BuRN: A Buffer Region Neural Network Approach for QM/MM Simulations of Complex Systems

Bettina Lier, a,b Peter Poliak, a,b Philipp Marquetand, c Julia Westermayr* d and Chris Oostenbrink,* a

The study of complex systems, such as proteins or interfaces, has been tremendously advanced by combined quantum mechanics / molecular mechanics (QM/MM) simulations. However, the splitting of a system into a quantum chemical and molecular mechanical regions can cause artefacts at their interface and the high computational costs of quantum chemical regions still limit QM/MM simulations considerably. In this work, we develop BuRN, a buffer-region neural network, which can overcome the limitations of existing QM/MM schemes by introducing a buffer region in-between the inner quantum region and the environment. The buffer region minimizes artefacts that occur at the interface between common QM and MM regions and facilitates a full polarizable embedding of the buffer while the most important inner region is described at first principle level. High computational efficiency can be obtained with deep learning for the interaction between the QM and the buffer region. We demonstrate the performance of BuRN by deep learning enhanced QM/MM simulations of the hexa-aqua iron complex, which is notoriously difficult to describe by molecular mechanics only. Importantly, we show that BuRN opens the way to the calculation of alchemical free-energy differences at a quantum mechanical energy level, a so-far, highly challenging task.

Introduction

The study of complex systems, such as materials, hybrid interfaces, or biomolecules is of fundamental importance for gaining a better understanding of their physicochemical properties and consequently, enabling a targeted design of novel systems, like (transition-metal-) dichalcogenides for optoelectronic devices, 1–2 electrocatalysts, 3, 4 or drugs. 5 Especially metalloproteins have been the focus of many biological studies as their centres have been identified as prominent targets in many diseases like cancer, bacterial infection, or AIDS. 6, 7 However, the exact mechanisms involved in processes often remain inaccessible. Theoretical simulations could help to unravel these mechanisms, but the complexity and size of such systems makes computational studies highly nontrivial. 8, 9

Conventional approaches that exist to tackle such problems, either describe a system fully classically or in a quantum-mechanical (QM) molecular mechanics (MM) embedding (QM/MM) scheme. The latter approach is very powerful and enables a quantum mechanical description of a small, important part of a system, while the rest is treated classically. 10 Different QM/MM embedding schemes exist, namely mechanical, electrostatic, and polarizable embedding, which all have different advantages and drawbacks. 11–14

Mechanical embedding is the simplest approach, in which interactions between the MM and the QM regions are described via classical point charges assigned to the atoms in the QM region. The electronic degrees of freedom in the QM region are not affected by the MM environment and consequently, it is the least accurate of all approaches. In contrast, electrostatic embedding is physically better motivated as the QM system is perturbed by the MM particles with the MM charge distribution being embedded in the QM Hamiltonian. Hence, QM particles see MM particles as fixed point charges, which neglects polarization in the MM region. To account for these polarization effects, polarized force fields can be used to describe the MM region, which could be based on Drude oscillators or induced dipoles, for instance. 15, 16

To date, the current gold standard of all embedding methods is considered to be the polarizable one. 17 All embedding schemes discussed above are limited by high computational costs of the quantum chemical calculation and issues at the QM-MM interface. 17 In mechanical embedding, the electronic degrees of freedom do not experience the MM environment at all, while for electrostatic and polarizable embedding MM point charges close to the boundary may lead to an overpolarization in the QM region. Particularly challenging are boundaries that cross covalent interactions, necessitating a linked atom approach and possibly ad-hoc assumptions to avoid such polarization effects. Furthermore discrepancies between the forces
derived for the QM region and for the MM region can lead to artificial crowding or depletion at the interface, when particles are allowed to change character during a QM/MM simulation.

Alternatively, the whole system can be treated accurately using machine-learned force fields based on ab initio data.18 Machine learning is especially powerful for molecular dynamics as it can learn the relation between a descriptor, i.e., the structure of a system, and a targeted output, i.e., energies and forces for molecular dynamics simulations with the accuracy of the reference method but much lower computational costs. Such machine-learned force fields are available for specific materials (e.g., amorphous carbon19 or disordered silicon20), often at Density Functional Theory (DFT) level of accuracy, or for small organic molecules,21 often with higher accuracy such as coupled cluster22 for the ground state, but also for excited state simulations of single systems.23, 24 However, universal machine-learned force fields for more complex systems, such as biomolecules, still pose a challenge and similarly to traditional QM/MM simulations, they are limited by the computational expenses of the reference calculations.11, 18

Very recently, machine-learned potentials have been combined with QM/MM concepts, and have been applied for example to calculate free energies or transition paths.25-31 One challenge is the nontrivial task of including the effects of the environment into the machine-learned potential of the QM region. Even though only the QM region has to be learned, machine learning models need to capture effects due to the MM region. The introduction of a cutoff, up to which the MM region is included in the machine learning input, has emerged as a solution,27-29, 32 but requires extensive sampling of the environment.27, 29 In addition, all currently existing approaches are based on an electrostatic or mechanical embedding scheme. Thus, machine-learned QM/MM schemes experience the same problems as any other electrostatic and mechanical embedding scheme.

In this work, we circumvent the aforementioned problems by proposing an alternative embedding scheme in which an additional buffer region is introduced that experiences full electronic polarization by the QM region. The buffer region is described at both the QM and the MM level. Effectively, the interactions between the QM and the buffer regions are calculated entirely at the QM level, avoiding any of the artefacts described above. Similarly, the interactions between the buffer region and the MM region are entirely described at the MM level, also avoiding artefacts at the interface between the buffer region and the MM region. Within the buffer region, the interactions are a combination of MM-interactions and the effect of the QM region on the electronic degrees of freedom of the buffer region.

While this approach minimizes the artefacts that arise from mixing two levels of theory, it comes at considerable computational costs. Two QM calculations are required, one of the combined QM+buffer region and an additional calculation for the buffer region only. This makes the direct use in a purely QM/MM based buffer region approach largely unfeasible in practical applications. However, by using machine learning to describe QM-derived energy surfaces, an elegant solution emerges. In this work, we train deep neural networks (NN) to directly predict the difference of the two required QM calculations. Thus, this scheme does not require any additional particles in the machine learning setup, making any model applicable for this task. The evaluation of all relevant energies can efficiently be done with a single evaluation of the NN. We refer to this approach as buffer region with neural networks, or short BuRNN.

2 Buffer Region Neural Network (BuRNN)

The system in the BuRNN approach is split into an inner region I, a buffer region B, and an outer region O. The scheme is illustrated in Fig. 1. The total potential energy, \( V_{\text{tot}} \), is computed in the following way: The QM energy of the buffer region, \( V_{\text{QM}}^{B+} \), is subtracted from

\[
V_{\text{tot}} = V_{\text{QM}}^{I+} + V_{\text{QM}}^{I} + V_{\text{MM}}^{I+} + V_{\text{MM}}^{I} - V_{\text{QM}}^{B+} - V_{\text{MM}}^{B+}
\]

\( \approx V_{\text{QM}}^{I+} + V_{\text{MM}}^{I+} \)

Figure 1: Scheme of the BuRNN approach, which distinguishes three regions: 1) the inner region I (orange), which is described entirely by quantum mechanics (QM); 2) the buffer region B (blue), which is described both by QM and classical mechanics (MM) and 3) the outer region O (grey) which is described entirely by a classical MM force field. The right panel shows the process of BuRNN in more detail, including the training set generation. At every time step during molecular mechanics, two neural networks (A and B) are compared. When predictions diverge, the training set is expanded by additional QM calculations.
the QM energy of the inner+buffer region, \( V_{1+\mathbb{B}}^{\text{QM}} \), and subsequently, the energy of the buffer region is added from MM, \( V_{\mathbb{B}}^{\text{MM}} \) together with the MM energies of the outer region, \( V_{\mathbb{O}}^{\text{MM}} \):

\[
V_{\text{tot}} = V_{1+\mathbb{B}}^{\text{QM}} - V_{\mathbb{B}}^{\text{QM}} + V_{\mathbb{O}}^{\text{MM}}. \tag{1}
\]

As it becomes clear from equation 1, the BuRNN approach requires two QM calculations: one of the inner-buffer region and one of only the buffer region. The subscripts indicate the regions for which the potential energy is calculated and the superscripts denote the calculation method with QM referring to quantum mechanics and MM to molecular mechanics.

