FITTING OF SPIN SYSTEM PARAMETERS TO NMR SPECTRA REVISITED

Stanislav Sykora, Extra Byte, Italy

Early history

Simulation of NMR spectra and fitting of spectral parameters to actual experimental data are time-honored NMR spectroscopy tasks.

1948: G.E. Pake, H.S. Gutowsky:
   elucidation of **dipolar couplings in solids** and of the **line narrowing in liquids**
1950: W.G. Proctor, F.C. Yu; W.C. Dickinson: **discovery of chemical shifts**
1951: W.G. Proctor, F.C. Yu: first **observation of scalar couplings** (unexplained)
1951 - 1957: H.S. Gutowsky, D.W. McCall, Slichter, Karplus M., et al:
   elucidation of the **functional form of scalar couplings** (J's)
1953 - 1956: **W.A. Anderson: introduction of the concept of spin-Hamiltonian**
1958 - 1959: J.D. Roberts, J.N. Schoolery, H.S. Gutowsky, et al:
   wrapping up the **role of liquid-state NMR in chemical structural analysis**
   creation of first **chemical compendia of NMR parameters**
1959: the **concept of spin Hamiltonian is now universally accepted**
   and **the way to simulate spectra is well known** (in Hilbert space).
1961: S. Castellano, J.S. Waugh:
   Fortran **software for simulation of NMR spectra**
1964: S. Castellano, A.A. Bothner-By:
   Fortran **software for fitting NMR parameters to spectra** (LAOCONOOR)

Early Czech history of NMR spectra simulation/fitting

My own role (with admiration for Weston Anderson and reading all his stuff):

1964 – 1965, Institute of Macromolecular Chemistry CSAV, Praha:
   Simulation package for Zuse 25 computer, with my own compiler.
   Max 5 generic spins (½). Run time of a generic 5-spin system: overnight.

1965 – 1967, Institute of Macromolecular Chemistry CSAV, Praha:
   Simulation package for Minsk 22 computer, with my own compiler.
   It was faster, but with the same maximal limits.
   My first fitting code (for Minsk 22).

Others active in this field at that time in CSR:

1962 - …, Institute of Organic Chemistry, Praha:
   V. Špirko: addressed the problem of the uniqueness of fitting results.
   J. Schraml: wrote his own simulation programs.

Later: ???

A call for help

Please, help me to identify others and complete this history chapter

What I did in later years

1974, Institute of Physics, University of Basel:
Fortran **NMR simulation package** independent of Castellano's,
eventually **merged with vibrational normal-coordinates calculations**.
Max 7 generic spins (½).

1975, Institute of Physics of University of Basel:
Fitting spectra of molecules dissolved in oriented nematic solvents.
For that, launched the **Integral Transforms fitting method** (IT).

1979, Bruker Spectrospin Italiana, Milano, and Spectrospin AG, Zurich:
Together with Jürg Vogt, we wrote the **Bruker program PANIC for
ASPECT mini-computers** (assembler, using our own floating point system).
It was of LAOCONOOR type approach with manual selection of peaks to fit.
It became very wide-spread and used until the end of the 80's.

2006 - 2007, Extra Byte, commissioned by Mestrelab Research (Mnova):
C++ SpinStar spectra simulation package, incorporating all my accumulated
experience. It runs on PC's over 10^6 times faster than Zuse 25, yet can do
max 10 - 11 generic spins (½) before resorting to fragmentation.

Further developments around the world

In 1975 - 1977, the **IT method** was explored by G. Binsch and D.S. Stephenson (Uni Muenchen) and **incorporated into the simulation and fitting package DAVINS**, published in QCPE, which gave it broader notoriety.

The IT approach was also **adopted by PERCH** into their fitting algorithms and through them it makes part of the current Bruker software.

Programs for **simulation of dynamic systems** ("chemical exchange") based on **Liouville space representations** were also written and made available (for example, Alex Bain's MEXICO and its successors). They are limited to 5-7 spins.

Current state-of-the-art

Simulations

Plain Hilbert-space handling (the same as Weston Anderson's) is still the universal approach. Current maximum size or fully manageable spin-systems is 10 - 12 spins (beyond that, execution times exceed 1'). Considering the development rate of computer hardware, this limit is likely to grow by about 1 spin every 10 years.

