Correlated sampling without reweighting, computing properties with size-independent variances

Roland Assaraf

Laboratoire de Chimie Théorique, CNRS-UMR 7616, Université Pierre et Marie Curie Paris VI, Case 137, 4, place Jussieu 75252 PARIS Cedex 05, France

BIRS 2012

Some perspective on Quantum Monte Carlo (QMC)

Many problem in Quantum physics at zero temperature

The Schroedinger equation,

$$H\Phi = (-\sum_{i=1}^{N} \Delta_i + V(\mathbf{r_1}, \mathbf{r_2} \ldots \mathbf{r_N}))\Phi = E\Phi \quad (1)$$

- N number of particles.
- $\mathbf{r}_i$, 3 spatial coordinates of particle i.
- E lowest eigenvalue, the groundstate energy.
- $\Phi(\mathbf{r}_1 \ldots \mathbf{r}_N)$ the lowest eigen vector, the groundstate
- Φ antisymetric for electrons (fermions).

Stochastic technics in principle adapted for solving the Schroedinger equation :

Solving the many problem in Quantum Physics

$<=>$

Computing integrals in large dimensions.

Example : variational energy

Variational energy

$$E_V \equiv \langle \Psi | \hat{H} | \Psi \rangle$$

Average on a probability distribution

$$\langle \Psi | \hat{H} | \Psi \rangle = \int d\mathbf{R} \Psi^2(\mathbf{R}) \ \frac{H\Psi}{\Psi}(\mathbf{R})$$

$$= \left\langle \frac{H\Psi}{\Psi}(\mathbf{R}) \right\rangle_{\Psi^2} = \left\langle e(\mathbf{R}) \right\rangle_{\Psi^2}$$

$\mathbf{R}$: 3N coordinates of the N interacting particles

$$E_v = \frac{1}{N} \sum^{N} e(\mathbf{R}_k)$$

Example : variational energy

Variational energy

$$E_V \equiv \langle \Psi | \hat{H} | \Psi \rangle$$

Average on a probability distribution

$$\langle \Psi | \hat{H} | \Psi \rangle = \int d\mathbf{R} \Psi^2(\mathbf{R}) \ \frac{H\Psi}{\Psi}(\mathbf{R})$$

$$= \left\langle \frac{H\Psi}{\Psi}(\mathbf{R}) \right\rangle_{\Psi^2} = \left\langle e(\mathbf{R}) \right\rangle_{\Psi^2}$$

$\mathbf{R}$: 3N coordinates of the N interacting particles

$$E_v = \frac{1}{N} \sum^{N} e(\mathbf{R}_k)$$

Example : variational energy

Variational energy

$$E_V \equiv \langle \Psi | \hat{H} | \Psi \rangle$$

Average on a probability distribution

$$\langle \Psi | \hat{H} | \Psi \rangle = \int d\mathbf{R} \Psi^2(\mathbf{R}) \ \frac{H\Psi}{\Psi}(\mathbf{R})$$

$$= \left\langle \frac{H\Psi}{\Psi}(\mathbf{R}) \right\rangle_{\Psi^2} = \left\langle e(\mathbf{R}) \right\rangle_{\Psi^2}$$

$\mathbf{R}$: 3N coordinates of the N interacting particles

$$E_v = \frac{1}{N} \sum^{N} e(\mathbf{R}_k)$$

In general

More generally

$$E_{\text{QMC}} = \left\langle e(\mathbf{R}) \right\rangle_{\Pi}$$

Depending on the QMC method, the nature of $\mathbf{R}$ might change :

- $3N$ particle coordinates (VMC,DMC..).
- Trajectories in the space of $3N$ particle coordinates (PDMC, PIMC, reptation...)

Accurate energies

No analytical integration

- Flexibility (choice of ψ in VMC).
- Weak limitation in system sizes.
- Possibility to improve “arbitrarily” the accuracy (“zero-variance zero-bias principle”, choice of ψ in VMC..).

In practice, reference methods for total energies on large systems (large N)

Quantites of physical interest

They are energy differences.

Exploiting accurate total energies

More tricky in QMC than in a deterministic method.

- Energy differences are usually very small.
- Statistical uncertainties.

Small statistical uncertainty on a total energy might be huge on a difference if energies are computed independently.

Why energy differences are usually small ?

Examples

- Binding energies, transition state energies. One, two particle gaps (electron affinities, ionization energies) . . .
- First order derivatives of the energy : Any observable (force, dipole, moment, densities...).
- Higher order derivatives : spectroscopic constants . . .

They are groundstate energies of similar systems

Paradigm : Calculation of an observable O

$$H_\lambda = H + \lambda O => \bar{O} = \frac{dE_\lambda}{d\lambda} \simeq \frac{E_\lambda - E_0}{\lambda} = \frac{\Delta_\lambda}{\lambda} \quad (2)$$

$$\Delta_\lambda = E_\lambda - E_0 \propto \lambda \text{ small}$$

Behavior as a function of the system size

$\lim_{N\to\infty} \Delta_\lambda(N) = K$ finite.

The perturbation λO depends usually on a few degrees of freedom.

Δ_λ has a locality property

In summary

Small λ and large N

$$\Delta_\lambda(N) \propto \lambda$$

Accuracy on Δ_λ in an independent energy calculation

$$\frac{\delta\Delta_\lambda}{\Delta_\lambda} \propto \frac{\delta E_0}{\lambda} \propto \frac{\sqrt{N}}{\lambda}$$

No locality property for the statistical uncertainty.

Comparison to total energy

$$\frac{\delta\Delta_\lambda}{\Delta_\lambda} \propto \frac{N^{\frac{3}{2}}}{\lambda}\frac{\delta E}{E}$$

Overview

1. Introduction
2. Correlated sampling with reweighting
   - The method
   - Statistical uncertainties
   - Numerical illustration
3. Correlated sampling with no reweighting
   - The method
   - Numerical illustration
4. Conclusion and perspectives

correlated sampling with reweighting

We have to compute the difference

$$E_\lambda - E_0 = \langle e_\lambda(\mathbf{R}) \rangle_{\pi_\lambda} - \langle e(\mathbf{R}) \rangle_\pi$$

Sampling the same distribution for the two energies

$$E_\lambda - E_0 = \frac{\left\langle e_\lambda \frac{\pi_\lambda}{\pi} \right\rangle_\pi}{\left\langle \frac{\pi_\lambda}{\pi} \right\rangle_\pi} - \langle e \rangle_\pi . \qquad (3)$$

weight w_λ

Different contexts

- Variational Monte Carlo $e_\lambda(\mathbf{R}) = \frac{H_\lambda \psi_\lambda}{\psi_\lambda}(\mathbf{R})$, $w_\lambda(\mathbf{R}) = \frac{\psi_\lambda^2}{\psi^2}(\mathbf{R})$
- Forward walking method inc ontext of DMC algorithms.
- ...

correlated sampling with reweighting

We have to compute the difference

$$E_\lambda - E_0 = \langle e_\lambda(\mathbf{R}) \rangle_{\pi_\lambda} - \langle e(\mathbf{R}) \rangle_\pi$$

