Coordinate Transformation and the Nuclear Vibrational Problem

Marat Sibaev

A thesis presented for the degree of

Doctor of Philosophy in Chemistry

Department of Chemistry

University of Canterbury

January, 2016

Contents

Acknowledgements

Thanks to my family for their support and Deb for being an awesome supervisor. So long and thanks for all the cake!

Abstract

This thesis describes the development and testing of a new suite of methods for solving the nuclear vibrational Schrödinger equation in order to calculate anharmonic fundamental vibrational frequencies for realistically sized chemical systems.

To get around the potential energy surface (PES) construction bottleneck, we present a new method based upon constructing the PES in a curvilinear coordinate system and transforming back to rectilinear normal mode coordinates to facilitate solving the vibrational problem.

We also implement and benchmark the performance of a screened vibrational configuration interaction method for calculating anharmonic fundamental frequencies. Both methods combined allow modelling of vibrational spectra for molecules with up to 20 atoms to be calculated routinely on a desktop computer, provided that the *ab initio* calculations required to construct the PES are computationally feasible.

Declaration

All the work presented in this thesis is my own except as otherwise stated, and has been undertaken under the supervision of Deborah Crittenden in the Department of Chemistry, University of Canterbury. Parts of this thesis have been previously published as original research papers or submitted as manuscripts:

– Chapter 2: M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* **2015**, *36*, 2200-2207, doi:{10.1002/jcc.24192};

– Chapter 3: M. Sibaev, D. L. Crittenden, *Comput. Phys. Commun.*, submitted;

– Chapter 6: M. Sibaev, D. L. Crittenden, *J. Phys. Chem. A*, **2015**, *119*, 13107-13112, doi:{10.1021/acs.jpca.5b11386}.

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 2: M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* 2015, 36, 2200-2207

Please detail the nature and extent (%) of contribution by the candidate:

1) All work (100%) and most of the writing (85%) was by the candidate

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Deborah Crittenden* Signature: D. Crittenden Date: *28/01/2016*

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 3: M. Sibaev, D. L. Crittenden, *Comput. Phys. Commun.*, submitted

Please detail the nature and extent (%) of contribution by the candidate:

All work (100%) and most of the writing (75%) was by the candidate

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Deborah Crittenden* Signature: D. Crittenden Date: *28/01/2016*

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 6: M. Sibaev, D. L. Crittenden, *J. Phys. Chem. A*, 2015, 119, 13107- 13112

Please detail the nature and extent (%) of contribution by the candidate:

Most work (90%) and most of the writing (65%) was by the candidate

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Deborah Crittenden* Signature: D. Crittenden Date: *28/01/2016*

Abbreviations

PEF – potential energy function, also referred to as potential

PES – potential energy hyper-surface (used interchangeably with PEF)

PES' – potential energy hyper-surfaces, plural of PES

QFF – quartic force field (PEF approximated as a 4^{th} order Taylor series expansion around equilibrium)

SFF – sextic force field (PEF approximated as a 6^{th} order Taylor series expansion around equilibrium)

KEO – kinetic energy operator

MC – mode coupling

MR – mode represenation

VCI – vibrational configuration interaction

VPT – vibrational perturbation theory

VPT2 – second order vibrational perturbation theory

MAE – mean absolute error

CCSD(T) – coupled cluster with single, double and perturbative triple excitations

HF – Hartree-Fock method

Chapter 1

Introduction

The experimental techniques of infrared and Raman spectroscopy are regularly used to study molecules and materials. Their spectra characterise molecular vibrations, which are determined by molecular structure and the strengths of intra- and intermolecular interactions. The high accuracy of modern spectrometers and the ability to study systems in gas, liquid and solid states, makes vibrational spectroscopy a key research tool in the design and development of new molecules and materials such as pharmaceuticals, photovoltaics and catalysts.

However, quantitative and unambiguous assignment of IR and Raman spectra becomes very challenging when studying systems of more than a few atoms. A purely experimental approach requires complicated procedures, where perturbations are introduced into the system and resultant changes in the spectrum are studied. Alternatively, or concurrently, accurate theoretical studies can help elucidate the link between the shape of the spectrum and dynamics of the system.

Normal mode analysis is the most widely used computational method for modelling vibrational spectra.[1] It assumes that vibrations occur along normal mode coordinates, which are linear in Cartesian space, and that the potential is harmonic in those coordinates. However, when accurate predictions are required, more rigorous approaches that account for anharmonicity in the potential must be used.

A more complete form of a quantum Hamiltonian incorporating anharmonicity has been derived by Watson.[2,3] It uses normal mode coordinates to ensure separability of the kinetic energy operator (KEO), provided less important angular

momentum terms and inverse moment of inertia terms are ignored, and the anharmonicity can be included via higher order power series expansion of the potential energy function (PEF). The simple form of the Watson Hamiltonian, together with the fact that normal mode coordinates can be uniquely and rather easily defined for any system, has made it the most useful starting point when developing efficient black box procedures for solving the vibrational problem. A multitude of methods have been proposed and implemented, and the main bottleneck restricting their application and accuracy has become constructing the PEF.

The rectilinear form of normal mode coordinates, although leading to a decoupled KEO, is not optimal for the description of the PEF. It has long been known that curvilinear coordinates, such as valence internal coordinates, lead to a less coupled and faster converging PEF.[4–9] However, their use has been limited to smaller molecules, mainly due to the complicated nature of the KEO. Additionally, the curvilinear coordinates for which the KEO can be derived in an analytical form convenient for subsequent numerical treatments, such as polyspherical harmonics, form a redundant set. This requires manual user intervention for defining a non-redundant subset. Therefore, the use of curvilinear coordinates for full treatment of the vibrational problem is not appropriate if a black box procedure is required.

However, construction of the PEF is completely separate from solving the vibrational problem. Thus, as long as the PEF can ultimately be transformed to normal mode coordinates, it can be constructed in any appropriate coordinate set. Curvilinear coordinates, defined as combinations of bond lengths, bond angles, and dihedral angles, are an obvious choice, as they are physically intuitive. Because the KEO does not have to be defined in the curvilinear coordinates, any non-redundant set can be chosen, e.g. the delocalised internal coordinates of Baker *et al.*[10] Constructing the PEF in curvilinear coordinates reduces the coupling in the PEF and therefore decreases the number of terms required for its accurate representation. Developing and testing this new approach is the main aim of this thesis.

Two things are required in order to assess the ability of the coordinate transformation procedure described above to generate accurate force fields for calculating fundamental frequencies:

- a library of benchmark analytical PEFs;
- implementation of an accurate method for solving the nuclear vibrational Schrödinger equation.

During method development it is essential to separate the various sources of error, so that the accuracy of the proposed approach can be quantified and any limitations properly understood. For that reason, accurate benchmark data in a form applicable to a wide range of treatments is essential for fostering future development of this field. Such benchmark data is spread around the literature in a form of small molecule studies, where highly accurate semi-global potentials have been constructed and the nuclear vibrational problem rigorously solved. In order to ensure that this data is preserved and can be used for future testing and benchmarking, we have developed a library containing a selection of such high quality potentials that is capable of providing energy and derivative information to 6^{th} order at any specified geometry, in some of the most commonly used coordinate systems. It is written mainly in Python, and symbolic differentiation algorithms are heavily used, ensuring that a wide range of potentials, in various forms, can be implemented as seamlessly as possible. Our Python potential energy surfaces library (PyPES library), is available freely online.[11,12]

Comparison between calculated and literature anharmonic frequencies was used to verify the implementation of the potential energy surfaces within the PyPES library. However, in the absence of any freely available and well documented variational nuclear vibrational structure codes it was necessary to develop our own. Our general and flexible vibrational configuration interaction (VCI)[13] algorithm is implemented primarily in Python for accessibility and customisability, with computationally intensive parts written in C.

However, including enough excited states in the VCI wavefunction expansion to obtain converged fundamental frequencies rapidly becomes computationally prohibitive for larger molecules. Therefore, we investigate two different approaches for obtaining anharmonic frequencies at a lower computational cost:

- selecting VCI configurations based upon their contribution to the second order vibrational perturbation theory (VPT2) energy expression;

- a very simple harmonic frequency correction method.

Unlike VCI, second order vibrational perturbation theory (VPT2) is computationally facile, and can provide accurate results when not affected by divergences due to near-degenerate states, which is a problem that commonly plagues perturbation theory. In recent research, a simpler and more robust algorithm has been proposed that involves combining VCI and VPT2 based methods in an iterative procedure,[14–17] in which a low order VCI wavefunction is used as a starting point for a VPT2 correction, which concurrently selects a larger VCI space containing any near-degenerate and strongly interacting states. These states are then explicitly included in the VCI matrix for the next iteration. Implementation and rigorous testing of a new way for performing this procedure forms another key part of this thesis.

Unfortunately, due to the complexity of rigorously including anharmonic effects, there is always a limit on the size of the system that can be effectively studied, beyond which normal mode analysis becomes the only computationally feasible method. In such cases, anharmonicity may be estimated using an empirical model. Most commonly, a simple linear scaling of harmonic frequencies is performed, with scaling factors parameterised against experimental data to capture both anharmonic effects and account for any deficiencies in the underlying electronic structure model.[18–24] However, we argue that these two independent sources of error in anharmonic frequencies should be accounted for independently. Using our benchmark library of potential energy surfaces to eliminate methodological incompleteness errors, we show that anharmonicity grows approximately quadratically as a function of harmonic frequencies. We therefore propose a simple one parameter quadratic correction model that is more accurate than its linear counterpart.

The thesis is structured so that each project is contained in a separate chapter in journal article format (Chapters 2-6). Chapters 7 and 8 outline directions for future work and summarise key results from this thesis.

References

[1] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, New York, **1955**.

[2] J. Watson, *Mol. Phys.* **1968**, *15*, 479–490.

[3] J. Watson, *Mol. Phys.* **1970**, *19*, 465–487.

[4] T. Carrington in *Encyclopedia of Computational Chemistry*, *Vol. 5*, (Ed.: P. Schleyer), Wiley, Chichester, **1998**, p. 3157.

[5] A. G. Császár in *Encyclopedia of Computational Chemistry*, *Vol. 1*, (Ed.: P. Schleyer), Wiley, Chichester, **1998**, p. 13.

[6] L. Joubert-Doriol, B. Lasorne, F. Gatti, M. Schröder, O. Vendrell, H.-D. Meyer, *Comp. Theor. Chem.* **2012**, *990*, 75–89.

[7] E. L. Sibert III, *J. Chem. Phys.* **1989**, *90*, 2672–2683.

[8] Y. Zhang, S. Klippenstein, R. Marcus, *J. Chem. Phys.* **1991**, *94*, 7319–7334.

[9] A. Capobianco, R. Borrelli, C. Noce, A. Peluso, *Theor. Chem. Acc.* **2012**, *131*, 1–10.

[10] J. Baker, A. Kessi, B. Delley, *J. Chem. Phys.* **1996**, *105*, 192–212.

[11] PyPES Library of Potential Energy Surfaces, `http://sourceforge.net/projects/pypes-lib/`.

[12] M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* **2015**, *36*, 2200–2207.

[13] M. Sibaev, D. L. Crittenden, *Comp. Phys. Comm.* **2016**, Submitted.

[14] G. Rauhut, *J. Chem. Phys.* **2007**, *127*, 184109.

[15] Y. Scribano, D. M. Benoit, *Chem. Phys. Lett.* **2008**, *458*, 384–387.

[16] M. Neff, G. Rauhut, *J. Chem. Phys.* **2009**, *131*, 124129.

[17] P. Carbonniere, A. Dargelos, C. Pouchan, *Theor. Chem. Acc.* **2010**, *125*, 543–554.

[18] M. P. Andersson, P. Uvdal, *J. Phys. Chem. A* **2005**, *109*, 2937–2941.

[19] S. G. Andrade, L. C. S. Goncalves, F. E. Jorge, *J. Mol. Struct. (THEOCHEM)* **2008**, *864*, 20–25.

[20] M. L. Laury, S. E. Boesch, I. Haken, P. Sinha, R. A. Wheeler, A. K. Wilson, *J. Comput. Chem.* **2011**, *32*, 2339–2347.

[21] J. A. Pople, A. P. Scott, M. W. Wong, L. Radom, *Isr. J. Chem.* **1993**, *33*, 345–350.

[22] A. P. Scott, L. Radom, *J. Phys. Chem.* **1996**, *100*, 16502–16513.

[23] J. P. Merrick, D. Moran, L. Radom, *J. Phys. Chem. A* **2007**, *111*, 11683–11700.

[24] P. Sinha, S. E. Boesch, C. M. Gu, R. A. Wheeler, A. K. Wilson, *J. Phys. Chem. A* **2004**, *108*, 9213–9217.

Chapter 2

The PyPES Library of High Quality Semi-Global Potential Energy Surfaces

2.1 Introduction

Since the advent of modern computational quantum chemistry, the development of new methods for solving the electronic Schrödinger equation has been facilitated by databases of benchmark molecular geometries and energies. Today, computational chemists can choose from a range of electronic structure methods to suit the chemical problem at hand, desired level of accuracy and computational resources available.

A lot of progress has been made in developing methods for solving the nuclear vibrational Schrödinger equation[1–31], which have been described in some recent reviews[32–37]. However, harmonic normal mode analysis remains the most widely used method for solving the nuclear vibrational Schrödinger equation. This can be attributed, in large part, to the difficulty and computational cost associated with constructing *ab initio* potential energy surfaces (PES). The problem of obtaining accurate PES representations becomes particularly acute when testing new approaches for solving the nuclear ro-vibrational Schrödinger equation. Quantifying and controlling for errors in *ab initio* derived potential energy surfaces is difficult, as errors can arise from a multitude of sources *e.g.* level of theory, basis set incompleteness,

and numerical differentiation discretization error. Therefore, it is hard to determine whether disagreements between calculated frequencies and gas phase spectra are due to inadequacies in the electronic structure model, or method used to solve the nuclear Schrödinger equation, or both. To circumvent this problem, access to a benchmark data set of highly accurate semi-global potential energy surfaces is required.

Fortunately, a lot of time and effort has been devoted to constructing spectroscopically accurate analytic potential energy surfaces and high-level *ab initio* internal coordinate force fields for small molecules. A selection of these are implemented within the PyPES library, which has been carefully compiled to contain molecules with up to 6 atoms that display a representative range of topologies and bonding patterns. The form of the potential varies between molecules. The simplest semi-global PES' included in the PyPES library are quartic Taylor series expansions in internal coordinates about the equilibrium geometry, while more complicated PES representations involve parameterized functions that describe multiple local minima, particularly for molecules with low-barrier rotations. These semi-global PES' describe regions close to the symmetry-equivalent global minima with high accuracy, but the accuracy deteriorates at longer range and/or around other stationary points on the global PES.

The accuracy of the potential energy surfaces within PyPES also vary, from spectroscopically accurate PES that reproduce experimental gas phase IR spectra to within 1 cm^{-1} compiled from the astrochemistry and astrophysics literature, to high quality *ab initio* derived surfaces that typically reproduce experimental results to within 5 cm^{-1}, where high enough quality gas phase experimental data are available for comparison.

Recognising that the form of the potential energy surface and choice of coordinate system vary between PES definitions and representations underlying nuclear vibrational structure models, the PyPES library provides a framework for evaluating energies and derivatives of the energy up to 6^{th} order with respect to a range of common coordinate systems, including curvilinear internal, Cartesian and normal mode coordinates. These benchmark data will enable sources of error in approximate nuclear vibrational models to be isolated and quantified in a statistically meaningful

way. This should prove particularly useful in the continuing development of 'black-box' nuclear vibrational structure theories that are scalable to larger molecules.

All potential energy surfaces within the library are accessible through a common user interface, and wrappers written in C, FORTRAN, MATLAB and Mathematica are provided to allow easy interfacing with other non-python based codes. All wrappers interface with PyPES via data files containing unique upper triangular elements of derivative matrices in text format. Code fragments are provided to supply coordinates to PyPES, evaluate the PES and its derivatives, and read the output data back into multidimensional arrays (FORTRAN, MATLAB, Mathematica) or appropriately indexed one-dimensional arrays (C).

For advanced users, the PyPES library has also been designed to be readily extensible. New surfaces can be easily implemented, either by supplying appropriate force constants, *i.e.* coefficients of Taylor series expansions in any of the implemented coordinate systems, or through symbolic differentiation of more complicated functional forms. Thus, we anticipate that PyPES will also function as a general platform for the implementation and distribution of high quality potential energy surfaces for a broader range of molecules with different sizes and topologies.

2.2 Methodology

2.2.1 Terminology

The following notation is used throughout: $\mathbf{S}$, $\mathbf{X}$ and $\mathbf{Q}$ represent vectors of internal, Cartesian and normal mode coordinates, respectively; while r, θ, τ and ω refer to bond length, bond angle, dihedral angle and out-of-plane coordinates.

Sets of derivatives are denoted using square brackets. For example, derivatives of the potential with respect to internal coordinates to arbitrary order are denoted $\left[\frac{\partial V}{\partial \mathbf{S}}\right]$, while derivatives of the set of internal coordinates with respect to Cartesian displacements are denoted $\left[\frac{\partial \mathbf{S}}{\partial \mathbf{X}}\right]$. Note that this notation implies the complete set of derivatives up to the maximum implemented capability of 6^{th} order.

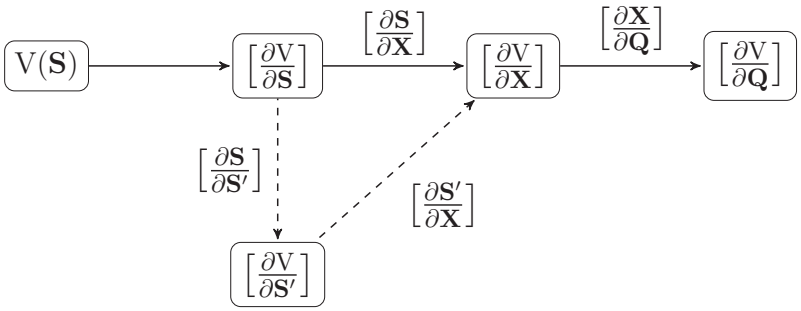


Figure 2.1: Outline of operations performed by PyPES. The first unannotated arrow denotes a differentiation step, while subsequent annotated arrows denote coordinate transformation steps. Each annotation specifies the derivative sets required to achieve the corresponding coordinate transformation. Dotted arrows indicate optional steps. V($\mathbf{S}$) represents the original PES as a function of implemented internal coordinates, $\mathbf{S}$, which is first differentiated to a chosen order with respect to the original coordinate set. This may be followed by transformation into an alternative internal coordinate set, $\mathbf{S}'$. Regardless of whether this option is chosen or not, the next step is transformation to give energy derivatives with respect to Cartesian coordinates, $\mathbf{X}$, followed by a linear transformation into normal mode coordinates, $\mathbf{Q}$. The process can be stopped at any step.

2.2.2 PyPES Outline

The flow chart illustrating the operations performed by PyPES is given in Figure 2.1.

The PyPES library contains only analytic potentials formulated in terms of the implemented internal coordinates listed in Table 2.1. This set ($S_1 + S_2$) contains most commonly used internal coordinates, giving the PyPES library broad utility as a framework for implementing and distributing a wide range of analytical PES'. For clarity, definitions of the 'fundamental' internal coordinates used by PyPES (set S_1) are provided as Supporting Information.

If V($\mathbf{S}$) is formulated as a Taylor series expansion, derivatives of the energy with respect to the internal coordinates at any geometry, $\left[\frac{\partial V}{\partial \mathbf{S}}\right]$, are generated using

Table 2.1: Summary of internal coordinates that have been implemented, where Δr and $\Delta\theta$ are deviations from equilibrium for r and θ coordinates, respectively; α is the Morse parameter; β_1, β_2, and β_3 are Carter-Handy parameters. The distinction between coordinates in S_1 and S_2 refers to the process by which their derivatives with respect to Cartesian coordinates are obtained, as outlined in Figure 2.2 and the text. Coordinate conversion operations M, SPF and CH, refer to transformation into Morse[38], Simons-Parr-Finlan[39] and Carter-Handy[40] coordinates, respectively.

S_1	$\xrightarrow{\hat{f}}$	S_2
r	M	$f_{\mathrm{M}}(r) = 1 - e^{-\alpha\Delta r}$
r	SPF	$f_{\mathrm{SPF}}(r) = \frac{\Delta r}{r}$
$\cos(\theta)$	arccos	θ
$\cos(\theta)$	arccos,CH	$f_{\mathrm{CH}}(\theta) = \beta_1\Delta\theta + \beta_2\Delta\theta^2 + \beta_3\Delta\theta^3$
τ	sin	$\sin(\tau)$
$\sin(\omega)$	arcsin	ω
$\sin(\omega')$	arcsin	ω'

custom-written code within PyPES. For more complicated PES representations, *e.g* involving symmetrized internal coordinates, derivatives both at and off equilibrium can only be found using the symbolic differentiation package SymPy. Derivatives up to 6^{th} order can be calculated at this step.

It is helpful to note that the PyPES process may be terminated at this point, with the derivative data set returned as output. This information is likely to be of limited use in general, as the internal coordinates are molecule and PES-specific, requiring customized code to construct a PES representation. However, if only 0^{th} derivatives are requested, the energy of the input structure is returned. Coordinates may be supplied in either Cartesian or internal coordinates, providing a convenient way of mapping out V($\mathbf{S}$) over a grid. In this way, PyPES can be run as a 'pseudo'-electronic structure code, providing highly accurate single point energies.

However, PyPES is designed to extract or derive far more information at each geometry than just single point energies, through its coordinate transformation code. The coordinate transformations implemented within PyPES can be formulated using the chain rule and some basic calculus. They have been thoroughly described in the literature[41–44], so will not be elaborated further here.

Once derivatives of the energy with respect to an initial set of internal coordinates, $\left[\frac{\partial V}{\partial \mathbf{S}}\right]$, have been generated, they can optionally be transformed into derivatives with respect to a different set of internal coordinates, $\left[\frac{\partial V}{\partial \mathbf{S}'}\right]$, following the dotted line in Figure 2.1. This transformation can only be carried out if all elements of the new internal coordinate set, $\mathbf{S}'$, can be expressed as straightforward functions of their counterparts in the original coordinate set, $\mathbf{S}$. This makes the required derivatives for each coordinate-pair $\left[\frac{\partial S'}{\partial S}\right]$ easy to derive and hard-code. Implemented transformations between coordinates are shown in Table 2.1, and can be carried out in either direction across each row of this table.

However, it is typically more useful to transform from simple bond length, angle, dihedral and out of plane angle coordinates to more sophisticated coordinates with appropriate asymptotic or periodic behaviour *e.g.* Morse or inverse bond length coordinates, trigonometric or polynomial functions of angular coordinates. The benefits of this procedure have been recently reviewed by Fortenberry *et al.*[45] When transforming into Morse or Carter-Handy coordinates, relevant parameters need to

be specified.

Regardless of whether the optional coordinate transformation step is carried out or not, the next step in the PyPES procedure is transformation into Cartesian coordinates, to obtain the derivative set $\left[\frac{\partial V}{\partial \mathbf{X}}\right]$. We note that if the order of the Cartesian coordinate derivative set, $\left[\frac{\partial V}{\partial \mathbf{X}}\right]$, is equal to or lower than the order of the internal coordinate derivative set, $\left[\frac{\partial V}{\partial \mathbf{S}}\right]$, the Cartesian derivatives will be independent of the choice of the internal coordinate system. In other words, there is nothing to be gained by pre-transforming to an alternative coordinate system, *i.e.* following the dashed arrows in Figure 2.1. However, if the order of the Cartesian coordinate derivative set exceeds that of the internal coordinates, the higher Cartesian derivatives will strongly reflect the asymptotic behaviour of $V(\mathbf{S})$, so pre-transforming to more sophisticated coordinates with appropriate limiting behaviour is advantageous. In either case, to achieve the subsequent transformation into Cartesian coordinates, the derivatives of each internal coordinate with respect to each Cartesian coordinate, $\left[\frac{\partial \mathbf{S}}{\partial \mathbf{X}}\right]$, must be evaluated.

For computational expedience, we divide our set of internal coordinates into a 'fundamental' set (S_1) and a 'derived' set (S_2). As the coordinates in set S_1 have been widely employed to define force fields used in molecular mechanics and vibrational spectroscopy, many articles have been devoted to efficient evaluation of their derivatives with respect to Cartesian coordinates.[46–53] In this work, we closely follow the formulation of Tuzun *et al.*[46] for r, $\cos(\theta)$, τ and $\sin(\omega)$ coordinates. We extend their expressions up to 5^{th} order derivatives with repetitive use of the chain rule, making no attempts at complicated simplifications. We have derived expressions for derivatives of the $\sin(\omega')$ coordinate independently, and report them to first order in the Supporting Information. Again, higher derivatives are derived by repeated application of the chain rule from these expressions.

Derivatives of coordinates from set S_2 with respect to Cartesian coordinates, $\left[\frac{\partial S_2}{\partial \mathbf{X}}\right]$, are obtained from $\left[\frac{\partial S_1}{\partial \mathbf{X}}\right]$ via coordinate transformation, as illustrated in Figure 2.2. This requires derivatives of each derived coordinate in S_2 with respect to its base coordinate in S_1, $\left[\frac{\partial S_2}{\partial S_1}\right]$, expressions for which have been hard-coded within PyPES to 6th order.

This completes the specification of the procedure required to obtain derivatives

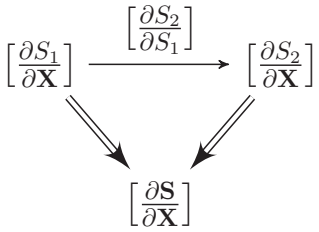


Figure 2.2: Internal coordinates are divided into sets S_1 and S_2 (Table 2.1) for the purpose of obtaining derivatives of internal coordinates with respect to Cartesian coordinates. Derivatives of coordinates in S_1 can be efficiently calculated directly via hard-coded expressions, while derivatives of coordinates in S_2 are most easily and efficiently obtained by coordinate transformation from derivatives in set S_1. Both pathways are generally used to populate $\left[\frac{\partial \mathbf{S}}{\partial \mathbf{X}}\right]$. This process is denoted by double-lined arrows.

of any internal coordinate with respect to Cartesian displacements, $\left[\frac{\partial \mathbf{S}}{\partial \mathbf{X}}\right]$, illustrated in Figure 2.2. Because $\left[\frac{\partial \mathbf{S}}{\partial \mathbf{X}}\right]$ have been implemented to 5^{th} order, transformation of $\left[\frac{\partial V}{\partial \mathbf{S}}\right]$ to $\left[\frac{\partial V}{\partial \mathbf{X}}\right]$ can be performed up to 5^{th} order off-equilibrium and up to 6^{th} order at the equilibrium geometry.

The final step of the PyPES procedure is transformation into normal mode coordinates. For completeness, the normal mode analysis code is included in PyPES, and all normal mode frequencies and normalized eigenvectors are pre-computed at the equilibrium geometry and stored separately for each molecule contained within the PyPES library. For generality, we provide the option for the user to specify alternative rectilinear basis vectors in Cartesian space via text file input.

One of the key features of the PyPES library is its generality and extensibility. To enable this, a range of approaches have been implemented or incorporated within PyPES for evaluating derivatives of functions, as described in Table 2.2. The SymPy interface is provided to enable easy evaluation of derivatives for any potential energy surface expressed as a function of implemented internal coordinates, no matter how complex the function is. Likewise, the custom written symbolic differentiation code enables straightforward implementation of new coordinates, provided those coordinates can be formulated in terms of operations on internuclear vectors, *e.g.* see reference [48]. However, this lies beyond the scope of the current paper, as

Table 2.2: Methods available within PyPES for evaluating derivatives. Code length and complexity are related to ease of implementation but inversely proportional to execution speed.

Method	Applicability	Ease of implementation	Execution speed
Hard-coded expressions	Specific to implemented functions only	Hard	Fast
Custom-written symbolic differentiation	Applicable to algebraic expressions for which derivatives of terms are available	Moderate	Moderate
SymPy	General	Easy	Slow

all derivative transformations performed on the PES' contained within the PyPES library rely on hard-coded expressions for maximum efficiency.

2.2.3 Code

The PyPES program and PES database is written in Python and is compatible with versions 2.7.0 and later. For full functionality, it requires the python packages NumPy, SciPy and SymPy. All time consuming steps were optimised and transformed to C code with Cython v0.21. Derivatives of the energy with respect to either internal coordinates, Cartesian coordinates or normal mode coordinates are output to file, with a separate file for each derivative order, and a summary file. Scripts for reloading data into MATLAB, Mathematica, C and FORTRAN and storing them in appropriate array structures are included in the package. Source code and user documentation are available from http://sourceforge.net/projects/pypes-lib.

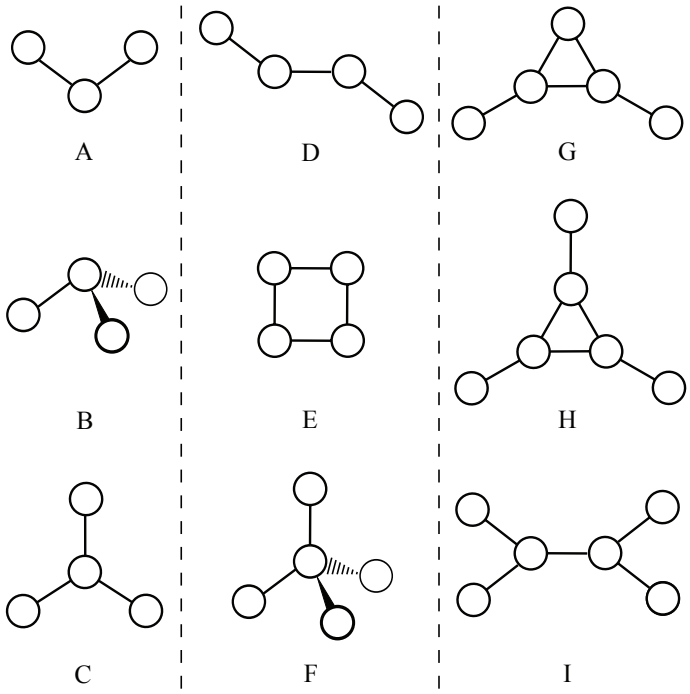


Figure 2.3: Types of molecular geometries included in the library. Circles represent atoms and lines indicate connectivity.

2.2.4 Library

Table 2.3 contains a list of potential energy surfaces included in the library. They represent 50 different molecules, of which 17 are triatomic, 21 are tetraatomic, 10 contain 5 atoms, plus a pair of 6 atom molecules. They cover a wide range of geometries and bonding patterns, as summarised in Figure 2.3.

Table 2.3: List of semi-global potential energy surfaces that have been implemented organised by the molecular system they represent and the type of equilibrium geometry, as shown in Figure 2.3. The third column contains the set of internal coordinates with respect to which the PES is differentiated in the first step.

Molecule	**Geometry**	**Internals**	**References**
H_2O	A	$\{r, \theta\}$	[54]
NH_2^-	A	$\{r, \theta\}$	[55]
HO_2^+ (X^3A'')	A	$\{r, \theta\}$	[54]
HO_2^+ (A^1A')	A	$\{r, \theta\}$	[54]
HOCl	A	$\{r, f_{CH}(\theta)\}$	[56]

Continued on next page

Table 2.3 – *Continued from previous page*

Molecule	**Geometry**	**Internals**	**References**
HOBr	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[56]
HOF	A	$\{r, \theta\}$	[57]
PF_2^+	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[58]
PO_2^-	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[58]
SO_2	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[58]
SiF_2	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[58]
F_2O	A	$\{r, \theta\}$	[57]
BrO_2 (X^2B_1)	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[59]
BrO_2 (A^2A_2)	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[59]
ClO_2 (X^2B_1)	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[59]
ClO_2 (A^2A_2)	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[59]
ClO_2^+	A	$\{r, f_{\mathrm{CH}}(\theta)\}$	[58]
BiH_3	B	$\{f_{\mathrm{M}}(r), \theta\}$	[60]
NF_3	B	$\{r, \theta\}$	[61]
NH_3	B	$\{f_{\mathrm{M}}(r), \theta\}$	[62]
PH_3	B	$\{f_{\mathrm{M}}(r), \theta\}$	[63]
SbH_3	B	$\{f_{\mathrm{M}}(r), \theta\}$	[60]
SiH_3^-	B	$\{r, \theta\}$	[64]
AlF_3	C	$\{r, \theta, \sin(\omega)\}$	[65]
BF_3	C	$\{r, \theta, \sin(\omega)\}$	[66]
CF_3^+	C	$\{r, \theta, \sin(\omega)\}$	[66]
SiF_3^+	C	$\{r, \theta, \sin(\omega)\}$	[65]
SO_3	C	$\{r, \theta, \sin(\omega')\}$	[67]
H_2CO	C	$\{r, \theta, \tau\}$	[68]
H_2SiO	C	$\{f_{\mathrm{SPF}}(r), \theta, \tau\}$	[69]
N_2H_2	D	$\{r, \theta, \tau\}$	[70]
H_2O_2	D	$\{f_{\mathrm{SPF}}(r), \theta, \tau\}$	[71]
HSOH	D	$\{f_{\mathrm{M}}(r), \theta, \tau\}$	[72]
cis-HSiOH	D	$\{r, \theta, \tau\}$	[73]

Continued on next page

Table 2.3 – *Continued from previous page*

Molecule	**Geometry**	**Internals**	**References**
trans-HSiOH	D	$\{r, \theta, \tau\}$	[73]
cis-HOCO	D	$\{r, \theta, \tau\}$	[74]
trans-HOCO	D	$\{r, \theta, \tau\}$	[75]
C_4	E	$\{r, \theta, \tau\}$	[76]
CF_4	F	$\{r, \theta\}$	[77]
CH_4	F	$\{r, \theta\}$	[78]
NH_4^+	F	$\{r, \theta\}$	[79]
SiF_4	F	$\{r, \theta\}$	[77]
SiH_4	F	$\{r, \theta\}$	[80]
SnH_4	F	$\{r, \theta\}$	[81]
$FClO_3$	F	$\{r, \theta\}$	[82]
OPH_3	F	$\{r, \theta\}$	[83]
SPH_3	F	$\{r, \theta\}$	[83]
C_3H_2	G	$\{r, \theta, \tau\}$	[84]
$C_3H_3^+$	H	$\{f_M(r), \cos(\theta), \sin(\omega)\}$	[85]
C_2H_4	I	$\{f_M(r), \theta, \tau\}$	[86]

2.2.5 Testing

As noted above, a range of methods are available within PyPES for evaluating derivatives. The correct implementation of each method was verified by cross-checking between methods, and also by comparison with derivatives obtained through numerical differentiation.

