Accelerated MR-PT2 with a Hybrid Stochastic/Deterministic Algorithm


A. Scemama, Y. Garniron, P.-F. Loos, M. Caffarel

22 May 2017

Lab. Chimie et Physique Quantiques, IRSAMC, UPS/CNRS, Toulouse, France
Introduction: CIPSI

- Old method (Bender and Davidson, 1969), lots of development in Toulouse in the past (Malrieu, Evangelisti, Daudey, Spiegelman, etc)
- CIPSI: Configuration Interaction using a Perturbative Selection made Iteratively
- A few years ago, we considered CIPSI wave functions for QMC (Ph.D E. Giner)
- CIPSI appears to be a good candidate for massively parallel wave function calculations (Ph.D Y. Garniron)
- Open-source Code: *Quantum Package*
Preliminaries

- **Red**: Variational wave function
- **Green**: External wave function (perturbation)
- Excitation degree: \( d(I, J) = \left( k : |J\rangle = \hat{T}_k |I\rangle \right) \)
1. Define a **reference** wave function:

\[ |\Psi\rangle = \sum_{l \in \mathcal{D}} c_l |l\rangle \]

2. Generate external determinants:

\[ A = \{ (\forall I \in \mathcal{D}) (\forall \hat{T} \in T_1 \cup T_2) : |\alpha\rangle = \hat{T} |I\rangle \} \]

3. Second order perturbative contribution of each \(|\alpha\rangle\):

\[ \delta E(\alpha) = \langle \Psi |H| \alpha \rangle \langle \alpha |H| \Psi \rangle - \langle \alpha |H| \alpha \rangle \]

4. Select the \(|\alpha\rangle\) with the largest \(\delta E(\alpha)\) and add them into \(\mathcal{D}\)

5. Diagonalize \(H\) in \(\mathcal{D} = \Rightarrow \) update \(|\Psi\rangle\) and \(E_{\text{var}}\)

6. Iterate 3
CIPSI algorithm

1. Define a reference wave function:

$$|\Psi\rangle = \sum_{I \in D} c_I |I\rangle$$

$$E_{\text{var}} = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

2. Generate external determinants:

$$\mathcal{A} = \left\{ (\forall I \in D) \left( \forall \hat{T} \in \mathcal{T}_1 \cup \mathcal{T}_2 \right) : |\alpha\rangle = \hat{T}|I\rangle \right\}$$
CIPSI algorithm

1. Define a \textit{reference} wave function: 

\[ |\psi\rangle = \sum_{l \in \mathcal{D}} c_l |l\rangle \quad \text{with} \quad E_{\text{var}} = \frac{\langle \psi | \mathcal{H} | \psi \rangle}{\langle \psi | \psi \rangle} \]

2. Generate \textit{external determinants}:

\[ \mathcal{A} = \left\{ (\forall l \in \mathcal{D}) \left( \forall \hat{T} \in \mathcal{T}_1 \cup \mathcal{T}_2 \right) : |\alpha\rangle = \hat{T} |l\rangle \right\} \]

3. Second order perturbative contribution of each \( |\alpha\rangle \):

\[ \delta E(\alpha) = \frac{\langle \psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \psi \rangle}{E_{\text{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle} \]
CIPSI algorithm

1. Define a reference wave function:

\[ |\psi\rangle = \sum_{l \in D} c_l |l\rangle \]

\[ E_{\text{var}} = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \]

2. Generate external determinants:

\[ \mathcal{A} = \{ (\forall l \in D) (\forall \hat{T} \in \mathcal{T}_1 \cup \mathcal{T}_2) : |\alpha\rangle = \hat{T}|l\rangle \} \]

3. Second order perturbative contribution of each \( |\alpha\rangle \):

\[ \delta E(\alpha) = \frac{\langle \psi | H | \alpha \rangle \langle \alpha | H | \psi \rangle}{E_{\text{var}} - \langle \alpha | H | \alpha \rangle} \]

4. Select the \( |\alpha\rangle \) with the largest \( \delta E(\alpha) \) and add them into \( D \)
CIPSI algorithm

1. Define a reference wave function:
\[ |\psi\rangle = \sum_{I \in D} c_I |I\rangle \quad E_{\text{var}} = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \]

2. Generate external determinants:
\[ \mathcal{A} = \left\{ (\forall I \in D) \left( \forall \hat{T} \in \mathcal{T}_1 \cup \mathcal{T}_2 \right) : |\alpha\rangle = \hat{T} |I\rangle \right\} \]

