

Classification of the HCN isomerization reaction dynamics in Ar buffer gas via machine learning

メタデータ	言語: English 出版者: AIP Publishing 公開日: 2023-10-03 キーワード (Ja): ポテンシャルエネルギー面, 機械学習, 分子間力, 異性化, 古典トラジェクトリ, 化学反応ダイナミクス, 遷移状態理論, 反応確率, 溶媒効果 キーワード (En): Potential energy surfaces, Machine learning, Intermolecular forces, Isomerization, Classical trajectory method, Chemical reaction dynamics, Transition state theory, Reaction probability, Solvent effect 作成者: Yamashita, Takefumi, Miyamura, Naoaki, Kawai, Shinnosuke メールアドレス: 所属:
URL	http://hdl.handle.net/10297/0002000035

RESEARCH ARTICLE | SEPTEMBER 28 2023

Classification of the HCN isomerization reaction dynamics in Ar buffer gas via machine learning **FREE**

Takefumi Yamashita ; Naoaki Miyamura; Shinnosuke Kawai  



J. Chem. Phys. 159, 124116 (2023)

<https://doi.org/10.1063/5.0156313>



View
Online



Export
Citation

CrossMark

Articles You May Be Interested In

Controlled subnanosecond isomerization of HCN to CNH in solution

J. Chem. Phys. (May 2005)

Investigation of the vibrational dynamics of the HCN/CNH isomers through high order canonical perturbation theory

J. Chem. Phys. (November 2000)

Simplification of the transition state concept in reactive island theory: Application to the HCN \rightleftharpoons CNH isomerization

J. Chem. Phys. (September 1994)

500 kHz or 8.5 GHz?
And all the ranges in between.

Lock-in Amplifiers for your periodic signal measurements



Find out more

 Zurich
Instruments

Classification of the HCN isomerization reaction dynamics in Ar buffer gas via machine learning

Cite as: J. Chem. Phys. 159, 124116 (2023); doi: 10.1063/5.0156313

Submitted: 28 April 2023 • Accepted: 7 September 2023 •

Published Online: 28 September 2023



View Online



Export Citation



CrossMark

Takefumi Yamashita,^{1,2,a)}  Naoaki Miyamura,¹ and Shinnosuke Kawai^{3,b)} 

AFFILIATIONS

¹Laboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

²Department of Physical Chemistry, School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-ku, Tokyo 142-8501, Japan

³Department of Chemistry, Faculty of Science, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka 422-8529, Japan

^{a)}Electronic mail: yamashita@lsbm.org

^{b)}Author to whom correspondence should be addressed: kawai.shinnosuke@shizuoka.ac.jp

ABSTRACT

The effect of the presence of Ar on the isomerization reaction $\text{HCN} \rightleftharpoons \text{CNH}$ is investigated via machine learning. After the potential energy surface function is developed based on the CCSD(T)/aug-cc-pVQZ level *ab initio* calculations, classical trajectory simulations are performed. Subsequently, with the aim of extracting insights into the reaction dynamics, the obtained reactivity, that is, whether the reaction occurs or not under a given initial condition, is learned as a function of the initial positions and momenta of all the atoms in the system. The prediction accuracy of the trained model is greater than 95%, indicating that machine learning captures the features of the phase space that affect reactivity. Machine learning models are shown to successfully reproduce reactivity boundaries without any prior knowledge of classical reaction dynamics theory. Subsequent analyses reveal that the Ar atom affects the reaction by displacing the effective saddle point. When the Ar atom is positioned close to the N atom (resp. the C atom), the saddle point shifts to the CNH (HCN) region, which disfavors the forward (backward) reaction. The results imply that analyses aided by machine learning are promising tools for enhancing the understanding of reaction dynamics.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0156313>

INTRODUCTION

Chemical reactions are dynamical processes in which atoms move from one configuration to another. Because the motion of atoms is governed by quantum mechanics, or its classical approximation, a chemical reaction can be understood as a solution to mechanical equations of motion. A molecular system may begin in a given initial state, and its evolution over time is determined by the rules of mechanics. Studies in the field of reaction dynamics¹ conducted over the last few decades have led to precise state-to-state understandings of chemical reactions. Examples of experimental evidence supporting the importance of dynamical viewpoints in studying chemical reactions include, but are not limited to, the control of reactions by selecting initial rotational and/or vibrational states of molecules,^{2–7} selectivity in biochemical synthesis,⁸ and reaction control via strong laser fields.^{9–16}

A notable finding achieved in the study of reaction dynamics is the existence of “reactivity boundaries”^{14–43} in the phase space that describes atomic motions. In classical mechanics, owing to the deterministic nature of the equations of motion, the initial positions and momenta of the atoms contained in a system uniquely determine whether the reaction will occur or not. Therefore, the phase space, an abstract space that is spanned by all the atomic positions and momenta as coordinates, is divided into two domains: one is the set of all “reactive” initial conditions that lead to a reaction, and the other is the set of all “non-reactive” initial conditions. Between these two domains lies a boundary, which is called a *reactivity boundary* in this study. It is also worth mentioning that quantum mechanical versions of reactivity boundaries have also been developed through various studies.^{15,16,41}

In this study, the reactivity of the isomerization reaction $\text{HCN} \rightleftharpoons \text{CNH}$ is investigated in the phase space under the

influence of one argon atom as a buffer gas. This isomerization reaction has long been drawing interest as one of the simplest isomerization reactions, as well as for applications in interstellar chemistry⁴⁴ and as a prototypical system for studying the phase space structures of chemical reactions.^{27,28} In gas-phase chemistry, the buffer atoms act as energy sources that activate target molecules through collisions. As the density of the system increases, they become solvents, influencing the reaction by changing the potential energy landscape experienced by the reacting molecules. The HCN–Ar cluster studied here may be regarded as a first-step model for investigating the solvent effect.

Recently, machine learning (ML) has attracted substantial attention in various fields ranging from business to the basic sciences, as studies have demonstrated that computers can solve pattern recognition problems, as humans do, via machine learning.^{45–48} Applications of machine learning to solve molecular problems are also being actively explored,^{49–51} in particular the prediction of rate constants⁵² and state-to-state cross sections.^{53,54} The performance of machine learning, which is known to occasionally outperform human recognition, would contribute to the evolution of computational analyses to improve the current understanding of molecular phenomena. Importantly, in some cases, machine learning can be used to extract essential dominant factors from data^{55–58} without resorting to any prior knowledge of physical laws.

In this study, a systematic method is presented for the analysis of the reaction dynamics to reveal the effects of the Ar atom, seen as a fundamental model for buffer gas or solvent, on the reactivity of HCN. The focus is given to the dynamics occurring in the vicinity of the saddle point lying on the potential energy surface (PES) between the reactant and product wells. The dynamics in this region is crucial to the branching of trajectories into the product and reactant wells. To explain it, the process of chemical reaction can be, for simplicity, viewed as three consecutive steps. First, the system in the reactant well is excited and climbs up the PES to reach the vicinity of the saddle point. Then, the dynamics in the region around the saddle point determine the bifurcation into the reactant or product well. Finally, the system falls into the product well (or back into the reactant well) and relaxes there. In the present work, we focus on the second step, that is, the dynamics occurring in the vicinity of the saddle point that determine the branching into the product or reactant well. The saddle region dynamics has been extensively investigated from the viewpoint of phase space geometry^{14–43} and, at least under certain conditions, it has been proved that there exist clear boundaries between reacting and non-reacting trajectories. The present work investigates how machine learning (ML) and accompanying analyses can extract such insights from trajectory data without prior knowledge of dynamics theory. Random forest (RF) classifiers learn the final state of the reaction, that is, whether the system ends in the HCN state or in the CNH state, as a function of the initial positions and momenta of the four atoms (H, C, N, and Ar). The method allows rational dimensional reductions to extract a small number of important coordinates whose values are essentially required to predict the reaction direction. After the dimensional reduction, molecular pictures are presented to interpret why those coordinates are important.

MATERIALS AND METHODS

Potential energy surface and trajectory calculation

The system is described by classical equations of motion derived from the following Hamiltonian:

$$H = \sum_i \frac{|\mathbf{p}_i|^2}{2m_i} + V(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N, \mathbf{r}_{Ar}), \quad (1)$$

where the index i refers to the atoms in the system (H, C, N, and Ar). The position of an atom is described by a three-dimensional vector $\mathbf{r}_i = (x_i, y_i, z_i)$, and its conjugate momentum is described by $\mathbf{p}_i = (p_{xi}, p_{yi}, p_{zi})$. The mass of each atom i is denoted as m_i . For the potential energy $V(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N, \mathbf{r}_{Ar})$, the following function is used in the present study:

$$V(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N, \mathbf{r}_{Ar}) = V_{\text{HCN}}(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N) + V_{\text{inter}}(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N, \mathbf{r}_{Ar}), \quad (2)$$

where $V_{\text{HCN}}(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N)$ is the intramolecular potential energy surface (PES) of HCN, for which the function in Ref. 59, constructed by fitting to spectroscopic data, is used in the present study. The second term, $V_{\text{inter}}(\mathbf{r}_H, \mathbf{r}_C, \mathbf{r}_N, \mathbf{r}_{Ar})$, describes the intermolecular interaction between HCN and Ar. To obtain V_{inter} , *ab initio* calculations were performed at the CCSD(T)/aug-cc-pVQZ level, and the intermolecular energies were obtained with counterpoise correction. The calculation points are distributed in a range including the regions close to the HCN minimum, the CNH minimum, and the saddle point between them. The concrete positions of these points are explicitly given in the supplementary material to this paper. Data of 1097 points with intermolecular energy less than 2 kJ mol⁻¹, which is regarded as the “typical” range of the Ar–HCN interaction considering the binding energy of ≈ 136 cm⁻¹ of the Ar–HCN complex, were least-squares fitted to the following analytical form:

$$V_{\text{inter}} = \sum_{i=1}^3 (A_i u_i^2 - B_i u_i) + \sum_{i \neq j} C_{ij} u_i u_j + \sum_{i=1}^3 \sum_{k=1}^3 F_{ik} u_i s_k, \quad (3)$$

$$s_1 = |\mathbf{r}_H - \mathbf{r}_C|^{-1}, \quad (4)$$

$$s_2 = |\mathbf{r}_N - \mathbf{r}_C|^{-1}, \quad (5)$$

$$s_3 = |\mathbf{r}_H - \mathbf{r}_N|^{-1}, \quad (6)$$

$$u_1 = |\mathbf{r}_{Ar} - \mathbf{r}_H|^{-6}, \quad (7)$$

$$u_2 = |\mathbf{r}_{Ar} - \mathbf{r}_C|^{-6}, \quad (8)$$

$$u_3 = |\mathbf{r}_{Ar} - \mathbf{r}_N|^{-6}, \quad (9)$$

where coefficients A_i , B_i , C_{ij} , and F_{ik} are the fitting parameters. The resulting fitting coefficients, as well as the structure and energy at the

data points used for the fitting, are available in the supplementary material to this paper. The data points are concentrated on the configurations with HCN adopting the HCN minimum structure, the CNH minimum structure, and the HCN–CNH saddle point structure. Points with displacements of $\cos \gamma$, where γ is the Jacobi angle (Fig. 2), by up to ± 0.2 from these three points were also included to cover a region around the saddle point rather than the single point. The deviation of ± 0.2 in $\cos \gamma$ roughly corresponds to the region where the dynamics is investigated in the present study, as will be expressed in Eqs. (14) and (15) below. The mean square residual error of the fitting was 0.18 kJ mol^{-1} , which is reasonably small compared to the range of the sampled intermolecular energy ($\sim \pm 2 \text{ kJ mol}^{-1}$) and the activation energy of 146 kJ mol^{-1} for the HCN \rightarrow CNH isomerization reaction. Figure 1 shows the contour plots of the PESs obtained in this study with respect to the position of the H atom for some fixed positions of the C, N, and Ar atoms.