Even though interactions in QM are not pairwise additive, it is instructive to consider the different terms of equation 1 in terms of hypothetical pairwise interactions within or between regions. Pairwise interactions are indicated with a comma-separated subscript. Using the concept of hypothetical pairwise interactions, the first part, \( V_{1+\mathbb{B}}^{\text{QM}} \), can be separated into three terms, i.e., the energy that results from interactions within the QM region, \( V_{1+1}^{\text{QM}} \), the energy that results from interactions between the QM and the buffer region, \( V_{1+\mathbb{B}}^{\text{QM}} \), and the energy that results from interactions within the buffer region, \( V_{\mathbb{B}}^{\text{QM}} \):

\[
V_{1+\mathbb{B}}^{\text{QM}} = V_{1+1}^{\text{QM}} + V_{1+\mathbb{B}}^{\text{QM}} + V_{\mathbb{B}}^{\text{QM}}. \tag{2}
\]

To emphasize the fact that interaction energies are not pairwise additive, we use the addition \((\mathbb{I} + \mathbb{B})\) in the subscript of the last term of equation 2. These interactions between atoms in the buffer region are thus computed in the context of a QM-calculation of the joint inner and buffer regions. Accordingly, we write the interaction energy of the buffer region that results from a QM calculation of the buffer region alone as: \( V_{\mathbb{B}}^{\text{QM}} = V_{\mathbb{B}}^{\text{QM}} \). The influence of the QM region on the interactions within the buffer region can then be seen as the polarization of the buffer region due to the QM region, which is denoted as \( V_{\Delta \mathbb{B}}^{\text{QMpol}} \):

\[
V_{\Delta \mathbb{B}}^{\text{QMpol}} = V_{\mathbb{B}}^{\text{QM}} - V_{\mathbb{B}}^{\text{QM}}. \tag{3}
\]

In the next step, the force field is used to also compute the energy of the buffer region, \( V_{\mathbb{B}}^{\text{MM}} \). Note that if QM and MM agree perfectly, the terms \( V_{\Delta \mathbb{B}}^{\text{QMpol}} \) and \( V_{\mathbb{B}}^{\text{MM}} \) cancel exactly. Interactions within the MM region are computed with the force field, denoted as \( V_{\Delta \mathbb{B}}^{\text{MM}} \). The interactions between the outer region and the other regions, \( V_{\Delta \mathbb{O}}^{\text{MM}} \) and \( V_{\Delta \mathbb{D}}^{\text{MM}} \), are also treated at the MM level, but with partial charges of the inner and buffer regions that are derived from the QM calculation. These interactions comprise the last part of equation 1:

\[
V_{\Delta \mathbb{O}}^{\text{MM}} = V_{\Delta \mathbb{O}}^{\text{MM}} + V_{\Delta \mathbb{O}}^{\text{MM}}. \tag{4}
\]

In this sense, the inner and the buffer regions interact with the outer region via mechanical embedding. In this work, the electrostatic interactions of the inner and buffer regions with the outer region \( V_{\Delta \mathbb{O}}^{\text{MM}} \) are enabled by a deep NN trained on partial charges of the atoms in the \( \mathbb{I} + \mathbb{B} \) region that are derived from the QM calculation of the \( \mathbb{I} + \mathbb{B} \) region.

The total energy in terms of hypothetical interaction energies can then be written:

\[
V_{\text{tot}} = V_{1+\mathbb{B}}^{\text{QM}} + V_{1+\mathbb{B}}^{\text{QM}} + V_{\mathbb{B}}^{\text{MM}} + V_{\Delta \mathbb{B}}^{\text{QMpol}} + V_{\Delta \mathbb{O}}^{\text{MM}} + V_{\Delta \mathbb{D}}^{\text{MM}}. \tag{5}
\]

As can be seen, the BuRNN scheme ensures that the interactions within one region and between neighbouring regions are computed at the appropriate levels. One of the benefits is that any artefacts in the electronic degrees of freedom due to the interface to the outer region will largely cancel in the difference \( V_{1+\mathbb{B}}^{\text{QM}} - V_{\mathbb{B}}^{\text{QM}} \), assuming that the relevant polarization of the buffer region predominantly takes place at the interface between the inner region and the buffer region. Any remaining artefacts in the electronic degrees of freedom will arise at the interface between the buffer region and the outer region, which is relatively far away from the inner region of interest.

The interactions between \( \mathbb{I}, \mathbb{O} \) as well as \( \mathbb{B}, \mathbb{O} \) are computed using a mechanical embedding scheme, which is appropriate because of the large distances between inner and outer regions. Direct electronic influences to the inner region due to the outer region will be relatively small and the interaction is largely electrostatic. Remaining artefacts may arise from the fact that particles moving from the outer region into the buffer region switch from a mechanically embedded MM interaction to a full QM interaction with the inner region. Also in this case, the interaction is at a relatively long distance from the QM region, where the \( \mathbb{I}, \mathbb{B} \) interaction will largely be of electrostatic nature, such that these artefacts can be expected to be small.

Despite the accuracy and benefit of this scheme, the burden lies in the high computational costs that remain because two computationally expensive QM calculations are required. To overcome this limitation, we describe the first two terms of equation 1 directly using a deep neural network:

\[
V_{\text{NN}}^{\mathbb{I}+\mathbb{O}} \equiv V_{1+\mathbb{B}}^{\text{QM}} - V_{\mathbb{B}}^{\text{QM}}, \tag{6}
\]

which is equal to

\[
V_{\text{NN}}^{\mathbb{I}+\mathbb{B}} = V_{1+\mathbb{B}}^{\text{QM}} + V_{\Delta \mathbb{B}}^{\text{QMpol}} \tag{7}
\]

after insertion of equation 3 into equation 2. It now becomes clear that BuRNN is akin to a polarizable embedding with the polarization described at the full QM level. As can be seen, the NN learns the energy of the inner region, its interaction with the buffer region, as well as the difference in energy due to the inner region working on the buffer region, \( \Delta \mathbb{B} \). The Delta is used for the difference between the interactions within the buffer region including the inner region and excluding the inner region. Thus, the NN represents the full interactions within the inner region, the interactions between the inner and buffer region, and the polarization of the buffer region due to the inner region in a single term, \( V_{\text{NN}}^{\mathbb{I}+\mathbb{B}} \). The total BuRNN energy can be rewritten considering the NN to:

\[
V_{\text{tot}} = V_{\text{NN}}^{\mathbb{I}+\mathbb{O}} + V_{1+\mathbb{O}}^{\text{MM}} + V_{\mathbb{B}}^{\text{MM}} + V_{\Delta \mathbb{O}}^{\text{MM}} + V_{\Delta \mathbb{D}}^{\text{MM}}. \tag{8}
\]

The remaining terms can be computed from an inexpensive force field calculation from which the \( \mathbb{I}, \mathbb{I} \) and \( \mathbb{I}, \mathbb{B} \) interactions are excluded (see the methods section). The BuRNN approach reduces the computational costs for this approach considerably and enables long time scale and highly accurate molecular dynamics simulations of complex systems.
3 Results and Discussion

Here, we demonstrate the use of BuRNN for the hexacoordinated water-ion $[\text{Fe(H}_2\text{O)}_6]^{3+}$ complex in water. The description of transition metal–water interactions is notoriously difficult to describe by force fields, as the coordinative bond is somewhat in between a covalent and an ionic bond. To ensure proper coordination, several models have been proposed, e.g., by using additional charge sites or interaction terms. In our approach, the Fe$^{3+}$ ion comprises the inner region, and water molecules up to a distance of 0.5 nm are treated as the buffer region. This roughly accounts for the first two solvation shells, where the first solvation shell is expected to be formed by the well-known hexacoordinated water molecules. We train two types of deep NNs: One for the interaction energy between the Fe$^{3+}$ ion and the water molecules in the buffer region (equations 6 and 7) and one for the charge distribution of the Fe$^{3+}$ ion and the buffered water molecules to subsequently perform extensive molecular dynamics simulations. During these simulations, the water molecules are freely diffusing between the buffer and outer regions, smoothly switching interactions between the NN (QC) level and the MM (FF) level of theory.

