The Journal of Chemical Physics

RESEARCH ARTICLE | MARCH 25 2024

Toward transferable empirical valence bonds: Making classical force fields reactive

Alice E. A. Allen ; Gábor Csányi

J. Chem. Phys. 160, 124108 (2024)
https://doi.org/10.1063/5.0196952

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Toward transferable empirical valence bonds: Making classical force fields reactive

Cite as: J. Chem. Phys. 160, 124108 (2024); doi: 10.1063/5.0196952
Submitted: 10 January 2024 • Accepted: 7 March 2024 •
Published Online: 25 March 2024

Alice E. A. Allen[1,2,a)] and Gábor Csányi[3]

AFFILIATIONS

[1] Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
[2] Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
[3] Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

[a)] **Author to whom correspondence should be addressed:** aallen@lanl.gov

ABSTRACT

The empirical valence bond technique allows classical force fields to model reactive processes. However, parametrization from experimental data or quantum mechanical calculations is required for each reaction present in the simulation. We show that the parameters present in the empirical valence bond method can be predicted using a neural network model and the SMILES strings describing a reaction. This removes the need for quantum calculations in the parametrization of the empirical valence bond technique. In doing so, we have taken the first steps toward defining a new procedure for enabling reactive atomistic simulations. This procedure would allow researchers to use existing classical force fields for reactive simulations, without performing additional quantum mechanical calculations.

I. INTRODUCTION

Classical force fields describe the potential energy of a molecular system with a well-defined, simplistic functional form.[1–4] This enables classical force fields to be used for large and long timescale simulations.[5–7] However, the simplistic functional form of classical force fields does not allow for the breaking and forming of new bonds. This means that reactive atomistic simulations must rely on more complex and slower techniques, such as the bond order-based force field Reaxff, machine learning potentials, and semi-empirical methods, such as Density Functional based Tight Binding (DFTB) or *ab initio* simulations.[8–15] Methods were developed in the 1980s to enable classical force fields to be used for reactive simulations with empirical valence bond (EVB) introduced in Ref. 12. EVB can describe a reaction path by redefining the energy of a system as a combination of the energy from the reactants force field, the product's force field, and a coupling term. The application of this technique has been wide ranging, from describing proton mobility in aqueous environments to proton transport across transmembrane protein channels.[10,16] However, while classical force fields are transferable and can be used "out of the box," the coupling terms for empirical valence bonds must be calculated for each specific system of interest.[10,16] This means that it is not a transferable technique for reactive simulations. By creating a predictive model for the parameters used in the empirical valence bond method, this paper takes the first steps toward developing a transferable form of empirical valence bonds that would enable classical force fields to be used for reactive simulations without system-specific parametrization to quantum mechanical data.

Using machine learning models (ML) to predict the parameters of semi-empirical methods has been previously carried out in several works.[17,18] However, the advantage of using EVB is that we can still use the underlying classical force fields as the primary model. With EVB, classical force fields that have been carefully parameterized to recreate experimental properties can describe reactive simulations without significant alteration. In addition, there is very little additional cost involved when EVB is combined with classical force fields. If a machine learning model could be produced that could accurately learn the parameters of EVB across chemical reaction space, then classical force fields could be readily used for reactive simulation with no additional complex parametrization steps involved, little change to the equilibrium potential energy surface compared to the classical force field, and only a small increase in computational speed.

Creating a transferable form of EVB is a multi-stage process. First, the coupling parameters need to be automatically calculated

for thousands of reaction paths. To do this, the EVB model was implemented inside the atomic simulation environment (ASE) software with AMBER used as the underlying classical force field.[19,20] The minimum energy path was then estimated with ASE using nudged elastic band (NEB) and the dimer method.[19,21,22] The coupling parameters were then optimized to recreate the quantum mechanical (QM) barrier height for the reaction, with the reaction paths and barrier heights taken from Ref. 23. With a database of these coupling terms produced, a machine learning model was created to learn the coupling terms. A message-passing neural network using the SMILEs strings as features was constructed to do this.[24] An alternative route for removing QM calculations from the parametrization process is to directly predict activation energies using a machine learning model. This was also attempted and shown to excellently recreate the QM activation energies, albeit with optimization of the EVB terms still required in the workflow.

