Title

Large-scale Electronic Structure Method Development

Author

Nguyen, Minh

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Large-Scale Electronic Structure Method Development

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Chemistry

by

Minh Nguyen

2024

ABSTRACT OF THE DISSERTATION

Large-scale Electronic Structure Method Development

by

Minh Nguyen

Doctor of Philosophy in Chemistry

University of California, Los Angeles, 2024

Professor Daniel Neuhauser, Chair

Electronic structure theory seeks to describe the behavior of electrons in atomic and molecular systems. Due to the intractable nature of solving the molecular Schrødinger's equation, approximations are made. The main challenge is to create methods that are accurate enough to gain insight while also being efficient enough to run calculations in a reasonable amount of time. In this balancing act, many strategies have been developed to allow for electronic structure calculations of large systems. Much progress has been made from calculating the states of isolated one-electron systems to now being able to simulate dynamic processes in large extended systems. This dissertation seeks to contribute to the development of novel methods to enable more efficient large-scale electronic structure calculations. A major theme of the dissertation is the use of stochastic techniques to reduce the computational scaling of methods.

Chapter 2 discusses these techniques and highlights the improvement in computational scaling when implemented with density functional theory (DFT) and many-body perturbation theory within the GW approximation. Many improvements to stochastic DFT (sDFT) have been made over the years, incorporating techniques such as embedding to reduce the required number of

statistical samples. Chapter 3 continues in the same line of work and introduces the concept of tempering and its application in sDFT. The core idea of tempering is to rewrite the electronic density into the sum of a cheaper "warm" term and a smaller more expensive "cold" term. This results in a significant reduction in the statistical fluctuations and systematic deviation compared to sDFT for the same computational effort.

Chapter 4 discusses the gapped filtering method and its application in the stochastic GW (sGW) approximation. In gapped-filtering, a short Chebyshev expansion accurately represents the density-matrix operator. The method optimizes the Chebyshev coefficients to give the correct density matrix at all energies except within the gapped region where there are no eigenstates. Gapped filtering reduces the number of required terms in the Chebyshev expansion compared to traditional expansion methods, as long as one knows or can efficiently determine the HOMO and LUMO positions such as in sGW.

Another direction in this dissertation is laying the foundations to implement the projector augmented wave (PAW) method into stochastic quantum methods. Compared to norm-conserving pseudopotentials (NCPP), PAW has the advantage of lower kinetic energy cutoffs and larger grid spacing at the cost of having to solve for non-orthogonal wavefunctions. Orthogonal PAW (OPAW) was earlier developed with DFT to allow the use of PAW when orthogonal wavefunctions are desired. To make OPAW viable for post-DFT stochastic methods, time-dependent wavefunctions are required. For this purpose, chapters 5 and 6 detail OPAW and its implementation in the real-time time-dependent (TD) DFT framework.

The dissertation of Minh Nguyen is approved.

Peter M. Felker

Benjamin J. Schwartz

Justin R. Caram

Daniel Neuhauser, Committee Chair

University of California, Los Angeles

2024

This thesis is dedicated to family, friends, mentors, and God for supporting me.

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I express much gratitude to my advisor Daniel Neuhauser, whose support and guidance has allowed me to flourished during my doctoral studies. Under his mentorship, my skills as a scientist and programmer have grown immensely. I also acknowledge the current and former members of the Neuhauser group for their friendship and the good memories they have provided. In particular, special thanks to Yangtao (Barry) Li and Tim Duong who I have both mentored and published papers with. The support of my family was instrumental in helping me finish my studies. Finally, I give thanks to my future wife Alery Tenorio who has been along side me in my journey since my undergraduate studies.

VITA

9/2015-6/2019	B.S. in Chemical Physics
	Department of Chemistry
	University of California, Davis
9/2019-9/2021	M.S. in Chemistry
	Department of Biochemistry and Chemistry
	University of California, Los Angeles
9/2019-present	Ph.D. in Chemistry
	Department of Biochemistry and Chemistry
	University of California, Los Angeles

PUBLICATIONS

1. Nguyen, M., Duong, T. & Neuhauser, D. Time-dependent density functional theory with the orthogonal projector augmented wave method. *The Journal of Chemical Physics* **160** (2024)

2. Bradbury, N. C., Allen, T., Nguyen, M. & Neuhauser, D. Deterministic/Fragmented-Stochastic Exchange for Large-Scale Hybrid DFT Calculations. *Journal of Chemical Theory and Computation* **19,** 9239–9247 (2023)

3. Bradbury, N. C., Allen, T., Nguyen, M., Ibrahim, K. Z. & Neuhauser, D. Optimized attenuated interaction: Enabling stochastic Bethe–Salpeter spectra for large systems. *The Journal of Chemical Physics* **158** (2023)

4. Nguyen, M. & Neuhauser, D. Gapped-filtering for efficient Chebyshev expansion of the density projection operator. *Chemical Physics Letters* **806,** 140036 (2022)

5. Yeo, C., Nguyen, M. & Wang, L.-P. Benchmarking Density Functionals, Basis Sets, and Solvent Models in Predicting Thermodynamic Hydricities of Organic Hydrides. *The Journal of Physical Chemistry A* **126,** 7566–7577 (2022)

6. Bradbury, N. C., Nguyen, M., Caram, J. R. & Neuhauser, D. Bethe–Salpeter equation spectra for very large systems. *The Journal of Chemical Physics* **157** (2022)

7. Nguyen, M., Li, W., Li, Y., Rabani, E., Baer, R. & Neuhauser, D. Tempering stochastic density functional theory. *The Journal of Chemical Physics* **155** (2021)

8. Jang, H., Qiu, Y., Hutchings, M. E., Nguyen, M., Berben, L. A. & Wang, L.-P. Quantum chemical studies of redox properties and conformational changes of a four-center iron CO 2 reduction electrocatalyst. *Chemical Science* **9,** 2645–2654 (2018)

1 Introduction

The modern electronic structure problem consists of solving the nonrelativistic Schrödinger equation for atoms and molecules under the Born-Oppenheimer approximation where the are nuclei fixed in space. The equation relates the Hamiltonian, $\hat{H}$, which contains information necessary for calculating the total energy of a many-body system, to the wavefunctions, Ψ, which provide mathematical descriptions of quantum states. For molecular systems, the time-independent Schrödinger equation under the Born-Oppenheimer approximation is

$$\hat{H}\Psi = [\hat{T} + \hat{V}_{ext} + \hat{V}_{ee}]\Psi = \left[-\sum_i \frac{1}{2}\nabla_i^2 + v_{ext} + \sum_{i<j} \frac{1}{r_i - r_j} \right]\Psi, \tag{1}$$

where i and j are indices over the atoms of the system, $\hat{T}$ is the kinetic energy operator, $\hat{V}_{ext}$ is the external potential that comes from the interaction of the electrons with the atomic nuclei and applied electric fields, and $\hat{V}_{ee}$ is the electron-electron interaction. All units used in this dissertation are atomic units unless otherwise noted. The many-electron wavefunction, $\Psi(r_1, ..., r_N)$, depends on the coordinates of all electrons in the system simultaneously. Because of the last term in the Hamiltonian, the many-body equation is not separable into one-body equations making, analytic solutions nearly impossible except for single electron systems such as H_2^+. Thus it is necessary to employ further approximations to make the problem more tractable.