Simulations with BuRNN

BuRNN simulations of $[\text{Fe(H}_2\text{O)}_6]^{3+}$ in water were carried out for 10 ns and compared to simulations using a pure force field description (denoted as “MM only”). To ensure that the simulation samples configurations that are well-described by the NN, a second NN that was independently trained from a selection of the same training data was evaluated every ps. The difference between the total energy evaluated by the two NNs was on average $-0.38 \pm 0.01$ kJ/mol, with a standard deviation of 1.17 kJ/mol. Figure 2a shows the Fe–O radial distribution functions, $g(r)$, of both simulations (pure MM in dark blue and BuRNN in orange). Both methods show a first distinct peak around 0.2 nm, corresponding to the first, coordinative solvation shell of 6 water molecules. In the force-field description, this peak is narrower and more pronounced. The second solvation shell is more pronounced in the BuRNN approach, corresponding to an average of 12.7 water molecules in the second solvation shell. It shows a maximum at 0.41 nm, while the pure MM simulation shows a broader peak with a maximum at 0.40 nm. Experimentally, the maximum of the second solvation shell was found at 0.415 nm and comprises 12 water molecules. Remarkably, the transition at 0.5 nm in the radial distribution function obtained with BuRNN, where the buffer region ends and the water molecules interact completely according to a force field description, is smooth and does not show any artefacts.

Figure S1 shows the profiles of the net forces on water molecules (panel a), the contribution of the NN treatment of the inner and buffer region (panel b and c), and their radial components (panels d–f). While the atom-wise forces (panel c and f) show a jump at the cutoff of 0.5 nm due to the switch from a purely point-charge based electrostatic interaction to a quantum mechanical interaction mediated by delocalized electron clouds, the net forces on complete water molecules as well as their NN contribution are smooth at the boundary. Figure S2 additionally compares the radial distribution functions for simulations, in which we used MM charges (force field charges, but referred to MM charges for consistency) and charges derived from quantum chemical calculations in the implicit solvent.

To account for the polarization due to a homogeneous solvent outside the buffer region on the charge distribution, an implicit solvent model was attempted. In these simulations, clear artefacts arise at 0.5 nm, demonstrating the relevance of using a consistent embedding scheme throughout.

During the 10-ns BuRNN simulation, the six coordinating water molecules did not exchange spontaneously. In contrast to water molecules within the first solvation shell, remaining water molecules in the buffer region, accounting for the second solvation shell, readily exchange with water molecules from the outer region. On average 18.7 water molecules are part of the buffer region. Out of the total number of 786 water molecules, all visit the buffer region at least once during the 10-ns simulation, with an average lifetime of 14.4 ps. This is in line with the estimates from NMR experiments that determine a lifetime which is below their resolution limit of 100 ps.

Figure 2: a) $\text{Fe}^{3+}$–O Radial distribution function, $g(r)$ of BuRNN (orange) and a purely MM description (dark blue). Intermediate colors from alchemical intermediates. b) Radial distribution functions for BuRNN with MM charges (orange) and the alchemical transition to a non-interacting dummy particle (in red); c) probability distribution of the improper dihedral defining the co-planarity of the Fe and a water molecule for BuRNN, a pure MM description and their alchemical intermediates; d) Development of the alchemical free-energy changes between BuRNN and MM only and BuRNN and a non-interacting dummy particle.
Experiments estimate a self-diffusion of the Fe$^{3+}$ ion in water of 0.55 - 0.68 10$^{-5}$ cm$^2$/ps.$^{36-39}$ We computed the self-diffusion rate for BuRNN simulations to be 1.09 10$^{-5}$ cm$^2$/ps, and for the pure MM simulations at 0.87 10$^{-5}$ cm$^2$/ps. The simulated data is in line with the observation that bulk SPC (simple point charge) water has a slightly too large diffusion constant.$^{40}$

During most of the simulation time, a hexa-coordination is maintained, which is also the case for pure MM simulations (see Fig. 3 a). We find that radial distances between the Fe-O (panel b) and O-Fe-O angles (panel d) are quite similar and agree well with the range of experimental estimates for the Fe-O bond lengths of 0.199 to 0.210 nm,$^{35, 41-47}$ and with the angles expected for an octahedral arrangement. However, the frequencies by which the Fe-O bonds vibrate are very different between BuRNN and the MM only simulations (panel c), implying that the Fe-O interaction is indeed not well captured by a purely non-bonded description. At the quantum level the vibrations take place at larger frequencies and are in good agreement with experiment.$^{48}$ Fig. 2 c shows the distribution of the O-H-H-Fe improper dihedral angle. A value of 0° for this angle, as predominant in pure MM simulations, implies that the water molecule and the Fe$^{3+}$ ion are in the same plane, while the larger values observed for BuRNN indicate a more tetrahedral arrangement in which the Fe interacts with the lone pairs on the oxygen. For comparison, a BP86-D3/def2-TZVP/COSMO estimate for this improper dihedral angle in [Fe(H$_2$O)$_{6}$]$^{3+}$ lies at 16°.$^{49}$ A more detailed discussion of the BuRNN simulations are provided in the SI.

**Free energy difference between BuRNN and force field**

The computationally inexpensive sampling of conformations with BuRNN opens the way to get sufficient sampling for alchemical free-energy calculations. The use of NN potentials to obtain robust free-energy differences at the QM level is an exciting possibility that is only just developing.$^{50, 51}$ In a first application, we have computed the free energy of turning the Fe$^{3+}$ ion in BuRNN into a pure force field (“MM only”) description. This allows for the creation of thermodynamic cycles in which free-energy differences between systems at the quantum level of theory are computed via a purely classical pathway.$^{52-56}$

The Bennett acceptance ratio (BAR) method$^{57}$ was used to compute the free-energy difference between the two states, which was evaluated to be 2619.4 ± 0.6 kJ/mol. Figure 2 d shows the development of the free-energy as a function of the coupling parameter $\lambda$, $\lambda = 1$ refers to a purely classical Hamiltonian, while $\lambda = 0$ refers to the BuRNN Hamiltonian. The overlap integral for the individual $\lambda$-segments is at least 0.80. From the Fe-O radial distribution functions in panel a, one can see very nicely how the BuRNN radial distribution function gradually changes into the pure MM version along the alchemical process. Similarly, the lower left-hand side panel c shows the distributions of the Fe-H-H-O improper dihedral, showing a gradual increase of the population at 0° and a decrease of the population around 25°. The positive value of the overall free-energy differences correlates with the fact that a force field description has intrinsically less degrees of freedom than a QM calculation to describe the interactions and also with the tighter coordination of water molecules observed in the “MM only” simulations.

**Fully disappearing a BuRNN particle**

The example above is a highly relevant, but relatively straightforward exercise, as the BuRNN and the “MM only” approaches sample distinct, but quite similar areas in phase space. A much bigger challenge for alchemical free-energy calculations at the QM level of theory is the complete disappearance of the Fe$^{3+}$ ion from the system, by perturbing it into a non-interacting dummy particle. For alchemical modifications involving dummy particles, there is a well-known end-state problem, which occurs if energies or energy derivatives need to be calculated for a configuration that is obtained when interactions are turned off, but using the energy function of the interacting particle. The non-interacting dummy particle may be overlapping with water molecules in the system, leading to infinite energies. In classical force fields, this problem is circumvented through the use of soft-core potentials, which are modifications of the Lennard-Jones and Coulombic functional forms for intermediate $\lambda$-values.$^{58}$ Unfortunately, the use of a soft-core interaction function is not possible with quantum chemistry. The use of NNs to approximate the QM interaction energies, as can be done in BuRNN, may even exacerbate the end-state problem, as the NN is typically not trained for configurations in which two particles overlap, leading to inappropriate predictions. However, recently König et al.$^{59}$ showed that minimum variance pathways$^{60-62}$ can be used to mimic soft-core potentials, without the need for modified interaction energy functions.

Here, we have used the minimum variance pathway approach to compute the free-energy associated with the alchemical removal of the central Fe$^{3+}$ ion, and constructed the total potential energy as a function of $\lambda$, as outlined in equation 10 in the Methods section. With increasing values of $\lambda$, the contribution of all energy terms within the inner region are reduced. As discussed above, strongly disrupted structures are sampled for larger $\lambda$ values, which is why the prediction of $V_{NN}^{BB}$ as well as charge predictions become unreliable. Therefore, we here used pure MM charges for the interactions of the

![Figure 3: Coordination of Fe$^{3+}$ by water molecules in simulations of BuRNN and for a straight MM description. a) number of water molecules observed within 0.25 nm; b) distribution of the Fe – O distance; c) power spectrum of the Fe-O coordinate bond; d) distribution of the O-Fe-O angles.](image-url)
inner and buffer regions with the outer region and were only able to simulate up to 0.2 ns for some $\lambda$-values.

