Using computed infrared intensities for fast computation of vibrational spectra

Vincent Le Bris, Marc Odunlami, Didier Bégué, Isabelle Baraille, Olivier Coulaud

Using computed infrared intensities for fast computation of vibrational spectra

Vincent Le Bris,[1, a)] Marc Odunlami,[1] Didier Bégué,[1] Isabelle Baraille,[1] and Olivier Coulaud[2]

[1)]*Université de Pau et des Pays de l'Adour/CNRS/E2S UPPA, Institut des Sciences Analytiques et de Physicochimie pour l'Environnement et les Matériaux, UMR5254, 64000 Pau, France*

[2)]*HiePACS project-team, Inria Bordeaux Sud-Ouest, 200, avenue de la Vieille Tour, F-33405 Talence Cedex*

(Dated: 16 March 2020)

The Adaptive Vibrational Configuration Interaction (A-VCI) algorithm is an iterative process able to compute the spectrum of an Hamiltonian operator, using a discretization basis as small as possible. In this work, we show how this algorithm can handle more sophisticated operators, which ro-vibrational Coriolis coupling terms. In order to overcome the increase of computing and storage resources needed due to this enrichment, the InfraRed (IR) intensities are computed and used as a criterion to select only the eigenstates corresponding to IR active vibrational states. The benefits of this new approach are presented for a few well studied molecular systems (H_2O, H_2CO, CH_2NH, CH_3CN, C_2H_4O), and it is ultimately applied to a 10-atom molecule ($C_4H_4N_2$).

[a)]Electronic mail: vincent.lebris@univ-pau.fr

CONTENTS

I. INTRODUCTION

For many decades, chemists have sought to identify and characterize molecules in different chemical environments with varying degrees of success. Among the experimental techniques commonly used, InfraRed (IR) vibrational spectroscopy has undoubtedly made the greatest progress in biochemistry, chemical reactivity, interstellar chemistry, complex matrix chemistry and materials chemistry. The complexity of vibrational data, particularly in spectral areas with high density of states (mid-IR) or with very low active signals (near IR), makes the use of predictive modeling based on quantum mechanics calculations an essential support to the interpretation of experimental spectra.

In order to reach a useful accuracy for these calculations, it is necessary to introduce, according to Watson's definition[1], the terms arising from vibration-rotation interactions, especially the rovibrational terms of Coriolis[2–4]. For low mass systems such as water, the correction of rotation-vibration coupling terms can reach[5] several tens of cm^{-1}.

In addition, the experimental data also contains the absolute intensities of the vibrational transitions. These are as difficult to measure as they are to calculate, but this information is mandatory for the full interpretation of the IR spectra. In fact, relatively a few codes propose such developments in the context of electrical anharmonicity[6–12]. We can mention Multimode[13,14] (based on a variational approach of the problem), Gaussian[15] which allows us to access the infrared intensities beyond the double-harmonic approximation within the VPT2 framework[16], and Molpro[17] in which the VSCF and VCI approaches are implemented.

A downside is that the inclusion of the Coriolis terms in the Hamiltonian and the calculation of the intensities entail an increase in the computational complexity and the memory footprint. For example, the impact of the Coriolis treatment is estimated at about 30% for the systems we have studied in this work.

The A-VCI algorithm[18,19] has been developed to effectively reduce the number of vibrational states used in the configuration interaction process. It constructs a nested basis for the discretization of the Hamiltonian operator and uses an *a-posteriori* error estimator (residue) to select the most relevant directions to expand the discretization space. The anharmonic contributions of the potential energy is expanded as a Taylor series in the normal coordinates of vibration up to the fourth order. Currently, the robustness and reliability of this method has been tested on molecules of 4 atoms (formaldehyde), 6 atoms (acetonitrile) and 7 atoms (ethylene oxide) for calculated vi-

brational spectra up to $3\,000\,\mathrm{cm}^{-1}$. In order to maintain the possibility to deal with larger systems and higher accuracy, the operator has to be as complete as possible and it is necessary to keep control over the memory footprint and the execution time of the A-VCI algorithm. In this work, we explore the idea that the information on IR intensities makes it possible to considerably reduce the number of target eigenvalues that need to be precisely calculated. The implementation of the intensities was carried out within the A-VCI algorithm to focus on the active vibrational states with the objective of effectively reducing the size of the problem without any loss of accuracy.

This report starts with a reminder of the context, in order to introduce the notations, with details on the Coriolis operator and the calculation of vibrational intensities. After a brief reminder of the method, we introduce the algorithm used to take advantage of the intensities to reach convergence more quickly. Finally, to highlight the benefits of these new developments, we present numerical results on a few molecular systems (H_2O, H_2CO, CH_2NH, CH_3CN, C_2H_4O). This method is then pushed to a 10-atom molecular system ($C_4H_4N_2$).

II. FORMALISM, ALGORITHM OVERVIEW AND NEW DEVELOPMENTS

A. Overview

Computing the vibrational frequencies of a molecule requires calculating the eigenvalues of the Hamiltonian operator $\mathscr{H}$. Given a N-atom molecular system with $D = 3N-6$ degrees of freedom, we consider the vibrational Hamiltonian $\mathscr{H}$

$$\mathscr{H}(\mathbf{q}) = \mathscr{H}_0(\mathbf{q}) + \mathscr{V}(\mathbf{q}) + \mathscr{C}(\mathbf{q}), \tag{1}$$

with $\mathscr{H}_0(\mathbf{q}) = \sum_{i=1}^{D} \frac{\omega_i}{2}\left(p_i^2 + q_i^2\right)$ the harmonic operator, $\mathscr{V}(\mathbf{q})$ the anharmonic Potential Energy Surface (PES), $\mathscr{C}(\mathbf{q})$ the second order Coriolis operator, $\mathbf{q} = (q_1, q_2, \ldots, q_D)$ the normal dimensionless coordinates and the conjugate momentum $p_i = -\mathrm{i}\dfrac{\partial}{\partial q_i}$. The PES operator is a Taylor expansion of order S, which is a sum of monomials products. The Coriolis operator is written as

$$\mathscr{C}(\mathbf{q}) = \sum_{\alpha=x,y,z} B_\alpha \sum_{i,j\neq i,k,l\neq k} \zeta_{ij}^{\alpha}\zeta_{kl}^{\alpha} q_i p_j q_k p_l \sqrt{\frac{\omega_j\omega_l}{\omega_i\omega_k}}, \tag{2}$$

with ω_i the harmonic frequency related to the coordinate q_i, B_α the rotational constant of the $\alpha = x,y,z$ axis (in cm^{-1}), and ζ_{kl}^{α} the Coriolis constant coupling q_k and q_l through rotation along the α axis with $\zeta_{kl}^{\alpha} = -\zeta_{lk}^{\alpha}$.

Let Π be the space spanned by the eigenfunctions of the harmonic operator, $\mathscr{H}_0$. These eigenvectors, $\phi^0_{\mathbf{n}}$, write as the product of D one-dimensional Hermite functions of degrees $\mathbf{n} = \{n_1, n_2, \ldots, n_D\}$. Let $\mathbf{d} = \{d_1, d_2, \ldots, d_D\}$ the maximal degree of these eigenfunctions, then we define the approximation space $\Pi_{\mathbf{d}}$ as a subspace of Π by

$$\Pi_{\mathbf{d}} = \left\{ \phi^0_{\mathbf{n}} \,/\, \mathbf{n} \in \prod_{i=1}^{D} [0, d_i] \right\}.$$

The calculation of the spectrum of the Hamiltonian operator leads to the computation of the eigenvalues of the matrix H, the discretization of the operator $\mathscr{H}$ in $\Pi_{\mathbf{d}}$. In the framework of the variational approach, the coefficients of the H matrix are written as

$$H_{(i,j)} = \langle \phi^0_{\mathbf{i}}(\mathbf{q}) | \left(\mathscr{H}_0(\mathbf{q}) + \mathscr{V}(\mathbf{q}) + \mathscr{C}(\mathbf{q}) \right) \phi^0_{\mathbf{j}}(\mathbf{q}) \rangle .$$

The calculation of the $\langle \phi^0_{\mathbf{i}}(\mathbf{q}) | \mathscr{C}(\mathbf{q}) \phi^0_{\mathbf{j}}(\mathbf{q}) \rangle$ is given by the formulae in Carbonniere et al.[20], whereas the formulae for the integrals involving the harmonic operator and the anharmonic PES are given in our previous work[19].

B. Calculation of infrared intensities

In addition to the precise vibrational frequencies, knowledge of IR related intensities represents additional information useful to complement experimental attributions for non-fundamental bands. The calculation of IR intensities beyond the electrical harmonicity has already been considered in the literature[12,21,22]. It is based on the non-linear dependence of the dipole moment with respect to the normal coordinates.

Let $(E_k, \mathbf{X}_k)$ be the k eigenpairs of the $m \times m$ matrix H. As shown in Ref. 21 the intensity I_k between the vibrational state k and the ground state $(E_0, \mathbf{X}_0)$ is

$$I_k = \frac{8\pi^3 N_{Av}}{3hc(4\pi\varepsilon_0)} (E_k - E_0) \sum_{\alpha=x,y,z} |R_{\alpha,k}|^2,$$

where N_{Av} is the Avogadro's number, $R_{\alpha,k}$ the transition dipole moment between the states 0 and k in the α-direction, E_0 and E_k the first and the $(k+1)^{th}$ eigenvalues. Noting the constant $C = \dfrac{8\pi^3 N_{Av}}{3hc(4\pi\varepsilon_0)}$ and converting it to standard units leads to the following formula

$$I_k = 16.194105 \times (E_k - E_0) \sum_{\alpha=x,y,z} |R_{\alpha,k}|^2 \quad \text{km/mol}, \tag{3}$$

with E_0 and E_k in cm^{-1}, and $R_{\alpha,k}$ in a.u. The transition dipole moments is written as

$$R_{\alpha,k} = \langle \mathbf{X}_0 | \mu_\alpha(\mathbf{q}) \mathbf{X}_k \rangle = \langle \mathbf{X}_k | \mu_\alpha(\mathbf{q}) \mathbf{X}_0 \rangle , \tag{4}$$

where $\mu = (\mu_x, \mu_y, \mu_z)$ is the dipole operator. Each dipole moment surface μ_α is a Taylor expansion with respect to the D-dimensional $\mathbf{q}$ variable and is written as

$$\mu_\alpha(\mathbf{q}) = \sum_{\|\mathbf{s}\|_1=1}^{\tilde{S}} C_{\alpha,\mathbf{s}} \mathbf{q}^{\mathbf{s}},$$

where $\|.\|_1$ is the usual 1-norm and $\tilde{S}$ is its maximal degree, which verifies $\tilde{S} \leq S-1$, since μ is the first derivative of the energy with respect to the electric field.

C. The A-VCI algorithm

1. Quick algorithm overview

The A-VCI algorithm[18,19] is an iterative procedure used to calculate the first F eigenpairs of the Hamiltonian spectrum in a very large space by guaranteeing the accuracy of the calculation. At the end of the computation we obtain a representation of the F smallest eigenpairs in a subset B of $\Pi_{\mathbf{d}}$.

