The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.

P. A. M. Dirac, 1929
Collaborators and funding

University of Southampton
Joern Werner
Giuseppe Pileio
Maria Concistre

University of Southampton
Luke Edwards
Zenawi Welderufael
Gareth Charnock
Dmitry Savostyanov
Andreas Biternas
David Goodwin

University of Oxford
Peter Hore
Hannah Hogben

University of Oxford

University of Nottingham
Walter Kockenberger
Alexander Karabanov

Nanyang University
Konstantin Pervushin

MPIBPC
Christian Griesinger
Donghan Lee
I think you and Uhlenbeck
have been very lucky to get your
spinning electron published and talked
about before Pauli heard of it. It
appears that more than a year ago
Kronig believed in the spinning electron
and worked out something; the first
person he showed it to was Pauli.
Pauli ridiculed the whole thing so
much that the first person became also
the last and no one else heard anything
of it. Which all goes to show that
the infallibility of the Deity does not
extend to his self-styled vicar on earth.
Prologue

Current simulation capabilities

up to five spins accurately (Liouville space) and up to ten spins approximately (Hilbert space)

Typical systems requiring simulation

protein spin systems (200+ sparsely coupled spins), large aromatic radicals (20+ densely coupled spins)
Magnetic resonance simulations - features

Long-range dynamics far from the ground state with a time-dependent Hamiltonian

\[ t_{\text{max}} \gg \left\| \hat{H} \right\|^{-1}, \quad E_i \ll kT \]

Very complicated 3D coupling graphs with multiple loops and complex anisotropies

\[ \hat{H}_{ij} = \left( \hat{L}^{(i)}_X, \hat{L}^{(i)}_Y, \hat{L}^{(i)}_Z \right) \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \hat{L}^{(j)}_X \\ \hat{L}^{(j)}_Y \\ \hat{L}^{(j)}_Z \end{pmatrix} \]

Strict requirement for correct phases through the trajectory

\[ \frac{\partial}{\partial t} \hat{\rho}(t) = -i \left[ \hat{H}, \hat{\rho}(t) \right] \implies \hat{\rho}(t + \Delta t) = \exp(-i\hat{H}\Delta t) \hat{\rho}(t) \exp(i\hat{H}\Delta t) \]

Significant spin interactions span ten orders of magnitude

\[ \hat{H}_{\text{Zeeman}} \sim \text{GHz} \quad \hat{H}_{\text{Quad}} \sim \text{MHz} \quad \hat{H}_{\text{DD}} \sim \text{kHz} \quad \hat{H}_j \sim \text{Hz} \]
Exponential scaling problem

The objective is to solve

\[
\frac{\partial \hat{\rho}}{\partial t} = -\frac{i}{\hbar} [\hat{H}, \hat{\rho}] \quad \text{or} \quad \frac{\partial \hat{\rho}}{\partial t} = -\frac{i}{\hbar} \hat{L} \hat{\rho} \quad \text{with} \quad \hat{L} = \hat{H} \otimes \hat{E} - \hat{E} \otimes \hat{H}^T
\]

With operator matrices generated as

\[
\hat{S}_{\pm}^{(k)} = E_1 \otimes \ldots E_{k-1} \otimes \sigma_\pm \otimes E_{k+1} \ldots \otimes E_N
\]

The matrix dimension is

\[
2^n \times 2^n \quad \text{in Hilbert space}
\]
\[
4^n \times 4^n \quad \text{in Liouville space}
\]

Fundamentally no way beyond about 15 spins. Furthermore...

...for systems with over 50 spins even a sparse Hamiltonian cannot so much as be stored, let alone diagonalized.

Yet somehow we can do it...

Quantum mechanical NMR simulation algorithm for protein-size spin systems

Luke J. Edwards a,b, D.V. Savostyanov b, Z.T. Welderufael b, Donghan Lee c, Ilya Kuprov b,*

a Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK
b School of Chemistry, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
c Department of NMR Based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Goettingen, Germany

As of 2015, ubiquitin simulations take a few hours.
Sparse matrices

For any large $n$-spin system, even if every spin is coupled to every other, the Hamiltonian is mostly zeroes...