It was previously pointed out⁶⁰ that exponential functions, rather than Lennard-Jones type functions, are required to describe the inner repulsive wall accurately. This is probably because their interest is in the dissociation reactions that occur with high energy, while the present study focuses on the isomerization reaction. Based on the accuracy of the fitting, we consider that the present PES correctly describes at least the essential features of the influence of the Ar atom on the HCN isomerization reaction. A more detailed examination of the effect of the form of repulsion is left for future work. Additionally, since we fit the intermolecular component V_{inter} separately from the intramolecular component V_{HCN} , one can replace the latter with a more recent *ab initio* potential function for HCN,⁶¹ which gives a higher barrier height of 200 kJ mol^{-1} compared to 146 kJ mol^{-1} of Ref. 59, for further quantitative investigations in the future while using the same intermolecular potential of this work. In the present work, the focus is given to the dynamics in the saddle region, which is the pivotal stage governing the branching of trajectories into the reactant or product well. While the barrier height may certainly affect the motion climbing up the PES in the reactant well toward the saddle region, the dynamics of the saddle region is determined by the local morphology of the PES near the saddle point, and its height relative to the reactant minimum does not actually play a significant role. As we fit the intermolecular potential with data points concentrated in the saddle region (see Fig. S2 in the supplementary material), we judge that the present treatment of the PES is sufficient for the purpose while leaving room for improvement in the future investigation on the reactant excitation stage.

From a given initial condition, the equations of motion were numerically integrated by the fourth-order Runge–Kutta method with variable time steps.⁶² To monitor the reaction channels, Jacobi angle γ was defined as the angle between the vector connecting the C and N atoms and that connecting the H atom from the CN mass center, as shown in Fig. 2. The linear HCN structure corresponds to $\gamma = 0$, and the CNH structure to $\gamma = \pi$. In the simulation, when the angle γ became larger than 1.5 rad , the system was judged to have fallen into the CNH product, and when the angle γ became smaller than 1.0 rad , the system was judged to have fallen into the HCN product.

Initial condition sampling

Datasets for machine learning were generated by sampling 200 000 initial conditions and simulating the trajectory to assign the reaction channel to each initial condition. The trajectory simulation was performed for the planar condition, wherein all the atoms were assumed to remain in the same plane for simplicity. Without loss of generality, the C and N atoms were initially placed on the x -axis, with their mass center at the origin (Fig. 3) and having zero velocity. The molecular plane was identified with the xy -plane. In the form of equations, the initial positions and momenta of H, C, N, and Ar atoms were sampled as follows:

$$\mathbf{r}_H = (x_H, y_H, 0), \quad \mathbf{p}_H = (p_{xH}, p_{yH}, 0), \quad (10)$$

$$\mathbf{r}_C = \left(-\frac{m_N}{m_C} x_N, 0, 0 \right), \quad \mathbf{p}_C = (-p_{xN}, -p_{yN}, 0), \quad (11)$$

$$\mathbf{r}_N = (x_N, 0, 0), \quad \mathbf{p}_N = (p_{xN}, p_{yN}, 0), \quad (12)$$

$$\mathbf{r}_{Ar} = (x_{Ar}, y_{Ar}, 0), \quad \mathbf{p}_{Ar} = (p_{xAr}, p_{yAr}, 0), \quad (13)$$

where each variable was independently and uniformly sampled in the following ranges:

$$x_H/10^{-10} \text{ m} \in [-0.65, -0.29], \quad (14)$$

$$y_H/10^{-10} \text{ m} \in [0.93, 1.29], \quad (15)$$

$$x_N/10^{-10} \text{ m} \in [0.48, 0.56], \quad (16)$$

$$x_{Ar}/10^{-10} \text{ m} \in [-2.5, 2.0], \quad (17)$$

$$y_{Ar}/10^{-10} \text{ m} \in [2.89, 3.56], \quad (18)$$

$$p_{xH}/10^{-24} \text{ kg m s}^{-1} \in [-21, 21], \quad (19)$$

$$p_{yH}/10^{-24} \text{ kg m s}^{-1} \in [-21, 21], \quad (20)$$

$$p_{xN}/10^{-24} \text{ kg m s}^{-1} \in [-46, 46], \quad (21)$$

$$p_{yN}/10^{-24} \text{ kg m s}^{-1} \in [-46, 46], \quad (22)$$

$$p_{xAr}/10^{-24} \text{ kg m s}^{-1} \in [-100, 100], \quad (23)$$

$$p_{yAr}/10^{-24} \text{ kg m s}^{-1} \in [-100, 100]. \quad (24)$$

The structure of the HCN molecule was sampled in a neighborhood of the saddle point ($\gamma \approx 1.2 \text{ rad} \approx 70^\circ$) for the isomerization.

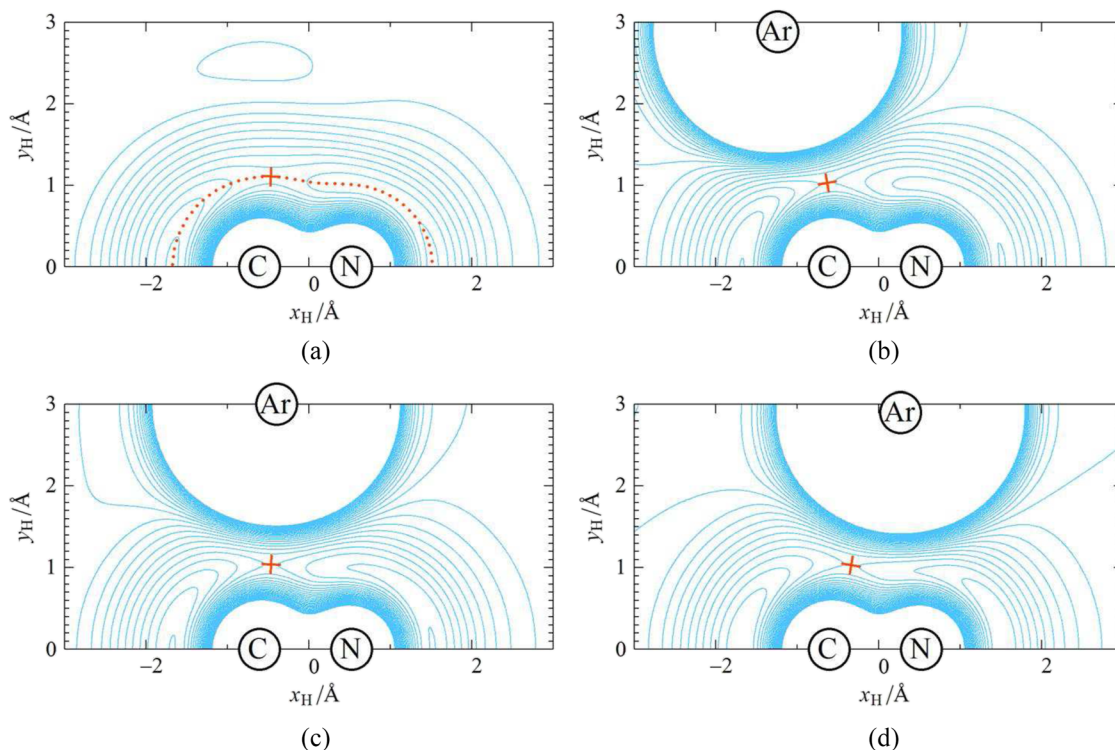


FIG. 1. Contour plots of the PES for some chosen positions of C, N, and Ar atoms. The saddle point on each surface is marked by a cross symbol whose arms show the directions of the normal modes at the saddle point. The top left panel shows the PES in the absence of Ar, and the dots depict the intrinsic reaction coordinate (IRC).

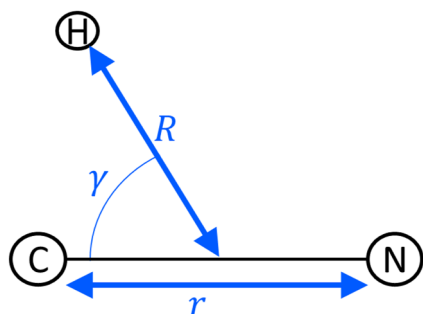


FIG. 2. Definition of Jacobi coordinates for the HCN system.

The Ar atom position was then sampled in a region near the HCN molecule. This simulated a situation in which the Ar atom collided with the HCN molecule in the middle of an isomerization reaction and affected the outcome of the proceeding reaction. Both forward and backward time propagations were calculated from each initial condition, and the reactant and product states were assigned. Consequently, there are four possible types of trajectories: forward reaction (HCN \rightarrow CNH), backward reaction (CNH \rightarrow HCN), no reaction from HCN (HCN \rightarrow HCN), and no reaction from CNH (CNH \rightarrow CNH). The type of trajectory was regarded as a function of the initial condition specified by the 11 variables ($x_H, y_H, p_{xH}, p_{yH}, x_N, p_{xN}, p_{yN}, x_{Ar}, y_{Ar}, p_{xAr},$ and p_{yAr}) as described earlier and was used as the input to the machine learning.

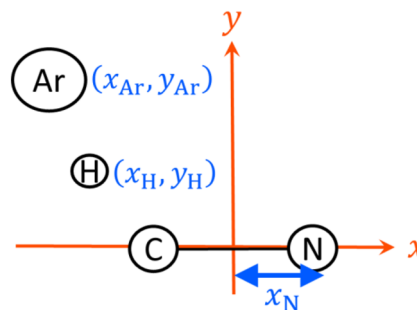


FIG. 3. Schematic illustration of the variables that parameterize the initial condition. Conjugate momenta are omitted to keep the figure simple.

