

# Size-Extensive Wave Functions for Quantum Monte Carlo: A Linear Scaling Generalized Valence Bond Approach

Francesco Fracchia,<sup>\*,†</sup> Claudia Filippi,<sup>\*,‡</sup> and Claudio Amovilli<sup>\*,†</sup>

<sup>†</sup>Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Risorgimento 35, 56126 Pisa, Italy

<sup>‡</sup>MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

## S Supporting Information

**ABSTRACT:** We propose a new class of multideterminantal Jastrow–Slater wave functions constructed with localized orbitals and designed to describe complex potential energy surfaces of molecular systems for use in quantum Monte Carlo (QMC). Inspired by the generalized valence bond formalism, we elaborate a coupling scheme between electron pairs which progressively includes new classes of excitations in the determinantal component of the wave function. In this scheme, we exploit the local nature of the orbitals to construct wave functions which have increasing complexity but scale linearly. The resulting wave functions are compact, can correlate all valence electrons, and are size extensive. We assess the performance of our wave functions in QMC calculations of the homolytic fragmentation of N–N, N–O, C–O, and C–N bonds, very common in molecules of biological interest. We find excellent agreement with experiments, and, even with the simplest forms of our wave functions, we satisfy chemical accuracy and obtain dissociation energies of equivalent quality to the CCSD(T) results computed with the large cc-pV5Z basis set.

## 1. INTRODUCTION

Quantum Monte Carlo (QMC) methods are a large class of techniques to stochastically solve the Schrödinger equation.<sup>1</sup> In both variational (VMC) and diffusion Monte Carlo (DMC) calculations, a key step is the construction of a trial correlated many-electron wave function, which is commonly chosen of the Jastrow–Slater form, namely, a Jastrow correlation factor explicitly depending on the interelectronic distances, multiplied by a linear combination of Slater determinants.<sup>2</sup> The use of a complete active space (CAS) expansion in the determinantal component offers a good description of static correlation and has been successfully employed also in QMC to obtain accurate absolute<sup>3</sup> and excitation<sup>4,5</sup> energies. Unfortunately, CAS wave functions have significant limitations in the description of complex potential energy surfaces (PES) since a reduced active space does not correlate all the electrons and can only be applied to a small portion of the PES, while a large active space quickly yields a very high number of determinants. Consequently, these large expansions must be truncated for use in QMC, and the truncation introduces an element of arbitrariness difficult to control when the molecules dissociate. Moreover, the complete active space self-consistent field (CASSCF) calculation to set up the Jastrow–Slater wave function may become a bottleneck in the QMC calculation if all valence electrons are correlated in molecules with more than a few atoms. In order to overcome these obstacles, an alternative route is to exploit the chemical intuition due to Lewis<sup>6</sup> and describe the molecular structures in terms of electron pairs on bonds or lone pairs. This idea has played a key role in the history of chemistry because it captures many important features of the electronic distribution in molecules. Wave functions constructed following this principle have been proposed since the early days of quantum chemistry<sup>7</sup> and have also been recently introduced in QMC calculations.

Anderson et al.<sup>8</sup> have used perfect pair generalized valence bond (GVB-PP) determinantal components in Jastrow–Slater wave functions, while Braida et al.<sup>9</sup> have employed valence bond (VB) functions constructed with nonorthogonal orbitals localized on the atoms. Both approaches have produced promising results but the number of determinants in these wave functions scales exponentially with respect to the number of electron pairs of the molecular system. Other pairing wave functions employed in QMC are the Pfaffians,<sup>10</sup> of which the antisymmetrized geminals power (AGP)<sup>11</sup> is a particular case. The single-determinant form of AGP leads to a lower computational cost but reduced variational flexibility. It is important to stress that the VB and CASSCF approaches are not in conflict, as a CASSCF wave function can be rewritten in terms of a simpler and compact VB form.<sup>12–15</sup> The pairing wave functions (VB, GVB-PP, AGP, etc.) are however more suitable for QMC calculations since they allow one to perform a balanced truncation for the treatment of complex chemical systems.

In this paper, we propose multideterminantal Jastrow–Slater trial functions for use in QMC, where the determinantal part is inspired by the GVB-PP form<sup>16,17</sup> but the number of configuration state functions (CSFs) scales linearly with the size of the system. In order to achieve linear scaling, we exploit the localization of the orbitals, which emerges naturally in the GVB method when the wave function is optimized. The localization of the orbitals allows us to classify the CSFs according to the electron pairs they correlate. Since we find that the correlation between adjacent pairs is most important, we adopt a geometric criterion for truncation of the wave function that respects size extensivity. We further classify the CSFs

Received: February 9, 2012

Published: May 11, 2012

according to the occupations of the bonding and antibonding orbitals of the electron pairs, and for a representative set of molecular systems, we identify a hierarchy of importance for these classes by analyzing the CISDTQ wave functions built with localized orbitals. The gradual introduction of these classes of CSFs allows the construction of an ordered set of wave functions of increasing complexity. To assess the performance of our wave functions, we compute the dissociation energy of  $\text{N}_2\text{H}_4$ ,  $\text{HNO}_2$ ,  $\text{CH}_3\text{OH}$ , and  $\text{CH}_3\text{NH}_2$ . The fragmentation of these molecules requires a balanced description of the N–N, N–O, C–O, and C–N bonds, which are among the most common bonds in biological molecules. In all cases, we find that, already with the simplest forms of our wave functions, we satisfy chemical accuracy and obtain dissociation energies in excellent agreement with experiments and with the CCSD(T) results computed with the large cc-pVSZ basis set.

## 2. METHOD

**2.1. Linear Scaling GVB Wave Functions.** For a molecular system with an even number  $N$  of electrons, the GVB-PP wave function can be expressed as the antisymmetrized product of geminal functions:

$$\Psi_{\text{GVB-PP}} = |(\varphi_{1u}\bar{\varphi}_{1v} - \bar{\varphi}_{1u}\varphi_{1v})(\varphi_{2u}\bar{\varphi}_{2v} - \bar{\varphi}_{2u}\varphi_{2v})\dots(\varphi_{(N/2)u}\bar{\varphi}_{(N/2)v} - \bar{\varphi}_{(N/2)u}\varphi_{(N/2)v})| \quad (1)$$

where each geminal  $j$  is a pairing function constructed with the Coulson–Fisher orbitals,<sup>18</sup>  $\varphi_{ju}$  and  $\varphi_{jv}$ . These orbitals are semilocalized atomic orbitals, i.e., they are localized on an atom but have small delocalization tails on other close atoms. In the most common formulation of the GVB-PP theory, the Coulson–Fisher orbitals of two different electron pairs are constrained to be orthogonal:

$$\begin{aligned} \langle \varphi_{ju} | \varphi_{j'v} \rangle &\neq 0, \\ \langle \varphi_{ju} | \varphi_{j'v} \rangle &= \delta_{jj'} \langle \varphi_{ju} | \varphi_{jv} \rangle \end{aligned} \quad (2)$$

This restriction is known as the strong orthogonality.

