

# Fast Calculation of the Excitation Energies for Compact Molecules by using Tensor-structured Methods

Venera Khoromskaia

(based on joint works with Boris Khoromskij, Peter Benner, Sergey Dolgov and Chao Yang)  
Annual meeting of the GAMM Activity Group MOANSI  
Zuse Institute Berlin, 25-26.10, 2018

Max-Planck-Institute for Mathematics in the Sciences, Leipzig  
Max-Planck-Institute for DCTS, Magdeburg



## Grid-based tensor methods for the Hartree-Fock equation and TEI calculus

**The Hartree-Fock** (HF) nonlinear 3D integral-differential spectral problem (EVP)

$$\mathcal{F}\varphi_i(x) \equiv \left(-\frac{1}{2}\Delta + V_c + V_H - \mathcal{K}\right)\varphi_i(x) = \lambda_i \varphi_i(x), \quad i = 1, \dots, N_{orb}.$$

The Fock operator  $\mathcal{F}$  depends on  $\tau(x, y) = 2\sum_{i=1}^{N_{orb}} \varphi_i(x)\varphi_i(y)$ ,

$$\mathcal{F}\varphi := \left[-\frac{1}{2}\Delta - \sum_{\nu=1}^{M_0} \frac{Z_\nu}{\|x - a_\nu\|} + \int_{\mathbb{R}^3} \frac{\tau(y, y)}{\|x - y\|} dy\right]\varphi - \frac{1}{2} \int_{\mathbb{R}^3} \frac{\tau(x, y)}{\|x - y\|} \varphi(y) dy.$$

**Challenges:** High accuracy, 3D convolutions, nuclear cusps, nonlinear EVP, post HF.

- Grid-based tensor methods for HF equation.  
[Khoromskij, Khoromskaia, Schneider, Flad, Andrae, '08-'14],
- Rank approximation of the two-electron integrals (TEI)  
[Khoromskaia, Khoromskij, Schneider, '13]
- Post Hartree-Fock: excitation energies via rank-structured approx. to BSE system  
[Benner, Khoromskaia, Khoromskij, Dolgov, Yang, '15-'18],
- $L \times L \times L$  lattice sums of electrostatic potentials,  $1/\|x\|$ ,  $x \in \mathbb{R}^3$ , in  $O(L)$ .  
[Khoromskaia, Khoromskij, '14-'16],
- Range-separated (RS) tensor format for long range interaction potentials.  
[Benner, Khoromskaia, Khoromskij, '16]

[Szabo, Ostlund '96], [Helgaker, Jørgensen and Olsen, 2000],

Using GTO basis set,  $\varphi_i(x) = \sum_{\mu=1}^{N_b} c_{\mu i} g_{\mu}(x)$ , we obtain the Galerkin system for coefficients matrix  $C = \{c_{\mu i}\} \in \mathbb{R}^{N_b \times N_{orb}}$ , and density matrix  $D = 2CC^* \in \mathbb{R}^{N_b \times N_b}$ ,

$$F(D)C = SC\mathcal{E}, \quad \mathcal{E} = \text{diag}(\varepsilon_1, \dots, \varepsilon_{N_{orb}}), \quad C^T SC = I_{N_b},$$

where  $F(D) = H_0 + J(D) + K(D)$  and  $S$  is the mass matrix.

Precomputed: core Hamiltonian  $H_0 = \{h_{\mu\nu}\}$ ,  $1 \leq \mu, \nu \leq N_b$ ,

$$h_{\mu\nu} = \frac{1}{2} \int_{\mathbb{R}^3} \nabla g_{\mu} \cdot \nabla g_{\nu} dx + \int_{\mathbb{R}^3} V_c(x) g_{\mu} g_{\nu} dx,$$

and the two-electron integrals (TEI) tensor  $\mathbf{B} = [b_{\mu\nu\kappa\lambda}]$ ,

$$b_{\mu\nu\kappa\lambda} = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{g_{\mu}(x) g_{\nu}(x) g_{\kappa}(y) g_{\lambda}(y)}{\|x - y\|} dx dy, \quad \mu, \nu, \kappa, \lambda = 1, \dots, N_b.$$

The nonlinear EVP is solved by DIIS iteration [Pulay '80] and updating

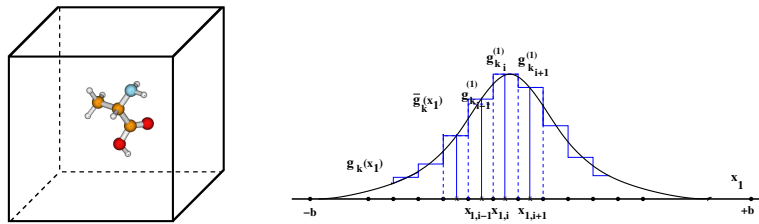
$$J(D)_{\mu\nu} = \sum_{\kappa, \lambda=1}^{N_b} b_{\mu\nu, \kappa\lambda} D_{\kappa\lambda}, \quad K(D) = -\frac{1}{2} \sum_{\kappa, \lambda=1}^{N_b} b_{\mu\lambda, \nu\kappa} D_{\kappa\lambda}.$$

Standard HF packages : all 3D integrals are analytically precomputed.

## Tensor-structured approach for 3D functions and operators

[Khoromskij & Khoromskaia '07-'12 ]

[Khoromskaia & Andrae & Khoromskij, CPC '12],



Box:  $[-b, b]^3$ ,  $b \approx 15 \text{ \AA}$ ,  $n \times n \times n$  3D Cartesian grids,  $n^3 \sim 10^{15}$   
the continuous basis functions  $g_{\mu}(x) : \mathbf{l}_0 : g_{\mu} \rightarrow \bar{g}_{\mu} := \sum_{i \in \mathcal{I}} g_{\mu}(x_i) \zeta_i(x)$ .

$$g_{\mu}(x) \approx \mathbf{G}_{\mu} = \mathbf{g}_{\mu}^{(1)} \otimes \mathbf{g}_{\mu}^{(2)} \otimes \mathbf{g}_{\mu}^{(3)}, \quad \mathbf{G}_{\mu} \in \mathbb{R}^{n \times n \times n}, \quad \mathbf{g}_{\mu} \in \mathbb{R}^n$$