In the sampling explained earlier, the center of mass of the four-atom system is not necessarily at the origin and can also have non-vanishing momentum. The sampling was performed uniformly in the rectangular region, as described earlier. This sampling method makes the explanatory variables independently distributed. This makes the interpretation of the role of each variable more accessible from the analyses described in the section titled Results and Discussion. As an example of correlated sampling, the results of microcanonical sampling are presented in [Appendix A](#).

In addition to the full-dimensional sampling, trajectory calculations on the points sampled with fixed Ar positions and momenta were also performed with the aim of providing further analyses on

the effect of the Ar atom on the reaction. Eight Ar positions were selected by seeking “representative” positions that illustrated the effect of the Ar atom as revealed by the machine learning results of the full-dimensional sample. For each Ar position, 200 000 trajectory data were generated.

Machine learning

In this study, we used a random forest (RF) classifier. RF is an ensemble learning method that constructs multiple decision trees using training data and classifies the input data based on majority voting.⁶³ The RF classifier was trained using 180 000 data points, and 20 000 data points were used for the prediction test. All the RF calculations were conducted using the scikit-learn library (ver. 0.23.2) with the following default hyperparameters:⁶⁴ number of trees ($n_{\text{estimators}}$) is 100, maximum depth of each tree (max_depth) is “none” (unlimited depth), minimum number of samples required to split an internal node (min_samples_split) is 2, minimum number of samples required to be at a leaf node (min_samples_leaf) is 1, and maximum number of features to consider for each split (max_features) is “auto” (square root of number of features). Before the test calculation, we checked the generalization performance of the trained model by using the fivefold cross-validation technique.⁶⁵ In the test calculation, we utilize “accuracy” to see the Random Forest model’s ability to correctly predict the reactivity of the HCN isomerization reaction in the presence of Ar based on the given initial conditions. Accuracy is a performance metric commonly used in machine learning to assess the model’s predictive ability. It measures the proportion of correctly predicted outcomes over the total number of predictions.

To evaluate the importance of the features in the prediction test, we employed permutation importance,⁶³ which quantifies the reduction in predictive performance resulting from the random shuffling of a specific feature. Specifically, for each feature, we randomly permuted its values across the dataset and measured the resulting decrease in prediction accuracy. This allowed us to assess the individual impact of each feature on the model’s performance.

In addition to the importance of permutation, we utilized the SHapley Additive exPlanation (SHAP) analysis, which is a game-theoretic method widely utilized for interpreting machine learning models.^{66,67} The SHAP analysis assigns a numerical value, known as the SHAP value, to each feature, representing its contribution to the prediction for a particular sample. The SHAP values provide insights into the direction and magnitude of the feature’s impact on the predicted reactivity of the HCN isomerization reaction in the presence of Ar. Positive SHAP values indicate that the feature positively contributes to the predicted reactivity, meaning that higher values of the feature are associated with an increased likelihood of the reaction occurring. Conversely, negative SHAP values indicate a negative contribution, indicating that lower values of the feature are associated with a higher likelihood of the reaction.

RESULTS AND DISCUSSION

Full dimensional learning and dimension reduction

Each sample point, specified with the 11 variables (x_{H} , y_{H} , $p_{x\text{H}}$, $p_{y\text{H}}$, x_{N} , $p_{x\text{N}}$, $p_{y\text{N}}$, x_{Ar} , y_{Ar} , $p_{x\text{Ar}}$, $p_{y\text{Ar}}$), was assigned one of

the four reaction channels ($\text{HCN} \rightarrow \text{CNH}$, $\text{CNH} \rightarrow \text{HCN}$, $\text{HCN} \rightarrow \text{HCN}$, $\text{CNH} \rightarrow \text{CNH}$) from the forward and backward trajectory calculations as explained in the last section. RF machine learning was then applied to predict the reaction channels from the values of the 11 variables. The results are shown in the first row of Table I. When all 11 variables are used, the accuracy is as high as 0.95. This indicates that the present eleven-dimensional (11D) RF model can correctly predict the reaction channels of the HCN reaction for 95% of the initial conditions.

With the aim of using the machine learning results to obtain molecular-level insights into the chemical reaction, it is here intended to eliminate some unimportant variables and predict the reaction channel by using as few variables as possible to obtain a simpler picture of the reactivity. First, we conducted a permutation analysis⁶³ to evaluate the importance of the variables (called “features” in the machine learning field). Briefly, the values of a variable were randomly permuted in the samples, and the prediction accuracy was then computed for these samples. The decrease in accuracy because of this permutation is regarded as indicating the importance of that variable for the prediction. The results shown in Fig. 4 reveal that the x -coordinate and its conjugate momentum of the H atom are the most important variables required for determining the reactivity. This is intuitively understandable because the title reaction $\text{HCN} \rightleftharpoons \text{CNH}$ can be approximately regarded as the migration of the H atom from the C side to the N side owing to the smaller mass of H compared to those of C and N. Next to the H variables, the most important variable required for the prediction is $p_{y\text{N}}$, which approximately represents the bending motion of the HCN molecule. In fact, the elimination of this variable from the RF model decreases the accuracy of prediction by 10% (10D model, the second row in Table I). Note that the 10D model was trained using the same full-dimensional data as those used for the 11D learning. This implies that the variable $p_{y\text{N}}$ has a non-negligible contribution to predicting the reactivity in the full-dimensional sample. By contrast, the decrease in accuracy induced by the elimination of x_{N} , $p_{x\text{N}}$, $p_{x\text{Ar}}$, and $p_{y\text{Ar}}$ is negligible; the 7D model (third row in Table I) shows a prediction accuracy of $\sim 95\%$. This result is consistent with the permutation importance results shown in Fig. 4, which demonstrate that the importance of x_{N} , $p_{x\text{N}}$, $p_{x\text{Ar}}$, and $p_{y\text{Ar}}$ is negligible. The elimination of $p_{y\text{N}}$ from the 7D model results in a 6D model. The fourth row in Table I shows that this model decreases the prediction accuracy to 86%, as can be expected from the results of the 10D model. The Ar positions x_{Ar} and y_{Ar} contribute to the prediction to some extent, whereas the contributions of the Ar momenta $p_{x\text{Ar}}$ and $p_{y\text{Ar}}$ are found to be negligible. The elimination of the Ar coordinates decreases the prediction accuracy by 2% (5D model, the last row in Table I). Figure 4 also shows that although the importance of the Ar position coordinates is not as high as that of $p_{y\text{N}}$, it is not negligible.

Molecular interpretation of the full dimensional results

Figure 1(a) shows the contours of the potential energy experienced by the H atom. The saddle point for the isomerization reaction and IRC are also shown. Near the saddle point, the direction of the IRC is approximately parallel to the x -direction, that is, the direction of the CN axis. Therefore, in this region, the chemical reaction

TABLE I. Accuracy of the machine learning results in the reaction channel assignment for the HCN reaction under the presence of Ar. Acc denotes the accuracy of the test, and Cross_max and Cross_min denote the maximum and minimum values of the cross-validation results, respectively.

Variables	Dimension	Acc	Cross_max	Cross_min
$x_H, y_H, p_{xH}, p_{yH}, x_N, p_{xN}, p_{yN}, x_{Ar}, y_{Ar}, p_{xAr}, p_{yAr}$	11	0.952	0.953	0.948
$x_H, y_H, p_{xH}, p_{yH}, x_N, p_{xN}, x_{Ar}, y_{Ar}, p_{xAr}, p_{yAr}$	10	0.860	0.860	0.859
$x_H, y_H, p_{xH}, p_{yH}, p_{yN}, x_{Ar}, y_{Ar}$	7	0.947	0.949	0.944
$x_H, y_H, p_{xH}, p_{yH}, x_{Ar}, y_{Ar}$	6	0.858	0.857	0.856
$x_H, y_H, p_{xH}, p_{yH}, p_{yN}$	5	0.927	0.926	0.924

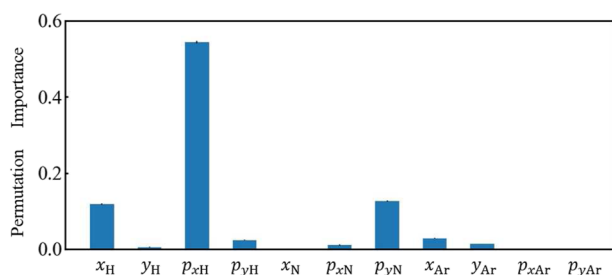


FIG. 4. Contribution of each variable to the prediction of reactivity in the full-dimensional model as evaluated via permutation importance.

almost corresponds to the movement of the H atom along the x -direction. This explains the high importance of x_H and p_{xH} as well as the low importance of y_H and p_{yH} .

To elucidate more concretely how x_H and p_{xH} affect the occurrence of the reaction, the reactivity boundaries in a section of phase space are shown in Fig. 5. For visualization, a two-dimensional plane spanned by x_H and p_{xH} is selected, with the other variables fixed at the following values: $y_H = 1.112 \text{ \AA}$, $x_N = 0.526 \text{ \AA}$, $x_{Ar} = -2.50 \text{ \AA}$, $y_{Ar} = 3.56 \text{ \AA}$, and $p_{yH} = p_{xN} = p_{yN} = p_{xAr} = p_{yAr} = 0$. As can be observed, the plane is divided into four domains in terms of the reaction channels: two reactive channels (HCN \rightarrow CNH and CNH \rightarrow HCN) and two non-reactive channels (HCN \rightarrow HCN and CNH \rightarrow CNH). The boundaries of these four regions are called the “reactivity boundaries.” In the calculation, the plane was divided into 200×200 grid points, and the boundaries were detected by assigning a reaction channel to each grid point. Reaction-channel assignment was performed both by the trajectory simulations and by the RF prediction, and their results are compared in the figure. It is seen that the RF prediction satisfactorily reproduces the reactivity boundaries obtained from trajectory simulations.

The positioning of the four regions and their boundaries shown in Fig. 5 is essentially the same as that found in previous studies conducted on dynamical chemical reaction theory.^{14–43} A large positive (negative) p_{xH} implies that the H atom has sufficiently high energy to overcome the reaction barrier from the C side to the N side (from the N side to the C side, resp.). On the other hand, x_H , the position of the H atom along the x -direction, mainly distinguishes the two non-reactive channels (HCN \rightarrow HCN and CNH \rightarrow CNH) for small $|p_{xH}|$;

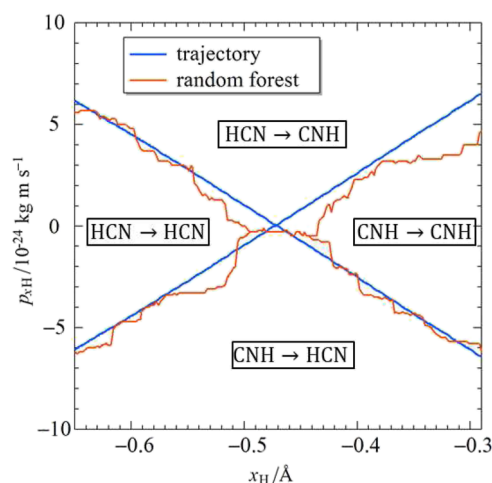


FIG. 5. Reactivity boundaries are illustrated in the two-dimensional section spanned by x_H and p_{xH} . Boundaries calculated by trajectory simulations (thick blue lines) and RF predictions (thin red lines) are compared.

hence, the kinetic energy is low and the system is confined within either the HCN or the CNH potential well.

