

*"Slater Type Orbital Project for Quantum Monte Carlo
large molecule simulations"*

**ELECTRONIC CORRELATION : NUMERICAL
STABILITY OF STO-nG AND ITS DERIVATIVES.**

Philip E. Hoggan

LASMEA UMR CNRS 6602
University of Clermont II, Aubière, France

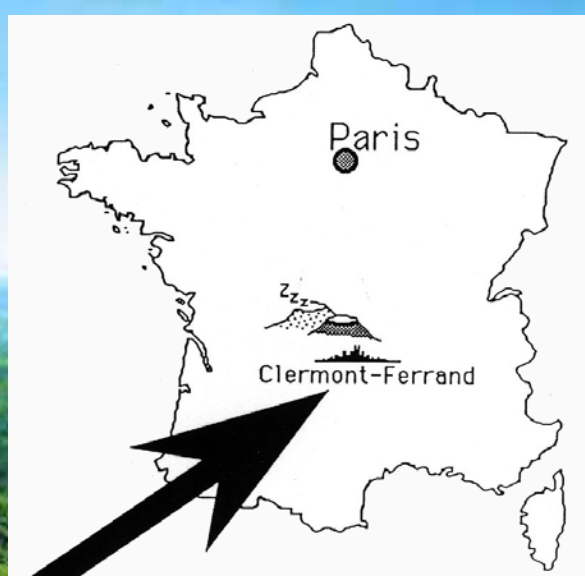
STOP-Qalms: Slater Type Orbital Project for Quantum Monte Carlo
large Molecule simulations

Philip E. Hoggan (PI)

In collaboration with:

*Julien Toulouse. For CHAMP code.
Université Pierre et Marie Curie, Paris*

*Roland Assaraf and Peter Reinhardt, LCT,
Université Pierre et Marie Curie, Paris.*



**University Clermont-Ferrand
France**

DEISA platforms:

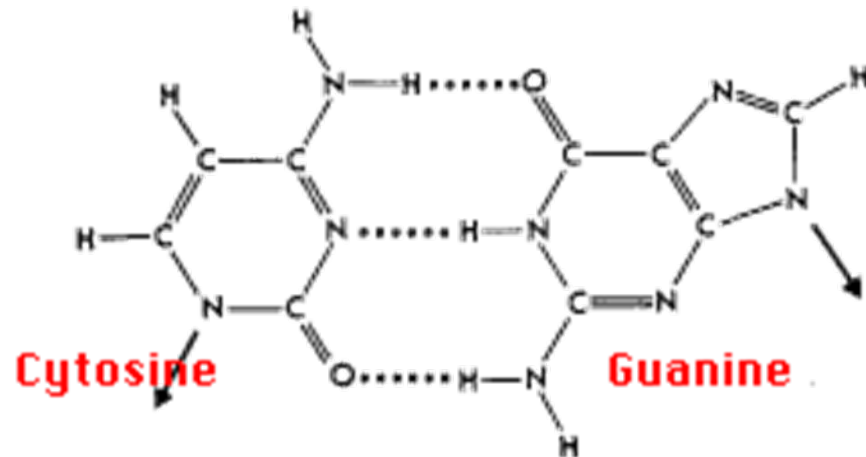
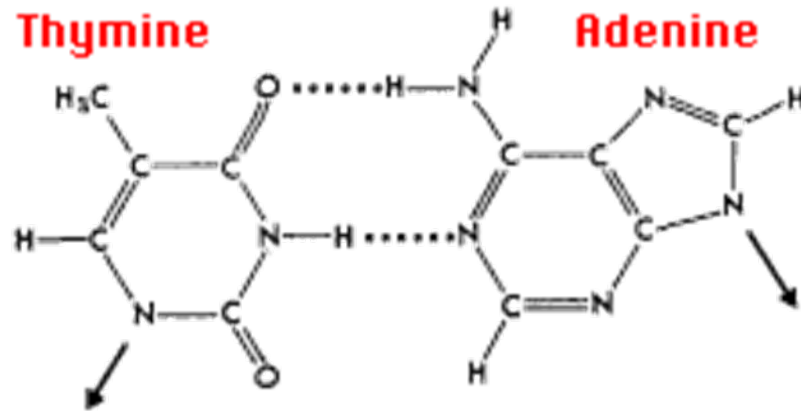
IDRIS (Orsay, France) for adapting code to BLUEGENE/P architectures.

JULICH and RZG (Germany) for production runs.

BLUEGENE/P was used for Quantum Monte Carlo (QMC) only, i.e. Only the CHAMP code was run on it.

Clearly, this provides access to much larger molecules than on our in-house clusters of 32 to 64 cores. (often limited to diatomics).

Some objectives: QMC Study of weakly bound systems
(each run requires about 75h on 2048 cores).



OUTLINE

Introduction

- A few words on how Quantum Monte Carlo works
- Variational and diffusion Monte Carlo.
Trial wavefunction, VMC, DMC.
- Adapting the CHAMP code to Bluegene
- Applications to symmetry-breaking in H-rings.
- Other applications.

Conclusion and perspectives

INTRODUCTION: Exponential decay of orbitals

Slater type functions are chosen as basis since they constitute a complete set (with the continuum states).

Certain combinations restore Hydrogen-like functions, which have the correct radial nodal structure for atoms.

An extensive study of numerical properties of the STO-nG expansion shows the implications for evaluating derivatives.

A preference for ETO trial wavefunctions in QMC is recommended.

Analytical solution of Coulomb integrals in molecules over ETOs is applied to build trial wavefunctions for Quantum Monte Carlo. Integral code is a time-consuming task in Slater orbitals.