The number of terms in the Hamiltonian is at most

$$\frac{n^2 + 3n}{2}$$

The maximum number of operators in the Hamiltonian

$$3n^2 + 5n$$

Upper bound on the number of non-zeros in the Hamiltonian

$$N_{NZ} \leq (4n^2 + 5n) \prod_{k=1}^{n} (2S_k + 1)$$

Upper bound on the Hamiltonian density

$$d_{\hat{H}} \leq \frac{N_{NZ}}{\left[ \prod_{k=1}^{n} (2S_k + 1) \right]^2} = \frac{4n^2 + 5n}{\prod_{k=1}^{n} (2S_k + 1)}$$

Liouvillian matrix of a typical liquid-state NMR system (blue dots denote non-zero elements). The density (fraction of non-zeros) of this matrix is 0.0005. A full matrix of this dimension would overflow the system memory.
## Scaling of elementary operations

<table>
<thead>
<tr>
<th>Operation</th>
<th>CPU cost (FLOPs)</th>
<th>Memory cost</th>
<th>Sparse cost</th>
<th>Parallel execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix-matrix multiplication</td>
<td>$O(n^3)$</td>
<td>$O(nnz)$</td>
<td>LOWER</td>
<td>EASY</td>
</tr>
<tr>
<td>Matrix-vector multiplication</td>
<td>$O(n^2)$</td>
<td>$O(nnz)$</td>
<td>LOWER</td>
<td>EASY</td>
</tr>
<tr>
<td>exp(matrix)-vector multiplication</td>
<td>$O(n^2)$</td>
<td>$O(nnz)$</td>
<td>LOWER</td>
<td>EASY</td>
</tr>
<tr>
<td>Matrix exponentiation</td>
<td>$O(n^3)$</td>
<td>$O(nnz)$</td>
<td>LOWER</td>
<td>EASY</td>
</tr>
<tr>
<td>Most matrix factorizations</td>
<td>$O(n^3)$</td>
<td>$O(n^3)$</td>
<td>SAME</td>
<td>HARD</td>
</tr>
</tbody>
</table>

Matrix factorizations (LU, Householder, SVD, diagonalization, etc.) are very expensive, do not benefit from sparse storage (the results are nearly always dense) and show poor parallelization scaling, particularly on cluster architectures. For algorithms based on diagonalization, the game is usually over at about ten spins.
Memory arrangement around the CPU

Time in a supercomputer is measured in CPU clock cycles ("clocks"). One clock is the time it takes a ~3 GHz CPU to tick – about 0.3 nanoseconds. A few multiplications, additions or other operations are performed per clock.

The speed of light in copper is about two thirds of its vacuum value, meaning that the round trip to the memory banks (about 20 cm) takes at least a 1.5 nanoseconds – five clocks. If memory access is not organized well, the CPU will spend most of the time waiting for data.

Linear sequential memory access is essential.

Front Side Bus architecture used in Intel Pentium chips.

Point-to-point interconnect used in modern Intel and AMD chips.

DDR3 memory chip.
**Memory bandwidth and latency**

Latency is the time it takes for the data requested by the CPU to start arriving.

Bandwidth is the rate at which the data arrives.

<table>
<thead>
<tr>
<th>Location</th>
<th>Latency</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU registers</td>
<td>1 clock</td>
<td>-</td>
</tr>
<tr>
<td>L1 cache</td>
<td>1 clock</td>
<td>150 GB/s</td>
</tr>
<tr>
<td>L2 cache</td>
<td>3 clocks</td>
<td>50 GB/s</td>
</tr>
<tr>
<td>L3/L4 cache</td>
<td>10 clocks</td>
<td>50 GB/s</td>
</tr>
<tr>
<td>Own RAM</td>
<td>100 clocks</td>
<td>10 GB/s</td>
</tr>
<tr>
<td>NUMA RAM</td>
<td>150 clocks</td>
<td>10 GB/s</td>
</tr>
<tr>
<td>Other node RAM</td>
<td>10k clocks</td>
<td>1 GB/s</td>
</tr>
<tr>
<td>Flash storage</td>
<td>10k clocks</td>
<td>1 GB/s</td>
</tr>
<tr>
<td>Magnetic storage</td>
<td>10k clocks</td>
<td>0.1 GB/s</td>
</tr>
</tbody>
</table>

Smaller programs and datasets are processed faster because they fit into a higher level cache.

