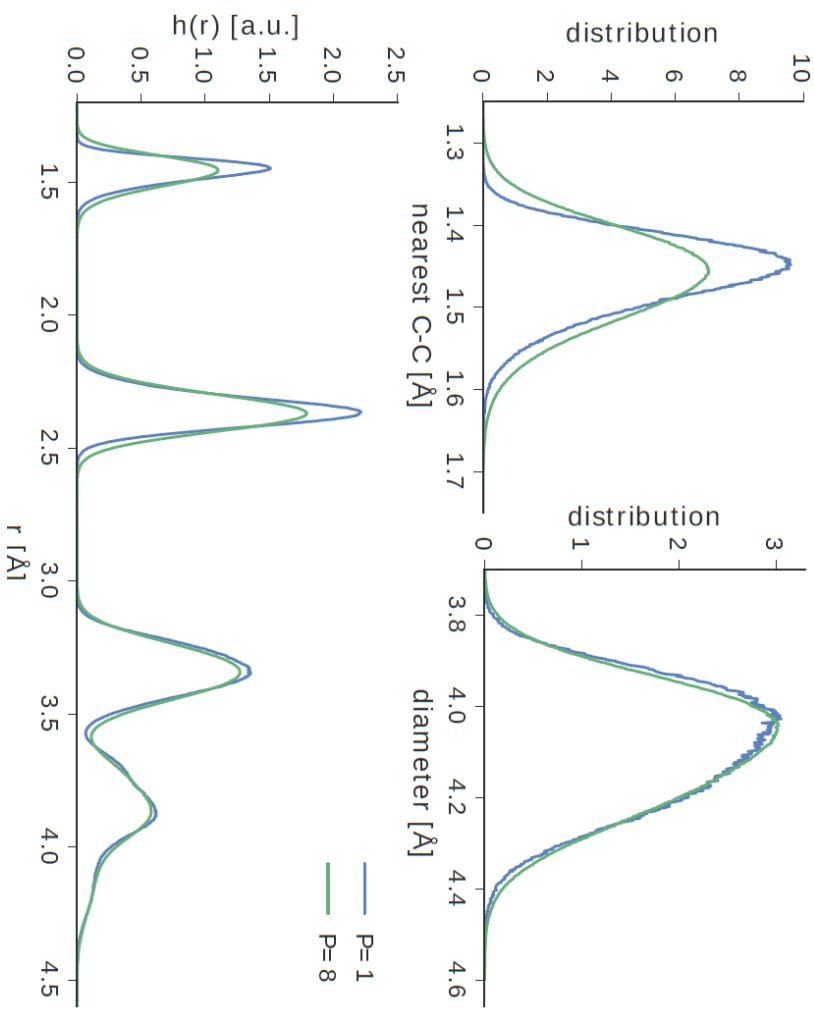
[1] Faber et al. Crystal structure representations for machine learning models of formation energies (2015). IJQC 115(16).

