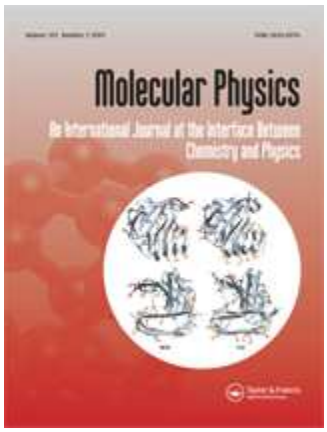
Experimental, semi-experimental and ab initio equilibrium structures

Jean Demaison



Experimental, semi-experimental and ab initio equilibrium structures

Journal:	*Molecular Physics*
Manuscript ID:	TMPH-2007-0302
Manuscript Type:	Invited Article
Date Submitted by the Author:	11-Oct-2007
Complete List of Authors:	Demaison, Jean; Université de Lille I, PhLAM
Keywords:	equilbrium structure, ab initio, spectroscopy

Experimental, semi-experimental and *ab initio* equilibrium structures

J. Demaison

Laboratoire de Physique des Lasers, Atomes, et Molécules, UMR CNRS 8523, Université de Lille I, F-59655 Villeneuve d'Ascq Cédex, France

E-mail: jean.demaison@univ-lille1.fr ; fax: +33 3 20 33 70 20

Number of pages: @
Number of tables: 21
Number of figures: 4

Abstract

The determination of equilibrium structures of molecules by spectroscopic methods or by quantum mechanical calculations is reviewed. The following structures are described in detail: experimental equilibrium structures, empirical structures, semi-experimental structures and *ab initio* structures. The approximations made by the different methods are discussed and their accuracies are compared.

1. Introduction

The concept of molecular structure was developed during the 19th century. Couper [1], around 1856, was one of the first, with Kekulé [2], to state that carbon atom is tetravalent and to form a concrete idea of molecular structure. In 1885, von Baeyer [3] published his theory of ring strain whose main conclusion is that bond angles may deviate from the tetrahedral value of 109°28'. In 1874, the concept of tetrahedral carbon was simulteanously introduced by le Bel [4]and van't Hoff [5].

The golden age of molecular structure research is the first half of the 20th century, perhaps starting with the discovery of X-ray diffraction around 1900 permitting to determine interatomic distances in crystals [6]. The structure of diamond was obtained in 1913 by the Bragg's, father and son [7]. Debye [8] was the first to determine the experimental structure of a molecule (CCl_4) in the gas phase by X-ray diffraction. At about the same time, around 1930, the first gas-phase electron diffraction experiment was performed by Mark and Wierl [9]. Although the accuracy of the measurements was not high, many molecules were studied in a relatively short time and the results were used to develop concepts concerning chemical bonding by Pauling [10] and others.

In quantum chemistry, the first successful attempt to calculate the structure of a molecule was by Burrau in 1927 on H_2^+ [11]. It was followed, in the same year, by the calculation of the bond length in H_2 by Heitler and London [12].

After the second world war microwave spectroscopy had arrived [13] which is generally considered to be the most precise technique for obtaining molecular geometries in the gas phase. This spectroscopy (as well as high resolution infrared spectroscopy) now determines rotational constants with a precision close to 1 in 10^8. If there was a simple relation between experimental rotational constants and equilibrium geometry, it would be possible to determine the structure of molecules with a tremendous accuracy. However, we will see that many factors limit this accuracy.

The apparition of the computer and its considerable increase in power are also extremely important because it permitted the development of *ab initio* electronic structure methods which can now provide accurate estimates of equilibrium structures.

There are many reviews devoted to structure determination. However, most of the time, they limit themselves either to a single experimental technique or to *ab initio* calculations. There are relatively few papers where a critical comparison of the experimental and *ab initio* techniques is made and where the interplay of these methods is emphasized. Furthermore, although the techniques are described in great detail, there is no thorough discussion of the accuracy really achievable.

The first question to answer is which accuracy is desirable. There are three reasons militing in favour of a high accuracy:

- Theoreticians need accurate structures to check their more and more sophisticated computations. Although the accuracy of *ab initio* calculations varies wildly, it is reasonable to define a range from 0.3 pm to better than 0.1 pm.
- Inspection of the range of a few X-Y bond lengths shows that it is rather small. For instance, it is only 6 pm for the CH and NH bonds in different molecules. Thus, to compare a particular structure in different molecules, an accuracy significantly better than 1 pm is required.
- The energy of a molecule is sensitive to its structure. Molecular mechanics programs for the calculation of properties of large molecules are parametrized against a small number of small molecules whose structure is assumed to be accurate. For instance, distortion of a C-C single bond by 2 pm "costs" 0.6 kJ mol^{-1}; distortion of a ∠(CCC) bond angle by 2°, about 0.4 kJ mol^{-1}; while the torsional distortion of a CCCC chain by 5° costs about 0.2 kJ mol^{-1} [14]. In order to be able to determine the relative energy of a molecule with an acceptable accuracy (a few kJ/mol), it is necessary to scale the molecular mechanics programs [15] with the aid of molecules whose molecular geometry is very accurately known.

The next step is to see under which conditions such an accuracy can be achieved. For this goal, we will first analyze the approximations which are made during the derivation of the molecular Hamiltonian.

2. The molecular Hamiltonian [16]

A molecule is a collection of nuclei and electrons, held together by certain forces and obeying the laws of quantum mechanics. The dominant forces are the Coulomb electrostatic forces. Gravitational forces are also present but a simple calculation shows that they are completely negligible compared to the electrostatic forces. For open-shell molecules or when heavy nuclei are present, the electron spin magnetic moment may interact with other magnetic moments (generated by the orbital motion of the electrons or of the nuclei, or magnetic moments of the other electrons). These interactions are much smaller and will be neglected in order to simplify the presentation. When they are not negligible, they can be taken into account without difficulty[1]. What is left is the Coulomb Hamiltonian. To write it, it is assumed that the nuclei are point masses and the relativistic effects are negligible. The first approximation is a very good one and will be further discussed in section 5.2. The second approximation is also a sound one. However, when heavy nuclei are present or during highly accurate electronic structure computations, relativistic effects become important and should be taken into account. This approximation will be further discussed in section 3.5. Using these approximations, the molecular Hamiltonian may be written as

$$\mathbf{H} = \mathbf{T}_\mathrm{N} + (\mathbf{T}_\mathrm{e} + \mathbf{V}_\mathrm{Ne} + \mathbf{V}_\mathrm{NN} + \mathbf{V}_\mathrm{ee}) = \mathbf{T}_\mathrm{N} + \mathbf{H}_\mathrm{e} \tag{1}$$

with $$\mathbf{T}_\mathrm{N} = -\hbar^2 \sum_{\alpha=1}^{N} \frac{\Delta_\alpha}{2M_\alpha} \tag{2a}$$

$$\mathbf{T}_\mathrm{e} = -\frac{\hbar^2}{m} \sum_{i=1}^{n} \frac{\Delta_i}{2} \tag{2b}$$

$$\mathbf{V}_\mathrm{Ne} = -\frac{1}{4\pi\varepsilon_0} \sum_{\alpha=1}^{N} \sum_{i=1}^{n} \frac{Z_\alpha}{r_{i\alpha}} \tag{2c}$$

$$\mathbf{V}_\mathrm{NN} = \frac{1}{4\pi\varepsilon_0} \sum_{\alpha=1}^{N} \sum_{\beta>\alpha}^{n} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \tag{2d}$$

$$\mathbf{V}_\mathrm{ee} = \frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{1}{r_{ij}} \tag{2e}$$

Δ is the Laplace operator, M_α and Z_α are the mass and atomic number of the nucleus α, m is the mass of the electron, $r_{i\alpha}$ is the distance between electron i and nucleus α, and similar definitions hold for r_{ij} and $R_{\alpha\beta}$.

[1] The interactions of the magnetic and electric moments of the nuclei with the other electric and magnetic moments in the molecule may lead to a splitting of the energy levels in several hyperfine components but they do not affect the following discussion.

2.1. The Born-Oppenheimer approximation and the electronic Hamiltonian [17]

This Hamiltonian is too complicated to be solved exactly. As proposed by Born and Oppenheimer (BO), the nuclear kinetic energy $\mathbf{T}_N$ is first separated when motion of the electrons is investigated. The justification of the BO approximation is that the heavy nuclei move much more slowly than the light electons (it also assumes that the momentum of the electrons and the nuclei is of the same order of magnitude), thus $\mathbf{T}_N$ is an almost constant term and can therefore be neglected in the differential equation. In the remaining electronic Hamiltonian $\mathbf{H}_e$, the nuclear positions **R** enter as parameters

$$\mathbf{H}_e\boldsymbol{\psi}^{(e)} = E^{BO}\boldsymbol{\psi}^{(e)} \quad (3)$$

Varying the position of **R** in small steps, one obtains E^{BO} as a function of **R**. This is the potential energy (hyper)surface (PES) : $E^{BO}(\mathbf{R})$. Its minimum corresponds to the equilibrium structure of the molecule. As the nuclear masses are absent in $\mathbf{H}_e$, E^{BO} is isotopically invariant. In other words, in this approximation all the isotopologues of an individual molecule have the same PES and the same equilibrium structure irrespective of the coordinate system used to represent **R**. It can be shown that the BO approximation can be trusted when the PESs corresponding to the different electronic states are well separated:

$$E_0^{BO} << E_1^{BO} << E_2^{BO} << \cdots \quad (4)$$

This is generally a good approximation for ground electronic states except for a few molecules. For instance, the first two excited electronic states of NO_3 are close to the ground state and all three levels interact via vibronic couplings [18].

The PES and the equilibrium structure can be calculated at different levels of electronic structure theory. This will be discussed in sections 3 and 6 together with the relativistic effects and the correction to the BO approximation.

2.2. Vibration-rotation Hamiltonian [19]

In the second step of the BO approximation the nuclear kinetic energy $\mathbf{T}_N$ is reintroduced and the Hamiltonian is written as

$$[\mathbf{T}_N + E^{BO}(\mathbf{R})]\phi(\mathbf{R}) \quad (5)$$

where $E^{BO}(\mathbf{R})$ may be considered as the potential.

The kinetic part of the vibration-rotation Hamiltonian was first derived by Wilson and Howard [20]. However, this original Hamiltonian is rather complicated (particularly due to the non-commutation of the various terms). Watson [21] could considerably simplify it. This led him to introduce a small correction term *U* which acts as a small mass-dependent effective

contribution to the potential energy. It comprises the translation of the center of mass described by three degrees of freedom (dof), the overall rotation of the molecule (three dofs) and the nuclear vibrations ($3n$ – 6 dofs). The separation of the translational motion is straightforward but the (approximate) separation of vibrational and rotational motion is not a trivial problem. It is treated in many textbooks [16, 19, 20].

After separation of rotation and vibration, a rotational Hamiltonian may be written for each vibrational state v as

$$\mathbf{H}_{\mathrm{R}}^{v} = \sum_{\xi=a,b,c} B_{v}^{\xi} \mathbf{J}_{\xi}^{2} + \mathbf{H}_{\mathrm{cd}}, \tag{6}$$

where B_v^ξ is the rotational constant in vibrational state v about the principal axis ξ, $\mathbf{J}_\xi^2$ is the component of the rotational angular momentum along the principal axis ξ and $\mathbf{H}_{\mathrm{cd}}$ is a small correction term called centrifugal distortion which is due to the fact that the molecule is not rigid and thus, upon rotation, the molecule is distorted by the stretching effects of the centrifugal forces.

In a perturbational treatment, the rotational constant B_v^ξ is given by

$$B_{v}^{\xi} = B_{\mathrm{e}}^{\xi} - \sum_{i} \alpha_{i}^{\xi} \left(v_i + \frac{d_i}{2} \right) + \sum_{i \geq j} \gamma_{ij}^{\xi} \left(v_i + \frac{d_i}{2} \right) \left(v_j + \frac{d_{ji}}{2} \right) + \cdots \tag{7}$$

The summations are over all vibrational states, each characterized by a quantum number v_i and a degeneracy d_i. The parameters α_i^ξ and γ_{ij}^ξ are called vibration-rotation interaction constants of different order. B_e^ξ is the equilibrium rotational constant. It is proportional to the inverse of the equilibrium moment of inertia which is itself a function of the coordinates at equilibrium,

$$B_{\mathrm{e}}^{\xi} = \frac{h^2}{8\pi^2 I_{\mathrm{e}}^{\xi}(\mathbf{r}_{\mathrm{e}})} \tag{8}$$

The convergence of the series expansion is usually fast, α_i^ξ being about two orders of magnitude smaller than B_e^ξ and γ_{ij}^ξ two orders of magnitude smaller than α_i^ξ, see Table 1.

The original rotational Hamiltonian cannot be used to analyze rovibrational spectra because it contains many terms which are not all independent. Watson [22] submitted the Hamiltonian to unitary transformations which allowed him to reduce the number of parameters. The rotational constants are only marginally affected by the transformation. The rotational constants of an asymmetric top obtained from a fit using the Hamiltonian of Watson are affected by a small centrifugal distortion contribution which depends on the choice of the

reduction and of the representation. Watson has shown that only the following linear combinations can be determined from the analysis of the spectra:

$$B_z = B_z^{(A)} + 2\Delta_J = B_z^{(S)} + 2D_J + 6d_2 \quad (9a)$$

$$B_x = B_x^{(A)} + 2\Delta_J + \Delta_{JK} - 2\delta_J - 2\delta_K = B_x^{(S)} + 2D_J + D_{JK} + 2d_1 + 4d_2 \quad (9b)$$

$$B_y = B_y^{(A)} + 2\Delta_J + \Delta_{JK} + 2\delta_J + 2\delta_K = B_y^{(S)} + 2D_J + D_{JK} - 2d_1 + 4d_2 . \quad (9c)$$

In eqs.(9) $B_\xi^{(A)}$ are the experimental constants in the so-called *A* reduction, $B_\xi^{(S)}$ are the experimental constants in the so-called *S* reduction, and B_ξ are the determinable constants (where $\xi = x, y, z$). However, these latter constants are still contaminated by the centrifugal distortion. As shown by Kivelson and Wilson [23], the rigid rotor constants B_ξ' are given by

$$B_x' = B_x + \frac{1}{2}\left(\tau_{yyzz} + \tau_{xyxy} + \tau_{xzxz}\right) + \frac{1}{4}\tau_{yzyz} \quad (10)$$

where B_y' and B_z' are obtained by cyclic permutation of *x, y,* and *z*.

Finally, Watson has shown that the rotational constants are affected by the mass-dependent contribution *U* to the potential energy. This contribution may be written as

$$B_x^{rigid} = B_x' - \frac{1}{8}\left[\tau_{xxxx} + \tau_{xxyy} + \tau_{xxzz}\right] \quad (11)$$

The problem is that the τ constants are experimentally determinable only for a planar molecule by means of the planarity relations of Dowling [24]. For a non-planar molecule they can be calculated from the harmonic force field [25].

These equations are approximate for three reasons: i) the result is sensitive to the particular choice of the rotational constants (equilibrium, ground state experimental, ground state corrected, etc.) used in the calculations ii) the higher-order terms (sextic terms) are neglected iii) there are still additional terms as shown by Chung and Parker [26] but whose expressions are not known to date. However, as noted by Aliev and Watson [27] these last corrections are indistinguishable from the effects of the breakdown of the BO approximations.

Compared to the other corrections, the centrifugal distortion correction is generally quite small except for very light molecules, see Table 2. Furthermore, it is different from zero only for asymmetric top molecules. However, in this case, it generally remains much larger than the experimental accuracy. In many practical cases, the spectroscopist will usually fit the line frequencies measured in order to obtain what he takes to be the experimental ground state rotational constants plus all quartic (for sufficiently high *J* also the sextic) centrifugal distortion constants. He then uses these rotational constants to determine the molecular

structure. It should be better to use the determinable constants, or still better to use the ground state rigid rotor corrected constants of Eq. (11). The spectroscopist seldom goes farther than to the determinable constants.

2.3. Magnetic correction [28]

Since the electrons tend to follow the motion of the nuclei, the bulk of the electronic contribution to the rotational constants can be taken into account by employing atomic rather than nuclear masses. This is a very good approximation for most molecules and it is about the only feasible one for polyatomic molecules. However, a small correction for unequal sharing of the electrons by the atoms and for nonspherical distribution of the electronic clouds around the atoms is sometimes nonnegligible and has to be taken into account.

The total angular momentum **J** of a molecule may be written as the sum of **N**, the angular momentum due to the rotation of the nuclei and **L**, the angular momentum of the electrons. The rotational Hamiltonian for the nuclear system plus the Hamiltonian for the unperturbed electronic energies may be written as

$$\begin{aligned}\mathbf{H} &= \frac{1}{2}\sum_\xi \frac{\mathbf{N}_\xi^2}{I_\xi} + \mathbf{H}_\mathrm{e} = \frac{1}{2}\sum_\xi \frac{\left(\mathbf{J}_\xi - \mathbf{L}_\xi\right)^2}{I_\xi} + \mathbf{H}_\mathrm{e} \\ &= \underbrace{\frac{1}{2}\sum_\xi \frac{\mathbf{J}_\xi^2}{I_\xi} + \mathbf{H}_\mathrm{e}}_{\mathbf{H}^0 = \mathbf{H}_\mathrm{R} + \mathbf{H}_\mathrm{e}} \underbrace{- \sum_\xi \frac{\mathbf{J}_\xi \mathbf{L}_\xi}{I_\xi}}_{\mathbf{H}'} + \frac{1}{2}\sum_\xi \frac{\mathbf{L}_\xi^2}{I_\xi}\end{aligned} \tag{12}$$

Since $\mathbf{L}_\xi$ is very small, the third term can be neglected and **H'** can be treated as a perturbation of $\mathbf{H}^0$. We now assume that the molecule is not in a pure $^1\Sigma$ state $\psi_0^{(0)}$ ($L = 0$) but in a perturbed state $\psi_0^{(1)}$ which has some electronic momentum. The correct effective rotational Hamiltonian is then

$$\mathbf{H}_\mathrm{eff} = \left\langle \psi_0^{(1)} \middle| \mathbf{H}_R + \mathbf{H}' \middle| \psi_0^{(1)} \right\rangle \tag{13}$$

A simple perturbation calculation up to second-order gives

$$\mathbf{H}_\mathrm{eff} = \frac{1}{2}\sum_\xi \mathbf{J}_\xi^2 \left(\frac{1}{I_\xi} - \frac{2}{I_\xi^2} \sum_{n\neq 0} \frac{\left|\left\langle n \middle| \mathbf{L}_\xi \middle| 0 \right\rangle\right|^2}{E_n - E_0} \right) \tag{14}$$

This is equivalent to the definition of an effective moment of inertia $(I_\xi)_\mathrm{eff}$ by

$$\frac{1}{(I_\xi)_\mathrm{eff}} = \frac{1}{I_\xi} - \frac{2}{I_\xi^2} \sum_{n\neq 0} \frac{\left|\left\langle n \middle| \mathbf{L}_\xi \middle| 0 \right\rangle\right|^2}{E_n - E_0} \tag{15}$$

where I_ξ on the right is calculated using the nuclear masses. This effective moment of inertia can be expressed as a function of the molecular rotational g factor in the principal axis system, whose definition is

$$g_{xx} = \frac{M_p}{I_x}\sum_i Z_i(y_i^2 + z_i^2) - \frac{2M_p}{mI_x}\sum_{n\neq 0}\frac{\left|\langle n|\mathbf{L}_x|0\rangle\right|^2}{E_n - E_0} \tag{16}$$

and g_{yy} and g_{zz} are obtained by cyclic permutation. In this equation M_p is the mass of the proton.

The effective rotational constant B_{eff} (obtained from the analysis of the rotational spectrum) is therefore

$$(B^\xi)_{\text{eff}} = B^\xi + \frac{m}{M_p} g_{\xi\xi} B^n, \tag{17}$$

where B^ξ is the rotational constant calculated with atomic masses and B^n the rotational constant calculated with nuclear masses.

The g factor can be obtained experimentally from the analysis of the Zeeman effect on the rotational spectrum [28, 29]. It can also be calculated *ab initio* [30]. A few typical results are given in Table 3. As expected, the correction is the largest for very light molecules (as LiH) and it rapidly decreases when the mass of the molecule increases. There are, however, a few exceptions. As the expression of g shows, see Eq. (16), g may become large when an electronic excited state is close to the ground state, (because the denominator $E_n - E_0$ is small). This is the case for ozone (O_3), where the magnetic correction is extremely large.

3. *Ab initio* electronic structure methods [31, 32]

3.1.The Hartree-Fock method

The electronic Hamiltonian is still too complicated to be solved exactly (except in a few special cases). Neglecting $\mathbf{V}_{NN}$ (which is a constant term), it may be written in atomic units as

$$\mathbf{H}_e = \sum_{i=1}^{n} h_i + \sum_{i=1}^{n}\sum_{j>i}^{n} \frac{1}{r_{ij}} \tag{18}$$

The first part is a sum of monoelectronic terms which are easy to solve. The simplest approximation is to assume that each electron moves in the field created by the other electrons. The second term will be approximated by a monoelectronic operator $u(r_i)$ which will take into acount the mean repulsion effect of all the other electrons on electron i

$$\begin{aligned}\mathbf{H}_e &= \sum_{i=1}^{n} \underbrace{\left[h(r_i) + u(r_i)\right]}_{F(r_i)} + \underbrace{\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{1}{r_{ij}} - \sum_{i=1}^{n} u(r_i)}_{V} \\ &= \sum_{i=1}^{n} F(r_i) + V = \mathbf{H}_0 + V\end{aligned} \tag{19}$$

If V is small (it is expected to be much smaller than V_{ee}), it is a good approximation to replace $\mathbf{H}_e$ by $\mathbf{H}_0$ which is monoelectronic and, thus, easily solvable. This is called the Hartree-Fock (HF) method. The total HF energy is obtained by adding the V_{NN} value to the eigenfunctions of $\mathbf{H}_0$. A monoelectronic wavefunction (called orbital) is attributed to each electron.

Generally a basis set expansion technique is used. The many-electron wavefunction is written as a product (more exactly a Slater determinant) of orthonormal one-electron functions called molecular orbitals (MOs)

$$\Psi = \det(\varphi_1\varphi_2\cdots\varphi_n) \tag{20}$$

Each of these MOs is expressed as a linear combination of basis functions

$$\varphi_i = \sum_k c_{ik}\chi_k \tag{21}$$

The unknown c_{ik} coefficients are determined by the variational method. For purely practical reasons, the χ_k functions are usually taken as Gaussians. A single basis function is composed of one or more primitive Gaussian functions. For example, an *s*-type basis function is

$$\chi(1s) = \sum_{i=1}^{N} d_i e^{-\alpha_i f^2 r^2} \tag{22}$$

N is the number of primitive functions, called degree of contraction, d_i the contraction coefficients, α the exponent and f a scale factor.

For a closed-shell molecule with n electrons, the $n/2$ orbitals of smallest energy (which is negative) are called occupied orbitals and the orbitals of higher energy are called virtual orbitals.

Typical errors are 1% in bond distances (which are typically underestimated). Although this result is remarkable, it is not accurate enough. The main weakness of the HF method is that it is a mean field approach which does not take into account the instanteneous interactions between the electrons: they are allowed to be at the same point in space.

3.2. Post-Hartree-Fock methods [33, 34]

The difference between the exact non-relativistic energy and the HF energy is called the correlation. If ψ_0 is the HF wavefunction, a better approximation of ψ is

$$\psi = a_0\psi_0 + a_1\psi_1 + a_2\psi_2 + \cdots \tag{23}$$

where ψ_i ($i > 0$) correspond to configurations involving virtual orbitals (excited determinants). If all the virtual orbitals corresponding to a given basis are included, the wavefunction is called the full configuration interaction (CI) wavefunction. In this FCI wavefunction, there are three categories of correlation corrections

1. excitations whose individual contributions are small, but their total contribution is large because of their great number. This is called *dynamic correlation*. It enables electrons to stay apart and it is usually the largest part of the correlation energy.
2. excitations required to provide a correct zeroth-order description as dictated by spin-symmetry considerations. It occurs for open-shell molecules and it is called *static correlation*.
3. excitations whose coefficients a_i are large. It is called *nondynamic correlation.*

The CI calculations are computer intensive, they are thus limited to small systems. Furthermore, they generally require a lot of experience.

Another method to estimate the correlation energy is provided by many-body perturbation theory with V being the perturbation term. The second-order theory gives, among others, the popular Møller-Plesset 2 (MP2) method [35], which recovers about 90% of the correlation energy. It is possible to use higher-order perturbation theory (MP3, MP4, …). The problem is that the perturbation series converges rather poorly if it converges at all.

To improve the accuracy, it is better to use the so-called coupled cluster (CC) method which takes into account the "instantaneous" interactions between the electrons [36, 37]. When one electron collides with another electron, it may be excited from an occupied orbital

φ_i to a virtual orbital φ_α. The most frequent "collision" is between two electrons which corresponds to the transition from two occupied orbitals i, j to two virtual orbitals α, β.

If Ψ_0 is the HF wavefunction and if $\mathbf{T}_2$ is the operator which induces this biexcitation,

$$\mathbf{T}_2\Psi_0 = \frac{1}{4}\sum_{ij\alpha\beta} t_{ij}^{\alpha\beta}\Psi_{ij}^{\alpha\beta} . \tag{24}$$

A better description is obtained by also taking into account single excitations, triple excitations, quadruple excitations, etc.,

$$\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \mathbf{T}_4 + \cdots, \tag{25}$$

and write

$$\begin{aligned}\Psi_{CC} &= e^{\mathbf{T}}\Psi_0 \\ &= \Psi_0 + \sum_{i<j}\sum_{\alpha<\beta} t_{ij}^{\alpha\beta}\Psi_{ij}^{\alpha\beta} + \cdots\end{aligned} \tag{26}$$

This is called the coupled cluster method. The CCSD method stops at double excitations (more exactly, all operators that involve more than two electrons are omitted) and reduces the error by a factor of three to four. The CCSDT method stops at triple excitations and further reduces the error by a factor of three to four.

If n is the number of basis functions and m is the order of the clusters (m = 2 for CCSD, m = 3 for CCSDT, m = 4 for CCSDTQ, the computation time is proportional to n^{2m+2}. To reduce this cost, the triple excitations may be taken into account perturbationally. A variant gives the CCSD(T) method whose cost is proportional to n^7 (instead of n^8 for CCSDT).

The CC methods estimate accurately the dynamic correlation but, when the nondynamic correlation is large (coefficients > 0.2), the accuracy is reduced and it may be necessary to use multireference methods. There are several diagnostics which predict the size of the nondynamic correlation [38], the most popular one being the T_1 diagnostics of Lee and Taylor [39]. In conclusion, when the nondynamic correlation is small, we have at our disposal a hierarchy of approximations of increasing accuracy :

HF < MP2 < CCSD < CCSD(T) < CCSDTQ < …

Helgaker et al. [40] compared the performances of the different methods for several bond lengths. The results are given in Table 4. The CCSD(T) method gives the best results, the next best method being the MP2 one. The question is now to check whether a larger basis set than the cc-pVQZ one or a method more sophisticated than CCSD(T) would improve the accuracy. Halkier et al. [41] studied the performance of the CCSDT method for structure computations. They found that it was generally less accurate than the cheaper CCSD(T) one. The small error

due to the perturbation calculation is nearly compensated by the error due to the neglect of quadruple excitations. Thus, the CCSD(T) method is cheaper and more accurate than the CCSDT method. The improvement brought by the CCSDTQ method (at complete basis set) was analyzed by several authors [42, 43]. The results are given in Table 5. It is seen that the CCSD(T) method performs quite well, the largest error, 0.34 pm being found for F_2 within the test set. This is true when the nondynamical correlation is small. When it is large, the situation is completely different as can be seen on the example of HONO (and other NOx molecules) where the O-N bond is 0.8 pm too short at the CCSD(T) level, see Table 6.

3.3. Choice of the basis sets

There is a great number of atom-centered Gaussian basis sets but it is recommended to use a hierarchy of basis sets which provide a systematic approach to the complete basis set (CBS) limit. One of the most popular ones are the correlation-consistent polarized valence basis sets of Dunning [44]: cc-pV*n*Z where n = D, T, Q, 5, … is the cardinal number and usually represents the highest spherical harmonic of the polarization functions. The cc-pVnZ basis sets are designed for correlation of valence electrons only. To correlate all electrons, the correlation-consistent core-valence cc-pCV*n*Z, basis sets [45] (or the correlation-consistent weighted core-valence basis sets, cc-pwCV*n*Z [46] which significantly improve the convergence with n) have to be used (Gaussians with large exponents are added to the cc-pV*n*Z basis sets). On the other hand, for an accurate description of the outer valence region, diffuse functions (Gaussians with small exponents) are added to the cc-pV*n*Z basis sets to give the aug(mented)-cc-pV*n*Z basis sets [47]. Thus the most general basis sets are aug-cc-pCV*n*Z.

It is important to have a look at the size of the basis sets because it determines, together with the method, the computation time. For the cc-pV*n*Z basis sets, the number of contracted functions $N_V(n)$ increases as the third power of the cardinal number n, see Table 7, which also shows that the core-valence sets are considerably larger than the valence sets and that the number of diffuse functions increase quadratically with the cardinal number.

3.4. Strategy to calculate a structure

Ideally, one should use the CCSD(T) method (or better) with the aug-cc-pwCV*n*Z basis sets ($n \geq 3$), all electrons being correlated. This is probably the best method for small molecules but, due to the large size of these basis sets, it is extremely computer intensive. Different procedures are used. One of the most common is to use the cc-pVnZ basis sets,

correlating only the valence electrons (the so-called frozen core approximation) and to increase the value of n, the cardinal number, as much as possible. The rate of convergence depends on the kind of bond. For instance, for the CH bond, convergence is almost achieved at $n = 3$. On the other hand, for the O···H bond (hydrogen bond), it is necessary to go at least up to $n = 6$, see Table 8.

Then, the effect of the core and core-valence correlation is calculated with the cc-pwCVQZ basis set [48 ,49]. If there is no heavy atom present ($Z < 20$), the MP2 method can be safely used to calculate this correction [50]. On the other hand, the use of the cc-pwCVQZ basis set somewhat limits the accuracy because the convergence is not yet fully achieved, see Table 9 which also shows that the correction calculated at the MP2 level is slightly too large (in absolute value). The error is quite small for first-row atoms but it becomes sizeable when a second-row atom is involved. For instance, for the C-Cl bond in ClCN, the MP2 method gives a correction of –0.43 pm whereas the CCSD(T) method gives –0.36 pm [51]. (As the cc-pwCVQZ basis set is quite large, it may be advantageous to use a completely decontracted cc-pVTZ basis set supplemented by an appropriate (1p3d2f) primitive set. This basis set called Martin-Taylor basis set (denoted as MT) is significantly smaller without any loss in accuracy [48].

Finally, if necessary, the contribution of the diffuse functions is added. It can also be calculated at a lower level of theory (e.g. MP2) [52]. The effect of the diffuse functions may be large if the basis set is small ($n \leq 3$) but it decreases rapidly when n increases and for $n \geq 5$, it is negligible in most cases, even for the C-F bond length in CH_3F, which is particularly sensitive to the effect of diffuse functions, see Table 10 (the van der Waals and hydrogen bonds are, however, important exceptions, see section 3.7). In conclusion, when the effect of the diffuse functions is calculated with a smaller basis set, the correction appears to be too large.

We have seen that there are two main sources of error: the basis set convergence error and the electronic structure method error. There are also two other approximations which may limit the accuracy: the relativistic effects and the BO approximation itself.

3.5. Relativistic effects [53]

The best method to take into account relativistic effects is to use a fully relativistic Dirac Hamiltonian. However, it is very demanding in computational resources and various methods have been developed to estimate relativistic effects.

When heavy atoms are present, the most widely used method is the pseudopotential approximation [54, 55], because it avoids the basis functions necessary for the description of the electronic core and for the inner nodal structure of the valence orbitals.

Since the effects of relativity are small for the great majority of usual atoms (they scale up to Z^4), perturbation theory may be successfully used. A number of approximate methods have been developed. One of the most widely used is the Douglas-Kroll-Hess (DKH) method [56, 57, 58] which recovers most of the scalar relativistic effects. A few typical values are given in Table 11. For chlorine and lighter atoms, this correction is smaller than 0.07 pm (value for SiH_4) but it becomes important for heavy atoms.

3.6. Correction to the Born-Oppenheimer approximation [59, 60]

A first-order Born-Oppenheimer correction which is diagonal in the electronic state is straightforwardly calculated

$$\Delta E^{\mathrm{DBOC}} = \left\langle \psi_0^{(e)} \middle| \mathbf{T_N} \middle| \psi_0^{(e)} \right\rangle \tag{27}$$

It is called adiabatic correction (or diagonal BO correction) and it is a good approximation if the ground electronic state is much lower than the electronic excited states. It is usually small (it is roughly proportional to Z^2). It is important to note that it is mass dependent. Table 12 gives a few values which show that this correction is normally negligible.

3.7. Accuracy of the coupled cluster methods

The two main sources of errors are the basis set incompletness error and the electronic structure method error. In Table 5, it is seen that the CCSD(T) method calculates a bond length in F_2 which is 0.34 pm too long (compared to CCSDTQ). Table 8 shows that the basis set convergence can be quite slow for some bonds. However, the errors are not additive. The bond length normally decreases with the size of the basis set whereas it increases with the level of theory. It is thus possible to make use of the concept of balanced calculation for which there is a near cancellation of the errors. For instance, in F_2, the experimental value of the bond length is 141.19 pm and the CCSD(T)/cc-pCVQZ value is 141.13 pm whereas the CCSD(T)/cc-pCV6Z value is much worse at 140.87 pm. This behaviour is quite general. However, a balanced calculation is difficult to use because neither the errors nor the true value are known in advance. A detailed study must be carried out in each case.

As shown in Table 4, the accuracy of the CCSD(T) method is 0.3 pm (or better) for a molecule without heavy atom ($Z < 18$) and with a small nondynamical correlation. Obviously,

it is possible to improve the accuracy. Inclusion of quadruple excitations (CCSDTQ level) changes the computed values for bond distances up to a few tenths of a picometer [42, 43, 61] and significantly improves the agreement of the theoretical results with the corresponding values. However, the computational cost is very high and the CCSDTQ method is still limited to small molecules, even with a small basis set.

Another difficulty, is the slow convergence of the basis set. Owing to the presence of the Coulomb potential, the molecular electronic Hamiltonian becomes singular when two electrons coincide in space and give rise to a characteristic cusp in the wave function [32] that the traditional basis sets are unable to describe correctly. Two solutions have been proposed

- basis set extrapolation (BSE) schemes which make use of well established extrapolation formulas for energies when using the correlation-consistent basis sets of Dunning [62].
- R12 methods which explicitely include the interatomic distance in the wave function ansatz which permit to correctly take into account the correlation cusp [63, 64].

However, there are contributions that are of similar order of magnitude as the BSE effects such as relativistic contributions and quintuple excitations.

3.8. Lower-level methods

For large molecules, the CCSD(T) method is still too expensive. It is therefore interesting to check the accuracy of lower-level *ab initio* calculations which are more readily accessible. The MP2 method has been shown to perform rather well, see Table 4. It also appeared that the remaining errors are generally mainly systematic and correction factors, or "offsets" can be derived empirically in order to predict molecular structures with an accuracy which is competitive with the best experimental methods (i.e. a few tenths of pm). The first obstacle is that the offset values are basis set dependent. Moreover, for a given basis set, the offset is not always constant, but may vary as the true equilibrium distance varies. In addition, the offset is sometimes a function of substituent effects. For instance, for the MP2/cc-pVTZ value of r(CC) in benzene: $\Delta r = r_e - r$(calc.) = –0.13 pm whereas the offset at the same level of theory is Δr = –0.65 pm for NC-CN. This large difference cannot be explained by the variation of the r_e value because they are quite close. Constancy of the offset value in a given type of bond implies that similar errors occur in the calculation of that value, i.e. the finite basis set creates the same error, the partial neglect of electron correlation has the same effect on the calculated bond length, etc. Consequently, it is not surprising that the magnitude of the offset value is at least somewhat responsive to environmental perturbations from the

surroundings of the bond. The conclusion of this discussion is that the offset method is useful but has to be used with caution.

A few typical offset values are given in Table 13. Offsets have also been determined for the CO bond (where the offset is negative and decreases in absolute value when the bond length increases) and the CN bond [65].

4. Least-squares method

4.1. Introduction

Most of the empirical and experimental methods which will be described use the least-squares method to determine the structural parameters. This method has been reviewed with emphasis on structure determination [66, 67]. For details, the reader is referred to these reviews. Therefore, we will limit ourselves to describe some shortcomings of this method and how to possibly circumvent them.

