

**Report on the Workshop**  
**“Developing High-Dimensional Potential Energy Surfaces:**  
**From the Gas Phase to Materials”**

**Göttingen, April 24-26, 2019**

**Organizers:**

**Jörg Behler, Georg-August-Universität Göttingen, Germany**

**Joel Bowman, Emory University, USA**

**Cábor Csányi, Cambridge University, UK**

**Alexander Kandratsenka, MPI für biophysikalische Chemie, Göttingen, Germany**

**Report**

The rapid progress in modern machine learning (ML) techniques has important consequences for almost all fields of science including chemical sciences and closely related fields like materials science and condensed matter physics. Among the most rapidly evolving applications of ML methods is their use in computer simulations with the aim to understand complex chemical reactions and to quantitatively predict properties of new materials. For this purpose, substantial progress has been made in the development of a new generation of accurate, ab initio based atomistic ML potentials, which provide a direct relation between the atomic configuration and the potential energy.

The workshop brought together leading researchers who develop and apply Machine Learning techniques with the common goal of determining the fundamental properties of "small" molecules, biomolecules, and materials. These properties include high-dimensional potential energy surfaces, atomic densities, and molecular properties, such as dipole moments and polarizabilities. Communities in the areas of materials, biomolecules, gas-phase molecules and complexes have formed over the past ten or so years and a major objective of the workshop was to bring these communities together to hear and learn from each other's experience. There were 19 invited talks and 31 poster contributions to the workshop. Talks were limited to 30 minutes, followed by 15 minutes of discussion. This format was adhered to and the discussions were very lively and informative.

The ML techniques that were described ranged from High-Dimensional Neural Networks (HDNN), Gaussian Process Regression (GPR), and Kernel Ridge Regression to Permutationally Invariant Polynomials (PIPs). HDNN and GPR approaches have been widely and successfully used in the area of materials, where atomic-based approaches are used to represent the potential energy surface and where density functional theory (DFT) is generally the electronic method of choice. DFT using ML and " $\Delta$ ML" to determine functionals was the focus of the talk by Burke. High dimensional NNs were described in a variety of contexts, including the interface between ZnO and water by Hellström, applications of atom-centered GPR using DFT were described for Si by Bernstein. Several talks focused on photochemistry and electronically non-adiabatic effects. These ranged from highly complex exciton processes in aggregates by Burghardt to long-time molecular dynamics propagation

by Marquetand and the challenges to fit the diabatic matrix, using PIP-NN, for a multiple electronic state application by Guo. Many-body ML strategies were also discussed and illustrated for materials using PIPS in talks by Dusson and Ortner and Kernel Ridge Regression by Rupp and for ion hydration by Paesani. ML using GPR for atomic densities, dipoles and polarizabilities was described by Wilkins. Several talks discussed force fields using ML, including a charge equilibration scheme, by Goedecker, and the ones by Meuwly, and Popelier focused on biomolecular force field. A direct approach using fast Tight Binding methods by described in the talk by Grimme. A talk focusing on aspects of fitting gradients and with  $\Delta$ ML to obtain highly accurate PESs especially for MD simulations was presented by Tkatchenko. The solid-liquid phase transition in water showed the power of HDNN in a talk by Cheng. GPR for extrapolation was presented in a talk by Krems. Using ML and Bayesian methods to navigate high-dimensional PESs was the focus of the talk by Todorović.



	Wednesday April 24, 2019	Thursday April 25, 2019	Friday April 26, 2019
08:00	Registration		
09:00	Welcome Address	Invited Talk 9 Noam Bernstein	Invited Talk 16 Roman Krems
	Invited Talk 1 Kieron Burke	Invited Talk 10 Stefan Grimme	Invited Talk 17 Francesco Paesani
10:00	Invited Talk 2 Matti Hellström	Coffee Break	Coffee Break
	Coffee Break	Invited Talk 11 David Wilkins	Invited Talk 18 Bingqing Cheng
11:00	Invited Talk 3 Irene Burghardt	Invited Talk 12 Paul Popelier	Invited Talk 19 Matthias Rupp
12:00	Invited Talk 4 Hua Guo	Lunch	Concluding Discussion
	Lunch	Invited Talk 13 Christoph Ortner	Lunch
13:00	Invited Talk 5 Markus Meuwly	Invited Talk 14 Geneviève Dusson	
14:00	Invited Talk 6 Philipp Marquetand	Invited Talk 15 Milica Todorovic	
15:00	Coffee Break	Coffee Break	
16:00	Invited Talk 7 Stefan Goedecker	transfer to city center	
	Invited Talk 8 Alexander Tkatchenko	Excursion	
17:00	Poster Session and Snacks		
22:00		Conference Dinner 19:30 - 22:00	

## List of Invited Talks

I1	Burke, Kieron	Machine learning of density functionals
I2	Hellström, Matti	Condensed-phase simulations using high-dimensional neural network potentials
I3	Burghardt, Irene	Ab initio based construction of molecular aggregate Hamiltonians for exciton dynamics
I4	Guo, Hua	Permutation Invariant Polynomial-Neural Network (PIP-NN) Approach to Nonadiabatic Potential Energy Surfaces Beyond the Born-Oppenheimer Approximation
I5	Meuwly, Markus	Insights into Chemical Reactivity from Empirical and Neural Network-Learned Reactive Force Fields
I6	Marquetand, Philipp	Simulation of ground and excited-state molecular dynamics with machine learning
I7	Goedecker, Stefan	A force field based on a charge equilibration scheme steered by a neural network
I8	Tkatchenko, Alexandre	Towards Exact Molecular Dynamics Simulations with Quantum Chemistry and Machine Learning
I9	Bernstein, Noam	Machine learning an interatomic potential without blood, sweat, and tears
I10	Grimme, Stefan	New Tight-binding Quantum Chemistry Methods
I11	Wilkins, David	Symmetry-Adapted Machine Learning: from Tensors to Charge Densities
I12	Cheng, Bingqing	Ab initio thermodynamics with the help of machine learning
I13	Ortner, Christoph	Interatomic Potentials from Linear Fits
I14	Dusson, Geneviève	Atomic Permutation-Invariant Potential
I15	Todorović, Milica	Navigating High-Dimensional Energy Landscapes
I16	Krems, Roman	Gaussian Processes for interpolation of PES and extrapolation of physical properties
I17	Paesani, Francesco	Modeling Hydration, One Water Molecule at a Time
I18	Popelier, Paul	FFLUX: on knowledgeable quantum atoms
I19	Rupp, Matthias	Exact representations of molecules and materials for accurate interpolation of ab initio simulations

## Abstracts of the Talks

I01

# Machine learning of density functionals

Kieron Burke

Departments of Chemistry and of Physics, UC Irvine

E-mail: [kieron@uci.edu](mailto:kieron@uci.edu)

I will discuss how to make new density functionals using kernel ridge regression. These functionals are currently limited to the systems they are trained on but, unlike standard semilocal approximations, have no difficulties with bond breaking or strong correlation. I will present our most recent results.

