QMC USING VERY LARGE CONFIGURATION INTERACTION-TYPE EXPANSIONS

arXiv:1510.00730v2 [physics.chem-ph]

A. Scemama[1], T. Applencourt, E. Giner, M. Caffarel

18/12/2015

[1]Lab. 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

[1] arXiv:1408.3672 [physics.chem-ph]
[2] arXiv:1409.3671 [physics.chem-ph]

- *Exact* solution in a given basis set
- Basis : All possible Slater determinants ($\mathcal{D}_k$) of *spin-orbitals*

$$\Psi_{\mathrm{FCI}} = \sum_k c_k \mathcal{D}_k \tag{1}$$

- Spin-Free formalism :

$$\mathcal{D}_k = D_k^{\uparrow} \times D_k^{\downarrow} \tag{2}$$

- Size of the FCI space :

$$\binom{N_{\mathrm{MO}}}{N_{\mathrm{elec}\uparrow}} \times \binom{N_{\mathrm{MO}}}{N_{\mathrm{elec}\downarrow}} \tag{3}$$

Cl in cc-pVDZ

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

$\Longrightarrow$ CIPSI algorithm to select Slater determinants *on the fly*.[3]

[3] `http://github.com/LCPQ/quantum_package`

Ψ can be expressed in a bi-linear form:

$$\Psi(\mathsf{R}) = \sum_k c_k \mathcal{D}_k = \sum_i \sum_j C_{ij} D_{i\uparrow}(\mathsf{R}_\uparrow) D_{j\downarrow}(\mathsf{R}_\downarrow) \tag{4}$$

- $\mathsf{D}_\uparrow(\mathsf{R}_\uparrow)$: vector of $N_{\mathrm{det}\uparrow}$ elements
- $\mathsf{D}_\downarrow(\mathsf{R}_\downarrow)$: vector of $N_{\mathrm{det}\downarrow}$ elements
- C : $N_{\mathrm{det}\uparrow} \times N_{\mathrm{det}\downarrow}$ matrix. The matrix contains N_{det} elements.

C is constant in a QMC calculation $\Longrightarrow$ preprocessing.

DETERMINATION OF C

1. Encoding of determinants[4] :

HF	:	(000011111111,000111111111)	=	(255,511)
Single ↑	:	(010001111111,000111111111)	=	(1151,511)
Double ↑↓	:	(010001111111,010101111111)	=	(1151,1407)

[4] arXiv:1311.6244 [physics.comp-ph]

DETERMINATION OF C

1. Encoding of determinants[4] :

HF	:	(000011111111,000111111111)	=	(255,511)
Single ↑	:	(010001111111,000111111111)	=	(1151,511)
Double ↑↓	:	(010001111111,010101111111)	=	(1151,1407)

2. Separate $D_\uparrow$ and $D_\downarrow$ of $\mathcal{D}_k$: two lists of N_{det}

000011111111	=	255	000111111111	=	511
010001111111	=	1151	000111111111	=	511
010001111111	=	1151	010101111111	=	1407

[4] arXiv:1311.6244 [physics.comp-ph]

DETERMINATION OF C

1. Encoding of determinants[4] :

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

   Single ↑ : (010001111111,000111111111) = (1151,511)

   Double ↑↓ : (010001111111,010101111111) = (1151,1407)

2. Separate $D_\uparrow$ and $D_\downarrow$ of $\mathcal{D}_k$: two lists of N_{det}

000011111111	=	255	000111111111	=	511
010001111111	=	1151	000111111111	=	511
010001111111	=	1151	010101111111	=	1407

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

000011111111	=	255	000111111111	=	511
010001111111	=	1151	010101111111	=	1407

[4] arXiv:1311.6244 [physics.comp-ph]

DETERMINATION OF C

1. Encoding of determinants[4] :

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

   Single ↑ : (010001111111,000111111111) = (1151,511)

   Double ↑↓ : (010001111111,010101111111) = (1151,1407)

2. Separate $D_\uparrow$ and $D_\downarrow$ of $\mathcal{D}_k$: two lists of N_{det}

000011111111	=	255	000111111111	=	511
010001111111	=	1151	000111111111	=	511
010001111111	=	1151	010101111111	=	1407

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

000011111111	=	255	000111111111	=	511
010001111111	=	1151	010101111111	=	1407

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

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

[4] arXiv:1311.6244 [physics.comp-ph]

REMARK ON ORDERING

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

List index	Decimal	Binary	Determinant
1	15	00001111	$\|1234\rangle$
2	23	00010111	$\|1235\rangle$
3	27	00011011	$\|1245\rangle$
4	29	00011101	$\|1345\rangle$
5	30	00011110	$\|2345\rangle$
6	39	00100111	$\|1236\rangle$
7	43	00101011	$\|1246\rangle$
8	45	00101101	$\|1346\rangle$
...	...	...	...

