

# Full-frequency GW without frequency

Sylvia J. Bintrim<sup>1</sup> and Timothy C. Berkelbach<sup>1,2, a)</sup>

<sup>1)</sup>Department of Chemistry, Columbia University, New York, New York 10027, USA

<sup>2)</sup>Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

Efficient computer implementations of the GW approximation must approximate a numerically challenging frequency integral; the integral can be performed analytically, but doing so leads to an expensive implementation whose computational cost scales as  $O(N^6)$  where  $N$  is the size of the system. Here we introduce a new formulation of the full-frequency GW approximation by exactly recasting it as an eigenvalue problem in an expanded space. This new formulation (1) avoids the use of time or frequency grids, (2) naturally precludes the common “diagonal” approximation, (3) enables common iterative eigensolvers that reduce the canonical scaling to  $O(N^5)$ , and (4) enables a density-fitted implementation that reduces the scaling to  $O(N^4)$ . We numerically verify these scaling behaviors and test a variety of approximations that are motivated by this new formulation. In this new formulation, the relation of the GW approximation to configuration interaction, coupled-cluster theory, and the algebraic diagrammatic construction is made especially apparent, providing a new direction for improvements to the GW approximation.

Green’s function approaches based on time-dependent many-body perturbation theory provide an economical description of excitation energies and spectral intensities. For the one-particle Green’s function, which describes electron addition and removal processes, the GW approximation to the self-energy<sup>1</sup> performs well for weakly correlated insulators and metals,<sup>2</sup> which has partially motivated its application to molecules<sup>3–6</sup> (here and throughout we are considering the common non-self-consistent  $G_0W_0$  approximation, which we call the GW approximation for simplicity). The size of systems that can be studied with the GW approximation is determined by the implementation, which can be characterized by its asymptotic scaling with the system size  $N$ , ranging from  $O(N^3)$  to  $O(N^6)$  with widely varying prefactors.<sup>7,8</sup>

GW implementations can be distinguished based on their handling of a numerically challenging frequency integral, which is relatively uncommon in quantum chemical methods. The earliest works used a generalized plasmon pole model to approximate the dielectric function and thus integrate analytically.<sup>2,9,10</sup> More sophisticated approaches treat the full frequency dependence using numerical integration techniques such as analytic continuation<sup>11–16</sup> and contour deformation.<sup>15–19</sup> These latter methods introduce numerical errors, but ones that in principle can be eliminated with increasing cost (e.g. the frequency integration grid or the fitting of the self-energy on the imaginary frequency axis). The final class of methods are numerically exact within a given single-particle basis set and require the explicit enumeration of all neutral excitations energies,<sup>3–5,20,21</sup> typically calculated within the random-phase approximation (RPA). This explicit enumeration, i.e. a sum over states, dominates the cost of such a GW calculation due to its  $O(N^6)$  scaling. This exact handling of the full frequency dependence is the type that we address in the present work. We note that this class of methods still constructs a frequency-dependent self-energy, which is used to solve the quasiparticle equation for each excitation. In this work, we present a new formulation of the GW approxima-

tion by recasting it as an eigenvalue problem in an expanded space, and a *frequency variable never appears*.

Within Green’s function theories, charged excitation energies  $E_n$ , i.e. ionization potentials (IPs) and electron affinities (EAs), are found as the poles of the one-particle Green’s function matrix  $\mathbf{G}(\omega)$  via the eigenvalue problem

$$[\mathbf{G}^{-1}(\omega = E_n)] \mathbf{R}_n = [\mathbf{f} + \Sigma_c(\omega = E_n)] \mathbf{R}_n = E_n \mathbf{R}_n, \quad (1)$$

where  $\mathbf{f} = \mathbf{h} + \mathbf{J} + \mathbf{K}$  is the Fock matrix,  $\mathbf{h}$  is the kinetic and external potential energy matrix,  $\mathbf{J}$  is the Hartree matrix,  $\mathbf{K}$  is the exchange matrix, and  $\Sigma_c(\omega)$  is the correlation part of the self-energy matrix. In practice, we typically work in a basis of orbitals  $\phi_p(\mathbf{x})$  that diagonalize a mean-field Green’s function, which serves as the reference and defines the orbital energies  $\varepsilon_p$ . As usual, the  $O$  occupied orbitals will be indexed by  $i, j, k, l$ , the  $V$  unoccupied orbitals by  $a, b, c, d$ , and generic orbitals by  $p, q, r, s$ . For simplicity, we will assume real orbitals. In this basis, we have  $\mathbf{f} = \varepsilon + \mathbf{K} - \mathbf{V}_{xc}$  where  $\mathbf{V}_{xc}$  is the exchange-correlation potential matrix. Note that for a HF reference,  $\mathbf{K} - \mathbf{V}_{xc} = 0$ .

In the GW approximation, the self-energy is calculated to lowest-order in the screened Coulomb interaction  $W$ , which gives rise to the aforementioned frequency integral,  $\Sigma_c(\omega) = (i/2\pi) \int d\omega' e^{i\eta\omega'} G(\omega + \omega') W_p(\omega')$  where  $W_p = W - v$  is the polarized part of the screened Coulomb interaction. When the polarizability that enters  $W_p$  is expressible by a spectral representation, then the frequency integration can be performed analytically to yield<sup>3–5,20,21</sup>

$$[\Sigma_c(\omega)]_{pq} = \sum_v \left\{ \sum_j \left[ \frac{M_{pj}^\nu M_{qj}^\nu}{\omega - (\varepsilon_j - \Omega_v) - i\eta} \right] + \sum_b \left[ \frac{M_{bp}^\nu M_{bq}^\nu}{\omega - (\varepsilon_b + \Omega_v) + i\eta} \right] \right\} \quad (2)$$

where  $M_{pq}^\nu = \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_p(\mathbf{x}_1) \phi_q(\mathbf{x}_2) r_{12}^{-1} \rho_\nu(\mathbf{x}_2)$ ,  $\Omega_v$  are neutral excitation energies, and  $\rho_\nu(\mathbf{x})$  are transition densities.