Fragmentation methods extend this to practically any system, without significantly affecting the precision, but the longest execution times for the «small molecules» of today (< 1000 Daltons) still get close to, or even exceed, 1 second, which is an unsatisfactory performance to embark on a practical fitting.

Liouville-space handling remains marginal (hindered by its strict size limitations).
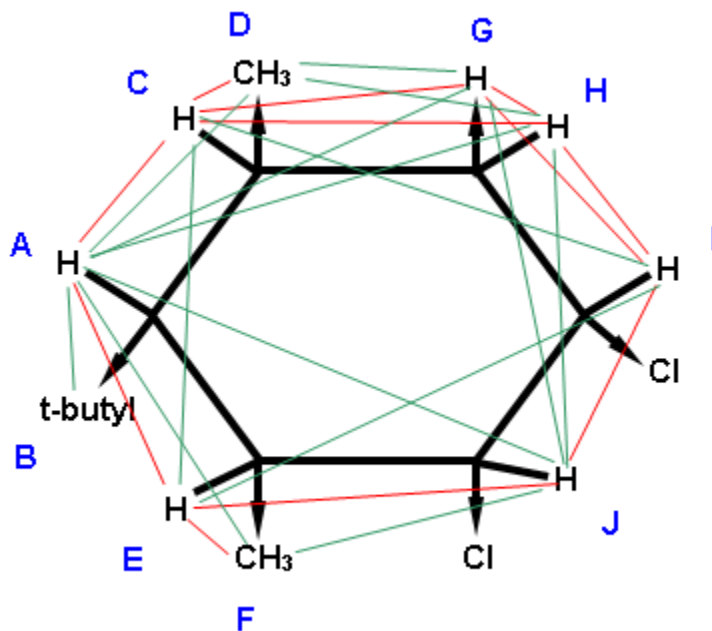
Fitting

The Integral Transforms approach dominates the fitting scene.
As far as I know, there is nothing better in use today.

Hence, **simulation is stuck at where it was over 50 years ago, and fitting at almost 40 years ago**. This is the worst development performance I ever heard about, considering that the current state of the art is not at all satisfactory !!!

A molecule whose spin system still can't be handled

22-spin system with 10 groups of equivalent-nuclei. A B9 C D3 E F3 G H I J.

This molecule **could exist**. The arrows indicate orientation with respect to the puckered ring "plane" (due to the chirality of all the ring carbons, there are 2^5 = 32 stereoisomers, each with L and R optical form; the arrows select just one of them). Red lines indicate 2- and 3-bond couplings (assumed to be all quite large). Green lines are the 4-bond couplings, all of which *could* be present.

The form of a spin-system Hamiltonian in liquid-state NMR of diamagnetic organic molecules
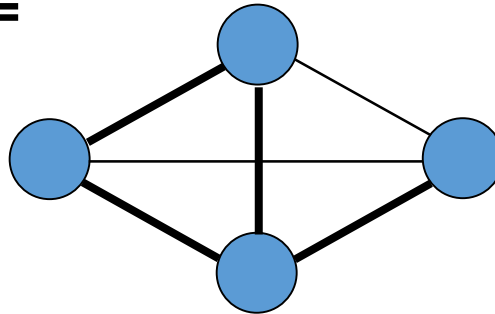
Just a reminder of what I am talking about:

$$H = \sum_i \delta_i S^z_i + \sum_{ij} J_{ij} (S_i \cdot S_j)$$

$$= \sum_i \delta_i S^z_i + \sum_{ij} J_{ij} (S^z_i S^z_j) + \sum_{i<j} J_{ij} (S^+_i S^-_j + S^+_j S^-_i)$$

For N nuclei we have:

N chemical shifts δ, and up to N(N-1)/2 coupling constants J,

some of which possibly negligibly small.

