Graphic characterization of determinant space from full configuration interaction quantum Monte Carlo

Lei Sun,\textsuperscript{1,\textbullet\textsuperscript{a)} Zixi Zhang,\textsuperscript{1,\textbullet\textsuperscript{a)} Tonghuan Jiang,\textsuperscript{1,\textbullet\textsuperscript{b)} Yilin Chen,\textsuperscript{2} and Ji Chen\textsuperscript{1,\textbullet\textsuperscript{3,4,5,\textbullet\textsuperscript{c)}}

\textsuperscript{1)}School of Physics, Peking University, Beijing 100871, P. R. China
\textsuperscript{2)}International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, P. R. China
\textsuperscript{3)}Collaborative Innovation Center of Quantum Matter, Beijing 100871, P. R. China
\textsuperscript{4)}Interdisciplinary Institute of Light-Element Quantum Materials and Research Center for Light-Element Advanced Materials, Peking University, Beijing 100871, P. R. China
\textsuperscript{5)}Frontiers Science Center for Nano-Optoelectronics, Peking University, Beijing 100871, P. R. China

(Dated: 27 September 2022)

Full configuration interaction quantum Monte Carlo (FCIQMC) is a powerful algorithm to compute electronic states of molecules by efficient sampling of the whole determinant space. FCIQMC simulation not only calculates the correlation energy at unprecedented accuracy, but also provides insightful information of the electronic structure of computed states, e.g. the population of determinants and their connections, which have not been fully explored. In this work, we devise a determinant configuration graph for visualizing the simulation results of FCIQMC, revealing the nature of molecule’s electronic structure in a new perspective. In addition, we propose two new analytical descriptors to quantify the extent of multi-configuration of the electronic states of molecules. The graph and descriptors provide us new understanding of the electronic structure of molecules, and will be useful methods for analyzing stochastic quantum chemistry calculations such as FCIQMC.

\textsuperscript{a)}These authors contributed equally to this work
\textsuperscript{b)}Electronic mail: jiangth1997@pku.edu.cn
\textsuperscript{c)}Electronic mail: ji.chen@pku.edu.cn

\textsuperscript{1}\textsuperscript{arXiv:2209.12570v1 [physics.chem-ph] 26 Sep 2022}
I. INTRODUCTION

Exact wave function of realistic systems, e.g. large molecules, described by Schrödinger’s equation is generally out of reach because it is NP-hard to compute and hence approximations are often used\textsuperscript{1,2}. In the state-of-the-art wave function theory for Fermionic system, one computational approach commonly adapted is based on expanding the many-body wave function by Slater determinants of single-particle orbitals, which are ultimately expanded in a given basis set. Theoretically, if one can include all possible determinant configurations unbiasly, i.e. include the full configuration interaction (FCI), then the solution represents the exact solution in the given basis set. Introduced by Booth et al. in 2009\textsuperscript{3}, full configuration interaction quantum Monte Carlo (FCIQMC) is becoming one of the most powerful algorithm in wave function theory, which employs an efficient stochastic sampling method to explore the full space of determinant configurations. FCIQMC provides a promising means to tackle correlated electronic systems, and the past decade has witnessed a rapid development of FCIQMC algorithm\textsuperscript{4–6}, as well as the range of applications\textsuperscript{7,8}.

In previous studies, the priority of FCIQMC development is mostly put on improving the accuracy and efficiency, namely getting better energy calculation with lower computational cost\textsuperscript{8}. Most analyses of FCIQMC have focused on the energy and other derived physical quantities, but analyses of wave function, i.e. configuration population in the Hilbert space, are relatively inadequate. Such information is key to further develop efficient deterministic wave function algorithms such as configuration interaction (CI), coupled cluster (CC), and many-body perturbation methods\textsuperscript{9–17}. Currently, analysis of FCIQMC simulations is limited to arrange the determinants according its population in a one-dimensional histogram\textsuperscript{18,19}. For most systems, such an analysis is enough to tell whether the system is single-configurational or multi-configurational, but it is often not sufficient to understand true nature of the determinant space. In particular, the so-called multi-configuration feature is rather vague, and there are still differences between two multi-configurational systems. Diagnostics of multi-reference character have been proposed in various forms of electronic structure calculations. However, these diagnostics may differ from one to another because they are often based on calculations with different approximations, e.g. density functional theory, multi-reference self consistent field and coupled cluster\textsuperscript{20}. Therefore, it is desirable to develop new analysis methods for FCIQMC simulations to provide deeper insights to the
electronic structure of molecules.