In QMC=Chem : We do *all-electron* moves.

At every MC step, we need to evaluate:

$$\Psi = (D_\uparrow{}^\dagger(C)D_\downarrow) \quad (5)$$

$$\nabla_i\Psi = \nabla_i D_\uparrow{}^\dagger.(CD_\downarrow) \text{ or } (D_\uparrow{}^\dagger C).\nabla_i D_\downarrow \quad (6)$$

$$\Delta_i\Psi = \Delta_i D_\uparrow{}^\dagger.(CD_\downarrow) \text{ or } (D_\uparrow{}^\dagger C).\Delta_i D_\downarrow \quad (7)$$

$$V_{\text{pseudo}}^{\text{non-loc}}\Psi = V_{\text{pseudo}}^{\text{non-loc}} D_\uparrow{}^\dagger.(CD_\downarrow) \text{ or } (D_\uparrow{}^\dagger C).V_{\text{pseudo}}^{\text{non-loc}} D_\downarrow \quad (8)$$

($\uparrow$ electrons and $\downarrow$ electrons)

Vector-matrix product $\mathsf{D}_{\uparrow}{}^{\dagger}.\mathsf{C}$: N_{det} operations, returns a $N_{\mathrm{det}\downarrow}$ vector

1. Dot product with $\mathsf{D}_{\downarrow}$: $N_{\mathrm{det}\downarrow}$ operations, produces a scalar
2. Matrix product with $\nabla\mathsf{D}_{\downarrow}$: $3N_{\mathrm{elec}\downarrow} \times N_{\mathrm{det}\downarrow}$ operations, produces a $3N_{\mathrm{elec}\downarrow}$ vector
3. Matrix product with $\Delta\mathsf{D}_{\downarrow}$: $N_{\mathrm{elec}\downarrow} \times N_{\mathrm{det}\downarrow}$ operations, produces a $N_{\mathrm{elec}\downarrow}$ vector
4. Matrix product with $V_{\mathrm{pseudo}}^{\mathrm{non-loc}}\mathsf{D}_{\downarrow}$: $N_{\mathrm{elec}\downarrow} \times N_{\mathrm{det}\downarrow}$ operations, produces a $N_{\mathrm{elec}\downarrow}$ vector

If one localizes the pseudo-potential on the CI wave function :

$$\hat{T} + \hat{V} + \hat{V}_{\mathrm{loc}}^{\mathrm{pseudo}} + \sum_{lm} V_{\mathrm{non-loc},l}^{\mathrm{pseudo}} \frac{|Y_{lm}\rangle\langle Y_{lm}|\Psi_{\mathrm{CI}}\rangle}{\Psi_{\mathrm{CI}}} \tag{9}$$

$$V_{\mathrm{non-loc}}^{\mathrm{pseudo},i} = \sum_{j} S_{ij}^{-1} \sum_{lm} Y_{lm}(\mathbf{r}_i)\langle Y_{lm}|\phi_j\rangle \tag{10}$$

- 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

- 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

Excitation	$D_\uparrow D_\downarrow$	$D_\uparrow$	$D_\downarrow$
0	1	1	1
1	38	90	88
2	2 177	1603	1520
3	43 729	7811	6507
4	308 045	8581	5071
5	351 182	2090	579
6	291 481	77	0
7	3 067	0	0
8	280	0	0