Although any theory of neutral excitations can be used to calculate the polarizability,<sup>22</sup> here we consider the Tamm-Dancoff approximation (TDA) to the (direct) RPA. Within

<sup>a)</sup>Electronic mail: [tim.berkelbach@gmail.com](mailto:tim.berkelbach@gmail.com)

the TDA, the neutral excitation energies and transition density moments are defined by  $\mathbf{AX}^v = \Omega_v \mathbf{X}^v$ , where

$$A_{ia,jb} = (\varepsilon_a - \varepsilon_i)\delta_{ab}\delta_{ij} + \langle ib|aj \rangle, \quad (3)$$

$\langle pq|rs \rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_p(\mathbf{x}_1)\phi_q(\mathbf{x}_2)r_{12}^{-1}\phi_r(\mathbf{x}_1)\phi_s(\mathbf{x}_2)$ ,  $\rho_v(\mathbf{x}) = \sum_{ia} X_{ia}^v \phi_i(\mathbf{x})\phi_a(\mathbf{x})$ , and  $M'_{pq} = \sum_{ia} X_{ia}^v \langle pi|qa \rangle$ . Diagrammatically, such a self-energy has screening due to infinite-order, *forward time-ordered* ring (or bubble) diagrams. The algebraic form Eq. (2) assumes that *all* eigenvalues and eigenvectors of the  $\mathbf{A}$  matrix have been calculated, which implies a canonical  $O(N^6)$  scaling, as discussed in the introduction.

In the GW community, RPA screening is much more commonly implemented without the TDA. Although the frequency-free implementation of the GW approximation that we present here is far simpler to formulate within the TDA, many of the same ideas can be applied for the case of RPA screening, which we discuss in the supplementary material. In particular, we show that a frequency-free formulation with RPA screening exists; however, it is less conducive to cost reductions. Moreover, in Fig. 1, we show that results obtained with TDA screening are of similar accuracy to those obtained with RPA screening, especially when based on a HF reference, empirically justifying our focus on TDA screening.

In order to make progress on a frequency-free implementation that avoids the explicit sum over states in Eq. (2), we define a vector space of excitations corresponding to one hole (1h), one particle (1p), two holes and one particle (2h1p), and two particles and one hole (2p1h). A vector in this space has elements  $\mathbf{R} = (r_i, r_a, r_{i|jb}, r_{[jb]a})$ . The notation of the 2h1p and 2p1h amplitudes indicates that the  $j \rightarrow b$  excitation is independent of the other particle or hole index, i.e. the amplitudes do not obey any antisymmetry as they do in determinantal approaches. We define a frequency-independent “super-

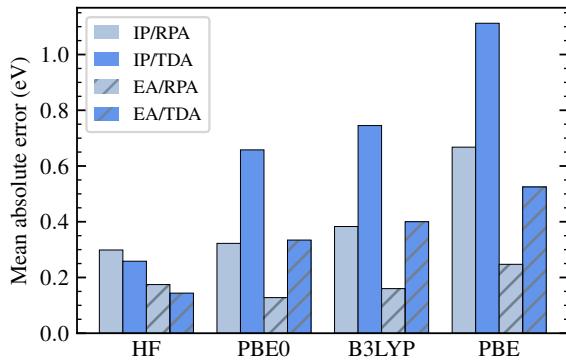


FIG. 1. Mean absolute errors in the first IP and first EA calculated using GW with TDA and RPA screening for various mean-field references. Using the  $O(N^6)$  implementation in PySCF,<sup>23,24</sup> we performed calculations on the smallest 91 molecules in the GW100 test set<sup>6</sup> in the def2-TZVPP basis.<sup>25</sup> Error is calculated with respect to to ACCSD(T) for the IP<sup>26</sup> and EOM-CCSD for the EA.<sup>27</sup>

matrix”  $\mathbf{H}$ ,

$$\mathbf{H} = \begin{pmatrix} \mathbf{f} & \mathbf{V}^{2h1p} & \mathbf{V}^{2p1h} \\ (\mathbf{V}^{2h1p})^\dagger & \mathbf{C}^{2h1p} & \mathbf{0} \\ (\mathbf{V}^{2p1h})^\dagger & \mathbf{0} & \mathbf{C}^{2p1h} \end{pmatrix}, \quad (4)$$

where  $\mathbf{C}^{2h1p} = \varepsilon^{1h} \oplus (-\mathbf{A})$  and  $\mathbf{C}^{2p1h} = \varepsilon^{1p} \oplus \mathbf{A}$  with matrix elements

$$V_{p,k|lc}^{2h1p} = \langle pc|kl \rangle \quad (5a)$$

$$V_{p,[kc]d}^{2p1h} = \langle pk|dc \rangle \quad (5b)$$

$$C_{i[ja],k[lc]}^{2h1p} = [(\varepsilon_i + \varepsilon_j - \varepsilon_a)\delta_{jl}\delta_{ac} - \langle jc|al \rangle] \delta_{ik} \quad (5c)$$

$$C_{[ia]b,[kc]d}^{2p1h} = [(\varepsilon_a + \varepsilon_b - \varepsilon_i)\delta_{ik}\delta_{ac} + \langle ak|ic \rangle] \delta_{bd}. \quad (5d)$$

