QMC USING VERY LARGE CONFIGURATION INTERACTION-TYPE EXPANSIONS

arXiv:1510.00730v2 [physics.chem-ph]

A. Scemama¹, T. Applencourt, E. Giner, M. Caffarel
18/12/2015

¹Lab. Chimie et Physique Quantiques, IRSAMC, UPS/CNRS, Toulouse, France
QMC=Chem:

- 2013: Oxygen atom, 100 000 determinants
- 2014: F$_2$ dissociation curve, 10 000 determinants$^1$
- 2014: Transition metals, 10 000 – 48 000 determinants$^2$

Today

- Water: 84 000 determinants [-76.4354(7)]
- G2 set: $\sim$ 200 000 determinants (w/wo pseudo-potentials)
- Cl atom: 750 000 determinants

• Exact solution in a given basis set

• Basis: All possible Slater determinants \((\mathcal{D}_k)\) of spin-orbitals

\[
\psi\text{_{FCI}} = \sum_k c_k \mathcal{D}_k
\]  

(1)

• Spin-Free formalism:

\[
\mathcal{D}_k = D^\uparrow_k \times D^\downarrow_k
\]  

(2)

• Size of the FCI space:

\[
\binom{N\text{_{MO}}}{N\text{_{elec}^{\uparrow}}} \times \binom{N\text{_{MO}}}{N\text{_{elec}^{\downarrow}}}
\]  

(3)
Cl in cc-pVDZ

- 9 ↑ and 8 ↓ electrons in 19 MOs
- 92378 ↑ determinants and 75582 ↓ determinants
- 6.98 \times 10^9 \text{ basis functions}
- The size of the space is huge, but the space is empty: \sim 10^6 \text{ coefficients } |c_k| > 10^{-12}

\implies \text{CIPSI algorithm to select Slater determinants on the fly.}^3

^3http://github.com/LCPQ/quantum_package
ψ can be expressed in a bi-linear form:

\[
\psi(R) = \sum_k c_k D_k = \sum_i \sum_j C_{ij} D_{i\uparrow}(R_{\uparrow}) D_{j\downarrow}(R_{\downarrow})
\] (4)

- \(D_{i\uparrow}(R_{\uparrow})\): vector of \(N_{\text{det}\uparrow}\) elements
- \(D_{j\downarrow}(R_{\downarrow})\): vector of \(N_{\text{det}\downarrow}\) elements
- \(C\): \(N_{\text{det}\uparrow} \times N_{\text{det}\downarrow}\) matrix. The matrix contains \(N_{\text{det}}\) elements.

\(C\) is constant in a QMC calculation \(\implies\) preprocessing.
1. Encoding of determinants$^4$:

HF : \((000011111111,000111111111) = (255,511)\)

Single $\uparrow$ : \((010001111111,000111111111) = (1151,511)\)

Double $\uparrow\downarrow$ : \((010001111111,010101111111) = (1151,1407)\)

\(^4\text{arXiv:1311.6244 [physics.comp-ph]}\)
1. Encoding of determinants:\(^4\):
   - HF: \((000011111111,000111111111) = (255,511)\)
   - Single \(\uparrow\): \((010001111111,000111111111) = (1151,511)\)
   - Double \(\uparrow\downarrow\): \((010001111111,010101111111) = (1151,1407)\)

2. Separate \(D\uparrow\) and \(D\downarrow\) of \(D_k\): two lists of \(N_{\text{det}}\):
   - \(000011111111 = 255\)
   - \(010001111111 = 1151\)
   - \(010001111111 = 1151\)

1. Encoding of determinants⁴:
   - HF : \((000011111111,000111111111) = (255,511)\)
   - Single \(\uparrow\) : \((010001111111,000111111111) = (1151,511)\)
   - Double \(\uparrow\downarrow\) : \((010001111111,010101111111) = (1151,1407)\)

2. Separate \(D_\uparrow\) and \(D_\downarrow\) of \(\mathcal{D}_k\): two lists of \(N_{\det}\)
   - \(000011111111 = 255\)
   - \(010001111111 = 1151\)

