Journal manuscript No.
(will be inserted by the editor)

Reaction Coordinates in Complex Systems – A Perspective

Jutta Rogal[a,1,2]

[1]Department of Chemistry, New York University, New York, NY 10003, USA
[2]Fachbereich Physik, Freie Universität Berlin, 14195 Berlin, Germany

Received: date / Accepted: date

Abstract In molecular simulations, the identification of suitable reaction coordinates is central to both the analysis and sampling of transitions between metastable states in complex systems. If sufficient simulation data are available, a number of methods has been developed to reduce the vast amount of high-dimensional data to a small number of essential degrees of freedom representing the reaction coordinate. Likewise, if the reaction coordinate is known, a variety of approaches has been proposed to enhance the sampling along the important degrees of freedom. Often, however, neither one nor the other is available. One of the key questions is therefore, how to construct reaction coordinates and evaluate their validity. Another challenges arises from the physical interpretation of reaction coordinates, which is often addressed by correlating physically meaningful parameters with conceptually well-defined but abstract reaction coordinates. Furthermore, machine learning based methods are becoming more and more applicable also to the reaction coordinate problem. This perspective highlights central aspects in the identification and evaluation of reaction coordinates and discusses recent ideas regarding automated computational frameworks to combine the optimization of reaction coordinates and enhanced sampling.

1 Reaction coordinates and rare events – Analysis and sampling

Molecular dynamics (MD) simulations have become an indispensable tool in the study of dynamical processes on the atomistic level. Over the years, tremendous progress has been achieved regarding the optimization and parallelization of MD algorithms on massively parallel supercomputers, enabling exceptionally large-scale simulations with up to a billion atoms [1, 2]. Furthermore, the calculation of accurate energies and forces for complex systems at affordable computational costs is continuously advanced, with machine learning (ML) potentials becoming increasingly popular [3–8]. Another challenge faced in MD simulations is the sampling and analysis of *rare* or *infrequent events*. Rare events are characterized by transitions between metastable states in phase space that require the system to overcome sizeable free energy barriers. As a consequence, the system spends most of the time within the metastable states whereas the actual transitions themselves take place on much shorter timescales. These transitions constitute, however, often the process of interest in a wide range of areas from chemistry and physics to materials science and biology, and for a multitude of systems from single molecules to condensed phase systems. Typical examples include conformational changes, protein folding, ion dissociation, nucleation, phase transformations, and defect diffusion in solids, to name a few. The limited timescales accessible in MD are, however, not the only challenge. Even if it were possible to extensively sample rare events directly with MD simulations, the analysis of the vast amount of data in the high-dimensional phase space remains far from trivial.

In order to analyze and sample rare events, the concept of a reaction coordinate (RC) is extremely useful. The terms reaction coordinate, collective variable (CV), and order parameter (OP) are not always used consistently throughout the literature. In general, these represent low-dimensional projections of the phase space. Similar to Peters [9], we refer here to any function of the full phase space coordinates as collective variable, an order parameter is a CV or combination of CVs that can distinguish between different metastable states, and a reaction coordinate is a function of CVs that accurately captures the progress of the transition between two states. The definition of RCs is essential

[a]e-mail: jutta.rogal@nyu.edu

for reaction-rate theories [10]. Within transition state theory [11, 12], an estimate of the rate constants for transitions between metastable states relies on the identification of a suitable RC, and the analysis of both the mechanisms and the kinetics of rare events can be greatly impacted by the choice of the RC.

A number of methods has been developed to facilitate the sampling of rare events that can be grouped roughly into two classes: (*i*) with and (*ii*) without applying a bias along a given set of CVs. The latter include accelerated MD approaches [13–18], replica exchange MD [19], and the family of path-based methods [20–30], which yield an ensemble of true dynamical trajectories of the transition. Methods that use a bias to enhance the sampling include umbrella sampling [31], metadynamics [32–35], hyperdynamics [36], adaptive biasing force [37], and (driven) adiabatic free energy dynamics [38–40] or temperature accelerated molecular dynamics [41], often employed to estimate free energy profiles. Approaches applying a bias can be very efficient, but rely on an *a priori* knowledge of a suitable reaction coordinate. Except for the simplest cases, an intuitive definition of an appropriate RC is, in general, highly error-prone. On the other hand, if sufficient sampling of the process of interest has been obtained, several methods, as discussed in this perspective, have been proposed to extract collective variables and assess their validity as RC. As noted in an earlier review on reaction coordinates [42], this interplay between sampling and analysis constitutes a *chicken-and-egg* problem that still poses a major challenge to the rare event community.

In this perspective, some fundamental aspects of how to construct and evaluate reaction coordinates in complex systems will be discussed. The ultimate method of choice does not only depend on the investigated problem, but also on the intended usage of the RC: should the RC provide *physical insight* into the mechanism, is it needed for sampling, is the free energy profile of interest, or can the transition state ensemble be inferred. Furthermore, ML based or supported approaches that are becoming increasingly accessible and applicable to the RC challenge will be discussed, as well as some recent ideas regarding an automated sampling and RC optimzation, before finishing with some concluding remarks.

2 What constitutes a *good* reaction coordinate?

Whether or not a reaction coordinate is considered useful depends, to a certain degree, on the question that is being asked. To obtain, for example, mechanistic insight and infer kinetic trends, a physically meaningful reaction coordinate is most helpful. However, reaction coordinates derived from intuition alone already assume specific mechanisms and require extensive knowledge about the system of interest. Even in seemingly obvious cases, an intuitive RC might not capture all important degrees of freedom, as was shown for ion pair dissociation in water more than 20 years ago [43]: here, the distance between the two ions does not provide an accurate RC, but additional solvent degrees of freedom need to be included to fully characterise the dissociation mechanism. On the other hand, mathematically derived RCs can be very accurate in describing the progress of a transition, but are often difficult to interpret.