# SchNet architecture: Application to Molecular Dynamics

Accurate prediction of vibrational frequencies

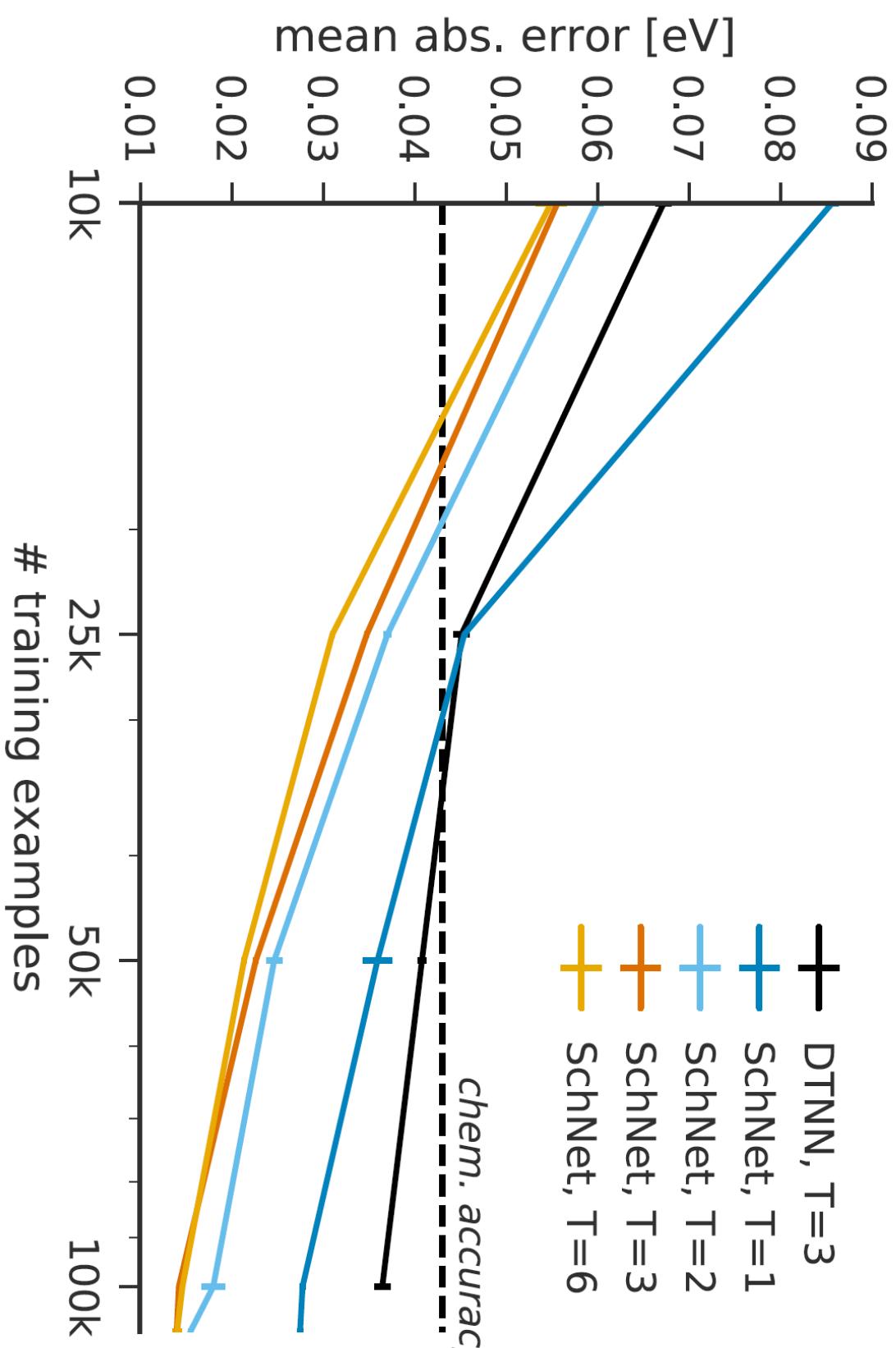


PIMD@SchNet shows delocalization of bonds



