To exact exchange and beyond!
Building a linear-scaling two-electron integral engine
for quantum chemistry in a generalized Wannier basis

James C. Womack,¹ Jacek Dziedzic,¹,² Chris-Kriton Skylaris¹

¹ University of Southampton, UK
² Gdańsk University of Technology, Poland

7th EuCheMS Chemistry Congress, August 2018
Preamble

Slides will soon be available at jcwomack.com

References — ask me, or see slides online
Electron repulsion integrals (ERIs) in quantum chemistry

- **Methods**
- **Computation**
  - Post-Hartree-Fock
  - SCF
  - $O(N^4)$ integrals
  - Expensive!

**Electron Repulsion Integrals**

$$(\varphi_\alpha \varphi_\delta | \varphi_\beta \varphi_\gamma) = \int dr dr' \varphi_\alpha^*(r) \varphi_\delta^*(r) \frac{1}{|r - r'|} \varphi_\beta(r') \varphi_\gamma(r')$$

**Many-body effects**

- Exchange
- Dispersion
- Multireference systems
Quantum chemistry with linear-scaling cost

Conventional methods are impractical for large systems

- Cost grows rapidly with system size, $N$
- Even if cost of ERIs is avoided/reduced
e.g. Kohn-Sham DFT is at least $O(N^3)$

But, $O(N)$ scaling is possible.

- One-particle density matrix is “nearsighted” [1]
e.g. Exponential decay for insulators [2]

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_i^N f_i \psi_i(\mathbf{r}) \psi_i^*(\mathbf{r}')$$

$$\rho(\mathbf{r}, \mathbf{r}') \sim \exp(-\gamma |\mathbf{r} - \mathbf{r}'|)$$

We can exploit the locality of $\rho(\mathbf{r}, \mathbf{r}')$ to achieve $O(N)$ scaling!
Practical $O(N)$ quantum chemistry with ONETEP

A framework for DFT built around the nearsightedness principle [3]:

- Strictly localized orbitals (NGWFs)
- $O(N)$ cost via truncation of $K$
- $E_{\text{total}}$ minimized wrt $K$ and $\{\varphi_\alpha\}$

$$\rho(r, r') = \varphi_\alpha(r) K^{\alpha\beta} \varphi^*_\beta(r')$$

$$R_{\alpha\beta} > r_{\text{cut}} \implies K^{\alpha\beta} = 0$$

Large-basis set accuracy is possible with a minimal set of NGWFs
Demonstration of $O(N)$ scaling in ONETEP

Large-scale DFT calculations are possible with $N \sim 10000$
Linear-scaling exact exchange (LS-EXX)

EXX is challenging for $O(N)$ DFT because ERIs are non-local

► (Semi-)local XC functionals do not require ERIs
► BUT the most accurate XC functionals include EXX [4]
► EXX corrects self-interaction error in Coulomb energy

\[
\langle \varphi_\alpha \varphi_\delta \varphi_\beta \varphi_\gamma \rangle
\]

\[
\int \, d\mathbf{r} d\mathbf{r}' \, \varphi_\alpha^*(\mathbf{r}) \varphi_\delta(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \varphi_\beta^*(\mathbf{r}') \varphi_\gamma(\mathbf{r}')
\]
Linear-scaling exact exchange (LS-EXX) with NGWFs

The strictly localized NGWF basis presents unique challenges

\[ E_{\text{EXX}} = -K^{\beta \alpha} (\varphi_{\alpha} \varphi_{\delta} | \varphi_{\beta} \varphi_{\gamma}) K^{\delta \gamma} \]
\[ = -K^{\beta \alpha} X_{\alpha \beta} \]

- NGWFs produce sparse matrices & simplified integrals

\[ \int \left[ A^{\alpha} \cdot D_{\delta} \right] \, dr \neq 0 \]
\[ \int \left[ B^{\beta} \cdot C_{\gamma} \right] \, dr = 0 \]