In the following we use the notation of these reviews which agrees with that of the tutorial by Albritton et al. [68]. If we want to determine p parameters β_j (j = 1, …, p) from n experimental data y_i (i = 1, …, $n > p$), the starting equations of the linear least-squares method (with unequally-weighted and/or correlated errors) are, after the model equations have been linearized in the neighborhood of the expected solution,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\Theta}_\mathbf{y} = \sigma^2\mathbf{M} \qquad (28)$$

where $\mathbf{X}$ is the Jacobian matrix, $\boldsymbol{\beta}$ is the vector of the parameters to be determined by the fit, and $\boldsymbol{\varepsilon}$ the vector of residuals (which should not be confused with the vibrational correction ε used elsewhere in this paper). $\boldsymbol{\Theta}_\mathbf{y}$ is the covariance matrix of the observations $\mathbf{y}$, σ the standard deviation of the fit. How $\boldsymbol{\Theta}_\mathbf{y}$ is subdivided into σ and the positive definite matrix $\mathbf{M}$ depends on the type of fit. For a general least-squares fit (weighted fit, with unequal and/or correlated errors of the observations), $\mathbf{M}$ is usually taken to be the covariance matrix of the experimentally measured observations $\boldsymbol{\Theta}_\mathbf{y} = \sigma_w^2\mathbf{M} = \sigma_w^2\boldsymbol{\Theta}_{\mathbf{y}\mathrm{exp}}$, then σ_w is dimensionless and should be near unity for a perfect model and a realistic assessment of the observational errors which, however, is very rarely the case in the spectroscopic applications discussed here (see remark on p. 39 of ref. [67]). For the equally (unity-) weighted and uncorrelated problem, $\mathbf{M}$ is chosen as the identity matrix $\mathbf{I}$, $\boldsymbol{\Theta}_\mathbf{y} = \sigma_u^2\mathbf{M} = \sigma_u^2\mathbf{I}$, then σ_u has the dimension of the

observations and is, in the present applications, often larger by one to two orders of magnitude than the experimental errors of the observations.

For a solution, Eq. (28) must be reduced to that of the equally-weighted and uncorrelated problem, where **M** is the identity matrix. For this purpose we define a square matrix **P** transforming the errors **ε** such that

$$\gamma = \mathbf{P}^{\mathrm{T}}\varepsilon \quad \text{and} \quad \left\langle \gamma\gamma^{\mathrm{T}} \right\rangle = \mathbf{P}^{\mathrm{T}}\left\langle \varepsilon\varepsilon^{\mathrm{T}} \right\rangle \mathbf{P} = \mathbf{P}^{\mathrm{T}}\Theta_{\mathbf{y}}\mathbf{P} = \sigma^2\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P} = \sigma^2\mathbf{I} \tag{29}$$

where **I** is the identity matrix. If **Λ** is the diagonal matrix of the eighenvalues of **M**, and **U** the corresponding matrix of the eigenvectors, it is easy to show that choosing

$$\mathbf{P} = \mathbf{U}\Lambda^{-1/2} \tag{30}$$

yields

$$\mathbf{M} = (\mathbf{P}^{\mathrm{T}})^{-1}\mathbf{P}^{-1} = \mathbf{P}\mathbf{P}^{\mathrm{T}} \tag{31}$$

left-mutiplying eq. (28) by $\mathbf{P}^{\mathrm{T}}$, we obtain the original problem with unequal variances (data of different precision) and/or correlated errors in the formulation of an equally-weighted uncorrelated problem

$$\underbrace{\mathbf{P}^T\mathrm{y}}_{\mathbf{y}'} = \underbrace{\mathbf{P}^T\mathbf{X}}_{\mathbf{X}'}\beta + \underbrace{\mathbf{P}^T\varepsilon}_{\gamma} \quad \text{i.e.} \quad \mathbf{y}' = \mathbf{X}'\boldsymbol{\beta} + \boldsymbol{\gamma}, \quad \Theta_{\mathbf{y}'} = \sigma^2\mathbf{I} \tag{32}$$

Thus, In the general case of unequal variances (data of different precision) or correlated errors, the problem is transformed to the unity-weighted uncorrelated problem.

with the solution (^ designates the least-squares estimator)

$$\hat{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{y} \tag{33}$$

and

$$\hat{\Theta}_{\hat{\beta}} = \hat{\sigma}^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{X}\right)^{-1} \tag{34}$$

with

$$\hat{\sigma}^2 = \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)^{\mathrm{T}}\mathbf{M}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)\Big/\left(n - p\right) \tag{35}$$

where $\mathbf{M}^{-1}$ is known as the generalized weight matrix. Thus, the following discussion is valid provided the transformed Jacobian, observations, and residuals are used.

4.2. Assumptions of the least-squares method

The application of the least-squares method requires certain assumptions (Gauss-Markov conditions). The distribution of the errors should be random and have zero mean and a finite second moment (the distribution need not be a "normal distribution"). If the observations to be fitted, the moments of inertia, were the highly accurate equilibrium

moments of inertia (without systematic errors), it would probably be true. But, since generally only approximations of the equilibrium moments of inertia can be used, contaminated by systematic errors, this assumption does not hold. The errors are not at all random, on the contrary, they are correlated (in fact they are roughly proportional to the inertial moments). There have been a few rare attempts to take this correlation into account [69], but it is not easy because neither the theoretical form of the correlation matrix is known nor are there enough experimental data to use estimation procedures.

In theory, the correlation should not bias the estimates, but this is true only in the limit of an infinite number of data. In the particular case of a structure determination, the number of experimental data is often not much greater than the number of parameters to be determined. Furthermore, as the mean of the errors is not zero, a bias is expected. Due to the nature of the system of equations, this bias can be large (see section 4.3).

4.3. Choice of the weighting

Generally the general least-squares method (weighted lsq, unequal and/or correlated errors) is used. Usually the weight of an observation is taken as the inverse square of the experimental uncertainty. However, as mentioned before, in the present applications the inherent model limitations due to the lack of sufficiently well known vibrational corrections generally cause errors which are much larger than the experimental errors of the inertial moments, although the latter may themselves differ greatly in magnitude, usually being larger for the less abundant species. This is disclosed, e. g., by a (possibly preliminary) weighted fit of the inertial moments using their experimental uncertainties as errors. Even when the moments come from well conducted spectral measurements, the standard deviation $\hat{\sigma}_w$ of this weighted fit may, nonetheless, come out one to two orders of magnitudes larger than unity, the value expected for a perfect model and correctly assessed errors. Multiplying, in a new weighted fit, all experimental errors with this standard deviation $\hat{\sigma}_w$ would then indeed result in a standard deviation of unity. Epple et al. [70] have proposed a method to assign weights which is hopefully more appropriate in this situation. They assign to each inertial moment $I^{\xi}_{\mathrm{exp}}(i)$ an additional "model-induced" error, Eq. (36), where $\Delta I^{\xi}_{\mathrm{rovib}}(i)$ is the rovibrational contribution to the inertial moment (for $\xi = a, b, c$, and isotopologue i) calculated (by any approximate theory) from the results of an earlier fit,

$$\mathrm{error}\left(I^{\xi}_{\mathrm{exp}}(i)\right) = \hat{\sigma}_w \cdot \sqrt{\frac{3\left(\Delta I^{\xi}_{\mathrm{rovib}}(i)\right)^2}{\sum_{\xi'=a,b,c}\left(\Delta I^{\xi'}_{\mathrm{rovib}}(i)\right)^2}} \quad \text{for} \quad \xi = a,b,c, \ \text{all} \ \ i \tag{36}$$

The true experimental errors of the inertial moments could then increase up to the values given by Eq. (36) representing the "apparent" (model-induced) errors, before the standard deviation of the fit of approximately unity would start to increase. In a crude form, this additional error even takes account of the different magnitude of the inertial moments for ξ = *a, b, c*. In practice, the square of Eq. (36) is added to the experimental variance of the respective moment on the diagonal of the covariance matrix of the observations. The procedure was used for the determination of the structure of 2-chloropropane [71], cyclopropylgermane [70], and several dicyanides [72], where the usual weighting scheme (by reciprocal squares of the experimental errors only) would have resulted in a molecular structure with very unbalanced errors due to very differently weighted moments. In all cases the additional "model-induced" variance was much larger than the experimental variance. Since the procedure does not change the covariances, the additional terms on the diagonal lead also to a substantial reduction of the correlations between the observations. From a theoretical standpoint, the method remains questionable because systematic (model) errors have been simulated by random errors.

The weights may also be calculated by the iteratively reweighted least-squares method [73] which, in contrast to Eq. (36), makes no assumptions regarding the dependence of the weights on the magnitude of the inertial moments and may hence be better suited when the model errors go back to the inadequate approximation of the vibrational correction ε. Originally, the reweighted least-squares method was devised to detect outliers due to faulty measurements etc. It is also appropriate to find suitable weights to compensate for the deficiencies of the model which, in principle, cannot be made perfect.

If possible, the weight finding by the iteratively reweighted least-squares method should be applied separately to each of the three component sets ξ = *a, b, c* of the inertial moments I_i^{ξ}, in particular, when the errors of the three sets are very different. An initial set of residuals

$$e_i^{\xi} = I_i^{\xi}(\mathrm{exp.}) - I_i^{\xi}(\mathrm{calc.}) \tag{37}$$

is obtained by the application of ordinary least-squares. Then weights are assigned to the experimental inertial moments which are functions of these residuals. Different weight functions are in use [73]. Some of them assign the weight zero to inertial moments whose

residual, Eq. (37), is larger than a certain threshold value, which means that the respective inertial moment is dropped from the fit as a suspected "outlier". For other weight functions, the weights decrease more gradually with increasing residuals. Since for the problems treated here the number of input data is usually not much larger than the number of parameters to be determined, one can hardly afford the loss of data (see Eq.(35)); therefore, the latter type of weight functions appears to be preferable. With the newly assigned weights a new (weighted) least-squares fit is performed which will result in new residuals and new weights. The treatment is iterated until consistence of the residuals is attained. The reweighted least-squares procedure was already used for the structure determination of a few molecules but there is not yet any systematic study of its performance. A number of problems remain, in particular the presence of autocorrelation of the input data (non-zero covariances between the errors of inertial moments) requires further study.

In conclusion, most assumptions of the least-squares method are not fullfilled by the problem as originally given. The difficulties can be circumvented, at least partly, by a judicious manipulation of the variances of the observations.

4.4. Collinearity or ill conditioning

It often happens that some parameters cannot be estimated with precision and that they are very sensitive to small perturbations in the data. This is due to "collinearity", or, more exactly, to the near-collinearity of the fit vector subspace spanned by p column vectors of the Jacobian matrix $\mathbf{X}$ in error space (the n-dimensional space of the vectors of the observations and of their residuals, $\mathbf{y}$ and $\boldsymbol{\varepsilon}$, respectively). This collinearity increases the variances of the estimated parameters and is responsible for important round-off errors. This problem was pointed out early by Kuchitsu et al. [74]. There are diagnostics that determine whether a collinearity exists and that can identify the near-dependent parameters affected. The correlation matrix is often employed for that purpose. But the absence of high correlations does not imply the absence of collinearity. Therefore many different procedures have been proposed.

Belsley [75] has critically reviewed these procedures and has concluded that "none is fully successful in diagnosing the presence of collinearity and variable involvement or in assessing collinearity's potential harm". To palliate the weaknesses of the existing diagnostics, he has introduced the condition indexes and has shown that they can be easily used to determine the strength and number of near-dependencies. First, the columns of the Jacobian matrix $\mathbf{X}$ are scaled to have unit length (each term of the vector column $\mathbf{X}_i$ is divided by the

norm $\|\mathbf{X}_i\|$). Then the singular values (square roots of the eigenvalues of $\mathbf{X}^T\mathbf{X}$) of the scaled $\mathbf{X}$ matrix are calculated: $\mu_1, \mu_2, \ldots \mu_p$. The singular-value decomposition of $\mathbf{X}$ may be written [76]

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{38}$$

where $\mathbf{U}$ is a $n \times p$ matrix with orthonormal columns, $\mathbf{D}$ is a $p \times p$ diagonal matrix whose elements are the singular values μ_i of $\mathbf{X}$, and $\mathbf{V}$ is a $p \times p$ orthonormal matrix. The "scaled condition indexes" of the scaled matrix $\mathbf{X}$ are defined by:

$$\eta_k = \frac{\mu_{\max}}{\mu_k} \quad k = 1, \cdots p \tag{39}$$

The highest condition index is the condition number $\kappa(\mathbf{X})$. It is an error magnification factor and it is used to determine whether a matrix is ill-conditioned or not: if the data are known to d significant figures and if the condition number of $\mathbf{X}$ is 10^r, then a small change in the data in its least significant digit can affect the solution in the $(d\text{-}r)$th place. If we consider a perturbation $\delta\mathbf{y}$ in $\mathbf{y}$, then

$$\left\| \frac{\delta\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}} \right\| \le \kappa(\mathbf{X}) \left\| \frac{\delta\mathbf{y}}{\mathbf{y}} \right\| \tag{40}$$

where $\hat{\boldsymbol{\beta}}$ is the least-squares solution and $\delta\hat{\boldsymbol{\beta}}$ the perturbation in $\hat{\boldsymbol{\beta}}$ due to $\delta\mathbf{y}$. Thus a large κ means a higher bound for the relative perturbation of the solution and may be responsible for a large systematic error.

The number of near-dependencies is equal to the number of high scaled condition indexes. To determine which parameters are involved in the collinearities, Belsley defines the variance-decomposition proportions. The variance-covariance matrix $\hat{\boldsymbol{\Theta}}(\hat{\boldsymbol{\beta}})$ of the least-squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is (where $\hat{\sigma}$ is the estimated standard deviation of the fit)

$$\begin{aligned} \hat{\boldsymbol{\Theta}}(\hat{\boldsymbol{\beta}}) &= \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1} = \hat{\sigma}^2\mathbf{V}\mathbf{D}^{-2}\mathbf{V}^T \\ \mathrm{var}(\hat{\beta}_k) &= \hat{\sigma}^2 \sum_j \left(\frac{v_{kj}}{\mu_j} \right)^2 \end{aligned} \tag{41}$$

Eq. (41) decomposes the k^{th} diagonal element $\mathrm{var}(\hat{\beta}_k)$ of $\hat{\Theta}(\hat{\beta})$ into a sum of components, each of which depends on a single μ_j. For a small μ_j the value of the component and hence its contribution to $\mathrm{var}(\hat{\beta}_k)$ will be large. Unusually large variance proportions of two (or more) parameter estimates, say $\hat{\beta}_k$ and $\hat{\beta}_{k'}$, concentrated in components associated with the same

small μ_j provide evidence that these parameters are involved in a near-dependency due to that small singular value μ_j . This can be more readily detected when variance-decomposition proportions have been defined as the elements $0 < \pi_{jk} \leq 1$ of a matrix $\boldsymbol{\pi}$,

The variance-decomposition proportions are

$$\pi_{jk} = \left(\frac{v_{kj}}{\mu_j}\right)^2 \Bigg/ \sum_{j'=1}^{p}\left(\frac{v_{kj'}}{\mu_{j'}}\right)^2 \qquad k, j = 1 \cdots p \tag{42}$$

Large elements in the same row j (of a small μ_j) for two (or more) columns, say k and k', point to a near dependency between parameters $\hat{\beta}_k$ and $\hat{\beta}_{k'}$. Near-dependency will degrade the parameter estimates involved. Belsley proposes the following rule of thumb: estimates are degraded when two or more variances have at least half of their magnitude associated with a scaled condition index of 30 or more. Evidently the probability of encountering the problem of ill-conditioning rapidly increases with the number of parameters to be determined. However, it is already present for small molecules. It is easy to explain why: quite often, an isotopic substitution induces a very small change in the moments of inertia, change which is sometimes even smaller than the change in ε.

4.5. Corrective action

Once the problem of ill-conditioning has been diagnosed, it is important to reduce its influence. The first idea is to include the moments of inertia of additional isotopologues. But the number of isotopic substitutions in a molecule is limited and, furthermore, these new data will possess near dependencies similar to those of the original data. Thus it is preferable to use additional information derived in some other way (electron diffraction, *ab initio* calculations, etc.). The different (and most useful) possibilities are described in section 5.

Unfortunately, it is not always possible to obtain new data which will substantially reduce the correlations. In such a case the introduction of appropriate "prior" information is the best solution. In the case of structure determination, the mixed estimation [75, 77] is often the most useful. Auxiliary information is added directly to the vector of observations and the Jacobian matrix. It may be a set of reasonable values of parameters based on similar molecules or from *ab initio* calculations (section 3) or from empirical correlations (section 9). For instance, if it is possible to find linear relations between the parameters which may be written in the form

$$\mathbf{c} = \mathbf{R}\beta + \eta \tag{43}$$

c is a vector of known values, **R** is a matrix of known constants and **η** is a random vector whose variance-covariance matrix is $\mathbf{V}(\boldsymbol{\eta}) = \boldsymbol{\Sigma}$. The estimation of Eq.(28) subject to the condition Eq.(43) proceeds by augmenting **y**, **X**, and **M** by **c**, **R**, and **Σ** respectively. The solution is then

$$\hat{\beta} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{X} + \mathbf{R}^{\mathrm{T}}\Sigma^{-1}\mathbf{R}\right]^{-1}\left[\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{y} + \mathbf{R}^{\mathrm{T}}\Sigma^{-1}\mathbf{c}\right] \tag{44}$$

When data of different origin are used, it is particularly important to check that the estimated weights are appropriate and that the data are compatible. Also, particular caution is indicated when one parameter appears to be determined by only one datum. This can be checked by an outlier analysis.

4.6. Outlier analysis

An outlier analysis which checks the compatibility of the data is particularly important. Several good books describe the different outlier diagnostics, they are summarized in Refs. [67, 75].

5. Experimental equilibrium structure [67, 78, 79]

5.1. General method

To determine an equilibrium structure, it is necessary to know the equilibrium rotational constants which are obtained from the ground state rotational constants and the rotational constants of all fundamentally excited vibrational states, see Eq.(7). In principle, microwave as well as infrared spectroscopies may be used to get the rotational constants B_v^{ξ} of the excited states. However, the intensity of a rotational transition is proportional to the population of its lower state which is given by the Boltzman law $\exp(-E''/kT)$. If the energy of the excited vibrational state is high, above 1000 - 2000 cm^{-1}, the population of its rotational levels will be small and the rotational transitions between them will be too weak to be observed. In this rather common case, only infrared spectroscopy may be used to determine the α constants.

It is thus obvious that the combination of microwave and infrared spectroscopies is well suited to obtain the equilibrium rotational constants. In theory, this is a simple task. However, unless the molecule has the formula XY_n and sufficiently high symmetry, the equilibrium rotational constants of only one isotopologue do not provide enough information

because the number of independent structural parameters is larger than the number of independent rotational constants (between one and three for one isotopologue). In this frequent case, the equilibrium rotational constants have to be determined for several isotopologues. Furthermore, it is often found that, in order to obtain an accurate structure (i.e. to avoid high correlation between the parameters), it is necessary to have the rotational constants of all singly substituted isotopologues. For a molecule with more than three atoms, this task requires a tremendous amount of data whose analysis is extremely time consuming. Moreover, this is often complicated by the fact that at least some excited states are not really isolated but are in resonance either by Coriolis interaction or anharmonic (Fermi, Darling-Dennison, etc.) resonance [80]. Unfortunately, this is frequently the case for a polyatomic molecule. It is then necessary to analyze the interactions between the excited states which is not easy even for small molecules. When there are only two interacting vibrational states, it is still not too complicated. If E_- is the energy of the lower interacting level and E_+ is that of the upper level, the rotational constants with the interaction taken into account may be written as

$$B_{\pm}^{\mathrm{eff}} = B_{\pm} \pm \frac{\left|\left\langle + \left| \mathbf{H}_{\pm} \right| - \right\rangle^2\right|}{E_+ - E_-}. \qquad (45)$$

Therefore, $B_+^{\mathrm{eff}} + B_-^{\mathrm{eff}} = B_+ + B_-$ and the interaction does not need to be treated accurately. For instance, there is a Fermi resonance between the $v_3 = 1$ and $v_2 = 2^0$ levels of many linear XYZ molecules. To cancel the effect of this resonance, the equilibrium rotational constants are usually calculated by $B_{\mathrm{e}} = \frac{1}{2}\left[5B_{000} - B_{100} - B_{001} - B_{02^2o}\right]$, which makes use of Eq. (45) [81].

Unfortunately, this simplification is no more valid when more than two states interact. For instance, in $^{13}CH_3F$, when the states $v_3 = 1$ and $v_6 = 1$ on one side and $v_2 = 1$ and $v_5 = 1$ on the other side are analyzed as two independent dyads, the equilibrium rotational constant is $B_{\mathrm{e}} = 25280.2$ MHz. This assumption may seem justified because these two dyads are rather far apart: $\nu_2 - \nu_6 = 276.7$ cm^{-1}. However, if the four states are analyzed as a single interacting tetrad, $B_{\mathrm{e}} = 25099.8$ MHz [82]. This discrepancy, as first sight surprising, seems to be rather general and was noted a long time ago by Saito [83] in the analysis of the excited states of SO_2.

For these reasons, the number of accurate equilibrium structures for polyatomic molecules is rather small and it is, up to now, mainly limited to molecules with three independent structural parameters (or less) i.e. to molecules with five atoms or less.

5.2. Diatomic molecules [84, 85]

The case of the diatomic molecules is worth a separate treatment. As they are much simpler, a more sophisticated theory may be used and the bond length is determinable with much higher accuracy, which permits to see the breakdown of the BO approximation, among other facts.

The vibrational potential energy $V(r)$ is usually approximated by a power series using the expansion parameter

$$\xi = \frac{r - r_e}{r_e} \tag{46}$$

$$V(r) = h a_0 \xi^2 \left(1 + \sum_{i=1}^{\infty} a_i \xi^i\right) \tag{47}$$

where r_e is the equilibrium value of the internuclear distance, a_i are Dunham potential coefficients and h is Planck's constant. Similarly to the potential function, the rovibrational energy of the molecule in a vibrational state v and a rotational state J may be written following Dunham [86] as

$$\frac{1}{h} E(v, J) = \sum_{l,k} Y_{lk} \left(v + \frac{1}{2}\right)^l J^k (J+1)^k \tag{48}$$

The first Dunham coefficient Y_{01} is approximately equal to the equilibrium rotational constant B_e, the exact relation being

$$Y_{01} = B_e + \Delta Y_{01}^{(D)} = \frac{\hbar}{4\pi\mu_e r_e^2} + \Delta Y_{01}^{(D)} \tag{49}$$

where $\mu_e = m_A m_B / (m_A + m_B)$ is the reduced nuclear mass of the molecule AB, m_A and m_B being the nuclear masses of atoms A and B, respectively, and $\Delta Y_{01}^{(D)}$ is a small correction called Dunham correction

$$\Delta Y_{01}^{(D)} = \frac{B_e^3}{2\omega_e^2} \left(15 + 14a_1 - 9a_2 + 15a_3 - 23a_1 a_2 + \frac{21}{2}\left[a_1^2 + a_1^3\right]\right) \tag{50}$$

where $\omega_e = 2\sqrt{a_0 B_e}$ is the harmonic vibrational frequency.

Within the Born-Oppenheimer approximation all the isotopologues of a molecule have the same molecular potential which results in a single bond distance. Actually, the bond distance is found to be slightly dependent of the isotopologic masses, see Table 14. The correction to the Born-Oppenheimer approximation yields a slight modification to the molecular potential which becomes dependent upon the nuclear masses. Y_{01} is now expressed as

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{ad})^2} + \Delta Y_{01}^{(D)}, \tag{51}$$

where r_e^{ad} depends on the nuclear masses. Watson [84] has shown that

$$r_e^{ad} = r_e^{BO}\left[1 + m_e\left(\frac{d_A^{ad}}{m_A} + \frac{d_B^{ad}}{m_B}\right)\right] \tag{52}$$

where r_e^{BO} is the BO interatomic distance, m_e is the mass of the electron, and d_A^{ad} and d_B^{ad} are two isotopic independent parameters.

A second correction, called non-adiabatic, has to be taken into account. It is due to the fact that the electronic ground state is slightly perturbed by the excited states (see section 2.3). This correction can be calculated with the molecular *g*-factor, and using now the atomic masses $M_X = m_X + Z_X m_e$ (X = A, B), Y_{01} may is equal to

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{ad})^2}\left[1 + \frac{\Delta Y_{01}^{(D)}}{B_e} + \frac{m_e}{M_p}g\right] \tag{53}$$

where M_p is the proton mass and $\mu = M_A M_B / (M_A + M_B)$ the reduced atomic mass. Finally, using Eq. (52), Y_{01} may be written in a more compact form as

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{BO})^2}\left[1 + m_e\left(\frac{\Delta_{01}^A}{M_A} + \frac{\Delta_{01}^B}{M_B}\right)\right] \tag{54}$$

where Δ_{01}^A and Δ_{01}^B are two isotopic independent parameters which are the sum of three terms: non-adiabatic, adiabatic, and higher order Dunham. Tiemann et al. [87] have studied in detail the order of magnitude of the different terms. It was found that the higher-order Dunham correction is almost negligible and that the non-adiabatic correction is the most important one. The adiabatic correction does not change very much for one specific atom by varying the chemical counterpart, and in general it is less than 30% of the total correction for the molecules studied.

In addition to the mass variation by isotopic substitution also the nuclear size will vary slightly giving rise to small changes in the Coulomb interaction between the electrons and the nuclei [88]. This isotope effect, called field shift in the theory of atomic spectra, slightly modifies Eq. (54) where the mean square nuclear charge radius $<r^2>_{A,B}$ is used as an expansion parameter and the new molecular parameter $V_{01}^{A,B}$ is introduced

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{BO})^2}\left[1 + m_e\left(\frac{\Delta_{01}^A}{M_A} + \frac{\Delta_{01}^B}{M_B}\right) + V_{01}^A\left\langle r^2\right\rangle_A + V_{01}^B\left\langle r^2\right\rangle_B\right]. \tag{55}$$

The parameter $V_{01}^{A,B}$ depends mainly on the electron density and its derivatives with respect to the internuclear distance at nucleus A or B. This finite nuclear size correction is only

significant when the accuracy of the measurements is extremely high and when the mass number of the atom is not too small (> 40).

5.3. Accuracy

The structure of diatomic molecules can be determined with a tremendous accuracy which is often better than 10^{-4} pm. When the structure of a polyatomic molecule can be obtained using the moments of inertia of only one isotopologue, the accuracy remains high. This is the case, for instance, for triatomic XY_2 molecules and tetraatomic C_{3v} molecules. SO_2 belongs to the first category, its equilibrium structure is $r_e(SO) = 143.078(2)$ pm and $\angle(OSO) = 119.330(3)°$ [89]. These results are in perfect agreement with a previous (and independent) determination [90] and with high level *ab initio* computations [91] (note, however, that the experimental values are likely to be more accurate). Another interesting example is linear CO_2 whose equilibrium structure could be determined in different ways from the equlibrium rotational constants of fourteen isotopologues giving: $r_e(CO) = 115.995884$ pm with a range of $75 \cdot 10^{-6}$ pm [92]. Two typical examples of the second category are phosphine and stibine where the equilibrium structure could be independently determined using the H_3 and D_3 isotopologues. Furthermore, for both molecules, the anharmonic force field was calculated in order to check the accuracy of the α constants. The results are given in Table 15 together with the corresponding *ab initio* structures. From inspection of data in Table 15, three conclusion can be drawn: i) the standard deviation of the parameters is much smaller than their range, thus indicating the presence of systematic errors, ii) their accuracy is nevertheless high, about 0.01 pm, and iii) the *ab initio* structure is likely to be less accurate.

When the moments of inertia of more than one isotopologue are needed, the accuracy usually drops to a few tenths of pm because the system of equations is no more well conditioned. There is, however, one important exception: linear triatomic molecules XYZ. The vibrational interactions of several XYZ molecules have been analyzed in great detail permitting the determination of very accurate equilibrium rotational constants for several isotopologues. Thus, the accuracy of the resulting equilibrium structures is high. A typical example is the structure of nitrous oxide, N_2O, [93] : $r_e(NN) + r_e(NO) = 231.2384(4)$ pm and $r_e(NN) - r_e(NO) = 5.779(7)$ pm. Another classic example is the structure of OCS [94] : $r_e(OC) + r_e(CS) = 271.756(3)$ pm and $r_e(OC) - r_e(CS) = 40.52(3)$ pm.

6. Semi-experimental equilibrium structure [37, 95, 96, 97, 98, 99, 100]

6.1. Method

In order to obtain equilibrium structures more easily, it was proposed quite early to calculate the vibration-rotation interaction constants from the force field [101]. This avoids the difficult problem of analyzing the resonances. Furthermore, it allows us to obtain the α constants of rare isotopologues without difficulty. Finally, it is quite useful when a vibrational state cannot be experimentally analyzed. However, it is extremely difficult to determine a cubic force field experimentally because the number of constants increases rapidly with the number of atoms and the loss of symmetry of the molecule.

Fortunately, it is now possible to calculate *ab initio* anharmonic force fields [93]. The molecular geometry is first calculated. Then, the associated harmonic force field is evaluated at the same level of theory (preferably analytically). (Note, however, that it may be advantageous to sue a better reference structure, see e.g. Ref. [102]). The cubic force constants can be most easily determined with the use of a finite difference procedure involving displacements along reduced normal coordinates [103]. This procedure is implemented in several computer packages [104]. Finally, the equilibrium rotational constants are derived from the experimental effective ground-state rotational constants and an *ab initio* cubic force field.

The high-level *ab initio* computation of a cubic force field may be time consuming. Fortunately, it has been found that the MP2 method with a basis set of triple zeta quality often gives results which are sufficiently accurate, especially for the cubic part of the force field [105]. This is partly due to the fact that the vibrational correction is only a small percentage of the rotational constants to be corrected and partly to the fact that the underlying reference structure is accurate. Thus, it needs not to be known accurately, at least in the more favorable cases. Indeed, a number of accurate equilibrium structures has been determined in this way, even for molecules as large as glycine [106] and proline [107].

However, the use of this method requires some experience. In particular, it is well advised to check that the derived force field is accurate enough, which may not always be the case. For instance, the semi-experimental structure of $H_2C{=}C{=}C$: was first determined using the CEPA-1 method giving r(C=C:) = 129.1(1) pm. It was later redetermined with the CCSD(T) method giving a bond length significantly shorter: r(C=C:) = 128.7 pm [108]. In this case, it is the choice of the treatment of electron correlation which is important but it may also happen that the choice of the basis set becomes important. For instance, in acetylene, the

harmonic bending frequency ω_4 and the corresponding anharmonicity $\omega_4 - \nu_4$ are both found to be extremely sensitive to the basis set used, in particular to the presence of a sufficient complement of diffuse functions [109]. Similar phenomena are observed more generally in bending modes for molecules that possess carbon-carbon multiple bonds [110]. Furthermore, the available experimental ground state rotational constants do not always permit to obtain a well-conditioned system of equations. In this rather frequent case, the determined parameters are not precise and may even be unreliable as will be seen in the next section.

Experimental parameters which may be used to check the quality of the force field are listed in Table 16. Examples of applications include the determination of the structure of the symmetric top $SiHF_3$ [111] or the asymmetric top FPO [112].

6.2. Comparison of the different equilibrium structures

A few structures are compared in Table 17. One conclusion is obvious: the standard deviations of the parameters (when determined) of the experimental and semi-experimental equilibrium structures are much too optimistic. This can be explained by two facts

- the number of available data is generally not much larger than the number of parameters to be determined; thus, the derived standard deviations are not reliable
- the assumptions of the least squares method are not obeyed (see section 4.2).

The experimental equilibrium structure is generally less accurate. However, for the r(SiCl) bond in SiH_3Cl, it is obviously the *ab initio* value which is not accurate. This is probably because it is difficult to calculate *ab initio* accurate interatomic distances between second-row atoms. This is confirmed by the result on the r(CCl) bond in O=CHCl. Finally, it is important to note that the semi-experimental equilibrium structure is perhaps not always the most accurate: in HOF, BHFOH, and BF_2OH the *ab initio* structures might be more accurate

7. Empirical structures [66, 67, 78, 113]

As it is extremely difficult to determine the equilibrium structure of a polyatomic molecule, particularly if there are many degrees of freedom, several empirical methods have been developed. Most of them only use the ground state moments of inertia. The simplest one gives the effective structure.

7.1. Effective structure (r_0)

The vibrational correction $\varepsilon = I_0 - I_e$ is quite small compared to I_0, less than 1% in most cases, see Table 18. Thus, it may seem to be a good approximation to neglect its contribution, the structural parameters being directly fitted to the ground state moments of inertia I_0 of a sufficiently large set of isotopologues. This method is simple and widely used. For diatomic molecules, its accuracy is about 1%. The accuracy remains about the same for polyatomic molecules, provided that the system of normal equations is well conditioned. Otherwise, the resulting r_0 structure may be widely different from the equilibrium structure, see for instance Table 19 which compares different structures of vinyl fluoride (H_2C=CHF). There are three different r_0 structures for which the angle $\angle(CCH_g)$ varies between 123.7° and 129.2°.

Furthermore, when hydrogen atoms are present, the determined bond lengths are affected by an error of 0.5 pm or more. Finally, an analysis of the residuals of the fit shows that they are highly autocorrelated instead of being random, see figure 1. It indicates that the model is not correct and that the standard deviations of the parameters are not reliable.

7.2. Substitution structure (r_s)

A better hypothesis than the neglect of ε is to assume that ε remains constant upon isotopic substitution, see Table 18. It is the implicit basis of Costain's [113] early suggestion to calculate the Cartesian coordinates of the substituted atom in the principal axis system by Kraitchman's equations [114](and their special forms for molecular symmetries [115, 116]), essentially from the difference of the inertial moments of the parent and the substituted isotopologue, the only two species involved.

The assumption ε = constant is obviously not valid in three cases:

- when an hydrogen atom is substituted by deuterium because the change of mass (and of ε) is large,
- when there are large axis rotation upon isotopic substitution,
- for most molecular complexes because the variation of ε upon substitution is large.

In other cases, the resulting r_s structure is believed to have a high degree of validity for heavy atoms. Unfortunately, even this is not always true. In the particular case of a linear molecule, the Cartesian coordinate z of the substituted atom may be written, using Kraitchman equation, as

$$z^2 = \frac{I_0' - I_0}{\mu} = \frac{\Delta I_e + \Delta\varepsilon}{\mu} \qquad (56)$$

with $\mu = \frac{M\Delta m}{M + \Delta m}$. (57)

This equation permits to estimate the uncertainty δz on z,

$$\delta z = \frac{1}{2}\frac{\Delta\varepsilon}{\mu}\frac{1}{|z|} \quad . \tag{58}$$

Costain [117] assumed that $\Delta\varepsilon/\mu$ is approximately constant and that the experimental error on the moments of inertia is negligible. Using the data for N_2O, he proposed the following empirical rule to estimate the error of a r_s coordinate (in pm)

$$\sigma(z) = \frac{12}{|z|} \tag{59}$$

Groner [78] discussed the usefulness of this relation. It cannot be taken for granted that Eq. (59) determines the true uncertainty of the coordinates. It is mainly useful to check the consistency of r_s coordinates. Actually, as $\mu \approx \Delta m$, i;e. about 1 or 2, Eq. (58) shows

- the uncertainty δz increases when z decreases; in other words, small coordinates are inaccurate, a frequent case in large molecules,
- the uncertainty is large when $\Delta\varepsilon$ is large, i.e. when hydrogen is substituted by deuterium or when there is a large axis rotation. These problems, already mentioned, are well known. But what is less known is that $\Delta\varepsilon$ roughly increases with the mass of the molecule (see figure 2), that is to say it is difficult to obtain an accurate substitution structure for a large molecule [118].

An illuminating example is the structure of C_5O: although the equilibrium Cartesian coordinate of the C_2 atom is as large as 492.2 pm, the error is $z_e - z_s = 2.56$ pm which is two orders of magnitude larger than predicted by Eq. (59).

To improve the accuracy of the substitution structure, a double substitution method (where two atoms are substituted at the same time) has been proposed [119] but it is difficult to apply. Futhermore, it was shown that the double substitution method is in general not much better than the single substitution method [120,121,122].

The r_0 method has been expanded to the r_{I_ε} method [123] where the experimental moments of inertia $I_0^\xi(i)$ of a sufficiently large set of isotopologues i are fitted to the structural parameters and to three constant, mass-independent parameters ε^ξ which hence do not depend on the isotopologue: $I_0^\xi(i) = I_{\text{rigid}}^\xi(i) + \varepsilon^\xi$, the $I_{\text{rigid}}^\xi(i)$ alone being functions of the structure. The method is a true r_0 derivative: the independent structure parameters must all either be free to be fitted or kept at fixed values, and the center- and product-of-mass conditions are fulfilled automatically. A similarity to the substitution method exists in so far as the difference

between respective moments of the parent and any isotopologue no longer contains any rovibrational contribution. The method has meanwhile been replaced by the methods with mass-dependent rovibrational contributions, as described in next section.

