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

Guillaume Reinisch,*,[†] Kenji Miki,[†] Gérard L. Vignoles,[‡] Bryan M. Wong,[§] and Chris S. Simmons[†]

[†]Predictive Engineering and Computational Sciences (PECOS), Institute for Computational Engineering and Sciences, University of Texas, 78712 Austin, United States

[‡]Laboratory for Thermostructural Composites (LCTS), UMR-5801 (CNRS-SNECMA-CEA-UB1), University Bordeaux 1, 33600 Pessac, France

[§]Sandia: Materials Chemistry Department, Sandia National Laboratories, Livermore, California 94551, United States

ABSTRACT: 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) itself. 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 for quantification of uncertainties associated with a PES model as well as the forward propagation of these uncertainties to the quantity of interest. A selection and generalization of some treatments accounting for the coupling of the LAM with other internal or external DOF are also presented. Finally, we discuss and validate the approach with two applications: the calculation of the partition function of 1,3-butadiene and the calculation of the high-pressure reaction rate of the $CH_3 + H \rightarrow CH_4$ recombination.

INTRODUCTION

Quantum chemical methods, as implemented in many software packages,[1,2] 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.[3]

A reliable prediction of such properties requires tackling two main issues: First, it is necessary to calculate accurate molecular energies at different 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 that these approximations fail when large amplitude motions are involved,[4,5] and a rigorous treatment is often required if an accurate prediction of statistical properties is required.[6–8] This is particularly the case when a torsional motion is present[7,9–11] or when a loose TS is involved in a dissociation/recombination reaction.[12–14] 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.[4] Recent

studies[7,9,11] 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, two- to five-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,[12,13,15] 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.[12,16,17]

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.[18,19] Even though the first-principle calculations have proven their ability to yield accurate electronic energies, their computational cost does not generally allow for on-the-fly calculations,[20] and therefore, an appropriate interpolation method has to be used.[20] Three approaches can be distinguished: (i) local (or weighted) interpolation,[19,21–23] (ii) global (or nonweighted) interpolation by a set of mathematical functions,[16,24,25] and (iii) global interpolation by a parametrized physical model.[26–29] 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 applicability. 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.[30−32] 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.[33] 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 uncertainties 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 two-fold: 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 propagatation 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 a large amplitude motion (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 DOF (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.[7,11]

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 $CH_3 + H \rightarrow CH_4$ reaction rate. The presented methods are parts of the C++ library openSOAMS,[34] currently under development.

■ THEORY

The overall methodology proposed in this paper is illustrated in Figure 1. The particular steps that are emphasized in this work are (i) the inference of a PES model (parametrized by the vector θ) 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 a LAM into the following four 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
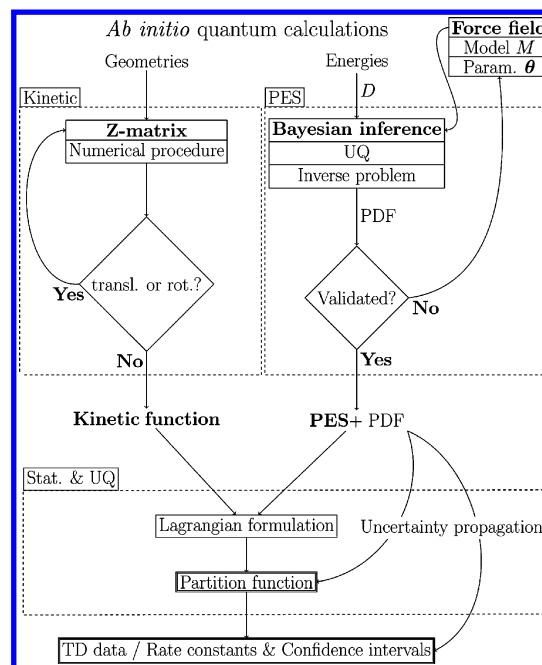

Figure 1. Flowchart of the thermal statistical computations in our approach. Acronyms and abbreviations: potential energy surface (PES); uncertainty quantification (UQ); probability density function (PDF); thermodynamic (TD); parameters (param.); translation (transl.); rotation (rot.); statistic (stat.).

translation, rigorously independent 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}, $Q_{HO}^{coupled}$, and $Q_{rot}^{coupled}$ in detail, and the partition function of the overall translation has been discussed extensively elsewhere.[35]

LAM. 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 **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:

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

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[36] allows a computation of the LAM partition function through a configurational integral involving the determinant of the kinetic matrix and the PES:

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

where k_B is the Boltzmann constant, T the temperature, n the dimension of $\mathbf{q}$, and δ the symmetry number of the motion (see the work of Fernández-Ramos et al.[37] 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 from 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

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

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.[38] An effective temperature-dependent kinetic constant is introduced by the following relation:

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

The Fourier grid Hamiltonian algorithm[39] is then employed to compute the eingenvalues $\varepsilon_i^{eff}(T)$ of the effective Hamiltonian defined by

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

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

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

Calibration of the PES. The calibration of a model can be viewed as the update of model parameters in order to get a better respresentation of observations. This model can then be used to make predictions of a quantity of interest for specific scenarios. When calibrating a physico-mathematical model of the PES 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' in Figure 1.

Check Model (In)adequacy. Let M designate a stochastic model class.[40,41] A stochastic model M is specified by a set of uncertain parameters $\mathbf{a}$, to which an additional uncertain parameter, variance (σ_{total}^2), is included as a measure of the total model error. Each stochastic model is then specified by the set $\theta = \mathbf{a} \cup \sigma_{total}^2 \in \Omega \subset \mathbb{R}_d$. One can use the data D to compute the posterior PDF $p(\theta|D,M)$ as defined by Bayes theorem:

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

where c is a normalization constant that makes the probability volume under the posterior PDF equal to unity, $p(D|\theta M)$ is the likelihood function, and $p(\theta|M)$ is the prior PDF for θ (always chosen to be uniform in this paper). The likelihood function expresses the probability of observing D based on the predictive PDF for the system output given by the set of parameters θ in the model M. To compute the likelihood function, the assumption of statistically independent error is used. We denote by D_j the jth 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:

$$r_{total} = Y_j - D_j \tag{11}$$

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 σ_{total}^2. Based on all these assumptions, the likelihood function reads

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

where N_d is the number of data points. The variance σ_{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 thus σ_{total}^2 will be referred to simply as σ_M^2. 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_{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 detail below.