### References:

- [1] Finding Density Functionals with Machine Learning Snyder, John C., Rupp, Matthias, Hansen, Katja, Müller, Klaus-Robert and Burke, Kieron, *Phys. Rev. Lett.* 108, 253002 (2012).
- [2] Understanding machine-learned density functionals Li, Li, Snyder, John C., Pelaschier, Isabelle M., Huang, Jessica, Niranjana, Uma-Naresh, Duncan, Paul, Rupp, Matthias, Müller, Klaus-Robert and Burke, Kieron, *International Journal of Quantum Chemistry* 16, 819–833 (2016).
- [3] Understanding kernel ridge regression: Common behaviors from simple functions to density functionals Vu, Kevin, Snyder, John C., Li, Li, Rupp, Matthias, Chen, Brandon F., Khelif, Tarek, Müller, Klaus-Robert and Burke, Kieron, *International Journal of Quantum Chemistry* 115, 1115–1128 (2015).
- [4] By-passing the Kohn-Sham equations with machine learning Brockherde, Felix, Vogt, Leslie, Li, Li, Tuckerman, Mark E, Burke, Kieron and Muller, Klaus-Robert, *Nature Communications* 8, 872 (2017).
- [5] Pure density functional for strong correlation and the thermodynamic limit from machine learning Li, Li, Baker, Thomas E., White, Steven R. and Burke, Kieron, *Phys. Rev. B* 94, 245129 (2016).
- [6] Can exact conditions improve machine-learned density functionals? Jacob Hollingsworth, Li, Li, Thomas E. Baker and Kieron Burke, *The Journal of Chemical Physics* 148, 241743 (2018).

# Condensed-phase simulations using high-dimensional neural network potentials

Matti Hellström<sup>1</sup>

<sup>1</sup> Software for Chemistry & Materials BV, Amsterdam, The Netherlands

E-mail: [hellstrom@scm.com](mailto:hellstrom@scm.com)

High-dimensional neural network potentials (HDNNPs) can be parameterized to accurately reproduce potential energy surfaces from electronic structure calculations. From dispersion-corrected DFT data, we parameterized HDNNPs for the full room-temperature solubility range of NaOH in water [1], and for the interface between ZnO and liquid water [2], and applied them to large-scale molecular dynamics simulations in which we characterized the rates and mechanisms of different proton transfer reactions [1,2], finding unexpected similarities between the two systems. Moreover, we explored the role of nuclear quantum effects in NaOH solutions [3] and how the ZnO surface structure influences the long-range Grotthuss-like proton transport at the ZnO/water interface [4].

This talk also discusses the challenging case of modeling metal oxides, for which the electronic band structure strongly influences defect formation energies [5] and even the coverage-dependence of adsorption energies on surfaces [6].

## References:

- [1] M. Hellström, J. Behler. *J. Phys. Chem. Lett.* **7** (2016) 3302
- [2] V. Quaranta, M. Hellström, J. Behler. *J. Phys. Chem. Lett.* **8** (2017) 1476
- [3] M. Hellström, M. Ceriotti, J. Behler. *J. Phys. Chem. B* **122** (2018) 10158
- [4] M. Hellström, V. Quaranta, J. Behler. *Chem. Sci.* **10** (2019) 1232
- [5] C. Freysoldt et al. *Rev. Mod. Phys.* **86** (2014) 253
- [6] M. Hellström, J. Behler. *Phys. Chem. Chem. Phys.* **19** (2017) 28731

# Ab initio based construction of molecular aggregate Hamiltonians for exciton dynamics

Irene Burghardt<sup>1</sup> and Robert Binder<sup>1</sup>

<sup>1</sup> Department of Physical and Theoretical Chemistry, Goethe University Frankfurt, Germany

E-mail: [burghardt@chemie.uni-frankfurt.de](mailto:burghardt@chemie.uni-frankfurt.de)

This talk addresses recent efforts towards the *ab initio* based construction of parametrized model Hamiltonians describing energy and charge transfer in molecular aggregates coupled to intra- and intermolecular vibrations. While the use of such Hamiltonians, including the Frenkel-Holstein Hamiltonian for exciton transport and related charge-transport Hamiltonians of Su-Schrieffer-Heeger (SSH) type, is well established, the direct parametrization of such models based on electronic structure data is attempted less frequently. Here, we report on two types of approaches that are capable to provide an accurate translation of electronic structure information to potential surfaces adapted to the relevant model Hamiltonians: (i) First, a mapping procedure which was developed for J-type and H-type homo-aggregate systems [1] as well as combined HJ-aggregates [2]; this procedure is based upon the analytic solution of an inverse eigenvalue problem for an effective Frenkel Hamiltonian with nearest-neighbor couplings. (ii) Second, the use of vibronic coupling models which are parametrized by diabaticization procedures in conjunction with the determination of linear or higher-order vibronic coupling parameters [3]. Several examples will be given of quantum dynamical treatments of ultrafast exciton transfer [4] and exciton dissociation [3,5] based upon these *ab initio* parametrized model Hamiltonians, and a perspective is given on further developments.

## References:

- [1] R. Binder, S. Römer, J. Wahl, I. Burghardt, *J. Chem. Phys.* **141** (2014) 014101.
- [2] R. Binder, M. Polkehn, T. Ma, I. Burghardt, *Chem. Phys.* **482** (2017) 16.
- [3] M. Polkehn, P. Eisenbrandt, H. Tamura, I. Burghardt, *Int. J. Quant. Chem.* **118** (2018) e25502.
- [4] R. Binder, D. Lauvergnat, I. Burghardt, *Phys. Rev. Lett.* **120** (2018) 227401.
- [5] M. Polkehn, H. Tamura, I. Burghardt, *J. Phys. B: At. Mol. Opt. Phys.* **51** (2018) 014003.

