

Sensitivity of semiclassical vibrational spectroscopy to potential energy surface accuracy: A test on formaldehyde

Riccardo Conte*, Giacomo Botti, Michele Ceotto

Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano (Italy)

Abstract

A set of permutationally invariant potential energy surfaces for the electronic ground state of formaldehyde is built at several levels of electronic theory and atomic orbital basis sets starting from a database of more than 34000 *ab initio* energies. Preliminarily, the reliability of the fitted surfaces is determined by comparing the calculated harmonic frequencies with the corresponding *ab initio* values. Then, semiclassical estimates of the quantum frequencies of vibration are presented, and their dependence on the employed level of theory, type of atomic orbital basis set, and complexity of the fit is investigated. Comparisons of semiclassical results to experimental data provide a further assessment of the quality of the analytical surfaces and show that anharmonic frequencies are influenced by the precision of the fit, while accurate frequency values are obtained also with density functional theory. Results and conclusions support the use of *ab initio* “on-the-fly” semiclassical dynamics as a means of spectroscopic investigation when high-level analytical potential energy surfaces are not available.

Keywords: Semiclassical Dynamics; Vibrational Spectroscopy; PES fitting; Formaldehyde.

* Corresponding author.

*Email address:*riccardo.conte1@unimi.it (Riccardo Conte)

1. Introduction

The accurate description of the potential energy surface (PES) of a molecule is crucial for a reliable theoretical spectroscopy investigation. Theoretical spectroscopy methods may require knowledge of the global PES or demand for just local information. In the former case the surface must be fitted by an analytical expression, which should be not only flexible enough to ensure a good precision over a large range of molecular geometries (including distorted ones) but also fast to be evaluated. In the other instance the potential and related quantities, for instance gradient components, are computed *ab initio* only at relevant geometries, a process which, in the case of dynamical approaches, is commonly labeled as “on-the-fly”.

The choice of a suitable analytical form for fitting the PES is not a trivial one. One possible approach consists in dividing the potential energy into a sum of different kinds of parametrized contributions. Their analytical expressions are usually chosen to be computationally cheap and harmonic or Morse oscillators based on internal coordinates are often employed. The parameters are then fitted to a large set of experimental and theoretical benchmark data to provide a general-purpose force field. In this way, the resulting force field can be used for fast application to a large number of chemical compounds, but its reliability may be insufficient for calculations aiming at spectroscopic (i.e. wavenumber) accuracy.

Much better accuracy can be achieved by constructing a specific PES for the molecule or system of interest at the cost of slower potential energy evaluations. For this purpose several approaches have been developed even recently. They include, among others, Shepard interpolation,[1] permutationally invariant fitting,[2] empirical valence bond theory,[3] neural networks,[4] and machine learning.[5] In these methods a database or training set of energies is prepared usually employing a very high level of electronic structure theory. The relevant geometries may be collected for instance through grid searches or by means of molecular dynamics runs (at low electronic theory level) to sample the phase space. Furthermore, a detailed analytical form for the surface is chosen to fit the data. Eventually, the outcome is a precise and targeted PES, which can be employed in very refined calculations including spectroscopy ones. Many of such surfaces have been constructed in recent years for rather small systems (less than 10 atoms), while current scientific effort is devoted to attain comparable accuracy for bigger systems keeping the burden of computational costs manageable.[6, 7]

Among the several quantum approaches to vibrational spectroscopy, we focus on semiclassical (SC) dynamics. SC vibrational spectroscopy is a dynamical method able to describe quantum effects with high accuracy starting from classical trajectories.[8–18] It is the mathematical structure of the SC propagator that allows to regain quantum effects from the information included in the classical trajectories,[19] and to avoid zero point energy leakage.[20] The method, in its initial value representation (IVR) version,[21] can be interfaced with both analytical PESs and “on-the-fly” potential evaluations. In the first case, the vibrational frequencies of interest are obtained through integration of the contributions to the SCIVR propagator coming from a distribution of trajectories. In the second case, due to the overwhelming computational cost of an “on-the-fly” SC simulation, very few trajectories (or even a single one) are employed. The initial conditions of these tailored trajectories are determined according to the recipe established by the multiple coherent states semiclassical initial value representation (MC SCIVR) approach.[22–24] Applications are not limited to small systems but also medium-large molecules and supramolecular aggregates can be studied.[25–32] This is permitted by the divide-and-conquer SCIVR approach (DC SCIVR), which allows one to define an appropriate and manageable set of low dimensional investigations rather than dealing with the inaccessible full-dimensional problem.[33–35] Finally, recent advances in the field have permitted to determine vibrational eigenfunctions,[36, 37] IR spectra,[38] decrease substantially the impact of deterministic chaos on SC simulations leading to more accurate and precise results,[39] and decrease the computational overhead related to Hessian calculations.[40–43]

The last aspect points out why “on-the-fly” SCIVR calculations are so computationally demanding. The reason is that *ab initio* electronic calculations must be performed not only for energies and gradients, but also for the Hessian matrix of the potential. The anticipated advances to ease this computational overhead consist in techniques able to decrease the number of Hessian calculations needed along the trajectory, while faster computation hardware, like GPUs, is also employable.[44] However, high levels of electronic theory, like second order Möller-Plesset perturbation theory (MP2) or coupled cluster (CCSD(T)), are still unsuitable for “on-the-fly” semiclassical spectroscopy, which is limited to simulations based on density functional theory (DFT).

