Cent. Eur. J. Chem. • 9(4) • 2011 • 549-556
DOI: 10.2478/s11532-011-0042-6

Central European Journal of **Chemistry**

Vibrational calculations in formaldehyde: the CH stretch system

Invited Paper

Svetoslav Rashev[1], David C. Moule[2]

[1]*Department of Optics&Spectroscopy, Institute of Solid State Physics,*
Bulgarian Academy of Sciences, Sofia 1784, Bulgaria

[2]*Department of Chemistry, Brock University, St. Catharines,*
ON, L2S3A1, Canada

Received 21 January 2011; Accepted 7 March 2011

Abstract: An alternative procedure for the calculation of highly excited vibrational levels in S0 formaldehyde was developed to apply to larger molecules. It is based on a new set of symmetrized vibrational valence coordinates. The fully symmetrized vibrational kinetic energy operator is derived in these coordinates using the Handy expression [Molec. Phys. 61, 207 (1987)]. The potential energy surface is expressed as a fully symmetrized quartic expansion in the coordinates. We have performed *ab initio* electronic computations using GAMESS to obtain all force constants of the S0 formaldehyde quartic force field. Our large scale vibrational calculations are based on a fully symmetrized vibrational basis set, in product form. The vibrational levels are calculated one by one using an artificial intelligence search/selection procedure and subsequent Lanczos iteration, providing access to extremely high vibrational energies. In this work special attention has been given to the CH stretch system by calculating the energies up to the fifth CH stretch overtone at ~16000 cm^{-1}, but the method has also been tested on two highly excited combination levels including other lower frequency modes.

Keywords: *Formaldehyde • Vibrational calculations • Vibrational overtones • Lanczos tridiagonalization*

© Versita Sp. z o.o.

1. Introduction

Formaldehyde is one of the most thoroughly spectroscopically studied polyatomic molecules [1-6] and has long served as a model for large scale vibrational calculations [7-19]. Using a variety of techniques (IR, stimulated emission pumping, dispersed fluorescence) 276 ground electronic state vibrational levels (up to 12500 cm^{-1}) have been observed and assigned [3-6]. Even higher CH stretch overtones have been observed by photoacoustic spectroscopy and analyzed theoretically [20].

There have been a number of *ab initio* determinations of the S_0 ground electronic state equilibrium geometry and force field [21-24], with the 1993 quartic force field of Martin, Lee and Taylor (MLT) [23] achieving almost spectroscopic precision. The first exact variational calculation of vibrational levels in formaldehyde using an exact analytical expression for the kinetic energy [25] and the MLT quartic field was carried out by Carter, Pinnavaia and Handy [11]. There have been numerous adjustments and refinements of the MLT quartic force field as well as equilibrium geometry parameters (two bond lengths and one interbond angle) based on extensive vibrational calculations [11-13]. Burleigh *et al*. [14] performed a calculation and adjustment to experimental data [6] of 138 vibrational levels in formaldehyde up to ~7600 cm^{-1}, using 6th order canonical Van Vleck perturbation theory. More recently, large scale vibrational calculations were carried out using a generalized discrete variable representation and adiabatic contraction techniques [15]. It is noteworthy that using the MLT quartic force field and their symmetrized coordinates, Luckhaus [15] has almost exactly reproduced all the vibrational frequencies up to ~5700 cm^{-1} previously calculated by Carter *et al.* [11], who used a totally different finite set variational procedure.

Bramley and Carrington [10] developed a general variational method based on the Lanczos algorithm without storage of the Hamiltonian matrix and calculated formaldehyde vibrational levels up to ~5700 cm^{-1}. Poulin *et al*. [13] employed the recursive residue generation

---

* *E-mail: rashev@issp.bas.bg*

method of Wyatt to carry out converged calculations up to about 6300 cm^{-1}. Mladenovic [19] carried out accurate large scale calculations on levels up to ~5200 cm^{-1} using discrete variable representation (DVR) and a contraction scheme resulting from several diagonalization/ truncation steps. Ribeiro *et al*. [16] developed a prediagonalization-perturbation Davidson algorithm to evaluate selected states up to 9500 cm^{-1}. Recently, Bernal and Lemus [17] calculated 260 levels up to 12500 cm^{-1} using an effective Hamiltonian based on an algebraic approach within a SU(2) representation of coordinates and momenta. Lee and Light [18] applied their iterative solutions/energy selected bases variational method to calculate 729 A_1 symmetry states up to 13500 cm^{-1} based on a slightly modified version of the Burleigh *et al*. adjusted potential surface [14]. This method seems very promising for calculations on larger molecules and/or higher energies.