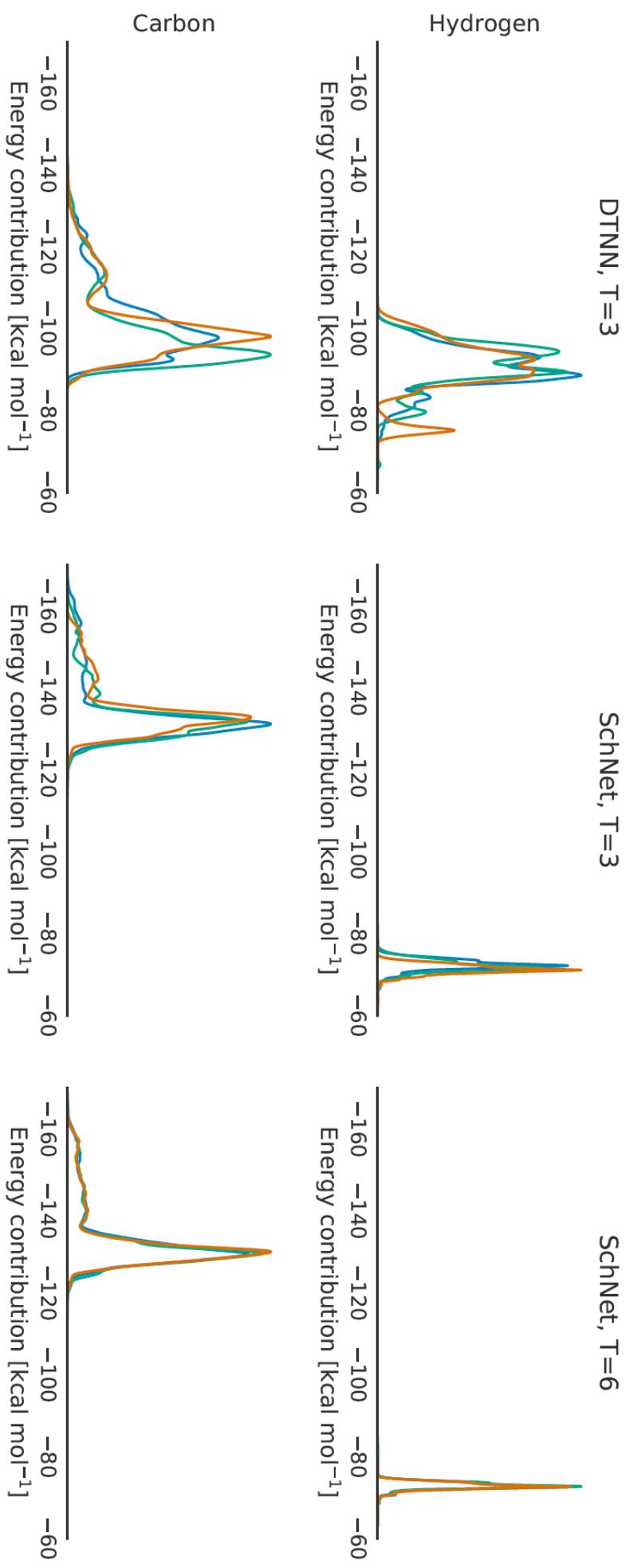
K.T. Schuett, H.E. Saucedo, P. J. Kindermans, S. Chmiela, A. Tkatchenko, K.-R. Mueller,  
*J. Chem. Phys.* **148**, 241722 (2018).