- may be parallelized (poor scaling)
- needs development before being ported to a BlueGene structure
- initial idea for this part was not realizable in due time

Introduction to quantum Monte Carlo simulations

This is based on stochastic approaches of the Schrödinger equation.
The time-dependent equation is transformed to a diffusion problem.

The initial (trial) wavefunction should be accurate and correlated. It must possess accurate (preferably analytical) first derivatives.

Here exponential basis functions are preferable to STO-nG expansions.

Fundamentals

Basic object= Walker

= Configuration in space of the electrons

= Vecteur- position: $R = f (r_1 , r_2 \dots r_n)$, n = number of electrons

Time evolution of walkers follows simple stochastic rules (stochastic MD)

Ensemble averages are obtained for the set of walkers.

QMC calculations *concrete* stages

i Choice of **trial wavefunction** ϕ_T .

ii **Optimisation of parameters** therein.

iii **VMC** (Variational Monte Carlo)

= calculation of intermediate quality (typically, 30% to 60% correlation.)

iv **FN - DMC** (Fixed - Node Diffusion Monte Carlo)

= calculation is “quasi - exact” (90 to 100 % correlation..)

v **Jastrow factor** = description of the Coulomb hole.

(dynamic correlation) Use explicitly correlated Sturmians, which orthogonalise the Coulomb repulsion: the e-e cusp is put into a basis.

Post Hartree-Fock methods describe this cusp with difficulty.

Monte-Carlo calculations need Slater orbitals for electronic structure: electrons close to the nuclei must be well-described as **gradients and 2nd derivatives are huge** in these regions:

Obtain ground-state energy (of each spin state -> acrolein)

subject to three sources of systematic error:

- finite number of walkers
- finite time step in the simulation run
- **fixed nodes of the wavefunction**

(changes sign on exchange of electrons, thus electrons for a walker are displaced within regions of the same sign)

- **independent walkers**, thus perfectly parallelizable small test case: H₆ ring

Variational Monte Carlo.

VMC permits sophisticated wavefunctions, with explicitly correlated terms (e-e, e-N, e-e-N terms)

The [expectation value](#) necessary can be written in the X representation as

$$\frac{\langle \Psi(a) | H | \Psi(a) \rangle}{\langle \Psi(a) | \Psi(a) \rangle} = \frac{\int |\Psi(X, a)|^2 \frac{H\Psi(X, a)}{\Psi(X, a)} dX}{\int |\Psi(X, a)|^2 dX}.$$

Following the [Monte Carlo method](#) for evaluating integrals, we can interpret

as a [probability distribution](#) function, sample it, and evaluate the energy expectation value $E(a)$ as the average of the local

function $\frac{H\Psi(X, a)}{\Psi(X, a)}$

$$\frac{|\Psi(X, a)|^2}{\int |\Psi(X, a)|^2 dX}$$

General Metropolis algorithm.

Usual choice for VMC:

For vector positions defined by x and y , the displacement probability:

$$P(x \text{ to } y) = c(t) \exp[-(y - x - b(x) t)^2 / 2t]$$

Depends on the drift vector b , defined by $\text{grad} (\Psi) / \Psi$.

Initially, Ψ is the trial wavefunction Ψ_T .

b = drift vector $b = \text{grad} (\Psi_T) / \Psi_T$.

t = a time-step chosen on the scale of the process.

Principles of Diffusion Monte Carlo.

The time dependent Schroedinger equation is re-written by defining an ‘imaginary time’ variable in the form of a diffusion equation:

$$i \frac{d\Psi(x, t)}{dt} = -\frac{1}{2} \frac{d^2\Psi(x, t)}{dx^2} + V(x)\Psi(x, t). \quad \frac{-d\Psi(x, t)}{dt} = (H - E_0)\Psi(x, t)$$

This is then solved by propagating a trial wave-function in imaginary time, and adjusting the guessed ground-state energy eigenvalue (E_0). Note that the exact value gives a zero derivative. **Excited state contributions fade out.**

Ground state produced is fully correlated.

Fixed-Node approximation.

The ground-state wave function for a given Hamiltonian is approached during the diffusion process by the product of a Trial wave-function and a Jastrow (correlation function).

Limitation: the ground-state is constrained to have the same zeros (nodes) as the trial wave-function.

Flexibility: the Jastrow factor generally contains 2- and 3- particle factors, Optimised variationally.

For electron pairs, the Coulomb hole is given by terms like: $\exp(\mathbf{r}_{12}/2)$.

Usually, electron pair+nuclear 3-particle terms are essential.

Adapting the QMC code CHAMP to a Bluegene architecture

CHAMP is a QMC code developed at Cornell by Cyrus Umrigar, Claudia Filippi, Julien Toulouse and others.

The code had already been tested on parallel machines up to several hundred processors, but had to be rewritten partly to comply with specificities of the DEISA IBM BlueGene architectures.

On these machines each processor has four cores, sharing 2 GB of main memory.

Each process should therefore occupy only 512 MB, which was achieved by changing parts written in FORTRAN77 with static memory allocation to FORTRAN90 dynamical allocation i.e. All static tables were removed. (work done by J. Toulouse).

Code ran on 8192 cores in VN mode i.e. 4 processes per cpu.

DMC: walkers for each core.

This process must be done by the program, without input-output to the disc.

The cores may then run independent DMC processes.

This increases the total number of walkers (N) and accuracy is improved by a factor of $N^{1/2}$.

Recall three steps of a calculation: **Input to CHAMP:**

- Hartree-Fock orbitals for starting wavefunction, with true Slater orbitals or Slater orbitals fitted to 10-15 Gaussians.

This uses conventional molecular codes, given in the input to CHAMP.

