Exhaustive state-to-state cross sections for reactive molecular collisions from importance sampling simulation and a neural network representation 🔵EP

🆔 Debasish Koner, 🆔 Oliver T. Unke, 🆔 Kyle Boe, et al.

COLLECTIONS

🔵EP This paper was selected as an Editor's Pick

🌐 View Online   ⬆ Export Citation   🔴 CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Unsupervised machine learning in atomistic simulations, between predictions and understanding
The Journal of Chemical Physics **150**, 150901 (2019); https://doi.org/10.1063/1.5091842

Dynamics in reactions on metal surfaces: A theoretical perspective
The Journal of Chemical Physics **150**, 180901 (2019); https://doi.org/10.1063/1.5096869

Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework
The Journal of Chemical Physics **150**, 220901 (2019); https://doi.org/10.1063/1.5092590
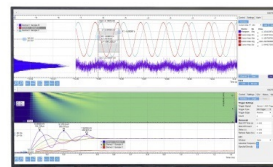
Exhaustive state-to-state cross sections for reactive molecular collisions from importance sampling simulation and a neural network representation EP

Debasish Koner,[1] iD Oliver T. Unke,[1] iD Kyle Boe,[2] iD Raymond J. Bemish,[3] iD and Markus Meuwly[1,a] iD

AFFILIATIONS

[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, Albuquerque, New Mexico 87117, USA

[a] m.meuwly@unibas.ch

ABSTRACT

High-temperature, reactive gas flow is inherently nonequilibrium 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 to model gas flow at the hypersonic regime. As an example, the state-to-state cross sections for $N(^4S) + NO(^2\Pi) \rightarrow O(^3P) + N_2(X^1\Sigma_g^+)$ 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 are 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 nonequilibrium ensembles for network-based simulations.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5097385

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., high-resolution transmission molecular absorption (HITRAN)[1] database] or state-to-state (STS) cross sections in reactive and nonreactive 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.[2] Another example is hypersonic flow around a space vehicle reentering the atmosphere. Temperatures can easily reach 20 000 K, for which reliable experimental data are 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.[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') rovibrational states at given relative translational energy E_t. For this, the $N(^4S) + NO(^2\Pi)(v, j) \rightarrow O(^3P) + N_2(X^1\Sigma_g^+)(v', j')$ reaction is considered because (a) it is relevant in the hypersonic flight regime characterized by temperatures $T \leq 20\,000$ K at which a multitude of the available rovibrational states are occupied and accessible and (b) an accurate, fully dimensional, and reactive potential energy surface (PES) is available.[4] Specifically, N_2-formation rates from simulations on the $^3A'$ and $^3A''$ PESs are in favorable agreement with experiments for temperatures $T \geq 2000$ K and N_2-formation below 5000 K is dominated by processes on the $^3A''$ surface.[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)[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.[6–8] 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)[9,10] 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_2O, 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 rovibrational 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 ~10%, an estimated ≥10^{13} trajectories would be required (10^7 trajectories to converge one cross section, see below; ~10^4 initial states; ~10^4 final states per initial state). Hence, using QCT simulations to directly sample all possible rovibrational initial states is computationally impractical, even for this simple 3-body system which also ignores the complexity of electronic states.[11] Thus, alternative approaches need to be explored.

To better define the QCT sampling problem, state-to-state cross sections for $\sigma_{v=6,j=30 \rightarrow v'j'}(E_t = 2.5$ eV) were considered. A total of 2×10^7 trajectories were 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 (8×10^4, 1.6×10^5, 1.6×10^6, and 9.6×10^6) finds 37%, 54%, 83%, and 90% of the final states, respectively, see Fig. 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 us to locally average computed cross sections according to

$$\sigma_{v,j \rightarrow v',j'}^{av} = \frac{1}{2n_v + 1} \frac{1}{2n_j + 1} \sum_{v*=v'-n_v}^{v'+n_v} \sum_{j*=j'-n_j}^{j'+n_j} \sigma_{v,j \rightarrow v*,j*}, \quad (1)$$