Each geminal can be rewritten in terms of orthogonal bonding and antibonding orbitals which are a linear combination of the Coulson–Fisher orbitals, namely,

$$\{\phi_{jb}, \phi_{ja}\} = \{\varphi_{ju}, \varphi_{jv}\} \begin{bmatrix} \frac{1}{\sqrt{2(1+S_j)}} & \frac{-1}{\sqrt{2(1-S_j)}} \\ \frac{1}{\sqrt{2(1+S_j)}} & \frac{1}{\sqrt{2(1-S_j)}} \end{bmatrix} \quad (3)$$

where

$$S_j = \langle \varphi_{ju} | \varphi_{jv} \rangle \quad (u \neq v) \quad (4)$$

In this form, the wave function above becomes

$$\Psi_{\text{GVB-PP}} = |\Phi_0\rangle + \sum_{j=1}^{N/2} q_j |\Phi_{ib\bar{a}}^{ia\bar{a}}\rangle + \sum_{i<j}^{N/2} q_i q_j |\Phi_{ib\bar{a},jb\bar{a}}^{ia\bar{a},ja\bar{a}}\rangle + \dots \quad (5)$$

where the indices  $b$  and  $a$  denote bonding and antibonding orbitals and the sum is extended to all possible pair-excitation determinants. The  $q_j$  parameters define the mixing of bonding and antibonding geminals in each pair and are given by

$$q_j = \frac{S_j - 1}{1 + S_j} \quad (6)$$

If we truncate the expansion of eq 5 to the double excitations, we obtain a function consisting of  $(N/2) + 1$  determinants:

$$\Psi_{\text{LGVB1}} = c_0 |\Phi_0\rangle + \sum_{i=1}^{N/2} c_i |\Phi_{ia\bar{a}}^{ib\bar{b}}\rangle \quad (7)$$

This wave function is extremely compact, scales linearly with respect to the size of the system, and can correlate all valence electrons of the molecule. We denote this function as LGVB1, and we use it as starting point to generate Jastrow–Slater wave functions of increasing complexity. Through the use of localized orbitals, we develop a coupling scheme between electron pairs which progressively includes new classes of excitations in the determinantal component of the wave function. In the most extended formulation, our wave functions correlate each couple of adjacent electron pairs in a CAS(4,4) form transcending the perfect pair approximation.

**2.2. Theory.** If we perform a MCSCF calculation including the CSFs of LGVB1 wave function, the localization of orbitals (bonding, antibonding, lone pair function, and diffuse lone pair function with nodes) emerges naturally. We find that the orbitals localized through this procedure are more suitable for our purposes than those obtained through Boys,<sup>19</sup> Ruedenberg,<sup>20</sup> or other similar localization schemes. The LGVB1 orbitals are in fact the result of a variational optimization in a wave function similar to the one we will use in our QMC calculations.

If we now use these orbitals in a CISDTQ calculation, the orbital localization allows us to classify excitations according to the mutual relations between the occupied bonding and antibonding orbitals. If we order the CSFs of the CISDTQ wave function with respect to decreasing absolute values of the coefficients, we observe a hierarchy in the classes of excitations. In tests performed on a representative set of molecules ( $\text{N}_2\text{H}_4$ ,  $\text{NH}_2$ ,  $\text{H}_2\text{O}_2$ ,  $\text{OH}$ , and  $\text{H}_2\text{O}$ ), we find that this hierarchy is roughly the same among the systems considered. The most important excitations are the double excitations from the bonding orbitals to the corresponding antibonding orbitals (E1). This is not surprising since the orbitals were generated in a MCSCF calculation which includes only the E1 excitations. As already mentioned, the number of E1 excitations grows linearly with the number of electron pairs. In hierarchical order emerging from the CISDTQ wave function, the second class of excitations is double excitations constructed as pairs of single excitations from bonding orbitals to the corresponding antibonding orbitals (E2). The number of E2 excitations grows quadratically with respect to the number of electron pairs. However, the CI calculations show that only excitations involving adjacent pairs, (i.e., sharing at least one atom) have significant coefficients. Excluding the nonadjacent couplings between electron pairs, the number of CSFs increases linearly also for this class of excitations.

The importance of the E1 and E2 excitations is common to all performed tests, regardless of the spin state of the molecules considered. For the subsequent classes, a hierarchy continues to be present but differs depending on the spin state of the molecule. The single excitations from bonding orbitals to the corresponding antibonding orbital (E3) are important for molecules in doublet states but to a lesser extent for molecules in singlet states. The E3 excitations grow linearly with respect

to the size of the system. Both for singlet and doublet states, an important class of excitations consists of double excitations in which both electrons in the same bonding orbital in the reference are promoted to the same antibonding orbital of another pair (E4). The E4 excitations grow quadratically with respect to the number of electron pairs but, as in the case of E2, not all pairs give a significant contribution. Also in this case, the most important E4 excitations are those in which the antibonding orbital is from an adjacent pair. If we consider only these excitations, the number of E4 excitations grows linearly with respect to the size of the system. For singlet states, the next class of excitations in order of importance consists of the quadruple in which two pairs of electrons are excited to their antibonding orbitals (E5). The E5 excitations grow quadratically, but again, only the couplings between adjacent pairs give rise to excitations whose coefficients are important. We can then keep only adjacent pairs to obtain linear scaling also for the E5 excitations. Similarly, we can proceed with the triple excitations built as a double excitation to the respective antibonding orbital coupled to a single excitation toward its antibonding orbital (E6). Using the modular truncation based on the contiguity of electron pairs, we are then dividing the global valence active space into a set of smaller active spaces, each consisting of a couple of electron pairs. The size of these active spaces is given by four electrons in four orbitals. If we now include the excitations E7–E10 (see Table 1) in addition

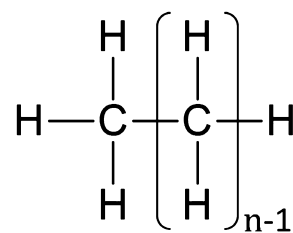
**Table 1. Classes of Excitations (E1–E10) That Complete a CAS(4,4) for Two Electron Pairs<sup>a</sup>**

code	type	occupation numbers $b_1, b_2, a_2, a_1$	$N(\text{CSF})$
E0	reference	2200	1
E1	D	0202 + 2020	$3n + 1$
E2	D	1111	$12n$
E3	S	1201 + 2110	$3n + 1$
E4	D	0220 + 2002	$12n$
E5	Q	0022	$6n$
E6	T	1021 + 0112	$12n$
E7	S	1210 + 2101	$12n$
E8	D	0211 + 2011	$12n$
E9	D	1120 + 1102	$12n$
E10	T	1012 + 0121	$12n$