VMC step: Takes input multiplied by Jastrow factor (for explicit correlation). Jastrow factor is optimised variationally.

DMC step: fixed node ground state propagated in imaginary time. Gives pure correlated ground state.

After this optimization of the code, runs up to 8192 cores (2048 processors) were possible, even on large molecular systems, for VMC and DMC work.

Scaling test. (Babel, IDRIS nr Paris)

This justified using Blugene/P and is applicable to other parallel systems.

Scaling of QMC: Identical jobs, optimization of Jastrow factor on a H_6 ring.

The number of cores was a multiple of 64. Times were linear.

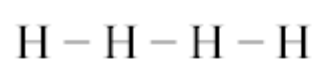
Cores (n times 64). Total cpu time (time for n/time for 2n ~ 2) in secs.

n=1	1892
n=2	952 (1.98)
n=4	482 (1.97)
n=8	247 (1.96)

Note this takes us to 512 cores but runs were on 2048-8192.

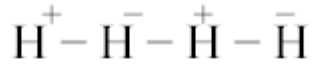
Symmetry breaking in metallic H2 rings

In previous work by P. Reinhardt and J.-P. Malrieu, it was shown that HF calculations on metallic hydrogen rings result, for large enough rings, in three symmetry-distinct wave functions: (density matrix structure is different):



a	b	c	-	-----
b	a	b	c	
c	b	a	b	c
	c	b	a	b
				/
				/
				/
				/

SA



a	c	-	-	-----
c	b	c		
	c	a	c	
		c	b	c
				/
				/
				/
				/

AC-SB



a	b	-	-	-----
b	a	c		
	c	a	b	
		b	a	
				/
				/
				/
				/

BC-SB

alternates:

on diagonal

on off-diagonal

In HF, the symmetry-broken (SB) solutions are lower in energy than a symmetry-adapted (SA) one.

Adding electron correlation inverted the energetic order of the wave function

The symmetry-adapted one is the ground state, as expected, but the energy differences between the three solutions persist.

We expect a unique solution (from quantum theory).

The existence of three stable solutions shows that the correlation methods used were not sufficient to restore the symmetry breaking.

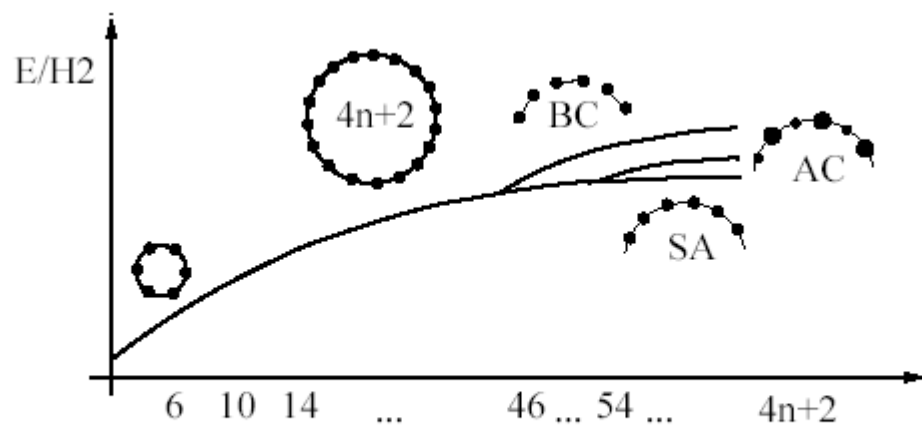


Fig. 1 Schematic drawing of the symmetry-breaking in metallic hydrogen rings. At a certain ring size a bond-centered (BC) molecular and an atom-centered (AC) ionic solution emerge from the continuous line of symmetry-adapted solutions.

The practical approach:

s Slater orbitals with exponent 1.17 on each hydrogen
H-H spacing 0.74747 Angstrom

Hartree-Fock calculation for obtaining orbitals

Trial wave-function: HF determinant multiplied
by a Jastrow factor with optimized parameters