8. Mass-dependent structures

8.1. Method

To improve the substitution method, the variation of ε with the moment of inertia I_0 was analyzed. It was found that ε approximately varies as $\sqrt{I_0}$ [124]. This is particularly true in the case of isotopic substitution as the analysis of the residuals of the r_0 and r_s structures shows. The ground state moment of inertia I_0 may be approximately written as

$$I_0^{\xi} = I_m^{\xi} + c_{\xi}\sqrt{I_m^{\xi}} \tag{60}$$

where I_m^{ξ} is an approximation of I_e^{ξ}. The structural parameters and the proportionality constants c_{ξ} (one for each principal axis) are determined by least squares fitting. The resulting structure was called $r_m^{(1)}$ by Watson et al. [125], "m" because it is mass-dependent (ε is a homogeneous function of degree 1/2 in the masses) and "1" because one extra parameter per axis is used. It is important to note that the number of fitted parameters is the same as for the r_{I_ε} method. The $r_m^{(1)}$ method significantly increases the accuracy of the derived structure, at least in most cases. However, it was found to give poor results when there is a small Cartesian coordinate. An analysis of several linear triatomic XYZ molecules, where the coordinate of Y is small showed that ε varies as $\frac{m_X m_Z}{M}$, where m_X and m_Z are the masses of atoms X and Z, respectively, see figure 3 and ref. [122].

The method was generalized by Watson et al. [125] who proposed the following expression for I_0

$$I_0^{\xi} = I_m^{\xi} + c_{\xi}\sqrt{I_m^{\xi}} + d_{\xi}\left[\frac{m_1 m_2 \cdots m_N}{M}\right]^{1/(2N-2)} \tag{61}$$

where N is the number of atoms of the molecule and c_{ξ} and d_{ξ} are fitting parameters (one on each axis). Using this equation in the fitting procedure gives the $r_m^{(2)}$ structure. Excellent agreement between the r_e and $r_m^{(2)}$ structure was found for many small molecules, even when a small coordinate (as in N_2O) is present.

Watson et al. [125] have extended the r_m method also to cases where large principal inertial axis rotations with respect to the parent molecule occur in individual species of the set of isotopologues. Such cases are mainly limited to near oblate molecules and require a more complicated formulation.

8.2. Laurie correction [126]

In almost all structures investigated, hydrogen atoms, if present, have caused anomalies. This is usually solved by using the "Laurie" correction [127], which assumes that the effective r(X-D) and r(X-H) bond lengths are different by a small quantity δr_D. In practice, a shrinkage of the X-H bond length upon deuteration is expected. This agrees with the idea of the heavier deuterium atom lying lower in the asymmetric potential well of the X-H vibrational mode than the hydrogen atom and, hence, nearer to atom X. It is worth pointing out that the ensuing bond length change has often been shown to remain restricted to the X-H bond length considered, hardly affecting the rest of the parameters, at least when the least-squares system is not ill-conditioned.

The correction method presented by Watson et al. assumes that the apparent lengths of both bonds, X-H as well as X-D, consist of a common major part r_m, which is identical for both bonds (and close to the r_e value) plus an additional bond elongation due to the H or D atom asymmetrically vibrating against the remainder of the molecule. This elongation is taken to be proportional to the respective vibration frequency, i.e. proportional to the reciprocal square root of the reduced mass m_{red} of the vibrator. The additional part of the bond length is hence different for the X-H and the X-D bond,

$$r_H = r_m + \delta\mu_H \quad (62a)$$

$$r_D = r_m + \delta\mu_{D,} \quad (62b)$$

with r_H and r_D being the apparent bond length values resulting from the least squares fit, δ is a common proportionality factor and $\mu = m_{red}$ is different for the H and D containing species:

$$\mu_{H/D} = \sqrt{\frac{M}{m_{H/D}(M - m_{H/D})}} \quad (63)$$

where M is the total mass of the isotopologue containing either H or D, m_H is the mass of hydrogen and m_D the mass of deuterium. In principle, r_H and r_D can be determined by a least squares fit and r_m can then be deduced. However, in most cases, the system of normal equations is very ill-conditioned and the derived parameters are meaningless (an exception is when the rotational constant of the H-, D-, and T-isotopologues are available). Therefore, a constraint has to be introduced to alleviate the problem. Either $r_H - r_D$ is constrained to a value

known from experience ("Laurie" correction, near 0.3 pm) or r_m is constrained to a r_e value known from other sources (e.g. *ab initio* calculations or empirical correlation, see sections 3 and 9).

8.3. Use of the mixed regression method

Besides using the *ab initio* value for r_m(XH), it may be useful to use other *ab initio* parameters as predicate observations in a mixed regression, see section 4.5. This, indeed, improves the conditioning and allows the determination of parameters with a better accuracy. This is furthermore a good way to check the accuracy of the *ab initio* structure. However, it is not always possible to calculate a reliable *ab initio* structure. Nevertheless, there is another way to improve the conditioning. Inspection of the constants c_ξ and d_ξ (when they are accurately determined) shows that they vary smoothly and slowly for a given series of molecules, see Table 20. Hence, it should be possible to use the c_ξ and d_ξ of a similar molecule in a mixed regression. For instance, it is easy to obtain an accurate $r_m^{(2)}$ structure for ClCN and BrCN but it fails in the case of FCN and ICN, in particular because of the lack of isotopic substitution for F or I. However, using the c and d values of ClCN/BrCN for FCN/ICN as predicate observations (with an accuracy of about 10%) considerably improves the conditioning and gives structures in good agreement with the r_e structure. However, these results are still preliminary and more work is needed to appreciate the validity of this method.

8.4. Accuracy

At each step of the improvement of the model ($r_0 \rightarrow r_m^{(1)} \rightarrow r_m^{(2)}$), the number of parameters increases. Thus, the conditioning of the system is expected to deteriorate. This is important because the r_m model is still an approximate one, see figure 4 which shows that the residuals are still not random. If the rovibrational contribution is roughly estimated as 0.5% of the inertial moment and its model error as 10% of its value (as derived from the standard deviations of the parameters c_ξ and d_ξ), then the model error is 0.05% of the inertial moment which is more, by orders of magnitude, than the experimental error in modern spectroscopy. This model error, with repercutions on the accuracy of the parameters, is often "amplified" by the ill-conditioning, see section 4.4.

In conclusion, the r_m method is well suited when the number of available rotational constants is large compared to the number of structural parameters, a condition which is usually fulfilled only for small molecules. Nevertheless, the structure of several molecules

(not all of them small) was recently determined using the r_m method and these experimental structures were found to be in good agreement with the corresponding *ab initio* structures (i.e. within 0.2-0.3 pm), see Table 21.

9. Empirical correlations

There are many correlations between structural parameters and other parameters [128]. Although most of them are difficult to use, there are a few relations which may be of great help for a structure determination.

The stretching force constants are related to the bond strength and one can expect a good correlation between the length of a bond and the corresponding diagonal stretching force constant, as shown for instance by Badger [129,130]. In principle, such a correlation might be used to determine bond lengths, but it is often more difficult to accurately determine force constants than bond lengths because the number of force constants, $(1/2)n_{vib}(n_{vib} + 1)$, is usually larger than the number of fundamental vibrations, n_{vib}, and additional information (vibrational frequencies of isotopic species, …) is required to determine the force constants. There are, however, a few cases in which the bond length vs. diagonal force constant correlation is relevant. One is when the molecule is simple enough and this correlation has been for instance used to determine the Au-Au bond length in several molecules [131].

More generally, when a vibrational mode r has characteristic frequency far from the others, r', it may be considered as isolated. In other words, the non-diagonal force constants $f_{rr'}$ are then negligible compared to the corresponding energy differences $E_r - E_{r'}$. Thus, in this particular case, a relationship between the bond length and the corresponding stretching vibrational frequency is to be expected. Such a relationship was first pointed out by Bernstein [132] and considerably developed by McKean [133] for CH bonds.

McKean used selective deuteriation, all CH bonds but the relevant CH one being deuteriated. One therefore decouples the CH group of interest from the rest of the molecule in order to make sure that the CH stretching frequency is not affected. The main difficulty is often to synthetize a molecule where all hydrogens but one have been replaced by deuterium. Furthermore, in a few cases the stretch vibration is perturbed by some resonance and, except if a detailed vibration-rotation analysis was carried out to provide a deperturbed stretching frequency, another method is required.

Another method for determining isolated XH stretching frequencies relies on data from overtone bands, i.e. resulting from the multiexcitation of a single vibrational mode. It is well known that the motions of the individual bonds become increasingly localized upon increasing excitation and thus decreasingly affected by internal couplings [134]. The band origin of as many overtones of the XH stretch as possible needs to be measured and inserted into a so-called Birge-Sponer plot [135] :

$$\nu = v\tilde{\omega} - v(v+1)\tilde{\omega}x \qquad (64)$$

where $\tilde{\omega}$ is the mechanical frequency and $\tilde{\omega}x$ the first anharmonic correction term. This procedure thus also allows the frequency of the "unperturbed" fundamental band, ν_{is}(XH), to be determined :

$$\nu_{is} = \tilde{\omega} - 2\tilde{\omega}x \qquad (65)$$

The major advantage of this second method is to avoid isotopic labeling. However, it does not always avoid the problem of resonance.

The accuracy of the correlation corresponds to a bond length change of about 0.1 pm for a shift in the isolated CH stretching fundamental frequency of 10 cm^{-1} [136], *r*(C-H) bond lengths can thus be determined with a precision of about 0.2 pm, from this correlation. This method was extended by McKean to SiH and GeH bond lengths [137]. Such a linear relationship was also shown to exist between *r*(OH) and ν(OH) [138] and between *r*(NH) and ν(NH) [139]. A non-linear relationship between *r*(N=O) and ν(N=O) was also pointed out [140].

There are other correlations of a bond length with other properties (electronegativity or another bond length) [141]. They are much less accurate but they may be useful to check whether the value of a bond length is correct or not.

10. Special molecules

10.1. Nonrigid molecules

Up to the present we have assumed that the molecule is semi-rigid, i.e. that the vibrational amplitudes are small. In principle, there is no difficulty to calculate the *ab initio* structure of a nonrigid molecule and this was done for many molecules. On the other hand, when there is a large amplitude motion the Taylor series expansion of the potential is no more valid and the calculated force constants (and α constants) may be not reliable. Thus, experimental and semi-experimental equilibrium structures might be doubtful. There are, up to this date, very few results. However, in the case of internal rotation of a symmetric rotor (a methyl group for instance) and when the potential is high enough, it seems that it is still possible to calculate the force constants using the assumption of small amplitude vibrations and, hence, a semi-experimental structure. One possible explanation of this nice behaviour is that the internal rotation potential is almost harmonic (at the bottom of the well). On the other hand, when the large amplitude motion is highly anharmonic, as for the inversion motion in formamide for instance, the results are very disappointing. For instance, the α constants of formamide have been recently calculated *ab initio* at the MP2/cc-pVTZ level of theory [59]. Although the $v_{12} = 1$ mode (the large-amplitude NH_2 inversion motion at 288.7 cm^{-1}) is well isolated, the calculated $\alpha_{12}^{A} = 2039$ MHz and $\alpha_{12}^{C} = 28.9$ MHz constants are in poor agreement with the corresponding experimental values, 978 MHz and –0.0099 MHz, respectively. However, it would be dangerous to draw any conclusion because the level of theory used is not high enough and the CCSD(T) method is likely to be required. Furthermore, it was possible to determine a $r_m^{(2)}$ structure of formamide which is in very good agreement with the *ab initio* structure [65].

There is another difficulty which may be encountered with non-rigid molecules. For instance, several molecules containing the CONH linkage seem to have a pyramidalized nitrogen at equilibrium and a double-minimum inversion potential with a very small inversion barrier allowing for an effectively planar ground-state structure [65]. Thus, the comparison of structures of different meaning may be meaningless.

10.2.Weakly bound cluster molecules

Weakly bound cluster molecules are a particular case of nonrigid molecules. Their study is a large field of research because the forces that hold the clusters together are responsible for the interactions of the molecules in liquid and solids. They also explain the stability of some conformers of large molecules (intramolecular hydrogen bonds).

The structure of complexes can also be calculated *ab initio*. For hydrogen bonded and van der Waals complexes, the choice of the *ab initio* method is not a problem. On the other hand, the basis set convergence is extremely slow, see Table 8. Furthermore, it is necessary to use a basis set with diffuse functions and the effect of the basis set superposition error (BSSE) has to be corrected [142]. This BSSE is the consequence of the use of finite basis sets. When the distance between the two monomers decreases, their basis functions overlap. Each monomer "borrows" functions from the other monomer, effectively increasing its basis set and improving the calculation of derived properties (of the monomers only). The calculated interaction energy becomes too large and the potential energy hypersurface distorted. A simple solution is the use of extremely large basis sets, which is not practical in most cases. The conventional way to correct for BSSE is based on the Boys-Bernardi counterpoise (CP) scheme. The CP-corrected potential energy surface of the dimer AB is

$$E^{\mathrm{CP}}(\mathrm{AB}) = E^{\mathrm{AB}}(\mathrm{AB}) + [E^{\mathrm{A}}(\mathrm{A}) + E^{\mathrm{B}}(\mathrm{B}) - E^{\mathrm{AB}}(\mathrm{A}) - E^{\mathrm{AB}}(\mathrm{B})]$$

where the superscript denotes the basis set used. Thus, one has to calculate five different energies at each geometry.

To date, there are experimental or *ab initio* equilibrium structures available for this class of molecules, all the experimental structures being empirical. The applicability of the r_{m} method was tested on several small complexes and satisfactory results were obtained [143]. On the other hand, to determine a satisfactory $r_{\mathrm{m}}^{(1)}$ structure for the very large complex phenylacetylene-argon [144], it was necessary to introduce constraints. This is not surprising because the number of structural parameters, 26, is very large (although also the number of isotopologues is large: 24). It is also probable that the large amplitude motion of the argon atom complicates the situation.

10.3. Large molecules

When the number of structural parameters is large, most of the methods previously described fail although lower level *ab initio* calculations are still possible. In this case, the molecule may often exist in different conformations and the primary aim is to determine which conformers are present (i.e. estimate the dihedral angles and more, if possible) as well as their relative energies. Microwave spectroscopy in combination with electronic structure calculations is well suited for this purpose.

The ground state rotational constants, which are easy to determine, are the most useful parameters but it often happens that the rotational constants of different conformers are quite

similar. In such a (frequent) case, the ambiguity can be cleared up using one (or several) of the following pieces of information:

- the components of the dipole moment, which are easily estimated by microwave spectroscopy (from Stark effect or relative intensities) usually change during a rotation around a bond even if the total dipole moment remains almost constant,
- isotopic substitution of one atom not too close to the center of mass often results in conclusive differences of the rotational constants of different conformers,
- nuclear quadrupole coupling constants are also useful because different orientations of the quadrupole tensor result in different coupling constants.

A very attractive example is 1-octene for which fifteen conformational isomers (from 131 predicted) have been measured by molecular beam Fourier transform microwave spectroscopy. The identification used the rotational constants, inertial defect, and relative intensities [145]. It is interesting to note that the quality of the *ab initio* calculations (MP2/6-31G*) was not sufficient to provide the correct energy ordering of the conformers. This is a problem frequently encountered in this kind of study.

Another fascinating example is the microwave study of 2-chloroethyl ethyl sulfide ($CH_3CH_2SCH_2CH_2Cl$) [146], for which many conformers are possible. The two conformers identified were assigned by comparing the experimental rotational constants to the *ab initio* values. However, this did not permit an unambiguous identification. For the most stable conformer, the experimental rotational constants of the ^{37}Cl isotopologue were used with those of the ^{35}Cl species to calculate the Kraitchman Cartesian coordinates of the chlorine atom (see section 7.2). The comparison with the *ab initio* coordinates permitted to identify the conformer. This identification was corroborated by comparison of the experimental relative intensities with those predicted *ab initio* (which are function of the components of the dipole moment). For the second conformer, it was the comparison of the experimental and *ab initio* quadrupole coupling constants which permitted its identification.

Another typical example is the glycolaldehyde-water complex, for which seventeen conformers are possible. The most stable conformer was identified without ambiguity in the same way as for 1-octene [147]. Plusquellic et al. [148] have summarized the work recently done on peptides and dipeptides. More examples of conformational analyses are given, for example, in Ref. [149].

11. Conclusions

Experimental equilibrium structures can be highly accurate for diatomic molecules and C_{2v} or $C_{\infty v}$ triatomic molecules. The accuracy is generally much lower for larger molecules. It is mainly limited by two factors: i) ill-conditioning of the system of equations, and ii) anharmonic resonances. The harmful influence of the ill-conditioning may be reduced by increasing the diversity and accuracy of the data. A thorough analysis of the resonances is in principle possible as shown for example for linear triatomic molecules. However, it is difficult and time consuming but an ab initio calculation of the anharmonic foce field should make the analysis much easier.

During the last 20 years, tremendous progress has been made in the ab initio computation of the equilibrium structures which is now much faster and much easier than their determination by experimental methods. The accuracy of the computed structures is comparable or even better to that observed in most spectroscopic studies. For instance, the accuracy of the CCSD(T) method is 0.3 pm or better. However, this relatively small error is due to a cancellation of errors due to the truncation of the coupled-cluster hierarchy and to the truncation of the basis set. Although it is possible to improve this accuracy, at least for small molecules, this is not an easy task.

The semi-experimental method is also much easier to use than the experimental method. It is also generally more accurate. However, it also suffers from the problem of ill-conditioning which may be solved in the same way as for the experimental method. Another problem is that it is difficult to estimate the accuracy of the computed rovibrational corrections. For this reason, an experimental determination of deperturbed α constants for at least the main isotopologue is desirable. To improve the accuracy, it is possible to calculate the higher-order vibration-rotation interaction constants γ_{ij}^{ξ} from the quartic force field. However, although their contribution is not negligible, they do not affect much the resulting structure.

The mass-dependent r_{m} methods are also limited to small molecules and they are significantly less accurate than the other methods. Nevertheless, they are quite useful because they are very easy to use and their accuracy is sometimes sufficient.

Acknowledgments

I thank Prof. H.D. Rudolph for many helpful suggestions, Prof. A.G. Császár, Dr. C. Gutle and Prof. J. Liévin for critical reading of the manuscript and Profs. J.E. Boggs and M. Herman for their encouragements.

References

[1] Couper, A.S., 1858, *Annales de chimie et de physique,* Série 3, **53**, 488.

[2] Kekulé, F.A., 1865, *Bull. Soc. Chim.,* **3**, 98.

[3] von Bayer, A., 1885, *Ber. Dtsch. Chem. Ges.*, **18**, 2269.

[4] Le Bel, A.J., 1874, *Bull. Soc. Chim. France*, **22**, 337.

[5] van't Hoff, J.H., 1875, *La Chimie dans l'Espace* (Rotterdam,: P.M. Bazendijk).

[6] von Laue, M., 1912, *Münchener Ber.*, 303.

[7] Bragg, W.H., and Bragg, W.L., 1914, *Proc. Roy. Soc. London*, Ser. A **89**, 468.

[8] Debye, P., Bewilogua, L., Ehrhardt, P., 1929, *Phys. Z.*, **30**, 84.

[9] Mark, H., and Wierl, R., 1930, *Naturwissenschaften*, **18**, 205.

[10] Pauling, L., 1960, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (Ithaca, NY: Cornell University Press).

[11] Burrau, Ø., 1927, *K. Danske Viden. Selskab*, **7**, 14.

[12] Heitler, W., and London, F., 1927, *Z. Physik*, **44**, 453.

[13] Townes, C.H., and Schawlow, A.L., 1955, *Microwave Spectroscopy* (New York: McGraw-Hill).

[14] Hargittai, I., and Levy, J.B., 1999, *Struct. Chem.* **10**, 387.

[15] Burkert, U., and Allinger, N.L., 1982, *Molecular Mechanics* (Washington, D.C.: American Chemical Society).

[16] Bunker, P.R., and Jensen, P., 1998, *Molecular Symmetry and Spectroscopy* (Ottawa: NRC Research Press).

[17] Bunker, P.R., and Jensen, P., 2000, *Computational Molecular Spectroscopy*, edited by P. Jensen and P.R. Bunker (Chichester: Wiley), p. 3.

[18] Stanton, J.F., 2007, *J. Chem. Phys.*, **126**, 134309.

[19] Papousek, D., and Aliev, M.R., 1982, *Molecular Vibrational-Rotational Spectra* (Amsterdam: Elsevier).

[20] Wilson, E.B., Decius, J.C., and Cross, P.C., 1955, *Molecular Vibrations* (New York: McGraw Hill).

[21] Watson, J.K.G., 1970, *Molec. Phys.*, **19**, 465.

[22] Watson, J.K.G., 1977, *Vibrational Spectra and Structure*, Vol. 6, edited by J.R. Durig (Amsterdam: Elsevier), p. 1.

[23] Kivelson, A., and Wilson, E.B., 1952, *J. Chem. Phys.,* **20**, 1575.

[24] Dowling, J.M., 1961, *J. Molec. Spectrosc.,* **6**, 550.

[25] Hedberg, L., and Mills, I.M., 1993, *J. Molec. Spectrosc.,* **160**, 117.

[26] Chung, K.T., and Parker, P.M., 1963, *J. Chem. Phys.,* **38**, 8.

[27] Aliev, M.R., and Watson, J.K.G., 1985, *Molecular Spectroscopy: Modern Research*, Vol. 3, edited by K.N. Rao (New York: Academic Press), p. 1.

[28] Gordy, W., and Cook, R.L., 1984, *Microwave Molecular Spectra*, (New York: Interscience).

[29] Flygare, W.H., 1974, *Chem. Rev.* **74**, 655.

[30] Gauss, J., Ruud, K., Helgaker, T., 1996, *J. Chem. Phys.,* **105**, 2804.

[31] Hehre, W.J., Radom, L., Schleyer, P.v.R., and Pople, 1986, J.A., *Ab Initio Molecular Orbital Theory* (New York: Wiley).

[32] Helgaker, T., Jørgensen, P., and Olsen, J., 2000, *Molecular Electronic Structure Theory* (New York: Wiley).

[33] Bartlett, R.J., and Stanton, J.F., 1994, *Rev. Comp. Chem.,* Vol. V, edited by K.B. Lipkowitz and D.B. Boyd (New Yok: VCH), p. 65.

[34] Helgaker, T., Ruden, T.A., Jørgensen, P., Olsen, J., and Klopper, W., 2004, *J. Phys. Org. Chem.* **17**, 913.

[35] Møller, C., and Plesset, M. S., 1934, *Phys. Rev*., **46**, 618.

[36] Purvis III, G. D., and Bartlett, R. J., 1982, *J. Chem. Phys*., **76**, 1910.

[37] Lee, T. J., and Scuseria, G. E., 1995, *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, edited by S. R. Langhoff (Dordrecht: Kluwer), pp. 47.

[38] Janssen, C.L., and Nielsen, I.M.B., 1998, *Chem. Phys. Lett.*, **290**, 423.

[39] Lee, T.J., and Taylor, P.R., 1989, *Int. J. Quant. Chem. Symp.*, **23**, 199.

[40] Helgaker, T., Gauss, J., Jørgensen, P., and Olsen, J., 1997, *J. Chem. Phys.,* **106**, 6430.

[41] Halkier, A., Jørgensen, P., Gauss, J., and Helgaker, T., 1997, *Chem. Phys. Lett.,* **274**, 235.

[42] Feller, D., Peterson, K.A., Crawford, T.D., 2006, *J. Chem. Phys.* **124**, 054107.

[43] Ruden, T.A., Helgaker, T., Jørgensen, P., and J. Olsen, *J. Chem. Phys.* **121**, 5874.

[44] Dunning Jr., T. H., 1989, *J. Chem. Phys*., **90**, 1007.

[45] Woon, D.E., and Dunning Jr., T.H., 1995, *J. Chem. Phys*., **103**, 4572.

[46] Peterson, K.A., and Dunning Jr., T.H., 2002, *J. Chem. Phys*., **117**, 10548.

[47] Kendall, R. A.; Dunning Jr., T. H.; Harrison, R.J., 1992, *J. Chem. Phys*., **96**, 6796.

[48] Martin, J.M.L., and Taylor, P.R., 1994, *Chem. Phys. Lett.*, **225**, 473.

[49] Császár, A. G., Allen, W. D., 1996, *J. Chem. Phys*. **104**, 2746.

[50] Margulès, L., Demaison, J., Rudolph, H.D., 2001, *J. Molec. Struct.* **599**, 23.

[51] Demaison, J., Boggs, J.E., and Rudolph, H.D., 2004, *J. Molec. Struct.* **695**, 145.

[52] Margulès, L., Demaison, J., and Boggs, J.E., 1999, *J. Phys. Chem. A*, **103**, 7632.

[53] Pyykkö, P., 1988, *Chem. Rev.* **88**, 563.

[54] Martin, J.M.L., Sundermann, A., 2001, *J. Chem. Phys*., **114**, 3408.

[55] Peterson, K.A., 2003, *J. Chem. Phys*., **119**, 11099.

[56] Douglas, M., and Kroll, N.M., 1974, *Ann. Phys. (N.Y.)*, **82**, 89.

[57] Hess, B.A., 1986, *Phys. Rev. A,* **33**, 3742.

[58] de Jong, W.A., Harrison, R.J., and Dixon, D.A., 2001, , *J. Chem. Phys*., **114**, 48.

[59] Handy, N.C., and Lee, A.M., 1996, *Chem. Phys. Lett.* **252**, 425.

[60] Gauss, J., Tajti, A., Kállay, M., Stanton, J.F., and Szalay, P.G., 2006, , *J. Chem. Phys*., **125**, 144111.

[61] Heckert, M., Kállay, M., Gauss, J., 2005, *Mol. Phys.* **103**, 2109.

[62] Heckert, M., Kállay, M., Tew, D.P., Klopper, W., Gauss, J., 2006, *J. Chem. Phys.* **125**, 044108.

[63] Klopper, W., Manby, F.R., Ten-No, S., Valeev, E.F., 2006, *Intern. Rev. Phys. Chem.* **25**, 427.

[64] Pawlowski, F., Halkier, A., Jørgensen, P., Bak, K.L., Helgaker, T., Klopper, W., 2003, *J. Chem. Phys.* **118**, 2539.

[65] Demaison, J., Császár, A.G., Kleiner, I., Møllendal, H., 2007, *J. Phys. Chem. A*, **111**, 2574.

[66] Rudolph, H.D., 1995, *Advances in Molecular Structure Research,* Vol. 1 edited by M. Hargittai and I. Hargittai (Greenwich, CT: JAI Press), p. 63.

[67] Demaison, J., Wlodarczak, G., and Rudolph, H.D., 1197, *Advances in Molecular Structure Research,* Vol. 3 edited by M. Hargittai and I. Hargittai (Greenwich, CT: JAI Press), p. 1.

[68] Albritton, D.L., Schmeltekopf, A.L., and Zare, R.N., 1976, *Molecular Spectroscopy: Modern Research*, Vol. II, edited by K.N. Rao (New York: Academic Press), p. 1.

[69] Stiefvater, O.L., and Rudolph, H.D., 1989, *Z. Naturforsch.,* **44a**, 95.

[70] Epple, K.J., and Rudolph, H.D., 1992, *J. Molec. Spectrosc.,* **152**, 355.

[71] Mayer, M., Grabow, J.-U., Dreizler, H., and Rudolph, H.D., 1992, *J. Molec. Spectrosc.,* **151**, 217.

[72] Demaison, J., Wlodarczak, G., Rück, H., Wiedenmann, K.H., and Rudolph, H.D., 1996, *J. Molec. Struct.,* **376**, 399.

[73] Hamilton, L.C., 1992, *Regression with Graphics* (Belmont, CA: Duxbury Press).

[74] Kuchitsu, K., Fukuyama, T., and Morino, Y., 1969, *J. Molec. Struct.,* **4**, 41.

[75] Belsley, D.A., 1991, *Conditioning Diagnostics* (New York: Wiley).

[76] Brodersen, S., 1990, *J. Molec. Spectrosc.,* **142**, 122.

[77] Bartell, L.S., Romanesko, D.J., and Wong, T.C., 1975, *Chemical Society Specialist Periodical Report N° 20: Molecular Structure by Diffraction Methods,* edited by G.A. Sim and L.E. Sutton (London: The Chemical Society), p. 72.

[78] Groner, P., 2000, *Vibrational Spectra and Structure*, Vol. 24, edited by J.R. Durig (Amsterdam: Elsevier), p. 165.

[79] Graner, G., 1992, *Accurate Molecular Structures*, edited by A. Domenicano and I. Hargittai (Oxford University Press), p. 65.

[80] Califano, S., 1976, *Vibrational States* (London: Wiley).

[81] Morino, Y., and Matsumura, C., 1967, *Bull. Chem. Soc. Japan,* **40**, 1095.

[82] Demaison, J., Breidung, J., Thiel, W., and Papousek, D., 1999, *Struct. Chem.* **10**, 129.

[83] Saito, S., 1969, *J. Molec. Spectrosc.,* **30**, 1.

[84] Watson, J.K.G., 1973, *J. Molec. Spectrosc.,* **45**, 99.

[85] Watson, J.K.G., 1980, *J. Molec. Spectrosc.,* **80**, 411.

[86] Dunham, J.L., 1932, *Phys. Rev.*, **41**, 721.

[87] Tiemann, E., Arnst, H., Stieda, W.U., Törring, T., and Hoeft, J., 1982, *Chem. Phys.*, **67**, 133.

[88] Tiemann, E., Knöckel, H., and Schlembach, J., 1982, *Ber. Bunsenges. Phys. Chem.* **86**, 821.

[89] Flaud, J.-M., and Lafferty, W., 1993, *J. Molec. Spectrosc.,* **161**, 396.

[90] Morino, Y., Tanimoto, M., and Saito, S., 1988, *Acta Chem. Scand. A*, **42**, 346.

[91] Martin, J.M.L., 1998, *J. Chem. Phys.*, **108**, 2791.

[92] Graner, G., Rossetti, C., and Bailly, D., 1986, *Molec. Phys.*, **58**, 627.

[93] Teffo, J.-L., and Chédin, A., 1989, *J. Molec. Spectrosc.,* **135**, 389.

[94] Lahaye, J.-G., Vandenhaute, R., and Fayt, A., 1987, *J. Molec. Spectrosc.,* **123**, 48.

[95] Clabo, D.A., Allen, W.D., Yamaguchi, Y., Remington, R.B., and Schaefer, H.F., 1988, *Chem. Phys.*, **123**, 187.

[96] Allen, W.D., Yamaguchi, Y., Császár, A.G., Clabo, D.A., Remington, R.B., and Schaefer, H.F., 1990, *Chem. Phys.*, **145**, 427.

[97] Botschwina, P., and Flügge, J., 1991, *Chem. Phys. Lett.,* **180**, 589.

[98] Bürger, H., and Thiel, W., 1997, *Vibration-Rotational Spectroscopy and Molecular Dynamics,* edited by D. Papousek (Singapore: World Scientific), p. 56.

[99] Császár, A.G., 1998, *Encyclopedia of Computational Chemistry*, edited by P. von Ragué Schleyer (Wiley: New York).

[100] Sarka, K., and Demaison, J., 2000, *Computational Molecular Spectroscopy*, edited by P. Jensen and P.R. Bunker (Chichester: Wiley), p. 255.

[101] Hoy, A.R., Mills, I.M., and Strey, G., 1972, *Molec. Phys.*, **24**, 1265.

[102] Allen, W.D., Cs´szár, A.G., 1993, *J. Chem. Phys.* **98**, 2983.

[103] Schneider, W., and Thiel, W., 1989, *Chem. Phys. Lett.*, 157, 367.

[104] Stanton, J. F.; Lopreore, C.; Gauss, J., 1998, *J. Chem. Phys.* **108**, 7190.

[105] Puzzarini, C., and Cazzoli, G., 2006, *J. Molec. Spectrosc.*, **240**, 260.

[106] Kasalová, V., Allen, W.D., Schaefer III, H.F., Czinki, E., and Császár, A.G., 2007, *J. Comput. Chem.*, **28**, 1373.

[107] Allen, W.D., Czinki, E., and Császár, A.G., 2004, *Chem. Eur. J.*, **10**, 4512.

[108] Gauss, J., and Stanton, J.F., 1999, *J. Molec. Struct.*, **485-486**, 43.

[109] Martin, J.M.L., Lee, J.J., and Taylor, P.R., 1998, *J. Chem. Phys.*, **108**, 676.

[110] Simandiras, E.D., Rice, J.E., Lee, T.J., Amos, R.D., and Handy, N.C., 1988, *J. Chem. Phys.*, **88**, 3187.

[111] Demaison, J., Margulès, L., Breidung, J., Thiel, W., and Bürger, H., 1999, *Molec. Phys.*, **97**, 1053.

[112] Beckers, H., Bürger, H., Paplewski, P., Bogey, M., Demaison, J., Dréan, P., Walters, A., Breidung, J., and Thiel, W., 2001, *Phys. Chem. Chem. Phys.*, **3**, 4247.

[113] Costain, C.C., 1958, *J. Chem. Phys.*, **29**, 864.

[114] Kraitchman, J., 1953, *Amer. J. Phys.*, **21**, 17.

[115] Chutjian, A., 1964, *J. Mol. Spectrosc.,*14, 361

[116] Nygaard, L., 1976, *J. Mol. Spectrosc.*, **62**, 292

[117] Costain, C.C., 1966, *Trans. Amer. Cryst. Ass.*, **2**, 157.

[118] Demaison, J., and Rudolph, H.D., 2002, *J. Mol. Spectrosc.*, **215**, 78.

[119] Pierce, L., 1959, *J. Mol. Spectrosc.*, **3**, 575.

[120] Demaison, J., Wlodarczak, G., Burie, J., and Bürger, H., 1990, *J. Molec. Spectrosc.*, **140**, 322.

[121] Sarka, K, Papousek, D., Demaison, J., Mäder, H., and Harder, H., 1997, *Vibration-rotational spectroscopy and molecular dynamics*, edited by D. Papousek (World Scinetific: Singapore), p. 116.

[122] Le Guennec, M., Wlodarczak, G., Demaison, J., Bürger, H., Litz, M., and Willner, H., 1993, *J. Mol. Spectrosc.* **157**, 419.

[123] Rudolph, H. D., 1991, *Struct.Chem.,* **2**, 581.

[124] Demaison, J., and Nemes, L., 1979, *J. Molec. Struct.*, **55**, 295.

[125] Watson, J. K. G., Roytburg, A., and Ulrich, W., 1999, *J. Molec. Spectrosc*., **196**, 102.

[126] Demaison, J., Margulès, L., Boggs, J.E., and Rudolph, H.D., 2001, *Struct. Chem.*, **12**, 1.

[127] Laurie, V.W., 1958, *J. Chem. Phys.*, **28**, 704.

[128] Mastryukov, V.S., and Simonsen, S.H., 1996, *Advances in Molecular Structure Research*, Vol. 2, edited by M. Hargittai and I. Hargittai (Greenwich, CT: JAI press), p. 163.

[129] Badger, R.M., 1934, *J. Chem. Phys,.* **2,** 128.

[130] Herschbach, D.R., and Laurie, V.W., 1961, *J. Chem. Phys,.* **35**, 458.

[131] Perreault, D., Drouin, M., Michel, A., Miskowski, V.M., Schaefer, W.P., and Harvey, P.D., 1992, *Inorg. Chem,.* **31,** 695.

[132] Bernstein, H.J., 1962, *Spectrochim. Acta,* **18,** 161.

[133] McKean, D.C., 1978, *Chem. Soc. Rev,.* **7,** 399.

[134] Henry, B.R., 1987, *Acc. Chem. Res,.* **20**, 429.

[135] Birge, R.T., and Sponer, H., 1926, *Phys. Rev,.* **28**, 259.

[136] Demaison, J., and Wlodarczak, G., 1994, *Struct. Chem.*, **5**, 57.

[137] McKean, D.C., 1981, *J. Molec. Struct,.* **113,** 251.

[138] Demaison, J., Herman, M., and Liévin, J., 2007, *Intern. Rev. Phys. Chem.*, **26**, 391.

[139] Demaison, J., Margulès, L., and Boggs, J.E., 2000, *Chem. Phys.*, **260**, 65.

[140] Turner, P.H., and Cox, A.P., 1978, *J. Chem. Soc. Faraday Trans.,* II **74,** 533.

[141] Demaison, J., Margulès, L., and Boggs, J.E., 2003, *Struct. Chem.*, **14**, 159.

[142] Boys, S.F., and Bernardi, F., 1970, *Mol. Phys.* **19**, 553.

[143] Kisiel, Z., 2003, *J. Mol. Spectrosc.*, **218**, 58.

[144] Dreizler, H., Hartke, B., Rudolph, H.D., 2006, *J. Molec. Struct.*, **825***,* 1

[145] Fraser, G.T., Suenram, R.D., and Lugez, C.L., 2001, *J. Chem. Phys. A*, **105**, 9859.

[146] Tubergen, M.J., Lesarri, A., Suenram, R.D., Samuels, A.C., Jensen, J.O., Ellzy, M.W., and Lochner, J.M., 2005, *J. Molec. Spectrosc.*, **233**, 180.