Core Hamiltonian (Laplace and nuclear potential operators):

$$H_0 = A_3 + V, \quad \text{where } a_{\mu\nu} = \langle \Delta_3 \mathbf{G}_{\mu}, \mathbf{G}_{\nu} \rangle, \quad v_{\mu\nu} = \langle \mathbf{G}_{\mu} \odot \mathbf{G}_{\nu}, \sum_{a=1}^A \mathbf{P}_{N,a} \rangle,$$

$$\Delta_3 = \Delta_1^{(1)} \otimes I^{(2)} \otimes I^{(3)} + I^{(1)} \otimes \Delta_1^{(2)} \otimes I^{(3)} + I^{(1)} \otimes I^{(2)} \otimes \Delta_1^{(3)}.$$

$\mathbf{P}_N = \sum_{q=1}^R \mathbf{p}_q^{(1)} \otimes \mathbf{p}_q^{(2)} \otimes \mathbf{p}_q^{(3)} \in \mathbb{R}^{2n \times 2n \times 2n}$  is the tensor representation of the Newton kernel developed in [Bertoglio, Khoromskij '08-'10].

## Example: 3D convolution integrals in 1D complexity (HF and KS eqn.)

[Khoromskij, Khoromskaia '08 (SISC 2009)]

$$J_{\mu\nu} := \int_{\mathbb{R}^3} \mathbf{g}_\mu(x) \mathbf{g}_\nu(x) V_H(x) dx, \quad \mu, \nu = 1, \dots, N_b \quad x \in \mathbb{R}^3,$$

$$V_H(x) := \int_{\mathbb{R}^3} \frac{\rho(y)}{\|x-y\|} dy \quad \rho(x) = 2 \sum_{i=1}^{N_{orb}} (\varphi_i)^2, \quad \varphi_i(x) = \sum_{\mu=1}^{N_b} c_{i\mu} \mathbf{g}_\mu(x), \quad i = 1, \dots, N_{orb}.$$

$$\rho \approx \Theta = \sum_{i=1}^{N_{orb}} \sum_{\mu=1}^{N_b} \sum_{\nu=1}^{N_b} C_{i\nu} C_{i\mu} (\mathbf{g}_\mu^{(1)} \odot \mathbf{g}_\nu^{(1)}) \otimes (\mathbf{g}_\mu^{(2)} \odot \mathbf{g}_\nu^{(2)}) \otimes (\mathbf{g}_\mu^{(3)} \odot \mathbf{g}_\nu^{(3)}).$$

C2T + T2C to reduce the rank,  $\Theta \rightarrow \Theta' := \sum_{t=1}^{R_\rho} \mathbf{u}_t^{(1)} \otimes \mathbf{u}_t^{(2)} \otimes \mathbf{u}_t^{(3)}$ .

Newton kernel  $\frac{1}{\|x\|} \Rightarrow \mathbf{P}_N = \sum_{q=1}^{R_N} \mathbf{p}_q^{(1)} \otimes \mathbf{p}_q^{(2)} \otimes \mathbf{p}_q^{(3)}$ .

$$V_H \approx \mathbf{V}_H = \Theta' * \mathbf{P}_N = \sum_{t=1}^{R_\rho} \sum_{q=1}^{R_N} c_m b_k (\mathbf{u}_t^{(1)} * \mathbf{p}_q^{(1)}) \otimes (\mathbf{u}_t^{(2)} * \mathbf{p}_q^{(2)}) \otimes (\mathbf{u}_t^{(3)} * \mathbf{p}_q^{(3)}),$$

Result :  $J_{\mu\nu} \approx \langle \mathbf{G}_\mu \odot \mathbf{G}_\nu, \mathbf{V}_H \rangle \quad \mu, \nu = 1, \dots, N_b.$

## Tensor-product convolution vs. 3D FFT

[Khoromskij& Khoromskaia, 2008 (SISC 2009)]

Tensor product convolution (1D FFT) ( $O(n \log n)$ ) instead of  $O(n^3 \log n)$  for 3D FFT.

$n^3$	$512^3$	$1024^3$	$2048^3$	$4096^3$	$8192^3$	$16384^3$
FFT <sub>3</sub>	5.4	~ 51.6	–	–	–	~ 7 days
$C * C$	1.5	8.8	20.0	61.0	157.5	299.2
C2T	5.6	6.9	10.9	20.0	37.9	86.0

Table shows CPU time (in sec) for the computation of  $V_H$  for  $H_2O$ .  
(3D FFT time for  $n \geq 1024$  is obtained by extrapolation).

Tensor approximation of the Newton kernel is based on the Laplace transform and sinc-quadrature approximation.

Theory: [Stenger '93], [Gavrilyuk, Hackbusch, Khoromskij '08],

Practice: Canonical tensor for the Newton kernel: [Bertoglio, Khoromskij '08]

$$\mathbf{P}_N = \sum_{q=1}^{R_N} \mathbf{p}_q^{(1)} \otimes \mathbf{p}_q^{(2)} \otimes \mathbf{p}_q^{(3)} \in \mathbb{R}^{n \times n \times n}$$

with the canonical rank,  $R_N \approx 30 \div 40$ .

[Khoromskaia & Khoromskij & Schneider, SISC'13]

$$b_{\mu\nu\kappa\lambda} = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{g_\mu(x)g_\nu(x)g_\kappa(y)g_\lambda(y)}{\|x-y\|} dx dy = \langle \mathbf{G}_\mu \odot \mathbf{G}_\nu, \mathbf{P}_N * (\mathbf{G}_\kappa \odot \mathbf{G}_\lambda) \rangle_{n \otimes 3}.$$

- $G^{(\ell)} = [G_\mu^{(\ell)} \odot G_\nu^{(\ell)}]_{1 \leq \mu, \nu \leq N_b} \in \mathbb{R}^{n \times N_b^2} \quad \ell = 1, 2, 3; n \approx 10^5, N_b^2 \approx 10^4.$
- $G^{(\ell)} \approx U^{(\ell)} V^{(\ell)T}, \quad U^{(\ell)} \in \mathbb{R}^{n \times R_\ell}, V^{(\ell)} \in \mathbb{R}^{N_b^2 \times R_\ell}, R_\ell \sim N_b.$
- Then for TEI  $\mathbf{B} = [b_{\mu\nu\kappa\lambda}]$ , its matrix

$$B := \text{mat}(\mathbf{B}) \approx B_\varepsilon := \sum_{k=1}^{R_N} \odot_{\ell=1}^3 V^{(\ell)} M_k^{(\ell)} V^{(\ell)T} \in \mathbb{R}^{N_b^2 \times N_b^2}$$

is computed as vectors of its Cholesky factors (columns  $B(:,j)$  and diagonal elements  $B(i,i)$ ), where the convolution matrix  $M_k^{(\ell)} = U^{(\ell)T} (P_k^{(\ell)} *_n U^{(\ell)}) \in \mathbb{R}^{R_\ell \times R_\ell}$ .

