Predicting Vibrational Spectra of Condensed Phase Systems



Martin Brehm Martin-Luther-Universität Halle–Wittenberg

https://brehm-research.de/

Outline

- **1.) Introduction**
- 2.) Wannier Localization and Voronoi Integration
- 3.) Computing VCD and ROA Spectra
- 4.) Compressing Volumetric Data
- 5.) Resonance Raman Spectroscopy

Vibrational Spectroscopy

"Standard"	"Chiral"
Infrared (IR)	Vibrational Circular Dichroism (VCD)
Raman	Raman Optical Activity (ROA)

Important to determine absolute configurations



Vibrational Spectroscopy



Ab initio prediction required!

Computing Vibrational Spectra

Static–Harmonic Approach	Molecular Dynamics Approach
Single molecule/cluster in vacuum	Possible for bulk phase systems
Optimized structure (one conformer)	Intrinsic conformer sampling
Harmonic approximation	Includes some anharmonic effects*
No explicit solvent influence	Full solvent influence (H-Bonds,)
Discrete line spectrum	Realistic band shapes
Used since ≈ 50 years	Use since ≈ year 2000

* e.g. approximative overtones, combination bands

Predicting Bulk Phase Spectra



Molecular Dynamics Approach

Example: Methanol

- Good agreement both in vacuum and bulk phase
- Solvent effects fully captured (line broadening)



M. Thomas, M. Brehm, R. Fligg, P. Vöhringer, B. Kirchner, Phys. Chem. Chem. Phys. 2013, 15, 6608–6622.

Molecular Dynamics Approach

General Idea: Compute molecular electromagnetic properties *(e.g. dipole moment)* along simulation trajectory

Mostly done with *ab initio* MD (AIMD):

- Numerical solution of the time-independent Schrödinger equation
- No empirical parameters \rightarrow *first principles* method



All spectra in this talk: No scaling / shifting / etc., directly as computed

→ Truly predictive method (also works reliably for "new" substances)

2. Wannier Localization and Voronoi Integration





- Molecular orbitals are not unique; every unitary transformation yields another set of valid MOs
- Idea: Find set of MOs with minimal (spatial) spread
- Those are called Wannier orbitals
- The center of mass of each Wannier orbital is called Wannier center
- Suggested by G. Wannier for solid-state systems in 1937



- Can be used to compute molecular dipole moments
- For only 1 molecule in vacuum, dipole from Wannier centers exactly matches "true" QM dipole (Berry phase)
- Standard algorithm uses Jacobi diagonalization to construct suitable unitary transformation; rather slow...
- CP2k offers a very modern and efficient method, called "Crazy Angle algorithm" ^(C)
- Virtually all IR / Raman spectra from AIMD rely on Wannier centers for molecular dipole moments



Wannier Centers: Example

Bulk phase [EMIm][OAc], liquid, 936 atoms, cell 21 x 21 x 21 Å, 350 K, BLYP–D3, DZVP–MOLOPT–SR, 2500 electrons, 8100 basis functions, 1 node (16 cores) Intel Xeon E5-2620 v4





Wannier Centers: Example





Wannier Centers: Example



LOCALIZA	FION	Computing	localizati	ion properties	for OCCUPIE	D ORBITALS.	Spin: 1
Spread	d Fund	ctional	sum_in -v	w_i ln(z_in ^:	2) sum_in	w_i(1- z_in	^2)
Initia	al Spi	read (Berry	7) :	992039.6	748362964	149090.84766	25035
CRAZY	Iter		value	gradient	Max. eval	limit	
CRAZY	1	759416.300	978268	0.1314E+05	0.2295E+01	0.2000E+00	
CRAZY	2	1825582.2	5180607	0.7230E+04	0.5686E+01	0.2000E+00	
CRAZY	3	3504667.50	291648	0.4232E+04	0.9659E+01	0.2000E+00	
CRAZY	4	5213901.5	7181955	0.3002E+03	0.1881E+02	0.2000E+00	
CRAZY	497	5872400.29	9576750	0.2857E+03	0.2367E+03	0.2000E+00	
CRAZY	498	5872334.68	3198040	0.3178E+03	0.2374E+03	0.2000E+00	
CRAZY	499	5872451.22	2848857	0.3350E+03	0.2378E+03	0.2000E+00	
CRAZY	500	5871893.20	651500	0.2990E+03	0.2360E+03	0.2000E+00	
Crazy	Wanni	er locali:	zation not	converged aft	er 500		
itera	tions	switching	g to jacobi	i rotations			
Local	izatio	on by itera	ative dist	ri			
		Ite	eration		Ainud		Time
			100		VIIIU		3.361
			200			3	3.361
			300	-74895.0539	325070	0.4188E+01	3.362
			1200	-74895.0559	064523	0.7365E-03	3.355
			1300	-74895.0559	064549	0.3717E-03	3.353
			1400	-74895.0559	064543	0.1876E-03	3.355
Localization for spin 1 converged in 1493 iterations							
Spread Functional sum_in -w_i ln(z_in ^2) sum_in w_i(1- z_in ^2)							
Total	Sprea	ad (Berry)	:	-22638.0	247818508	-74895.05590	64557

Average frame times:

Standard AIMD: 47.9 seconds AIMD + Wannier: 139.3 seconds

 \rightarrow Wannier localization takes 91.4 seconds on average!