Rings from H₆ to H₈₂

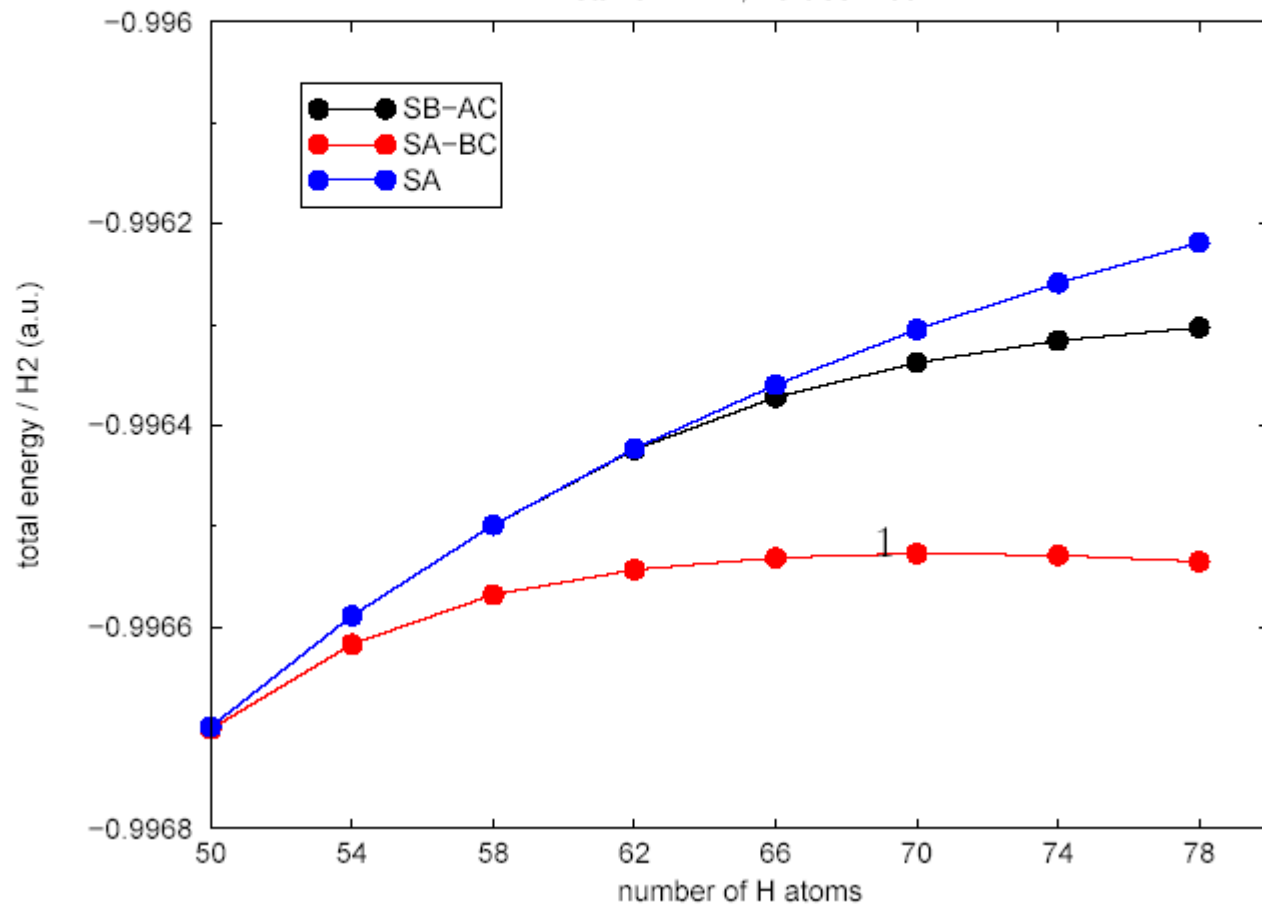
Symmetry breaking sets in at 46 H-atoms (HF,MP2).

Reference data: HF, MP2, (ACPF = averaged coupled pair functional,
like Full-CI, but N⁶ scaling in the system size)

