Supplementary Information: Real-time coupled-cluster approach for the cumulant Green's function

F. D. Vila, J. J. Rehr, and J. J. Kas

Department of Physics, University of Washington, Seattle, WA 98195

K. Kowalski

William R. Wiley Environmental Molecular Sciences Laboratory,

Battelle, Pacific Northwest National Laboratory,

K8-91, P. O. Box 999, Richland, Washington 99352

B. Peng

Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354

(Dated: September 24, 2025)

Abstract

This Supplementary Information presents results that are too detailed for the main manuscript, including more detailed discussion of the theory derivations in the main manuscript, computational details, results for other basis sets, and more results that are too technical for the main manuscript.

I. THEORY

A. Retarded Green's function in spin-orbital basis

The real-space, real-time retarded Green's function is defined as

$$\begin{aligned} G(x,x') = G(rt,r't') = \\ -i\Theta(t-t')\left\langle 0\left|\left\{\psi(rt),\psi^\dagger(r't')\right\}\right|0\right\rangle \end{aligned} \tag{1}$$

where $|0\rangle$ is the ground state of the system, $\psi^\dagger(r't')$ and $\psi(rt)$ are, respectively, the creation and annhilation field operators at $r't'$ and rt, and $\{,\}$ indicates the anticonmutation operator. By introducing a basis set of single-particle spin-orbitals $\{\phi_p(r)\}$, we can express any two-positions, two-times operator M as:

$$M(rt,r't') = \sum_{pq}\phi_p^*(r)M_{pq}(t,t')\phi_q(r') \tag{2}$$

where the matrix elements are defined as:

$$M_{pq}(t,t') = \int drdr'\phi_p^*(r)M(rt,r't')\phi_q(r'). \tag{3}$$

Inserting these definitions into Eq. (1) we obtain the Green's function matrix expressed in the spin-orbital basis $\{\phi_p(r)\}$:

$$G_{pq}(t,t') = -i\Theta(t-t')\left\langle 0\left|\left\{a_p(t),a_q^\dagger(t')\right\}\right|0\right\rangle, \tag{4}$$

where the creation and annhilation operators $a_p^\dagger(t)$ and $a_q(t')$ are associated with the spin-orbitals ϕ_p and ϕ_q respectively.

B. From the real-space time-domain Dyson equation to the cumulant ansatz in a spin-orbital basis

The real-space, time-domain form of the Dyson equation

$$\begin{aligned} G(x,x') = G^0(x,x')+ \\ \int dx_1dx_2G^0(x,x_1)\Sigma(x_1,x_2)G(x_2,x') \end{aligned} \tag{5}$$

can also be cast in matrix form (after including time translation invariance):

$$\hat{G}(t) = \hat{G}^0(t) + \int dt_1dt_2\hat{G}^0(t-t_1)\hat{\Sigma}(t_1-t_2)\hat{G}(t_2). \tag{6}$$

Expanding the matrix form of the cumulant ansatz

$$\hat{G}(t) = \hat{G}^0(t)e^{\hat{C}(t)}, \tag{7}$$

and Eq. (6) to first order

$$\hat{G}(t) = \hat{G}^0(t) + \hat{G}^0(t)\hat{C}(t) + ... \tag{8}$$

$$\hat{G}(t) = \hat{G}^0(t) + \int dt_1 dt_2 \hat{G}^0(t-t_1)\hat{\Sigma}(t_1-t_2)\hat{G}^0(t_2) + ... \tag{9}$$

and equating the second terms in the right hand sides we get:

$$\hat{G}^0(t)\hat{C}(t) = \int dt_1 dt_2 \hat{G}^0(t-t_1)\hat{\Sigma}(t_1-t_2)\hat{G}^0(t_2) \tag{10}$$

For a typical matrix element we obtain

$$\sum_r G^0_{pr}(t)C_{rq}(t) = \sum_{rs}\int dt_1 dt_2 G^0_{pr}(t-t_1)\Sigma_{rs}(t_1-t_2)G^0_{sq}(t_2), \tag{11}$$

Shifting the ω integration variable and using the Fourier transform of the cumulant we have:

$$C_{pq}(\omega) = iG^0_{pp}(\omega+\epsilon_p)\Sigma_{pq}(\omega+\epsilon_p)G^0_{qq}(\omega+\epsilon_p). \tag{12}$$

Finally, introducing the frequency domain form of $G^0_{pp} = (\omega - \epsilon_p + i\delta)^{-1}$:

$$C_{pq}(\omega) = \frac{i\Sigma_{pq}(\omega+\epsilon_p)}{(\omega+i\delta)(\omega+\epsilon_p-\epsilon_q+i\delta)}. \tag{13}$$

If we approximate $\Sigma_{pq}(\omega) \simeq \Sigma_{pp}(\omega)\delta_{pq}$, then

$$C_{pp}(\omega) = \frac{i\Sigma_{pp}(\omega+\epsilon_p)}{(\omega+i\delta)^2}, \tag{14}$$

or, in the time domain

$$C_{pp}(t) = \int \frac{d\omega}{2\pi}\frac{i\Sigma_{pp}(\omega+\epsilon_p)}{(\omega+i\delta)^2}e^{-i\omega t}. \tag{15}$$

We can now return to the time domain, inserting into Eq. (7) and taking into account that the matrix exponential is now diagonal:

$$G_{pp}(t) = i\Theta(t)e^{-i\epsilon_p t + C_{pp}(t)}, \tag{16}$$

which is the standard form of the diagonal cumulant.

