# An Effective Semiclassical Approach to IR Spectroscopy

Marco Micciarelli,* Fabio Gabas, Riccardo Conte, and Michele Ceotto†

*Dipartimento di Chimica, Università degli Studi di Milano,*

*via C. Golgi 19, 20133 Milano, Italy.*

## Abstract

We present a novel approach to calculate molecular IR spectra based on semiclassical molecular dynamics. The main advance from a previous semiclassical method [M. Micciarelli, R. Conte, J. Suarez, M. Ceotto *J. Chem. Phys.* **149**, 064115 (2018)] consists in the possibility to avoid state-to-state calculations making applications to systems characterized by sizable densities of vibrational states feasible. Furthermore, this new method accounts not only for positions and intensities of the several absorption bands which make up the IR spectrum, but also for their shapes. We show that accurate semiclassical IR spectra including quantum effects and anharmonicities for both frequencies and intensities can be obtained starting from semiclassical power spectra. The approach is first tested against the water molecule, and then applied to the 10-atom glycine aminoacid.

———

* marco.micciarelli@unimi.it

† michele.ceotto@unimi.it

## I.  INTRODUCTION

Infrared (IR) spectroscopy is commonly employed for chemical characterization owing to some peculiar features which include its cheapness, portability to different environments, and generally short execution times. The main features of an IR spectrum, from which important information about molecular structure and interactions can be obtained, include the frequencies of vibrational transitions, the shapes of absorption bands, and their intensities. However, experimental measures can be of difficult interpretation. This is especially true when the IR spectrum consists of many overlapping absorption bands, which make it hard to assign fingerprint and stretch vibrations. It is in this setting that theoretical simulations of IR spectra may be crucial, allowing to decompose the spectrum into specific molecular motions. This is usually achieved first relying on the normal modes of vibration as a basic harmonic approximation, and then refining the theory by including anharmonicity effects.[? ? ]

The frequencies of vibrational transitions can be calculated reliably and accurately through Semiclassical (SC) molecular dynamics.[? ? ? ] In fact, a SC propagator is able to regain quantum effects from a classical Hamiltonian dynamics and several SC methods have been introduced to estimate quantum frequencies of vibration upon calculation of power spectra.[? ? ? ? ? ? ? ? ] In practice, a quantum reference state is chosen and the vibrational eigenenergies are obtained from the Fourier transform of its survival amplitude. It is then trivial to calculate the vibrational transition frequencies by difference with respect to the zero-point energy (ZPE).[? ? ? ]
Recent advances have permitted to get SC power spectra of systems characterized by many degrees of freedom. This has been achieved in our group by developing innovative methodologies like the multiple coherent (MC) and the divide-and-conquer (DC) semiclassical initial value representation (SCIVR). MC SCIVR is based on a tailored choice of reference state and dynamics initial conditions. In this way, accurate results are collected running just a few or even a single classical trajectory.[? ? ? ? ] Such a reduction in the needed computational effort has opened up the possibility to employ ab initio on-the-fly dynamics and to apply the semiclassical formalism to systems with many degrees of freedom.[? ? ? ? ] When dealing with high dimensional systems, though, it is not always possible to get a sensible spectroscopic signal with a full dimensional SC approach. In the case of a system-

bath model, one can employ a mixed semiclassical approach with an accurate semiclassical propagator for the system and a less accurate one for the bath.[? ? ?] However, in general, DC SCIVR has been introduced to overcome the curse of dimensionality issue.[? ] The technique, still based on full dimensional classical molecular dynamics, allows to compute semiclassical power spectra in reduced dimensionality within a set of appropriately chosen subspaces. The total spectrum is eventually obtained by collecting the low dimensional spectra calculated in the subspaces. Some representative applications of these techniques include fullerene,[? ] glycine,[? ] benzene,[? ] water clusters,[? ] the protonated glycine dimer and $H_2$-tagged protonated glycine.[? ]

Regarding band shapes, they arise naturally in dynamical approaches, like SC ones, from the Fourier transform and include the effect of any interactions experienced along the dynamics. This is different from common stick spectra in which a single central transition is representative of the whole absorption band whose shape is modeled by means of an *ad hoc* Lorentzian function.

Finally, the accurate estimate of spectral intensities is still a partially open issue and the missing tile for a complete semiclassical simulation of IR spectra. The straightforward approach to the problem deals with the calculation of the dipole autocorrelation, since its Fourier transform returns the correct estimates for both transition frequencies and intensities.[? ? ? ? ? ? ? ? ? ] Strictly speaking, though, the presence of two propagators leads semiclassically to a double phase-space integration, which makes the calculation very hard to converge due to the oscillations produced by the phase differences between trajectories with different initial conditions.

Several methods have been developed to try to overcome this issue. A possible strategy consists in easing the calculation by means of a filter able to damp the oscillations, hopefully without spoiling the results.[? ? ?] Another possibility is represented by the linearization approximation, which leads to a semiclassical expression based on a single phase space integration and formally equal to its classical counterpart, but with Wigner functions replacing classical ones.[? ? ? ? ?] This kind of approximation has the drawback that it is unable to account for quantum coherence between distinct trajectories and is affected by the zero point energy leakage problem, which is not the case for the original SCIVR formulation.[? ] A more accurate approach would be based on the rearrangement of the dipole autocorrelation in a suitable way for application of the forward-backward SC formula obtained by

stationary phase approximating the primitive SC expression. In this way a single phase space integration is needed, quantum coherences are included, and oscillations are quenched due to the evolution of trajectories first forward and then backward in time.[? ? ? ? ? ] In all cases, though, the challenge is to go beyond model systems and to be able to get semiclassical IR spectra even for molecules of sizable dimensionality, i.e. the goal already reached for power spectra.

For this purpose a SC methodology able to reproduce IR spectra has been introduced very recently.[? ] It is based on the calculation of oscillator strengths from vibrational eigenfunctions represented as linear combinations of harmonic functions. The coefficients of the linear combinations are obtained after collecting information from an appropriate set of SC power spectrum simulations in which harmonic states are employed as reference states. The approach takes advantage from the techniques developed for power spectra, and applications to systems with many degrees of freedom are in principle feasible. Results for the water molecule are indeed in excellent agreement with the quantum benchmark, but the method has the drawback to require calculation of contributions from all states involved in the transitions that make up the several absorption bands in the IR spectrum, which is unpractical for systems characterized by large densities of vibrational states.

The principal aim of the present paper is to take SC IR spectroscopy to the same level of applicability as SC power spectrum investigations. The goal is achieved by decomposing the IR spectrum into a sum of dynamical correlation functions, which can be calculated from SC power spectra and allows to avoid the undesired state-to-state computations. The paper outline is as follows: In Section II we detail the theory behind the new method; Section III is devoted to two representative applications. Specifically, application to water yields results of excellent accuracy, while application to the high energy fundamentals of glycine points out the importance of electrical and mechanical anharmonicity, and demonstrates the possibility to get reliable IR spectra for higher dimensional systems. A summary and some conclusions end the paper.

## II. THEORY

**Eigenvalues, eigenfunctions and transition dipoles** For a molecular system governed by the vibrational Hamiltonian operator

$$\hat{H} = \hat{T} + \hat{V} \tag{1}$$

with potential energy surface (PES) $V(\boldsymbol{R})$ given by the electronic Born-Oppenheimer adiabatic energy, we are interested in studying the spectral decomposition of $\hat{H}$ in terms of vibrational bound states beyond the harmonic approximation, i.e. in solving the eigenvalue problem

$$\hat{H} \left| e_n \right\rangle = E_n \left| e_n \right\rangle . \tag{2}$$

An analytical solution to Eq.(2) is in general not available, so numerical and approximate strategies are needed. For this purpose, the theoretical foundation of the formalism employed in this work lies on the adoption of the time propagation operator

$$\hat{\mathcal{P}}(t) = e^{-\frac{i}{\hbar}\hat{H}t} = \sum_n e^{-\frac{i}{\hbar}E_n t} \left| e_n \right\rangle \left\langle e_n \right| \tag{3}$$

to compute the time dependent survival amplitude of an arbitrary reference state $\left| \chi \right\rangle$

$$
\begin{aligned}
I_\chi(t) &\equiv \left\langle \chi \right| \hat{\mathcal{P}}(t) \left| \chi \right\rangle = \\
&= \sum_n e^{-\frac{i}{\hbar}E_n t} \left| \left\langle \chi | e_n \right\rangle \right|^2 ,
\end{aligned}
\tag{4}
$$

where the second equality is obtained upon introduction of the representation of the propagator in the basis of the Hamiltonian eigenvectors.

Eq. (4) implies that both squared projections $\left| \left\langle \chi | e_n \right\rangle \right|^2$ of the reference state onto the eigenvectors and eigenvalues $E_n$ can be determined respectively from peak amplitudes and positions of the following power spectrum

$$
\begin{aligned}
\tilde{I}_\chi(E) &= \frac{1}{2\pi\hbar} \int_{-\tau}^{\tau} dt I_\chi(t) e^{\frac{i}{\hbar}Et} = \\
&= \frac{1}{\pi\hbar} \, Re \left[ \int_0^{\tau} dt \, I_\chi(t) e^{\frac{i}{\hbar}Et} \right] = \\
&= \sum_n \left| \left\langle \chi | e_n \right\rangle \right|^2 \mathcal{D}(E - E_n; \Gamma_\tau).
\end{aligned}
\tag{5}
$$

The dynamical convolution function $\mathcal{D}$ is a delta-nascent function centered on $E_n$ with amplitude $\Gamma_\tau$ approaching zero as $\tau \to \infty$. It has the analytical form of a sinc function if the exact propagator is employed.

