JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

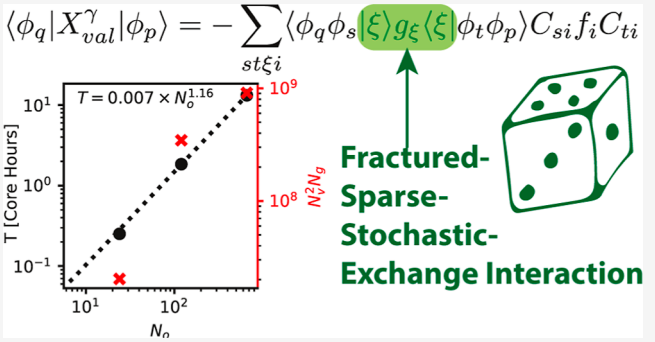
Deterministic/Fragmented-Stochastic Exchange for Large-Scale Hybrid DFT Calculations

Nadine C. Bradbury, Tucker Allen, Minh Nguyen, and Daniel Neuhauser*

Cite This: J. Chem. Theory Comput. 2023, 19, 9239−9247 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: We develop an efficient approach to evaluate range-separated exact exchange for grid- or plane-wave-based representations within the generalized Kohn−Sham−density functional theory (GKS−DFT) framework. The Coulomb kernel is fragmented in reciprocal space, and we employ a mixed deterministic-stochastic representation, retaining long-wavelength (low-k) contributions deterministically and using a sparse ("fragmented") stochastic basis for the high-k part. Coupled with a projection of the Hamiltonian onto a subspace of valence and conduction states from a prior local-DFT calculation, this method allows for the calculation of the long-range exchange of large molecular systems with hundreds and potentially thousands of coupled valence states delocalized over millions of grid points. We find that even a small number of valence and conduction states is sufficient for converging the HOMO and LUMO energies of the GKS−DFT. Excellent tuning of long-range separated hybrids (RSH) is easily obtained in the method for very large systems, as exemplified here for the chlorophyll hexamer of Photosystem II with 1320 electrons.



INTRODUCTION

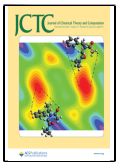
The introduction of hybrid exchange and long-range hybrid functionals into density functional theory (DFT) dramatically improved their accuracy.[1−7] These improvements, now 30 years old, enabled the rapid growth of DFT as a standard tool in the chemistry lab, with the establishment of many popular commercial and open-source pieces of software. Unfortunately, it is this key improvement in functional design, exact exchange, that limits the size of computation that is feasible for most researchers with a set budget of computing power and time. Traditional Hartree−Fock-type exchange requires the generation of all 2-electron integrals on a given basis, scaling naively as $O(N_o^4)$ for N_o spatially occupied orbitals.

The most substantial advancement in improving the computational cost of exact exchange in ab initio DFT has come in the form of the so-called "resolution of the identity" (RI) methods.[8] Now widely adopted, these methods reduce the memory cost of exact exchange and, in practice, dramatically reduce the computational scaling. For the entire set of 2e-integrals, $\langle pq|rs\rangle$, one expands the identity in another auxiliary basis, β, reducing a 4-center integral tensor to a product of two 3-center integral tensors, $\langle pq|rs\rangle = \sum_\beta \langle pq|\beta\rangle\langle\beta|rs\rangle$. Such auxiliary basis sets are either pretabulated or optimized on the fly from primary basis functions.[9] With this intelligent design, one can cap the number of β to be comparable to the number of atomic orbitals needed for the calculation,[10] though this auxiliary basis traditionally still scales with system size. In practice, several advancements such as pair atomic RI, auxiliary density RI, and atomic concentric RI have reduced computational cost for many different kinds of DFT codes.[11−14] Most of these advancements depend on using atomic basis sets.

Other efforts involve the power of parallel computing, such as fragmented systems, localized auxiliary orbitals, and sparse matrix algorithms.[15,16] In large finite systems, the sparsity of overlap integrals allows for highly optimized localized auxiliary orbitals and near-linear scaling.[17,18] Multilevel fragmented approaches have also recently improved scaling, especially in spatially localized cases.[16] Modern graphical processing units also contribute to unlocking larger and larger calculations with RI methods.[19,20]

Separately, we introduced a stochastic formalism for Hartree−Fock or long-range exchange for grid-based DFT codes.[21] In this formalism, the exchange becomes a projection to a stochastic occupied orbital, which is a random linear combination of all occupied orbitals represented on a grid basis

Received: September 8, 2023
Revised: November 6, 2023
Accepted: November 14, 2023
Published: December 5, 2023



times a random amplitude due to the Coulomb potential. A statistical average over multiple random vectors converges to the matrix elements of the exchange operator. In this case, each random orbital covers the entire eigenbasis of the molecule, and the number of such operators typically does not grow with system size and occasionally shrinks due to self-averaging.[21]

In this work, we employ a different strategy whereby the individual molecular states are treated deterministically. However, the usual cost of making all the matrix elements of the Coulomb interaction is reduced by orders of magnitude (and its scaling made constant) by the fragmented-stochastic compression approach we developed in a different context, stochastic GW.[22] Basically, we have shown that data over a large grid can be efficiently represented by a stochastic basis made of many small "fragments". Beyond a small threshold, the error does not depend on the fragment size, only on the number of fragments, so a large number of short fragments can be used to efficiently represent data on a giant grid.

In this work, we combine the best of sparse stochastic basis compression with the RI technique. In short, we split the Coulomb kernel for the exchange calculation to two sets (see also ref 23). The first is the large interaction at few low wavevectors (small k) which is treated deterministically. The remainder, the interaction at the very many (often millions of) high k's, is represented here cheaply and accurately by fragmented stochastic compression, i.e., by representing the interaction through a small number (few thousands, here) of short stochastic vectors, and this number does not increase with system size.

The second ingredient to the present deterministic/fragmented-stochastic approach is to represent the hybrid-DFT Hamiltonian on the basis of molecular orbital states (MOs) near the Fermi energy (near-gap) from local-DFT. Specifically, we first perform a local- (or semilocal) functional DFT calculation by any efficient basis-set or plane-wave method. We then divide the resulting local-DFT MOs into core, valence, and conduction as well as high virtual orbitals which are ignored.