Random hits to the main RAM can make a program very slow.

(Physical layout of IBM zEnterprise 196 hub chip.)

(approximate data for an Intel Nehalem system)
Why exponentiate a matrix?

The formal solution to the time-independent Schrödinger equation is

$$\frac{\partial}{\partial t} |\psi\rangle = -\frac{i}{\hbar} \hat{H} |\psi\rangle \quad \Rightarrow \quad |\psi\rangle = e^{\frac{-i}{\hbar} \hat{H} t} |\psi_0\rangle$$

In the time-dependent case, a piecewise-constant approximation is often employed

$$|\psi\rangle = e^{\frac{-i}{\hbar} \hat{H}(t_n) t} e^{\frac{-i}{\hbar} \hat{H}(t_{n-1}) t} \cdots e^{\frac{-i}{\hbar} \hat{H}(t_2) t} e^{\frac{-i}{\hbar} \hat{H}(t_1) t} |\psi_0\rangle$$

For the Liouville - von Neumann equation therefore

$$|\psi\rangle \langle \psi | = e^{\frac{-i}{\hbar} \hat{H}(t_n) t} \cdots e^{\frac{-i}{\hbar} \hat{H}(t_2) t} \langle \psi_0 | e^{\frac{i}{\hbar} \hat{H}(t_1) t} \cdots e^{\frac{i}{\hbar} \hat{H}(t_n) t}$$

...meaning that a single time propagation step in a typical NMR simulation requires the calculation of at least one exponential of a VERY LARGE matrix. The way this used to be done in NMR is:

$$\exp[-iHt] = \exp[-iS^{-1}DS t] = S^{-1} \exp[-iDt] S$$

that is, a diagonalization (a very expensive operation) was required at every step. It eventually became clear that for large systems this is unrealistic, therefore...
Krylov propagation

It may be seen directly (from the Taylor expansion above) that all m-th degree polynomial approximations of the matrix exponential are elements of mth Krylov subspace:

\[ K_m(tA, v) = \text{Span} \left\{ v, (tA)v, \ldots, (tA)^{m-1}v \right\} \]

which is often a lot smaller than the full state space (remarkably, even for 16k+ matrices this dimension is usually around 30 if the time step is chosen correctly).

Basis screening using conservation laws

Any spin operator that commutes with the Hamiltonian is conserved:

\[ [\hat{H}, \hat{A}] = 0 \quad \Leftrightarrow \quad \hat{H}\hat{A} = 0 \]

Any state violating this can be weeded out. (analytic checks using only index information is normally possible). Furthermore, if:

\[ \hat{A}\hat{\rho}(0) = [\hat{A}, \hat{\rho}(0)] = a\hat{\rho}(0) \]

then the corresponding eigenvalue \( a \) is also a conserved property, because

\[ \hat{A}\hat{\rho}(t) = [\hat{A}, e^{-i\hat{H}t}\hat{\rho}(0)e^{i\hat{H}t}] = e^{-i\hat{H}t}[\hat{A}, \hat{\rho}(0)]e^{i\hat{H}t} = a\hat{\rho}(t) \]

Each operator in the null space of the Hamiltonian superoperator yields an invariant against which the basis set may be screened.

Example: secular Hamiltonians (very popular in NMR, quite frequent in EPR)

\[ \left[ \hat{H}, \hat{L}_Z^{(\text{total})} \right] = 0 \quad \Rightarrow \quad \text{only} \]

\[ \hat{T}_{l_1,m_1} \otimes \hat{T}_{l_2,m_2} \otimes \ldots \otimes \hat{T}_{l_n,m_n} \]

with \( \sum_n m_n = \left( \sum_n m_n \right)_{t=0} \)

Destination state screening

Only the states that evolve into or affect detectable states need in practice be simulated.