While eigenvalues are obtained straightforwardly, determination of eigenvectors requires a more elaborated strategy. In a recent paper[?] we have shown how vibrational eigenstates can be derived from SC power spectra. For a generic system with $N_v$ vibrational degrees of freedom, the starting point is the representation of the eigenvectors as a linear combination of $N_v$-dimensional harmonic states $|\phi_\mathbf{K}\rangle$

$$|e_n\rangle = \sum_{\mathbf{K}=1}^{N} C_{n,\mathbf{K}} |\phi_\mathbf{K}\rangle. \tag{6}$$

$N$ is the number of states in the basis set, $C_{n,\mathbf{K}} = \langle\phi_\mathbf{K}|e_n\rangle$ are real coefficients, and $|\phi_\mathbf{K}\rangle$ is the generic $\mathbf{K}$-th element of the complete and orthonormal N-dimensional basis set $\{|\phi_\mathbf{K}\rangle\}$. Each element $|\phi_\mathbf{K}\rangle$ of the basis set is obtained as the Hartree product of one-dimensional harmonic states

$$|\phi_\mathbf{K}\rangle = |\phi_{K_1}, \phi_{K_2}, \ldots, \phi_{K_{Nv}}\rangle =$$
$$= |\phi_{K_1}\rangle \ldots |\phi_{K_{Nv}}\rangle, \tag{7}$$

and

$$\phi_{K_\alpha}(Q_\alpha) = \langle Q_\alpha|\phi_{K_\alpha}\rangle = \frac{1}{\sqrt{2^{K_\alpha}K_\alpha!}} \left(\frac{\omega_\alpha}{\pi\hbar}\right)^{\frac{1}{4}} \times$$
$$e^{-\frac{\omega_\alpha Q_\alpha^2}{2\hbar}} h_{K_\alpha}\left(\sqrt{\frac{\omega_\alpha}{\hbar}} Q_\alpha\right), \tag{8}$$

where $Q_\alpha = q_\alpha - q_{eq,\alpha}$, $\omega_\alpha$, and $K_\alpha$ are, respectively, the $\alpha$-th normal mode coordinate displacement from equilibrium, frequency, and quantum number. $h_{K_\alpha}$ indicates the $K_\alpha$-th order Hermite polynomial in the variable $Q_\alpha$. The square modulus of coefficients $C_{n,\mathbf{K}}$ can be computed considering that they are proportional to the intensity, at the eigenvalue of the vibrational Hamiltonian, of the Fourier transform of the survival amplitude with reference harmonic state $\phi_\mathbf{K}$, i.e. $|C_{n,\mathbf{K}}|^2 \propto \tilde{I}_{\phi_\mathbf{K}}(E_n)$. As shown in details in our previous work,[?] the signed coefficients in Eq. (6) can be calculated from survival amplitudes using the following working formula

$$C_{n,\mathbf{K}} = \frac{\Delta \tilde{I}_{\phi_0, \phi_\mathbf{K}}(E_n)}{2\sqrt{\tilde{I}_{\phi_0}(E_n)}}, \tag{9}$$

where $\phi_0$ is the harmonic ground state, $\tilde{I}_{\phi_k}(E_n)$ is the value at energy $E_n$ of the power spectrum obtained with the harmonic state $|\phi_\mathbf{K}\rangle$, and

$$\Delta \tilde{I}_{\phi_{\mathbf{K}_1}, \phi_{\mathbf{K}_2}}(E) \equiv \tilde{I}_{\phi_{\mathbf{K}_1} + \phi_{\mathbf{K}_2}}(E) - \tilde{I}_{\phi_{\mathbf{K}_1}}(E) - \tilde{I}_{\phi_{\mathbf{K}_2}}(E). \tag{10}$$

The same coefficients can be exploited to evaluate the transition dipole between an initial state $|e_n\rangle$ and a final state $|e_m\rangle$

$$\langle e_n | \hat{\mu}_{0N}^\epsilon | e_m \rangle = \sum_{\mathbf{K}, \mathbf{K}'} C_{n,\mathbf{K}} C_{m,\mathbf{K}'} \langle \phi_\mathbf{K} | \hat{\mu}_{0N}^\epsilon | \phi_{\mathbf{K}'} \rangle, \tag{11}$$

which is needed for calculating the absorption spectrum. In Eq.(11) $\epsilon = x, y, z$ and $\hat{\boldsymbol{\mu}}_{0N}(\mathbf{R}) = \hat{\boldsymbol{\mu}}_N(\mathbf{R}) + \hat{\boldsymbol{\mu}}_{e0}(\mathbf{R})$ is the molecular dipole made of two contributions: $\boldsymbol{\mu}_N(\mathbf{R}) = \sum_i Z_i \mathbf{R}_i$ is the nuclear part; $\boldsymbol{\mu}_{e0}(\boldsymbol{R}) = \int d\mathbf{r} \, |\varphi_0(\mathbf{r}; \mathbf{R})|^2 \boldsymbol{\mu}_e(\mathbf{r})$ is the electronic dipole with $\varphi_0(\mathbf{r}; \mathbf{R})$ representing the adiabatic electronic ground state wavefunction for a given nuclear configuration. Calculation of the nuclear contribution is trivial, while the electronic one requires a Monte Carlo estimate.[?] The latter can be avoided by means of the widely employed linearization approximation to the dipole

$$\boldsymbol{\mu}_{0N}(\mathbf{q}) - \boldsymbol{\mu}_{0N}(\mathbf{q}_{eq}) \simeq \sum_{\alpha=1}^{N_v} \left. \frac{\partial \boldsymbol{\mu}_{0N}}{\partial q_\alpha} \right|_{\mathbf{q}_{eq}} (q_\alpha - q_{eq,\alpha}). \tag{12}$$

Eq.(11) then can be rearranged as

$$\langle e_n | \hat{\mu}_{0N}^\epsilon | e_m \rangle = \sum_{\mathbf{K}, \mathbf{K}'} C_{n,\mathbf{K}} C_{m,\mathbf{K}'} \sum_\alpha Z_{\epsilon,\alpha} \langle \phi_{K_\alpha} | \hat{Q}_\alpha | \phi_{K'_\alpha} \rangle, \tag{13}$$

where $Z_{\epsilon,\alpha} = \left. \frac{\partial \mu_{0N}^\epsilon}{\partial q_\alpha} \right|_{\mathbf{q}_{eq}}$ is a quantity easy to compute and routinely returned by the most popular electronic structure softwares. Eq.(13) represents the fully anharmonic estimate of transition dipoles within the linearized dipole approximation. Other, more approximate

formulae can be adopted to ease the calculation. For instance, the initial state $|e_n\rangle$ can be approximated by a single harmonic state $|\phi_{\bar{\mathbf{K}}}\rangle$, i.e. $|e_n\rangle \simeq \sum_{\mathbf{K}} \delta_{\mathbf{K}\bar{\mathbf{K}}} |\phi_{\mathbf{K}}\rangle$. This leads to

$$\langle e_n|\hat{\mu}_{0N}^\epsilon|e_m\rangle = \sum_{\mathbf{K}'} C_{m,\mathbf{K}'} \sum_\alpha Z_{\epsilon,\alpha} \langle\phi_{\bar{K}_\alpha}| \hat{Q}_\alpha |\phi_{K'_\alpha}\rangle . \tag{14}$$

We refer to Eq.(14) as the semi-anharmonic transition dipole. Finally, the harmonic approximation can be invoked also for the final state, so that $|e_m\rangle \simeq |\phi_{\bar{\mathbf{K}}'}\rangle$ and the transition dipole takes the simpler form

$$\langle e_n|\hat{\mu}_{0N}^\epsilon|e_m\rangle = \sum_\alpha Z_{\epsilon,\alpha} \langle\phi_{\bar{K}_\alpha}| \hat{Q}_\alpha |\phi_{\bar{K}'_\alpha}\rangle . \tag{15}$$

This is the case commonly known as the harmonic electrical approximation, which is often coupled in basic spectroscopy calculations to its mechanical counterpart (i.e. the harmonic estimate of frequencies) under the collective name of double harmonic approximation.

A matrix element like the one in Eq.(15) can be readily evaluated. In fact

$$\langle\phi_{\mathbf{K}}| \hat{Q}_\alpha |\phi_{\mathbf{K}'}\rangle = \left( \prod_{\beta \neq \alpha}^{N_v} \delta_{K_\beta, K'_\beta} \right) \sqrt{\frac{1}{2\omega_\alpha}} \times$$
$$\left( \delta_{K_\alpha, K'_\alpha+1}\sqrt{K'_\alpha + 1} + \delta_{K_\alpha, K'_\alpha-1}\sqrt{K'_\alpha} \right) , \tag{16}$$

which is obtained starting from $\hat{Q}_\alpha = \sqrt{\frac{1}{2\omega_\alpha}} \left( \hat{a}_\alpha^\dagger + \hat{a}_\alpha \right)$, with $\hat{a}_\alpha^\dagger$ and $\hat{a}_\alpha$ being the harmonic oscillator creation and annihilation operators, respectively, for normal mode $\alpha$. Eq.(16) incorporates the harmonic selection rules, which permit to simplify Eq.(13) by neglecting many zero-valued terms

$$\langle e_n|\hat{\mu}_{0N}^\epsilon|e_m\rangle = \mu_{nm}^\epsilon = \sum_{\mathbf{K}} C_{n,\mathbf{K}} \sum_\alpha C_{m,\mathbf{K}^{(\alpha)}} \mu_{\epsilon,\mathbf{K}^{(\alpha)}}. \tag{17}$$