Many electric structure theories start with the Hartree-Fock (HF) method where Ψ is approximated by a single Slater determinant and the total energy minimized. The HF method is an example of a mean-field theory in which the complicated electron-electron interactions are replaced with an electron interacting with an effective potential created by the other electrons. This scheme reduces the many-body problem into one-body equations, greatly simplifying the calculation. In HF, the wavefunctions are solutions to the HF equations, but the equations themselves are defined by the wavefunctions i.e. the equations require self-consistency. The self-consistent field (SCF) method

iteratively solves the HF equations by creating a first-iteration guess Hamiltonian to calculate a set of wavefunctions which then are used to calculate a second-iteration Hamiltonian for the next SCF cycle. This cycle repeats until the wavefunctions are converged, meaning an error metric such as the difference in the total energy between cycles does not change significantly between SCF cycles.

There are many theories that use the HF method as their starting point including configuration interaction, Møller-Plesset perturbation theory, and coupled cluster. These theories tend to be more suitable for smaller systems such as gas-phase molecules while other theories have to be used for larger systems. Density functional theory (DFT), another mean-field theory, has found large success in calculating properties and processes in solids and larger molecular systems.

1.1 Density functional theory

The core idea of DFT is that ground state properties can be exactly determined by the electronic density, $n(r)$. The foundations of DFT are the Hohenberg-Kohn theorems[9]. The first theorem states that the ground state of any interacting particle system is a functional of the density. The second theorem shows that the ground state density minimizes the energy functional. Although exact in theory, the explicit form of the energy functional has not been discovered, thereby, necessitating approximations. The orbital-free DFT approach tries to directly approximate the energy functional; however, it suffers from a lack of accuracy mainly due to the lack of good approximations to the kinetic energy functional.

Kohn and Sham (KS) proposed an alternative method in their seminal work in which they convert the many-body problem to a set of fictitious KS states, $|\psi_i\rangle$, interacting in a mean-field whose density is equal to that of the fully interacting problem [10]. With these fictitious states, the accuracy of approximating the kinetic energy was greatly improved compared to orbital-free DFT.

The total electronic energy in the KS framework can be expressed as

$$E[n] = -\frac{1}{2}\sum_i \langle\psi_i|\nabla^2|\psi_i\rangle + E_H[n] + E_{ext}[n] + E_{XC}[n], \tag{2}$$

where $E_H[n]$ is the Hartree energy that is derived from classical electrostatic interactions, $E_{ext}[n]$ is the energy from the interaction of electrons with the nuclei and external electric fields, and $E_{XC}[n]$ encapsulates the remaining energy. The majority of $E_{XC}[n]$ comes from electron-electron exchange and correlation effects and some from the difference between the true kinetic energy and the fictitious one.

Much work has been done to develop approximations to the exchange correlation functional, the simplest being the local density approximation (LDA) where the functional is a simple function of the density. More complex approximations include non-local information such as the generalized gradient approximations (GGA) which uses the gradient of the density, and meta-GGAs which further includes the kinetic energy density.

KS-DFT has been very successful in a wide range of problems in chemistry, physics, and biology [11, 12]. The awarding of the Nobel prize to Walter Kohn for his work highlights the significance and impact of DFT [13]. The KS equations are solved for the KS states and are derived through taking functional derivatives of the energy functional in Eq. (2):

$$\hat{H}_{KS}\,|\psi_i\rangle = \epsilon_i\,|\psi_i\rangle\,, \tag{3}$$

where the KS Hamiltonian is

$$\hat{H}_{KS} = -\frac{1}{2}\nabla^2 + v_{ext} + V_H[n] + V_{XC}[n]. \tag{4}$$

The potential terms are functional derivatives of their corresponding energy functionals. From

these KS states, the electronic density is defined as

$$n(r) = \sum_i f_i |\psi_i(r)|^2, \tag{5}$$

where f_i is the occupation number of the i'th state. For spin unpolarized systems, $f_i = 2$ for $i \leq N_{\text{occ}}$ where N_{occ} is the number of occupied states.

The KS equations are solved using a self consistent field similar to HF. In the traditional KS formulation, matrix diagonalization of the Hamiltonian is employed to solve for the eigenstates. This diagonalization gives a nominal scaling of $O(N^2)-O(N^3)$ where N is the number of particles in the system.

1.2 GW Approximation

For applications in which electron correlation is significant such as strongly correlated materials and excited states, more sophisticated theories than DFT are needed. For example, DFT calculations of the fundamental band gap tend to underestimate experimental values as much as 1-3 eV or more [14]. The fundamental band gap is the energy difference of between adding and subtracting one electron from a system, i.e., the difference between the first ionization energy and the first electron affinity of a neutral solid. Band gaps are experimentally measured using (inverse) photoemission spectroscopy in which electrons are (added to) ejected from a system. In these processes, the independent particle picture breaks down due to strong Coulomb interactions. Post-DFT methods such as many-body perturbation theory (MBPT) provide better estimates of the gap.

Before expanding on MBPT, it is useful to introduce the concept of quasiparticles (QPs). QPs are fictitious particles that describe the collective phenomena of interacting systems and behave similarly to independent particles. For an example of a QP, consider a photoemission experiment in which an electron is ejected. Due to exchange and correlation effects such as electron repulsion, the resulting collective behavior of the remaining electrons create a Coulomb hole that can be

treated as a positively charged independent particle. The interaction between holes is attenuated by the reduction in total charge and thus the interaction between QPs are screened and weaker than the bare Coulomb interaction between electrons. The effective interaction is small enough that the independent particle approximation is justified for QPs.

The MBPT approach describes the behavior of QPs and accounts for electron correlation by treating it as a perturbation of the HF states. A key quantity in MBPT is the Green's function, G, which describes the evolution of system when particles are added and removed from it. Hedin derived a set of equations (Hedin's equations) that expand many-body Green's function in terms of simpler quantities, in particular, the non-interacting Green's function G_0 [15]. From Hedin's equations comes the QP equation:

$$\left(-\frac{1}{2}\nabla^2 + v_{ext}(r) + v_H(r) \right)\psi_n^{QP}(r) + \int \Sigma(r, r'; \epsilon_n^{QP})\psi_n^{QP}(r')dr' = \epsilon_n^{QP}\psi_n^{QP}(r) \tag{6}$$

where the QP energies, ϵ_n^{QP}, describe the energy to add or remove a particle and are used to estimate the fundamental band gap. The self-energy, $\Sigma(r, r', \omega)$, contains the many-body exchange and correlation effects beyond the electrostatic Hartree potential and accounts for the dynamic many-electron processes.

