

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

Marat Sibaev, Deborah L Crittenden

Department of Chemistry, University of Canterbury, Christchurch, New Zealand

Abstract

The PyVCI program package is a general purpose open-source code for simulating accurate molecular spectra, based upon force field expansions of the potential energy surface in normal mode coordinates. It includes harmonic normal coordinate analysis and vibrational configuration interaction (VCI) algorithms, implemented primarily in Python for accessibility but with time-consuming routines written in C. Coriolis coupling terms may be optionally included in the vibrational Hamiltonian. Non-negligible VCI matrix elements are stored in sparse matrix format to alleviate the diagonalization problem. CPU and memory requirements may be further controlled by algorithmic choices and/or numerical screening procedures, and recommended values are established by benchmarking using a test set of 44 molecules for which accurate analytical potential energy surfaces are available.

Force fields in normal mode coordinates are obtained from the PyPES library of high quality analytical potential energy surfaces (to 6th order) or by numerical differentiation of analytic second derivatives generated using the GAMESS quantum chemical program package (to 4th order).

Keywords: vibrational configuration interaction, Coriolis coupling, infrared spectroscopy, normal mode expansion, sextic force field, excitation order, sparse matrix

*Corresponding author

Email address: `deborah.crittenden@canterbury.ac.nz` (Deborah L Crittenden)

1. Program Summary

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

Authors: M. Sibaev and D. L. Crittenden

5 *Program Title:* PyVCI

Journal Reference:

Catalogue identifier:

Licensing provisions: Public Domain release

Programming language: Python, C

10 *Computer:* PC

Operating system: Linux, MacOSX, Windows

RAM: varies widely

Number of processors used: 1-16

Supplementary material: User Manual and examples (tutes) included

15 *Keywords:* Vibrational configuration interaction, nuclear vibrational problem, harmonic oscillator basis, sparse matrix, parallel

Classification: 16.3 Molecular Vibrations

External routines/libraries: Numpy, Scipy, Cython

20 *Nature of problem:* The simulation of accurate molecular vibrational spectra is a significant and long-standing problem in computational chemistry. There are two major challenges: constructing an accurate *ab initio* potential energy surface and solving the nuclear vibrational problem. Both scale poorly with respect to molecular size, requiring large amounts of CPU time and memory.

25 *Solution method:* We have implemented a straightforward numerical differentiation algorithm to construct quartic force fields in normal mode coordinates using second derivatives of the energy with respect to nuclear displacement obtained from *ab initio* quantum chemical calculations, for nuclear vibrational structure algorithm development and testing purposes. We have also provided an interface to the PyPES library
30 of high quality semi-global potential energy surfaces, which enable quantitative prediction of molecular vibrational spectra. To solve the nuclear vibrational problem, we use a vibrational configuration interaction algorithm in a harmonic oscillator basis.

35 *Unusual features:* One of the unusual features of our code is its flexibility, with multiple ways of generating or supplying force field data, dynamic memory allocation, adjustable screening thresholds, and explicit user control over terms in the VCI wavefunction (maximum excitation level and extent of mode-coupling). We employ sparse matrix linear algebra libraries to reduce the memory required for VCI matrix storage
40 and diagonalization, and provide for parallel VCI matrix construction to reduce required wall times.

Running time: varies widely

45 2. 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).

50 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, 2, 3, 4, 5, 6, 7, 8, 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
55 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
60 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.

65 Advances in hardware and software capability [10, 11, 12, 13, 14, 15, 16, 17, 18, 4, 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. Recently, we have developed a coordinate transformation procedure that enables sextic force fields in rectilinear normal
70 mode coordinates to be generated in a numerically stable and computationally efficient manner, taking advantage of the rapid convergence of force field expansions in curvilinear internal coordinates.[20] This makes simulating fundamental modes of infrared spectra possible for a larger range of chemically interesting molecules.

75 A recent review by Roy and Gerber [21] 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
80 energy integrals that can be evaluated analytically.

A hierarchy of approximations yield a series of methods including normal mode analysis (NMA) [22], vibrational self-consistent field theory (VSCF) [23, 24, 25, 26, 27, 28, 18, 29, 30], vibrational perturbation theory (VPT) [31, 32, 33, 34, 30, 35, 36, 37, 38, 39, 40, 41, 42], vibrational configuration interaction (VCI)
85 [43, 30, 44] and vibrational coupled cluster theory (VCC) [45, 30, 46, 47, 48].

Analogous to their electronic structure theory counterparts, the quality of each method is generally commensurate with its computational cost.[30, 16, 48]

90 Despite the extensive efforts that have gone into developing anharmonic nuclear vibrational structure theories,[31, 32, 33, 34, 49, 30, 35, 36, 37, 38, 39, 40, 41, 42, 43, 30, 44, 45, 30, 46, 47, 48] they remain under-utilized within the wider computational chemistry community. This can be traced back to a number of factors including: the ubiquity, ease of use and relatively low computational cost of harmonic normal mode analysis and, conversely, the complexity and computational cost associated with obtaining appropriate PES representations
95 and interfacing with specialized nuclear structure codes [50, 28, 51, 52, 53]; 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.

