

Advances in Computer-Assisted Evaluation of 1D NMR Spectra

Stanislav Sykora, Extra Byte, Italy

www.ebyte.it

Juan Carlos Cobas Gómez, Mestrelab, Spain

www.mestrelab.com

Why bother about 1D spectra

when 2D/3D ones contain more information ?

- Because 1D spectra are much more accessible.
- Because they are low-cost and fast to acquire.
- **Because we want to find the real limits of 1D NMR.**
- Because 1D's have still a lot to say

The NMR software house

Software tends to form hierarchical structures
(this is why programmers use the term *software layers*)

Since we do not want to talk about any specific
software but about the state of the art,
we will talk about **levels**, or **floors**,
and move from the basement up to the roof.

Basement and ground-floor

Basement (gray boxes):

Online software mounted on instruments and essential for the hardware operator.

Ground floor:

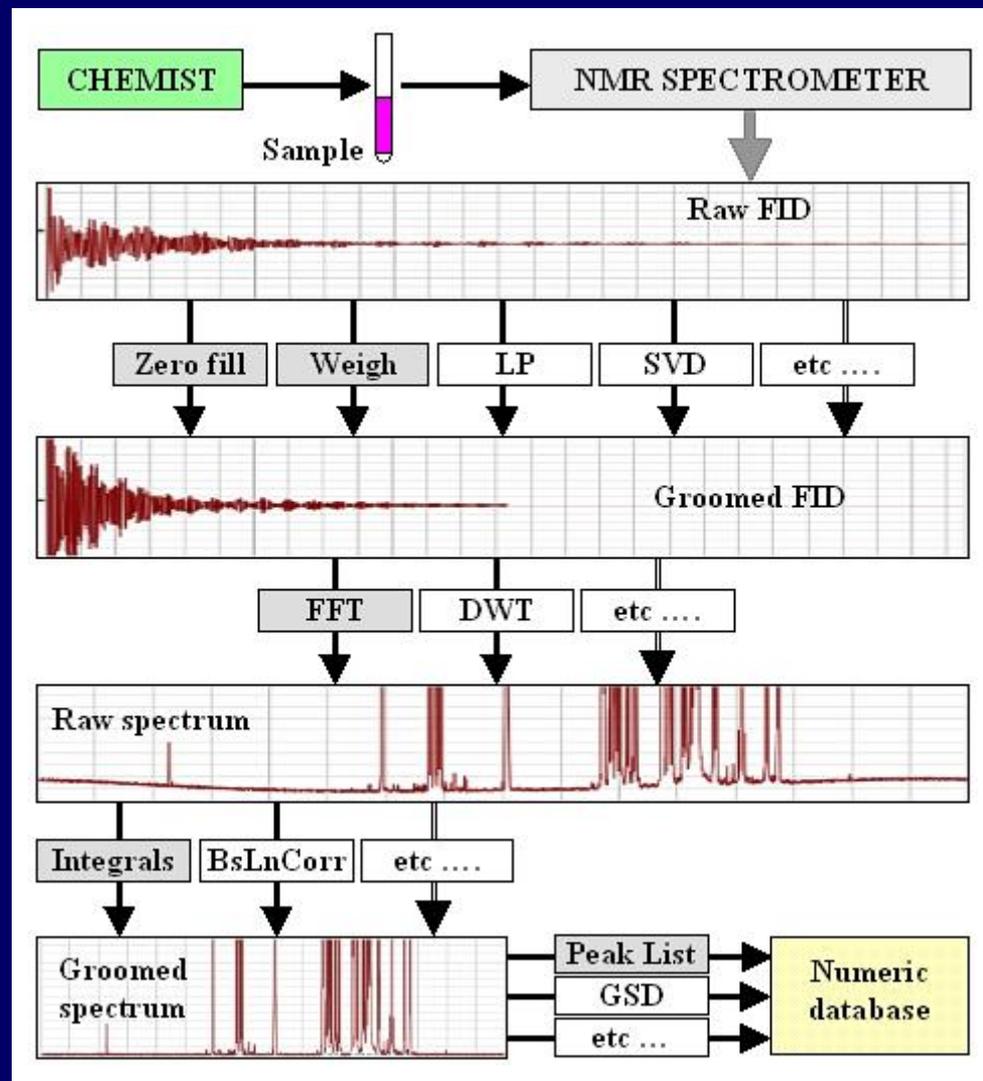
Software used mostly for post-acquisition evaluation. It may be also part of online packages but that is not essential for data acquisition.

The separation between levels is *fuzzy*

At these levels, the software includes very little specific NMR knowledge. Most of the procedures can be applied to almost any experimental data set.

Today, major advances are infrequent.
A recent example:

GSD ... Global Spectrum Deconvolution



First floor: the NMR department

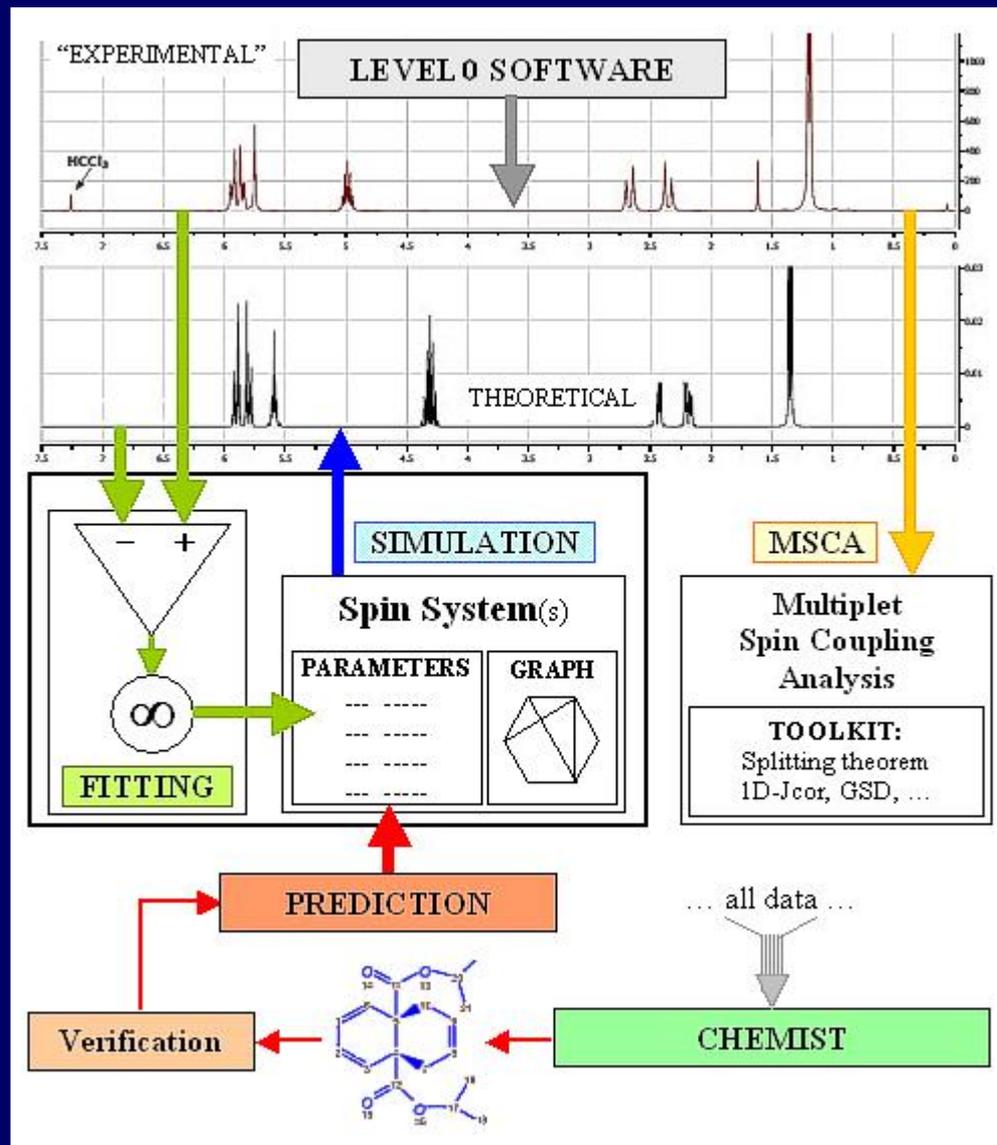
Here we find software which relies on NMR expertise.