Indeed, an open question about semiclassical methods for vibrational spectroscopy concerns their sensitivity to the treatment of the underlying

electronic structure problem. The issue is not trivial and related to the dualism between analytical-surface-based and “on-the-fly” approaches. On the one hand an analytical PES may be based on data calculated at very high levels of electronic theory, but introduction of a fitting error is unavoidable; on the other hand “on-the-fly” approaches are free of fitting errors but limited to lower levels of electronic theory.

The goal of this paper is to move a first step toward the understanding of the influence of these two competing effects on the accuracy of SC frequency estimates. For this purpose, as a test case, we built a set of permutationally invariant analytical PESs for the electronic ground state of H₂CO at several levels of electronic theory and atomic orbital basis sets. The surfaces were also fitted by several functional forms of different complexity to tune the fitting error. In Section 2 we report the details of the methodologies employed. Section 3 is dedicated to the presentation of results. Finally, in Section 4, we briefly discuss the work and in Section 5 we come to the conclusions.

2. Theory

Our first step in the construction of a permutationally invariant potential energy surface for the electronic ground state of formaldehyde consisted in the collection of an appropriate database of energies. This is usually performed by sampling the phase space by means of *ab initio* direct dynamics trajectories evolved at low level of electronic structure theory. The single energies of the various geometries are then refined adopting a higher level of theory. In Table 1 the different levels of electronic calculations undertaken in this work are reported.

Upon determination of all energies we fitted the PESs. We employed the permutationally invariant polynomial technique of Braams and Bowman.[2] Finally, the analytical surfaces were used for our semiclassical vibrational spectroscopy calculations. More details on these two techniques are given in the following subsections.

2.1. Permutationally invariant fit

Invariance of the potential energy under translations and rigid rotations of the molecular geometry is among the properties of a PES. It may be enforced by defining the potential energy in terms of functions of scalar and cross products of the internuclear Cartesian distances. One possible way to build a fitting basis set is to begin with monomials dependent on the

Table 1: Levels of theory of the constructed energy databases and potential energy surfaces.

Theory level	Basis set
DFT	6-31G**
	cc-pVDZ
	cc-pVTZ
MP2	6-31G**
	cc-pVDZ
	cc-pVTZ
CCSD(T)	6-31G**
	cc-pVDZ

internuclear distances, and then combine them into polynomials of various complexity up to a chosen maximum order. This ensemble of polynomials constitutes the fitting basis set. The mathematical expression for the variables on which monomials and polynomials depend should be driven by some physical-chemical insight. A proper choice for an isolated molecule is represented by Morse variables

$$M_k = \exp(-r_k/a), \quad (1)$$

where r_k defines the k -th internuclear distance, and a is a constant. The value of a could be optimized, but it is commonly chosen to be equal to 2 Bohr.[2] Morse variables are certainly suitable to describe molecular motion and distorted geometries including bond elongation and dissociation. However, another key feature of the PES is missing and should be introduced. Specifically, the potential energy must be invariant under permutation of like atoms. This is also the case for the H₂CO molecule investigated here. Because of the two hydrogen atoms, formaldehyde belongs to the A₂BC permutational group and particular care must be taken in symmetrizing the polynomials appropriately. The process is exemplified by Figure 1, which shows how internuclear distances are labeled, and by Eq. (2), which presents some of the permutationally invariant polynomials p .

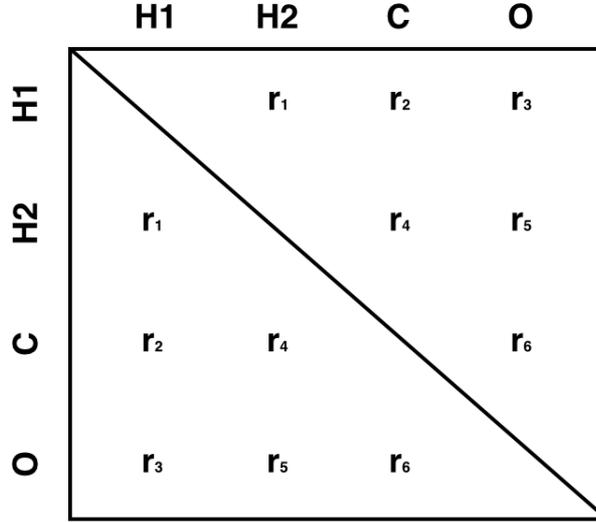


Figure 1: Definition of the internuclear distances for H_2CO . The two hydrogen atoms are labeled as H_1 and H_2 .

$$\begin{aligned}
p(0) &= 1 \\
p(1) &= \exp(-r_1/a) \\
p(2) &= \exp(-r_2/a) * \exp(-r_4/a) \\
p(3) &= \exp(-r_3/a) * \exp(-r_5/a) \\
p(4) &= \exp(-r_6/a) \\
p(5) &= \exp(-r_2/a) + \exp(-r_4/a) \\
p(6) &= \exp(-r_3/a) + \exp(-r_5/a)
\end{aligned} \tag{2}$$

The first five polynomials in Eq. (2) are actually some of the possible symmetrized monomials, while $p(5)$ and $p(6)$ are binomials. Furthermore, $p(2)$ and $p(3)$ are second order polynomials (monomials) and, consequently, they would not be part of the fitting basis set if a maximum polynomial order of 1 were chosen. Appropriate combinations of symmetrized monomials and polynomials define the whole permutationally invariant fitting basis set, whose complexity depends on the maximum polynomial order allowed.[45, 46] Dedicated software for this purpose has been developed by Xie and Bowman.[47]

A linear coefficient is associated to each polynomial, and the fit is performed by means of a minimum least squares algorithm. More elaborated

Table 2: Dependence of the number of permutationally invariant polynomials on the maximum polynomial order for H₂CO.

max. polynomial order	2	3	4	5	6	7	8	9
No. polynomials	18	50	120	256	502	918	1589	2625

basis sets provide more precise fits, but potential calls are slower, and overfitting should be carefully avoided. Table 2 reports the number of coefficients (i.e. permutationally invariant polynomials) for the A₂BC permutation group as a function of the maximum polynomial order.

