Accelerated MR-PT2 with a Hybrid Stochastic/Deterministic Algorithm

arXiv:1703.05347 [physics.chem-ph]

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 : $d(I, J) = \left(k : |J\rangle = \hat{T}_k |I\rangle\right)$

CIPSI algorithm

1. Define a *reference* wave function:

$$|\Psi\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad\qquad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

CIPSI algorithm

1. Define a *reference* wave function:

$$|\Psi\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

2. Generate external determinants:

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

CIPSI algorithm

1. Define a *reference* wave function:

$$|\Psi\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

2. Generate external determinants:

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

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

$$\delta E(\alpha) = \frac{\langle \Psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \Psi \rangle}{E_{\text{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle}$$

CIPSI algorithm

1. Define a *reference* wave function:

$$|\Psi\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

2. Generate external determinants:

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

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

$$\delta E(\alpha) = \frac{\langle \Psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \Psi \rangle}{E_{\text{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle}$$

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

CIPSI algorithm

1. Define a *reference* wave function:

$$|\Psi\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

2. Generate external determinants:

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

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

$$\delta E(\alpha) = \frac{\langle \Psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \Psi \rangle}{E_{\text{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle}$$

4. Select the $|\alpha\rangle$ with the largest $\delta E(\alpha)$ and add them into $\mathcal{D}$
5. Diagonalize $\mathcal{H}$ in $\mathcal{D} \Longrightarrow$ update $|\Psi\rangle$ and E_{var}

CIPSI algorithm

1. Define a *reference* wave function:

$$|\Psi\rangle = \sum_{I \in \mathcal{D}} c_I |I\rangle \qquad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

2. Generate external determinants:

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

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

$$\delta E(\alpha) = \frac{\langle \Psi | \mathcal{H} | \alpha \rangle \langle \alpha | \mathcal{H} | \Psi \rangle}{E_{\text{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle}$$

4. Select the $|\alpha\rangle$ with the largest $\delta E(\alpha)$ and add them into $\mathcal{D}$
5. Diagonalize $\mathcal{H}$ in $\mathcal{D} \Longrightarrow$ update $|\Psi\rangle$ and E_{var}
6. Iterate

Illustration

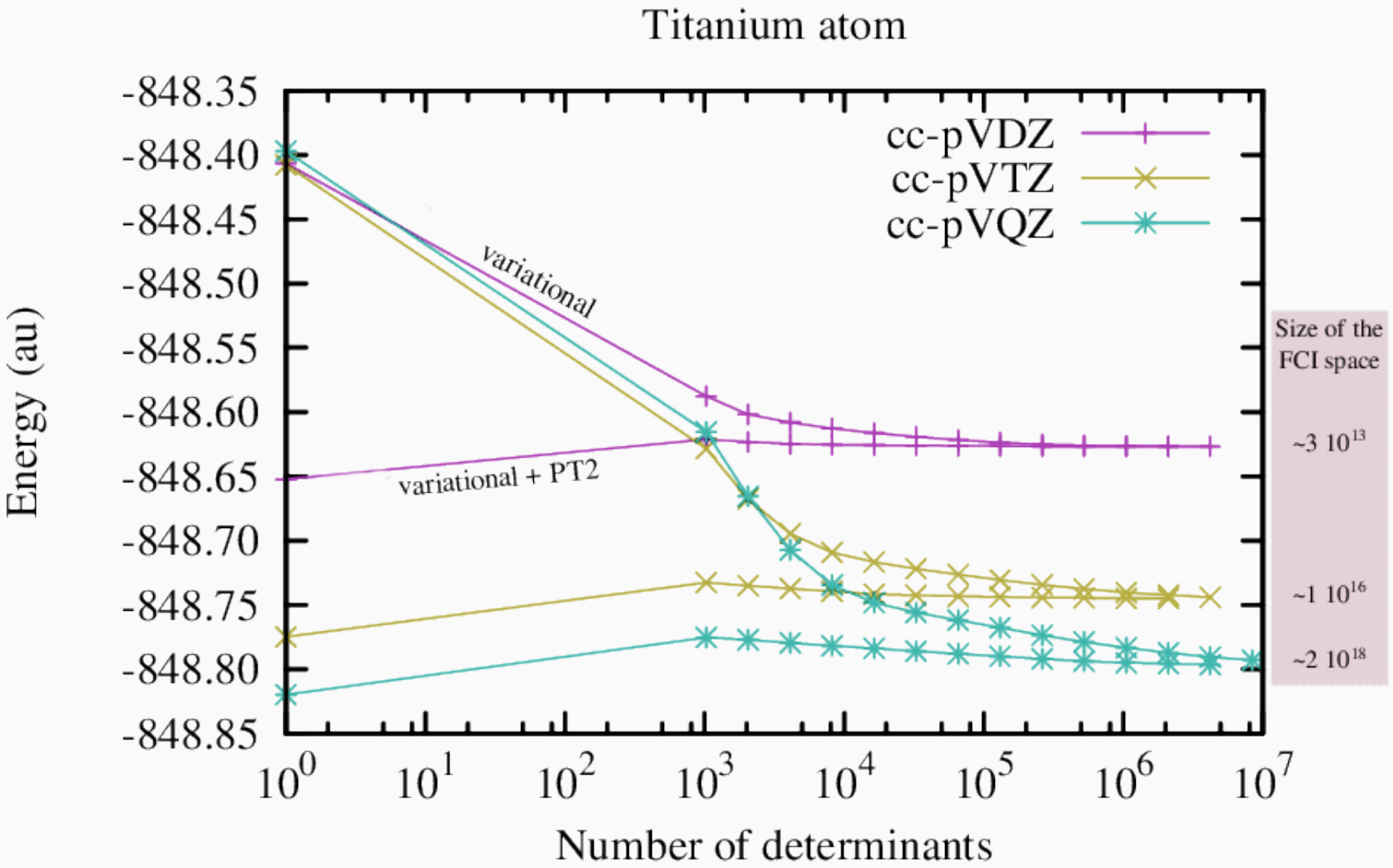


Figure 1: Titanium atom, CIPSI.

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

Practical considerations

At any time, $E_{\text{PT2}} = \sum_\alpha \delta E(\alpha)$ estimates the distance to the solution

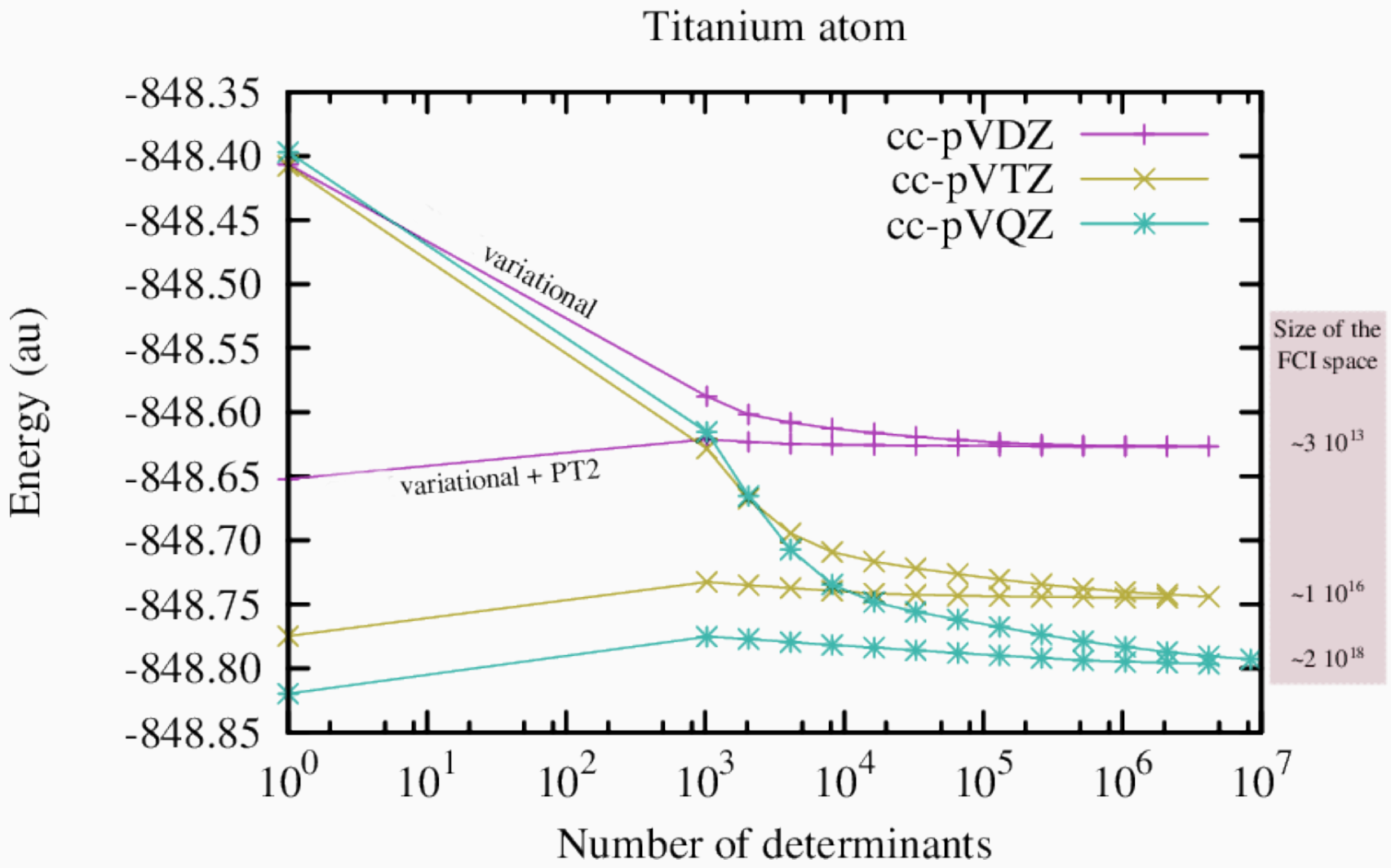
- The $|\alpha\rangle$ with largest $\delta E(\alpha)$ have been added to Ψ in the past
  $\Longrightarrow$ Only small contributions remain
- $\mathcal{A}$ increases with $\mathcal{D}$
- $\Longrightarrow$ a *very* large number of *very* small contributions

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

This work:

Fast and unbiased E_{PT2}

Illustration



General Problem of MR-PT2 calculations

Epstein-Nesbet Second order correction

Consider a wave function Ψ expanded on an *arbitrary* set $\mathcal{D}$ of N_{det} orthonormal Slater determinants,

$$\Psi = \sum_{I \in \mathcal{D}} c_I |I\rangle, \quad E_{\text{var}} = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$$

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

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

The set $\mathcal{A}$ contains the Slater determinants

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

Formal Scaling

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

- Size of $\mathcal{A}$: size of $(\hat{T}_1 + \hat{T}_2)|\Psi\rangle$
- Number of non-zero terms : $d(I, \alpha) \leq 2$

  $\sim N_{\text{det}} \times \left[\left(N_{\text{elec}}^{\uparrow} \times (N_{\text{MO}} - N_{\text{elec}}^{\uparrow}) \right)^2 \right]$
- 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 $\mathcal{D}$ to consider only contributions of large c_I
   Truncation $\longrightarrow$ bias : E_{PT2} is a sum of negative values.
2. Monte Carlo sampling in $\mathcal{A}$. Statistical error decreases as $\mathcal{O}\left(1/\sqrt{N_{\text{samples}}}\right) \Longrightarrow$ Difficult to get $10^{-5} a.u$ precision.
3. Parallelism