To verify that the PES' within the PyPES library are implemented correctly, fundamental vibrational transition frequencies are calculated and compared to values from original publications. We used the vibrational configuration interaction (VCI) method, based upon the Watson operator in normal mode coordinates with zero rotational angular momentum, including Coriolis coupling through the leading term in the RSPT1 expansion of the vibrational angular momentum operator[87,88], and

ignoring the Watson correction term.[89] The wavefunction is expanded in a basis of harmonic oscillator functions in normal mode coordinates and all integrals are evaluated analytically. The potential is expanded in normal mode coordinates up to 6^{th} order in full form with no reduction in mode-mode coupling. We have chosen this approach as the simplest 'black-box' variational method likely to be extensible to larger molecules yet accurate enough for benchmarking purposes. However, we note that the choice to represent the PES as a 6^{th} order expansion in normal mode coordinates is likely to limit the accuracy of the calculated fundamental frequencies, particularly for high frequency stretching modes.

All configurations with a sum of vibrational quantum numbers up to a specified value are included in the VCI matrix. This excitation level is increased until all fundamental frequencies are converged to within 1 cm^{-1}. Data illustrating convergence with respect to excitation level are provided as Supporting Information. Full details of our VCI implementation will be discussed in Chapter 3.

To ensure correct representation of PES' up to energy regimes important for VCI calculations of fundamental frequencies, all PES' originally formulated as quartic force fields in 'fundamental' internal coordinates are transformed to a coordinate set with appropriate asymptotic behaviour. In particular, r is transformed to $f_{\text{M}}(r)$, θ to $\cos(\theta)$, and τ to $\sin(\tau)$. The required Morse parameter is estimated using the procedure recently reviewed by Fortenberry *et al.*[45].

An exception to the above rule is made for C_3H_2, which instead requires transformation to the alternative $\{f_M(r), \theta, \tau\}$ coordinate system to avoid the VCI frequencies diverging at high excitation levels, due to unphysically low angle bending and torsional rotation barriers in the $\{f_M(r), \cos(\theta), \sin(\tau)\}$-transformed potential, and incorrect asymptotic behaviour in the torsional coordinate, resulting in a spurious additional minimum. Slices of the PES along these coordinates are provided in the Supporting Information.

Derivative data required for Taylor series expansion of the PES in asympotically correct internal coordinates and normal mode coordinates are all calculated at the equilibrium geometry and stored as part of the library for each molecule.

2.3 Results

Converged VCI fundamental frequencies are summarised in Table 2.4, and compared to values from original publications. Degenerate modes are not repeated and all fundamentals are listed in ascending order (based on harmonic frequencies), with no regard for symmetry. The full set of results with all excitation levels and degenerate modes is included as Supporting Information. The fundamentals agree with reported literature values to within 3 cm^{-1} for most molecules, except as discussed below.

Table 2.4: Summary of VCI results from this work compared to reported fundamental vibrational frequencies from original publications. Fundamentals were ordered by frequency and only unique values were included. D is used to denote deuterated hydrogen. Values in brackets for NH_3 and D_2O_2 correspond to splitting due to low barrier inversion modes. The full set of results demonstrating convergence with respect to excitation level and symmetry preservation in symmetry-equivalent modes, is attached as Supporting Information. The labels (a), (b) and (c) refer to the results section, where explanations for significant deviations from literature values are given. Values are given in cm^{-1}.

Molecule	Current	Literature	Molecule	Current	Literature	Molecule	Current	Literature
H_2O	1596.5	1596.3	HO_2^+	1059.8	1058.4	ClO_2	449.9	449.9
(b)	3660.7	3656.1	(X^3A'')	1379.2	1379.0	(X^2B_1)	940.7	940.7
	3759.4	3753.4	(b)	3027.0	3021.7		1105.5	1105.5
HOF	897.7	898.0	HO_2^+	1373.0	1372.9	ClO_2	280.5	280.5
	1359.8	1360.0	(A^1A')	1443.7	1443.2	(A^2A_2)	436.7	436.5
	3602.0	3600.0	(b)	2964.9	2959.8		698.6	698.4
HOCl	724.7	724.7	ClO_2^+	496.2	496.1	BrO_2	316.1	316.1
	1239.4	1238.1		1004.8	1005.1	(X^2B_1)	794.7	794.7
	3615.5	3614.0		1271.0	1271.7		856.4	856.4
HOBr	624.0	624.0	PF_2^+	411.3	411.3	BrO_2	208.2	208.2
	1166.0	1164.0		1017.9	1017.8	(A^2A_2)	453.6	453.6
	3622.2	3621.1		1058.4	1058.4		631.6	631.6
F_2O	465.4	465.0	NH_2^-	1447.8	1447.8	SO_2	515.4	515.6
	845.1	845.0	(b)	3122.2	3118.5		1146.2	1146.3
	936.7	936.0		3191.8	3186.5		1349.6	1349.7
SiF_2	343.6	343.6	PO_2^-	460.3	460.3			

Continued on next page

Table 2.4 – *Continued from previous page*

Molecule	Current	Literature	Molecule	Current	Literature	Molecule	Current	Literature
	857.6	857.5		1059.6	1059.6			
	873.2	873.1		1212.5	1212.6			
NH_3	951.5	934.2	NF_3	490.5	491.0	CF_3^+	592.9	592.6
(c)		(969.7)		645.8	645.9		812.4	812.6
	1628.4	1628.8		866.7	866.7		1044.8	1044.5
		(1630.0)		1018.5	1018.6		1683.6	1682.8
	3341.4	3342.2						
		(3343.2)						
	3449.9	3449.3						
		(3449.7)						
PH_3	991.8	991.9	BF_3	480.8	480.6	SiF_3^+	307.1	307.2
	1118.8	1118.9		695.8	696.2		356.7	356.7
	2322.7	2321.0		887.8	887.6		852.8	853.3
	2328.0	2325.8		1470.2	1469.6		1187.2	1187.9
SbH_3	799.0	798.9	AlF_3	241.3	241.4	SiH_3^-	843.2	844.1
	836.8	836.8		301.2	301.2		939.9	937.8
	1894.1	1893.8		689.0	689.5		1824.7	1821.5
	1899.8	1899.1		951.2	951.9		1840.0	1840.7
BiH_3	733.9	733.9	SO_3	496.5	498.6	C_4	303.2	300.5
	759.3	759.5		528.3	528.1		522.9	520.4
	1742.6	1742.4		1067.0	1067.0		931.8	928.2
	1746.8	1746.3		1396.4	1396.3		1003.1	1002.0
							1256.9	1256.1
							1318.3	1316.0
N_2H_2	1291.8	1294.2	H_2CO	1165.7	1166.1	H_2SiO	680.3	680.1
(a)	1317.4	1317.5		1245.7	1245.6		691.0	690.9
	1519.7	1519.3		1499.2	1499.1		994.4	994.3
	1579.7	1579.4		1744.7	1744.6		1207.0	1206.9
	3037.8	3033.3		2781.8	2781.7		2171.2	2171.0
	3115.5	3125.0		2843.0	2842.4		2191.7	2191.3
D_2O_2	273.5	210.1	*cis*-	469.5	446.9	*cis*-	458.9	450.2
(b),(c)		(251.9)	DOCO	541.6	539.8	DSiOD	529.3	523.4
	870.2	869.3	(b)	968.2	960.9	(a),(b)	721.8	718.2
	952.1	945.4		1123.5	1123.1		838.7	838.4
	1026.1	1026.2		1827.2	1827.5		1375.5	1372.8
	2664.1	2666.6		2550.3	2551.6		2708.9	2713.1

Continued on next page

Table 2.4 – *Continued from previous page*

Molecule	Current	Literature	Molecule	Current	Literature	Molecule	Current	Literature
	2663.4	2667.2						
HSOH	484.5	443.0	*trans-*	402.7	368.0	*trans-*	476.9	469.2
(b)	763.0	760.0	DOCO	593.3	590.1	DSiOD	576.6	573.4
	999.1	1007.7	(b)	906.3	902.6	(a),(b)	708.9	705.3
	1179.9	1174.0		1088.2	1086.4		835.2	834.7
	2542.6	2544.4		1859.6	1859.8		1426.1	1423.6
	3608.4	3625.9		2686.9	2685.1		2706.5	2709.7
CH_4	1309.0	1312.7	SiH_4	912.3	915.1	SnH_4	683.2	681.3
(a)	1529.5	1533.1	(a)	968.9	968.3	(a)	754.3	753.6
	2914.6	2911.1		2187.2	2185.0		1904.7	1901.4
	3015.1	3011.6		2190.0	2185.2		1910.1	1909.9
CF_4	435.4	435.4	SiF_4	264.2	264.2	NH_4^+	1442.9	1446.2
	631.2	631.1		388.4	388.4	(a)	1687.6	1690.8
	909.0	909.1		800.9	800.6		3235.7	3230.9
	1284.4	1283.7		1031.5	1031.4		3343.2	3339.1
OPH_3	846.6	846.6	SPH_3	677.1	676.7	$FClO_3$	404.3	405.8
(a)	1116.2	1117.1		723.8	723.8	(a)	552.9	552.8
	1146.8	1146.7		1100.7	1102.2		582.8	590.7
	1263.7	1262.6		1118.0	1118.0		730.2	730.3
	2354.2	2353.1		2348.6	2348.5		1074.5	1074.5
	2360.7	2356.4		2366.1	2366.2		1336.1	1335.1
C_3H_2	776.7	771.1	$C_3H_3^+$	758.9	757.1	C_2H_4	823.7	822.4
(a)	884.2	879.1		927.7	927.0		935.3	934.3
	887.8	882.8		1003.0	1002.0		950.7	949.5
	973.1	971.3		1041.3	1040.6		1025.9	1024.9
	1057.1	1056.2		1295.5	1296.2		1224.9	1224.3
	1274.5	1271.0		1621.1	1622.1		1342.9	1342.5
	1592.1	1591.4		3130.8	3134.8		1441.8	1441.1
	3111.7	3115.3		3172.7	3175.4		1625.4	1624.4
	3133.0	3134.3					2985.5	2985.4
							3019.2	3019.0
							3079.4	3079.9
							3101.3	3101.7

For the purpose of PES verification, disagreements of this magnitude can be

considered acceptable, as this lies within the error margins arising from the implementation of slightly different VCI algorithms, potentially varying in approach to generating VCI excitations, choice of Hamiltonian operator, and PES representation.

There are three main reasons for larger disagreements:

(a) the difference in the method used to solve the nuclear vibrational problem *e.g.* VPT2 vs. VCI;

(b) not being able to accurately describe the relevant energy space using a 6^{th} order expansion of the potential in normal mode coordinates;

(c) inadequacy of the Watson Hamiltonian for describing systems with multiple minima.

Differences due to (a) arise when comparing our VCI results with original results obtained using low order vibrational perturbation theory, which is non-variational. However, fundamentals are within 10 cm^{-1} in all cases, and we observe full convergence with respect to excitation level. An average deviation of 5 cm^{-1} between VPT2 and VCI results has been observed in the literature[45,69,74,75,90,91], with larger deviations possible for highly anharmonic and less rigidly bound systems[92]. In the current work, the N_2H_2 PES contains the largest difference between VCI results and literature VPT2 results, of 10 cm^{-1} in the asymmetric NH stretching frequency. Therefore, we independently verified the implementation of this surface by comparing our values for $\left[\frac{\partial V}{\partial \mathbf{X}}\right]$ with those supplied in the original publication.[70]

There are two groups of cases where 6^{th} order expansion of the PES in normal mode coordinates is inadequate, leading to reason (b).

For H_2O, NH_2^-, and both electronic states of HO_2^+, the high frequency stretching fundamentals are overestimated by up to 6 cm^{-1}, reflecting the artificially steep walls imposed by the truncated Taylor series expansion in normal mode coordinates. Therefore, it is most likely that the PES is correctly implemented and this small error only reflects the nature of the PES representation chosen for use in our VCI implementation.

For tetra-atomics with low frequency torsional modes (H_2O_2, HSOH, cis- and trans- HSiOH, cis- and trans- HOCO), the problem is more pronounced. In these

cases, incorrect description of the potential along the torsional coordinates leads to significant overestimation of the torsional frequencies, and strong coupling of torsional coordinates with OH stretching leads to divergence in fundamental OH stretching frequencies with respect to excitation level. Fortunately, potential energy surfaces for some molecules in this class can be independently verified. The cis- and trans- HSiOH surfaces were validated by reproduction of Cartesian derivatives that were included with the paper describing the original PES implementation,[73] while HSOH was tested against an energy map provided in the literature by the original authors of the PES.[72]

Deuteration provides an alternative strategy for assessing the implemented surfaces in a lower energy regime, to reduce the effect of errors arising from inappropriate asymptotic behaviour of the potential energy surface expansion. Therefore, we have calculated vibrational frequencies for deuterated isotopologues of H_2O_2, cis- and trans- HSiOH, and cis- and trans- HOCO. In all cases, the VCI wavefunction converges with respect to excitation level, and agreement with the published fundamental frequencies[71,73–75,93] improves. This consistent pattern of behaviour across this class of molecules, combined with independent verification of the cis- and trans- HSiOH and HSOH surfaces, gives us confidence that these PES' are correctly implemented within PyPES.

Ammonia (NH_3) and hydrogen peroxide (H_2O_2) have low barrier inversion modes, with symmetry-equivalent vibrationally accessible minima leading to energy level tunneling splitting that cannot be properly captured in our treatment. For NH_3, the inversion mode does not strongly couple to the other modes, so we are able to reproduce the frequencies of all other modes to within 2 cm^{-1} of both split levels, and observe full convergence in the VCI wavefunction with respect to excitation level. For H_2O_2 and D_2O_2 the error in the fundamental frequencies is dominated by the inadequacy of the normal mode PES expansion for describing the PES along the torsional coordinate, as discussed above. Therefore, we again conclude that the deviation from published results arises from use of the Watson Hamiltonian in conjunction with normal mode coordinates for expanding the potential and solving the VCI problem, rather than errors in our implementation of the potential energy surfaces themselves.

2.4 Conclusions

The PyPES library provides a rigorously tested implementation of 50 highly accurate semi-global potential energy surfaces for a range of small molecules with a wide variety of geometries and bonding patterns. We anticipate that this library will find widespread use in benchmarking new methods for solving the nuclear Schrödinger equation, particularly methods designed to scale to larger molecules and systems. For maximum user-friendliness, we provide a series of wrappers that allow the user to treat PyPES as a 'black-box' tool for obtaining accurate energies and derivatives of the energy to 6^{th} order with respect to internal coordinates, Cartesian coordinates and normal mode coordinates at arbitrary geometries.

PyPES also provides a platform for the implementation and distribution of analytic potential energy surfaces. Most common coordinates are implemented within PyPES. Force fields formulated as Taylor series expansions up to 6^{th} order in implemented internal coordinates are readily incorporated by providing the appropriate force constants, and PES' that are more complicated functions of internal coordinates can be readily differentiated using SymPy. PyPES also provides a powerful customized symbolic differentiation algorithm to facilitate implementation of new coordinates if and when required.

The PyPES library is available for free download from http://sourceforge.net/projects/pypeslib.

2.4.1 Acknowledgements

Thanks to X. Huang for providing the original code implementing the $C_3H_3^+$ force field. We would also like to thank R.C. Fortenberry and T.J. Lee for helpful discussions.

References

[1] G. Chaban, J. Jung, R. Gerber, *J. Chem. Phys.* **1999**, *111*, 1823–1829.

[2] N. Matsunaga, G. M. Chaban, R. B. Gerber, *J. Chem. Phys.* **2002**, *117*, 3541–3547.

[3] K. Yagi, S. Hirata, K. Hirao, *Phys. Chem. Chem. Phys.* **2008**, *10*, 1781–1788.

[4] D. Strobusch, C. Scheurer, *J. Chem. Phys.* **2011**, *135*, 124102.

[5] L. Norris, M. Ratner, A. Roitberg, R. Gerber, *J. Chem. Phys.* **1996**, *105*, 11261–11267.

[6] O. Christiansen, *J. Chem. Phys.* **2003**, *119*, 5773–5781.

[7] K. Yagi, S. Hirata, K. Hirao, *J. Chem. Phys.* **2007**, *127*, 034111.

[8] O. Christiansen, *J. Chem. Phys.* **2004**, *120*, 2149–2159.

[9] O. Christiansen, *J. Chem. Phys* **2004**, *120*, 2140–2148.

[10] C. König, O. Christiansen, *J. Chem. Phys.* **2015**, *142*, 144115.

[11] S. Carter, J. M. Bowman, N. C. Handy, *Theor. Chem. Acc.* **1998**, *100*, 191–198.

[12] P. Cassam-Chenai, J. Liévin, *Int. J. Quant. Chem.* **2003**, *93*, 245–264.

[13] M. Neff, G. Rauhut, *J. Chem. Phys.* **2009**, *131*, 124129.

[14] D. Bégué, N. Gohaud, C. Pouchan, P. Cassam-Chenai, J. Liévin, *J. Chem. Phys.* **2007**, *127*, 164115.

[15] Y. Scribano, D. M. Benoit, *Chem. Phys. Lett.* **2008**, *458*, 384–387.

[16] Y. Scribano, D. M. Lauvergnat, D. M. Benoit, *J. Chem. Phys.* **2010**, *133*, 094103.

[17] A. Samsonyuk, C. Scheurer, *J. Comput. Chem.* **2013**, *34*, 27–37.

[18] G. Rauhut, *J. Chem. Phys.* **2007**, *127*, 184109.

[19] P. Carbonniere, A. Dargelos, C. Pouchan, *Theor. Chem. Acc.* **2010**, *125*, 543–554.

[20] D. Strobusch, C. Scheurer, *J. Chem. Phys.* **2011**, *135*, 144101.

[21] F. Pfeiffer, G. Rauhut, *J. Chem. Phys.* **2014**, *140*, 064110.

[22] C. Lin, A. Gilbert, P. Gill, *Theor. Chem. Acc.* **2008**, *120*, 23–35.

[23] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Phys. Chem. A* **2012**, *116*, 3691–3709.

[24] J. Bloino, V. Barone, *J. Chem. Phys.* **2012**, *136*, 124108.

[25] M. Sparta, D. Toffoli, O. Christiansen, *Theor. Chem. Acc.* **2009**, *123*, 1–17.

[26] E. Matito, D. Toffoli, O. Christiansen, *J. Chem. Phys.* **2009**, *130*, 134104.

[27] P. Seidler, T. Kaga, K. Yagi, O. Christiansen, K. Hirao, *Chem. Phys. Lett.* **2009**, *483*, 138–142.

[28] S. Hirata, M. Keçeli, K. Yagi, *J. Chem. Phys.* **2012**, *137*, 204118.

[29] X. Cheng, R. P. Steele, *J. Chem. Phys.* **2014**, *141*.

[30] B. Thomsen, K. Yagi, O. Christiansen, *Chem. Phys. Lett.* **2014**, *610*, 288–297.

[31] B. Thomsen, K. Yagi, O. Christiansen, *J. Chem. Phys.* **2014**, *140*, 154102.

[32] O. Christiansen, *Phys. Chem. Chem. Phys.* **2007**, *9*, 2942–2953.

[33] O. Christiansen, *Phys. Chem. Chem. Phys.* **2012**, *14*, 6672–6687.

[34] S. Hirata, K. Yagi, *Chem. Phys. Lett.* **2008**, *464*, 123–134.

[35] D. Oschetzki, M. Neff, P. Meier, F. Pfeiffer, G. Rauhut, *Croat. Chem. Acta* **2012**, *85*, 379–390.

[36] A. G. Császár, C. Fabri, T. Szidarovszky, E. Matyus, T. Furtenbacher, G. Czako, *Phys. Chem. Chem. Phys.* **2012**, *14*, 1085–1106.

[37] J. M. Bowman, T. Carrington, H.-D. Meyer, *Mol. Phys.* **2008**, *106*, 2145–2182.

[38] C. E. Dateo, T. J. Lee, D. W. Schwenke, *J. Chem. Phys.* **1994**, *101*, 5853–5859.

[39] G. Simons, R. G. Parr, M. J. Finlan, *J. Chem. Phys.* **1973**, *59*, 3229–3234.

[40] S. Carter, N. C. Handy, *J. Chem. Phys.* **1987**, *87*, 4294–4301.

[41] W. D. Allen, A. G. Császár, *J. Chem. Phys.* **1993**, *98*, 2983–3015.

[42] W. D. Allen, A. G. Császár, V. Szalay, I. M. Mills, *Mol. Phys.* **1996**, *89*, 1213–1221.

[43] A. G. Császár in *Encyclopedia of Computational Chemistry*, *Vol. 1*, (Ed.: P. Schleyer), Wiley, Chichester, **1998**, p. 13.

[44] A. G. Császár, *WIRES Comput. Mol. Sci.* **2012**, *2*, 273–289.

[45] R. C. Fortenberry, X. Huang, A. Yachmenev, W. Thiel, T. J. Lee, *Chem. Phys. Lett.* **2013**, *574*, 1–12.

[46] R. E. Tuzun, D. W. Noid, B. C. Sumpter, *Macromol. Theory Simul.* **1996**, *5*, 771–788.

[47] A. R. Hoy, I. M. Mills, G. Strey, *Mol. Phys.* **1972**, *24*, 1265–1290.

[48] T. Schlick, *J. Comput. Chem.* **1989**, *10*, 951–956.

[49] K. J. Miller, R. J. Hinde, J. Anderson, *J. Comput. Chem.* **1989**, *10*, 63–76.

[50] D. Noid, B. Sumpter, B. Wunderlich, G. Pfeffer, *J. Comput. Chem.* **1990**, *11*, 236–241.

[51] H. Bekker, H. Berendsen, W. van Gunsteren, *J. Comput. Chem.* **1995**, *16*, 527–533.

[52] S.-H. Lee, K. Palmo, S. Krimm, *J. Comput. Chem.* **1999**, *20*, 1067.

[53] R. E. Tuzun, D. W. Noid, B. G. Sumpter, *J. Comput. Chem.* **2000**, *21*, 553–561.

[54] X. Huang, T. J. Lee, *J. Chem. Phys.* **2008**, *129*, 044312.

[55] X. Huang, T. J. Lee, *J. Chem. Phys.* **2009**, *131*, 104301.

[56] K. A. Peterson, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **1997**, *53*, 1051–1064.

[57] J. Breidung, W. Thiel, J. Gauss, J. F. Stanton, *J. Chem. Phys.* **1999**, *110*, 3687.

[58] Y. Pak, R. C. Woods, *J. Chem. Phys.* **1996**, *104*, 5547.

[59] K. A. Peterson, *J. Chem. Phys.* **1998**, *109*, 8864.

[60] S. N. Yurchenko, W. Thiel, P. Jensen, *J. Mol. Spectrosc.* **2006**, *240*, 174–187.

[61] R. Tarroni, P. Palmieri, M. L. Senent, A. Willetts, *Chem. Phys. Lett.* **1996**, *257*, 23–30.

[62] S. N. Yurchenko, J. Zheng, H. Lin, P. Jensen, W. Thiel, *J. Chem. Phys.* **2005**, *123*, 134308.

[63] R. I. Ovsyannikov, W. Thiel, S. N. Yurchenko, M. Carvajal, P. Jensen, *J. Chem. Phys.* **2008**, *129*, 044309.

[64] K. Aarset, A. G. Császár, E. L. Sibert III, W. D. Allen, H. F. Schaefer III, W. Klopper, J. Noga, *J. Chem. Phys.* **2000**, *112*, 4053–4063.

[65] Y. Pak, E. L. Sibert, R. C. Woods, *J. Chem. Phys.* **1997**, *107*, 1717–1724.

[66] Y. Pak, R. C. Woods, *J. Chem. Phys.* **1997**, *106*, 6424.

[67] J. Martin, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **1999**, *55*, 709–718.

[68] A. Yachmenev, S. N. Yurchenko, P. Jensen, W. Thiel, *J. Chem. Phys.* **2011**, *134*, 244307–244307.

[69] J. Koput, S. Carter, N. C. Handy, *Chem. Phys. Lett.* **1999**, *301*, 1–9.

[70] J. M. L. Martin, P. R. Taylor, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **1997**, *53*, 1039–1050.

[71] P. Malyszek, J. Koput, *J. Comput. Chem.* **2013**, *34*, 337–345.

[72] S. N. Yurchenko, A. Yachmenev, W. Thiel, O. Baum, T. F. Giesen, V. V. Melnikov, P. Jensen, *J. Mol. Spectrosc.* **2009**, *257*, 57–65.

[73] J. Martin, *J. Phys. Chem. A* **1998**, *102*, 1394–1404.

[74] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, *J. Chem. Phys.* **2011**, *135*, 214303.

[75] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, *J. Chem. Phys.* **2011**, *135*, 134301.

[76] X. Wang, X. Huang, J. M. Bowman, T. J. Lee, *J. Chem. Phys.* **2013**, *139*, 224302.

[77] X.-G. Wang, E. L. Sibert, J. M. L. Martin, *J. Chem. Phys.* **2000**, *112*, 1353–1366.

[78] T. J. Lee, J. M. L. Martin, P. R. Taylor, *J. Chem. Phys.* **1995**, *102*, 254–261.

[79] J. Martin, T. Lee, *Chem. Phys. Lett.* **1996**, *258*, 129–135.

[80] J. Martin, K. Baldridge, T. Lee, *Mol. Phys.* **1999**, *97*, 945–953.

[81] L. Fusina, G. Nivellini, T. Salzillo, M. Lamarra, R. Tarroni, *J. Chem. Phys.* **2012**, *137*, 204316.

[82] E. Cane, L. Fusina, M. Lamarra, R. Tarroni, K. Burczyk, *J. Phys. Chem. A* **2008**, *112*, 13729–13736.

[83] M. Lamarra, R. Tarroni, *Mol. Phys.* **2011**, *109*, 2095–2104.

[84] T. J. Lee, X. Huang, C. E. Dateo, *Mol. Phys.* **2009**, *107*, 1139–1152.

[85] X. Huang, P. R. Taylor, T. J. Lee, *J. Phys. Chem. A* **2011**, *115*, 5005–5016.

[86] T. Delahaye, A. Nikitin, M. Rey, P. Szalay, V. Tyuterev, *J. Chem. Phys.* **2014**, *141*, 104301.

[87] V. Barone, *J. Chem. Phys.* **2005**, *122*, 014108.

[88] M. Neff, T. Hrenar, D. Oschetzki, G. Rauhut, *J. Chem. Phys.* **2011**, *134*, 064105.

[89] J. Vázquez, M. E. Harding, J. F. Stanton, J. Gauss, *J. Chem. Theory Comput.* **2011**, *7*, 1428–1442.

[90] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, *J. Chem. Phys* **2012**, *136*, 234309.

[91] N. Inostroza, X. Huang, T. Lee, *J. Chem. Phys* **2011**, *135*, 244310.

[92] W. J. Morgan, R. C. Fortenberry, *Spectrochim. Acta Part A: Mol. Biomol. Spectrosc.* **2015**, *135*, 965–972.

[93] X. Huang, R. C. Fortenberry, Y. Wang, J. S. Francisco, T. D. Crawford, J. M. Bowman, T. J. Lee, *J. Phys. Chem. A* **2013**, *117*, 6932–6939.

Chapter 3

PyVCI: a flexible open-source code for calculating accurate molecular infrared spectra

3.1 Introduction

The simulation of accurate molecular vibrational spectra has historically been limited by the difficulty and computational cost associated with modelling how the energy changes as the molecule vibrates, *i.e.* constructing multidimensional anharmonic potential energy surfaces (PES).

High quality *semi-global* potential energy surfaces for small molecules are available in the literature, and a number of these have been compiled into PES libraries.[1–9] However, the scalability of this approach is primarily limited by the need to construct an appropriate curvilinear internal coordinate set in which to represent the PES that appropriately accounts for molecular symmetry. Choosing and parameterising appropriate PES functional forms is also non-trivial.

A more pragmatic approach is to focus on simulating only the fundamental vibrational transitions required to model experimental infrared spectra. Quantitative assignment of IR spectra is an important and longstanding problem of widespread interest within the general chemistry community. Predicting fundamental frequencies requires only the low energy region of the PES in the vicinity of the minimum

to be accurately described. This can be achieved using a *local* expansion of the potential energy surface about equilibrium, which does not necessarily need to be formulated in internal coordinates.

Recent hardware and software advances[4,10–19] now enable quartic force field expansions in orthonormal rectilinear coordinate sets to be routinely generated for larger molecules, in a straightforward although time-consuming manner. This makes simulating fundamental modes of infrared spectra possible for a larger range of chemically interesting molecules.

A recent review by Roy and Gerber[20] provides a comprehensive overview of methods based upon expanding the nuclear vibrational wavefunction in terms of products of single-mode functions in normal mode coordinates. Formulating the nuclear vibrational Schrödinger equation in normal mode coordinates confers two major advantages; separability of the kinetic energy operator and potential energy integrals that can be evaluated analytically.

A hierarchy of approximations yield a series of methods including normal mode analysis (NMA)[21], vibrational self-consistent field theory (VSCF)[18,22–28], vibrational perturbation theory (VPT)[28–40], vibrational configuration interaction (VCI)[28,41,42] and vibrational coupled cluster theory (VCC)[28,43–46]. Analogous to their electronic structure theory counterparts, the quality of each method is generally commensurate with its computational cost.[16,28,46]

Despite the extensive efforts that have gone into developing anharmonic nuclear vibrational structure theories,[28–46] they remain under-utilized within the wider computational chemistry community. This can be traced back to a number of factors including: lack of general availability of anharmonic vibrational structure codes[26,47]; the ubiquity, ease of use and relatively low computational cost of harmonic normal mode analysis; experimental reference data biased by interactions with solvent molecules limiting the utility of highly accurate gas phase vibrational structure models, and; the need to customize anharmonic nuclear vibrational theories for larger molecules to make them computationally tractable.

In this paper, we primarily address the first of these issues. Although some nuclear vibrational structure methods are included in some quantum chemical soft-

ware packages, they vary in both how the PES is represented and how the nuclear vibrational problem is solved. This makes comparing results between different programs both difficult and time-consuming. To ensure reproducibility, it is necessary to be able to specify and/or control both the PES representation and the nuclear vibrational algorithm.

Therefore, we present the PyVCI package, a general and open-source vibrational configuration interaction code in which the potential energy surface is represented as a Taylor series expansion up to 6th order in normal mode coordinates. PyVCI can import force field data from the PyPES library of analytical potential energy surfaces, or generate quartic force fields by numerical differentiation of second derivative data obtained from *ab initio* calculations. Currently, only the GAMESS quantum chemistry program package is supported. Although the VCI method we have implemented within PyVCI is general, it is not completely unique; other variants are possible through different algorithmic choices and alternative approaches to truncating the full ro-vibrational Hamiltonian. Therefore, in this paper we provide complete details of our VCI implementation before benchmarking its performance using the PyPES library of potential energy surfaces.[1,2]

3.2 VCI theory and algorithm

Nuclear vibrational structure theories are defined and differentiated by the form of the Hamiltonian operator and representation of the wavefunction. In the interests of computational efficiency, we employ the Watson Hamiltonian and expand both the wavefunction and potential energy surface in terms of normal mode coordinates about the global minimum. Coriolis rotational coupling terms may be optionally included:

$$\hat{H} = \hat{H}_{\text{vib}} + \hat{H}_{\text{Cor}} \tag{3.1}$$

$$\hat{H}_{\text{vib}} = -\frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial Q_i^2} + V(Q_1, ..., Q_M) \tag{3.2}$$

$$\hat{H}_{\text{Cor}} = -\sum_{\alpha} B_\alpha \sum_{i<j}\sum_{k<l} \zeta_{ij}^{\alpha}\zeta_{kl}^{\alpha}\left(Q_i\frac{\partial}{\partial Q_j} - Q_j\frac{\partial}{\partial Q_i}\right)\left(Q_k\frac{\partial}{\partial Q_l} - Q_l\frac{\partial}{\partial Q_k}\right) \tag{3.3}$$

The kinetic energy operator is separable in normal mode coordinates, and the potential energy surface is given as a Taylor series expansion:

$$\begin{aligned} V =& V_{\text{ref}} + \sum_i F_i Q_i + \frac{1}{2!}\sum_{i,j} F_{ij} Q_i Q_j + \frac{1}{3!}\sum_{i,j,k} F_{ijk} Q_i Q_j Q_k + \\ & \frac{1}{4!}\sum_{i,j,k,l} F_{ijklmn} Q_i Q_j Q_k Q_l + \frac{1}{5!}\sum_{i,j,k,l,m} F_{ijklm} Q_i Q_j Q_k Q_l Q_m + \\ & \frac{1}{6!}\sum_{i,j,k,l,m,n} F_{ijklmn} Q_i Q_j Q_k Q_l Q_m Q_n + \ldots \end{aligned} \tag{3.4}$$

where the summation indices run from 1 through to the number of vibrational modes. The force constants, F, are the derivatives of the potential with respect to the normal coordinates. The normal coordinates are defined as linear combinations of Cartesian displacements that diagonalize the Hessian in mass-weighted Cartesian coordinates.[21] This produces a coordinate system in which the first order and off-diagonal second order force constants are zero at equilibrium. For completeness, the details of our implementation are provided in Appendix A.

