

Exploring non-linear correlators on AGP

Armin Khamoshi*

Department of Physics and Astronomy, Rice University, Houston, TX 77005-1892

Guo P. Chen

Department of Chemistry, Rice University, Houston, TX 77005-1892

Thomas M. Henderson and Gustavo E. Scuseria

Department of Chemistry, Rice University, Houston, TX 77005-1892 and

Department of Physics and Astronomy, Rice University, Houston, TX 77005-1892

(Dated: January 15, 2022)

Single-reference methods such as Hartree-Fock-based coupled cluster theory are well known for their accuracy and efficiency for weakly correlated systems. For strongly correlated systems, more sophisticated methods are needed. Recent studies have revealed the potential of the antisymmetrized geminal power (AGP) as an excellent initial reference for the strong correlation problem. While these studies improved on AGP by linear correlators, we explore some non-linear exponential ansätze in this paper. We investigate two approaches in particular. Similar to *Phys. Rev. B* **91**, 041114(R) (2015), we show that the similarity transformed Hamiltonian with a Hilbert-space Jastrow operator is summable to all orders and can be solved over AGP by projecting Schrödinger's equation. The second approach is based on approximating the unitary pair-hopper ansatz recently proposed for application on a quantum computer. We report benchmark numerical calculations against the ground state of the pairing Hamiltonian for both of these approaches.

I. INTRODUCTION

Many-body methods in electronic structure theory frequently start from a mean-field reference. Commonly, a single-reference Slater determinant obtained by Hartree-Fock (HF) calculations is used for post-HF methods that make correction to this mean-field [1]. Such methods include configuration interaction (CI) and coupled cluster theory (CC). For weakly correlated systems, single-reference CC is often regarded as the *gold standard* and calculations are routinely performed using CC with single and double excitations (CCSD) very accurately at an affordable cost of $\mathcal{O}(M^6)$ where M is the size of the system [1–3]. Unfortunately, the same cannot be said for the strong correlation problem. When many-body interactions are strong, more than one Slater determinant becomes important and methods based on a single HF reference often break down [4–6]. Although multireference methods exist, they have certain limitations and are not always straightforward to use [7–10]. As such, a new way of thinking and more sophisticated methods are desirable.

A useful way of organizing Hilbert space in the context of strong correlation is by *seniority* [11, 12]. Loosely speaking, seniority (Ω) is the number of unpaired fermions in some appropriate pairing scheme. For example, a two-body Hamiltonian can generally be written as [13]

$$\mathbf{H} = \mathbf{H}^{\delta\Omega=0} + \mathbf{H}^{\delta\Omega=2} + \mathbf{H}^{\delta\Omega=4} \quad (1)$$

where the superscripts indicate the portion of the Hamiltonian that couples those determinants whose seniority differs by 0, 2, etc. Numerical evidence has shown that the low-seniority portions of the Hilbert space can recover a very large part of the correlation energy and have the correct qualitative behavior [12, 14]. The seniority-zero subspace can be solved exactly by doubly occupied configuration interaction (DOCI) but it scales combinatorially with system size [15–17]. Remarkably, CC restricted to paired double excitation (pCCD) [18], also known as AP1roG [19], is often nearly exact in this subspace with a very low computational cost for realistic molecular Hamiltonians [20–22]. However, this correspondence is not universal and it has been shown that pCCD and its extensions break down for some problems including the attractive pairing Hamiltonian [5, 13, 23–25],

$$\mathbf{H} = \sum_p \epsilon_p \mathbf{N}_p - G \sum_{pq} \mathbf{P}_p^\dagger \mathbf{P}_p, \quad (2)$$

where $\epsilon_p = p\Delta\epsilon$ are equally spaced, doubly degenerate energy levels where $\Delta\epsilon$ is the level spacing, G tunes the strength of the infinite-range pairwise interactions, and

$$\mathbf{P}_p^\dagger = c_p^\dagger c_{\bar{p}}^\dagger, \quad (3a)$$

$$\mathbf{N}_p = c_p^\dagger c_p + c_{\bar{p}}^\dagger c_{\bar{p}}, \quad (3b)$$

such that c_p^\dagger creates a fermion in spin-orbital p , and \bar{p} is its “paired” companion. Generally, in the seniority-zero subspace, all operators can be organized as a linear combination of terms like $\mathbf{P}_p^\dagger \dots \mathbf{N}_q \dots \mathbf{P}_r$ [26], where the *nilpotent* operators \mathbf{P}_p^\dagger and \mathbf{P}_p together with \mathbf{N}_p are gen-

* Correspondence email address: armin.khamoshi@rice.edu

erators of a $su(2)$ Lie algebra

$$[\mathbf{P}_p, \mathbf{P}_q^\dagger] = \delta_{pq} (1 - \mathbf{N}_p), \quad (4a)$$

$$[\mathbf{N}_p, \mathbf{P}_q^\dagger] = 2\delta_{pq} \mathbf{P}_q^\dagger. \quad (4b)$$

The breakdown of CC in the pairing Hamiltonian can be understood in terms of the spontaneous breaking of number symmetry in the electronic mean-field, which occurs at some critical value of G , $G_c > 0$. As a result, the mean-field solution gives rise to the number-broken Bardeen–Cooper–Schrieffer (BCS) wavefunction [27] for $G > G_c$ and a number-conserving Slater determinant for $G < G_c$ including all $G < 0$ where the interaction is repulsive. Indeed, it is in the attractive regime where CC as well as a whole host of other conventional methods struggle [25]. It is noteworthy that the pairing Hamiltonian is exactly solvable [23, 28].

Recently, we proposed to use the AGP wavefunction as the initial reference for this and potentially more general problems [25, 26, 29–32]. Along similar lines of research, Johnson et al. have developed correlated methods by explicitly using the pairing Hamiltonian’s eigenvectors, Richardson–Gaudin states, as the initial reference [33–36]. The main advantage of AGP for the seniority-zero sector is that on the one hand, it puts coefficients on each Slater determinant in the DOCI while still conserving seniority, and on the other hand, AGP is computationally straightforward. To put it differently, the natural orbitals of AGP, HF, pCCD, and DOCI are all the same if the same pairing scheme is chosen, differing only in occupation. The AGP wavefunction can then be expanded in its seniority-zero natural orbital determinants and can occupy all $\binom{M}{N}$ such determinants, where N is the number of fermion pairs. It thus encompasses all possible configurations in the seniority-zero space, albeit with inexact coefficients. However, unlike DOCI, AGP can be optimized at mean-field cost $\mathcal{O}(M^3)$ [37]. It also encompasses the HF state, so it has a much richer structure as an initial reference. Moreover, we have shown recently that the reduced density matrices (RDM) over AGP can be computed efficiently as all higher order RDMs are decomposable in terms of lower order ones. [26] These realizations permit us to envision *post*-AGP methods—analogue to *post*-HF—to capture the remaining error and provide accurate ansätze in both weak and strong correlation. A case can be made for using AGP in the seniority non-zero spaces where pairs are broken, but we concentrate on the seniority-zero case at this stage.

Recent studies in our group have developed CI-based methods over AGP and numerical calculations show excellent accuracy when applied to the pairing Hamiltonian. These methods include orthogonal excitations [25, 29] and those based on the generators of the Lie algebra Eq. (4) [31]. It has been shown that these models are formally equivalent and can be viewed as “geminal replacement” methods [31]. Encouraged by the accuracy of linear models, we would like to explore non-linear correlator ansätze over AGP. Non-linear models in principle

can approximate the DOCI coefficient differently and it remains to be seen whether they provide an advantage over the linear case. Developing a CC theory over AGP is not trivial and remains an open problem [25].

In this paper, we explore two exponential correlators over AGP which we build from number-conserving combinations of the generators of the Lie algebra, Eq. (4). The first is the exponential of Jastrow [38] operators on AGP (JAGP). Jastrow–Gutzwiller correlators are ubiquitous in physics and chemistry and are widely used in Monte Carlo calculations. [38–49] It was shown in Ref. [50, 51] that the similarity transformation of any fermionic Hamiltonian, $e^{-\mathbf{J}} \mathbf{H} e^{\mathbf{J}}$, where \mathbf{J} is a Hilbert-space Jastrow operator, is analytically summable to all orders. Inspired by that, we extend the formalism to the pairing $su(2)$ algebra Eq. (4) and show that we can efficiently solve the projected Schrödinger equation over AGP. In a way, this could be viewed as a form of *generalized* coupled cluster theory, but instead of the traditional particle-hole excitations, correlation is mediated by number operators that create excitations on AGP—a possibility that does not occur for single Slater determinants as they are eigenfunctions of \mathbf{J} .

The second ansatz is the unitary *pair-hopper* that we proposed recently for applications in quantum computers [32]. Unitary ansätze can be implemented efficiently on a quantum computer, [52] but some approximations must be made in order to implement them on a classical computer. Here, we approximate the unitary transformation of the Hamiltonian using the Baker–Campbell–Hausdorff (BCH) formula by truncating it at the 4th commutator. Affordable computational scaling is made possible by the *reconstruction formulae* [26] with which we can eliminate the bottleneck of computing and storing higher rank density matrices from our calculations. As an alternative to the brute-force expansion of the BCH formula, we apply the *canonical transformation* theory as formulated in Ref. [53–57] to recursively sum the transformed Hamiltonian. We report benchmark calculations of all methods against the ground state energy of the pairing Hamiltonian.

The organization of this paper is as follows. To familiarize the reader with our notation, in Sec. II we review some basic properties of AGP and its density matrices. In Sec. III, we present our formalism for solving similarity transformed JAGP as well as its application to the pairing Hamiltonian. Sec. IV discusses the truncated commutator expansion and canonical transformation with the unitary pair-hoppers ansatz over the pairing Hamiltonian. Lastly, some discussion and concluding remarks are provided in Sec. V.