<sup>a</sup>The excitations with respect to the reference (E0) are labelled as single (S), double (D), triple (T), and quadruple (Q). In the occupation number, the order of orbitals is bonding for the pair 1 ( $b_1$ ), bonding for the pair 2 ( $b_2$ ), antibonding for the pair 2 ( $a_2$ ), and antibonding for the pair 1 ( $a_1$ ). We consider the most common case of a reference (E0), where all the bonding orbitals (or lone pairs functions) are doubly occupied, and all antibonding orbitals (or diffuse lone pairs functions with nodes) are empty. We also list the total number of CSFs [ $N(\text{CSF})$ ] for all adjacent electron pairs of a molecule consisting of  $n$  atoms of the second period, not cyclic, which does not contain unsaturations.

to the ones already considered, we are correlating each couple of adjacent electron pairs in a CAS(4,4). In Table 1, we collect all the excitations  $E_n$ . Each class of excitations scales linearly with the number of atoms in the molecule. Therefore, a function which uses any combination of these excitations also scales linearly.

In the following, we consider the aliphatic alkanes shown in Figure 1 to discuss the size and scaling of the wave function by means of quantitative relations. The number of CSFs for the alkanes is listed as a function of the number of carbon atoms in



$$N_{\text{pair}} = 3n + 1$$

$$N_{\text{pair\_int}} = (9n^2 + 3n)/2$$

$$N_{\text{pair\_close\_int}} = 6n$$

**Figure 1.** Generic linear alkane with  $n$  carbon atoms. We list the number of the electronic pairs ( $N_{\text{pair}}$ ), the global number of the interactions between the electron pairs ( $N_{\text{pair\_int}}$ ), and the number of interactions between adjacent electron pairs ( $N_{\text{pair\_close\_int}}$ ). The number of CSFs in the LGVB wave functions increases linearly because they depend on  $N_{\text{pair\_close\_int}}$  and not  $N_{\text{pair\_int}}$ .

Table 1, and these relations are valid for all molecules containing elements of the second period and hydrogen, not cyclic, and not containing unsaturations. Both the presence of multiple bonds and cycles reduce the number of adjacent electron pairs and therefore the number of CSFs to be considered, while the presence of hypervalent atoms increases the number of adjacent electron pairs and therefore the number of CSFs. In general, the deviations from the relations obtained for the alkanes are minimal, and the proportionality constants are the same or change slightly.

The inclusion of all  $E_n$  excitations gives rise to a wave function which can be written as

$$\Psi_{\text{LGVB10}} = \sum_{ij}^{6n} C_{ij} |\Phi_{ij}\rangle^{\text{CAS}(4,4)} \quad (8)$$

where  $ij$  are adjacent electron pairs in an alkane molecule. As shown in Figure 1, the number of adjacent electron pairs is equal to  $6n$ , while the total number of interactions between the electron pairs is equal to  $(9n^2 + 3n)/2$ , where  $n$  is the number of carbon atoms. For every interaction between electron pairs, the number of CSFs which completes the CAS(4,4) is 19, so the total number of CSFs is  $114n + 1$ . In this way, we have however counted more times than necessary the E1 and E3 excitations, whose number does not depend on the interaction between pairs but only on the number of pairs. If we correct for this, the number of CSFs is equal to  $96n + 3$ .

The description of electronic correlation between adjacent pairs with a CAS(4,4) allows us to construct a modular wave function, balanced with respect to fragmentation and able to properly account for nondynamical correlation. If we combine this determinantal component with a Jastrow factor to account for dynamical correlation, we can build accurate Jastrow–Slater wave functions to use in QMC calculations. These Jastrow–Slater wave functions are here denoted as J-LGVB $n$  and include progressively the  $E_n$  excitations, according to the hierarchy identified in the CISDTQ calculations, up to the complete CAS(4,4) for each couple of adjacent electron pairs:

$$\Psi_{\text{trial}} = J\Psi_{\text{LGVB}n} \quad (9)$$

We obtain therefore 10 Jastrow–Slater wave functions and report the composition of each wave function and the number of CSFs in Table 2. Importantly, our formulation of LGVB $n$

Table 2. Composition of the LGVB $n$  Wave Functions<sup>a</sup>

$\Psi$	CSFs	$N(\text{CSF})$
LGVB1	E0 + E1	$3n + 2$
LGVB2	LGVB1 + E2	$15n + 2$
LGVB3	LGVB2 + E3	$18n + 3$
LGVB4	LGVB3 + E4	$30n + 3$
LGVB5	LGVB4 + E5	$36n + 3$
LGVB6	LGVB5 + E6	$48n + 3$
LGVB7	LGVB6 + E7	$60n + 3$
LGVB8	LGVB7 + E8	$72n + 3$
LGVB9	LGVB8 + E9	$84n + 3$
LGVB10	LGVB9 + E10	$96n + 3$

<sup>a</sup>The excitations  $E_n$  are defined in Table 1. We consider a molecule consisting of  $n$  atoms of the second period, not cyclic, which does not contain unsaturations and where all bonding orbitals (or lone pair) are doubly occupied and all antibonding orbitals (or diffuse lone pair with nodes) are empty in the reference (E0).

wave functions is different from the original GVB-PP scheme already at the fourth order, since we include excitations from bonding orbitals of an electron pair to antibonding of another electron pair (E4, E7–E10). Moreover, the linear parameters of all CSFs are completely free, while in the GVB-PP scheme, the parameters of the higher excitations are related to those of the double excitations (eq 5). A further important difference is the removal of any orbital orthogonality constraint.

The use of localized orbitals and the modular scheme of the excitations illustrated above ensures the size extensivity of the wave function. For the same reason, size consistency is always respected if the dissociation does not lead to breaking of chemical bonds (e.g., van der Waals, hydrogen bond). In breaking single bonds, size consistency is satisfied by the use of J-LGVB3 and higher wave functions since these functions include a CAS(2,2) for each bond. The correct dissociation of the double bonds is achieved with the J-LGVB10 wave functions because they include all the CAS(4,4) between couple of adjacent electron pairs. The breaking of a triple bond is instead properly described with a CAS(6,6) which is not included in the J-LGVB10 function, which only correlates the six electrons of the triple bond with three CAS(4,4) generated by the three possible pairings among the three electron pairs  $\pi_x$ ,  $\pi_y$ , and  $\sigma$ .

The J-LGVB $n$  wave functions can be applied to all molecular systems that can be described by a dominant Lewis structure. In addition, they are also suitable for molecules characterized by multicentric bonds, which are common in transition states and electron-deficient systems. However, when resonance between different Lewis structures is important, our formulation must be modified to include all the relevant pairing schemes. This change will lead to an increased number of CSFs but will not spoil the linear scaling of the method. For chemical systems where the concept of electronic pair fails (e.g., a cluster of Li atoms where the directionality of the bonds is lost), the method requires to be thoroughly reviewed.

