THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

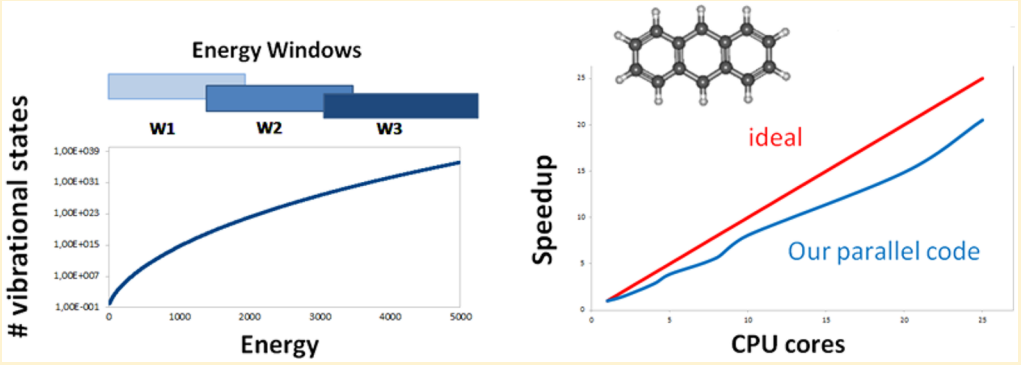
# An Efficient Computational Approach for the Calculation of the Vibrational Density of States

Chiara Aieta,[†] Fabio Gabas,[†,§] and Michele Ceotto*[,†]

[†]Dipartimento di Chimica, Università degli Studi di Milano, via C. Golgi 19, 20133 Milano, Italy

[§]CINECA - Interuniversity Computing Center Supercomputing Applications and Innovation Department - SCAI Via R. Sanzio, 4, 20090 Segrate, Milan, Italy

Supporting Information




**ABSTRACT:** We present an optimized approach for the calculation of the density of fully coupled vibrational states in high-dimensional systems. This task is of paramount importance, because partition functions and several thermodynamic properties can be accurately estimated once the density of states is known. A new code, called paradensum, based on the implementation of the Wang−Landau Monte Carlo algorithm for parallel architectures is described and applied to real complex systems. We test the accuracy of paradensum on several molecular systems, including some benchmarks for which an exact evaluation of the vibrational density of states is doable by direct counting. In addition, we find a significant computational speedup with respect to standard approaches when applying our code to molecules up to 66 degrees of freedom. The new code can easily handle 150 degrees of freedom. These features make paradensum a very promising tool for future calculations of thermodynamic properties and thermal rate constants of complex systems.

## INTRODUCTION

The vibrational density of states of a chemical system is the number of vibrational states per unit of energy. This quantity is of paramount importance in molecular and chemical physics. All information about energetics, thermodynamics at equilibrium, as well as molecular spectroscopy can be obtained from the knowledge of the density of states $\rho(E)$. More specifically, the vibrational density of states yields the vibrational partition function upon Laplace transform

$$Q_{vib}(\beta) = \int \rho_{vib}(E) e^{-\beta E} dE \quad (1)$$

where $\beta = 1/k_B T$, $k_B$ being the Boltzmann constant and $T$ the temperature of the system. $Q_{vib}(\beta)$ is used to compute not only specific heats but also thermal rate constants. More generally, statistical kinetic theories of dynamical processes, such as energy transfer and reactivity,[1−6] employ the density of states in their formulation.

Unfortunately, quantum densities of states can be calculated exactly by direct counting of quantum states only for low-dimensional systems or can be obtained only for a restricted number of separable systems. When the BS algorithm is combined with the Stein Rabinovitch (SR) one, the combination of the two (BSSR) enables very fast calculations for separable models of high-dimensional systems.[7] Such models may include not only harmonic degrees of freedom (consisting of uncoupled harmonic oscillators), but also highly anharmonic ones, such as Morse oscillators or hindered internal rotations. An alternative route consists in inverting eq 1 and calculate $\rho_{vib}(E)$ as the Inverse Laplace Transform (ILT) of the canonical partition function. Clearly, this approach is advantageous when the partition function can be determined via computationally affordable Monte Carlo methods. In this paper, we will test the ILT approach and show that its successful application is limited in dimensionality and in energy range, as already pointed out elsewhere.[8] A more widely used

**Special Issue:** Piergiorgio Casavecchia and Antonio Lagana Festschrift

**Received:** December 17, 2015
**Revised:** February 3, 2016
**Published:** February 3, 2016






DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

method is based on Steepest Descent approximations.[9,10] Nevertheless, ILT is a viable tool, which has already been employed in several cases to estimate microcanonical densities of states.[11−14] Other more manageable (but approximated) approaches have been developed in the past. For example, in the so-called Simple Perturbation Theory (SPT) of Isaacson et al.,[15] an effective harmonic frequency for each mode is defined as

$$\bar{\nu}_i = \nu_i + 2X_{ii} + \frac{1}{2}\sum_{j\neq i} X_{ij} \tag{2}$$

where $\nu_i$ is the harmonic frequency of the $i$-th mode, and $X_{ij}$ is the $i,j$ element of the matrix of anharmonic couplings. SPT approximates $\rho_{vib}(E)$ as that of a set of uncoupled harmonic oscillators of frequencies provided by eq 2.

In general, to calculate $\rho_{vib}(E)$ directly and with high accuracy for large systems is a difficult task. The challenge ahead of us is to include not only anharmonicities but also intermode couplings and quantum effects in the calculation of $\rho_{vib}(E)$. This goal can be reached, at least in part, by using a perturbative approach, where the energy of the quantized vibrational levels is expanded around a stationary point up to the second order, including the anharmonic coupling terms as follows

$$E = \sum_{i=1}^{nvib} \omega_i\left(v_i + \frac{1}{2}\right) + \sum_{i,j=1}^{nvib} X_{ij}\left(v_i + \frac{1}{2}\right)\left(v_j + \frac{1}{2}\right) \tag{3}$$

where $E$ is the vibrational energy, nvib is the number of vibrational degrees of freedom, $\omega_i$ is the $i$-th element of the vector $\boldsymbol{\omega}$ of harmonic frequencies, $v_i$ is the vibrational quantum number of the $i$-th mode, and $X_{ij}$ is the $i,j$ element of the X-matrix of anharmonic couplings. In principle, adoption of eq 3 would permit the calculation of $\rho_{vib}(E)$ via Monte Carlo sampling.[16−19] However, a straightforward application of the Monte Carlo approach is hindered by a drastic increase of the variance with the number of degrees of freedom.[18,19] As an alternative, Basire et al.[20] first, and Nguyen and Barker[21] later, showed that the Wang−Landau (WL) method,[22] in which Monte Carlo walker visits are driven by the criterion of uniform histogram of visits, can successfully be applied to calculate not only classical but also quantum $\rho_{vib}(E)$. In particular, Barker et al.[23] implemented the WL approach for $\rho_{vib}(E)$ calculations in the code adensum of the MultiWell suite. This program suite can also solve the internal energy master equation for complex unimolecular reactions systems, calculate exact $\rho_{vib}(E)$ (doloops code) and approximate ones using the BSSR algorithm (code DenSum), and estimate thermal rate constants and other thermodynamic quantities.[24]

Quite recently, there has been a significant boost in the WL application provided by its parallelization.[25−31] Vogel et al.[25] showed how the WL algorithm can be systematically implemented for application to large-dimensional problems. Their parallelization idea is based on the splitting of the total energy range into smaller windows with large overlaps between adjacent windows. Multiple independent walkers are sampled in each energy window, and replica exchanges between walkers across overlapping windows are allowed. There is no *a priori* limit on the number of windows, and this parallel WL scheme can scale up to thousands of CPUs. We find this hierarchical parallel framework potentially advantageous for vibrational density of states calculations.