C. Coupled Cluster Green's function in time

In this section we derive a compact form for the full Coupled Cluster Green's function that can be used for further derivation of time-domain approximations. Starting with Eq. (17) in Ref. 1 (from now on referred as "PK") we have

$$G^R_{pq}(\omega) = \left\langle \Phi \left| (1+\Lambda) e^{-T} a_q^\dagger \left(\omega + (H - E_0) - i\delta\right)^{-1} a_p e^T \right| \Phi \right\rangle \tag{17}$$

and inserting the $I = e^{-T}e^T$ we get

$$G^R_{pq}(\omega) = \left\langle \Phi \left| (1+\Lambda) \bar{a_q^\dagger} \left(\omega + \bar{H}_N\right)^{-1} \bar{a}_p \right| \Phi \right\rangle \tag{18}$$

where $\bar{O} = e^{-T}Oe^T$ is the similarity transformed form of the O operator, H_N is the normal ordered hamiltonian, and using the Baker-Campbell-Hausdorff (BCH) relation we have that

$$\bar{a}_p = a_p + [a_p, T], \tag{19}$$

$$\bar{a_q^\dagger} = a_q^\dagger + [a_q^\dagger, T]. \tag{20}$$

For simplicity, we now make the convergence factor $-i\delta$ implicit in the energy ω.

Following PK, we introduce the $X_p(\omega)$, $Z_q(\omega)$, and $W_q(\omega)$ operators, which are solutions to the following equations:

$$X_p(\omega) \left|\Phi\right\rangle = \left(\omega + \bar{H}_N\right)^{-1} \bar{a}_p \left|\Phi\right\rangle, \tag{21}$$

$$\left\langle\Phi\right| Z_q(\omega) = \left\langle\Phi\right| (1+\Lambda) \bar{a_q^\dagger} \left(\omega + \bar{H}_N\right)^{-1}, \tag{22}$$

and

$$\left\langle\Phi\right| (1+\Lambda) W_q(\omega) = \left\langle\Phi\right| (1+\Lambda) \bar{a_q^\dagger} \left(\omega + \bar{H}_N\right)^{-1} \tag{23}$$

These operators have the following expansions in the $N-1$ Fock space:

$$\begin{aligned} X_p(\omega) = &\sum_i x^i(\omega)_p a_i + \\ &\frac{1}{2!} \sum_{ij,a} x^{ij}_a(\omega)_p a_a^\dagger a_j a_i + ... \end{aligned} \tag{24}$$

$$\begin{aligned} Z_q(\omega) = &\sum_i z_i(\omega)_q a_i^\dagger + \\ &\frac{1}{2!} \sum_{ij,a} z^a_{ij}(\omega)_p a_i^\dagger a_j^\dagger a_a + ... \end{aligned} \tag{25}$$

and

$$W_q(\omega) = \sum_i w_i(\omega)_q a_i^\dagger + \frac{1}{2!}\sum_{ij,a} w_{ij}^a(\omega)_p a_i^\dagger a_j^\dagger a_a + ... \tag{26}$$

Eqs. (17)-(26) are a summary of the formulation in PK. Now, using

$$\text{IFT}\left[\left(\omega + \bar{H}_N - i\delta\right)^{-1}\right] = i\Theta(-t)e^{i\bar{H}_N t} \tag{27}$$

where IFT is the inverse Fourier transform, we can write $G_{pq}^R(\omega)$ in the time-domain as:

$$G_{pq}^R(t) = i\left\langle \Phi \left| (1+\Lambda)\bar{a_q^\dagger} e^{i\bar{H}_N t}\bar{a}_p \right| \Phi \right\rangle \tag{28}$$

where from now on we assume that $t < 0$ to remove all the $\Theta(-t)$ functions. We can also convert the equations defining the $X_p(\omega)$, $Z_q(\omega)$, and $W_q(\omega)$ operators as:

$$X_p(t)\left|\Phi\right\rangle = ie^{i\bar{H}_N t}\bar{a}_p\left|\Phi\right\rangle, \tag{29}$$

$$\left\langle\Phi\right| Z_q(t) = i\left\langle\Phi\right|(1+\Lambda)\bar{a_q^\dagger}e^{i\bar{H}_N t} \tag{30}$$

and

$$\left\langle\Phi\right|(1+\Lambda)W_q(t) = i\left\langle\Phi\right|(1+\Lambda)\bar{a_q^\dagger}e^{i\bar{H}_N t} \tag{31}$$

where the $X_p(t)$, $Z_q(t)$, and $W_q(t)$ are defined simply by the inverse Fourier transform of their coefficients in their excitation expansions (Eqs. (24)-(26)).

Based on the properties of the $\bar{H}_N$, i.e., where one assumes that the CC equations $Q\bar{H}_N|\Phi\rangle = 0$ are satisfied, where Q denotes the projection operator onto the space spanned by excitations with respect to $|\Phi\rangle$ Slater determinants, we can prove that, in full analogy with the frequency representation, the $X_p(t)$ operator can be expressed in terms of connected diagrams only. To this end we expand $X_p(t)$ in term of powers of the $\bar{H}_N$ operator

$$X_p(t)|\Phi\rangle = ie^{i\bar{H}_N t}\bar{a_p}|\Phi\rangle, \tag{32}$$

$$= i\{\sum_{n=0}^{\infty}\frac{1}{n!}(i)^n t^n(\bar{H}_N)^n\bar{a_p}\}|\Phi\rangle \tag{33}$$

Now we assume that we are dealing with the exact CC theory. This assumption plays a crucial role in proving the connected character of $X_p(t)$ since all approximate approaches