The key point of the approach is the decomposition of the image of a subspace $B^{(j)}$ at iteration j, noted $\mathscr{H}(B^{(j)})$, by the operator $\mathscr{H}$ in the direct sum of two orthogonal spaces $B^{(j)}$ and $B_R^{(j)}$. Our estimator on the difference between an eigenpair $(E, \mathbf{X})$ of the operator discretized in $B^{(j)}$ and the corresponding eigenvalue $\tilde{E}$ in $\mathscr{H}(B^{(j)})$ is

$$|E - \tilde{E}| \leq \|{B_R^{(j)}}^T \mathscr{H} B^{(j)} \mathbf{X}\|_2,$$

where $\|.\|_2$ is the usual Euclidian norm.

We briefly recall the main steps of the A-VCI algorithm. After defining an initial basis $B^{(0)}$ belonging to the $\Pi_{\mathbf{d}}$ space, we construct the sparse structures of the matrices $H^{(0)} = {B^{(0)}}^T \mathscr{H} B^{(0)}$ and $H_R^{(0)} = {B_R^{(0)}}^T \mathscr{H} B^{(0)}$, *i.e.* only the row and column indices of the non-zero elements. During this step we also build the basis of admissible nodes $B_R^{(0)} = \mathscr{H}(B^{(0)}) \backslash B^{(0)}$, which is needed to compute the sparse structure of the matrix $H_R^{(0)}$. The iterative procedure begins by calculating the terms of the Hamiltonian matrix, then the first F eigenpairs are computed by an iterative eigensolver.

To check the convergence of the algorithm, we evaluate the scaled residue for all eigenpairs $(E_\ell^{(j)}, \mathbf{X}_\ell^{(j)})$

$$\mathbf{r}_\ell^{(j)} = H_R^{(j)} \mathbf{X}_\ell^{(j)} / E_\ell^{(j)}. \tag{5}$$

If the maximum value of the norms $\|\mathbf{r}_\ell^{(j)}\|_2$ for $\ell = 1, \ldots, F$ is lower than the target threshold ε, the method has converged. This evaluation means computing each coefficient of the rectangle matrix $H_R^{(j)}$. If the convergence is not reached, we build the new active space, $B^{(j+1)}$, by adding directions selected in $B_R^{(j)}$ to the $B^{(j)}$ basis. Finally, we update the sparse structure of the two matrices with to the newly added basis elements.

When the convergence is reached, the eigenpairs are not only the eigenpairs of the Hamiltonian discretized in $B^{(j)}$, but also a good approximation of those of the Hamiltonian discretized in $B^{(j)} \oplus B_R^{(j)}$.

2. Sparse structure of the Coriolis matrix

As described in Ref. 19, we have shown that the method is efficient if we can quickly determine the sparse structure of the Hamiltonian matrix. We now describe how the indices (rows and columns) of the non-zero elements of the Coriolis contribution are efficiently computed. Since the Coriolis operator contains first and second derivatives, the determination of its sparse structure is more complex than the computation of the PES operator's graph. Therefore, we introduce a pseudo-operator dependent only on the monomials $\mathbf{q}^\mathbf{s}$, which has the same non-zero elements as the real Coriolis operator (2).

In the Coriolis operator, each term $\zeta_{ij}^\alpha \zeta_{kl}^\alpha q_i p_j q_k p_l$ involves at most four variables with indices i, j, k, l, where $1 \leq i, j, k, l \leq D$. In addition, since $j \neq i$ and $l \neq k$ the maximal number of times each index appears in such product is $S_C = 2$. For a given basis function $\phi_\mathbf{n}^0$, the set of all nodes connected to $\mathbf{n}$ through a term of the Coriolis surface is

$$C_{i,j,k,l}(\mathbf{n}) = \{\phi_\mathbf{m}^0 \ / \ \langle q_i p_j q_k p_l \phi_\mathbf{n}^0 | \phi_\mathbf{m}^0 \rangle \neq 0\}.$$

Finally, the set of all nodes connected to $\mathbf{n}$ is obtained by iterating on all the elements of the Coriolis operator:

$$C_{Cor}(\mathbf{n}) = \bigcup_{i,j \neq i,k,l \neq k} C_{i,j,k,l}(\mathbf{n}).$$

The number of nodes in $C_{Cor}(\mathbf{n})$ is bounded by $4(S_C+1)N_C = 12N_C$, where N_C is the number of elements of the Coriolis surface. The complexity of calculating $C_{Cor}(\mathbf{n})$ grows linearly with N_C, and is directly related to the number of Coriolis constants ζ_{ij}^{α}.

Let us consider two one-dimensional Hermite functions ψ_n^0 and ψ_m^0 related to the 1-D normal dimensionless coordinate q and the corresponding conjugate momentum p. Thanks to the properties of the Hermite functions, we have

$$\{\psi_m^0 \,/\, \langle q\,\psi_n^0|\psi_m^0\rangle \neq 0\} = \{\psi_m^0 \,/\, \langle p\,\psi_n^0|\psi_m^0\rangle \neq 0\},$$

and

$$\begin{aligned}\{\psi_m^0 \,/\, \langle q^2\,\psi_n^0|\psi_m^0\rangle \neq 0\} &= \{\psi_m^0 \,/\, \langle qp\,\psi_n^0|\psi_m^0\rangle \neq 0\} \\ &= \{\psi_m^0 \,/\, \langle pq\,\psi_n^0|\psi_m^0\rangle \neq 0\} \\ &= \{\psi_m^0 \,/\, \langle p^2\,\psi_n^0|\psi_m^0\rangle \neq 0\}.\end{aligned}$$

Indeed, $\langle q\,\psi_n^0|\psi_m^0\rangle$ and $\langle p\,\psi_n^0|\psi_m^0\rangle$ are non-zero for the same values of n and m. It is also true for the integrals with q^2, qp, pq and p^2 operators. Since $S_C = 2$, the Coriolis term $\zeta_{ij}^{\alpha}\zeta_{kl}^{\alpha}q_ip_jq_kp_l$ only involves products of type q, p, q^2, qp, pq and p^2. Consequently, we have the following proposition.

Proposition 1 *Consider the normal dimensionless coordinate $\mathbf{q} = (q_a)_{a=1,D}$ and the multi-index (i,j,k,l), where $j \neq i$ and $l \neq k$. Let introduce the multi-index $\tilde{\mathbf{s}}(i,j,k,l) = (\tilde{s}_a)_{a=1,D}(i,j,k,l)$, where $\tilde{s}_a(i,j,k,l)$ is the number of times the index 'a' appears in the multi-index (i,j,k,l). Then*

$$C_{i,j,k,l}(\mathbf{n}) = C_{CPS}(\tilde{\mathbf{s}}(i,j,k,l),\mathbf{n}) = \{\phi_{\mathbf{m}}^0 \,/\, \langle \mathbf{q}^{\tilde{\mathbf{s}}(i,j,k,l)}\phi_{\mathbf{n}}^0|\phi_{\mathbf{m}}^0\rangle \neq 0\}.$$

Let introduce the following Coriolis Pseudo-Surface (CPS) set

$$C_{CPS}(\mathbf{n}) = \bigcup_{i,j\neq i,k,l\neq k} C_{CPS}(\tilde{\mathbf{s}}(i,j,k,l),\mathbf{n}).$$

Thanks to Prop. 1, it is easy to see that the two sets $C_{CPS}(\mathbf{n})$ and $C_{Cor}(\mathbf{n})$ are identical. Now the $C_{CPS}(\mathbf{n})$ set is considered to decrease the complexity of finding the non-zero elements of the Coriolis matrix. Thus, the computation of the Coriolis matrix sparse structure is conducted in the same way as for the PES.

III. SCREENING BY THE INTENSITY

A. Quick overview of the Collective Component-Wise (CCW) strategy

The performance of our hierarchical method for calculating eigenvalues depends on how the active space increases at each iteration of the algorithm. In the previous paper[19], we have introduced several strategies to limit the number of terms to be added at each iteration while ensuring the decrease of the residual norm. We now briefly recall the CCW(p) approach used in this document.

At iteration j of the AVCI algorithm, let $K^{(j)}$ be the set of the non-converged eigenpairs $(E_\ell^{(j)}, \mathbf{X}_\ell^{(j)})$, such that $K^{(j)} = \left\{ \ell \in \{1,\dots,F\} \text{ with } \|\mathbf{r}_\ell^{(j)}\|_2 > \varepsilon \right\}$, where $\mathbf{r}_\ell^{(j)}$ is the scaled residual defined in (5). We introduce the mean residual vector $\mathbf{R}^{(j)}$ such that each component is

$$R_i^{(j)} = \frac{1}{k^{(j)}} \sum_{\ell \in K^{(j)}} |(\mathbf{r}_\ell^{(j)})_i| \text{ with } i = 1,\dots,m_R^{(j)},$$

where $m_R^{(j)}$ (resp. $k^{(j)}$) is the size of space $B_R^{(j)}$ (resp. $K^{(j)}$). The main objective is to select the most relevant components in the vector $\mathbf{R}^{(j)}$ to define the direction to be added to the active space so that we trade-off convergence speed and control of the search space growth. To do that, we denote by

$$M^{(j)} = \left\{ i \in \{1,\dots,m_R^{(j)}\} \text{ such that } |R_i^{(j)}| > \frac{\varepsilon}{\sqrt{m_R^{(j)}}} \right\}$$

the set of the indices of the admissible nodes or directions of $B_R^{(j)}$ and we define the generalized average with respect to $M^{(j)}$ by

$$mean(p) = \sqrt[p]{\frac{1}{m_r^{(j)}} \sum_{i \in M^{(j)}} |R_i^{(j)}|^p}.$$

In the CCW(p) strategy, the nodes to add to $B^{(j)}$ are selected by using a general mean criterion on the components of $\mathbf{R}^{(j)}$. The set of nodes to add, $A^{(j)}$, is

$$A^{(j)} = \{\mathbf{n}_i \text{ such that } \langle \mathbf{R}^{(j)} | \mathbf{n}_i \rangle > mean(p)\}.$$

The objective is to select p to obtain the best compromise between the growth rate of the basis size and the convergence rate of the algorithm. We refer the reader to our previous work[19] for a discussion on the choice of this strategy and the involved parameters.

B. The new intensity-based selection strategy

The CCW approach is used to select basis elements from B_R, thus restricting considerably the number of nodes to be added at each iteration. However, this approach takes into account every non-converged eigenvalues. The idea here is to apply this method to a smaller subset of eigenvalues by selecting only the vibrational states with a large enough intensity when calculating the residues in order to reduce the number of residues to evaluate.

We compute the residues and only check the convergence on eigenvalues corresponding to an intensity larger than the given threshold ε_I. We consider the space of the eigenpairs we want to converge, defined by

$$K_{\varepsilon_I} = \{(E_j, \mathbf{X}_j), \text{ such that } I_j > \varepsilon_I \quad j = 1, \dots F\},$$

where I_j is the the intensity of eigenpair $(E_j, \mathbf{X}_j)$ defined in (3). We construct the scaled residues (5) only for the eigenpairs of K_{ε_I} to check if they have converged, then we apply the CCW approach on the non-converged eigenpairs of K_{ε_I}.

As can be seen in the equation (4), the computation of the intensities I_k for every eigenpair $(E_k, \mathbf{X}_k)$ depends on the first eigenvector $\mathbf{X}_0$. Since our process is iterative, it is necessary to have a good approximation of the first eigenvector $\mathbf{X}_0$ to expect accurate computation of these intensities. We need, therefore, to ensure the convergence of the first eigenpair before using the intensity based selection strategy.