This approach yields the diagonal second order force constants, but the higher order derivatives are usually calculated *via* numerical differentiation. For this reason, the PES expansion is usually truncated at fourth order, to keep the computational cost of generating the force field manageable. However, we have implemented a library of analytical potential energy surfaces for benchmarking and testing, so also employ sextic force field expansions for these molecules.

Equilibrium rotational and Coriolis coupling constants, B_α and ζ_{ij}^{α}, about each principal axis, α, are required to calculate the overall Coriolis coupling. Rotational constants are obtained by diagonalising the inertia tensor[21] and $\boldsymbol{\zeta}$ matrices are calculated according to the method of Meal and Polo.[48,49] These algorithms are detailed in Appendix B.

For computational expedience, we construct the VCI wavefunction from Hartree products of harmonic oscillator basis functions:

$$\Phi_{\mathbf{n}}\left(Q_1, ..., Q_M\right) = \prod_{i=1}^{M} \phi_{n_i}(Q_i) \tag{3.5}$$

where $\mathbf{n}$ is a string of quantum numbers $n_1, ..., n_i, ...n_M$, specifying the vibrational state across all M modes. The strings that define the VCI basis states are generated

Table 3.1: Fundamental Coriolis coupling integrals for vibrational mode i, $\langle\phi_{n_i'}(Q_i)|\,\hat{O}\,|\phi_{n_i}(Q_i)\rangle$ involving operators $\hat{O} = \dfrac{\partial}{\partial Q_i}, Q_i\dfrac{\partial}{\partial Q_i}, \dfrac{\partial}{\partial Q_i}Q_i$

$n_i - n_i'$	$\frac{\partial}{\partial Q_i}$	$Q_i\frac{\partial}{\partial Q_i}$	$\frac{\partial}{\partial Q_i}Q_i$
-2	0	$-\frac{\sqrt{n'(n'-1)}}{2}$	$-\frac{\sqrt{n'(n'-1)}}{2}$
-1	$-\sqrt{\frac{n'}{2\omega_i}}$	0	0
0	0	$-\frac{1}{2}$	$\frac{1}{2}$
1	$\sqrt{\frac{n}{2\omega_i}}$	0	0
2	0	$\frac{\sqrt{n(n-1)}}{2}$	$\frac{\sqrt{n(n-1)}}{2}$

by specifying a maximum value for the sum of the vibrational quantum numbers, which will henceforth be referred to as the excitation level, with its value denoted in round brackets, *e.g.* VCI(8) matrix is indexed by configurations with a sum of vibrational quantum numbers of 8 or less.

Finally, it remains to evaluate and store the VCI matrix elements:

$$\langle\Phi_{\mathbf{n}}\,(Q_1,...,Q_M)|\,\hat{H}_{\text{vib}} + \hat{H}_{\text{Cor}}\,|\Phi_{\mathbf{n}'}\,(Q_1,...,Q_M)\rangle \tag{3.6}$$

for all unique combinations of Hartree product basis functions, then diagonalise the VCI matrix. The fundamental integrals required to evaluate both the Coriolis and vibrational correction terms are given in Tables 3.1 and 3.2. Sparse matrix array structures are used to store the VCI matrix elements.

The final VCI wavefunction for each state $\Psi_{\mathbf{n}}\,(Q_1,...,Q_M)$ is a linear combination of Hartree product basis states $\Phi_{\mathbf{n}'}\,(Q_1,...,Q_M)$. The coefficients, $c_{\mathbf{n},\mathbf{n}'}$, are the eigenvectors of the VCI matrix, and the energy levels, $\epsilon_{\mathbf{n}}$, its eigenvalues. These solutions are generated using the sparse matrix diagonalization routines implemented in SciPy, to minimize memory and CPU time requirements.

$$\Psi_{\mathbf{n}}\,(Q_1,...,Q_M) = \sum_{n'} c_{\mathbf{n},\mathbf{n}'}\Phi_{\mathbf{n}'}\,(Q_1,...,Q_M) \tag{3.7}$$

The VCI fundamental frequencies are identified according to the extent of wavefunction overlap with Hartree product singly excited basis states. Leading coefficients in the VCI wavefunction for all states with frequencies below 4000 cm^{-1} are printed by default and may be used to resolve ambiguities in state assignments when they arise.

Table 3.2: Fundamental kinetic and potential energy integrals for vibrational mode i, $\langle\phi_{n_i'}(Q_i)|\,\hat{O}\,|\phi_{n_i}(Q_i)\rangle$ involving operators $\hat{O} = \frac{\partial^2}{\partial Q_i^2}, Q_i, Q_i^2, Q_i^3, Q_i^4, Q_i^5, Q_i^6$. Only $n_i > n_i'$ cases are shown as all operators are Hermitian.

$n_i - n_i'$	$\frac{\partial^2}{\partial Q_i^2}$	Q_i	Q_i^2	Q_i^3	Q_i^4	Q_i^5	Q_i^6
0	$-\omega_i(n+\frac{1}{2})$	0	$\frac{(2n+1)}{2\omega_i}$	0	$\frac{(6n^2+6n+3)}{4\omega_i^2}$	0	$\frac{5(4n^3+6n^2+8n+3)}{8\omega_i^3}$
1	0	$\frac{\sqrt{n}}{(2\omega_i)^{1/2}}$	0	$\frac{3n\sqrt{n}}{(2\omega_i)^{3/2}}$	0	$\frac{5(2n^2+1)\sqrt{n}}{(2\omega_i)^{5/2}}$	0
2	$\frac{\omega_i\sqrt{n(n-1)}}{2}$	0	$\frac{\sqrt{n(n-1)}}{2\omega_i}$	0	$\frac{2(2n-1)\sqrt{n(n-1)}}{4\omega_i^2}$	0	$\frac{15(n^2-n+1)\sqrt{n(n-1)}}{8\omega_i^3}$
3	0	0	0	$\frac{\sqrt{n(n-1)(n-2)}}{(2\omega_i)^{3/2}}$	0	$\frac{5(n-1)\sqrt{n(n-1)(n-2)}}{(2\omega_i)^{5/2}}$	0
4	0	0	0	0	$\frac{\sqrt{n(n-1)(n-2)(n-3)}}{4\omega_i^2}$	0	$\frac{3(2n-3)\sqrt{n(n-1)(n-2)(n-3)}}{8\omega_i^3}$
5	0	0	0	0	0	$\frac{\sqrt{n(n-1)(n-2)(n-3)(n-4)}}{(2\omega_i)^{5/2}}$	0
6	0	0	0	0	0	0	$\frac{\sqrt{n(n-1)(n-2)(n-3)(n-4)(n-5)}}{8\omega_i^3}$

3.3 Methods

Our test set comprises 44 polyatomic molecules containing up to 6 atoms for which accurate quartic and sextic force fields in normal mode coordinates are available through the PyPES library.[1,2] This test set is used to explore convergence of calculated fundamental frequencies with respect to:

- number of configurations in the VCI expansion,
- threshold for storage of non-negligible VCI matrix elements,
- threshold for inclusion of non-negligible force constants in the PES expansion, and
- order of the PES expansion.

A further 6 tetra-atomics with low barrier torsional modes are used to quantify the effect of excluding divergent modes from the VCI expansion *a priori*.

VCI calculations include all configurations with an 'excitation level' (specified sum of vibrational quantum numbers) up to a maximum of 10, denoted VCI(10). Coriolis coupling terms are included in the Hamiltonian throughout. The excitation level is increased until all fundamental frequencies are converged to within 1 cm^{-1} for each molecule. The screening threshold for storing non-negligible VCI matrix elements is set to 10^{-15} E_h for benchmark calculations and tested at a range of values between 1×10^{-7} and 5×10^{-5} E_h. The screening threshold for retaining non-negligible force constants in dimensionless normal mode coordinates is set to zero during benchmark calculations, then tested at values of 0.1, 0.5, 1 and 2 cm^{-1}. The VCI matrix is diagonalised using sparse matrix diagonalization routines implemented in SciPy.

Sextic force fields are used in all convergence and threshold testing calculations. Benchmark results for quartic force fields are also generated without screening.

Statistical data are summarized using box-and-whisker plots, with boxes extending one quartile in each direction from the median, and whiskers extending out by 1.5 $\times$ the interquartile range in each direction, or to the limits of the data, whichever

comes first. Any data points outside this range are considered outliers and marked using crosses. Frequency data are expressed in units of reciprocal centimeters (cm^{-1}) throughout.

The PyVCI program package may be freely downloaded from:
http://sourceforge.net/projects/PyVCI.

3.4 Results

3.4.1 VCI convergence with respect to excitation level

Reference results were generated using a sextic force field at VCI(10) for triatomics and C_3H_2, VCI(9) for all other 4 and 5 atom molecules, and VCI(8) for $C_3H_3^+$ and C_2H_4, to ensure convergence of all fundamental frequencies to within 1 cm^{-1}.

Errors in fundamental frequencies at each excitation level, n, are then calculated as:

$$\Delta^{\mathrm{SFF}}_{\mathrm{VCI}(n)} = \nu_{\mathrm{ref}} - \nu^{\mathrm{SFF}}_{\mathrm{VCI}(n)} \tag{3.8}$$

$$\Delta^{\mathrm{QFF}}_{\mathrm{VCI}(n)} = \nu_{\mathrm{ref}} - \nu^{\mathrm{QFF}}_{\mathrm{VCI}(n)} \tag{3.9}$$

Mean and maximum absolute errors across all fundamental frequencies of the 44 molecules with restricted torsional motion in the PyPES library, for each excitation level, are presented in Table 3.3.

For excitation levels lower than 4, errors are large and convergence behaviour erratic, for both QFF- and SFF-derived results. At these lower excitation levels, QFF and SFF force fields provide equally accurate descriptions of the PES, and therefore produce similar fundamental frequencies that have similar errors.

At higher excitation levels, differences between quartic and sextic force fields become apparent in the calculated VCI fundamental frequencies. Between VCI(4) and VCI(6), both QFF and SFF-derived fundamental frequencies appear to be converging to their respective VCI limits, as illustrated in Figures 3.1 and 3.2.

Beyond VCI(6), SFF-derived fundamental frequencies all eventually converge to a limiting value. Most are converged to within 1 cm^{-1} by VCI(8), with the

Table 3.3: Mean and maximum absolute errors due to VCI wavefunction truncation, using both sextic (SFF) and quartic (QFF) force field expansions in normal mode coordinates to represent the potential energy surface. Reference results are generated using a sextic force field, increasing excitation level until all fundamental frequencies are converged to within 1 cm^{-1}.

n	$\left\langle \lvert\Delta^{\mathrm{SFF}}_{\mathrm{VCI}(n)}\rvert \right\rangle$ (cm^{-1})	$\lvert\Delta^{\mathrm{SFF}}_{\mathrm{VCI}(n)}\rvert_{\max}$ (cm^{-1})	$\left\langle \lvert\Delta^{\mathrm{QFF}}_{\mathrm{VCI}(n)}\rvert \right\rangle$ (cm^{-1})	$\lvert\Delta^{\mathrm{QFF}}_{\mathrm{VCI}(n)}\rvert_{\max}$ (cm^{-1})
1	92.9	515.5	92.4	498.2
2	16.2	85.9	16.8	92.8
3	31.1	139.1	32.8	140.6
4	5.6	29.4	6.9	51.0
5	1.0	9.1	5.0	39.7
6	0.6	6.0	4.8	38.9
7	0.2	6.6	5.2	54.9
8	0.04	2.4	6.3	219.0*

* N-H stretching mode in NH_3 starts diverging

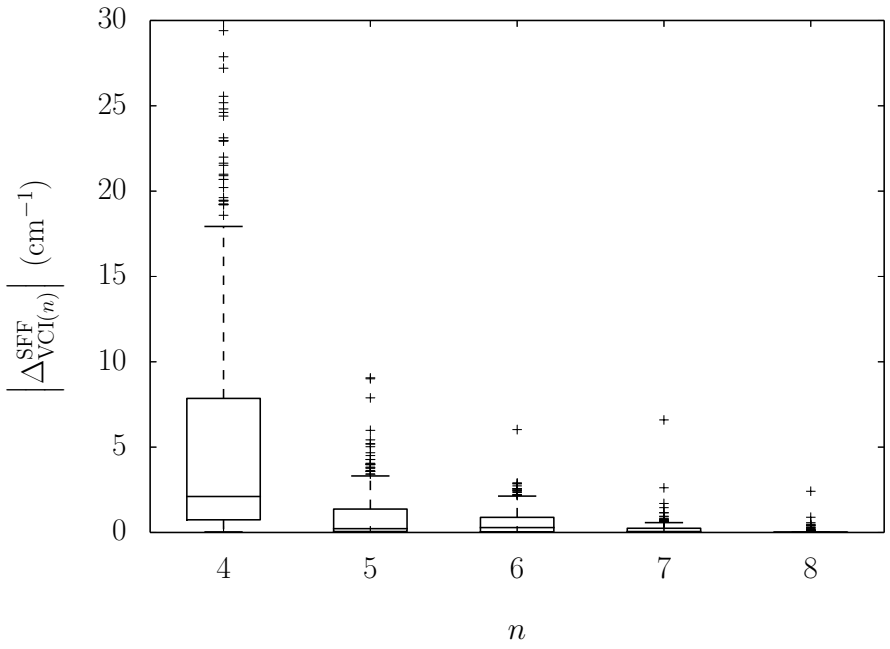


Figure 3.1: The magnitude and range of absolute errors in fundamental frequencies calculated using truncated VCI expansions with a sextic force field (SFF) are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, at each excitation level.

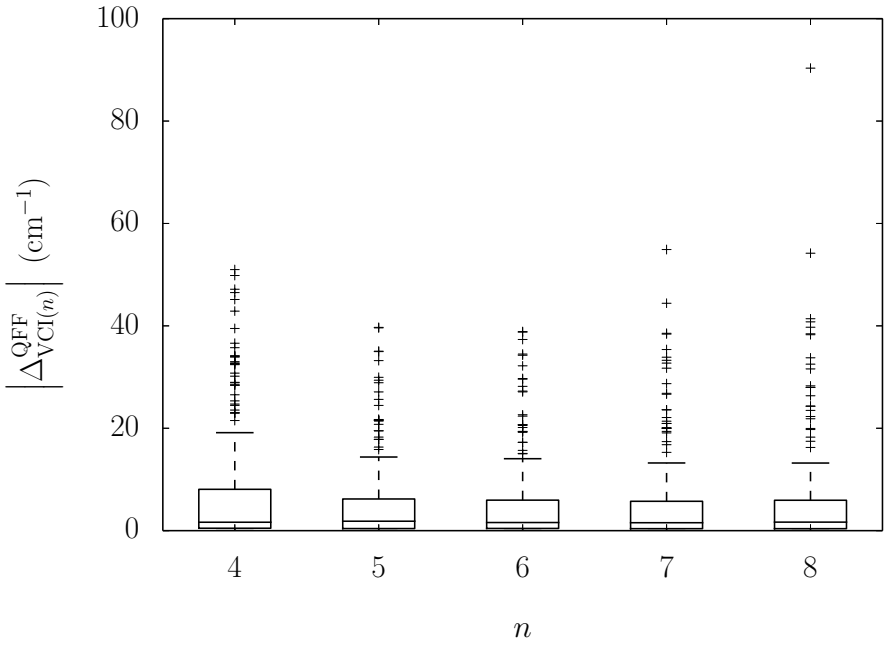


Figure 3.2: The magnitude and range of absolute errors in fundamental frequencies calculated using truncated VCI expansions with a quartic force field (QFF) are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, at each excitation level. The y axis maximum is restricted to 100 cm^{-1}, excluding an additional outlying data point at VCI(8) with a value of 219.0 cm^{-1}.

exception of the symmetric C–H stretching mode of C_3H_2 and the inversion mode of NH_3. In general, C–H stretches exhibit slowest convergence with respect to VCI excitation level. This is due to their highly anharmonic nature with strong coupling to molecular bending modes.

Although the majority of QFF-derived fundamental frequencies converge to a limiting value, divergence is observed in a small number of pathological cases. These include the low frequency bending modes of C_3H_2, the low frequency 'ring-breathing' mode of $C_3H_3^+$, and the inversion and N–H stretching modes of NH_3. Excluding these pathological cases, QFF-derived frequencies deviate from SFF results by 4.7 cm^{-1} on average, and up to 39 cm^{-1} at most.

3.4.2 Negligible VCI matrix element threshold testing

The major computational bottleneck in VCI calculations on larger molecules at higher excitation levels is the memory required to store and diagonalize the VCI matrix. This can potentially be reduced by taking advantage of sparse matrix storage and diagonalization routines, provided enough negligible matrix elements can be excluded by VCI matrix screening.

The accuracy implications of discarding negligible elements from VCI matrices generated using sextic force fields with a range of different screening thresholds, are summarized in Table 3.4 and Figure 3.3, where:

$$\Delta_{\text{screen}}^{\text{VCI}(n)} = \nu_{\text{VCI}(n)}^{\text{SFF}} - \nu_{\text{VCI}(n),\text{screened}}^{\text{SFF}} \tag{3.10}$$

The fractional reduction in number of matrix elements across the test data set is represented in Figure 3.4 and broken down by molecule size in Figure 3.5.

Overall, a screening threshold of 2×10^{-5} E_h achieves a good balance between accuracy and computational cost, reducing the number of matrix elements to be stored by more than half, while introducing errors of 0.2 cm^{-1} on average and 1.7 cm^{-1} at most. Even more fortunately, the fraction of non-negligible elements decreases as the molecule size increases (Figure 3.5). Therefore, matrix element screening becomes more useful the larger the VCI matrix and harder the diagonalization problem.

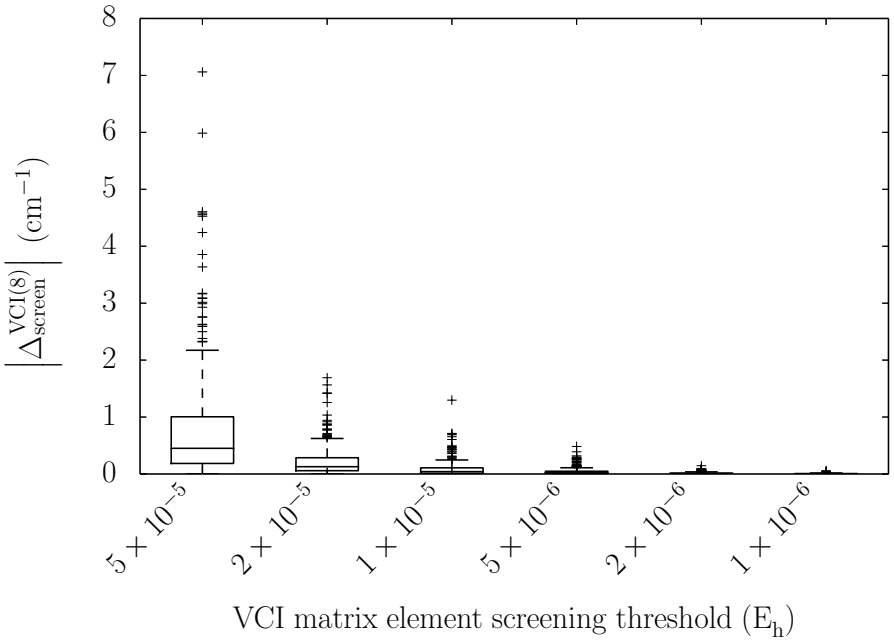


Figure 3.3: The magnitude and range of absolute errors in fundamental frequencies due to matrix element screening are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, for each threshold value.

Table 3.4: Mean and maximum absolute errors arising from VCI matrix element screening with tabulated threshold values.

Threshold (E_h)	$\left\langle \lvert\Delta^{\mathrm{VCI(8)}}_{\mathrm{screen}}\rvert \right\rangle$ (cm^{-1})	$\lvert\Delta^{\mathrm{VCI(8)}}_{\mathrm{screen}}\rvert_{\mathrm{max}}$ (cm^{-1})	$\left\langle \lvert\Delta^{\mathrm{VCI(4)}}_{\mathrm{screen}}\rvert \right\rangle$ (cm^{-1})	$\lvert\Delta^{\mathrm{VCI(4)}}_{\mathrm{screen}}\rvert_{\mathrm{max}}$ (cm^{-1})
5×10^{-5}	0.8	7.1	0.8	11.7
2×10^{-5}	0.2	1.7	0.2	3.1
1×10^{-5}	0.09	1.3	0.08	1.0
5×10^{-6}	0.04	0.5	0.04	0.4
2×10^{-6}	0.01	0.1	0.01	0.1
1×10^{-6}	0.006	0.06	0.004	0.05
5×10^{-7}	0.002	0.02	0.002	0.02
2×10^{-7}	0.0006	0.005	0.0004	0.004
1×10^{-7}	0.0003	0.002	0.0001	0.002

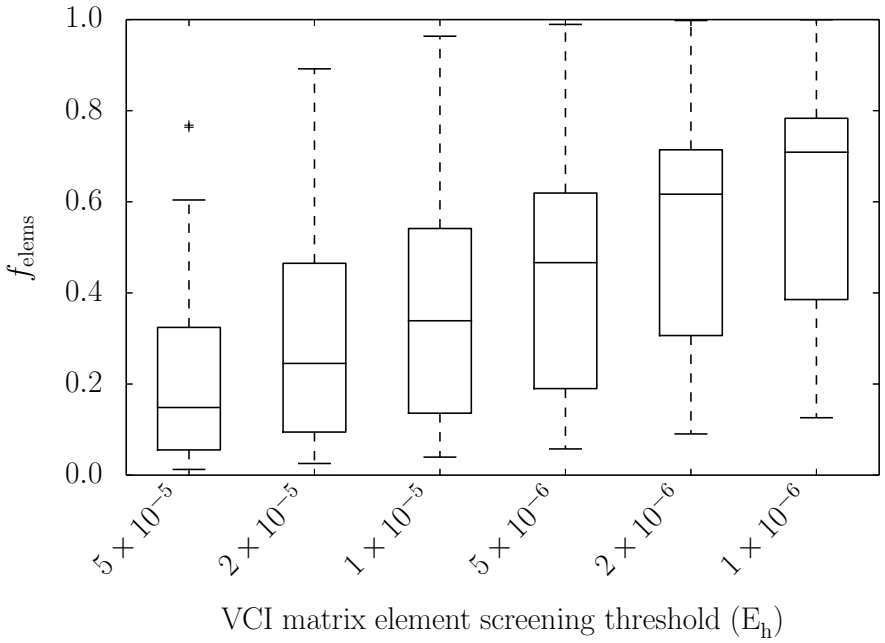


Figure 3.4: The fractional number of VCI matrix elements to be stored varies significantly with screening threshold (x-axis) and across all molecules without low-barrier torsional modes within the PyPES library, as indicated using the boxplot format.

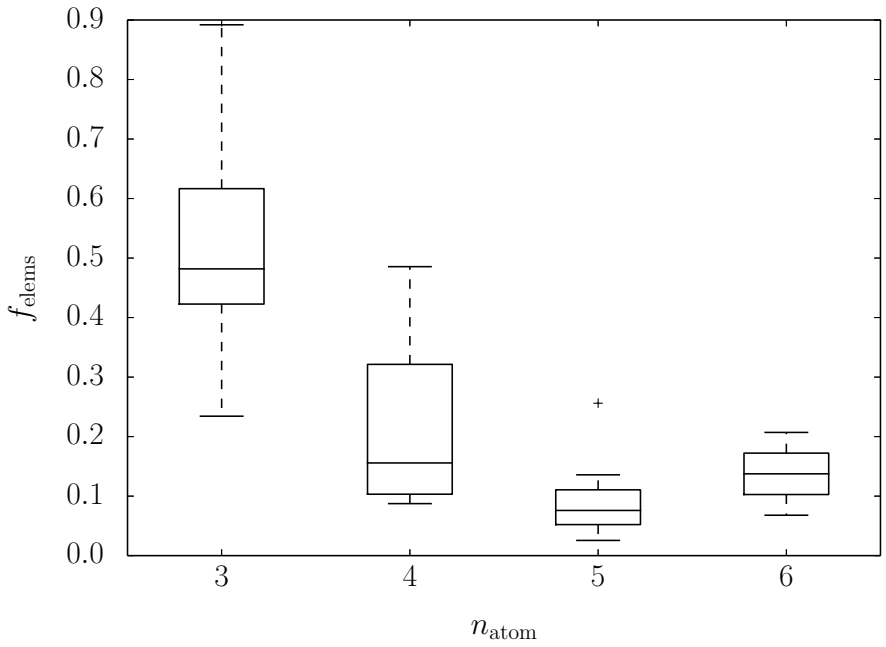


Figure 3.5: The fractional number of VCI matrix elements to be stored at a screening threshold of 2×10^{-5} $\mathrm{E_h}$ is broken down by molecular size (x-axis), with collated data represented in boxplot format.

However, the size of the VCI matrix grows much faster with molecule size than the extent of screening, particularly at high excitation levels. Even with matrix element screening, VCI(8) calculations are not practicable for molecules with more than 7 or 8 atoms. For larger molecules, it will be necessary to truncate the VCI expansion at a lower level, or use a more sophisticated screening approach for selecting VCI matrix elements. Therefore, we test the transferability of our threshold screening value recommendations by repeating the screening threshold testing at VCI(4).

The results presented in Table 3.4 confirm that screening errors are only weakly dependent on VCI level, with negligible differences between average errors due to screening in VCI(8) and VCI(4) calculations.

3.4.3 Negligible force constant threshold testing

Although VCI matrix diagonalization is the computational bottleneck of the VCI algorithm, determining whether a calculation is feasible or not, the majority of a

Table 3.5: Mean and maximum absolute errors arising from force constant screening with tabulated threshold values, using a VCI(8) wavefunction expansion with a sextic force field.

Threshold (cm^{-1})	$\langle\lvert\Delta^{SFF}_{screen}\rvert\rangle$ (cm^{-1})	$\lvert\Delta^{SFF}_{screen}\rvert_{max}$ (cm^{-1})
2	0.1	1.0
1	0.05	0.4
0.5	0.02	0.2
0.1	0.003	0.02

job's runtime is often taken up in VCI matrix construction. A straightforward way to reduce runtime is to pre-screen the force field, removing negligible force constants.

Errors in fundamental frequencies, due to force-constant screening during VCI(8) calculations with a sextic force field, are calculated as:

$$\Delta^{SFF}_{screen} = \nu^{SFF}_{VCI(8)} - \nu^{SFF,screened}_{VCI(8)} \tag{3.11}$$

Statistical analysis of the combined results, for a series of different threshold values, are presented in Table 3.5 and illustrated in Figure 3.6.

The data presented in Table 3.5 and Figure 3.6 show that force constant screening errors converge rapidly and monotonically to the unscreened limit, with maximum errors around half the threshold value and average errors ~ 20 times lower.

Errors due to force constant screening are expected to be largely independent of VCI level; this is supported by our data which return the average errors listed in Table 3.5 for all VCI excitation levels greater than 1. Maximum errors are slightly more dependent on VCI excitation level, but only for higher screening thresholds and at low excitation levels. For example, with a screening threshold of 2 cm^{-1}, maximum errors are constant for $4 < n < 8$, decreasing to 0.8 at $n = 3$ and 0.6 at $n = 2$.

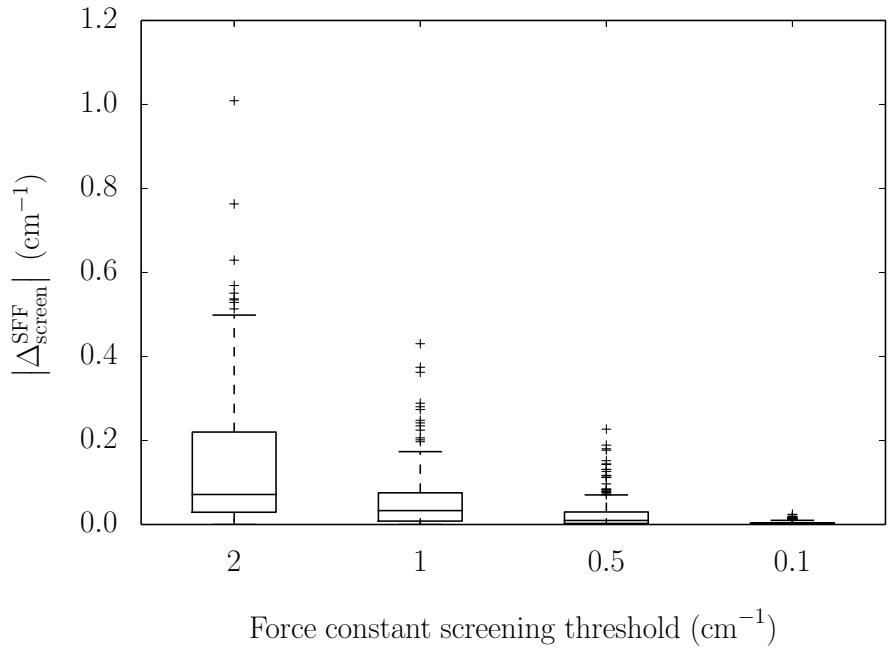


Figure 3.6: The magnitude and range of absolute errors in fundamental frequencies due to force constant screening are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, for each threshold value.

3.4.4 Effect of excluding low-barrier torsional modes

Low-barrier torsional modes are problematic when solving the nuclear vibrational problem in normal mode coordinates, due to both inaccurate representation of the potential energy surface over a large amplitude torsional range, and the inadequacy of the Watson Hamiltonian for describing flexible systems with multiple low barrier minima. The most expedient solution would be to exclude these modes entirely from the VCI expansion, but this could impact upon the accuracy of the remaining fundamental frequencies, particularly if there is strong coupling between torsional modes and others.

To quantify this effect, we have performed VCI(8) calculations using SFFs with the torsional mode excluded from the VCI expansion for each of the molecules listed in Table 3.6. Errors are calculated relative to benchmark literature values:

$$\Delta_{\text{tors}} = \nu_{\text{ref}} - \nu_{\text{VCI(8)}} \tag{3.12}$$

In all cases, the mean absolute error is less than 10 cm^{-1}. Larger mean and maximum errors for *cis*-HSiOH and *trans*-HSiOH likely arise from the inaccuracy of the VPT2-derived reference data, as well as the neglect of torsional modes within our VCI calculations, and so should be considered absolute worse-case values. HOOH is the most anharmonic and strongly coupled, and this is reflected in slightly poorer predictions of fundamental frequencies with the torsional mode excluded. It also has the lowest torsional barrier, resulting in torsional splitting of the other modes that our VCI formulation is unable to capture. This effect also contributes to the relatively large observed Δ_{tors} errors for HOOH.

3.4.5 Recommendations

Final algorithmic recommendations are presented in Table 3.7. In making these choices, we have first aimed to balance force field and VCI method accuracy, and then ensure that errors due to screening are around an order of magnitude lower than errors inherent in the choice of force field representation and VCI excitation level. Overall, we consider that these combinations will provide an optimal balance between accuracy and computational resource demand. The calculated cumulative

Table 3.6: Mean and maximum absolute deviations in calculated fundamental frequencies for all modes except low-barrier torsions, calculated at VCI(8) using a sextic force field and excluding the torsional mode from the VCI expansion. Reference data are collated from the literature, and are generated using a range of different methods, as implemented in; a) RVIB4, b) TROVE[50], c) MULTIMODE[26] and d) SPECTRO[47]. ‡ For HOOH, strong tunnelling splitting leads to doublet spectral peaks. Reference data are derived by averaging split peak positions.

Molecule	Reference method, [citation]	$\langle\lvert\Delta_{\text{tors}}\rvert\rangle$ (cm^{-1})	$\lvert\Delta_{\text{tors}}\rvert_{\text{max}}$ (cm^{-1})
HOOH	VCI[a][51]	7.3‡	13.4‡
HSOH	VCI[b][52]	3.4	6.8
cis-HOCO	VCI[c][53]	3.2	7.8
trans-HOCO	VCI[c][54]	3.2	9.6
cis-HSiOH	VPT2[d][55]	9.2	20.2
trans-HSiOH	VPT2[d][55]	5.7	13.2

Table 3.7: Recommended algorithmic choices and settings

FF order	VCI level	matrix element screening threshold (E_h)	force constant screening threshold (cm^{-1})	cumulative average error (cm^{-1})	cumulative maximum error (cm^{-1})
4 (QFF)	5	2×10^{-5}	2	5.3	43.4
6 (SFF)	7	5×10^{-6}	1	0.3	7.5

errors in Table 3.7 are likely to be a significant overestimate, as some error cancellation is to be expected.

We note that errors arising from discarding negligible matrix elements and force constants may increase somewhat for larger molecules, as the number of small matrix elements and force constants is likely to increase. Therefore, their combined contribution to the overall accuracy of the calculated fundamentals may also increase, requiring a lower screening threshold. If accuracy is paramount, we recommend repeating the screening analysis at VCI(4) to tailor screening thresholds for a given molecule.