The results of these simulations can be seen in Figure 2 b, in which the gradual loss of local structure is shown for selected $\lambda$-values. The artifact in the RDF for BuRNN at distances of 0.5 nm are due to the use of MM charges for the interactions of the inner and buffer regions with the outer region (see also Figure S2). The free energy between this simulation and a simulation in which the partial charges for these interactions were derived from a NN was estimated using BAR\(^{57}\) to be $-29.3 \pm 0.4$ kJ/mol. The free-energy difference of subsequently turning off the BuRNN interactions completely is shown as a function of $\lambda$ in panel d (orange to red function). The free energy of fully removing the Fe\(^{3+}\) ion was estimated at $+5952 \pm 4$ kJ/mol, with a minimum overlap integral in the BAR calculation of 0.58.

The full free energy of removing the BuRNN particle is hence estimated at $-29 + 5952 = 5923$ kJ/mol. This can be compared to the sum of turning the BuRNN particle into a classical particle obtained in the previous paragraph (2619 kJ/mol) and the free energy of removing an entirely classically decribed Fe\(^{3+}\) ion from the system, which was estimated at 3410 kJ/mol, leading to 6029 kJ/mol. Considering the short simulations on which the alchemical free energy changes are currently based, the agreement is reasonable. The negative of this value can theoretically be compared to the experimental single ion solvation free energy, which is reported to be between $-4416$ and $-4265$ kJ/mol.\(^{63}\) However, it is well known that charge-changing free-energy calculations are heavily influenced by finite size artefacts in the treatment of the electrostatic interactions.\(^{64,65}\) It is not straightforward to determine the appropriate correction terms for such a heavily charged particle in the context of the BuRNN potential energy. The aim of our current work is to show that difficult alchemical modifications become feasible with BuRNN. The challenges in describing charge changes, how to properly include unphysical structures visited for larger $\lambda$-values in the NN and the optimization of the parameters $s$ and $V_{\text{offset}}$ in equation 10 will be the subject of future research.

BuRNN is robust with respect to external perturbations

Lastly, we sought to test if the BuRNN approach is robust with respect to alternative simulation settings. Therefore, we performed simulations at elevated temperatures of 320 K, 340 K, 360 K, 380 K, and 400 K. No spontaneous exchanges of water molecules were observed during these simulations, but an increased occurrence of penta-coordinated configurations was observed from 0.43 % at 300K to 3.61 % at 380 K and 1.83 % at 400 K, see Fig. S3, due to the fluctuations in the Fe-O distances to values slightly larger than the cutoff distance of 0.25 nm.

To quantify the strength of the Fe-O interaction and to test if the BuRNN approach is robust with respect to external perturbations, we next pulled one of the coordinating water molecules away from the Fe\(^{3+}\) ion. This was done in 21 simulations with a harmonic distance restraint that was gradually increased from 0.2 nm to 0.6 nm with 0.02 nm increments and a force constant of 2.5 \textit{M} kJ/mol/nm\(^2\). For every restraining distance 10 ps of equilibration was followed by 50 ps of data sampling. The weighted histogram analysis method was used to compute the free-energy profile as a function of the Fe-O distance.\(^{66,67}\) Figure 4 shows the free-energy profile along with the number of coordinating waters, using a cutoff of 0.25 nm. As the water molecule is pulled away from the iron, the free energy goes up to a barrier with a maximum at of 43.8 kJ/mol at 0.34 nm. This distance corresponds to the minimum in the radial distribution function between the first and second solvation shells. Note that while the water molecule is pulled through the second solvation shell (from 0.34 to 0.5 nm), other water molecules spontaneously find their way to the Fe\(^{3+}\) ion and the six-fold coordination is restored.

After the second solvation shell, the free-energy profile levels off to around 16 kJ/mol, which represents the hysteresis of the process and the underlying methodology. As the initial and final states are identical (hexa-coordinated iron in water), one would expect an overall free-energy difference of 0 kJ/mol. However, as a sixth coordinating water only enters the coordination shell of the iron when the initial water is already at relatively large distances, the energy gain from restoring the hexa-coordination is not reflected in the distance restraint on the initial water anymore and a free-energy offset remains. This could potentially be addressed by longer simulations or alternative sampling schemes, for instance using Hamiltonian replica exchange to ensure reversibility. For the current application, we were quite pleased to see the spontaneous rebinding of a water molecule (see supplementary movie), in which the complex transition from a penta-coordinated trigonal bipyramid to a hexa-coordinated octahedral arrangement is visited during dynamics (see also the insets of Fig. 4).

Throughout all simulations discussed in this section, two NNs agreed very well (see section S2.3).
Deep neural networks

The aforementioned BuRNN simulations were enabled by replacing costly quantum chemical calculations with computationally inexpensive deep NN predictions. In this work, a message-passing deep convolutional NN, namely SchNet, is trained on the interaction energies between the QM and the buffer region to yield $V_{NN}$. The training data contains 11,430 data points for the interaction energies and interaction forces as described in equations 6 and 7 and 11,445 data points for charges of the inner and buffer region. The training data is based on Density Functional Theory (DFT) and was generated via initial sampling based on snapshots from purely molecular mechanics simulations of the targeted system and was extended using adaptive sampling. The workflow of the training data expansion is illustrated in Figure 1 in the right panel. Two initially trained NNs, A and B, predict interaction energies, interaction forces, and charges that are used to carry out QM/MM simulations. Every picosecond the interaction energies of the two NNs are compared. Whenever they are similar, i.e., when $V_{NN(A)} = V_{NN(B)}$ at $t$, NN predictions are deemed accurate and dynamics are carried on. Whenever the predictions start to diverge stronger from each other and exceed a predefined threshold, two reference QM calculations ($V_{QM}^{A}$ and $V_{QM}^{B}$) are executed and their difference is added to the training set. This process was carried out until dynamics could be run up to 1 nanosecond (ns) without any further interruptions. This exact thresholds and details are discussed in more details in section S1.2.

In total, 4 adaptive sampling runs were carried out and final models are trained on 9,500 data points for training and validating, while the rest are kept for testing. The mean absolute error of a model trained directly on the interaction energy and interaction forces is $1.7 \pm 0.25$ kJ/mol and $8.4 \pm 0.4$ kJ/mol/nm, respectively. The errors on each model separately are given in Table S1.

To facilitate a mechanical embedding approach for the interactions of the inner and buffer regions with the outer region, we extend the NN setup to train an additional NN on partial atomic charges of the full QM region, i.e., the inner and the buffer region. Details on the network architecture used and the screened hyperparameters for training can be found in section S1.1.

The learning curves, in which each point shows the mean and the standard deviation of the mean absolute error (MAE) for 4-20 models trained on the interaction energy, interaction forces, and atomic partial charges, are shown in Fig. 5, panels a, c, and e, respectively. The corresponding scatter plots of 5 models trained on interaction energies and interaction forces using the largest training set size are illustrated in panels b and d, respectively. For charges, only one model is used for dynamics, hence only one model is shown in panel f. The scatter plots are shown for the test set. In each plot, the systems for which the QM calculations are performed are indicated at the bottom. As can be seen from the scatter plots, the model learns interaction energies and forces accurately. Except for a few data points with larger errors (that scatter stronger), model predictions almost perfectly match the reference values. Data points with larger errors are related to energetically unfavourable structures that were visited mainly during the last adaptive sampling run, mostly during dynamics with BuRNN on long time scales. As can be seen from the learning curves, all models learn properly, as a linear relation is observed when plotting the data on a log-log scale for the number of data points used for training and the MAEs of

![Image](image_url)

Figure 5: Learning curves for (a) interaction energies, (b) interaction forces, and (c) atomic partial charges that show the mean of 4-20 models trained on a given training set size including their standard deviation. Learning curves are shown in a logarithmic scale. Scatter plots for (d) interaction energies, (e) interaction forces, and (f) atomic partial charges of models trained on the largest training set size are shown. For adaptive sampling, at least two NN models are used for energies and forces, while only one model is used for atomic partial charges, hence only one model is shown. The interaction or region that is modelled is indicated at the bottom of each panel.

4 Conclusions

We have introduced a buffered QM/MM embedding scheme termed BuRNN that minimizes artefacts at the interface between regions that are treated at different levels of theory. The scheme ensures that interactions that go over the boundaries are treated at a consistent level of theory and includes the energies and forces of a full quantum mechanical polarization in the buffer region. Inconsistencies in the description of e.g. polarization at the edges of the buffer region i) can be expected to cancel in the difference between the two QM terms and ii) are far removed from the inner region of interest. These advantages come at the cost of an additional QM calculation for the buffer region only, a problem that is elegantly solved by training deep neural networks directly on the difference of the QM energy terms. A single evaluation of the NN is
needed to evaluate the BuRNN energies and a second NN is used to
derive charges for inner and buffer regions to classically describe
the interactions with the outer region. BuRNN allows for fast simulations
and extensive sampling of an enhanced QM/MM system.

