NMR Spectra
Processing, Verification and Elucidation:
challenges and current progress

Stanislav Sykora, Extra Byte, Italy
www.ebyte.it

Juan Carlos Cobas Gómez, Mestrelab, Spain
www.mestrelab.com

The abstract and slides of this talk are available at
www.ebyte.it/stan/Talk_Valtice_2008.html

I would like to touch upon:

- ➢ **"Simulation"** of 1D spectra **(spin system $\Rightarrow$ spectrum)**

- ➢ **"Fitting"** **(spin system $\Leftrightarrow$ spectrum)**

- ➢ Problems with automatic **verification** & **elucidation**

- ➢ Verification & elucidation **master flowchart**

- ➢ Emerging **tools**:

    J-Correlator, Resolution Booster, GSD, Digital J-Correlator

- ➢ Necessity of a **two-step approach**:

    (1) Spectrum $\leftrightarrow$ Spin System (NMR physics & know-how)
    (2) Spin System $\leftrightarrow$ Molecule (chemical know-how)

What is a spin system

Molecule(s)

CH$_3$ 4

1 H

CH$_2$ 3

NC

H 2

CH$_3$ 4

ROOC

CH$_2$ 3

1 H

H 2

… etc …

Unique,
'trivial'

Multiple,
software
nonexistent

Spin system

3
2
4 3   1 1
1
2

**Mathematical Graph
(König)**

Edges   … weighed by J's

Vertices … equivalent groups
            … weighed by δ's
            … have multiplicities
            … have spin **I**
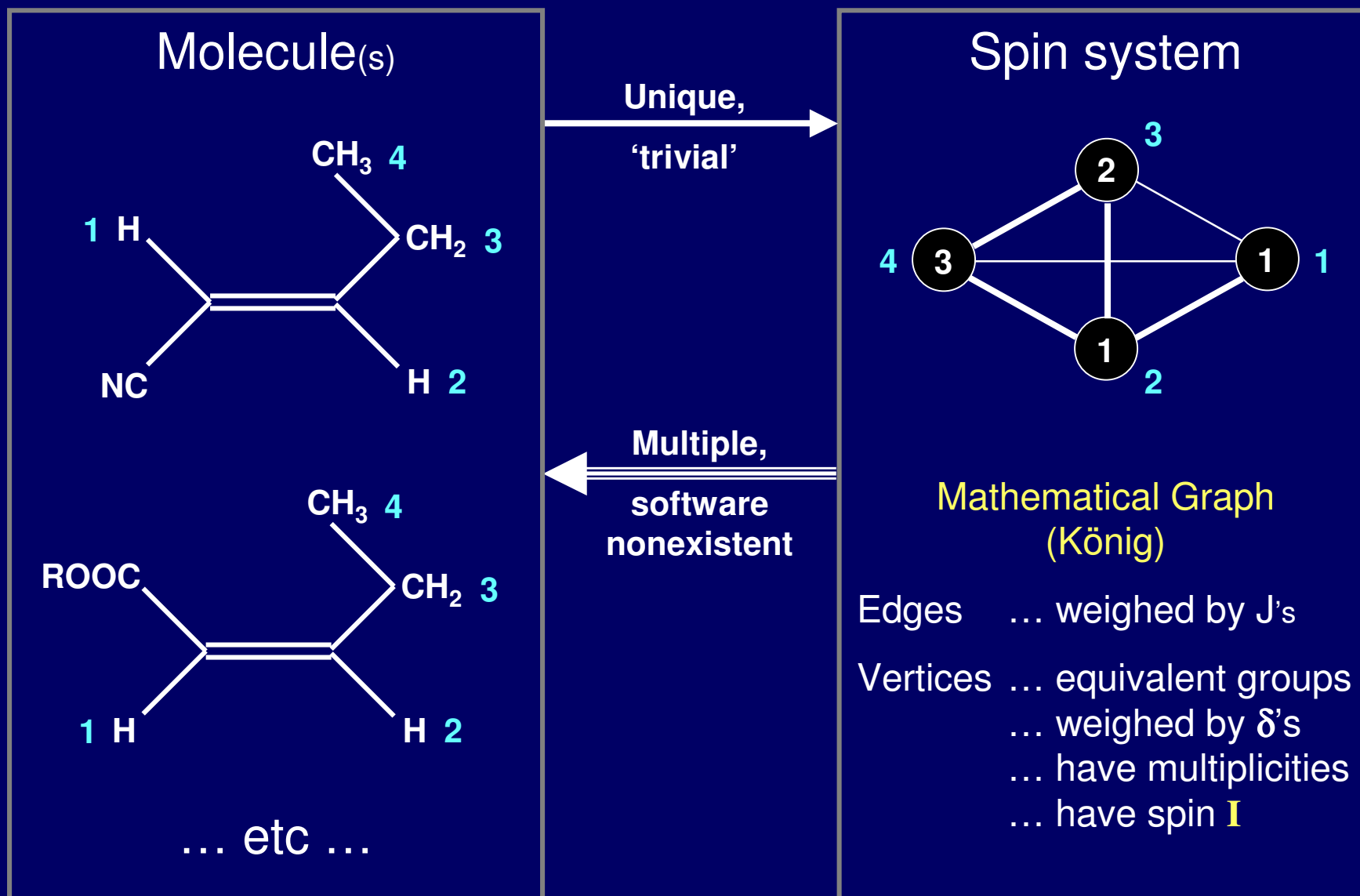
Simulation of NMR spectra

One can only 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

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

or axially oriented:

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

The current engine covers isotropic and axially oriented spin systems with any spin

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
   at present, feasible only for very small spin systems

3.4.5 …Other methods:

   work in progress: there are several alternatives

   We concentrate on the Hilbert-space approach

Dimensions of the problem

for N nuclides with spin S = ½

Maximum matrix …
the largest matrix to diagonalize

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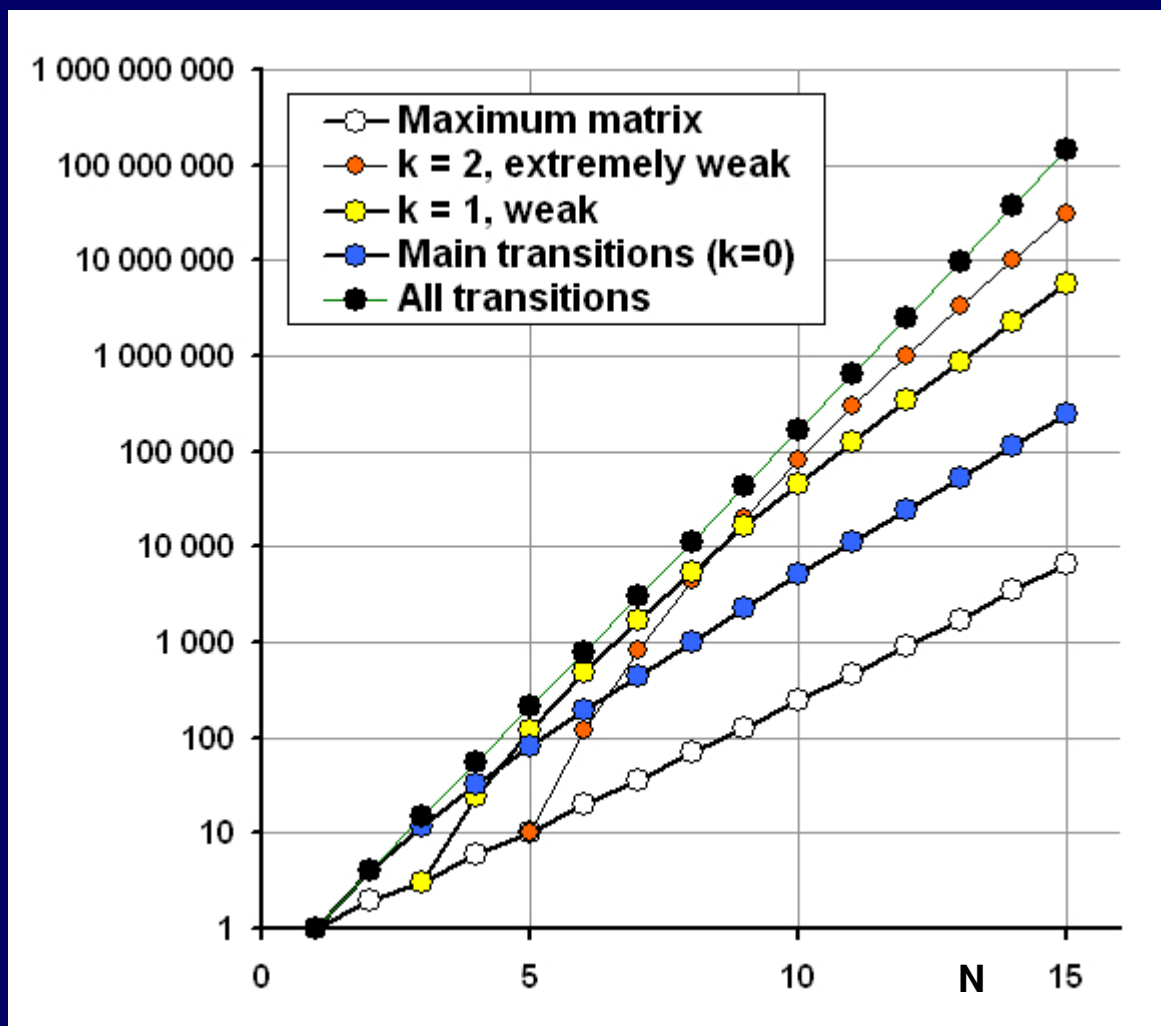one spin goes up, while k pair
of spins undergo exchange

Weak coupling limit:
transitions with k > 0
have zero intensity.

Strongly coupled systems:
transitions with k = 1
must be considered!

The numbers are HUGE !

- —○— Maximum matrix
- —●— k = 2, extremely weak
- —●— k = 1, weak
- —●— Main transitions (k=0)
- —●— All transitions

Simulation performance

for N nuclides with spin S = ½

Execution load expressed

in 100 MFlop units
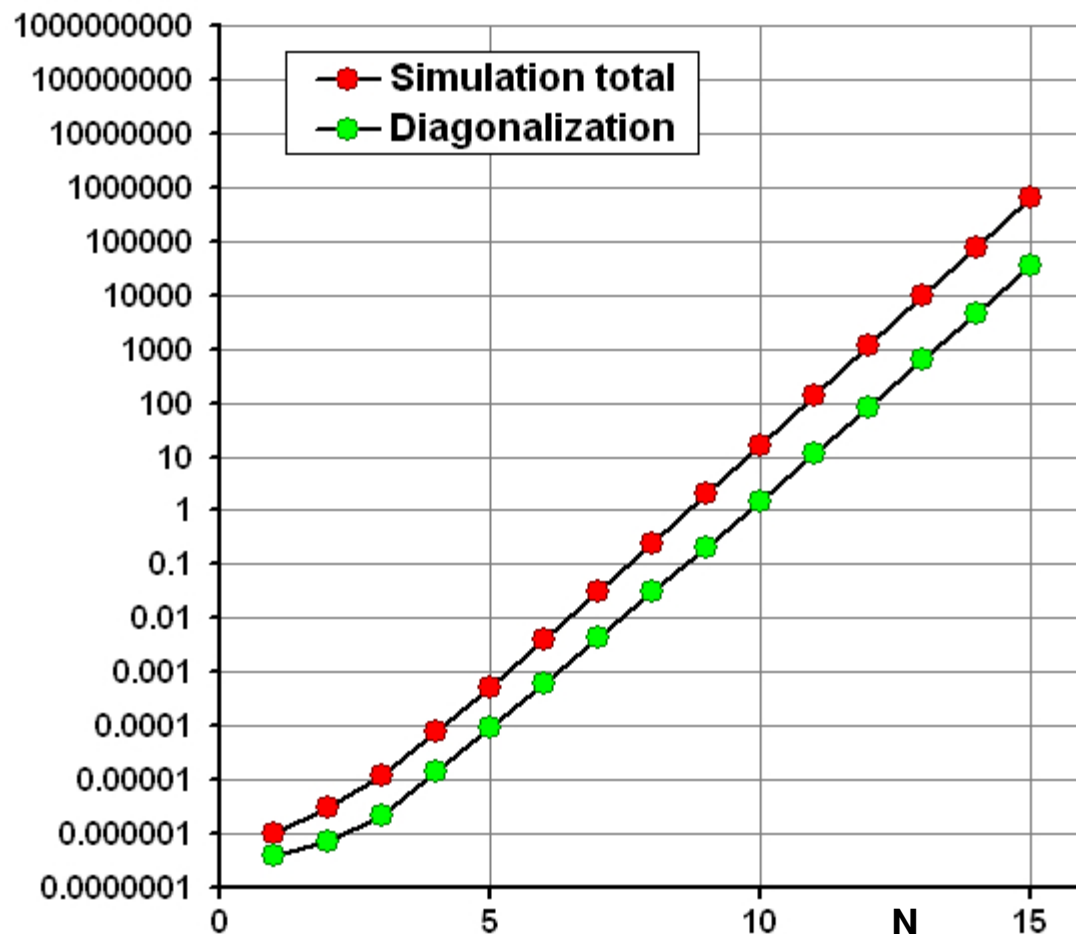
On a 1 GHz, single-core PC

100 MFlops take about 1 second

Simulation total =
diagonalization of matrices +
handling of transitions

The times are HUGE

despite careful optimization!

Fragmentation of the spin system

Isn't the idea obvious ?
Yes and Not !

Rules of the game:

➤ Fragment the <u>spin-system</u>, not the molecule!

➤ There are many strategies to do it

➤ It is an approximation, not a dirty trick:

  $\Rightarrow$ one must prove that final errors are negligible

  $\Rightarrow$ proper fragmentation requires NMR knowledge

Fragmentation at work

Cholesterol:	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	
C1 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	

Simulation is essentially solved

? What about fitting ?

