Exhaustive State-to-State Cross Sections for Reactive Molecular Collisions from Importance Sampling Simulation and a Neural Network Representation

Debasish Koner, Oliver T. Unke, Kyle Boe, Raymond J. Bemish, and Markus Meuwly

1) Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland
2) Boston College, Institute for Scientific Research, Chestnut Hill, Massachusetts 02467, USA
3) Air Force Research Laboratory, Space Vehicles Directorate, Kirtland AFB, New Mexico 87117, USA

High-temperature, reactive gas flow is inherently non-equilibrium in terms of energy and state population distributions. Modeling such conditions is challenging even for the smallest molecular systems due to the extremely large number of accessible states and transitions between them. Here, neural networks (NNs) trained on explicitly simulated data are constructed and shown to provide quantitatively realistic descriptions which can be used in mesoscale simulation approaches such as direct simulation Monte Carlo (DSMC) to model gas flow at the hypersonic regime. As an example, the state-to-state cross sections for the N(4S)+NO(2Π) → O(3P)+N2(X1Σg+) are computed from quasiclassical trajectory (QCT) simulations. By training NNs on a sparsely sampled noisy set of state-to-state cross sections it is demonstrated that independently generated reference data is predicted with high accuracy. State-specific and total reaction rates as a function of temperature from the NN are in quantitative agreement with explicit QCT simulations and confirm earlier simulations and the final state distributions of the vibrational and rotational energies agree as well. Thus, NNs trained on physical reference data can provide a viable alternative to computationally demanding explicit evaluation of the microscopic information at run time. This will considerably advance the ability to realistically model non-equilibrium ensembles for network-based simulations.
There are numerous situations in physical chemistry which involve a potentially large number of states and transitions between them. Examples include complete line lists for polyatomic molecules in hot environments (e.g. HITRAN\textsuperscript{1}) or state-to-state cross sections in reactive and non-reactive molecular collisions. Exhaustively probing and enumerating all relevant combinations or creating high-dimensional analytical representations is usually impossible. On the other hand, it has been shown for spectroscopic applications that omission of certain crucial states makes prediction or modeling of the spectroscopic band difficult or even impossible.\textsuperscript{2} Another example is hypersonic flow around a space vehicle reentering the atmosphere. Temperatures can easily reach 20000 K where reliable experimental data is sparse and the energy distributions are out of equilibrium. In such an environment, the space vehicle is exposed to a collisionally dense environment which generates an immensely diverse population of rovibrational states between which collisions take place.\textsuperscript{3} An accurate, fast, and reliable method is required to include this information in more coarse grained models to study the associated dynamics. The question thus is, how to best probe and represent a function with $> 10^7$ values for discrete input data, without explicitly computing each entry which may require thousands to millions of samples to statistically converge each of the entries.

The present work attempts to develop such a model for state-to-state cross sections $\sigma_{v,j \rightarrow v',j'}(E_t)$ between initial $(v,j)$ and final $(v',j')$ ro-vibrational states at given relative translational energy $E_t$. For this, the $\text{N}(^4\text{S})+\text{NO}(^2\Pi)(v,j) \rightarrow \text{O}(^3\text{P})+\text{N}_2(X^1\Sigma^+_g)(v',j')$ reaction is considered because a) it is relevant in the hypersonic flight regime characterized by temperatures $T \leq 20000$ K at which a multitude of the available ro-vibrational states are occupied and accessible, and b) an accurate, fully-dimensional and reactive potential energy surface (PES) is available.\textsuperscript{4} Specifically, $\text{N}_2$-formation rates from simulations on the $^3A'$ and $^3A''$ PESs are in favourable agreement with experiments for temperatures $T \geq 2000$ K and $\text{N}_2$-formation below 5000 K is dominated by processes on the $^3A''$ surface.\textsuperscript{4} State-to-state cross sections are typically required when modeling the reactive flow around a re-entry object with macroscopic dimensions using techniques such as direct simulation Monte Carlo (DSMC)\textsuperscript{5} as the flows are usually not in thermal equilibrium. However, it should be noted that the present ansatz will be applicable to other relevant simulations and hypersonic flow is merely chosen because the occupation number of the available states is high and the
number of states is large as well.