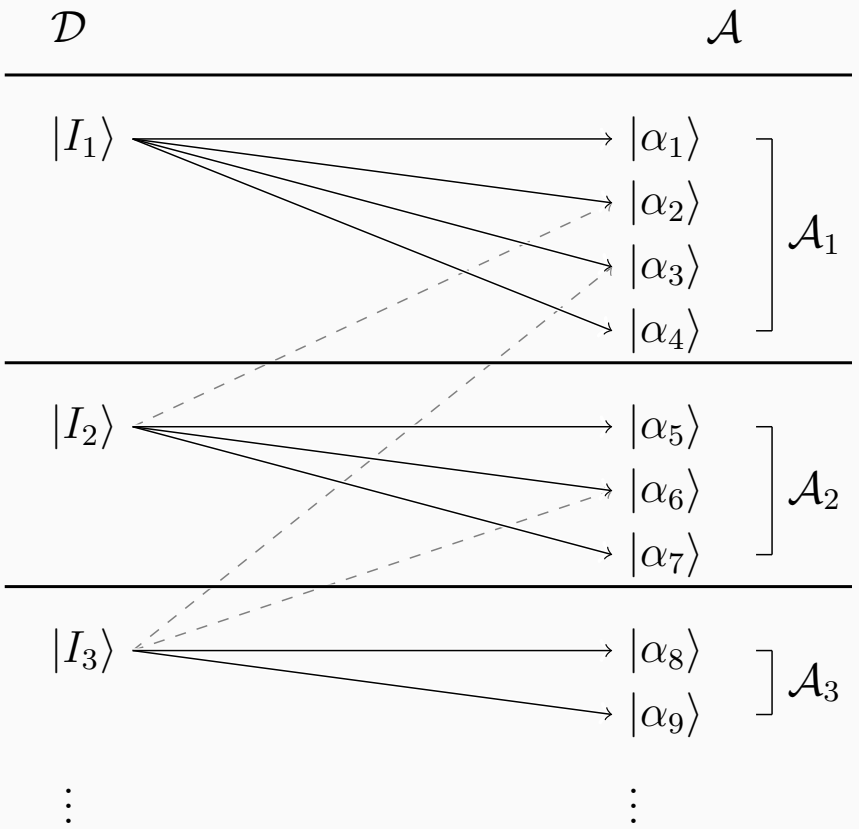
MR-PT2 Algorithm

Central idea

1. Choose an arbitrary ordering of $|I\rangle$. Natural choice:

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

2. Make *disjoint* groups $\mathcal{A}_I$ of $|\alpha\rangle$ originating from the same generator $|I\rangle$
3. Each $\mathcal{A}_I$ has its own contribution ϵ_I to E_{PT2}



Central idea

$$
\begin{aligned}
E_{\mathrm{PT2}} &= \sum_{\alpha \in \mathcal{A}} \frac{\left(\langle \Psi | \mathcal{H} | \alpha \rangle\right)^2}{E_{\mathrm{var}} - \langle \alpha | \mathcal{H} | \alpha \rangle} \\
&= \sum_{I \in \mathcal{D}} \sum_{\alpha_I \in \mathcal{A}_I} \frac{\left(\langle \Psi | \mathcal{H} | \alpha_I \rangle\right)^2}{E_{\mathrm{var}} - \langle \alpha_I | \mathcal{H} | \alpha_I \rangle} \\
&= \sum_{I \in \mathcal{D}} \epsilon_I
\end{aligned}
$$

Consequences

$$\epsilon_I = \sum_{\alpha_I \in \mathcal{A}_I} \frac{\left(\langle \Psi | \mathcal{H} | \alpha_I \rangle\right)^2}{E_{\text{var}} - \langle \alpha_I | \mathcal{H} | \alpha_I \rangle}$$

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 varational 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

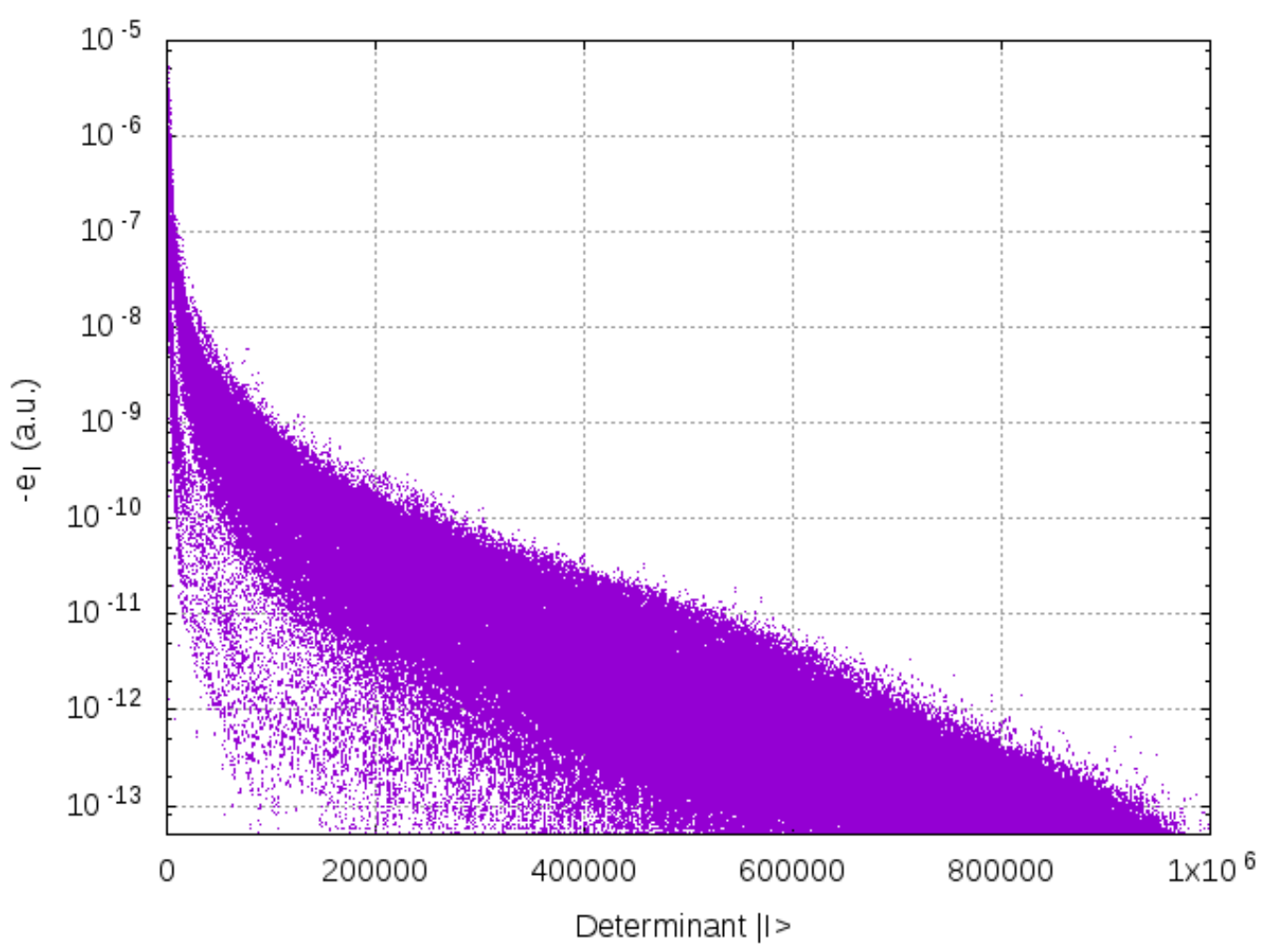


Figure 2: F_2, cc-pVDZ, 10^6 determinants in the variational space

Consequences

N_{det} contributions $\epsilon_I \longrightarrow$ 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

$$E_{\mathrm{PT2}} = \sum_{I \in \mathcal{D}} \epsilon_I = \sum_{I \in \mathcal{D}} p_I \frac{\epsilon_I}{p_I} = \left\langle \frac{\epsilon_I}{p_I} \right\rangle_{p_I}$$

- Draw a generator determinant $|I\rangle$ with probability $p_I = {c_I}^2 / \left(\sum_{J \in \mathcal{D}} p_J\right)$
- Increment n_I, the number of evaluations of ϵ_I
- If ϵ_I is not already computed, compute it and store its value
- $E_{\mathrm{PT2}} \sim \sum_{I \in \mathcal{D}} \frac{n_I}{N_{\mathrm{samples}}} \frac{\epsilon_I}{p_I}$
- Statistical error : $\mathcal{O}\left(1/\sqrt{N_{\mathrm{samples}}}\right)$
- Lazy evaluation : Acceleration (wall-clock time)