Spin-system as a weighted (König) graph

H =

Edges … weighed by J's

Vertices … equivalent groups
… weighed by δ's
… have multiplicities
… have spin S

Dimensions of the problem
for N nuclides with spin S = ½

Maximum matrix …
 the largest matrix to diagonalize

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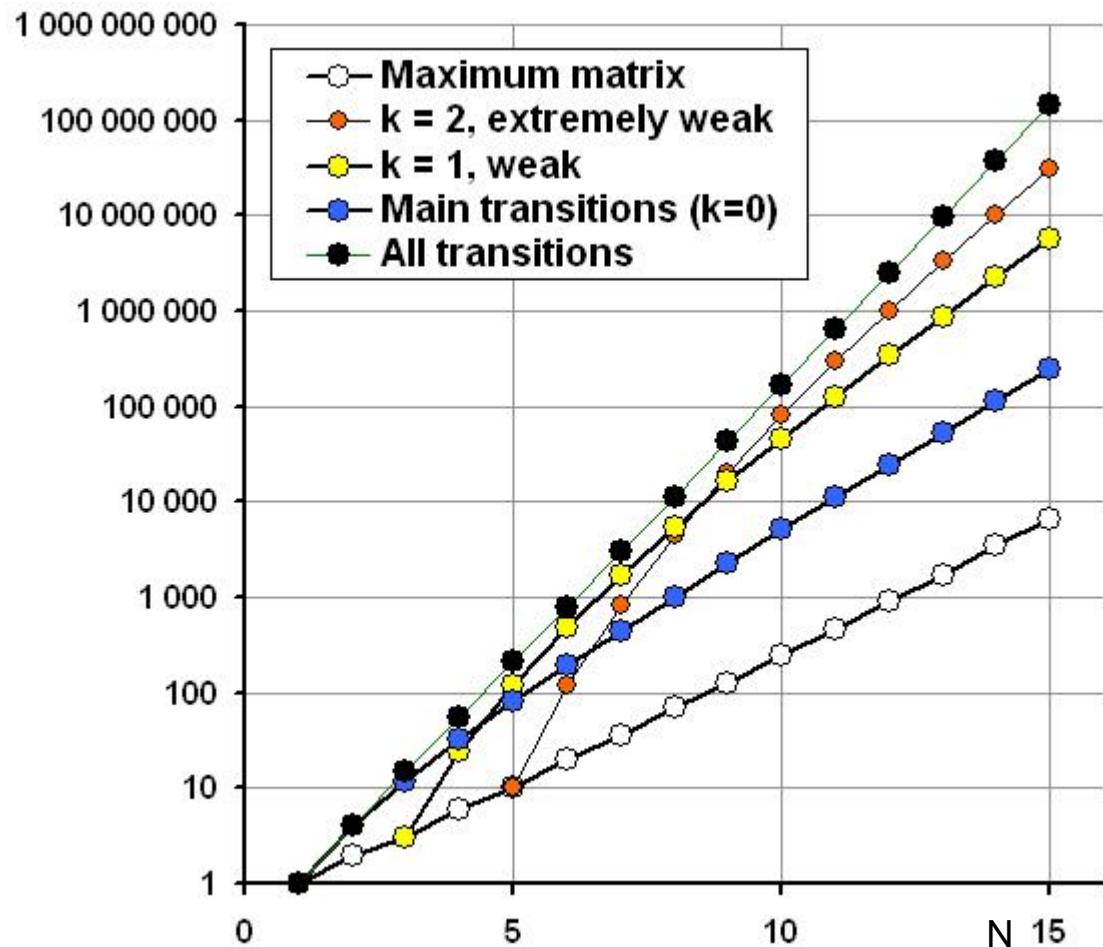 one spin goes up, while k pair
 of spins undergo exchange

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

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

The counts are HUGE !

Legend:
- Maximum matrix
- k = 2, extremely weak
- k = 1, weak
- Main transitions (k=0)
- All transitions