Thus, this new strategy, called Selection by Intensity (**SI**), has two main steps: in the first one, the algorithm uses the classical CCW approach on the first F eigenpairs to enrich the basis until the convergence of the first eigenvector $\mathbf{X}_0$ is reached. In the second step, we continue the algorithm to achieve the convergence by using the CCW approach only for the eigenpairs with an intensity greater than ε_I.

A variant of the SI procedure, called Global Selection by Intensity (**GSI**), considers the space K_{ε_I} for all iterations without any condition on the convergence of the eigenpair $(E_0, \mathbf{X}_0)$.

C. Fast evaluation of the intensity

Now, we detail how the computation of the intensity is performed in A-VCI. Calculating these intensities (3) requires a quick assessment of the transition dipole moments (4).

The elements of the transition dipole moment matrix M^α in the $\alpha-$direction are defined by

$$M^\alpha_{i,j} = \langle \phi^0_{\mathbf{i}}(\mathbf{q}) | \mu_\alpha(\mathbf{q}) \phi^0_{\mathbf{j}}(\mathbf{q}) \rangle \, .$$

Only the elements belonging to the graph G of M^α will participate to the calculation of the transition moments $R_{\alpha,k}$.

Algorithm 1: Intensities evaluation algorithm

Data: $(E_k, \mathbf{X}_k)_{k=0}^{F-1}$ the eigenpairs

Result: I, the intensity vector for the eigenvalues E_k when $k > 0$

Build the graph G ; 1

Construct on the fly the vectors $\mathbf{Y}_0^\alpha = M^\alpha \mathbf{X}_0$ for $\alpha = x, y, z$; 2

for $k = 1$ **to** $F-1$ **do**

  Evaluate: $R_\alpha = \langle \mathbf{Y}_0 | \mathbf{X}_k \rangle$ for $\alpha = x, y, z$; 3

  Build: $I_k = C_I (E_k - E_0)(R_x^2 + R_y^2 + R_z^2)$;

Algorithm 1 explains how the intensities are computed. As we have the matrices M^α, we consider the same structure of the matrices for all α. Let C_{DMS} be the set

$$C_{DMS} = \{ \mathbf{s} \ / \ \text{ if } \exists \alpha \text{ such that } C_{\alpha,\mathbf{s}} \neq 0 \}.$$

Then, we define the scalar pseudo operator $\mu(\mathbf{q}) = \sum_{\mathbf{s} \in C_{DMS}} \mathbf{q}^{\mathbf{s}}$ to construct the combined graph G of every operator μ_α. Using this operator, we save memory by building only one graph, but we can calculate some zero terms that would not exist if we had considered a graph per operator. Once the graph is calculated, we evaluate three sparse matrix-vector products to obtain $\mathbf{Y}_0^\alpha$ from $\mathbf{X}_0$ and the coefficients of the matrices M^α calculated on the fly (line 2). Finally, for each calculated eigenvalue, we construct the three dipole transition moments (line 3) and add them together to obtain the intensity via the formula (3).

Let NNZ be the number of non zero elements of the graph G, m the size of an eigenvector and $N_{\tilde{S}}$ the maximal number of terms in the scalar pseudo operator μ. As the cost to evaluate a coefficient of the matrices M^α is $\mathscr{O}(D)$, the complexity of the algorithm becomes $\mathscr{O}(N_{\tilde{S}} D \, \mathrm{NNZ} + mF)$.

IV. RESULTS AND DISCUSSION

The introduction of the Coriolis couplings and the infrared intensities in A-VCI was tested on molecules with 3 to 10 atoms. The code is implemented is C++ and takes advantage of the OpenMP shared-memory paradigm for parallelization. We consider Arpack[23] to solve the eigenvalue problem at each iteration.

The results were obtained on two types of computers:

(i) A 96-core Intel Xeon E7-8890 node running at 2.2 GHz with 1 TB of shared memory. The Intel compiler (2019.0.117) with the following options: `-O3 -qopenmp` is used. We refer to it as brise in the sequel.

(ii) 32-core Intel Sandy Bridge E5-4650 nodes at 2.7 GHz with 256 GB of shared memory. The Intel compiler (2017 update 1) with the options `-O3 -DNDEBUG -qopenmp` was used. We refer to it as ada in the sequel.

A. Coriolis and intensities

With the Coriolis couplings introduced in the A-VCI procedure, we computed the energy levels and their intensities on 4 benchmark molecules: H_2O, H_2CO, CH_2NH and CH_3CN. The PES were approximated by a Taylor series in normal dimensionless coordinates and Table I shows the levels of calculation used to obtain the PES force constants, the dipole moments and Coriolis coefficients. For H_2O, we used an in-house 6-order PES, given in the supplementary material.

TABLE I. Level of electronic calculations for the 4 benchmarks molecules

Molecule	PES	Dipole moment and Coriolis
H_2O	CCSD(T)/d-aug-cc-pVTZ (6^{th} order)	b3lyp/cc-pVTZ[15]
H_2CO	CCSD(T)/aug-cc-pVTZ (4^{th} order)[18]	b3lyp/aug-cc-pVTZ[15]
CH_2NH	MP2/6-311G** (4^{th} order)[24–26]	MP2/6-311G**[15]
CH_3CN	CC-B3/cc-pVTZ (4^{th} order)[19,27]	b3lyp/cc-pVTZ[15]

Moreover, a reference calculation for each system giving at least the 0–3 200 cm^{-1} frequency range was performed. It corresponds to the first $F = 10$ eigenvalues for H_2O, the first $F = 20$ for

H_2CO, the first $F = 30$ for CH_2NH and the first $F = 300$ for CH_3CN. The entire results of these reference calculations can be found in the supplementary material.

The initial A-VCI basis was built from the first $F+1$ functions of $\Pi_{\mathbf{d}}$ sorted by ascending energies. Unlike Refs.18 and 19 where the harmonic energies are considered for classification, the basis elements are ranked according to the diagonal of the operator (1)

$$\langle \Phi_{\mathbf{n}}(q)|\mathscr{H}\Phi_{\mathbf{n}}(q)\rangle = \sum_{i=1}^{D} \omega_i \left(\frac{1}{2}+n_i\right) + \langle \Phi_{\mathbf{n}}(q)|\mathscr{V}\Phi_{\mathbf{n}}(q)\rangle + \sum_{\alpha=x,y,z} B_\alpha \sum_{i,j>i} (\zeta_{ij}^{\alpha})^2 \left[(n_i+0.5)(n_j+0.5)\frac{\omega_i^2+\omega_j^2}{\omega_i\omega_j} - 0.5\right].$$

The calculations were performed on ada with 4 cores for H_2O and H_2CO, and with 16 cores for CH_2NH while for CH_3CN, 96 cores on brise were used.

1. Convergence on eigenvalues with the Coriolis correction

The effect of the Coriolis couplings can be significant on vibrational frequencies, as shown in the literature[28], in particular for fundamental states involving out-of-plane motions, and for combinations between out-of-plane bendings and stretchings. This influence is particularly well-known in the case of the fundamental frequencies of water[29], and our results available in supplementary information agree with this statement. These computations led to the results showed in Table II where a VCI(24) computation serves as a reference, with a binomial basis $\left(\sum_{i=1}^{3} n_i \leq 24\right)$ of 2925 elements.

For H_2CO, the maximum error on the frequencies including the Coriolis correction was evaluated with respect to a reference calculation realized in a VCI(21) binomial basis of 296010 elements. The first 20 eigenvalues were computed and cover the 0–3600 cm^{-1} frequency domain. As shown on Table III, we have a very good agreement with respect to the reference calculation with a basis containing less than 12000 elements. Along with the calculations done for H_2O, these results validate the approach of integrating the Coriolis correction in the Hamiltonian of the A-VCI iterations.

We introduced two larger systems (CH_2NH and CH_3CN) to the previous ones to study the behavior of the A-VCI algorithm with Coriolis couplings. We calculated for the 4 systems a fixed number of eigenvalues ($F = 30$) at different precisions, in the approximation space induced by

TABLE II. Convergence of A-VCI frequencies for H_2O with the Coriolis contribution (ε is the convergence criterion). The parameters are $F = 10$, $E_{max} = 24\,000\,\mathrm{cm}^{-1}$ and $p = 2$. The maximum absolute error Err is evaluated with respect to the VCI(24) calculation.

	Err (cm^{-1})	Final basis size	Number of iterations	Time (s)
VCI(24)	0	2925	0	0.42
A-VCI($\varepsilon = 1.0 \times 10^{-3}$)	0.35	600	13	0.24
A-VCI($\varepsilon = 2.5 \times 10^{-3}$)	0.37	444	12	0.19
A-VCI($\varepsilon = 5.0 \times 10^{-3}$)	0.46	294	10	0.11
A-VCI($\varepsilon = 7.5 \times 10^{-3}$)	0.68	236	9	0.09

TABLE III. Convergence of A-VCI frequencies for H_2CO with the Coriolis contribution (ε is the convergence criterion). The parameters are $F = 20$, $E_{max} = 20\,000\,\mathrm{cm}^{-1}$ and $p = 2$. The maximum absolute error Err is evaluated with respect to the VCI(21) calculation.

	Err (cm^{-1})	Final basis size	Number of iteration	Time (s)
VCI(21)	0	296010	0	163.5
A-VCI($\varepsilon = 1.0 \times 10^{-3}$)	0.04	11021	10	9.0
A-VCI($\varepsilon = 2.5 \times 10^{-3}$)	0.05	4631	8	2.7
A-VCI($\varepsilon = 5.0 \times 10^{-3}$)	0.16	2777	8	1.7
A-VCI($\varepsilon = 7.5 \times 10^{-3}$)	0.36	1703	7	0.9

$E_{max} = 20\,000\,\mathrm{cm}^{-1}$. We used the CCW(3) enlargement strategy and the following thresholds: 7.5×10^{-3}, 5.0×10^{-3}, 2.5×10^{-3} and 1.0×10^{-3}. These test cases are compared to a reference calculation done with $\varepsilon = 5.0 \times 10^{-4}$ in the same conditions.

Figure 1 shows how the maximal absolute error on the eigenvalues decreases with respect to the ε threshold. These convergence properties of the A-VCI method have already been observed for the computation of a PES-based Hamiltonian operator[18,19]. They are also valid when adding the Coriolis correction to the operator. In addition, the final basis size decrease becomes increasingly important as the molecular system grows in size. Even with the Coriolis couplings, the adaptive approach of A-VCI remains an appropriate way to study large systems.