Monte Carlo with Lazy Evaluation

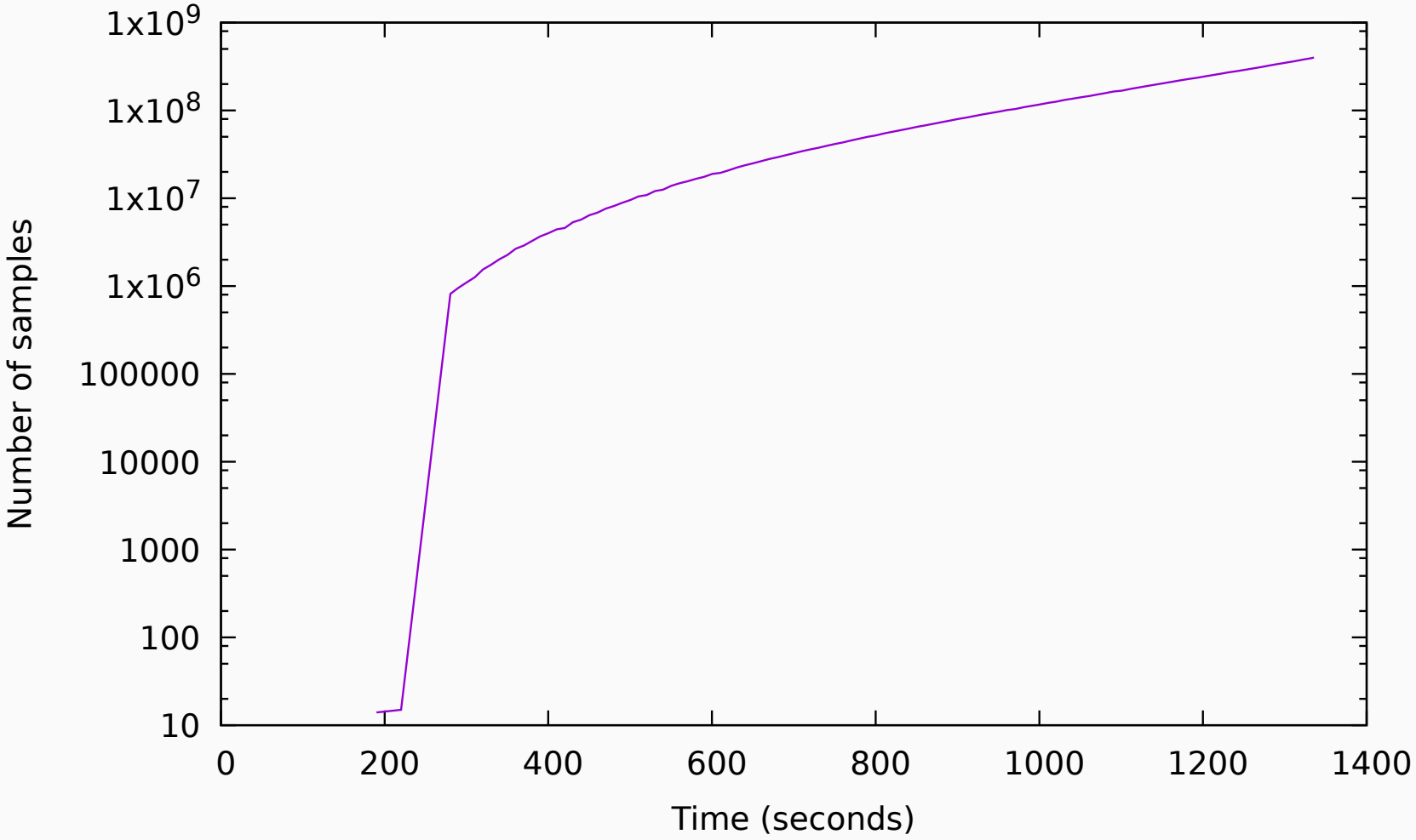


Figure 3: F_2, cc-pVQZ, $5\,10^6$ determinants in the variational space

Monte Carlo with Lazy Evaluation

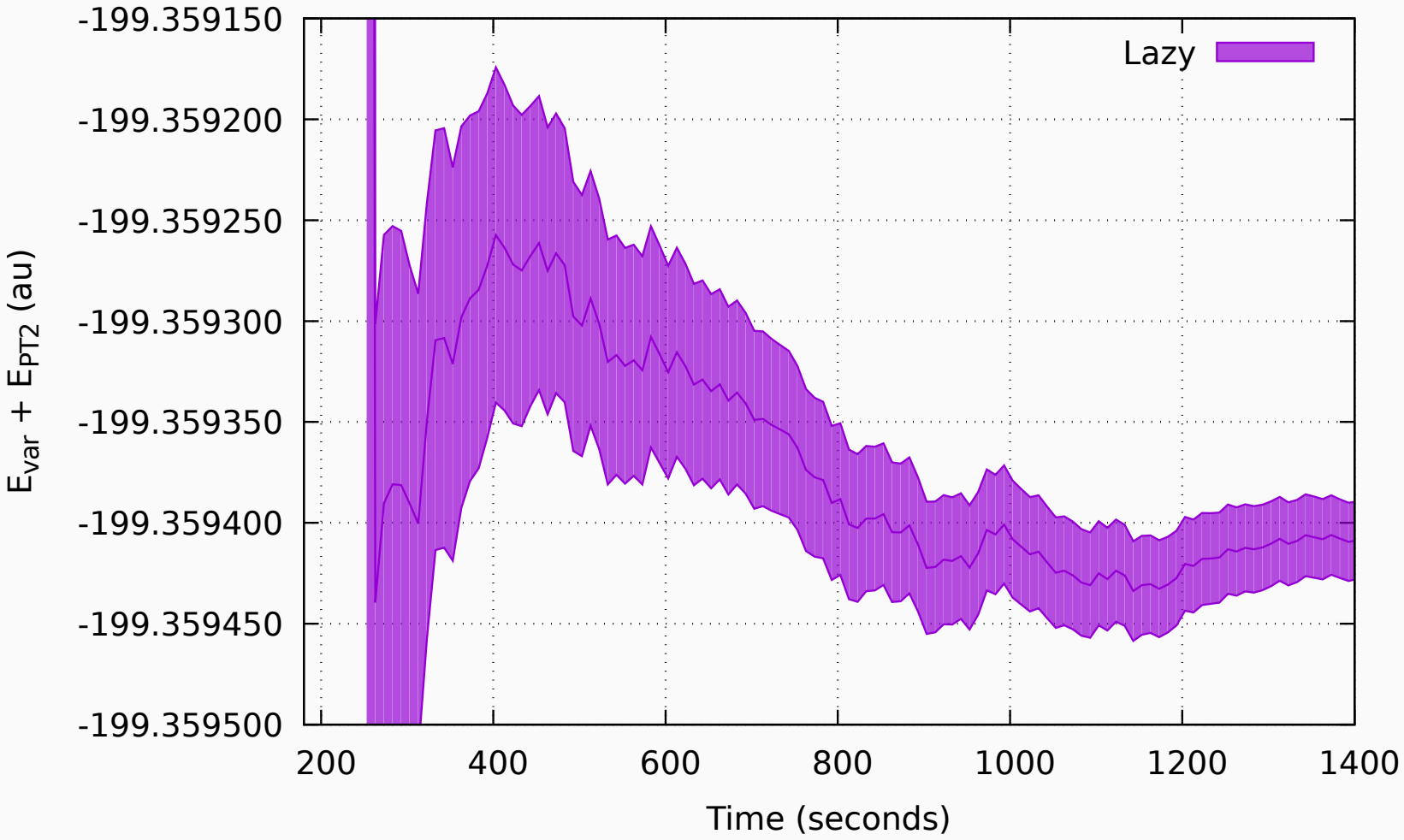


Figure 4: F_2, cc-pVQZ, $5\,10^6$ determinants in the variational space

Monte Carlo with Lazy Evaluation

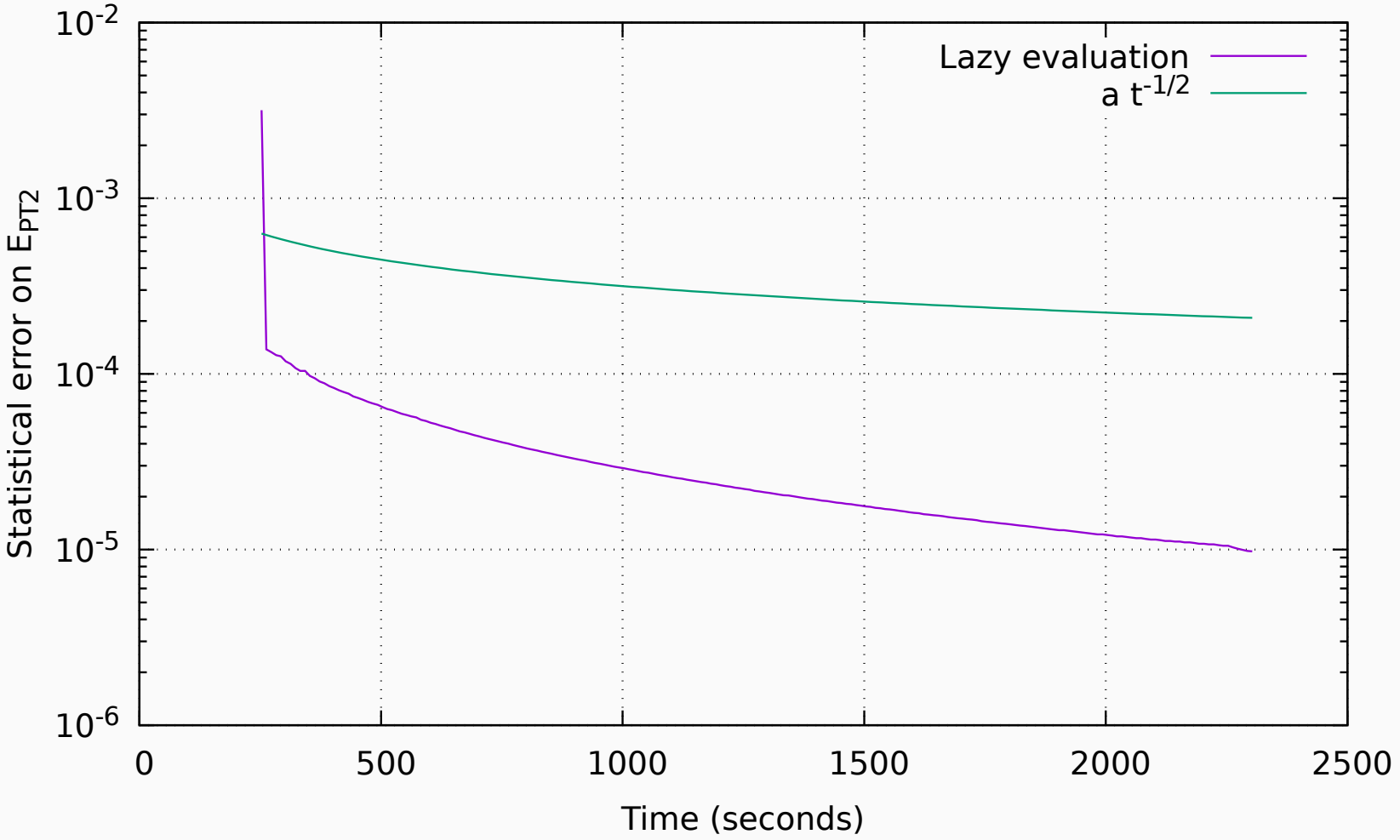
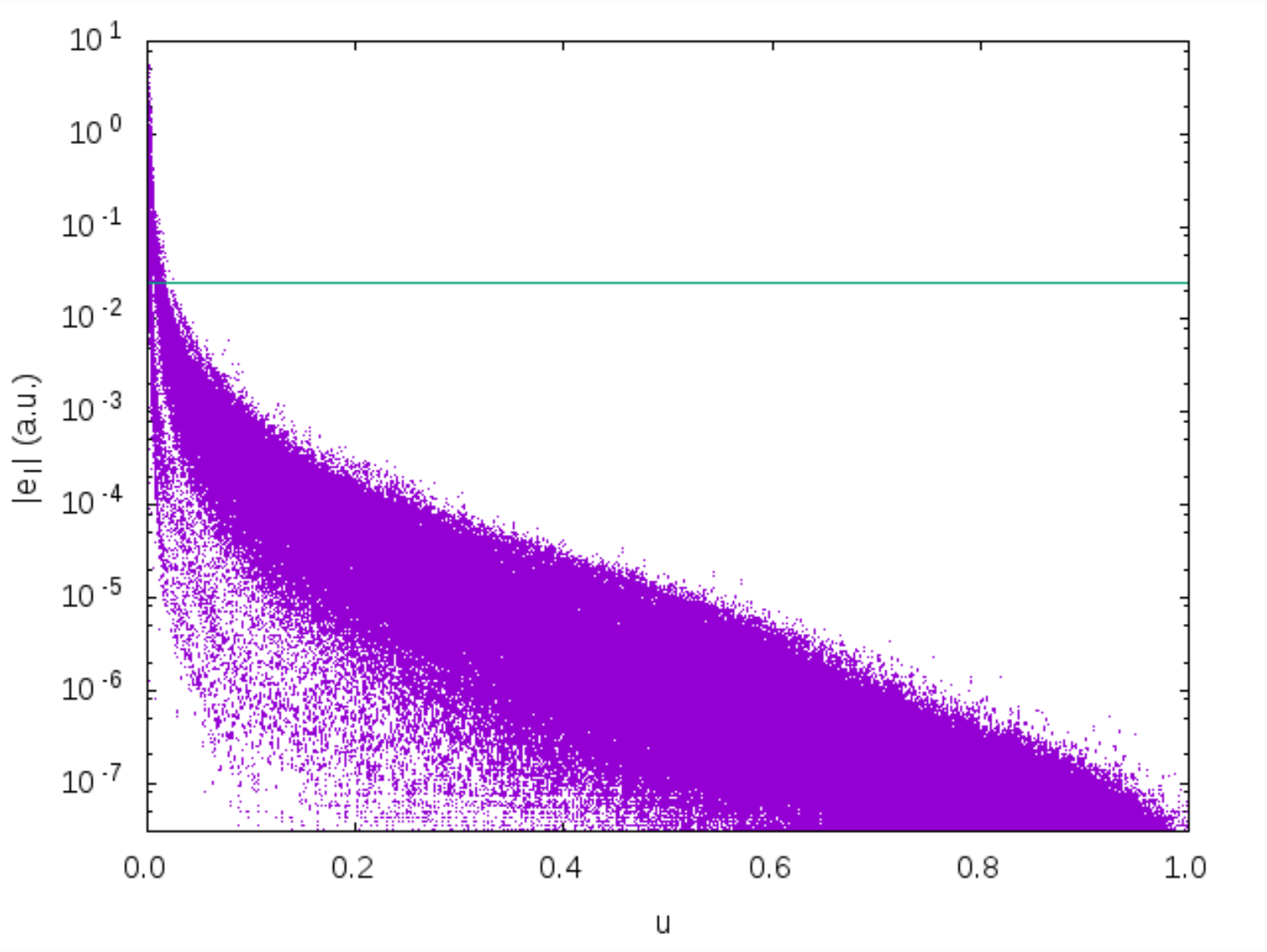


