



**HAL**  
open science

# An Efficient and Accurate Formalism for the Treatment of Large Amplitude Intramolecular Motion

Guillaume Reinisch, Kenji Miki, G. Vignoles, Bryan M Wong, Chris B. Simmons

► **To cite this version:**

Guillaume Reinisch, Kenji Miki, G. Vignoles, Bryan M Wong, Chris B. Simmons. An Efficient and Accurate Formalism for the Treatment of Large Amplitude Intramolecular Motion. *Journal of Chemical Theory and Computation*, 2012, 8, pp.2713 - 2724. 10.1021/ct300278x . hal-01761329

**HAL Id: hal-01761329**

**<https://hal.science/hal-01761329>**

Submitted on 9 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# An efficient and accurate formalism for the treatment of large amplitude intramolecular motion.

G. Reinisch, K. Miki, G. L. Vignoles, B. M. Wong, C. Simmons

We propose a general approach to describe Large Amplitude Motions (LAM) with multiple Degrees Of Freedom (DOF) in molecules or reaction intermediates, which is useful for the computation of thermochemical or kinetic data. The kinetic part of the LAM Lagrangian is derived using a Z-matrix internal coordinate representation within a new numerical procedure. This derivation is exact for a classical system, and the uncertainties on the prediction of observable quantities largely arise from uncertainties on the LAM Potential Energy Surface (PES). In order to rigorously account for these uncertainties, we present an approach based on Bayesian theory to infer a parametrized physical model of the PES using *ab initio* calculations. This framework allows a quantification of uncertainties associated with a PES model, as well as a propagation scheme for the final quantity of interest. A selection and generalization of some treatments accounting for the coupling of the LAM with other internal or external degrees of freedom are also presented. Finally, we discuss and validate the approach with two applications: the calculation of the partition function of the 1,3-butadiene, and the calculation of the high-pressure reaction rate of the  $\text{CH}_3 + H \rightarrow \text{CH}_4$  recombination.

## I. INTRODUCTION

Quantum chemical methods, as implemented in many software packages<sup>1,2</sup>, nowadays allow for routine calculations of various properties of molecules. Among them, thermodynamic and kinetic data are crucial for the understanding and prediction of chemical processes<sup>3</sup>.

A reliable prediction of such properties requires tackling two main issues. First, it is necessary to calculate accurate molecular energies at least for some specific geometries defining the energetic minima (stable states) and saddle points (Transition States, TS) on the Potential Energy Surface (PES). Modern *ab-initio* methods are well-adapted and efficient in addressing these issues. A second issue arises when one wishes to compute temperature-dependent properties (entropies, heat capacities, reaction rates) for which Densities Of States (DOS) and/or partition functions need to be computed. This is usually achieved using the simple Rigid Rotor Harmonic Oscillator (RRHO) approximation. Indeed, most of the internal vibrations are of small amplitude and are very well described under this approximation. However, it has been shown long ago that these approximations fail when large amplitude motions are involved<sup>4,5</sup>, and a rigorous treatment is often required if an accurate prediction of statistical properties is required<sup>6-8</sup>. This is particularly the case when a torsional motion is

present<sup>7,9-11</sup> or when a loose transition state is involved in a dissociation/recombination reaction<sup>12-14</sup>. Numerous studies have been dedicated to the development of statistical methods in the context of one or the other application, while, to our knowledge, none has been presented for the general case. In cases of torsional motions, the proposed treatments are usually derived from the one-dimensional hindered rotor (1DHR) model, originally proposed by Pitzer *et.al.*<sup>4</sup>. Recent studies<sup>7,9,11</sup> show the importance of a rigorous treatment, where quantum effects, multidimensionality, as well as coupling with other internal or external motions, can have a non-negligible influence. In the context of dissociation reactions involving a loose complex, 2- to 5-dimensional large amplitude motions need to be described. Analytical expressions for the kinetic energy of two rigid counterparts have already been derived in previous work<sup>12,13,15</sup>, and it has been shown that this formalism gives a very good approximation of the reaction rate, as long as the used PES is determined with a high level of accuracy<sup>12,16,17</sup>.

Since the kinetic energy operator has an analytic form, the main and most critical issue involved in studies of large amplitude motion is the determination of the PES<sup>18,19</sup>. Even though first principle calculations have proven their ability to yield accurate electronic energies, their computational cost does not generally allow for on-the-fly calculations<sup>20</sup> and, therefore, an appropriate interpolation method has to be used<sup>20,21</sup>. Three approaches can be distinguished: (i) local (or weighted) interpolation<sup>19,22-24</sup>, (ii) global (or non-weighted) interpolation by a set of mathematical functions<sup>16,25,26</sup>, and (iii) global interpolation by a parametrized physical model<sup>27-30</sup>. The first two solutions are general in principle but tedious to set up in practice when the dimension of the problem increases. The practical issues involved, along with the large amount of *ab initio* energy computations required, are important limitations to their utilization. The third solution is a global interpolation by a parametrized physical model of the interaction energy, and is widely referred to as a 'force field'. This particular formalism is extensively used to study the dynamics of large atomic systems<sup>31-33</sup>. Explicit introduction of the physical contributions to the interaction energy allows for a considerable decrease in the amount of information required to set up the model. However, because this simplification can introduce large model errors, it is of primary importance to estimate the confidence expected on a final Quantity of Interest (QoI). Only very recently has this issue been addressed by Cailliez and Pernot<sup>34</sup>. By using a Bayesian approach, they have calculated the parameter uncertainties associated with a Van der Waals type PES, and have evaluated their influence on the un-

certainties for the thermodynamic data. A major conclusion of their work is that the PES model error is indeed a critical feature when one wants to predict thermodynamic properties.

The purpose of this paper is twofold. First, we present a general procedure based on a Bayesian framework to infer a system-specific force field. This procedure allows the computation of Probability Density Functions (PDFs) of the model parameters using a set of *ab initio* data, and also allows their propagation to the desired QoI (here, partition function or reaction rate). The other concern of this paper is the presentation of a general approach for characterizing the statistical properties of an LAM. The restriction of previous methods to specific types of motion, largely due to an *ad hoc* derivation of the kinetic energy of the system, is overcome by introducing a new numerical method based on a functional Z-Matrix, in which the Degrees Of Freedom (DOFs) are written as functions of generalized coordinates. This approach provides several advantages: (i) the dynamical variables can be of any kind (bond length, bending or dihedral angle, reaction path-like coordinate, etc.), (ii) the geometry relaxation as well as the constraints are easily expressed in the Z-Matrix formalism, and (iii) the numerical implementation is straightforward and robust. Other issues, such as quantum effects and couplings between LAMs with the overall rotation and the internal HO, are approximately taken into account using selected methods from the literature<sup>7,11</sup>.

To illustrate the advantages and accuracies of the proposed methodology, two calculation tests are presented: (i) the partition function of 1,3-butadiene, and (ii) the high-pressure limit of the  $\text{CH}_3+\text{H} \rightarrow \text{CH}_4$  reaction rate. The presented methods are part of the currently developed C++ library openSOAMS<sup>35</sup>.

## II. THEORY

The overall methodology proposed in this paper is illustrated in figure 1. The particular steps which are emphasized in this work are (i) the inference of a PES model (parametrized by the vector  $\theta$ ) from a set of *ab initio* energies ('PES' box), (ii) the implementation of a Z-matrix coordinate system for the computation of the kinetic function ('Kinetic' box), and (iii) the Lagrangian analysis of the LAM and PES uncertainty propagation to the QoI ('Stat. & UQ' box). In the canonical ensemble, the partition function is the central quantity for computing thermodynamic data and rate constants. In order to compute the partition function, we propose to split the total Hamiltonian of a system containing an LAM into the following 4 contributions: (i) a large amplitude motion contribution, from which overall rotation and translation are rigorously removed, (ii) the small amplitude vibrations, eventually considered loosely coupled to the LAM, (iii) the overall rotation, loosely coupled to the LAM, and (iv) the overall translation, rigorously inde-

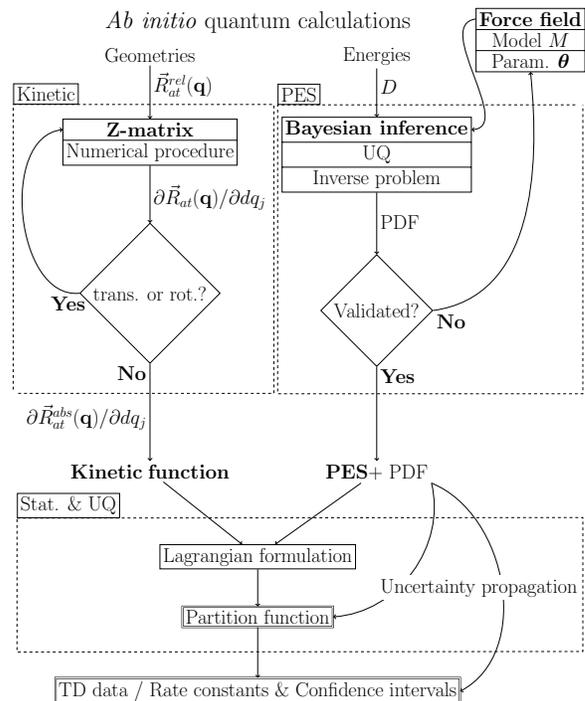


FIG. 1. Flowchart of the thermal statistical computations in our approach. Acronyms and abbreviations: PES: Potential Energy Surface; UQ: Uncertainty Quantification; PDF: Probability Density Function; TD: ThermoDynamic; param.: parameters; trans.: translation; rot.:rotation; stat.: statistic.

pendent of all the other modes. The total partition function of the system is then trivially obtained by simple multiplication of all the contributions:

$$Q_{tot}(T) = Q_{LM}(T)Q_{HO}^{coupled}(T)Q_{rot}^{coupled}(T)Q_{trans}(T) \quad (1)$$

where  $T$  is the temperature and  $Q_{LM}$ ,  $Q_{HO}^{coupled}$ ,  $Q_{rot}^{coupled}$  and  $Q_{trans}$  are the partition function of the contributions (i) to (iv), respectively. In this section, we will address the calculation of  $Q_{LM}$ ,  $Q_{HO}^{coupled}$ ,  $Q_{rot}^{coupled}$  in detail, and the partition function of the overall translation can be found elsewhere<sup>36</sup>.