65% of total computer time goes into localization only...

Even if CRAZY <u>would</u> converge in every step, localization would take ≈ 30 seconds per frame.

It can even happen that JACOBI does not converge; then no Wannier centers at all are available for that frame...

Benzene IR spectrum: Artificial peak



M. Thomas, M. Brehm, B. Kirchner, Phys. Chem. Chem. Phys. 2015, 17, 3207–3213.



M. Thomas, M. Brehm, B. Kirchner, Phys. Chem. Chem. Phys. 2015, 17, 3207–3213.

Benzene Raman: No spectrum at all, only noise...



M. Thomas, M. Brehm, B. Kirchner, Phys. Chem. Chem. Phys. 2015, 17, 3207–3213.

Most IR / Raman spectra from AIMD use Wannier centers.

Limitations:

- Huge computational overhead (for systems with ≈ 1000 atoms: around 65% of the total CPU time!)
- Not guaranteed to converge at all
- Severe problems in aromatic systems (artificial bands; polarizability not accessible)
- Only works for electric dipole; can't reproduce quadrupole *(required for ROA)*

Our idea: Completely drop Wannier localization; Integrate molecular dipole via Voronoi instead.

Our idea: Completely drop Wannier localization; Integrate molecular dipole via Voronoi instead.

Voronoi Tessellation (G. Voronoi, 1908):





Our idea: Completely drop Wannier localization; Integrate molecular dipole via Voronoi instead.

$$\mathbf{p}^{\text{Mol}} = \sum_{n=1}^{N_{\text{Mol}}} q_n \mathbf{r}_n - \int_{\text{Mol}} \rho\left(\mathbf{r}\right) \mathbf{r} \, d^3 \mathbf{r}$$
$$\mathbf{Q}_{ij}^{\text{Mol}} = \sum_{n=1}^{N_{\text{Mol}}} q_n \left(3\mathbf{r}_{n,i}\mathbf{r}_{n,j} - \|\mathbf{r}_n\|^2 \, \delta_{ij}\right) - \int_{\text{Mol}} \rho\left(\mathbf{r}\right) \left(3\mathbf{r}_i \mathbf{r}_j - \|\mathbf{r}\|^2 \, \delta_{ij}\right) d^3 \mathbf{r}$$



M. Thomas, M. Brehm, B. Kirchner, Phys. Chem. Chem. Phys. 2015, 17, 3207–3213.

Average Timings

Wannier-based:

47.9 s AIMD 91.4 s Localization 139.3 s per frame

Voronoi Integration:

47.9 s AIMD

- 10.0 s Write CUBE (~100 MiB)
- 2.0 s Voronoi Integration

59.9s per frame

Saves more than a factor of 2 in total CPU time!

Our idea: Completely drop Wannier localization; Integrate molecular dipole via Voronoi instead

- Spectra look almost identical
- Saves lots of computational time (Factor > 2)
- All problems due to Wannier localization are gone
- Works also for higher multipole moments



Literature:

M. Thomas, M. Brehm, B. Kirchner, Phys. Chem. Chem. Phys. 2015, 17, 3207–3213.

M. Brehm, M. Thomas, *Molecules* **2021**, *26* (7), 1875.

Since 2021, the Voronoi integration is available directly inside of CP2k *(from v8.1)* via *"libvori"*!

&DFT &PRINT &VORONOI ... &END VORONOI &END PRINT &END DFT



 \rightarrow see CP2k manual

Literature:

M. Thomas, M. Brehm, B. Kirchner, Phys. Chem. Chem. Phys. 2015, 17, 3207–3213.

M. Brehm, M. Thomas, *Molecules* **2021**, *26 (7)*, 1875.

Predicting Bulk Phase Spectra

"Standard"	"Chiral"
Infrared (IR)	Vibrational Circular 7 Dichroism (VCD)
Raman 🗸	Raman Optical 7 Activity (ROA)

3. Computing VCD and ROA Spectra



Why is VCD / ROA from MD so Hard?