The core orbitals of this preliminary calculation are assumed to be a good representation of the core orbitals in the eventual hybrid calculation. We therefore assume that the valence and conduction orbitals of the hybrid case can be expanded from the valence and conduction MOs of the local-DFT calculation. This restriction to top valence and bottom conduction orbitals is of course routinely done in beyond-DFT methods, such as RPA, TD-DFT, and the Bethe–Salpeter Equation.

With the introduction of sparse stochastic compression to the plane-wave auxiliary basis, the scaling of the resulting approach is very gentle with system size, so that in practice, the hybrid exchange correction costs less than the underlying local-DFT calculation. Further, the approach is easily parallelizable. We label it as near-gap Hybrid DFT (ngH-DFT).

In the sections below, we develop the ngH-DFT formalism, benchmark its convergence for naphthalene and fullerene, and then show the method's power by solving for a hexamer dye complex, a large system of biological significance. The proper inclusion of exact exchange here in such a large biomolecule is promising for the future use of general post-DFT methods in giant systems.

METHODOLOGY

Hybrid DFT in the Valence-Conduction Subspace. We begin with the Kohn–Sham (KS) orbitals $\{\phi_s\}$ and associated eigenvalues $\{\varepsilon_s\}$ of a ground-state DFT calculation approximately satisfying $h_0\phi_s \approx \varepsilon_s\phi_s$, where h_0 is the starting local functional KS-DFT Hamiltonian. It is not necessary that the starting calculation be fully converged, and it can originate from LDA, PBE, or whichever DFT flavor of choice, but, for simplicity, it will be denoted here as LDA–DFT.

The MOs from the LDA–DFT calculation, denoted by ϕ, are then divided into four set of states: N_{core} core, $N_{\mathrm{v}}(= N_{\mathrm{o}} - N_{\mathrm{core}})$ valence, N_{c} conduction, and the remainder, which are high conduction states which are neglected. We stress that the orbitals denoted as the core here are simply lower-energy MOs that are less important to the chemistry of large molecules; these are not the atomic-core electrons that are usually removed from the DFT Hamiltonian by the use of pseudopotentials.

We then approximate that the core states from the LDA calculation are unchanged under the GSK Hamiltonian, as least as far as their effect on their valence orbital energies is concerned, i.e.

$$\psi_{\mathrm{f}} = \phi_{\mathrm{f}}, \qquad f \in \text{core} \tag{1}$$

where ψ refers to a GKS MO. Therefore, the $M \equiv N_{\mathrm{v}} + N_{\mathrm{c}}$ GKS near-gap (i.e., valence + conduction) states are assumed to be described by the valence-conduction LDA states, i.e.

$$\psi_s(r) = \sum_p \phi_p(r) C_{ps} \tag{2}$$

where s, p, and q are indices over the M near-gap states.

The converged LDA–DFT Hamiltonian is expressed as (using atomic units throughout)

$$h_0 = -\frac{1}{2}\nabla^2 + v_{\mathrm{eN}}^{\mathrm{NL}} + v^0[n_0](r) \tag{3}$$

with the respective terms being the kinetic energy, the nonlocal component of electron–nucleus interaction, and the local KS potential. The latter is a functional of the LDA density, $n_0(r)$, and contains the local electron–nucleus interaction, Hartree potential, and local exchange–correlation (XC) potential, taken here to be PW-LDA[24]

$$v^0[n_0](r) = v_{\mathrm{eN}}^{\mathrm{local}}(r) + \int \frac{n_0(r')}{|r - r'|}\mathrm{d}r' + v_{\mathrm{XC}}^0[n_0](r) \tag{4}$$

The electron–nucleus interaction is handled with Troullier–Martins norm-conserving pseudopotentials.[25] Additionally, the Martyna–Tuckerman approach is used to avoid the effect of periodic images in our simulations.[26]

We now turn to the GKS Hamiltonian. Here, we employ a long-range hybrid, though the same formulation also applies to any other form, such as short-range or Becke-type fractional exchange. Note that to avoid a cluttering of indices, we write here only of the closed-shell GKS formalism, but the GKS Hamiltonian would generally be spin-selective (unlike the LDA–DFT). In fact, the tuning procedure we use to yield the correct γ requires a spin-selective Hamiltonian, as discussed later.

The starting point is the long-range part of the Coulomb interaction, defined as $u^\gamma(|r - r'|) = \mathrm{erf}(\gamma|r - r'|)/|r - r'|$, so for the exchange, the Coulomb kernel in position space is split as[7]

$$\frac{1}{|r - r'|} = \frac{\mathrm{erfc}(\gamma|r - r'|)}{|r - r'|} + u^\gamma(|r - r'|) \tag{5}$$

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239–9247

The first term dominates at short distances and is treated locally, while the second, long-range, term is accounted for explicitly.

Hybrid functionals, such as B3LYP for example, are better than local or semilocal functionals for transfer and excitonic effects due to the addition of a $-C/|r-r'|$ asymptotic behavior of the exchange term. Range-separated hybrid functionals, where the constant has the correct value, $C = 1$, mostly alleviate the nonphysical long-range self-repulsion in the LDA potential. Further, in optimally tuned range-separated hybrids, the charge-transfer characteristics are further improved by tuning the exchange to obey the ionization potential (IP) theorem. The range−separation parameter γ is best obtained by enforcing piece-wise linearity of the energy with the electron number. For further details on the tuning procedure in general, see ref 27

The GKS Hamiltonian is then

$$h = -\frac{1}{2}\nabla^2 + v_{eN}^{NL} + v^{\gamma}(r) + X_{val}^{\gamma} + X_{core}^{\gamma} \tag{6}$$

where γ refers to one or more parameters of the hybrid exchange. The γ-dependent Kohn−Sham potential is

$$v^{\gamma}(r) = v_{eN}^{local}(r) + \int \frac{n(r')}{|r-r'|}dr' + v_{XC}^{SR,\gamma}[n] \tag{7}$$

where SR denotes short-range and $n(r)$ is the overall density, made from a sum of core and valence densities

$$n(r) = n^{core}(r) + n^{val}(r) \tag{8}$$

where $n^{core}(r) = 2\sum_{f\in core}|\phi_f(r)|^2$. The valence density is

$$n^{val}(r) = 2\sum_i f_i |\psi_i(r)|^2 = 2\sum_{pq} \phi_p(r) P_{pq} \phi_q(r) \tag{9}$$

where the density matrix is $P_{pq} = \sum_i C_{pi} f_i C_{qi}$. Here, the sum runs over all occupied (or partially occupied) valence GKS MOs, and f_i is the occupation, which can be fractional

$$f_i(\varepsilon_i;\mu) = \frac{1}{1+e^{(\varepsilon_i-\mu)/k_BT}} \tag{10}$$

The action of the valence (short-hand val) component of the γ-dependent exact exchange operator on a general function η is

$$(X_{val}^{\gamma}\eta)(r) = -\sum_i f_i \psi_i^*(r) \int u^{\gamma}(|r-r'|)\eta(r')\psi_i(r')dr' \tag{11}$$

The contribution of the core states to the exchange part of the Hamiltonian will be done perturbatively, as discussed later. The LDA → GKS rotation matrix, eq 2, is initially $C_{ps} = \delta_{ps}$ and is then iterated in the SCF procedure.