One of the most successful approximations for the self-energy is the GW approximation in which the time-domain self-energy is the product of the Green's function the screened Coulomb interaction, W:

$$\Sigma(r, r', t) \approx iG(r, r', t)W(r, r', t). \tag{7}$$

The time-domain $\Sigma(r, r', t)$ is related to energy-domain $\Sigma(r, r', \omega)$ via a Fourier transform:

$$\Sigma(r, r', t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Sigma(r, r', \omega)e^{-i\omega t}d\omega. \tag{8}$$

The screened Coulomb interaction describes the attenuated interaction between QPs. Another

widely used approximation is the single-shot G_0W_0 approximation in which the non-interacting Green's function, $G_0(r, r', t)$, and the first-iteration screened Coulomb interaction, $W_0(r, r', t)$, is used to further simplify the calculation of Σ [16]. Specifically, $W_0(r, r', t)$ is the potential at point r and time t due to a QP introduced at time $t = 0$ at point r'. It is composed of an instantaneous Coulomb term and a time-dependent polarization term:

$$W_0(r, r', t) = u_C(|r - r'|)\delta(t) + W_P(r, r', t), \tag{9}$$

where $u_C(r) = r^{-1}$ is the Coulomb potential. The latter term is the polarization potential due to the density perturbation of the QP:

$$W_P(r, r', t) = \int\int u_C(|r - r''|)\chi(r'', r''', t)u_C(|r''' - r'|)dr''dr''', \tag{10}$$

where $\chi(r, r', t)$ is the reducible polarization function.

In the G_0W_0 picture, the QP energies can be treated as perturbative corrections to the KS-DFT eigenvalues, ϵ_i^{KS} [17]:

$$\epsilon_i^{QP} = \epsilon_i^{KS} + \langle\psi_i^{KS}|\Sigma^P(\omega = \epsilon_i^{QP}) + \Sigma^X - v_{XC}|\psi_i^{KS}\rangle\,, \tag{11}$$

where Σ^P is the polarization self-energy, Σ^X is the exchange self-energy, v_{XC} is the DFT exchange-correlation potential, and $|\psi_i^{KS}\rangle$ are the converged eigenstates of a KS-DFT calculation. The first two terms come from Eq. (9) which results in the expectation value of the self-energy to be the sum of an instantaneous and a time-dependent contribution [18]:

$$\langle\psi_i^{KS}|\Sigma|\psi_i^{KS}\rangle = \langle\psi_i^{KS}|\Sigma^X|\psi_i^{KS}\rangle + \langle\psi_i^{KS}|\Sigma^P|\psi_i^{KS}\rangle\,. \tag{12}$$

The instantaneous exchange self-energy can be written as

$$\langle \psi_i^{KS} | \Sigma^X | \psi_i^{KS} \rangle = - \int \int \psi_i^{KS}(r) u_C(|r - r'|) \rho^{KS}(r, r') \psi_i^{KS}(r') dr dr', \tag{13}$$

where $\rho^{KS}(r, r') = -iG_0(r, r', 0^-) = \sum_n f_n \psi_n^{KS}(r) \psi_n^{KS}(r')$ is the KS density matrix where 0^- is a time infinitesimally earlier than 0.

Hence, the exchange self-energy is the expectation value of the Fock exchange operator. The polarization part is given by

$$\langle \psi_i^{KS} | \Sigma^P | \psi_i^{KS} \rangle = \int \int \psi_i^{KS}(r) iG_0(r, r', t) W_P(r, r', t^+) \psi_i^{KS}(r') dr dr', \tag{14}$$

where t^+ is a time infinitesimally later than t.

The G_0W_0 approximation has been shown to greatly improve the calculated band gap for many systems [19–21]. In traditional applications of GW, the calculation of the screened Coulomb potential is especially demanding, leading to $\mathcal{O}(N^4)$ scaling [22–25].

1.3 Outline of dissertation

The main theme of this dissertation is developing large-scale electronic structure methods primarily through using stochastic sampling to reduce the computational scaling. Each project involves developing novel numerical techniques to advance the computational efficiency of these stochastic quantum chemistry methods. The core idea behind these methods is the stochastic resolution of identity, expanded upon in chapter 2.1. Applications of this technique are discussed in the context of stochastic density functional theory (sDFT) and the stochastic GW (sGW) methods in sections 2.2 and 2.3 respectively.

The first major direction of this dissertation is to improve upon sDFT with embedding techniques to expand the density using operators of varying temperatures. This expansion allows for

the use of more stochastic sampling than would be possible in traditional sDFT and is discussed in chapter 3. The second direction is in a similar vein with the goal of shortening the Chebyshev expansion of the occupation operator given knowledge of the band gap. The gapped-filtering technique greatly improves computational efficiency with a very small overhead of matrix diagonalization and is discussed in chapter 4.

The final project is to expand on the previously developed orthogonal projector augmented wave method (OPAW). The PAW method offers the advantage of better computational costs compared to the norm-conserving pseudopotential approach with the trade-off of complexity in implementing it in theory and code. OPAW was developed to simplify PAW and was successfully implemented with DFT. Chapter 6 details the development of OPAW in the time-dependent DFT (TDDFT) framework. The OPAW-TDDFT method serves as a precursor to implementing PAW into other stochastic quantum chemistry methods namely sGW and the stochastic Bethe-Salpeter equation (sBSE).

1.4 Reprint of published articles and manuscripts

Chapter 3 is adapted with permission from: Nguyen, M., Li, W., Li, Y., Rabani, E., Baer, R. & Neuhauser, D. Tempering stochastic density functional theory. *The Journal of Chemical Physics* **155** (2021). Chapter 4 is adapted with permission from: Nguyen, M. & Neuhauser, D. Gapped-filtering for efficient Chebyshev expansion of the density projection operator. *Chemical Physics Letters* **806,** 140036 (2022). Chapter 6 is adapted with permission from: Nguyen, M., Duong, T. & Neuhauser, D. Time-dependent density functional theory with the orthogonal projector augmented wave method. *The Journal of Chemical Physics* **160** (2024).