What can fitting really achieve ?

One can only fit the **numeric parameters**
(shifts, coupling constants, …)
of a **given** **spin system**
(not a molecule)

! This is NOT what chemists intend !

Even so, there are many obstacles
(even if the spectra were perfect)

➢ **Exasperating sluggishness:**

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

➢ **Frequent lack of convergence:**

> We have solved this one

➢ **Possible multiple solutions:**

> Should be automatically detected

➢ **False local minima:**

> Proper choice of the the fitting-target function is essential

Fitting performance

for N nuclides with spin S = ½

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
<u>spin-system fragmentation</u>**

Exe times are still HUGE !

Should be cut by 1000

Legend:
- Diagonalization
- Simulation total
- Fitting; full
- Fitting; J3
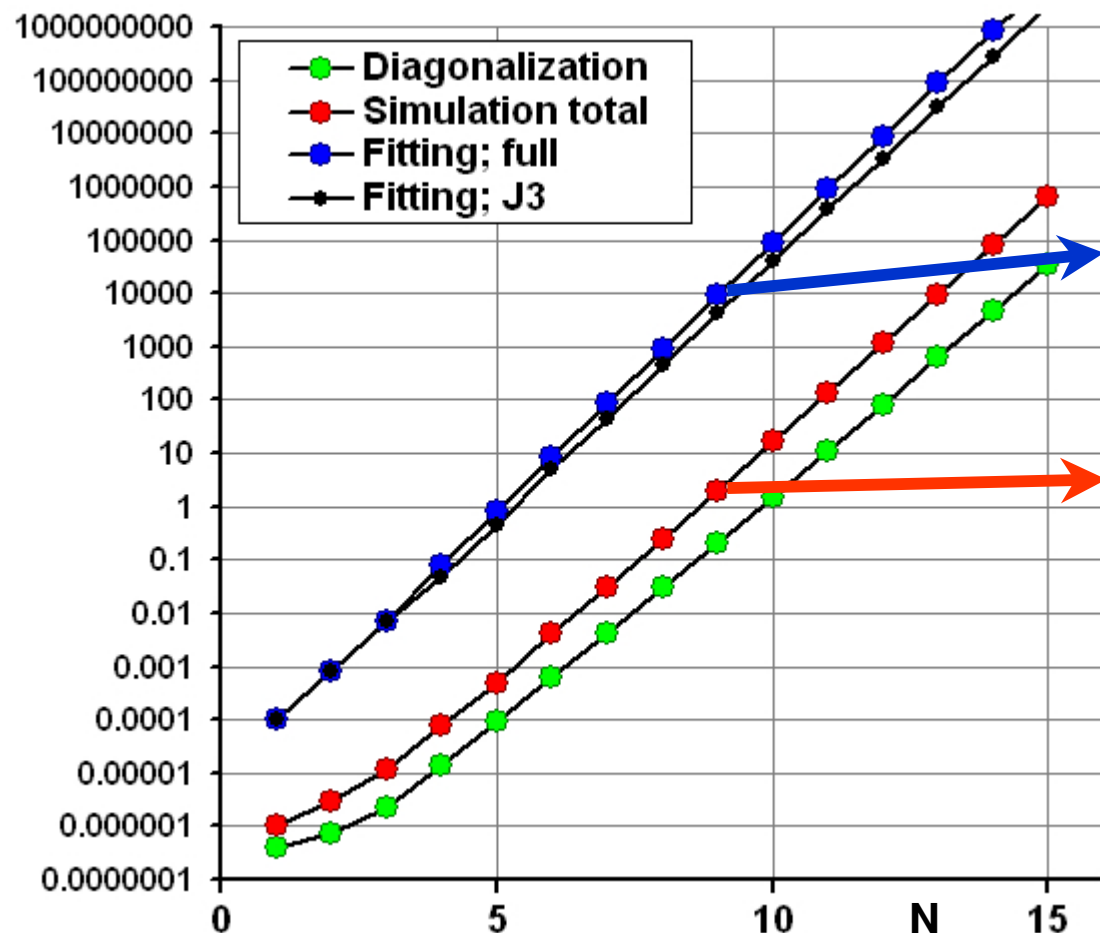
Fitting target functions
(how does one compare two spectra)

➤ **Laocoon** (Castellano '60s, Sykora 4x Assembler '60s, PANIC 1979, …)

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, Binsch, Davins 1979)

Designed to avoid multiple local minima.
Great freedom in choosing the basis of functions for the IT's:
   Powers, Lorentzians, Hermite functions, harmonic functions, …

➤ **Spectral intervals and interval functions** (work in progress)

Criteria like *"in this interval, the tallest peak is at 2.33 ppm"*

➤ **Spectral transforms** (work in progress)

New general-purpose tools

➢ **RB: Resolution Booster ™**

    Poster at ENC 2008: see **ebyte.it\stan\SS_Posters.html**

➢ **JC: J-Correlator ™ : Analog, Digital (dJC)**

    Poster at ENC 2008: see **ebyte.it\stan\SS_Posters.html**

➢ **GSD: Global Spectral Deconvolution**

    work in progress

Resolution Booster ™

RB at a glance: Improved elucidation of coupling patterns

The following Figure shows the results of applying RB to the ^1H spectrum of *dimethyl pyridine-2,5-dicarboxylate* acquired at 250 MHz. In the untreated spectrum (lower trace), the long range coupling $^5J_{5-2}$ cannot be estimated because of the lack of resolution.