Once the coefficients have been determined, the analytical surface is ready for use and the potential energy, at a given molecular geometry, can be calculated as the scalar product between the arrays of coefficients and permutationally invariant polynomials evaluated at that specific geometry.

2.2. Semiclassical spectroscopy

Determination of the frequencies of vibration of a molecule is a problem at the heart of quantum mechanics, even if approximations based on harmonic or classical estimates are widely (sometimes inappropriately) employed. Semiclassical dynamics is a quantum dynamical approach that allows one to get quantum estimates starting from classical trajectories. Once an analytical PES is available, SC calculations can be undertaken by means of Kaledin and Miller’s time-averaged (TA) SCIVR formula for semiclassical power spectra[48, 49]

$$I(E) = \frac{1}{(2\pi\hbar)^{N_{vib}}} \int \int d\mathbf{p}_0 d\mathbf{q}_0 \frac{1}{2\pi\hbar T} \left| \int_0^T dt e^{i[S_t + \phi_t + Et]/\hbar} \langle \mathbf{p}_t, \mathbf{q}_t | \chi \rangle \right|^2. \quad (3)$$

$I(E)$ is the energy-dependent density of vibrational states, which is characterized by signals (ideally Dirac deltas) with maximum intensity located at the eigenvalues of the vibrational Hamiltonian. From the power spectrum it is trivial to get by difference the quantum frequency associated with the transition between any two states. In particular, the frequencies of the transitions between the vibrational ground state at zero point energy (ZPE) and the states corresponding to a single excitation of the modes of vibration are called fundamental frequencies. In Eq. (3) N_{vib} is the number of vibrational degrees of freedom; \mathbf{p}_0 and \mathbf{q}_0 determine the phase space points from which

the classical trajectories start; T is the total evolution time; S_t is the instantaneous classical action; ϕ_t is a phase; $\langle \mathbf{p}_t, \mathbf{q}_t | \chi \rangle$ is the quantum mechanical overlap between the time-evolved coherent state $|\mathbf{p}_t, \mathbf{q}_t\rangle$ and the quantum reference state $|\chi\rangle$. The phase, which is related to the Herman-Kluk SC propagator, is defined as

$$\phi_t(\mathbf{p}_0, \mathbf{q}_0) = phase \left[\sqrt{\left| \frac{1}{2} \left(\frac{\partial \mathbf{q}_t}{\partial \mathbf{q}_0} + \Gamma^{-1} \frac{\partial \mathbf{p}_t}{\partial \mathbf{p}_0} \Gamma - i\hbar \frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_0} \Gamma + \frac{i\Gamma^{-1}}{\hbar} \frac{\partial \mathbf{p}_t}{\partial \mathbf{q}_0} \right) \right|} \right], \quad (4)$$

where the partial derivatives are the block elements of the stability (monodromy) matrix and Γ is the coherent state width matrix.[50] The physical meaning of the latter is clarified by the coordinate representation of a coherent state, i.e.

$$\langle \mathbf{q} | g_t(\mathbf{p}_0, \mathbf{q}_0) \rangle = \left(\frac{det(\Gamma)}{\pi^{N_{vib}}} \right)^{1/4} \exp \left[-(\mathbf{q} - \mathbf{q}_t)^T \frac{\Gamma}{2} (\mathbf{q} - \mathbf{q}_t) + \frac{i}{\hbar} \mathbf{p}_t^T (\mathbf{q} - \mathbf{q}_t) \right]. \quad (5)$$

Γ is usually a diagonal matrix with elements equal to the harmonic frequencies of vibration ω_{harm} , which are easily obtained upon diagonalization of the Hessian matrix of the potential energy.

Eq. (3) requires first to sample the phase space to determine the initial conditions for the classical dynamics, then to run the dynamics, and eventually to achieve Monte Carlo convergence. As for the initial sampling the common procedure consists in selecting phase space points $(\mathbf{p}_0, \mathbf{q}_0)$ from a Husimi distribution centered at the equilibrium geometry and momenta with harmonic zero-point energy components. The Gaussian widths of the distribution for the i -th generic vibrational mode can be tailored on the simulation, but they are usually chosen to be $\sqrt{\omega_{harm,i}/2}$ for the coordinates and $\sqrt{1/(2\omega_{harm,i})}$ for the momenta, in agreement with SC calculations based on the parent Herman-Kluk propagator. The dynamics is evolved by means of a 4-th order symplectic integrator,[51] and convergence is achieved by running a few thousand trajectories per vibrational degree of freedom.

The last aspect of Eq. (3) still to be discussed concerns the role of the quantum reference state $|\chi\rangle$. This is an arbitrary state which is decomposed onto the complete basis of the eigenfunctions of the vibrational Hamiltonian by the semiclassical calculation. The consequence is twofold: on the one

hand, from an SC power spectrum it is possible to collect the whole spectrum of the operator (i.e. all eigenvalues); on the other hand, the intensity of signals can be tuned according to the reference state adopted. Indeed, it is possible to select as reference states some appropriate linear combinations of coherent states that are able to enhance specific spectral signals according to the parity or group symmetry of the target eigenstate.[52]

3. Results

The work has been divided into two parts. First we constructed a set of potential energy surfaces for H₂CO as anticipated in Table 1, and then performed our semiclassical vibrational spectroscopy simulations. The PES of formaldehyde had already been studied in details.[53, 54] Apart from the global minimum it is characterized by two isomers denoted as trans-HCOH and cis-HCOH. However, for our purposes, in this work we focused only on the spectroscopic features of the global minimum well.