### A. The large amplitude motion

Let a LAM of a molecular system be described by the positions  $\vec{R}_{at}$  of the atomic nuclei in a system of generalized coordinates  $\mathbf{q}$ . The Lagrangian of the system is written:

$$L(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) + V(\mathbf{q}) \quad (2)$$

where  $V(\mathbf{q})$  is the electronic PES and  $T(\mathbf{q}, \dot{\mathbf{q}})$  the kinetic energy, given by the relation:

$$\begin{aligned} T(\mathbf{q}, \dot{\mathbf{q}}) &= \frac{1}{2} \sum_{at} m_{at} \left( \sum_i \frac{\partial \vec{R}_{at}}{\partial q_i} \dot{q}_i \right)^2 \\ &= \frac{1}{2} \dot{\mathbf{q}}^T \underline{\mathbf{A}}(\mathbf{q}) \dot{\mathbf{q}} \end{aligned} \quad (3)$$

The matrix  $\underline{\mathbf{A}}$  is called the kinetic matrix, its elements are defined by:

$$A_{ij} = \frac{1}{2} \sum_{at} m_{at} \frac{\partial \vec{R}_{at}}{\partial q_i} \cdot \frac{\partial \vec{R}_{at}}{\partial q_j} \quad (4)$$

The theorem of Aston and Eidinoff<sup>37</sup> allows a computation of the LAM partition function through a configurational integral involving the determinant of the kinetic matrix and the PES:

$$Q(T) = \frac{1}{\delta} (2\pi k_B T)^{n/2} \int |\underline{\mathbf{A}}(\mathbf{q})|^{1/2} \exp(-V(\mathbf{q})/k_B T) d\mathbf{q} \quad (5)$$

where  $k_B$  is the Boltzmann constant,  $T$  the temperature,  $n$  the dimension of  $\mathbf{q}$ , and  $\delta$  the symmetry number of the motion (see the work of Fernández-Ramos *et al.*<sup>38</sup> for a thorough discussion). The term  $K(\mathbf{q}) = |\underline{\mathbf{A}}(\mathbf{q})|$ , called the kinetic function here, completely describes the kinetic part of the LAM in a statistical point of view. This integral is restricted within configurational space and allows a simple numerical procedure for its evaluation for routine applications. The partition function defines the statistical properties of the LAM in the canonical ensemble; specifically, the probability density that the system occupies a particular position  $\mathbf{q}$  in the classical formulation is given by:

$$P(\mathbf{q}, T) = \frac{1}{\delta} (2\pi k_B T)^{n/2} \frac{|\underline{\mathbf{A}}(\mathbf{q})|^{1/2} \exp(-V(\mathbf{q})/RT)}{Q(T)} \quad (6)$$

In the case of one-dimensional motion, the computation of the partition function is improved by using quantum statistics. The method is similar to the one presented by Reinisch *et al.*<sup>39</sup>. An effective temperature-dependent kinetic constant is introduced by the following relation:

$$K_{eff}(T) = \int K(q) P(q, T) dq \quad (7)$$

The Fourier Grid Hamiltonian algorithm<sup>40</sup> is then employed to compute the eigenvalues  $\varepsilon_i^{eff}(T)$  of the effective Hamiltonian defined by:

$$\hat{H}_{eff}^{1D} = -\frac{\hbar^2}{2K_{eff}} \frac{\partial^2}{\partial q^2} + \hat{V}(q) \quad (8)$$

The partition function is finally obtained by a direct count of the eigenvalues  $\varepsilon_i^{eff}(T)$ :

$$Q_{eff}^{1D}(T) = \frac{1}{\delta} \sum_i \exp(-\varepsilon_i^{eff}(T)/k_B T) \quad (9)$$

## 1. Calibration of the Potential Energy Surface

The calibration of a model can be viewed in the narrow sense of adapting some of the model parameters in order to get a better resemblance between observations and major end-predictions in specific situations. When calibrating a physico-mathematical model of the potential energy surface with respect to *ab initio* data, uncertainties are associated with the “*physical model uncertainties*”, arising from inadequacies of the physical model (due to underlying assumptions and simplifications). As a result, there are “*model parameter uncertainties*,” arising from uncertain values of the model parameters. These uncertainties are simultaneously quantified here through the solution of statistical inverse problems based on a Bayesian approach, as illustrated by the block ‘PES’ on figure 1.

*a. Check model (in)adequacy*. Let  $M$  designate a stochastic model class<sup>41,42</sup>. A stochastic model  $M$  is specified by a set of uncertain parameters  $\mathbf{a}$ , to which an additional uncertain parameter, variance ( $\sigma_{total}^2$ ), is included as a measure of the total model error. Each stochastic model is then specified by the set  $\boldsymbol{\theta} = \mathbf{a} \cup \sigma_{total}^2 \in \Omega \subset \mathbb{R}^d$ . One can use the data  $\mathbf{D}$  to compute the posterior PDF  $p(\boldsymbol{\theta}|\mathbf{D}, M)$  as defined by the Bayes theorem:

$$p(\boldsymbol{\theta}|\mathbf{D}, M) = \frac{1}{c} p(\mathbf{D}|\boldsymbol{\theta}, M) p(\boldsymbol{\theta}|M) \quad (10)$$

where  $c$  is a normalization constant that makes the probability volume under the posterior PDF equal to unity,  $p(\mathbf{D}|\boldsymbol{\theta}, M)$  is the likelihood function, and  $p(\boldsymbol{\theta}|M)$  is the prior PDF for  $\boldsymbol{\theta}$  (always chosen to be uniform in this paper). The likelihood function expresses the probability of observing  $\mathbf{D}$  based on the predictive PDF for the system output given by the set of parameters  $\boldsymbol{\theta}$  in the model  $M$ . To compute the likelihood function, the assumption of statistically independent error is used. We denote by  $D_j$  the  $j^{th}$  data point and by  $Y_j$  the model output computed for the same scenario as  $D_j$ . We also consider an additive error based on the assumption that the error is independent from the value  $D_j$ :

$$r_{total} = Y_j - D_j \quad (11)$$

In this study we assume that there is no data error, e.g. that the *ab initio* values  $D_j$  calculated at a particular level of theory are perfectly converged with respect to all the numerical parameters of the quantum method used. The error is modeled in this paper as a Gaussian deviate with zero mean and variance  $\sigma_{total}^2$ . Based on all these assumptions, the likelihood function reads

$$\begin{aligned} p(\mathbf{D}|\mathbf{M}, \boldsymbol{\theta}) &= \frac{1}{\sqrt{(2\pi\sigma_{total}^2)^{N_d}}} \\ &\exp \left[ -\frac{1}{2} \sum_{j=1}^{N_d} \frac{(D_j - Y_j)^2}{\sigma_{total}^2} \right] \end{aligned} \quad (12)$$

where  $N_d$  is the number of data point. The variance  $\sigma_{\text{total}}^2$  is treated as an unknown and thus needs to be calibrated as well. Because no data error is assumed, the calibrated variance is a measure of the model adequacy and we call  $\sigma_{\text{total}}^2$  as  $\sigma_M^2$  hereafter. The criterion for accepting a model as “not-invalidated” (hereafter, we simply say “validated”) is subjective ; it requires a metric to compare the predicted quantities produced by the calibration and the data used for the calibration. If the data agree within the acceptable tolerance limit, the model (denoted as  $M_{\text{val}}$ ) is then directly used for the forward problem: the PDF model parameters are propagated to the quantity of interest. This is illustrated in the right part of the block ‘Stat. & UQ’ in figure 1 and discussed in what follows.

*b. Uncertainty propagation* . One of the most important objectives of performing the above analysis is to make robust predictions about the QoI from a data set of the system of interest. Based on a candidate  $M$ , all the probabilistic information for the prediction of a vector of quantities of interest  $\mathbf{Q}$  is contained in the posterior predictive PDF for  $M$  given by the Theorem of Total Probability:

$$p(\mathbf{Q}|\mathbf{D}, M) = \int p(\mathbf{Q}|\boldsymbol{\theta}, \mathbf{D}, M)p(\boldsymbol{\theta}|\mathbf{D}, M)d\boldsymbol{\theta} \quad (13)$$

The above equation obtains the prediction  $p(\mathbf{Q}|\mathbf{D}, M)$  of a vector of quantities of interest  $\mathbf{Q} \in \mathbf{R}^q$  by summing up the prediction  $p(\mathbf{Q}|\boldsymbol{\theta}, \mathbf{D}, M)$  of each model specified by  $\boldsymbol{\theta} \in \Omega$  weighted by its posterior probability  $p(\boldsymbol{\theta}|\mathbf{D}, M)d\boldsymbol{\theta}$ . The evaluation of the multi-dimensional integrals of Eq. (13) cannot usually be done analytically. A common numerical approach often used is based on simulating samples  $\boldsymbol{\theta}^{(k)}$ ,  $k=1,2,\dots,K$ , (called posterior samples) from the posterior PDF  $p(\boldsymbol{\theta}|\mathbf{D}, M)$ . The posterior PDF  $p(\boldsymbol{\theta}|\mathbf{D}, M)$  can be approximated by using these samples:

$$p(\boldsymbol{\theta}|\mathbf{D}, M) \approx \frac{1}{K} \sum_{k=1}^K \delta(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) \quad (14)$$

We use in this paper the Adaptive Multilevel Stochastic Simulation Algorithm presented by Cheung and Beck<sup>43</sup> to generate the posterior samples. The integral in Eq. (13) is then approximated by:

$$p(\mathbf{Q}|\mathbf{D}, M) \approx \frac{1}{K} \sum_{k=1}^K \delta(\mathbf{Q} - \mathbf{Q}^{(k)}) \quad (15)$$

where  $\mathbf{Q}^{(k)}$  is a sample simulated from  $p(\mathbf{Q}|\boldsymbol{\theta}^{(k)}, \mathbf{D}, M)$ . The Monte Carlo simulation method<sup>44</sup> can be used to simulate this sample. If  $\mathbf{Q}$  is a deterministic function of  $\boldsymbol{\theta}$  (i.e.,  $\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta})$ ), then  $\mathbf{Q}^{(k)} = \mathbf{Q}(\boldsymbol{\theta}^{(k)})$ . Estimates for important statistical moments of  $\mathbf{Q}$  conditioned on  $M$  and  $\mathbf{D}$  can be obtained using the samples  $\mathbf{Q}^{(k)}$ ,  $k = 1, \dots, K$ . For instance, the posterior mean is calculated

as follows:

$$E(\mathbf{Q}|\mathbf{D}, M) \approx \frac{1}{K} \sum_{k=1}^K \mathbf{Q}^{(k)}. \quad (16)$$

The 95% confidence interval, usually taken to represent the confidence expected on a predicted outcome, is the interval  $I$  defined by :

$$Prob(\mathbf{Q} \in I) = 95\% \quad (17)$$

In this paper, the mean and the 95% confidence interval on the predicted posterior of  $\mathbf{Q}$  are noted by  $\langle \mathbf{Q} \rangle$  and  $[\mathbf{Q}]$  respectively. The detailed explanation on how to calculate other higher moments can be found elsewhere<sup>45</sup>.

The library QUESO<sup>46</sup> is used to solve the inverse problem and to compute the posteriors  $p(\boldsymbol{\theta}|\mathbf{D}, M)$ . The current numerical methodology is very efficient and feasible for various engineering applications (e.g.<sup>47-51</sup>).

## 2. The kinetic function

The computation of the kinetic function  $|\underline{A}(\mathbf{q})|$  defined by Eq. (4) is achieved using the internal Z-Matrix coordinates. As illustrated in the ‘Kinetic’ box on figure 1, the objective is to compute the absolute displacement of the atoms with respect to variations in the generalized coordinates from the atomic relative positions. The contribution of the overall rotation and translation is separable through Eq. (1), and the kinetic function is calculated so that the total angular ( $\vec{J}$ ) and linear ( $\vec{P}$ ) momenta associated to any variation  $d\mathbf{q}$  are null.

