Semiclassical Vibrational Spectroscopy with Hessian Databases

Riccardo Conte,[1, a)] Fabio Gabas,[1] Giacomo Botti,[1] Yu Zhuang,[2, b)] and Michele Ceotto[1, c)]

[1)] *Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy.*

[2)] *Department of Computer Science, Texas Tech University, Lubbock, Texas 79409-3104, USA.*

We report on a new approach to ease the computational overhead of ab initio "on-the-fly" semiclassical dynamics simulations for vibrational spectroscopy. The well known bottleneck of such computations lies in the necessity to estimate the Hessian matrix for propagating the semiclassical pre-exponential factor at each step along the dynamics. The procedure proposed here is based on the creation of a dynamical database of Hessians and associated molecular geometries able to speed up calculations while preserving the accuracy of results at a satisfactory level. This new approach can be interfaced to both analytical potential energy surfaces and on-the-fly dynamics, allowing one to study even large systems previously not achievable. We present results obtained for semiclassical vibrational power spectra of methane, glycine, and N-acetyl-L-phenylalaninyl-L-methionine-amide, a molecule of biological interest made of 46 atoms.

[a)]Electronic mail: riccardo.conte1@unimi.it

[b)]Electronic mail: yu.zhuang@ttu.edu

[c)]Electronic mail: michele.ceotto@unimi.it

I. INTRODUCTION

Several computational methods for calculating quantum molecular vibrational spectra have been established. To name some popular ones, we recall the vibrational self-consistent field and virtual space configuration interaction (VSCF/VCI) method,[1–4] the multi-configuration time dependent Hartree (MCTDH) approach,[5–9] the use of contracted basis functions with an eigensolver,[10,11] collocation methods,[12,13] and second-order vibrational perturbation theory (VPT2).[14] Despite the advance these methods provide in dealing with the exponential scaling of quantum approaches with the number of vibrational degrees of freedom, the investigation of large molecules or complex supra-molecular systems remains quite challenging. In this case classical simulations or ad-hoc scaling of harmonic estimates are the approaches commonly employed in spite of the drawback of totally neglecting or not properly describing quantum effects (such as zero-point energies, quantum couplings, and overtones). This aspect has been recently confirmed by a semiclassical (SC) investigation,[15] which has explained experimental findings due to quantum effects, in contrast to previous non conclusive theoretical attempts.[16,17] Furthermore, recent SC techniques which are capable of accurately calculating the intensities of vibrational spectral transitions have demonstrated that the widely employed double harmonic approximation often provides only a rough estimate, which has to be corrected by incorporating quantum couplings and anharmonicities.[18,19]

Semiclassical dynamics is indeed gaining more and more attention due to some of its peculiar features, which include the possibility of accounting for quantum effects in an accurate way starting from short-time classical trajectories,[20–45] the advantage of not being limited to a single-well picture,[46–48] and the possibility of adopting an ab initio "on-the-fly" approach.[47,49–53] These characteristics make the range of applicability of semiclassical spectroscopy enormously vast, ranging from small molecules, for which very precise analytical potential energy surfaces (PES) can be constructed, to larger molecules or even supra-molecular systems whose dynamics can be integrated "on-the-fly" from ab initio calculations. Clearly, a very different computational effort is required depending on whether calls to the potential are analytical or ask for ab initio electronic structure calculations as in the case of "on-the-fly" simulations.[54–58] In SC dynamics this difference translates mainly into the very different number of trajectories on which SC calculations are based. In fact,

when working with an analytical PES and exploiting time-average (TA) filtering techniques, a proper Monte Carlo convergence in the evaluation of the semiclassical phase-space integration can be reached.[59,60] A tailored approach, known as Multiple Coherent Semiclassical Initial Value Representation (MC SCIVR), must be employed instead when adopting "on-the-fly" dynamics, so that even a single tailored trajectory allows one to regain the main spectral features.[32,47]

Application of standard SC techniques is hampered by a few well-known drawbacks. One is the so-called "curse of dimensionality" which concerns the impossibility of obtaining a resolved spectroscopic signal when the dimensionality of the system becomes too large. Another one is related to the chaotic behavior of classical dynamics. Chaotic trajectories make numerical integration of the monodromy matrix elements (needed to evaluate the SC frozen Gaussian propagator) inaccurate because of finite machine precision. The consequence is that, if the contribution of these trajectories is neither discarded nor reshaped, the entire calculation may be spoiled. However, even in the case of non-chaotic trajectories, propagation of the monodromy matrix elements represents a computational bottleneck due to the necessity to evaluate the Hessian matrix (matrix of second derivatives of the potential with respect to the nuclear coordinates) at each step along the dynamics.

Recent advances in the semiclassical field have permitted researchers to make progress on these issues and further research is still being carried out. For instance, a "divide-and-conquer" strategy (DC SCIVR) has been employed to extend semiclassical spectroscopy to systems as large as fullerene;[48,61,62] appropriate approximations have been introduced to deal with chaotic trajectories;[47,63–65] and a compact finite-difference (CFD) approximation to the Hessian has been able to reduce the number of Hessian calculations required.[66–68] In particular, CFD-Bofill schemes permit one to preserve the accuracy in the propagation of monodromy matrix elements quite well and, using them, accurate semiclassical spectra have been obtained for several small molecules including an "on-the-fly" application to CO_2.[69]

So far this introduction has been focused on SC frozen Gaussian propagators, which were employed in the numerical applications presented in this paper. However, Hessian calculations are ubiquitous in semiclassical dynamics and, in fact, they are required also by another variety of SC propagators, known as thawed Gaussian propagators.[70–74] The main difference between the two families is that in a frozen Gaussian the width of the Gaussians is a constant in time, while in a thawed Gaussian the width evolves in time.

Furthermore, SC thawed Gaussian calculations are based on single-trajectory dynamics and are known to be less accurate than their frozen Gaussian counterparts even in the case of MC SCIVR single trajectory simulations. Interpolation schemes have been developed for thawed Gaussian propagators,[52,53,75–78] and, at the time of writing, an approach based on a single, initial Hessian evaluation has just been proposed and applied to SC thawed Gaussian vibro-electronic (vibronic) calculations.[79]

The main goals of this paper are the introduction of a new strategy to slash the computational overhead associated with Hessian matrix calculations and its application to precise semiclassical frozen Gaussian vibrational spectroscopy of polyatomic molecular systems. The approach is based on the dynamical construction of a database of Hessian matrices dependent on the molecular geometry visited along the dynamics. It is not as computationally inexpensive as the thawed Gaussian single Hessian scheme, but more accurate. We show that this Hessian approximation is accurate, and able to extend the range of applicability of "on-the-fly" SC vibrational spectroscopy to larger chemical species. The paper introduces the theoretical details of the new approach in Section II. Section III is devoted to applications to methane, glycine and N-acetyl-L-phenylalaninyl-L-methionine-amide. Finally, in Section IV, a summary and some conclusions are reported.

II.   THEORETICAL AND COMPUTATIONAL DETAILS

The semiclassical calculation of vibrational power spectra of molecular species is a long established technique, which does not suffer from zero point energy (ZPE) leakage[80] and has recently evolved into a very powerful computational tool able to deal with systems involving dozens of degrees of freedom. In this Section we introduce the SC theoretical methodology on which our results are based, and then focus on the description of the new database strategy.