3.1. PES construction

Potential energies for H₂CO were calculated with NWChem,[55] for the several levels of electronic theory and atomic orbital basis sets reported in Table 1, at the geometries included in a pre-existing database made of about 34000 points.[56] We maintained the full set of points for DFT and two sets of MP2 evaluations, while CCSD(T) calculations were performed on a reduced database of 20000 points. Another exception was the MP2/cc-pVDZ database (28575 points). Data were fitted by means of permutationally invariant basis sets featuring maximum polynomial orders ranging from 6 (502 polynomials) to 9 (2625 polynomials). Table 3 summarizes the root mean square error (RMSe) of the several fits, calculated according to

$$\text{RMSe} = \sqrt{\frac{\sum_{i=1}^{N_{db}} (E_{i,db} - E_{i,PES})^2}{N_{db}}}, \quad (6)$$

where N_{db} is the number of points in the database, $E_{i,db}$ is the *ab initio* energy of the i -th point in the database, and $E_{i,PES}$ is the corresponding energy calculated on the analytical PES.

From Table 3 it is clear that there is an improvement in fit precision when increasing the number of polynomials in the fitting basis set. The more elaborated fits return RMS errors of about 0.5 kcal/mol in most cases.

Table 3: Root mean square errors associated to the fitting of several analytical surfaces. Values are in kcal/mol.

No. polys	DFT			MP2			CCSD(T)	
	6-31G**	cc-pVDZ	cc-pVTZ	6-31G**	cc-pVDZ	cc-pVTZ	6-31G**	cc-pVDZ
502	1.464	1.444	1.389	1.546	0.986	1.428	0.791	0.842
918	0.869	0.853	0.812	0.925	0.735	0.837	0.751	0.804
1589	0.604	0.590	0.557	0.653	0.645	0.579	0.741	0.793
2625	0.531	0.518	0.487	0.571	0.619	0.501	0.732	0.784

Table 4: Mean Absolute Errors (cm^{-1}) of harmonic frequencies calculated with analytical PESs with respect to *ab initio* values. The surfaces have been fitted at several levels of electronic theory with basis sets made of 502 and 2625 polynomials.

No. polys	DFT			MP2			CCSD(T)	
	6-31G**	cc-pVDZ	cc-pVTZ	6-31G**	cc-pVDZ	cc-pVTZ	6-31G**	cc-pVDZ
502	35	31	33	50	20	49	28	30
2625	27	14	18	12	13	13	20	25

As anticipated, this was not the case in 3 instances, and to get comparable RMSe values for all levels of theory and basis sets the database of energies has been reduced by getting rid of the points at higher energy and associated to the most distorted geometries compared to the equilibrium one.

To better assess the accuracy of the fitted PESs we have calculated the harmonic frequencies for each of them. Table 4 reports the mean absolute errors (MAE) of the six harmonic frequencies of each PES with respect to the corresponding *ab initio* value.

The complete list of harmonic frequencies can be found in the supplementary material. As expected, by increasing the dimensionality of the fitting basis set we were able to get values closer to the *ab initio* ones. The magnitude of the improvement of estimates moving from the simpler fit (502 polynomials) to the more elaborated one (2625 polynomials) is strictly related to the decrease in the fitting RMSe. Thereby at the harmonic level differences are smaller for the surfaces constructed starting from the reduced databases, which cover a restricted portion of the configuration space and allow for a better fit when a small fitting basis set is employed. The enhanced accuracy of a fit based on 2625 polynomials comes at the price of slower potential energy evaluations. However the scaling in cpu times is linear with the number of coefficients, a feature that allows one to adopt the

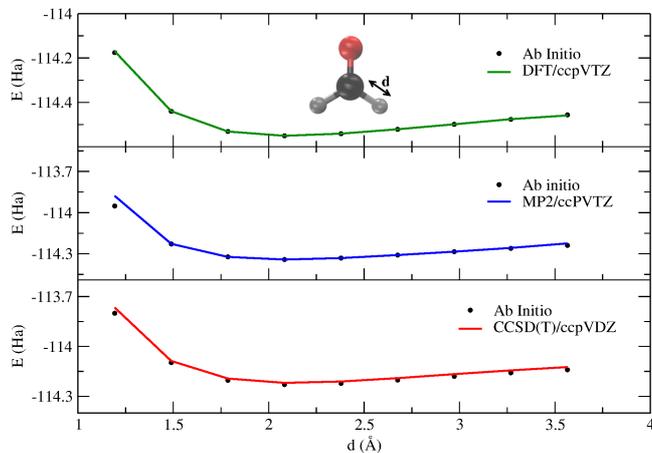


Figure 2: Potential cuts for CH bond elongation. *Ab initio* values (black dots) are compared to analytical surface evaluations at DFT/ccpVTZ (green), MP2/ccpVTZ (blue), and CCSD(T)/ccpVDZ (red) level.

more precise fit also in SC calculations. A Table illustrating this aspect can be found in the supplementary material.