This super-matrix can be downfolded into the 1h+1p space, leading to a frequency-dependent eigenvalue problem of the form Eq. (1), with

$$\Sigma(\omega) = \mathbf{V}^{2h1p} \left[ \omega \mathbf{1} - \mathbf{C}^{2h1p} \right]^{-1} [\mathbf{V}^{2h1p}]^\dagger + \mathbf{V}^{2p1h} \left[ \omega \mathbf{1} - \mathbf{C}^{2p1h} \right]^{-1} [\mathbf{V}^{2p1h}]^\dagger. \quad (6)$$

It is straightforward to check that this frequency-dependent matrix, arising from the downfolding of the 2h1p and 2p1h spaces, is precisely the correlation part of the GW self-energy. The above presentation closely follows the algebraic diagrammatic construction (ADC) literature.<sup>28,29</sup> In particular, the above theory, i.e. the GW approximation with TDA screening, is a strict but severe approximation to the ADC(3) and 2p1h-TDA methods.<sup>30</sup> Diagrammatically, the latter two theories include many vertex corrections beyond the GW approximation, including ladder and exchange diagrams. An analogous approach was also used recently to formulate an efficient renormalized second-order Green’s function theory<sup>31</sup> and similar conceptual ideas were discussed in the context of double excitations in time-dependent density functional theory.<sup>32</sup>

Importantly, the frequency-independent super-matrix form of the GW approximation enables the use of iterative eigensolvers that lower the computational scaling. Matrix-vector multiplication is given by  $\mathbf{H}\mathbf{R} = \sigma$ , with

$$\sigma_i = \sum_j f_{ij}r_j + \sum_b f_{ib}r_b + \sum_{klc} \langle ic|kl \rangle r_{k[lc]} + \sum_{kcd} \langle ik|dc \rangle r_{[kc]d} \quad (7a)$$

$$\sigma_a = \sum_j f_{aj}r_j + \sum_b f_{ab}r_b + \sum_{klc} \langle ac|kl \rangle r_{k[lc]} + \sum_{kcd} \langle ak|dc \rangle r_{[kc]d} \quad (7b)$$

$$\sigma_{i[ja]} = \sum_k \langle ka|ij \rangle r_k + \sum_b \langle ba|ij \rangle r_b + (\varepsilon_i + \varepsilon_j - \varepsilon_a)r_{i[ja]} - \sum_{lc} \langle jc|al \rangle r_{i[lc]} \quad (7c)$$

$$\begin{aligned} \sigma_{[ia]b} &= \sum_j \langle ji | ba \rangle r_j + \sum_c \langle ci | ba \rangle r_c \\ &+ (\varepsilon_a + \varepsilon_b - \varepsilon_i) r_{[ia]b} + \sum_{kc} \langle ak | ic \rangle r_{[kc]b}, \end{aligned} \quad (7d)$$

where all indices correspond to spin-orbitals. For a restricted, closed-shell reference, spin-free equations are straightforward to derive and are given in the supplementary material. Clearly, the above equations have no worse than  $O(N^5)$  scaling (specifically  $O^2 V^3$  for moderately sized basis sets), which is a significant improvement over the  $O(N^6)$  scaling exhibited by the sum-over-states implementation. Furthermore, because only Coulomb-type electron repulsion integrals are used in the direct TDA (or RPA), the scaling of the most expensive contractions can be easily reduced by density-fitting. For example, if the ERIs are approximated as  $\langle pq | rs \rangle \approx \sum_Q B_{pq}^Q B_{rs}^Q$  then the worst-scaling  $O(N^5)$  term can be calculated by

$$\sigma_{[ia]b} = \sum_Q \sum_{kc} B_{ai}^Q B_{kc}^Q r_{[kc]b}, \quad (8)$$

which has two steps that scale as  $O(N_{\text{aux}} OV^2)$  or  $O(N^4)$ .

Although we will not show results here, we briefly describe how spectral quantities can also be obtained iteratively with identical scalings. Using a spectral resolution of  $\mathbf{H}$ , the full Green's function is given by  $G_{pq}(\omega) = \sum_n (r_p^n r_q^n) / (\omega - E_n)$ , i.e. the quasiparticle weight is given simply in terms of the 1p+1h elements of the solution vector  $\mathbf{R}$ ; this formulation naturally precludes the common diagonal approximation  $\Sigma_{pq}(\omega) \approx \delta_{pq} \Sigma_{pp}(\omega)$ . The matrix  $\mathbf{H}$  can also be used iteratively (without diagonalization) to calculate the frequency-dependent self-energy  $\Sigma(\omega) = \mathbf{V}^{2\text{h}1\text{p}} \mathbf{Z}^{2\text{h}1\text{p}}(\omega) + \mathbf{V}^{2\text{p}1\text{h}} \mathbf{Z}^{2\text{p}1\text{h}}(\omega)$  where  $\mathbf{Z}(\omega)$  is a matrix that solves the linear systems of equations, e.g.  $[\omega \mathbf{1}^{2\text{h}1\text{p}} - \mathbf{C}^{2\text{h}1\text{p}}] \mathbf{Z}^{2\text{p}1\text{h}}(\omega) = [\mathbf{V}^{2\text{h}1\text{p}}]^\dagger$ , which can be solved with iterative methods such as conjugate gradient or the generalized minimum residual method. Similarly, the Green's function can be calculated as  $\mathbf{G}(\omega) = \mathbf{P} \mathbf{Z}(\omega)$  where  $\mathbf{Z}(\omega)$  solves  $[\omega \mathbf{1} - \mathbf{H}] \mathbf{Z}(\omega) = \mathbf{P}^\dagger$  and  $\mathbf{P}$  is a matrix that projects onto the 1p+1h space.

