The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics†

Kun Yao, John E. Herr, David W. Toth, Ryker Mckintyre and John Parkhill*

Traditional force fields cannot model chemical reactivity, and suffer from low generality without re-fitting. Neural network potentials promise to address these problems, offering energies and forces with near \textit{ab initio} accuracy at low cost. However a data-driven approach is naturally inefficient for long-range interatomic forces that have simple physical formulas. In this manuscript we construct a hybrid model chemistry consisting of a nearsighted neural network potential with screened long-range electrostatic and van der Waals physics. This trained potential, simply dubbed “TensorMol-0.1”, is offered in an open-source Python package capable of many of the simulation types commonly used to study chemistry: geometry optimizations, harmonic spectra, open or periodic molecular dynamics, Monte Carlo, and nudged elastic band calculations. We describe the robustness and speed of the package, demonstrating its millihartree accuracy and scalability to tens-of-thousands of atoms on ordinary laptops. We demonstrate the performance of the model by reproducing vibrational spectra, and simulating the molecular dynamics of a protein. Our comparisons with electronic structure theory and experimental data demonstrate that neural network molecular dynamics is poised to become an important tool for molecular simulation, lowering the resource barrier to simulating chemistry.

1 Introduction

Statistical models from machine learning experienced growing popularity in many areas of chemistry, such as in reducing the cost of simulating chemical systems, improving the accuracy of quantum methods, generating force field parameters, predicting molecular properties and designing new materials. Neural network model chemistries (NNMCs) are one of the most powerful methods among this class of models. They are shown to be capable of generating high quality potential energy surfaces (PESs) with different schemes such as summing over atoms or bonds, many-body expansions and permutation invariant polynomials. They are also used to predict the properties of materials, and even to find new drugs. In spite of their growing popularity, traditional force fields remain more popular than NNMCs, even for screening and reactive applications. This paper develops an open-source, transferable neural network model chemistry called TensorMol-0.1 (Fig. 1). This model hybridizes a NNMC with the physical contributions to molecular energies that are familiar from Molecular Mechanics and corrections to Density Functional Theory (DFT). This approach yields a predictable reproduction of physical long-range forces, and features a linear-scaling

1 Introduction

Statistical models from machine learning experienced growing popularity in many areas of chemistry, such as in reducing the cost of simulating chemical systems, improving the accuracy of quantum methods, generating force field parameters, predicting molecular properties and designing new materials. Neural network model chemistries (NNMCs) are one of the most powerful methods among this class of models. They are shown to be capable of generating high quality potential energy surfaces (PESs) with different schemes such as summing over atoms or bonds, many-body expansions and permutation invariant polynomials. They are also used to predict the properties of materials, and even to find new drugs. In spite of their growing popularity, traditional force fields remain more popular than NNMCs, even for screening and reactive applications. This paper develops an open-source, transferable neural network model chemistry called TensorMol-0.1 (Fig. 1). This model hybridizes a NNMC with the physical contributions to molecular energies that are familiar from Molecular Mechanics and corrections to Density Functional Theory (DFT). This approach yields a predictable reproduction of physical long-range forces, and features a linear-scaling

1 Introduction

Statistical models from machine learning experienced growing popularity in many areas of chemistry, such as in reducing the cost of simulating chemical systems, improving the accuracy of quantum methods, generating force field parameters, predicting molecular properties and designing new materials. Neural network model chemistries (NNMCs) are one of the most powerful methods among this class of models. They are shown to be capable of generating high quality potential energy surfaces (PESs) with different schemes such as summing over atoms or bonds, many-body expansions and permutation invariant polynomials. They are also used to predict the properties of materials, and even to find new drugs. In spite of their growing popularity, traditional force fields remain more popular than NNMCs, even for screening and reactive applications. This paper develops an open-source, transferable neural network model chemistry called TensorMol-0.1 (Fig. 1). This model hybridizes a NNMC with the physical contributions to molecular energies that are familiar from Molecular Mechanics and corrections to Density Functional Theory (DFT). This approach yields a predictable reproduction of physical long-range forces, and features a linear-scaling

Fig. 1 A schematic graph of TensorMol-0.1. Each element has its own charge network and energy network. The charge network predicts the atomic charges that yield the \textit{ab initio} dipole moment. An atom index matrix is used to reassemble the molecular energies/dipoles from atom energies/charges. The Behler–Parinello type energy network produces a short-range embedded atomic energy, which is summed with the electrostatic energy and van der Waals energy to predict the total atomization energy of molecules at and away from equilibrium.

