

2023-12-22



D5.2 – Report on machine learning results delivered for water systems

Version 1.0

GA no 952165

Dissemination Level

- PU: Public
- PP: Restricted to other programme participants (including the Commission)
- RE: Restricted to a group specified by the consortium (including the Commission)
- CO: Confidential, only for members of the consortium (including the Commission)

Document Information

Project Title	Targeting Real Chemical Accuracy at the EXascale
Project Acronym	TREX
Grant Agreement No	952165
Instrument	Call: H2020-INFRAEDI-2019-1
Topic	INFRAEDI-05-2020 Centres of Excellence in EXascale computing
Start Date of Project	01-10-2020
Duration of Project	36 months originally, extended to 42 months
Project Website	https://trex-coe.eu/
Deliverable Number	D5.2
Deliverable title	D5.2 – Report on machine learning results delivered for water systems, as seen in GA
Due Date	M38 – 2023-11-30; extended to 2023-12-22 (from GA)
Actual Submission Date	2023-12-22
Work Package	WP5 – Demonstrations
Lead Author (Org)	Matthias Rupp (LIST)
Contributing Author(s) (Org)	Michele Casula (CNRS)
Reviewers (Org)	Kasia Pernal (TUL), Ivan Stich (SAV)
Version	1.0
Dissemination level	PU
Nature	Report
Draft / final	Final
No. of pages including cover	vi (front matter), 10 (main text), II (back matter)



Disclaimer



TREX: Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizon 2020 research and innovation program under Grant Agreement No. 952165.

The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.



Versioning

Version	Date	Authors	Notes
1.0	2023-12-22	Matthias Rupp (LIST)	First Official Release



Abbreviations

CoE	Centre of Excellence
DFT	Density Functional Theory
EU	European Union
KRR	Kernel Ridge Regression
MD	Molecular Dynamics
ML	Machine Learning
MLP	Machine Learning Potential
PIMD	Path-Integral Molecular Dynamics
PIOD	Path-Integral Ornstein–Uhlenbeck Dynamics
PT	Proton Transfer
QMC	Quantum Monte Carlo
TREX	Targeting REal chemical accuracy at the eXascale
SOAP	Smooth Overlap of Atomic Positions
WP	Work Package



Table of Contents

Document Information	i
Disclaimer	ii
Versioning	iii
Abbreviations	iv
Table of Contents	v
List of Figures	vi
<i>Executive summary</i>	1
1 Introduction	2
Scope	2
2 Water	3
Outputs	5
3 Hydrogen	5
Outputs	6
4 Machine-learning potentials	7
Outputs	8
5 Summary and outcomes	8
References	I



List of Figures

1	Different configurations of the protonated water hexamer	3
2	Radial distribution functions in classical simulations of the protonated water hexamer .	4
3	Structures considered for the low-temperature high-pressure phase diagram of hydrogen	6
4	Learning rate of Δ -learning MLP for hydrogen	8



Executive summary

The TREX EU Centre of Excellence investigates implementations of Quantum Monte Carlo (QMC) calculations optimized for exascale high-performance computing. These calculations are high-accuracy quantum-chemical and materials simulations that are inherently parallelizable and computationally demanding. Thus, they are uniquely positioned to utilize and explore the upcoming exascale supercomputer architectures. TREX focuses on the development and promotion of an open-source, high-performance software platform of inter-operable flagship codes and exascale-ready libraries.

This scope includes, in work package 5, applications of these QMC methods to atomistic systems that are highly and directly relevant for technological progress and society. One of these systems is water, the “liquid of life.” In addition, TREX investigates Machine Learning Potentials (MLPs) to greatly accelerate QMC dynamics simulations in work package 4, enabling running more and longer simulations with larger unit cells, a task that will remain computationally unfeasible using only QMC calculations for the foreseeable future.

This Periodic Activity Report D5.2 centers on results obtained via MLPs for water. Because the study of water H_2O is strongly linked to the physics and chemistry of hydrogen H, results for hydrogen under pressure are included as well. This report is related to deliverables D4.4, D5.3, and D5.4.

Michele Casula’s group (CNRS) investigated hydrogen’s role in hydrogen bonds of water by exploring electronic properties affecting bond dynamics. Their study on protonated water hexamers using QMC methods revealed temperature-dependent proton behavior. To extend findings, they developed an MLP for water clusters. Ongoing work focuses on improving agreement and studying QMC noise effects on MLP quality and long-range interactions’ impact on charged systems.

They also studied the phase diagrams of hydrogen (H) and hydrogen-rich materials due to H’s relationship with water and high-temperature superconductivity found in H-rich materials. These phase diagrams are very rich, with many competing phases. Resolving them is highly challenging and requires coupling QMC calculations for electrons with path-integral molecular dynamics or path-integral Monte Carlo for quantum nuclei. Lower levels of theory cannot predict these phase diagrams. One of H’s most accurate phase diagrams was calculated with the TREX code TurboRVB within work package 5.