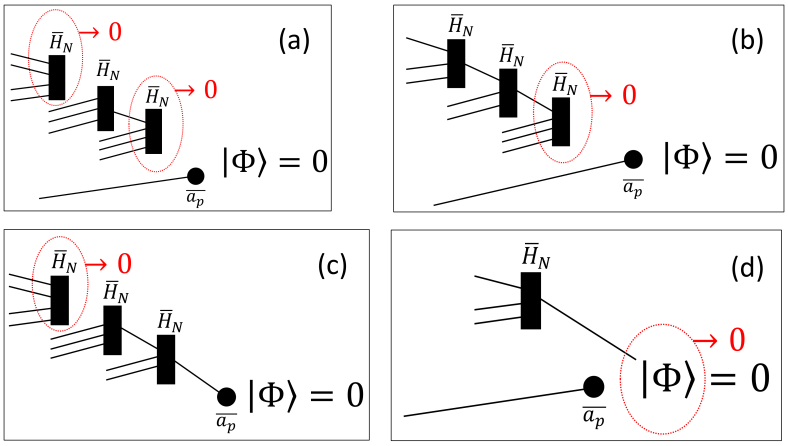

FIG. 1: Typical examples of diagrams not contributing to the general term in Eq. (34).

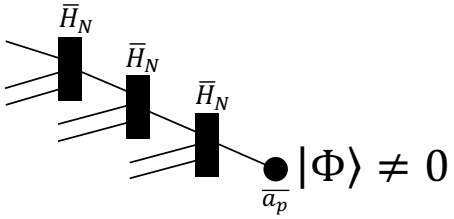

FIG. 2: Example of a connected diagram that contributes to the general term in Eq. (34).

can be build using connected properties of the exact formulation. Using Wick's theorem for the particle-hole formalism to analize a general term in the expansion in Eq. (33):

$$(\bar{H}_N)^n \bar{a_p}|\Phi\rangle, \tag{34}$$

we can identify several classes of diagrams contributing to $(\bar{H}_N)^n \bar{a_p}|\Phi\rangle$ (see Figs. (1) and (2)). It can be easily verified that all disconnected diagrams (typical examples of these are shown in Fig. (1a-1d)) disappear due to the existence of vertices representing projections of CC equations (i.e., $\bar{H}_N$ matrix elements with all particle-hole creation lines, that is, all "legs", located to the left of the corresponding matrix element or diagrammatic vertex, see Fig. (1a-1c)) or due to the existence of the uncontracted particle-hole line (lines) that annihilates the reference function $|\Phi\rangle$ as shown in Fig. (1d). Consequently, the only diagrams contributing to $X_p(t)$ are connected diagrams (Fig.(2)) which can be symbolically denoted as

$$X_p(t)|\Phi\rangle = i\{e^{i\bar{H}_N t}\bar{a_p}\}_C|\Phi\rangle \ , \tag{35}$$

where the subscript "C" designates connected part of a given operator expression.

Using the equations above for $X_p(t)$, $Z_q(0^-)$, and $W_q(0^-)$, we can re-write $G^R_{pq}(t)$ as:

$$\begin{aligned} G^R_{pq}(t) &= \left\langle \Phi \left| (1+\Lambda)\bar{a}^\dagger_q X_p(t) \right| \Phi \right\rangle \\ &= -i \left\langle \Phi \left| Z_q(0^-) X_p(t) \right| \Phi \right\rangle \\ &= -i \left\langle \Phi \left| (1+\Lambda) W_q(0^-) X_p(t) \right| \Phi \right\rangle \end{aligned} \tag{36}$$

The first equality is just the time-domain version of Eq. (27) in PK. Note that the Z_q and W_q operators are computed as the limit $t \to 0^-$ to avoid any issues with the poorly defined nature of $\Theta(-t)$ at 0.

We now proceed to expand the second equality in Eq. (36). We choose to expand this form of $G^R_{pq}(t)$ because it does not include the Λ terms (which are implicit in the definition of $Z_q(0^-)$) and is thus simpler. Inserting the time dependent definitions of $X_p(t)$ and $Z_q(0^-)$ into $G^R_{pq}(t)$ we have

$$\begin{aligned} iG^R_{pq}(t) = \langle \Phi | \Big(\sum_i z_i(0^-)_q \{a^\dagger_i\} + \frac{1}{2!} \sum_{ij,a} z^a_{ij}(0^-)_p \{a^\dagger_i a^\dagger_j a_a\} + ... \Big) \\ \Big(\sum_k x^k(t)_p \{a_k\} + \frac{1}{2!} \sum_{kl,b} x^{kl}_b(t)_p \{a^\dagger_b a_l a_k\} + ... \Big) | \Phi \rangle \end{aligned} \tag{37}$$

where, to simplify the rest of the calculations, we also introduced the normal ordered forms of the operators, indicated using Bartlett's style notation $\{...\}$ rather than the PK one with $N[...]$. Before going any further it is helpful to analyze the general form of a generic matrix element for the different products:

$$\langle \Phi | \underbrace{\{a^\dagger_i a^\dagger_j ... a_b a_a\}}_{n \text{ ops}} \underbrace{\{a^\dagger_c a^\dagger_d ... a_l a_k\}}_{m \text{ ops}} | \Phi \rangle = \sum_{\text{FC}} \langle \Phi | \{ \overbrace{a^\dagger_i a^\dagger_j ... a_b a_a a^\dagger_c a^\dagger_d ... a_l a_k}^{\text{FC}} \} | \Phi \rangle \tag{38}$$

where, according to the generalized Wick's theorem (GWT), we compute the Fermi vacuum expectation value of the product of two normal ordered operator sets by summing over all possible full contractions (FC). However, for $n \neq m$, the are no possible full contractions between the two operator sets, thus, the "cross" terms in Eq. (37) are all zero. For example, we are only left with terms of the form:

$$\langle \Phi | \{a^\dagger_i\} \{a_k\} | \Phi \rangle = \langle \Phi | \{ \overbrace{a^\dagger_i a_k}^{\text{FC}} \} | \Phi \rangle = \delta_{ik} \tag{39}$$

for 0 excitation order (*i.e.* order 1 in PK),

$$\begin{aligned} \langle \Phi | \{a^\dagger_i a^\dagger_j a_a\} \{a^\dagger_b a_l a_k\} | \Phi \rangle = \langle \Phi | \{ \overbrace{a^\dagger_i a^\dagger_j a_a a^\dagger_b a_l a_k}^{\text{FC}} \} | \Phi \rangle + \langle \Phi | \{ \overbrace{a^\dagger_i a^\dagger_j a_a a^\dagger_b a_l a_k}^{\text{FC}} \} | \Phi \rangle = \\ - \delta_{il} \delta_{jk} \delta_{ab} + \delta_{ik} \delta_{jk} \delta_{ab} \end{aligned} \tag{40}$$

for 1 excitation order (*i.e.* order 2 in PK), etc. The pattern is fairly clear: each order produces $h! \times p!$, where h and p are the number of hole and particle operators in the products. Thus PK order 1 produces one term, order 2 produces 2, order 3 would produce 12, etc. Thus we can write:

$$iG^R_{pq}(t) = \sum_i z_i(0^-)_q x^i(t)_p + \frac{1}{4}\sum_{ij,a} z^a_{ij}(0^-)_p x^{ij}_a(t)_p - \frac{1}{4}\sum_{ij,a} z^a_{ij}(0^-)_p x^{ji}_a(t)_p + ... \tag{41}$$

This sum can be generalized by recognizing that the $x^{ij}_a(t)_p$ coefficients are antisymmetric with respect to index swaps, and that $x^{ii}_a(t)_p = 0$, thus

$$\begin{aligned} iG^R_{pq}(t) = & \\ & \sum_i z_i(0^-)_q x^i(t)_p + \\ & + \frac{1}{2}\sum_{ij,a} z^a_{ij}(0^-)_p x^{ij}_a(t)_p + \\ & + \frac{1}{6}\sum_{ijk,ab} z^{ab}_{ijk}(0^-)_p x^{ijk}_{ab}(t)_p + ... \end{aligned} \tag{42}$$

D. Perturbation theory of $G^R_{pq}(t)$

In this section we demonstrate that the full form of the Coupled Cluster Green's function can be reduced to a cumulant form when approximated as a perturbations series. Although this analysis is valid for any CC level (CCD, CCSD, CCSDT, etc.) with HF orbitals and Moller Plesset MBPT, here for simplicity we limit the CC expansion to doubles ($T = T_2 = \frac{1}{4}\sum_{ijab} t^{ab}_{ij} a^\dagger_a a^\dagger_b a_j a_i$, $\Lambda = \Lambda_2 = \frac{1}{4}\sum_{ijab} \lambda^{ab}_{ij} a^\dagger_i a^\dagger_j a_b a_a$) and expand the $\bar{a}_p$ and $\bar{a^\dagger_q}$ operators into their connected forms:

$$\begin{aligned} G^R_{pq}(\omega) = & \\ & \left\langle \Phi \left| (1+\Lambda_2)(a^\dagger_q + (a^\dagger_q T_2)_C) X_p(\omega) \right| \Phi \right\rangle + \\ & \left\langle \Phi \left| (1+\Lambda_2)(a_p + (a_p T_2)_C) Y_q(\omega) \right| \Phi \right\rangle \end{aligned} \tag{43}$$

where the X_p and Y_q operators are

$$X_p(\omega)\left|\Phi\right\rangle = (\omega + \bar{H}_N + i\delta)^{-1}\bar{a}_p\left|\Phi\right\rangle, \tag{44}$$

$$Y_q(\omega)\left|\Phi\right\rangle = (\omega - \bar{H}_N + i\delta)^{-1}\bar{a}^\dagger_q\left|\Phi\right\rangle, \tag{45}$$

with components

$$X_p(\omega) = \sum_i x^i(\omega)_p a_i + \frac{1}{2!}\sum_{ij,a} x^{ij}_a(\omega)_p a^\dagger_a a_j a_i = X_{1,p}(\omega) + X_{2,p}(\omega), \tag{46}$$

$$Y_q(\omega) = \sum_a y_a(\omega)_q a^\dagger_a + \frac{1}{2!}\sum_{i,ab} y^i_{ab}(\omega)_q a^\dagger_a a^\dagger_b a_i = Y_{1,q}(\omega) + Y_{2,q}(\omega). \tag{47}$$

We proceed in the frequency domain because this simplifies the perturbation analysis. Each term in Eq. (43) generates six possible terms, but, analyzing their corresponding diagrams we can see that there are only three non-zero terms when $p, q \in occ$:

$$G^R_{pq}(\omega) = \left\langle \Phi \left| a^\dagger_q X_{1,p}(\omega) \right| \Phi \right\rangle + \left\langle \Phi \left| \Lambda_2 (a^\dagger_q T_2)_C X_{1,p}(\omega) \right| \Phi \right\rangle + \left\langle \Phi \left| \Lambda_2 a_p Y_{2,q}(\omega) \right| \Phi \right\rangle. \tag{48}$$

After computing each term we get:

$$G^R_{pq}(\omega) = x^q(\omega)_p - \frac{1}{2}\sum_{ijab} \lambda^{ab}_{ij} t^{qj}_{ab} x^i(\omega)_p - \frac{1}{2}\sum_{iab} \lambda^{ab}_{pi} y^i_{ab}(\omega)_q. \tag{49}$$