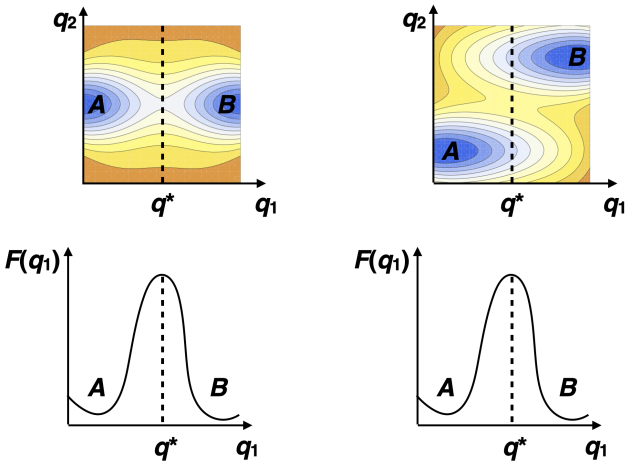

Fig. 1 Both energy landscapes on the top yield the same free energy profile and presumed transition state q^* when projected onto q_1 (bottom graphs). For the energy landscape on the left, q_1 represents a good RC and configurations at q^* are representative of the transition state. On the right energy landscape, however, configurations with $q_1 = q^*$ and small values of q_2 are fully committed to state A, whereas configuration with large q_2 values are committed to B.

The danger of projecting a high-dimensional space onto a single or a few CVs is often exemplified with the 2D energy landscapes shown in Fig. 1. When projected onto the coordinate q_1, both landscapes result in the same free energy profile, $F(q_1)$, with a maximum at q^* marking the presumed transition state (TS) for the transition between states A and B. For the energy landscape on the left, this is a suitable representation and configurations with $q_1 = q^*$ constitute the transition state ensemble. On the right energy landscape, however, configurations with $q_1 = q^*$ and small values of q_2 lie in the basin of attraction of state A, whereas configurations with large q_2 values are committed to state B. Correspondingly, configurations at q^* do not comprise the transition state ensemble and q_1 would clearly be an ill-chosen RC.

To ensure that an RC is suitable to describe the progress of a transition between two states, three criteria were proposed in Ref. [9]: (*i*) the RC is a function of only the instantaneous point in configuration space; in this definition, the velocities are not included in the RC; (*ii*) the value of the RC should change monotonically between two states and the corresponding isosurfaces yield a set of non-intersecting

dividing surfaces in configuration space; (*iii*) a free energy profile can be projected along the RC and the reduced dynamics along this RC are still consistent with the dynamics in the full phase space. These criteria are generally applicable to systems with different dynamics, ranging from overdamped to inertial dynamics.

3 The committor

The committor $p_B(\mathbf{r})$ is defined as the probability that a trajectory initiated at point $\mathbf{r}$ in configuration space will reach state B before A. It provides a statistical measure of the progress of a transition between two states and, naturally, represents an ideal reaction coordinate [44–46]. The concept has been introduced as early as the 1930s [47] and has also been termed splitting or commitment probability or p_{fold} in the case of protein folding. A brief historical recapitulation is, for example, given in Ref. [48].

Conceptually, the committor fulfils all criteria of an optimal RC, and configurations on the isocommittor surface with $p_B(\mathbf{r}) = 0.5$ constitute the transition state ensemble, as these configurations have an equal probability to either proceed to the final state B or return to the initial state A. From a practical or computational point of view, however, the committor is not quite as optimal. It lacks any direct connection to physical observables and is, thus, difficult to interpret. Furthermore, to obtain the committor, a large number of MD simulations need to be performed for each configuration $\mathbf{r}$, rendering this approach computationally rather demanding [22]. Still, the committor is extremely useful in the validation as well as in the identification of suitable RCs and is central to several approaches discussed in this perspective.

4 Evaluation of reaction coordinates

Since the committor is the ideal RC, any collective variable that represent a suitable RC must exhibit a strong correlation with the committor. More precisely, all configurations with the same CV value must lie on the same isocommittor surface. The quality of any given CV $q(\mathbf{r})$ as RC can be evaluated by computing the committor distribution [22, 23]

$$P(\hat{p}_B|\hat{q}) = \frac{\langle \delta(p_B(\mathbf{r}) - \hat{p}_B)\delta(q(\mathbf{r}) - \hat{q})\rangle}{\langle \delta(q(\mathbf{r}) - \hat{q})\rangle} \quad , \tag{1}$$

where $\delta(x)$ is the Dirac delta function and $\langle \dots \rangle$ denotes an ensemble average. If $\hat{q} = q^*$ marks the transition state, the committor distribution should be sharply peaked around $\hat{p}_B = 0.5$. For the projection of the energy landscape onto q_1 in the left graph of Fig. 1, this would be the case, whereas the committor distribution for the presumed transition state q^* on the right would yield two peaks around $\hat{p}_B = 0.0$ and 1.0 for small and large q_2 values, respectively.

The committor can also be used to *quantitatively* compare the quality of different RCs and identify the optimal RC from a given set of CVs. The first systematic approach to achieve this was based on genetic neural networks (GNN) [46]. Here, the input to the NN is a set of CVs and the output is the predicted p_B value of a given configuration. The corresponding loss function to train the NN is the root mean square (RMS) error in the predicted p_B. A genetic algorithm was employed to find the optimal combination of CVs, where again the RMS error in the predicted p_B determines the fitness of the population in each generation. With this approach, a large number of trial CVs could be compared quantitatively to identify the best approximation to the RC. The training of the NN does, however, require the computation of the committor for a large number of configurations spanning the entire range of p_B values. A variation of the method was recently suggested where, instead of the committor, the NN predicts atomic coordinates of configurations along the transition [50]. In order to improve the fitting, a second genetic algorithm was employed in an initial step to optimize the architecture of the NN.