The usual strategy of training an MLP directly on QMC reference data fails as it is computationally too expensive to generate enough QMC training data. Instead, QMC corrections to a computationally cheaper physical baseline method, such as Density Functional Theory (DFT), were employed. This “ Δ -learning” approach requires less QMC training data. The group of Sandro Sorella (SISSA) developed a Δ -learning MLP, enabling them to train an accurate model using only 684 QMC calculations. They used this model to study the liquid-liquid phase transition of high-pressure hydrogen.

To enable further studies, the groups of Matthias Rupp (UKON, LIST) and Michele Casula (CNRS) collaborate to determine whether so-called “ultra-fast potentials” trained on DFT reference data can be used as baseline potential for the Δ -learning approach. This would enable a computational speed-up by several orders of magnitude, paving the way to a more extended and comprehensive study of the phase diagrams of H and H-rich materials. Further efforts were made towards improved workflows for training set generation.

Overall, nine scientific studies were published that acknowledge TREX funding, in journals including Nature Physics and Nature Communications.



1 Introduction

Targeting REal chemical accuracy at the eXascale (**TREX**) is a European Union (EU) Centre of Excellence (CoE) in the field of high-accuracy quantum chemical and materials simulations with a special focus on QMC approaches to the solution of the quantum many-body problem at the heart of atomistic physics, chemistry and materials science. Due to their inherent parallelizability and high computational cost, QMC approaches, and thus TREX, are uniquely positioned to fully exploit the massive parallelism of the upcoming exascale supercomputer architectures. Work in TREX focuses on the development and promotion of an open-source, high-performance software platform of interoperable flagship codes and exascale-ready libraries in its area of applications.

These **applications** include the computational investigation of atomistic systems that are highly and directly relevant for technological progress and society. Work Package (WP) 5 tests and demonstrates the effectiveness of the QMC software developed in TREX on a selection of such systems relevant for energy storage, conversion, and transport. Specifically, these are two-dimensional materials for energy conversion in future opto-electronic systems, water as the basis of all organic life on our planet, quantum magnetism and high-temperature superconductors, as well as strained and excited functional van-der-Waals materials beyond graphene.

In addition, TREX investigates the applicability of **Machine Learning (ML)** methods to accelerate QMC calculations in WP 4. This was done primarily by constructing MLPs, which are data-driven surrogate models of QMC potential energy surfaces. Trained on a set of reference QMC calculations for a specific atomistic system, MLPs accurately approximate these systems' potential energy surfaces at a fraction of the computational cost. The resulting acceleration by multiple orders of magnitude greatly extends the reach of QMC approaches. TREX focuses on MLP-accelerated Molecular Dynamics (MD) simulations. For these, MLPs enable running more and longer simulations with larger unit cells, a task that will remain computationally unfeasible using QMC calculations alone for the foreseeable future.

Scope

This Periodic Activity Report D5.2 centers on results obtained via MLPs for water and hydrogen under pressure. The reported work is linked to Task T4.3 “Workflows to machine learn QMC accuracy” (which was led by Anatole von Lilienfeld of UNIVIE, who left the consortium as of 2021-10-13, and is now lead by Matthias Rupp of UKON, and then LIST, who joined the consortium as of 2021-10-01 and changed affiliation from UKON to LIST as of 2022-10-01). Report D5.2 is also affected by the untimely death of Prof. Dr. Sandro Sorella and the subsequent termination of SISSA as consortium member as of 2023-09-30.

This report D5.2 is also related to deliverables D4.4 “Report on release of transferable QMC-quality ML models” in WP 4 via the developed MLPs, D5.3 “Final report on all demonstrations” in WP 5 through task T5.2 on the “accurate and reliable description of water” and task T5.3 on high-temperature superconductors, as well as D5.4 on “Datasets made available for benchmarking and ML modelling” in WP 5 via the datasets created in the work reported here.

As the study of water H₂O is intrinsically strongly linked to the physics and chemistry of hydrogen (H), we also present application results for hydrogen in this report.

A list of TREX publications related to this report is available in Section 5.



2 Water

One of the activities of the group of Michele Casula (CNRS-Paris), in the realm of the WP5 TREX work package, is to study water by using advanced QMC methods, optimized and ported to HPC clusters, as demonstrator of QMC capabilities. **Water** has been hailed as the “liquid of life”, whose unique characteristics stem from the chemical properties of the hydrogen bond (H-bond). In this respect, there is a tight link between this application and our investigation of hydrogen presented in Section 3. The role played by hydrogen in the H-bond is fundamentally affected by electronic properties, such as charge transfer, electron polarizability, dispersive interactions, together with nuclear quantum effects (NQE), which profoundly modify the classical potential energy landscape.