Earlier, Luckhaus *et al*. [20] studied the higher CH stretching overtones up to about 16000 cm^{-1} both experimentally and by using normal mode modeling based on *ab initio* potential energy and dipole moment surfaces computation.

In this work we describe an alternative method for H_2CO frequencies calculation equally suitable for low and highly excited vibrational states. However, the quartic PES is not of global character and will certainly not hold for too high excitation energies; it has been tested for the CH stretch overtones up to about 16000 cm^{-1}. Its most important characteristics are: (i) employment of an exact kinetic energy operator (derived from the Handy [25] expression) expressed as a completely symmetrized product in terms of our vibrational coordinates; (ii) employment of our own *ab initio* quartic force field (slightly adjusted to reproduce the 6 fundamental frequencies exactly); (iii) a variational calculation based on a set of symmetrized basis functions in product form, including a symmetrized local mode (LM) representation for the CH stretch modes; (iv) employment of an AI/SS procedure to obtain the most economic Hamiltonian matrix, complemented by a Lanczos matrix manipulation.

2. Theoretical derivation

2.1. Vibrational coordinates and Hamiltonian.

The atom configuration and notation is given in Supplementary Fig. 1. These are somewhat different from the MLT coordinates S_k [23], especially the angular coordinates.

The vibrational coordinates Q_k for the formaldehyde S_0 ground electronic state are the Simons-Parr-Finlan stretching coordinates [26] ρ_{CH}, $\rho_{CH'}$ and ρ_{CO}, as in MLT [23]:

$$Q_1(A_1) = S_1(A_1) = \frac{\rho_{CH} + \rho_{CH'}}{\sqrt{2}},\ Q_5(B_2) = S_5(B_2) = \frac{\rho_{CH} - \rho_{CH'}}{\sqrt{2}},$$

$$Q_2(A_1) = S_2(A_1) = \rho_{CO},$$

where $\rho_{CH} = \frac{\Delta r_{CH}}{r_{CH}},\ \rho_{CO} = \frac{\Delta r_{CO}}{r_{CO}};\ r_{CH} = \Delta r_{CH} + r_0(CH),\ r_{CO} = \Delta r_{CO} + {}$

${} + r_0(CO)$, $\Delta r_{CH}\,(\Delta r_{CH'})$ and Δr_{CO} are the stretches of the C–H (C–H') and C–O bonds from their equilibrium lengths r_0(CH) and r_0(CO).

Instead of the "book" dihedral angle τ, connecting the HCO and H'CO planes we use the angle $Q_4=\varphi$ between the C=O bond and the HCH' plane as an out-of-plane coordinate.

We use the projections ψ_{OCH} and $\psi_{OCH'}$ of θ_{OCH} and $\theta_{OCH'}$ on the HCH' plane instead of the two angles θ_{OCH} and $\theta_{OCH'}$ [23]. We define symmetrized combinations of the ψ_{OCH} and $\psi_{OCH'}$ angles of symmetry species A_1 and B_2:

$$Q_3(A_1)\sqrt{2} = \theta = 2\pi - \psi_{OCH} - \psi_{OCH'},\ Q_6(B_2)/\sqrt{2} = \beta = \frac{1}{2}(\psi_{OCH} - \psi_{OCH'})$$

Our angular coordinates φ, θ and β are related to the MLT angular coordinates $S_3= \frac{1}{\sqrt{2}}(\theta_{OCH}+\theta_{OCH'})$, $S_4=\tau$, and $S_6=\frac{1}{\sqrt{2}}(\theta_{OCH}-\theta_{OCH'})$, through the equations:

$$\sin S_4 = \frac{\sin\varphi \sin\theta}{\sqrt{\left[1-\cos^2\varphi\cos^2(\theta/2-\beta)\right]\left[1-\cos^2\varphi\cos^2(\theta/2+\beta)\right]}}$$

$$\cos\theta_{OCH} = \cos\left(\frac{S_3+S_6}{\sqrt{2}}\right) = -\cos\varphi\cos\left(\frac{\theta}{2}-\beta\right) \qquad (1)$$

(1)

$$\cos\theta_{OCH'} = \cos\left(\frac{S_3-S_6}{\sqrt{2}}\right) = -\cos\varphi\cos\left(\frac{\theta}{2}+\beta\right)$$

Our angular coordinates φ, θ, β are related to the three angular coordinates in Handy's [25] full kinetic energy operator T: θ_1 (=θ), θ_2 (=$\theta_{OCH'}$), and ϕ (=the dihedral angle between the OCH and OCH' planes), as follows:

$$\sin\theta_2 \sin\phi = \sin\varphi\ , \qquad \cos\theta_2 = -\cos\varphi\cos\left(\frac{\theta}{2}-\beta\right) \qquad (2)$$

Using Eqs. 2, we have transformed Handy's expression [25] for the full vibrational kinetic energy operator T into our coordinates φ, θ, β, in explicit completely symmetrized form as a sum of totally symmetric terms only, each a product of functions of only one symmetrized coordinate (Supplementary Information).