Let the internal coordinates of the Z-Matrix be noted by the vector  $\mathbf{z}$  of dimension  $N_Z$ , where  $N_Z = 3N - 5$  for a linear molecule or  $3N - 6$  otherwise ( $N$  being the number of atoms). The configurations of an  $n$ -dimensional motion parameterized by the generalized coordinates  $\mathbf{q}$  will then be described by  $N_Z$  functions  $z_i(\mathbf{q})$ . The connectivity scheme of the Z-Matrix along with the functions  $z_i(\mathbf{q})$  is referred to as a functional Z-Matrix. It is the definition of the kinetic energy of the system. The (relative) positions of the atoms obtained using the Z-Matrix definition are denoted by  $\vec{Z}_{at}(\mathbf{q})$ , and their construction rule is straightforward: the first three atoms are defining the orientation of the structure, and the origin of the Cartesian coordinates coincides with the center of gravity of the system. Given an initial structure  $\vec{R}_{at}(\mathbf{q}^0)$  directly constructed from the Z-Matrix (e.g.  $\vec{R}_{at}(\mathbf{q}^0) = \vec{Z}_{at}(\mathbf{q}^0)$ ), the  $\partial \vec{R}_{at} / \partial q_i$  terms in Eq. (4) are calculated by determining the structure  $\vec{R}_{at}(\mathbf{q}^0 + d\mathbf{q})$  generated by any displacement  $d\mathbf{q}$  and associated to  $\vec{J} = \vec{P} = \vec{0}$ . While the condition  $\vec{P} = \vec{0}$  is satisfied by construction in the Z-Matrix (because the center of mass of the system is fixed at the origin), this is not the case for the condition  $\vec{J} = \vec{0}$ . By consequence, the structure  $\vec{R}_{at}(\mathbf{q}^0 + d\mathbf{q})$  is related to  $\vec{Z}_{at}(\mathbf{q}^0 + d\mathbf{q})$  by an overall rotation around its center of gravity:

$$\vec{R}_{at}(\mathbf{q}^0 + d\mathbf{q}) = \underline{\mathbf{M}}_{rot} \vec{Z}_{at}(\mathbf{q}^0 + d\mathbf{q}) \quad (18)$$

where  $\underline{\mathbf{M}}_{rot}$  is a rotational matrix defined by the condition:

$$\vec{J} = \sum_{at} m_{at} \left[ \underline{\mathbf{M}}_{rot} \vec{Z}_{at}(\mathbf{q}_0 + d\mathbf{q}_\alpha) - \vec{Z}_{at}(\mathbf{q}_0) \right] \times \vec{Z}_{at}(\mathbf{q}_0) = \vec{0} \quad (19)$$

The matrix  $\underline{\mathbf{M}}_{rot}$  is optimized by an iterative procedure, generalizing the one presented in our previous paper<sup>39</sup> which was restricted to a fixed axis of rotation. The iterative method is initiated by setting  $\underline{\mathbf{M}}_{rot}^{i=0} = \underline{\mathbf{Id}}$ . At each step  $i$ , the angular momentum  $\vec{J}^i$  is calculated using the current  $\underline{\mathbf{M}}_{rot}^i$  according to:

$$\vec{J}^i = \sum_{at} m_{at} \left[ \underline{\mathbf{M}}_{rot}^i \vec{Z}_{at}(\mathbf{q}_0 + d\mathbf{q}_\alpha) - \vec{Z}_{at}(\mathbf{q}_0) \right] \times \vec{Z}_{at}(\mathbf{q}_0) \quad (20)$$

The matrix  $\underline{\mathbf{M}}_{rot}^i$  is then corrected at the next iteration by:

$$\underline{\mathbf{M}}_{rot}^{i+1} = \underline{\mathbf{M}}_{rot}^i \times \underline{\mathbf{M}}_{rot}^{\vec{J}^i, d\theta} \quad (21)$$

where  $\underline{\mathbf{M}}_{rot}^{\vec{J}^i, d\theta}$  is the matrix of rotation around the axis  $\vec{J}^i$  (current overall angular momentum) with the counter rotational angle  $d\theta$  defined by:

$$d\theta = -d\alpha \frac{\vec{J}^i}{\vec{J}^{ext}(d\alpha)} \text{sign}(\vec{J}^i \cdot \vec{J}^{ext}) \quad (22)$$

where  $d\alpha$  is an elementary angle (taken in this paper as  $10^{-4}$  radian),  $\vec{J}^{ext}$  is the angular momentum associated with a rotation of  $d\alpha$  around  $\vec{J}^i$  of the structure  $\underline{\mathbf{M}}_{rot}^i \vec{Z}_{at}(\mathbf{q}_0 + d\mathbf{q}_\alpha)$ .

This procedure leads to a progressive annihilation of  $\vec{J}^i$  associated with a decrease of the kinetic function. The procedure is stopped in this paper when the kinetic function is converged within a relative factor of  $10^{-5}$ .

## B. Coupled motions

### 1. Flexible Rotor (FR) partition function

The presence of the large amplitude motion leads to some coupling with the overall rotation, and the usual rigid rotor approximation needs to be overcome. The partition function is computed by considering a loose coupling between the LAM and the overall rotation, as proposed by Vansteenkiste *et al.*<sup>11</sup>. The partition function of the 'Flexible' Rotor  $Q_{FR}(T)$  is based on the Rigid Rotor

expression (RR)<sup>52</sup>, calculated at each configuration point of the LAM, and averaged using the probability density defined in relation (6) :

$$Q(T) = \int Q_{RR}(T, \mathbf{q}) P(\mathbf{q}, T) d\mathbf{q} \quad (23)$$

where  $Q_{RR}(T, \mathbf{q})$  is the partition function of a rigid rotor<sup>52</sup> of the structure  $\vec{Z}_{at}(\mathbf{q})$ .

### 2. Flexible Oscillator (FO) partition function

As for the overall rotation, the presence of the LAM may lead to some coupling with small internal vibrations. Wong *et al.*<sup>9</sup> and Vansteenkiste *et al.*<sup>11</sup> proposed to perform an integration similar to Eq. (23) by using a Harmonic Oscillator approximation of the small amplitude vibration at each configuration point. This procedure leads to an excellent approximation of the partition function, however it requires the costly evaluation of the Hessian over all the configurational space. Here, we propose to restrict the integration to the stable configurations involved in the LAM. The partition function of the coupled small amplitude vibration under this assumption given by:

$$Q(T) = \sum_i Q_{HO}(T, q_i) P(q_i, T) \quad (24)$$

where the sum is carried out on all the stable configurations of the LAM,  $q_i$  is the associated configurational point,  $Q_{HO}(T, q_i)$  is the partition function of the small amplitude vibration in the HO approximation<sup>36</sup> of the structure  $\vec{Z}_{at}(q_i)$ , and  $P(q_i, T)$  is the weighting factor defined by:

$$P(q_i, T) = \frac{|A(q_i)|^{1/2} \exp(-V(q_i)/RT)}{\sum_j |A(q_j)|^{1/2} \exp(-V(q_j)/RT)} \quad (25)$$

## III. APPLICATIONS

### A. 1,3-butadiene partition function

The first application concerns the computation of the partition function of 1,3-butadiene. The geometry of the molecule at the minimum of the PES is illustrated on figure 2 along with the atom numbering used hereafter. The molecule contains a torsional motion defined by the relative rotation of the two  $\text{H}_2\text{CH}$  parts around the central  $\text{C}_5\text{-C}_2$  bond. Previous studies<sup>9,11</sup> have shown that it involves a highly asymmetric internal rotor associated with a non-negligible geometry relaxation effect as well as a non-negligible coupling between the torsion and the other degrees of freedom.

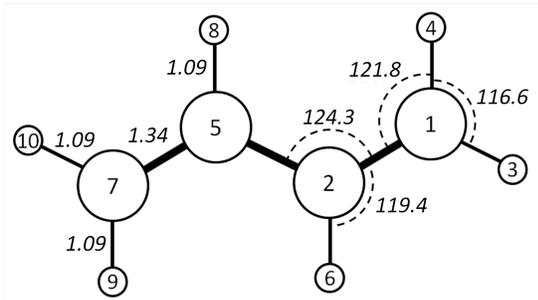


FIG. 2. Illustration of the geometry of 1,3-butadiene at the global minimum of the PES. The large circles represent the carbon atoms, and the small circles represent the hydrogen atoms. Bond lengths are in Angstroms and angles in degrees. The atom numbering used in the paper are indicated in the circles.

### 1. Electronic structure calculation

Ground state electronic energies are calculated using the *ab initio* methods UB3LYP<sup>53</sup>/6-31G(d)<sup>54</sup> (hereafter referred to as DFT) for qualitative calculations, and RHF-CCSD(T)/aug-ccdpvz<sup>55,56</sup> (hereafter referred as CC) for accurate quantitative calculations. Geometry optimizations and a normal mode analysis are performed at the UB3LYP/6-31G(d) level of theory. The torsional potential energy surface is obtained by computing the relaxed geometries with respect to the dihedral  $C_1C_2C_5C_7$  using a  $10^\circ$  step. We use a local quadratic interpolation to obtain *ab initio* values of the energy at other torsional angles. All the electronic calculations have been performed using the GAMESS code<sup>2</sup>.

### 2. Kinetic function

Table I presents the functional Z-Matrix used to compute the kinetic function. The functions  $f_i(\theta)$  ( $i = 1, \dots, 8$ ) define the relaxation of the degrees of freedom with respect to the generalized coordinate  $\theta$ . The DoFs contributing to less than 1% to the kinetic function have been considered constant. The functions  $f_i$  are obtained by Fourier series development of the optimized internal coordinates, and are presented in Supplementary Information 1. The kinetic function calculated using our numerical approach is presented on figure 3 (solid line), and compared to those reported by Vansteenkiste *et al.*<sup>11</sup> and Wong *et al.*<sup>9</sup>. A very good agreement is observed between the 3 studies, the difference from the Eckart method at  $\theta \approx 180^\circ$  being most likely attributed to the different level of theory used in the study of Wong *et al.* for geometry optimization (MP2). We also present in figure 3 the kinetic function computed when the relaxation of the geometry with respect to  $\theta$  is not taken into account ( $f_i(\theta) = 0 \forall i$ , dashed line). As it can be seen, the kinetic function is strongly affected and gives values in close agreement with those presented by Wong *et al.*<sup>9</sup>

$C_1$					
$C_2$	1	1.34			
$H_3$	1	1.09	2	121.8	
$H_4$	1	1.09	3	116.6	2 $180 + f_1(\theta)$
$C_5$	2	$1.46 + f_2(\theta)$	1	$124.3 + f_3(\theta)$	3 $180 + f_4(\theta)$
$H_6$	2	1.09	1	119.4	3 $f_5(\theta)$
$C_7$	5	1.34	2	$124.3 + f_3(\theta)$	1 $-180 + \theta$
$H_8$	5	1.09	2	116.2	1 $\theta + f_7(\theta)$
$H_9$	7	1.09	5	121.5	2 $f_8(\theta)$
$H_{10}$	7	1.09	5	121.8	2 $180 + f_9(\theta)$

TABLE I. Functional Z-Matrix definition of the torsional motion of the 1,3-Butadiene. The definition of the relaxation functions  $f_i(\theta)$  are given as supplementary informations. Lengths are in Angstrom and angles in degrees.