The second sum in Eq.(13) is restricted in Eq.(17) to the $2N_v$ basis set elements obtained by exciting or de-exciting the $\alpha$-th degree of freedom, i.e. the states $|\phi_{\mathbf{K}^{(\alpha)}}\rangle$ such that $|\phi_{\mathbf{K}^{(\alpha)}}\rangle \propto \hat{a}_\alpha^\dagger |\phi_{\mathbf{K}}\rangle$ or $|\phi_{\mathbf{K}^{(\alpha)}}\rangle \propto \hat{a}_\alpha |\phi_{\mathbf{K}}\rangle$. Furthermore

$$
\mu_{\epsilon,\mathbf{K}^{(\alpha)}} = \begin{cases} Z_{\epsilon,\alpha}\sqrt{\frac{K_\alpha+1}{2\omega_\alpha}} & \text{if } |\phi_{\mathbf{K}^{(\alpha)}}\rangle \propto \hat{a}_\alpha^\dagger |\phi_{\mathbf{K}}\rangle \\[2mm] Z_{\epsilon,\alpha}\sqrt{\frac{K_\alpha}{2\omega_\alpha}} & \text{if } |\phi_{\mathbf{K}^{(\alpha)}}\rangle \propto \hat{a}_\alpha |\phi_{\mathbf{K}}\rangle \quad \text{and} \quad \hat{a}_\alpha |\phi_{\mathbf{K}}\rangle \neq 0 \\[2mm] 0 & \text{if } |\phi_{\mathbf{K}^{(\alpha)}}\rangle \propto \hat{a}_\alpha |\phi_{\mathbf{K}}\rangle \quad \text{and} \quad \hat{a}_\alpha |\phi_{\mathbf{K}}\rangle = 0 \end{cases} \tag{18}
$$

**The IR spectrum from power spectra** Our target is the calculation of the following spectrum

$$
S_{\epsilon,n}(\omega) = \sum_{m \neq n} |\mu_{nm}^\epsilon|^2 \, \omega \, \mathcal{D}(\omega - E_m + E_n; \Gamma_\tau), \tag{19}
$$

in which the sum runs over all the spectral lines from the $n$-th to the $m$-th vibrational states. This is an approximate expression for the IR absorption of a system initially lying on the $n$-th pure Hamiltonian eigenstate and perturbed by a radiation of frequency $\omega$ polarized along the $\epsilon$ direction, which takes the energy of the system to $E = \omega + E_n$. The spectral lines are broadened over a finite range of energies by means of the dynamical convolution function $\mathcal{D}$. Details on this approach and derivation of Eq.(19) are presented in Appendix A.

It is convenient to start by considering the following IR spectrum in which the transition dipoles are treated by means of the semi-anharmonic approximation (see Eq.(14))

$$
S_{\epsilon,\phi_{\bar{\mathbf{K}}}}(\omega) = \sum_m |\mu_{\bar{\mathbf{K}}m}^\epsilon|^2 \, \omega \, \mathcal{D}(\omega - E_m + E_n; \Gamma_\tau), \tag{20}
$$

where

$$
\mu_{\bar{\mathbf{K}}m}^\epsilon = \langle \phi_{\bar{\mathbf{K}}} | \, \hat{\mu}_{0N}^\epsilon \, | e_m \rangle. \tag{21}
$$

Inserting Eq.(17) (with $C_{n\bar{\mathbf{K}}} = \delta_{n,\bar{\mathbf{K}}}$) into Eq.(20) and expanding the square in the sum, we get

$$
S_{\epsilon,\phi_{\bar{\mathbf{K}}}}(\omega) = \sum_m \left( \sum_\alpha C_{m,\bar{\mathbf{K}}^{(\alpha)}}^2 (\mu_{\bar{\mathbf{K}}^{(\alpha)}}^\epsilon)^2 + \right.
$$
$$
\left. +2\sum_{\beta<\alpha} C_{m,\bar{\mathbf{K}}^{(\alpha)}} C_{m,\bar{\mathbf{K}}^{(\beta)}} \mu_{\bar{\mathbf{K}}^{(\alpha)}}^\epsilon \mu_{\bar{\mathbf{K}}^{(\beta)}}^\epsilon \right) \omega \, \mathcal{D}(\omega - E_m + E_n; \Gamma_\tau), \tag{22}
$$

where the indexes $\alpha$ and $\beta$ run over the normal modes. Proper combination of Eq.(5), Eq.(6), and Eq.(10) permits to rearrange Eq.(22) in terms of power spectra only. In fact, by means of

$$\tilde{I}_{\phi_{\boldsymbol{K}}}(E) = \sum_m |C_{m,\boldsymbol{K}}|^2 \mathcal{D}(E - E_m; \Gamma_\tau)$$

$$\Delta\tilde{I}_{\phi_{\mathbf{K}_1}, \phi_{\mathbf{K}_2}}(E) = \sum_m 2 C_{m,\boldsymbol{K}_1} C_{m,\boldsymbol{K}_2} \mathcal{D}(E - E_m; \Gamma_\tau), \tag{23}$$

and inverting in Eq.(22) the sum over $m$ (that we do not want to compute) and the sum over $\alpha$ (that we want to keep explicit), after some straightforward algebra we get

$$S_{\epsilon,\phi_{\bar{\boldsymbol{K}}}}(\omega) = \omega \sum_\alpha \left[ \mu^2_{\bar{\mathbf{K}}(\alpha)} \tilde{I}_{\phi_{\bar{\boldsymbol{K}}(\alpha)}}(\omega + E_n) + \sum_{\beta<\alpha} \mu_{\bar{\mathbf{K}}(\alpha)} \mu_{\bar{\mathbf{K}}(\beta)} \Delta\tilde{I}_{\phi_{\bar{\boldsymbol{K}}(\alpha)}, \phi_{\bar{\boldsymbol{K}}(\beta)}}(\omega + E_n) \right]. \tag{24}$$

It is worth noting that in Eq.(24) the dependence of the power spectra on the radiation frequency has been explicitly indicated, as derived from the anticipated key relation $E = \omega + E_n$.

Eventually, the (fully anharmonic) IR spectrum defined in Eq.(19) can be obtained (see Appendix B for derivation details) as

$$S_{\epsilon,n}(\omega) = \sum_{\mathbf{K}} C^2_{n,\mathbf{K}} S_{\epsilon,\phi_{\mathbf{K}}}(\omega) +$$

$$+ \omega \sum_{\mathbf{K}'<\mathbf{K}} \sum_{\alpha,\alpha'} C_{n,\mathbf{K}} C_{n,\mathbf{K}'} \Delta\tilde{I}_{\phi_{\mathbf{K}(\alpha)}, \phi_{\mathbf{K}'(\alpha')}}(\omega + E_n) \mu^\epsilon_{\mathbf{K}(\alpha)} \mu^\epsilon_{\mathbf{K}'(\alpha')}, \tag{25}$$

where in the sum over $\mathbf{K}' < \mathbf{K}$ the elements of the basis set are sorted in some arbitrary way. Eq.(24) and Eq.(25) demonstrate that a state-to-state computation is not required, and that, if power spectra are calculated exactly, it is only the fully anharmonic estimate that demands for coefficients (i.e. for knowledge of the eigenfunction of the initial state).

**Semiclassical power spectra** In semiclassical dynamics the quantum propagator is usually approximated by means of the Herman-Kluk (HK) expression

$$\hat{\mathcal{P}}(t) \propto \int\int d\mathbf{Q}_0 d\mathbf{p}_0 \, C_t(\mathbf{Q}_0, \mathbf{p}_0) e^{\frac{i}{\hbar}S_t(\mathbf{Q}_0, \mathbf{p}_0)} |\mathbf{Q}_t, \mathbf{p}_t\rangle \langle\mathbf{Q}_0, \mathbf{p}_0|, \tag{26}$$

where $\mathbf{Q}_t$ and $\mathbf{p}_t$ are the classical normal mode displacement and momentum vectors at time t, obtained from the classical propagation of the trajectory started at $(\mathbf{Q}_0, \mathbf{p}_0)$ under the classical vibrational Hamiltonian. $|\mathbf{Q}_t, \mathbf{p}_t\rangle$ are coherent states of the form

$$\langle \mathbf{x}|\mathbf{Q}_t, \mathbf{p}_t \rangle = \left( \frac{det(\boldsymbol{\gamma})}{\pi} \right)^{\frac{N_v}{4}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{Q}_t)^T \boldsymbol{\gamma}(\mathbf{x}-\mathbf{Q}_t) + \frac{i}{\hbar} \mathbf{p}_t^T (\mathbf{x}-\mathbf{Q}_t)}, \tag{27}$$

where $\boldsymbol{\gamma}$ is a $N_v \times N_v$ diagonal matrix with diagonal elements equal to the harmonic frequencies $\{\omega_\lambda\}_{\lambda=1}^{N_v}$; $S_t$ is the classical action at time $t$ computed along the trajectory, and, finally, $C_t(\mathbf{Q}_0, \mathbf{p}_0)$ is the HK prefactor at time $t$ that accounts for second order quantum fluctuations around each classical path and which is defined as

$$C_t(\mathbf{Q}_0, \mathbf{p}_0) = \sqrt{\frac{1}{2^{N_v}} \left| \frac{\partial \mathbf{Q}_t}{\partial \mathbf{Q}_0} + \boldsymbol{\gamma}^{-1} \frac{\partial \mathbf{p}_t}{\partial \mathbf{p}_0} \boldsymbol{\gamma} - i\hbar \frac{\partial \mathbf{Q}_t}{\partial \mathbf{p}_0} \boldsymbol{\gamma} + \frac{i\boldsymbol{\gamma}^{-1}}{\hbar} \frac{\partial \mathbf{p}_t}{\partial \mathbf{Q}_0} \right|}. \tag{28}$$

The multi-dimensional integral over the initial phase space conditions is usually performed by means of Monte Carlo techniques, and the method has been applied successfully in many instances, yielding accurate results.[? ? ? ? ? ] However, the number of different classical trajectories to run is often prohibitively high for an effective interface to ab initio on-the-fly evaluations of energies and gradients.