This paper contains a number of new contributions. First, carrying out automated parametrization of EVB for large datasets of reaction paths has not previously been performed. Therefore, we provide an analysis of the accuracy of the activation energies with EVB compared to the QM values. In addition, we have developed new methods and workflows for automated parametrization for EVB. This is alongside the creation of the machine learning model used to predict the parameters of EVB and the proposal of a new methodology for reactive potentials. The steps we have taken in this work are the first toward a new methodology for reactive simulations—one where unreactive classical potentials could readily be coupled together, without additional parametrization, to simulate reactive processes.

II. METHOD

A. Empirical valence bond

The formulas used in EVB are as follows.[12] The matrix Hamiltonian is given by

$$V(x) = \begin{bmatrix} E_A(x) & V_{AB} \\ V_{AB} & E_B(x) \end{bmatrix},$$

where V_A/V_B is the energy of the system with the reactant/product classical force field and V_{AB} is a function to be chosen.[12] The corresponding EVB energies are the solutions to $\det[V(x) - V_i(x)I] = 0$. Two solutions emerge describing the ground state (V_0) and excited state energies (V_1). In this work, we are interested only in the ground state energy given by

$$V_0(x) = \frac{1}{2}(E_A(x) + E_B(x)) - \frac{1}{2}(\Delta E(x)^2 + 4V_{AB}^2)^{1/2},$$

where $\Delta E(x) = E_A(x) - E_B(x)$.[12,25] An illustration of the empirical valence bond method with a harmonic potential and constant V_{AB} parameter is shown in Fig. 1. The empirical valence bond method allows a smooth transition between the two potentials. With EVB, a reaction can be described as a smooth transition between the potential energy surface for two distinct molecules. For example, a classical force field could be used to model butane, and a classical force field with different parameters could be used to model methylpropane. However, classical force fields would not allow a description of the transition between the two molecules and the breaking and forming of bonds. By using EVB, a smooth transition between the two potentials is possible and reactions and transition states can be simulated.

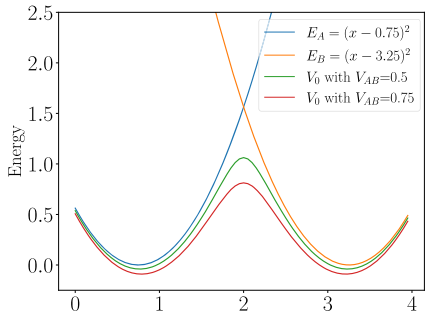

FIG. 1. An illustration of EVB for two simplistic harmonic energy curves.

Many different forms for the off-diagonal components have been proposed:

- The simplest form is V_{AB} = constant. However, if V_{AB} does not change across the potential energy surface, then the energies around the equilibrium structures will also be affected. This is apparent in Fig. 1 and is not desirable as the regions around the equilibrium structure will be well described by the classical model.
- Alternatively, far more complex models can be used such as the distributed Gaussian from Ref. 25. The advantage of this approach is that complex reaction paths can be recreated. However, it requires multiple QM calculations along the reaction path and is not suitable for our current purposes.
- The V_{AB} term can also be expressed as a function of the energy difference as seen in Ref. 26,

$$V_{AB} = Ae^{-B(E_A(x)-E_B(x))^2}. \quad (1)$$

This is the functional form that we will use in this work as it requires just two parameters and, as we will see, is often sufficient to exactly recreate the activation energy. Improving the functional form of V_{AB} is a direction for further work.

The classical force used as the underlying model in this work is AMBER parametrized with GAFF.[20,27] A calculator in the ASE code was then written to implement the EVB method with AMBER.[19,20] An additional approximation present in EVB is that off-diagonal element parametrized in the gas phase can be transferred to condensed phases simulation. This approximation was tested in Ref. 28 and shown to be valid for certain reaction classes. Testing this assumption further is outside the scope of this work but is an important consideration.

B. Minimum energy path

One of the most challenging aspects of creating a transferable form of EVB is automating the parametrization process for EVB. This requires the minimum energy path to be found and then parameters optimized to recreate the QM barrier height as closely as possible.