This is the form of the Green's function which can be used to approximate G^R_{pq} using perturbation theory. For this purpose we write each of the coefficients in Eq. (49) up to second order in a perturbation parameter λ, e.g. $t^{rs}_{pq} = t^{(0)rs}_{pq} + \lambda t^{(1)rs}_{pq} + \lambda^2 t^{(2)rs}_{pq}$, etc. By keeping only terms up to second order we get:

$$G^{(0)R}_{pq}(\omega) = x^{(0)q}(\omega)_p \tag{50}$$

$$G^{(1)R}_{pq}(\omega) = x^{(1)q}(\omega)_p - \frac{1}{2}\sum_{iab} \lambda^{(1)ab}_{pi} y^{(0)i}_{ab}(\omega)_q, \tag{51}$$

$$G^{(2)R}_{pq}(\omega) = x^{(2)q}(\omega)_p - \frac{1}{2}\sum_{ijab} \lambda^{(1)ab}_{ij} t^{(1)qj}_{ab} x^{(0)i}(\omega)_p - \frac{1}{2}\sum_{iab} (\lambda^{(1)ab}_{pi} y^{(1)i}_{ab}(\omega)_q + \lambda^{(2)ab}_{pi} y^{(0)i}_{ab}(\omega)_q) \tag{52}$$

It is easy to prove that $x^{(1)q}(\omega)_p = 0$ and that $x^{(0)q}(\omega)_p = x^{(0)p}(\omega)_p \delta_{pq}$. It can also be easily proven that $Y^{(0)}_q(\omega) = 0$, resulting in $G^{(1)R}_{pq}(\omega) = 0$. After this, the retarded Green's function to second order is:

$$G^R_{pq}(\omega) = x^{(0)p}(\omega)_p \delta_{pq} + x^{(2)q}(\omega)_p - \frac{1}{2}\sum_{iab} \lambda^{(1)ab}_{pi} (t^{(1)qi}_{ab} x^{(0)p}(\omega)_p + y^{(1)i}_{ab}(\omega)_q) \tag{53}$$

Here we skip the derivation of each of the coefficients in term of the two-particle integrals and HF eigenvalues and simply list expressions:

$$x^{(0)p}(\omega)_p = \frac{1}{(\omega - \epsilon_p)} \tag{54}$$

$$x^{(2)q}(\omega)_p = \frac{1}{2(\omega-\epsilon_q)}\left[\sum_{ija} v_{ij}^{qa} x_a^{(1)ij}(\omega)_p + \frac{1}{(\omega-\epsilon_p)}\sum_{iab} v_{ab}^{pi} t_{ab}^{(1)qi}\right] \tag{55}$$

$$x_a^{(1)ij}(\omega)_p = \frac{v_{pa}^{ij}}{(\omega-\epsilon_p)(\omega+\epsilon_a-\epsilon_i-\epsilon_j)} \tag{56}$$

$$t_{ab}^{(1)ij} = \lambda_{ij}^{(1)ab} = \frac{v_{ab}^{ij}}{(\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b)} = \frac{v_{ab}^{ij}}{\epsilon_{ab}^{ij}} \tag{57}$$

$$x_a^{(1)ij}(\omega)_p = \frac{v_{pa}^{ij}}{(\omega-\epsilon_p)(\omega+\epsilon_a-\epsilon_i-\epsilon_j)} \tag{58}$$

The final Dyson and cumulant form of $G_{pq}^R(t)$ becomes pparent after inserting these expressions into Eq. (53) and proceeding as described in the main paper.

E. One-particle simplification of the EOM-CCS equations

Given the computational demand of the full EOM-CCS method, it is of interest to see if approximations other than those explored in the main paper are possible. In particular, approximations arising from effective one-body Hamiltonians since these are common in the works of Hedin, Nozieres, Langreth, etc. The full EOM-CCS method has the following matrix element for the computation of the amplitude variation:

$$\begin{aligned}\langle\phi_i^a|\,\bar{H}_N(t)\,|\phi\rangle =& f_{ai} + \sum_b f_{ab}t_i^b - \sum_j f_{ji}t_j^a - \sum_{jb} f_{jb}t_i^b t_j^a \\ &+ \sum_{jb} v_{aj}^{ib}t_j^b - \sum_{jkb} v_{ib}^{jk}t_j^a t_k^b + \sum_{jbc} v_{aj}^{bc}t_i^b t_j^c - \sum_{jkbd} v_{jk}^{bd}t_i^b t_j^a t_k^d,\end{aligned} \tag{59}$$

where the f terms come from the one-particle part of H and the v terms from the two-particle part. A simple, yet extreme approximation would be to only retain the first line in Eq. (59). However, this approximation leaves out the important linear term in v. We now partition that term as follows:

$$\sum_{jb} v_{aj}^{ib}t_j^b = v_{ia}^{ia}t_i^a + \sum_j v_{aj}^{ia}t_j^a + \sum_b v_{ai}^{ib}t_i^b + \sum_{jb(\neq ia)} v_{aj}^{ib}t_j^b \tag{60}$$

Discarding the last term and introducing the approximate sum into the matrix element expression above while only keeping the f terms we obtain:

$$\langle\phi_i^a|\,\bar{H}_N(t)\,|\phi\rangle = f_{ai} + v_{ia}^{ia}t_i^a + \sum_b (f_{ab} - v_{aj}^{bi})t_i^b - \sum_j (f_{ji} - v_{aj}^{ia})t_j^a - \sum_{jb} f_{jb}t_i^b t_j^a \tag{61}$$