In a pivotal paper[59] Kaledin and Miller demonstrated that by time averaging the Herman-Kluk semiclassical propagator, it is possible to work out a computationally advantageous expression for the power spectrum of a survival amplitude evolved according to a vibrational Hamiltonian. Their starting point, the Herman-Kluk (HK) propagator, provides a semiclassical approximation to the exact quantum propagator. In particular, for an N_{vib}-dimensional system, the survival amplitude of a generic reference state $|\chi\rangle$ can be written as

$$\left\langle \chi \left| e^{-i\hat{H}t/\hbar} \right| \chi \right\rangle_{HK} = \left(\frac{1}{2\pi\hbar} \right)^{N_{vib}} \iint d\mathbf{p}_0 d\mathbf{q}_0 C_t(\mathbf{p}_0, \mathbf{q}_0) e^{\frac{i}{\hbar}S_t(\mathbf{p}_0, \mathbf{q}_0)} \langle \chi | \mathbf{p}_t, \mathbf{q}_t \rangle \langle \mathbf{p}_0, \mathbf{q}_0 | \chi \rangle. \quad (1)$$

The Monte Carlo integration needed to evaluate the rhs of Eq. (1) is performed by generating a set of phase space points $(\mathbf{p}_0, \mathbf{q}_0)$ according to the distribution function defined by the overlap $\langle \mathbf{p}_0, \mathbf{q}_0 | \chi \rangle$. $|\mathbf{p}, \mathbf{q}\rangle$ indicates a coherent state, which has a Gaussian representation in both position and momentum space. Specifically,

$$\langle \mathbf{x} | \mathbf{p}, \mathbf{q} \rangle = \left(\frac{\det(\mathbf{\Gamma})}{\pi^{N_{vib}}} \right)^{1/4} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{q})^T \Gamma(\mathbf{x}-\mathbf{q}) + i\mathbf{p}^T(\mathbf{x}-\mathbf{q})/\hbar}, \quad (2)$$

where $\mathbf{\Gamma}$ is the $N_{vib} \times N_{vib}$ coherent state width matrix that is generally chosen to be diagonal with elements equal to the harmonic frequencies of vibrations. The select phase space points $(\mathbf{p}_0, \mathbf{q}_0)$ serve as starting conditions for classical dynamics runs that permit one to evaluate the whole integrand. In fact, at every time t along the dynamics, the coherent state overlap $\langle \chi | \mathbf{p}_t, \mathbf{q}_t \rangle$ is readily estimated; $S_t(\mathbf{p}_0, \mathbf{q}_0)$ is the instantaneous classical action, and $C_t(\mathbf{p}_0, \mathbf{q}_0)$ is the Herman-Kluk pre-exponential factor defined as

$$C_t(\mathbf{p}_0, \mathbf{q}_0) = \sqrt{\left| \frac{1}{2} \left(\frac{\partial \mathbf{q}_t}{\partial \mathbf{q}_0} + \mathbf{\Gamma}^{-1}\frac{\partial \mathbf{p}_t}{\partial \mathbf{p}_0}\mathbf{\Gamma} - i\hbar\frac{\partial \mathbf{q}_t}{\partial \mathbf{p}_0}\mathbf{\Gamma} + \frac{i\mathbf{\Gamma}^{-1}}{\hbar}\frac{\partial \mathbf{p}_t}{\partial \mathbf{q}_0} \right) \right|}. \quad (3)$$

The Fourier transform of Equation (1) yields the SC power spectrum (i.e. the spectral density) associated to the vibrational Hamiltonian $\hat{H}$

$$I(E)_{HK} = \frac{Re}{\pi\hbar} \int_0^{+\infty} dt \, e^{iEt/\hbar} \left\langle \chi \left| e^{-i\hat{H}t/\hbar} \right| \chi \right\rangle_{HK}. \quad (4)$$

The SC estimates of the molecular eigenenergies of vibration correspond to the energy values at which the peaks in the power spectrum are localized. The ground state energy, i.e. the zero point energy, is commonly shifted to zero and, in this case, the peaks in the power spectrum are centered at the quantum (semiclassical) transition frequencies of fundamentals and overtones. Unfortunately, convergence of the phase space Monte Carlo integration for the Herman Kluk propagator is too slow for application to sizeable molecular systems due to the oscillatory nature of its integrand. This issue is overcome by Kaledin and Miller's time average filter, which leads to a more efficient formula for the spectral density known as

the time averaged semiclassical initial value representation (TA SCIVR)

$$I(E)_{TASCIVR} = \left(\frac{1}{2\pi\hbar}\right)^{N_{vib}} \iint d\mathbf{p}_0 d\mathbf{q}_0 \frac{1}{2\pi\hbar T} \left| \int_0^T e^{\frac{i}{\hbar}[S_t(\mathbf{p}_0,\mathbf{q}_0)+Et+\phi_t(\mathbf{p}_0,\mathbf{q}_0)]} \langle\chi|\mathbf{p}_t,\mathbf{q}_t\rangle dt \right|^2 . \quad (5)$$

In Eq.(5) $\phi_t(\mathbf{p}_0,\mathbf{q}_0)$ is the phase of the original complex-valued Herman-Kluk pre-exponential factor introduced owing to the so-called "separable" approximation,[59] and T is the total time of the dynamics. The key advantage of Eq. (5) is that the integrand is positive definite and this facilitates the numerical convergence of the integration. The derivation of Eq. (5) is discussed extensively in Ref. 59, and the interested reader can find all the details therein.

For the purposes of this work it is important to point out that Eq. (5) is the basic working formula for our spectral simulations. It has been applied previously to several molecules and further developed to interface the TA-SCIVR formalism to "on-the-fly" dynamics for application to much larger molecular systems. Specifically, on-the-fly simulations must rely on a limited number of trajectories to be computationally affordable. This is achieved by adopting for each vibrational state a few tailored trajectories (i.e. SC propagators) and reference states $|\chi\rangle$, a technique known as multiple coherent semiclassical initial value representation (MC SCIVR). The trajectories are run at an energy corresponding to the harmonic estimate for the generic k-th state to investigate, while the corresponding reference state is chosen as

$$\left|\chi^{(k)}\right\rangle = \prod_{i=1}^{N_{vib}} \sum_{j=1}^{N_\alpha} \varepsilon_{ij}^{(k)} \left| p_{eq,i}^{(k)}, q_{eq,i}^{(k)} \right\rangle . \quad (6)$$

The coherent states in Eq.(6) are centered at specific positions and momenta $(\mathbf{p}_{eq}^{(k)}, \mathbf{q}_{eq}^{(k)})$, which determine the initial conditions and the energy of the classical trajectories needed by the semiclassical simulation. N_α accounts for the possible duplication of coherent states, which are linearly combined through the array of integers $(\boldsymbol{\varepsilon})$, to enforce parity and molecular symmetries. Parity symmetry allows to enhance specific peaks in the power spectrum (i.e. the ground state, specific fundamental transitions, or overtones) and it is obtained by inverting the sign of the momentum in the duplicate coherent state. Molecular symmetry (detection of peaks corresponding to vibrations belonging to a specific irreducible repre-

sentation of the symmetry group of the molecule) is introduced by inverting the sign of the (equilibrium) positions in the duplicate coherent state.[32,60] The greatest novelty of the MC SCIVR procedure is that accurate results can be obtained by running just one tailored trajectory per state with the effect of alleviating the computational cost substantially.

Another problem related to Eq. (5) is that the coherent state overlap $\langle\chi|\mathbf{p}_t,\mathbf{q}_t\rangle$ is more and more likely to give a negligible contribution as N_{vib} increases, so that a spectrally resolved Fourier signal is harder and harder to collect. A very recent advance able to overcome this problem is the divide-and-conquer semiclassical initial value representation (DC SCIVR), which is based on performing semiclassical calculations in lower dimensional subspaces. The dynamics is still performed in full dimensionality permitting one to recover at least partially the interaction between different subspaces, and Eq. (5) remains the working formula. The several quantities appearing in the integrand of Eq. (5) are readily projected onto the subspaces with the exception of the action, which requires a redefinition of the projected potential energy due to its non-separability in general. For an M-dimensional subspace

$$\tilde{I}(E)_{DCSCIVR} = \left(\frac{1}{2\pi\hbar}\right)^M \iint d\tilde{\mathbf{p}}_0 d\tilde{\mathbf{q}}_0 \frac{1}{2\pi\hbar T} \left| \int_0^T e^{\frac{i}{\hbar}\left[\tilde{S}_t(\tilde{\mathbf{p}}_0,\tilde{\mathbf{q}}_0)+Et+\tilde{\phi}_t(\tilde{\mathbf{p}}_0,\tilde{\mathbf{q}}_0)\right]} \langle\tilde{\chi}|\tilde{\mathbf{p}}_t,\tilde{\mathbf{q}}_t\rangle dt \right|^2 . \quad (7)$$

$$\tilde{S}_t(\tilde{\mathbf{p}}_0,\tilde{\mathbf{q}}_0) = \int_0^t dt' \left(\tilde{T}(t') - \tilde{V}(t')\right),$$
$$\tilde{V}(t') = V(\tilde{q}_M(t'); q_{N_{vib}-M}^{eq}) + \lambda(t');$$
$$\lambda(t) = V(\tilde{q}_M(t); q_{N_{vib}-M}(t)) - [V(\tilde{q}_M^{eq}; q_{N_{vib}-M}(t) + V(\tilde{q}_M(t); q_{N_{vib}-M}^{eq})];$$
$$\tilde{V}(t') = V(\tilde{q}_M(t'); q_{N_{vib}-M}(t')) - V(\tilde{q}_M^{eq}; q_{N_{vib}-M}(t')). \quad (8)$$