100 In this paper, we primarily address the first of these issues. Although some nuclear vibrational structure methods are included in some quantum chemical software packages, *e.g.* VPT2 is implemented in GAUSSIAN[38], VSCF and VPT2 in GAMESS[54, 31] and VCI in Q-Chem[16], they vary in both how the PES is represented and how the nuclear vibrational problem is solved. This
105 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.

As a starting point, we present the PyVCI package, a general and open-source vibrational configuration interaction code in which the potential energy surface
110 is represented as a Taylor series expansion up to 6th order in normal mode coordinates. PyVCI is unique in providing a number of user-friendly ways of generating or specifying potential energy surfaces. PyVCI can import force field data from the PyPES library of analytical potential energy surfaces,[1, 2, 55, 20] or directly generate quartic force fields by numerical differentiation of
115 second derivative data obtained from *ab initio* calculations. Currently, only the GAMESS quantum chemistry program package is supported, as it is the only freely available quantum chemical program package that outputs high precision derivative data by default.

Although the VCI method we have implemented within PyVCI is general, it
120 is not completely unique; other variants are possible through different algorithmic choices and alternative approaches to truncating the full ro-vibrational Hamiltonian.[56, 28, 57]

For example, the wavefunction may be expanded in a basis of normal mode harmonic oscillator functions[56, 57] or VSCF modals[56, 28], the complexity
125 of the wavefunction expansion may be limited by either limiting the dimensionality of the system[56], or restricting the extent of mode-coupling within the wavefunction[28] or the maximum extent of excitation within the VCI basis states[57, 28]. Similarly, the Hamiltonian may include[56, 28] or omit[57] Coriolis

coupling and/or Watson terms.

130 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. 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:

$$\hat{H} = \hat{H}_{\text{vib}} + \hat{H}_{\text{Cor}} \quad (1)$$

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

$$\hat{H}_{\text{Cor}} = -\sum_{\alpha} B_{\alpha} \sum_{i < j} \sum_{k < l} \zeta_{ij}^{\alpha} \zeta_{kl}^{\alpha} \left(Q_i \frac{\partial}{\partial Q_j} - Q_j \frac{\partial}{\partial Q_i} \right) \left(Q_k \frac{\partial}{\partial Q_l} - Q_l \frac{\partial}{\partial Q_k} \right) \quad (3)$$

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

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

135 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.[22] 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 140 Appendix A.

This approach yields the diagonal second order force constants, but the higher order derivatives must generally be calculated *via* numerical differentiation. For this reason, the PES expansion is usually truncated at fourth order, to keep the

145 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[22] and ζ matrices calculated according to the method of Meal and Polo.[58, 59] These algorithms are detailed in Appendix B.

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

$$\Phi_{\mathbf{n}}(Q_1, \dots, Q_M) = \prod_{i=1}^M \phi_{n_i}(Q_i) \quad (5)$$

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

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

$$\langle \Phi_{\mathbf{n}}(Q_1, \dots, Q_M) | \hat{H}_{\text{vib}} + \hat{H}_{\text{Cor}} | \Phi_{\mathbf{n}'}(Q_1, \dots, Q_M) \rangle \quad (6)$$

for all unique combinations of Hartree product basis functions, then diagonalise the VCI matrix. The fundamental integrals required to evaluate both the vibrational and Coriolis correction terms are given in Tables 1 and 2. Sparse matrix array structures are used to store the VCI matrix elements.

The final VCI wavefunction for each state $\Psi_{\mathbf{n}}(Q_1, \dots, Q_M)$ is a linear combination of Hartree product basis states $\Phi_{\mathbf{n}'}(Q_1, \dots, Q_M)$. 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.

$$\Psi_{\mathbf{n}}(Q_1, \dots, Q_M) = \sum_{\mathbf{n}'} c_{\mathbf{n}, \mathbf{n}'} \Phi_{\mathbf{n}'}(Q_1, \dots, Q_M) \quad (7)$$

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.

4. 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} for benchmark calculations and tested at a range of values between 1×10^{-7} and 5×10^{-5} . 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} .

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>.

200 5. Results

5.1. VCI convergence with respect to excitation level

Reference results were generated using a sextic force field at VCI(10) for triatomics and C₃H₂, VCI(9) for all other 4 and 5 atom molecules, and VCI(8) for C₃H₃⁺ and C₂H₄, 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:

$$\Delta_{\text{VCI}(n)}^{\text{SFF}} = \nu_{\text{ref}} - \nu_{\text{VCI}(n)}^{\text{SFF}} \quad (8)$$

$$\Delta_{\text{VCI}(n)}^{\text{QFF}} = \nu_{\text{ref}} - \nu_{\text{VCI}(n)}^{\text{QFF}} \quad (9)$$

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 1.