In this work, we propose a new scheme of analysis based on the determinant configuration graph using not only their populations but also the Hamiltonian matrix elements that represent the “interaction” between determinants. The analysis is presented by a graph of nodes distributed in a two-dimensional manner. We further propose new descriptors to characterize the nature of electronic structure that can be related to the extent of multi-configuration, or in another word resembling the extent of strong correlation. In the following section we present details about FCIQMC simulations, determinant configuration graph, and proposed descriptors. Using the new analyses, we further discuss the ground state of several molecular systems including H\textsubscript{4}, B\textsubscript{2}, C\textsubscript{2} and N\textsubscript{2}, where distinct features are illustrated for molecules upon the variation of atomic configurations.

II. METHODS

A. FCIQMC

The theoretical foundation of FCIQMC is the full configuration interaction (FCI) method, in which the wave function of a many-electron system can be written as

\[
\Psi = \sum_i C_i D_i, \quad D_i(x) = \frac{1}{\sqrt{N!}} \det\begin{vmatrix}
\psi_{i1}(x_{i1}) & \psi_{i2}(x_{i1}) & \cdots & \psi_{iN}(x_{i1}) \\
\psi_{i1}(x_{i2}) & \psi_{i2}(x_{i2}) & \cdots & \psi_{iN}(x_{i2}) \\
\vdots & \vdots & \ddots & \vdots \\
\psi_{i1}(x_{iN}) & \psi_{i2}(x_{iN}) & \cdots & \psi_{iN}(x_{iN})
\end{vmatrix}
\]  

(1)

where \( D_i \) is the i-th configuration formed by a Slater determinant of molecular orbitals, \( C_i \) is the corresponding expansion coefficient, and \( N \) is the number of electrons. If we start with the Hartree-Fock state, represented by one determinant formed with only occupied molecular orbitals, then one can interpret FCI as a method considering all configurations excited from the Hartree-Fock state, namely including all the electron correlations neglected in Hartree-Fock wave function.

FCIQMC employs a stochastic algorithm to solve FCI wave function. The deterministic calculation on FCI wave function is too challenging\textsuperscript{3}. The FCIQMC algorithm follows the Schrödinger equation and its imaginary time evolution, which resembles the diffusion of classical particles. The evolution of wave function is discretized into many steps of propagation
of Monte Carlo walkers. The master equation is written as,

$$C_i(\tau + \delta \tau) = (1 - \delta \tau (H_{ii} - S))C_i(\tau) - \delta \tau \sum_{j \neq i} H_{ij} C_j(\tau)$$

(2)

where $H$ is the Hamiltonian, $S$ is a parameter to be adjusted during the evolution for population control. Furthermore, each Monte Carlo step can be decomposed into three processes: (i) spawning, where every configuration spawns a number of new walkers at another configuration, described by $-\delta \tau \sum_{j \neq i} H_{ij} C_j(\tau)$; (ii) cloning/death, where walkers of each configuration clone or die, described by $(1 - \delta \tau (H_{ii} - S))C_i(\tau)$. (iii) annihilation, where walkers with different signs on each configuration cancel each other so that only one type of walkers remains. The Monte Carlo simulation converges to the ground state of many-body electronic wave function as the imaginary time evolution proceeds $\tau \to \infty$.

Our FCIQMC calculations are performed with the NECI package. The cc-pVTZ basis set was used in our calculations on H$_4$, B$_2$, C$_2$ and N$_2$, respectively. The molecular orbitals for the subsequent FCIQMC calculations were obtained with restricted Hartree-Fock, performed with the PySCF package. The initiator (i-FCIQMC) and adaptive shift approach (as-FCIQMC) were used in all FCIQMC calculations, with initiator threshold $n_a = 3$. The time-step was updated using the TAU-SEARCH facility of NECI. The semi-stochastic method was used, and the size of deterministic space was set to 100. A trial wave function was used to obtain the projected energy estimate, and the size of the trial space was set to 100.