[147] Aviles-Moreno, J.-R., Demaison, J., and Huet, T.R., 2006, *J. Am. Chem. Soc.*, **128**, 10467.

[148] Plusquellic, D.F., Kleiner, I., Demaison, J., Suenram, R.D., Lavrich, R.J., Lovas, F.J., Fraser, G.T., and Ilyushin, V.V., 2006, *J. Chem. Phys.*, **125**, 104312.

[149] Møllendal, H., 1993, *Structures and Conformations of Non-Rigid Molecules*, edited by J. Laane, M. Dakkouri, B. van der Veken, and H. Oberhammer (Dortrecht: Kluwer), p. 277.

[150] Demaison, J., Hübner, H., Wlodarczak, G., 1998, *Molecular Constants Mostly from Microwave, Molecular beam, and Sub-Doppler Laser Spectroscopy*, edited by W. Hüttner (Berlin: Landolt-Bönstein, Springer), Vol. 24A.

[151] Demaison, J., Margulès, L., Boggs, J.E., 2003, *Phys. Chem. Chem. Phys.* **5**, 3359.

[152] Demaison, J., Herman, M., Liévin, J., 2007, *J. Chem. Phys.* **126**, 164305.

[153] Demaison, J., Hübner, H., Hüttner, W., Vogt, J., Wlodarczak, G., 1998, *Molecular Constants Mostly from Microwave, Molecular beam, and Sub-Doppler Laser Spectroscopy*, edited by W. Hüttner (Berlin: Landolt-Bönstein, Springer), Vol. 24C.

[154] Demaison, J., Császár, A.G., and Dehayem-Kamadjeu, A., 2006, *J. Phys. Chem. A* **110**, 13609.

[155] Schafer, A., Horn, H., and Ahlrichs, R., 1992, *J. Chem. Phys.* **97**, 2571.

[156] Császár, A.G., Czakó, G., Furtenbacher, T., Tennyson, J., Szalay, V., Shirin, S.V., Zobov, N.F., and Polyansky, O.L., 2005, *J. Chem. Phys.* **122**, 214305.

[157] Dyall, K.G., Taylor, P.R., Faegri, K., and Partridge, H., 1991, *J. Chem. Phys.* **95**, 2583.

[158] Margulès, L., Demaison, J., and Boggs, J.E., 2000, *J. Mol. Struct. (Theochem)* **500**, 245.

[159] Margulès, L., Demaison, J., and Boggs, J.E., 2000, *Struct. Chem.* **11**, 145.

[160] Authier, N., Balland, N., and Le Floch, A., 1993, *J. Molec. Spectrosc.*, **160**, 590.

[161] Ogilvie, J.F., 1994, *J. Phys. B*, **27**, 47.

[162] Canè, E., Fusina, L., Bürger, H., Jerzembeck, W., Brünken, S., Lewen, F., and Winnewisser, G., 2002, *J. Molec. Spectrosc.*, **215**, 1.

[163] Canè, E., Di Lonardo, G., Fusina, L., Jerzembeck, W., Bürger, H., Breidung, J., and Thiel, W., 2006, *J. Molec. Struct.*, **780-781**, 98.

[164] Canè, E., Di Lonardo, G., Fusina, L., Jerzembeck, W., Bürger, H., Breidung, J., and Thiel, W., 2005, *Mol. Phys.*, **103**, 557.

[165] Watson, J.K.G., 2006, *J. Mol. Struct.* **795**, 263.

[166] Watson, J.K.G., 2005, *J. Mol. Struct.* **742**, 91.

[167] Breidung, J., Demaison, J., D'Eu, J.-F., Margulès, L., Collet, D., Mkadmi, E.B., Perrin, A., and Thiel, W., 2004, *J. Mol. Spectrosc.* **228**, 7.

[168] Demaison, J., Sormova, H., Bürger, H., Margulès, L., Constantin, F.L., and Ceausu-Velcescu, A., 2005, *J. Mol. Spectrosc.* **232**, 323.

[169] Demaison, J., 2006, *J. Mol. Spectrosc.* **239**, 201.

Table 1. Vibrational contribution to the rotational constants (in MHz) [a].

Molecule	B_e	$C = \sum \alpha_i d_i / 2$	C/B (%)	$D = \sum \gamma_{ij} d_i d_j / 4$	D/C (%)
HCN	44511.620	198.137	0.45	2.395	1.21
FCN	10586.782	32.604	0.31	–0.248	–0.76
ClCN	5982.8975	12.0644	0.20	–0.0207	–0.17
BrCN	4126.5059	6.2838	0.15	0.0596	0.95
ICN	3329.0568	3.5084	0.11	–0.0049	–0.14

[a] Data from Ref. [150].

Table 2. Centrifugal contribution to the rotational constants of NH_2 and HCOOH (in MHz)

	exp.	Eqs.(I) [a]	Eqs.(II) [b]	Eqs.(III) [c]	ΔB=I+II+III	$\Delta B_\xi/B_\xi$ (%)
NH_2 [d]						
A	712634.609(5)	61.906	–69.363	199.621	192.163	0.03
B	388213.098(9)	–122.779	–88.220	–19.726	–230.725	–0.06
C	244843.726(9)	2.577	177.916	17.007	197.500	0.08
HCOOH [e]						
A	77512.2354(11)	0.0200	–0.0643	0.7805	0.7362	0.000
B	12055.1065(2)	–0.1556	0.0154	–0.0057	–0.1460	0.000
C	10416.1151(2)	0.0231	0.0454	–0.0079	0.0605	0.000

[a] Eqs. (9).
[b] Eq. (10).
[c] Eq. (11).
[d] Ref. [151].
[e] Ref. [152].

Table 3. Electronic contribution to the rotational constants (B_0 and ΔB in MHz) [a]

molecule	B_0	g	ΔB
LiH	229965.07	-0.6584	82.467
CO	57635.966	-0.2691	8.448
CO_2	11698.4721	-0.05508	0.351
OCS	6081.4921	-0.028839	0.096
OCSe	4017.6537	-0.01952	0.043
CS_2	3271.4882	-0.02274	0.041
HCN	44315.9757	-0.098	2.365
$FC^{15}N$	10186.2903	-0.0504	0.280
$ClC^{15}N$	5748.0527	-0.0385	0.121
$BrC^{15}N$	3944.8441	-0.0385	0.083
$IC^{15}N$	3225.5485	-0.0325	0.057
HC≡C-CN	4549.067	-0.0213	0.053
O_3, *A*	105536.235	0.642	-36.903
O_3, *B*	13349.2547	-0.119	0.865
O_3, *C*	11834.3614	-0.061	0.393

[a] Data from Ref. [153].

Table 4. Accuracy of the ab initio methods for bond lengths [40] (in pm, basis set cc-pVQZ, all electrons correlated).

	mean error [a]	standard deviation	mean absolute error	maximum absolute error
	$\frac{1}{n}\sum_{i=1}^{n} e_i$	$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(e_i-\bar{e}\right)^2}$	$\frac{1}{n}\sum_{i=1}^{n}\lvert e_i\rvert$	$\max_i\lvert e_i\rvert$
HF	–2.91	2.25	2.91	8.45
MP2	–0.26	0.67	0.54	1.67
CCSD	–0.89	0.79	0.89	3.07
CCSD(T)	–0.19	0.30	0.22	1.20

[a] error: $e_i = r_i^{\text{calc}} - r_i^{\text{exp}}$

Table 5. Comparison of the CCSD(T) and CCSDTQ methods (in pm)

Molecule	Δr [a]	Ref.
N_2	0.07	42
CS	0.15	42
CS_2	0.14	42
HF	0.02	43
N_2	0.07	43
F_2	0.34	43
CO	0.06	43

[a] $\Delta r = r$[CCSD(T)] – r[CCSDTQ]

Table 6. Constrained optimization of the *r*(N–O) bond length (in pm) in *trans*-HONO using various levels of theory [154]

Method [a]	*r*(N–O) [b]
CCSD	140.1
CCSD(T)	142.1
CCSDT	142.4
CCSDTQ	142.9

[a] The pVDZ basis of Ahlrichs for the middle two atoms and the VDZ basis of Ahlrichs for the two terminal atoms have been used [155]. The frozen-core approximation is employed throughout.

[b] The other structural parameters were fixed at *r*(O–H) = 96.34 pm, *r*(N=O) = 116.90 pm, ∠(ONO) = 110.7642°, and ∠(HNO) = 102.0827°.

Table 7. Number of contracted functions in the cc-pVnZ, aug- cc-pVnZ, and cc-pCVnZ basis sets.

		cardinal number				
Atoms	basis	D	T	Q	5	6
H-He	N [a]	5	14	30	55	91
	+aug [b]	4	9	16	25	36
Li-F	N [a]	14	30	55	91	140
	+aug [b]	9	16	25	36	49
	+C [c]	4	13	29	54	90
Na-Cl	N [a, d]	18	34	59	95	144
	+aug [b]	9	16	25	36	49
	+C [c]	9	25	50	70	-

[a] standard cc-pVnZ basis set
[b] number to add to the previous (N) line to get the number of functions in aug-cc-pVnZ set
[c] number to add to the previous (N) line to get the number of functions in cc-pCVnZ set
[d] the use of the cc-pV(n+d)Z is recommended for secondrow atoms, it increases the number of basis functions by 5.

Table 8. Convergence of the basis sets: *r*[aug-cc-pVnZ] – *r*[aug-cc-pV(n-1)Z] (in pm, CCSD(T) method, unless otherwise stated)[a].

	HCCH		HNO		HOOH		HF	$H_2O.HF$
n	*r*(CH)	*r*(CC)	*r*(OH)	*r*(NO)	*r*(OH)	*r*(OO)	*r*(HF)	r(O···H) [b]
T	-1.470	-1.985	-1.060	-1.050	-0.240	0.060	-0.314	-3.760
Q	-0.040	-0.336	-0.190	-0.350	-0.300	-0.750	-0.327	-0.470
5	-0.043	-0.101	-0.020	-0.130	-0.110	-0.200	-0.040	-0.398
6	-0.004	-0.031	-0.010	-0.080	-0.050	-0.050	-0.005	-0.327

[a] This work.
[b] MP2 method with counterpoise correction, see text.

Table 9. Calculation of the core and core-valence correlation in acetylene (in pm) [a].

method	basis set	Δr(CC) [b]	Δr(CH) [b]
CCSD(T)	cc-pwCVDZ	0.076	0.052
	cc-pwCVTZ	0.221	0.107
	cc-pwCVQZ	0.262	0.126
	cc-pwCV5Z	0.274	0.132
	MT [c]	0.26	0.12
MP2	cc-pCVQZ	0.29	0.13

[a] This work.
[b] $\Delta r = r$(frozen core) – r(all electrons correlated)
[c] Martin-Taylor basis set, see text.

Table 10. Effect of the diffuse functions: *r*[aug-cc-pVnZ] – *r*[cc-pVn)Z] (distances in pm, angles in degrees, CCSD(T) method) [a].

	CH_3F			HCCH	
n	*r*(CF)	*r*(CH)	∠(HCF)	*r*(CC)	*r*(CH)
D	2.46	0.318	-1.156	0.134	-0.028
T	0.78	0.031	-0.463	0.051	0.026
Q	0.29	0.022	-0.174	0.037	0.019
5	0.08	0.017	-0.039	0.014	0.010
6				0.008	0.007

[a] This work.

Table 11. Relativistic effects (distances in pm, angles, ∠, in degrees)

Molecule	Δr [a]	method [b]	ref.
N_2	-0.02	DKH	42
CS	0.05	DKH	42
CS_2	0.02	DKH	42
F_2	0.03	DKH	58
Cl_2	0.05	DKH	58
Br_2	–0.32	DKH	58
ClF	0.06	DKH	58
BrF	0.13	DKH	58
BrCl	–0.04	DKH	58
CF_4	0.00	DKH	58
SiF_4	–0.05	DKH	58
Br_2CO, r(CO)	–0.05	DKH	58
Br_2CO, r(CBr)	–0.23	DKH	58
Br_2CO, ∠(BrCBr)	–0.10	DKH	58
H_2O, r(OH)	0.016	Breit	156
H_2O, ∠(HOH)	–0.074	Breit	156
CH_4	–0.013	DHF	157
SiH_4	–0.066	DHF	157
GeH_4	–0.70	DHF	157
SnH_4	–2.06	DHF	157
PbH_4	–7.33	DHF	157

[a] $\Delta r = r$[relativistic] – r[non-relativistic]
[b] DKH = Douglas-Kroll-Hess method ; Breit = Breit-Pauli; DHF = Dirac-Hartree-Fock

Table 12. Diagonal Born-Oppenheimer contribution (distances in pm, angles in degrees)

Molecule	DBOC	Ref.
H_2O, r(OH)	0.003	156
H_2O, ∠(HOH)	0.015	156
H_2	0.0215	59
HF	0.0023	59
N_2	0.0012	59
F_2	0.0001	59

Table 13. Typical offset values $\Delta r = r_e - r$[MP2] for the MP2 method (distances in pm and angles in degrees).

Bond	basis	mean	median	std	max	min	Ref.
angle	cc-pVTZ	0.38 [a]	0.31 [a]		–1.37		158
	cc-pVQZ	0.25 [a]	0.21 [a]		1.09		158
CC [b]	cc-pVTZ	–0.46	–0.40	0.52	0.65	–1.72	159
	cc-pVQZ	–0.17	–0.12	0.50	0.77	–1.32	159
NH	cc-pVTZ	–0.09	–0.07	0.19	0.25	–0.57	139
	cc-pVQZ	0.01	0.00	0.19	0.40	–0.39	139
OH	cc-pVTZ	–0.31	–0.31	0.10	0.54		138
	cc-pVQZ	0.15	0.15	0.11	0.38		138

[a] for absolute values of the residuals
[b] note that the offset for the CC bond depends on the environment, see text.

Table 14. Adiabatic and Born-Oppenheimer equilibrium structures of carbon monoxide (CO) and hydrogen chloride (HCl). All values in pm.

type	isotope	*r*	isotope	*r*
adiabatic	$^{12}C^{16}O$	112.833 634 6	$H^{35}Cl$	127.457 168 7
	$^{12}C^{17}O$	112.833 399 7	$H^{37}Cl$	127.457 160 7
	$^{13}C^{16}O$	112.833 219 6	$D^{35}Cl$	127.459 403 0
	$^{12}C^{18}O$	112.833 182 5	$D^{37}Cl$	127.459 394 8
	$^{14}C^{16}O$	112.832 881 7	$T^{35}Cl$	127.460 225 8
	$^{13}C^{18}O$	112.832 767 6	$T^{35}Cl$	127.460 117 5
BO		112.8230(1) [a]		127.460 40(11) [b]

[a] Ref. [160].
[b] Ref. [161].

Table 15. Equilibrium structures of phosphine and stibine (distances in pm and angles in degees).

X	from	r_e(X-H)	$\angle_e$(HXH)	Ref.
P	PH_3	141.1607(16)	93.4184(21)	162
	PD_3	141.17846(16)	93.42523(19)	162
	CCSD(T)/wCVQZ	141.107	93.498	this work
	CCSD(T)/wCV5Z	141.098	93.501	this work
Sb	$^{121}SbH_3$	170.0013(1)	91.5586(1)	163
	$^{121}SbD_3$	170.0909(2)	91.6713(2)	163
	$^{123}SbD_3$	170.0897(3)	91.6728(3)	164
	CCSD(T)/SDB-AVQZ	170.15	91.76	164

Table 16. main experimental sources of information on the force constants [19, 100].

parameter	symbol	order of force constant
harmonic wave number	ω	2
equilibrium quartic centrifugal distortion constant	D_e	2
Coriolis coupling constant	ζ	2
inertial defect (planar molecule)	Δ	2
mean amplitude of vibration	u	2
vibration-rotation constant	α	2, 3
sextic centrifugal distortion constant	H or Φ	2, 3
second-order Coriolis	F	2, 3
isotopic effect on r_z (average structure)	δr_z	2, 3
ℓ-type doubling constants (symmetric molecule)	q_t, r_t	2, 3
diagonal Coriolis coupling (symmetric molecule)	η_J, η_K	2, 3
Fermi resonance constant	W	3
vibrational frequency or	ν	2, 3, 4
vibrational anharmonicity	x	2, 3, 4
octic centrifugal distortion constant [a]	L	2, 3, 4
vibrational dependence of the quartic centrifugal distortion constant [b]	$D_v - D_e$	2, 3, 4
Darling-Dennison resonance constant	W	2, 3, 4

[a] Ref. [165].
[b] Ref. [166].

Table 17. Comparison of ab initio, experimental and semi-experimental equilibrium parameters for a few molecules (distances in pm and angles in degees).

Molecule	parameter	ab initio	experimental	semi-experimental	Ref.
HOF	*r*(OH)	96.64	96.57(16)	96.862(8)	138
	r(OF)	143.26	143.50(31)	143.447(11)	
	∠(HOF)	97.97	97.54(50)	97.86(2)	
BHFOH	*r*(OH)	95.69		95.853(6)	138
BF_2OH	*r*(OH)	95.74		94.30(3)	167
	∠(BOH)	113.14		114.1(2)	
O=CHCl	*r*(CCl)	176.17		176.30(2)	51
SiH_3Cl	*r*(SiH)	146.95	146.84	146.96	168
	r(SiCl)	204.77	204.58	204.58	
HNO	*r*(NH)	105.18	106.3(3)	105.20(6)	154
	r(N=O)	120.76	120.6(3)	120.86(7)	
	∠(HNO)	108.10	109.1(2)	108.16(18)	

Table 18. Variation of the vibrational correction ε as a function of the moment of inertia I_0 (in uÅ^2) [a].

Molecule	number of species	I_0	ε mean	ε range	ε/I_0 (%)
HCN	11	11.707	0.049(1)	0.007	0.4
N_2O	12	40.232	0.206(4)	0.015	0.5
OCS	12	83.101	0.250(6)	0.018	0.3
OCSe	27	125.019	0.349(9)	0.026	0.3

[a] Data from Ref. [150].

Table 19. Value of the angle ∠(CCH$_g$) in H$_2$C=CH$_g$F (in degrees) [169].

year	method	value
1958	r_0	123.7
1961	r_0	129.2
1961	r_0	120.9
1974	r_g [a]	127.7(7)
1979	r_z [a]	130.8(25)
1989	r_s	124.35(63)
1992	r_0	127.6(42)
2006	r_e	125.95(20)

[a] from electron diffraction

Table 20. *c* and *d* parameters of the $r_m^{(2)}$ method for a few XCN molecules (X = H, F, Cl, Br, and I) [a].

X	c / u$^{1/2}$Å	d / u$^{1/2}$Å^2
H	0.0272	-0.0055
F	0.0198	0.0155
Cl	0.0137	0.0186
Br	0.0110	0.0207
I	0.0072	0.0211

[a] This work.

Table 21. Molecules for which a $r_m^{(1)}$ or a $r_m^{(2)}$ structure was determined.

Molecule	structural parameters	isotopologues	symmetry	ref.
SH_3^+	2	5	C_{3v}	a
ClNO	3	6	planar	b
$BrNO_2$	3	6	planar	c
$(NO)_2$	3	3	planar	d
CH_3NC	4	12	C_{3v}	e
HONO	5	5	planar	b
HCOCl	5	9	planar	f
HNSO	5	5	planar	g
HC_5N	6	14	linear	h
$(CH_3)_2O$	7	7	C_{2v}	i
N_2O_3	7	7	planar	j
1,2,5-oxadiazole	8	9	planar	k
H_2SO_4	8	4	C_2	l
$HCONH_2$	9	16	planar	i
CH_2=CHBr	9	16	planar	m
CH_2=CHCl	9	7	planar	n
CH_2=CHCN	11	14	planar	o
C_6H_5F	13	9	planar	p
$C_6H_5C\equiv CH$	13	39	planar	q

[a] Tinti, F., Bizzocchi, L., Degli Esposti, C., and Dore, L., 2006, *J. Molec. Spectrosc.*, **240**, 202.
[b] Demaison, J., Császár, A.G., and Dehayem-Kamadjeu, A., 2006, *J. Phys. Chem. A* **110**, 13609..
[c] Kwabia Tchana, F., Orphal, J., Kleiner, I., Rudolph, H.D., Willner, H., Garcia, P., Bouba, O., Demaison, J., and Redlich, B., 2004, *Molec. Phys.*, **102**, 1509.
[d] McKellar, A.R.W., Watson, J.K.G., and Howard, B.J., 1995, *Molec. Phys.*, **86**, 273.
[e] Margulès, L., Demaison, J., Rudolph, H.D., 2001, *J. Molec. Struct.* **599**, 23.
[f] Demaison, J., Boggs, J.E., and Rudolph, H.D., 2004, *J. Molec. Struct.* **695**, 145.
[g] Demaison, J., Margulès, L., Boggs, J.E., and Rudolph, H.D., 2001, *Struct. Chem.*, **12**, 1.
[h] Bizzocchi, L., Degli Esposti, C., and Botschwina, P., 2004, *J. Molec. Spectrosc.*, **225**, 145.
[i] Demaison, J., Császár, A.G., Kleiner, I., Møllendal, H., 2007, *J. Phys. Chem. A*, **111**, 2574.
[j] Demaison, J., Herman, M., Liévin, J., Margulès, L., and Møllendal, H., 2007, submitted.
[k] Vázquez, J., Demaison, J., López-González, J.J., Boggs, J.E., and Rudolph, H.D., 2001, *J. Molec. Spectrosc.*, **207**, 224.
[l] Demaison, J., Herman, M., Liévin, J., and Rudolph, H.D., 2007, *J. Phys. Chem. A*, **111**, 2602.
[m] Zvereva-Loëte, N., Demaison, J., and Rudolph, H.D., 2006, *J. Molec. Spectrosc.*, **236**, 248.
[n] Demaison, J., Møllendal, H., Perrin, A., Orphal, J., Kwabia Tchana, F., Rudolph, H.D., Willaert, F., 2005, *J. Molec. Spectrosc.*, **232**, 174.
[o] Askeland, E., Møllendal, H., Uggerud, E., Guillemin, J.-C., Aviles Moreno, J.-R., Demaison, J., and Huet, T., 2006, *J. Phys. Chem. A*, **110**, 12572.
[p] Kisiel, Z., Bialkowska-Jaworska, E., and Pszczólkowski, L., 2005, *J. Molec. Spectrosc.*, **232**, 47.
[q] Dreizler, H., Rudolph, H.D., and Hartke, B., 2004, *J. Molec. Struct.*, **698**, 1.

Figure 1. Residuals of the r_0 structure: plot of $1000(I_{exp.} - I_{calc.})$ as a function of $I_{exp.}$ for OCSe [122] (all values in uÅ^2).

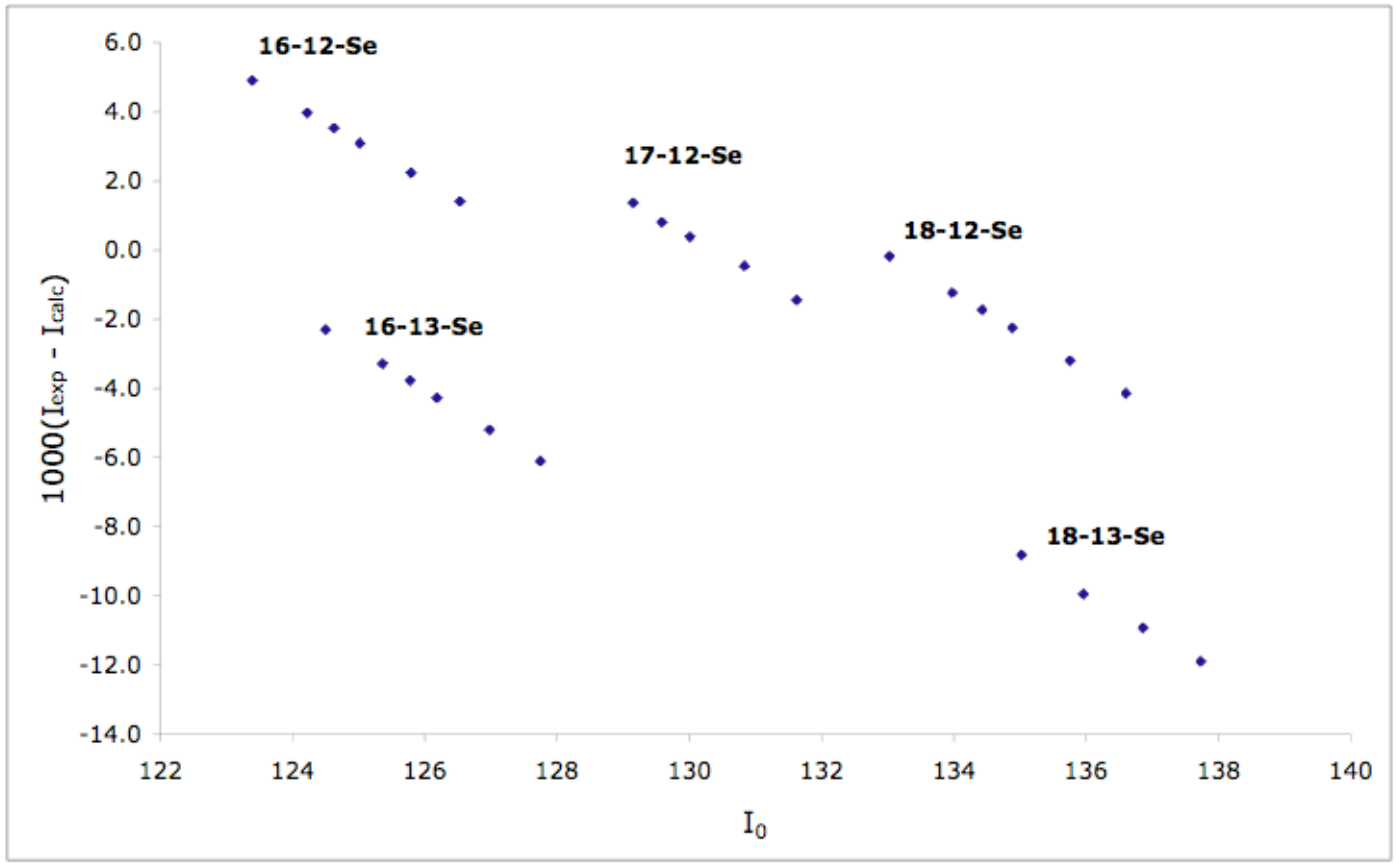


Figure 2. Plot of the variation of ε with isotopic substitution $X^{81}Br \leftarrow X^{79}Br$ for the bromide derivatives of Table 6 of Ref. [118].

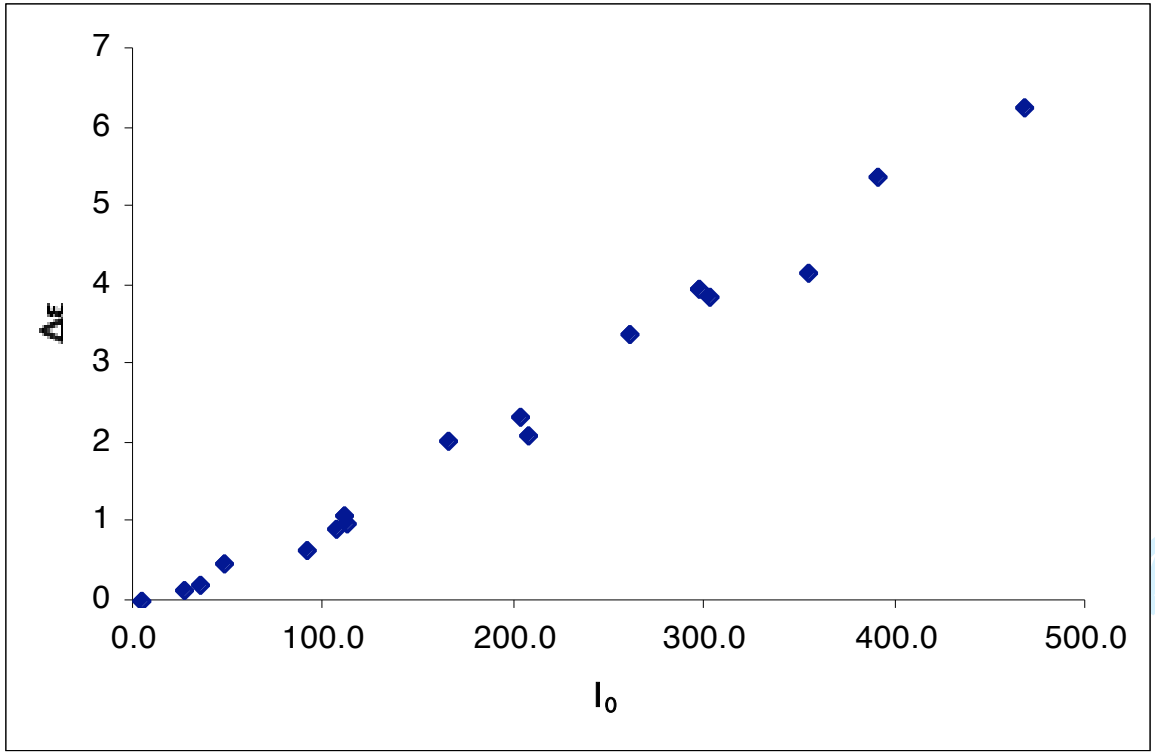


Figure 3. Plot of $\varepsilon = I_0 - I_e$ (uÅ^2) as a function of $m_x m_z/M$ for the XYZ molecule NNO [122].

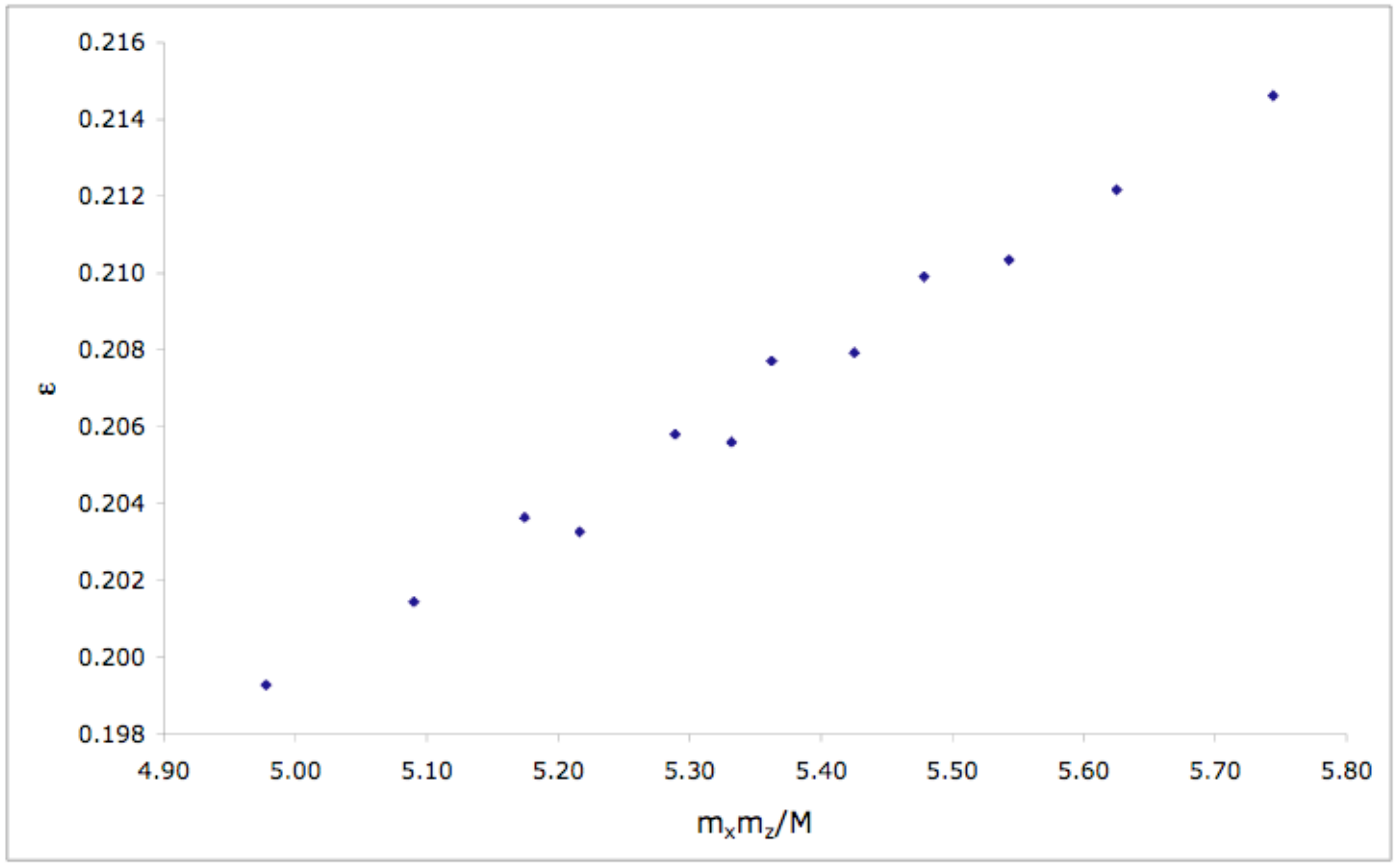
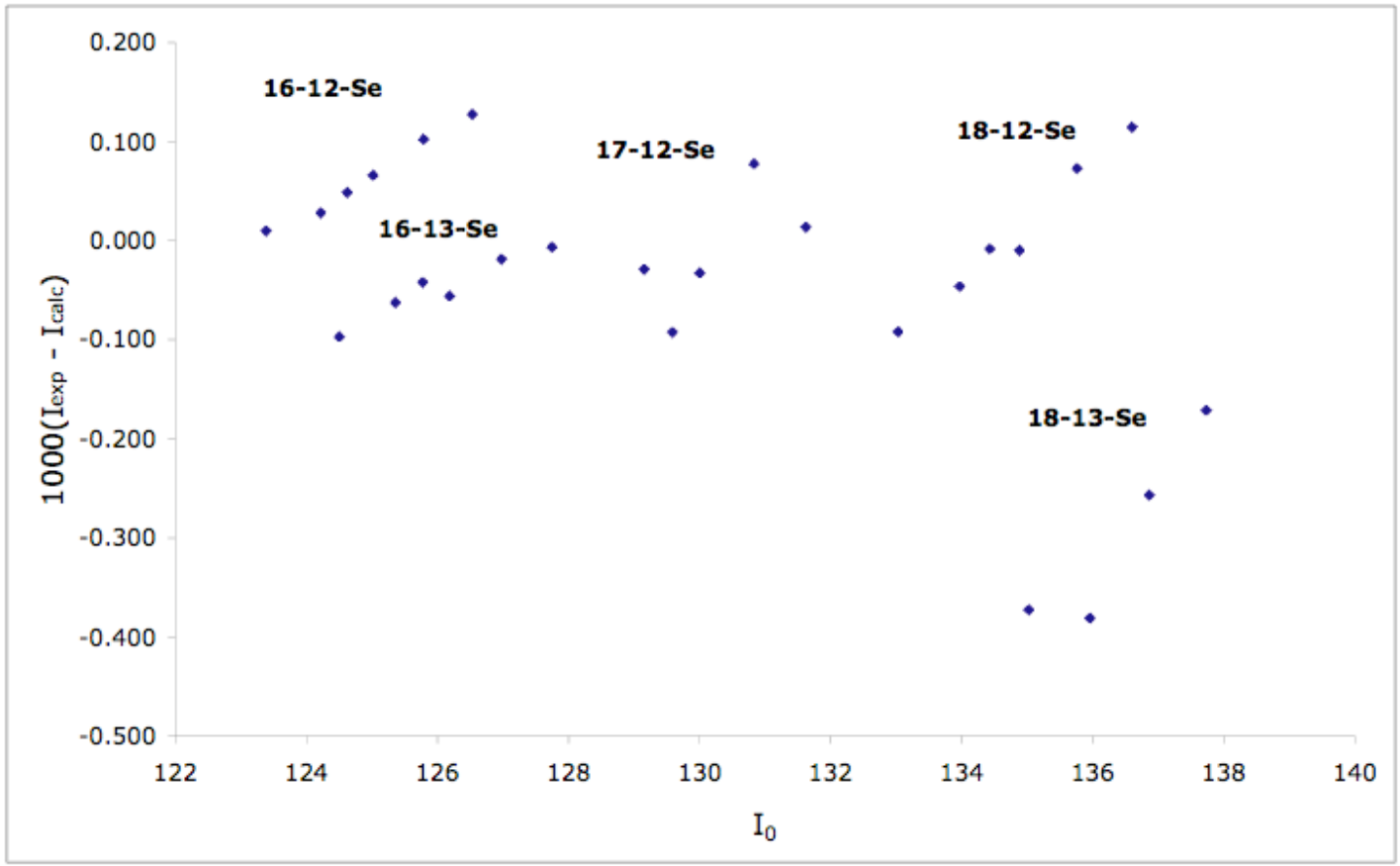


Figure 4. Residuals of the $r_m^{(2)}$ structure: plot of $1000(I_{exp.} - I_{calc.})$ as a function of $I_{exp.}$ for OCSe [122] (all values in uÅ^2).