Table 1: 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	$\langle \Delta_{\text{VCI}(n)}^{\text{SFF}} \rangle$ (cm ⁻¹)	$ \Delta_{\text{VCI}(n)}^{\text{SFF}} _{\text{max}}$ (cm ⁻¹)	$\langle \Delta_{\text{VCI}(n)}^{\text{QFF}} \rangle$ (cm ⁻¹)	$ \Delta_{\text{VCI}(n)}^{\text{QFF}} _{\text{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

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 1 and 2.

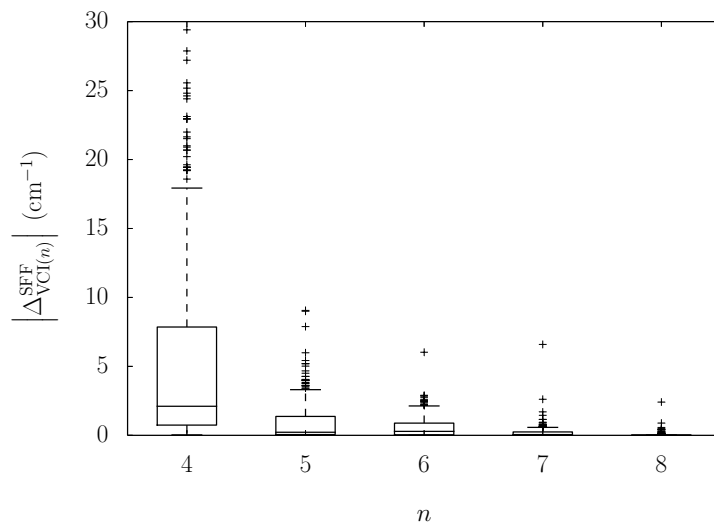


Figure 1: The magnitude and range of absolute errors in fundamental frequencies calculated using truncated VCI expansions with a sextic force field (SFF) are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, at each excitation level.

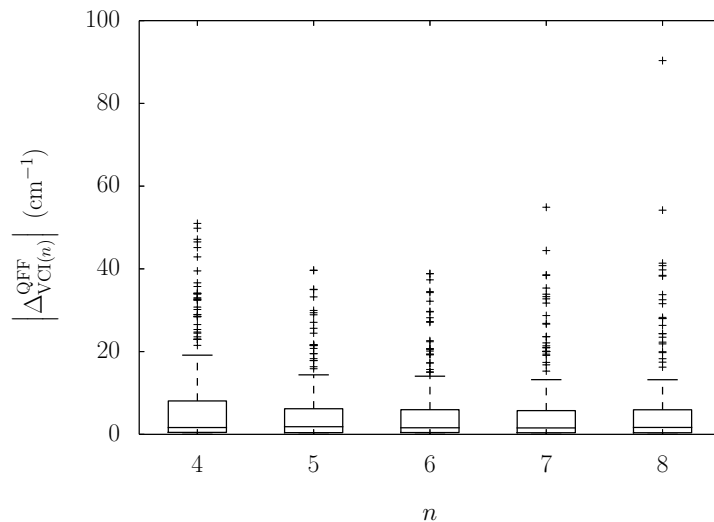


Figure 2: The magnitude and range of absolute errors in fundamental frequencies calculated using truncated VCI expansions with a quartic force field (QFF) are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, at each 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} .

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 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.

We note that this behaviour is consistent with previous observations that VPT2 often outperforms VCI when the PES expansion is truncated at fourth order.[12] However, a detailed analysis of this observation is outside the scope of the current work.

5.2. 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 2 and Figure 3, where:

$$\Delta_{\text{screen}}^{\text{VCI}(n)} = \nu_{\text{VCI}(n)}^{\text{SFF}} - \nu_{\text{VCI}(n),\text{screened}}^{\text{SFF}} \quad (10)$$

The fractional reduction in number of matrix elements across the test data set is represented in Figure 4 and broken down by molecule size in Figure 5.

Overall, a screening threshold of 2×10^{-5} 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 5). Therefore, matrix element screening becomes more useful the larger the VCI matrix and harder the diagonalization problem.

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

Table 2: Mean and maximum absolute errors arising from VCI matrix element screening with tabulated threshold values.