# Permutation Invariant Polynomial-Neural Network (PIP-NN) Approach to Nonadiabatic Potential Energy Surfaces Beyond the Born-Oppenheimer Approximation

Hua Guo<sup>1</sup> and Changjian Xie<sup>1</sup>

<sup>1</sup> Department of Chemistry and Chemical Biology, University of New Mexico, USA

E-mail: [hguo@unm.edu](mailto:hguo@unm.edu)

It is well established that the Born-Oppenheimer adiabatic potential energy surface (PES) is invariant under permutation of identical nuclei in a molecule. The enforcement of the permutation invariance is not only essential to correctly characterize spectra and dynamics, but also numerically beneficial. One approach to enforce permutation symmetry is to use the atomistic neural network (NN) method of Behler, in which each atom type is associated with a NN.<sup>[1]</sup> This can also be achieved using permutation invariant polynomials (PIPs) either as the interpolating basis<sup>[2]</sup> or as the first layer of a global NN.<sup>[3]</sup> The latter is the so-called PIP-NN approach, and it has been demonstrated to be accurate and efficient for few-atom systems. In this talk, I will discuss the extension of the PIP-NN method to constructing diabatic potential energy matrices (PEMs), particularly when conical intersections are involved. It is shown that the elements of the diabatic PEM are not always permutationally invariant, and system-dependent symmetry considerations are needed to ensure the permutation symmetry in the corresponding adiabatic PESs.<sup>[4]</sup> Dynamics on the PIP-NN PEMs confirm the accuracy of this approach.

## References:

- [1] J. Behler, *Int. J. Quant. Chem.* **2015**, *115*, 1032-1050.
- [2] B. J. Braams, J. M. Bowman, *Int. Rev. Phys. Chem.* **2009**, *28*, 577-606.
- [3] B. Jiang, J. Li, H. Guo, *Int. Rev. Phys. Chem.* **2016**, *35*, 479-506.
- [4] C. Xie, X. Zhu, D. R. Yarkony, H. Guo, *J. Chem. Phys.* **2018**, *149*, 144107.

I05

# Insights into Chemical Reactivity from Empirical and Neural Network-Learned Reactive Force Fields

Markus Meuwly<sup>1</sup>

<sup>1</sup> Department of Chemistry, University of Basel, Switzerland

E-mail: m.meuwly@unibas.ch

Following and characterizing the dynamics accompanying chemical reactions in gas phase and in solution is one of the essential aims of chemistry. Although *ab initio* and mixed quantum/classical molecular dynamics simulations are in principle viable approaches they are typically too slow to comprehensively sample configurational space for converging reaction probabilities or free energies of reaction. In this talk I will present recent efforts to use empirical and neural network-based force fields to investigate the energetics of chemical reactions in gas phase and in solution. A particular emphasis is on comparing with experimental data and providing molecular-level insight beyond what analysis of state-of-the art experiments can obtain.

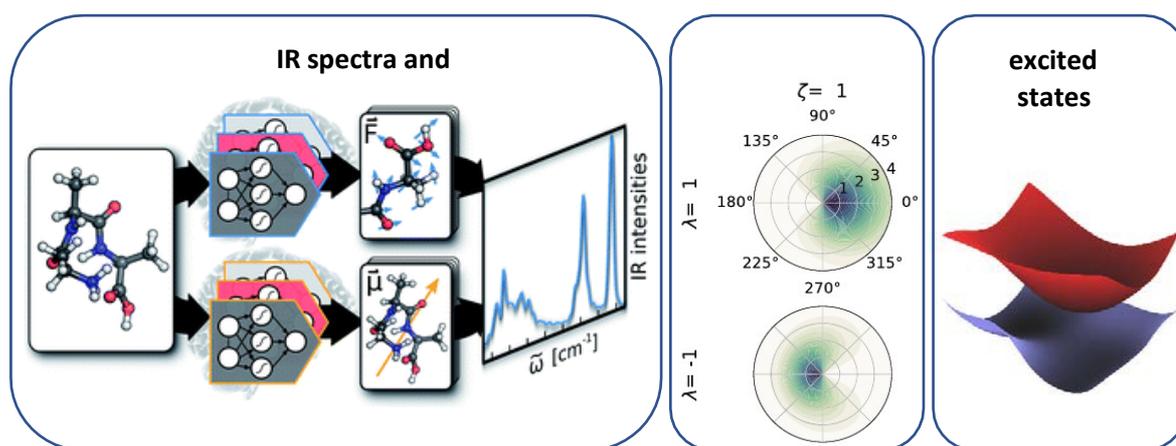
# Simulation of ground and excited-state molecular dynamics with machine learning

Philipp Marquetand<sup>1</sup>

<sup>1</sup> Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Austria

E-mail: [philipp.marquetand@univie.ac.at](mailto:philipp.marquetand@univie.ac.at)

Different applications and developments of machine learning potentials will be presented, aiming at fast and accurate molecular dynamics simulations. As a first example, the simulation of infrared spectra computed including anharmonic effects is chosen [1,2]. Here, high-dimensional neural networks exploit chemical locality [3,4] and predict accurate potentials from only a few training points chosen with an adaptive sampling scheme [1,5]. Emphasis is put on the intrinsic charge model derived from molecular dipole moments [1,2]. Furthermore, weighted atom-centered symmetry functions (wACSFs) are discussed as descriptors [6]. Finally, machine learning potentials are used to carry out excited-state molecular dynamics and open up the possibility to reach time scales in the nanosecond regime [7].



## References:

- [1] M. Gastegger, J. Behler, P. Marquetand, *Chem. Sci.*, **8** (2017) 6924-6935.
- [2] M. Gastegger, P. Marquetand, arXiv:1812.07676 [physics.chem-ph] (2018).
- [3] J. Behler, M. Parrinello, *Phys. Rev. Lett.*, **98** (2007) 146401.
- [4] M. Gastegger, C. Kauffmann, J. Behler, P. Marquetand, *J. Chem. Phys.*, **144** (2016) 194110.
- [5] J. Behler, *J. Phys.: Condens. Matter*, **26** (2014) 183001.
- [6] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, P. Marquetand, *J. Chem. Phys.*, **148** (2018) 241709.
- [7] J. Westermayr, M. Gastegger, M. Menger, S. Mai, L. González, P. Marquetand, arXiv:1811.09112 [physics.chem-ph] (2018).