Projected quantities in Eqs. (7) and (8) are indicated by tildes. Upon parametrization of the degrees of freedom external to the M-dimensional subspace and correction by means of the external time-dependent field $\lambda(t)$, the projected potential is obtained as a difference between the instantaneous potential and the potential at the corresponding unrelaxed subspace minimum. Combination of the multiple coherent and divide-and-conquer techniques,

named MC-DC SCIVR is feasible and effective. A complete treatment of DC SCIVR can be found in Refs. 61,62, where appropriate strategies for determining the best suited subspaces have been illustrated and discussed.

A final issue concerning Eq. (5) is that evolution of the phase $\phi_t(\mathbf{p}_0, \mathbf{q}_0)$, which bears key quantum effects, requires to calculate the Hessian matrix along the whole classical trajectory and the computational effort therefore increases rapidly with the number of degrees of freedom. This is the bottleneck of semiclassical calculations plaguing both full dimensional and DC-SCIVR approaches. To overcome the problem it is necessary to find a way to reduce the number of Hessian calculations while keeping a sufficient accuracy for vibrational frequencies. A first effort in this direction has been made by developing a compact finite-difference approximation.[69] If the Hessian is calculated at the i-th step of the dynamics ($\mathcal{H}_i$), then the approximation estimates $\mathcal{H}_j$ ($j=i+1,...,i+N$-1) at the following N-1 steps according to the formula

$$\mathcal{H}_j = \mathcal{H}_{j-1} + (1-\lambda)\frac{\boldsymbol{R}\boldsymbol{R}^T}{\boldsymbol{R}^T\Delta\boldsymbol{q}} + \lambda\left(\frac{\Delta\boldsymbol{q}\boldsymbol{R}^T + \boldsymbol{R}^T\Delta\boldsymbol{q}}{||\Delta\boldsymbol{q}||^2} - \frac{\boldsymbol{R}^T\Delta\boldsymbol{q}}{||\Delta\boldsymbol{q}||^4}\Delta\boldsymbol{q}\Delta\boldsymbol{q}^T\right),\qquad(9)$$

where λ is a parameter identifying different families of approximations according to its value; $\Delta\mathbf{q} = \mathbf{q}_j$ - $\mathbf{q}_i$, and $\mathbf{R} = 2(\mathbf{g}_j$ - $\mathbf{g}_i$ - $\mathcal{H}_i\Delta\mathbf{q})$ with $\mathbf{g}_i$ being the gradient at the i-th geometry. The simplest form of the approximation employs the value $\lambda = 1$, and it is known as the Hessian Update (HU) Powell Symmetric Broyden scheme. A more refined choice of λ leads to the HU Bofill approximation. Calculation of the Hessian matrix is performed only at intervals of N steps, while, according to Eq. (9), only the computationally cheaper gradient must be evaluated for all the other steps. These schemes have been shown to work well for small-to-medium size molecules.[66,67,69]

In this work we introduce an alternative strategy to decrease the number of Hessian calculations. It is based on the dynamical creation of a database of Hessians (DBH) and related geometries, which can be exploited whenever the instantaneous molecular geometry is close enough to one of the N_{db} records already present in the database. An estimate of the "distance" between two molecular configurations can be obtained either by evaluating the root mean square difference between vibrational normal mode coordinates in the two configurations,

$$\sigma\left(\mathbf{q},\mathbf{q}_{db}\right)=\sqrt{\frac{\sum_{l=1}^{N_{vib}}(q_l-q_{l,db})^2}{N_{vib}}}, \tag{10}$$

where q_l and $q_{l,db}$ are the values of the l-th normal mode for the instantaneous geometry and one included in the database respectively, or by estimating the absolute difference for each mode

$$\Theta_l\left(\mathbf{q},\mathbf{q}_{db}\right)=|q_l-q_{l,db}|, \quad l=1,\dots,N_{vib}. \tag{11}$$

To determine whether the two configurations are enough close or not, an arbitrary threshold ρ is introduced. It is straightforward to demonstrate that $\sigma\left(\mathbf{q},\mathbf{q}_{db}\right)<\rho$ is a weaker condition than having the $\Theta\left(\mathbf{q},\mathbf{q}_{db}\right)$ test passed for each mode, i.e. $\Theta_l\left(\mathbf{q},\mathbf{q}_{db}\right)<\rho \quad \forall l\in[1,N_{db}]$. For this reason we employ Eq. (11) to determine the set of suitable records in the database, while Eq.(10) is used to identify the closest among database candidates ($\mathbf{q}_{db}^{closest}$) to the instantaneous geometry.

The search for the best match among the N_{db} records $\mathbf{q}_{db}$ in the database for the instantaneous geometry $\mathbf{q}$ is another important aspect of the method since it may become computationally costly when the size of the database grows large. To accelerate the search, mode-1 components of geometries in the database are sorted in increasing order. The algorithm starts with calculation of $\Theta_1\left(\mathbf{q},\mathbf{q}_{db}\right)$ along the database and check of the $\Theta_1\left(\mathbf{q},\mathbf{q}_{db}\right)<\rho$ condition. Once the test is passed, the search based on the first normal mode is stopped as soon as $\Theta_1\left(\mathbf{q},\mathbf{q}_{db}\right)>\rho$. Then, the $\Theta_l\left(\mathbf{q},\mathbf{q}_{db}\right)<\rho$ condition is evaluated sequentially for the other normal modes only on the restricted set of records that have passed all previous $\Theta_{l'}\left(\mathbf{q},\mathbf{q}_{db}\right)<\rho$, $l'=1,\dots,l-1$ tests.

If, after all N_{vib} modes have been examined, there are one or more configurations still left, then the Hessian associated with the geometry with the smallest $\sigma\left(\mathbf{q},\mathbf{q}_{db}\right)$ value, i.e. $\mathbf{q}_{db}^{closest}$, is selected to approximate the instantaneous Hessian in the SC simulation. Otherwise, a new Hessian calculation is performed and added to the database in the correct k-th position as determined by its $\mathbf{q}_{1,db}$ value. Figure 1 shows the flow chart for the method.

Some preliminary tests on systems of increasing dimensionality have been undertaken to check on the performance of the approach depending on database size, threshold ρ, and system complexity. As a first test, we employed analytical water[81] and methane[82] potentials to construct three databases of different sizes by running different numbers of trajectories.

9

Then, an additional trajectory was run to evaluate, at a specific value of ρ: i) the number of geometries for which a suitable match in the database was found; ii) the root mean square deviation (σ_{Hess}) of the elements of the associated Hessian in the database from those calculated at the trajectory instantaneous geometry. σ_{Hess} is defined as

$$\sigma_{Hess} = \sqrt{\frac{\sum_m^{N_{vib}} \sum_n^{N_{vib}} (\mathcal{H}_{mn} - \mathcal{H}_{mn,db})^2}{N_{vib} - 1}}, \tag{12}$$

where $\mathcal{H}_{mn}$ is the (m,n) component of the Hessian matrix. A single test trajectory is representative of the general case of interest for SC simulations in which initial conditions are sampled around the equilibrium geometry.