II. BACKGROUND

The AGP wavefunction with N pairs corresponds to a condensate where all $2N$ fermions are placed in the same

geminal. Mathematically it can be written as [58, 59]

$$|\text{AGP}\rangle = \frac{1}{N!} \left(\mathbf{\Gamma}^\dagger \right)^N |\text{vac}\rangle, \quad (5)$$

where $|\text{vac}\rangle$ is the physical vacuum and $\mathbf{\Gamma}^\dagger$ is the geminal creation operator whose general expression is

$$\mathbf{\Gamma}^\dagger = \sum_{pq} \eta_{pq} c_p^\dagger c_q^\dagger. \quad (6)$$

Here, p, q denote spin-orbitals and η is a skew-symmetric matrix known as the *geminal coefficient*. To make a connection with Eq. (3) and for mathematical convenience, we perform an orbital rotation that brings the matrix of geminal coefficients into a block-diagonal form [60], that is

$$\eta = \bigoplus_{p=1}^M \begin{pmatrix} 0 & \eta_p \\ -\eta_p & 0 \end{pmatrix}. \quad (7)$$

Therefore, without loss of generality, we obtain the geminal operator in the *natural orbital* basis of the geminal as

$$\mathbf{\Gamma}^\dagger = \sum_{p=1}^M \eta_p c_p^\dagger c_{\bar{p}}^\dagger = \sum_{p=1}^M \eta_p \mathbf{P}_p^\dagger. \quad (8)$$

where we made use of the notation in Eq. (3) and M denotes the number of orbitals. In this formalism, we can express the AGP wavefunction as

$$|\text{AGP}\rangle = \sum_{1 \leq p_1 < \dots < p_N \leq M} \eta_{p_1} \dots \eta_{p_N} \mathbf{P}_{p_1}^\dagger \dots \mathbf{P}_{p_N}^\dagger |\text{vac}\rangle. \quad (9)$$

Similarly, we can write down the equation for DOCI,

$$|\text{DOCI}\rangle = \sum_{1 \leq p_1 < \dots < p_N \leq M} D_{p_1 \dots p_N} \mathbf{P}_{p_1}^\dagger \dots \mathbf{P}_{p_N}^\dagger |\text{vac}\rangle. \quad (10)$$

From these expressions, it is evident that the two wavefunctions have the same structure but differ in the coefficients of each determinant; in DOCI the amplitude tensor is fully connected, whereas in AGP it is fully factorized.

Despite being a linear combination of a combinatorial number of determinants, reduced density matrices of many-body fermionic operators over AGP can be computed efficiently at polynomial cost and are very sparse in the natural orbital basis [26]. To see this, first note that for a generic many-body operator \hat{O} we have

$$\langle \text{AGP} | \hat{O} | \text{AGP} \rangle = \langle \text{AGP} | \hat{O}^{\delta\Omega=0} | \text{AGP} \rangle. \quad (11)$$

In other words, since AGP is a seniority-zero wavefunction, all many-body density matrices over AGP can be written in terms of $\langle \text{AGP} | \mathbf{P}_p^\dagger \dots \mathbf{N}_q \dots \mathbf{P}_r \dots | \text{AGP} \rangle$ where the overlaps associated with particle-number breaking or seniority non-conserving operators are identically zero.

Therefore, we can organize all nonzero reduced density matrices in ascending rank as follows

$$Z_p^{(1,1)} = \langle \text{AGP} | \mathbf{N}_p | \text{AGP} \rangle, \quad (12a)$$

$$Z_{pq}^{(0,2)} = \langle \text{AGP} | \mathbf{P}_p^\dagger \mathbf{P}_q | \text{AGP} \rangle, \quad (12b)$$

$$Z_{pq}^{(2,2)} = \langle \text{AGP} | \mathbf{N}_p \mathbf{N}_q | \text{AGP} \rangle, \quad (12c)$$

$$Z_{pqr}^{(1,3)} = \langle \text{AGP} | \mathbf{P}_p^\dagger \mathbf{N}_q \mathbf{P}_r | \text{AGP} \rangle, \quad (12d)$$

$$Z_{pqr}^{(3,3)} = \langle \text{AGP} | \mathbf{N}_p \mathbf{N}_q \mathbf{N}_r | \text{AGP} \rangle, \quad (12e)$$

...

where the first integer in the superscript indicates the number of \mathbf{N}_p operators in the middle and the second integer is the rank of the density matrix. When indices of an RDM are different, we refer to it as an *irreducible* density matrix. This is because $\mathbf{N}_p = \mathbf{N}_p^2/2 = 2\mathbf{P}_p^\dagger \mathbf{P}_p$ together with $\mathbf{P}_p^\dagger \mathbf{P}_p = \mathbf{P}_p^\dagger \mathbf{N}_p = 0$ in the seniority-zero space imply that RDMs with repeated indices are either zero or can be written in terms of lower rank density matrices.

Every matrix element of an AGP RDM, regardless of the rank, can be computed directly at $\mathcal{O}(M^2)$ cost using a method based on elementary symmetric polynomials (ESP) [26]. Even more efficiently, one can take advantage of the *reconstruction formulae* to express any irreducible RDM as a linear combination of lower rank ones, which can be performed all the way down to occupation numbers and geminal coefficients [26]. This allows us to compute any density matrix of rank 2 or higher with $\mathcal{O}(1)$ cost per element. Finally, we can take advantage of the equivalence between AGP and the number-projected BCS wavefunction [11, 61] to extract AGP density matrices via grid integrations of BCS transition density matrices [62], at $\mathcal{O}(N_{\text{grid}})$ cost per element, where N_{grid} is the size of the numerical quadrature for gauge integration [37]. While less efficient than the reconstruction formulae, this final approach can be very efficient when used to contract density matrices with other factorized quantities. In the subsequent sections, we use these methods to compute all density matrices in our calculations.

III. JASTROW-AGP

A. Analytic properties

Let the k -body, or rank- k , Hilbert-space Jastrow operator [63] be written as

$$\mathbf{J}_k = \frac{1}{2^k} \sum_{p_1 < \dots < p_k} \alpha_{p_1 \dots p_k} \mathbf{N}_{p_1} \mathbf{N}_{p_2} \dots \mathbf{N}_{p_k}, \quad (13)$$

where the sums run over all orbitals, α is a symmetric tensor and is invariant under the exchange of its indices, and $1/2^k$ is introduced for convenience. Importantly, since $\mathbf{N}_p^2 = 2\mathbf{N}_p$ in the seniority-zero space, we impose that α is zero if any two indices are the same and eliminate these terms from the sum in Eq. (13).

From the k -body Jastrow operator we can define the k -body Jastrow-AGP (J_k AGP) wavefunction as

$$|J_k\text{AGP}\rangle = e^{J_k}|\text{AGP}\rangle \quad (14)$$

As we prove in Appendix. A, we do not need to include lower body Jastrow operators in this J_k AGP wavefunction, as they are contained inside J_k . Thus, for example, we may write

$$e^{J_1+J_2}|\text{AGP}\rangle = e^{J'_2}|\text{AGP}\rangle. \quad (15)$$

where J'_2 is another two-body Jastrow operator.

Slater determinants built from the same orbitals used to define the Hilbert-space Jastrow operators are eigenfunctions of J_k , that is

$$J_k|\Phi_\mu\rangle = J_{k,\mu}|\Phi_\mu\rangle, \quad (16a)$$

$$J_{k,\mu} = \frac{1}{2^k} \sum_{p_1 < \dots < p_k} \alpha_{p_1 \dots p_k} N_{p_1,\mu} \dots N_{p_k,\mu} \quad (16b)$$

$$= \sum_{i_1 < \dots < i_k} \alpha_{i_1 \dots i_k}, \quad (16c)$$

where $|\Phi_\mu\rangle$ is a determinant, $N_{p,\mu}$ is the occupation number of level p in determinant μ , and the indices $i_1 \dots i_k$ run over the occupied orbitals in $|\Phi_\mu\rangle$. From this, it follows that

$$e^{J_k} \sum c_\mu |\Phi_\mu\rangle = \sum e^{J_{k,\mu}} c_\mu |\Phi_\mu\rangle. \quad (17)$$

For a complex α , the magnitude of the coefficient of a given determinant is modified by the Hermitian part of J_k , i.e. the real part of α . The anti-Hermitian part of J_k (the imaginary part of α) only adjusts the phase of each coefficient. It is thus apparent that for an N -body initial wavefunction $|\Psi\rangle$ in which all determinants $|\Phi_\mu\rangle$ have non-zero coefficients, $\exp(J_N)|\Psi\rangle$ can be made an exact eigenstate of a chosen Hamiltonian provided that J_N is non-Hermitian.

AGP is, in general, such an initial wavefunction, so $\exp(J_N)$ acting on it can be made exact for the seniority zero space. For any $k \leq N$, it follows from Eq. (16) and Eq. (17) that

$$e^{J_k}|\text{AGP}\rangle = \sum_{p_1 < \dots < p_N} \exp \left(\sum_{1 \leq j_1 < \dots < j_k \leq N} \alpha_{p_{j_1} \dots p_{j_k}} \right) \eta_{p_1} \dots \eta_{p_N} \mathbf{P}_{p_1}^\dagger \dots \mathbf{P}_{p_N}^\dagger |\text{vac}\rangle. \quad (18)$$

In contrast, we can see how the linear wavefunction $|\Psi\rangle = J_k|\text{AGP}\rangle$, which we refer to as J_k CI AGP [31], modifies AGP as follows

$$|\Psi\rangle = \sum_{p_1 < \dots < p_N} \left(\sum_{1 \leq j_1 < \dots < j_k \leq N} \alpha_{p_{j_1} \dots p_{j_k}} \right) \eta_{p_1} \dots \eta_{p_N} \mathbf{P}_{p_1}^\dagger \dots \mathbf{P}_{p_N}^\dagger |\text{vac}\rangle. \quad (19)$$

Thus, the way in which the linear and non-linear J_k correlators factorize the DOCI coefficients can be easily deduced from Eq. (18) and Eq. (19). Together with Eq. (10), we can readily see that

$$D_{p_1 \dots p_N}^{(\text{Linear})} \approx \sum_{1 \leq j_1 < \dots < j_k \leq N} \alpha_{p_{j_1} \dots p_{j_k}} \eta_{p_1} \dots \eta_{p_N}, \quad (20a)$$

$$D_{p_1 \dots p_N}^{(\text{Non-linear})} \approx \prod_{1 \leq j_1 < \dots < j_k \leq N} e^{\alpha_{p_{j_1} \dots p_{j_k}}} \eta_{p_1} \dots \eta_{p_N}. \quad (20b)$$

Indeed the two factorizations are different, and Eq. (20a) can be understood as the linear approximation to Eq. (20b) up to a constant shift of all amplitudes. An implication can be made, for example, about the sign of each coefficient; if α is real, the DOCI coefficient in Eq. (20b) can be negative only if some η 's are negative, whereas the same restriction is not implied in (20a).