A schematic is shown in Fig. 6 to interpret the importance of p_{yN} . Suppose a case with small H atom momentum. If CN does not move, the H atom is reflected by the reaction barrier and falls back into the HCN potential well [panel (a)]. If the N atom has some positive momentum along the y direction, the CN axis rotates counterclockwise, as shown in the figure [panel (b)]. The N atom then approaches the moving H atom. Finally, the H atom is caught by the N atom, thereby forming the CNH product. To verify this interpretation quantitatively, the reactivity boundaries are drawn in Fig. 7. The figure shows a two-dimensional section of the phase space spanned by p_{yN} and p_{xH} , with the other variables fixed at the following values: $x_H = -0.574 \text{ \AA}$, $y_H = 1.112 \text{ \AA}$, $x_N = 0.526 \text{ \AA}$, $x_{Ar} = y_{Ar} = 100 \text{ \AA}$, and $p_{yH} = p_{xN} = p_{xAr} = p_{yAr} = 0$. As observed in Fig. 5, the RF prediction satisfactorily reproduces the reactivity boundaries obtained via the trajectory simulations. It is noted that the boundaries slant downward as p_{yN} increases, enlarging the region of the HCN \rightarrow CNH reaction. This indicates that for positive p_{yN} , the HCN \rightarrow CNH

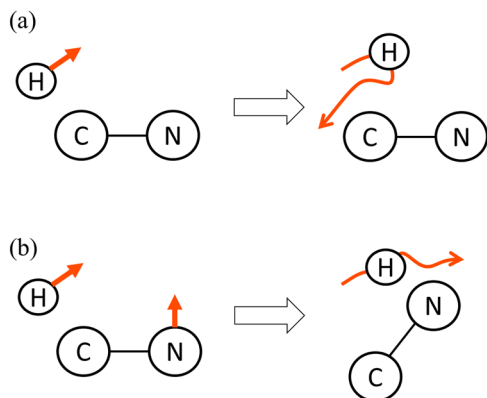


FIG. 6. Illustrations to elucidate the role of the N atom's motion in determining the reactivity. (a) Consider the case where the H atom initially has a small momentum. It is reflected back into the reactant (HCN) region by the reaction barrier. (b) For the same initial momentum of H, if the N atom is initially moving upward in the figure, the H atom can easily be captured by the N atom to form the CNH product.

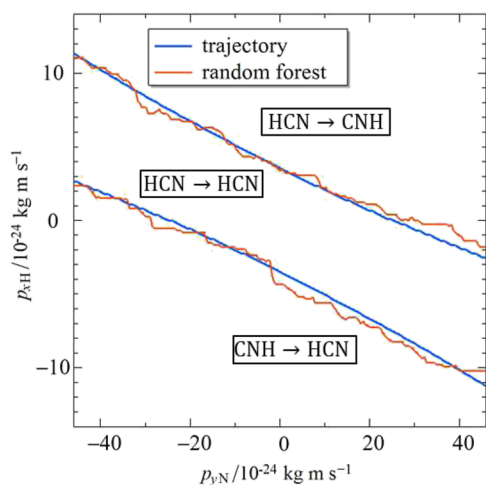


FIG. 7. Reactivity boundaries are illustrated in a two-dimensional section spanned by p_{yN} and p_{xH} . Boundaries calculated by trajectory simulations (thick blue lines) and RF predictions (thin red lines) are compared. The region of HCN \rightarrow CNH reactive initial conditions expands with increasing p_{yN} , which implies that the motion of the N atom in the positive y -direction assists the HCN \rightarrow CNH reaction.

reaction occurs more easily, which confirms the interpretation illustrated in Fig. 6.

The mechanical picture obtained here, as shown in Fig. 6, is intuitively reasonable and is computationally confirmed by plotting the reactivity boundaries as in Fig. 7. This plot is, however, possible only after one has chosen the coordinates (in this case, p_{yN} and p_{xH}) to plot the boundaries in the phase space. The machine learning helped to choose the coordinates for the plot by identifying the coordinates that have a significant influence on the reaction outcome, as done by inspecting the permutational importance in Fig. 4.

By contrast, x_N and p_{xN} , which approximately represent the CN stretching motion, are not important for predicting the reaction. This implies that the isomerization reaction is dynamically independent of the CN stretching motion.

The effect of Ar atom position analyzed by SHAP

Machine learning and the subsequent analyses via permutation importance have revealed that the presence of Ar affects the reactivity of the HCN system. To obtain further insights into the role of the Ar atom, the SHAP values^{66,67} of the variables x_{Ar} and y_{Ar} were evaluated for the 7D model here. Briefly, the SHAP value of a variable approximately describes the extent of the variable's contribution to making a prediction. Figure 8 shows the SHAP values of x_{Ar} and y_{Ar} plotted against the two-dimensional plane of (x_{Ar}, y_{Ar}) . For example, it is observed from the top right panel of Fig. 8 that the variable x_{Ar} makes a large positive contribution in predicting that the given initial condition corresponds to the CNH \rightarrow CNH channel when the Ar atom is located at the points $(x_{Ar}, y_{Ar}) \approx (-1.0 \text{ \AA}, 2.9 \text{ \AA})$.

Eight regions can be roughly recognized in the (x_{Ar}, y_{Ar}) -plane from the plots of SHAP values in Fig. 8. To clarify, the top right panel of Fig. 8 is reproduced in the left panel of Fig. 9, wherein the right panel shows the division of the left-panel figure into eight regions. Regions (F) and (G) are characterized by the greatest effect of the Ar atom position (x_{Ar}, y_{Ar}) on the reactivity. These two regions also show the most contrasting SHAP values because the effect of the Ar-atom positions experienced in the region (F) is in the opposite direction to that experienced in the region (G). For example, for the CNH \rightarrow CNH channel, the variable x_{Ar} has a large positive effect in the region (F), whereas it has a large negative effect in the region (G). Compared to regions (F) and (G), the effect of the Ar-atom position is weaker in regions (A)–(E) and (H), and the contrast between these regions is not necessarily clear. In the molecular picture, these regions correspond to the configuration wherein the Ar atom is located rather far from the HCN molecule; therefore, they may exhibit weaker intermolecular interactions.

Machine learning with single initial Ar position

To obtain further insights into the effect of the Ar-atom position, one representative point was chosen from each of the eight regions identified in the previous subsection, as shown in Fig. 9. Hereafter, these points are denoted by **a, b, . . . , h** in bold Roman letters. Machine learning was performed on datasets in which the initial position of the Ar atom was set at each of the points **a–h**, and the remaining variables were randomly sampled in the region described in the section titled Materials and Methods. The learning results obtained for different Ar-atom positions were compared to assess the influence of the Ar-atom position on the reaction. In the sampling, the initial momentum of the Ar atom was fixed at zero because the initial momentum of the Ar atom (p_{xAr}, p_{yAr}) was found to have negligible effects on reactivity in full-dimensional learning. Based on the high performance of the “7D model” with the input $(x_H, y_H, p_{xH}, p_{yH}, p_{yN}, x_{Ar}, y_{Ar})$ found for the full-dimensional data, machine learning was applied here with respect to the five variables x_H, y_H, p_{xH}, p_{yH} , and p_{yN} , since x_{Ar} and y_{Ar} were not used as variables in this analysis. In addition, machine learning was also performed for the configuration in which the Ar atom was placed sufficiently

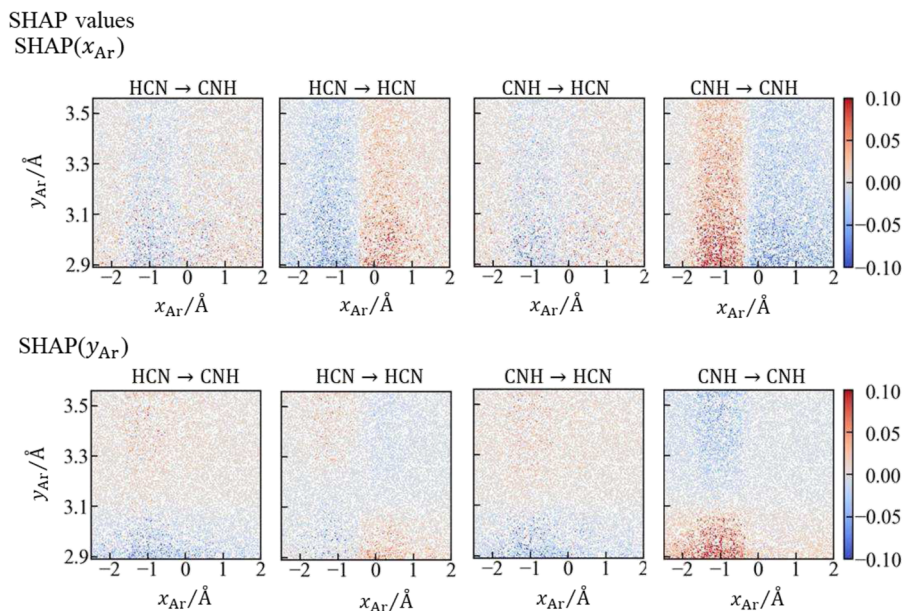


FIG. 8. The contributions of the variables x_{Ar} and y_{Ar} are evaluated by using SHAP values. The SHAP value is calculated for each point in the learning dataset. Each point is then plotted by its (x_{Ar}, y_{Ar}) values. The color indicates the SHAP value of x_{Ar} (top panel) or y_{Ar} (bottom panel) for each point.

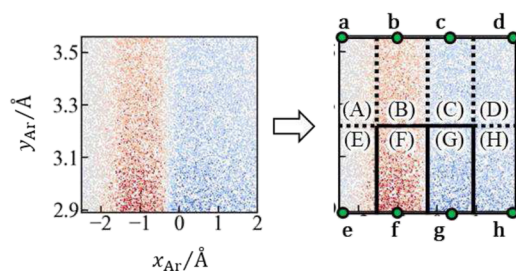


FIG. 9. The left panel reproduces the right top panel of Fig. 8. As shown in the right panel, eight regions can be recognized according to the direction and extent of the effect of Ar on the reaction.

distant from the HCN molecule so that the intermolecular interaction between them was negligible. This condition is denoted as “ Ar_{∞} .”

