Geometrical and properties of (bio)molecules by Quantum Monte Carlo

Leonardo Guidon

Dipartimento di Scienze Fische e Chimiche Università degli Studi de L'Aquila, Italy

QMC on the Apuan Alps – 29th July 2014



Dipartimento di Fisica Sapienza – Università di Roma

Outline

- Why Quantum Monte Carlo (QMC) for (bio)molecules ?
- VMC with JAGP at work: geometries, polarizabilities, reaction paths, vibrations, electron density,...
- The structure of small (and large!) organic conjugated molecules @ VMC level
- The important of geometries in photoreceptors properties
- Combining QMC structures with MBPT excitations for protein chomophores
- Perspectives and challenges in the biophysics of transitionmetal complexes

Molecular simulations in biophysics and biochemistry

X-rays, NMR, Electron Microscopy



Computational Molecular Biophysics



Compute Molecular properties Other techniques in Biophysics, Biochemistry and Molecular Biology

- Protein engineering
- Dynamics and folding exp.
- Enzyme Kinetics
- InfraRed Spectroscopy
- Raman Spectroscopy
- EPR,
- Light/Protein interactions: absorption, emission, fluo-rescence,
- photoreactions, ...
- Electron Transfer properties
- Mass spectroscopy
- Neutron scattering
- XAS and many others ...

We need Calculations Accurate (Quantum Mechanics) - Large systems (10⁵ atoms)

MULTISCALE MODELLING Quantum Mechanics /Molecular Mechanics (QM/MM) schemes

Mixed Quantum-Mechanics / Molecular-Mechanics (QM/MM) simulations and Density Functional Theory

Schrödinger Equation – Ab initio MD

Chemical reactions in vacuo Small Clusters 100-1000 atoms 10-100 ps



Newton's Equations - Classical MD Structural and dynamical properties of biomolecules 10⁴-10⁵ atoms **1** μs



QM / MM 10⁴-10⁵ atoms for 10-100 ps



MM(enzyme+solvent) "classical" atoms



QM(active site) "quantum" atoms

A. Laio et al., J. Chem. Phys. 116, 6941 (2002) M. Colombo et al. CHIMIA 56, 11 (2002)

An example: the Biophysics of photosynthesis

Light capture



Energy conversion and stabilization



 $\begin{array}{c} h\nu \\ 2 \text{ H}_2\text{O} \xrightarrow{} 4 \text{ H}^+ + \text{O}_2 + 4 \text{ e}^- \end{array}$

Energy stabilization by secondary processes to avoid charge recombination Synthesis and Export of stable products (ATP)



Quantum Mechanics / Molecular Mechanics (DFT level) of the water splitting reaction in photosynthesis



D. Bovi, D. Narzi and L. Guidoni, *Angewandte Chemie* 125, 11960-11965. (2013) D. Narzi, D. Bovi, L. Guidoni, *PNAS*, in press (2014)

	Accuracy and Density Functional Theory (DFT) (current) limitations							
	Limitations	Biochemistry/Biophysics	Chemistry & Materials Science					
•	Excited states (conjugates systems, charge transfer excitations)	Light-matter interactions in photoactive proteins and nucleic acids	Optical properties of organic materials					
Radicals and di-radicals		Oxidative stress	Radicals in catalysis					
•	Nearly-half-filled transition metals (like Cr, Mn, Fe, Ni)	Heme-proteins, iron-sulfur proteins, bioenergetics	Transition metal oxydes Magnetic materials Catalysis					
•	Missing (ab-initio) Dispersion	Drug-protein and protein- protein iteractions	Molecular crystals, self- assembling materials					
•	Underestimation of correlation effects (i.e. at reaction barriers)	Accurate reaction rate calculations in enzymes	Reaction mechanisms					

Accuracy: any alternative method?

Method	number of electrons N	Pro 😊	Contra 🛞		
DFT	≤ N ³	Simple to use, large system, ab initio MD,	Many systems are still a challenge for current XC functionals		
MP2,MP4, CCSD, CCSD(T),CI,	N ⁵ -N ⁷ or higher	Very accurate	High accuracy is possible only on small systems		
Quantum Monte Carlo	N ^{3 -} N ⁴	<u>Very accurate</u> good scalabiliy Intrinsically parallel	Large prefactor Stochastic error Forces (?) Lack of "standards"		

Calculating π by blind dards game





Variational Monte Carlo and Stochastic Integration





M: number of sampling points N: number of dimensions of the domain

Quadrature methods to calculate the integral I are

- Very precise for small N
- M increase badly with the dimension N

Stochastic (Monte Carlo) methods



independently to the dimension N of the of the system



Variational Monte Carlo

Finding the minumum of the energy functional E_{T}

$$E_{T} = \min_{\{\alpha_{k}\}} \frac{\left\langle \psi_{T}(\left\{\vec{R}_{i}\right\}, \left\{\alpha_{k}\right\}) \middle| \hat{H} \middle| \psi_{T}(\left\{\vec{R}_{i}\right\}, \left\{\alpha_{k}\right\}) \right\rangle}{\left\langle \psi_{T}(\left\{\vec{R}_{i}\right\}, \left\{\alpha_{i}\right\}) \middle| \psi_{T}(\left\{\vec{R}_{i}\right\}, \left\{\alpha_{i}\right\}) \right\rangle} \ge E_{0}$$

Integrals in 3N-dim space N: number of electrons of coordinates {R_i}

The wave function Ψ_T can be simple (HF) or a complex many body wave function In the 3*N electrons coordinate space

Energy Minimization methods:

- Stochastic reconfiguration [S.Sorella Phys. Rev. B 2005, 71 241103]
- Hessian methods [C. Umrigar et al Phys.Rev.Lett. 2007, 98 110201]

Monte Carlo evaluation of integrals with importance sampling

 $I = \int d\mathbf{R} g(\mathbf{R}) \Big|^{\mathrm{im}}$

portance function

$$\mathcal{P}(\mathbf{R}) \ge 0,$$

 $\int d\mathbf{R} \mathcal{P}(\mathbf{R}) = 1$
 $I = \int d\mathbf{R} f(\mathbf{R}) \mathcal{P}(\mathbf{R})$
 $f(\mathbf{R}) \equiv g(\mathbf{R}) / \mathcal{P}(\mathbf{R})$