Figure 5: F_2, cc-pVQZ, $5\,10^6$ determinants in the variational space

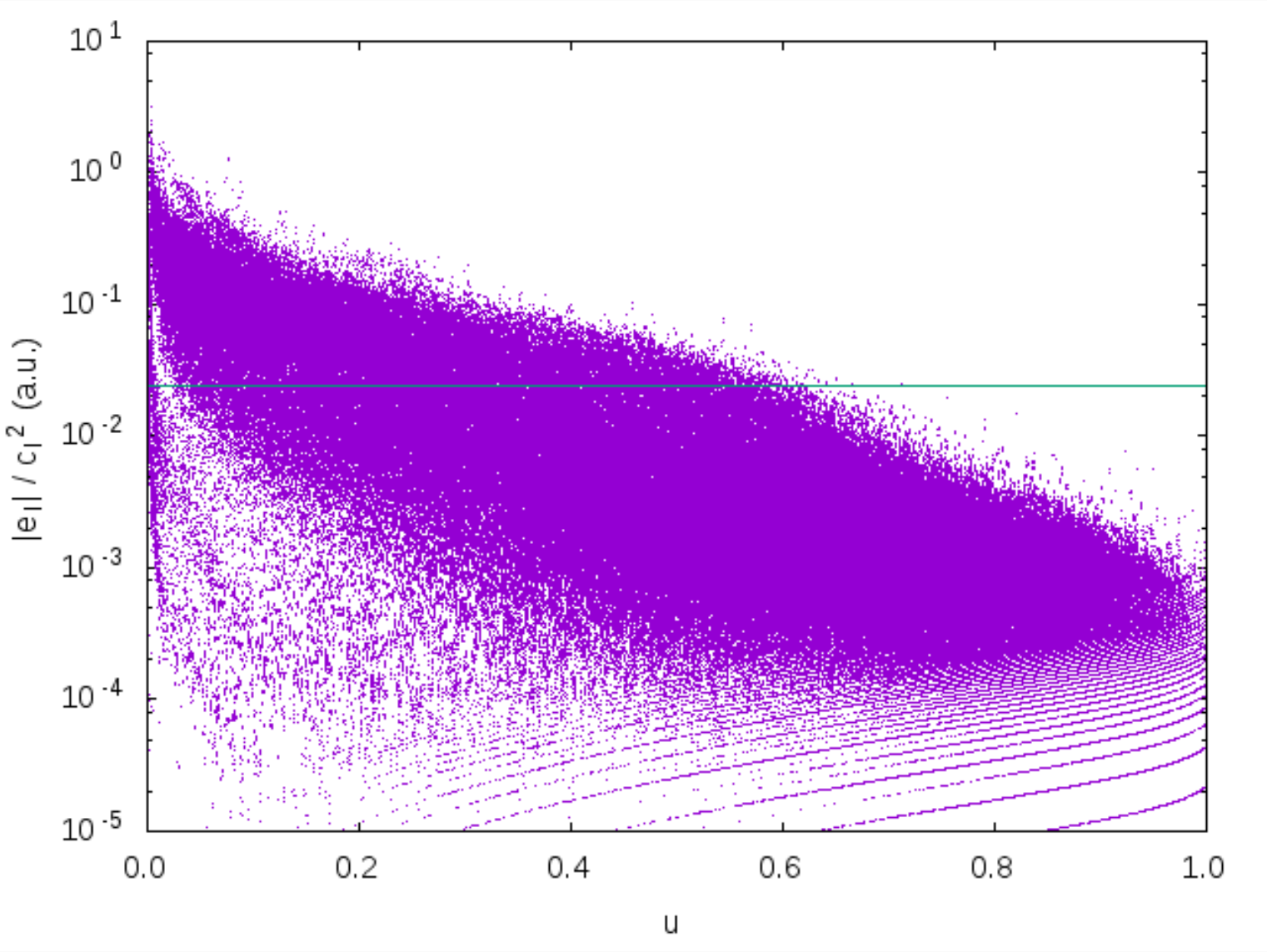
Variance reduction

Uniform sampling : $p_I = 1/N_{\text{det}}$, $E_{\text{PT2}} = \langle N_{\text{det}} \times \epsilon_I \rangle_{p_I}$



Variance reduction

Improved sampling : $p_I = {c_I}^2$, $E_{\mathrm{PT2}} = \langle \epsilon_I / p_I \rangle_{p_I}$



Variance reduction

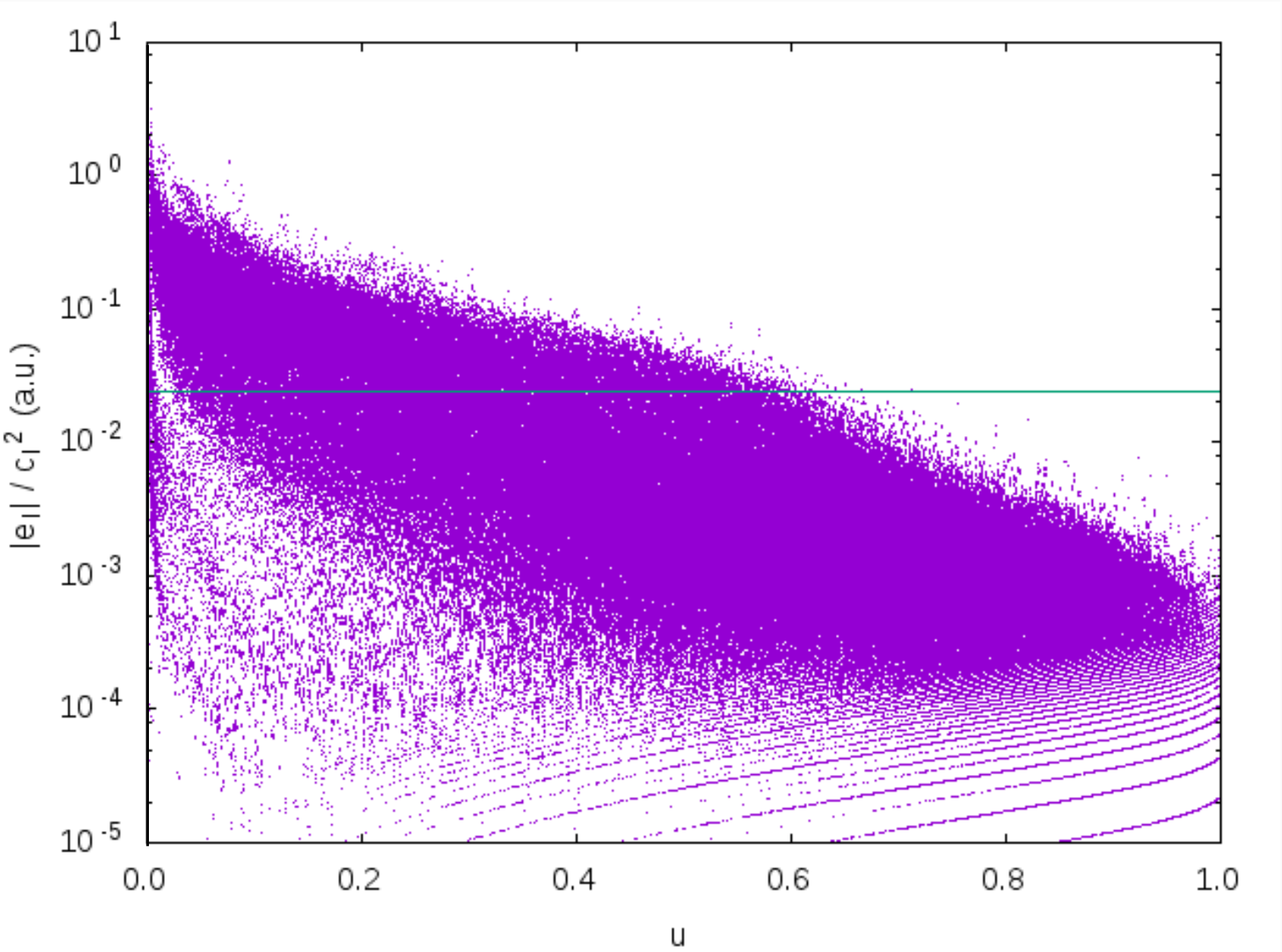
- Noise can be smoothed out by averaging
- Split $\mathcal{D}$ into M *equiprobable* sets : Comb

$$E_{\mathrm{PT2}} = \sum_{I \in \mathcal{D}} \epsilon_I = \sum_{k=1}^{M} \sum_{I_k \in \mathcal{D}_k} \epsilon_{I_k}$$