Hilbert-space spin-system simulation performance
for N nuclides with spin S = ½
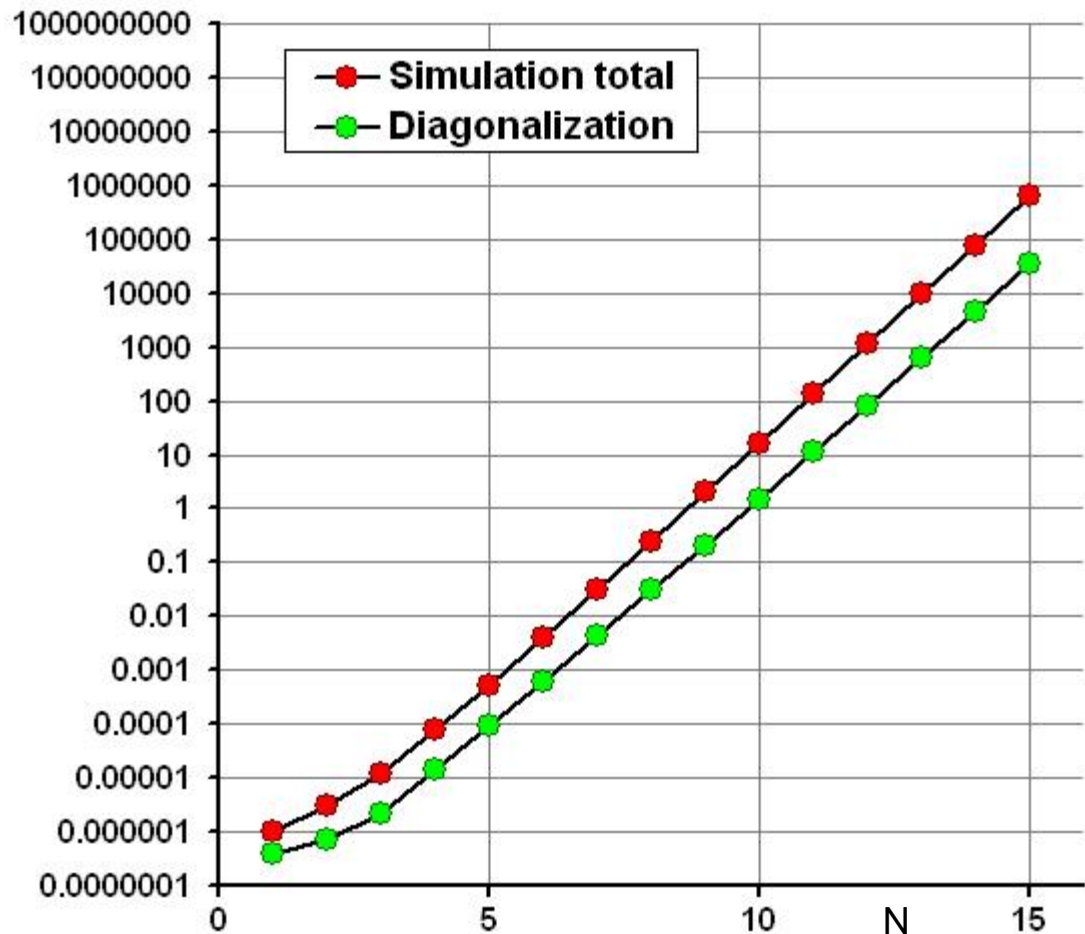
Execution load expressed

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 !
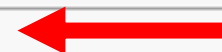
Spin system fragmentation at work

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

Todays requirements

- Need to handle very **high throughput**. Especially in pharmacological industries (but not only) they must analyze **tens to hundreds of samples per day**. Any software that takes more than 10 minutes per sample is useless.

- There must exist a **fully automatic mode**, as well as of a **computer-aided mode**.

- One may not rely of sample purity, nor on high quality of the spectra! The **presence of impurities and artifacts is a rule** and all evaluation algorithms – including fitting – must be «resistant» to them (an IT advantage).

- Any **pre-editing of spectral peaks must be fully automatic** (GSD is The Best).

- All **numeric results must include probable error estimates**. The widths of confidence interval of coupling constants, in particular, vary enormously.

- The **algorithms must be able to analyze the uniqueness of solutions**.

