

Exploiting sum of products structure to compute vibrational spectra: molecules with up to 18 atoms

Phillip Thomas and Tucker Carrington

Chemistry Department,
Queen's University

14 juin 2018

- Develop new methods for solving the time-independent Schroedinger equation to compute vibrational spectra, i.e., many energy levels, of molecules with 5, 6, 7, \dots , 18, \dots atoms

- Easy to use (e.g. second-order PT is in Gaussian)
- Nearly degenerate levels cause perturbation theory to break down.
- The density of states increases with energy.
- Methane vibrational levels in the Octad ($\sim 4000 \text{ cm}^{-1}$ above the ZPE) computed with fourth and sixth order perturbation theory differ by about 8 cm^{-1} .

Standard recipe for solving the Schroedinger equation

- represent wavefunctions with basis functions

$$\psi_n(\mathbf{q}) = \sum_k c_{kn} f_k(\mathbf{q})$$

- compute eigenvalues and eigenvectors of the Hamiltonian matrix

$$\mathbf{H}\mathbf{U} = \mathbf{U}\mathbf{E}$$

Basis sets are huge

Often one uses a direct product basis. Each basis function is a product :

$$f_{k_1, k_2, \dots} = \theta_{k_1}(q_1) \theta_{k_2}(q_2) \cdots \theta_{k_D}(q_D)$$

If n basis functions are required for each coordinate and there are D coordinates then the size of the basis is n^D .

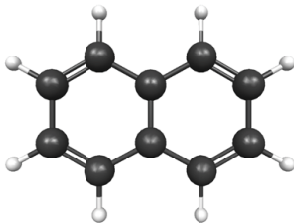
About 10 1-D functions required for each coordinate.

To compute vibrational levels $> 10^{3N-6}$ multi-d basis functions are required.

- Iterative eigensolvers are advantageous

One would like to compute spectra for larger molecules,
e.g. Naphthalene

$$N_{atoms} = 18, D = 48, \text{ if } n = 10$$



A single direct-product vector has $\sim 10^{48}$ elements
and requires $\sim 8 \times 10^{39}$ GB

This talk : A method that does not require storing (even) vectors with as many components as basis functions

- It can only be used if the Hamiltonian is a sum of products (SOP).
- The kinetic energy operator (KEO) is often a SOP.
- The terms of a SOP potential need not be products of powers of coordinates.

Understanding the origin of the memory barrier

- The vectors represent basis functions (in a primitive direct product basis).
- A *general* basis function is

$$\psi_{gen}(q_1, \dots, q_D) = \sum_{i_1=1}^{n_1} \dots \sum_{i_D=1}^{n_D} F_{i_1 i_2 \dots i_D} \prod_{j=1}^D \theta_{i_j}^j(q_j) .$$

- The corresponding vector is $F_{i_1 i_2 \dots i_D}$. To store it one must store n^D numbers.

The memory required to represent a *product* function scales linearly, and not exponentially, with D

$\phi_1(q_1)\phi_2(q_2)$ can be represented as

$$\left(\sum_{i_1=1}^n f_{i_1}^{(1)} \theta_{i_1}^1(q_1) \right) \left(\sum_{i_2=1}^n f_{i_2}^{(2)} \theta_{i_2}^2(q_2) \right) .$$

In this case, $F_{i_1 i_2}$ has the special form,

$$F_{i_1 i_2} = f_{i_1}^{(1)} f_{i_2}^{(2)}$$

All n^2 $F_{i_1 i_2}$ can be computed from the n $f_{i_1}^{(1)}$ values and the n $f_{i_2}^{(2)}$ values.

This works only because the (basis) function being represented is a product of univariate functions.

In general, the memory required scales as Dn

The same ideas apply in more than two dimensions

$$F_{i_1 i_2 \dots i_D} = \prod_{j=1}^D f_{i_j}^{(j)} ;$$

the memory cost is Dn .

Okay, but ...

It is possible to reduce the cost of storing vectors,
but that only works if the basis functions are products of univariate
functions and products are poor basis functions.

- Sums of products (SOP) are good basis functions and **also** require little memory.
- A SOP vector will be written

$$F_{i_1 i_2 \dots i_D} = \sum_{\ell=1}^{R_\psi} \prod_{j=1}^D f_{i_j}^{(\ell, j)},$$

- To represent a sum of R_ψ products, the memory cost is $R_\psi Dn$.