B. Configuration graph

We design a configuration graph of the determinant space. The schematic diagram is shown in FIG.1(a). The main elements of the graph are nodes with different size and color, which are connected by lines varying in color and width. Each node represents a configuration $D_i$ and its size is proportional to the corresponding CI coefficient from FCIQMC simulation. The color of node is related to the energy of the configuration itself, i.e., the diagonal elements $H_{ii}$ of the Hamiltonian matrix. Higher energy leads to darker nodes. The length of lines is negatively correlated to the absolute value of the Hamiltonian matrix element between two configurations, i.e. the absolute value of off-diagonal elements $|H_{ij}|$. The width and transparency of lines are positively correlated to $|H_{ij}|$. The color of lines is determined by
the sign of $H_{ij}$, which takes orange if positive and cyan if negative. The distance between two nodes is negatively correlated to $|H_{ij}|$. Hence, the configurations with greater Hamiltonian matrix element tend to cluster together. In FCIQMC simulation a greater $|H_{ij}|$ means a greater chance of walkers spreading between the corresponding configurations.

This graph provides a direct visualization of the FCI determinant space and shows the relationship between configurations. A graph with nodes averagely scattered and connected by weak lines means the Hamiltonian matrix is highly diagonal and the walkers barely spread. The leading configuration would have a CI coefficient of approximately 1, while all the other determinants are rarely populated. In this case the exact FCI wave function would be highly similar to the Hartree-Fock wave function. A similar situation occurs when the wave function is computed using single-determinant density functional theory. For particular systems we can see the gathering of nodes, where nodes are much closer to each other and
have wider connected lines. We call such group a “cluster”. In such a scenario, walkers could spread fast within a cluster and slowly between clusters, and the system would be multi-configurational and the exact wave function would be far different from the single-determinant wave function.

It is worth noting that each pair of nodes has independent $H_{ij}$ and if we intend to draw a two-dimensional graph the distance between each pair is entangled and can not be determined separately. In this work, we use the two-dimensional (2D) Kamada-Kawai layout to get the position of nodes, which leads to an optimal determination of node separations following the value of $H_{ij}$. A cross validation using the four-dimensional Kamada-Kawai layout in conjunction with the t-distributed stochastic neighbor embedding (t-SNE) algorithm shows that the main feature of the graph is correctly characterized by the 2D Kamada-Kawai layout.

C. Multi-configuration descriptor

In FCIQMC, the possibility of walkers spawning among configurations is proportional to Hamiltonian matrix elements. The configuration graph is a direct indicator of the possible dynamic behavior of walkers. The distance between nodes is negatively correlated with the Hamiltonian matrix element, the farther the nodes are separated, the smaller the spawning probability is. If the configurations form clusters, walkers would have greater chance to spawn within the cluster. Such features are likely to be related to the performance of traditional methods such as configuration interaction and coupled cluster, and can be useful to characterize the electronic states of molecules.

The concept of clustering coefficient in graph theory is designed to describe the degree of the gathering of nodes, and there are multiple ways to define it. Here we propose two descriptors as defined in Eq. 3 and Eq. 4.

$$\gamma_e = \sum_{u \neq v}(C_u)^2(C_v)^2 |H_{uv}|$$ (3)

$$\gamma_t = \sum_{u \neq v \neq w}(C_u)^2(C_v)^2(C_w)^2 |H_{uv}H_{uw}H_{vu}|^{\frac{1}{2}}$$ (4)
$C_u$ is the CI coefficient at node $u$, and $H_{uv}$ is the Hamiltonian matrix element of node $u$ and node $v$. The introduction of CI coefficient guarantees the convergence of descriptors as a function of the number of configurations included. $\gamma_e$ characterizes a pair of nodes in the graph, whereas $\gamma_t$ characterizes the triangles as shown in FIG. 1(b). The form $H_{uv}H_{vu}H_{wu}$ means the summation is based on triangles and each term represents an edge of the triangle composed of node $u$, $v$ and $w$. When several nodes in a graph form clusters have large CI coefficient and are connected by large $H_{uv}$, descriptors $\gamma_e$ and $\gamma_t$ are large, indicating a multi-configurational system.

