

# Green's function coupled cluster formulations utilizing extended inner excitations

Cite as: J. Chem. Phys. **149**, 214102 (2018); <https://doi.org/10.1063/1.5046529>

Submitted: 28 June 2018 . Accepted: 15 November 2018 . Published Online: 05 December 2018

Bo Peng, and Karol Kowalski



View Online



Export Citation



CrossMark

## ARTICLES YOU MAY BE INTERESTED IN

[Perspective: Computational chemistry software and its advancement as illustrated through three grand challenge cases for molecular science](#)

The Journal of Chemical Physics **149**, 180901 (2018); <https://doi.org/10.1063/1.5052551>

[Exact and approximate adiabatic connection formulae for the correlation energy in multireference ground and excited states](#)

The Journal of Chemical Physics **149**, 204101 (2018); <https://doi.org/10.1063/1.5048988>

[Perspective: Multireference coupled cluster theories of dynamical electron correlation](#)

The Journal of Chemical Physics **149**, 030901 (2018); <https://doi.org/10.1063/1.5039496>



# Green's function coupled cluster formulations utilizing extended inner excitations

Bo Peng<sup>a)</sup> and Karol Kowalski<sup>b)</sup>

William R. Wiley Environmental Molecular Sciences Laboratory, Battelle, Pacific Northwest National Laboratory, K8-91, P.O. Box 999, Richland, Washington 99352, USA

(Received 28 June 2018; accepted 15 November 2018; published online 5 December 2018)

In this paper, we analyze new approximations of the Green's function coupled cluster (GFCC) method where locations of poles are improved by extending the excitation level of inner auxiliary operators. These new GFCC approximations can be categorized as the GFCC- $i(n, m)$  method, where the excitation level of the inner auxiliary operators ( $m$ ) used to describe the ionization potential and electron affinity effects in the  $N - 1$  and  $N + 1$  particle spaces is higher than the excitation level ( $n$ ) used to correlate the ground-state coupled cluster wave function for the  $N$ -electron system. Furthermore, we reveal the so-called " $n + 1$ " rule in this category [or the GFCC- $i(n, n + 1)$  method], which states that in order to maintain size-extensivity of the Green's function matrix elements, the excitation level of inner auxiliary operators  $X_p(\omega)$  and  $Y_q(\omega)$  cannot exceed  $n + 1$ . We also discuss the role of the moments of coupled cluster equations that in a natural way assures these properties. Our implementation in the present study is focused on the first approximation in this GFCC category, i.e., the GFCC- $i(2,3)$  method. As our first practice, we use the GFCC- $i(2,3)$  method to compute the spectral functions for the  $N_2$  and CO molecules in the inner and outer valence regimes. In comparison with the Green's function coupled cluster singles, doubles results, the computed spectral functions from the GFCC- $i(2,3)$  method exhibit better agreement with the experimental results and other theoretical results, particularly in terms of providing higher resolution of satellite peaks and more accurate relative positions of these satellite peaks with respect to the main peak positions. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5046529>

## I. INTRODUCTION

Green's function formalism<sup>1-6</sup> is a broadly used tool to calculate properties such as total energies, densities, ionization potentials (IPs), electron affinities (EAs), and neutral excitations for molecules, clusters, nano-structures, and solids. This flexibility of the Green's function formalism, stemming from the possibility of characterizing many-body systems in terms of electron removal and/or addition processes, has motivated numerous developments towards providing a theoretical framework that can efficiently combine accuracy and relatively low numerical cost. Various computational schemes based on the algebraic and diagrammatic perturbative approaches have been developed so far. Typical examples include the outer valence Green's function (OVGF) approach,<sup>7-10</sup> the diagonal third-order self-energy approximation (D3) (for the review of diagonal Green function approximations, see Ref. 11), the non-diagonal renormalized second-order approach (NR2),<sup>12</sup> the two-hole-one-particle Tamm-Dancoff approximation ( $2h - p$  TDA),<sup>13</sup> the third-order quasiparticle method P3,<sup>14</sup> the algebraic-diagrammatic construction (ADC),<sup>8,15,16</sup> and the algebraic approaches based on the intermediate-state representations.<sup>17,18</sup> In addition, the state-of-the-art  $GW$

approximations (one-body Green's function  $G$  and dynamically screened Coulomb interaction  $W$ )<sup>19-24</sup> permeated many areas of computational chemistry and materials sciences.<sup>25-27</sup>

Recently, one could witness an increasing interest in developing Green's function approaches accounting for higher-order correlation effects, which turns out to be indispensable in describing correlated behavior of electrons. These efforts serve multiple purposes and have to meet certain requirements. For example, in order to support the analysis and interpretation of photoelectron and various X-ray spectroscopies, the Green's function formulations have to efficiently deal with various energy regimes. Also, as can be seen from a class of Green's function applications that are motivated by the development of various embedding methods including dynamical mean field theories (DMFT)<sup>28-34</sup> and self-energy embedding theory (SEET)<sup>35-38</sup> approaches, very accurate representations of Green's function and/or corresponding self-energies are necessary in describing the so-called impurity regions, which are usually associated with the presence of strong correlation effects.

Hierarchical approximations have been recently proposed to describe self-energies using perturbative techniques for one-particle many-body Green's function (MBGF)<sup>39,40</sup> that have been used to re-derive linked cluster and irreducible-diagram theorems as well as to provide algorithms for the general order component of the self-energy. A lot of attention has also been attracted by the Green's function formulation

<sup>a)</sup>peng398@pnnl.gov

<sup>b)</sup>karol.kowalski@pnnl.gov

proposed by Nooijen and Snijders,<sup>41–43</sup> who successfully employed the coupled cluster (CC) bi-orthogonal formalism to express Green's function matrix in terms of the cluster operator ( $T$ ) and the so-called  $\Lambda$  operator, which is frequently used in linear response CC (LRCC) theory.<sup>44,45</sup> In particular, the latter approach provides a natural elimination of the problems associated with normalization of the ground-state wave function in exponential CC representation. More recently, algorithmic developments for calculating Green's function CC (GFCC) have been focused on the design and solving of the auxiliary  $\omega$ -dependent ionization potential/electron affinity equation-of-motion CC (IP/EA-EOM-CC) type  $X_p(\omega)$  and  $Y_q(\omega)$  operators acting in the  $N - 1$  and  $N + 1$  electron spaces (see Refs. 46 and 47 for details). This representation can be used to prove/derive basic properties of GFCC including connected character of Green's function matrix elements and their arbitrary-order  $\omega$ -derivatives.<sup>48,49</sup> The possibility of calculating the GFCC matrix and their  $\omega$ -derivatives, through the Dyson equations, extends in a natural way to the corresponding self-energies. These properties enable one to calculate spectral functions, pole strengths, and other properties using the GFCC method.

In this paper, we further extend the analysis carried out in the previous papers.<sup>46–50</sup> In particular, motivated by the accuracy attainable by the GFCC model with singles, doubles, and triples (GFCCSDT),<sup>47</sup> we explore the possibilities of further improving accuracies of calculated pole locations by only increasing the excitation level in the auxiliary inner operators  $X_p(\omega)$  and  $Y_q(\omega)$ . To this end, we discuss the general class of GFCC- $i(n, m)$  approximations, where  $n$  is the rank of excitations included in the  $T$  and  $\Lambda$  operators, while  $m$  designates the excitation level used to define  $X_p(\omega)$  and  $Y_q(\omega)$  operators. Also, using consistency conditions between  $N$  and  $N - 1/N + 1$  particle systems introduced by Nooijen and Snijders,<sup>42</sup> for the first time ever, we analyze results of the combined GFCC- $i(n, m)$  formalisms from the point of view of properties of approximate GFCC spectral functions. We will demonstrate that, if  $m > n + 1$ , the disconnected components in the equations for  $X_p(\omega)$  and  $Y_q(\omega)$  will lead to the appearance of disconnected components in the corresponding Green's function. These disconnected components originate in the non-vanishing higher-order moments of the CC equations for the  $N$ -electron ground-state problem. Moreover, if  $m = n + 1$ , the equations for  $X_p(\omega)$  and  $Y_q(\omega)$  operators are defined in terms of connected diagrams, which leads in turn to the fully connected representation of the corresponding GFCC- $i(n, n+1)$  Green's function. We will refer to this property of the GFCC- $i(n, m)$  formulation as an “ $n + 1$ ”-rule. We will illustrate the performance of the GFCC- $i(n, n + 1)$  formalism by applying the GFCC- $i(2,3)$  method to the computation of the spectral functions of the  $N_2$  and CO benchmark systems in the inner and outer valence regime, where several challenging satellite peaks have been identified. The GFCC- $i(2,3)$  spectral functions/pole locations will be compared with those obtained with GFCCSD, IP-EOM-CC, configuration interaction (CI), and the experimental results. We will also discuss the favorable numerical scaling of the GFCC- $i(2,3)$  method, which obviates the need of correlating the ground state of the  $N$ -electron at the CCSDT level that is associated with very high memory

requirements limiting the application area of the GFCCSDT method.

## II. METHODOLOGY

The standard form for the matrix elements of the one-body Green function written in terms of a normalized ground-state wave function  $|\Psi\rangle$  for the  $N$ -electron system can be represented as

$$G_{pq}(\omega) = \langle \Psi | a_q^\dagger(\omega + (H - E_0) - i\eta)^{-1} a_p | \Psi \rangle + \langle \Psi | a_p(\omega + (H - E_0) - i\eta)^{-1} a_q^\dagger | \Psi \rangle. \quad (1)$$

To construct the GFCC formulation, the direct use of the CC parameterization of the wave function is prohibitive, as in the general case, the exponential form of the wave function cannot be easily normalized. To overcome the normalization problem, Nooijen and Snijders<sup>41–43</sup> first employed the biorthogonal CC formalism (see, for example, Refs. 51 and 52 for recent reviews of the biorthogonal CC method) to Eq. (1). Similarly, we can also introduce the bi-variational CC formalism into Eq. (1). In the bi-variational formalism, the parameterizations of the bra ( $\langle \Psi |$ ) and ket ( $|\Psi\rangle$ ) are defined as follows:

$$\langle \Psi | = \langle \Phi | (1 + \Lambda) e^{-T}, \quad (2)$$

$$|\Psi\rangle = e^T |\Phi\rangle. \quad (3)$$

Plugging Eqs. (2) and (3) to Eq. (1) and introducing resolution of identity  $1 = e^{-T} e^T$ , the Green's function CC formulation can then be expressed as

$$G_{pq}(\omega) = \langle \Phi | (1 + \Lambda) \bar{a}_q^\dagger(\omega + \bar{H}_N - i\eta)^{-1} \bar{a}_p | \Phi \rangle + \langle \Phi | (1 + \Lambda) \bar{a}_p(\omega - \bar{H}_N + i\eta)^{-1} \bar{a}_q^\dagger | \Phi \rangle, \quad (4)$$

where  $|\Phi\rangle$  is the reference function, the  $\omega$  parameter denotes the frequency, and the imaginary part  $\eta$  is often called the broadening factor. The similarity transformed operators  $\bar{H}_N$  (in its norm product representation),  $\bar{a}_p$ , and  $\bar{a}_q^\dagger$  are defined as

$$\bar{H}_N = e^{-T} H e^T - E_0, \quad (5)$$

$$\bar{a}_p = e^{-T} a_p e^T = a_p + [a_p, T], \quad (6)$$

$$\bar{a}_q^\dagger = e^{-T} a_q^\dagger e^T = a_q^\dagger + [a_q^\dagger, T]. \quad (7)$$

Here,  $E_0$  is the CC energy,  $T$  is the cluster operator, and  $\Lambda$  is the de-excitation operator.  $T$  and  $\Lambda$  are defined as

$$T = \sum_{n=1}^N \frac{1}{(n!)^2} \sum_{\substack{i_1, \dots, i_n \\ a_1, \dots, a_n}} t_{a_1 \dots a_n}^{i_1 \dots i_n} a_{a_1}^\dagger \dots a_{a_n}^\dagger a_{i_1} \dots a_{i_n}, \quad (8)$$

$$\Lambda = \sum_{n=1}^N \frac{1}{(n!)^2} \sum_{\substack{i_1, \dots, i_n \\ a_1, \dots, a_n}} \lambda_{a_1 \dots a_n}^{i_1 \dots i_n} a_{i_1}^\dagger \dots a_{i_n}^\dagger a_{a_1} \dots a_{a_n}, \quad (9)$$

with  $t_{a_1 \dots a_n}^{i_1 \dots i_n}$  and  $\lambda_{a_1 \dots a_n}^{i_1 \dots i_n}$  being the antisymmetric amplitudes and the indices  $i, j, k, \dots$  ( $i_1, i_2, \dots$ ) and  $a, b, c, \dots$  ( $a_1, a_2, \dots$ ) corresponding to occupied and unoccupied spin-orbitals in the reference function  $|\Phi\rangle$ , respectively.  $E_0$ ,  $T$ ,  $\Lambda$  can be obtained from the conventional CC equations

$$Q e^{-T} H e^T |\Phi\rangle = 0, \quad (10)$$

$$\langle \Phi | e^{-T} H e^T |\Phi\rangle = E_0, \quad (11)$$

$$\langle \Phi | (1 + \Lambda) e^{-T} H e^T Q = E_0 \langle \Phi | (1 + \Lambda) Q, \quad (12)$$

where the  $Q$  is a projection operator,

$$Q = \sum_{n=1}^N \frac{1}{(n!)^2} \sum_{\substack{i_1, \dots, i_n; \\ a_1, \dots, a_n}} |\Phi_{i_1 \dots i_n}^{a_1 \dots a_n}\rangle \langle \Phi_{i_1 \dots i_n}^{a_1 \dots a_n}|, \quad (13)$$

representing the projection onto the subspace spanned by excited configurations  $|\Phi_{i_1 \dots i_n}^{a_1 \dots a_n}\rangle$  defined as  $a_{a_1}^\dagger \dots a_{a_n}^\dagger a_{i_1} \dots a_{i_n} |\Phi\rangle$ . Note that other parameterizations of the  $N$ -electron wave function can also be used to generate their corresponding Green's function formulations.