2 Stochastic Electronic Structure Methods

2.1 Quantum Stochastic Techniques

The stochastic framework seeks to reduce the scaling of electronic structure methods, using statistical sampling to replace the explicit calculation and use of thousands of eigenstates or more. The stochastic resolution of the identity (sRI) is a numerical technique that serves as the foundation for these stochastic quantum chemistry methods [26]. With the sRI, the summation over a large number of eigenstates is replaced by an average of a much smaller number of stochastic states.

A stochastic state, $\chi(r) = \langle r|\chi\rangle$, has its value at every point, r_i, as an independent random variable, $\eta_i = \chi(r_i)$. For real variables, η_i is drawn from $\{\pm 1\}$ with uniform probability and for complex variables, η_i is drawn from the complex unit circle $e^{i\theta}$ where θ is a random angle from $\{0, 2\pi\}$. The average of η_i over an infinite number of $|\chi\rangle$ samples, $\mathrm{E}[\eta_i]_\chi$, is equal 0 as for every η_i there is an equal probability of drawing $-\eta_i$. Now consider the average of the product of $\eta_i^\dagger = \langle\chi|r_i\rangle$ and $\eta_j = \langle r_j|\chi\rangle$, $\mathrm{E}[\eta_i^\dagger\eta_j]_\chi$. For $r_i = r_j$, $\mathrm{E}[\eta_i^\dagger\eta_i]_\chi = 1$ while for $r_i \neq r_j$, $\mathrm{E}[\eta_i^\dagger\eta_i]_\chi = \mathrm{E}[\eta_i^\dagger]_\chi\mathrm{E}[\eta_j]_\chi = 0$. The latter relationship is derived from the fact that η_i and η_j are independent random variables. Thus $\mathrm{E}[\eta_i^\dagger\eta_j] = \delta_{ij}$. Another way to express this idea is that the average of the auto-correlation function of a set of stochastic states is equal to the identity operator in the position basis:

$$\mathrm{E}[\eta_i\eta_j^\dagger]_\chi = \mathrm{E}[\langle r_i|\chi\rangle\,\langle\chi|r_j\rangle]_\chi = \langle r_i|\hat{I}|r_j\rangle\,. \tag{15}$$

From the above equation, we extract that the identity operator is

$$\hat{I} = \mathrm{E}[|\chi\rangle\,\langle\chi|]_\chi. \tag{16}$$

With this sRI, the trace of operators is then expressed with stochastic states as [27]

$$\text{Tr}[A] = E[\langle\chi|A|\chi\rangle]_\chi. \tag{17}$$

These stochastic states are random linear combinations of all the occupied and unoccupied eigenstates of the Hamiltonian:

$$|\chi\rangle = \sum_i a_i \, |\psi_i\rangle \, , \tag{18}$$

where the coefficients, $a_i = \langle\chi|\psi_i\rangle$, are random variables that have the property $E[a_i a_j]_\chi = \delta_{ij}$.

In practice, a discrete grid is used so η_i is multiplied by a factor of $dv^{-1/2}$ for normalization where dv is the volume element. In addition, a finite average has to be performed leading to statistical errors that are proportional to $N_\chi^{-1/2}$ where N_χ is the number of samples. For acceptable statistical errors, values of N_χ in the tens or hundreds tend to be sufficient. The identity operator can of course also be expressed with the complete set of orthonormal eigenstates of the Hamiltonian:

$$\hat{I} = \sum_i |\psi_i\rangle \, \langle\psi_i| \, . \tag{19}$$

Compared to the sRI, the number of basis states can be in the thousands for large systems making the stochastic framework more efficient.

2.2 Stochastic Density Functional Theory

Stochastic DFT (sDFT) uses these techniques to greatly improve scaling. The first step in deriving sDFT is to express the electronic density (assuming unpolarized molecules) as a trace:

$$n(r) = 2\text{Tr}[\hat{\rho}\delta(r - r')]. \tag{20}$$

The operator $\hat{\rho}$ is an occupation operator that can be expressed as the sum of projector functions of the occupied KS states and can also be written as a low-pass energy filter:

$$\hat{\rho} = \sum_{i \leq N_{\text{occ}}} |\psi_i\rangle \langle\psi_i| = \Theta(\mu - \hat{H}_{KS}). \tag{21}$$

Using the sRI, we rewrite the electronic density as

$$n(r) = 2\text{E}\Big[\langle r|\sqrt{\hat{\rho}}|\chi\rangle \langle\chi|\sqrt{\hat{\rho}}|r\rangle \Big]_{\chi} = 2\text{E}[\langle r|\xi\rangle \langle\xi|r\rangle] = 2\text{E}[|\xi(r)|^2]_{\chi}, \tag{22}$$

where $|\xi\rangle = \sqrt{\hat{\rho}}\,|\chi\rangle$ are the occupied stochastic states, made from projecting out the unoccupied contributions from $|\chi\rangle$. The square root of $\hat{\rho}$ is used so the calculated $n(r)$ is positive at every grid point when the absolute value of $\xi(r)$ is taken. To apply $\sqrt{\hat{\rho}}$, which is a function of the Hamiltonian, a polynomial expansion is used. The chemical potential, μ, is chosen so that the density integrates to the number of electrons in the system,

$$N_e = \int n(r) dr. \tag{23}$$

More details on expanding $\sqrt{\hat{\rho}}$ and determining μ are discussed in chapter 3.

Because in practice a finite number of samples is used, the density has an associated statistical error which in turn produces error in quantities such as the total energy. Despite this error, it has been shown that sDFT can still achieve sublinear scaling [28, 29]. Part of this success is in a "self-averaging" phenomena where fewer stochastic samples are required as the system grows in size [28]. There has been previous work to reduce the error such as embedding the sDFT calculation with smaller deterministic KS-DFT calculations [30, 31]. The work outlined in chapter 3 follows in the same theme of seeking to reduce the statistical error using embedding techniques.

2.3 Stochastic GW Approximation

These stochastic techniques have also been used to develop post-DFT methods, specifically the GW method. In sGW, the sRI is used multiple times to efficiently calculate G_0 and W_0. To derive sGW, first consider the KS Green's function which can expressed as

$$iG_0(r,r',t) = \langle r|\, e^{-i\hat{H}_{KS}t}\big[(\hat{I}-\hat{\rho})\theta(t)+\hat{\rho}\theta(-t)\big]\,|r'\rangle\,, \tag{24}$$

where θ is the Heaviside step function. Inserting the sRI into the above expression gives

$$\begin{aligned} iG_0(r,r',t) &= \mathrm{E}\Big[\,\langle r|\, e^{-i\hat{H}_{KS}t}\big[(\hat{I}-\hat{\rho})\theta(t)+\hat{\rho}\theta(-t)\big]\,|\chi\rangle\,\langle\chi|r'\rangle\,\Big]_\chi, \\ &= \mathrm{E}\big[\zeta_\mu(r,t)\zeta(r')\big]_\chi \end{aligned} \tag{25}$$

where $\zeta_\mu(r,t<0) = \langle r|e^{-i\hat{H}_{KS}t}\hat{\rho}|\chi\rangle$ and $\zeta_\mu(r,t>0) = -\langle r|e^{-i\hat{H}_{KS}t}(\hat{I}-\hat{\rho})|\chi\rangle$ are projected occupied and unoccupied stochastic states propagated backward and forward in time respectively.