Sampling the same distribution for the two energies

$$E_\lambda - E_0 = \frac{\left\langle e_\lambda \frac{\pi_\lambda}{\pi} \right\rangle_\pi}{\left\langle \frac{\pi_\lambda}{\pi} \right\rangle_\pi} - \langle e \rangle_\pi . \qquad (3)$$

weight w_λ

Different contexts

- Variational Monte Carlo $e_\lambda(\mathbf{R}) = \frac{H_\lambda \psi_\lambda}{\psi_\lambda}(\mathbf{R})$, $w_\lambda(\mathbf{R}) = \frac{\psi_\lambda^2}{\psi^2}(\mathbf{R})$
- Forward walking method inc ontext of DMC algorithms.
- ...

correlated sampling with reweighting

We have to compute the difference

$$E_\lambda - E_0 = \langle e_\lambda(\mathbf{R}) \rangle_{\pi_\lambda} - \langle e(\mathbf{R}) \rangle_\pi$$

Sampling the same distribution for the two energies

$$E_\lambda - E_0 = \frac{\left\langle e_\lambda \frac{\pi_\lambda}{\pi} \right\rangle_\pi}{\left\langle \frac{\pi_\lambda}{\pi} \right\rangle_\pi} - \langle e \rangle_\pi . \qquad (3)$$

weight w_λ

Different contexts

- Variational Monte Carlo $e_\lambda(\mathbf{R}) = \frac{H_\lambda \psi_\lambda}{\psi_\lambda}(\mathbf{R})$, $w_\lambda(\mathbf{R}) = \frac{\psi_\lambda^2}{\psi^2}(\mathbf{R})$
- Forward walking method inc ontext of DMC algorithms.
- ...

General expression

Compact expression

$$\Delta_\lambda = E_\lambda - E_0 = \langle e_\lambda - e \rangle_\pi + \frac{\mathrm{cov}(e_\lambda, w_\lambda)}{\langle w_\lambda \rangle_\pi} \tag{4}$$

λ dependence

$$E_\lambda - E_0 = \lambda \frac{\partial E_\lambda}{d\lambda}|_{\lambda=0} + o(\lambda).$$

$$E'_\lambda = \langle e'_\lambda \rangle_\pi + \mathrm{cov}(e_\lambda, w'_\lambda)$$

Zero-Variance (ZV) estimator

Pulay correction

Finite statistical uncertainty on $E'_\lambda \Longrightarrow \frac{\delta\Delta_\lambda}{\Delta_\lambda} = K + o(\lambda)$

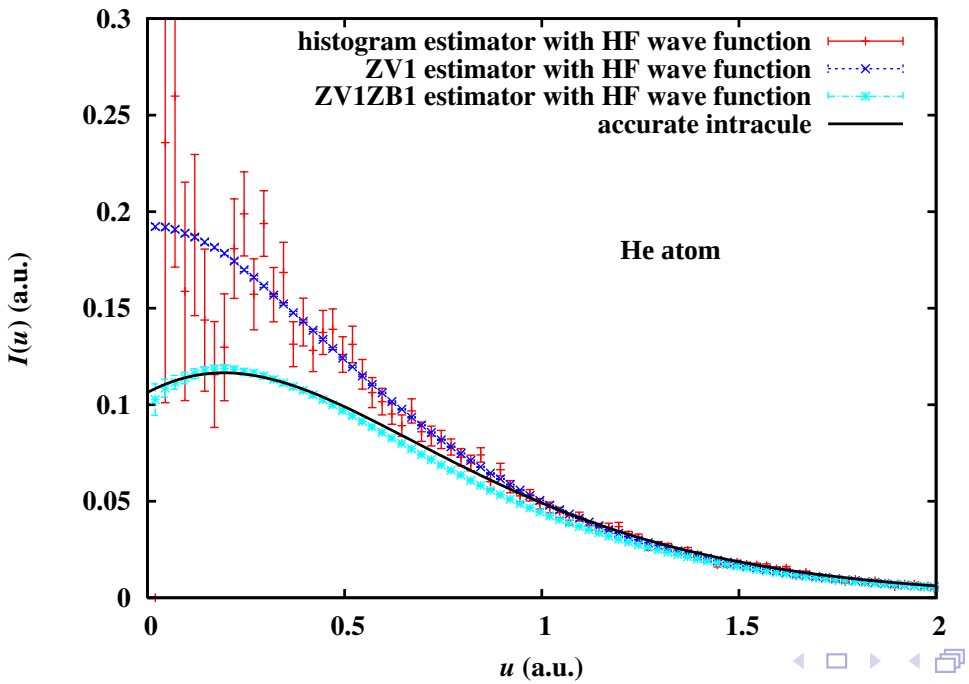
Pair correlation function

[J. Toulouse, R. Assaraf, C. J. Umrigar. J. Chem. Phys. **126** 244112 (2007)]

$O_u = \sum_{i<j} \delta(r_{ij} - u)$

Probability density to find a pair of electrons at distance u

ZV term : $\frac{dE_\lambda}{d\lambda} = \langle O_u + \frac{(H-e)\psi'}{\psi_0} \rangle_{\psi_0^2} = \langle O_u \rangle_{\psi_0^2}$



N-dependence

R. Assaraf, D. Domin, W. Lester.

Model of two separated (non interacting) subsystems

Particles coordinates $\mathbf{R}^l$ and $\mathbf{R}^u$. $H_\lambda = H^l_\lambda + H^u$

Variational Monte Carlo

- $\mathbf{R} = (\mathbf{R}^l, \mathbf{R}^u)$
- $\Psi_\lambda(\mathbf{R}) = \Psi_\lambda(\mathbf{R}^l, \mathbf{R}^u) = \Psi^l_\lambda(\mathbf{R}^l)\Psi^u(\mathbf{R}^u)$
- Local energy $e_\lambda(\mathbf{R}) = e^l_\lambda(\mathbf{R}^l) + e^u(\mathbf{R}^u)$

$$E_\lambda - E = \langle e^l_\lambda - e^l_0 \rangle + \frac{\mathrm{cov}(e_\lambda, w^l)}{\langle w^l \rangle} \qquad (5)$$

First term (ZV)

$\langle e_\lambda^l - e_0^l \rangle$ depends only on $\mathbf{R}^l$

$\Longrightarrow$ Locality property of its variance

Pulay term

$$\frac{\mathrm{cov}(e_\lambda, w^l)}{\langle w^l \rangle} = \frac{\mathrm{cov}(e^l_\lambda, w^l)}{\langle w^l \rangle} + \frac{\mathrm{cov}(e^u, w^l)}{\langle w^l \rangle} \tag{6}$$

Local

Non local

- The non local contribution is 0 (e^u and w^l independant)!
- Its variance on a finite sample $(e^u(\mathbf{R}^u_i), w^l(\mathbf{R}^l_i))_{i\in[1..M]}$:

$$\propto V(e^u) \propto N$$

.