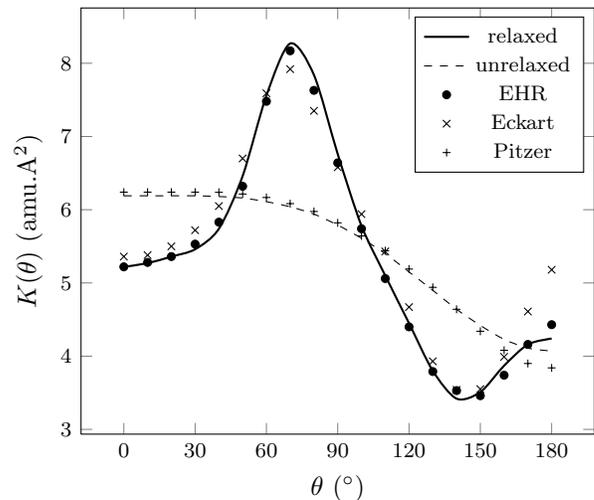


FIG. 3. Comparison between the kinetic function calculated in this work (lines) and the kinetic function obtained using the Extended Hindered Rotor model<sup>11</sup> (EHR), the Eckart method<sup>9</sup> and the Pitzer method<sup>57</sup> (presented by Wong *et al.*<sup>9</sup>). Solid line: flexible torsion ( $f_i(\theta) \neq 0$ ); dashed line: rigid rotation ( $f_i(\theta) = 0$ ).

calculated using the Pitzer method<sup>57</sup>. Because the Pitzer values are also based on relaxed geometries, this demonstrates that the Pitzer method actually fails to take into account the dynamic influence of the relaxation on the kinetic energy.

### 3. Calibration of the potential energy surface

The model used to describe the PES of the torsional motion accounts for two physical contributions: (i) the contribution of the intrinsic torsional energy of the  $C_5-C_2$

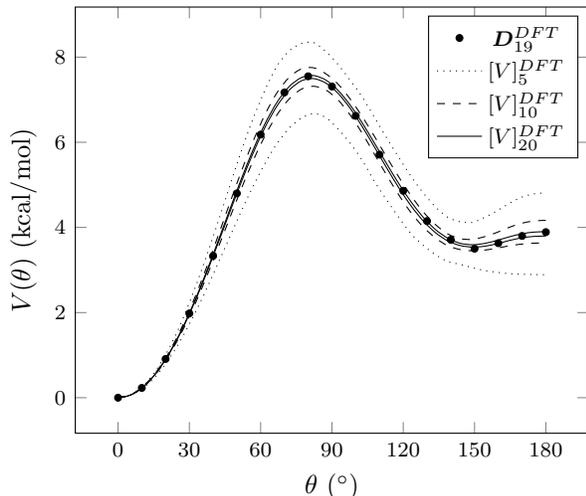


FIG. 4. 95% confidence interval of the posterior PES using the data set  $\mathbf{D}_n^{DFT}$  with  $n = 5$  (dotted line),  $n = 10$  (dashed line) and  $n = 20$  (solid line). Filled circles represent the data set  $\mathbf{D}_{19}^{DFT}$ .

$sp^2$  chemical bond, and (ii) the contribution of a repulsive part, representing the steric repulsion between the atoms  $H_9$  and  $H_4$  during the torsion. Using the definition of the repulsive part of a Lennard-Jones type potential, and a modified cosine function to account for the  $sp^2$  torsion energy, the PES is defined using the 3 parameters  $a_0$ ,  $a_1$ , and  $a_2$  by:

$$V(\theta) = \frac{a_1}{2} (1 - \cos(2\theta g_{a_0}(\theta))) + a_2 \left( \frac{d_{VDW}}{|\vec{R}_9(\theta) - \vec{R}_4(\theta)|^9} - \frac{d_{VDW}}{|\vec{R}_9(0) - \vec{R}_4(0)|^9} \right) \quad (26)$$

where  $d_{VDW}$  is taken equal to  $2.4 \text{ \AA}$  (sum of the van der Waals radii of the hydrogen atoms), and  $g_{a_0}(\theta) = 1 - a_0(1 - \cos(2\theta))$ . The function  $g_{a_0}$  adds degrees of freedom to model the bonding contribution of the  $sp^2$  torsional energy, and one can verify that  $g_{a_0}(\theta) = g_{a_0}(-\theta)$  as well as  $g_{a_0}(-180) = g_{a_0}(180) = g_{a_0}(0) = 0$ . The term  $-d_{VDW}/|\vec{R}_9(0) - \vec{R}_4(0)|^9$  in the equation ensures the condition  $V(0) = 0$ . The stochastic PES model  $M$  contains four parameters to calibrate: the three involved in Eq.(26) ( $a_0, a_1, a_2$ ), and the variance ( $\sigma_M^2$ ).

In order to study how the *ab initio* data inform the PES parameters, different datasets are used for the inference process. They are noted  $\mathbf{D}_n^X$  where  $X \in \{DFT, CC\}$  stands for the level used for the electronic energy calculations, and  $n$  is the number of values included (in uniform repartition in  $[0; 180^\circ]$ ). The prior PDFs, the 95% confidence intervals (CIs) based on the posterior PDFs, and the posterior means are presented in table II.

We first comment on the results obtained using the DFT datasets. It is shown in rows 3 to 8 that the 95% CI is considerably narrowed around the mean value for each parameter each time the dataset size increases. Also, the posterior mean value  $\langle a_i \rangle$  is already accurately predicted when only 5 points are used. However, the parameter uncertainties remain large at this level. The mean of the variance shows a different behavior: its value keeps decreasing when the amount of data increases. It reaches  $1.25 \times 10^{-3} \text{ kcal}^2/\text{mol}^2$  (e.g.  $\langle \sigma_M \rangle = 3.5 \times 10^{-2} \text{ kcal/mol}$ ) when 20 points are used. The 95% CIs  $[V]_n^{DFT}$  ( $n = 5, 10, 20$ ) are presented in figure 4 and compared to the  $\mathbf{D}_{19}^{DFT}$  data set. The figure confirms the quality of the proposed model and is able to reproduce almost perfectly the *ab initio* results when a sufficient amount of data points is used. It is worth mentioning that the uncertainty on the posterior PES (maximum here at  $\approx 90^\circ$  and  $\approx 180^\circ$ ) highlights the most relevant volume of the configurational space in which data are needed to improve the calibration process.

We now look at the posterior PDFs of model parameters obtained using the data set  $\mathbf{D}_n^{CC}$  ( $n = 2, 3, 5$ ) presented on the four last rows of table II. In this case, we calibrate model parameters except for  $\sigma_M^2$ , which is set to be  $1.25 \times 10^{-3} \text{ kcal}^2/\text{mol}^2$ . In other words, we assume that the model error obtained by the DFT level of theory is close to the realistic estimation of the true model error. As can be seen on table II, this assumption is reasonable since  $[V]_5^{CC}$  is already very narrow (and actually converged) for all the parameters even when 5 points are used. It is also shown that using 2 points (at  $0^\circ$  and  $180^\circ$ ) is not enough to infer  $a_0$  and  $a_2$  (posterior PDFs approximately equal to the prior PDFs), while a third point at  $90^\circ$  allows a considerable reduction of the uncertainty on  $a_2$ . These results are illustrated on figure 5, which presents the uncertainty domain of the posterior PES for these three datasets. Comparing  $[V]_5^{CC}$  a posteriori with the  $\mathbf{D}_{19}^{CC}$  dataset demonstrates that the DFT level of theory is fully able to estimate the absolute model error for this case as the data points are almost perfectly encapsulated in  $[V]_5^{CC}$  (assuming that the CCSDT/aug-cc-dvpz level of theory gives the exact PES).

#### 4. Forward Problem: Partition function

The 95% CI of the posterior partition function  $[Q_{LAM}]^{DFT}$  using the  $\mathbf{D}_{n=5, 10, 20}^{DFT}$  datasets are presented in figure 6. The results are normalized to the partition function obtained using an HO approximation. The mean value of the posterior partition function using  $\mathbf{D}_5^{DFT}$  is also presented. As for the PES results, the 95% CI becomes very narrow as soon as 10 points are used, and the mean value  $\langle Q \rangle_5^{DFT}$  is already almost converged. The fact that the  $[Q]_n^X$  are centered around 1 at low temperature ( $T < 400 \text{ K}$ ) is important because it demonstrates that the method presented here to treat the 1D LAM has a sufficient capability to perfectly re-

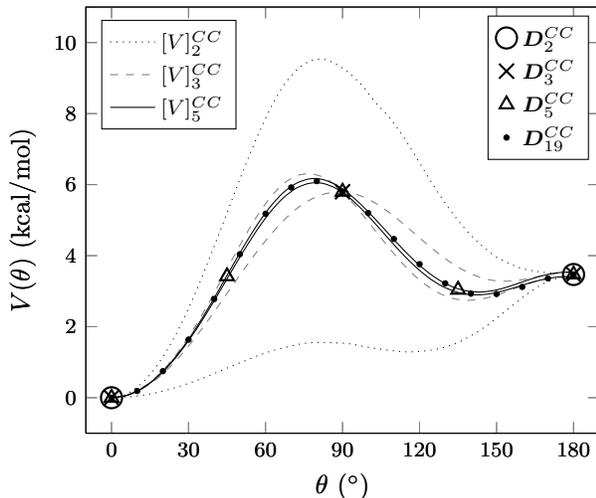


FIG. 5. 95% CI of the posterior PES using the data set  $D_n^{CC}$  with  $n = 2$  (dotted line),  $n = 3$  (dashed line) and  $n = 5$  (solid line). The circles, crosses and empty circles represent the corresponding data set respectively. The filled circles represent the data set  $D_{19}^{CC}$ .

	$a_0 \times 10^2$	$a_1$	$a_2$	$\sigma_M^2 \times 10^3$
Prior	[1;10]	[1;10]	[1;10]	[0;2000]
$[.]_5^{DFT}$	[2.22;9.15]	[2.32;3.87]	[6.65;8.34]	[25.8;1332.2]
$[.]_{10}^{DFT}$	[4.82;6.84]	[2.92;3.35]	[7.28;7.72]	[0.34;154.8]
$[.]_{20}^{DFT}$	[5.54;5.99]	[3.05;3.14]	[7.46;7.53]	[0.18;3.9]
$\langle . \rangle_5^{DFT}$	5.95	3.15	7.45	196.02
$\langle . \rangle_{10}^{DFT}$	5.92	3.17	7.5	16.30
$\langle . \rangle_{20}^{DFT}$	5.79	3.10	7.49	1.25
$\sigma_M^2 = \text{cst}$				
$[.]_2^{CC}$	[1.36;9.52]	[2.73;2.84]	[1.53;9.59]	1.25
$[.]_3^{CC}$	[1.42;9.6]	[2.74;2.84]	[5.72;6.28]	1.25
$[.]_5^{CC}$	[6.89;7.78]	[2.75;2.85]	[6.02;6.14]	1.25
$\langle . \rangle_5^{CC}$	7.33	2.80	6.08	1.25

TABLE II. Definition of the parameter's priors, and description of the parameter's posterior for each dataset used in the paper. Units:  $a_0$ : no unit,  $a_1$ : kcal/mol,  $a_2$ : kcal/mol,  $\sigma_M$ : kcal/mol.

produce the harmonic approximation results in the low temperature limit.

The total partition function of 1,3-butadiene is obtained by multiplication of the LAM partition function by the FR and FO partition functions. To compute the FHO partition function, we consider the system constituted by two stable complexes, characterized by the normal mode frequencies at  $\theta = 0^\circ$  and  $\theta = 145^\circ$  (the normal mode frequencies are presented as supplementary information).