3. Second order perturbative contribution of each \( |\alpha\rangle \) :
\[ \delta E(\alpha) = \frac{\langle \psi | H | \alpha \rangle \langle \alpha | H | \psi \rangle}{E_{\text{var}} - \langle \alpha | H | \alpha \rangle} \]

4. Select the \( |\alpha\rangle \) with the largest \( \delta E(\alpha) \) and add them into \( D \)

5. Diagonalize \( H \) in \( D \) \( \Longrightarrow \) update \( |\psi\rangle \) and \( E_{\text{var}} \)
CIPSI algorithm

1. Define a **reference** wave function:

   \[ |\psi\rangle = \sum_{l \in \mathcal{D}} c_l |l\rangle \quad \text{and} \quad E_{\text{var}} = \frac{\langle \psi | \mathcal{H} | \psi \rangle}{\langle \psi | \psi \rangle} \]

2. Generate external determinants:

   \[ \mathcal{A} = \left\{ (\forall l \in \mathcal{D}) \left( \forall \hat{T} \in \mathcal{T}_1 \cup \mathcal{T}_2 \right) : |\alpha\rangle = \hat{T} |l\rangle \right\} \]

3. Second order perturbative contribution of each \( |\alpha\rangle \):

   \[ \delta E(\alpha) = \frac{\langle \psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \psi \rangle}{E_{\text{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle} \]

4. Select the \( |\alpha\rangle \) with the largest \( \delta E(\alpha) \) and add them into \( \mathcal{D} \)

5. Diagonalize \( \mathcal{H} \) in \( \mathcal{D} \rightleftharpoons \) update \( |\psi\rangle \) and \( E_{\text{var}} \)

6. Iterate
Figure 1: Titanium atom, CI-PSI.
Remarks

- When all $|\alpha\rangle$ are selected the Full-CI is obtained.
- CIPSI is more an algorithm than a method.
- Rules on the generation of $|\alpha\rangle$ defines wave function methods.
- Any WF method can be realized with the CIPSI algorithm.
At any time, $E_{PT2} = \sum_\alpha \delta E(\alpha)$ estimates the distance to the solution.

- The $|\alpha\rangle$ with largest $\delta E(\alpha)$ have been added to $\Psi$ in the past $\implies$ Only small contributions remain
- $A$ increases with $D$
- $\implies$ a very large number of very small contributions

In practice, not all $|\alpha\rangle$ are generated: good for selection but $E_{PT2}$ is biased.

This work:
Fast and unbiased $E_{PT2}$
General Problem of MR-PT2 calculations
Consider a wave function $\Psi$ expanded on an *arbitrary* set $D$ of $N_{\text{det}}$ orthonormal Slater determinants,

$$\Psi = \sum_{I \in D} c_I |I\rangle,$$

$$E_{\text{var}} = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

The Epstein-Nesbet 2nd order correction to the energy is

$$E_{\text{PT2}} = \sum_{\alpha \in A} \frac{\langle \Psi | H | \alpha \rangle \langle \alpha | H | \Psi \rangle}{E_{\text{var}} - \langle \alpha | H | \alpha \rangle}$$

The set $A$ contains the Slater determinants

- which belong to the target space (FCI, CAS, MR-CI, etc)
- which are not in $D$
- for which $d(I, \alpha) = 1$ or $2$ for at least one pair $(I, \alpha)$
Formal Scaling

\[ E_{PT2} = \sum_{\alpha \in A} \frac{\langle \Psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \Psi \rangle}{E_{var} - \langle \alpha | \mathcal{H} | \alpha \rangle} = \sum_{\alpha \in A} \frac{\left( \sum_{I \in D} c_I \langle I | \mathcal{H} | \alpha \rangle \right)^2}{E_{var} - \langle \alpha | \mathcal{H} | \alpha \rangle} \]

- Size of \( A \): size of \( (\hat{T}_1 + \hat{T}_2) | \Psi \rangle \)
- Number of non-zero terms: \( d(I, \alpha) \leq 2 \)
  \[ \sim N_{\text{det}} \times \left[ \left( N_{\text{elec}}^\uparrow \times (N_{\text{MO}} - N_{\text{elec}}) \right)^2 \right] \]
- Expensive
Solutions to make simulations possible

“Non-general” solutions:

- Partition the MO space into different classes (active, virtual, inactive, etc)
- Complete active space
- Use another zeroth-order Hamiltonian (CAS-PT2, NEV-PT2)

Solutions applicable to any wave function

1. Truncation of $D$ to consider only contributions of large $c_i$
   - Truncation $\rightarrow$ bias: $E_{PT2}$ is a sum of negative values.