$\Longrightarrow \delta\Delta_\lambda(N) \propto \sqrt{N}$ for large N.

Non locality property of the Pulay term.

Conclusion

$$\frac{\delta\Delta_\lambda}{\Delta_\lambda} \propto \sqrt{N} \tag{7}$$

Correlated sampling with reweighting solves the small λ difficulty but not the large N one

Illustration on H_n chains

Is the analysis for non interacting subsystems holds for interacting systems ?

- Hydrogen chains, metallic and insulating
- Calculation of the force on the first nucleus : derivative of the energy with respect to the position of the first nucleus
- Variational calculation
- ψ is a single determinant (Restricted Hartree Fock)

Metallic hydrogen chains

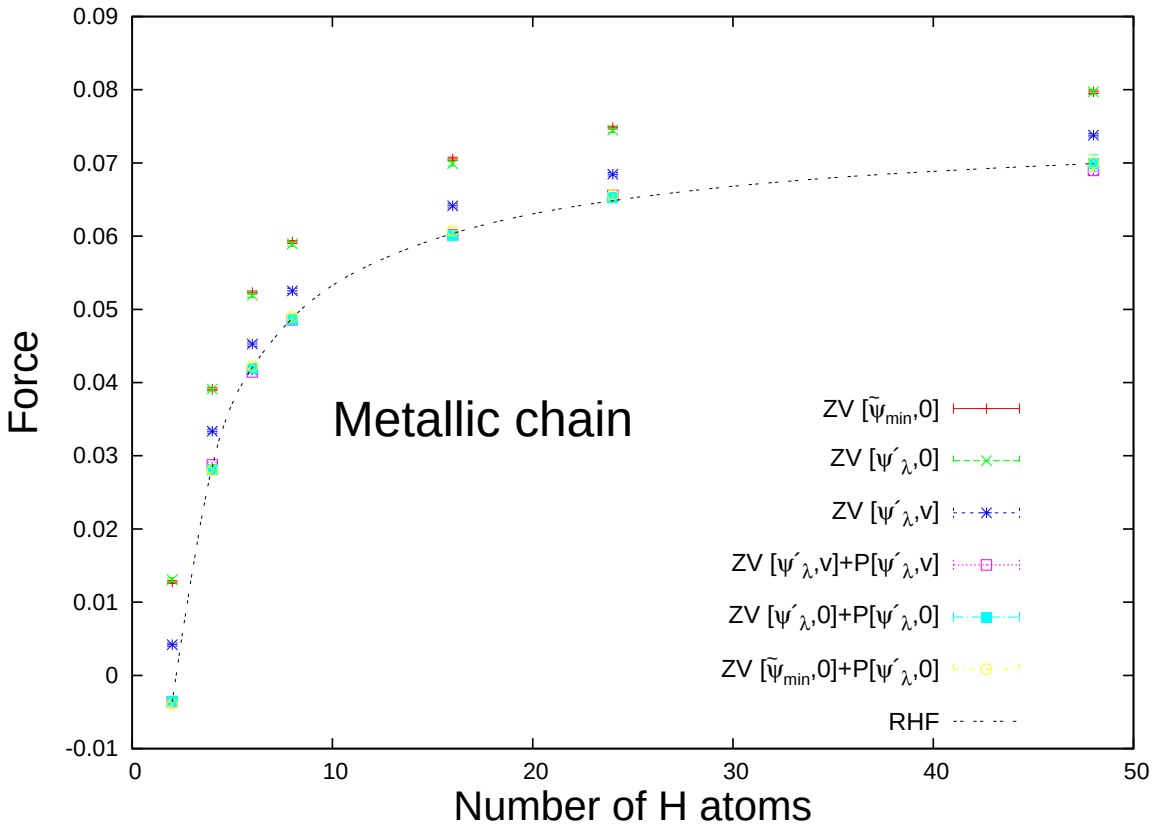


FIG.: Energy derivative, different estimators

Insulating hydrogen chains

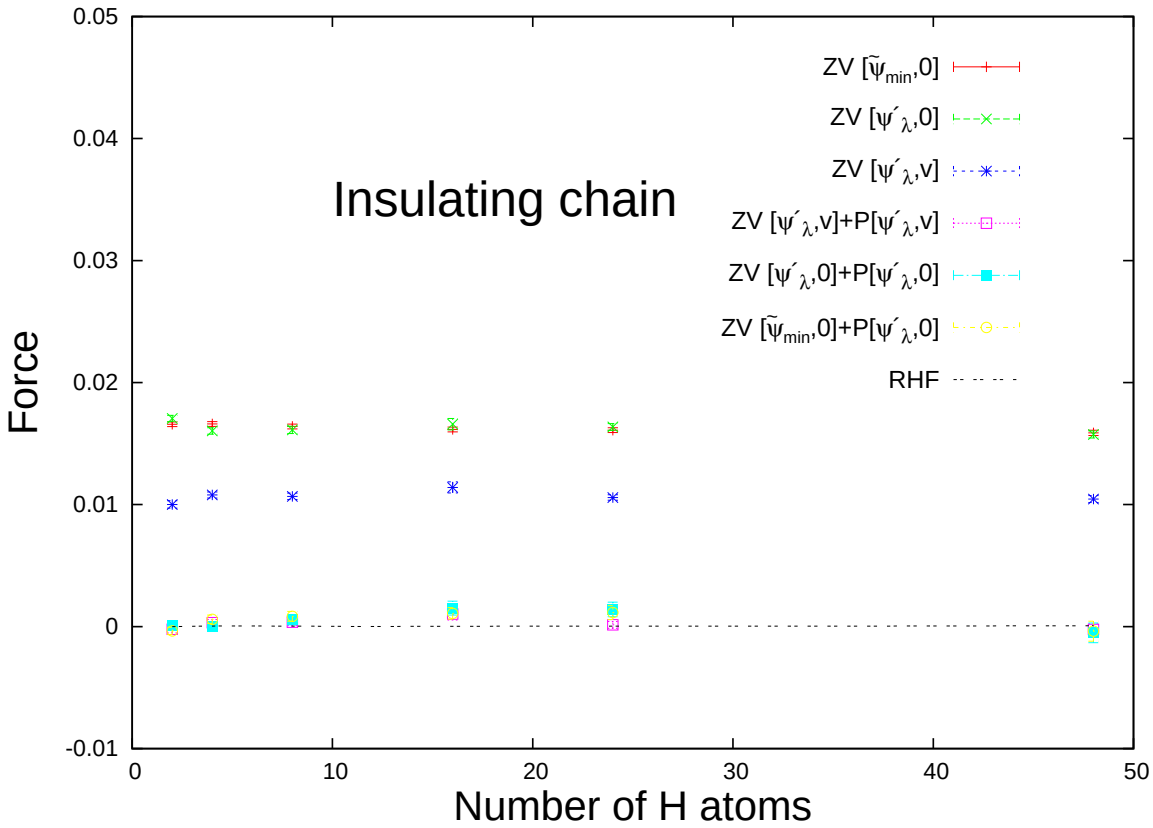


FIG.: Energy derivative, different estimators

Histogram of the ZV term, metallic chain

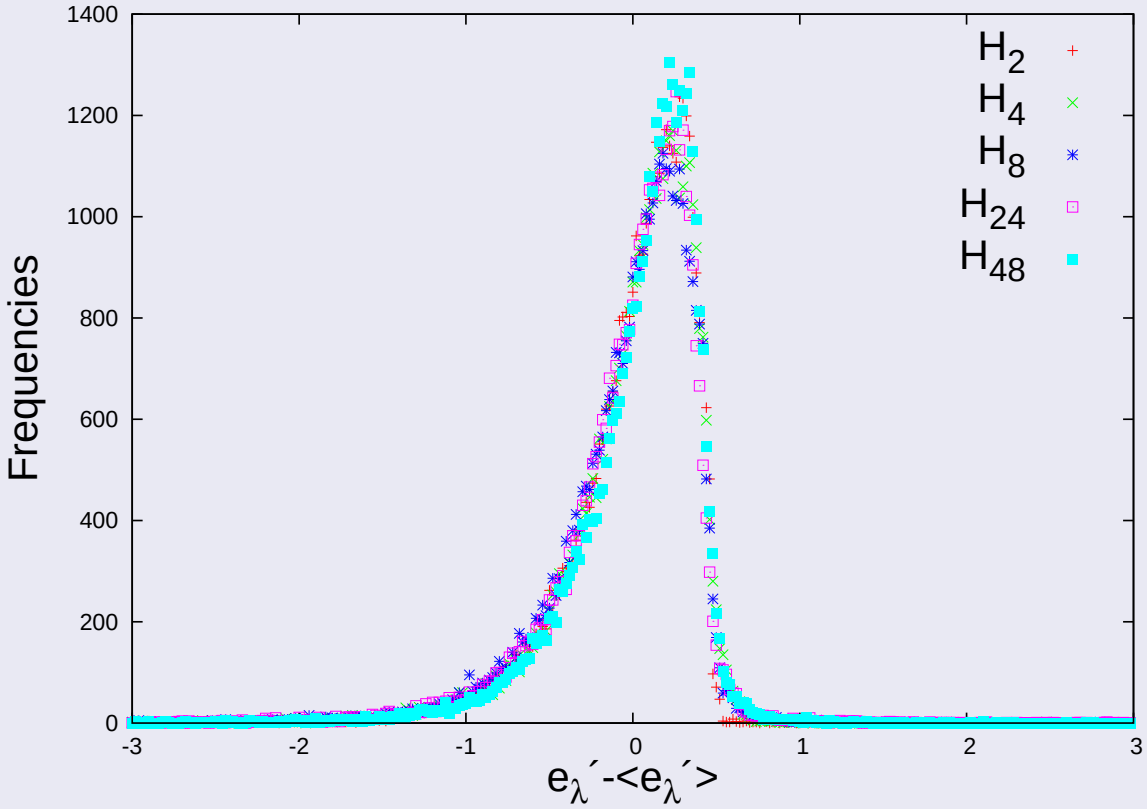


FIG.: Histogram of the energy derivative in the H_n chain

ZV contribution has the local property!!