We have implemented the GW techniques described above in the PySCF software package.<sup>23,24</sup> To compare their costs and verify their asymptotic scaling, we have calculated the first IP of a series of linear alkanes in the def2-SVP basis<sup>25</sup> up to  $\text{C}_{37}\text{H}_{76}$ , which has 898 basis functions. The execution timings of the  $O(N^6)$  sum-over-states,  $O(N^5)$  frequency-free, and  $O(N^4)$  density-fitted frequency-free implementations are shown in Fig. 2; all calculations were performed on a single core of an Intel Xeon Gold 6126 2.6 GHz (Skylake) CPU and density-fitted calculations used the def2-SVP-JKFIT auxiliary basis set.<sup>33</sup> As can be seen, all methods exhibit the expected asymptotic scaling. Comparing the absolute execution times of the sum-over-states and density-fitted frequency-free implementations, we obtained a speed-up of four orders of magnitude for the  $\text{C}_{10}\text{H}_{22}$  calculation. For our largest system with almost one thousand basis functions, the density-fitted implementation required only two hours on a single core, demonstrating the immense savings available with the advances described here. The use of density fitting was found to introduce a negligible error of around 0.01 eV.

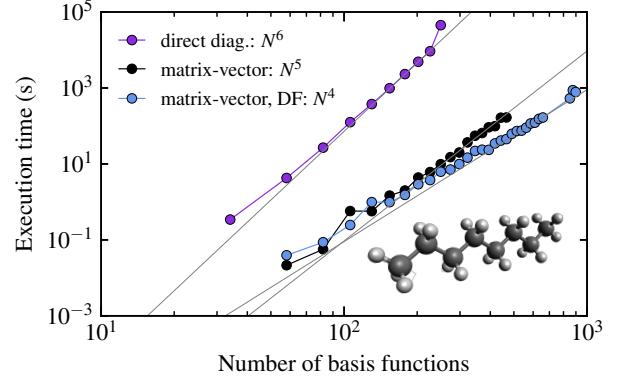


FIG. 2. Timings of the sum-over-states (direct diag.), frequency-free (matvec), and density-fitted frequency-free (matvec, DF) implementations of the GW approximation for a series of linear alkanes in the def2-SVP basis, up to  $\text{C}_{37}\text{H}_{76}$ . For the sum-over-states implementation, we timed the full diagonalization of the TDA matrix; for the frequency-free implementations, we timed the 6-8 matrix-vector multiplications required for convergence of the Davidson algorithm.

One challenge with using iterative eigensolvers on  $\mathbf{H}$  is that the eigenvalues of typical interest (valence ionization potentials and electron affinities) are interior eigenvalues. Therefore, they must be found using energy-targeting methods like shift-and-invert or ones which maximize eigenvector overlap with a given guess vector. We have found the latter to work well, in conjunction with Davidson diagonalization,<sup>34,35</sup> for valence IPs and EAs. However, two simple alternatives exist by introducing additional approximations.

In a first approach, one can make the diagonal approximation to the Green's function and the self-energy, seeking the self-consistent solution  $E_n$  of the algebraic equation  $f_{pp} + \Sigma_{pp}(\omega = E_n) = E_n$ . This can be solved by iterative diagonalization of a modified matrix  $\mathbf{H}^{(p)}$  which has deleted all 1p+1h rows and columns except that of orbital  $p$ . Although the principal eigenvalue of interest is still an interior eigenvalue, this approach eliminates all other quasiparticle energies, which can facilitate energy- or overlap-targeting procedures. In a second approach, one can perturbatively decouple the IP and EA parts of  $\mathbf{H}$ , which will make the valence IPs and EAs into extremal eigenvalues. For example, for the calculation of IPs, we perturbatively eliminate the 1p and 2p1h subspaces based on their lowest-order influence on the 1h subspace, and likewise for EAs. This leads to modified Fock matrix elements

$$F_{ii} = f_{ii} + \sum_b \frac{|f_{ib}|^2}{f_{ii} - f_{bb}} + \sum_{kcd} \frac{|\langle ik | dc \rangle|^2}{\varepsilon_i + \varepsilon_k - \varepsilon_c - \varepsilon_d} \quad (9a)$$

$$F_{aa} = f_{aa} + \sum_j \frac{|f_{aj}|^2}{f_{aa} - f_{jj}} + \sum_{klc} \frac{|\langle ac | kl \rangle|^2}{\varepsilon_a + \varepsilon_c - \varepsilon_k - \varepsilon_l} \quad (9b)$$

which are used in place of  $f_{ii}$  and  $f_{aa}$  during the matrix-vector product. This approach has the added benefit of reducing the size of the vector spaces for the IP and EA problems,  $\mathbf{R}^{\text{IP}} = (r_i, r_{i[jb]})$  and  $\mathbf{R}^{\text{EA}} = (r_a, r_{a[jb]})$ , and reducing the scal-