This paper describes how to best implement the parallel WL scheme in a code (called paradensum) for application to large molecular systems. First, we show the details of this implementation and describe the paradensum pseudo code. Then, we test paradensum for analytical and exactly solvable systems up to hundreds of degrees of freedom. Finally, we apply paradensum to real molecules with up to 66 vibrational degrees of freedom, and we show that it can potentially be employed for even larger systems. Whenever possible, results are compared with exact calculations. In the last section, we briefly discuss results and conclude the paper.

## THE PARADENSUM ALGORITHM

The quantum molecular density of vibrational states $\rho_{vib}(E)$ can be estimated by counting the number of vibrational eigenstates $\Delta N_{vib}(E)$ per energy interval $\Delta E$ within the interval $[E, E + \Delta E]$, i.e.

$$\rho_{vib}(E) = \Delta N_{vib}(E)/\Delta E \tag{4}$$

To perform the counting, the exact vibrational eigenvalues should be known in advance. When eigenvalues are calculated with grid methods, such as Discrete Variable Representation methods,[32−34] the complexity of the system is usually confined to a few atoms. For higher-dimensional systems, one can either rely on semiclassical approximations[35−43] or perturbation expansions around the molecular minimum geometry.[44]

The perturbation approach approximates the vibrational energy with the well-known Dunham expansion.[16,17] Here, nvib normal modes are coupled to each other and quantum effects are included by the quantum numbers $\mathrm{v} = (v_1,v_2,...,v_i,...,v_{nvib})$. Since the first occupied quantum state is at the zero point energy value ($E_{ZPE}$), a convenient expression for the vibrational energy relative to the zero-point one ($E'$) can be derived from eq 3

$$E' = E - E_{ZPE} = \sum_{i=1}^{nvib} \omega_i\left(v_i + \frac{1}{2}\right) + \sum_{i,j=1}^{nvib} X_{ij}\left(v_i + \frac{1}{2}\right)\left(v_j + \frac{1}{2}\right) - E_{ZPE} \tag{5}$$

Harmonic frequencies and the X-matrix have to be supplied as input and can be computed with quantum chemistry codes (see, for instance, refs 45,46). All combinations of vibrational quantum numbers v providing bounded states, i.e. with energy in between the ZPE and the dissociation energy, are counted. However, eq 5 is not accurate near the dissociation threshold or above. The number of states $\Delta N_{vib}(E)$ for each interval of vibrational energies between $E$ and $E+\Delta E$ will provide the density of states $\rho_{vib}(E)$ according to eq 4.

The doloops code in MultiWell is based on a direct counting of the number of vibrational states by using as many nested do cycles on the vibrational quantum numbers as the number of degrees of freedom. The algorithm is useful for small-size systems (up to 4 atoms), whereas for larger systems, the computational overhead quickly becomes unaffordable. For this reason, Basire et al.[20] first, and Nguyen and Barker[21] later, implemented the calculation of $\rho_{vib}(E)$ within a Wang Landau Monte Carlo scheme.[22] The algorithm is based upon the observation that the histogram of visits $H(E)$ will be flat when the probability of visiting each energy level for a random walk in the space of the quantum numbers is proportional to $1/\rho_{vib}(E)$. In this way, the random walk is forced to visit regions


DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

with low density of states and is not trapped in high density zones. The Wang Landau algorithm is rigorously proven to converge,[47] and it has been shown to be even suitable for building the density of states of rough energy landscapes.[22] After a starting guess $\rho_{vib}(E) = 1$, random walks are performed in energy space with an acceptance probability $p$ from a vibrational state with energy $E_i$ to another at $E_f$ given by

$$p(E_i \rightarrow E_f) = \min\left(\frac{\rho_{vib}(E_i)}{\rho_{vib}(E_f)}, 1\right) \quad (6)$$

Once a given energy bin $E_i$ is visited, $\rho_{vib}(E_i)$ is multiplied by the modification factor $f$. Initially, $f_0 = e$. During each random walk, the histogram $H(E)$ of visits of each energy level is updated together with $\rho_{vib}(E)$. The procedure halts when the flatness of $H(E)$ is within a cutoff value. Once the flatness criterion for the j-th iteration is satisfied, the modification factor is reduced using a monotonically decreasing function of the type $f_{j+1} = \sqrt{f_j}$, and a new random walk begins starting from the last estimate of the density of states $\rho_{vib,j}(E)$. A sort of WL rule of thumb[22,25] is to take the whole algorithm completed after 21 such iterations, because after that, the modification factor has reached a value really close to 1 and the density of states does not change anymore when it is updated in the limit of numerical precision used in molecular system calculations. The main advantage of the WL algorithm is that it can be employed for the calculation of the density of states of any system.[48] Regarding the problem of calculating the density of vibrational states of a molecule, $\rho_{vib}(E)$ is normalized by setting the width of the first ground state energy bin to contain only the ground vibrational state.

The adensum code of the MultiWell suite[23,24] represents the state-of-the-art for molecular density of states calculations. Each quantum number $v_k$ of the $k$-th degree of freedom has an upper limit provided by the equation $\partial E/\partial v_k = 0$, which enforces the system to be bound along the $k$-th degree of freedom.[21] Vibrational modes associated with hindered rotations are convoluted as rotations in the vibrational states counting. Nguyen et al.[21] also tested the possible expression of the transition probability of eq 6 and they finally set $p = \mathrm{MIN}(1/N, 0.25)$, since the result standard deviation is quite stable with respect to the choice of $p$. The critical issues of the adensum code are that the number of random walkers over the entire energy space is restricted to a single one and that there is no flatness requirements over the histogram, since the total number of Monte Carlo iterations is fixed either at $10^2$, $10^3$, $10^4$, or $10^5$ times the energy bin number.[21]

To overcome the limitations of the MultiWell adensum code, we modify the present algorithm structure and then implement it for parallel architectures by using the MPI (Message Passing Interface) API (Application Programming Interface). The code structure does not allow for a straightforward parallelization, since the outer *do cycle* is the evaluation of $\rho_{vib}(E)$, with the $j$-th iteration starting from the converged $\rho_{vib,j-1}(E)$ value of the $(j-1)$-th iteration. Instead, the parallelization strategy we employ is to divide the energy range into windows and set the calculations of $\rho_{vib}(E)$ for each window on a single processor. This strategy is dictated by the observation that less random sampling events and reduced computational effort is needed for Monte Carlo convergence in a restricted energy range. We choose all windows to have the same width. The number of processors is given by the choice of windowing. Since the WL density of states is notoriously biased at the edges of small energy windows,[25] we always keep a percentage of overlap between neighbor windows. We call the new code paradensum, and its structure is sketched in Figure 1. The code supports the possibility to run multiple walkers for each energy window. A preliminary single walk is performed on the entire energy range to pick up the initial configuration of each walker in each window. Given the uniformly distributed probability in energy space, this single walk is quite efficient in providing initial quantum numbers configurations. After choosing the starting configuration in each window as described above, unique random number generator seeds are created from the processor ID for each task and random walks start. In the adensum code, $f$ is updated after a fixed number of Monte Carlo sweeps. Instead, in paradensum the WL flatness criterion is applied and monitored separately for each window. In case of multiple walkers, they average out their estimate of $\rho_{vib}(E)$ before the update of the modification factor $f$ and the beginning of a new iteration. This setup makes paradensum quite flexible and able to fit the energy domain better than the adensum code. As a general strategy to better handle the data in this parallelization scheme, we set arrays of dimensionality equal to the number of windows. $\rho_{vib}(E)$ and $H(E)$ are represented as three-dimensional arrays where the indexes define the number of grains per window, the number of walkers per window and, a third index, the window number. In this way, each processor accesses the array part corresponding to its own window without interferring with other tasks. At the end of the random walks, MPI reduction is invoked to merge each processor array into a global unique one that includes the results of all walkers. This strategy is depicted in Figure 2. Following Barker et al.,[21] we set the number of energy bins for the lowest energy window such that the lowest energy bin contains just a single quantum state, i.e. the ZPE one. In this way $\rho_{vib}(E_{ZPE}) = 1/\Delta E$ for that energy bin and the values of $\rho_{vib}(E)$ for the other bins are rescaled accordingly. A matching between the overlapping energy window bins allows to extend this normalization to all windows. The final rescaled density profile $T_{tot}(E)$ is built by joining the $\rho_{vib}(E)$ of different windows where the inverse microcanonical temperature $\mathrm{d}\ln(\rho_{vib}(E))/\mathrm{d}E$ matches more

