Report progetto CINECA-INAF B 18
Modelling anharmonic spectra of Polycyclic Aromatic Hydrocarbons at high temperatures

Giacomo Mulas [*]
Shubhadip Chakraborty [†]

30 October 2018

Abstract

Anharmonicity plays a crucial role in determining the band profile and position of the interstellar Polycyclic Aromatic Hydrocarbons. Here we modeled the infrared spectrum of pyrene ($C_{16}H_{10}$) using ab initio simulations. We used our new AnharmoniCaOs code to describe the detailed structures of hot band transitions up to moderately high temperatures (600K), using Monte Carlo sampling of the states but beyond that due to high computational cost it is difficult to extend such fancy calculation up to very high temperature relevant for astrophysics. IR spectra of $C_{16}H_{10}$ beyond 600K was calculated using Density functional based tight binding molecular dynamics simulation. A gradual red shift of the band position and increasing band width is observed with increasing the temperature. The band positions at different temperatures were fitted with a linear regression and the anharmonicity factors were retrieved from the linear fits. Theoretical anharmonicity factors were compared to the recent laboratory results and previous gas phase analysis.

1 Introduction

Polycyclic Aromatic Hydrocarbons (PAHs) are a family of organic compounds composed of aromatic rings containing carbon atoms, and whose peripheral bonds are saturated by hydrogen atoms. PAHs (or closely related species) are thought to be the carriers of the so-called Aromatic Infrared Bands (AIBs) at ∼3.3, 6.7, 7.7, 8.6, 11.3 μm [15, 1]. These features are among the strongest emission features observed in the interstellar medium [26], therefore the infrared (IR) spectroscopy of PAHs is of paramount importance in astrophysics. PAHs absorb starlight in the visible and ultraviolet domain via electronic transitions

---

[*] INAF - Osservatorio Astronomico di Cagliari
[†] Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse

and re-emit most of this energy in the mid-IR in a vibrational de-excitation cascade. In such model PAHs should emit most of the flux in the AIBs from highly vibrationally excited states corresponding to a thermal temperature of the order of ∼1000 K [9].

Each AIB results from the superposition of the emission of molecules at different temperatures, each of which being the superposition of a large number of individual vibrational transitions from a statistical distribution of excited states. Due to anharmonicity, each of these states is expected to have a shifted position with respect to the corresponding 1-0 fundamental transition. Fundamental transitions are also shifted with respect to their frequencies in the harmonic approximation. Including anharmonicity in AIB models is therefore crucial if one wants to account for the observed band positions and profiles and to link them to the physics of the process of emission.

We developed and implemented a code [20] which implements the generalized second-order vibrational perturbation theory (GVPT2) [10, 17, 3, 21] to compute anharmonic vibrational spectra from any arbitrary vibrational state of a given molecule. This code can be very naturally used to obtain a Monte Carlo sampling of the vibrational states using the Wang-Landau approach, enabling us to obtain temperature dependent spectra at finite, non-zero temperatures. We tested its behaviour benchmarking its calculations of the 0 K spectrum against both experimental measurements for some small-medium PAHs and very high-level VMFCI calculations for a smaller species, namely ethylene oxyde, then with experimental spectra of two PAHs, namely Pyrene ($C_{16}H_{10}$) and Coronene ($C_{24}H_{12}$).

2 Computational method

We made use of the Van Vleck approach to perturbation theory applied to molecular vibrations [25, 23, 18, 24]. In this method, one applies a contact transformation to the Hamiltonian, represented in an expansion of infinitesimal contact transformations. Each order of the infinitesimal contact transformation, and thus each individual infinitesimal contact transformation in the expansion, is determined by the condition that non-diagonal terms of the transformed hamiltonian be negligible up to that order. Therefore, the 1st order contact transformation will leave non-diagonal terms of 2nd or higher order, the 2nd order contact transformation will leave non-diagonal terms of 3rd or higher order, and so on. The diagonal terms of the transformed Hamiltonian therefore yield the energies of the states corrected up to the order of the transformation used. One can also apply the same transformation to operators other than the Hamiltonian, such as e. g. the electric dipole moment, and thus obtain them correct to the same order. All the gory algebraic details can be found in the references given above. A simple recap can also be found in our paper [19]. Since this is, at the core, one of the formal ways to apply the Perturbation Theory, it suffers of the same problems: if there are degeneracies, exact or approximate, in the states one applies the method to, this results in divergent terms. The solution is to identify the terms, in the perturbed Hamiltonian, which couple such degenerate states, i. e. *resonances*, and remove them from the Van Vleck treatment. These terms are not removed by the contact transformation, and one is left with a block-diagonal effective Hamiltonian, which must be numerically diagonalised to obtain eigenstates end eigenvalues. Since the resulting complete blocks to diagonalise can easily become prohibitively large, one has to find a simple and effective way to truncate them without unacceptably affecting the accuracy. We implemented, to this effect, a "cost model".