Table I reports the results. Each trajectory was evolved in full dimensionality with a symplectic algorithm for a total of 2,500 steps, i.e. it is made of 2,500 geometries. The larger the database becomes, the easier it is to find a suitable record in it and the more accurate the Hessian database estimate.
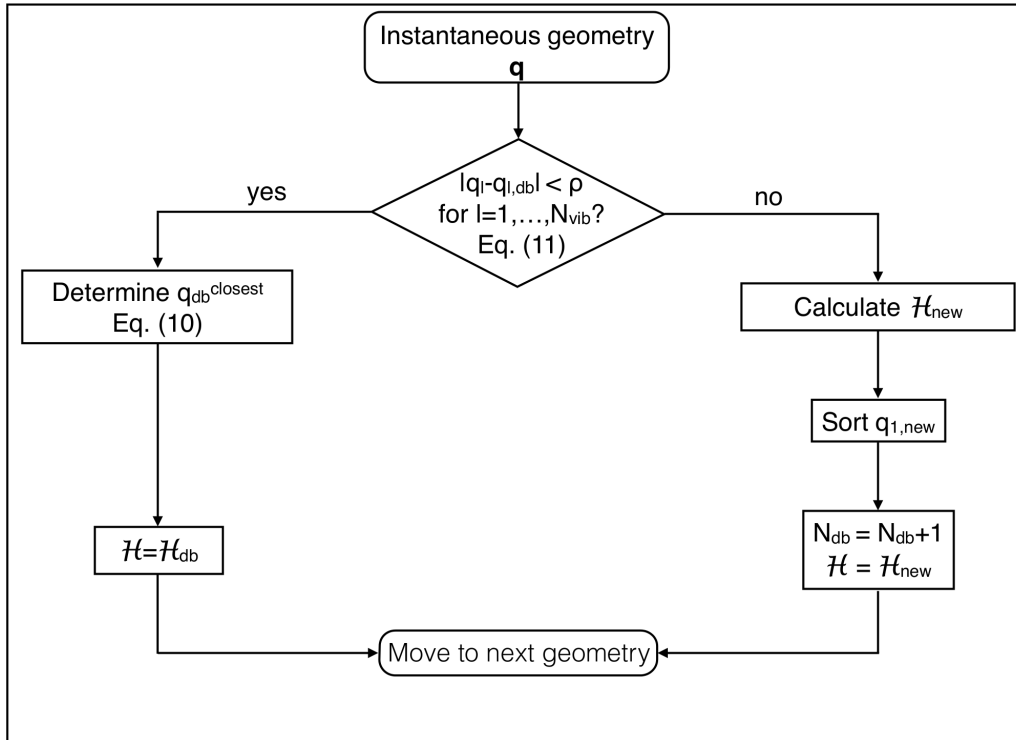

Figure 1. Flow chart for the DBH method.

Table I. Comparison of different databases for H_2O and CH_4. The threshold parameter is $\rho = 1$ for H_2O and $\rho = 10$ for CH_4. N^{db} indicates the number of geometries constituting the database; N^{mg} is the number of geometries of the test trajectory for which a matching database entry is found; σ_{Hess} is the root mean square deviation of the Hessian elements for the matched geometries, as defined in Eq. (12).

H_2O			CH_4		
N^{db}	N^{mg}	σ_{Hess}	N^{db}	N^{mg}	σ_{Hess}
2500	295	$9.4\,10^{-6}$	2500	118	$5.3\,10^{-5}$
25000	2157	$8.5\,10^{-6}$	25000	1209	$4.1\,10^{-5}$
250000	2500	$3.0\,10^{-6}$	250000	2500	$2.6\,10^{-5}$

For given H_2O and CH_4 databases, constructed by running ten trajectories made of 2,500 steps each (i.e. a database made of 25,000 records), the same quantities have been estimated for different values of the threshold parameter. The results are shown in Table II. It is not surprising to find that, for a given database, a stricter threshold value leads to fewer matched geometries and also to enhanced accuracy.

Table II. Number of matched geometries (N^{mg}) and root mean square Hessian deviation σ_{Hess} at various threshold values for a 25000-point database for H_2O and CH_4.

H_2O			CH_4		
ρ	N^{mg}	σ_{Hess}	ρ	N^{mg}	σ_{Hess}
10	2500	$9.4\,10^{-6}$	50	2500	$4.4\,10^{-5}$
1	2157	$8.5\,10^{-6}$	10	1209	$4.1\,10^{-5}$
0.1	17	$1.2\,10^{-6}$	5	45	$2.3\,10^{-5}$

Similar trends have been obtained in an application to glycine, which requires "on-the-fly" (i.e. direct) dynamics. The database had to be constructed from a single trajectory due to the computational overhead of direct dynamics. Specifically, given a 5000-step trajectory started with harmonic ZPE, we employed the first 2500 steps to build the database and the final 2500 steps as our test trajectory. The trajectory was calculated at DFT/B3LYP level of theory with an aug-cc-pVDZ basis set. An interesting feature in an "on-the-fly" application is that the cost of the search algorithm versus the time needed to run a single trajectory and build a database with *ab initio* calculated Hessians is negligible. This is different from what happens for smaller molecules for which a fast-to-evaluate PES is available. The same conclusion is valid and reinforced when applying DBH to the much larger Ac-Phe-Met-NH$_2$

molecule at DFT-B3LYP-D level of theory and 6-31G* basis set. Table III reports some insightful data.

Table III. Comparison between the time needed to run a first trajectory (single core, 2500 steps) with creation of the associate ordered database ($t_{\text{traj+db}}$) and the time needed to search (t_{search}) in the database during a second test trajectory. Times are reported for molecules of increasing dimensionality.

Molecule	N_{vib}	$t_{\text{traj+db}}(s)$	$t_{\text{search}}(s)$	$t_{\text{search}}/t_{\text{traj+db}}$
H_2O	3	0.984	0.5	0.508
CH_4	9	4.663	1.5	0.322
Glycine	24	$> 10^5$	< 10	≈ 0
Ac-Phe-Met-NH$_2$	132	$> 10^6$	< 60	≈ 0

The way the Hessian database is built and employed depends on the type of dynamics adopted. When an analytical PES is available, many trajectories can be run to integrate Eq. (5) and each step of the dynamics is evolved by means of a 4-step symplectic algorithm.[83] The search for a suitable Hessian in the database (or the calculation of a new Hessian, according to the flow chart in Fig. 1) is performed at the first of the four symplectic steps. For the three remaining steps, the HU Bofill scheme is invoked. Clearly, HU Bofill can be used along the dynamics to restrict the search in the database to just a small fraction of steps. Therefore, we adopt the expression HU=N to indicate that the search is performed every N dynamics steps.

In the case of an *ab initio* on-the-fly simulation the symplectic velocity-Verlet integrator is used for the dynamics and Hessian calculations are undertaken only once the entire dynamics is complete. This allows the construction of a "predictor" code that, given the dynamics and a chosen threshold ρ, and following the flow chart of Fig. 1, returns the geometries at which the Hessian must be calculated *ab initio*. For all other geometries, a suitable match in the database is available ($\mathcal{H} = \mathcal{H}_{db}$).

III.   RESULTS

Methane

First we present an application of the Hessian database approach to the small methane molecule, for which an analytical potential energy surface is available.[82] We performed several

simulations, in which DBH was employed either at each dynamics step (DBH; HU=1) or every N dynamics steps when interfaced to Hessian Update schemes (DBH; HU=N, N>1). As anticipated in Section II, if no suitable database record was available, then a new Hessian matrix was calculated by finite differences at the first of the four symplectic steps and added to the database. For comparison purposes, additional simulations were also undertaken without using DBH (no DBH). Each simulation consisted of 33000 trajectories evolved for 2500 steps with a time step of 10 a.u. (for a total of about 600 fs). Chaotic trajectories were discarded before the end of the propagation on the basis of the value of the determinant of the monodromy matrix M. Specifically, a trajectory was rejected whenever the deviation of the determinant of $M^T M$ from its expected value of unity[60] had become greater than 1%. The different databases were constructed from the first 2000 trajectories of each simulation employing the search threshold $\rho = 5$, which led to database sizes of several thousand records as reported in Table IV.

Table IV. Application of DBH to methane with threshold $\rho = 5$ and different periodicities N of the Hessian Update Bofill scheme (HU=N). The size of the database is reported in the second column (DB Size); column 3 shows the percentage of trajectory rejection (% Rejection) given a $\det(M^T M)$ tolerance of 1%; in the last column, the percentage of rejection is reported for a simulation that does not employ DBH.