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:

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

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

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

C

We use in this paper the adaptive multilevel stochastic simulation algorithm presented by Cheung and Beck[42] to generate the posterior samples. The integral in eq 13 is then approximated by

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

where $\mathbf{Q}^{(k)}$ is a sample simulated from $p(\mathbf{Q}|\theta^{(k)}, D, M)$. The Monte Carlo simulation method[43] can be used to simulate this sample. If $\mathbf{Q}$ is a deterministic function of q (i.e., $\mathbf{Q} = \mathbf{Q}(\theta)$), then $\mathbf{Q}^{(k)} = \mathbf{Q}(\theta^{(k)})$. Estimates for important statistical moments of $\mathbf{Q}$ conditioned on M and D can be obtained using the samples $\mathbf{Q}^{(k)}$, $k = 1, ..., K$. For instance, the posterior mean is calculated as follows:

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

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

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

In this paper, the mean and the 95% CI on the predicted posterior of $\mathbf{Q}$ are noted by $<\mathbf{Q}>$ and $[\mathbf{Q}]$, respectively. The library QUESO[44] is used to solve the inverse problem and to compute the posteriors $p(\theta|D, M)$. The current numerical methodology is very efficient and feasible for various engineering applications (e.g. refs 45−49).

Kinetic Function. The computation of the kinetic function |$\mathbf{A}(\mathbf{q})$| defined by eq 4 is achieved using the internal Z-matrix coordinates. As illustrated in the kinetic box in 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 parametrized 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 $\vec{J} = \vec{P} = 0$. 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:

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

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

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

The matrix $\underline{\mathbf{M}}_{rot}$ is optimized by an iterative procedure, generalizing the one presented in our previous paper[38] 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

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

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

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

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

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

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}.

Coupled Motions. *Flexible Rotor (FR) Partition Function.* The presence of the large amplitude motion leads to some coupling with the overall rotation, and the usual RR 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.[11] The partition function of the flexible rotor $Q_{FR}(T)$ is based on the rigid rotor (RR) expression,[50] calculated at each configuration point of the LAM and averaged using the probability density defined in relation (eq 6):

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

where $Q_{RR}(T, \mathbf{q})$ is the partition function of a RR[50] of the structure $\vec{Z}_{at}(\mathbf{q})$.

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.[9] and Vansteenkiste et al.[11] proposed to perform an integration similar to eq 23 by using a HO 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 is given by

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

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[35] of the structure $\vec{Z}_{at}(q_i)$, and $P(q_i, T)$ is the weighting factor defined by

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

■ APPLICATIONS

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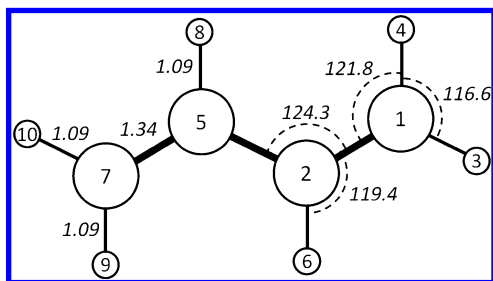
Figure 2. Illustration of the geometry of 1,3-butadiene at the global minimum of the PES. The large circles represent the carbon atoms, and the small circles represent the hydrogen atoms. Bond lengths are in angstroms and angles in degrees. The atom numbering used in the paper is indicated in the circles.

numbering used hereafter. The molecule contains a torsional motion defined by the relative rotation of the two H_2CH parts around the central C_5-C_2 bond. Previous studies[9,11] 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 DOF.

Electronic Structure Calculation. Ground-state electronic energies are calculated using the ab initio methods UB3LYP[51]/6-31G(d)[52] (hereafter referred to as DFT) for qualitative calculations, and RHF-CCSD(T)/aug-cc-pVDZ[53,54] (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 PES is obtained by computing the relaxed geometries with respect to the dihedral $C_1C_2C_5C_7$ using a 10° 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.[2]

Kinetic Function. The Table 1 presents the functional Z-matrix used to compute the kinetic function. The functions $f_i(\theta)$ ($i = 1, ..., 8$) define the relaxation of the DOF with respect to the generalized coordinate θ. The DOFs contributing to less than 1% of the kinetic function have been considered constant. The functions $f_i(\theta)$ are obtained by Fourier series fit to the optimized internal coordinates and are presented in Appendix A. The kinetic function calculated using our numerical approach is presented in Figure 3 (solid line) and compared to those reported by Vansteenkiste et al.[11] and Wong et al.[9] A very good agreement is observed between the three studies, the difference from the Eckart method at $\theta \approx 180°$ being most likely attributed to the different level of theory used in the study of Wong et al. for geometry optimization (MP2). We also

Table 1. Functional Z-Matrix Definition of the Torsional Motion of the 1,3-Butadiene[a]

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	**1243** + $f_3 (\theta)$	3	**180** + $f_4 (\theta)$
H_6	2	1.09	1	119.4	3	$f_5 (q)$
C_7	5	1.34	2	**124.3** + $f_6 (\theta)$	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** + $f_9 (\theta)$

[a]The definition of the relaxation functions $f_i(\theta)$ is given in Appendix A. Lengths are in angstroms and angles in degrees.
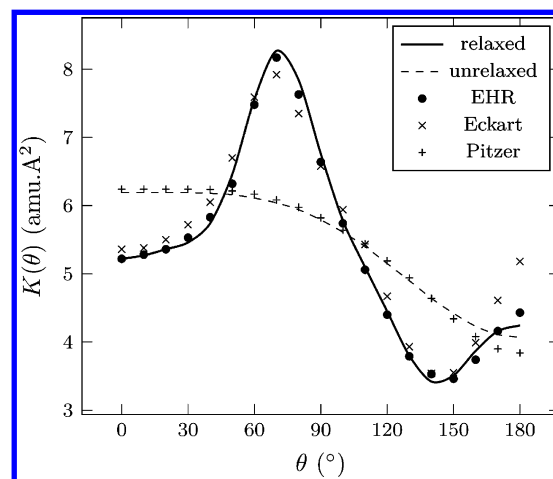
Figure 3. Comparison between the kinetic function calculated in this work (lines) and the kinetic function obtained using the EHR model[11] and the Eckart[9] and the Pitzer[55] methods (presented by Wong et al.).[9] Solid line: flexible torsion ($f_i(\theta) \neq 0$); dashed line: rigid rotation ($f_i(\theta) = 0$).

present in Figure 3 the kinetic function computed when the relaxation of the geometry with respect to θ 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.[9] calculated using the Pitzer method.[55] 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.