Here  $P^{(\ell)} \in \mathbb{R}^{n \times R_N}$  are the factor matrices in rank- $R_N$  canonical tensor  $\mathbf{P}_N \in \mathbb{R}^{n \times n \times n}$ ,  $n \approx 10^5$  representing the Newton kernel,  $\frac{1}{\|x\|}$ .

Then TEI is represented by Cholesky decomposition ( $\varepsilon$ -approx. )

$$B \approx LL^T, \quad L \in \mathbb{R}^{N_b^2 \times R_B} \quad R_B \sim N_b$$

## Tensor-based Hartree-Fock solver

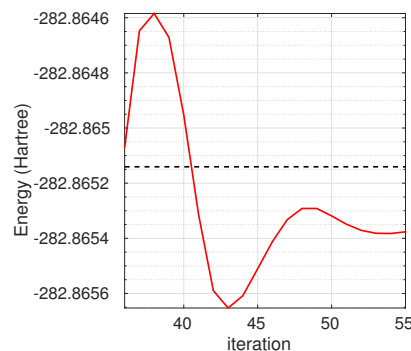
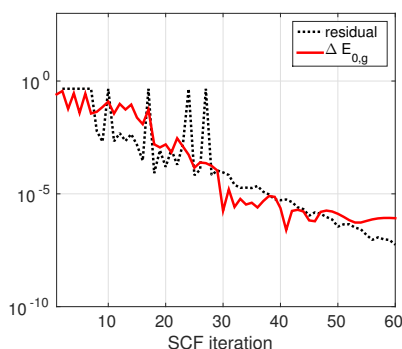
Tensor-based Electronic Structure Calculations (TESC) package  
[Khoromskaia & Khoromskij, 2008-2014]

Hartree-Fock solver: all parts of the Fock operator are computed by tensor-structured numerical methods (using 3D Cartesian grids but all operations in 1D complexity)

$$F(C)C = SC\mathcal{E}, \quad F = H_0 + J(C) - K(C),$$

▷ Coulomb matrix: given  $\bar{D} = \text{vec}(D)$ ,  $\text{vec}(J) = B\bar{D} \approx L(L^T\bar{D})$ .

▷ HF exchange:  $K(D)_{\mu\nu} = - \sum_{i=1}^{N_{orb}} \sum_{k=1}^{R_B} (\sum_\lambda L_{\mu\lambda k} C_{\lambda i}) (\sum_\kappa C_{\kappa i} L_{\kappa\nu k})$ ,



Glycine: last  $k + 27$  iterations; grid for TEI:  $n^3 = 131072^3$ .

The Bethe-Salpeter equation provides calculation of the excitation energies of molecular structures.

=> optical properties of molecules, nano-structures and solids: in photoluminescence, light-emitting devices, laser technology, and photovoltaics.

=> the BSE model was developed during several decades (Many-Body Perturbation Theory, Green Function Approximation for fermions in Quantum Electrodynamics, etc.) : [Dyson 1949], [Feynmann 1949], [Salpeter, Bethe 1951], [Hedin 1965], [Louie, Rohlfing 1998], [Runge, Gross 1984], [Casida 1995], [Onida, Reining, Rubio 2002],...

The Bethe-Salpeter equation is an eigenvalue problem (EVP) w.r.t. Hamiltonian of size  $N^2 \times N^2$ , where  $N = N_b$  is the number of atomic orbital basis functions,

$$\mathcal{H}\psi = E\psi.$$

Alternative: Full Configuration Interaction, Coupled Clusters, etc. (extremely computationally consuming).

## Calculation of excitation energies for molecules

The Bethe-Salpeter equation (BSE)

$$\mathcal{H}\psi = E\psi.$$

The main problem:  $O(N^6)$  scaling of EVP.

Our approach with computational (and storage) cost  $O(N^2) \div O(N^3)$ :

- 1 Self-consistent ab-initio calculation of the ground state energy:  
⇒ one-electron orbitals and energies (in our case Hartree-Fock equation).
- 2 Construction of BSE system matrix in molecular orbitals basis using the output of HF solver and low-rank approximation to involved quantities.
- 3 Fast iterative solution of BSE, obtaining the lowest (in magnitude) part of spectrum.

[Benner & Khoromskaia & Khoromskij, Mol.Phys.'16]

We use the formulation of the BSE model in the framework of noninteracting Green's function in the MO basis as in [Ribolini, Toulouse, Savin 2013].  
( the part of it corresponding to one-particle Green's function).

## Prerequisites from Hartree-Fock calculations

The prerequisites for generating matrices in the BSE EVP are ab-initio Hartree-Fock calculations: [\[Khoromskaia CMAM '14\]](#)

- 1 Full set of ground state energies (eigenvalues of the Hartree-Fock EVP)

$$\varepsilon_1, \dots, \varepsilon_{N_b}.$$

- 2 The full set of Galerkin coefficients in the expansion of molecular orbitals in Gaussian basis,

$$C = \{c_{\mu i}\} \in \mathbb{R}^{N_b \times N_b}.$$

- 3 TEI matrix  $B = [b_{\mu\nu, \kappa\lambda}] \in \mathbb{R}^{N_b^2 \times N_b^2}$  is precomputed as a low-rank Cholesky factorization,

$$B \approx LL^T, \quad L \in \mathbb{R}^{N_b^2 \times R_B}, \quad R_B = O(N_b)$$

presented in molecular orbitals basis, [\[Khoromskaia & Khoromskij, CPC '14\]](#)

$$B \mapsto \mathbf{V}_{MO} = [v_{iajb}], \quad \text{where } v_{iajb} = \sum_{\mu, \nu, \kappa, \lambda=1}^{N_b} c_{\mu i} c_{\nu a} c_{\kappa j} c_{\lambda b} b_{\mu\nu, \kappa\lambda}.$$

Here indices  $i, j \in \mathcal{I}_o$  correspond to occupied orbitals,  $\mathcal{I}_o := \{1, \dots, N_{orb}\}$ , and  $a, b \in \mathcal{I}_v$  to virtual  $\mathcal{I}_v := \{N_{orb}, \dots, N_b\}$ . Denote  $N_v = N_b - N_{orb}$ , and  $N_{ov} = N_{orb}N_v$ .