- **Simulation of spectra**
- **Fitting of spectral parameters**
- **Prediction of parameters**
- **Spin-coupling analysis**

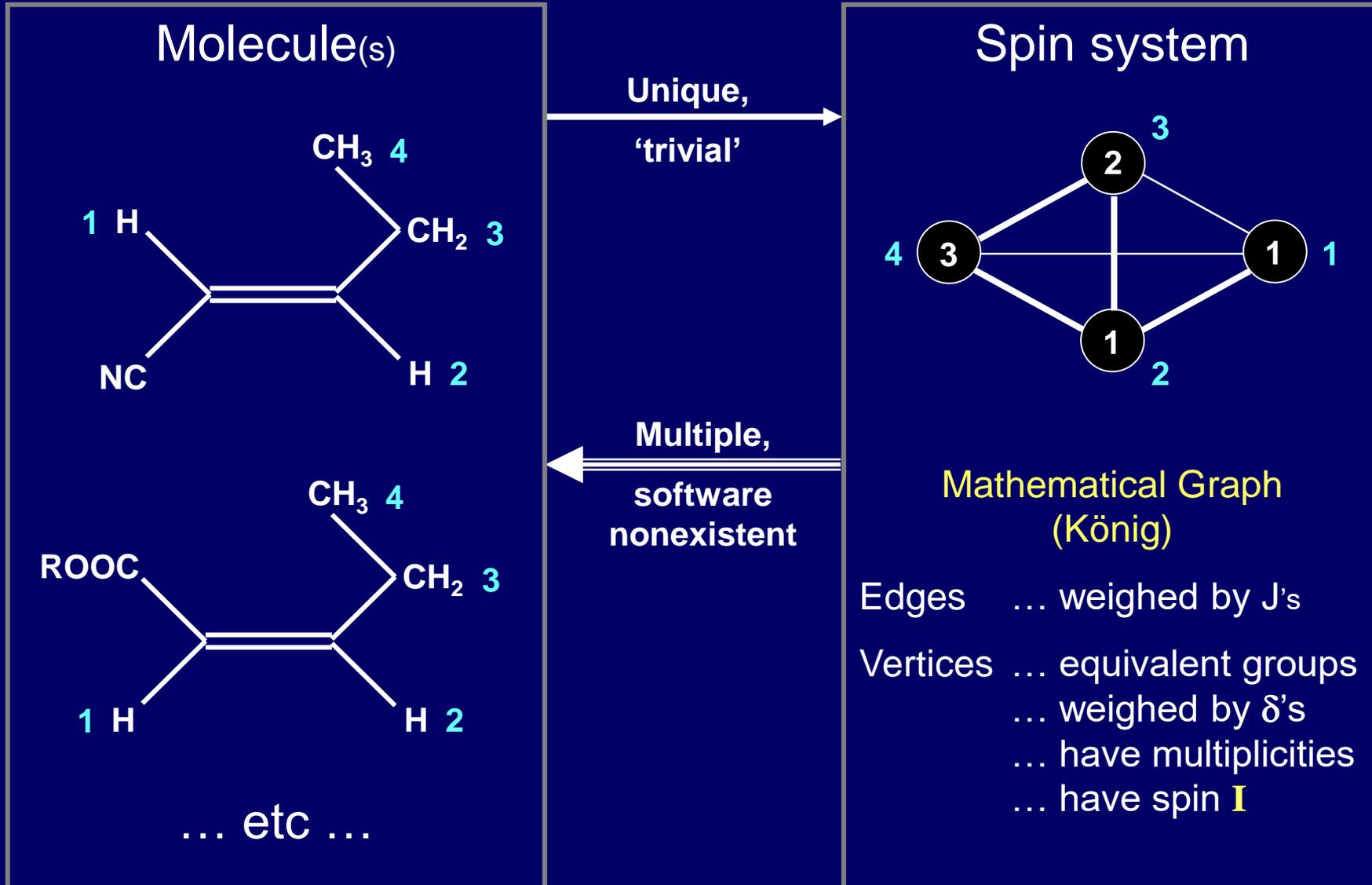
Presently, there is a growing interest, both academic and industrial, and the field is again on a fast upswing.

There are still formidable barriers to be overcome, but there is also no lack of tools and ideas to try.

“Direct” and “inverse” tasks



What is a spin system



Subtle shifts of perspective

NMR spectra reflect spin systems, not molecules

The analyzed spectrum is a primary input.
It determines one or more compatible spin systems.
A spin system determines compatible molecules.