The potential energy is taken to be a quartic expansion in terms of our symmetrized coordinates Q_1, Q_2, Q_3, Q_4, Q_5, Q_6:

$$V = E_0 + \sum_{i\ge j} F_{ij}Q_iQ_j + \sum_{i\ge j\ge k} F_{ijk}Q_iQ_jQ_k + \sum_{i\ge j\ge k\ge l} F_{ijkl}Q_iQ_jQ_kQ_l \qquad (3)$$

2.2. Symmetrized vibrational basis set.

We define a symmetrized orthogonal, normalized unrestricted vibrational basis set, with each basis function a product of symmetrized factors – functions of one individual vibrational coordinate. For the three stretching modes Δr_{CH}, $\Delta r_{CH'}$, Δr_{CO} we employ Morse oscillator basis wavefunctions $\psi^{M}_{n}(\Delta r_{CH})$, $\psi^{M}_{n'}(\Delta r_{CH})$ and $\psi^{M}_{n2}(\Delta r_{CO})$, while for the three angular motions – coordinates $Q_3(A_1) = \Delta\theta/\sqrt{2}$ [$\Delta\theta = \theta - \theta_0$(HCH')], $Q_4(B_1) = \varphi$, $Q_6(B_2) = \beta\sqrt{2}$ (defined above), we use harmonic oscillator basis sets: $\psi^{H}_{n3}(Q_3)$, $\psi^{H}_{n4}(Q_4)$, $\psi^{H}_{n6}(Q_6)$. The vibrational basis functions (as well as the vibrational coordinates themselves) employed here are defined with respect to the equilibrium molecular structure and therefore are not suitable for the description of extremely large deviations such as those associated with isomerization.

Our symmetrized CH stretch wavefunctions are linear combinations of individual Morse oscillator eigenfunctions $\psi^{M}_{n}(\Delta r_{CH})$ and $\psi^{M}_{n'}(\Delta r_{CH'})$ corresponding to the bond stretches CH and CH'. These symmetrized LM wavefunctions (dependent on two quantum numbers n and n') replace the conventional NM basis set using symmetrized coordinates Q_1 and Q_5. Thus, instead of using symmetrized coordinates we symmetrize the basis set as follows.

If excitation in both CH stretches is equal n=n', the basis wavefunction is of symmetry A_1, and it is written as: $\Psi^{A_1}_{n,n}(\Delta r_{CH}, \Delta r_{CH'}) = \psi^{M}_{n}(\Delta r_{CH})\,\psi^{M}_{n}(\Delta r_{CH'})$. For n≠n' there are two wavefunctions of symmetry A_1 and B_2:

$$\Psi^{A_1}_{n,n'}(\Delta r_{CH}, \Delta r_{CH'}) = \frac{\psi^{M}_{n}(\Delta r_{CH})\psi^{M}_{n'}(\Delta r_{CH'}) + \psi^{M}_{n'}(\Delta r_{CH})\psi^{M}_{n}(\Delta r_{CH'})}{\sqrt{2}}$$
$$\Psi^{B_2}_{n,n'}(\Delta r_{CH}, \Delta r_{CH'}) = \frac{\psi^{M}_{n}(\Delta r_{CH})\psi^{M}_{n'}(\Delta r_{CH'}) - \psi^{M}_{n'}(\Delta r_{CH})\psi^{M}_{n}(\Delta r_{CH'})}{\sqrt{2}} \qquad (4)$$

These basis eigenfunctions are orthogonal and normalized. The full 6-D vibrational basis set wavefunction is then:

$$\Psi^{S}_{k} = \left| n, n'(S_{CH}); n_2, n_3, n_4, n_6(S) \right\rangle = \Psi^{S_{CH}}_{n,n'}(\Delta r_{CH}, \Delta r_{CH'})$$
$$\psi^{M}_{n2}(\Delta r_{CO})\,\psi^{H}_{n3}(Q_3),\ \psi^{H}_{n4}(Q_4),\ \psi^{H}_{n6}(Q_6)$$

S_{CH} is the symmetry of the LM CH stretch part ($S_{CH}=A_1$ or B_2); S is the overall basis wavefunction symmetry ($S=A_1$, A_2, B_1 or B_2).

