I. INTRODUCTION

Accurate calculation of optical spectra for large systems is essential for future novel optical and electronic devices, in fields ranging from photovoltaics,[1] photocatalysis,[2] organic semiconductors,[3, 4] to even the mechanisms of UV damage on DNA.[5] The ubiquitous method of the field is time dependent density functional theory (TD-DFT), but unfortunately this method lacks the accuracy needed for predictive power in most electronically complex systems. The Bethe-Salpeter Equation (BSE) formalism is an increasingly popular alternative for calculating electronic spectra.[6] The success of the BSE is due to the proper inclusion of an effective long range exchange kernel, which solves the failures of TD-DFT in accurately describe charge transfer excitations and avoided crossings.[7, 8]

Current conventional methods for solving the BSE are substantially more computationally demanding than most implementations of TD-DFT due the explicit calculation of a large number of occupied and virtual electronic states, yielding a typical scaling of $O(n_o^4)$, for $n_o$ valence orbitals.[9][11] TD-DFT with local exchange functionals has a naive scaling of $O(n_o^5)$. However, progress in the field has reduced the scaling with techniques that only requires the occupied orbitals which implicitly interact with all conduction states, what we later term a “mixed representation”. [12][20] While very important for TD-DFT, the speedup gained from use of an approximate iterative method versus direct diagonalization is only worthwhile in BSE if solving for the absorption is the algorithmic bottleneck. To go beyond TD-DFT to BSE requires constructing the effective Coulombic interaction, $W$, the most computationally expensive step. To overcome this issue, we adopt our previous stable iterative methods[14][15], and pull from our previous works in both TD-DFT/BSE[14][15][21][22] and stochastic GW[23][24] to present an efficient stochastic generation of $W$ within an iterative BSE technique. Our combined approach uses stochastic time-dependent propagation to obtain the action of $W$ on each required term in linear scaling.[23] Overall, this results in an efficient method with at most cubic scaling with respect to system size. The method and its application to a large organic semiconductor are detailed below.

II. METHOD

Overall, every method for solving the BSE has two numerical parts. Construction of the “kernels”, i.e., the action of the effective interaction $W$ on a given transition, and then diagonalizing or iterating the resulting Bethe-Salpeter Hamiltonian-like operator for the excitons. The full algorithm is covered here for completeness.

The starting point is a closed shell molecular system with $2N_{occ}$ electrons. We are interested in excitons composed of a mixture of $n_o$ occupied (valence) orbitals, $\phi_1, \phi_2, \ldots$, times $n_v$ conduction (virtual) ones written as $\phi_a, \phi_b, \ldots$. Typically $n_o \ll N_{occ}$ states are considered. The orbitals are eigenstates of a zero-order Hamiltonian $H_0$, with occupied-state eigenvalues $\varepsilon_1, \varepsilon_2, \ldots$ and virtual-state eigenvalues $\varepsilon_a, \varepsilon_b, \ldots$. Formally the zero-order Hamiltonian eigenvalues come from a very accurate method, specifically self-consistent GW. Additionally, we will use the well established Tamm-Dancoff approximation.

For singlet excitations, the excitation energies of the system are the eigenvalues of the $(n_o n_v \times n_v n_o)$ Tamm-Dancoff matrix $A$ that couples excitons, i.e., occupied-virtual pairs:

$$A(i; a; j; b) = (\varepsilon_a^{GW} - \varepsilon_j^{GW})\delta_{a,b}\delta_{i,j} + 2(i\langle a|b\rangle - (\phi_a|\phi_b)W|\phi_i\phi_j)$$

with exchange elements

$$i\langle a|b\rangle = \int \phi_i(r)\phi_a(r)\delta(r-r')\phi_b(r')drdr'$$

while $W \equiv W(\omega = 0)$ refers to the static effective Coulombic interaction approximation, and its matrix elements are

$$(\phi_a|\phi_b)W|\phi_i\phi_j) \equiv \int \phi_a(r)\phi_b(r)W(r,r')\phi_i(r')\phi_j(r')drdr'.$$