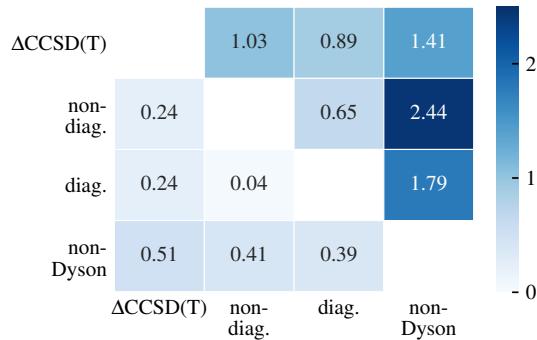


FIG. 3. Impact of various approximations on the first IP of the molecules in the GW100 test set,<sup>6</sup> using a HF reference. Mean absolute deviations are given in the lower triangle, and maximum absolute deviations are given in the upper triangle (all in eV).

ing of the IP matrix-vector product to be  $O(O^3V^2)$ . Finally, we point out that the valence IPs can be made into the *lowest* eigenvalues by negating the matrix,  $\mathbf{H}^{\text{IP}} \rightarrow -\mathbf{H}^{\text{IP}}$ . With all of these changes, the calculation of IPs and EAs within the GW approximation looks quite similar to that within the IP/EAs-EOM-CCSD approximation.<sup>36,37</sup> The decoupling of the IP and EA spaces is also common in the ADC literature, and referred to as a non-Dyson approach.<sup>38</sup> We will use the same language and evaluate the performance of the Dyson (coupled IP and EA) and non-Dyson (perturbatively decoupled IP and EA) GW approximation.

To assess the effect of the diagonal approximation and perturbative decoupling, we used our frequency-free  $O(N^5)$  GW implementation to calculate the first IP of all the molecules in the GW100 test set.<sup>6</sup> In Fig. 3, we compare the non-diagonal, diagonal, and non-diagonal perturbatively decoupled (“non-Dyson”) GW results among themselves and to ΔCCSD(T) results,<sup>26</sup> using a HF reference. As shown in Fig. 1, the GW results have a mean absolute error of 0.24 eV, with respect to CCSD(T). On average, the diagonal approximation has negligible effect (less than 0.1 eV), although a maximum deviation of 0.65 eV is observed, indicating the potential importance of off-diagonal elements of the self-energy for some molecules. Perturbative decoupling (the non-Dyson GW approximation) changes the results by 0.41 eV on average (and by as much as 2.44 eV) and increases the mean absolute error from 0.24 eV to 0.51 eV, suggesting that it is a relatively severe approximation. We have performed the same analysis (not shown) for the EA, as well as for a PBE reference; all results are qualitatively similar.

To summarize, we have shown that the typical Dyson equation formulation of the GW approximation can be exactly reformulated as a frequency-independent eigenvalue problem in an expanded space. In addition to providing a new conceptual framework for the GW approximation and related Green’s function theories, the new formulation was used to reduce the computational scaling from  $O(N^6)$  to  $O(N^4)$ . Based on our preliminary results, we expect that this frequency-free formulation of the GW approximation will be readily applicable to systems with hundreds or thousands of atoms, likely limited

by the memory needed to store three-index quantities.

We anticipate that this eigenvalue formulation of the GW approximation will lead to new methodological developments inspired by quantum chemical methods with similar structure. For example, we are exploring the use of partitioning schemes to mitigate the cost of large basis sets,<sup>39-41</sup> the introduction of vertex corrections through the algebraic diagrammatic construction,<sup>28,29</sup> the use of renormalization and compression to perform self-consistent GW,<sup>31</sup> and the extension towards strongly correlated systems with multi-reference techniques.<sup>42,43</sup> Finally, the ideas presented in this work and these latter extensions can be applied to treat the frequency dependence of the Bethe-Salpeter equation for neutral excitations.<sup>44,45</sup>

## SUPPLEMENTARY MATERIAL

See the supplementary material for spin-free matrix-vector product equations and for derivation of the frequency-free GW approximation with RPA screening.

## ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1644869 (S.J.B.) and by the National Science Foundation under Grant No. CHE-1848369 (T.C.B.). We acknowledge computing resources from Columbia University’s Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded April 15, 2010. The Flatiron Institute is a division of the Simons Foundation.

## DATA AVAILABILITY STATEMENT

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

- <sup>1</sup>L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- <sup>2</sup>M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* **34**, 5390 (1986).
- <sup>3</sup>M. L. Tiago and J. R. Chelikowsky, *Phys. Rev. B* **73**, 205334 (2006).
- <sup>4</sup>F. Bruneval, *J. Chem. Phys.* **136**, 194107 (2012).
- <sup>5</sup>M. J. van Setten, F. Weigend, and F. Evers, *J. Chem. Theory Comput.* **9**, 232 (2013).
- <sup>6</sup>M. J. van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, *J. Chem. Theory Comput.* **11**, 5665 (2015).
- <sup>7</sup>D. Foerster, P. Koval, and D. Sánchez-Portal, *J. Chem. Phys.* **135**, 074105 (2011).
- <sup>8</sup>D. Golze, M. Dvorak, and P. Rinke, *Front. Chem.* **7**, 377 (2019).
- <sup>9</sup>R. W. Godby and R. J. Needs, *Phys. Rev. Lett.* **62**, 1169 (1989).
- <sup>10</sup>P. Larson, M. Dvorak, and Z. Wu, *Phys. Rev. B* **88**, 125205 (2013).
- <sup>11</sup>M. M. Rieger, L. Steinbeck, I. White, H. Rojas, and R. Godby, *Comput. Phys. Commun.* **117**, 211 (1999).
- <sup>12</sup>F. Giustino, M. L. Cohen, and S. G. Louie, *Phys. Rev. B* **81**, 115105 (2010).