Our recent study [1] of the **protonated water hexamer** focused on the temperature dependence of the hydrated proton and its dynamics, including Proton Transfer (PT). The structure of the cluster is shown in Fig. 1, where it is clear how the two limiting structures of the PT dynamics, i.e., the Zundel (leftmost panels of Fig. 1) and the Eigen configurations (rightmost panel of Fig. 1), are fully included in the possible cluster geometries. The protonated water hexamer is the smallest water cluster including both limiting structures, being the central core solvated by a complete water shell. Thus, the hydrated proton in the middle of the cluster is not far from the bulk water limit, and one can still afford accurate but expensive QMC simulations to study the PT dynamics.

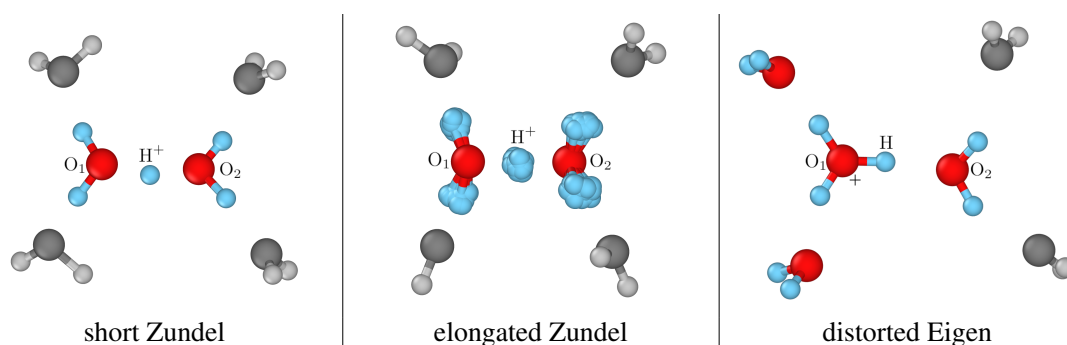


Figure 1: *Different configurations of the protonated water hexamer during molecular dynamics simulations. Short Zundel is the most likely configuration at low temperature (< 150 K), elongated Zundel is the one at intermediate temperatures until room temperature, and distorted Eigen is a high-temperature configuration. Labels O_1 , O_2 , H^+ are also used in Fig. 2.*

This led to the discovery of a remarkably small thermal expansion of the cluster core, and of an optimal temperature range for the PT, located in the 250–300 K temperature window, where the PT rate shows a maximum. These results have been obtained by using a combination of variational QMC to compute the electronic wave function and ionic forces on the fly, during a Path-Integral Molecular Dynamics (PIMD) simulation driven by the Path-Integral Ornstein–Uhlenbeck Dynamics (PIOUD) algorithm[2], able to cope with the QMC noise, and to exploit it for system thermalization, as implemented in the **TREX code** TurboRVB. Such coupling of QMC with PIMD is a computational challenge, achieved here thanks to the relatively small size of the system. Thus, to extend these findings to larger clusters and to bulk water, we would like to train an efficient MLP on the trajectories generated during this study.

Therefore, we developed an **MLP for water clusters** that combines the accuracy of QMC calculations with the computational efficiency of ML. We achieved this by using the FCHL19 [3, 4] representation and Kernel Ridge Regression (KRR). FCHL19 is a computationally efficient and suitable representation for operator kernel regression used by us for force learning. Our aim is to use an MLP trained directly on the QMC forces, which enables the MLP to enter into the MD schemes as a force field, in place of heavy *ab initio* calculations. This approach bypasses any other intermediate DFT steps, and it exploits “operator quantum machine learning” [4] capabilities.