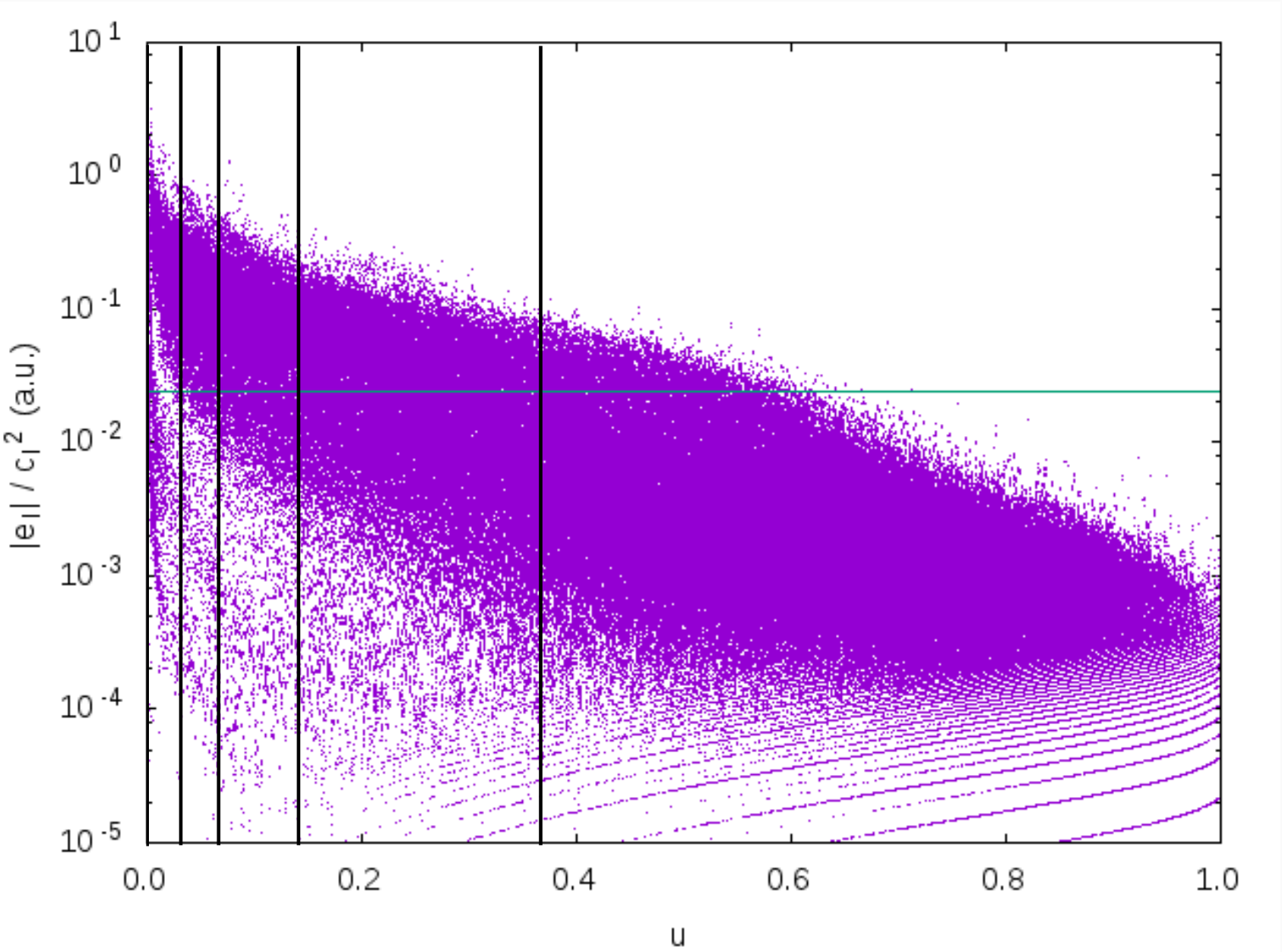
- New Monte Carlo estimator:

$$E_{\mathrm{PT2}} = \left\langle \frac{1}{M} \sum_{k=1}^{M} \frac{\epsilon_{I_k}}{p_{I_k}} \right\rangle_{(p_{I_1},\ldots,p_{I_M})}$$