Calibration of the PES. 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 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 three parameters a_0, a_1, and a_2 by

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

$$\left. - \left(\frac{d_{vdW}}{|\vec{R}_9(0) - \vec{R}_4(0)|}\right)^9\right) \tag{26}$$

where d_{vdW} is taken equal to 2.4 Å (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 DOF 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 $g_{a_0}(-180) = g_{a_0}(180) = g_{a_0}(0) = 1$. 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 (σ_M^2).

In order to study how the ab initio data inform the PES parameters, different data sets are used for the inference process. They are denoted by 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 $[0;180°]$). The prior PDFs, the 95% CIs based on the posterior PDFs, and the posterior means are presented in Table 2.

Table 2. Definition of the Parameter's Prior and Description of the Parameter's Posterior for Each Data Set[a]

	$a_0 \times 10^2$	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]
$<\ >_5^{DFT}$	5.95	3.15	7.45	196.02
$<\ >_{10}^{DFT}$	5.92	3.17	7.5	16.30
$<\ >_{20}^{DFT}$	5.79	3.10	7.49	1.25
$\sigma_M^2 = cst$				
$[\ ]_2^{CC}$	[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
$<\ >_5^{CC}$	7.33	2.80	6.08	1.25

[a]Units: a_0, no unit; and a_1, a_2, and σ_M, kcal/mol.

We first comment on the results obtained using the DFT data sets. It is shown in rows 3−8 that the 95% CI is considerably narrowed around the mean value for each parameter each time the data set size increases. Also, the posterior mean value $<a_i>$ is already accurately predicted when only five 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 1.25×10^{-3} kcal^2/mol^2 (e.g., $<\sigma_M>$ = 3.5×10^{-2} 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 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 ≈90° and ≈180°) 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 D_n^{CC} (n = 2, 3, 5) presented in the four last rows of Table 2. In this case, we calibrate model parameters except for σ_M^2, which is set to be 1.25×10^{-3} kcal^2/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
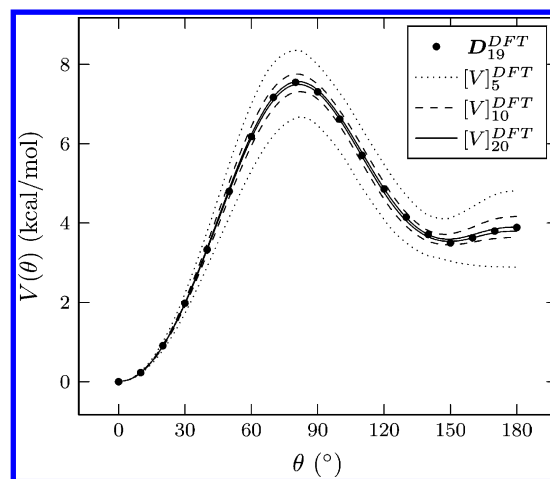

Figure 4. 95% CI of the posterior PES using the data set D_n^{DFT} with n = 5, 10, and 20 (dotted, dashed, and solid lines, respectively), and ● represents the data set D_{19}^{DFT}.

the true model error. As can be seen in Table 2, this assumption is reasonable since $[V]_5^{CC}$ is already very narrow (and actually converged) for all the parameters even when five points are used. It is also shown that using two points (at 0° and 180°) is not enough to infer a_0 and a_2 (posterior PDFs approximately equal to the prior PDFs), while a third point at 90° allows a considerable reduction of the uncertainty on a_2. These results are illustrated in Figure 5, which presents the uncertainty
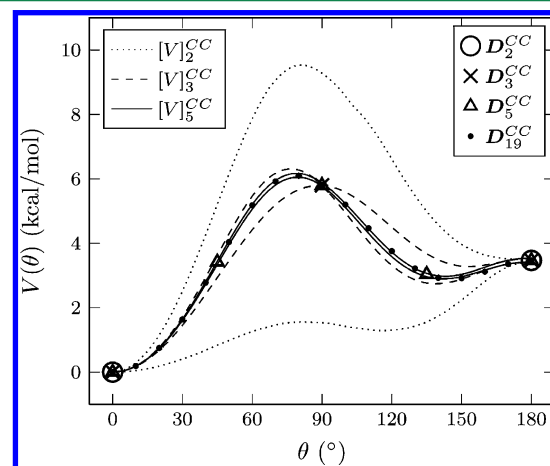

Figure 5. 95% CI of the posterior PES using the data set D_n^{CC} with n = 2, 3, and 5 (dotted, dashed, and solid lines, respectively). The ○, ×, and △ represent the corresponding data set, respectively, and ● represents the data set D_{19}^{CC}.

domain of the posterior PES for these three data sets. Comparing $[V]_5^{CC}$ *a posteriori* with the D_{19}^{CC} data set 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-pVDZ level of theory gives the exact PES).

Forward Problem: Partition Function. The 95% CI of the posterior partition function $[Q_{LAM}]^{DFT}$ using the $D_{n=5,10,20}^{DFT}$ data sets is 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 D_5^{DFT} is also presented. As for the PES results, the 95% CI becomes
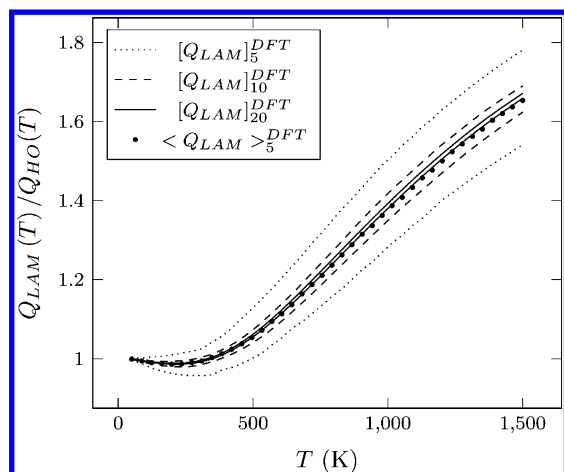
Figure 6. 95% CI of the posterior torsional partition function using the data set D_n^{DFT} with $n = 5$, 10, and 20 (dotted, dashed, and solid lines, respectively), and ● mean of the posterior torsional partition function using D_5^{DFT}. Results are normalized by the HO partition function.