3. Vibrational calculation procedure

3.1. Hamiltonian matrix elements calculation

We should be able to calculate matrix elements $H_{ik}=\left\langle \Psi^{S}_{i} \middle| H \middle| \Psi^{S}_{k} \right\rangle$ of the Hamiltonian operator H=T+V for any pair of basis functions Ψ^{S}_{i} and Ψ^{S}_{k} belonging to the same symmetry species S of point group C_{2v}. Since H (both T and V) are the sum of terms that are products of functions of one internal coordinate only, and all basis functions Ψ^{S}_{i} and Ψ^{S}_{k} are in the same form, the matrix element $\left\langle \Psi^{S}_{i} \middle| H \middle| \Psi^{S}_{k} \right\rangle$ is obtained as the sum of products of one-dimensional integrals only.

All required one-dimensional integrals $\left\langle \Psi_{n_k}(Q_k) \middle| f(Q_k) \middle| \Psi_{n'_k}(Q_k) \right\rangle$ for all $f(Q_k)$ functions in both T and V operators were computed at the beginning of each calculation and stored. These were computed numerically using Gauss-Hermite or Gauss-Laguerre integration [27]. As a result the calculation of a matrix element H_{ik} was reduced to a series of multiplications and summations, which greatly accelerated the calculation.

3.2. Artificial intelligence search selection procedure (AI/SS)

Even for a molecule with as few as four atoms, the vibrational level density at 16000 cm^{-1} is too high for a conventional variational calculation. There exist a number of alternative approaches capable of producing converged molecular vibrational levels with high energy and high level densities. Recent progress in this field has been almost entirely based on the employment of curvilinear vibrational coordinates and DVR, sequential diagonalization and truncation methods, or energy selected basis schemes and iterative methods such as modified Lanczos or Davidson algorithms [15,16,18,19,28-34]. We have chosen to apply a different but well established strategy for selecting a reduced dimensionality vibrational basis, namely the artificial intelligence search selection (AI/SS) procedure [30,35-41], combined with conventional Lanczos iteration (tridiagonalization) [31,42,43]. Thus, our method does not use any prediagonalization as do most modern iterative methods. After Tietz and Chu [38] introduced using artificial intelligence (AI) for molecular vibrational state selection, Chang *et al*. [39] obtained accurate highly excited vibrational energies by diagonalizing small Hamiltonian matrices based on a smaller subset (active space - AS) selected from a much larger original basis states space. A number of developments mainly aimed at the exploration of intramolecular dynamical processes have since been published [30,35,36,40-42]. In general, there is an infinite-dimensional primitive space (PS) of vibrational states belonging to a single electronic state. Having defined a basis set for this space we would like to calculate all possible Hamiltonian matrix elements connecting the basis states as well as their energies. We would then diagonalize the huge Hamiltonian matrix and obtain the vibrational eigenvalues and eigenvectors. Unfortunately, this is not possible for most practical systems. Therefore, the aim is to select a reduced dimensionality active subspace (AS) that is feasible to manipulate but preserves the important features of the PS.

The basis set definition, based on the choice of vibrational coordinates, is crucial. First, a completely symmetrized vibrational basis set reduces the original PS dimensionality by three fourths. Next, the PS can always be subdivided into submanifolds A_i of basis states effectively isolated from each other, *i.e.*, each state from one submanifold is substantially coupled (either directly or indirectly) to all basis states from the same submanifold, while negligibly coupled to all states from the remaining submanifolds. The dimensionality of various submanifolds can be quite different, and some may contain only one or a small number of basis states. The possibility of dividing the PS and the submanifolds' dimensionalities strongly depends on the choice of vibrational basis set. Deriving the Hamiltonian matrix based on those basis states restricted to one particular submanifold and diagonalizing that matrix yields vibrational eigenvalues and eigenvectors for not only this submanifold but for the global molecular vibrational problem as well because of the effective submanifold isolation.

From our PS global basis set and a suitable initial vibrational basis state from one of the submanifolds, we identify all (or most) of the remaining basis states belonging to this submanifold. We apply an AI/SS procedure [30,35-37,40,42]. The original molecular vibrational PS is a ”net”, the basis states are the “nodes”, and the Hamiltonian coupling matrix elements are the “links” of the nodes. Starting at a given node – the “root” - our aim is to build a “subnet” incorporating the “nodes” (basis states) from the same submanifold, simultaneously preserving the substantial links among them while discarding those that are negligibly small. Since we are looking for a subnet rather than a tree a traditional decision tree pruning algorithm based on subsequent tiers [37,40] is precluded.

Our technique of search/selection is based on two main principles: the vibrational quantum number representation of the basis states and the definition of an evaluation function *f* ascribed to each basis state. The initial state in the search is given the value of $f=1$, all subsequently probed states will then be assigned $0<f<1$. A parameter f_0 is defined (typically 10^{-4} or smaller) and the algorithm selects only those basis states that have $f>f_0$ and rejects those with $f<f_0$. A second search parameter is a window width ΔE_0 centered around the initial state energy, usually chosen as 3000 cm^{-1} or larger.