This detectability requirement (formally expressed using Lie group orbits) leads to further exact reduction in matrix dimension.

Table 1. Matrix dimension statistics for Liouville space simulations of the high-field pulse-acquire ESR experiment on several common organic radicals in liquid state.

<table>
<thead>
<tr>
<th>Radical and symmetry</th>
<th>Full state space dimension</th>
<th>Basis set</th>
<th>Basis dimension</th>
<th>A_{1g} irrep dimension</th>
<th>Source state screening Dimension after ZTE</th>
<th>Source state screening Dimension after ZTE and PT</th>
<th>Destination state screening Dimension after ZTE</th>
<th>Destination state screening Dimension after ZTE and PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>methyl, S_1</td>
<td>256</td>
<td>complete</td>
<td>256</td>
<td>80</td>
<td>9</td>
<td>1×1, 2×4</td>
<td>4</td>
<td>1×4</td>
</tr>
<tr>
<td>methyl, S_1</td>
<td>256</td>
<td>complete</td>
<td>256</td>
<td>256</td>
<td>19</td>
<td>3×1, 2×8</td>
<td>8</td>
<td>1×8</td>
</tr>
<tr>
<td>phenyl, S_2 ⊗ S_2</td>
<td>4096</td>
<td>ESR-1</td>
<td>128</td>
<td>72</td>
<td>39</td>
<td>3×1, 2×18</td>
<td>18</td>
<td>1×18</td>
</tr>
<tr>
<td>pyrene, S_4 ⊗ S_4 ⊗ S_2</td>
<td>4⋅10^6</td>
<td>ESR-1</td>
<td>4096</td>
<td>300</td>
<td>153</td>
<td>3×1, 2×75</td>
<td>75</td>
<td>1×75</td>
</tr>
</tbody>
</table>

*a ESR-1 basis set: complete on the electrons, T_{10} states only on nuclei, adapted to high-field ESR simulations [8].

Graph-theoretical block-diagonalization

Not everything is connected to everything and not every loop overlaps with every other...

The basis set expansion of the initial condition can be traced through the Liouvillian to determine the independent evolution subspaces.

Many subspaces are not in practice populated and may be dropped from the state space, resulting in further reduction.

*Disjoint subgraph detection (using Tarjan’s implementation of depth-first search) scales linearly with the number of non-zeroes in \( L \).

*Number of disjoint subspaces in typical NMR cases: 4 to 400

Sparse array cleanup procedure

In practice, the non-zero index is inspected and small non-zeroes are dropped from the index. In Matlab syntax:

\[ H = H \cdot (\text{abs}(H) > 1e-6) \]

(a superficially equivalent command \[ H(\text{abs}(H) < 1e-6) = 0 \] would overflow the system memory because the logical matrix in brackets is dense).

Zero track elimination

Matrix dimensions are reduced from astronomical to manageable and NMR simulations for systems with hundreds of spins can be performed on a desktop workstation in a matter of minutes (1D experiments) or hours (2D and 3D).

State space partitioning into single, double, triple, etc. spin correlation orders clearly indicates that high orders are not getting populated.

If we neglect the coherences of order $k < n$...

$$4^k C_n^k = 4^k \frac{n!}{k!(n-k)!} = O\left(\left(\frac{4n}{k}\right)^k\right)$$

...the scaling becomes polynomial in the total number of spins $n$.

Exact applicability range can be obtained for a given Hamiltonian and a given relaxation superoperator:

$$k > 2\sqrt{\frac{\hbar}{\xi}} \text{erfc}^{-1} \left( \text{erfc} \left( \frac{1}{2} \sqrt{\frac{r}{\hbar}} \right) \right)$$
We can partition the state space into subspaces containing spin coherences of a given rank:

$$\mathcal{L} = \mathcal{L}_0 \oplus \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \ldots \oplus \mathcal{L}_N$$

and partition the Liouvillian into single-spin and two-spin terms:

$$\hat{H}_1 \mathcal{L}_k \subset \mathcal{L}_k$$
$$\hat{H}_2 \mathcal{L}_k \subset \mathcal{L}_{k-1} \oplus \mathcal{L}_k \oplus \mathcal{L}_{k+1}$$
$$\hat{R}_1 \mathcal{L}_k \subset \mathcal{L}_k$$
$$\hat{R}_2 \mathcal{L}_k \subset \mathcal{L}_{k-1} \oplus \mathcal{L}_k \oplus \mathcal{L}_{k+1}$$