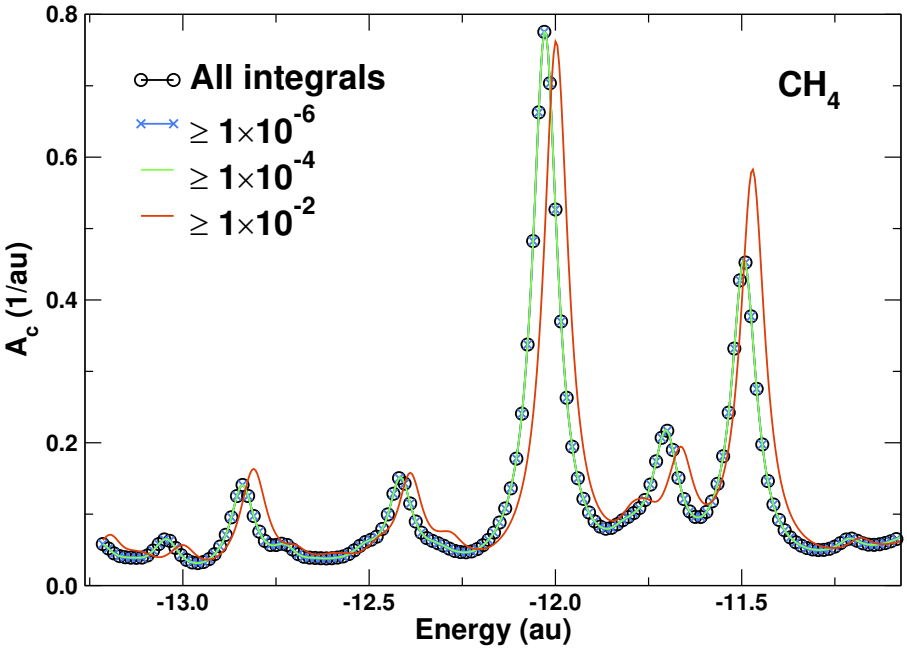

FIG. 3: Comparison of the satellite region of the core spectral function of CH_4 computed with the cc-pVDZ basis set and the full EOM-CCS method, as a function of potential integral cutoff.

Therefore, with this approximation the form of the one-particle H is preserved, but with modified f elements, which now include a correction for the particular (i, a) valence-valence excitation. We are currently exploring the possibility of using this simplified propagation form with f and v parameters from effective Hamiltonians.

II. RESULTS

A. Effect of potential integral trimming

As discussed in the main manuscript, in order to reduce the storage requirements and computational demands of the EOM-CC method, we screen those v_{pq}^{rs} integrals below a certain threshold after the SCF is fully converged. Figure 3 shows a comparison between the spectral function of a typical system (CH_4) for different values of the cutoff parameter. We find that for the cutoff used in this paper (1×10^{-4} au) the EOM-CC results are indistinguishable from those obtained with all the integrals. With this cutoff, however, the performance of the method is increase by a factor of 10. This cutoff is also used in the DSE2 calculations,

TABLE I: Comparison of the experimental core binding energies (in eV) to those obtained with the DZVP basis set, using the L and NL approximations to the cumulant and the 1-3 approximations of the EOM-CCS method, and their mean absolute errors (MAE).

System	**KT**	**DSE2**	$\mathbf{1}_L$	$\mathbf{2}_L$	$\mathbf{3}_L$	$\mathbf{1}_{NL}$	$\mathbf{2}_{NL}$	$\mathbf{3}_{NL}$	**Expt**	**Ref**
CH_4	304.744	291.881	286.990	287.425	286.994	290.412	290.679	290.415	290.703	[5]
NH_3	422.523	405.466	400.603	400.815	400.198	405.057	405.177	404.816	405.520	[6]
H_2O	559.003	538.597	534.795	534.390	533.705	539.498	539.248	538.843	539.700	[7]
HF	714.753	692.127	689.876	688.904	688.313	694.174	693.549	693.178	694.200	[8]
Ne	890.987	868.010	867.661	866.444	866.109	870.935	870.076	869.842	870.200	[9]
MAE	18.34	1.32	4.08	4.47	5.00	0.34	0.32	0.65		

where it also has little effect in the accuracy, but produced only a modest improvement in performance. It should be noted that this integral trimming is not used in the GFCCSD and GFCC-i(2,3) calculations.

B. Quasiparticle properties with the DZVP and cc-pVDZ basis sets

Tables I and II summarize the core binding energies for the different systems and methods calculated with the DZVP and cc-pVDZ basis sets, respectively. While the KT, DSE2, and EOM-CC results show little dependence on the basis set, for the GFCC methods the augmented Dunning basis set seems not able to improve the results. To reduce the discrepancy of the GFCC results, the employment of bare Dunning basis sets seems slightly better. For the GFCCSD results, employing the aug-cc-pVDZ basis gives the MAE of 4.24 eV, in comparison to the MAE of 3.57 eV brought by employing the bare cc-pVDZ basis. Furthermore, the triple-ζ cc-pVTZ basis can systematically reduce the discrepancies to even below 1 eV ($\sim$0.74 eV), which agrees with basis set discussion in the previous EOM-CC and GFCC results for the core ionizations of small molecules.[2–4]

Tables III and IV summarize the core quasiparticle strengths for the different systems and methods calculated with the DZVP and cc-pVDZ basis sets, respectively.

TABLE II: Comparison of the experimental core binding energies (in eV) to those obtained with the cc-pVDZ basis set, using the L and NL approximations to the cumulant and the L_c, Q and F approximations of the EOM-CCS method, and their mean absolute errors (MAE). The KT results are obtained with Koopmans' Theorem.