1. We start from the zero order (harmonic) states describing the starting state of the molecule and the states connected to it by permitted transitions, each of them is the first member of a set of states which we subsequently call a *polyad*. Each of these states is initially assigned a "credit" of 1.

2. We run a loop over resonances. Each resonance can add states to each polyad, the ones connected to it by that resonance. To each resonance we attach a "cost", directly proportional to the energy difference between the near-degenerate states it connects and inversely proportional to the Hamiltonian term involved in the resonance. Thus, strong resonances are less expensive than weak ones. Added states are assigned the credit of the state, already in the polyad, they connect to, minus the cost. Only states with a positive credit are added to a polyad. If two polyads try to add the same additional state, they are merged in a single polyad.

3. This addition-merging cycle is iterated until no more states are added.

Each resulting polyad is associated to a block in the effective Hamiltonian. Its diagonalisation yields the correct (up to the given perturbative order) energies as eigenstates, and the corresponding eigenstates described as a linear combination of the harmonic states included in the polyad.

A typical run of the code thus proceeds through three steps:

1. initialisation of the data, reading hamiltonian terms, selection of which terms are to be considered resonances, construction of the infinitesimal contact transformations, construction of the transformed operators (currently the Hamiltonian and the electric dipole);

2. for each (harmonic) "starting state" in a list, defined by the number of quanta in each normal vibrational mode, build the polyads to be solved;

3. for each "starting state" processed in the previous step, actually diagonalise the effective Hamiltonian blocks, thus determining the states and their energies; compute all permitted transitions from the (diagonalised) starting state, build discretised spectra averaging the results according to the energy of the starting state, following the Wang-Landau scheme to uniformly sample energies.

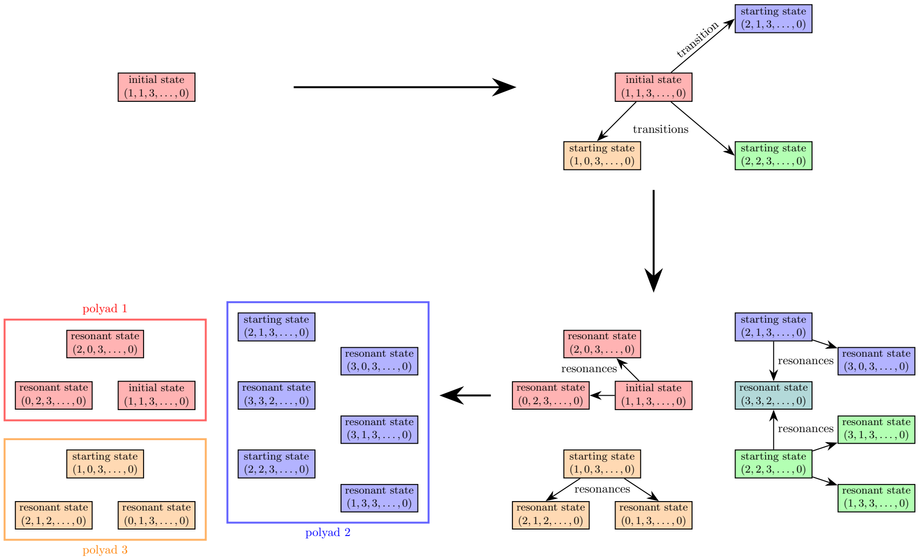

Figure 1: Schematic algorithm of polyad definition

Tunable parameters enable the user to define thresholds for which terms are to be considered resonances and which not. Considering more terms as resonances increases the accuracy of the calculation as well as its cost. Other tunable parameters define the truncation criteria for polyads. Again, including more states in a polyad makes the calculation more accurate but also more expensive. The first step above only has to be performed once for each given molecule and set of parameters. It has not been parallelised, but even for very large molecules and aggressive parameters it only takes a few hours to complete on a single core. It could in principle be parallelised, but it is not worth the effort. The second step is also not parallel, but requires fairly little resources both in terms of memory (∼1 Gbyte) and computational power (a few minutes on a single core per starting state). It could be parallelised, but it would be fairly difficult and error-prone, involving complicated syncronisations among MPI processes. Again, in the end it is not worth the effort. Instead, the first two steps have been split from the third one, so that they can be run separately on single-core jobs.