Spin-systems simulation and fitting performance with/without fragmentation

for N nuclides with spin S = ½

Execution load expressed
  in 100 MFlop units

On a 1 GHz PC
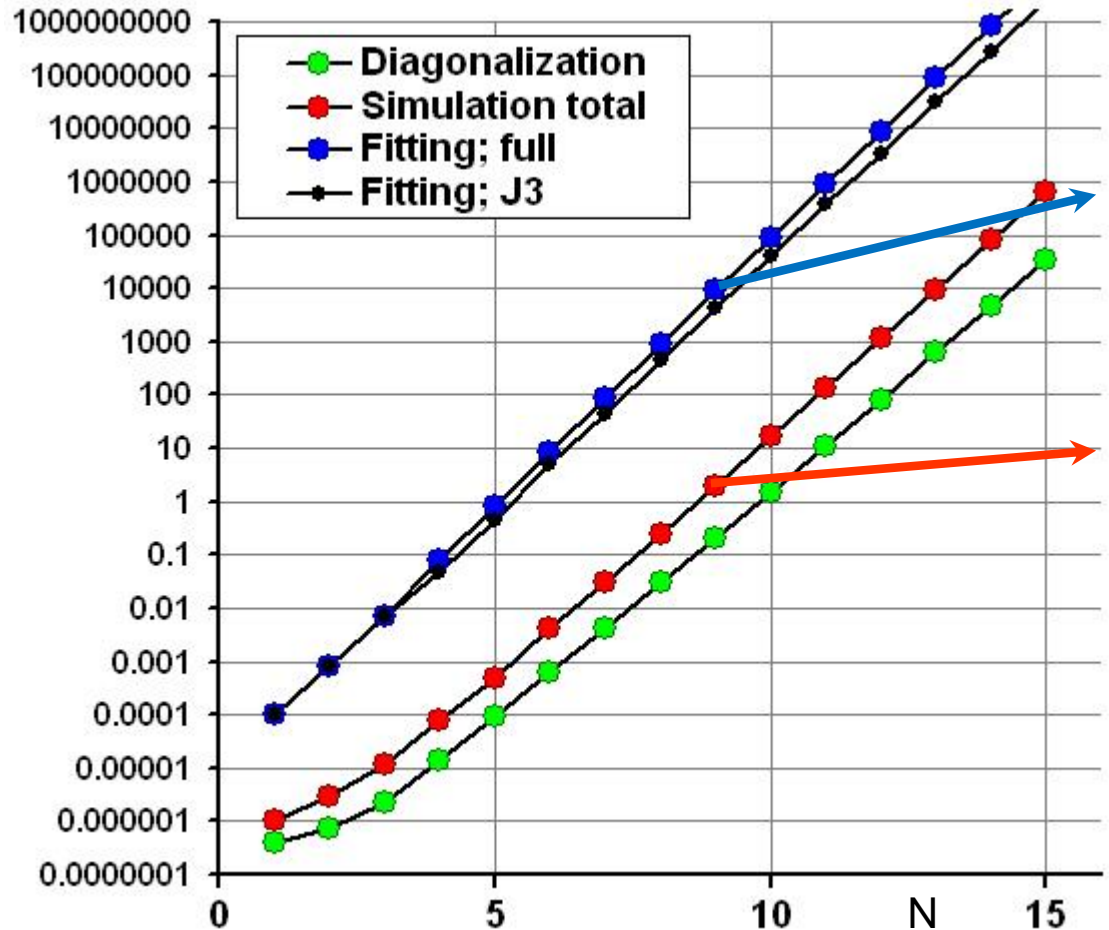 100 MFlops take about 1 second

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 !
They should be cut by 1000