Another approach to obtain optimal RCs is based on likelihood maximization [51, 52]. The basic idea in maximum likelihood estimation (MLE) is to find a model that best describes the observed data by maximizing the likelihood function. The data are, in this case, obtained from aimless shooting transition path sampling (TPS) simulations [51–53], and the model is a function that represents the committor, usually a sigmoid function of the RC in the range 0 to 1. In aimless shooting, a new transition path is created by selecting a configuration close to the TS from the current trajectory, randomizing the velocities, and integrating forward and backward in time. The corresponding shooting point can thus be considered as an instantaneous evaluation of the committor, and the likelihood for the shooting point data is given by [45, 51]

$$\mathscr{L} = \prod_{\mathbf{x}_i \to B} \tilde{p}_B(r_c(\mathbf{x}_i)) \prod_{\mathbf{x}_i \to A} (1 - \tilde{p}_B(r_c(\mathbf{x}_i))) \quad , \tag{2}$$

where the product runs over all shooting points $\mathbf{x}_i$ leading to states B and A, respectively, and $\tilde{p}_B$ is the committor modeled by a sigmoid function. The reaction coordinate was initially approximated as a linear combination of CVs, $r_c(\mathbf{x}) = \sum \alpha_i q_i(\mathbf{x}) + \alpha_0$, and the parameters α_i were optimized by maximizing the likelihood. The combination of CVs that maximizes the Bayesian information criterion (BIC) [54], taking into account added benefit due to increasing model complexity with larger numbers of variables, represents the best RC.

The likelihood in Eq. (2) is valid for systems with diffusive dynamics. For inertial dynamics, not only the value of the RC, but also the velocity along the RC will impact the probability to commit to the final state. The *inertial* likelihood maximization approach [55] extends the previous idea

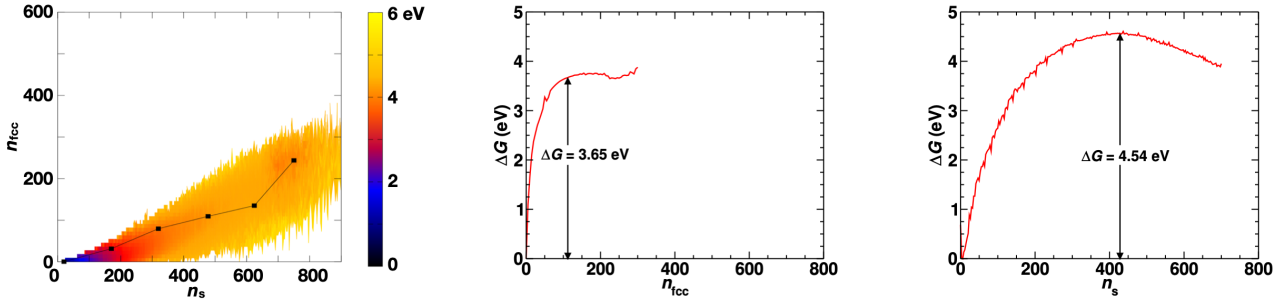

Fig. 2 Free energy during nucleation in Ni projected from the reweighted path ensemble: projection in the n_{fcc},n_s plane together with the optimized string RC (*left*), projection onto n_{fcc} (*middle*), and projection onto n_s (*right*); the increase in free energy associated with the initial formation of the precursor is not captured by the projection onto n_{fcc}. Adapted with permission from Reference [49]. Copyright 2018 American Chemical Society.

by also including the velocities in the likelihood expression. The RC, however, is still only a function of configuration space since the particle velocities solely enter through their projection onto the RC.

The use of data from the aimless shooting algorithm in Eq. (2) restricts the analysis of the RC to the TS region, whereas in complex systems information along the entire transition process might be needed to unravel the mechanism. In addition, a nonlinear approximation to the RC might be necessary instead of a simple linear combination of CVs. Both aspects can be addressed by combining maximum likelihood estimation with data from the reweighted path ensemble (RPE) and a projection onto a string in CV space that approximates the RC [56, 57]. The string connects states A and B and its position in CV space is optimized during the likelihood maximization. Again, the BIC is used to determine the best combination of CVs and string representation and identify the optimal RC.

An illustrative example of how a quantitative evaluation of proposed RCs can provide insight into the transition mechanism is nucleation from supercooled liquids. A maximum likelihood analysis of RCs describing solidification in Ni revealed that the number of solid particles in the growing cluster n_s is a much better RC than the number of face-centred cubic (fcc) particles n_{fcc} [49], even though the bulk phase crystallises in an fcc structure. The reason for this is that in the initial stage of the nucleation process a precursor is formed in the supercooled liquid from which the crystalline phase emerges, which is not captured by n_{fcc}. This precursor formation is associated with a sizeable energy barrier. Consequently, the free energy profile projected onto n_{fcc} yields a nucleation barrier that is much too low as it does not incorporate the initial pre-ordering in the liquid. The 2D projection of the free energy in the n_{fcc},n_s plane together with the optimized string RC and the projections onto the individual coordinates are shown in Fig. 2. This also demonstrates that a poor RC does not necessarily increase the apparent free energy barrier of the process, but the projection might also result in a barrier that is too low. The same effect was observed for a completely different system, examining the free energy barrier for the phosphodiester hydrolysis reaction catalyzed by the RNase H enzyme [58]. Some other examples, where a committor analysis or likelihood maximization was used to find optimal RCs, include Refs. [59–72].

The methods discussed in this section can evaluate the quality of RCs constructed from a set of proposed CVs by correlating them with the committor. If, however, an important degree of freedom cannot be captured within the provided set of CVs, it will also not appear in the analysis and will be missed.

5 Constructing reaction coordinates

There are essentially two ways to construct possible RCs: (*i*) inferred from physical intuition and/or prior knowledge about the process or similar systems; and (*ii*) inferred from analyzing extensive simulation data of the process. Some of the central ideas are highlighted below.