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7sc04934j
inductive charge model. The charges do not depend on a quadratic-scaling polarization equation like a Thole-model," instead they are not fixed and respond well to geometry changes as we will show with reproduction of IR spectra.

Our group is one of several who have been pursuing transferable and black-box neural network model chemistries.16,46,50,54,82,83 The state of the art in this field is progressing rapidly. Readers may not appreciate that a model can achieve chemical accuracy for energies but have useless noisy forces. Models that provide energies at equilibrium, and those treating a fixed molecule or stoichiometry, are nowadays reliably produced.39 We will show that TensorMol-0.1 yields usefully accurate predictions of forces out-of-equilibrium by showing the reproduction of infrared spectra that closely approximate our source model chemistry (omega97X-D, 6-311G**) and molecular dynamics. We outline several tricks that are required to ensure the stability of long-time molecular dynamics.

This force model is implemented in an open-source package that uses the TensorFlow tensor algebra system to compute descriptors and forces. The methodology can be used to propagate the dynamics of large molecules (10^5 atoms) on simple laptop computers. No force field refinement, atom assignment, or other interventions are needed to apply the method to a molecule of interest, so long as the elements are supported. The package is also interfaced with the I-PI path integral package45 to allow for quantum simulations and enhanced sampling.

2 Methods

The community of neural network model chemistry developers is rapidly improving the accuracy and generality of these reactive force fields.4,48,49,51,52,82,86 The model developed in this paper includes several components that were the subject of recent developments in other groups.50,54,57,82,85 We will describe the details here from the bottom up, citing prior studies. Our notational convention will be that i, j, k... are the indices of atoms, q, is the charge on atom i, x, y and z are atomic numbers, A, B and C are molecules, and α, β... are the indices of basis functions which are the products of radial and angular functions. If a function depends on all the atomic coordinates of a molecule it will be written as a vector, and those functions which depend on only a few coordinates will be given explicit indices. The energy of TensorMol-0.1 is expressed as a sum of a short-range embedded n-body potential,46 long-range electrostatic potential and van der Waals force:

\[
E(\mathbf{R}) = \sum_i E_{\text{BF}}^i \left( S_a(\mathbf{R}_i) \right) + \sum_{i<j} E_{\text{DSF}}^{ij} \left( q_i S_\alpha(\mathbf{R}_i), R_i, R_j \right) + E_{\text{VDW}}(\mathbf{R}_{ij})
\] (1)

In the above expression \(E_a\) is a Behler–Parinello type energy network for the element \(z\) for atom \(i\). This \(n\)-body potential takes as its argument \(S_a\), the modified symmetry functions of Isayev and coworkers.42

\[S_a(\text{radial}) = \sum_{j \neq i} e^{-r(R_i-R_j)^2} f_i(R_j)
\] (2)

\[S_a(\text{angular}) = 2^{1-\zeta} \sum_{j \neq i, k \neq i} (1 + \cos(\theta_{ijk} - \theta_i))^\zeta \times e^{-\frac{(R_{ik} + R_{ik} - R_i)}{2}} f_j(R_{ij}) f_k(R_{ik})
\] (3)

Modern machine learning frameworks provide automatic differentiation of tensor algebraic expressions, allowing a force field developer to obtain the gradient of a molecular potential \(\frac{dE(\mathbf{R})}{d\mathbf{R}}\) in a single line of code, once the expression for \(E(\mathbf{R})\) has been written. An important feature of our code is that this symmetry function is coded within the TensorFlow system,88 so all the parameters of this descriptor can be optimized alongside the network weights to minimize error. Our implementation of the symmetry function employs a list of nearest-pairs and triples within radial cutoffs such that the scaling of the overall network is asymptotically linear. On an ordinary laptop equipped with only a CPU, a force/energy call on 20 000 atoms takes less than a minute.