Spectra come first, molecules last

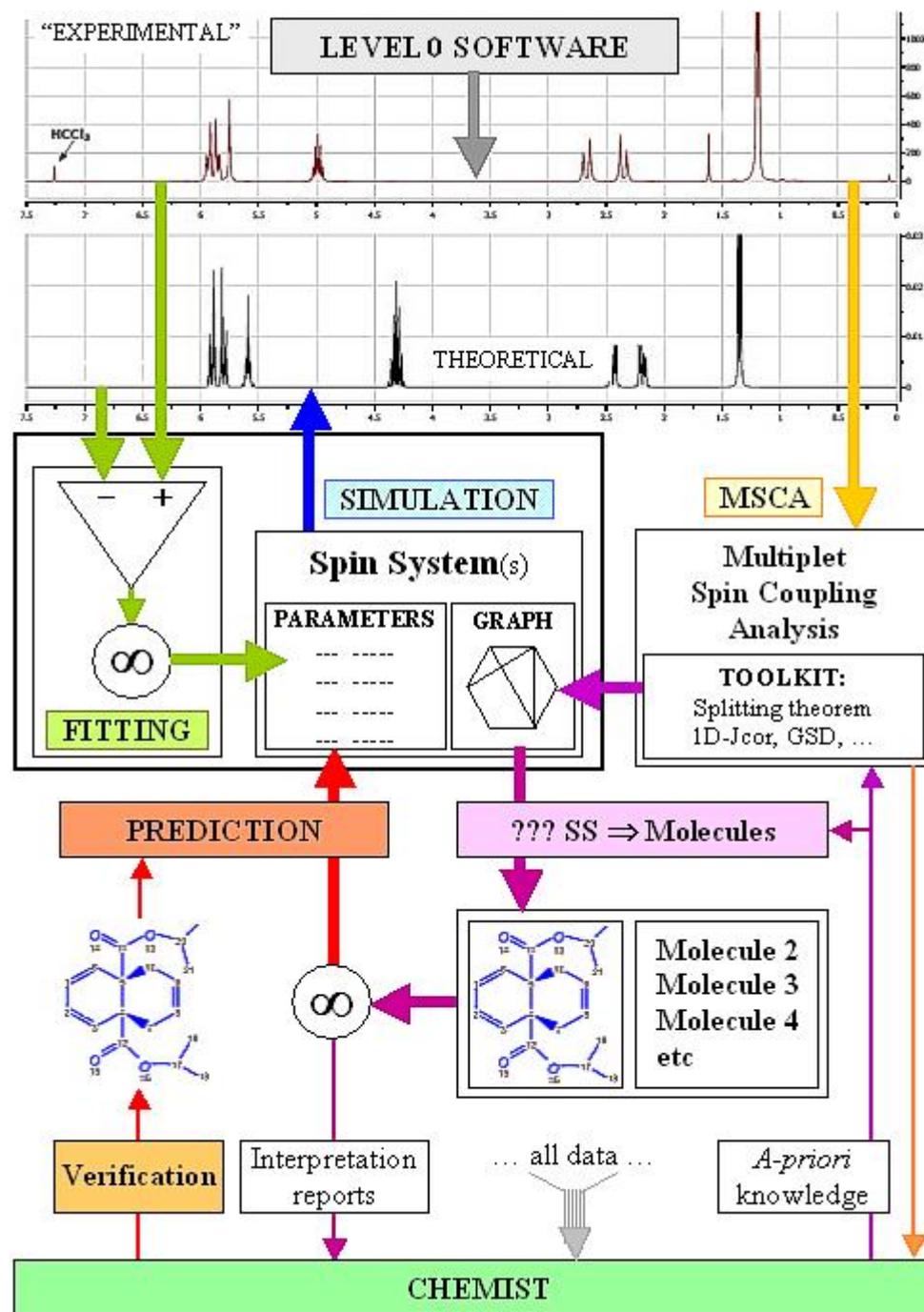


Full House

A complete NMR software
for molecular structure
applications

should combine

all the mentioned components



Simulation of NMR spectra

One can simulate the spectrum of a spin system with its structure graph and its parameters δ_i , J_{ij} and D_{ij}

Quantum – mechanical treatment is a must

The Hamiltonian:

static, motionally averaged, isotropic

$$\mathbf{H} = \sum_i \delta_i \mathbf{I}_i^z + \sum_{ij} J_{ij} (\mathbf{I}_i \cdot \mathbf{I}_j) = \sum_i \delta_i \mathbf{I}_i^z + \sum_{ij} J_{ij} (\mathbf{I}_i^z \mathbf{I}_j^z) + \sum_{i < j} J_{ij} (\mathbf{I}_i^+ \mathbf{I}_j^- + \mathbf{I}_j^+ \mathbf{I}_i^-)$$

or axially oriented:

$$\mathbf{H} = \sum_i \delta_i \mathbf{I}_i^z + \sum_{ij} (J_{ij} + D_{ij}) (\mathbf{I}_i^z \mathbf{I}_j^z) + \sum_{i < j} (J_{ij} - 2D_{ij}) (\mathbf{I}_i^+ \mathbf{I}_j^- + \mathbf{I}_j^+ \mathbf{I}_i^-)$$

Various approaches to simulation

1. Hilbert space of spin-states:

simplest & fastest

handles rigorously only static problems

dynamics and relaxation can be introduced empirically

2. Liouville space of spin-operators:

more difficult (dimensions are squared)

rigorous treatment of dynamic problems and relaxation

3.4.5 ...Other methods:

work in progress: there are alternatives

Let us concentrate on the Hilbert-space approach

Dimensions of the problem

for N nuclides with spin $S = \frac{1}{2}$

Maximum matrix ...

the largest matrix to diagonalize

k ...

the transition combination index:
one spin goes up, while k pair
of spins undergo exchange

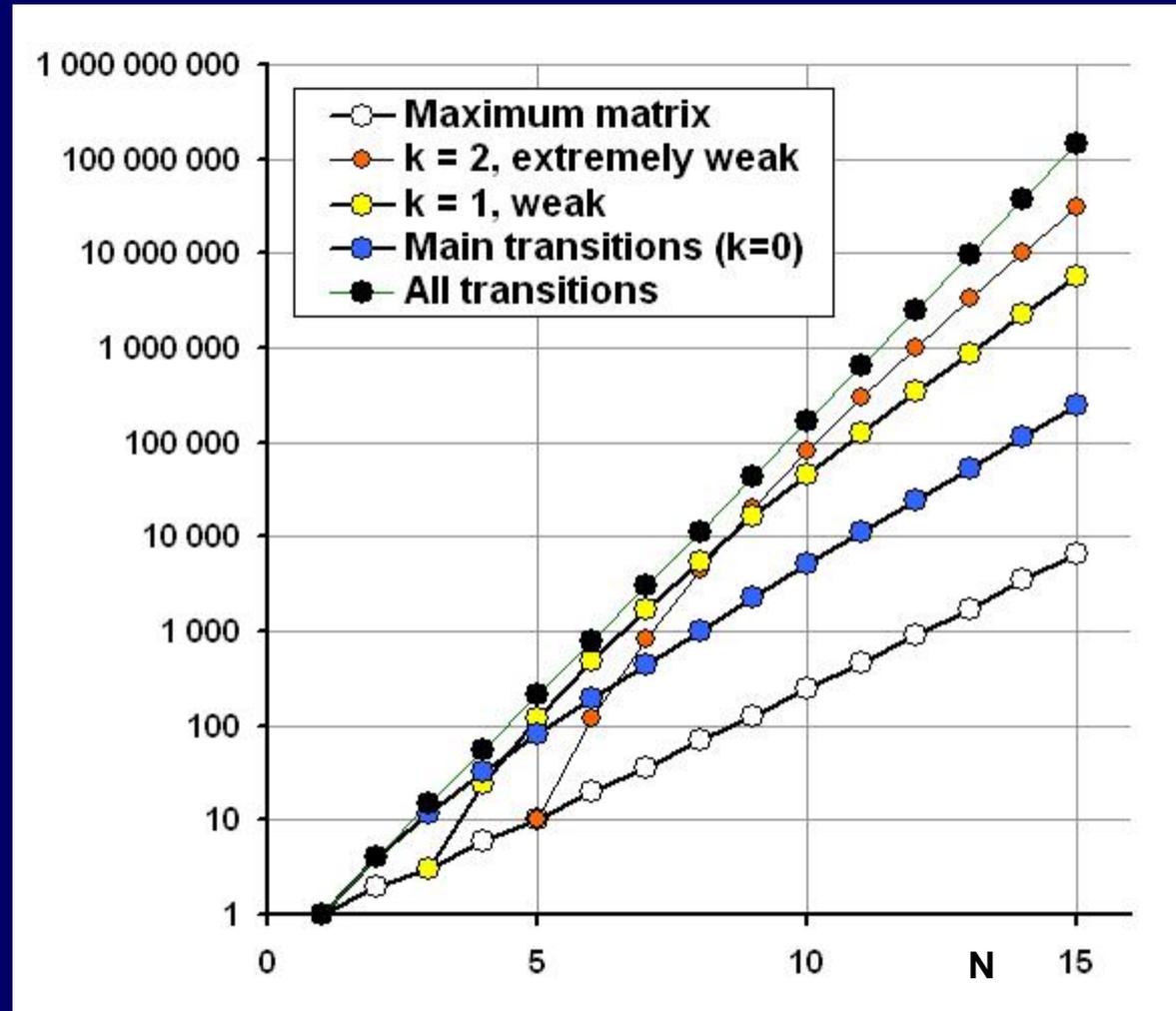
Weak coupling limit:

transitions with $k > 0$
have zero intensity.

Strongly coupled systems:

transitions with $k = 1$
must be considered

The numbers are HUGE !