Current work:
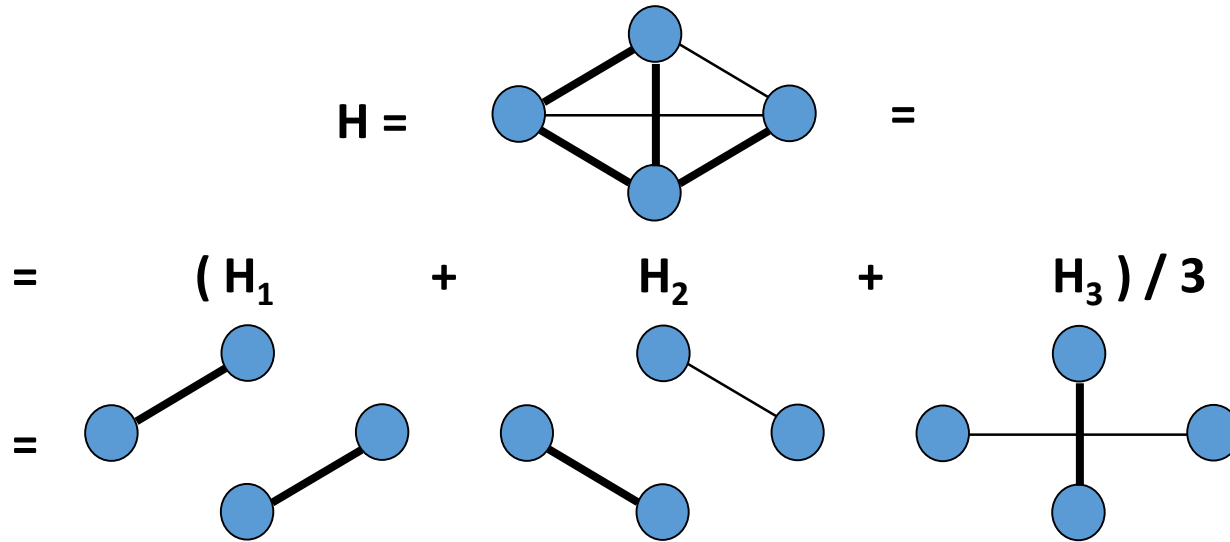# New algorithms for a faster and better fitting

- Simulation must be done MUCH faster. This is an absolute imperative and at least 3 orders of magnitude are needed!

- While Integral Transforms (IT) have many advantages (reduction of false minima, robustness with respect to impurities and other artifacts), they nevertheless need to be refined (work in progress).
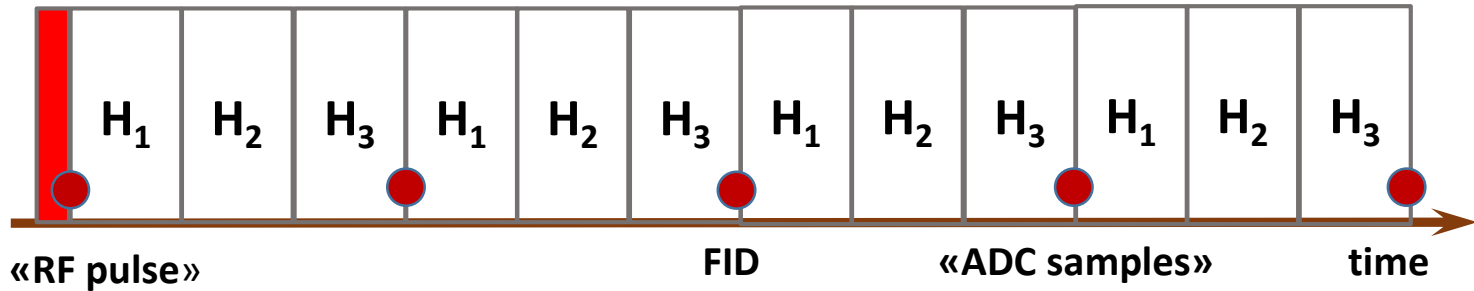
The Goal:

After 60 years of standard simulations, and 40 years of Integral Transforms, it is high time to do finally something new.

I am after a truly major performance jump.

Accelerating spin-system simulations: Hamiltonian time slicing and averaging

$$H = \quad = \quad (H_1 \quad + \quad H_2 \quad + \quad H_3) / 3$$

Note: All J's in H_1, H_2, H_3 are 3 times larger than in H

| H_1 | H_2 | H_3 | H_1 | H_2 | H_3 | H_1 | H_2 | H_3 | H_1 | H_2 | H_3 |

«RF pulse»   FID   «ADC samples»   time

The recipe

- **Express the spin system as the average of much simpler sub-systems** composed of only K mutually disjoined two-spin, or even single-spin, graphs with shifts and J's modified so as to guarantee that the average value is the actual full Hamiltonian H. Note that this is always possible, but also quite easy.

- **Assign a slice of time to every of the K sub-graphs**.

- **Simulate the time-domain signal** (FID) as a **cyclic progression of disjoined evolution intervals**, using within each of them the trivial evolution operator of the corresponding sub-Hamiltonian H_k.