Monte Carlo estimate: finite sampling

set of random vectors from the distribution $P(\mathbf{R})$ (Metropolis algorithm)

(1

from

$$I \approx \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{R}_{m})$$
variance: $\sigma_{f}^{2} \approx 1$

$$I = \lim_{M \to \infty} \left\{ \frac{1}{M} \sum_{m=1}^{\infty} f(\mathbf{R}_m) \right\}$$

Μ

variance:
$$\frac{\sigma_f^2}{M} \approx \frac{1}{M(M-1)} \sum_{m=1}^M \left[f(\mathbf{R}_m) - \frac{1}{M} \sum_{n=1}^M f(\mathbf{R}_n) \right]^2$$

error bar:

r: $\pm \sigma_f / \sqrt{M}$

Foulkes et al.: Quantum Monte Carlo simulations of solids, Rev. Mod. Phys., 2001



 $\{\mathbf{R}_m\}$ is a set of point in configurational space distributed like $P(\mathbf{R})$ and generated by Metropolis algorithm

You do not waste time in configurations where P(R) (i.e ψ^2) ≈ 0



Diffusion Monte Carlo (DMC)

$$|\psi_0\rangle = \lim_{\tau \to \infty} e^{-\tau(\hat{H} - E^*)} |\psi_T\rangle$$

DMC gives the lowest energy with the constraints to have the same nodal surface of the VMC (energy optimized) trial wave function (Fixed node approximation) ψ_{0}

 ψ_T

Use of pseudopotentials [Burkatzi,M., Filippi, C. Dolg,M. J.Chem.Phys 2007, 126, 234105]

0

We used Lattice Regularized Diffusion Monte Carlo (LRDMC). This garantee that the use of pseudopotentials preserves the variational principle [Casula, Filippi, Sorella Phys.Rev. Lett 2005, 95, 100204]

Variational Wave Function

$$\psi_T = J(\overline{x}_{1,...,\overline{x}_N}) \cdot D(\overline{x}_{1,...,\overline{x}_N})$$

- $D(\bar{x}_{1,...,\bar{x}_{N}})$
- Fully antisimmetric function. Typically a Slater determinant, or a linear combination of Slater determinants For instance HF, MRSCF, CASSCF, CI.

$$J(\bar{x}_{1,}...,\bar{x}_{N})$$

Jastrow term. Symmetric function with respect to particle exchange. Typically a product of terms explicitly dependent on inter-particle distances $\mathbf{r}_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$

$$F(\overline{x}_{1,...,\overline{x}_{N}}) = \exp(f_{ee} + f_{en} + f_{een})$$

Electron-nucleus terms: e-n electron-electron terms: e-e electron-electron-nuclei : e-e-n

 ψ_T has explicit electron correlation terms

Jastrow Antisymmetrized Geminal Power (JAGP)

Variational wave function based on Pauling's resonating Valence Bonds idea

The chemical bond is described as a singlet pair of electrons.
The quantum state is a superposition or *resonance* of many valence bond states

Singlet between site A and B

 $\Phi(r^{\uparrow}, r^{\downarrow}) = \phi(r^{\uparrow}, r^{\downarrow}) 1 / \sqrt{2} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$



Huckel's Delocalized n-MO Picture

Pauling-Wheland's Resonating Picture



 $\Psi = c_1(K_1 + K_2) + c_2(D_1 + D_2 + D_3)$

 $c_1 > c_2$

Molecular orbital theory Delocalization of π electrons

Valence Bond theory Resonance between different Lewis' structures

Picture from Shaik et al. Reviews in Computational Chemistry, Volume 20, 2004 John Wiley & Sons,

Jastrow Antisymmetrized Geminal Power (JAGP) Determinant

Variational wave function based on Pauling's resonating Valence Bonds idea

- The chemical bond is described as a singlet pair of electrons.
- The quantum state is a superposition or **resonance** of many valence bond states

Singlet between site A and B

В

Picture from Shaik et al. Reviews in Computational Chemistry, Volume 20, 2004 John Wiley & Sons,

Jastrow Antisymmetrized Geminal Power (JAGP) Determinant

A general and fully correlated wavefunction based on Resonating Valence Bond can be used to treat realistic systems in a quantitative way by QMC [1]

$$\Psi_{JAGP}(\mathbf{r}_1,...,\mathbf{r}_N) = \Psi_{AGP}(\mathbf{r}_1,...,\mathbf{r}_N) J(\mathbf{r}_1,...,\mathbf{r}_N)$$

The AGP is the antisymmetrized product of spin singlets

$$\Psi_{AGP} = \hat{A}[\Phi(\mathbf{r}_1^{\uparrow}, \mathbf{r}_1^{\downarrow}) \cdots \Phi(\mathbf{r}_{N/2}^{\uparrow}, \mathbf{r}_{N/2}^{\downarrow})]$$

$$\Phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \psi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) 1/\sqrt{2} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$$

The coordinate part $\psi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow})$ can be expanded on atomic orbitals

$$\psi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \sum_{a,b} \psi_{a,b}(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow})$$

 $\psi_{a,b}(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \sum_{l,m} \lambda_{l,m}^{a,b} \phi_{a,l}(\mathbf{r}^{\uparrow}) \phi_{b,m}(\mathbf{r}^{\downarrow})$

I,m run over different orbitals centered on nuclei *a,b*

m-

a

 $\{\lambda\}$

b

[1] M.Casula et al. J.Chem.Phys. 121,7110 (2004)

AGP rewritten as a multi-determinant wave function in term of molecular (natural) orbitals

$$\Psi_{AGP} = c_0 |\Psi_0\rangle + \sum_{i=1}^{N_p} \sum_{a=N_p+1}^{L} c_{ii}^{aa} |\Psi_{ii}^{aa}\rangle + \sum_{\substack{i,j=1\\i\neq j}}^{N_p} \sum_{\substack{a,b=N_p+1\\a\neq b}}^{L} c_{iijj}^{aabb} |\Psi_{iijj}^{aabb}\rangle + \dots$$