# A force field based on a charge equilibration scheme steered by a neural network

Stefan Goedecker<sup>1</sup>

<sup>1</sup> Department of Physics, University of Basel, Switzerland

E-mail: [Stefan.Goedecker@unibas.ch](mailto:Stefan.Goedecker@unibas.ch)

Most existing atomic environment descriptors are short range, in the sense that they probe only a limited number of neighboring atoms within a finite sphere around the central atom whose environment has to be characterized. While in many cases chemical properties depend only on local properties, this is not true anymore if long range charge transfer is important in the system. I will present a machine learning force field, that is based on a charge equilibration scheme and that can therefore describe long range charge transfer or even ionized systems. I will describe some structure prediction results based on this force field. I will also discuss the resolution power of various atomic fingerprints when they are used to distinguish different atomic environments.

References:

Ghasemi, S. Alireza and Hofstetter, Albert and Saha, Santanu and Goedecker, Stefan: Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network, Phys. Rev. **92** 045131 (2015)

# Towards Exact Molecular Dynamics Simulations with Quantum Chemistry and Machine Learning

Alexandre Tkatchenko<sup>1</sup>

<sup>1</sup> Physics and Materials (PhyMS) Research Unit, University of Luxembourg, Luxembourg

Email: alexandre.tkatchenko@uni.lu

I will discuss an efficient symmetric gradient-domain machine learning (sGDML) approach for constructing next-generation CCSD(T)-quality interatomic force fields for molecules. The sGDML implementation is able to reproduce global potential-energy surfaces of intermediate-sized molecules with an accuracy of 0.1 kcal/mol using only few hundred molecular conformations for training. We demonstrate this accuracy for *ab initio* molecular dynamics (AIMD) trajectories of molecules, including toluene, naphthalene, ethanol, uracil, and aspirin. The GDML approach enables quantitative molecular dynamics simulations with quantum electrons and nuclei for molecules at a fraction of cost of explicit AIMD calculations, thereby allowing to achieve unprecedented insights into (thermo)dynamics of molecules [1,2].

References:

- [1] S. Chmiela, A. Tkatchenko, H.E. Sauceda, I. Poltavsky, K.T. Schütt, and K.-R. Müller, Machine Learning of Accurate Energy-Conserving Molecular Force Fields. **Science Adv.** 3, 1603015 (2017).
- [2] S. Chmiela, H. E. Sauceda, K. R. Mueller, and A. Tkatchenko, Towards exact molecular dynamics simulations with machine-learned force fields. **Nature Commun.** 9, 3887 (2018).

# Machine learning an interatomic potential without blood, sweat, and tears

**Noam Bernstein**,<sup>1</sup> **Gábor Csányi**,<sup>2</sup> and **Volker L. Deringer**<sup>2</sup>

<sup>1</sup> Center for Materials Physics and Technology, U. S. Naval Research Laboratory, USA

<sup>2</sup> Department of Engineering, University of Cambridge, UK

E-mail: [noam.bernstein@nrl.navy.mil](mailto:noam.bernstein@nrl.navy.mil)

Defining interatomic potentials using ideas from machine learning that treat the problem as a high-dimensional fit of the reference (usually density functional theory) potential energy surface is an exciting new approach for developing accurate potentials. However, because of their variational freedom, such potentials require large fitting datasets, with large amounts of manual selection and tuning of configurations by the researcher [1]. We present an iterative method, where a preliminary potential is used to carry out a number of random-structure searches, and selected configurations from the searches are used to fit the next iteration's potential. We test the method on a number of elements with different bonding types, including an insulator, a semiconductor, and a metal. We show how the process converges in a few iterations, and how the resulting potentials reproduce the reference DFT values on a number of bulk and defect properties.

References:

[1] "Machine Learning a General-Purpose Interatomic Potential for Silicon," A. P. Bartók *et al.*, *Phys. Rev. X* **8** (2018) 041048.

# New Tight-binding Quantum Chemistry Methods

Stefan Grimme<sup>1</sup>

<sup>1</sup> Mulliken Center for Theoretical Chemistry, University Bonn, Germany

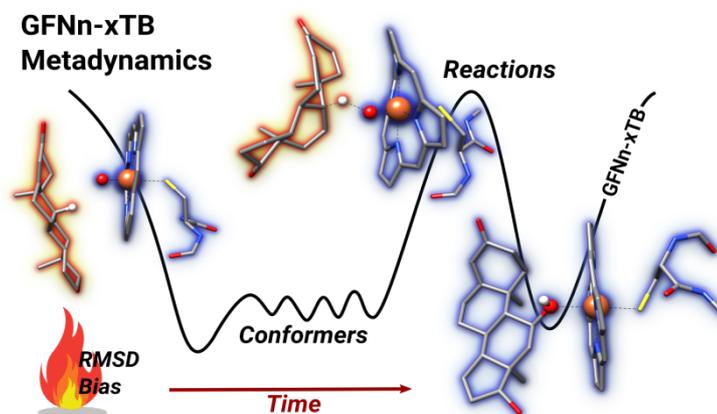
E-mail: [grimme@thch.uni-bonn.de](mailto:grimme@thch.uni-bonn.de)

The GFN-xTB family of semi-empirical tight-binding methods, which are variants of the well-known DFTB approaches, is introduced. The methods follow a global and element-specific parameters only strategy and are consistently parameterized for all elements through radon. Their original purpose and main target for the parameter optimization has been the computation of molecular geometries, vibrational frequencies, and non-covalent interactions. Very recently, the original GFN-xTB has been extended by including multipole electrostatic as well as one-center exchange-correlation terms leading to higher accuracy (at lower empiricism) specifically for non-covalent interactions and conformational energies[1]. A new, much faster (speedup of 3-5) first-order form employing classical electrostatics is briefly described (GFN0-xTB) The most sophisticated GFN2-xTB approach which furthermore employs density dependent D4 dispersion, is effectively used in the framework of meta-dynamics (MTD) to globally explore chemical compound, conformer, and reaction space[2]. The biasing potential given as a sum of Gaussian functions is expressed with the RMSD in Cartesian space as a metric for the collective variables. For typical conformational search problems of drug-like organic molecules, the new MTD(RMSD) algorithm yields lower energy structures and more complete conformer ensembles at reduced computational effort. Because TB methods (when combined with the Fermi-smearing technique) can also describe difficult electronic situations at least qualitatively correct, chemical reaction space exploration in a virtual nanoreactor also for transition metal containing systems is routinely possible.