Having shown analytically how J_k acts on AGP, we concern ourselves with optimizing the energy over JAGP in the subsequent sections.

B. Energy optimization with JAGP

Consider $k = 1$ for which the J_1 AGP wavefunction is just an AGP,

$$e^{J_1}|\text{AGP}\rangle = \sum_{p_1 < \dots < p_N} e^{\sum_j \alpha_{p_j}} \left(\prod_k \eta_{p_k} \mathbf{P}_{p_k}^\dagger \right) |\text{vac}\rangle \quad (21a)$$

$$= \sum_{p_1 < \dots < p_N} \left(\prod_k \eta_{p_k} e^{\alpha_{p_k}} \mathbf{P}_{p_k}^\dagger \right) |\text{vac}\rangle. \quad (21b)$$

Evidently, the action of J_1 just changes the geminal coefficient η_p to $\eta_p e^{\alpha_p}$. For $k = 1$, J_1 AGP can optimize an AGP, but it cannot correlate it; we need $k > 1$ to go beyond an AGP mean-field. Unfortunately, except for $k = 1$, the norm and density matrices over $e^{J_k}|\text{AGP}\rangle$ are not trivial to compute, and aside from quantum Monte Carlo methods, [40, 42, 46, 47, 64, 65] we are not aware of an efficient non-stochastic algorithm to compute the overlaps. This for examples makes the variational optimization of JAGP energy

$$E_{\text{var-JAGP}} = \frac{\langle \text{AGP} | e^{J^\dagger} H e^J | \text{AGP} \rangle}{\langle \text{AGP} | e^{J^\dagger} e^J | \text{AGP} \rangle}, \quad (22)$$

out of reach for the existing non-stochastic methods.

However, a case can be made for the similarity transformed Hamiltonian using J_2 , that is

$$E_{\text{ST-J}_2\text{AGP}} = \langle \text{AGP} | e^{-J_2} H e^{J_2} | \text{AGP} \rangle. \quad (23)$$

The two-body Jastrow correlator has a long history in electronic structure methods and has shown to be capable of recovering a significant portion of the correlation energy. [41, 46, 48] Previous work in our group has shown

that the similarity transformation of fermionic Hamiltonians using a two-body Jastrow operator is summable to all orders and the amplitudes can be solved by left projection via components of \mathbf{J}_2 acting on the HF reference [50, 51]. Following the same line of reasoning, we extend that formalism to the pairing $su(2)$ algebra Eq. (4) and show how to obtain the working equations for AGP in the next section.

The main idea relies on the fact that the similarity transformation of the Hamiltonian reduces the rank of \mathbf{J}_2 by one, thus the remaining one-body Jastrow operator can be subsequently absorbed into the reference. The same idea can be applied to the unitary transformation of the Hamiltonian with an anti-Hermitian $i\mathbf{J}_2$ operator

$$E_{uJ_2AGP} = \langle AGP | e^{-i\mathbf{J}_2} \mathbf{H} e^{i\mathbf{J}_2} | AGP \rangle, \quad (24)$$

wherein α is assumed to be real. Based on the argument in the previous section, the Jastrow correlators with purely imaginary amplitudes only contribute to the phase of each Slater determinant in AGP and do not alter their magnitudes. This can also be seen from Eq. (18) by $\alpha \rightarrow i\alpha$. The phase contributions could be important for Hamiltonians that contain complex terms including those that break time-reversal symmetry. For other Hamiltonians, the mere change in phase is insufficient to make the wavefunction exact.

C. Similarity transformation with JAGP

In this section, we extend the results of Ref. [50, 51] to the AGP wavefunction and our $su(2)$ algebra. For pedagogical reasons, we keep this section self-contained. Suppose one had a generic Jastrow operator \mathbf{J} . In a manner similar to coupled cluster theory, we intend to solve for the ground state energy by similarity transforming the Hamiltonian,

$$\bar{\mathbf{H}} = e^{-\mathbf{J}} \mathbf{H} e^{\mathbf{J}}, \quad (25a)$$

$$\bar{E} = \langle AGP | \bar{\mathbf{H}} | AGP \rangle. \quad (25b)$$

Similarity transformation does not change the spectrum of \mathbf{H} , but $\bar{\mathbf{H}}$ is no longer Hermitian and \bar{E} is not variational. Therefore we obtain the residual equations by left projecting the Schrödinger equation to get

$$\langle \mathbf{N}_{p_1} \dots \mathbf{N}_{p_k} | \bar{\mathbf{H}} \rangle - \bar{E} \langle \mathbf{N}_{p_1} \dots \mathbf{N}_{p_k} \rangle = 0 \quad \forall p_1 < \dots < p_k \quad (26)$$

where the expectation values are taken with respect to AGP. In traditional coupled cluster theory using particle-hole excitations, one expresses the similarity transformation by the Baker-Campbell-Hausdorff (BCH) expansion whose series naturally truncates at the 4th commutator for a two-body Hamiltonian. In contrast, the transformation using the Jastrow operator does not truncate, but it is summable to all orders for any Hamiltonian.

Our ultimate goal is to use a two- or higher-body Jastrow operators to similarity transform the Hamiltonian

and provide correlation. However, it will prove useful to begin by working through the case in which we use a one-body operator, for which the algebra is much simpler. Therefore, consider the \mathbf{J}_1 operator

$$\mathbf{J}_1 = \frac{1}{2} \sum_p \alpha_p \mathbf{N}_p. \quad (27)$$

Using Eq. (4), it is easy to show that

$$e^{-\mathbf{J}_1} \mathbf{P}_p^\dagger e^{\mathbf{J}_1} = e^{-\alpha_p} \mathbf{P}_p^\dagger, \quad (28a)$$

$$e^{-\mathbf{J}_1} \mathbf{P}_p e^{\mathbf{J}_1} = e^{\alpha_p} \mathbf{P}_p, \quad (28b)$$

$$e^{-\mathbf{J}_1} \mathbf{N}_p e^{\mathbf{J}_1} = \mathbf{N}_p. \quad (28c)$$

Thus, the similarity-transformed Hamiltonian simply adjusts the coefficients of the \mathbf{P}^\dagger and \mathbf{P} operators.

Alternatively, we can use Eq. (21) to absorb $\exp(\mathbf{J}_1)$ into AGP. Although we have seen how this can be done, we will extract the same result in a different fashion. Notice by Eq. (5) that

$$e^{\mathbf{J}_1} | AGP \rangle = \frac{1}{N!} \left(e^{\mathbf{J}_1} \Gamma^\dagger e^{-\mathbf{J}_1} \right)^N | vac \rangle, \quad (29)$$

where we used the fact that $\exp(\mathbf{J}_1) | vac \rangle = | vac \rangle$. From the algebra, Eq. (4), it follows

$$e^{\mathbf{J}_1} \Gamma^\dagger e^{-\mathbf{J}_1} = \sum_p \eta_p e^{\alpha_p} \mathbf{P}_p^\dagger, \quad (30a)$$

$$e^{\mathbf{J}_1} \Gamma e^{-\mathbf{J}_1} = \sum_p \eta_p e^{-\alpha_p} \mathbf{P}_p, \quad (30b)$$

from which it is clear that $e^{\mathbf{J}_1} | AGP \rangle$ leads to $\eta_p \rightarrow \eta_p e^{\alpha_p}$ for all geminal coefficients in the ket, and $\langle AGP | e^{-\mathbf{J}_1}$ results in $\eta_p \rightarrow \eta_p e^{-\alpha_p}$ in the bra.

We now proceed to derive the equations for similarity transformation with \mathbf{J}_2 . Define $\mathbf{J}_1(p) = \sum_q \alpha_{pq} \mathbf{N}_p / 4$ and recall $\alpha_{pp} = 0$, then as in Eq. (28) we obtain

$$e^{-\mathbf{J}_2} \mathbf{P}_p^\dagger e^{\mathbf{J}_2} = e^{-\mathbf{J}_1(p)} \mathbf{P}_p^\dagger e^{-\mathbf{J}_1(p)}, \quad (31a)$$

$$e^{-\mathbf{J}_2} \mathbf{P}_p e^{\mathbf{J}_2} = e^{\mathbf{J}_1(p)} \mathbf{P}_p e^{\mathbf{J}_1(p)}, \quad (31b)$$

$$e^{-\mathbf{J}_2} \mathbf{N}_p e^{\mathbf{J}_2} = \mathbf{N}_p \quad (31c)$$

where we used $[\mathbf{J}_1(p), \mathbf{P}_p^\dagger] = 0$ to symmetrically distribute the exponentials in the right hand side of Eq. (31) on either side of each operator. In fact, using these equations, we can transform any many-body operator of the form $\mathbf{P}_p^\dagger \dots \mathbf{N}_q \dots \mathbf{P}_r \dots$ in a symmetric way. In general, it can be shown that

$$e^{-\mathbf{J}_2} \mathbf{P}_{p_1}^\dagger \dots \mathbf{P}_{p_n}^\dagger \mathbf{N}_{q_1} \dots \mathbf{N}_{q_m} \mathbf{P}_{r_1} \dots \mathbf{P}_{r_n} e^{\mathbf{J}_2} = e^{\sum_{i=1}^n \mathbf{J}_1(r_i) - \mathbf{J}_1(p_i)} \mathbf{P}_{p_1}^\dagger \dots \mathbf{P}_{r_n} e^{\sum_{i=1}^n \mathbf{J}_1(r_i) - \mathbf{J}_1(p_i)}. \quad (32)$$