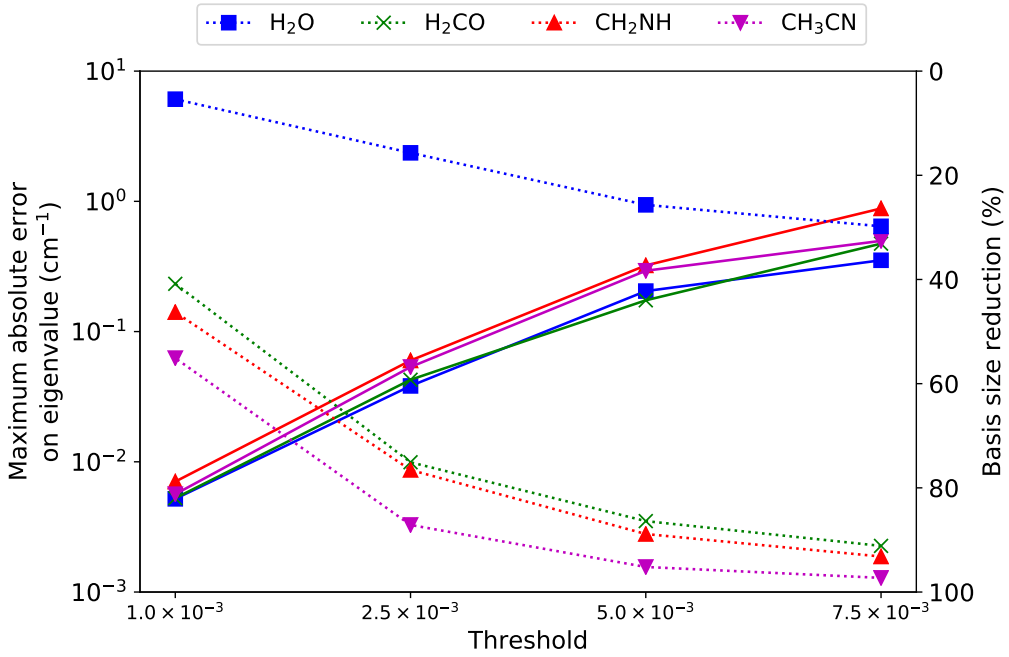

FIG. 1. Maximum absolute errors ($\log_{10}$ scale) on frequencies (———) and final basis size reduction percentage (·····) for H_2O, H_2CO, CH_2NH and CH_3CN. Calculations are performed for $F = 30$, $p = 3$ and the active space defined by $E_{max} = 20\,000\,\mathrm{cm}^{-1}$. Errors and reductions are computed with respect to the calculation done for $\varepsilon = 5.0 \times 10^{-4}$.

2. *Calculating intensities with the Coriolis correction*

In this subsection, we use the calculations previously performed for $F = 30$, $p = 3$ and $E_{max} = 20\,000\,\mathrm{cm}^{-1}$ to study the intensities obtained at the convergence of the A-VCI algorithm. As shown in Fig .2, the maximum absolute error on the intensities decreases with the value of the threshold, even though the decrease rate depends on the system. This indicates that the quality of the eigenvectors is also ensured by the A-VCI algorithm when the Coriolis contribution is taken into account.

In these calculations, the CPU time needed to compute the intensities represents less than 1% of the total A-VCI algorithm time. However, we see on Fig. 3 how the complexity of the intensity algorithm increases with the system dimension. This means that for very large systems, the computational time might become significant, depending primarily on the number of terms in the dipole moment surface.

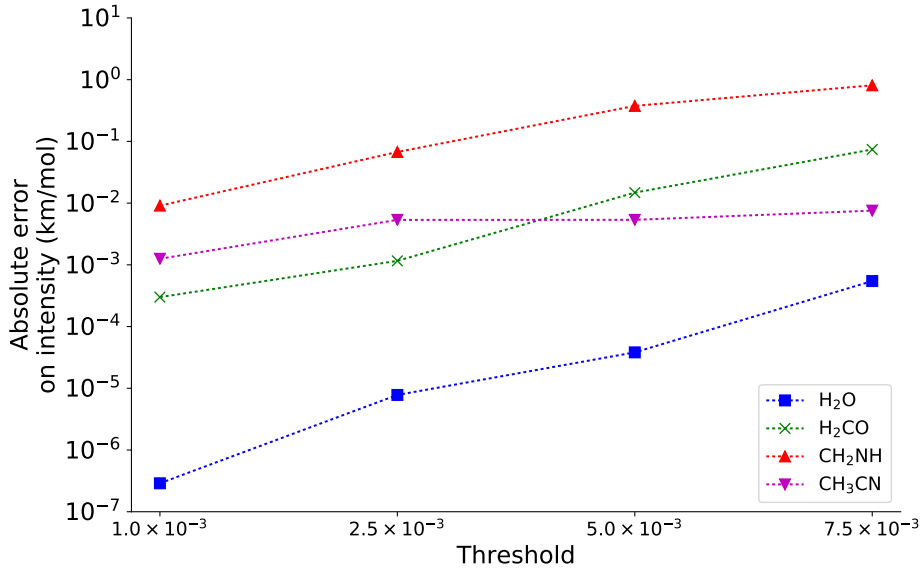

FIG. 2. Maximum absolute errors ($\log_{10}$ scale) on intensities with respect to ε. Calculations are done for H_2O, H_2CO, CH_2NH and CH_3CN with the following parameters: $F = 30$, $p = 3$ and $E_{max} = 20000\,\mathrm{cm}^{-1}$. Errors are computed with respect to the calculation done for $\varepsilon = 5.0 \times 10^{-4}$.

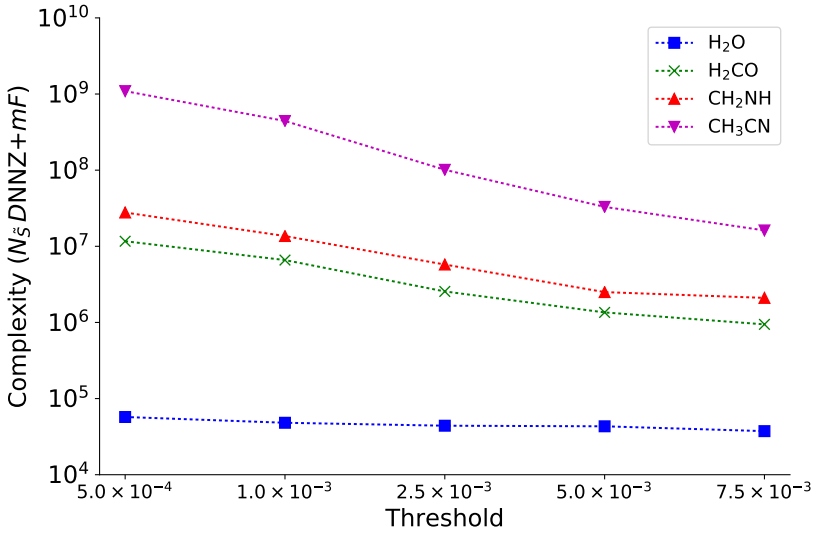

FIG. 3. Complexity of the intensity algorithm calculation with respect to ε. Calculations are done for H_2O, H_2CO, CH_2NH and CH_3CN with the following parameters: $F = 30$, $p = 3$ and $E_{max} = 20000\,\mathrm{cm}^{-1}$.

We compare in Fig. 4 the results obtained when using the PES operator with or without the

Coriolis operator for CH_3CN and $\varepsilon = 5.0 \times 10^{-3}$. We only represent intensities greater than 0.5 km/mol. For H_2O, H_2CO and CH_2NH, the intensity diagrams similar to Fig. 4 are provided in the supplementary material. For these systems, the Coriolis contribution has no tangible effect on the intensities. The only difference is the result of the frequency shift.

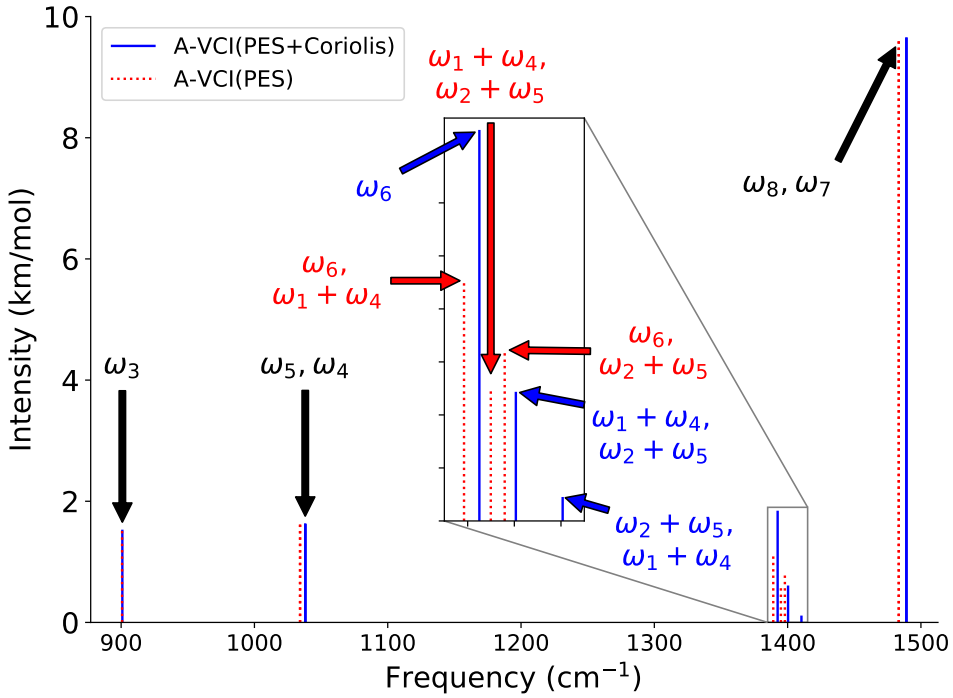

FIG. 4. CH_3CN intensities with and without the Coriolis contributions. Calculations are done for $F = 30$, $p = 3$, $\varepsilon = 5.0 \times 10^{-3}$ and the active space defined by $E_{max} = 20000\,\text{cm}^{-1}$. Only the intensities larger than 0.1 km/mol are represented. Intensity bands are annotated by their main attribution.

Regarding the role of the Coriolis contribution on CH_3CN, some numerical values are affected only to a very small extend (Table IV). The Coriolis effect on the fundamental frequencies ω_4, ω_5, ω_7 and ω_8 is lower than $6\,\text{cm}^{-1}$, and the intensity variation does not exceed 0.05 km/mol. The ω_3 fundamental frequency is not affected by the Coriolis operator. In this case, the frequencies seem to be more affected by the Coriolis couplings than their related intensities.

However, the influence of the Coriolis effect is more important for the ω_6 fundamental frequency (cf. Table V). Without the Coriolis contribution, the $\omega_6(A_1)$ state is accidentally near the combination of the degenerate states $\omega_1, \omega_2(E)$ and $\omega_4, \omega_5(E)$. According to point group symmetry, $E \times E$ contains contributions from A_1 ($E \times E = A_1 + A_2 + E$). Since their energies are very

TABLE IV. CH_3CN frequencies and intensities slightly affected by the Coriolis contribution. Calculations are done for $F = 30$, $p = 3$, $\varepsilon = 5.0 \times 10^{-3}$ and the active space defined by $E_{max} = 20000\,\mathrm{cm}^{-1}$. The main atributions are provided in the last two columns with the corresponding coefficient in the eigenvector.

Frequency	**(cm^{-1})**	**Intensity**	**(km/mol)**	**Assignment**	
PES	PES + Coriolis	PES	PES + Coriolis	PES	PES + Coriolis
900.71	900.92	1.53	1.53	$\omega_3(0.95)$	$\omega_3(0.95)$
1034.24	1038.12	1.67	1.64	$\omega_5(0.97)$	$\omega_5(0.97)$
1034.22	1038.33	1.65	1.62	$\omega_4(0.97)$	$\omega_4(0.97)$
1483.37	1489.01	9.61	9.66	$\omega_8(0.97)$	$\omega_8(0.97)$
1483.37	1489.24	9.61	9.66	$\omega_7(0.97)$	$\omega_7(0.97)$

TABLE V. CH_3CN quasi-degenerate frequencies and intensities around $1400\,\mathrm{cm}^{-1}$ with and without the Coriolis contributions. Calculations are done for $F = 30$, $p = 3$, $\varepsilon = 5.0 \times 10^{-3}$ and the active space defined by $E_{max} = 20000\,\mathrm{cm}^{-1}$. The main atributions are provided with the corresponding coefficient in the eigenvector.