System	**KT**	**DSE2**	GFCCSD	GFCC-i(2,3)	$\mathbf{1}_{NL}$	$\mathbf{2}_{L}$	$\mathbf{3}_{L}$	$\mathbf{1}_{L}$	$\mathbf{2}_{NL}$	$\mathbf{3}_{NL}$	**Expt**	**Ref**
CH_4	305.17	292.56	292.80	293.45	286.98	287.84	287.44	290.54	291.08	290.83	290.703	[5]
NH_3	422.78	406.26	407.68	408.61	400.67	401.35	400.81	405.13	405.55	405.23	405.520	[6]
H_2O	559.25	539.30	542.21	543.41	534.53	534.74	534.15	539.32	539.46	539.10	539.700	[7]
HF	715.09	692.64	696.78	698.19	689.27	688.97	688.45	693.78	693.59	693.27	694.200	[8]
Ne	891.59	868.17	874.52	874.52	866.50	865.87	865.57	870.16	869.73	869.52	870.200	[9]
MAE	18.71	1.32	2.73	3.57	4.48	4.31	4.78	0.28	0.35	0.53		

TABLE III: Comparison of the quasiparticle strengths obtained with the DZVP basis set, using the L and NL approximations to the cumulant and the 1-3 approximations of the EOM-CCS method.

System	**DSE2**	$\mathbf{1}_{L}$	$\mathbf{2}_{L}$	$\mathbf{3}_{L}$	$\mathbf{1}_{NL}$	$\mathbf{2}_{NL}$	$\mathbf{3}_{NL}$
CH_4	0.79	0.60	0.61	0.59	0.70	0.71	0.69
NH_3	0.77	0.60	0.61	0.58	0.71	0.71	0.69
H_2O	0.75	0.63	0.62	0.59	0.73	0.72	0.70
HF	0.76	0.68	0.66	0.64	0.76	0.74	0.72
Ne	0.78	0.76	0.73	0.72	0.80	0.78	0.77

C. Full comparison of the spectral function as a function of basis set, level of CCS approximation and cumulant form

Figures 4 and 5 shows a comparison of the core spectral function of the $10e$ systems computed with the DZVP and cc-pVDZ basis sets, respectively, and the 3_{NL} approach, as a function of EOM-CCS approximation.

TABLE IV: Comparison of the quasiparticle strengths obtained with the cc-pVDZ basis set, using the L and NL approximations to the cumulant and the 1-3 approximations of the EOM-CCS method.

System	**DSE2**	**GFCCSD**	**GFCC-i(2,3)**	$\mathbf{1}_L$	$\mathbf{2}_L$	$\mathbf{3}_L$	$\mathbf{1}_{NL}$	$\mathbf{2}_{NL}$	$\mathbf{3}_{NL}$
CH_4	0.80	0.77	0.81	0.60	0.63	0.61	0.70	0.72	0.71
NH_3	0.78	0.76	0.81	0.61	0.63	0.61	0.71	0.73	0.71
H_2O	0.78	0.71	0.82	0.64	0.65	0.63	0.74	0.74	0.72
HF	0.79	0.79	0.84	0.70	0.69	0.67	0.77	0.76	0.75
Ne	0.81	1.00	1.00	0.77	0.76	0.75	0.82	0.81	0.80

TABLE V: Scissors corrections used in Fig. 6.

System	$\mathbf{1}_{NL}$	$\mathbf{2}_{NL}$
CH_4	1.9	1.1
NH_3	6.2	1.6
H_2O	8.8	1.8
HF	10.4	2.1
Ne	10.2	1.5

D. Effect of the CCS approach on the gap between the quasiparticle and the satellites

Figure 6 shows a comparison of the spectral functions with a scissors correction applied to the 1_{NL} and 2_{NL} approaches in such a way that the satellite regions become aligned with those in the 3_{NL} approach. After the correction is applied, the 2_{NL} approximation is shown to give results that are almost identical to the full approach. Despite showing some noticeable differences, 1_{NL} approximation shows reasonable agreement in the overall distribution of the satellite intensity. The scissors corrections for each system and method are shown in Table V. Interestingly, while the corrections required for the 1_{NL} approximation are clearly system-dependent, in the case of the 2_{NL} approximation the corrections are almost constant, suggesting that an overall scissors correction can be used to simulate the results of the more expensive full approach.

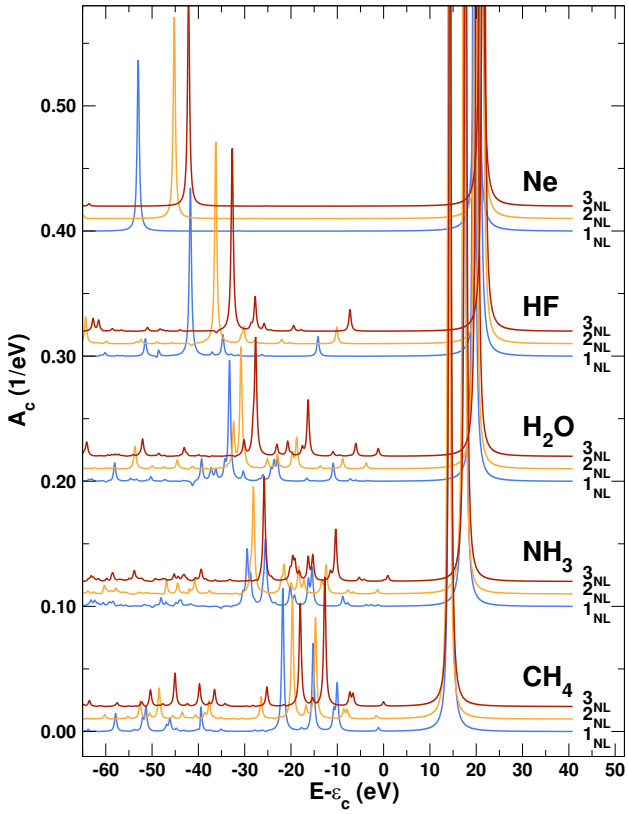

FIG. 4: Core spectral function A_c of the 10e systems computed with the DZVP basis set, as a function of EOM-CCS approximation 1_{NL} (blue), 2_{NL} (orange), and 3_{NL} (red).