Up to 2016: No VCD / ROA spectra from MD in literature... **Why?**

VCD and ROA requires magnetic dipole moment m

Magnetic moment results from electric current...

But: Simulations (mostly) run in Born–Oppenheimer approximation

 \rightarrow No electronic currents occur

"One cannot do it with BOMD."

Solutions?

- a) Use Perturbation Theory (*e.g.* NVPT)
- b) Use Classical Approach

Idea: When considering total electron density of 2 subsequent BOMD steps, how could the current have flown to cause this change?

 ∂t

Start with the continuity equation:

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{j}(\mathbf{r}, t) = 0$$

 $\mathbf{j}(\mathbf{r}, t) = -\rho(\mathbf{r}, t) \nabla \alpha(\mathbf{r}, t)$

Approximate current as a simple product:

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = \nabla \rho(\mathbf{r}, t) \cdot \nabla \alpha(\mathbf{r}, t) + \rho(\mathbf{r}, t) \Delta \alpha(\mathbf{r}, t)$$

Schematic Illustration (2D)



Input 1:

 $\rho(\mathbf{r}, t)$

"Total Density"

Input 2: $\partial
ho({f r},\,t)$

 ∂t

"Change"

Result:

 $\mathbf{j}(\mathbf{r}, t)$

"Current"

Idea: When considering total electron density of 2 subsequent BOMD steps, how could the current have flown to cause this change?

Resulting PDE is discretized and solved on a 3D grid.

→ Huge linear system (≈ 10⁷ equations & unknowns; 10⁷ × 10⁷ matrix), but sparsity can be exploited.

Use iterative method ("BiCGStab") to solve the linear system.

Magnetic moments:
$$\mathbf{m}^{\text{Mol}} = \frac{1}{2} \sum_{n=1}^{N_{\text{Mol}}} q_n \left(\mathbf{r}_n \times \mathbf{v}_n \right) - \frac{1}{2} \int_{\text{Mol}} \mathbf{r} \times \mathbf{j} \left(\mathbf{r} \right) \, d^3 \mathbf{r}$$

Limitations	Advantages
Purely classical (no QM)	Very fast (≈10s per step)
No eddy currents can flow	Only requires total electron density (no modifications to QM code)
Not derived from exact theory	Not limited to DFT (can be applied to any electron structure method)
(πο justification why it should work)	Works well for > 1000 atoms

Additional time required is much smaller than that of a standard AIMD!

Predicted VCD spectrum of liquid (R)-propylene oxide



→ Classical approach indeed works!

The first *ab initio* prediction of a bulk phase VCD spectrum in literature.

M. Thomas et al., J. Phys. Chem. Lett. 2016, 7, 509–513.
Computation of ROA spectra requires three properties:

- Electric dipole electric dipole polarizability
- Electric quadrupole electric dipole polarizability
- Magnetic dipole electric dipole polarizability

The spectrum can be obtained as FT of cross-correlations of those properties along the trajectory...

Sounds ", easy", but required ≈ 1 year of additional effort.

Polarizabilities from finite differences (external electric field):

 $\alpha^{\text{Mol}} = \frac{d}{d\mathbf{E}} \mathbf{p}^{\text{Mol}}$ $\mathcal{A}^{\text{Mol}} = \frac{d}{d\mathbf{E}} \mathbf{Q}^{\text{Mol}}$ $\mathcal{E} \text{lectric dipole} - \mathcal{E} \text{lectric quadrupole Polarizability}$ $\mathcal{G}'^{\text{Mol}} = \frac{d}{d\mathbf{E}} \mathbf{m}^{\text{Mol}}$ $\mathcal{E} \text{lectric dipole} - \mathcal{M} \text{agnetic dipole Polarizability}$

Obtain the required polarizabilities:

 $A^{\mathrm{Mol}} = \mathcal{A}^{\mathrm{Mol}}$ Electric quadrupole – Electric dipole Polarizability $G'^{\mathrm{Mol}} = -(\mathcal{G}'^{\mathrm{Mol}})^T$ Magnetic dipole – Electric dipole Polarizability

D. A. Long: *"The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules"*, John Wiley & Sons Ltd., **2002.**



















Molecular electric dipole / quadrupole from classical expressions:

$$\mathbf{p}^{\text{Mol}} = \sum_{n=1}^{N_{\text{Mol}}} q_n \mathbf{r}_n - \int_{\text{Mol}} \rho(\mathbf{r}) \mathbf{r} \, d^3 \mathbf{r}$$
$$\mathbf{Q}_{ij}^{\text{Mol}} = \sum_{n=1}^{N_{\text{Mol}}} q_n \left(3\mathbf{r}_{n,i} \mathbf{r}_{n,j} - \|\mathbf{r}_n\|^2 \, \delta_{ij} \right) - \int_{\text{Mol}} \rho(\mathbf{r}) \left(3\mathbf{r}_i \mathbf{r}_j - \|\mathbf{r}\|^2 \, \delta_{ij} \right) d^3 \mathbf{r}$$

Electric current from purely classical PDE on grid (see above).