The second term of our energy expression is the damped-shifted force (DSF) Coulomb energy of Gezelter and coworkers.89 The charges are obtained from a sub-network which reproduces molecular dipole moments. Previous studies90,91 included electrostatic energy by training the networks to learn Hirshfeld charges. Our charge model enforces conservation of total charge by evenly spreading any required neutralizing charge over the entire molecule or unit cell. The damped-shifted force ensures the long range continuity and differentiability of the effective Coulomb potential with smooth cutoffs. We modify the DSF kernel at short range with an “elu” type non-linearity,92 such that the forces within the radius of the Behler–Parinello symmetry function smoothly approach zero, avoiding singularities and interference with the Behler–Parinello many-body potential. The range separation concept has a long history in chemistry whenever two models of a physical force have complementary cost or accuracy range advantages.42

The energy of the DSF kernel is expressed as:

\[
\begin{cases}
E_{\text{DSF}} = E_{\text{DSF(original)}} & R > R_{\text{switch}} \\
q_i q_j (\alpha_{\text{elu}}e^{R_{\text{switch}}-R_{ij}} + \beta_{\text{elu}}) & R < R_{\text{switch}}
\end{cases}
\] (4)

where \(E_{\text{DSF(original)}}\) is the energy of the DSF kernel92 and \(R_{\text{switch}}\) is the short range cutoff for the “elu” kernel. \(\alpha_{\text{elu}}\) and \(\beta_{\text{elu}}\) are chosen so that the value and the gradient of \(E_{\text{DSF}}\) are continuous at \(R_{\text{switch}}\). The damped-shifted force is well-suited to being combined with neural network models because it requires no Fourier transformation to treat periodic systems with linear scaling, and maps well onto TensorFlow. The last term is the van der Waals energy, which is calculated by following Grimme’s C6 scheme.88

We employed a two step training approach. First, the charge networks are trained to learn the atom charges that predict the dipole moment. The loss function can be written as follows:
After the charge training is converged, we train the energy network. During the energy network training, the weights in charge networks are kept frozen, but they are still evaluated to calculate the electrostatic energy that is added to construct the total energy. Our Behler–Parinello many-body potential also absorbs the shape of the transition between the many-body and electrostatic regions. The learning target for the energy network includes both the DFT energy and DFT force. The loss function for the energy network training is:

$$ L_{\text{energy}} = \sum_{A} \left( \frac{E_{A}^{\text{DFT}} - E_{A}^{\text{NN}}}{N_{\text{atom}}} \right)^2 + \gamma \sum_{A} \left( \frac{F_{A}^{\text{DFT}} - F_{A}^{\text{NN}}}{N_{\text{atom}}} \right)^2 $$

where $E_{A}^{\text{NN}}$ is obtained according to eqn (1) and $F_{A}^{\text{NN}}$ is calculated by taking the gradient of $E_{A}^{\text{NN}}$ with respect to the coordinates of the atoms. $N_{\text{atom}}$ is the number of the atoms in the system and $\gamma$ is a parameter that controls the portion of force loss. We employ $\gamma = 0.05$. We trained two neural networks based on two sets of data (see Table 1). One network (the “water network”) was trained on a database that includes $\sim 370\,000$ water clusters with 1 to 21 water molecules. The other network was trained on $\sim 3\,000\,000$ different geometries of $\sim 15\,000$ different molecules that only contain C, H, O and N and up to 35 atoms. Since these 15 000 molecules were sampled randomly from the Chemspider database, we will refer to this network as the “Chemspider network” in the following text. The training geometries were sampled with metadynamics\(^\text{**}\) and their energies calculated using the QChem\(^\text{**}\) program and oB97X-D\(^\text{*}\) exchange correlation functional and a 6-311G\(^*\) basis set.