To evaluate the  $G_{pq}(\omega)$  matrix elements from Eq. (4), one can formally diagonalize the non-Hermitian effective Hamiltonian  $\bar{H} = e^{-T} H e^T$  in the  $(N \pm 1)$ -particle space. For the evaluation of the  $G_{pq}(\omega)$  matrix elements in the CCSD level, the dimension of the secular matrix ( $\bar{H}_N$ ) is  $n_o + n_o^2 n_v$  for the retarded part and  $n_v + n_v^2 n_o$  for the advanced part, where  $n_o$  and  $n_v$  denote the numbers of occupied molecular orbitals and virtual molecular orbitals, respectively. When systems of interest are getting larger, the increasing number of orbitals prohibits the direct diagonalization of the secular matrix, and an iterative algorithm such as *Krylov* subspace methods is often used. However, since only part of the spectrum is solved (or converged) by the conventional iterative subspace diagonalization methods, the introduction of the incomplete eigenpairs to the GFCC formulation would lead to a sum-over-states representation of the GFCC formulation (see Ref. 47), which is usually unfavored in a sense that there is unclearness of how many states should be involved in the representation. Besides, for systems that exhibit a significant many-body effect in the spectral region of interest, solving the secular equations by the conventional iterative diagonalization is often subject to nontrivial construction of initial vectors and convergence problems. To circumvent these problems, approximations such as core-valence separation (CVS)<sup>53</sup> have been introduced in the high-order Green's function method to compute the K-shell ionization spectra of small and medium-size molecules.<sup>54,55</sup> Essentially, the CVS approximation neglects the coupling between core- and valence-excited states (set certain types of Coulomb integrals to zero in practice) and, therefore, reduces the dimension of the effective Hamiltonian for the problem of interest. The reduction of dimension of the effective Hamiltonian becomes more significant for core-electron ionization than for the valence-electron ionization, and the associated error was estimated to be 0.5–1.0 eV for the former and much smaller for the latter.<sup>56</sup> In particular, for the core-electron ionization, the performance of the convergence would be greatly improved when the CVS is employed, and vast majority states of interest would then be easily positioned. In the context of wave-function-based or propagator-based *ab initio* methods, the CVS approximation has been applied to the ADC, CC2, CCSD, CC3, and CCSDR(3) methods see Ref. 57 for a recent review). Recently, on the computational side, more algorithms have also been proposed for iterative matrix-free eigensolvers [such as the asymmetric Lanczos-chain-driven subspace algorithm,<sup>58</sup> energy-specific Davidson algorithm,<sup>59</sup> Generalized Preconditioned Locally Harmonic Residual (GPLHR) method,<sup>60</sup> etc.] and have been proposed to work with the LRCC or EOM-CC method to solve not only the lowest/highest eigenvalues but also the inner eigenvalues

embedded deeply in the spectrum of the large Jacobian matrix such that accurate resolution can be obtained for the frequency window of interest, and viable approximation can even be generated for the entire spectrum. For the GFCC calculation, as mentioned in a recent study of the optical potential calculation for the nucleon-nucleon scattering,<sup>61</sup> the advantage of using the Lanczos-based method lies in the fact that the resolution of the tridiagonal matrix representing the normal-ordered Hamiltonian matrix in the Lanczos basis only needs to be done once for all frequencies ( $\omega$ 's). Alternatively, we recently proposed to first solve a set of linear equations for IP/EA-EOMCC type vectors (using their zeroth and first order terms as starting points) and then to contract these vectors with the converged amplitudes of the CC  $\Lambda$  de-excitation operators to get the GFCC matrix elements.<sup>46,47</sup> The computational approach was designed to calculate GFCC for the whole complex plane; therefore, it includes all poles of the GFCC structure and serves naturally for, for example, an embedding scheme when working with low-level methods for large-scale applications. Facilitated by this approach, we then were able to prove the connected character of the diagrams contributing to GFCC matrix elements, as well as the connected character of its  $n$ th order derivative with respect to the energy and the corresponding CC self-energy operators,<sup>48,49</sup> which provided a useful guidance for designing and analyzing new and size-extensive GFCC approximation schemes. Furthermore, due to its algebraic structure, the proposed method is highly scalable and is capable of computing the spectral function for a given molecular system in any energy region. We recently demonstrated this capability for several typical molecular systems (such as H<sub>2</sub>O, N<sub>2</sub>, CO, 1,3-butadiene, benzene, and adenine molecules).<sup>50</sup> Consistent with previous ADC results,<sup>62–65</sup> satellite peaks have been observed from the computed spectral functions within the GFCCSD framework in the energy regions where the many-body effect becomes significant and the single particle picture of ionization often breaks down.<sup>66,67</sup> It should be emphasized here that solving a set of linear equations and diagonalizing the secular matrix are purely two different computational procedures with similar scaling (i.e.,  $\mathcal{O}(N^6)$  for CCSD) and a different computational procedure in the same GFCC theoretical context will not add anything new to the theory itself. Therefore, the peak positions obtained from the two computational procedures should be exactly the same within the same truncated subspace. Also, it needs to point out that the CVS approximation mentioned above is in principle applicable to either a diagonalization routine or a linear solver since similar tensor contractions need to be performed in both routines (see  $\bar{H}_N X_p$  in the equations below). Nevertheless, it should be pointed out (also see the discussion in Sec. III) that the disadvantage of solving linear equations in comparison to the diagonalization routine lies in these facts that (i) a much smaller frequency interval is necessary in order to explore some detailed information (in particular, the state of interest with weak intensity) which unavoidably increases the computation cost and (ii) it fails to identify any dark state. Despite these disadvantages, solving linear equations has been chosen by some other studies recently for the computation of the spectral function of, for example, uniform electron gas,<sup>68</sup> light atoms,<sup>69</sup> heavy metal atoms,<sup>70</sup> and simple 1-D periodic systems.<sup>71</sup> A key step

in our procedure is to, in a computationally tractable way, introduce  $\omega$ -dependent IP/EA-EOMCC type operators  $X_p(\omega)$  in an  $(N-1)$ -electron Hilbert space and  $\omega$ -dependent EA-EOMCC type operators  $Y_q(\omega)$  in an  $(N+1)$ -electron Hilbert space,

$$X_p(\omega) = \sum_i x^i(p, \omega) a_i + \sum_{i < j, a} x_a^{ij}(p, \omega) a_a^\dagger a_j a_i + \dots, \quad (14)$$

$$Y_q(\omega) = \sum_i y^a(q, \omega) a_a^\dagger + \sum_{i, a < b} y_{ab}^i(q, \omega) a_a^\dagger a_b^\dagger a_i + \dots, \quad (15)$$

which satisfy

$$(\omega + \bar{H}_N - i\eta) X_p(\omega) |\Phi\rangle = \bar{a}_p |\Phi\rangle, \quad (16)$$

$$(\omega - \bar{H}_N + i\eta) Y_q(\omega) |\Phi\rangle = \bar{a}_q^\dagger |\Phi\rangle, \quad (17)$$

and can be used to write a compact expression for the GFCC,

$$G_{pq}(\omega) = \langle \Phi | (1 + \Lambda) \bar{a}_q^\dagger X_p(\omega) | \Phi \rangle + \langle \Phi | (1 + \Lambda) \bar{a}_p Y_q(\omega) | \Phi \rangle. \quad (18)$$

Typical approximations in the GFCC formulation discussed in early papers<sup>46–50</sup> are defined by limiting the rank of excitation included in the all  $T$ ,  $\Lambda$ ,  $X_p(\omega)$ , and  $Y_q(\omega)$  operators. This general approximation procedure determines the class of correlation effects used to describe the ground state of the  $N$ -electron system but also impacts the quality of pole locations described by  $X_p(\omega)$  and  $Y_q(\omega)$  for  $(N-1)$ - and  $(N+1)$ -electron systems. For example, in the GFCCSD approximation (GFCC with singles and doubles), the expansions for these operators are truncated at the singles and doubles level, i.e.,

$$T \approx \sum_{i, a} t_a^i a_a^\dagger a_i + \sum_{i < j, a < b} t_{ab}^{ij} a_a^\dagger a_b^\dagger a_j a_i, \quad (19)$$

$$\Lambda \approx \sum_{i, a} \lambda_i^a a_i^\dagger a_a + \sum_{i < j, a < b} \lambda_{ij}^{ab} a_i^\dagger a_j^\dagger a_b a_a, \quad (20)$$

$$X_p(\omega) \approx \sum_i x^i(p, \omega) a_i + \sum_{i < j, a} x_a^{ij}(p, \omega) a_a^\dagger a_j a_i, \quad (21)$$

$$Y_q(\omega) \approx \sum_i y^a(q, \omega) a_a^\dagger + \sum_{i, a < b} y_{ab}^i(q, \omega) a_a^\dagger a_b^\dagger a_i. \quad (22)$$

$X_p(\omega)$  and  $Y_q(\omega)$  operators are determined from

$$\begin{aligned} & (\mathcal{Q}_1^{(N-1)} + \mathcal{Q}_2^{(N-1)})(\omega + \bar{H}_N - i\eta)(\mathcal{Q}_1^{(N-1)} + \mathcal{Q}_2^{(N-1)}) X_p(\omega) |\Phi\rangle \\ & = (\mathcal{Q}_1^{(N-1)} + \mathcal{Q}_2^{(N-1)}) \bar{a}_p |\Phi\rangle, \end{aligned} \quad (23)$$

$$\begin{aligned} & (\mathcal{Q}_1^{(N+1)} + \mathcal{Q}_2^{(N+1)})(\omega - \bar{H}_N + i\eta)(\mathcal{Q}_1^{(N+1)} + \mathcal{Q}_2^{(N+1)}) Y_q(\omega) |\Phi\rangle \\ & = (\mathcal{Q}_1^{(N+1)} + \mathcal{Q}_2^{(N+1)}) \bar{a}_q^\dagger |\Phi\rangle, \end{aligned} \quad (24)$$

where subspaces on which Eqs. (23) and (24) are projected on are chosen to reflect excitation included in the CCSD  $X_p(\omega)$  and  $Y_q(\omega)$  operators. This leads to the following for the CCSD Green's function:

$$\begin{aligned} G_{pq}(\omega) & = \langle \Phi | (1 + \Lambda_1 + \Lambda_2) \bar{a}_q^\dagger (X_{p,1}(\omega) + X_{p,2}(\omega)) | \Phi \rangle \\ & + \langle \Phi | (1 + \Lambda_1 + \Lambda_2) \bar{a}_p (Y_{q,1}(\omega) + Y_{q,2}(\omega)) | \Phi \rangle, \end{aligned} \quad (25)$$

which has been used in several studies to determine the efficiency of the GFCCSD approximations.

As discussed in Ref. 47, the GFCC formalism offers flexibility in treating the correlation effect. In particular, different

levels of approximations can be used to describe ground-state effects for the  $N$ -electron systems ( $T$  and  $\Lambda$  operators) and location of poles for  $N-1$  and  $N+1$  systems ( $X_p(\omega)$  and  $Y_q(\omega)$  operators). In this paper, we will establish a simple rule that assures that the basic properties of the GFCC formalism, such as connected character of contributing diagrams, are still preserved. We will refer to these schemes as GFCC- $i(n, m)$ , where  $n$  and  $m$  define the level of excitations used to define approximate operators  $T/\Lambda$  and  $X_p(\omega)/Y_q(\omega)$ , respectively, and 'i' stands for the inner space indicating that two-levels of approximations are used. An interesting feature of the GFCC- $i(n, m)$  is the fact that, when  $m = N$ , the  $X_p(\omega)$  and  $Y_q(\omega)$  operators reproduce exact location of poles in the  $N-1$  and  $N+1$  spaces. This is a straightforward consequence that similarity transformation  $e^{-T} H e^T$  does not change the spectral properties of the electronic Hamiltonian  $H$ .