 $c_0 = \prod_i^{N_p} \lambda_i; \qquad c_{ii}^{aa} = c_0 \frac{\lambda_a}{\lambda_i}; \qquad c_{iijj}^{aabb} = c_0 \frac{\lambda_a \lambda_b}{\lambda_i \lambda_j};$

(constrained) coefficients

leading closed-shell Slater determinant

excitation Slater determinants

$$\begin{split} |\Psi_0\rangle &= \hat{\mathcal{A}} \left\{ \left[\prod_i^{N_p} \psi_i(\mathbf{r}_i) \alpha(i) \right] \left[\prod_j^{N_p} \psi_j(\mathbf{r}_{N_p+j}) \beta(j) \right] \right\} \\ &\left| \Psi_{ii}^{aa} \right\rangle \quad , \quad \dots \end{split}$$

Zero-seniority subset (with constrained coefficients) of the Full CI expansion. Only paired excitations included

Jastrow Antisymmetrized Geminal Power (JAGP) Jastrow 1&2

$$\Psi_{JAGP}(\mathbf{r}_1,...,\mathbf{r}_N) = \Psi_{AGP}(\mathbf{r}_1,...,\mathbf{r}_N) J(\mathbf{r}_1,...,\mathbf{r}_N)$$

The **Jastrow part** is composed by 3 different terms $J=J_1J_2J_3$ All of them introduces some penalty factor to the wave function that depends on the distances between electronic coordinates r_i and r_j

 J_1 and J_2 care about the electron-nuclei correlation (one body term) and The electron-electron correlation (two-body term). For instance:

$$J_2(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \exp\left(\sum_{i< j}^N u(r_{ij})\right)$$

Where $u(r_{ii}) < 0$ is a proper pairing function that satisfies the cusp conditions

Jastrow Antisymmetrized Geminal Power (JAGP) Jastrow 3&4

$$\Psi_{JAGP}(\mathbf{r}_1,...,\mathbf{r}_N) = \Psi_{AGP}(\mathbf{r}_1,...,\mathbf{r}_N) J(\mathbf{r}_1,...,\mathbf{r}_N)$$

AGP

$$\Psi_{AGP} = \hat{A}[\Phi(\mathbf{r}_1^{\uparrow}, \mathbf{r}_1^{\downarrow}) \cdots \Phi(\mathbf{r}_{N/2}^{\uparrow}, \mathbf{r}_{N/2}^{\downarrow})] \qquad \Phi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) = \psi(\mathbf{r}^{\uparrow}, \mathbf{r}^{\downarrow}) 1/\sqrt{2}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$$

Jastrow J=J₁J₂J₃

 $J_{3/4}$ is an electron-electron-nucleus-nucleus

l,m run over different orbitals centered on nuclei *a,b*

J_{3/4} introduces instantaneous correlated polarization

Dispersive van der Waals interactions

 J_{3-4} term introduces also off diagonal terms $a \neq b$ that modulate the electron-electon correlation depending on atoms **a** and **b** also at large |a-b|separation

$$J_{3}(\mathbf{r}_{1},...,\mathbf{r}_{N}) = \exp\left(\sum_{i < j} \Phi^{J}(\mathbf{r}_{i},\mathbf{r}_{j})\right)$$
$$\Phi^{J} = \sum_{a,b} \Phi^{J}_{a,b}$$
$$\Phi^{J}_{a,b}(\mathbf{r}_{i},\mathbf{r}_{j}) = \sum_{l,m} g^{a,b}_{l,m} \phi^{J}_{a,l}(\mathbf{r}_{i}) \phi^{J}_{b,m}(\mathbf{r}_{j})$$





SAPT (Symmetry Adapted Perturbation Theory): from ref. [1]
DCACP (Dispersion Corrected Atom Centered Potential): plane wave125 Ry using the scheme described in [2]
[1] Misquitta, A. J. et al., J. Chem. Phys. 2005, 123, 214103.
[2] von Lilienfeld, O. et al. Phys. Rev. Lett. 2004, 93, 153004.

F. Sterpone, S. Spanu, L. Ferraro, S. Sorella, LG JCTC 4, 1428 (2008)

Wave function optimization and Forces

Wave function optimization

<u>All wavefunction parameters, including exponents and coefficients of atomic orbitals</u> of the Jastrow and Determinant part are optimised

- Stochastic reconfiguration [S.Sorella Phys. Rev. B 2005, 71 241103]
- Hessian methods [C. Umrigar et al Phys.Rev.Lett. 2007, 98 110201]

VMC forces

Numerical derivative of the local energy E_{LOC}

$$F(R) = -\langle \lim_{|\Delta R| \to 0} \frac{\Delta}{\Delta R} E_L \rangle_{|\psi_T|^2} +$$

$$+ 2\langle H \rangle_{|\psi_T|^2} \langle \lim_{|\Delta R| \to 0} \frac{\Delta}{\Delta R} log(J^{1/2}\psi_T) \rangle_{|\psi_T|^2} - 2\langle H \lim_{|\Delta R| \to 0} \frac{\Delta}{\Delta R} log(J^{1/2}\psi_T) \rangle_{|\psi_T|^2}$$

ΔE

ΔR

Space-Warp transformation C. Filippi and C. J. Umrigar. Phys. Rev. B, 61:R16291, 2000 Attaccalite C and Sorella S. Phys. Rev. Lett. 100, 114501 (2008) O. Valsson and C. Filippi, JCTC 6, 1275–1292 (2010) Computational cost: **x 3 x number of nuclei**

S. Sorella, L. Capriotti J.Chem.Phys 133, 234111 (2010) Algorithmic differentiation and the calculation of forces by quantum Monte Carlo

Forces calculations just 6 x energy calculations, independently on the number of nuclei



TurboRVB - Quantum Monte Carlo Package S. Sorella et al.

http://people.sissa.it/~sorella/web/index.html

QMC scaling with the number of electrons

N: number of electrons

System		CPU core-hours/ WF point (h)	CPU core-hours/ GEO point (h)
$\mathrm{C_{2}H_{2}}$	10	2.8	8.6
$C_5H_6NH_2^+$	32	27.2	124.3
RPSB	120	631.6	4333.9
$\mathrm{RPSB} + \mathrm{H_2O} + \mathrm{Glu113} + \mathrm{Glu181}$	182	1818.9	10574.7

Total computer time for Wavefunction / Geometry Optimization: $\approx N^{(3.2 \div 3.5)}$

Thanks to High-Performance-Computing geometry relaxation of organic molecules

up to about 100 atoms (with pseudopotential) are possible

Coccia and Guidoni J.Comp.Chem 33, 2332-2339 (2012)

How far we can go with JAGP on Molecules?