The superscript in $\varepsilon^{GW}$ denotes, as mentioned, that high quality $GW$ single particle energies should be used. In practice, and especially for medium sized systems
of a few dozen electrons, it is sufficient to use the DFT eigenstates plus a GW derived correction, called a scissors approximation. We calculate the HOMO and LUMO GW energies by the linear-scaling stochastic-GW (sGW) method [24–29] and use the scissors approximation: $\varepsilon_{\alpha}^{GW} \simeq \varepsilon_{\alpha} + \delta_{\alpha}$, $\varepsilon_{\alpha}^{GW} \simeq \varepsilon_{\alpha} + \delta_{\alpha}$, where $\delta_{\alpha} = \varepsilon_{\alpha}^{HOMO} - \varepsilon_{\alpha}^{HOMO}$ and analogously for $\delta_{\alpha}$. Further, for higher accuracy we use the self-consistent $\Delta GW_0$ approach where the sGW HOMO and LUMO corrections are a posteriori shifted self-consistently; for large systems this approach was found to be an excellent approximating to eigenvalue-only self-consistent $GW_0$ and to experiment. [27]

A. Mixed Representation Iterative Solution

The simplest derivation of an iterative method for the BSE spectrum starts with the linear-response time-dependent Hartree-Fock (TD-HF) equation. [15, 30, 31] For an initially real occupied state perturbed along the $x$ axis, $\psi_j(r,t) = e^{-i\omega t} \phi_j$, one performs a linear response expansion in $\omega$, $\psi_j(r,t) \simeq e^{-i\omega t}(\phi_j(r) - i\alpha f_j(r,t))$, where $f_j(r,t) = x\phi_j(r)$. The formally non-linear TD-HF equation for $\psi_j$ then converts, for small $\alpha$, to a linear equation for $f_j$. In the Tamm-Dancoff approximation, (where $f_j$ is not coupled to $f_j$) this evolution equation reads

$$i|f_j\rangle = A|f_j\rangle$$

(2)

where

$$A \equiv Q\tilde{A}$$

(3)

and

$$\tilde{A}|f_j\rangle = (H_0 + \Delta - \varepsilon_j)|f_j\rangle + (\delta v - \delta X)|\phi_j\rangle.$$  

(4)

Here we introduced several terms. $\Delta \equiv \Delta \phi_0$ is the $\Delta GW_0$ shift, the exciton Coulomb potential is $\delta v(r,t) = \int f(r - r')\delta n_0(r')d^3r'$, where the exciton density is $\delta n_0(r,t) = \sum_i \phi_i(r)f_i(r,t)$ and the sum extends over the occupied states. The exciton exchange $\delta X$ is defined analogously, again under the Tamm-Dancoff approximation.

$$\langle r|\delta X(t)|\phi_j\rangle = \sum_i f_i(r,t) \int |r - r'|^{-1} \phi_i(r')\phi_j(r')d^3r'.$$

(5)

Finally, $Q$ is a projection operator that ensures that the excited functions $f_j$ have no overlap with the occupied space, i.e., $Q = I - P$, with $P \equiv \sum_{\alpha \leq N_{occ}} |\phi_\alpha\rangle \langle \phi_\alpha|.$

The BSE equation results then when the static effective interaction $W$ replaces the Coulombic interaction in the exchange operator, yielding eventually (hiding the time-dependence of $f$):

$$\langle r|A|f_j\rangle \equiv (H_0 + \Delta - \varepsilon_j)f_j(r)$$

$$+ \delta v(r)\phi_j(r) - \sum_i f_i(r)W_{ij}(r),$$

(6)

where the action of $W$ on the occupied-occupied term is $W_{ij}(r) = \int W(r, r'; \omega = 0)\phi_i(r')\phi_j(r')d^3r$.