Each charge network and energy network contains three fully-connected hidden layers with 500 hidden neurons in each layer. For the Chemspider network, a network with three hidden layers, with 2000 hidden neurons in each layer, is used for each charge network and energy network. L2 regularization is used for both networks and dropout\(^\text{**}\) on the last layer was used for the Chemspider network to prevent overfitting, with a dropout probability of 0.3. We chose a soft-plus function $\ln(1.0 + e^{x})/\alpha$ with $\alpha = 100$ as the non-linear activation function and used the Adaptive moment solver (Adam)\(^\text{**}\) to fix the weights of the network. The test sets were separated from the training data by randomly choosing 20% of the molecules at the outset, which were kept independent throughout. Besides water, we will present calculations from molecules strictly absent from either the training or test set.

To obtain linear scaling, TensorMol uses atom neighbor lists within cutoffs. This allows double precision energy, charge and force calculations of up to 24 000 atoms to be executed in less than 90 seconds on a 2015 Intel i7 2.5 GHz MacBook Pro (Fig. 2). Periodic evaluations are achieved by tessellation of a unit cell with summation of the energies of atoms within the cell. Periodic calculations require about three times more wall time to execute. Speedups greater than a factor of two are obtained automatically when using computers with GPUs (Fig. S7†) or single-precision calculations.

### 3 Results

The root mean square error (RMSE) on the independent test set of the energy is 0.054 kcal mol$^{-1}$ atom$^{-1}$ and the RMSE of the force is 0.49 kcal mol$^{-1}$ Å$^{-1}$. The top panel of Fig. 3 is a plot of the potential energy surface (PES) of a water trimer when one of the water molecules is pulled away from the other two. One can see our neural network PES is not only in good agreement with the PES of the target method but is also smooth. To achieve this we use a variation of the soft-plus neuron rather than the rectified linear units that are popular in computer science. The latter train more efficiently, but produce discontinuous forces.

The bottom panel shows the fractional contribution of each of the three energy components in eqn (1) to the binding energy along the trimer dissociation coordinate. At short range, most of the binding energy is contributed by the $n$-body neural network potential. When the distance between the monomer and the dimer approaches the cutoff distance of the neural network, the contribution of the neural network potential starts to decrease and the contribution of the electrostatic potential.

<table>
<thead>
<tr>
<th></th>
<th>Water network</th>
<th>Chemspider network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of training cases</td>
<td>370 844</td>
<td>2 979 162</td>
</tr>
<tr>
<td>Training time (days)$^a$</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>Energy RMSE</td>
<td>0.054</td>
<td>0.24</td>
</tr>
<tr>
<td>Gradient RMSE</td>
<td>0.49</td>
<td>2.4</td>
</tr>
<tr>
<td>Dipole RMSE</td>
<td>0.0082</td>
<td>0.024</td>
</tr>
</tbody>
</table>

$^a$ Training was done on a single Nvidia K40 GPU.
increases. After 6 Å, where the neural network symmetry functions on the atoms in the monomer have no contribution from the dimer, the neural network force drops smoothly to zero and the electrostatic interaction dominates. The small difference in the energy at 7 Å is due to the difference between the Madelung energy given by the learned charges, and the genuine physical cohesive forces at this distance. The dimer and monomer are beyond the symmetry function sensory radius, and so the charges are constant in this region. Future iterations of the charge network will use local-field information to improve this region of the PES. The learned inductive charges are of high quality considering their linear scaling cost. Fig. 4 shows the PES and dipole change of a water dimer when the hydrogen bond is broken by rotating the OH bond. Both the PES and dipole change fit well with the DFT results.

Given the increased dimensions of the Hessian, it is naturally a more stringent test to reproduce forces and infrared spectra than it is to simply produce energies. The top panel and bottom panel of Fig. 5 show the optimized geometries and IR spectra of a 10 water cluster and a 20 water cluster, respectively, generated using our force field and DFT. Each method uses its own equilibrium geometry, so this also tests the ability of TensorMol-0.1 to reproduce non-covalent geometries. Our force field quantitatively reproduces the IR spectra generated using DFT, both in terms of frequencies and intensities, especially for the water bend modes and inter-monomer modes. The Mean Absolute Error (MAE) of the frequencies in those two regions is
33.2 cm\(^{-1}\) for the 10 water cluster and 16.2 cm\(^{-1}\) for the 20 water cluster. The error is slightly larger in the water OH stretching region with a MAE of 34.2 cm\(^{-1}\) and 13.1 cm\(^{-1}\) for the 10 and 20 water clusters, respectively. This accuracy is comparable to high quality polarizable water force fields.\(^a\)