III. RESULTS

A. Configuration graph analyses

We first discuss the configuration graph of a model system composed of four hydrogen atoms on the vertexes of a rectangle. The structure is determined by two parameters, namely (i) the angle of two diagonals ($\theta$) and (ii) the length of half diagonal ($R$), as shown in the inset of Fig. 2(b). Via tuning the two structure parameters, we can cover a wide range of electronic structures from weak to strong correlation. Using FCIQMC, we calculated the ground state of the rectangle $H_4$ system at various structures, $R$ ranging from 1.5 Bohr to 8.0 Bohr and $\theta$ ranging from $20^o$ to $80^o$, and the energy results are shown in Fig. 2(a). Fig. 2(b) is a one-dimensional plot of the 20 most populated determinant configurations of three selected structures. Fig. 2(c-e) shows the corresponding configuration graphs, which represent the typical graphs observed for all $H_4$ structures calculated. Graphs at other structures are presented in Fig. S1.

Among the three typical graphs, in Fig. 2(c) and Fig. 2(d), the HF node is very much the single dominating configuration, with only weak edges connecting the HF node and others in both graphs. In the graph of Fig. 2(c), the HF determinant locates far from the other nodes, which are of much higher energy. The death rate of walkers on high energy determinant is rather high, which along with the slow spawning rate suppresses the growth of walkers on these non-HF determinants. Therefore, most walkers are trapped in the HF node and rarely spawn to the other determinants. In Fig. 2(d), the HF node is close to a group of nodes even without obvious edges connected, indicating some kind of connection which enables
FIG. 2. **FCIQMC calculation and configuration graph of H$_4$.** (a) The FCIQMC energy of H$_4$ at different structures. (b) The one-dimensional plot of 20 most populated CI coefficient for three selected structures. The inset shows the definition of structural parameters. (c, d, e) Configuration graphs corresponding to the three selected structures, respectively. The largest yellow nodes in panels (c-e) correspond to the Hartree Fock determinant.

walkers to spread between nodes. Such connection is stronger than Fig. 2(c), hence some of the nodes have higher walker populations.

In contrast, Fig. 2(e) is a typical strongly correlated and multi-configurational graph. The main cluster contains the HF determinant and a dozen of other determinants. All the nodes in this main cluster are low in energy and most of the nodes have a significant number of walkers populated. Spawning among the cluster is very frequent during FCIQMC simulation. It is also worth noting that the cluster can simultaneously contain nodes with opposite signs. On both sides of the main cluster are two symmetrical clusters of the second lightest color and second largest size on average. The axial symmetry of the two clusters is related to the spin symmetry. Further analyses are reported in Fig. S10, which show that the nodes of the two clusters are in one-to-one correspondence, each couple has the same space orbitals
but the opposite spin direction, lying on roughly axial symmetric position of the graph. Another larger cluster of darker and smaller nodes lies right down below the main cluster. These three clusters accompany the main cluster and characterizes the main feature of this system. Edges within these clusters are mostly positive and wide, walkers spawn fast inside each cluster, making the node size within each cluster approximately the same. Edges connecting different clusters are weaker, along which walkers spawn slower, thus the number of walkers can be prominently different between clusters.

Having demonstrated the main characters of the designed configuration graph on the model H$_4$ system, we now discuss three realistic molecules including B$_2$, C$_2$, N$_2$. FCIQMC calculations of the energy of diatomic molecules have been reported in previous works as good examples to discuss the electronic correlation effects in molecules,$^{26-28}$ and in this study we only focus the graph analyses. To highlight the change of complexity of electronic structure, we plot in Fig. 3(a,e,i) the energy of FCIQMC. The energy curve shows equilibrium bond length of the molecules and where they are to dissociate. From the second to fourth column of Fig. 3 we show two graphs for each molecule, with one structure around the equilibrium and another one in the strongly stretched region. More graphs at other bond lengths are enclosed in Fig. S2-S4. In Fig. 3(b,j), the molecules are single-reference while there are two prominent configurations in FIG. 3(f), which can’t be perceived as single-reference. In Fig. 3(d,h,l), the molecules become heavily stretched and the graphs show multiple clusters of nodes. Interestingly, the graph patterns are quite different for three molecules, but they are common in the multi-configurational feature and suggest strong correlation.