As shown in Table II, the machine learning results exhibit $\sim 95\%$ or higher prediction accuracy for every case, while the Ar-position dependence of accuracy is negligible. This demonstrates the consistent performance and effectiveness of the model in predicting the reactivity for each Ar position. Moreover, the permutation-importance results shown in Fig. 10 indicate high similarity between results achieved for **a–e**, **h**, and Ar_{∞} . This further supports the molecular view that regions (A)–(E) and (H) are characterized by weak intermolecular interactions, and the reactions there are almost the same as those in the isolated HCN without Ar buffer gas. By contrast, the permutation-importance results obtained for points **g** and **h** show the increased importance of the motions of the hydrogen atom in the y -direction, parameterized by y_H and p_{y_H} .

TABLE II. Accuracy of the results of machine learning performed on the HCN reaction with a single Ar initial position. Acc denotes the accuracy of the test, and Cross_max and Cross_min denote the maximum and minimum values of the cross-validation results, respectively.

	x_{Ar} (Å)	y_{Ar} (Å)	Acc	Cross_max	Cross_min
a	−2.5	3.56	0.963	0.962	0.958
b	−1.2	3.56	0.962	0.961	0.960
c	0.3	3.56	0.959	0.962	0.960
d	2.0	3.56	0.960	0.961	0.958
e	−2.5	2.89	0.963	0.961	0.958
f	−1.2	2.89	0.968	0.970	0.967
g	0.3	2.89	0.947	0.949	0.945
h	2.0	2.89	0.960	0.960	0.959
Ar_{∞}	100	0	0.961	0.960	0.959

Cross tests

To further demonstrate the distinction between points **f**, **g** and the other cases, the results of “cross tests” were examined. A cross-test means that the result of learning in one case is used to predict the reaction channel in another case, and the prediction accuracy is evaluated. As shown in Table III, the result of learning by using the dataset Ar_{∞} can predict the reaction outcome for points **a–e** and **h** with almost perfect accuracy, further supporting the view that reactions that occur in the regions (A)–(E) and (H) are essentially similar to the unimolecular reaction of isolated HCN. By contrast, the learning result obtained for the unimolecular reaction shows a low performance in the prediction of reaction outcomes for

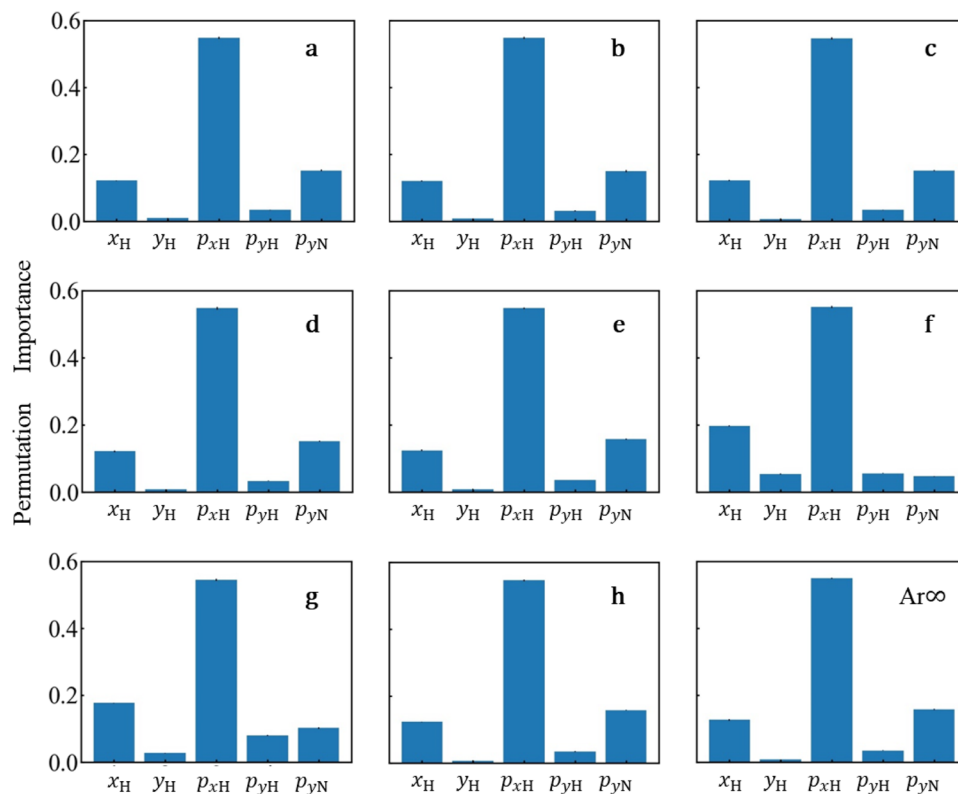


FIG. 10. Contribution of each variable to the prediction of reactivity is evaluated via permutation importance. The result is shown for each Ar-atom position defined in Fig. 9.

TABLE III. Accuracy of the random forest evaluated in the cross-prediction tests. Machine learning is performed on one dataset (denoted as “Training”), and then the result is applied to the prediction in another dataset (denoted as “Test”).

Training	Ar ∞	a	b	c	d	e	f	g	h
Test									
Ar ∞	0.961	0.960	0.961	0.960	0.961	0.960	0.616	0.772	0.961
a	0.961	0.963	0.958	0.961	0.961	0.960	0.619	0.780	0.961
b	0.960	0.960	0.962	0.958	0.959	0.961	0.632	0.765	0.959
c	0.959	0.959	0.957	0.959	0.960	0.958	0.622	0.773	0.959
d	0.960	0.962	0.959	0.959	0.960	0.961	0.624	0.773	0.960
e	0.962	0.961	0.963	0.960	0.962	0.963	0.622	0.762	0.960
f	0.617	0.616	0.624	0.613	0.619	0.622	0.968	0.479	0.616
g	0.773	0.774	0.764	0.776	0.769	0.765	0.479	0.947	0.773
h	0.961	0.960	0.959	0.959	0.960	0.959	0.613	0.768	0.960

f and **g**, and the cross-test results between **f** and **g** exhibit the worst performance.

Interestingly, the permutation importance evaluated for the prediction of **f** by the learning of **g** and vice versa (Fig. 11) shows negative values for x_H and p_{yH} . This implies that the information on x_H or p_{yH} worsens the prediction accuracy if the reactivity at **f** is predicted by using the learning results obtained for **g**. In other words,

the roles played by x_H and p_{yH} in determining the reaction channels are likely to be opposite in direction between regions (F) and (G).

Molecular interpretation for the effect of the Ar position

Figures 1(b) and 1(d) show the potential energy contours experienced by the H atom for the Ar-atom positions corresponding to points **f** and **g**, respectively. The saddle point is marked by a cross symbol whose arms show the accompanying normal directions. As can be observed, the position of the saddle point shifts along the x -direction, mainly because of the repulsive interaction with the Ar atom. Qualitatively, a saddle point on the PES divides the space into the “reactant” and “product” regions (at least approximately for sufficiently low energies). Therefore, the initial conditions wherein p_{xH} is so low that the reaction barrier cannot be overcome are classified into the HCN \rightarrow HCN (CNH \rightarrow CNH) channel if $x_H < x_{sdl}$ ($x_H > x_{sdl}$, resp.). Here, x_{sdl} denotes the x -coordinate of the saddle point. Because the value of x_{sdl} changes significantly between points **f** and **g**, classifying the trajectories via learning on one of the points results in significant errors when applied to the other point. The error may be so large that it is better not to use the information in x_H at all. This explains the negative importance found in the cross-test results obtained for **f** and **g**. This interpretation can be confirmed by plotting the reactivity boundaries, as shown in Fig. 12, for the

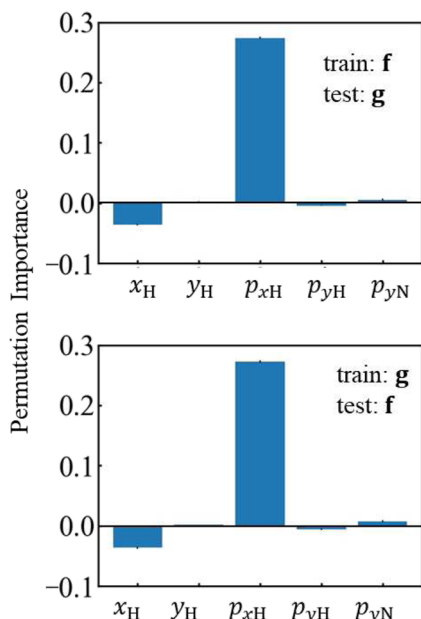


FIG. 11. Contribution of each variable is evaluated via the permutation importance when the result of learning the dataset at **f** (**g**) is applied to predict the reaction channels at **g** (**f**, resp.). A negative contribution from a variable implies that its use likely leads to the wrong prediction. In other words, the effect of the variable on the reaction is opposite in direction for points **f** and **g**.

Ar-atom positions **f** and **g**. The topological positioning of the four regions is similar to that shown in Fig. 5, which is drawn with the Ar atom located at **a** and using the learning results for full-dimensional data. A comparison of the two panels in Fig. 12 with each other

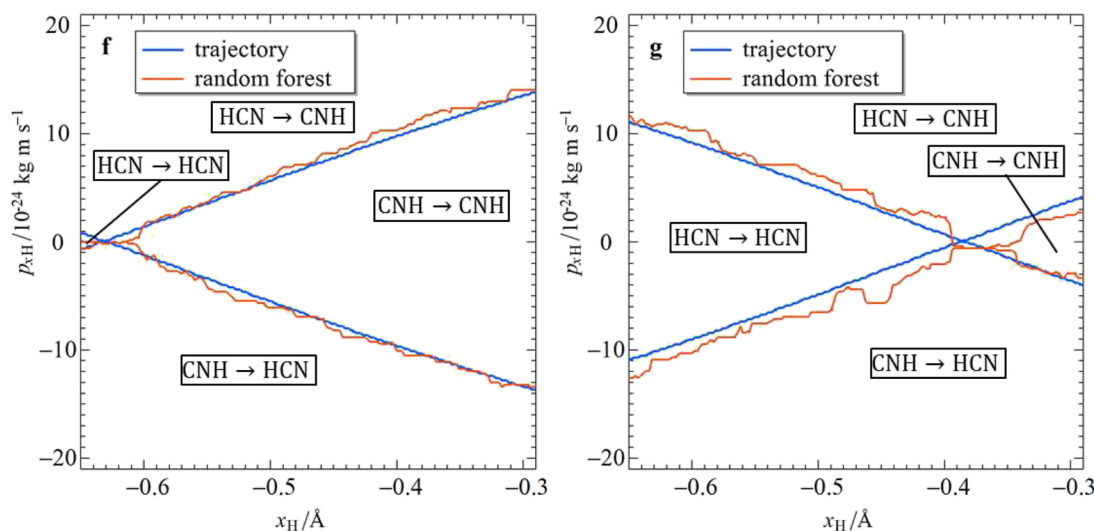


FIG. 12. Reactivity boundaries are illustrated in a two-dimensional section spanned by x_H and p_{xH} . Left panel: the Ar atom is initially positioned at the point **f**. Right panel: the Ar atom is initially positioned at the point **g**.

and with Fig. 5 reveals the shift of the reactivity boundaries in the direction of x_H in accordance with the saddle point.