As a specific example, let us focus on the GFCC- $i(2,3)$  method. In this case, in addition to singly ( $X_{p,1}(\omega)/Y_{q,1}(\omega)$ ) and doubly ( $X_{p,2}(\omega)/Y_{q,2}(\omega)$ ) excited components of  $X_p(\omega)$  and  $Y_q(\omega)$  operators, we add three-body terms ( $X_{p,3}(\omega)/Y_{q,3}(\omega)$ ), which leads to the following definitions of these operators:

$$\begin{aligned} X_p(\omega) & \approx \sum_i x^i(p, \omega) a_i + \sum_{i < j, a} x_a^{ij}(p, \omega) a_a^\dagger a_j a_i \\ & + \sum_{i < j < k, a < b} x_{ab}^{ijk}(p, \omega) a_a^\dagger a_b^\dagger a_k a_j a_i, \end{aligned} \quad (26)$$

$$\begin{aligned} Y_q(\omega) & \approx \sum_i y^a(q, \omega) a_a^\dagger + \sum_{i, a < b} y_{ab}^i(q, \omega) a_a^\dagger a_b^\dagger a_i \\ & + \sum_{i < j, a < b < c} y_{abc}^{ij}(q, \omega) a_a^\dagger a_b^\dagger a_c^\dagger a_j a_i. \end{aligned} \quad (27)$$

At the same time, the  $T$  and  $\Lambda$  operators in the GFCC- $i(2,3)$  are given by Eqs. (19) and (20). The working equations for the  $X_p(\omega)$  and  $Y_q(\omega)$  in the GFCC- $i(2,3)$  method are given by expressions

$$\begin{aligned} & (\mathcal{Q}_1^{(N-1)} + \mathcal{Q}_2^{(N-1)} + \mathcal{Q}_3^{(N-1)})(\omega + \bar{H}_N - i\eta)(X_{p,1}(\omega) + X_{p,2}(\omega) \\ & + X_{p,3}(\omega)) |\Phi\rangle = (\mathcal{Q}_1^{(N-1)} + \mathcal{Q}_2^{(N-1)} + \mathcal{Q}_3^{(N-1)}) \bar{a}_p |\Phi\rangle, \end{aligned} \quad (28)$$

$$\begin{aligned} & (\mathcal{Q}_1^{(N+1)} + \mathcal{Q}_2^{(N+1)} + \mathcal{Q}_3^{(N+1)})(\omega - \bar{H}_N + i\eta)(Y_{q,1}(\omega) + Y_{q,2}(\omega) \\ & + Y_{q,3}(\omega)) |\Phi\rangle = (\mathcal{Q}_1^{(N+1)} + \mathcal{Q}_2^{(N+1)} + \mathcal{Q}_3^{(N+1)}) \bar{a}_q^\dagger |\Phi\rangle. \end{aligned} \quad (29)$$

The key role in assuring the connected character of the CC Green's function matrix elements was played by the fact that the  $X_p(\omega)$  and  $Y_q(\omega)$  operators are expressible in terms of connected diagrams. In Refs. 48 and 49, we demonstrated that this feature is a consequence of the connected character of equations for the  $X_p(\omega)$  and  $Y_q(\omega)$  operators. It is interesting to analyze this property from the point of view of Eqs. (28) and (29). By closer inspection [see Figs. 1(a) and 1(b)], one can notice that the only disconnected diagrams contributing to Eq. (28) correspond to diagrams that involve elements of  $\bar{H}_N$  corresponding to CCSD equations [examples on these diagrams corresponding to projections on triples are also shown in Figs. 1(a) and 1(b); analogous behavior can also be found for the  $Y_q(\omega)$  equations] and numerically disappear once the CCSD ground state converges. Therefore, the equations for  $X_p(\omega)$  and  $Y_q(\omega)$  GFCC- $i(2,3)$  amplitudes [i.e., Eqs. (28)

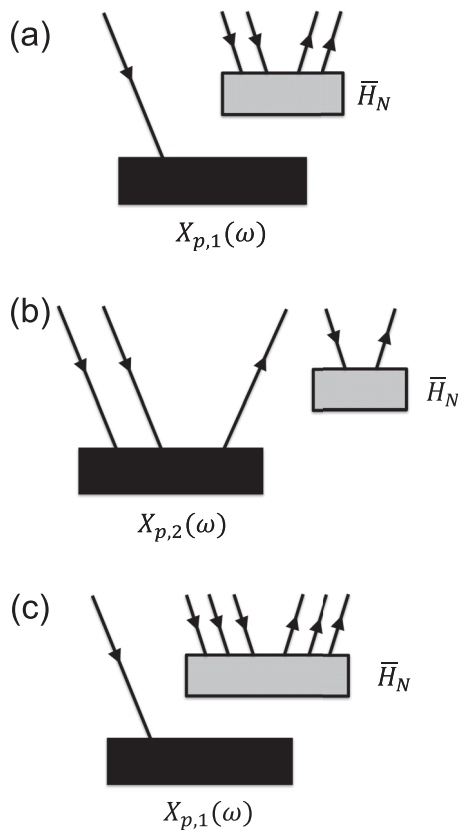


FIG. 1. Disconnected contributions in the equations for triples ( $X_{p,3}(\omega)$ ) in the GFCC-i(2,3) method [panels (a) and (b)] and in the equations for quadruples ( $X_{p,4}(\omega)$ ) in the GFCC-i(2,4) method [panel (c)].

and (29)] involve only connected diagrams. Given the fact that GFCC-i(2,3) matrix elements are given by the same expression, Eq. (25), as in the GFCCSD case [with the difference that in the GFCC-i(2,3) case,  $X_{p,i}(\omega)$  and  $Y_{q,i}(\omega)$  ( $i = 1, 2$ ) are iterated in the presence of triply excited amplitudes  $X_{p,3}(\omega)$  and  $Y_{q,3}(\omega)$ ], their diagrammatic expansion contains connected diagrams only.

However, one should proceed with caution when extending GFCC-i(2,3) to higher  $m$ 's. For example, in the GFCC-i(2,4), a disconnected diagram shown in Fig. 1(c) appears in the equations for quadruples ( $X_{p,4}(\omega)$  equations), which, as explained in Fig. 2, gives rise during the iterations to disconnected components entering  $X_p(\omega)$  operators that correspond to lower order excitations. The provenance of these problems is associated with the non-vanishing character of higher-order moments of the CCSD equations [see Refs. 72 and 73], which are not used to define CCSD equations. Therefore, the only GFCC-i( $n, m$ ) methods that guarantee the connectivity of the corresponding matrix elements of the Green's function are of the GFCC-i( $n, n + 1$ ) type.

Generally speaking, the GFCC-i( $n, m$ ) formalism can be viewed as an extension of previous studies of approximate inclusion of higher-order excitations in the IP/EA-EOMCC methods represented by a broad class of iterative and non-iterative methodologies including EOMIP-CCSD\*,<sup>74–76</sup> CCSDT- $n$  ( $n = 1, 2, 3$ ),<sup>77–81</sup> CC $n$  ( $n = 2, 3$ ),<sup>82,83</sup> EOM-CC( $m, n$ ),<sup>84–86</sup> and EOM-CCSD(T)(a)\* formalisms.<sup>87,88</sup> Moreover, the EOM-CC( $m, n$ ) approach introduced by Hirata *et al.* combines different levels of approximations for the

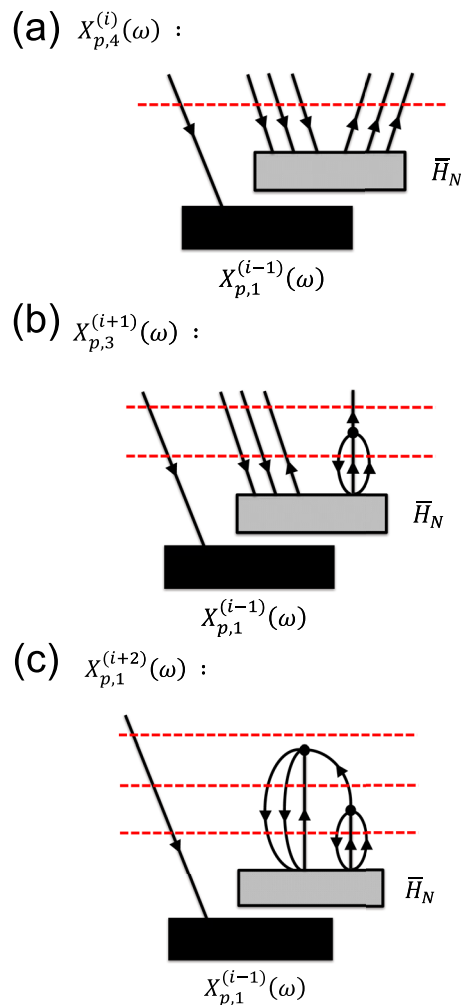


FIG. 2. The mechanism of formation of disconnected contributions in the GFCC-i(2,4) formalism: (a) Disconnected yet linked contribution to the  $X_{p,4}(\omega)$  operator in the  $i$ th iterative cycle. (b) Disconnected yet linked contribution to the  $X_{p,3}(\omega)$  operator in the  $(i + 1)$ -th iteration. Diagram shown in step (b) refers to typical  $Q_3^{(N-1)}V_N X_{p,4}(\omega)|\Phi\rangle$  ( $V_N$  refers to the two-body part of electronic Hamiltonian in normal product form). (c) Disconnected diagram (containing a disconnected closed part) contributing to  $X_{p,1}(\omega)$  in the  $(i + 2)$ -th iteration through the coupling term  $Q_1^{(N-1)}V_N X_{p,3}(\omega)|\Phi\rangle$ . Red dashed lines represent symbolically zeroth order resolvents.

ground-state CC ansatz and equation-of-motion CC operators  $R_K$ , for vertical excitation energies, ionization potential, and electron affinities have already been applied to molecular systems. Several problems of the EOM-CC( $m, n$ ) class of methods stemming from possible inconsistencies between approximation levels of the  $T$  and  $R_K$  operator approaches and resulting in the lack of the size-intensivity of the calculated excitation energies have been discussed by Splipchenko and Krylov.<sup>86</sup> In this paper, for the first time, we demonstrate that similar procedures applied to  $X_p(\omega)$  and  $Y_q(\omega)$  operators (which in contrast to the  $R_K$  operators are determined by solving linear equations) may violate size-extensive character of the corresponding Green's function if the difference between excitation levels used to approximate  $X_p(\omega)/Y_q(\omega)$  operators and  $T$  operators is bigger than one.

In this paper, we will focus on the GFCC-i(2,3) method which introduces triple excitations in describing correlation effects needed for the accurate determination of the poles of

the Green's function. In contrast to the GFCCSDT method, the GFCC-i(2,3) formalism bypasses memory demands for storing  $T_3$  amplitudes (proportional to  $n_o^3 n_v^3$ , where  $n_o$  and  $n_v$  refer to the total number of occupied and virtual orbitals, respectively), and the maximum memory demand in the GFCC-i(2,3) formalism comes from storing  $X_{p,3}(\omega)$  ( $Y_{q,3}(\omega)$ ) amplitudes that is proportional to  $n_o^3 n_v^2$  ( $n_o^2 n_v^3$ ).

To obtain the working equations for the GFCC-i(2,3) method, we have chosen a form of Eq. (28), in which the projections on singles and doubles are the same as in GFCCSDT [see Eqs. (30) and (31), where the full form of  $X_{p,3}(\omega)$ -dependent coupling terms including  $Q_1^{(N-1)} V_N X_{p,3}(\omega) |\Phi\rangle$ ,  $Q_2^{(N-1)} V_N X_{p,3}(\omega) |\Phi\rangle$ , and  $Q_2^{(N-1)} (V_N T_1 X_{p,3}(\omega))_C |\Phi\rangle$  is maintained],

$$\begin{aligned} Q_1^{(N-1)} (\omega + \bar{H}_N - i\eta) (X_{p,1}(\omega) + X_{p,2}(\omega) + X_{p,3}(\omega)) |\Phi\rangle \\ = Q_1^{(N-1)} \bar{a}_p |\Phi\rangle, \end{aligned} \quad (30)$$

$$\begin{aligned} Q_2^{(N-1)} (\omega + \bar{H}_N - i\eta) (X_{p,1}(\omega) + X_{p,2}(\omega) + X_{p,3}(\omega)) |\Phi\rangle \\ = Q_2^{(N-1)} \bar{a}_p |\Phi\rangle. \end{aligned} \quad (31)$$

As to the projections on triples, in contrast to GFCCSDT, we chose to maintain the leading contributing  $X_{p,1}$ ,  $X_{p,2}$ , and  $X_{p,3}$  operators, i.e.,

$$\begin{aligned} Q_3^{(N-1)} \{((\omega + F_N - i\eta) X_{p,3}(\omega))_C + (V_N X_{p,2}(\omega))_C \\ + (V_N T_2^{(1)} X_{p,1}(\omega))_C\} |\Phi\rangle \approx Q_3^{(N-1)} \bar{a}_p |\Phi\rangle, \end{aligned} \quad (32)$$

where the one-particle part  $F_N = \sum_r \epsilon_r N[a_r^\dagger, a_r]$  and the two-particle part  $V_N = \frac{1}{4} \sum_{p,q,r,s} v_{pq}^{rs} N[a_p^\dagger a_q^\dagger a_s a_r]$ , with  $v_{pq}^{rs}$  denoting antisymmetrized two-electron integrals and  $N[\dots]$  designating normal ordered form of a given second-quantized expression.