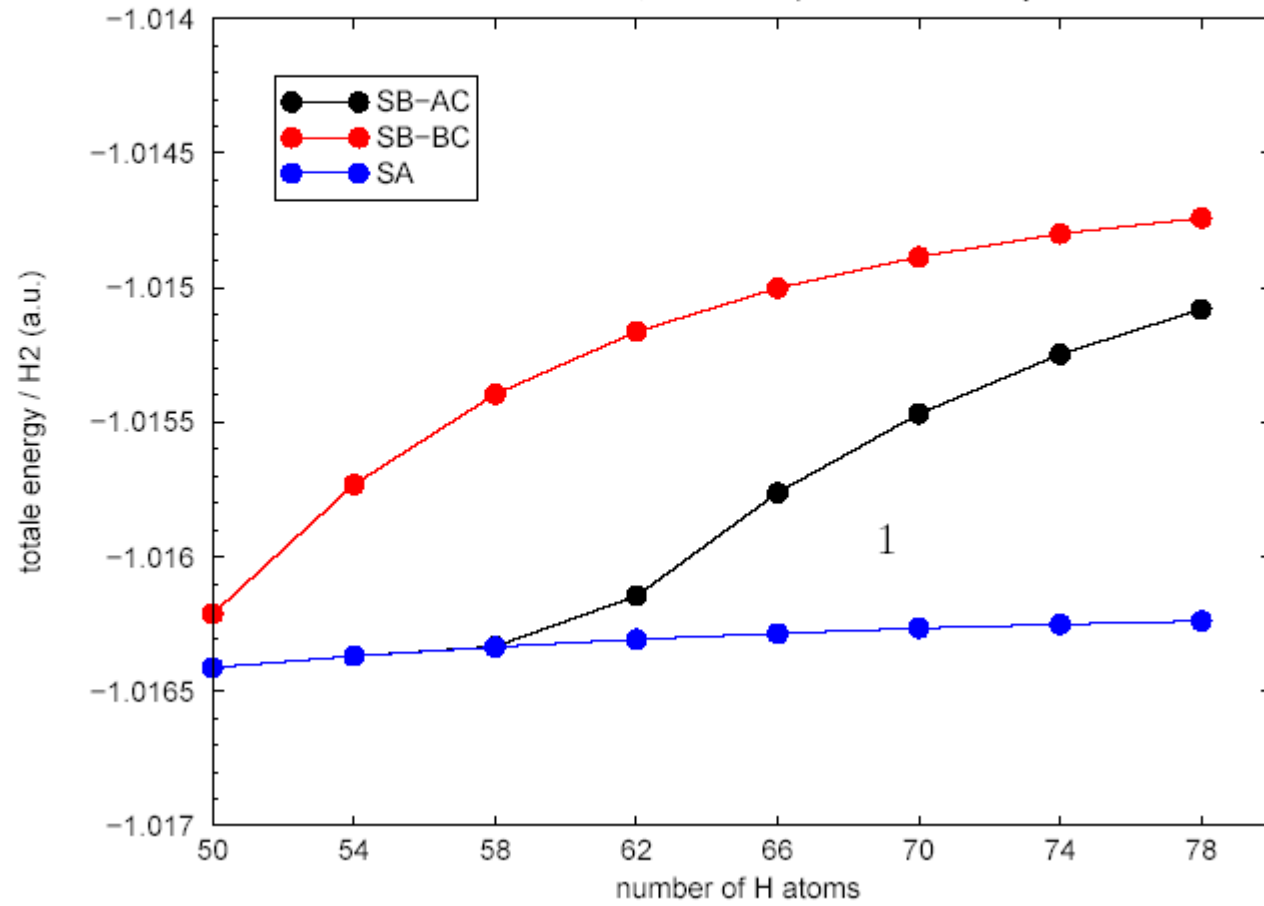
HF : little splitting between solutions

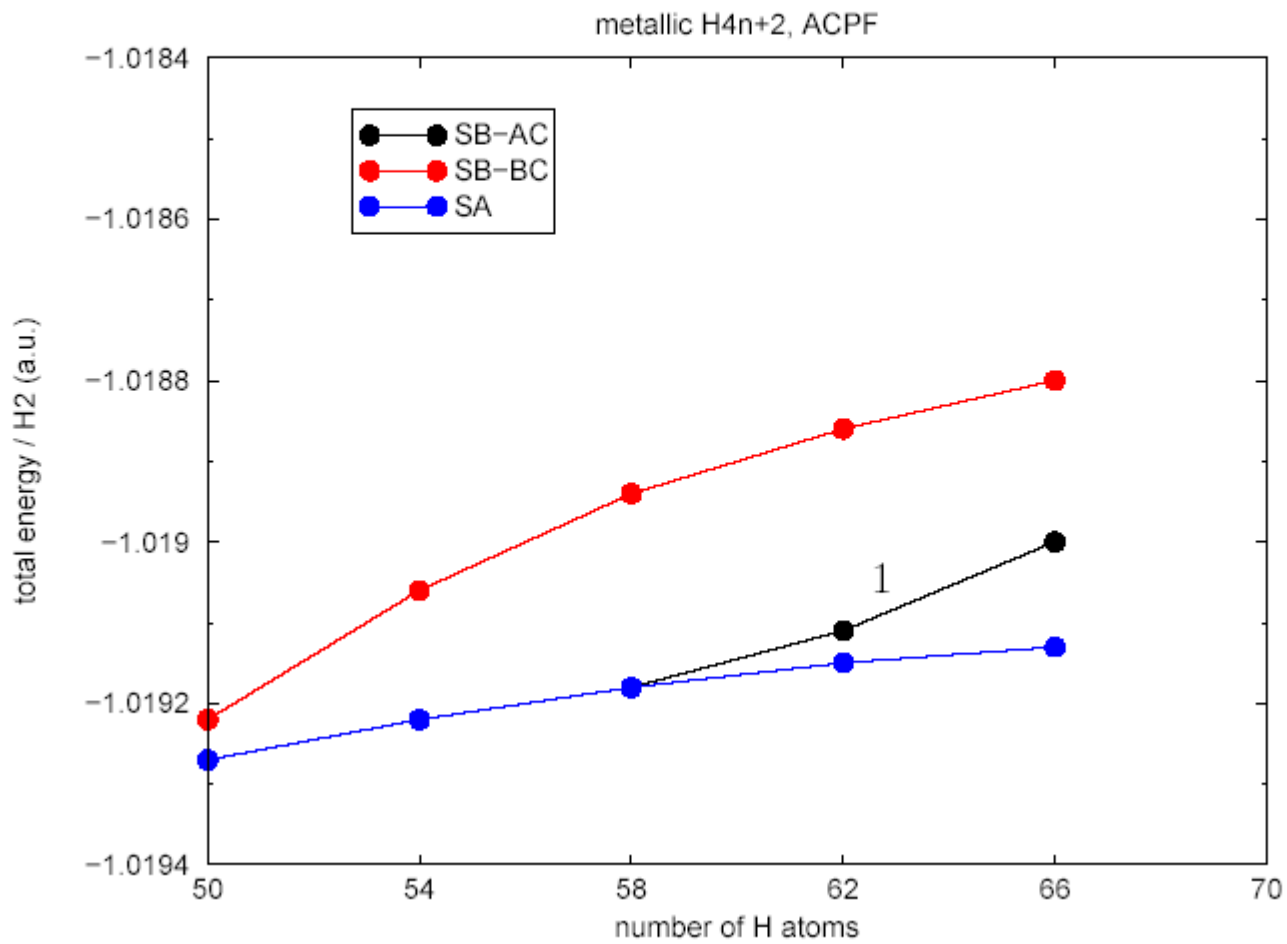
MP2: significantly larger splitting between solutions