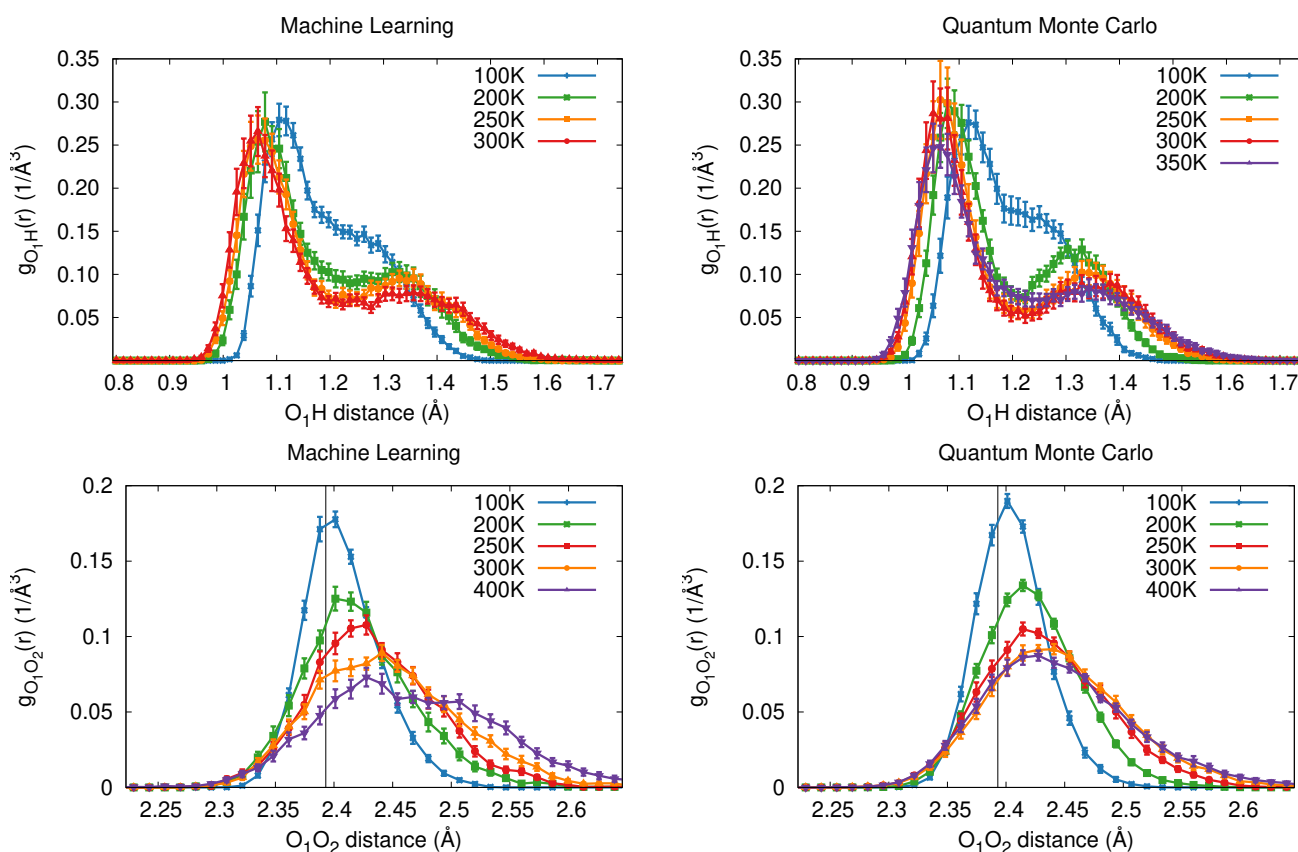


Figure 2: Radial distribution functions in classical simulations of the protonated water hexamer. OH^+ is the distance between the central proton and the neighboring oxygen atoms, while O_1O_2 is the distance between the two inner water molecules bridged by the central proton. The vertical lines in the bottom panels show the equilibrium $d_{\text{O}_1\text{O}_2}$ distance at zero temperature.

Preliminary **results**, based on the trajectories generated during our protonated water hexamer study, are satisfactory (Figure 2). The MD simulations run using the MLP are stable and produce pair distribution functions in reasonable agreement with the original QMC-driven MD. Some work is still needed to improve the agreement below a tolerance of 50 K. The latter is desirable in order to properly describe proton transfer processes in aqueous environments and to extend this analysis to larger water clusters. We presently study the effects of QMC noise on the MLP quality and the impact of including long-range interactions on top of the MLP to correctly describe charged systems, such as protonated clusters.

Outputs

- Félix Mouhat, Matteo Peria, Tommaso Morresi, Rodolphe Vuilleumier, Antonino Marco Saitta, Michele Casula: *Thermal dependence of the hydrated proton and optimal proton transfer in the protonated water hexamer*, Nature Communications 14: 6930, 2023. [\[doi\]](#)

3 Hydrogen

Hydrogen and hydrogen-rich materials have seen increasing interest in the last decade due to the discovery of high-temperature superconductivity found first in H_3S , and then in LaH_{10} , showing critical temperatures T_c as high as 250K, i.e., close to the room-temperature superconductivity dream. These T_c values are larger than previous records measured in strongly correlated materials, such as cuprates and pnictides, where superconductivity is unconventional. Superconductors in the hydrogen class are instead driven by a strong electron-phonon coupling linked with large phonon vibrations and a metallic behavior of the material. These conditions are usually met at a very high pressure. Superconductivity there is thus mediated by lattice vibrations. Efforts are currently made to reduce the pressure at which superconductivity appears to make this remarkable phenomenon technologically exploitable by exploring the chemistry of hydrogen.

The resulting **phase diagram** of hydrogen-based materials is usually very rich, with many competing phases (Figure 3). This is due to the interplay of electronic and vibrational effects. Both electrons and nuclei must be accounted for to predict the various phases emerging at different pressure and temperature conditions. Resolving these phase diagrams is useful to detect the presence of possible high-temperature superconducting phases, but it also has other applications. One example is determining the equations of state of gas planets, usually made of a mixture of hydrogen with other light elements such as helium. Another example is the field of energy storage materials, which is of critical importance for a green economy and a much coveted green transition.