The excitation scheme shown in Table 1 is not the only one possible. If a bonding orbital is singly occupied, we have the scheme shown in Table 3, while the scheme in Table 4 holds when an antibonding orbital that is singly occupied. In these cases, we do not have a CAS(4,4) for adjacent electron pairs but a CAS(3,4) and CAS(5,4), respectively. In the molecules considered here, only these three types of pairings are present, but in general, others are possible if the occupation numbers of

Table 3. Classes of Excitations ( $E_n$ ) That Complete a CAS(3,4) for 3 Electrons Occupying Two Bonding Orbitals in the Reference (2100)<sup>a</sup>

code	type	occupation numbers $b_1, b_2, a_2, a_1$
E0	reference	2100
E1	D	0102
E2	D	1011
E3	S	1101 + 2010
E4	D	0120
E5	Q	—
E6	T	0012
E7	S	1110 + 2001
E8	D	0111
E9	D	1020 + 1002
E10	T	0021

<sup>a</sup>The excitations are labelled as single (S), double (D), triple (T), and quadruple (Q) with respect to the reference (E0). In the occupation numbers, the order of the orbitals is bonding 1 ( $b_1$ ), bonding 2 ( $b_2$ ), antibonding 2 ( $a_2$ ), and antibonding 1 ( $a_1$ ). For each 2100 coupling, the corresponding CAS(3,4) produces 16 CSFs.

Table 4. Classes of Excitations ( $E_n$ ) That Complete a CAS(5,4) for 5 Electrons Occupying a Couple of Bonding–Antibonding Orbitals in the Reference (2210)<sup>a</sup>

code	type	occupation numbers $b_1, b_2, a_2, a_1$
E0	reference	2210
E1	D	0212
E2	D	1121
E3	S	1211 + 2120
E4	D	2012
E5	Q	—
E6	T	0122
E7	S	1220 + 2111
E8	D	0221 + 2021
E9	D	1112
E10	T	1022

<sup>a</sup>The excitations are labelled as single (S), double (D), triple (T), and quadruple (Q) with respect to the reference (E0). In the occupation numbers, the order of orbitals is bonding 1 ( $b_1$ ), bonding 2 ( $b_2$ ), antibonding 2 ( $a_2$ ), antibonding 1 ( $a_1$ ). For each 2210 coupling, the corresponding CAS(5,4) produces 19 CSFs.

a couple of bonding–antibonding orbitals in the reference are 1100, 2211, 2220, 2221, 1110, 1111, or 2110, where the order of orbitals is bonding of the first electron pair, bonding of the second electron pair, antibonding of the first electron pair, and antibonding of the second electron pair.

**2.3. Computational Details.** In the construction of the Jastrow–Slater LGVB1 wave functions (J-LGVB1), we optimize the orbitals, the coefficients of the CSFs, and the parameters of Jastrow factor by energy minimization within VMC using the linear method.<sup>21</sup> As initial guess for the orbitals, we use the orbitals obtained in a MCSCF calculation that includes the same CSFs present in LGVB1 and perform the MCSCF calculations with the package GAMESS-US.<sup>22,23</sup> As discussed above, the MCSCF calculations with this selection of CSFs yield localized orbitals, and the localization of the orbitals is preserved also after the optimization in the presence of the Jastrow factor without imposing any symmetry constraints. We generate the J-LGVB $n$  wave functions by adding progressively the  $E_n$  excitations, according to the hierarchy that emerges from the CISDTQ calculations (Table 1). For the J-LGVB $n$  ( $n$



**Table 5.** VMC and DMC Dissociation Energies ( $\Delta E$ , kcal/mol) of Hydrazine in Amino Radicals, Computed with Several Trial Wave Functions<sup>a</sup>

wave function		VMC			DMC		
NH <sub>2</sub>	N <sub>2</sub> H <sub>4</sub>	NH <sub>2</sub>	N <sub>2</sub> H <sub>4</sub>	$\Delta E$	NH <sub>2</sub>	N <sub>2</sub> H <sub>4</sub>	$\Delta E$
1 det	1 det	−11.0683	−22.2438	67.3	−11.0798	−22.2724	70.8
1 det	CAS(2,2)	−11.0683	−22.2460	68.7	−11.0798	−22.2738	71.6
CAS(3,3)	CAS(6,6)	−11.0707	−22.2502	68.3	−11.0812	−22.2754	70.9
CAS(7,7)	CAS(14,14)	−11.0707	−22.2558	71.8	−11.0806	−22.2780	73.3
J-LGVB1	J-LGVB1	−11.0715	−22.2538	69.5	−11.0809	−22.2762	71.8
J-LGVB2	J-LGVB2	−11.0718	−22.2565	70.8	−11.0812	−22.2777	72.4
J-LGVB3	J-LGVB3	−11.0719	−22.2566	70.8	−11.0813	−22.2777	72.2
J-LGVB4	J-LGVB4	−11.0723	−22.2580	71.2	−11.0814	−22.2787	72.7
J-LGVB5	J-LGVB5	−11.0723	−22.2580	71.1	−11.0815	−22.2788	72.6
J-LGVB6	J-LGVB6	−11.0724	−22.2581	71.2	−11.0815	−22.2789	72.7
J-LGVB7	J-LGVB7	−11.0724	−22.2581	71.1	−11.0816	−22.2789	72.6
J-LGVB8	J-LGVB8	−11.0724	−22.2581	71.2	−11.0816	−22.2790	72.6
J-LGVB9	J-LGVB9	−11.0724	−22.2584	71.3	−11.0817	−22.2792	72.6
J-LGVB10	J-LGVB10	−11.0726	−22.2585	71.1	−11.0817	−22.2794	72.8
1 det/VQZ	1 det/VQZ	−11.0695	−22.2494	69.2	−11.0798	−22.2740	71.8
1 det/VQZ	CAS(2,2)/VQZ	−11.0695	−22.2511	70.3	−11.0798	−22.2748	72.3
J-LGVB1/VQZ	J-LGVB1/VQZ	−11.0721	−22.2583	71.6	−11.0811	−22.2776	72.4
J-LGVB2/VQZ	J-LGVB2/VQZ	−11.0726	−22.2592	71.6	−11.0815	−22.2792	72.9
experimental		73.39			73.39		

<sup>a</sup>We also report the total energies (hartree) and do not include zero-point energy correction. The experimental reference is active thermochemical tables (ATcT)<sup>44</sup> data corrected for zero-point energy, spin–orbit interaction, and Born–Oppenheimer approximation.<sup>41</sup> The statistical error on the VMC and DMC total energies is 0.0001 au, while the statistical error on the corresponding dissociation energies is 0.1 kcal/mol. Unless explicitly stated, we use the VTZ basis set.