Threshold (E_h)	$\langle \Delta_{\text{screen}}^{\text{VCI}(8)} \rangle$ (cm^{-1})	$ \Delta_{\text{screen}}^{\text{VCI}(8)} _{\text{max}}$ (cm^{-1})	$\langle \Delta_{\text{screen}}^{\text{VCI}(4)} \rangle$ (cm^{-1})	$ \Delta_{\text{screen}}^{\text{VCI}(4)} _{\text{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

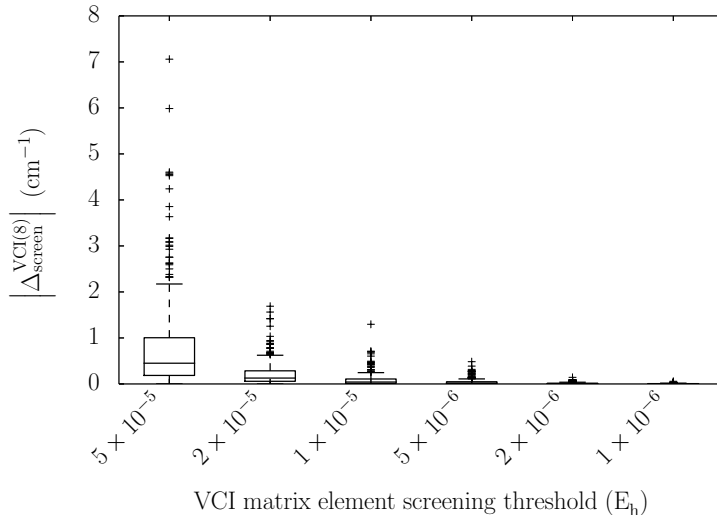


Figure 3: The magnitude and range of absolute errors in fundamental frequencies due to matrix element screening are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, for each threshold value.

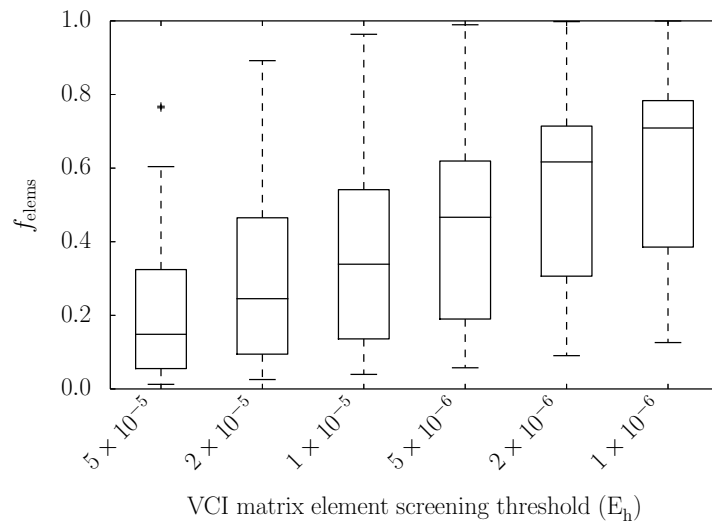


Figure 4: The fractional number of VCI matrix elements to be stored varies significantly with screening threshold (x -axis) and across all molecules without low-barrier torsional modes within the PyPES library, as indicated using the boxplot format.

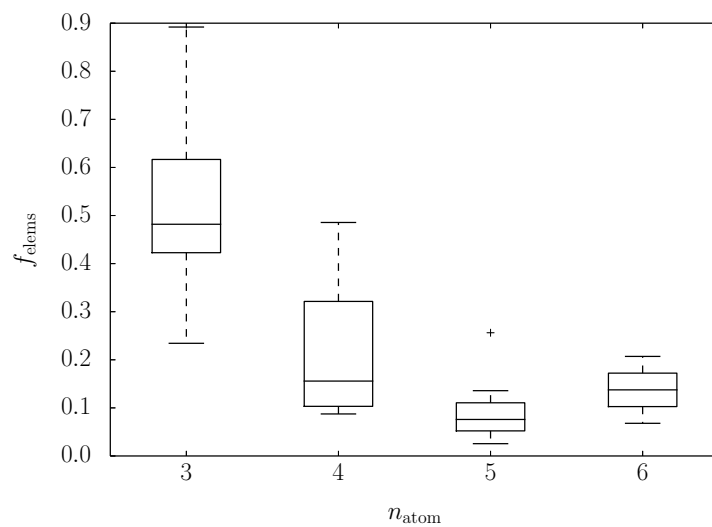


Figure 5: The fractional number of VCI matrix elements to be stored at a screening threshold of 2×10^{-5} is broken down by molecular size (x -axis), with collated data represented in boxplot format.

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 2 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.

5.3. 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 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:

$$\Delta_{\text{screen}}^{\text{SFF}} = \nu_{\text{VCI}(8)}^{\text{SFF}} - \nu_{\text{VCI}(8)}^{\text{SFF,screened}} \quad (11)$$

Statistical analysis of the combined results, for a series of different threshold values, are presented in Table 3 and illustrated in Figure 6.