Molecular magnetic dipole from canonical expression:

$$\mathbf{m}^{\mathrm{Mol}} = \frac{1}{2} \sum_{n=1}^{N_{\mathrm{Mol}}} q_n \left(\mathbf{r}_n \times \mathbf{v}_n \right) - \frac{1}{2} \int_{\mathrm{Mol}} \mathbf{r} \times \mathbf{j} \left(\mathbf{r} \right) \, d^3 \mathbf{r}$$

We always use the molecular center of mass as coordinate origin.

Compute the ROA invariants via FT of cross-correlations:

$$\begin{split} aG'(\tilde{\nu}) &= 2\pi c\tilde{\nu}_{\rm in} \int_{-\infty}^{\infty} \left\langle \frac{\dot{\alpha}_{xx}(\tau) + \dot{\alpha}_{yy}(\tau) + \dot{\alpha}_{zz}(\tau)}{3} \frac{G'_{xx}(\tau+t) + G'_{yy}(\tau+t) + G'_{zz}(\tau+t)}{3} \right\rangle_{\tau} \cdot \exp(-2\pi i c\tilde{\nu} t) \, \mathrm{d} t \\ \gamma_{G'}^2(\tilde{\nu}) &= 2\pi c\tilde{\nu}_{\rm in} \int_{-\infty}^{\infty} \left[\frac{1}{2} \left\langle (\dot{\alpha}_{xx}(\tau) - \dot{\alpha}_{yy}(\tau)) \left(G'_{xx}(\tau+t) - G'_{yy}(\tau+t) \right) \right\rangle_{\tau} \right. \\ &+ \frac{1}{2} \left\langle (\dot{\alpha}_{yy}(\tau) - \dot{\alpha}_{zz}(\tau)) \left(G'_{yy}(\tau+t) - G'_{zz}(\tau+t) \right) \right\rangle_{\tau} \\ &+ \frac{1}{2} \left\langle (\dot{\alpha}_{zz}(\tau) - \dot{\alpha}_{xx}(\tau)) \left(G'_{zz}(\tau+t) - G'_{xx}(\tau+t) \right) \right\rangle_{\tau} \\ &+ \frac{3}{2} \left\langle \dot{\alpha}_{xy}(\tau) \left(G'_{xy}(\tau+t) + G'_{yx}(\tau+t) \right) \right\rangle_{\tau} + \frac{3}{2} \left\langle \dot{\alpha}_{yz}(\tau) \left(G'_{yz}(\tau+t) + G'_{zy}(\tau+t) \right) \right\rangle_{\tau} \end{split}$$

$$+\frac{3}{2}\left\langle \dot{\alpha}_{zx}(\tau) \left(G'_{zx}(\tau+t) + G'_{xz}(\tau+t) \right) \right\rangle_{\tau} \right] \exp(-2\pi i c \tilde{\nu} t) dt$$
$$\gamma_A^2(\tilde{\nu}) = \pi c \tilde{\nu}_{\rm in} \int_{-\infty}^{\infty} \left[\left\langle \left(\dot{\alpha}_{yy}(\tau) - \dot{\alpha}_{xx}(\tau) \right) \dot{A}_{z,xy}(\tau+t) \right\rangle_{\tau} \right]$$

$$+ \left\langle \left(\dot{\alpha}_{xx}(\tau) - \dot{\alpha}_{zz}(\tau) \right) \dot{A}_{y,zx}(\tau+t) \right\rangle_{\tau} + \left\langle \left(\dot{\alpha}_{zz}(\tau) - \dot{\alpha}_{yy}(\tau) \right) \dot{A}_{x,yz}(\tau+t) \right\rangle_{\tau} \right. \\ \left. + \left\langle \dot{\alpha}_{xy}(\tau) \left(\dot{A}_{y,yz}(\tau+t) - \dot{A}_{z,yy}(\tau+t) + \dot{A}_{z,xx}(\tau+t) - \dot{A}_{x,xz}(\tau+t) \right) \right\rangle_{\tau} \right. \\ \left. + \left\langle \dot{\alpha}_{yz}(\tau) \left(\dot{A}_{z,zx}(\tau+t) - \dot{A}_{x,zz}(\tau+t) + \dot{A}_{x,yy}(\tau+t) - \dot{A}_{y,yx}(\tau+t) \right) \right\rangle_{\tau} \right. \\ \left. + \left\langle \dot{\alpha}_{zx}(\tau) \left(\dot{A}_{y,zz}(\tau+t) - \dot{A}_{z,zy}(\tau+t) + \dot{A}_{x,xy}(\tau+t) - \dot{A}_{y,xx}(\tau+t) \right) \right\rangle_{\tau} \right]$$