PES				**PES + Coriolis**			
Frequency (cm^{-1})	Intensity (km/mol)	State Assignment	State Symmetry	Frequency (cm^{-1})	Intensity (km/mol)	State Assignment	State Symmetry
1389.22	1.133	$\omega_6(0.73)$, $\omega_1+\omega_4(-0.45)$	A_1	1392.51	1.844	$\omega_6(0.94)$, $\omega_6+\omega_{10}(0.19)$	A_1
1394.95	0.628	$\omega_1+\omega_4(0.68)$, $\omega_2+\omega_5(-0.45)$	E	1400.33	0.609	$\omega_1+\omega_4(0.68)$, $\omega_2+\omega_5(-0.68)$	A_1
1394.81	0.014	$\omega_2+\omega_4(0.73)$, $\omega_1+\omega_5(0.63)$	E	1402.71	0.013	$\omega_1+\omega_5(0.68)$, $\omega_2+\omega_4(0.68)$	A_2
1395.03	0.001	$\omega_1+\omega_5(-0.73)$, $\omega_2+\omega_4(0.64)$	A_2	1410.29	0.001	$\omega_2+\omega_4(-0.68)$, $\omega_1+\omega_5(0.68)$	E
1397.94	0.806	$\omega_6(0.63)$, $\omega_2+\omega_5(0.52)$	A_1	1410.36	0.113	$\omega_2+\omega_5(0.66)$, $\omega_1+\omega_4(0.66)$	E

close, ω_6 has almost the same weight as $(\omega_1,\omega_2)+(\omega_4,\omega_5)$ for two distinct eigenpairs. The Cori-

olis contribution removes the inadvertent degeneracy. We can now identify a state with a main contribution from ω_6, and four combinations developed on ω_1, ω_2(E) and ω_4, ω_5(E). In this case, the Coriolis part of the operator is essential to identify the ω_6 fundamental.

B. Screening by intensity

We will now validate the selection of nodes by considering only eigenvectors with an intensity on two well studied molecules (acetonitrile and ethylene oxide), and then use it on a more challenging molecule with 10 atoms (pyrazine). In these calculations, the threshold ε_I is decreased from 1 km/mol to 0.01 km/mol in order to study the behavior of the selection technique, its effect on the final basis size and on the computational time.

1. Acetonitrile molecule, CH_3CN

In this section we analyze the performance of the SI and GSI strategies on the CH_3CN molecule compared to the classical A-VCI approach with no intensity selection. The reference calculations for different thresholds ε (parameters: $F = 300$, $p = 4$ and $E_{max} = 20\,000\,\mathrm{cm}^{-1}$) are fully described in supplementary material.

Since the number of frequencies with an intensity greater than ε_I is much smaller than the total number of eigenvalues F, this selection procedure leads to a significant CPU time reduction of at least 50%, as shown in the supplementary material. For a given threshold ε, the errors on selected frequencies are below $1\,\mathrm{cm}^{-1}$ as presented in Fig. 5. We see that for a required A-VCI precision, the errors on the frequencies have the same order of magnitude regardless of the intensity selection threshold ε_I.

Fig. 6 compare the two strategies SI and GSI in terms of iterations to reach the convergence. The number of iterations in the GSI selection is larger than when we have no selection. Since the algorithm only seeks to converge a small number of eigenpairs, the number of iterations required to reach the convergence of the first eigenpair $(E_0, \mathbf{X}_0)$ increases. Therefore, as long as the first eigenvector is poorly estimated, so are all the computed intensities. Moreover, the convergence iteration number for the first eigenpair strongly depends on the intensity threshold. The higher this threshold, the more slowly it converges.

Conversely, the number of iterations decreases using the SI strategy. This strategy converges

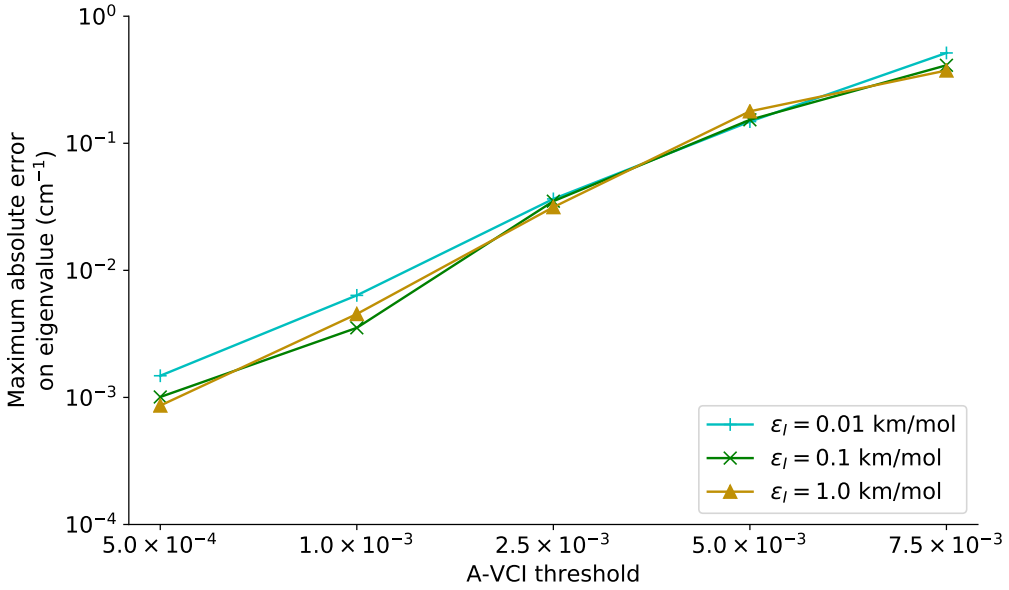

FIG. 5. Maximum absolute errors ($\log_{10}$ scale) on CH_3CN frequencies when using the SI strategy for different values of ε and ε_I. Calculations are done for $F = 300$, $p = 4$ and $E_{max} = 20\,000\,\text{cm}^{-1}$. Errors are computed with respect to the calculations done without selection by intensity.

the first eigenpair $(E_0, \mathbf{X}_0)$ faster than GSI, at the speed of the strategy with no intensity selection. This means that when the selection is activated, it is based on a good approximation of the first eigenvector $\mathbf{X}_0$, and, as a result, of the intensities. Another advantage of SI over GSI is that the convergence iteration number of the first eigenpair does not depend on ε_I for a required A-VCI precision. The key point of the SI strategy is to ensure the convergence of the first eigenpair $(E_0, \mathbf{X}_0)$ prior to the selection process. Therefore the quality of all eigenvectors $\mathbf{X}_k$ (and thus their intensities I_k) is improved at each iteration until the first eigenvector has converged. This explains why we obtain a faster convergence and a better accuracy on both frequencies and intensities. On the other hand, the GSI procedure calculates intensities with a non-converged $\mathbf{X}_0$ that can lead to incorrect intensities and therefore wrong residues are involved in the enrichment of our basis.

The GSI approach is very penalizing and does not efficiently reduce the A-VCI basis compared to the SI strategy, as shown in Figure 7. The best basis reduction obtained with the GSI strategy is around 32% for a reference basis of 57 386 elements, whereas with the SI strategy reductions of more than 60% are obtained for a basis of 1 652 790 elements and for any value of ε_I. This confirms the importance of converging $(E_0, \mathbf{X}_0)$ in the first place for intensity selection methods,

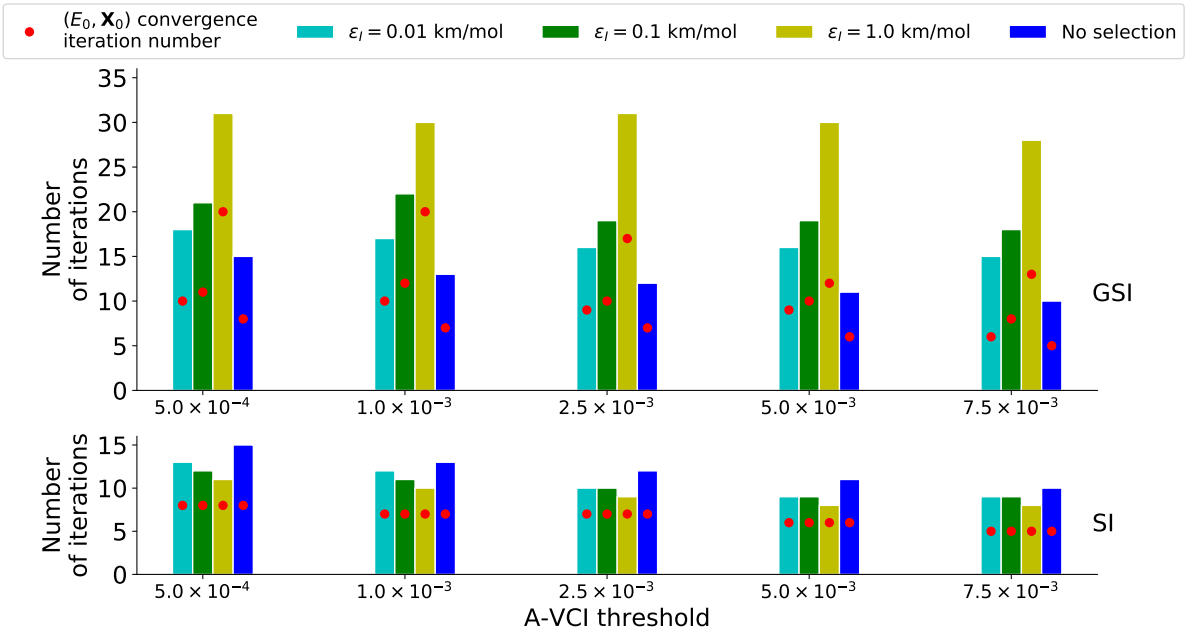

FIG. 6. Number of A-VCI iterations for CH_3CN when using GSI, SI ($\varepsilon_I = 0.01$, 0.1 and 1.0 km/mol) and the classical A-VCI strategy with no selection by intensity. Iteration numbers are plotted as functions of the A-VCI threshold. Iteration numbers corresponding to the convergence of the first eigenpair $(E_0, \mathbf{X}_0)$ are also reported.

as done in the SI strategy.

In Figure 8, we plotted the numerical spectra for different values of ε_I with $F = 300$, $p = 4$, $E_{max} = 20000\,\text{cm}^{-1}$ and $\varepsilon = 5 \times 10^{-3}$. The absorbance corresponds to the normalized $\log_{10}$ of the computed intensities, which is set to be the maximum of a Lorentzian profile with a half-width at half-maximum of 25 cm^{-1} (arbitrary value set using experimental data). The horizontal lines reflect the respective values of ε_I in this normalized logarithmic scale.

For $\varepsilon_I = 1.0$, Figure 8 shows that a few frequencies were not taken into account by the selection algorithm. This absence represents a substantial change to the final profile. However, for $\varepsilon_I = 0.1$, the selection has not significantly affected the profile compared to the one obtained without the selection. This leads us to conclude that the $\varepsilon_I = 0.1$ threshold is a good trade-off between efficiency and precision.