> 1) wave functions, we do not further optimize the orbitals but use the orbitals obtained in the J-LGVB1 wave functions. For the monodeterminantal wave functions, we optimize the orbitals and the parameters of the Jastrow factor by energy minimization within VMC, using B3LYP orbitals as an initial guess. We use energy minimization within VMC also to simultaneously optimize the CSF coefficients, the orbitals, and the Jastrow factor parameters of the CAS wave functions, starting from canonical orbitals. In the optimization of the orbitals in both the monodeterminantal and CAS wave functions, we maintain the constraints of symmetry on the orbitals.

The QMC calculations are performed with the CHAMP program.<sup>24</sup> We employ the BFD pseudopotentials<sup>25,26</sup> and the VTZ basis set specifically developed for these pseudopotentials. The pseudopotentials are treated beyond the locality approximation.<sup>27</sup> For the fragmentation of N<sub>2</sub>H<sub>4</sub>, we also perform tests with the VQZ basis set without the *g* functions on the nitrogen atoms. The Jastrow factors contain electron–nuclear, electron–electron, and electron–electron–nuclear terms.<sup>28</sup> We use a time step of 0.05 au in the DMC fixed-node calculations and verify that a time step of 0.025 au yields the same energy differences in the dissociation of N<sub>2</sub>H<sub>4</sub> with the J-LGVB1 and J-LGVB2 wave functions.

The coupled cluster calculations are performed in the frozen-core approximation with the CFOUR code.<sup>29</sup> The DFT results with the B3LYP,<sup>30–33</sup> PBE0,<sup>34–36</sup> M06-2X,<sup>37</sup> and M08-HX<sup>38</sup> functionals are obtained with GAMESS-US, while for the double-hybrid B2PLYP<sup>39</sup> functional, we use Gaussian09.<sup>40</sup> For the coupled cluster calculations, we employ the cc-pVnZ (*n* = T, Q, 5) basis set and, for the DFT calculations, the aug-cc-pVQZ basis set. The molecular geometries are taken from the

work by Martin et al.<sup>41</sup> and were optimized within CCSD(T)/cc-pV(Q + d)Z.

### 3. ILLUSTRATIVE EXAMPLES

The accurate estimate of the energy in an homolytic bond breaking is a severe challenge for all methods in quantum chemistry. Since the number of electronic pairs is not conserved in the reaction, we cannot rely on error cancellations to accurately compute the energy of reaction, and it is therefore essential that the computational method is capable to provide a good estimate of the electron correlation. Moreover, since these reactions involve the fragmentation of a molecule, the method must also be size extensive.<sup>42</sup> If we want to extend the applicability of QMC to molecules of biological interest, the respect of size extensivity is crucial. In a nonperiodic context, it is appropriate to adopt the one proposed by Nooijen et al. as the definition of size extensivity:<sup>43</sup> “One has to grow a repeating unit in the form of a rigid chain and show that the change in energy becomes independent of the chain length in the limit.” The modular structure of the J-LGVBn wave functions ensures size extensivity. Here, we verify the size extensivity of our J-LGVBn wave functions by calculating the QMC dissociation energy of the N–N, N–O, C–N, and C–O bonds in the N<sub>2</sub>H<sub>4</sub>, HNO<sub>2</sub>, CH<sub>3</sub>NH<sub>2</sub>, and CH<sub>3</sub>OH molecules. Bonds of this type are among the most common ones in molecules of biological interest. For comparison, we also perform the same calculations using CCSD(T) and various DFT methods.

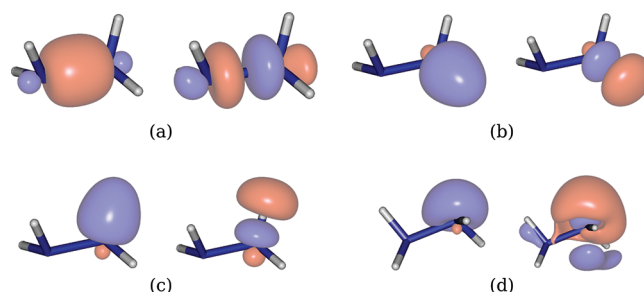
For the fragmentation of N<sub>2</sub>H<sub>4</sub> in two NH<sub>2</sub> molecules, we present a comprehensive analysis of the performance of the J-LGVBn wave functions and summarize the results in Table 5. The VMC and DMC energies of both NH<sub>2</sub> and N<sub>2</sub>H<sub>4</sub> converge rapidly with increasing order of the wave function. For the

$\text{N}_2\text{H}_4$  molecule, the difference between the VMC energies of the J-LGVB4 and J-LGVB10 wave functions is only 0.5 mhartree. Clearly, the excitations  $E_n$  for  $n > 4$  are not important to describe nondynamical correlation, except for E9 responsible for almost all of the energy difference between J-LGVB4 and J-LGVB10. The DMC energies display the same trend as VMC. The rapid convergence of the dissociation energy of the N–N bond, with increasing the order of the wave function, is even more evident. The binding energy obtained with the J-LGVB $n$  wave function remains unchanged passing from  $n = 4$  to 10. Moreover, already for  $n = 2$ , the dissociation energy is comparable to the CCSD(T)/cc-pV5Z result (see Table 9).

Since simple one-determinant Jastrow–Slater wave functions are often used within QMC, we also report the energies obtained with a single determinant. We observe that the absolute energies obtained with the J-LGVB $n$  wave functions are significantly lower than the energies of the monodeterminantal wave functions. For  $\text{N}_2\text{H}_4$ , the difference between the VMC energy obtained with J-LGVB1 is 10 mhartree lower than the value obtained with 1 determinant, while the difference at the DMC level is 4 mhartree. Compared to the monodeterminantal wave functions, J-LGVB1 yields an improvement of 1 kcal/mol at the DMC level and more than 2 kcal/mol at VMC level for the dissociation energy of the N–N bond.