The third step is the really heavy one: for big enough molecules, having many resonances, even moderate choices for the tunable parameters result in having to diagonalise matrices which can easily be of the order of 100000×100000. This step has been parallelised, both in the building of the matrices and in diagonalising them. The code implements three different eigensolvers, which can be chosen at run time by the user. The first one is provided by the SLEPc library [13, 14, 12, 22], which offers a number of iterative solvers for sparse eigenproblems. Indeed, especially for large polyads, the eigenproblems are very sparse, with at most a few tens of non-zero elements per row (or column) of the matrix to be diagonalised. Also, only a relatively small subset of the eigensolutions are required, namely enough to cover the vector subspace of the starting states with good enough accuracy. As an indication, for a polyad with 100000 states one may need at most a few hundred eigensolutions, since they are the only ones involved in significantly intense transitions. However, unfortunately, all currently available algorithms for the iterative solving of sparse eigenproblems select the solutions based on the eigenvalues, while we need to select them based on the eigenvectors themselves (maximising the projection on the subspace of starting states). As a result, while SLEPc is very convenient from the point of view of memory efficiency, it is not competitive in terms of the computational effort required to obtain all the needed eigensolutions, since one gets them randomly mixed with unneeded ones. One ends up having to obtain all, or almost all, eigenvectors, and in this case direct eigensolvers are much more efficient than iterative ones.

We therefore also developed support in the code for more standard eigensolvers of dense matrices, using either the ScaLAPACK [4] or the ELPA [2, 16] libraries. Some care is needed to make sure that long integers are used to index matrix elements, since default integers would overflow when indexing elements in matrices larger than 32768×32768. The drawback of using standard solvers is that they need enough memory to accomodate the whole matrix being diagonalised, i. e. $N_{\mathrm{pol}}^2 \times \mathrm{sizeof(double)}$, and not only non-zero elements, as well as for all eigenvectors (same space as for the matrix) and some work space.

3 Tests and validation

To assess the accuracy of the code, and specifically of the Van Vleck approach to generalised vibrational 2$^{\mathrm{nd}}$ level perturbation theory (GVPT2), we ran a battery of calculations using as starting state the ground state of Ethylene Oxyde, Pyrene, and Coronene. In the first case, the molecule is small enough to make it tractable also with much more advanced (and expensive) computational techniques. We therefore obtained from P. Cassam-Chenaï a very accurate calculation performed using the Vibrational Mean Field Configuration Interaction (VMFCI) method [7, 6] using exactly the same quartic force field Hamiltonian as we used with AnharmoniCaOs. This enabled us to obtain some insight on how to tune the parameters of AnharmoniCaOs to get the desired accuracy. In the case of Pyrene and Coronene, instead, a VMFCI calculation would be prohibitively expensive, but we have experimental spectra available, and could compare with them. On the one hand, this enables us to assess the overall accuracy of our calculations, on the other hand we cannot distinguish whether the limiting factor is in our GVPT2 implementation or, instead, in the underlying quartic force field Hamiltonian, which we had previously computed using the Density Functional Theory at the B97-1/TZ2P level of theory [11, 27]. Moreover, Pyrene and Coronene are PAHs, hence belong to the class of molecules we want to actually study. As such, they have a large number of clustered vibrational frequencies, e. g. all aromatic C-H stretches around 3000 cm^{-1}, all out of plane bends around 800 cm^{-1}, and so on, which result in numerous close resonances and, consequently, in big polyads which merge in even bigger ones.