Making the SOP basis

- To make basis vectors, we use a (shifted) block power method, and apply $\tilde{\mathbf{H}} = \mathbf{H} - \sigma \mathbf{I}$ to a set of B \mathbf{F} vectors.
- If the Hamiltonian is a SOP, and \mathbf{F} is in CP format, then the vector $\mathbf{G} = \mathbf{H} \mathbf{F}$ will also be CP. To store it, one needs to store $TR_\psi nD$ numbers; T is the number of terms in the Hamiltonian; R_ψ is the number of terms in \mathbf{F} .
- We have called this the RRBPM, rank reduced block power method

After applying **H** to **F**, the rank must be reduced

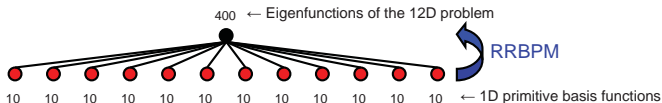
- After N_{pow} matrix-vector products the rank is $T^{N_{pow}} R_{\psi}$. The rank increases after each matrix-vector product and must be reduced to R_{ψ}
- Individual basis vectors may be compromised by reduction error

These ideas shatter the memory barrier

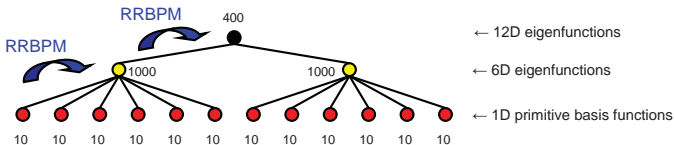
The lowest 70 eigenstates of CH_3CN can be computed using less than 1 GB of memory (cf. 8000 GB)
(Leclerc and Carrington, JCP (2014)).

It is even better to combine the RRBPM with successive layers of contractions

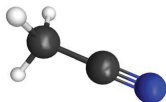
The pure RRBPM is



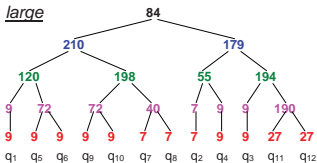
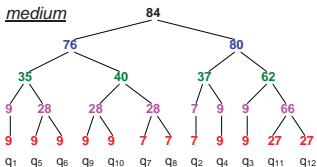
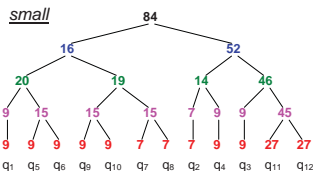
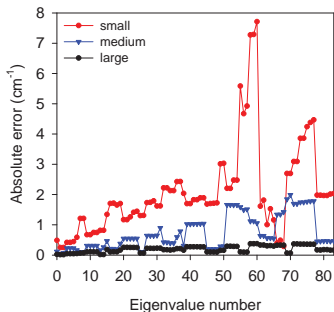
Instead, use it in successive layers :



Acetonitrile (12D), contraction reduces cost factor of 10^5



Calculation	CPU time	Memory
small	44 s	6.7 MB
medium	11 min	29 MB
large	3.2 h	139 MB



We use

$$\begin{aligned}\hat{H} = & \frac{\omega_c}{2} \left(\sum_{c=1}^D -\frac{\partial^2}{\partial q_c^2} + q_c^2 \right) \\ & + \frac{1}{6} \sum_{c_1=1}^D \sum_{c_2=1}^D \sum_{c_3=1}^D \phi_{c_1 c_2 c_3}^{(3)} q_{c_1} q_{c_2} q_{c_3} \\ & + \frac{1}{24} \sum_{c_1=1}^D \sum_{c_2=1}^D \sum_{c_3=1}^D \sum_{c_4=1}^D \phi_{c_1 c_2 c_3 c_4}^{(4)} q_{c_1} q_{c_2} q_{c_3} q_{c_4} ,\end{aligned}$$