2   RB Spectrum

1   Normal Spectrum

9.26   8.42   8.19
f1 (ppm)
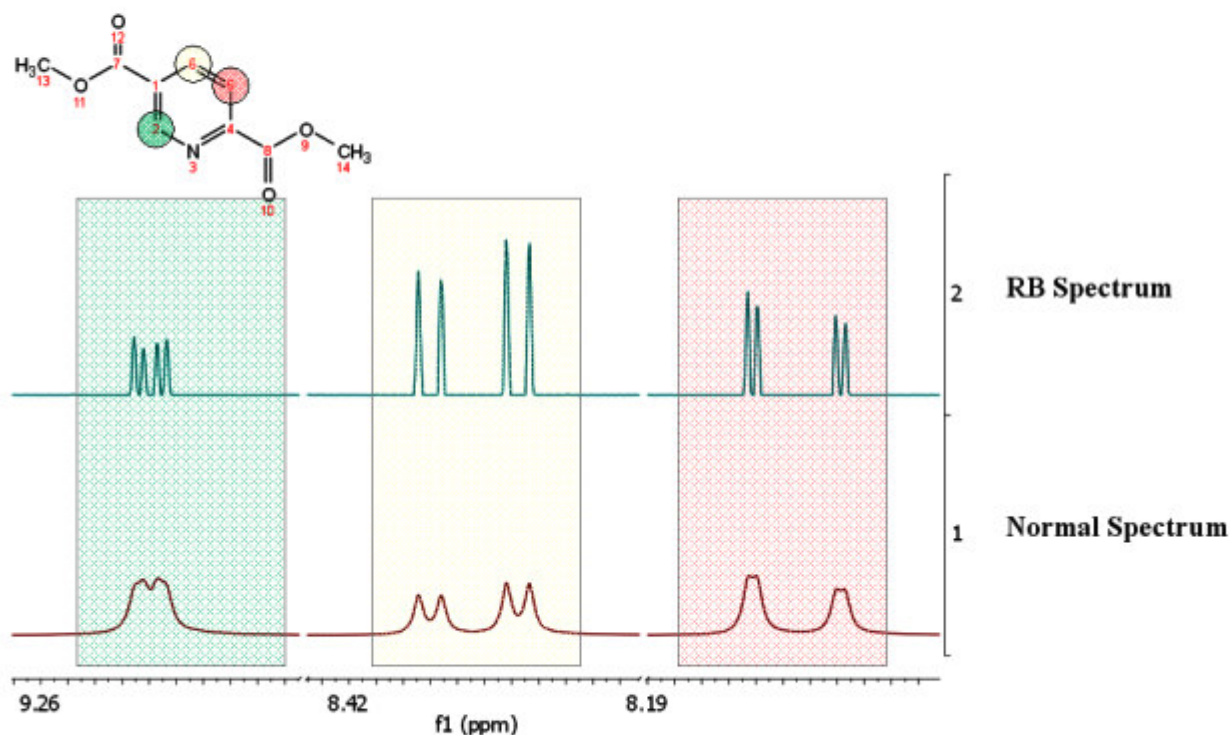
Applying RB to the same data (upper trace), spectral resolution is increased by over 200%, thus making it possible to estimate the value of the long range coupling $^5J_{5-2}$, equal to 0.82 Hz.
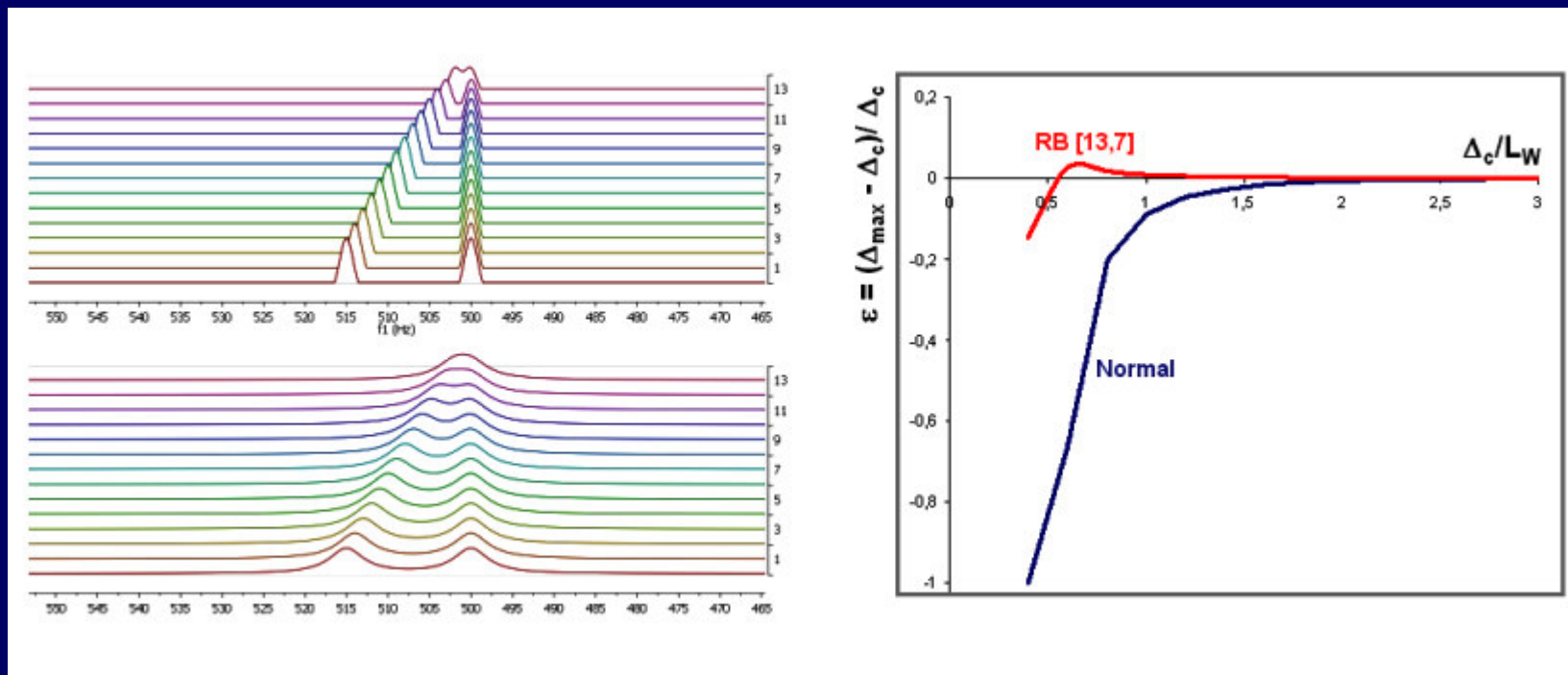
Resolution Booster™

RB at a glance: Improved recognition of spectral lines

The Figure below shows a very small fraction of a **metabolomic spectrum** to which we have applied an automated RB procedure and, subsequently, a peak-picking routine. In this way, over 900 peaks were identified in the whole spectrum and the positions of their centers were marked, all in a single command (to simplify the visualization, we show only the peak marks here, not the whole RB peaks).
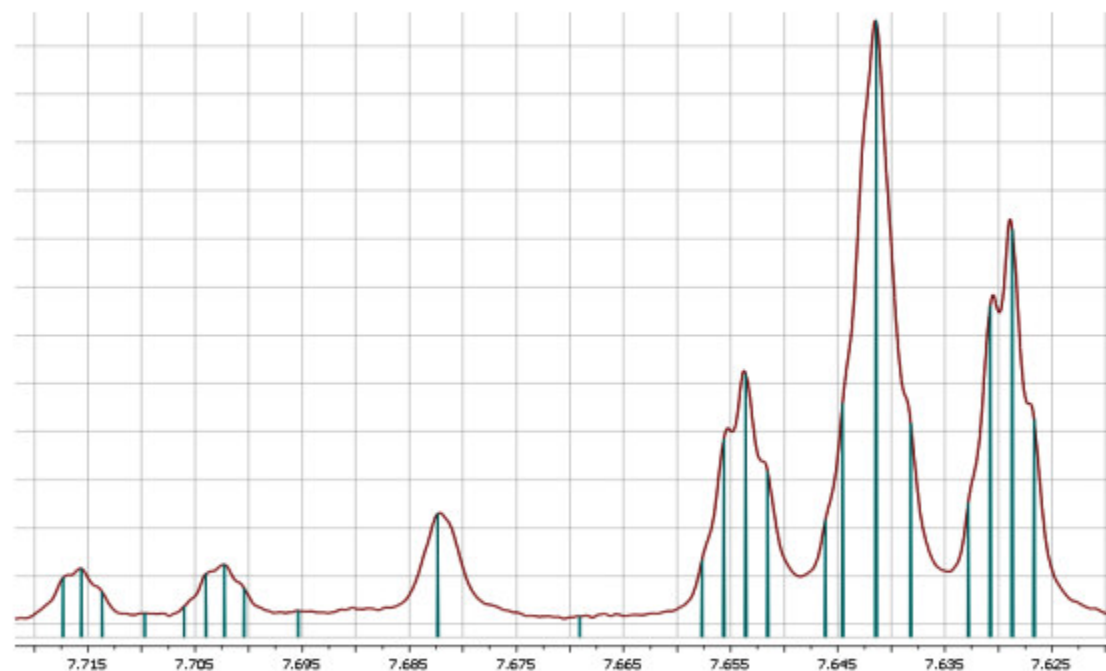
Notice that, thanks to RB, **barely resolved shoulders are correctly identified** as peaks. To human eye, some of the peak markers appear slightly off-center but we will explain below that the RB line center estimates are actually remarkably correct.