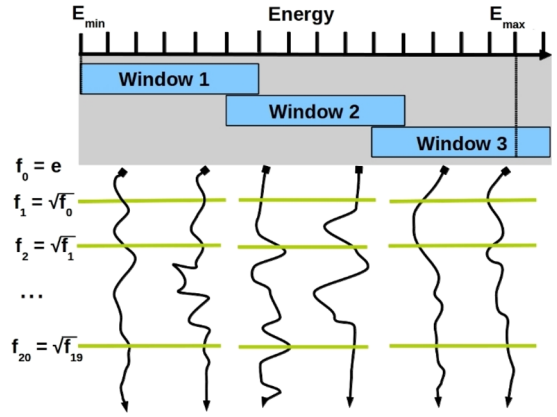



**Figure 1.** Paradensum code structure. $E_{min}$ and $E_{max}$ are, respectively, the minimum and maximum energy $E$ for $\rho_{vib}(E)$ evaluation. The energy range is divided into overlapping windows, and each energy window contains the same number of bins. Black continuous lines represent random Monte Carlo walks, and green lines the averaging of $\rho_{vib}(E)$ and the rescaling of the modification factor.



DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

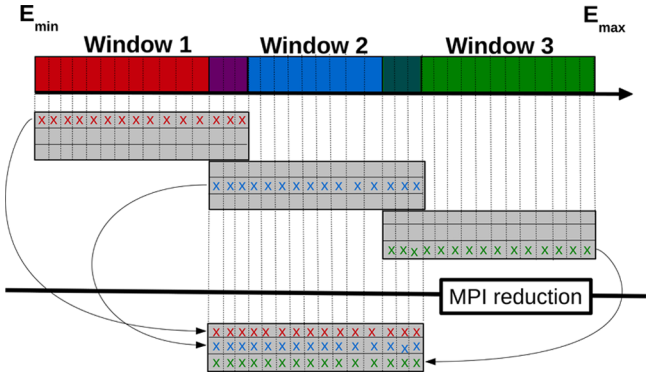



**Figure 2.** MPI reduction scheme.

accurately, as described in ref 25. The paradensum pseudo code is reported in Figure 3.

## RESULTS AND DISCUSSION

In this section, we describe the results of the calculation of $\rho_{vib}(E)$ for several systems. First, we consider model systems to check the correctness and the scalability of the code paradensum. Then, we justify the choice to perform calculations by means of a single walker per energy window with strict requirement for the histogram flatness. Finally, we turn into real and complex molecule calculations to show the computational advantages of paradensum with respect to other codes for density of states calculations.

**Algorithm 1** ParAdensum pseudo-code.

**Require:** $n$, number of windows
**Require:** $w$, number of walkers
**Require:** $o$, overlap between neighbour windows
**Require:** $fp$, flatenss percentage

Divide the whole energy range in $n$ windows with overlap $o$
Initialize: $f = e$, $g_{n,w}(E) = 0$, $H_{n,w}(E) = 0\ \forall\ n, w$
Associate the window $n$ to the process $n$
**for** $i = 1, 21$ **do**
  **for** $j = 1, w$ **do**
    Initialize random number generator with process dependent seed
    **if** $i = 1$ **then**
      Find an acceptable **v** in the window by WL in whole energy range
    **else**
      Retrieve walker information
    **end if**
    **while** $\min[H_{n,w}(E)] > \overline{H_{n,w}(E)} * fp$ **do**
      Select new **v** varying old one
      **if** **v** has an energy in the process window **then**
        Accept or reject the step
        $g_{n,w}(E) = g_{n,w}(E) + \ln(f)$ and $H_{n,w}(E) = H_{n,w}(E) + 1$
      **end if**
      **if** number of iterations exceeds a cut off value **then**
        Exit from while loop
      **end if**
    **end while**
    Save information of the walker
  **end for**
  Average $g_{n,w}(E)$ over all walkers in the window
  Redistribute $\overline{g_{n,w}(E)}$ to all $w$ walkers in the window
  $f = \sqrt{f}$
**end for**
Reinitialize walkers information
Normalize $g_{n,w}(E)$ respect to the first non-zero element in the overlap to obtain $T_{n,w}(E)$
MPI reduction of $g_{n,w}(E)$, $H_{n,w}(E)$ and $T_{n,w}(E)$
Scale $T_{n,w}(E)$ sequentially
Join $T_{n,w}(E)$ from all windows to make $T_{tot}(E)$ with derivatives method

**Figure 3.** Paradensum pseudo code. Here the variable $g$ indicates the logarithm of the vibrational density of states.

**Model Systems.** The calculation of the density of states $\rho_{vib}(E)$ for harmonic model systems, i.e. uncoupled harmonic oscillators, is only in part analytical. While the harmonic partition function is analytical, $\rho_{vib}(E)$ is obtained by a convolution in energy space. This integration is not analytical unless the single mode density of states is approximated. This is the case of the classical approximation where the one-dimensional harmonic $\rho_{vib}(E) = 1/\hbar\omega$ is convoluted into the $N$-dimensional expression to obtain

$$\rho_{vib}(E) = \frac{E^{N-1}}{(N-1)!\prod_{j=1}^{N}\hbar\omega_j} \tag{7}$$

which is famously employed in the classical RRKM theory.[49] Alternatively, we can consider that the cumulative sum of states $N_{vib}(E)$ is related to the partition function by the inverse Laplace transform:

$$N_{vib}(E) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}\frac{e^{\beta E}}{\beta 2^N\prod_{j=1}^{N}\sinh\left(\frac{1}{2}\beta\hbar\omega_j\right)}d\beta \tag{8}$$

and that $\rho_{vib}(E) = dN_{vib}(E)/dE$. To exploit this route, we perform the inverse Laplace transform numerically, using an algorithm based on an expansion in Fourier Series.[50] We found this approach useful to manage the cumulative sum of states up to 20 uncoupled harmonic oscillators. The main limitation of this approach is that the inversion of the Laplace transform is numerically stable only within a small energy range, and reliable results of $N_{vib}(E)$ are limited to energies up to around 1500 $cm^{-1}$ above the zero point energy. Given the inaccuracy of eq 7,[9,10,51,52] an alternative approach for higher energies, the Whitten and Rabinovitch semiempirical approximation, can be employed.[53,54]

In Figure 4, we consider a 1-d harmonic oscillator. Paradensum correctly reproduces the results provided by the exact counting of states. The classical results, labeled as CL, reproduce on average the discrete quantum mechanical counting given by the straight line. The ILT can manage the

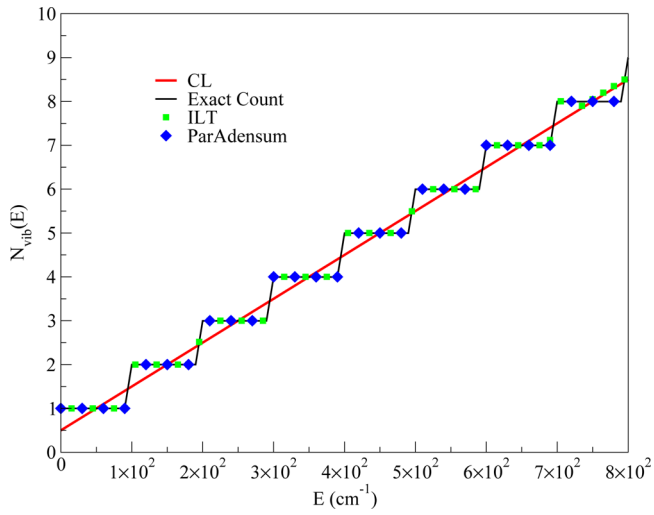



**Figure 4.** Cumulative sum of the density of states $N_{vib}(E) = \int_0^E \rho_{vib}(E')dE'$ for the one-dimensional harmonic oscillator. Exact results are the "Exact Count" ones. CL stands for classical, paradensum for the present WL calculations and ILT for the inverse Laplace transform approach.


DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

exact counting up to an energy threshold. For higher energy, ILT results are on the top of the CL results. However, within this limitation, ILT is going to be useful for higher-dimensionality systems where the "Exact Count" approach becomes unfeasible.

We then consider the calculation of the cumulative sum of vibrational states $N_{vib}(E)$ for uncoupled harmonic systems of increasing dimensionality. The harmonic frequencies have been chosen randomly within a range from 100 to 4000 $cm^{-1}$ which is representative of typical molecular vibrations, and the anharmonicity matrix elements $X_{ij}$ are set to 0. Up to four dimensions, exact calculations can be performed. The typical quantum mechanical staircase profile of $N_{vib}(E)$ is still visible for such a low dimensionality. As reported on the top panel of Figure 5, paradensum and ILT faithfully reproduce the staircase

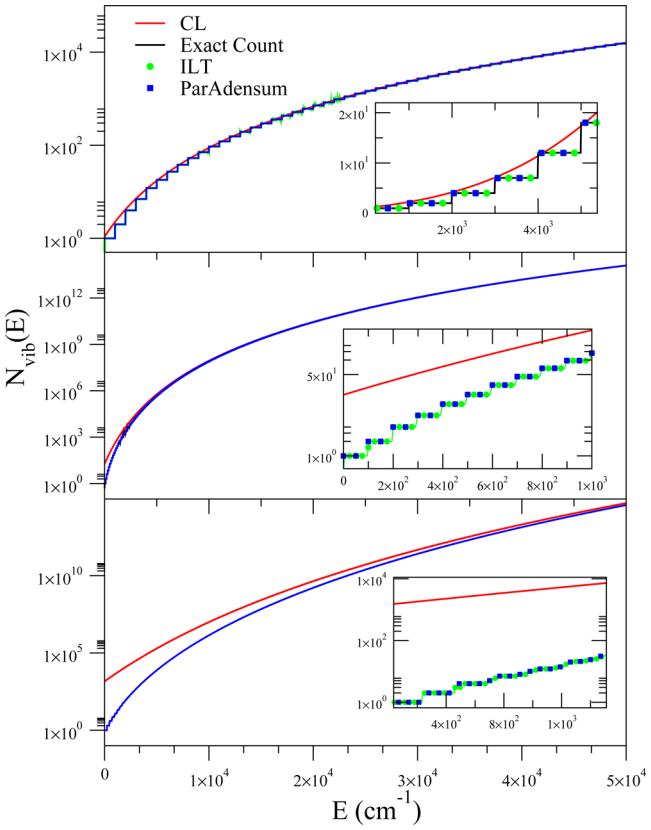



**Figure 5.** Cumulative sum of states $N_{vib}(E)$ calculations. Four (upper panel), 10 (middle panel) and 20 (lower panel) uncoupled harmonic oscillators. "Exact Count" are the exact results (continuous black line), CL stands for the classical approximation of eq 7 (continuous red line), and ILT for the Numerical Inversion of Laplace transform of eq 8 (green diamonds). Paradensum results are reported as blue squares.

exact results. Instead, CL calculations overestimate $\rho_{vib}(E)$ at low energies while reproducing the exact results at higher energy ranges. The reasons of this deviation are explained by the approximation employed to obtain eq 7. The convolution summation over all possible quantum numbers is approximated as an integration over a continuum of energies and corresponding fictitious fractional quantum numbers. This overcounting is severe at low energies where the density of states is low, while it is moderated at higher energies where the density of states is higher and the variation over the quantum numbers generates a quasi-continuum of states. A comparison between the three panels in Figure 5 also shows that the CL approximation at low energies becomes less accurate as the dimensionality is increased. Paradensum and ILT are in excellent agreement up to 20 dimensions. When the dimensionality is further increased, the ILT approach can no longer be applied, and an implementation of the SPT of eq 2 is not even possible.

In Figure 6, we look at the cumulative sum of states $N_{vib}(E)$ for a system of 80 (Figure 6a,b) and 150 (Figure 6c,d) uncoupled harmonic oscillators. We check how paradensum performs for different windowing strategy and which WL flatness criterion is more suitable. For this purpose, we compare the left panels calculations where a 80% (Figure 6a,c) of flatness criterion for the histogram $H(E)$ is applied to right ones where a 95% (Figure 6b,d) flatness criterion is employed. For each panel, $N_{vib}(E)$ is calculated with different numbers of windows and reported in logarithmic scale. The inset of each panel reports the percentage deviations for different choices of window numbers. An increase in WL flatness criterion significantly reduces standard deviations of the windowing and guarantees that the results are independent of the windowing choice. We will show below how computational costs decrease when increasing the number of the windows. The original adensum code is limited to systems of about 100 degrees of freedom, while paradensum can easily allocate 150 degrees of freedom, and this is not yet an upper bound, since further increment in the dimensionality of the problem is achievable.

The main conclusion of this section is that paradensum is not only able to reproduce exact cumulative vibrational sums of states for low-dimensionality systems, as calculated by means of the "Exact Count" approach, but it can also be successfully applied to significantly more complex systems.

**Influence of Flatness Criterion and Multiple Walkers.** We now investigate the possibility to exploit multiple walkers as a possible parallelization strategy. We chose the water molecule as a test case, since the exact $\rho_{vib}(E)$ can be obtained by direct counting. We employed the experimental vibrational data for the exact counting.[55] Figure 7 reports the percent error deviation of different paradensum setups. For all simulations reported, the energy range has been divided into 20 windows made of 338 energy bins each (including the overlap). We observe that the accuracy is roughly linear with respect to the energy variation, except for a few bins. However, accuracy is still preserved within 35%.

We wonder if greater accuracy can be achieved by increasing the number of walkers. We performed these tests for a given 80% flatness of the WL algorithm and compare the results to a single-walker 95% flatness simulation. Indeed, the more numerous the walkers employed, the greater the accuracy. However, Figure 7 clearly shows that a single-walker simulation with a 95% flatness criterion provides by far more accurate results than simulations based on a larger number of walkers, but with a weaker flatness constraint. Computational overheads must also be taken into account for a fair assessment of the most efficient setup. A typical WL simulation for water at 80% of flatness requires $\sim 10^4$ Monte Carlo sweeps for each $f$ iteration, for a total of $2.1 \times 10^5$ sweeps. Considering a parallelization strategy where multiple walkers are placed on different cores with different seeds, the total computational cost is a multiple of the number of walkers. For instance, if 100 walkers are employed, the number of total Monte Carlo sweeps is $2.1 \times 10^7$. In comparison, when simulating the same molecule


DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

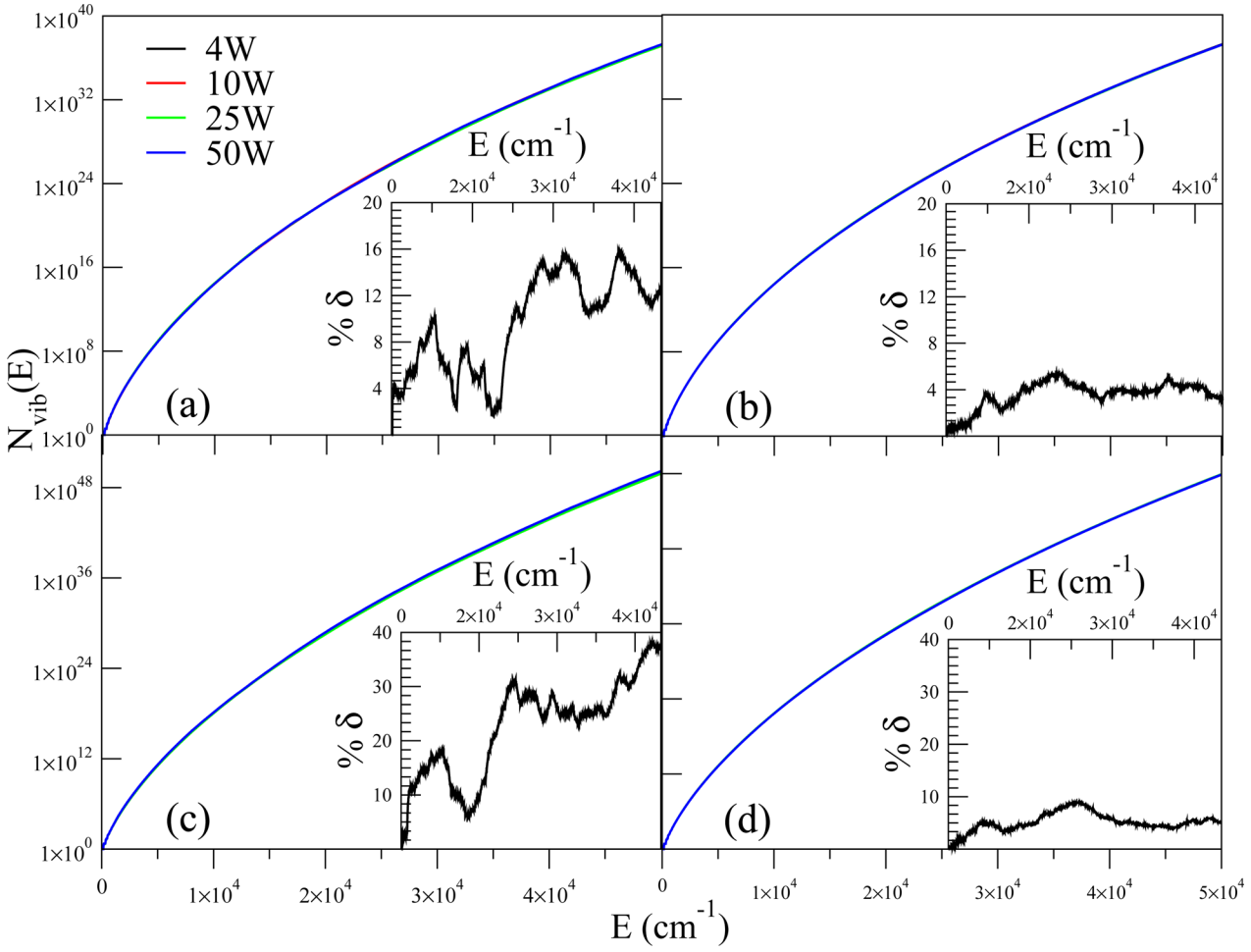



**Figure 6.** Paradensum code results for uncoupled harmonic oscillators systems with increasing degrees of freedom. (a) and (b) refer to 80 oscillators, and (c) and (d) refer to 150. Flatness criterion is 80% for (a) and (c), 95% for (b) and (d). Insets report the values of the percentage deviation of $N_{vib}(E)$ calculated with different window partitioning $\delta(E) = 100 \times \sqrt{\sum_{i=1}^{W} (N_{vib,i}(E) - \overline{N_{vib}(E)})^2} / \overline{N_{vib}(E)}$ where $\overline{N_{vib}(E)} = \frac{1}{W} \sum_{i=1}^{W} N_{vib,i}(E)$ and $W$ is the number of windows.

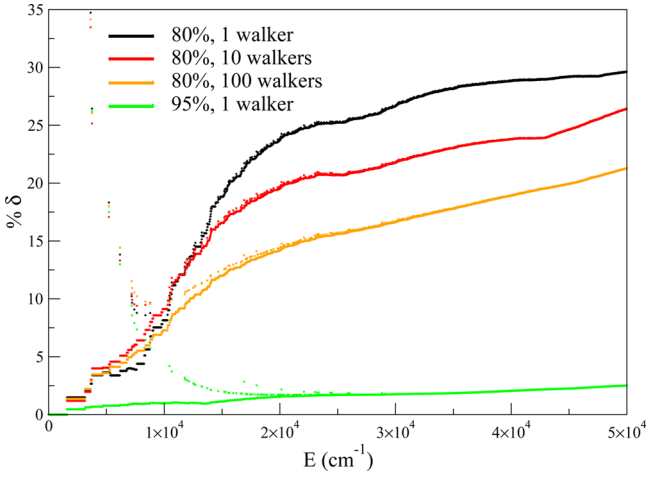



**Figure 7.** Percentage error deviation for $N_{vib}(E)$ calculation of water molecule. Black, red, and orange dots are for WL 80% flatness simulations employing respectively 1, 10, and 100 walkers. Green dots are for a single walker WL simulation at 95% flatness criterion.

with the much stricter flatness criterion of 95% but with a single walker, after 21 iterations, we find a total number of Monte Carlo sweeps of $\sim 3 \times 10^4$, which is almost 3 orders of magnitude smaller than the 100-walker simulation. With respect to these considerations, it is without a doubt more important to adopt a parallelization strategy based on energy windowing, as we have done in our code paradensum, than one based on multiple walkers.

**Molecules.** We now turn to real and more complex molecule calculations, aiming at estimating the cumulative sum of the number of vibrational states in the bound-state energy region. We have investigated several molecules of increasing complexity, i.e. the HOCO radical, *N*-methylmethanimine ($CH_2$═$NCH_3$), naphtalene ($C_{10}H_8$), triethylphosphine (P-($CH_3CH_2$)$_3$), and anthracene ($C_{14}H_{10}$). The number of vibrational degrees of freedom involved are 6, 18, 48, 60, and 66 respectively. Results are reported in Figure 8, where each panel shows the cumulative sum of vibrational states and the percentage deviation of the windowing strategy as defined above (see caption of Figure 6).

The needed harmonic frequencies $\boldsymbol{\omega}$ and anharmonic $X_{ij}$ coupling parameters have been calculated using either the Gaussian09[45] or Cfour[46] codes and are reported in the Supporting Information. In particular, Gaussian09 has been employed to calculate frequencies and anharmonic coupling parameters at the MP2/cc-pvdz level for $CH_2$═$NCH_3$ and


DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

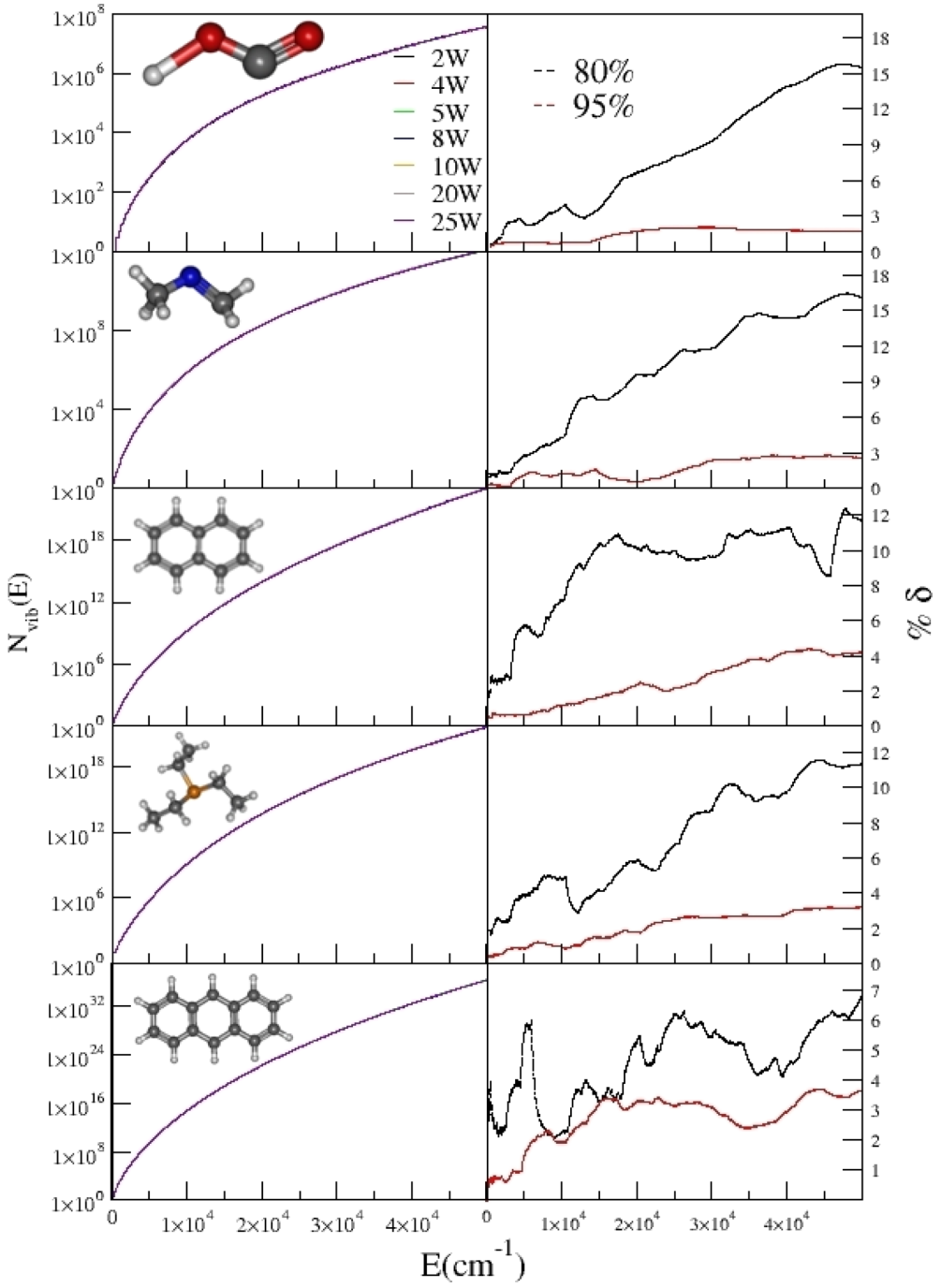



**Figure 8.** Cumulative sum of states for five molecules of increasing complexity: HOCO radical, *N*-methylmethanimine, naphtalene, triethylphosphine, and anthracene. In the right column, the percent deviations of $N_{vib}(E)$ are reported for a WL flatness choice of 80% (black dashed) and 95% (red dashed). A single walker per energy window has been employed.

$C_{10}H_8$, and B3LYP/cc-pvdz level for $P(CH_3CH_2)_3$ and $C_{14}H_{10}$. Furthermore, we have employed Cfour at the CCSD(T)/ANO1 level for the HOCO radical. Results for different numbers of windows are within 10% using a flatness criterion of 80%, and the agreement is stricter, i.e. within 5%, for a 95% flatness choice. This value is comparable to what is expected from the WL Monte Carlo statistical deviation, so it clears the field from any possible systematic error that may have been introduced by windowing or parallelization. In other words, such a statistical interval of confidence for systems of dimensionality up to 66 degrees of freedom proves the reliability of the parallelization strategy adopted in the paradensum code. Higher-energy ranges are not plotted because eq 5 cannot be applied.

**Timing and Scalability.** We now look at the computational time scaling of the code paradensum with the number of cores employed for the systems presented in the previous Sections. We have performed single-walker simulations with a 95% flatness constraint. The computational speed up generated by partitioning the energy range into windows is reported on the two upper panels of Figure 9.

The efficiency of the strategy based on multiple windows is analyzed on the lower panels of the same figure, and it is calculated as the ratio between the corresponding windowing



DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

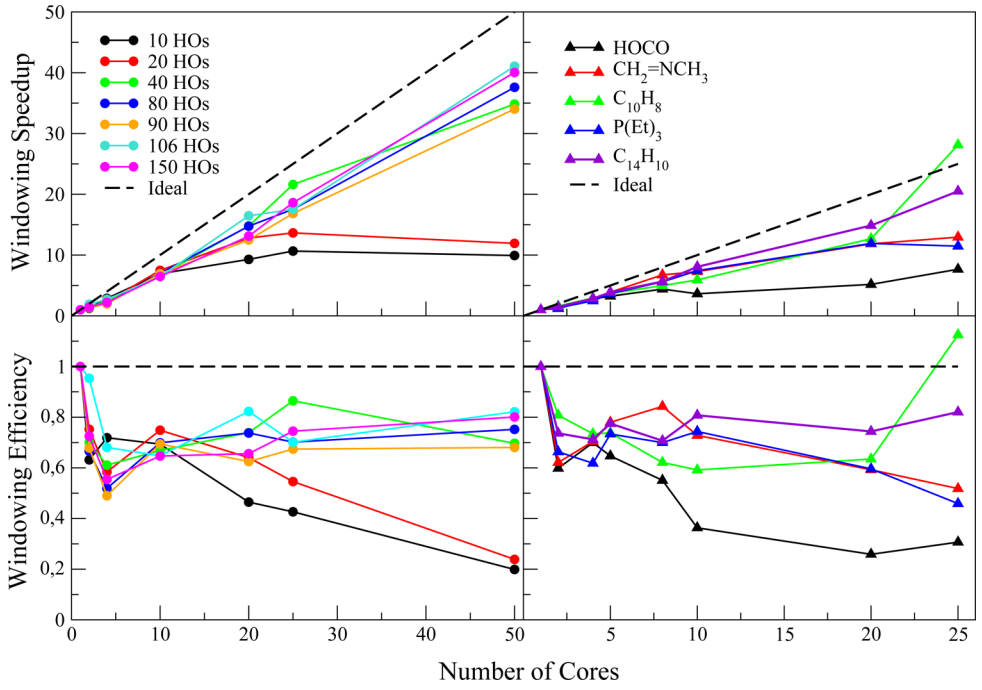



**Figure 9.** Computational speedup and efficiency with respect to the number of cores for several systems. Left panels report results for uncoupled harmonic oscillators, and the right panels are for molecules.

speedup and the number of processors. We start by looking at the harmonic oscillators simulations reported on the left panels. Here the computational speed up is almost ideal (dashed line), except for the relatively low-dimensional 10- and 20-harmonic oscillator systems. The almost ideal scaling makes sense by considering that the energy space for a set of uncoupled harmonic oscillators is quite trivial. The unfavorable scaling of the lower-dimensional systems can be explained by the more accentuated sparsity of their energy levels at low energy. As a consequence, bins within the low-energy windows are characterized by bigger differences in the density of states, thus requiring a higher number of Monte Carlo sweeps before the flatness criterion is satisfied, and creating a bottleneck for the whole calculation. Indeed, the lower left panel confirms that paradensum outperforms serial codes like adensum, even if its efficiency deteriorates when too many energy windows are employed for the lower-dimensional systems. When considering real molecules, the rationalization of the results is more cumbersome, because one needs to consider that the spacing of the vibrational levels is quite different for different molecules. We observe again that for small molecules, the parallelization is not very efficient, as reported on the lower right panel. From Figure 9, it is clear that the strategy adopted in paradensum becomes more and more convenient as the number of degrees of freedom of the molecule increases. However, the speedup does not monotonically increase with the number of degrees of freedom. For example, the parallelization efficiency for naphtalene ($C_{10}H_8$) is greater than that of both $P(CH_3CH_2)_3$ and $C_{14}H_{10}$, even if the vibrational space dimensionality of naphtalene is smaller. Actually, naphtalene presents superscalability, as it scales with parallelization even better than a set of uncoupled harmonic oscillators. We believe that the reason for naphtalene superscalability lies in the rigidity of this molecule, which is responsible for its quite high vibrational frequencies if compared to those of other molecules where internal floppy modes are present. As described above, the frequencies of the uncoupled harmonic oscillators were chosen randomly in a range of frequencies that includes floppy modes. This explains the better scalability of naphtalene even with respect to the harmonic systems tested.