Finding the minimum energy path is a multi-stage process. The first step is to find an approximate solution to the path. One such approximation is to simply linearly interpolate between the reactant and product. However, the characteristics of the EVB can also be exploited for the proposed path. The method developed is as follows. Starting from the reactant, the molecule is minimized with respect to the sum of the energy of the $E_A + E_B$, and several images along this path are taken. Then, starting from the product, the molecule is minimized with respect to the sum of the energy. The two optimization paths are then combined together to give an approximation to the path. The method avoids any paths being proposed that have extremely large $E_A + E_B$ components and this is desirable as a large part of V_0 is the sum component. If the minimum energy path fails to converge with this linear path, then this EVB-specific path is attempted.

After an initial path is proposed, the nudged elastic band method (NEB) is then used to find an approximation to the minimum energy path.[22] There are a maximum of 500 optimization stages to NEB performed, with 20 images taken along the path. The spring constant is set to 0.100 and the improved tangent method from Ref. 29 is employed with the FIRE optimizer. If convergence (with the maximum force of 0.12 eV/Å) with these settings fails, the BFGS optimizer is instead used.

However, even with a converged NEB path, it is still possible to miss the transition state in the reaction path. Fortunately, the properties of EVB can again be exploited to try to identify this. An important point on the reaction path is the $E_A = E_B$ point where $\Delta E = 0$. There is no guarantee that this will be the transition state position, but in practice, it often is as the $\Delta E = 0$ point occurs at a high energy. Therefore, to ensure this point is correctly identified, the following steps occur:

- The two images on either side of the $\Delta E = 0$ energy are found.
- These two images are minimized with respect to ΔE to find the nearest $\Delta E = 0$ point, and the lowest energy image is identified.
- To find the saddle point, the dimer method is used with this image.[21] The dimer method is limited to 50 steps; this is to prevent points being identified that do not correspond to the transition state. If convergence fails after the first attempt, the process is repeated up to a maximum of three attempts with different displacement vectors used at each attempt.

The addition of the dimer method greatly improves the recreation of the barrier height and is an essential part of the process. An example of this is shown in Fig. 2 with the recreation of the barrier height improved by the inclusion of the dimer method.

C. Checking the transition state

It is important that the proposed transition state is sufficiently accurate. The transition state can be inaccurate due to convergence problems. There are two methods that we use to check the accuracy of the transition state. First, the maximum force for the structure proposed by the dimer method can be used. Large forces indicate that the true saddle point on the surface has not been found. Second, the minimum ΔE along the MEP can be calculated. Although

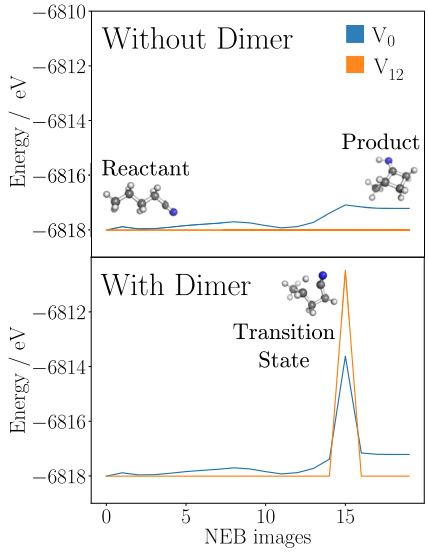

FIG. 2. A comparison of the MEP found with NEB only and using the dimer method alongside NEB. The transition state is more accurately recreated with the inclusion of the dimer method and this is reflected by the contribution from the V_{AB} component. Only a small contribution is seen when only the NEB method is used. The black dashed line shows the quantum mechanical barrier height. The V_{AB} component is shifted so that the zero points align with the energy of the first image.

it is not always the transition state point, the $\Delta E = 0$ point is often extremely close to the transition state. If the minimum energy path and dimer method fail to explore this low ΔE region entirely, it indicates that there could be problems with the accuracy of the path proposed. In addition, if the dimer method suggests a structure with atom distances that are greater than 2 Å from the original structure, the structure was not included (Fig. 3).

Both of these approaches were incorporated into the custom optimizer discussed below to ensure that the optimal parameters proposed were not the result of inaccurate MEP. Inaccuracy MEP are particularly problematic as it will reduce the performance of the ML model and provide an inaccurate test set.