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 1(a) and 1(b) report the raw data from 1.6×10^6 trajectories and locally
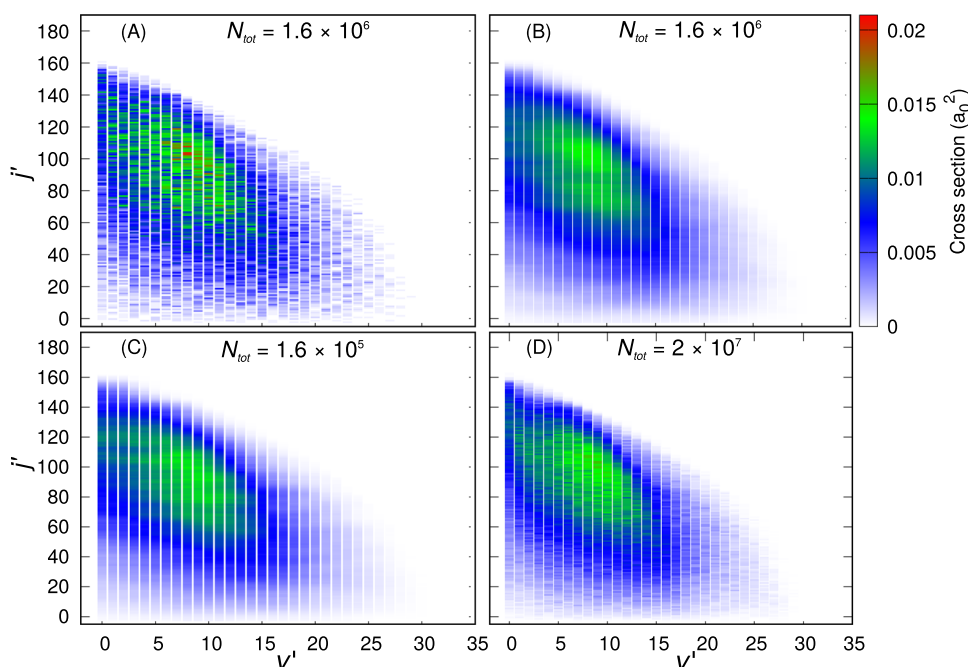
FIG. 1. State-to-state cross sections $(\sigma_{v,j \rightarrow v',j'})$ for N + NO($v = 6$, $j = 30$) → O + $N_2(v', j')$ at $E_t = 2.5$ eV computed from different N_{tot}. Cross sections shown in the top right (b) and bottom left (c) panels are averaged using Eq. (1). In panel (c), cross sections are obtained from trajectories sampled in b space via importance sampling which are close to those in panel (d) from 2×10^7 trajectories.

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 Fig. 1(d) illustrates the benefit of local averaging; see also Fig. S4, for example ($v = 10$, $j = 60$ and $E_t = 1.8$ eV). Convergence of the state-to-state cross sections with respect to N_{tot} and the effect of local averaging are shown in Figs. 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 $\sigma_{v,j \to v'j'} > 0$ for unsampled transitions and reduces local oscillations while correctly describing the broad features (see Fig. 1 and Fig. 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_{max}^2 with $0 \le b \le b_{\text{max}}$ (see Fig. 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 $0 \le b \le b_{\text{max}}$ (see Fig. S5). The number of reactive trajectories as a function of b is fitted to

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

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

The performance of IS is illustrated in Fig. 1(c) and Fig. S4(C) which report the locally averaged cross sections from 1.6×10^5 samples for $\sigma_{v=6,j=30 \to v'j'}(E_t = 2.5$ eV) and $\sigma_{v=10,j=60 \to v'j'}(E_t = 1.8$ eV), respectively. For these two examples, 42 193 and 38 665 reactive trajectories are found using IS from $N_{\text{tot}} = 1.6 \times 10^5$ compared with 44 737 and 44 979 reactive trajectories for $N_{\text{tot}} = 1.6 \times 10^6$ 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 Figs. S2 and S3.