Finally, we note that the accuracy of calculated VCI frequencies will strongly reflect the quality of the *ab initio* method used to construct the force field. Here, we have circumvented this problem by using a library of high quality analytic potential energy surfaces. Nonetheless, the benchmark data presented herein will enable the user to make informed choices of VCI level and screening threshold for the quality of the force field in hand. It also provides a useful starting point for estimating contributions to the overall error in calculated fundamental frequencies from approximations inherent in both electronic and nuclear vibrational structure models.

3.5 Summary of program capabilities

Our freely available, open-source PyVCI package provides:

- normal mode analysis with projection of contaminant translational and rotational modes arising from numerical imprecision and/or incomplete geometry optimization
- vibrational configuration theory based upon states with an 'excitation level' (maximum sum of vibrational quantum numbers) of up to 10.
- expansion of the potential up to 6th order in normal mode coordinates and/or any orthonormal linear combination of normal modes
- Coriolis coupling corrections
- optional exclusion of selected vibrational modes – typically low-barrier torsional modes – from the VCI expansion
- parallel sparse matrix construction and diagonalization

Appendix A: Normal mode analysis

All quantities reported herein are in atomic units, unless explicitly stated otherwise. The atomic unit of energy is Hartree, the atomic unit of length is the Bohr radius, and the electron mass is the fundamental atomic unit of mass ($m_e = 1$). The conversion factor to convert between a.m.u. (g mol^{-1}) and atomic units of mass (m_e molecule^{-1}) is 1822.888482.

1. Form the mass-weighted Cartesian matrix $\mathbf{f}^{\mathrm{MWC}}$ from the second derivatives of the energy with respect to nuclear displacements (Hessian matrix, $\mathbf{f}^{\mathrm{Hess}}$):

$$f_{ij}^{\mathrm{MWC}} = f_{ij}^{\mathrm{Hess}} / \sqrt{m_i m_j} \tag{3.13}$$

2. Project out translational and rotational modes.

Construct and diagonalize the inertia tensor, $\mathbf{I}$:

$$\mathbf{I} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} = \begin{pmatrix} \sum_i m_i(y_i^2 + z_i^2) & -\sum_i m_i(x_i y_i) & -\sum_i m_i(x_i z_i) \\ -\sum_i m_i(y_i x_i) & \sum_i m_i(x_i^2 + z_i^2) & -\sum_i m_i(y_i z_i) \\ -\sum_i m_i(z_i x_i) & -\sum_i m_i(z_i y_i) & \sum_i m_i(x_i^2 + y_i^2) \end{pmatrix} \tag{3.14}$$

where $1 < i < N_{\mathrm{atom}}$, producing a 3×3 matrix, $\mathbf{X}$ of normalized eigenvectors of $\mathbf{I}$.

Calculate the centre of mass, $\mathbf{R}^{\mathrm{COM}}$ from the atomic position vectors $\mathbf{r}_i$:

$$\mathbf{R}^{\mathrm{COM}} = \frac{\sum_i m_i \mathbf{r}_i}{\sum_i m_i} \tag{3.15}$$

and shift the coordinates of each atom, indexed by i, so the centre of mass is at the origin:

$$\mathbf{R}_i = \mathbf{r}_i - \mathbf{R}^{\mathrm{COM}} \tag{3.16}$$

Assemble vectors corresponding to translation and rotation in mass-weighted Cartesian coordinates, storing them as $3 \times 3N$ matrices $\mathbf{D}^{\mathrm{trans}}$ and $\mathbf{D}^{\mathrm{rot}}$, respectively:

$$D_{ij}^{\mathrm{trans}} = \delta_{ik} \sqrt{m_n} \tag{3.17}$$

where $1 < i < 3$, $1 < j < 3N$, n is the atom number for a given value of j [$n = \mathrm{ceiling}(j/3)$] and k the Cartesian component index [$k = \mathrm{modulo}(j, 3) +$

$3\times\text{floor}(j/3)]$; $1 = x$, $2 = y$, $3 = z$.

$$D_{ij}^{\text{rot}} = \left[(\mathbf{R}_n \cdot \mathbf{X}_l^{\text{T}})X_{im} - (\mathbf{R}_n \cdot \mathbf{X}_m^{\text{T}})X_{il}\right] \times \sqrt{m_n} \tag{3.18}$$

where i, j, k and n are defined as above, and k, l, m are cyclic permutations of x, y, z, *i.e.* if $k = 1 = x$ then $l = 2 = y$ and $m = 3 = z$; if $k = 2 = y$ then $l = 3 = z$ and $m = 1 = x$, and if $k = 3 = z$ then $l = 1 = x$ and $m = 2 = y$.

Append $\mathbf{D}^{\text{trans}}$, $\mathbf{D}^{\text{rot}}$ and the identity matrix, $\mathbf{D}^{\text{I}}$, to form the $(3N + 6) \times 3N$ projection matrix $\mathbf{D}$. Orthonormalize $\mathbf{D}$ using the Gram-Schmidt process, deleting the 6 redundant rows at the end, to yield a final $(3N \times 3N)$ $\mathbf{D}$ matrix.

Transform the mass-weighted Hessian to an 'internal' Cartesian coordinate set with translational and rotational modes rigorously projected out, by extracting the bottom right $N_{\text{vib}} \times N_{\text{vib}}$ sub-matrix $\mathbf{f}^{\text{INT}}$:

$$\mathbf{f}^{\text{INT}} = \mathbf{D}\mathbf{f}^{\text{MWC}}\mathbf{D}^{\text{T}} \tag{3.19}$$

3. Diagonalize the projected force constant matrix, to obtain the eigenvalues $\mathbf{\Lambda}$ and eigenvectors, $\mathbf{L}$:

$$\mathbf{f}^{\text{INT}}\mathbf{L} = \mathbf{\Lambda}\mathbf{L} \tag{3.20}$$

4. Calculate frequencies in units of reciprocal centimetres, reduced masses and normalized Cartesian displacement vectors for each vibrational mode

$$\tilde{\nu}_i = \sqrt{\lambda_i} \times 219474.6435 \tag{3.21}$$

where λ_i are the diagonal elements of $\mathbf{\Lambda}$, and $\tilde{\nu}_i$ are the frequencies in cm^{-1}.

The unnormalized Cartesian displacement vectors that diagonalize $\mathbf{f}^{\text{Hess}}$ are given by:

$$\mathbf{l}_{\text{CART}} = \mathbf{M}\mathbf{D}_{\text{INT}}^{\text{T}}\mathbf{L} \tag{3.22}$$

where $\mathbf{D}_{\text{INT}}$ is the $\mathbf{D}$ matrix with the first 6 rows corresponding to rotational and translational vectors omitted, $\mathbf{M}$ is a diagonal matrix with diagonal elements $M_n = 1/\sqrt{m_n}$ for $1 < j < 3N_{\text{atom}}$ and n identifies the atom number for each block of x, y, z coordinates, as defined above.

Normalization factors (N_n) and reduced masses (μ_n) are straightforward to calculate for each vibrational mode:

$$\mu_n = \left(\sum_{j=1}^{3N} (l_{jn}^{\mathrm{CART}})^2 \right)^{-1} \tag{3.23}$$

$$N_n = \sqrt{\mu_n} \tag{3.24}$$

where n is the vibrational mode index.

Appendix B: Coriolis coupling constants

Alongside the **D** and **L** matrices from above, enforcing Eckart conditions and transformation from Cartesian to normal mode coordinates, respectively, 3 rotational vector product matrices ($\mathscr{M}^x$, $\mathscr{M}^y$ and $\mathscr{M}^z$) are also required. Each matrix comprises N_{atom} identical 3×3 submatrices (one for each atom) along the main diagonal. These diagonal blocks have the form:

$$(\mathscr{M}^x) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, (\mathscr{M}^y) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (\mathscr{M}^z) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.25}$$

The eigenvectors of the moment of inertia tensor stored in the matrix **X** are also required to rotate the coordinate system to align the molecule with its principal axes of inertia. This is acheived by composing a $(3N \times 3N)$ matrix $\mathscr{X}$ with N_{atom} identical 3×3 **X** blocks on the diagonal.

Coriolis coupling coefficient matrices, ζ^x, ζ^y and ζ^z are then calculated as:

$$\zeta^\alpha = (\mathscr{X}^{\mathrm{T}} \mathbf{D}_{\mathrm{INT}}^{\mathrm{T}} \mathbf{L})^{\mathrm{T}} \mathscr{M}^\alpha \mathscr{X}^{\mathrm{T}} \mathbf{D}_{\mathrm{INT}}^{\mathrm{T}} \mathbf{L} \tag{3.26}$$

where $\alpha = x$, y, or z.

References

[1] PyPES Library of Potential Energy Surfaces, `http://sourceforge.net/projects/pypes-lib/`.

[2] M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* **2015**, *36*, 2200–2207.

[3] PES Database, `http://pes-database.theochem.uni-stuttgart.de/surfaces/index.php`.

[4] G. Rauhut, *J. Chem. Phys.* **2004**, *121*, 9313–9322.

[5] PotLib Potential Energy Surface Library, `http://comp.chem.umn.edu/potlib/`.

[6] R. J. Duchovic, Y. L. Volobuev, G. C. Lynch, D. G. Truhlar, T. C. Allison, A. F. Wagner, B. C. Garrett, J. C. Corchado, *Comp. Phys. Comm.* **2002**, *144*, 169–187.

[7] ezPES Library of Potential Energy Surfaces, `http://iopenshell.usc.edu/downloads/ezpes/`.

[8] J. M. Bowman, B. J. Braams, S. Carter, C. Chen, G. Czako, B. Fu, X. Huang, E. Kamarchik, A. R. Sharma, B. C. Shepler, Y. Wang, Z. Xie, *J. Phys. Chem. Lett.* **2010**, *1*, 1866–1874.

[9] B. J. Braams, J. M. Bowman, *Int. Rev. Phys. Chem.* **2009**, *28*, 577–606.

[10] R. Ramakrishnan, G. Rauhut, *J. Chem. Phys.* **2015**, *142*, 154118.

[11] M. Ringholm, D. Jonsson, R. Bast, B. Gao, A. J. Thorvaldsen, U. Ekström, T. Helgaker, K. Ruud, *J. Chem. Phys.* **2014**, *140*, 034103.

[12] R. C. Fortenberry, X. Huang, A. Yachmenev, W. Thiel, T. J. Lee, *Chem. Phys. Lett.* **2013**, *574*, 1–12.

[13] G. Rauhut, B. Hartke, *J. Chem. Phys.* **2009**, *131*, 014108.

[14] M. Neff, G. Rauhut, *J. Chem. Phys.* **2009**, *131*, 124129.

[15] E. Matito, D. Toffoli, O. Christiansen, *J. Chem. Phys.* **2009**, *130*, 134104.

[16] C. Y. Lin, A. T. B. Gilbert, P. M. W. Gill, *Theor. Chem. Acc.* **2008**, *120*, 23–25.

[17] K. Yagi, S. Hirata, K. Hirao, *Theor. Chem. Acc.* **2007**, *118*, 681–691.

[18] K. Yagi, K. Hirao, T. Taketsugu, M. Schmidt, M. Gordon, *J. Chem. Phys.* **2004**, *121*, 1383–1389.

[19] S. Dressler, W. Thiel, *Chem. Phys. Lett.* **1997**, *273*, 71–78.

[20] T. K. Roy, R. B. Gerber, *Phys. Chem. Chem. Phys.* **2013**, *15*, 9468–9492.

[21] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, New York, **1955**.

[22] J. Bowman, *Acc. Chem. Res.* **1986**, *19*, 202–208.

[23] S. Carter, S. Culik, J. Bowman, *J. Chem. Phys.* **1997**, *107*, 10458–10469.

[24] S. Carter, J. Bowman, N. Handy, *Theor. Chem. Acc.* **1998**, *100*, 191–198.

[25] N. Wright, R. Gerber, *J. Chem. Phys.* **2000**, *112*, 2598–2604.

[26] J. Bowman, S. Carter, X. Huang, *Int. Rev. Phys. Chem.* **2003**, *22*, 533–549.

[27] D. Benoit, *J. Chem. Phys.* **2004**, *120*, 562–573.

[28] O. Christiansen, *Phys. Chem. Chem. Phys.* **2007**, *9*, 2942–2953.

[29] N. Matsunaga, G. Chaban, R. Gerber, *J. Chem. Phys.* **2002**, *117*, 3541–3547.

[30] O. Christiansen, *J. Chem. Phys.* **2003**, *119*, 5773–5781.

[31] V. Barone, *J. Chem. Phys.* **2004**, *120*, 3059–3065.

[32] V. Barone, *J. Chem. Phys.* **2005**, *122*, 014108.

[33] P. Cassam-Chenai, J. Lievin, *Int. J. Quant. Chem.* **2003**, *93*, 245–264.

[34] W. Mizukami, D. P. Tew, *J. Chem. Phys.* **2013**, *139*, 194108.

[35] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Phys. Chem. A* **2012**, *116*, 3691–3709.

[36] J. Bloino, M. Biczysko, V. Barone, *J. Chem. Theor. Comp.* **2012**, *8*, 1015–1036.

[37] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Chem. Phys.* **2014**, *141*, 234114.

[38] A. M. Rosnik, W. F. Polik, *Mol. Phys.* **2014**, *112*, 261–300.

[39] V. Barone, M. Biczysko, J. Bloino, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.

[40] M. Piccardo, J. Bloino, V. Barone, *Int. J. Quant. Chem.* **2015**, *115*, 948–982.

[41] S. Carter, N. Handy, *Comp. Phys. Rep.* **1986**, *5*, 115–172.

[42] C. Fabri, T. Furtenbacher, A. G. Csaszar, *Mol. Phys.* **2014**, *112*, 2462–2467.

[43] O. Christiansen, *J. Chem. Phys.* **2004**, *120*, 2149–2159.

[44] G. Rauhut, G. Knizia, H.-J. Werner, *J. Chem. Phys.* **2009**, *130*, 054105.

[45] P. Seidler, M. Sparta, O. Christiansen, *J. Chem. Phys.* **2011**, *134*, 054119.

[46] O. Christiansen, *Phys. Chem. Chem. Phys.* **2012**, *14*, 6672–6687.

[47] J. F. Gaw, A. Willetts, W. H. Green, N. C. Handy, *Advances in Molecular Vibrations and Collision Dynamics A Research Annual* **1991**, *1B*, 169.

[48] J. Meal, S. Polo, *J. Chem. Phys.* **1956**, *24*, 1119–1125.

[49] J. Meal, S. Polo, *J. Chem. Phys.* **1956**, *24*, 1126–1133.

[50] S. N. Yurchenko, W. Thiel, P. Jensen, *J. Mol. Spectrosc.* **2007**, *245*, 126–140.

[51] P. Malyszek, J. Koput, *J. Comp. Chem.* **2013**, *34*, 337–345.

[52] S. N. Yurchenko, A. Yachmenev, W. Thiel, O. Baum, T. F. Giesen, V. V. Melnikov, P. Jensen, *J. Mol. Spectrosc.* **2009**, *257*, 57–65.

[53] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, *J. Chem. Phys.* **2011**, *135*.

[54] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, *J. Chem. Phys.* **2011**, *135*.

[55] J. Martin, *J. Phys. Chem. A* **1998**, *102*, 1394–1404.

Chapter 4

Efficient construction of anharmonic vibrational force fields by coordinate transformation from curvilinear to rectilinear normal mode coordinates

4.1 Introduction

Infrared and Raman spectroscopy are widely used chemical characterisation techniques that interrogate the vibrational structure of matter, providing information about molecular structure and bonding. However, they fall short of other spectroscopic methods, such as nuclear magnetic resonance or X-ray diffraction, in the amount of structural information that may be deduced from each spectrum. This is largely due to the mismatch between the accuracy to which IR spectrum can be easily modelled using normal mode analysis and the much higher accuracy needed for quantitative assignment.

A variety of black box methods for calculating anharmonic vibrational frequencies have been developed. They include vibrational perturbation theory (VPT),[1–14] vibrational self consistent field,[15–20] vibrational coupled cluster[21–23] and vibrational configuration interaction (VCI).[24–27] A series of recent review articles provide a good overview of the field.[28–34]

All black box nuclear vibrational structure models are based upon the Watson Hamiltonian,[35,36] which is formulated in rectilinear normal mode coordinates. The main advantages of this coordinate set are that it can be uniquely defined, has a simple form, and to a good approximation, decouples the kinetic energy operator (KEO). When the potential energy function is expanded as a power series in normal mode coordinates, then the Hamiltonian can be integrated analytically over a harmonic oscillator or distributed Gaussian function product basis, thus allowing efficient evaluation of Hamiltonian matrix elements. However, the rectilinear nature of normal mode coordinates, while leading to a simple KEO, introduces artificial coupling in the potential and increases the computational cost of its construction.

Curvilinear coordinates, *i.e.* coordinates that are non-linear in Cartesian space, more naturally describe intramolecular motion and lead to reduced coupling in the potential.[37–42] However, derivation of the KEO becomes rather complicated,[43–45] although it has been done for some types of curvilinear coordinates.[46–48] Most notably, recent advances with polyspherical harmonic coordinates now allow derivation of analytical expressions of the KEO with any practical set of valence internal coordinates.[49–52] However, even when analytical expressions are available they are still far more complicated than those obtained from the Watson Hamiltonian, and applications have been limited to smaller systems.

Ideally, the potential should be constructed using curvilinear coordinates to minimise the number of computationally intensive *ab initio* calculations required, and then transformed to rectilinear normal mode coordinates, so that the vibrational problem can be more easily solved using the Watson Hamiltonian. The issue of formulating a non-redundant set of curvilinear coordinates from a redundant set of valence internals has been addressed in research on geometry optimisation.[53,54] The delocalised internal coordinates proposed by Baker et al.[55] can be defined automatically and form a non-redundant set. They are commonly used in geometry optimisation and give faster convergence than Cartesian coordinates. Defining curvilinear normal mode coordinates from delocalised internal coordinates allows the potential to be generated by numerical differentiation. Subsequent coordinate transformation to rectilinear normal modes is straightforward, although the implementation is non-trivial.

4.2 Description of the Method

4.2.1 Formulation of Curvilinear Normal Mode Coordinates

Assuming, for now, that a redundant set of N_{int} internal coordinates have been defined, $S = (s_1, s_2, \ldots, s_{N_{\text{int}}})^T$, we now wish to perform a linear transformation to a non-redundant set of internal coordinates, $\Sigma = (\sigma_1, \sigma_2, \ldots, \sigma_{N_{\text{mode}}})^T$, where $N_{\text{mode}} = 3n - 6$ is the number of vibrational degrees of freedom, or $(3n - 5)$ for linear molecules, and n is the number of atoms. This can be represented with an $(N_{\text{mode}} \times N_{\text{int}})$ matrix $\mathbf{U}$,

$$\Sigma = \mathbf{U}S. \tag{4.1}$$

Various approaches for defining the matrix $\mathbf{U}$ have been proposed, and in our studies we use the delocalised internal coordinates formulation of Baker et al.[55] It requires construction and subsequent diagonalisation of an $(N_{\text{int}} \times N_{\text{int}})$ matrix,

$$\mathbf{G} = \mathbf{B}\mathbf{B}^t, \tag{4.2}$$

where $\mathbf{B}$ is the usual Wilson B matrix,[56] with elements $B_{ij} = (\partial s_i / \partial x_j)$ evaluated at equilibrium. After diagonalisation of $\mathbf{G}$, the N_{mode} eigenvectors with non-zero eigenvalues form the $\mathbf{U}$ matrix.

With the usual procedure,[56,57] rectilinear normal mode coordinates, $Q = (q_1, q_2, \ldots, q_{N_{\text{mode}}})^T$, can be defined and their relationship to Cartesian coordinates, X, established,

$$\Delta X = \mathbf{L}\Delta Q, \tag{4.3}$$

where $\mathbf{L}$ contains elements $L_{ij} = (\partial x_i / \partial q_j)$ defining the one-to-one mapping form normal mode to Cartesian space.

The curvilinear normal modes, $\tilde{Q} = (\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_{N_{\text{mode}}})^T$, are defined as the linear combination of Σ that reproduces their rectilinear counterparts, Q, to first order in Cartesian space,

$$\mathbf{R}^{(\sigma)} = (\mathbf{B}^{(\sigma)}\mathbf{L})^{-1} \tag{4.4}$$

where $\mathbf{B}^{(\sigma)}$ the Wilson B matrix, as defined above, with the non-redundant set of internals, and $\mathbf{R}^{(\sigma)}$ is an $(N_{\text{mode}} \times N_{\text{mode}})$ matrix with elements $R_{ij}^{(\sigma)} = (\partial \tilde{q}_i / \partial \sigma_j)$.

To simplify subsequent transformations, it is convenient to relate $\tilde{Q}$ directly to the redundant set of internals,

$$\mathbf{R} = \mathbf{R}^{(\sigma)}\mathbf{U}, \tag{4.5}$$

where matrix $\mathbf{R}$ contains elements $R_{ij} = (\partial\tilde{q}_i/\partial s_j)$.

The curvilinear and rectilinear normal modes can be related via a non-linear transformation,

$$\begin{aligned}\Delta\tilde{q}_i &= \sum_r \tilde{L}_{i,r}\Delta q_r + \frac{1}{2!}\sum_{r,t}\tilde{L}_{i,rt}\Delta q_r\Delta q_t + \ldots \\ &= \Delta q_r + \frac{1}{2!}\sum_{r,t}\tilde{L}_{i,rt}\Delta q_r\Delta q_t + \ldots,\end{aligned} \tag{4.6}$$

where

$$\tilde{L}_{i,rt} = \frac{\partial^2\tilde{q}_i}{\partial q_t \partial q_r},$$

with similar notation adopted for higher order derivatives. The simplification in the first term of equation (4.6) arises because the curvilinear normal mode coordinates ($\tilde{Q}$) and rectilinear normal mode coordinates (Q) are the same to first order, by construction. That is, $\tilde{L}_{i,r} = \delta_{ir}$, where δ_{ir} is the Kronecker delta.

4.2.2 Coordinate Transformations

With curvilinear normal mode coordinates now defined, $\tilde{q}_i = \sum_j R_{ij}s_j$, the potential energy function can be constructed as a Taylor series expansion in $\tilde{Q}$,

$$\tilde{V} = \frac{1}{2!}\sum_{i,j}F_{ij}\Delta\tilde{q}_i\Delta\tilde{q}_j + \frac{1}{3!}\sum_{i,j,k}F_{ijk}\Delta\tilde{q}_i\Delta\tilde{q}_j\Delta\tilde{q}_k + \frac{1}{4!}\sum_{i,j,k,l}F_{ijkl}\Delta\tilde{q}_i\Delta\tilde{q}_j\Delta\tilde{q}_k\Delta\tilde{q}_l + \ldots, \tag{4.7}$$

where force constants F represent derivatives at equilibrium with respect to curvilinear normal modes. The first derivative is zero at equilibrium and the absolute energy is irrelevant in this context, so those terms were omitted.

Due to the non-linear relationship between curvilinear and rectilinear normal modes, equation (4.6), the coordinate transformation of the potential is also non-linear. The necessary expressions have been described in detail elsewhere[1,38,58,59] and only the expression for the normal mode coordinate force constants of third

order, K_{rtv}, is included here for illustration,

$$\begin{aligned} K_{rtv} &= \sum_{i,j,k} F_{ijk}\tilde{L}_{i,r}\tilde{L}_{j,t}\tilde{L}_{k,v} + \sum_{i,j} F_{ij}(\tilde{L}_{i,rt}\tilde{L}_{j,v} + \tilde{L}_{i,rv}\tilde{L}_{j,t} + \tilde{L}_{i,tv}\tilde{L}_{j,r}) \\ &= F_{rtv} + \sum_{i,j} F_{ij}(\tilde{L}_{i,rt}\tilde{L}_{j,v} + \tilde{L}_{i,rv}\tilde{L}_{j,t} + \tilde{L}_{i,tv}\tilde{L}_{j,r}). \end{aligned} \tag{4.8}$$

The derivatives $\tilde{L}$ can be obtained by two linear transformations. Only transformations up to third order are shown, and higher order transformations can be easily derived by application of the chain rule.

Firstly, derivatives of internal coordinates S with respect to X are transformed to Q,

$$\begin{aligned} \frac{\partial s_i}{\partial q_r} &= \sum_{\alpha} \frac{\partial s_i}{\partial x_\alpha} L_{\alpha r} \\ \frac{\partial^2 s_i}{\partial q_t \partial q_r} &= \sum_{\alpha\beta} \frac{\partial^2 s_i}{\partial x_\beta \partial x_\alpha} L_{\alpha r} L_{\beta t} \\ \frac{\partial^3 s_i}{\partial q_v \partial q_t \partial q_r} &= \sum_{\alpha\beta\gamma} \frac{\partial^3 s_i}{\partial x_\gamma \partial x_\beta \partial x_\alpha} L_{\alpha r} L_{\beta t} L_{\gamma v}, \end{aligned} \tag{4.9}$$

where $L_{\alpha r} = (\partial x_\alpha / \partial q_r)$ are elements of the $\mathbf{L}$ matrix, and the summation is over all Cartesian coordinates that were used in definition of s_i. Lastly, derivatives of $\tilde{Q}$ with respect to Q are defined as linear combinations of derivatives of S with respect to Q,

$$\begin{aligned} \tilde{L}_{i,r} &= \sum_{\alpha} \frac{\partial s_\alpha}{\partial q_r} R_{i\alpha} \\ \tilde{L}_{i,rt} &= \sum_{\alpha} \frac{\partial^2 s_\alpha}{\partial q_t \partial q_r} R_{i\alpha} \\ \tilde{L}_{i,rtv} &= \sum_{\alpha} \frac{\partial^3 s_\alpha}{\partial q_v \partial q_t \partial q_r} R_{i\alpha}, \end{aligned} \tag{4.10}$$

where $R_{i\alpha} = (\partial \tilde{q}_i / \partial s_\alpha)$ are elements for the previously defined matrix $\mathbf{R}$.

Once the $\tilde{L}$ derivatives have been evaluated, the transformation can be performed. The opposite transformation, from rectilinear to curvilinear normal modes, can also be performed, and the necessary derivatives can be obtained from $\tilde{L}$, as described in the Appendix. This can be useful if the reverse transformation form rectilinear to curvilinear normal modes is required. For example, when comparing different sets of curvilinear coordinates, by changing S, and using the potential in

Q as an intermediate for the exact transformation into a potential in $\tilde{Q}$ of the same derivative order.

In practice, the potential energy function has to be truncated at some derivative level, d, and it is also common to use the reduced mode representation,[60] wherein mode coupling at all derivative levels is truncated at a specified value, m. Any force constants which involve more than m different coordinates are excluded. In the following discussion, the potential will be specified as $V(m,d)$, indicating its mode representation and highest derivative level, and when $m = d$ than only derivative level will be specified, $V(d)$. Expansion of the potential in curvilinear normal mode coordinates, $\tilde{Q}$, will be indicated with a tilde, $\tilde{V}$. Potentials obtained by coordinate transformation will be indicated by giving the starting potential with a right arrow pointing to the final potential. For example, $\tilde{V}(m_c,d_c) \rightarrow V(m_r,d_r)$ represents a potential in $\tilde{Q}$ with mode representation m_c and derivative level d_c, transformed to Q with mode representation m_r and derivative level d_r. The subscripts c and r pertain to differentiate curvilinear and rectilinear normal mode coordinates, respectively.

To achieve exact coordinate transformation, not all derivative levels, d_l, and mode representations, m_l, in derivatives of mode q_i in $\tilde{L}$ are necessary. In particular, since the first derivative of the potential is zero at equilibrium it does not need to be transformed, hence $d_l = d_r - 1$, see equation (4.8). Also, the mode representation does not need to be larger than in $V(m_r,d_r)$.

When valence internal coordinates are used, then the scaling for transformation of all internal coordinate derivatives with respect to Cartesian coordinates into derivatives with respect to rectilinear normal modes, equation (4.9), is $O(N_{\text{mode}}^{m_l} N_{\text{int}})$. The scaling from equation (4.10) is also $O(N_{\text{mode}}^{m_l} N_{\text{int}})$. On the other hand, nonlinear transformation of the potential scales as $O(N_{\text{mode}}^{(m_c+m_r)})$ and it simplifies to $O(N_{\text{mode}}^{(m_c+m_r-1)})$ when $d_c = d_r$. As a result, coordinate transformation of the potential energy function, rather than construction of the required intermediates, is the bottleneck of this procedure.

4.2.3 Choice of Internal Coordinates

If we restrict ourselves to valence coordinates based on bond lengths, bond angles and dihedral angles, there are a few functional forms that are commonly used.

For bond lengths, both Morse,[61] f_{M}, and Simons-Parr-Finlan,[62] f_{SPF}, coordinates have appropriate asymptotic behaviour and lead to fast convergence of the potential energy function. They are defined as,

$$f_{\mathrm{M}}(r) = 1 - e^{-a\Delta r}, \tag{4.11}$$

$$f_{\mathrm{SPF}}(r) = \frac{\Delta r}{r}, \tag{4.12}$$

where r is the bond length, Δr is its deviation from equilibrium, and a is an empirical parameter controlling the shape of the Morse coordinate. We choose f_{SPF} in preference to f_{M} because it is parameter free.

For the angular coordinates, the geometrically defined angles, θ and τ, for bond angle and dihedral angle, respectively, are commonly used. Alternatively, $\cos(\theta)$ and $\sin(\tau)$ or $\cos(\tau)$ are often favoured due to their more physically defined periodic behaviour.[63] The periodic nature of sine and cosine functions allows for more accurate description of potentials with multiple minima, connected via dihedral angle or bond angle coordinates, which is one of the reasons for their popularity.

However, when defining curvilinear normal modes it is important that they have correct limits. In case of sine and cosine functions, only values between 0 and 1 have a physical meaning, which restricts working range of $\tilde{Q}$, if they are used. This is particularly problematic for the dihedral angle coordinates, since the equilibrium geometry of an ordinary molecule is likely to have $\sin(\tau)$ and $\cos(\tau)$ close to their turning points. Further, since we are restricted to a single minimum by the Watson Hamiltonian, use of more localised θ and τ functions is more appropriate. The $\{f_{\mathrm{SPF}}, \theta, \tau\}$ coordinate set is implemented within the PyPES library[64] along with all higher order derivatives necessary to define $\tilde{L}$.

4.3 Computational Details

Our test set consists of 25 molecules, ranging in size from 5 atoms to 8 atoms. From the PyPES library,[64] we include methane (CH_4), ammonium ion (NH_4^+), phosphine oxide (OPH_3), cyclopropenylidene (C_3H_2), cyclopropenyl cation ($C_3H_3^+$) and ethene (C_2H_4) molecules. The rest of potentials are calculated at HF/6-311G**, via numerical differentiation. The following molecules are included: fluoroethene, cyclopropenone, 1,2-diazacyclobutadiene, 1,3-diazacyclobutadiene, 1H-pentazole, 1H-azirine, 4H-1,2-oxazete, cyclopropene, 1,2,5-oxadiazole, 1,2,4-oxadiazole, oxirane, thiirane, 2(3H)-azetone, aziridine, cyclobutadiene, cyclopropanone, methylenecyclopropene, 1,3-oxazole, and 2H-oxete. Graphical representation of geometries for all molecules used in this study is included in the Supplementary Information.

For molecules from the PyPES library, the full mode representation $V(4)$ potential is obtained analytically. For other molecules, numerical differentiation is performed using analytical second derivatives and a variable step size in rectilinear normal modes. Step sizes recommended by Boese et al.[65] are used:

$$\Delta q_i = c \times \sqrt{\mu_i} \times \sqrt{\frac{1000\ \text{cm}^{-1}}{\omega_i}} \tag{4.13}$$

where $c = 0.04$ is a constant, ω_i is the harmonic frequency for mode i in cm^{-1}, and μ_i is the reduced mass for mode i, $\mu_i = \left(\sum_j L_{ij}^2\right)^{-1}$. The $V(4)$ potential is then evaluated using previously reported expressions:[38,66]

$$\begin{aligned}
K_{ijk} =& \frac{1}{2\Delta q_i}\left[K_{jk}(+\Delta q_i) - K_{jk}(-\Delta q_i)\right] \\
K_{iijk} =& \frac{1}{(\Delta q_i)^2}\left[K_{jk}(+\Delta q_i) + K_{jk}(-\Delta q_i) - 2K_{jk}\right] \\
K_{ijkl} =& \frac{1}{2\Delta q_i \Delta q_j}[K_{kl}(+\Delta q_i, +\Delta q_j) + K_{kl}(-\Delta q_i, -\Delta q_j) + 2K_{kl} \\
& - K_{kl}(+\Delta q_i) - K_{kl}(-\Delta q_i) - K_{kl}(+\Delta q_j) - K_{kl}(-\Delta q_j)],
\end{aligned} \tag{4.14}$$

where K_{ij} refers to (i, j) elements of the Hessian matrix obtained analytically at displaced geometry. Values in brackets indicate displacement from equilibrium, and refer to values at equilibrium when absent. Hessian matrices are evaluated on a full grid of all unique $(\pm q_i)$ and $(\pm q_i, \pm q_j)$ displacements, and averaged values for any higher order force constants that could be evaluated in multiple ways are used. For

example, K_{ijk} can be evaluated from displacements along q_i, q_j or q_k, in this case all three are calculated and then averaged. $V(4)$ potentials are then transformed to curvilinear normal modes, $\tilde{V}(4)$, analytically, using our coordinate transformation procedure.