Another wave function widely used in QMC calculations is a CAS expansion in the determinantal component. Therefore, we also compare our results to those obtained with different choices of CAS space. For the N–N bond breaking in  $\text{N}_2\text{H}_4$ , we test three types of CAS functions: (i) a minimum active space of (2,2) for  $\text{N}_2\text{H}_4$  and a single determinant for  $\text{NH}_2$ , (ii) an intermediate active space of (6,6) for  $\text{N}_2\text{H}_4$  and (3,3) for  $\text{NH}_2$ , and (iii) an active space including all valence electrons and an equal number of orbitals,<sup>45</sup> namely, (14,14) for  $\text{N}_2\text{H}_4$ , and (7,7) for  $\text{NH}_2$ . We perform a truncation only on the CSF coefficients for the largest active space. The dissociation energy obtained with the small CAS is comparable with the value of J-LGVB1. This finding can be explained with the fact that the two orbitals of the active space (2,2) of  $\text{N}_2\text{H}_4$  have substantially the shape of the bonding and antibonding orbitals of the N–N bond. This wave function is therefore capable to describe the dissociation of this particular bond but no other bond in the molecule, in contrast to J-LGVB1 that provides a good description of the dissociation of all bonds. The intermediate active space yields lower absolute energies but a worse dissociation energy than the smaller CAS. This highlights the difficulty of identifying balanced active spaces to describe multiple dissociative processes in the same molecule. For the largest active space, we are forced to truncate the expansion since the numbers of CSFs in the CAS(7,7) of  $\text{NH}_2$  and CAS(14,14) of  $\text{N}_2\text{H}_4$  are 196 and 1 381 095, respectively. As criterion for truncation, we use the sum of the squares of the coefficients and include the CSFs with greater absolute values of the coefficients with threshold for the sum of the squares of the coefficients set to 0.974. This value is selected to obtain less than 250 determinants for  $\text{N}_2\text{H}_4$ . The dissociation energy obtained with the truncated CAS(7,7) and CAS(14,14) wave functions is closer to the experimental value than the J-LGVB $n$  results, but the absolute energies are worse than the energies of the J-LGVB $n$  wave functions of higher order. Therefore, the good result obtained for the dissociation energy is not a consequence of smaller errors in the calculation of the correlation energies.

In Table 5, we also include the energies calculated with a larger basis set (VQZ). We employ this basis set only for the monodeterminantal, the smallest CAS, and the J-LGVB1 and J-LGVB2 wave functions. The total energies for all four wave functions decrease significantly, and also the DMC dissociation energies benefit from the increased size of the basis set as the errors in the binding compared to the VTZ basis are reduced by 1.0, 0.7, 0.6, and 0.5 kcal/mol for the monodeterminantal, small CAS, J-LGVB1, and J-LGVB2 wave functions, respectively. For  $\text{N}_2\text{H}_4$ , we plot in Figure 2 the orbitals obtained in the VMC



**Figure 2.** Orbitals of  $\text{N}_2\text{H}_4$  obtained in the optimization of the J-LGVB1 wave function. The (a) pair represents the bonding and antibonding orbitals for the N–N bond. The (b) and (c) pairs are the bonding and antibonding orbitals for the two types of N–H bonds in the molecule. The (d) pair represents the lone pair on the nitrogen and its relative diffuse orbital. We show only four of the seven pairs of orbitals because the remaining three pairs can be generated by rotation of the (b–d) pairs with respect to the  $C_2$  axis of symmetry of the molecule.

optimization of the J-LGVB1 wave function and also used in the higher-order J-LGVB $n$  wave functions. We can see that, for each electron pair, we obtain bonding and antibonding orbitals which are very well localized in space.

We collect the results for the dissociation of nitrous acid ( $\text{HNO}_2$ ) in hydroxyl radical (OH) and nitric oxide (NO) in Table 6. The improvement in the VMC energies passing from a monodeterminantal wave function to J-LGVB1 is drastic. In particular, the energy of  $\text{HNO}_2$  decreases by 19 mhartree, and the error in the dissociation energy is reduced by 3.5 kcal/mol.

**Table 6.** VMC and DMC Dissociation Energies ( $\Delta E$ , kcal/mol) of Nitrous Acid in Hydroxyl Radical and Nitric Oxide, Computed with Several Trial Wave Functions<sup>a</sup>

method	wave function	energy			$\Delta E$
		OH	NO	$\text{HNO}_2$	
VMC	1 det	−16.5513	−25.8863	−42.5103	45.6
	J-LGVB1	−16.5543	−25.8967	−42.5292	49.1
	J-LGVB2	−16.5546	−25.8981	−42.5339	50.9
DMC	1 det	−16.5630	−25.9152	−42.5564	49.1
	J-LGVB1	−16.5648	−25.9220	−42.5661	49.8
	J-LGVB2	−16.5651	−25.9222	−42.5693	51.5
experimental					51.82

<sup>a</sup>We also report the total energies (hartree) and do not include zero-point energy correction. The experimental reference is ATcT data corrected for zero-point energy, spin–orbit interaction, and Born–Oppenheimer approximation.<sup>41</sup> The statistical error on the VMC and DMC total energies is 0.0001 au, while the statistical error on the corresponding dissociation energies is 0.1 kcal/mol.

Compared to the J-LGVB1 wave function, the use of J-LGVB2 further reduces the error in the dissociation energy by 1.8 kcal/mol. Even though these differences are washed out at the DMC level, we can still observe the same trend in DMC. The DMC energy of  $\text{HNO}_2$  decreases by 9.7 mhartree when going from the monodeterminantal to the J-LGVB1 wave function, and the improvement in the dissociation energy is of 0.7 kcal/mol. A significant increase in accuracy is obtained with the J-LGVB2 wave functions, which reduce the error with respect to the experimental dissociation energy to only 0.3 kcal/mol.

In Table 7, we show the results for the dissociation of methylamine ( $\text{CH}_3\text{NH}_2$ ) molecule in amino ( $\text{NH}_2$ ) and methyl

**Table 7. VMC and DMC Dissociation Energies ( $\Delta E$ , kcal/mol) of Methylamine in Amino and Methyl Radicals, Computed with Several Trial Wave Functions<sup>a</sup>**

method	wave function	energy			$\Delta E$
		$\text{CH}_3$	$\text{NH}_2$	$\text{CH}_3\text{NH}_2$	
VMC	1 det	-7.4065	-11.0683	-18.6162	88.7
	J-LGVB1	-7.4081	-11.0715	-18.6239	90.6
	J-LGVB2	-7.4087	-11.0718	-18.6256	91.0
DMC	1 det	-7.4150	-11.0798	-18.6397	90.9
	J-LGVB1	-7.4155	-11.0809	-18.6423	91.6
	J-LGVB2	-7.4159	-11.0812	-18.6434	91.8

experimental 92.05

<sup>a</sup>We also report the total energies (hartree) and do not include zero-point energy correction. The experimental reference is ATcT data corrected for zero-point energy, spin-orbit interaction, and Born-Oppenheimer approximation.<sup>41</sup> The statistical error on the VMC and DMC total energies is 0.0001 au, while the statistical error on the corresponding dissociation energies is 0.1 kcal/mol.

radicals ( $\text{CH}_3$ ). Compared to the fragmentations discussed above, the results obtained with the monodeterminantal wave functions give smaller errors. The monodeterminantal, J-LGVB1, and J-LGVB2 wave functions result in an error in the binding energy of 3.3, 1.4, and 1.0 kcal/mol in VMC and an error 1.1, 0.4, and 0.2 kcal/mol in DMC.