 $\cdot \exp(-2\pi i c \tilde{\nu} t) dt$

Assemble the ROA spectrum as linear combination of invariants:

$$\Delta I(\tilde{\nu}) = \frac{h}{8\varepsilon_0^2 c k_{\rm B} T} \cdot \frac{(\tilde{\nu}_{\rm in} - \tilde{\nu})^4}{\tilde{\nu} \left(1 - \exp\left(-\frac{hc\tilde{\nu}}{k_{\rm B} T}\right)\right)} \cdot \frac{1}{90} \left(X \cdot aG'(\tilde{\nu}) + Y \cdot \gamma_{G'}^2(\tilde{\nu}) + Z \cdot \gamma_A^2(\tilde{\nu})\right)$$

Coefficient values from literature:

Scattering angle	Polarization	X	Y	Z
0°	$\Delta I^{\perp} = \Delta I^{\parallel}$	360	8	-8
0°	ΔI	720	16	-16
90°	ΔI^{\perp}	180	28	4
90°	ΔI^{\parallel}	0	24	-8
90°	ΔI	180	52	-4
180°	$\Delta I^{\perp} = \Delta I^{\parallel}$	0	48	16
180°	ΔI	0	96	32

M. Brehm, M. Thomas, J. Phys. Chem. Lett. 2017, 8 (14), 3409–3414.

ROA Spectrum



The goal is completely achieved 😳

First predicted bulk phase ROA spectrum.

Published in Summer 2017:





pubs.acs.org/JPCL

Computing Bulk Phase Raman Optical Activity Spectra from *ab initio* Molecular Dynamics Simulations

Martin Brehm*[®] and Martin Thomas

Institut für Chemie - Theoretische Chemie, Martin-Luther-Universität Halle-Wittenberg, Von-Danckelmann-Platz 4, 06120 Halle (Saale), Germany

S Supporting Information

ABSTRACT: We present our novel methodology for computing Raman optical activity (ROA) spectra of liquid systems from *ab initio* molecular dynamics (AIMD) simulations. The method is built upon the recent developments to obtain magnetic dipole moments from AIMD and to integrate molecular properties by using radical Voronoi tessellation. These techniques are used to calculate optical activity tensors for large and complex periodic bulk phase systems. Only AIMD simulations are required as input, and no time-consuming perturbation theory is involved. The approach relies only on the total electron density in each time step and can readily be combined with a wide range of electronic structure methods. To the best of our knowledge, these are the first computed ROA spectra for a periodic bulk phase system. As an example, the experimental ROA spectrum of liquid (R)-propylene oxide is reproduced very well.



M. Brehm, M. Thomas: "Computing Bulk Phase Raman Optical Activity Spectra from *ab initio* Molecular Dynamics Simulations", J. Phys. Chem. Lett. **2017**, 8 (14), 3409–3414.

Computational Resources

Used core hours (c.h.) on Intel Xeon "Haswell"@2.4 GHz:

2350 c.h. for equilibration
25500 c.h. for MD *(incl. 3 external field directions)*2000 c.h. to solve the PDE

 \rightarrow Approx 30 000 c.h. in total.

Takes ≈ 3 weeks on a "small" server with 64 cores.

Available for $\approx 5000 \notin$ today.



Computational Resources

What about disk storage?

We require the electron density (CUBE file) in every MD step...

Electron density grid is $160 \times 160 \times 160$ in our case.

CUBE file is 52 MiB per frame.

We require 4×65000 frames...

→ 13 Terabyte of data for one ROA spectrum



Two possible solutions:

- **1.)** On-the-fly processing
- 2.) Compress electron density data (lossless!)





Computational Resources

What about compression tools (bzip2)?

Can achieve ratio of $\approx 4.5 : 1 \rightarrow \text{still } 3 \text{ TiB...}$

Compression time is longer than AIMD simulation

 \rightarrow We needed a better solution.

4. Compression of Volumetric Data



What are Volumetric Data Trajectories?

Answer: A consecutive sequence of 3D Cartesian grid frames with real numbers

Examples:

- Electron Density
- Molecular Orbitals
- Electrostatic Potential



Who needs Volumetric Data Trajectories?

Typical Applications:

- Partial Charges from Electron Density
- Molecular Electric Dipole / Quadrupole Moments
- Vibrational Spectroscopy (IR, Raman, VCD, ROA)



How are Volumetric Data Trajectories stored?