The Hamiltonian matrix elements in the valence-conduction basis are

$$h_{pq} = \langle\phi_p|h_0 + \delta v + X_{val}^{\gamma}|\phi_q\rangle \tag{12}$$

where $\delta v \equiv v^{\gamma}(r) - v^0(r)$ is the difference between the current GKS and the initial estimate KS potentials.

Traditionally, the matrix elements of the valence exact exchange are produced from a generalized 4-index integral tensor by starting with

$$\langle\phi_q|X_{val}^{\gamma}|\phi_p\rangle = -\sum_i f_i (\phi_q\psi_i|u^{\gamma}(|r-r'|)|\psi_i\phi_p) \tag{13}$$

and inserting the expanded wave function gives

$$\langle\phi_q|X_{val}^{\gamma}|\phi_p\rangle = -\sum_{st} (\phi_q\phi_s|u^{\gamma}|\phi_t\phi_p)P_{st} \tag{14}$$

where real-valued orbitals are used with the chemists' convention of $(rr|r'r')$.

Deterministic/Fragmented-Stochastic Representation of the Coulomb Kernel. Our starting point is the exchange kernel in eq 14, which requires a generic convolution form, written schematically as $w(r) = \int u^{\gamma}(r-r')y(r')dr'$. This form is diagonal in reciprocal space, and for finite grids, it reads

$$w(k) = \frac{1}{V}\sum_k u^{\gamma}(k)y(k) \tag{15}$$

In the Martyna−Tuckerman approach, V is the overall volume including full padding in each direction (i.e., V is $2^3 = 8$ times the wave function volume). Further, $u^{\gamma}(k)$ is not necessarily positive due to the Martyna−Tuckerman construct.

Since $u^{\gamma}(k)$ is large at low k, its action is evaluated deterministically below an assigned cutoff, k_{cut} (the results are correct upon convergence for any k_{cut} as this parameter only affects the speed of convergence). Specifically, for a given k_{cut}, we divide k-space into 3 subspaces: "low"—values of k below k_{cut}; "high$^+$"—values above k_{cut} where $u^{\gamma}(k)$ is positive; and "high$^-$"—values above k_{cut} where $u^{\gamma}(k)$ is negative. The number of points in each space is denoted, respectively, as $N_{k_{low}}$,$N_{k_{high^+}}$, and $N_{k_{high^-}}$. Formally, we then write the identity operator in the reciprocal space as

$$I = \sum_{k_{low}} |k_{low}\rangle\langle k_{low}| + \sum_{k_{high^+}} |k_{high^+}\rangle\langle k_{high^+}| + \sum_{k_{high^-}} |k_{high^-}\rangle\langle k_{high^-}| \tag{16}$$

The Coulomb long-range operator is then

$$u^{\gamma} = \sum_{k_{low}} |k_{low}\rangle u^{\gamma}(k_{low})\langle k_{low}| + \sum_{k_{high^+}} \sqrt{u^{\gamma}(k_{high^+})}\,|k_{high^+}\rangle\langle k_{high^+}|\sqrt{u^{\gamma}(k_{high^+})} - \sum_{k_{high^-}} \sqrt{|u^{\gamma}(k_{high^-})|}\,|k_{high^-}\rangle\langle k_{high^-}|\sqrt{|u^{\gamma}(k_{high^-})|} \tag{17}$$

Next, we introduce stochastic fragmented bases[22] for the positive and negative high-k components. We detail the discussion for the high$^+$ space, and it follows analogously for the high$^-$ space.

A set of N_{α^+} short vectors is chosen, where each is randomly positive and negative in a "strip", also labeled as "fragment"

$$\alpha^+(k_{high^+}) = \pm\sqrt{\frac{N_{k^+}}{L}} A_{\alpha^+}(k_{high^+}) \tag{18}$$

Here, $A_{\alpha^+}(k)$ is a projection to a randomly placed fragment α^+ of length L, i.e., is 1 within the fragment and 0 outside, so $\alpha^+(k_{high^+})$ is randomly positive or negative in the fragment and vanishes outside. The strip length, L, is the same for each fragment. The fragments thus randomly and uniformly sample the entire $\{|k_{high^+}\rangle\}$ space.

The constant factor in eq 18 ensures that with sufficient sampling, the α^+ vectors form an orthonormal set, as explained below. A technical point is that fragments that start near the

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247

edge of the k_{high^+} space, i.e., their starting point is larger than $N_{k_{\text{high}^+}} - L$, need to wrap around; alternately, one can zero pad the space of $N_{k_{\text{high}^+}}$ points by L points on both sides, and then the constant square root factor in eq 18 needs to be slightly modified.

The strip length L and the number of stochastic vectors N_{α^+} are chosen such that each k point in the high$^+$ space is sufficiently "covered", i.e., will be adequately visited by the stochastic basis α^+. Specifically, we choose a coverage parameter, cov, which samples how often, on average, each point is sampled. The number of chosen stochastic vectors is then

$$N_{\alpha^+} = \frac{\text{cov} \cdot N_{k_{\text{high}^+}}}{L} \tag{19}$$

In the limit that this coverage parameter is large, the stochastic fragments form an orthonormal basis, i.e.

$$\{\alpha^+(k_{\text{high}^+})\alpha^+(k'_{\text{high}+})\} = \delta_{k_{\text{high}^+}k'_{\text{high}+}} \tag{20}$$

where the large curly brackets denote a stochastic sampling with, formally, cov $\to \infty$. In practice, it is enough to use cov $\simeq$ 5.