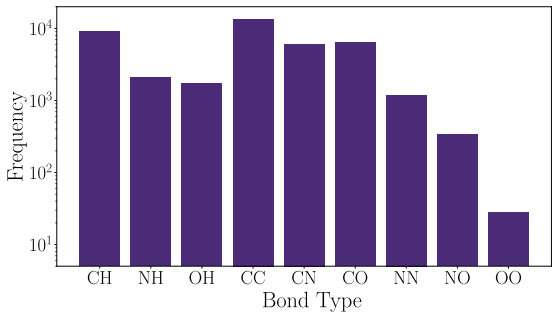

FIG. 3. The bond types in the reacting atoms identified by RDkit for the reaction dataset in Ref. 23.

D. Optimization

The optimization is performed with the following loss function:

$$Loss = |AE_{EVB} - AE_{QM}| + |S_m| + \lambda(|V_{AB}(x_A)| + |V_{AB}(x_B)| + |A| + |B|),$$

Where AE is the activation energy, λ is a constant set at 0.001, $V_AB(x_A)$ is the off-diagonal term for the first NEB image, and S_m is to prevent the appearance of new spurious minima,

$$S_m = \begin{cases} n, & \text{if } n < 0, \\ 0, & \text{otherwise,} \end{cases}$$

with $n = \min(E_{NEB}) - \min(V(x_A), V(x_B))$.

The final regularization terms for A and B push A and B toward zero if these terms do not influence the height of the activation barrier and also ensure that the loss function does not become flat if the A and B terms no longer influence the height. The off-diagonal elements A and B are the only terms optimized in this process, with the ΔH terms being obtained directly from the QM or ML model.

Optimizers obtained through the SciPy library were originally used for the optimization. However, standard optimizers were found to be prone to becoming stuck in local minima. This is because the optimization process can encounter issues, for example, with convergence of the dimer method. Additionally, the ability of NEB and the dimer method to converge is dependent on the parameters used. Therefore, a custom optimizer was produced, which used a method with a changeable step length based on the difference in the QM and EVB barrier height. The custom optimizer also takes into account that the parameter A is primarily responsible for the barrier height. Additionally, parameters that result in bad convergence of the dimer method are not included as possible solutions.

E. QM dataset

The dataset used to optimize the EVB parameters is taken from Ref. 23. This dataset contains energies for the reactants, products, and transition states for 11 961 different reactions. The reactions were sampled with the single-ended growing string method that allows for just a reactant and driving coordinates to be specified.[30] The structures were then calculated at the ωB97X-D3/def2-TZVP level of theory (all in the singlet state) with QChem.[31] Due to the sampling method used, the dataset does not describe just one reaction class but a variety of different reaction classes as discussed in Ref. 23.

F. Machine learning model

In Ref. 24, a message-passing neural network model was constructed that took SMILES strings as an input and then predicted activation energies—this was an extension of the model proposed in Ref. 32 and allowed for reaction properties to be predicted. To do this, fingerprints were constructed from the reactants and products, and the difference between these two fingerprints was then calculated. A directed message-passing neural network model then used these fingerprints to predict the activation energies.

Given its previous success, we have taken this existing machine-learning architecture and applied it to the prediction of EVB parameters. In this work, several properties were predicted with this model, AE, ΔH, A, and B—where the difference in energy between the reactant and product is defined as ΔH. As in Ref. 24, we used joint fitting with AE and ΔH fit together and AE, ΔH and A fit together. The model was trained for 250 epochs with a batch size of 50.

III. RESULTS

A. Performance of EVB across the dataset

Before we investigate the possibility of a transferable form of EVB, we will begin by analyzing how well EVB can recreate QM activation energies and examining the performance of our proposed workflow.

Of the 11 961 reactions investigated, the EVB parameters can be accurately calculated for 9799 of the reactions. The remaining cases have problems converging either due to NEB or the dimer method, with 934 failing due to NEB convergence and 1227 failing due to the dimer method. Failure to converge is both a consequence of limitations in the algorithms used as well as limitations in the force field used. Of those reaction paths that successfully converge, 9630 have an error under 0.05 eV, while 169 have an error above 0.05 eV. This is equivalent to 98% of cases accurately recreating the QM barrier. Figure 4(a) shows the ability of the EVB technique to recreate the QM activation energy.

