

Coupled Anharmonic Thermochemistry from Stratified Monte-Carlo Integration

Gabriel Rath,^{*,†,‡} Wassja A. Kopp,[†] and Kai Leonhard^{*,†}

[†]*Institute of Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany*

[‡]*Software for Chemistry & Materials, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands*

E-mail: gabriel.rath@itt.rwth-aachen.de; kai.leonhard@itt.rwth-aachen.de

Phone: +49 (0)241 80 98174. Fax: +49 (0)241 80 92255

Abstract

This study presents CIMCI, a new semi-classical method for handling fully coupled anharmonicity in gas-phase thermodynamics that promises to be black-boxable, to be applicable for all kinds of anharmonicity, and to scale better at higher dimensionality than other methods for handling gas-phase molecular anharmonicity. The method does so by using automatically and recursively stratified, simultaneous Monte Carlo integration of multiple functions. For the small systems analyzed by this study, the method’s anharmonic corrections match reference data better than those of other black-box anharmonic methods, *e.g.* VPT2. This holds even when sampling with CIMCI is done with primitive force fields, *e.g.* UFF, while the competing methods are used with proper, comprehensive potentials, *e.g.* the M06-2X meta-hybrid DFT functional. With further refinements in Monte Carlo sampling efficiency, in the quality of fast potentials practical for Monte Carlo sampling, and in automatic detection of which stereoisomers should be included during sampling, CIMCI has the potential to be the

ideal anharmonic treatment for larger molecules where the large number of conformers and the high dimensionality of coupled torsions present major difficulties for other, existing treatments for anharmonicity.

Introduction

Accurate thermochemical data are needed to model atmospheric, industrial, and combustion processes. Due to their influence on energy balances and reaction equilibria, significant scientific effort has been invested in their prediction.¹ Such chemical quantities depend on contributions from both electronic states and nuclear motions. Lingering gross inaccuracies in either can cascade into gross inaccuracies in thermodynamic quantities.

By today’s standards, electronic energies can often be calculated very accurately. This is mainly thanks to electron correlation-including methods like Møller-Plesset perturbation theory of second-order (MP2) and coupled-cluster (CC) methods.² Even with modern density functional theory (DFT) methods, barrier heights are accessible from (double-hybrid) density functionals at an uncertainty below 8 kJ mol^{-1} .³ Since most computational chemistry focuses on electronic contributions to potential energy surfaces (PESs), electronic methods have become good enough that their errors can be overshadowed by errors in nuclear contributions.

Methods to calculate nuclear contributions have not fared so well. The commonly used harmonic oscillator (HO) model can be especially inaccurate for motions that are not small-amplitude, linear oscillations. If a species’ nuclear motions do not conform well enough to the expectations of the employed model, thermochemical predictions can be inaccurate, even when using a high-performance electronic structure method for the PES.⁴⁻⁶ For example, every torsion wrongly treated as a HO can influence a species’ entropy by as much as $1.8 \text{ J mol}^{-1} \text{ K}^{-1}$.⁴ This error can ripple onto Gibbs free energies, which determine rate constants and reaction equilibria. To remedy this, various approximations to the exact solutions of the nuclear Schrödinger Equation (SE) have been proposed. Many of these are used in hybrid schemes, combining spectroscopic experimental data with statistical mechanical models.

As of writing, these schemes reach benchmark accuracy for small molecules, especially when coupled anharmonicity has been accounted for.^{7–11} However, due to the “curse of dimensionality”,¹² detailed anharmonicity treatments are rarely practical for larger molecules. For example, constructing a hyperrectangular grid with 10 points assigned to each dimension for a 12-atomic system using some DFT method would require more computational time than a multiple of the current age of the universe.¹³ A different avenue is to describe the motion of the nuclei classically and apply some correction scheme, *e.g.* the Pitzer-Gwinn (PG) approximation,¹⁴ to correct for quantum effects.¹⁵ This way allows for the use of Monte Carlo (MC) integration, an integration technique well-known to handle high dimensionality better than grid-based methods.

This study presents Configuration Integral Monte Carlo Integration (CIMCI), a semiclassical method for handling nuclear motion that combines the framework of a detailed treatment of anharmonicity with the beneficial scaling of MC integration. The approach converges to the exact quantum solution at higher temperatures, all the while avoiding numerical issues from summing a vast amount of small, high-energy contributions. It also does not need the excessive amount of wavefunctions ordinarily required in a quantum approach at these temperatures. Crucial to its function is MISER,¹⁶ a generic sampling strategy for MC integration that dramatically reduces numerical uncertainties. With slight modification, MISER has been found to be highly suitable for CIMCI. CIMCI relies on integrating many different functions in the same phase space simultaneously, and MISER’s method of automatically biasing the otherwise uniformly random sampling of MC integration can be tuned to accommodate for this.

The rest of the paper is presented in three parts. First, to present CIMCI in context, there is a brief review of the strengths and limitations of common methods for treating anharmonicity, both quantum-mechanical and semiclassical ones. Next come descriptions for CIMCI’s physical basis and the specific implementation of its MC integration. Finally, results obtained from CIMCI are compared to recent benchmark computations for anharmonicity

at intermediate-to-high temperatures.

State-of-the-art

Table 1: Comparison of selected, state-of-the-art treatments for anharmonicity

Method	Quantum?	Anharmonicities from Asymmetry?	Anharmonicities from Torsions?	Possible Superposition Errors?
VPT2 ¹⁷	Yes	Yes	No	No
VSCF ^{18–20}	Yes	PES-Dependent	PES-Dependent	No
Multiconformer ^{21,22} / Multistructural ²³	Method-Dependent	Same as Base Method	Yes	Vibrational State Overlap
RRMO ²⁴	Yes	Yes	No	No
1DHR ²⁵ / Nx1DHR ^{5,26,27}	Yes	Same as Base Method	Uncoupled Only	No
2DHR ⁶	Yes	Same as Base Method	Uncoupled and Singly-coupled Only	No
FR / QP Interpolation ^{14,28–40} (<i>e.g.</i> McClurg ^{31,33})	Only at Low Temperature	Same as Base Method	Yes	No
Path Integral MC ^{41–43}	Yes	Yes	Yes	No
MCPSI ^{44,45}	Only with PG	Yes	Yes	Same as PG Reference
CIMCI	Only with PG	Yes	Yes	Same as PG Reference

Quantum Methods

Several quantum mechanics-based anharmonic nuclear Hamiltonians have been developed, and they all have different benefits and drawbacks (summarized in 1). While some very-accurate models have been developed for specific species,⁴⁶ only generic methods will be mentioned, *i.e.* methods that can be applied to all stable chemical species without needing significant changes on a case-by-case basis.

Vibrational perturbation theory (VPT2)¹⁷ was originally developed to deal with anharmonicities in asymmetric, linear molecules. It works well for treating anharmonicities caused by asymmetry, *e.g.* those from bond stretches and small-amplitude bends, but it can encounter problems with treating other kinds of anharmonicities, such as motions within bounded and/or periodic coordinate systems, *e.g.* torsions and large-amplitude bends. VPT2 also has other limitations, such as needing third- and fourth-order derivatives of the PES that make accurate treatment of PESs without analytical second derivatives very difficult, as well as having issues with Fermi resonances, which are surprisingly common, *e.g.* in CO₂.

Multiconformer methods^{21,22} and multistructural methods²³ combine results from multiple optimized minima. These are usually rigid rotor harmonic oscillator (RRHO) results, but they can be something more involved, *e.g.* in Zheng *et al.*²³ Their accuracy depends on how complete the conformer handling is, as well as on the accuracy of the model(s) used for each individual result. With larger molecules, computational cost can quickly explode due to the number of optimizations that need to be made, as each newly added dimension tends to exponentially increase the amount of available conformers at a given temperature.

Full SE solution methods attempt to capture the complete SE for nuclear motion, primarily through variational techniques. Vibrational self-consistent field (VSCF)¹⁸ is a well-known full SE method itself roughly equivalent to Hartree-Fock (HF) for nuclear wavefunctions, and it has numerous variants that are roughly equivalent to various post-HF methods for nuclear wavefunctions, including vibrational configurational interaction (VCI),¹⁸ Vibrational Møller-Plesset perturbation theory (VMP),¹⁹ and vibrational coupled cluster (VCC).²⁰ In theory, these methods can treat all kinds of anharmonicity. In practice, their accuracy varies not only based on the particular variant being used, but also based on how appropriate the selected PES and basis set are to the system being studied. Historically, PESs based on a local expansion around a minimum and basis sets based on HO normal modes have been the most popular choices. This limits the types of nuclear motion that can be well-represented by these calculations, and is *e.g.* the impetus behind recent work by Klinting *et al.*⁴⁷ and by

Carrington *et al.*⁴⁸ Other full SE solution software, *e.g.* TNUM and ElVibRot,^{49,50} numerically evaluate kinetic energy operators and compute energy levels with Lanczos solvers. More recently, analytical expressions of kinetic energy operators for these methods were shown to also be possible.⁵¹ Von Horsten employs TNUM in his wave propagation code MrPropa¹² and provides a reduced-dimensional approach⁵⁰ as a possible means of dealing with the “curse of dimensionality” while still working within the full SE solution space. Path integral methods^{41–43} are another full SE method that treats the motion of nuclei in the Feynman path integral formalism. They are very accurate, but also enormously time-consuming because they require an extremely high-dimensional numerical integration. As a result, they have only been demonstrated on tri- and tetraatomic systems so far.

Inherently non-polynomial methods are a class of methods that replace HO in the RRHO model with a different, non-polynomial model, often one that is based on an analytical solution to the SE. Their accuracy depends on how appropriate the new model is to a given system, and their speed primarily depends on how many single-point energy (SPE) calculations are necessary to construct their new model. Rigid rotor Morse oscillator (RRMO)²⁴ replaces HO with a Morse oscillator. This makes it most suitable for anharmonicities from asymmetry, like VPT2, but unlike VPT2, it only requires third-order gradients of the PES. One-dimensional hindered rotor (1DHR)²⁵ and its extension superposition of one-dimensional hindered rotations ($N \times 1\text{DHR}$)^{5,26,27} replace HOs with hindered rotors (HRs), but only for torsional coordinates. While they are based on the quantum pendulum model, which has an analytical SE solution, in practice, the HRs are constructed by numerically fitting a general Fourier series to a series of relaxed or unrelaxed PES scans along torsional coordinates. They are thus best suited to treating anharmonicities caused by Cartesian normal modes being unsuitable to model torsional motions. However, because of their reliance on a series of uncoupled PES scans, they can suffer from being unable to capture coupled torsional motions. Examples of this misbehavior appear in hydrogen-bonded clusters^{52,53} and in the transition state (TS) for hydrogen abstraction from oxygen-containing fuels by HO₂.^{13,54–56}

Two-dimensional hindered rotor (2DHR)⁶ is a natural extension of 1DHR that is able to capture coupling across any two different torsional motions. This comes at the cost of needing exponentially more SPE calculations in PES scans, and it is still not able to capture coupling across more than two torsional motions. However, by far the most popular non-polynomial methods, possibly even the most popular methods for modeling anharmonicity, are HO-free rotor (FR) interpolations.^{14,28–40} These are based on the idea at low temperatures, HRs behave more like HOs, while at high temperatures, they behave more like FRs, so a decent way of treating HRs is to use some interpolation between FR and HO quantities^{14,28–39} or VPT2 quantities.⁴⁰ HRs with a simple cosine potential, also known as quantum pendulum (QP), can also be used as part of the interpolation, something incorporated by McClurg *et al.*^{31,33} to great effect. Interpolation-based methods tend to be significantly faster than all other methods listed so far, especially the ones that only employ HO frequencies and/or the geometry from a single minimum to calculate their interpolation factor(s), usually by way of estimating barrier heights for torsional motions. While the accuracy of these methods often leaves much to be desired compared to the previously outlined methods for treating torsional anharmonicities (and of course, they are not meant for treating non-torsional anharmonicities), they are still used regularly due to the speed and ease at which they can be run in a blind-box fashion.