# Data Efficiency and Robustness of Deep Networks



K.T. Schuett, H.E. Sauceda, P. J. Kindermans, S. Chmiela, A. Tkatchenko, K.-R. Mueller,  
*J. Chem. Phys.* **148**, 241722 (2018).

# Data Efficiency and Robustness of Deep Networks

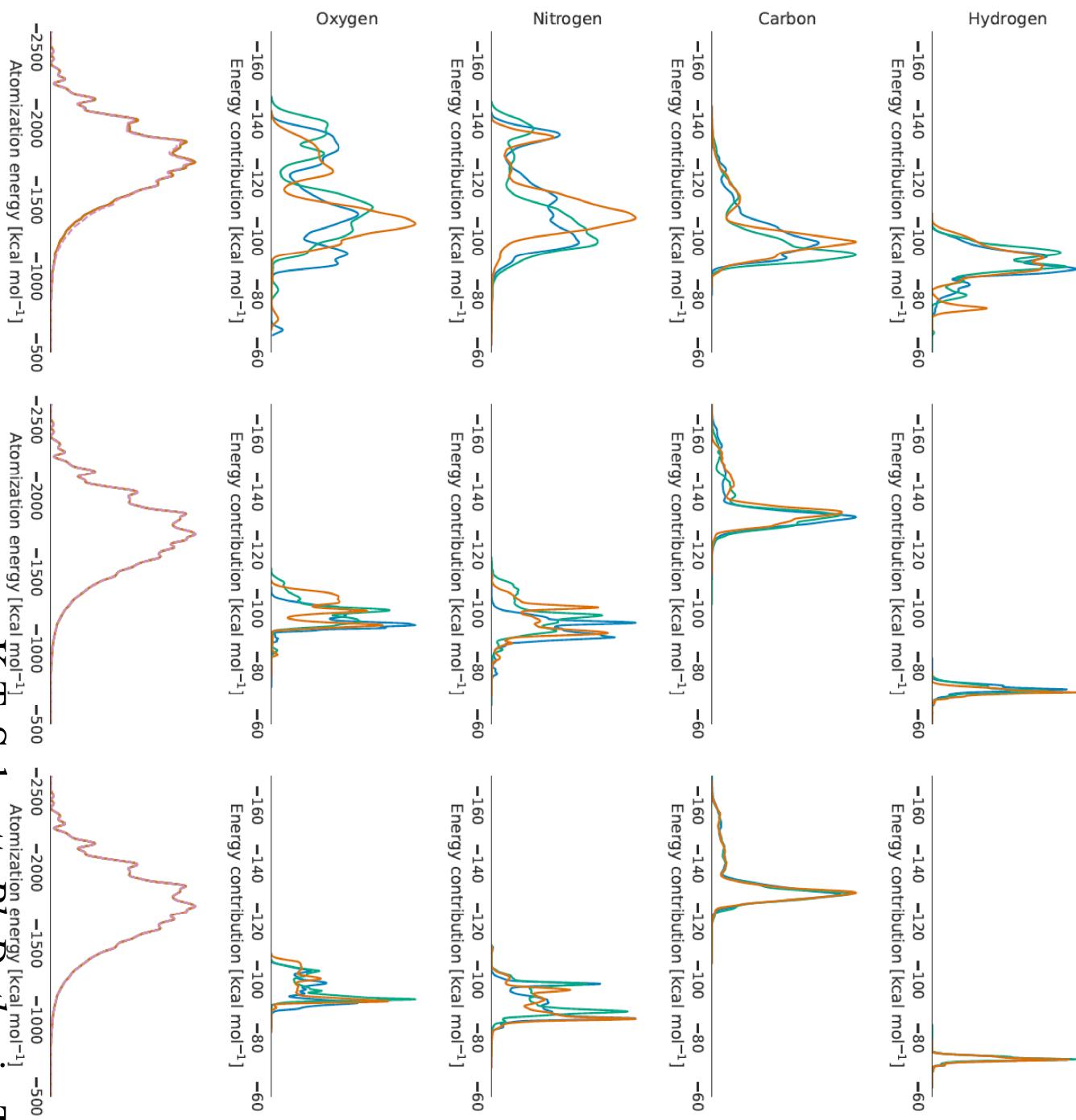


# Data Efficiency and Robustness of Deep Networks

DTNN, T=3

SchNet, T=3

SchNet, T=6



# Towards Exact Molecular Dynamics: Gradient-Domain Machine Learning (GDML)

# Gradient-Domain Machine Learning (GDML) for Conformational Dynamics

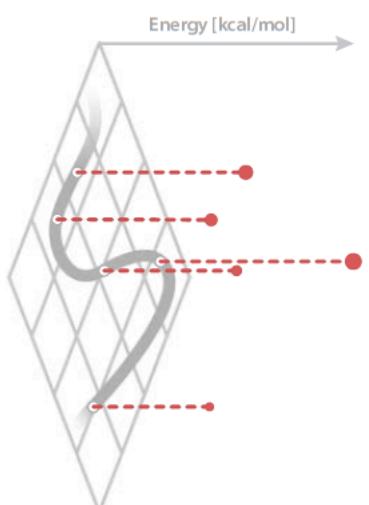
B Energy domain



Kernel



Energy [kcal/mol]



Energy samples  $V_{BO}$

Prediction

C Force domain

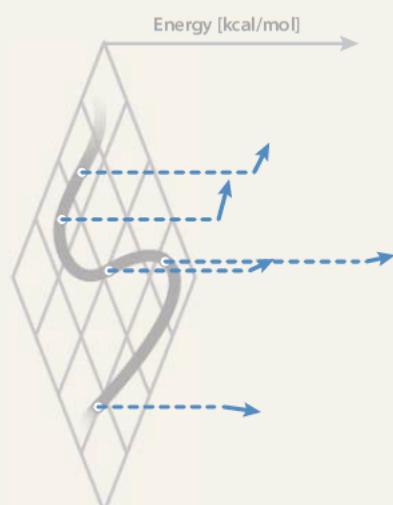


$K_{Hess}(\kappa)$



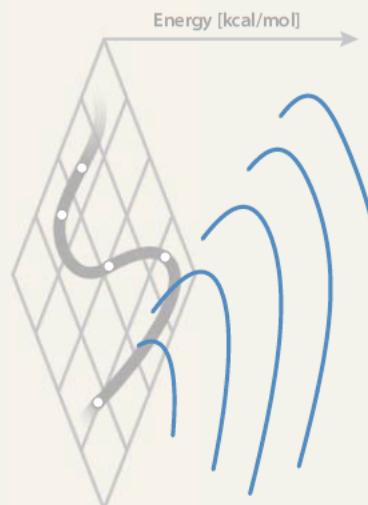
ML

Energy [kcal/mol]