The computational cost required for this kind of simulations can be much decreased by employing Kaledin and Miller's time average filter with separable approximation to the prefactor.[? ] This approximation consists in imposing that the amplitude of the HK prefactor is constant in time, i.e.

$$C_t(\mathbf{Q}_0, \mathbf{p}_0) \simeq e^{i\phi_t(\mathbf{Q}_0, \mathbf{p}_0)}, \tag{29}$$

a condition that is exactly fulfilled in the case of the harmonic oscillator. In this way, the TA SCIVR power spectrum is

$$\tilde{I}_\chi(E) \propto \iint d\mathbf{Q}_0 d\mathbf{p}_0 \frac{1}{\tau} \left| \int_0^\tau dt \langle \chi|\mathbf{Q}_t, \mathbf{p}_t \rangle e^{i[S_t(\mathbf{Q}_0, \mathbf{p}_0) + \phi_t(\mathbf{Q}_0, \mathbf{p}_0) + Et]/\hbar} \right|^2 \tag{30}$$

Calculation of the time averaged power spectrum still requires to perform a multidimensional integration but yields converged results orders of magnitude faster with a loss in accuracy

in peak positions of just a few cm$^{-1}$. Notwithstanding, the computational overhead of ab initio on-the-fly simulations demands for a further reduction of the number of trajectories to be run.

This goal has been reached by means of the MC-SCIVR approach in which rather than relying on a full Monte Carlo sampling of the phase space, the SC time propagator is built using only a handful of tailored classical trajectories. The trajectories are ideally selected according to the Einstein-Brillouin-Keller (EBK) quantization rules

$$\oint_{H(\mathbf{Q}_t^{(n)},\mathbf{P}_t^{(n)})=E_n} p_i^{(n)} dq_i^{(n)} = \hbar\left(\nu_i + \frac{a_i}{2} + \frac{b_i}{4}\right), \tag{31}$$

where $\nu_i$ are positive integers, while $a_i$ and $b_i$ are Maslov indexes. Indeed the EBK quantization condition is exact for the harmonic oscillator with Maslov indexes $a_i = 1$ and $b_i = 0$. In this particular case

$$\frac{1}{2}\left(p_i^{(n)}(t)\right)^2 + \frac{1}{2}\omega_i^2\left(Q_i^{(n)}(t)\right)^2 = \left(\frac{1}{2} + \nu_i\right)\hbar\omega_i, \tag{32}$$

so that the classical trajectories have total energies (and energy partition) corresponding to the harmonic oscillator spectral energies

$$\mathcal{E}_{\boldsymbol{\nu}}^{HO} = \sum_i \left(\frac{1}{2} + \nu_i\right)\hbar\omega_i. \tag{33}$$

The EBK quantization conditions are still exact when the PES anharmonicity preserves the generalized periodicity of the motion (i.e. when each mode performs a periodic motion). In fact, Eq.(32) and Eq.(33) become

$$\sum_i \frac{1}{2}\left(p_i^{(n)}(t)\right)^2 + V\left(\boldsymbol{Q}^{(n)}(t)\right) = \sum_i \left(\frac{1}{2} + \nu_i\right)\hbar\,\Xi_i(E)$$

$$E_{n(\boldsymbol{\nu})} = \sum_i \left(\frac{1}{2} + \nu_i\right)\hbar\,\Xi_i(E_{n(\boldsymbol{\nu})}), \tag{34}$$

where $\Xi_i(E)$ are the classical frequencies of the generalized periodic dynamics, which depend on the energy since the PES is not harmonic.

The trajectories employed in SC dynamics require a short time evolution (1 ps or less) without any preliminary equilibration. Therefore, the harmonic EBK quantization conditions of Eq.(32) can be easily fulfilled at the initial step by setting

$$Q_i^n(0) = 0$$
$$p_i^n(0) = \sqrt{(2\nu_i + 1)\hbar\omega_i}. \tag{35}$$

These conditions still realize a good approximation to the proper EBK quantization condition of Eq.(34) if the energy dependence of the anharmonic frequencies is moderate enough, i.e. if $\Xi_i(E_{n(\boldsymbol{\nu})}) \simeq \Xi_i(\mathcal{E}_{\boldsymbol{\nu}}^h)$. The MC-SCIVR power spectrum is eventually computed as

$$\tilde{I}_{\boldsymbol{\nu},\chi}(E) \propto Re\left[\int_0^\tau dt \, \langle\chi|\hat{\mathcal{P}}_{n(\boldsymbol{\nu})}^{MC}(t)|\chi\rangle \, e^{\frac{i}{\hbar}Et}\right] =$$
$$= \frac{1}{\tau}\left|\int_0^\tau dt \, \langle\chi|\mathbf{Q}_t^n, \mathbf{p}_t^n\rangle \, e^{i[S_t(\mathbf{Q}_0^n, \mathbf{p}_0^n) + \phi_t(\mathbf{Q}_0^n, \mathbf{p}_0^n) + Et]/\hbar}\right|^2. \tag{36}$$

In calculations where $|\chi\rangle = |\mathbf{K}\rangle$ the term $\langle\chi|\mathbf{Q}_t^n, \mathbf{p}_t^n\rangle$ is analytical.[?  ] We have indicated with $n(\boldsymbol{\nu})$ the $n$-th Hamiltonian eigenvalue corresponding to the $\boldsymbol{\nu}$ vector of integers via Eq.(35). The label $n(\boldsymbol{\nu})$ has been added explicitly to the MC propagator to indicate that it gives a reliable approximation to the exact propagator only in the region of the energy spectrum close to $E_n$, i.e.

$$\hat{\mathcal{P}}_{n(\boldsymbol{\nu})}^{MC}(t) = \sum_m W_{n,m}^2 \, e^{-\frac{i}{\hbar}E_m^{(\mathcal{P}_n)}t} \, |e_m^{(\mathcal{P}_n)}\rangle \, \langle e_m^{(\mathcal{P}_n)}| \tag{37}$$

with

$$E_n^{(\mathcal{P}_n)} \simeq E_n \tag{38}$$

and

$$W_n^2 \, |e_n^{(\mathcal{P}_n)}\rangle \simeq |e_n\rangle, \tag{39}$$

where the constants $W_n^2 = W_{n,n}^2$ account for both the potential loss of amplitude due to the separable approximation and the fact that, having used a single trajectory, the amplitudes of the different states are not converged uniformly.

In practice, the initial EBK conditions are chosen according to the state to be investigated, and each different trajectory is used to build a different propagator specialized in the energy range around the eigenenergy of the target state. The MC-SCIVR approach has been numerically tested in several applications returning accurate eigenvalues and eigenvectors.[? ? ] The typical strategy consists in obtaining a first estimate of the SC frequencies of the fundamental transitions by means of the ground state propagator. Calculations are then refined by employing a different and tailored MC-SCIVR propagator (i.e. a different trajectory) for each state. Using this approach, hence, the full SC power spectrum can be composed piece after piece as a collection of different single-trajectory propagators. In particular, for a given state $|\chi\rangle$, the MC-SCIVR power spectrum obtained after a run of classical molecular dynamics of length $\tau$ can be written as

$$\tilde{I}_\chi(E \sim E_n) \simeq \frac{1}{W_n^2}\tilde{I}_{\boldsymbol{\nu},\chi}(E \sim E_n) \simeq |\langle\chi|e_n\rangle|^2 \mathcal{D}(E - E_n; \Gamma_\tau), \tag{40}$$

where, as already anticipated, the broadening functions $\mathcal{D}$ have the shape of squared sync functions. The constants $W_n^2$ can be eventually derived enforcing the normalization of the eigenstates, i.e.

$$1 = \langle e_n|e_n\rangle = \sum_{\mathbf{K}} |C_{n,\mathbf{K}}|^2. \tag{41}$$

In fact, by selecting in Eq.(40) $|\chi\rangle = |\phi_{\mathbf{K}}\rangle$, $E = E_n$, and $\mathcal{D}(0; \Gamma_\tau) = 1$ (see Appendix A for a justification for this choice)

$$W_n^2 = \sum_{\mathbf{K}} \tilde{I}_{\boldsymbol{\nu},\phi_{\mathbf{K}}}(E = E_n), \tag{42}$$

which means that the constant for the $n$-th state can be calculated as the sum of the (non-negligible) intensities at energy $E_n$ of all power spectra obtained using the harmonic states of the basis set as reference states. For systems with sizable densities of states, many different states and transitions may contribute to power spectra and absorption bands. To avoid

calculating all needed normalization constants, they can be considered to be the same for all vibrational states in the confidence energy range of each propagator. This is justified by some tests which show that only a very mild discrepancy ($\sim 1\%$) in the value of normalization constants comes from a change in the reference energy position within the same confidence energy window. This approximation is instead no longer valid when different energy ranges and/or different propagators are taken into consideration.