Semi-Classical Methods

A potentially attractive alternative to resolving anharmonic nuclear motions within quantum mechanics is to instead resolve them within classical mechanics and then apply a corrective approximation to recapture quantum-specific behavior that is lost in the classical picture. The classical partition function of a system is⁵⁷

$$Q_{\text{class}} = \frac{1}{h^{3N}\sigma} \int \cdots \int \exp\left(-\frac{H(\mathbf{p}, \mathbf{X})}{k_{\text{B}}T}\right) d\mathbf{p} d\mathbf{X} \quad (1)$$

Where h is Planck's constant, N is the number of nuclei in the system, σ is the system's overall symmetry number, k_B is Boltzmann's constant, T is temperature, H is the overall Hamiltonian, and \mathbf{p} and \mathbf{X} are respectively the momenta and Cartesian coordinate vectors of the system. Within the classical framework, the overall Hamiltonian can be split into a kinetic Hamiltonian that only depends on momenta and a potential Hamiltonian that only depends on positions. This allows the overall integral itself to be split into two separate integrals, one over momenta and one over positions:

$$Q_{\text{class}} = \frac{1}{h^{3N}\sigma} \int \cdots \int \exp\left(-\frac{H_{\text{kin}}(\mathbf{p})}{k_B T}\right) d\mathbf{p} \times \int \cdots \int \exp\left(-\frac{H_{\text{pot}}(\mathbf{X})}{k_B T}\right) d\mathbf{X} \quad (2)$$

The major benefit of the semi-classical approach is that both multidimensional integrals in eq. 2 can be handled and solved separately. In the case of the total partition function, *i.e.* where all kinetic contributions are included in the partition function, the kinetic integral has an easy-to-calculate analytical solution:

$$H_{\text{kin}}(\mathbf{p}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} \quad (3)$$

$$\int \cdots \int \exp\left(-\frac{H_{\text{kin}}(\mathbf{p})}{k_B T}\right) d\mathbf{p} = \prod_{i=1}^N \iiint_{-\infty}^{+\infty} \exp\left(-\frac{\mathbf{p}_i^2}{2m_i k_B T}\right) d\mathbf{p}_i \quad (4)$$

$$= (2\pi k_B T)^{1.5N} \prod_{i=1}^N m_i^{1.5N} \quad (5)$$

Substituting eq. 5 into eq. 2, it can be seen that for the total classical partition function, only a single term remains that is not trivial to solve:

$$Q_{\text{class}} = P \cdot C_{\text{I}} \quad (6)$$

$$P := \frac{1}{\sigma} \left(\frac{2\pi k_{\text{B}} T}{h^2} \right)^{1.5N} \prod_{i=1}^N m_i^{1.5} \quad (7)$$

$$C_{\text{I}} := \int \cdots \int \exp \left(-\frac{H_{\text{pot}}(\mathbf{X})}{k_{\text{B}} T} \right) d\mathbf{X} \quad (8)$$

C_{I} is the system’s configuration integral, and finding it is the primary objective of semi-classical methods of modeling anharmonic nuclear motion.

Finding the total partition function is not strictly necessary however. If the scheme seeks only to correct some elements of RRHO, the job of calculating the configuration integral can become much easier. Early semi-classical methods^{14,28,29} focused on only treating torsional motions and kept RRHO in place for other types of motion. While this approach does mean that a more complicated kinetic term needs to be used instead of eq. 3, they, too, can still be found analytically for free internal rotors.⁵⁸ Formulæ for HR potentials’ kinetic and potential components are also available, such as those used by Pitzer *et al.* to construct tables of thermodynamic terms for HRs.¹⁴ More recent developments that intend to capture said tables’ values³⁰ and/or use them more as benchmarks^{31,32,34} are also available, though these were primarily designed as part of quantum methods that interpolate between HO and FR values.

A more recently documented approach is to still calculate the total partition function, but to obtain the configuration integral from a numerical integration over positional space instead of trying to use an analytical solution. The key issue with determining this numerical integral is its high dimensionality. Since a non-monatomic, non-diatomic system with N nuclei will require a numerical integration over $3N - 6$ dimensions (the system has $3N$ positional coordinates, but 6 of them can be integrated analytically, as explained later), the number of dimensions over which integration is to occur can be exceedingly large even

for smaller molecules. Grid-based numerical integration methods, *e.g.* rectangular or trapezoidal integration, cannot reasonably deal with this high dimensionality. With N points used for a d -dimensional integration, these methods' worst case scenario uncertainties scale as $\mathcal{O}\left(N^{-\frac{2}{d}}\right)$ (for Simpson's Rule, $\mathcal{O}\left(N^{-\frac{4}{d}}\right)$). By contrast, MC integration techniques' uncertainties scale significantly better for higher dimensional integrals, with the most primitive technique scaling with $\mathcal{O}\left(N^{-\frac{1}{2}}\right)$. They are thus an ideal choice for numerical integration of the configuration integral. However, on its own, this improved scaling does not eliminate the need for a large number of sampled points in MC integration. This is because absolute uncertainties will still start at a higher amount purely due to the volume of the to-be-integrated space increasing faster than that of the relevant parts of the integrand. High-dimensional integrands that do not fill up their integration volume completely, such as those of configuration integrals, still require an overwhelming number of samples to obtain any kind of reasonable integral estimate. Thankfully, this issue can be addressed with the help of various alternate sampling strategies developed over the past few decades. These include: 1) using specially-designed numerical sequences instead of pseudorandom numbers during sampling, such as Sobol sequences;⁵⁹ 2) using importance sampling to concentrate samples in regions that contribute the most to the integral, such as the VEGAS^{60,61} and Suave⁶² methods, and; 3) using stratified sampling to allocate the overall sampling budget to regions with higher statistical uncertainty, such as the MISER method.¹⁶

The first detailed description of using MC integration to numerically obtain configuration integrals was written by Kamarchik, Jasper, *et al.*^{44,45} While their method, Monte Carlo Phase Space Integration (MCPSI), formally integrates state counts at fixed energies instead of configuration integrals directly, it is still a good demonstration of the potentials and pitfalls of MC integration for this application. The method was demonstrated to be able to treat all kinds of anharmonicities in all kinds of systems, but a very large number of samples (numbering well into the millions and billions) is still expected to be required to achieve adequate precision for anything but the smallest systems. To remedy this, MCPSI

employs some techniques to reduce statistical uncertainty and/or the total amount of time spent on a calculation. To ensure that such a large amount of points can be computed in a timely manner, SPEs needed by MCPSI are calculated with a PES fitted to many thousand coupled-cluster singles, doubles, and perturbative triples (CCSD(T)) SPEs, which was shown to give reasonable results at the cost of some setup time and risks involved with fitting and extrapolation.⁴⁴ Stratified sampling is used in the form of a fixed number of nested hyper-rectangles tuned to not split integration regions in inefficient places,⁴⁵ though this approach is still not as automatic or as tuned to the number of points in the calculation compared to generic MC integration methods like MISER. A sanity filtering system helps save time and weed out gross errors in the fitted PESs by generating a test energy for each sample and setting the integral contributions of those samples with extreme energies to 0.⁴⁵ Sampling was performed in curvilinear/internal coordinates instead of Cartesian normal mode displacement coordinates, which was shown to reduce the amount of points necessary for a given level of precision up to 10-fold;⁴⁴ however, MCPSI requires that the necessary Jacobians be computed numerically within the already-numerical estimation of a configuration integral.

So far, the methods discussed obtain only classical results. This is adequate for thermodynamic quantities at high temperatures and when the difference between zero-point energies (ZPEs) and PESs is not too large. However, for temperature ranges where classical mechanical results are no longer close to quantum mechanical ones, or when ZPEs are large enough to be important at all temperatures, additional treatment is necessary. Often, this will come in the form of some type of approximation to transform classical mechanical partition functions Q_{class} into approximations of quantum mechanical ones Q_{QM} . The most well-known of these approximations is the PG approximation,²⁸ which requires only classical and quantum mechanical results from the same method, $Q_{\text{class,ref}}$ and $Q_{\text{QM,ref}}$, to act as a reference:

$$\frac{Q_{\text{QM}}}{Q_{\text{class}}} \approx \frac{Q_{\text{QM,ref}}}{Q_{\text{class,ref}}} \quad (9)$$

Most commonly, RRHO results are what are used because they are so readily available and give adequate results at most temperature ranges of interest.^{28,31,41,44} Therefore, for a system with N frequencies ν :

$$\begin{aligned}
Q_{\text{QM}} &\approx Q_{\text{class}} \frac{Q_{\text{QM,HO}}}{Q_{\text{class,HO}}} \\
&= Q_{\text{class}} \exp\left(-\frac{\Delta E_{\text{ZPE}}}{k_{\text{B}}T}\right) \left(\frac{k_{\text{B}}T}{2h}\right)^N \prod_{i=1}^N \frac{\text{csch}\left(\frac{h\nu_i}{2k_{\text{B}}T}\right)}{\nu_i}
\end{aligned} \tag{10}$$

However, anharmonic effects that cannot be represented in classical results, *e.g.* the shifting of zero-point energies as a result of the shape of the PES, are not accounted for if RRHO results are used. The approximation itself also relies on the assumption that the relationship between the reference method's quantum and classical results is similar to that of between the given Q_{class} and its quantum equivalent. Situations where this does not hold true can produce gross inaccuracies in the approximation.⁶³

Methods

CIMCI is a semi-classical method for obtaining anharmonic partition functions developed in this study. It is similar in its overall approach to MCPSI, but aims to address some of its shortcomings. It also is set up to require less manual work in its setup in order to be usable as a black-box method, and it uses a framework that allows generic strategies for improving MC integration to be incorporated more easily.

CIMCI's core is the numerical estimate of a system's configuration integral as expressed in eq. 5 by way of MC integration across all positional coordinates of the system's nuclei. To do this, eq. 8 can be expanded as follows:

$$C_I = \int \cdots \int \exp \left(-\frac{U(\mathbf{q}) - U(\mathbf{q}_0)}{k_B T} \right) \|\mathbf{J}_{\mathbf{q}}\| d\mathbf{q} \quad (11)$$

Where $U(\mathbf{q})$ is the value of the PES at some nuclear coordinates \mathbf{q} , and $U(\mathbf{q}_0)$ is the value of the PES at its global minimum geometry \mathbf{q}_0 . The new term $\|\mathbf{J}_{\mathbf{q}}\|$ is the Jacobian of the coordinate system used to express \mathbf{q} , and amounts to 1 when \mathbf{q} is expressed in Cartesian coordinates. However, CIMCI by default performs its integration in curvilinear or internal coordinates, so $\|\mathbf{J}_{\mathbf{q}}\| \neq 1$. Instead, in the case of curvilinear coordinates, the Jacobian is as follows:⁶⁴

$$\|\mathbf{J}_{\mathbf{q}}\| = \sin(\beta) \prod_{i=1}^{N-1} r_i^2 \prod_{j=2}^{N-1} \sin(\phi_j) \quad (12)$$

Where β is the second Eulerian angle for overall rotation of the entire frame, r_i is the radial coordinate for the i th nucleus, and ϕ_j is the angular coordinate of the j th nucleus. Substituting eq. 12 into eq. 11, one obtains the following for the configuration integral in

curvilinear coordinates:

$$C_I = \int_{R_z=0}^Z \int_{R_y=0}^Y \int_{R_x=0}^X \int_{\gamma=0}^{2\pi} \int_{\beta=0}^{\pi} \int_{\alpha=0}^{2\pi} \int_{(\mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\theta})} \exp \left(-\frac{U(\mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\theta}) - U(\mathbf{r}_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{k_B T} \right) \sin(\beta) \prod_{i=1}^{N-1} r_i^2 \prod_{j=2}^{N-1} \sin(\phi_j) d\mathbf{r} d\boldsymbol{\phi} d\boldsymbol{\theta} d\alpha d\beta d\gamma dR_x dR_y dR_z \quad (13)$$

Where X , Y , and Z constitute the total volume in which the frame is allowed to move, R_x , R_y , and R_z are Cartesian displacements for the entire frame, and α , β , and γ are Eulerian angles for overall rotation of the entire frame. R_x , R_y , and R_z are traditionally anchored to the frame's center-of-mass, but any anchor is sufficient for CIMCI. The integrations over R_x , R_y , R_z , α , β , and γ can be solved analytically and do not need to be part of the MC integration. Doing so for a nonlinear system with 3 or more nuclei leads to:

$$C_I = V_{\text{trans}} 8\pi^2 \int_{(\mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\theta})} \exp \left(-\frac{U(\mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\theta}) - U(\mathbf{r}_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{k_B T} \right) \prod_{i=1}^{N-1} r_i^2 \prod_{j=2}^{N-1} \sin(\phi_j) d\mathbf{r} d\boldsymbol{\phi} d\boldsymbol{\theta} \quad (14)$$

Where V_{trans} is the volume in which the system is allowed to move. For diatomic systems, eq. 14 is almost the same, but the $8\pi^2$ term becomes 4π instead, as there is no integration over γ . In turn, V_{trans} can be obtained for an ideal gas from the system's temperature T and pressure p as:

$$V_{\text{trans}} = \frac{k_B T}{p} \quad (15)$$

This leaves an integration over $3N - 6$ curvilinear coordinates ($3N - 5$ for diatomics) that needs to be solved numerically with MC integration:

$$Q_{\text{class}} = \frac{8\pi^2 V_{\text{trans}}}{\sigma} \left(\frac{2\pi k_B T}{h^2} \right)^{1.5N} \prod_{i=1}^N m_i^{1.5N} \cdot C_{\text{MC}} \quad (16)$$

$$C_{\text{MC}} := \int_{(\mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\theta})} \exp \left(-\frac{U(\mathbf{r}, \boldsymbol{\phi}, \boldsymbol{\theta}) - U(\mathbf{r}_0, \boldsymbol{\phi}_0, \boldsymbol{\theta}_0)}{k_B T} \right) \prod_{i=1}^{N-1} r_i^2 \prod_{j=2}^{N-1} \sin(\phi_j) d\mathbf{r} d\boldsymbol{\phi} d\boldsymbol{\theta} \quad (17)$$

Eqs. 14, 16, and 17 hold true for non-diatomic, linear input geometries, as positions where

the system is not completely linear still need to be incorporated into the total partition function. They could, however, potentially present an issue in calculating eq. 16 through eq. 17. Due to the Jacobian (eq. 12), the integrand value for C_{MC} for geometries close to the input geometry will be either 0 or extremely close to 0, and geometries close to given input geometries would tend to be the ones that contribute the most to the integral. Such potential issues have not been analyzed in full detail for this study.