## SUMMARY AND CONCLUSIONS

This paper introduces a computational approach for the calculation of the vibrational density of states of molecular systems. To exploit the possibility of calculations on parallel architectures and open up the possibility to calculate the density of states of high-dimensional systems, we have adapted a WL algorithm parallelization strategy.[25] After describing in detail the implementation of the algorithm (see Figure 3) that we call paradensum, we have tested it on a common ground with other codes and assessed its accuracy. We have then applied the new code to real molecules and examined its computational speedup and efficiency. We have found paradensum able to exhibit almost ideal efficiency and to be significantly advantageous with respect to other codes commonly employed to deal with high-dimensional systems. Remarkably, in the case of naphtalene, paradensum exhibits a superscalability trend. We believe that the method described in this paper is an important and useful tool for the physical chemists community, since the density of states is ubiquitous and of paramount importance in physical chemistry applications. To foster widespread diffusion and application of paradensum, the code is freely available upon request to the authors[56] and it is implemented in MultiWell suite of codes.[24]

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.5b12364.

Input data of program paradensum consisting of harmonic frequencies, X-matrixes, and separable degrees



DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

of freedom for all the molecules treated in the paper (PDF)

## AUTHOR INFORMATION


### Corresponding Author

*E-mail: michele.ceotto@unimi.it. Tel: +39 0250314258.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We acknowledge the CINECA and the Regione Lombardia award under the LISA initiative (grant SURGREEN) for the availability of high-performance computing resources. We also thank the Chemistry Department for funding through the Development Plan of Athenaeum grant−line B1 (UNIAGI 17777). Dr. Riccardo Conte is warmly thanked for useful discussions and for revising the manuscript.

## REFERENCES


(1) Nordholm, S.; Freasier, B. C.; Jolly, D. L. Ergodic Collision Theory of Intermolecular Energy Transfer. *Chem. Phys.* **1977**, *25*, 433−449.

(2) Freasier, B. C.; Jolly, D. L.; Nordholm, S. Ergodic Collision Theory of Intermolecular Energy Transfer II. Quantum Effects in the Harmonic Approximation. *Chem. Phys.* **1978**, *32*, 161−168.

(3) Lenzer, T.; Luther, K.; Nilsson, D.; Nordholm, S. J. PECT Model Analysis and Predictions of Experimental Collisional Energy Transfer Probabilities P(E′,E) and Moments ⟨ΔE⟩ for Azulene and Biphenylene. *J. Phys. Chem. B* **2005**, *109*, 8325−8331.

(4) Conte, R.; Houston, P. L.; Bowman, J. M. Collisional Energy Transfer in Highly Excited Molecules. *J. Phys. Chem. A* **2014**, *118*, 7742−7757.

(5) Houston, P. L.; Conte, R.; Bowman, J. M. A Model for Energy Transfer in Collisions of Atoms with Highly Excited Molecules. *J. Phys. Chem. A* **2015**, *119*, 4695−4710.

(6) Greene, S. M.; Shan, X.; Clary, D. C. Reduced-Dimensionality Semiclassical Transition State Theory: Application to Hydrogen Atom Abstraction and Exchange Reactions of Hydrocarbons. *J. Phys. Chem. A* **2015**, *119*, 12015−12027.

(7) Beyer, T.; Swinehart, D. Algorithm 448: Number of Multiply-restricted Partitions. *Commun. ACM* **1973**, *16*, 379. Stein, S.; Rabinovitch, B. Accurate Evaluation of Internal Energy Level Sums and Densities Including Anharmonic Oscillators and Hindered Rotors. *J. Chem. Phys.* **1973**, *58*, 2438−2445.

(8) Hüpper, B.; Pollak, E. Numerical Inversion of the Laplace Transform. *J. Chem. Phys.* **1999**, *110*, 11176−11186.

(9) Forst, W. *Theory of Unimolecular Reactions*; Academic Press: New York, 1973.

(10) Forst, W. *Unimolecular Reactions. A Concise Introduction*; Cambridge University Press: Cambridge, 2003.

(11) Parneix, P.; Van-Oanh, N. T.; Bréchignac, P. Estimation of the Anharmonic Quantum Density of States in Large Systems. *Chem. Phys. Lett.* **2002**, *357*, 78−84.

(12) Börjesson, L. E. B.; Nordholm, S.; Andersson, L. L. Quantized Vibrational Densities of States and Ergodic Energy Transfer in Molecular Collisions. *Chem. Phys. Lett.* **1991**, *186*, 65−72.

(13) Krems, R.; Nordholm, S. Z. A Thermodynamic Method of Estimating Anharmonic Molecular Densities of States. *Z. Phys. Chem.* **2000**, *214*, 1467−1477.

(14) Wadi, H.; Pollak, E. Accurate Computation of Quantum Densities of States and RRKM Rate Constants for Large Polyatomic Molecules: The STAIR Method. *J. Chem. Phys.* **1999**, *110*, 8246−8253.

(15) Isaacson, A. D.; Truhlar, D. G.; Scanlan, K.; Overend, J. Tests of Approximation Schemes for Vibrational Energy Levels and Partition Functions for Triatomics: H2O and SO2. *J. Chem. Phys.* **1981**, *75*, 3017−3024. Truhlar, D. G.; Isaacson, A. D. Simple Perturbation Theory Estimates of Equilibrium Constants from Force Fields. *J. Chem. Phys.* **1991**, *94*, 357−359.

(16) Barker, J. R. Sums of Quantum States of Nonseparable Degrees of Freedom: Muiltidimensional Monte Carlo Integration. *J. Phys. Chem.* **1987**, *91*, 3849−3854.

(17) Toselli, B. M.; Barker, J. R. RO-Vibrational State Densities Based on Spectroscopic Data for Non-separable Systems. *Chem. Phys. Lett.* **1989**, *159*, 499−504.

(18) Noid, D. W.; Koszykowski, M. L.; Tabor, M.; Marcus, R. M. Properties of Vibrational Energy Levels in the Quasi Periodic and Stochastic Regimes. *J. Chem. Phys.* **1980**, *72*, 6169−6175.

(19) Farantos, S. C.; Murrell, J. N.; Hajduk, J. C. Monte Carlo Calculations of Classical Density of States for Non-separable Polyatomic Potential Energy Surfaces. *Chem. Phys.* **1982**, *68*, 109−117.

(20) Basire, M.; Parneix, P.; Calvo, F. Quantum Anharmonic Densities of States Using the Wang−Landau Method. *J. Chem. Phys.* **2008**, *129*, 081101.

(21) Nguyen, T. L.; Barker, J. R. Sums and Densities of Fully Coupled Anharmonic Vibrational States: A Comparison of Three Practical Methods. *J. Phys. Chem. A* **2010**, *114*, 3718−3730.