## III.   RESULTS AND DISCUSSION

**H$_2$O molecule** The first test application we propose concerns the non-rotating water molecule in vacuum. MC SCIVR was already applied to this system in our previous study[? ] and results for both eigenenergies and eigenstates were in excellent agreement with reference calculations performed using the Grid Time-Dependent Schrödinger Equation (GTDSE) computational package[? ]. From that work we borrowed the same, pre-existing analytical H$_2$O PES[? ], and the same pre-existing dipole surface.[? ]

The initial step of any SC approach consists in providing a harmonic estimate of vibrational frequencies. To this end, the Hessian matrix at the equilibrium geometry has been diagonalized to get the three harmonic frequencies of vibration, which are related to the symmetric stretch ($\omega_s = 3831$ cm$^{-1}$), the bending ($\omega_b = 1650$ cm$^{-1}$), and the asymmetric stretch ($\omega_a = 3941$ cm$^{-1}$) motions. Consistently with the MC-SCIVR methodology presented above, to investigate the 5 lowest-lying vibrational states we selected the appropriate harmonic EBK initial conditions and then generated five classical trajectories to build five MC-SCIVR propagators. The trajectories were associated to the following triplets of harmonic quantum numbers (in increasing order of energy): $\boldsymbol{\nu}_{(n=1,...,5)} = \{(0,0,0); (0,1,0); (0,2,0); (1,0,0); (0,0,1)\}$. Each trajectory was propagated for a total of 1.2 ps with Hessians calculated at each step along the dynamics to evaluate the time evolution of the Herman-Kluk prefactor (specifically its phase). We then applied Eq.(36) to get 5 distinct MC-SCIVR power spectra $\tilde{I}_{\boldsymbol{\nu}_n,\phi_{\boldsymbol{\nu}}}(E)$. The final, total power spectrum has been obtained as a direct sum of the 5 single power spectra.

In panel (a) of Fig. 1 we report the total MC-SCIVR power spectrum. As discussed above, the Hamiltonian eigenenergies correspond to the positions of the different peaks. In panel (b) of the same Figure we present the total power spectrum obtained from the

5 normalized power spectra, i.e. $\frac{1}{W_n^2}\tilde{I}_{\boldsymbol{\nu}_n,\phi_{\boldsymbol{\nu}}}(E)$. The constants $W_n$ have been obtained from Eq.(42) by employing a truncated harmonic basis set made of all possible harmonic functions with quantum numbers not greater than 10, i.e. the total number of harmonic states in the basis set was $11^3 = 1331$. The different intensities of peaks between panel (a) and (b) in Fig. 1 point out the importance of the normalization factors $W_n^2$ for this new SC approach very clearly. It is also worth noting that the different peak intensities in panel (b) of Fig. 1 are due to anharmonicity effects, which are related to the squared projections $|\langle\phi_{\boldsymbol{\nu}_n}|e_n\rangle|^2$. However, as observed in our previous work,[?] the main contribution to a generic anharmonic state $|e_n\rangle$ is given by its purely harmonic counterpart. Therefore all peak intensities are indeed close to 1, since most of the character of each anharmonic state is given by its harmonic associate. This is particularly true for the ground and first bunch of excited states, while anharmonicity increases in the excitation of bond stretches (corresponding to the two peaks at the highest energies in Figure 1). For the latter, in fact, the normalized peak intensities become smaller. The effect is larger for the symmetric stretch (blue curve) than for the asymmetric one (green line). The reason is that a Fermi resonance between the symmetric stretch $|100\rangle$ and the bending overtone $|020\rangle$ is present.

In panel (c) of Fig. 1 we show the fully anharmonic semiclassical IR spectrum of water under the effect of unpolarized light $(S_0(E) = \sum_{\epsilon=x,y,z} S_{\epsilon,0}(E))$ obtained using the ground state as a reference state. This is, hence, the IR absorption spectrum at temperature $T = 0$ K. The ground state eigenfunction was expanded in terms of the harmonic states already employed in our previous work,[?] and the related coefficients employed in Eq.(25). We note that the intensity of the bending transition, located at 1587 cm$^{-1}$, is correctly almost twice as intense as the two stretching ones at 3707 cm$^{-1}$ and 3811 cm$^{-1}$ respectively. Furthermore, the bending overtone transition, estimated at 3162 cm$^{-1}$, is very weak but not exactly 0 because of the anharmonicity of the PES.

The exact anharmonic intensities of these absorption peaks have been derived in our previous work by means of a DVR approach.[?] They are reported as vertical black lines in Fig. 2 for the bright vibrational fundamental transitions of water and compared to the peaks of the SC IR spectrum. The agreement is excellent and all three absorption intensities are perfectly reproduced within a negligible error due to the dipole linearization and/or the semiclassical approximation. Fig. 2 also points out the enhanced accuracy of fully anharmonic IR spectra with respect to the semi-anharmonic ones reported in dashed

lines.

These results validate the proposed approach and show that it is equivalent to the direct state-to-state calculation of the oscillator strengths (i.e. the square moduli of transition dipoles). However, the decrease in computational overhead is evident already at this low dimensionality. In fact, the IR spectrum reported above asks only for the dipole derivative with respect to nuclear displacements at the equilibrium geometry in addition to what is needed to get a SC power spectrum. Any subsequent dipole evaluation is not required.

**Glycine** We move to the 10-atom glycine molecule in its neutral form in gas phase. Being the smallest among all aminoacids, this molecule has both a great biological relevance and a manageable size, so several theoretical methods have been applied to calculate its vibrational spectrum beyond the harmonic approximation.[? ? ? ? ? ? ] In a recent work,[? ] MC-SCIVR power spectra have been calculated for the four main conformers of glycine using on-the-fly ab initio molecular dynamics. Semiclassical energies are in very good agreement with other theoretical calculations as well as experimental data.[? ] Here we extend the previous study by evaluating intensities and absorption bands for the high-energy fundamental transitions, i.e. the $CH_2$ and $NH_2$ stretches and the OH vibration, of the global minimum conformer (Conf I). This spectroscopic region is of great interest and the key target of investigation in bigger aggregates because it is influenced by the hydrogen bonding responsible for the structural stability of the complexes.[? ]

We performed ab initio on-the-fly molecular dynamics runs at DFT-B3LYP level of theory with aug-cc-pVDZ basis set using the NWChem[? ] suite of codes. The structure obtained for the global minimum conformer is reported in ball-and-stick representation in Fig. 3. The molecular dipole derivatives, which are necessary for calculating the harmonic transition dipoles, the set of 24 normal mode coordinates and the harmonic spectrum have been computed at this molecular geometry with the same level of theory and basis set. Starting from the double harmonic approximation thus obtained, as anticipated, we focused on the anharmonic corrections to the highest energetic fundamentals in the energy range between 3000 cm$^{-1}$ and 4000 cm$^{-1}$, for which harmonic estimates are known to be inaccurate.

In Fig. 4 we report the five unpolarized light semi-anharmonic IR spectra $S_{\phi_0}(\omega)$ obtained using five different EBK trajectories. For each trajectory, the initial conditions were determined by means of the harmonic EBK quantization rule with vectors $\boldsymbol{\nu}$ obtained giving one quantum of excitation to modes 20-24 as indicated in the legend of the different figure

panels. All trajectories have been evolved for 5000 time steps ($dt = 10$ $a.u.$) giving a total evolution time of $\sim 1.2$ $ps$. It is worth mentioning that, according to Eq.(24), for glycine the total number of power spectra to be computed in order to get the semi-anharmonic IR spectrum $S_{\phi_0}(\omega)$ would add up to $N_v(N_v + 1)/2 = 300$ for each trajectory (i.e. for each SC propagator). However, for many of the 24 single-excited harmonic reference states the power spectra $\tilde{I}_{\phi_{\mathbf{0}(\alpha)}}$ appearing in Eq.(24) give no contribution in the energy range of interest. Furthermore, in these cases the $\Delta \tilde{I}_{\phi_{\mathbf{0}(\alpha)}, \phi_{\mathbf{0}(\beta)}}$ terms vanish for all $\beta$, thus decreasing substantially the total number of power spectra to be evaluated. In practice, the sole contributors to the high energy bands investigated are the six modes from 19 to 24, decreasing the total number of power spectra calculated to just 21.

Interestingly, energy shifts of the order of 100 cm$^{-1}$ and variations in band shapes are observed among different spectra, but, consistently with the MC-SCIVR recipe, we consider each IR spectrum reliable only within a given confidence energy range. These ranges are selected in a way that they contain the absorption band located where vibrational states have a significant component on the harmonic state corresponding to the EBK trajectory adopted. The energy windows are revealed by considering the (principal) band of the survival probability power spectrum $\tilde{I}_{\phi_{\mathbf{K}}}(\omega)$, with $\mathbf{K} = \boldsymbol{\nu}$ (reported with black dashed lines in each panel) and have been highlighted in Fig. 4. As it can be seen from panel (a) and (b) of Fig. 4, the confidence regions relative to modes 20 and 21 (corresponding to the symmetric and asymmetric CH$_2$ stretches) coincide. Furthermore, the two IR spectra obtained are equivalent within method accuracy, and hence any of the two is representative of the IR spectrum in that specific energy region, i.e. it is the result of the sum of contributions from both CH$_2$ absorption bands. A different situation applies in the case of the two NH$_2$ stretches, whose confidence regions and absorption bands are not equal. Specifically, data shown in panel (c) and (d) are obtained from the EBK trajectories for the symmetric and asymmetric stretch, respectively. From the survival probability power spectra, two non-overlapping confidence windows can be determined. In panel (c) a double peak structure is present and we assign the symmetric stretch to the peak at lower energy. In panel (d) a single band conceals both states and its peak is taken as the reference for the absorption of the asymmetric stretch fundamental, in agreement with the underlying EBK trajectory. In this way, though, the intensity of this band is expected to be overestimated because contributions from the symmetric stretch are also included. However, it turns out that in

this specific case they are small. Finally, a well defined single band characterizes the OH stretch fundamental in panel (e).

