Philipp Schienbein, Jörg Behler

Atomistic simulations with machine learning: From dynamics to vibrational spectroscopy

Machine learning potentials

Machine learning is revolutionizing all fields of science, drastically accelerating research from high-throughput screening and materials design via laboratory automation to planning and analysis of experiments. Similarly, a paradigm change is taking place in the field of atomistic simulations by the increasing use of machine learning potentials (MLP), which combine the accuracy of electronic structure calculations, such as density functional theory (DFT), with the efficiency of classical force fields [1, 2, 3]. These advances nowadays allow to perform molecular dynamics (MD) simulations of systems with constantly increasing complexity going far beyond the size accessible by direct *ab initio* MD.

The core idea of MLPs is to replace demanding electronic structure calculations by trained machine learning (ML) models of the potential energy surface. Still, for a reliable description of many systems a combination of ML and physical concepts has turned out to be crucial. Consequently, many different types of MLPs have been proposed in recent years, and the underlying architectures may be classified into different generations to specify the physical interactions they are able to describe and their applicability to different types of systems [4]. The first generation of MLPs has been introduced already three decades ago, but their restriction to very small systems has strongly limited their applicability. Large condensed systems containing thousands of atoms can be studied with second-generation MLPs, such as second-generation high-dimensional neural network potentials (2G-HDNNP, see Fig. 1), which are based on the locality of a major part of the atomic interactions [5]. Long-range interactions beyond the local atomic environments are included in third-generation MLPs, most importantly electrostatics and dispersion. Finally, global phenomena like long-range charge transfer are accessible in fourth-generation MLPs [6].

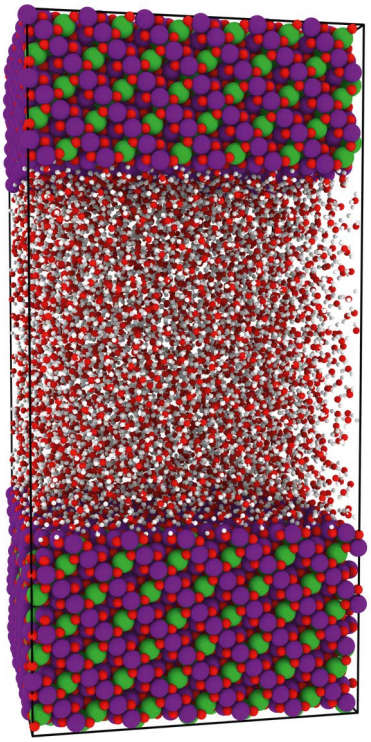


Fig. 1: Typical system size for the simulation of liquid water at the MnO_2-terminated (001) surface of the Li intercalation compound $LiMn_2O_4$ using a second-generation high-dimensional neural network potential [7].

Machine learning for vibrational spectroscopy

Modern MLPs can reproduce reference energies and forces with remarkable accuracy, extending electronic-structure simulations to time and length scales that would otherwise be inaccessible. A natural next step is to learn complementary electronic quantities that connect trajectories to experimental observables. A particularly attractive target is vibrational spectroscopy. Peak positions and intensities reflect the distribution of molecular configurations, while line shapes also contain information on fluctuations and relaxation dynamics. Quantitative agreement with experimental spectra can therefore build confidence in a simulation and thus supports exploring quantities that are not directly accessible experimentally. The ex-

Dr. Philipp Schienbein (main author)
Lehrstuhl für Theoretische Chemie II
Ruhr-Universität Bochum
D-44780 Bochum
Research Center Chemical Sciences and Sustainability
Research Alliance Ruhr
D-44780 Bochum
philipp.schienbein@ruhr-uni-bochum.de
https://schienbein.eu

Prof. Dr. Jörg Behler
Lehrstuhl für Theoretische Chemie II
Ruhr-Universität Bochum
D-44780 Bochum
Research Center Chemical Sciences and Sustainability
Research Alliance Ruhr
D-44780 Bochum
joerg.behler@ruhr-uni-bochum.de
https://www.theochem2.ruhr-uni-bochum.de/

DOI-Nr.: 10.26125/ahhp-k814

change between experiment and theory is equally valuable in the opposite direction: low-frequency (THz), multidimensional, and time-dependent spectra often contain overlapping, correlated contributions and can be difficult to interpret atomistically. Here, simulations can disentangle these contributions, assign spectral features to molecular motions, and distinguish between competing mechanisms [8].