Remarkably, analogous to CCSDT-n ( $n = 1, 2, 3$ ) and CC3 methods, in Eq. (32), the triply excited component  $X_{p,3}(\omega)$  only contracts with zeroth order terms; i.e., the  $3h2p - 3h2p$  coupling is only represented by the lowest-order contribution stemming from the Fock operator (note that full inclusion of  $\bar{H}_N$  in the  $3h2p - 3h2p$  coupling would significantly increase the numerical overhead from  $\mathcal{O}(N^6)$  to  $\mathcal{O}(N^7)$  per iteration), which allows the on-the-fly determination of  $X_{p,3}(\omega)$  as a function of  $X_{p,1}(\omega)$  and  $X_{p,2}(\omega)$  (hereafter, we call  $X_{p,3}(\omega)$  of this type internal  $X_{p,3}(\omega)$  or internal triple). That is, based on Eq. (32), the components of  $X_{p,3}(\omega)$ ,  $x_{ab,p}^{ijk}(\omega)$  (including both the real and imaginary parts) can be obtained through

$$\begin{aligned} \Re(x_{ab,p}^{ijk}(\omega)) = \frac{1}{\Delta\epsilon_{ab}^{ijk}(\omega)^2 + \eta^2} [\Delta\epsilon_{ab}^{ijk}(\omega) \Re(U_{ab,p}^{ijk}(\omega)) \\ - \eta \Im(U_{ab,p}^{ijk}(\omega))], \end{aligned} \quad (33)$$

$$\begin{aligned} \Im(x_{ab,p}^{ijk}(\omega)) = \frac{1}{\Delta\epsilon_{ab}^{ijk}(\omega)^2 + \eta^2} [\Delta\epsilon_{ab}^{ijk}(\omega) \Im(U_{ab,p}^{ijk}(\omega)) \\ + \eta \Re(U_{ab,p}^{ijk}(\omega))], \end{aligned} \quad (34)$$

where

$$\begin{aligned} U_{ab,p}^{ijk}(\omega) = A_1 \{v_{la}^{ij} x_{b,p}^{lk}(\omega) - t_{jk}^{ca} v_{ic}^{db} x_{d,p}(\omega)\} \\ - A_2 \{v_{c,p}^{ab} x_{c,p}^{jk}(\omega) + t_{lk}^{ab} v_{ij}^{cl} x_{c,p}(\omega)\} \end{aligned} \quad (35)$$

and  $\Delta\epsilon_{ab}^{ijk}(\omega) = \omega + \epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j - \epsilon_k$ , with  $A_1\{\dots\}$  and  $A_2\{\dots\}$  representing the corresponding antisymmetric permutation operators and  $\epsilon_r$  denoting the  $r$ th Hartree-Fock (HF) orbital energy. In the above expressions,  $i, j, k, l, \dots$  denote the occupied Hartree-Fock orbital indices,  $a, b, c, d, \dots$  denote the virtual Hartree-Fock orbital indices, and  $p, q, r, s, \dots$  denote the general Hartree-Fock orbital indices. Note that the computational cost of Eqs. (33) and (34), for a given HF orbital  $p$  and frequency  $\omega$ , only scales as  $n_o^3 n_v^2$ , while the cost of Eq. (35) scales as  $n_o^2 n_v^3$ .

Once  $X_{p,3}(\omega)$  is generated, the major goal is then to solve Eqs. (30) and (31), which can alternatively be represented using the following compact form:

$$(\omega + \bar{\mathbf{H}}_N - i\eta) \mathbf{x}_p(\omega) = \mathbf{b}_p, \quad (36)$$

in which the index  $p$  runs over the entire spin-orbital domain and  $\mathbf{b}_p$  is a real free term. As discussed in our previous work, by representing  $\mathbf{x}_p(\omega)$  as a vector

$$\mathbf{x}_p(\omega) = \begin{bmatrix} \Re(\mathbf{x}_p(\omega)) \\ \Im(\mathbf{x}_p(\omega)) \end{bmatrix}, \quad (37)$$

then the matrix form (36) becomes

$$\begin{bmatrix} \omega + \bar{\mathbf{H}}_N & \eta \\ -\eta & \omega + \bar{\mathbf{H}}_N \end{bmatrix} \begin{bmatrix} \Re(\mathbf{x}_p(\omega)) \\ \Im(\mathbf{x}_p(\omega)) \end{bmatrix} = \begin{bmatrix} \mathbf{b}_p \\ \mathbf{0} \end{bmatrix}. \quad (38)$$

Equation (38) can be approximately solved by iterative procedures (i.e., linear equation solvers). Once the one-body and two-body parts of  $\mathbf{x}_p(\omega)$  are converged, the GFCC matrix element can then be obtained through Eq. (25).

The overall procedure of executing a GFCC-i(2,3) calculation for the retarded part of  $G(\omega)$  can then be summarized as three parts, i.e., ( $\Lambda$ )-CCSD calculations that scale as  $n_o^2 n_v^4$ , the GFCC-i(2,3) iterative part in which each iteration step scales as  $n_o^3 n_v^3$  iteration for a given pair of  $p$  and  $\omega$ , and one-shot contraction [Eq. (25)] that scales as  $n_o^2 n_v^2$  for a given pair of  $p$  and  $\omega$ . It is worth mentioning that, in comparison with the GFCCSDT method, the GFCC-i(2,3) method does not require the cumbersome  $n_o^3 n_v^5$  ( $\Lambda$ )-CCSDT calculations. Furthermore, in spite of having the same numerical scalings during the iteration, the iteration part in the GFCC-i(2,3) method bears a much smaller prefactor than that in the GFCCSDT method. Also, in terms of storage requirement, the GFCC-i(2,3) method eases off the storage of  $X_{p,3}(\omega)$  at the end of the iteration. Besides, in general, due to the inherent independencies in the GFCC formalism (i.e., the calculation of  $X_p(\omega)$  for a given pair of  $p$  and  $\omega$  is independent of the calculation for another pair), our GFCC method is a strong scaling method that allows the efficient utilization of process groups in the calculation to significantly reduce time-to-solution and is suitable for the simulation of larger systems with large basis sets.<sup>46</sup>

As above mentioned, most of the computation cost in the GFCC-i(2,3) calculation is spent on solving the linear equation, Eq. (38), which is closely related to the linear equation solvers used in the calculation. In the present study, we chose to solve Eq. (38) by exploiting the so-called direct inversion in the iterative subspace (DIIS) procedure. Different from the conventional DIIS approach, here we adapt the DIIS to the complex space. Suppose we have  $K$  micro-iteration iterates

$\mathbf{x}_{p,j+1}(\omega), \dots, \mathbf{x}_{p,j+K}(\omega)$ , then the new  $\mathbf{x}_{p,j+K+1}(\omega)$  is obtained through a linear combination of the  $K$  micro-iteration iterates

$$\mathbf{x}_{p,j+K+1}(\omega) = \sum_{k=1}^K c_k \tilde{\mathbf{x}}_{p,j+k+1}(\omega) = \sum_{k=1}^K c_k (\mathbf{x}_{p,j+k}(\omega) + \mathbf{r}_{p,j+k}(\omega)) \text{ with } \sum_{k=1}^K c_k = 1, \quad (39)$$

where  $\mathbf{r}_{p,j+k}(\omega)$  represents the residual vector of  $\mathbf{x}_{p,j+k}(\omega)$ ,  $\tilde{\mathbf{x}}_{p,j+k+1}(\omega)$  is an intermediate update of  $\mathbf{x}_{p,j+k+1}(\omega)$ , and the coefficient vector  $\mathbf{c} = [c_1, \dots, c_k, \dots, c_K]^T$  is defined in the complex space. The goal is to minimize the DIIS least square functional  $J_{\text{DIIS}} := \|\sum_{k=1}^K c_k \mathbf{r}_{p,j+k}\|_2$  by solving the following system of linear equations:

$$\begin{bmatrix} \mathbf{B} & -\mathbf{1} \\ -\mathbf{1}^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{c} \\ \zeta \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ -1 \end{bmatrix}, \quad (40)$$

where  $\mathbf{1} = [1, \dots, 1]$  is a constant vector of length  $K$ ,  $\zeta$  is a complex *Lagrangian* multiplier, and the  $m \times m$  matrix  $\mathbf{B}$  is

constructed through

$$\mathbf{B} = \begin{bmatrix} \mathbf{r}_{p,j+1}^T(\omega) \\ \vdots \\ \mathbf{r}_{p,j+K}^T(\omega) \end{bmatrix} \cdot [\mathbf{r}_{p,j+1}(\omega) \dots \mathbf{r}_{p,j+K}(\omega)]. \quad (41)$$

If we assume that the  $\bar{\mathbf{H}}_N$  matrix is diagonal dominant, the correction vector at the  $(j+k)$ -th step,  $\mathbf{r}_{p,j+k}(\omega)$ , can be approximated through

$$\begin{bmatrix} \Re(\mathbf{r}_{p,j+k}(\omega)) \\ \Im(\mathbf{r}_{p,j+k}(\omega)) \end{bmatrix} \approx \begin{bmatrix} \omega + \mathbf{D} & \eta \\ -\eta & \omega + \mathbf{D} \end{bmatrix}^{-1} \left\{ \begin{bmatrix} \mathbf{b}_p \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \omega + \bar{\mathbf{H}}_N & \eta \\ -\eta & \omega + \bar{\mathbf{H}}_N \end{bmatrix} \begin{bmatrix} \Re(\mathbf{x}_{p,j+k}(\omega)) \\ \Im(\mathbf{x}_{p,j+k}(\omega)) \end{bmatrix} \right\}, \quad (42)$$

with  $\mathbf{D}$  being the orbital-energy dependent diagonal part of the  $\bar{\mathbf{H}}_N$  matrix.

Figure (3) exhibits the convergence performance of our linear equation solver (DIIS adapted to complex space) in the frameworks of both GFCCSD and GFCC-i(2,3) methods for the calculation of  $\mathbf{x}_p$ 's over several frequencies that are close

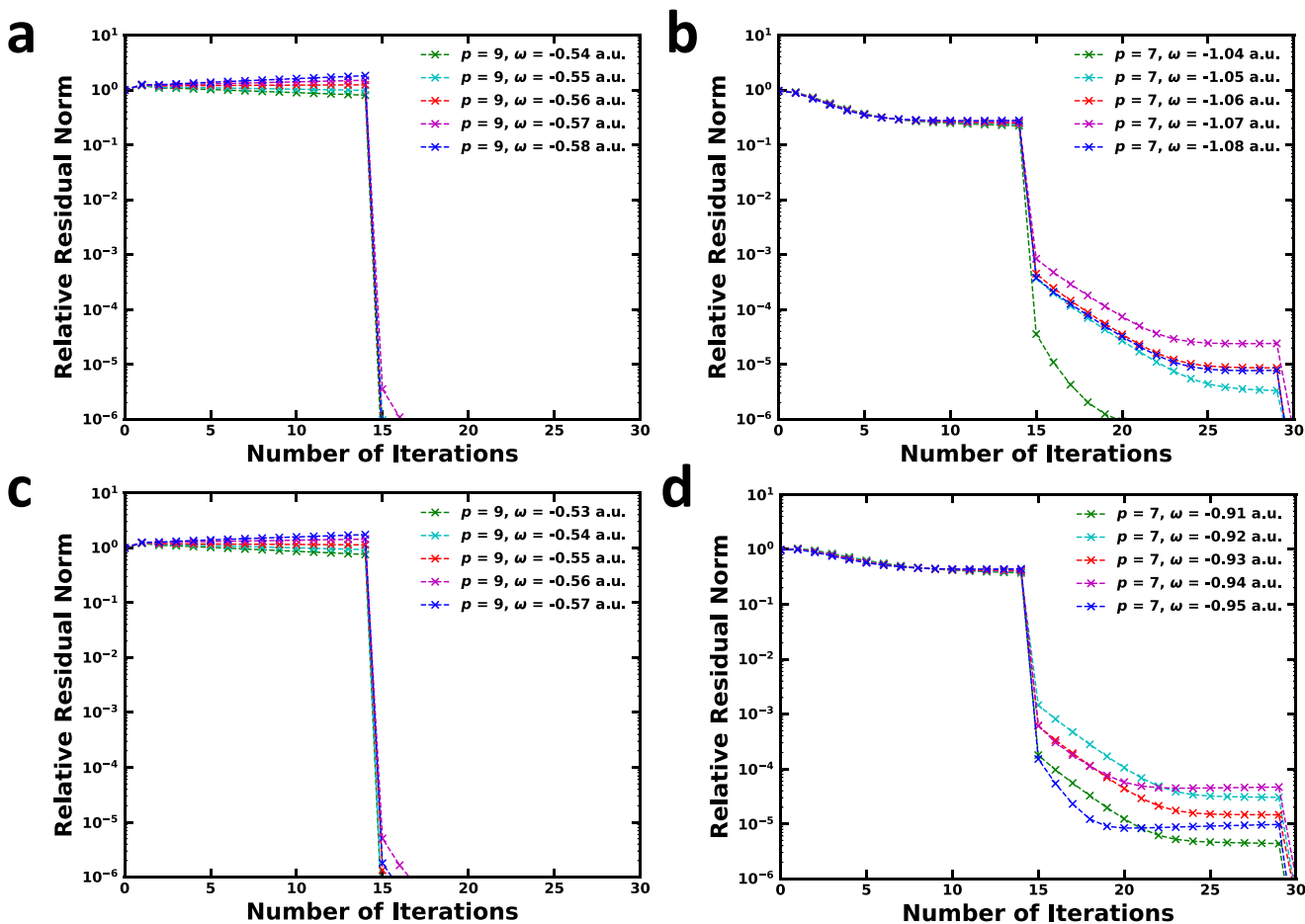


FIG. 3. Convergence performance test of the linear equation solver (DIIS adapted to complex space) used in the calculations of  $\mathbf{x}_p(\omega)$ 's for an  $\text{N}_2$  molecule at chosen orbital indices  $p$ 's and frequency domains. The number of iterates used in the test is 15.  $\Delta\omega = 0.01$  a.u. (0.272 eV) applies to all the chosen frequency domains. (a) and (c) correspond to the GFCCSD and GFCC-i(2,3) calculations of the  $\mathbf{x}_p(\omega)$ 's attributing to the first main peak of the spectral function. (b) and (d) correspond to the GFCCSD and GFCC-i(2,3) calculations of the  $\mathbf{x}_p(\omega)$ 's attributing to the first satellite peak of the spectral function.

to the first main peak and first satellite peak in the spectral function of an  $N_2$  molecule. The position of the first main peak is predicted to be at  $\omega = -0.558$  a.u. ( $-15.185$  eV) by GFCCSD and at  $\omega = -0.550$  a.u. ( $-14.968$  eV) by GFCC-i(2,3), respectively. The position of the first satellite peak is predicted to be at  $\omega = -1.058$  a.u. ( $-28.792$  eV) by GFCCSD and at  $\omega = -0.930$  a.u. ( $-25.309$  eV) by GFCC-i(2,3), respectively. Given the number of micro iterates to be 15, Figs. 3(a) and 3(c) show that the relative residual norm of  $\mathbf{x}_p$  of the frequencies that are close to the main peak can be quickly reduced to be  $<10^{-5}$  after one DIIS update in both the GFCCSD and GFCC-i(2,3) calculations, while the convergence of  $\mathbf{x}_p$  of the frequency domains that include the satellite peak is a bit slower [see Figs. 3(b) and 3(d)], i.e., to reach the same accuracy as for the main peak, at least one more DIIS

update would be needed in either GFCCSD or GFCC-i(2,3) calculation.