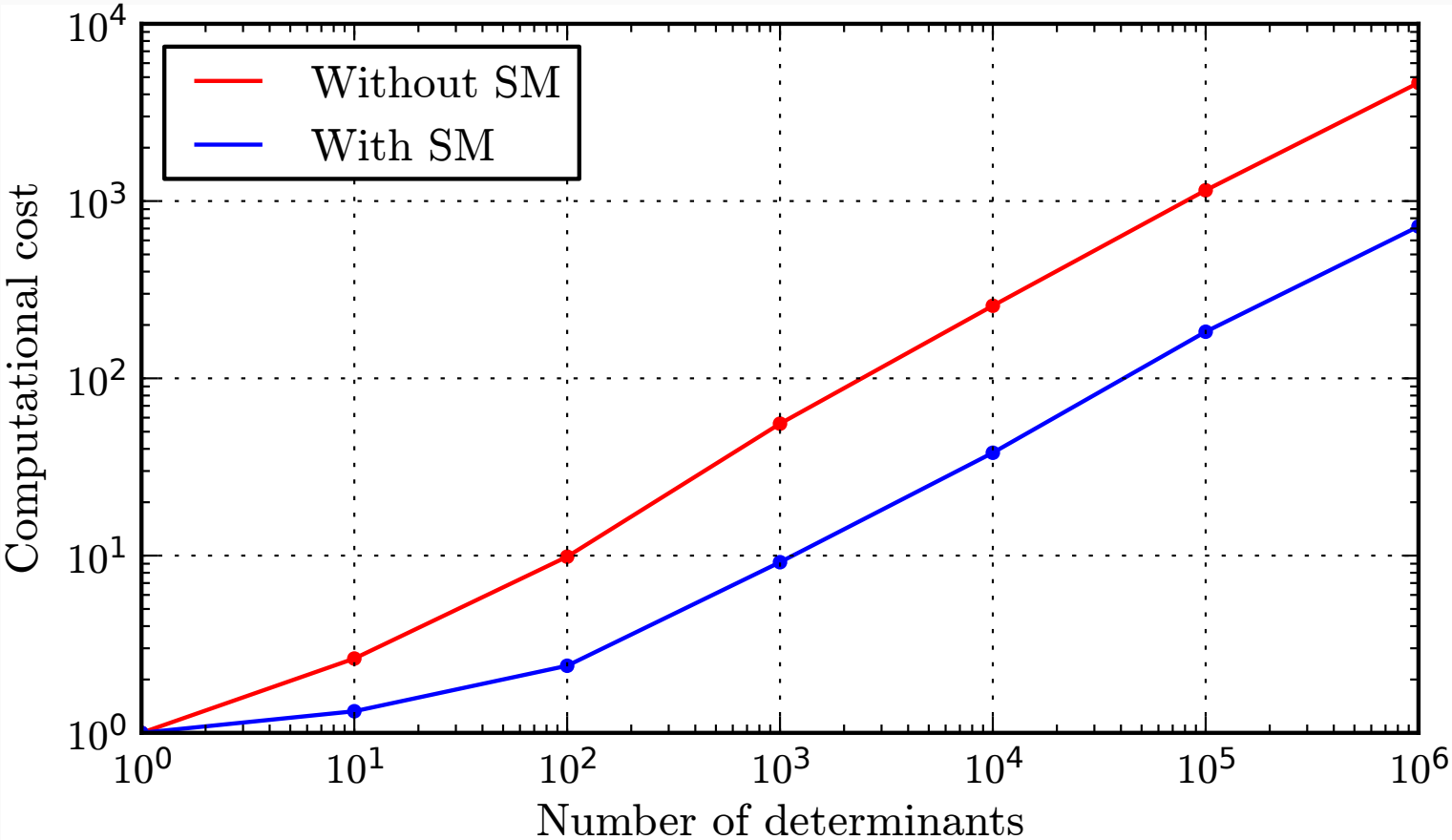
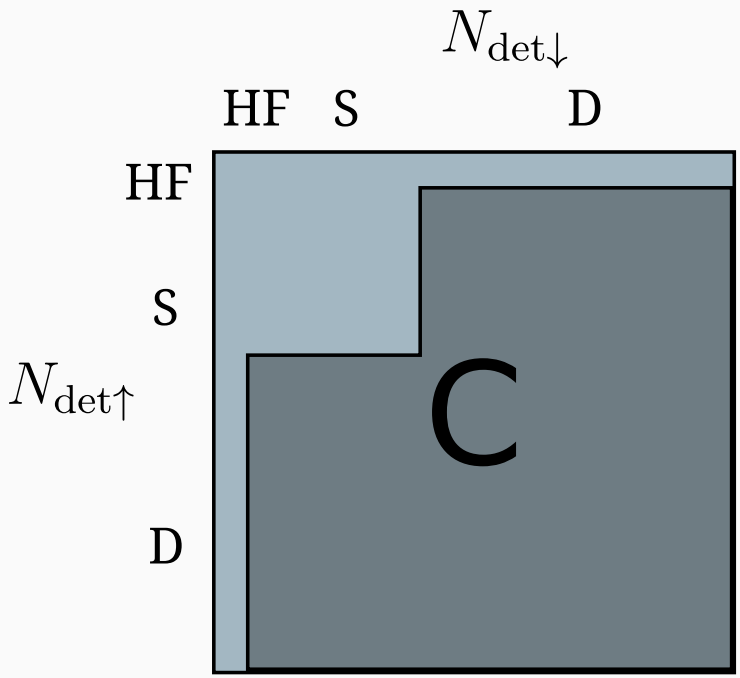


Figure 1: Cl : Scaling is $\sim \sqrt{N_{\text{det}}}$. SM updates $\times 7$.