As discussed in the theory Section and already pointed out for water, the relative intensities of the different bands in Fig. 4 are not directly comparable to each other in absence of a preliminary normalization. Each IR spectrum in Fig. 4 has been globally normalized by applying Eq.(42) at the energy corresponding to the maximum of the absorption band in the confidence region (i.e. in correspondence of the vertical dashed lines reported in Fig. 4). The relative intensities of absorption bands in confidence regions obtained with different propagators are hence meaningful. In order to determine the normalization constants via Eq.(42), the sum over $\mathbf{K}$ has been performed over a truncated set of harmonic basis set elements. The truncation strategy adopted consisted in considering all possible 24-dimensional direct products of 1-dimensional harmonic eigenstates up to the total quantum number $k_{max} = 6$ with a maximum of 2 modes simultaneously excited. The truncated basis set obtained in this way contains 10081 harmonic states. Not surprisingly, looking at the amplitude of the expansion coefficients associated to each peak, in all cases the largest coefficient belongs to the harmonic excited state corresponding to the absorption band. However, dozens of other relevant contributions are also present and many of them are due to (dark) harmonic states in resonance with the principal one. So, while within the harmonic approximation at T=0 K only fundamental transitions give non-zero contributions to the IR spectrum, multiple Fermi resonances mixing the bright harmonic states with many others imply that the number of non-dark states becomes much larger in the real anharmonic case. The absorption bands obtained in Fig. 4, hence, are made of multiple energetically close vibrational transitions. This is different from the simpler picture of absorption bands as made of a single broadening function around a central bright transition. Any state-by-state approach would fail to describe this phenomenology, unless all oscillator strengths relative to transitions from a given reference state to all states under the absorption band are taken into account. Apart from the computational overhead required, this procedure would be numerically unfeasible.

In Fig. 5 we report our estimate for the zero temperature fully anharmonic IR spectrum $S_0(\omega)$. In the same figure, the semi-anharmonic spectrum $S_{\phi_0}(\omega)$ is also presented in dashed lines. The different colors (the same adopted in Fig. 4) identify the single trajectory propagator used in each confidence energy range and, for all bands reported, the relative intensities are meaningful. The spectrum $S_0(\omega)$ has been obtained by sorting the elements

of the harmonic basis set in order of relevance (i.e. absolute value) of the corresponding ground state expansion coefficients and then applying Eq.(25) with the sum over harmonic states performed over the first 50 basis set elements. With this cutoff, all coefficients with absolute value bigger than $\sim 0.03$ in the ground state expansion (see Fig. S1 in Supporting Information) are considered. However, in order to keep under control the number of power spectra to evaluate, we have limited the sum over normal mode contributions to the same set of coordinates used for the calculation of $S_{\phi_0}(\omega)$. The total number of power spectra taken into account in this way, in fact, already adds up to $\sim 2.6 \times 10^6$. As shown in Fig. S2 of the Supporting Information, inclusion of more and more harmonic contributions does have an observable effect on the intensity of all bands with convergence obtained only after the inclusion of the $40^{th}$ basis set element. The observed increment in absorption intensities, however, is proportionally almost equivalent for all bands. Hence, after a global re-normalization, the overall effects of taking into account the anharmonicity of the vibrational ground state (in terms of change in relative band intensities) become minor.

It is worth noting that for absorption bands composed of more than one transition line, the maximum may not formally coincide with the maximum of a given survival amplitude power spectrum. A not negligible change in the estimate of the position of the maximum of absorption bands is indeed observed for the semi-anharmonic spectra of glycine in panels (a)-(d) of Fig. 4. Furthermore, an additional even if less significant shift of most band maxima is observed ($\lesssim 5$ cm$^{-1}$) when ground state anharmonicity is also taken into account (see Fig. S2 in Supporting Information). In Fig. 5 we report also the experimental spectrum recorded in Argon matrix at low temperature.[? ] A satisfactory overall agreement between our theoretical prediction and the experimental spectrum is observed. The key difference is in the region between 3150 and 3250 cm$^{-1}$, where the experimental signature of another conformer, not included in our theoretical investigation, stands out. The most intense absorption band is centered at 3560 cm$^{-1}$ and associated to the OH stretch. The second most intense experimental band is located in the energy range between 2900 and 3000 cm$^{-1}$, with an overall absorption intensity $\sim 6$ times smaller than the OH one. This band features a barely distinguishable bimodal shape and is associated to the two (symmetric and asymmetric) CH$_2$ stretches. The bimodality of this band is not resolved in our semiclassical spectrum in which the broadening functions resulting from the dynamics are indeed too large to recover this effect. Notably, however, a secondary much less intense band located between

3050 and 3150 cm$^{-1}$ appears in the computed spectrum. In fact, as revealed by the power spectra reported in Fig. 4 (black dashed curves), in this spectral energy range vibrational states with relevant components on the harmonic excited states of the two CH$_2$ stretches are present, most probably originated by the presence of Fermi resonances coupling them with (dark) harmonic excited states. The intensity of this secondary band, as expected, is much reduced compared to the principal one in the semi-anharmonic spectrum, but its oscillator strength increases upon inclusion of ground state anharmonicities making it non negligible. However, the signal is not enough intense to be assigned in the experimental spectrum and to be discernible from noise or signals coming from other conformers. Finally, another absorption band (assigned to the NH$_2$ stretches) in the experimental spectrum spans the energy range between 3370 and 3480 cm$^{-1}$. Its intensity, even if an order of magnitude smaller than that of the OH stretch, makes it clearly distinguishable from the experimental baseline and well matched in position and intensity by the semiclassical prediction.

| Assignment (Modes) | MC-TA-SCIVR | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | Harmonic | | Anharmonic | | | |
| | (B3LYP/aug-cc-pVDZ) | | | | | |
| | $\omega$ | $I$ | $\omega$ | $I$ | | |
| CH$_2$ (20;21) | 3051; 3089 | 0.08; 0.03 | 2900 | 0.07 | Expt. | |
| NH$_2$ (22;23) | 3495; 3568 | 0.01; 0.02 | 3345; 3430 | 0.03; 0.01 | | |
| OH (24) | 3735 | 0.25 | 3563 | 0.25 | $\omega$ | $I$ |
| GVPT2 | | | | | 2935 | 0.04 |
| | Harmonic | | Anharmonic | | 3410; 3450 | 0.03; 0.01 |
| | (B3LYP/N07D) | | | | 3560 | 0.25 |
| | $\omega$ | $I$ | $\omega$ | $I$ | | |
| CH$_2$ (20;21) | 3044; 3079 | 0.08; 0.03 | 2938; 2929 | 0.09; 0.04 | | |
| NH$_2$ (22;23) | 3509; 3582 | 0.01; 0.03 | 3387; 3407 | 0.01; 0.02 | | |
| OH (24) | 3750 | 0.25 | 3568 | 0.25 | | |

Table I. Main features of the SC IR spectrum of glycine most stable conformer in the high energy range are compared to experiments[?] as well as second order Generalized Vibrational Perturbation Theory (GVPT2) calculations performed at the B3LYP/N07D level of theory.[?] Energy of maximum/maxima associated to each IR band are reported in cm$^{-1}$ while intensities are scaled in order to level the intensity of the OH stretch band up to the experimental one in all cases. Results from the double harmonic approximation are also reported.

Energies and intensities obtained from the semiclassical IR spectrum are reported in Table I and compared to both experimental values and theoretical estimates provided by the double harmonic approximation and by perturbative methods at a similar level of electronic structure theory.[?] In order to facilitate the comparison, all intensities have been globally re-normalized by leveling off the intensities assigned by the different approaches to the intense OH stretch band. This reference value is set equal to the experimental value of 0.25. From Table I it is clear that a double harmonic approximation is not efficient especially in estimating the frequencies of vibration, while a better prediction is obtained for intensities.

Results are neatly improved moving to the SC and GVPT2 calculations. Both methods are in excellent agreement with the experiment for all modes, taking alternatively the lead as the most accurate approach. This demonstrates that the SC approach works appropriately and accurately if compared to benchmark calculations in both low (water) and high (glycine) dimensionality, opening the way to the quantum simulation of IR spectra of systems currently not achievable.

## IV.  SUMMARY AND CONCLUSIONS

We have presented a new semiclassical approach to simulate anharmonic IR spectra. The method is able to deal with medium-large molecular systems with sizable densities of vibrational states. At the heart of the strategy are MC-SCIVR power spectra, which are combined linearly to give, in addition to frequencies of vibration, IR intensities and band shapes.

Meaningful IR intensities are indeed obtained exploiting a proper decomposition of the absorption spectrum in terms of survival amplitude power spectra. As for band shapes the method relies on the fact that, as the number of absorption lines under a given band becomes large and in the limit in which the width of the band is much larger than the width of the single lines, the shape of the band is independent of the broadening function adopted. This allows to employ the convolution function associated to the finite-time survival amplitude Fourier transform as a broadening function in lieu of the usual Gaussian or Lorentzian envelopes. Furthermore, this feature permits to avoid undesired state-to-state oscillator strength calculations, whose number becomes huge for large systems making their computation hardly feasible. In fact, even though MC-SCIVR power spectra need a proper normalization to allow comparisons between different energy regions, normalization constants refer to the central peak of absorption bands and barely vary within it. Thus, only one constant per absorption band has to be computed, a figure that in large systems is much smaller than the actual number of lines composing each band.

The approach is also based on the linearization of the molecular dipole, but intensities of absorption spectra can still be calculated at the double harmonic, semi-anharmonic, or fully anharmonic level. The difference depends on the way the states contributing to the linearized transition dipoles are treated. The accuracy of the method has been tested against exact

results for the water molecule with more than satisfactory outcome. Furthermore, the study of the IR spectrum of glycine has provided results in agreement with both experiments and previous VPT2 calculations and demonstrated the effectiveness of the method in dealing with a 24-dimensional system. Investigation of glycine focused on the absorption bands of high energy fundamentals, an energy region where anharmonicities are relevant. The importance of anharmonicities at both electrical and mechanical level is evident after a comparison to the widely employed double harmonic approximation. In both cases explored, the double harmonic approximation is off the mark, which raises serious questions about the legitimacy of employing such a rough guess in a black box fashion.