3. Sort both lists, remove consecutive repetitions:
   - One list of \(N_{\det\uparrow}\), one list of \(N_{\det\downarrow}\)

1. Encoding of determinants⁴:
   HF : \((000011111111,000111111111) = (255,511)\)
   Single ↑ : \((010001111111,000111111111) = (1151,511)\)
   Double ↑↓ : \((010001111111,010101111111) = (1151,1407)\)

2. Separate \(D_↑\) and \(D_↓\) of \(\mathcal{D}_k\): two lists of \(N_{\text{det}}\)
   \[
   \begin{align*}
   000011111111 & = 255 & 000111111111 & = 511 \\
   010001111111 & = 1151 & 000111111111 & = 511 \\
   010001111111 & = 1151 & 010101111111 & = 1407
   \end{align*}
   \]

3. Sort both lists, remove consecutive repetitions: One list of \(N_{\text{det}_↑}\),
   one list of \(N_{\text{det}_↓}\)
   \[
   \begin{align*}
   000011111111 & = 255 & 000111111111 & = 511 \\
   010001111111 & = 1151 & 010101111111 & = 1407
   \end{align*}
   \]

4. For each \(\mathcal{D}_k\), find the indices \(i \) and \(j \) of \(D_↑\) and \(D_↓\) (binary search
   \(\mathcal{O}(\log N_{\text{det}_↑})\), and set \(C_{ij} = c_k\)

   Takes 3 seconds on a single CPU core for \(10^6 \mathcal{D}_k\), \(\mathcal{O}(N_{\text{det}})\).

**Remark on Ordering**

$D_{i-1}$ and $D_i$ are likely to differ by a single substitution:

<table>
<thead>
<tr>
<th>List index</th>
<th>Decimal</th>
<th>Binary</th>
<th>Determinant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>00001111</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>00010111</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>00011011</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>00011101</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>00011110</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>39</td>
<td>00100111</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td>00101011</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>00101101</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
In QMC=Chem: \textit{We do all-electron moves.}

At every MC step, we need to evaluate:

\[ \Psi = (D_{\uparrow}^\dagger (C)D_{\downarrow}) \]  

(5)

\[ \nabla_i \Psi = \nabla_i D_{\uparrow}^\dagger . (C)D_{\downarrow} \text{ or } (D_{\uparrow}^\dagger C) . \nabla_i D_{\downarrow} \]  

(6)

\[ \Delta_i \Psi = \Delta_i D_{\uparrow}^\dagger . (C)D_{\downarrow} \text{ or } (D_{\uparrow}^\dagger C) . \Delta_i D_{\downarrow} \]  

(7)

\[ V_{\text{non-loc pseudo}} \Psi = V_{\text{non-loc pseudo}} D_{\uparrow}^\dagger . (C)D_{\downarrow} \text{ or } (D_{\uparrow}^\dagger C) . V_{\text{non-loc pseudo}} D_{\downarrow} \]  

(8)

(\uparrow \text{ electrons and } \downarrow \text{ electrons})
Vector-matrix product $D^{†}_\uparrow . C : N_{\text{det}}$ operations, returns a $N_{\text{det}}$ vector

1. Dot product with $D_{\downarrow} : N_{\text{det}}$ operations, produces a scalar
2. Matrix product with $\nabla D_{\downarrow} : 3N_{\text{elec}} \times N_{\text{det}}$ operations, produces a $3N_{\text{elec}}$ vector
3. Matrix product with $\Delta D_{\downarrow} : N_{\text{elec}} \times N_{\text{det}}$ operations, produces a $N_{\text{elec}}$ vector
4. Matrix product with $V_{\text{pseudo}}^{\text{non-loc}} D_{\downarrow} : N_{\text{elec}} \times N_{\text{det}}$ operations, produces a $N_{\text{elec}}$ vector
If one localizes the pseudo-potential on the CI wave function:

\[ \hat{T} + \hat{V} + \hat{V}_{\text{pseudo loc}} + \sum_{lm} V_{\text{pseudo non-loc, } l} \frac{|Y_{lm}\rangle\langle Y_{lm}| \Psi_{\text{CI}}}{\Psi_{\text{CI}}} \]  