very narrow as soon as 10 points are used, and the mean value $< Q >_5^{\mathrm{DFT}}$ is already almost converged. The fact that the $[Q]_n^X$ is centered around 1 at low temperature ($T < 400\ K$) is important because it demonstrates that the method presented here to treat the 1D LAM has a sufficient capability to perfectly reproduce 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 $\theta = 0°$ and 145° (the normal-mode frequencies are presented in Appendix B). It should be noted that the PES uncertainty also affects the FR and FO partition functions through eqs 6 and 25. The 95% CI of the posterior partition function of 1,3-butadiene using the D_5^{CC} (with σ_M^2 kept fixed at 1.25×10^{-3} kcal^2/mol^2) and D_{20}^{DFT} data sets are presented in Figure 7 and compared to the results presented by Wong and Raman[56] (Pert model) and those presented by Vansteenkiste et al.[11] [extended hindered rotor
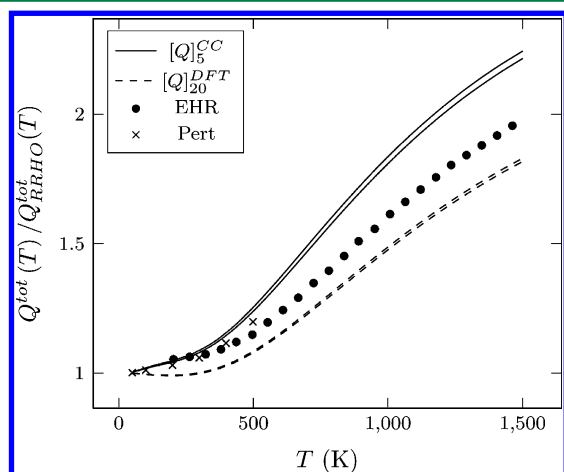
(EHR) model]. The results are normalized by the partition function obtained using a RRHO approximation. We first compare our prediction using D_5^{CC} with the low-temperature results presented by Wong and Raman[56] using their Pert method. The comparison is meaningful because the same level of theory 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.[11] (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 (0.1 kcal/mol). As can be seen in the figure, the behavior of the partition function curves is 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.[11] and Wong and Raman.[56] Also, this efficient formalism, in terms of computational time, further allows it to be used in conjunction with an uncertainty quantification algorithm.

Kinetic Rate of $CH_3 + H \rightarrow CH_4$ in the High-Pressure Limit. The reaction rate of the $CH_3 + H$ 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 parametrized using one pivot point $\vec{p}$ attached to the CH_3 part, around which the approaching H* atom is allowed to rotate. The optimal DS (noted DS*), defining the TS, is optimized for every temperature in a way that it is associated to a minimal partition function:[57,58]

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

where s is the reaction coordinate (separation $\vec{p}$−H*), and $Q_{\mathrm{DS}}^{\vec{p},s}(T)$ 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[50]

$$k(T) = \frac{k_{\mathrm{B}}T}{\hbar} \frac{Q_{\mathrm{DS}*}^{\vec{p},x}}{Q_{CH_3}(T)Q_H(T)} \exp(V*/k_{\mathrm{B}}T) \qquad (28)$$

where $\hbar$ is Planck's constant, Q_{CH_3} and Q_H the partition functions of the methyl and hydrogen radicals, respectively, and $V*$ is the electronic barrier height associated with the $CH_3 + H$ DS*. Over the small amplitude motion involved in the system, the CH_3 umbrella motion changes quite substantially and may

Figure 7. Comparison between the 95% CI of the posterior partition function of the butadiene using D_{20}^{DFT} (dashed line) and D_5^{CC} (solid line, $\sigma_M^2 = 1.25 10^{-3}$); ×, Pert method;[56] and ●, EHR method.[11] Results normalized to the RRHO partition function.

need to be included in a FO partition function. However, Klippenstein et al.[16] 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 to

$$k(T) = \frac{k_B T}{\hbar} \frac{Q^{2D}_{DS*}(T) Q^{FR}_{DS*}}{Q^{trans\&rot}_{CH_3}(T) Q_H(T)} \exp(V*/k_B T) \quad (29)$$

where Q^{2D}_{DS*} is the partition function of the H* motion on the DS*, Q^{FR}_{DS*} the effective overall rotation partition function of the DS, and $Q^{trans\&rot}_{CH_3}$ the partition function of the overall translational and rotational motion of CH_3.

Electronic Structure Calculation. Ground-state electronic energies are calculated using the CR-CC(2,3)[59,60]/aug-cc-pVDZ[53,54] 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.[61] Geometry optimizations are carried out at the UB3LYP[51]/6-31G(d)[52] level of theory. All the electronic calculations have been performed using the GAMESS code.[2]

Kinetic Function. The kinetic energy of the DS is obtained by defining the functional Z-matrix associated with the H* 2D motion. Because of the C_3 symmetry of CH_3, the DS is parametrized by two parameters. The first parameter is the distance x between the pivot point $\vec{p}$, lying in the C_3 axis of symmetry of 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 H* atom and the pivot point. The functional Z-matrix which describes the DS is presented in Table 3. The C_3 symmetry is imposed to the

Table 3. Z-Matrix Definition of the Dividing Surface Parameterized by the Reaction Coordinate s and the Pivot Point x[a]

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

[a]Lengths in angstroms, angles in degrees.

system, and the relaxation of the CH_3 part is taken into account to some extent by using the functions $a(|\overrightarrow{CH}*|)$ (interpolation between the optimized values on the MEP).