## Construction of BSE matrix

[\[Benner & Khoromskaia & Khoromskij, MolPhys '16\]](#)

The construction of BSE matrix includes computation of several auxiliary quantities.

- The diagonal "energy" matrix

$$\Delta\varepsilon = [\Delta\varepsilon_{ia,jb}] \in \mathbb{R}^{N_{ov} \times N_{ov}}, \quad \text{where } \Delta\varepsilon_{ia,jb} = (\varepsilon_a - \varepsilon_i)\delta_{ij}\delta_{ab}.$$

$\varepsilon_a$  and  $\varepsilon_i$  correspond to virt. and occ. one-electron energies.

Energy matrix can be represented in the Kronecker product form

$$\Delta\varepsilon = I_o \otimes \text{diag}\{\varepsilon_a : a \in \mathcal{I}_v\} - \text{diag}\{\varepsilon_i : i \in \mathcal{I}_o\} \otimes I_v,$$

where  $I_o$  and  $I_v$  are the identity matrices on respective index sets.

Matrix  $\Delta\varepsilon$  is invertible if the *homo lumo* gap of the system is positive,

$$\varepsilon_a - \varepsilon_i > \delta > 0, \quad a \in \mathcal{I}_v, i \in \mathcal{I}_o,$$

- Using  $\Delta\varepsilon$  and  $V = [v_{ia,jb}]$  the *dielectric function*  $Z = [z_{pq,rs}] \in \mathbb{R}^{N_{ov} \times N_{ov}}$  is defined by

$$z_{pq,rs} := \delta_{pr}\delta_{qs} - v_{pq,rs}[\chi_0(\omega = 0)]_{rs,rs},$$

where  $\chi_0(\omega)$  is the matrix of the so-called Lehmann representation to the *response function*

$$\chi_0(0) = - \begin{pmatrix} \Delta\varepsilon^{-1} & \mathbf{0} \\ \mathbf{0} & \Delta\varepsilon^{-1} \end{pmatrix}.$$

Let  $\mathbf{1} \in \mathbb{R}^{N_{ov}}$  and  $\mathbf{d}_\varepsilon = \text{diag}\{\Delta\varepsilon^{-1}\} \in \mathbb{R}^{N_{ov}}$  be the all-ones and diagonal vectors of  $\Delta\varepsilon^{-1}$ , respectively, specifying the rank-1 matrix  $\mathbf{1} \cdot \mathbf{d}_\varepsilon^T$ .  
In this notations the matrix  $Z = [z_{pq,rs}]$  takes a compact form

$$Z = I_o \otimes I_v + V \odot (\mathbf{1} \cdot \mathbf{d}_\varepsilon^T).$$

- The *static screened interaction* matrix (tensor) defined by

$$W = [w_{pq,rs}] : \quad w_{pq,rs} := \sum_{t \in \mathcal{I}_v, u \in \mathcal{I}_o} z_{pq,tu}^{-1} v_{tu,rs}. \quad (1)$$

Eq. (1) is considered on conventional index set  $\{p, r \in \mathcal{I}_o\} \cup \{q, s \in \mathcal{I}_v\}$ .  
For example, we have the following matrix factorization of  $W := [w_{ia,jb}]$ ,

$$W = Z^{-1} V \quad \text{provided that} \quad a, b \in \mathcal{I}_v, \quad i, j \in \mathcal{I}_o.$$

Here we follow the BSE scheme for  $H_2$  from [\[Ribolini & Toulouse & Savin 2013\]](#), accomplished by [factorized TEI](#) (usually not available in standard packages),  
⇒ thus their HF-BSE scheme becomes valid for larger molecular systems.

## The Bethe-Salpeter eigenvalue problem for rank-structured matrices

EVP for  $2N_{ov} \times 2N_{ov}$ -BSE matrix to determine the excitation energies  $\omega_n$

$$F_1 \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix} \equiv \begin{pmatrix} A & B \\ -B^* & -A^* \end{pmatrix} \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix} = \omega_n \begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{pmatrix}, \quad (2)$$

where the matrix blocks are defined in the index notation by

$$\begin{aligned} a_{ia,jb} &:= \Delta\varepsilon_{ia,jb} + v_{ia,jb} - w_{ij,ab}, \\ b_{ia,jb} &:= v_{ia,bj} - w_{ib,aj}, \quad a, b \in \mathcal{I}_v, \quad i, j \in \mathcal{I}_o. \end{aligned}$$

- In the matrix form

$$A = \Delta\varepsilon + V - \overline{W}, \quad B = V - \widetilde{W}.$$

For the matrix elements in  $\overline{W} = [\overline{w}_{ia,jb}]$ :  $\overline{w}_{ia,jb} = w_{ij,ab}$ .

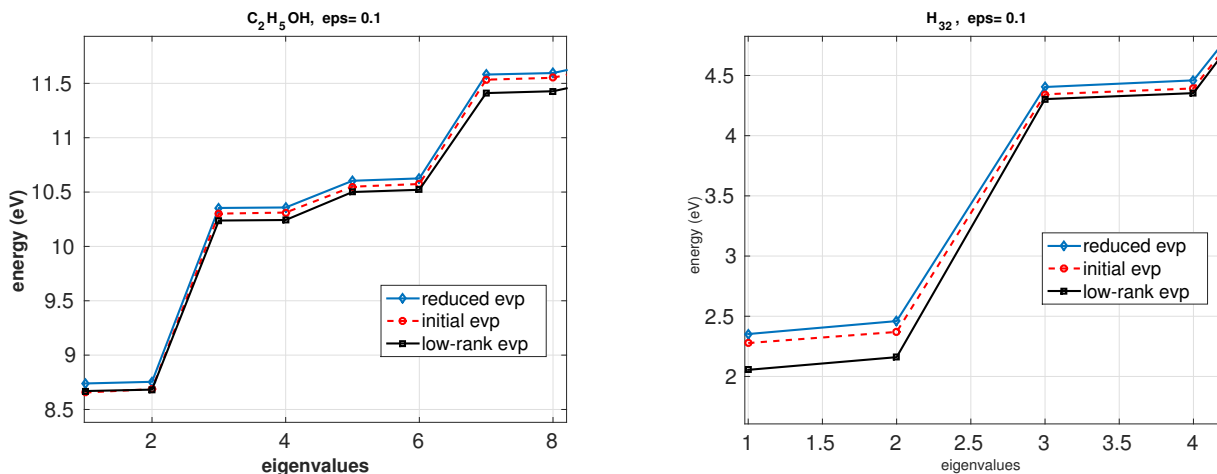
$\widetilde{W}$  is defined by permutation of  $W$ :  $\widetilde{W} = [\widetilde{w}_{ia,jb}] = [w_{ib,aj}]$ .