FCIQMC is an accurate but expensive approach because one need to converge the CI coefficient to a very high precision in order to capture the correlation energy to enough accuracy. However, the main characters of the configuration graph can be reproduced without converging the CI coefficient to a very high precision, which will extend the use of FCIQMC in theoretical analyses of molecular electronic structure. The construction of graph is determined by two factors: (i) the Hamiltonian matrix elements which are determined in the self-consistent-field procedure, and (ii) the configurations chosen to draw the graph. The information required from FCIQMC mainly comes from highly populated configurations, and those with little population would have negligible effect on the graphical features of the wavefunction. In this work, the first 100 most populated configurations are chosen to construct the graph. The exact value of CI coefficients is to determine the size of nodes,
FIG. 3. (a,e,i) FCIQMC energy of B$_2$, C$_2$, N$_2$. (b,c,d), (f,g,h), (j,k,l) are typical graphs of B$_2$, C$_2$, N$_2$ respectively.

which can be seen as secondary to our analysis when compared to the distribution of nodes and doesn’t affect the character of the graph.

B. Multi-configuration descriptor analyses

In order to understand the change of electronic structure in a more quantitative manner, we further analyze the FCIQMC calculations using two multi-configuration descriptors as defined in the Method section. The calculated values of the two descriptors as a function of molecular bond length are plotted in Fig. 4(a,b). A larger value of the descriptors corresponds to a more multi-configuration electronic structure. Overall the two descriptors behave similarly when bond length changes. The most obvious observation is the rise of descriptors at large bond lengths, which is in line with the configuration graphs displayed in Fig. 3 and Fig. S2-S4. It is also a common feature for three molecules that both descriptors would decline at medium bond length, i.e. 3-5 Bohr, when molecules are about to dissociate,
FIG. 4. Descriptor $\gamma_t$ (a) and $\gamma_e$ (b) of three molecules at different bond lengths. (c) The correlation between the two descriptors, the positive and roughly linear correlation shows the consistency of $\gamma_t$ and $\gamma_e$.

before the system eventually turn multi-configurational. Fig. (c) shows the two descriptors are positively correlated, which further demonstrates the consistency.

There are also differences for different molecules. Comparing the three molecules, $N_2$ is generally less multi-configurational than $B_2$ and $C_2$. However, $N_2$ has the most sharp increase of the descriptors in the 3-4 Bohr range, which is known to be the most challenging regime of $N_2$ with electronic structure methods\textsuperscript{29}. Moreover, for the $B_2$ and $N_2$ molecules, both descriptors vanish at small bond length, but for $C_2$ the value of both descriptors are large at small bond length. This clearly shows that $C_2$ is very different from the other two molecules, even without being strongly stretched $C_2$ is still very multi-configurational (Fig. S3). This is likely related to the existence of many low lying states in $C_2$ at small bond lengths\textsuperscript{30}.

The new descriptors proposed in this work is based on the fundamental information of the determinant space, whereas many existing diagnostics are based on energy calculations. Here, we can further compare the other diagnostics to $\gamma_t$ and $\gamma_e$. To this end, we have calculated 12 widely used descriptors on $B_2$, $C_2$, and $N_2$. These include three descriptors obtained from complete active space self-consistent field (CASSCF) calculation, namely the leading weight $(C_0^2)$\textsuperscript{31-33}, the occupation number of the lowest unoccupied molecular orbital ($n_{LUMO}[CAS]$) and the occupation number of the highest occupied molecular orbital ($n_{HOMO}[CAS]$)\textsuperscript{34,35}.

Two similar descriptors are from Second-order Møller-Plesset perturbation theory (MP2), namely $n_{LUMO}[MP2]$ and $n_{HOMO}[MP2]$\textsuperscript{30}. Three descriptor are obtained from coupled cluster theory: $T_1$ is the Frobenius norm of the single excitation amplitude vector $t_1$\textsuperscript{31}, $D_1$ is the largest eigenvalue of the matrix $t_1t_1^\dagger$\textsuperscript{37}, and $D_2$ is the largest eigenvalue of the matrix $t_2t_2^\dagger$\textsuperscript{38}. 