Table 3: 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)	$\langle \Delta_{\text{screen}}^{\text{SFF}} \rangle$ (cm ⁻¹)	$ \Delta_{\text{screen}}^{\text{SFF}} _{\text{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

The data presented in Table 3 and Figure 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 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 $n = 2$.

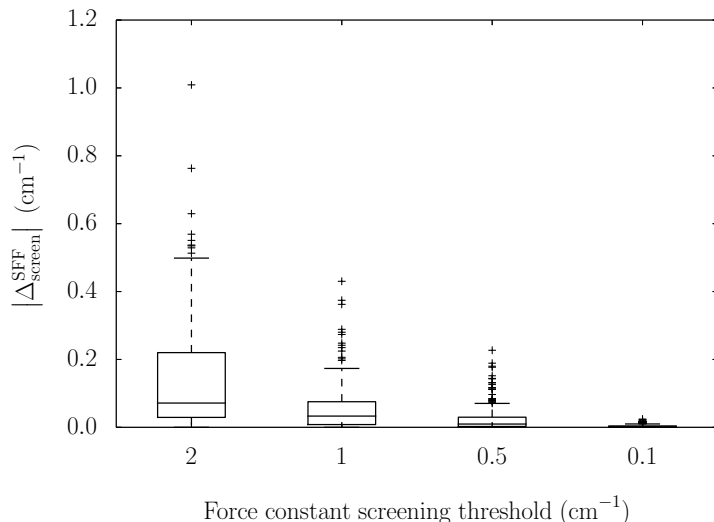


Figure 6: The magnitude and range of absolute errors in fundamental frequencies due to force constant screening are shown using the boxplot format. Data are aggregated across all molecules without low-barrier torsional modes within the PyPES library, for each threshold value.

5.4. Effect of excluding low-barrier torsional modes

Low-barrier torsional modes are problematic when solving the nuclear vibra-
 280 tional 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 delocalized vibrational modes. The most expedient solution would be to exclude these modes entirely from the VCI expansion, but this could decrease how accurately the
 285 remaining fundamental frequencies could be predicted, particularly if there is strong coupling between torsional modes and others.

To quantify this effect, we have performed VCI(8) calculations on a SFF excluding the torsional mode from the VCI expansion for each of the molecules listed in Table 4. Errors are calculated relative to benchmark literature values:

$$\Delta_{\text{tors}} = \nu_{\text{ref}} - \nu_{\text{VCI}(8)} \quad (12)$$

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
 290 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
 295 also contributes to the relatively large observed Δ_{tors} errors for HOOH.

Table 4: 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 [60], c) MULTIMODE [28] and d) SPECTRO [50]. ‡ For HOOH, strong tunnelling splitting leads to doublet spectral peaks. Reference data are derived by averaging split peak positions.

Molecule	Reference method, [citation]	$\langle \Delta_{\text{tors}} \rangle$ (cm^{-1})	$ \Delta_{\text{tors}} _{\text{max}}$ (cm^{-1})
HOOH	VCI ^a [61]	7.3 [‡]	13.4 [‡]
HSOH	VCI ^b [62]	3.4	6.8
<i>cis</i> -HOCO	VCI ^c [63]	3.2	7.8
<i>trans</i> -HOCO	VCI ^c [64]	3.2	9.6
<i>cis</i> -HSiOH	VPT2 ^d [65]	9.2	20.2
<i>trans</i> -HSiOH	VPT2 ^d [65]	5.7	13.2

5.5. Recommendations

Final algorithmic recommendations are presented in Table 5. 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
 300 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 errors in Table 5 are likely to be a significant overestimate, as some error cancellation is to be expected.

305 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,
 310 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
 315 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
 320 nuclear vibrational structure models.

Table 5: 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

6. 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

Table 6: 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	$\frac{\sqrt{n(n-1)(n-2)(n-3)(n-4)(n-5)}}{8\omega_i^3}$

Table 7: Fundamental Coriolis coupling integrals for vibrational mode i , $\langle \phi_{n'_i}(Q_i) | \hat{O} | \phi_{n_i}(Q_i) \rangle$ involving operators $\hat{O} = \frac{\partial}{\partial Q_i}, Q_i \frac{\partial}{\partial Q_i}, \frac{\partial}{\partial Q_i} Q_i$