GEOMETRIES	Barborini, Sorella and Guidoni, JCTC 2012, 8, 1260							
TRANSITION STATES	Barborini and Guidoni, J. Chem. Phys. 2012, 137, 224309							
POLARIZABILITY	Coccia, Chernomor, Barborini, Sorella and Guidoni, JCTC 2012, 8, 1952							
DENSITY	Varsano, Barborini and Guidoni, J. Chem. Phys. 2014, 140, 054102							
VIBRATIONAL PROPER	TIES Zen, Zhelyazov, Guidoni 2012, JCTC 8:4204–4215 Zen, Luo, Sorella, Guidoni 2013 JCTC 9:4332–4350.							
RADICALS	Zen, Trout, and Guidoni J.Chem.Phys. 2014							
DI-RADICALS	Zen, Coccia, Luo, Sorella, and Guidoni 2014, JCTC 10, 1048-1061.							
CONJUGATED (BIO) MO	DLECULES Coccia, Guidoni <i>J. Comp. Chem 2012,</i> 33:2332–2339 Coccia, Varsano, Guidoni 2014, <i>JCTC</i> 10:501–506 Barborini, Guidoni 2014, in preparation							
QUANTUM MONTE CA	RLO / MOLECULAR MECHANICS							
	Coccia, Varsano, Guidoni 2013, <i>JCTC</i> 9, 8-12 Varsano, Coccia, Pulci, Conte, and Guidoni, 2014 Comp.Theo.Chem							
MOLECULAR DYNAMI	MOLECULAR DYNAMICS							
	Mazzola, Luo, Zen, Guidoni, Sorella, 2014 in preparation							

Variational Monte Carlo and Density Functional Theory

DFT orbitals from QMC density Using the ZMP procedure [Zhao, Morrison, Parr, Phys.Rev.A, 50, 2138 (1994)]

VMC density \rightarrow KS Vxc

Kohn-Sham Orbitals from Quantum Monte Carlo Densities: H₂ dissociation



Variational Monte Carlo and Density Functional Theory Kohn-Sham Orbitals from Quantum Monte Carlo Densities: H ₂ O									
Train chain orbitato nom quantum monte ourio Derionited. H20									
1a ₁	$2a_1$	1b ₁	3	a ₁	1b ₂				
H_2O Eigenvalues									
Density	$\int \rho v_{xc} d\tau$	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5			
I						-0.4663			
${ m CI}^{a,d,f}$	-14.626	-19.117	-1.180	-0.721	-0.584	-0.507			
$\mathrm{CI}^{a,e,f}$	-14.630	-19.073	-1.184	-0.725	-0.588	-0.511			
$\mathrm{CI}^{a,c}$	-14.628	-19.007	-1.181	-0.725	-0.588	-0.511			
$\mathrm{VMC}^{b,e,f}$	-14.287	-19.072	-1.168	-0.701	-0.567	-0.483			
$\mathrm{VMC}^{b,c}$	-13.852	-18.882	-1.151	-0.706	-0.567	-0.490			
$VMC+ACS^{c}$	<i>.g</i> -14.789	-19.204	-1.148	-0.699	-0.560	-0.481			

VMC results converge much faster than CI with respect the basis set: TZ vs 6Z

Calculation of "Exact" exchange correlation potential should be possible also for large molecules using 3D grids

Varsano, Barborini and Guidoni, J. Chem. Phys. 2014, 140, 054102



M.Barborini, S. Sorella, LG, submitted

		Singlet-Triplet gap				
		Vertical Adiaba				
		$^{1}A_{g} \rightarrow {}^{3}B_{1u}$ [eV]	$^{1}A_{g} \rightarrow {}^{3}A_{1}$ [eV]			
		Results				
	VMC(all-ele)	4.598(16)	2.843(16)			
	VMC(SR-ECP)	4.626(5)	2.862(5)			
1-00	LRDMC(all-ele)	4.642(19)	2.919(19)			
	LRDMC(SR-ECP)	4.626(10)	2.892(10)			
121.30	Quan	Quantum Chemistry calculations				
	MP2 ³¹	4 549	2,970			
	B3LYP ³¹	4.462	2.667			
1.34	MR-CI ²⁵	4.597	2.780			
121.30	CCSD(T) ³⁰	4.514	2.845			
	DMC(HF all) ³¹	4.501(26)	2.875(18)			
	DMC(GVB) ²⁸	4.6027(13)	2.9102(61)			
1.09		Experimental val	ues			
	PD ^(b)		2 52(13)			
	EELS ^(c)	4.2	2.52(15)			
	$\Pi S^{(d)}$	4.3				
	EELS ^(e)	4.32				
	$OS^{(f)}$	4.60				

M.Barborini, S. Sorella, L.Guidoni, JCTC 8, 1260-1269 (2012) CCSD(T) but higher than experiment