Including both electrons and nuclei at the high **accuracy** needed to resolve the phase boundaries of hydrogen materials is a very challenging problem. While QMC is beneficial to predicting accurate internal electronic energies, coupling QMC calculations for electrons with path-integral molecular dynamics or path-integral Monte Carlo for quantum nuclei is out of reach with the present computational capabilities. Thus, studying the most relevant hydrogen-rich systems is a formidable task. It is only very recently that one of the most accurate phase diagrams of hydrogen has been computed in the high-pressure region (400 GPa–700 GPa), where hydrogen is supposed to become a room-temperature superconductor, thanks to extensive QMC calculations performed with the **TREX code** TurboRVB within WP5. [5] In the latter calculations, the nuclear part has been computed with the self-consistent harmonic approximation, which is computationally cheaper but less accurate than path-integral methods. A one-to-one comparison between the self-consistent harmonic approximation and the PIMD has been recently done in Ref. [6], for the study of the H_3S high-temperature superconductor.

Computationally more efficient but less accurate theories than QMC, such as DFT with simple semi-local functionals, but also more elaborated and more expensive hybrid functionals, failed to predict and explain these rich and complex phase diagrams. As elaborated in the Demonstrations (WP5), we envisage that QMC at the exascale will open new possibilities for fundamental studies of hydrogen and hydrogen-rich materials. However, the overall **computational cost** would remain a severe bottleneck, even for the most efficient and well ported algorithms. Thus, the acceleration provided by MLPs is even more important for these classes of materials.



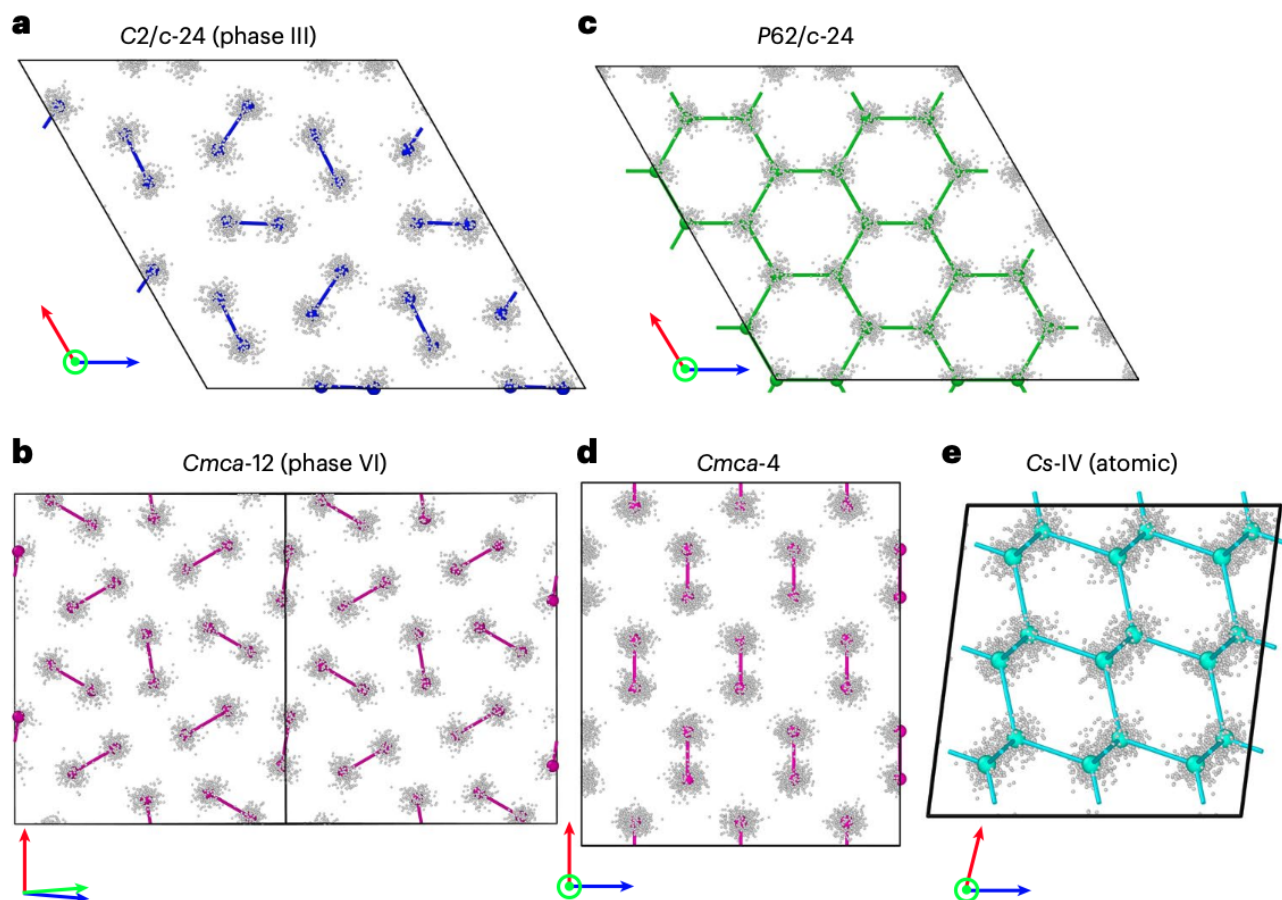


Figure 3: Structures considered for the low-temperature high-pressure phase diagram of hydrogen. Coloured balls are the average centroid positions, sticks represent H_2 molecules and the cloud of smaller grey balls is a set of 250 configurations that sample the quantum probability distribution at 0 K. All the structures, apart from the atomic one, are made of layers, out of which only one is shown. $P62/c-24$ is made of alternating layers, one with atoms arranged in a honeycomb lattice (b), and the other with molecular H_2 in a $C2/c-24$ arrangement (not shown here).