It has been shown that a limitation of the spherical dividing surfaces is that they overaccount artificial contributions to the reaction rate when multiple reaction paths are present.[12,13] To avoid this overestimation, a multifaceted dividing surface is 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 H* addition on CH_3, a typical dividing surface for the reaction is illustrated in Figure 8. To compute the DS partition function, the integration domain defined in eq 5 has to be restricted to
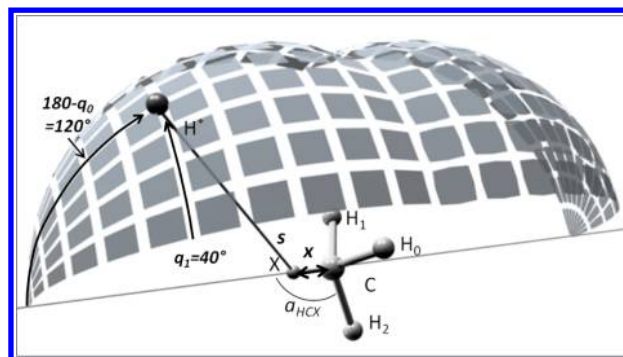

Figure 8. Illustration of a typical bifaceted dividing surface of the CH_3 + H → CH_4 reaction.

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

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 CH_3 and the DS.

Calibration of the PES. The PES of the CH_3/H* interaction accounts for three contributions: (i) the stretching energy of the C−H* bond (noted V_{str}), (ii) the bending energy of the H_iCH* DOF (noted V_{bend}), and (iii) the steric repulsion energy between the H_i and H*. The stretching energy V_{str} is written as a Morse-like function of $|\overrightarrow{CH}*|$ with an origin of energy taken at the products state:

$$V^{a_0,a_1}_{str} = D_e[1 - \exp(-a_0(|\overrightarrow{CH}*| - r_{eq})^{a_1})]^2 - D_e \quad (31)$$

where $D_e = 107.75$ kcal/mol is the dissociation energy, $r_{eq} = 1.09$ Å is the equilibrium bond length, and a_0 and a_1 are the first two parameters of the PES model. Note that the function is defined only for $|\overrightarrow{CH}*| > r_{eq}$. The bending energy is modeled by a simple cosine function, with a barrier height equal to the dissociation energy at the angle $\alpha_{XCH*} = \pm \pi/2$:

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

Finally the PES also accounts for the spherical repulsion between the passive hydrogen atoms H_i and H*:

$$V^{a_2,a_3}_{rep} = a_2 \sum_{i=0,1,2} \left[\left(\frac{\sigma_{HH}}{|\overrightarrow{H_iH}*|}\right)^{a_3} - \left(\frac{\sigma_{HH}}{|\overrightarrow{H_iH*}_0|}\right)^{a_3} \right] \quad (33)$$

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

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

The geometries used to compute the electronic energies are defined by the Z-matrix presented in Table 3 with $x = \varepsilon = 10^{-5}$

H

dx.doi.org/10.1021/ct300278x | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 4. Prior PDFs of the Parameters and 95% CIs ([ ])^a

	a_0	a_1	a_2	a_3	σ_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

^{a}Based on the posterior PDFs, and the posterior mean values (< . >). Units: a_0, Å$^{-a_1}$; a_1: no unit; and a_2, a_3, and σ_M, kcal/mol.

(in order to preserve a nonambiguous definition of the C_3 axis). The 50 data points collected are divided in two groups: (i) one is used to infer the PES model (noted D^*), and (ii) the second is used to validate it for extrapolation (noted D). The data set D^* contains 25 points in the relevant space of the PES for reaction rate calculation and is constituted by:

(i) 10 points on the MEP ($q_0 = q_1 = 0$) from $s = 2.0–3.8$ Å.
(ii) 15 points out of the MEP defined by the combination $s \in \{2.0, 2.4, 2.8 \text{ Å}\}$ and $q_0 \in \{20,40,60,80,100°\}$ with $q_1 = 0$.

The D data set collects the energies of the corresponding geometries of (ii) at $q_1 = 30°$ as well as 10 other points at $s = 1.6$ Å. The prior intervals, the posteriors means, and 95% CIs are presented in Table 4, and the 4 posteriors $p(a|D^*,M)$ are presented in figure Figure 9. We confirm here that the Morse
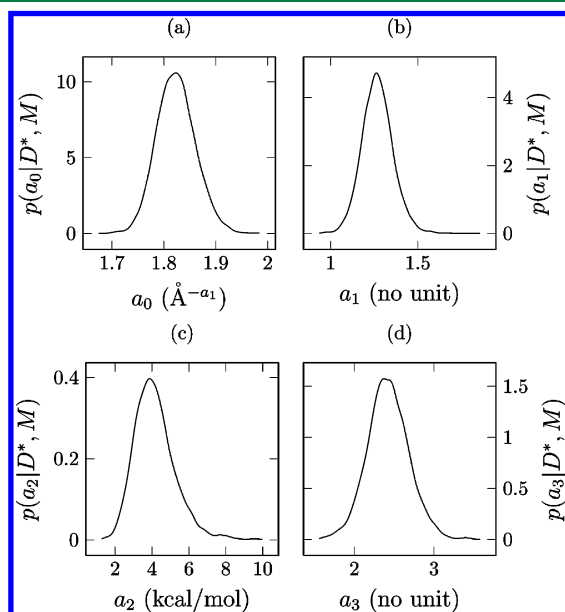

Figure 10. Comparison between the predicted 95% CI of the MEP ([V]), the data sets D and D^* and the ab initio results presented in literature. Lines, [V]; ●, D^*; and ○, D. Gray symbols, literature results: *, CCSD(T)/6-31G(d);[17] †+, CASPT2/aug-cc-pVQZ;[17] and ×, full CI/6-31G(d).[62] Origin of energy: $V(4 \text{ Å},0,0)$.

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 shown in Figure 11 for out-of-MEP situations. The symbols still represent the

Figure 9. Posterior PDF, $p(a_i|D^*,M)$, (a–d) for respectively $i = 0, 1, 2$, and 3.

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.