Ground state and triplet excited state of ethylene

	1	30	3 .	3D 1 4	3 . 1 .
Basis set	$^{1}A_{g}$	$^{S}\mathbf{B}_{1u}$	$^{3}A_{1}$	$^{3}\mathbf{B}_{1u} \leftarrow ^{1}\mathbf{A}_{g}$	${}^{J}A_1 \leftarrow {}^{I}A_g$
	[Hartree]	[Hartree]	[Hartree]	[eV]	[eV]
			VMC		
8s4p1d (n=8)/(n=9)	8s4p1d (n=8)/(n=9) -78.5217(4) -		-78.4171(4)	4.528(16)	2.846(16)
9s5p2d (n=8)/(n=9)	-78.5249(4)	-78.3578(4)	78.3578(4) -78.4190(4)		2.882(16)
8s4p1d (n=12)	-78.5287(4)	-78.3595(4)	-78.4182(4)	4.604(16)	3.007(16)
9s5p2d (n=12)	-78.5303(4)	-78.3613(4)	-78.4207(4)	4.598(16)	2.982(16)
8s4p1d (AGP) -78.5246(4)		-78.3561(4)	-78.4173(4)	4.585(16)	2.920(16)
			LRDMC		_
8s4p1d (n=8)/(n=9)	-78.5663(5)	-78.3999(6)	-78.4588(6)	4.527(21)	2.925(21)
9s5p2d (n=8)/(n=9) -78.5685(6)		-78.4015(5)	-78.4612(5)	4.544(21)	2.919(21)
8s4p1d (n=12)	-78.5717(6)	-78.4013(5)	-78.4600(5)	4.637(21)	3.039(21)
9s5p2d (n=12)	-78.5743(5)	-78.4037(3)	-78.4619(3)	4.642(19)	3.058(19)
8s4p1d (AGP)	-78.5726(5)	-78.4019(5)	-78.4631(5)	4.645(19)	2.979(19)
GVB-2(cc-pVTZ) ^(b)	-78.56236(18)	-78.39401(18)	-78.45124(18)	4.581(7)	3.024(7)
HF(cc-pVDZ/cusp) ^(c)	<u> </u>	-	-	4.501(17)	3.014(26)
MR-CISD+Q $^{(d)}$	-78.451288	-78.285126	-	4.52	-
MR-AQCC ^(d)	-78.432429	-78.266543	-	4.51	-

Table 3: All-electron energy calculations^(a)

Both at VMC and DMC level the JAGP variational wave function has a significantly lower energy with respect other approaches

M.Barborini, S. Sorella, LG, submitted

How to get harmonic frequencies and Precise geometries from error-affected Potential Energy Surfaces?

- Aim: vibrational spectroscopy by Quantum Monte Carlo
- Multidimensional fitting of the potential energy surface and forces
- Normal modes in internal coordinates
- Extract harmonic frequencies and equilibrium geometry
- Anharmonic coefficients (up to 4° order) can be also calculated and compared to experiments

 $V(q) = F + F_i q^i + F_{ik} q^i q^k + F_{ikl} q^i q^k q^l + F_{iklm} q^i q^k q^l q^m$



The VMC optimised structure of the water monomer

Table 4. VMC Evaluation of the Equilibrium Configuration, the Harmonic and the Fundamental Frequencies of the Water Molecule, Compared with Other Accurate *Ab Initio* Evaluations and Experimental Results^{*a*}

	equilibrium structure		harmonic freq. [cm ⁻¹]			fundamental freq. [cm ⁻¹]		
	<i>r</i> ₀ [Å]	ϕ_0 [deg]	ω_2	ω_1	ω ₃	$\nu_2[010]$	$\nu_1[100]$	$\nu_{3}[001]$
function/core								
JDFT/ECP	0.95497(3)	104.49(2)	1664(2)	3882(2)	3995(3)	1610(1)	3693(2)	3787(3)
JSD/ECP	0.95426(3)	104.74(1)	1670(2)	3892(3)	4006(3)	1617(1)	3702(3)	3794(2)
JAGPn*/ECP	0.95612(8)	104.17(2)	1710(3)	3896(6)	3990(7)	1654(1)	3710(3)	3800(7)
JAGP/ECP	0.95550(4)	104.41(1)	1669(1)	3872(3)	3974(4)	1613.3(6)	3677(2)	3772(2)
JSD/NCP	0.95536(3)	104.85(1)	1668(2)	3889(2)	4001(3)	1613.1(9)	3700(2)	3796(3)
JAGP/NCP	0.95668(3)	104.52(1)	1663(2)	3869(2)	3973(3)	1610.4(7)	3679(2)	3767(3)
method/basis								
BLYP/aug-cc-pVTZ ^b	0.9719	104.47	1596	3655	3757	1543	3480	3567
B3LYP/aug-cc-pVTZ ^b	0.9619	105.08	1627	3796	3899	1575	3631	3720
FC MP2/aug-cc-pVTZ ^b	0.9614	104.11	1628	3822	3948	1578	3653	3767
$CISD/(13,8,4,2/8,4,2)^{c}$	0.952	104.8	1676.1	3947.3	4050.5			
CCSD/(13,8,4,2/8,4,2) ^c	0.956	104.4	1662.5	3870.9	3977.8			
$FC CCSD(T)/aug-cc-pV7Z^d$	0.95831	104.452	1649.83	3835.55	3946.05	1595.58	3659.31	3757.45
experiment ^e	0.95721(30)	104.522(50)	1648.47	3832.17	3942.53	1594.59	3656.65	3755.79

^{*a*}For the VMC results, the equilibrium configuration, the harmonic frequencies ω_i and the fundamental frequencies ν_i have been evaluated from the PES fitted using the VMC forces; see details in the text and in Zen et al.^{51 *b*}From ref 99. ^{*c*}From ref 74. ^{*d*}From ref 75. ^{*e*}From ref 68.



Zen, Zhelyazov, Guidoni 2012, *JCTC* 8:4204–4215 Zen, Luo, Sorella, Guidoni 2013 *JCTC* 9:4332–4350.


Diffusion Monte Carlo further improves

Energies and forces along the asymmetric stretching mode



Zen, Zhelyazov, Guidoni 2012, *JCTC* 8:4204–4215 Zen, Luo, Sorella, Guidoni 2013 *JCTC* 9:4332–4350.





Conjugated Molecules and Bond Length alternation

- Biological Pigment / Dyes / Chromophore are mostly conjugated organic molecules ... C-C=C-C=C- ...
- Spectral Tuning
 - By changing the environment
 - By changing the geometry
- Electronic excitations \propto Bond lenght alternation



 Difference between average single and average double bond lenghs

- Difference between central single and central double bond in polyacetilene fragments

JAGP seems is an appropriate ansatz



^aThe three chemical configurations separated also by structural conformations: (a) ground state configuration;
(b) resonating electronic configuration with identical carbon bonds¹; (c) inversion of the single and double bonds, diradical state, with two localized electrons on the ending CH₂ groups².

Polyacetilene with Gutzwiller ansatz

- Horsch, P. Phys. Rev. B 1981, 24, 7351.
- Baeriswyl, D.; Maki, K. Phys. Rev. B 1985, 31, 6633.