2. Monte Carlo sampling in $A$. Statistical error decreases as $O\left(1/\sqrt{N_{\text{samples}}}\right)$ $\Rightarrow$ Difficult to get $10^{-5} \text{a.u}$ precision.

3. Parallelism
MR-PT2 Algorithm
1. Choose an arbitrary ordering of $|I\rangle$. Natural choice:

$$w_I = \frac{c_I^2}{\langle \Psi | \Psi \rangle}$$

2. Make disjoint groups $A_I$ of $|\alpha\rangle$ originating from the same generator $|I\rangle$

3. Each $A_I$ has its own contribution $\epsilon_I$ to $E_{PT2}$
Central idea

\[ E_{PT2} = \sum_{\alpha \in A} \frac{(\langle \Psi | H | \alpha \rangle)^2}{E_{var} - \langle \alpha | H | \alpha \rangle} \]

\[ = \sum_{I \in D} \sum_{\alpha_I \in A_I} \frac{(\langle \Psi | H | \alpha_I \rangle)^2}{E_{var} - \langle \alpha_I | H | \alpha_I \rangle} \]

\[ = \sum_{I \in D} \epsilon_I \]
Consequences

\[ \epsilon_I = \sum_{\alpha_I \in A_I} \frac{(\langle \Psi | \mathcal{H} | \alpha_I \rangle)^2}{E_{\text{var}} - \langle \alpha_I | \mathcal{H} | \alpha_I \rangle} \]

1. \( \langle \Psi | \mathcal{H} | \alpha_I \rangle = \sum_{J \geq I} c_J \langle J | \mathcal{H} | \alpha_I \rangle \)

2. \( \langle \alpha_I | \mathcal{H} | \alpha_I \rangle \) is always large (otherwise \( |\alpha_I\rangle \) would be better in the variational space, and PT is questionable)

- \( \forall I \in \mathcal{D} : \epsilon_I \leq 0 \)
- \( |\epsilon_I| \) is expected to decrease as \( c_I^2 \)
- The computational cost decreases with \( I \)
Consequences

Figure 2: \( F_2, \text{cc-pVDZ, } 10^6 \) determinants in the variational space
$N_{\text{det}}$ contributions $\epsilon_i \rightarrow$ can be stored in memory

**Lazy Evaluation** (Wikipedia)

*In programming language theory, lazy evaluation, or call-by-need is an evaluation strategy which delays the evaluation of an expression until its value is needed (non-strict evaluation) and which also avoids repeated evaluations (sharing).*
Monte Carlo with Lazy Evaluation

\[ E_{PT2} = \sum_{I \in D} \frac{\epsilon_I}{p_I} \sum_{I \in D} \frac{\epsilon_I}{p_I} = \left\langle \frac{\epsilon_I}{p_I} \right\rangle_{p_I} \]

- Draw a generator determinant |I⟩ with probability
  \[ p_I = \frac{c_I^2}{\left( \sum_{J \in D} p_J \right)} \]
- Increment \( n_I \), the number of evaluations of \( \epsilon_I \)
- If \( \epsilon_I \) is not already computed, compute it and store its value
- \( E_{PT2} \sim \sum_{I \in D} \frac{n_I}{N_{\text{samples}}} \frac{\epsilon_I}{p_I} \)
- Statistical error : \( O\left(\frac{1}{\sqrt{N_{\text{samples}}}}\right) \)
- Lazy evaluation : Acceleration (wall-clock time)
Figure 3: $F_2$, cc-pVQZ, $5 \times 10^6$ determinants in the variational space
Monte Carlo with Lazy Evaluation

Figure 4: F$_2$, cc-pVQZ, $5 \times 10^6$ determinants in the variational space
Monte Carlo with Lazy Evaluation

Figure 5: $F_2$, cc-pVQZ, $5 \times 10^6$ determinants in the variational space
Variance reduction

Uniform sampling: \( p_I = 1 / N_{\det} \), \( E_{PT2} = \langle N_{\det} \times \epsilon_I \rangle_{p_I} \)
Variance reduction

Improved sampling: \( p_I = c_I^2, \ E_{PT2} = \langle \epsilon_I / p_I \rangle_{p_I} \)
Variance reduction