We start our search with the basis state $\left|s_0\right\rangle$, which is the zeroth-order representation of a required molecular vibrational state, say $\left|s_0\right\rangle=\left|n,n'(S_{CH});n_2,\ldots\right\rangle$, of a symmetry species S. This state is given the value $f=1$. Next, the algorithm proceeds to generate all basis states $\left|s_m\right\rangle=\left|m,m'(S_{CH});m_2,\ldots\right\rangle$ of the same symmetry S that differ by one vibrational quantum from $\left|s_0\right\rangle$ ($\Delta=\Sigma_i|n_i-m_i|=1$). For each newly generated state the energy $E_k=\left\langle s_k\right|H\left|s_k\right\rangle$ and the coupling matrix element $H_{0,k}=\left\langle s_0\right|H\left|s_k\right\rangle$ are calculated, as well as the evaluation function $f=|H_{0,k}/(E_0-E_k)|\times R$, where $R=1$ if the state is located within ΔE_0 of the root state or $R=|\Delta E_0/(E_0-E_k)|$ if it is further away ($|\Delta E_0/(E_0-E_k)|$ is also taken as 1 if it exceeds 1). Each newly generated state that satisfies $f>f_0$ is stored. Next, the algorithm proceeds to probe the states separated from the initial state by two vibrational quanta ($\Delta=2$), processing the newly generated states in the same way. The search of more and more distant quantum number space states is carried on at successively larger Δ until a Δ_0 is reached where all the newly generated states with $\Delta=\Delta_0$ have $f<f_0$. Search beyond Δ_0 shows that no states with $f>f_0$ occurred at $\Delta>\Delta_0$. At that point the search from the first state $\left|s_0\right\rangle$ is terminated and the same procedure is started from $\left|s_1\right\rangle$.

Now the evaluation function of a newly generated $\left|s_k\right\rangle$ is the product $f_1\times|H_{1,k}/(E_1-E_k)|\times R$, where $H_{1,k}=\left\langle s_1\right|H\left|s_k\right\rangle$ and f_1 is the evaluation function of $\left|s_1\right\rangle$. In this way the search consecutively explores each of $\left|s_2\right\rangle$, $\left|s_3\right\rangle$, … which leads in general to the selection of more states. Each new state is checked whether it has already been examined. In general, the newly selected states have steadily diminishing *f*-values. This leads to termination of the search when no new states with $f>f_0$ can be selected on exploration of $\left|s_{N-1}\right\rangle$.

This AI/SS procedure selects an active space A of *N* basis states of the same symmetry S as the initial $\left|s_0\right\rangle$, consisting of those basis states $\left|s_k\right\rangle$ involved in nonnegligible coupling to $\left|s_0\right\rangle$. Simultaneously, the $N\times N$ symmetric Hamiltonian matrix *H* of this submanifold is stored.

N depends on the values of f_0 and E_0. In our CH stretch overtones calculations we used $E_0=3000$ cm^{-1}, although $E_0=2000$ cm^{-1} yielded satisfactory results (within 2 cm^{-1} of the $E_0=3000$ cm^{-1} result) for even the highest excited levels explored. It also gave strongly reduced *N* and CPU times. The more important parameter is f_0; for most calculations we have used $f_0=10^{-5}$, but in some cases even smaller f_0 values had to be used to achieve convergence.

Next, we carry out Lanczos tridiagonalization [43] of the Hamiltonian matrix. The Lanczos recursion starts with the same initial vector as the AI/SS procedure with 1 as a first component and zeros in all other places. The Hamiltonian matrix obtained is rather sparse since the too small matrix elements were discarded. Thus *H* can be presented as a 1D array containing only the nonzero matrix elements instead of the usual 2D array containing many zeros. Such a representation greatly accelerates the matrix×vector multiplication, which is the major time-consuming step in a Lanczos recursion, by reducing it to one run over the 1D array.

In the course of the recursion, the first component of each successively generated Lanczos vector after the starting one is set to zero, making this vector orthogonal to the initial one. This makes the Lanczos procedure more stable, especially with regard to the eigenvalue corresponding to the starting vector; it has other important advanages in addition. After the tridiagonal Lanczos matrix has been obtained we fully diagonalize it using *tqli*() from Numerical recipes [27]. This routine has been simplified to yield the eigenvalues and only the first eigenvector component C_1. The implicit orthogonalization of all Lanczos vectors with respect to the starting one makes these C_1 coincide with the first components of the Krylov vectors - approximations of the original Hamiltonian matrix eigenvectors. This means that we do not have to store all the Lanczos vectors needed to calculate eigenvectors as long as we are satisfied with the first components only. The eigenenergy of $|s_0\rangle$ is readily recognized by having one of the largest $|C_1|$ values in the spectral range of interest.