The distribution of the A and B parameters is shown in Figs. 4(b) and 4(c). The mean value of A is 5.36, and 94.4% of the values are less than 10. For the B parameter, for 69.3% of paths, 1.0 is sufficient in the optimization process.

B. Constant A, B

The mean value of A = 5.36 and the mode value of B = 1.0 were selected as a constant EVB model to compare more sophisticated approaches against. The results are shown in Fig. 7(a) for all of

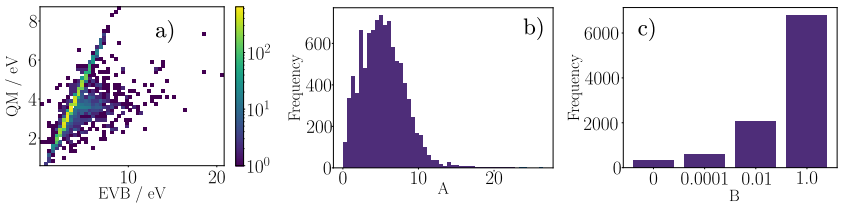

FIG. 4. The (a) recreation of the QM activation energy by the EVB method, (b) distribution of the A parameter, and (c) the distribution of the B parameter.

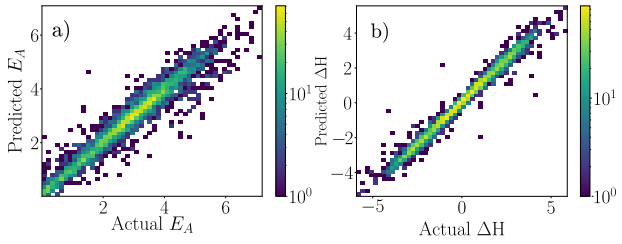

FIG. 5. The (a) recreation *AE* by the NN model and (b) the recreation of the Δ*H* by the NN model. There are 3918 reactions in the test set. The RMSE is 0.35 and 0.27 eV for *AE* and Δ*H*, respectively.

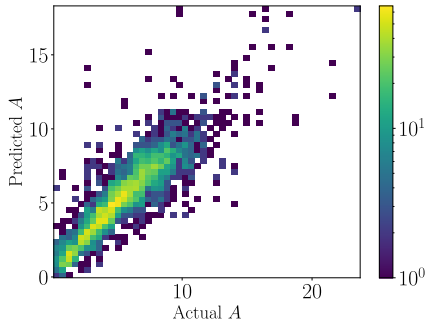

FIG. 6. The recreation of *A* by the NN model. There are 3918 reactions in the test set. The test set RMSE for the recreation of *A* is 1.33.

the reaction paths in the dataset that successfully converge. From Fig. 7(a), we can see that this simplistic method is insufficient for recreating the QM barrier and there is a large deviation from the correct value with an RMSE of 2.12 eV. Therefore, more sophisticated approaches must be used.

C. Predicting *AE* and Δ*H*

One approach to remove QM calculations from the EVB parametrization process is to predict the *AE* and Δ*H* directly rather than use QM. Creating predictive models for these two quantities greatly speeds up the parametrization process as only the optimization of the *A* and *B* terms is required. This methodology was tested as a route for EVB parametrization.

The prediction of *AE* and Δ*H* is shown in Fig. 5. Even with a relatively limited set of 15 674 reactions to train to (both forward and backward reactions are included), an RMSE error of 0.35 and 0.27 eV can be reached. The predicted values of *AE* and Δ*H* can then be used to produce EVB parameters using the optimization workflow previously described.

The EVB model with *A* and *B* parameters optimized with the predicted E_A and Δ*H* terms is compared to the QM activation energy in Fig. 7(c). Excellent recreation of the QM activation energies is possible with an RMSE of 0.42 eV. This is only 0.07 eV greater than the direct prediction of E_A. Therefore, removing QM calculations from EVB parametrization is possible through the use of ML models.

D. Predicting *A*

While predicting *AE* does remove direct QM calculations from the workflow, parametrization of the *A* and *B* terms is still required. Therefore, direct prediction of the *A* and *B* terms is the preferred route. Given that the *B* parameters are *B* = 1.0 for the majority of values, we focus on predicting just the *A* parameters with the results shown in Fig. 6.