"Standard" (since ≈40 years): Gaussian Cube Files

- Simple text file format
- Used by many programs (Gaussian, Orca, CP2k, CPMD, TurboMole, etc.)
- Trajectory is simple concatenation of single frames
- Typical resolution: 100 × 100 × 100 to 300 × 300 × 300

-Qui	ckstep-	-																	
ELE	CTRON)ENS	ITY																
32	9 O.	. 000	000	I	0.0	0000	00	. ().0	000	00								
16	9 O.	.184	053	I	0.0	0000	30	- (3.0	000	00								
16	0 O.	. 000	000	I	0.1	8405	53	. (0.0	000	00								
16	0 O.	. 000	000	I	0.0	0000	30	(9.1	840	53								
	5 O.	. 000	000	2	0.1	02 07	2	- 20).0	466	46	5	.1049	61					
	5 O.	. 000	000	2	1.8	1831	15	- 19).5	691	34	3	.0530	13					
	5 O.	. 000	000	18	8.1	7390	35	22	2.1	252	53	4	.9391	10					
	1 0.	. 000	000	2	0.8	1123	34	- 19	9.8	342	54	7	.0845	41					
	1 0.	. 000	000	1	8.4	4061	11	ť	5.0	737	38	9	.4747	05					
	1 0.	. 000	000		7.9	4988	39	7	.5	499	63	12	.6434	19					
:	3 Ø.	. 000	000	1	8.7	8956	6 O	11	1.4	024	96	7	.6581	00					
0.	79844E-	- 02	0.	7582	5E-	02	0.	6941	ŧ7E	- 02		0.615	31E-0	2	0.52995	E-02	0.1	44648	3E-02
0.:	37078E-	- 02	0.3	3062:	2E-	02	0.3	2541	I ØE	- 02		0.214	34E-0	2	0.18602	E-02	0.1	16791	1E-02
0.	15863E-	- 02	0.	15679	9E-	02	0.	1610	30E	- 02		0.169	88E-0	2	0.18205	E-02	0.1	19611	1E-02
0.:	21068E-	- 02	0.3	2244	4E-	02	0.3	2362	21E	- 02		0.245	05E-0	2	0.25035	E-02	0.2	25189	9E-02
0.:	24984E-	- 02	0.3	2446	8E-	02	0.3	2371	16E	- 02		0.228	08E-0	2	0.21822	E-02	0.2	20823	3E-02
0.	19854E-	- 02	0.	1893)	7E-	02	0.	1807	7 ØE	- 02		0.172	39E-0	2	0.16422	E-02	0.1	15592	4E-02
0.	14735E-	- 02	0.	1383)	7E-	02	0.	1290	95E	- 02		0.119	61E-0	2	0.11036	E-02	0.1	10174	4E-02
0.9	74286E-	- 03	0.	88572	2E-	03	0.	8521	ŧ3E	- 03		0.849	99E-0	3	0.88608	E-03	0.9	96908	3E-03
0.	11082E-	- 02	0.	1313!	5E-	02	0.	1595	53E	- 02		0.196	31E-0	2	0.24228	E-02	0.2	29723	3E-02
0.	35962E-	- 02	0.	4261	0E-	02	0.	4913	39E	- 02		0.548	74E-0	2	0.59111	E-02	0.0	61287	7E-02
0.0	51145E-	- 02	0.	5880!	5E-	02	0.	5471	ŧØE	- 02		0.496	42E-0	2	0.44264	E-02	0.3	39296	5E-02

Volumetric Data Structure

What we need:

• Efficient **lossless** compression algorithm *(to the accuracy of the input data)*

What we know:

 Volumetric data is spatially continuous (smooth, no sharp edges)

 \rightarrow can be exploited

- If from a MD simulation, data is also temporally continuous (*no abrupt jumps*)
 - \rightarrow can be exploited

General Idea

- Before processing each bin, estimate (*extrapolate*) its value based on earlier values (*both space and time*)
- Then look at the true value and only store the deviation from the estimate (*"residue"*)
- Finite precision: Deviation can be stored as integer numbers
- This converts the input grid of real values into a stream of integers
- If the extrapolation is good, the integers are small in value (reduction of information entropy!)
- Finally: Compress integer stream by entropy encoding

Polynomial Extrapolation

Is easy in 1D:



Polynomial Extrapolation

But how to proceed in more dimensions?