We compare in Figure 10 the 95% CI of the posterior MEP ([V]) with the data points of D^* (solid circles) and D (open circles) included in the MEP. We also report in this figure the ab initio results at the CASPT2/aug-cc-pVQZ,[17] full-CI/6-31G(d),[62] and CCSD(T)/6-31G(d)[17] level of theories presented in the literature. The uncertainty of the PES model is under 2 kcal/mol for separations higher than 2.0 Å and coincidently encapsulates the results obtained at the full CI and CASPT2 level of theory. As already reported,[17] the CCSD(T)
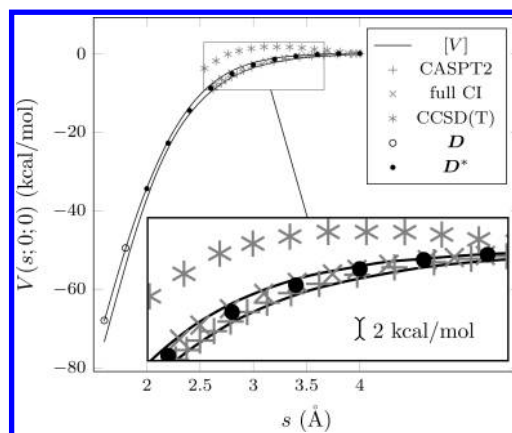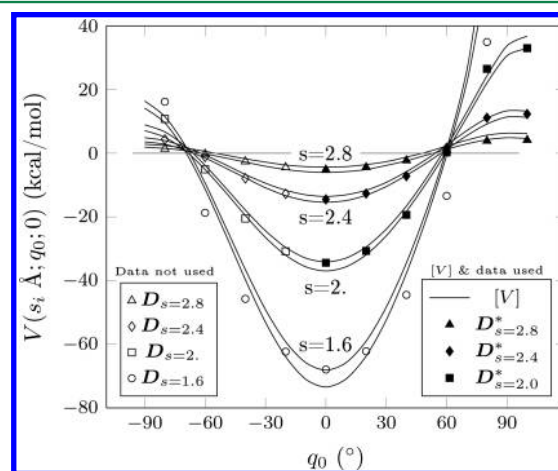

Figure 11. Comparison between the predicted 95% CI of the PES ([V]) with the data sets D and D^*. Lines, [V]; filled symbols, D^*; and empty symbols, D. Origin of energy: $V(4 \text{ Å},0,0)$.

D^* and D data sets respectively. The approximation of the PES model is very satisfying. Even for extrapolated values at low separation ($s = 1.6$ Å), the model still performs well with the hindering domain being correctly predicted within 10%.

Forward Propagation to the High-Pressure Recombination Rate. The 95% CI ([k]) and the mean (<k>) of the posterior high-pressure reaction are presented in Figure 12 with the available experimental measurements[63−65] and the VTST-VRC theoretical predictions presented by Klippenstein et al.[16]
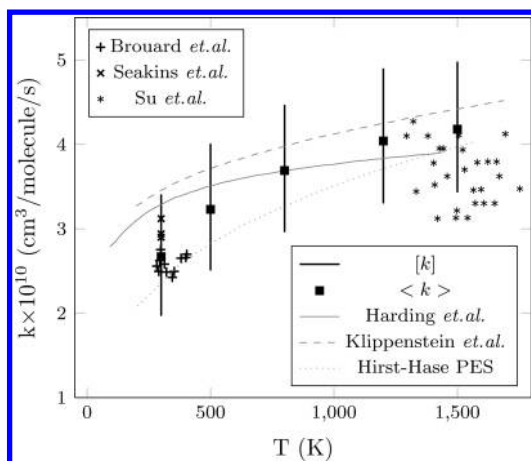
Figure 12. Comparison of the high pressure CH_3 + H recombination rate predicted by this work (black lines, $[k]$; ■, $<k>$) with other theoretical calculations (gray lines) and the experimental values (+,[64] *,[65] ×[63]). Gray lines, calculations using different representation of the PES; full line, on the fly calculation;[17] dashed line, 4D mathematical fit;[16] and dotted line, Hirst−Hase PES.

and Harding et al.[17] The experimental measures have been converted from the k_0 values of the original experimental work in the same way presented by Klippenstein et al.[16] The CIs 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 dividing surface does not constitute a perfect DS and is associated with an overestimation of the reaction rate. The works of Klippenstein et al.[16] have shown that by using both VTST-VRC and a direct dynamic method, the spherical DS for the CH_3 + H → 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,[12] 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 CI at the s = 2.8 Å case in 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.[17] and Klippenstein et al.[16] In the works of Klippenstein et al.,[16] a combined Fourier series/3D spline fitting procedure is used to obtain a 4D 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 3D 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.,[17] 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 integrate 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.

■ 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 QoI(s). Two typical and different test cases have been considered for assessing the methodology: (i) the study of 1,3-butadiene, involving nontrivial features, such as coupling of modes or a highly variable kinetic function; and (ii) the study of the CH_3 + 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 DOFs are defined with respect 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 describe 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, and 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 CI 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 CH_3 + 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 studies. 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-type potentials for bond stretching, cosine-like functions for bending and torsional motion, and steric repulsion for nonbonded atoms. We believe that these types of potentials should hold for the majority of torsional and simple 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 a prior estimation of the model error allows to considerably reduce the required amount of data needed to obtain a given accuracy on the QoI, this property would be of particular interest for the calibration of high-level PESs at a minimum computational cost.

■ APPENDIX

A. Fourier Series Development of the Functions $f_i(\theta)$ for 1,3-Butadiene (see Table 1)

$$f_1 = -8.15\cos(\theta)10^{-2} + 2.05\sin(\theta)10^{-1} + 3.55\sin(2\theta)10^{-1} + 7.38\sin(3\theta)10^{-1} - 3.42\sin(4\theta)10^{-1}$$

$$f_2 = -3.42\cos(\theta)10^{-3} + 1.05\sin(\theta)10^{-5} - 1.08\cos(2\theta)10^{-2} - 4.21\cos(3\theta)10^{-3}$$

$$f_3 = -1.06\cos(\theta) - 8.48\sin(\theta)10^{-4} + 4.93\cos(2\theta)10^{-1} - 3.10\cos(3\theta)10^{-1} + 3.43\cos(4\theta)10^{-1}$$

$$f_4 = -5.95\cos(\theta)10^{-2} - 1.05\sin(\theta) - 2.28\sin(2\theta) - 2.49\sin(3\theta) + 3.92\sin(4\theta)10^{-1} + 3.93\sin(5\theta)10^{-1} + 2.12\sin(6\theta)10^{-1}$$

$$f_5 = -7.91\cos(\theta)10^{-2} - \sin(\theta) + 9.31\cos(2\theta)10^{-3} + 3.57\sin(2\theta) - 8.77\cos(3\theta)10^{-3} + 7.90\sin(3\theta)10^{-1} + 5.39\cos(4\theta)10^{-2} - 5.46\sin(4\theta)10^{-1} + 2.47\cos(5\theta)10^{-2} - 3.27\sin(5\theta)10^{-1} - 9.89\cos(6\theta)10^{-3} + 6.02\sin(6\theta)10^{-3} - 1.41\cos(7\theta)10^{-2} + 1.1\sin(7\theta)10^{-1} + 5.32\cos(8\theta)10^{-3} + 8.94\sin(8\theta)10^{-2} + 6.05\cos(9\theta)10^{-2}$$