The results of these tests and validation were recently published on the Journal of Chemical Physics [19]. As an example, Figure 2 shows results for different calculations of the C-H stretch bands of pyrene computed in absorption at T=0K for varying values of the tunable parameter r acting as a threshold for resonance detection. Calculations performed with a smaller r include more and more resonances. As a result, one can see that even at 0K the spectrum goes from the relatively crude approximation with only the few IR-active fundamentals to a rich structure with several tens of combination bands activated by the many resonances. Comparison with experiment shows that, surprisingly, the spectrum with $r = 0.1$ is the closest one to laboratory data. This is unexpected, since from the theoretical point of view the calculation should become more and more accurate (at the price of increasing computational cost) when more and more resonances are included. This is likely the consequence of the inaccuracy of the quartic force field Hamiltonian we used, resulting from a DFT calculation, and of an accidental cancellation of (small) errors. In any case, the results appear to be extremely accurate, for an *ab initio* calculation.

Figure 3 shows the ratio of anharmonic versus harmonic frequencies of computed transitions, again for varying values of r, and compares them with the empyrical scaling factors [5] commonly adopted to match experimental band positions with harmonic vibrational analyses at the DFT level we used. Again, the agreement is excellent.

4 High temperature calculations for Pyrene

After validating our code, we proceeded to run a large number of calculations (about ten thousands of individual starting states foreseen) to obtain an adequate Monte Carlo sampling of the spectrum of Pyrene as a function of vibrational energy. The temperature-dependent spectrum can be obtained from this by means of a numerical Laplace transform. We were unfortunately unable to complete our calculations within the first B-class project, therefore we applied for a second one. We do have, however, some preliminary results. On the basis of our validation work, we decided to use a very conservative value $r = 0.3$ for the resonance detection parameter, to minimise the computational costs and make the Monte Carlo sampling feasible. We had determined that overall band positions were already quite accurate with $r = 3$, and detailed band structures would be lost anyway with increasing temperature. Figures 4 and 5 show the computed evolution of band profiles for increasing temperatures. The overall effect, in agreement with experiments, is a progressive broadening of the bands, and of a slight shift to the red of the position of the peak.

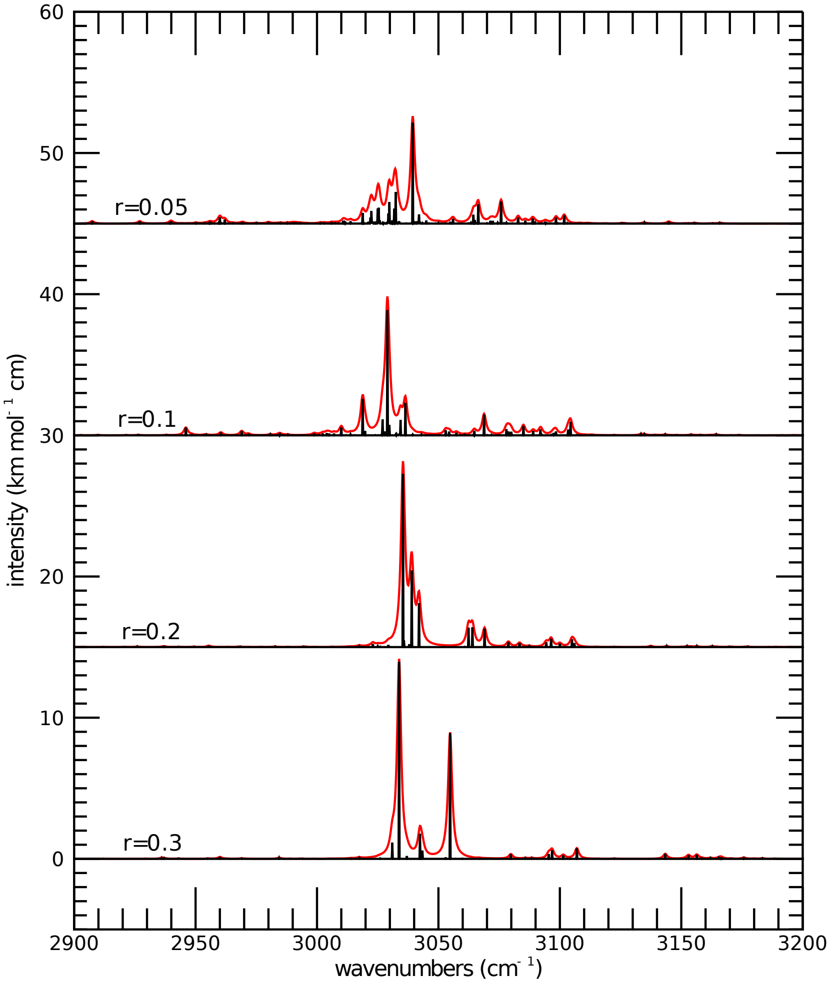

Figure 2: Evolution of the C-H stretch bands of Pyrene computed in absorption at T=0K for varying values of the threshold parameter for resonance detection r. From bottom to top an increasing number of resonances is included.