Overall, IS and local averaging lead to an estimated reduction of the required number of QCT trajectories by a factor of ~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 neural network (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 ($v = 0, 3, 6, 9, 12, 15, 19, 23, 28,$ and 34) and 12 initial j-states ($j = 0, 25, 50, 75$ 100, 125, 145, 160, 175, 190, 200, and 210) were sampled. The relative translational energies (E_t) for the N + NO
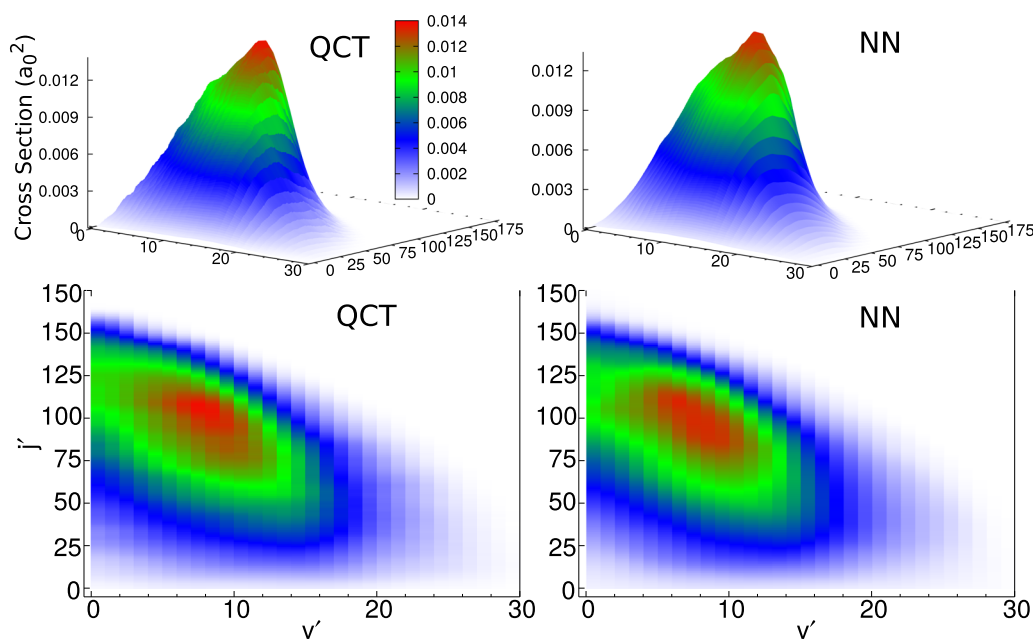
FIG. 2. 3D surface (upper panel) and contour color map (lower panel) of QCT calculated and NN-STS predicted state-to-state cross sections for N + NO($v = 6$, $j = 30$) → O + N$_2$(v', j') at E_t = 2.5 eV.

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, 1.6×10^5 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.[13] The NN transforms its input through four identical residual blocks,[13] after which a linear transformation followed by a scaled sigmoid function is used to obtain the final output; see Fig. S6. Before they can be used as input to the residual blocks, the raw input features $\mathbf{f} \in \mathbb{R}^{N_{in}}$ are first transformed by a linear transformation $\mathbf{x} = \mathbf{W}\mathbf{f} + \mathbf{b}$ to a vector $\mathbf{x} \in \mathbb{R}^F$ with the same dimensionality F as the hidden features in the residual blocks in order to simplify the formulation of skip connections.[13] Here, $\mathbf{W} \in \mathbb{R}^{N_{in} \times F}$ (weight matrix) and $\mathbf{b} \in \mathbb{R}^F$ (bias vector) are parameters to be optimized. The residual blocks consist of two dense layers with the same number of nodes F (see Fig. S6) and transform their input $\mathbf{x}^l$ according to

$$\mathbf{x}^{l+2} = \mathbf{x}^l + \text{ReLU}[\mathbf{W}^{l+1}\text{snasinh}(\mathbf{W}^l\mathbf{x}^l + \mathbf{b}^l) + \mathbf{b}^{l+1}], \quad (3)$$

where the superscript l denotes parameters or feature representations of layer l and $\mathbf{W} \in \mathbb{R}^{F \times F}$ and $\mathbf{b} \in \mathbb{R}^F$ are parameters. Two different activation functions, one for rectified linear units (ReLU)[14] and a self-normalizing[15] inverse hyperbolic sine (snasinh),[16] are used in the residual blocks. Use of residual blocks improves the flow of gradients between layers[13,17] and helps alleviate the "vanishing gradients

problem."[18] After the last residual block, the final output is obtained from

$$y = C \times \text{sig}(\mathbf{W}^o\mathbf{x}^l + b^o), \quad (4)$$

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 $\mathbf{W}^o \in \mathbb{R}^{F \times 1}$ and $b^o \in \mathbb{R}$ corresponding to the output layer and the hidden features $\mathbf{x}^l$ obtained after the last residual block, respectively.