(22) Wang, F.; Landau, D. P. Efficient, Multiple-range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053. Wang, F.; Landau, D. P. Determining the Density of States for Classical Statistical Models: A Random Walk Algorithm to Produce a Flat Histogram. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2001**, *64*, 056101.

(23) Barker, J. R. Multiple-Well, Multiple-path Unimolecular Reaction Systems. I. MultiWell Computer Program Suite. *Int. J. Chem. Kinet.* **2001**, *33*, 232−245.

(24) Barker, J. R.; Ortiz, N. F.; Preses, J. M.; Lohr, L. L.; Maranzana, A.; Stimac, P. J.; Nguyen, T. L.; Kumar, T. J. D. *MultiWell Program Suite*, Version 2014.1b; University of Michigan: Ann Arbor, MI, 2014; http://aoss-research.engin.umich.edu/multiwell/ (accessed June 25, 2014).

(25) Vogel, T.; Li, Y. W.; Wüst, T.; Landau, D. P. Generic, Hierarchical Framework for Massively Parallel Wang-Landau Sampling. *Phys. Rev. Lett.* **2013**, *110*, 210603.

(26) Yin, J.; Landau, D. P. Massively Parallel Wang-Landau Sampling on Multiple GPUs. *Comput. Phys. Commun.* **2012**, *183*, 1568−1573.

(27) Poulain, P.; Calvo, F.; Antoine, R.; Broyer, M.; Dugourd, P. Performances of Wang-Landau Algorithms for Continuous Systems. *Phys. Rev. E* **2006**, *73*, 056704.

(28) Valentim, A.; Rocha, J. C. S.; Tsai, S.-H.; Li, Y. W.; Eisenbach, M.; Fiore, C. E.; Landau, D. P. Exploring Replica-exchange Wang-Landau Sampling in Higher-dimensional Parameter Space. *J. Phys.: Conf. Ser.* **2015**, *640*, 012006.

(29) Pattanasiri, B.; Li, Y. W.; Wust, T.; Landau, D. P. Effect of Surface Attractive Strength on Structural Transitions of a Confined HP Lattice Protein. *J. Phys.: Conf. Ser.* **2015**, *640*, 012015.

(30) Shi, G.; Wüst, T.; Li, Y. W.; Landau, D. P. Protein Folding of the HOP Model: A Parallel Wang—Landau Study. *J. Phys.: Conf. Ser.* **2015**, *640*, 012017.

(31) Koh, Y. W.; Sim, A. Y. L.; Lee, H. K. Dynamical Traps in Wang-Landau Sampling of Continuous Systems: Mechanism and Solution. *Phys. Rev. E* **2015**, *92*, 023306.

(32) Colbert, D. T.; Miller, W. H. A Novel Discrete Variable Representation for Quantum Mechanical Reactive Scattering Via the S-matrix Kohn Method. *J. Chem. Phys.* **1992**, *96*, 1982−1991.

(33) Meyer, H.-D.; Gatti, F.; Worth, G. A. *Multidimensional Quantum Dynamics: MCTDH Theory and Applications*; Wiley-VCH: Weinheim, 2009.

(34) Manthe, U. The Multi-configurational Time-dependent Hartree Approach Revisited. *J. Chem. Phys.* **2015**, *142*, 244109.

(35) Miller, W. H. The Semiclassical Initial Value Representation: A Potentially Practical Way for Adding Quantum Effects to Classical Molecular Dynamics Simulations. *J. Phys. Chem. A* **2001**, *105*, 2942. Miller, W. H. Quantum Dynamics of Complex Molecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6660.




DOI: 10.1021/acs.jpca.5b12364
J. Phys. Chem. A 2016, 120, 4853−4862

(36) Conte, R.; Pollak, E. Comparison Between Different Gaussian Series Representations of the Imaginary Time Propagator. *Phys. Rev. E* **2010**, *81*, 036704.
(37) Kay, K. G. Semiclassical Initial Value Treatments of Atoms and Molecules. *Annu. Rev. Phys. Chem.* **2005**, *56*, 255−280.
(38) Conte, R.; Aspuru-Guzik, A.; Ceotto, M. Reproducing Deep Tunneling Splittings, Resonances, and Quantum Frequencies in Vibrational Spectra From a Handful of Direct Ab Initio Semiclassical Trajectories. *J. Phys. Chem. Lett.* **2013**, *4*, 3407−3412.
(39) Tamascelli, D.; Dambrosio, F. S.; Conte, R.; Ceotto, M. Graphics Processing Units Accelerated Semiclassical Initial Value Representation Molecular Dynamics. *J. Chem. Phys.* **2014**, *140*, 174109.
(40) Zhuang, Y.; Siebert, M. R.; Hase, W. L.; Kay, K. G.; Ceotto, M. Evaluating the Accuracy of Hessian Approximations for Direct Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 54−64.
(41) Ceotto, M.; Dell'Angelo, D.; Tantardini, G. F. Multiple Coherent States Semiclassical Initial Value Representation Spectra Calculations of Lateral Interactions for CO on Cu (100). *J. Chem. Phys.* **2010**, *133*, 054701.
(42) Sulc, M.; Vanicek, J. Accelerating the Calculation of Time-resolved Electronic Spectra with the Cellular Dephasing Representation. *Mol. Phys.* **2012**, *110*, 945−955.
(43) Conte, R.; Pollak, E. Continuum Limit Frozen Gaussian Approximation for the Reduced Thermal Density Matrix of Dissipative Systems. *J. Chem. Phys.* **2012**, *136*, 094101.
(44) Santoro, F.; Cappelli, C.; Barone, V. Effective Time-independent Calculations of Vibrational Resonance Raman Spectra of Isolated and Solvated Molecules Including Duschinsky and Herzberg−Teller Effects. *J. Chem. Theory Comput.* **2011**, *7*, 1824−1839.
(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H. et al. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2009.
(46) *CFOUR*, a quantum chemical program written by Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. with contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J. et al. For the current version, see http://www.cfour.de (accessed July, 2015).
(47) Zhou, C.; Bhatt, R. N. Understanding and Improving the Wang-Landau Algorithm. *Phys. Rev. E* **2005**, *72*, 025701(R).
(48) Landau, D. P.; Tsai, S. H.; Exler, M. A New Approachto Monte Carlo Simulations in Statistical Physics: Wang-Landau Sampling. *Am. J. Phys.* **2004**, *72*, 1294−1302.
(49) Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. *Chemical Kinetics and Dynamics*, 2nd ed.; Prentice-Hall: Upper Saddle River, NJ, 1999; pp 94−97.
(50) D'Amore, L.; Laccetti, G.; Murli, A. Algorithm 796: A Fortran Software Package for the Numerical Inversion of the Laplace Transform Based on a Fourier Series Method. *ACM T MATH SOFTWARE* **1999**, *25*, 306−315.
(51) Holbrook, K. A.; Pilling, M. J.; Robertson, S. H. *Unimolecular Reactions*, 2nd ed.; Wiley: Chichester, 1996.
(52) Robinson, P. J.; Holbrook, K. A. *Unimolecular Reactions*; Wiley-Interscience: London; New York, 1972.
(53) Whitten, G. Z.; Rabinovitch, B. S. Accurate and Facile Approximation for Vibrational Energy-Level Sums. *J. Chem. Phys.* **1963**, *38*, 2466−2473.
(54) Whitten, G. Z.; Rabinovitch, B. S. Approximation for Rotation—Vibration Energy Level Sums. *J. Chem. Phys.* **1883**, *1964*, 41.
(55) Benedict, W. S.; Gailar, N.; Plyler, E. K. Rotation-Vibration Spectra of Deuterated Water Vapor. *J. Chem. Phys.* **1956**, *24*, 1139−1165.
(56) Ceotto Group Website. http://users.unimi.it/ceotto/ (accessed January, 2016).


DOI: 10.1021/acs.jpca.5b12364
*J. Phys. Chem. A* 2016, 120, 4853−4862