Force samples  $F$

Integration

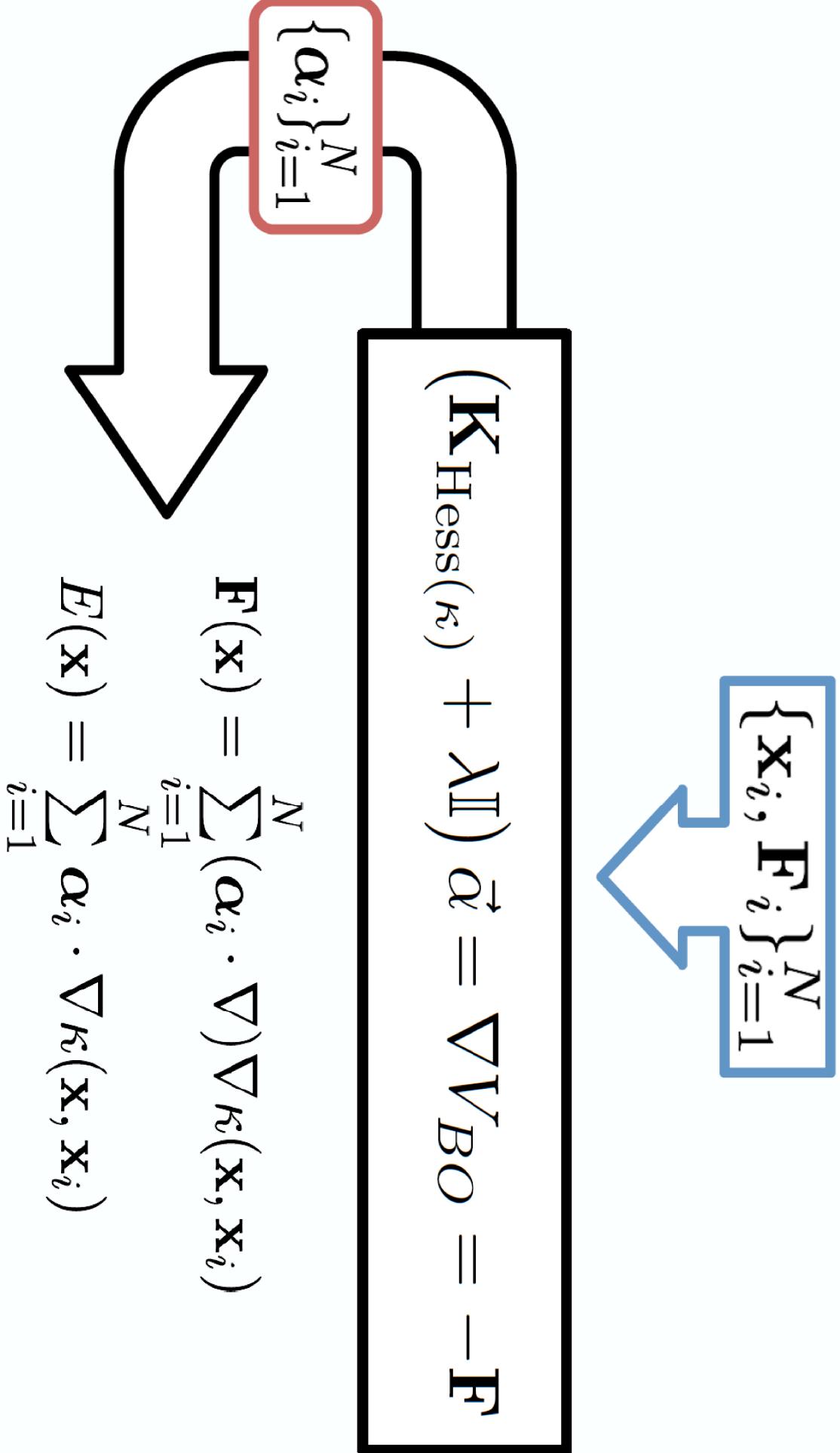


Prediction

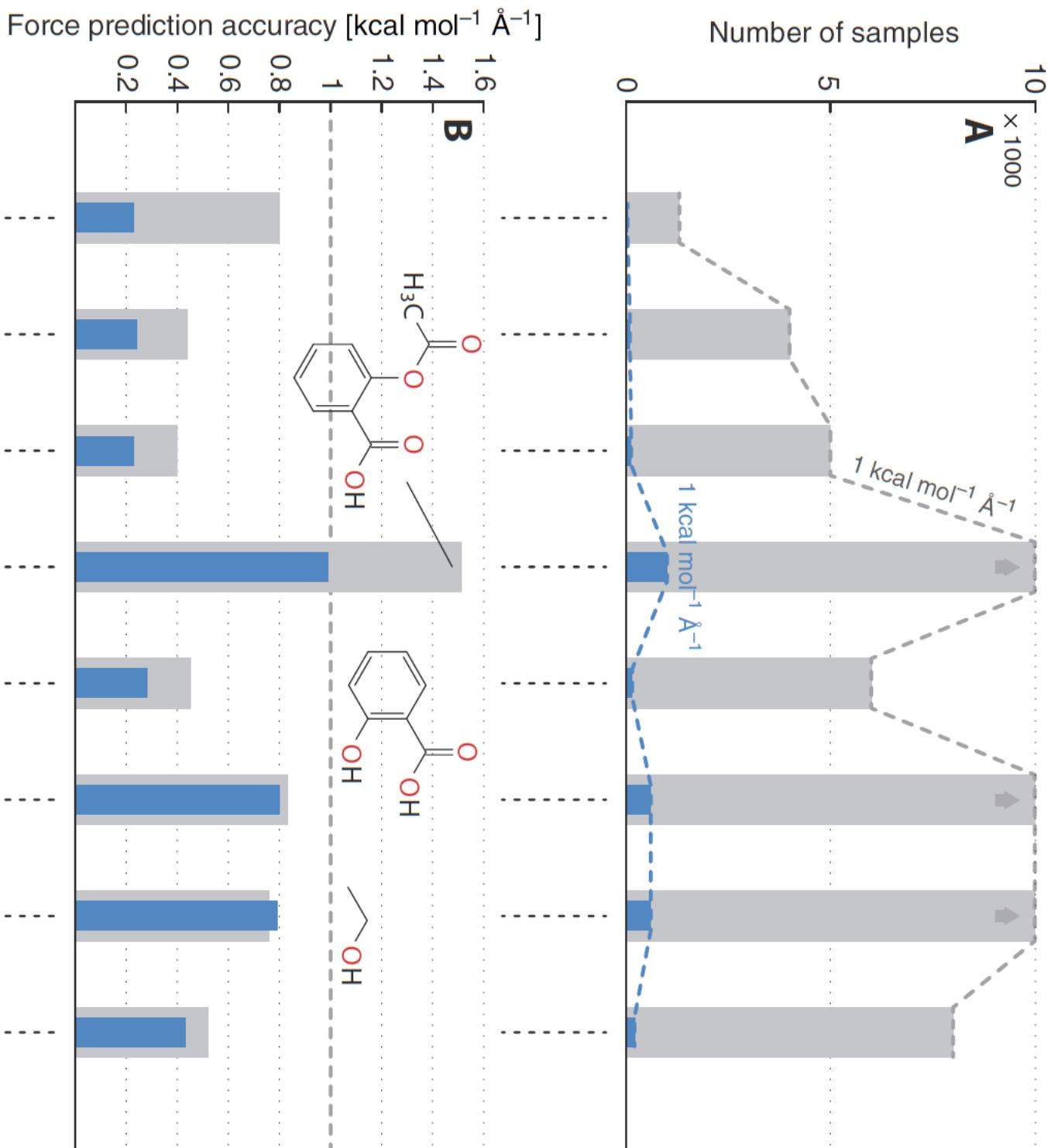


S. Chmiela, A. Tkatchenko, H. Sauceda, I. Poltavsky, K. T. Schuett, K.