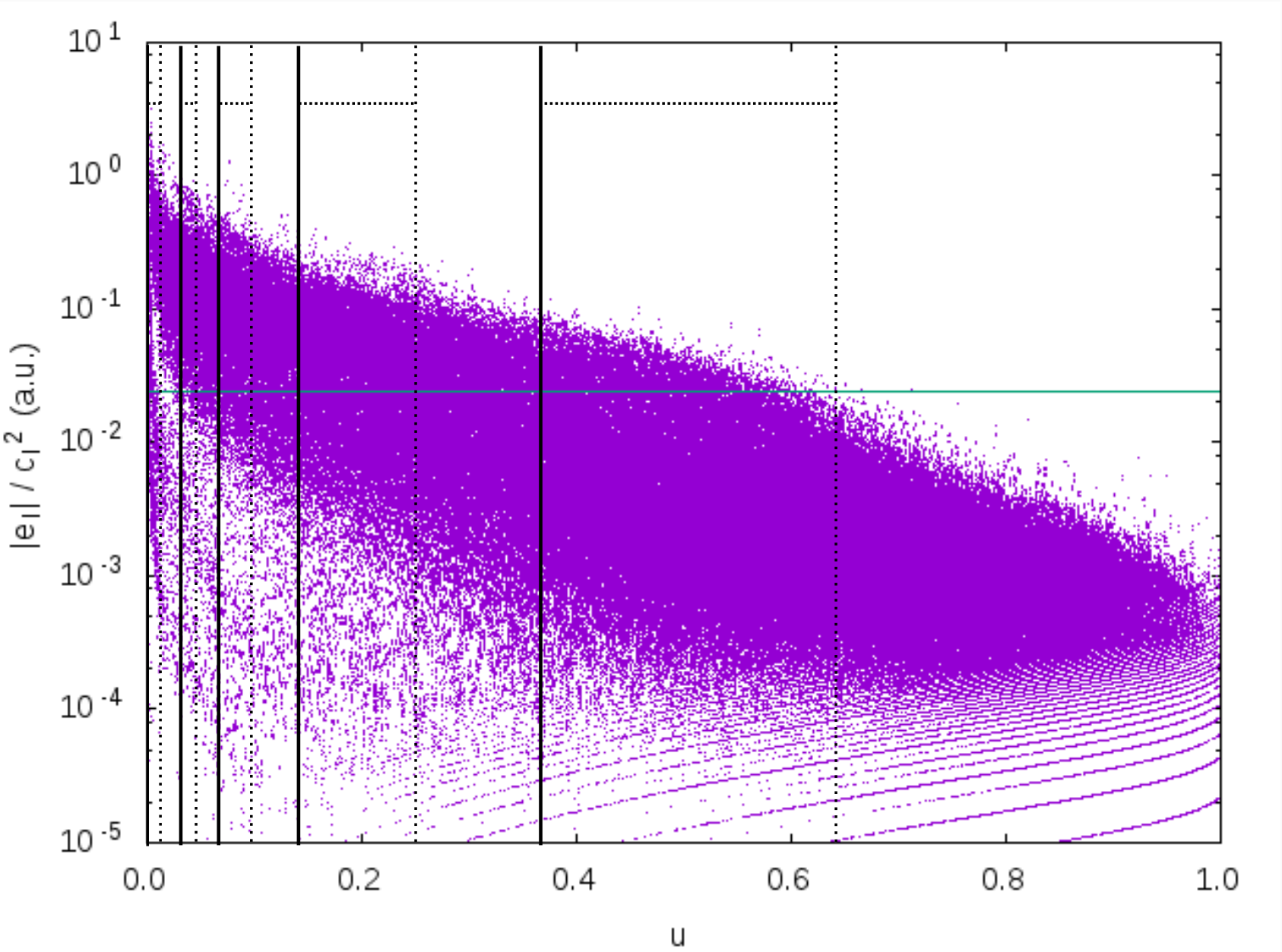
Variance reduction



Variance reduction

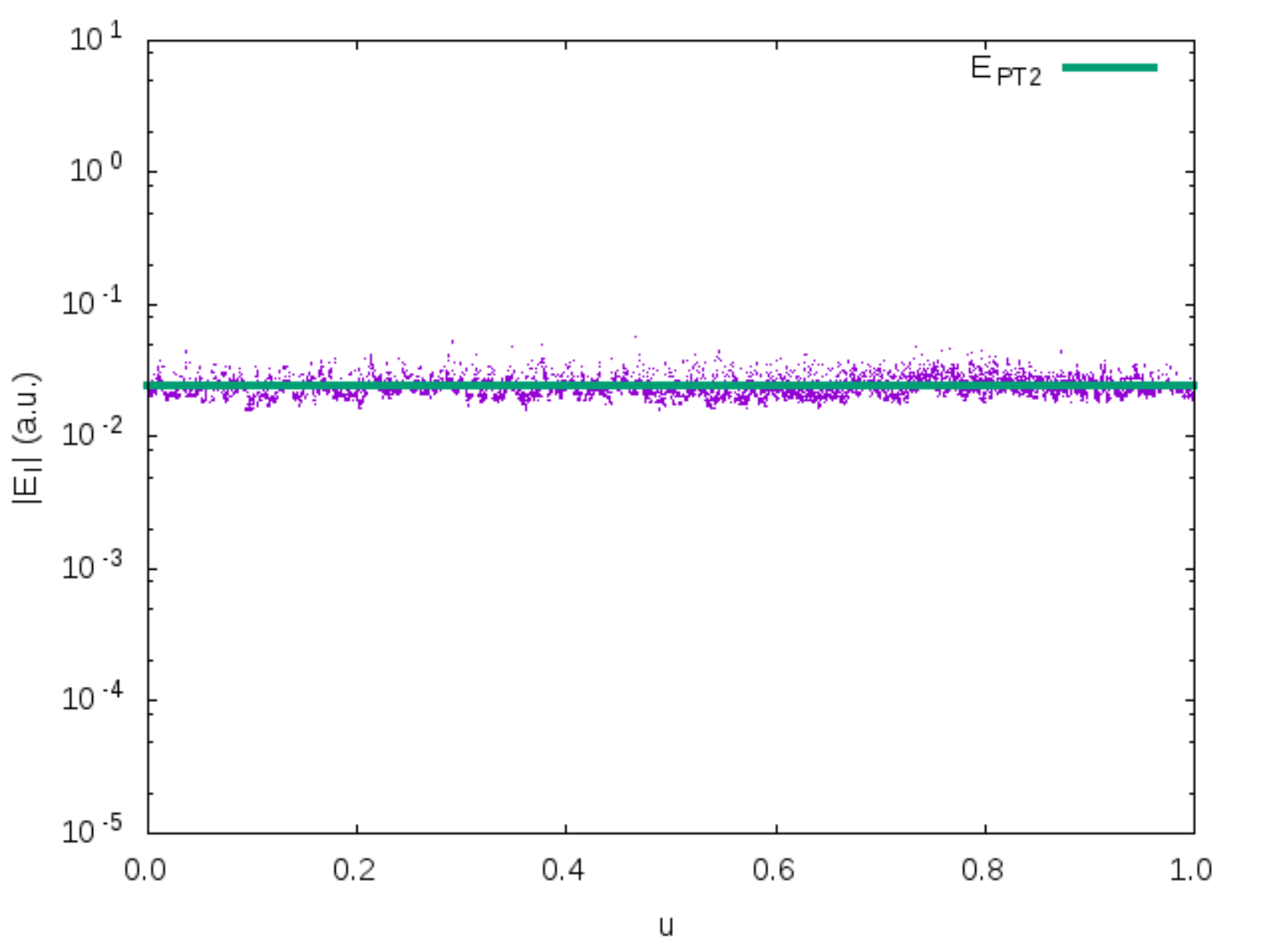


Variance reduction

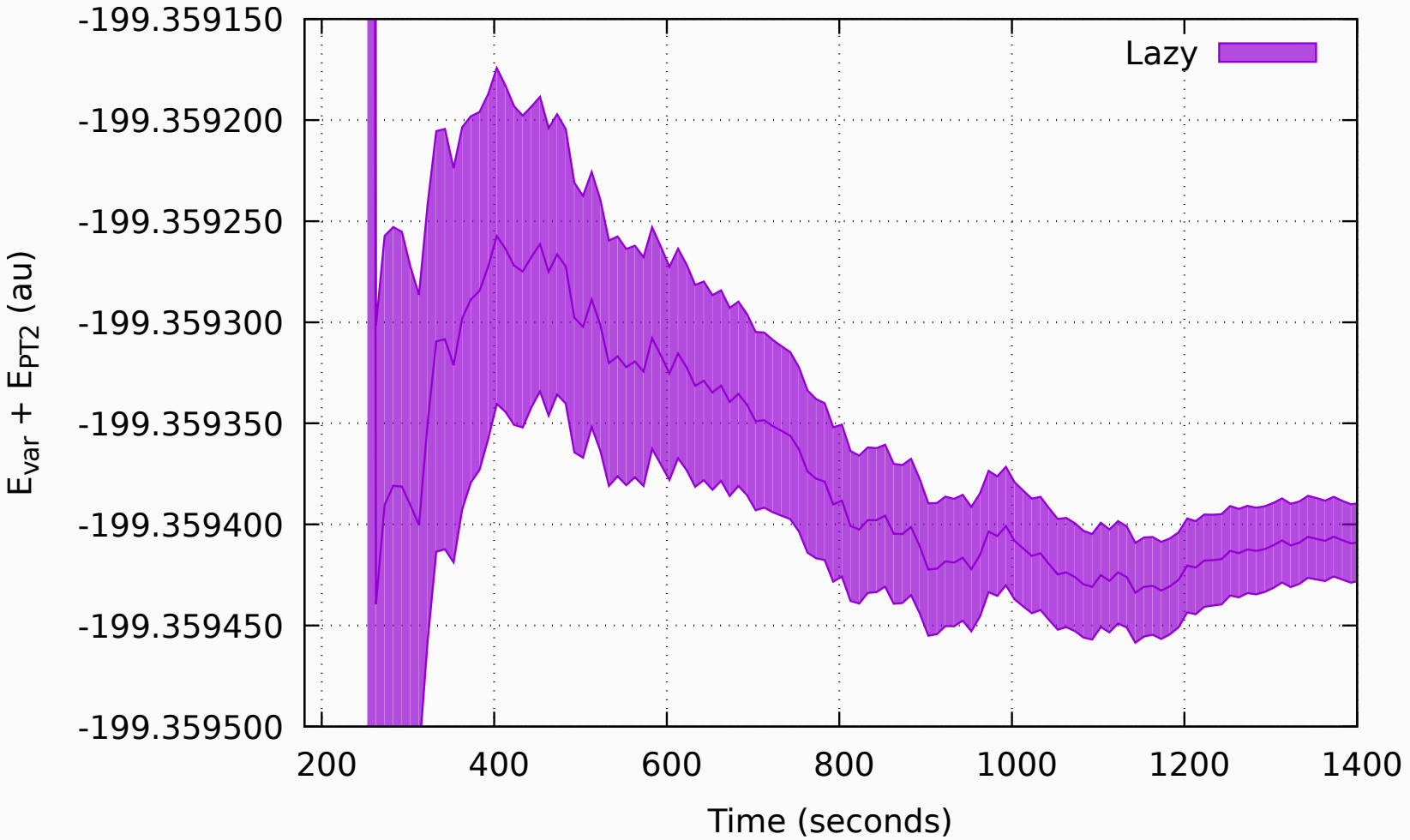


Variance reduction

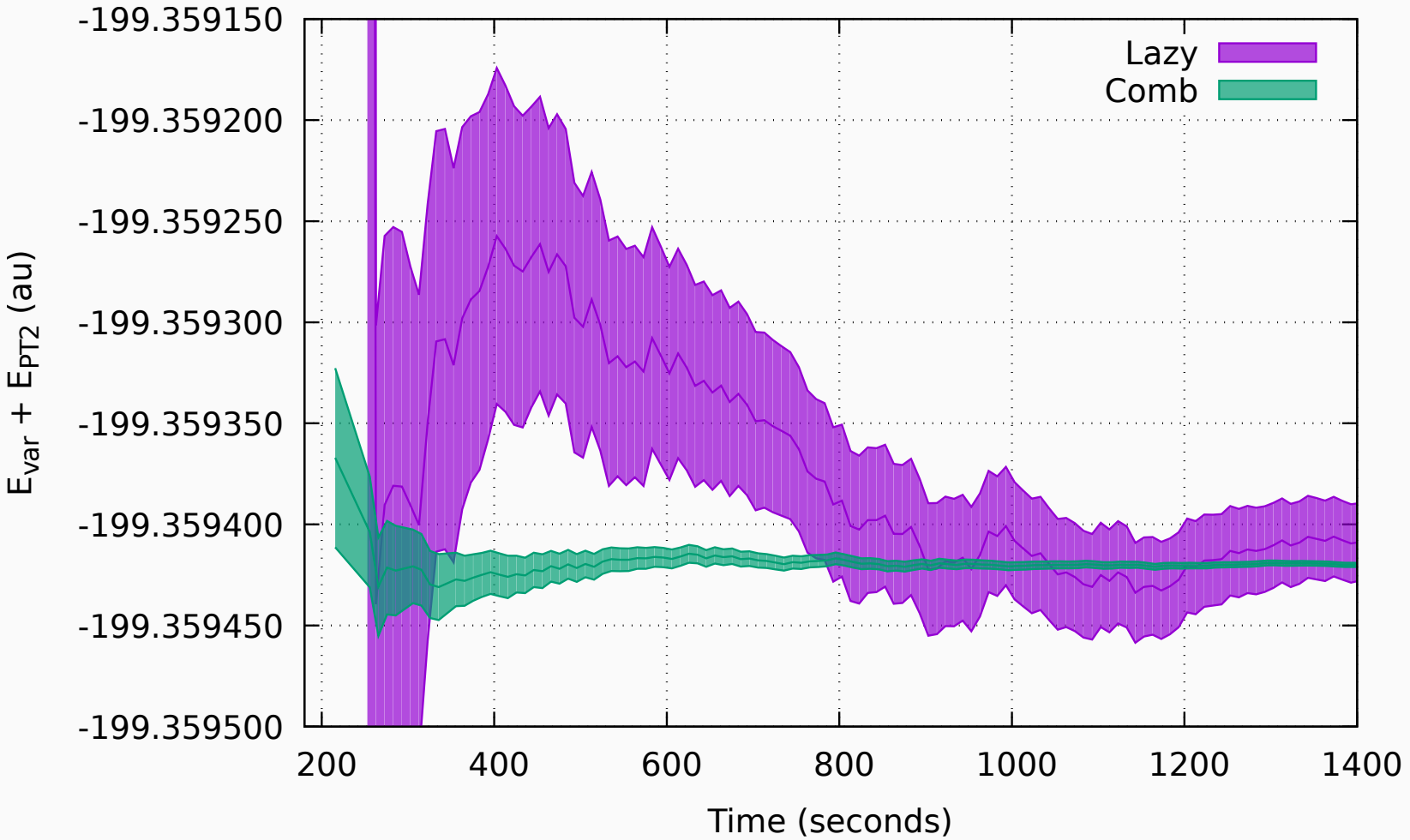
With $M = 100$



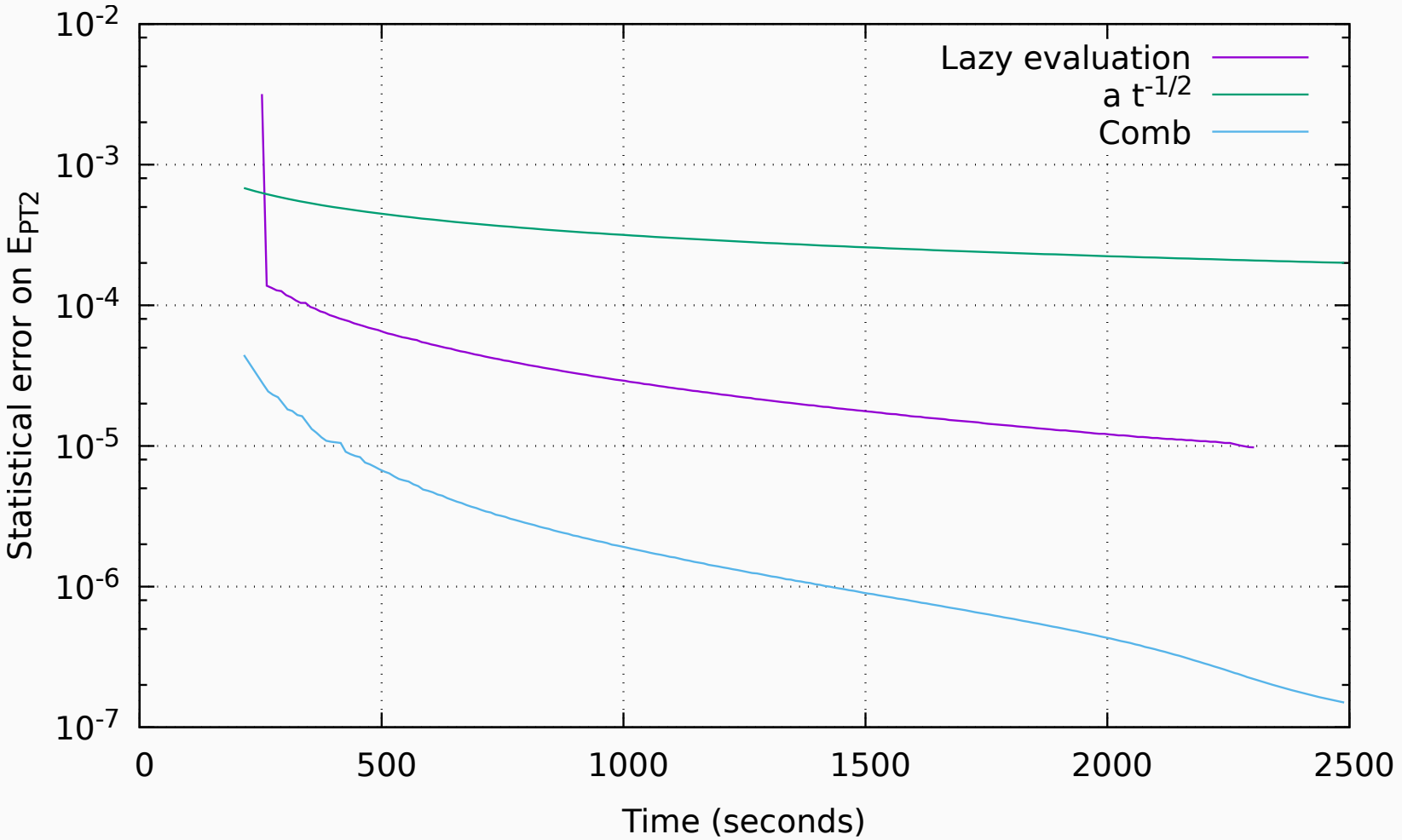
Variance reduction



Variance reduction



Variance reduction



Hybrid deterministic/stochastic scheme

- When all the determinants have been drawn, the *exact* E_{PT2} can be computed
- $\Longrightarrow$ 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 ϵ_I, and perform the MC in the rest

$$E_{\text{PT2}} = \sum_{I \in \mathcal{D}_D} \epsilon_I + \left\langle \frac{1}{M} \sum_{k=1}^{M} \frac{\epsilon_{I_k}}{p_{I_k}} \right\rangle_{(p_{I \in \mathcal{D}_S})}$$

Hybrid deterministic/stochastic scheme

Make the deterministic part grow during the calculation.