metallic H_{4n+2}, Hartree-Fock



metallic H_{4n+2}, 2nd-order perturbation theory



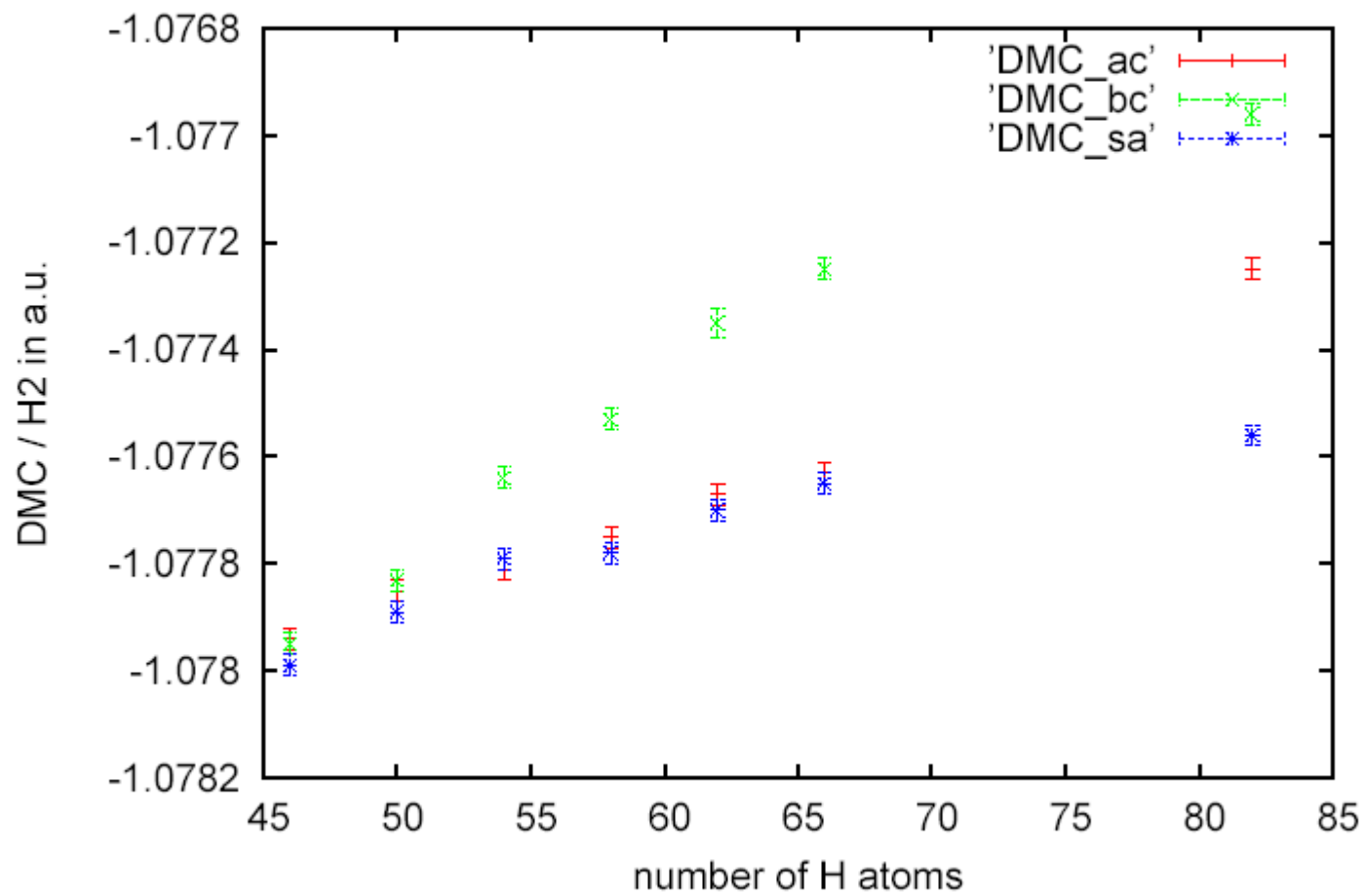


VMC results after optimization of the Jastrow factor

More correlation than previous:less splitting.

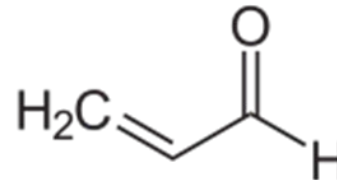
N	ac	bc	sa	error (micro Hartree)
46	-1.06375	-1.06373	-1.06376	21
50	-1.06361	-1.06340	-1.06361	20
54	-1.06352	-1.06290	-1.06352	18
58	-1.06347	-1.06259	-1.06347	17
62	-1.06325	-1.06236	-1.06341	16
66	-1.06296	-1.06228	-1.06337	15
82	-1.06236	-1.06187	-1.06328	12

metallic hydrogen rings



Singlet-triplet excitation in acrolein

Following the work of Caffarel and coworkers on QMC calculations of the n to π transition in the CO moiety of acrolein,



we redid QMC calculations with compact Slater basis sets instead of large Gaussian basis sets.

It was found to be advantageous, the excitation energy being much stabler with respect to changes in the basis size and in the number of Slater determinants.

In particular, a state-average optimization was found to be no longer necessary, i.e. a separate optimization of each state was possible.

Ongoing discussion with Caffarel is expected to lead to a publication.

Some results. Acrolein singlet-triplet (UV) transition.

1) Variational results are below those of the largest GTO basis for a compact (Ema vb1) Slater basis set.

2) The Slater type orbital basis State Specific gap results are (exp =3.6ev):

$$\text{E-DMC (CAS(2,2) VB1) = gap = 3.77(2)}$$

$$\text{E-DMC (CAS(6,5) VB1) = gap = 3.61(2)}$$

Whereas the previous (cusp-corrected gaussian) State Specific results are:

$$\text{E-DMC (CAS(2,2) basis 3) = gap= 3.29(12)}$$

$$\text{E-DMC (CAS(6,5) basis 3) = gap= 4.03(14)}$$

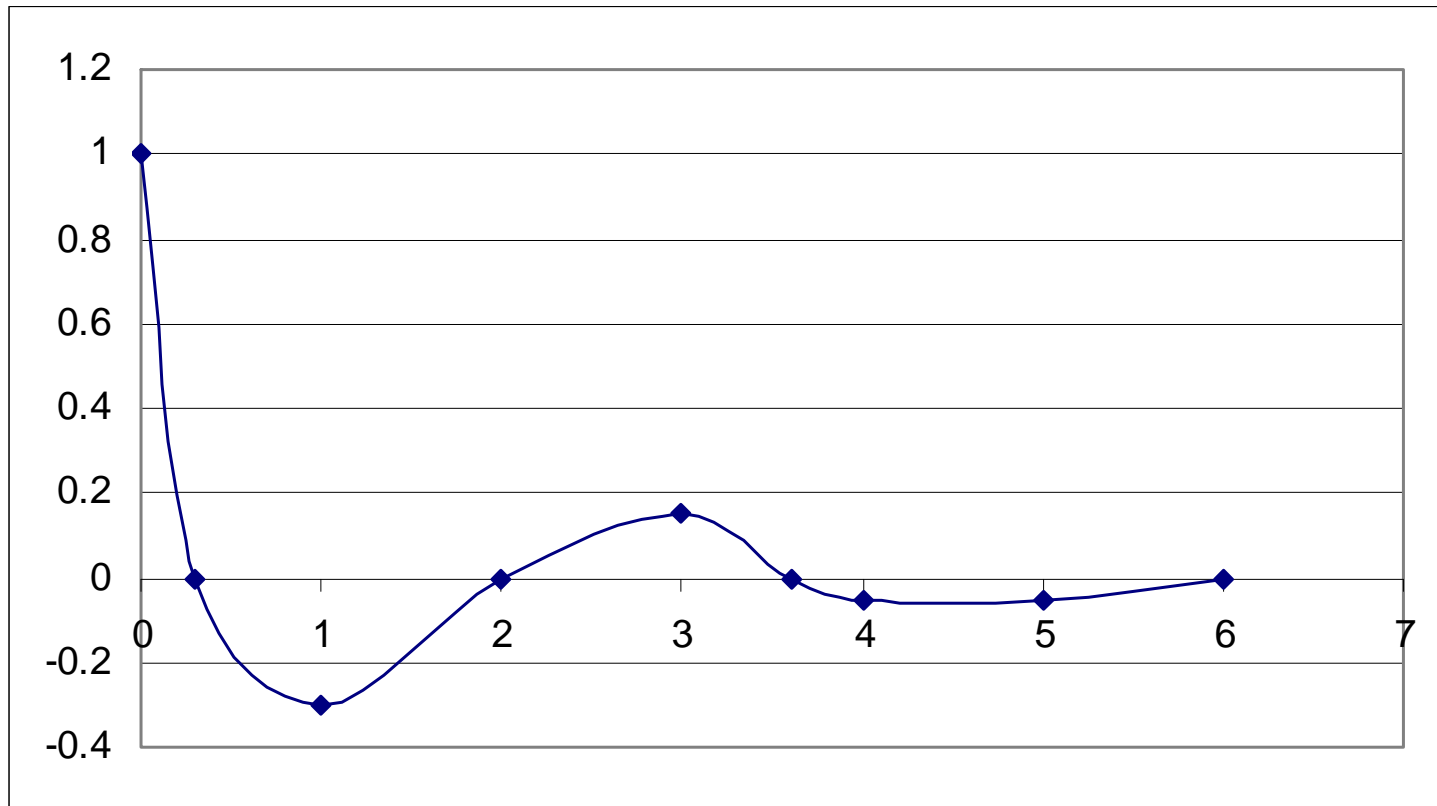
Publications (in preparation):

Symmetry-breaking in metallic hydrogen rings:
a Diffusion Monte-Carlo study, *P. Reinhardt, J. Toulouse,*
R. Assaraf, P. E. Hoggan

Benchmark evaluation of the CO singlet-triplet excitation in acrolein,
J. Toulouse, R. Assaraf, C. Umrigar, M. Caffarel, P. E. Hoggan

Difference between a 1s Sturmian and the STO-nG

Global optimisation of exponents and expansion coefficients with CFSQP.



Universal curve with $2n$ roots

Approximating an exponential

In the space $L^2([0, +\infty[, r^2 dr)$, $\langle f, g \rangle$ denotes the scalar product of two functions and $\| f \|$ the norm of a function.

Problem

For a given $N > 0$, find the best approximation in $L^2([0, +\infty[, r^2 dr)$ of the function

$E_1 = 2e^{-x}$ (normalised to 1) by a sum of the form

$$S_N = \sum_{i=1, N} c_i G_{\alpha_i}(x).$$

Where all α_i are strictly positive (but distinct from Slater exponents)

$G_{\alpha}(x) = 2^{7/4} \alpha^{3/4} / \pi^{1/4} e^{-\alpha x^2}$ Are gaussians normalised to 1.

Denoting by A the matrix N by N with general element

$$A_{i,j} = \langle G\alpha_i, G\alpha_j \rangle = 2^{3/2} \alpha_i^{3/4} \alpha_j^{3/4} / (\alpha_i + \alpha_j)^{3/2}$$

And by F the vector of length N and components $F_i = \langle E_i, G\alpha_i \rangle$;
And by C the vector of constants c_i , it can be shown that

$$J = \| E_1 - s_N \|^2 = 1 - 2C^T F + C^T A C.$$

The minimum J_0 of J is sought when c_i and α_i vary.
Since the α_i are fixed, J is quadratic in the c_i with minimum

$$J = 1 - C^T F \text{ for } A C = F.$$

$$\text{Therefore } J = 1 - F^T A^{-1} F.$$

The problem is thus an optimisation (non linear) in the α_i .