11
where $t_2$ is the double excitation vector. Using density functional theory (DFT) with fractional occupation, one can define the so-called dynamic ($I_D$) and none-dynamic ($I_{ND}$) correlation based on occupation numbers. We include $I_{ND}$ and the portion of none-dynamic correlation, $r_{ND} = \frac{I_{ND}}{I_{ND} + I_D}$, calculated using two different exchange correlation functionals. A complete plot of the calculated descriptors are presented in Fig. S9. While these descriptors may be capable of describing multi-configuration character to some extent, they do not always show a consistent behavior to the determinant space information obtained in FCIQMC simulation. This is not a surprise because most traditional descriptors of wavefunction theory simply measure the deviation of multi-configuration wavefunction from Hartree-Fock, and do not capture features of connection between excited Slater determinants. There are many cases that some descriptors may work very well, but the performance varies from one system to another. In contrast, our descriptors borrows graph connectivity measurements from graph theory to capture these features, and are closely related to the intrinsic nature of the determinant space, and correspond well with the visualized configuration graphs.

In Fig. 5, we plot the correlation coefficient (Pearson’s r) between these descriptors and our graph based multi-configurational descriptors. For $B_2$, value of Pearson’s r is relatively large for almost all circumstances, which means that all the descriptors work well and can be used to capture the electronic structure change. For $N_2$, Pearson’s r of descriptors from MP2 and CASSCF are close to zero and can not predict well the multi-configurational character. The other descriptors still show good consistency with our descriptors. The most challenging case is $C_2$. As we discussed above, even around equilibrium bond length $C_2$ is strongly correlated. Most descriptors are qualitatively wrong with negative Pearson’s r.

IV. CONCLUSIONS

In this study, we propose new analysis methods of the electronic structure of molecules based on FCIQMC simulation, in particular to reveal the multi-configurational nature of different electronic states. The analyses include (i) a configuration graph for direct visualization of the population and connections between important configurations of the full determinant space; (ii) two descriptors for quantitative prediction of the multi-configuration character. Our methods can also be used with other wave function calculations based on exploring the
FIG. 5. Correlation coefficient (Pearson’s r) between traditional descriptors and $\gamma_e$ (a) and $\gamma_t$ (b). The values of descriptors are reported in Fig. 4 and Fig. S9. Positive (negative) Pearson’s r means the two descriptors are positively (negatively) correlated.

determinant space. The insights provided by these analyses can assist further developments of electronic structure methods, especially those aimed at tackling the most challenging strongly correlated system. In previous methods, e.g. selected CI\textsuperscript{10} and heat-bath CI\textsuperscript{14}, selection criterion is often based on a single threshold computed. Our graph analysis may provide an additional rule based on the clustering feature, i.e. selecting a whole cluster or adapting the previous selection criteria within clusters. Furthermore, the graphical symmetry observed may also be useful information when considering the rules in selected CI methods.

V. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grant No. 92165101 and No. 11974024, the National Key R&D Program of China under Grant No. 2021YFA1400500, and the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDB33000000. We are grateful for computational resources provided by Peking University, the TianHe-1A supercomputer, Shanghai Supercomputer Center, and Songshan Lake Materials Lab.
REFERENCES

1. J. A. Pople, Rev. Mod. Phys. 71, 1267 (1999).
2. W. Kohn, Rev. Mod. Phys. 71, 1253 (1999).
3. G. H. Booth, A. J. W. Thom, and A. Alavi, The Journal of Chemical Physics 131, 054106 (2009).
4. D. Cleland, G. H. Booth, and A. Alavi, The Journal of Chemical Physics 132, 041103 (2010).
5. N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, The Journal of Chemical Physics 143, 134117 (2015).
6. K. Ghanem, A. Y. Lozovoi, and A. Alavi, The Journal of Chemical Physics 151, 224108 (2019).
7. G. H. Booth, A. Gruneis, G. Kresse, and A. Alavi, Nature 493, 365 (2013).
8. K. Guther, R. J. Anderson, N. S. Blunt, N. A. Bogdanov, D. Cleland, N. Dattani, W. Dobrautz, K. Ghanem, P. Jeszenszki, N. Liebermann, G. L. Manni, A. Y. Lozovoi, H. Luo, D. Ma, F. Merz, C. Overy, M. Rampp, P. K. Samanta, L. R. Schwarz, J. J. Shepherd, S. D. Smart, E. Vitale, O. Weser, G. H. Booth, and A. Alavi, The Journal of Chemical Physics 153, 034107 (2020).
9. B. Huron, J. P. Malrieu, and P. Rancurel, The Journal of Chemical Physics 58, 5745 (1973).
10. J. Ivanic and K. Ruedenberg, Theoretical Chemistry Accounts 106, 339 (2001).
11. P. J. Knowles, Molecular Physics 113, 1655 (2015).
12. J. B. Schriber and F. A. Evangelista, The Journal of Chemical Physics 144, 161106 (2016).
13. W. Liu and M. R. Hoffmann, Journal of Chemical Theory and Computation 12, 1169 (2016).
14. A. A. Holmes, N. M. Tubman, and C. J. Umrigar, Journal of Chemical Theory and Computation 12, 3674 (2016).
15. N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, The Journal of Chemical Physics 145, 044112 (2016).
16. S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, Journal of Chemical Theory and Computation 13, 1595 (2017).
17. J. E. Deustua, I. Magoulas, J. Shen, and P. Piecuch, **The Journal of Chemical Physics** **149**, 151101 (2018).