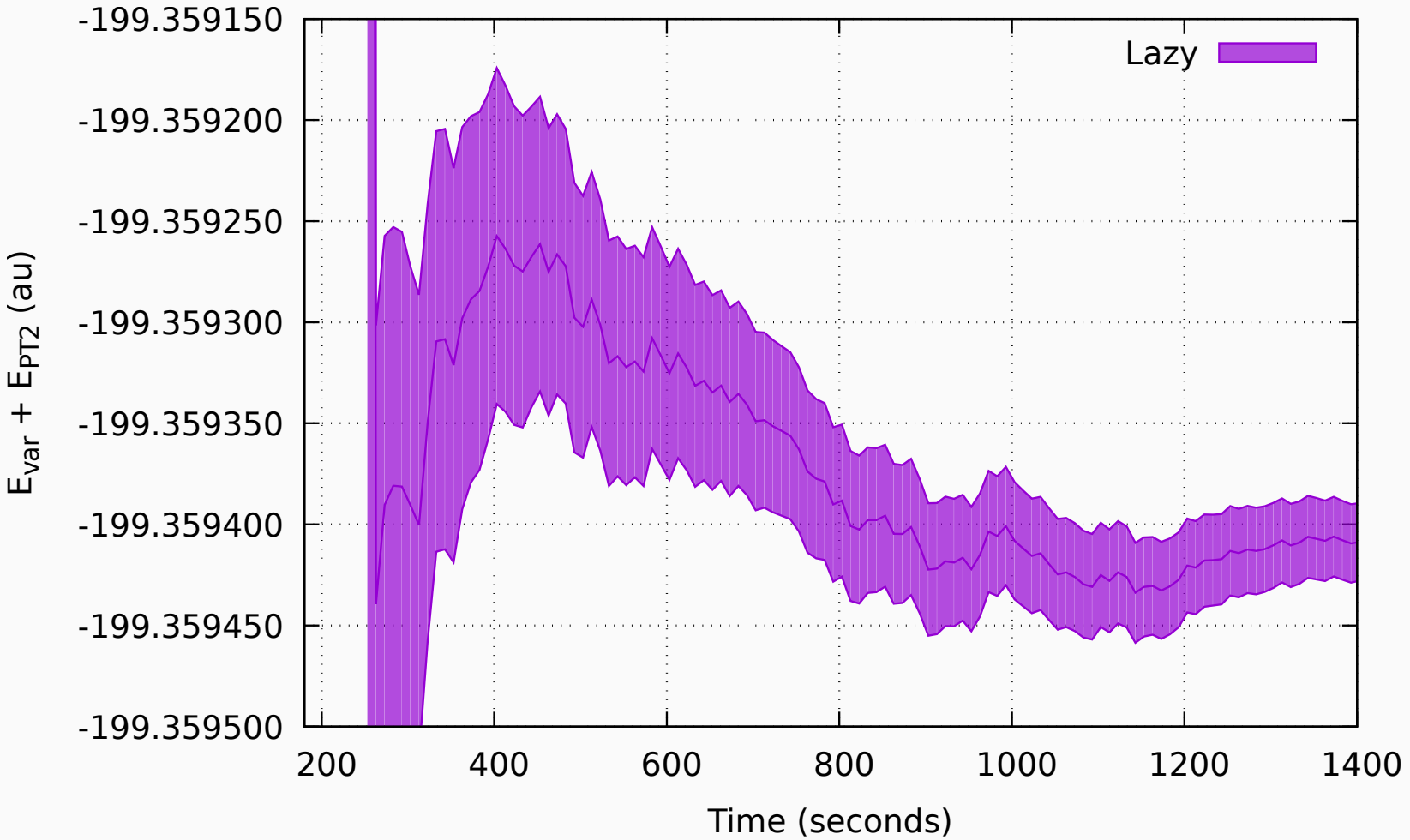
At each MC step:

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

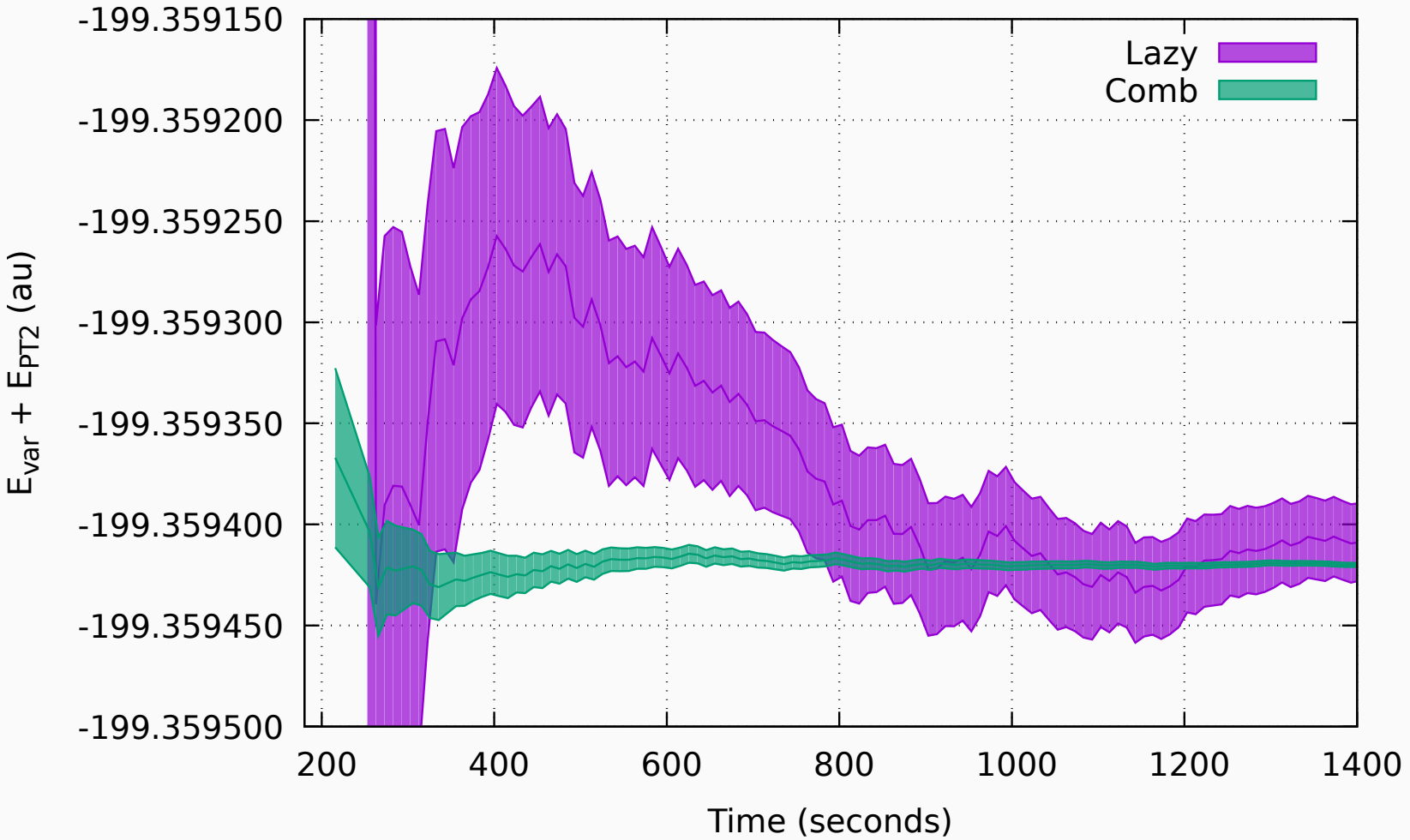
At any time:

$$E_{\text{PT2}}(t) = \sum_{I \in \mathcal{D}_D(t)} \epsilon_I + \sum_{I \in \mathcal{D}_S(t)} \frac{1}{M(t)} \frac{n_I(t)}{N_{\text{samples}}(t)} \frac{\epsilon_I}{p_I}$$