We then define N_{α^+} states, $|\xi^+\rangle$, with components

$$\langle k_{\text{high}^+}|\xi^+\rangle = \sqrt{u^\gamma(k_{\text{high}^+})}\,\alpha^+(k_{\text{high}^+}) \tag{21}$$

We repeat the whole procedure for the high$^-$ space and end up with N_α-states for the negative high-k portion of the exchange kernel

$$\langle k_{\text{high}^-}|\xi^-\rangle = \sqrt{|u^\gamma(k_{\text{high}^-})|}\,\alpha^-(k_{\text{high}^-}) \tag{22}$$

We now define a combined set of states, of size $N_\xi = N_{k_{\text{low}}} + N_{\alpha^+} + N_{\alpha^-}$, which is glued together via direct summation

$$|\xi\rangle = \{\sqrt{|u^\gamma(k_{\text{low}})|}\,|k_{\text{low}}\rangle\} \oplus \{|\xi^+\rangle\} \oplus \{|\xi^-\rangle\} \tag{23}$$

We similarly define a sign vector of length N_ξ

$$g_\xi = \{\text{sign}(u^\gamma(k_{\text{low}}))\} \oplus \{1\} \oplus \{-1\} \tag{24}$$

i.e., in addition to the sign of the interaction for the low-k components, g is composed of N_{α^+} values of 1 and N_{α^-} values of -1.

With these definitions, we now reach the stochastic fragmented basis representation of the exchange operator

$$u^\gamma = \sum_\xi |\xi\rangle g_\xi \langle\xi| \tag{25}$$

This is the central equation of the deterministic/stochastic-fragment representation of the Coulomb interaction. As mentioned, it is used here only for the exchange component and not for the direct Coulomb interaction.

Inserting this form of u^γ in the matrix element of eq 14

$$\langle\phi_q|X^\gamma_{\text{val}}|\phi_p\rangle = -\sum_{st\xi i} \langle\phi_q\phi_s|\xi\rangle g_\xi \langle\xi|\phi_t\phi_p\rangle C_{si} f_i C_{ti} \tag{26}$$

and defining

$$u_{\xi pi} \equiv \sum_t C_{ti}\langle\xi|\phi_t\phi_p\rangle \tag{27}$$

yields the final expression for the exact exchange matrix elements

$$\langle\phi_q|X^\gamma_{\text{val}}|\phi_p\rangle = -\sum_{i\xi} u^*_{\xi qi} f_i g_\xi u_{\xi pi} \tag{28}$$

Note that for a spin-resolved calculation, the only difference is that in addition to the amplitudes C_{ti} and the exchange correlation potential δv, the transformed exchange vectors $u_{\xi p}$ and the X^γ_{val} matrix would also gain a spin index.

Core State Correction to the Exchange. In the previous sections, the core state contributions to the exact exchange were neglected. We will account for it by a perturbative correction to the KS eigenvalues $\varepsilon_s \to \varepsilon_s + \Delta_s$, where

$$\Delta_s = \langle\psi_s|X^\gamma_{\text{core}}|\psi_s\rangle \tag{29}$$

is evaluated as

$$\Delta_s = -\sum_{f\in\text{core}} \langle\psi_s\phi_f|u^\gamma|\phi_c\psi_f\rangle \tag{30}$$

Since in this work, we are most interested in the highest occupied MO (HOMO) and lowest unoccupied MO (LUMO) energies, we calculate the correction for these two states only, labeled as Δ_{occ} and Δ_{unocc}. The core corrections stabilize the frontier orbital eigenvalues and bandgap even when the number of active valence and conduction orbitals included in the GKS-Hamiltonian is dramatically reduced. Computationally, these core corrections are very cheap as they are only added in the last iteration, and they are calculated as explicit convolution integrals (Figure 1).

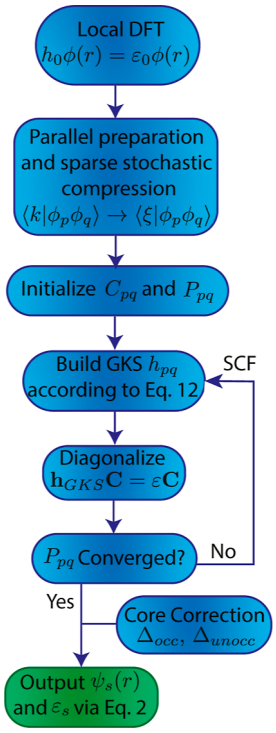


Figure 1. Flowchart of the complete algorithm.

RESULTS

We test the ngH-DFT method with three molecular systems of increasing size: naphthalene (N_o = 24), fullerene (N_o = 120), and a hexamer dye complex (N_o = 660). An initial PW-LDA DFT calculation is performed for all systems. The large dye system's nuclear coordinates, optimized at the PBE/def2-

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247

Table 1. Fundamental Gaps for Naphthalene, Fullerene, and a 476 Atom Hexamer Dye Complex[a]

system	N_o	N_v^{max}	N_c^{max}	optimal γ (Bohr^{-1})	plane-wave LDA−DFT	atomic basis-set LDA−DFT	ngH-DFT	atomic basis-set RSH−DFT
naphthalene	24	24	104	0.285	3.34	3.34	8.63	8.54
fullerene	120	120	480	0.189	1.63	1.64	5.42	5.40
hexamer	660	200	400	0.120	1.23		3.81	

[a]Also shown are the total number of occupied states, the maximum numbers of valence and conduction states, and the range−separation parameter for each system. All energies are in eV. The atomic basis-set calculation uses the NWChem package. Both ngH-DFT and the atomic basis-set RSH−DFT use the BNL XC functional.

TZVP-MM level, were taken from previous studies.[28,29] All simulations use a generous box size that extends 6 Bohr beyond the extent of the molecule in each direction, with real-space grids (before the Martyna−Tuckerman expansion) of N_g = 50,688, 216,000, and 2,273,280 points, respectively, and uniform grid spacings $dx = dy = dz = 0.5$ Bohr. The range-separated hybrid (RSH)−DFT studies use the Baer−Neuhauser−Livshits (BNL) XC functional.