Scattering cross sections can be determined from quasiclassical trajectory (QCT) simulations. Here, Hamilton’s equations of motion are solved using a fourth order Runge-Kutta numerical integration in reactant Jacobi coordinates. The analytical representation of the $^3A'$ potential energy surface (PES) is a reproducing kernel Hilbert space (RKHS) based on MRCI+Q/aug-cc-pVTZ calculations. The two possible reactive channels are (I) nitrogen exchange (NO+N' → N'O+N) and (II) N$_2$ molecule formation (NO+N→N$_2$+O). Besides these two channels, the elastic or inelastic collisions can also lead to kinetically or internally excited reactants.

For the $^3A'$ surface of N$_2$O a maximum of 47 and 57 vibrational states are available for NO and N$_2$, respectively, while the maximum rotational quantum numbers for NO and N$_2$ are 241 and 273. Altogether, there are 6329 ro-vibrational states for the N+NO channel, and 8733 states for the O+N$_2$ channel. For one defined initial state $(v, j)$ there are $>10^4$ possible final states $(v', j')$. To converge each of the $\sigma_{v,j\rightarrow v'j'}(E_t)$ for given $E_t$ with a statistical significance of $\sim 10 \%$ an estimated $\geq 10^{13}$ trajectories would be required ($10^7$ trajectories to converge one cross section, see below; $\sim 10^4$ initial states; $\sim 10^4$ final states per initial state). Hence, using QCT simulations to directly sample all possible ro-vibrational initial states is computationally impractical, even for this simple 3-body system which also ignores the complexity of electronic states. Thus, alternative approaches need to be explored.

To better define the QCT sampling problem, state-to-state cross sections for $\sigma_{v=6,j=30\rightarrow v'j'}(E_t = 2.5 \text{ eV})$ were considered. A total of $2 \times 10^7$ trajectories was run initially, which is considered as the reference. Out of the 3784 energetically accessible states, 3420 final states are found as products, i.e. 90.4 %. Compared to this, running fewer trajectories ($8 \times 10^4$, $1.6 \times 10^5$, $1.6 \times 10^6$ and $9.6 \times 10^6$) finds 37 %, 54 %, 83 % and 90% of the final states, respectively, see Figure S1, which converges at approximately the cube root of the number of trajectories. In addition, running too few trajectories leads to highly oscillatory cross sections due to the limited statistics of the final state.

Some computational gain can be obtained from exploring the fact that cross sections often
vary smoothly with \( v \) and \( j \). This allows to locally average computed cross sections according to

\[
\sigma_{v,j \rightarrow v',j'}^{\text{av}} = \frac{1}{2n_v + 1} \frac{1}{2n_j + 1} \sum_{v'=v'-n_v}^{v'+n_v} \sum_{j'=j'-n_j}^{j'+n_j} \sigma_{v,j \rightarrow v',j'}. 
\]

(1)

where \( n_v \) and \( n_j \) are the bin widths for vibration and rotation over which the state to state cross sections are averaged. Figures 1A and B report the raw data from \( 1.6 \times 10^6 \) trajectories and locally averaged cross sections from the same data set with \( n_v = 2 \) and \( n_j = 3 \), respectively. Comparison with the unaveraged result from \( N_{\text{tot}} = 2 \times 10^7 \) trajectories in Figure 1D illustrate the benefit of local averaging, see also Figure S4 for another example (\( v = 10, j = 60 \) and \( E_t = 1.8 \) eV). Convergence of the state-to-state cross sections with respect to \( N_{\text{tot}} \) and the effect of local averaging are shown in Figure S2 and S3. Averaging over neighboring states reduces the noise and also decreases the number of trajectories required approximately by the square root of the number of states averaged over, i.e. \( \sqrt{(2n_v + 1)(2n_j + 1)} = \sqrt{35} \). Besides that, local averaging also provides values for \( \sigma_{v,j \rightarrow v',j'} > 0 \) for unsampled transitions and reduces local oscillations while correctly describing the broad features (see Figures 1 and S4). However, sharp resonances, for which the width is comparable to the size of the averaging window, are washed out in this approach and depending on the application this point would need to be considered separately.

Concerning the sampling strategy for choosing initial conditions it is worthwhile to note that large impact parameters \( b \) mostly lead to nonreactive collisions. A straightforward approach to sampling \( b \) is to draw it from \( b^2/b_{\text{max}}^2 \) with \( 0 \leq b \leq b_{\text{max}} \) (see Figure S5) and the number of trajectories sampled in the interval \( b + db \) increases with increasing \( b \), which leads to a larger fraction of nonreactive trajectories when \( b \) increases. For such situations, importance sampling (IS)\(^{12}\) can provide a more advantageous protocol as those values of \( b \) for which reactive trajectories are more likely to occur are chosen with higher probability, which causes the cross section of all of the exit channels to converge at the same rate.