18. J. Chen, N. A. Bogdanov, D. Usvyat, W. Fang, A. Michaelides, and A. Alavi, **The Journal of Chemical Physics** **153**, 204704 (2020).

19. T. Jiang, Y. Chen, N. A. Bogdanov, E. Wang, A. Alavi, and J. Chen, **The Journal of Chemical Physics** **154**, 164302 (2021).

20. C. Duan, F. Liu, A. Nandy, and H. J. Kulik, **Journal of Chemical Theory and Computation** **16**, 4373 (2020).

21. Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, “Pyscf: the python-based simulations of chemistry framework,” (2017).

22. K. Ghanem, K. Guther, and A. Alavi, **The Journal of Chemical Physics** **153**, 224115 (2020).

23. F. R. Petruzielo, A. A. Holmes, H. J. Changlani, M. P. Nightingale, and C. J. Umrigar, **Phys. Rev. Lett.** **109**, 230201 (2012).

24. T. Kamada and S. Kawai, **Information Processing Letters** **31**, 7 (1989).

25. L. van der Maaten and G. Hinton, **Journal of Machine Learning Research** **9**, 2579 (2008), pagination: 27.

26. G. H. Booth, D. Cleland, A. J. W. Thom, and A. Alavi, **The Journal of Chemical Physics** **135**, 084104 (2011).

27. D. Cleland, G. H. Booth, C. Overy, and A. Alavi, **Journal of Chemical Theory and Computation** **8**, 4138 (2012).

28. T. Jiang, W. Fang, A. Alavi, and J. Chen, “General analytical nuclear force and molecular potential energy surface from full configuration interaction quantum Monte Carlo,” (2022), arXiv:2204.13356 [physics].

29. W. Ren, W. Fu, and J. Chen, “Towards the ground state of molecules via diffusion Monte Carlo on neural networks,” (2022).

30. J. R. Stallcop, H. Partridge, A. Pradhan, and E. Levin, **Journal of Thermophysics and Heat Transfer** **14**, 480 (2000).

31. T. J. Lee and P. R. Taylor, **International Journal of Quantum Chemistry** **36**, 199–207 (2009).

32. J. S. Sears and C. D. Sherrill, **The Journal of Physical Chemistry A** **112**, 6741–6752 (2008).
33 S. R. Langhoff and E. R. Davidson, International Journal of Quantum Chemistry 8, 61–72 (1974).
34 U. R. Fogueri, S. Kozuch, A. Karton, and J. M. L. Martin, Theoretical Chemistry Accounts 132, 1291 (2013).
35 O. Tishchenko, J. Zheng, and D. G. Truhlar, Journal of Chemical Theory and Computation 4, 1208–1219 (2008).
36 C. Møller and M. S. Plesset, Physical Review 46, 618–622 (1934).
37 C. L. Janssen and I. M. Nielsen, Chemical Physics Letters 290, 423–430 (1998).
38 I. M. Nielsen and C. L. Janssen, Chemical Physics Letters 310, 568–576 (1999).
39 E. Ramos-Cordoba, P. Salvador, and E. Matito, Physical Chemistry Chemical Physics 18, 24015–24023 (2016).
40 E. Ramos-Cordoba and E. Matito, Journal of Chemical Theory and Computation 13, 2705–2711 (2017).
41 M. K. Kesharwani, N. Sylvetsky, A. Köhn, D. P. Tew, and J. M. L. Martin, The Journal of Chemical Physics 149, 154109 (2018).