Finally, we wanted to test the analytical surfaces on geometries not included in our databases. To this end we performed some one-dimensional PES cuts describing the stretching of the CH and CO bonds, and the displacement of the carbon atom above and below the molecular plane. In Fig. 2 we show potential energy estimates at several values of the CH bond length for the three levels of theory and biggest basis sets, and compare them to the corresponding *ab initio* values. The other two cuts at the same levels of theory and three additional cuts involving the smaller 6-31G** basis set are reported in the supplementary material. The general agreement is quite satisfactory. Accuracy deteriorates occasionally for the CCSD(T) fits as a consequence of the fact that they have been built starting from reduced and more localized databases, which make these fits less suitable for the description of geometries far from equilibrium.

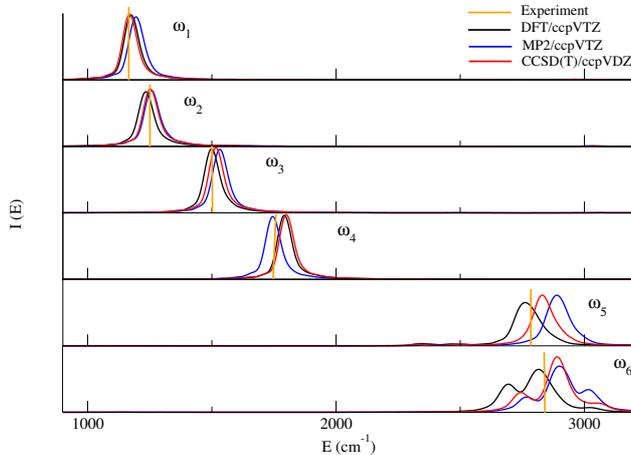


Figure 3: Semiclassical power spectra for H_2CO . Experimental values of the six fundamental frequencies (orange vertical bar) are compared to results at DFT/ccpVTZ (black), MP2/ccpVTZ (blue), and CCSD(T)/ccpVDZ (red) level of theory.

3.2. Semiclassical spectra

After constructing and testing the potential energy surfaces, we moved to the TA-SCIVR spectroscopy calculations. We performed a different simulation for each of the six fundamental frequencies employing a tailored reference state able to enhance the desired spectroscopic signal. Specifically we chose

$$|\chi\rangle = \prod_{i=1}^6 (|p_i^{\text{harm}}, q_i^{\text{eq}}\rangle + \varepsilon_i | -p_i^{\text{harm}}, q_i^{\text{eq}}\rangle), \quad (7)$$

with $\varepsilon_i = -1$ for the i -th fundamental we were interested in, and $\varepsilon_i = +1$ otherwise. Fig. 3 illustrates the results for three different levels of theory and compares them to experimental results.[57]

The corresponding, detailed numerical results are reported in Table 5. The calculated frequencies reproduce the experimental data often accurately. The complete set of vibrational frequencies obtained for all surfaces constructed is reported in the supplementary material.

TA-SCIVR returns MAE values for the three levels of theory close to each other with a minimum for the surface built at the DFT/ccpVTZ level. In Fig.

Table 5: Fundamental frequencies of vibration (cm^{-1}) for H_2CO calculated with three different potential energy surfaces. MAE stands for Mean Absolute Values.

	DFT/ccpVTZ	MP2/ccpVTZ	CCSD(T)/ccpVDZ	Experiment
ω_1	1175	1194	1168	1167
ω_2	1234	1256	1252	1249
ω_3	1499	1531	1516	1500
ω_4	1791	1744	1800	1746
ω_5	2761	2838	2830	2782
ω_6	2814	2898	2890	2843
MAE	19	29	27	-

4 the DFT fundamental frequencies are compared to harmonic estimates and experimental values, thus allowing one to fully appreciate the anharmonicity of the vibrational motions.

4. Discussion

An internal check to verify the effectiveness of the adopted fitting procedure consists in comparing the harmonic frequencies calculated with the analytical surfaces to their *ab initio* counterparts. Under this aspect, all the constructed PESs have passed the test with an improvement in accuracy moving from surfaces based on 502 terms to those characterized by 2625 polynomials. The potential cuts, which involve geometries not included in the databases used for the fits, further demonstrate the reliability of the surfaces. In fact, very good agreement between fits and *ab initio* data is generally found, with the exception of a few points associated to distorted geometries. In particular this is the case (see Supporting material) for the cut in which a carbon atom is displaced above and below the molecular plane and the CCSD(T) surfaces fitted on a restricted database are employed.

An estimate of the absolute accuracy of the PESs can be obtained by comparing calculated frequencies with experimental values. We have performed this test at both harmonic and semiclassical level and the outcome is reported in Table 6.

An immediate feature to be noticed is that, on average, the harmonic frequencies are closer to the experiment in the case of the less accurate 502-polynomial-based fits. However, this is only apparently puzzling because experiments certainly take into account the anharmonicity of the molecular

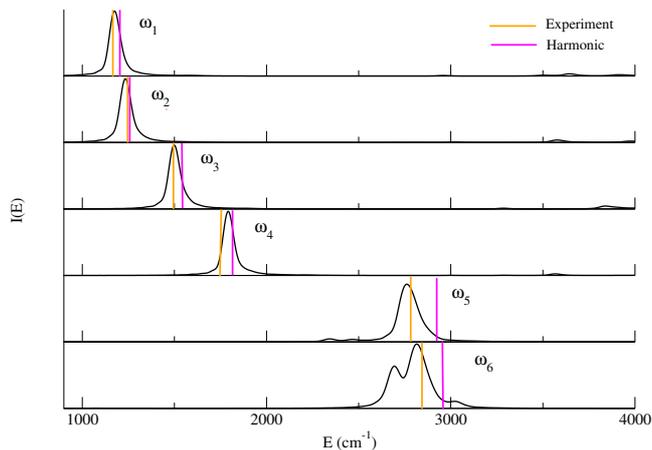


Figure 4: Semiclassical power spectra for H_2CO calculated at the DFT/ccpVTZ level. Harmonic estimates (pink vertical bars) and experimental frequencies (orange vertical bars) complete the graphs.