- BUT cannot use conventional ERI evaluation methods
- Straightforward EXX evaluation (via FFTs) is impractical [5, 6]

An alternative scheme is necessary!
Linear-scaling exact exchange (LS-EXX) with NGWFs

Defining features of the scheme [5]

\[ X_{\alpha\beta} = (\varphi_{\alpha\varphi_{\delta}}|f_p) V^{pq} (f_q|\varphi_{\beta\varphi_{\gamma}}) K^{\delta\gamma} \]

- Spherical wave (SW) [7]
- Resolution-of-the-identity (RI)
- Distance-based exchange cutoff

Expand NGWF products in SWs

- Strictly localized in a sphere
- Analytic Coulomb potentials
- “2-centre” fitting scheme

\[ f_p(r) = \begin{cases} 
    j_{l_p}(q_p r) Z_{l_p m_p}(\hat{r}) & r < a \\
    0 & r \geq a 
\end{cases} \]
Demonstration of $O(N)$ scaling EXX

Polyethylene chains
$H-(\text{CH}_2)_n-H$

Cost in core-hours
$N_{\text{CPU}} \times \text{walltime}$

Data from 2013 [5]
Some improvements since

Linear-scaling is achieved, but with a large prefactor. . .
A more recent calculation...

Single-point DFT energy with and without EXX

57 atom monomer from organic photovoltaic polymer

BLYP (GGA) 0.3 h
B3LYP (hybrid) 62 h
5 nodes / 120 cores
“Thomas” MMM Hub
Building a linear-scaling ERI engine

LS-ERI evaluation at large $N$ would have immediate benefits
  ▶ Large scale hybrid (TD)DFT calculations, useful for e.g
    ▶ Computing optical spectra and solvatochromism [8]
    ▶ Modelling charge transfer excitations [9, 10]

A LS-ERI engine would support new $O(N)$ approaches
  ▶ Post-Hartree-Fock correlated methods
  ▶ Multireference methods

First, we must overcome two major obstacles...
  1. Parallelization must scale to 1000s of cores
  2. Very costly Coulomb metric integrals over SWs
The Coulomb metric is a costly bottleneck

\[ V_{A p, B q} = \int \text{d}r f_p(r_A) g_q(r_B) \]

\[ g_q(r) = \int \text{d}r' \frac{f_q(r')}{|r - r'|} \]

- Integrand is highly oscillatory
- Closed-form solutions only for same-centre case \((A = B)\)
- Must resort to numerical integration for “off-site” case \(A \neq B\)
The Coulomb metric is a costly bottleneck

Original implementation (3Dc) [5]
- Polynomial expansion in 3-D sphere
- Integrate expansion analytically

Very computationally intensive
- Memory scales as $O(N_{\text{node}}^3)$
- $N_{\text{node}} = 12^2 \implies \sim 3 \times 10^6$ coeffs
  i.e. 22 MiB per SW expansion
- $\times 100$s of SWs per atomic centre!

$$V_{Ap,Bq} = \int_V \ dx dy dz \ f_p(x_A, y_A, z_A) g_q(x_B, y_B, z_B)$$
A new scheme for fast Coulomb metric evaluation

2-D numerical, 1-D analytic (2Dn-1Da)

- Spherical polar coordinates
- Align z-axis with $\mathbf{R}_{AB}$
- Rotation matrices needed [11, 12]

Significant reduction in cost

- Numerical integral on 2-D half disc
- Memory scales as $O(N_{\text{node}}^2)$.
- $N_{\text{node}} = 12^2 \implies \sim 2 \times 10^4$ coeffs
  i.e. 160 KiB per SW expansion

\[
\int_{0}^{\pi} d\theta \int_{0}^{a} dr \, r^2 \sin(\theta)f_{1,p}(r, \theta)g_{1,q}(r_B, \theta_B) \left\{ \int_{0}^{2\pi} d\phi \, [f_{2,p}g_{2,q}](\phi) \right\}
\]
Coulomb metric evaluation: Some promising results

Full $V$ evaluation for a $H_2$ molecule with typical size SW basis. . .