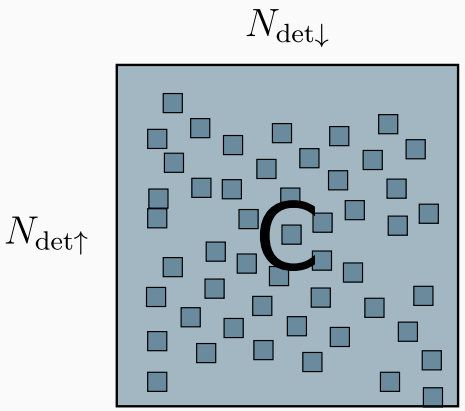
a priori truncation of the FCI space

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



Usual truncation

- Based on the $|c_k|$ (or CSF coefficients)
- Removes mostly same-spin doubles in CISD $\Longrightarrow$ 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

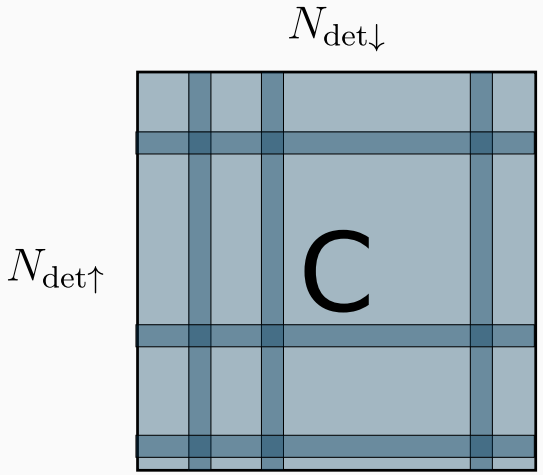


Alternate truncation scheme

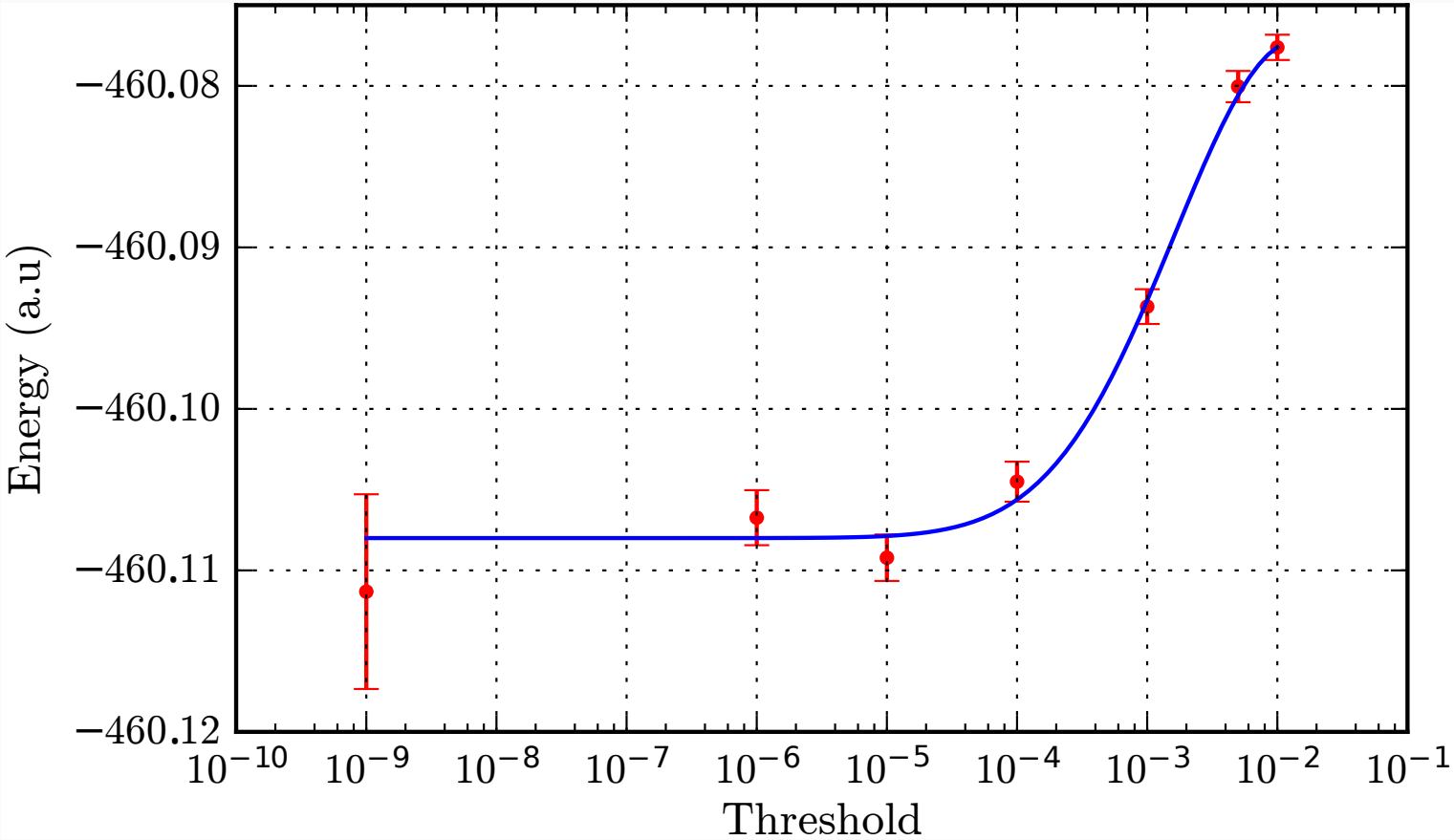
- Contribution of $D_\uparrow$ and $D_\downarrow$ to the norm of Ψ