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

References

- 335 [1] The PyPES Library of Potential Energy Surfaces.
URL <http://pypes-lib.sourceforge.net/>
- [2] M. Sibae, D. L. Crittenden, The PyPES library of high quality semi-global potential energy surfaces, J. Comp. Chem. 36 (2015) 2200–2207.
- [3] PES Database.
340 URL <http://pes-database.theochem.uni-stuttgart.de/surfaces/index.php>
- [4] G. Rauhut, Efficient calculation of potential energy surfaces for the generation of vibrational wave functions, J. Chem. Phys. 121 (2004) 9313–9322. doi:{10.1063/1.1804174}.
- 345 [5] PotLib Potential Energy Surface Library.
URL <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, POTLIB: A potential energy surface library for chemical systems, Comp. Phys. Comm. 144 (2002) 169–187. doi:{10.1016/S0010-4655(01)00437-4}.
- 350 [7] ezPES Library of Potential Energy Surfaces.
URL <http://iopshell.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, Ab-initio-based potential energy surfaces for complex molecules and molecular complexes, *J. Phys. Chem. Lett.* 1 (2010) 1866–1874.
- [9] B. J. Braams, J. M. Bowman, Permutationally invariant potential energy surfaces in high dimensionality, *Int. Rev. Phys. Chem.* 28 (2009) 577–606. doi:{10.1080/01442350903234923}.
- [10] R. Ramakrishnan, G. Rauhut, Semi-quartic force fields retrieved from multi-mode expansions: Accuracy, scaling behavior, and approximations, *J. Chem. Phys.* 142 (2015) 154118. doi:10.1063/1.4918587.
- [11] M. Ringholm, D. Jonsson, R. Bast, B. Gao, A. J. Thorvaldsen, U. Ekström, T. Helgaker, K. Ruud, Analytic cubic and quartic force fields using density-functional theory, *J. Chem. Phys.* 140 (2014) 034103. doi:{10.1063/1.4861003}.
- [12] R. C. Fortenberry, X. Huang, A. Yachmenev, W. Thiel, T. J. Lee, On the use of quartic force fields in variational calculations, *Chem. Phys. Lett.* 574 (2013) 1–12. doi:{10.1016/j.cplett.2013.03.078}.
- [13] G. Rauhut, B. Hartke, Modeling of high-order many-mode terms in the expansion of multidimensional potential energy surfaces: Application to vibrational spectra, *J. Chem. Phys.* 131 (2009) 014108. doi:{10.1063/1.3160668}.
- [14] M. Neff, G. Rauhut, Toward large scale vibrational configuration interaction calculations, *J. Chem. Phys.* 131 (2009) 124129. doi:{10.1063/1.3243862}.
- [15] E. Matito, D. Toffoli, O. Christiansen, A hierarchy of potential energy surfaces constructed from energies and energy derivatives calculated on grids, *J. Chem. Phys.* 130 (2009) 134104. doi:{10.1063/1.3092921}.
- [16] C. Y. Lin, A. T. B. Gilbert, P. M. W. Gill, Calculating molecular vibrational spectra beyond the harmonic approximation, *Theor. Chem. Acc.* 120 (2008) 23–25. doi:{10.1007/s00214-007-0292-8}.
- [17] K. Yagi, S. Hirata, K. Hirao, Multiresolution potential energy surfaces for vibrational state calculations, *Theor. Chem. Acc.* 118 (2007) 681–691. doi:{10.1007/s00214-007-0363-x}.
- [18] K. Yagi, K. Hirao, T. Taketsugu, M. Schmidt, M. Gordon, Ab initio vibrational state calculations with a quartic force field: Applications to H₂CO, C₂H₄, CH₃OH, CH₃CCH, and C₆H₆, *J. Chem. Phys.* 121 (2004) 1383–1389. doi:{10.1063/1.1764501}.

- 390 [19] S. Dressler, W. Thiel, Anharmonic force fields from density functional theory, *Chem. Phys. Lett.* 273 (1997) 71–78. doi:{10.1016/S0009-2614(97)00552-6}.
- [20] M. Sibae, D. L. Crittenden, Efficient construction of anharmonic vibrational force fields by coordinate transformation from curvilinear to rectilinear normal mode coordinates, *Phys. Chem. Chem. Phys.* (2016) Submitted.
- 395 [21] T. K. Roy, R. B. Gerber, Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications, *Phys. Chem. Chem. Phys.* 15 (2013) 9468–9492. doi:{10.1039/c3cp50739d}.
- 400 [22] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, Dover Publications, Inc., 1980.
- [23] J. Bowman, The self-consistent field approach to polyatomic vibrations, *Acc. Chem. Res.* 19 (1986) 202–208. doi:{10.1021/ar00127a002}.
- 405 [24] A. Wierzbicki, J. M. Bowman, GVSCF: a general code to perform vibrational self-consistent field calculations, *Comp. Phys. Comm.* 51 (1988) 225.
- [25] S. Carter, S. Culik, J. Bowman, Vibrational self-consistent field method for many-mode systems: A new approach and application to the vibrations of CO adsorbed on Cu(100), *J. Chem. Phys.* 107 (1997) 10458–10469. doi:{10.1063/1.474210}.
- 410 [26] S. Carter, J. Bowman, N. Handy, Extensions and tests of “multi-modes”: a code to obtain accurate vibration/rotation energies of many-mode molecules, *Theor. Chem. Acc.* 100 (1998) 191–198. doi:{10.1007/s002140050379}.
- 415 [27] N. Wright, R. Gerber, Direct calculation of anharmonic vibrational states of polyatomic molecules using potential energy surfaces calculated from density functional theory, *J. Chem. Phys.* 112 (2000) 2598–2604. doi:{10.1063/1.480833}.
- 420 [28] J. Bowman, S. Carter, X. Huang, MULTIMODE: a code to calculate rovibrational energies of polyatomic molecules, *Int. Rev. Phys. Chem.* 22 (2003) 533–549. doi:{10.1080/0144235031000124163}.
- [29] D. Benoit, Fast vibrational self-consistent field calculations through a reduced mode-mode coupling scheme, *J. Chem. Phys.* 120 (2004) 562–573. doi:{10.1063/1.1631817}.
- 425 [30] O. Christiansen, Vibrational structure theory: new vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties, *Phys. Chem. Chem. Phys.* 9 (2007) 2942–2953. doi:{10.1039/b618764a}.