For testing purposes, this approach is favoured over construction of the potential directly in curvilinear normal modes, mainly because step size recommendations for rectilinear normal modes are available in the literature, but also because rectilinear normal modes are uniquely defined.

The redundant set of valence internal coordinates is constructed using the usual procedure outlined by Bakken and Helgaker.[67] Bonded pairs of atoms are defined according to their interatomic distance being less than 1.3 times the sum of the respective covalent radii. The bond angle coordinate is formed by any three atoms A, B and C, where both A and C are bonded to B. The threshold for linearity is set at 165°, and only bond angles less than that are included. Dihedral angles are defined for any set of four atoms A, B, C and D, for which A-B-C and B-C-D bond angles are defined.

The vibrational problem is solved using the PyVCI program, as described in Chapter 3. The Coriolis coupling term is not included in the Watson Hamiltonian, since it does not depend on the potential. The harmonic oscillator basis functions are restricted by the sum of vibrational quantum numbers for a given product, referred to as the excitation level. In this study, the excitation level was set to 6, unless stated otherwise, with no reduction in mode-mode coupling of the basis set. In keeping with our previous nomenclature, this is referred to as VCI(6).

Before the calculation, the force constants are screened and only significant values are retained, using a threshold of 1.0 cm^{-1} in dimensionless normal mode coordinates. We have previously shown that the error from ignoring some force constants does not vary strongly with the excitation level and VCI(4) calculations can be used for its estimation.[68] The mean absolute error (MAE) associated with excluding force constants smaller than 1.0 cm^{-1} is 0.2 cm^{-1} and maximum error is 1.2 cm^{-1} across all molecules in our data set (data available in Supplementary Information). The threshold for storing non-negligible elements of the Hamiltonian matrix was set to 1.0×10^{-8} E_h and should be accurate to better than 3 d.p.

4.4 Results and Discussion

In the following discussion, common reference will be made to the low and high frequency regions, the separation of which we define at 2200 cm^{-1}. Consequently, the high frequency region refers to the fundamentals of modes dominated by stretching motion of XH groups, where X is a heavy atom. The error is defined as the fundamental frequency calculated with the approximate potential minus the corresponding fundamental with the reference potential. For consistency, errors are plotted as a function of harmonic frequencies, ω. In all calculations with quartic force fields (QFF), 1H-pentazole had to be excluded because it diverged after VCI(4).

4.4.1 Errors due to coordinate system

To verify that our completely general procedure for defining curvilinear normal mode coordinates forms an appropriate basis in which the PES may be accurately expanded, fundamental frequencies calculated using this approach are compared to those generated using exact sextic force fields (SFF) generated from the PyPES library. The results are illustrated in Figure 4.1 and summarised in Table 4.1.

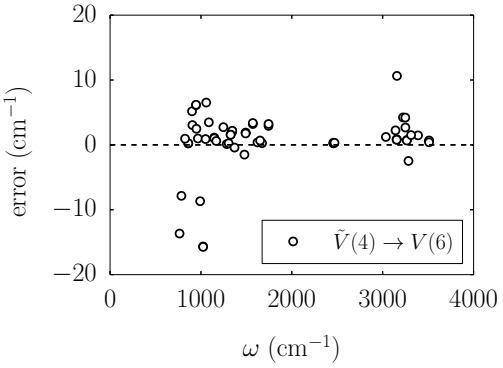


Figure 4.1: Accuracy of $\tilde{V}(4) \rightarrow V(6)$ potential against exact $V(6)$ potential for molecules from PyPES library.

Potentials for CH_4, NH_4^+ and OPH_3 are in the form of quartic force fields in the $\{f_M, \cos(\theta)\}$ internal coordinate system; C_3H_2 and $C_3H_3^+$ are also quartic force fields with $\{f_M, \theta, \tau\}$ and $\{f_M, \cos(\theta), \sin(\omega)\}$ coordinate systems, respectively, where ω is an out-of-plane coordinate; and C_2H_4 is represented as sextic force field in the

Table 4.1: Break down of errors in Figure 4.1 by contributions from each molecule, summarised using mean absolute error (MAE) values.

Molecule	MAE (cm^{-1})
CH_4	1.8
NH_4^+	1.6
OPH_3	0.5
C_3H_2	3.8
$C_3H_3^+$	5.9
C_2H_4	1.3
Aggregate	2.9

$\{f_{\mathrm{M}}, \cos(\theta), \sin(\tau)\}$ coordinate system. In all cases, potentials are transformed into SFFs in normal mode coordinates, $V(6)$, as input into VCI(6) calculations.

Good agreement is observed for all vibrational modes of CH_4, NH_4^+ and OPH_3, with a maximum error of 3.4 cm^{-1} and MAEs of 1.8, 1.6 and 0.5 cm^{-1}, respectively. This shows that the $\{f_{\mathrm{SPF}}, \theta\}$ and $\{f_{\mathrm{M}}, \cos(\theta)\}$ coordinate systems compare well for rigid systems.

The errors are considerably larger for C_3H_2 and $C_3H_3^+$, with MAEs of 3.8 and 5.9 cm^{-1}, respectively. The largest deviations come from torsional modes, showing errors of -7.9 and -8.7 cm^{-1} for C_3H_2, and -13.7 and -15.7 cm^{-1} for $C_3H_3^+$. For C_3H_2, both coordinate sets have the same functional form, except for the stretching coordinates, and the errors are solely due to using a redundant set of dihedral angle coordinates. On the other hand, curvilinear normal modes for $C_3H_3^+$ are constructed from quite different coordinate sets, especially for out of plane motion where a special $\sin(\omega)$ coordinate is used in the reference potential.

The vibrational frequencies of C_2H_4 have a MAE of 1.3 cm^{-1}, with a maximum error of 10.6 cm^{-1} in one of the stretching modes. To discern whether deviations are mainly due to truncation of the potential in $\tilde{Q}$, or use of different coordinate sets, another $V(6)$ potential was generated from the reference force field, by truncating it

Table 4.2: Summary of MAE values for comparison against V(4) potential in Figure 4.2.

	MAE (cm^{-1})
$V(2,4)$	21.8
$V(3,4)$	6.5
$\tilde{V}(2,4) \rightarrow V(4)$	6.0
$\tilde{V}(3,4) \rightarrow V(4)$	0.3

at fourth order. Results of VCI(6) calculations with the new potential are included in the Supplementary Information. They show good agreement with our $\tilde{V}(4) \rightarrow V(6)$ potential, with a MAE of 1.3 cm^{-1} and maximum error of 2.3 cm^{-1}. This indicates that the $\{f_{\mathrm{SPF}}, \theta, \tau\}$ and $\{f_{\mathrm{M}}, \cos(\theta), \sin(\tau)\}$ coordinate sets are equally appropriate for expanding the PEF, and that fifth and sixth order force constants in internal coordinates, that are typically excluded since they are too costly to generate, can still have significant contributions.

4.4.2 Errors due to reduced mode representation of QFFs

In this section, we compare the convergence behaviour of QFFs in both curvilinear and rectilinear normal mode coordinates with respect to mode representation. Rectilinear $V(4)$ potentials are truncated directly, while $\tilde{V}(4)$ potentials are truncated and then transformed to $V(4)$, before solving the nuclear vibrational problem. We note that the untruncated $\tilde{V}(4)$ potential transforms exactly to $V(4)$ and vice versa. Results are summarised in Figure 4.2 and Table 4.2.

Rectilinear normal mode coordinates converge slowly with respect to mode representation level with MAEs of 21.8 and 6.5 cm^{-1} for $V(2,4)$ and $V(3,4)$ potentials, respectively, across all fundamental frequencies of all molecules in our test set. Both potentials have a tendency to underestimate anharmonicity in the low frequency region, while overestimating it in the high frequency region.

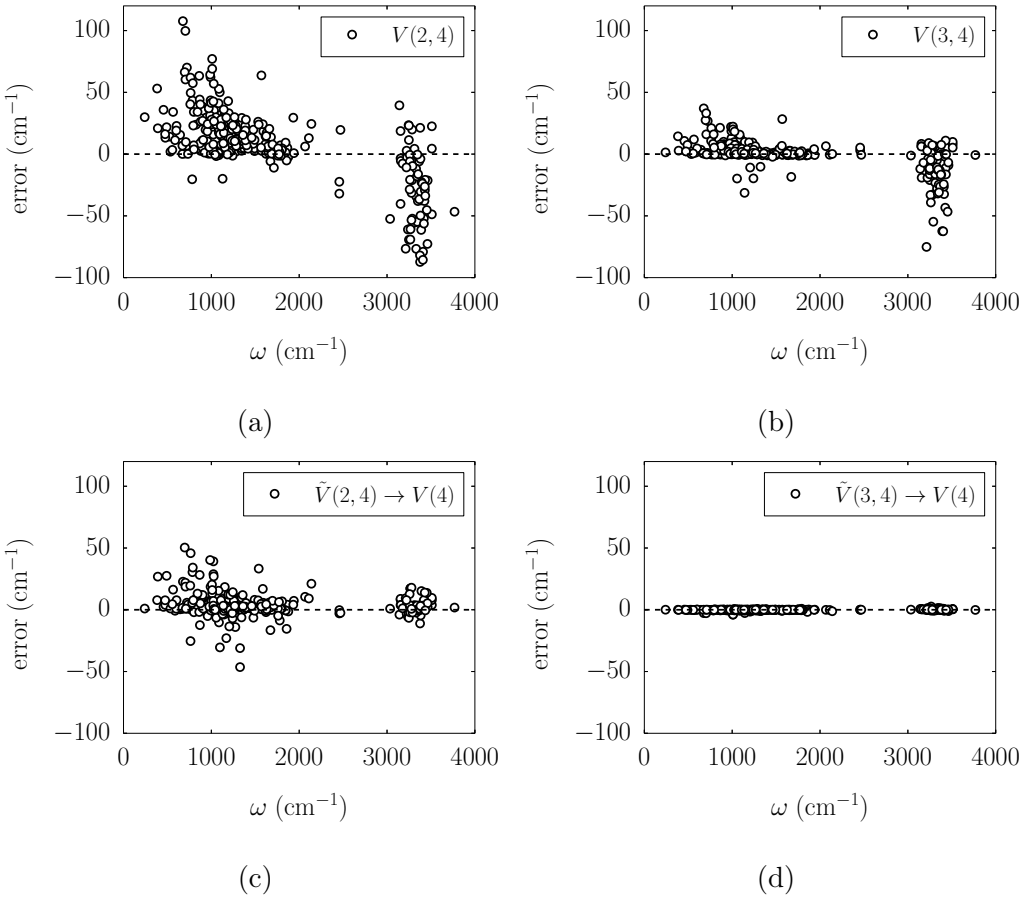


Figure 4.2: Accuracy of 2 and 3 mode representation quartic force fields in rectilinear and curvilinear normal mode coordinates. Subfigures (a), (b), (c) and (d) correspond to $V(2,4)$, $V(3,4)$, $\tilde{V}(2,4) \to V(4)$ and $\tilde{V}(3,4) \to V(4)$ potentials, respectively.

On the other hand, $V(4)$ potentials generated from reduced mode representation $\tilde{V}(4)$ potentials show much faster convergence with respect to mode representation level, with MAEs of 6.0 and 0.3 cm^{-1} for $\tilde{V}(2,4) \rightarrow V(4)$ and $\tilde{V}(3,4) \rightarrow V(4)$ potentials, respectively. Thus, a three mode representation in $\tilde{Q}$ is sufficient to effectively capture all of the mode coupling in $V(4)$, and is a big improvement over the corresponding potential constructed using the same number of *ab initio* calculations in rectilinear space. It is also interesting to note that, with the $\tilde{V}(2,4) \rightarrow V(4)$ potential, all of the fundamentals in the high frequency region are within 20 cm^{-1} of the reference values, but the low frequency region is less well described with deviations as large as 50 cm^{-1}.

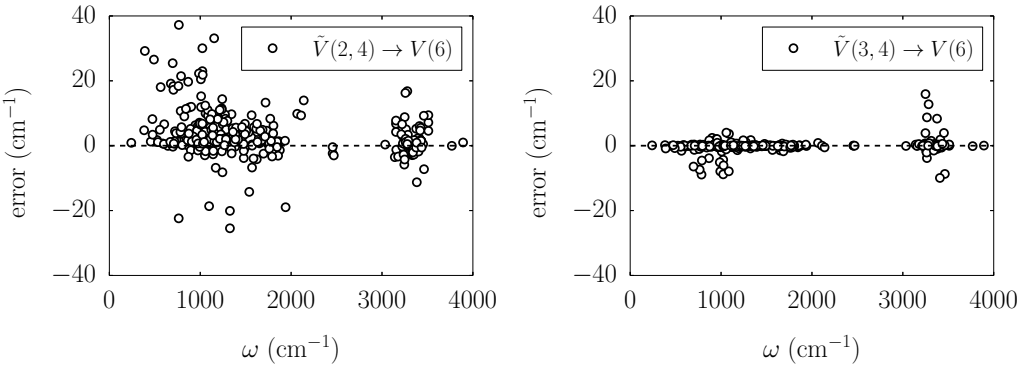
4.4.3 Optimising efficiency in generating SFFs

In theoretical investigations of small molecules, QFFs in internal coordinates have long been used for accurate prediction of fundamental vibrational frequencies.[61,63,69–74] Our previous work implementing a selection of such force fields and other semi-global potentials as part of a larger library, has shown that a sixth order expansion of the potential in normal mode coordinates, $V(6)$, is generally sufficient for describing the fundamentals to within 3 cm^{-1}, benchmarked against more rigorous treatments.[64] Given the fast convergence for the expansion of the potential in $\tilde{Q}$ with respect to its mode representation, it is reasonable to assert that $\tilde{V}(4) \rightarrow V(6)$ potentials can, by analogy, be used to calculate fundamentals with relatively small errors.

The aim of this section is to optimise efficiency for constructing sextic force fields in rectilinear normal mode coordinates, while incurring minimal losses in accuracy, by truncating both curvilinear and rectilinear normal mode potentials at an appropriate mode representation level.

We first look at errors due to truncating curvilinear normal mode potentials, by testing sextic force fields constructed from $\tilde{V}(2,4)$ and $\tilde{V}(3,4)$ against $\tilde{V}(4) \rightarrow V(6)$ potential as a reference. Results are summarised in Figure 4.3 and Table 4.3.

Neglecting 4-mode coupling in $\tilde{Q}$ has little effect, with a MAE of 0.7 cm^{-1}, across all vibrational modes of all molecules in our data set, and most fundamentals within 3 cm^{-1} of the reference values, with only a few difficult cases where deviations up



(a) (b)

Figure 4.3: Accuracy of (a) $\tilde{V}(2,4) \rightarrow V(6)$, and (b) $\tilde{V}(3,4) \rightarrow V(6)$ potentials for calculating fundamental frequencies, compared to $\tilde{V}(4) \rightarrow V(6)$ reference data.

Table 4.3: Summary of MAE values for fundamental frequencies calculated using reduced mode representation potentials compared to $\tilde{V}(4) \rightarrow V(6)$.

	MAE (cm^{-1})
$\tilde{V}(2,4) \rightarrow V(6)$	4.7
$\tilde{V}(3,4) \rightarrow V(6)$	0.7
$\tilde{V}(4) \rightarrow V(3,6)$	4.7
$\tilde{V}(4) \rightarrow V(4,6)$	0.6
$\tilde{V}(4) \rightarrow V(5,6)$	0.03

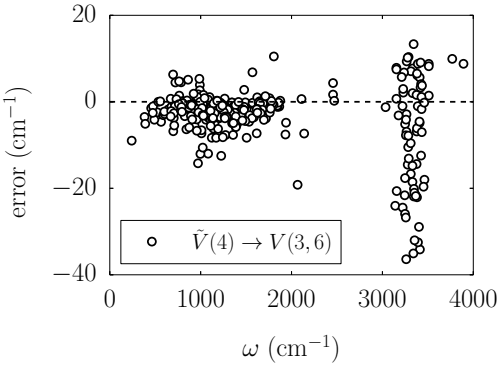
to 15.9 cm^{-1} are observed. The deviations in the low frequency region are due to inaccurate description of some torsional modes, which has a knock-on effect on the high frequency stretches by increasing their coupling to those torsional modes. In particular, the C_3H_2 molecule has deviations of -7.9 and -8.9 cm^{-1} in the torsional modes, which leads to splitting of the wavefunction for high frequency stretches, resulting in errors of 15.9 and 12.8 cm^{-1}.

Two mode representation in $\tilde{Q}$ leads to a MAE value of 4.7 cm^{-1}, across the fundamental frequencies of all molecules in our test set. The anharmonicity in the low frequency region is generally underestimated, with errors up to 40 cm^{-1}. Errors in the high frequency region are more randomly distributed, with a maximum deviation of 16.8 cm^{-1}. Large errors in the low frequency region are due to both bending and torsional modes, and all deviations greater than 20 cm^{-1} come from vibrational modes of only three molecules: cyclopropenium cation, cyclopropene and methylenecyclopropene. Errors in transformed $V(4)$ and $V(6)$ potentials generated from $\tilde{V}(2,4)$ are similar, as a comparison of Figures 4.2(c) and 4.3(a) illustrates. Therefore, the majority of this error is due to neglect of important force constants from the $\tilde{V}(4)$ potential.

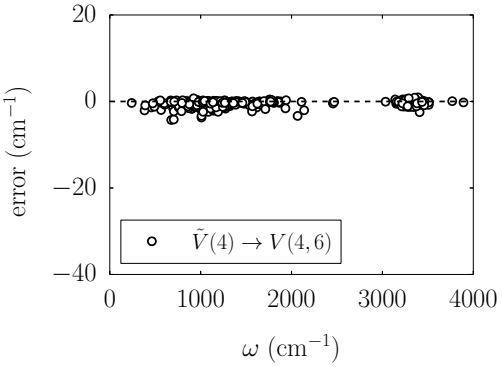
Overall, three mode representation in $\tilde{Q}$ gives the best compromise between computational demand and accuracy, being able to reproduce the fundamentals calculated from four mode representation potentials to within 3 cm^{-1} for all but a handful of difficult cases.

We now look at reducing the cost of the coordinate transformation procedure by investigating the errors introduced from truncation of the rectilinear normal mode potential. $\tilde{V}(4) \rightarrow V(m_r, 6)$ potentials with $m_r = 5$, 4 and 3 were generated and compared to full sextic contributions with $m_r = 6$ and the results are summarised in Figure 4.4 and Table 4.3.

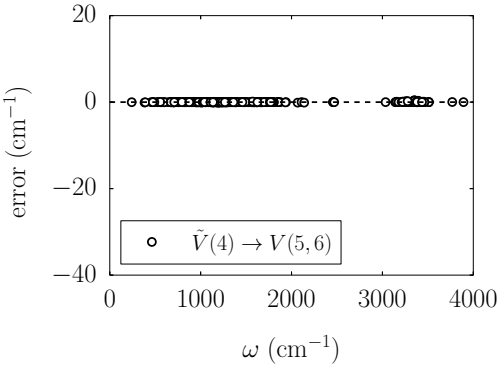
At three mode representation of the SFF in rectilinear normal mode coordiantes, the $\tilde{V}(4) \rightarrow V(3,6)$ potential produces fundamentals with a MAE of 4.7 cm^{-1} and shows large errors in the stretching region with deviations up to 37 cm^{-1}. The low frequency region is much better described, with fundamentals deviating by less than 15 cm^{-1}, apart from a CO stretching mode that appears in this region from cyclopropenone which has an error of -19.2 cm^{-1}, at a harmonic frequency of 2066.9



(a)



(b)



(c)

Figure 4.4: Accuracy of 3, 4 and 5 mode representation approximations in rectilinear coordinates with $\tilde{V}(4) \rightarrow V(6)$ potential. Subfigures (a), (b) and (c) correspond to 3, 4 and 5 mode representation, respectively.

cm^{-1}. Most of the fundamentals in the low frequency region are in error by less than 10 cm^{-1}. Five modes with errors between 10 and 15 cm^{-1} correspond to various torsional modes in azetinone and oxazole, and an NH bending mode in aziridine. The anharmonicity is generally overestimated.

At four mode representation, the $\tilde{V}(4) \rightarrow V(4,6)$ potential results in a MAE of 0.6 cm^{-1} across all modes of all molecules, with all errors less than 5 cm^{-1}. The $\tilde{V}(4) \rightarrow V(5,6)$ potential has a MAE of only 0.03 cm^{-1}, and can be considered converged. However, it is worth noting that while including six mode coupling does not have a significant effect on the fundamental frequencies, it can alter the resonance structure of highly anharmonic modes, the resultant changes in relative contributions of a given harmonic oscillator product will likely become more significant in calculating transition intensities. Overall, at least four mode representation in rectilinear normal modes is required for quantitative assignment of fundamentals.

Finally, we can test the error from approximating both the curvilinear and rectilinear normal mode potentials concurrently. The quality of the $\tilde{V}(2,4) \rightarrow V(4,6)$ and $\tilde{V}(3,4) \rightarrow V(4,6)$ potentials is assessed according to the calculated fundamentals in Figure 4.5 and Table 4.4, using the $\tilde{V}(4) \rightarrow V(6)$ results as a reference. The error is also plotted as a function of anharmonicity,

$$\Delta_{\text{anh}} = \omega - \nu, \tag{4.15}$$

where ν is the frequency from VCI calculation with the reference potential.

In general, the distribution of error for $\tilde{V}(2,4) \rightarrow V(4,6)$ and $\tilde{V}(3,4) \rightarrow V(4,6)$ potentials is largely unaffected by truncation of $V(6)$, with MAEs of 4.5 and 1.0 cm^{-1}, respectively. Plotted as a function of anharmonicity, it is clear that the $\tilde{V}(2,4) \rightarrow V(4,6)$ potential does not always improve on the harmonic approximation in the low frequency region, as indicated by points falling above the black diagonal line and below the red anti-diagonal line ($|\text{error}| > \Delta_{\text{anh}}$). However, it shows good improvement in the high frequency region, with a maximum deviation of 15.8 cm^{-1}. Therefore, this approach could possibly be used for qualitative modelling of high frequency XH stretching modes, where X is a heavy atom, and would be particularly efficient coupled with a lower quality *ab initio* or density functional method for generating the potential energy surface.

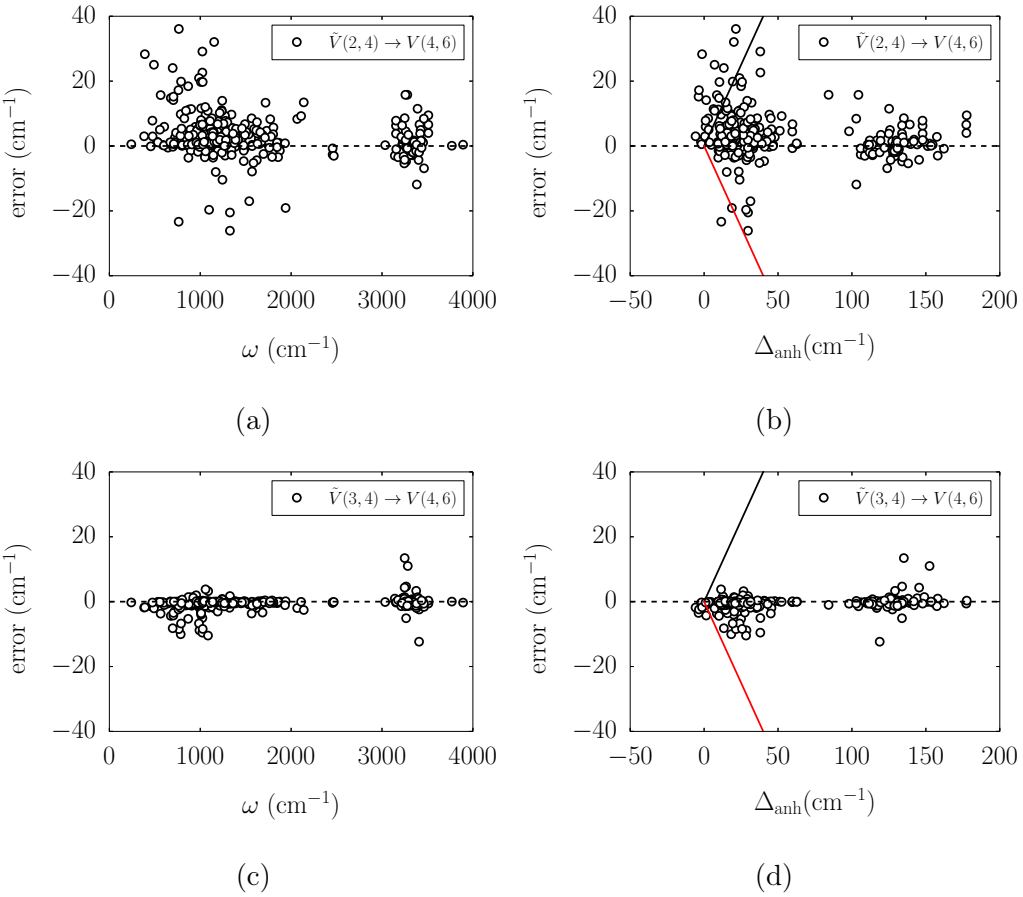


Figure 4.5: Accuracy of (a,b) $\tilde{V}(2,4) \to V(4,6)$ and (c,d) $\tilde{V}(3,4) \to V(4,6)$ potentials for calculating fundamental frequencies with $\tilde{V}(4) \to V(6)$ as a reference. Black and red solid lines correspond to diagonal and anti-diagonal lines, respectively. Anharmonicity, Δ_{anh}, is defined as in equation (4.15).

Table 4.4: Summary of MAE values for potentials compared to $\tilde{V}(4) \to V(6)$, in Figures 4.6 and 4.5.

	MAE (cm^{-1})
$\tilde{V}(2,4) \to V(4,6)$	4.5
$\tilde{V}(3,4) \to V(4,6)$	1.0
$V(3,4)$	8.6
$V(4)$	9.9

The $\tilde{V}(3,4) \rightarrow V(4,6)$ potential produces results that show good agreement across the whole frequency range, with a maximum deviation of 13.5 cm^{-1} and most fundamentals within 5 and 2 cm^{-1} in the high and low frequency ranges, respectively. It can be used, together with a high level *ab initio* method, for accurate prediction of fundamental vibrational frequencies.

4.4.4 Existing methods

It is a common practice to use $V(3,4)$ and $V(4)$ potentials when studying the vibrational structure of molecules,[75–84] usually coupled with methods based on second order vibrational perturbation theory. This is mostly due to the computational cost associated with calculation of higher order force constants, and increasing error at higher orders of numerical differentiation. The quality of these potentials is assessed by the error data illustrated in Figure 4.6 and summarised in Table 4.4, using frequencies calculated based on $\tilde{V}(4) \rightarrow V(6)$ potential as a reference.

Frequencies calculated from both $V(3,4)$ and $V(4)$ potentials show significant deviations from the reference data with MAEs of 8.6 cm^{-1} and 9.9 cm^{-1}, respectively. The low frequency region is better described with the $V(3,4)$ potential, which compensates for its worse performance in the high frequency region, resulting in a smaller MAE.

The improved performance of the $V(3,4)$ potential over $V(4)$ is almost certainly due to cancellation of errors. Ignoring four mode coupling leads to underestimation of anharmonicity relative to $V(4)$ (Figure 4.2(b)), thus cancelling out some of the error arising from ignoring higher order force constants that tends to lead to overcorrection (Figure 4.6(c)). The deviations at low frequencies are particularly troubling when viewed as a function of anharmonicity in Figures 4.6(b) and 4.6(d). A considerable number of points lie below the red anti-diagonal line, which corresponds to overcorrection errors that give frequencies that are worse than the original harmonic estimates, but lower than the true values.

In this context, it is surprising how accurately experimental frequencies can be predicted using semi-quartic force fields, $V(3,4)$, in combination with second order perturbation theory (VPT2). Agreement to within 10.0 cm^{-1} of experiment

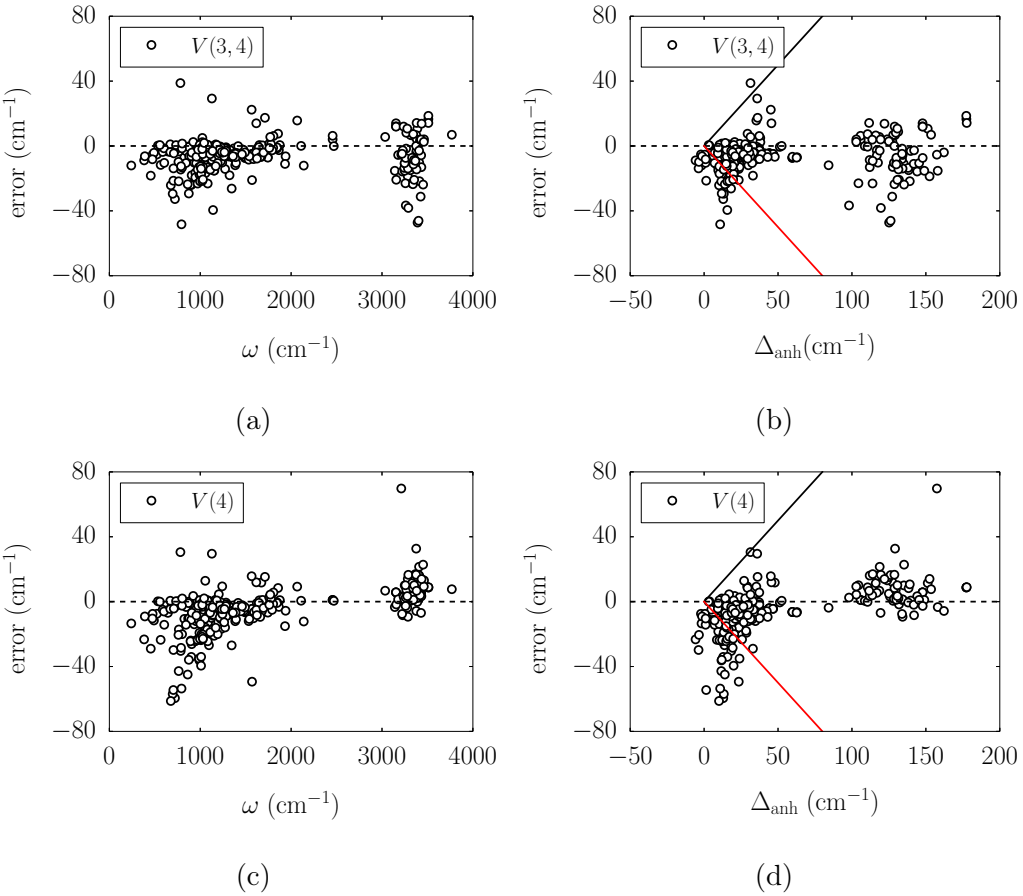


Figure 4.6: Accuracy of (a,b) $V(3,4)$ and (c,d) $V(4)$ potentials for calculating fundamental vibrational frequencies with $\tilde{V}(4) \rightarrow V(6)$ data as a reference. Black and red solid lines correspond to diagonal and anti-diagonal lines, respectively. Anharmonicity, Δ_{anh}, is defined in equation (4.15).

is generally observed for small molecules,[61,63,69–74] provided high quality *ab initio* methods are used to construct the potential and resonances are corrected for. As is common with perturbation theory based methods, this approach suffers from instabilities due to near degenerate states, which require special treatment. This is likely to occur more frequently with larger systems.[85] Nonetheless, black box implementations have been developed and are commonly used.[5,86,87] As such, our conclusions from Figure 4.6 are mainly concerned with more rigorous variational treatments, as in VCI. It would be interesting to investigate whether VPT2 still provides accurate results for larger molecules, with more near degeneracies and more complicated resonance structures with multiple connected resonances, compared to more rigorous variational approaches like VCI, but that is beyond the scope of this work.

4.4.5 Timing and Scaling

The limiting steps in the overall procedure are construction of the $\tilde{V}(m_c, 4)$ potential from *ab initio* data and subsequent transformation to $V(4, 6)$ in rectilinear normal modes. The scaling and feasibility of the first step depends on the method used to solve the electronic structure problem, which is beyond our control. The second step has formal scaling of $O(N_{\text{mode}}^6)$ and $O(N_{\text{mode}}^7)$ operations for $m_c = 2$ and 3, respectively. The calculation can be sped up by pre-screening of the $\tilde{V}(m_c, 4)$ potential, and/or implementing a parallel algorithm. Even more than timing, memory requirements play a big role in determining the feasibility of a computation. Most of the memory during coordinate transformation goes into storing the objects $V(4, 6)$ and $\tilde{L}$, which scale as $O(N_{\text{mode}}^4)$ and $O(N_{\text{mode}}^5)$ in size, respectively.

In order to estimate the range of system sizes that this approach can be effectively used on, blank calculations for the coordinate transformation step, where all operations are performed using dummy data arrays, were performed on a desktop machine with a 2.7 GHz Intel Core i5 processor and 16 GB 1600 MHz DDR3 memory. This corresponds to a worst case scenario where no screening is used. The timings and total memory are summarised in Table 4.5.