The smallest system: butadiene

Bond length alternation (BLA) of 1,3-butadiene

s-trans-1,3-butadiene



The B.L.A. = $R_{C-C} - R_{C=C}$ obtained with VMC and the JAGP or JSD ansätze is 0.01 Å greater than the one obtained through CCSD(T) calculations.

Are the single and double bonds well described?

Geometries of s-trans-1,3-butadiene with ab initio methods^a.

	R_{C-C} [Å]	$R_{C-C}[Å]$	BLA [Å]
JAGP (AE)	1.3345(5)	1.4593(8)	0.1247(10)
JSD (AÉ)	1.3318(5)	1.4585(8)	0.1267(10)
$J_1 AGP_1$ (ECP)	1.3326(10)	1.4583(10)	0.1257(20)
J ₁ SD ₂ (ECP)	1.3309(10)	1.4554(10)	0.1246(20)
$J_1 AGP_2$ (ECP)	1.3323(10)	1.4582(10)	0.1259(20)
$J_2 AGP_2$ (ECP)	1.3326(10)	1.4577(10)	0.1251(20)
$J_3^-AGP_2^-$ (ECP)	1.3328(10)	1.4572(10)	0.1244(20)
CCSD ^c	1.3429	1.4663	0.1234
CCSD(T) ^c	1.3494	1.4648	0.1154
CCSD(T) (CBS) ^b	1.3389	1.4549	0.1160
Semi-Exp	1.3376(10)	1.4539(10)	0.1163(14)
SED	1.3439(5)	1.4672(13)	0.1233(14)
ED	1.341(2)	1.463(3)	0.122(3)
ED	1.349(1)	1.467(2)	0.118(2)
MFTS	1.337	1.467	0.130
ED	1.337	1.483	0.146

^{*a*}All-electron (AE) and pseudopotential (ECP), JAGP and JSD calculations, with the various basis sets described in Table 1S; ^{*b*}CCSD(T)/CBS calculations with various corrections from ref. 1; ^{*c*}This Work, with the 6-31G^{*} basis set;

The smallest system: butadiene

	$R_{C=C}$ [Å]	\mathbb{R}_{C-C} [Å]	BLA [Å]	$MP3^d$	1.3396	1.4646	0.1250
$J_1AGP_1 (ECP)^a$	1.3326(2)	1.4583(4)	0.1257(4)) MP4(SDTQ) ^{d}	1.3493	1.4639	0.1227
$CASSCF^{b}$	1.345	1.463	0.118	$B3LYP^{c}$	1.339	1.456	0.117
$RASSCF^{b}$	1.351	1.468	0.117	B3LYP ^f	1.3339	1.4527	0.1188
$CASSCF^{c}$	1.344	1.467	0.123	B3LYP ^h	1.3344	1.4527	0.1183
CASPT2 ^c	1.348	1.454	0.106	Semi-Exp ⁱ	1.3376(10)	1,4539(10)	0.1163(14)
HF^{d}	1.3225	1.4676	0.1450	SED^{j}	1.3439(5)	1.4672(13)	0.1233(14)
CISD^d	1.3316	1.4631	0.1315	ED^k	1.341(2)	1.463(3)	0.122(3)
CISDT^d	1.3340	1.4627	0.1287	ED^l	1.349(1)	1.467(2)	0.118(2)
$CCSD(T)/CBS^{e}$	1.3389	1.4549	0.1160	$MFTS^m$	1.337	1.467	0.130
CCSD^{d}	1.3429	1.4663	0.1234	ED ¹	1.33/	1.483	0.146
$CCSD(T)^d$	1.3494	1.4648	0.1154	^a Variational Monte	a Carlo regults from	ref 13. ^b CASSC	$F(4.8)/(6.31C^*+3n.3)$

0.1151

0.1146

0.1229

0.1147

0.113

0.1132

0.1137

 $CCSDT^{d}$

 $OCISD^d$

 $MP2^{c}$

 $MP2^{f}$

 $MP2^{g}$

 $\text{CCSDT}(2)_{O}^{d}$

 $QCISD(T)^d$

1.3496

1.3500

1.3436

1.3503

1.343

1.3401

1.3425

1.4648

1.4646

1.4664

1.4650

1.456

1.4533

1.4562

^aVariational Monte Carlo results from ref. 13; ^bCASSCF(4,8)/6-31G*+3p and RASSCF(22,9+5+12)[1,1]/6-31G*+3p are taken from ref. 9; ^cCalculations with 6-31G* basis set from ref. 41; ^dThis Work, with the 6-31G* basis set; ^eCCSD(T)/CBS calculations with various corrections from ref. 10; ^fCalculations with the cc-pVTZ basis set from ref. 7; ^gCalculations with 6-31G* basis set from ref. 12; ^hCalculations with auc-cc-pVTZ basis set from ref. 11; ⁱSemi-experimental results from ref. 7; ^jSelection electron diffraction from ref. 6; ^kElectron diffraction from ref. 5; ^lElectron diffraction from ref. 4; ^mMicrowave Fourier transform spectrum from ref. 3; ⁿElectron diffraction from ref. 2; ^oCalculations from ref. 14 with the 6-31G* basis set.

The smallest system: butadiene

Why CCSD(T) BLA is 0.01 Å shorter?