To balance the cost between the deterministic low-k and sparse stochastic high-k components of the exchange, we set, as mentioned, the size of the sparse basis, N_α, to be equal to the number of deterministic k-vectors, $N_{k_{low}}$. The k_{cut} parameter, separating the deterministic and fragmented-stochastic terms, is adjusted so that for most of our simulations (except for a few reported in Table 4), a constant $N_{k_{low}} \simeq N_\alpha = 5000$ is used, so the auxiliary basis size is $N_\xi \simeq 10{,}000$. The associated k_{cut} values (in atomic units) are, respectively, 1.8, 1.1, and 0.5.

Table 2. Naphthalene Frontier Orbital Eigenvalues, Fundamental Gaps, and Core Corrections for Different Numbers of Valence to Conduction States[a]

N_v:N_c	ε_H	ε_L	gap	Δ_{occ}	Δ_{unocc}
24:104	−8.77	−0.14	8.63		
20:40	−8.78	−0.15	8.63	−0.07	−0.04
10:20	−8.72	−0.08	8.64	−0.23	−0.03

[a]All energies are in eV. The first row includes all occupied states, so it has no core correction.

Table 3. Fullerene Frontier Orbital Eigenvalues, Fundamental Gap, and Core Corrections for Different N_v:N_c[a]

N_v:N_c	ε_H	ε_L	gap	Δ_{occ}	Δ_{unocc}
120:480	−8.26	−2.84	5.42		
40:80	−8.20	−2.78	5.42	−0.15	−0.12
20:40	−8.23	−2.76	5.47	−0.42	−0.29
20:20	−8.23	−2.77	5.46	−0.42	−0.29
10:10	−8.25	−2.83	5.42	−1.12	−0.63

[a]All energies are in eV.

Note that at these values and for the tuned values of γ listed below (0.285, 0.189, and 0.12 Bohr^{-1}, respectively), the high-k interaction is very small, as $v_\gamma(k) \propto \exp(-k^2/4\gamma^2)/k^2$ (although it is numerically somewhat larger in the Martyna−Tuckerman approach). For a preliminary study of the potential usefulness of the approach for other types of hybrid functionals, where $v_\gamma(k)$ is not so tiny at high k, we also include later results at lower k_{cut}.

Before showing the promise of using only a fraction of near-gap states, we report in Table 1 the fundamental gaps obtained for naphthalene, fullerene, and the hexamer, using a large number of valence and conduction states (including all N_o occupied states for the two smaller systems). For naphthalene and fullerene, we benchmark vs an all-electron calculation that uses the NWChem package,[30] with a Gaussian aug-cc-pvdz basis containing 302 atomic basis functions for naphthalene and 1380 for fullerene. The fundamental gaps agree well between ngH-DFT and NWCHEM, and we demonstrate below that this agreement is maintained even when we significantly reduce the size of the valence-conduction near-gap space.

Table 4. Fundamental Gap and Its Standard Deviation, σ, for Three Test Systems (in eV), for Different Numbers of Deterministic Low-k Terms, $N_k$low, and Sizes of the Sparse Stochastic Basis, N_α

system	N_v	N_c	$N_{k_{low}}$	N_α	gap	σ
naphthalene	20	40	501	500	8.6329	0.0122
			501	5000	8.6373	0.0077
			4987	5000	8.6344	0.0004
fullerene	40	80	515	500	5.4209	0.0066
			515	5000	5.4226	0.0051
			4945	5000	5.4228	0.0001
hexamer	40	80	503	500	3.7914	0.0286
			503	5000	3.8018	0.0152
			4785	5000	3.8032	0.0002

Both the ngH-DFT and RSH−DFT calculations use the same optimal range−separation parameter γ obtained by systematic tuning of the HOMO energies, enforcing consistency with the IP theorem as required by the BNL functional. In practice, the IP theorem is enforced by tuning γ such that the HOMO energy does not change when the system is slightly ionized, and we use here $\varepsilon_{HOMO}^{neutral} = \varepsilon_{HOMO}^{+0.1}$.[27] The ngH-DFT for the charged system is done via an open-shell calculation.

A side note is that to ensure rapid convergence with the valence basis size N_v, we find it important to do the initial LDA calculation with the right charge as this ensures that the core eigenstates are correctly polarized. Thus, the charged system ngH-DFT requires an initial basis-set ϕ_s from an LDA SCF with fractional occupation $f_{HOMO} = 1-0.1$ (though done in a nonspin-selective calculation) rather than relying on the ϕ_s from the neutral LDA.

In Table 2, we provide the HOMO and LUMO eigenvalues and gap for naphthalene for a chosen number of valence and conduction states. The first row in the table includes all occupied and a large number of unoccupied states, while the following two use a reduced valence-conduction space. Reduction of this active space necessitates core corrections of the HOMO and LUMO eigenvalues. The gap does not change much when the valence-conduction basis-set size is made smaller.

As Table 3 shows, the convergence is even better for the next bigger system, fullerene. The number of included valence

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247

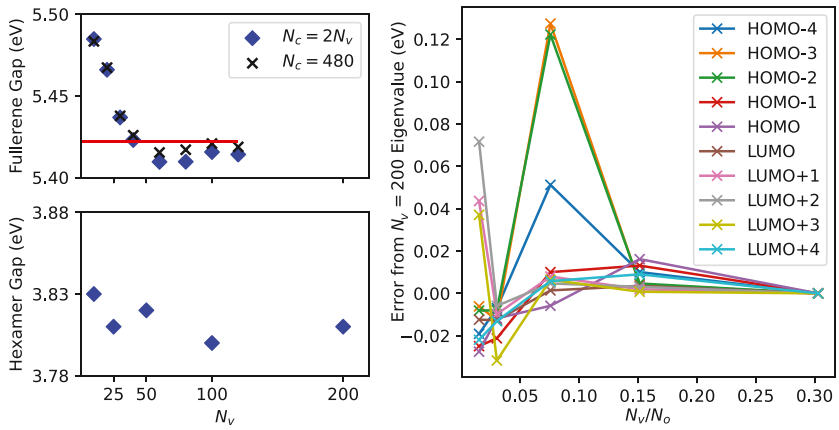


Figure 2. (Top) convergence of the fundamental gaps of fullerene and (bottom) hexamer with the number of valence states, N_v, and the number of conduction states, N_c, either chosen as $N_c = 2N_v$ (blue diamonds) or fixed at $N_c = 480$ (black x). The red line is the reference value of the fullerene gap including all occupied states, $N_v = N_o = 120$ and $N_c = 480$. (Right) convergence of the 10 states nearest to the gap with N_v for the hexamer system relative to n_o.

and conduction states can now be much smaller than n_o. This rapid convergence with N_v is also shown in Figure 2a. The figure further shows that the results converge rapidly with the conduction basis size so that $N_c = 2N_v$ gives essentially the same result as using a very large value of N_c.