The linear form of the time-dependent equation readily implies that the frequency-dependent spectrum can be obtained from the dipole-dipole correlation function, where up to a constant

$$\sigma(\omega) = \langle f^0|\delta(A - \omega)|f^0\rangle = \omega\langle f^0|f(\omega)\rangle$$

(7)

where $f^0 = \langle r|Q|f_j(t = 0)\rangle = \langle r|Q|\phi_j\rangle$, and the (smoothed) delta function is readily expressed using a Chebyshev expansion in $A$.

$$|f(\omega)\rangle \equiv \delta(A - \omega)|f^0\rangle \simeq \sum_n g_n(\omega)|f^n\rangle$$

(8)

where $|f^n\rangle$ are obtained by iteratively applying a scaled $A$, starting from $|f^0\rangle$, while $g_n(\omega)$ are numerical coefficients. [32, 34] The spectrum evaluation therefore reduces to calculation of the residues, $R_n \equiv \langle f^n|f^n\rangle$. Numerically, one just requires the application of $A$ on an arbitrary exciton vector $f_j(r)$, i.e., $f \rightarrow Af$.

While here we use Eq. [1], based on the mixed hole representation $f_j(r)$, we note that in many cases one would want to use an explicit electron-hole basis. For example, for systems such as large quantum dots, where $N_{occ}$ is very large and we are interested in a smaller number of conduction states relative to the total number of occupied electron states, $n_c < N_{occ}$, it is numerically better to replace $Q$ by a projection to the $n_c$ conduction states, $Q = \sum_{c \leq n_c} |\phi_c\rangle \langle \phi_c|$. Then, the fundamental iterated object is the electron-hole basis coefficients, $f_{ia} \equiv \langle \phi_a|f_i\rangle$. In the electron-hole basis, the initial state is simply the $x$-dipoles elements, $f_{ia}^0 \equiv f_{ia}(t = 0) = \langle \phi_a|x|\phi_i\rangle$. Further, it is easy to see that the iterative application of $A$ on $f_i$ as given in Eqs. [3, 6], becomes in the electron-hole basis: $(Af)_{ia} = \sum_{j} A(i, a; j, b)f_{jb}$ and $A$ is here exactly the BSE matrix from Eq. [1]. Practically the action by $A$ would be done then as:

$$(Af)_{ia} = (\varepsilon_a - \varepsilon_i + \Delta)f_{ia} + \langle \phi_a|\delta v - \delta X|\phi_j\rangle.$$  

(9)

i.e., given the set of coefficients $f_{jb}$, one would calculate the mixed representation vectors $f_j(r)$, from which $\delta v$ and $\delta X$ follows, and then dot product per the equation above. Note that using the electron-hole basis coefficients reduces the spectral range of $A$, which will be controlled now by highest conduction state included, instead of the (much larger) highest eigenvalue of $H_0$, thereby reducing the number of required Chebyshev terms. Also, the same formalism carries over trivially to localized orthogonal basis sets, where $\varepsilon_a$ and $\varepsilon_i$ would be replaced by the Hamiltonian matrices within the electron and hole spaces, respectively. The expressions for non-orthogonal basis sets can be similarly derived.

B. Stochastic Evaluation of the action of $W$

The main numerical task in our formalism is the preparation of all $n_c(n_c + 1)/2$ functions $W_{ij}(r)$. $W$
is made from a static Coulomb part and a polarization component, \( W = v + v \chi v \equiv v + W_{\text{pol}} \) (where \( v(r,r') = |r-r'|^{-1} \)), so \( W_{ij}(r) = q_{ij} + W_{ij}^{\text{pol}} \), where
\[
q_{ij} \equiv \int |r-r'|^{-1} \phi_i(r') \phi_j(r') \, dr'.
\]

As is well-known, the action of \( W_{\text{pol}} \) can be obtained by time-dependent Hartree (TD-H) calculations. Specifically, for each pair of occupied functions \( 1 \leq i,j \leq n \), one calculates the source “potential” \( q_{ij}(r) \) due to the \( \phi_i \phi_j \) density-like source term. Then the full set of all occupied states is perturbed, \( \psi_s(r,t = 0) = e^{-i\omega_s(r) t} \phi_s(r), s = 1, \ldots, N_{\text{occ}}, \) where \( \alpha \approx 10^{-6} - 10^{-4} \) is a small perturbation strength, just as in the linear-response TD-HF derivation above. Note that to avoid a plethora of indices we do not denote the dependence of \( \psi_s \) on \( i,j \).