- Basis set < 0.001 Å
- All electron / pseudo ≈ 0.001 Å
- Dynamical correlation

- Static correlation



The VMC structure is lower in erergy with respect to The CCSD(T) structure at JSD and JAGP LRDMC level



Static correlation

Expanded the AGP in active space, like for CASSCF

$$\Phi_{G}^{mol}(\mathbf{x}_{i};\mathbf{x}_{j}) = \sum_{l,k} \tilde{\lambda}_{lk} \varphi_{k}(\mathbf{r}_{i}) \varphi_{l}(\mathbf{r}_{j}) |0,0\rangle$$

Orbital symmetry is kept during optimization



Static correlation							
CASSCF and NEVPT2		$R_{C=C}$ [Å]	\mathbb{R}_{C-C} [Å]	BLA [Å]	E [Hartree]		
	CASSCF						
	(2,2)(2,4)(2,6)(2,8)(4,4)(6,6)(8,8)	1.3257 1.3257 1.3257 1.3257 1.3257 1.3467 1.3467 1.3474 1.3469	1.4708 1.4707 1.4709 1.4709 1.4664 1.4863 1.4650	0.1451 0.1450 0.1452 0.1452 0.1452 0.1196 0.1389 0.1181	-154.96138 -154.96573 -154.96655 -154.96658 -154.98935 -155.01049 -155.02658		
	(10,10) (12,12)	1.3490	1.465 <i>3</i> 1.4639	0.1164	-155.03860		
		NEVPT2					
	(2,2) (2,4) (2,6) (2,8) (4,4) (6,6)	1.3533 1.3438 1.3426 1.3419 1.3542 1.3525	1.4688 1.4740 1.4745 1.4748 1.4627 1.4744	0.1155 0.1302 0.1319 0.1329 0.1084 0.1219	-155.48081 -155.47734 -155.47193 -155.46602 -155.48646 -155.48003		
	(8,8)	1.3535	1.4640	0.1105	-155.48132		
Static correlation matters							

		Static c	orrelatio	n	
JAGP _{AS}		$\mathbf{R}_{C=C}$ [Å]	\mathbb{R}_{C-C} [Å]	BLA [Å]	E [Hartree]
		from HF	orbitals and str	uctures	
	JSD JAGPmo _(4,4) JAGPmo _(4,13) JAGPmo _(22,22)	1.3307(4) 1.3339(3) 1.3327(3) 1.3330(3)	1.4561(7) 1.4508(7) 1.4551(4) 1.4557(4)	0.1254(8) 0.1168(8) 0.1224(5) 0.1227(5)	-26.2636(1) -26.2699(1) -26.2732(1) -26.2733(1)
	from B3LYP orbitals and structures				
	JSD JAGPmo _(4,4) JAGPmo _(4,13) JAGPmo _(22,22)	1.3302(2) 1.3341(2) 1.3330(3) 1.3327(4)	1.4549(6) 1.4502(6) 1.4555(5) 1.4553(5)	0.1247(6) 0.1161(6) 0.1225(6) 0.1226(6)	-26.2637(1) -26.2699(1) -26.2732(1) -26.2732(1)
	from JAGP projected orbitals and structures				
	JSD JAGPmo _(4,4) JAGPmo _(4,13) JAGPmo _(22,22)	1.3305(3) 1.3335(3) 1.3328(3) 1.3329(3)	1.4556(6) 1.4537(9) 1.4560(4) 1.4560(5)	0.1251(7) 0.1202(9) 0.1232(5) 0.1231(6)	-26.2634(1) -26.2719(1) -26.2732(1) -26.2731(1)
	JAGP reference				
	J_3AGP_2	1.3328(3)	1.4572(5)	0.1244(6)	-26.2734(1)
Static correlation matters					

The smallest system: butadiene

Why CCSD(T) BLA is 0.01 Å shorter?

- Basis set < 0.001 Å
- All electron / pseudo ≈ 0.001 Å
- Dynamical correlation VMC energies lower then CCSD(T)
 - LRDMC prefers VMC structure
- Static correlation is important but not trivial

We are confident we have the right wavefunction ansatz and enough dynamic correlation!



A Strongly correlated problem: Bond Lenght Alternation (BLA) in polyacetilene



Bond Lenght Alternation (BLA) in polyacetilene fragments by QMC





Rhodopsin Photocycle: structure & spectroscopy

The light cycle can be monitored following the peak of the absorbed light. Detailed structural information are known only for the starting



A minimal model for the retinal





Quantum Monte Carlo / Molecular Mechanics (QMC/MM)

Quantum Monte Carlo on Retinal in Rhodopsin and gas phase



10 13 \oplus 15 Lys296





E. Coccia, D. Varsano and L.Guidoni, JCTC (2013)

Quantum Monte Carlo on Rhodopsin





Coccia, Varsano and Guidoni, JCTC (2013)



The Peridinin-Chlorophyll-a-Protein (PCP)

- The PCP complex is a water-soluble antenna from dinoflagellate Amphidinium carterae
- **ReFolded** PCP (RFPCP) → X-rays resolution at 1.4 Å [Schulte, PNAS, 2009]
- It has the highest Carotenoid:chlorophyll ratio \rightarrow Peridinin: Chlorophyll = 4:1
- It contains:
- > 32 kDa protein chain of 313 aa arranged in 16
- α -helices
- ▶ 8 peridinins (PID)
- ▶ 2 chlorophylls-a (CHL)



[1] T. Schulte et al., PNAS, 1-6, 2009

Α

B



The roles of Peridinin (PID): light harvesting and photoprotection



PID: main responsible of the light absorption in PCP

Förster Resonance Energy Transfer (FRET) from donor (PID) to acceptor (CHL)

PID613 Photoprotection role: Quenching of CHL triplet states Dexter Triplet-Triplet Energy Transfer from CHL to PID

- S_1 : dark state
 - PID excitation $S_0 \rightarrow S_2$
 - Energy transfer from PID to CHL $S_2 \rightarrow Q_x (\approx 25 \%)$
 - PID relaxation to S_1/S_{ICT} (*internal charge transfer*)

• Energy transfer from PID to CHL $S_1/S_{ICT} \rightarrow Q_v (\approx 63 \%)$

Energy transfer channels
 Intramolecular relaxation
 Chl-a Fluorescence

67

[2] D. Zigmantas et al., PNAS, 99, 16760-16765, 2002

QMC geometry optimization on Peridinin in gas phase

- All-trans polyenes (C_{2h} symmetry):
 - 1A_g ground state
 - 2Ag excited state
 - 1B_u excited state ($\pi \rightarrow \pi^*$)
- $E(2A_g) < E(1B_u)$ for polyenes with no. of double bonds ≥ 4
- Single-double inversion in $\pi \to \pi^*$ due to:
 - double character $mm \rightarrow n'n'$ coupling with covalent configurations (lowering 2A_g energy)
 - Iarge singlet-triplet splitting (pushing 1B_u up in energy)
- TDDFT deficiencies for excitation with double-excitation character (CASPT2, DFT/MRCI, SAC-CI, EOM-CCSD ok?)
- "...many attempts have been made to find a consistent theoretical treatment of the different excited states in linear polyenes that is also applicable to larger systems, such as carotenoids. But this is still a challenging task for the available quantum chemical methods." (C. König and J. Neugebauer, ChemPhysChem, vol. 13, p. 386 (2012))