$$f_6 = -1.056\cos(\theta) - 2.61\sin(\theta)10^{-3} + 4.87\cos(2\theta)10^{-1} - 3.08\cos(3\theta)10^{-1} + 3.47\cos(4\theta)10^{-1}$$

$$f_7 = 8.36\cos(\theta)10^{-2} - 5.69\sin(\theta)10^{-2} - 7.99\cos(2\theta)10^{-2} - 5.7\sin(2\theta) + 4.04\cos(3\theta)10^{-2} - 3.28\sin(3\theta) - 1.14\cos(4\theta)10^{-3} + 1.03\sin(4\theta) - 1.33\cos(5\theta)10^{-2} + 6.22\sin(5\theta)10^{-1}$$

$$f_8 = 5.9\cos(\theta)10^{-3} - 1.2170\sin(\theta) - 1.78\cos(2\theta)10^{-2} - 2.68\sin(2\theta) + 1.2\cos(3\theta)10^{-3} - 3.27\sin(3\theta) + 1.03\cos(4\theta)10^{-2} + 7.85\sin(4\theta)10^{-1} + 5.8\cos(5\theta)10^{-3} + 4.46\sin(5\theta)10^{-1} + 2.42\sin(6\theta)10^{-1} - 4.9\cos(7\theta)10^{-3} - 1.9\sin(7\theta)10^{-1} + 6.83\cos(8\theta)10^{-4} - 1.52\sin(8\theta)10^{-1} - 1.10\cos(9\theta)10^{-2}$$

$$f_9 = 2.5\cos(\theta)10^{-3} - 1.11\sin(\theta) - 2.26\sin(2\theta) - 2.49\sin(3\theta) + 3.66\sin(4\theta)10^{-1} + 3.81\sin(5\theta)10^{-1} + 1.81\sin(6\theta)10^{-1}$$

$$(35)$$

B. Normal Mode Frequencies of 1,3-Butadiene at $\theta = 0$ and 145°

List of the normal mode frequencies of 1,3-butadiene for $\theta = 0°$:

$$f(cm^{-1}) = \{176.97(torsion), 297.8, 515.78, 539.96,$$
$$781.87, 907.07, 927.71, 931.51, 1002.42,$$
$$1010.99, 1063.85, 1240.02, 1328.87,$$
$$1331.52, 1434.8, 1496.2, 1676.49, 1729.1,$$
$$3127.6, 3143.88, 3168.83, 3176.04,$$
$$3233.98, 3255.78\} \tag{36}$$

List of the normal mode frequencies of the 1,3-butadiene for $\theta = 145°$:

$$f(cm^{-1}) = \{186.12(torsion), 273.93, 476.51, 625.72,$$
$$761.55, 890.42, 934.59, 936.96, 1024.91,$$
$$1041.86, 1078.38, 1112.56, 1323.89, 1356.82,$$
$$1455.35, 1483.65, 1689.57, 1715.05, 3109.54,$$
$$3142.19, 3150.97, 3160.79, 3232.96, 3239.67\} \tag{37}$$

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: guillaume@ices.utexas.edu.

Notes
The authors declare no competing financial interest.

■ REFERENCES

(1) Frisch, M. J. et al. *Gaussian 03*, revision C.02; Guassian, Inc.: Wallingford, CT, XXXX.
(2) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(3) Benson, S. W. *The Foundations of Chemical Kinetics*; R.E. Kriegler: Malabar, FL, 1982.
(4) Pitzer, K. S.; Gwinn, W. D. *J. Chem. Phys.* **1942**, *10*, 428−440.
(5) Kilpatrick, J. E.; Pitzer, K. S. *J. Chem. Phys.* **1949**, *17*, 1064−1075.
(6) Katzer, G.; Sax, A. F. *J. Phys. Chem. A* **2002**, *106*, 7204−7215.
(7) Pfaendtner, J.; Yu, X.; Broadbelt, L. J. *Theor. Chem. Acc.* **2007**, *118*, 881−898.
(8) East, A. L. L.; Radom, L. *J. Chem. Phys.* **1997**, *106*, 6655−6674.
(9) Wong, B. M.; Thom, R. L.; Field, R. W. *J. Phys. Chem. A* **2006**, *110*, 7406−7413.
(10) Gang, J.; Pilling, M. J.; Robertson, S. H. *Chem. Phys.* **1998**, *231*, 183−192.
(11) Vansteenkiste, P.; Van Neck, D.; Van Speybroack, V.; Waroquier, M. *J. Chem. Phys.* **2006**, *124*, 044314.
(12) Georgievskii, Y.; Klippenstein, S. J. *J. Chem. Phys.* **2003**, *118*, 5442−5456.
(13) Robertson, S. H.; Wagner, A. F.; Wardlaw, D. M. *J. Phys. Chem. A* **2002**, *106*, 2598−2613.
(14) Smith, S. C. *J. Chem. Phys.* **1999**, *111*, 1830−1842.
(15) Robertson, S. H.; Wagner, A. F.; Wardlaw, D. M. *J. Chem. Phys.* **2000**, *113*, 2648−2661.
(16) Klippenstein, S. J.; Georgievskii, Y.; Harding, L. B. *Proc. Combust. Inst.* **2002**, *29*, 1229−1236.
(17) Harding, L. B.; Klippenstein, S. J.; Jasper, A. W. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4055−4070.
(18) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514−1527.
(19) Schatz, G. C. *Reaction and Molecular Dynamics*; Springer: Berlin, Germany, 2000; Vol. 75, pp 15−32.
(20) Schatz, G. C. *Rev. Mod. Phys.* **1989**, *61*, 669−688.
(21) Hoffman, D. K.; Wei, G. W.; Zhang, D. S.; Kouri, D. J. *Phys. Rev. E.* **1998**, *57*, 6152−6160.

(22) Ishida, T.; Schatz, G. C. *J. Chem. Phys.* **1997**, *107*, 3558−3568.