As can be observed in Figs. 1(b) and 1(d), the direction of the reaction coordinate (by normal mode approximation at the saddle) has a small y -component for points **f** and **g**, whereas it is nearly parallel to the x -direction in the absence of the Ar atom. This highlights the importance of p_{yH} in determining the reactivity of points **f** and **g**. Moreover, the sign of the y -component of the reaction coordinate direction for **f** is opposite to that for **g**. This implies that the contributions of p_{yH} to the reactivity are opposite in direction at points **f** and **g**, explaining the negative importance of p_{yH} found in the cross-tests.

SUMMARY AND OUTLOOK

In the present study, the reaction dynamics of the isomerization reaction $\text{HCN} \rightleftharpoons \text{CNH}$ was investigated under the effect of an Ar atom as the buffer gas. Reaction trajectories were simulated on the potential energy function newly constructed with CCSD(T)/aug-cc-pVQZ level *ab initio* calculations. The four reaction channels ($\text{HCN} \rightarrow \text{HCN}$, $\text{HCN} \rightarrow \text{CNH}$, $\text{CNH} \rightarrow \text{CNH}$, and $\text{CNH} \rightarrow \text{HCN}$) assigned by the trajectory simulation were regarded as a function of the initial positions and momenta of all the atoms in the system. Machine learning by the RF method was performed on the data obtained by the trajectory simulation. Its prediction accuracy was $\sim 95\%$, which indicated that the RF model captured the features of the phase space affecting the reactivity. Subsequent analyses of the contributions of the variables via permutation importance and SHAP values enabled the extraction of variables essential to determine reactivity. Corresponding molecular pictures of the roles played by these variables were provided. The primary importance of the position and momentum of the H atom along the x -direction, where the x -axis is taken parallel to the CN bond, can be understood from the fact that the reaction is essentially the motion of the H atom

because of its small mass and that the IRC is nearly parallel to the x -axis near the saddle point. Machine learning results revealed that p_{yN} is also important in predicting reactivity, which corresponds to the initial rotation of the CN. This is interpreted as meaning that the motion of the H and N atoms toward each other facilitates the formation of the HN bond. Central to the interest of the present study, the effect of the Ar atom was also elucidated through the analyses of the machine-learning results, and it was successfully interpreted in terms of the displacement of the saddle point by the repulsive interaction between the H and Ar atoms.

It is noted that the machine learning found out those variables essential to the dynamics of the reaction without any prior knowledge of this reaction. The data used for the learning were only the trajectory simulation results obtained for the sampled initial conditions. Starting from the set of all the variables in the system, the machine learning and subsequent analyses extracted a reduced number of variables regarded as important for the reaction. While the molecular interpretations provided for the roles played by these variables are clear after the extraction of the variables, it was not *a priori* trivial that only those variables were of essential importance before the machine learning analyses. The present results imply that thorough statistical analyses provided by machine learning can further advance the fundamental understanding of polyatomic reaction dynamics. It is known in many reactions that differences in initial conditions lead to significant differences in reaction dynamics.^{68–73} It would be interesting to apply the machine learning analysis of the present study to elucidate the mechanism of such chemical reactions.

In this present study, we have chosen a relatively simple system of short-time passage over an isomerization saddle and focused on the applicability of machine learning to extract insights into the dynamics. For the purpose of machine learning, the initial condition was sampled uniformly in the rectangular region with the planar geometry of the four-atom system. In fact, it is shown in Appendix A that this uniform sampling is more efficient for the purpose of evaluating the importance of each variable than microcanonical sampling, which holds the total energy of the system constant. Concerning the planarity, a partial investigation of the effect of the out-of-plane motions is presented in Appendix B. In the range investigated, the results do not change the main conclusion of this study. Moreover, a complete picture of a chemical reaction beyond the passage over the saddle will include excitation in the reactant basin before entering the saddle region and relaxation in the product basin after the passage over the saddle. Indeed, in the case of the HCN isomerization considered in the present study, the system will keep passing back and forth between the HCN and CNH wells if relaxation does not occur in the well after the passage over the saddle. More complicated behaviors like fractal dependence on initial conditions may arise in those dynamics in the well or when one proceeds to analyze larger molecular systems with the methods presented in this work. Those problems would present intriguing challenges in the technical aspects of machine learning in future investigations. Technical developments in machine learning combined with the analysis methods proposed in our current study would open the way to elucidate the dynamics of such complicated systems. For example, the use of curvilinear coordinates in contrast to the cartesian coordinates as used in the present study may offer advantages in terms of interpretability. In this respect, it is encouraging that machine learning can handle analytical formulas,⁵⁶ which may enable it to find the best

curvilinear coordinates to describe the reaction dynamics. It is also noteworthy that the concept of mapping the initial phase space by analyzing the minimum dynamic path (MDP), as proposed by Unke *et al.*,⁷⁴ holds promise for enabling efficient sampling of reactive trajectories, which can further enhance the effectiveness of our machine learning-based analysis. However, the precise relationship between the MDP analysis and the present ML-based analysis is not yet clear and warrants further investigation as an intriguing topic for future research from a physical viewpoint. Since the reaction path is necessarily curvilinear, it would be interesting to compare the results of machine learning using curvilinear coordinates with their findings in future investigations.

Another interesting future direction may be to validate the present results through experiments, e.g., scattering experiments with molecular beams.¹ The energy and direction of the collision can be controlled by aligned molecular beams.^{1,68,75} The initial state of the molecule can be prepared by laser excitation. With the precise control of the initial state by these experimental techniques, the reaction outcome can be controlled by selecting the reactive or non-reactive initial conditions revealed by the present study. While the present study primarily focused on the saddle region dynamics rather than the entire reaction process including the initial excitation, the obtained insights already carry some implications for experimental investigations. For instance, it was found that the position of the Ar atom affects the branching of trajectories in the saddle region, directing them into the reactant or product well. This suggests that for the same given excitation method (e.g., laser irradiation), isomerization can be promoted by colliding the Ar atom from the C side rather than the N side of the HCN molecule (see Fig. 12). This direction would open an interesting avenue for controlling chemical reactions based on clear molecular insights.

SUPPLEMENTARY MATERIAL

See the supplementary material for the data of the electronic state calculation and the potential energy surface fitting.

ACKNOWLEDGMENTS

This research was supported by the JSPS KAKENHI (Grant No. 16KT0050). S.K. was also supported by JSPS KAKENHI (Grant No. 16K17852). Part of the computation was performed by the supercomputers of ACCMS, Kyoto University, and TSUBAME (Grant Nos. hp210141 and hp220134). T.Y. also acknowledges the support from Scientific Research [Grant Nos. JSPS KAKENHI (C) 21K03482 and (C)18K05025], the GAP Fund (UTokyo), and the Program for Promoting Research on the Supercomputer Fugaku (Application of Molecular Dynamics Simulation to Precision Medicine Using Big Data Integration System for Drug Discovery, Grant Nos. JPMXP1020200201, hp210172, and hp220164).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Takefumi Yamashita: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal). **Naoaki Miyamura:** Data curation (supporting). **Shinnosuke Kawai:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the supplementary material.

APPENDIX A: MACHINE LEARNING RESULTS FOR MICROCANONICAL SAMPLING

In the main text, machine learning was performed on data points uniformly sampled in the rectangular region of initial conditions [Eqs. (14)–(24)]. In this sampling, all the explanatory variables are independent of each other. One may consider more chemically “natural” ways of sampling such as the microcanonical ensemble, where energy conservation is taken into account, sampling with constant energy and angular momentum, or even a quasi-classical sampling, where all the quantum numbers are held constant in the initial condition. As a primary investigation for such sampling, here we report the results of machine learning on the microcanonical sampling.

To obtain sample points with a constant energy, a 60 ns-long trajectory simulation of HCN was performed, and the coordinates and momenta were extracted every 60 fs. Then 5000 points with $1.0 < \gamma < 1.5$ were randomly selected. The energy of the trajectory was set at 69.4 kJ mol^{-1} above the saddle point, which roughly allows excitation of the vibrational mode orthogonal to the reaction coordinate with one quantum number, considering that the highest-frequency normal mode at the saddle point has the frequency $3100 \text{ cm}^{-1} \approx 35 \text{ kJ mol}^{-1}$. The total angular momentum was set to zero. The momenta were re-sampled every 20 fs during the trajectory. After the microcanonical sample of HCN was

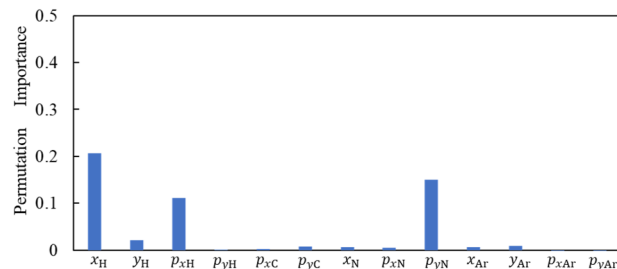


FIG. 13. Contribution of each variable to the prediction of reactivity in the microcanonical sample as evaluated via permutation importance.

obtained, the Ar atom was put at random positions in the regions $-1.24 < x_{\text{Ar}}/10^{-10} \text{ m} < 0.27$ and $2.89 < y_{\text{Ar}}/10^{-10} \text{ m} < 3.56$, and the momenta $p_{x\text{Ar}}$ and $p_{y\text{Ar}}$ were sampled by the normal distribution with a standard deviation of $2.0 \times 10^{-23} \text{ kg m s}^{-1}$.

The results of the machine learning on the microcanonical sample are summarized in Table IV, corresponding to Table I in the main text, and the values of permutation importance are shown in Fig. 13, corresponding to Fig. 4 in the main text. Somewhat counter-intuitively, a non-negligible contribution of y_{H} is found. This is in contrast to what was found on the rectangular sampling shown in the main text (Fig. 4), and it is counterintuitive from the fact that the reaction coordinate is nearly parallel to the x -direction and the motion along y is a vibrational mode orthogonal to the reaction direction [Fig. 1(a)]. This can be interpreted as follows: As the vibration along the y -direction is excited, the energy available in the reaction direction is reduced due to the constancy of the total energy. Hence, the magnitude of $p_{x\text{H}}$ is inversely correlated with y_{H} in the sample, and a higher value of y_{H} implies a smaller $|p_{x\text{H}}|$, which makes the initial condition less likely to be reactive. From this interpretation, the reactivity itself is directly affected by the value of $p_{x\text{H}}$, but due to the correlation between y_{H} and $|p_{x\text{H}}|$ existing in the sample, the machine simply finds a correlation between y_{H} and the reactivity. Therefore, for the purpose of obtaining molecular insights from the results of machine learning, it is recommended to prepare rectangular sample points so that the explanatory variables are independently distributed. Nevertheless, it would be interesting for future work to devise methodologies to extract sensible interpretations even from correlated samples.