For IR spectroscopy, the central electronic property is the dipole moment. For finite systems such as gas-phase clusters, IR spectra have been calculated by learning the total dipole moment [9]. In periodic systems, the corresponding polarization is multivalued. This challenge can be addressed using ML representations of atomic polar tensors (APTs) that rigorously assign the dipole response to individual atoms without the ambiguity of an artificial charge partitioning scheme [10, 11]. Raman spectroscopy can be treated similarly using polarizability gradient tensors (PGTs) [11]. These models require comparatively few reference calculations and reproduce explicit *ab initio* spectra with high accuracy. For typical system sizes, calculations that would require months of direct *ab initio* MD can be completed within hours on a standard workstation. Most importantly, the atom resolved response tensors enable a rigorous, quantitative assignment of spectral features. The THz spectrum of liquid water provides an illustrative example: the calculated spectrum can be decomposed into additive contributions from specific atoms and different kinds of motion, revealing directly how H-bond vibrations and water librations combine to produce the overall signal (Fig. 2).

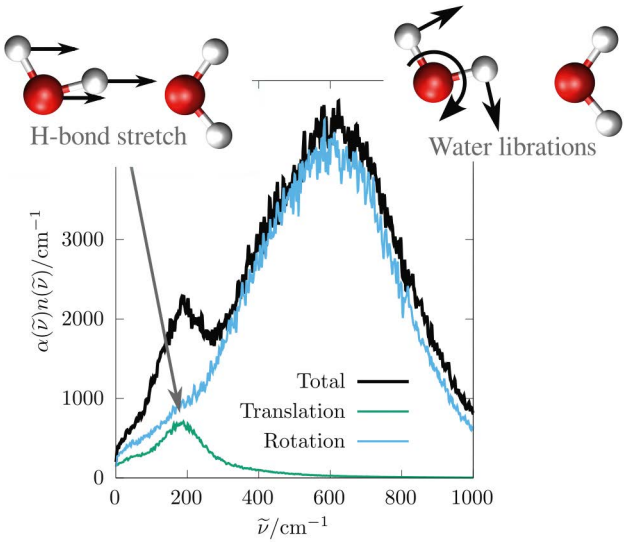


Fig. 2: Direct microscopic assignment of the THz spectrum of liquid water. Learned atom-resolved response tensors (APTs) allow the calculated total spectrum (black) to be decomposed into additive translational (green) and rotational (blue) contributions. The feature near 200 cm^{-1} is dominated by collective H-bond stretching, whereas the broad band around 600 cm^{-1} arises mainly from librations.

Driving molecular dynamics with pulsed electric fields

The learned APTs have a second purpose: they rigorously provide the first-order coupling between the atoms and an external electric field by delivering the field-induced forces. Static and time-dependent electric fields can therefore be incorporated into machine learning MD simulations [12-14]. Applying static fields, the field strength can be used as a thermodynamic control parameter, providing access to field-dependent transport properties, such as ionic conductivities in aqueous solution [13]. Time-dependent fields allow the electric-field component of a light pulse to be modeled explicitly [14, 15]. The system's response can then be followed with atomistic resolution during and after the pulse. This opens a route towards simulating pump-probe and multidimensional spectroscopies, with the pulse frequency, duration, polarization, and intensity chosen to match the experiment as closely as possible.