However, obtaining a numerical estimate of Q_{class} is not enough for most thermodynamic quantities, as many of them depend on $\frac{\partial Q_{\text{class}}}{\partial T}$ and/or $\frac{\partial^2 Q_{\text{class}}}{\partial T^2}$. This means that $\frac{\partial C_{\text{MC}}}{\partial T}$ and $\frac{\partial^2 C_{\text{MC}}}{\partial T^2}$ must also be found in addition to C_{MC} . The derivative is standard calculus and the result given in the Supporting Information (SI) along with the textbook formulas for thermodynamic functions used by us.¹³

Ordinarily, a separate MC integration would need to be performed for each of these three quantities, and this would need to be repeated at each temperature of interest. This would consume a lot of extra time and computational resources. However, the computational cost of a MC integration of any of these functions lies disproportionately in the SPE calculations needed by all of these functions. During testing, at least 99.7% of the computational time of a CIMCI run is spent on obtaining SPEs, and that was using the relatively fast UFF⁶⁵ potential for SPEs. However, all integrals that need to occur happen (or can happen) over the same space and on the same system. The only difference between them is how their integrands process SPE. In order to save on time, CIMCI evaluates every integral, *i.e.* eqs. 17, S-12, and S-13 for every temperature, at the same time. The SPE of a single, randomized geometry is passed onto all three integrands (more if more than one temperature is being looked at), whose results are then calculated, stored, and processed individually.

To calculate the aforementioned integrals, CIMCI has been built to work with any standard MC integration technique that can be adjusted to compute and process multiple integrals across the same set of random samples. Two such techniques have been tested so far: “naïve” MC integration, which uses no special, alternate sampling technique for reducing sta-

tistical variance, and a slightly modified version of MISER,¹⁶ which uses a sampling technique that revolves around recursive stratification of the integration space. Use of quasi-random number sequences, *e.g.* Sobol sequences,⁵⁹ was not tested because it was not guaranteed that they would reduce statistical variance for the integrals of interest for CIMCI. Importance sampling techniques, *e.g.* VEGAS^{60,61} and Suave,⁶² were not tested because it was difficult to determine how they should be modified to evaluate multiple integrals simultaneously without sacrificing the quality of any one integral, which is vital for CIMCI. “Naïve” MC integration obtains a numerical estimate for a definite integral $F(\mathbf{x})$ over some ranges $\boldsymbol{\omega}$ with a total volume of V using N uniformly randomly sampled points:

$$\begin{aligned}
F(\mathbf{x}) &= \int_{\boldsymbol{\omega}} f(\mathbf{x}) \, d\mathbf{x} \\
&\cong \frac{V}{N} \sum_{n=1}^N f(\mathbf{x}_n) \\
&= V \langle f(\mathbf{x}) \rangle
\end{aligned} \tag{18}$$

$$\begin{aligned}
\delta F(\mathbf{x}) &= \frac{V}{\sqrt{N}} \sqrt{\frac{\sum_{n=1}^N (f(\mathbf{x}_n) - \langle f(\mathbf{x}) \rangle)^2}{N-1}} \\
&= \frac{V \sigma_N}{\sqrt{N}}
\end{aligned} \tag{19}$$

The number of samples used N can be increased to reduce the uncertainty in the final estimate $\delta F(\mathbf{x})$, but the reduction in uncertainty will be rather slow if this is the only strategy used. MISER¹⁶ increases the reduction in uncertainty with the help of recursive stratified sampling. With a single stratification, a single, large, definite integral over $\boldsymbol{\omega}$ is split up into two smaller definite integrals over $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$, resulting in the following numerical

estimate:

$$\begin{aligned}
F(\mathbf{x}) &= \int_{\omega_A} f(\mathbf{x}) \, d\mathbf{x} + \int_{\omega_B} f(\mathbf{x}) \, d\mathbf{x} \\
&\cong \frac{V_A}{N_A} \sum_{n=1}^{N_A} f(\mathbf{x}_n) + \frac{V_B}{N_B} \sum_{n=1}^{N_B} f(\mathbf{x}_n) \\
&= V_A \langle f_A(\mathbf{x}) \rangle + V_B \langle f_B(\mathbf{x}) \rangle
\end{aligned} \tag{20}$$

$$\begin{aligned}
\delta F(\mathbf{x}) &= \sqrt{\delta F_A(\mathbf{x})^2 + \delta F_B(\mathbf{x})^2} \\
&= \sqrt{\frac{V_A^2 \sigma_{N_A}^2}{N_A} + \frac{V_B^2 \sigma_{N_B}^2}{N_B}}
\end{aligned} \tag{21}$$

The uncertainty in eq. 21 will never be greater than that in eq. 19, and, in the case of $V_A = V_B$, it is minimized when:¹⁶

$$\frac{N_A}{N_A + N_B} = \frac{\sigma_{N_A}}{\sigma_{N_A} + \sigma_{N_B}} \tag{22}$$

MISER applies this splitting recursively and attempts to guess each σ_{N_A} and σ_{N_B} by allocating some of the sampling budget to pre-sampling the integration region at each recursion. It also uses this same pre-sampling guesses for σ_{N_A} and σ_{N_B} to decide along which dimension it should split the integration region into two smaller regions (MISER always only splits the integration region along a single dimension at a time). It does so by looking at which split would minimize the following heuristic:

$$\sigma_{N_{\text{Guess}}}^2 \propto \left({}^{1+\alpha}\sqrt{\sigma_{N_A}^2} + {}^{1+\alpha}\sqrt{\sigma_{N_B}^2} \right)^{1+\alpha} \tag{23}$$

Where α is an arbitrary parameter that is said to produce a good heuristic when $\alpha = 2$.¹⁶ However, because CIMCI evaluates multiple integrals simultaneously while MISER only analyzes a single set of uncertainties at each recursion, MISER needs to be slightly modified in order to work with CIMCI. At each recursion, the pre-sampling step proceeds as normal,

albeit with pre-sampling results collected for all integrals instead of just one. For the heuristic step though, the modified version of MISER first determines a slightly different heuristic based on eqs. 23 and 20. For each integral, the following heuristic is calculated for each dimension of the integration region:

$$\sigma_{N_{\text{MultiGuess}}}^2 \propto \frac{\left({}^{1+\alpha}\sqrt{\sigma_{N_A}^2} + {}^{1+\alpha}\sqrt{\sigma_{N_B}^2} \right)^{1+\alpha}}{V_A \langle f_A(\mathbf{x}) \rangle + V_B \langle f_B(\mathbf{x}) \rangle} \quad (24)$$

Different integrands can have vastly different magnitudes within a given integration region, and therefore different absolute magnitudes of uncertainty. This step compensates for these differences in absolute magnitude and makes the heuristic work off of relative uncertainties instead of absolute ones. However, there is still more that needs to be done to make MISER work for multiple simultaneous integrations. Each integral will still have a different shape within the integration space. Simply assigning the average of eq. 24 across all integrands to each dimension and then selecting to split the dimension with the lowest average can result in splittings that are good for most integrands, but horrible for a select few. Instead, the modified version of MISER used by CIMCI assigns the highest value of eq. 24 across all integrands, and then it will choose to split the dimension that has the lowest heuristic assigned to it, like with the unmodified version. In effect, MISER always chooses to split the dimension for which the worst heuristic value is the least worst out of all worst values. This makes sure that no MISER bisection will be outright horrible for any of the integrations being performed. While it is not necessarily the best compromise solution under all circumstances, it is a safe one if all integrals to be evaluated are of equal value. It also maintains MISER's automatic, black-boxable nature, which is a major benefit for CIMCI's intended purpose compared to performing a stratified sampling strategy by hand (recursive or otherwise).

Integrating over curvilinear coordinates introduces an additional challenge in the way random geometries are generated for MC integration. Blindly generating uniform numbers for each individual curvilinear coordinate will result in the randomly sampled geometries

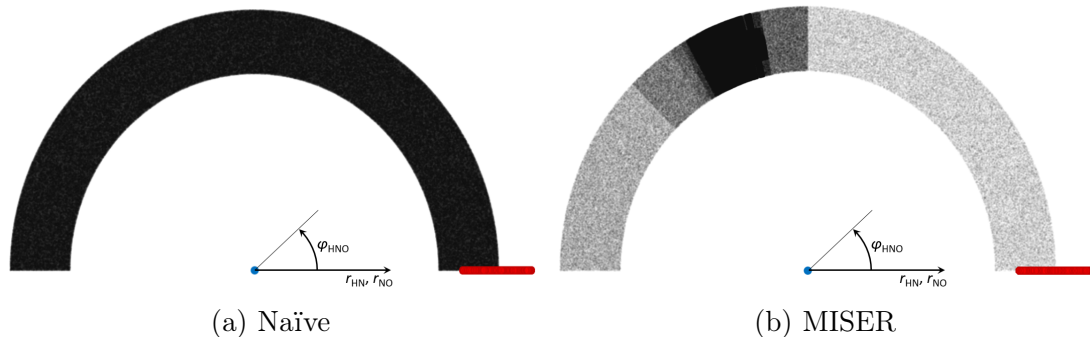


Figure 1: The overlaid atomic coordinates of all one million generated geometries taken during naïve (1a) and MISER (1b) MC integrations of HNO. H is in black, N is in blue (in the bottom center), and O is in red. For completeness, geometries under MISER that were used purely for pre-sampling and heuristics are also included.

biasing heavily towards ones with smaller r values and values of ϕ near the poles at 0 or π radians. Coordinates are therefore randomized with a biasing that is based on what curvilinear coordinate they are and to which nucleus are they assigned, as given in Table 2.

Table 2: Conversion of integration ranges to correct point distributions generated for curvilinear coordinates of the i th atom to yield uniform distribution in Cartesian coordinates.

	$i = 1$	$i = 2$	$i > 2$
r_i	$[a; b]$	$b\sqrt{[\frac{a^2}{b^2}; 1]}$	$b\sqrt[3]{[\frac{a^3}{b^3}; 1]}$
ϕ_i	-	$[a; b]$	$\cos^{-1}([\cos a; \cos b])$
θ_i	-	-	$[a; b]$

The resulting geometry distributions have been checked for smaller molecules (up to 5 nuclei). For di- and triatomic systems, one can still easily check geometries by visual inspection, such as in fig. 1, which shows the distribution of one million points among the internal coordinate space of HNO. Relative to N nuclei (blue), positions for O nuclei are shown in red, and positions for H are shown in black. In both cases, color value represents concentration, *i.e.* the more geometries that were generated with a particular nucleus at a given coordinate, the brighter/darker that nucleus' color appears in the figure. One can see from both subfigures that neither sampling technique shows a bias towards smaller distances or towards poles of the angles. In addition, one can see from the abrupt changes in point distribution in plot 1b how the automatic stratification scheme in MISER leads to meaningful

cuts in the H–N–O bending angle space, as well as slightly less visible cuts in the H–N bond distance space. It is worth noting that while MISER is a stratified sampling algorithm and not an importance sampling one, the concentration of its generated geometries can end up focusing on regions of high importance anyway. Geometries are predominantly generated at a bending angle near 114° , which is very near the minimum of the PES for HNO and thus is the primary contributor to its partition function and its partition function temperature derivatives. Unfortunately, adequate benchmark thermochemistry data for HNO, whether experimental or high-level theoretical, could not be found owing to its instability, so the performance of CIMCI could not be appropriately discussed or judged.

The presampling within the MISER algorithm leads to a smaller overall variance compared to a naïve calculation using the same number of points. When calculating variances for eq. 19, there is a risk of loss of precision due to floating point arithmetic. The traditional method of calculating the variance of evaluations of some function $f(\mathbf{x})$ involves collecting a sum of the squares of the evaluations and subtracting the evaluations’ average value squared from that sum. This subtraction is well-known to potentially result in catastrophic cancellation.⁶⁶ To avoid this, CIMCI uses the variant of the Youngs-Cramer method for evaluating the massive sum in eq. 19 that is outlined in a review by Schubert and Gertz.⁶⁶ During test runs of CIMCI on the types of systems present in this study, the difference in σ_N , *i.e.* the square root of the variance, between the Youngs-Cramer method and the traditional method were usually around 7-15%: not major, but not insignificant.