To obtain a closed-form expression for the energy and the residual equations, we need to further absorb each $\mathbf{J}_1(p)$

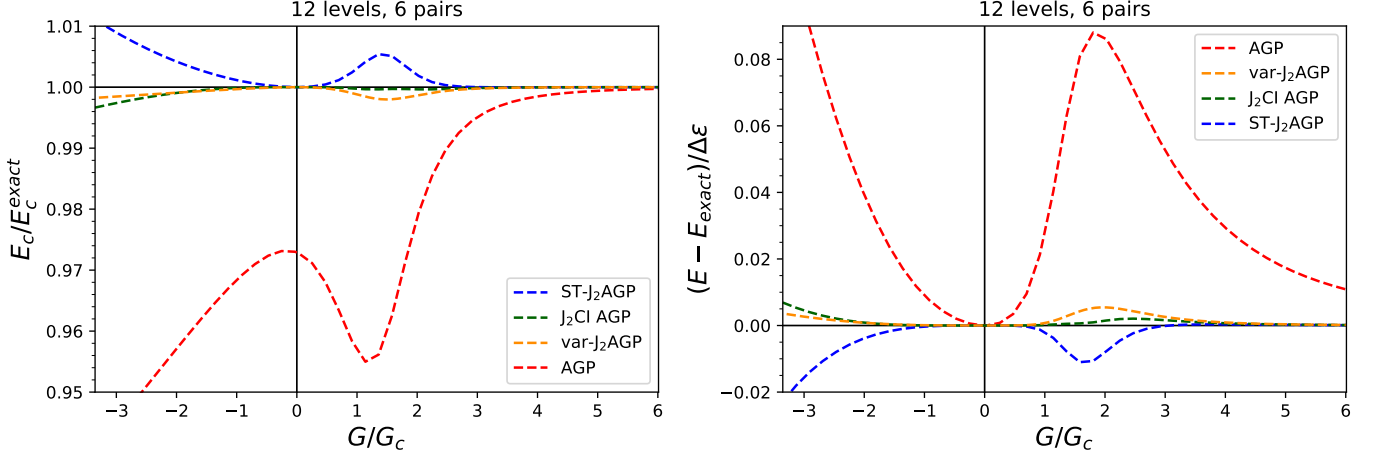


FIG. 1. Comparison of energy obtained using ST-J2AGP at 12 levels, half-filled with J2-CI and var-J2AGP. The left figure is the correlation energy as the difference with HF and the right is the total energy error.

into AGP. Following the same steps as in Eq. (30), we can see that

$$\begin{aligned} e^{J_1(q)-J_1(p)} \mathbf{\Gamma}^\dagger e^{J_1(p)-J_1(q)} \\ = \sum_r \left(\eta_r e^{(\alpha_{qr}-\alpha_{pr})/2} \right) \mathbf{P}_r^\dagger = \tilde{\mathbf{\Gamma}}_{pq}^\dagger. \end{aligned} \quad (33)$$

This together with Eq. (32) show that we can evaluate the expectation value of any similarity transformed operator of the form $\mathbf{P}_p^\dagger \dots \mathbf{N}_q \dots \mathbf{P}_r \dots$ with \mathbf{J}_2 by scaling the geminal coefficients appropriately.

The same basic steps follow for even higher-body Jastrow operators; the initial similarity-transformation of the Hamiltonian leads to modified matrix elements along with exponentials of Jastrow operators whose ranks are reduced by one. However, only \mathbf{J}_2 permits easy evaluation of the remaining expectation values.

For unitary JAGP, as in Eq. (24), one follows the same steps as above but replaces $\alpha \rightarrow i\alpha$.

D. Application to the pairing Hamiltonian

We now show how to apply the equations above to similarly transform the pairing Eq. (2). It follows from Eq. (32) that

$$\begin{aligned} e^{-J_2} \mathbf{H} e^{J_2} = \sum_p \epsilon_p \mathbf{N}_p \\ - G \sum_{pq} \left(e^{J_1(q)-J_1(p)} \mathbf{P}_p^\dagger \mathbf{P}_q e^{J_1(p)-J_1(q)} \right). \end{aligned} \quad (34)$$

For every $\mathbf{P}_p^\dagger \mathbf{P}_q$, define $|\widetilde{\text{AGP}}_{pq}\rangle = e^{J_1(q)-J_1(p)} |\text{AGP}\rangle$, and note that $|\widetilde{\text{AGP}}_{pq}\rangle$ is an AGP whose geminal coefficients have been rescaled by $\tilde{\eta}_r = \eta_r e^{(\alpha_{qr}-\alpha_{pr})/2}$ for all

orbitals r from Eq. (33). As such, we can obtain the energy,

$$\begin{aligned} \bar{E} = \langle e^{-J_2} \mathbf{H} e^{J_2} \rangle = \sum_p \epsilon_p \langle \text{AGP} | \mathbf{N}_p | \text{AGP} \rangle \\ - G \sum_{pq} \langle \widetilde{\text{AGP}}_{pq} | \mathbf{P}_p^\dagger \mathbf{P}_q | \widetilde{\text{AGP}}_{pq} \rangle. \end{aligned} \quad (35)$$

Following similar steps as above, it is straightforward to get closed-form expressions for the residual equations, Eq. (26), as well. We simply use

$$\begin{aligned} \langle \mathbf{N}_r \mathbf{N}_s e^{-J_2} \mathbf{H} e^{J_2} \rangle = \sum_p \epsilon_p \langle \text{AGP} | \mathbf{N}_p \mathbf{N}_r \mathbf{N}_s | \text{AGP} \rangle \\ - G \sum_{pq} \langle \widetilde{\text{AGP}}_{pq} | \mathbf{N}_r \mathbf{N}_s \mathbf{P}_p^\dagger \mathbf{P}_q | \widetilde{\text{AGP}}_{pq} \rangle \end{aligned} \quad (36)$$

which can then be normal ordered in terms of $\mathbf{P}_p^\dagger \dots \mathbf{N}_q \dots \mathbf{P}_r \dots$.

As outlined in Sec. II, it is a property of AGP-based RDMs that $Z_p^{(1,1)} = \langle \text{AGP} | \mathbf{N}_p | \text{AGP} \rangle$ is sufficient to construct all higher-body density matrices, where the cost of building $Z^{(1,1)}$ is $\mathcal{O}(N_{\text{grid}} M)$ using grid integration in the number-projected BCS representation of AGP. Of course the same principle applies to RDMs using $|\widetilde{\text{AGP}}_{pq}\rangle$, except that $\tilde{Z}_{r;pq}^{(1,1)} = \langle \widetilde{\text{AGP}}_{pq} | \mathbf{N}_r | \widetilde{\text{AGP}}_{pq} \rangle$ needs to be evaluated for every pq in the Hamiltonian, as demonstrated for Eq. (35). The cost of building $\tilde{Z}_{r;pq}^{(1,1)}$ for all pqr is $\mathcal{O}(N_{\text{grid}} M^3)$ and is the most expensive step in evaluating the ST-J2AGP energy. The cost of evaluating the residuals using the reconstruction formulae is $\mathcal{O}(M^4)$. Without the reconstruction formulae, the cost would be $\mathcal{O}(M^6)$.

In Fig. 1, we show the energy error and the fraction of correlation energy captured by ST-J2AGP. All calculations were carried out using the energy optimized AGP

for \mathbf{H} . The correlation energy is computed as the difference with that of HF. As shown in Fig. 1, ST-J₂AGP over-correlates near the recoupling regime, $G \approx G_c$ and smaller, including $G < 0$. In the strongly attractive correlated limit, $G \gg G_c$, ST-J₂AGP is highly accurate and becomes exact for sufficiently large G/G_c . It is noteworthy, however, that AGP is an eigenstate of the pairing Hamiltonian for $G/G_c \rightarrow \infty$. For comparison purposes, we have plotted the results for variational J₂AGP where we carried out the optimization using an in-house DOCI code. We have also included results using a linear wavefunction $\mathbf{J}_2|\text{AGP}\rangle$ in which the energy is determined variationally; this J₂CI-AGP has $\mathcal{O}(M^6)$ scaling, or $\mathcal{O}(M^8)$ without reconstruction formulae.

For larger systems, the over-correlation of ST-J₂AGP near the recoupling regime worsens while the linear variational theory remains well-behaved. Thus, although ST-J₂AGP has a lower cost than other AGP-based methods developed recently in our group, neither var-J₂AGP nor ST-J₂AGP yield as accurate energies for the pairing Hamiltonian. While correlators build from particle-hole excitations generally benefit from being put in an exponential operator, it is far from clear that the same is true of correlators built from Hilbert space Jastrow operators. Thus in the next section we discuss an exponential ansatz that we build from different generators.

IV. UNITARY PAIR-HOPPER CORRELATOR

An alternative way of creating an exponential ansatz based on two-body, number preserving operators with the generators of the algebra is to use a linear combination of terms like $\mathbf{P}_p^\dagger \mathbf{P}_q$. Unlike the Jastrow operators discussed in Sec. III, a unitary or similarity transformation using $\mathbf{P}_p^\dagger \mathbf{P}_q$ operators is not easily summable nor does it truncate naturally the way it does in CC theory. This is a direct consequence of the commutation relations Eq. (4). As a result, some form of approximation must be made. Recently, we proposed a unitary ansatz that we call *unitary pair-hopper* for an implementation on a quantum computer [32]. Benchmark calculations against the ground state energy of the pairing-Hamiltonian show highly accurate results.

In this section, we seek approximation schemes to this model that can be implemented affordably on a classical computer. In particular, we implement two approaches: First, we use the BCH formula and truncate the commutator expansion at some high order. Second, we apply the canonical transformation theory as studied extensively in Ref. [53–57] to this problem later in this section.