Compared to full machine-learned force fields based on \textit{ab initio}
simulations that require an external description or adaption for long-
range interaction and are limited by the costs of accurate reference
data that are not available for arbitrarily large system, BuRNN has the
advantage of being generally applicable to any system and useable
with any existing machine learning model that exists and can
describe molecular systems.

We have demonstrated the BuRNN approach at the hand of the
hexaqua iron complex \([\text{Fe(H}_2\text{O)}]^{3+}\) in water. BuRNN leads to
realistic simulations of this system, is amenable to consistent
alchemical free-energy calculations and robust to perturbations due
to higher temperatures or external forces on the configuration.

The BuRNN approach opens the way for many additional applications
or lines of research. For instance, BuRNN can be applied to transition
metals in metal-organic frameworks or in metalloproteins. The use
of alchemical modifications at a QM level and in a QM/MM context
is highly interesting in many research areas. For instance, in drug
discovery, the definition of a force field for new chemical moieties
can be bypassed by directly training a NN for the molecular
interactions, while still relying on established force fields for the
protein degrees of freedom. Proper training of neural networks in
the presence of a buffer region could also facilitate to description
of chemical reactions in a BuRNN setting and allow for extensive
sampling of reactive pathways in the condensed phase.

5 Computational Details
5.1 Quantum Chemical Calculations
Total electronic energies and gradients of the training data points
were all calculated in the Gaussian 16 program package\textsuperscript{71} using spin-
unrestricted density functional theory (DFT) in the gas phase. The
iron-containing structures were calculated in the sextet state with a
net charge of +3. Based on the benchmark calculations of iron
complexes,\textsuperscript{49} the OPTX exchange\textsuperscript{72} with PBE correlation functional\textsuperscript{73}
(OPBE) were used in def2SVPP basis set.\textsuperscript{74} The self-consistent field
(SCF) procedure was performed using the ‘yqc’ option. The energy
minimisations were carried out by the Gaussian 16 default Berny
optimisation algorithm. The spin contamination of the sextet species
was below 10^{-5} a.u. The use of the input orientation was enforced
using the ‘nosymm’ keyword. Partial atomic charges were obtained
from single-point calculations on the QM+buffer region by fitting to
the electrostatic potential using the Merz-Singh-Kollman scheme.\textsuperscript{71, 75, 76}

To account for the effect of the solvation on the charge
distribution, implicit solvation was included using the SMD variant
of the integral-equation-formalism polarizable continuum model.\textsuperscript{77} The
partial atomic charges in the water droplet of 1.0 nm diameter
calculated by this approach are very similar to the charges of the SPC
water model, therefore their adjustment for this difference was not
needed.

5.2 Machine Learning
Deep NNs used in this work are based on the continuous-filter
convolutional convolutional NN SchNet\textsuperscript{59, 60} which is a message-passing NN that
learns the molecular descriptor in addition to its relation to target
properties. In this work, we use SchNet to train on the interaction
energies, \(\mathcal{V}_{\text{NN}}^{\text{I+AB}}\) (see equations 6 and 7) and the corresponding
interaction forces, \(\mathbf{F}_{\text{NN}}^{\text{I+AB}}\), calculated as derivatives of the NN
potentials:

\[
\mathbf{F}_{\text{NN}}^{\text{I+AB}} = -\frac{\partial \mathcal{V}_{\text{NN}}^{\text{I+AB}}}{\partial \mathbf{r}}
\]

\(\mathbf{r}\) denotes the atomic positions of all atoms within the QM region
(inner and buffer region). The loss function used for training contains
both terms, i.e., interaction energies and interaction forces, and a
tradeoff is applied to weigh properties differently. In this work, we
used a tradeoff of 0.05 for interaction energies, which means the
 corresponding forces have a weight of 0.95 (1.0-0.05) (see SI for further
details). In the evaluation of the NN, the forces are computed as the
analytical derivative of the predicted energies, to ensure detailed
balance in the simulations.

To describe atomic partial charges, another SchNet model was
trained on the full QM region (inner + buffer region) that was
adapted such that all atomic partial charges could be fit in one NN.
To obtain the reference partial charges, the full QM calculations of
the inner and the buffer region was used, hence no interaction
charges were trained. Each atom described in its chemical and
structural environment by the NN gave rise to a corresponding
atomic partial charge. Therefore, the last pooling layer, which usually
sums or averages over all atomic contributions for a target property,
was removed. During the simulations, the predicted charges were
adjusted to sum up to +3 exactly by homogeneously distributing any
charge deficit or surplus (typically 5\%) over all particles in the inner
and buffer regions.\textsuperscript{78}

Hyperparameters were assessed for both, interaction energy and
interaction force models and atomic partial charge models
separately. A random grid search was applied to obtain optimal
model parameters. Details on the range of sampled
hyperparameters can be found in the S1.1.

To assess whether models can learn interaction energies, interaction
forces, and atomic partial charges properly, we computed learning
curves, which show the mean absolute error of several individually
trained models for a given training set size in logarithmic scale. For
proper learning a linear declines is expected for the loss function on
a double logarithmic scale. The number of models that were trained
for a given training set size was chosen such that the standard
deviation of the MAE was within 1 kJ/mol (less than 10 meV).

5.3 Molecular Dynamics Simulations
Conventional MD
All molecular dynamics simulations were performed with a modified version of the GROMOS simulation package, with a direct interface to SchNetPack modules using the pybind11 library. As the code is deeply integrated into GROMOS, it will be part of the next release and will become freely available at www.gromos.net. Until then, the code is available upon request. Example input files to run a BuRNN simulation are provided.

Simulations were performed in a periodic cubic box with box-edge lengths of 2.91 nm, containing one Fe$^{3+}$ ion, and 786 SPC water molecules. Temperature was maintained at 298 K using Nosé Hoover chains scheme by coupling the centre-of-mass motion and internal/rotational degrees of freedom to two separate temperature baths with coupling time of 0.1 ps and four chains. Bond lengths and angles in SPC water were constrained using the SHAKE algorithm. Unless stated otherwise, a timestep of 2 fs was used. Lenov-Jaard parameters of the Fe$^{3+}$ atom were taken from Zou et al., and the effective charge was set to $+3.0$. For the nonbonded MM interactions we applied a cutoff of 1.4 nm based on the distances between the heavy atoms (charge-group based cutoff) with a reaction-field of relative permittivity of 61 to account for a homogeneous medium outside the cutoff. The pairlist was updated every step.

The BuRNN buffer region was created by applying a charge-group based cutoff of 0.5 nm from the central ion. Energies and forces were obtained from the previously trained SchNet model. We simultaneously used two models trained on the same data sets with different training and validation splits. The second model was used to validate the first model by comparing the energy predictions on-the-fly. Their mutual disagreement was used to pick snapshots for the next round of adaptive learning. Our implementation allows to monitor this quantity throughout the simulation. Partial atomic charges were obtained from another, adapted atomistic SchNet model and applied to the Fe$^{3+}$ ion and the water molecules inside the buffer region. To avoid double-counting of the water-water interactions inside the buffer, their standard MM parameters and partial charges were used. The NN attributed charges were used only for the Coulombic interactions of the inner and buffer regions with the outer region. The coordination number of the iron ion was determined using a 0.25 nm cutoff, based on an analysis of initial radial distribution functions. Hydrogen bonds were determined using a geometric criterion and were considered present if any hydrogen atom was within 0.25 nm of the oxygen atom of a neighbouring water molecule, while the O–H–O angle was at least 135°. Diffusion of the iron ion was determined using the Einstein equation by applying a linear fit to the mean square displacement over timescales of 0–1 ns.

Free-energy perturbations

Two types of alchemical free energy calculations are performed. First, the BuRNN Fe$^{3+}$ ion was gradually perturbed into a Fe$^{2+}$ ion that was described by MM only. For this, we created a system of two Fe$^{3+}$ ions, with half the physical mass each. These were connected by a strong distance restraint with reference distance 0 and a harmonic force constant of $10^5$ kJ/mol/nm$^2$. The distance between the two Fe$^{3+}$ particles remained at 0.041 ± 0.006 nm. A coupling parameter $\lambda$ was defined that linearly reduces the BuRNN potential energy on the first iron to zero and that increases the classical interactions on the second particle from a non-interacting dummy particle to a fully interacting classical Fe$^{3+}$ ion. Eleven independent 1-ns simulations were performed for 1 ns at equidistant $\lambda$-values and energies for neighbouring $\lambda$-values were recalculated every 0.1 ps. The two particles were kept together with a tight distance restraint, of which the distance remained at 0.041 ± 0.006 nm. The average difference between two independent NNs remained at values of $-0.14 \pm 0.02$ kJ/mol at $\lambda = 0.5$ to $-0.58 \pm 0.03$ kJ/mol at $\lambda = 0.9$. Only the standard deviation of this measure increased slightly from 1.3 kJ/mol at $\lambda = 0.0$ to 2.3 kJ/mol at $\lambda = 1.0$. This means that the NN is still able to make consistent predictions for conformations that were generated from a purely classical Hamiltonian ($\lambda = 1.0$).