Table 6: Mean absolute errors (cm^{-1}) of harmonic and semiclassical frequencies with respect to experimental values. The number of polynomials employed for the fit are reported in parenthesis.

Calculation	DFT			MP2			CCSD(T)	
	6-31G**	cc-pVDZ	cc-pVTZ	6-31G**	cc-pVDZ	cc-pVTZ	6-31G**	cc-pVDZ
Harm (502)	72	68	59	98	84	76	137	100
SC (502)	32	36	31	50	27	34	65	30
Harm (2625)	104	62	69	126	103	112	124	87
SC (2625)	21	30	19	57	27	29	61	27

vibrations. Thereby, moving to the more realistic semiclassical estimates, it is the more refined and computationally demanding 2625-polynomial-based surfaces that return the best estimates and overperform neatly the harmonic values, demonstrating the importance of having an accurate description not only near the equilibrium geometry but also in the anharmonic part of the potential well. Furthermore data for MP2 and CCSD(T) show the importance of choosing an adequately large atomic orbital basis set, a characteristic which appears to be more influential than the level of electronic theory.

A deeper investigation of SC power spectra points out three main aspects. One is that different levels of theory with sufficiently large basis sets provide similar results with more evident discrepancies for the two highest frequencies. This is related to the second aspect, i.e., frequency estimates are worse for ω_5 and ω_6 . A possible cause lies in the fact that the database was originally designed for a surface at a level of theory and basis set not considered here (see Ref. 56). One could refine the database around the equilibrium geometry and re-fit the surface possibly improving the estimates of the highest frequency modes, which are rapidly influenced by small changes in the topology of the surface near the equilibrium structure. Finally, SC power spectra demonstrates clearly the role of anharmonicities showing not only that harmonic estimates are often badly off the mark but also that the amount of anharmonicity is different mode by mode, which is a clue suggesting that scaled harmonic approaches should not be employed for accurate vibrational investigations.

5. Conclusions

We have presented a study of formaldehyde vibrational frequencies by means of semiclassical power spectra. The investigation was based on the construction of several potential energy surfaces at different levels of theory, atomic orbital basis set, and complexity of the analytical surface, followed by the calculation of harmonic and semiclassical anharmonic frequencies. From the analysis of the results, we draw a twofold conclusion. First, an adequately large basis set is necessary to obtain accurate results. This is evident for MP2 and CCSD(T) calculations. The role of the basis set overshadows the choice of the level of electronic theory, because results for anharmonic frequencies are accurate also when employing DFT surfaces. Secondly, PESs based on a larger number of polynomials fit better the *ab initio* points in the database and return more accurate semiclassical estimates.

In this work we have studied a very small molecule. Certainly the role of the *ab initio* level of theory gains importance when moving to larger and more complex systems for which density functional theory is not able to ensure a level of accuracy similar to post Hartree-Fock methods. However, results suggest that when high levels of electronic theory are not computationally affordable, DFT with a suitable basis set can be a valid alternative, provided that the potential energy surface is a precise description of the *ab initio* energy landscape not only close to the well minimum but also far in the anharmonic region. This substantiates the success in many applications of DFT-level *ab initio* “on-the-fly” semiclassical dynamics, which is, by definition, free of fitting errors.

Acknowledgments

This work was supported by the Italian Ministry for Education, University, and Research (MIUR) [FARE programme No. R16KN7XBRB - project QURE]. The authors thank also CINECA (Italian Supercomputing Center) for allocation of computational time under IskraC project “heavyTUN”.

- [1] M. A. Collins, Molecular potential energy surfaces constructed from interpolation of systematic fragment surfaces, *J. Chem. Phys.* 127 (2) (2007) 024104.
- [2] B. J. Braams, J. M. Bowman, Permutationally invariant potential energy surfaces in high dimensionality, *Int. Rev. Phys. Chem.* 28 (4) (2009) 577–606.
- [3] A. Warshel, R. M. Weiss, An empirical valence bond approach for comparing reactions in solutions and in enzymes, *J. Am. Chem. Soc.* 102 (20) (1980) 6218–6226.
- [4] B. Jiang, H. Guo, Permutation invariant polynomial neural network approach to fitting potential energy surfaces. III. Molecule-surface interactions, *J. Chem. Phys.* 141 (3) (2014) 034109.
- [5] A. P. Bartok, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csanyi, M. Ceriotti, Machine learning unifies the modeling of materials and molecules, *Sci. Adv.* 3 (12).