This performance difference can be attributed to different peak features that impact the complication of solving the coupled linear equations [Eq. (38)]. Note that different from the main peaks which are dominated by the one-electron process, the satellite peaks have significant contributions from double excitations (two-electron process). Thus, in order to solve for  $\mathbf{x}_p$ , in the frequency domain for main peaks, the iterating update would be basically on the one-body part of  $\mathbf{x}_p$  whose dimension is only  $n_o$ , while in the frequency domain for satellite peaks, the iterating update would be instead on the larger two-body part of  $\mathbf{x}_p$  whose dimension is  $n_o^2 n_v$ . Apparently, more variables would usually require more iterations to help the linear equations converge.

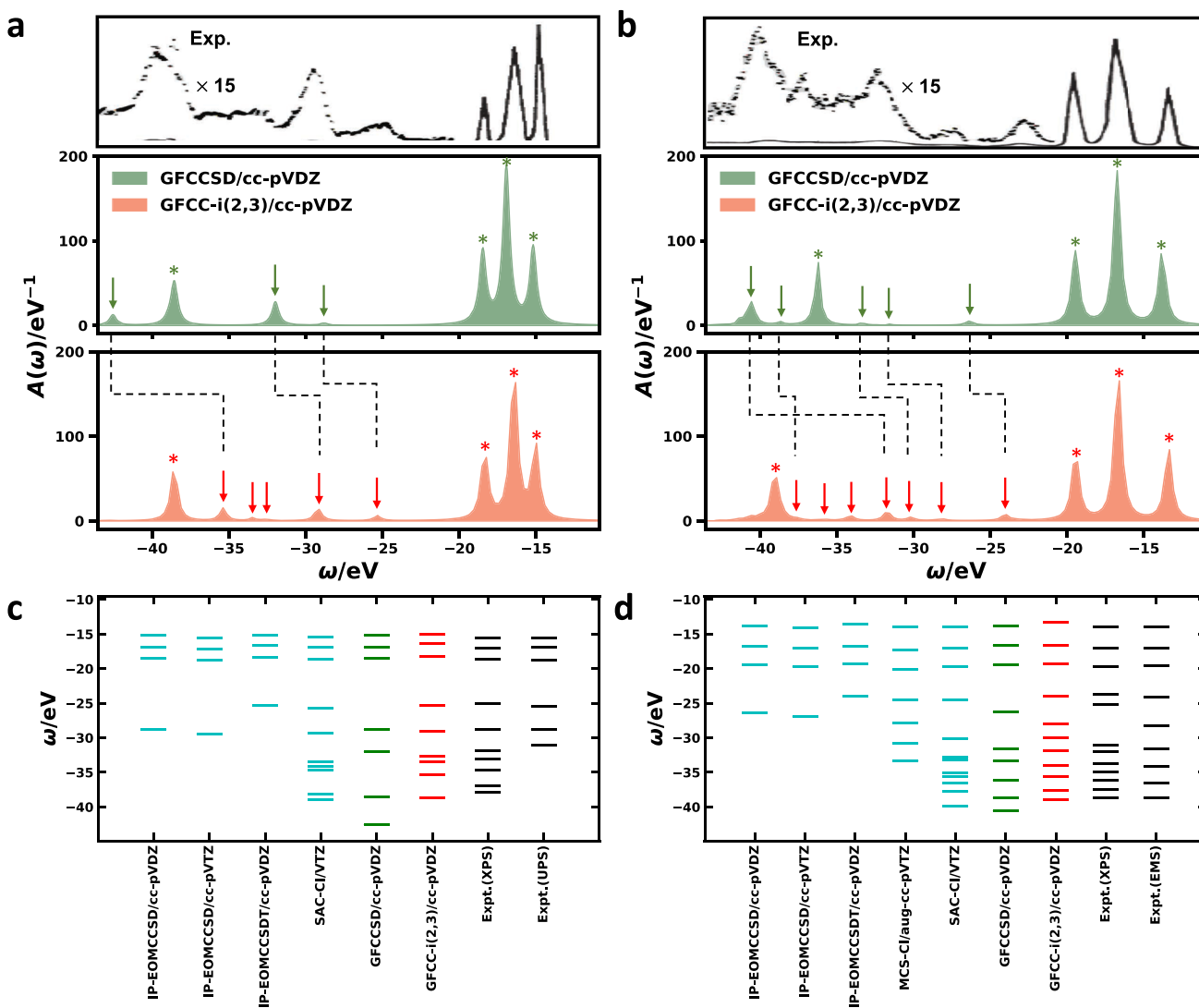


FIG. 4. Spectral functions,  $A(\omega)$ 's, of  $N_2$  (a) and CO (b) molecules in the energy regime between  $-1.6$  a.u. ( $-43.542$  eV) and  $-0.4$  a.u. ( $-10.886$  eV) computed by the GFCCSD and GFCC-i(2,3) methods with  $\eta = 0.01$  a.u. ( $0.272$  eV) and  $\Delta\omega \leq 0.01$  a.u. ( $\leq 0.272$  eV), in comparison with the peak positions extracted from the experimental spectra and other theoretical results [(c) and (d)]. In (a) and (b), main peaks (dominated by the one-electron process) are labeled by asterisks, and satellite peaks are labeled by arrows. In (c) and (d), the IP-EOMCC results for the two molecules are from Ref. 85, the symmetry-adapted-cluster configuration interaction (SAC-CI) results are collected from Ref. 91, the multichannel Schwinger configuration interaction (MCS-CI) results for the CO molecule are from Ref. 92, and the experimental spectra and values are collected from Refs. 93–95, in which the incident photoelectron energies ( $h\nu$ ) were reported to be 50.3 eV and 55.1 eV, respectively. Note that the relative intensities between the different final ionic state manifolds exhibited in the photoelectron spectrum are supposed to be different from those observed from the calculated spectral functions (see discussion in the text).

### III. RESULTS AND DISCUSSION

Our preliminary accuracy test of the GFCC-i(2,3) method is focused on the computed spectral functions of the N<sub>2</sub> and CO molecules and their comparisons with the GFCCSD spectral function as well as other theoretical results and experimental observations. The experimental geometries<sup>89</sup> of the N<sub>2</sub> and CO molecules were used in all the calculations. The conventional (Λ-)CCSD calculations of the two molecules were performed first by using the NWChem suite of quantum chemical codes.<sup>90</sup> The converged *T* and *Λ* amplitudes, as well as the two-electron integral tensors, were then used by our pilot code to compute the IP-EOMCCSD type vector  $\mathbf{x}_p(\omega)$  and the  $\omega$ -dependent retarded GFCC matrix elements. The spectral function is then given by the imaginary part of the retarded GFCC matrix,

$$A(\omega) = -\frac{1}{\pi} \text{Tr}[\Im(\mathbf{G}^R(\omega))] = -\frac{1}{\pi} \sum_p \Im(G_{pp}^R(\omega)). \quad (43)$$

The computed spectral functions of the N<sub>2</sub> and CO molecules are shown in Fig. 4, and detailed information of the peaks is given in Tables I and II. The Hartree-Fock electronic configurations of the ground states of the N<sub>2</sub> and CO molecules can be described as  $(1\sigma_g)^2 (1\sigma_u)^2 (2\sigma_g)^2 (2\sigma_u)^2 (3\sigma_g)^2 (1\pi_u)^4$  and  $(1\sigma)^2 (2\sigma)^2 (3\sigma)^2 (4\sigma)^2 (1\pi)^4 (5\sigma)^2$ , respectively.

For both N<sub>2</sub> and CO molecules, it has been well known that strong correlation effects exist in their inner-valence ionization process and lead to the breakdown of the MO picture. A typical example is the splitting of the single  $2\sigma_g$  line of N<sub>2</sub> (similarly a  $3\sigma$  line of CO) into several satellite lines (see Ref. 96 for an early study). In the present study, as shown in Figs. 4(a) and 4(b), the GFCCSD/cc-pVDZ and GFCC-i(2,3)/cc-pVDZ calculations almost give the same main peak positions (labeled by asterisks) above  $-0.735$  a.u. ( $\sim -20$  eV). Below this energy regime, several satellite peaks started to emerge. Take N<sub>2</sub>, for example, and the GFCCSD/cc-pVDZ calculation exhibits three satellite peaks between  $-43$  eV and  $-20$  eV which

are dominated by the two-hole-one-particle process, but still show significant  $2\sigma^{-1}$  configuration contributions. However, in comparison to previous experimental observations,<sup>93–95</sup> the positions of these satellite peaks are not well reproduced by the GFCCSD calculations. For the first satellite peak in the calculated spectral function of N<sub>2</sub> (similar to the CO molecule case), the GFCCSD calculation predicted its position at  $\sim -28.79$  eV, which is about 3 eV below the experimental value.<sup>93,95</sup> For all the following satellite peaks, a systematic overestimation (blue shift) of the satellite peak positions can be observed from the GFCCSD results. On the other hand, the GFCC-i(2,3) calculation significantly red-shifts the satellite peaks to be more close to the experimental observations. The largest red-shift for the N<sub>2</sub> molecule (0.264 a.u. or 7.184 eV) occurs to the  $K^2\Sigma_g^+$  state that is dominated by a  $1\pi_u^{-1}2\sigma_u^{-1}1\pi_g$  two-electron process (see Table I). For the CO molecule, the largest red-shift is 0.320 a.u. (8.708 eV) and is associated with the  $I^2\Sigma^+$  state that is dominated by a  $5\sigma^{-1}1\pi^{-1}1\sigma^*$  two-electron process (see Table II). Besides, in comparison with the GFCCSD results, new satellite peaks are resolved from the GFCC-i(2,3) calculation. As shown in the spectral functions computed by the GFCC-i(2,3)/cc-pVDZ approach in Figs. 4(a) and 4(b), despite exhibiting relatively weak intensity, two new satellite peaks at  $-1.200$  a.u. and  $-1.230$  a.u. (that is, at  $-32.657$  eV and  $-33.473$  eV, assigned to the  $G^2\Pi_u$  and  $I^2\Sigma_u^+$  states, respectively; see Table I) can be easily observed for the N<sub>2</sub> molecule, and two new satellite peaks at  $-1.250$  a.u. and  $-1.310$  a.u. (that is, at  $-34.017$  eV and  $-35.650$  eV, assigned to the  $J^2\Sigma^+$  and  $K^2\Sigma^+$  states, respectively, see Table II) are observed for the CO molecule.

The significant red-shifts and emergence of these inner satellite peaks then distinguish the GFCCSD result and GFCC-i(2,3) result in terms of the agreement with the experimental results. As can be seen from the comparison of the peak positions obtained from the spectral function computed by the GFCC methods with those peak positions read from the experimental photoelectron spectra [Figs. 4(c) and 4(d)], the

TABLE I. Significant peaks in the spectral functions,  $A(\omega)$ 's, of the N<sub>2</sub> molecule calculated in valence (including outer and inner valence) regimes by the GFCC-i(2,3) method with cc-pVDZ basis. For each peak, its relative intensity, significant Green's function matrix element contributors, and main configurations, as well as the weight of all the  $|2h, p\rangle$  configurations (as a net effect of all contributing Slater determinants) are given.