4. Results and Discussion

4.1. Ab initio computation of the quartic force field of S$_0$ formaldehyde in terms of our vibrational coordinates.

We performed large scale *ab initio* electronic computations to establish a quartic force field for the S$_0$ ground electronic state of formaldehyde using the GAMESS code [44-46] with a cc-pVTZ (correlation consistent polarized valence triple zeta) basis set. Electron correlation was accounted for by the CCSD(T) correction. This is a coupled cluster method with all singles and doubles substitutions augmented by quasiperturbative estimates of the connected triplet excitations. To establish the potential energy surface we generated total energy data points from displacements from the equilibrium geometry: r_0(CO)=1.2096082 A; r_0(CH)=1.1032935 A; θ_0(HCH')=116.1803202 deg. Step sizes of 0.05 Å were used for the symmetric bond and antisymmetric displacements $\Delta r_{CH}+\Delta r_{CH'}$, Δr_{CO} and $\Delta r_{CH}-\Delta r_{CH'}$. The three angular coordinates were stepped through 5 degree increments. In all, 425 points were used to establish the grid. The quadratic, cubic and quartic potential constants of Supplementary Table 1 came from fitting the energy data points to Eq. 3 by the program NLREG after converting to Simons-Parr-Finlan coordinates [26].

Using the computed force constants and equilibrium parameters and the above calculation procedure, we calculated the six fundamental frequencies of S$_0$ formaldehyde (Supplementary Table 2). The calculated frequencies are reasonably close to the experimental values, implying that the calculated quartic force field is satisfactory. Neverthless, before starting calculations on higher excited vibrational levels, we carried out a restricted adjustment on the six diagonal harmonic force constants, F11, F22, F33, F44, F55 and F66, to achieve precise coincidence of these calculated frequencies with the experimental values. Their adjusted values are also in Supplementary Table 1.

4.2. Large scale calculations on highly excited CH stretch overtone levels.

Next, using the *ab initio* quartic force field and adjusted constants we performed calculations on the S_0 formaldehyde CH stretch high overtones. A feature that distinguishes our approach from all other formaldehyde vibrational calculations known to us is the employment of a symmetrized LM to describe the CH stretch modes instead of the traditional NM treatment. Reisner *et al*. [5] concluded in 1984 that these modes they are better described as LM than NM [47] due to their strong diagonal anharmonicity and comparatively good localization. Neverthless, even in the recent work of Luckhaus *et al*. [20], the CH stretch system is modeled in terms of NM. However, despite the NM representation employed, their derived energy level structure [20] is characteristic of LM behavior that gets more pronounced at higher energies. Lee and Light [18] also found stronger LM behavior with increasing excitation [47].

Using two variable parameters our calculations converge to less than 0.1 cm^{-1} for the lower excited levels and to within about 2 cm^{-1} for the higher ones. Actually, only f_0 was varied; E_0=3000 cm^{-1} was suitable for all cases (Supplementary Table 3). While for the lowest levels active spaces of N~2000 were selected, for the highest excited levels (about 16000 cm^{-1}) N>25000 is usual.

The CPU times grow steeply mainly with energy but they are also state specific, being somewhat shorter for the two purely LM CH stretches |n,0(S=A$_1$,B$_2$)> of each overtone manifold. The Lanczos tridiagonalization and subsequent full diagonalization take an insignificant portion of the time required for the full calculation (few minutes for the largest N). The AI/SS procedure is the most time consuming part. It takes a few minutes for the fundamentals, about an hour for the states at ~10000 cm^{-1} and grows to about 60h for those near 16000 cm^{-1} (2.66 GHz processor).

Supplementary Table 4 contains our calculated energies for the CH stretches up to the 5th overtone, 6ν_{CH}. The levels display typical behavior for a LM structure [47]: within each overtone manifold n the two levels of purely local character, $|n,0(S_{CH});0,...\rangle$ and (S_{CH}=A$_1$,B$_2$), are located at the lowest energy, are well separated from the remaining levels, and get closer together with increasing n.

In general, our calculated CH stretch frequencies are closer to the experimental data [6,20], than those calculated (and scaled) by Luckhaus *et al.* [20], especially at the comparatively lower and at the higher vibrational energies. At the higher levels two sources of uncertainty must be recognized: first, the quartic force field may not be adequate due to the large deviations from equilibrium; and secondly, the experimental measurements themselves are uncertain because of weak absorption and poor signal to noise ratio [20].