- Make each sub-Hamiltonian H_k act on spin functions which are combinations of its own eigenfunctions.

- When passing from interval H_k to H_{k+1}, **project the final state of the previous interval into the eigenfunctions base of the next interval** (the new starting state).

- **Every time a full cycle of slices was completed, sample the «signal».**
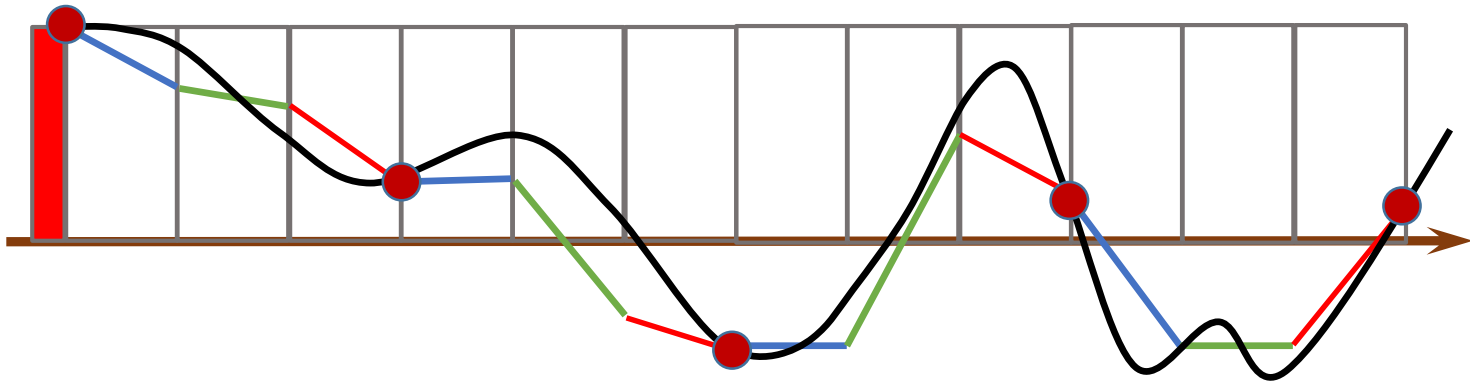
Pros

1. **All steps of the recipe are low-dimensional, extremely fast, and easy.**

2. There are no large matrices to handle. Indeed, **there are NO matrices**.

3. All **formulas are explicit**.

4. The task **grows approximately linearly with the number of nuclei N**, not with 2^(2N) as in Hilbert space, nor with 2^(4N) as in Liouville space.

5. The starting Hamiltonian H can itself change during the evolution, which is **an easy way to handle dynamic systems**. The capability to do so is implicit in the method and takes almost no extra computational effort. The same applies to NMR diffusion experiments, fast kinetics, etc.

6. RF pulses can be part of the set of slices. Therefore, **the same software can be used to simulate any NMR experiment** – nD, selective pulses, $T_{1\rho}$, decoupling, many types of artifacts, etc.

7. The idea might yet spill over into general Quantum Mechanics of many body systems and I might become really famous.

Cons

1. **The method is of course an approximation**. What happens is something similar to moving along a generic path while constrained to a rectangular grid. It is also vaguely reminiscent of CORDIC algorithms. Preliminary tests show that, as an approximation, it is amazingly good, probably much better than any fragmentation (proving this is work in progress).

2. The time-varying part of the time-sliced Hamiltonian might give rise to **fake relaxation phenomena** (a false contribution to relaxation rates). These should be kept smaller than the true relaxation rates.

I am not aware of any other potential problems.

Why should it work

Simulating FID's of mutually isolated AB pairs takes: max 10 ns per pair per point

In a 40 spin system we might have 20 AB pairs per slice: 200 ns/slices

Should we need 10 slices per point, that would mean: 2 us/point

Thus, a 32K FID simulation of a 40 spin system might take: 64 ms

In a conventional way, the same system would take about: 10^{25} s (3e14 millennia)

In a conventional way with fragmentation (current state):   1 – 5 s

This is not yet what I want, but it comes close.

Especially since the estimates are based on preliminary tests with non-optimized code!

Thank you for your patience and attention