The convergence with N_v further improves for the biggest system, the hexamer, as shown in Figure 2b. The gaps shown all agree within ±0.02 eV, even for very small N_v and N_c. This implies that very large systems could be used with a small valence-conduction space.

Figure 3 shows, for the hexamer, the convergence of the range−separation parameter, as well as the core corrections.

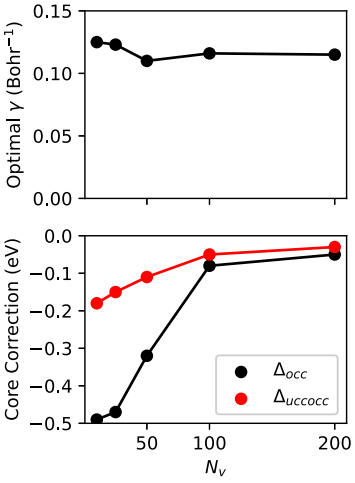


Figure 3. (Top) convergence of γ and (bottom) core corrections as a function of N_v for the hexamer system.

The extracted γ values are consistent, even with a valence-conduction space of only ten valence and ten conduction orbitals. This implies that optimal tuning of long-RSHs of giant systems could be done rather cheaply.

The single-run stochastic error, i.e., the standard deviation of the energy, is shown in Table 4. It is estimated from the results of ten independent runs. As mentioned, for $N_{k_{low}} \simeq 5000$, k_{cut} is large for each of the three studied systems so that the values of $v_\gamma(k)$ are very small for the stochastically sampled high-k spaces. We therefore also include results with a smaller k_{cut} so that $N_{k_{low}} \simeq 500$, for $N_\alpha = 500$ and $N_\alpha = 5000$ (i.e., $N_\xi \simeq 1000$, 5500). As shown, the statistical error is still quite small, about 0.01−0.03 eV, and is lower than or similar to the low stochastic error associated with using a small value of N_v.

Algorithm Cost. In addition to the underlying local-DFT, the algorithm cost is mostly due to preparing the $\langle\phi_q\phi_s|\xi\rangle$ and then calculating in each SCF iteration the exact exchange matrix elements. The steps are as follows:

- First, one Fourier-transforms, i.e., prepares, $\langle\phi_q\phi_s|k\rangle$ from $\phi_q(r)\phi_s(r)$, which costs $O(M^2N\log N)$ operations, where N is the number of the total number of grid and k points. This is the dominant expense of the entire method.
- Next, one dot-products $\langle\phi_q\phi_s|k\rangle$ with the $N_\alpha(\equiv N_{\alpha^+} + N_{\alpha^-})$ fragmented stochastic orbitals of length L each to yield $u_{\xi qs} = \langle\phi_q\phi_s|\xi^{\pm}\rangle$, at a cost of $O(M^2N_\alpha L)$ operations. For simplicity, we choose here $N_\alpha = N_{k_{low}} = N_\xi/2$. Therefore, the dot product cost is $O(M^2\cdot\mathrm{cov}\cdot N_\xi)$.
- Finally, in each of the N_{scf} iterations, one prepares the matrix elements via eqs 27 and 28, at a cost of $M^2N_vN_\xi$ operations each.

The overall cost is therefore

$$O(M^2(N_\xi\mathrm{cov} + N\log N + N_{scf}N_vN_\xi)) \tag{31}$$

Since N_ξ does not grow with system size, as demonstrated below, the dominant operation cost of the method is the production of $u_{\xi pi}$ at a cost $O(M^2N\log N) = O(N_v^2N\log N)$, i.e., a formally quadratic scaling in N_v, which is indeed observed in the left panel of Figure 4 in which we vary N_v for the largest hexamer system.

While ostensibly the $O(N_v^2N\log N)$ scaling would have indicated a cubic scaling if N_v grows linearly with system size, the actual scaling is much gentler. This is because in practice, N_v typically does not grow much with system size as the active MOs become increasingly concentrated near the HOMO−LUMO gap. When we optimize N_v to give a fixed error of the gap, leading to a system-dependent N_v, the observed scaling is, therefore, very gentle.

The gentle scaling is shown on the right panel of Figure 4, which details the operation count and total CPU time for all

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247

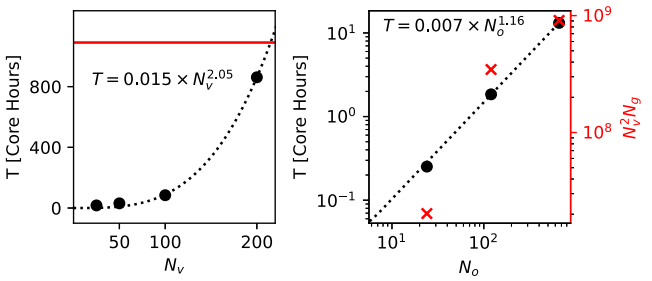


Figure 4. (Left) CPU-core hours required for the ngH-DFT method for the large hexamer complex. Parabolic scaling with the number of valence states (for a given grid) is shown. The red line indicates the core hours required for the initial LDA-DFT calculation. (Right) timing performance between systems included in this paper. N_v for each system was minimized such that the error in the gap was less than 0.01 eV, resulting in N_v = 20, 40, and 20 for the three systems in increasing N_o order, respectively.

three systems when using a minimal N_v such that the gap is converged to 0.01 eV. The number of operations and wall time both scale quasilinearly with system size.

We expect that for truly giant systems of many thousands of electrons, it would be beneficial to form a localized combination of the N_v MOs to keep this gentle scaling, but note that at the size of the hexamer system (N_o = 660 active orbitals), there was still no need for localization.