To test the performance of our method on other highly excited levels in S$_0$ formaldehyde involving non CH stretch vibrational modes, we performed test calculations on Bouwens' *et al.* [6] two highest experimentally observed levels with A$_1$ symmetry. These are $|1\rangle$=$2_6 4_2$ (our notation: $|0,0(A_1);6,2,0,0(A_1)\rangle$) measured at 12415.4 cm^{-1} and $|2\rangle$=$1_1 2_3 4_4$ (our notation: $|1,0(A_1);3,4,0,0(A_1)\rangle$), at 12470.2 cm^{-1} [6]. Their vibrational mixing was found to be much more extensive than for the pure CH stretches, especially for $|2\rangle$. For the components of the 5th CH stretch overtone at ~16000 cm^{-1} the first calculated eigenvector component C_1 exceeded by far all the C_1 values for the remaining eigenvalues. This made the calculated level assignment straightforward and unique. However, this was not the case for the non-CH stretch combinations $|1\rangle$ and $|2\rangle$ which showed more highly structured C_1 coefficient distributions. While for $|1\rangle$ the C_1 intensity was shared mostly between two levels with comparable values, for $|2\rangle$ the intensity was distributed among a much larger number of levels, with C_1 not exceeding 0.06 for the most intense components. Despite the calculations' convergence, a unique assignment in terms of basis states was more difficult. To test the convergence and assignments, we performed a series of large scale calculations (progressively diminishing f_0) starting from each of the zeroth (basis) states $|1\rangle$ and $|2\rangle$. Supplementary Table 5 shows the components with the largest intensities along with their C_1 values. We conclude that $|1\rangle$ should be assigned to 12261 cm^{-1} and $|2\rangle$ to 12276 cm^{-1}. It must be noted that the calculated energies are quite far from the experimental values, probably due to the defects of the quartic potential field.

5. Conclusions

We have described a variational procedure for calculating highly excited vibrational energy levels in S$_0$ formaldehyde, based on artificial intelligence selection and Lanczos iteration, and employing the complete vibrational kinetic energy operator in analytical, symmetrized form. For the potential energy, we have computed a quartic *ab initio* force field in internal symmetrized coordinates. It has been slightly adjusted to reproduce the six fundamental frequencies of S$_0$ formaldehyde exactly. Using these we have carried out converged vibrational calculations on the CH stretch overtone system up to about 16000 cm^{-1}, which is the highest value achieved so far by variational calculations. The results were compared to previous model calculations as well as to experiment [20].

We have also demonstrated the method for non CH-stretch combination levels at ~12000 cm^{-1}, which were more strongly fragmented than the CH stretch overtones.

The most important development is the possibility of performing converged large scale variational (nonperturbative) vibrational calculations in S$_0$ formaldehyde at very high vibrational excitation energies where the vibrational level density is extremely large.

In forthcoming work we plan to publish calculations using this approach on a much larger number and variety of highly excited vibrational levels in formaldehyde, including overtone and combination bands of all remaining molecular modes (the CO stretch and the three angular modes).

Our calculations extend up to about 16000 cm^{-1}, although in principle our procedure allows them to be extended even further. To our knowledge, this is the first fully variational calculation that can probe such highly excited regions using no simplifications and employing the exact kinetic energy operator (without rotation). The model is restricted to the quartic potential field which is expected to behave regularly at the energies considered, but our approach also allows employment of more sophisticated potential energy functions. In forthcoming work we shall introduce a more convenient and well behaved compact semi-global PES functional form. It also substantially accelerates the calculations.

We also plan to develop a common set of internal coordinates to describe the vibrational motions in the three lowest electronic states: S$_0$, T$_1$ and S$_1$. As the T$_1$ and S$_1$ states are nonplanar at equilibrium, the φ angular coordinate must span a range of at least +50 to −50 degrees. The book angle definition used by MLT limits the out of plane displacement to 180.0 − <OCH, or 62.5 degrees. The present definition for φ allows wagging displacements of +90 to −90 degrees.