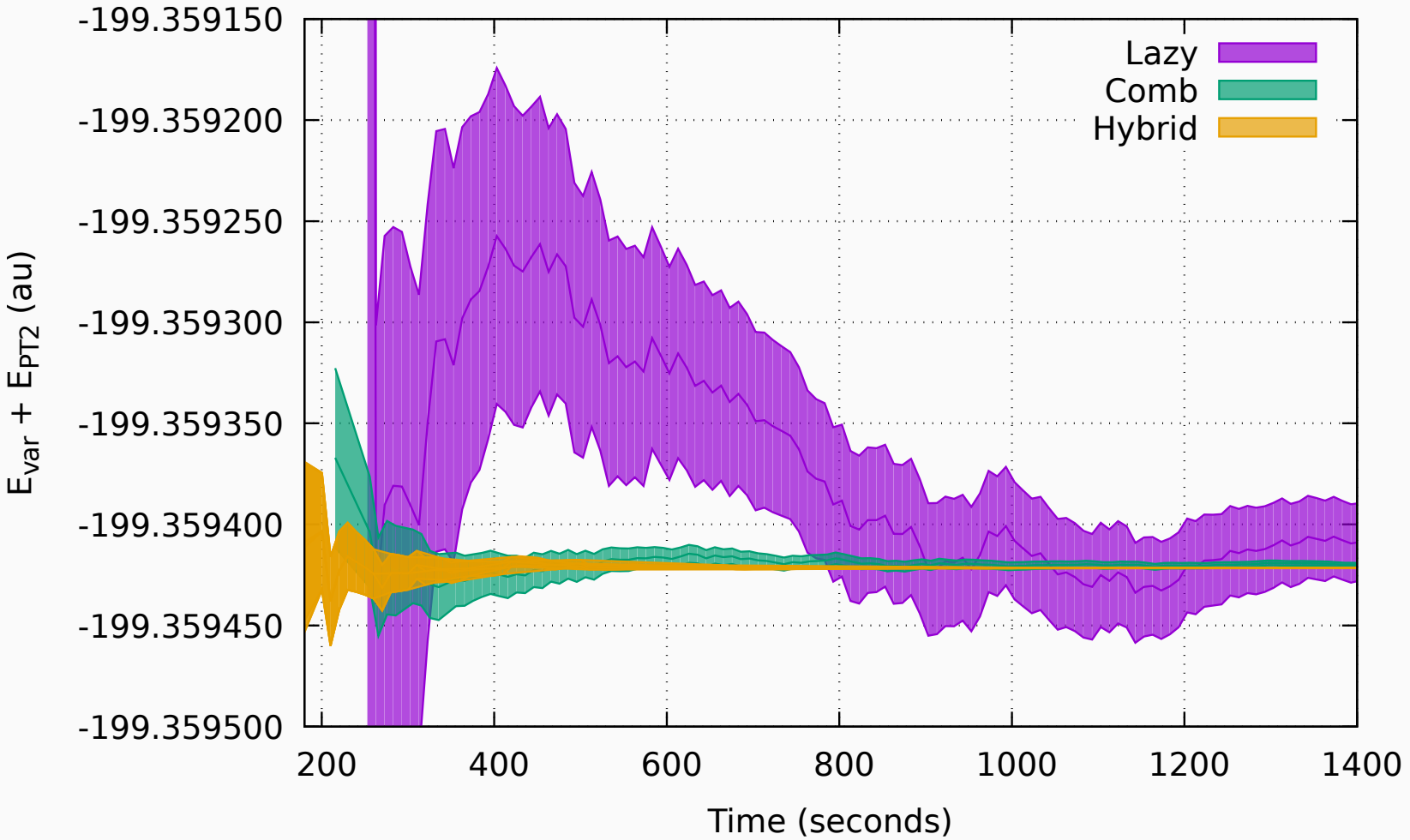
Hybrid deterministic/stochastic scheme



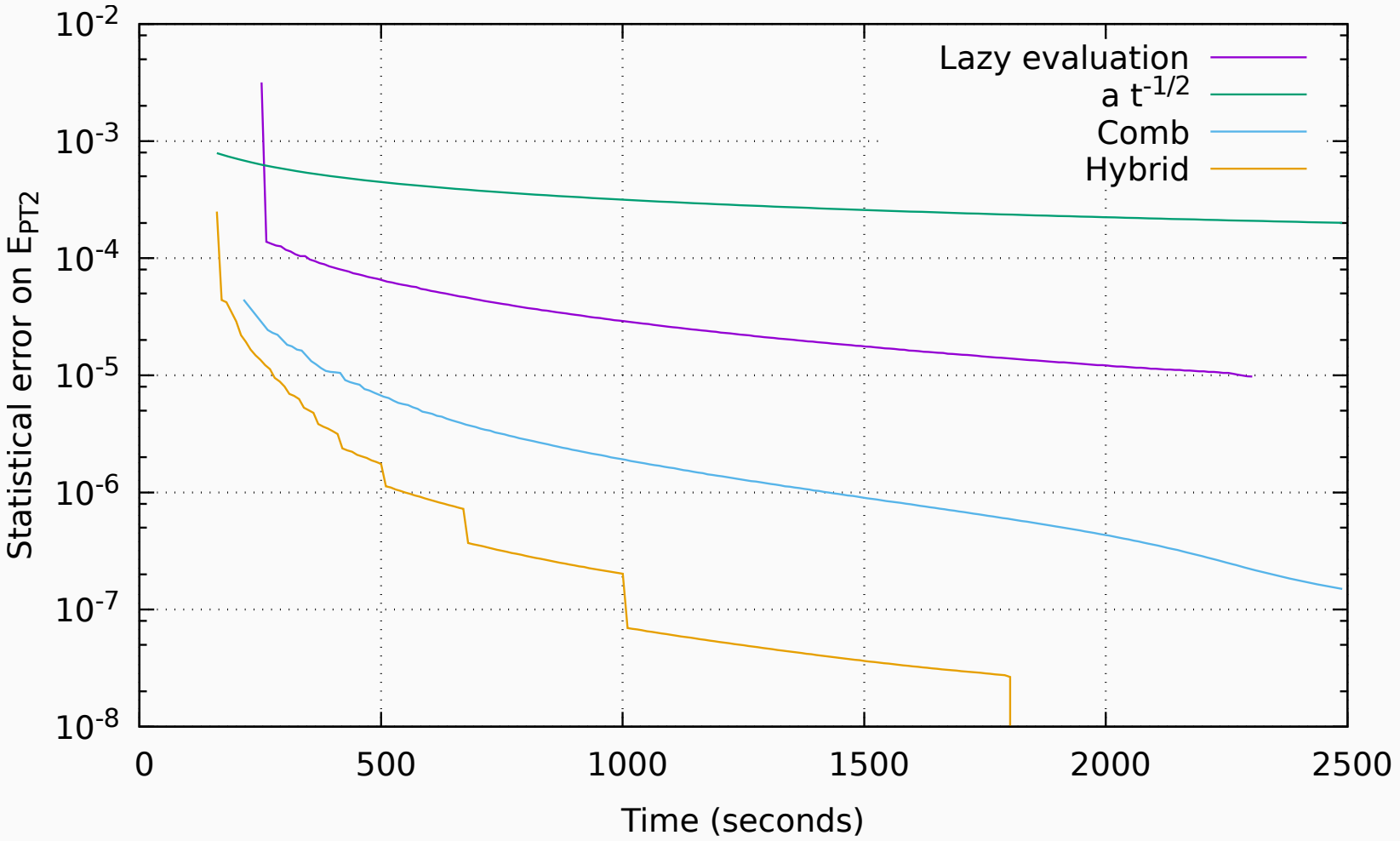
Hybrid deterministic/stochastic scheme



Hybrid deterministic/stochastic scheme



Hybrid deterministic/stochastic scheme



A few results

F_2

- 16 correlated electrons
- RHF MOs
- cc-pVDZ : 28 MOs
- cc-pVTZ : 60 MOs
- cc-pVQZ : 110 MOs
- Empirical rule: $|E_{\mathrm{FCI}} - E_{\mathrm{PT2}}| < |E_{\mathrm{PT2}} - E_{\mathrm{var}}|/5$

	N_{det}	E_{var}	$E_{\mathrm{var}} + E_{\mathrm{PT2}}$	E_{PT2}
cc-pVDZ	$2\,10^6$	$-199.098\,015$	$-199.099\,412$	$-0.001\,397$
			$-199.099\,41(9)$	
cc-pVTZ	$2\,10^6$	$-199.286\,288$	$-199.298\,119(1)$	$-0.011\,831(1)$
			$-199.297\,7(1)$	
cc-pVQZ	$1\,10^7$	$-199.349\,290$	$-199.361\,355(1)$	$-0.012\,065(1)$
			$-199.359\,8(2)$	

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 : $2\,10^7$ determinants

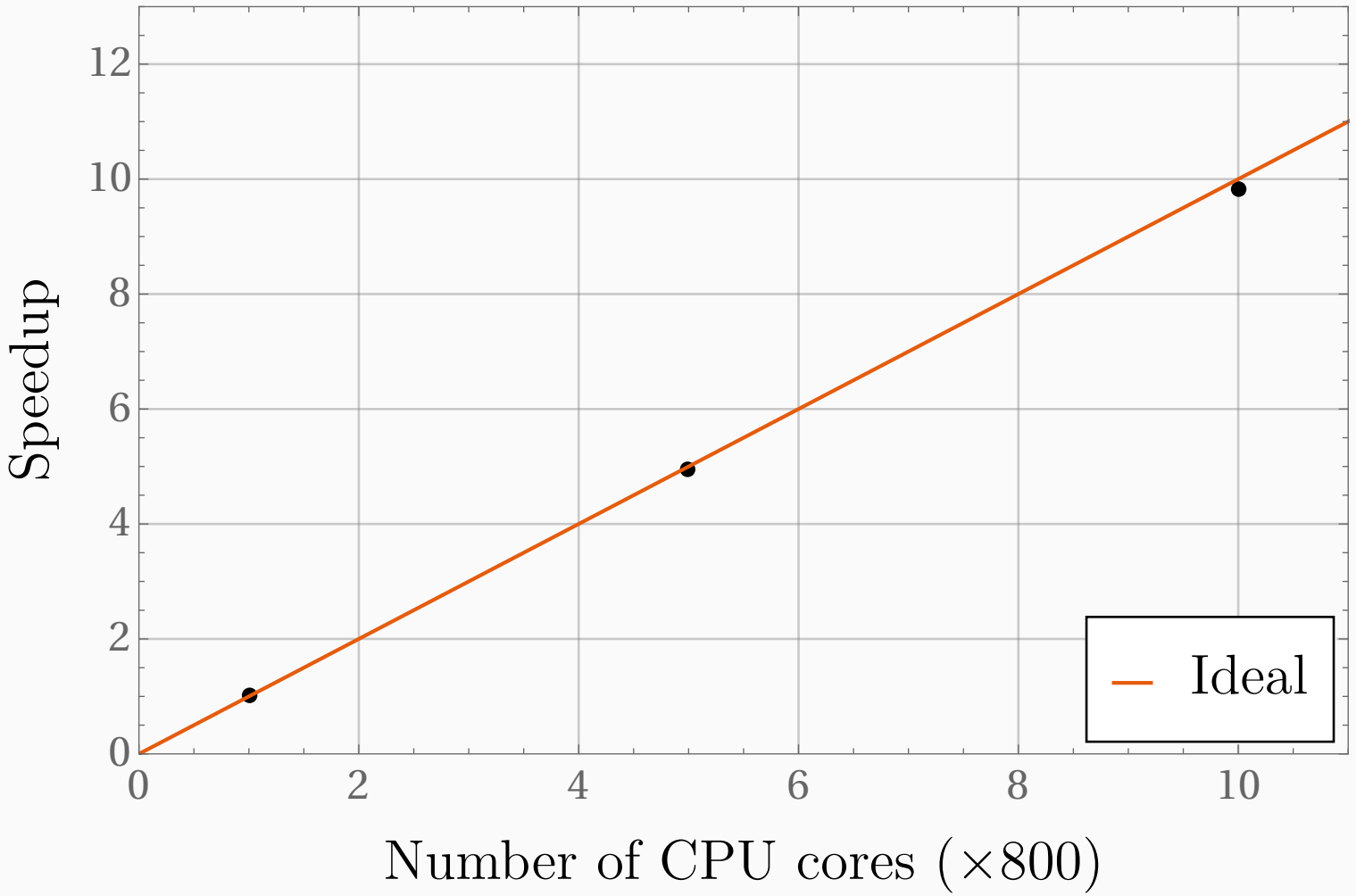
Cr_2, $2\,10^7$ determinants, 800 CPU cores

Basis	E_{PT2}	Wall-clock time
cc-pVDZ	$-0.068\,3(1)$	14 min
	$-0.068\,36(1)$	55 min
	$-0.068\,361(1)$	2.4 hr
	$-0.068\,360\,604$	3 hr
cc-pVTZ	$-0.124\,4(5)$	19 min
	$-0.124\,7(1)$	58 min
	$-0.124\,63(1)$	3.5 hr
	$-0.124\,642(1)$	8.7 hr
	—	$\sim$ 15 hr (estimated)
cc-pVQZ	$-0.155\,8(5)$	56 min
	$-0.155\,9(1)$	2.5 hr
	$-0.155\,95(1)$	9.0 hr
	$-0.155\,952(1)$	18.5 hr
	—	$\sim$ 29 hr (estimated)

Cr_2

	N_{det}	E_{var}	$E_{var} + E_{PT2}$	E_{PT2}
cc-pVDZ				
CAS	$3\,10^5$	$-2086.650\,896$	$-2087.287\,47(5\ )$	$-0.636\,57$
MR-CISD	$2\,10^6$	$-2087.156\,659$	$-2087.273\,230(4)$	$-0.116\,571$
CIPSI	$6\,10^7$	$-2087.244\,972$	$-2087.298\,530$	$-0.053\,558$
cc-pVTZ				
CAS(12,12)	$5\,10^5$	$-2086.655\,594$	$-2087.588\,64(5\ )$	$-0.933\,05(5\ )$
CIPSI	$2\,10^7$	$-2087.449\,781$	$-2087.574\,423(1)$	$-0.124\,676(1)$
cc-pVQZ				
CAS(12,12)	$6\,10^5$	$-2086.643\,210$	$-2087.690\,55(4\ )$	$-1.047\,34(4\ )$
CIPSI	$2\,10^7$	$-2087.513\,373$	$-2087.669\,330(1)$	$-0.155\,957(1)$

Parallel efficiency



Cr_2/cc-pVQZ, $2\,10^7$ determinants

Perspectives

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

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 $\rightarrow$ LCT)
- Yann Garniron (Ph.D LCPQ)
- Thomas Applencourt (Ph.D LCPQ $\rightarrow$ post-doc Argonne)
- Jean-Paul Malrieu (LCPQ)
- Pierre-François Loos (LCPQ)

References

- Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, "Hybrid stochastic-deterministic calculation of the second-order perturbative contribution of multireference perturbation theory," *arXiv:1703.05347 [physics.chem-ph]*, 2017.
- A. Scemama and E. Giner, "An efficient implementation of Slater-Condon rules," *arXiv:1311.6244 [physics.comp-ph]*, 2013.
- A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, "Quantum Monte Carlo with very large multideterminant wavefunctions," *J. Comput. Chem.*, vol. 37, pp. 1866–1875, jun 2016.
- E. Giner, C. Angeli, Y. Garniron, A. Scemama, and J.-P. Malrieu, "A Jeziorski-Monkhorst fully uncontracted multi-reference perturbative treatment I: principles, second-order versions and tests on ground state potential energy curves," *ArXiv e-prints*, p. 1702.03133, Feb. 2017.