Experimental, semi-experimental and *ab initio* equilibrium structures

J. Demaison

Laboratoire de Physique des Lasers, Atomes, et Molécules, UMR CNRS 8523, Université de Lille I, F-59655 Villeneuve d'Ascq Cédex, France

E-mail: jean.demaison@univ-lille1.fr ; fax: +33 3 20 33 70 20

Number of pages: @

Number of tables: 21

Number of figures: 4

Abstract

The determination of equilibrium structures of molecules by spectroscopic methods or by quantum mechanical calculations is reviewed. The following structures are described in detail: experimental equilibrium structures, empirical structures, semi-experimental structures and *ab initio* structures. The approximations made by the different methods are discussed and their accuracies are compared.

1. Introduction

The concept of molecular structure was developed during the 19th century. Couper [1], around 1856, was one of the first, with Kekulé [2], to state that carbon atom is tetravalent and to form a concrete idea of molecular structure. In 1885, von Baeyer [3] published his theory of ring strain whose main conclusion is that bond angles may deviate from the tetrahedral value of 109°28'. In 1874, the concept of tetrahedral carbon was simulteanously introduced by le Bel [4]and van't Hoff [5].

The golden age of molecular structure research is the first half of the 20th century, perhaps starting with the discovery of X-ray diffraction around 1900 permitting to determine interatomic distances in crystals [6]. The structure of diamond was obtained in 1913 by the Bragg's, father and son [7]. Debye [8] was the first to determine the experimental structure of a molecule (CCl_4) in the gas phase by X-ray diffraction. At about the same time, around 1930, the first gas-phase electron diffraction experiment was performed by Mark and Wierl [9]. Although the accuracy of the measurements was not high, many molecules were studied in a relatively short time and the results were used to develop concepts concerning chemical bonding by Pauling [10] and others.

In quantum chemistry, the first successful attempt to calculate the structure of a molecule was by Burrau in 1927 on H_2^+ [11]. It was followed, in the same year, by the calculation of the bond length in H_2 by Heitler and London [12].

After the second world war microwave spectroscopy had arrived [13] which is generally considered to be the most precise technique for obtaining molecular geometries in the gas phase. This spectroscopy (as well as high resolution infrared spectroscopy) now determines rotational constants with a precision close to 1 in 10^8. If there was a simple relation between experimental rotational constants and equilibrium geometry, it would be possible to determine the structure of molecules with a tremendous accuracy. However, we will see that many factors limit this accuracy.

The apparition of the computer and its considerable increase in power are also extremely important because it permitted the development of *ab initio* electronic structure methods which can now provide accurate estimates of equilibrium structures.

There are many reviews devoted to structure determination. However, most of the time, they limit themselves either to a single experimental technique or to *ab initio* calculations. There are relatively few papers where a critical comparison of the experimental and *ab initio* techniques is made and where the interplay of these methods is emphasized. Furthermore, although the techniques are described in great detail, there is no thorough discussion of the accuracy really achievable.

The first question to answer is which accuracy is desirable. There are three reasons militating in favour of a high accuracy:

- Theoreticians need accurate structures to check their more and more sophisticated computations. Although the accuracy of *ab initio* calculations varies widely, it is reasonable to define a range from 0.3 pm to better than 0.1 pm.
- Inspection of the range of a few X-Y bond lengths shows that it is rather small. For instance, it is only 6 pm for the CH and NH bonds in different molecules. Thus, to compare a particular structure in different molecules, an accuracy significantly better than 1 pm is required.
- The energy of a molecule is sensitive to its structure. Molecular mechanics programs for the calculation of properties of large molecules are parametrized against a small number of small molecules whose structure is assumed to be accurate. For instance, distortion of a C-C single bond by 2 pm "costs" 0.6 kJ mol^{-1}; distortion of a ∠(CCC) bond angle by 2°, about 0.4 kJ mol^{-1}; while the torsional distortion of a CCCC chain by 5° costs about 0.2 kJ mol^{-1} [14]. In order to be able to determine the relative energy of a molecule with an acceptable accuracy (a few kJ/mol), it is necessary to scale the molecular mechanics programs [15] with the aid of molecules whose molecular geometry is very accurately known.

The next step is to see under which conditions such an accuracy can be achieved. For this goal, we will first analyze the approximations which are made during the derivation of the molecular Hamiltonian.

2. The molecular Hamiltonian [16]

A molecule is a collection of nuclei and electrons, held together by certain forces and obeying the laws of quantum mechanics. The dominant forces are the Coulomb electrostatic forces. Gravitational forces are also present but a simple calculation shows that they are completely negligible compared to the electrostatic forces. For open-shell molecules or when heavy nuclei are present, the electron spin magnetic moment may interact with other magnetic moments (generated by the orbital motion of the electrons or of the nuclei, or magnetic moments of the other electrons). These interactions are much smaller and will be neglected in order to simplify the presentation. When they are not negligible, they can be taken into account without difficulty[1]. What is left is the Coulomb Hamiltonian. To write it, it is assumed that the nuclei are point masses and the relativistic effects are negligible. The first approximation is a very good one and will be further discussed in section 5.2. The second approximation is also a sound one. However, when heavy nuclei are present or during highly accurate electronic structure computations, relativistic effects become important and should be taken into account. This approximation will be further discussed in section 3.5. Using these approximations, the molecular Hamiltonian may be written as

$$\mathbf{H} = \mathbf{T}_\mathrm{N} + (\mathbf{T}_\mathrm{e} + \mathbf{V}_\mathrm{Ne} + \mathbf{V}_\mathrm{NN} + \mathbf{V}_\mathrm{ee}) = \mathbf{T}_\mathrm{N} + \mathbf{H}_\mathrm{e} \tag{1}$$

with

$$\mathbf{T}_\mathrm{N} = -\hbar^2 \sum_{\alpha=1}^{N} \frac{\Delta_\alpha}{2M_\alpha} \tag{2a}$$

$$\mathbf{T}_\mathrm{e} = -\frac{\hbar^2}{m} \sum_{i=1}^{n} \frac{\Delta_i}{2} \tag{2b}$$

$$\mathbf{V}_\mathrm{Ne} = -\frac{1}{4\pi\varepsilon_0} \sum_{\alpha=1}^{N} \sum_{i=1}^{n} \frac{Z_\alpha}{r_{i\alpha}} \tag{2c}$$

$$\mathbf{V}_\mathrm{NN} = \frac{1}{4\pi\varepsilon_0} \sum_{\alpha=1}^{N} \sum_{\beta>\alpha}^{N} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \tag{2d}$$

$$\mathbf{V}_\mathrm{ee} = \frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{n} \sum_{j>i}^{n} \frac{1}{r_{ij}} \tag{2e}$$

Δ is the Laplace operator, M_α and Z_α are the mass and atomic number of the nucleus α, m is the mass of the electron, $r_{i\alpha}$ is the distance between electron i and nucleus α, and similar definitions hold for r_{ij} and $R_{\alpha\beta}$.

[1] The interactions of the magnetic and electric moments of the nuclei with the other electric and magnetic moments in the molecule may lead to a splitting of the energy levels in several hyperfine components but they do not affect the following discussion.

2.1. The Born-Oppenheimer approximation and the electronic Hamiltonian [17]

This Hamiltonian is too complicated to be solved exactly. As proposed by Born and Oppenheimer (BO), the nuclear kinetic energy $\mathbf{T}_N$ is first separated when the motion of the electrons is investigated. The justification of the BO approximation is that the heavy nuclei move much more slowly than the light electrons (it also assumes that the momentum of the electrons and the nuclei is of the same order of magnitude), thus $\mathbf{T}_N$ is an almost constant term and can therefore be neglected in the differential equation. In the remaining electronic Hamiltonian $\mathbf{H}_e$, the nuclear positions $\mathbf{R}$ enter as parameters

$$\mathbf{H}_e\boldsymbol{\psi}^{(e)} = E^{BO}\boldsymbol{\psi}^{(e)} \quad (3)$$

Varying the position of $\mathbf{R}$ in small steps, one obtains E^{BO} as a function of $\mathbf{R}$. This is the potential energy (hyper)surface (PES) : $E^{BO}(\mathbf{R})$. Its minimum corresponds to the equilibrium structure of the molecule. As the nuclear masses are absent in $\mathbf{H}_e$, E^{BO} is isotopically invariant. In other words, in this approximation all the isotopologues of an individual molecule have the same PES and the same equilibrium structure irrespective of the coordinate system used to represent $\mathbf{R}$. It can be shown that the BO approximation can be trusted when the PESs corresponding to the different electronic states are well separated:

$$E_0^{BO} << E_1^{BO} << E_2^{BO} << \cdots \quad (4)$$

This is generally a good approximation for ground electronic states except for a few molecules. For instance, the first two excited electronic states of NO_3 are close to the ground state and all three levels interact via vibronic couplings [18].

The PES and the equilibrium structure can be calculated at different levels of electronic structure theory. This will be discussed in sections 3 and 6 together with the relativistic effects and the correction to the BO approximation.

2.2. Vibration-rotation Hamiltonian [19]

In the second step of the BO approximation the nuclear kinetic energy $\mathbf{T}_N$ is reintroduced and the Hamiltonian is written as

$$[\mathbf{T}_N + E^{BO}(\mathbf{R})]\phi(\mathbf{R}) \quad (5)$$

where $E^{BO}(\mathbf{R})$ may be considered as the potential.

The kinetic part of the vibration-rotation Hamiltonian was first derived by Wilson and Howard [20]. However, this original Hamiltonian is rather complicated (particularly due to the non-commutation of the various terms). Watson [21] could considerably simplify it. This

led him to introduce a small correction term U which acts as a small mass-dependent effective contribution to the potential energy. It comprises the translation of the center of mass described by three degrees of freedom (dof), the overall rotation of the molecule (three dofs) and the nuclear vibrations ($3n-6$ dofs). The separation of the translational motion is straightforward but the (approximate) separation of vibrational and rotational motion is not a trivial problem. It is treated in many textbooks [16, 19, 20].

After separation of rotation and vibration, a rotational Hamiltonian may be written for each vibrational state v as

$$\mathbf{H}_{\mathrm{R}}^{v} = \sum_{\xi=a,b,c} B_{v}^{\xi} \mathbf{J}_{\xi}^{2} + \mathbf{H}_{\mathrm{cd}}, \tag{6}$$

where B_v^ξ is the rotational constant in vibrational state v about the principal axis ξ, $\mathbf{J}_\xi^2$ is the component of the rotational angular momentum along the principal axis ξ and $\mathbf{H}_{\mathrm{cd}}$ is a small correction term called centrifugal distortion which is due to the fact that the molecule is not rigid and thus, upon rotation, the molecule is distorted by the stretching effects of the centrifugal forces.

In a perturbational treatment, the rotational constant B_v^ξ is given by

$$B_{v}^{\xi} = B_{\mathrm{e}}^{\xi} - \sum_{i} \alpha_{i}^{\xi}\left(v_i + \frac{d_i}{2}\right) + \sum_{i\geq j} \gamma_{ij}^{\xi}\left(v_i + \frac{d_i}{2}\right)\left(v_j + \frac{d_j}{2}\right) + \cdots \tag{7}$$

The summations are over all vibrational states, each characterized by a quantum number v_i and a degeneracy d_i. The parameters α_i^ξ and γ_{ij}^ξ are called vibration-rotation interaction constants of different order. B_{e}^ξ is the equilibrium rotational constant. It is proportional to the inverse of the equilibrium moment of inertia which is itself a function of the coordinates at equilibrium,

$$B_{\mathrm{e}}^{\xi} = \frac{h^2}{8\pi^2 I_{\mathrm{e}}^{\xi}(\mathbf{r}_{\mathrm{e}})} \tag{8}$$

The convergence of the series expansion is usually fast, α_i^ξ being about two orders of magnitude smaller than B_{e}^ξ and γ_{ij}^ξ two orders of magnitude smaller than α_i^ξ, see Table 1.

The original rotational Hamiltonian cannot be used to analyze rovibrational spectra because it contains many terms which are not all independent. Watson [22] submitted the Hamiltonian to unitary transformations which allowed him to reduce the number of parameters. The rotational constants are only marginally affected by the transformation. The rotational constants of an asymmetric top obtained from a fit using the Hamiltonian of Watson are affected by a small centrifugal distortion contribution which depends on the choice of the

reduction and of the representation. Watson has shown that only the following linear combinations can be determined from the analysis of the spectra:

$$B_z = B_z^{(A)} + 2\Delta_J = B_z^{(S)} + 2D_J + 6d_2 \quad (9a)$$

$$B_x = B_x^{(A)} + 2\Delta_J + \Delta_{JK} - 2\delta_J - 2\delta_K = B_x^{(S)} + 2D_J + D_{JK} + 2d_1 + 4d_2 \quad (9b)$$

$$B_y = B_y^{(A)} + 2\Delta_J + \Delta_{JK} + 2\delta_J + 2\delta_K = B_y^{(S)} + 2D_J + D_{JK} - 2d_1 + 4d_2 . \quad (9c)$$

In eqs.(9) $B_\xi^{(A)}$ are the experimental constants in the so-called *A* reduction, $B_\xi^{(S)}$ are the experimental constants in the so-called *S* reduction, and B_ξ are the determinable constants (where ξ = *x, y, z*). However, these latter constants are still contaminated by the centrifugal distortion. As shown by Kivelson and Wilson [23], the rigid rotor constants B_ξ' are given by

$$B_x' = B_x + \frac{1}{2}\left(\tau_{yyzz} + \tau_{xyxy} + \tau_{xzxz}\right) + \frac{1}{4}\tau_{yzyz} \quad (10)$$

where B_y' and B_z' are obtained by cyclic permutation of *x, y,* and *z*.

Finally, Watson has shown that the rotational constants are affected by the mass-dependent contribution *U* to the potential energy. This contribution may be written as

$$B_x^{rigid} = B_x' - \frac{1}{8}\left[\tau_{xxxx} + \tau_{xxyy} + \tau_{xxzz}\right] \quad (11)$$

The problem is that the τ constants are experimentally determinable only for a planar molecule by means of the planarity relations of Dowling [24]. For a non-planar molecule they can be calculated from the harmonic force field [25].

These equations are approximate for three reasons: i) the result is sensitive to the particular choice of the rotational constants (equilibrium, ground state experimental, ground state corrected, etc.) used in the calculations ii) the higher-order terms (sextic terms) are neglected iii) there are still additional terms as shown by Chung and Parker [26] but whose expressions are not known to date. However, as noted by Aliev and Watson [27] these last corrections are indistinguishable from the effects of the breakdown of the BO approximations.

Compared to the other corrections, the centrifugal distortion correction is generally quite small except for very light molecules, see Table 2. Furthermore, it is different from zero only for asymmetric top molecules. However, in this case, it generally remains much larger than the experimental accuracy. In many practical cases, the spectroscopist will usually fit the line frequencies measured in order to obtain what he takes to be the experimental ground state rotational constants plus all quartic (for sufficiently high *J* also the sextic) centrifugal distortion constants. He then uses these rotational constants to determine the molecular

structure. It should be better to use the determinable constants, or still better to use the ground state rigid rotor corrected constants of Eq. (11). The spectroscopist seldom goes farther than to the determinable constants.

2.3. Magnetic correction [28]

Since the electrons tend to follow the motion of the nuclei, the bulk of the electronic contribution to the rotational constants can be taken into account by employing atomic rather than nuclear masses. This is a very good approximation for most molecules and it is about the only feasible one for polyatomic molecules. However, a small correction for unequal sharing of the electrons by the atoms and for nonspherical distribution of the electronic clouds around the atoms is sometimes nonnegligible and has to be taken into account.

The total angular momentum **J** of a molecule may be written as the sum of **N**, the angular momentum due to the rotation of the nuclei and **L**, the angular momentum of the electrons. The rotational Hamiltonian for the nuclear system plus the Hamiltonian for the unperturbed electronic energies may be written as

$$\begin{aligned}\mathbf{H} &= \frac{1}{2}\sum_{\xi}\frac{\mathbf{N}_{\xi}^2}{I_{\xi}} + \mathbf{H}_{\mathrm{e}} = \frac{1}{2}\sum_{\xi}\frac{\left(\mathbf{J}_{\xi}-\mathbf{L}_{\xi}\right)^2}{I_{\xi}} + \mathbf{H}_{\mathrm{e}} \\ &= \underbrace{\frac{1}{2}\sum_{\xi}\frac{\mathbf{J}_{\xi}^2}{I_{\xi}} + \mathbf{H}_{\mathrm{e}}}_{\mathbf{H}^0=\mathbf{H}_{\mathrm{R}}+\mathbf{H}_{\mathrm{e}}} \underbrace{-\sum_{\xi}\frac{\mathbf{J}_{\xi}\mathbf{L}_{\xi}}{I_{\xi}}}_{\mathbf{H}'} + \frac{1}{2}\sum_{\xi}\frac{\mathbf{L}_{\xi}^2}{I_{\xi}}\end{aligned} \tag{12}$$

Since $\mathbf{L}_{\xi}$ is very small, the third term can be neglected and **H'** can be treated as a perturbation of $\mathbf{H}^0$. We now assume that the molecule is not in a pure $^1\Sigma$ state $\psi_0^{(0)}$ ($L = 0$) but in a perturbed state $\psi_0^{(1)}$ which has some electronic momentum. The correct effective rotational Hamiltonian is then

$$\mathbf{H}_{\mathrm{eff}} = \left\langle \psi_0^{(1)} \middle| \mathbf{H}_R + \mathbf{H}' \middle| \psi_0^{(1)} \right\rangle \tag{13}$$

A simple perturbation calculation up to second-order gives

$$\mathbf{H}_{\mathrm{eff}} = \frac{1}{2}\sum_{\xi}\mathbf{J}_{\xi}^2\left(\frac{1}{I_{\xi}} - \frac{2}{I_{\xi}^2}\sum_{n\neq 0}\frac{\left|\left\langle n\middle|\mathbf{L}_{\xi}\middle|0\right\rangle\right|^2}{E_n - E_0}\right) \tag{14}$$

This is equivalent to the definition of an effective moment of inertia $(I_{\xi})_{\mathrm{eff}}$ by

$$\frac{1}{(I_{\xi})_{\mathrm{eff}}} = \frac{1}{I_{\xi}} - \frac{2}{I_{\xi}^2}\sum_{n\neq 0}\frac{\left|\left\langle n\middle|\mathbf{L}_{\xi}\middle|0\right\rangle\right|^2}{E_n - E_0} \tag{15}$$

where I_ξ on the right is calculated using the nuclear masses. This effective moment of inertia can be expressed as a function of the molecular rotational g factor in the principal axis system, whose definition is

$$g_{xx} = \frac{M_p}{I_x}\sum_i Z_i(y_i^2 + z_i^2) - \frac{2M_p}{mI_x}\sum_{n\neq 0}\frac{\left|\langle n|\mathbf{L}_x|0\rangle\right|^2}{E_n - E_0} \tag{16}$$

and g_{yy} and g_{zz} are obtained by cyclic permutation. In this equation M_p is the mass of the proton.

The effective rotational constant B_{eff} (obtained from the analysis of the rotational spectrum) is therefore

$$(B^\xi)_{\text{eff}} = B^\xi + \frac{m}{M_p} g_{\xi\xi} B^n, \tag{17}$$

where B^ξ is the rotational constant calculated with atomic masses and B^n the rotational constant calculated with nuclear masses.

The g factor can be obtained experimentally from the analysis of the Zeeman effect on the rotational spectrum [28, 29]. It can also be calculated *ab initio* [30]. A few typical results are given in Table 3. As expected, the correction is the largest for very light molecules (as LiH) and it rapidly decreases when the mass of the molecule increases. There are, however, a few exceptions. As the expression of g shows, see Eq. (16), g may become large when an electronic excited state is close to the ground state, (because the denominator $E_n - E_0$ is small). This is the case for ozone (O_3), where the magnetic correction is extremely large.

3. *Ab initio* electronic structure methods [31, 32]

3.1.The Hartree-Fock method

The electronic Hamiltonian is still too complicated to be solved exactly (except in a few special cases). Neglecting $\mathbf{V}_{NN}$ (which is a constant term), it may be written in atomic units as

$$\mathbf{H}_e = \sum_{i=1}^{n} h_i + \sum_{i=1}^{n}\sum_{j>i}^{n} \frac{1}{r_{ij}} \tag{18}$$

The first part is a sum of monoelectronic terms which are easy to solve. The simplest approximation is to assume that each electron moves in the field created by the other electrons. The second term will be approximated by a monoelectronic operator $u(r_i)$ which will take into acount the mean repulsion effect of all the other electrons on electron i

$$\begin{aligned}\mathbf{H}_e &= \sum_{i=1}^{n} \underbrace{\left[h(r_i) + u(r_i)\right]}_{F(r_i)} + \underbrace{\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{1}{r_{ij}} - \sum_{i=1}^{n} u(r_i)}_{V} \\ &= \sum_{i=1}^{n} F(r_i) + V = \mathbf{H}_0 + V\end{aligned} \tag{19}$$

If V is small (it is expected to be much smaller than V_{ee}), it is a good approximation to replace $\mathbf{H}_e$ by $\mathbf{H}_0$ which is monoelectronic and, thus, easily solvable. This is called the Hartree-Fock (HF) method. The total HF energy is obtained by adding the V_{NN} value to the eigenfunctions of $\mathbf{H}_0$. A monoelectronic wavefunction (called orbital) is attributed to each electron.

Generally a basis set expansion technique is used. The many-electron wavefunction is written as a product (more exactly a Slater determinant) of orthonormal one-electron functions called molecular orbitals (MOs)

$$\Psi = \det(\varphi_1 \varphi_2 \cdots \varphi_n) \tag{20}$$

Each of these MOs is expressed as a linear combination of basis functions

$$\varphi_i = \sum_k c_{ik} \chi_k \tag{21}$$

The unknown c_{ik} coefficients are determined by the variational method. For purely practical reasons, the χ_k functions are usually taken as Gaussians. A single basis function is composed of one or more primitive Gaussian functions. For example, an *s*-type basis function is

$$\chi(1s) = \sum_{i=1}^{N} d_i e^{-\alpha_i f^2 r^2} \tag{22}$$

N is the number of primitive functions, called degree of contraction, d_i the contraction coefficients, α the exponent and f a scale factor.

For a closed-shell molecule with n electrons, the $n/2$ orbitals of smallest energy (which is negative) are called occupied orbitals and the orbitals of higher energy are called virtual orbitals.

Typical errors are 1% in bond distances (which are typically underestimated). Although this result is remarkable, it is not accurate enough. The main weakness of the HF method is that it is a mean field approach which does not take into account the instanteneous interactions between the electrons: they are allowed to be at the same point in space.

3.2. Post-Hartree-Fock methods [33, 34]

The difference between the exact non-relativistic energy and the HF energy is called the correlation. If ψ_0 is the HF wavefunction, a better approximation of ψ is

$$\psi = a_0\psi_0 + a_1\psi_1 + a_2\psi_2 + \cdots \tag{23}$$

where ψ_i ($i > 0$) correspond to configurations involving virtual orbitals (excited determinants). If all the virtual orbitals corresponding to a given basis are included, the wavefunction is called the full configuration interaction (CI) wavefunction. In this FCI wavefunction, there are three categories of correlation corrections

1. excitations whose individual contributions are small, but their total contribution is large because of their great number. This is called *dynamic correlation*. It enables electrons to stay apart and it is usually the largest part of the correlation energy.
2. excitations required to provide a correct zeroth-order description as dictated by spin-symmetry considerations. It occurs for open-shell molecules and it is called *static correlation*.
3. excitations whose coefficients a_i are large. It is called *nondynamic correlation.*

The CI calculations are computer intensive, they are thus limited to small systems. Furthermore, they generally require a lot of experience.

Another method to estimate the correlation energy is provided by many-body perturbation theory with V being the perturbation term. The second-order theory gives, among others, the popular Møller-Plesset 2 (MP2) method [35], which recovers about 90% of the correlation energy. It is possible to use higher-order perturbation theory (MP3, MP4, …). The problem is that the perturbation series converges rather poorly if it converges at all.

To improve the accuracy, it is better to use the so-called coupled cluster (CC) method which takes into account the "instantaneous" interactions between the electrons [36, 37]. When one electron collides with another electron, it may be excited from an occupied orbital

φ_i to a virtual orbital φ_α. The most frequent "collision" is between two electrons which corresponds to the transition from two occupied orbitals *i, j* to two virtual orbitals α, β.

If Ψ_0 is the HF wavefunction and if $\mathbf{T}_2$ is the operator which induces this biexcitation,

$$\mathbf{T}_2\Psi_0 = \frac{1}{4}\sum_{ij\alpha\beta} t_{ij}^{\alpha\beta}\Psi_{ij}^{\alpha\beta}. \tag{24}$$

A better description is obtained by also taking into account single excitations, triple excitations, quadruple excitations, etc.,

$$\mathbf{T} = \mathbf{T_1} + \mathbf{T_2} + \mathbf{T_3} + \mathbf{T_4} + \cdots, \tag{25}$$

and write

$$\begin{aligned}\Psi_{CC} &= e^{\mathbf{T}}\Psi_0 \\ &= \Psi_0 + \sum_{i<j}\sum_{\alpha<\beta} t_{ij}^{\alpha\beta}\Psi_{ij}^{\alpha\beta} + \cdots\end{aligned} \tag{26}$$

This is called the coupled cluster method. The CCSD method stops at double excitations (more exactly, all operators that involve more than two electrons are omitted) and reduces the error by a factor of three to four. The CCSDT method stops at triple excitations and further reduces the error by a factor of three to four.

If *n* is the number of basis functions and *m* is the order of the clusters (*m* = 2 for CCSD, *m* = 3 for CCSDT, *m* = 4 for CCSDTQ, the computation time is proportional to n^{2m+2}. To reduce this cost, the triple excitations may be taken into account perturbationally. A variant gives the CCSD(T) method whose cost is proportional to n^7 (instead of n^8 for CCSDT).

The CC methods estimate accurately the dynamic correlation but, when the nondynamic correlation is large (coefficients > 0.2), the accuracy is reduced and it may be necessary to use multireference methods. There are several diagnostics which predict the size of the nondynamic correlation [38], the most popular one being the T_1 diagnostics of Lee and Taylor [39]. In conclusion, when the nondynamic correlation is small, we have at our disposal a hierarchy of approximations of increasing accuracy :

HF < MP2 < CCSD < CCSD(T) < CCSDTQ < …

Helgaker et al. [40] compared the performances of the different methods for several bond lengths. The results are given in Table 4. The CCSD(T) method gives the best results, the next best method being the MP2 one. The question is now to check whether a larger basis set than the cc-pVQZ one or a method more sophisticated than CCSD(T) would improve the accuracy. Halkier et al. [41] studied the performance of the CCSDT method for structure computations. They found that it was generally less accurate than the cheaper CCSD(T) one. The small error

due to the perturbation calculation is nearly compensated by the error due to the neglect of quadruple excitations. Thus, the CCSD(T) method is cheaper and more accurate than the CCSDT method. The improvement brought by the CCSDTQ method (at complete basis set) was analyzed by several authors [42, 43]. The results are given in Table 5. It is seen that the CCSD(T) method performs quite well, the largest error, 0.34 pm being found for F_2 within the test set. This is true when the nondynamical correlation is small. When it is large, the situation is completely different as can be seen on the example of HONO (and other NOx molecules) where the O-N bond is 0.8 pm too short at the CCSD(T) level, see Table 6.

3.3. Choice of the basis sets

There is a great number of atom-centered Gaussian basis sets but it is recommended to use a hierarchy of basis sets which provide a systematic approach to the complete basis set (CBS) limit. One of the most popular ones are the correlation-consistent polarized valence basis sets of Dunning [44]: cc-pV*n*Z where n = D, T, Q, 5, … is the cardinal number and usually represents the highest spherical harmonic of the polarization functions. The cc-pVnZ basis sets are designed for correlation of valence electrons only. To correlate all electrons, the correlation-consistent core-valence cc-pCV*n*Z, basis sets [45] (or the correlation-consistent weighted core-valence basis sets, cc-pwCV*n*Z [46] which significantly improve the convergence with n) have to be used (Gaussians with large exponents are added to the cc-pV*n*Z basis sets). On the other hand, for an accurate description of the outer valence region, diffuse functions (Gaussians with small exponents) are added to the cc-pV*n*Z basis sets to give the aug(mented)-cc-pV*n*Z basis sets [47]. Thus the most general basis sets are aug-cc-pCV*n*Z.

It is important to have a look at the size of the basis sets because it determines, together with the method, the computation time. For the cc-pV*n*Z basis sets, the number of contracted functions $N_V(n)$ increases as the third power of the cardinal number n, see Table 7, which also shows that the core-valence sets are considerably larger than the valence sets and that the number of diffuse functions increase quadratically with the cardinal number.

3.4. Strategy to calculate a structure

Ideally, one should use the CCSD(T) method (or better) with the aug-cc-pwCV*n*Z basis sets ($n \geq 3$), all electrons being correlated. This is probably the best method for small molecules but, due to the large size of these basis sets, it is extremely computer intensive. Different procedures are used. One of the most common is to use the cc-pVnZ basis sets,

correlating only the valence electrons (the so-called frozen core approximation) and to increase the value of n, the cardinal number, as much as possible. The rate of convergence depends on the kind of bond. For instance, for the CH bond, convergence is almost achieved at $n = 3$. On the other hand, for the O···H bond (hydrogen bond), it is necessary to go at least up to $n = 6$, see Table 8.

Then, the effect of the core and core-valence correlation is calculated with the cc-pwCVQZ basis set [48 ,49]. If there is no heavy atom present ($Z < 20$), the MP2 method can be safely used to calculate this correction [50]. On the other hand, the use of the cc-pwCVQZ basis set somewhat limits the accuracy because the convergence is not yet fully achieved, see Table 9 which also shows that the correction calculated at the MP2 level is slightly too large (in absolute value). The error is quite small for first-row atoms but it becomes sizeable when a second-row atom is involved. For instance, for the C-Cl bond in ClCN, the MP2 method gives a correction of –0.43 pm whereas the CCSD(T) method gives –0.36 pm [51]. (As the cc-pwCVQZ basis set is quite large, it may be advantageous to use a completely decontracted cc-pVTZ basis set supplemented by an appropriate (1p3d2f) primitive set. This basis set called Martin-Taylor basis set (denoted as MT) is significantly smaller without any loss in accuracy [48].

Finally, if necessary, the contribution of the diffuse functions is added. It can also be calculated at a lower level of theory (e.g. MP2) [52]. The effect of the diffuse functions may be large if the basis set is small ($n \leq 3$) but it decreases rapidly when n increases and for $n \geq 5$, it is negligible in most cases, even for the C-F bond length in CH_3F, which is particularly sensitive to the effect of diffuse functions, see Table 10 (the van der Waals and hydrogen bonds are, however, important exceptions, see section 3.7). In conclusion, when the effect of the diffuse functions is calculated with a smaller basis set, the correction appears to be too large.

We have seen that there are two main sources of error: the basis set convergence error and the electronic structure method error. There are also two other approximations which may limit the accuracy: the relativistic effects and the BO approximation itself.

3.5. Relativistic effects [53]

The best method to take into account relativistic effects is to use a fully relativistic Dirac Hamiltonian. However, it is very demanding in computational resources and various methods have been developed to estimate relativistic effects.

When heavy atoms are present, the most widely used method is the pseudopotential approximation [54, 55], because it avoids the basis functions necessary for the description of the electronic core and for the inner nodal structure of the valence orbitals.

Since the effects of relativity are small for the great majority of usual atoms (they scale up to Z^4), perturbation theory may be successfully used. A number of approximate methods have been developed. One of the most widely used is the Douglas-Kroll-Hess (DKH) method [56, 57, 58] which recovers most of the scalar relativistic effects. A few typical values are given in Table 11. For chlorine and lighter atoms, this correction is smaller than 0.07 pm (value for SiH_4) but it becomes important for heavy atoms.

3.6. Correction to the Born-Oppenheimer approximation [59, 60]

A first-order Born-Oppenheimer correction which is diagonal in the electronic state is straightforwardly calculated

$$\Delta E^{\mathrm{DBOC}} = \left\langle \psi_0^{(e)} \middle| \mathbf{T_N} \middle| \psi_0^{(e)} \right\rangle \tag{27}$$

It is called adiabatic correction (or diagonal BO correction) and it is a good approximation if the ground electronic state is much lower than the electronic excited states. It is usually small (it is roughly proportional to Z^2). It is important to note that it is mass dependent. Table 12 gives a few values which show that this correction is normally negligible.

3.7. Accuracy of the coupled cluster methods

The two main sources of errors are the basis set incompletness error and the electronic structure method error. In Table 5, it is seen that the CCSD(T) method calculates a bond length in F_2 which is 0.34 pm too long (compared to CCSDTQ). Table 8 shows that the basis set convergence can be quite slow for some bonds. However, the errors are not additive. The bond length normally decreases with the size of the basis set whereas it increases with the level of theory. It is thus possible to make use of the concept of balanced calculation for which there is a near cancellation of the errors. For instance, in F_2, the experimental value of the bond length is 141.19 pm and the CCSD(T)/cc-pCVQZ value is 141.13 pm whereas the CCSD(T)/cc-pCV6Z value is much worse at 140.87 pm. This behaviour is quite general. However, a balanced calculation is difficult to use because neither the errors nor the true value are known in advance. A detailed study must be carried out in each case.

As shown in Table 4, the accuracy of the CCSD(T) method is 0.3 pm (or better) for a molecule without heavy atom ($Z < 18$) and with a small nondynamical correlation. Obviously,

it is possible to improve the accuracy. Inclusion of quadruple excitations (CCSDTQ level) changes the computed values for bond distances up to a few tenths of a picometer [42, 43, 61] and significantly improves the agreement of the theoretical results with the corresponding values. However, the computational cost is very high and the CCSDTQ method is still limited to small molecules, even with a small basis set.

Another difficulty, is the slow convergence of the basis set. Owing to the presence of the Coulomb potential, the molecular electronic Hamiltonian becomes singular when two electrons coincide in space and give rise to a characteristic cusp in the wave function [32] that the traditional basis sets are unable to describe correctly. Two solutions have been proposed

- basis set extrapolation (BSE) schemes which make use of well established extrapolation formulas for energies when using the correlation-consistent basis sets of Dunning [62].
- R12 methods which explicitely include the interatomic distance in the wave function ansatz which permit to correctly take into account the correlation cusp [63, 64].

However, there are contributions that are of similar order of magnitude as the BSE effects such as relativistic contributions and quintuple excitations.

3.8. Lower-level methods

For large molecules, the CCSD(T) method is still too expensive. It is therefore interesting to check the accuracy of lower-level *ab initio* calculations which are more readily accessible. The MP2 method has been shown to perform rather well, see Table 4. It also appeared that the remaining errors are generally mainly systematic and correction factors, or "offsets" can be derived empirically in order to predict molecular structures with an accuracy which is competitive with the best experimental methods (i.e. a few tenths of pm). The first obstacle is that the offset values are basis set dependent. Moreover, for a given basis set, the offset is not always constant, but may vary as the true equilibrium distance varies. In addition, the offset is sometimes a function of substituent effects. For instance, for the MP2/cc-pVTZ value of r(CC) in benzene: $\Delta r = r_e - r(\text{calc.}) = -0.13$ pm whereas the offset at the same level of theory is $\Delta r = -0.65$ pm for NC-CN. This large difference cannot be explained by the variation of the r_e value because they are quite close. Constancy of the offset value in a given type of bond implies that similar errors occur in the calculation of that value, i.e. the finite basis set creates the same error, the partial neglect of electron correlation has the same effect on the calculated bond length, etc. Consequently, it is not surprising that the magnitude of the

offset value is at least somewhat responsive to environmental perturbations from the surroundings of the bond. The conclusion of this discussion is that the offset method is useful but has to be used with caution.

A few typical offset values are given in Table 13. Offsets have also been determined for the CO bond (where the offset is negative and decreases in absolute value when the bond length increases) and the CN bond [65].

4. Least-squares method

4.1. Introduction

Most of the empirical and experimental methods which will be described use the least-squares method to determine the structural parameters. This method has been reviewed with emphasis on structure determination [66, 67]. For details, the reader is referred to these reviews. Therefore, we will limit ourselves to describe some shortcomings of this method and how to possibly circumvent them.