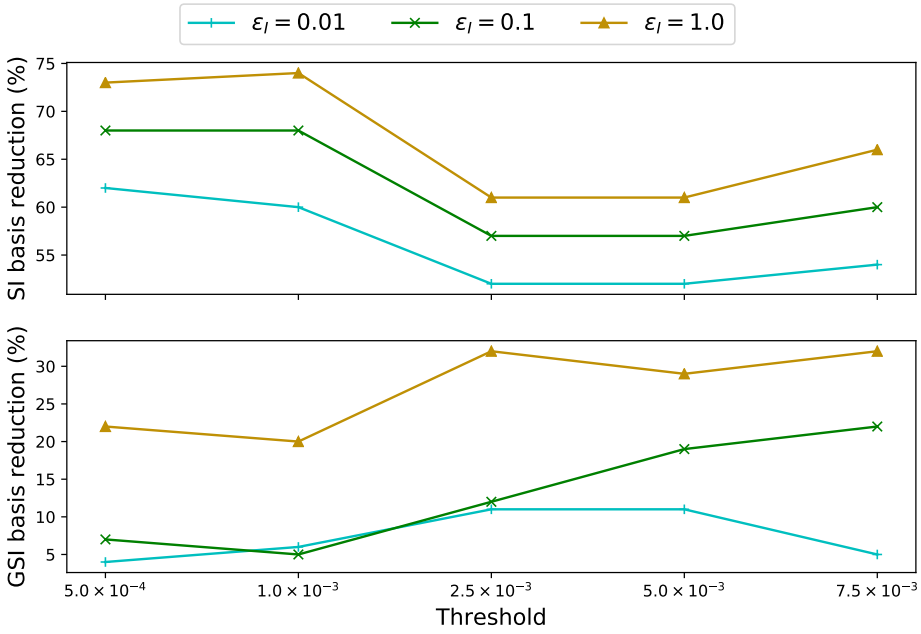

FIG. 7. Basis reductions (percentage) obtained by the SI and the GSI strategy on CH_3CN with respect to the A-VCI approach without selection by intensity.

2. *Ethylene oxide, C_2H_4O*

Due to the large dimensions (≥ 15) of the next calculations, we changed two options in A-VCI. First, we used the 64 bits interface of LAPACK/ARPACK for integers and second, to limit the memory footprint we did not store in memory the coefficients of the H_R matrix. This means we computed the scaled residues on the fly.

To reach the convergence threshold with the Coriolis operator is challenging compared to the PES operator. It seems that ARPACK has some difficulty to converge when the basis set is around 4 millions. This is not the case for the PES operator. We used the PES described in the work of Bégué *et al.*[30]. The 15 harmonic terms of this PES are obtained at the CCSD(T)/cc-pVTZ level. The anharmonic part of this potential is composed of 180 cubic terms and 445 quartic terms, computed at the B3LYP/6-31+G** level. The Coriolis coefficients, and the dipole moments surface terms are also computed at the B3LYP/6-31+G** level.

Let us first compare in Table VI the efficiency of the SI strategy for an accuracy of 5×10^{-3} on the 200 first eigenpairs. The last computed frequency is at 3219.62 cm^{-1}. At the 5^{th} iteration, the first eigenvector $\mathbf{X_0}$ has reached convergence, and the SI strategy starts at the next iteration. As

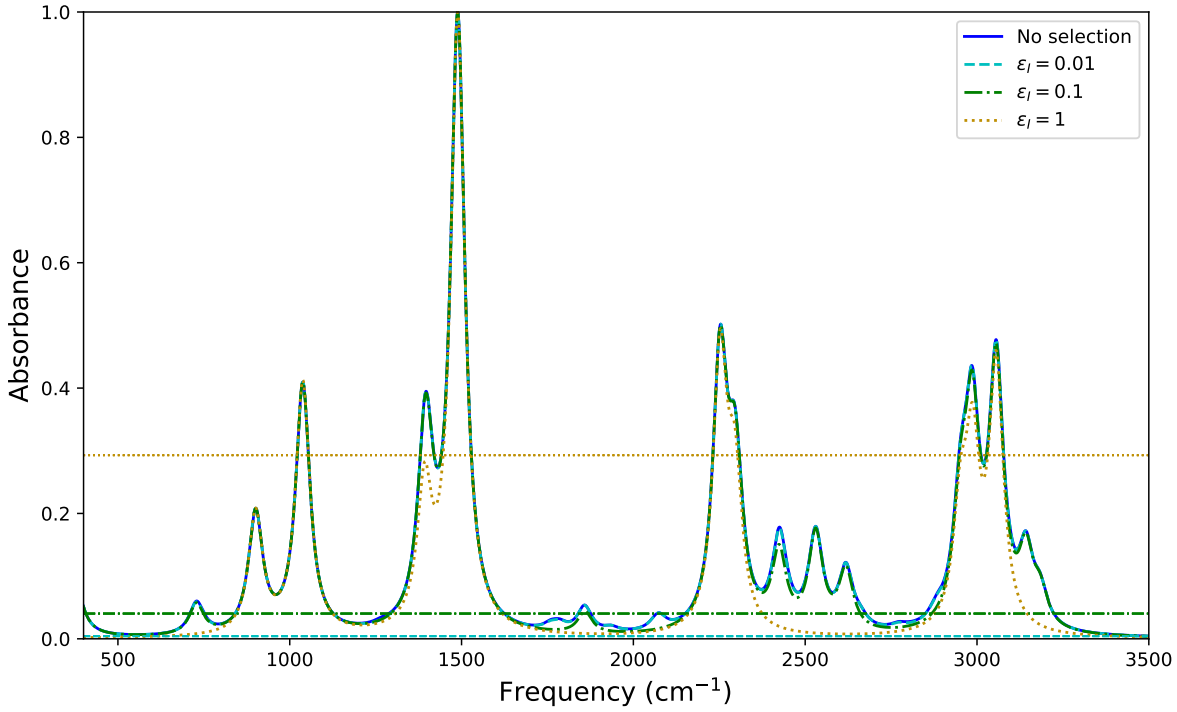

FIG. 8. Influence of the ε_I parameter using the SI strategy for CH_3CN. The log_{10} of the intensities has been normalized and bandshapes have been created using a Lorentzian profile with a half-width at half-maximum of 25 cm^{-1}. Horizontal lines represent the corresponding scaled thresholds ε_I.

the number of frequencies is reduced by the selection algorithm, we obtain smaller basis sizes and execution times.

TABLE VI. Intensity screening for C_2H_4O. The parameters are $\varepsilon = 5 \times 10^{-3}$, $F = 200$, $p = 4$ and $E_{max} = 20000\,\text{cm}^{-1}$.

ε_I	Selected intensities	Final basis size	Total time (s)	Number of iterations
0.0	200	2011865	20207	11
1.0	25	874559	6050	10
0.1	70	1191764	10213	11
0.01	112	1385089	15070	11

It is important to note that the number of frequencies with an intensity above a given threshold ε_I (*i.e.* selected by the algorithm) is slighly different from the number of frequencies with an

intensity above the same threshold when no selection is made. The intensities corresponding to these missing frequencies have the same order of magnitude as the threshold ε_I. When the selection begins (second step of the algorithm), the corresponding eigenvectors are not well represented in the basis and lead to an inaccurate approximation of their intensities. Since the value of the resulting intensities are near ε_I, but slightly lower, the corresponding states are not selected by the algorithm. For these particular states, the added information thereafter does not entail the convergence of their eigenvectors. This is fixed by lowering the value of the A-VCI threshold ε so that the algorithm gives a better approximation of these eigenvectors when the convergence of the first eigenpair is achieved.

Figure 9 presents the numerical spectra derived from these computations. As previously mentioned, they are drawn using a Lorentzian profile with a half-width at half-maximum of 25 cm^{-1}. These lineshapes are centered on the frequencies with an intensity above ε_I, and the peak heights correspond to a normalization of the $\log_{10}$ of the corresponding intensities. The horizontal lines also represent the different values of the normalized intensity threshold ε_I. As in the case of CH_3CN, $\varepsilon_I = 0.1$ is a good compromise. The final basis size and the total time are almost halved without any significant impact on the spectrum profile.

Figure 10 presents the scaled residues for all eigenvalues with an accuracy of 0.0035. The red dots correspond to the 71 eigenvalues with an intensity higher than $\varepsilon_I = 0.1$, while the blue triangle is for those with an intensity lower than ε_I. The run takes 8 h 47 min to converge in 12 iterations. At the convergence, only 101 eigenvalues have a residue greater than $\varepsilon = 0.0035$. This means that 28 eigenvalues with an intensity lesser than 0.1 km/mol have also converged. The final basis size is 2 785 552.

3. Pyrazine, $C_4H_4N_2$

To push the limits of the method, we consider a larger molecular system. The harmonic coefficients of the PES generated were calculated at the CCSD(T)/cc-pVTZ level. The anharmonic part of the potential (composed of 367 cubic terms and 666 quartic terms), the Coriolis coefficients, and the dipole moments surface terms at were all computed at the B3LYP/6-31+G** level. We consider an intensity criterion ε_I set to 0.1 which appears as a good compromise between computational time, accuracy on the eigenvalues concerned, and relevant experimental data.

To compute 140 eigenvalues, with an accuracy of $\varepsilon = 5.0 \times 10^{-3}$, the algorithm takes 15 h 13

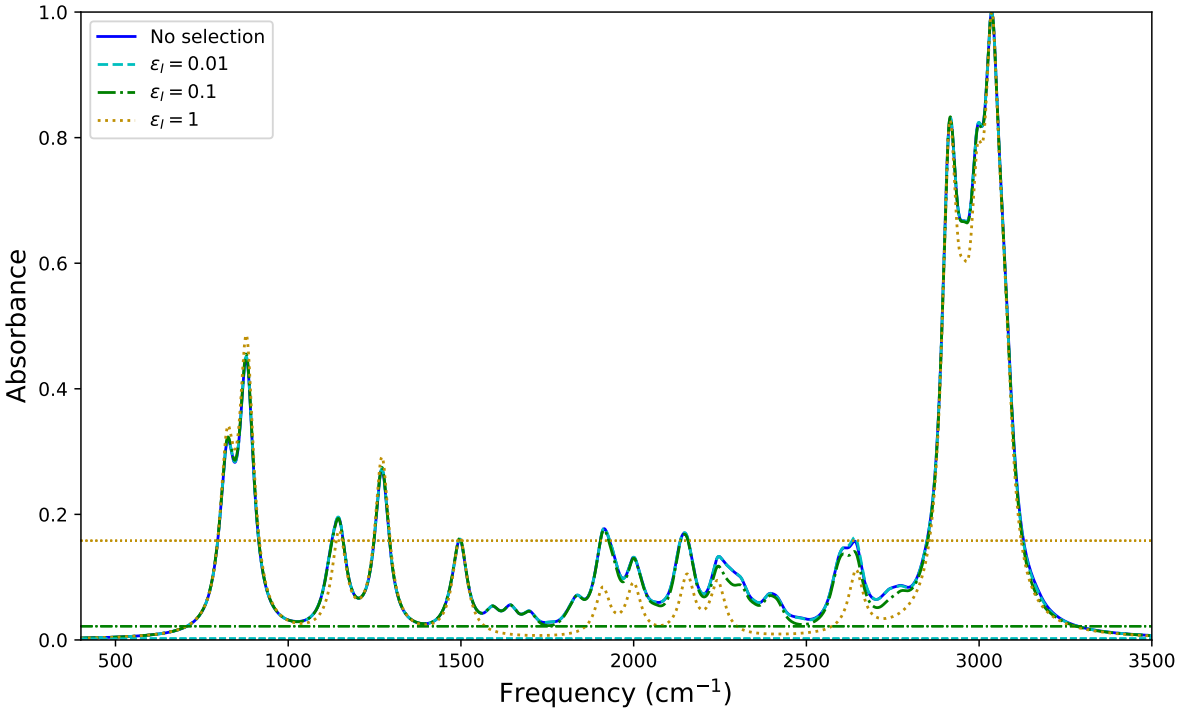

FIG. 9. Influence of the ε_I parameter using the SI strategy for C_2H_4O. The log_{10} of the intensities has been normalized and bandshapes have been created using a Lorentzian profile with a half-width at half-maximum of 25 cm^{-1}. Horizontal lines represent the corresponding scaled thresholds ε_I.