Figure 6 shows that we are able to predict the *A* EVB parameter. To the best of our knowledge, this is the first time that this has been shown. The next step is to test how well these predicted *A* parameters can recreate the *QM* activation energy. The Δ*H* predictive message passing neural network model previously created was used to define the energy difference between the product and reactant and the predicted *A* parameter used.

Figure 7(b) shows the ability of predicted *A* parameters to recreate the QM barrier height. The RMSE error is ~1 eV lower than the constant A, B values but higher than parametrizing EVB using predicted *AE* values. Figure 7(d) shows the error distribution for the three different approaches. With constant *A* and *B* terms, 49% of barrier heights are recreated within 1 eV. This rises to 81% when *A* values are predicted and 96% when *AE* is predicted. Therefore, predicting the *A* EVB parameters is possible and does greatly improve the activation energy recreation in comparison to using constant

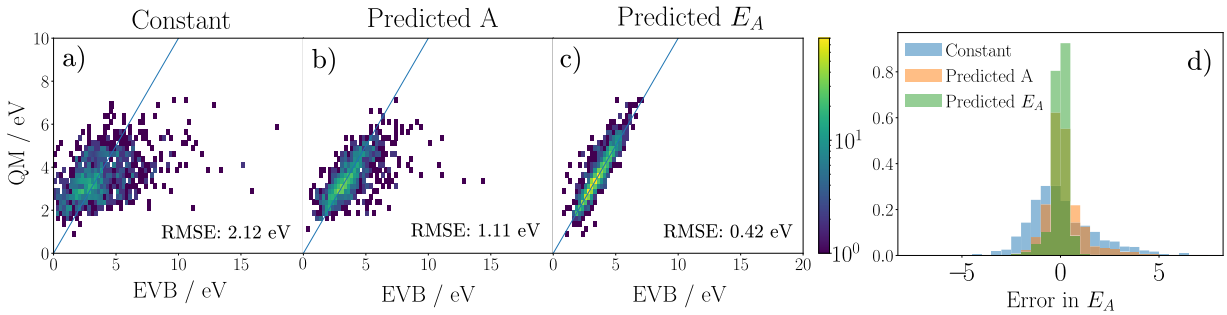

FIG. 7. The activation energy for EVB using (a) constant *A* = 5.36 and *B* = 1.00 (tested and converged on 1199 reactions), (b) predicted *A* values and *B* = 1.00 (trained on 15 675 reactions, tested and converged on 1173 reactions), and (c) predicted *AE* values compared to the QM results (trained on 15 675 reactions, tested and converged on 1527 reactions). The x axis is restricted to a maximum of 20 eV and only results that have fully converged are included. The ML models are trained on backward and forwards reactions from the original dataset, while the test set contains only the forward reaction.

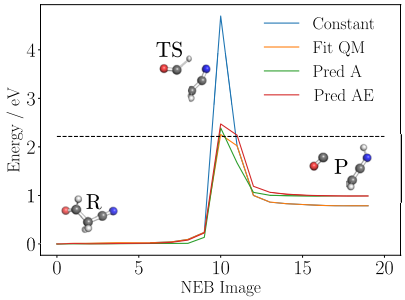

FIG. 8. The recreation of the MEP by the four different parameterization approaches used. The *A* value is set to 5.36, 7.80, 7.77, and 7.69 for the constant approach, fitting to QM, predicting the *A* model, and predicted the activation energy, respectively. For all the methods, *B* is set to 1.00.

EVB parameters. A comparison of an MEP for the four different fitting approaches is shown in Fig. 8.

IV. CONCLUSION

A fundamental limitation of classical force fields is their inability to perform reactive simulations.[3,4] The empirical valence bond method can overcome this limitation and allow reactions to be performed with classical force fields; however, it is required that expensive QM calculations, or experimental measurements, are carried out to find the activation energies for the systems of interest.[12] In this work, we show that by using ML techniques, we can remove the need for QM calculations in the EVB parametrization process. This development opens new routes for reactive simulations and suggests that it may be possible to give any classical force field reactive capabilities with minimal computational effort.