State (Symmetry)	Energy (eV)	Significant $G_{pp}^R$ (Contribution $\geq 0.05$ )	Main configurations ( $ C ^2 \geq 0.02$ )	Weight of $ 2h, p\rangle$ 's
$X(^2\Sigma_g^+)$	-14.968	$G_{55}^R(0.93)$	$3\sigma_g^{-1}(0.91)$	0.02
$A(^2\Pi_u)$	-16.328	$G_{66}^R(0.48), G_{77}^R(0.48)$	$1\pi_u^{-1}(0.94)$	0.01
$B(^2\Sigma_u^+)$	-18.233	$G_{44}^R(0.94)$	$2\sigma_u^{-1}(0.90), 1\pi_u^{-1}3\sigma_g^{-1}1\pi_g(0.03)$	0.04
$C(^2\Sigma_u^+)$	-25.309	$G_{44}^R(0.95)$	$2\sigma_u^{-1}(0.20), 1\pi_u^{-1}3\sigma_g^{-1}1\pi_g(0.70),$ $1\pi_u^{-1}3\sigma_g^{-1}2\pi_g(0.06)$	0.75
$F(^2\Sigma_g^+)$	-29.119	$G_{33}^R(0.93), G_{55}^R(0.05)$	$2\sigma_g^{-1}(0.32), 3\sigma_g^{-1}(0.04), 1\pi_u^{-1}2\sigma_u^{-1}1\pi_g(0.46),$ $2\sigma_g^{-1}3\sigma_u^{-1}1\pi_g(0.08), 1\pi_u^{-1}2\sigma_u^{-1}2\pi_g(0.02)$	0.63
$G(^2\Sigma_g^+, ^2\Sigma_u^+)$	-32.657	$G_{33}^R(0.16), G_{44}^R(0.22),$ $G_{66}^R(0.30), G_{77}^R(0.30)$	$2\sigma_g^{-1}(0.14), 2\sigma_u^{-1}(0.03), 2\sigma_u^{-1}3\sigma_g^{-1}1\pi_g(0.55),$ $1\pi_u^{-1}3\sigma_g^{-1}1\pi_g(0.17), 1\pi_u^{-1}2\sigma_u^{-1}1\pi_g(0.02),$ $2\sigma_u^{-1}3\sigma_g^{-1}2\pi_g(0.02)$	0.79
$I(^2\Sigma_g^+, ^2\Sigma_u^+)$	-33.473	$G_{33}^R(0.11), G_{44}^R(0.84)$	$2\sigma_g^{-1}(0.08), 2\sigma_u^{-1}(0.08), 1\pi_u^{-1}3\sigma_g^{-1}1\pi_g(0.68),$ $1\pi_u^{-1}2\sigma_u^{-1}1\pi_g(0.02), 3\sigma_g^{-2}4\sigma_u(0.03)$	0.78
$K(^2\Sigma_g^+)$	-35.378	$G_{33}^R(0.98)$	$2\sigma_g^{-1}(0.38), 1\pi_u^{-1}2\sigma_u^{-1}1\pi_g(0.54), 2\sigma_u^{-1}3\sigma_g^{-1}4\sigma_u(0.02),$	0.60
$(^2\Sigma_g^+)$	-38.643	$G_{33}^R(0.99)$	$2\sigma_g^{-1}(0.85), 1\pi_u^{-1}2\sigma_u^{-1}1\pi_g(0.08)$	0.13
$(^2\Sigma_g^+, ^2\Sigma_u^+)$	-413.07	$G_{11}^R(0.46), G_{22}^R(0.54)$	$1\sigma_g^{-1}(0.44), 1\sigma_u^{-1}(0.51)$	0.05

TABLE II. Significant peaks in the spectral functions,  $A(\omega)$ 's, of the CO molecule calculated in valence (including outer and inner valence) energy regimes by the GFCC-i(2,3) methods with cc-pVDZ basis. For each peak, its relative intensity, significant Green's function matrix element contributors, and main configurations, as well as the weight of all the  $|2h, p\rangle$  configurations (as a net effect of all contributing Slater determinants) are given.

State (Symmetry)	Energy/eV	Significant $G_{pp}^R$ (Contribution $\geq 0.05$ )	Main configurations ( $ C ^2 \geq 0.02$ )	Weight of $ 2h, p\rangle$ 's
$X(^2\Sigma^+)$	-13.335	$G_{77}^R(0.98)$	$5\sigma^{-1}(0.96)$	0.02
$A(^2\Pi)$	-16.600	$G_{55}^R(0.50), G_{66}^R(0.50)$	$1\pi^{-1}(0.96)$	0.02
$B(^2\Sigma^+)$	-19.322	$G_{44}^R(0.97)$	$4\sigma^{-1}(0.91)$	0.05
$C(^2\Sigma^+)$	-23.948	$G_{33}^R(0.09), G_{44}^R(0.65),$ $G_{77}^R(0.21)$	$4\sigma^{-1}(0.17), 5\sigma^{-1}(0.05), 5\sigma^{-1}1\pi^{-1}1\pi^*(0.62),$ $5\sigma^{-1}1\pi^{-1}2\pi^*(0.04)$	0.70
$E, G(^2\Sigma^+, ^2\Pi)$	-28.030	$G_{33}^R(0.05), G_{44}^R(0.06),$ $G_{55}^R(0.43), G_{66}^R(0.43)$	$3\sigma^{-1}(0.06), 4\sigma^{-1}(0.04), 1\pi^{-1}(0.02),$ $5\sigma^{-1}(0.16), 4\sigma^{-1}5\sigma^{-1}1\pi^*(0.12),$ $1\pi^{-2}1\pi^*(0.20), 5\sigma^{-2}1\pi^*(0.04),$ $4\sigma^{-1}1\pi^{-1}1\pi^*(0.05), 5\sigma^{-1}1\pi^{-1}1\pi^*(0.17),$	0.67
$F(^2\Sigma^+)$	-29.935	$G_{33}^R(0.28), G_{44}^R(0.34),$ $G_{77}^R(0.33)$	$3\sigma^{-1}(0.02), 4\sigma^{-1}(0.03), 5\sigma^{-1}(0.08),$ $5\sigma^{-1}1\pi^{-1}1\pi^*(0.59), 4\sigma^{-1}1\pi^{-1}1\pi^*(0.10),$ $5\sigma^{-2}1\sigma^*(0.02)$	0.80
$I(^2\Sigma^+)$	-31.840	$G_{33}^R(0.84), G_{44}^R(0.14)$	$3\sigma^{-1}(0.29), 4\sigma^{-1}(0.03), 4\sigma^{-1}1\pi^{-1}1\pi^*(0.46),$ $5\sigma^{-1}1\pi^{-1}1\pi^*(0.04)$	0.64
$J(^2\Sigma^+)$	-34.017	$G_{33}^R(0.80), G_{44}^R(0.11),$ $G_{77}^R(0.08)$	$3\sigma^{-1}(0.17), 4\sigma^{-1}1\pi^{-1}1\pi^*(0.65),$ $5\sigma^{-1}1\pi^{-1}1\pi^*(0.03), 4\sigma^{-1}5\sigma^{-1}6\sigma(0.02)$	0.79
$K(^2\Sigma^+)$	-35.650	$G_{33}^R(0.33), G_{55}^R(0.30),$ $G_{66}^R(0.30)$	$3\sigma^{-1}(0.27), 4\sigma^{-1}1\pi^{-1}1\pi^*(0.03),$ $5\sigma^{-1}1\pi^{-1}1\pi^*(0.54), 1\pi^{-2}1\pi^*(0.02)$	0.64
$(^2\Sigma^+)$	-37.555	$G_{33}^R(0.83), G_{33}^R(0.09)$	$3\sigma^{-1}(0.31), 4\sigma^{-1}5\sigma^{-1}1\sigma^*(0.50), 5\sigma^{-2}1\sigma^*(0.03)$	0.61
$(^2\Sigma^+)$	-38.916	$G_{33}^R(0.99)$	$3\sigma^{-1}(0.85), 4\sigma^{-1}1\pi^{-1}1\pi^*(0.04)$	0.14
$(^2\Sigma^+)$	-298.26	$G_{22}^R(0.99)$	$2\sigma^{-1}(0.92)$	0.08
$(^2\Sigma^+)$	-546.73	$G_{11}^R(0.99)$	$1\sigma^{-1}(0.97)$	0.03

relative peak positions obtained from the GFCC-i(2,3)/cc-pVDZ calculation show better agreement with the experimental results.<sup>95,97</sup> For example, for the  $N_2$   $C^2\Sigma_u^+$  and CO  $C^2\Sigma^+$  ionic states, the GFCC-i(2,3)/cc-pVDZ results are  $\sim 0.13$  a.u. ( $\sim 3.538$  eV) and  $\sim 0.09$  a.u. ( $\sim 2.449$  eV) more close to the experiment results, respectively, than the GFCCSD/cc-pVDZ results.

It is worth emphasizing that, in Fig. 4, the comparisons of the computed spectral functions and the experimental photoelectron spectra are made in terms of peak positions, while the relative intensities between the different final ionic state manifolds exhibited in the photoelectron spectra are supposed to be different from those observed from the calculated spectral functions. This is because the intensity of the  $n$ th ionic state in the photoelectron spectrum is given by<sup>98,99</sup>

$$P_n = \sum_{ij} \theta_{i,n} \theta_{j,n}^* \sum_k s_{k,i} s_{k,j}^* \delta(\epsilon_k + I_n - h\nu), \quad (44)$$

where  $\epsilon_k$  is the kinetic energy of the emitted electron,  $I_n$  is the  $n$ th ionization potential,  $h\nu$  is the energy of a photon beam, and  $\delta(\dots)$  is the Dirac delta function.  $\theta_{i,n} = \langle \Phi_n^{N-1} | a_i | \Phi_0^N \rangle$  determines the relative intensity of the  $n$ th ionic state in the spectral function for removing an electron from orbital  $\phi_i$  [ $|\Phi_0^N\rangle$  is the initial state of the  $N$ -electron system and  $|\Phi_n^{N-1}\rangle$  is the  $n$ th state of the  $(N-1)$ -electron system], and  $s_{k,i} = \langle \phi_k | \hat{r} | \phi_i \rangle$  is the dipole matrix element for photoionization ( $\hat{r}$  is the position operator, and  $\phi_k$  and  $\phi_i$  are the continuum one-particle state and bound one-particle state, respectively). In the single particle picture (that is, there is only one  $\theta_{i,n}$  contributing

significantly), the above equation can be simplified to

$$P_n = |\theta_{i,n}|^2 \sum_k |s_{k,i}|^2 \delta(\epsilon_k + I_n - h\nu), \quad (45)$$

where  $|\theta_{i,n}|^2$  is the so-called pole strength whose convoluted sums compose the spectral function. Therefore,  $P_n$  is approximately proportional to  $|\theta_{i,n}|^2$  only when  $h\nu$  is much larger than  $I_n$  (such that the dependence of  $|s_{k,i}|^2$  on  $\epsilon_k$  is weak) and there are no appreciable initial state correlation effects (i.e., in the single particle picture). Also, as can be seen from Fig. 4, the computed spectral function includes the effects of orbital degeneracy. Here, due to the degeneracy of  $1\pi$  orbitals, the corresponding intensity at  $-16.300$  eV for  $N_2$  ( $-16.600$  eV for CO) is almost twice as large as that corresponding to the ionization of the adjacent  $\sigma$  orbitals.

In Figs. 4(c) and 4(d), we also compare the positions of the peaks in the spectral function computed by our GFCC methods with other theoretical ionization potentials. As can be seen, early IP-EOM-CC calculations<sup>85</sup> only reported the computed ionization potentials up to  $\sim -1.0$  a.u. ( $\sim -27.214$  eV). Using the same cc-pVDZ basis set, the peak positions in this regime computed by the GFCCSD are almost the same as the ionization potentials computed by the IP-EOMCCSD method. By including the internal triple ( $X_{p,3}(\omega)$ ) in the calculation, the peak positions from the GFCC-i(2,3) calculations are very close to those ionization potentials computed by the IP-EOMCCSDT method. Beyond this regime, the GFCC-i(2,3) results, in comparison with the GFCCSD results, show better agreement with the SAC-CI results in terms of peak positions and associated configurations. The exploitation of a larger basis set in the

GFCC calculations would be expected to follow the trends observed in the IP-EOM-CC studies<sup>85</sup> that larger basis sets in the GFCC method would make the calculated results in the inner valence regime gradually deviate from the experimental values.

To have a more detailed examination, we found that, in the previous theoretical IP-EOM-CC and SAC-CI studies, a satellite peak classified as the  $D^2\Pi$  state was also mentioned for the  $N_2$  and CO molecules to be located slightly above the  $C^2\Sigma^+$  state. However, the intensity of the  $D^2\Pi$  state was reported to be extremely low. According to the previous SAC-CI study,<sup>91</sup> the ratio of the intensity of the  $D^2\Pi$  state to its adjacent satellite  $C^2\Sigma^+$  state is  $\sim 1/520$  for the  $N_2$  molecule and  $\sim 1/13$  for the CO molecule. Given the already low intensity of the  $C^2\Sigma^+$  state, the direct observation of the  $D^2\Pi$  state from the computed spectral function would be hard [in Figs. 4(a) and 4(b), the  $D^2\Pi$  state cannot be clearly observed from the computed spectral function with  $\Delta\omega = 0.01$  a.u. or 0.272 eV]. In fact, this state was originally even missing in the XPS spectra and can only be identified in the UPS (He II) and de-excitation (DES) spectra,<sup>95</sup> which favor different final states in comparison with the photoelectron spectra.