A. Truncated unitary pair-hopper

Recall the definition of the anti-Hermitian pair-hopper operator [32]

$$\mathcal{T} = \sum_{p < q} \tau_{pq} (\mathbf{P}_p^\dagger \mathbf{P}_q - \mathbf{P}_q^\dagger \mathbf{P}_p), \quad (37)$$

where τ_{pq} is the amplitude and is antisymmetric. \mathcal{T} is derived from

$$\mathbf{K}_{pq} - \mathbf{K}_{pq}^\dagger \propto \mathbf{P}_p^\dagger \mathbf{P}_q - \mathbf{P}_q^\dagger \mathbf{P}_p, \quad (38)$$

where $\mathbf{K}_{pq}|\text{AGP}\rangle = 0$ is the two-body killer of AGP [29]. Its exact expression is

$$\begin{aligned} \mathbf{K}_{pq} = & \eta_p^2 \mathbf{P}_p^\dagger \mathbf{P}_q + \eta_q^2 \mathbf{P}_q^\dagger \mathbf{P}_p + \\ & \frac{1}{2} \eta_p \eta_q (\mathbf{N}_p \mathbf{N}_q - \mathbf{N}_p - \mathbf{N}_q). \end{aligned} \quad (39)$$

Notice that $e^{\mathcal{T}}|\text{HF}\rangle$ is equivalent to paired unitary CC with generalized indices (UGCC) [18, 66, 67].

Define the unitary pair-hopper ansatz as $e^{\mathcal{T}}|\text{AGP}\rangle$ which we denote as uP-AGP. We want to solve for the amplitudes by minimizing the energy, that is

$$E(\tau) = \langle \text{AGP} | e^{-\mathcal{T}} \mathbf{H} e^{\mathcal{T}} | \text{AGP} \rangle \quad (40a)$$

$$\tau^* = \underset{\tau_{pq} \in \mathbb{R}}{\text{argmin}} E(\tau). \quad (40b)$$

However, since the BCH expansion of the unitary transformed Hamiltonian does not truncate naturally, we choose to truncate the expansion at the highest order that gives us an affordable scaling, i.e. $\mathcal{O}(M^6)$. Again, using the reconstruction formulae, the expectation value of all two-body or higher density matrices can be accessed at $\mathcal{O}(1)$ cost per element. This means the asymptotic scaling of the commutator expansion is dominated by the many-body contractions. For example, for a two-body Hamiltonian, the first commutator, $\langle [\mathbf{H}, \mathcal{T}] \rangle$ contains three-body contractions; the second commutator, $\langle [[\mathbf{H}, \mathcal{T}], \mathcal{T}] \rangle$ contains four-body contractions, and so on. By truncating the expansion at the 4th commutator, we obtain a scaling of $\mathcal{O}(M^6)$ for the energy. In other words, we can state

$$\begin{aligned} E(\tau) \approx & \underbrace{\langle \mathbf{H} \rangle}_{\mathcal{O}(M^2)} + \underbrace{\langle [\mathbf{H}, \mathcal{T}] \rangle}_{\mathcal{O}(M^3)} + \frac{1}{2!} \underbrace{\langle [[\mathbf{H}, \mathcal{T}], \mathcal{T}] \rangle}_{\mathcal{O}(M^4)} \\ & + \frac{1}{3!} \underbrace{\langle [[[\mathbf{H}, \mathcal{T}], \mathcal{T}], \mathcal{T}] \rangle}_{\mathcal{O}(M^5)} + \frac{1}{4!} \underbrace{\langle [[[[\mathbf{H}, \mathcal{T}], \mathcal{T}], \mathcal{T}], \mathcal{T}] \rangle}_{\mathcal{O}(M^6)}, \end{aligned} \quad (41)$$

such that addition of every term in the expansion multiplies the leading scaling by a factor of M . Obtaining the equations above analytically is cumbersome but not prohibitive. We use our home-built algebraic manipulator software, *drudge* [68], to generate this and the analytic

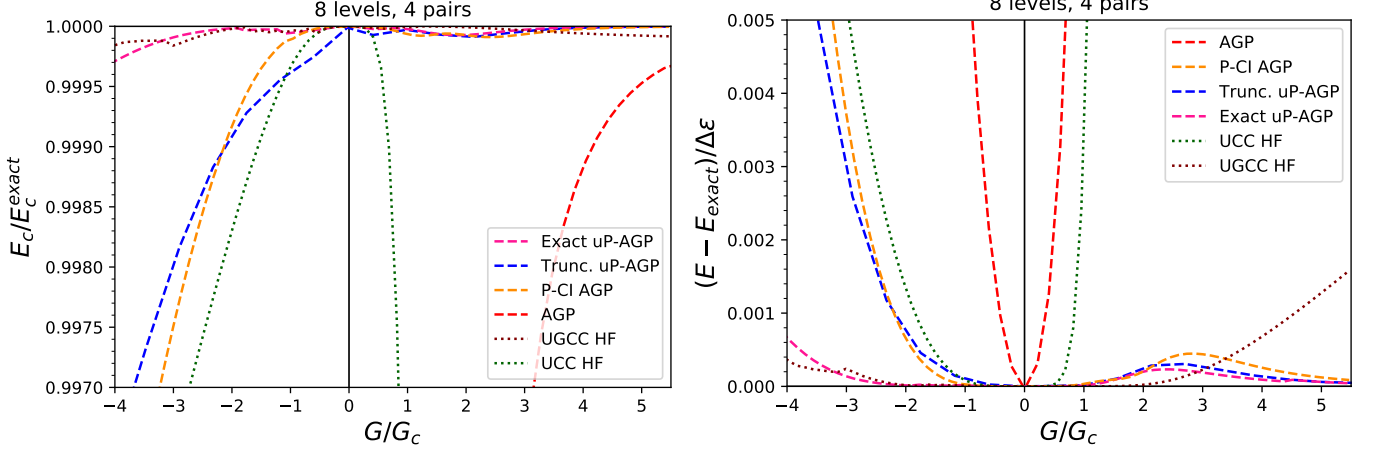


FIG. 2. Fraction of the correlation energy and the energy error of the unitary pair-hopper ansatz, with and without truncation, in comparison with other methods in a small system of 8 levels, half-filled. In addition to AGP based-methods, the plot shows the results for the single-reference unitary coupled cluster on HF. The left figure shows the correlation energy as the difference with HF and the right is the total energy error.

gradient of the energy. As a result, we can compute both $E(\tau)$ and $\nabla E(\tau)$ at $\mathcal{O}(M^6)$ cost albeit with a relatively large prefactor.

We minimize the energy as in Eq. (40) using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm [69] where we provide the gradients using our analytical expressions. We could alternatively find the solutions to $\nabla E(\tau) = 0$ using non-linear root-finding algorithms, but we find that the former is typically more robust when an accurate initial guess is provided.

Because truncating the commutator expansion yields a non-variational energy expression, we must take care with the minimization of the energy. This becomes particularly important in larger systems for which the numerical algorithms may go below the exact energy and eventually blow up. Fortunately, this failure mode is easy to identify as it is accompanied by extremely large amplitudes in τ . To avoid this problem, we introduce an L_2 regularization [70]

$$\tilde{E}(\tau) = E(\tau) + \frac{\lambda}{2} \|\tau\|^2 \quad (42a)$$

$$\implies \nabla \tilde{E}(\tau) = \nabla E(\tau) + \lambda \tau \quad (42b)$$

where $\lambda > 0$ is the regularization parameter and is sufficiently small to allow a robust convergence. In effect, the regularization penalizes the energy optimization in proportion to the magnitude of the amplitudes, thereby preventing it from blowing up. Although this compromises the accuracy of final energy, we gain a smoother convergence. To find λ , we pick a small value e.g. 10^{-3} , converge to an optimal solution for τ , and use this solution as the initial guess of another optimization in which a smaller λ is used. The first initial guess for τ can be set to zero. We repeat this process until we can no longer

converge or the change in energy is negligible. The convergence threshold of the gradient is set to 10^{-8} , and we have not observed significant dependence of the final results on the initial guess for τ or λ in our calculations.

In Fig. 2 we plot the results for the pairing Hamiltonian in a small system in which we can compare the accuracy of our truncated commutator expansion with that of exact uP-AGP which we obtain from a full configuration interaction (FCI) code. As we can see in the plot, the agreement between the exact and truncated uP-AGP is excellent on the attractive side while at the same time being slightly more accurate than the linear wavefunction $\mathcal{T}|\text{AGP}\rangle$ denoted as P-CI AGP. On the repulsive side, the results for the truncated uP-AGP is comparable to P-CI and becomes more accurate only at sufficiently small G/G_c .

For benchmarking purposes, we also plot HF-based unitary coupled cluster (UCC) [71, 72] as well as that with generalized indices, UGCC, in Fig. 2. The results are obtained from a FCI code. As expected UCC works well in the repulsive regime, but it breaks down as G approaches G_c . On the other hand, UGCC is remarkably more accurate than UCC, but it too eventually breaks down for large G/G_c . In contrast, our AGP based methods are well behaved in all correlation regimes.