Histogram of the local energy, metallic chain

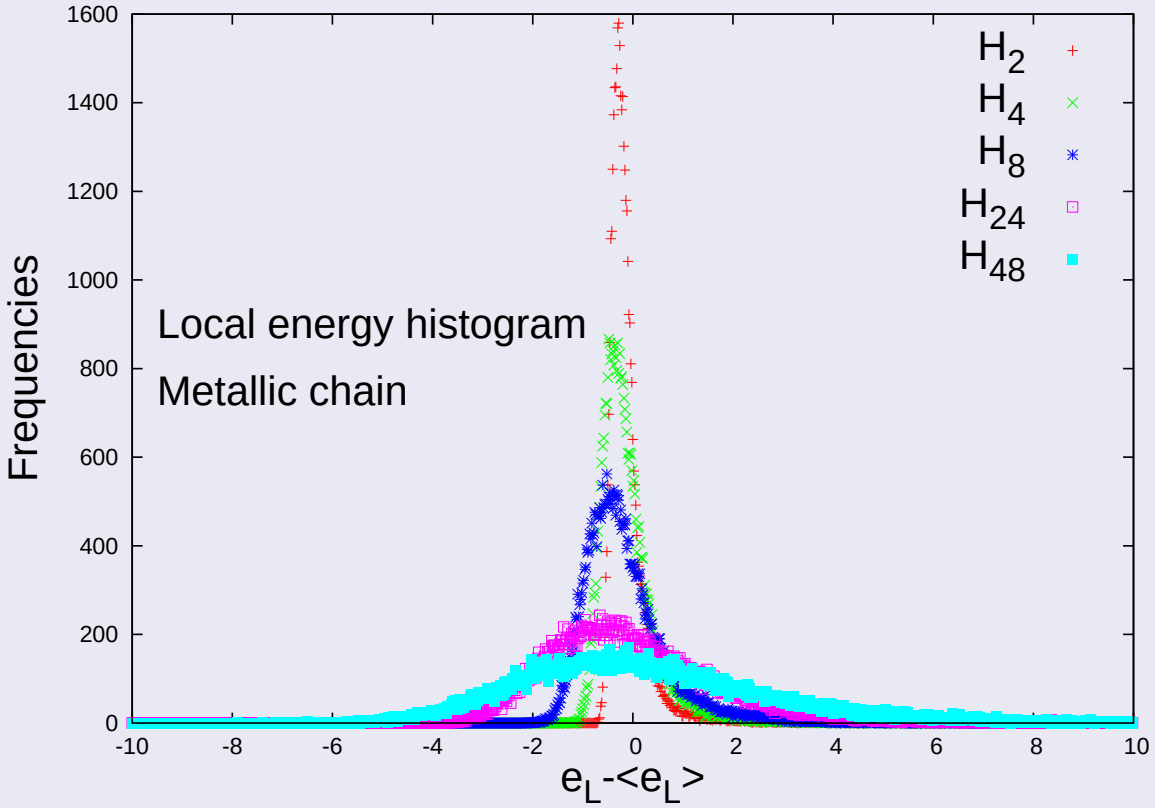


FIG.: Histogram of the local in the H_n chain

The local energy has not the local property

Statistical uncertainties

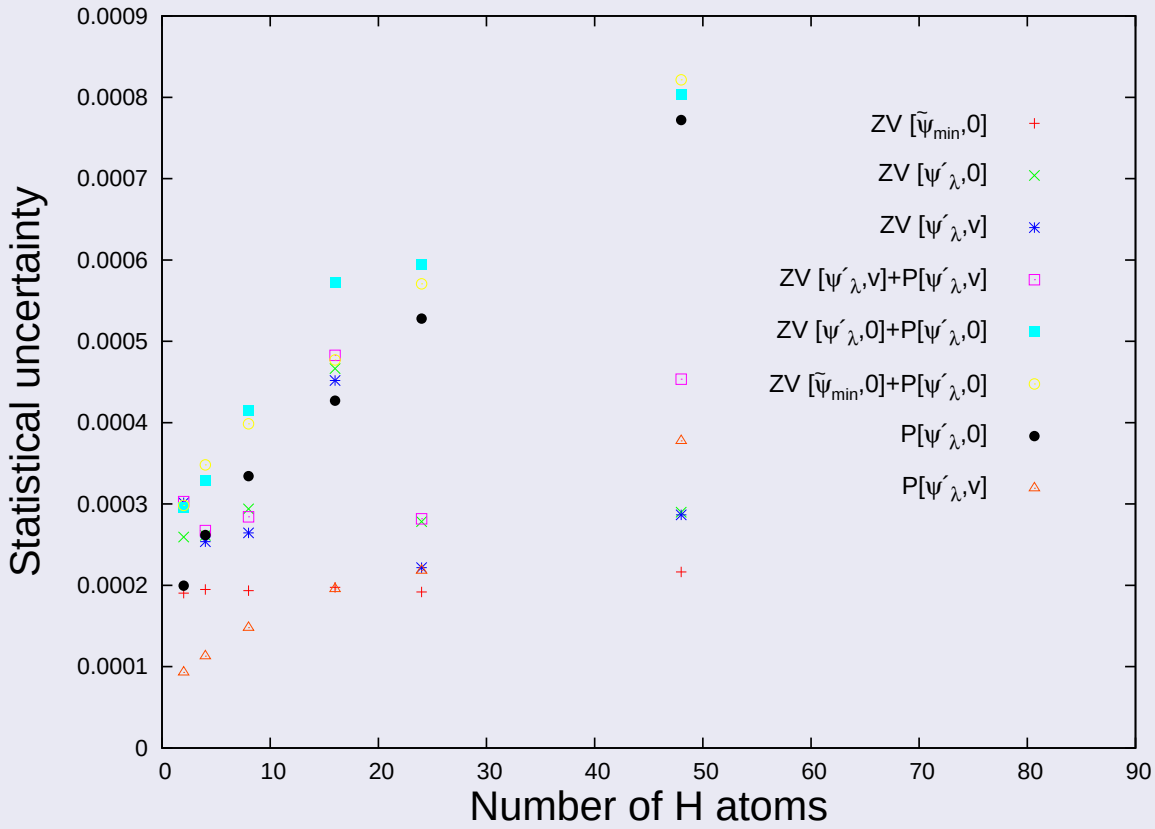


FIG.: Statistical uncertainties in the insulating H_n chain

Statistical uncertainties

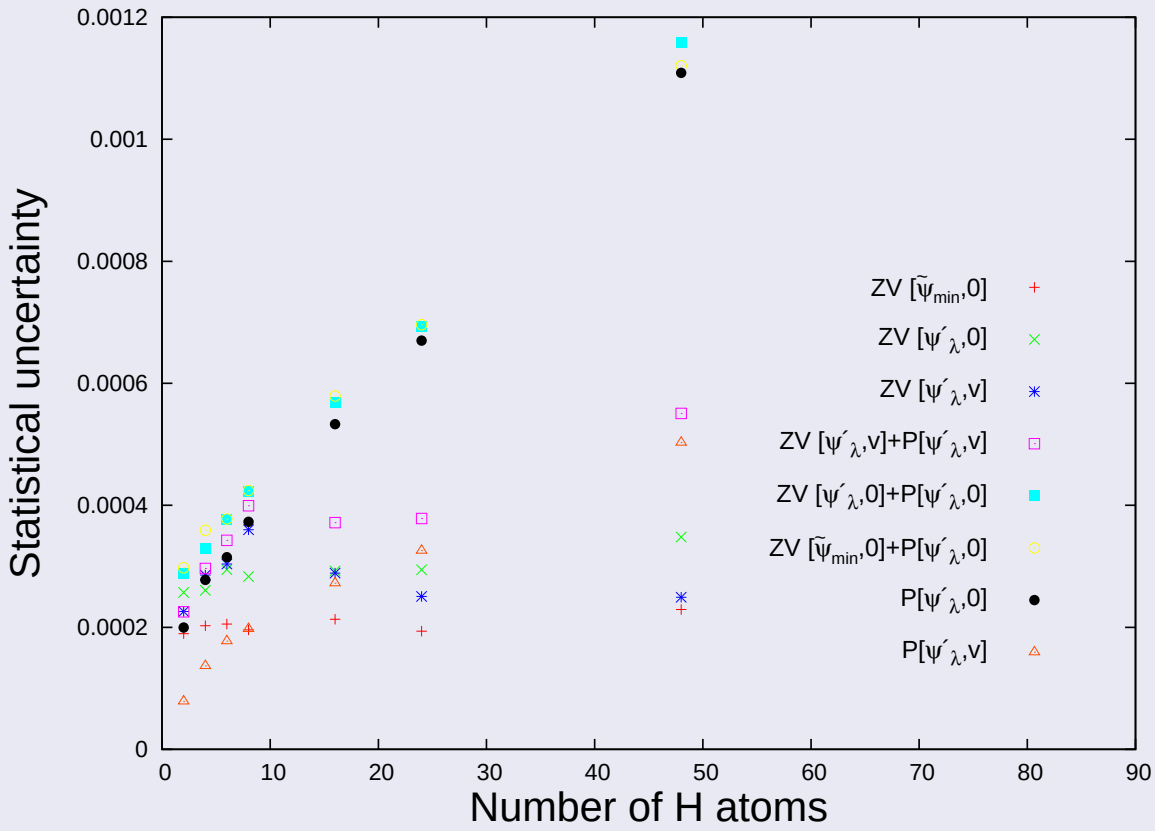


FIG.: Statistical uncertainties in the metallic H_n chain

The method

Assaraf, Caffarel, Kollias 2011

Basic idea

$$\langle e_\lambda(\mathbf{R})\rangle_{\pi_\lambda} - \langle e(\mathbf{R})\rangle_{\pi} = \langle e_\lambda(\mathbf{R}_\lambda) - e(\mathbf{R})\rangle_{\Pi(\mathbf{R},\mathbf{R}_\lambda)}$$

- Marginal distributions of $\Pi(\mathbf{R},\mathbf{R}_\lambda)$ must be $\pi(\mathbf{R}), \pi_\lambda(\mathbf{R}_\lambda)$.
- differences of the order of λ, $\langle(\mathbf{R}_\lambda) - \mathbf{R})^2\rangle = K\lambda^2$

How to build such a process

- Choosing close stochastic processe, L, L_λ having π and π_λ as stationary states.
- Stability versus chaos. Two trajectories with the different initial conditions and same pseudo random numbers meet exponentially fast.
- Insures that close processes will produce close trajectories.