TABLE IV. Accuracy of the machine learning results in the reaction channel assignment for the HCN reaction with microcanonically sampled initial conditions. Acc denotes the accuracy of the test, and Cross_max and Cross_min denote the maximum and minimum values of the cross-validation results, respectively.

Variables	Dimension	Acc	Cross_max	Cross_min
$x_{\text{H}}, y_{\text{H}}, p_{x\text{H}}, p_{y\text{H}}, p_{x\text{C}}, p_{y\text{C}}, x_{\text{N}}, p_{x\text{N}}, p_{y\text{N}}, x_{\text{Ar}}, y_{\text{Ar}}, p_{x\text{Ar}}, p_{y\text{Ar}}$	13	0.943	0.944	0.936
$x_{\text{H}}, y_{\text{H}}, p_{x\text{H}}, p_{y\text{H}}, p_{x\text{C}}, p_{y\text{C}}, x_{\text{N}}, p_{x\text{N}}, x_{\text{Ar}}, y_{\text{Ar}}, p_{x\text{Ar}}, p_{y\text{Ar}}$	12	0.936	0.940	0.932
$x_{\text{H}}, y_{\text{H}}, p_{x\text{H}}, p_{y\text{H}}, x_{\text{Ar}}, y_{\text{Ar}}, p_{x\text{Ar}}, p_{y\text{Ar}}$	8	0.928	0.926	0.919
$x_{\text{H}}, y_{\text{H}}, p_{x\text{H}}, p_{y\text{H}}, x_{\text{N}}, x_{\text{Ar}}, y_{\text{Ar}}$	7	0.933	0.931	0.926
$x_{\text{H}}, y_{\text{H}}, p_{x\text{H}}, p_{y\text{H}}, x_{\text{Ar}}, y_{\text{Ar}}$	6	0.933	0.933	0.926
$x_{\text{H}}, y_{\text{H}}, p_{x\text{H}}, p_{y\text{H}}$	4	0.919	0.917	0.912

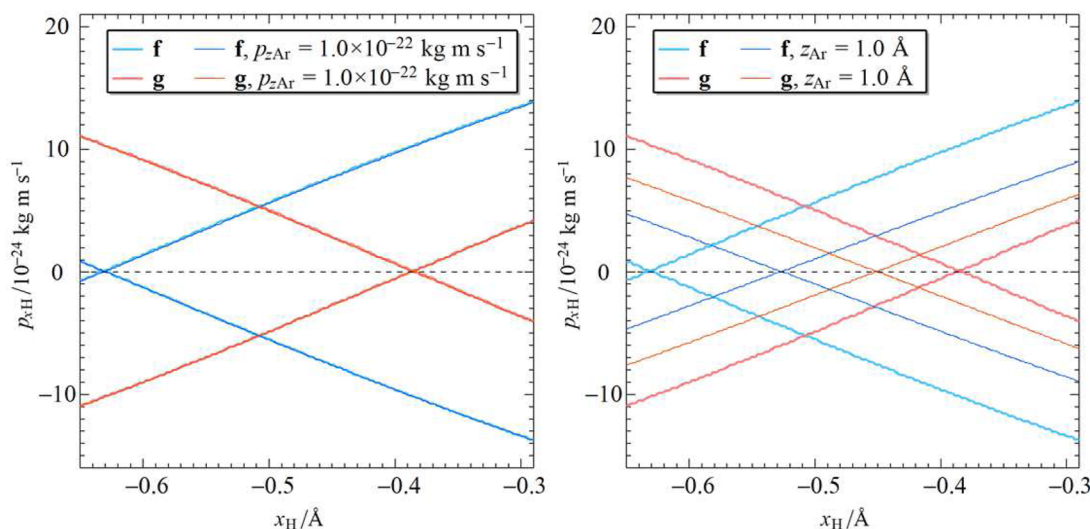


FIG. 14. Plots of the reactivity boundaries as in Fig. 12 in the main text. The Ar atom is placed at points **f** and **g**, where the effect of the Ar atom was best observed in the main text. In the left panel, the Ar atom is given initial momentum $p_{zAr} = 1.0 \times 10^{-22} \text{ kg m s}^{-1}$. In the right panel, the Ar atom is initially off the HCN plane with $z_{Ar} = 1.0 \text{ \AA}$.

APPENDIX B: EFFECTS OF THE z -DIRECTION

In the main text, the sampling was performed with all four atoms in the x - y plane [Eqs. (10)–(13)] because it is expected that the effect of the Ar atom can be best observed when it is on the same plane with HCN. In this paragraph, we show some preliminary results for the effect of displacing the Ar atom in the z -direction. Figure 14 shows plots of the reactivity boundaries in the x_H - p_{xH} plane, as in Fig. 12 in the main text. In the left panel, the Ar atom is given the initial momentum of $p_{zAr} = 1.0 \times 10^{-22} \text{ kg m s}^{-1}$, which is the same as the maximum value given to p_{xAr} and p_{yAr} in the sampling conducted in the main text [Eqs. (23) and (24)]. It is seen that the effect of the initial p_{zAr} on the reactivity boundaries is negligible. In the right panel of the same figure, the Ar atom is initially displaced off the HCN plane with $z_{Ar} = 1.0 \text{ \AA}$. Here, the reactivity boundaries for **f** and **g** approach each other and become more similar to those shown in Fig. 5, which was drawn for the case **a**, where the Ar effect is weak. This suggests that the effect of the Ar atom diminishes as it moves away from the HCN in the z -direction. Therefore, in either case, the effect of the z -direction is easily understandable. We consider that this provides a good reason to focus, at least as the first-step of the investigation, on the motion confined to the x - y plane.

REFERENCES

- R. D. Levine, *Molecular Reaction Dynamics* (Cambridge University Press, Cambridge, 2005).
- B. C. Dian, A. Longarte, and T. S. Zwier, "Conformational dynamics in a dipeptide after single-mode vibrational excitation," *Science* **296**(5577), 2369–2373 (2002).
- W. Zhang, H. Kawamata, and K. Liu, "Stretching excitation in the early barrier F + CHD₃ reaction inhibits CH bond cleavage," *Science* **325**(5938), 303–306 (2009).
- K. Nakai, H. Kono, Y. Sato, N. Niitsu, R. Sahnoun, M. Tanaka, and Y. Fujimura, "Ab initio molecular dynamics and wavepacket dynamics of highly charged

fullerene cations produced with intense near-infrared laser pulses," *Chem. Phys.* **338**(2–3), 127–134 (2007).

⁵K. Iwata, "Ultrafast bimolecular radical reaction between S₁p-terphenyl and carbon tetrachloride: Mode-specific acceleration of vibrational dephasing in reactant molecule," *J. Raman Spectrosc.* **39**(11), 1512–1517 (2008).

⁶A. Gutiérrez-González, F. F. Crim, and R. D. Beck, "Bond selective dissociation of methane (CH₃D) on the steps and terraces of Pt(211)," *J. Chem. Phys.* **149**(7), 074701 (2018).

⁷N. Gerrits, J. Geweke, D. J. Auerbach, R. D. Beck, and G.-J. Kroes, "Highly efficient activation of HCl dissociation on Au(111) via rotational preexcitation," *J. Phys. Chem. Lett.* **12**(30), 7252–7260 (2021).

⁸Y. J. Hong and D. J. Tantillo, "Biosynthetic consequences of multiple sequential post-transition-state bifurcations," *Nat. Chem.* **6**, 104–111 (2014).

⁹K. Yamanouchi, "The next Frontier," *Science* **295**(5560), 1659–1660 (2002).

¹⁰S. A. Rice and M. Zhao, *Optical Control of Molecular Dynamics* (John Wiley & Sons, New York, 2000).

¹¹M. Shapiro and P. Brumer, "Quantum control of bound and continuum state dynamics," *Phys. Rep.* **425**(4), 195–264 (2006).

¹²T. Brabec and F. Krausz, "Intense few-cycle laser fields: Frontiers of nonlinear optics," *Rev. Mod. Phys.* **72**(2), 545–591 (2000).

¹³A. D. Bandrauk, "Molecules in laser fields," in *Frontiers of Chemical Dynamics, NATO ASI Series Vol. 470*, edited by E. Yurtsever (Springer, Dordrecht, 1995), pp.131–150.

¹⁴S. Kawai, A. D. Bandrauk, C. Jaffé, T. Bartsch, J. Palacián, and T. Uzer, "Transition state theory for laser-driven reactions," *J. Chem. Phys.* **126**(16), 164306 (2007).

¹⁵S. Kawai and T. Komatsuzaki, "Quantum reaction boundary to mediate reactions in laser fields," *J. Chem. Phys.* **134**(2), 024317 (2011).

¹⁶S. Kawai and T. Komatsuzaki, "Laser control of chemical reactions by phase space structures," *Bull. Chem. Soc. Jpn.* **85**(8), 854–861 (2012).

¹⁷*Geometrical Structures of Phase Space in Multidimensional Chaos: Applications to Chemical Reaction Dynamics in Complex Systems, Advances in Chemical Physics Vols. 130A and 130B*, edited by M. Toda, T. Komatsuzaki, T. Konishi, R. S. Berry, and S. A. Rice (John Wiley & Sons, New York, 2005).

¹⁸*Advancing Theory for Kinetics and Dynamics of Complex, Many-Dimensional Systems: Clusters and Proteins, Advances in Chemical Physics Vol. 145*, edited by T. Komatsuzaki, R. S. Berry, and D. M. Leitner (John Wiley & Sons, New York, 2011).