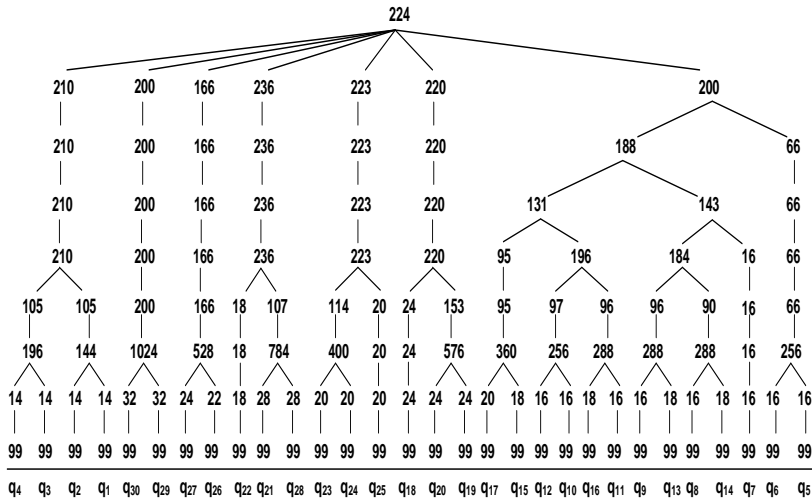
- Lump terms in a node to reduce the rank of the operators
- Univariate basis functions are products of eigenfunctions of 1D cut Hamiltonians

Uracil ($\text{C}_4\text{H}_4\text{O}_2\text{N}_2$)

Force constants are from S. V. Krasnoshchekov, N. Vogt,
and N. F. Stepanov,

Potential modified to eliminate holes which cause spurious
eigenvalues.

Tree, group by type of motion



Low-lying levels

R= 400	R= 500	R= 600	
B1	B2	B3	Assignment
18993.14	18993.07	18993.02	ZPE
143.90	143.87	143.84	ν_{30}
159.20	159.17	159.15	ν_{29}
296.53	296.47	296.41	$2\nu_{30}$
303.56	303.48	303.42	$\nu_{30} + \nu_{29}$
324.39	324.30	324.24	$2\nu_{29}$
385.18	385.14	385.12	ν_{21}
386.37	386.31	386.27	ν_{28}
511.49	511.45	511.43	ν_{20}
528.56	528.32	528.14	$\nu_{30} + \nu_{28}$
529.83	529.62	529.49	$\nu_{30} + \nu_{21}$
533.04	532.83	532.68	ν_{19}
536.79	536.67	536.56	ν_{18}
545.08	544.95	544.87	$\nu_{29} + \nu_{21}$
551.42	551.24	551.10	$\nu_{29} + \nu_{28}$
655.56	655.45	655.37	$\nu_{30} + \nu_{20}$
671.25	671.24	671.12	$\nu_{30} + \nu_{19}$

Higher-lying levels

R= 400	R=500	R=600	Assignment
B1	B2	B3	
921.93	919.71	919.29	$\nu_{21} + \nu_{19}$
925.86	924.79	923.61	$\nu_{21} + \nu_{18}$
929.64	927.56	927.15	$\nu_{29} + \nu_{17}$
950.77	949.54	949.15	$\nu_{30} + \nu_{26}$
952.43	951.72	950.77	$\nu_{30} + \nu_{23}$
962.00	960.94	960.34	ν_{16}
965.80	965.03	964.69	$\nu_{29} + \nu_{26}$
975.61	972.40	971.14	ν_{22}
997.61	996.30	995.56	ν_{15}
1030.83	1027.40	1026.27	$2\nu_{20}$
1050.63	1048.80	1048.94	$\nu_{20} + \nu_{19}$
1136.77	1134.35	1128.43	$\nu_{27} + \nu_{21}$
1149.46	1146.22	1144.15	$\nu_{21} + \nu_{24}$
1149.34	1146.84	1146.03	$\nu_{28} + \nu_{24}$
1163.85	1159.03	1156.53	$\nu_{28} + \nu_{17}$
1205.22	1198.78	1196.00	$\nu_{26} + \nu_{21}$
1208.62	1207.24	1205.80	ν_{13}