The results for a larger system is show in Fig. 3 where we compare the accuracy of the truncated uP-AGP with ST-J₂AGP and P-CI. Although all of these methods have the same number of parameters, they clearly differ considerably in their accuracy. The observation that uP-AGP could be more accurate than P-CI in principle is a testament to its non-linear nature. Previously, it was shown that CI methods based on the generators of the algebra, namely J₂-CI, P-CI, and K-CI, yield identical results [31]. The other advantage of the non-linear model

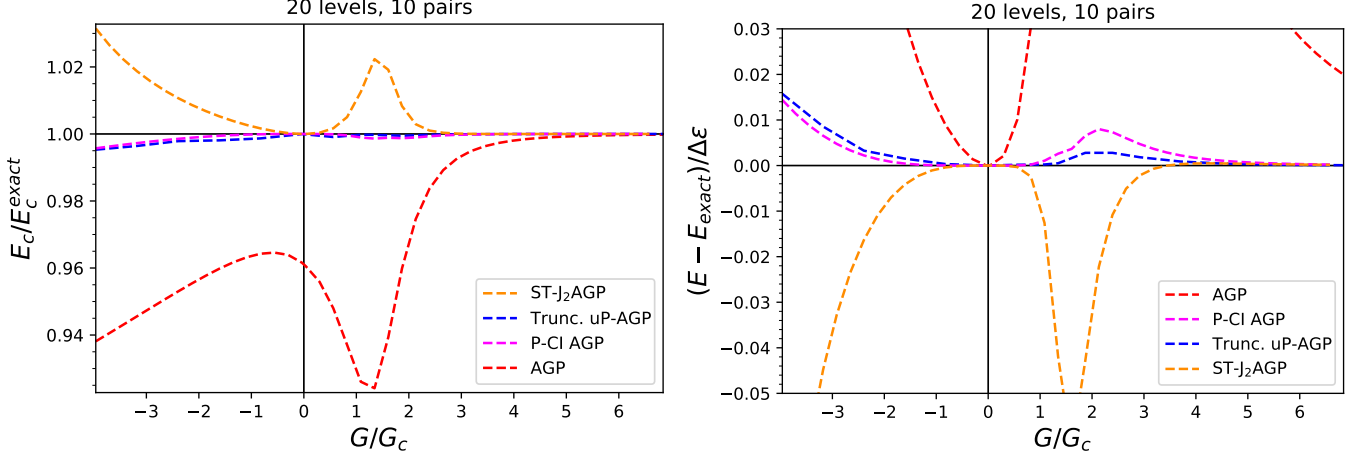


FIG. 3. Accuracy of the truncated unitary pair-hopper as compared with P-CI AGP and ST-J2AGP at 20 levels, half-filled. The left figure shows the correlation energy as the difference with HF and the right is the total energy error.

is that it is agnostic to the linear dependencies of the ansatz. This is true for both ST-J2AGP and the unitary pair-hopper. Although removing the linear dependencies by diagonalizing the metric can yield faster convergence, our numerical results show very little difference in the final energy.

The downside of the truncated uP-AGP is that it has a relatively large prefactor which is a result of the brute-force expansion of the BCH formula. In the next section we apply canonical transformation theory as another means for approximating the unitary transformed Hamiltonian.

B. AGP-based canonical transformation

In canonical transformation (CT) theory, as formulated originally by Yanai and Chan [53, 54], the commutator expansion of a unitary transformed Hamiltonian,

$$e^{-\mathbf{A}}\mathbf{H}e^{\mathbf{A}} = \mathbf{H} + [\mathbf{H}, \mathbf{A}] + \frac{1}{2!}[[\mathbf{H}, \mathbf{A}], \mathbf{A}] + \dots, \quad (43)$$

where \mathbf{A} is an anti-Hermitian operator, is recursively summed by systematically replacing all higher-body operators obtained from each commutator by an approximate, lower-body operator. For example, if \mathbf{A} and \mathbf{H} are each two-body operators, one approximates $[\mathbf{H}, \mathbf{A}] \rightarrow [\mathbf{H}, \mathbf{A}]_{(1,2)}$ wherein the three-body operator resulting from $[\mathbf{H}, \mathbf{A}]$ is approximated in terms of one- and two-body operators. This process yields an effective Hamiltonian, $\bar{\mathbf{H}} \approx e^{-\mathbf{A}}\mathbf{H}e^{\mathbf{A}}$, that, in its simplest form, can be written as

$$\bar{\mathbf{H}} = \mathbf{H} + [\mathbf{H}, \mathbf{A}]_{(1,2)} + \frac{1}{2!} [[\mathbf{H}, \mathbf{A}]_{(1,2)}, \mathbf{A}]_{(1,2)} + \dots, \quad (44)$$

where we have retained up to two-body terms in the effective Hamiltonian. Eq. (44) in particular can be summed recursively by

$$\bar{\mathbf{H}}^{(n+1)} = \frac{1}{n} [\bar{\mathbf{H}}^{(n)}, \mathbf{A}]_{(1,2)}, \quad (45a)$$

$$\bar{\mathbf{H}} = \sum_{n=0} \bar{\mathbf{H}}^{(n)}, \quad (45b)$$

where $\bar{\mathbf{H}}^{(0)} = \mathbf{H}$, and the sum is continued until the change in the parameters of the effective Hamiltonian is sufficiently small. The energy is then defined as $E = \langle \Psi | \bar{\mathbf{H}} | \Psi \rangle$ where $|\Psi\rangle$ is some mean-field reference. The minimization of the energy can be carried out by solving CC-style amplitude equations $\langle \Psi | \hat{\mathbf{A}}_i^\dagger \bar{\mathbf{H}} | \Psi \rangle$ which can be rearranged to give

$$R_i = \langle \Psi | [\bar{\mathbf{H}}, \hat{\mathbf{A}}_i] | \Psi \rangle = 0, \quad \forall \hat{\mathbf{A}}_i \in \mathbf{A}, \quad (46)$$

where $\hat{\mathbf{A}}_i$ are the operator parts of \mathbf{A} .

In essence, the goal of CT theory is to make the effective Hamiltonian more and more diagonal by successively approximating higher order excitations. It can be viewed as a kind of renormalization group approach and is rooted in an earlier work by White [73] as well as the flow renormalization approach by Wegner [74] and Glazek and Wilson [75]. In this section, however, we very closely follow the formalism presented in Ref. [53–57] by Yanai, Chan, and Neuscamman.

The chief difference between our approach and those of earlier papers is in the operator decomposition or *down-folding* scheme. Instead of invoking the cumulant decomposition of RDMs [76] as a means to get lower-body operators, we use AGP and our reconstruction formulae. Just as in cumulant decomposition, the reconstruction formulae are statements about the RDMs, but we can extend them to operators as an approximation.

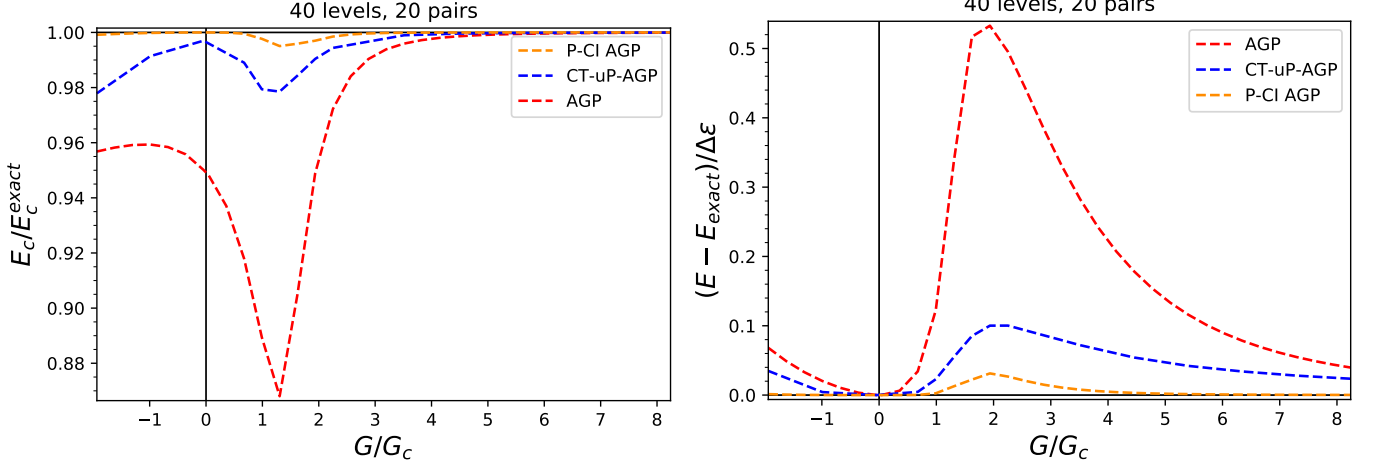


FIG. 4. Accuracy of canonical transformation with the unitary pair-hoppers, CT-uP-AGP, and that of P-CI AGP. The left figure is the correlation energy as the difference with HF and the right is the total energy error.

For example, consider the three-body irreducible density matrix $Z_{pqr}^{(3,3)}$. From the reconstruction formulae [26], it follows that

$$\langle \mathbf{N}_p \mathbf{N}_q \mathbf{N}_r \rangle = \Lambda_{qp} \Lambda_{rp} \langle \mathbf{N}_p \rangle + \Lambda_{pq} \Lambda_{rq} \langle \mathbf{N}_q \rangle + \Lambda_{pr} \Lambda_{qr} \langle \mathbf{N}_r \rangle, \quad (47)$$

where $\Lambda_{qp} = 2\eta_q^2/(\eta_q^2 - \eta_p^2)$. For CT, we propose to approximate $\mathbf{N}_p \mathbf{N}_q \mathbf{N}_r$ by removing the expectation values to get

$$\mathbf{N}_p \mathbf{N}_q \mathbf{N}_r \rightarrow \Lambda_{qp} \Lambda_{rp} \mathbf{N}_p + \Lambda_{pq} \Lambda_{rq} \mathbf{N}_q + \Lambda_{pr} \Lambda_{qr} \mathbf{N}_r. \quad (48)$$

Note that since the reconstruction formulae apply only to irreducible RDMs, we limit our downfolding to *irreducible operators*, i.e. $\mathbf{P}_p^\dagger \dots \mathbf{N}_q \dots \mathbf{P}_r$ where all indices are different. Indeed, all reducible operators (RDMs) can be written as a sum of irreducible ones.