- ¹⁹R. Hernandez and W. H. Miller, "Semiclassical transition state theory. A new perspective," *Chem. Phys. Lett.* **214**(2), 129–136 (1993).
- ²⁰R. Hernandez, "A combined use of perturbation theory and diagonalization: Application to bound energy levels and semiclassical rate theory," *J. Chem. Phys.* **101**(11), 9534–9547 (1994).
- ²¹T. Komatsuzaki and R. S. Berry, "Regularity in chaotic reaction paths. I. Ar₆," *J. Chem. Phys.* **110**(18), 9160–9173 (1999).
- ²²T. Komatsuzaki and R. S. Berry, "Dynamical hierarchy in transition states: Why and how does a system climb over the mountain?," *Proc. Natl. Acad. Sci. U. S. A.* **98**(14), 7666–7671 (2001).
- ²³S. Wiggins, L. Wiesenfeld, C. Jaffé, and T. Uzer, "Impenetrable barriers in phase-space," *Phys. Rev. Lett.* **86**(24), 5478–5481 (2001).
- ²⁴T. Uzer, C. Jaffé, J. Palacián, P. Yanguas, and S. Wiggins, "The geometry of reaction dynamics," *Nonlinearity* **15**(4), 957 (2002).
- ²⁵T. Komatsuzaki and R. S. Berry, "Chemical reaction dynamics: Many-body chaos and regularity," *Adv. Chem. Phys.* **123**, 79–152 (2002).
- ²⁶S. Kawai, H. Teramoto, C.-B. Li, T. Komatsuzaki, and M. Toda, "Dynamical reaction theory based on geometric structures in phase space," in *Advancing Theory for Kinetics and Dynamics of Complex, Many-Dimensional Systems: Clusters and Proteins, Advances in Chemical Physics Vol. 145*, edited by T. Komatsuzaki, R. S. Berry, and D. M. Leitner (John Wiley & Sons, New York, 2011), pp.123–169.
- ²⁷C.-B. Li, Y. Matsunaga, M. Toda, and T. Komatsuzaki, "Phase-space reaction network on a multisaddle energy landscape: HCN isomerization," *J. Chem. Phys.* **123**(18), 184301 (2005).
- ²⁸C.-B. Li, M. Toda, and T. Komatsuzaki, "Bifurcation of no-return transition states in many-body chemical reactions," *J. Chem. Phys.* **130**(12), 124116 (2009).
- ²⁹T. Bartsch, R. Hernandez, and T. Uzer, "Transition state in a noisy environment," *Phys. Rev. Lett.* **95**(5), 058301 (2005).
- ³⁰T. Bartsch, T. Uzer, and R. Hernandez, "Stochastic transition states: Reaction geometry amidst noise," *J. Chem. Phys.* **123**(20), 204102 (2005).
- ³¹T. Bartsch, T. Uzer, J. M. Moix, and R. Hernandez, "Identifying reactive trajectories using a moving transition state," *J. Chem. Phys.* **124**(24), 244310 (2006).
- ³²T. Bartsch, "Phase-space geometry of the generalized Langevin equation," *J. Chem. Phys.* **131**(12), 124121 (2009).
- ³³R. Hernandez, T. Uzer, and T. Bartsch, "Transition state theory in liquids beyond planar dividing surfaces," *Chem. Phys.* **370**(1-3), 270–276 (2010).
- ³⁴T. Bartsch, J. M. Moix, R. Hernandez, S. Kawai, and T. Uzer, "Time-dependent transition state theory," *Adv. Chem. Phys.* **140**, 191–238 (2008).
- ³⁵S. Kawai and T. Komatsuzaki, "Dynamic pathways to mediate reactions buried in thermal fluctuations. I. Time-dependent normal form theory for multidimensional Langevin equation," *J. Chem. Phys.* **131**(22), 224505 (2009).
- ³⁶S. Kawai and T. Komatsuzaki, "Dynamic pathways to mediate reactions buried in thermal fluctuations. II. Numerical illustrations using a model system," *J. Chem. Phys.* **131**(22), 224506 (2009).
- ³⁷S. Kawai and T. Komatsuzaki, "Hierarchy of reaction dynamics in a thermally fluctuating environment," *Phys. Chem. Chem. Phys.* **12**(27), 7626–7635 (2010).
- ³⁸S. Kawai and T. Komatsuzaki, "Nonlinear dynamical effects on reaction rates in thermally fluctuating environments," *Phys. Chem. Chem. Phys.* **12**(27), 7636–7647 (2010).
- ³⁹S. Kawai and T. Komatsuzaki, "Dynamic reaction coordinate in thermally fluctuating environment in the framework of the multidimensional generalized Langevin equations," *Phys. Chem. Chem. Phys.* **12**(47), 15382–15391 (2010).
- ⁴⁰S. Kawai and T. Komatsuzaki, "Robust existence of a reaction boundary to separate the fate of a chemical reaction," *Phys. Rev. Lett.* **105**(4), 048304 (2010).
- ⁴¹H. Waalkens, R. Schubert, and S. Wiggins, "Wigner's dynamical transition state theory in phase space: Classical and quantum," *Nonlinearity* **21**(1), R1 (2008).
- ⁴²A. Goussev, R. Schubert, H. Waalkens, and S. Wiggins, "The quantum normal form approach to reactive scattering: The cumulative reaction probability for collinear exchange reactions," *J. Chem. Phys.* **131**(14), 144103 (2009).
- ⁴³Y. Nagahata, F. Borondo, R. M. Benito, and R. Hernandez, "Identifying reaction pathways in phase space via asymptotic trajectories," *Phys. Chem. Chem. Phys.* **22**(18), 10087–10105 (2020).
- ⁴⁴J.-C. Loison, V. Wakelam, and K. M. Hickson, "The interstellar gas-phase chemistry of HCN and HNC," *Mon. Not. R. Astron. Soc.* **443**(1), 398–410 (2014).
- ⁴⁵J. Stallkamp, M. Schlipfing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks* **32**, 323–332 (2012).
- ⁴⁶M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science* **349**(6245), 255–260 (2015).
- ⁴⁷S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artif. Intell. Rev.* **26**(3), 159–190 (2006).
- ⁴⁸B. Mahesh, "Machine learning algorithms: A review," *Int. J. Sci. Res.* **9**, 381–386 (2020).
- ⁴⁹K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature* **559**(7715), 547–555 (2018).
- ⁵⁰F. Noé, A. Tkatchenko, K. R. Müller, and C. Clementi, "Machine learning for molecular simulation," *Annu. Rev. Phys. Chem.* **71**(1), 361–390 (2020).
- ⁵¹M. Haghghatari and J. Hachmann, "Advances of machine learning in molecular modeling and simulation," *Curr. Opin. Chem. Eng.* **23**, 51–57 (2019).
- ⁵²P. L. Houston, A. Nandi, and J. M. Bowman, "A machine learning approach for prediction of rate constants," *J. Phys. Chem. Lett.* **10**(17), 5250–5258 (2019).
- ⁵³D. Koner, O. T. Unke, K. Boe, R. J. Bemish, and M. Meuwly, "Exhaustive state-to-state cross sections for reactive molecular collisions from importance sampling simulation and a neural network representation," *J. Chem. Phys.* **150**(21), 211101 (2019).
- ⁵⁴J. Arnold, J. C. San Vicente Veliz, D. Koner, N. Singh, R. J. Bemish, and M. Meuwly, "Machine learning product state distributions from initial reactant states for a reactive atom–diatom collision system," *J. Chem. Phys.* **156**(3), 034301 (2022).
- ⁵⁵B. Chen, K. Huang, S. Raghupathi, I. Chandratreya, Q. Du, and H. Lipson, "Automated discovery of fundamental variables hidden in experimental data," *Nat. Comput. Sci.* **2**(7), 433–442 (2022).
- ⁵⁶S.-M. Udrescu and M. Tegmark, "AI Feynman: A physics-inspired method for symbolic regression," *Sci. Adv.* **6**(16), eaay2631 (2020).
- ⁵⁷P. Zhang, H. Shen, and H. Zhai, "Machine learning topological invariants with neural networks," *Phys. Rev. Lett.* **120**(6), 066401 (2018).
- ⁵⁸J. Liang and X. Zhu, "Phillips-inspired machine learning for band gap and exciton binding energy prediction," *J. Phys. Chem. Lett.* **10**(18), 5640–5646 (2019).
- ⁵⁹J. N. Murrell, S. Carter, and L. O. Halonen, "Frequency optimized potential energy functions for the ground-state surfaces of HCN and HCP," *J. Mol. Spectrosc.* **93**(2), 307–316 (1982).
- ⁶⁰S. P. J. Rodrigues and A. J. C. Varandas, "Dynamics study of the reaction Ar + HCN → Ar + H + CN," *J. Phys. Chem. A* **102**(31), 6266–6273 (1998).
- ⁶¹T. van Mourik, G. J. Harris, O. L. Polyansky, J. Tennyson, A. G. Császár, and P. J. Knowles, "Ab initio global potential, dipole, adiabatic, and relativistic correction surfaces for the HCN–HNC system," *J. Chem. Phys.* **115**(8), 3706–3718 (2001).
- ⁶²W. H. Press, W. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge, 2007).
- ⁶³L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
- ⁶⁴F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- ⁶⁵T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer, New York, 2009).
- ⁶⁶S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.* **30**, 4768–4777 (2017).
- ⁶⁷S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.* **2**(1), 56–67 (2020).
- ⁶⁸D. H. Parker and R. B. Bernstein, "Oriented molecule beams via the electrostatic hexapole: Preparation, characterization, and reactive scattering," *Annu. Rev. Phys. Chem.* **40**(1), 561–595 (1989).
- ⁶⁹A. Sinha, M. C. Hsiao, and F. F. Crim, "Controlling bimolecular reactions: Mode and bond selected reaction of water with hydrogen atoms," *J. Chem. Phys.* **94**(7), 4928–4935 (1991).

⁷⁰F. J. Aoiz, L. Bañares, V. J. Herrero, V. Sáez-Rábanos, K. Stark, and H.-J. Werner, "The $F + HD \rightarrow DF(HF) + H(D)$ reaction revisited: Quasiclassical trajectory study on an *ab initio* potential energy surface and comparison with molecular beam experiments," *J. Chem. Phys.* **102**(23), 9248–9262 (1995).

⁷¹F. J. Aoiz, L. Bañares, V. J. Herrero, V. S. Sáez-Rábanos, K. Stark, I. Tamarro, and H.-J. Werner, "The $F + HD$ reaction: Cross sections and rate constants on an *ab initio* potential energy surface," *Chem. Phys. Lett.* **262**(3–4), 175–182 (1996).

⁷²S. Kawai, Y. Fujimura, O. Kajimoto, and T. Yamashita, "Quasiclassical trajectory study of $O(^1D) + N_2O \rightarrow NO + NO$: Classification of reaction paths and vibrational distribution," *J. Chem. Phys.* **124**(18), 184315 (2006).

⁷³S. Kawai, Y. Fujimura, O. Kajimoto, T. Yamashita, C.-B. Li, T. Komatsuzaki, and M. Toda, "Dimension reduction for extracting geometrical structure of multidimensional phase space: Application to fast energy exchange in the reaction $O(^1D) + N_2O \rightarrow NO + NO$," *Phys. Rev. A* **75**(2), 022714 (2007).

⁷⁴O. T. Unke, S. Brickel, and M. Meuwly, "Sampling reactive regions in phase space by following the minimum dynamic path," *J. Chem. Phys.* **150**(7), 074107 (2019).

⁷⁵V. Aquilanti, M. Bartolomei, F. Pirani, D. Cappelletti, F. Vecchiocattivi, Y. Shimizu, and T. Kasai, "Orienting and aligning molecules for stereochemistry and photodynamics," *Phys. Chem. Chem. Phys.* **7**(2), 291–300 (2005).