In Table 8, we show the results for the dissociation of methanol ( $\text{CH}_3\text{OH}$ ) in hydroxyl ( $\text{OH}$ ) and methyl radicals ( $\text{CH}_3$ ). Methanol and methylamine molecules are isoelectronic, and although the C-O bond is slightly stronger than the C-N bond, the performance of the J-LGVB $n$  wave function is rather similar to the one observed in the previous fragmentations. The VMC error with respect to the experimental dissociation energy is 4.4 kcal/mol for the monodeterminantal wave functions. The use of J-LGVB1 wave functions reduces this error to 1.7 kcal/mol, while the error is only 1.4 kcal/mol with the J-LGVB2 wave functions. At DMC level, the deviation from experiments is 1.0 kcal/mol with the monodeterminantal wave functions and 0.4 and 0.1 kcal/mol with the J-LGVB1 and J-LGVB2 wave functions, respectively.

In Table 9, we collect the results of the dissociation energies calculated with CCSD(T)/cc-pV $n$ Z ( $n = \text{T, Q, 5}$ ) and with different density functionals theory (DFT) approximations and the aug-cc-pVQZ basis set. For the DFT calculations, we consider the widely tested functionals B3LYP and PBE0, the most recent M06-2X and M08-HX recommended for thermochemical estimates, and the double-hybrid functional B2PLYP which was shown to be the most accurate for the

**Table 8. VMC and DMC Dissociation Energies ( $\Delta E$ , kcal/mol) of Methanol in Hydroxyl and Methyl Radicals, Computed with Several Trial Wave Functions<sup>a</sup>**

method	wave function	energy			$\Delta E$
		$\text{CH}_3$	$\text{OH}$	$\text{CH}_3\text{OH}$	
VMC	1 det	-7.4065	-16.5513	-24.1073	93.8
	J-LGVB1	-7.4081	-16.5543	-24.1160	96.5
	J-LGVB2	-7.4087	-16.5546	-24.1176	96.8
DMC	1 det	-7.4150	-16.5630	-24.1329	97.2
	J-LGVB1	-7.4155	-16.5648	-24.1362	97.8
	J-LGVB2	-7.4159	-16.5651	-24.1373	98.1

experimental 98.20

<sup>a</sup>We also report the total energies (hartree) and do not include zero-point energy correction. The experimental reference is ATcT data corrected for zero-point energy, spin-orbit interaction, and Born-Oppenheimer approximation.<sup>41</sup> The statistical error on the VMC and DMC total energies is 0.0001 au, while the statistical error on the corresponding dissociation energies is 0.1 kcal/mol.

description of bond fragmentation in the recent work by Karton et al.<sup>41</sup> For our set of fragmentations, we find that DFT results are very far from chemical accuracy, with errors as large as 5 kcal/mol. Even the use of the double hybrid functional B2PLYP gives unsatisfactory mean-average and maximum deviations of 2.0 and 2.8 kcal/mol, respectively. The CCSD(T) results are instead very good and show a systematic reduction in the errors with increasing size of the basis set. For completeness, we also compute the CCSD(T) dissociation energies with the BFD pseudopotentials used in the QMC calculations and obtain that the small differences compared to the experimental data suggest that the use of pseudopotential only introduces minor errors in the QMC calculations (see also ref 46). When compared to the CCSD(T) results, the DMC energies of bond breaking computed with the J-LGVB1 wave functions have an accuracy better than CCSD(T)/cc-pVTZ but worse than CCSD(T)/cc-pVQZ. The use of J-LGVB2 wave functions give better results than CCSD(T)/cc-pV5Z.

#### 4. CONCLUSIONS

In this paper, we develop a new type of Jastrow-Slater wave functions (J-LGVB $n$ ) for use in QMC calculations. The J-LGVB $n$  wave functions are inspired to the GVB functions and constructed with localized orbitals optimized together with the Jastrow factor and CSF coefficients. These wave functions are size extensive, have a compact structure, and scale linearly with respect to the size of the molecule. Importantly, they allow one to correlate all valence electrons with a small number of CSFs and are therefore suitable to describe large portions of complex potential energy surfaces. The determinantal part of the wave functions is built by progressively including 10 classes of excitations that complete an active space of 4 electrons in 4 orbitals, defined by the antibonding and bonding orbitals of each of the adjacent electron pairs in the molecule. Our method relies on chemical intuition to build the electronic wave function, is based on a flexible choice of variational parameters, and allows us to treat correlation at different levels in different parts of the system. We extensively test the performance of our wave function on the fragmentation of  $\text{N}_2\text{H}_4$  and show that the convergence of the total and relative QMC energies is very fast with increasing order of the wave functions. Already with a



**Table 9. QMC, DFT, and CCSD(T) Dissociation Energies, Mean Absolute Deviation (MAD), and Maximum Absolute Deviation (MAX) (kcal/mol) for All the Systems Taken into Account<sup>a</sup>**

method	D(N <sub>2</sub> H <sub>4</sub> )	D(HNO <sub>2</sub> )	D(CH <sub>3</sub> NH <sub>2</sub> )	D(CH <sub>3</sub> OH)	MAD	MAX
VMC 1 det	67.3	45.6	88.7	93.8	5.1	6.2
VMC J-LGVB1	69.5	49.1	90.6	96.5	2.5	3.9
VMC J-LGVB2	70.8	50.9	91.0	96.8	1.5	2.5
DMC 1 det	70.8	49.1	90.9	97.2	1.9	2.7
DMC J-LGVB1	71.8	49.8	91.6	97.8	1.1	2.0
DMC J-LGVB2	72.4	51.1	91.8	98.1	0.4	1.0
B3LYP/aug-cc-pVQZ	70.2	52.5	88.2	95.4	2.6	3.8
PBE0/aug-cc-pVQZ	74.9	54.3	95.4	100.8	1.7	4.3
M08-HX/aug-cc-pVQZ	78.0	56.1	92.7	98.7	3.7	5.5
M06-2X/aug-cc-pVQZ	78.9	53.7	95.7	101.9	3.5	4.6
B2PLYP/aug-cc-pVQZ	70.8	51.3	89.2	96.1	2.0	2.8
CCSD(T)-FC/cc-pVTZ	69.9	49.6	89.2	95.4	2.8	3.5
CCSD(T)-FC/cc-pVQZ	71.7	51.3	90.7	97.2	1.1	1.6
CCSD(T)-FC/cc-pVSZ	72.4	51.9	91.2	97.8	0.6	1.0
CCSD(T)-BFD/VSZ	73.3	51.9	91.8	97.5	0.3	0.7
experimental	73.39	51.82	92.05	98.20		

<sup>a</sup>We do not include zero-point-energy correction. The experimental references are ATcT data corrected for zero-point energy, spin-orbit interaction, and Born–Oppenheimer approximation.<sup>41</sup> The statistical error on the VMC and DMC dissociation energies is 0.1 kcal/mol.