Because of the blind way geometries are generated, CIMCI can sample geometries that are very far away from the equilibrium, potentially even being geometries of chemical isomers instead of those of the same species. It would be unnecessary and/or undesirable to fully evaluate these geometries for the integration, either because their contribution to the integral would be completely negligible or because they do not represent a configuration that the starting species could realistically adopt at the specified temperatures. To address this, CIMCI employs a “fast filtering” system that tests every sampled geometry before it

is passed onto the full method to be used by a given CIMCI calculation. If any two nuclei in a given geometry are closer than 0.3 Bohr (about 0.16 Angstrom), the geometry fails the test immediately. Otherwise, the geometry’s bond orders are resolved with the help of a fast method (Mayer bond orders⁶⁷ with GFN1-xTB⁶⁸ if the full method is classified as “expensive”, *i.e.* DFT or slower, Pauling bond orders⁶⁹ with UFF⁶⁵ otherwise) and compared with the bond orders of the input geometry. If any new bond is present in the sampled geometry where there is none in the input geometry or vice versa, the sampled geometry fails the test as well. Geometries are only passed on to the full method if they do not fail the test; if they do, their contribution to all integrals is set to 0 and CIMCI moves on to another geometry. In case the base method is the same as the method used for “fast filtering” (only possible if the base method is UFF⁶⁵), then the bond checks are still made, but CIMCI just takes the total energy of the system from the “fast filtering” result to calculate the geometry’s potential instead of requesting another calculation. While this “fast filtering” setup works well for the systems analyzed in this study, this study also acknowledges that it does have two notable pitfalls that would need to see improvement before it can be used in a black box fashion: very high temperatures and stereoisomers. At very high temperatures, interatomic distances that would be flagged as bond breakages can still contribute non-negligibly to a species’ thermodynamics, so excluding their contributions to the configuration integral and its temperature derivatives would be erroneous. As for stereoisomers, because the “fast filtering” system only checks for changes in bond orders and not for changes in stereochemistry, CIMCI can generate geometries that are stereoisomers of the inputted species and that will pass through the “fast filtering” system without issue. When a user is only expecting the configuration integral to be evaluated for the inputted species, this pollution by stereoisomers can heavily distort CIMCI’s results.

The code for this study was written exclusively in Fortran2008 and compiled with Intel Fortran Compiler 19.1.3 using Intel MPI Library 2019 Update 9 for parallelization. CIMCI was developed as a job type within the AMS software suite,⁷⁰ which allows the program to

use any SPE calculation method within the suite. It is made to work off of standard methods that are available in the AMS software package, specifically, force field (FF) methods, reactive FF methods like ReaxFF,⁷¹ machine learning potentials, and density functional-based tight-binding (DFTB)⁷² methods. Only parameterless methods were used for this study, namely UFF⁶⁵ and GFN1-xTB.^{68,73} DFT calculations with CIMCI were obtained with the M06-2X functional in ADF^{74–80} using ADF2020’s TZ2P basis set⁸¹ and standard, linear ZORA relativistic scaling.^{82–84} Random numbers for geometry randomization were generated with the random number generator (RNG) that is present in AMS; it is a xoshiro256**^{85,86} generator with specially set RNG seed values to help avoid bad initializations (“zeroland”⁸⁶). In order to avoid floating-point underflow in high-dimensional species due to the Jacobian being very small when evaluating the integral in SI units, the code also progressively multiplies in the kinetic pre-factor term (progressively because doing so all at once can cause an overflow due to the pre-factor being so big). Coordinate bounds for the integration procedure can be specified manually or left to be determined automatically; all calculations for this study used the latter. When coordinate bounds are determined automatically, bond angles and dihedrals are automatically detected and always set to their full ranges: 0 to π radians for angles, 0 to 2π radians for dihedrals. Automatic bond distance bounds are based on the input geometry and the value of the greatest temperature specified for the calculation, with the assumption that the input geometry is at the global minimum of the PES. They range from $\frac{(\text{input length})}{(f)}$ to $(\text{input length}) \times (f)$, where $(f) = 1 + 0.25 \times \frac{\sqrt{g(\mathbf{T})+g(\mathbf{T})^4}}{1+g(\mathbf{T})^2}$ and $g(\mathbf{T}) = \frac{\max(\mathbf{T})}{1500 \text{ K}}$. This value is based on observations of CIMCI integrand values with respect to (w.r.t.) bond length of weak F–F bonds obtained from running the program on F_2 at 298.15 K to 1500 K and making a conservative estimate for the temperature domains where available bond stretch potentials are still parabolic vs. where they cease to be parabolic. The range itself is asymmetric w.r.t. equilibrium bond lengths by design, and so the region encompassing bond lengthening is significantly larger than that encompassing bond contraction.

Results

In this section, CIMCI-derived thermodynamic quantities of small systems are first compared to those of experiment and high-level computations, as well as to quantities derived by VPT2, 1DHR/ $N\times$ 1DHR, and McClurg-style HR corrections where appropriate. The quantities obtained from these other methods were done so with the Gaussian16 Revision C.01 software package.⁸⁷ VPT2 results were obtained with the “freq=anharm” keyword. McClurg-style HR results were obtained with the “freq=hinderedrotor” keyword. 1DHR and $N\times$ 1DHR results were obtained by first performing relaxed PES scans for each torsional coordinate and then processing the results with the Tamkin Python library.⁸⁸ Afterwards, the scaling behavior and obtained numerical uncertainties of CIMCI will be discussed.

Anharmonic Contributions to Entropy

Water is maybe the most important chemical species known to mankind. Its thermochemistry is governed by anharmonic motion of many hydrogen-bonded water clusters, and so water clusters tend to be the popular subject of study,^{89–91} mainly their structure, energetics, and frequencies.^{92,93} Recently, however, rovibrational levels of the water monomer⁹⁴ and its isotopologues⁹⁵ in the ideal gas state have been measured and calculated. Of about one million bound rovibrational states, approximately 20000 have been determined experimentally. Furtenbacher *et al.*⁹⁴ give a complete list of rovibrational energy levels up to 9724 cm^{-1} in their W2020 database. Being a small system with such extensive reference data, gas-phase water serves as a good starting point for comparing the performance of CIMCI to other anharmonic treatments.

Anharmonic effects on the thermochemistry of water are minor compared to molecules with torsional degrees of freedom (DOFs), *e.g.* H_2O_2 , which will be discussed later. The vibrational energy level subset of the W2020 database provides an excellent experimental benchmark to which to compare results of CIMCI and other anharmonic treatments, as it

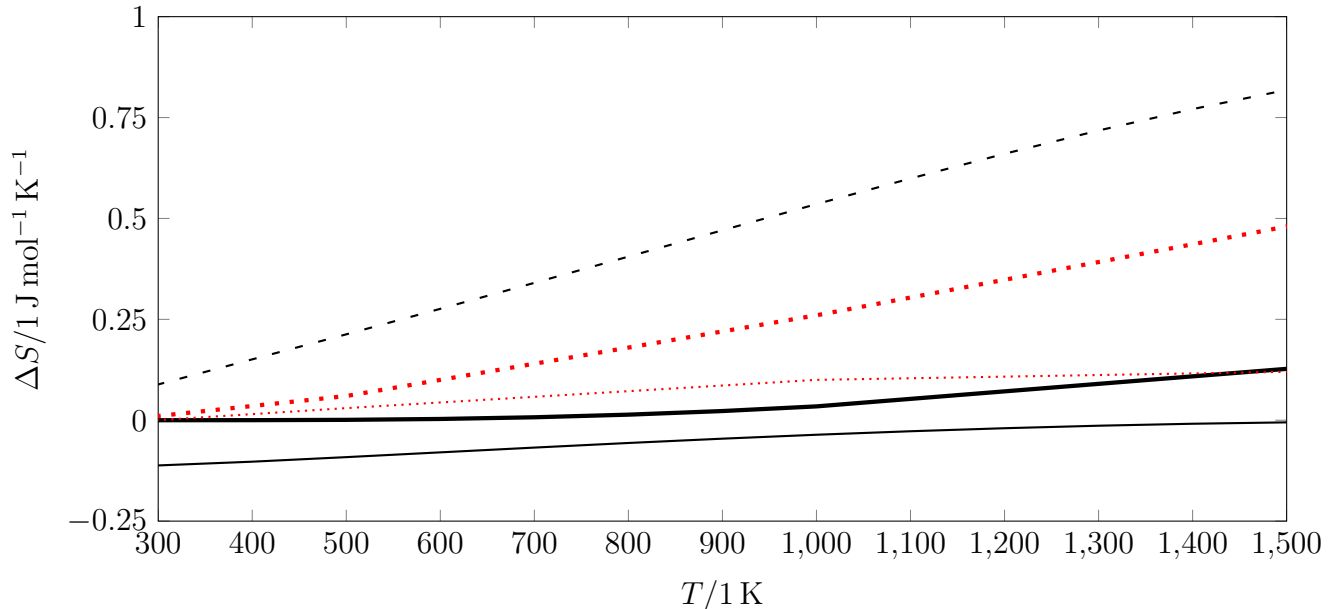


Figure 2: $S_{\text{anharm}} - S_{\text{HO}}$ of $\text{H}_2\text{O}(\text{g})$ at 101 325 Pa with CIMCI at UFF (—) and GFN1-xTB (- -), VPT2 at UFF (.....) and M06-2X/cc-pTVZ (- . - .), and the W2020 measurement compilation (—).⁹⁴ CIMCI results are averages of 5×10^6 sample runs.

is certain that its quantities are not merely the result of applying experimentally obtained force constants to a RRHO model. Fig. 2 compares the anharmonic corrections of CIMCI done on UFF, GFN1-xTB, and M06-2X/DZP potentials to those of VPT2 with UFF, VPT2 with M06-2X/cc-pVTZ, and the W2020 database. With UFF, CIMCI is able to reproduce a similar trend as the W2020 values and at the same magnitudes, but with a slight vertical shift of around $-0.1 \text{ J mol}^{-1} \text{ K}^{-1}$. By contrast, VPT2 with UFF does not have this vertical shift while also appearing to have the exact same trend at the exact same magnitudes as CIMCI. The anharmonic correction for CIMCI with GFN1-xTB not only has a shift of about $0.1 \text{ J mol}^{-1} \text{ K}^{-1}$, but also has a much steeper trend than either the W2020 data or any of the other two methods with UFF. VPT2 with M06-2X/cc-pVTZ, while having no vertical shift, also displays a much steeper trend than the W2020 data, though not as steep as CIMCI with GFN1-xTB. VPT2 with M06-2X/cc-pVTZ culminates in an entropy correction at 1500 K that is about three times greater than it should be, while CIMCI with GFN1-xTB gives a correction that is about six times greater than it should be. These shifts and different

trends between different potentials and/or between using VPT2 or CIMCI demonstrate how these methods are sensitive to the shape of the PES. VPT2 is sensitive to the shape of the PES at the minimum being looked at, whereas CIMCI is also sensitive to its shape in areas further away from the minimum. In the absence of significant enough anharmonicity, like in the case of H_2O , inaccuracies in the broad shape of the PES used can cause greater errors in CIMCI than in methods that only rely on the shape at or around stationary points, such as VPT2.

For larger molecules, anharmonicities can be quite substantive. One well-known example for this is hydrogen peroxide (H_2O_2), a species with two symmetric minima separated by hindered rotation around the $\text{H}-\text{O}-\text{O}-\text{H}$ torsion. H_2O_2 is important for study because it is used as bleaching agent and disinfectant and it plays a paramount role in atmospheric and combustion chemistry.⁹ Hybrid experimental data are reported by Chase in the NIST-JANAF Thermochemical Tables monograph⁹⁶ of 1998 and later by Dorofeeva.⁹⁷ These data combine spectroscopic force constants with a RRHO treatment and data from Pitzer and Gwinn’s tables on internal rotation.¹⁴ As a result, this source is not expected to supply benchmark-quality information on anharmonicity. In 2013, Malyszczek and Koput provided a PES for H_2O_2 from CCSD(T)-F12b/aug-cc-pV5Z energies.⁹ They solved the SE in 9D rovibrational coordinates with a variational approach and reported 173 energy levels up to 48 kJ mol^{-1} in their Supporting Information. With their reported force constants, 91 combinations of HO energy levels up to 48 kJ mol^{-1} , have been determined. It was found by this study that above 700 K, the cutoff at 48 kJ mol^{-1} leads to increasing entropy underestimation, from $0.3\text{ J mol}^{-1}\text{ K}^{-1}$ at 700 K to $2.3\text{ J mol}^{-1}\text{ K}^{-1}$ at 1000 K. This correlates with the kink in the Malyszczek curve in Figs. 3 and 4, which is why the line is switched to a dashed one from 700 K.