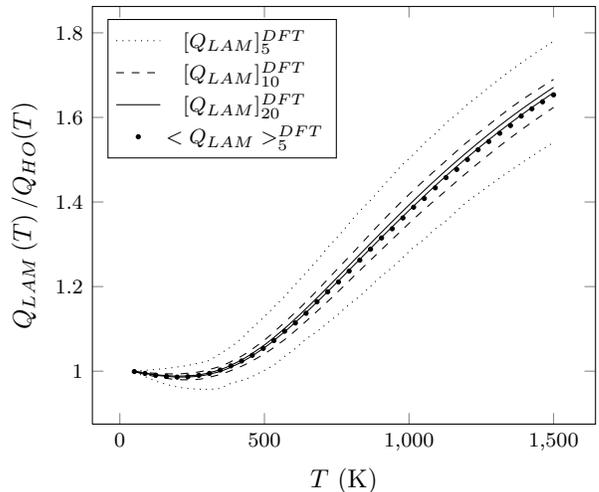


FIG. 6. 95% CI of the posterior torsional partition function using the data set  $D_n^{DFT}$  with  $n = 5$  (dotted line),  $n = 10$  (dashed line) and  $n = 20$  (solid line). Filled circles: mean of the posterior torsional partition function using  $D_5^{DFT}$ . Results are normalized by the HO partition function.

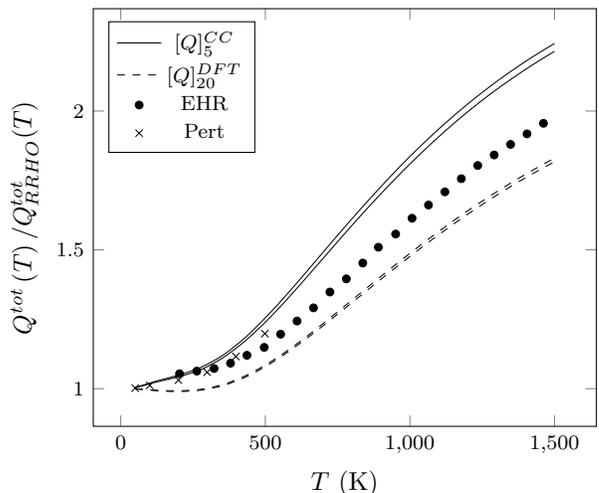


FIG. 7. Comparison between the 95% confidence interval of the posterior partition function of the butadiene using  $D_{20}^{DFT}$  (dashed line) and  $D_5^{CC}$  ( $\sigma_M^2 = 1.25 \cdot 10^{-3}$ ). Crosses: Pert method<sup>58</sup>, circles: EHR method<sup>11</sup>. Results normalized to the RRHO partition function.

It should be noted that the PES uncertainty also affects the FR and FO partition functions through Eq. (6) and Eq. (25). The 95% CI of the posterior partition function of 1,3-butadiene using the  $D_5^{CC}$  (with  $\sigma_M^2$  kept fixed at  $1.25 \cdot 10^{-3}$  kcal<sup>2</sup>/mol<sup>2</sup>) and  $D_{20}^{DFT}$  datasets are presented on figure 7 and compared to the results presented by Wong and Raman<sup>58</sup> (Pert model) and those presented by

Vansteenkiste *et al.*<sup>11</sup> (EHR model). The results are normalized by the partition function obtained using a Rigid Rotor Harmonic Oscillator approximation (RRHO). We first compare our prediction using  $D_5^{CC}$  with the low temperature results presented by Wong and Raman<sup>58</sup> using their Pert method. The comparison is meaningful because the same level of theory has been used to compute the PES, and also because the Pert method used represents most likely the best model available in the literature. However, this fully coupled quantum method is computationally expensive, which explains why no results for temperatures higher than 500 K have been presented. Although our predictions slightly overestimate the partition function, the behavior of the curves is very similar, and the results stay very close to each other. The small overestimation comes from either the simplification introduced by our treatment and/or from differences in the RRHO reference partition function. We now compare the results obtained by Vansteenkiste *et al.*<sup>11</sup> (EHR method) and our results using the  $D_{20}^{DFT}$  data set. The same DFT method used to compute the PES is shared by the two studies; however, Vansteenkiste *et al.* consider a larger basis set in their work (6-311G(d,p)). The two PESs are nevertheless similar, as reflected by the small difference of the torsional barrier height (0.1kcal/mol). As can be seen in the figure, the behavior of the partition function curves are very similar, essentially differing by a small translation factor. This factor seems to be an artifact involved in the EHR method as the partition function does not converge to 1 at the limit of low temperatures.

The model 1D LAM method proposed in this paper is able to account for 1D torsional motion with an accuracy comparable to the best methods presented until now. It is worth recalling that only 2 normal mode analyses have been realized here, while they have been conducted all over the configurational space in the studies of Vansteenkiste *et al.*<sup>11</sup> and Wong and Raman<sup>58</sup>. Also, this efficient formalism, in terms of computational time, further allows it to be used in conjunction with an uncertainty quantification algorithm.

## B. Kinetic rate of $\text{CH}_3 + \text{H} \rightarrow \text{CH}_4$ in the high pressure limit

The reaction rate of the  $\text{CH}_3 + \text{H}$  recombination is computed at the high pressure limit using Variational Transition State Theory (VTST) with Variational Reaction Coordinate (VRC) and spherical Dividing Surfaces (DS) in the canonical ensemble. In the present case, the DS is parameterized using one pivot point  $\vec{p}$  attached to the  $\text{CH}_3$  part, around which the approaching  $\text{H}^*$  atom is allowed to rotate. The optimal DS (noted  $\text{DS}^*$ ), defining the transition state, is optimized for every temperature in a way that it is associated to a minimal partition

function<sup>59,60</sup>:

$$Q_{\text{DS}^*}(T) = \min_{\vec{p}, s} [Q_{\text{DS}}^{\vec{p}, s}(T)] \quad (27)$$

where  $s$  is the reaction coordinate (separation  $\vec{p}-\text{H}^*$ ), and  $Q_{\text{DS}}^{\vec{p}, s}(T)$  is the partition function of the DS defined by  $s$  and  $\vec{p}$ . The high pressure reaction rate is calculated using the standard TST assumptions and is given by<sup>52</sup>:

$$k(T) = \frac{k_B T}{h} \frac{Q_{\text{DS}^*}^{\vec{p}, x}}{Q_{\text{CH}_3}(T) Q_{\text{H}}(T)} \exp(V^*/k_B T) \quad (28)$$

where  $h$  is Planck's constant,  $Q_{\text{CH}_3}$  and  $Q_{\text{H}}$  the partition functions of the methyl and hydrogen radicals respectively, and  $V^*$  is the electronic barrier height associated with the  $\text{CH}_3 + \text{H}$   $\text{DS}^*$ . Over the small amplitude motion involved in the system, the  $\text{CH}_3$  umbrella motion changes quite substantially and may need to be included in a FO partition function. However, Klippenstein *et al.*<sup>16</sup> have shown that its influence is negligible with at most a 2 % increase of the reaction rate for temperatures below 2400 K. In this work, the small amplitude motions are then supposed uncoupled to the reactional motion, and the reaction rate expression is simplified into:

$$k(T) = \frac{k_B T}{h} \frac{Q_{\text{DS}^*}^{\text{D}}(T) Q_{\text{DS}^*}^{\text{FR}}}{Q_{\text{CH}_3}^{\text{trans\&rot}}(T) Q_{\text{H}}(T)} \exp(V^*/k_B T) \quad (29)$$

where  $Q_{\text{DS}^*}^{\text{D}}$  is the partition function of the  $\text{H}^*$  motion on the  $\text{DS}^*$ ,  $Q_{\text{DS}^*}^{\text{FR}}$  the effective overall rotation partition function of the DS, and  $Q_{\text{CH}_3}^{\text{trans\&rot}}$  the partition function of the overall translational and rotational motion of  $\text{CH}_3$ .

### 1. Electronic structure calculation

Ground state electronic energies are calculated using the CR-CC(2,3)<sup>61,62</sup>/aug-cc-pvdz<sup>55,56</sup> level of theory. The CR-CC(2,3) method is an improvement over the CCSD(T) approach to overcome its deficiencies in describing systems involving biradical character<sup>63</sup>. Geometry optimizations are carried out at the UB3LYP<sup>53</sup>/6-31G(d)<sup>54</sup> level of theory. All the electronic calculations have been performed using the GAMESS code<sup>2</sup>.

### 2. Kinetic function

The kinetic energy of the DS is obtained by defining the functional Z-Matrix associated with the  $\text{H}^*$  2D motion. Because of the  $C_3$  symmetry of  $\text{CH}_3$ , the DS is parameterized by 2 parameters. The first parameter is the distance  $x$  between the pivot point  $\vec{p}$ , lying in the  $C_3$  axis of symmetry of  $\text{CH}_3$ , and the carbon atom, taken as the origin. We denote from now on  $\vec{p} = x$ . The second parameter is the distance  $s$  between the approaching  $\text{H}^*$  atom and the pivot point. The functional Z-Matrix which describes the DS is presented in table III. The  $C_3$

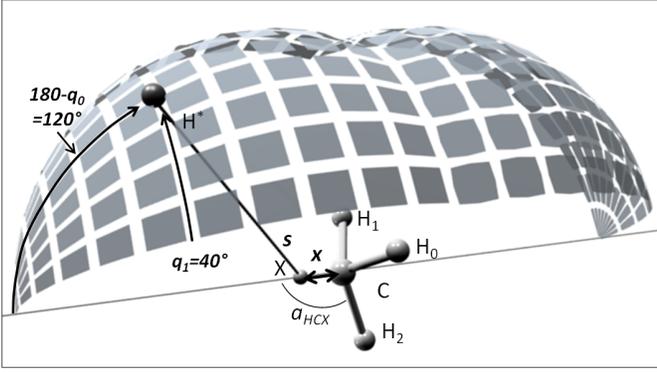


FIG. 8. Illustration of a typical bi-faceted dividing surface of the  $\text{CH}_3+\text{H} \rightarrow \text{CH}_4$  reaction.

symmetry is imposed to the system, and the relaxation of the  $\text{CH}_3$  part is taken into account to some extent by using the functions  $a(|\vec{\text{CH}}^*|)$  (interpolation between the optimized values on the MEP).

It has been shown that a limitation of the spherical dividing surfaces is that they can account for artificial contributions to the reaction rate when multiple reaction paths are present<sup>12,13</sup>. To avoid this overestimation, a multifaceted dividing surface are used, composed by the envelope of spherical DSs centered around a reactive channel specific pivot point. Considering the equivalence of the two association channels for the  $\text{H}^*$  addition on  $\text{CH}_3$ , a typical dividing surface for the reaction is illustrated on figure 8. To compute the DS partition function, the integration domain defined in Eq. (5) has to be restricted to:

$$\begin{aligned} q_0 &\in [q_0^{\min}, \pi] \\ q_1 &\in [0, \pi/3] \end{aligned} \quad (30)$$

where  $q_0^{\min} = \arcsin(x/s)$  is the solid angle from  $X$  defining the intersection of the two spherical DS. The overall symmetry number of the irreducible integration domain is then 12. Since this number takes into account the two reaction paths, no symmetry number has to be considered in the overall rotational partition function of both  $\text{CH}_3$  and the DS.

C						
X	1	$x$				
$H_0$	1	1.09	2	$a( \vec{\text{CH}}^* )$		
$H_1$	1	1.09	2	$a( \vec{\text{CH}}^* )$	3	120
$H_2$	1	1.09	2	$a( \vec{\text{CH}}^* )$	4	120
$H^*$	2	$s$	1	$q_0$	3	$q_1$

TABLE III. Z-Matrix definition of the Dividing Surface parameterized by the reaction coordinate  $s$  and the pivot point  $x$ . Lengths in Angstrom, angles in degrees.