In this work, two separate approaches for parametrizing EVB without QM were proposed, predicting the activation energy and optimizing EVB from this predicted value or directly predicting the EVB parameters themselves. Predicting the *AE* terms was shown to result in a more accurate recreation of the QM activation energy, and there are a number of reasons for this. First, the *A* parameters of the EVB model are dependent on the classical force fields used. Information about the classical force field is not included in the input to the ML model and this prevents the ML model from accounting for the classical force field parameters used. In addition, there will be inaccuracies in the classical force field, and this will further reduce the accuracy possible when predicting the *A* parameters. The use of a constant *B* parameter will also limit the performance of the model.

Given these points, the EVB model could be improved in various ways. Changing the classical force field from the class I force field AMBER to a class II force field, such as COMPASS, which has a more complex functional form and parametrization, would improve performance as deficiencies in the classical force field would be corrected.[20,33] Along this, increasing the dataset size would reduce the test set error for the ML model as the set of 12 000 reactions currently used is relatively limited. However, even with the current performance of the model seen, we can still state the prediction of the EVB parameters is possible, with the RMSE error being reduced by 50% compared to constant EVB parameters.

Many EVB parameterizations are based upon the underlying assumption that the off-diagonal elements can be transferred between phases, and this is a valid assumption for S_N2 reactions.[28] In this work, we have parametrized our model to gas phase reactions. However, gas phase reactions are not sufficient for certain simulations such as enzyme catalysis.[34] This is a limitation of the current work. However, we expect equivalent performance for an ML model trained to EVB parameters optimized to datasets containing *ab initio* reactions in solvents or experimental datasets as to gas phase data. This forms a future direction for this work and would enable a wider range of reactions to be modeled.

An additional remaining challenge is the applicability of the method to multiple possible product states. To do this, we will use the method described in this work to consider the different reaction paths possible and calculate the EVB coupling terms for each reaction. The different paths will then be combined together using the weighting scheme presented in Ref. 35 to provide a new route to multi-reaction EVB.

Unreactive simulations can be performed quickly and easily with classical force fields. However, when reactions are required, the simulation process becomes significantly more complex with semi-empirical, *ab initio*, machine learning, or bond-order methods required. In this work, we looked to develop a new methodology for parametrizing EVB which could allow reactive simulations to be performed quickly, without the need for complex, and expensive, techniques.

ACKNOWLEDGMENTS

A. E. A. Allen acknowledged the Center for Nonlinear Studies, LA-UR-23-34051.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Alice E. A. Allen: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Gábor Csányi**: Conceptualization (equal); Methodology (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data used for this is work is available on request (and will be made available on publication) and from the stated references within the work.

REFERENCES

[1] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A new force field for molecular mechanical simulation of nucleic acids and proteins," J. Am. Chem. Soc. **106**, 765–784 (1984).
[2] W. L. Jorgensen and J. Tirado-Rives, "The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," J. Am. Chem. Soc. **110**, 1657–1666 (1988).