- The diagonal + low-rank sparsity in  $\Delta\varepsilon + V$  and  $\widetilde{W}$  can be recognized.





## Surprising result I: Two-sided bounds for excitation energies (red. bl. $\overline{W}$ )



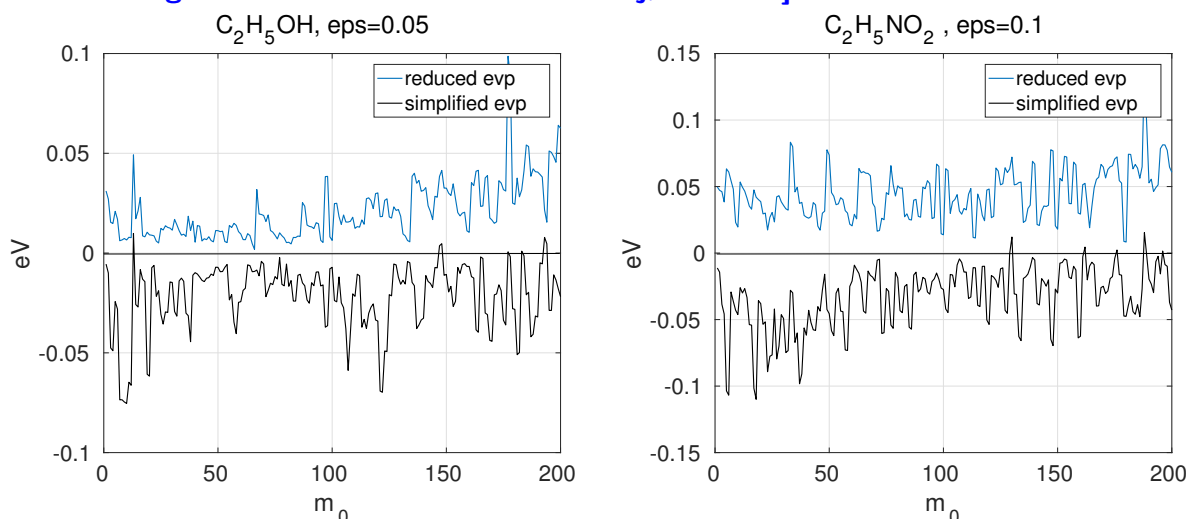
Errors (in eV) for simplified and reduced BSE eigenvalues for Ethanol ( $C_2H_5OH$ ) molecule and  $H_{32}$ -chain.

Lower bound is hard to prove: open question!

► Physical interpretation: open question!

## Two-sided bounds for excitation energies (reduced block $\overline{W}$ )

[Benner & Dolgov & Khoromskaia & Khoromskij, JCP '17]



Errors (in eV) in  $m_0$  smallest eigenvalues for simplified and reduced schemes: ethanol  $C_2H_5OH$  (left), Glycine amino acid  $C_2H_5NO_2$  (right).

Molecule	$H_2O$	$N_2H_4$	$C_2H_5OH$	$H_{32}$	$C_2H_5NO_2$	$H_{48}$	$C_3H_7NO_2$
TDA size	$180^2$	$657^2$	$1430^2$	$1792^2$	$3000^2$	$4032^2$	$4488^2$
$ \overline{\gamma}_1 - \omega_1 $	0.02	0.03	0.08	0.07	0.05	0.10	0.1

Errors (in eV) for reduced-block approximation to BSE eigenvalues ( $\varepsilon = 0.1$ ).

## [Benner & Dolgov & Khoromskaia & Khoromskij, JCP '17]

► Compute eigenvalues with smallest magnitudes taking advantage from the rank-structured approximation of the matrix inverse  $A_0^{-1}$  and  $F_0^{-1}$ . Define

$$\begin{aligned} A_0 &= \mathbf{\Delta}\epsilon + PQ^\top, & P &= [L_V \quad L_W], & Q &= [L_V \quad -L_W], \\ B_0 &= \Phi\Psi^\top, & \Phi &= [L_V \quad Y], & \Psi &= [L_V \quad -Z], & r &= R_V. \end{aligned}$$

► First consider the Tamm-Duncoff approximation (TDA),

$$A_0 \mathbf{x}_n = \mu_n \mathbf{x}_n, \quad \mathbf{x}_n \in \mathbb{R}^{N_{ov}} \quad A_0 \in \mathbb{R}^{N_{ov} \times N_{ov}}$$

Inverting a (block) diagonal plus low rank matrix: use the S-M-W formula for  $A_0^{-1}$

$$A_0^{-1} = \mathbf{\Delta}\epsilon^{-1} - \mathbf{\Delta}\epsilon^{-1}P \left( I + Q^\top \mathbf{\Delta}\epsilon^{-1}P \right)^{-1} Q^\top \mathbf{\Delta}\epsilon^{-1}.$$

► The inner  $2r \times 2r$  matrix  $K = (I + Q^\top \mathbf{\Delta}\epsilon^{-1}P)^{-1}$  is small and can be computed explicitly at  $\mathcal{O}(r^3 + r^2 N_{ov})$  operations.

► Matrix-vector product  $A_0^{-1} \mathbf{x}_n$  for the diagonal  $\mathbf{\Delta}\epsilon^{-1}$  and low-rank matrix in the second summand at the overall cost  $\mathcal{O}(N_{ov}r)$ .

## Linear scaling method: diagonal + low rank + reduced block structure

Times (s) for eig-solvers using reduced block representation of TDA and BSE systems.