This gives us equations for the parts of the density matrix belonging to each subspace:

$$\hat{\rho} = \hat{\rho}_0 + \hat{\rho}_1 + \hat{\rho}_2 + \ldots + \hat{\rho}_N$$
$$\hat{\rho}_k \in \mathcal{L}_k$$
$$\hat{\rho}_{\text{eq}} \in \mathcal{L}_1$$

and, in particular, for the corresponding norms (aka probabilities):

$$|\hat{\rho}|^2 = |\hat{\rho}_0|^2 + |\hat{\rho}_1|^2 + |\hat{\rho}_2|^2 + \ldots + |\hat{\rho}_N|^2$$

A. Karabanov et al., Journal of Chemical Physics, submitted (arXiv:1104.3866v1)
Diagonalization-free algorithms

The integral in the master equation of BRW relaxation theory:

\[
\hat{R} = -\sum_{kmpq} \int_0^\infty G_{kmpq}(\tau) \hat{Q}_{km} e^{-i\hat{H}_0 \tau} \hat{Q}_{pq} e^{i\hat{H}_0 \tau} d\tau
\]

is normally evaluated by diagonalizing the Hamiltonian and expanding the Q operators in its eigenbasis. For large spin systems this is not possible -- eigenvectors of sparse matrices are normally dense. However:

\[
M = \begin{pmatrix}
A_{11} & A_{12} & 0 & 0 & 0 \\
0 & A_{22} & A_{23} & 0 & 0 \\
0 & 0 & A_{33} & \ddots & 0 \\
0 & 0 & 0 & \ddots & A_{k-1,k} \\
0 & 0 & 0 & 0 & A_{kk}
\end{pmatrix}, \quad \exp(Mt) = \begin{pmatrix}
B_{11} & B_{12} & B_{13} & \cdots & B_{1k} \\
0 & B_{22} & B_{23} & \cdots & B_{2k} \\
0 & 0 & B_{33} & \ddots & \vdots \\
0 & 0 & 0 & \ddots & B_{k-1,k} \\
0 & 0 & 0 & 0 & B_{kk}
\end{pmatrix}
\]

\[
B_{1k} = \int_0^{t_1} dt_1 \int_0^{t_2} dt_2 \cdots \int_0^{t_{k-2}} dt_{k-1} \left\{ e^{A_{11}(t-t_1)} A_{12} e^{A_{22}(t_1-t_2)} A_{23} \cdots A_{k-1,k} e^{A_{kk}t_{k-1}} \right\}
\]

Some of the scariest integrals in relaxation theory, average Hamiltonian theory, pulsed field gradient simulations, etc. may be taken in one line of Matlab!

D.L. Goodwin, I. Kuprov, arXiv:1506.00628
Basis set construction strategies

**Basis Set**

- **Uniform truncation** -- ignore all spin orders higher than $k$.
- **Adaptive truncation** -- the local spin order cut-off is equal to the number of nearest neighbours on the coupling graph.
- **Exact truncation** -- populated subspaces are identified using algebraic techniques.
- **On-the-fly truncation** -- the basis set is dynamically shaped around a running simulation.

---

**Basis set specification**

The most convenient initial basis set is irreducible spherical tensors (many transformations may be carried out analytically using index information).

\[
\hat{T}_{l_1 m_1} \otimes \hat{T}_{l_2 m_2} \otimes \ldots \otimes \hat{T}_{l_n m_n}
\]

---

**Basis set construction: notes**

1. The full basis set cannot be stored -- it must be generated directly in the reduced form.
2. A representation of spin operators must be constructed directly in the reduced basis.
3. The reduced basis must be closed with respect to commutation and temporal propagation.
4. The scaling of all procedures should ideally be linear with respect to basis set size.