In order to determine the necessary weighting function \( w(b) \) for a particular trajectory the following strategy is used. For given \( (v,j, E_t) \) first a few thousand trajectories are run by uniformly sampling \( 0 \leq b \leq b_{\text{max}} \) (see Figure S5). The number of reactive trajectories as a
FIG. 1. State-to-state cross sections ($\sigma_{v,j \rightarrow v',j'}$) for $N + NO(v = 6, j = 30) \rightarrow O + N_2(v', j')$ at $E_t = 2.5$ eV computed from different $N_{tot}$. Cross sections shown in the top right and bottom left panels are averaged using Eq. 1. In the bottom left panel (C) cross sections are obtained from trajectories sampled in $b$ space via importance sampling which are close to those in panel (D) from $2 \times 10^7$ trajectories.

The function of $b$ is fitted to

$$n(b) = a_0 \exp(-a_1 b) + a_2 \exp\left(-\frac{\ln(2)b^2}{a_3^2}\right).$$

(2)

From this, the distribution of the cross sections, $g(b) = 2\pi b n(b)$ is determined. Next, $1.6 \times 10^5$ initial conditions are sampled from $g(b)$ (see Figure S5), the trajectories are explicitly run and the weight of each trajectory is calculated according to $w(b) = f(b)/g(b)$ where $f(b) = 2b/b_{max}^2$.

The performance of IS is illustrated in Figures 1C and S4C which report the locally averaged cross sections from $1.6 \times 10^5$ samples for $\sigma_{v=6,j=30\rightarrow v',j'}(E_t = 2.5$ eV) and $\sigma_{v=10,j=60\rightarrow v',j'}(E_t = 1.8$ eV), respectively. For these two examples, 42193 and 38665 reactive trajectories are
FIG. 2. 3D surface (upper panel) and contour color map (lower panel) of QCT calculated and NN-STS predicted state-to-state cross sections for $N + NO(v = 6, j = 30) \rightarrow O + N_2(v', j')$ at $E_t = 2.5$ eV.

found using IS from $N_{tot} = 1.6 \times 10^5$ compared with 44737 and 44979 reactive trajectories for $N_{tot} = 1.6 \times 10^6$ from conventional sampling, respectively. This leads to an efficiency increase by one order of magnitude when IS is used. The effect of IS on the convergence of state-to-state cross section can also be seen in Figure S2 and S3.

Overall, IS and local averaging lead to an estimated reduction of the required number of QCT trajectories by a factor of $\sim 60$ which will be explored next to cover the entire state space for state-to-state cross sections for different selected initial states. Those will then be used for training a NN to construct a model to compute state-to-state cross sections.

To compute the necessary reference data to train the NN, 10 initial $v$–states ($v = 0, 3, 6, 9, 12, 15, 19, 23, 28, \text{and} 34$) and 12 initial $j$–states ($j = 0, 25, 50, 75, 100, 125, 145, 160, 175, 190, 200 \text{and} 210$) were sampled. The relative translational energies ($E_t$) for the N+NO collision were 0.05, 0.1, 0.25, 0.5, 0.8, 1.2, 1.6, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 and 5.5 eV. Thus, for a total of 1232 initial conditions in the $(v, j, E_t)$ space, for each of them $1.6 \times 10^5$ QCT trajectories were run with IS of $b$. The cross sections from these $\sim 10^8$ trajectories
constitute the training set for learning the NN to predict $\sigma_{v,j \to v',j'}(E_t)$ for any initial $(v, j, E_t)$.