		DBH	No DBH
	DB Size	% Rejection	% Rejection
HU=1	59277	58.3	54.7
HU=2	55082	63.7	65.6
HU=5	24997	79.6	80.8
HU=10	9899	94.4	94.8

The Table lists the percentage of rejected trajectories for simulations involving the HU Bofill scheme in the presence and absence of the DBH procedure. The percentages are similar in the two instances.

Figure 2 shows a comparison between a simulation in which the Hessian has been calculated at all dynamics steps (no DBH; HU=1) and the four DBH simulations sketched in Table IV. In all these simulations, a Husimi distribution of initial phase-space conditions around the harmonic zero point energy was employed, while a coherent state centered at the equilibrium geometry and at the momentum corresponding to the harmonic ZPE was chosen as a reference state.
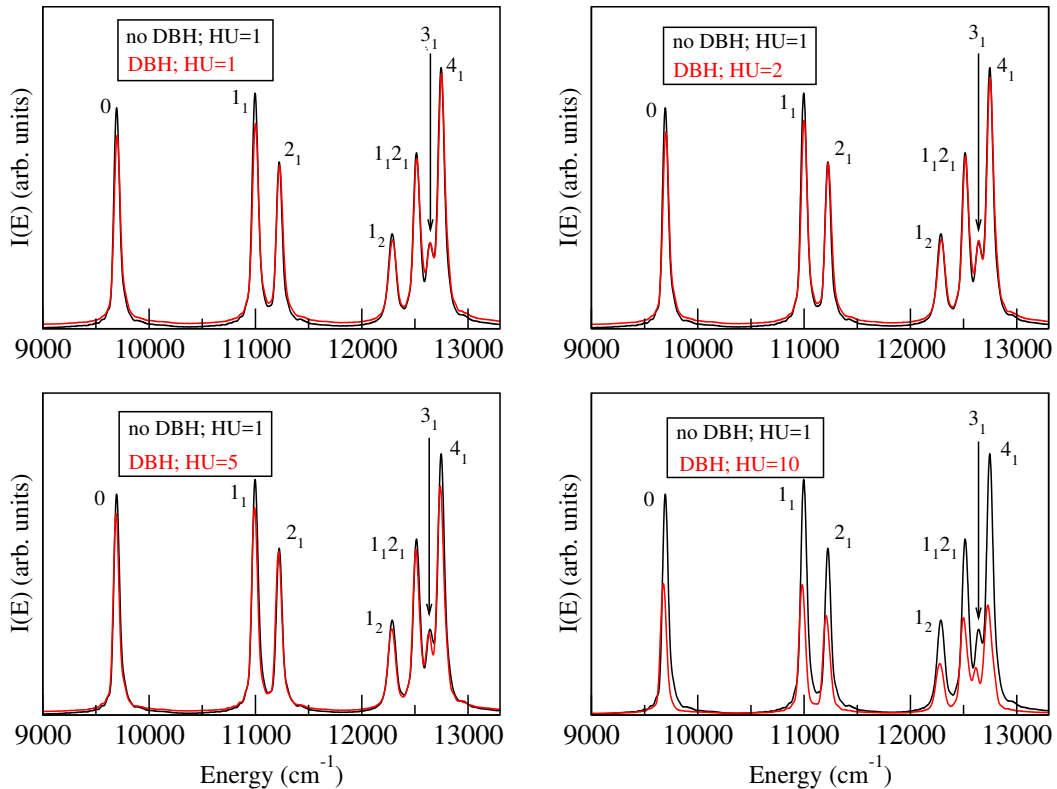
Figure 2. Comparison between the power spectrum obtained for CH_4 calculating all Hessians (no DBH; HU=1, black line) and those employing the DBH technique (top left panel, red line) and DBH in connection with Hessian Update schemes with different periodicities (other panels, red lines).

An examination of the frequencies of vibration reported in Table V leads to the observation that results obtained by means of the database technique are in good agreement with the standard approach in which all Hessians are calculated. The accuracy starts to diminish, especially regarding the ZPE value, for the DBH; HU=10 case. This was already clear from the bottom right panel of Figure 2, in which a loss in signal intensity is detected, an aspect related to the high percentage of rejected trajectories. The full widths at half maximum (FWHM) of the peaks are reported to provide a rough estimate of data uncertainty, and frequency estimates are generally well within the tolerance interval.

14

Table V. Vibrational frequencies and ZPE for CH_4 in cm^{-1}. FWHM values (cm^{-1}) are given in parentheses. Column 1 indicates the vibrational transition (0 represents the ground state). Column 2 is reserved to the quantum mechanical (QM) benchmark values from Ref. 84. Column 3 reports the outcomes of the all-Hessian simulation. The last four columns contain the results of DBH simulations. In the last row the mean absolute error (MAE) in cm^{-1} with respect to QM results is given. FWHM values are not available for the $0 \rightarrow 3_1$ transition.

Transition	QM	no DBH; HU=1	DBH; HU=1	DBH; HU=2	DBH; HU=5	DBH; HU=10
$0 \rightarrow 1_1$	1313	1302 (68)	1301 (72)	1301 (73)	1303 (69)	1304 (71)
$0 \rightarrow 2_1$	1535	1529 (69)	1528 (77)	1528 (75)	1529 (71)	1529 (80)
$0 \rightarrow 1_2$	2624	2592 (87)	2590 (95)	2591 (92)	2592 (92)	2599 (91)
$0 \rightarrow 1_1 2_1$	2836	2820 (78)	2818 (83)	2818 (80)	2819 (78)	2820 (85)
$0 \rightarrow 3_1$	2949	2948	2945	2943	2942	2938
$0 \rightarrow 4_1$	3053	3053 (76)	3051 (79)	3050 (80)	3050 (81)	3050 (93)
ZPE	9707	9694 (65)	9699 (71)	9699 (69)	9690 (65)	9677 (68)
MAE	-	11.3	12.1	12.4	13.1	14.3

Glycine

Moving to "on-the-fly" applications, we employed the database approach to study the vibrational frequencies of glycine in full dimensionality (24 degrees of freedom). Calculations were based on a single, 2500-step long trajectory with a time step of 10 a.u. *Ab initio* molecular dynamics was performed at harmonic ZPE at DFT-B3LYP level of theory with aVDZ basis set, in agreement with a previous semiclassical study of glycine.[47] For the test of the determinant of the monodromy matrix we used a threshold equal to 10^{-2}. Furthermore, a well-established regularization technique for the monodromy matrix[65] was adopted to avoid discard of the trajectory before its scheduled end. As anticipated in Section II, when performing SC spectroscopy "on-the-fly", Hessian matrices are calculated once the entire trajectory has already been determined. This allows one to exploit DBH to reduce the number of Hessian calculations similarly to HU schemes but in a more flexible way.

Table VI. ZPE energies and frequencies of fundamental transitions of glycine (cm^{-1}). MAE values (cm^{-1}) are determined with respect to the SC simulation with all Hessians calculated *ab initio* (all Hess). Columns 3-5 report the outcomes of DBH calculations based on several sets of *ab initio* Hessians. The last three columns provide estimates obtained with the HU scheme. Numbers in parentheses are the FWHM (cm^{-1}) of the associate Fourier transform signal and their average is reported in the last row.