### 3. Calibration of the PES

The PES of the  $\text{CH}_3/\text{H}^*$  interaction accounts for three contributions: (i) the stretching energy of the C- $\text{H}^*$  bond (noted  $V_{str}$ ), (ii) the bending energy of the  $\text{H}_i\text{CH}^*$  degree of freedom (noted  $V_{bend}$ ), and (iii) the steric repulsion energy between the  $\text{H}_i$  and  $\text{H}^*$ . The stretching energy  $V_{str}$  is written as a Morse-like function of  $|\vec{\text{CH}}^*|$  with an origin of energy taken at the products state:

$$V_{str}^{a_0, a_1} = D_e \left[ 1 - \exp(-a_0(|\vec{\text{CH}}^*| - r_{eq})^{a_1}) \right]^2 - D_e \quad (31)$$

where  $D_e = 107.7$  kcal/mol is the dissociation energy,  $r_{eq} = 1.09$  Å is the equilibrium bond length,  $a_0$  and  $a_1$  are the first two parameters of the PES model. Note that the function is defined only for  $|\vec{\text{CH}}^*| > r_{eq}$ . The bending energy is modeled by a simple cosine function, with a barrier height equal to the dissociation energy at the angle  $\alpha_{\text{XCH}^*} = \pm\pi/2$ :

$$V_{bend} = \frac{D_e - V_{str}(q_0, q_1)}{2} (1 - \cos(\alpha_{\text{XCH}^*})) \quad (32)$$

Finally the PES also accounts for the spherical repulsion between the passive hydrogen atoms  $\text{H}_i$  and  $\text{H}^*$ :

$$V_{rep}^{a_2, a_3} = a_2 \sum_{i=0,1,2} \left[ \left( \frac{\sigma_{HH}}{|\text{H}_i\text{H}^*|} \right)^{a_3} - \left( \frac{\sigma_{HH}}{|\text{H}_i\text{H}_0^*|} \right)^{a_3} \right] \quad (33)$$

where  $\sigma_{HH} = 2.4$  Å with  $a_2$  and  $a_3$  being the last two parameters, and  $\text{H}_0^*$  is the corresponding position of  $\text{H}^*$  on the MEP (at the same C -  $\text{H}^*$  separation). The term  $\left( \frac{\sigma_{HH}}{|\text{H}_i\text{H}_0^*|} \right)^{a_3}$  allows a cancellation of the steric interaction energy on the MEP. The PES is the sum of these three contributions, and are defined through the four parameters  $a_i$ ,  $i = 0, \dots, 3$ :

$$V^{a_i}(\mathbf{q}) = V_{str}^{a_0, a_1}(\mathbf{q}) + V_{bend}(\mathbf{q}) + V_{rep}^{a_2, a_3}(\mathbf{q}) \quad (34)$$

The geometries used to compute the electronic energies are defined by the Z-Matrix presented in table III with  $x = \epsilon = 10^{-5}$  (in order to preserve a non-ambiguous definition of the  $\text{C}_3$  axis). The 50 data points collected are divided in two groups: one is used to infer the PES model (noted  $\mathbf{D}^*$ ), and the second is used to validate it for extrapolation (noted  $\mathbf{D}$ ). The dataset  $\mathbf{D}^*$  contains 25 points in the relevant space of the PES for reaction rate calculation and is constituted by:

- 10 points on the MEP ( $q_0 = q_1 = 0$ ) from  $s = 2.0$  to  $s = 3.8$  Å
- 15 points out of the MEP defined by the combination  $s \in \{2.0, 2.4, 2.8\text{Å}\} \otimes q_0 \in \{20, 40, 60, 80, 100^\circ\} \otimes q_1 = 0$

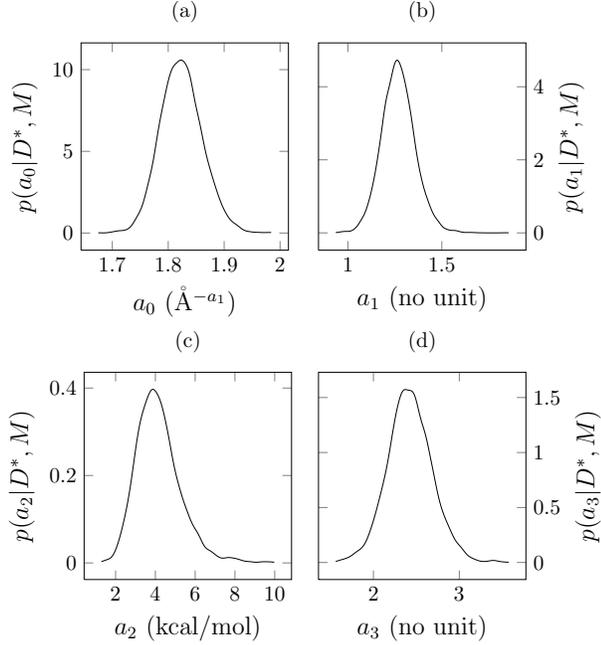


FIG. 9. Posterior PDF,  $p(\mathbf{a}_i|\mathbf{D}^*, M)$ , (a), (b), (c), (d) for respectively  $i=0, 1, 2, 3$ .

The  $\mathbf{D}$  data set collects the energies of the corresponding geometries of (b) at  $q_1 = 30^\circ$  as well as 10 other points at  $s = 1.6 \text{ \AA}$ . The prior intervals, the posteriors means, and 95 % confidence intervals are presented in table IV, and the 4 posteriors  $p(\mathbf{a}|\mathbf{D}^*, M)$  are presented in figure 9. We confirm here that the Morse potential (defined for  $a_1 = 1$ ) is not the most appropriate to describe the MEP, and a value of  $a_1 = 1.25$  is the most probable value. The mean value of  $a_3$  is surprisingly low for a repulsion term: in a Van Der Waals force field, the exponent of the repulsive part is usually taken between 9 and 12.

	$a_0$	$a_1$	$a_2$	$a_3$	$\sigma_M^2$
Prior	[0;10]	[0;10]	[0;10]	[0;10]	[0;10]
[.]	[1.75;1.87]	[1.22;1.43]	[2.13;6.04]	[2.12;2.87]	[1.6;2.8]
< . >	1.82	1.27	4.12	2.53	2.59

TABLE IV. The prior PDFs of the parameters, and 95% CIs ([.]) based on the posterior PDFs, and the posterior mean values (< . >). Units:  $a_0$ :  $\text{\AA}^{-a_1}$ ,  $a_1$ : no unit,  $a_2$ : kcal/mol,  $a_3$ : kcal/mol,  $\sigma_M$ : kcal/mol.

We compare on figure 11 the 95 % CI of the posterior MEP ( $[V]$ ) with the data points of  $\mathbf{D}^*$  (black symbols) and  $\mathbf{D}$  (white symbols) included in the MEP. We also report on this figure the *ab initio* results at the CASPT2/aug-cc-pVQZ<sup>17</sup>, full-CI/6-31G(d)<sup>64</sup> and CCSD(T)/6-31G(d)<sup>17</sup> level of theories presented in the

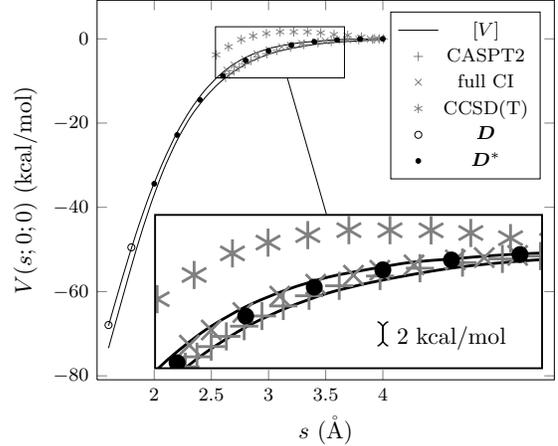


FIG. 10. Comparison between the predicted 95% confidence interval of the MEP ( $[V]$ ), the datasets  $\mathbf{D}$  and  $\mathbf{D}^*$  and the *ab initio* results presented in literature. Lines:  $[V]$ ; black circles:  $\mathbf{D}^*$ , empty circles:  $\mathbf{D}$ . Grey symbols: literature results; asterisks: CCSD(T)/6-31G(d)<sup>17</sup>, crosses: CASPT2/aug-ccVQZ<sup>17</sup>, times: full CI/6-31G(d)<sup>64</sup>. Origin of energy:  $V(4 \text{ \AA}, 0, 0)$ .

literature. The uncertainty of the PES model is under 2 kcal/mol for separations higher than 2.0  $\text{\AA}$ , and coincidentally encapsulates the results obtained at the full CI and CASPT2 level of theory. As already reported<sup>17</sup>, the CCSD(T) level of theory is not able to properly estimate the energy of the system for intermediate separation. The comparison between the PES model and the direct *ab initio* data is pursued in figure 11 for out-of-MEP situations. The black and white symbols still represent the  $\mathbf{D}^*$  and  $\mathbf{D}$  data sets respectively. The approximation of the PES model is very satisfying. Even for extrapolated values at low separation ( $s = 1.6 \text{ \AA}$ ) the model still performs well with the hindering domain being correctly predicted within 10%.

#### 4. Forward propagation to the high-pressure recombination rate

The 95 % CI ( $[k]$ ) and the mean ( $\langle k \rangle$ ) of the posterior high pressure reaction are presented in figure 12 with the available experimental measurements<sup>65–67</sup> and the VTST-VRC theoretical predictions presented by Klippenstein *et al.*<sup>16</sup> and Harding *et al.*<sup>17</sup>. The experimental measures have been converted from the  $k_0$  values of the original experimental works in the same way presented by Klippenstein *et al.*<sup>16</sup>. The confidence intervals predicted by our approach are in very good agreement with the experimental values as well as with the theoretical calculations. Our calculations are mainly limited by two methodological factors. First, the representation of the

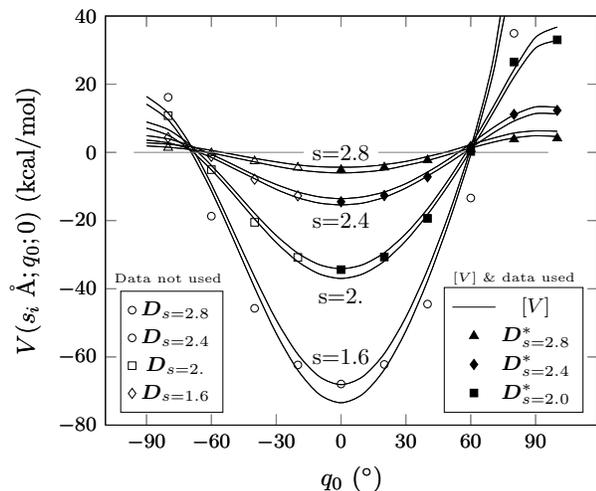


FIG. 11. Comparison between the predicted 95% confidence interval of the PES ( $[V]$ ) with the datasets  $D$  and  $D^*$ . Lines:  $[V]$ ; black circles:  $D^*$ , empty circles:  $D$ . Origin of energy:  $V(4 \text{ Å}, 0, 0)$ .

dividing surface does not constitute a perfect DS and it is associated with an overestimation of the reaction rate. The works of Klippenstein *et al.*<sup>16</sup> have shown that by using both VTST-VRC and a direct dynamic method, the spherical DS for the  $\text{CH}_3+\text{H} \rightarrow \text{CH}_4$  reaction is associated with a 9% overestimation, approximately independent of the temperature. The doubly faceted DS used here leads to an even lower recrossing factor. The second limitation is the restriction of the statistical study to the canonical ensemble. In other works<sup>12</sup>, the associated error is evaluated at approximately 20%. The overestimation of the reaction rate coming from the canonical analysis is probably compensated to some extent by the PES model which predicts a slightly higher hindrance effect at high separation (the *ab initio* data are close to the bottom boundary of the confidence interval at the  $s = 2.8 \text{ Å}$  case on figure 11).