Simulation performance

for N nuclides with spin $S = \frac{1}{2}$

Execution load expressed

in 100 MFlop units

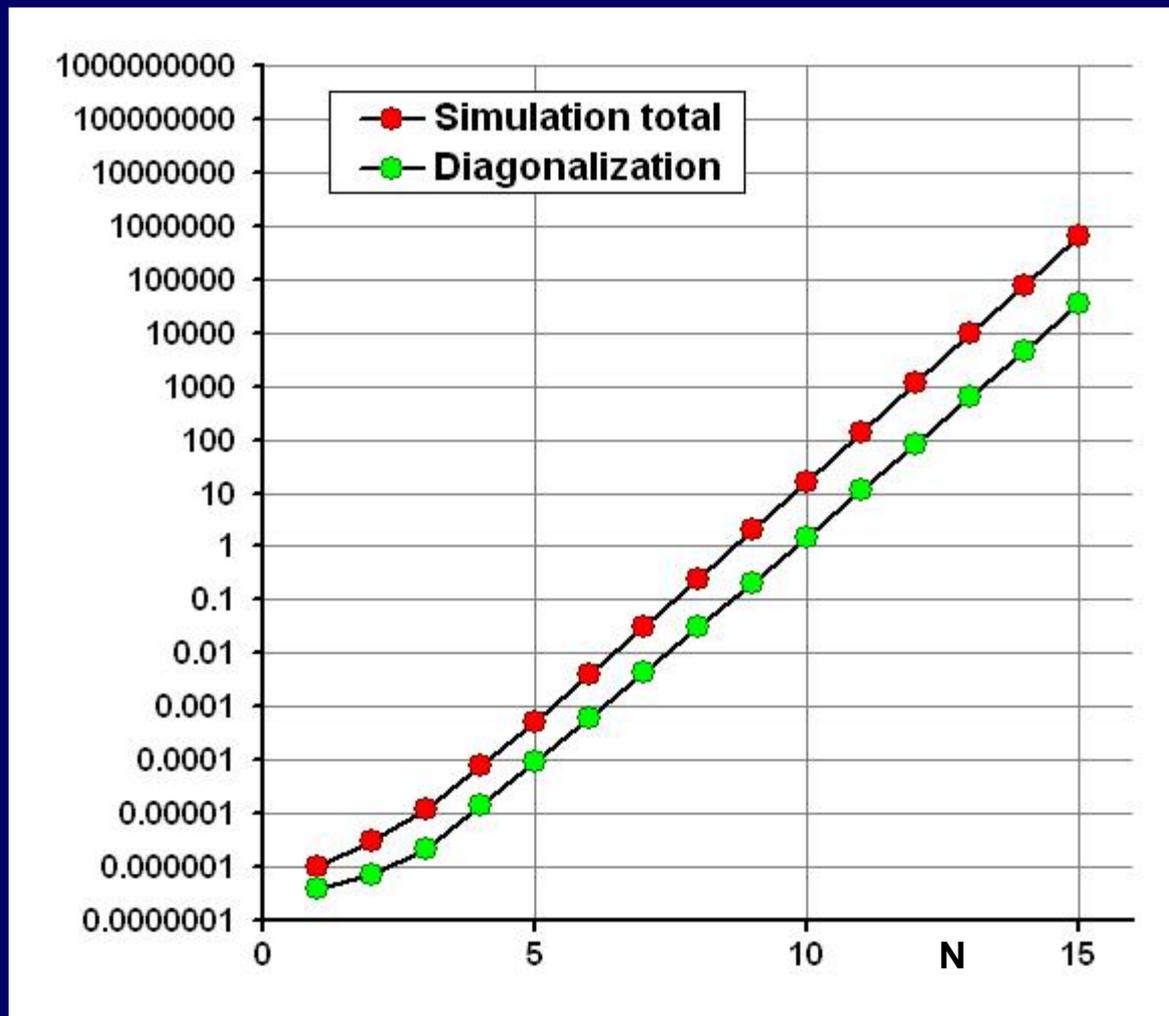
On a 1 GHz PC

100 MFlops take about 1 second

Simulation total =

diagonalization of matrices +
handling of transitions

The times are HUGE !



Fragmentation of the spin system

Is the idea not obvious ?

Yes and not !

Rules of the game:

- Fragment the spin-system, not the molecule
- There are various strategies to do it (options)
- Handle fragment overlays (just bookkeeping)
- It is an approximation, not a dirty trick:
⇒ one must prove that the errors are negligible

Fragmentation at work (strychnine)

No fragmentation

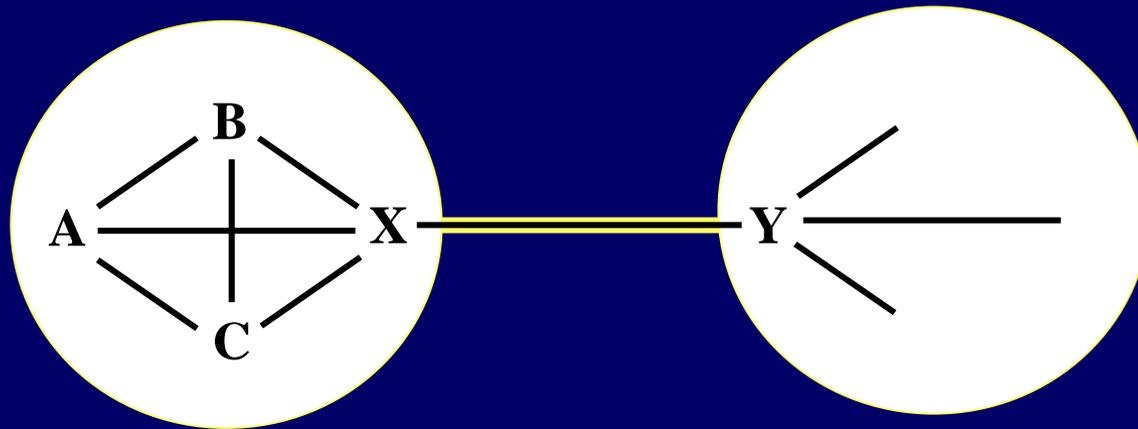
Method 1

Metod 2

Spin System Profile	Eqv.Group Properties	Interactions	es	Interactions	es	Interactions
UserId / SubId	2 / 0		3 / 0		4 / 0	
Parent UserId/SubId	User-defined system		User-defined system		User-defined system	
Equivalent Groups	22		22		22	
Enabled Equivalent Groups	22		22		22	
Weight	1.00000		1.00000		1.00000	
Equivalent size	22		22		22	
Total spin	22/2		22/2		22/2	
Runtime flags	0		0		0	
Splitting Level	2 (Graph components)		2 (Graph components)		2 (Graph components)	
Branches	2		2		2	
Subsystems / Exe.subs	3 / 2		12 / 10		18 / 16	
Weight factor	1		1		1	
Number of manifolds	24		114		89	
Maximum dimension	352716 x 352716		3432 x 3432		70 x 70	
Number of states	164322		72066		706	
Main transitions	22020097		495041		2249	
Cl transitions	1045954560		9382800		8262	
All transitions	2468501365		152615148		16596	
Execution time [s]	321927158.799		309937.606		0.282	
Minimum RAM [MByte]	26898.475		167.794		0.081	

An old fragmentation theorem

Sykora, *Approximate Methods in Analysis of NMR Spectra. Application of Perturbation Theory to Decomposition of Many-Spin Systems*, Coll.Czechoslov.Chem.Comm. **33**, 3073, 1968. Available at www.ebyte.it



The effect of J_{XY} on the A-transitions is described by

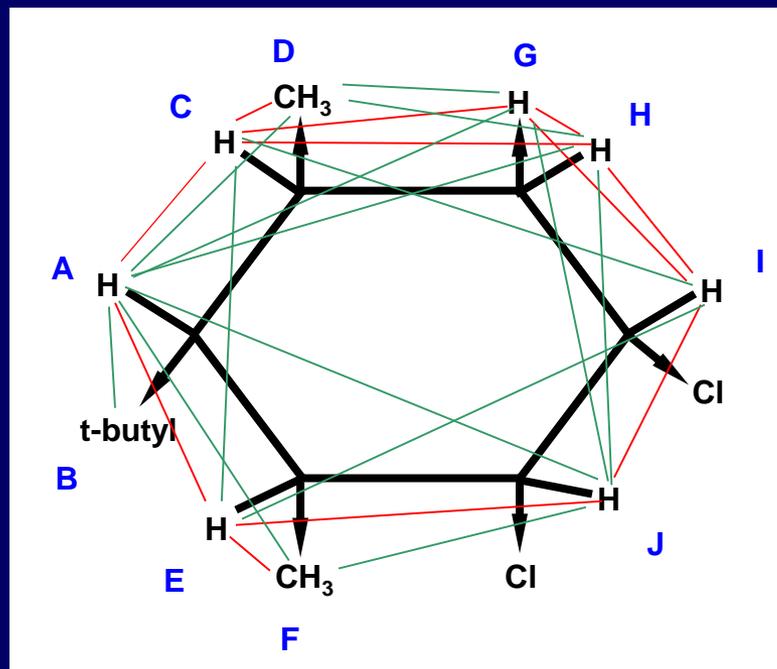
$$\Delta\nu_A = J_{XY} (\Delta J_X^z) (J_Y^z)$$

and there is also an analogous formula for intensities

The formula is **additive** and describes **maximum effect**

Are there still limits ?