In the following we use the notation of these reviews which agrees with that of the tutorial by Albritton et al. [68]. If we want to determine p parameters β_j (j = 1, …, p) from n experimental data y_i (i = 1, …, $n > p$), the starting equations of the linear least-squares method (with unequally-weighted and/or correlated errors) are, after the model equations have been linearized in the neighborhood of the expected solution,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\Theta}_\mathbf{y} = \sigma^2\mathbf{M} \qquad (28)$$

where $\mathbf{X}$ is the Jacobian matrix, $\boldsymbol{\beta}$ is the vector of the parameters to be determined by the fit, and $\boldsymbol{\varepsilon}$ the vector of residuals (which should not be confused with the vibrational correction ε used elsewhere in this paper). $\boldsymbol{\Theta}_\mathbf{y}$ is the covariance matrix of the observations $\mathbf{y}$, σ the standard deviation of the fit. How $\boldsymbol{\Theta}_\mathbf{y}$ is subdivided into σ and the positive definite matrix $\mathbf{M}$ depends on the type of fit. For a general least-squares fit (weighted fit, with unequal and/or correlated errors of the observations), $\mathbf{M}$ is usually taken to be the covariance matrix of the experimentally measured observations $\boldsymbol{\Theta}_\mathbf{y} = \sigma_w^2\mathbf{M} = \sigma_w^2\boldsymbol{\Theta}_{\mathbf{y}\mathrm{exp}}$, then σ_w is dimensionless and should be near unity for a perfect model and a realistic assessment of the observational errors which, however, is very rarely the case in the spectroscopic applications discussed here (see remark on p. 39 of ref. [67]). For the equally (unity-) weighted and uncorrelated problem, $\mathbf{M}$ is chosen as the identity matrix $\mathbf{I}$, $\boldsymbol{\Theta}_\mathbf{y} = \sigma_u^2\mathbf{M} = \sigma_u^2\mathbf{I}$, then σ_u has the dimension of the

observations and is, in the present applications, often larger by one to two orders of magnitude than the experimental errors of the observations.

For a solution, Eq. (28) must be reduced to that of the equally-weighted and uncorrelated problem, where **M** is the identity matrix. For this purpose we define a square matrix **P** transforming the errors **ε** such that

$$\gamma = \mathbf{P}^{\mathrm{T}}\varepsilon \quad \text{and} \quad \left\langle \gamma\gamma^{\mathrm{T}} \right\rangle = \mathbf{P}^{\mathrm{T}}\left\langle \varepsilon\varepsilon^{\mathrm{T}} \right\rangle \mathbf{P} = \mathbf{P}^{\mathrm{T}}\Theta_{\mathbf{y}}\mathbf{P} = \sigma^2\mathbf{P}^{\mathrm{T}}\mathbf{M}\mathbf{P} = \sigma^2\mathbf{I} \tag{29}$$

where **I** is the identity matrix. If **Λ** is the diagonal matrix of the eighenvalues of **M**, and **U** the corresponding matrix of the eigenvectors, it is easy to show that choosing

$$\mathbf{P} = \mathbf{U}\Lambda^{-1/2} \tag{30}$$

yields

$$\mathbf{M} = (\mathbf{P}^{\mathrm{T}})^{-1}\mathbf{P}^{-1} = \mathbf{P}\mathbf{P}^{\mathrm{T}} \tag{31}$$

left-mutiplying eq. (28) by $\mathbf{P}^{\mathrm{T}}$, we obtain the original problem with unequal variances (data of different precision) and/or correlated errors in the formulation of an equally-weighted uncorrelated problem

$$\underbrace{\mathbf{P}^T\mathrm{y}}_{\mathbf{y}'} = \underbrace{\mathbf{P}^T\mathbf{X}}_{\mathbf{X}'}\beta + \underbrace{\mathbf{P}^T\varepsilon}_{\gamma} \quad \text{i.e.} \quad \mathbf{y' = X'\beta + \gamma}, \quad \mathbf{\Theta_{y'} = \sigma^2 I} \tag{32}$$

Thus, In the general case of unequal variances (data of different precision) or correlated errors, the problem is transformed to the unity-weighted uncorrelated problem.

with the solution (^ designates the least-squares estimator)

$$\ddot{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{y} \tag{33}$$

and

$$\ddot{\Theta}_{\ddot{\beta}} = \ddot{\sigma}^2\left(\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{X}\right)^{-1} \tag{34}$$

with

$$\ddot{\sigma}^2 = \left(\mathbf{y} - \mathbf{X}\ddot{\beta}\right)^{\mathrm{T}}\mathbf{M}^{-1}\left(\mathbf{y} - \mathbf{X}\ddot{\beta}\right)\Big/\left(n - p\right) \tag{35}$$

where $\mathbf{M}^{-1}$ is known as the generalized weight matrix. Thus, the following discussion is valid provided the transformed Jacobian, observations, and residuals are used.

4.2. Assumptions of the least-squares method

The application of the least-squares method requires certain assumptions (Gauss-Markov conditions). The distribution of the errors should be random and have zero mean and a finite second moment (the distribution need not be a "normal distribution"). If the observations to be fitted, the moments of inertia, were the highly accurate equilibrium

moments of inertia (without systematic errors), it would probably be true. But, since generally only approximations of the equilibrium moments of inertia can be used, contaminated by systematic errors, this assumption does not hold. The errors are not at all random, on the contrary, they are correlated (in fact they are roughly proportional to the inertial moments). There have been a few rare attempts to take this correlation into account [69], but it is not easy because neither the theoretical form of the correlation matrix is known nor are there enough experimental data to use estimation procedures.

In theory, the correlation should not bias the estimates, but this is true only in the limit of an infinite number of data. In the particular case of a structure determination, the number of experimental data is often not much greater than the number of parameters to be determined. Furthermore, as the mean of the errors is not zero, a bias is expected. Due to the nature of the system of equations, this bias can be large (see section 4.3).

4.3. Choice of the weighting

Generally the general least-squares method (weighted lsq, unequal and/or correlated errors) is used. Usually the weight of an observation is taken as the inverse square of the experimental uncertainty. However, as mentioned before, in the present applications the inherent model limitations due to the lack of sufficiently well known vibrational corrections generally cause errors which are much larger than the experimental errors of the inertial moments, although the latter may themselves differ greatly in magnitude, usually being larger for the less abundant species. This is disclosed, e. g., by a (possibly preliminary) weighted fit of the inertial moments using their experimental uncertainties as errors. Even when the moments come from well conducted spectral measurements, the standard deviation σ_w of this weighted fit may, nonetheless, come out one to two orders of magnitudes larger than unity, the value expected for a perfect model and correctly assessed errors. Multiplying, in a new weighted fit, all experimental errors with this standard deviation σ_w would then indeed result in a standard deviation of unity. Epple et al. [70] have proposed a method to assign weights which is hopefully more appropriate in this situation. They assign to each inertial moment $I^{\xi}_{\mathrm{exp}}(i)$ an additional "model-induced" error, Eq. (36), where $\Delta I^{\xi}_{\mathrm{rovib}}(i)$ is the rovibrational contribution to the inertial moment (for $\xi = a, b, c$, and isotopologue i) calculated (by any approximate theory) from the results of an earlier fit,

$$\mathrm{error}\left(I^{\xi}_{\mathrm{exp}}(i)\right) = \ddot{\omega}_w \cdot \sqrt{\frac{3\left(\Delta I^{\xi}_{\mathrm{rovib}}(i)\right)^2}{\sum_{\xi'=a,b,c}\left(\Delta I^{\xi'}_{\mathrm{rovib}}(i)\right)^2}} \quad \text{for} \quad \xi = a,b,c, \ \text{all} \ \ i \tag{36}$$

The true experimental errors of the inertial moments could then increase up to the values given by Eq. (36) representing the "apparent" (model-induced) errors, before the standard deviation of the fit of approximately unity would start to increase. In a crude form, this additional error even takes account of the different magnitude of the inertial moments for ξ = *a, b, c*. In practice, the square of Eq. (36) is added to the experimental variance of the respective moment on the diagonal of the covariance matrix of the observations. The procedure was used for the determination of the structure of 2-chloropropane [71], cyclopropylgermane [70], and several dicyanides [72], where the usual weighting scheme (by reciprocal squares of the experimental errors only) would have resulted in a molecular structure with very unbalanced errors due to very differently weighted moments. In all cases the additional "model-induced" variance was much larger than the experimental variance. Since the procedure does not change the covariances, the additional terms on the diagonal lead also to a substantial reduction of the correlations between the observations. From a theoretical standpoint, the method remains questionable because systematic (model) errors have been simulated by random errors.

The weights may also be calculated by the iteratively reweighted least-squares method [73] which, in contrast to Eq. (36), makes no assumptions regarding the dependence of the weights on the magnitude of the inertial moments and may hence be better suited when the model errors go back to the inadequate approximation of the vibrational correction ε. Originally, the reweighted least-squares method was devised to detect outliers due to faulty measurements etc. It is also appropriate to find suitable weights to compensate for the deficiencies of the model which, in principle, cannot be made perfect.

If possible, the weight finding by the iteratively reweighted least-squares method should be applied separately to each of the three component sets ξ = *a, b, c* of the inertial moments I^{ξ}_i, in particular, when the errors of the three sets are very different. An initial set of residuals

$$e^{\xi}_i = I^{\xi}_i(\mathrm{exp.}) - I^{\xi}_i(\mathrm{calc.}) \tag{37}$$

is obtained by the application of ordinary least-squares. Then weights are assigned to the experimental inertial moments which are functions of these residuals. Different weight functions are in use [73]. Some of them assign the weight zero to inertial moments whose

residual, Eq. (37), is larger than a certain threshold value, which means that the respective inertial moment is dropped from the fit as a suspected "outlier". For other weight functions, the weights decrease more gradually with increasing residuals. Since for the problems treated here the number of input data is usually not much larger than the number of parameters to be determined, one can hardly afford the loss of data (see Eq.(35)); therefore, the latter type of weight functions appears to be preferable. With the newly assigned weights a new (weighted) least-squares fit is performed which will result in new residuals and new weights. The treatment is iterated until consistence of the residuals is attained. The reweighted least-squares procedure was already used for the structure determination of a few molecules but there is not yet any systematic study of its performance. A number of problems remain, in particular the presence of autocorrelation of the input data (non-zero covariances between the errors of inertial moments) requires further study.

In conclusion, most assumptions of the least-squares method are not fullfilled by the problem as originally given. The difficulties can be circumvented, at least partly, by a judicious manipulation of the variances of the observations.

4.4. Collinearity or ill conditioning

It often happens that some parameters cannot be estimated with precision and that they are very sensitive to small perturbations in the data. This is due to "collinearity", or, more exactly, to the near-collinearity of the fit vector subspace spanned by p column vectors of the Jacobian matrix $\mathbf{X}$ in error space (the n-dimensional space of the vectors of the observations and of their residuals, $\mathbf{y}$ and $\boldsymbol{\varepsilon}$, respectively). This collinearity increases the variances of the estimated parameters and is responsible for important round-off errors. This problem was pointed out early by Kuchitsu et al. [74]. There are diagnostics that determine whether a collinearity exists and that can identify the near-dependent parameters affected. The correlation matrix is often employed for that purpose. But the absence of high correlations does not imply the absence of collinearity. Therefore many different procedures have been proposed.

Belsley [75] has critically reviewed these procedures and has concluded that "none is fully successful in diagnosing the presence of collinearity and variable involvement or in assessing collinearity's potential harm". To palliate the weaknesses of the existing diagnostics, he has introduced the condition indexes and has shown that they can be easily used to determine the strength and number of near-dependencies. First, the columns of the Jacobian matrix $\mathbf{X}$ are scaled to have unit length (each term of the vector column $\mathbf{X}_i$ is divided by the

norm $\|\mathbf{X}_i\|$). Then the singular values (square roots of the eigenvalues of $\mathbf{X}^T\mathbf{X}$) of the scaled $\mathbf{X}$ matrix are calculated: μ_1, μ_2, … μ_p. The singular-value decomposition of $\mathbf{X}$ may be written [76]

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{38}$$

where $\mathbf{U}$ is a $n \times p$ matrix with orthonormal columns, $\mathbf{D}$ is a $p \times p$ diagonal matrix whose elements are the singular values μ_i of $\mathbf{X}$, and $\mathbf{V}$ is a $p \times p$ orthonormal matrix. The "scaled condition indexes" of the scaled matrix $\mathbf{X}$ are defined by:

$$\eta_k = \frac{\mu_{\max}}{\mu_k} \qquad k = 1, \cdots p \tag{39}$$

The highest condition index is the condition number $\kappa(\mathbf{X})$. It is an error magnification factor and it is used to determine whether a matrix is ill-conditioned or not: if the data are known to d significant figures and if the condition number of $\mathbf{X}$ is 10^r, then a small change in the data in its least significant digit can affect the solution in the $(d\text{-}r)$th place. If we consider a perturbation $\delta\mathbf{y}$ in $\mathbf{y}$, then

$$\left\| \frac{\delta\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}} \right\| \le \kappa(\mathbf{X}) \left| \frac{\delta\mathbf{y}}{\mathbf{y}} \right| \tag{40}$$

where $\hat{\boldsymbol{\beta}}$ is the least-squares solution and $\delta\hat{\boldsymbol{\beta}}$ the perturbation in $\hat{\boldsymbol{\beta}}$ due to $\delta\mathbf{y}$. Thus a large κ means a higher bound for the relative perturbation of the solution and may be responsible for a large systematic error.

The number of near-dependencies is equal to the number of high scaled condition indexes. To determine which parameters are involved in the collinearities, Belsley defines the variance-decomposition proportions. The variance-covariance matrix $\hat{\boldsymbol{\Theta}}(\hat{\boldsymbol{\beta}})$ of the least-squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is (where $\hat{\sigma}$ is the estimated standard deviation of the fit)

$$\begin{aligned} \hat{\boldsymbol{\Theta}}(\hat{\boldsymbol{\beta}}) &= \hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1} = \hat{\sigma}^2\mathbf{V}\mathbf{D}^{-2}\mathbf{V}^T \\ \mathrm{var}(\hat{\beta}_k) &= \hat{\sigma}^2 \sum_j \left(\frac{v_{kj}}{\mu_j} \right)^2 \end{aligned} \tag{41}$$

Eq. (41) decomposes the k^{th} diagonal element $\mathrm{var}(\ddot{\beta}_k)$ of $\ddot{\Theta}(\ddot{\beta})$ into a sum of components, each of which depends on a single μ_j. For a small μ_j the value of the component and hence its contribution to $\mathrm{var}(\ddot{\beta}_k)$ will be large. Unusually large variance proportions of two (or more) parameter estimates, say $\ddot{\beta}_k$ and $\ddot{\beta}_{k'}$, concentrated in components associated with the same

small μ_j provide evidence that these parameters are involved in a near-dependency due to that small singular value μ_j . This can be more readily detected when variance-decomposition proportions have been defined as the elements $0 < \pi_{jk} \leq 1$ of a matrix $\boldsymbol{\pi}$,

The variance-decomposition proportions are

$$\pi_{jk} = \left(\frac{v_{kj}}{\mu_j}\right)^2 \Bigg/ \sum_{j'=1}^{p}\left(\frac{v_{kj'}}{\mu_{j'}}\right)^2 \qquad k, j = 1 \cdots p \tag{42}$$

Large elements in the same row j (of a small μ_j) for two (or more) columns, say k and k', point to a near dependency between parameters $\ddot{\beta}_k$ and $\ddot{\beta}_{k'}$. Near-dependency will degrade the parameter estimates involved. Belsley proposes the following rule of thumb: estimates are degraded when two or more variances have at least half of their magnitude associated with a scaled condition index of 30 or more. Evidently the probability of encountering the problem of ill-conditioning rapidly increases with the number of parameters to be determined. However, it is already present for small molecules. It is easy to explain why: quite often, an isotopic substitution induces a very small change in the moments of inertia, change which is sometimes even smaller than the change in ε.

4.5. Corrective action

Once the problem of ill-conditioning has been diagnosed, it is important to reduce its influence. The first idea is to include the moments of inertia of additional isotopologues. But the number of isotopic substitutions in a molecule is limited and, furthermore, these new data will possess near dependencies similar to those of the original data. Thus it is preferable to use additional information derived in some other way (electron diffraction, *ab initio* calculations, etc.). The different (and most useful) possibilities are described in section 5.

Unfortunately, it is not always possible to obtain new data which will substantially reduce the correlations. In such a case the introduction of appropriate "prior" information is the best solution. In the case of structure determination, the mixed estimation [75, 77] is often the most useful. Auxiliary information is added directly to the vector of observations and the Jacobian matrix. It may be a set of reasonable values of parameters based on similar molecules or from *ab initio* calculations (section 3) or from empirical correlations (section 9). For instance, if it is possible to find linear relations between the parameters which may be written in the form

$$\mathbf{c} = \mathbf{R}\beta + \eta \tag{43}$$

c is a vector of known values, **R** is a matrix of known constants and **η** is a random vector whose variance-covariance matrix is $\mathbf{V}(\boldsymbol{\eta}) = \boldsymbol{\Sigma}$. The estimation of Eq.(28) subject to the condition Eq.(43) proceeds by augmenting **y,** **X**, and **M** by **c**, **R**, and **Σ** respectively. The solution is then

$$\ddot{\boldsymbol{\beta}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{X} + \mathbf{R}^{\mathrm{T}}\Sigma^{-1}\mathbf{R}\right]^{1}\left[\mathbf{X}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{y} + \mathbf{R}^{\mathrm{T}}\Sigma^{-1}\mathbf{c}\right] \qquad (44)$$

When data of different origin are used, it is particularly important to check that the estimated weights are appropriate and that the data are compatible. Also, particular caution is indicated when one parameter appears to be determined by only one datum. This can be checked by an outlier analysis.

4.6. Outlier analysis

An outlier analysis which checks the compatibility of the data is particularly important. Several good books describe the different outlier diagnostics, they are summarized in Refs. [67, 75].

5. Experimental equilibrium structure [67, 78, 79]

5.1. General method

To determine an equilibrium structure, it is necessary to know the equilibrium rotational constants which are obtained from the ground state rotational constants and the rotational constants of all fundamentally excited vibrational states, see Eq.(7). In principle, microwave as well as infrared spectroscopies may be used to get the rotational constants B_v^{ξ} of the excited states. However, the intensity of a rotational transition is proportional to the population of its lower state which is given by the Boltzman law exp(–*E''*/*kT*). If the energy of the excited vibrational state is high, above 1000 - 2000 cm^{-1}, the population of its rotational levels will be small and the rotational transitions between them will be too weak to be observed. In this rather common case, only infrared spectroscopy may be used to determine the α constants.

It is thus obvious that the combination of microwave and infrared spectroscopies is well suited to obtain the equilibrium rotational constants. In theory, this is a simple task. However, unless the molecule has the formula XY_n and sufficiently high symmetry, the equilibrium rotational constants of only one isotopologue do not provide enough information

because the number of independent structural parameters is larger than the number of independent rotational constants (between one and three for one isotopologue). In this frequent case, the equilibrium rotational constants have to be determined for several isotopologues. Furthermore, it is often found that, in order to obtain an accurate structure (i.e. to avoid high correlation between the parameters), it is necessary to have the rotational constants of all singly substituted isotopologues. For a molecule with more than three atoms, this task requires a tremendous amount of data whose analysis is extremely time consuming. Moreover, this is often complicated by the fact that at least some excited states are not really isolated but are in resonance either by Coriolis interaction or anharmonic (Fermi, Darling-Dennison, etc.) resonance [80]. Unfortunately, this is frequently the case for a polyatomic molecule. It is then necessary to analyze the interactions between the excited states which is not easy even for small molecules. When there are only two interacting vibrational states, it is still not too complicated. If E_- is the energy of the lower interacting level and E_+ is that of the upper level, the rotational constants with the interaction taken into account may be written as

$$B_{\pm}^{\mathrm{eff}} = B_{\pm} \pm \frac{\left|\left\langle + \left| \mathbf{H}_{\pm} \right| - \right\rangle^2\right|}{E_+ - E_-}. \qquad (45)$$

Therefore, $B_+^{\mathrm{eff}} + B_-^{\mathrm{eff}} = B_+ + B_-$ and the interaction does not need to be treated accurately. For instance, there is a Fermi resonance between the $v_3 = 1$ and $v_2 = 2^0$ levels of many linear XYZ molecules. To cancel the effect of this resonance, the equilibrium rotational constants are usually calculated by $B_{\mathrm{e}} = \frac{1}{2}\left[5B_{000} - B_{100} - B_{001} - B_{02^20}\right]$, which makes use of Eq. (45) [81].

Unfortunately, this simplification is no more valid when more than two states interact. For instance, in $^{13}CH_3F$, when the states $v_3 = 1$ and $v_6 = 1$ on one side and $v_2 = 1$ and $v_5 = 1$ on the other side are analyzed as two independent dyads, the equilibrium rotational constant is $B_{\mathrm{e}} = 25280.2$ MHz. This assumption may seem justified because these two dyads are rather far apart: $\nu_2 - \nu_6 = 276.7$ cm^{-1}. However, if the four states are analyzed as a single interacting tetrad, $B_{\mathrm{e}} = 25099.8$ MHz [82]. This discrepancy, as first sight surprising, seems to be rather general and was noted a long time ago by Saito [83] in the analysis of the excited states of SO_2.

For these reasons, the number of accurate equilibrium structures for polyatomic molecules is rather small and it is, up to now, mainly limited to molecules with three independent structural parameters (or less) i.e. to molecules with five atoms or less.

5.2. Diatomic molecules [84, 85]

The case of the diatomic molecules is worth a separate treatment. As they are much simpler, a more sophisticated theory may be used and the bond length is determinable with much higher accuracy, which permits to see the breakdown of the BO approximation, among other facts.

The vibrational potential energy $V(r)$ is usually approximated by a power series using the expansion parameter

$$\xi = \frac{r - r_e}{r_e} \tag{46}$$

$$V(r) = h a_0 \xi^2 \left(1 + \sum_{i=1}^{\infty} a_i \xi^i\right) \tag{47}$$

where r_e is the equilibrium value of the internuclear distance, a_i are Dunham potential coefficients and h is Planck's constant. Similarly to the potential function, the rovibrational energy of the molecule in a vibrational state v and a rotational state J may be written following Dunham [86] as

$$\frac{1}{h} E(v, J) = \sum_{l,k} Y_{lk} \left(v + \frac{1}{2}\right)^l J^k (J+1)^k \tag{48}$$

The first Dunham coefficient Y_{01} is approximately equal to the equilibrium rotational constant B_e, the exact relation being

$$Y_{01} = B_e + \Delta Y_{01}^{(D)} = \frac{\hbar}{4\pi\mu_e r_e^2} + \Delta Y_{01}^{(D)} \tag{49}$$

where $\mu_e = m_A m_B / (m_A + m_B)$ is the reduced nuclear mass of the molecule AB, m_A and m_B being the nuclear masses of atoms A and B, respectively, and $\Delta Y_{01}^{(D)}$ is a small correction called Dunham correction

$$\Delta Y_{01}^{(D)} = \frac{B_e^3}{2\omega_e^2}\left(15 + 14a_1 - 9a_2 + 15a_3 - 23a_1a_2 + \frac{21}{2}\left[a_1^2 + a_1^3\right]\right) \tag{50}$$

where $\omega_e = 2\sqrt{a_0 B_e}$ is the harmonic vibrational frequency.

Within the Born-Oppenheimer approximation all the isotopologues of a molecule have the same molecular potential which results in a single bond distance. Actually, the bond distance is found to be slightly dependent of the isotopologic masses, see Table 14. The correction to the Born-Oppenheimer approximation yields a slight modification to the molecular potential which becomes dependent upon the nuclear masses. Y_{01} is now expressed as

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{ad})^2} + \Delta Y_{01}^{(D)}, \tag{51}$$

where r_e^{ad} depends on the nuclear masses. Watson [84] has shown that

$$r_e^{ad} = r_e^{BO}\left[1 + m_e\left(\frac{d_A^{ad}}{m_A} + \frac{d_B^{ad}}{m_B}\right)\right] \tag{52}$$

where r_e^{BO} is the BO interatomic distance, m_e is the mass of the electron, and d_A^{ad} and d_B^{ad} are two isotopic independent parameters.

A second correction, called non-adiabatic, has to be taken into account. It is due to the fact that the electronic ground state is slightly perturbed by the excited states (see section 2.3). This correction can be calculated with the molecular *g*-factor, and using now the atomic masses $M_X = m_X + Z_X m_e$ (X = A, B), Y_{01} may is equal to

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{ad})^2}\left[1 + \frac{\Delta Y_{01}^{(D)}}{B_e} + \frac{m_e}{M_p}g\right] \tag{53}$$

where M_p is the proton mass and $\mu = M_A M_B / (M_A + M_B)$ the reduced atomic mass. Finally, using Eq. (52), Y_{01} may be written in a more compact form as

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{BO})^2}\left[1 + m_e\left(\frac{\Delta_{01}^A}{M_A} + \frac{\Delta_{01}^B}{M_B}\right)\right] \tag{54}$$

where Δ_{01}^A and Δ_{01}^B are two isotopic independent parameters which are the sum of three terms: non-adiabatic, adiabatic, and higher order Dunham. Tiemann et al. [87] have studied in detail the order of magnitude of the different terms. It was found that the higher-order Dunham correction is almost negligible and that the non-adiabatic correction is the most important one. The adiabatic correction does not change very much for one specific atom by varying the chemical counterpart, and in general it is less than 30% of the total correction for the molecules studied.

In addition to the mass variation by isotopic substitution also the nuclear size will vary slightly giving rise to small changes in the Coulomb interaction between the electrons and the nuclei [88]. This isotope effect, called field shift in the theory of atomic spectra, slightly modifies Eq. (54) where the mean square nuclear charge radius $<r^2>_{A,B}$ is used as an expansion parameter and the new molecular parameter $V_{01}^{A,B}$ is introduced

$$Y_{01} = \frac{\hbar}{4\pi\mu(r_e^{BO})^2}\left[1 + m_e\left(\frac{\Delta_{01}^A}{M_A} + \frac{\Delta_{01}^B}{M_B}\right) + V_{01}^A\left\langle r^2\right\rangle_A + V_{01}^B\left\langle r^2\right\rangle_B\right]. \tag{55}$$

The parameter $V_{01}^{A,B}$ depends mainly on the electron density and its derivatives with respect to the internuclear distance at nucleus A or B. This finite nuclear size correction is only

significant when the accuracy of the measurements is extremely high and when the mass number of the atom is not too small (> 40).

5.3. Accuracy

The structure of diatomic molecules can be determined with a tremendous accuracy which is often better than 10^{-4} pm. When the structure of a polyatomic molecule can be obtained using the moments of inertia of only one isotopologue, the accuracy remains high. This is the case, for instance, for triatomic XY_2 molecules and tetraatomic C_{3v} molecules. SO_2 belongs to the first category, its equilibrium structure is $r_e(SO) = 143.078(2)$ pm and $\angle(OSO) = 119.330(3)°$ [89]. These results are in perfect agreement with a previous (and independent) determination [90] and with high level *ab initio* computations [91] (note, however, that the experimental values are likely to be more accurate). Another interesting example is linear CO_2 whose equilibrium structure could be determined in different ways from the equlibrium rotational constants of fourteen isotopologues giving: $r_e(CO) = 115.995884$ pm with a range of $75 \cdot 10^{-6}$ pm [92]. Two typical examples of the second category are phosphine and stibine where the equilibrium structure could be independently determined using the H_3 and D_3 isotopologues. Furthermore, for both molecules, the anharmonic force field was calculated in order to check the accuracy of the α constants. The results are given in Table 15 together with the corresponding *ab initio* structures. From inspection of data in Table 15, three conclusion can be drawn: i) the standard deviation of the parameters is much smaller than their range, thus indicating the presence of systematic errors, ii) their accuracy is nevertheless high, about 0.01 pm, and iii) the *ab initio* structure is likely to be less accurate.

When the moments of inertia of more than one isotopologue are needed, the accuracy usually drops to a few tenths of pm because the system of equations is no more well conditioned. There is, however, one important exception: linear triatomic molecules XYZ. The vibrational interactions of several XYZ molecules have been analyzed in great detail permitting the determination of very accurate equilibrium rotational constants for several isotopologues. Thus, the accuracy of the resulting equilibrium structures is high. A typical example is the structure of nitrous oxide, N_2O, [93] : $r_e(NN) + r_e(NO) = 231.2384(4)$ pm and $r_e(NN) - r_e(NO) = 5.779(7)$ pm. Another classic example is the structure of OCS [94] : $r_e(OC) + r_e(CS) = 271.756(3)$ pm and $r_e(OC) - r_e(CS) = 40.52(3)$ pm.

6. Semi-experimental equilibrium structure [37, 95, 96, 97, 98, 99, 100]

6.1. Method

In order to obtain equilibrium structures more easily, it was proposed quite early to calculate the vibration-rotation interaction constants from the force field [101]. This avoids the difficult problem of analyzing the resonances. Furthermore, it allows us to obtain the α constants of rare isotopologues without difficulty. Finally, it is quite useful when a vibrational state cannot be experimentally analyzed. However, it is extremely difficult to determine a cubic force field experimentally because the number of constants increases rapidly with the number of atoms and the loss of symmetry of the molecule.

Fortunately, it is now possible to calculate *ab initio* anharmonic force fields [93]. The molecular geometry is first calculated. Then, the associated harmonic force field is evaluated at the same level of theory (preferably analytically). (Note, however, that it may be advantageous to sue a better reference structure, see e.g. Ref. [102]). The cubic force constants can be most easily determined with the use of a finite difference procedure involving displacements along reduced normal coordinates [103]. This procedure is implemented in several computer packages [104]. Finally, the equilibrium rotational constants are derived from the experimental effective ground-state rotational constants and an *ab initio* cubic force field.

The high-level *ab initio* computation of a cubic force field may be time consuming. Fortunately, it has been found that the MP2 method with a basis set of triple zeta quality often gives results which are sufficiently accurate, especially for the cubic part of the force field [105]. This is partly due to the fact that the vibrational correction is only a small percentage of the rotational constants to be corrected and partly to the fact that the underlying reference structure is accurate. Thus, it needs not to be known accurately, at least in the more favorable cases. Indeed, a number of accurate equilibrium structures has been determined in this way, even for molecules as large as glycine [106] and proline [107].

However, the use of this method requires some experience. In particular, it is well advised to check that the derived force field is accurate enough, which may not always be the case. For instance, the semi-experimental structure of $H_2C{=}C{=}C$: was first determined using the CEPA-1 method giving r(C=C:) = 129.1(1) pm. It was later redetermined with the CCSD(T) method giving a bond length significantly shorter: r(C=C:) = 128.7 pm [108]. In this case, it is the choice of the treatment of electron correlation which is important but it may

also happen that the choice of the basis set becomes important. For instance, in acetylene, the harmonic bending frequency ω_4 and the corresponding anharmonicity $\omega_4 - \nu_4$ are both found to be extremely sensitive to the basis set used, in particular to the presence of a sufficient complement of diffuse functions [109]. Similar phenomena are observed more generally in bending modes for molecules that possess carbon-carbon multiple bonds [110]. Furthermore, the available experimental ground state rotational constants do not always permit to obtain a well-conditioned system of equations. In this rather frequent case, the determined parameters are not precise and may even be unreliable as will be seen in the next section.

Experimental parameters which may be used to check the quality of the force field are listed in Table 16. Examples of applications include the determination of the structure of the symmetric top $SiHF_3$ [111] or the asymmetric top FPO [112].

6.2. Comparison of the different equilibrium structures

A few structures are compared in Table 17. One conclusion is obvious: the standard deviations of the parameters (when determined) of the experimental and semi-experimental equilibrium structures are much too optimistic. This can be explained by two facts

- the number of available data is generally not much larger than the number of parameters to be determined; thus, the derived standard deviations are not reliable
- the assumptions of the least squares method are not obeyed (see section 4.2).

The experimental equilibrium structure is generally less accurate. However, for the r(SiCl) bond in SiH_3Cl, it is obviously the *ab initio* value which is not accurate. This is probably because it is difficult to calculate *ab initio* accurate interatomic distances between second-row atoms. This is confirmed by the result on the r(CCl) bond in O=CHCl. Finally, it is important to note that the semi-experimental equilibrium structure is perhaps not always the most accurate: in HOF, BHFOH, and BF_2OH the *ab initio* structures might be more accurate

7. Empirical structures [66, 67, 78, 113]

As it is extremely difficult to determine the equilibrium structure of a polyatomic molecule, particularly if there are many degrees of freedom, several empirical methods have been developed. Most of them only use the ground state moments of inertia. The simplest one gives the effective structure.

7.1. Effective structure (r_0)

The vibrational correction $\varepsilon = I_0 - I_e$ is quite small compared to I_0, less than 1% in most cases, see Table 18. Thus, it may seem to be a good approximation to neglect its contribution, the structural parameters being directly fitted to the ground state moments of inertia I_0 of a sufficiently large set of isotopologues. This method is simple and widely used. For diatomic molecules, its accuracy is about 1%. The accuracy remains about the same for polyatomic molecules, provided that the system of normal equations is well conditioned. Otherwise, the resulting r_0 structure may be widely different from the equilibrium structure, see for instance Table 19 which compares different structures of vinyl fluoride (H_2C=CHF). There are three different r_0 structures for which the angle $\angle(CCH_g)$ varies between 123.7° and 129.2°.

Furthermore, when hydrogen atoms are present, the determined bond lengths are affected by an error of 0.5 pm or more. Finally, an analysis of the residuals of the fit shows that they are highly autocorrelated instead of being random, see figure 1. It indicates that the model is not correct and that the standard deviations of the parameters are not reliable.

7.2. Substitution structure (r_s)

A better hypothesis than the neglect of ε is to assume that ε remains constant upon isotopic substitution, see Table 18. It is the implicit basis of Costain's [113] early suggestion to calculate the Cartesian coordinates of the substituted atom in the principal axis system by Kraitchman's equations [114](and their special forms for molecular symmetries [115, 116]), essentially from the difference of the inertial moments of the parent and the substituted isotopologue, the only two species involved.

The assumption ε = constant is obviously not valid in three cases:

- when an hydrogen atom is substituted by deuterium because the change of mass (and of ε) is large,
- when there are large axis rotation upon isotopic substitution,
- for most molecular complexes because the variation of ε upon substitution is large.

In other cases, the resulting r_s structure is believed to have a high degree of validity for heavy atoms. Unfortunately, even this is not always true. In the particular case of a linear molecule, the Cartesian coordinate z of the substituted atom may be written, using Kraitchman equation, as

$$z^2 = \frac{I_0' - I_0}{\mu} = \frac{\Delta I_e + \Delta\varepsilon}{\mu} \qquad (56)$$

with $\mu = \frac{M\Delta m}{M + \Delta m}$. (57)

This equation permits to estimate the uncertainty δz on z,

$$\delta z = \frac{1}{2}\frac{\Delta\varepsilon}{\mu}\frac{1}{|z|} \quad . \tag{58}$$

Costain [117] assumed that $\Delta\varepsilon/\mu$ is approximately constant and that the experimental error on the moments of inertia is negligible. Using the data for N_2O, he proposed the following empirical rule to estimate the error of a r_s coordinate (in pm)

$$\sigma(z) = \frac{12}{|z|} \tag{59}$$

Groner [78] discussed the usefulness of this relation. It cannot be taken for granted that Eq. (59) determines the true uncertainty of the coordinates. It is mainly useful to check the consistency of r_s coordinates. Actually, as $\mu \approx \Delta m$, i;e. about 1 or 2, Eq. (58) shows

- the uncertainty δz increases when z decreases; in other words, small coordinates are inaccurate, a frequent case in large molecules,
- the uncertainty is large when $\Delta\varepsilon$ is large, i.e. when hydrogen is substituted by deuterium or when there is a large axis rotation. These problems, already mentioned, are well known. But what is less known is that $\Delta\varepsilon$ roughly increases with the mass of the molecule (see figure 2), that is to say it is difficult to obtain an accurate substitution structure for a large molecule [118].

An illuminating example is the structure of C_5O: although the equilibrium Cartesian coordinate of the C_2 atom is as large as 492.2 pm, the error is $z_e - z_s = 2.56$ pm which is two orders of magnitude larger than predicted by Eq. (59).

To improve the accuracy of the substitution structure, a double substitution method (where two atoms are substituted at the same time) has been proposed [119] but it is difficult to apply. Futhermore, it was shown that the double substitution method is in general not much better than the single substitution method [120,121,122].

The r_0 method has been expanded to the $r_{I\varepsilon}$ method [123] where the experimental moments of inertia $I_0^{\xi}(i)$ of a sufficiently large set of isotopologues i are fitted to the structural parameters and to three constant, mass-independent parameters ε^{ξ} which hence do not depend on the isotopologue: $I_0^{\xi}(i) = I_{\text{rigid}}^{\xi}(i) + \varepsilon^{\xi}$, the $I_{\text{rigid}}^{\xi}(i)$ alone being functions of the structure. The method is a true r_0 derivative: the independent structure parameters must all either be free to be fitted or kept at fixed values, and the center- and product-of-mass conditions are fulfilled

automatically. A similarity to the substitution method exists in so far as the difference between respective moments of the parent and any isotopologue no longer contains any rovibrational contribution. The method has meanwhile been replaced by the methods with mass-dependent rovibrational contributions, as described in next section.

8. Mass-dependent structures

8.1. Method

To improve the substitution method, the variation of ε with the moment of inertia I_0 was analyzed. It was found that ε approximately varies as $\sqrt{I_0}$ [124]. This is particularly true in the case of isotopic substitution as the analysis of the residuals of the r_0 and r_s structures shows. The ground state moment of inertia I_0 may be approximately written as

$$I_0^{\xi} = I_m^{\xi} + c_{\xi}\sqrt{I_m^{\xi}} \tag{60}$$

where I_m^{ξ} is an approximation of I_e^{ξ}. The structural parameters and the proportionality constants c_{ξ} (one for each principal axis) are determined by least squares fitting. The resulting structure was called $r_m^{(1)}$ by Watson et al. [125], "m" because it is mass-dependent (ε is a homogeneous function of degree 1/2 in the masses) and "1" because one extra parameter per axis is used. It is important to note that the number of fitted parameters is the same as for the $r_{I\varepsilon}$ method. The $r_m^{(1)}$ method significantly increases the accuracy of the derived structure, at least in most cases. However, it was found to give poor results when there is a small Cartesian coordinate. An analysis of several linear triatomic XYZ molecules, where the coordinate of Y is small showed that ε varies as $\frac{m_X m_Z}{M}$, where m_X and m_Z are the masses of atoms X and Z, respectively, see figure 3 and ref. [122].

The method was generalized by Watson et al. [125] who proposed the following expression for I_0

$$I_0^{\xi} = I_m^{\xi} + c_{\xi}\sqrt{I_m^{\xi}} + d_{\xi}\left[\frac{m_1 m_2 \cdots m_N}{M}\right]^{1/(2N-2)} \tag{61}$$

where N is the number of atoms of the molecule and c_{ξ} and d_{ξ} are fitting parameters (one on each axis). Using this equation in the fitting procedure gives the $r_m^{(2)}$ structure. Excellent agreement between the r_e and $r_m^{(2)}$ structure was found for many small molecules, even when a small coordinate (as in N_2O) is present.

Watson et al. [125] have extended the r_m method also to cases where large principal inertial axis rotations with respect to the parent molecule occur in individual species of the set of isotopologues. Such cases are mainly limited to near oblate molecules and require a more complicated formulation.

8.2. Laurie correction [126]

In almost all structures investigated, hydrogen atoms, if present, have caused anomalies. This is usually solved by using the "Laurie" correction [127], which assumes that the effective r(X-D) and r(X-H) bond lengths are different by a small quantity δr_D. In practice, a shrinkage of the X-H bond length upon deuteration is expected. This agrees with the idea of the heavier deuterium atom lying lower in the asymmetric potential well of the X-H vibrational mode than the hydrogen atom and, hence, nearer to atom X. It is worth pointing out that the ensuing bond length change has often been shown to remain restricted to the X-H bond length considered, hardly affecting the rest of the parameters, at least when the least-squares system is not ill-conditioned.

The correction method presented by Watson et al. assumes that the apparent lengths of both bonds, X-H as well as X-D, consist of a common major part r_m, which is identical for both bonds (and close to the r_e value) plus an additional bond elongation due to the H or D atom asymmetrically vibrating against the remainder of the molecule. This elongation is taken to be proportional to the respective vibration frequency, i.e. proportional to the reciprocal square root of the reduced mass m_{red} of the vibrator. The additional part of the bond length is hence different for the X-H and the X-D bond,

$$r_H = r_m + \delta\mu_H \tag{62a}$$

$$r_D = r_m + \delta\mu_D, \tag{62b}$$

with r_H and r_D being the apparent bond length values resulting from the least squares fit, δ is a common proportionality factor and $\mu = m_{red}$ is different for the H and D containing species:

$$\mu_{H/D} = \sqrt{\frac{M}{m_{H/D}(M - m_{H/D})}} \tag{63}$$

where M is the total mass of the isotopologue containing either H or D, m_H is the mass of hydrogen and m_D the mass of deuterium. In principle, r_H and r_D can be determined by a least squares fit and r_m can then be deduced. However, in most cases, the system of normal equations is very ill-conditioned and the derived parameters are meaningless (an exception is when the rotational constant of the H-, D-, and T-isotopologues are available). Therefore, a

constraint has to be introduced to alleviate the problem. Either $r_H - r_D$ is constrained to a value known from experience ("Laurie" correction, near 0.3 pm) or r_m is constrained to a r_e value known from other sources (e.g. *ab initio* calculations or empirical correlation, see sections 3 and 9).

8.3. Use of the mixed regression method

Besides using the *ab initio* value for r_m(XH), it may be useful to use other *ab initio* parameters as predicate observations in a mixed regression, see section 4.5. This, indeed, improves the conditioning and allows the determination of parameters with a better accuracy. This is furthermore a good way to check the accuracy of the *ab initio* structure. However, it is not always possible to calculate a reliable *ab initio* structure. Nevertheless, there is another way to improve the conditioning. Inspection of the constants c_ξ and d_ξ (when they are accurately determined) shows that they vary smoothly and slowly for a given series of molecules, see Table 20. Hence, it should be possible to use the c_ξ and d_ξ of a similar molecule in a mixed regression. For instance, it is easy to obtain an accurate $r_m^{(2)}$ structure for ClCN and BrCN but it fails in the case of FCN and ICN, in particular because of the lack of isotopic substitution for F or I. However, using the c and d values of ClCN/BrCN for FCN/ICN as predicate observations (with an accuracy of about 10%) considerably improves the conditioning and gives structures in good agreement with the r_e structure. However, these results are still preliminary and more work is needed to appreciate the validity of this method.

8.4. Accuracy

At each step of the improvement of the model ($r_0 \rightarrow r_m^{(1)} \rightarrow r_m^{(2)}$), the number of parameters increases. Thus, the conditioning of the system is expected to deteriorate. This is important because the r_m model is still an approximate one, see figure 4 which shows that the residuals are still not random. If the rovibrational contribution is roughly estimated as 0.5% of the inertial moment and its model error as 10% of its value (as derived from the standard deviations of the parameters c_ξ and d_ξ), then the model error is 0.05% of the inertial moment which is more, by orders of magnitude, than the experimental error in modern spectroscopy. This model error, with repercutions on the accuracy of the parameters, is often "amplified" by the ill-conditioning, see section 4.4.

In conclusion, the r_m method is well suited when the number of available rotational constants is large compared to the number of structural parameters, a condition which is

usually fulfilled only for small molecules. Nevertheless, the structure of several molecules (not all of them small) was recently determined using the r_m method and these experimental structures were found to be in good agreement with the corresponding *ab initio* structures (i.e. within 0.2-0.3 pm), see Table 21.

9. Empirical correlations

There are many correlations between structural parameters and other parameters [128]. Although most of them are difficult to use, there are a few relations which may be of great help for a structure determination.

The stretching force constants are related to the bond strength and one can expect a good correlation between the length of a bond and the corresponding diagonal stretching force constant, as shown for instance by Badger [129,130]. In principle, such a correlation might be used to determine bond lengths, but it is often more difficult to accurately determine force constants than bond lengths because the number of force constants, $(1/2)n_{vib}(n_{vib} + 1)$, is usually larger than the number of fundamental vibrations, n_{vib}, and additional information (vibrational frequencies of isotopic species, …) is required to determine the force constants. There are, however, a few cases in which the bond length vs. diagonal force constant correlation is relevant. One is when the molecule is simple enough and this correlation has been for instance used to determine the Au-Au bond length in several molecules [131].

More generally, when a vibrational mode r has characteristic frequency far from the others, r', it may be considered as isolated. In other words, the non-diagonal force constants $f_{rr'}$ are then negligible compared to the corresponding energy differences E_r - $E_{r'}$. Thus, in this particular case, a relationship between the bond length and the corresponding stretching vibrational frequency is to be expected. Such a relationship was first pointed out by Bernstein [132] and considerably developed by McKean [133] for CH bonds.

McKean used selective deuteriation, all CH bonds but the relevant CH one being deuteriated. One therefore decouples the CH group of interest from the rest of the molecule in order to make sure that the CH stretching frequency is not affected. The main difficulty is often to synthetize a molecule where all hydrogens but one have been replaced by deuterium. Furthermore, in a few cases the stretch vibration is perturbed by some resonance and, except if a detailed vibration-rotation analysis was carried out to provide a deperturbed stretching frequency, another method is required.

Another method for determining isolated XH stretching frequencies relies on data from overtone bands, i.e. resulting from the multiexcitation of a single vibrational mode. It is well known that the motions of the individual bonds become increasingly localized upon increasing excitation and thus decreasingly affected by internal couplings [134]. The band origin of as many overtones of the XH stretch as possible needs to be measured and inserted into a so-called Birge-Sponer plot [135] :

$$\nu = v\tilde{\omega} - v(v+1)\tilde{\omega}x \qquad (64)$$

where $\tilde{\omega}$ is the mechanical frequency and $\tilde{\omega}x$ the first anharmonic correction term. This procedure thus also allows the frequency of the "unperturbed" fundamental band, ν_{is}(XH), to be determined :

$$\nu_{is} = \tilde{\omega} - 2\tilde{\omega}x \qquad (65)$$

The major advantage of this second method is to avoid isotopic labeling. However, it does not always avoid the problem of resonance.

The accuracy of the correlation corresponds to a bond length change of about 0.1 pm for a shift in the isolated CH stretching fundamental frequency of 10 cm^{-1} [136], *r*(C-H) bond lengths can thus be determined with a precision of about 0.2 pm, from this correlation. This method was extended by McKean to SiH and GeH bond lengths [137]. Such a linear relationship was also shown to exist between *r*(OH) and ν(OH) [138] and between *r*(NH) and ν(NH) [139]. A non-linear relationship between *r*(N=O) and ν(N=O) was also pointed out [140].

There are other correlations of a bond length with other properties (electronegativity or another bond length) [141]. They are much less accurate but they may be useful to check whether the value of a bond length is correct or not.

10. Special molecules

10.1. Nonrigid molecules

Up to the present we have assumed that the molecule is semi-rigid, i.e. that the vibrational amplitudes are small. In principle, there is no difficulty to calculate the *ab initio* structure of a nonrigid molecule and this was done for many molecules. On the other hand, when there is a large amplitude motion the Taylor series expansion of the potential is no more valid and the calculated force constants (and α constants) may be not reliable. Thus, experimental and semi-experimental equilibrium structures might be doubtful. There are, up to this date, very few results. However, in the case of internal rotation of a symmetric rotor (a methyl group for instance) and when the potential is high enough, it seems that it is still possible to calculate the force constants using the assumption of small amplitude vibrations and, hence, a semi-experimental structure. One possible explanation of this nice behaviour is that the internal rotation potential is almost harmonic (at the bottom of the well). On the other hand, when the large amplitude motion is highly anharmonic, as for the inversion motion in formamide for instance, the results are very disappointing. For instance, the α constants of formamide have been recently calculated *ab initio* at the MP2/cc-pVTZ level of theory [59]. Although the v_{12} = 1 mode (the large-amplitude NH_2 inversion motion at 288.7 cm^{-1}) is well isolated, the calculated α_{12}^{A} = 2039 MHz and α_{12}^{C} = 28.9 MHz constants are in poor agreement with the corresponding experimental values, 978 MHz and –0.0099 MHz, respectively. However, it would be dangerous to draw any conclusion because the level of theory used is not high enough and the CCSD(T) method is likely to be required. Furthermore, it was possible to determine a $r_m^{(2)}$ structure of formamide which is in very good agreement with the *ab initio* structure [65].

There is another difficulty which may be encountered with non-rigid molecules. For instance, several molecules containing the CONH linkage seem to have a pyramidalized nitrogen at equilibrium and a double-minimum inversion potential with a very small inversion barrier allowing for an effectively planar ground-state structure [65]. Thus, the comparison of structures of different meaning may be meaningless.

10.2.Weakly bound cluster molecules

Weakly bound cluster molecules are a particular case of nonrigid molecules. Their study is a large field of research because the forces that hold the clusters together are responsible for the interactions of the molecules in liquid and solids. They also explain the stability of some conformers of large molecules (intramolecular hydrogen bonds).

The structure of complexes can also be calculated *ab initio*. For hydrogen bonded and van der Waals complexes, the choice of the *ab initio* method is not a problem. On the other hand, the basis set convergence is extremely slow, see Table 8. Furthermore, it is necessary to use a basis set with diffuse functions and the effect of the basis set superposition error (BSSE) has to be corrected [142]. This BSSE is the consequence of the use of finite basis sets. When the distance between the two monomers decreases, their basis functions overlap. Each monomer "borrows" functions from the other monomer, effectively increasing its basis set and improving the calculation of derived properties (of the monomers only). The calculated interaction energy becomes too large and the potential energy hypersurface distorted. A simple solution is the use of extremely large basis sets, which is not practical in most cases. The conventional way to correct for BSSE is based on the Boys-Bernardi counterpoise (CP) scheme. The CP-corrected potential energy surface of the dimer AB is

$$E^{\mathrm{CP}}(\mathrm{AB}) = E^{\mathrm{AB}}(\mathrm{AB}) + [E^{\mathrm{A}}(\mathrm{A}) + E^{\mathrm{B}}(\mathrm{B}) - E^{\mathrm{AB}}(\mathrm{A}) - E^{\mathrm{AB}}(\mathrm{B})]$$

where the superscript denotes the basis set used. Thus, one has to calculate five different energies at each geometry.

To date, there are experimental or *ab initio* equilibrium structures available for this class of molecules, all the experimental structures being empirical. The applicability of the r_{m} method was tested on several small complexes and satisfactory results were obtained [143]. On the other hand, to determine a satisfactory $r_{\mathrm{m}}^{(1)}$ structure for the very large complex phenylacetylene-argon [144], it was necessary to introduce constraints. This is not surprising because the number of structural parameters, 26, is very large (although also the number of isotopologues is large: 24). It is also probable that the large amplitude motion of the argon atom complicates the situation.

10.3. Large molecules

When the number of structural parameters is large, most of the methods previously described fail although lower level *ab initio* calculations are still possible. In this case, the molecule may often exist in different conformations and the primary aim is to determine which conformers are present (i.e. estimate the dihedral angles and more, if possible) as well as their relative energies. Microwave spectroscopy in combination with electronic structure calculations is well suited for this purpose.

The ground state rotational constants, which are easy to determine, are the most useful parameters but it often happens that the rotational constants of different conformers are quite

similar. In such a (frequent) case, the ambiguity can be cleared up using one (or several) of the following pieces of information:

- the components of the dipole moment, which are easily estimated by microwave spectroscopy (from Stark effect or relative intensities) usually change during a rotation around a bond even if the total dipole moment remains almost constant,
- isotopic substitution of one atom not too close to the center of mass often results in conclusive differences of the rotational constants of different conformers,
- nuclear quadrupole coupling constants are also useful because different orientations of the quadrupole tensor result in different coupling constants.

A very attractive example is 1-octene for which fifteen conformational isomers (from 131 predicted) have been measured by molecular beam Fourier transform microwave spectroscopy. The identification used the rotational constants, inertial defect, and relative intensities [145]. It is interesting to note that the quality of the *ab initio* calculations (MP2/6-31G*) was not sufficient to provide the correct energy ordering of the conformers. This is a problem frequently encountered in this kind of study.

Another fascinating example is the microwave study of 2-chloroethyl ethyl sulfide ($CH_3CH_2SCH_2CH_2Cl$) [146], for which many conformers are possible. The two conformers identified were assigned by comparing the experimental rotational constants to the *ab initio* values. However, this did not permit an unambiguous identification. For the most stable conformer, the experimental rotational constants of the ^{37}Cl isotopologue were used with those of the ^{35}Cl species to calculate the Kraitchman Cartesian coordinates of the chlorine atom (see section 7.2). The comparison with the *ab initio* coordinates permitted to identify the conformer. This identification was corroborated by comparison of the experimental relative intensities with those predicted *ab initio* (which are function of the components of the dipole moment). For the second conformer, it was the comparison of the experimental and *ab initio* quadrupole coupling constants which permitted its identification.

Another typical example is the glycolaldehyde-water complex, for which seventeen conformers are possible. The most stable conformer was identified without ambiguity in the same way as for 1-octene [147]. Plusquellic et al. [148] have summarized the work recently done on peptides and dipeptides. More examples of conformational analyses are given, for example, in Ref. [149].

11. Conclusions

Experimental equilibrium structures can be highly accurate for diatomic molecules and C_{2v} or $C_{\infty v}$ triatomic molecules. The accuracy is generally much lower for larger molecules. It is mainly limited by two factors: i) ill-conditioning of the system of equations, and ii) anharmonic resonances. The harmful influence of the ill-conditioning may be reduced by increasing the diversity and accuracy of the data. A thorough analysis of the resonances is in principle possible as shown for example for linear triatomic molecules. However, it is difficult and time consuming but an ab initio calculation of the anharmonic foce field should make the analysis much easier.

During the last 20 years, tremendous progress has been made in the ab initio computation of the equilibrium structures which is now much faster and much easier than their determination by experimental methods. The accuracy of the computed structures is comparable or even better to that observed in most spectroscopic studies. For instance, the accuracy of the CCSD(T) method is 0.3 pm or better. However, this relatively small error is due to a cancellation of errors due to the truncation of the coupled-cluster hierarchy and to the truncation of the basis set. Although it is possible to improve this accuracy, at least for small molecules, this is not an easy task.

The semi-experimental method is also much easier to use than the experimental method. It is also generally more accurate. However, it also suffers from the problem of ill-conditioning which may be solved in the same way as for the experimental method. Another problem is that it is difficult to estimate the accuracy of the computed rovibrational corrections. For this reason, an experimental determination of deperturbed α constants for at least the main isotopologue is desirable. To improve the accuracy, it is possible to calculate the higher-order vibration-rotation interaction constants γ_{ij}^{ξ} from the quartic force field. However, although their contribution is not negligible, they do not affect much the resulting structure.

The mass-dependent r_{m} methods are also limited to small molecules and they are significantly less accurate than the other methods. Nevertheless, they are quite useful because they are very easy to use and their accuracy is sometimes sufficient.

Acknowledgments

I thank Prof. H.D. Rudolph for many helpful suggestions, Prof. A.G. Császár, Dr. C. Gutle and Prof. J. Liévin for critical reading of the manuscript and Profs. J.E. Boggs and M. Herman for their encouragements.

References

[1] Couper, A.S., 1858, *Annales de chimie et de physique,* Série 3, **53**, 488.

[2] Kekulé, F.A., 1865, *Bull. Soc. Chim.,* **3**, 98.

[3] von Bayer, A., 1885, *Ber. Dtsch. Chem. Ges.*, **18**, 2269.

[4] Le Bel, A.J., 1874, *Bull. Soc. Chim. France*, **22**, 337.

[5] van't Hoff, J.H., 1875, *La Chimie dans l'Espace* (Rotterdam,: P.M. Bazendijk).

[6] von Laue, M., 1912, *Münchener Ber.*, 303.

[7] Bragg, W.H., and Bragg, W.L., 1914, *Proc. Roy. Soc. London*, Ser. A **89**, 468.

[8] Debye, P., Bewilogua, L., Ehrhardt, P., 1929, *Phys. Z.*, **30**, 84.

[9] Mark, H., and Wierl, R., 1930, *Naturwissenschaften*, **18**, 205.

[10] Pauling, L., 1960, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (Ithaca, NY: Cornell University Press).

[11] Burrau, Ø., 1927, *K. Danske Viden. Selskab*, **7**, 14.

[12] Heitler, W., and London, F., 1927, *Z. Physik*, **44**, 453.

[13] Townes, C.H., and Schawlow, A.L., 1955, *Microwave Spectroscopy* (New York: McGraw-Hill).

[14] Hargittai, I., and Levy, J.B., 1999, *Struct. Chem.* **10**, 387.

[15] Burkert, U., and Allinger, N.L., 1982, *Molecular Mechanics* (Washington, D.C.: American Chemical Society).

[16] Bunker, P.R., and Jensen, P., 1998, *Molecular Symmetry and Spectroscopy* (Ottawa: NRC Research Press).

[17] Bunker, P.R., and Jensen, P., 2000, *Computational Molecular Spectroscopy*, edited by P. Jensen and P.R. Bunker (Chichester: Wiley), p. 3.

[18] Stanton, J.F., 2007, *J. Chem. Phys.*, **126**, 134309.

[19] Papousek, D., and Aliev, M.R., 1982, *Molecular Vibrational-Rotational Spectra* (Amsterdam: Elsevier).

[20] Wilson, E.B., Decius, J.C., and Cross, P.C., 1955, *Molecular Vibrations* (New York: McGraw Hill).

[21] Watson, J.K.G., 1970, *Molec. Phys.*, **19**, 465.

[22] Watson, J.K.G., 1977, *Vibrational Spectra and Structure*, Vol. 6, edited by J.R. Durig (Amsterdam: Elsevier), p. 1.

[23] Kivelson, A., and Wilson, E.B., 1952, *J. Chem. Phys.,* **20**, 1575.

[24] Dowling, J.M., 1961, *J. Molec. Spectrosc.,* **6**, 550.

[25] Hedberg, L., and Mills, I.M., 1993, *J. Molec. Spectrosc.,* **160**, 117.

[26] Chung, K.T., and Parker, P.M., 1963, *J. Chem. Phys.,* **38**, 8.

[27] Aliev, M.R., and Watson, J.K.G., 1985, *Molecular Spectroscopy: Modern Research*, Vol. 3, edited by K.N. Rao (New York: Academic Press), p. 1.

[28] Gordy, W., and Cook, R.L., 1984, *Microwave Molecular Spectra*, (New York: Interscience).

[29] Flygare, W.H., 1974, *Chem. Rev.* **74**, 655.

[30] Gauss, J., Ruud, K., Helgaker, T., 1996, *J. Chem. Phys.,* **105**, 2804.

[31] Hehre, W.J., Radom, L., Schleyer, P.v.R., and Pople, 1986, J.A., *Ab Initio Molecular Orbital Theory* (New York: Wiley).

[32] Helgaker, T., Jørgensen, P., and Olsen, J., 2000, *Molecular Electronic Structure Theory* (New York: Wiley).

[33] Bartlett, R.J., and Stanton, J.F., 1994, *Rev. Comp. Chem.,* Vol. V, edited by K.B. Lipkowitz and D.B. Boyd (New Yok: VCH), p. 65.

[34] Helgaker, T., Ruden, T.A., Jørgensen, P., Olsen, J., and Klopper, W., 2004, *J. Phys. Org. Chem.* **17**, 913.

[35] Møller, C., and Plesset, M. S., 1934, *Phys. Rev*., **46**, 618.

[36] Purvis III, G. D., and Bartlett, R. J., 1982, *J. Chem. Phys*., **76**, 1910.

[37] Lee, T. J., and Scuseria, G. E., 1995, *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*, edited by S. R. Langhoff (Dordrecht: Kluwer), pp. 47.

[38] Janssen, C.L., and Nielsen, I.M.B., 1998, *Chem. Phys. Lett.*, **290**, 423.

[39] Lee, T.J., and Taylor, P.R., 1989, *Int. J. Quant. Chem. Symp.*, **23**, 199.

[40] Helgaker, T., Gauss, J., Jørgensen, P., and Olsen, J., 1997, *J. Chem. Phys.,* **106**, 6430.

[41] Halkier, A., Jørgensen, P., Gauss, J., and Helgaker, T., 1997, *Chem. Phys. Lett.,* **274**, 235.

[42] Feller, D., Peterson, K.A., Crawford, T.D., 2006, *J. Chem. Phys.* **124**, 054107.

[43] Ruden, T.A., Helgaker, T., Jørgensen, P., and J. Olsen, *J. Chem. Phys.* **121**, 5874.

[44] Dunning Jr., T. H., 1989, *J. Chem. Phys*., **90**, 1007.

[45] Woon, D.E., and Dunning Jr., T.H., 1995, *J. Chem. Phys*., **103**, 4572.

[46] Peterson, K.A., and Dunning Jr., T.H., 2002, *J. Chem. Phys*., **117**, 10548.

[47] Kendall, R. A.; Dunning Jr., T. H.; Harrison, R.J., 1992, *J. Chem. Phys*., **96**, 6796.

[48] Martin, J.M.L., and Taylor, P.R., 1994, *Chem. Phys. Lett.*, **225**, 473.

[49] Császár, A. G., Allen, W. D., 1996, *J. Chem. Phys*. **104**, 2746.

[50] Margulès, L., Demaison, J., Rudolph, H.D., 2001, *J. Molec. Struct.* **599**, 23.

[51] Demaison, J., Boggs, J.E., and Rudolph, H.D., 2004, *J. Molec. Struct.* **695**, 145.

[52] Margulès, L., Demaison, J., and Boggs, J.E., 1999, *J. Phys. Chem. A*, **103**, 7632.

[53] Pyykkö, P., 1988, *Chem. Rev.* **88**, 563.

[54] Martin, J.M.L., Sundermann, A., 2001, *J. Chem. Phys*., **114**, 3408.

[55] Peterson, K.A., 2003, *J. Chem. Phys*., **119**, 11099.

[56] Douglas, M., and Kroll, N.M., 1974, *Ann. Phys. (N.Y.)*, **82**, 89.

[57] Hess, B.A., 1986, *Phys. Rev. A,* **33**, 3742.

[58] de Jong, W.A., Harrison, R.J., and Dixon, D.A., 2001, , *J. Chem. Phys*., **114**, 48.

[59] Handy, N.C., and Lee, A.M., 1996, *Chem. Phys. Lett.* **252**, 425.

[60] Gauss, J., Tajti, A., Kállay, M., Stanton, J.F., and Szalay, P.G., 2006, , *J. Chem. Phys*., **125**, 144111.

[61] Heckert, M., Kállay, M., Gauss, J., 2005, *Mol. Phys.* **103**, 2109.

[62] Heckert, M., Kállay, M., Tew, D.P., Klopper, W., Gauss, J., 2006, *J. Chem. Phys.* **125**, 044108.

[63] Klopper, W., Manby, F.R., Ten-No, S., Valeev, E.F., 2006, *Intern. Rev. Phys. Chem.* **25**, 427.

[64] Pawlowski, F., Halkier, A., Jørgensen, P., Bak, K.L., Helgaker, T., Klopper, W., 2003, *J. Chem. Phys.* **118**, 2539.

[65] Demaison, J., Császár, A.G., Kleiner, I., Møllendal, H., 2007, *J. Phys. Chem. A*, **111**, 2574.

[66] Rudolph, H.D., 1995, *Advances in Molecular Structure Research,* Vol. 1 edited by M. Hargittai and I. Hargittai (Greenwich, CT: JAI Press), p. 63.

[67] Demaison, J., Wlodarczak, G., and Rudolph, H.D., 1197, *Advances in Molecular Structure Research,* Vol. 3 edited by M. Hargittai and I. Hargittai (Greenwich, CT: JAI Press), p. 1.

[68] Albritton, D.L., Schmeltekopf, A.L., and Zare, R.N., 1976, *Molecular Spectroscopy: Modern Research*, Vol. II, edited by K.N. Rao (New York: Academic Press), p. 1.

[69] Stiefvater, O.L., and Rudolph, H.D., 1989, *Z. Naturforsch.,* **44a**, 95.

[70] Epple, K.J., and Rudolph, H.D., 1992, *J. Molec. Spectrosc.,* **152**, 355.

[71] Mayer, M., Grabow, J.-U., Dreizler, H., and Rudolph, H.D., 1992, *J. Molec. Spectrosc.,* **151**, 217.

[72] Demaison, J., Wlodarczak, G., Rück, H., Wiedenmann, K.H., and Rudolph, H.D., 1996, *J. Molec. Struct.,* **376**, 399.

[73] Hamilton, L.C., 1992, *Regression with Graphics* (Belmont, CA: Duxbury Press).

[74] Kuchitsu, K., Fukuyama, T., and Morino, Y., 1969, *J. Molec. Struct.,* **4**, 41.

[75] Belsley, D.A., 1991, *Conditioning Diagnostics* (New York: Wiley).

[76] Brodersen, S., 1990, *J. Molec. Spectrosc.,* **142**, 122.

[77] Bartell, L.S., Romanesko, D.J., and Wong, T.C., 1975, *Chemical Society Specialist Periodical Report N° 20: Molecular Structure by Diffraction Methods,* edited by G.A. Sim and L.E. Sutton (London: The Chemical Society), p. 72.

[78] Groner, P., 2000, *Vibrational Spectra and Structure*, Vol. 24, edited by J.R. Durig (Amsterdam: Elsevier), p. 165.

[79] Graner, G., 1992, *Accurate Molecular Structures*, edited by A. Domenicano and I. Hargittai (Oxford University Press), p. 65.

[80] Califano, S., 1976, *Vibrational States* (London: Wiley).

[81] Morino, Y., and Matsumura, C., 1967, *Bull. Chem. Soc. Japan,* **40**, 1095.

[82] Demaison, J., Breidung, J., Thiel, W., and Papousek, D., 1999, *Struct. Chem.* **10**, 129.

[83] Saito, S., 1969, *J. Molec. Spectrosc.,* **30**, 1.

[84] Watson, J.K.G., 1973, *J. Molec. Spectrosc.,* **45**, 99.

[85] Watson, J.K.G., 1980, *J. Molec. Spectrosc.,* **80**, 411.

[86] Dunham, J.L., 1932, *Phys. Rev.*, **41**, 721.

[87] Tiemann, E., Arnst, H., Stieda, W.U., Törring, T., and Hoeft, J., 1982, *Chem. Phys.*, **67**, 133.

[88] Tiemann, E., Knöckel, H., and Schlembach, J., 1982, *Ber. Bunsenges. Phys. Chem.* **86**, 821.

[89] Flaud, J.-M., and Lafferty, W., 1993, *J. Molec. Spectrosc.,* **161**, 396.

[90] Morino, Y., Tanimoto, M., and Saito, S., 1988, *Acta Chem. Scand. A*, **42**, 346.

[91] Martin, J.M.L., 1998, *J. Chem. Phys.*, **108**, 2791.

[92] Graner, G., Rossetti, C., and Bailly, D., 1986, *Molec. Phys.*, **58**, 627.

[93] Teffo, J.-L., and Chédin, A., 1989, *J. Molec. Spectrosc.,* **135**, 389.

[94] Lahaye, J.-G., Vandenhaute, R., and Fayt, A., 1987, *J. Molec. Spectrosc.,* **123**, 48.

[95] Clabo, D.A., Allen, W.D., Yamaguchi, Y., Remington, R.B., and Schaefer, H.F., 1988, *Chem. Phys.*, **123**, 187.

[96] Allen, W.D., Yamaguchi, Y., Császár, A.G., Clabo, D.A., Remington, R.B., and Schaefer, H.F., 1990, *Chem. Phys.*, **145**, 427.

[97] Botschwina, P., and Flügge, J., 1991, *Chem. Phys. Lett.,* **180**, 589.

[98] Bürger, H., and Thiel, W., 1997, *Vibration-Rotational Spectroscopy and Molecular Dynamics,* edited by D. Papousek (Singapore: World Scientific), p. 56.

[99] Császár, A.G., 1998, *Encyclopedia of Computational Chemistry*, edited by P. von Ragué Schleyer (Wiley: New York).

[100] Sarka, K., and Demaison, J., 2000, *Computational Molecular Spectroscopy*, edited by P. Jensen and P.R. Bunker (Chichester: Wiley), p. 255.

[101] Hoy, A.R., Mills, I.M., and Strey, G., 1972, *Molec. Phys.*, **24**, 1265.

[102] Allen, W.D., Császár, A.G., 1993, *J. Chem. Phys.* **98**, 2983.

[103] Schneider, W., and Thiel, W., 1989, *Chem. Phys. Lett.*, 157, 367.

[104] Stanton, J. F.; Lopreore, C.; Gauss, J., 1998, *J. Chem. Phys.* **108**, 7190.

[105] Puzzarini, C., and Cazzoli, G., 2006, *J. Molec. Spectrosc.*, **240**, 260.

[106] Kasalová, V., Allen, W.D., Schaefer III, H.F., Czinki, E., and Császár, A.G., 2007, *J. Comput. Chem.*, **28**, 1373.

[107] Allen, W.D., Czinki, E., and Császár, A.G., 2004, *Chem. Eur. J.*, **10**, 4512.

[108] Gauss, J., and Stanton, J.F., 1999, *J. Molec. Struct.*, **485-486**, 43.

[109] Martin, J.M.L., Lee, J.J., and Taylor, P.R., 1998, *J. Chem. Phys.*, **108**, 676.

[110] Simandiras, E.D., Rice, J.E., Lee, T.J., Amos, R.D., and Handy, N.C., 1988, *J. Chem. Phys.*, **88**, 3187.

[111] Demaison, J., Margulès, L., Breidung, J., Thiel, W., and Bürger, H., 1999, *Molec. Phys.*, **97**, 1053.

[112] Beckers, H., Bürger, H., Paplewski, P., Bogey, M., Demaison, J., Dréan, P., Walters, A., Breidung, J., and Thiel, W., 2001, *Phys. Chem. Chem. Phys.*, **3**, 4247.

[113] Costain, C.C., 1958, *J. Chem. Phys.*, **29**, 864.

[114] Kraitchman, J., 1953, *Amer. J. Phys.*, **21**, 17.

[115] Chutjian, A., 1964, *J. Mol. Spectrosc.,*14, 361

[116] Nygaard, L., 1976, *J. Mol. Spectrosc.*, **62**, 292

[117] Costain, C.C., 1966, *Trans. Amer. Cryst. Ass.*, **2**, 157.

[118] Demaison, J., and Rudolph, H.D., 2002, *J. Mol. Spectrosc.*, **215**, 78.

[119] Pierce, L., 1959, *J. Mol. Spectrosc.*, **3**, 575.

[120] Demaison, J., Wlodarczak, G., Burie, J., and Bürger, H., 1990, *J. Molec. Spectrosc.*, **140**, 322.

[121] Sarka, K, Papousek, D., Demaison, J., Mäder, H., and Harder, H., 1997, *Vibration-rotational spectroscopy and molecular dynamics*, edited by D. Papousek (World Scinetific: Singapore), p. 116.

[122] Le Guennec, M., Wlodarczak, G., Demaison, J., Bürger, H., Litz, M., and Willner, H., 1993, *J. Mol. Spectrosc.* **157**, 419.

[123] Rudolph, H. D., 1991, *Struct.Chem.,* **2**, 581.

[124] Demaison, J., and Nemes, L., 1979, *J. Molec. Struct.*, **55**, 295.

[125] Watson, J. K. G., Roytburg, A., and Ulrich, W., 1999, *J. Molec. Spectrosc.*, **196**, 102.

[126] Demaison, J., Margulès, L., Boggs, J.E., and Rudolph, H.D., 2001, *Struct. Chem.*, **12**, 1.

[127] Laurie, V.W., 1958, *J. Chem. Phys.*, **28**, 704.

[128] Mastryukov, V.S., and Simonsen, S.H., 1996, *Advances in Molecular Structure Research*, Vol. 2, edited by M. Hargittai and I. Hargittai (Greenwich, CT: JAI press), p. 163.

[129] Badger, R.M., 1934, *J. Chem. Phys,.* **2,** 128.

[130] Herschbach, D.R., and Laurie, V.W., 1961, *J. Chem. Phys,.* **35**, 458.

[131] Perreault, D., Drouin, M., Michel, A., Miskowski, V.M., Schaefer, W.P., and Harvey, P.D., 1992, *Inorg. Chem,.* **31,** 695.

[132] Bernstein, H.J., 1962, *Spectrochim. Acta,* **18,** 161.

[133] McKean, D.C., 1978, *Chem. Soc. Rev,.* **7,** 399.

[134] Henry, B.R., 1987, *Acc. Chem. Res,.* **20**, 429.

[135] Birge, R.T., and Sponer, H., 1926, *Phys. Rev,.* **28**, 259.

[136] Demaison, J., and Wlodarczak, G., 1994, *Struct. Chem.*, **5**, 57.

[137] McKean, D.C., 1981, *J. Molec. Struct,.* **113,** 251.

[138] Demaison, J., Herman, M., and Liévin, J., 2007, *Intern. Rev. Phys. Chem.*, **26**, 391.

[139] Demaison, J., Margulès, L., and Boggs, J.E., 2000, *Chem. Phys.*, **260**, 65.

[140] Turner, P.H., and Cox, A.P., 1978, *J. Chem. Soc. Faraday Trans.,* II **74,** 533.

[141] Demaison, J., Margulès, L., and Boggs, J.E., 2003, *Struct. Chem.*, **14**, 159.

[142] Boys, S.F., and Bernardi, F., 1970, *Mol. Phys.* **19**, 553.

[143] Kisiel, Z., 2003, *J. Mol. Spectrosc.*, **218**, 58.

[144] Dreizler, H., Hartke, B., Rudolph, H.D., 2006, *J. Molec. Struct.*, **825**, 1

[145] Fraser, G.T., Suenram, R.D., and Lugez, C.L., 2001, *J. Chem. Phys. A*, **105**, 9859.

[146] Tubergen, M.J., Lesarri, A., Suenram, R.D., Samuels, A.C., Jensen, J.O., Ellzy, M.W., and Lochner, J.M., 2005, *J. Molec. Spectrosc.*, **233**, 180.

[147] Aviles-Moreno, J.-R., Demaison, J., and Huet, T.R., 2006, *J. Am. Chem. Soc.*, **128**, 10467.

[148] Plusquellic, D.F., Kleiner, I., Demaison, J., Suenram, R.D., Lavrich, R.J., Lovas, F.J., Fraser, G.T., and Ilyushin, V.V., 2006, *J. Chem. Phys.*, **125**, 104312.

[149] Møllendal, H., 1993, *Structures and Conformations of Non-Rigid Molecules*, edited by J. Laane, M. Dakkouri, B. van der Veken, and H. Oberhammer (Dortrecht: Kluwer), p. 277.

[150] Demaison, J., Hübner, H., Wlodarczak, G., 1998, *Molecular Constants Mostly from Microwave, Molecular beam, and Sub-Doppler Laser Spectroscopy*, edited by W. Hüttner (Berlin: Landolt-Bönstein, Springer), Vol. 24A.

[151] Demaison, J., Margulès, L., Boggs, J.E., 2003, *Phys. Chem. Chem. Phys.* **5**, 3359.

[152] Demaison, J., Herman, M., Liévin, J., 2007, *J. Chem. Phys.* **126**, 164305.

[153] Demaison, J., Hübner, H., Hüttner, W., Vogt, J., Wlodarczak, G., 1998, *Molecular Constants Mostly from Microwave, Molecular beam, and Sub-Doppler Laser Spectroscopy*, edited by W. Hüttner (Berlin: Landolt-Bönstein, Springer), Vol. 24C.

[154] Demaison, J., Császár, A.G., and Dehayem-Kamadjeu, A., 2006, *J. Phys. Chem. A* **110**, 13609.

[155] Schafer, A., Horn, H., and Ahlrichs, R., 1992, *J. Chem. Phys.* **97**, 2571.

[156] Császár, A.G., Czakó, G., Furtenbacher, T., Tennyson, J., Szalay, V., Shirin, S.V., Zobov, N.F., and Polyansky, O.L., 2005, *J. Chem. Phys.* **122**, 214305.

[157] Dyall, K.G., Taylor, P.R., Faegri, K., and Partridge, H., 1991, *J. Chem. Phys.* **95**, 2583.

[158] Margulès, L., Demaison, J., and Boggs, J.E., 2000, *J. Mol. Struct. (Theochem)* **500**, 245.

[159] Margulès, L., Demaison, J., and Boggs, J.E., 2000, *Struct. Chem.* **11**, 145.

[160] Authier, N., Balland, N., and Le Floch, A., 1993, *J. Molec. Spectrosc.*, **160**, 590.

[161] Ogilvie, J.F., 1994, *J. Phys. B*, **27**, 47.

[162] Canè, E., Fusina, L., Bürger, H., Jerzembeck, W., Brünken, S., Lewen, F., and Winnewisser, G., 2002, *J. Molec. Spectrosc.*, **215**, 1.

[163] Canè, E., Di Lonardo, G., Fusina, L., Jerzembeck, W., Bürger, H., Breidung, J., and Thiel, W., 2006, *J. Molec. Struct.*, **780-781**, 98.

[164] Canè, E., Di Lonardo, G., Fusina, L., Jerzembeck, W., Bürger, H., Breidung, J., and Thiel, W., 2005, *Mol. Phys.*, **103**, 557.

[165] Watson, J.K.G., 2006, *J. Mol. Struct.* **795**, 263.

[166] Watson, J.K.G., 2005, *J. Mol. Struct.* **742**, 91.

[167] Breidung, J., Demaison, J., D'Eu, J.-F., Margulès, L., Collet, D., Mkadmi, E.B., Perrin, A., and Thiel, W., 2004, *J. Mol. Spectrosc.* **228**, 7.

[168] Demaison, J., Sormova, H., Bürger, H., Margulès, L., Constantin, F.L., and Ceausu-Velcescu, A., 2005, *J. Mol. Spectrosc.* **232**, 323.

[169] Demaison, J., 2006, *J. Mol. Spectrosc.* **239**, 201.

Figure 1. Residuals of the r_0 structure: plot of $1000(I_{exp.} - I_{calc.})$ as a function of $I_{exp.}$ for OCSe [122] (all values in uÅ^2).

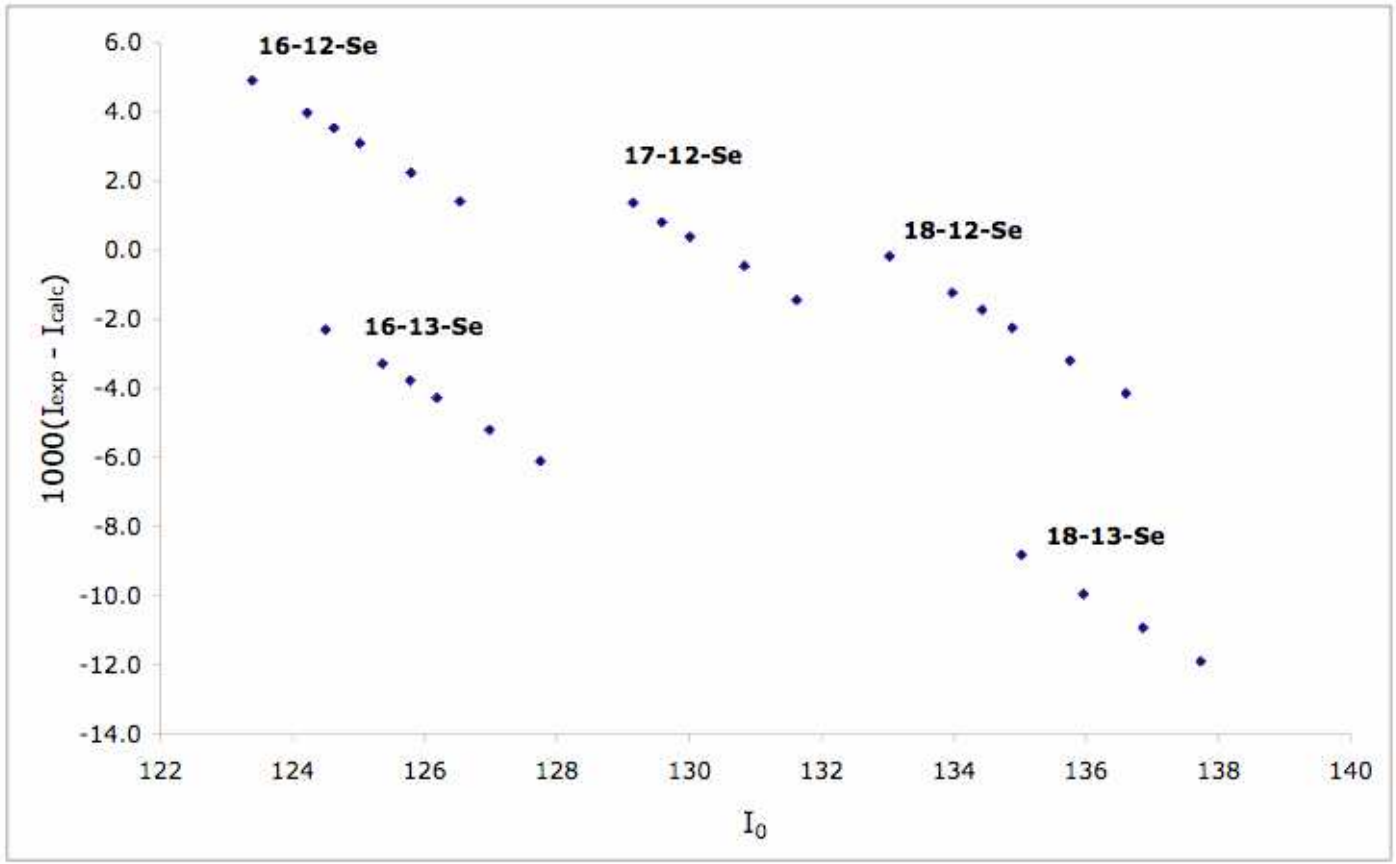


Figure 2. Plot of the variation of ε with isotopic substitution $X^{81}Br \leftarrow X^{79}Br$ for the bromide derivatives of Table 6 of Ref. [118].

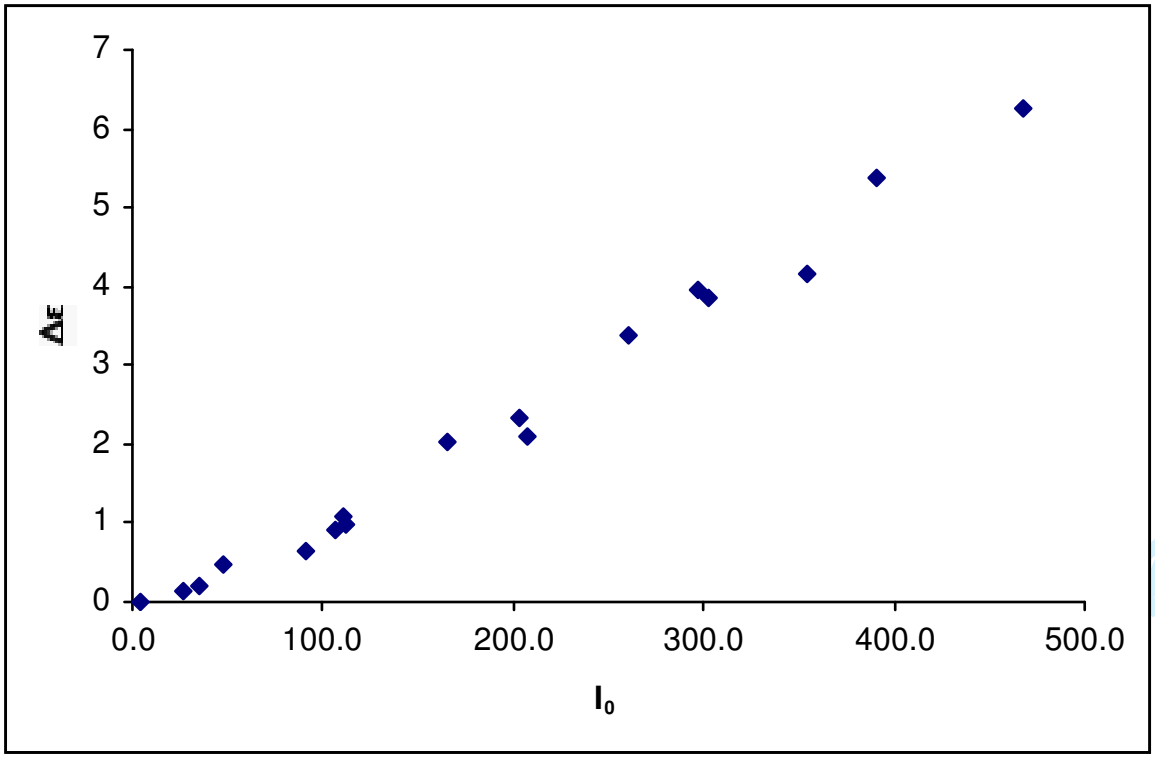


Figure 3. Plot of $\varepsilon = I_0 - I_e$ (uÅ^2) as a function of $m_x m_z/M$ for the XYZ molecule NNO [122].

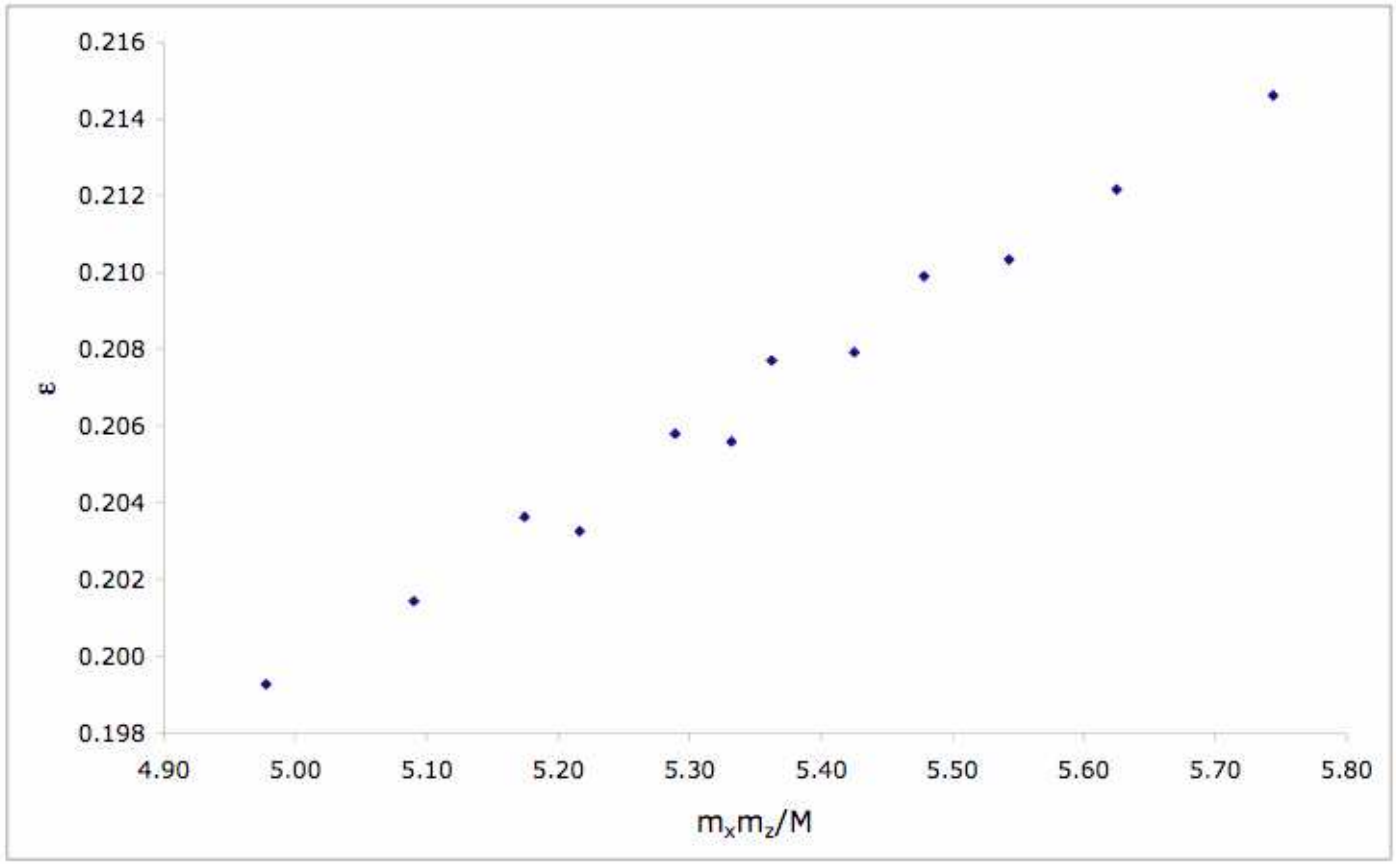


Figure 4. Residuals of the $r_m^{(2)}$ structure: plot of $1000(I_{exp.} - I_{calc.})$ as a function of $I_{exp.}$ for OCSe [122] (all values in uÅ^2).

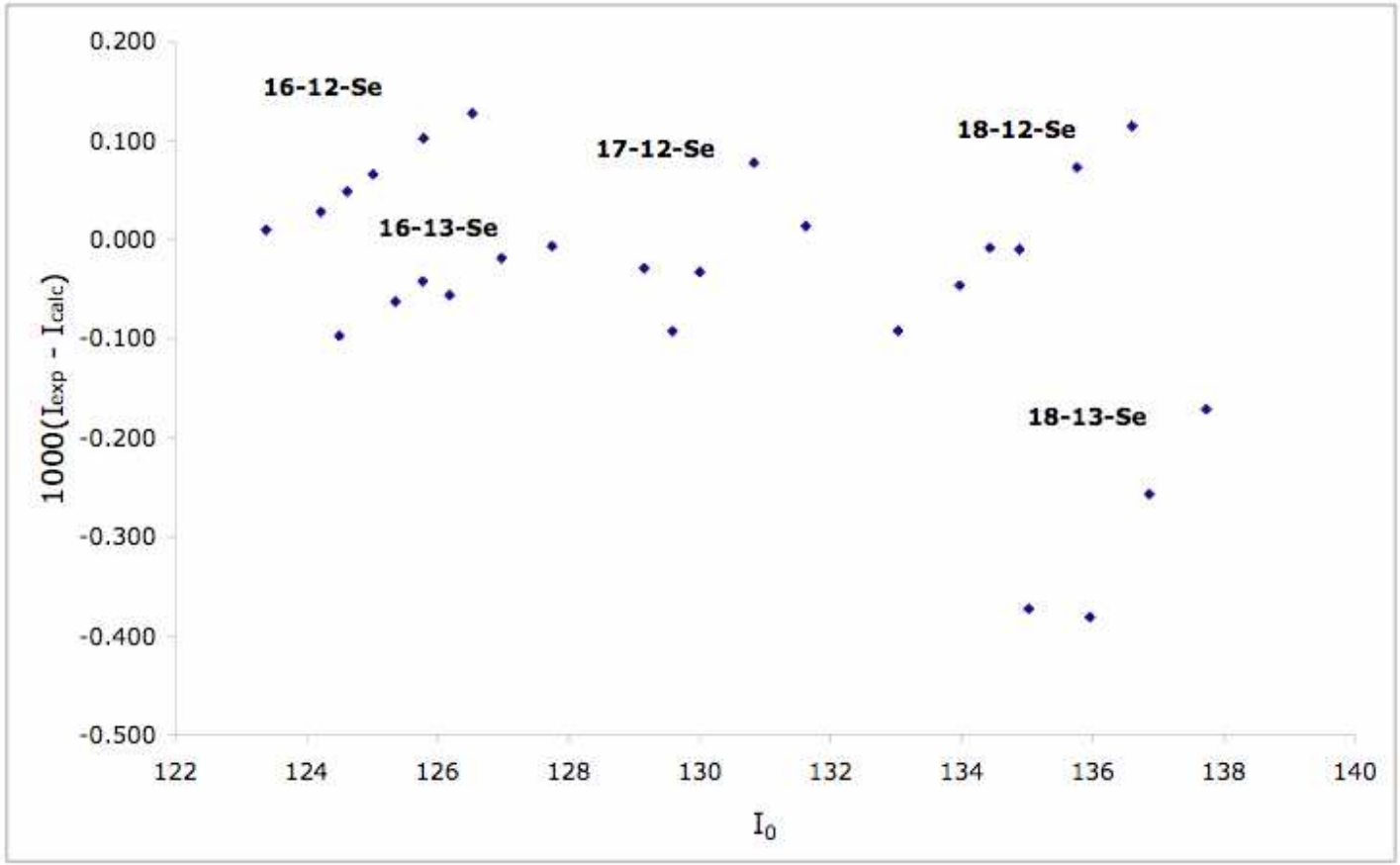


Table 1. Vibrational contribution to the rotational constants (in MHz) [a].

Molecule	B_e	$C = \sum \alpha_i d_i / 2$	C/B (%)	$D = \sum \gamma_{ij} d_i d_j / 4$	D/C (%)
HCN	44511.620	198.137	0.45	2.395	1.21
FCN	10586.782	32.604	0.31	–0.248	–0.76
ClCN	5982.8975	12.0644	0.20	–0.0207	–0.17
BrCN	4126.5059	6.2838	0.15	0.0596	0.95
ICN	3329.0568	3.5084	0.11	–0.0049	–0.14

[a] Data from Ref. [150].

Table 2. Centrifugal contribution to the rotational constants of NH_2 and HCOOH (in MHz)

	exp.	Eqs.(I) [a]	Eqs.(II) [b]	Eqs.(III) [c]	ΔB=I+II+III	$\Delta B_\xi / B_\xi$ (%)
NH_2 [d]						
A	712634.609(5)	61.906	–69.363	199.621	192.163	0.03
B	388213.098(9)	–122.779	–88.220	–19.726	–230.725	–0.06
C	244843.726(9)	2.577	177.916	17.007	197.500	0.08
HCOOH [e]						
A	77512.2354(11)	0.0200	–0.0643	0.7805	0.7362	0.000
B	12055.1065(2)	–0.1556	0.0154	–0.0057	–0.1460	0.000
C	10416.1151(2)	0.0231	0.0454	–0.0079	0.0605	0.000

[a] Eqs. (9).
[b] Eq. (10).
[c] Eq. (11).
[d] Ref. [151].
[e] Ref. [152].

Table 3. Electronic contribution to the rotational constants (B_0 and ΔB in MHz) [a]

molecule	B_0	g	ΔB
LiH	229965.07	-0.6584	82.467
CO	57635.966	-0.2691	8.448
CO_2	11698.4721	-0.05508	0.351
OCS	6081.4921	-0.028839	0.096
OCSe	4017.6537	-0.01952	0.043
CS_2	3271.4882	-0.02274	0.041
HCN	44315.9757	-0.098	2.365
$FC^{15}N$	10186.2903	-0.0504	0.280
$ClC^{15}N$	5748.0527	-0.0385	0.121
$BrC^{15}N$	3944.8441	-0.0385	0.083
$IC^{15}N$	3225.5485	-0.0325	0.057
HC≡C-CN	4549.067	-0.0213	0.053
O_3, *A*	105536.235	0.642	-36.903
O_3, *B*	13349.2547	-0.119	0.865
O_3, *C*	11834.3614	-0.061	0.393

[a] Data from Ref. [153].

Table 4. Accuracy of the ab initio methods for bond lengths [40] (in pm, basis set cc-pVQZ, all electrons correlated).

	mean error [a]	standard deviation	mean absolute error	maximum absolute error
	$\frac{1}{n}\sum_{i=1}^{n} e_i$	$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(e_i-\bar{e})^2}$	$\frac{1}{n}\sum_{i=1}^{n}\lvert e_i\rvert$	$\max_i\lvert e_i\rvert$
HF	–2.91	2.25	2.91	8.45
MP2	–0.26	0.67	0.54	1.67
CCSD	–0.89	0.79	0.89	3.07
CCSD(T)	–0.19	0.30	0.22	1.20

[a] error: $e_i = r_i^{\text{calc}} - r_i^{\text{exp}}$

Table 5. Comparison of the CCSD(T) and CCSDTQ methods (in pm)

Molecule	Δr [a]	Ref.
N_2	0.07	42
CS	0.15	42
CS_2	0.14	42
HF	0.02	43
N_2	0.07	43
F_2	0.34	43
CO	0.06	43

[a] $\Delta r = r$[CCSD(T)] – r[CCSDTQ]

Table 6. Constrained optimization of the r(N–O) bond length (in pm) in *trans*-HONO using various levels of theory [154]

Method [a]	r(N–O) [b]
CCSD	140.1
CCSD(T)	142.1
CCSDT	142.4
CCSDTQ	142.9

[a] The pVDZ basis of Ahlrichs for the middle two atoms and the VDZ basis of Ahlrichs for the two terminal atoms have been used [155]. The frozen-core approximation is employed throughout.

[b] The other structural parameters were fixed at r(O–H) = 96.34 pm, r(N=O) = 116.90 pm, ∠(ONO) = 110.7642°, and ∠(HNO) = 102.0827°.

Table 7. Number of contracted functions in the cc-pVnZ, aug- cc-pVnZ, and cc-pCVnZ basis sets.

		cardinal number				
Atoms	basis	D	T	Q	5	6
H-He	N [a]	5	14	30	55	91
	+aug [b]	4	9	16	25	36
Li-F	N [a]	14	30	55	91	140
	+aug [b]	9	16	25	36	49
	+C [c]	4	13	29	54	90
Na-Cl	N [a, d]	18	34	59	95	144
	+aug [b]	9	16	25	36	49
	+C [c]	9	25	50	70	-

[a] standard cc-pVnZ basis set
[b] number to add to the previous (N) line to get the number of functions in aug-cc-pVnZ set
[c] number to add to the previous (N) line to get the number of functions in cc-pCVnZ set
[d] the use of the cc-pV(n+d)Z is recommended for secondrow atoms, it increases the number of basis functions by 5.

Table 8. Convergence of the basis sets: *r*[aug-cc-pVnZ] – *r*[aug-cc-pV(n-1)Z] (in pm, CCSD(T) method, unless otherwise stated)[a].

	HCCH		HNO		HOOH		HF	$H_2O.HF$
n	*r*(CH)	*r*(CC)	*r*(OH)	*r*(NO)	*r*(OH)	*r*(OO)	*r*(HF)	r(O···H) [b]
T	-1.470	-1.985	-1.060	-1.050	-0.240	0.060	-0.314	-3.760
Q	-0.040	-0.336	-0.190	-0.350	-0.300	-0.750	-0.327	-0.470
5	-0.043	-0.101	-0.020	-0.130	-0.110	-0.200	-0.040	-0.398
6	-0.004	-0.031	-0.010	-0.080	-0.050	-0.050	-0.005	-0.327

[a] This work.
[b] MP2 method with counterpoise correction, see text.

Table 9. Calculation of the core and core-valence correlation in acetylene (in pm) [a].

method	basis set	Δr(CC) [b]	Δr(CH) [b]
CCSD(T)	cc-pwCVDZ	0.076	0.052
	cc-pwCVTZ	0.221	0.107
	cc-pwCVQZ	0.262	0.126
	cc-pwCV5Z	0.274	0.132
	MT [c]	0.26	0.12
MP2	cc-pCVQZ	0.29	0.13

[a] This work.
[b] $\Delta r = r$(frozen core) – r(all electrons correlated)
[c] Martin-Taylor basis set, see text.

Table 10. Effect of the diffuse functions: *r*[aug-cc-pVnZ] – *r*[cc-pVn)Z] (distances in pm, angles in degrees, CCSD(T) method) [a].

	CH_3F			HCCH	
n	*r*(CF)	*r*(CH)	∠(HCF)	*r*(CC)	*r*(CH)
D	2.46	0.318	-1.156	0.134	-0.028
T	0.78	0.031	-0.463	0.051	0.026
Q	0.29	0.022	-0.174	0.037	0.019
5	0.08	0.017	-0.039	0.014	0.010
6				0.008	0.007

[a] This work.

Table 11. Relativistic effects (distances in pm, angles, ∠, in degrees)

Molecule	Δr [a]	method [b]	ref.
N_2	-0.02	DKH	42
CS	0.05	DKH	42
CS_2	0.02	DKH	42
F_2	0.03	DKH	58
Cl_2	0.05	DKH	58
Br_2	–0.32	DKH	58
ClF	0.06	DKH	58
BrF	0.13	DKH	58
BrCl	–0.04	DKH	58
CF_4	0.00	DKH	58
SiF_4	–0.05	DKH	58
Br_2CO, *r*(CO)	–0.05	DKH	58
Br_2CO, *r*(CBr)	–0.23	DKH	58
Br_2CO, ∠(BrCBr)	–0.10	DKH	58
H_2O, *r*(OH)	0.016	Breit	156
H_2O, ∠(HOH)	–0.074	Breit	156
CH_4	–0.013	DHF	157
SiH_4	–0.066	DHF	157
GeH_4	–0.70	DHF	157
SnH_4	–2.06	DHF	157
PbH_4	–7.33	DHF	157

[a] Δr = *r*[relativistic] – *r*[non-relativistic]
[b] DKH = Douglas-Kroll-Hess method ; Breit = Breit-Pauli; DHF = Dirac-Hartree-Fock

Table 12. Diagonal Born-Oppenheimer contribution (distances in pm, angles in degrees)

Molecule	DBOC	Ref.
H_2O, r(OH)	0.003	156
H_2O, ∠(HOH)	0.015	156
H_2	0.0215	59
HF	0.0023	59
N_2	0.0012	59
F_2	0.0001	59

Table 13. Typical offset values $\Delta r = r_e - r$[MP2] for the MP2 method (distances in pm and angles in degrees).

Bond	basis	mean	median	std	max	min	Ref.
angle	cc-pVTZ	0.38 [a]	0.31 [a]		–1.37		158
	cc-pVQZ	0.25 [a]	0.21 [a]		1.09		158
CC [b]	cc-pVTZ	–0.46	–0.40	0.52	0.65	–1.72	159
	cc-pVQZ	–0.17	–0.12	0.50	0.77	–1.32	159
NH	cc-pVTZ	–0.09	–0.07	0.19	0.25	–0.57	139
	cc-pVQZ	0.01	0.00	0.19	0.40	–0.39	139
OH	cc-pVTZ	–0.31	–0.31	0.10	0.54		138
	cc-pVQZ	0.15	0.15	0.11	0.38		138

[a] for absolute values of the residuals
[b] note that the offset for the CC bond depends on the environment, see text.

Table 14. Adiabatic and Born-Oppenheimer equilibrium structures of carbon monoxide (CO) and hydrogen chloride (HCl). All values in pm.

type	isotope	*r*	isotope	*r*
adiabatic	$^{12}C^{16}O$	112.833 634 6	$H^{35}Cl$	127.457 168 7
	$^{12}C^{17}O$	112.833 399 7	$H^{37}Cl$	127.457 160 7
	$^{13}C^{16}O$	112.833 219 6	$D^{35}Cl$	127.459 403 0
	$^{12}C^{18}O$	112.833 182 5	$D^{37}Cl$	127.459 394 8
	$^{14}C^{16}O$	112.832 881 7	$T^{35}Cl$	127.460 225 8
	$^{13}C^{18}O$	112.832 767 6	$T^{35}Cl$	127.460 117 5
BO		112.8230(1) [a]		127.460 40(11) [b]

[a] Ref. [160].
[b] Ref. [161].

Table 15. Equilibrium structures of phosphine and stibine (distances in pm and angles in degees).

X	from	r_e(X-H)	$\angle_e$(HXH)	Ref.
P	PH_3	141.1607(16)	93.4184(21)	162
	PD_3	141.17846(16)	93.42523(19)	162
	CCSD(T)/wCVQZ	141.107	93.498	this work
	CCSD(T)/wCV5Z	141.098	93.501	this work
Sb	$^{121}SbH_3$	170.0013(1)	91.5586(1)	163
	$^{121}SbD_3$	170.0909(2)	91.6713(2)	163
	$^{123}SbD_3$	170.0897(3)	91.6728(3)	164
	CCSD(T)/SDB-AVQZ	170.15	91.76	164

Table 16. main experimental sources of information on the force constants [19, 100].

parameter	symbol	order of force constant
harmonic wave number	ω	2
equilibrium quartic centrifugal distortion constant	D_e	2
Coriolis coupling constant	ζ	2
inertial defect (planar molecule)	Δ	2
mean amplitude of vibration	u	2
vibration-rotation constant	α	2, 3
sextic centrifugal distortion constant	H or Φ	2, 3
second-order Coriolis	F	2, 3
isotopic effect on r_z (average structure)	δr_z	2, 3
ℓ-type doubling constants (symmetric molecule)	q_t, r_t	2, 3
diagonal Coriolis coupling (symmetric molecule)	η_J,, η_K	2, 3
Fermi resonance constant	W	3
vibrational frequency or	ν	2, 3, 4
vibrational anharmonicity	x	2, 3, 4
octic centrifugal distortion constant [a]	L	2, 3, 4
vibrational dependence of the quartic centrifugal distortion constant [b]	$D_v - D_e$	2, 3, 4
Darling-Dennison resonance constant	W	2, 3, 4

[a] Ref. [165].
[b] Ref. [166].

Table 17. Comparison of ab initio, experimental and semi-experimental equilibrium parameters for a few molecules (distances in pm and angles in degees).

Molecule	parameter	ab initio	experimental	semi-experimental	Ref.
HOF	*r*(OH)	96.64	96.57(16)	96.862(8)	138
	r(OF)	143.26	143.50(31)	143.447(11)	
	∠(HOF)	97.97	97.54(50)	97.86(2)	
BHFOH	*r*(OH)	95.69		95.853(6)	138
BF_2OH	*r*(OH)	95.74		94.30(3)	167
	∠(BOH)	113.14		114.1(2)	
O=CHCl	*r*(CCl)	176.17		176.30(2)	51
SiH_3Cl	*r*(SiH)	146.95	146.84	146.96	168
	r(SiCl)	204.77	204.58	204.58	
HNO	*r*(NH)	105.18	106.3(3)	105.20(6)	154
	r(N=O)	120.76	120.6(3)	120.86(7)	
	∠(HNO)	108.10	109.1(2)	108.16(18)	

Table 18. Variation of the vibrational correction ε as a function of the moment of inertia I_0 (in uÅ^2) [a].

Molecule	number of species	I_0	ε mean	ε range	ε/I_0 (%)
HCN	11	11.707	0.049(1)	0.007	0.4
N_2O	12	40.232	0.206(4)	0.015	0.5
OCS	12	83.101	0.250(6)	0.018	0.3
OCSe	27	125.019	0.349(9)	0.026	0.3

[a] Data from Ref. [150].

Table 19. Value of the angle ∠(CCH$_g$) in H$_2$C=CH$_g$F (in degrees) [169].

year	method	value
1958	r_0	123.7
1961	r_0	129.2
1961	r_0	120.9
1974	r_g [a]	127.7(7)
1979	r_z [a]	130.8(25)
1989	r_s	124.35(63)
1992	r_0	127.6(42)
2006	r_e	125.95(20)

[a] from electron diffraction

Table 20. *c* and *d* parameters of the $r_m^{(2)}$ method for a few XCN molecules (X = H, F, Cl, Br, and I) [a].

X	c / u$^{1/2}$Å	d / u$^{1/2}$Å^2
H	0.0272	-0.0055
F	0.0198	0.0155
Cl	0.0137	0.0186
Br	0.0110	0.0207
I	0.0072	0.0211

[a] This work.

Table 21. Molecules for which a $r_m^{(1)}$ or a $r_m^{(2)}$ structure was determined.

Molecule	structural parameters	isotopologues	symmetry	ref.
SH_3^+	2	5	C_{3v}	a
ClNO	3	6	planar	b
$BrNO_2$	3	6	planar	c
$(NO)_2$	3	3	planar	d
CH_3NC	4	12	C_{3v}	e
HONO	5	5	planar	b
HCOCl	5	9	planar	f
HNSO	5	5	planar	g
HC_5N	6	14	linear	h
$(CH_3)_2O$	7	7	C_{2v}	i
N_2O_3	7	7	planar	j
1,2,5-oxadiazole	8	9	planar	k
H_2SO_4	8	4	C_2	l
$HCONH_2$	9	16	planar	i
CH_2=CHBr	9	16	planar	m
CH_2=CHCl	9	7	planar	n
CH_2=CHCN	11	14	planar	o
C_6H_5F	13	9	planar	p
$C_6H_5C\equiv CH$	13	39	planar	q

[a] Tinti, F., Bizzocchi, L., Degli Esposti, C., and Dore, L., 2006, *J. Molec. Spectrosc.*, **240**, 202.
[b] Demaison, J., Császár, A.G., and Dehayem-Kamadjeu, A., 2006, *J. Phys. Chem. A* **110**, 13609..
[c] Kwabia Tchana, F., Orphal, J., Kleiner, I., Rudolph, H.D., Willner, H., Garcia, P., Bouba, O., Demaison, J., and Redlich, B., 2004, *Molec. Phys.*, **102**, 1509.
[d] McKellar, A.R.W., Watson, J.K.G., and Howard, B.J., 1995, *Molec. Phys.*, **86**, 273.
[e] Margulès, L., Demaison, J., Rudolph, H.D., 2001, *J. Molec. Struct.* **599**, 23.
[f] Demaison, J., Boggs, J.E., and Rudolph, H.D., 2004, *J. Molec. Struct.* **695**, 145.
[g] Demaison, J., Margulès, L., Boggs, J.E., and Rudolph, H.D., 2001, *Struct. Chem.*, **12**, 1.
[h] Bizzocchi, L., Degli Esposti, C., and Botschwina, P., 2004, *J. Molec. Spectrosc.*, **225**, 145.
[i] Demaison, J., Császár, A.G., Kleiner, I., Møllendal, H., 2007, *J. Phys. Chem. A*, **111**, 2574.
[j] Demaison, J., Herman, M., Liévin, J., Margulès, L., and Møllendal, H., 2007, submitted.
[k] Vázquez, J., Demaison, J., López-González, J.J., Boggs, J.E., and Rudolph, H.D., 2001, *J. Molec. Spectrosc.*, **207**, 224.
[l] Demaison, J., Herman, M., Liévin, J., and Rudolph, H.D., 2007, *J. Phys. Chem. A*, **111**, 2602.
[m] Zvereva-Loëte, N., Demaison, J., and Rudolph, H.D., 2006, *J. Molec. Spectrosc.*, **236**, 248.
[n] Demaison, J., Møllendal, H., Perrin, A., Orphal, J., Kwabia Tchana, F., Rudolph, H.D., Willaert, F., 2005, *J. Molec. Spectrosc.*, **232**, 174.
[o] Askeland, E., Møllendal, H., Uggerud, E., Guillemin, J.-C., Aviles Moreno, J.-R., Demaison, J., and Huet, T., 2006, *J. Phys. Chem. A*, **110**, 12572.
[p] Kisiel, Z., Bialkowska-Jaworska, E., and Pszczólkowski, L., 2005, *J. Molec. Spectrosc.*, **232**, 47.
[q] Dreizler, H., Rudolph, H.D., and Hartke, B., 2004, *J. Molec. Struct.*, **698**, 1.