5.1 Intuition-based reaction coordinates

The main advantage of intuition-based RCs is that these are usually simple and have a well-defined physical interpretation. For instance, conformational changes in molecules can often be described in terms of dihedral angles, the number of native contacts is often used in protein folding, changes in local coordination or symmetry for structural transformations in solids, or cluster sizes to describe nucleation. As discussed, a CV can essentially be any function of configuration space and multiple CVs can be combined into a reaction coordinate. If the CVs comprising the RC are to be used, however, in enhanced sampling simulations, they need to be differentiable with respect to the particle positions. Another aspect to be considered when deriving intuition-based RCs

for enhanced sampling is the computational cost of obtaining the derivatives. Since forces are needed in every simulation step, expensive calculations of additional derivatives will substantially slow down the simulation. Even though useful and suitable for many systems, intuition-based RCs are always trial-and-error and their validity needs to be carefully scrutinized for every new problem.

5.2 Dimensionality reduction

The search for an adequate RC is essentially a dimensionality reduction problem from the high-dimensional phase space to a low-dimensional space of a few collective variables. The underlying assumption in applying dimensionality reduction schemes to identify RCs is, that the configurations relevant to the process can be found on a low-dimensional manifold embedded in the high-dimensional space. A number of ML based methods has been used to obtain RCs through linear and nonlinear dimensionality reduction. Several comprehensive reviews with extensive lists of references can be found in Refs. [42, 48, 73–76]. Here, we will only briefly mention frequently used approaches. The oldest and probably simplest linear dimensionality reduction method is principle component analysis (PCA), which aims to find a linear combination of variables that optimally captures the variance of the dataset. Several dimensionality reduction techniques determine a low-dimensional space in such a way that pairwise distances between points projected in the low-dimensional space reproduce the pairwise distances in the high-dimensional space. In multidimensional scaling (MDS) [77], Euclidean distances are used in the high-dimensional space. Nonlinearity is introduced by employing nonlinear functions of Euclidean distances in kernel PCA [78] or geodesic distances in Isomap [79]. Sketch-maps [80] apply a nonlinear transformation to both the distances in the high- and low-dimensional space, and can thus focus on a particular range of distances. A dynamical distance measure is used in diffusion maps [81] defining a diffusion distance in the high-dimensional space.

The coordinates obtained from dimensionality reduction are well-defined. There is, however, one major drawback: these coordinates are not directly connected to any physical variable which makes it difficult to analyse or infer mechanisms. Similar to the committor, an interpretation is often attempted by correlating a set of physically meaningful CVs with the identified coordinates that comprise the optimal RC space.

5.3 Path collective variables

A path collective variable $S(\mathbf{r})$ defines the progress along a given initial trajectory or sequence of n states in configuration space [82]

$$S(\mathbf{r}) = \frac{1}{n-1} \frac{\sum_{i=1}^{n} (i-1) \exp(-\lambda D(\mathbf{r}^{(i)}, \mathbf{r}))}{\sum_{i=1}^{n} \exp(-\lambda D(\mathbf{r}^{(i)}, \mathbf{r}))} \quad , \tag{3}$$

where $D(\mathbf{r}^{(i)}, \mathbf{r})$ is a distance metric between the reference configuration $\mathbf{r}^{(i)}$ and the instantaneous configuration $\mathbf{r}$. In addition, the function $Z(\mathbf{r})$

$$Z(\mathbf{r}) = -\frac{1}{\lambda} \ln \left(\sum_{i=1}^{n} \exp(-\lambda D(\mathbf{r}^{(i)}, \mathbf{r})) \right) \tag{4}$$

describes the distance from the reference path. $S(\mathbf{r})$ increases monotonically from 0 to 1 along the path and the hypersurfaces with a constant value of $S(\mathbf{r})$ are locally perpendicular to the path. Path CVs can be used together with enhanced sampling schemes to explore the free energy surface. The function $Z(\mathbf{r})$ can either be used to promote sampling perpendicular to the reference path or to restrain the sampling close to the proposed path.

In the original formulation [82], the distance $D(\mathbf{r}^{(i)}, \mathbf{r})$ was defined as the mean square displacement, but other metrics might be more suitable depending on the investigated system. Recently, an approach to optimize the distance metric based on a weighted combination of a set of CVs was proposed [83]. Alternatively, a reference path can also be defined in CV space, resulting in a path CV that effectively provides a nonlinear combination of CVs along the transition [84, 85]. It was furthermore shown, that an efficient sampling of the free energy can already be achieved by considering only the end-points of the path [86] in Eq. (3). In this case, just two reference configurations are needed, one in each of the metastable states.

Path CVs are particularly useful if no suitable RC can be readily defined, but an initial guess of a trajectory describing the transition can be obtained. Together with the exploration of the free energy landscape, a good approximation of the free energy is achieved even if the initial path is not optimal.

5.4 Machine learning based classification

The dimensionality reduction schemes discussed in Sec. 5.2 require extensive simulation data along the entire transition, which is not always readily available. However, if the end-states of a process are known, sampling within these metastable states is often feasible. The data can then be used in supervised ML approaches for classification to learn the decision boundary between the two states and the decision function of the classifier serves as a collective variable.

In Ref. [87], support vector machines (SVM), logistic regression (LR), and neural networks (NN) were used for classification. The trajectory data from each state are first projected onto a set of CVs that serves as the input feature vector for the ML classification. The list of CVs can

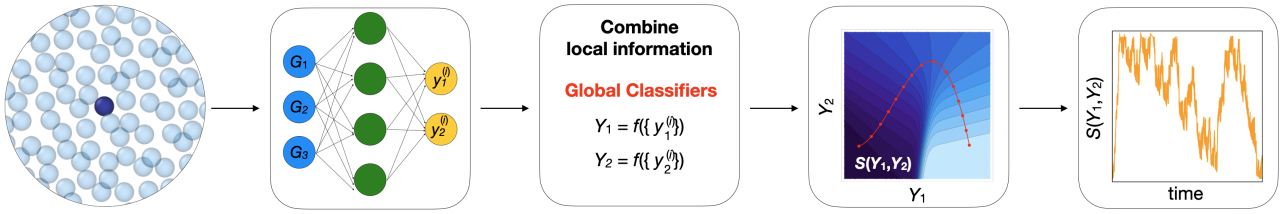

Fig. 3 Schematic representation of constructing a path CV based on local structure classification: for each atom i, a set of local descriptors G is fed into an NN that classifies the local structural environment. The local environments $y_k^{(i)}$ are combined into global classifiers Y_k and transitions are sampled along a path CV $S(\mathbf{Y})$ in the global classifier space.