Compared with traditional force fields, one major advantage of TensorMol is its reactivity. TensorMol is able to simulate a concerted proton transfer in a water hexamer, finding a minimum energy transition path. The PES calculated using a nudged elastic band (NEB) method\(^\) with the TensorMol force field and the PES calculated using DFT are shown in Fig. 6. The barrier height predicted by TensorMol is 29.7 kcal mol\(^{-1}\), which is 6.7 kcal mol\(^{-1}\) lower than the prediction from DFT, which is remarkable considering the dearth of transition structures in the training data. Our method of sampling molecular geometries uses a meta-dynamics procedure described elsewhere,\(^9\) so these proton transfers do occur in the training data although extremely infrequently.

Encouraged by our results from studying water, we developed a force field with applicability across the chemical space spanned by C, N, O and H. The Chemsider dataset that we used to train our force field covers a vast area of chemical space containing 15 thousand different molecules and 3 million geometries. The geometries are generated using a meta-dynamics procedure,\(^{10}\) which ensures that each new geometry is a fresh part of chemical space; energies up to 400\(k_bT\) are sampled in the data. We describe the details of this meta-dynamics sampling algorithm, which we have found vital to achieving robust and transferable force fields elsewhere.\(^{11}\) The diversity of structures makes learning the Chemsider dataset a much harder task for the TensorMol-0.1 network; the test set RMSE of energy is 0.24 kcal mol\(^{-1}\) atom\(^{-1}\) and the RMSE of force is 2.4 kcal mol\(^{-1}\) atom\(^{-1}\). More importantly, the model usefully reproduces several elements of molecular structures, at and away from equilibrium, for molecules outside its training set. It robustly optimizes the geometries of typical organic molecules to structures that match DFT well, and yields infrared frequencies and intensities that are in good agreement with \textit{ab initio} calculations. It is a black-box method that does not rely on any specific atom type, connectivity, etc., which one would need to specify in a traditional classical force field. The few proteins we have examined remain stable and near their experimental structures when optimized or propagated at room temperature using the TensorMol-0.1 force field.

Morphine was not included in our training set. The top right panel of Fig. 7 shows the geometry of morphine that is optimized with our force field. The RMSE of the bond lengths predicted by our force field is 0.0067 Å and the RMSE of the angles is 1.04 degrees, compared with the source DFT model chemistry. The upper left panel plots the harmonic IR spectra generated using DFT and the TensorMol force field, at their respective optimized geometries. One can see the IR spectrum generated using our force field is in good agreement with the DFT-generated IR spectrum. The MAE in our force field frequencies is 13.7 cm\(^{-1}\) compared with the DFT frequencies, which is about half of the MAE in the prediction using MMFF94\(^\dagger\) (Fig. S3\(^\dagger\)). Fig. 8 shows comparisons of the IR spectra that are generated using these two methods for aspirin, tyrosine, caffeine and cholesterol. All four of these molecules were not included in the training set. The MAE in the frequencies predicted by our field is less than 20 cm\(^{-1}\) for all four molecules, compared with the target DFT frequencies. As Fig. S4\(^\dagger\) shows, the MAE in the frequencies calculated using MMFF94 are 2 to 3 times larger than the MAE in the frequencies calculated using our force field for these four molecules. The intensities calculated using MMFF94 are also qualitatively different to the DFT intensities. The concept of a chemical bond and force constant are not enforced in any way, yet good agreement with DFT is obtained at a tiny fraction of the original cost.

Traditional harmonic vibrational spectra require quadratic computational effort, which works against the speed advantage of a NNMC. For large systems, one can use the molecular dynamics functionality of TensorMol to simulate infrared

Fig. 6 The reaction energy profile converged from a nudged elastic band method along the reaction coordinate of conservative proton transfer in a water hexamer cluster. The reaction coordination is defined as \(R_{\text{OH}} - R_{\text{In}}\)\((R_{\text{final}} - R_{\text{In}})\).