(9)

\[ V_{\text{pseudo, } i}^{\text{non-loc}} = \sum_j S_{ij}^{-1} \sum_{lm} Y_{lm}(r_i) \langle Y_{lm}|\phi_j\rangle \]  

(10)

- The pseudo-potential integrals can be computed analytically
- \( \langle Y_{lm}|\phi_j\rangle \) is pre-computed (once) on a grid
- No quadrature is needed: \( \times 16, \times 27 \) speed-up
- Tested on G2 set: Same MAD with/without pseudo-potentials
- Cost: Same as Laplacian
- Also applicable to single-determinant
DETERMINANTS AND THEIR DERIVATIVES

- First determinant: Standard technique as in single-determinant
- All other determinants: Sherman-Morrison updates

Determinant $D_i$ is updated wrt $D_{i-1}$ by successive column updates. Depends on the ordering.

Cl atom, cc-pVDZ, $10^6$ Slater determinants:
- Average excitation degree wrt HF: 3.46
- Average number of column substitutions: 1.78

<table>
<thead>
<tr>
<th>Excitation</th>
<th>$D_{\uparrow \downarrow}$</th>
<th>$D_{\uparrow}$</th>
<th>$D_{\downarrow}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>38</td>
<td>90</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>2 177</td>
<td>1603</td>
<td>1520</td>
</tr>
<tr>
<td>3</td>
<td>43 729</td>
<td>7811</td>
<td>6507</td>
</tr>
<tr>
<td>4</td>
<td>308 045</td>
<td>8581</td>
<td>5071</td>
</tr>
<tr>
<td>5</td>
<td>351 182</td>
<td>2090</td>
<td>579</td>
</tr>
<tr>
<td>6</td>
<td>291 481</td>
<td>77</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3 067</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>280</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 1: Cl : Scaling is $\sim \sqrt{N_{\text{det}}}$. SM updates $\times 7$. 
approximation to the FCI solution

- CAS: good for re-use of $D_{\uparrow}$ and $D_{\downarrow}$
- CISD, CISDT, ...: Poor re-use

Diagram:

```
    N_{det\downarrow}   
      \   \           
    HF S D           
      \   \           
    HF S             
      \   \           
    HF              
      \   \           
    C
```
Usual truncation

- Based on the $|c_k|$ (or CSF coefficients)
- Removes mostly same-spin doubles in CISD $\rightarrow$ some gain is observed.
- Makes the matrix $C$ more sparse
- For CASSCF or FCI: Needs a severe truncation to remove $D_\uparrow$ or $D_\downarrow$: the whole column must be zero
Alternate truncation scheme

• Contribution of $D_\uparrow$ and $D_\downarrow$ to the norm of $\Psi$

\[
\mathcal{N} = \sum_{i=1}^{N_{\text{det}_\uparrow}} \sum_{j=1}^{N_{\text{det}_\downarrow}} C_{ij}^2 = \sum_{i=1}^{N_{\text{det}_\uparrow}} \mathcal{N}_i^\uparrow = \sum_{j=1}^{N_{\text{det}_\downarrow}} \mathcal{N}_j^\downarrow \quad (11)
\]

• Remove $D_{i\uparrow}$ when $\mathcal{N}_i^\uparrow = \sum_{j=1}^{N_{\text{det}_\downarrow}} C_{ij}^2 < \epsilon$
CONVERGENCE OF THE FN-DMC ENERGY

Cl FCI/cc-pVTZ
Converged FCI/cc-pVTZ FN-DMC energy with $\epsilon = 10^{-5}$: $7.1\times$ the cost of a single-determinant
### TOTAL CPU TIME PER MC STEP