Mode	All Hess	DBH = 3.89	DBH = 7.62	DBH = 12.69	HU = 4	HU = 8	HU=13
24	3639 (55)	3639 (56)	3639 (60)	3639 (59)	3639 (55)	3636 (60)	3637 (73)
23	3372 (59)	3372 (83)	3369 (60)	3369 (48)	3372 (60)	3363 (70)	3370 (88)
22	3375 (58)	3378 (59)	3375 (61)	3378 (67)	3375 (59)	3372 (64)	3370 (82)
21	2904 (59)	2904 (59)	2904 (61)	2904 (62)	2901 (59)	2898 (61)	2895 (72)
20	2904 (53)	2907 (53)	2901 (55)	2904 (57)	2904 (54)	2901 (57)	2902 (78)
19	1779 (54)	1779 (53)	1782 (61)	1779 (58)	1779 (54)	1779 (60)	1777 (60)
18	1662 (52)	1662 (52)	1662 (46)	1668 (47)	1662 (50)	1656 (51)	1645 (60)
17	1404 (52)	1404 (53)	1404 (66)	1401 (57)	1404 (54)	1401 (63)	1405 (61)
16	1380 (66)	1383 (69)	1386 (68)	1383 (78)	1380 (67)	1383 (66)	1378 (62)
15	1344 (55)	1347 (56)	1350 (59)	1347 (60)	1344 (56)	1347 (56)	1342 (57)
14	1287 (85)	1287 (95)	1281 (68)	1287 (100)	1284 (86)	1275 (75)	1258 (78)
13	1158 (56)	1161 (57)	1164 (67)	1158 (64)	1158 (57)	1161 (60)	1159 (60)
12	1122 (53)	1125 (53)	1131 (57)	1134 (43)	1122 (53)	1125 (53)	1120 (51)
11	1098 (59)	1101 (60)	1101 (62)	1098 (68)	1098 (59)	1098 (63)	1096 (71)
10	900 (55)	900 (55)	903 (63)	900 (61)	900 (55)	900 (60)	899 (60)
9	879 (54)	879 (53)	882 (70)	876 (64)	879 (51)	882 (62)	880 (62)
8	798 (53)	798 (54)	801 (61)	798 (59)	798 (53)	798 (58)	796 (59)
7	654 (53)	654 (57)	663 (57)	657 (64)	654 (54)	657 (53)	652 (50)
6	621 (53)	621 (52)	621 (59)	621 (57)	621 (52)	621 (57)	619 (58)
5	489 (56)	492 (56)	492 (65)	498 (63)	489 (57)	489 (59)	491 (65)
4	456 (59)	456 (59)	456 (67)	456 (68)	456 (60)	453 (65)	454 (72)
3	252 (50)	252 (49)	252 (56)	252 (53)	252 (50)	249 (56)	254 (62)
2	201 (75)	201 (83)	204 (68)	204 (105)	201 (76)	198 (75)	193 (87)
1	78 (54)	78 (54)	81 (65)	78 (56)	78 (55)	81 (59)	79 (53)
ZPE	17164 (46)	17188 (55)	17185 (62)	17230 (57)	17161 (55)	17146 (58)	17046 (63)
MAE	-	2.1	3.8	4.7	0.4	3.8	8.8
MFWHM	57	59	62	63	58	61	66

All the results presented in Table VI have excellent accuracy with respect to the benchmark semiclassical calculation based on 2500 *ab initio* Hessians. FWHM data further demonstrate the reliability of the various approximations. Due to the length of the dynamics ($\approx$ 0.6 ps), the lower bound to the FWHM values of the Fourier transform signal is about 30 cm^{-1}. Results show that most of them are actually in the 50-60 cm^{-1} range and occasionally larger. The main reasons for such an enlargement, which apply also to the case of methane, lie in nearby states contributing to the power spectrum and to spurious rotations due to the adoption for SC calculations of a normal mode reference frame based on the equilibrium geometry. Both the database approach and the HU Bofill scheme are accurate, but DBH appears more stable and deteriorates in accuracy more slowly as fewer Hessian calculations are performed *ab initio*. To better appreciate this point and the comparison among the two approaches, the ratio (R) between the maximum of 2500 Hessians and the number of those actually calculated *ab initio* is indicated in the header of DBH columns (DBH=R). More specifically, data in column 3 of Table VI are based on 643 *ab initio* Hessians, the ones in column 4 on 328 Hessians, and those in column 5 on 197 Hessians. The corresponding saving with respect to the total simulation time (including generation of the dynamics) of the all-Hessian simulations is equal to about 70%, 82%, and 88% respectively.

N-Acetyl-L-Phenylalaninyl-L-Methionine Amide

Our final spectroscopic study was dedicated to the challenging 46-atom N-acetyl-L-phenylalaninyl-L-methionine amide. The L-phenylalaninyl-L-methionine dipeptide (Phe-Met) is a prototypical system of proteins in which several hydrogen bonds govern the secondary and tertiary structures. In particular, there are three different H-bonds which are crucial in the stabilization of the folded structure: Two of them, an NH---π in the phenylalaninyl and an NH---S in the methionine, are confined within the lateral chain of the aminoacids, while the third one is established between the NH$_2$ and C=O of the terminal groups. Biswal et al. have studied recently the conformation and the vibrational spectrum of the amide of this dipeptide capped with acetyl (Ac-Phe-Met-NH$_2$). They presented quantum DFT-D calculations together with gas-phase experiments, showing that the hydrogen bond involving the sulfur atom has a strength similar to that of the intrabackbone NH---O=C hydrogen bond.[85]

Based on the satisfactory results obtained for the previous systems, we decided to employ the DBH approach. In fact, such a big and complex molecule could not be investigated with reasonable computational effort by means of the standard, all-Hessian SC procedure. DBH made the computational overhead affordable, and we were able to undertake a study based on *ab initio* "on-the-fly" semiclassical dynamics at DFT-B3LYP-D level of theory and 6-31G* basis set of the high energy quantum fundamentals of vibration. These include the 2 stretches involving the NH_2 group (sNH_2 and aNH_2) plus the two NH stretches, labeled NH(I) and NH(II). We identified the conformation illustrated in Fig. 3 as the most stable one, and normal modes were determined for this geometry.
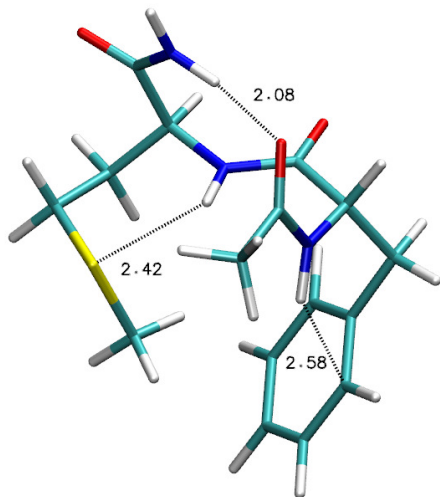
Figure 3. Global minimum of N-acetyl-L-phenylalaninyl-L-methionine amide at DFT-B3LYP-D level of theory with 6-31G* basis set. Hydrogen bonds are indicated by dashed lines.

As in the case of glycine, we performed *ab initio* calculations by means of the NWChem suite of codes,[86] but this time the MC-DC-SCIVR approach was employed. Following the standard MC-DC-SCIVR recipe, we ran four simulations tailored on the four normal modes of interest. The starting conditions were chosen to be the equilibrium geometry for positions, while initial momenta were assigned according to the harmonic ZPE with an additional quantum of harmonic excitation for the mode under investigation. The reference states were chosen as described in Eq. (6) to enhance the signal corresponding to the target mode. All the trajectories were evolved for a total of about 0.6 ps. The subspace partition was obtained in agreement with the Hessian criterion[62] and monodimensional subspaces were determined for all four modes. DBH was employed with $\rho = 0.15$, which led to the

construction of databases of less than 300 Hessians.