To finish this study, we comment on the difference in the PES representation used here and in the works presented by Harding *et al.*<sup>17</sup> and Klippenstein *et al.*<sup>16</sup>. In the works of Klippenstein *et al.*<sup>16</sup>, a combined Fourier series/3D spline fitting procedure is used to obtain a four dimensional PES representation (the umbrella motion is also explicitly considered). If the fit is almost exact, it has been achieved using 798 *ab initio* calculations in the three dimensional space analyzed here. Klippenstein *et al.* have also used the analytical PES presented by Hirst-Hase to compute the reaction rate. This model, in addition to having required the manual optimization of almost 20 parameters, does not provide any indication of the associated model error. In the study of Harding *et al.*<sup>17</sup>, an on-the-fly method is used to compute the PES. The number of quantum calculations has not been presented, but, assuming that ten points are used to inte-

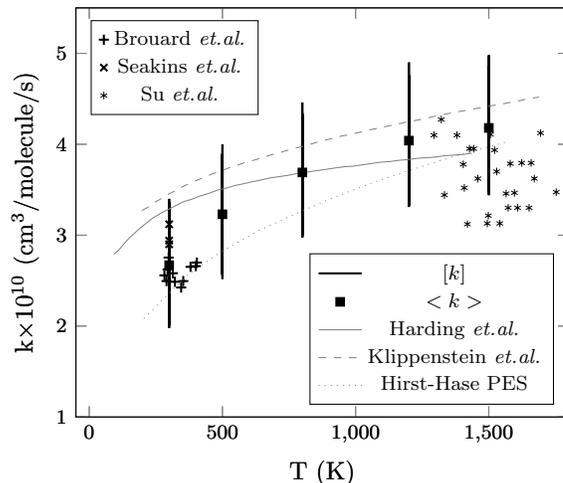


FIG. 12. Comparison of the high pressure  $\text{CH}_3+\text{H}$  recombination rate predicted by this work (black lines:  $[k]$ , squares:  $\langle k \rangle$ ) with other theoretical calculations (gray lines) and the experimental values (crosses<sup>66</sup>, asterisks<sup>67</sup>, times<sup>65</sup>). Gray lines: calculations using different representation of the PES; full line: on the fly calculation<sup>17</sup>, dashed line: 4D mathematical fit<sup>16</sup>, dotted line Hirst-Hase PES.

grate the partition function along  $q_0$  and  $q_1$  and 10 points to optimize  $ds$  and  $dx$ , this already results in 10,000 *ab initio* calculations. We recall that our model, even if associated with a 1.5 uncertainty factor on the reaction rate, is based on 4 parameters and has been set using 25 *ab initio* calculations for the calibration step (and 25 to examine the extrapolation capabilities). The restriction to the canonical ensemble is another issue which does not need more PES calculations to be overcome.

#### IV. CONCLUSIONS

We have presented in this work an original approach for the computation of statistical properties of molecular systems involving a large amplitude motion. The objectives were to propose: (i) a simple and general procedure to compute the kinetic energy of a LAM, (ii) a general procedure to calibrate an analytical PES using *ab initio* data, and (iii) a rigorous quantification of the uncertainty of the PES model, and their propagation to the  $\text{QoI}(s)$ . Two typical and different test cases have been considered for assessing the methodology: (i) the study of 1-3 butadiene, involving non trivial features like coupling of modes or a highly variable kinetic function, and (ii) the study of the  $\text{CH}_3+\text{H}$  recombination, for which a VTST-VRC approach is needed to compute an accurate reaction rate.

We proposed to compute the kinetic energy of a LAM based on a functional Z-Matrix formalism, e.g. a Z-Matrix, in which the internal DOF are defined with re-

spect to some generalized coordinates. The results obtained are exact within numerical precision and are in total agreement with previous exact methods. The particular advantage that our approach offers is its practical convenience. Indeed, for a typical LAM, the internal Z-Matrix coordinates naturally describes the configurational space of the motion and are much more suited than Cartesian coordinates. Also, the possibility to include ghost atoms in the Z-Matrix to represent a virtual pivot point is an additional important advantage when one wishes to study dissociation reactions involving loose complexes. Finally, the method applies equally regardless of the number of generalized coordinates or their type (length, angle, reaction path-like coordinate).

Furthermore, the calibration of parameters for the analytical PES from *ab initio* calculations has been achieved using Bayesian theory. The two examples treated have allowed us to point out different and interesting features of this approach. The most important one, which is critical to the derivation of analytical models, is that the PES model uncertainty is properly evaluated, and can be propagated to the QoI. We have also shown that, providing an adequate PES model is used, narrowing the QoI confidence interval needs significantly fewer data points than other methods, which does not exploit any particular physical contribution of the interaction energy. This is particularly true when the dimension of the problem increases. For instance, in the study of the  $\text{CH}_3+\text{H}$  recombination rate, we were able to use only 25 data points to calibrate the 3D PES in order to compute a reaction rate within an uncertainty factor of 1.5 (coming from the PES model). This corresponds to a typical reduction of one or two orders of magnitude in the amount of data points needed with respect to previous study. However, it is clear that this approach is relevant when the condition that a satisfactory model of the interaction energy is provided. Even for complex force fields, we have shown here that for the two studied applications, accurate models can be based on simple contributions: Morse-like type potentials for bond stretching, cosine-like functions for bending and torsional motion, and steric repulsion for non bonded atoms. We believe that these types of potentials should hold for the majority of torsional and bond breaking context studies. Finally, for the 1,3 butadiene application, we have discussed the possibility of performing a dual level inference process. While it is not possible to generalize the results, it was shown that the B3LYP/6-31G(d) level of theory was almost perfectly able to compute the true (or absolute) model error, even if the PES is not accurately rendered at this level. As knowing the model error is a useful information, which considerably reduces the required amount of data needed to obtain a given accuracy on the QoI, this property would be of a particular interest for the calibration of high level PESs at a minimum computational cost.

## Appendix A: Supplementary Information

### 1. SI 1. Fourier series development of the functions $f_i(\theta)$ for Butadiene (see table I)

$$\begin{aligned}
 f_1 &= -8.15 \cos(q)10^{-2} + 2.05 \sin(q)10^{-1} + 3.55 \sin(2q)10^{-1} \\
 &\quad + 7.38 \sin(3q)10^{-1} - 3.42 \sin(4q)10^{-1} \\
 f_2 &= -3.42 \cos(q)10^{-3} + 1.05 \sin(q)10^{-5} - 1.08 \cos(2q)10^{-2} - 4.21 \cos(3q)10^{-3} \\
 f_3 &= -1.06 \cos(q) - 8.48 \sin(q)10^{-4} + 4.93 \cos(2q)10^{-1} - 3.10 \cos(3q)10^{-1} \\
 &\quad + 3.43 \cos(4q)10^{-1} \\
 f_4 &= -5.95 \cos(q)10^{-2} - 1.05 \sin(q) - 2.28 \sin(2q) - 2.49 \sin(3q) \\
 &\quad + 3.92 \sin(4q)10^{-1} + 3.93 \sin(5q)10^{-1} + 2.12 \sin(6q)10^{-1} \\
 f_5 &= -7.91 \cos(q)10^{-2} - \sin(q) + 9.31 \cos(2q)10^{-3} + 3.57 \sin(2q) \\
 &\quad - 8.77 \cos(3q)10^{-3} + 7.90 \sin(3q)10^{-1} + 5.39 \cos(4q)10^{-2} - 5.46 \sin(4q)10^{-3} \\
 &\quad + 2.47 \cos(5q)10^{-2} - 3.27 \sin(5q)10^{-1} - 9.89 \cos(6q)10^{-3} + 6.02 \sin(6q)10^{-3} \\
 &\quad - 1.41 \cos(7q)10^{-2} + 1.1 \sin(7q)10^{-1} + 5.32 \cos(8q)10^{-3} + 8.94 \sin(8q)10^{-2} \\
 &\quad + 6.05 \cos(9q)10^{-2} \\
 f_6 &= -1.056 \cos(q) - 2.61 \sin(q)10^{-3} + 4.87 \cos(2q)10^{-1} - 3.08 \cos(3q)10^{-1} \\
 &\quad + 3.47 \cos(4q)10^{-1} \\
 f_7 &= 8.36 \cos(q)10^{-2} - 5.69 \sin(q)10^{-2} - 7.99 \cos(2q)10^{-2} - 5.7 \sin(2q) \\
 &\quad + 4.04 \cos(3q)10^{-2} - 3.28 \sin(3q) - 1.14 \cos(4q)10^{-3} + 1.03 \sin(4q) \\
 &\quad - 1.33 \cos(5q)10^{-2} + 6.22 \sin(5q)10^{-1} \\
 f_8 &= 5.9 \cos(q)10^{-3} - 1.2170 \sin(q) - 1.78 \cos(2q)10^{-2} - 2.68 \sin(2q) \\
 &\quad + 1.2 \cos(3q)10^{-3} - 3.27 \sin(3q) + 1.03 \cos(4q)10^{-2} + 7.85 \sin(4q)10^{-1} \\
 &\quad + 5.8 \cos(5q)10^{-3} + 4.46 \sin(5q)10^{-1} + 2.42 \sin(6q)10^{-1} - 4.9 \cos(7q)10^{-3} \\
 &\quad - 1.9 \sin(7q)10^{-1} + 6.83 \cos(8q)10^{-4} - 1.52 \sin(8q)10^{-1} - 1.10 \cos(9q)10^{-2} \\
 f_9 &= 2.5 \cos(q)10^{-3} - 1.11 \sin(q) - 2.26 \sin(2q) - 2.49 \sin(3q) + 3.66 \sin(4q)10^{-1} \\
 &\quad + 3.81 \sin(5q)10^{-1} + 1.81 \sin(6q)10^{-1}
 \end{aligned} \tag{A1}$$

### 2. SI 2. Normal mode frequencies of the 1,3 Butadiene at $\theta = 0$ and $\theta = 145^\circ$

List of the normal mode frequencies of the 1,3 Butadiene for  $\theta = 0^\circ$ :

$$\begin{aligned}
 f \text{ (cm}^{-1}\text{)} &= \{176.97 \text{ (torsion)}, 297.8, 515.78, 539.96, 781.87, 907.07, 927.71, \\
 &\quad 931.51, 1002.42, 1010.99, 1063.85, 1240.02, 1328.87, 1331.52, \\
 &\quad 1434.8, 1496.2, 1676.49, 1729.1, 3127.6, 3143.88, 3168.83, \\
 &\quad 3176.04, 3233.98, 3255.78\}
 \end{aligned} \tag{A2}$$

List of the normal mode frequencies of the 1,3 Butadiene for  $\theta = 145^\circ$ :

$$\begin{aligned}
 f \text{ (cm}^{-1}\text{)} &= \{186.12 \text{ (torsion)}, 273.93, 476.51, 625.72, 761.55, 890.42, 934.59, \\
 &\quad 936.96, 1024.91, 1041.86, 1078.38, 1112.56, 1323.89, 1356.82, \\
 &\quad 1455.35, 1483.65, 1689.57, 1715.05, 3109.54, 3142.19, 3150.97, \\
 &\quad 3160.79, 3232.96, 3239.67\}
 \end{aligned} \tag{A3}$$