be rather exhaustive as during the optimization the ML algorithms can automatically reduce the complexity of the model and highlight important features. Since the resulting RC is a combination of all input CVs, biasing along the RC will simultaneously enhance all important degrees of freedom. It should be noted that for the approach to work with enhanced sampling, analytical derivatives with respect to the atomic coordinates need to be available. Consequently, the SVM output itself was not suitable, but instead the distance to the SVM's hyperplane was used as RC. For LR and NN, the probability output can be employed directly. The approach was also extended to multiple stable states using a multiclass SVM model.

Similarly, linear discriminant analysis (LDA) has been employed to classify metastable states and construct a reaction coordinate [84]. The performance was not entirely satisfying, but could be improved by using the harmonic instead of arithmetic average of the covariance within each state to define the scatter matrix. This was rationalised by the nature of the data for metastable states of rare events, as states with smaller fluctuations should be less easy to enter and exit and should therefore have a larger weight in computing the discriminant. The LDA approach is suitable for a rather small number of input CVs and only provides a linear combination of the CVs. To alleviate these limitations, the approach was further combined with an NN [88]. A large set of physical CVs is fed into the NN which performs a nonlinear transformation of the data and reduces the dimensionality. LDA is then performed on the final layer of the NN and the RC is constructed from a projection of the NN final layer onto the LDA eigenvector.

In condensed phase systems, a local classification for each atom can also be used, for example, to describe transformations between different crystalline structures. The information for each atom can then be combined into global classifiers that characterize the state of the entire system. In Ref. [85], an NN was used for local structure classification, where the NN input features are local descriptors for each atom and the output is the probability that a given local environment corresponds to any of the investigated bulk crystalline phases. As for the other approaches discussed in this section, only trajectory data from the metastable states, in this case the bulk phases, are needed to train the NN. The global classifiers were chosen as the average over the NN output of each atom, effectively corresponding to a phase fraction. In principle, the global classifiers could directly be used as CVs in enhanced sampling. However, in this case, the global classifiers were not independent and individually biasing each component turned out to be inefficient. Instead a path CV, Eq. (3), was defined in the global classifier space to capture the transition. A schematic representation of constructing a path CV based on local structure classification is shown in Fig. 3.

In all approaches discussed in this section, it is still necessary to define descriptors as input to the ML classification method. However, these can be many and are combined in a nonlinear fashion by the ML algorithms. The resulting RC might not necessarily represent the optimal one, but is suitable to sample the transition between the metastable states, even though it is constructed from data obtained only within the metastable states. Additional data acquired from sampling the transition can subsequently be used to further refine the RC with the corresponding approaches discussed above.

6 Automated sampling and reaction coordinate optimization

The sampling of rare events and the identification of suitable RCs are strongly interdependent. For new systems, for which neither sampling data nor a good approximation to the RC are available, the initial exploration often requires significant human input based on trial-and-error and iterative sampling, data analysis, and RC improvement. Very recently, a few approaches have been proposed that *automatically* perform the sampling and optimize RCs. Two of the methods discussed in the following use autoencoders together with biased sampling in an iterative manner, the third one is based on ML predicted committors and transition path sampling. In particular for ML-based approaches, the availability of sufficient and representative data is crucial. However, for sampling rare events, data are generally sparse and

have to be generated in the first place. The approaches outlined in this section aim to iteratively sample rare events and train ML models to facilitate sampling, thus successively providing more and more data. An important aspect in these methods is the critical assessment of the convergence of the iterative workflow.

6.1 Molecular enhanced sampling with autoencoders

The central idea in molecular enhanced sampling with autoencoders (MESA) [89, 90] is an automated framework that iteratively employs an autoencoder to extract CVs from the available data and accelerated sampling in the corresponding CV space. The initial data can, for example, be taken from an initial unbiased simulation. The autoencoder optimizes the low-dimensional projection into the bottleneck layer which is used as CVs. The number of nodes in the bottleneck layer determines the dimensionality of the CV space and needs to be optimized as well. Using the trained autoencoder, the trajectory data are projected into the CV space defined by the bottleneck nodes to identify the region with non-vanishing probability density. This region is subsequently divided into a number of cells and umbrella sampling is performed in each cell. The new sampling data are again fed into the autoencoder and a new set of CVs is determined. The process is repeated until convergence is reached. It was recently remarked that the data obtained from biased simulations should be properly reweighted before being passed to the autoencoder again [91].

Some care must be taken when preparing the input data for the autoencoder from cartesian coordinates: translational and rotational invariance must be removed or the rotational invariance must be build into the NN architecture, respectively. Alternatively, for molecular systems, which were the focus of the current studies, internal degrees of freedom could be used. In condensed phase systems, additional challenges arise in the representation of the input data, as permutational invariance needs to be considered as well. Another concern are CVs that are intrinsically periodic as they increase the dimensionality of the bottleneck region. In this case, it was suggested to employ circular activation functions for pairs of coupled bottleneck nodes [90]. As in other dimensionality reduction schemes, one drawback is the interpretability of the obtained CVs. This still needs to be performed *by hand*, for example, by correlating physically meaningful CVs with the low-dimensional projection of the autoencoder.