The perturbed states are then numerically propagated with a TD-H Hamiltonian,
\[
i\dot{\psi}_s(r,t) = (H_0 + u_{ij}(r,t)) \psi_s(r,t),
\]
where \( u_{ij}(r,t) \) is the potential due to the time-dependent density perturbation,
\[
u_{ij}(r,t) \equiv \int |r-r'|^{-1} (n^\alpha(r',t) - n^\alpha=0(r',t=0)) \, dr',
\]
where \( n^\alpha(r,t) = 2 \sum_s |\psi_s(r,t)|^2 \) is the density due to the propagated perturbed orbitals. This potential, used to propagate the time-dependent Hamiltonian, is then scaled to give the result of acting with the time-dependent effective potential, \( W(t) \), i.e.,
\[
\langle r | W_{\text{pol}}(t) | \phi_i \phi_j \rangle = \alpha^{-1} u_{ij}(r,t).
\]

Finally, the desired action of the static polarization is obtained by damped integration of the action of the time-dependent polarization, \( W_{\text{pol}}(\omega) = 0 \)
\[
\int_0^\infty e^{-\gamma \tau^2 / 2} W_{\text{pol}}(t) \, dt,
\]
i.e.,
\[
W_{ij}^{\text{pol}}(r) = \langle r | W_{\text{pol}}(\omega = 0) | \phi_i \phi_j \rangle = \alpha^{-1} \int_0^\infty e^{-\gamma \tau^2 / 2} u_{ij}(r,t) \, dt,
\]
where \( \gamma \) is a numerical convergence parameter.

The one caveat in this overall approach is that the propagation of the full set of occupied orbitals is very expensive for large systems, and is in fact the most expensive portion of other large system BSE codes. We therefore use here our stochastic-TD-H approach, which leads to a stochastic \( W \), outlined below. (Note that this is exactly the same approach we use in our stochastic GW method, with a small improvement detailed later). Briefly, in stochastic-TD-H (or stochastic-TD-DFT in the general case) we replace the full set of occupied orbitals by a few random-sign combinations of all occupied states,
\[
\eta_s(r) = L^{-\frac{1}{2}} \sum_{s \leq N_{\text{occ}}} (\pm 1) \phi_s(r),
\]
where \( L = 1, \ldots, L \) and for large systems a very small number of states is sufficient, \( L \ll N_{\text{occ}} \). The \( L \) stochastic occupied states are then treated in the TD-H procedure as if they were the full set of \( N_{\text{occ}} \) molecular orbitals, i.e., they are perturbed \( \langle \eta_s(r,t = 0^+) = e^{-i\omega_s(r) t} \eta_s(r) \rangle \) and propagated with \( H_0 + u_{ij}(r,t) \), where the time-dependent density used in constructing \( u \) is now obtained from \( n(r,t) = 2 \sum_{s \leq N_{\text{occ}}} |\eta_s(r,t)|^2 \), etc. Note that two sets need to be propagated, the perturbed \( |\eta_{s=0}^p(t)\rangle \) and unperturbed \( |\eta_{s=0}^n(t)\rangle \) stochastic orbitals.

At long times, this simple stochastic TD-H approach would eventually become unstable, due to “contamination” by occupied states. This means that the excited component, \( \eta_{s=0}^p(r,t) - \eta_{s=0}^n(r,t) \), has in it an occupied-states’ amplitudes. For regular TD-H propagation of all states this is not a problem since in the overall density the “contamination” of a propagated state \( \psi_j(r,t) \) by an occupied \( \phi_i(r) \) is exactly cancelled by the “contamination” of the opposite pair. However, in our stochastic occupied orbitals there is no such cancellation. Luckily the instability gets tamed as the system size gets bigger, but we did find that it affects the results here if untreated for medium system sizes.

To prevent the instability we simply “clean” the stochastic orbitals periodically, so after every \( M \)th time step we write:
\[
|\eta_{s=0}^p(t)\rangle \rightarrow |\eta_{s=0}^p(t)\rangle + Q |\eta_{s=0}^n(t)\rangle - Q |\eta_{s=0}^n(t)\rangle,
\]
with \( t = 0,Mdt,2Mdt, \ldots \). This does not increase the scaling of the method since the required cleaning fre-
frequency decreases (i.e., a larger $M$ is possible) with increasing system size. Also note that after each cleaning step we renormalize each $|\psi(t)\rangle$ orbital so it keeps its initial norm. In our calculations detailed below, we found $M = 10$ suffices for a total time of $t = 300$.