The initial states of the reactants are described by $N_{in} = 12$ input features $\mathbf{f}$, 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) and (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) were defined as

$$L_f = \frac{1}{N}\sum_{1}^{N}[\log(y' + 1.0) - \log(y' + |y - y'| + 1.0)]^2, \quad (5)$$

where y' and y 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
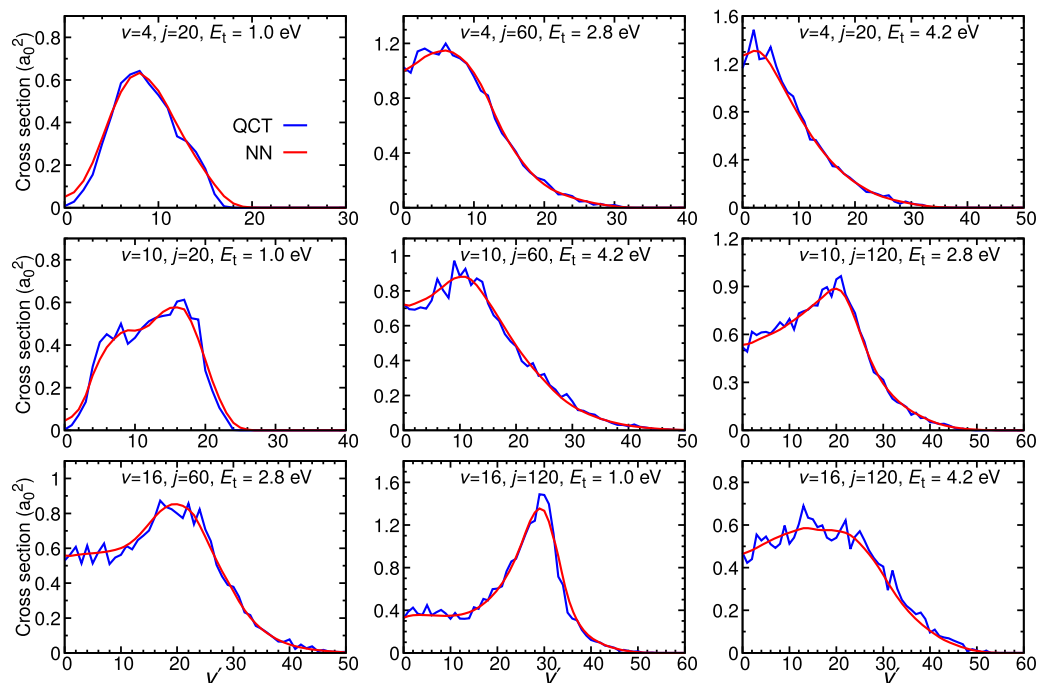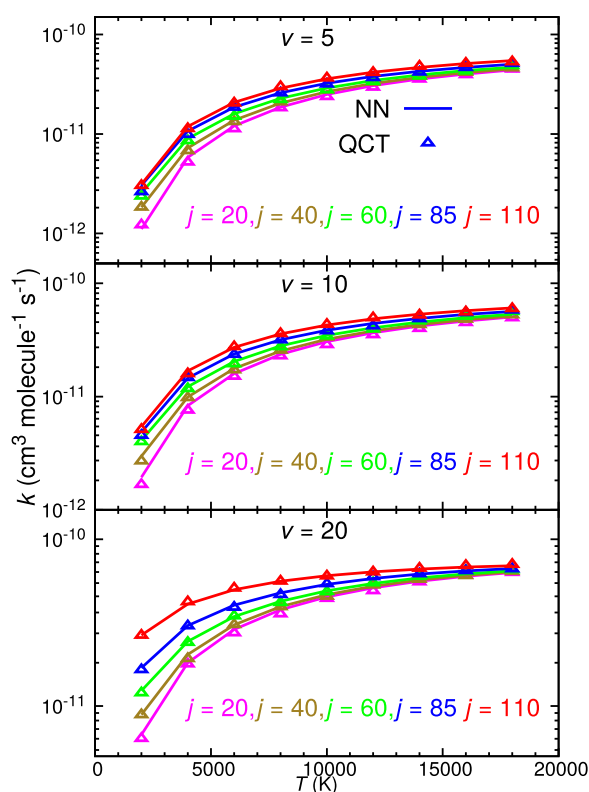

FIG. 3. Rotationally averaged, vibrational distributions of the cross sections, $\sigma(v)$, calculated from QCT (blue) and predicted from NN-STS (red) for the N + NO(v, j) → O + N$_2$(v') reaction.