Molecular syst.	H <sub>2</sub> O	N <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>5</sub> OH	H <sub>32</sub>	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	H <sub>48</sub>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>
TDA size	180 <sup>2</sup>	657 <sup>2</sup>	1430 <sup>2</sup>	1792 <sup>2</sup>	3000 <sup>2</sup>	4032 <sup>2</sup>	4488 <sup>2</sup>
EIG( $A_0$ )	0.02	0.5	4.3	9.8	37.6	91	127.4
TDA: EIGS( $A_0$ )	0.09	0.33	2.8	0.77	16.1	3.0	30
TDA: EIGS( $A_0^{-1}$ )	0.07	0.09	0.25	0.77	0.54	3.0	1.0
BSE size	360 <sup>2</sup>	1314 <sup>2</sup>	2860 <sup>2</sup>	3584 <sup>2</sup>	6000 <sup>2</sup>	8064 <sup>2</sup>	8976 <sup>2</sup>
EIG( $F_0$ )	0.08	4.2	33.7	68.1	274	649	903
BSE: EIGS( $F_0^{-1}$ )	0.21	0.37	1.11	1.10	2.4	2.92	4.6

Note: accuracy is better by the order of magnitude compared with “diag + low-rank” structure.

**CPU time scales linearly in the system size  $O(N_{ov})$  !**

[Benner & Dolgov & Khoromskaia & Khoromskij, JCP '17]

Fast QTT-structured BSE solvers based on low-rank + reduced block + diagonal format.

**Proof of concept: estimating QTT ranks:**  $r_{QTT} \lesssim N_o$

Molecular sys.	H <sub>2</sub> O	H <sub>16</sub>	N <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>5</sub> OH	H <sub>32</sub>	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>
$N_o$	5	8	9	13	16	20	24
$r_{QTT}$ of $L_V$	5.4	7	9.1	12.7	14	17.5	21
$r_{QTT}$ of eig-vect.	5.3	7.6	9.1	12.7	13.6	17.2	20.9
$N_{ov}$	180	448	657	1430	1792	3000	4488

QTT ranks of column vectors in  $L_V$  for  $m_0 = 30$  "minimal" TDA-eigenvectors.

H <sub>24</sub> , $N_o = 12$	$N_b$	72	96	144	168
	size BSE	1440 <sup>2</sup>	2016 <sup>2</sup>	3168 <sup>2</sup>	3744 <sup>2</sup>
	QTT ranks	9.5	11.6	11.8	12.7

Average QTT ranks of column vectors in  $L_V$  factor vs.  $N_o$  for Hydrogen chains: weak dependence on the number of basis functions  $N_b$ :  $r_{QTT} \lesssim N_o$ .

## Surprising result II: Lower bound on the algebraic complexity of BSE, $\mathcal{O}(N_o^2)$

Usual relation:  $N_b/N_o \geq C_{GTO} \approx 10$ , i.e.  $N_{ov} \approx C_{GTO} N_o^2$

⇒ quasi-optimal complexity of QTT solver

**Hypothesis:** Estimate on the lower bound on the asymptotic algebraic complexity of the large scale BSE eigenvalue problem

$$\mathcal{W}_{BSE} = \mathcal{O}(\log(N_{ov})r_{QTT}^2) = \mathcal{O}(\log(N_o)N_o^2).$$

determines the irreducible lower bound on the asymptotic algebraic complexity of the large scale BSE eigenvalue problem.

Molecular syst.	C <sub>2</sub> H <sub>5</sub> OH	H <sub>32</sub>	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	H <sub>48</sub>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>
TDA size	1430 <sup>2</sup>	1792 <sup>2</sup>	3000 <sup>2</sup>	4032 <sup>2</sup>	4488 <sup>2</sup>
time QTT eig	0.14	0.23	0.32	0.28	0.63
abs. error (eV)	0.08	0.19	0.17	0.14	0.00034

Times (s) and abs. error (eV) for QTT-DMRG eigensolvers for TDA.

DOS for symmetric matrices [L. Lin, Y. Saad, C. Yang, SIAM Review '16]

$$\phi(t) = \frac{1}{n} \sum_{j=1}^n \delta(t - \lambda_j), \quad t, \lambda_j \in [0, a]; \quad Au_j = \lambda_j u_j, \quad A = A^T.$$

A Lorentzian broadening of the Tamm-Dancoff (TDA) absorption spectrum (DOS).

► Blurring (regularization) by Lorentzians:

$$\delta(t) \rightsquigarrow L_\eta(t) := \frac{1}{\pi} \frac{\eta}{t^2 + \eta^2} = \frac{1}{\pi} \text{Im} \left( \frac{1}{t - i\eta} \right)$$

Reduction to the trace calculation

$$\phi(t) \mapsto \phi_\eta(t) := \frac{1}{n} \sum_{j=1}^n L_\eta(t - \lambda_j) = \frac{1}{n\pi} \text{Trace} [((tI - A)^2 + \eta^2 I)^{-1}].$$

$$\phi(t) \mapsto \phi_\eta(t) := \frac{1}{n\pi} \text{Im} \sum_{j=1}^n \frac{1}{(t - \lambda_j) - i\eta} = \frac{1}{n\pi} \text{Im} \text{Trace} [(tI - A - i\eta I)^{-1}].$$

Gaussian blurring:

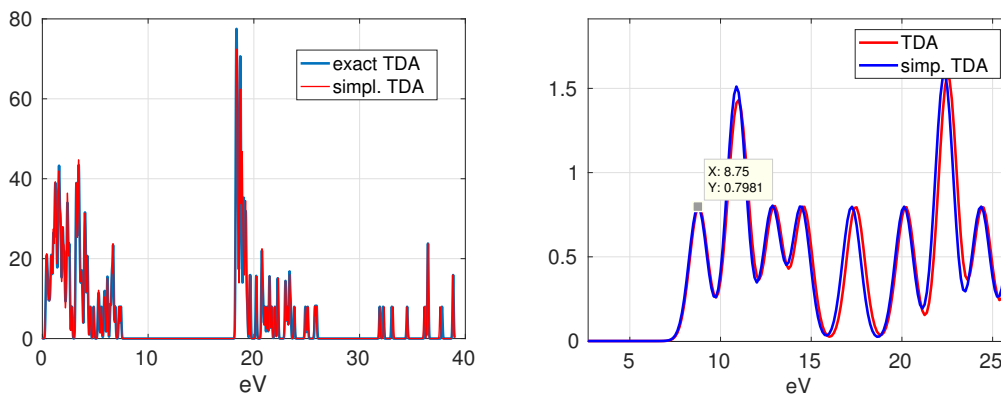
$$\delta(t) \rightsquigarrow g_\eta(t) = \frac{1}{\sqrt{2\pi\eta}} \exp\left(-\frac{t^2}{2\eta^2}\right) \Rightarrow \phi(t) \approx \phi_\eta(t) := \frac{1}{n} \sum_{j=1}^n g_\eta(t - \lambda_j).$$