-R. Mueller, *Science Adv.* 3, e1603015 (2017).

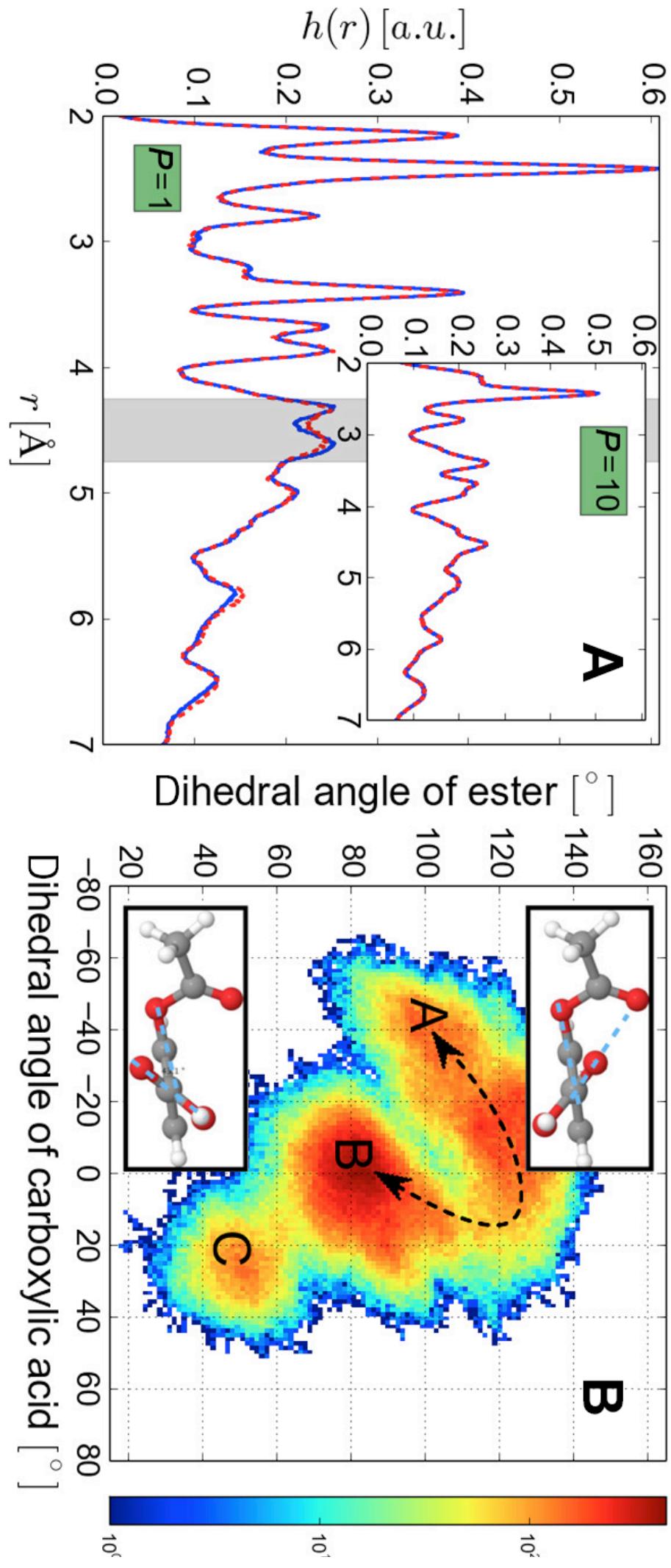
# Gradient-Domain Machine Learning (GDML) for Conformational Dynamics