As shown in Fig. 3, at standard conditions, anharmonicity in $\text{H}_2\text{O}_2(\text{g})$ leads to an entropy increase of nearly $R\ln 2$ in Malyszczek and Koput’s results, coinciding with the two equivalent minima. However, this effect decreases with increasing temperature as hindered

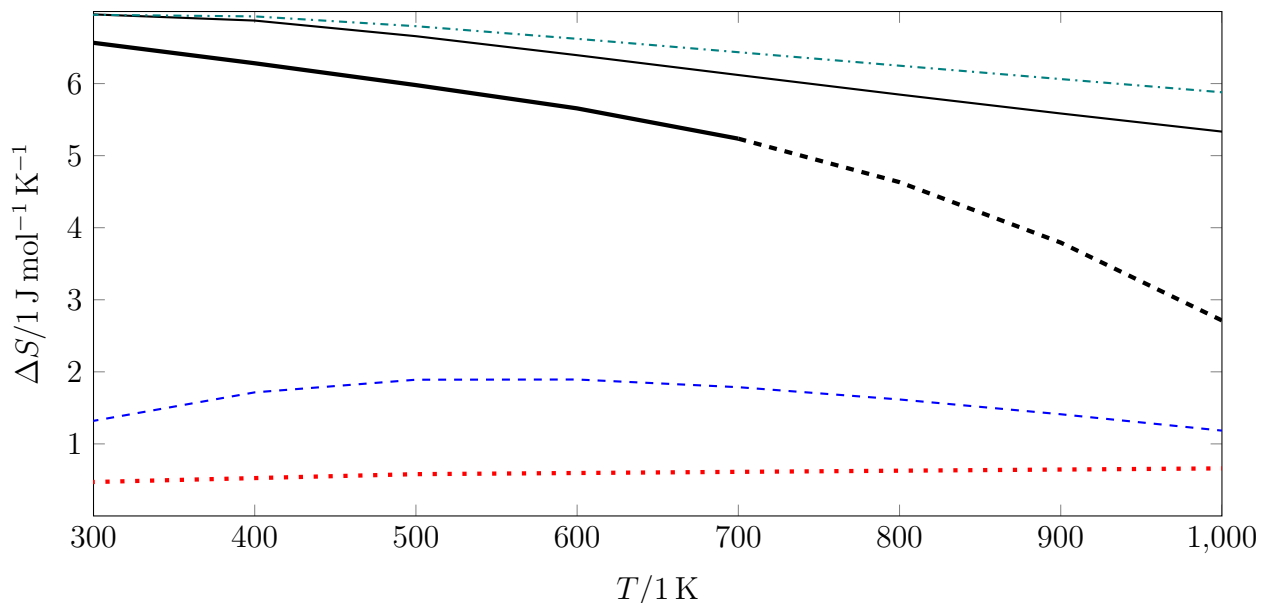


Figure 3: $S_{\text{anharm}} - S_{\text{HO}}$ of $\text{H}_2\text{O}_2(\text{g})$ at 101 325 Pa computed with anharmonicity schemes 1DHR with Tamkin⁸⁸ at UFF (---), McClurg HR^{31,33} at UFF (···), VPT2¹⁷ at UFF (···), CIMCI at UFF (—), and the calculations of Malyszek and Koput⁹ (—, ---). CIMCI results are averages of 5×10^7 sample runs.

rotation between the two minima becomes more possible. CIMCI on a UFF PES is able to replicate both the trend and the magnitude of this anharmonicity reasonably well. For Tamkin’s 1DHR results on a UFF PES, 4000 wavefunctions and energy levels were used to solve the full quantum-mechanical SE for one-dimensional H–O–O–H torsional rotation. At lower temperatures, they are almost identical to CIMCI’s results, though CIMCI’s are still slightly closer to Malyszek and Koput’s results. This changes at higher temperatures though: both CIMCI’s and Malyszek and Koput’s measured anharmonicities decrease more than Tamkin 1DHR’s. The most likely cause would be anharmonic effects other than just that of H–O–O–H hindered rotation, probably asymmetries within H–O–O bends. The McClurg HR model on a UFF PES severely underestimates anharmonicity and yields entropy corrections less than $2 \text{ J mol}^{-1} \text{ K}^{-1}$ at all considered temperatures. VPT2 on a UFF PES fares even worse, consistently yielding entropy corrections of less than $1 \text{ J mol}^{-1} \text{ K}^{-1}$.

Fig. 4 shows that if a higher-level PES like M06-2X/cc-pVTZ is used for anharmonicity calculations, VPT2 performs slightly better and correctly determines the overall order of

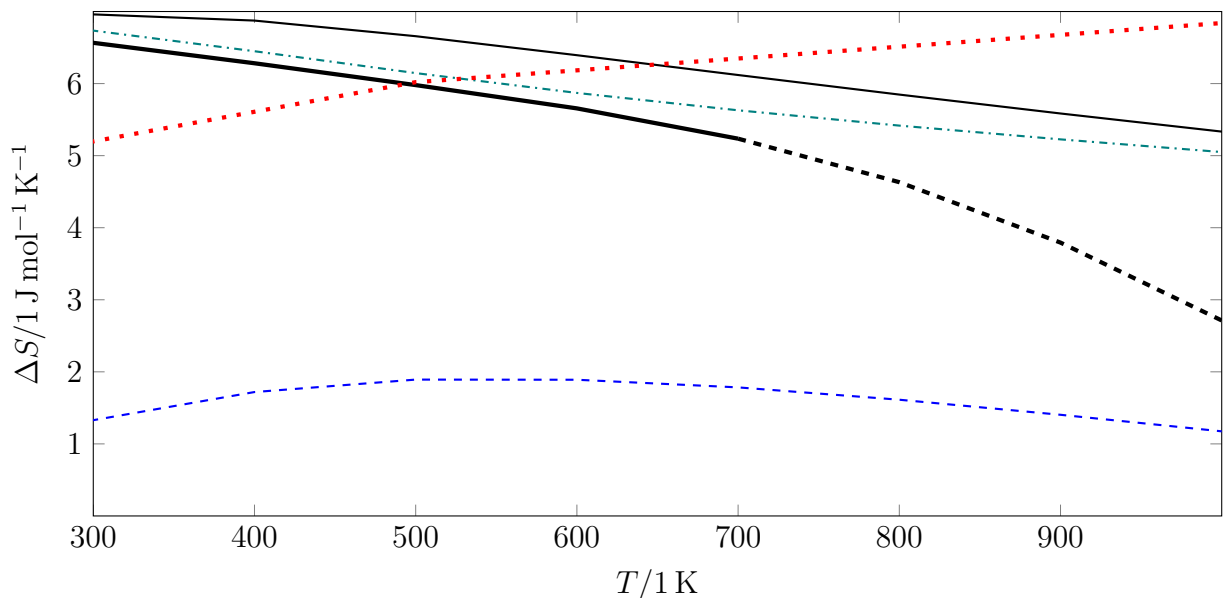


Figure 4: $S_{\text{anharm}} - S_{\text{HO}}$ of $\text{H}_2\text{O}_2(\text{g})$ at 101 325 Pa computed with anharmonicity schemes 1DHR with Tamkin⁸⁸ at M06-2X/cc-pVTZ (---), McClurg HR^{31,33} at M06-2X/cc-pVTZ (---), VPT2¹⁷ at M06-2X/cc-pVTZ (···), CIMCI at UFF (—), and the calculations of Malyszczek and Koput⁹ (—, ---). CIMCI results are averages of 5×10^7 sample runs.

magnitude of anharmonicity. However, it shows a completely opposite trend in anharmonicity at increasing temperature compared to the results of 1DHR at M06-2X/cc-pVTZ, CIMCI at UFF, and Malyszczek and Koput. The 1DHR results themselves are in very close agreement with Malyszczek and Koput’s results when a higher-level PES like M06-2X/cc-pVTZ is used, improving upon both 1DHR with UFF and CIMCI with UFF. This agreement only lasts until around up to 600 K, and the same, shallower decrease in anharmonic corrections is still observed throughout. The McClurg HR model’s results remain high unchanged at M06-2X/cc-pVTZ compared to UFF.

Conspicuously absent from the H_2O_2 studies compared to the H_2O ones are CIMCI results computed with some DFTB PES, *e.g.* GFN1-xTB. With every single DFTB method tested by this study, it was found that the global minimum for H_2O_2 was a completely incorrect, C_{2h} symmetry structure. It was assumed that with a global minimum geometry so dramatically incorrect, obtained thermochemical values would be grossly incorrect as well for any anharmonic treatment.

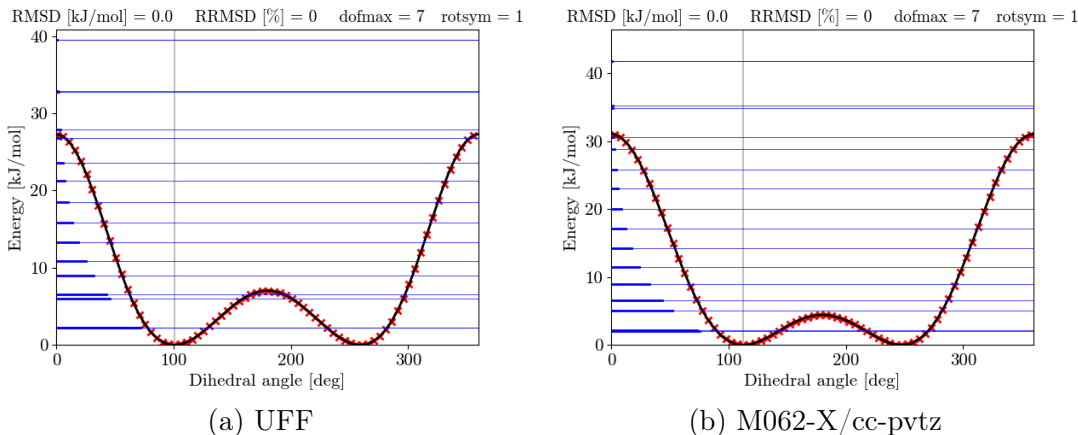


Figure 5: Potential energy scan for the dihedral angle of H_2O_2 at two levels of theory. Computed points are shown in red, the Fourier fit in black, and the blue horizontal lines are the energy levels from the one-dimensional SE. The bold parts of the horizontal lines represent the population of each energy level at standard conditions.

Potential energy curves along the H–O–O–H dihedral are shown in Fig. 5 for UFF (5a) and M06-2X/cc-pVTZ (5b). Both methods compute the rotational barrier to the same order of magnitude, and therefore produce vibrational level populations that are similar.

Table 3 shows that the lowest-frequency HO vibrational mode ($\nu = 1$), which is the one that corresponds the most to H–O–O–H torsion, has a similar frequency for both methods. By contrast, the bending modes ($\nu = 3, 4$ for UFF, $\nu = 2, 3$ for M06-2X/cc-pVTZ) show very different frequencies between the two methods. This is probably why VPT2 results differed so much between the two levels of theory. Nevertheless, anharmonicity in $\text{H}_2\text{O}_2(\text{g})$ still stems primarily from the H–O–O–H torsional mode, and so H–O–O bends should not be given a large weight when correcting for anharmonicity. CIMCI is able to correctly account for this even on a UFF PES, as its complete exploration of phase space takes all modes into account at proper magnitudes.

Table 3: HO vibrational frequencies for H₂O₂ at two levels of theory. All values are in cm⁻¹.

ν	UFF	M06-2X/cc-pVTZ
1	397.9	387.5
2	1417.6	1046.7
3	1789.6	1357.8
4	1914.2	1468.6
5	3732.1	3843.9
6	3750.8	3845.7

The trend of decreasing anharmonicity in molar entropy with increasing temperature is also found for other molecules that show more complex anharmonicity than the smaller ones treated so far in this study. For example, whereas H₂O₂’s anharmonicity comes from a single, uncoupled torsion connecting two equivalent minima, NH₂OH has two unequivalent minima that are connected through a torsion and an inversion that are coupled with each other.⁴⁵ Energy levels up to 20 000 cm⁻¹ have been experimentally determined for NH₂OH by Luckhaus, and he reports corroborating *ab-initio*-derived values as well.⁹⁸ From these energy levels, NH₂OH entropies including anharmonic effects have been calculated, along with entropies from the RRHO model based on the nine fundamental frequencies reported by Luckhaus. These calculated entropies are given in the Supporting Information. The resulting anharmonic entropy correction shows a decreasing trend with increasing temperature similar to the one shown for H₂O₂ in this study. The authors of this study therefore believe that NH₂OH and NH₂NH₂ (which has a pair of minima connected by a similarly coupled torsion and inversion⁴⁵) are interesting targets for future investigations with CIMCI.