Yes, one can invent molecules whose spin systems can not be efficiently fragmented.
However, if I did not tell you, you would have probably never found out.



System of 22-spins with 10 equivalent groups of the type (A B9 C D3 E F3 G H I J).
The smallest fragment for A contains 21-spins (A B9 C D3 E F3 G H J).
But this happens only if almost all 4-bond J's are present!

Fitting of spectral parameters

Once a simulation method has been chosen, one can try and adjust the spin system parameters (not the graph) to fit a given experimental spectrum

Methods differ by a combination of the following:

- Target function $Q(\mathbf{p})$ to be optimized
- Optimization algorithm

Optimization algorithms

These are general-purpose classical algorithms.
Some require knowledge of the derivatives dQ/dp_i , some don't

- **DS** ... Downhill simplex (no der)
- **LM** ... Loewenberg_Marquardt (der)
- **CG** ... Powell's conjugate gradients (der)

All these methods have huge convergence problems for more than
about 10 parameters but,

for DS, we have solved this problem

We prefer DS also because it does not need derivatives,
giving us more freedom to experiment with various target functions Q .

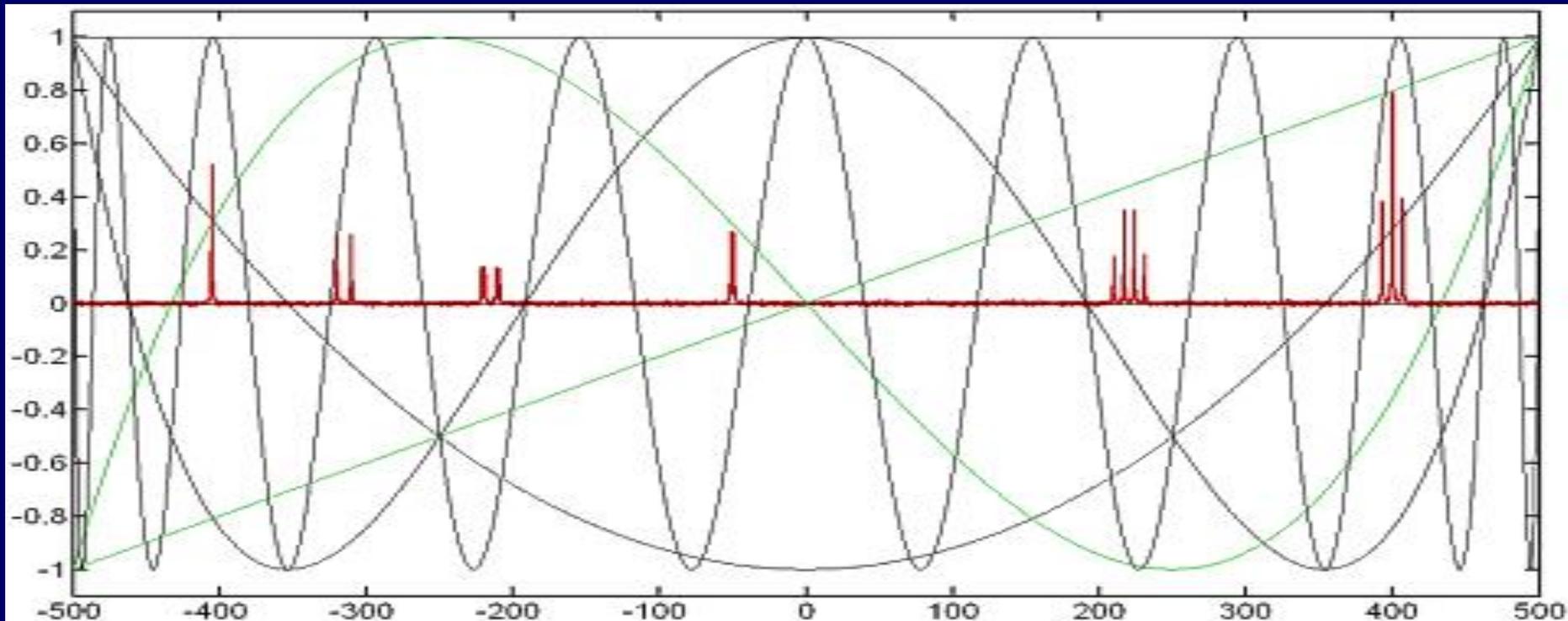
The target functions

- **Laocoön** (Castellano '60s)
Matches the frequencies of selected (line-transition) pairs.
Assignments usually need to be done manually.
Completely ruled out today.
- **Integral transforms** (Sykora, Vogt, Diehl 1974)
Designed to avoid multiple local minima.
Great freedom in choosing the basis of functions for the IT's
- **Innovative interval functions** (work in progress)
Criteria like *"in this interval, the tallest peak is at 2.33 ppm"*

Integral Transforms base functions

We prefer **Chebyshev polynomials T_n** because they have many nice properties:

- Are extremely easy to compute [$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$]
- Exhibit a natural transition from power-like behaviour to a harmonic one
- Have bounded amplitudes (± 1) and easily adapt to any spectral window



Fitting obstacles

➤ Exasperating sluggishness

In large systems it takes > 100 iterations per parameter
Should be improved by a factor of 1000
Work in progress; good expectations

➤ Lack of convergence

We have solved this one

➤ False local minima

Proper choice of the base is essential
Cleverly designed target functions are essential
Work in progress (not particularly worrisome)

➤ Experimental artifacts

Impact of systematic intensity errors
False lines (sidebands, solvent, impurities, ^{13}C satellites)