$$\mathcal{N} = \sum_{i=1}^{N_{\det\uparrow}} \sum_{j=1}^{N_{\det\downarrow}} C_{ij}^2 = \sum_{i=1}^{N_{\det\uparrow}} \mathcal{N}_i^{\uparrow} = \sum_{j=1}^{N_{\det\downarrow}} \mathcal{N}_j^{\downarrow} \tag{11}$$

- Remove $D_{i\uparrow}$ when $\mathcal{N}_i^{\uparrow} = \sum_{j=1}^{N_{\det\downarrow}} C_{ij}^2 < \epsilon$

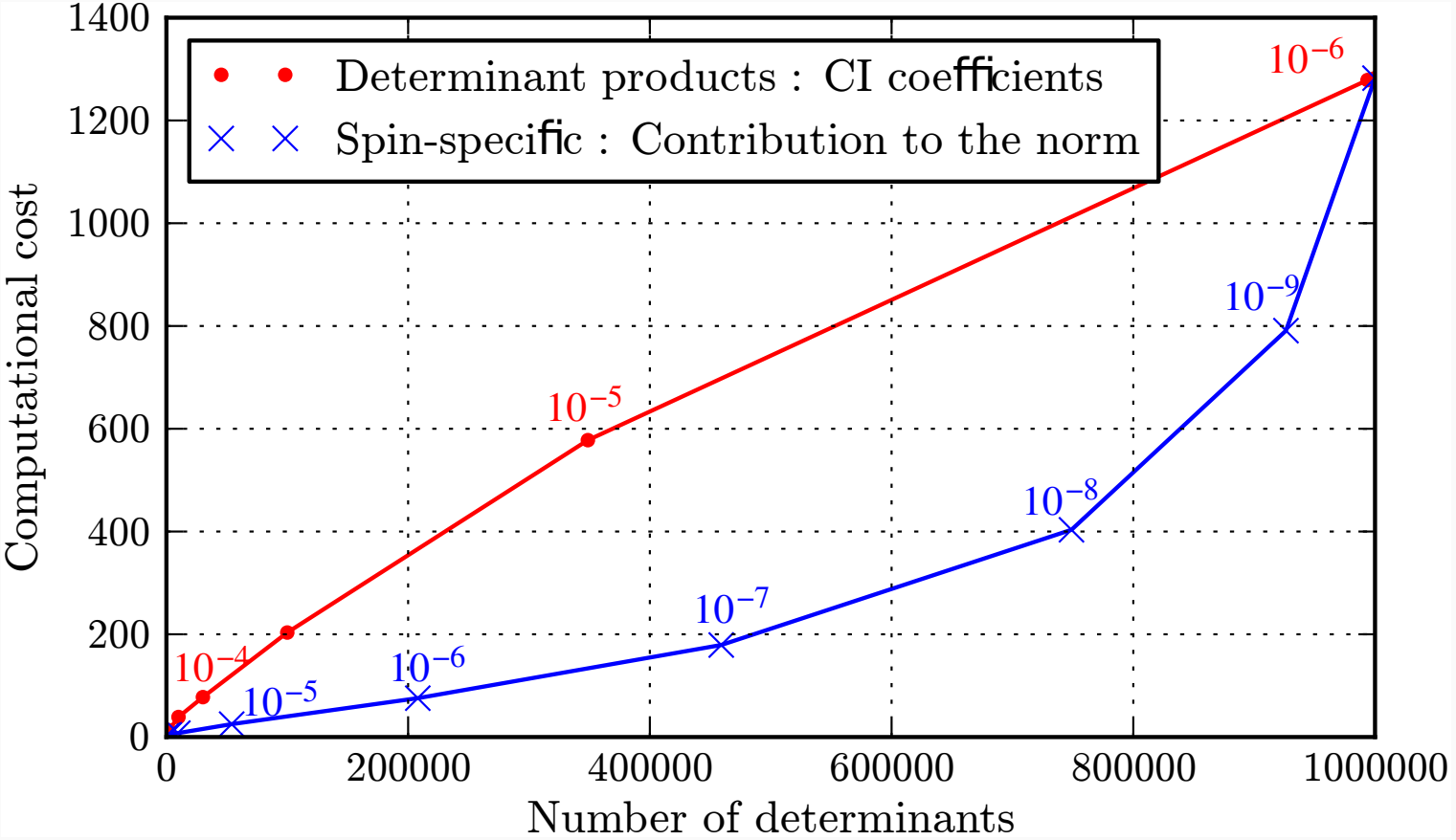


CONVERGENCE OF THE FN-DMC ENERGY



Cl FCI/cc-pVTZ

COMPUTATIONAL COST



Converged FCI/cc-pVTZ FN-DMC energy with $\epsilon = 10^{-5}$: 7.1× the cost of a single-determinant