A particularly interesting case was provided by experiments in which liquid water was excited with an intense 12.3 THz pump pulse [16]. The interpretation of this measurement sparked a debate about if the observed response reflected a genuine

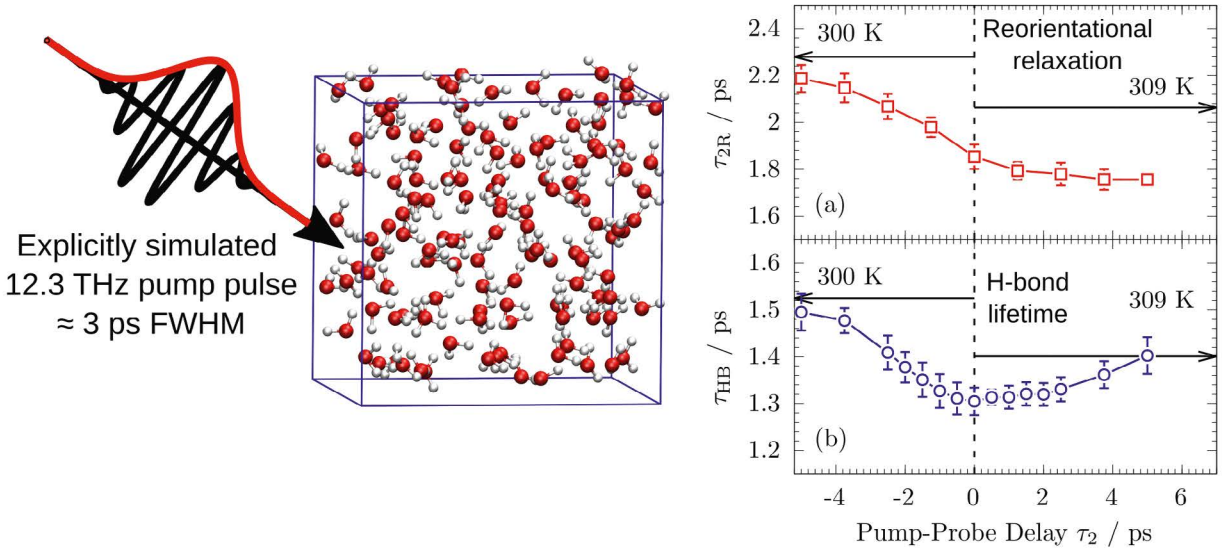


Fig. 3: Example of an explicitly simulated THz pump-probe experiment. A 12.3 THz pulse excites liquid water, while molecular reorientation (top) and H-bond lifetime (bottom) are monitored using machine learning MD simulations. Their different relaxation behaviors reveal a transient response beyond simple heating. The dashed vertical line marks the pulse maximum in time. Adapted from Ref. [14].

non-equilibrium response of water dynamics or if it is caused predominantly by thermal effects [17, 18]. The molecular response was therefore simulated directly by explicitly applying a closely matching THz pulse to liquid water in machine learning MD simulations [14]. Even with ML acceleration, obtaining statistically converged results required 1,600 independent non-equilibrium trajectories of 20 ps each – effectively repeating the pump pulse 1600 times – generating a total of 32 ns of trajectory data at *ab initio* accuracy. The simulations show that the energy is initially deposited into molecular librations, before being redistributed into translational and intramolecular vibrations. Simultaneously, rotational relaxation and H-bond dynamics evolve differently in time: the former does not relax back to equilibrium in the simulated time window, while the H-bond lifetime returns to equilibrium at about 5 ps after the pulse maximum (Fig. 3). The effect of the pulse therefore extends beyond simple heating and confirms the initial interpretation of a transient perturbation of the H-bond network.

In conclusion, modern ML-based methods allow experiment and simulations to advance hand in hand: measurements provide stringent tests of the simulations, while atomistic trajectories reveal the molecular motions behind the observed signals. This partnership holds a particular promise for multidimensional spectroscopies whose signals encode a wealth of structural and dynamical information, but are exceptionally difficult to interpret. By bringing *ab initio* accuracy to the time scales and complexity of these experiments, ML allows theory to close this gap – and to uncover molecular dynamics that measurements alone cannot reveal directly.