---

Semantic symmetry factorization

Symmetry factorization may be performed using just index information:

A rotation-friendly general (any spin) basis is provided by direct products of irreducible spherical tensors:

\[ S = \prod_{l,m} \hat{T}_{l,m} = \hat{T}_{l_1,m_1} \otimes \hat{T}_{l_2,m_2} \otimes \ldots \otimes \hat{T}_{l_n,m_n} \]

In the IST representation, only the index table needs to be stored and manipulated:

\[ \hat{P} \left( \begin{array}{c} l_1 \\ m_1 \\ l_2 \\ m_2 \\ \vdots \\ l_n \\ m_n \end{array} \right) = \left( \begin{array}{c} l_{P(1)} \\ m_{P(1)} \\ l_{P(2)} \\ m_{P(2)} \\ \vdots \\ l_{P(n)} \\ m_{P(n)} \end{array} \right) \]

Result: basis set symetrization with very low storage and CPU requirements for 100+ spin systems and all common symmetry groups.

Multiple groups of equivalent spins yield more complex symmetries...

The character table for the group direct product is fairly straightforward to generate:

\[ \chi^{a \times b} \left( \hat{A}_k \hat{B}_l \right) = \chi^a \left( \hat{A}_k \right) \chi^b \left( \hat{B}_l \right) \]

Ubiquitin spin system data

Ubiquitin: 76 amino acids, 1060 magnetic nuclei, about 60,000 spin-spin couplings.

Isotropic chemical shifts: BMRB database, ubiquitin is fully assigned.
CSA tensors: literature data rotated into the local reference frame, where available.
Dipolar couplings: extracted automatically from atomic coordinates.
Atomic coordinates: PDB database, rigid structure assumed.
J-couplings: an MD-like “field”, parameterized by the literature data.
Quadrupolar couplings: DFT values where necessary.
Rotational correlation time: global rotational diffusion only at the moment, 5.0 ns.
Relaxation superoperator: computed automatically from the data listed above.

NOESY of ubiquitin

The anisotropic Hamiltonian has between 60,000 and 200,000 terms, depending on the distance cut-off radius. Dipolar relaxation is very expensive.

NOESY calculations get very hard very fast as the distance cut-off and maximum correlation order are increased. Realistic limit for ubiquitin is IK-1(4,2).

The anisotropic Hamiltonian has between 60,000 and 200,000 terms, depending on the distance cut-off radius. Dipolar relaxation is very expensive.

Experimental data from Donghan Lee
1. 3D NMR experiments can be simulated at essentially the same cost as 2D.
2. Phase cycles are not necessary - coherence selection is performed algebraically.
3. Explicit decoupling is not necessary - the Hamiltonian may be edited instead.
4. Quadrature detection is achieved in one pass - L_+ state may be detected directly.
For DD-CSA cross-correlation effect to appear, $^{15}$N CSA tensors must be either supplied or reasonably guessed (their eigenaxes are linked to local geometry).

HSQC of ubiquitin

The differences between the two spectra are due to: (A, B) rapid exchange of HN protons in GLU24 and GLY53 with the deuterium of the solvent – the corresponding signals are lost in the noise in the experimental data; (C) aliasing of arginine Nε-Hε signals from their position at ~90 ppm 15N chemical shift that is outside the spectral window – simulated spectra are free of this artefact; (D) slow exchange of Hε protons in GLN41 with the deuterium of the solvent – the corresponding pair of signals is attenuated in the experiment.

Experimental data from Donghan Lee
HSQC of ubiquitin

Theoretical HSQC, RSS IK-1(4,1)

Experimental HSQC

Experimental data from Joern Werner
HNCO of ubiquitin

The few missing signals (or extra signals) can all be tracked to deuterated residues and folded peaks. In all other respects the agreement is quantitative.