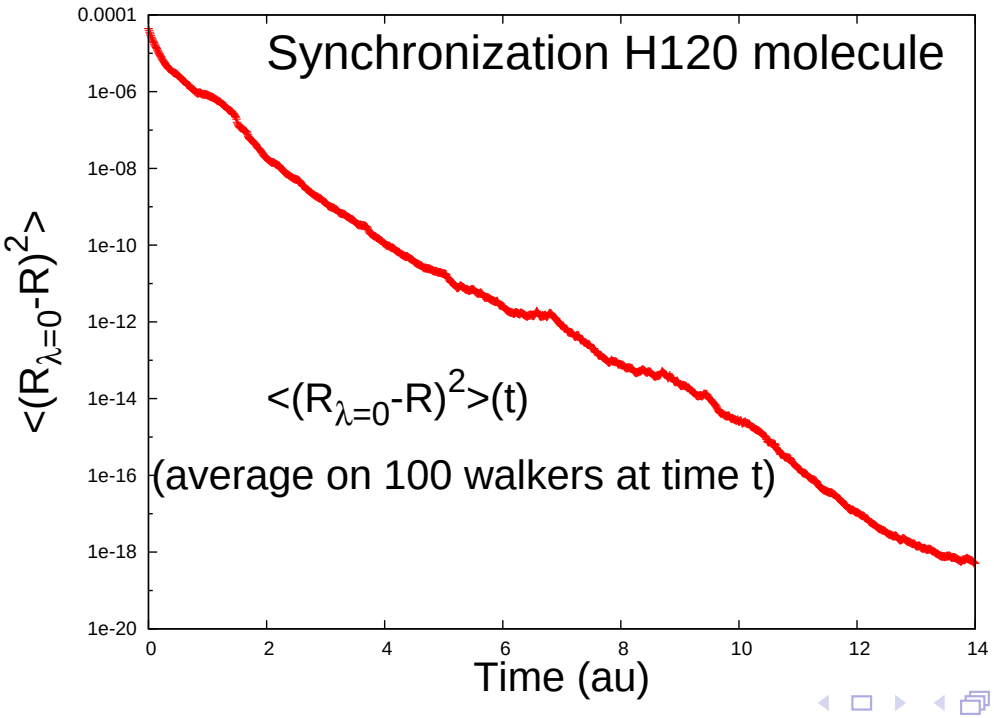
For example, with the overdamped Langevin process one would have

$$\mathbf{R}(t+dt) = \mathbf{R}(t) + \mathbf{b}\left[\mathbf{R}(t)\right] dt + \mathbf{dW} \quad (8)$$

$$\mathbf{R}_\lambda(t+dt) = \mathbf{R}_\lambda(t) + \mathbf{b}_\lambda\left[\mathbf{R}_\lambda(t)\right] dt + \mathbf{dW} \quad (9)$$

Stability of the process versus chaos

- Chain of 120 Hydrogens (120 electrons).
- Same process but different initial conditions.
- Perturbed system one atom displaced of $\lambda = 10^{-4} a.u$ (finite difference derivative).



Independance of the uncertainties on λ

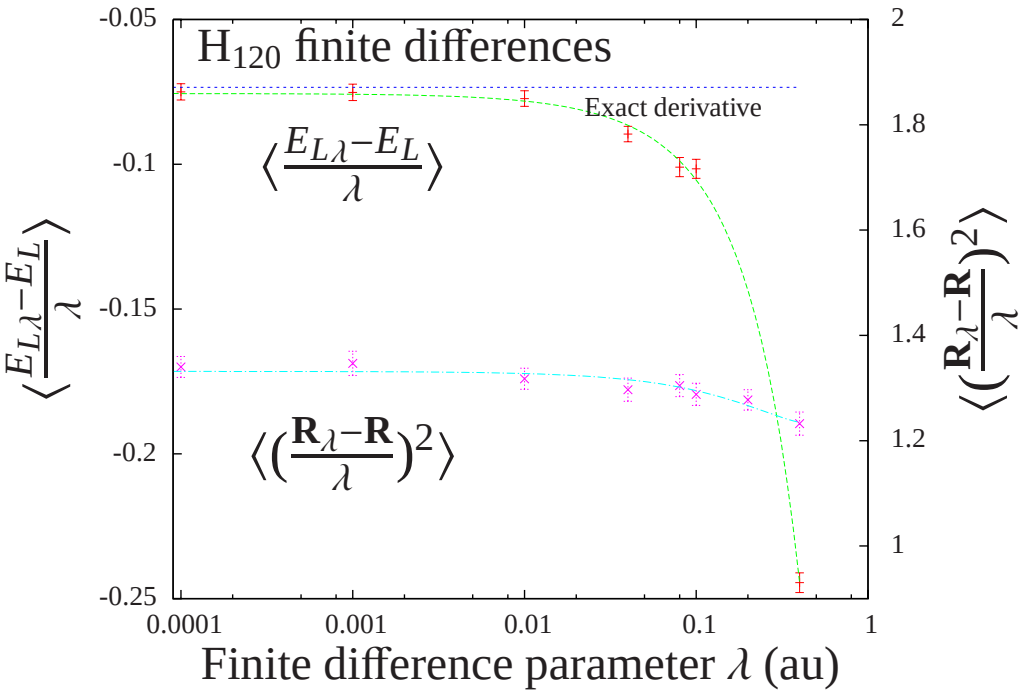


FIG.: Quadratic distances betwen the two processes

Locality of the algorithm

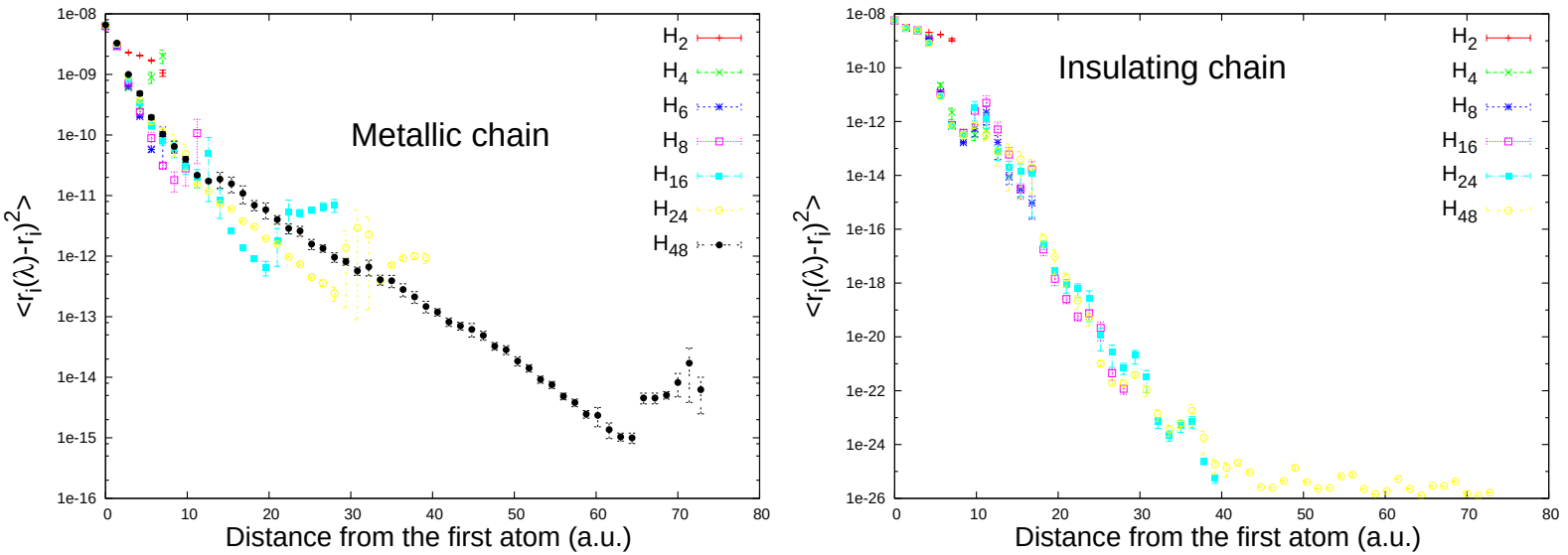


FIG.: Square average of the inter electron distance at a given distance from the first atom