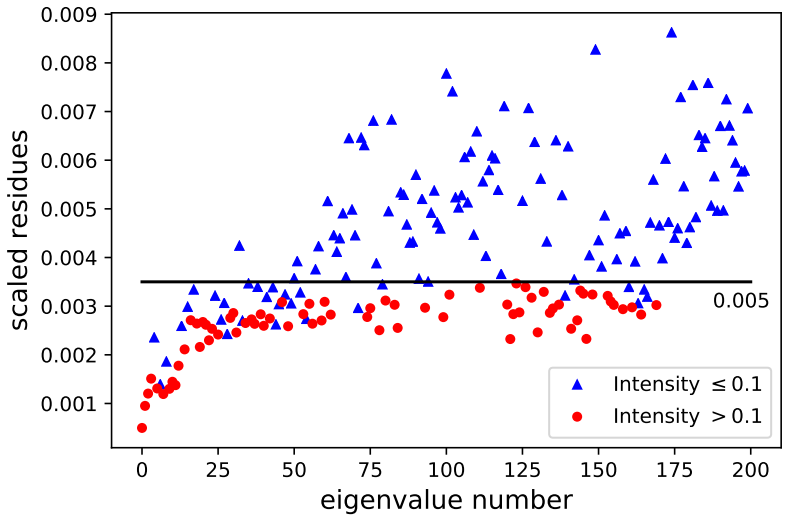

FIG. 10. Scaled residues using the SI strategy ($\varepsilon_I = 0.1$ km/mol). Calculations are done for C_2H_4O with the following parameters: $F = 200$, $p = 4$, $\varepsilon = 3.5 \times 10^{-3}$ and $E_{max} = 20000\,cm^{-1}$.

min to converge in 11 iterations, with a final basis size of 2622918. As seen in the section above, this method requires the convergence of the first eigenpair before entering the intensity selection process. The first eigenpair (E_0, $\mathbf{X_0}$) converges in 8 iterations. Only 21 eigenvalues out of 140 have an intensity above $\varepsilon_I = 0.1$ (see Table VII).

TABLE VII: Coriolis calculation for the first $F = 140$ eigenvalues of $C_4H_4N_2$ with the SI method ($\varepsilon_I = 0.1$ km/mol). The parameters are $p = 4$, $E_{max} = 15000\,\mathrm{cm}^{-1}$ and $\varepsilon = 5.0 \times 10^{-3}$. Only the eigenpairs selected by the SI method are provided. Their position number in the full spectrum is given in parentheses, as well as the eigenvector coefficients used to make the attributions.

Number	**Frequency (cm^{-1})**	**Intensity (km/mol)**	**Assignment**
0(0)	16558.14	–	ZPE
1(2)	414.24	26.41	$\omega_2(0.97), \omega_2+\omega_{10}(0.16)$
2(8)	770.20	31.23	$\omega_6(-0.96), \omega_6+\omega_{10}(-0.17)$
3(17)	1014.22	31.17	$\omega_{11}(-0.96), \omega_{10}+\omega_{11}(-0.19)$
4(18)	1037.22	0.19	$\omega_1+\omega_4(0.96), \omega_1+\omega_4+\omega_{10}(0.16)$
5(19)	1056.04	8.86	$\omega_{12}(0.9), \omega_1+\omega_5(-0.33)$
6(21)	1090.47	2.04	$\omega_1+\omega_5(-0.9), \omega_{12}(-0.33)$
7(24)	1126.63	6.56	$\omega_{13}(0.9), \omega_2+\omega_5(0.33)$
8(25)	1143.23	4.25	$\omega_{14}(0.93), \omega_{10}+\omega_{14}(0.25)$
9(27)	1173.27	0.52	$\omega_2+\omega_5(0.91), \omega_{13}(-0.33)$
10(34)	1268.79	0.12	$\omega_1+\omega_8(0.92), \omega_2+\omega_7(-0.16)$
11(37)	1320.17	0.80	$\omega_2+\omega_7(-0.93), \omega_2+\omega_7+\omega_{10}(-0.15)$
12(44)	1358.81	0.21	$\omega_2+\omega_8(0.95), \omega_2+\omega_8+\omega_{10}(0.16)$
13(50)	1404.36	30.09	$\omega_{17}(-0.94), \omega_{10}+\omega_{17}(-0.2)$
14(62)	1474.30	0.20	$\omega_{18}(0.95), \omega_{10}+\omega_{18}(0.2)$
15(88)	1602.74	0.15	$\omega_3+\omega_{11}(0.95), \omega_3+\omega_{10}+\omega_{11}(0.19)$
16(93)	1643.31	0.39	$\omega_3+\omega_{12}(0.71), \omega_6+\omega_7(-0.54)$
17(96)	1650.91	1.44	$\omega_6+\omega_7(0.67), \omega_3+\omega_{12}(0.51)$

Continued on next page

TABLE VII – *Continued from previous page*

Number	Frequency (cm^{-1})	Intensity (km/mol)	Assignment
18(108)	1687.32	1.84	$\omega_6+\omega_8(-0.86), \omega_7+\omega_9(-0.26)$
19(118)	1717.84	0.26	$\omega_5+\omega_9(-0.89), \omega_6+\omega_7(0.21)$
20(119)	1731.21	0.16	$\omega_3+\omega_{14}(0.92), \omega_3+\omega_{10}+\omega_{14}(0.25)$
21(137)	1777.42	0.13	$\omega_6+\omega_{10}(0.91), \omega_6+2\omega_{10}(0.26)$

Figure 11 presents the scaled residues for all eigenvalues with an accuracy of 0.005. Although the convergence of the method only concerns these 21 eigenvalues (red dots), 13 additional eigenvalues located at the bottom of the spectrum have also converged.

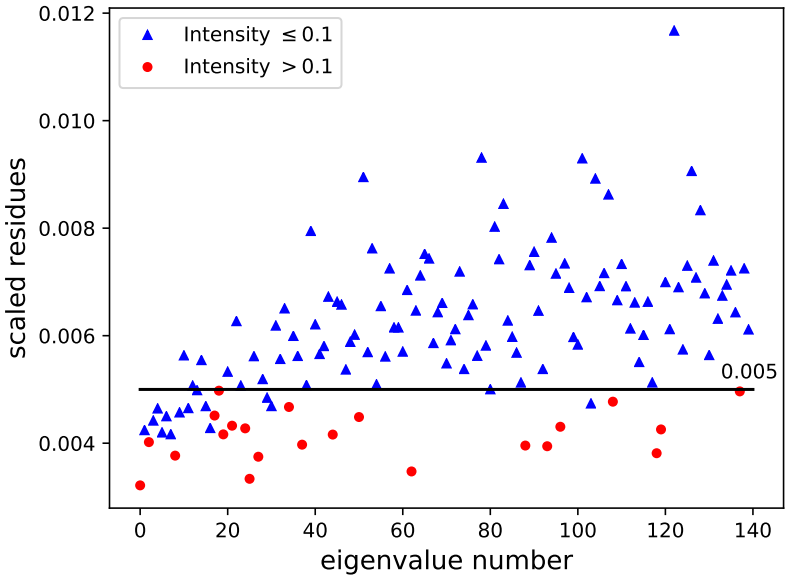

FIG. 11. Scaled residues using the SI strategy ($\varepsilon_I = 0.1\,$km/mol). Calculations are done for $C_4H_4N_2$ with the following parameters: $F = 140$, $p = 4$, $\varepsilon = 5\times10^{-3}$ and $E_{max} = 15\,000\,cm^{-1}$.

Figure 12 shows how these results are consistent with experimental data[31,32]. The A-VCI results are largely in agreement with these data, except for a few specific bands : ω_6 ($770.20\,cm^{-1}$), $\omega_6+\omega_7$ ($1650.91\,cm^{-1}$) and $\omega_6+\omega_8$ ($1687.32\,cm^{-1}$). All these discrepancies involve the normal mode ω_6, and they can be found using PES generated with different levels of theory (CCSD(T)