References:

[1] C. Bannwarth, S. Ehlert, S. Grimme *J. Chem. Theory Comput.* DOI: 10.1021/acs.jctc.8b01176

[2] S. Grimme, DOI: 10.26434/chemrxiv.7660532.v1



# Symmetry-Adapted Machine Learning: from Tensors to Charge Densities

Andrea Grisafi,<sup>1</sup> David M. Wilkins,<sup>1</sup> and Michele Ceriotti<sup>1</sup>

<sup>1</sup> Department of Materials Science, École Polytechnique Fédérale de Lausanne, Switzerland

E-mail: [david.wilkins@epfl.ch](mailto:david.wilkins@epfl.ch)

A full description of a physical system requires the knowledge not only of scalar properties that are invariant under transformations such as translations and rotations, but also of properties that transform in specified ways when these operations are applied. Examples include response tensors and scalar fields, which are rotated when a rigid-body rotation is applied to the system. In order to learn these properties, we require a method that takes into account these transformations.

I describe a formalism that extends standard Gaussian process methods to learn tensor properties, giving predictions that rotate covariantly with the reference frame [1]. This framework requires kernels that are adapted to this symmetry, realized as a generalization of the smooth overlap of atomic positions (SOAP) kernels. This method, called symmetry-adapted Gaussian process regression (SA-GPR), is able to learn tensors of arbitrary rank for systems of arbitrary complexity.

In particular, I describe two recent applications of SA-GPR: the first of these is the prediction of molecular polarizabilities  $\alpha$ , using a model trained on a dataset of small molecules [2]: this model is able to predict the coupled-cluster polarizability of a molecule with an accuracy better than density functional theory, even when extrapolated to larger and more complex molecules; the prediction of the molecular  $\alpha$  opens the way to the design of polarizable forcefields and the simulation of spectroscopic experiments.

A further application is the prediction of the electronic charge density of a molecule without the need for electronic structure calculations, learning this scalar field as a decomposition into atom-centred spherical harmonic components. Once again, a model trained on smaller molecules can be extrapolated to give good predictions of the electron density of larger molecules, showing the versatility and power of the SA-GPR method. I will also describe the outlook for this framework and its future applications.

## References:

- [1] A. Grisafi, D. M. Wilkins, G. Csányi, M. Ceriotti, *Phys. Rev. Lett.* **120** (2018) 036002.
- [2] D. M. Wilkins, A. Grisafi, Y. Yang, K.-U. Lao, R. A. DiStasio, M. Ceriotti, *Proc. Natl. Acad. Sci.* **116** (2019) 3401.
- [3] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, M. Ceriotti, *ACS Cent. Sci.* **5** (2019) 57.

# Ab initio thermodynamics with the help of machine learning

**Bingqing Cheng,<sup>1</sup> Michele Ceriotti<sup>2</sup>**

<sup>1</sup> Department of Physics, University of Cambridge, UK

<sup>2</sup> Department of Materials Science and Engineering, EPFL, Switzerland

E-mail: [tonicbq@gmail.com](mailto:tonicbq@gmail.com)

A central goal of computational physics and chemistry is to predict material properties using first principles methods based on the fundamental laws of quantum mechanics. However, the high computational costs of these methods typically prevent rigorous predictions of macroscopic quantities at finite temperatures, such as heat capacity, density, and chemical potential.

In this talk, I will discuss how to enable such predictions by combining advanced free energy methods with data-driven machine learning interatomic potentials. I will show that, for the omnipresent and technologically essential system of water, a first-principles thermodynamic description not only leads to excellent agreement with experiments, but also reveals the crucial role of nuclear quantum fluctuations in modulating the thermodynamic stabilities of different phases of water.

## References:

- [1] B. Cheng, J. Behler, M. Ceriotti, *Journal of Physical Chemistry Letters* 7 (2016) 2210-2215.
- [2] B. Cheng, M. Ceriotti, *Physical Review B* 97 (2018) 054102.
- [3] B. Cheng, E. A. Engel, J. Behler, C. Dellago, M. Ceriotti, *Proceedings of the National Academy of Sciences* 116 (2019) 1110-1115.

# I13

## Interatomic Potentials from Linear Fits

Christoph Ortner

Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL,  
UK

E-mail: [C.Ortner@warwick.ac.uk](mailto:C.Ortner@warwick.ac.uk)

I will discuss different approaches to constructing Interatomic Potentials from linear fits, including the roles of descriptors, basis sets, regression methods and in particular regularisation. Potentials in this category include GAP (Bartok, Csanyi, ...), MTPs (Shapeev), PIPs (Bownman, Braams), SNAP (Thompson, ...), and others. I will also explore the connection between RMSE and errors in quantities of interest (e.g. defect formation energy).

**Atomic Permutation-Invariant Potential**

**Gábor Csányi<sup>1</sup>, Geneviève Dusson<sup>2</sup>, Christoph Ortner<sup>2</sup> and Cas Van der Ord<sup>1</sup>**

<sup>1</sup> Department of Engineering, University of Cambridge, United Kingdom

<sup>2</sup> Department of Mathematics, University of Warwick, United Kingdom

E-mail: [g.dusson@warwick.ac.uk](mailto:g.dusson@warwick.ac.uk)

In this talk, I will present a potential targeted for materials simulations combining a physically motivated functional form: an atomic body-order expansion, with a data-driven linear fitting procedure. On top of being systematically improvable, this potential is constructed to satisfy the rotation and permutation invariance of the problem.

I will report our first experience with a practical implementation, demonstrating a competitive accuracy vs cost ratio, as well as some additional useful properties, such as the possibility of regularising the functional, leading to an accurate prediction of some physical properties.

# Navigating High-Dimensional Energy Landscapes

**Milica Todorović**<sup>1</sup>

<sup>1</sup> Department of Applied Physics, Aalto University, Finland

E-mail: [milica.todorovic@aalto.fi](mailto:milica.todorovic@aalto.fi)

Atomistic structure search at organic/inorganic interfaces is made complex by the many degrees of freedom and the need for accurate but costly density-functional theory (DFT) simulations. To accelerate and simplify structure determination in such heterogeneous functional materials, we developed the Bayesian Optimization Structure Search (BOSS).