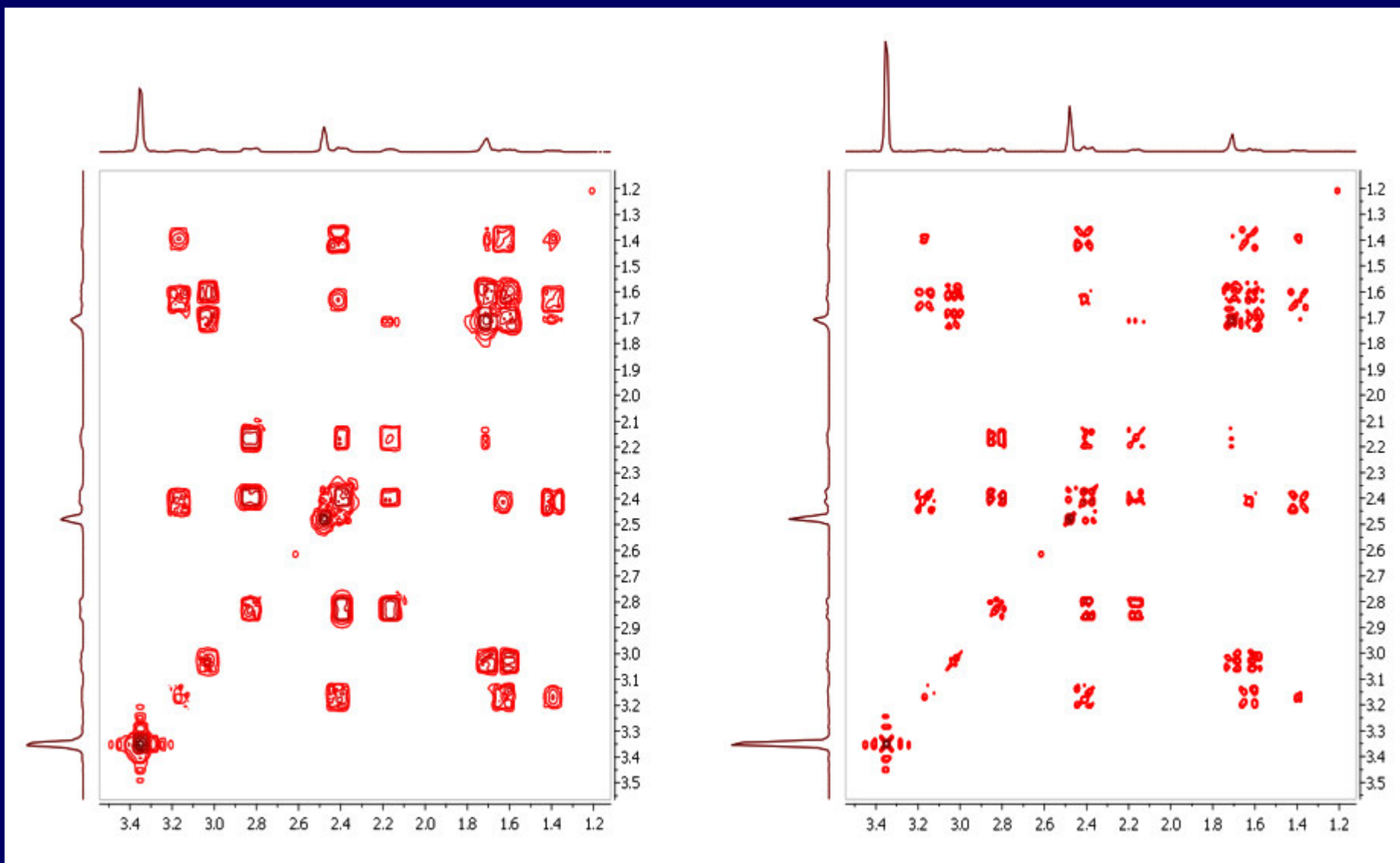# Resolution Booster™ resolving power

Δ_c … **Distance between true line centers**

L_w … **Linewidth**

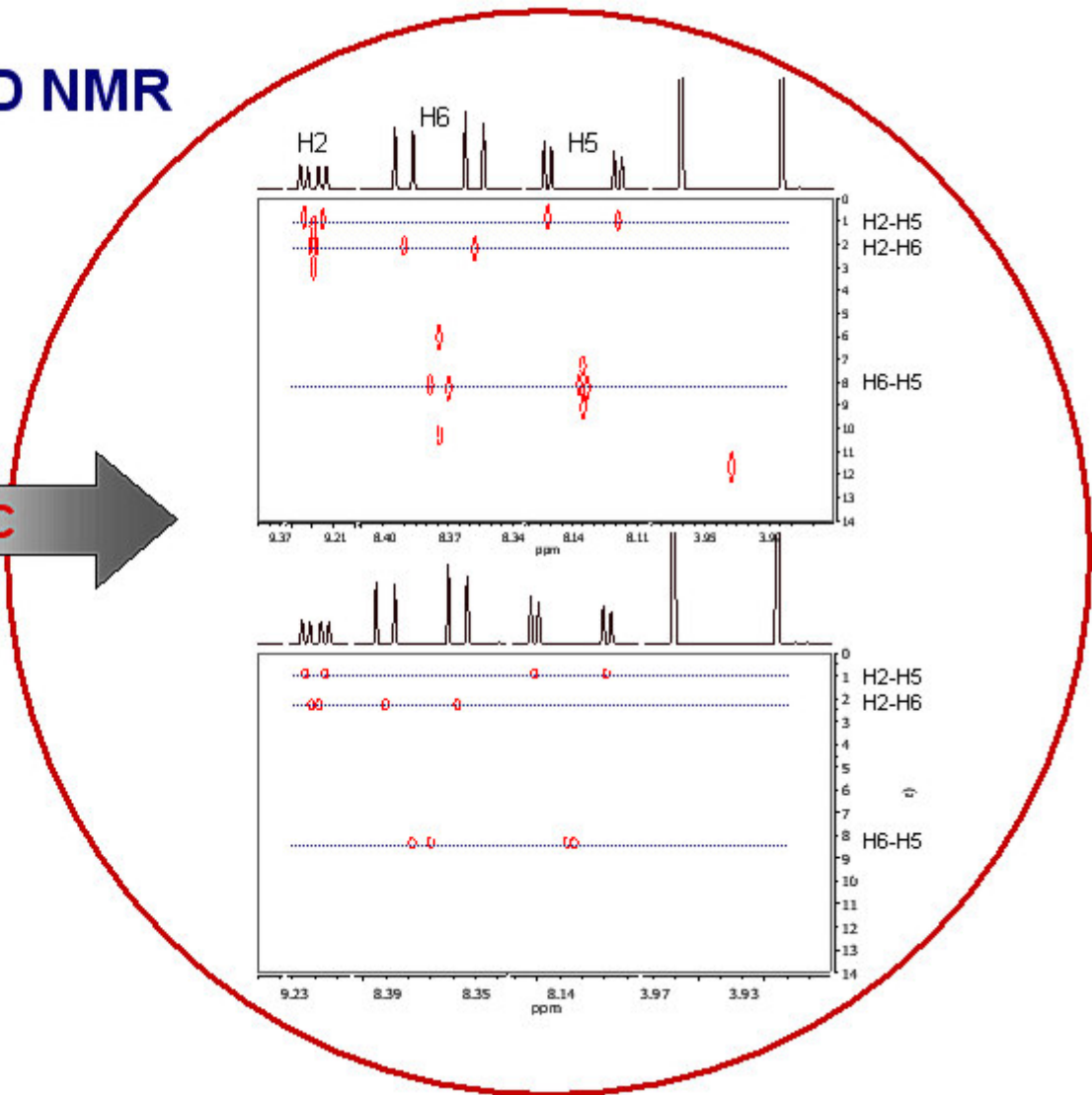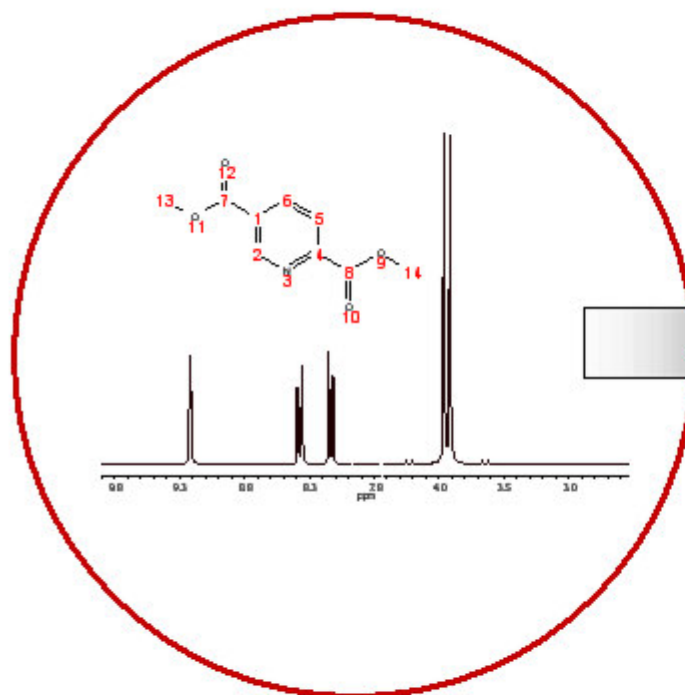ε … **Relative coalescence error**

Resolution Booster™ in 2D

J-Correlator ™

A splitting Theorem

Basic level-diagram rule for a spin-system fragment