[3]J. W. Ponder and D. A. Case, "Force fields for protein simulations," in *Protein Simulations*, *Advances in Protein Chemistry* (Academic Press, 2003), Vol. 66, pp. 27–85.
[4]W. L. Jorgensen and J. Tirado-Rives, "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems," Proc. Natl. Acad. Sci. U. S. A. **102**, 6665–6670 (2005).
[5]M. J. Robertson, J. Tirado-Rives, and W. L. Jorgensen, "Improved peptide and protein torsional energetics with the OPLS-AA force field," J. Chem. Theory Comput. **11**, 3499–3509 (2015).
[6]R. Best, N.-V. Buchete, and G. Hummer, "Are current molecular dynamics force fields too helical?," Biophys. J. **95**, L07–L09 (2008).
[7]J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, "Long-timescale molecular dynamics simulations of protein structure and function," Curr. Opin. Struct. Biol. **19**, 120–127 (2009).
[8]A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, "Reaxff: A reactive force field for hydrocarbons," J. Phys. Chem. A **105**, 9396–9409 (2001).
[9]S. Zhang, M. Makoś, R. Jadrich, E. Kraka, K. Barros, B. Nebgen, S. Tretiak, O. Isayev, N. Lubbers, R. Messerly, and J. Smith, "Exploring the frontiers of chemistry with a general reactive machine learning potential," Nat. Chem. (published online) (2024).
[10]K. Park, W. Lin, and F. Paesani, "A refined MS-EVB model for proton transport in aqueous environments," J. Phys. Chem. B **116**, 343–352 (2012).
[11]M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," Phys. Rev. B **58**, 7260–7268 (1998).
[12]A. Warshel and R. M. Weiss, "An empirical valence bond approach for comparing reactions in solutions and in enzymes," J. Am. Chem. Soc. **102**, 6218–6226 (1980).
[13]B. J. Braams and J. M. Bowman, "Permutationally invariant potential energy surfaces in high dimensionality," Int. Rev. Phys. Chem. **28**, 577–606 (2009).
[14]R. G. Parr, "Density functional theory of atoms and molecules," in *Horizons of Quantum Chemistry*, edited by K. Fukui and B. Pullman (Springer Netherlands, Dordrecht, 1980), pp. 5–15.
[15]J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," Angew. Chem., Int. Ed. **56**, 12828–12840 (2017).
[16]A. M. Smondyrev and G. A. Voth, "Molecular dynamics simulation of proton transport through the influenza a virus M2 channel," Biophys. J. **83**, 1987–1996 (2002).
[17]T. Zubatiuk, B. Nebgen, N. Lubbers, J. S. Smith, R. Zubatyuk, G. Zhou, C. Koh, K. Barros, O. Isayev, and S. Tretiak, "Machine learned hückel theory: Interfacing physics and deep neural networks," J. Chem. Phys. **154**, 244108 (2021).
[18]G. Zhou, N. Lubbers, K. Barros, S. Tretiak, and B. Nebgen, "Deep learning of dynamically responsive chemical Hamiltonians with semiempirical quantum mechanics," Proc. Natl. Acad. Sci. U. S. A. **119**, e2120333119 (2022).
[19]A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, 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. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—A python library for working with atoms," J. Phys.: Condens. Matter **29**, 273002 (2017).
[20]D. A. Case *et al.*, *Amber* (University of California, San Francisco, 2022).
[21]G. Henkelman and H. Jónsson, "A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives," J. Chem. Phys. **111**, 7010–7022 (1999).
[22]H. Jonsson, G. Mills, and K. W. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions," in *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific, 1998), pp. 385–404.
[23]C. A. Grambow, L. Pattanaik, and W. H. Green, "Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry," Sci. Data **7**, 137 (2020).
[24]C. A. Grambow, L. Pattanaik, and W. H. Green, "Deep learning of activation energies," J. Phys. Chem. Lett. **11**, 2992–2997 (2020).
[25]H. B. Schlegel and J. L. Sonnenberg, "Empirical valence-bond models for reactive potential energy surfaces using distributed Gaussians," J. Chem. Theory Comput. **2**, 905–911 (2006).
[26]B. Hartke and S. Grimme, "Reactive force fields made simple," Phys. Chem. Chem. Phys. **17**, 16715–16718 (2015).
[27]J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," J. Comput. Chem. **25**, 1157–1174 (2004).
[28]G. Hong, E. Rosta, and A. Warshel, "Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases," J. Phys. Chem. B **110**, 19570–19574 (2006).
[29]G. Henkelman and H. Jónsson, "Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points," J. Chem. Phys. **113**, 9978–9985 (2000).
[30]P. M. Zimmerman, "Single-ended transition state finding with the growing string method," J. Comput. Chem. **36**, 601–611 (2015).
[31]Y. Shao *et al.*, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," Mol. Phys. **113**, 184–215 (2015).
[32]K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, and R. Barzilay, "Analyzing learned molecular representations for property prediction," J. Chem. Inf. Model. **59**, 3370–3388 (2019).
[33]H. Sun, Z. Jin, C. Yang, R. L. C. Akkermans, S. H. Robertson, N. A. Spenley, S. Miller, and S. M. Todd, "COMPASS II: Extended coverage for polymer and drug-like molecule databases," J. Mol. Model. **22**, 47 (2016).
[34]S. C. L. Kamerlin and A. Warshel, "The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions," Faraday Discuss. **145**, 71–106 (2010).
[35]P. Pinski and G. Csányi, "Reactive many-body expansion for a protonated water cluster," J. Chem. Theory Comput. **10**, 68–75 (2014).