![Fig. 7](image_url)
spectra, Fourier transforming the dipole–dipole correlation function of conservative Newtonian dynamics, whose cost grows linearly with the size of the system. The lower left panel of Fig. 7 shows the infrared spectrum produced by propagation in TensorMol-0.1, also showcasing the good energy conservation of TensorMol. Unlike when using a traditional force field, in this case it is non-trivial to obtain smoothly differentiable NNMCs. 64-bit precision needs to be used as the network cannot be made too flexible, and smooth versions of the typical rectified linear units need to be used. Our package can be used in this way to simulate IR spectra of large systems with linear cost.

TensorMol-0.1 uses a relatively simple treatment of the electrostatic and van der Waals forces, which we would like to augment in the future with a many-body dispersion scheme. However, a main advantage of the approach used by TensorMol-0.1 is that no self-consistent polarization equation is solved even though the charges are inductive, which results in linear scaling and ease of inexpensively calculating the electrostatic energies of even very large molecules. At shorter ranges, non-covalent interactions like hydrogen bonds are dealt with by the Behler–Parinello portion of the network. The Chemspider training data include some examples of dimers and intra-molecular hydrogen bonds. To our surprise, the treatment of the inter-molecular interactions that were not targets for TensorMol-0.1 are satisfactory. Fig. 9 shows the optimized geometries and binding energies of two DNA base pairs calculated using our force field. The target DFT method predicts a binding energy of 18.3 kcal mol$^{-1}$ for the thymine–adenine (TA) pair and a binding energy of 32.4 kcal mol$^{-1}$ for the guanine–cytosine (GC) pair. The prediction using our force field is 1.2 kcal mol$^{-1}$ smaller for the TA pair and 2.0 kcal mol$^{-1}$ larger for the GC pair relative to DFT.

One holy grail in the field of neural network model chemistries is to simulate biological chemistry without QM-MM or bespoke force fields. Protein simulation also demonstrates several important features of neural network model chemistry: reasonable inter-molecular forces, stability, scalability and generalization far from small-molecule training data. TensorMol-0.1 was not trained on any peptide polymers and includes no biological data of any sort. To our pleasant surprise, even this first iteration of neural network model chemistry is accurate enough to perform rudimentary studies on small proteins. Fig. 10 shows geometries sampled from a 1 picosecond, periodic, 300 K TensorMol dynamics NVT trajectory in explicit solvent. The initial structure (included in the supplementary information) was generated from the PDB structure 2MZX using OpenMM’s automatic solvation and hydrogenation scripts, but includes nothing but atom coordinates. This short alpha-helix is stable, both in optimizations and dynamics, and the structures sampled during the dynamics simulation superficially resemble the solution NMR structure. Traditional force fields will always be less expensive (by some prefactor) than NNMCs, yet the reactivity advantages of NNMCs and the ease of set up will probably lead to a rapid adoption of these methods in the biological community.
4 Discussion and conclusions

We have presented a transferable neural network model chemistry, TensorMol-0.1, with long-range coulombic physics and a short-range n-body potential. The model is integrated in an open-source Python package, which provides many of the types of simulation commonly used in chemistry. The method can be used to scan conformational and chemical space along the singlet neutral potential energy surface with high throughput using bare atomic coordinates.

TensorMol-0.1 is not the final iteration of a neural network model chemistry, although it shows that DFT-quality predictions can be made by models with five orders of magnitude lower cost. Inexpensive post-DFT corrections such as many-body dispersion\(^{100}\) will become even more powerful when integrated with these potentials, opening the door to quantitative treatments of large systems. NNMCs may compete strongly with DFT packages, and provide an interesting complement to QM-MM-type simulations in the near future.

There are several clear paths to extend this work:

- generalize the descriptors to encode other physical atom properties besides charge (spin or polarizability)
- develop accurate descriptors whose cost grows linearly with the number of elements treated
- extend the range of the n-body embedding
- explore the hierarchy of physical detail between force fields and semi-empirical electronic structures

These are the directions of continuing study in our group and others.

Conflicts of interest

There are no conflicts to declare.

References