Experimental data from Donghan Lee
Summary of Spinach capabilities

- Automatic long-lived state detection
- Tensor train formalism
- Spin system trajectory analysis
- Coupling tensor visualization
- Spin Chemistry experiments
- Open-source, extensively commented code
- General rotations module
- Isotropic, axial and rhombic rotational diffusion correlation functions
- Hilbert space, Liouville space, Fokker-Planck space
- Accurate thermal equilibria and thermalization
- Polynomial complexity scaling in Liouville space

Spinach capabilities

- Exotica
- NMR
- General
- Hyperpolarization
- EPR
- Optimal Control

- 2D: COSY, NOESY, HSQC, etc.
- 3D: HNCO, HNCOCA, HNCA, etc.
- Parallel powder averages
- Partially aligned systems
- Redfield relaxation theory
- Floquet and Fokker-Planck MAS
- Stochastic Liouville equation
- Simultaneous shaped pulses
- Chemical kinetics superoperator
- Overhauser DNP
- Cross effect DNP
- Solid effect DNP
- Field/frequency sweep DNP
- Parahydrogenation
- CIDNP and CIDEH
- Krotov algorithm
- Zhu-Rabitz algorithm
- BFGS-GRAPe algorithm
- Penalty functional library
- Waveform basis translation

http://spindynamics.org
Summary of algorithms in Spinach

Most simulation stages (Hamiltonian generation, powder averages, propagation, time increments for indirect dimensions, etc.) are parallelized and tested up to 128 cores.

http://www.spindynamics.org
Hardware, software and time requirements

<table>
<thead>
<tr>
<th>Simulation (ubiquitin)</th>
<th>Time / hours (24 cores)</th>
<th>Memory / GB (incl. Matlab)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMQC, 256x256, IK-1(2,1)</td>
<td>0.5</td>
<td>40</td>
</tr>
<tr>
<td>HMQC, 256x256, IK-1(2,2) + DD/CSA relaxation theory</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>HMQC, 256x256, IK-1(3,1)</td>
<td>5.5</td>
<td>60</td>
</tr>
<tr>
<td>HMQC, 256x256, IK-1(3,2) + DD/CSA relaxation theory</td>
<td>22</td>
<td>60</td>
</tr>
<tr>
<td>HMQC, 256x256, IK-1(4,1)</td>
<td>11</td>
<td>90</td>
</tr>
<tr>
<td>HMQC, 256x256, IK-1(4,2) + DD/CSA relaxation theory</td>
<td>96</td>
<td>90</td>
</tr>
</tbody>
</table>

If Moore’s law keeps going, your laptop would run this in about five years’ time.

<table>
<thead>
<tr>
<th>Number of spins</th>
<th>State space dimension</th>
<th>Reduced Liouvillian density, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete</td>
<td>All 4-spin</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>2.8×10</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>3.2×10^{32}</td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>2.6×10^{61}</td>
</tr>
<tr>
<td></td>
<td>198</td>
<td>1.6×10^{119}</td>
</tr>
</tbody>
</table>

Dell PowerEdge R820, 32 Intel Sandy Bridge cores, 1024 GB of RAM, 6 TB of storage, approx. £50,000.

Matlab Distributed Computing Server, 32 worker licenses, approx. £6,000.
The choice of language and platform

Linux / C++
- Coding speed: very slow
- Runtime speed: very fast
- TTI (Spinach): unrealistic
- Entry barrier: unrealistic
- Maintenance: difficult
- License costs: modest to none

Windows / Matlab
- Coding speed: very fast
- Runtime speed: fast
- TTI (Spinach): six months
- Entry barrier: none
- Maintenance: easy
- License costs: £1,000+

FEC of brain time: between £30 and £100 per hour, depending on qualification.
FEC of CPU time: £0.06 per core-hour, improves exponentially with time.
Collaborators and funding

**University of Southampton**
- Joern Werner
- Giuseppe Pileio
- Maria Concistre

**University of Oxford**
- Peter Hore
- Hannah Hogben

**Nanyang University**
- Konstantin Pervushin

**University of Southampton**
- Luke Edwards
- Zenawi Welderufael
- Gareth Charnock
- Dmitry Savostyanov
- Andreas Biternas
- David Goodwin

**University of Nottingham**
- Walter Kockenberger
- Alexander Karabanov

**MPIBPC**
- Christian Griesinger
- Donghan Lee

---

[Logos for Oxford Supercomputing Centre, University of Southampton, EPSRC, and SEVENTH FRAMEWORK PROGRAMME]