6.2 Reweighted autoencoded variational Bayes for enhanced sampling

In its original formulation, the reweighted autoencoded variational Bayes for enhanced sampling (RAVE) approach [92, 93] employed a variational autoencoder to project trajectory data onto a single latent space coordinate. Instead of biasing directly along this coordinate, a trial RC is proposed as a linear combination of a set of CVs (as in the MLE approach, Sec. 4). The corresponding coefficients are optimized by matching the probability distribution of the trajectory data along the bottleneck coordinate and the trial RC. From the probability distribution, a bias potential is constructed along the RC and sampling is performed on the biased energy landscape. The obtained simulation data are unbiased before being used as input to the variational autoencoder in a next iteration, and a new trial RC and biasing potential are constructed. Iterations are continued until convergence of the desired properties is obtained.

RAVE was extended by estimating the predictive information bottleneck (PIB), which is then associated with the RC [94, 95]. Here, the encoder projects data at a time t onto the bottleneck coordinate, whereas the decoder predicts the data at a time $t+\Delta t$. In this respect, the RC is the coordinate that is maximally predictive of a trajectory's future evolution based on the current configuration. The encoder was chosen as a simple linear combination of the input values, so that the bottleneck coordinate is directly interpreted as the RC being a linear combination of the input CVs. The decoder was implemented as a stochastic deep NN, which allows for enough flexibility in the ML model to evaluate the usefulness of various input features. The iterative scheme remains the same, a bias potential is constructed from the probability distribution which is then used in enhanced sampling to produce a new set of data.

The remaining *human input* in RAVE is the proposed set of trial CVs that are combined into the RC. The main advantage is that the results can be directly interpreted, in particular, since the coefficients in the linear combination indicate the importance of each CV. Furthermore, the input CVs can incorporate any type of invariance inherent to the system. The obvious disadvantage is that important degrees of freedom might not be captured by the set of trial CVs. An indication for this would be if the added bias potential does not lead to any acceleration in the sampling of the rare event, necessitating the introduction of additional trial CVs.

6.3 Concurrent transition path sampling and committor prediction

In transition path sampling, an ensemble of unbiased trajectories connecting two metastable states is created by a Monte Carlo sampling in trajectory space. The efficiency of the sampling hinges on the generation of new transition paths (TP) from existing ones. Since the probability of sampling a transition path is related to the committor [96], knowledge of the committor can accelerate the generation of TPs [97].

On the basis of this, an automated framework as been proposed that learns both the committor and an optimized sampling of TPs during the simulation [98, 99]. As in the MLE approach (Sec. 4), the committor is modelled as a sigmoid function of the RC and the likelihood of the shooting points is given by Eq. (2). The negative logarithm of the likelihood is used as the loss function in training an NN, where the input is a set of CVs and the output is the predicted p_B value. Training the NN on existing shooting point data yields an optimal nonlinear combination of input CVs to represent the RC. Furthermore, the trained NN can predict p_B values for any configuration, which is used to optimize the shooting point selection for the generation of new TPs. As the sampling progresses, the NN is concurrently retrained with new shooting point data, but this is only necessary if the number of generated and expected TPs based on the predicted p_B values differ significantly.

In order to regain physical interpretability of the optimized RC, a sensitivity analysis of the NN was combined with symbolic regression [98, 99]. The sensitivity analysis filters out a small subset of CVs from all input variables to the NN. Symbolic regression is then used to identify simple mathematical expressions for this subset of CVs that best represent the nonlinear transformations encoded in the NN. Similar to RAVE, an important *human input* is the chosen feature representation of the investigated system serving as input CVs to the NN, which still requires a certain level of physical intuition about important degrees of freedom.

7 Concluding remarks

The sampling and analysis of rare events continues to pose a central challenge to the molecular simulation community. The identification of reaction coordinates to both enhance the sampling and facilitate the interpretation of transition mechanisms is a key step in this endeavour. Ideally, the task of an exhaustive sampling and analysis would be performed by fully automated computational frameworks with minimal human input. As machine learning and data-driven approaches become more and more accessible, considerable efforts have been made in this direction. Still, we are not yet at a point where we can simply provide an algorithm with one or several snapshots of a system without any additional information, in particular regarding the representation of configurations in terms of features that, for example, reflect certain symmetries and highlight important degrees of freedom. Furthermore, to derive physical interpretations and infer mechanistic trends, we generally rely on parameters that have some sort of meaning within the physical model of the investigated system. We can evaluate the importance of each parameter in describing the process and we can optimally combine them in linear and nonlinear ways, which is critical for our understanding, but we still need to come up with a set of trial parameters ourselves. Another aspect in the era data-driven science that might become beneficial is the setup of databases. Collecting information about important CVs in a wide variety of systems could potentially be used to propose sets of CVs for systems with similar characteristics, guiding the selection of trial CVs beyond human intuition.

Acknowledgements JR acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) through the Heisenberg Programme project 428315600.

References

1. J. Jung, W. Nishima, M. Daniels, G. Bascom, C. Kobayashi, A. Adedoyin, M. Wall, A. Lappala, D. Phillips, W. Fischer, C. Tung, T. Schlick, Y. Sugita, K.Y. Sanbonmatsu, J. Comput. Chem. **40**(21), 1919 (2019)
2. J. Jung, C. Kobayashi, K. Kasahara, C. Tan, A. Kuroda, K. Minami, S. Ishiduki, T. Nishiki, H. Inoue, Y. Ishikawa, M. Feig, Y. Sugita, J. Comput. Chem. **42**(4), 231 (2021)
3. J. Behler, J. Chem. Phys. **145**(17), 170901 (2016)
4. V.L. Deringer, M.A. Caro, G. Csányi, Adv. Mater. **31**(46), 1902765 (2019)
5. Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A.V. Shapeev, A.P. Thompson, M.A. Wood, S.P. Ong, J. Phys. Chem. A **124**(4), 731 (2020)
6. F. Noé, A. Tkatchenko, K.R. Müller, C. Clementi, Annu. Rev. Phys. Chem. **71**(1), 361 (2020)
7. M. Ceriotti, C. Clementi, O. Anatole von Lilienfeld, J. Chem. Phys. **154**(16), 160401 (2021)
8. J. Behler, G. Csányi, Eur. Phys. J. B **94**(7), 142 (2021)
9. B. Peters, Annu. Rev. Phys. Chem. **67**(1), 669 (2016)
10. P. Hänggi, P. Talkner, M. Borkovec, Rev. Mod. Phys. **62**(2), 251 (1990)
11. D.G. Truhlar, B.C. Garrett, S.J. Klippenstein, J. Phys. Chem. **100**(31), 12771 (1996)
12. E. Vanden-Eijnden, F.A. Tal, J. Chem. Phys. **123**(18), 184103 (2005)
13. A.F. Voter, Phys. Rev. B **57**, R13985 (1998)
14. M.R. Sørensen, A.F. Voter, J. Chem. Phys. **112**(21), 9599 (2000)
15. A.F. Voter, F. Montalenti, T.C. Germann, Annu. Rev. Mater. Res. **32**, 321 (2002)
16. D. Perez, B.P. Uberuaga, Y. Shim, J.G. Amar, A.F. Voter, (Elsevier, 2009), pp. 79–98
17. R.J. Zamora, D. Perez, E. Martinez, B.P. Uberuaga, A.F. Voter, in *Handbook of Materials Modeling*, ed. by W. Andreoni, S. Yip (Springer International Publishing, Cham, 2020), pp. 1–28