$?\alpha\alpha?+...$

A_1

B_1

$?\beta\alpha?+...$

$?\alpha\beta?+...$

B_0

A_0

$?\beta\beta?+...$
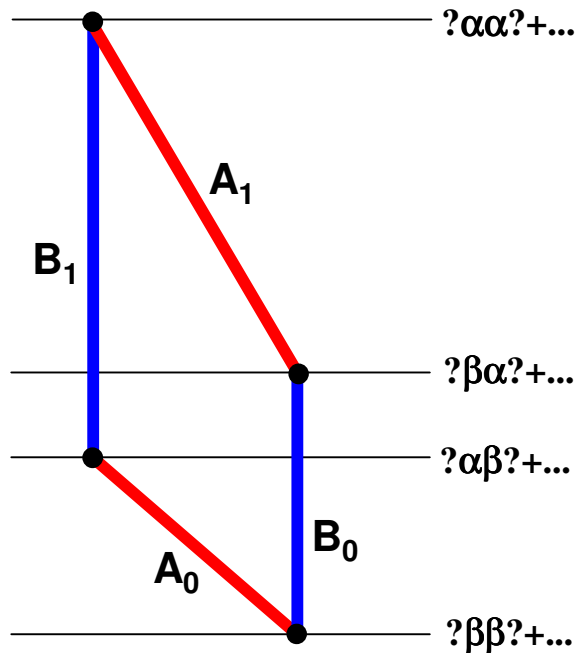
Theorem:
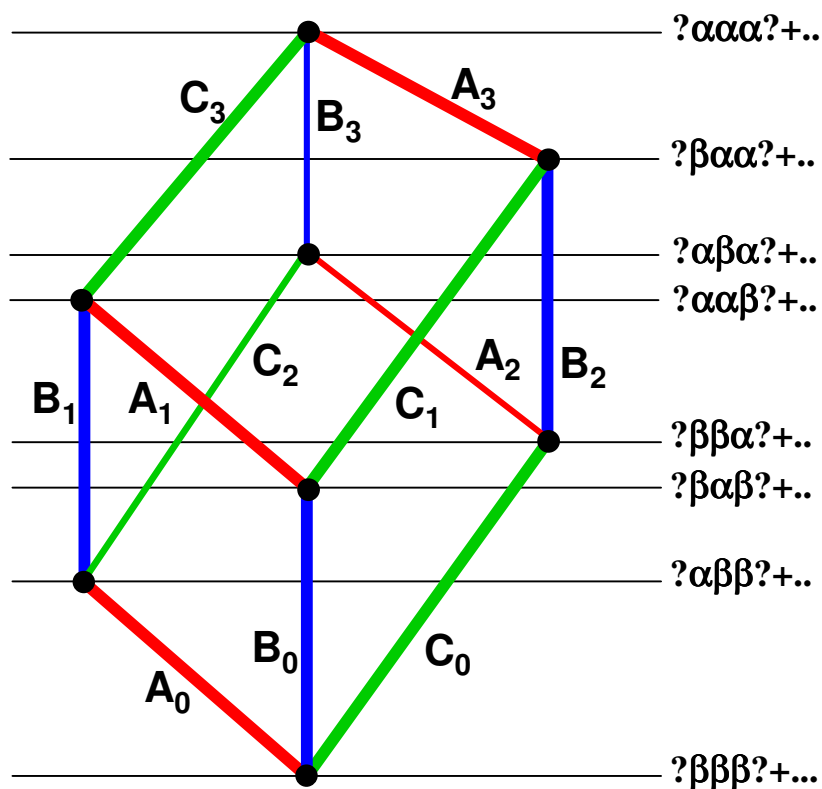
$$A_1 - A_0 = B_1 - B_0$$

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

ABC Splitting Theorem

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 !!!