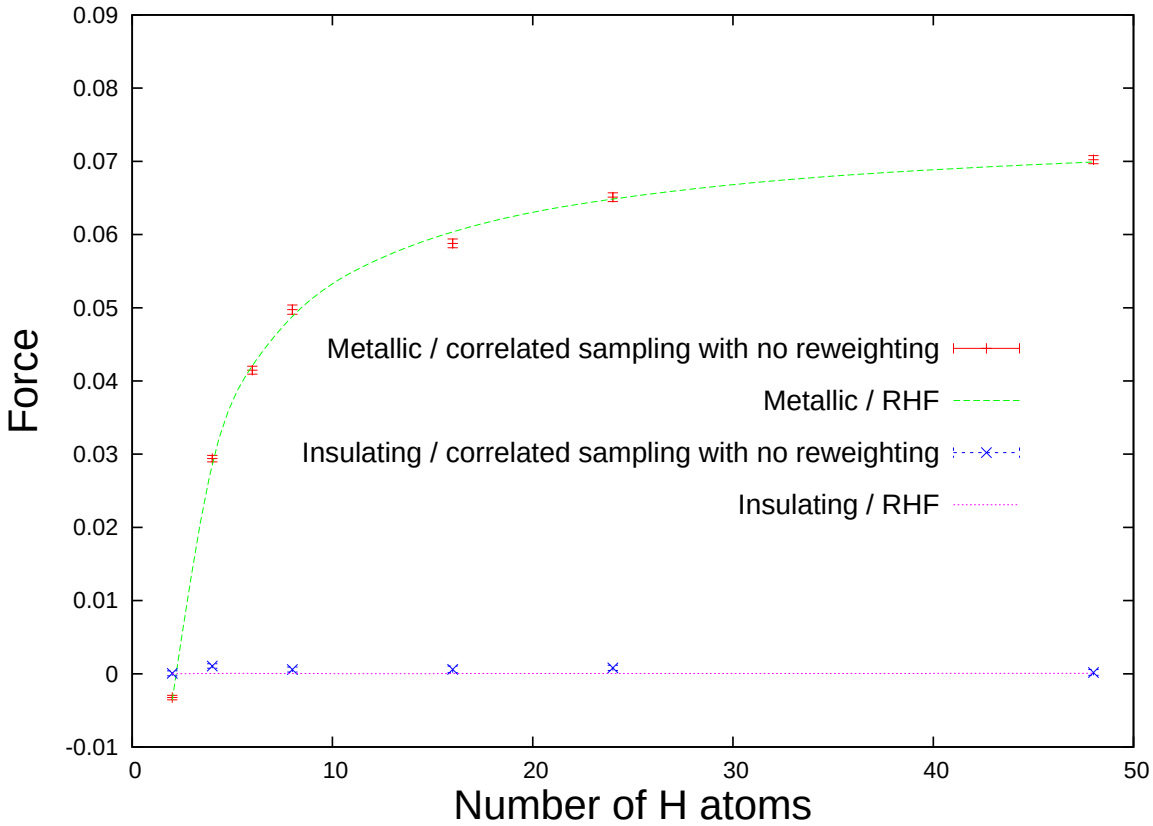


FIG.: Energy derivative with the correlated sampling with no reweighting

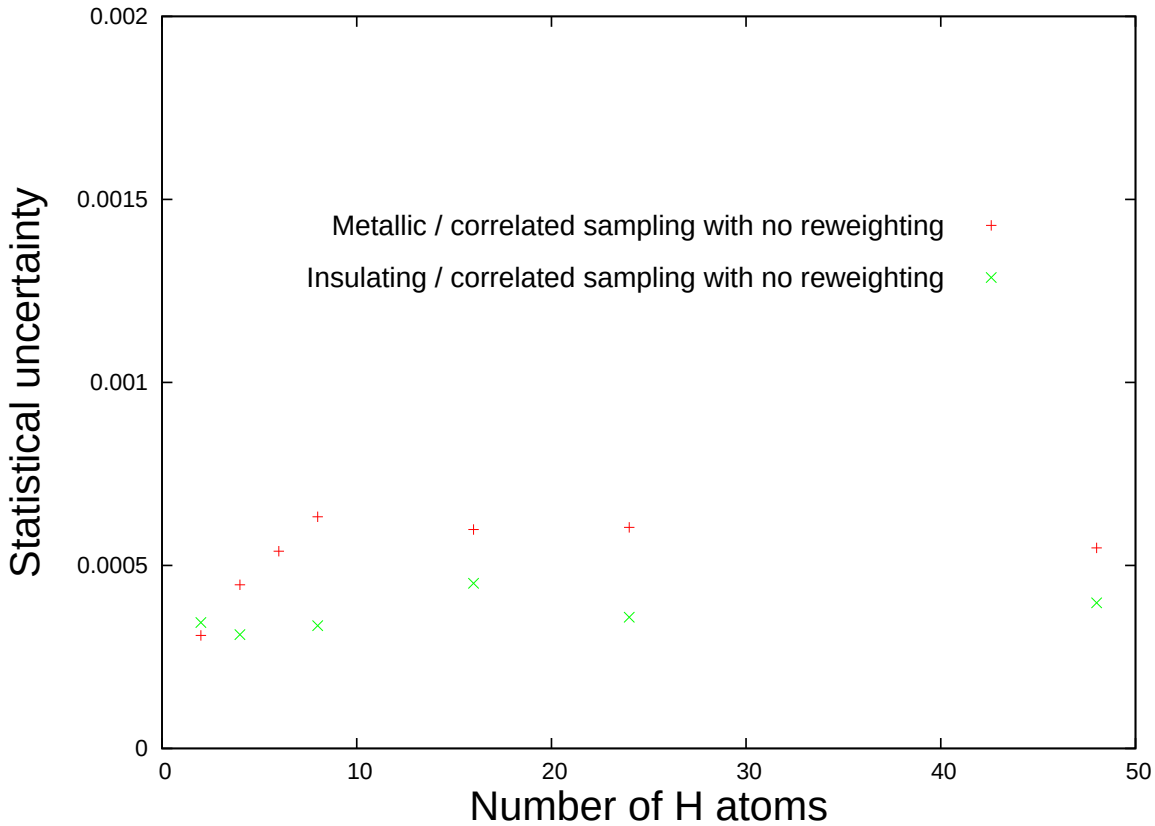


FIG.: Uncertainty as a function of N, metallic chains

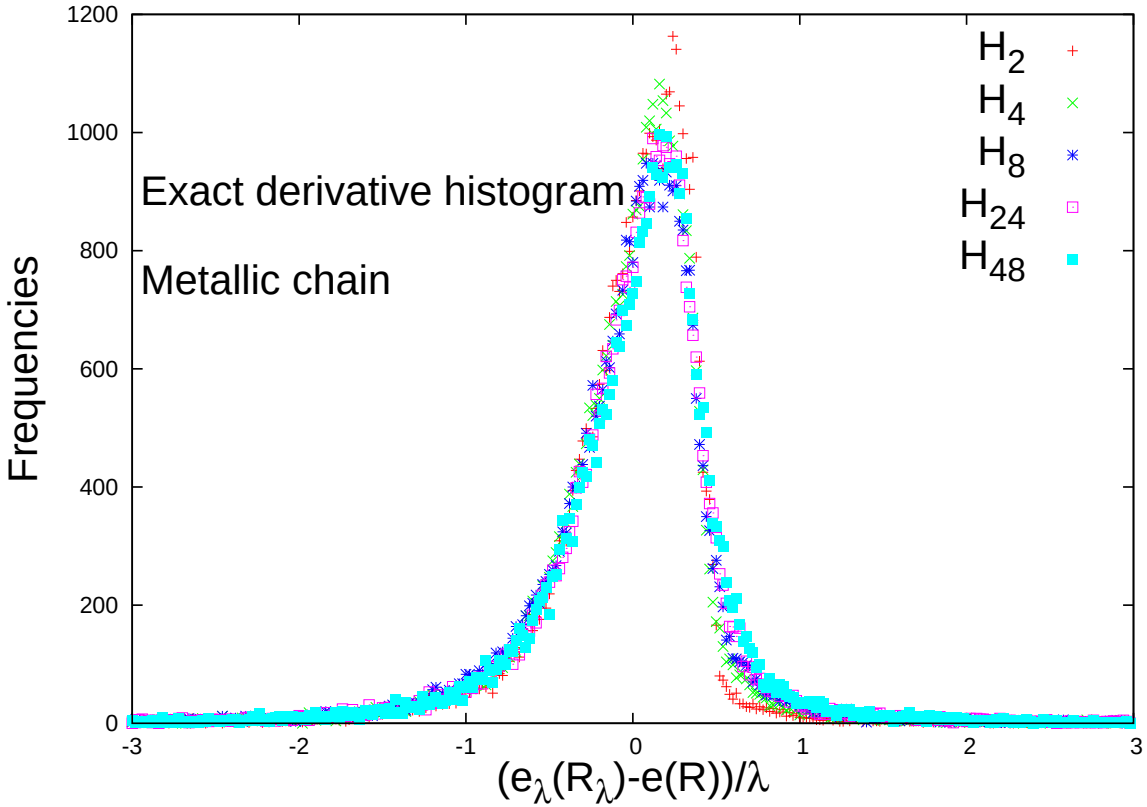


FIG.: Histogram of the correlated difference metallic chain

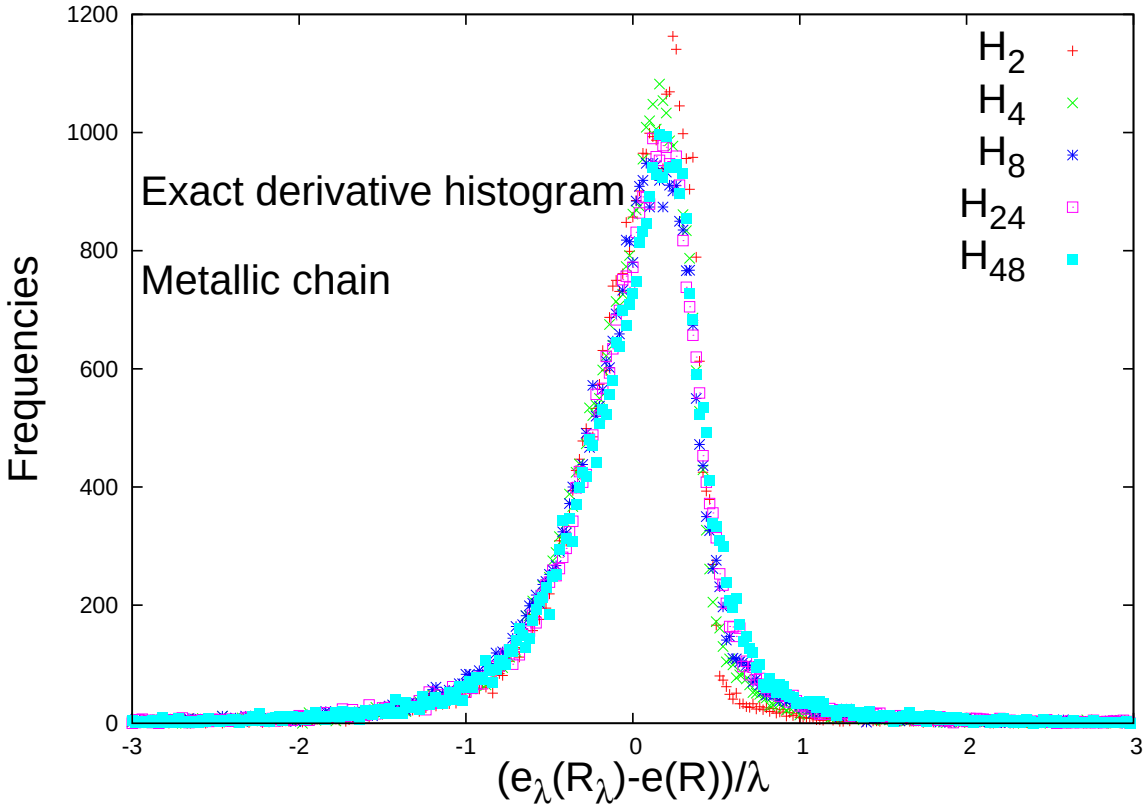


FIG.: Histogram of the correlated difference metallic chain

Reweighting introduces statistitical fluctuations difficult to control

- Solves the small perturbation problem (λ small).
- Sometimes large prefactors in the variance.
- Same large N behavior as independent energy calculations.

Correlated sampling with no reweighting

- Solves the small λ and large N undesirable behavior.
- Perspective to obtain small energy differences with comparable accuracy to the energy.
- Relies on some particular dynamics (stability with respect to the chaos).

Possible to build such stable dynamics

- At the core of perfect sampling (criteria of time convergence, see Fahy, Krauth...).
- Building such dynamics for general molecules is underway.
- Vast subject (numerically, mathematically). Collaborations are welcome...