- [6] C. Qu, J. M. Bowman, A fragmented, permutationally invariant polynomial approach for potential energy surfaces of large molecules: Application to n-methyl acetamide, *J. Chem. Phys.* 150 (14) (2019) 141101.
- [7] A. Nandi, C. Qu, J. M. Bowman, Full and fragmented permutationally invariant polynomial potential energy surfaces for trans and cis n-methyl acetamide and isomerization saddle points, *J. Chem. Phys.* 151 (8) (2019) 084306.
- [8] E. J. Heller, The semiclassical way to molecular spectroscopy, *Acc. Chem. Res.* 14 (12) (1981) 368–375.
- [9] K. G. Kay, Semiclassical propagation for multidimensional systems by an initial value method, *J. Chem. Phys.* 101 (3) (1994) 2250–2260.
- [10] J. Shao, N. Makri, Forward- Backward Semiclassical Dynamics without Prefactors, *J. Phys. Chem. A* 103 (39) (1999) 7753–7756.
- [11] D. V. Shalashilin, M. S. Child, Multidimensional quantum propagation with the help of coupled coherent states, *J. Chem. Phys.* 115 (12) (2001) 5367–5375.
- [12] D. H. Zhang, E. Pollak, Coherent classical-path description of deep tunneling, *Phys. Rev. Lett.* 93 (14) (2004) 140401.
- [13] F. Grossmann, A semiclassical hybrid approach to many particle quantum dynamics, *J. Chem. Phys.* 125 (1).
- [14] J. Liu, W. H. Miller, Linearized semiclassical initial value time correlation functions using the thermal Gaussian approximation: Applications to condensed phase systems, *J. Chem. Phys.* 127 (2007) 114506.
- [15] E. M. Heatwole, O. V. Prezhdo, Analytic dynamics of the Morse oscillator derived by semiclassical closures, *J. Chem. Phys.* 130 (24) (2009) 244111.
- [16] R. Conte, A. Aspuru-Guzik, M. Ceotto, Reproducing Deep Tunneling Splittings, Resonances, and Quantum Frequencies in Vibrational Spectra From a Handful of Direct Ab Initio Semiclassical Trajectories, *J. Phys. Chem. Lett.* 4 (20) (2013) 3407–3412.

- [17] M. S. Church, S. V. Antipov, N. Ananth, Validating and implementing modified filinov phase filtration in semiclassical dynamics, *J. Chem. Phys.* 146 (23) (2017) 234104.
- [18] A. Patoz, T. Begusic, J. Vanicek, On-the-fly ab initio semiclassical evaluation of absorption spectra of polyatomic molecules beyond the condon approximation, *J. Phys. Chem. Lett.* 9 (9) (2018) 2367–2372.
- [19] W. H. Miller, An alternate derivation of the herman kluk (coherent state) semiclassical initial value representation of the time evolution operator, *Mol. Phys.* 100 (4) (2002) 397–400.
- [20] M. Buchholz, E. Fallacara, F. Gottwald, M. Ceotto, F. Grossmann, S. D. Ivanov, Herman-kluk propagator is free from zero-point energy leakage, *Chem. Phys.* 515 (2018) 231–235.
- [21] W. H. Miller, Classical-limit quantum mechanics and the theory of molecular collisions, *Adv. Chem. Phys.* 25 (1) (1974) 69–177.
- [22] M. Ceotto, S. Atahan, G. F. Tantardini, A. Aspuru-Guzik, Multiple coherent states for first-principles semiclassical initial value representation molecular dynamics, *J. Chem. Phys.* 130 (23) (2009) 234113.
- [23] M. Ceotto, D. Dell’ Angelo, G. F. Tantardini, Multiple coherent states semiclassical initial value representation spectra calculations of lateral interactions for CO on Cu (100), *J. Chem. Phys.* 133 (5) (2010) 054701.
- [24] M. Ceotto, S. Atahan, S. Shim, G. F. Tantardini, A. Aspuru-Guzik, First-principles semiclassical initial value representation molecular dynamics, *Phys. Chem. Chem. Phys.* 11 (2009) 3861–3867. doi:10.1039/B820785B.
- [25] F. Gabas, R. Conte, M. Ceotto, On-the-fly ab initio Semiclassical Calculation of Glycine Vibrational Spectrum, *J. Chem. Theory Comput.* 13 (2017) 2378.
- [26] X. Ma, G. Di Liberto, R. Conte, W. L. Hase, M. Ceotto, A quantum mechanical insight into sn2 reactions: Semiclassical initial value representation calculations of vibrational features of the cl-ch3cl pre-reaction complex with the venus suite of codes, *J. Chem. Phys.* 149 (16) (2018) 164113.

- [27] F. Gabas, G. Di Liberto, M. Ceotto, Vibrational investigation of nucleobases by means of divide and conquer semiclassical dynamics, *J. Chem. Phys.* 150 (22) (2019) 224107.
- [28] F. Gabas, G. Di Liberto, R. Conte, M. Ceotto, Protonated glycine supramolecular systems: the need for quantum dynamics, *Chem. Sci.* 9 (2018) 7894–7901.
- [29] M. Buchholz, F. Grossmann, M. Ceotto, Mixed semiclassical initial value representation time-averaging propagator for spectroscopic calculations, *J. Chem. Phys.* 144 (2016) 094102.
- [30] M. Buchholz, F. Grossmann, M. Ceotto, Application of the mixed time-averaging semiclassical initial value representation method to complex molecular spectra, *J. Chem. Phys.* 147 (16) (2017) 164110.
- [31] M. Buchholz, F. Grossmann, M. Ceotto, Simplified approach to the mixed time-averaging semiclassical initial value representation for the calculation of dense vibrational spectra, *J. Chem. Phys.* 148 (11) (2018) 114107.
- [32] G. Bertaina, G. Di Liberto, M. Ceotto, Reduced rovibrational coupling cartesian dynamics for semiclassical calculations: Application to the spectrum of the zundel cation, *J. Chem. Phys.* 151 (11) (2019) 114307. doi:10.1063/1.5114616.
- [33] M. Ceotto, G. Di Liberto, R. Conte, Semiclassical "Divide-and-Conquer" Method for Spectroscopic Calculations of High Dimensional Molecular Systems, *Phys. Rev. Lett.* 119 (1) (2017) 010401.
- [34] G. Di Liberto, R. Conte, M. Ceotto, "divide and conquer" semiclassical molecular dynamics: A practical method for spectroscopic calculations of high dimensional molecular systems, *J. Chem. Phys.* 148 (1) (2018) 014307.
- [35] G. Di Liberto, R. Conte, M. Ceotto, "divide-and-conquer" semiclassical molecular dynamics: An application to water clusters, *J. Chem. Phys.* 148 (10) (2018) 104302.