GSD: Global Spectrum Deconvolution
What is it?

Essentially, it the idea is very simple:

Automatic multiplet deconvolution of the whole spectrum

Notice: the peaks do not be "real"

Problems to overcome:

Initial recognition of all significant lines (RB)
Good Initial parameters (RB+)
Sliding-window fitting

Results:

List of peaks (position, height, width, phase, shape parameters, flags)
Array of residues

GSD: Global Spectrum Deconvolution
## What can I do with it ?

- ➤ Peak-by-peak editing
- ➤ Precise phase corrections
- ➤ Generation of synthetic spectra
    - without baseline
    - with all linewiths set to the same value (for fitting)
    - with reduced linewidths (resolution enhancement)
- ➤ Explicit (non-numeric) integration over intervals
- ➤ Multiplet intensity corrections
- ➤ Explicit calculation of integral transforms (for fitting)

etc ……

Master flowchart, section 1:
automatic, basic evaluation

From **FID** to **Phased Spectrum**

(sorry, I will not talk about this)

Flowchart, section 2: automatic, basic evaluation
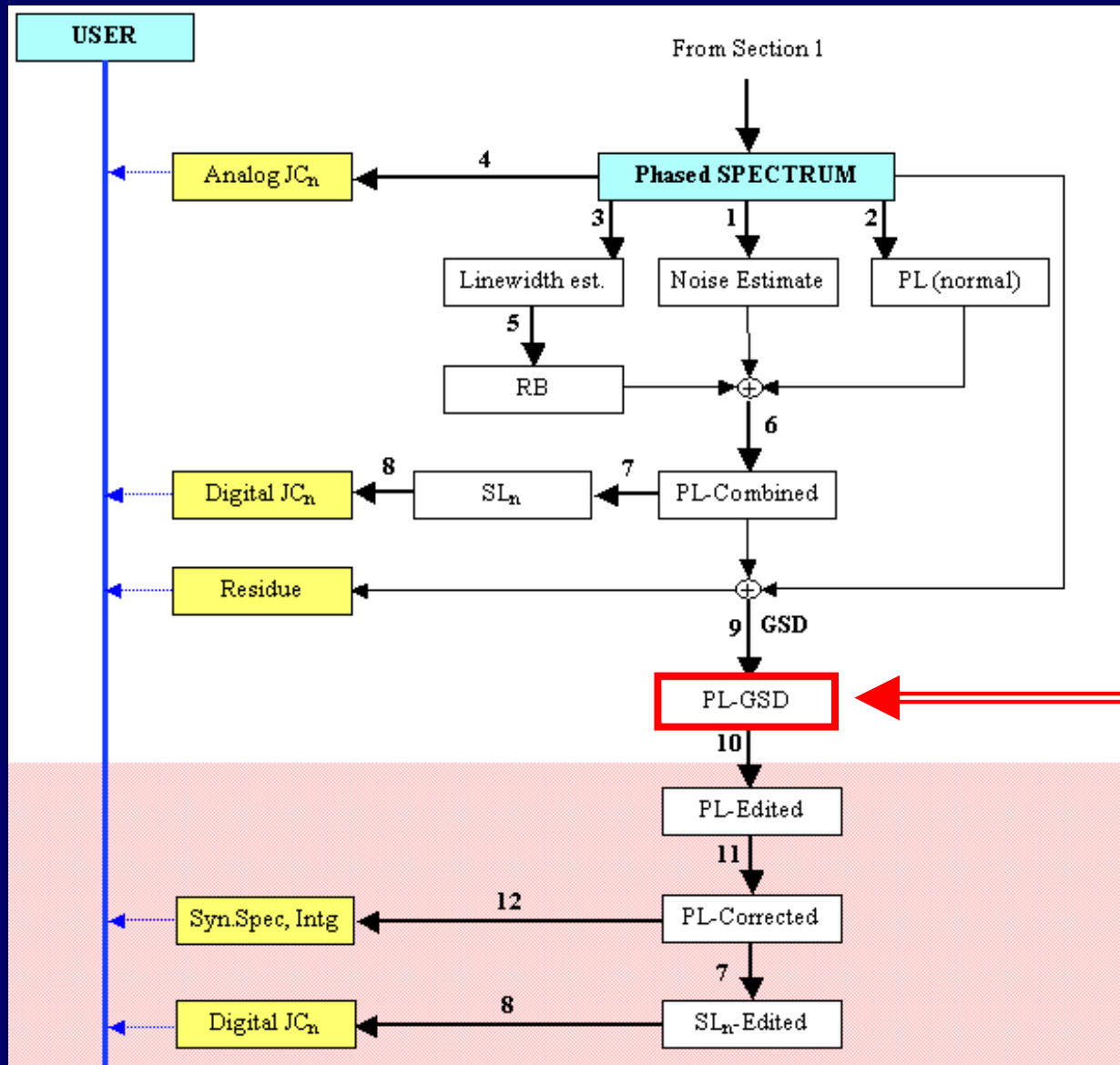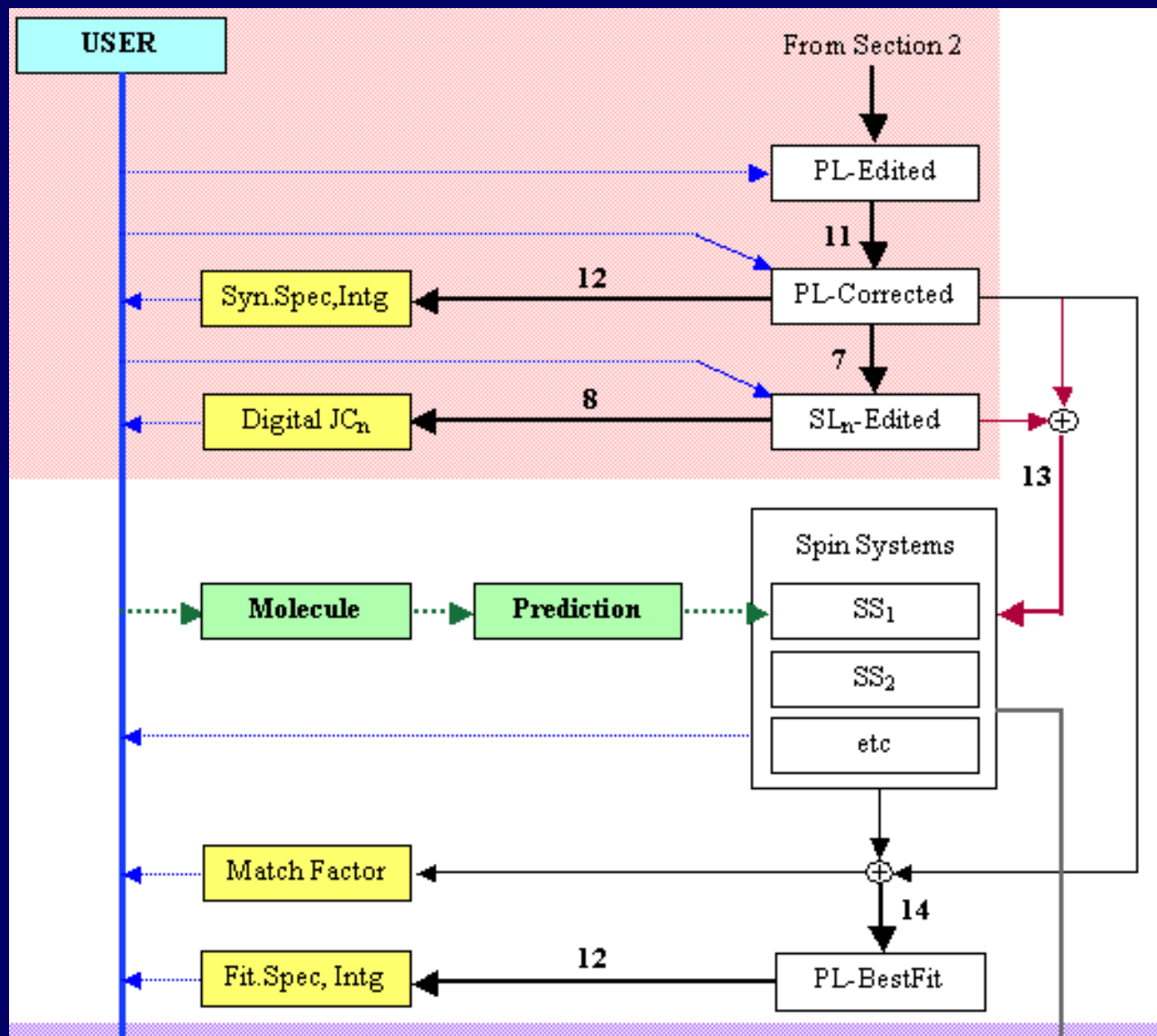