QMC geometry optimization on Peridinin in gas phase



QMC geometry optimization on Peridinin in gas phase

• S₁ and S₂ states

S ₀ Geometry	Excitations	ΔE (eV)	f
VMC	TDDFT/B3LYP	2.52	2.95
		3.19	0.52
VMC	MBGFT	2.62	large
		3.12	small
B3LYP	TDDFT/TDA/SVWN	2.26	1.00
		2.41	1.87
B3LYP	TDDFT/SVWN	2.09	2.61
		2.33	0.46
B3LYP	DFT/MRCI	2.26	0.28
		2.54	3.16
	Exp	2.0-2.3	dark
		2.72	intense

Coccia, Varsano and Guidoni, JCTC 10, 501 (2013)



Ab initio molecular dynamics by Quantum Monte Carlo

Error on forces is anyway too high for deterministic integration algorithms

Second order Langevin dynamics $\dot{\boldsymbol{v}} = -\gamma \boldsymbol{v} + \boldsymbol{f}(\boldsymbol{R}) + \boldsymbol{\eta}(t)$ $\dot{\boldsymbol{R}} = \boldsymbol{v}$

Successfully developped and applied by Sandro Sorella's group

G Mazzola, A Zen, and S Sorella Finite-temperature electronic simulations without the Born-Oppenheimer constraint *The Journal of chemical physics* 137 (13), 134112

G Mazzola, S Yunoki, and S Sorella Unexpectedly high pressure for molecular dissociation in liquid hydrogen by electronic simulation *Nature Communications* 5, 3487, 2014
Quantum Monte Carlo on $Fe_2S_2((SH)^{2-})_4$

				Tabl repo ener	Table. Energies and $\langle S^2 \rangle$ values of Ferodoxin states. The values of $\langle S^2 \rangle$ are reported in <i>italic</i> below the total energies of the molecule in Hartree. The total energies are obtained from pseudopotential wave functions.				
Table. Compa	rison between	optimized Ferre	edoxin				HS Structure	BS Structure	
structures.	R _{Fe} –Fe	R _{Fe-S} High Spin	θ_{S-Fe-S}	B3LYP		HS	-311.621962 30.0408	-311.618621 <i>30.0423</i>	
JSD B3LYP	3.0033(2) 3.02642	2.2904(2) 2.29860	98.05(1) 97.6567		BS	-311.634858 <i>4.8110</i>	-311.637555 <i>4.7420</i>		
JSD	BROKEN SYMMETRY JSD 2.8505(3) 2.2572(2) 101.64(1)					Δ <i>E</i> [eV] J [cm ⁻¹]	0.3509 -112.17	0.5152 -164.23	
B3LYP PBE ² PBE+U ²	2.84386 2.68 2.78	2.25175 2.20 2.24	101.6805	J+SD	HS	VMC results -311.5905(3) .30.0300(1)	-311.5860(3) 30.0322(1)		
PBE ²	2.62	Low Spin 2.17			J+SD	BS	-311.5989(3) 4.8704(5)	-311.6017(3) <i>4.8488(5)</i>	
pbe+0² B3lyp1	2.72 2.81	2.22 2.26	102.79			∆ <i>E</i> [eV] J [cm ^{−1}]	0.228(12) -73(8)	0.427(12) -136(8)	
The B3LYP value of J in Low spin state is -186 cm ⁻¹¹ . CASSCF(BS+U) -137 cm ⁻¹ , CASSCF(BS+U) -159 cm ⁻¹ . The experimental value for the [Fe ₂ S ₂]Cys ₄ in protein is of -182(20) cm ^{-11,2} .									

(1) Bovi D. and Guidoni L. J. Chem. Phys. 2012, 137, 114107; (2) Nair N. N. et al.; J. Chem. Theory Comput. 2010, 6, 569

"E pur si muove (and yet it moves)" (Galileo Galilei)



Second order Langevin dynamics of 32 water molecules at 300K with Variational Monte Carlo G. Mazzola, A. Zen, Y. Luo, L. Guidoni, and S. Sorella in preparation (2014)

QMC water radial distribution function





JAGP alleviates **QMC** historical limitations

Variational MC based on JAGP wavefunction in TurboRVB

Computational Cost (large prefactor)

Ionic Forces

Lack of standards (Wavefunction ansatze, basis set, properties) Multiconfigurational at prize of 1 determinant

- Compactness and full optimization
- Forces at 6x the price of energies
- Dynamics!
- Standards are now set up with JAGP
- Molecular properties
- Static correlation / size consistency

Fixed node is an approximation

• High-level Variational Wavefunction: in the mayority of the cases DMC is not necessary!

Conclusion and Perspectives

- Accurate QMC energies, geometries, polarizabilities, harmonic and anharmonic frequencies
- QMC: Strong scalability on new generation parallel machines. Geometry optimization100 atoms are now feasible (with PRACE)
- JAGP wavefunction is a compact and precise correlated ansanz even at the variational level. radicals, diradicals
- QMC/MM in combination with MBGFT/MM as reference methods for biological chromophores
- Kohn-Sham orbitals from QMC densities
- QMC ab initio MD

Perspectives:

- Transition metal complexes,
- Embedding method for mixed QMC/DFT and QMC/DFT/MM



Computational Biophysics Biochemistry and Chemistry Group

Dipartimento di Scienze Fische e Chimiche Università degli Studi de L'Aquila, Italy Dipartimento di Fisica Sapienza Università di Roma, Italy

http://bio.phys.uniroma1.it

Emanuele Coccia Andrea Zen Daniele Narzi Daniele Bovi Shibing Chu Barbara Gregori Maria Montagna Matteo Barborini Fabio Pitari

Many thanks to Sandro Sorella and the TurboRVB team!

Mohammad Islam Giorgia Pinna Andrea De Luca Pietro De Gaetano Gyamfi Kwame Atta Caterina de Franco







END