In order to resolve this state, we switch to a smaller frequency interval,  $\Delta\omega = 0.002$  a.u. (0.054 eV), in our GFCC calculations followed by a decomposition analysis. Figure 5 shows the computed spectral function of the CO molecule using the GFCC-i(2,3)/cc-pVDZ method in the energy regime of  $[-0.92, -0.84]$  a.u. ( $[-25.037, -22.860]$  eV) as well as the imaginary part of the individual diagonal Green's function matrix elements that significantly contributes to the computed spectral function in this regime. From Fig. 5(a), slightly above the peak (at  $-0.88$  a.u. or  $-23.948$  eV) classified as the  $C^2\Sigma^+$

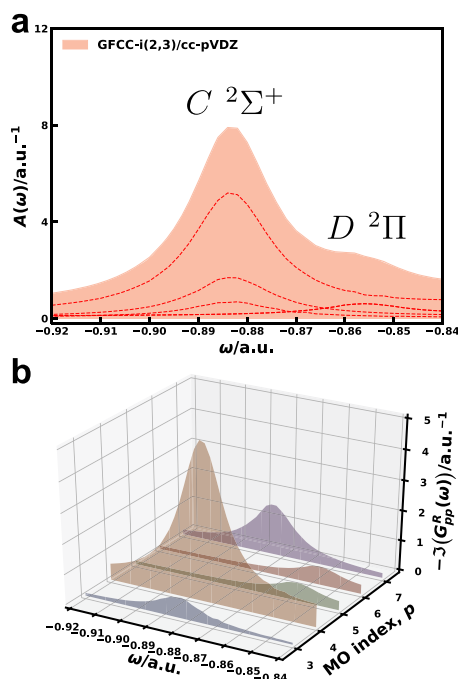


FIG. 5. (a) Spectral functions,  $A(\omega)$ 's, and (b) its major contributions (i.e., the imaginary part of the diagonal Green's function matrix elements,  $-\Im(G_{pp}^R(\omega))$ ) of the CO molecule in the energy regime between  $-0.92$  a.u. and  $-0.84$  a.u. (that is, between  $-25.037$  eV and  $-22.860$  eV) computed by the GFCC-i(2,3) method with  $\eta = 0.01$  a.u. (0.272 eV) and  $\Delta\omega = 0.002$  a.u. (0.054 eV).

state, a small shoulder bump can be seen being centered at  $\sim -0.856$  a.u. ( $-23.295$  eV). From the decomposition analysis in Fig. 5(b), we can further identify the bump as a satellite peak rather than a numerical fluctuation, and the major contributions to this shoulder satellite peak are from  $-\Im(G_{55}^R)$  and  $-\Im(G_{66}^R)$ , which are associated with  $N2p$  orbitals. This then confirms the classification of this satellite peak as the  $D^2\Pi$  state. The energy difference between the  $C^2\Sigma^+$  state and the  $D^2\Pi$  state in the GFCC-i(2,3), i.e., 0.024 a.u. ( $\sim 0.653$  eV), is slightly less than the experimental observation ( $\sim 1.0$  eV) and SAC-CI results (0.88 eV) and more close to the IP-EOM-CCSDT/cc-pVDZ result (0.77 eV). The main configurations of this satellite peak include  $5\sigma^{-2}1\pi^*$  ( $\sim 64\%$ ),  $4\sigma^{-1}5\sigma^{-1}1\pi^*$  ( $\sim 21\%$ ), and  $5\sigma^{-2}2\pi^*$  ( $\sim 6\%$ ), which are consistent with MCS-CI and SAC-CI results.<sup>91,92</sup>

Compared to the early Green's function/CI studies of the inner-valence ionization of  $N_2$  and CO molecules,<sup>91,92,96,100,101</sup> the GFCC-i(2,3) calculation exhibits better agreement than the GFCCSD calculation in terms of the overall spectral function profile (that is, the consistent prediction of the significant satellite peaks). Note that the previous Green's function study<sup>96</sup> is able to provide more detailed information of the inner valence ionization by providing the positions of weaker satellite peaks which might not be observable in our present GFCC calculations. For example, in the same energy regime as used in Fig. 4, there were more than 10 ionization potentials (with corresponding Green's function residual  $> 0.01$ ) reported in Ref. 96, while only six significant peaks were observed from the GFCC-i(2,3) spectral function in the present study. As discussed in Sec. II, this is because the previous Green's function methods (such as  $2h - p$  TDA and ADC methods) are based on the diagonalization scheme which is in principle able to provide all the  $1h$  and  $2h - p$  poles for the given theoretical level within the energy window of interest. The diagonalization scheme will be especially useful to resolve poles that have different characters but are lying very close to each other. In the present study, solving the linear equations and scanning over the energy points in the energy window of interest are on the other hand targeting the overall profile, and to explore more detailed information for a specific energy regime, a smaller energy interval and decomposition analysis of the peaks need to be employed, as evidenced from our above effort to identify the weak  $D^2\Pi$  state in the CO spectral function. However, even with the aid of a smaller energy interval and the decomposition analysis, it is still impossible to find dark states (corresponding to zero pole strength) and may still be difficult to find (or identify) the states with extremely low intensity. The missing of the states with low intensity would have direct impact on the accurate calculation of some properties (like the vibronic coupling constant) in which the corresponding energies are indispensable. In such cases, it would be more reliable to combine the GFCC results with the results from recently proposed iterative subspace methods<sup>58-60</sup> for the narrow energy window of interest to resolve all the possible states.

Since our GFCC approach is able to target any energy regime, we also give the core ionization potentials in Tables I and II. Compared to our previous GFCCSD/cc-pVDZ results,<sup>50</sup> the inclusion of internal triple in the GFCC framework does not significantly change the main ionic

TABLE III. Vertical ionization potentials of  $N_2$  and CO molecules obtained from GFCC-i(2,3) and various EOM-CC (CCLR) based methods. The Aug-cc-pVTZ basis is used in all the calculations. The CCSDT-n and CC3 results are collected from Ref. 80.

State	GFCC-i(2,3)	CCSDT-n				CC3	Exp.
		n = 1	n = 1b	n = 2	n = 3		
$N_2$							
$2\sigma_g^{-1}$	15.67	15.85	15.84	15.89	15.58	15.54	15.60
$1\pi_u^{-1}$	17.28	17.35	17.34	17.40	16.95	16.88	16.98
$2\sigma_u^{-1}$	18.91	19.02	19.02	19.09	18.87	18.82	18.78
CO							
$5\sigma^{-1}$	14.17	14.17	14.15	14.25	13.94	13.86	14.01
$1\pi^{-1}$	17.09	17.33	17.35	17.36	17.10	17.09	16.91
$4\sigma^{-1}$	19.81	19.88	19.88	19.91	19.77	19.76	19.72

peak positions, and both the GFCCSD and GFCC-i(2,3) results show  $\sim 3$  eV deviation from the experimental observation<sup>102,103</sup> and early fourth-order ADC study.<sup>96,104</sup> However, similar to what we observed in Ref. 50, employing a larger basis set is able to reduce this deviation. For example, switching the cc-pVDZ basis set to Aug-cc-pVTZ basis, the position of the  $N$   $1\sigma^{-1}$  peak will be red-shifted to  $-15.09$  a.u. ( $\sim -410.62$  eV), which is much closer to the experimental value (the deviation is  $< 1$  eV). Note that the satellite peaks in the core regime are also an interesting topic, and there were efforts focusing on computing the satellite peaks accompanying the K-shell ionization of  $N_2$  and CO molecules employing the CVS-ADC(4) method.<sup>96,104</sup> However, the computation of the satellite peaks in the core regime is not covered in the present study, but will be one of the topics in our future work.

Finally, we would like to compare our GFCC-i(2,3) approach with other approximate EOM-CC (CCLR)-based methods where the treatment of triples is involved in a different manner. Table III lists the first three vertical ionization potentials of the  $N_2$  and CO molecules computed by GFCC-i(2,3), CCSDT-n ( $n = 1, 2, 3$ ), and CC3 methods. Generally speaking, the CCSDT-3 and CC3 are able to provide the more accurate numbers than the CCSDT-n ( $n = 1, 2$ ) and GFCC-i(2,3) methods, due to their more elaborate treatment of the triple equation, and the deviations from the experimental values for both methods are  $< 0.1$  eV. The quality of the GFCC-i(2,3) numbers lies between the CCSDT-3/CC3 and CCSDT-n ( $n = 1, 2$ ), and the largest deviation in these GFCC-i(2,3) numbers is  $\sim 0.3$  eV (for the  $N_2$   $1\pi_u^{-1}$  state). Note that both CCSDT-n ( $n = 1, 2$ ) and GFCC-i(2,3) exclude the orbital relaxation associated with the  $T_1$  cluster amplitude. In comparison to the CCSDT-n ( $n = 1, 2$ ) which totally ignores the triple-single terms (that is, the triple-single block of the associate Jacobian matrix is set to zero), the  $(V_N T_2 X_{p,1})_C$  term in the GFCC-i(2,3) approach contributes at least to the second order correction to the triple equation. Since the GFCC-i(2,3) approach does not require  $T_3$  cluster amplitude, its computational cost is even cheaper than the CCSDT-1 method.

#### IV. CONCLUSION

In this paper, we propose a new category of the GFCC approximations, i.e., the GFCC-i( $n, m$ ) method, to improve

the pole locations of the CC Green's function. In the GFCC-i( $n, m$ ) method, the level of excitation ( $m$ ) used to describe the ionization potential and electron affinity effects in the  $N - 1$  and  $N + 1$  particle spaces is higher than the level of excitation ( $n$ ) used to correlate the ground-state coupled cluster wave function for the  $N$ -electron system. We demonstrate that in order to maintain size-extensivity of the GFCC matrix elements,  $m$  cannot be larger than  $n + 1$ . As our first practice in this direction, we implement the GFCC-i(2,3) method, in which the cluster operators  $T$  and the de-excitation operator  $\Lambda$  are kept at the CCSD level, while the auxiliary operators  $X_p(\omega)$  and  $Y_q(\omega)$  include the inner three-body terms (or internal triples) that can be obtained from one-body and two-body  $X_p(\omega)$  and  $Y_q(\omega)$  terms in an on-the-fly manner. The computational cost and storage requirement of the GFCC-i(2,3) method are significantly lower than the corresponding GFCCSDT method. Our preliminary accuracy test has been focused on the accurate computing of the spectral functions of the  $N_2$  and CO molecules in the inner and outer valence regime. In comparison with the GFCCSD results, the computed spectral functions by the GFCC-i(2,3) method are greatly improved and exhibit better agreement with the experimental results and other theoretical results. The improvement is particularly significant in terms of providing higher resolution of satellite peaks and more accurate relative positions of these satellite peaks with respect to the main peak positions. A similar strategy could also be applied to other CC methods. For example, it is feasible to combine the CCSD treatment of the ground state with excitation calculation that employs an extended biorthogonal CC expansion manifold comprising single ( $1h$ ), double ( $2h - 1p$ ), and zeroth order triple ( $3h - 2p$ ) excitations. Such a treatment will be able to treat the one-hole ( $1h$ ) main states through third order and give a distinctly better description of the  $2h - 1p$  satellite states. Finally, it is worth mentioning that vibronic coupling is another important factor for accurately simulating spectral functions.<sup>105,106</sup> Recent efforts include the development of similarity-transformed EOM vibrational coupled-cluster theory<sup>107</sup> and relevant studies for benzene and adenine molecules.<sup>108,109</sup> Our future development will also be focusing on this direction.

#### ACKNOWLEDGMENTS

This work was supported by the Center for Scalable, Predictive methods for Excitation and Correlated phenomena (SPEC), which is funded by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, the Division of Chemical Sciences, Geosciences, and Biosciences. All calculations have been performed using the Molecular Science Computing Facility (MSCF) in the Environmental Molecular Sciences Laboratory (EMSL) at the Pacific Northwest National Laboratory (PNNL). EMSL is funded by the Office of Biological and Environmental Research in the U.S. Department of Energy. PNNL is operated for the U.S. Department of Energy by the Battelle Memorial Institute under Contract No. DE-AC06-76RLO-1830. B.P. acknowledges the Linus Pauling Postdoctoral Fellowship from PNNL. B.P. and K.K. acknowledge the constructive comments from the reviewers.