With this stochastic representation for G_0, the exchange self-energy in Eq. (13) is

$$\langle\psi_i^{KS}|\Sigma^X|\psi_i^{KS}\rangle = -\mathrm{E}\bigg[\int \psi_i^{KS}(r)\zeta_\mu^*(r,0^-)v_\zeta^{\text{aux}}(r)dr\bigg]_\chi, \tag{26}$$

where the auxiliary potential is

$$v_\zeta^{\text{aux}}(r) = \int u_C(|r-r'|)\zeta(r')\psi_i^{KS}(r')dr'. \tag{27}$$

The polarization self-energy from Eq. (10) can be rewritten as

$$\langle\psi_i^{KS}|\Sigma_P|\psi_i^{KS}\rangle = \mathrm{E}\bigg[\int \psi_i^{KS}(r)\zeta_\mu^*(r,t)u(r,t)dr\bigg]_\chi. \tag{28}$$

where the response potential $u(r,t)$ is due to the perturbation of the pseudo charge density $\zeta(r')\psi_i^{KS}(r')$

at time $t = 0$:

$$u(r,t) = \int W_P(r,r',t)\zeta(r')\psi_i^{KS}(r')dr'. \tag{29}$$

To calculate the action of $W_P(r,r',t)$ on $\zeta(r')\psi_i^{KS}(r')$, time-dependent Hartree (TDH) propagation is used. TDH can only provide description for the retarded $u^r(r,r',t)$ rather than the full time-ordered $u(r,r',t)$. An explanation of how to convert from $u^r(r,r',t)$ to $u(r,r',t)$ will be given after some derivation for the former.

In the traditional TDH formulation, the entire occupied space is used, but with stochastic techniques, this can be avoided with stochastic TDH (sTDH). For each $|\chi\rangle$, a set of $\sim 5-30$ completely random orbitals, $\{\xi(r)\}$, are generated and filtered $|\varphi\rangle = \hat{\rho}\,|\xi\rangle$ to perform sTDH. A set of perturbed states are made from $|\varphi\rangle$:

$$\varphi^\lambda(r) = e^{-i\Delta(r)\lambda}\varphi(r), \tag{30}$$

where

$$\Delta(r) = \int u_C(|r-r'|)\zeta(r')\phi_i^{KS}(r')dr', \tag{31}$$

and λ is a small factor, typically $10^{-4}E_h^{-1}$. These perturbed stochastic states, when averaged, are equivalent to perturbing all the occupied states simultaneously. The perturbed (λ) and unpertubed ($\lambda = 0$) states are then propagated in time under the sTDH equations [32]:

$$-i\frac{\partial}{\partial t}\varphi^\lambda(r,t) = \left[\hat{H}_{KS} + V_H[n^\lambda(r,t)](r) - V_H[n^\lambda(r,t=0)](r)\right]\varphi^\lambda(r,t), \tag{32}$$

where $V_H[n(r)](r) = \int u_C(|r-r'|)n(r')dr'$ and $n^\lambda(r,t) = 2\mathrm{E}[|\varphi^\lambda(r,t)|^2]_\xi$. From these propagated orbitals, the retarded response function is calculated:

$$u^r(r,t) = \frac{V_H[n^\lambda(r,t)](r) - V_H[n^{\lambda=0}(r,t)](r)}{\lambda}. \tag{33}$$

The full response is obtained through a Fourier transform of $u^r(r,t)$ to ω space, then calculating

$$u(r,\omega) = \begin{cases} u^r(r,\omega), & \omega \geq 0 \\ u^r(r,\omega)^*, & \omega < 0 \end{cases}. \tag{34}$$

An inverse Fourier transform is then performed on $u(r,\omega)$ to obtain $u(r,t)$.

With all these ingredients, sGW has been shown to efficiently calculate quasiparticle energies, reproducing the results of traditional GW while achieving linear and even sub-linear scaling, being applied to study systems with more than 10,000 electrons [32, 33].

3 Tempering Stochastic DFT

Kohn-Sham density functional theory (KS-DFT) is widely used for calculating the properties of molecular and extended systems [34]. In particular, the method is useful for determining the structure based on the estimates it provides for the forces on the corresponding nuclei [35–37]. However, applying KS-DFT for systems with hundreds or thousands of atoms is challenging due to the high scaling of computational costs with a system's size (potentially quadratic but eventually cubic for large systems). Lower scaling implementations of the theory have been developed for systems that have a density matrix that is fairly sparse so that only a linear-scaling near-diagonal portion of the matrix needs to be processed. Because of the restriction to near-localized density matrices, the use of such methods is often limited to low-dimensional structures [38, 39] or systems with strictly localized electrons [40, 41].

In previous work, we introduced stochastic density functional theory (sDFT) [28] which avoids the costly diagonalization step in KS-DFT without the need to make a locality assumption; instead, the density matrix is approximated statistically. Specifically, the density matrix in sDFT is viewed as a correlation function of stochastic functions, each of which is, in essence, a random combination of the occupied states. While the method scales linearly, the tradeoff is the introduction of statistical uncertainties in the density and other observables. The statistical errors can be reduced by using an embedded-fragmented (ef-sDFT) technique [42–44] which is based on dividing the system into fixed-size fragments and expressing the total electron density, $n(\boldsymbol{r})$, as the sum of fragment densities plus a correction term which is evaluated stochastically. This technique reduces the statistical fluctuations in the estimates of the atomic forces and the energies [42, 43], and the magnitude of this reduction is controlled by varying the size of the fragments and the number of stochastic realizations. An additional approach for mitigating the statistical errors is the energy-window sDFT (ew-sDFT) scheme [29] and its combination with the embedded-fragmented technique [42, 45].

Here we propose a tempering method, referred to as t-sDFT, as a complementary technique for reducing the statistical noise. In t-sDFT, the density for the desired temperature is calculated using a higher-temperature reference density with smaller correction(s). This idea has been implemented before within the energy renormalization group in the context of telescopically expanding the Hamiltonian matrix in a series [46]. In Sec. 3.1, we describe the t-sDFT method and in Sec. 3.2 we benchmark and analyze its efficacy using large hydrogenated silicon clusters.