Fitting performance

for N nuclides with spin $S = \frac{1}{2}$

Execution load expressed

in 100 MFlop units

On a 1 GHz PC

100 MFlops take about 1 second

Simulation total

= diagonalization of matrices
+ handling of transitions

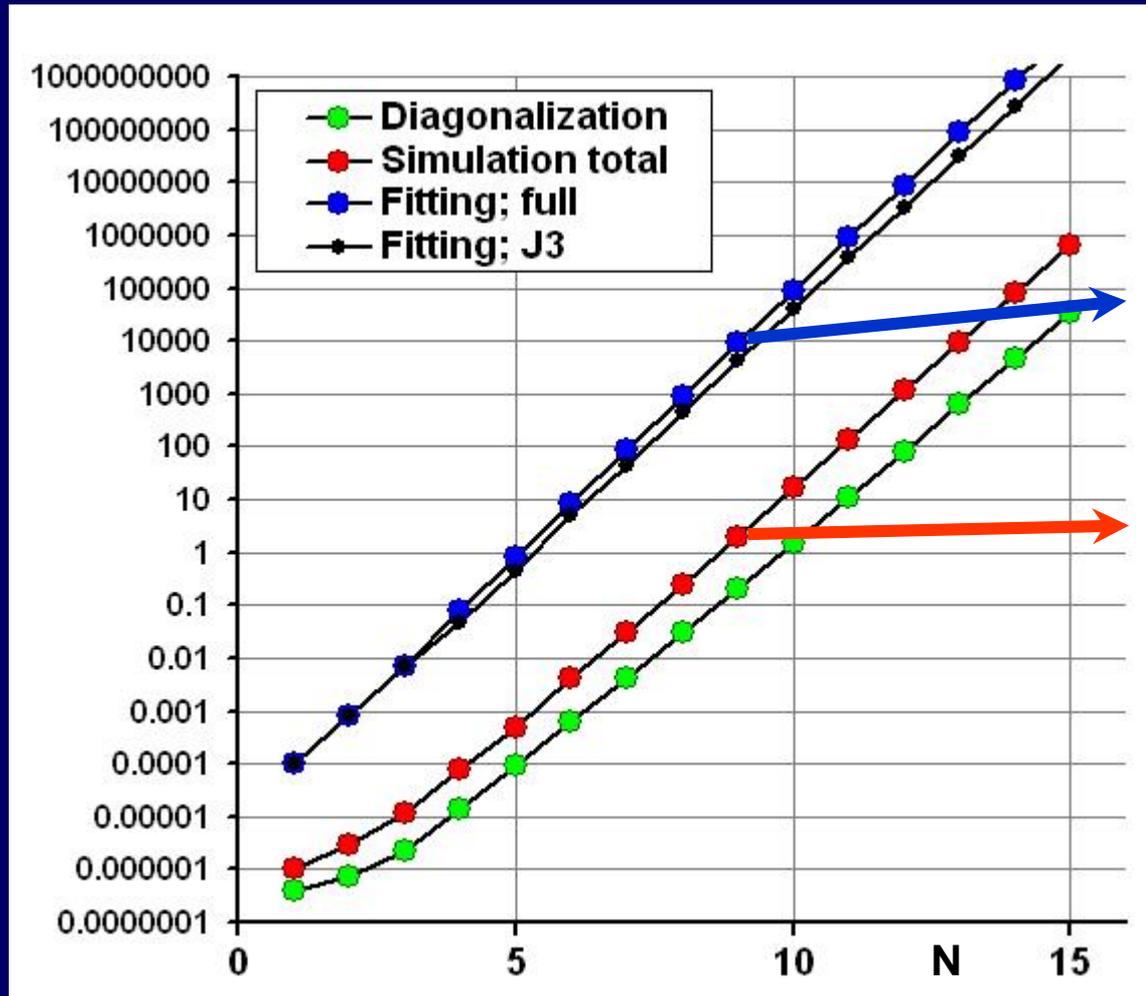
Fitting load estimates:

full ... totally coupled spin graph
J3 ... un to 3-bond couplings

Thick arrows show the effect of
spin-system fragmentation

Exe times are still HUGE !

Should be cut by 1000



MSCA

Multiplet Spin-Coupling Analysis

Loose definition:

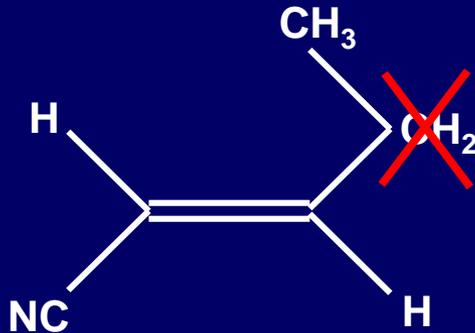
A set of tools, both know-how rules and algorithms, which make it possible to determine all spin systems compatible with a given experimental spectrum

Composition:

- **A Toolkit**
of algorithms and NMR rules (pieces of knowledge)
- **An Artificial Intelligence**
applying the tools to an NMR task

Subspectra due to equivalent groups

(an abstract MSCA tool)



One of the subsystems has the H₂ replaced by a pseudo-nucleus with spin 0.

Consequently, if the J⁴ coupling is negligible, we must have somewhere in the spectrum:

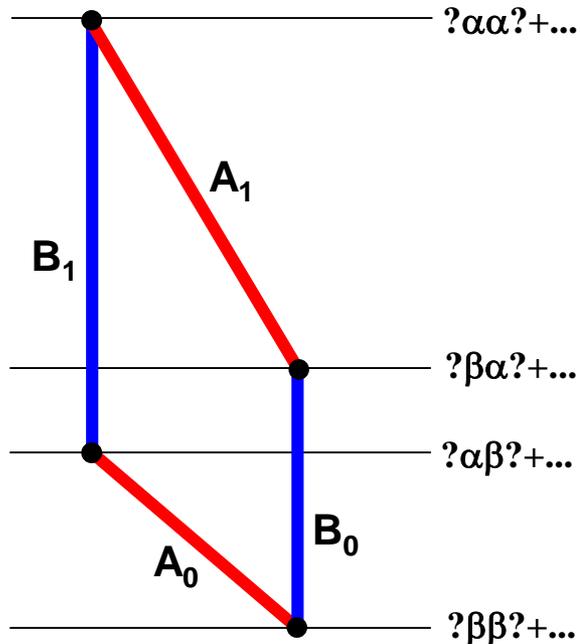
- A methyl singlet
- A pure AB-type quadruplet

When the J⁴ is resolved, the singlet becomes a narrow doublet and half of the quadruplet lines split into narrow quadruplets.

Splitting Theorem

(an abstract MSCA tool)

Basic level-diagram rule for a spin-system fragment



Theorem:

$$A_1 - A_0 = B_1 - B_0$$

Corollary:

When multiplets A and B are coupled, there is a splitting in A and a splitting in B which are exactly the same, no matter how complex is the system and how strongly it is coupled

Note: in strongly coupled systems, splitting \neq coupling

MSCA Toolkit algorithms

(based on the Splitting Theorem)

➤ **1D JCor (algorithm and utility)**

Converts a 1D spectrum into a 2D-like plot in which it is easy to correlate multiplets containing the same splitting. Great at ruling out a coupling between too multiplets. Suitable also for human inspection.

➤ **Peak list coincidence analyzer (algorithm)**

Does a similar thing numerically, starting from a peak list

➤ **GSD list coincidence analyzer (algorithm)**

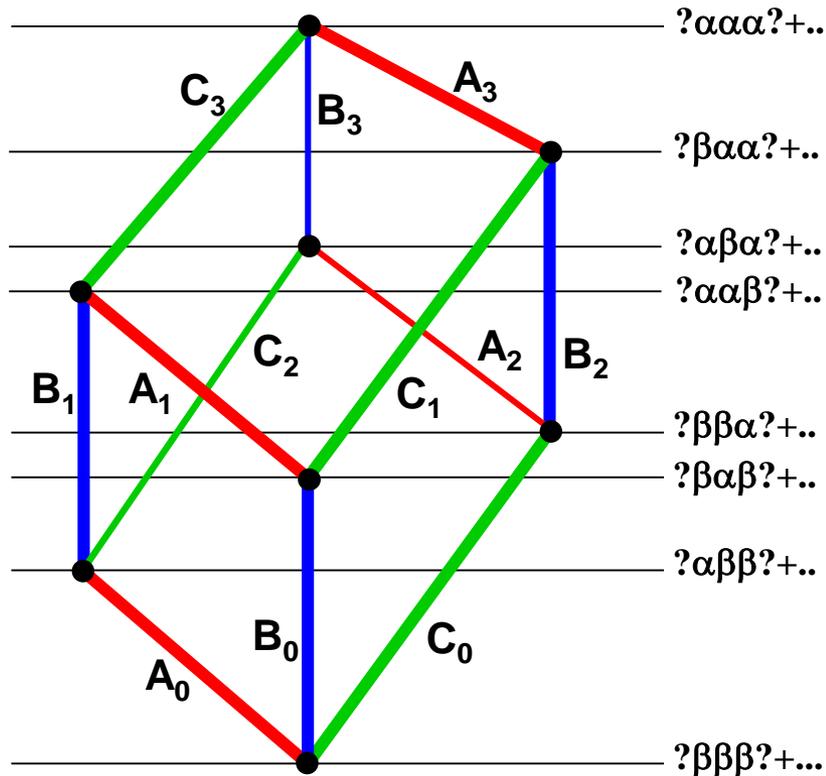
Again the same, using a Global Spectrum Deconvolution list