- <sup>1</sup>J. Paldus and J. Čížek, *J. Chem. Phys.* **60**, 149 (1974).
- <sup>2</sup>J. Paldus and J. Čížek, "Time-independent diagrammatic approach to perturbation theory of fermion systems," in *Advances in Quantum Chemistry*, edited by P.-O. Löwdin (Academic Press, 1975), Vol. 9, pp. 105–197.
- <sup>3</sup>R. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem*, Dover Books on Physics, 2nd ed. (Dover Publications, 2012).
- <sup>4</sup>A. Abrikosov, L. Gorkov, I. Dzyaloshinski, and R. Silverman, *Methods of Quantum Field Theory in Statistical Physics*, Dover Books on Physics (Dover Publications, 2012).
- <sup>5</sup>A. Fetter and J. Walecka, *Quantum Theory of Many-Particle Systems*, Dover Books on Physics (Dover Publications, 2012).
- <sup>6</sup>A. Migdal, *Theory of Finite Fermi Systems and Application to Atomic Nuclei*, Interscience Monographs and Texts in Physics and Astronomy (Wiley-Interscience Publishers, 1967).
- <sup>7</sup>L. S. Cederbaum, *J. Phys. B: At. Mol. Phys.* **8**, 290 (1975).
- <sup>8</sup>W. von Niessen, J. Schirmer, and L. Cederbaum, *Comput. Phys. Rep.* **1**, 57 (1984).
- <sup>9</sup>J. V. Ortiz, "The electron propagator picture of molecular electronic structure," in *Computational Chemistry: Reviews of Current Trends*, edited by J. Leszczynski (World Scientific, Singapore, 1997), Vol. 2.
- <sup>10</sup>J. V. Ortiz, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **3**, 123 (2013).
- <sup>11</sup>V. G. Zakrzewski, O. Dolgounitcheva, A. V. Zakjevskii, and J. Ortiz, *Int. J. Quantum Chem.* **110**, 2918 (2010).
- <sup>12</sup>J. V. Ortiz, *J. Chem. Phys.* **108**, 1008 (1998).
- <sup>13</sup>W. von Niessen, G. H. F. Diercksen, and L. S. Cederbaum, *J. Chem. Phys.* **67**, 4124 (1977).
- <sup>14</sup>J. V. Ortiz, *J. Chem. Phys.* **104**, 7599 (1996).
- <sup>15</sup>J. Schirmer, *Phys. Rev. A* **26**, 2395 (1982).
- <sup>16</sup>A. Dreuw and M. Wormit, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **5**, 82 (2015).
- <sup>17</sup>F. Mertins and J. Schirmer, *Phys. Rev. A* **53**, 2140 (1996).
- <sup>18</sup>F. Mertins, J. Schirmer, and A. Tarantelli, *Phys. Rev. A* **53**, 2153 (1996).
- <sup>19</sup>L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- <sup>20</sup>J. J. Rehr and R. C. Albers, *Rev. Mod. Phys.* **72**, 621 (2000).
- <sup>21</sup>M. van Schilfgaarde, T. Kotani, and S. Faleev, *Phys. Rev. Lett.* **96**, 226402 (2006).
- <sup>22</sup>G. Samsonidze, M. Jain, J. Deslippe, M. L. Cohen, and S. G. Louie, *Phys. Rev. Lett.* **107**, 186404 (2011).
- <sup>23</sup>J. J. Kas, J. J. Rehr, and L. Reining, *Phys. Rev. B* **90**, 085112 (2014).
- <sup>24</sup>L. Reining, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **8**, e1344 (2018).
- <sup>25</sup>G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002).
- <sup>26</sup>S. V. Faleev, M. van Schilfgaarde, and T. Kotani, *Phys. Rev. Lett.* **93**, 126406 (2004).
- <sup>27</sup>J. B. Neaton, M. S. Hybertsen, and S. G. Louie, *Phys. Rev. Lett.* **97**, 216405 (2006).
- <sup>28</sup>A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996).
- <sup>29</sup>G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, *Rev. Mod. Phys.* **78**, 865 (2006).
- <sup>30</sup>D. Vollhardt, *Ann. Phys.* **524**, 1 (2012).
- <sup>31</sup>P. Werner and A. J. Millis, *Phys. Rev. B* **74**, 155107 (2006).
- <sup>32</sup>P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, *Phys. Rev. Lett.* **97**, 076405 (2006).
- <sup>33</sup>D. Zgid and G. K.-L. Chan, *J. Chem. Phys.* **134**, 094115 (2011).
- <sup>34</sup>D. Zgid, E. Gull, and G. K.-L. Chan, *Phys. Rev. B* **86**, 165128 (2012).
- <sup>35</sup>A. A. Kananenka, E. Gull, and D. Zgid, *Phys. Rev. B* **91**, 121111 (2015).
- <sup>36</sup>T. N. Lan, A. A. Kananenka, and D. Zgid, *J. Chem. Phys.* **143**, 241102 (2015).
- <sup>37</sup>A. A. Rusakov and D. Zgid, *J. Chem. Phys.* **144**, 054106 (2016).
- <sup>38</sup>T. N. Lan and D. Zgid, *J. Phys. Chem. Lett.* **8**, 2200 (2017).
- <sup>39</sup>S. Hirata, M. R. Hermes, J. Simons, and J. V. Ortiz, *J. Chem. Theory Comput.* **11**, 1595 (2015).
- <sup>40</sup>S. Hirata, A. E. Doran, P. J. Knowles, and J. V. Ortiz, *J. Chem. Phys.* **147**, 044108 (2017).
- <sup>41</sup>M. Nooijen and J. G. Snijders, *Int. J. Quantum Chem.* **44**, 55 (1992).
- <sup>42</sup>M. Nooijen and J. G. Snijders, *Int. J. Quantum Chem.* **48**, 15 (1993).
- <sup>43</sup>M. Nooijen and J. G. Snijders, *J. Chem. Phys.* **102**, 1681 (1995).
- <sup>44</sup>H. J. Monkhorst, *Int. J. Quantum Chem.* **12**, 421 (1977).
- <sup>45</sup>H. Koch and P. Jørgensen, *J. Chem. Phys.* **93**, 3333 (1990).
- <sup>46</sup>K. Kowalski, K. Bhaskaran-Nair, and W. A. Shelton, *J. Chem. Phys.* **141**, 094102 (2014).
- <sup>47</sup>K. Bhaskaran-Nair, K. Kowalski, and W. A. Shelton, *J. Chem. Phys.* **144**, 144101 (2016).
- <sup>48</sup>B. Peng and K. Kowalski, *Phys. Rev. A* **94**, 062512 (2016).
- <sup>49</sup>B. Peng and K. Kowalski, *Mol. Phys.* **116**, 561 (2018).
- <sup>50</sup>B. Peng and K. Kowalski, *J. Chem. Theory Comput.* **14**, 4335 (2018).
- <sup>51</sup>J. Schirmer and F. Mertins, *Theor. Chem. Acc.* **125**, 145 (2010).
- <sup>52</sup>T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, 2014).
- <sup>53</sup>L. S. Cederbaum, W. Domcke, and J. Schirmer, *Phys. Rev. A* **22**, 206 (1980).
- <sup>54</sup>G. Angonoa, O. Walter, and J. Schirmer, *J. Chem. Phys.* **87**, 6789 (1987).
- <sup>55</sup>J. Wenzel, M. Wormit, and A. Dreuw, *J. Chem. Theory Comput.* **10**, 4583 (2014).
- <sup>56</sup>A. B. Trofimov, T. É. Moskovskaya, E. V. Gromov, N. M. Vitkovskaya, and J. Schirmer, *J. Struct. Chem.* **41**, 483 (2000).
- <sup>57</sup>P. Norman and A. Dreuw, *Chem. Rev.* **118**, 7208 (2018).
- <sup>58</sup>S. Coriani, T. Fransson, O. Christiansen, and P. Norman, *J. Chem. Theory Comput.* **8**, 1616 (2012).
- <sup>59</sup>B. Peng, P. J. LeStrange, J. J. Goings, M. Caricato, and X. Li, *J. Chem. Theory Comput.* **11**, 4146 (2015).
- <sup>60</sup>D. Zuev, E. Vecharynski, C. Yang, N. Orms, and A. I. Krylov, *J. Comput. Chem.* **36**, 273 (2014), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.23800>.
- <sup>61</sup>J. Rotureau, P. Danielewicz, G. Hagen, F. M. Nunes, and T. Papenbrock, *Phys. Rev. C* **95**, 024315 (2017).
- <sup>62</sup>A. B. Trofimov and J. Schirmer, *J. Chem. Phys.* **123**, 144115 (2005).
- <sup>63</sup>M. S. Deleuze and S. Knippenberg, *J. Chem. Phys.* **125**, 104309 (2006).
- <sup>64</sup>O. Dolgounitcheva, V. G. Zakrzewski, and J. V. Ortiz, *J. Phys. Chem. A* **113**, 14630 (2009).
- <sup>65</sup>O. Plekan, V. Feyrer, R. Richter, M. Coreno, M. de Simone, K. C. Prince, A. B. Trofimov, E. V. Gromov, I. L. Zaytseva, and J. Schirmer, *Chem. Phys.* **347**, 360 (2008).
- <sup>66</sup>L. S. Cederbaum, J. Schirmer, W. Domcke, and W. von Niessen, *J. Phys. B: At. Mol. Phys.* **10**, L549 (1977).
- <sup>67</sup>L. S. Cederbaum, W. Domcke, J. Schirmer, and W. von Niessen, *Phys. Scr.* **21**, 481 (1980).
- <sup>68</sup>J. McClain, J. Lischner, T. Watson, D. A. Matthews, E. Ronca, S. G. Louie, T. C. Berkelbach, and G. K.-L. Chan, *Phys. Rev. B* **93**, 235139 (2016).
- <sup>69</sup>H. Nishi, T. Kosugi, Y. Furukawa, and Y.-i. Matsushita, *J. Chem. Phys.* **149**, 034106 (2018).
- <sup>70</sup>T. Kosugi, H. Nishi, Y. Furukawa, and Y.-i. Matsushita, *J. Chem. Phys.* **148**, 224103 (2018).
- <sup>71</sup>Y. Furukawa, T. Kosugi, H. Nishi, and Y.-i. Matsushita, *J. Chem. Phys.* **148**, 204109 (2018).
- <sup>72</sup>K. Jankowski, J. Paldus, and P. Piecuch, *Theor. Chim. Acta* **80**, 223 (1991).
- <sup>73</sup>K. Kowalski and P. Piecuch, *J. Chem. Phys.* **113**, 18 (2000).
- <sup>74</sup>J. F. Stanton and J. Gauss, *Theor. Chim. Acta* **93**, 303 (1996).
- <sup>75</sup>J. C. Saeh and J. F. Stanton, *J. Chem. Phys.* **111**, 8275 (1999).
- <sup>76</sup>P. U. Manohar, J. F. Stanton, and A. I. Krylov, *J. Chem. Phys.* **131**, 114112 (2009).
- <sup>77</sup>M. Urban, J. Noga, S. J. Cole, and R. J. Bartlett, *J. Chem. Phys.* **83**, 4041 (1985).
- <sup>78</sup>J. D. Watts and R. J. Bartlett, *Chem. Phys. Lett.* **233**, 81 (1995).
- <sup>79</sup>J. D. Watts and R. J. Bartlett, *Chem. Phys. Lett.* **258**, 581 (1996).
- <sup>80</sup>J. F. Stanton and J. Gauss, *J. Chem. Phys.* **111**, 8785 (1999).
- <sup>81</sup>Y. J. Bomble, J. C. Saeh, J. F. Stanton, P. G. Szalay, M. Kállay, and J. Gauss, *J. Chem. Phys.* **122**, 154107 (2005).
- <sup>82</sup>H. Koch, O. Christiansen, P. Jørgensen, A. M. S. de Mers, and T. Helgaker, *J. Chem. Phys.* **106**, 1808 (1997).
- <sup>83</sup>O. Christiansen, H. Koch, and P. Jørgensen, *Chem. Phys. Lett.* **243**, 409 (1995).
- <sup>84</sup>S. Hirata, M. Nooijen, and R. J. Bartlett, *Chem. Phys. Lett.* **326**, 255 (2000).
- <sup>85</sup>M. Kamiya and S. Hirata, *J. Chem. Phys.* **125**, 074111 (2006).
- <sup>86</sup>L. V. Slipchenko and A. I. Krylov, *J. Chem. Phys.* **123**, 084107 (2005).
- <sup>87</sup>D. A. Matthews and J. F. Stanton, *J. Chem. Phys.* **145**, 124102 (2016).
- <sup>88</sup>T.-C. Jagau, *J. Chem. Phys.* **148**, 024104 (2018).
- <sup>89</sup>G. Herzberg, *Molecular Spectra and Molecular Structure* (Read Books Limited, 2013), Vol. 1.
- <sup>90</sup>M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, *Comput. Phys. Commun.* **181**, 1477 (2010).
- <sup>91</sup>M. Ehara, M. Ishida, and H. Nakatsuji, *Collect. Czech. Chem. Commun.* **70**, 881 (2005).
- <sup>92</sup>M. Lebeck, J. C. Houver, G. Raseev, A. S. dos Santos, D. Dowek, and R. R. Lucchese, *J. Chem. Phys.* **136**, 094303 (2012).

- <sup>93</sup>S. Krummacher, V. Schmidt, and F. Wuilleumier, *J. Phys. B: At. Mol. Phys.* **13**, 3993 (1980).
- <sup>94</sup>S. Krummacher, V. Schmidt, F. Wuilleumier, J. M. Bizau, and D. Ederer, *J. Phys. B: At. Mol. Phys.* **16**, 1733 (1983).
- <sup>95</sup>S. Svensson, M. Carlsson-Göthe, L. Karlsson, A. Nilsson, N. Mårtensson, and U. Gelius, *Phys. Scr.* **44**, 184 (1991).
- <sup>96</sup>J. Schirmer, L. Cederbaum, W. Domcke, and W. von Niessen, *Chem. Phys.* **26**, 149 (1977).
- <sup>97</sup>A. Potts and T. Williams, *J. Electron Spectrosc. Relat. Phenom.* **3**, 3 (1974).
- <sup>98</sup>L. S. Cederbaum and W. Domcke, "Theoretical aspects of ionization potentials and photoelectron spectroscopy: A green's function approach," in *Advances in Chemical Physics* (John Wiley & Sons, Ltd., 2007), pp. 205–344.
- <sup>99</sup>J. Schirmer and L. S. Cederbaum, *J. Phys. B: At. Mol. Phys.* **11**, 1889 (1978).
- <sup>100</sup>P. S. Bagus and E.-K. Viinikka, *Phys. Rev. A* **15**, 1486 (1977).
- <sup>101</sup>P. S. Bagus, C. Sousa, and F. Illas, *J. Chem. Phys.* **145**, 144303 (2016).
- <sup>102</sup>A. A. Bakke, H.-W. Chen, and W. L. Jolly, *J. Electron Spectrosc. Relat. Phenom.* **20**, 333 (1980).
- <sup>103</sup>J. Stöhr, *NEXAFS Spectroscopy*, Springer Series in Surface Sciences (Springer, Berlin, Heidelberg, 2013).
- <sup>104</sup>J. Schirmer, G. Angonoa, S. Svensson, D. Nordfors, and U. Gelius, *J. Phys. B: At. Mol. Phys.* **20**, 6031 (1987).
- <sup>105</sup>L. Cederbaum and H. Köppel, *Chem. Phys. Lett.* **87**, 14 (1982).
- <sup>106</sup>H. Köppel, W. Domcke, and L. S. Cederbaum, "Multimode molecular dynamics beyond the Born-Oppenheimer approximation," *Advances in Chemical Physics* (Wiley-Blackwell, 2007), pp. 59–246.
- <sup>107</sup>J. A. Fauchaux, M. Nooijen, and S. Hirata, *J. Chem. Phys.* **148**, 054104 (2018).
- <sup>108</sup>H. Köppel, L. S. Cederbaum, and W. Domcke, *J. Chem. Phys.* **89**, 2023 (1988).
- <sup>109</sup>K. B. Bravaya, O. Kostko, S. Dolgikh, A. Landau, M. Ahmed, and A. I. Krylov, *J. Phys. Chem. A* **114**, 12305 (2010).