3.1 Methodology

3.1.1 Stochastic Density Functional Theory

Our starting point is the following expression for the electron density, $n(\boldsymbol{r})$ (assuming a spin-unpolarized system) [28]:

$$n(\boldsymbol{r}) = 2 \times \mathrm{Tr}\left[\sqrt{\hat{\rho}_\beta}\,|\boldsymbol{r}\rangle\,\langle\boldsymbol{r}|\,\sqrt{\hat{\rho}_\beta}\right], \tag{35}$$

where $\boldsymbol{r}$ is a point on a 3D grid that spans the space containing the electron density of the system and has a volume element dV. The operator

$$\hat{\rho}_\beta = f_{\beta\,\mu}\left(\hat{h}\right) \tag{36}$$

is a low band-pass Fermi-Dirac (FD) filter. Our main interest in this work is zero-temperature DFT; however, to efficiently represent the density matrix, a smooth step function must be used, and the simplest is a Fermi-Dirac distribution, $f_{\beta\,\mu}(\varepsilon) = \left(1 + e^{\beta(\varepsilon-\mu)}\right)^{-1}$, which blocks high energies ($\varepsilon > \mu + \beta$). Here, the chemical potential, μ, must be adjusted such that the integrated density equals the number of electrons, N_e,

$$\int n(r)dV = \sum_{\boldsymbol{r}} n(\boldsymbol{r})\,dV = N_e, \tag{37}$$

while β is the inverse temperature (note that the filter, $\hat{\rho}_\beta$, depends on the chemical potential but to avoid a plethora of indices we do not explicitly note this dependence below). In the low-temperature limit ($\beta\varepsilon_{\mathrm{gap}} \gg 1$, where $\varepsilon_{\mathrm{gap}}$ is the fundamental KS gap), $n(\boldsymbol{r})$ indeed converges to the ground state KS-density. Further, as our goal here is not finite-temperature DFT, we use the usual zero-temperature DFT exchange-correlation functionals.

In Eq. (36), the Kohn-Sham Hamiltonian is:

$$\hat{h} = \hat{t} + \hat{v}[n](\boldsymbol{r}) \tag{38}$$

where $\hat{t}$ is the kinetic energy operator and $\hat{v}[n](\boldsymbol{r})$ is the density-dependent KS potential that is composed of electron-nuclei, Hartree, and exchange-correlation components. Eqs. (35)-(38) must be solved simultaneously to yield the self-consistent electron density.

In sDFT, we introduce a stochastic resolution of the identity [26] which transforms the trace in Eq. (35) to an expectation value [28]

$$n(\boldsymbol{r}) = 2\left\langle \left| \left\langle \boldsymbol{r} \left| \sqrt{\hat{\rho}_\beta} \right| \chi \right\rangle \right|^2 \right\rangle_\chi , \tag{39}$$

where $|\chi\rangle$ is a stochastic orbital taking the randomly signed values $\langle \boldsymbol{r} | \chi \rangle = \pm (dV)^{-1/2}$.

To apply the filter, $\sqrt{\hat{\rho}_\beta}$, we use a Chebyshev expansion of length K [47]

$$\sqrt{\hat{\rho}_\beta} |\chi\rangle = \sum_{k=0}^{K} c_k(\beta, \mu) \left| \zeta^{(k)} \right\rangle , \tag{40}$$

where $\left|\zeta^{(k)}\right\rangle = T_k(\hat{h}_s)|\chi\rangle$ are defined by the Chebyshev polynomial recursion relations: $\left|\zeta^{(0)}\right\rangle = |\chi\rangle$, $\left|\zeta^{(1)}\right\rangle = |\chi\rangle$ and $\left|\zeta^{(k+1)}\right\rangle = 2\hat{h}_s \left|\zeta^{(k)}\right\rangle - \left|\zeta^{(k-1)}\right\rangle$. Here, $\hat{h}_s = \left(\hat{h} - \bar{\varepsilon}\right)/\Delta\varepsilon$ is a normalized KS Hamiltonian where $\bar{\varepsilon}$ and $\Delta\varepsilon$ are chosen such that the spectrum of $\hat{h}_s$ lies within the interval $[-1, 1]$, $T_k(x)$ is the k'th Chebyshev polynomial, and $c_k(\beta, \mu)$ are the Chebyshev expansion coefficients of the filter $\sqrt{f_\beta(\varepsilon)}$[47]. The expansion length K terminates the series when $|c_{k>K}|$ is smaller than

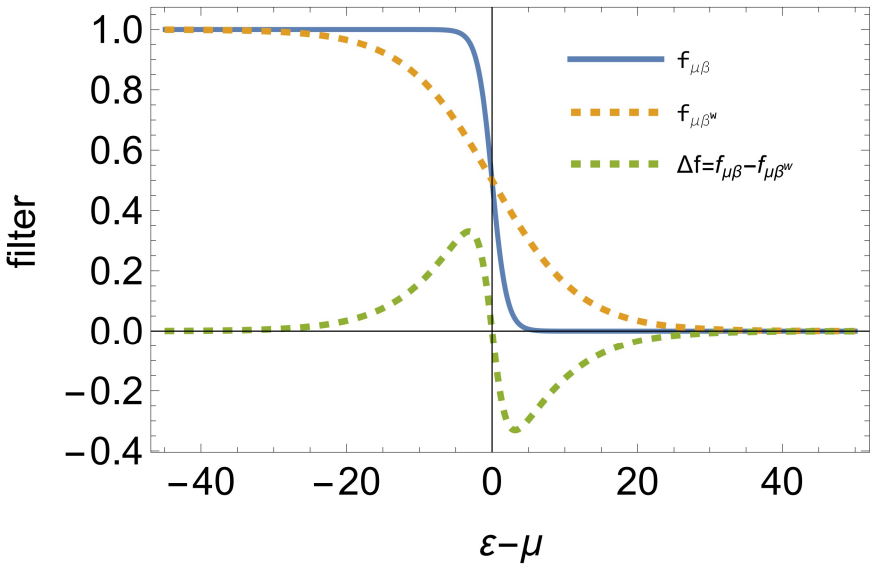

Figure 1: The desired low-temperature filter, $f_{\beta\,\mu}\left(\varepsilon\right)$, the high-temperature filter $f_{\beta^w\,\mu}\left(\varepsilon\right)$, and the correction for $\beta = 6\beta^w$.

a predetermined cutoff value.