The network architecture used here is inspired by ResNet. The NN transforms its input through four identical residual blocks, after which a linear transformation followed by a scaled sigmoid function is used to obtain the final output, see Figure S6. Before they can be used as input to the residual blocks, the raw input features $f \in \mathbb{R}^{N_{in}}$ are first transformed by a linear transformation $x = Wf + b$ to a vector $x \in \mathbb{R}^F$ with the same dimensionality $F$ as the hidden features in the residual blocks in order to simplify the formulation of skip connections. Here, $W \in \mathbb{R}^{N_{in} \times F}$ (weight matrix) and $b \in \mathbb{R}^F$ (bias vector) are parameters to be optimized. The residual blocks consist of two dense layers with the same number of nodes $F$ (see Figure S6) and transform their input $x^l$ according to

$$x^{l+2} = x' + \text{ReLU} [W^{l+1}\text{snasinh}(W^l x^l + b^l) + b^{l+1}]$$

where the superscript $l$ denotes parameters or feature representations of layer $l$, and $W \in \mathbb{R}^{F \times F}$ and $b \in \mathbb{R}^F$ are parameters. Two different activation functions, one for rectified linear units (ReLU) and a self-normalizing inverse hyperbolic sine (snasinh) are used in the residual blocks. Use of residual blocks improves the flow of gradients between layers and helps alleviate the “vanishing gradients problem”. After the last residual block, the final output is obtained from

$$y = C \times \text{sig}(W^o x^f + b^o),$$

where $C$ is a scaling constant, $\text{sig}(x) = (1 + e^{-x})^{-1}$ denotes the sigmoid function and the superscripts $o$ and $l$ denote parameters $W^o \in \mathbb{R}^{F \times 1}$ and $b^o \in \mathbb{R}$ corresponding to the output layer and the hidden features $x^f$ obtained after the last residual block, respectively.

The initial states of the reactants are described by $N_{in} = 12$ input features $f$, namely: (i) internal energy, (ii) vibrational energy, (iii) vibrational quantum number, (iv) rotational energy, (v) rotational quantum number, (vi) angular momentum of the diatom, (vii) relative translational energy, (viii) relative velocity, (ix, x) turning periods of the diatom, (xi) rotational barrier height and (xii) vibrational time period of the diatom. For state-to-state cross sections, the same 12 features for the final states of the products are also included as
input (i.e., \( N_{in} = 24 \)). The loss functions \( (L_f) \) was defined as
\[
L_f = \frac{1}{N} \sum_{i=1}^{N} \left[ \log(y'_i + 1.0) - \log(y'_i + |y_i - y'_i| + 1.0) \right]^2,
\]
where \( y'_i \) and \( y_i \) are the reference (QCT) and predicted values (NN), respectively. This loss function penalizes all relative errors in the prediction of cross sections approximately equally irrespective of the absolute magnitude of the reference value. This is important, because for small values of the cross section, a small absolute error corresponds to a large relative error.

All parameters of the NN are initialized according to the Glorot initialization scheme\(^{18}\) and optimized using Adam optimization\(^{19}\) with an exponentially decaying learning rate. From all state-to-state cross sections \( (\sim 8 \times 10^6) \), different numbers \( (N_{train}) \) of training samples were randomly chosen for training and \( N_{train}/10 \) data was used for validation. From the remaining data, \( 2 \times 10^5 \) values constitute the test data set. The convergence of the model with training set size is reported in Figure S7. Training the NN for state-to-state cross sections (referred to as NN-STS) was done with \( N_{in} = 24, F = 24, \) and \( C = 0.4 \).

Since trajectories were calculated for different initial reactant states \( (v, j, E_t) \) the total cross sections \( (\sigma_{tot}(v, j, E_t)) \) for the \( N + NO \rightarrow N_2 + O \) reaction are also available. On this data, another NN (with \( N_{in} = 12, F = 8, \) and \( C = 85.0 \)) was trained using 1122 training data and 110 validation data to get a model for the \( \sigma_{tot}(v, j, E_t) \) (referred to as NN-Tot). Both networks were trained using TensorFlow\(^{20}\). The models with the lowest validation losses (Eq. 5) are selected and used to predict different observables in the following and compared with values from explicit QCT simulations.

The entire reference state-to-state data set \( (\sim 10^7) \) has an average cross section of 0.00246 a\(^2\). To test the quality of model NN-STS, \( 2 \times 10^5 \) state-to-state cross sections were randomly chosen from the test data. The RMSE is 0.000328 a\(^2\) for those data with a maximum deviation of 0.008 a\(^2\) (for \( \sigma_{QCT} = 0.211 \) a\(^2\), \( \sigma_{NN-STS} = 0.219 \) a\(^2\)) and the correlation coefficient is \( R^2 = 0.993 \), see Figure S8. A similar analysis was carried out for model NN-Tot which has an average cross section of 16.18 a\(^2\) for all 1232 data points. The RMSE is found to be 0.1552 a\(^2\) and the correlation between NN-Tot and QCT is \( R^2 = 0.9997 \), see Figure S9, indicative of a high quality of the fit.
FIG. 3. Rotationally averaged, vibrational distributions of the cross sections, $\sigma(v)$, calculated from QCT (blue) and predicted from NN-STS (red) for the $N + NO(v,j) \rightarrow O + N_2(v')$ reaction.