Finally, a well-known technical aspect is that due to the use of a finite grid box size, the scissors parameter needs to be shifted down by the $W(k \to 0)$ term. We use a variation of the existing procedures to find this term.\[38, 39\] We repeat the iterative calculation for a medium size (i.e., a lesser $n_o$ than the one eventually used) at several different box sizes, and for each run find the average Kohm-Sham potential on the box boundary, $\phi_{bndry}$. Then we approximate $W(k \to 0) = \epsilon_{eff}^{-1} \phi_{bndry}$, and the parameter $\epsilon_{eff}^{-1}$, playing the role of an inverse dielectric constant, is fit so that the spectra from the different box-sizes runs overlap.

Summarizing the resulting algorithm: we use stochastic TD-H to prepare $W_{ij}(r)$ for $n_o(n_o + 1)/2$ occupied-state pairs. Separately we use sGW to calculate the self-consistent $GW$ scissors shift $\Delta$ and subtract from it the $W(k \to 0)$ term. Then for each polarization we start with the dipole exciton state. The Tamm-Dancoff operator, Eq. (6), is successively applied and the Chebyshev residues $R_o$ are used to calculate the absorption frequencies. As usual, if one wants to characterize the different peaks then the Chebyshev expansion of the frequency-resolved exciton state, Eq. (8), can be used (potentially with filter-diagonalization\[40, 41\] for resolving different sub-peaks).

The two parts of the method, preparing $W_{ij}$ and applying the $A$ operator, both scale as $O(n^2 g n_g)$ for $n_g$ grid points, more gentle than current methods. Formally, this is cubic scaling with system size but in practice the scaling is better since $n_o$ often rises only gently with $N_{occ}$. In addition, the number of grid points would be reduced in future studies as we have shown that very sparse grids suffice when using orthogonal projected augmented waves (OPAW) instead of pseudo-potentials.\[42\]

### III. RESULTS

We demonstrate the algorithm on a characteristic large organic semiconducting system, a carbon nanohoop-fullerene complex. In the last decade, cyclopara-phenylenes (CPPs), also known as carbon nanohoops, have emerged as highly structurally tunable emitters, with rich size-dependent opto-electronic properties and host-guest chemistry.\[43\] While substantial DFT modeling has been completed on CPP+fullerene complexes,\[34, 44, 46\] extraction of optical properties at this level of theory is difficult due to the characteristic charge transfer states in CPPs. Further, it has already been established that the BSE formalism is very accurate in predicting the properties of other fullerene-polymer complexes.\[47\] Here we present detailed results for the smallest such CPP+fullerene “pea-pod,”\[48, 49\] [10]-CPP+C$_{60}$.

To simulate [10]-CPP+C$_{60}$, we use a generous box of $(N_x, N_y, N_z) = (100, 100, 84)$ with a grid spacing of 0.5Bohr, using norm-conserving pseudo-potentials (NCPP).\[50, 51\] The effective inverse dielectric constant was found to be $0.30$, which gives, at this grid size, a shift of $-W(k \to 0) = -0.29$eV.

The DFT gap is $1.02$eV. This gap is corrected with stochastic GW by an amount of $1.24$eV, and after applying the self-consistent self-consistent $\Delta GW_{0}$ this gap correction rises to $1.33$ eV (i.e., a fundamental $\Delta GW_{0}$ gap of $2.35$eV). Combined with $-W(k \to 0)$, the overall scissors shift used is $\Delta = 1.04$.

The calculations of the action of $W$, Eqs. (10)- (13), were done with a broadening of $\gamma = 0.1$ Hartree. A time-step $dt \approx 0.1$ a.u. was used for a split-operator propagation of $L = 10$ stochastic states, and the cleaning was done every $M = 10$ steps. The runs were done for $n_o = 100$ valence states, requiring $n_o(n_o + 1)/2 = 5005$ actions of $W$.