Outputs

- Lorenzo Monacelli, Michele Casula, Kousuke Nakano, Sandro Sorella, Francesco Mauri: *Quantum phase diagram of high-pressure hydrogen*, Nature Physics 19(6): 845, 2023. [\[doi\]](#)
- Romain Taureau, Marco Cherubini, Tommaso Morresi, Michele Casula: *Quantum symmetrization transition in superconducting sulfur hydride from quantum Monte Carlo and path integral molecular dynamics*, arXiv: 2307.15684, 2023. [\[url\]](#)

4 Machine-learning potentials

To address the problem of the high computational cost of QMC calculations and thus to enable efficient MD simulations at larger length and time scales but with QMC accuracy, MLPs based on QMC reference calculations were developed in TREX. Due to the high computational cost of QMC calculations, the usual strategy of training an MLP directly on QMC reference data fails as it is too expensive to generate enough QMC training data. Instead, QMC corrections to a computationally cheaper physical baseline method, such as DFT, are employed. This approach is called “ Δ -learning” [7] and has the advantage that learning the correction is an easier learning problem, and thus requires less QMC training data.

The group of Sandro Sorella (SISSA) developed such a Δ -learning MLP that provides the accuracy of QMC calculations at the cost of DFT calculations. Specifically, they used kernel least-squares regression with Smooth Overlap of Atomic Positions (SOAP) features to learn the difference between DFT and QMC, as well as sparsification via farthest-point sampling. [8] The effect of sparsification is shown in Figure 4. Together, this enabled them to train an accurate model using only 684 QMC calculations of 128-atom configurations of H as reference in total.

As an application we present a benchmark study of the liquid-liquid transition of high-pressure hydrogen and show the quality of our MLP, by emphasizing the importance of high accuracy for this very debated subject, where experiments are difficult in the laboratory, and theory is still far from being conclusive.

As a first application, they present a benchmark study of the liquid-liquid transition of high-pressure hydrogen. These results have been published in Ref. [8]. Their reported MLP runs essentially at the speed of the underlying DFT baseline method since the computational cost of the MLP is negligible in comparison. While at least one order of magnitude faster than the QMC reference method, this is still the same computational effort that most current MLPs based on DFT reference data try to avoid. Matthias Rupp’s and Michele Casula’s groups are collaborating, together with a PhD student at SISSA formerly supervised by Sandro Sorella, with the aim of investigating if the so-called “ultra-fast potentials” (see below) trained on DFT reference data can be used as baseline potential for the Δ -learning approach. This would enable a computational speed-up by several orders of magnitude, this time, and pave the way to a more extended and comprehensive study of phase diagrams of hydrogen and hydrogen-rich materials, with a larger stretch in temperatures and pressures, and by preserving the QMC accuracy across the spanned phases. While the “ultra-fast potentials” developed as MLP models for QMC internal energies is still work in progress, the achievement made with the Δ -learning approach to develop QMC-quality MLP is remarkable, as proved in Refs. [8, 9]. In parallel, we are studying an improved strategy to generate unbiased training sets with QMC, made of internal energies, ionic forces and pressures for a given set of nuclear configurations, computed for the time being at the variational Monte Carlo level. This will require on the one hand an accurate wave function optimization per nuclear configuration, on the other hand it will need the development of improved workflows and workflow managers to control in an automated way the process of generating a set of QMC reference values on a given pool of nuclear configurations constituting the ML training set.