The function $D(\alpha)$

The cylindrical parabolic function $U(a, x)$ can be defined for $x > 0$ to be the solution of the differential equation :

$$f''(x) - (x^2/4 + a) f(x) = 0$$

Satisfying the conditions (initial limit)

$$U(a, 0) = \pi^{1/2} / (2^b \Gamma(b + 1/2))$$

$$U'(a, 0) = \pi^{1/2} / (2^{b-1/2} \Gamma(b))$$

With $2b = a + 1/2$

The function $D(\alpha)$

Expanding e^{-x} as a series, one obtains the expression

$$D(\alpha) = \sum_{n=0, +\infty} a_n, \text{ with } a_n = (-1)^n \Gamma(p) / 2n! \alpha^{-p} \text{ with } p = (n+3)/2$$

The rule due to d'Alembert shows the series is convergent for all $\alpha > 0$
(formula due to Stirling)

$$|a_{n+1}/a_n| = \Gamma(n/2+2) / (\Gamma((n+3)/2) \alpha^{1/2} (n+1)) = 1 / (2n \alpha)^{1/2} + o(1/n^{1/2})$$

The coefficients a_n satisfy recurrence relations:

$$a_0 = \pi^{1/2}/4 \alpha^{-3/2} \quad \text{and} \quad a_1 = -1/(2 \alpha^2).$$

$$a_{n+2} = (n+3) a_n / (2 \alpha (n+1)(n+2)).$$

This makes calculating numerical values of the sum of the series easy.

CFSQP

CFSQP (C code for Feasible Sequential Quadratic Programming) is a set of functions in C to solve the following optimisation problem, with one or more cost functions.

$$\min_x \text{ in } X \max_i \text{ in } I f_i(x) , \text{ for } X \text{ in } \mathbb{R}^n .$$

with

- (i) limits of the variables $x_{\min} < x < x_{\max}$,
- (ii) constraints : non-linear inequalities $g_i(x) < 0$,
 - linear inequalities $< c_i, x > -d_i < 0$,
 - non-linear equalities $h_i(x) = 0$,
 - linear equalities $< e_i(x), x > - f_i(x) = 0$.

For each iteration, two steps are carried out.....

1. From an acceptable point, search for
 - (i) an acceptable descent direction of the form

$$d = (1 - \rho) d_{SQP} + \rho d_{RANDOM};$$

- (ii) a correction of second order $d'(x, d, H)$.

2. Search for a minimum on the arc $x + td + t^2 d'$

- (i) by the method due to Armijo : decrease in maximum 4 iterations,
 - (ii) decrease at each iteration.

The result is a global minimum and convergence super-linear.

The cost function

The CFSQP user must supply cost functions (c.fn)

- (i) Functions describing the constraints,
- (ii) Variable limits
- (iii) optional: functions calculating partial derivatives of c. fn (Hessian).

The efficiency of optimisation depends on suitable input parameters for the programme for problems considered.

First results

The function optimised is

$$J_N = \left\| e^{-x} - \sum_{i=1, N} d_i e^{-\alpha_i x^2} \right\|$$

Where the variables of optimisation are the $\alpha_i, i = 1 \dots N$.

N	J_N
2	$7.895242 \cdot 10^{-4}$
3	$8.262505 \cdot 10^{-5}$
4	$1.094058 \cdot 10^{-5}$
5	$1.722431 \cdot 10^{-6}$

The cost function J_N is present in the form of a difference of two terms of the same order of magnitude thus requiring accuracy when the it is small
For $N > 5$ it becomes necessary to calculate J_N with extended precision

With MPFR (Multiple Precision Floating-Point Reliable Library) in C, the optimisation is then carried out over $2N$ variables

$$d_i \quad i = 1 \dots N$$

and

$$\alpha_i \quad i = 1 \dots N.$$

N	J_N
6	$3.093592 \cdot 10^{-7}$
7	$6.187166 \cdot 10^{-8}$
8	$1.352305 \cdot 10^{-8}$
9	$4.157609 \cdot 10^{-9}$
10	$0.900932 \cdot 10^{-9}$

J_N decreases exponentially with N .

Nano-hartree accuracy may be obtained with $N = 10-12$.

Asymptotic behaviour of solutions

The coefficients d_i $i = 1, \dots, N$ are all positive.

If the exponents satisfy $\forall N, \alpha_1 < \alpha_2 < \dots < \alpha_N$

Then for $1 < i < N_0$, the quotient α_{N-i}/α_N for $N > N_0$ converges rapidly to a limit as N increases.

N	α_N/α_{N-1}	α_N/α_{N-2}
3	$1.821520 \cdot 10^{-1}$	$4.929780 \cdot 10^{-2}$
4	$1.830743 \cdot 10^{-1}$	$5.087484 \cdot 10^{-2}$
5	$1.846768 \cdot 10^{-1}$	$5.181421 \cdot 10^{-2}$
6	$1.830072 \cdot 10^{-1}$	$5.115294 \cdot 10^{-2}$
7	$1.833940 \cdot 10^{-1}$	$5.133951 \cdot 10^{-2}$
8	$1.834157 \cdot 10^{-1}$	$5.136016 \cdot 10^{-2}$

We deduce a method for determination of initial values for the optimisation with $N + 2$ knowing the optimum for N .

Universal curve

For the optimal values of (α_i, d_i) satisfying $\alpha_1 < \dots < \alpha_N$.

$$f_N(x) = e^{-x} - \sum_{i=1, N} d_i e^{-\alpha_i x^2}$$

Then

$$\phi_N(x) = 1/f_N(0) \times f_N((x/\alpha_N)^{1/2})$$

Conjecture: ϕ_N tends to a limit $\phi(x)$ when N is large. ϕ_N has $2N$ real roots all positive definite, denoted x_{Nj} $j = 1, \dots, 2N$ in increasing order

The sequences x_{Nj} , $N > 2$ converge rapidly.

In the large x limit, the maximae of $|\phi(x)|$ follow an asymptotic curve of the form:

$$|\phi(x)| = K e^{-cx^{1/5}}$$

with $0.62 < c < 0.70$.

This allows us to study the limit curve $\phi(x)$.