This illustrates that both transformations can be performed routinely on a mod-

Table 4.5: Timing and total memory (RAM) requirement for coordinate transformation step of $\tilde{V}(m_c, 4) \rightarrow V(4, 6)$ potential. The memory does not vary strongly with m_c, since the quartic force field is small in comparison to $V(4, 6)$ and $\tilde{L}$, so only a common value rounded to 2 d.p. is reported. Blank calculations were run on a desktop machine with a 2.7 GHz Intel Core i5 processor and 16 GB 1600 MHz DDR3 memory.

	time (hours)		
n_{atoms}	$m_c = 2$	$m_c = 3$	Memory (MB)
10	0.01	0.1	100
15	0.15	3.2	250
20	1.1	30.4	900
25	4.8	182.3	2800

ern desktop computer, within about 5 hours for $m_c = 2$ and 8 days for $m_c = 3$, for a 25 atom system. Screening of the $\tilde{V}(m_c, 4)$ potential can provide significant speed up and the main limitation in studying larger systems is the memory required to store the $\tilde{L}$ derivatives.

4.5 Conclusions

We have implemented and tested a coordinate transformation procedure for interconverting force field expansions between rectilinear and curvilinear normal mode coordinates. Transforming from low order, low mode representation force fields constructed in curvilinear coordinates to higher order, higher mode expansion force fields in rectilinear normal mode coordinates minimises the number of costly *ab initio* calculations that are required to construct the force field but enables the nuclear vibrational problem to be easily solved, as the kinetic energy operator takes a simple form in rectilinear normal mode coordinates. We overall recommended transforming from $\tilde{V}(3, 4)$ to $V(4, 6)$, which we show introduces errors of only 1 cm^{-1}, on average, across our test set of molecules.

This is, to the best of our knowledge, the first completely general procedure that enables sextic force fields in rectilinear normal mode coordinates to be generated for realistically sized molecules, requiring only sufficient *ab initio* calculations to construct the 3MR QFF in curvilinear normal modes.

Although we have chosen to use curvilinear normal mode coordinates for simplicity and ease of implementation in this proof-of-principle study, our coordinate transformation procedure is completely general and will work with any uniquely defined, non-redundant set of curvilinear internal coordinates. Indeed, convergence with respect to the extent of mode coupling in $\tilde{V}$ may be further improved by using coordinates that even more appropriately represent independent modes of molecular motion.

Appendix

Consider two complete and non-redundant sets of coordinates: curvilinear coordinates $\{\eta_1, \eta_2, \ldots, \eta_N\}$, and rectilinear coordinates $\{d_1, d_2, \ldots, d_N\}$. Curvilinear coordinate η_i can be represented as a Taylor series expansion in $\{d_\alpha\}$ according to

$$\eta_i = \sum_\alpha B_{i,\alpha} \Delta d_\alpha + \frac{1}{2!} \sum_{\alpha,\beta} B_{i,\alpha\beta} \Delta d_\alpha \Delta d_\beta + \frac{1}{3!} \sum_{\alpha,\beta,\gamma} B_{i,\alpha\beta\gamma} \Delta d_\alpha \Delta d_\beta \Delta d_\gamma + \cdots , \quad (4.16)$$

where the necessary derivatives calculated at equilibrium are defined as

$$B_{i,\alpha_1 \ldots \alpha_n} = \frac{\partial^n \eta_i}{\partial d_{\alpha_1} \ldots \partial d_{\alpha_n}},$$

and are assumed to be available.

The reverse relationship, representing d_α as a Taylor series in $\{\eta_i\}$ can also be established,

$$d_\alpha = \sum_i A_{\alpha,i} \Delta \eta_i + \frac{1}{2!} \sum_{i,j} A_{\alpha,ij} \Delta \eta_i \Delta \eta_j + \frac{1}{3!} \sum_{i,j,k} A_{\alpha,ijk} \Delta \eta_i \Delta \eta_j \Delta \eta_k + \cdots , \quad (4.17)$$

however the necessary values of

$$A_{\alpha,i_1 \ldots i_n} = \frac{\partial^n d_\alpha}{\partial \eta_{i_1} \ldots \partial \eta_{i_n}},$$

cannot be calculated directly. Instead they can be obtained from elements of B derivatives. This approach has been thoroughly described by Allen and Császár in Ref. [58], and the necessary relationships are summarised below.

$$A_{\alpha,ij} = -\sum_{\beta,\gamma} A_{\beta,i} A_{\gamma,j} F_{\alpha,\beta\gamma}, \tag{4.18}$$

$$A_{\alpha,ijk} = -\sum_{\beta,\gamma} (A_{\beta,i} A_{\gamma,jk} + A_{\beta,j} A_{\gamma,ik} + A_{\beta,k} A_{\gamma,ij}) F_{\alpha,\beta\gamma} - \sum_{\beta,\gamma,\delta} A_{\beta,i} A_{\gamma,j} A_{\delta,k} F_{\alpha,\beta\gamma\delta}, \tag{4.19}$$

where

$$F_{\alpha,\beta\gamma} = \sum_{i} A_{\alpha,i} B_{i,\beta\gamma}, \tag{4.20}$$

$$F_{\alpha,\beta\gamma\delta} = \sum_{i} A_{\alpha,i} B_{i,\beta\gamma\delta}. \tag{4.21}$$

In case of curvilinear normal modes, we replace $\{\eta_i\}$ with $\{\tilde{q}_i\}$ and $\{d_\alpha\}$ with $\{q_\alpha\}$ in the above, and use $\tilde{L}$ instead of B for consistency with the main discussion, than the expressions for A can be simplified by noting that, $\tilde{L}_{i,\alpha} = \delta_{i\alpha}$ where $\delta_{i,\alpha}$ is the Kronecker delta. Leading to,

$$A_{\alpha,ij} = -\tilde{L}_{\alpha,ij}, \tag{4.22}$$

$$A_{\alpha,ijk} = \sum_{\beta} (\tilde{L}_{\beta,jk} \tilde{L}_{\alpha,i\beta} + \tilde{L}_{\beta,ik} \tilde{L}_{\alpha,j\beta} + \tilde{L}_{\beta,ij} \tilde{L}_{\alpha,k\beta}) - \tilde{L}_{\alpha,ijk} \tag{4.23}$$

References

[1] D. Papousek, M. R. Aliev, *Molecular vibrational-rotational spectra*, Elsevier Scientific, Amsterdam, **1982**.

[2] N. Matsunaga, G. Chaban, R. Gerber, *J. Chem. Phys.* **2002**, *117*, 3541–3547.

[3] O. Christiansen, *J. Chem. Phys.* **2003**, *119*, 5773–5781.

[4] V. Barone, *J. Chem. Phys.* **2004**, *120*, 3059–3065.

[5] V. Barone, *J. Chem. Phys.* **2005**, *122*, 014108.

[6] P. Cassam-Chenai, J. Lievin, *Int. J. Quant. Chem.* **2003**, *93*, 245–264.

[7] W. Mizukami, D. P. Tew, *J. Chem. Phys.* **2013**, *139*, 194108.

[8] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Phys. Chem. A* **2012**, *116*, 3691–3709.

[9] J. Bloino, M. Biczysko, V. Barone, *J. Chem. Theor. Comp.* **2012**, *8*, 1015–1036.

[10] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Chem. Phys.* **2014**, *141*, 234114.

[11] A. M. Rosnik, W. F. Polik, *Mol. Phys.* **2014**, *112*, 261–300.

[12] V. Barone, M. Biczysko, J. Bloino, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.

[13] M. Piccardo, J. Bloino, V. Barone, *Int. J. Quant. Chem.* **2015**, *115*, 948–982.

[14] J. L. Davisson, N. R. Brinkmann, W. F. Polik, *Mol. Phys.* **2012**, *110*, 2587–2598.

[15] J. Bowman, *Acc. Chem. Res.* **1986**, *19*, 202–208.

[16] S. Carter, S. Culik, J. Bowman, *J. Chem. Phys.* **1997**, *107*, 10458–10469.

[17] N. Wright, R. Gerber, *J. Chem. Phys.* **2000**, *112*, 2598–2604.

[18] J. Bowman, S. Carter, X. Huang, *Int. Rev. Phys. Chem.* **2003**, *22*, 533–549.

[19] K. Yagi, K. Hirao, T. Taketsugu, M. Schmidt, M. Gordon, *J. Chem. Phys.* **2004**, *121*, 1383–1389.

[20] D. Benoit, *J. Chem. Phys.* **2004**, *120*, 562–573.

[21] O. Christiansen, *J. Chem. Phys.* **2004**, *120*, 2149–2159.

[22] P. Seidler, M. Sparta, O. Christiansen, *J. Chem. Phys.* **2011**, *134*, 054119.

[23] P. Seidler, O. Christiansen, *J. Chem. Phys.* **2009**, *131*, 234109.

[24] S. Carter, N. Handy, *Comp. Phys. Rep.* **1986**, *5*, 115–172.

[25] S. Carter, J. Bowman, N. Handy, *Theor. Chem. Acc.* **1998**, *100*, 191–198.

[26] G. Rauhut, *J. Chem. Phys.* **2007**, *127*, 184109.

[27] D. Bégué, N. Gohaud, C. Pouchan, P. Cassam-Chenai, J. Liévin, *J. Chem. Phys.* **2007**, *127*, 164115.

[28] O. Christiansen, *Phys. Chem. Chem. Phys.* **2007**, *9*, 2942–2953.

[29] O. Christiansen, *Phys. Chem. Chem. Phys.* **2012**, *14*, 6672–6687.

[30] S. Hirata, K. Yagi, *Chem. Phys. Lett.* **2008**, *464*, 123–134.

[31] D. Oschetzki, M. Neff, P. Meier, F. Pfeiffer, G. Rauhut, *Croat. Chem. Acta* **2012**, *85*, 379–390.

[32] A. G. Császár, C. Fabri, T. Szidarovszky, E. Matyus, T. Furtenbacher, G. Czako, *Phys. Chem. Chem. Phys.* **2012**, *14*, 1085–1106.

[33] J. M. Bowman, T. Carrington, H.-D. Meyer, *Mol. Phys.* **2008**, *106*, 2145–2182.

[34] T. K. Roy, R. B. Gerber, *Phys. Chem. Chem. Phys.* **2013**, *15*, 9468–9492.

[35] J. Watson, *Mol. Phys.* **1968**, *15*, 479–490.

[36] J. Watson, *Mol. Phys.* **1970**, *19*, 465–487.

[37] T. Carrington in *Encyclopedia of Computational Chemistry*, *Vol. 5*, (Ed.: P. Schleyer), Wiley, Chichester, **1998**, p. 3157.

[38] A. G. Császár in *Encyclopedia of Computational Chemistry*, *Vol. 1*, (Ed.: P. Schleyer), Wiley, Chichester, **1998**, p. 13.

[39] L. Joubert-Doriol, B. Lasorne, F. Gatti, M. Schröder, O. Vendrell, H.-D. Meyer, *Comp. Theor. Chem.* **2012**, *990*, 75–89.

[40] E. L. Sibert III, *J. Chem. Phys.* **1989**, *90*, 2672–2683.

[41] Y. Zhang, S. Klippenstein, R. Marcus, *J. Chem. Phys.* **1991**, *94*, 7319–7334.

[42] A. Capobianco, R. Borrelli, C. Noce, A. Peluso, *Theor. Chem. Acc.* **2012**, *131*, 1–10.

[43] N. C. Handy, *Mol. Phys.* **1987**, *61*, 207–223.

[44] S. M. Colwell, N. C. Handy, *Mol. Phys.* **1997**, *92*, 317–330.

[45] X. Chapuisat, C. Iung, *Phys. Rev. A* **1992**, *45*, 6217–6235.

[46] J. H. Frederick, C. Woywod, *J. Chem. Phys.* **1999**, *111*, 7255–7271.

[47] J. Makarewicz, A. Skalozub, *Chem. Phys. Lett.* **1999**, *306*, 352–356.

[48] O. Vendrell, F. Gatti, D. Lauvergnat, H.-D. Meyer, *J. Chem. Phys.* **2007**, *127*, 184302.

[49] F. Gatti, C. Iung, *Phys. Rep.* **2009**, *484*, 1–69.

[50] M. Ndong, L. Joubert-Doriol, H.-D. Meyer, A. Nauts, F. Gatti, D. Lauvergnat, *J. Chem. Phys.* **2012**, *136*, 034107.

[51] M. Ndong, A. Nauts, L. Joubert-Doriol, H.-D. Meyer, F. Gatti, D. Lauvergnat, *J. Chem. Phys.* **2013**, *139*, 204107.

[52] F. Richter, F. Thaunay, D. Lauvergnat, P. Carbonnière, *J. Phys. Chem. A* **2015**, *119*, 11719–11728.

[53] H. B. Schlegel, *J. Comput. Chem.* **2003**, *24*, 1514–1527.

[54] H. B. Schlegel, *WIREs Comput. Mol. Sci.* **2011**, *1*, 790–809.

[55] J. Baker, A. Kessi, B. Delley, *J. Chem. Phys.* **1996**, *105*, 192–212.

[56] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, New York, **1955**.

[57] L. Hedberg, I. M. Mills, *J. Mol. Spectrosc.* **1993**, *160*, 117–142.

[58] W. D. Allen, A. G. Császár, *J. Chem. Phys.* **1993**, *98*, 2983–3015.

[59] W. D. Allen, A. G. Császár, V. Szalay, I. M. Mills, *Mol. Phys.* **1996**, *89*, 1213–1221.

[60] J. Kongsted, O. Christiansen, *J. Chem. Phys.* **2006**, *125*, 124108.

[61] C. E. Dateo, T. J. Lee, D. W. Schwenke, *J. Chem. Phys.* **1994**, *101*, 5853–5859.

[62] G. Simons, R. G. Parr, M. J. Finlan, *J. Chem. Phys.* **1973**, *59*, 3229–3234.

[63] R. C. Fortenberry, X. Huang, A. Yachmenev, W. Thiel, T. J. Lee, *Chem. Phys. Lett.* **2013**, *574*, 1–12.

[64] M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* **2015**, *36*, 2200–2207.

[65] D. A. Boese, W. Klopper, J. M. L. Martin, *Mol. Phys.* **2005**, *103*, 863–876.

[66] A. Gambi, C. Puzzarini, G. Cazzoli, L. Dore, P. Palmieri, *Mol. Phys.* **2002**, *100*, 3535–3543.

[67] V. Bakken, T. Helgaker, *J. Chem. Phys.* **2002**, *117*, 9160–9174.

[68] M. Sibaev, D. L. Crittenden, *Comp. Phys. Comm.* **2016**, Submitted.

[69] X. Huang, T. J. Lee, *J. Chem. Phys.* **2008**, *129*, 044312.

[70] R. C. Fortenberry, X. Huang, T. D. Crawford, T. J. Lee, *J. Phys. Chem. A* **2013**, *117*, 9324–9330.

[71] X. Wang, X. Huang, J. M. Bowman, T. J. Lee, *J. Chem. Phys.* **2013**, *139*, 224302.

[72] X. Huang, P. R. Taylor, T. J. Lee, *J. Phys. Chem. A* **2011**, *115*, 5005–5016.

[73] T. J. Lee, J. M. L. Martin, C. E. Dateo, P. R. Taylor, *J. Phys. Chem.* **1995**, *99*, 15858–15863.

[74] T. J. Lee, J. M. Bowman, B. Gazdy, C. E. Dateo, *J. Phys. Chem.* **1993**, *97*, 8937–8943.

[75] K. Yagi, K. Hirao, T. Taketsugu, M. W. Schmidt, M. S. Gordon, *J. Chem. Phys.* **2004**, *121*, 1383.

[76] K. Yagi, S. Hirata, K. Hirao, *Phys. Chem. Chem. Phys.* **2008**, *10*, 1781–1788.

[77] K. Yagi, H. Karasawa, S. Hirata, K. Hirao, *ChemPhysChem* **2009**, *10*, 1442–1444.

[78] P. Daněček, P. Bouř, *J. Comput. Chem.* **2007**, *28*, 1617–1624.

[79] J. Hudecová, V. Profant, P. Novotná, V. Baumruk, M. Urbanová, P. Bouř, *J. Chem. Theory Comput.* **2013**, *9*, 3096–3108.

[80] P. Carbonniere, C. Pouchan, *Theor. Chem. Acc.* **2012**, *131*, 1–8.

[81] P. Carbonniere, C. Pouchan, *Chem. Phys. Lett.* **2008**, *462*, 169–172.

[82] P. Carbonniere, I. Ciofini, C. Adamo, C. Pouchan, *Chem. Phys. Lett.* **2006**, *429*, 52–57.

[83] D. Begue, C. Pouchan, J.-C. Guillemin, A. Benidar, *Theor. Chem. Acc.* **2012**, *131*, 1–11.

[84] C. Pouchan, M. Aouni, D. Bégué, *Chem. Phys. Lett.* **2001**, *334*, 352–356.

[85] S. Krasnoshchekov, N. Craig, P. Boopalachandran, J. Laane, N. Stepanov, *J. Phys. Chem. A* **2015**, *119*, 10706–10723.

[86] V. Barone, M. Biczysko, J. Bloino, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.

[87] V. Barone, M. Biczysko, C. Puzzarini, *Acc. Chem. Res.* **2015**, *48*, 1413–1422.

Chapter 5

Balancing Accuracy and Efficiency in Screened Vibrational Configuration Interaction Calculations

5.1 Introduction

Advances in electronic structure methods and computer technology enable increasingly accurate modelling of electronic properties for ever larger and more complex chemical systems. In contrast to the sustained and ongoing progress in the development and application of methods for solving the electronic Schrödinger equation, harmonic normal mode analysis remains the only widely used method for solving the nuclear Schrödinger equation for realistically sized chemical systems.

This can largely be attributed to the difficulty and computational cost associated with modelling how the energy changes as the molecule vibrates, *i.e.* constructing multidimensional anharmonic potential energy surfaces (PES). As already described, we have recently implemented a coordinate transformation procedure to overcome this problem, based upon obtaining concise, low order PES expansions in internal coordinates and then transforming to lengthier, higher order expansions in normal mode coordinates, to facilitate solving the nuclear vibrational problem. This minimizes the number of computationally intensive *ab initio* calculations required to accurately describe the PES.

Therefore, we now turn our attention to solving the nuclear vibrational problem in a general, scalable and accurate manner. Any two of these constraints can be readily satisfied by standard methods; harmonic normal mode analysis[1] and vibrational self-consistent field theory (VSCF)[2–9] are general and scalable, but not particularly accurate, vibrational configuration interaction (VCI)[9–11] and vibrational coupled cluster (VCC)[9,12–15] methods are accurate and general, but scale badly with molecular size. Methods based upon expressing both the kinetic energy operator and potential energy surface in internal coordinates[16,17] are highly accurate and potentially scalable, but lack generality. Hence, in our opinion normal mode coordinates are the only practical choice for a black box nuclear vibrational structure model.[18]

Second-order vibrational perturbation theory (VPT2)[9,19–29] displays a promising balance of generality, scalability and accuracy, but naïve implementations suffer from numerical instability due to accidental near-degeneracies leading to divergence of the VPT2 energy expression. More sophisticated implementations avoid this problem by constructing and diagonalizing VCI sub-matrices for sets of resonant frequencies. However, the highly efficient analytical implementations that rely on pre-determining sets of resonant states do not always capture all relevant states and their couplings.[30,31] More rigorous methods have recently been developed that identify resonant states using an approximate anharmonic wavefunction ansatz.[23,32] While these methods are expected to be more reliable, they are still not designed to converge to the full VCI limit.

An alternative, although less popular,[33–38] way of combining VCI and VPT2 is to first perform a VCI calculation and then apply a VPT2 correction. This is conceptually appealing, as VCI naturally captures strong resonances, and VPT2 is designed to capture small perturbations from a reference state. Combining these two approaches in a simple but robust iterative procedure enables strongly coupled terms identified by large contributions to the VPT2 energy expression to be included in the VCI matrix, ensuring convergence to the VCI limit.[34–37]

Even though this procedure reduces the number of elements explicitly included in the VCI matrix, previous studies have found that memory requirements can still become prohibitive, even when performing case studies on small molecules. Therefore, a range of strategies have been employed to reduce the number of VCI

matrix elements calculated and stored, including; using a state-specific approach that involves performing separate VCI calculations for each vibrational state of interest,[34,36,37] restricting the maximum excitation level, below the convergence limit,[34,36,38] restricting the extent of mode-coupling in excited states, particularly for larger molecules,[35,37] and mode-by-mode customization of the VCI basis.[33] Disadvantages associated with these strategies include increasing run-time, decreasing accuracy or forfeiting accessibility by requiring expert user input.

Here, we address the memory bottleneck problem differently, using a modified VPT2 screening procedure to minimize the number of configurations selected, and using sparse matrices to store only non-negligible matrix elements, in conjunction with sparse matrix diagonalization routines. The efficiency of this process enables us to quantify memory requirements as a function of molecule size across a test set of 45 molecules, ranging in size from 3 - 10 atoms, and establish the computational scaling behaviour.

Previous studies have also indicated that time requirements can also become substantial.[33–38] To reduce overall run time, the most commonly employed strategy is to decrease the time required to generate and evaluate the potential energy surface, either by limiting the dimensionality of the potential energy surface and/or using a truncated series expansion to represent the PES, typically a quartic force field.[33–38] However, not only does this decrease accuracy, but it can destabilize the VCI wavefunction at higher excitation orders, leading to divergence in calculated frequencies.[39] Therefore, in the present work, we employ sextic force fields throughout,[40] and explicitly quantify the effect of using reduced mode representations on the accuracy of calculated fundamentals and time required.

The overall aim of the present work is to establish optimal user-adjustable parameters controlling the description of the PES, and the construction of the wavefunction that minimize errors for a given computational cost, eliminating errors due to premature truncation of the wavefunction and/or force field expansion. To achieve this, we quantify errors and computational resource requirements associated with each approximation, truncation or numerical screening procedure used during VPT2-screened VCI calculations to the convergence limit, and explore the optimal balance between including configurations in the VCI expansion explicitly or

capturing their contribution using a post-hoc VPT2 correction.

5.2 Theory and algorithms

5.2.1 VPT2

The vibrational Hamiltonian must be partitioned into a primary analytically solvable component and a secondary weak perturbation. We follow the usual approach of separating out the kinetic energy and harmonic potential terms, leaving the anharmonic potential correction and, optionally, the Coriolis coupling operator.

$$\hat{H} = \hat{H}_0 + \Delta\hat{H} \tag{5.1}$$

$$\hat{H}_0 = -\frac{1}{2}\sum_{i=1}^{M}\frac{\partial^2}{\partial Q_i^2} + \frac{1}{2!}\sum_{i,j} F_{ij}Q_iQ_j \tag{5.2}$$

$$\begin{aligned}\Delta\hat{H} = &\frac{1}{3!}\sum_{i,j,k} F_{ijk}Q_iQ_jQ_k + \frac{1}{4!}\sum_{i,j,k,l} F_{ijklmn}Q_iQ_jQ_kQ_l \\ &+ \frac{1}{5!}\sum_{i,j,k,l,m} F_{ijklm}Q_iQ_jQ_kQ_lQ_m \\ &+ \frac{1}{6!}\sum_{i,j,k,l,m,n} F_{ijklmn}Q_iQ_jQ_kQ_lQ_mQ_n + \left(\hat{H}_{\text{Cor}}\right)\end{aligned} \tag{5.3}$$

$$\hat{H}_{\text{Cor}} = -\sum_{\alpha} B_\alpha \sum_{i<j}\sum_{k<l} \zeta_{ij}^{\alpha}\zeta_{kl}^{\alpha}\left(Q_i\frac{\partial}{\partial Q_j} - Q_j\frac{\partial}{\partial Q_i}\right)\left(Q_k\frac{\partial}{\partial Q_l} - Q_l\frac{\partial}{\partial Q_k}\right) \tag{5.4}$$

For computational expedience, basis states are constructed as Hartree products of harmonic oscillator basis functions:

$$\Phi_{\mathbf{n}}(Q_1, ..., Q_M) = \prod_{i=1}^{M}\phi_{n_i}(Q_i) \tag{5.5}$$

where $\mathbf{n}$ is a string of quantum numbers $n_1, ..., n_i, ...n_M$, specifying the vibrational state across all M modes. The strings that define the complete set of possible VPT2 and VCI basis states are generated by specifying a maximum value for the sum of the vibrational quantum numbers, which will henceforth be referred to as the excitation level, with its value denoted in round brackets, *e.g.* VCI(8) matrix is indexed by configurations with a sum of vibrational quantum numbers of 8 or less.

The VCI wavefunction is expressed as a linear combination of basis states:

$$\Psi_{\mathbf{n}}(Q_1, ..., Q_M) = \sum_{n'} c_{\mathbf{n},\mathbf{n}'}\Phi_{\mathbf{n}'}(Q_1, ..., Q_M) \tag{5.6}$$

with associated energies $\epsilon_{\mathbf{n}}$. We have previously described in detail how VCI matrix elements are evaluated and coefficients obtained.[39]

The harmonic energy levels for each Hartree product basis state $\Phi_{\mathbf{n'}}(Q_1, ..., Q_M)$ are denoted by $\epsilon_{\mathbf{n'},0}$. These are calculated by the usual harmonic normal mode analysis procedure.[39]

It is also helpful to partition the set of all excited Hartree product basis states defined by a given maximum excitation level into mutually exclusive VCI and VPT2 expansion sets.

The general expression for the second order perturbation theory energy correction is then:

$$\Delta\epsilon_{\mathbf{n}} = \sum_{\mathbf{n'}} \frac{|\langle \Psi_{\mathbf{n}}(Q_1, ..., Q_M)| \Delta\hat{H} |\Phi_{\mathbf{n'}}(Q_1, ..., Q_M)\rangle|^2}{\epsilon_{\mathbf{n}} - \epsilon_{\mathbf{n'},0}} \tag{5.7}$$

The sum runs over all excited states in the VPT2 expansion set but not the VCI set. This reduces to the usual non-degenerate VPT2 energy expression if the unperturbed wavefunction is harmonic *i.e.* $\Psi_{\mathbf{n}} = \Phi_{\mathbf{n}}$, $\epsilon_{\mathbf{n}} = \epsilon_{\mathbf{n},0}$ and the VPT2 expansion set contains all harmonic excited states except $\mathbf{n}$.

For future reference, we define the VPT2 pair energy, $\xi_{\mathbf{n},\mathbf{n'}}$, as the contribution each VPT2 basis state, $\mathbf{n'}$, makes to the overall VPT2 energy correction for a given VCI state, $\mathbf{n}$:

$$\xi_{\mathbf{n},\mathbf{n'}} = \frac{|\langle \Psi_{\mathbf{n}}(Q_1, ..., Q_M)| \Delta\hat{H} |\Phi_{\mathbf{n'}}(Q_1, ..., Q_M)\rangle|^2}{\epsilon_{\mathbf{n}} - \epsilon_{\mathbf{n'},0}} \tag{5.8}$$

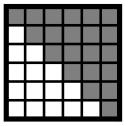
5.2.2 VPT2 based screening

The primary aim of our screening algorithm is to select a VCI sub-matrix that contains only the configurations required to describe the ground and fundamental excited states of the system *i.e.* configurations that couple strongly to each of these states, or are responsible for strong coupling between them. In the process, we also determine VPT2 corrections for each state for each VCI sub-matrix generated.

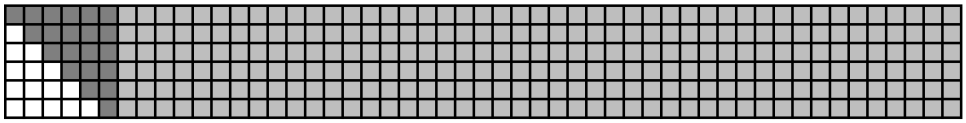
Our iterative VPT2 based screening algorithm proceeds as described below. In the illustrations, each grid point represents an individual VCI matrix element, dark grey shading denotes elements explicitly selected or generated for inclusion in VCI

sub-matrix, light grey points represent elements calculated during VPT2 screening which may be subsequently also included in the VCI sub-matrix, and a thick black line denotes a VCI sub-matrix to be diagonalized.

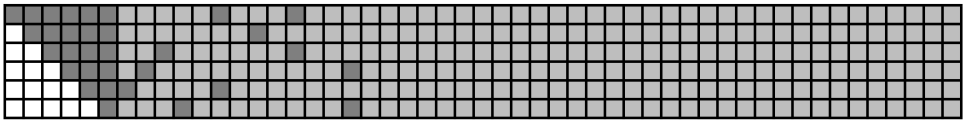
1. Generate the upper triangular elements of the square and symmetric VCI(1) sub-matrix, of leading dimension $N_{\text{inc}} = M + 1$, and diagonalize to obtain the initial VCI wavefunction for the ground and fundamental excited states.



2. Generate the remaining unique elements in the top $N_{\text{inc}} \times N_{\text{total}}$ block of the full VCI matrix.

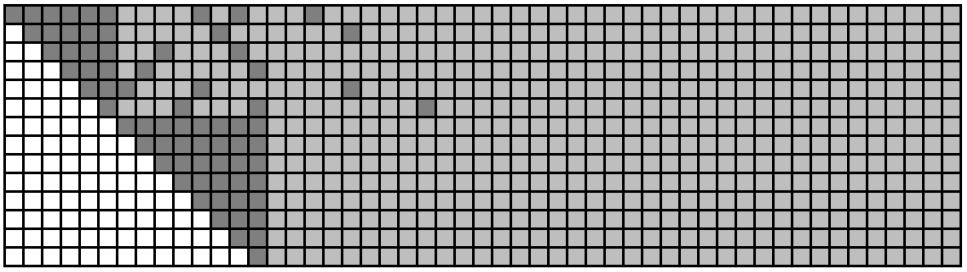


3. Select configurations to include in the VCI sub-matrix to be diagonalized based upon the VPT2 pair energy (5.8) exceeding a pre-set threshold *for the ground state, fundamental excited states and states strongly resonant with fundamentals*, noting that the full VCI matrix elements are also intermediates in the evaluation of the VPT2 pair energy, as $\langle \Psi_{\mathbf{n}} | \, \Delta\hat{H} \, | \Phi_{\mathbf{n'}} \rangle = \langle \Psi_{\mathbf{n}} | \, \hat{H} \, | \Phi_{\mathbf{n'}} \rangle$ if $\mathbf{n} \neq \mathbf{n'}$.



4. Add new configurations to VCI sub-matrix, $N_{\text{inc}} = N_{\text{inc}} + N_{\text{selected}}$, and diagonalize to update VCI wavefunction.

5. Repeat steps 3 and 4 until no new configurations are selected.



6. Repeat steps 3, 4 and 5, decreasing the screening threshold, until fundamental VPT2-corrected VCI frequencies are converged.

In our algorithm, a resonant state is defined as any state that has a squared coefficient value from one of the fundamentals in its VCI wavefunction larger than a pre-set threshold, here set to 0.15. When strong resonances occur, assignment of the fundamental becomes unclear and inclusion of resonant states in the VCI matrix during each VPT2 step ensures that all relevant basis functions are selected.

Sequentially lowering the VPT2 screening threshold value ensures that the VCI wavefunction is iteratively improved so that it provides an appropriately accurate reference state for selecting new configurations for inclusion at each VPT2 step. Similarly, multiple configuration selection iterations are carried out at each VPT2 screening threshold, again to ensure that all relevant strongly coupled configurations - and no irrelevant ones - are included in the VCI matrix. This double iterative procedure minimizes the number of configurations included in the VCI matrix to be diagonalized, and therefore results in a more memory efficient algorithm.

5.3 Methods

Our test set comprises 45 polyatomic molecules containing up to 10 atoms for which sextic force fields (SFFs) in normal mode coordinates are available[40–42]. For smaller molecules (3-6 atoms), accurate SFFs of at least CCSD(T)/cc-pVTZ quality have been compiled from the literature[40,41]. For larger molecules (6-10 atoms), HF/6-311G** quality SFFs in normal mode coordinates have been generated via a coordinate transformation procedure from quartic force fields (QFFs) constructed in curvilinear internal coordinates.[42] A full description of the coordinate transformation and numerical differentiation procedures can be found in Chapter 4. Sixth-order expansions in normal mode coordinates ensure that the PES is accurately represented in the energy regime relevant to calculating fundamental vibrational frequencies.[42] Full details of the potential energy surfaces for all molecules in our data set are available as Supporting Information.

Benchmark VCI results are generated by including in the VCI expansion all configurations with a specified sum of vibrational quantum numbers to a maximum of 6, for all molecules with 8 or fewer atoms. We have previously shown that VCI(6) predicts fundamental frequencies to well within 1 cm^{-1} of benchmark values, on average, across a chemically diverse data set.[39]. For 9 and 10 atom molecules, the time and memory requirements for the naïve VCI(6) algorithm become prohibitive, and it is necessary to explore strategies for decreasing the number of VCI matrix elements that must be calculated and stored.

The first, and simplest strategy, is to restrict the number of vibrational modes that may be concurrently excited when forming the VCI basis states. Here, we implement reduced mode coupling by excluding all configurations with more than 3, 4 or 5 modes concurrently excited, regardless of the total sum of vibrational quantum numbers for each basis state.

VPT2-based screening is then carried out using the algorithm described above, iterating over different choices of screening threshold ranging from 1×10^{-4} to 1×10^{-7} E_h, tabulated in Supporting Information. To minimize memory requirements, all non-negligible VCI matrix elements ($> 1 \times 10^{-8}$ E_h) are stored in sparse matrix format and the VCI sub-matrix is diagonalized using the sparse matrix routines

implemented within SciPy.

To further reduce run time, a reduced mode representation of the potential may be used. In previous work, we have demonstrated that omitting force constants with more than 4 unique indices from the SFF does not substantially change the calculated fundamental frequencies, introducing errors of less than 1 cm^{-1}, on average. We also used a threshold of 1 cm^{-1} to screen the force constants, as it was previously shown to incorporate mean error of only 0.2 cm^{-1}.[42] Although errors due to truncating the PES expansion and wavefunction expansion are expected to be only weakly correlated, we test this explicitly by repeating reduce mode coupling calculations and VPT2-based screening calculations using 4 mode representation sextic force fields.