- [36] M. Micciarelli, R. Conte, J. Suarez, M. Ceotto, Anharmonic vibrational eigenfunctions and infrared spectra from semiclassical molecular dynamics, *J. Chem. Phys.* 149 (6) (2018) 064115.
- [37] M. Ceotto, S. Valleau, G. F. Tantardini, A. Aspuru-Guzik, First principles semiclassical calculations of vibrational eigenfunctions, *J. Chem. Phys.* 134 (23) (2011) 234103. doi:10.1063/1.3599469.
- [38] M. Micciarelli, F. Gabas, R. Conte, M. Ceotto, An effective semiclassical approach to ir spectroscopy, *J. Chem. Phys.* 150 (18) (2019) 184113.
- [39] R. Conte, L. Parma, C. Aieta, A. Rognoni, M. Ceotto, Improved semiclassical dynamics through adiabatic switching trajectory sampling, *J. Chem. Phys.* 151 (21) (2019) 214107.
- [40] M. Ceotto, Y. Zhuang, W. L. Hase, Accelerated direct semiclassical molecular dynamics using a compact finite difference Hessian scheme, *J. Chem. Phys.* 138 (5) (2013) 054116.
- [41] R. Conte, F. Gabas, G. Botti, Y. Zhuang, M. Ceotto, Semiclassical vibrational spectroscopy with hessian databases, *J. Chem. Phys.* 150 (24) (2019) 244118.
- [42] K. Karandashev, J. Vanicek, A combined on-the-fly/interpolation procedure for evaluating energy values needed in molecular simulations, *J. Chem. Phys.* 151 (17) (2019) 174116.
- [43] Y. Zhuang, M. R. Siebert, W. L. Hase, K. G. Kay, M. Ceotto, Evaluating the Accuracy of Hessian Approximations for Direct Dynamics Simulations, *J. Chem. Theory Comput.* 9 (1) (2012) 54–64.
- [44] D. Tamascelli, F. S. Dambrosio, R. Conte, M. Ceotto, Graphics processing units accelerated semiclassical initial value representation molecular dynamics, *J. Chem. Phys.* 140 (17) (2014) 174109.
- [45] R. Conte, P. L. Houston, J. M. Bowman, Communication: A benchmark-quality, full-dimensional ab initio potential energy surface for Ar-HOCO, *J. Chem. Phys.* 140 (15) (2014) 151101.
- [46] R. Conte, C. Qu, J. M. Bowman, Permutationally Invariant Fitting of Many-Body, Non-covalent Interactions with Application to Three-Body Methane-Water-Water, *J. Chem. Theory Comput.* 11 (2015) 1631–1638.

- [47] Z. Xie, J. M. Bowman, Permutationally invariant polynomial basis for molecular energy surface fitting via monomial symmetrization, *J. Chem. Theory Comput.* 6 (1) (2010) 26–34.
- [48] A. L. Kaledin, W. H. Miller, Time averaging the semiclassical initial value representation for the calculation of vibrational energy levels, *J. Chem. Phys.* 118 (16) (2003) 7174–7182.
- [49] A. L. Kaledin, W. H. Miller, Time averaging the semiclassical initial value representation for the calculation of vibrational energy levels. ii. application to h2co, nh3, ch4, ch2d2, *J. Chem. Phys.* 119 (6) (2003) 3078–3084.
- [50] G. Di Liberto, M. Ceotto, The Importance of the Pre-exponential Factor in Semiclassical Molecular Dynamics, *J. Chem. Phys.* 145 (2016) 144107.
- [51] M. L. Brewer, J. S. Hulme, D. E. Manolopoulos, Semiclassical dynamics in up to 15 coupled vibrational degrees of freedom, *J. Chem. Phys.* 106 (12) (1997) 4832–4839.
- [52] M. Ceotto, G. F. Tantardini, A. Aspuru-Guzik, Fighting the curse of dimensionality in first-principles semiclassical calculations: Non-local reference states for large number of dimensions, *J. Chem. Phys.* 135 (21) (2011) 214108. doi:10.1063/1.3664731.
- [53] X. Zhang, S. Zou, L. B. Harding, J. M. Bowman, A global ab initio potential energy surface for formaldehyde, *J. Phys. Chem. A* 108 (41) (2004) 8980–8986.
- [54] X. Wang, P. L. Houston, J. M. Bowman, A new (multi-reference configuration interaction) potential energy surface for h2co and preliminary studies of roaming, *Philos. Trans. R. Soc., A* 375 (2092) (2017) 20160194.
- [55] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, W. de Jong, NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations, *Comput. Phys. Commun.* 181 (9) (2010) 1477–1489.
- [56] C. Qu, Q. Yu, B. L. Van Hoozen, J. M. Bowman, R. A. Vargas-Hernandez, Assessing gaussian process regression and permutationally

invariant polynomial approaches to represent high-dimensional potential energy surfaces, *J. Chem. Theory Comput.* 14 (7) (2018) 3381–3396.

- [57] T. Nakanaga, S. Kondo, S. Saeki, Infrared band intensities of formaldehyde and formaldehyded2, *J. Chem. Phys.* 76 (8) (1982) 3860–3865.