References

[1] J. Behler, *J. Chem. Phys.,* 2016, **145**, 170901.

[2] P. Friederich, F. Häse, J. Proppe, A. Aspuru-Guzik, *Nat. Mater.,* 2021, **20**, 750.

[3] S. Käser, L. I. Vazquez-Salazar, M. Meuwly, K. Töpfer, *Digital Discovery,* 2023, **2**, 28.

[4] J. Behler, *Chem. Rev.,* 2021, **121**, 10037.

[5] J. Behler and M. Parrinello, *Phys. Rev. Lett.,* 2007, **98**, 146401.

[6] T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, *Nat. Commun.,* 2021, **12**, 398.

[7] K. N. Lausch, R. El Haouari, P. Schienbein, and J. Behler. *Angew. Chem. Int. Ed.* **accepted** (DOI: https://doi.org/10.1002/anie.3712910)

[8] P. Schienbein and D. Marx. *Chem. Rev.,* 2026, **126**, 7465.

[9] R. Beckmann, F. Brieuc, C. Schran, and D. Marx, *J. Chem. Theory Comput.,* 2022, **18**, 5492.

[10] P. Schienbein. *J. Chem. Theory Comput.,* 2023, **19**, 705.

[11] P. Schienbein. *J. Chem. Theory Comput.,* 2026, **22**, 4626.

[12] K. Joll, P. Schienbein, K. M. Rosso, and J. Blumberger. *Nat. Commun.,* 2024, **15**, 8192.

[13] K. Joll, P. Schienbein, K. M. Rosso, and J. Blumberger. *J. Phys. Chem. B,* 2026, **130**, 2817.

[14] K. Joll and P. Schienbein. *J. Phys. Chem. Lett.*, 2025, **16**, 9183.

[15] E. Stocco, C. Carbogno, and M. Rossi, *npj Comput. Mater.,* 2025, **11**, 304.

[16] F. Novelli, C. Hoberg, E. M. Adams, J. M. Klopf, and M. Havenith, *Phys. Chem. Chem. Phys.,* 2022, **24**, 653.

[17] A. F. G. van der Meer, *Phys. Chem. Chem. Phys.,* 2022, **24**, 13411.

[18] F. Novelli, C. Hoberg, E. M. Adams, J. M. Klopf, and M. Havenith, *Phys. Chem. Chem. Phys.,* 2022, **24**, 13413.

Dr. Philipp Schienbein



©RUB, Marquard

Philipp Schienbein studied chemistry at Ruhr-Universität Bochum, where he completed his PhD in Theoretical Chemistry in 2019. He subsequently held postdoctoral positions at University College London and Imperial College London, supported by fellowships from the German National Academy of Sciences Leopoldina and the German Research Foundation (DFG). In 2024, he established an independent research group at Ruhr-Universität Bochum. His research combines *ab initio* molecular dynamics, vibrational spectroscopy, and machine learning to connect vibrational spectra to molecular structural dynamics in condensed systems.

Prof. Dr. Jörg Behler



©RUB, Marquard

Jörg Behler studied chemistry in Dortmund and obtained his PhD in 2004 at the Fritz-Haber-Institut in Berlin. After a postdoctoral stay at ETH Zürich, in 2007 he established an independent research group at Ruhr-Universität Bochum funded by a Liebig fellowship, the Emmy Noether program and a Heisenberg fellowship. In 2017, he became W3 professor for theoretical chemistry in Göttingen. In 2022, he joined the Research Center Chemical Sciences and Sustainability of the Research Alliance Ruhr and founded the new Chair of Theoretical Chemistry II at Ruhr-Universität Bochum. His main research interest is the development of machine learning potentials with applications in chemistry and materials science.