<sup>13</sup>X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, *New J. Phys.* **14**, 053020 (2012).

<sup>14</sup>J. Wilhelm, M. Del Ben, and J. Hutter, *J. Chem. Theory Comput.* **12**, 3623 (2016).

<sup>15</sup>D. Golze, J. Wilhelm, M. J. van Setten, and P. Rinke, *J. Chem. Theory Comput.* **14**, 4856 (2018).

<sup>16</sup>T. Zhu and G. K.-L. Chan, ArXiv200703148 Cond-Mat Physicsphysics (2020), arXiv:2007.03148 [cond-mat, physics:physics].

<sup>17</sup>R. W. Godby, M. Schlüter, and L. J. Sham, *Phys. Rev. B* **37**, 10159 (1988).

<sup>18</sup>S. Lebègue, B. Arnaud, M. Alouani, and P. E. Bloechl, *Phys. Rev. B* **67**, 155208 (2003).

<sup>19</sup>M. Govoni and G. Galli, *J. Chem. Theory Comput.* **11**, 2680 (2015).

<sup>20</sup>L. Hedin, *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* **308**, 169 (1991).

<sup>21</sup>F. Bruneval, T. Rangel, S. M. Hamed, M. Shao, C. Yang, and J. B. Neaton, *Comput. Phys. Commun.* **208**, 149 (2016).

<sup>22</sup>A. M. Lewis and T. C. Berkelbach, *J. Chem. Theory Comput.* **15**, 2925 (2019).

<sup>23</sup>Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1340 (2018).

<sup>24</sup>Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, *J. Chem. Phys.* **153**, 024109 (2020).

<sup>25</sup>F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).

<sup>26</sup>K. Krause, M. E. Harding, and W. Klopper, *Mol. Phys.* **113**, 1952 (2015).

<sup>27</sup>M. F. Lange and T. C. Berkelbach, *J. Chem. Theory Comput.* **14**, 4224 (2018).

<sup>28</sup>J. Schirmer, L. S. Cederbaum, and O. Walter, *Phys. Rev. A* **28**, 1237 (1983).

<sup>29</sup>W. von Niessen, J. Schirmer, and L. S. Cederbaum, *Comput. Phys. Rep.* **1**, 57 (1984).

<sup>30</sup>J. Schirmer and L. S. Cederbaum, *J. Phys. B: At. Mol. Phys.* **11**, 1889 (1978).

<sup>31</sup>O. J. Backhouse, M. Nusspickel, and G. H. Booth, *J. Chem. Theory Comput.* **16**, 1090 (2020).

<sup>32</sup>P. Romaniello, D. Sangalli, J. A. Berger, F. Sottile, L. G. Molinari, L. Reinig, and G. Onida, *The Journal of Chemical Physics* **130**, 044108 (2009).

<sup>33</sup>F. Weigend, *J. Comp. Chem.* **29**, 167 (2008).

<sup>34</sup>E. R. Davidson, *J. Comput. Phys.* **17**, 87 (1975).

<sup>35</sup>A. R. Tackett and M. D. Ventra, *Phys. Rev. B* **66**, 245104 (2002).

<sup>36</sup>J. F. Stanton and J. Gauss, *J. Chem. Phys.* **101**, 8938 (1994).

<sup>37</sup>A. I. Krylov, *Annu. Rev. Phys. Chem.* **59**, 433 (2008).

<sup>38</sup>J. Schirmer, A. B. Trofimov, and G. Stelter, *J. Chem. Phys.* **109**, 4734 (1998).

<sup>39</sup>M. Nooijen and J. G. Snijders, *J. Chem. Phys.* **102**, 1681 (1995).

<sup>40</sup>J. F. Stanton and J. Gauss, *J. Chem. Phys.* **103**, 1064 (1995).

<sup>41</sup>M. F. Lange and T. C. Berkelbach, *Mol. Phys.* **119**, e1808726 (2020).

<sup>42</sup>A. Y. Sokolov, *J. Chem. Phys.* **149**, 204113 (2018).

<sup>43</sup>K. Chatterjee and A. Y. Sokolov, *J. Chem. Theory Comput.* **15**, 5908 (2019).

<sup>44</sup>J. C. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, *Phys. Rev. Lett.* **86**, 472 (2001).

<sup>45</sup>F. Bruneval, S. M. Hamed, and J. B. Neaton, *J. Chem. Phys.* **142**, 244101 (2015).