- [31] N. Matsunaga, G. Chaban, R. Gerber, Degenerate perturbation theory corrections for the vibrational self-consistent field approximation: Method and applications, *J. Chem. Phys.* 117 (2002) 3541–3547. doi:{10.1063/1.1494978}.
- [32] O. Christiansen, Moller-Plesset perturbation theory for vibrational wave functions, *J. Chem. Phys.* 119 (2003) 5773–5781. doi:{10.1063/1.1601593}.
- [33] V. Barone, Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation, *J. Chem. Phys.* 120 (2004) 3059–3065. doi:{10.1063/1.1637580}.
- [34] V. Barone, Anharmonic vibrational properties by a fully automated second-order perturbative approach, *J. Chem. Phys.* 122 (2005) 014108. doi:{10.1063/1.1824881}.
- [35] P. Cassam-Chenai, J. Lievin, Alternative perturbation method for the molecular vibration-rotation problem, *Int. J. Quant. Chem.* 93 (2003) 245–264. doi:{10.1002/qua.10556}.
- [36] W. Mizukami, D. P. Tew, A second-order multi-reference perturbation method for molecular vibrations, *J. Chem. Phys.* 139 (2013) 194108. doi:{10.1063/1.4830100}.
- [37] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, Numerical-Analytic Implementation of the Higher-Order Canonical Van Vleck Perturbation Theory for the Interpretation of Medium-Sized Molecule Vibrational Spectra, *J. Phys. Chem. A* 116 (2012) 3691–3709. doi:{10.1021/jp211400w}.
- [38] J. Bloino, M. Biczysko, V. Barone, General Perturbative Approach for Spectroscopy, Thermodynamics, and Kinetics: Methodological Background and Benchmark Studies, *J. Chem. Theor. Comp.* 8 (2012) 1015–1036. doi:{10.1021/ct200814m}.
- [39] S. V. Krasnoshchekov, E. V. Isayeva, N. F. Stepanov, Criteria for first- and second-order vibrational resonances and correct evaluation of the Darling-Dennison resonance coefficients using the canonical Van Vleck perturbation theory, *J. Chem. Phys.* 141 (2014) 234114. doi:{10.1063/1.4903927}.
- [40] A. M. Rosnik, W. F. Polik, VPT2+K spectroscopic constants and matrix elements of the transformed vibrational Hamiltonian of a polyatomic molecule with resonances using Van Vleck perturbation theory, *Mol. Phys.* 112 (2014) 261–300. doi:{10.1080/00268976.2013.808386}.
- [41] V. Barone, M. Biczysko, J. Bloino, Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation, *Phys. Chem. Chem. Phys.* 16 (2014) 1759–1787. doi:{10.1039/c3cp53413h}.