Next, a single BuRNN Fe$^{3+}$ ion was perturbed into a non-interacting dummy particle. We constructed a $\lambda$-dependent potential energy function that mimics a soft-core interaction through the use of an EDS-like potential. The value of $\lambda$ corresponds to the lowest value of $V_{\text{NN}}$ in the training set, incremented by 800 kJ/mol to ensure that favourably interacting configurations contribute significantly to the term in the straight brackets. This functional form mimics a soft-core potential and ensures that at $\lambda = 0$, $V_{\text{NN}}(\lambda)$ is identical to equation 1 while at $\lambda = 1$ only the MM terms for the buffer and outer regions of equation 1 remain. Simulations of 0.5 ns were performed at $\lambda$-values separated by 0.02 in the range from 0 to 0.3 and subsequently by 0.1 until the final value of 1.0. As $\lambda$ progresses, the forces due to the NN become smaller and the water molecules start to show overlap with the Fe$^{3+}$ ion. As these intermediate conformations are highly unphysical and strongly disrupted, $V_{\text{NN}}$ values typically become very large (about 1600 kJ/mol higher than $V_{\text{NN}}$) and through the exponential formulation, the contribution of this energy term quickly becomes negligible, mimicking the soft-core potential. Unfortunately, as these conformations were not part of the training set for the NN, in rare cases (about 2% of the time) extremely low values (below $V_{\text{NN}}$) are predicted for $V_{\text{NN}}$ leading to unphysically large forces and a failure to maintain the bond-length constraints of the water molecules involved. For this reason, the simulations at larger values occasionally fail. Only simulations were included in these calculations that ran successfully for at least 0.2 ns. Similarly, to avoid the assignment of random partial charges at values of $\lambda$, where configurations are sampled that are unknown to the NN, these calculations were performed with constant force field charges for the inner and buffer regions to compute $V_{\text{NN}}$ and $V_{\text{NN}}$.

Data Availability

The training data in the Atomic Simulation Environment (ASE-db) format and the trained NNs are freely available.
on figshare. In addition, we include input configurations and parameters to perform a BuRNN simulation.

Author Contributions

Bettina Lier: conceptualization, investigation, data curation, funding acquisition, software, writing – review and editing, visualization; Peter Poliak: investigation, data curation, methodology, software, writing – original draft, review and editing; Philipp Marquetand: methodology, supervision, writing – review and editing; Julia Westermayr: investigation, data curation, methodology, writing – original draft, review and editing, visualization; Chris Oostenbrink: conceptualization, investigation, data curation, methodology, writing – original draft, review and editing, supervision, resources.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Michael Gastegger for discussion regarding the training process of ML models. B.L. is a recipient of a DOC Fellowships of the Austrian Academy of Sciences (OAW) at the Institute for Molecular Modeling and Simulation at the University of Natural Resources and Life Sciences, Vienna (grant no. 25743). The project was further supported by the Austrian Science Fund, FWF doctoral program BioToP (W1224) and the Erwin-Schrödinger project J 4522-N (J.W.). The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC).

Table of Content Graphic

Notes and references

1. Q. Zeng and Z. Liu, Advanced Electronic Materials, 2018, 4, 1700335.
2. B. W. H. Baugher, H. O. H. Churchill, Y. Yang and P. Jarillo-Herrero, Nature Nanotechnology, 2014, 9, 262-267.
3. Y. Li, S. Li, A. V. Nagarajan, Z. Liu, S. Nevins, Y. Song, G. Mpourmpakis and R. Jin, Journal of the American Chemical Society, 2021, 143, 11102-11108.
4. N. Govindarajan, G. Kastlunger, H. H. Heenen and K. Chan, Chemical Science, 2021, DOI: 10.1039/D1SC04775B.
5. A. Y. Chen, R. N. Adamek, B. L. Dick, C. V. Credille, C. N. Morrison and S. M. Cohen, Chemical Reviews, 2019, 119, 1323-1455.
6. L. Riccardi, V. Genna and M. De Vivo, Nature Reviews Chemistry, 2018, 2, 100-112.
7. D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, Nature Reviews Drug Discovery, 2004, 3, 935-949.
8. W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt and H. B. Yu, Angewandte Chemie International Edition, 2006, 45, 4064-4092.
9. M. Karplus and J. Kuriyan, Proceedings of the National Academy of Sciences of the United States of America, 2005, 102, 6679-6685.
10. M. J. Field, P. A. Bash and M. Karplus, Journal of Computational Chemistry, 1990, 11, 700-733.
11. Q. Cui, T. Pal and L. Xie, The Journal of Physical Chemistry B, 2021, 125, 689-702.
12. E. Brunck and U. Rothlisberger, Chemical Reviews, 2015, 115, 6217-6263.
13. H. M. Snn and W. Thiel, Angewandte Chemie International Edition, 2009, 48, 1198-1229.
14. J. G. and D. G. Truhlar, Annual Review of Physical Chemistry, 2002, 53, 467-505.
15. A. Ganguly, E. Boulanger and W. Thiel, Journal of Chemical Theory and Computation, 2017, 13, 2954-2961.
16. D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci and J.-P. Piquemal, Journal of Chemical Theory and Computation, 2017, 13, 4025-4033.
17. M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini and B. Mennucci, Physical Chemistry Chemical Physics, 2020, 22, 14433-14448.
18. O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, Chemical Reviews, 2021, DOI: 10.1021/acs.chemrev.0c01111.
19. V. L. Deringer and G. Csányi, Physical Review B, 2017, 95, 094203.
20. V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold and S. R. Elliott, Nature, 2021, 589, 59-64.
21. J. S. Smith, O. Isayev and A. E. Roitberg, Chem. Sci., 2017, 8, 3192-3203.
22. H. E. Sauceda, S. Chmiela, I. Poltavsky, K.-R. Müller and A. Tkatchenko, The Journal of Chemical Physics, 2019, 150, 114102.
23. J. Westermayr, M. Gastegger, D. Vörös, L. Panzenboeck, F. Joerg, L. González and P. Marquetand, arXiv, 2021, 2108.04373.
72. W.-M. Hoe, A. J. Cohen and N. C. Handy, *Chemical Physics Letters*, 2001, **341**, 319-328.
73. J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865-3868.
74. F. Weigend and R. Ahlrichs, *Physical Chemistry Chemical Physics*, 2005, **7**, 3297-3305.
75. U. C. Singh and P. A. Kollman, *Journal of Computational Chemistry*, 1984, **5**, 129-145.
76. B. H. Besler, K. M. Merz Jr. and P. A. Kollman, *Journal of Computational Chemistry*, 1990, **11**, 431-439.
77. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *The Journal of Physical Chemistry B*, 2009, **113**, 6378-6396.
78. M. Gasteiger, J. Behler and P. Marquetand, *Chemical Science*, 2017, **8**, 6924-6935.
79. W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger and W. F. van Gunsteren, *The Journal of Physical Chemistry A*, 1999, **103**, 3596-3607.
80. J. Wenzel, J. Rhinelander and D. Moldovan, pybind11 - Seamless operability between C++11 and Python, https://rgl.epfl.ch/software/pybind11).
81. Westermayr, Julia (2021): BuRNN. figshare. Dataset. https://doi.org/10.6084/m9.figshare.17088770.v1
82. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, in *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, ed. B. Pullman, Springer Netherlands, Dordrecht, 1981, DOI: 10.1007/978-94-015-7658-1_21, pp. 331-342.
83. W. G. Hoover, *Physical Review A*, 1985, **31**, 1695-1697.
84. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, **1977**, **23**, 327-341.
85. C. Zou, M. Larisika, G. Nagy, J. Srajer, C. Oostenbrink, X. Chen, W. Knoll, B. Liedberg and C. Nowak, *The Journal of Physical Chemistry B*, 2013, **117**, 9606-9614.
86. I. G. Tironi, R. Sperb, P. E. Smith and W. F. van Gunsteren, *J. Chem. Phys.*, 1995, **102**, 5451-5459.
87. A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubai, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütz, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.
Supplementary Information for: BuRNN: A Buffer Region Neural Network Approach for QM/MM Simulations of Complex Systems