A side note is that the application of the exchange interaction formally costs $O(M^2 N_v N_\xi)$ due to matrix rotations in eq 28. However, since $N_v \ll N_g$, the rotations to apply the exchange are much faster than the generation of the integral kernels, and as such, the whole SCF procedure is only a small fraction of the total cost. In addition, the computation also requires a memory of $O(N_v^2 N_\xi)$ terms, which is actually quite small, since an advantage of the stochastic method is that the stochastic basis size, N_ξ, is, in practice, limited to about 500−5000.

DISCUSSION

We developed and demonstrated here a new method, ngH-DFT, for incorporating exact exchange within a GKS−DFT framework. Long-wavelength (low k) components of the exchange are evaluated deterministically, and high momenta are represented by a sparse stochastic basis. Using an underlying MO basis from a preliminary LDA calculation, the frontier eigenvalues converge with a small number of included valence and conduction orbitals. Given that we use MOs at a lower level of DFT as a basis set for further calculations, atomic orbital-based DFT codes can also be used to generate the initial input orbitals.

We reiterate that this method has stochasticity only in its handling of the high momenta components of the exchange, which are not as physically important as the low components. Treating less relevant degrees of freedom stochastically works very well here when combined with the sparse compression technique.

Future work will expand the method in several directions:

First, the stochastic compression gave equal weight to all high-k components and could be replaced by preferred sampling of points with relatively higher $u^\gamma(k)$ within the high$^\pm$ spaces, either explicitly or by division to several subspaces.

Next, a relatively simple extension would be to construct random combinations of the core states that would be used to calculate the core exchange. This would reduce the memory requirements since the full set of core states would not need to be stored.[21] Further, for the corrections of other states, we could use a rigid scissor approximation,[31] where all the occupied and unoccupied subspaces are shifted by the respective HOMO and LUMO orbital expectation values of X^γ_{core}, or, better yet, we could sample a few more states to determine an energy-dependent core-state contribution, analogous to our GW matrix elements.[32,33] Since it will be applied only to the core states, the contribution would be small and therefore so will its underlying stochastic error. Additionally, other approaches can be used, such as the plane-augmented wave method for pseudopotentials that more explicitly treat the core, as they only require modification of the underlying local DFT in h_0."

The present near-gap approach method will be useful for many-body perturbation theory (MBPT). In MBPT methods, having access to exact exchange-corrected eigenstates gives an improved starting point for methods such as one-shot G_0W_0 where the quality of the beginning canonical states is very important.[34,35]

Our formalism will also apply to time-dependent hybrid-DFT, where, like in GKS−DFT SCF, the $\langle \phi_q \phi_s | \xi \rangle$ vectors would be evaluated once while the exchange matrix, eqs 27 and 26, will be updated repeatedly, here once per time step. It will be useful both for real-time TDDFT and for frequency-resolved TDDFT and BSE.[36,37] We also expect applications within atomic orbital basis set-based DFT codes, where the wave function is eventually represented on a complete grid. Additionally, we anticipate that this method will have applications in auxiliary field quantum Monte Carlo methods, where the bulk of the computational effort also lies in evaluating exchange energy on many Slater determinants.[38−40]

Finally, the underlying LDA−DFT approach could be efficiently done with stochastic DFT,[41,42] so very large systems could be used, with tens of thousands of electrons or more. Eigenstates are not produced automatically in stochastic DFT, so the set of $N_v + N_c$ near-gap eigenstates, required for ngH-DFT, would be then extracted by filter-diagonalization.[43]

AUTHOR INFORMATION

Corresponding Author

Daniel Neuhauser − *Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States*; Email: dxn@ucla.edu

Authors

Nadine C. Bradbury − *Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States*; orcid.org/0000-0002-1214-113X

Tucker Allen − *Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States*

Minh Nguyen − *Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00987

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This paper was supported by the Center for Computational Study of Excited State Phenomena in Energy Materials (C2SEPEM), which is funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, via contract no. DE-AC02-05CH11231, as part of the Computational Materials Sciences Program. Computational resources were provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operating under contract no. DE-AC02-05CH11231. NCB acknowledges the NSF Graduate Research Fellowship Program under grant DGE-2034835.