To carry out the CT procedure, we take the unitary pair-hopper $e^{\mathcal{T}}$ ansatz to be the generator of our transformation, and we seek an effective Hamiltonian of the form

$$\bar{\mathbf{H}} = \sum_p h_p \mathbf{N}_p + \sum_{p \neq q} w_{pq} \mathbf{N}_p \mathbf{N}_q + \sum_{pq} v_{pq} \mathbf{P}_p^\dagger \mathbf{P}_q, \quad (49)$$

where w_{pq} is symmetric and v_{pq} is Hermitian. Note that Eq. (49) is the general form of a two-body Hamiltonian in the seniority-zero space [13]. In order to get a relatively more accurate approximation in CT, we would like to delay the downfolding as much possible. To this end, following Ref. [56], we split $\bar{\mathbf{H}} = \bar{\mathbf{H}}_1 + \bar{\mathbf{H}}_2$ such that $\bar{\mathbf{H}}_1$ and $\bar{\mathbf{H}}_2$ are the one- and two-body parts of the effective Hamiltonian respectively, and the downfolding is delayed

until four-body operators appear. As such, one obtains

$$\begin{aligned} \bar{\mathbf{H}}_1 &= \mathbf{H}_1 + [\mathbf{H}_1, \mathcal{T}] + \frac{1}{2!} [[\mathbf{H}_1, \mathcal{T}], \mathcal{T}]_{(1,2)} \\ &\quad + \frac{1}{3!} [[[\mathbf{H}_1, \mathcal{T}], \mathcal{T}], \mathcal{T}]_{(1,2)} + \dots \end{aligned} \quad (50a)$$

$$\begin{aligned} \bar{\mathbf{H}}_2 &= \mathbf{H}_2 + [\mathbf{H}_2, \mathcal{T}]_{(1,2)} + \frac{1}{2!} [[\mathbf{H}_2, \mathcal{T}], \mathcal{T}]_{(1,2)} \\ &\quad + \frac{1}{3!} [[[\mathbf{H}_2, \mathcal{T}], \mathcal{T}]_{(1,2)}, \mathcal{T}]_{(1,2)} + \dots \end{aligned} \quad (50b)$$

Similarly, to get accurate residual equations, we split $R_{pq} = R_{1,pq} + R_{2,pq}$ and evaluate the residuals associated with the one- and two-body parts recursively

$$R_{1,pq} = \langle [\mathbf{H}_1, \mathcal{T}_{pq}] \rangle + \frac{1}{2!} \langle [[\mathbf{H}_1, \mathcal{T}_{pq}], \mathcal{T}_{pq}]_{(1,2)} \rangle + \dots, \quad (51a)$$

$$R_{2,pq} = \langle [\mathbf{H}_2, \mathcal{T}_{pq}]_{(1,2)} \rangle + \frac{1}{2!} \langle [[\mathbf{H}_2, \mathcal{T}_{pq}], \mathcal{T}_{pq}]_{(1,2)} \rangle + \dots \quad (51b)$$

In both Eq. (50) and Eq. (51) we retain all two body-terms that naturally appear in the commutator expansion and absorb their coefficients into w_{pq} and v_{pq} in Eq. (49). All higher-body operators are downfolded directly into one-body with the intent of making the Hamiltonian more diagonal. The cost of evaluating the effective Hamiltonian and the residuals is $\mathcal{O}(M^4)$.

Having the effective Hamiltonian and the residual equations, one can solve the amplitudes using a non-linear root-finding algorithm. In so doing, we can provide an approximation to the Jacobian matrix as follows

$$J_{pq,rs} = \langle [[\bar{\mathbf{H}}, \mathcal{T}_{rs}], \mathcal{T}_{pq}] \rangle. \quad (52)$$

The cost of building the Jacobian is also $\mathcal{O}(M^4)$ by using the reconstruction formula.

We apply the CT as described above to the pairing Hamiltonian Eq. (2). As encountered in other implementations of CT, we find that the optimization over the amplitude equations is ill-conditioned [55–57]. These numerical difficulties are attributed to near-zero eigenvalues of the Jacobian and are said to be similar to the intruder states in the second order perturbation theory [56]. Several techniques have been proposed to mitigate these difficulties [56, 57]. In our implementation, we find that shifting of the amplitude equations [57]

$$\tilde{R}_\mu = R_\mu + \lambda \tau_\mu \quad (53a)$$

$$\implies \tilde{J}_{\mu\nu} = J_{\mu\nu} + \lambda \delta_{\mu\nu}, \quad (53b)$$

where μ, ν are flattened indices, to be the most effective. This is analogous to the regularization introduced in Eq. (42). The shifting parameter $\lambda > 0$ can be chosen *ad hoc* [57]. While we may follow the same steps as in the previous section to find optimal values for λ , for convenience, we picked some sufficiently small values with which the equations can converge reliably.

The results for 40 levels, 20 pairs is shown in Fig. 4. We picked $\lambda = 1$ for $G > 0$ and a larger $\lambda = 2.5$ for $G < 0$. CT with the unitary pair-hopper on AGP (CT-uP-AGP) is considerably less expensive than the truncated uP-AGP as well as the linear P-CI AGP. However, the approximation made to higher order excitations compromises the accuracy of CT-uP-AGP. Systematic improvements can be made to CT-uP-AGP by delaying the downfolding even further, but we leave these improvements for future work.

V. CONCLUSIONS

We explore several non-linear exponential ansätze based on AGP in this paper. First, we study Hilbert-space Jastrow correlators on AGP. We show analytically that while a non-Hermitian JAGP exponential ansatz can in principle be made exact, practical computational limitations with non-stochastic methods compel us to make certain approximations. To this end, we extend the formalism of Ref. [50] to AGP and the $su(2)$ algebra, and show how to similarity transform any given Hamiltonian in the seniority-zero space and solve for the amplitudes in a CC-style manner for AGP. Benchmark calculations of this ST-J₂AGP on the ground state of the pairing Hamiltonian show significant improvement on AGP but the resulting energies are less accurate compared to their CI counter-part, J₂-CI AGP. The variational ansatz, var-J₂AGP, is generally more accurate than the similarity transformed one, but it is out of reach for non-stochastic methods to the best of our knowledge. The cost of ST-J₂AGP is $\mathcal{O}(M^4)$.

In the second half of the paper, we sought an approximation to the unitary pair-hopper ansatz. Owing to its non-linear nature, exact variational calculations show uP-AGP can be considerably more accurate than the lin-

ear P-CI AGP. However, due to its non truncating commutator expansion, some approximation must be made to implement it efficiently on a classical computer. For an approximation to uP-AGP, we expand $e^{-\mathcal{T}H\mathcal{T}}$ using the well-known BCH formula up to the highest affordable computational scaling $\mathcal{O}(M^6)$. Numerical results indicate improvement over P-CI AGP on the attractive regime of the pairing Hamiltonian.

The downside of truncated uP-AGP is that it has a relatively large prefactor in its asymptotic scaling. To resolve this, we implement a slight variation of canonical transformation theory [56] wherein we carry out the operator decomposition from the reconstruction formulae [26] instead of the commonly used cumulant decomposition [76]. Our CT-uP-AGP is significantly less expensive than either truncated uP-AGP or P-CI AGP, but the approximation made to the higher order excitations noticeably compromises its accuracy.

Considering the pros and cons of each method, it is far from clear whether exponential ansätze for post-AGP methods are necessarily more advantageous than linear correlators. While there are non-linear ansätze that are more accurate than their linear counterparts, practical considerations for implementing them could limit their scope. So far, the linear ansätze have shown remarkable accuracy, they are straightforward to implement and seem rather resilient to system size, at least in the pairing Hamiltonian.

In electronic structure theory, exponential ansätze typically enjoy more favorable size-extensivity properties. However, we intentionally avoided discussing size-extensivity in this paper since the interactions in the pairing Hamiltonian are infinite-range. This makes the energy a non-linear function of system size and complicates the discussion of size-extensivity. Nevertheless, it must be noted that Neuscamman has shown [77] variational JAGP is fully size-consistent. Also, earlier work in our group has shown a significant but incomplete restoration of size-consistency using K-CI and linearized-CC correlators on AGP [25] with the latter exhibiting more favorable size-consistency. A more thorough analysis regarding size-extensivity of post-AGP methods will be reported in due time.

VI. ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation under Grant No. CHE-1762320. G.E.S. is a Welch Foundation Chair (Grant No. c-0036). A.K. is thankful to Yiheng Qiu for useful discussions regarding Jastrow AGP and to Gaurav Harsha for assistance with the drudge software. We thank Rishab Dutta for letting us use his JCI-AGP code.

VII. DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix A: Proof for J_k CI Manifold Containing Lower-Rank JCI Manifolds over AGP

The J_k CI-AGP manifold contains the J_{k-1} CI-AGP and thus all lower-rank JCI-AGP manifolds. This has been observed numerically in Ref. [31]. Here we provide a proof for the $k = 2$ case, which can be readily generalized to higher ranks. Along the same lines, J_k AGP does not need to include lower-rank Jastrow operators, which is shown as a corollary.

For an AGP with N pairs of electrons

$$\begin{aligned}
 & \mathbf{J}_1|\text{AGP}\rangle \\
 &= \frac{1}{2N} \sum_p \mathbf{N}_p \mathbf{J}_1|\text{AGP}\rangle \\
 &= \frac{1}{4N} \sum_{pq} \alpha_q \mathbf{N}_p \mathbf{N}_q |\text{AGP}\rangle \\
 &= \frac{1}{4N} \sum_{p<q} (\alpha_p + \alpha_q) \mathbf{N}_p \mathbf{N}_q |\text{AGP}\rangle + \frac{1}{4N} \sum_p \alpha_p \mathbf{N}_p^2 |\text{AGP}\rangle \\
 &= \frac{1}{4N} \sum_{p<q} (\alpha_p + \alpha_q) \mathbf{N}_p \mathbf{N}_q |\text{AGP}\rangle + \frac{1}{2N} \sum_p \alpha_p \mathbf{N}_p |\text{AGP}\rangle.
 \end{aligned}$$

It follows that

$$\mathbf{J}_1|\text{AGP}\rangle = \frac{1}{4} \sum_{p<q} \frac{\alpha_p + \alpha_q}{N-1} \mathbf{N}_p \mathbf{N}_q |\text{AGP}\rangle. \quad (\text{A1})$$

Namely, $\mathbf{J}_1|\text{AGP}\rangle$ can be expressed as a linear combination of states in J_2 CI.

From Eq. (A1), we can show Eq. (15) by defining

$$\alpha'_{pq} = \frac{\alpha_p + \alpha_q}{N-1} + \alpha_{pq} \quad (\text{A2})$$

and

$$\mathbf{J}'_2 = \frac{1}{4} \sum_{p<q} \alpha'_{pq} \mathbf{N}_p \mathbf{N}_q. \quad (\text{A3})$$

Generalization to higher-rank cases is straightforward.