second-order form (J-LGVB2), we obtain excellent agreement with the experimental dissociation. This finding is confirmed in further tests we performed for the dissociation of the HNO<sub>2</sub>, CH<sub>3</sub>NH<sub>2</sub>, and CH<sub>3</sub>OH molecules. Both in VMC and DMC, J-LGVB1 wave functions yield always better results than monodeterminantal wave functions, and a further systematic improvement is observed when switching from J-LGVB1 to J-LGVB2. Compared to the CAS trial wave functions, the use of J-LGVB<sub>n</sub> wave functions eliminates the arbitrariness of the truncation on the expansion and is applicable to larger systems. For the dissociations considered here, the DMC estimates with the J-LGVB2 wave functions satisfy the chemical accuracy, and only the CCSD(T) values computed with the large cc-pVSZ basis set are of comparable quality. These promising results encourage us to further study the J-LGVB<sub>n</sub> wave functions and, thanks to their linear scaling, to apply them to larger molecules at a manageable computational cost.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

The geometries of the considered molecules and the CCSD(T) and DFT energies are given. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [francesco.fracchia@for.unipi.it](mailto:francesco.fracchia@for.unipi.it); [c.filippi@utwente.nl](mailto:c.filippi@utwente.nl); [amovilli@dcci.unipi.it](mailto:amovilli@dcci.unipi.it)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was carried out under the HPC-EUROPA2 project (project no. 228398) with the support of the European Commission Capacities Area - Research Infrastructures Initiative. This work is part of a national research project cofunded by the Italian Ministry of Research and University (PRIN 2009).

## ■ REFERENCES

- (1) Lüchow, A. *WIREs Comput. Mol. Sci.* **2011**, *1*, 388–402.
- (2) Reynolds, P.; Ceperley, D.; Alder, B.; Lester, W., Jr. *J. Chem. Phys.* **1982**, *77*, 5593–5603.
- (3) Toulouse, J.; Umrigar, C. *J. Chem. Phys.* **2008**, *128*, 174101–174115.
- (4) Valsson, O.; Filippi, C. *J. Chem. Theory Comput.* **2010**, *6*, 1275–1292.
- (5) Filippi, C.; Zaccheddu, M.; Buda, F. *J. Chem. Theory Comput.* **2009**, *5*, 2074–2087.
- (6) Lewis, G. N. *J. Am. Chem. Soc.* **1916**, *38*, 762–785.
- (7) Heitler, W.; London, F. Z. *Phys.* **1927**, *44*, 455–472.
- (8) Anderson, A. G.; Goddard, W. A. *J. Chem. Phys.* **2010**, *132*, 164110–164122.
- (9) Braïda, B.; Toulouse, J.; Caffarel, M.; Umrigar, C. J. *J. Chem. Phys.* **2011**, *134*, 084108–084118.
- (10) Bajdich, M.; Mitas, L.; Drobný, G.; Wagner, L.; Schmidt, K. *Phys. Rev. Lett.* **2006**, *96*, 130201–130204.
- (11) Casula, M.; Attaccalite, C.; Sorella, S. *J. Chem. Phys.* **2004**, *121*, 7110–7135.
- (12) McDouall, J. J.; Robb, M. *Chem. Phys. Lett.* **1987**, *142*, 131–135.
- (13) Thorsteinsson, T.; Cooper, D. L.; Gerratt, J.; Karadakov, P. B.; Raimondi, M. *Theor. Chim. Acta* **1996**, *93*, 343–366.
- (14) Amovilli, C.; Floris, F. M.; Mennucci, B. *Int. J. Quantum Chem.* **1999**, *74*, S9–67.
- (15) Amovilli, C.; Floris, F. M.; Solà, M.; Tomasi, J. *Organometallics* **2001**, *20*, 1310–1316.
- (16) Hay, P. J.; Hunt, W. J.; Goddard, W. A. *J. Am. Chem. Soc.* **1972**, *94*, 8293–8301.
- (17) Goddard, W. A.; Dunning, T. H.; Hunt, W. J.; Hay, P. J. *Acc. Chem. Res.* **1973**, *6*, 368–376.
- (18) Coulson, C.; Fischer, I. *Philos. Mag.* **1949**, *40*, 386–393.
- (19) Foster, J.; Boys, S. *Rev. Mod. Phys.* **1960**, *32*, 303–304.
- (20) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
- (21) Umrigar, C. J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. G. *Phys. Rev. Lett.* **2007**, *98*, 110201–110203.
- (22) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (23) Ivanic, J.; Ruedenberg, K. *Theor. Chem. Acc.* **2001**, *106*, 339–351.



- (24) CHAMP is a quantum Monte Carlo program package written by Umrigar, C. J.; Filippi, C. and collaborators.
- (25) Burkatzki, M.; Filippi, C.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 234105–234112.
- (26) For the hydrogen atom, we use a more accurate BFD pseudopotential and basis set. Dolg, M.; Filippi, C., private communication.
- (27) Casula, M. *Phys. Rev. B* **2006**, *74*, 161102–161105.
- (28) Filippi, C.; Umrigar, C. J. *J. Chem. Phys.* **1996**, *105*, 213–226.
- (29) CFOUR, a quantum chemical program package written by J. F. Stanton, J. Gauss, M. E. Harding, P. G. Szalay with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W. J. Lauderdale, D. A. Matthews, T. Metzroth, D. P. O'Neill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkiewicz, A. Tajti, J. Vázquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see <http://www.cfour.de> (accessed July 2, 2010).
- (30) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (31) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (32) Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B* **1988**, *37*, 785–789.
- (33) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (34) Perdew, J.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982–9985.
- (35) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (36) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.
- (37) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215–241.
- (38) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.
- (39) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108–034122.
- (40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
- (41) Karton, A.; Daon, S.; Martin, J. M. *Chem. Phys. Lett.* **2011**, *510*, 165–178.
- (42) Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.
- (43) Nooijen, M.; Shamasundar, K. R.; Mukherjee, D. *Mol. Phys.* **2005**, *103*, 2277–2298.
- (44) Ruscic, B.; Pinzon, R.; Morton, M.; von Laszewski, G.; Bittner, S.; Nijsure, S.; Amin, K.; Minkoff, M.; Wagner, A. *J. Phys. Chem. A* **2004**, *108*, 9979–9997.
- (45) For N<sub>2</sub>H<sub>4</sub>, a CAS(14,12) expansion corresponds to the valence CAS space obtained by correlating all valence electrons in the set of complete valence orbitals (obtained from the 4 atomic 2s and 2p basis functions for nitrogen and 1 atomic s basis for hydrogen). Here, we consider a CAS(14,14) to be compatible with our J-LGVBn wave functions and include also the antibonding orbital for the two lone pairs. Similarly, we augment the valence CAS(7,6) space of NH<sub>2</sub> with the antibonding lone pair on the nitrogen.
- (46) Fracchia, F.; Amovilli, C. *Chem. Phys. Lett.* **2012**, *521*, 20–25.