- Noise can be smoothed out by averaging
- Split $\mathcal{D}$ into $M$ equiprobable sets: Comb

\[
E_{PT2} = \sum_{l \in \mathcal{D}} \epsilon_l = \sum_{k=1}^{M} \sum_{l_k \in \mathcal{D}_k} \epsilon_{l_k}
\]

- New Monte Carlo estimator:

\[
E_{PT2} = \left\langle \frac{1}{M} \sum_{k=1}^{M} \frac{\epsilon_{l_k}}{p_{l_k}} \right\rangle (p_{l_1}, \ldots, p_{l_M})
\]
Variance reduction
Variance reduction
Variance reduction
Variance reduction

With $M = 100$
Variance reduction

![Graph showing variance reduction over time](image)
Variance reduction

![Graph showing variance reduction over time](image)

- **Evar** + **EPT2** (au)
- **Time (seconds)**

Legend:
- **Lazy**
- **Comb**
Variance reduction

Statistical error on E_{PT2}

Time (seconds)
Lazy evaluation
a t^{-1/2}
Comb

Statistical error on E_{PT2}

Time (seconds)
Hybrid deterministic/stochastic scheme

- When all the determinants have been drawn, the exact $E_{PT2}$ can be computed.
- The result with zero statistical error can be reached in a finite time.
- In typical wave functions, 90% of the norm is on a few determinants.
- Compute the few first contributions $\epsilon_I$, and perform the MC in the rest.

$$E_{PT2} = \sum_{I \in D_D} \epsilon_I + \left\langle \frac{1}{M} \sum_{k=1}^{M} \frac{\epsilon_{I_k}}{p_{I_k}} \right\rangle (p_{I \in D_S})$$
Hybrid deterministic/stochastic scheme

Make the deterministic part grow during the calculation.

At each MC step:

- Draw a random number
- Find the determinants selected by the comb (increment $n_I$’s)
- Compute the $\epsilon_I$ which have not been yet computed
- Compute deterministically the first non-computed determinant
- If a tooth of the comb is completely filled $\Rightarrow$ Deterministic

At any time:

$$E_{PT2}(t) = \sum_{I \in D_D(t)} \epsilon_I + \sum_{I \in D_S(t)} \frac{1}{M(t)} \frac{n_I(t)}{N_{\text{samples}}(t)} \frac{\epsilon_I}{p_I}$$
Hybrid deterministic/stochastic scheme

![Graph showing the hybrid deterministic/stochastic scheme](image-url)
Hybrid deterministic/stochastic scheme
Hybrid deterministic/stochastic scheme

The graph shows the evolution of the energy $E_{\text{var}} + E_{\text{Pr2}}$ (in atomic units) over time (in seconds) for different schemes:

- **Lazy**
- **Comb**
- **Hybrid**

The data points and lines represent the energy values at various time intervals. The y-axis is labeled $E_{\text{var}} + E_{\text{Pr2}}$ (au), and the x-axis is labeled Time (seconds). The graph covers a range of time from 200 to 1400 seconds.
Hybrid deterministic/stochastic scheme

![Graph showing statistical error on EPT2 vs time (seconds) with different algorithms: Lazy evaluation, $a t^{-1/2}$, Comb, and Hybrid.]
A few results
- 16 correlated electrons
- RHF MOs
- cc-pVDZ: 28 MOs
- cc-pVTZ: 60 MOs
- cc-pVQZ: 110 MOs
- Empirical rule: $|E_{\text{FCI}} - E_{\text{PT2}}| < |E_{\text{PT2}} - E_{\text{var}}|/5$

<table>
<thead>
<tr>
<th></th>
<th>$N_{\text{det}}$</th>
<th>$E_{\text{var}}$</th>
<th>$E_{\text{var}} + E_{\text{PT2}}$</th>
<th>$E_{\text{PT2}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>$2 \times 10^6$</td>
<td>$-199.098,015$</td>
<td>$-199.099,412$</td>
<td>$-0.001,397$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$-199.099,41(9)$</td>
<td></td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>$2 \times 10^6$</td>
<td>$-199.286,288$</td>
<td>$-199.298,119(1)$</td>
<td>$-0.011,831(1)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$-199.297,7(1)$</td>
<td></td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>$1 \times 10^7$</td>
<td>$-199.349,290$</td>
<td>$-199.361,355(1)$</td>
<td>$-0.012,065(1)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$-199.359,8(2)$</td>
<td></td>
</tr>
</tbody>
</table>