TOTAL CPU TIME PER MC STEP

ϵ	N_{det}	$N_{\mathrm{det}}^{\uparrow}$	$N_{\mathrm{det}}^{\downarrow}$	CPU time(ms)
10^{-2}	1	1	1	0.02450
5.10^{-3}	3	3	2	0.02688
10^{-3}	86	21	17	0.03899
5.10^{-4}	214	29	28	0.04636
10^{-4}	1361	93	74	0.08808
5.10^{-5}	2424	120	89	0.1054
10^{-5}	9485	234	166	0.1855
10^{-6}	54016	772	523	0.5960
10^{-7}	207995	2279	1389	1.740
10^{-8}	459069	5797	3291	4.196
10^{-9}	748835	14456	8054	9.724
10^{-10}	926299	30320	16571	19.18
0	1000000	52433	26833	31.92

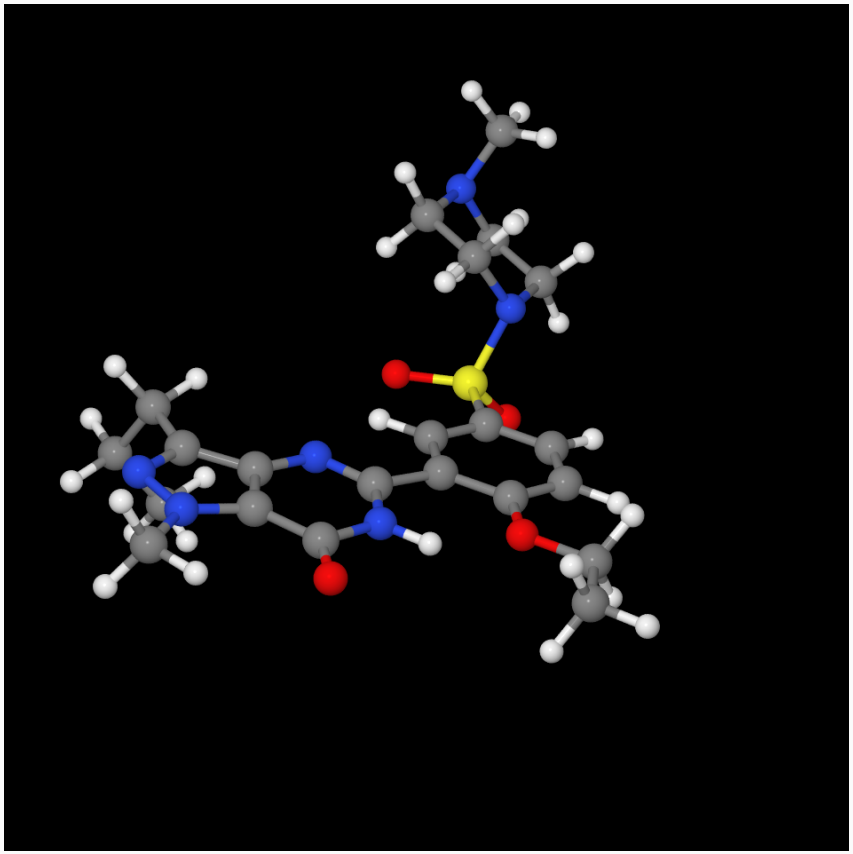
Order of magnitude :