# Supplementary material: Full-frequency GW without frequency

Sylvia J. Bintrim<sup>1</sup> and Timothy C. Berkelbach<sup>1,2, a)</sup>

<sup>1</sup>Department of Chemistry, Columbia University, New York, New York 10027, USA

<sup>2</sup>Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

## I. SPIN-FREE EQUATIONS

For a restricted, closed-shell reference system, the spin-free matrix-vector equations for the GW approximation with TDA screening are

$$\begin{aligned}\sigma_i &= \sum_j f_{ij} r_j + \sum_b f_{ib} r_b \\ &+ 2 \sum_{klc} \langle ic|kl \rangle r_{k[lc]} + 2 \sum_{kcd} \langle ik|dc \rangle r_{[kc]d}\end{aligned}\quad (\text{S1a})$$

$$\begin{aligned}\sigma_a &= \sum_j f_{aj} r_j + \sum_b f_{ab} r_b \\ &+ 2 \sum_{klc} \langle ac|kl \rangle r_{k[lc]} + 2 \sum_{kcd} \langle ak|dc \rangle r_{[kc]d}\end{aligned}\quad (\text{S1b})$$

$$\begin{aligned}\sigma_{i[ja]} &= \sum_k \langle ka|ij \rangle r_k + \sum_b \langle ba|ij \rangle r_b \\ &+ (\varepsilon_i + \varepsilon_j - \varepsilon_a) r_{i[ja]} - 2 \sum_{lc} \langle jc|al \rangle r_{i[lc]}\end{aligned}\quad (\text{S1c})$$

$$\begin{aligned}\sigma_{[ia]b} &= \sum_j \langle ji|ba \rangle r_j + \sum_c \langle ci|ba \rangle r_c \\ &+ (\varepsilon_a + \varepsilon_b - \varepsilon_i) r_{[ia]b} + 2 \sum_{kc} \langle ak|ic \rangle r_{[kc]b}.\end{aligned}\quad (\text{S1d})$$

## II. RPA SCREENING

As discussed in the main text, the GW approximation is more commonly implemented without the Tamm-Dancoff approximation (TDA) to the random-phase approximation (RPA). Here we present the frequency-free version of the GW approximation with RPA screening. We show that a frequency-free formulation exists by recasting the GW approximation as an eigenvalue problem in an expanded space; however, the use of RPA screening precludes the cost reductions discussed in the main text.

The direct RPA leads to the generalized eigenvalue equation

$$\mathbf{M} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} = \mathbf{N} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}_+ & 0 \\ 0 & -\boldsymbol{\Omega}_+ \end{pmatrix} \quad (\text{S2})$$

where

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \quad (\text{S3a})$$

$$\mathbf{N} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \quad (\text{S3b})$$

<sup>a)</sup>Electronic mail: [tim.berkelbach@gmail.com](mailto:tim.berkelbach@gmail.com)

$$A_{ia,jb} = (\varepsilon_a - \varepsilon_i) \delta_{ab} \delta_{ij} + \langle ib|aj \rangle, \quad (\text{S3c})$$

$$B_{ia,jb} = \langle ij|ab \rangle \quad (\text{S3d})$$

and  $\boldsymbol{\Omega}_+$  is a diagonal matrix of positive excitation energies. The correlation part of the GW self-energy is

$$\begin{aligned}[\Sigma_c(\omega)]_{pq} &= \sum_{\nu}^{>0} \left\{ \sum_j \left[ \frac{M_{pj}^{\nu} M_{qj}^{\nu}}{\omega - (\varepsilon_j - \Omega_{\nu}) - i\eta} \right] \right. \\ &\quad \left. + \sum_b \left[ \frac{M_{bp}^{\nu} M_{bq}^{\nu}}{\omega - (\varepsilon_b + \Omega_{\nu}) + i\eta} \right] \right\}.\end{aligned}\quad (\text{S4})$$

where

$$\rho_{\nu}(\mathbf{x}) = \sum_{ai} (X_{ia}^{\nu} + Y_{ia}^{\nu}) \phi_i(\mathbf{x}) \phi_a(\mathbf{x}), \quad (\text{S5})$$

$$M_{pq}^{\nu} = \sum_{ia} (X_{ia}^{\nu} + Y_{ia}^{\nu}) \langle pi|qa \rangle, \quad (\text{S6})$$

$$\delta_{\mu\nu} = \sum_{ia} (X_{ia}^{\mu} X_{ia}^{\nu} - Y_{ia}^{\mu} Y_{ia}^{\nu}) \quad (\text{S7})$$

and we emphasize that only the positive-energy excitations are included in the summation.

The GW approximation with RPA screening can be implemented in a manner that is similar to that presented in the main text, with a few changes. First, the trial vector amplitudes must be expanded to include the deexcitation space of the RPA matrix,

$$\mathbf{R} = (r_i, r_a, r_{i[jb]}, \bar{r}_{i[jb]}, r_{[jb]a}, \bar{r}_{[jb]a}) \quad (\text{S8})$$

Second, because the GW self-energy expression (S4) only includes positive excitation energies, we introduce a matrix modification that eliminates the influence of the negative excitation energies,  $\mathbf{M} \rightarrow \tilde{\mathbf{M}}$  with

$$\tilde{\mathbf{M}} = \mathbf{M} + \eta \mathbf{N} \Theta(-\mathbf{NM}) \quad (\text{S9})$$

where  $\eta$  is a large number and  $\Theta$  is the step function. This gives a modified eigenvalue equation

$$\tilde{\mathbf{M}} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} = \mathbf{N} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}_+ & 0 \\ 0 & \eta - \boldsymbol{\Omega}_+ \end{pmatrix}, \quad (\text{S10})$$

i.e. the negative excitation energies are shifted to large positive values. Again we define a super-matrix,

$$\mathbf{H} = \begin{pmatrix} \mathbf{f} & \mathbf{V}^{2h1p} & \mathbf{V}^{2h1p} & \mathbf{V}^{2p1h} & \mathbf{V}^{2p1h} \\ (\mathbf{V}^{2h1p})^{\dagger} & \mathbf{C}^{2h1p} & & \mathbf{0} & \\ (\mathbf{V}^{2h1p})^{\dagger} & & \mathbf{0} & & \\ (\mathbf{V}^{2p1h})^{\dagger} & & & \mathbf{C}^{2p1h} & \\ (\mathbf{V}^{2p1h})^{\dagger} & & & & \end{pmatrix}, \quad (\text{S11})$$

where  $\mathbf{C}^{2h1p} = \varepsilon^{1h} \oplus (-\tilde{\mathbf{M}})$  and  $\mathbf{C}^{2p1h} = \varepsilon^{1p} \oplus \tilde{\mathbf{M}}$ , which requires a super-metric

$$\mathcal{N} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \oplus \mathcal{N} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \oplus \mathcal{N} \end{pmatrix}. \quad (\text{S12})$$