Ansatz (example):

 $F(x,y) = c_0 + c_x \cdot x + c_y \cdot y + c_{xy} \cdot xy + c_{x^2} \cdot x^2$

$$F(x,y) = c_0 + c_x \cdot x + c_y \cdot y + c_{xy} \cdot xy + c_{x^2} \cdot x^2$$

\rightarrow Linear System of Equations (7 equations, 5 unknowns):

$$\begin{aligned} F(x_0 - 1, y_0 - 2) &= c_0 + c_x(x_0 - 1) + c_y(y_0 - 2) + c_{xy}(x_0 - 1)(y_0 - 2) + c_{x^2}(x_0 - 1)^2 \\ F(x_0, y_0 - 2) &= c_0 + c_x x_0 + c_y(y_0 - 2) + c_{xy} x_0 + (y_0 - 2) + c_{x^2} x_0^2 \\ F(x_0 + 1, y_0 - 2) &= c_0 + c_x(x_0 + 1) + c_y(y_0 - 2) + c_{xy}(x_0 + 1)(y_0 - 2) + c_{x^2}(x_0 + 1)^2 \\ F(x_0 - 1, y_0 - 1) &= c_0 + c_x(x_0 - 1) + c_y(y_0 - 1) + c_{xy}(x_0 - 1)(y_0 - 1) + c_{x^2}(x_0 - 1)^2 \\ F(x_0, y_0 - 1) &= c_0 + c_x x_0 + c_y(y_0 - 1) + c_{xy} x_0 + (y_0 - 1) + c_{x^2} x_0^2 \\ F(x_0 + 1, y_0 - 1) &= c_0 + c_x(x_0 + 1) + c_y(y_0 - 1) + c_{xy}(x_0 + 1)(y_0 - 1) + c_{x^2}(x_0 + 1)^2 \\ F(x_0 - 1, y_0) &= c_0 + c_x(x_0 - 1) + c_y y_0 + c_{xy}(x_0 - 1) y_0 + c_{x^2}(x_0 - 1)^2 \end{aligned}$$

System is **over-determined**. → no exact solution, only least squares.

How to solve it? Bring it in matrix form!

 $A \cdot \mathbf{x} = \mathbf{b}$

M. Brehm, M. Thomas, J. Chem. Inf. Model. 2018, 58 (10), 2092–2107.



Space–Filling Hilbert Curve



This saves ~ 10% of space!

M. Brehm, M. Thomas, J. Chem. Inf. Model. 2018, 58 (10), 2092–2107.

Entropy Encoding

Starting with a stream of integers:

- Apply Burrows–Wheeler transformation
- Apply move-to-front transformation
- Apply run-length encoding
- Apply multi-table canonical Huffman coding
- \rightarrow Obtain a compressed bit-stream.

This is all very similar to bzip2.

Result: Volumetric Data

Electron Density, 36 [EMIm][OAc], 936 Atoms, Avg. over 1000 frames, Grid 216 x 216 x 216, $\Delta t = 0.5$ fs.



M. Brehm, M. Thomas, J. Chem. Inf. Model. 2018, 58 (10), 2092–2107.

Single Volumetric Frames

If there is ony 1 cube frame to compress, no temporal extrapolation is possible.

But spatial extrapolation can go to a higher order then $\ensuremath{\mathfrak{S}}$

- \rightarrow We still achieve a ratio of $\approx 20:1$
- \rightarrow Also very efficient for single cube frames.

Position Trajectories

Would this also work for *"normal"* position trajectories?

Use temporal extrapolation from last atom positions.

Result: Position Trajectory

36 [EMIm][OAc], 936 Atoms, Precision 10⁻⁵ Angstrom, Avg. over 1000 Frames, $\Delta t = 0.5$ fs.



M. Brehm, M. Thomas, J. Chem. Inf. Model. 2018, 58 (10), 2092–2107.

For both position trajectories (".xyz") and volumetric trajectories (".cube"), our format has by far the best compression ratio, but is still fast to read / write.

That's nice 🙂

Example: The 13 TiB of volumetric data from the ROA spectrum are now merely 350 GiB.

 \rightarrow Goal completely achieved.

The bqb File Format

Compressed data is saved in the newly developed bqb file format:

Space-efficient compression format



- Open source and well documented (in future ^(C))
- Can store all required data (cell vectors, atom labels, charges, comment lines, velocities, etc.)
- Contains headers & checksums \rightarrow Robust
- Contains index \rightarrow Fast seeking and random access

We hope that the bqb format is adopted in many programs and will be widely used.
The bqb File Format

Why didn't we use an existing format such as HDF5?



HDF5 is a general-purpose format with a **huge flexibility** for all different applications.

bqb is specifically designed for simulation trajectories, and aims at **maximum compression ratio**.

bqb stores bit streams and does not care for byte boundaries \rightarrow not a single bit is wasted.

 \rightarrow The two formats are not at all competitors.

Article

Article published in 2018:



M. Brehm, M. Thomas: "An Efficient Lossless Compression Algorithm for Trajectories of Atom Positions and Volumetric Data", *J. Chem. Inf. Model.* **2018**, *58* (10), 2092–2107.