18. B.P. Uberuaga, D. Perez, A.F. Voter, in *Computational Materials, Chemistry, and Biochemistry: From Bold Initiatives to the Last Mile*, vol. 284, ed. by S. Shankar, R. Muller, T. Dunning, G.H. Chen (Springer International Publishing, Cham, 2021), pp. 137–156
19. Y. Sugita, Y. Okamoto, Chem. Phys. Lett. **314**(1-2), 141 (1999)
20. C. Dellago, P.G. Bolhuis, F.S. Csajka, D. Chandler, J. Chem. Phys. **108**(5), 1964 (1998)
21. C. Dellago, P.G. Bolhuis, D. Chandler, J. Chem. Phys. **108**(22), 9236 (1998)
22. C. Dellago, P.G. Bolhuis, P.L. Geissler, Adv. Chem. Phys. **123**, 1 (2002)
23. P.G. Bolhuis, D. Chandler, C. Dellago, P.L. Geissler, Ann. Rev. Phys. Chem. **53**(1), 291 (2002)
24. T.S. van Erp, P.G. Bolhuis, J. Comp. Phys. **205**(1), 157 (2005)
25. T.S. van Erp, Phys. Rev. Lett. **98**, 268301 (2007)
26. P.G. Bolhuis, J. Chem. Phys. **129**, 114108 (2008)
27. R.J. Allen, P.B. Warren, P.R. ten Wolde, Phys. Rev. Lett. **94**(1), 018104 (2005)
28. R.J. Allen, D. Frenkel, P.R. ten Wolde, J. Chem. Phys. **124**(2), 024102 (2006)
29. R.J. Allen, D. Frenkel, P.R. ten Wolde, J. Chem. Phys. **124**(19), 194111 (2006)
30. R.J. Allen, C. Valeriani, P.R. Ten Wolde, J. Phys.: Condens. Matter **21**(46), 463102 (2009)
31. G. Torrie, J. Valleau, J. Comput. Phys. **23**(2), 187 (1977)
32. A. Laio, M. Parrinello, Proc. Natl. Acad. Sci. USA **99**, 12562 (2002)
33. A. Laio, A. Rodriguez-Fortea, F.L. Gervasio, M. Ceccarelli, M. Parrinello, J. Phys. Chem. B **109**(14), 6714 (2005)
34. A. Laio, F.L. Gervasio, Rep. Prog. Phys. **71**(12), 126601 (2008)
35. A. Barducci, G. Bussi, M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008)
36. A.F. Voter, Phys. Rev. Lett. **78**, 3908 (1997)
37. E. Darve, D. Rodríguez-Gómez, A. Pohorille, J. Chem. Phys. **128**(14), 144120 (2008)
38. L. Rosso, P. Mináry, Z. Zhu, M.E. Tuckerman, J. Chem. Phys. **116**, 4389 (2002)
39. L. Rosso, M.E. Tuckerman, Mol. Sim. **28**(1-2), 91 (2002)
40. J.B. Abrams, M.E. Tuckerman, J. Phys. Chem. B **112**, 15742 (2008)
41. L. Maragliano, E. Vanden-Eijnden, Chem. Phys. Lett. **426**(1), 168 (2006)
42. M.A. Rohrdanz, W. Zheng, C. Clementi, Annu. Rev. Phys. Chem. **64**(1), 295 (2013)
43. P.L. Geissler, C. Dellago, D. Chandler, J. Phys. Chem. B **103**(18), 3706 (1999)
44. W. E, W. Ren, E. Vanden-Eijnden, Chem. Phys. Lett. **413**(1-3), 242 (2005)
45. R.B. Best, G. Hummer, Proc. Natl. Acad. Sci. USA **102**(19), 6732 (2005)
46. A. Ma, A.R. Dinner, J. Phys. Chem. B **109**(14), 6769 (2005)
47. L. Onsager, Phys. Rev. **54**(8), 554 (1938)
48. W. Li, A. Ma, Mol. Simul. **40**(10-11), 784 (2014)
49. G. Díaz Leines, J. Rogal, J. Phys. Chem. B **122**(48), 10934 (2018)
50. F. Hooft, A. Pérez de Alba Ortíz, B. Ensing, J. Chem. Theory Comput. **17**(4), 2294 (2021)
51. B. Peters, B.L. Trout, J. Chem. Phys. **125**(5), 054108 (2006)
52. B. Peters, G.T. Beckham, B.L. Trout, J. Chem. Phys. **127**, 034109 (2007)
53. R.G. Mullen, J.E. Shea, B. Peters, J. Chem. Theory Comput. **11**(6), 2421 (2015)
54. G. Schwarz, Ann. Stat. **6**, 461 (1978)
55. B. Peters, Chem. Phys. Lett. **554**, 248 (2012)
56. J. Rogal, W. Lechner, J. Juraszek, B. Ensing, P.G. Bolhuis, J. Chem. Phys. **133**, 174109 (2010)
57. W. Lechner, J. Rogal, J. Juraszek, B. Ensing, P.G. Bolhuis, J. Chem. Phys. **133**(17), 174110 (2010)
58. E. Rosta, H.L. Woodcock, B.R. Brooks, G. Hummer, J. Comput. Chem. **30**(11), 1634 (2009)
59. D. Moroni, P.R. ten Wolde, P.G. Bolhuis, Phys. Rev. Lett. **94**(23), 235703 (2005)
60. J. Juraszek, P.G. Bolhuis, Proc. Natl. Acad. Sci. USA **103**(43), 15859 (2006)
61. M. Grünwald, C. Dellago, J. Chem. Phys. **131**(16), 164116 (2009)
62. J. Vreede, J. Juraszek, P.G. Bolhuis, Proc. Natl. Acad. Sci. USA **107**(6), 2397 (2010)
63. G.T. Beckham, B. Peters, J. Phys. Chem. Lett. **2**(10), 1133 (2011)
64. J.D. Chodera, V.S. Pande, Phys. Rev. Lett. **107**(9), 098102 (2011)
65. W. Lechner, C. Dellago, P.G. Bolhuis, Phys. Rev. Lett. **106**(8), 085701 (2011)
66. W. Lechner, C. Dellago, P.G. Bolhuis, J. Chem. Phys. **135**(15), 154110 (2011)
67. A.J. Ballard, C. Dellago, J. Phys. Chem. B **116**(45), 13490 (2012)
68. S. Jungblut, A. Singraber, C. Dellago, Mol. Phys. **111**(22-23), 3527 (2013)
69. R.G. Mullen, J.E. Shea, B. Peters, J. Chem. Theory Comput. **10**(2), 659 (2014)
70. C. Leitold, W. Lechner, C. Dellago, J. Phys.: Condens. Matter **27**(19), 194126 (2015)
71. G. Menzl, M.A. Gonzalez, P. Geiger, F. Caupin, J.L.F. Abascal, C. Valeriani, C. Dellago, Proc Natl Acad Sci USA **113**(48), 13582 (2016)