A still open issue is that, due to the short duration of the semiclassical propagation, the SC absorption bands obtained for glycine are wider than the low-temperature experimental ones. This effect could be less relevant for spectra at higher temperature, where experimental bands are expected to extend over a larger interval of energies. Furthermore, the quality of the SC calculation could be improved by reducing the impact of spurious rotations, which may contribute to the enlargement of peaks and are due to the adoption of a normal mode reference frame. Advances on this aspect are currently being undertaken.

We conclude by remarking that SC spectroscopy, through power spectra calculations, had already well established two of its hallmark features. They are the possibility of application to large dimensional molecular systems, which is often precluded to other quantum approaches, and the capability to reproduce and explain experimental findings where other theoretical approaches (like scaled harmonic or classical ones) fail. Now we have increased the appeal of semiclassical spectroscopy by demonstrating that semiclassical IR spectra are also achievable without any restrictions due to the density of states. This gives SC approaches the potential to be reference methods for IR spectroscopy of medium-large systems.

**SUPPLEMENTARY MATERIAL**

See supplementary material for additional data about the expansion coefficients of the ground state vibrational eigenfunction of glycine and an additional plot about convergence of the full-anharmonic IR spectrum of glycine as a function of the number of expansion coefficients taken into account.

**Appendix A**

From quantum linear response theory in its sum-over-state version, the functional form of the IR spectrum of isotropic and homogeneous molecular systems is

$$S(\omega, T) = \sum_n \sum_{m \neq n} [P_n(T) - P_m(T)] \, \Omega_{nm} \, \frac{1}{3} \, |\boldsymbol{\mu}_{nm}|^2 \, \delta(\omega - \Omega_{nm}), \qquad (A.1)$$

where $\Omega_{nm} = E_m - E_n$ is the difference between the vibrational energies of a given transition, $\boldsymbol{\mu}_{nm} = \langle e_n | \hat{\boldsymbol{\mu}}_{0N} | e_m \rangle$ is the corresponding transition dipole, and $P_l = e^{-\frac{E_l}{k_B T}}/Z$ is the $l^{th}$ vibrational state population at a given temperature T (with $Z = \sum_n e^{-\frac{E_n}{k_B T}}$ being the partition function). In this work we are interested in the modeling of the IR spectrum in the region of the high energy fundamentals ($\omega \gtrsim 2500 \; cm^{-1}$) of a system at a given (reasonable) temperature $T$. Hence, $\Omega_{nm} \gg K_B T$ and it can be assumed that all the arrival state populations $P_m(T)$ are zero. Eq.(A.1) simplifies to

$$S(\omega, T) = \sum_n \sum_{m \neq n} P_n(T) \, \Omega_{nm} \, \frac{1}{3} \, |\boldsymbol{\mu}_{nm}|^2 \, \delta(\omega - \Omega_{nm}) =$$
$$= \sum_n P_n(T) S_n(\omega), \qquad (A.2)$$

where

$$S_n(\omega) = \sum_{m \neq n} \Omega_{nm} \, \frac{1}{3} \, |\boldsymbol{\mu}_{nm}|^2 \, \delta(\omega - \Omega_{nm}). \qquad (A.3)$$

In experimental IR spectra each spectral line is broadened over a finite range of energies. This is due to several known effects. For instance thermal Doppler, collisional, and Stark

25

effect broadening all play an important role in a gas phase environment. These effects can be accounted for phenomenologically by assuming that the transition probability distributions of the system are broadened, so that in Eq.(A.3) the Dirac delta function can be substituted by some bell shaped function $\mathcal{L}(\omega - \Omega_{nm}; \Gamma)$ of finite amplitude $\Gamma$. Eq. (A.3) in this case becomes

$$S_n(\omega) = \sum_{m \neq n} \frac{1}{3} \, |\boldsymbol{\mu}_{nm}|^2 \, \omega \, \mathcal{L}(\omega - E_m + E_n; \Gamma). \tag{A.4}$$

It is straightforward to notice from a comparison between Eq. (A.3) and Eq. (A.4) that the term $\Omega_{nm}$, characteristic of a transition between two single and well defined states, has been substituted by $\omega$ in the case of an absorption band. This is necessary to account for the gain of a quantum $\hbar\omega$ of energy by the system (and, correspondingly, for the loss of an equal amount of energy by the electromagnetic field) in the neighborhood of $\Omega_{nm}$ with probability density given by $\frac{1}{3} \, |\boldsymbol{\mu}_{nm}|^2 \, \mathcal{L}(\omega - E_m + E_n; \Gamma)$. Depending on the most relevant phenomenological effect to take into account, $\mathcal{L}$ is generally described by either a Gaussian or a Lorentzian function. In this work $\mathcal{L}$ is substituted by the dynamical convolution function $\mathcal{D}(\omega - E_m + E_n; \Gamma_\tau)$, a legitimate procedure as explained in the Summary and Conclusions Section of the paper. Therefore the calculated spectrum in presence of a radiation polarized along the direction $\epsilon$, $S_{\epsilon,n}(\omega)$, descends from Eq.(19).

It is common practice to use normalized broadening functions that integrate to unity over the energy axis. This choice, in fact, has the advantage to preserve the oscillator strength summation rules. The most convenient choice for our purposes, instead, is to normalize them by setting $\mathcal{D}(0, \Gamma_\tau) = 1$. This choice has the advantage to allow a straightforward calculation of the normalization constants of the eigenstates expanded on the harmonic basis via Eq.(42). These two normalization choices are in any case equivalent because the resulting absorption spectra just differ in a multiplicative constant.

**Appendix B**

We derive the decomposition of the fully anharmonic IR spectrum of Eq.(19) in terms of power spectra, i.e. the result reported in Eq.(25). First, Eq.(17) for transition dipoles has to be inserted into Eq.(19), leading to

$$S_{\epsilon,n}(\omega) = \omega \sum_{m \neq n} \left| \sum_{\mathbf{K}} C_{n,\mathbf{K}} \sum_{\alpha} C_{m,\mathbf{K}(\alpha)} \mu_{\mathbf{K}(\alpha)}^{\epsilon} \right|^2 \mathcal{D}(\omega - E_m + E_n; \Gamma_\tau). \qquad \text{(B.1)}$$

At this point the square modulus is expanded to allow inversion of the order of the sums over $\alpha$ and over $m$. From an algebraic point of view this procedure is not straightforward because of the presence of multiple sums, so we perform it starting from the following general expansion

$$\left| \sum_{\mathbf{K}} C_{n,\mathbf{K}} A_{m,\mathbf{K}} \right|^2 = \sum_{\mathbf{K}} \left[ C_{n,\mathbf{K}}^2 A_{m,\mathbf{K}}^2 + \sum_{\mathbf{K}' < \mathbf{K}} 2 C_{n,\mathbf{K}} C_{n,\mathbf{K}'} A_{m,\mathbf{K}} A_{m,\mathbf{K}'} \right]. \qquad \text{(B.2)}$$

If

$$A_{m,\mathbf{K}} = \sum_{\alpha} C_{m,\mathbf{K}(\alpha)} \mu_{\mathbf{K}(\alpha)}^{\epsilon}, \qquad \text{(B.3)}$$

then

$$A_{m,\mathbf{K}} A_{m,\mathbf{K}'} = \sum_{\alpha,\alpha'} C_{m,\mathbf{K}(\alpha)} C_{m,\mathbf{K}'(\alpha')} \mu_{\mathbf{K}(\alpha)}^{\epsilon} \mu_{\mathbf{K}'(\alpha')}^{\epsilon}, \qquad \text{(B.4)}$$

and upon substitution into Eq.(B.1)

$$\begin{aligned} S_{\epsilon,n}(\omega) =& \omega \sum_{\mathbf{K}} C_{n,\mathbf{K}}^2 \Bigg\{ \sum_{\alpha} \Bigg[ \left( \sum_{m \neq n} C_{m,\mathbf{K}(\alpha)}^2 \mathcal{D}(\omega - E_m + E_n) \right) (\mu_{\mathbf{K}(\alpha)}^{\epsilon})^2 + \\ &+ \sum_{\beta < \alpha} \mu_{\mathbf{K}(\alpha)}^{\epsilon} \mu_{\mathbf{K}(\beta)}^{\epsilon} \left( \sum_{m \neq n} 2 C_{m,\mathbf{K}(\alpha)} C_{m,\mathbf{K}(\beta)} \mathcal{D}(\omega - E_m + E_n) \right) \Bigg] + \\ &+ \sum_{\mathbf{K}' < \mathbf{K}} \sum_{\alpha,\alpha'} C_{n,\mathbf{K}} C_{n,\mathbf{K}'} \left( \sum_{m \neq n} 2 C_{m,\mathbf{K}(\alpha)} C_{m,\mathbf{K}'(\alpha')} \mathcal{D}(\omega - E_m + E_n) \right) \mu_{\mathbf{K}(\alpha)}^{\epsilon} \mu_{\mathbf{K}'(\alpha')}^{\epsilon} \Bigg\}, \end{aligned} \qquad \text{(B.5)}$$