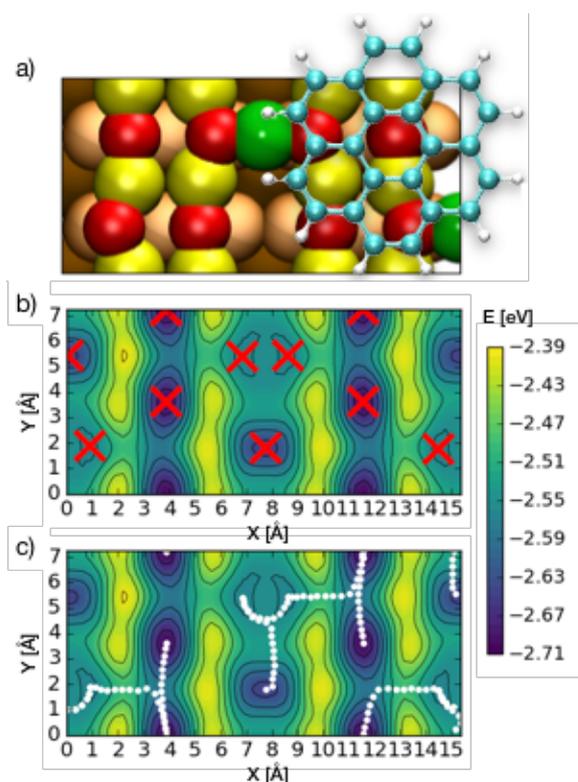
Bayesian optimization was employed to build N-dimensional surrogate models for the energy landscapes [1] and infer global minima. The models were iteratively improved by sequentially sampling DFT data points that are promising and/or have high information content. We represented heterogeneous materials with chemical ‘building blocks’, which reduced dimensionality and allowed us to search the configurational phase space, instead of the chemical phase space.

We applied this active learning scheme to molecular surface adsorption, as illustrated in Figure 1. Global minima were identified with reasonable computational efficiency [2]. BOSS produced chemically-intuitive adsorption energy landscapes. These could be parsed for local minima and the minimum energy paths between them, allowing us to extract energy barriers and the corresponding atomistic pathways in N-dimensions.

References:

[1] M.U. Gutmann and J. Corander, *J. Mach. Learn. Res.* **17** (2016) 1.

[2] M. Todorović, M.U. Gutmann, J. Corander and P. Rinke, *npj Comput. Mater.* **5** (2019) 35.



**Figure 1.** BOSS inference of coronene adsorption on Cu(110)-O c(6x2). A) DFT model; b) adsorption energy landscape with local minima; c) minimum energy paths.

# Gaussian Processes for interpolation of PES and extrapolation of physical properties

Roman Krems, Rodrigo Vargas-Hernandez, and Yun-wen Mao

Department of Chemistry, University of British Columbia, Vancouver, Canada

E-mail: [rkrems@chem.ubc.ca](mailto:rkrems@chem.ubc.ca)

This talk will discuss three different but related problems [1-4]. First, I will discuss the advantages and limitations of fitting multi-dimensional potential energy surfaces (PES) by Gaussian Process (GP) regression. In particular, I will argue that GPs can be used to construct accurate PES with a small number of ab initio points [2] and automate the construction of multi-dimensional PES. The latter can be used to implement Bayesian optimization for the inverse scattering problem [3]. I will illustrate an algorithm where Bayesian optimization is trained to modify the PES in response to feedback from scattering calculations, with the goal of producing a PES yielding an agreement with desired quantum dynamics results. Second, I will discuss the role of sampling the configuration space for the construction of accurate PES by means of GP regression. I will use Bayesian optimization to determine the optimal sampling distributions for a variety of polyatomic systems and compare the results with those obtained using random sampling. Finally, I will discuss the feasibility of using GP regression for extrapolation rather than interpolation. In order to build GP models capable of extrapolation, it is necessary to construct complex kernels capturing the physical evolution of the properties of interest. Such kernels can be constructed by an iterative procedure combining simple kernels into linear combinations of products [4] and using the Bayesian information criterion for kernel selection [5]. I will show that GP models thus built are capable of extrapolating the properties of complex quantum systems across quantum phase transitions [6].

References:

- [1] J. Cui and R. V. Krems, "Gaussian Process Model for Collision Dynamics of Complex Molecules", *Phys. Rev. Lett.* **115**, 073202 (2015).
- [2] J. Cui and R. V. Krems, "Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes", *J. Phys. B* **49**, 224001 (2016)
- [3] R. Vargas-Hernandez, Y. Guan, D. H. Zhang, and R. V. Krems, "Bayesian optimization for the inverse scattering problem in quantum reaction dynamics", *New J. Phys.* **21**, 022001 (2019)
- [4] D. K. Duvenaud, H. Nickisch, and C. E. Rasmussen, "Additive Gaussian Processes", *Adv. Neural Inf. Process. Syst.* **24**, 226 (2011).
- [5] D. K. Duvenaud, J. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, "Structure Discovery in Nonparametric Regression through Compositional Kernel Search", *Proc. Mach. Learn. Res.* **28**, 1166 (2013).
- [6] R. Vargas-Hernandez, J. Sous, M. Berciu, and R. V. Krems, "Extrapolating quantum observables with machine learning: Inferring multiple phase transitions from properties of a single phase", *Phys. Rev. Lett.* **121**, 255702 (2018).

# Modeling Hydration, One Water Molecule at a Time

**Francesco Paesani,<sup>1</sup> Marc Riera,<sup>1</sup> Colin Egan,<sup>1</sup> Pushp Bajaj, Debbie Zhuang,<sup>1</sup> and Brandon Bizzarro<sup>1</sup>**

<sup>1</sup> Department of Chemistry and Biochemistry, University of California, San Diego, USA

E-mail: [fpaesani@ucsd.edu](mailto:fpaesani@ucsd.edu)

Two of the most challenging problems at the intersection of electronic structure theory and molecular dynamics simulations are the accurate representation of intermolecular interactions and the development of reduced-scaling algorithms applicable to large systems. To some extent, these two problems are antithetical, since the accurate calculation of non-covalent interactions typically requires correlated, post-Hartree-Fock methods whose computational scaling with respect to system size precludes the application of these methods to large systems. I will describe our many-body molecular dynamics (MB-MD) methodology<sup>1</sup> for aqueous systems that overcomes these limitations and enables computer simulations from the gas to the condensed phase, with chemical and spectroscopic accuracy. MB-MD is a unified molecular dynamics framework that combines many-body representations for potential energy, dipole moment, and polarizability surfaces that are derived entirely from correlated electronic structure data-driven approaches,<sup>2</sup> with quantum dynamics methods that explicitly account for nuclear quantum effects.<sup>3-7</sup> I will discuss the accuracy and predictive ability of the MB-MD methodology in the context of molecular modeling of complex aqueous systems, from gas-phase clusters to bulk solutions and interfaces, with a particular focus on the relationships between structural and dynamical properties and vibrational spectra.<sup>8-15</sup>