(I cheated a bit! No rotations are needed in this case: $H_2$ is aligned along Cartesian $z$ axis.)

- Max. abs. diff. between schemes is $\sim 10^{-5}$

<table>
<thead>
<tr>
<th></th>
<th>Min time over 3 repetitions / s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>PBE (no $V$-matrix)</td>
<td>0.85</td>
</tr>
<tr>
<td>PBE0 (new scheme)</td>
<td>8.24</td>
</tr>
<tr>
<td>PBE0 (old scheme)</td>
<td>141.07</td>
</tr>
</tbody>
</table>

Total: Time for fixed NGWF calculation ($K$ optimization)

Off-site blocks: Time spent evaluating $V$ blocks with $A \neq B$

Significant speed-up while maintaining numerical agreement!
But isn’t $H_2$ a bit small?

Some $V$ evaluation results for larger systems...  

$H_{10}$ linear chain

- Aligned along $z$ axis
- Max. abs. diff. $\sim 10^{-5}$

<table>
<thead>
<tr>
<th></th>
<th>V-matrix / s</th>
</tr>
</thead>
<tbody>
<tr>
<td>New scheme</td>
<td>9.70</td>
</tr>
<tr>
<td>Old scheme</td>
<td>4154.71</td>
</tr>
</tbody>
</table>

57 atom monomer

Old: 9 h  
New*: 140 s 

5 nodes / 120 cores

*Not including rotations
To exact exchange and beyond!

1. We are building a LS-ERI engine to enable fast non-local XC models within ONETEP

2. ONETEP’s existing LS-EXX method provides a starting point, but further work is necessary

3. Our new scheme for Coulomb metric evaluation should resolve a major bottleneck
Acknowledgements

Funding and compute resources

▶ EPSRC-funded CCP9 flagship ([EP/P02209X/1](#))
▶ UK MMM Hub “Thomas” ([EP/P020194/1](#))
▶ IRIDIS HPC Facility (Southampton)

CCP9 flagship collaborators

▶ Nick Hine (Warwick, PI)
▶ Peter Haynes (Imperial, Co-I)
▶ Arash Mostofi (Imperial, Co-I)
▶ Chris-Kriton Skylaris (Southampton, Co-I)
▶ Jacek Dziedzic (Southampton/Gdańsk, RCo-I)
▶ Jolyon Aarons (Warwick, PDRA)
▶ Joseph Prentice (Imperial, PDRA)
Numerical comparison of Coulomb metric schemes

Full \( V \) evaluation for a \( \text{H}_2 \) molecule with typical size \( \text{SW} \) basis. . .

(I cheated a bit! No rotations are needed in this case: \( \text{H}_2 \) is aligned along Cartesian \( z \) axis.)

<table>
<thead>
<tr>
<th>Scheme Comparison</th>
<th>Norm of diff.</th>
<th>Max abs. diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3Dc (OT) vs 2Dn-1Da (OT)</td>
<td>( 1.93 \times 10^{-04} )</td>
<td>( 4.12 \times 10^{-05} )</td>
</tr>
<tr>
<td>3Dc (OT) vs 2Dn-1Da (SM)</td>
<td>( 1.93 \times 10^{-04} )</td>
<td>( 4.12 \times 10^{-05} )</td>
</tr>
<tr>
<td>2Dn-1Da (OT) vs 2Dn-1Da (SM)</td>
<td>( 3.31 \times 10^{-06} )</td>
<td>( 5.27 \times 10^{-07} )</td>
</tr>
</tbody>
</table>

OT: ONETEP implementation
SM: Sage Math prototype
3Dc: old scheme
2Dn-1Da: new scheme

“\[O\]btaining the elements of \( V \) to an accuracy of the sixth decimal [is] sufficient for stable calculations”

Bibliography I


Bibliography II