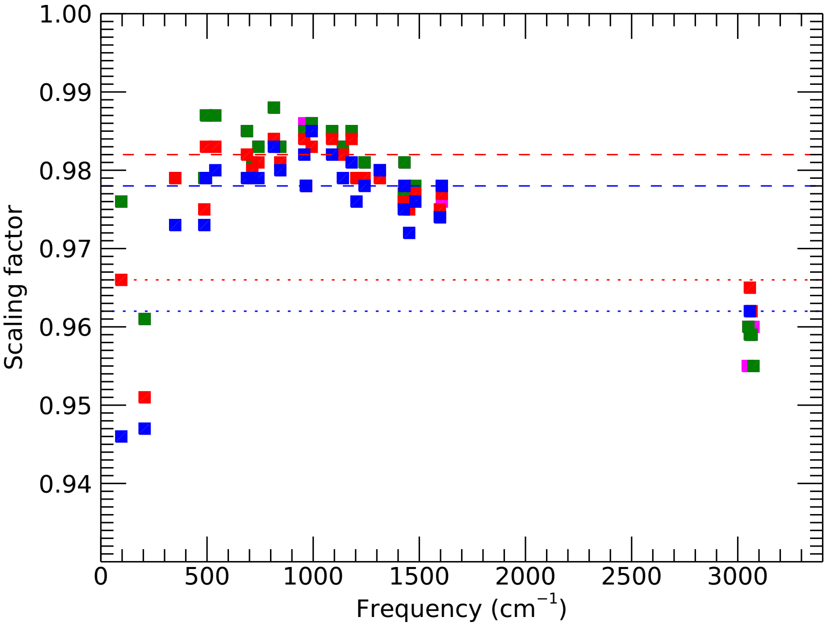

Figure 3: Ratios of the anharmonic vs. harmonic frequencies of computed transitions, for varying values of the resonance detection parameter r. Magenta squares are for $r = 0.3$, green ones for $r = 0.2$, red ones for $r = 0.1$, blue ones for $r = 0.05$. The overlaid lines represent the scaling factors for C-H stretches (dotted) and for all other bands (dashed). Red lines are empyrical scaling factors [5], namely 0.966 for C-H stretches and 0.982 for all the other bands, blue lines mark scaling factors obtained from our best anharmonic calculation for pyrene.

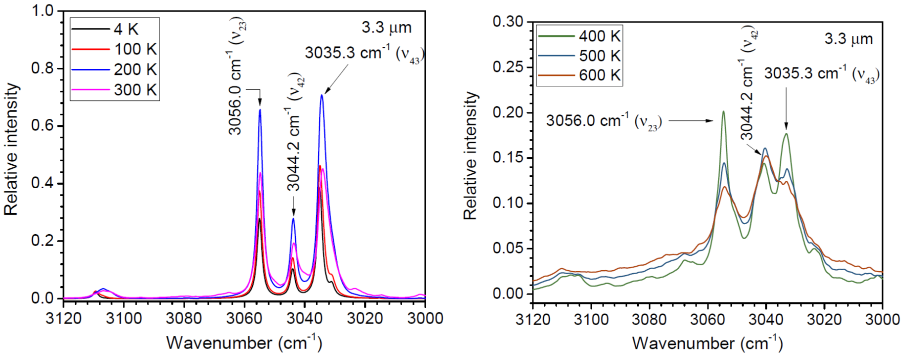

Figure 4: C-H stretch bands of Pyrene, all computed with resonance detection parameter $r = 0.3$