# GDML model: Results for Small Molecules



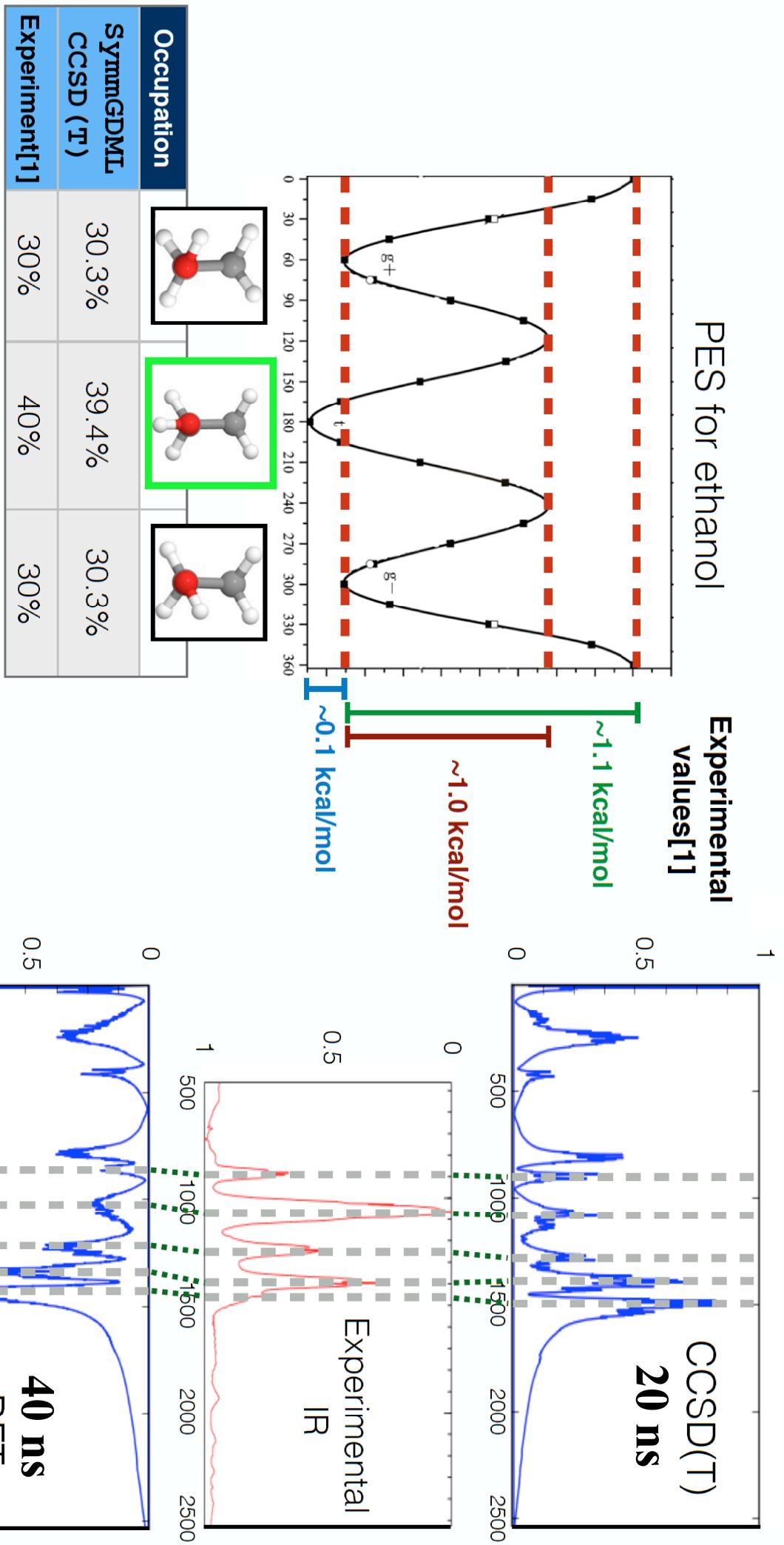
# GDMC model: Dynamics of Aspirin



S. Chmiela, A. Tkatchenko, H. Sauceda, I. Poltavsky, K. T. Schuett, K.-R. Mueller, *Science Adv.* 3, e1603015 (2017).

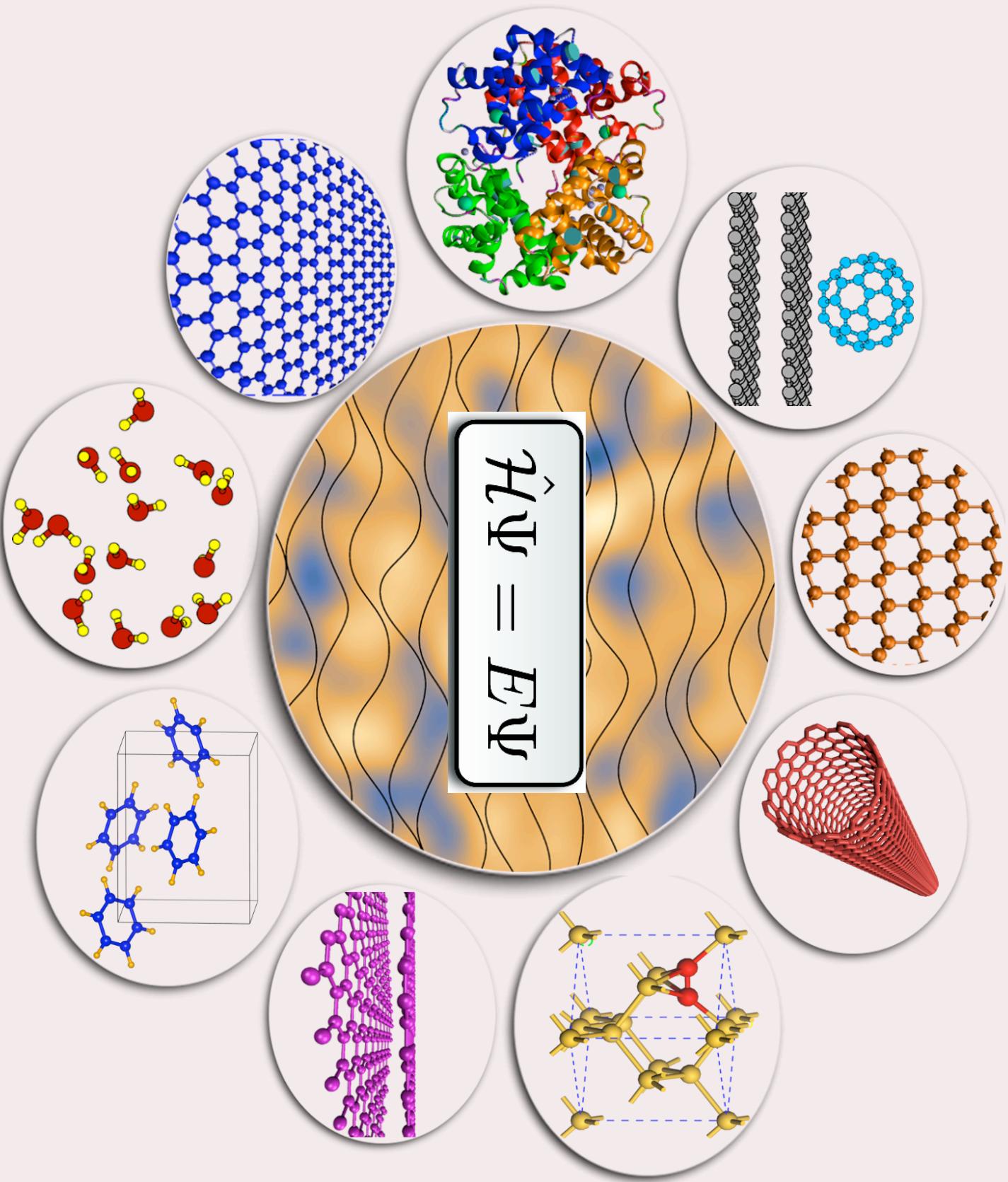
# Fully Quantum Dynamics for Ethanol: Quantized Electrons [CCSD(T)] and Nuclei [PiMD]

S. Chmiela, H. Sauceda, K.-R. Mueller,  
and A. Tkatchenko, *Nature Comms*, in press (2018).



# Towards Machine Learning Model of Chemical Compound Space

- DTNN excels for chemical bonding and GDMIL for molecular dynamics
- DTNN is accurate and insightful, but requires to be trained on big data
- GDMIL is accurate, requires little training data, but is specific to dynamics of a certain molecule
- How to best combine DTNN and GDMIL towards a unified model for the whole chemical compound space?



# Grand Challenges for Machine Learning in Physics/Chemistry

- *What is chemical space:* descriptors of molecules and materials, metric?
- *How to learn intensive properties:* energy levels, excited states, spectra?
- How to combine ML with physical laws (symmetries) and interaction models?
- Can we learn (approximate) Hamiltonians?
- Can ML suggest better approximations for SE?
- More and better (big) data

*Towards rational design of molecules and  
materials in chemical space*