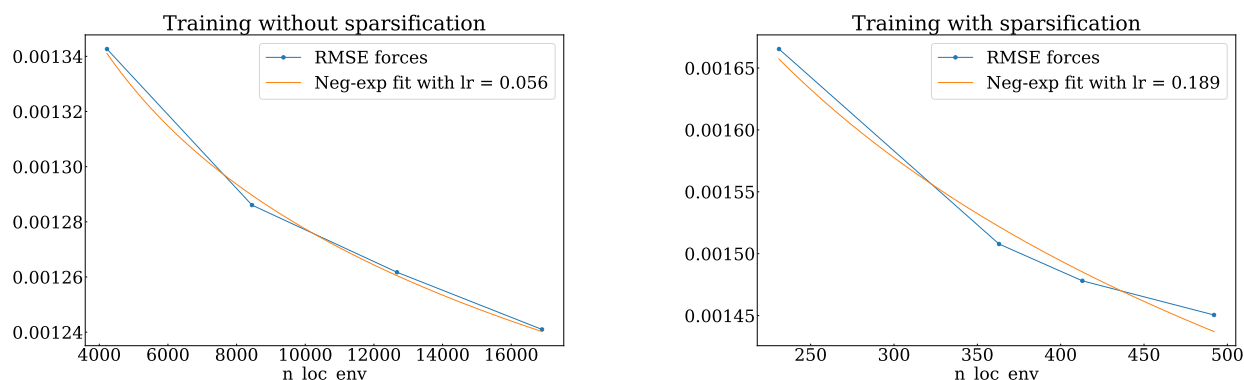


Figure 4: Learning rate of Δ -learning MLP for hydrogen. Shown are the RMSE of predicted forces as a function of the number of training local atom environments for liquid hydrogen, from which we estimate the learning rate, both with sparsification (right panel) and without (left panel). With sparsification, the model error approaches the error of the full model, but using one order of magnitude fewer training data.

Outputs

- Stephen R. Xie, Matthias Rupp, Richard G. Hennig: Ultra-Fast Interpretable Machine-Learning Potentials, arXiv: 2110.00624, 2021. *An extremely data-efficient MLP that enables training on QMC datasets.*
- Andrea Tirelli, Giacomo Tenti, Kousuke Nakano, Sandro Sorella: High-Pressure Hydrogen by Machine Learning and Quantum Monte Carlo, Physical Review B 106(4): L041105, American Physical Society, 2022. [\[doi\]](#)

5 Summary and outcomes

The open-source, high-performance, exascale-ready, inter-operable flagship QMC codes and libraries developed by the TREX consortium were used to study water and hydrogen. These two related materials are of high and direct relevance for technological progress and thus society.

Michele Casula's group (CNRS) investigated hydrogen's role in hydrogen bonds of water by exploring electronic properties affecting bond dynamics. Their study on protonated water hexamers using QMC methods revealed temperature-dependent proton behavior. To extend findings, they developed an MLP for water clusters.

TREX partners also studied the phase diagrams of hydrogen (H) and hydrogen-rich materials. The rich phase diagrams of these systems can not be computed with lower levels of theory and requires coupling QMC calculations for electrons with path-integral molecular dynamics or path-integral Monte Carlo for quantum nuclei. One of H's most accurate phase diagrams was calculated with the TREX code TurboRVB.



Training an MLP directly on QMC reference data fails as it is computationally too expensive to generate enough QMC training data. Instead, “ Δ -learning” was employed by the group of the late Sandro Sorella (SISSA) to train an accurate model for a liquid-liquid phase transition of high-pressure hydrogen using only 684 QMC calculations.

The groups of Matthias Rupp (UKON, LIST) and Michele Casula (CNRS) collaborate to determine whether so-called “ultra-fast potentials” trained on DFT reference data can be used as baseline potential for the Δ -learning approach for further acceleration by several orders of magnitude. Further efforts were made towards improved workflows for training set generation.

The following peer-reviewed published works (nine) and upcoming publications (one) are outputs related to this report and acknowledge TREX funding:

- Lorenzo Monacelli, Michele Casula, Kousuke Nakano, Sandro Sorella, Francesco Mauri: *Quantum phase diagram of high-pressure hydrogen*, Nature Physics 19(6): 845, 2023. DOI [10.1038/s41567-023-01960-5](https://doi.org/10.1038/s41567-023-01960-5)
- Félix Mouhat, Matteo Peria, Tommaso Morresi, Rodolphe Vuilleumier, Antonino Marco Saitta, Michele Casula: Thermal Dependence of the Hydrated Proton and Optimal Proton Transfer in the Protonated Water Hexamer, Nature Communications 14: 6390, 2023. DOI [10.1038/s41467-023-42366-4](https://doi.org/10.1038/s41467-023-42366-4)
- Marcel F. Langer, Florian Knoop, Christian Carbogno, Matthias Scheffler, Matthias Rupp: Heat Flux for Semilocal Machine-Learning Potentials, Physical Review B 108(10): L100302, American Physical Society, 2023. DOI [10.1103/physrevb.108.l100302](https://doi.org/10.1103/physrevb.108.l100302)
- Stephen R. Xie, Matthias Rupp, Richard G. Hennig: Ultra-Fast Interpretable Machine-Learning Potentials, npj Computational Materials 9: 162, Springer, 2023. DOI [10.1038/s41524-023-01092-7](https://doi.org/10.1038/s41524-023-01092-7)
- Giacomo Tenti, Andrea Tirelli, Kousuke Nakano, Michele Casula, Sandro Sorella: Principal Deuterium Hugoniot via Quantum Monte Carlo and Δ -learning, arXiv:2301.03570, 2023. DOI [10.48550/arXiv.2301.03570](https://doi.org/10.48550/arXiv.2301.03570)
- Andrea Tirelli, Giacomo Tenti, Kousuke Nakano, Sandro Sorella: High-Pressure Hydrogen by Machine Learning and Quantum Monte Carlo, Physical Review B 106(4): L041105, American Physical Society, 2022. DOI [10.1103/physrevb.106.l041105](https://doi.org/10.1103/physrevb.106.l041105)
- Dominik Lemm, Guido Falk von Rudorff, O. Anatole von Lilienfeld: Machine Learning based Energy-Free Structure Predictions of Molecules, Transition States, and Solids, Nature Communications 12: 4468, 2021. DOI [10.1038/s41467-021-24525-7](https://doi.org/10.1038/s41467-021-24525-7)
- Jan Weinreich, Nicholas J. Browning, O. Anatole von Lilienfeld: Machine Learning of Free Energies in Chemical Compound Space using Ensemble Representations: Reaching Experimental Uncertainty for Solvation, Journal of Chemical Physics 154(13): 134113, 2021. DOI [10.1063/5.0041548](https://doi.org/10.1063/5.0041548)
- Bing Huang, O. Anatole von Lilienfeld: Ab Initio Machine Learning in Chemical Compound Space, Chemical Reviews 121(16): 10001, 2021. DOI [10.1021/acs.chemrev.0c01303](https://doi.org/10.1021/acs.chemrev.0c01303)