36 ms/step $\equiv 10^{11}$ steps in 10 hours on 100 000 cores

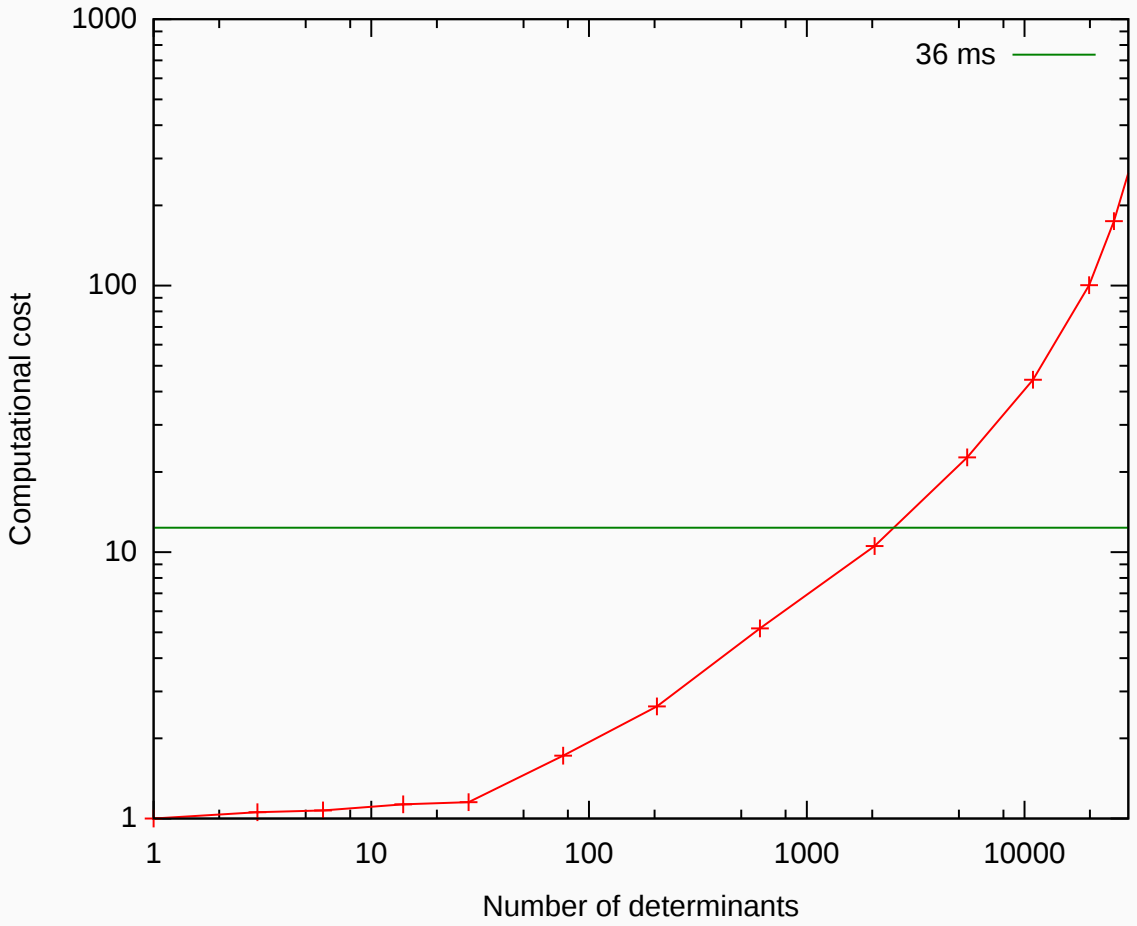
CAN BE APPLIED TO LARGER SYSTEMS

Sildenafil

- 63 atoms
- 126 $\uparrow$ and 126 $\downarrow$ electrons
- 649 basis functions
- MP2 wave function