References

[1] D.C. Moule, A.D. Walsh, Chem. Rev. 75, 67 (1975)
[2] D.J. Clouthier, D.A. Ramsay, Annu. Rev. Phys. Chem. 34, 31 (1983)
[3] J.L. Hardwick, S.M. Till, J. Chem. Phys. 70, 2340 (1979)
[4] L.R. Brown, R.H. Hunt, A.S. Pine, J. Molec. Spectrosc. 75, 406 (1979)
[5] D.E. Reisner, R.W. Field, J.L. Kinsey, H.L. Dai, J. Chem. Phys. 80, 5968 (1984)
[6] R.J. Bouwens, J.A. Hammerschmidt, M.M. Grzeskowiak, T.A. Stegink, P.M. Yorba, W.F. Polik, J. Chem. Phys. 104, 460 (1996)
[7] H. Romanowski, J.M. Bowman, L.B. Harding, J. Chem. Phys. 82, 4156 (1985)
[8] M. Aoyagi, S.K. Gray, M.J. Davis, J. Opt. Soc. Am. B7, 1859 (1990)
[9] M. Aoyagi, S.K. Gray, J. Chem. Phys, 94, 195 (1991)
[10] M.J. Bramley, T. Carrington, J. Chem. Phys, 99, 8519 (1993)
[11] S. Carter, N. Pinnavaia, N.C. Handy, Chem. Phys. Lett. 240, 400 (1995)
[12] S. Carter, N.C. Handy, J. Molec. Spectrosc. 179, 65 (1996)
[13] N.M. Poulin, M.J. Bramley, T. Carrington Jr., H.G. Kjaergaard, R. Henry, J. Chem. Phys. 104, 7807 (1996)
[14] D.C. Burleigh, A.B. McCoy, E.L. Sibert III, J. Chem. Phys. 104, 480 (1996)
[15] D. Luckhaus, J. Chem. Phys. 113, 1329 (2000)
[16] F. Ribeiro, C. Iung, C. Leforestier, Chem. Phys. Letters 362, 199 (2002)
[17] R. Bernal, R. Lemus, J. Molec. Spectrosc. 235, 218 (2006)
[18] H.-S. Lee, J.C. Light, J. Chem. Phys. 120, 4626 (2004)
[19] M. Mladenovic, Spectrochim. Acta A58, 809 (2002)
[20] D. Luckhaus, M.J. Coffey, M.D. Fritz, F.F. Crim, J. Chem. Phys. 104, 3472 (1996)
[21] L.B. Harding, W.C. Ermler, J. Comp. Chem. 6, 13 (1985)
[22] D.A. Clabo, Jr, W.D. Allen, R.B. Remington, Y. Yamaguchi, H.F. Schaeffer, III, Chem. Phys. 123, 187 (1988)
[23] J.M.L. Martin, T.J. Lee, P.R. Taylor, J. Molec. Spectrosc. 160, 105 (1993)
[24] K. Yagi,C. Oyanagi,T. Taketsugu, K. Hirao, J. Chem. Phys. 118, 1653 (2003)
[25] N.C. Handy, Molec. Phys. 61, 207 (1987)
[26] G. Simons, R.G. Parr, J.M. Finlan, J. Chem. Phys. 59, 3229 (1973)
[27] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes in C (Cambridge University Press, Cambridge, 1988)
[28] A. Nauts, R.E.Wyatt, Phys. Rev. Lett. 51, 2238 (1983)
[29] A. Nauts, R.E.Wyatt, Phys. Rev. A 30, 872 (1984)
[30] C. Iung, C. Leforestier, R.E.Wyatt, J. Chem. Phys. 98, 6722 (1993)
[31] M. Gruebele, R. Bigwood, Int. Rev. Phys. Chem. 17, 91 (1998)
[32] J.M. Bowman, T. Carrington, H.-D. Meyer, Molec. Phys. 106, 2145 (2008)
[33] Z. Bacic, J.C. Light, Annu. Rev. Phys. Chem. 40, 469 (1989)
[34] J.C. Light, T. Carrington, Adv. Chem. Phys. 114, 265 (2000)
[35] S.M. Lederman,R.A.Marcus, J. Chem. Phys. 88, 6312 (1988)
[36] Y. Zhang, R.A.Marcus, J. Chem. Phys. 96, 6065 (1992)
[37] P.H. Winston, Artificial intelligence, 3rd edition (Addison-Wesley, Reading, Massachusets, 1992)
[38] J.V. Tietz, S.I. Chu, Chem. Phys. Letters 101, 446 (1983)
[39] J. Chang, N. Moiseyev, R.E. Wyatt, J. Chem. Phys. 84, 4997 (1986)
[40] J. Chang, R.E.Wyatt, J. Chem. Phys. 85, 1826 (1986)
[41] R.E. Wyatt, C. Iung, C. Leforestier, J. Chem. Phys. 97, 3458 (1992)
[42] R.E. Wyatt, Adv. Chem. Phys. 73, 231 (1989)
[43] J.K. Cullum, R.A. Willowghby, Lanczos Algorithms for Large Symmetric Eigenvalue Computations (Birkhauser, Boston, 1985) Vols. I, II
[44] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.J. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J.Comput.Chem. 14, 1347 (1993)

Acknowledgement

This research was supported by the National Research and Engineering Council of Canada.

[45] P. Piecuch, M. Wloch, J. Chem. Phys. 123, 224105/1 (2005)
[46] M. Wloch, J.R. Gour, P. Piecuch, J. Phys. Chem. A 111, 11359 (2007)
[47] M.S. Child, L. Halonen, Adv. Chem. Phys. 57, 1 (1984)