- Thomas Bischoff, Bastian Jäckl, Matthias Rupp: Hydrogen under Pressure as a Benchmark for Machine-Learning Potentials, *in preparation*, to be published in 2024.

Results presented at the international workshop on Crash TEsting machine learning force fields: Applicability, best practices, limitations (TEA), University of Luxembourg, Luxembourg, October 23–25, 2023. [URL `tea-uni-lu.github.io`](https://tea-uni-lu.github.io)



References

- [1] F. Mouhat, M. Peria, T. Morresi, R. Vuilleumier, A. M. Saitta, and M. Casula, “Thermal dependence of the hydrated proton and optimal proton transfer in the protonated water hexamer,” *Nature Communications*, vol. 14, p. 6930, 2023.
- [2] F. Mouhat, S. Sorella, R. Vuilleumier, A. M. Saitta, and M. Casula, “Fully quantum description of the zundel ion: combining variational quantum monte carlo with path integral langevin dynamics,” *Journal of chemical theory and computation*, vol. 13, no. 6, pp. 2400–2417, 2017.
- [3] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, “Alchemical and structural distribution based representation for universal quantum machine learning,” *The Journal of Chemical Physics*, vol. 148, no. 24, p. 241717, 2018.
- [4] A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. A. von Lilienfeld, “FCHL revisited: Faster and more accurate quantum machine learning,” *The Journal of Chemical Physics*, vol. 152, no. 4, p. 044107, 2020.
- [5] L. Monacelli, M. Casula, K. Nakano, S. Sorella, and F. Mauri, “Quantum phase diagram of high-pressure hydrogen,” *Nature Physics*, vol. 19, pp. 845–850, 2023.
- [6] R. Taureau, M. Cherubini, T. Morresi, and M. Casula, “Quantum symmetrization transition in superconducting sulfur hydride from quantum Monte Carlo and path integral molecular dynamics,” *arXiv*, vol. 2307.15684, 2023. [Online]. Available: <https://arxiv.org/abs/2307.15684>
- [7] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, “Big data meets quantum chemistry approximations: The Δ -machine learning approach,” *Journal of Chemical Theory and Computation*, vol. 11, no. 5, pp. 2087–2096, 2015.
- [8] A. Tirelli, G. Tenti, K. Nakano, and S. Sorella, “High-pressure hydrogen by machine learning and quantum Monte Carlo,” *Physical Review B*, vol. 106, no. 4, p. L041105, 2022.
- [9] G. Tenti, A. Tirelli, K. Nakano, M. Casula, and S. Sorella, “Principal deuterium Hugoniot via quantum Monte Carlo and Δ -learning,” *arXiv*, vol. 2301.03570, 2023. [Online]. Available: <https://arxiv.org/abs/2301.03570>
- [10] M. F. Langer, F. Knoop, C. Carbogno, M. Scheffler, and M. Rupp, “Heat flux for semilocal machine-learning potentials,” *Physical Review B*, vol. 108, no. 10, p. L100302, 2023.
- [11] S. R. Xie, M. Rupp, and R. G. Hennig, “Ultra-fast interpretable machine-learning potentials,” *arXiv*, p. 2110.00624, 2021.
- [12] D. Lemm, G. F. von Rudorff, and O. A. von Lilienfeld, “Machine learning based energy-free structure predictions of molecules, transition states, and solids,” *Nature Communications*, vol. 12, p. 4468, 2021.



- [13] J. Weinreich, N. J. Browning, and O. A. von Lilienfeld, “Machine learning of free energies in chemical compound space using ensemble representations: Reaching experimental uncertainty for solvation,” *The Journal of Chemical Physics*, vol. 154, no. 13, p. 134113, 2021.
- [14] B. Huang and O. A. von Lilienfeld, “Ab initio machine learning in chemical compound space,” *Chemical Reviews*, vol. 121, no. 16, pp. 10 001–10 036, 2021.