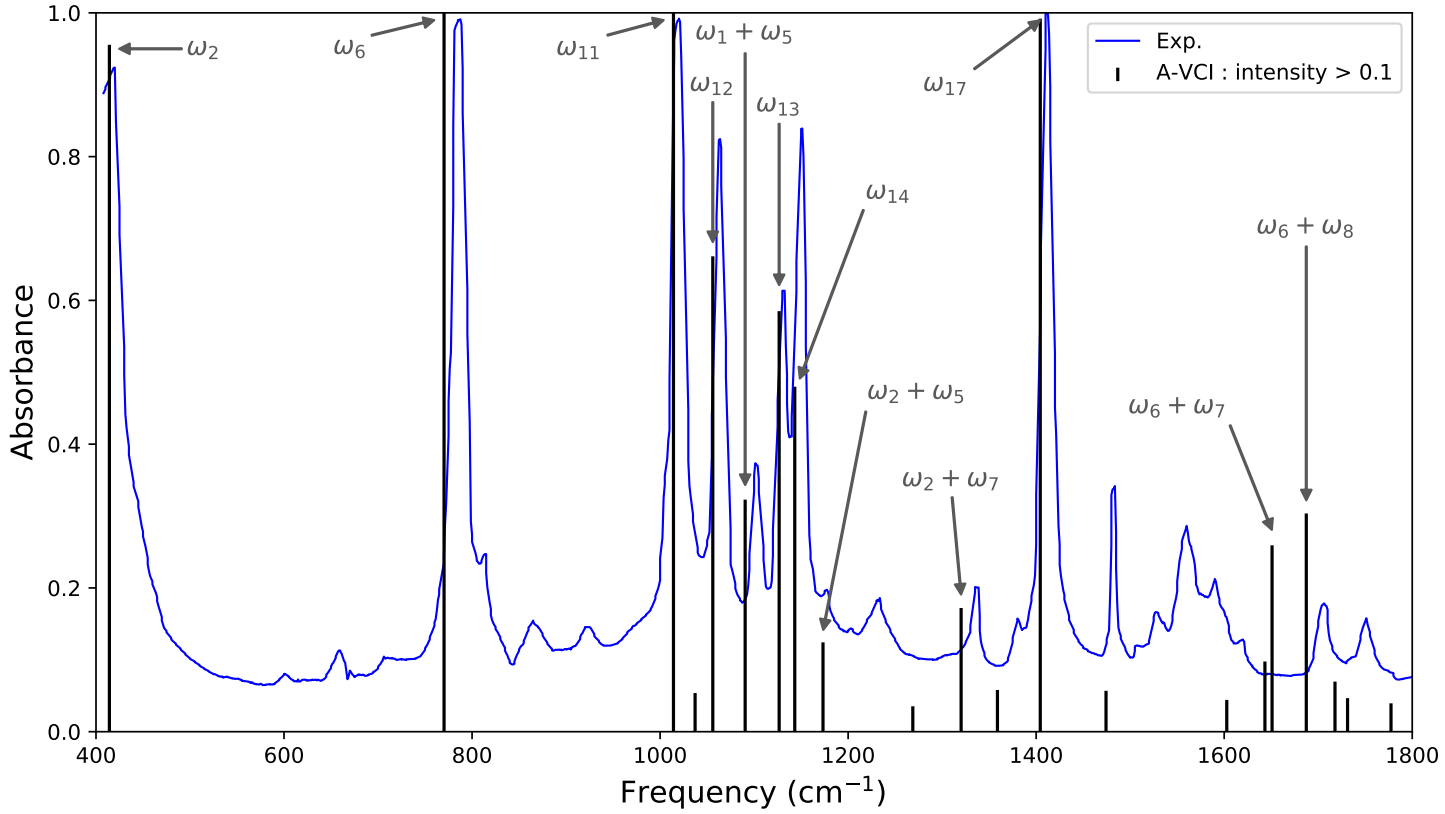

FIG. 12. Comparison between experimental data for $C_4H_4N_2$ and an A-VCI computation using SI ($\varepsilon_I =$ 0.1 km/mol). The parameters are $F = 140$, $p = 4$, $E_{max} = 15\,000\,\text{cm}^{-1}$ and $\varepsilon = 5.0 \times 10^{-3}$.

and B3LYP). Therefore, these differences between computed and experimental bands must be due to incompleteness in the underlying Taylor expansion of the PES. Moreover, the quality of this quartic force field representation prevents us from using a lower ε and/or searching for more eigenvalues. It is likely that a full (or four mode representation) sextic force field is necessary to address this problem.

V. CONCLUSION

One of the main challenges for the computation of anharmonic spectra resides in the choice of a suitable basis to discretize a given vibrational Hamiltonian. We have shown previously that the A-VCI method provides the beginning of a solution by carefully selecting basis functions in a hierarchical way, allowing some control over the accuracy of the computed eigenvalues.

In this report, we have presented how to take into account the Coriolis terms in the Hamiltonian operator for the A-VCI method leading to a more accurate vibrational modeling in order to compare our results with experimental data. The introduction of these terms has two consequences. First of all, the matrix representing the Hamiltonian operator is less sparse than in the case of the PES alone and therefore leads to high memory pressure and much higher computation times. Secondly, when more accuracy is needed, the ARPACK solver has difficulty achieving convergence. We have also introduced the calculation of the IR intensities in order to better characterize the computed frequencies. Since the frequencies detected experimentally are only those with an intensity, we took advantage of this fact by selecting only the vibrational states interacting with IR active frequencies at each iteration. This strategy reduces the size of the discretization space, and thus the memory footprint, as well as speeding up the calculation time.

These developments have been validated on systems with 3 to 7 atoms, without any loss in the accuracy of the computed eigenvalues. As a conclusion we show that the approach based on intensity selection made it possible to quickly calculate 140 eigenvalues of a molecule with 10 atoms. However, for large molecules (more than 8 atoms) obtaining a calculation relevant for experimental interpretation would require improving the quality of the Hamiltonian operator by having better surfaces (PES, Coriolis and dipole moment), with better convexity, to successfully reach high energy frequencies. In addition, algorithmic improvements are also needed to improve the efficiency of iterative solvers, reduce the memory footprint of graph construction either by parallelizing through a distributed memory approach or by using another formalism such as tensors.

SUPPLEMENTARY MATERIAL

The supplementary material contains the PES, Coriolis and dipole moment coefficients used in this work for H_2O, H_2CO, CH_2NH, CH_3CN, C_2H_4O and $C_4H_4N_2$. Figures comparing intensities of H_2O, H_2CO and CH_2NH when using the PES operator with or without the Coriolis operator are provided. We also present in the supplementary material reference calculations for the molecular systems considered in this work. All frequencies and intensities of the reference calculations are computed using the Coriolis correction. Finally, we provide the full attributions of the Selection by Intensity calculation done for $C_4H_4N_2$ in this work.

ACKNOWLEDGMENTS

This work was granted access to the HPC resources of IDRIS under the allocations 2017-A0030810099 and 2018-A0050810638 made by GENCI. Some experiments presented in this document were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr/). Numerical computations were also carried out using the MCIA platform (Mésocentre de Calcul Intensif Aquitain, see www.mcia.univ-bordeaux.fr). Finally we acknowledge the Direction du Numérique of the Université de Pau et des Pays de l'Adour for the computing facilities it provided us.

REFERENCES

[1]J. K. Watson, "Simplification of the molecular vibration-rotation hamiltonian," Mol. Phys. **15**, 479–490 (1968).

[2]W. D. Allen, Y. Yamaguchi, A. G. Császár, D. Clabo, R. B. Remington, and H. F. Schaefer, "A systematic study of molecular vibrational anharmonicity and vibration-rotation interaction by self-consistent-field higher-derivative methods. linear polyatomic molecules," Chem. Phys. **145**, 427 – 466 (1990).

[3]P. Cassam-Chenaï and J. Liévin, "Alternative perturbation method for the molecular vibration-rotation problem," Int. J. Quantum Chem. **93**, 245–264 (2003).

[4]S. Banik and M. Durga Prasad, "On the spectral intensities of vibrational transitions in polyatomic molecules: role of electrical and mechanical anharmonicities," Theor. Chem. Acc. **131**, 1282 (2012).

[5]M. Neff, T. Hrenar, D. Oschetzki, and G. Rauhut, "Convergence of vibrational angular momentum terms within the watson hamiltonian," J. Chem. Phys. **134**, 064105 (2011).

[6]A. P. Charmet, P. Stoppa, N. Tasinato, S. Giorgianni, V. Barone, M. Biczysko, J. Bloino, C. Cappelli, I. Carnimeo, and C. Puzzarini, "An integrated experimental and quantum-chemical investigation on the vibrational spectra of chlorofluoromethane," J. Chem. Phys. **139**, 164302 (2013).

[7]P. E. Maslen, N. C. Handy, R. D. Amos, and D. Jayatilaka, "Higher analytic derivatives. iv. anharmonic effects in the benzene spectrum," J. Chem. Phys. **97**, 4233–4254 (1992).

[8]J. Váquez and J. F. Stanton, "Simple(r) algebraic equation for transition moments of fundamental transitions in vibrational second-order perturbation theory," Mol. Phys. **104**, 377–388 (2006).

[9]S. V. Krasnoshchekov, E. V. Isayeva, and 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**, 3691–3709 (2012).

[10]O. Bludský, K. L. Bak, P. Jørgensen, and V. Špirko, "Ab initio calculations of anharmonic vibrational transition intensities of trans-2,3-dideuteriooxirane," J. Chem. Phys. **103**, 10110–10115 (1995).

[11]A. Willetts, N. C. Handy, W. H. Green, and D. Jayatilaka, "Anharmonic corrections to vibrational transition intensities," J. Phys. Chem. **94**, 5608–5616 (1990).

[12]H. G. Kjaergaard, B. R. Henry, H. Wei, S. Lefebvre, T. Carrington, O. Sonnich Mortensen, and M. L. Sage, "Calculation of vibrational fundamental and overtone band intensities of h2o," J. Chem. Phys. **100**, 6228–6239 (1994).

[13]J. Bowman, T. Carrington, and H.-D. Meyer, "Variational quantum approaches for computing vibrational energies of polyatomic molecules," Mol. Phys. **106**, 2145–2182 (2008).

[14]S. Carter, J. M. Bowman, and N. C. Handy, "Extensions and tests of "multimode": a code to obtain accurate vibration/rotation energies of many-mode molecules," Theor. Chem. Acc. **100**, 191–198 (1998).

[15]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 Revision D.01," Gaussian Inc. Wallingford CT 2009.

[16]V. Barone, J. Bloino, C. Guido, and F. Lipparini, "A fully automated implementation of vpt2 infrared intensities," Chem. Phys. Lett. **496**, 157–161 (2010).

[17]M. Neff and G. Rauhut, "Toward large scale vibrational configuration interaction calculations," J. Chem. Phys. **131**, 124129 (2009).

[18]R. Garnier, M. Odunlami, V. Le Bris, D. Bégué, I. Baraille, and O. Coulaud, "Adaptive vibrational configuration interaction (a-vci): A posteriori error estimation to efficiently compute anharmonic ir spectra," J. Chem. Phys. **144**, 204123 (2016).

[19]M. Odunlami, V. Le Bris, D. Bégué, I. Baraille, and O. Coulaud, "A-VCI: A flexible method to efficiently compute vibrational spectra," J. Chem. Phys. **146**, 214108 (2017).

[20]P. Carbonnière, A. Dargelos, and 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**, 543–554 (2010).

[21]R. Burcl, S. Carter, and N. C. Handy, "Infrared intensities from the multimode code," Chem. Phys. Lett. **380**, 237 – 244 (2003).

[22]D. Bégué, I. Baraille, P. A. Garrain, A. Dargelos, and T. Tassaing, "Calculation of IR frequencies and intensities in electrical and mechanical anharmonicity approximations: application to small water clusters." J. Chem. Phys. **133**, 34102 (2010).

[23]R. B. Lehoucq, D. C. Sorensen, and C. Yang, "ARPACK Users ' Guide : Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods ." Communication **6**, 147 (1998).

[24]C. Pouchan and K. Zaki, "Ab initio configuration interaction determination of the overtone vibrations of methyleneimine in the region 2800-3200 cm-1," .

[25]G. Avila and T. Carrington, "Reducing the cost of using collocation to compute vibrational energy levels: Results for ch2nh," J. Chem. Phys. **147**, 064103 (2017).

[26]T. Halverson and B. Poirier, "Calculation of exact vibrational spectra for p2o and ch2nh using a phase space wavelet basis," J. Chem. Phys. **140**, 204112 (2014).

[27]G. Avila and T. Carrington Jr., "Using nonproduct quadrature grids to solve the vibrational schrödinger equation in 12d," J. Chem. Phys. **134**, 054126 (2011).

[28]P. Carbonniere and V. Barone, "Coriolis couplings in variational computations of vibrational spectra beyond the harmonic approximation: implementation and validation," Chem. Phys. Lett. **392**, 365 – 371 (2004).

[29]R. J. Coker DF and W. RO, "The infrared absorption spectrum of water," Australian Journal of Physics **35**, 623 (1982).

[30]D. Bégué, N. Gohaud, C. Pouchan, P. Cassam-Chenai, and J. Liévin, "A comparison of two methods for selecting vibrational configuration interaction spaces on a heptatomic system: Ethylene oxide," J. Chem. Phys. **127**, 164115 (2007).

[31]S. V.K., S. D.W., K. W.P., and H. H.W., NIST Standard Reference Simulation Website (National Institute of Standards and Technology, Gaithersburg, MD, 2017).

[32]S. A.L., The Coblentz Society Desk Book of Infrared Spectra (The Coblentz Society Desk Book of Infrared Spectra, Second Edition, The Coblentz Society, Kirkwood, MO, 1982).