■ REFERENCES

(1) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(2) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(3) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(4) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207−8215.
(5) Krukau, A. V.; Scuseria, G. E.; Perdew, J. P.; Savin, A. Hybrid functionals with local range separation. *J. Chem. Phys.* **2008**, *129*, 124103.
(6) Baer, R.; Neuhauser, D. Density Functional Theory with Correct Long-Range Asymptotic Behavior. *Phys. Rev. Lett.* **2005**, *94*, 043002.
(7) Leininger, T.; Stoll, H.; Werner, H.-J.; Savin, A. Combining long-range configuration interaction with short-range density functionals. *Chem. Phys. Lett.* **1997**, *275*, 151−160.
(8) Ren, X.; Rinke, P.; Blum, V.; Wieferink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-identity approach to Hartree−Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys.* **2012**, *14*, 053020.
(9) Aquilante, F.; Pedersen, T. B.; Lindh, R. Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. *J. Chem. Phys.* **2007**, *126*, 194106.
(10) Jung, Y.; Sodt, A.; Gill, P. M. W.; Head-Gordon, M. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6692−6697.
(11) Hollman, D. S.; Schaefer, H. F.; Valeev, E. F. Semi-exact concentric atomic density fitting: Reduced cost and increased accuracy compared to standard density fitting. *J. Chem. Phys.* **2014**, *140*, 064109.
(12) Rebolini, E.; Izsák, R.; Reine, S. S.; Helgaker, T.; Pedersen, T. B. Comparison of Three Efficient Approximate Exact-Exchange Algorithms: The Chain-of-Spheres Algorithm, Pair-Atomic Resolution-of-the-Identity Method, and Auxiliary Density Matrix Method. *J. Chem. Theory Comput.* **2016**, *12*, 3514−3522.
(13) Förster, A.; Franchini, M.; van Lenthe, E.; Visscher, L. A Quadratic Pair Atomic Resolution of the Identity Based SOS-AO-MP2 Algorithm Using Slater Type Orbitals. *J. Chem. Theory Comput.* **2020**, *16*, 875−891.
(14) Spadetto, E.; Philipsen, P. H. T.; Förster, A.; Visscher, L. Toward Pair Atomic Density Fitting for Correlation Energies with Benchmark Accuracy. *J. Chem. Theory Comput.* **2023**, *19*, 1499−1516.
(15) Damle, A.; Lin, L.; Ying, L. Compressed Representation of Kohn−Sham Orbitals via Selected Columns of the Density Matrix. *J. Chem. Theory Comput.* **2015**, *11*, 1463−1469.
(16) Giovannini, T.; Koch, H. Fragment Localized Molecular Orbitals. *J. Chem. Theory Comput.* **2022**, *18*, 4806−4813.
(17) Prentice, J. C. A.; Aarons, J.; Womack, J. C.; Allen, A. E. A.; Andrinopoulos, L.; Anton, L.; Bell, R. A.; Bhandari, A.; Bramley, G. A.; Charlton, R. J.; et al. The ONETEP linear-scaling density functional theory program. *J. Chem. Phys.* **2020**, *152*, 152.
(18) Graf, D.; Beuerle, M.; Schurkus, H. F.; Luenser, A.; Savasci, G.; Ochsenfeld, C. Accurate and Efficient Parallel Implementation of an Effective Linear-Scaling Direct Random Phase Approximation Method. *J. Chem. Theory Comput.* **2018**, *14*, 2505−2515.
(19) Ufimtsev, I. S.; Martínez, T. J. Quantum Chemistry on Graphical Processing Units. 1. Strategies for Two-Electron Integral Evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.
(20) Kussmann, J.; Laqua, H.; Ochsenfeld, C. Highly Efficient Resolution-of-Identity Density Functional Theory Calculations on Central and Graphics Processing Units. *J. Chem. Theory Comput.* **2021**, *17*, 1512−1521.
(21) Neuhauser, D.; Rabani, E.; Cytter, Y.; Baer, R. Stochastic Optimally Tuned Range-Separated Hybrid Density Functional Theory. *J. Phys. Chem. A* **2016**, *120*, 3071−3078.
(22) Vlček, V.; Li, W.; Baer, R.; Rabani, E.; Neuhauser, D. Swift GW beyond 10,000 electrons using sparse stochastic compression. *Phys. Rev. B* **2018**, *98*, 075107.
(23) Dou, W.; Chen, M.; Takeshita, T. Y.; Baer, R.; Neuhauser, D.; Rabani, E. Range-separated stochastic resolution of identity: Formulation and application to second-order Green's function theory. *J. Chem. Phys.* **2020**, *153*, 074113.
(24) Perdew, J. P.; Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244−13249.
(25) Troullier, N.; Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1991**, *43*, 1993−2006.
(26) Martyna, G. J.; Tuckerman, M. E. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. *J. Chem. Phys.* **1999**, *110*, 2810−2821.
(27) Baer, R.; Livshits, E.; Salzner, U. Tuned Range-Separated Hybrids in Density Functional Theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85−109.
(28) Förster, A.; Visscher, L. Quasiparticle Self-Consistent GW-Bethe-Salpeter Equation Calculations for Large Chromophoric Systems. *J. Chem. Theory Comput.* **2022**, *18*, 6779−6793.
(29) Sirohiwal, A.; Pantazis, D. A. The Electronic Origin of Far-Red-Light-Driven Oxygenic Photosynthesis. *Angew. Chem.* **2022**, *134*, No. e202200356.
(30) Aprà, E.; Bylaska, E. J.; de Jong, W. A.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; van Dam, H. J. J.; Alexeev, Y.; Anchell, J.; et al. NWChem: Past, present, and future. *J. Chem. Phys.* **2020**, *152*, 184102.
(31) Vlček, V.; Baer, R.; Rabani, E.; Neuhauser, D. Simple eigenvalue-self-consistent ΔGW_0. *J. Chem. Phys.* **2018**, *149*, 174107.
(32) Neuhauser, D.; Gao, Y.; Arntsen, C.; Karshenas, C.; Rabani, E.; Baer, R. Breaking the Theoretical Scaling Limit for Predicting Quasiparticle Energies: The Stochastic GW Approach. *Phys. Rev. Lett.* **2014**, *113*, 076402.
(33) Vlček, V.; Rabani, E.; Baer, R.; Neuhauser, D. Nonmonotonic band gap evolution in bent phosphorene nanosheets. *Phys. Rev. Mater.* **2019**, *3*, 064601.
(34) Bruneval, F.; Marques, M. A. L. Benchmarking the Starting Points of the GW Approximation for Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 324−329.
(35) McKeon, C. A.; Hamed, S. M.; Bruneval, F.; Neaton, J. B. An optimally tuned range-separated hybrid starting point for GW plus Bethe−Salpeter equation calculations of molecules. *J. Chem. Phys.* **2022**, *157*, 074103.

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247

(36) Bradbury, N. C.; Nguyen, M.; Caram, J. R.; Neuhauser, D. Bethe−Salpeter equation spectra for very large systems. *J. Chem. Phys.* **2022**, *157*, 031104.
(37) Bradbury, N. C.; Allen, T.; Nguyen, M.; Ibrahim, K. Z.; Neuhauser, D. Optimized attenuated interaction: Enabling stochastic Bethe−Salpeter spectra for large systems. *J. Chem. Phys.* **2023**, *158*, 154104.
(38) Rom, N.; Charutz, D.; Neuhauser, D. Shifted-contour auxiliary-field Monte Carlo: circumventing the sign difficulty for electronic-structure calculations. *Chem. Phys. Lett.* **1997**, *270*, 382−386.
(39) Carlson, J.; Gubernatis, J. E.; Ortiz, G.; Zhang, S. Issues and observations on applications of the constrained-path Monte Carlo method to many-fermion systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 12788−12798.
(40) Zhang, S. *Handbook of Materials Modeling*; Springer International Publishing, 2018, pp 1−27.
(41) Baer, R.; Neuhauser, D.; Rabani, E. Self-Averaging Stochastic Kohn-Sham Density-Functional Theory. *Phys. Rev. Lett.* **2013**, *111*, 106402.
(42) Neuhauser, D.; Baer, R.; Rabani, E. Communication: Embedded fragment stochastic density functional theory. *J. Chem. Phys.* **2014**, *141*, 041102.
(43) Wall, M. R.; Neuhauser, D. Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal mode frequencies from a small number of residues or a short-time segment of a signal. I. Theory and application to a quantum-dynamics model. *J. Chem. Phys.* **1995**, *102*, 8011−8022.

https://doi.org/10.1021/acs.jctc.3c00987
J. Chem. Theory Comput. 2023, 19, 9239−9247