<sup>1</sup>M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi,

- C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, "Gaussian 03, Revision C.02." Gaussian, Inc., Wallingford, CT, 2004.
- <sup>2</sup>M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery *J. Comput. Chem.*, vol. 14, p. 1347, 1993.
- <sup>3</sup>S. W. Benson, *The foundations of chemical kinetic*. Krieger, Malabar, 1982.
- <sup>4</sup>K. S. Pitzer and W. D. Gwinn *J. Chem. Phys.*, vol. 10, p. 428, 1942.
- <sup>5</sup>J. E. Kilpatrick and K. S. Pitzer *J. Chem. Phys.*, vol. 17, p. 1064, 1949.
- <sup>6</sup>G. Katzer and A. F. Sax *J. Phys. Chem. A*, vol. 106, p. 7204, 2002.
- <sup>7</sup>J. Pfäendtner, X. Yu, and L. BroadBelt *Theor. Chem. Acc.*, vol. 118, p. 881, 2007.
- <sup>8</sup>A. L. East and L. Radom *J. Chem. Phys.*, vol. 106, p. 6655, 1997.
- <sup>9</sup>B. M. Wong, R. L. Thom, and R. W. Field *J. Phys. Chem. A*, vol. 110, pp. 7406–7413, 2006.
- <sup>10</sup>J. Gang, M. J. Pilling, and S. H. Robertson *Chemical Physics*, vol. 231, pp. 183–192, 1998.
- <sup>11</sup>P. Vansteenkiste, V. V. S. D. Van Neck, and M. Waroquier *J. Chem. Phys.*, vol. 124, p. Art. No. 044314, 2006.
- <sup>12</sup>Y. Georgievskii and S. J. Klippenstein *J. Chem. Phys.*, vol. 118, no. 12, pp. 5542–5455, 2003.
- <sup>13</sup>S. Robertson, A. F. Wagner, and D. M. Wardlaw *J. Phys. Chem. A*, vol. 106, pp. 2598–2613, 2002.
- <sup>14</sup>S. C. Smith *J. Phys. Chem.*, vol. 111, no. 1999, p. 1830.
- <sup>15</sup>S. Robertson, A. F. Wagner, and D. M. Wardlaw *J. Chem. Phys.*, vol. 113, p. 2648, 2000.
- <sup>16</sup>S. J. Klippenstein, Y. Georgievskii, and L. B. Harding *Proceedings of the Combustion Institute*, vol. 29, pp. 1229–1236, 2002.
- <sup>17</sup>L. B. Harding, S. J. Klippenstein, and A. W. Jasper *Phys. Chem. Chem. Phys.*, vol. 9, pp. 4055–4070, 2007.
- <sup>18</sup>H. B. Schlegel *J. Comput. Chem.*, vol. 24, pp. 1514–1527, 2002.
- <sup>19</sup>G. C. Schatz, *Reaction and Molecular Dynamics*, vol. 75 of *Lecture Notes in Chemistry*. Springer, Berlin, 2000.
- <sup>20</sup>G. C. Schatz *Rev. Mod. Phys.*, vol. 61, p. 669, 1989.
- <sup>21</sup>G. Ochoa de Aspuru and M. L. Hernandez, *Reaction and Molecular Dynamics*. Lecture Notes in Chemistry, Springer, Berlin, 2000.
- <sup>22</sup>D. K. Hoffman, G. W. Wei, D. S. Zhang, and K. D. J. *Phys. Rev. E*, vol. 57, p. 6152, 1998.
- <sup>23</sup>T. Ishida and G. C. Schatz *J. Chem. Phys.*, vol. 107, p. 3558, 1997.
- <sup>24</sup>G. C. Maisuradze, D. L. Thompson, A. F. Wagner, and M. Minkoff *J. Chem. Phys.*, vol. 119, p. 10002, 2003.
- <sup>25</sup>G. K. Chawla, G. C. McBane, P. L. Houston, and G. C. Schatz *J. Chem. Phys.*, vol. 88, p. 5481, 1988.
- <sup>26</sup>G. C. Schatz, A. Papaioannou, L. A. Pederson, L. B. Harding, T. Hollebeek, T. S. Ho, and H. Rabitz *J. Chem. Phys.* 107, vol. 107, p. 2340, 1997.
- <sup>27</sup>A. J. C. Varandas *Adv. Chem. Phys.*, vol. 74, p. 255, 1988.
- <sup>28</sup>J. T. Murrel, S. Carter, S. C. Farantos, P. Huxley, and A. J. C. Varandas, *Molecular Potential Energy Functions*. Wiley, 1988.
- <sup>29</sup>A. Lagana, G. O. Aspuru, and E. Garcia *J. Chem. Phys.*, vol. 108, p. 3886, 1998.
- <sup>30</sup>A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff *J. Am. Chem. Soc.*, vol. 114, p. 10035, 1992.
- <sup>31</sup>J. W. Ponder and D. A. Case *Adv. Prot. Chem.*, vol. 66, pp. 27–85, 2003.
- <sup>32</sup>D. Leckband and J. Israelachvili *Quart. Rev. Biophys.*, vol. 34, pp. 105–267, 2001.
- <sup>33</sup>W. Xie, M. Orozco, D. J. Thrular, and J. Gao *J. Chem. Theory Comput.*, vol. 5, p. 459, 2009.
- <sup>34</sup>F. Cailliez and P. Pernot *J. Chem. Phys.*, vol. 134, no. 054124, 2011.
- <sup>35</sup>G. Reinisch, "opensoams, a C++ library for the study of the statistic of atomic and molecular systems," 2012.
- <sup>36</sup>E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular vibrations*. McGraw-Hill: New York, 1955.
- <sup>37</sup>J. G. Aston and M. L. Eidinoff *J. Chem. Phys.*, vol. 3, p. 379, 1935.
- <sup>38</sup>A. Fernández-Ramos, B. A. Ellingson, M.-P. Ruben, J. M. C. Marques, and D. G. Thrular *The*, vol. 118, p. 813, 2007.
- <sup>39</sup>G. Reinisch, J.-M. Leyssale, and G. L. Vignoles *J. Chem. Phys.*, vol. 133, p. 154112, 2010.
- <sup>40</sup>C. C. Martson and G. G. Balint-Kurti *J. Chem. Phys.*, vol. 91, p. 3571, 1989.
- <sup>41</sup>J. L. Beck and L. S. Katafygiotis, "Updating of a model and its uncertainties utilizing dynamic test data," *Proc. 1st International Conference on Computational Stochastic Mechanics*, pp. 125–136, 1991.
- <sup>42</sup>J. L. Beck and L. S. Katafygiotis, "Updating models and their uncertainties. I: Bayesian statistical framework," *ASCE Journal of Engineering Mechanics*, vol. 124, pp. 455–461, 1998.
- <sup>43</sup>S. H. Cheung and J. L. Beck, *Updating Reliability of Monitored Nonlinear Structural Dynamic Systems Using Real-time Data*. Proc. Inaugural International Conference of the Engineering Mechanics Institute (EM08), University of Minnesota, Minneapolis, Minnesota, USA, May 18–21, 2008.
- <sup>44</sup>G. S. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications*. Springer, New York, 1995.
- <sup>45</sup>E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.
- <sup>46</sup>E. E. Prudencio and K. Schulz, "The Parallel C++ Statistical Library 'QUESO': Quantification of Uncertainty for Estimation, Simulation and Optimization," 2011 (Submitted).
- <sup>47</sup>S. H. Cheung and J. L. Beck, "Calculation of posterior probabilities for Bayesian model class assessment and averaging from posterior samples based on dynamic system data," *Computer Methods in Applied Mechanics and Engineering*, pp. 304–321, 2010.
- <sup>48</sup>K. Miki, M. Panesi, E. E. Prudencio, and S. Prudhomme, "Probabilistic models and uncertainty quantification for the ionization reaction rate of atomic nitrogen," *Journal of Computational Physics*, vol. 231, pp. 3871–3886, 2012.
- <sup>49</sup>S. H. Cheung, T. A. Oliver, E. E. Prudencio, S. Prudhomme, and R. Moser, "Bayesian uncertainty analysis with applications to turbulence modeling," *Reliability Engineering and System Safety*, vol. 96, pp. 1137–1149, 2011.
- <sup>50</sup>R. R. Upadhyay, K. Miki, O. Ezekoye, and J. Marschall, "Uncertainty quantification of a graphite nitridation experiment using a Bayesian approach," *Experimental Thermal and Fluid Science*, vol. 35, pp. 1588–1599, 2011.
- <sup>51</sup>K. Miki, M. Panesi, E. E. Prudencio, and S. Prudhomme, "Estimation of the nitrogen ionization reaction rate using electric arc shock tube data and bayesian model analysis," *Physics of Plasmas*, vol. 19, 2012. 023507.
- <sup>52</sup>K. A. Holbrook, M. J. Pilling, and S. H. Robertson, *Unimolecular Reactions, second edition*. John Wiley and Sons: New York, 1996.
- <sup>53</sup>A. D. Becke *J. Chem. Phys.*, vol. 98, p. 1372, 1993.
- <sup>54</sup>M. M. Francl, W. H. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. Defrees, and J. A. Pople *J. Chem. Phys.*, vol. 77, p. 3654, 1982.
- <sup>55</sup>W. H. Stevens, H. Basch, M. Krauss, and P. C. Jasien *J. Chem.*, vol. 70, p. 612, 1992.
- <sup>56</sup>T. R. Cundari and W. H. Stevens *J. Chem. Phys.*, vol. 98, p. 5555, 1993.
- <sup>57</sup>K. S. Pitzer *J. Chem. Phys.*, vol. 14, pp. 239–243, 1946.

- <sup>58</sup>B. M. Wong and S. Raman *J. Comp. Chem.*, vol. 28, no. 4, p. 760.
- <sup>59</sup>S. J. Klippenstein *Chem. Phys. Lett.*, vol. 214, p. 418, 1993.
- <sup>60</sup>W. H. Miller *J. Phys. Chem.*, vol. 102, p. 793, 1998.
- <sup>61</sup>P. Piecuch and M. W. J. *Chem. Phys.*, vol. 123, p. 224105, 2005.
- <sup>62</sup>M. W., J. R. Gour, and P. Piecuch *J. Phys. Chem. A.*, vol. 111, p. 11359, 2007.
- <sup>63</sup>J. Zheng, J. R. Gour, J. J. Lutz, M. W., P. Piecuch, and D. G. Truhlar *J. Chem. Phys.*, vol. 128, p. 044108, 2008.
- <sup>64</sup>A. Dutta and C. D. Sherrill *J. Chem. Phys.*, vol. 118, p. 1610, 2003.
- <sup>65</sup>P. W. Seakins, S. H. Robertson, M. J. Pilling, D. M. Wardlaw, F. L. Nesbitt, R. P. Thorn, W. A. Payne, and L. J. Stief *J. Phys. Chem. A*, vol. 101, p. 9974, 1997.
- <sup>66</sup>M. Brouard, M. T. Macpherson, and M. J. Pilling *J. Chem. Phys.*, vol. S23, p. 199, 1989.
- <sup>67</sup>M. C. Su and J. V. Michael *Proc. Combust. Inst.*, vol. 29, p. 1219, 2002.