The **Fragmentation Theorem** is another example of an abstract MSCA tool

ABC Splitting Theorem

(an abstract MSCA tool)

Level diagram and its consequences for ABC system



Constraints on the 12 main transitions:

$$A_1 - A_0 = B_1 - B_0$$

$$A_2 - A_0 = C_2 - C_0$$

$$A_3 - A_1 = C_3 - C_1$$

$$A_3 - A_2 = B_3 - B_2$$

$$B_2 - B_0 = C_1 - C_0$$

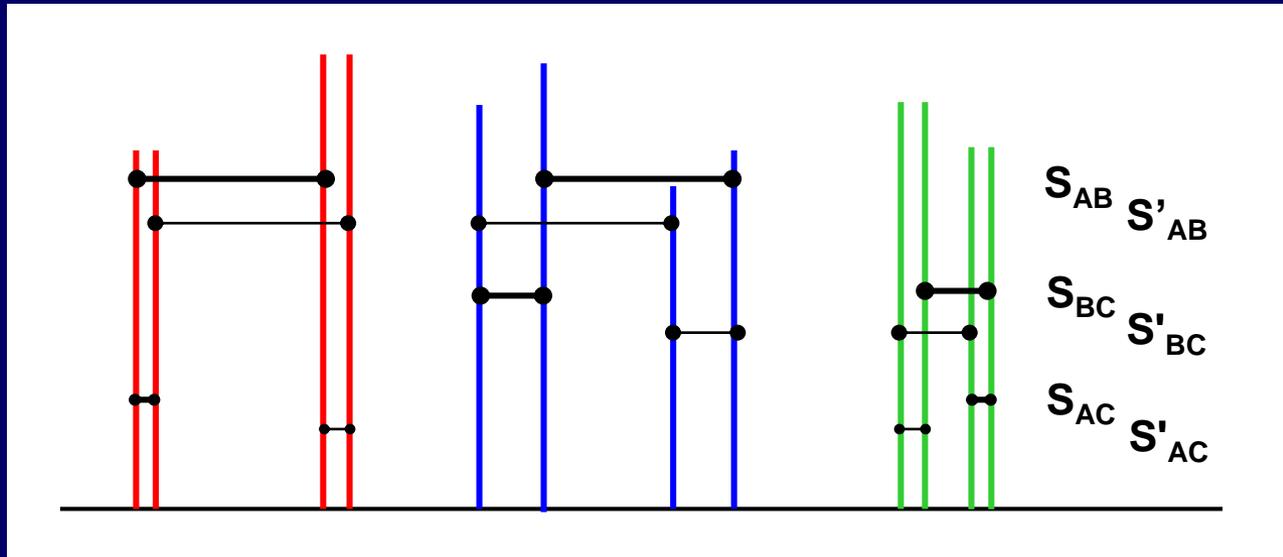
$$B_3 - B_1 = C_3 - C_2$$

These equalities hold no matter how strong are the couplings !!!

MSCA algorithms based on this Theorem are under development

Nature loves first-order splittings

Corollary of the **ABC** Splittings Theorem:



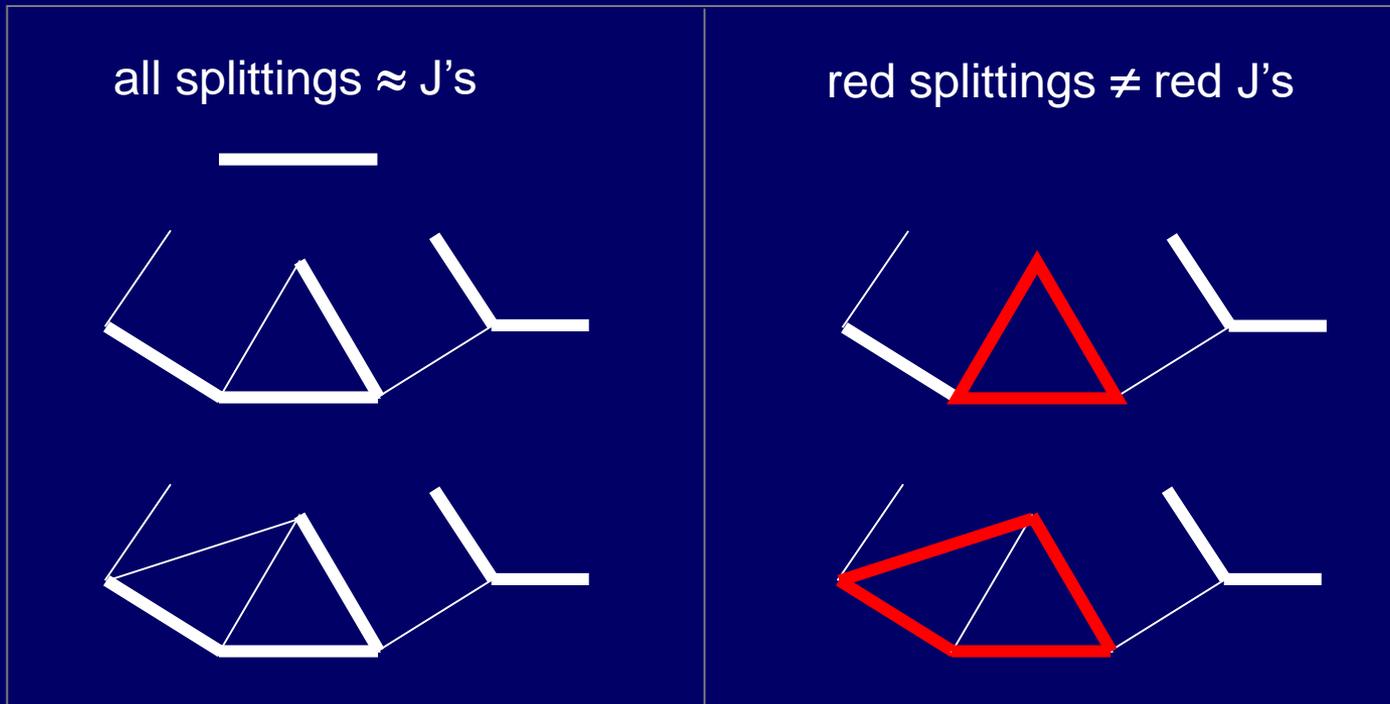
When **all** the couplings are strong, $S'_{ik} - S_{ik} = \xi \neq 0$, but there is *only one* ξ

When **any** of the couplings is weak, $S'_{ik} = S_{ik} = J_{ik}$ for **any** i,k !