Once an optimal force constant and wavefunction screening protocol has been established, it is deployed to carry out reduced mode-coupling, VPT2-screened VCI(6) calculations for all molecules in our test set, denoted $VCI_{scr}(6,m\text{MC})$, where m = mode coupling level. The same calculation with a VPT2 correction applied to account for all configurations not explicitly included in the screened VCI wavefunction is denoted $VCI_{scr}(6,m\text{MC})$+VPT2. If $m = 6$, *i.e.* no reduction in mode coupling is applied, the mode coupling specifier is omitted, *e.g.* $VCI_{scr}(6)$.

The accuracy of screened VCI(6) calculations, with or without VPT2 corrections, for molecules containing up to 8 atoms is established by comparison with unscreened, full SFF VCI(6) calculations. Although the accuracy of screened VCI(6) calculations on 9 and 10 atom systems cannot be directly established, the results of screened VCI(6) calculations are nonetheless important for quantifying scaling of computational resource requirement.

Computational efficiency is assessed by comparing number of VCI matrix elements evaluated during a screened calculation compared to unscreened to quantify CPU time savings, and analogous statistics on stored VCI matrix elements to quantify the reduction in memory requirements.

5.4 Results & Discussion

5.4.1 Reduced mode coupling in VCI expansions

Errors in VCI(6) anharmonic frequencies due to applying reduced mode coupling approximations are illustrated in Figure 5.1 and summarized in Table 5.1. Errors are calculated by subtracting approximate frequencies from reference data obtained at VCI(6) with a full SFF.

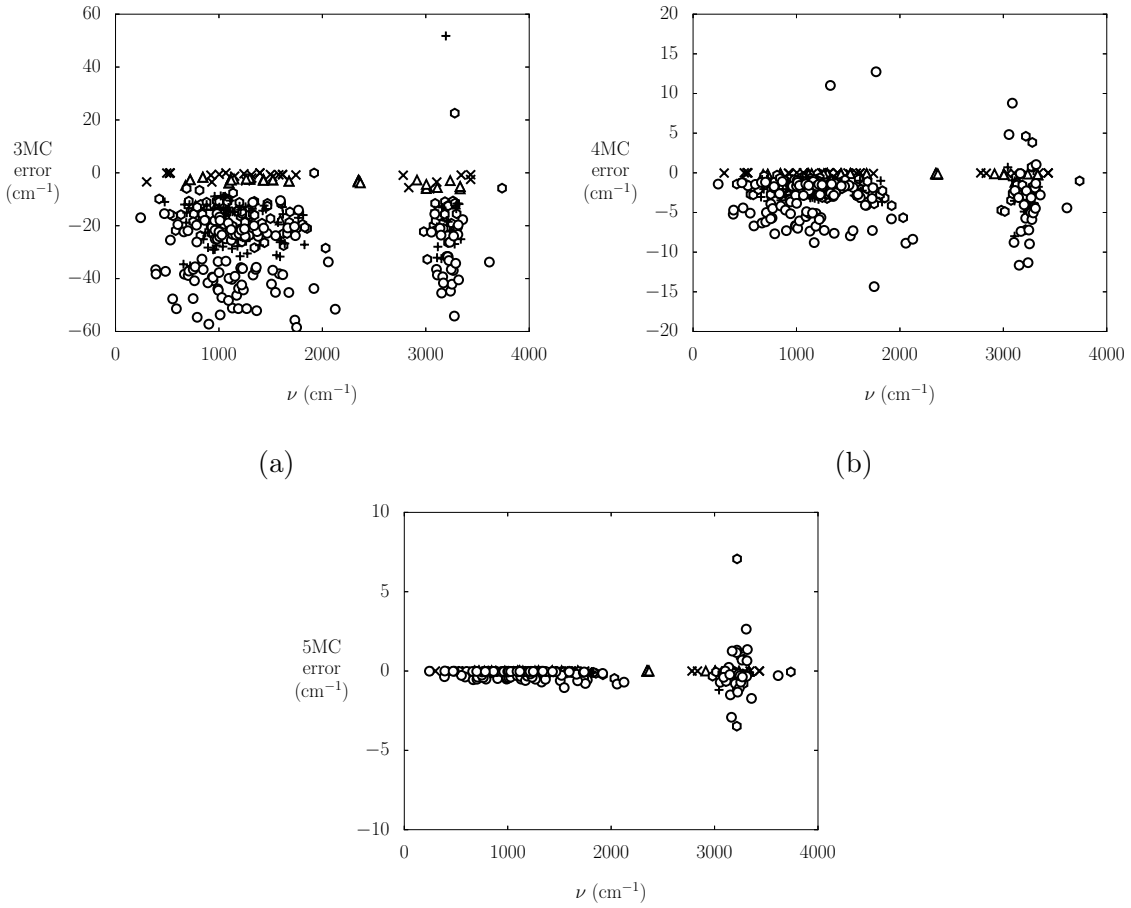


Figure 5.1: Errors in VCI(6) fundamental frequencies resulting from reducing the extent of mode-coupling in the wavefunction: a) 3-mode coupling; b) 4-mode coupling and c) 5-mode coupling, marked according to molecular size: 4 atoms (×), 5 atoms (△), 6 atoms(⬡), 7 atoms (+), 8 atoms(○).

The data illustrated in Figure 5.1 show that errors decrease as the extent of mode coupling increases, as expected. The error distribution also changes. At lower mode

Table 5.1: Mean absolute errors (cm^{-1}) in fundamental vibrational frequencies as a function of molecule size, due to reduced mode coupling (MC) in the VCI(6) wavefunction.

n_{atom}	3MC	4MC	5MC
4	1.5	0.04	0.00004
5	5.3	0.5	0.01
6	16.8	2.1	0.2
7	19.9	2.9	0.1
8	32.9	4.2	0.3

coupling levels, the majority of errors are < 0, *i.e.* the approximate frequencies are larger than the reference values. This indicates that the anharmonicity in each fundamental frequency has not been completely captured.

At 5 mode coupling, the errors become more randomly distributed, suggesting that the remaining errors are due to neglect of indirect coupling. The large deviations at higher frequencies are due to coupling with resonant states, which are difficult to capture correctly. Errors due to resonant states that result in underestimation of anharmonicities are less pronounced at 3 and 4 MC, mainly because they are of the same magnitude as errors arising from non-resonant states. However, errors in resonant states that lead to overestimation of anharmonicities (error > 0) can clearly be distinguished in all panels of Figure 5.1.

Upon inspection of Figure 5.1, it seems that larger molecules tend to have larger errors, particularly at lower mode coupling levels. This effect is quantified in the mean absolute error data presented in Table 5.1.

From Table 5.1, it is clear that 3MC errors are both large and strongly correlated with molecular size. Therefore, despite the significant reduction in computational cost due to truncating the VCI wavefunction at 3 mode coupling, the deteriorating accuracy of results as a function of molecule size makes this approach inappropriate for larger molecules.

On the other hand, 5 mode coupling truncation of the wavefunction affords

Table 5.2: Mean absolute errors (cm^{-1}) in fundamental vibrational frequencies for all molecules containing up to 8 atoms, resulting from VPT2 screening of the VCI wavefunction terminating at a range of different screening threshold values, both with and without a subsequent VPT2 correction accounting for weakly coupled configurations not explicitly included in the VCI wavefunction.

Screening threshold (E_h)	$VCI_{scr}(6)$	$VCI_{scr}(6)$ + VPT2
1×10^{-4}	65.5	10.1
1×10^{-5}	24.9	3.6
1×10^{-6}	10.2	1.3
1×10^{-7}	4.2	0.4

highly accurate fundamental frequencies with errors that are only weakly dependent on molecule size. However, the efficiency gains from using 5MC are modest, so do not significantly extend the limits of applicability beyond the molecules that may otherwise be modelled using full VCI(6).

Therefore, we overall recommend 4 mode coupling pre-screening of the VCI wavefunction for an acceptable balance between accuracy loss and efficiency gain. However, we do note that this can introduce up to 15 cm^{-1} errors in fundamental frequencies in the worse case scenario of highly anharmonic, strongly coupled and resonant modes. This approximation also routinely introduces errors around 5 cm^{-1}, particularly for larger molecules.

5.4.2 VPT2 screening

Mean absolute errors in VPT2-screened VCI frequencies, with an optional VPT2 correction, are presented as a function of final screening threshold in Table 5.2. In both cases, errors in calculated fundamental frequencies decrease as the screening threshold decreases and more configurations are included in the VCI sub-matrix.

However, as the VPT2 correction provides substantially more accurate results for a minimal increase in computational cost, we recommend that it is always applied.

Therefore, all subsequent VPT2 screening results presented in this work will include the VPT2 correction. We also note, in passing, that the VPT2 correction does not alter the rate of convergence with respect to screening threshold, but rather the associated pre-factor.

To assess the scalability of the VPT2 screening approach implemented here, it is useful to examine the relationship between error and molecule size, as illustrated in Figure 5.2, with associated summary statistics presented in Table 5.3.

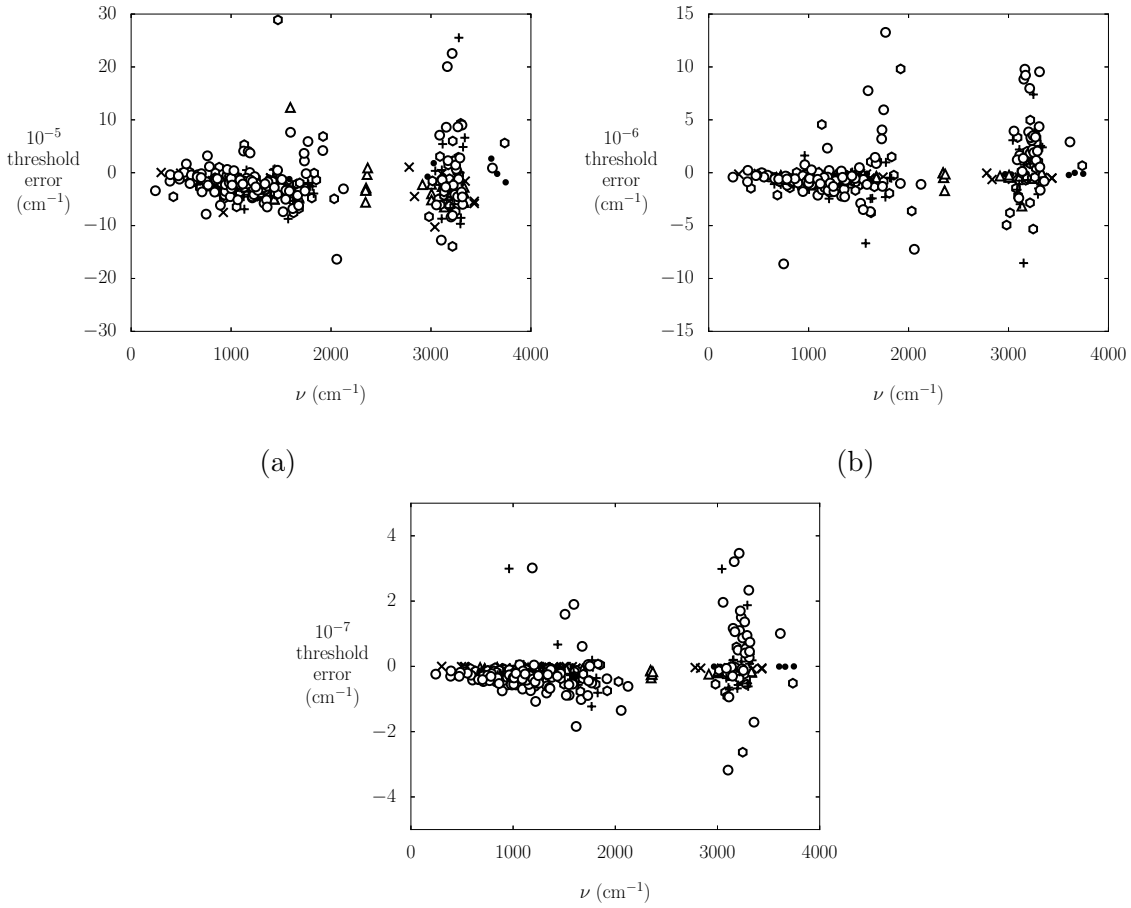


Figure 5.2: Errors in $\mathrm{VCI_{scr}}(6)$ + VPT2 with different thresholds at which the VPT2 screening procedure is terminated: a) 1×10^{-5} $\mathrm{E_h}$; b) 1×10^{-6} $\mathrm{E_h}$ and c) 1×10^{-7} $\mathrm{E_h}$. Data points are marked according to molecular size: 3 atoms (•), 4 atoms (×), 5 atoms (△), 6 atoms(⬡), 7 atoms (+), 8 atoms(○).

Unlike the errors associated with *a priori* reduction in mode coupling, VPT2 screening errors are randomly distributed for all choices of screening threshold value.

Table 5.3: Breakdown of mean absolute errors (cm^{-1}) in fundamental frequencies by molecule size. Frequencies are calculated using $VCI_{scr}(6)$ + VPT2, with a range of different final screening threshold values.

	Screening threshold		
n_{atom}	10^{-5} E_h	10^{-6} E_h	10^{-7} E_h
3	1.6	0.1	0.005
4	2.7	0.3	0.03
5	3.4	0.5	0.1
6	3.8	1.3	0.3
7	3.8	1.4	0.4
8	3.7	1.9	0.6

Further, at the highest acceptable screening threshold of 10^{-5} E_h, errors are only weakly related to molecular size, which makes this approach a particularly attractive way of reducing both the time and memory demands of a full VCI(6) calculation.

However, it should be noted that maximum errors associated with applying the 10^{-5} E_h screening threshold can be significant, with outliers in Figure 5.2 at up to 30 cm^{-1}, which are due to states with strong resonances. It may be safer to use a screening threshold of 10^{-6} E_h, which predicts fundamental frequencies to within 15 cm^{-1} in all cases, in line with the maximum error associated with applying the 4 mode coupling approximation. However, this will come at increased computational cost, as discussed in more detail later.

Using a screening threshold of 10^{-7} E_h produces the most accurate results, but this gain in accuracy is not worth the associated increase in computational cost. As using this screening threshold will not be practical for larger molecules, we do not investigate it any further here.

5.4.3 Reduced mode representation of the potential

We have previously shown that reducing the mode representation (MR) of potential energy surface expansions is an effective method for decreasing computational run

Table 5.4: Mean absolute errors (cm^{-1}) in fundamental frequencies as a function of molecule size, excluding triatomics, using 4 mode coupling in the VCI(6) wavefunction, VCI(6,4MC), with both full (6 mode) and reduced (4 mode) representations of the sextic force field potential.

	VCI(6,4MC)	
n_{atom}	Full SFF	4MR SFF
4	0.04	0.05
5	0.5	0.6
6	2.1	2.5
7	2.9	3.2
8	4.2	4.9
Aggregate	2.6	2.9

time without significantly sacrificing accuracy (see Chapter 4). Here, we investigate the extent of coupling between wavefunction truncation and PES truncation errors.

For the larger molecules in our data set (6-10 atoms), we employ a coordinate transformation procedure to obtain 4MR SFFs in normal coordinates from 3MR QFFs constructed in curvilinear internal coordinates.[42] This approach minimizes the computational cost of both generating the *ab initio* data required to construct the force field, and performing the coordinate transformation process, both of which become particularly important for larger molecules.

For the smaller molecules (4-6 atoms, excluding triatomics for which the reduced mode representation makes no change) in our data set, 4MR SFFs are generated more directly by simply truncating the SFF expansion to include only force constants with at most 4 different indices.

The data presented in Tables 5.4 and 5.5 confirm that the 4MR SFF provides an excellent approximation to the full SFF at much reduced computational cost. Further, we confirm that errors due to truncation of the force field expansion are not amplified by also truncating the wavefunction expansion.

Table 5.5: Mean absolute errors (cm^{-1}) in fundamental frequencies as a function of molecule size, excluding triatomics, using $VCI_{scr}(6)$ (screening threshold = 10^{-5} and 10^{-6} E_h) + VPT2, with both full (6 mode) and reduced (4 mode) representations of the sextic force field potential.

	$VCI_{scr}(6)$+VPT2, 10^{-5} E_h threshold		$VCI_{scr}(6)$+VPT2, 10^{-6} E_h threshold	
n_{atom}	Full SFF	4MR SFF	Full SFF	4MR SFF
4	2.7	2.7	0.3	0.3
5	3.4	3.2	0.5	0.5
6	3.8	3.9	1.3	1.4
7	3.8	3.2	1.4	1.3
8	3.7	3.9	1.9	1.9
Aggregate	3.6	3.6	1.3	1.4

5.4.4 Combined screening algorithm - accuracy and computational scaling

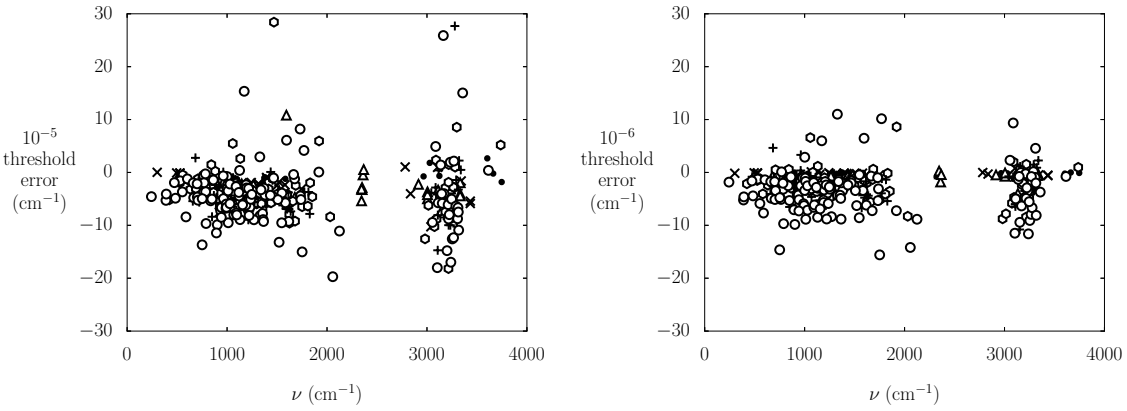
The data presented above allow us to make some clear and unambiguous choices that reduce computational cost with minimal accuracy loss. Throughout the remaining results and discussion, the PES will always be represented as a 4MR SFF and the VPT2 correction will always be applied to the VPT2-screened wavefunction.

Larger errors arise due to applying the reduced mode coupling approximation and using the VPT2 screening procedure, particularly when the mode coupling level is low and the VPT2 screening threshold is high. In this case, it is necessary to weigh up the accuracy loss with the efficiency gain.

Here, we quantify the errors associated with using both reduced mode coupling and VPT2 screening in tandem, and establish the scaling laws that apply to computational time and memory requirements.

Errors in fundamental frequencies are illustrated in Figure 5.3 and summarized in columns 4 and 6 of Table 5.6. For clarity, reference data from analogous reduced

mode coupling (column 2) or VPT2 screening calculations with thresholds of 10^{-5} and 10^{-6} E_h (columns 3,5) are reproduced in this Table.



(a) (b)

Figure 5.3: Errors in VCI_{scr}(6,4MC) + VPT2 with different thresholds at which the VPT2 screening procedure is terminated: a) 1×10^{-5} E_h; b) 1×10^{-6} E_h. Data points are marked according to molecular size: 3 atoms (•), 4 atoms (×), 5 atoms (△), 6 atoms(⬡), 7 atoms (+), 8 atoms(○).

At a VPT2 screening threshold of 10^{-5} E_h with 4-mode coupling, screening errors dominate for smaller molecules (up to 6 atoms), while errors due to screening and reduced mode coupling become cumulative for larger molecules. Overall, the error grows as a function of molecule size, reaching an average of 6 cm^{-1} across the fundamental frequencies of all 8 atom molecules. Inspection of Figure 5.3 reveals maximum errors of up to 30 cm^{-1}, which can be attributed to the VPT2 screening procedure rather than reduced mode coupling pre-screening

Decreasing the screening threshold to 10^{-6} E_h changes the error distribution substantially. At this screening threshold, errors in calculated fundamental frequencies are primarily due to the 4 mode coupling approximation rather than the VPT2 screening, as evident by comparing columns 2 and 6 of Table 5.6 and Figures 5.3(b) and 5.1(b).

In both cases, the 4MC approximation is responsible for the component of the error that increases as a function of molecule size. Therefore, any approach based upon 4MC pre-screening is liable to continue losing accuracy as molecule size increases.

Table 5.6: Mean absolute errors (cm^{-1}) in fundamental frequencies as a function of molecule size, using $VCI_{scr}(6,mMC)$+VPT2, with both full (m=6 mode) and reduced (m=4 mode) coupling in the wavefunction, with screening thresholds of 10^{-5} and 10^{-6} E_h. VCI(6,4MC) data is also provided for reference. The potential energy surface is described by a 4MR SFF in normal mode coordinates throughout.

	No screening	10^{-5} E_h threshold		10^{-6} E_h threshold	
n_{atom}	4MC	Full 6MC	4MC	Full 6MC	4MC
3	0	1.6	1.6	0.1	0.1
4	0.05	2.7	2.7	0.3	0.4
5	0.6	3.2	3.6	0.5	1.0
6	2.5	3.9	5.1	1.4	2.7
7	3.2	3.2	5.9	1.3	3.7
8	4.9	3.9	6.0	1.9	4.7

Therefore, it is also worth investigating the accuracy and computational cost associated with performing $VCI_{scr}(6)$+VPT2 calculations without any pre-truncation of the wavefunction expansion.

As a rough measure of accuracy, not accounting for molecule size-dependence, mean absolute errors associated with each proposed combination of reduced mode coupling level and screening threshold are evaluated across the full data set and reported in Table 5.7.

Associated empirical scaling laws for the total number of VCI matrix elements calculated and the number of configurations selected for inclusion in the VCI sub-matrix during the VPT2 screening procedure, both with and without reduced mode coupling pre-screening, are given in Tables 5.8 and 5.9. The total number of matrix elements calculated determines the run time, while the number of matrix elements selected for inclusion in the VCI sub-matrix determines the memory requirements. The data from which the scaling laws are derived are available as Supporting Information.

Although the errors in fundamental frequencies at $V_{scr}(6,4MC)$ + VPT2 with a

Table 5.7: Mean absolute error (cm^{-1}) in VPT2-screened and VPT2-corrected VCI(6) frequencies, both with and without *a priori* reduction in mode coupling (rows 1 and 2, respectively), and using loose and tight VPT2 screening thresholds (columns 1 and 2, respectively). The potential energy surface is described by a 4MR SFF in normal mode coordinates throughout.

	10^{-5} E_h	10^{-6} E_h
4MC	5.1 cm^{-1}	3.1 cm^{-1}
Full 6MC	3.6 cm^{-1}	1.3 cm^{-1}

Table 5.8: Empirical scaling laws for the total number of matrix elements evaluated during VPT2-screened and VPT2-corrected VCI(6) calculations, both with and without *a priori* reduction in mode coupling (rows 1 and 2, respectively), and using loose and tight VPT2 screening thresholds (columns 1 and 2, respectively). The potential energy surface is described by a 4MR SFF in normal mode coordinates throughout.

	10^{-5} E_h	10^{-6} E_h
4MC	$O(N_{\text{mode}}^{5})$	$O(N_{\text{mode}}^{6})$
Full 6MC	$O(N_{\text{mode}}^{7-8})$	$O(N_{\text{mode}}^{8-9})$

Table 5.9: Empirical scaling laws for the number of elements selected for inclusion in the VCI sub-matrix during VPT2-screened and VPT2-corrected VCI(6) calculations, both with and without *a priori* reduction in mode coupling (rows 1 and 2, respectively), and using loose and tight VPT2 screening thresholds (columns 1 and 2, respectively). The potential energy surface is described by a 4MR SFF in normal mode coordinates throughout.

	10^{-5} E_h	10^{-6} E_h
4MC	$O(N_{\text{mode}}^{1-2})$	$O(N_{\text{mode}}^{2-3})$
Full 6MC	$O(N_{\text{mode}}^{2})$	$O(N_{\text{mode}}^{2-3})$

loose screening threshold (10^{-5} E_h) push the upper limits of what may be considered acceptable, the major advantage of this approach is that its memory requirements scale between linearly and quadratically with molecule size. The exact scaling behaviour is determined by the chemical nature of the molecule, and the extent of coupling between vibrational modes. Further, significant speed-ups in run time are achieved using this screening procedure, with run time scaling as N_{mode}^{5}. By comparison, unscreened VCI(6) calculations formally scale as N_{mode}^{12}. The favourable combination of low memory and runtime scaling requirements mean this approach may be the only realistic option for performing the VPT2-screened VCI calculations on larger molecules.

However, if more accurate results are desired, it is necessary to either increase the extent of mode-coupling or decrease the VPT2 screening threshold, or both. If runtime, but not memory, is the limiting factor, *e.g.* on a shared memory symmetric multiprocessing (SMP) supercomputer with limited walltime, VCI_{scr}(6,4MC) + VPT2 with a tighter screening threshold of 10^{-6} E_h is recommended.

Conversely, if memory, but not runtime, is the limiting factor, *e.g.* on a standard desktop machine, performing VPT2 screening (threshold = 10^{-5} E_h) without reduced mode coupling gives significantly more accurate results only a small increase in memory requirements compared to pre-truncating the wavefunction at 4MC.

Finally, if computational resources are abundant, the safest approach that guarantees high accuracy even for larger molecules is $VCI_{scr}(6)$+VPT2 with the tighter screening threshold of 10^{-6} E_h.

5.5 Conclusions

Vibrational configuration interaction is a simple, robust, easy to implement and easy to parallelize variational nuclear structure method. However, the need to include highly excited states in the wavefunction expansion even to describe fundamental frequencies results in poor scaling properties, and generally precludes its application to realistically sized molecules of chemical interest.

Therefore, we have implemented and rigorously tested a range of strategies to reduce the computational cost associated with VCI calculations for a collection of molecules containing up to 10 atoms.

Truncating the sextic force field expansion at 4 mode representation gives substantial speed-ups with minimal accuracy loss, whether screening procedures are applied to the VCI wavefunction or not. *A priori* reduction of mode coupling in the wavefunction also significantly reduces runtime, but at the cost of systematically underestimating anharmonicities and introducing errors that grow as a function of molecule size. Iterative VPT2 screening of the VCI matrix decreases both runtime and memory requirements. Both the extent of computational efficiency gain and magnitude of error introduced depend on the final screening threshold. Applying a subsequent VPT2 correction always improves the calculated fundamental frequencies at minimal additional computational cost.

Overall, we recommend the $VCI_{scr}(6,4MC)$+VPT2 procedure with a screening threshold of 10^{-5} E_h, as the most efficient method that is capable of yielding reasonably accurate results, subject to the caveat that some highly anharmonic, highly coupled modes may incur substantial errors of up to 30 cm^{-1}, and that errors are likely to increase slightly as a function of molecular size.

For reliably accurate results, independent of molecular size, but still at much reduced computational cost compared to VCI(6), we recommend $VCI_{scr}(6)$+VPT2

with a screening threshold of 10^{-6} E_h.

References

[1] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, New York, **1955**.

[2] J. Bowman, *Acc. Chem. Res.* **1986**, *19*, 202–208.

[3] S. Carter, S. Culik, J. Bowman, *J. Chem. Phys.* **1997**, *107*, 10458–10469.

[4] S. Carter, J. Bowman, N. Handy, *Theor. Chem. Acc.* **1998**, *100*, 191–198.

[5] N. Wright, R. Gerber, *J. Chem. Phys.* **2000**, *112*, 2598–2604.

[6] J. Bowman, S. Carter, X. Huang, *Int. Rev. Phys. Chem.* **2003**, *22*, 533–549.

[7] K. Yagi, K. Hirao, T. Taketsugu, M. Schmidt, M. Gordon, *J. Chem. Phys.* **2004**, *121*, 1383–1389.

[8] D. Benoit, *J. Chem. Phys.* **2004**, *120*, 562–573.

[9] O. Christiansen, *Phys. Chem. Chem. Phys.* **2007**, *9*, 2942–2953.

[10] S. Carter, N. Handy, *Comp. Phys. Rep.* **1986**, *5*, 115–172.

[11] C. Fabri, T. Furtenbacher, A. G. Csaszar, *Mol. Phys.* **2014**, *112*, 2462–2467.

[12] O. Christiansen, *J. Chem. Phys.* **2004**, *120*, 2149–2159.

[13] P. Seidler, M. Sparta, O. Christiansen, *J. Chem. Phys.* **2011**, *134*, 054119.

[14] P. Seidler, O. Christiansen, *J. Chem. Phys.* **2009**, *131*, 234109.

[15] O. Christiansen, *Phys. Chem. Chem. Phys.* **2012**, *14*, 6672–6687.

[16] S. N. Yurchenko, W. Thiel, P. Jensen, *J. Mol. Spectrosc.* **2007**, *245*, 126–140.

[17] Y. Scribano, D. M. Lauvergnat, D. M. Benoit, *J. Chem. Phys.* **2010**, *133*, 094103.

[18] T. K. Roy, R. B. Gerber, *Phys. Chem. Chem. Phys.* **2013**, *15*, 9468–9492.

[19] V. Barone, *J. Chem. Phys.* **2004**, *120*, 3059–3065.

[20] V. Barone, *J. Chem. Phys.* **2005**, *122*, 014108.

[21] J. Bloino, M. Biczysko, V. Barone, *J. Chem. Theor. Comp.* **2012**, *8*, 1015–1036.

[22] W. Mizukami, D. P. Tew, *J. Chem. Phys.* **2013**, *139*, 194108.

[23] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Phys. Chem. A* **2012**, *116*, 3691–3709.

[24] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, *J. Chem. Phys.* **2014**, *141*, 234114.

[25] A. M. Rosnik, W. F. Polik, *Mol. Phys.* **2014**, *112*, 261–300.

[26] V. Barone, M. Biczysko, J. Bloino, *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.

[27] M. Piccardo, J. Bloino, V. Barone, *Int. J. Quant. Chem.* **2015**, *115*, 948–982.

[28] N. Matsunaga, G. Chaban, R. Gerber, *J. Chem. Phys.* **2002**, *117*, 3541–3547.

[29] O. Christiansen, *J. Chem. Phys.* **2003**, *119*, 5773–5781.

[30] S. Krasnoshchekov, E. Isayeva, N. Stepanov, *J. Chem. Phys* **2014**, *141*, 234114.

[31] S. Krasnoshchekov, N. Craig, P. Boopalachandran, J. Laane, N. Stepanov, *J. Phys. Chem. A* **2015**, *119*, 10706–10723.

[32] K. Yagi, S. Hirata, K. Hirao, *Phys. Chem. Chem. Phys.* **2008**, *10*, 1781–1788.

[33] N. C. Handy, S. Carter, *Mol. Phys.* **2004**, *102*, 2201–2205.

[34] G. Rauhut, *J. Chem. Phys.* **2007**, *127*, 184109.

[35] Y. Scribano, D. M. Benoit, *Chem. Phys. Lett.* **2008**, *458*, 384–387.

[36] M. Neff, G. Rauhut, *J. Chem. Phys.* **2009**, *131*, 124129.

[37] P. Carbonniere, A. Dargelos, C. Pouchan, *Theor. Chem. Acc.* **2010**, *125*, 543–554.

[38] C. Konig, O. Christiansen, *J. Chem. Phys.* **2015**, *142*, 144115.

[39] M. Sibaev, D. L. Crittenden, *Comp. Phys. Comm.* **2016**, Submitted.

[40] M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* **2015**, *36*, 2200–2207.

[41] PyPES Library of Potential Energy Surfaces, `http://sourceforge.net/projects/pypes-lib/`.

[42] M. Sibaev, D. L. Crittenden, *J. Chem. Theory Comput.* **2016**, In Preparation.

Chapter 6

Quadratic Corrections to Harmonic Vibrational Frequencies Outperform Linear Models

6.1 Introduction

It is widely reported[1–7] and commonly accepted in the scientific literature that *ab initio* derived harmonic vibrational frequencies tend to be larger than experimentally observed fundamentals, due to the combined effects of anharmonicity and methodological incompleteness. Assuming that these effects are uniform across the spectral range motivates the use of parameterized scaling factors to improve the agreement between predicted and observed fundamental frequencies.[1–7] Within this empirical framework, separate scaling factors are required for different *ab initio* method and basis set combinations.[1–7]

However, there are two lines of evidence in the literature to suggest that straightforward linear scaling of normal coordinate force constants may not always be justified.

Firstly, internal coordinate force constant scaling approaches[8] achieve higher accuracies than simple normal coordinate frequency scaling. However, this comes at the cost of having to define appropriate internal coordinate sets and parameterize different scaling factors for each internal coordinate. Although somewhat lacking in

generality, this approach nonetheless illustrates the importance of different scaling factors for different types of molecular motion.

Even within the literature on normal coordinate force constant scaling, there are strong indications that anharmonicity and methodological incompleteness effects are not, in fact, constant across the entire spectral range, with different scaling factors required in low and high frequency regimes.[5,6] Down-scaling is always recommended for high frequencies, but scaling factors for low frequencies vary significantly in both magnitude and direction across different levels of theory, with a median recommended value around 1.0, corresponding to no scaling correction.

These observations raise a number of questions:

- Can an alternative relationship between anharmonicity and harmonic normal mode frequency be empirically established?
- Could this underpin a more accurate and/or robust anharmonic correction model that retains the simplicity and generality of a frequency scaling approach?
- Are the low frequency scaling factors primarily accounting for methodological incompleteness rather than anharmonicity?