Scaling

MC integration is known for its ability to scale better with sample counts at higher dimensions than traditional, rectangular or trapezoidal integration. In theory, the absolute uncertainty of naïve MC integration scales with $\mathcal{O}\left(N^{-\frac{1}{2}}\right)$, and the scaling of MISER should be no worse than naïve MC integration. In the case of CIMCI however, there is an extra complication. CIMCI obtains configuration integrals and their temperature derivatives, not

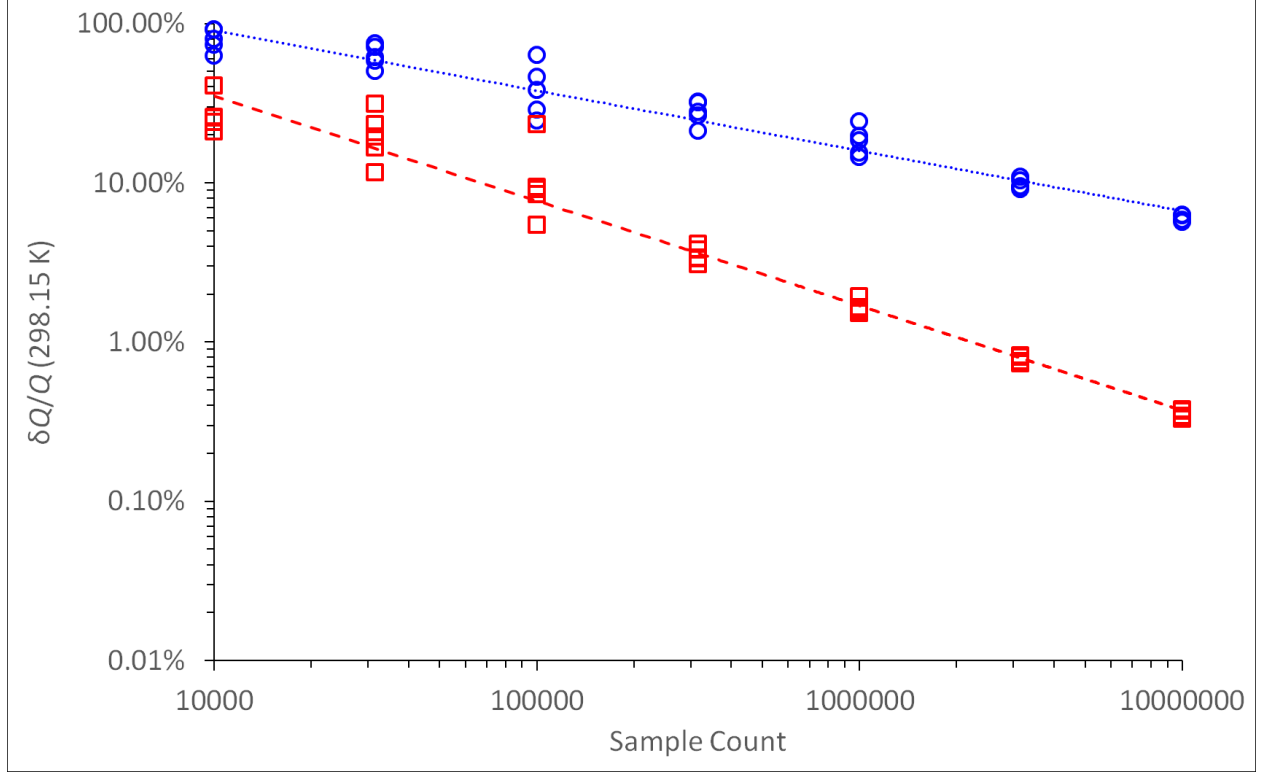


Figure 6: Relative uncertainties in PG-corrected $\text{H}_2\text{O}_2(\text{g})$ partition functions at standard conditions with CIMCI at UFF. Naïve MC integrations are blue circles, MISER MC integrations are red squares. The average relative statistical uncertainty scales for naïve integration (dotted) with sample counts x as $29x^{-0.38}$ ($R^2 = 0.95$) and for MISER (dashed) as $150x^{-0.66}$ ($R^2 = 0.97$). Each point represents the result of a separate, independent CIMCI run.

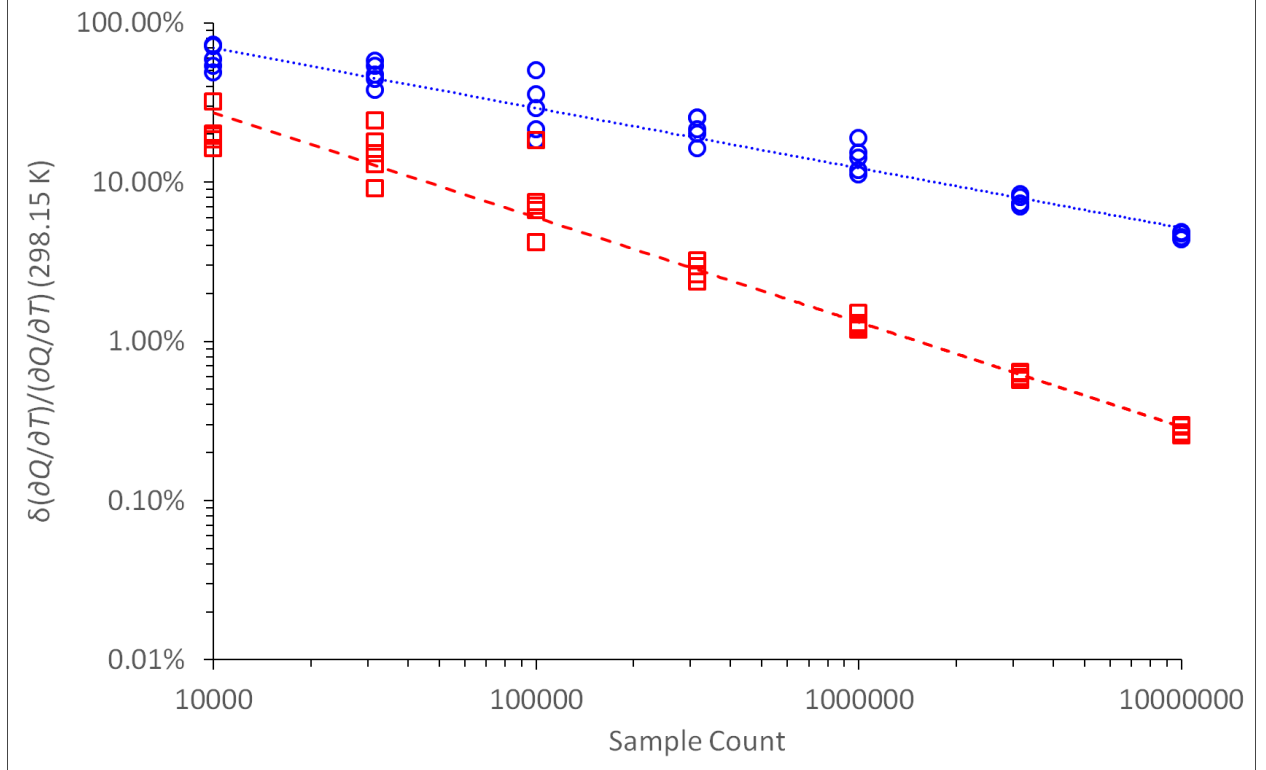


Figure 7: Relative uncertainties in PG-corrected $\text{H}_2\text{O}_2(\text{g})$ first derivatives of the partition function w.r.t. temperature at standard conditions with CIMCI at UFF. Naïve MC integrations are blue circles, MISER MC integrations are red squares. The average relative statistical uncertainty scales for naïve integration (dotted) with sample counts x as $22x^{-0.38}$ ($R^2 = 0.94$) and for MISER (dotted) as $116x^{-0.66}$ ($R^2 = 0.96$). Each point represents the result of a separate, independent CIMCI run.

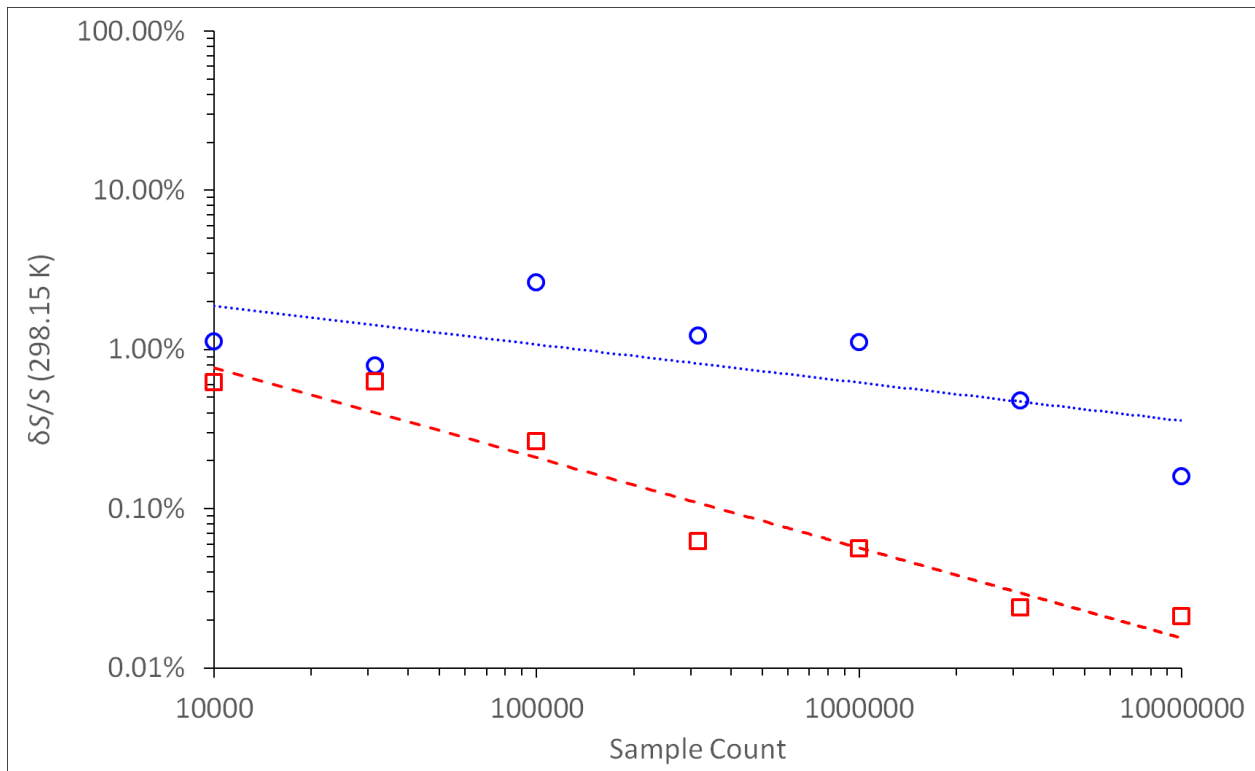


Figure 8: Average relative uncertainties in PG-corrected $\text{H}_2\text{O}_2(\text{g})$ standard molar entropies with CIMCI at UFF. Naïve MC integrations are blue circles, MISER MC integrations are red squares. The average relative statistical uncertainty scales for naïve integration with sample counts x as $0.17x^{-0.24}$ ($R^2 = 0.46$) and for MISER as $1.4x^{-0.57}$ ($R^2 = 0.94$). Each point represents the Bessel-corrected standard deviation in molar entropy of five separate, independent CIMCI runs relative to the same runs' average.

thermodynamic values like entropy or enthalpy. This means that the scaling of uncertainties in CIMCI-derived thermodynamic quantities is subject to propagations of uncertainty that are not completely straightforward. Scaling in these quantities’ uncertainties may therefore be different.

Figs. 6 and 7 show the naïve and MISER scaling behavior obtained by this study for the relative statistical uncertainty in the PG-corrected partition function and its first temperature derivative of $\text{H}_2\text{O}_2(\text{g})$ obtained from CIMCI done on a UFF PES. Both integration schemes show a polynomial decrease of the statistical uncertainty with increasing sample counts, where MISER’s uncertainty decreases roughly twice as fast as the naïve integration’s. For 10^7 points, MISER yields statistical uncertainties of around 0.35% for the partition function and 0.27% for the first temperature derivative, whereas the naïve scheme yields ones of around 6.0% for the former and 4.6% for the latter.

However, the scaling behavior of uncertainties is markedly different for entropies. Fig. 8 shows the scaling behavior of standard molar entropies derived from the runs and quantities shown in Figs. 6 and 7. Because propagation of uncertainty for entropy relies on derivatives that CIMCI does not compute, *e.g.* $\frac{\partial Q}{\partial(\frac{\partial Q}{\partial T})}$, the figure instead shows the Bessel-corrected standard deviation of the molar entropies calculated with the five runs at each sample count shown in Figs. 6 and 7. Compared to results for the partition function and its first temperature derivative, uncertainties in standard molar entropies seem to scale much slower, though the polynomial regression fit is quite poor for the naïve case. The starting relative uncertainties are significantly lower though. For MISER, sub-1% uncertainty is already achieved with 10^4 samples, whereas more than 10^6 samples are required for the same level of certainty in the partition function. This represents an interesting dynamic in the performance of CIMCI based on the way its results would be used. Depending on which thermodynamic quantities are of interest, adequate certainty in those quantities can be achieved from vastly more uncertain partition function results.