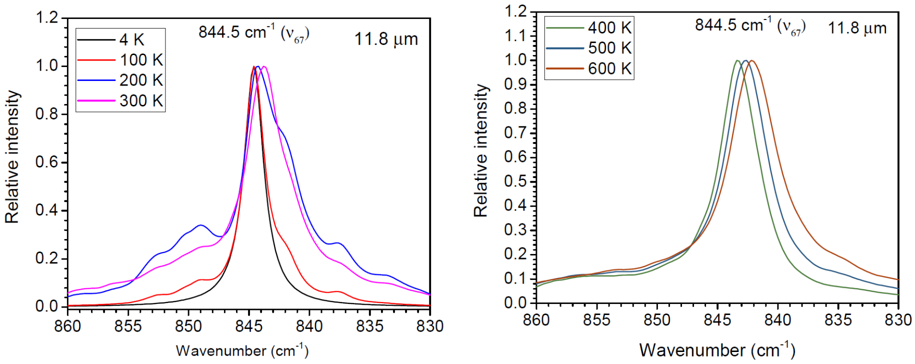

Figure 5: Out of plane C-H bend bands of Pyrene, all computed with resonance detection parameter $r = 0.3$

5 Discussion and Conclusions

Our preliminary results show that our approach can quantitatively account for the dependence of positions and profiles of PAH infrared bands on vibrational temperature. We were hindered, in our progress, by the lack of eigensolvers tailored to our problem. Indeed, we have to find a relatively small subset of the eigensolutions of very sparse, but rather large, real symmetric matrices. However, none of the available iterative eigensolvers for sparse matrices enables to selectively find the solutions we want, since these are determined by a condition of eigenvectors and not on eigenvalues. It was therefore much more efficient (or less inefficient) to use standard direct algorithms, like e. g. the Relatively Robust Representations [8], as implemented in well-tested libraries like ScaLAPACK or the relatively new ELPA. On Marconi we could use the optimised version of ScaLAPACK provided by the Intel Math Kernel Library (MKL), which enabled us to use both MPI across nodes and multithreading within a node for maximum efficiency. Nonetheless, as far as pure performance is involved, ELPA outperformed the MKL when it worked properly: it is faster and requires far less memory, since MKL needs quite some work space *for each thread*, which is dynamically allocated at runtime and not predictable beforehand. Indeed, a few job runs involving large polyads crashed just because at some point the MKL tried to allocate far more memory than we had estimated would be necessary. Unfortunately, in a number of apparently random (but reproducible) cases ELPA either failed gracefully (e.g. returning only eigenvalues but no eigenvectors, and issuing an error) or hung in an apparently endless internal loop. ScaLAPACK offers two functions suitable for our needs, namely `pdsyevr()` and `pdsyevx()`, the latter available only in the latest ScaLAPACK releases. While `pdsyevr()` sometimes failed gracefully to produce eigenvectors, `pdsyevx()` always returned complete, correct results in all our tests and subsequent production runs, except for the aforementioned cases in which the MKL crashed trying to allocate unavailable memory. We therefore settled for using `pdsyevx()` as a default.

We already published a first paper [19], validating our method, we are now in the process of finishing our runs, using our second, follow-up class B project, to produce our fully theoretical temperature-dependent spectra of a PAH, which will be submitted for publication within one or two months.

References

[1] L. J. Allamandola, A. G. G. M. Tielens, and J. R. Barker. Polycyclic aromatic hydrocarbons and the unidentified infrared emission bands - auto exhaust along the milky way. *Astrophys. J.*, 290:L25–L28, 1985.

[2] T. Auckenthaler, V. Blum, H.-J. Bungartz, T. Huckle, R. Johanni, L. Krämer, B. Lang, H. Lederer, and P.R. Willems. Parallel solution of partial symmetric eigenvalue problems from electronic structure calculations.

Parallel Computing, 37(12):783 – 794, 2011. 6th International Workshop on Parallel Matrix Algorithms and Applications (PMAA'10).

[3] Vincenzo Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.*, 122:014108, 2005.

[4] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[5] Elisabetta Cané, Andrea Miani, and Agostino Trombetti. Anharmonic force fields of naphthalene-h8 and naphthalene-d8. *J. Phys. Chem. A*, 111:8218–8222, 2007.

[6] P. Cassam-Chenaï and A. Ilmane. Frequently asked questions on the mean field configuration interaction method. i-distinguishable degrees of freedom. *J. Math. Chem.*, 50:652–667, 2012.