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 were used for validation. From the remaining data, 2×10^5 values constitute the test data set. The convergence of the model with training set size is reported in Fig. 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 these 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 $\sigma_{tot}(v, j, E_t)$ (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.
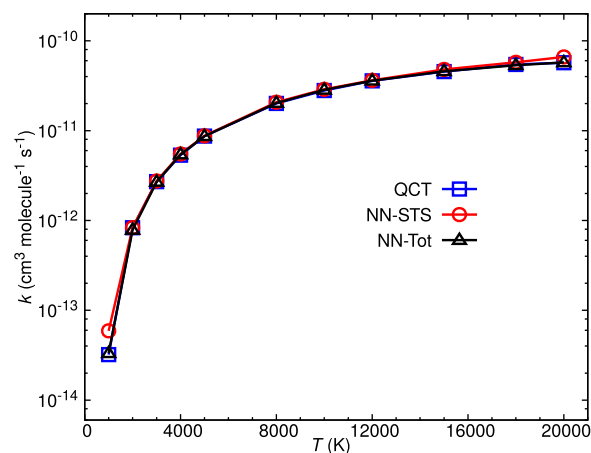
The entire reference state-to-state data set ($\sim 10^7$) has an average cross section of 0.00246 a_0^2. To test the quality of model NN-STS, 2×10^5 state-to-state cross sections were randomly chosen from the test data. The root-mean-square error (RMSE) is 0.000 328 a_0^2 for those data with a maximum deviation of 0.008 a_0^2 (for σ_{QCT} = 0.211 a_0^2, σ_{NN-STS} = 0.219 a_0^2), and the correlation coefficient is R^2 = 0.993; see Fig. S8. A similar analysis was carried out for model NN-Tot which has an average cross section of 16.18 a_0^2 for all 1232 data points. The RMSE is found to be 0.1552 a_0^2, and the correlation between NN-Tot and QCT is R^2 = 0.9997, see Fig. S9, indicative of the high quality of the fit.

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 Figs. 2 and 3 and Table S1. Again, the NN results describe the explicit QCT simulations which validate the use of such a model to predict microscopic information for such a reaction.

According to the method described in Ref. 11 (see also the supplementary material), initial state specific rates for the reaction were also calculated at temperatures between 2000 and 20 000 K for a few selected reactant states using QCT simulations and compared with the rates obtained from the NN models in Fig. 4 and Fig. 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 Fig. S11.

As another test, total thermal rates $k(T)$ were calculated from QCT simulations and compared with the results obtained from the NN models; see Fig. 5 and the supplementary material. The NN rates are calculated from integrating the NN cross sections over the (v, j)-state and translational energy phase space using Monte

FIG. 4. Initial state specific rates calculated and predicted from QCT and NN, respectively, for the N + NO(v, j) $\rightarrow$ 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 panels 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.