[1] B. Peng and K. Kowalski, Phys. Rev. A **94**, 062512 (2016).

[2] S. Sen, A. Shee, and D. Mukherjee, Mol. Phys. **111**, 2625 (2013).

[3] S. Coriani and H. Koch, J. Chem. Phys. **145**, 149901 (2016).

[4] B. Peng and K. Kowalski, J. Chem. Theory Comput. **14**, 4335 (2018), pMID: 29957945.

[5] T. Karlsen, K. J. Børve, L. J. Sæthre, K. Wiesner, M. Bässler, and S. Svensson, Journal of the American Chemical Society **124**, 7866 (2002).

[6] T. Buttersack, P. E. Mason, R. S. McMullen, T. Martinek, K. Brezina, D. Hein, H. Ali, C. Kolbeck, C. Schewe, S. Malerz, et al., Journal of the American Chemical Society **141**, 1838 (2019).

[7] F. Viñes, C. Sousa, and F. Illas, Phys. Chem. Chem. Phys. **20**, 8403 (2018).

[8] Atomic Data and Nuclear Data Tables **31**, 433 (1984), ISSN 0092-640X.

[9] G. P. Williams, *X-Ray Data Booklet: Section 1.1 ELECTRON BINDING ENERGIES* (Lawrence

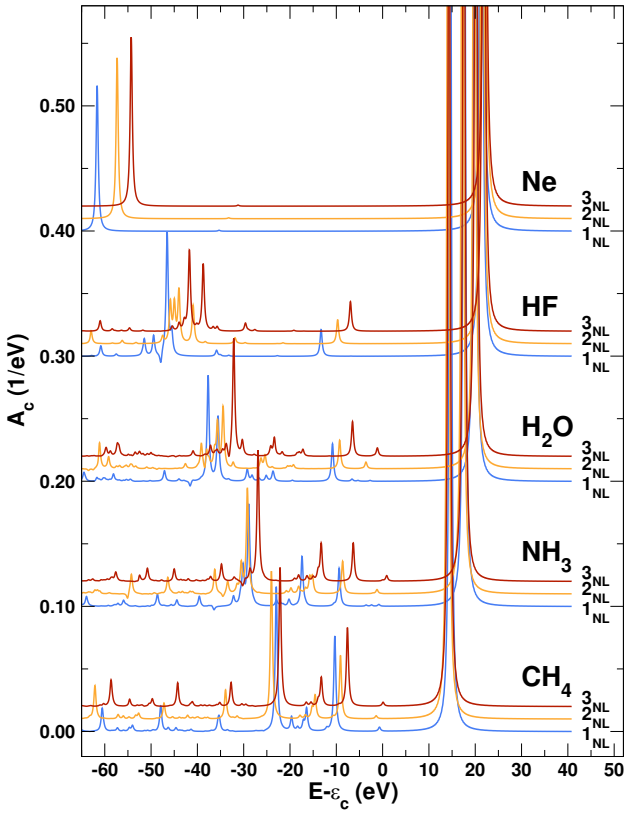

FIG. 5: Core spectral function A_c of the 10e systems computed with the cc-pVDZ basis set, as a function of EOM-CCS approximation 1_{NL} (blue), 2_{NL} (orange), and 3_{NL} (red).

Berkeley National Laboratory, Berkeley, 2009).

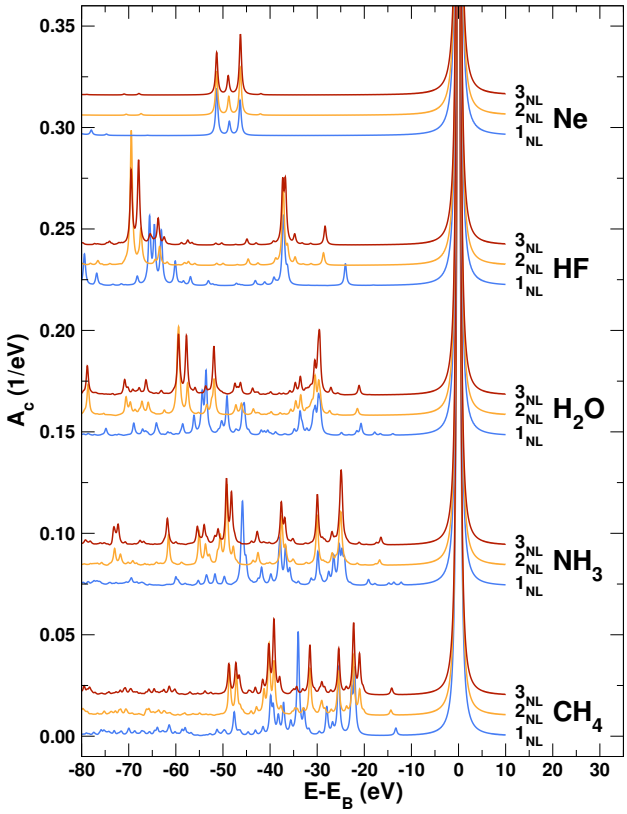

FIG. 6: Comparison of the spectral functions with a scissors correction (see Table V) applied to the 1_{NL} (blue) and 2_{NL} (orange) approximations in such a way that the satellite regions become aligned with those of 3_{NL} (red).