[7] P. Cassam-Chenaï and J. Liévin. The VMFCI method: A flexible tool for solving the molecular vibration problem. *J. Comput. Chem.*, 27:627–640, 2006.

[8] I. Dhillon. A new o(n^2) algorithm for the symmetric tridiagonal eigenvalue/eigenvector problem. Technical Report UCB/CSD-97-971, University of California, Berkeley, CA, May 1997.

[9] B.T. Draine. Astronomical models of pahs and dust. *EAS Publications Series*, 46:29–42, 2011.

[10] F. Gaw, A. Willets, N. Handy, and W. Green. *Advances in Molecular Vibrations and Collisions Dynamics*, volume 1B, page 169. JAI Press Inc., Greenwich (CT), 1991.

[11] Fred A. Hamprecht, Aron J. Cohen, David J. Tozer, and Nicholas C. Handy. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.*, 109:6264–6271, 1998.

[12] V. Hernandez, J. E. Roman, and V. Vidal. SLEPc: Scalable Library for Eigenvalue Problem Computations. *Lect. Notes Comput. Sci.*, 2565:377–391, 2003.

[13] Vicente Hernandez, Jose E. Roman, and Vicente Vidal. SLEPc - the scalable library for eigenvalue problem computations. http://slepc.upv.es/.

[14] Vicente Hernandez, Jose E. Roman, and Vicente Vidal. SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems. *ACM Trans. Math. Software*, 31(3):351–362, 2005.

[15] A. Léger and J. L. Puget. Identification of the 'unidentified' IR emission features of interstellar dust? *Astron. Astrophys.*, 137:L5–L8, 1984.

[16] A Marek, V Blum, R Johanni, V Havu, B Lang, T Auckenthaler, A Heinecke, H-J Bungartz, and H Lederer. The elpa library: scalable parallel eigenvalue solutions for electronic structure theory and computational science. *Journal of Physics: Condensed Matter*, 26(21):213201, 2014.

[17] Jan M. L. Martin, Timothy J. Lee, Peter R. Taylor, and Jean-Pierre François. The anharmonic force field of ethylene, C_2H_4, by means of accurate ab initio calculations. *J. Chem. Phys.*, 103:2589, 1995.

[18] Anne B. McCoy and III Sibert, Edwin L. An algebraic approach to calculating rotation-vibration spectra of polyatomic molecules. *Molecular Physics*, 77:697–708, January 1992.

[19] G. Mulas, C. Falvo, P. Cassam-Chenaï, and C. Joblin. Anharmonic vibrational spectroscopy of polycyclic aromatic hydrocarbons (PAHs). *J. Chem. Phys.*, 149(14):144102, October 2018.

[20] Giacomo Mulas and Cyril Falvo. AnharmoniCaOs - Cagliari-Orsay model for anharmonic molecular spectra in 2nd order pt. https://sourceforge.net/projects/anharmonica/.

[21] Matteo Piccardo, Julien Bloino, and Vincenzo Barone. Generalized vibrational perturbation theory for rotovibrational energies of linear, symmetric and asymmetric tops: Theory, approximations, and automated approaches to deal with medium-to-large molecular systems. *Int. J. Quantum Chem.*, 115:948–982, 2015.

[22] J. E. Roman, C. Campos, E. Romero, and A. Tomas. SLEPc users manual. Technical Report DSIC-II/24/02 - Revision 3.10, D. Sistemes Informàtics i Computació, Universitat Politècnica de València, 2018.

[23] Edwin L. Sibert. Vanvlk: An algebraic manipulation program for canonical van vleck perturbation theory. *Comput. Phys. Commun.*, 51:149–160, 1988.

[24] Edwin L. Sibert. Dressed local mode hamiltonians for ch stretch vibrations. *Mol. Phys.*, 111:2093–2099, 2013.

[25] Edwin L. Sibert, III. Theoretical studies of vibrationally excited polyatomic molecules using canonical van vleck perturbation theory. *J. Chem. Phys.*, 88:4378–4390, 1988.

[26] A. G. G. M. Tielens. Interstellar polycyclic aromatic hydrocarbon molecules. *Annu. Rev. Astro. Astrophys.*, 46:289–337, 2008.

[27] E. Van Lenthe and E. J. Baerends. Optimized slater-type basis sets for the elements 1–118. *J. Comput. Chem.*, 24:1142–1156, 2003.