Bettina Lier, †a Peter Poliak, †a,b Philipp Marquetand,c Julia Westermayr*d and Chris Oostenbrink,*a

Contents

S1 Deep neural networks .................................................................................................................. 1
S1.1 Neural network training ............................................................................................................ 1
S1.2 Data set generation .................................................................................................................... 3
S2 BuRNN simulations ....................................................................................................................... 4
S2.1 BuRNN forces ............................................................................................................................ 4
S2.2 Radial distribution function ...................................................................................................... 5
S2.3 BuRNN is robust with respect to external perturbations .......................................................... 6
References .......................................................................................................................................... 7

S1 Deep neural networks

S1.1 Neural network training

For training energies and forces, we use SchNet1,2 and adapt it to train partial atomic charges. As mentioned in the main text, besides default parameters, we use a cutoff of 1.0 nm and a batch size of 8. The network parameters were validated on a random grid using a training:validation:test split of approximately 8:1:1 and an initial training set of 7,864 data points. As SchNet is a message passing neural network (NN), which automatically generates a tailored representation for a given system, it can be seen as a connection of two NNs. The size of the network that models the descriptor based on the structural inputs and elemental charges is defined by 6 interaction layers, 256 features to represent the atoms and 50 Gaussian functions. The number of Gaussian functions was set to 100 for modelling atomic partial charges. The interaction layers were sampled from 4-8, the features from 128-512, and the number of Gaussian functions placed on each atom from 25-200. The learning rate was varied between 0.001 and 0.00001, whereas the default of 0.0001 was most appropriate for training the final data set using 11,430 data points. A larger learning rate up to 0.001 was used when training smaller training set sizes. The hidden layers to map the descriptor to the output energies was kept at 3 with more layers not improving training.

The training set contained outliers, i.e., unfavourable structures, e.g., 1-4 coordinated aqua-Fe-complexes, with large energies and forces (but not necessarily large interaction energies and interaction forces). To allow a training of a diverse data set including such “outliers”, we adapted SchNet and used an adapted version of the smooth L1 loss function, which allows to reduce the influence of outliers on the model training. Therefore, the model switches between L1 and L2 loss during training:

\[ L_{\text{batch}} = \begin{cases} 
L_2 & \text{if } \max(|v_{NN} - (v_{QM} - v_{QM})|) < 3 \text{ mean}(|v_{NN} - (v_{QM} - v_{QM})|) \\
L_1 & \text{if } \max(|v_{NN} - (v_{QM} - v_{QM})|) \geq 3 \text{ mean}(|v_{NN} - (v_{QM} - v_{QM})|) 
\end{cases} \]

with
\[ L_2 = t \left\| v^{NN}_{i+\Delta R} - (v^{QM}_{i+\Delta R} - v^{QM}_B) \right\|^2 + (1 - t) \left\| -\frac{\partial v^{NN}_{i+\Delta R}}{\partial R} + \left( \frac{\partial v^{QM}_{i+\Delta R}}{\partial R} - \frac{\partial v^{QM}_B}{\partial R} \right) \right\|^2 \]

and

\[ L_1 = t \left\| v^{NN}_{i+\Delta R} - (v^{QM}_{i+\Delta R} - v^{QM}_B) \right\| + (1 - t) \left\| -\frac{\partial v^{NN}_{i+\Delta R}}{\partial R} + \left( \frac{\partial v^{QM}_{i+\Delta R}}{\partial R} - \frac{\partial v^{QM}_B}{\partial R} \right) \right\| \]

The tradeoff, \( t \), is used to weigh forces and energies during training and was sampled between 0.001 and 100. The optimal value was found to be 0.05. \( R \) denotes the atomic positions. The forces are treated as derivatives of the NN interaction energies. Note that we train forces and not gradients, hence the negative sign.

Training directly on interaction energies and interaction forces has several advantages compared to training on the separate terms arising from the inner+buffer region and the buffer region alone. On one hand, our setup allows for higher accuracy and consequently, better data efficiency, as the interaction energies span a smaller energy window than the total energies of the inner+buffer region and the buffer region. While the mean absolute error (MAE) on total energies and forces for models trained on 9,500 data points is in the range of 500 kJ/mol and 5,000 kJ/mol/nm, respectively (resulting in an error of around 700 kJ/mol and 7,000 kJ/mol/nm for interaction energies and interaction forces, respectively), training directly on interaction energies and interaction forces leads to errors about 100-1000 times smaller, i.e., 1.7±0.25 kJ/mol and 8.4±0.4 kJ/mol/nm, respectively. The MAEs and root mean squared errors (RMSE) on a holdout test set for models trained on interaction energies and interaction forces are shown in Table S1.

### Table S1: Mean absolute error (MAE) and root mean squared error (RMSE) for energies and forces of 5 models trained on 9,500 data points corresponding to Fig. 2 (a-d) in the main text

| Neural Network Model | Energy MAE (RMSE) [kJ/mol] | Forces MAE (RMSE) [kJ/mol/nm] |
|----------------------|-----------------------------|--------------------------------|
| Model 1              | 1.73 (8.32)                 | 8.35 (21.29)                   |
| Model 2              | 2.17 (7.94)                 | 9.05 (26.16)                   |
| Model 3              | 1.51 (8.68)                 | 7.94 (23.02)                   |
| Model 4              | 1.60 (5.55)                 | 8.68 (24.17)                   |
| Model 5              | 1.46 (4.63)                 | 8.08 (19.65)                   |

In contrast to interaction energies and interaction forces, the atomic partial charges, \( q_a \), with \( a \) indicating an atom in the whole system containing \( N_a \) atoms, were modelled using an \( L_2 \) loss:

\[ L_2^{\text{charges}} = \sum_a^{N_a} \left\| q^{NN}_a - q^{QM}_a \right\|^2. \]

Using a smooth \( L_1 \) function instead of the \( L_2 \) loss did not improve training. As can be seen, multiple values are treated in one NN. MAEs and RMSEs of 5 independently trained models on 9,500 data points are shown in Table S2.
Table S 2: Mean absolute error (MAE) and root mean squared error (RMSE) for charges of 5 models trained on 9,500 data points corresponding to Fig. 2 (e-f) in the main text.

| Neural Network Model | Charges MAE (RMSE) [a.u.] |
|----------------------|---------------------------|
| Model 1              | 0.026 (0.059)             |
| Model 2              | 0.027 (0.060)             |
| Model 3              | 0.026 (0.057)             |
| Model 4              | 0.027 (0.057)             |
| Model 5              | 0.027 (0.059)             |

S1.2 Data set generation
The initial data set was generated from a molecular dynamics (MD) simulation (pure molecular mechanics description) set up as described in the Methods section 5.3. From an 8-ns MD simulation, snapshots were selected every 4 ps. Water molecules with the oxygen atom further than 0.5 nm from the iron atom were discarded, yielding a total of 2,000 configurations, for which single point QM calculations of the inner+buffer and for the buffer region alone were performed using the SMD variant of the integral-equation-formalism polarizable continuum model, and as further described in the Methods section 5.1. Configurations were compared by energies, forces and spin contamination and obvious outliers were discarded leading to an initial set of 1,734 configurations. After training on the initial training set, short MD simulations and optimizations with the BuRNN approach were performed on a system containing inner, buffer, and outer regions. These simulations did not lead to stable hexacoordinated water configurations, but another 100 configurations were generated. To enhance accuracy close to the minimum, 21 lowest energy configurations from the BuRNN simulation were extracted and minimized on the QM level. From these minimization trajectories, 3,953 new configurations were extracted. The initial model was retrained on the updated data set, which led to much more stable BuRNN simulations, but we still observed irregular behaviour in the coordination sphere and the boundary region and large model deviations during simulations. Therefore, postprocessed all configurations once more by a) applying SHAKE to ensuring that the conformations of the water molecules adhered to the SPC configuration; b) remove any water molecules that in the QM optimization moved outside of the 0.5 nm cutoff; c) removed the implicit solvent model. The latter modification is crucial as the buffer region in BuRNN interacts explicitly with the water in the outer region and the effect of an implicit solvent leads to double counting of the interaction of solvent molecules. The NNs that were trained after this step led to stable simulations in the BuRNN approach, with differences between two independent NNs on the order of 10 kJ/mol. We next generated one additional set of conformations a) by picking the conformations for which the disagreement between the two independent NNs was largest during a BuRNN simulation (1,082 configurations), b) by running a BuRNN MD simulation with applying a biasing potential derived from the disagreement between the two independent NNs to sample the undertrained conformational space as