To quantify the accuracy of the NN, additional QCT calculations were performed for independent initial conditions at fixed $E_t$. Total QCT cross sections are then compared with the NN predictions, see Figures 2 and 3 and Table S1. Again, the NN results describe the explicit QCT simulations which validates the use of such a model to predict microscopic information for such a reaction.

According to the method described in Ref. 11 (see also supporting information) initial state specific rates for the reaction were also calculated at temperatures between 2000 and 20000 K for a few selected reactant states using QCT simulations and compared with the rates obtained from the NN models in Figures 4 and S10. The NN models successfully capture the trends as well as the magnitudes of the rates from QCT. Although maximum relative errors of $\sim 17\%$ are found for $v = 5, j = 85$ and $T = 2000$ K from model NN-STS and $\sim 13\%$ for $v = 5, j = 25$ and $T = 2000$ K from NN-Tot, in most cases the relative errors are $< 5\%$. Good agreement between the QCT and NN rates can be seen in the correlation diagrams shown in Figure S11.
FIG. 4. Initial state specific rates calculated and predicted from QCT and NN, respectively, for the \( \text{N + NO}(v,j) \rightarrow \text{O + N}_2 \) reaction at different temperatures. The symbols represent the QCT results while the solid lines are the NN-STS results. The top, middle and bottom panel show the results for \( v = 5, 10 \) and 20 while the magenta, olive, green, blue and red colors represent \( j = 20, 40, 60, 85 \) and 110, respectively.

As another test, total thermal rates \( k(T) \) were calculated from QCT simulations and compared with results obtained from NN the models, see Figure 5 and the supporting information. The NN rates are calculated from integrating the NN cross sections over the \((v,j)\)–state and translational energy phase space using Monte Carlo integration. For NN-Tot the agreement with QCT is particularly striking whereas it is still good for NN-STS except for temperatures 1000 K and 20000 K. Here it should be mentioned that only a range of 0.05 to 5.5 eV in translational energy is covered by the QCT calculations to generate the state-to-state data. Sampling a broader range of energies will most likely further improve rates determined from the NN model in the low and high temperature regions.
FIG. 5. Total rates calculated from QCT (blue) and predicted by the NN models (NN-STS - red and NN-Tot - black) for the $^3A'$ state of the $N + NO \rightarrow O + N_2$ reaction between 1000 and 20000 K. $N_2$-formation below 5000 K is dominated by the $^3A''$ PES which is indicated by the rapid falloff of the present data. The present rates agree quantitatively with Ref. 4.

As a final validation of the robustness of NNs, the distribution of the final vibrational and rotational states and the rovibrational energies of $N_2$ after $N+NO$ collisions at different temperatures are calculated from QCT and compared with those from NN-STS, see Figure 6. The NN correctly captures the shape of all distributions but lacks the oscillatory features, in particular for the rotational distribution. Thus, the NN provides a physically robust model based on validated, microscopic data from which information about nonequilibrium systems can be obtained, obviating the construction of models based on simple, empirical expressions.\textsuperscript{21}

In this work an NN-based model for state-to-state cross sections has been constructed. For this purpose a total of $\sim 8 \times 10^6$ state-to-state cross sections have been explicitly determined from QCT simulations for selectively chosen 1232 initial states. Local averaging over $(v',j')$ reduces the noise of the data set and IS to sample impact parameters for the trajectories further accelerates the convergence of the cross sections. Typical training times for the NN-STS models are a few days and evaluation time of the NN for $10^6$ state-to-state cross
FIG. 6. Distributions of product vibrational, rotational states and ro-vibrational energies calculated from QCT (blue) and predicted by model NN-STS (red), respectively for the $^3A'$ state of the N + NO → O + N$_2$ at different temperatures.

sections is 24.4 s on a 64 bit 2.40 GHz Intel E5-2620 v3 CPU using a single processor. This makes the technique suitable for direct use in DSMC simulations where a large number of collision cross sections are required to model hypersonic air flow. The average error from the NN compared with the reference QCT data is $\sim 5 \%$. This compares with errors ranging from 25 \% to 60 \% for vibrational relaxation rates and state-specific dissociation rates from a maximum entropy model for O$_2$ + O.\textsuperscript{22} Unfortunately, the error for the cross sections is not reported and can not be compared here.