-
- [1] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular electronic-structure theory* (Wiley, Chichester; New York, 2000).
 - [2] G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, *J. Chem. Phys.* **89**, 7382 (1988).
 - [3] R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.* **79**, 291 (2007).
 - [4] I. W. Bulik, T. M. Henderson, and G. E. Scuseria, *J. Chem. Theory Comput.* **11**, 3171 (2015).
 - [5] M. Degroote, T. M. Henderson, J. Zhao, J. Dukelsky, and G. E. Scuseria, *Phys. Rev. B* **93**, 125124 (2016).
 - [6] Y. Qiu, T. M. Henderson, J. Zhao, and G. E. Scuseria, *J. Chem. Phys.* **147**, 064111 (2017).
 - [7] B. O. Roos, P. R. Taylor, and P. E. M. Sigbahn, *Chemical Physics* **48**, 157 (1980).
 - [8] J. Olsen, *Int. J. Quant. Chem.* **111**, 3267 (2011).
 - [9] P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, *Chem. Rev.* **112**, 108 (2012).
 - [10] D. I. Lyakh, M. Musiał, V. F. Lotrich, and R. J. Bartlett, *Chem. Rev.* **112**, 182 (2012).
 - [11] P. Ring and P. Schuck, *The Nuclear Many-Body Problem*, Theoretical and Mathematical Physics, The Nuclear Many-Body Problem (Springer-Verlag, Berlin Heidelberg, 1980).
 - [12] L. Bytautas, T. M. Henderson, C. A. Jiménez-Hoyos, J. K. Ellis, and G. E. Scuseria, *J. Chem. Phys.* **135**, 044119 (2011).
 - [13] T. M. Henderson, I. W. Bulik, and G. E. Scuseria, *J. Chem. Phys.* **142**, 214116 (2015).
 - [14] L. Bytautas, G. E. Scuseria, and K. Ruedenberg, *J. Chem. Phys.* **143**, 094105 (2015).
 - [15] A. Veillard and E. Clementi, *Theoret. Chim. Acta* **7**, 133 (1967).
 - [16] M. Couty and M. B. Hall, *J. Phys. Chem. A* **101**, 6936 (1997).
 - [17] C. Kollmar and B. A. Heß, *J. Chem. Phys.* **119**, 4655 (2003).
 - [18] T. Stein, T. M. Henderson, and G. E. Scuseria, *J. Chem. Phys.* **140**, 214113 (2014).
 - [19] P. A. Limacher, P. W. Ayers, P. A. Johnson, S. De Baerdemacker, D. Van Neck, and P. Bultinck, *J. Chem. Theory Comput.* **9**, 1394 (2013).
 - [20] P. A. Johnson, P. W. Ayers, P. A. Limacher, S. D. Baerdemacker, D. V. Neck, and P. Bultinck, *Comput. Theor. Chem.* **1003**, 101 (2013).
 - [21] T. M. Henderson, I. W. Bulik, T. Stein, and G. E. Scuseria, *J. Chem. Phys.* **141**, 244104 (2014).
 - [22] J. J. Shepherd, T. M. Henderson, and G. E. Scuseria, *J. Chem. Phys.* **144**, 094112 (2016).
 - [23] J. Dukelsky, S. Pittel, and G. Sierra, *Rev. Mod. Phys.*

- 76**, 643 (2004).
- [24] T. M. Henderson, G. E. Scuseria, J. Dukelsky, A. Signoracci, and T. Duguet, *Phys. Rev. C* **89**, 054305 (2014).
 - [25] T. M. Henderson and G. E. Scuseria, *J. Chem. Phys.* **153**, 084111 (2020).
 - [26] A. Khamoshi, T. M. Henderson, and G. E. Scuseria, *J. Chem. Phys.* **151**, 184103 (2019).
 - [27] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).
 - [28] R. W. Richardson and N. Sherman, *Nucl. Phys.* **52**, 221 (1964).
 - [29] T. M. Henderson and G. E. Scuseria, *J. Chem. Phys.* **151**, 051101 (2019).
 - [30] G. Harsha, T. M. Henderson, and G. E. Scuseria, *J. Chem. Phys.* **153**, 124115 (2020).
 - [31] R. Dutta, T. M. Henderson, and G. E. Scuseria, *J. Chem. Theory Comput.* **16**, 6358 (2020).
 - [32] A. Khamoshi, F. A. Evangelista, and G. E. Scuseria, *Quantum Sci. Technol.* **6**, 014004 (2020).
 - [33] P. A. Johnson, C.-E. Fecteau, F. Berthiaume, S. Cloutier, L. Carrier, M. Gratton, P. Bultinck, S. De Baerdemacker, D. Van Neck, P. Limacher, and P. W. Ayers, *J. Chem. Phys.* **153**, 104110 (2020).
 - [34] C.-E. Fecteau, H. Fortin, S. Cloutier, and P. A. Johnson, *J. Chem. Phys.* **153**, 164117 (2020).
 - [35] C.-E. Fecteau, F. Berthiaume, M. Khalfoun, and P. A. Johnson, *J. Math. Chem.* (2020), <https://doi.org/10.1007/s10910-020-01197-0>.
 - [36] P. A. Johnson, H. Fortin, S. Cloutier, and C.-E. Fecteau, *arXiv:2012.10477* (2020).
 - [37] G. E. Scuseria, C. A. Jiménez-Hoyos, T. M. Henderson, K. Samanta, and J. K. Ellis, *J. Chem. Phys.* **135**, 124108 (2011).
 - [38] R. Jastrow, *Phys. Rev.* **98**, 1479 (1955).
 - [39] F. D. M. Haldane, *Phys. Rev. Lett.* **60**, 635 (1988).
 - [40] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
 - [41] M. Casula and S. Sorella, *J. Chem. Phys.* **119**, 6500 (2003).
 - [42] M. Casula, C. Attaccalite, and S. Sorella, *J. Chem. Phys.* **121**, 7110 (2004).
 - [43] N. Umezawa and S. Tsuneyuki, *J. Chem. Phys.* **119**, 10015 (2003).
 - [44] S. Tsuneyuki, *Prog. Theor. Phys.* **176**, 134 (2008).
 - [45] T. M. Henderson and G. E. Scuseria, *J. Chem. Phys.* **139**, 234113 (2013).
 - [46] E. Neuscamman, *J. Chem. Phys.* **139**, 181101 (2013).
 - [47] E. Neuscamman, *J. Chem. Theory Comput.* **12**, 3149 (2016).
 - [48] C. Genovese, A. Meninno, and S. Sorella, *J. Chem. Phys.* **150**, 084102 (2019).
 - [49] K. Nakano, R. Maezono, and S. Sorella, *J. Chem. Theory Comput.* **15**, 4044 (2019).
 - [50] J. M. Wahlen-Strothman, C. A. Jiménez-Hoyos, T. M. Henderson, and G. E. Scuseria, *Phys. Rev. B* **91**, 041114 (2015).
 - [51] J. M. Wahlen-Strothman and G. E. Scuseria, *J. Phys.: Condens. Matter* **28**, 485502 (2016).
 - [52] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, *Chem. Rev.* **119**, 10856 (2019).
 - [53] T. Yanai and G. K.-L. Chan, *J. Chem. Phys.* **124**, 194106 (2006).
 - [54] T. Yanai and G. K.-L. Chan, *J. Chem. Phys.* **127**, 104107 (2007).
 - [55] E. Neuscamman, T. Yanai, and G. K.-L. Chan, *J. Chem. Phys.* **130**, 124102 (2009).
 - [56] E. Neuscamman, T. Yanai, and G. K.-L. Chan, *Int. Rev. Phys. Chem.* **29**, 231 (2010).
 - [57] T. Yanai, Y. Kurashige, E. Neuscamman, and G. K.-L. Chan, *Phys. Chem. Chem. Phys.* **14**, 7809 (2012).
 - [58] A. J. Coleman, *J. Math. Phys.* **6**, 1425 (1965).
 - [59] P. R. Surján, *Top. Curr. Chem.* **203**, 63 (1999).
 - [60] L.-K. Hua, *Am. J. Math.* **66**, 531 (1944).
 - [61] J. Dukelsky, S. Pittel, and C. Eсеbbag, *Phys. Rev. C* **93**, 034313 (2016).
 - [62] R. Dutta, G. P. Chen, T. M. Henderson, and G. E. Scuseria, *arXiv:2101.08911* (2021).
 - [63] E. Neuscamman, *J. Chem. Phys.* **139**, 194105 (2013).
 - [64] H. Wei and E. Neuscamman, *J. Chem. Phys.* **149**, 184106 (2018).
 - [65] A. J. Cohen, H. Luo, K. Guthrie, W. Dobrutz, D. P. Tew, and A. Alavi, *J. Chem. Phys.* **151**, 061101 (2019).
 - [66] M. Nooijen, *Phys. Rev. Lett.* **84**, 2108 (2000).
 - [67] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, *J. Chem. Theory Comput.* **15**, 311 (2019).
 - [68] J. Zhao, *Symbolic solution for computational quantum many-body theory development*, *Thesis* (2018).
 - [69] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *SIAM J. Sci. Comput.* **16**, 1190 (1995).
 - [70] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer Science & Business Media, New York, NY, 2006).
 - [71] R. J. Bartlett, S. A. Kucharski, and J. Noga, *Chem. Phys. Lett.* **155**, 133 (1989).
 - [72] A. G. Taube and R. J. Bartlett, *Int. J. Quantum Chem.* **106**, 3393 (2006).
 - [73] S. R. White, *J. Chem. Phys.* **117**, 7472 (2002).
 - [74] F. Wegner, *Ann. Phys.* **506**, 77 (1994).
 - [75] S. D. Glazek and K. G. Wilson, *Phys. Rev. D* **49**, 4214 (1994).
 - [76] W. Kutzelnigg and D. Mukherjee, *J. Chem. Phys.* **110**, 2800 (1999).
 - [77] E. Neuscamman, *Phys. Rev. Lett.* **109**, 203001 (2012).