References:

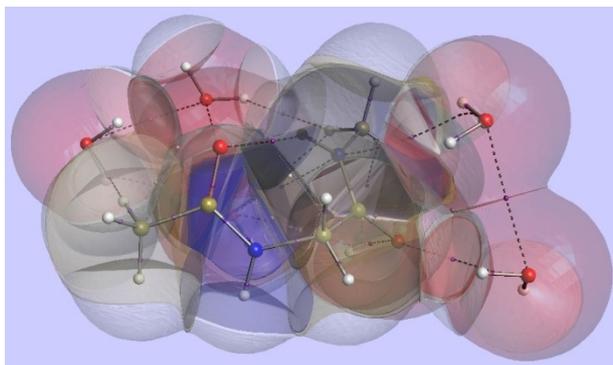
- [1] F. Paesani, *Acc. Chem. Res.* **49** (2016) 1844.
- [2] T.T. Nguyen, E. Szekely, G. Imbalzano, J. Behler, G. Csanyi, M. Ceriotti, F. Paesani, *J. Chem. Phys.* **148** (2018) 241725.
- [3] V. Babin, C. Leforestier, F. Paesani, *J. Chem. Theory Comput.* **9** (2013) 5395.
- [4] V. Babin, G.R. Medders, F. Paesani, *J. Chem. Theory Comput.* **10** (2014) 1599.
- [5] G.R. Medders, V. Babin, F. Paesani, *J. Chem. Theory Comput.* **10** (2014) 2906.
- [6] G.R. Medders, F. Paesani, *J. Chem. Theory Comput.* **11** (2015) 1145.
- [7] G.R. Medders, F. Paesani, *J. Am. Chem. Soc.* **138** (2016) 3912.
- [8] P. Bajaj, A.W. Götz, F. Paesani, *J. Chem. Theory Comput.* **12** (2016) 2698.
- [9] M. Riera, N. Mardirossian, P. Bajaj, A.W. Götz, F. Paesani, *J. Chem. Phys.* **147** (2017) 161715.
- [10] C.K. Egan, F. Paesani, *J. Chem. Theory Comput.* **14** (2018) 1982.
- [11] B.B. Bizzarro, C.K. Egan, F. Paesani, (2019). DOI: 10.26434/chemrxiv.7613042.v2
- [12] P. Bajaj, X. Wang, T. Carrington, Jr., F. Paesani, *JCP* **148** (2018) 102321.
- [13] P. Bajaj, J.O. Richardson, F. Paesani, *Nat. Chem.*, in press (2019).
- [14] M. Riera, S.E. Brown, F. Paesani, *J. Phys. Chem. A* **122** (2018) 5811.
- [15] D. Zhuang, M. Riera, G.K. Schenter, J.L. Fulton, F. Paesani, *J. Phys. Chem. Lett.* **10** (2019) 406.

## FFLUX: on knowledgeable quantum atoms

**Paul Popelier<sup>1</sup>, Matthew Burn<sup>1</sup>, Benjamin Symons<sup>1</sup>, Alex Wilson<sup>1</sup>  
and Zak Hughes<sup>2</sup>**

<sup>1</sup> MIB 131 Princess Street, M1 7DN, and School of Chemistry, Univ. of Manchester, Britain.

<sup>2</sup> School of Chemistry and Biosciences, Univ. of Bradford, Bradford BD7 1DP, Britain.  
E-mail: [pla@manchester.ac.uk](mailto:pla@manchester.ac.uk)



FFLUX<sup>1,2</sup> is a force field under development that ventures beyond next-generation force fields, such as AMOEBA or SIBFA. FFLUX breaks with the tradition of perturbation theory in force field design. Instead, FFLUX invokes Quantum Chemical Topology (QCT)<sup>3,4</sup> to define both intra- and interatomic properties and uses Gaussian Process regression (aka kriging) to capture how these properties change with the nuclear coordinates of the atom's environment. A minimal, consistent and fully integrated framework lies at the heart of FFLUX, covering polarization (e.g.<sup>5</sup>), multipolar electrostatics<sup>6</sup>, dynamic correlation energy<sup>7-9</sup> (i.e. dispersion but also *within* a molecule), exchange energy (i.e. bonding and conjugation effects) and steric repulsion<sup>10</sup>.

After successful tests with geometry optimisation<sup>11, 12</sup> liquid water simulations are being carried out (which will lead to a novel solvation model), adaptive sampling is being explored and oligopeptides will be investigated soon.

### References

1. P. L. A. Popelier, *Int.J.Quant.Chem.* **115**, 1005–1011 (2015).
2. P. L. A. Popelier, *Phys.Scr.* **91**, 033007 (2016).
3. P. L. A. Popelier, "On Quantum Chemical Topology", Eds R. Chauvin, C. Lepetit, E. Alikhani and B. Silvi (Springer, Switzerland, 2016), pp. 23-52.
4. R. F. W. Bader, *Atoms in Molecules. A Quantum Theory.*, Oxford, Britain, 1990.
5. T. L. Fletcher, S. J. Davie and P. Popelier, *J.Chem.Theory Comput.* **10**, 3708 (2014).
6. S. Cardamone, T. J. Hughes and P. L. A. Popelier, *PCCP*, **16**, 10367-10387 (2014).
7. J. McDonagh, A. Silva, M. Vincent and P. Popelier, *JPCL* **8**, 1937 (2017).
8. J. L. McDonagh, M. A. Vincent and P. Popelier, *CPL* **662**, 228-234 (2016).
9. J. McDonagh, A. Silva, M. Vincent and P. Popelier, *JCTC* **14**, 216 (2018).
10. A. L. Wilson and P. L. A. Popelier, *J.Phys.Chem.A* **120**, 9647-9659 (2016).
11. J. C. R. Thacker, A. L. Wilson, Z. E. Hughes, M. J. Burn, P. I. Maxwell and P. L. A. Popelier, *Mol.Simul.* **44**, 881-890 (2018).
12. Z. E. Hughes, J. Thacker, A. Wilson and P. Popelier, *JCTC* **15**, 116 (2019).