Notice the **Roof Effect** which is still another MSCA tool

Stan's Conjecture

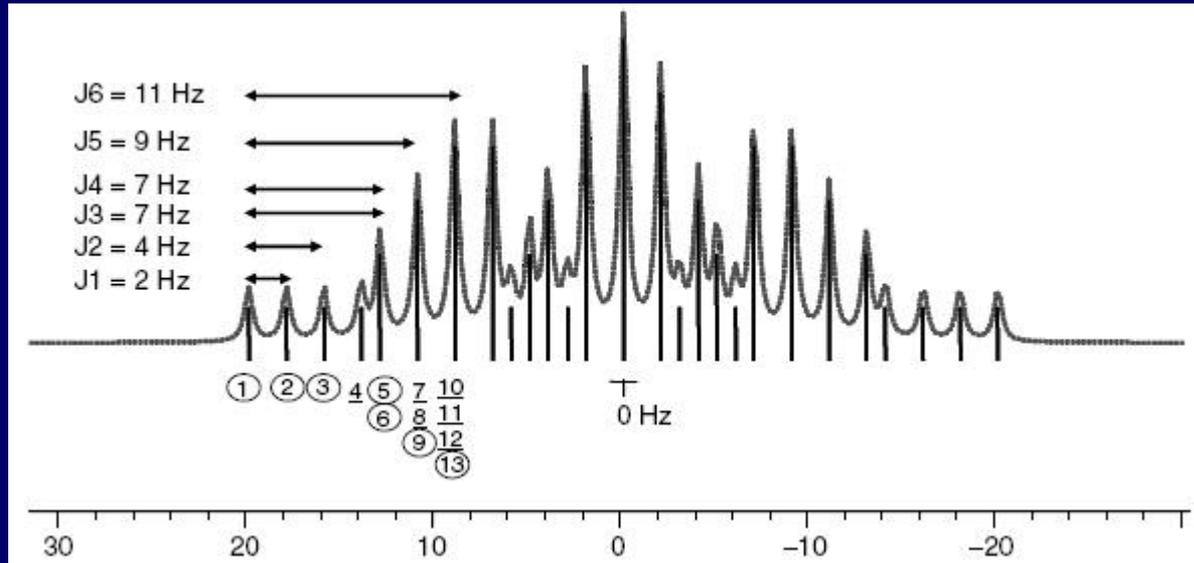
Splittings differ significantly from J's only when the involved couplings are all strong and form a closed loop



An old MSCA algorithm

boosted up by Stan's Conjecture

The first-order, single-multiplet analysis
is after all still good in all but very few situations



Cobas et al, *A two-stage approach to automatic determination of ¹H NMR coupling constants*, *Magnetic Resonance in Chemistry*, **43**, 843-848 (2005)

So, again, what is MSCA?

One answer:

It is the **top NMR Expert in the NMR Software House**

To simulate and/or fit a spectrum one does not really need to know NMR. Once one has encoded a very limited number of formulae, it is automatic.

To predict shifts and J's one must know quantum chemistry, not NMR.

But to assign a spin system to the spectrum of an unknown, one needs all the NMR tricks and treats one can put one's hands on!

Moreover, we want to build an **Artificial Intelligence**. We must therefore break all NMR know-how into encodable chunks and teach the AI how to use them.

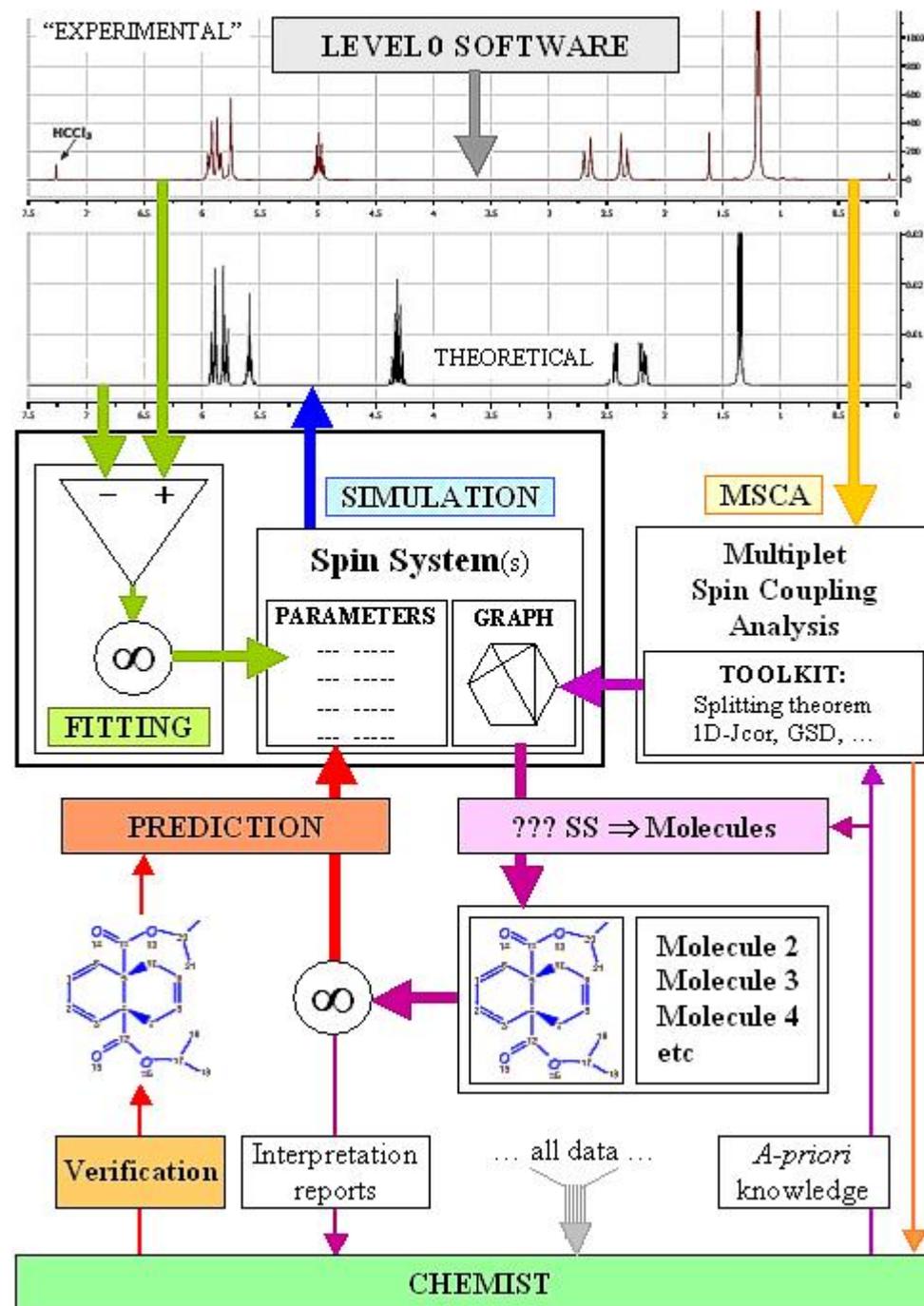
Since spectra are imperfect the AI will use fuzzy logic, assigning and combining probabilities on various counts.

How far are we with this baby?

The crib is under preparation, conception is imminent

Current status

- **Simulation:**
Almost no limits (2006 - 2007)
Impossible systems are very rare
- **Fitting:**
Limited to about 20 parameters
Breakthrough in optimization (May)
Work in progress on target functions
Work in progress on IT windowing
- **Spin-Coupling Analysis:**
Tools are being collected
Some are nice & very usable
- **Prediction:**
Mean errors: ± 0.25 ppm / ± 0.1 Hz
Troublesome cases still frequent
- **Spin System \Rightarrow Molecules:**
So far just a concept
A name is needed (public contest)



Collaborators

The *House of NMR Software* is a public project open to all.

Personally, the Authors participate in a venture with three **Contractors**, each with its own **BrickMasters** whose help we gratefully acknowledge:

Mestrelab:

- **Santiago Dominguez** (the one with the big whip)
- **Nikolay Larin** (mathematician & code-master with a small whip)
- **Isaac Iglesias** (physicist & code developer)
- **Felipe Seaone** (physicist & code developer)
- **Maruxa Sordo** (physicist & a rose in the desert)
- **Santi Fraga** (dot-net physicist)
- **Pablo Monje** (applications chemist - fits spectra to the theory)
- **Cristina Geada** (keeps the shop running with a magic wand)
- **Roberto Cobas** (keeps Stan running on Galician *pulpo* and wine)

Modgraph:

- **Ray Abraham** (decides what δ 's and J's should each molecule have)
- **Ernö Pretsch** (sentences those molecules which do not comply)
- **Mike Wainwright** (encodes the sentences in harsh C++)

Extra Byte:

- **Silvestro** (Stan's cat responsible for the most extravagant ideas) →