\[ V_{bias} = \begin{cases} 
K_{bias} \left[ (V_{NN}^{I+\Delta B} - V_{NN, val}^{I+\Delta B})^2 - V_{val, thresh}^2 \right], & V_{NN}^{I+\Delta B} - V_{NN, val}^{I+\Delta B} > V_{val, thresh} \\
0, & V_{NN}^{I+\Delta B} - V_{NN, val}^{I+\Delta B} \leq V_{val, thresh} 
\end{cases} \]

where \( K_{bias} \) was set to \(-0.01 \text{ kJ}^3 \text{ mol}^3 \), \( V_{val, thresh} \) to 1.0 kJ/mol, and \( V_{NN, val}^{I+\Delta B} \) is a predicted energy from the validation model (499 configurations), c) selecting conformations from a simulation in which one water molecule was artificially pulled away from the Fe\(^{3+}\) (300 configurations). This was done to ensure that the NNs learned energies and partial charges with a water molecule between the first and second solvation shell as well. Overall, the training set contained 11,430 configurations.
S2 BuRNN simulations

S2.1 BuRNN forces

Figure S1 shows the profiles of the net forces on water molecules in panel a and their radial components in panel d. The net forces are smooth at the boundary (0.5 nm), which can be attributed to the fact that at the boundary the interaction with the inner region is small in comparison to the rest of the system. Contrary, the atom-wise contributions (panels c and f) exhibit jumps at the boundary, reflecting the transition from an electrostatic interaction based on localized point charges to a quantum mechanical interaction, mediated by delocalized electron densities. For complete molecules, these jumps disappear, see also panels b and e in which the contribution of the NN to the forces is displayed. Within the buffer region, the forces reflect the observations in the radial distribution functions.

Figure S1. Forces as a function of the distance to the iron atom (dots). The line represents the median and the band corresponds to the 2\textsuperscript{nd} and 3\textsuperscript{rd} quartile. Net forces on water molecules (a), contribution of the NN treatment of the inner and buffer region ($F_{NN}^{i+\Delta B} + F_{MM}^{i+\Delta B} + p_{i+\Delta B}^{MM}$) to the net forces on the water molecules (b) and on the oxygen (red) and hydrogen (green) atoms (c). Panels d, e and f show radial components of the respective forces. Negative value means attraction.
S2.2 Observed structure

Figure S2: Iron-oxygen radial distribution function for different simulations settings. The blue curve shows the radial distribution function for a purely classical simulation (MM only), the orange one for a regular BuRNN simulation, in which the partial charges for the inner and buffer region were determined from the QM calculations in vacuum. The black dashed curve shows the radial distribution function for a simulation in which partial charges for the inner and buffer region were estimated from QM calculations in an implicit solvent. The red dotted curve shows the radial distribution function for a simulation in which the partial charges for the inner and buffer regions were assigned according to the classical force field.

Figure S2 shows the radial distribution functions for a simulation with BuRNN, with pure MM and for BuRNN with different charge distributions on the inner and buffer regions in an overlay. During the 10-ns BuRNN simulation the six coordinating water molecules did not exchange spontaneously. In contrast, out of the total number of 786 water molecules, all visit the buffer region at least once during the 10-ns simulation. In total, 8,847 distinct visits to the buffer region were monitored (not counting the six coordinating water molecules), with an average lifetime of 14.4 ps. 8,579 visits had a lifetime of 5 ps or less; 268 visits to the buffer region had lifetimes longer than 5 ps, up to a maximum of 2,240 ps.

To assess differences in pure MM and BuRNN simulations, Fig. 3 in the main text summarizes the geometries that are observed of the hexa-aqua iron complex. Based on the radial distribution functions, any Fe-O distance shorter than 0.25 nm was considered to be a coordinative bond. Panel a shows that both BuRNN and the pure MM simulations largely maintain a hexa-coordinated conformation. While in the MM description, a very small number of larger clusters appears, in BuRNN, few individual water molecules occasionally vibrate beyond the 0.25 nm cutoff, leading to coordination numbers smaller than 6. No complete exchanges of the coordinating water molecules were observed in these simulations. As illustrated in panel b, the Fe-O distance of the coordinating waters fluctuates around 0.209 nm, that agrees well with the range of experimental estimates of 0.199 to 0.210 nm.

The distributions of distances in are quite comparable between BuRNN and MM only simulations, but the dynamics of the Fe-O bond is rather distinct. Panel c illustrates the power spectrum calculated as the fourier transform of the autocorrelation function of the Fe-O bond lengths. Simulations using timesteps of 0.5 fs and of 2 fs show that the frequencies by which the Fe-O bonds vibrate are very
different between BuRNN and the MM only, implying that the Fe-O interaction is indeed not well captured by a purely non-bonded description. At the quantum level the vibrations take place at lower frequencies and are in good agreement with experiment. For both sets of simulations, the fastest vibrational motion has a period of 50 to 60 fs, confirming that a timestep of 2 fs is suitable. The distribution of the O-Fe-O angles in Fig. 3 d is quite comparable for both the BuRNN and the pure MM simulations. The expected 80:20 relation to angles of 90° and 180°, corresponding to an octahedral arrangement is observed in both cases.

Larger differences between BuRNN and pure MM simulations are observed in the distribution of the improper dihedral angle, which describes the co-planarity of the Fe³⁺ and a water molecule. As shown in Figure 2 c in the main text, the pure MM simulation leads to a distribution that is centered around 0°, while BuRNN leads to a much wider distribution with similar occurrences in the range from 0° to 25°. A value of 0° for this improper dihedral angle implies that the water molecule and the Fe³⁺ ion are in the same plane, while a purely tetrahedral arrangement around the O corresponds to a value of 35.26°. In the MM description, the hydrogen atoms of individual coordinating water molecules repel each other, leading to a mostly planar orientation of the individual water molecules. However, by describing the Fe-H₂O interaction at the quantum level of theory, the Fe prefers to interact with the lone pairs on the oxygen, more often leading to a tetrahedral arrangement. For comparison, a BP86-D3/def2-TZVP/COSMO estimate for this improper dihedral angle in [Fe(H₂O)₆]³⁺ lies at 16°.

S2.3 BuRNN is robust with respect to external perturbations

During BuRNN simulations at elevated temperatures, the disagreement between two NNs was monitored. The difference between the two NNS remained at an average value of $-0.39 \pm 0.02$ kJ/mol with a standard deviation that increased from 1.17 kJ/mol at 300 K to 1.45 kJ/mol at 400 K, which is still very small and stays below chemical accuracy of 1 kcal/mol.

![Image](figure_s3.png)

**Figure S3.** Iron-oxygen radial distribution functions for BuRNN simulations at different temperatures (left) and time series of the number of water molecules within 0.25 nm of the Fe³⁺ ion (with a small offset to visualize multiple simulations (right). Due to the size of the marker, the bars at a coordination of 6 seem continuous, but a small number of configurations is observed for which only 5 water molecules are within 0.25 nm.
Regarding simulations at which we pulled a water molecule away from Fe, we evaluated NN differences every 0.05 ps. The disagreement was found to be in the range of \(-0.68 \pm 0.10 \text{ kJ/mol}\) with a standard deviation of 1.47 kJ/mol, suggesting that the NNs are quite robustly sampling pentacoordinated iron complexes as well.

Movie S1: The file movie_S1.mp4 shows the spontaneous binding of a water molecule (cyan) to a pentacoordinated Fe\(^{3+}\)-water complex. The configuration switches from a trigonal bipyramidal arrangement to an octahedral complex in the process. Fe\(^{3+}\) as orange sphere, coordinating waters in sticks, buffer waters in blue, water in the outer region in grey.

References

1. K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, The Journal of Chemical Physics, 2018, 148, 241722.

2. K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko and K. R. Müller, Journal of Chemical Theory and Computation, 2019, 15, 448-455.

3. R. Caminiti and M. Magini, Chem. Phys. Lett., 1979, 61.

4. F. Böhm, V. Sharma, G. Schwaab and M. Havenith, Physical Chemistry Chemical Physics, 2015, 17, 19582-19591.

5. M. Radoñ, Physical Chemistry Chemical Physics, 2019, 21, 4854-4870.