6.2 Methods

To eliminate the confounding effects of methodological incompleteness, we use the PyPES library of high quality semi-global potential energy surfaces (PES).[9] This enables us to obtain benchmark anharmonic vibrational frequencies *and* their harmonic counterparts for 226 unique fundamental vibrational modes. Although reference anharmonic vibrational frequencies are available in the literature for the PES contained within the PyPES library, benchmark harmonic frequencies calculated from these surfaces have not all been available until now. The 50 molecules within the PyPES library vary in size from three to six atoms, and contain a range of different atom types, bonding patterns and molecular topologies. Hence, the vibrational modes of these molecules are expected to form a representative set. All benchmark

anharmonic and harmonic frequencies used in this work are provided as Supporting Information.

For larger molecules, where high level *ab initio* calculations to obtain accurate potential energy surfaces are not feasible, previous work[5,6,10] suggests that density functionals incorporating around 20% Hartree-Fock exchange reliably recover harmonic frequencies comparable to those obtained at much higher levels of theory. We therefore benchmark the ability of the B3LYP[11], B3PW91[12], PBE0[13], EDF2[10], M05[14] and M06[15] functionals to recover benchmark harmonic frequencies for all molecules in the PyPES library. We also assess the ability of the quadratic correction model defined above to predict anharmonic frequencies from DFT harmonic frequencies.

All calculations are carried out in the atomic orbital basis sets that were used for parameterizing each functional[16–20], augmenting each basis set with diffuse functions for anionic molecules if not already included.[21,22] The CRENBL ECP basis[23,24] with the accompanying relativistic effective core potential (ECP) is used for atoms larger than Kr. All (TD-)DFT geometry optimizations and subsequent frequency calculations are carried out using the *Q-Chem 4.2* program package[25], employing a Euler-Maclaurin-Lebedev product quadrature grid comprising 75 radial points and 302 angular points per radial point, with an SCF convergence threshold of 10^{-8} and geometry optimization thresholds decreased by an order of magnitude from their default values.

6.3 Results and Discussion

6.3.1 Anharmonicity Model

For clarity and consistency, we recast the scale factor approach of Radom *et al.*[5,6] (6.1) as a linear correction model (6.2).

$$\nu \approx \lambda \nu_{\mathrm{e}} \tag{6.1}$$

$$\nu \approx (1 - c_1)\nu_{\mathrm{e}} \tag{6.2}$$

in which ν represents the benchmark anharmonic frequency we wish to approximate and ν_e its harmonic equivalent. This enables us to recast the problem of minimizing the difference between scaled and benchmark frequencies as a problem of approximating anharmonicities as a function of harmonic frequencies:

$$\nu_e - \nu \approx c_1 \nu_e \tag{6.3}$$

Or, equivalently,

$$\Delta_{anh} \approx c_1 \nu_e \tag{6.4}$$

The optimal coefficient, c_1, is determined by least-squares fitting to experimentally derived or benchmark anharmonicities, *i.e.* by linear regression with Δ_{anh} as the response variable and ν_e as the independent variable, as illustrated in Figure 6.1(a). As per equations (6.1) and (6.2), the coefficient, c_1, derived in this manner is the complement of the scaling factor, λ, defined by Radom *et al.* The quality of the model is more evident upon examining the residual differences between predicted and benchmark frequencies:

$$\Delta_{res} = \nu_{predicted} - \nu \tag{6.5}$$

shown in Figure 6.1(b), noting that if no correction is applied to the harmonic frequencies then $\nu_{predicted} = \nu_e$ and $\Delta_{res} = \Delta_{anh}$.

From Figures 6.1(a) and 6.1(b), it is clear that a single parameter linear correction model significantly overestimates anharmonicity corrections in the low frequency, low anharmonicity regime while simultaneously underestimating anharmonicity corrections at higher frequencies. Indeed, low frequencies are often overcorrected to such an extent that the 'improved' frequency estimates are, in fact, further from the experimental values than the original harmonic estimates. Points below the red anti-diagonal line on Figure 6.1(b) fall into this category. Ideally, the trend line on Figure 6.1(a) would provide a closer fit to the benchmark anharmonicity data, resulting in residual errors in Figure 6.1(b) narrowly and randomly clustered around the dashed horizontal line.

The tendency of the single parameter linear scaling model to overcorrect low frequency modes is also reflected in the summary statistics presented in Table 6.1.

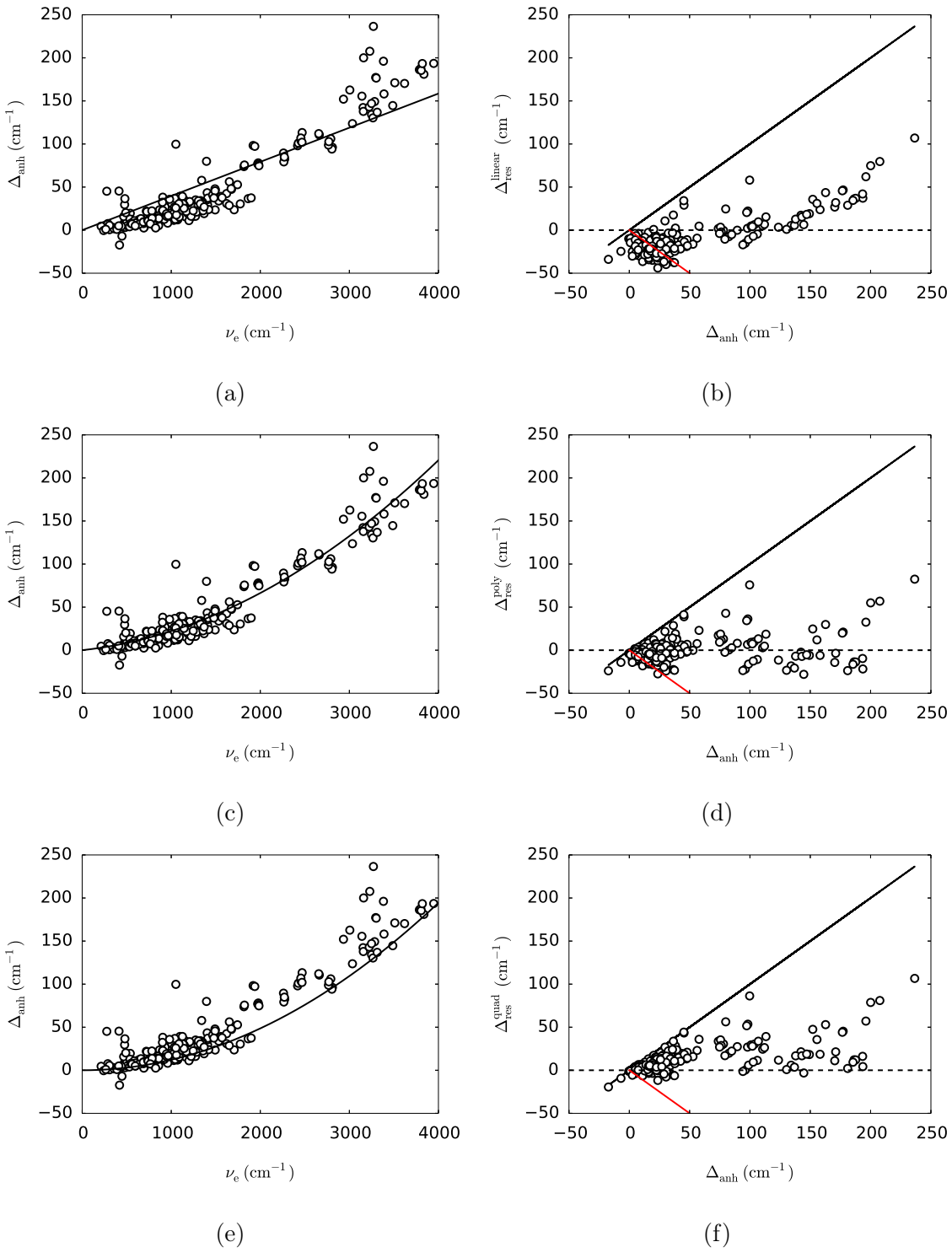


Figure 6.1: Left: Benchmark anharmonicities as a function of frequency, with trend-lines representing; a) single-parameter linear model, c) dual-parameter polynomial model, and e) single-parameter quadratic model. Right: Corresponding residual errors as a function of anharmonicity, for b) linear, d) polynomial and f) quadratic correction models.

Table 6.1: Mean unsigned, mean signed and maximum signed errors in predicted anharmonic frequencies (Δ_{res}), according to models defined by equation 6.6 and the parameters given. All values reported in units of cm^{-1}.

approx	c_1	c_2	$\langle\lvert\Delta_{\text{res}}\rvert\rangle$			$\lvert\Delta_{\text{res}}\rvert_{\text{max}}$	
			all	$\nu_{\text{model}} > \nu$	$\nu_{\text{model}} < \nu$	$\nu_{\text{model}} > \nu$	$\nu_{\text{model}} < \nu$
harmonic	0.0	0.0	47.0	47.6	8.2	236.5	17.3
linear	0.039640	0.0	19.6	24.7	18.2	106.9	44.0
polynomial	0.011214	0.000010982	9.9	12.9	8.1	82.5	28.0
quadratic	0.0	0.00001215	13.3	15.0	3.7	106.6	19.4

Although the mean absolute error decreases from 47.0 to 19.6 cm^{-1}, the average error associated with underpredicted frequencies increases from 8.2 to 18.2 cm^{-1}, with corresponding maximum error increasing from 17.3 to 44.0 cm^{-1}. The c_1 value of 0.039640 corresponds to a λ value of 0.96036, in very good agreement with existing scale factors parameterized for high level correlated *ab initio* methods across a larger data set; 0.9639 for CCSD(T)/6-311+G(d,p).[6]

It is now evident that the dual scaling factor recommendation of Radom *et al.*[5,6] implies that at least a bi-linear model is required to describe trends in anharmonicity as a function of harmonic frequency. However, they do not give an exact prescription for mapping scale factor to frequency range. To complete the specification of their model, further optimization to determine the optimal 'cross-over' point would be required.

Given the relatively straightforward relationship between anharmonicity and harmonic frequency apparent upon visual inspection of Figure 6.1(a), this approach seems needlessly complicated. Instead, we propose a second-order polynomial model:

$$\Delta_{\text{anh}} \approx c_1\nu_{\text{e}} + c_2\nu_{\text{e}}^2 \tag{6.6}$$

This produces a much closer fit to the anharmonicity data, as illustrated in Figure 6.1(c) and summarized in Table 6.1. Although the polynomial model produces universally more accurate estimates of the benchmark frequencies than the single-parameter linear model, there remains a cluster of outliers in the low frequency,

high anharmonicity region, a single outlier at 1052 cm^{-1} and Δ_{anh} = 99 cm^{-1}, and another outlying pair of modes with $\Delta_{anh} << 0$.

Modes with anomalously high anharmonicities all represent cases in which the assumption of low amplitude vibrations about a single minimum on a PES expanded in normal coordinates breaks down; for low barrier torsional modes (the low frequency, high anharmonicity cluster) and the NH_3 inversion mode (the lone outlier at 1052 cm^{-1}). In these cases, correcting for anharmonicity by scaling normal coordinate force constants is inappropriate, as internal-coordinate based approaches for expanding the PES and solving the nuclear vibrational Schrödinger equation are required.

The two cases in which anharmonicity *increases* the fundamental frequencies correspond to antisymmetric stretching modes of excited state ClO_2 and BrO_2. Early studies attributed this behaviour to C_s-distortion of the equilibrium geometry producing a very shallow double minimum in the potential.[26] However, more extensive recent work has concluded that the negative (according to the sign convention adopted here) anharmonicity corrections arise from strong anharmonic coupling between symmetric and asymmetric stretching modes.[27]. Again, a normal coordinate force constant scaling approach is ill-suited to capturing these effects.

This is reflected in the residual error data illustrated in Figure 6.1(d). The polynomial scaling model fails to allow harmonic frequencies to increase toward their anharmonic counterparts, resulting in residual errors as large as, or even worse than, the original harmonic estimates, *i.e.* $|\Delta_{res}| > |\Delta_{anh}|$ when $\Delta_{anh} < 0$. In these cases, the 'least worst' prediction would be no change from the harmonic approximation.

Otherwise, excluding torsional and inversion mode outliers, residual errors tend to be randomly and narrowly scattered about $\Delta_{res} = 0$. Although the polynomial model significantly outperforms the linear model on this metric, it has the disadvantage of requiring an additional empirical parameter. Further, there remain a number of points below the red anti-diagonal line on Figure 6.1(d), indicating that although the magnitude and extent of 'worse-than-harmonic' overcorrection errors have decreased, they have not been completely eliminated.

Therefore, we seek a model that; minimizes overcorrection errors rather than

minimizing the overall error, requires only a single parameter, and outperforms the single-parameter linear model in every metric. This combination of constraints yields the quadratic model illustrated in Figures 6.1(e) and 6.1(f), with $c_2 = 0.00001215$. A major advantage of this model is that it provides a lower bound estimate of the anharmonicity in most cases. In other words, it generally corrects harmonic frequencies down toward but not beyond their experimental values. This is particularly important when error *direction* is as important, if not more important, than error *magnitude*.

For example, corrected frequencies that are higher than their true values will yield lower bounds for derived thermochemical parameters such as enthalpies and entropies. Further, this leads to lower total errors in calculated thermochemical parameters, as anharmonic frequencies that are too high result in smaller errors than frequencies that are too low by the same amount, due to the inverse exponential ansatz.

There remain a handful of cases in which the quadratic model overpredicts the anharmonicity correction, but in each of these cases, the error is small. Overcorrection errors are less than 12 cm^{-1} in all cases, averaging 3.0 cm^{-1}. Like the polynomial model, the quadratic model fails to account for the rare cases in which the true frequencies are higher than the harmonic frequencies and a negative anharmonicity correction is required. In these cases, the quadratic model does not significantly compound this error, but instead returns frequencies similar within 3 cm^{-1} of the original harmonic frequencies.

6.3.2 DFT Frequencies

(TD-)DFT frequencies are calculated using a range of commonly-used gradient-corrected functionals including B3LYP[11], B3PW91[12], PBE0[13], M05[14] and M06[15]. The EDF2 functional, which was explicitly parametrized to reproduce anharmonic vibrational frequencies,[10] is also used. Errors in (TD-)DFT harmonic frequencies are calculated with reference to benchmark values:

$$\Delta_{\text{harm}} = \nu_{\text{e}}^{\text{DFT}} - \nu_{\text{e}} \tag{6.7}$$

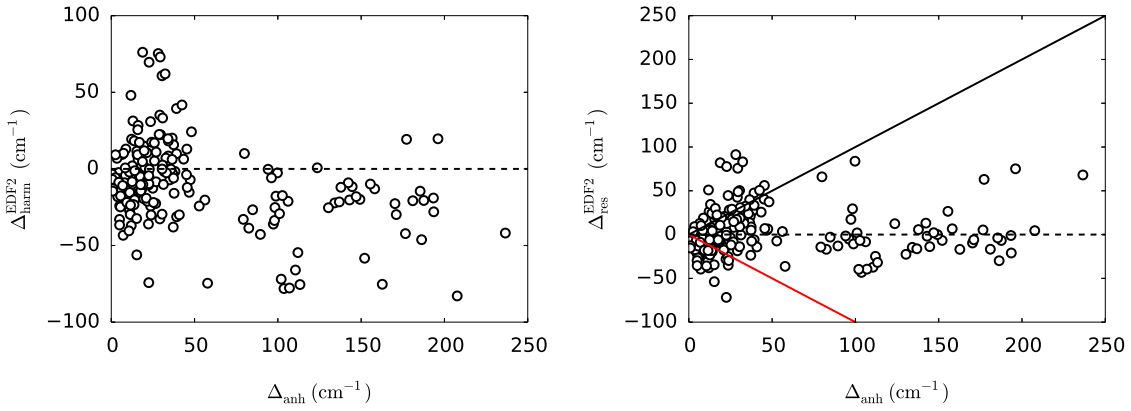
Table 6.2: Mean and maximum errors in DFT harmonic frequencies (Δ_{harm}), excluding excited states and molecules containing atoms larger than Kr. All values reported in units of cm^{-1}. [a] 6-31+G(d,p) or [b] 6-311+G(d,p) or [c] aug-cc-pVTZ basis used for anions.

Method	$\langle\lvert\Delta_{\mathrm{harm}}\rvert\rangle$			$\lvert\Delta_{\mathrm{harm}}\rvert_{\mathrm{max}}$	
	all	$\nu_{\mathrm{e}}^{\mathrm{DFT}} > \nu_{\mathrm{e}}$	$\nu_{\mathrm{e}}^{\mathrm{DFT}} < \nu_{\mathrm{e}}$	$\nu_{\mathrm{e}}^{\mathrm{DFT}} > \nu_{\mathrm{e}}$	$\nu_{\mathrm{e}}^{\mathrm{DFT}} < \nu_{\mathrm{e}}$
B3LYP/6-31G(d,p)[a]	23.4	18.4	26.3	72.0	138.3
B3PW91/6-31G(d,p)[a]	20.0	19.9	20.2	102.5	94.1
PBE0/6-311G(d,p)[b]	22.2	25.2	19.2	132.9	90.4
M05/6-311+G(2df,2p)	30.4	34.5	27.6	162.3	125.7
M06/6-311+G(2df,2p)	27.5	32.5	23.7	146.4	121.2
EDF2/cc-pVTZ[c]	21.1	21.3	21.1	76.1	82.9

Mean and maximum absolute and signed errors in DFT harmonic frequencies are reported in Table 6.2. Excited states and molecules containing atoms larger than Kr are excluded from statistical analysis, because using TD-DFT or effective core potentials introduces additional approximations beyond those inherent in the parameterization of each functional, which may further decrease the accuracy of the calculated $\nu_{\mathrm{e}}^{\mathrm{DFT}}$. For completeness, the full set of results is provided as Supporting Information.

The data presented in Table 6.2 are broadly consistent with previous studies that report mean or RMSD errors in harmonic frequencies of 30 – 40 cm^{-1} using B3LYP[28] and PBE0[28,29] with triple zeta basis sets. The minor discrepancy between the literature results and those reported here arises from our use of a larger and more representative test set of molecules and our choice to use the basis sets in which each functional was parameterized.

Of the functionals investigated here, EDF2 is generally the most accurate. This is to be expected, as it was explicitly parameterized to recover CCSD(T)/cc-pVTZ harmonic frequencies.[10] Nonetheless, significant errors in harmonic frequencies are observed, with a mean absolute deviation of 21.1 cm^{-1} and maximum absolute error of 82.9 cm^{-1}. Statistically, errors are randomly distributed across the data set.



(a) (b)

Figure 6.2: (a) Errors in EDF2/cc-pVTZ harmonic frequencies as a function of benchmark anharmonicity, and (b) corresponding errors in predicted anharmonic frequencies using the quadratic correction model.

However, upon visual inspection of Figure 6.2(a), it is clear that EDF2 systematically underestimates the frequencies of highly anharmonic modes.

Errors in quadratically corrected DFT-derived anharmonic frequencies are presented in Table 6.3. Comparing Tables 6.2 and 6.3 reveals a strong correlation between mean absolute and maximum errors in DFT harmonic frequencies, and corresponding errors in DFT-derived anharmonic frequencies. This implies that residual errors in predicted anharmonic frequencies derive primarily from the inaccuracy of the DFT harmonic frequencies rather than inadequacy of the anharmonicity correction model. This observation is supported by existing literature results, in which anharmonic corrections are calculated using vibrational perturbation theory. Even using this significantly more time consuming and rigorous procedure to account for anharmonicity, errors in calculated anharmonic frequencies are strongly correlated with errors in the underlying harmonic frequencies.[29]

For low frequency modes, the quadratic model predicts only small anharmonicity corrections by construction, and therefore errors in DFT harmonic frequencies translate almost directly into residual errors in predicted anharmonic frequencies, as anticipated above. This behaviour is evident comparing the low anharmonicity regions of Figures 6.2(a) and 6.2(b).

Table 6.3: Mean and maximum errors in anharmonic frequencies ($\Delta_{\text{res}}^{\text{quad}}$) predicted from DFT harmonic frequencies using the quadratic correction model (equation 6.6, $c_1 = 0$, $c_2 = 0.00001215$), excluding excited states and molecules containing atoms larger than Kr. All values reported in units of cm^{-1}. [a] 6-31+G(d,p) or [b] 6-311+G(d,p) or [c] aug-cc-pVTZ basis used for anions.

Method	$\langle\lvert\Delta_{\text{res}}^{\text{quad}}\rvert\rangle$			$\lvert\Delta_{\text{res}}^{\text{quad}}\rvert_{\text{max}}$	
	all	$\nu_{\text{model}} > \nu$	$\nu_{\text{model}} < \nu$	$\nu_{\text{model}} > \nu$	$\nu_{\text{model}} < \nu$
B3LYP/6-31G(d,p)[a]	23.1	23.1	23.0	123.18	133.9
B3PW91/6-31G(d,p)[a]	23.9	26.5	18.5	115.5	91.0
PBE0/6-311G(d,p)[b]	26.5	31.6	16.4	133.4	87.4
M05/6-311+G(2df,2p)	32.5	39.2	18.4	161.7	102.0
M06/6-311+G(2df,2p)	29.3	34.7	16.4	152.4	100.3
EDF2/cc-pVTZ[c]	19.8	23.8	16.0	91.3	71.7

For high frequency modes, the predicted anharmonic frequencies are scattered randomly about $\Delta_{\text{res}} = 0$, as shown in Figure 6.2(b). This is a consequence of error cancellation, with the quadratic correction model systematically overestimating anharmonic frequencies as it was designed to do, and the EDF2 functional systematically underestimating harmonic frequencies. Although it would be possible to reparameterize the quadratic correction model to reinstate the upper bound behaviour for high frequencies, or further optimize it to achieve maximum error cancellation, we consider it preferable to control for anharmonicity and methodological errors separately so we do not pursue this approach.

6.4 Conclusions

Overall, we recommend using the quadratic correction model in conjunction with high level *ab initio* harmonic frequencies, due to its simplicity, accuracy and ability to provide semi-bounded lower estimates of anharmonicities. This approach recovers anharmonic frequencies within ~ 13 cm^{-1} of benchmark values, on average, across a diverse range of chemical species. We note that low barrier torsional and inversion

modes should be excluded *a priori* due to the inappropriateness of normal modes for describing these types of motion.

Where high level *ab initio* harmonic frequency calculations are not feasible, quadratically corrected DFT frequencies reasonably approximate anharmonic stretching frequencies, with mean absolute errors in the 20 – 30 cm^{-1} range. However, DFT-derived estimates of anharmonic frequencies are less reliable for lower frequency torsional and bending modes, due to these regions of the potential energy surface being poorly described by DFT methods. In these cases, errors in anharmonic fundamental frequencies predicted by both simple empirical correction models and more rigorous nuclear vibrational structure theories (VPT2) are both dominated by relatively large errors in the DFT harmonic frequencies.

References

[1] M. P. Andersson, P. Uvdal, *J. Phys. Chem. A* **2005**, *109*, 2937–2941.

[2] S. G. Andrade, L. C. S. Goncalves, F. E. Jorge, *J. Mol. Struct. (THEOCHEM)* **2008**, *864*, 20–25.

[3] M. L. Laury, S. E. Boesch, I. Haken, P. Sinha, R. A. Wheeler, A. K. Wilson, *J. Comput. Chem.* **2011**, *32*, 2339–2347.

[4] J. A. Pople, A. P. Scott, M. W. Wong, L. Radom, *Isr. J. Chem.* **1993**, *33*, 345–350.

[5] A. P. Scott, L. Radom, *J. Phys. Chem.* **1996**, *100*, 16502–16513.

[6] J. P. Merrick, D. Moran, L. Radom, *J. Phys. Chem. A* **2007**, *111*, 11683–11700.

[7] P. Sinha, S. E. Boesch, C. M. Gu, R. A. Wheeler, A. K. Wilson, *J. Phys. Chem. A* **2004**, *108*, 9213–9217.

[8] J. Baker, A. A. Jarzecki, P. Pulay, *J. Phys. Chem. A* **1998**, *102*, 1412–1424.

[9] M. Sibaev, D. L. Crittenden, *J. Comput. Chem.* **2015**, *36*, 2200–2207.

[10] C. Y. Lin, M. W. George, P. M. W. Gill, *Aust. J. Chem.* **2004**, *57*, 365–370.

[11] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[12] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[13] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[14] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Phys.* **2005**, *123*, 161103.

[15] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[16] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. A. Rassolov, J. A. Pople, *J. Chem. Phys.* **1998**, *109*, 7764–7776.

[17] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265–3269.

[18] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650–654.

[19] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[20] D. E. Woon, T. H. Dunning, *J. Chem. Phys.* **1993**, *98*, 1358–1371.

[21] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, *J. Comp. Chem.* **1983**, *4*, 294–301.

[22] R. A. Kendall, T. H. Dunning, R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[23] L. A. LaJohn, P. A. Christiansen, R. B. Ross, T. Atashroo, W. C. Ermler, *J. Chem. Phys.* **1987**, *87*, 2812–2824.

[24] R. B. Ross, J. M. Powers, T. Atashroo, W. C. Ermler, L. A. LaJohn, P. A. Christiansen, *J. Chem. Phys.* **1990**, *93*, 6654–6670.

[25] Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kus, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stueck, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A.

Goddard, III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, M. Head-Gordon, *Mol. Phys.* **2015**, *113*, 184–215.

[26] J. B. Coon, F. A. Cesani, C. M. Loyd, *Discuss. Faraday Soc.* **1963**, *35*, 118–123.

[27] K. A. Peterson, *J. Chem. Phys.* **1998**, *109*, 8864.

[28] M. Biczysko, P. Panek, G. Scalmani, J. Bloino, V. Barone, *J. Chem. Theor. Comp.* **2010**, *6*, 2115–2125.

[29] D. A. Boese, W. Klopper, J. M. L. Martin, *Mol. Phys.* **2005**, *103*, 863–876.

Chapter 7

Future Work

In Chapter 4, it was shown that use of curvilinear normal mode coordinates reduces inter-mode coupling, leading to faster convergence of the PEF. Because this is the first time such an approach has been implemented and tested, we used a very simple algorithm to generate a reasonable set of non-redundant internal coordinates. Namely, the delocalised internal coordinates of Baker *et. al.*[1] However, it is generally agreed that optimal decoupling of coordinates can be achieved when the natural internal coordinates of Pulay *et al.* are used.[2–4] They exploit local pseudo-symmetry for groups of atoms within the system to generate non-redundant linear combinations of valence internal coordinates. Their main drawback is the complexity of implementation, and lack of generality, since a non-redundant set cannot always be defined.

Both issues have been addressed by von Arnim and Ahlrichs in their definition of generalised natural internal coordinates.[5] They propose a simpler automatic algorithm for defining natural internal coordinates, and in cases where redundancy issues arise, delocalised internals are used to describe problematic regions, while the usual natural internals are used for the rest of the system. To further reduce coordinate coupling when delocalised internals are necessary, they propose to separate valence internals of different types. That is, instead of having delocalised coordinates as linear combinations of the full set of bond lengths, bond angles and torsion angles, each delocalised internal is generated as a linear combination of only bond lengths, or bond angles, or torsion angles. Use of generalised natural internal coordinates

has the benefit of being completely general, like delocalised internal coordinates, but also exhibiting the separability of natural internal coordinates. In the same publication,[5] generalised natural internal coordinates were shown to outperform delocalised internals when used in geometry optimisation. Similarly, they are likely to lead to even greater decoupling and faster convergence when used for construction of the PEF.

As was already noted in Chapter 4, a three mode representation quartic force field, $V(3,4)$, coupled with standard VPT2 is known to provide quite accurate results for small molecules, when resonances are accounted for. However, with larger molecules, resonances become more numerous and occurrence of complicated connected resonances is also more likely. The standard treatment of resonances, to the best of my knowledge, has not been benchmarked against a more rigorous VCI or VCI+VPT2 method on large molecules (>10 atoms), where their error is likely to be the greatest. This information would be quite valuable when deciding what level of *ab initio* theory to use when constructing the potential, to ensure that errors in calculated fundamental frequencies introduced by the level of *ab initio* theory used in construction of the PEF are comparable to errors arising from the VPT2 procedure. On that note, it would also be interesting to quantify the accuracy of $\tilde{V}(2,4) \rightarrow V(3,4)$ potential combined with VPT2 for modelling fundamentals.

There are some algorithmic improvements that need to be tested and included in the code. The program for potential construction in curvilinear normal mode coordinates would benefit from further:

- setting screening thresholds for force constants in curvilinear normal mode coordinates;
- using sparse array structures for storing the derivatives of curvilinear normal mode coordinates with respect to rectilinear normal mode coordinates ($\tilde{L}$ matrix, Chapter 4).

The VCI+VPT2 implementation could be improved by:

- using more efficient sparse matrix diagonalisation routines;

- investigating whether preconditioning of the VCI matrix for diagonalisation can be improved by using information from the previous iteration.

Another logical step is to further benchmark both potential construction and VCI+VPT2 approaches with higher quality PEFs, enabling comparison to experiment. Because highly accurate methods for solving the electronic problem, like CCSD(T) with a large basis set, scale poorly with respect to molecular size, it is not practical to construct full anharmonic force fields with them for molecules with more than 6 atoms. Usually a hybrid approach is used, where highly accurate methods are applied to obtain harmonic frequencies and lower levels of theory are used to construct the anharmonic part of potential using the more accurate normal mode coordinates.[6,7] It would be interesting and novel to look into adopting composite methods, that are usually used only for calculating energies, to obtain the derivative data necessary for constructing the PEF.

Composite methods use multiple computations to estimate higher accuracy results at lower computational cost, overall. The well known Gaussian-n (G-n) series of methods by Pople and co-workers,[8–10] and more recently correlation consistent composite approach (ccCA) by Wilson and co-workers,[11] are two examples of highly accurate composite methods. However, they are not designed for calculating derivative data and their straightforward application is complicated by the fact that the equilibrium geometry varies with the level of theory, which will introduce artificial errors into the combined Hessian from the non-zero gradient. Arguably it would be more appropriate to use different equilibrium geometry for each level of theory, to make sure that the gradient terms are zero and the Hessian is positive definite. Of course, geometry optimisations are computationally demanding, and it is an open question whether such an approach would be computationally viable.

Finally, once the vibrational wavefunction has been calculated it can be used to evaluate other properties. For example, intensities of vibrational transitions, vibrational effects on the molecular electrostatic potential and on NMR chemical shifts. Implementing procedures to calculate these effects would provide useful tools for the rest of scientific community.

References

[1] J. Baker, A. Kessi, B. Delley, *J. Chem. Phys.* **1996**, *105*, 192–212.

[2] P. Pulay, G. Fogarasi, F. Pang, J. E. Boggs, *J. Am. Chem. Soc.* **1979**, *101*, 2550–2560.

[3] G. Fogarasi, X. Zhou, P. W. Taylor, P. Pulay, *J. Am. Chem. Soc.* **1992**, *114*, 8191–8201.

[4] F. Eckert, P. Pulay, H.-J. Werner, *J. Comput. Chem.* **1997**, *18*, 1473–1483.

[5] M. von Arnim, R. Ahlrichs, *J. Chem. Phys.* **1999**, *111*, 9183–9190.

[6] D. Begue, P. Carbonniere, C. Pouchan, *J. Phys. Chem. A* **2005**, *109*, 4611–4616.

[7] K. Yagi, S. Hirata, K. Hirao, *Theor. Chem. Acc.* **2007**, *118*, 681–691.

[8] J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, L. A. Curtiss, *J. Chem. Phys.* **1989**, *90*, 5622–5629.

[9] L. A. Curtiss, K. Raghavachari, G. W. Trucks, J. A. Pople, *J. Chem. Phys.* **1991**, *94*, 7221–7230.

[10] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, J. A. Pople, *J. Chem. Phys.* **1998**, *109*, 7764–7776.

[11] N. J. DeYonker, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *124*, 114104.

Chapter 8

Conclusions

Accurate modelling of vibrational spectra requires methods that go beyond the harmonic approximation, and take into account anharmonicity in the potential energy surface (PES). We have developed two methods that together enable rigorous and efficient calculation of anharmonic vibrational frequencies.

The first method reduces the computational cost of constructing the PES by using an appropriate set of curvilinear normal mode coordinates during its evaluation, and subsequently transforming to rectilinear normal mode coordinates so that the vibrational problem can be solved efficiently.

The second method reduces the computational cost of solving the vibrational problem, using the computationally efficient second order vibrational perturbation theory (VPT2) to select an appropriate basis for the vibrational configuration interaction (VCI) procedure and concurrently account for energetic contributions from the excluded configurations, all in an iterative manner.

Together, these methods allow for treatment of large molecules (up to 20 atoms), with fine control over the balance between accuracy and computational expense.

For even larger molecules where rigorously accounting for anharmonicity becomes unfeasible, we introduced an empirical quadratic correction model for approximating fundamental frequencies. Using our carefully compiled collection of high-quality analytic potential energy surfaces (PyPES library) to eliminate methodological incompleteness errors, we showed that the effects of anharmonicity can generally not

be described as a linear function of the harmonic frequencies, as is usually assumed, but a quadratic one.

Finally, we also addressed the issue of validating our new vibrational methods during the development stage, using analytic potentials from the PyPES library for benchmarking and testing our coordinate transformation procedure, and implementing a rigorous VCI algorithm, called PyVCI, to obtain reference data when solving the vibrational problem.