## Basic techniques and numerical proof of concept

[P. Benner & Khoromskaia & B.Khoromskij & C. Yang, arXiv'18]

► The proposed approach relies on the following techniques:

- using the low-rank BSE matrix structure which allows for each fixed  $t \in [0, a]$  the direct matrix inversion and computation of the respective traces,  $A \mapsto A_0$ ,  $F \mapsto F_0$ .
- the low-rank QTT tensor interpolation of the function  $\phi_\eta(t)$  sampled on large uniform grid  $\{t_1, \dots, t_M\}$  in the whole spectral interval  $t \in [0, a]$ .



DoS for H<sub>2</sub>O: exact TDA vs simplified TDA (left), the zoom in optical interval of spectrum.

[P. Benner & Khoromskaia & B.Khoromskij & C. Yang, arXiv'18]

$$A = E_0 + PQ^T, \quad \text{with } P, Q \in \mathbb{R}^{n \times R}, \quad E_0 = \text{blockdiag}\{B_0, D_0\}. \quad (5)$$

$R$  is a small rank parameter, block size is small.

Let  $S(t)$  denote the diagonal shift of  $A$  depending on the parameter  $t$ ,

$$S(t) = tI - E_0 - PQ^T - i\eta I =: E(t) - PQ^T. \quad (6)$$

The block-diagonal part  $E_0$  is modified by the diagonal shift,

$$E(t) = -E_0 + tI - i\eta I \equiv \text{blockdiag}\{B(t), D(t)\}$$

corresponding to the complex case

$$B(t) = tI_B - i\eta I_B - B_0, \quad D(t) = tI_D - i\eta I_D - D_0. \quad (7)$$

How to compute the trace of the structured matrix resolvent in  $O(n)$  op.

$$\text{trace}[S(t)^{-1}].$$

The cost of trace calculations is estimated to be  $O(nR^2)$ .

### Theorem

Let the matrix family  $S(t)$ ,  $t \in [0, a]$ , be given by (6), with

$$E(t) = \text{blockdiag}\{B(t), D(t)\},$$

where  $B(t), D(t)$  are defined in (7). Then the trace of the matrix inverse  $S(t)^{-1}$  can be calculated explicitly by

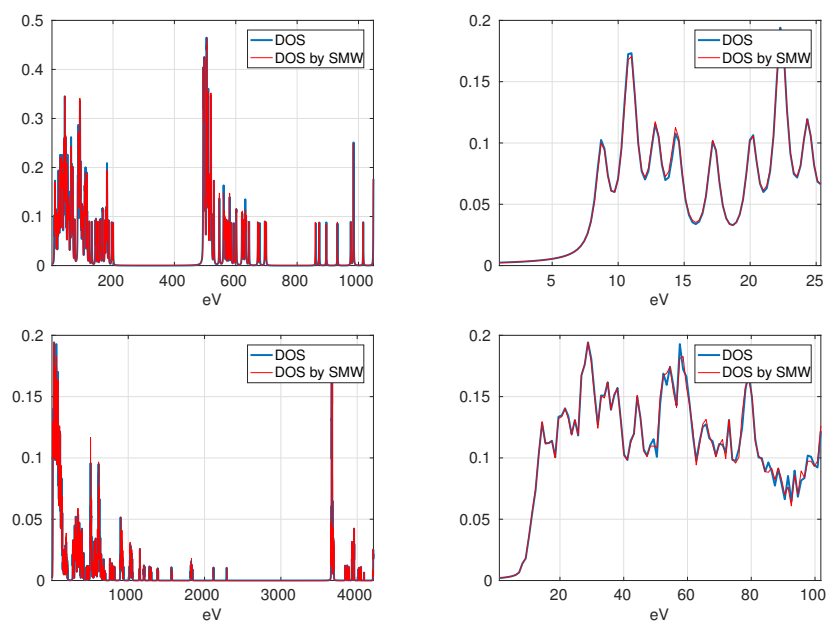
$$\text{trace}[S(t)^{-1}] = \text{trace}[B(t)^{-1}] + \text{trace}[D(t)^{-1}] - \mathbf{1}_n^T (U(t) \odot V(t)) \mathbf{1}_R,$$

where  $U(t) = E(t)^{-1}PK(t)^{-1} \in \mathbb{R}^{n \times R}$ ,  $V(t) = E(t)^{-1}Q \in \mathbb{R}^{n \times R}$ , and

$$K(t) = I_R + Q^T E(t)^{-1}(t)P$$

is a small  $R \times R$  matrix. For fixed  $t \in [0, a]$ , assume that  $n_B = O(n^\alpha)$  with  $\alpha \leq 1/3$ , then the numerical cost is estimated by  $O(nR^2)$ .

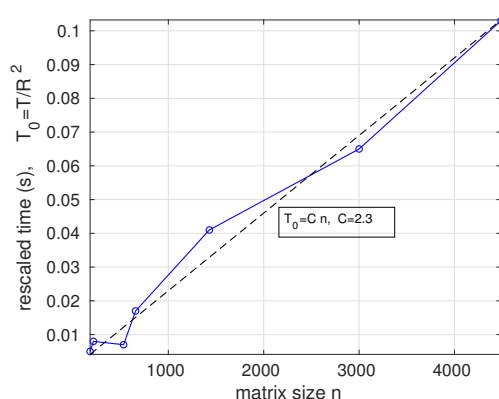
[P. Benner & Khoromskaia & B.Khoromskij & C. Yang, arXiv'18]



Left: DOS for H<sub>2</sub>O (top) and Ethanol (bottom), using full set of TDA eigenvalues (blue) and DOS by rank-structured approximations (red). Right: zoom in the optical part of spectrum.