# Exact representations of molecules and materials for accurate interpolation of ab initio simulations

Marcel Langer,<sup>1</sup> Alex Goeßmann,<sup>1,2</sup> and Matthias Rupp<sup>1,3</sup>

<sup>1</sup> Fritz Haber Institute of the Max Planck Society, Berlin, Germany

<sup>2</sup> Technische Universität Berlin, Germany

<sup>3</sup> Citrine Informatics, Redwood City, California, USA

E-mail: [mrupp@citrine.io](mailto:mrupp@citrine.io)

Computational study of molecules and materials from first principles is a cornerstone of physics, chemistry and materials science, but limited by the cost of accurate quantum-mechanical simulations. In settings involving many simulations, machine learning can reduce these costs significantly by rapidly and accurately interpolating between reference simulations. This requires a single Hilbert space, or *representation*, that accommodates any molecule or material and supports interpolation.

We review, discuss and benchmark state-of-the-art representations and relations between them, [1] including smooth overlap of atomic positions, many-body tensor representation and symmetry functions. [2] For this, we use a unified mathematical framework [3] based on many-body functions, group averaging and tensor products, and empirically compare predictive accuracy in controlled numerical experiments on datasets of organic molecules and binary alloys. In our benchmark, we control for (i) data distribution by multivariate stratification, and (ii) regression method by employing the same machine learning approach for all representations, including (iii) hyperparameter optimization. For the latter we use a consistent and fully automated procedure to optimize both numerical and architectural free parameters based on sequential model-based optimization with tree-structured Parzen estimators. [4] Our findings hint at current state-of-the-art representations implementing, based on the same principles, different trade-offs between computational complexity and modeled interaction orders.

References:

[1] M. Langer, A. Goeßmann, M. Rupp, *in preparation* (2019).

[2] a) A.P. Bartók, R. Kondor, G. Csányi, *Phys Rev B* 87 (2013) 184115; b) H. Huo, M. Rupp, *arXiv* (2017) 1704.06439v3; c) J. Behler, *J Chem Phys* 134 (2011) 074106.

[3] M.J. Willatt, F. Musil, M. Ceriotti, *arXiv* (2018) 1807.00408.

[4] a) J.S. Bergstra et al., *NIPS* 24 (2011) 2546. b) J.S. Bergstra et al., *ICML* 30 (2013) I-115.

## List of Participants

Allen, Alice; University of Cambridge, United Kingdom  
Batzner, Simon Lutz; Harvard University Cambridge, Massachusetts, USA  
Behler, Jörg; Georg-August-Universität Göttingen, Germany  
Bernstein, Noam; Naval Research Lab, Washington, D.C., USA  
Bertaina, Gianluca, University of Milano, Italy  
Beseda, Martin; Ostrava University, Czech Republic  
Bowman, Joel; Emory University, Georgia, USA  
Braams, Bastiaan; Centrum Wiskunde & Informatica (CWI) Amsterdam, Netherlands  
Burghardt, Irene; Universität Frankfurt, Germany  
Burke, Kieron; University of California Irvine, California, USA  
Caro, Miguel A.; Aalto University, Finland  
Cheng, Bingqing; University of Cambridge, United Kingdom  
Christiansen, Mads-Peter Verner; Aarhus University, Denmark  
Csányi, Gábor, University of Cambridge, United Kingdom  
Daelman, Nathan; Barcelona Institute of Science and Technology, Spain  
Deringer, Volker; University of Cambridge, United Kingdom  
Dusson, Geneviève; University of Warwick, United Kingdom  
Eckhoff, Marco; Georg-August-Universität Göttingen, Germany  
Egan, Colin; University of California San Diego, California, USA  
Ehlert, Sebastian; Universität Bonn, Germany  
Fekete, Adam; University of Cambridge, United Kingdom  
Galindo, Valentin Vassilev; University of Luxembourg, Luxembourg  
Glensk, Albert; EPFL Lausanne, Switzerland  
Goedecker, Stefan; University of Basel, Switzerland  
Grimme, Stefan; Universität Bonn, Germany  
Guo, Hua; University of New Mexico, USA  
Hellström, Matti; SCM Amsterdam, Netherlands  
Herbold; Marius; Georg-August-Universität Göttingen, Germany  
Järvi, Jari; Aalto University, Finland  
Kandratsenka, Alexander, MPI Biophysikalische Chemie Göttingen, Germany  
Kermode, James; University of Warwick, United Kingdom  
Knoll; Alexander; Georg-August-Universität Göttingen, Germany  
Ko, Tsz Wai; Georg-August-Universität Göttingen, Germany  
Koch, Werner; Universität Frankfurt, Germany  
Krems, Roman; University of British Columbia, Canada  
Langer, Marcel; Fritz-Haber-Institut Berlin, Germany  
Li, Jun; MPI Biophysikalische Chemie Göttingen, Germany  
Liebtrau, Martin; Georg-August-Universität Göttingen, Germany  
Long, Teng; Technische Universität Darmstadt, Germany  
Marquetand, Philipp; University of Vienna, Austria  
Mata, Ricardo; Georg-August-Universität Göttingen, Germany  
Meuwly, Markus; University of Basel, Switzerland  
Musil, Felix; EPFL Lausanne, Switzerland  
Nandi, Apurba; Emory University, Georgia, USA  
Oren, Eyal; Negev University, Israel  
Ortner, Christoph; University of Warwick, United Kingdom  
Oswald, Rainer; Georg-August-Universität Göttingen, Germany  
Paesani, Francesco; University of California San Diego, California, USA  
Paleico, Martin; Georg-August-Universität Göttingen, Germany  
Parsaeifard, Behnam; University of Basel, Switzerland  
Popelier, Paul; University of Manchester, United Kingdom  
Qu, Chen; NIST Gaithersburg, Maryland, USA

Riera, Marc; University of California San Diego, California, USA  
Rupp, Matthias; Fritz-Haber-Institut Berlin, Germany  
Shanavas Rasheeda, Dilshana; Georg-August-Universität Göttingen, Germany  
Szekely, Eszter; University of Cambridge, United Kingdom  
Tkatchenko Alexandre; University of Luxembourg, Luxembourg  
Todorovic, Milica; Aalto University, Finland  
van der Oord, Cas; University of Cambridge, United Kingdom  
Weinreich, Jan; Georg-August-Universität Göttingen, Germany  
Wilkins, David; EPFL Lausanne, Switzerland  
Wille, Sebastian; MPI Biophysikalische Chemie Göttingen, Germany

## **Acknowledgements**

We gratefully acknowledge support by

Psi-K  
ICASEC  
Georg-August-Universität Göttingen  
Max-Planck-Institut für biophysikalische Chemie  
Fonds der Chemischen Industrie e.V.  
The Journal of Chemical Physics