In gray: i-FCI-QMC results of Cleland et al, JCTC 2012
Cr$_2$

- 38 correlated electrons (10 frozen MOs)
- cc-pVDZ : 86 MOs
- cc-pVTZ : 136 MOs
- cc-pVQZ : 186 MOs
- CAS-SCF (12,12) MOs
- Selected Full-CI : $2 \times 10^7$ determinants
Cr$_2$, 2 $10^7$ determinants, 800 CPU cores

<table>
<thead>
<tr>
<th>Basis</th>
<th>$E_{PT2}$</th>
<th>Wall-clock time</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>$-0.0683(1)$</td>
<td>14 min</td>
</tr>
<tr>
<td></td>
<td>$-0.06836(1)$</td>
<td>55 min</td>
</tr>
<tr>
<td></td>
<td>$-0.068361(1)$</td>
<td>2.4 hr</td>
</tr>
<tr>
<td></td>
<td>$-0.068360604$</td>
<td>3 hr</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>$-0.1244(5)$</td>
<td>19 min</td>
</tr>
<tr>
<td></td>
<td>$-0.1247(1)$</td>
<td>58 min</td>
</tr>
<tr>
<td></td>
<td>$-0.12463(1)$</td>
<td>3.5 hr</td>
</tr>
<tr>
<td></td>
<td>$-0.124642(1)$</td>
<td>8.7 hr</td>
</tr>
<tr>
<td></td>
<td>$\sim 15$ hr (estimated)</td>
<td></td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>$-0.1558(5)$</td>
<td>56 min</td>
</tr>
<tr>
<td></td>
<td>$-0.1559(1)$</td>
<td>2.5 hr</td>
</tr>
<tr>
<td></td>
<td>$-0.15595(1)$</td>
<td>9.0 hr</td>
</tr>
<tr>
<td></td>
<td>$-0.155952(1)$</td>
<td>18.5 hr</td>
</tr>
<tr>
<td></td>
<td>$\sim 29$ hr (estimated)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$N_{\text{det}}$</td>
<td>$E_{\text{var}}$</td>
</tr>
<tr>
<td>----------------</td>
<td>------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>cc-pVDZ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAS</td>
<td>$3 \times 10^5$</td>
<td>$-2086.650,896$</td>
</tr>
<tr>
<td>MR-CISD</td>
<td>$2 \times 10^6$</td>
<td>$-2087.156,659$</td>
</tr>
<tr>
<td>CIPSI</td>
<td>$6 \times 10^7$</td>
<td>$-2087.244,972$</td>
</tr>
<tr>
<td><strong>cc-pVTZ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAS(12,12)</td>
<td>$5 \times 10^5$</td>
<td>$-2086.655,594$</td>
</tr>
<tr>
<td>CIPSI</td>
<td>$2 \times 10^7$</td>
<td>$-2087.449,781$</td>
</tr>
<tr>
<td><strong>cc-pVQZ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAS(12,12)</td>
<td>$6 \times 10^5$</td>
<td>$-2086.643,210$</td>
</tr>
<tr>
<td>CIPSI</td>
<td>$2 \times 10^7$</td>
<td>$-2087.513,373$</td>
</tr>
</tbody>
</table>
Parallel efficiency

Cr$_2$/cc-pVQZ, $2 \times 10^7$ determinants
Many things that will be done:

1. Large systems: Use JM-MRPT2 (Giner et al) instead of Epstein-Nesbet
   - Needs a partition of the MO space (CAS)
   - Perturbers are Slater determinants (decontracted formalism)
   - Size-consistent
   - Less sophisticated than NEV-PT2, but of comparable quality

2. CIPSI
   - Speed-up selection
   - Stochastic Shifted-$B_k$ method

3. Stochastic Multi-Reference Coupled Cluster
   - Use same algorithm for triples and quadruples

4. Parallelism
   - Target: 1M CPU cores
   - Hybrid CPU/GPU
Quantum Package: Wave function methods implemented with the CIPSI algorithm

- Full-CI
- CAS+SD (MR-CI)
- MR-CCSD(T)
- DDCI
- “Coupled-clusterized” DDCI
- PT2 to take account of the non-selected determinants
People involved

- Michel Caffarel (LCPQ)
- Emmanuel Giner (Ph.D LCPQ → LCT)
- Yann Garniron (Ph.D LCPQ)
- Thomas Applencourt (Ph.D LCPQ → post-doc Argonne)
- Jean-Paul Malrieu (LCPQ)
- Pierre-François Loos (LCPQ)