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 20 000 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.
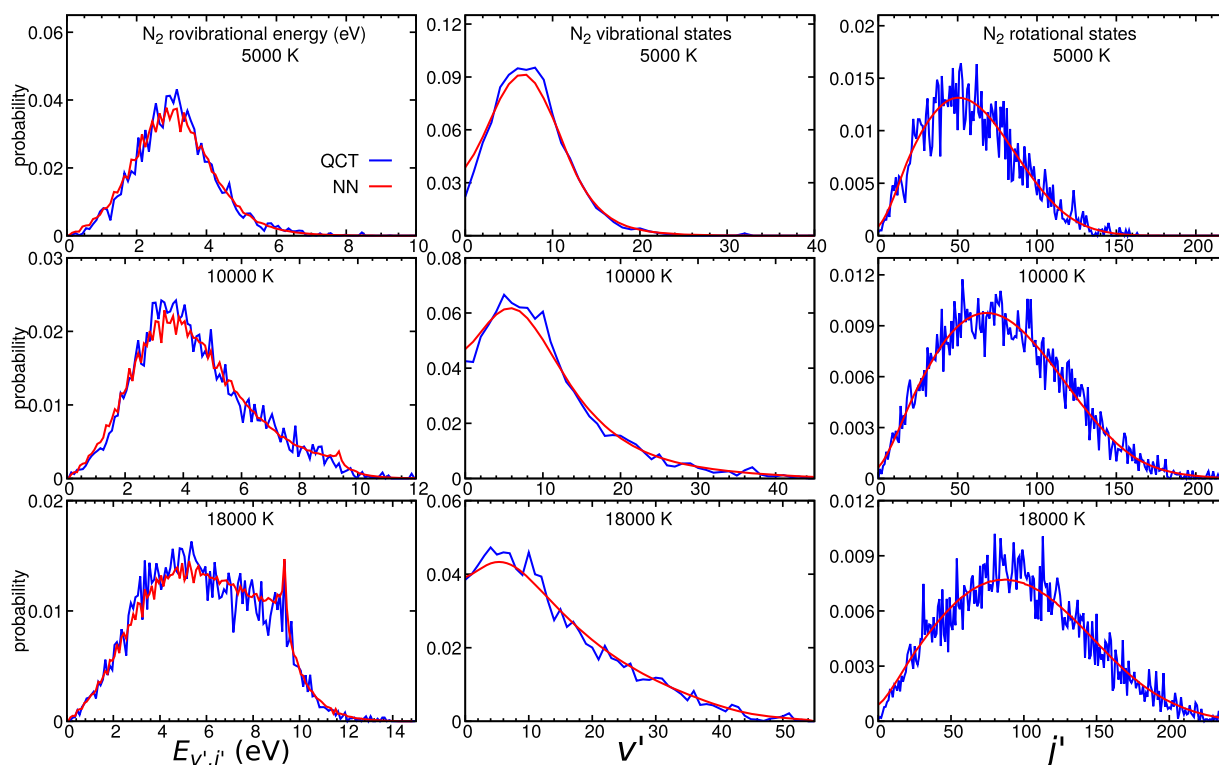
FIG. 6. Distributions of product vibrational and rotational states and rovibrational energies calculated from QCT (blue) and predicted by model NN-STS (red), respectively, for the $^3A'$ state of N + NO → O + N_2 at different temperatures.

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 20 000 K. Here, it should be mentioned that only a range of 0.05–5.5 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.

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 Fig. 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.[21]

In this work, an NN-based model for state-to-state cross sections has been constructed. For this purpose, a total of ~8 × 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 on a 64 bit 2.40 GHz Intel E5-2620 v3 central processing unit (CPU) workstation using 8 processors, and the evaluation time of the NN for 10^6 state-to-state cross sections is 24.4 s on the same computer using only 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 ~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.[22] Unfortunately, the error for the state-to-state cross sections is not reported and cannot be compared here.

In summary, the state-to-state cross sections for a reactive collision relevant to the hypersonic flight regime has been successfully modeled 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 computations.

See the supplementary material for the methodologies used to compute QCT and NN rate coefficients, Figs. S1–S11 and Table S1.

The codes for training the NN-model and computing the cross sections from the NN-model are freely available from https://github.com/MeuwlyGroup/NNcross together with a sample training data set.

REFERENCES

[1]L. 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, J. Quant. Spectrosc. Radiat. Transfer **96**, 139 (2005).

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

[3]R. 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).

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

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

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

[7]D. G. Truhlar and J. T. Muckerman, in *Atom—Molecule Collision Theory*, edited by R. B. Bernstein (Springer US, 1979), pp. 505–566.

[8]N. E. Henriksen and F. Y. Hansen, *Theories of Molecular Reaction Dynamics* (Oxford, 2011).

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

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

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

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

[13]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), p. 770.

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

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

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

[17]K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European Conference on Computer Vision*, e-print arXiv:1603.05027.

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

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

[20]M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, Software available from tensorflow.org.

[21]N. Singh and T. Schwartzentruber, Proc. Natl. Acad. Sci. U. S. A. **115**, 47 (2017).

[22]M. Kulakhmetov, M. Gallis, and A. Alexeenko, J. Chem. Phys. **144**, 174302 (2016).