This leads to the generalized eigenvalue equation  $\mathbf{HR} = \mathcal{N}\mathbf{RE}$  or the non-Hermitian eigenvalue equation  $\mathcal{N}\mathbf{HR} = \mathbf{RE}$  with matrix-vector multiplication  $\mathcal{N}\mathbf{HR} = \boldsymbol{\sigma}$  given by

$$\begin{aligned} \sigma_i = & \sum_j f_{ij} r_j + \sum_b f_{ib} r_b + \sum_{klc} \langle ic|kl \rangle r_{k[lc]} + \sum_{kcd} \langle ik|dc \rangle r_{[kc]d} \\ & + \sum_{klc} \langle ic|kl \rangle \bar{r}_{k[lc]} + \sum_{kcd} \langle ik|dc \rangle \bar{r}_{[kc]d} \end{aligned} \quad (\text{S13a})$$

$$\begin{aligned} \sigma_a = & \sum_j f_{aj} r_j + \sum_b f_{ab} r_b + \sum_{klc} \langle ac|kl \rangle r_{k[lc]} + \sum_{kcd} \langle ak|dc \rangle r_{[kc]d} \\ & + \sum_{klc} \langle ac|kl \rangle \bar{r}_{k[lc]} + \sum_{kcd} \langle ak|dc \rangle \bar{r}_{[kc]d} \end{aligned} \quad (\text{S13b})$$

$$\begin{aligned} \sigma_{i[ja]} = & \sum_k \langle ka|ij \rangle r_k + \sum_b \langle ba|ij \rangle r_b \\ & + \varepsilon_i r_{i[ja]} - \sum_{kb} [\mathbf{NM}]_{ja,kb}^{\text{xx}} r_{i[kb]} - \sum_{kb} [\mathbf{NM}]_{ja,kb}^{\text{xd}} \bar{r}_{i[kb]} \end{aligned} \quad (\text{S13c})$$

$$\begin{aligned} \bar{\sigma}_{i[ja]} = & - \sum_k \langle ka|ij \rangle r_k - \sum_b \langle ba|ij \rangle r_b \\ & + \varepsilon_i \bar{r}_{i[ja]} - \sum_{kb} [\mathbf{NM}]_{ja,kb}^{\text{dx}} r_{i[kb]} - \sum_{kb} [\mathbf{NM}]_{ja,kb}^{\text{dd}} \bar{r}_{i[kb]} \end{aligned} \quad (\text{S13d})$$

$$\begin{aligned} \sigma_{[ia]b} = & \sum_j \langle ji|ba \rangle r_j + \sum_c \langle ci|ba \rangle r_c \\ & + \varepsilon_b r_{[ia]b} + \sum_{jc} [\mathbf{NM}]_{ia,jc}^{\text{xx}} r_{[jc]b} + \sum_{jc} [\mathbf{NM}]_{ia,jc}^{\text{xd}} \bar{r}_{[jc]b} \end{aligned} \quad (\text{S13e})$$

$$\begin{aligned} \bar{\sigma}_{[ia]b} = & - \sum_j \langle ji|ba \rangle r_j - \sum_c \langle ci|ba \rangle r_c \\ & + \varepsilon_b \bar{r}_{[ia]b} + \sum_{jc} [\mathbf{NM}]_{ia,jc}^{\text{dx}} r_{[jc]b} + \sum_{jc} [\mathbf{NM}]_{ia,jc}^{\text{dd}} \bar{r}_{[jc]b}. \end{aligned} \quad (\text{S13f})$$

In the above equations, we have broken up the RPA  $\mathbf{NM}$  matrix into its four blocks associated with the excitation (x) and deexcitation (d) spaces. An exact evaluation of the above equations can be done by diagonalizing the RPA matrix in order to calculate  $\tilde{M}$ ; this diagonalization has  $O(N^6)$  asymptotic scaling. However, we have implemented the above in order to verify that it yields eigenvalues equal to those of the GW approximation with RPA screening.

In an effort to reduce the scaling, we have attempted an iterative implementation of the product of  $\Theta(-\mathbf{NM})$  with a trial vector. In particular, we approximate this matrix-vector product using Arnoldi iteration, with requires only the Krylov space generated by matrix-vector products of  $\mathbf{NM}$  that can each be carried out with  $O(N^5)$  cost or with  $O(N^4)$  cost if density fitting is used. Although this approach was found to work for small systems, we encountered numerical issues on larger systems. For example, the number of Arnoldi iterations needed to approximate the matrix-vector product was prohibitively large and the Davidson algorithm exhibited convergence issues. We consider an iterative, low-scaling, frequency-free implementation of the GW approximation with RPA screening to be an interesting area for future work.