However, should partition function data be directly of interest, the scaling of the direct

output of CIMCI itself still poses a problem: in order to obtain statistical uncertainties of less than 1% for a six-dimensional integration, about 4×10^6 samples are required with MISER. That is a very large amount of samples, too many to feasibly perform the calculations on even DFT PESs. However, generic potentials that are fast enough to be feasible with CIMCI, *e.g.* force fields and DFTB methods, are often not of good enough quality to get the shape of the PES right. During the creation of this study, one of the previously highlighted interesting molecules, hydrazine (NH_2NH_2) was shown to be one such system for which both UFF and GFN1-xTB PESs seemed to have inaccuracies that would significantly impact an application of CIMCI. Not only that, but the number of samples necessary to reach this less than 1% uncertainty threshold in entropy seems to increase rapidly with larger molecules. Preliminary exploration of NH_2OH with CIMCI that is outside the scope of the study, as well as sporadic benchmarking done during the H_2O analyses, seems to show that every additional atom, *i.e.* every three additional dimensions, increases the number of samples necessary for this level of precision by about a factor of 12. This would mean that for even smaller molecules of potential interest, *e.g.* octane with 26 atoms, approximately 2×10^{30} samples would be required to get full MISER CIMCI partition function results with less than 1% numerical uncertainty. Even if one could calculate a million SPEs every second, a calculation of this size would still take more than a million times the current age of the universe. Thankfully, there are plenty of promising approaches which could get this number down to a more manageable size. In a practical sense, CIMCI need not be used to treat every single dimension of the configuration integral, only those representing types of motion that are not handled properly by RRHO, namely, torsions involving heavy atoms. The other positional coordinates could be solved by analytically integrating over harmonic normal mode potentials. Separating the two could be done with the special normal mode projector-based method proposed by Ayala and Schlegel.³² This would get a CIMCI handling of octane down to needing only about 5×10^8 samples for less than 1% uncertainty, which is still fairly large, but achievable on human timescales with enough computational resources. Alternative MC integration

sampling strategies, some of which were mentioned previously, are also being investigated by the authors of this study, as well as improvements to the core MISER algorithm itself to make it use allocated sampling budgets more efficiently.

Conclusions

The CIMCI method presented in this study computes classical thermochemistry with anharmonic effects accurately and in a manner that requires no special user input other than the global minimum geometry, the desired method for obtaining SPEs, and a user-specified symmetry number. With the application of a correction scheme, *e.g.* the PG approximation, CIMCI is able to obtain semi-classical thermodynamic quantities that match high-quality reference data better than other popular, black-box methods that use the same potential. For H_2O_2 , entropies from CIMCI correspond well — *i.e.* within $1 \text{ J mol}^{-1} \text{ K}^{-1}$ from 298.15 K to 700 K — to entropies computed from *N*-D and 1-D SEs for potential energy surfaces of different quality. By contrast, the McClurg hindered rotor scheme and the VPT2 treatment fail to capture even the basic trend in anharmonic corrections with temperature. CIMCI, by covering large parts of the PES yet at a low level of theory, correctly computed the trend and order of magnitude of anharmonic effects. The MISER scheme for recursive stratified sampling greatly increases the efficiency of CIMCI without requiring any additional setup, and can do so regardless of how many temperatures are being evaluated at once with CIMCI. For H_2O_2 , convergence with MISER was twice as fast as with the naïve scheme. As a side effect, despite not being an importance sampling strategy, MISER was nevertheless able to identify PES subspaces that contribute the most to a species’ classical partition function and its temperature derivatives. This could allow MISER sampling information to also be used to focus and target regions of a parameterisable PES that are the most important for thermochemistry.

The current primary obstacles in CIMCI are the speed at which the many, necessary SPE samples are obtained, the automatic exclusion of enantiomers and other undesirable stereoisomers from the configuration integral, and the still-significant amount of samples that are needed for low enough numerical uncertainties of larger molecules. However, there are promising ways forward for all three of these roadblocks. While these issues do prevent CIMCI from currently being used in the role for which it was originally imagined, with

further study and development, CIMCI has the potential to be a black-boxable, broadly applicable, and reasonably accurate treatment for gas-phase anharmonicity that scales well for large molecules.

Acknowledgement

Simulations were performed with computing resources granted by RWTH Aachen University under project rwth0652. The authors acknowledge funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 814143.

Supporting Information Available

The following files are available free of charge.

- MCSIA.pdf: contains formulæ for thermochemistry from partition functions
- watercomparisonW2020.ods: computation of water entropy from energy levels reported in the W2020 database⁹⁴
- H2O2comparisonMalyszek.ods: computation of H₂O₂ entropy from energy levels reported by Malyszek and Koput⁹
- H2O2_HO.xlsx: contains combined full-HO energy levels for H₂O₂
- NH2OHcomparisonLuckhaus.ods: computation of NH₂OH entropy from energy levels reported by Luckhaus⁹⁸

References

- (1) vom Lehn, F.; Cai, L.; Pitsch, H. Sensitivity analysis, uncertainty quantification, and optimization for thermochemical properties in chemical kinetic combustion models. *Proc. Combust. Inst.* **2019**, *37*, 771 – 779.
- (2) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 theory. *J. Chem. Phys.* **2007**, *126*, 084108.
- (3) Zheng, J.; Zhao, Y.; Truhlar, D. G. The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *J. Chem. Theory Comput.* **2009**, *5*, 808–821.
- (4) East, A. L. L.; Smith, B. J.; Radom, L. Entropies and Free Energies of Protonation and Proton-Transfer Reactions. *J. Am. Chem. Soc.* **1997**, *119*, 9014–9020.
- (5) Van Speybroeck, V.; Vansteenkiste, P.; Van Neck, D.; Waroquier, M. Why Does the Uncoupled Hindered Rotor Model Work Well for the Thermodynamics of *n*-Alkanes? *Chem. Phys. Lett.* **2005**, *402*, 479–484.
- (6) Fernández-Ramos, A. Accurate Treatment of Two-Dimensional Non-Separable Hindered Internal Rotors. *J. Chem. Phys.* **2013**, *138*, 134112.
- (7) Chen, W.; Poirier, B. Quantum Dynamical Calculation of all Rovibrational States of HO₂ for total Angular Momentum $J = 0$ –10. *J. Theor. Comput. Chem.* **2010**, *09*, 435–469.
- (8) Petty, C.; Chen, W.; Poirier, B. Quantum Dynamical Calculation of Bound Rovibrational States of HO₂ up to Largest Possible Total Angular Momentum, $J \leq 130$. *J. Phys. Chem. A* **2013**, *117*, 7280–7297.
- (9) Malyszczek, P.; Koput, J. Accurate ab initio potential energy surface and vibration-rotation energy levels of hydrogen peroxide. *J. Comput. Chem.* **2013**, *34*, 337–345.

- (10) Polyansky, O. L.; Kozin, I. N.; Ovsyannikov, R. I.; Małyszczek, P.; Koput, J.; Tennyson, J.; Yurchenko, S. N. Variational Calculation of Highly Excited Rovibrational Energy Levels of H_2O_2 . *J. Phys. Chem. A* **2013**, *117*, 7367–7377.
- (11) Coelho, D. V.; Brandão, J. A full dimensional potential for H_2O_2 ($X^1\text{A}$) covering all dissociation channels. *Phys. Chem. Chem. Phys.* **2017**, *19*, 1378–1388.
- (12) von Horsten, H. F. Quantendynamik groesserer Molekuele anhand reduziert-dimensionaler Modelle: Methoden und Anwendungen eines universellen Zugangs. Ph.D. thesis, Christian-Albrechts-Universitaet zu Kiel, 2008.
- (13) Kopp, W. A. Coupled large-amplitude motions in oxygenated biofuel candidates. Dissertation, RWTH Aachen University, Aachen, 2018; Veröffentlicht auf dem Publikationsserver der RWTH Aachen University; Dissertation, RWTH Aachen University, 2018.
- (14) Pitzer, K. S.; Gwinn, W. D. Energy Levels and Thermodynamic Functions for Molecules with Internal Rotation. I: Rigid Frame with Attached Tops. *J. Chem. Phys.* **1942**, *10*, 428–440.
- (15) Ellingson, B. A.; Lynch, V. A.; Mielke, S. L.; Truhlar, D. G. Statistical Thermodynamics of Bond Torsional Modes: Tests of Separable, Almost-Separable, and Improved Pitzer–Gwinn Approximations. *J. Chem. Phys.* **2006**, *125*, 084305.
- (16) Press, W. H.; Farrar, G. R. Recursive Stratified Sampling for Multidimensional Monte Carlo Integration. *Computers in Physics* **1990**, *4*, 190–195.
- (17) Barone, V. Vibrational Zero-Point Energies and Thermodynamical Functions beyond the Harmonic Approximation. *J. Chem. Phys.* **2004**, *120*, 3059–3065.
- (18) Gerber, R. B.; Ratner, M. A. *Advances in Chemical Physics*; John Wiley & Sons, Ltd, 2007; pp 97–132.

- (19) Christiansen, O. Møller–Plesset perturbation theory for vibrational wave functions. *The Journal of Chemical Physics* **2003**, *119*, 5773–5781.
- (20) Christiansen, O. Vibrational coupled cluster theory. *The Journal of Chemical Physics* **2004**, *120*, 2149–2159.
- (21) DeTar, D. F. Theoretical ab Initio Calculation of Entropy, Heat Capacity, and Heat Content. *J. Phys. Chem. A* **1998**, *102*, 5128–5141.
- (22) Pracht, P.; Grimme, S. Calculation of Absolute Molecular Entropies and Heat Capacities Made Simple. **2021**,
- (23) Zheng, J.; Truhlar, D. G. Quantum Thermochemistry: Multistructural Method with Torsional Anharmonicity Based on a Coupled Torsional Potential. *J. Chem. Theory Comput.* **2013**, *9*, 1356–1367, PMID: 26587598.
- (24) Burkhardt, C. E.; Leventhal, J. J. Vibration-rotation coupling in a Morse oscillator. *American Journal of Physics* **2007**, *75*, 686–689.
- (25) East, A. L. L.; Radom, L. Ab initio Statistical Thermodynamical Models for the Computation of Third-law Entropies. *J. Chem. Phys.* **1997**, *106*, 6655–6674.
- (26) Vansteenkiste, P.; Van Speybroeck, V.; Marin, G. B.; Waroquier, M. Ab Initio Calculation of Entropy and Heat Capacity of Gas-Phase *n*-Alkanes Using Internal Rotations. *J. Phys. Chem. A* **2003**, *107*, 3139–3145.
- (27) Li, Y.-P.; Bell, A. T.; Head-Gordon, M. Thermodynamics of Anharmonic Systems: Uncoupled Mode Approximations for Molecules. *J. Chem. Theory Comput.* **2016**, *12*, 2861–2870, PMID: 27182658.
- (28) Pitzer, K. S. Energy Levels and Thermodynamic Functions for Molecules with Internal Rotation: II. Unsymmetrical Tops Attached to a Rigid Frame. *J. Chem. Phys.* **1946**, *14*, 239–243.

- (29) Kilpatrick, J. E.; Pitzer, K. S. Energy Levels and Thermodynamic Functions for Molecules with Internal Rotation. III. Compound Rotation. *J. Chem. Phys.* **1949**, *17*, 1064–1075.
- (30) Truhlar, D. G. A simple approximation for the vibrational partition function of a hindered internal rotation. *J. Comput. Chem.* **1991**, *12*, 266–270.
- (31) McClurg, R. B.; Flagan, R. C.; Goddard III, W. A. The hindered rotor density-of-states interpolation function. *J. Chem. Phys.* **1997**, *106*, 6675–6680.
- (32) Ayala, P. Y.; Schlegel, H. B. Identification and Treatment of Internat Rotation in Normal Model Virbrational Analysis. *J. Chem. Phys.* **1998**, *108*, 2314–2325.
- (33) McClurg, R. B. Comment on: "The hindered rotor density-of-states interpolation function" [J. Chem. Phys. 106, 6675 (1997)] and "The hindered rotor density- of-states" [J. Chem. Phys. 108, 2314 (1998)]. *J. Chem. Phys.* **1999**, *111*, 7163–7164.
- (34) Chuang, Y.-Y.; Truhlar, D. G. Statistical thermodynamics of bond torsional modes. *J. Chem. Phys.* **2000**, *112*, 1221.
- (35) Chuang, Y.-Y.; Truhlar, D. G. Erratum: "Statistical thermodynamics of bond torsional modes". *J. Chem. Phys.* **2004**, *121*, 7036.
- (36) Cuang, Y.-Y.; Truhlar, D. G. Erratum: "Statistical thermodynamics of bond torsional modes". *J. Chem. Phys.* **2006**, *124*, 179903.
- (37) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Chem. Eur. J.* **2012**, *18*, 9955–9964.
- (38) Li, Y.-P.; Gomes, J.; Mallikarjun Sharada, S.; Bell, A. T.; Head-Gordon, M. Improved Force-Field Parameters for QM/MM Simulations of the Energies of Adsorption for

- Molecules in Zeolites and a Free Rotor Correction to the Rigid Rotor Harmonic Oscillator Model for Adsorption Enthalpies. *The Journal of Physical Chemistry C* **2015**, *119*, 1840–1850.
- (39) Döntgen, M. Analytic energy-level densities of separable harmonic oscillators including approximate hindered rotor corrections. *AIP Advances* **2016**, *6*, 095318.
- (40) Skouteris, D.; Calderini, D.; Barone, V. Methods for Calculating Partition Functions of Molecules Involving Large Amplitude and/or Anharmonic Motions. *Journal of Chemical Theory and Computation* **2016**, *12*, 1011–1018, PMID: 26765363.
- (41) Lynch, V. A.; Mielke, S. L.; Truhlar, D. G. Accurate vibrational-rotational partition functions and standard-state free energy values for H₂O₂ from Monte Carlo path-integral calculations. *J. Chem. Phys.* **2004**, *121*, 5148–5162.
- (42) Mielke, S. L.; Truhlar, D. G. Improved Methods for Feynman Path Integral Calculations of Vibrational-Rotational Free Energies and Application to Isotopic Fractionation of Hydrated Chloride Ions. *J. Phys. Chem. A* **2009**, *113*, 4817–4827.
- (43) Mielke, S. L.; Truhlar, D. G. Accelerating the Convergence and Reducing the Variance of Path Integral Calculations of Quantum Mechanical Free Energies by Using Local Reference Potentials. *J. Chem. Theor. Comput.* **2012**, *8*, 1589–1596.
- (44) Kamarchik, E.; Jasper, A. W. Anharmonic state counts and partition functions for molecules via classical phase space integrals in curvilinear coordinates. *J. Chem. Phys.* **2013**, *138*, 194109.
- (45) Jasper, A. W.; Gruen, Z. B.; Harding, L. B.; Georgievskii, Y.; Klippenstein, S. J.; Wagner, A. F. Anharmonic Rovibrational Partition Functions for Fluxional Species at High Temperatures via Monte Carlo Phase Space Integrals. *J. Phys. Chem. A* **2018**, *122*, 1727–1740, PMID: 29356534.