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>$N_{\text{det}}$</th>
<th>$N_{\uparrow \text{det}}$</th>
<th>$N_{\downarrow \text{det}}$</th>
<th>CPU time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-2}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.02450</td>
</tr>
<tr>
<td>$5 \times 10^{-3}$</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0.02688</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>86</td>
<td>21</td>
<td>17</td>
<td>0.03899</td>
</tr>
<tr>
<td>$5 \times 10^{-4}$</td>
<td>214</td>
<td>29</td>
<td>28</td>
<td>0.04636</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>1361</td>
<td>93</td>
<td>74</td>
<td>0.08808</td>
</tr>
<tr>
<td>$5 \times 10^{-5}$</td>
<td>2424</td>
<td>120</td>
<td>89</td>
<td>0.1054</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>9485</td>
<td>234</td>
<td>166</td>
<td>0.1855</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>54016</td>
<td>772</td>
<td>523</td>
<td>0.5960</td>
</tr>
<tr>
<td>$10^{-7}$</td>
<td>207995</td>
<td>2279</td>
<td>1389</td>
<td>1.740</td>
</tr>
<tr>
<td>$10^{-8}$</td>
<td>459069</td>
<td>5797</td>
<td>3291</td>
<td>4.196</td>
</tr>
<tr>
<td>$10^{-9}$</td>
<td>748835</td>
<td>14456</td>
<td>8054</td>
<td>9.724</td>
</tr>
<tr>
<td>$10^{-10}$</td>
<td>926299</td>
<td>30320</td>
<td>16571</td>
<td>19.18</td>
</tr>
<tr>
<td>0</td>
<td>1000000</td>
<td>52433</td>
<td>26833</td>
<td>31.92</td>
</tr>
</tbody>
</table>

Order of magnitude:
36 ms/step $\equiv 10^{11}$ steps in 10 hours on 100 000 cores
Sildenafil

- 63 atoms
- 126 $\uparrow$ and 126 $\downarrow$ electrons
- 649 basis functions
- MP2 wave function
Sildenafil: 31 ms/step $\Leftrightarrow \epsilon = 2 \times 10^{-5}$ ($N_{\text{det}} = 2049$)

100 dets only 2\times more expensive than 1 det.
MEMORY PER CPU CORE

The graph shows the relationship between the number of determinants and the RAM (MiB) for KNL with 280 threads/16GiB. The RAM usage increases as the number of determinants increases.
• Bi-linear expression: scaling as $\sqrt{N_{\text{det}}}$
• The only $N_{\text{det}}$-scaling step is a matrix-vector product: tiny pre-factor, independent of $N_{\text{elec}}$, negligible
• Sherman-Morrison: $\sim \times N_{\text{elec}}$
• Ordering $\Rightarrow D_i / D_{i-1}$: less work to do than $D_i / D_0$: $\sim \times 2$
• Localizing the pseudo-potential on the determinantal component: $\times 16 - \times 27$
• Memory: scales as $N_{\text{elec}} \times N_{\text{det}}$
• $10^6$ determinants are possible for small systems (pseudo and all-electron)
• $\sim 2000$ determinants are possible for a 250-electron system using a (difficult) CISD basis
Table method: store a 2D array where each element is the result of a dot product

- Dot product: $N_{\text{elec}}$ FMA operations
- Modern CPUs can do 16 FMAs ($a = a + b*c$) per CPU cycle
- Replace FMA by a random memory access: L1 (4 cycles), L2 (12 cycles), L3 (44 cycles), RAM (357 cycles)
- And then compute a small determinant
- L1 $\leftrightarrow$ 64 ↑ electrons
- L2 $\leftrightarrow$ 192 ↑ electrons
- L3 $\leftrightarrow$ 704 ↑ electrons
- RAM $\leftrightarrow$ 5712 ↑ electrons
COMPARISON WITH TABLE METHOD

Measurements on Cl, 1M determinants:

<table>
<thead>
<tr>
<th></th>
<th>$D_{\uparrow}$</th>
<th>$D_{\downarrow}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table method</td>
<td>2110 cycles</td>
<td>2431 cycles</td>
</tr>
<tr>
<td>This work</td>
<td>544 cycles</td>
<td>359 cycles</td>
</tr>
</tbody>
</table>

But:
The table method allows to avoid the storage of all Slater matrices when doing single-electron moves: very good!

We prefer all-electron moves because they use more efficiently the hardware (same as BLAS2 vs BLAS3).