Code is available

You can find the implementation and documentation on

www.brehm-research.de/bqb

• **bqbtool**: Command line tool for working with bqb files

What will be available soon?

- **libbqb**: C++ library to include in other projects
- Technical documentation of the bqb file format

All licensed under GNU <u>L</u>GPL v3.

Implemented in TRAVIS

Whole implementation is also included in TRAVIS (you can use TRAVIS instead of bqbtool).



Availability in CP2k

Since 2021, the bqb compression of total electron density is available directly inside of CP2k (from v8.1)!

```
&DFT

&PRINT

&E_DENSITY_BQB

...

&END E_DENSITY_BQB

&END PRINT

&END DFT
```

 \rightarrow see CP2k manual



Literature:

M. Brehm, M. Thomas, J. Chem. Inf. Model. 2018, 58 (10), 2092–2107.

Now we can compute bulk phase IR, Raman, VCD, ROA.

Everything finished? Of course not!

Now we can compute bulk phase IR, Raman, VCD, ROA.

Everything finished? Of course not!

5. Resonance Raman Spectroscopy

Main problem with Raman: Low sensitivity (1:10⁶)

Observation: If laser wavelength matches an electronic excitation, signal is enhanced by a factor of ≈ 1000

→ "Resonance Raman effect" (used since 1960s)

Coupling between vibrations and electronic excitations: Only normal modes which displace atoms involved in the electronic excitation are enhanced...

- Advantage: In solutions, only solute modes are enhanced, not solvent
 → Raman of diluted solutions becomes possible.
- Disadvantage: Intensity ratio between peaks changes drastically... → Non-resonant Raman predictions are useless.

 \rightarrow Need methods for <u>bulk phase</u> resonance Raman prediction

Up to 2019: No such spectrum has been predicted in literature...We developed a method to do so.

- Raman spectrum obtained from cross-correlation of static (f → 0) polarizability tensor along trajectory
- Resonance Raman spectrum requires the dynamic polarizability tensor... How to compute it?
- 1. From linear response TDDFT (LR-TDDFT)
 - Captures only the lowest *n* electronic transitions
 - Only discrete lines (linear response approximation)
- 2. From real-time TDDFT (RT-TDDFT)
 - Captures <u>all</u> electronic transitions at once
 - Realistic band shape of dynamic polarizability



Apply external field step in X/Y/Z direction, evaluate dipole time series in X/Y/Z direction

→ Obtain all 9 components of dynamic polarizability tensor for all frequencies at once

- **1.** Run a standard BOMD trajectory of the system
- 2. Take a snapshot each 4 fs and run RT-TDDFT trajectories
- **3.** Obtain the full dynamic polarizability tensor for each snapshot
- **4.** Compute the resonance Raman spectrum via temporal cross-correlation of the dynamic polarizability





Uracil is a worthwhile model system:

- **1.** Important and ubiquitous in biological systems
- **2.** Extended π system \rightarrow Absorption in near UV
- Forms strong hydrogen bonds to water
 → Strong solvent influence:





The first *ab initio* prediction of a bulk phase resonance Raman spectrum.

M. Brehm et al., J. Chem. Theor. Comput. 2019, 15 (7), 3901–3905.



All possible resonance Raman spectra in one plot! ("Excitation Profile")

Reveals details on vibronic coupling

→ Allows to design interesting experiments



Vormalized Raman Spectrum

M. Brehm et al., J. Chem. Theor. Comput. 2019, 15 (7), 3901–3905.

Conclusions

- The "full set" of vibrational spectra (Infrared, Raman, VCD, ROA) can be routinely computed for complex bulk phase systems now
- Since 2019, even resonance Raman spectra are possible
- Includes full solvent effect, realistic line shapes, some anharmonic effects (approx. overtones, combination bands)
- All methods are implemented in the TRAVIS program package: http://www.travis-analyzer.de
- The full workflow "Idea → Spectrum" only relies on free software (CP2k, bqbtool, TRAVIS)
- A detailed **step-by-step tutorial** is available online:

https://brehm-research.de/spectroscopy

Conclusions

CP2k + TRAVIS is a powerful combination for computing vibrational spectra.



My personal "wish list":

- Find a way to apply PERIODIC_EFIELD to systems which require mixing and Fermi–Dirac smearing, so that e.g. Raman spectra on metal surfaces can be predicted.
- Make AIMD with k-points possible by implementing wave function extrapolation with k-points.

Acknowledgment



MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

"AK Brehm"





Prof. Daniel Sebastiani

https://brehm-research.de/ Funding by DFG ("Eigene Stelle" + 1 PhD)