The molecule is made of 132 vibrational degrees of freedom with nine of them characterized by harmonic frequencies below 100 cm^{-1}, 16 in the 1500-1800 cm^{-1} range, and 19 between 3000 and 3215 cm^{-1}. It is then clear that the density of vibrational states is very high and diffuse in the 3300-3600 cm^{-1} interval investigated in this work, which includes (in addition to the four high frequency fundamentals) a large number of overtones and combination bands. Consequently some bands, due to resonances and dependent on the SC propagator (i.e. the trajectory) and the reference state, rather than well isolated peaks are expected in the power spectra. Furthermore, we found that for mode NH(I) the spectrum was very noisy due to the many couplings to other modes, and a reliable frequency estimate was not achievable. For this mode we ran a trajectory with tailored reference state at the lower harmonic ZPE to help weaken the couplings. It is known that such a simulation yields generally less accurate (even if still acceptable) frequencies, but the outcome was in this case proved to be satisfactory. Power spectra obtained from the several simulations previously described are reported in Fig. 4, where labels indicating the vibrational modes are written next to the relevant spectral signal.
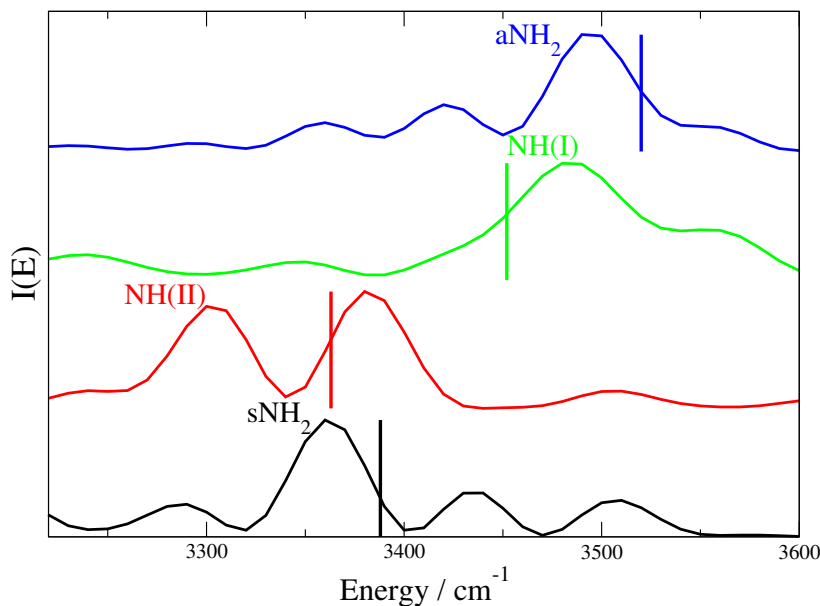

Figure 4. MC-DC-SCIVR power spectra for the 4 highest frequency fundamentals of Ac-Phe-Met-NH$_2$. Vertical bars indicate experimental values.

The corresponding semiclassical frequencies are reported in Tab. VII together with experimental results and harmonic estimates. FWHM values are given in parentheses. While harmonic values are substantially off the mark, the discrepancy between semiclassical values and the experiment is generally about 30 cm^{-1}. The distance is bigger for the NH(II) mode, which features a double peak. The lower frequency signal has been assigned to NH(II), while the other peak has been assigned to the signal of the coupled sNH$_2$ mode estimated in an approximate way using the SC propagator tailored for the NH(II) mode.

Table VII. Experimental, MC-DC-SCIVR, and harmonic frequencies of Ace-Phe-Met-NH$_2$. FWHM data are reported in parenthesis. All values are in cm^{-1}.

mode	Exp.[85]	MC-DC SCIVR	Harm
aNH$_2$	3520	3490 (47)	3682
NH(I)	3452	3480 (70)	3607
NH(II)	3363	3300 (45)	3568
sNH$_2$	3388	3360 (45)	3535
MAE	-	37	167

IV.   SUMMARY AND CONCLUSIONS

We have introduced an innovative strategy to ease the calculation of Hessian matrices required by semiclassical spectroscopy. It is based on the idea that similar Hessians are likely to be derived from similar geometries, and it consists in the dynamical construction of a database of Hessian matrices and associated geometries. The other Hessians needed are approximated by database records according to a given threshold parameter accounting for the similarity of the instantaneous geometry to those in the database. This novel approach is midway between the single-Hessian approximation recently introduced in thawed Gaussian semiclassical vibronic spectroscopy, which is generally not sufficiently accurate for vibrational spectroscopy, and the basic full-Hessian SC calculation, which is too computationally expensive for applications to medium-large systems.

For the same purpose some Hessian extrapolation schemes based on finite differences have been developed. They have been shown to work efficiently and satisfactorily for small molecules, but they depend on a linear (monodimensional) approximation and their accuracy may deteriorate in high dimensional applications. Furthermore, these Hessian update

schemes require new Hessian calculations rigorously at a constant pace along the trajectory lacking the flexibility of the database approach.

We applied the DBH technique to methane, interfacing the method with the Hessian Update Bofill scheme. This has permitted us to keep excellent accuracy while calculating only about 6% of the Hessians needed by a regular simulation. However, the small dimensionality of the molecule did not lead to a substantial speed up since the time needed by the search algorithm offsets the time saved by the reduced number of Hessian calculations. In this regard, the database approach has demonstrated its power in applications to bigger molecules for which, in absence of a precise analytical PES, *ab initio* "on-the-fly" dynamics is necessary. A first investigation of this kind involved glycine and showed that DBH permits the preservation of very good accuracy, reducing the number of Hessian matrices by a factor of about 13 and total computational times by about 88%. Hessian update schemes provided good results too, but their accuracy deteriorated faster. The ultimate demonstration of the upgrade provided by DBH is represented by the final study of the high frequency power spectrum for Ac-Phe-Met-NH$_2$. DBH has permitted the semiclassical analysis of this 46-atom system, which otherwise would have not been attainable.

Upon analysis of the results, we notice that the eigenvalues estimated by DBH are losing accuracy much faster than frequencies which are obtained as a difference between eigenvalues. This aspect does not constitute a severe drawback when moving to high dimensional systems for which the DC-SCIVR technique must be employed. In fact, "divide and conquer" SCIVR approaches are themselves less accurate in estimating ZPE energies compared to frequencies of vibration, which are the actual target of investigation.

In perspective DBH will permit semiclassical studies of complex molecular and supramolecular systems. In particular, the increased size of calculable systems will enable the investigation of molecular solvation and molecular adsorption on surfaces.

ACKNOWLEDGMENTS

REFERENCES

[1] J. M. Bowman, S. Carter, and X. Huang, Int. Rev. Phys. Chem. **22**, 533 (2003).

[2] J. M. Bowman, T. Carrington, and H.-D. Meyer, Molecular Physics **106**, 2145 (2008).

[3] X. Huang, S. Habershon, and J. M. Bowman, Chem. Phys. Lett. **450**, 253 (2008).

[4] J. S. Mancini and J. M. Bowman, J. Phys. Chem. Lett. **5**, 2247 (2014).

[5] H.-D. Meyer, U. Manthe, and L. S. Cederbaum, Chem. Phys. Lett. **165**, 73 (1990).

[6] H.-D. Meyer and G. A. Worth, Theor. Chem. Acc. **109**, 251 (2003).

[7] O. Vendrell, F. Gatti, and H.-D. Meyer, J. Chem. Phys. **127**, 184303 (2007).

[8] D. Picconi, J. A. Cina, and I. Burghardt, J. Chem. Phys. **150**, 064111 (2019).

[9] D. Picconi, J. A. Cina, and I. Burghardt, J. Chem. Phys. **150**, 064112 (2019).

[10] J. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* (Birkhauser, Boston, 2015).

[11] E. R. Davidson, J. Comp. Phys. **17**, 87 (1975).

[12] S. Manzhos and T. Carrington, J. Chem. Phys. **145**, 224110 (2016).

[13] G. Avila and T. CarringtonJr., J. Chem. Phys. **147**, 144102 (2017).

[14] V. Barone, J. Chem. Phys. **122**, 014108 (2005).

[15] F. Gabas, G. Di Liberto, R. Conte, and M. Ceotto, Chem. Sci. **9**, 7894 (2018).

[16] A. Masson, E. R. Williams, and T. R. Rizzo, J. Chem. Phys. **143**, 104313 (2015).

[17] H.-B. Oh, C. Lin, H. Y. Hwang, H. Zhai, K. Breuker, V. Zabrouskov, B. K. Carpenter, and F. W. McLafferty, J. Am. Chem. Soc. **127**, 4076 (2005).

[18] M. Micciarelli, R. Conte, J. Suarez, and M. Ceotto, J. Chem. Phys. **149**, 064115 (2018).

[19] M. Micciarelli, F. Gabas, R. Conte, and M. Ceotto, J. Chem. Phys. **150** (2019).

[20] W. H. Miller, J. Chem. Phys. **61**, 1823 (1974).

[21] E. J. Heller, J. Chem. Phys. **75**, 2923 (1981).