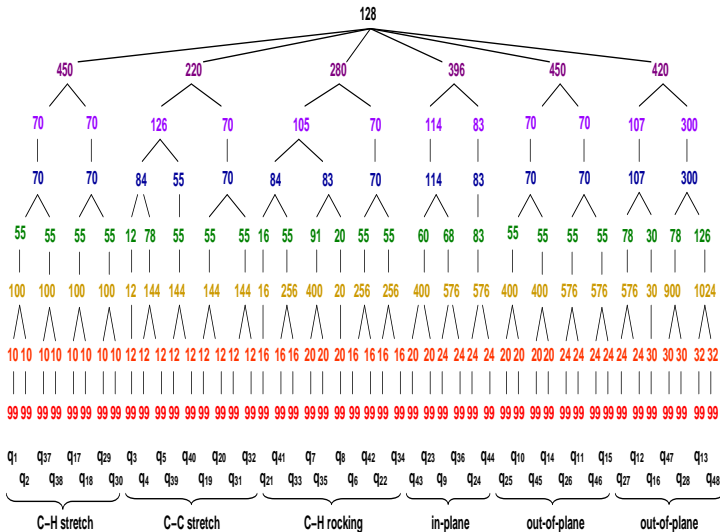
The largest calculation takes 16 days and requires 181 GB, which can be reduced to 4 GB.

Naphthalene (C_8H_{10})

Force constants are from E. Cané, A. Miani, and A. Trombetti.

No holes !

Tree



Low-lying levels

R= 600	R= 700	R= 1000	
E	F	G	
31766.87	31766.50	31766.03	ZPE
164.92	164.79	164.60	ν_{48}
178.51	178.36	178.18	ν_{13}
330.17	329.82	329.41	$2\nu_{48}$
342.76	342.47	342.02	$\nu_{13} + \nu_{48}$
355.05	354.85	354.44	ν_{24}
358.52	358.20	357.66	$2\nu_{13}$
387.95	387.84	387.71	ν_{16}
463.87	463.71	463.47	ν_{28}
472.83	472.63	472.41	ν_{47}
496.71	496.14	495.50	$3\nu_{48}$
505.81	505.87	505.64	ν_{44}
510.18	509.61	508.43	$\nu_{13} + 2\nu_{48}$
513.31	512.95	512.32	ν_9
521.72	521.14	520.31	$\nu_{24} + \nu_{48}$

Higher-lying levels

R = 600	R = 700	R = 1000	
795.67	793.66	793.54	ν_{23}
800.63	799.73	798.51	$\nu_{28} + 2\nu_{48}$
804.53	803.25	802.17	$\nu_{36} + \nu_{13}$
807.88	806.78	805.87	$\nu_{12} + \nu_{13}$
813.30	813.88	812.49	$\nu_{47} + 2\nu_{48}$
815.74	814.40	812.63	$\nu_{28} + \nu_{13} + \nu_{48}$
824.32	823.63	821.00	$\nu_{24} + \nu_{28}$
827.72	826.48	824.15	$\nu_{47} + \nu_{13} + \nu_{48}$
829.04	829.78	827.44	$\nu_{28} + 2\nu_{13}$
833.04	832.12	829.58	$\nu_{24} + \nu_{47}$
834.79	832.66	830.65	ν_{11}
840.61	839.33	838.23	$\nu_{47} + 2\nu_{13}$
846.24	846.72	841.94	$\nu_{44} + 2\nu_{48}$
853.59	851.53	848.19	$\nu_9 + 2\nu_{48}$
856.43	854.56	850.51	$5\nu_{48}$
860.61	857.81	854.09	$\nu_{44} + \nu_{13} + \nu_{48}$
861.97	860.26	858.45	$\nu_{13} + 4\nu_{48}$
865.45	864.87	858.93	$\nu_{24} + 3\nu_{48}$
868.27	866.84	860.27	$\nu_{16} + \nu_{28}$
870.46	868.47	865.36	$\nu_9 + \nu_{13} + \nu_{48}$
872.57	872.98	866.89	$\nu_{24} + \nu_{44}$
875.10	875.72	870.84	$2\nu_{13} + 3\nu_{48}$

The largest calculation takes 35 days and requires 371 GB, which can be reduced to 4 GB.

Storing the direct-product matrix for Naphthalene would require 10^{87} GB. Storing one vector would require 10^{39} GB.

- Using established methods, variational calculations of spectra of molecules with 6,7, \dots atoms are limited by the memory required
- The memory cost can be effectively eliminated by using SOP basis functions and rank reduction
 - The memory scales linearly with D
 - vectors with as many components as there are direct product basis functions are neither calculated nor stored.
 - the memory required can be reduced to the point that the vibrational levels of a molecule with 18 atoms can be done on a desktop computer or with GPUs.
 - First variational calculations on a molecule with 18 atoms

This work has been supported by :

the Natural Sciences and Engineering Research Council,
the Canadian Foundation for Innovation, and
the Canada Research Chairs programme.