- [42] M. Piccardo, J. Bloino, V. Barone, Generalized vibrational perturbation theory for rovibrational energies of linear, symmetric and asymmetric tops: Theory, approximations, and automated approaches to deal with medium-to-large molecular systems, *Int. J. Quant. Chem.* 115 (2015) 948–982. doi:{10.1002/qua.24931}.
- [43] S. Carter, N. Handy, The variational method for the calculation of rovibrational energy levels, *Comp. Phys. Rep.* 5 (1986) 115–172.
- [44] C. Fabri, T. Furtenbacher, A. G. Csaszar, A hybrid variational-perturbational nuclear motion algorithm, *Mol. Phys.* 112 (2014) 2462–2467. doi:{10.1080/00268976.2014.921341}.
- [45] O. Christiansen, Vibrational coupled cluster theory, *J. Chem. Phys.* 120 (2004) 2149–2159. doi:{10.1063/1.1637579}.
- [46] G. Rauhut, G. Knizia, H.-J. Werner, Accurate calculation of vibrational frequencies using explicitly correlated coupled-cluster theory, *J. Chem. Phys.* 130 (5) (2009) 054105. doi:{10.1063/1.3070236}.
- [47] P. Seidler, M. Sparta, O. Christiansen, Vibrational coupled cluster response theory: A general implementation, *J. Chem. Phys.* 134 (2011) 054119. doi:{10.1063/1.3536499}.
- [48] O. Christiansen, Selected new developments in vibrational structure theory: potential construction and vibrational wave function calculations, *Phys. Chem. Chem. Phys.* 14 (2012) 6672–6687. doi:{10.1039/c2cp40090a}.
- [49] P. Daněček, P. Bouř, Comparison of the numerical stability of methods for anharmonic calculations of vibrational molecular energies, *J. Comp. Chem.* 28 (2007) 1617–1624.
- [50] J. F. Gaw, A. Willetts, W. H. Green, N. C. Handy, Spectro: a program for the derivation of spectroscopic constants from provided quartic force fields and cubic dipole fields, *Advances in Molecular Vibrations and Collision Dynamics, A Research Annual 1B* (1991) 169.
- [51] E. Kamarchik, J. Bowman, A. Krylov, ezMode.
URL <http://iopenshell.usc.edu/downloads/ezmode>
- [52] P. Daněček, gvib.
URL <http://gvib.sourceforge.net>
- [53] P. Carbonniere, A. Dargelos, C. Pouchan, The VCI-P code: an iterative variation-perturbation scheme for efficient computations of anharmonic vibrational levels and IR intensities of polyatomic molecules, *Theor. Chem. Acc.* 125 (2010) 543–554. doi:{10.1007/s00214-009-0689-7}.

- [54] G. Chaban, J. Jung, R. Gerber, Ab initio calculation of anharmonic vibrational states of polyatomic systems: Electronic structure combined with vibrational self-consistent field, *J. Chem. Phys.* 111 (1999) 1823–1829. doi:{10.1063/1.479452}.
- [55] The Extensible PyPES Library of Potential Energy Surfaces.
URL <http://pypes-lib-ext.sourceforge.net/>
- [56] H. Romanowski, J. M. Bowman, L. B. Harding, Vibrational energy levels of formaldehyde, *J. Chem. Phys.* 82 (1985) 4155–4165.
- [57] N. Gohaud, D. Begue, C. Darrigan, C. Pouchan, New parallel software (p_anhar) for anharmonic vibrational calculations: Application to (ch3li)2, *J. Comp. Chem.* 26 (2005) 743–754.
- [58] J. Meal, S. Polo, Vibration-rotation interaction in polyatomic molecules .1. Zeta-matrices, *J. Chem. Phys.* 24 (1956) 1119–1125. doi:{10.1063/1.1742728}.
- [59] J. Meal, S. Polo, Vibration-rotation interaction in polyatomic molecules .2. Determination of Coriolis coupling coefficients, *J. Chem. Phys.* 24 (1956) 1126–1133. doi:{10.1063/1.1742729}.
- [60] S. N. Yurchenko, W. Thiel, P. Jensen, Theoretical ROVibrational Energies (TROVE): A robust numerical approach to the calculation of rovibrational energies for polyatomic molecules, *J. Mol. Spec.* 245 (2007) 126–140. doi:{10.1016/j.jms.2007.07.009}.
- [61] P. Malyszczek, J. Koput, Accurate ab initio potential energy surface and vibration-rotation energy levels of hydrogen peroxide, *J. Comp. Chem.* 34 (5) (2013) 337–345. doi:{10.1002/jcc.23137}.
- [62] S. N. Yurchenko, A. Yachmenev, W. Thiel, O. Baum, T. F. Giesen, V. V. Melnikov, P. Jensen, An ab initio calculation of the vibrational energies and transition moments of HSOH, *J. Mol. Spec.* 257 (1) (2009) 57–65. doi:{10.1016/j.jms.2009.06.010}.
- [63] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, Vibrational frequencies and spectroscopic constants from quartic force fields for cis-HOCO: The radical and the anion, *J. Chem. Phys.* 135 (21). doi:{10.1063/1.3663615}.
- [64] R. C. Fortenberry, X. Huang, J. S. Francisco, T. D. Crawford, T. J. Lee, The trans-HOCO radical: Quartic force fields, vibrational frequencies, and spectroscopic constants, *J. Chem. Phys.* 135 (13). doi:{10.1063/1.3643336}.
- [65] J. Martin, Anharmonic force fields and accurate thermochemistry of H2SiO, cis-HSiOH, and trans-HSiOH, *J. Phys. Chem. A* 102 (8) (1998) 1394–1404. doi:{10.1021/jp9731718}.