## DOS-Algorithm in real arithmetics: linear scaling in $n$

[P. Benner & Khoromskaia & B.Khoromskij & C. Yang, arXiv'18]



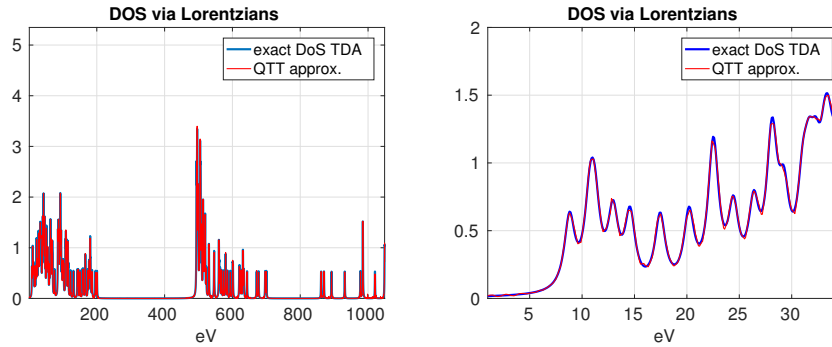
Real arithmetics: CPU time (left) and the rescaled time  $T/R^2$  vs.  $n$ .

Molecule	H <sub>2</sub> O	NH <sub>3</sub>	H <sub>2</sub> O <sub>2</sub>	N <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>5</sub> OH	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>
$N_{ov}$	180	215	531	657	1430	3000	4488
ranks	36	30	68	54	74	129	147
Time (s)	4	5	23	34	151	812	1782
Scaled time	0.1	0.16	0.33	0.62	2.1	4.8	12.1

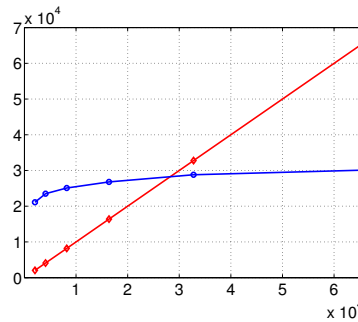
Rank and time characteristics of Alg. in real arithmetics.

# QTT interpolation to DOS: log-scaling in the grid-size $N$

[P. Benner & Khoromskaia & B.Khoromskij & C. Yang, arXiv'18]



DOS for  $H_2O$  via Lorentzians (blue), its QTT approximation (red) (left). Zoom on the left most part of the spectrum (right).  $\varepsilon=0.04$ ,  $r_{QTT} = 10.5$ .



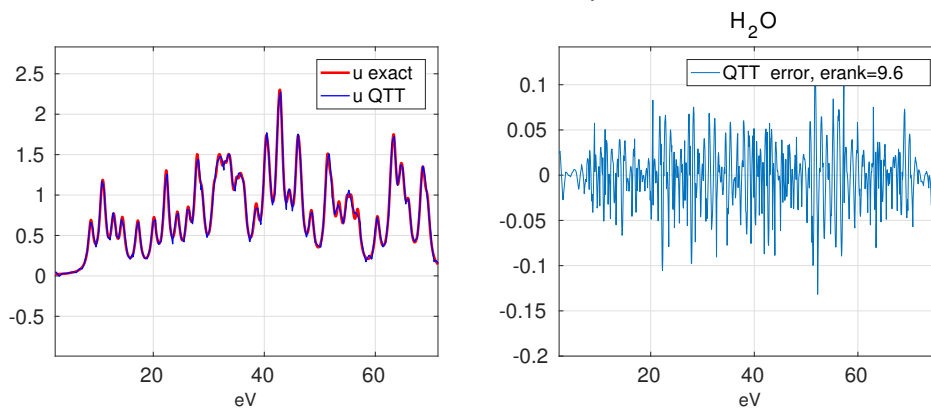
DOS  $H_2O$  via Lorentzians: number of funct. calls for QTT-cross interp. vs. grid size  $N$ .

# QTT interpolation to DOS: log-scaling in the grid-size $N$

► The ACA QTT tensor interpolation reduces the number of functional calls, i.e.,  $M < N$ , if the QTT rank parameters (or threshold  $\varepsilon > 0$ ) are chosen to satisfy

$$M = C_s r_{qtt}^2 \log_2 N \leq N.$$

(An estimate on the number of function evaluations).



QTT ACA interpolation of the DOS for  $H_2O$ : zoom into a small spectral interval.

QTT interpolation procedure: by TT-Toolbox (Oseledets et al.)

[P. Benner &amp; Khoromskaia &amp; B.Khoromskij &amp; C. Yang, arXiv'18]

## Theorem

Assume that the effective support of the shifted Gaussians  $g_\eta(t - \lambda_j)$ ,  $j = 1, \dots, n$ , is included in  $[-a, a]$ . Then the QTT  $\varepsilon$ -rank of the vector  $\mathbf{p}_\eta$  is bounded by

$$\text{rank}_{\text{QTT}}(\mathbf{p}_\eta) \leq C a \log^{3/2}(|\log \varepsilon|),$$

$C = O(|\log \eta|) > 0$  depends only logarithmically on the regularization parameter  $\eta$ .

Molecule	H <sub>2</sub> O	NH <sub>3</sub>	H <sub>2</sub> O <sub>2</sub>	N <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>5</sub> OH	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub>	C <sub>3</sub> H <sub>7</sub> NO <sub>2</sub>
$n = N_{ov}$	180	215	531	657	1430	3000	4488
QTT ranks	11	11	12	11	15	16	13

QTT ranks of Lorentzians-DOS for some molecules;  $\varepsilon = 0.04$ ,  $\eta = 0.4$ ,  $N = 16384$ .

► The QTT tensor rank remains almost independent of the molecular size! The weak dependence of the rank parameter on the molecular geometry.

## Summary: Rank-structured DOS calculations for optical spectra of molecules

Tensor-structured approach presents:

- Ab-initio grid-based tensor-structured HF solver.
- The two-electron integrals tensor (TEI) by Cholesky factorization.
- Excitation energies via BSE eig-problem with linear cost in  $n$ .
- Log-scaling BSE solver by QTT tensor approximation depending only on the number of molecular orbitals, as  $O(N_o^2)$ .
- DOS for optical spectra of molecules (TDA): linear cost in  $n$  per interpolation point (not  $n^3!$ ).
- Toward log-scaling in  $n$  by QTT-ACA interpolation.

<http://personal-homepages.mis.mpg.de/vekh>

<http://personal-homepages.mis.mpg.de/bokh>

Venera Khoromskaia, Boris N. Khoromskij.  
Tensor Numerical Methods in Quantum Chemistry  
De Gruyter, Berlin, 2018.

Thank you for attention !