(23) Maisuradze, G. G.; Thompson, D. L.; Wagner, A. F.; Minkoff, M. *J. Chem. Phys.* **2003**, *119*, 10002−10014.

(24) Chawla, G. K.; McBane, G. C.; Houston, P. L.; Schatz, G. C. *J. Chem. Phys.* **1988**, *88*, 5481−5488.

(25) Schatz, G. C.; Papaioannou, A.; Pederson, L. A.; Harding, L. B.; Hollebeek, T.; Ho, T. S.; Rabitz, H. *J. Chem. Phys.* **1997**, *107*, 2340−2350.

(26) Varandas, A. J. C. *Adv. Chem. Phys.* **1988**, *74*, 255−338.

(27) Murrel, J. N.; Carter, S.; Huxley, P.; Farantos, S. C.; Varandas, A. J. C. *Molecular Potential Energy Functions*; John Wiley and Sons: Hoboken, NJ, 1988.

(28) Laganá, A.; Aspuru, G. O. d.; Garcia, E. *J. Chem. Phys.* **1998**, *108*, 3886−3896.

(29) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−1035.

(30) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27−85.

(31) Leckband, D.; Israelachvili, J. *Q. Rev. Biophys.* **2001**, *34*, 105−267.

(32) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459−467.

(33) Cailliez, F.; Pernot, P. *J. Chem. Phys.* **2011**, *134*, 054124.

(34) Reinisch, G. *OpenSOAMS, a C++ library for the Statistics Of Atomic and Molecular Systems*, 2012. http://opensoams.googlecode.com/svn/trunk/docs/html/index.html/.

(35) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill, New York, 1955, 11−22.

(36) Eidinoff, M. L.; Aston, J. G. *J. Chem. Phys.* **1935**, *3*, 379−383.

(37) Fernández-Ramos, A.; Ellingson, B. A.; Meane-Pañeda, R.; Marques, J. M. C.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *118*, 813−826.

(38) Reinisch, G.; Leyssale, J.-M.; Vignoles, G. L. *J. Chem. Phys.* **2010**, *133*, 154112.

(39) Marston, C. C.; Balint-Kurti, G. G. *J. Chem. Phys.* **1989**, *91*, 3571−3576.

(40) Beck, J. L.; Katafygiotis, L. S. Proceedings of the 1st European Congress on Computational Mechanics, Munich, Germany, August 31−September 3, 1999,1991; pp 125−136.

(41) Katafygiotis, L. S.; Beck, J. L. *J. Eng. Mech. Div., Am. Soc. Civ. Eng.* **1998**, *1244*, 455−461.

(42) Cheung, S. H.; Beck, J. L. Updating Reliability of Monitored Nonlinear Structural Dynamic Systems Using Real-time Data. Proceedings of the Inaugural International Conference of the Engineering Mechanics Institute (EM08), Minneapolis, MN, May 18−21, 2008; University of Minnesota: Minneapolis, MN, 2008.

(43) Fishman, G. S. *Monte Carlo: Concepts, Algorithms, and Applications*; Springer-Verlag: New York, 1996, 145−255.

(44) Prudencio, E.; Schulz, K. The parallel C++ statistical library QUESO: Quantification of Uncertainty for Estimation, Simulation and Optimization. In *Euro-Par 2011 Workshops, Part I*; Alexander et al., Ed.; Springer-Verlag: Berlin Heidelberg, 2012; Vol. *7155*, pp 398−407.

(45) Cheung, S. H.; Beck, J. L. *Comput. Methods Appl. Mech. Engin.-Aided Civ. Inf.* **2010**, 304−321.

(46) Miki, K.; Panesi, M.; Prudencio, E. E.; Prudhomme, S. *J. Comput. Phys.* **2012**, *231*, 3871−3886.

(47) Cheung, S. H.; Oliver, T. A.; Prudencio, E. E.; Prudhomme, S.; Moser, R. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 1137−1149.

(48) Upadhyay, R. R.; Miki, K.; Ezekoye, O.; Marschall, J. *Exp. Therm. Fluid Sci.* **2011**, *35*, 1588−1599.

(49) Miki, K.; Panesi, M.; Prudencio, E. E.; Prudhomme, S. *Phys. Plasmas* **2012**, *19*, 023507, 023507.

(50) Holbrook, K. A.; Pilling, M. J.; Robertson, S. H. *Unimolecular Reactions*, 2nd ed.; John Wiley: Chichester, U.K., 1996; pp 39−113.

(51) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372−1377.

(52) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.

(53) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *J. Chem. Phys.* **1992**, *70*, 612−630.

(54) Cundari, T. R.; Stevens, W. J. *J. Chem. Phys.* **1993**, *98*, 5555−5565.

(55) Pitzer, K. S. *J. Chem. Phys.* **1946**, *14*, 239−243.

(56) Wong, B. M.; Raman, S. *J. Comput. Chem.* **2007**, *28*, 759−766.

(57) Klippenstein, S. J. *Chem. Phys. Lett.* **1993**, *214*, 418−424.

(58) Miller, W. H. *J. Phys. Chem. A* **1998**, *102*, 793−806.

(59) Piecuch, P.; Włoch, M. *J. Chem. Phys.* **2005**, *123*, 224105.

(60) Włoch, M.; Gour, J. R.; Piecuch, P. *J. Phys. Chem. A* **2007**, *111*, 11359−11382.

(61) Zheng, J.; Gour, J. R.; Lutz, J. J.; Włoch, M.; Piecuch, P.; Truhlar, D. G. *J. Chem. Phys.* **2008**, *128*, 044108.

(62) Dutta, A.; Sherrill, C. D. *J. Chem. Phys.* **2003**, *118*, 1610−1619.

(63) Seakins, P. W.; Robertson, S. H.; Pilling, M. J.; Wardlaw, D. M.; Nesbitt, F. L.; Thorn, R. P.; Payne, W. A.; Stief, L. J. *J. Phys. Chem. A* **1997**, *101*, 9974−9987.

(64) Brouard, M.; Macpherson, M. T.; Pilling, M. J. *J. Phys. Chem.* **1989**, *93*, 4047−4059.

(65) Su, M. C.; Michael, J. V. *Proc. Combust. Inst.* **2002**, *29*, 1219−1227.

L

dx.doi.org/10.1021/ct300278x | *J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX