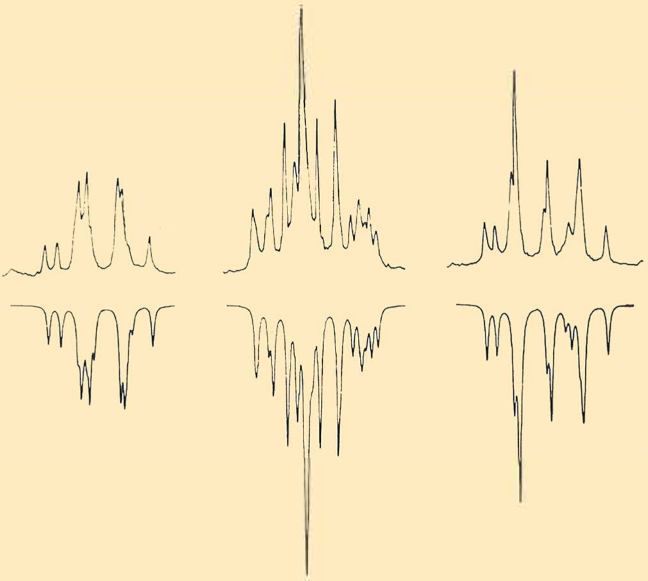
# COMPUTERIZED NMR: DEVELOPMENTS AND APPLICATIONS OF ANALYSIS AND INTERPRETATION METHODS OF HIGH-RESOLUTION SPECTRA

By

REINO LAATIKAINEN



JYVÄSKYLÄ 1979

# COMPUTERIZED NMR: DEVELOPMENTS AND APPLICATIONS OF ANALYSIS AND INTERPRETATION METHODS OF HIGH-RESOLUTION SPECTRA

By

REINO LAATIKAINEN

Department of chemistry,
university of Jyväskylä

To be presented, by permission of the Faculty of
Mathematics and Natural Sciences of the University
of Jyväskylä
for public examination in Blomstedt's Hall, Villa
Rana, on December 14, 1979, at 12 o'clock noon

Jyväskylä, Finland
1979

**URN:ISBN:978-951-39-9878-3**
**ISBN 978-951-39-9878-3 (PDF)**

**Jyväskylän yliopisto, 2023**

**ISBN 951-99239-3-4**

Kuopio 1979 — Antin Paino Oy

To my parents

# PREFACE

The present studies were carried out in the Department of Chemistry, University of Jyväskylä and in the Department of Chemistry, University of Kuopio during the years 1976-1979.

I wish to express my gratitude to Professor Jaakko Paasivirta, University of Jyväskylä, and to Associate Professor Pertti Äyräs, University of Turku, for their very valuable advices, encouragements, ideas and cooperation. I also want to thank Heikki Joela, Ph.D., for many encouraging discussions.

I am very grateful to Simo Lötjönen, M.Sc., and to Juha Knuutinen, Ph.Lic., for the good cooperation and interesting discussions. (And as to the ultimate questions of Science and Human life, I also want to thank them, as well as Harri Häkli, Ph.D., and Erkki Kolehmainen, M.Sc., for many good discussions).

My sincere thanks are due to Jukka Jokisaari, Ph.D., University of Oulu, for reading my manuscript and for his very valuable chriticism.

I am indebted to Roger Price, M.Sc., for inspecting the language of some of the manuscripts and to Miss Tuula Hartikainen for typing the final text.

This work was supported by a grant from the Emil Aaltonen Foundation, which is gratefully acknowledged.

Finally my deepest gratitude I owe to my wife for the understanding, support and also typing of the first version of this dissertation during the most beautiful summer days in 1979.

In Kuopio, December 1979

# CONTENTS

7.1. About the Reliability, Efficiency, and Meaning of the Error Estimates of a LAOCOON3-Type NMR Analysis System

R. Laatikainen, *J. Magn. Reson.* **27**(1977)169.
https://doi.org/10.1016/0022-2364(77)90066-X

7.2. An Automatically Assigning LAOCOON3-Type Algorithm in the Analysis of High-Resolution Proton Coupled $^{13}C$ NMR Spectra

R. Laatikainen. Submitted to J. Magn. Reson.

7.3. The $^{13}C$,H Coupling Constants in Structural and Conformational Analysis IV. The Rotational Dependence of the Long-Range CH Couplings of Aldehyde and Hydroxyl Protons in Salicylaldehyde

P. Äyräs, R. Laatikainen and S. Lötjönen, *Org. Magn. Reson.*, in press.
https://doi.org/10.1002/mrc.1270130520

7.4. An Experimental and Theoretical FPT/INDO Study on $^{n}J(\mathrm{O\underline{H}},\underline{\mathrm{C}})$, $^{n}J(\mathrm{O\underline{H}},\underline{\mathrm{H}})$, $^{n}J$(CH) and $^{n}J$(HH) Couplings of Methanol, Ethanol, 2-Propanol, *tert*-Butyl Alcohol and Acetylacetone

R. Laatikainen, S. Lötjönen, P. Äyräs. Submitted to Acta Chem. Scand. *A*.
https://doi.org/10.3891/acta.chem.scand.34a-0249

7.5. Optimized Parametrizations for the Calculations of the Spin-Spin Coupling Constants by the INDO/FPT and INDO/SOS Methods: A Comparison of the Methods

R. Laatikainen, a manuscript

7.6. A Statistical Study of the Additivities of the Substitutional Effects on the $^{13}C$ NMR Chemical Shifts in Hydroxy-and Chloro-Substituted Benzenes

J. Knuutinen, R. Laatikainen and J. Paasivirta. Submitted to Org. Magn. Reson.
https://doi.org/10.1002/mrc.1270140507

7.7. A CNDO/2 Study on the Additivity and the Nature of the Non-Additivity of the Substituent Effects on $^{13}C$ NMR Shifts in Chlorobenzenes and Chlorophenols

R. Laatikainen. Submitted to Org. Magn. Reson.
https://doi.org/10.1002/mrc.1270140508

# 1. INTRODUCTION

## 1.1. NMR spectroscopy

The first observed property of matter arising from the magnetic properties of nuclei, the hyperfine structure of atomic spectra (Michelson 1891[1]), was explained by Pauli in 1924[2]. The use of the atomic and molecular beam techniques led to the measurements of accurate magnetic moments during the following decades. And it was in 1945 when nuclear magnetic resonance was detected in bulk matter[3]. The principle of chemical shift and spin-spin coupling was documented in 1950[4].

The rapid development of instrumentation led to the very intense application of NMR in all areas of chemistry. Especially, the progress in the pulse Fourier transformation techniques has been very fast during the 1970s, and at present, in 1979, the possibilities of supraconductivity offer very interesting opportunities.

However, the use of NMR in chemistry is still greatly based on many empirically observed trends and rules. Although a great deal of theoretical work has been done after the perturbation theory treatment of Ramsay (1950-1953)[5], no practically useful quantitative theory has been developed to cover all the essential relationships of the chemical structures and NMR parameters. Anyway, the recent very rapid growth of NMR information has emphasized the need for such a theoretical framework.

## 1.2. Computerized NMR

There would be no modern spectroscopy or quantitative quantum mechanics without modern computer and microprocessor techniques.

In NMR spectroscopy there are three main areas in which the aid of computers is essential: the instrumentation, the analysis of high-resolution NMR spectra, and the theoretical considerations.

One very notable trend in the development of chemistry has been the growth of the use of the molecular orbital theoretical aspects in the interpretations of chemical properties. In my opinion, this development is important because the language of molecular orbital theory is needed to combine the very badly diverged theories and empirical models of different areas of chemistry.

The aim of these studies has been to build up and to test a useful program library for spectral analysis and calculations of theoretical NMR parameters, on the grounds of semiempirical SCF MO theory. The present work is composed of a theoretical part and seven separate papers or manuscripts reproduced in chapters 7.1.-7.7. (see also the table of contents). The studies are carried out partly at the University of Jyväskylä and partly at the University of Kuopio during 1976-1979.

The purpose of the first five sections is to discuss some trends that are not included or that are incompletely discussed in chapters 7.1. - 7.7.

## 2. THE PROPERTIES AND ANALYSIS OF NMR SPECTRA[6]

### 2.1. Nuclei in magnetic fields

The quantized states of nuclei in the magnetic field can be described by the states of the angular momentum. The states of the intrinsic spin angular momentum are denoted for spin 1/2 particles by $|\alpha>$ and $|\beta>$.

Chemically non-equivalent nuclei of a molecule may experience slightly different strengths of magnetic field, due to shielding by electrons. Furthermore, there may occur interactions between the nuclei. The determination of the shielding parameters (the chemical shifts) and the interaction parameters (the spin-spin coupling constants) is the purpose of the spectral analysis.

Let us consider protons A and B in the magnetic field. When the nuclei are chemically different and there are no interactions between them, the states of the system can be described by the basic product functions $|\alpha\alpha>$, $|\alpha\beta>$, $|\beta\alpha>$ and $|\beta\beta>$. If the coupling between the nuclei is small enough when compared with the difference between the chemical shifts of the molecule, a satisfactory description of the system is still obtained by using the basic product functions.

When the coupling constant $|J_{AB}|$ is comparable to the difference between the chemical shifts $|(v_A-v_B)|$, the states of the system can be described by using the mixed wave functions:

$$\psi_i = \sum_{i=1}^{4} c_{ij} \phi_j \qquad 2.1$$

where $\phi_j$:s are the basic product functions and $c_{ij}$:s are the constants that can be solved by using the standard perturbation method$^6$.

The interactions of the magnetic moments of the nuclei can be divided into two independent parts: the interactions of the z-components of the nuclei and those of the rotating xy-components. However, only the latter perturbation is able to mix the basic product functions. In other words, if the interactions of xy-components can be neglected, the basic product functions are eigenfunctions of the total hamiltonian, and the spectral analysis can be carried out directly by the first-order methods.

When the eigenfunctions and eigenvalues of the AB system are determined accurately, including also the mixing perturbations, the following equations are obtained for the relations between the spectral parameters and the observed spectral lines:

$$\begin{aligned} v_A - v_B &= \sqrt{(E_1 - E_4)(E_2 - E_3)} \\ J_{AB} &= (E_1 - E_2) = (E_3 - E_4) \end{aligned} \qquad 2.2$$

The absorptions $E_1$ and $E_2$ can be said to be due to the change of the state of nucleus A, and $E_3$ and $E_4$ correspondingly due to B. When $|v_A - v_B|$ is large enough, $(v_A - v_B) \simeq (E_1 + E_2 - E_3 - E_4)/2$, and thus the spectrum can be analysed graphically. This corresponds to the neglecting of the second-order perturbations. As a rule, when $|v_A - v_B| < |10J_{AB}|$ it is necessary to use Eqs. 2.2 if exact $(v_A - v_B)$ is desired. When the spectrum is of type $A_2$ (t.i. $v_A = v_B$), $J_{AA}$ cannot be observed because the intensities of $E_1$ and $E_4$ are zero. When, $|v_A - v_B| << |J_{AB}|$, $E_2$ and $E_3$ come very near to each other and the relative accuracy of $(E_2 - E_3)$ becomes poor, appearing as a large inaccuracy of $(v_A - v_B)$. If a spectrum contains only a single line one cannot decide on the basis of the spectrum whether it is a coupled, non-coupled, or a strongly coupled AB spectrum. The details and more interesting examples of the phenomenon are discussed in chapter 6.1.

The phenomenon can be illustrated with a more concrete analogy: in a molecule of type B-A-B´ one cannot observe infrared absorptions that would purely correspond to the vibrations of the single B-A or A-B´ bonds if the force constant $k_{BB'} \neq 0$. Also in this case, one must describe the total vibrations using the mixed vibrations.

Correspondingly, in $A_2$ NMR spectra the total states must be described by mixing the basic functions $|\alpha\beta>$ and $|\beta\alpha>$ completely:

$$\psi_2 = (|\alpha\beta> + |\beta\alpha>)/\sqrt{2}$$

$$\psi_3 = (|\alpha\beta> - |\beta\alpha>)/\sqrt{2} \qquad 2.3$$

In the B-A-X molecule one can separate the pure B-A and A-X vibrations only when the differences of the single bond vibration frequencies are large enough or the force constant $k_{BX}$ is small enough. The total vibration eigenfunctions must also in this case be described by more or less mixed functions (called normal coordinates). Thus, the analysis of the coupled AB NMR spectra is analogical with the analysis of the vibration frequencies of single B-A and A-X bonds and the force constant $k_{BX}$.

## 2.2. Explicit methods of spectral analysis

The AB system is the most complicated general spin system for which the spectral parameters can be expressed with practically useful, explicit formula. However, in many cases the mixing power of the second-order perturbations can be neglected or the hamiltonian can be reordered by symmetry considerations. Before the routine use of computers, approximative methods were developed for very many spin systems by using the second-order perturbation formula and the moment methods[7,8]. Nowadays they are almost completely replaced by the use of computers, although the standard perturbation formula is also found useful in the computer applications[9].

The principle of X approximation was explained in the previous chapter.

The consideration of the symmetry of the spin systems leads, for example, to the explicit analytical schemes of $AB_n$ and AA'BB' systems[10]. Both the symmetry considerations and X approximation are nowadays connected to computer algorithms[11].

A computer program cannot totally replace the explicit methods because the trial parameters for the numerical methods must be found. For example, in the case of the AA'BB' system a short study of the literature may reveal the type of the spectrum, and thus helps in the choice of the trial parameters.

## 2.3. Numerical methods

The principle of all the numerical methods is the same: the trial values of the spectral parameters are guessed on the basis of a supposed structure and the appearance of the spectrum (an approximate explicit analysis is not usually necessary), the spectrum is presented by a group of numbers(by places of the absorption maxima or in a digital form), and the trial parameters are refined iteratively to give the least square fit to the observations.

The first numerical method was that of Swalen and Reilly[12]. The principal idea of their program NMREN/NMRIT is: the eigenvalues of the Hamiltonian can be calculated from the observed spectra, and the corresponding diagonals of the undiagonalized hamiltonian can be calculated from these eigenvalues by using the eigenvectors which are calculated with the trial parameters. More exact parameters are now calculated from the diagonal elements of the undiagonalized hamiltonian (which can be presented as linear functions of the spectral parameters), and these new parameters are used as the new trial parameters. The iteration is continued until a sufficient convergence is obtained. The method is based on one of the main properties of the perturbation theory: first-order wave functions predict the energy of a system correct to even the third order. The principle is also exploited in SCF MO calculations. - However, the convergence of the method is not comparable to that of the Newton-Raphson method and the method is hardly used nowadays[11].

The most effecient method of iteration is the Newton-Raphson scheme. Generally, the problem of the spectral analysis is reduced to the solution of the group of equations to the least square criterium:

$$\begin{aligned} F_1(\bar{x}) &= O_1 \\ F_2(\bar{x}) &= O_2 \\ &\cdots\cdots \\ F_n(\bar{x}) &= O_n \end{aligned} \qquad 2.4$$

in which an observation $O_i$ is expressed as the function $F_i$ of parameter vector $\bar{x}$. In the classical NMREN/NMRIT[12] and LAOCOON III[13] methods $\bar{x}$ contains only the spectral parameters. In the more present developments[9,14] $\bar{x}$ also contains the line-shape parameters.

Generally, the methods can be divided into three classes:

(I) $O_i$ is the frequency of the transition $i$[13,15,16] or an eigenvalue[12]. $O_i$:s might also include the intensities of the transitions[16], although there is no document in the literature that such a method has been realized.

(II) $O_i$ is the intensity of the spectrum at frequency $v_i$. In other words the digitalized spectrum is directly iterated[9].

(III) $O_i$ is an integral transformation of the digitalized spectrum[14]:

$$O_i = \int_{a-d}^{a+d} W_i(v) \; I(v) \; dv \qquad 2.5$$

in which I(v) is the intensity of the observed spectrum as the function of the frequency v. $W_i$ is a member of a certain basis, for example:

$$W_i(v) = (d/2)^{-1/2} \cos\left[(i\pi/d)(v-a)\right] \qquad 2.6$$

in which d is the width of the frequency scale and a is the midpoint of it. When also the sine functions are included, a complete orthonormalized trigonometric basis is obtained. The trigonometric basis can be replaced by many other bases.

The method of the iteration is described in chapter 7.1[15]. The methods differ from each other in respect of the speed of convergence to the best solution: this depends greatly on the linearity of the functions $F_i(\bar{x})$ in Eq. 2.4. The linearity is best in the type I methods and the convergence is typically observed within 1-2 iteration cycles. In the types II and III the convergence is found to be quite slow[9,14] because of the non-linearity of the line-shape functions and integral transformation bases, see also chapters 7.1. and 7.2. The most troublesome part of the type I analysis has been the assigning of the observed lines. In the present work, an algorithm was developed to deal with that difficulty, see chapter 7.2.

There is still one important aspect. The dimensions of the hamiltonian sub-matrices grow very rapidly when the number of the spins grows. For this reason, the systems greater than the general eight spins are unreliable even for the largest available computers. However, the systems of even 11 spins (see 7.4.) are analysed by using computers, if (i) the spectrum can be simplified by using the double resonance techniques or the spectrum is divided to sub-spectra, (ii) the hamitonian is simplified by using the X approximation, or (iii) symmetry considerations. The third principle is the most important. The first algorithm for this was presented by Ferguson and Marguardt[17], and the same algorithm is also applied in the programs NMR-LAOCN-4A[18] and LAME[11]. See also the chapters 5.1. and 7.4.

## 3. THE INTERPRETATION OF SPIN-SPIN COUPLINGS

### 3.1. The perturbation approach

When the direct dipolar interaction of two nuclei is averaged to zero in isotropic media, there still occurs an indirect coupling due to the polarization in the electronic environments of the nuclei [5,19]:

$$E_{AB} = h \; J_{AB} \; \vec{I}_A \cdot \vec{I}_B \qquad 3.1$$

$E_{AB}$ is the interaction energy (in Joule) of the nuclei A and B, $J_{AB}$ is the coupling constants (in Herz) and $\vec{I}_A$ and $\vec{I}_B$ are the spin angular moments of the nuclei A and B (in units of $\hbar$). A standard way to calculate the interaction energy and the coupling constant is to use the perturbation theory: the interaction is described by letting the interactions of the nuclear magnetic dipoles and electrons to mix the ground state of the molecule with the triplet excited states of the molecule. A further description of the phenomenon is found when one thinks that the magnetic nuclei are able to induce small spin densities into the electron cloud of the molecule, and that the coupling of the nuclei A and B is due to the interaction of the magnetic momentum of A with the spin-density caused by B (and vica versa).

The interactions, and thus also the perturbing part of the total hamiltonian $\hat{H}$ can be divided into three independent parts[5,19]:

$$\hat{H} = \hat{H}^{o} + \hat{H}^{orb} + \hat{H}^{dip} + \hat{H}^{cont} \qquad 3.2$$

$\hat{H}^{orb}$ and $\hat{H}^{dip}$ are due to the interactions of the dipolar field of the nuclei with the orbital motion of electrons, and respectively, with the spins of the electrons. The contact part $\hat{H}^{cont}$ is due to the direct Fermi contact interaction of the nuclei and the electron spins.

The operator $\hat{H}^{cont}$ for one electron (i) and the nucleus A is written

$$\hat{H}^{cont} = (2\mu_o/3)g(\beta/\hbar)\gamma_A\vec{I}_A \cdot \vec{S}_i\delta(\vec{r}_{iA}) \qquad 3.3$$

where $\mu_o$ is the permeability of the space, g is the g-value of the electron, $\beta$ is the Bohr magneton, $\gamma_A$ is the gyromagnetic ratio of the nucleus A, $\vec{S}_i$ is an operator that measures the spin angular momentum of electron i, and $\delta(\vec{r}_{iA})$ is the Dirac delta function which measures the square amplitude of the wave function at the nucleus. When the interaction is summed over all the electrons the second-order formula is obtained for this coupling:

$$J_{AB}^{cont} = -\frac{\hbar}{3\pi}(2\mu_o g\beta/3\hbar)^2\,\gamma_A\gamma_B \sum_n \Big[ \langle 0|\sum_p \delta(\vec{r}_{pA})\,\vec{S}_p|n\rangle \cdot \langle n|\sum_q \delta(\vec{r}_{qB})\,\vec{S}_q|0\rangle \Big] \Big/ {}^3\Delta E_{0\to n} \qquad 3.4$$

This is the most important mechanism of the coupling when one of the coupled nuclei, A or B, is hydrogen.

Corresponding expressions can be derived also for $J_{AB}^{orb}$ and $J_{AB}^{dip}$. The magnitude of the second-order part of $J_{AB}^{orb}$ (there exists also a first-order part) depends on the magnitude of the orbital angular momenta and thus, on the multiple bond nature of the bonds between A and B. For this reason, the mechanism is not important when one of the coupled nuclei is a proton. But in many cases, for example, in CC, CF, CP and FP couplings, the mechanism is supposed to be quite important[19,20].

$J_{AB}^{dip}$ is due to the dipolar interactions of the nuclear and electron spin magnetic momenta. It can be shown[19] to be dependent on the double bond and p-electron nature of the AB bond. It is also supposed to be less important than the dipole-orbital and Fermi contact mechanisms[19,20].

## 3.2. The independent electron model

The simplest description of the molecular orbitals is obtained when the orbitals $\psi_i$ are described by the linear combinations of the atomic orbitals $\Phi_u$ (LCAO:s):

$$\psi_i = \sum_u c_{ui} \Phi_u \qquad 3.5$$

in which the coefficients $c_{ui}$ are obtained after one variation calculation, by completely neglecting the internal correlations of the electron motions. These kinds of the methods, like the Hückel and extended Hückel methods, are called also the independent electron models. By using the LCAO:s and Eq. (3.4) a simple expression for $J_{AB}^{cont}$ can be derived[19]. Firstly,

$$J_{AB}^{cont} = -\frac{\hbar}{2\pi} (2\mu_o g\beta/3)^2 \gamma_A \gamma_B \sum_i^{occ.} \sum_j^{unocc.} \sum_{rstu} c_{ri} c_{sj} c_{tj} c_{ui} \times$$

$$\langle\Phi_r|\delta(\vec{r}_A)|\Phi_s\rangle \langle\Phi_t|\delta(\vec{r}_B)|\Phi_u\rangle \;/\; {}^3\Delta E_{i\to j} \qquad 3.6$$

Very many of these integrals will be small because some pairs of the orbitals will only have a small product in the neighbourhood of the nuclei A and B. The neglecting of the integrals when $r \neq s$ and $t \neq u$ is thus a reasonable approximation:

$$J_{AB}^{cont} = \frac{\hbar}{2\pi}(2\mu_o g\beta/3)^2 \quad \gamma_A\gamma_B \quad s_A^2(0)\ s_B^2(0) \times (-4) \sum_i^{occ.} \sum_j^{unocc.}$$

$$c_{sAi}c_{sAj}c_{sBi}c_{sBj} \;/\; {}^3\Delta E_{i\to j} \qquad 3.7$$

or more simply

$$J_{AB}^{cont} = \text{CONSTANT} \times \Pi_{sA,sB} \qquad 3.8$$

where $s_A^2(0)$ is the $s_A$-electron density in the nucleus A and $\Pi_{sA,sB}$ is called the mutual polarizability of the orbitals $s_A$ and $S_A$. ${}^3\Delta E_{i\to j}$ is the triplet excitation energy for the states i and j (for the transition of the electron from the i:th occupied orbital to the j:th unoccupied orbital). The values of $s_A^2(0)$ can be calculated theoretically, or more likely, they are chosen to give the least square fit with the experiments. Furthermore, the value of $s_A^2(0)$ depends on the effective charge of the nucleus A and so also on the electron density at A. However, this defect is supposed to be quite impractical in the calculations of coupling constants[19], in contrast to chemical shift calculations.

The greatest difficulty in the application of Eq. (3.8) are the uncertainties of available theoretical (or experimental) ${}^3\Delta E_{i\to j}$:s. The roughest approximation is that ${}^3\Delta E_{i\to j}$ can be replaced with a "mean excitation energy", ΔE, which is typical of the molecule[19]. However, some quite important features of couplings were discovered by using this method[19,21]; for example, the features affecting the sign and magnitude of geminal couplings were characterized[21].

The second stage of the approximation is that ${}^3\Delta E_{i\to j}$ is set equal with $(e_j - e_i)$ in which $e_i$ is the eigenvalue of the one-electron hamiltonian, corresponding to the i:th molecular orbital. This approximation is shown to be quite realistic in many cases[22], and also to interpret the signs and trends in the geminal couplings[23]. Eq. (3.8) has also been applied to SCF wave functions, but the discussion on it is carried out in chapter 3.5.

Formulae corresponding to Eqs. (3.7) and (3.8) can also be derived for $J^{orb}$ and $J^{dip}$ (see Ref. 19).

## 3.3. The self consistent field molecular orbital theory

The general formula for the one-electron hamiltonian is

$$\hat{F} = -\frac{\hbar^2}{2m}\nabla_p^2 + \hat{V}(p) \qquad 3.9$$

in which the Hartree-Fock hamiltonian operator $\hat{F}_p$ is used instead of $\hat{H}_p$ if the potential energy operator $\hat{V}(p)$ also accounts for the interactions of the electron p and the other electrons of the molecule. $\nabla_p$ is the momentum operator for the electron p. $\hat{V}(p)$ depends on the other electrons of the molecule, and that is why the corresponding molecular orbitals must be known before $\hat{V}(p)$ can be defined exactly. In practice, $\hat{V}(p)$ can be estimated by using good independent electron calculations: after which the new improved wave functions and eigenvalues are recalculated by using the usual variation method:

$$\hat{F}\,\psi_i = e_i\,\psi_i \qquad 3.10$$

The solution of the eigenvalue problem leads to an improved $\hat{V}(p)$ which is applied in the following calculation. The iteration can be continued until the wave functions and eigenvalues are not significantly changed, or in other words, until the wave functions are self-consistent. (See also 2.3., the principle of program NMREN/NMRIT[12]).

If the electronic state of the molecule is described with a single Hartree product

$$\Psi(1,2,\ldots,n) = \psi_1(1)\alpha(1)\ \psi_1(2)\beta(2)\ldots\psi_n(n)\beta(n) \qquad 3.11$$

( $\alpha(i)$ describing the spin +1/2 electron states), one finds that it does not satisfy the Pauli's antisymmetry principle for fermions, for an interchange of a pair of electrons does not change the wave function (because the electrons are identical particles). That is why the total many-electron wave function must be described by the Slater determinant.

Applying the operator $\hat{F}$ for the total many-electron wave function and defining the one-electron wavefunctions orthogonal, one obtains the general formula of the total energy

$$E = \sum_{uv} P_{uv}H_{uv} + \frac{1}{2}\sum_{uvrs} P_{uv}P_{rs}\left[ <uv|rs> - \frac{1}{2}<ur|vs> \right] \qquad 3.12$$

in which $H_{uv}$ is a matrix element of the core hamiltonian (that is the hamiltonian describing the coulombic interactions of the electrons and nuclei), and $P_{uv}$ is a matrix element of the density matrix:

$$P_{uv} = 2 \sum_{i}^{occ.} c^*_{ui}c_{vi} \qquad 3.13$$

The application of the variation principle to Eq. (3.12) leads finally to the Roothan equations and to the standard eigenvalue problem also in this case:

$$\sum_{v} (F_{uv} - e_i S_{uv})c_{vi} = 0 \qquad 3.14$$

($S_{uv}$ is an overlap integral of atomic orbitals u and v). The matrix element of the Fock matrix (F) is

$$F_{uv} = H_{uv} + \sum_{rs} P_{rs}\left[<uv|rs> - \frac{1}{2}<ur|vs>\right] \qquad 3.15$$

$F_{uv}$ depends on the other orbitals through the elements $P_{rs}$, which are approximated, as the first guess, with the independent electron wave functions. The improving of $P_{rs}$:s leads to the iteration process sketched in the beginning of this chapter.

## 3.4. Semiempirical calculation methods

The Roothan equations can be solved accurately, ab initio, although the calculations become very cumbersome and demand a plenty of time if the molecule contains more than 4-5 heavy atoms. That is why many approximative methods are developed to imitate the ab initio calculations. At this moment, the most practical and popular calculations are carried out by using the CNDO/INDO stage approximation and the variations of it. However, quite a large number of ab initio calculations have been carried out for many small molecules. There are also variations in the ab initio methods: the number and quality of the atomic orbital basis can be varied in many ways.

In the semiempirical methods are some of the integrals appearing in the Eqs. (3.14) and (3.15) neglected or replaced with semiempirical parameters. For example, in the CNDO/2 method[24] are all the integrals $\langle uv|rs\rangle$ one neglected if $u \neq v$ and $r \neq s$, and the coulombic integrals between the orbitals of the atoms A and B are set independent of the types of the atomic orbitals. In the INDO method some exchange integrals $\langle uv|uv\rangle$ are accounted by using a semiempirical parametrization, to separate the different states arising from the same electronic configuration. This is important when unrestricted MO calculations are carried out and thus is important in the calculations of spin-spin couplings.

## 3.5. The calculation of spin-spin couplings from SCF wave functions

The Sum-Over-States method (Eq. 3.7) is valid also for SCF wave functions, if the excitation energies and the coefficients of the LCAO:s of the excited states can be calculated. Unfortunately, the complete or even an approximate account of the configuration interactions (CI:s) is difficult[25] and the increase in the calculation times makes them impractical if compared with the FPT method (cf. later). In addition, it has been found[26] that even a very complete account of the CI:s does not necessarily improve the results. However, it is shown that SOS+CI method allows the prediction of negative geminal couplings[27]. The simplest way to take some CI:s into account is to apply the first-order CI:s to the excitation energies, only[25]:

$$^{3}\Delta E_{i\rightarrow j} = e_j - e_i - \underline{J}_{ij} \qquad 3.16$$

where $e_j$ and $e_i$ are eigenvalues of the Fock matrix and $\underline{J}_{ij}$ is the bielectronic coulombic integral between the corresponding one electron wave functions. This methods was tested in our study on the couplings of hydroxyl protons,although without any success (see 7.4.).

However, in many cases like in the estimation of the direct and vicinal couplings,the account of CI:s is not necessary if the most important trends are wanted to be interpreted[22]. See also 7.3.-7.5.

In the SOS method the spin-spin couplings are calculated by using the standard formula of the second-order perturbation energy. The alternative method is to calculate the variations in the wave functions. The small mixing with the triplet states leads to small spin-densities in the electron cloud of the molecule. The calculation of these spin-densities can be carried out by a numerical method, called Finite Perturbation Theory (FPT)[28].

(i) The wave functions are described in the unrestricted way:

$$\psi_i^{\alpha} = \sum_u c_{ui}^{\alpha} \Phi_u$$

$$\psi_i^{\beta} = \sum_u c_{ui}^{\beta} \Phi_u \qquad 3.17$$

When there are no perturbations, the coefficients $c_{ui}^{\alpha} = c_{ui}^{\beta}$.

(ii) If the magnetic momenta of the nucleus B and, say, β-electron are opposite, the contact interaction between them leads to a lowered energy of β-electrons, and vice versa, to a higher energy of α-electrons. The energy change ($h_B$) depends on the s-nature of the orbital and can be written for one $s_B$-electron:

$$h_B = (8\pi/3)\ \mu_B\ s_B^2(0) \qquad 3.18$$

(where $\mu_B$ is the magnetic momentum of the nucleus A). In the SCF scheme this can be presented as $\pm h_B$ changes in the α- and β-core hamiltonian diagonal elements, corresponding to $s_B$-orbitals. This small perturbation induces the very small spin-densities around the other nuclei.

(iii) If the spin-density due to the $s_A$ orbitals is denoted by $\rho_{sA,sA}$ (= a diagonal element of the spin-density matrix), $J_{AB}^{cont}$ can be interpreted as the interaction of $\rho_{sA,sA}$ and the nucleus A:

$$J_{AB}^{cont} = \frac{\hbar}{2\pi}(2\mu_o g\beta/3)^2\ \gamma_A\gamma_B\ s_A^2(0)\ s_B^2(0)\ \frac{\rho_{sA,sA}(h_B)}{h_B} \qquad 3.19$$

In practice, the real $h_B$ is too small to be used in the computations, and be replaced with a more observable value.

However, the value of $h_B$ used must be small in order to prevent higher order defects in the results[28]. That is why the FPT calculations demand much more computer time than the calculations with the SOS method (without CI:s). In addition, only the couplings of one nucleus to all the others can be calculated in one run.

In the limit of the independent electron model FPT is identical with the SOS method[29]. When compared at the SCF stage, FPT contains CI:s implicitly, alhough it has been shown[30] that FPT cannot introduce singly-excited states into the perturbed wave functions without at the same time introducing the doubly-excited states in a restricted way. - Generally, FPT can be supposed to be more efficient than the SOS method without CI:s in the cases where the CI:s are important.

FPT is applied also using ab initio[31] and MINDO3[32] stage SCF calculations, although without any practical success when compared with the results by INDO stage calculations. In the chapter 7.5. it is shown that the standard INDO/FPT parametrization[24] cannot be improved to give practically more than about 40% more quantitative values of couplings, without making deeper changes in the INDO approximations.

FPT especially succeeds in the prediction of negative (often geminal) couplings[28,33], although the trends are not as good as in the case of the other type of the couplings[34-38]. These examples[34-38] also show that FPT is able to predict conformational and substitutional trends in the couplings: this is tested and applied in our study on the conformations of some aliphatic alcohols (see chapter 7.4.). FPT can also be used in the calculation of couplings induced by the orbital-dipole and spin-dipole mechanisms[38].

See also chapters 7.3.-7.5.

# 4. THE INTERPRETATION OF CHEMICAL SHIFTS

## 4.1. The perturbation approach

The spin-spin couplings arise from the interactions of nuclei and electrons, and thus their values are independent of the magnitude of the magnetic field. In contrast, a chemical shift can be interpreted as an interaction of the magnetic field and the nuclei via the electrons of the molecule. The most concrete explanation of the phenomenon is found in terms of the classical electricity theory: the magnetic field induces electric currents in the molecule, and thus creates new , often opposite, small local magnetic fields which interact with the nuclei. When the initial field is denoted with $B_0$ the shielding by the electrons can be described by the shielding constant $\sigma_i$:

$$B_i = (1 - \sigma_i)\, B_0 \qquad 4.1$$

in which i refers to the shielded nucleus. The shielding is usually given in the ppm:s of the original field and in respect to a standard shielding: this new parameters is the chemical shift $\delta_i$.

The theoretical model of the shielding phenomenon is obtained by mixing the ground state of the molecule with the singlet excited states. The perturbation is composed of two parts: from the interaction of the electron and the magnetic field, and from the interactions of the electrons and the dipolar field of the nucleus. These magnetic fields can be described in this case with vector potentials, generally of the form $\vec{B} \times \vec{r}$. If the vector potential due to the external field is $\vec{A}_{ex}$ and the field due to the nucleus is $\vec{A}_{nuc}$, the total hamiltonian of an electron in the magnetic field is:

$$\hat{H} = \frac{1}{2\,m}\,(\,\hat{p} + e\,\vec{A}_{ex} + e\,\vec{A}_{nuc})^{2} + \hat{V} \qquad 4.2$$

where $\hat{p}$ is the usual momentum operator ( $\hbar\nabla/i$ ). The application of this hamiltonian to the basic state of the molecule leads to two independent terms which are called the diamagnetic and paramagnetic shieldings ($\sigma^{d}$ and $\sigma^{p}$, respectively)[5,39,40,41]:

$$\sigma^{d} = \frac{e^{2}\mu_{o}}{12\,\pi\,m}\,\langle\frac{1}{r}\rangle \qquad 4.3$$

$$\sigma^{p} = -\,\frac{e^{2}\,\mu_{o}}{12\,\pi\,m^{2}}\,\sum_{n}\,\langle 0|\hat{L}|n\rangle \cdot \langle n|\frac{\hat{L}}{r^{3}}|0\rangle \,/\,{}^{1}\Delta E_{n\to 0} \qquad 4.4$$

($\hat{L}$ is the angular momentum operator).

By Applying Eq. (4.3) to LC AO:s a convenient formula of $\sigma^{d}$ is obtained [5,39,40]:

$$\sigma^{d}_{A} = \frac{e^{2}\,\mu_{o}}{12\pi\,m}\,\sum_{i}^{occ.}\,P_{ii}\,\langle i|\frac{1}{r_{iA}}|i\rangle \qquad 4.5$$

where the diagonal element of the density matrix P describes the electron density at orbital i. This term opposes the external field and can be interpreted as an electronic current induced in the orbitals. Its magnitude depends on the electronic density around the nucleus and it consists of two parts arising from the atom's own electrons and those of the other atoms. The diamagnetic part is supposed to be the most important term only in the case of the proton chemical shifts, although also the neighbour paramagnetic term may be important. In the case of the chemical shifts of the second and third row elements the diamagnetic term is less important because its variation range is usually very narrow when compared with that of the paramagnetic term[41,42]. Also the accounting of the substituent diamagnetism is easy because the effects are very additive[42].

Unfortunately, the application of the formula (4.4) to the standard LCAO:s leads to the results that depend on the origin used for the coordinates of the atoms[40,43]. That is why the LCAO:s must be formed from the so-called gauge-invariant atomic orbitals (GIAO:s)[40,44,45]:

$$\chi_u = \Phi_u \, e^{-\frac{ie}{\hbar c}\vec{A}_u \cdot \vec{r}} \qquad 4.6$$

where $\Phi_u$ is the usual AO and $\vec{A}_u$ is the vector potential $\frac{1}{2}\,\vec{B} \times \vec{R}_u$ in which $\vec{R}_u$ is the position vector of atom u. Using this kind of GIAO:s in the LCAO:s the following expression can be derived:

$$\sigma^p_A = -\frac{\mu_o\, e^2\hbar^2}{6\,\pi\, m^2} \langle r^{-3}\rangle_{2p} \sum_i^{occ.} \sum_j^{unocc.} \frac{1}{{}^1\Delta E_{i\to j}} \times$$

$$\begin{aligned}\big[\, &(c_{yAi}c_{zAj} - c_{zAi}c_{yAj}) \sum_B (c_{yBi}c_{zBj} - c_{zBi}c_{yBj}) + \\ &(c_{zAi}c_{xAj} - c_{xAi}c_{zAj}) \sum_B (c_{zBi}c_{xBj} - c_{xBi}c_{zBj}) + \\ &(c_{xAi}c_{yAj} - c_{yAi}c_{xAj}) \sum_B (c_{xBi}c_{yBj} - c_{yBi}c_{xBj})\,\big]\end{aligned} \qquad 4.7$$

where $c_{xAi}$ refers to the coefficient of $p_x$-atom orbital of atom A in the i:th LCAO. The second, third and fourth row of the equation define the diagonal elements of the shielding tensor ($\sigma_{xx}$, $\sigma_{yy}$ and $\sigma_{zz}$, respectively). $\sigma^p$ can be said to be due to the induced currents from one orbital to another, and the term is opposite to the diamagnetic term. The excitation energy is a singlet excitation energy and for SCF wave functions the value, with the first-order CI:s, is

$$^1\Delta E_{i\to j} = e_j - e_i - J_{ij} + 2\,K_{ij} \qquad 4.8$$

in which $K_{ij}$ is the bielectronic exchange integral[41]. The estimation of $^1\Delta E_{i\to j}$ is a serious difficulty, because even a small absolute inaccuracy can be serious when the absolute value of ${}^1\Delta E_{i \rightarrow j}$ is small, too. The calculations giving good results have been rare up to date. However, by using ab initio calculations with quite a large 4-31G basis , Ditchfield[45] has obtained reliable values for many small molecules. More recently, a serie of calculations have been published to show many experimental trends to be interpreted by using CNDO/S[46] or INDO/S approximation[41,47]. These calculations have been quite promising and it is obvious that they will become more popular in the near future.

Because of the lack of the realiable theoretical methods, many empirical relationships are developed to explain the correlations of molecular structures and chemical shifts. However, the most important trends can be explained on the basis of Eqs. (4.5) and (4.7). These trends are discussed in the next two chapters.

## 4.2. The mean excitation approximation

Also in the case of chemical shifts some practicall useful trends can be obtained when ${}^1\Delta E_{i \rightarrow j}$ in Eqs. (4.4) and (4.7) is replaced with a constant which is equal to the lowest $\sigma \rightarrow \sigma^*$ $\pi \rightarrow \pi^*$ transition energy of the molecule. The typical values of $\Delta E$ for alkanes, alkenes and ketones are 10, 8 and 7 eV[39,48], while the corresponding ${}^{13}C$ shifts are 0-50, 70-100 and about 200 ppm. However, the approximation is quantitatively useless, because the excitation energies can not be determined with sufficient accuracy.

The coefficient part of Eq. (4.7) can be developed into a qualitatively convenient expression by using the $\Delta E$ approximation[39]:

$$(\sigma_A)_{zz} = - \frac{\mu_o e^2 \hbar^2}{6\,\pi m^2} \langle r^{-3} \rangle \left[ (Q_{AA})_{zz} + \sum_{B \neq A} (Q_{AB})_{zz} \right] / \Delta E \qquad 4.9$$

where

$$(Q_{AA})_{zz} = 2-2(P_{xAxA} - 1)(P_{yAyA} - 1) + 2\,P^2_{xAyA} \qquad 4.10$$

$$(Q_{AB})zz = -2P_{xAxB}\,P_{yAyB} + 2P_{xAyB}\,P_{yAxB}$$

in which $P_{ij}$ is an element of the density matrix. $Q_{AA}$ represents the electron density at the nucleus and describes the current density/ $(1/\Delta E)$-unit. Correspondingly, $Q_{AB}$ describes the currents induced at the nucleus A by the paramagnetic currents at B. Thus $Q_{AB}$ depends on the neighbour atoms bonded to atom A and the number of the multiple bonds between them. $Q_{AB}$ is zero if B does not use p- or d-orbitals in the bondings. The variations in $Q_{AB}$ explain, for example, the high field $^{13}C$ chemical shift of acetylene and the differences in primary, secondary and tertiary carbons.

## 4.3. Electron density correlations

$\sigma^d$ depends on the electron density at the nucleus via term $<r^{-1}>$. A reduced electron density leads to smaller diamagnetism at the nucleus and to a low field shift which is, however, compensated partly by the shrinking of the atomic orbitals when the effective charge of the nucleus grows. The $<r^{-1}>$ variations are of importance in the case of the proton chemical shifts, although in the $^{13}C$ shifts the effects are hidden under the much greater changes in the paramagnetic $<r^{-3}>$ term.

$\sigma^p$ depends on the elecron density via two mechanisms:

(i) $Q_{AA}$, in Eq. (4.9) depends on the electron density. The smaller the electron density the smaller is the low field paramagnetic shift.

(ii) According to Karplus and Pople[39] the electron density dependence of $<r^{-3}>$ of 2p-orbitals can be presented by the expression

$$\langle r^{-3}\rangle_{2p} = 34.44\ a_o^{-3}\ (1 - 0.323\ q_A) \qquad 4.11$$

where $a_o$ is the Bohr radius of the 2p-orbital and $-eq_A$ is the net electronic charge on atom A.

The $\langle r^{-3}\rangle$ dependence is the most important part and a reduced electron density leads to low field shifts.

### 4.4. Substituent effects

The most valuable tool in the interpretation of $^{13}C$ spectra is the use of substituent effects, and some corrective terms for interactions of the substituents. Enormous numbers of papers have been published on the matter and at this moment the literature covers quite well all the common structure types[49].

The nature and additivity of the substituent effects are discussed in chapters 7.6. and 7.7.

## 5. COMPUTERIZED NMR: PROGRAMS FOR HIGH-RESOLUTION NMR SPECTROSCOPY

The purpose of this section is to describe the programs used or developed in this work, and which are available as Univac 1100 or PDP 11/70 versions. The listings of all the programs are available from the author.

### 5.1. The spectral analysis programs MAOCON-SC and NMR-LAOCN-4B

MAOCON-SC (see chapters 7.1. and 7.2) is a self-assigning version of LAOCOON3[13]. It also works in a mode in which the spectrum can be given in a digital form. The program is written for general spin systems of up to seven spins, and a version of it is also modified to PDP 11/70 computer (the modification demands quite an inconvenient segmentation of the program).

NMR-LAOCN-4B was expanded from NMR-LAOCN-4A[18]. The expansion demanded that some less important output sections were removed and two cards were added to account for the X approximation before the QCPE version worked correctly in the analysis of the coupled $^{13}C$ spectra (see chapter 7.4.)

The calculated spectra can be visualized by using line-diagram representation by the program PICASS, at a line printer; or by using the complete line-shape simulation by the program LORENZ.

## 5.2. The molecular geometry programs KOORDI, CHANGE and LANTA

KOORDI was written to change internal coordinates to cartesians. It contains the subroutine DEWAR from MINDO3[50]. KOORDI is used in the conversating mode.

CHANGE can be used to change the position of a molecule with repect to the coordinate axis. The new position of the molecule must be defined by giving the following information about the new position: (i) all three coordinates of one atom, (ii) two coordinates of the second atom, (iii) one coordinate of the third atom, and (iv) approximative coordinates of an additional atom. The program is based on the Newton-Raphson algorithm that seeks out the transformation matrix for the rotation by using the conditions given above and the orthogonality rules of the transformation matrix[51].

LANTA is a program meant for analysis of lanthanide shift data by using the dipolar equation[52]. It also works by the Newton-Raphson algorithm. The program is used in some conformational studies carried out in the University of Jyväskylä[53].

## 5.3. The molecular orbital programs

The programs MINDO3[50] and CNINDO[54] were first introduced to Univac 1108 system by Dr. Timo Nyrönen and Mr. Kallio. MINDO3 was used without any important modifications.

In order to improve the convenience of the calculations, CNINDO was provided by a mode in which it accepts internal coordinates. In addition, our version of CNINDO contains the following small modifications:

(i) The order in which the electrons are fed to the one-electrons orbitals after the first independent electron calculation can be defined in the input[55]. The same ordering subroutine is used also in FPT mode to ensure the correspondence of the $\alpha$- and $\beta$-wave functions; this mode is necessary when the couplings are desired to be specified to the different LCAO:s (see chapter 7.5.).

(ii) The parametrization of the method can be changed by the input.

(iii) The spin-spin coupling constants can be computed in the subroutine HESSU by using the SOS method. The same subroutine prints out, if wanted, the specification of the couplings to the occupied LCAO:s.

As presented in chapter 7.5., a "self-optimizing" version of CNINDO was also written. The program, called SUPER, contains CNINDO as a subroutine and the principle of the main program is explained in the chapter 7.5.

## 6. SUMMARY

The studies, of which this dissertation is composed, are reproduced in chapters 7.1.-7.7. The studies can be divided into three principal parts: (i) the developments in the methods of spectral analysis (see 7.1.-7.2.), (ii) the applications of the spectral analysis and quantum mechanical calculations to the conformational analysis of some hydroxyl substituted compounds (see 7.3.-7.4.), and (iii) a study on the efficiency of the substituent effect models in the interpretation of proton decoupled $^{13}C$ NMR spectra (see 7.6.-7.7.).

The uniqueness of the solution of the high resolution NMR spectra has always been a problem. The matter was studied and discussed in the first part of the dissertation. The problem was later (7.3.) also met in the analysis of the proton coupled $^{13}C$ NMR spectra: it is often difficult to assign the splittings of the spectra to the correct nuclei without decoupling experiments, which in this case are less attractive. However, if the spectrum is even slightly of the second-order, the best solution can often be decided on the basis of the suitability of the possible solutions. The program MAOCON-SC was developed to aid the assigning of the spectral lines, and thus make the checking of the possible solution as fast as possible. MAOCON-SC was developed during the spectral analysis reported in this dissertation.

MAOCON-SC and the extented version of NMR-LAOCN-4A[18] were applied to the spectral analysis of the spectra discussed in chapters 7.3.-7.4. In these studies also semiempirical INDO/CNDO calculations were tested and applied to the interpretation of the observed trends. In all the cases, the INDO/FPT method was found to allow the prediction of the conformational and substitutional dependences at least qualitatively. The INDO/SOS method was found to interpret the trends in other than geminal couplings. On the basis of the spectra and the molecular orbital calculations, it was possible to characterize the coupling behaviour of the hydroxyl proton. Generally, it was found that it greatly resembles that of $CH_3$ protons. On the basis of the experience obtained during the calculations, an attempt (7.5.) was made to find an improved parametrization of the INDO method as regards the calculations of spin-spin couplings. When a set of typical HH and CH couplings was tested, the optimization of the INDO/FPT and INDP/SOS methods improved the standard deviations between the experimental and observed couplings by 35-75%. Unfortunately, the smallest change was observed in the case of the INDO/FPT method. The optimization method developed in this study can be used in any further developments of the semiempirical MO methods.

In the third part of the studies the interpretation methods of the $^{13}C$ shifts were studied. It was found that, although the general correlation between the chemical shifts and electron densities is poor also in this case (33 chloro- and hydroxy substituted benzenes), there exists good correlation between the sums of the chemical shifts and the corresponding sums of the electron excesses at the substituents. The results also suggest that s-electrons have very pronounced effects on $^{13}C$ chemical shifts of p-electrons. This correlation also explains the different behaviour of the $\sigma$- and $\pi$-back-bonding substituents.

## REFERENCES

1. Tolansky, S. *"High Resolution Spectroscopy"*, Methuen, London 1947.
2. Pauli, W. *Naturwisschaften* **12**(1924)741.
3. Bloch, F., Hansen, W.W. and Packard, M., *Phys. Rev.* **69**(1946) 127; Purcell, E.M. Torrey, H.C. and Pound, R.V. *ibid*. **69**(1946) 37.
4. Proctor, W.G. and Yu, F.G., *Phys. Rev.* **81**(1951)20.
5. Ramsey, N.F., *Phys. Rev.* **85**(1952)60; *ibid.* **91**(1953)303.
6. Paasivirta, J. and Laatikainen, R., "Kemia X, NMR-spektrien analyysi ja tulkinta", University of Jyväskylä, Department of Chemistry, 1976.
7. Pople, J.A., Schneider, W.G. and Bernstein, H.J., *"High Resolution Nuclear Magnetic Resonance"* McGraw-Hill 1959, USA.
8. Corio, P.L. *"Structures of High-Resolution NMR Spectra"*, Academic Press, New York, 1966.
9. Heinzer, J., *J. Magn. Reson.* **26**(1977)301.
10. Abraham, R.J., *"Analysis of High-Resolution NMR Spectra"*, Elsevier Publishing Company, New York, 1971.
11. Haigh, C.W., in *"Annual Reports on NMR Spectroscopy"* (E.F. Mooney, Ed.), Vol. 4, p. 311, Academic Press, London, 1971.
12. Swalen, J.D. and Reilly, C.A., *J. Chem. Phys.* **34**(1962)295.
13. Castellano, S. and Bothner-By, A.A., *J. Chem. Phys.* **41**(1964) 3863; S. Castellano, *"Computer Programs for Chemistry"* (DeTar, D.F., Ed.), Vol. 1, W.A. Benjamin, New York, 1968.
14. Diehl, P., Sykora, S. and Vogt, J., *J. Magn. Reson.* **19**(1975) 67; Diehl, P. and Vogt, J., *Org. Magn. Reson.* **8**(1976)638.
15. Laatikainen, R., *J. Magn. Reson.* **27**(1977)169.
16. Arata, Y., Shimizu, H. and Fujiwara, S., *J. Chem. Phys.* **36**(1962) 1951; Yamamoto, T. and Fujiwara, S., *Bull. Chem. Soc. Jap.* **39** (1966)333.
17. Ferguson, R.C. and Marquardt, D.W., *J. Chem. Phys.* **41**(1964)2087.

18. Musso, J.A. and Isaia, A., *J. Chim. Phys.* **66**(1969)1676; Musso, J.A., *NMR-LAOCN-4A, QCPE 232.*

19. Pople, J.A. and Santry, D.P., *Mol. Phys.* **8**(1964)1.

20. Blizzard, A.C. and Santry, D.P., *J. Chem. Phys.* **55**(1971)950. Hirao, K., Nakatsuji, H., Kato, H. and Yonezawa, T., *J. Am. Chem. Soc.* **94**(1972)4078; Hansen, P.E., Berg, A. and Jakobsen, H.J., Manzara, A.P. and Michl, J., *Org. Magn. Reson.* **10**(1977) 179; Hansen, P.E., *ibid.* **11**(1978)215; Khin, T. and Webb, G.A., *ibid.* **10**(1977)175; *J. Magn. Reson.,* **33**(1979)159.

21. Pople, J.A. and Santry, D.P., *Mol. Phys.* **9**(1965)311.

22. Andre, J.M., Nagy, J.B., Derouane, E.G., Fripiat, J.G. and Vercauteren, D.P., *J. Magn. Reson.* **26**(1977)317; Joela, H., *Org. Magn. Reson.* **9**(1977)338.

23. Fahey, R.C., Graham, G.C. and Piccioni, R.L., *J. Am. Chem. Soc.* **88**(1966)193.

24. Pople, J.A. and Beveridge, D.L., *"Approximate Molecular Orbital Theory",* McGraw-Hill, New York, 1970; Murrell, J.N. and Harget, A.J., *"Semi-empirical Self-Consistent-Field Molecular Orbital Theory of Molecules",* Wiley-Interscience, Belfast, 1972.

25. Salem, L., *"The Molecular Orbital Theory of Conjugated Systems",* Benjamin, New York, 1966; Pople, J.A., *Proc. Phys. Soc. (London)* **A68**(1955)81.

26. Kowalewski, J., Roos, B., Siegbahn, P. and Vestin, R., *Chem. Phys.* **3**(1974)70; *ibid.* **9**(1975)29.

27. Loeve, P. and Salem, L., *J. Chem. Phys.* **43**(1965)3402; Ditchfield, R. and Murrell, J.N., *Mol. Phys.* **14**(1968)481; Ditchfield, R., *ibid.,* **17**(1969)33.

28. Pople, J.A., McIver, Jr., J.W. and Ostlund, N.S., *J. Chem. Phys.* **49**(1968)2960,2965.

29. Ellis, P.D. and Ditchfield, R., in *"Topics in Carbon-13 NMR Spectroscopy"* (Levy, G.C., Ed.), Vol. 2, Wiley-Interscience, USA, 1976.

30. Ditchfield, R., Ostlund, N.S., Murrell, J.N. and Turpin, M.A., *Mol. Phys.,* **18**(1970)433.

31. Ostlund, N.S., Newton, M.D., McIver, Jr., J.W. and Pople, J.A., *J. Magn. Reson.* **1**(1969)298.

32. Dewar, M.J.S., Landman, D., Suck, S.H. and Weiner, P.K., *J. Am. Chem. Soc.* **99**(1977)3951.

33. Maciel, G.E., McIver, Jr., J.W., Ostlund, N.S. and Pople, J.A., *J. Am. Chem. Sos.* **92**(1970)4151.

34. Maciel, G.E., McIver, Jr., J.W., Ostlund, N.S. and Pople, J.A., *J. Am. Chem. Soc.* **92**(1970)4497; Wasylishen, R. and Schaefer, T., *Can. J. Chem.* **51**(1973)961.

35. Maciel, G.E., McIver, Jr., J.W., Ostlund, N.S. and Pople, J.A., *J. Am. Chem. Soc.* **92**(1970)1.

36. Maciel, G.E., McIver, Jr., J.W., Ostlund, N.S. and Pople, J.A., *J. Am. Chem. Soc.* **92**(1970)4506.

37. Ernst, L., Wray, V., Chertkov, V.A. and Sergeyev, N.M., *J. Magn. Reson.* **25**(1977)123; Wray, V., Ernst, L. and Lustig, E., *ibid.* **27**(1977)1; Long, S.A.T. and Memory, J.D., *ibid.,* **29**(1978) 119.

38. Albright, T.A., *Org. Magn. Reson.* **8**(1976)489; Beer, M.D. and Grinter, R., *J. Magn. Reson.* **31**(1978)187.

39. Pople, J.A., *Disc. Faraday Soc.* **34**(1962)7; *Mol. Phys.* **7**(1964) 301.

40. Pople, J.A., *J. Chem. Phys.* **37**(1962)53,60.

41. Ebraheem. K.A.K. and Webb, G.A., *J. Magn. Reson.* **25**(1977)399; *ibid.,* **30**(1978)211; Jallali-Heravi, M. and Webb, G.A., *ibid.* **32**(1978)429; Ebraheem, K.A.K. and Webb, G.A., *Org. Magn. Reson.* **9**(1977)241,248; *ibid.* **10**(1977)70; Jallali-Heravi, M. and Webb, G.A., *ibid.* **11**(1978)524.

42. Grinter, R. and Mason, J., *J. Chem. Soc.* (*A*), (1970)2196; Mason, J., *ibid.* (1971)1038.

43. Amos, A.T. and Roberts, H.G.F., *J. Chem. Phys.* **50**(1969)2375; Arrighini, G.P., Maestro, M. and Moccia, R., *ibid.* **54**(1971)825.

44. Ellis, P.D., Maciel, G.E. and McIver, Jr., J.W., *J. Am. Chem. Soc.* **94**(1972)4069.

45. Ditchfield, R., *Mol. Phys.* **27**(1974)789.

46. Del Bene, J. and Jaffe, H.H., *J. Chem. Phys.* **48**(1968)1807; Ellis, R.L., Kuehnlenz, G. and Jaffe, H.H., *Theor. Chim. Acta* (Berl) **26**(1972)131.

47. Garber, A.R., Ellis, P.D., Seidman, K. and Schade, K., *J. Magn. Reson.* **34**(1979)1.

48. Karplus, M. and Pople, J.A., *J. Chem. Phys.* **38**(1963)2803.

49. See for example, Wehrli, F.W. and Wirthlin, T., *"Interpretation of Carbon-13 NMR Spectra"*, Heyden, Bristol, 1978; or Martin, G.J., Martin, M.L. and Odiot, S., *Org. Magn. Reson.* **7**(1975)1.

50. Dewar, M.J.S., Metiu, H., Student, P.J., Brown, A., Bingham, R.C., Lo, D.H., Ramsden, C.A., Kollmar, H., Weiner, P. and Bishof, P.K., *MINDO/3, QCPE* *279*.

51. Margenau, H. and Murphy, G.M., *"The Mathematics of Physics and Chemistry"*, van Nostrand, USA, 1968.

52. Wing, R.M., Uebel, J.J. and Andersen, K.K., *J. Am. Chem. Soc.* **95**(1973)6046.

53. Talvitie, A., *"Structure Determination of Some Sesquiterpenoids by Shift Reagent NMR"*, Department of Chemistry, University of Jyväskylä, Research Report Series, No. 6, 1979.

54. Dobosh, P.A. and Ostlund, N.S., *CNINDO/74, QCPE* *281*.

55. Pyykkö, P. and Joela, H., *Chem. Phys. Lett.* **31**(1975)574.

JOURNAL OF MAGNETIC RESONANCE 27, 169–180 (1977)

# About the Reliability, Efficiency, and Meaning of the Error Estimates of a LAOCOON3-Type NMR Analysis System

R. LAATIKAINEN

*Department of Chemistry, University of Jyväskylä, SF-40100 Jyväskylä 10, Finland*

Received December 13, 1976

https://doi.org/10.1016/0022-2364(77)90066-X

# An Automatically Assigning LAOCOON3-Type Algorithm in the Analysis of High-Resolution Proton Coupled $^{13}$C NMR Spectra

Reino Laatikainen
Department of Chemistry
University of Kuopio
P.O. Box 138
SF-70101 Kuopio
Finland

Abstract An algorithm that simulates a typical computer assisted analysis of NMR spectra is tested. The program seeks the self-consistent assignments of the spectral lines by assigning and checking the assignments of the observed lines during the run. The ability of the program to check also the analyzabilities of the spectral parameters ensures that the analysis can be started with very incomplete assignments, and without symmetry definitions. The speed of the convergence to self-consistent assignments, even from a bad guess, is found to be quite satisfactory. These properties are discussed in two examples. The algorithm is suitable especially for analysis of coupled $^{13}$C NMR spectra.

## INTRODUCTION

Only the $^{13}C$ part of proton-coupled $^{13}C$ NMR spectrum is normally observed. Thus the assignation of the couplings (relating the spectral splittings and real couplings between the nuclei in the molecule) is often impossible without decoupling experiments. And the time needed for obtaining a good coupled $^{13}C$ spectrum is usually long and therefore the decoupling experiments are not very attractive. On the other hand, in many $^{13}C$ NMR spectra there occur large $^{1}J_{CH}$ couplings that are able to induce very remarkable second-order effects on the spectrum. These second-order effects can be used in the assignation of the couplings. They also prevent an accurate first-order analysis of the couplings.

If a coupled $^{13}C$ NMR spectrum contains, say, three couplings that cannot be assigned with confidence, the only method to ensure the correctness of the result is to test all the six possible combinations of the assignments and then to decide if one of the solutions is the best. If the sign of a coupling is unknown, the number of the possible solutions is doubled.

In any case, the assignation of the observed lines is often a slow procedure, and the correction of input errors and wrong assignments demands a lot of time. In this report, an algorithm simulating the normal assignation-reassignation procedure is presented. The assignation, reassignation and the elimination of questionable assignments are based solely on the least square criteria, and the elimination procedure operates at the 90 or 95 % confidence level.

The idea of the automatic assignation was first presented by Keller & al. (1) in the program ASSIGN, and it was later developed by connecting the program with LAOCOON3 algorithm (2). ASSIGN is based on the general splitting rules, for example, on the regular quartets occuring in the general ABC spectrum. An assignation procedure is also connected to NMR-LAOCON-4A (3), but its algorithm does not contain any test of analyzabilities of parameters or reassignation and elimination procedures. However, these three aspects are necessary for a fluent process. To my knowledge, the usefulness of an automatical assignation procedure has not previously been discussed in detail.

A further solution to the problem of the troublesome assignations is the 'automatic NMR analysis' based on the direct analysis of the digital spectra (4,5,6). Anyway, there are problems also in these methods: practical difficulties in transferring the spectra from a spectrometer to a computer; slowness of the convergence if the trial parameters are not good; separation of extra signals (due, for example, to reflections in the FT spectrum when the spectrum is measured using a narrow spectral width); and noise phase, lineshape and double-resonance defects that reduce the reproducability of intensities of the spectrum. As to the testing of all the possible solutions, the situation in all three methods is the same: one must test them one by one.

## PROGRAM MAOCON-SC (SELF-CONSISTENT MAOCON (7))

The main structure of the assignation and reassignation procedures of MAOCON-SC are shown by the flow-diagram in Fig. 1. The program is modified from MAOCON (7) and still contains some routines in common with the original LAOCOON3 (8).

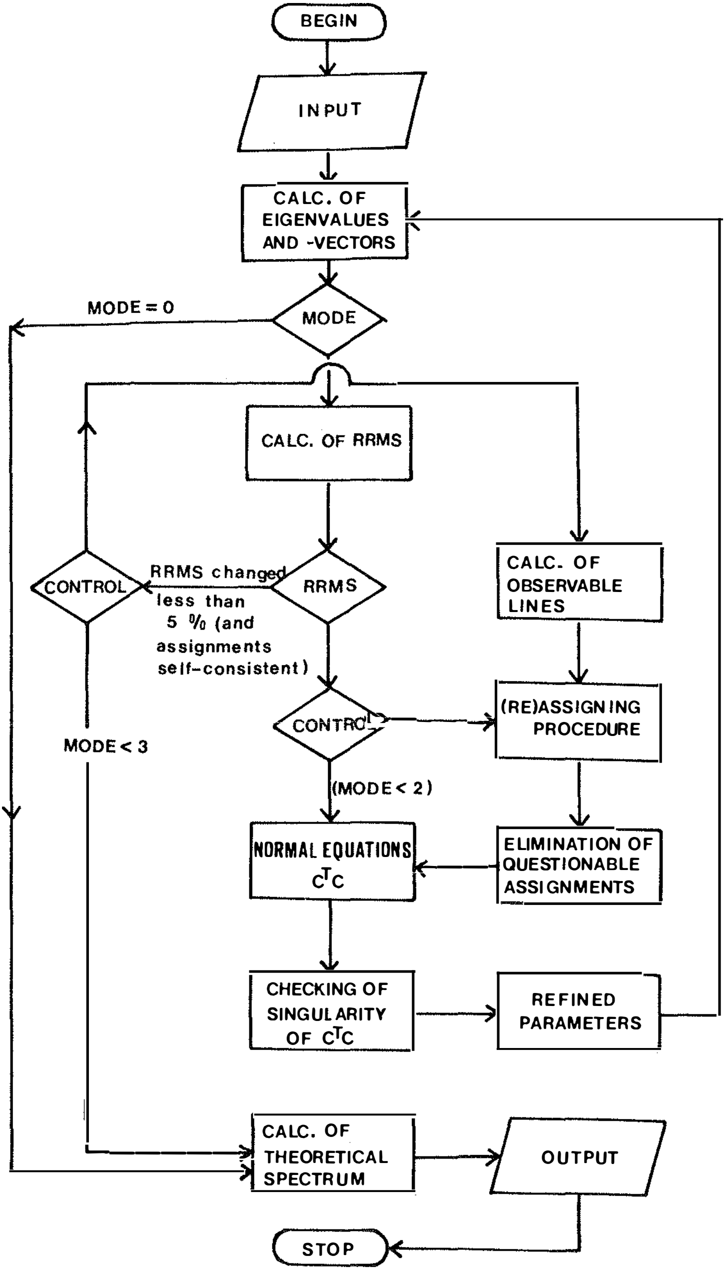


Fig. 1. The flow-diagram of MAOCON-SC.

The most remarkable modifications are:

i) The checking of analyzabilities of the spectral parameters.

ii) The checking of assignments and the elimination of questionable assignments.

iii) The assignation and reassignation procedure.

These modifications are discussed in more detail in the following chapters. Furthermore, the program can be used in 5 modes. For example, in mode 1, the program behaves like the classical LAOCOON3. In mode 4, all the modifications are applied and the condition equation corresponding to line $E_i^{calc}$ is weighed by a factor 1/N, in which N is the number of lines assigned to the observed signal to which also $E_i^{calc}$ is assigned; thus every observed signal has equal weight on the least square criteria. In mode 5, the program operates on digital spectra: the principle has previously been discussed by Heinzer (6).

Calculation times are reduced by using a reasonable diagonalization criterium (default = 0.05 + 0.2 times the smallest absolute non-zero value of the couplings) and calculating the differentials $\partial\lambda/\partial P_j$ only when one of the couplings or the differences of the chemical shifts is changed remarkably (default = 20 %, $\lambda_i$ = an eigenvalue). The latter approximation is strongly founded on the fact that the differentials depend on eigenvectors, and those are known to be quite independent of the order of the spectrum (9). When the refinements on the parameters are small, the eigenvalues can be estimated using the previously calculated differentials:

$$\lambda_i = \lambda_i^0 + \sum_{j=1}^{m} \frac{\partial\lambda_i}{\partial P_j} dP_j \qquad [1]$$

($\lambda_i^0$ is the old eigenvalue and $\lambda_i$ is the new one, after $P_j^0$:s are refined by $dP_j$:s). This approximation is very useful in the analysis of digital spectra, because in this case the non-linearity of the equations of conditions is almost completely due to non-linearity of the Lorenzian line-shape function. The approximation is also useful if the new parameters calculated after reassignation are not changed significantly. This approximation depends on the accuracy of the computer, and the recalculation of accurate eigenvalues should be carried out after the approximation is applied 3-10 times or a parameter is changed more than 20 %.

## Checking of the analyzability of spectral parameters

If a spectrum is independent of parameter $P_i$ or the assigned lines do not contain any information about it, the diagonal element $(C^TC)_{ii}$ of the normal equation matrix $C^TC$ is small. In this program, an extra equation

$$dP_i = 0.0 \qquad [2]$$

is added to the normal equations (by the procedure presented in the description of MAOCON (7)) if

$$\sqrt{(C^TC)_{ii}/(n-m)} \leqq \text{CRIT1} \qquad [3]$$

n = the number of the assigned lines and m = the number of the parameters to be optimized. The default value of criterium CRIT1 = 0.05. In other words, $P_i$ is not iterated if 1 Hz change in $P_i$ does not change the RRMS (Residual Root Mean Square) more than 0.05 Hz which is a typical RRMS value.

The internal correlation of $P_i$ and $P_j$ is defined by correlation coefficient $K_{ij}$:

$$K_{ij} = \frac{(C^TC)_{ij}}{\sqrt{(C^TC)_{ii}\ (C^TC)_{jj}}} \qquad [4]$$

In this algorithm, an extra equation (for $P_i<P_j$)

$$\frac{P_i}{P_j}\ (C^TC)_{ij}\ dP_i = (C^TC)_{ij}\ dP_j \qquad [5]$$

is added to the normal equations if

$$|K_{ij}| \geqq \mathrm{CRIT2} \qquad [6]$$

The default value of CRIT2 = 0.99. For example, if $K_{ij} = 1.0$ and $P_i = P_j$, $dP_i$ is constrained to be equal with $dP_j$. Coefficient $P_i/P_j$ prevents the change of the ratio of $P_i$ and $P_j$, and $(C^TC)_{ij}$ on the both sides of the equation is only a least-square constraint.

The extra equations remove the singularity of the normal equations if the assignments do not give any information about a parameter or the ratio of a couple of parameters.

## Checking of assignments and elimination of questionable assignments

After the first RRMS is calculated, the program seeks out the lines for which

$$|E_i^{obs} - E_i^{calc}| > 2 \times \mathrm{RRMS} \qquad [7]$$

and neglects them in the immediately following refinement of the parameters. The risk of eliminating a correct assignment is about 5 %. If there are no questionable assignments on the basis of this test, 5 % of the assignments will be eliminated, anyway. This procedure is necessary if there are several equally bad assignments. The criterium can be changed during the run; this is useful if the reassignation process appears to be slow.

At this stage, it is also possible to reassign the spectrum by seeking out the best possible signal for each theoretical line defined in the guess or in the assignation procedure. Checking, elimination and reassignation can be continued until the self-consistence of the assignments is obtained.

## Assignation and reassignation procedure

If a spectrum is incompletely assigned, the rest of the observed signals can be assigned automatically after the spectral parameters are optimized with respect to the incompletely assigned spectrum. The procedure seeks out the best possible observed signal for each observable line, if such a signal exists within limits $E_i^{calc} \pm$ CRIT3. The default value of CRIT3 is 1.0 Hz. This criterium is used if the spectrum is only partly observed, because of an impurity or some other disturbing signals.

The observed spectrum is given by format

IND(L), OBS(L), NUM(L).

IND(L) is the assignment of the observed signal at OBS(L) Hz; IND(L) is set to zero, if the assignment is not given. NUM(L) is the number of the lines which are allowed to be assigned to OBS(L);

default = 1000. If NUM(L) < 0, the line IND(L) is weighed in the normal equations by |NUM(L)| and NUM(L) is set 1. When NUM(L) = 1, the assignment of OBS(L) is not changed during the reassignations.

After the procedure described above, it is not sure at all that every signal observed is assigned: if the number of the reassignments in the previous process is zero or the reduction in the RRMS is less than 2 %, the program ensures that every signal observed is assigned to at least one theoretical line. The assigning of an observed signal is carried out if there is a theoretical line within ± CRIT4 (default: 2 x RRMS + 0.1 Hz).

## About programming

The program was written for 32 K memory of PDP 11/70, although the following calculations are partly carried out by a version for Univac 1100 system. The program is able to handle asymmetrical 7 spin spectra. The cpu-time of a single simulation of the ABCDEFX spectrum, presented in the following example, was 9 seconds in Univac 1100. When the diagonalization limit was reduced from the default value (about 0.1) to 0.001, the time was increased to 20 seconds: the results are, however, practically identical. The complete run, consisting of about ten iteration cycles, checking and reassignation procedures (cf. later), took about 70 seconds.

## AN EXAMPLE: THE PROTON COUPLED $^{13}$C NMR SPECTRUM OF CARBONYL CARBON OF 2-HYDROXY-BENZALDEHYDE

Fig. 2 presents stages of a typical spectral analysis carried out using MAOCON-SC. The spectrum derives from a study on the stereo-selectivity of the couplings of carbonyl and hydroxyl protons to aromatic carbons (<u>10</u>).

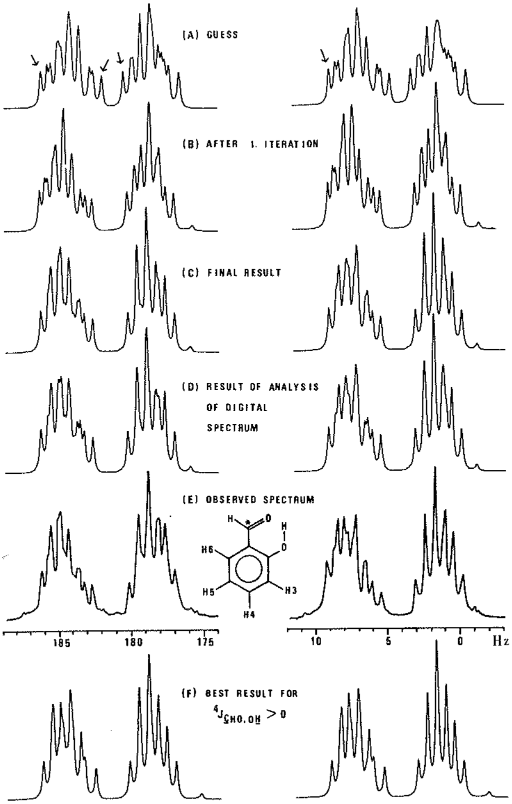


Fig. 2. The stages of the analysis of the proton coupled $^{13}C$ NMR spectrum of the carbonyl carbon of 2-hydroxy-benzaldehyde. (A) The guess: $^1J_{CH} = 177.0$, $^4J_{\underline{C}HO,O\underline{H}} = -0.7$, $^4J_{\underline{C}HO,H3} = 1.5$, $^5J_{\underline{C}HO,H4} = 0.4$, $^4J_{\underline{C}HO,H5} = 1.1$ and $^3J_{\underline{C}HO,H6} = 6.0$ Hz; the lines marked by arrows are assigned in the input. (B) The result of the first cycle: couplings are 177.04, -0.57, 1.23, 0.33, 0.90 and 6.40 Hz, respectively. (C) The final result: 177.02, -0.70, 1.32. 0.49, 0.52 and 6.39 Hz. (D) The result of the analysis of the digital spectrum: 177.02, -0.68, 1.33, 0.45, 0.56 and 6.42 Hz. (E) The observed spectrum (10): the data-point resolution is 0.061 Hz. (F) The best result with $^4J_{\underline{C}HO,O\underline{H}} > 0$: 177.05, 0.68, 1.27, 0.55, 0.49 and 6.38. The resolution used in the Lorenzian line-shape simulation was 0.23 Hz.

The X part of the ABCDEFX spectrum contains one large direct coupling of about 177 Hz, and one 6 Hz $^{3}J_{CH}$ coupling to 6-proton. The rest of the couplings are between 0.5 and 1.5 Hz. The small coupling of carbonyl proton to 3-proton and another of hydroxyl to 4-proton, together with the large direct coupling cause the second order nature of the spectrum. $^{4}J_{\underline{C}HO,O\underline{H}}$ can be estimated to be about ± 0.7 Hz, by using chemical decoupling with acid. If the sign of $^{4}J_{\underline{C}HO,O\underline{H}}$ is supposed to be unknown, the number of comparable solutions is 12. If one depends on INDO MO calculations and Pople-Santry equation for couplings (11,12), the most reliable guess is

$$^{4}J_{\underline{C}HO,H3} > {}^{4}J_{\underline{C}HO,H5} > {}^{5}J_{\underline{C}HO,H4}$$

Using the INDO MO values (1.5, 1.1 and 0.4 Hz, respectively) for these three couplings, $^{4}J_{\underline{C}HO,O\underline{H}} = -0.7$ and the first order values for the others, one obtains the guess presented in Fig. 2A. And one is certainly able to assign the lines marked with arrows; the rest of the 34 measurable absorption maxima or shoulders are given in input without indexes. If we try to iterate the chemical shift of the carbon and $^{n}J_{CH}$ couplings on the basis of this assignation, the checking procedure of parameter analyzabilities sets $dJ_{\underline{C}HO,H5} = 0.73\, dJ_{\underline{C}HO,H3}$, $dJ_{\underline{C}HO,H4} = 0.27\, dJ_{\underline{C}HO,H3}$ and $dJ_{\underline{C}HO,O\underline{H}} = -0.47\, dJ_{\underline{C}HO,H3}$ in order to prevent the divergence of RRMS. The result of the optimization of the parameters with respect to this incomplete assignation is presented in Fig. 2B.

The automatic assignation can now be carried out. The iteration process can be described by following the development of RRMS and the number of reassignations:

- After assigning 68 lines RRMS is 0.180 Hz and all the assignments are acceptable within risk of 5 %. However, about 5 % of the lines are always eliminated: after the elimination of the four most questionable assignments, RRMS is 0.163 Hz.
- After the first iteration, the reassignation of 5 lines and the elimination of 4 lines, RRMS is 0.143 Hz.
- During the 2nd - 7th iterations the total RRMS is reduced to 0.064 Hz, and the RRMS with 4 neglected assignments is 0.047 Hz. The numbers of the reassignments are 7 (RRMS is 0.120 Hz), 8 (0.086 Hz), 3(0.064), 1(0.054), 3(0.050) and 2(0.047), respectively.
- After the 8th refinement of the parameters the assignments are self-consistent. The method of assigning is now modified (and mode 4 applied): after the attempt of the reassignation, it is also ensured that every signal observed is assigned to at least one calculated line. The assignments are self-consistent also in respect of this criterium.
- After the 10th iteration, RRMS is completely converged. The total RRMS is 0.082 Hz and the RRMS with 4 neglected lines is 0.040 Hz. As to the neglected lines, three of them is hidden between the observed signals.

In the final result (Fig. 2C) the assigned lines are weighed with inverse number of the lines assigned to the corresponding observed signals. This mode is applied in connection with the modified method of assignation (in the present example, after the 7th iteration cycle). If this mode is not used, the iteration result is not as good as above because some of the observed signals are broad, containing several signals, and the lines assigned to these signals have an unreasonable weight in the least-square criterium.

The total cpu-time was only 65 seconds because the accurate calculation of the eigenvalues and eigenvectors was carried out only four times. Futhermore, the cpu-time can be reduced to about 55 seconds controlling the elimination and reassignation criteria during the run.

The iteration leads, in this case, to the correct solution although the order of the absolute trial parameters is not correct (the trial $|{}^4J_{\underline{C}HO,O\underline{H}}| < |{}^4J_{\underline{C}HO,H5}|$). This is due to the strong second-order nature of the spectrum. If the spectrum were of the first-order, the iteration would lead to a minimum, in which the order of the couplings is the same as in the guess.

In a case like this, it would be advisable to fix the parameters which are already exactly known. In this case, the fixing of ${}^4J_{\underline{C}HO,O\underline{H}}$ would be possible. It would also be sensible to assign at least one splitting corresponding to ${}^4J_{\underline{C}HO,O\underline{H}}$: this prevents it from changing before the automatic assigning procedure.

Fig. 2D presents the result when the spectrum is analysed by using the program in the digital mode. The result is quite similar to Fig. 2C and it is difficult to decide whether it is better than the result of the previous analysis. The differences in the optimized parameters are also quite meaningless (see text of Fig. 2). The time used in the digital mode was much longer although very good trial parameters were used.

The corresponding solution with ${}^4J_{\underline{C}HO,O\underline{H}} > 0$ can be obtained by using the iterated values for the rest of the parameters and carrying out the (automatic) assignation of the spectrum without iteration of the incompletely assigned spectrum. The best result obtained (RRMS = 0.055) is presented in Fig. 2F. The former solution with $^{4}J_{\underline{C}HO,O\underline{H}} < 0$ is clearly better: this is confirmed by inspection of the combination lines at -1.0 and 175.0 Hz. The other possible solutions may also be checked in the same way. The most interesting possible solution is the one in which $^{4}J_{\underline{C}HO,H5} < {}^{4}J_{\underline{C}HO,H4}$: however, both the reassignation method and the digital method always lead to the same result that is previously defined.

In general, the iteration always seems to lead to a minimum for which the RRMS is reasonable. For example, if the trial $^{5}J_{\underline{C}HO,H4}$ is less than 0.3 Hz and the rest of the parameters are the same as in Fig. 2A, the iteration leads to a local minimum (RRMS = 0.055) for which $^{5}J_{\underline{C}HO,H4} = -0.144$. Small variations of the trial parameters also lead to practically identical results, although deviations of 0.005 - 0.015 Hz may occur in the optimized couplings. This is in accordance to the 95 % confidence limits of the parameters which are 0.012 - 0.017 Hz.

## CONDITIONS OF THE CONVERGENCE

In order to give more concrete conditions of the convergence, a simple example is discussed: let us consider the X part of the ABCDEX spectrum shown in Fig. 3B. The spectral parameters are given in the text of the figure.

The most arbitrary guess for the spectrum is $J_{AX} = J_{BX} = J_{CX} = J_{DX} = J_{EX} = 5.0$ Hz. However, this guess (when the accurate parameters of the ABCDE part are supposed to be known) is sufficiently good for the assignation of the outmost lines of the multiplet. Using this parameter set as a guess, the program gives the result shown in Fig. 3C. The RRMS of the result is very good, although the result is

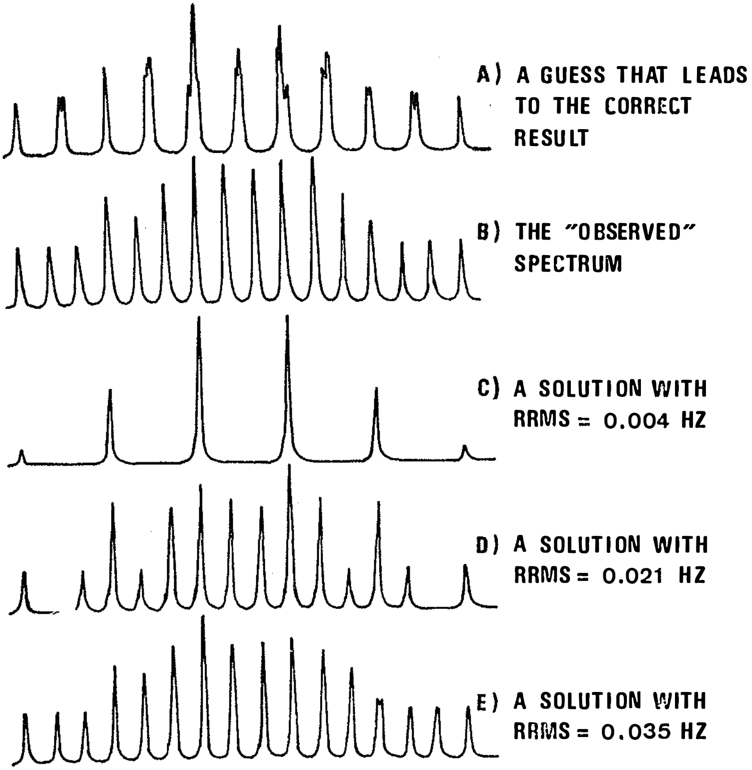


Fig. 3. The analysis of the (completely arbitary) ABCDEX spectrum. (A) A guess that leads to the exactly correct solution: the spectral parameters are given in the legend. (B) The X part of the spectrum: chemical shifts of the ABCDE part are 500, 520, 540, 560 and 580 Hz (for A, B, C, D and E, respectively) and the couplings are $J_{AX} = 2.00$, $J_{BX} = 4.00$, $J_{CX} = 6.00$, $J_{DX} = 8.00$, $J_{EX} = 10.00$, $J_{AB} = J_{BC} = J_{CD} = J_{DE} = 8.00$, $J_{AC} = J_{AE} = J_{BD} = J_{CE} = 3.00$, $J_{AD} = J_{BE} = 1.00$ Hz. (C) An iteration result: all the couplings of X are 6.00 Hz. (D) An iteration result: $J_{AX} = 3.90$, $J_{BX} = 6.07$, $J_{CX} = 5.97$, $J_{DX} = 5.99$ and $J_{EX} = 8.05$ Hz. (E) A solution with $J_{AX} = 4.06$, $J_{BX} = 1.83$, $J_{CX} = 6.05$, $J_{DX} = 8.00$ and $J_{EX} = 10.04$ Hz. The resolution used in the spectra is 0.2 Hz.

completely unreliable. By using guess $J_{AX} = 4$, $J_{BX} = 5$, $J_{CX} = 6$, $J_{DX} = 7$ and $J_{EX} = 8$ Hz, one obtains the result in Fig. 3D. There is a deficiency in these two results: every signal observed is not used in the assignments of the calculated lines. This can be ensured by using a non-default value for CRIT4. However, the use of this parameter does not necessarily lead to the correct result but results as in Fig. 3E can be obtained.

The correct result is obtained, for example, with the guess $J_{AX} = 2.9$, $J_{BX} = 3.1$, $J_{CX} = 6.0$, $J_{DX} = 7.1$ and $J_{EX} = 9.1$ Hz. This guess is shown in Fig. 3A. If trials $J_{AX}$ and $J_{BX}$ are interchanged the iteration leads to the result in Fig. 3E, in which also the values of $J_{AX}$ and $J_{BX}$ are reversed when compared with the correct result. To summarize, the program seeks the nearest possible solution.

If the ABCDE part were replaced with the corresponding pure first-order part, the X part would have 120 equally good solutions (if all the couplings are positive):

| $J_{AX}$ | $J_{BX}$ | $J_{CX}$ | $J_{DX}$ | $J_{EX}$ |
|---|---|---|---|---|
| 2 | 4 | 6 | 8 | 10 |
| 4 | 2 | 6 | 8 | 10 |
| 6 | 4 | 2 | 8 | 10, etc. |

However, in the second-order problem of Fig. 3B only the first combination gives the fit with RRMS = 0 Hz. The solution in Fig. 3E corresponds to the second first-order solution and has RRMS = 0.035 Hz. Supposing the reading inaccuracy of the lines is about 0.035 Hz the total RRMS is about 0.05 Hz ( $= (0.035^2 + 0.035^2)^{1/2}$), which is not notably bigger than the pure reading inaccuracy, and in practice, it may be difficult to decide if the solution as in Fig. 3E is wrong. The most remarkable differences between the two solutions are the small differences in the heights of the signals, due to the small second-order splittings.

## CONCLUSIONS

The simple algorithm described above works surprisingly well and the results are also comparable with the results of the direct analysis of the digital spectra. The trouble of assignation and manual reassignation is greatly avoided: in the first example it was necessary to give the program only the list of 34 observed signals and assign four of them (the total number of observable theoretical lines is about 70). And the first assignation was made on the basis of quite a bad guess. When compared with the more "modern" methods (4,5,6), this modification of the classical LAOCOON3 method does not suffer from the problems discussed in the introduction chapter. The approximations applied in the algorithm make also the computing times very reasonable.

Any method of analysis in the world will not remove the problem of the uniqueness of the solutions of NMR spectra (7). The problem is very serious also in the case of the coupled $^{13}C$ NMR spectra, as shown above. However, the method described above makes the checking of the possible solutions convenient if there is any possibility that the spectrum is sufficiently of the second-order to be analysed completely.

The program was also written to operate on digital spectra. This mode may be useful in the very final refinements of the result: for example, in the case of ABCDEX spectrum (in Fig. 3) it would be sensible to use the digital mode in testing the goodness of the solution in Fig. 3E.

## REFERENCES

1. W. D. Keller, T. R. Lusebrink and C. H. Sederholm, J. Chem. Phys. 44, 782 (1966).
2. P. Diehl, H. Kellerhals and E. Lusting, in "NMR, Basic Principles and Progress" (P. Diehl, E. Fluck and R. Kosfeld, Eds.), Vol. 6. p. 48, Springer, New York, 1972.
3. J. A. Musso and A. Isaia, J. Chim. Phys. 66, 1676 (1969); J. A. Musso, NMR-LAOCN-4A, QCPE 232.
4. P. Diehl, S. Sykora and J. Vogt. J. Magn. Resonance 19, 67, (1975).
5. P. Diehl and J. Vogt, Org. Magn. Resonance 8, 638 (1976).
6. J. Heinzer, J. Magn. Resonance 26, 301, (1977).
7. R. Laatikainen, J. Magn. Resonance 27, 169 (1977).
8. S. Castellano and A. A. Bothner-By, LAOCN3, QCPE 111.
9. J. D. Swalen and C. A. Reilly, J. Chem. Phys. 34 (1962) 295.
10. P. Äyräs, R. Laatikainen and S. Lötjönen, Org. Magn. Resonance 1979, in press.
11. J. A. Pople and D. P. Santry, Mol.Phys. 8 (1964) 1.
12. H. Joela, Org. Magn. Resonance 9, 338 (1977).

## AKNOWLEDGEMENTS

This work was supported by the Emil Aaltonen Foudation which is gratefully acknowledged. I am also very grateful to Phil. Lic. Simo Lötjönen and Associate Professor Pertti Äyräs for their co-operation.

# The $^{13}$C,H Coupling Constants in Structural and Conformational Analysis IV[x]. The Rotational Dependence of the Long-range CH Couplings of the Aldehyde and Hydroxyl Protons in Salicylaldehyde.

Pertti Äyräs,[xx] Reino Laatikainen[xxx] and Simo Lötjönen
Department of Chemistry
University of Kuopio
P.O. Box 138
SF-70101 Kuopio
Finland

https://doi.org/10.1002/mrc.1270130520

---

[x] For Part III, see Ref. 3.

[xx] Author to whom all correspondence should be addressed. Present address: Department of Chemistry, University of Turku, SF-20500 Turku, Finland

[xxx] Present address: Department of Chemistry, University of Jyväskylä, SF-40100 Jyväskylä, Finland

# An Experimental and Theoretical FPT/INDO Study on $^{n}J(O\underline{H},\underline{C})$, $^{n}J(O\underline{H},\underline{H})$, $^{n}J(CH)$ and $^{n}J(HH)$ Couplings of Methanol, Ethanol, 2-Propanol, tert-Butyl Alcohol and Acetylacetone

R. Laatikainen[x], S. Lötjönen and P. Äyräs[xx]
Department of Chemistry
University of Kuopio
P.O. Box 138
SF-70101 Kuopio
Finland

https://doi.org/10.3891/acta.chem.scand.34a-0249

x) Author to whom the correspondence should be addressed

xx) Present address: Department of Chemistry, University of Turku
SF-20500 Turku, Finland

# Optimized Parametrizations for the Calculations of the Spin-Spin Coupling Constants by INDO/FPT and INDO/SOS Methods: A comparison of the Methods

Reino Laatikainen[x]
Department of Chemistry
University of Jyväskylä
SF-40100 Jyväskylä 10
Finland

Abstract: An optimization method for the parametrizations of semi-empirical molecular orbital methods is developed and discussed. The algorithm is applied to the calculation of the spin-spin couplings by the INDO/FPT and INDO/SOS methods. However, the calculations show that the standard deviation between the observed and calculated HH and CH couplings of typical hydrocarbons can be improved only about 50% by optimizing only the parametrizations.

[x] The present address: University of Kuopio, P.O.B. 138, Kuopio 10, Finland

## INTRODUCTION

Two principles are used in calculations of spin-spin coupling constants in the terms of the molecular orbital theory: the sum-over-states (SOS) method [1] and the finite perturbation theory (FPT) [2]. In the SOS method, the second-order energy arising from the mixing of the ground state of the molecule with the excited states is calculated, when the mixing perturbation is the contact interaction. Shortly, the theory gives a convenient formula for the coupling [1]:

$$J_{AB} = K_{SOS}(AB)\ \pi_{AB} \tag{1}$$

in which $K_{SOS}(AB)$ is the constant depending on the types of the nuclei A and B, and $\pi_{AB}$ is the mutual polarizability of the s-orbitals of the atoms A and B:

$$-\pi_{AB} = 4 \sum_{i}^{occ.} \sum_{j}^{unocc.} c_{sAi}c_{sAj}c_{sBi}c_{sBj}/{}^{3}\Delta E_{i\rightarrow j} \tag{2}$$

where $c_{sAi}$ is the coefficient of atomic s-orbital $s_A$ in the i:th occupied molecular orbital, and ${}^{3}\Delta E_{i\rightarrow j}$ is the triplet excitation energy between the ground state and the excited state when the electron from the i:th occupied orbital is shifted to the j:th unoccupied orbital. The principal difficulty of the SOS method is the estimation of the excitation energies. In the case of Hückel type of calculations ${}^{3}\Delta E_{i\rightarrow j}$ can be, as the first approximation, set equal with the energy difference of the one-electron orbitals: ${}^{3}\Delta E_{i\rightarrow j} = e_j - e_i$. In the SCF orbitals, also the interactions of the electrons are taken into account, and the difference of these interactions in the excited and ground states should be included. The first-order correction on ${}^{3}\Delta E_{i\rightarrow j}$:s is easily obtained [3]:

$${}^{3}\Delta E_{i\rightarrow j} = e_j - e_i - J_{ij} \tag{3}$$

where $J_{ij}$ is the coulombic integral between the electrons in orbitals i and j. However, it is shown that reasonable results are obtained without any configuration interaction (CI) corrections [4] and furthermore, it is also shown that even a very complete inclusion of the configuration interactions does not lead to any practical improvements in the results [5].

In the FPT method the second-order perturbations on the wave functions are calculated [2]: the coupling constant can be connected with the spin-density $\rho_{sB,sB}$ at nucleus B when the mixing perturbation $h_A$ is the contact interaction of the electron and the magnetic momentum of nucelus A:

$$J_{AB} = K_{FPT}(AB)\ \rho_{sB,sB}(h_A) \qquad (4)$$

$K_{FPT}(AB)$ is a constant that depends on the types of A and B. The calculations must be carried out numerically [2] and demand much more computer capacity and time than the corresponding calculations by the SOS method.

The INDO method was mainly developed to imitate the ab initio calculations and especially to interpret the properties of the outer electrons of molecules [6]. Thus it is reasonable to believe that there exists a parametrization that would better interpret the spin-spin couplings, which are properties of the electrons in the energetically lowest orbitals. It would also be possible to modify some approximations made in the method; for example, to parametrize the s- and p-orbital Slater exponents separately ( as done in program MINDO3 [7]). However, the number of such possibilities is too large to be considered at this stage. The purpose of this work was to find out if there exists a convenient method to optimize the parametrization within reasonable computational efforts, and after that, to test the method on the optimizing of the parametrization in respect of SOS and FPT methods.

The simplest practical problem was chosen: the INDO parameters of hydrogen and carbon were optimized in respect of a typical set of couplings in common hydrocarbons. Methane, ethane, ethene, acetylene, cyclopropane and benzene were chosen to represent typical structures in hydrocarbons. The inclusion of also formaldehyde, methanol and acetaldehyde was considered, but the study of the internal correlations of the semiempirical parameters suggested that the using them was of no advantage. Furthermore, only pure hydrocarbons were studied, because the variations in the electronic densities around the nuclei are small in these compounds and thus have small effect on the sizes of the atomic orbitals [1]. For ethane, the non-averaged couplings were taken into account on the basis of the simple Karplus relationship [8]. In the case of the SOS method, also the proportionality constant $K_{SOS}(AB)$ was optimized.

## THE OPTIMIZATION METHOD

The problem was dealt with as a solution of a group of non-linear equations:

$$\begin{aligned} F_1(p_1,p_2,\ldots.p_m) &= J_1 \\ F_2(p_1,p_2,\ldots.p_m) &= J_2 \\ \ldots\ldots \\ F_n(p_1,p_2,\ldots.p_m) &= J_n \end{aligned} \qquad (5)$$

in which the calculated couplings $J_k$ are presented as functions ($F_k$, which, however, in this case are not explicit) of the semiempirical parameters ($p_i$) of the INDO method. The purpose was to find the best possible fit to the experimental couplings ($J_k^{obs}$), within the least squares criterium.

The problem was solved by using the Newton-Raphson iteration that is well-known for NMR spectroscopists from the LAOCOON3 method for spectral analysis [9,10]. The main principles of the method are (i) to guess the trial parameters ($p_i^o$) and, (ii) then to improve them until the best possible fit is obtained. The direction to which the parameters are changed can be determined from the partial differentials $\partial J_k/\partial p_i$

The first difficulty in the task is that the differentials can not be calculated directly from the wave functions of the molecule. In this case they must be calculated numerically: this means that the differentation of the couplings of one molecule in respect of parameters demands (n+1) complete INDO calculations, which in its turn demands a tremendous length of computer time. Furthermore, if the equations are far from linear, as they are in this case, the iteration does not converge quickly and the number of the iterations grows quickly if a complete convergation is desired.

Secondly, the most serious difficulty was that if the standard Newton-Raphson [9,10] was used, the iteration did not converge at all. This problem is due to the very bad internal correlations between the parameters. The internal correlation between the parameters $p_i$ and $p_j$ is described by the correlation coefficient $K_{ij}$:

$$K_{ij} = (D^TD)_{ij} / \sqrt{(D^TD)_{ii} (D^TD)_{jj}} \qquad (6)$$

in which $(D^TD)_{ij}$ is the i,j element of $D^TD$ matrix; D is the partial differential matrix defined, for example, in Table 1. If $K_{ij}$ is near to ±1, $p_i$ and $p_j$ are badly correlated internally and the same effects on the calculated couplings are observed if either $p_i$ or $p_j$ is changed. In practice, this means that the solution of the improvements ($dp_i$ and $dp_j$) on the parameters may give reasonable values for ($dp_i \pm dp_j$), but that the individual dp:s may be completely impractical. And this is a big problem in this case. Table 2 shows that the internal correlations in the case of the FPT method are often very near to $\pm 1.0$.

Finally, there was also the problem of the uniqueness of the solutions. In the case of FPT, two equally good minima were found, if the only criterium was RRMS. This is of course a very serious problem and it cannot be avoided: the best solution must be chosen on the basis of other criteria.

The problems in the calculation of the partial differentials and in the slow convergence were greatly reduced by using the following algorithms:

(i) The partial differentials were calculated only for every 5-10th cycle of iteration and, furthermore, the matrix was always stored in the mass memory of the computer. It was found that there is no necessity to recalculate the differentials if the parameters are not changed more than 20-30 % (except the Slater exponents). By using the stored matrix, it was possible to test the convergence and then to develop the following algorithms.

(ii) The absolute values of the correction steps $dp_i$ were maintained to be smaller than $(0.10 \times (|p_i| + 0.3)$, in order to prevent the divergence. In other words, a damping factor 0.001 -1.000 was applied.

(iii) In order to find the most reliable minimum, it is sensible to optimize the most doubtful parameters. In this case, it is sensible to optimize $\beta_H$ and $\beta_C$ first, because they are very inaccurately given in the original parametrization [6].

## PROGRAMMING AND CALCULATIONS

The calculations were carried out by UNIVAC 1100 computer. The program SUPER was composed from a self-written iteration algorithm and a subroutine modified from program CNINDO/74 [11]. The accurate geometries [6,12,13] and the litterature values of the couplings were used [6,13,14].

The differentation of the couplings given in Table 1 in respect of the INDO parametrization, and at the SOS level, demands about 6 cpu (central processing unit) minutes. One iteration cycle, without the recalculation of the differentials, demands about 50 cpu seconds. A reasonable convergence is obtained within 30-60 cpu minutes (15-30 iteration cycles). The times needed in the case of the FPT calculations are about five times longer. However, some useful trends can obtained by only iterating the parameters for 5-30 minutes (at the SOS and FPT levels, respectively), if the partial differentials are previously calculated and stored in the computer.

## RESULTS AND DISCUSSION

Table 1 contains the partial differentials for FPT couplings. The information of the table can be summarized in two paragraphs:

(i) The calculated couplings are not very dependent on the parametrization: a change of a parameter with 1% very seldom produces more than 0.1 Hz change in a coupling. The most dependent on the parametrization are the $^1J_{CH}$:s: thus they have the greatest influence on the result of the optimization. The largest

Table 1. The normalized partial differentials of the calculated couplings in respect of the semiempirical parameters of FPT/INDO method. The values of the differentials express the changes in the couplings if 1% changes are applied to the parameters.

| | | | | Normalized differentials (for $p_i$:s)[a] | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. | Obs. | Calc. | $r_H$ | $r_C$ | $F^2_C$ | $G^1_C$ | $\alpha_H$ | $\alpha_{s,C}$ | $\alpha_{p,C}$ | $\beta_H$ | $\beta_C$ |
| Methane (ref. 6) | | | | | | | | | | | | |
| $^2J_{HH}$ | 1 | -12.40 | -6.13 | 0.016 | -0.200 | -0.060 | -0.035 | 0.033 | 0.197 | -0.096 | 0.042 | 0.062 |
| $^1J_{CH}$ | 2 | 125.00 | 122.91 | 1.596 | 2.118 | 0.055 | 0.301 | 0.018 | -0.160 | 0.029 | -0.618 | -1.518 |
| Ethane (ref. 6) | | | | | | | | | | | | |
| $^2J_{HH}$ | 3 | -13.00 | -6.53 | -0.032 | -0.178 | -0.068 | -0.034 | 0.032 | 0.251 | -0.117 | 0.044 | 0.058 |
| $^3J_{HH}$(av.) | 4 | 8.00 | 9.73 | 0.163 | -0.167 | 0.009 | 0.016 | 0.017 | 0.002 | -0.016 | -0.123 | -0.079 |
| $^3J_{HH}$(anti) | 5 | 16.00 | 21.77 | 0.369 | -0.301 | 0.006 | 0.040 | 0.042 | 0.133 | -0.085 | -0.273 | -0.225 |
| $^3J_{HH}$(gauche) | 6 | 4.00 | 3.72 | 0.060 | -0.100 | 0.011 | 0.004 | 0.004 | -0.063 | 0.019 | -0.048 | -0.006 |
| $^1J_{CH}$ | 7 | 125.00 | 122.19 | 1.462 | 2.679 | 0.061 | 0.301 | -0.114 | -0.073 | 0.097 | -0.653 | -1.571 |
| $^2J_{CH}$ | 8 | -4.50 | -7.84 | -0.175 | -0.065 | -0.035 | -0.042 | -0.004 | 0.103 | -0.033 | 0.087 | 0.122 |
| Ethene (ref. 6) | | | | | | | | | | | | |
| $^2J_{HH}$ | 9 | 2.30 | 3.76 | 0.039 | 0.190 | -0.066 | -0.023 | -0.006 | 0.490 | -0.180 | -0.052 | -0.169 |
| $^3J_{HH}$(cis) | 10 | 11.50 | 9.38 | 0.096 | -0.040 | 0.024 | 0.011 | -0.003 | -0.159 | -0.057 | -0.101 | -0.019 |
| $^3J_{HH}$(trans) | 11 | 19.10 | 25.75 | 0.399 | 0.014 | 0.014 | 0.073 | 0.046 | -0.229 | -0.123 | -0.314 | -0.349 |
| $^1J_{CH}$ | 12 | 156.20 | 159.48 | 1.685 | 3.774 | 0.074 | 0.325 | -0.131 | -0.077 | 0.107 | -0.773 | -1.981 |
| $^2J_{CH}$ | 13 | -2.40 | -11.92 | -0.248 | -0.541 | -0.062 | -0.094 | 0.017 | 0.155 | -0.067 | 0.135 | 0.237 |
| Acetylene (ref. 6) | | | | | | | | | | | | |
| $^2J_{HH}$ | 14 | 9.80 | 10.63 | 0.234 | 0.147 | 0.002 | 0.073 | 0.033 | 0.262 | -0.118 | -0.123 | -0.271 |
| $^1J_{CH}$ | 15 | 248.70 | 233.00 | 2.031 | 5.723 | 0.049 | 0.325 | -0.119 | 0.017 | 0.060 | -0.106 | -2.589 |
| $^2J_{CH}$ | 16 | 49.70 | 2.42 | -0.052 | -0.154 | -0.057 | -0.058 | 0.059 | 0.314 | -0.161 | 0.027 | -0.088 |
| Cyclopropane (ref. 13) | | | | | | | | | | | | |
| $^2J_{HH}$ | 17 | -4.34 | 0.92 | 0.006 | 0.277 | -0.067 | -0.035 | -0.001 | 0.491 | -0.184 | -0.031 | -0.151 |
| $^3J_{HH}$(cis) | 18 | 8.97 | 8.39 | 0.189 | -0.220 | 0.007 | 0.008 | 0.026 | -0.050 | -0.005 | -0.107 | -0.032 |
| $^3J_{HH}$(trans) | 19 | 5.58 | 7.86 | 0.111 | -0.011 | 0.004 | 0.016 | 0.007 | 0.030 | -0.018 | -0.091 | -0.086 |
| $^1J_{CH}$ | 20 | 160.45 | 165.96 | 1.832 | 3.940 | 0.045 | 0.329 | -0.204 | 0.108 | 0.086 | -0.842 | 2.209 |
| $^2J_{CH}$ | 21 | -2.60 | -5.77 | -0.154 | -0.099 | -0.023 | -0.033 | -0.010 | 0.045 | -0.008 | 0.065 | 0.098 |
| Benzene (refs. 12 and 14) | | | | | | | | | | | | |
| $^3J_{HH}$ | 22 | 7.54 | 8.28 | 0.220 | -0.316 | 0.012 | -0.002 | 0.036 | -0.139 | 0.013 | -0.107 | 0.004 |
| $^4J_{HH}$ | 23 | 1.37 | 2.16 | 0.064 | -0.122 | -0.013 | 0.013 | 0.015 | 0.130 | -0.049 | -0.031 | -0.052 |
| $^5J_{HH}$ | 24 | 0.69 | 1.17 | -0.002 | 0.107 | 0.008 | 0.004 | -0.006 | 0.014 | -0.007 | -0.012 | -0.039 |
| $^1J_{CH}$ | 25 | 158.43 | 141.00 | 0.892 | 5.028 | 0.068 | 0.230 | -0.428 | 0.036 | 0.266 | -0.612 | -1.859 |
| $^2J_{CH}$ | 26 | 1.15 | -5.01 | -0.106 | -0.173 | -0.049 | -0.031 | -0.033 | 0.248 | -0.045 | 0.043 | 0.073 |
| $^3J_{CH}$ | 27 | 7.62 | 9.49 | 0.077 | 0.255 | 0.018 | 0.040 | 0.028 | 0.070 | -0.070 | -0.055 | -0.235 |
| $^4J_{CH}$ | 28 | -1.31 | -2.31 | -0.019 | -0.327 | -0.021 | -0.018 | 0.003 | 0.014 | 0.016 | 0.017 | 0.106 |

[a] $\alpha_u = (I_u + A_u)/2$, where $I_u$ is the ionization potential of orbital u and $A_u$ is the corresponding electron affinity |6|.

deviation between the observed and calculated values in coupling No. 16. It is also very independent of the parametrization, and its refinement would demand a very large refinement in the parametrization. That is why it was neglected in the refining of the parameters.

(ii) The internal correlations of the parameters appear also from this table: for example, the differentials of $\beta_H$ and $\beta_C$ are of the same sign in most cases and the absolute value of the former is significantly smaller. As seen from Table 2, the calculated correlation coefficient of the parameters is 0.98.

Table II contains the internal correlations of the parameters in the case of the FPT method and the present data. The table reveals, for example, that $\beta_C$ correlates badly with many other parameters. In practice, this means that in order to test the possibilities to improve the model, it is only necessary to optimize $\beta_C$ and few other parameters which are not correlated with it. Table 3 reveals that there are not so many bad correlations in the case of the SOS method, and this is why the convergence is expected to be much better if all the parameters are simultanously iterated.