The final Eq.(25) is obtained starting from Eq.(B.5) by noting that: i) the terms in brackets are the semi-anharmonic spectra $S_{\epsilon,\phi_{\mathbf{K}}}(E)$; ii) in the last line, the terms in parentheses are also easily related to survival amplitudes of the kind of Eq.(23).
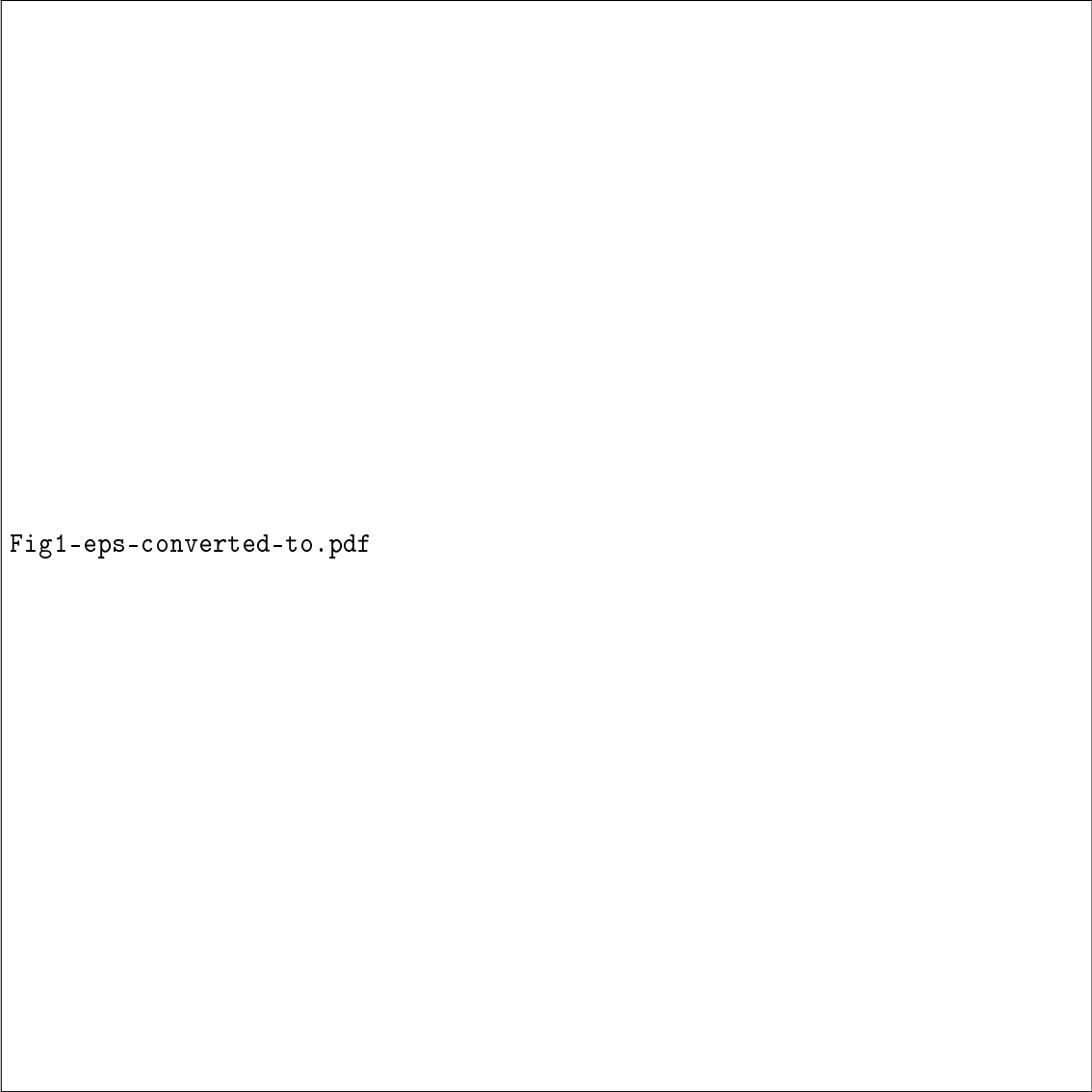
Fig1-eps-converted-to.pdf

Figure 1. Semiclassical power spectra for a non-rotating water molecule using the MC approach based on single-trajectory propagators. The peaks represent the ground state (brown), first bending excitation (red), first bending overtone (magenta), and first symmetric (blue) and asymmetric (green) excitations. In panel (a) power spectra from harmonic reference states (i.e. $\boldsymbol{\nu} = \mathbf{K}$) are reported. In panel (b) the same spectra are shown upon normalization. Panel (c) illustrates the semiclassical IR spectrum at 0K for unpolarized light. The inset, panel (d), zooms in on the bending overtone.
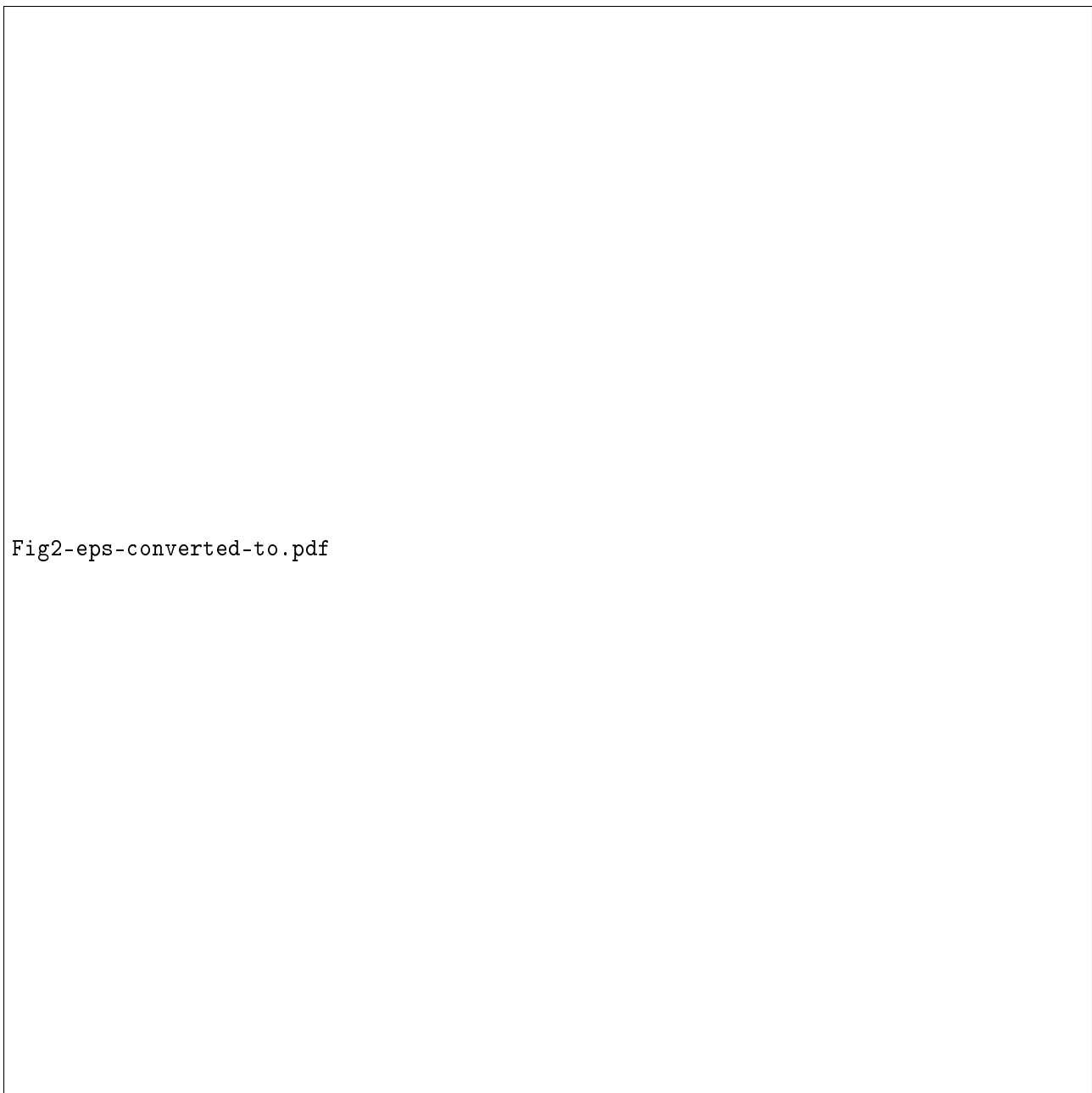
Fig2-eps-converted-to.pdf

Figure 2. Detail of the three bright IR peaks of water reported in Fig. 1. The exact (vertical black line), SC semi-anharmonic (colored dashed line), and SC fully anharmonic (colored solid line) intensity estimates are reported in each plot.
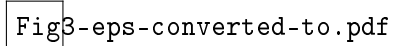
Fig3-eps-converted-to.pdf

Figure 3. Ball-and-stick representation of the global minimum energy structure of glycine in vacuum.
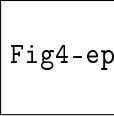
Fig4-eps-converted-to.pdf

Figure 4. The semi-anharmonic IR absorption spectra of glycine in vacuum $S_{\phi_0}(\omega)$ (continuous color lines) are reported together with survival amplitude power spectra $\tilde{I}_{\phi_{\mathbf{K}}}(\omega)$ (black dash lines) with harmonic reference states $\phi_{\mathbf{K}}$ corresponding to the EBK single trajectory used for the semiclassical propagator (i.e. $\boldsymbol{\nu} = \mathbf{K}$). Results have been obtained using the propagator associated to the first harmonic excitations of modes 20-24 (panels from (a) to (e)). Dashed vertical lines are located at the frequencies of the maxima of the absorption bands corresponding to the fundamental transitions. For each propagator, the energy confidence window is derived by the comparison of power and IR spectra and is highlighted by reporting it in a region with white background.
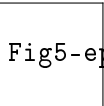
Fig5-eps-converted-to.pdf

Figure 5. Comparison between experiment (black line, top) and the fully anharmonic zero temperature IR spectrum of glycine in vacuum (continuous lines, bottom). The semi-anharmonic spectrum (dashed lines, bottom) is also reported. The fully and semi-anharmonic spectra have been obtained by means of single trajectory propagators, each one valid in its confidence energy range. This is indicated by the different colors chosen in agreement with Fig. 4.