- (46) Koput, J.; Carter, S.; Handy, N. C. Potential energy surface and vibrational-rotational energy levels of hydrogen peroxide. *J. Phys. Chem. A* **1998**, *102*, 6325–6330.
- (47) Klinting, E. L.; Lauvergnat, D.; Christiansen, O. Vibrational Coupled Cluster Computations in Polyspherical Coordinates with the Exact Analytical Kinetic Energy Operator. *J. Chem. Theory Comput.* **2020**, *16*, 4505–4520, PMID: 32422041.
- (48) Carrington, T. Perspective: Computing (ro-)vibrational spectra of molecules with more than four atoms. *J. Chem. Phys.* **2017**, *146*, 120902.
- (49) Nauts, A.; Lauvergnat, D. Quantum Dynamics of Floppy Molecular Systems with ELVIBROT and TNUM. INTERNATIONAL CONFERENCE OF COMPUTATIONAL METHODS IN SCIENCES AND ENGINEERING 2009 (ICCMSE 2009). 2012; pp 948–952, 7th International Conference on Computational Methods in Science and Engineering (ICCMSE), Rhodes, GREECE, SEP 29-OCT 04, 2009.
- (50) Lauvergnat, D.; Nauts, A. Torsional energy levels of nitric acid in reduced and full dimensionality with ElVibRot and Tnum. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8405–8412.
- (51) Kopp, W. A.; Leonhard, K. General formulation of rovibrational kinetic energy operators and matrix elements in internal bond-angle coordinates using factorized Jacobians. *J. Chem. Phys.* **2016**, *145*, 234102.
- (52) Umer, M.; Leonhard, K. Ab Initio Calculations of Thermochemical Properties of Methanol Clusters. *J. Phys. Chem. A* **2013**, *117*, 1569–1582.
- (53) Umer, M.; Kopp, W. A.; Leonhard, K. Efficient yet Accurate Approximations for Ab Initio Calculations of Alcohol Cluster Thermochemistry. *J. Chem. Phys.* **2015**, *143*, 214306.

- (54) Seal, P.; Papajak, E.; Truhlar, D. G. Kinetics of the Hydrogen Abstraction from Carbon-3 of 1-Butanol by Hydroperoxyl Radical: Multi-Structural Variational Transition-State Calculations of a Reaction with 262 Conformations of the Transition State. *J. Phys. Chem. Lett.* **2012**, *3*, 264–271.
- (55) Alecu, I. M.; Zheng, J.; Papajak, E.; Yu, T.; Truhlar, D. G. Biofuel Combustion. Energetics and Kinetics of Hydrogen Abstraction from Carbon-1 in n-Butanol by the Hydroperoxyl Radical Calculated by Coupled Cluster and Density Functional Theories and Multistructural Variational Transition-State Theory with Multidimensional Tunneling. *J. Phys. Chem. A* **2012**, *116*, 12206–12213.
- (56) Kopp, W. A.; Langer, R. T.; Döntgen, M.; Leonhard, K. Hydrogen Abstraction from *n*-Butyl Formate by H \cdot and HO $_2$. *J. Phys. Chem. A* **2013**, *117*, 6757–6770.
- (57) Lucas, K. *Molecular Models for Fluids*; Cambridge University Press: Cambridge, 2007.
- (58) Eidinoff, M. L.; Aston, J. G. The Rotational Entropy of Nonrigid Polyatomic Molecules. *J. Chem. Phys.* **1935**, *3*, 379–383.
- (59) Caflisch, R. E. Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* **1998**, *7*, 1–49.
- (60) Lepage, P. G. A new algorithm for adaptive multidimensional integration. *Journal of Computational Physics* **1978**, *27*, 192 – 203.
- (61) Lepage, G. P. *VEGAS - an adaptive multi-dimensional integration program*; techreport CLNS-447, 1980.
- (62) Hahn, T. Cuba—a library for multidimensional numerical integration. *Computer Physics Communications* **2005**, *168*, 78 – 95.
- (63) Schmalz, F.; Kopp, W. A.; Kröger, L. C.; Leonhard, K. Correcting Rate Constants

- from Anharmonic Molecular Dynamics for Quantum Effects. *ACS Omega* **2020**, *5*, 2242–2253.
- (64) Frederick, J. H.; Woywod, C. General formulation of the vibrational kinetic energy operator in internal bond-angle coordinates. *J. Chem. Phys.* **1999**, *111*, 7255–7271.
- (65) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.
- (66) Schubert, E.; Gertz, M. Numerically Stable Parallel Computation of (Co-)Variance. Proceedings of the 30th International Conference on Scientific and Statistical Database Management. New York, NY, USA, 2018.
- (67) Bridgeman, A. J.; Cavigliasso, G.; Ireland, L. R.; Rothery, J. The Mayer bond order as a tool in inorganic chemistry. *J. Chem. Soc., Dalton Trans.* **2001**, 2095–2108.
- (68) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86). *Journal of Chemical Theory and Computation* **2017**, *13*, 1989–2009, PMID: 28418654.
- (69) Pauling, L. Atomic Radii and Interatomic Distances in Metals. *Journal of the American Chemical Society* **1947**, *69*, 542–553.
- (70) Rüger, R.; Franchini, M.; Trnka, T.; Yakovlev, A.; van Lenthe, E.; Philipsen, P.; van Vuren, T.; Klumpers, B.; Soini, T. AMS 2020. 2020; <http://www.scm.com>, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands.
- (71) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

- (72) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of Tight-Binding-Like Potentials on the Basis of Density-Functional Theory: Application to Carbon. *Phys. Rev. B* **1995**, *51*, 12947–12957.
- (73) Rüger, R.; Yakovlev, A.; Philipsen, P.; Borini, S.; Melix, P.; Oliveira, A. F.; Franchini, M.; van Vuren, T.; Soini, T.; de Reus, M.; Asl, M. G.; Teodoro, T. Q.; McCormack, D.; Patchkovskii, S.; Heine, T. AMS DFTB 2020. 2020; <http://www.scm.com>, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands.
- (74) Baerends, E. et al. ADF 2020. 2020; <http://www.scm.com>, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands.
- (75) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931–967.
- (76) Fan, L.; Ziegler, T. Application of density functional theory to infrared absorption intensity calculations on transition-metal carbonyls. *The Journal of Physical Chemistry* **1992**, *96*, 6937–6941.
- (77) Fan, L.; Ziegler, T. Application of density functional theory to infrared absorption intensity calculations on main group molecules. *The Journal of Chemical Physics* **1992**, *96*, 9005–9012.
- (78) Wolff, S. K. Analytical second derivatives in the Amsterdam density functional package. *International Journal of Quantum Chemistry* **2005**, *104*, 645–659.
- (79) Jacobsen, H.; Bérces, A.; Swerhone, D. P.; Ziegler, T. Analytic second derivatives of molecular energies: a density functional implementation. *Computer Physics Communications* **1997**, *100*, 263–276.

- (80) Bércecs, A.; Dickson, R. M.; Fan, L.; Jacobsen, H.; Swerhone, D.; Ziegler, T. An implementation of the coupled perturbed Kohn-Sham equations: perturbation due to nuclear displacements. *Computer Physics Communications* **1997**, *100*, 247–262.
- (81) Van Lenthe, E.; Baerends, E. J. Optimized Slater-type basis sets for the elements 1–118. *Journal of Computational Chemistry* **2003**, *24*, 1142–1156.
- (82) Lenthe, E. v.; Baerends, E. J.; Snijders, J. G. Relativistic regular two-component Hamiltonians. *The Journal of Chemical Physics* **1993**, *99*, 4597–4610.
- (83) van Lenthe, E.; Baerends, E. J.; Snijders, J. G. Relativistic total energy using regular approximations. *The Journal of Chemical Physics* **1994**, *101*, 9783–9792.
- (84) van Lenthe, E.; Ehlers, A.; Baerends, E.-J. Geometry optimizations in the zero order regular approximation for relativistic effects. *The Journal of Chemical Physics* **1999**, *110*, 8943–8953.
- (85) Vigna, S. An Experimental Exploration of Marsaglia’s Xorshift Generators, Scrambled. *ACM Trans. Math. Softw.* **2016**, *42*.
- (86) Vigna, S. Further scramblings of Marsaglia’s xorshift generators. *Journal of Computational and Applied Mathematics* **2017**, *315*, 175 – 181.
- (87) Frisch, M. J. et al. Gaussian~16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.
- (88) Ghysels, A.; Verstraelen, T.; Hemelsoet, K.; Waroquier, M.; van Speybroeck, V. TAMkin: A Versatile Package for Vibrational Analysis and Chemical Kinetics. *J. Chem. Inf. Model.* **2010**, *50*, 1736–1750.
- (89) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.

- (90) Temelso, B.; Shields, G. C. The Role of Anharmonicity in Hydrogen-Bonded Systems: The Case of Water Clusters. *J. Chem. Theory Comput.* **2011**, *7*, 2804–2817.
- (91) Wang, Y.; Babin, V.; Bowman, J. M.; Paesani, F. The Water Hexamer: Cage, Prism, or Both. Full Dimensional Quantum Simulations Say Both. *J. Am. Chem. Soc.* **2012**, *134*, 11116–11119, PMID: 22731508.
- (92) Howard, J. C.; Tschumper, G. S. Benchmark Structures and Harmonic Vibrational Frequencies Near the CCSD(T) Complete Basis Set Limit for Small Water Clusters: (H₂O)_n = 2, 3, 4, 5, 6. *J. Chem. Theory Comput.* **2015**, *11*, 2126–2136, PMID: 26574415.
- (93) Malloum, A.; Fifen, J. J.; Dhaouadi, Z.; Nana Engo, S. G.; Conradie, J. Structures, relative stability and binding energies of neutral water clusters, (H₂O)_{2–30}. *New J. Chem.* **2019**, *43*, 13020–13037.
- (94) Furtenbacher, T.; Tóbiás, R.; Tennyson, J.; Polyansky, O. L.; Császár, A. G. W2020: A Database of Validated Rovibrational Experimental Transitions and Empirical Energy Levels of H₂¹⁶O. *J. Phys. Chem. Ref. Data* **2020**, *49*, 033101.
- (95) Furtenbacher, T.; Tóbiás, R.; Tennyson, J.; Polyansky, O. L.; Kyuberis, A. A.; Ovsyanikov, R. I.; Zobov, N. F.; Császár, A. G. The W2020 Database of Validated Rovibrational Experimental Transitions and Empirical Energy Levels of Water Isotopologues. II. H₂¹⁷O and H₂¹⁸O with an Update to H₂¹⁶O. *J. Phys. Chem. Ref. Data* **2020**, *49*, 043103.
- (96) Chase, J., M.W. NIST-JANAF Thermochemical Tables. *J. Phys. Chem. Ref. Data, Monograph 9* **1998**, *4*, 1–1951.
- (97) Dorofeeva, O. V.; Iorish, V. S.; Novikov, V. P.; Neumann, D. B. NIST-JANAF Thermochemical Tables. II. Three Molecules Related to Atmospheric Chemistry: HNO₃, H₂SO₄, and H₂O₂. *J. Phys. Chem. Ref. Data* **2003**, *32*, 879–901.

- (98) Luckhaus, D. The rovibrational spectrum of hydroxylamine: A combined high resolution experimental and theoretical study. *J. Chem. Phys.* **1997**, *106*, 8409–8426.