The inspection of Tables 1 and 2 reveals still a very interesting bit of information: the couplings depend on the parametrization in these two methods in very different ways. For example, the correlations between $r_H$ and $r_C$ are 0.91 and -0.55, in FPT and SOS methods, respectively. This means that if an improvement in the fit of the FPT method is obtained by changing $(r_H+r_C)$, the improvement in the SOS method is observed when $(r_H-r_C)$ is changed. In order to show the nature of the difference more deeply, Table 4 was included in this report. This table contains the specifications of the couplings of methanol to the individual (occupied) orbitals.

Table 2 The internal correlations of the semiempirical parameters used in FPT/INDO method.

| | $r_H$ | $r_C$ | $F^2$ | $G^1$ | $\alpha_H$ | $\alpha_{sC}$ | $\alpha_{pC}$ | $\beta_H$ | $\beta_C$ |
|---|---|---|---|---|---|---|---|---|---|
| $r_H$ | 1.00 | 0.91 | 0.64 | 0.98 | -0.60 | -0.04 | 0.38 | -0.99 | -0.98 |
| $r_C$ | 0.91 | 1.00 | 0.62 | 0.92 | -0.83 | -0.01 | 0.52 | -0.93 | -0.97 |
| $F^2$ | 0.64 | 0.62 | 1.00 | 0.74 | -0.53 | -0.65 | 0.78 | -0.65 | -0.62 |
| $G^1$ | 0.98 | 0.92 | 0.74 | 1.00 | -0.66 | -0.12 | 0.48 | -0.97 | -0.97 |
| $\alpha_H$ | -0.60 | -0.83 | -0.53 | -0.66 | 1.00 | 0.04 | -0.69 | 0.65 | 0.73 |
| $\alpha_{sC}$ | -0.04 | -0.01 | -0.65 | -0.12 | 0.04 | 1.00 | -0.74 | 0.00 | -0.02 |
| $\alpha_{pC}$ | 0.38 | 0.52 | 0.78 | 0.48 | -0.69 | -0.74 | 1.00 | -0.38 | -0.42 |
| $\beta_H$ | -0.99 | -0.93 | -0.65 | -0.97 | 0.65 | 0.00 | -0.38 | 1.00 | 0.98 |
| $\beta_C$ | -0.98 | -0.97 | -0.62 | -0.97 | 0.74 | -0.02 | -0.42 | 0.98 | 1.00 |

Table 3 The internal correlations of the semiempirical parameters used in SOS/INDO method.

| | $r_H$ | $r_C$ | $F^2$ | $G^1$ | $\alpha_H$ | $\alpha_{sC}$ | $\alpha_{pC}$ | $\beta_H$ | $\beta_C$ |
|---|---|---|---|---|---|---|---|---|---|
| $r_H$ | 1.00 | -0.55 | -0.54 | 0.54 | 0.78 | -0.03 | -0.61 | 0.05 | 0.24 |
| $r_C$ | -0.55 | 1.00 | 0.43 | -0.53 | -0.90 | 0.10 | 0.63 | -0.75 | -0.93 |
| $F^2$ | -0.54 | 0.43 | 1.00 | -0.89 | -0.59 | -0.73 | 0.91 | -0.07 | -0.23 |
| $G^1$ | -0.54 | -0.53 | -0.89 | 1.00 | 0.73 | 0.68 | -0.98 | 0.17 | 0.37 |
| $\alpha_H$ | 0.78 | -0.90 | -0.59 | 0.73 | 1.00 | 0.02 | -0.80 | 0.49 | 0.73 |
| $\alpha_{sC}$ | -0.03 | 0.10 | -0.73 | 0.68 | 0.02 | 1.00 | -0.62 | -0.17 | -0.14 |
| $\alpha_{pC}$ | -0.61 | 0.63 | 0.91 | -0.98 | -0.80 | -0.62 | 1.00 | -0.24 | -0.45 |
| $\beta_H$ | 0.05 | -0.75 | -0.07 | 0.17 | 0.49 | -0.17 | -0.24 | 1.00 | 0.91 |
| $\beta_C$ | 0.24 | -0.93 | -0.23 | 0.37 | 0.73 | -0.14 | -0.45 | 0.91 | 1.00 |

The results show that it is very difficult to interpret the FPT results in similar terms of electron excitations, as it is often done in the case of the SOS method [15]. Furthermore, the SOS method is often claimed to suffer from the fact that the total couplings are the sums of very many comparable terms of opposite signs [6]. Table 4 shows that the problem is at least as serious in the FPT method.

During the study it appeared that it would be interesting to carry out the four following different attempts on the parametrization: to optimize the parameters in respect of

(I) FPT method,
(II) SOS method,
(III) SOS method, to reproduce the absolute values of the couplings correct, and
(IV) SOS method, reproduce a good fit with other than geminal couplings.

Tables 5 and 6 contain the optimized parameter sets and the couplings calculated by the methods. The goodness of the fits are described by standard deviations between the observed and calculated values, and by RRMS (Residual Root Mean Square = the sum of the squares/(the number of the couplings - the number of the optimized parameters)). RRMS represents the estimate within which the new parametrization predicts the new values of the couplings.

The results show that in all the cases some improvements are obtained. The greatest difficulties appear to be, expectedly in the geminal couplings: in the case of the SOS method, no practical improvements are obtained in them. In case I, the new geminal couplings are clearly better than before. As mentioned before, in case I

Table 4 The portions (Hz) due to the individual orbitals in the cases of ${}^{2}J_{O\underline{H},\underline{C}}$ and ${}^{3}J_{O\underline{H},\underline{H}}$ of methanol.[a]

| | ${}^{2}J_{O\underline{H},\underline{C}}$ | | ${}^{3}J_{O\underline{H},\underline{H}}$ | |
|---|---|---|---|---|
| Orbital No. | FPT | SOS | FPT | SOS |
| 1 | 1.30 | 0.27 | 0.60 | 2.79 |
| 2 | -6.22 | 0.55 | 1.81 | -7.56 |
| 3 | -0.83 | -0.13 | 28.29 | -10.67 |
| 4[b] | 0.00 | 0.00 | 0.00 | 0.00 |
| 5 | 1.68 | 0.20 | -54.70 | 12.70 |
| 6 | -3.39 | -0.36 | 40.71 | 15.12 |
| 7[b] | 0.00 | 0.00 | 0.00 | 0.00 |
| Total (Hz) | -7.41 | 0.52 | 16.59 | 12.38 |
| Observed[a] (Hz) | -3.00 | -3.00 | 10.25 | 10.25 |

[a] Ref.[17].

[b] π-type of orbitals.

Table 5 The original [6] and optimized sets of the INDO parameters of hydrogen and carbon.

| | | Optimized | | | |
|---|---|---|---|---|---|
| | Original | I | II | III | IV |
| $r_H$ | 1.2000 | 1.2411 | 1.0064 | 1.2119 | 1.5167 |
| $r_C$ | 1.6250 | 1.6749 | 1.4840 | 1.4600 | 1.4334 |
| $F_C^2$ | 0.1737 | 0.2359 | -0.3476 | 0.0071 | 0.2198 |
| $G_C^1$ | 0.2677 | 0.0266 | 0.0200 | -0.0148 | -0.2225 |
| $\alpha_H$ | 7.1761 | 4.8786 | -3.7092 | 0.1316 | 0.3203 |
| $\alpha_{s,C}$ | 14.0510 | 12.4620 | 9.7445 | 14.2678 | 16.2216 |
| $\alpha_{s,C}$ | 5.5720[a] | 5.5720 | 5.5720 | 5.5720 | 5.5720 |
| $\beta_H$ | 9.0000 | 8.7493 | 3.2444 | 6.5183 | 6.5263 |
| $\beta_C$ | 21.0000 | 19.8117 | 18.4127 | 18.4033 | 16.8821 |
| $K_{SOS}$(HH) | - | - | 582.1 | 450.0 | 449.8 |
| $K_{SOS}$(CH) | - | - | 780.5 | 803.3 | 832.1 |

[a] $\alpha_{s,C}$ was not optimized: it was found during the iteration that one of the α-parameters must be fixed, or the iteration is diverged (this can be explained with the role of α:s in the diagonals of the Fock matrix).

Table 6 The observed, non-optimized and by the optimized parametrizations calculated couplings.

| | | calculated | | optimized | | | |
|---|---|---|---|---|---|---|---|
| No.[a] | Obs. | FPT | SOS | I | II | III | IV |
| 1 | -12.40 | -6.13 | 2.16 | -8.67 | -1.19 | 0.12 | - |
| 2 | 125.00 | 122.91 | 123.36 | 116.73 | 123.43 | 122.37 | 125.53 |
| 3 | -13.00 | -6.53 | 2.32 | -10.48 | -2.64 | 0.67 | - |
| 4 | 8.00 | 9.73 | 7.16 | 8.08 | 8.06 | 7.25 | 7.71 |
| 5 | 16.00 | 21.77 | 16.57 | 16.61 | 18.01 | 15.78 | 16.29 |
| 6 | 4.00 | 3.72 | 2.46 | 3.82 | 3.08 | 2.99 | 3.43 |
| 7 | 125.00 | 122.19 | 116.15 | 119.55 | 121.98 | 120.67 | 121.12 |
| 8 | -4.50 | -7.84 | 0.91 | -7.26 | -0.10 | 0.07 | - |
| 9 | 2.30 | 3.76 | 9.14 | -0.52 | 0.49 | 2.30 | - |
| 10 | 11.50 | 9.38 | 7.86 | 10.44 | 11.15 | 10.21 | 11.15 |
| 11 | 19.10 | 25.75 | 20.21 | 19.22 | 20.43 | 17.97 | 18.74 |
| 12 | 156.20 | 159.48 | 161.89 | 162.04 | 158.56 | 160.17 | 157.73 |
| 13 | -2.40 | -11.92 | 2.09 | -9.95 | 0.22 | 0.27 | - |
| 14 | 9.80 | 10.63 | 7.95 | 4.62 | 3.84 | 6.27 | 9.11 |
| 15 | 248.70 | 233.00 | 261.22 | 246.28 | 249.53 | 250.14 | 248.58 |
| 16[b] | 48.70 | 2.42 | - | - | - | - | - |
| 17 | -4.34 | 0.92 | 6.49 | -2.97 | -0.99 | 7.28 | - |
| 18 | 8.97 | 8.39 | 7.10 | 7.53 | 7.28 | 7.43 | 8.81 |
| 19 | 5.58 | 7.86 | 6.52 | 6.63 | 7.26 | 6.31 | 6.73 |
| 20 | 160.45 | 165.96 | 157.92 | 167.91 | 159.60 | 162.59 | 161.63 |
| 21 | -2.60 | -5.77 | 0.23 | -4.97 | -0.07 | 0.07 | - |
| 22 | 7.54 | 8.28 | 7.20 | 8.00 | 8.75 | 7.82 | 8.33 |
| 23 | 1.37 | 2.16 | 2.06 | 0.23 | 0.28 | 1.60 | 2.69 |
| 24 | 0.69 | 1.17 | 0.52 | 1.01 | 1.06 | 0.66 | 0.39 |
| 25 | 158.43 | 141.00 | 140.24 | 156.28 | 159.25 | 155.25 | 158.50 |
| 26 | 1.15 | -5.01 | 2.10 | -4.67 | 0.32 | 3.37 | - |
| 27 | 7.62 | 9.49 | 6.35 | 5.65 | 3.75 | 5.31 | 5.61 |
| 28 | -1.31 | -2.31 | 0.08 | -0.98 | 0.25 | 0.49 | 1.06 |
| Std. dev.[b] | | 5.97 | 5.86 | 3.69 | 3.69 | 4.20 | 1.33[c] |
| RRMS [b] | | 7.11 | 7.00 | 4.41 | 4.40 | 5.01 | 1.75[c] |

---

[a] See Table 1. [b] Observation No. 16 is not included in optimizations

[c] The original RRMS was 7.60 Hz

also another minimum was found, when the iteration was begun without a preliminary optimization of $\beta_H$, $\beta_C$, $F_C^2$ and $G_C^1$. However, the new parameters were very different from the originals and the improvement of the fit was completely due to improvements in the other than the geminal couplings.

## CONCLUSIONS

The results can be summarized in the following points:

(i) There exists a method to be used in the developments of more sophisticated and quantitative semiempirical models for the calculation of spin-spin couplings. And this can be done with reasonable computational efforts, if compared, for example, with typical MINDO3 and ab initio calculations.

(ii) The original INDO parametrization [6] can be improved to give a fit, from 35% (case I) to 75% (case IV) better, between the observed and calculated values (if the decrease in the RRMS value or standard deviation is considered). The greatest problems are, expectedly, in the geminal couplings: the attempts II - III show that it is not possible to obtain any practical improvements in them within the INDO/SOS method.

(iii) The optimized parameters, except in the case of FPT, deviate unreasonably from the originals.

## REFERENCES

1. Pople, J.A. and Santry, D.P., *Mol. Phys.* **8**(1964)1.
2. Pople, J.A., McIver, Jr., J.W. and Ostlund, N.S., *J. Chem. Phys.* **49**(1968)2965.
3. Salem, L., *"The Molecular Orbital Theory of Conjugated Systems"*, Benjamin, New York, 1966; Pople, J.A., *Proc. Phys. Soc. (London)* **A68**(1955)81; Ditchfield, R., *Mol. Phys.* **17**(1969) 33.
4. Andre, J.M., Nagy, J.B., Derouane, E.G., Fripiat, J.G. and Vercauteren, D.P., *J. Magn. Reson.* **26**(1977)317.
5. Kowalewski, J., Roos, B., Siegbahn, P. and Vestin, R., *Chem. Phys.* **3**(1974)70, *ibid.* **9**(1975)29.
6. Pople, J.A. and Beveridge, D.L., *"Approximate Molecular Orbital Theory"*, McGraw-Hill, New York, 1970.
7. Bingham, R.C., Dewar, M.J.S. and Lo, D.H., *J. Am. Chem. Soc.* **97**(1975)1285.
8. Karplus, M., *J. Chem. Phys.* **30**(1959)11; *J. Am. Chem. Soc.* **85** (1963)2870.
9. Castellano, S. and Bothner-By, A.A., *J. Chem. Phys.* **41**(1964) 3863.
10. Laatikainen, R., *J. Magn. Reson.* **27**(1977)169.
11. Dobosh, P.A. and Ostlund, N.S., *CNINDO/74, QCPE 281.*
12. Bacon, G.E., Curry, N.A. and Wilson, S.A., *Proc. Roy. Soc. Ser.* **A279**(1964)98.
13. Snyder, L.C. and Meiboom, S., *J. Chem. Phys.* **47**(1967)1480; Watts, V.S. and Goldstein, J.H., *ibid.* **46**(1967)4165.
14. Ernst, L., Wray, V., Chertkov, V.A. and Sergeyev, N.M., *J. Magn. Reson.* **25**(1977)123.
15. See for example, Barfield, M. and Johnston, Jr., M.D., *Chem. Rev.* **73**(1973)53.
16. Rader, C.P. *J. Am. Chem. Soc.* **91**(1969)3248.

# A Statistical Study on the Additivities of the Substitutional Effects in the $^{13}$C NMR Chemical Shifts of Hydroxy- and Chloro-substituted Benzenes

Juha Knuutinen, Reino Laatikainen and Jaakko Paasivirta
Department of Chemistry
University of Jyväskylä
SF-40100 Jyväskylä 10
Finland

https://doi.org/10.1002/mrc.1270140507

# A CNDO/2 Study on the Additivity and the Nature of the Non-Additivity of the Substituent Effects on $^{13}C$ NMR Shifts in Chlorobenzenes and Chlorophenols

Reino Laatikainen
Department of Chemistry
University of Jyväskylä
SF-40100 Jyväskylä 10
Finland

https://doi.org/10.1002/mrc.1270140508