72. Y. Liang, G. Díaz Leines, R. Drautz, J. Rogal, J. Chem. Phys. **152**(22), 224504 (2020)
73. J. Wang, A.L. Ferguson, Mol. Simul. **44**(13-14), 1090 (2018)
74. M. Ceriotti, J. Chem. Phys. **150**(15), 150901 (2019)
75. H. Sidky, W. Chen, A.L. Ferguson, Mol. Phys. **118**(5), e1737742 (2020)
76. A. Glielmo, B.E. Husic, A. Rodriguez, C. Clementi, F. Noé, A. Laio, Chem. Rev. p. acs.chemrev.0c01195 (2021)
77. M.A.A. Cox, T.F. Cox, in *Handbook of Data Visualization* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), pp. 315–347
78. B. Schölkopf, A. Smola, K.R. Müller, Neural Comput. **10**(5), 1299 (1998)
79. J.B. Tenenbaum, V. de Silva, J.C. Langford, Science **290**(5500), 2319 (2000)
80. G.A. Tribello, M. Ceriotti, M. Parrinello, Proc. Natl. Acad. Sci. USA **109**(14), 5196 (2012)
81. R.R. Coifman, I.G. Kevrekidis, S. Lafon, M. Maggioni, B. Nadler, Multiscale Model. Simul. **7**(2), 842 (2008)
82. D. Branduardi, F.L. Gervasio, M. Parrinello, J. Chem. Phys. **126**(5), 054103 (2007)
83. L. Hovan, F. Comitani, F.L. Gervasio, J. Chem. Theory Comput. **15**(1), 25 (2019)
84. D. Mendels, G.M. Piccini, M. Parrinello, J. Phys. Chem. Lett. **9**(11), 2776 (2018)
85. J. Rogal, E. Schneider, M.E. Tuckerman, Phys. Rev. Lett. **123**(24), 245701 (2019)
86. M.A. Cuendet, D.T. Margul, E. Schneider, L. Vogt-Maranto, M.E. Tuckerman, J. Chem. Phys. **149**(7), 072316 (2018)
87. M.M. Sultan, V.S. Pande, J. Chem. Phys. **149**(9), 094106 (2018)
88. L. Bonati, V. Rizzi, M. Parrinello, J. Phys. Chem. Lett. **11**(8), 2998 (2020)
89. W. Chen, A.L. Ferguson, J. Comput. Chem. **39**(25), 2079 (2018)
90. W. Chen, A.R. Tan, A.L. Ferguson, J. Chem. Phys. **149**(7), 072312 (2018)
91. Z. Belkacemi, P. Gkeka, T. Leliévre, G. Stoltz, arXiv:2104.11061 [physics.bio-ph] (2021)
92. J.M.L. Ribeiro, P. Bravo, Y. Wang, P. Tiwary, J. Chem. Phys. **149**(7), 072301 (2018)
93. J.M.L. Ribeiro, P. Tiwary, J. Chem. Theory Comput. **15**(1), 708 (2019)
94. Y. Wang, J.M.L. Ribeiro, P. Tiwary, Nat. Commun. **10**(1), 3573 (2019)
95. Y. Wang, P. Tiwary, J. Chem. Phys. **152**(14), 144102 (2020)
96. G. Hummer, J. Chem. Phys. **120**(2), 516 (2004)
97. H. Jung, K. Okazaki, G. Hummer, J. Chem. Phys. **147**(15), 152716 (2017)
98. H. Jung, R. Covino, G. Hummer, arXiv:1901.04595 [physics.chem-ph] (2019)
99. H. Jung, R. Covino, A. Arjun, P.G. Bolhuis, G. Hummer, arXiv:2105.06673 [physics.chem-ph] (2021)