The BSE iterative Chebyshev procedure was then done using the $n_o = 100$ valence states. The Chebyshev expansion of $\delta(A - \omega)$ is evaluated with a Gaussian broadening of $0.08$ eV; this does not effect the spectrum significantly as it is naturally broadened due to the large number of excitons. The runs took a total of 4000 core hours, i.e., 40 wall hours on a 100-core AMD Milan cluster.

In Figure 2 we show the calculated absorption spectrum, both total and separated to in-plane ($x$ and $y$), and perpendicular ($z$) polarizations. While there are no gas phase spectra of $[10]$-CPP+C$_{60}$ due to the fullerene slipping out of its “pea-pod”, for the lowest exciton energy, we get reasonable agreement with just $[10]$-CPP.\[52, 53\] and a stabilizing $[10]$-CPP+C$_{60}$-[2]-Rotaxane complex.\[54\] For just $[10]$-CPP in the gas phase the lowest strong transition sits $0.4$ eV higher than shown in Figure 2. Qualitatively, with the addition of the fullerene in the middle in our simulations, the overall dielectric constant would increase, thereby lowering the energy of the first exciton state. This shift is consistent with the shift to lower energies seen in the stabilized $[10]$-CPP+C$_{60}$-[2]-Rotaxane complex in solution.

In Figure 3, we show the exciton density of four prominent excitons as labeled in Figure 2 and, for the sake of analysis we also extract the exciton density in the electron-hole state basis. Specifically, we calculate $|f_j(\omega)|$ at the four exciton peak frequencies and then calculate the overlap onto to a set of unoccupied wave functions, giving us $f_{\text{ua}}$ for as many $a$ as we desire. For the lowest peak most of the exciton density is concentrated at near-gap $a$ states. However, for the largest spectrum peak at $6.83$ eV, labeled (iv) in Figure 2, one ought to go beyond the figure and use (100,250) states to capture the same amount of density, a basis size that’s very substantial. This demonstrates the strengths of the stochastic resolution of the action of $W$ in our present approach, as the full unoccupied space is sampled rather than just a subset of conduction states.
### IV. FORWARD PERSPECTIVE

Our results show that even larger systems are feasible. This is evident by the fact that the runs were not optimised. For example, while we used \( n_0 = 100 \), a smaller number \( n_0 = 70 \) would have sufficed, reducing the cost by a factor of two. Similarly, the box size was very large, and a grid with almost half the points (or even less with OPAW[T2]) would have sufficed. With optimized parameters, one can therefore easily reach systems with 1000-2000 electrons.

An interesting feature of the method is that for larger systems it becomes less and less sensitive to stochastic errors. Those errors appear primarily in the sGW calculation of the scissors shift, but this calculation scales sub-linearly with system size. Already here the sGW calculation took less than 20% of the total calculation and had an error of 0.05eV, so for larger systems an even higher accuracy would be obtained with a small fraction of the total cost.

An interesting question is how to go to huge systems, with many thousands of electrons. For this we note that that while this implementation of the BSE scales formally cubically, it holds promise to give eventually quadratic scaling. To achieve this, one would need to implement localized occupied basis sets sets (see references [55, 56]) for reducing the number of \( i, j \) pairs for which \( W_{ij}(r) \) needs to be calculated, and perhaps even stochastically sample these \( \phi_i(r)\phi_j(r) \) pairs. Work along this lines would be reported in future publications.

### ACKNOWLEDGMENTS

NCB would like to acknowledge the National Science Foundation Graduate Research Fellowship Program under grant DGE-2034835. JRC would like to acknowledge the ACS Petroleum Research Fund, grant 62717-DNI6. NCB, MN, and DN are grateful for support by NSF grant CHE-1763176. Computational resources were supplied through the XSEDE allocation TG-CHE170058.

---

[1] B. Kippelen and J.-L. Brédas, Energy & Environmental Science **2**, 251 (2009).

[2] R. Improta, F. Santoro, and L. Blancafort, Chemical Reviews **116**, 3540 (2016).
[55] A. S. Banerjee, L. Lin, W. Hu, C. Yang, and J. E. Pask, 
The Journal of Chemical Physics 145, 154101 (2016).
[56] L. Peng, D. Peng, F. L. Gu, and W. Yang, Journal of 
Chemical Theory and Computation 18, 2975 (2022)