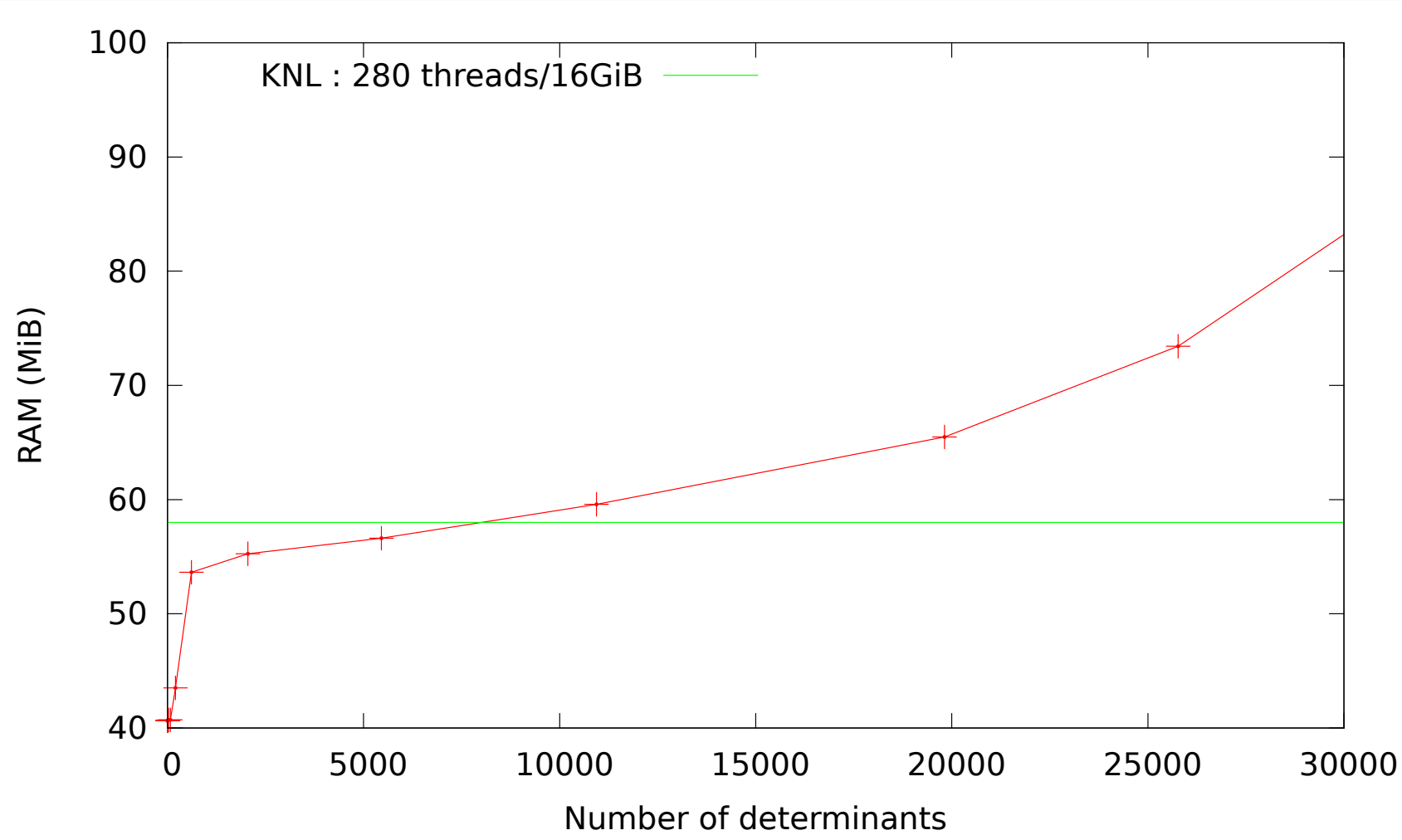
TOTAL TIME



Sildenafil : 31 ms/step $\Leftrightarrow \epsilon = 2\,10^{-5}$ ($N_{\mathrm{det}} = 2049$)

100 dets only 2$\times$ more expensive than 1 det.

MEMORY PER CPU CORE



SUMMARY

- Bi-linear expression : scaling as $\sqrt{N_{\mathrm{det}}}$
- The only N_{det}-scaling step is a matrix-vector product : tiny pre-factor, independent of N_{elec}, negligible
- Sherman-Morrison : $\sim \times N_{\mathrm{elec}\uparrow}$
- 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 $N_{\mathrm{elec}\uparrow} \times N_{\mathrm{det}\uparrow}$
- 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
- More details here : *arXiv:1510.00730v2 [physics.chem-ph]*

COMPARISON WITH TABLE METHOD

Table method : store a 2D array where each element is the result of a dot product

- Dot product : $N_{\mathrm{elec}\uparrow}$ 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 $\uparrow$ electrons
- L2 $\Leftrightarrow$ 192 $\uparrow$ electrons
- L3 $\Leftrightarrow$ 704 $\uparrow$ electrons
- RAM $\Leftrightarrow$ 5712 $\uparrow$ electrons

COMPARISON WITH TABLE METHOD

Measurements on Cl, 1M determinants:

	$D_\uparrow$	$D_\downarrow$
Table method	2110 cycles	2431 cycles
This work	544 cycles	359 cycles

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).