In summary, the state-to-state cross sections for a reactive collision relevant to the hypersonic flight regime has been successfully modelled using a neural network based on explicitly calculated data from QCT simulations on an accurate, fully dimensional reactive PES. Such an approach is general and demonstrates that for situations in which large amounts of data constitute the relevant state space, subsampling and subsequent machine learning can provide a viable, accurate and computationally tractable alternative to exhaustive compu-
tations.
ACKNOWLEDGMENT

Support by the Swiss National Science Foundation through grants 200021-117810, the NCCR MUST (to MM), and the University of Basel is also acknowledged. Part of this work was supported by the United State Department of the Air Force which is gratefully acknowledged (to MM).

DATASET AND CODE AVAILABILITY

The code for training the NN-model is freely available from https://github.com/MeuwlyGroup/NNcross together with input data and a sample training set.

REFERENCES

1L. Rothman, D. Jacquemart, A. Barbe, D. C. Benner, M. Birk, L. Brown, M. Carleer, C. Chackerian, K. Chance, L. Coudert, V. Dana, V. Devi, J.-M. Flaud, R. Gamache, A. Goldman, J.-M. Hartmann, K. Jucks, A. Maki, J.-Y. Mandin, S. Massie, J. Orphal, A. Perrin, C. Rinsland, M. Smith, J. Tennyson, R. Tolchenov, R. Toth, J. V. Auwera, P. Varanasi, and G. Wagner, Journal of Quantitative Spectroscopy and Radiative Transfer 96, 139 (2005).

2S. N. Yurchenko, J. Tennyson, J. Bailey, M. D. J. Hollis, and G. Tinetti, Proc. Natl. Acad. Sci. 111, 9379 (2014).

3R. Celiberto, I. Armenise, M. Cacciatore, M. Capitelli, F. Esposito, P. Gamallo, R. K. Janev, A. Laganà, V. Laporta, and A. Laricchiuta, Plasma Sources Sci. Technol. 25, 033004 (2016).

4O. Denis-Alpizar, R. J. Bemish, and M. Meuwly, Phys. Chem. Chem. Phys. 19, 2392 (2017).

5G. A. Bird, Molecular Gas Dynamics and the Direct Simulation of Gas Flows (Clarendon Press, 1994).

6M. Karplus, R. N. Porter, and R. D. Sharma, J. Chem. Phys. 43, 3259 (1965).

7D. G. Truhlar and J. T. Muckerman, in Atom - Molecule Collision Theory, edited by R. B. Bernstein (Springer US, 1979) pp. 505–566.
8N. E. Henriksen and F. Y. Hansen, *Theories of Molecular Reaction Dynamics* (Oxford, 2011).

9T.-S. Ho and H. Rabitz, J. Chem. Phys. **104**, 2584 (1996).

10O. T. Unke and M. Meuwly, J. Chem. Inf. Model **57**, 1923 (2017).

11D. Koner, R. J. Bemish, and M. Meuwly, J. Chem. Phys. **149**, 094305 (2018).

12R. Srinivasan, *Importance Sampling Applications in Communications and Detection* (Springer, 2002).

13K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. , 770 (2016).

14V. Nair and G. E. Hinton, in *Proceedings of the 27th International Conference on Machine Learning* (2010) pp. 807–814.

15G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, arXiv preprint arXiv:1706.02515 (2017).

16O. T. Unke and M. Meuwly, J. Chem. Phys. **148**, 241708 (2018).

17K. He, X. Zhang, S. Ren, and J. Sun, Identity Mappings in Deep Residual Networks. European Conference on Computer Vision. , 630 (2016).

18X. Glorot and Y. Bengio, in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (2010) pp. 249–256.

19D. Kingma and J. Ba, arXiv preprint arXiv:1412.6980 (2014).

20https://www.tensorflow.org/.

21N. Singh and T. Schwartzentruber, Proc. Natl. Acad. Sci. **115**, 47 (2017).

22M. Kulakhmetov, M. Gallis, and A. Alexeenko, J. Chem. Phys. **144** (2016).