PL   Peak list
RB   Resolution Booster
SL   Splittings List
JC   J-Correlator
GSD   Global Spectral
        Deconvolution

MAIN RESULT

⇐ Overlap with Section 3
  Default editing (void)
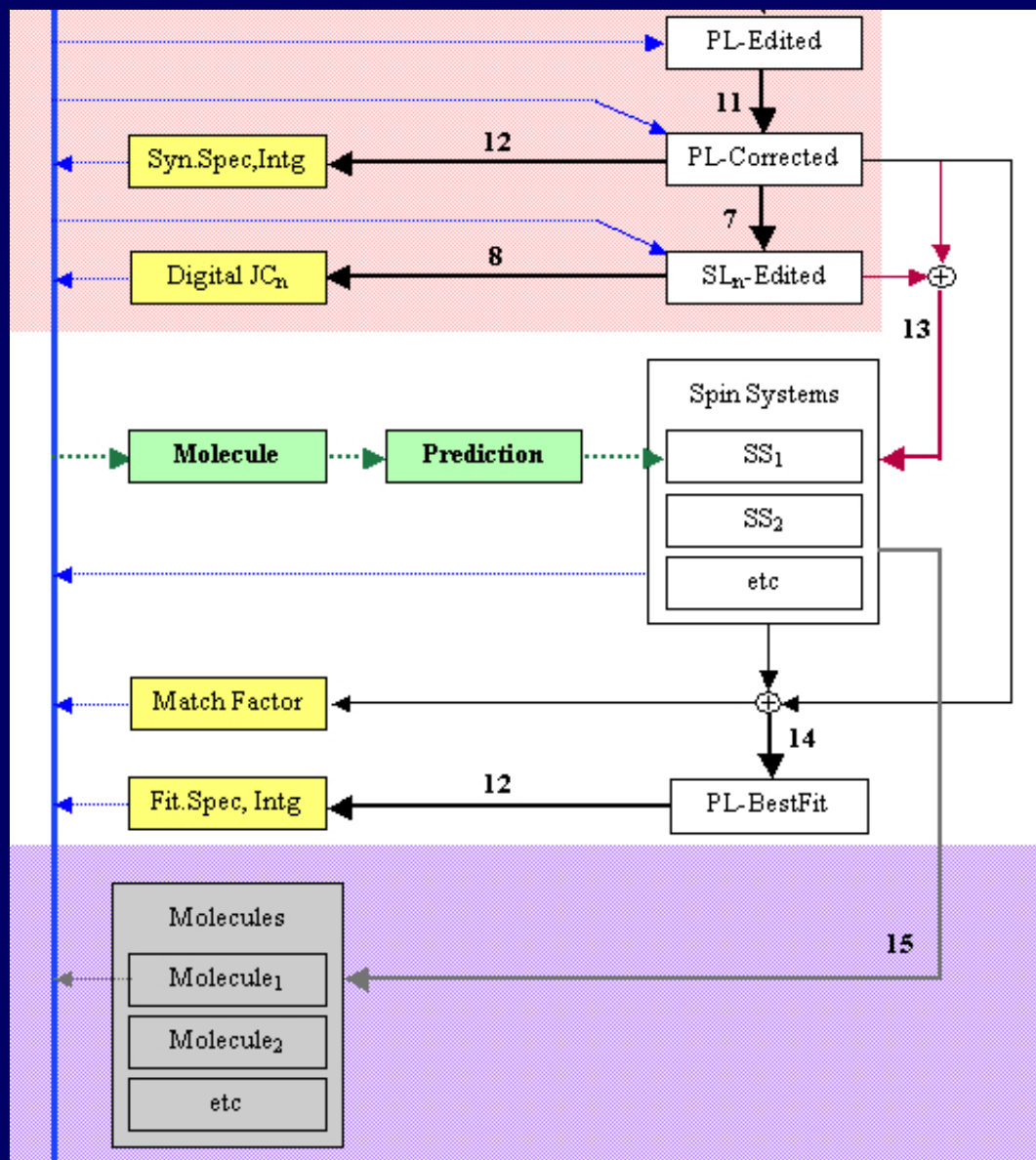
Flowchart, section 3: interactive NMR software

Iterations involve the User
(data editing)

Green … Verification Path

⇐ Simulation / Fitting

⇐ Overlap with Section 4

Flowchart, Section 4: interactive chemical software

⇐ Automatic proposal
of matching Spin Systems

⇐ List of matching Spin Systems

⇐ Section 4
from Spin Systems to Molecules

Valuable spin-offs

- ➢ **Spin-system fragmentation theorems (NMR)**

- ➢ **Splitting theorems (NMR)**

- ➢ **improved downhill-simplex algorithm (optimization theory)**

- ➢ **New noise estimators (evaluation of noisy data)**

- ➢ **New average-linewidth estimator (evaluation of noisy data)**

- ➢ **New metric for sets of real-number n-tuplets (mathematics)**

Collaborators

Mestrelab:

➢ **Santiago Dominguez** (the only one who likes deadlines)

➢ **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 Gaeda** (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) ➡️