[22] M. F. Herman and E. Kluk, Chem. Phys. **91**, 27 (1984).

[23] W. H. Miller, J. Chem. Phys. **95**, 9428 (1991).

[24] K. G. Kay, J. Chem. Phys. **101**, 2250 (1994).

[25] A. R. Walton and D. E. Manolopoulos, Chem. Phys. Lett. **244**, 448 (1995).

[26] Y. Elran and K. Kay, J. Chem. Phys. **110**, 3653 (1999).

[27] F. Grossmann, Comments At. Mol. Phys. **34**, 141 (1999).

[28] D. V. Shalashilin and M. S. Child, J. Chem. Phys. **115**, 5367 (2001).

[29] W. H. Miller, Proc. Natl. Acad. Sci. USA **102**, 6660 (2005).

[30] K. G. Kay, Annu. Rev. Phys. Chem. **56**, 255 (2005).

[31] J. Liu and W. H. Miller, J. Chem. Phys. **127**, 114506 (2007).

[32] M. Ceotto, S. Atahan, G. F. Tantardini, and A. Aspuru-Guzik, J. Chem. Phys. **130**, 234113 (2009).

[33] M. Ceotto, D. Dell' Angelo, and G. F. Tantardini, J. Chem. Phys. **133**, 054701 (2010).

[34] M. Ceotto, G. F. Tantardini, and A. Aspuru-Guzik, J. Chem. Phys. **135**, 214108 (2011).

[35] D. Tamascelli, F. S. Dambrosio, R. Conte, and M. Ceotto, J. Chem. Phys. **140**, 174109 (2014).

[36] M. Wehrle, S. Oberli, and J. Vaníček, J. Phys. Chem. A **119**, 5685 (2015).

[37] M. Buchholz, F. Grossmann, and M. Ceotto, J. Chem. Phys. **148**, 114107 (2018).

[38] X. Ma, G. Di Liberto, R. Conte, W. L. Hase, and M. Ceotto, J. Chem. Phys. **149**, 164113 (2018).

[39] R. Conte and M. Ceotto, *Semiclassical Molecular Dynamics for Spectroscopic Calculations* (Wiley, book chapter, accepted).

[40] M. S. Church, S. V. Antipov, and N. Ananth, J. Chem. Phys. **146**, 234104 (2017).

[41] S. Bonella, D. Montemayor, and D. F. Coker, Proc. Natl. Acad. Sci. **102**, 6715 (2005).

[42] G. Tao and W. H. Miller, J. Chem. Phys. **135**, 024104 (2011).

[43] G. Tao and W. H. Miller, J. Chem. Phys. **137**, 124105 (2012).

[44] G. Tao, Theor. Chem. Acc. **133**, 1448 (2014).

[45] G. Tao, J. Phys. Chem. A **117**, 5821 (2013).

[46] R. Conte, A. Aspuru-Guzik, and M. Ceotto, J. Phys. Chem. Lett. **4**, 3407 (2013).

[47] F. Gabas, R. Conte, and M. Ceotto, J. Chem. Theory Comput. **13**, 2378 (2017).

[48] G. Di Liberto, R. Conte, and M. Ceotto, J. Chem. Phys. **148**, 104302 (2018).

[49] M. Ceotto, S. Atahan, S. Shim, G. F. Tantardini, and A. Aspuru-Guzik, Phys. Chem. Chem. Phys. **11**, 3861 (2009).

[50] J. Tatchen and E. Pollak, J. Chem. Phys. **130**, 041103 (2009).

[51] M. Ceotto, S. Valleau, G. F. Tantardini, and A. Aspuru-Guzik, J. Chem. Phys. **134**,

234103 (2011).

[52] M. Wehrle, M. Sulc, and J. Vanicek, J. Chem. Phys. **140**, 244114 (2014).

[53] A. Patoz, T. Begusic, and J. Vanicek, J. Phys. Chem. Lett. **9**, 2367 (2018).

[54] S. Pratihar, X. Ma, Z. Homayoon, G. L. Barnes, and W. L. Hase, J. Am. Chem. Soc. **139**, 3570 (2017).

[55] D. Marx and J. Hutter, *Ab initio molecular dynamics: basic theory and advanced methods* (Cambridge University Press, 2009).

[56] D. Marx and M. Parrinello, J. Chem. Phys. **104**, 4077 (1996).

[57] D. R. Galimberti, A. Milani, M. Tommasini, C. Castiglioni, and M.-P. Gaigeot, Journal of chemical theory and computation **13**, 3802 (2017).

[58] M.-P. Gaigeot, Phys. Chem. Chem. Phys. **12**, 3336 (2010).

[59] A. L. Kaledin and W. H. Miller, J. Chem. Phys. **118**, 7174 (2003).

[60] A. L. Kaledin and W. H. Miller, J. Chem. Phys. **119**, 3078 (2003).

[61] M. Ceotto, G. Di Liberto, and R. Conte, Phys. Rev. Lett. **119**, 010401 (2017).

[62] G. Di Liberto, R. Conte, and M. Ceotto, J. Chem. Phys. **148**, 014307 (2018).

[63] K. G. Kay, J. Chem. Phys. **100**, 4432 (1994).

[64] R. Gelabert, X. Giménez, M. Thoss, H. Wang, and W. H. Miller, J. Phys. Chem. A **104**, 10321 (2000).

[65] G. Di Liberto and M. Ceotto, J. Chem. Phys. **145**, 144107 (2016).

[66] H. Wu, M. Rahman, J. Wang, U. Louderaj, W. Hase, and Y. Zhuang, J. Chem. Phys. **133**, 074101 (2010).

[67] Y. Zhuang, M. R. Siebert, W. L. Hase, K. G. Kay, and M. Ceotto, J. Chem. Theory Comput. **9**, 54 (2012).

[68] M. Ahmadian, Y. Zhuang, W. L. Hase, and Y. Chen, Big Data Research **9**, 57 (2017).

[69] M. Ceotto, Y. Zhuang, and W. L. Hase, J. Chem. Phys. **138**, 054116 (2013).

[70] E. J. Heller, J. Chem. Phys. **62**, 1544 (1975).

[71] M. Baranger, M. A. de Aguiar, F. Keck, H.-J. Korsch, and B. Schellhaass, J. Phys. A **34**, 7227 (2001).

[72] E. Pollak and S. Miret-Artés, J. Phys. A **37**, 9669 (2004).

[73] F. Grossmann, J. Chem. Phys. **125** (2006).

[74] R. Conte and E. Pollak, Phys. Rev. E **81**, 036704 (2010).

[75] M. Buchholz, F. Grossmann, and M. Ceotto, J. Chem. Phys. **144**, 094102 (2016).

[76] M. Buchholz, F. Grossmann, and M. Ceotto, J. Chem. Phys. **147**, 164110 (2017).

[77] M. Buchholz, C.-M. Goletz, F. Grossmann, B. Schmidt, J. Heyda, and P. Jungwirth, J. Phys. Chem. A **116**, 11199 (2012).

[78] T. Begusic, J. Roulet, and J. Vanicek, J. Chem. Phys. **149**, 244115 (2018).

[79] T. Begusic, M. Cordova, and J. Vanicek, J. Chem. Phys. **150**, 154117 (2019).

[80] M. Buchholz, E. Fallacara, F. Gottwald, M. Ceotto, F. Grossmann, and S. D. Ivanov, Chemical Physics **515**, 231 (2018).

[81] J. M. Bowman, A. Wierzbicki, and J. Zuniga, Chem. Phys. Lett. **150**, 269 (1988).

[82] T. J. Lee, J. M. Martin, and P. R. Taylor, J. Chem. Phys. **102**, 254 (1995).

[83] M. L. Brewer, J. S. Hulme, and D. E. Manolopoulos, J. Chem. Phys. **106**, 4832 (1997).

[84] S. Carter, H. M. Shnider, and J. M. Bowman, J. Chem. Phys. **110**, 8417 (1999).

[85] H. S. Biswal, E. Gloaguen, Y. Loquais, B. Tardivel, and M. Mons, J. Phys. Chem. Lett. **3**, 755 (2012).

[86] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, Comput. Phys. Commun. **181**, 1477 (2010).