In practice, the expected value appearing in Eq. (39) is evaluated approximately by taking a finite sample of N_s stochastic orbitals:

$$n\left(\boldsymbol{r}\right) \approx \frac{2}{N_s}\sum_{i=1}^{N_s}\left|\left\langle \boldsymbol{r}\left|\sqrt{\hat{\rho}_\beta}\right|\chi_i\right\rangle\right|^2. \tag{41}$$

Furthermore, to ensure $N_e = \sum_{\boldsymbol{r}} n\left(\boldsymbol{r}\right) dV$, we tune the chemical potential μ to satisfy the relation:

$$N_e = 2\sum_{k=0}^{K} b_k\left(\beta,\mu\right) M_k, \tag{42}$$

where

$$M_k = \frac{1}{N_s}\sum_{i=1}^{N_s}\left\langle \chi_i\left|\zeta_i^{(k)}\right.\right\rangle \tag{43}$$

are the stochastic estimates of the Chebyshev moments [48] and $b_k\left(\beta,\mu\right)$ are the Chebyshev expansion coefficients of the function $f_{\beta\,\mu}\left(\varepsilon\right)$ as opposed to $c_k(\beta,\mu)$ which are the expansion coefficients

Table 1: The Chebyshev expansion lengths, K and K^w, and the number of stochastic orbitals, N_s^Δ and N_s^w, used in our simulations. We also show the required numerical work, W, defined as the number of Hamiltonian operations. Note that for each system we increase the number of high-temperature orbitals (N_s^w) with temperature (i.e., with increasing β/β^w) such that the total work W^{tot} is independent of β/β^w.

			Correction filter			Warm filter			
System	Band gap (eV)	β (eV^{-1})	K	N_s^Δ	$W^\Delta = KN_s^\Delta$	K^w	N_s^w	$W^w = K^w N_s^w$	W^{tot}
$Si_{35}H_{36}$	3.4	1.83	2000	6	$12,000$	$K \times \beta^w/\beta$	$24 \times \beta/\beta^w$	$48,000$	$60,000$
$Si_{87}H_{76}$	2.5	2.94	3200	6	$19,200$	$K \times \beta^w/\beta$	$24 \times \beta/\beta^w$	$76,800$	$96,000$
$Si_{353}H_{196}$	1.6	4.60	5000	6	$30,000$	$K \times \beta^w/\beta$	$24 \times \beta/\beta^w$	$120,000$	$150,000$

of $\sqrt{f_{\beta\,\mu}(\varepsilon)}$.

As a result of using stochastic orbitals, the sDFT density and associated observables have two additional types of errors. One is the usual stochastic fluctuations that scale as $\mathrm{O}(N_s^{-\frac{1}{2}})$, but in addition, there is a systematic deviation which scales as $\mathrm{O}(N_s^{-1})$ that appears due to the nonlinear SCF procedure (the filtering operator applied on each orbital depends on the density, which itself depends on the set of filtered orbitals).

Increasing the number of sampling orbitals, N_s, will decrease both types of errors, at the cost of additional work.

To measure the computational cost of a sDFT or t-sDFT calculation, we use a numerical "work" quantity, W, which is the total number of Hamiltonian operations performed per SCF cycle (i.e., action by the Hamiltonian on a function), which for sDFT is approximately

$$W \simeq KN_s.$$

In practice, the work needs to be multiplied by a factor of about 1.7 due to the need to determine μ based on Eq. (42), but since this factor is common to all our methods here we do not include it.

3.1.2 Tempering Stochastic Density Functional Theory

We now describe the tempering method, designed to reduce the statistical errors in sDFT without increasing the overall computational effort. Consider, the decomposition of the desired filter, $\hat{\rho}_\beta$ (see Eq. (36)), into a higher temperature filter, $\hat{\rho}_{\beta^w}$ ($\beta^w < \beta$), with the correction term:

$$\Delta\hat{\rho} = \hat{\rho}_\beta - \hat{\rho}_{\beta^w}, \tag{44}$$

which is shown in Fig. (1) for a typical case of $\beta/\beta^w = 6$. Note that (a) all values of $\Delta\hat{\rho}$ are much smaller than unity and (b) the high-temperature filter, $\hat{\rho}_{\beta^w}$, is smoother than the low-temperature one, so its Chebyshev expansion is shorter (the Chebyshev expansion length of $\hat{\rho}_\beta$ is proportional to β [28]).

The electron density in Eq. (39) can therefore be written as $n(\boldsymbol{r}) = n_{\beta^w}(\boldsymbol{r}) + \Delta n(\boldsymbol{r})$, where the two terms are evaluated separately using two distinct independent sets of stochastic orbitals: χ_i^w, $i = 1, \ldots, N_s^w$ for the warmer density,

$$n_{\beta^w}(\boldsymbol{r}) = \frac{2}{N_s^w} \sum_{i=1}^{N_s^w} \left| \left\langle \boldsymbol{r} \left| \sqrt{\hat{\rho}_{\beta^w}} \right| \chi_i^w \right\rangle \right|^2 \tag{45}$$

and χ_i^Δ, $i = 1, \ldots, N_s^\Delta$ for the correction term,

$$\Delta n(\boldsymbol{r}) = \frac{2}{N_s^\Delta} \sum_{i=1}^{N_s^\Delta} \left(\left| \left\langle \boldsymbol{r} \left| \sqrt{\hat{\rho}_\beta} \right| \chi_i^\Delta \right\rangle \right|^2 - \left| \left\langle \boldsymbol{r} \left| \sqrt{\hat{\rho}_{\beta^w}} \right| \chi_i^\Delta \right\rangle \right|^2 \right). \tag{46}$$

As demonstrated in Fig. 1, the correction density, $\Delta n(\boldsymbol{r})$, is much smaller than the warm density, $n_{\beta^w}(\boldsymbol{r})$, which is similar in overall magnitude to the total density. This gives the key for the efficiency of the tempering approach as compared to the original sDFT calculation. Specifically, compared to an sDFT calculation with polynomial expansion length K and N_s stochastic orbitals and aiming for the same overall work as in sDFT, we get that:

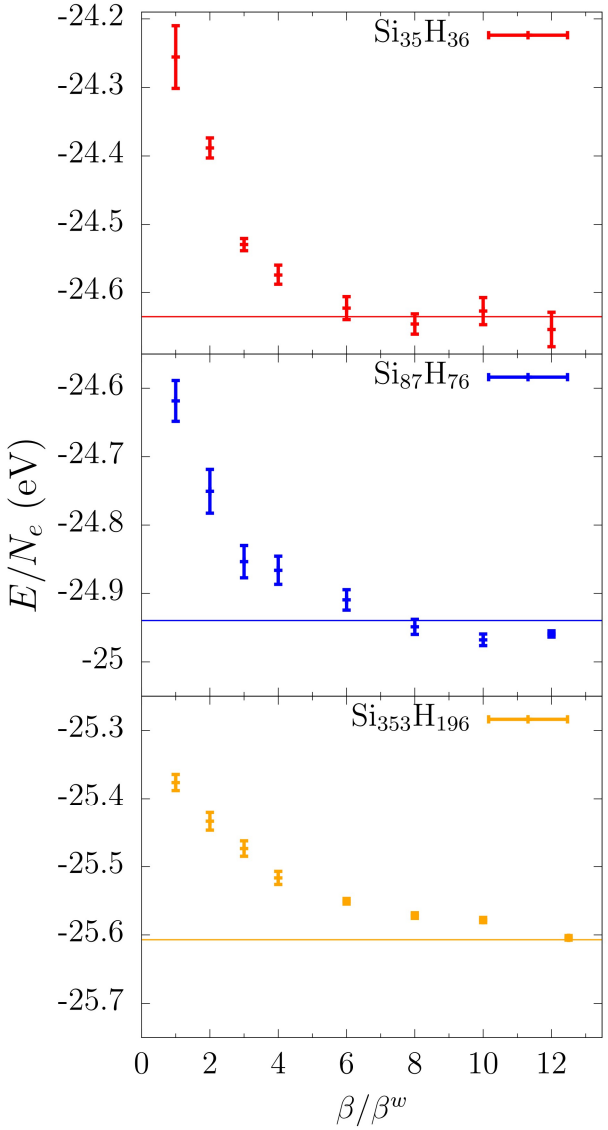

Figure 2: The energy per electron (in eV) as a function of β/β^w for three cluster sizes, based on $N_{\rm ind} = 10$ independent runs. Also included is the deterministic value in each system (horizontal line). The numerical work W, i.e., the number of Hamiltonian operations, is independent of β/β^w (see Table 1 for details). The leftmost point in each graph, $\beta/\beta^w = 1$, corresponds to sDFT (no tempering). Since the number of orbitals used is very small ($N_s = 30$ for sDFT), these sDFT results show marked systematic deviation (i.e., deviation of the average energy from the deterministic value) and fluctuation errors. Both types of statistical errors decrease markedly in t-sDFT, especially when $\beta/\beta^w \sim 7 - 10$, due to the much larger number of stochastic orbitals used in the main (warm) density part.

- The computational work required to calculate the warm density is $W^w \simeq K^w N_s^w$. Since the Chebyshev expansion lengths are proportional to β, the warmer temperature density $n_{\beta^w}(\boldsymbol{r})$ requires a much *shorter* Chebyshev expansion length than the original sDFT density ($K^w = \frac{\beta^w}{\beta} K$), so many more stochastic orbitals can be used to evaluate it for the same overall computational cost (i.e., $N_s^\Delta \gg N_s$).

- The computational work for correction term, $\Delta n(\boldsymbol{r})$, is $W^\Delta \simeq K N_s^\Delta$ as both terms in the RHS of Eq. (46) use the same set of $\left|\zeta_i^{\Delta,(k)}\right\rangle = T_k(\hat{h}_s)\left|\chi_i^\Delta\right\rangle$ (the two terms differ in their expansion coefficients). Since the numerical magnitude of the correction term is much smaller than that of the overall density, its standard deviation is correspondingly much smaller. Therefore the numerical effort (i.e., the number of samples required for a given accuracy), which is proportional to the squared standard deviation, is much smaller for the correction term, so it is sufficient to use fewer stochastic orbitals ($N_s^\Delta \ll N_s$) to achieve a similar statistical error.

Here, K^w and K are respectively the Chebyshev expansion lengths for the warm reference density and the correction term, with $\frac{K^w}{K} = \frac{\beta^w}{\beta} \ll 1$ (note that since the correction term involves the original low-temperature density, the number of Chebyshev terms it requires, K, is the same as in the original sDFT). Overall, the partitioning of the filter into a larger component at a higher temperature with a shorter Chebyshev expansion, and a smaller correction term, offers an additional knob to control the statistical error by using $N_s^w \gg N_s$ without increasing the overall computational effort.

The use of tempering modifies how the chemical potential is calculated. Instead of fulfilling the single-sum sDFT constraint on the residues (Eq. (42)), the chemical potential is adjusted to satisfy the following relation which consists of two summation terms that each has its own Chebyshev

expansion:

$$N_e = 2\sum_{k=0}^{K^w} b_k\left(\beta^w,\mu\right) M_k^w + 2\sum_{k=0}^{K}\left[b_k\left(\beta,\mu\right) - b_k\left(\beta^w,\mu\right)\right] M_k^{\Delta}. \tag{47}$$

The corresponding Chebyshev moments, defined in analogy to Eq. (43), are:

$$\begin{aligned} M_k^w &= \frac{1}{N_s^w}\sum_{i=1}^{N_s^w}\left\langle \chi_i^w \left| \zeta_i^{w,(k)} \right\rangle, \\ M_k^{\Delta} &= \frac{1}{N_s^{\Delta}}\sum_{i=1}^{N_s^{\Delta}}\left\langle \chi_i^{\Delta} \left| \zeta_i^{\Delta,(k)} \right\rangle. \end{aligned} \tag{48}$$

In practice, finding the chemical potential in both sDFT and t-sDFT involves a straightforward single-variable search. In t-sDFT, the chemical potential (Eq. (47)) is not strictly guaranteed to be monotonic with the number of electrons, since it is a summation of two terms, one of which is not necessarily positive; but in practice, we find that it is always monotonic so the determination of μ from the residues is instantaneous.

3.2 Results

We studied three hydrogen-terminated silicon nanocrystals of different sizes, $Si_{35}H_{36}$, $Si_{87}H_{76}$, and $Si_{353}H_{196}$. Note that such nanocrystals are a convenient test ground for stochastic methods since they have small band gaps which decrease with increasing system size. As such they place a more stringent test on the method than clusters of molecules with large gaps. Metals, in contrast, require much smaller temperatures and are therefore not ideal for stochastic applications.

An LDA functional [49] was applied with norm-conserving pseudopotentials [50] using the Kleinman-Bylander form [51], and we used the Martyna-Tuckerman reciprocal-space method for

treating long-range interactions [52]. The grid spacing was $0.55a_0$, and the energy cutoff was $15\,\mathrm{Hartree}$ for all systems. To gather sufficient statistics, $N_{\mathrm{ind}} = 10$ independent runs with different stochastic numbers were used for each of the calculations below.

For each system, we performed calculations for several β/β^w ratios. As these systems are semiconductors, we are simply interested in the limit where the Fermi-Dirac distribution is effectively a step function. We, therefore, replace the Fermi-Dirac distribution by the complementary error function, $f_{\beta\,\mu}(\hat{h}) = \frac{1}{2}\mathrm{erfc}(\beta(\hat{h}-\mu))$, which looks similar to the Fermi-Dirac distribution but does not require a very small β to be effectively a step function.

The numerical parameters for the runs are summarized in Table 1. There are several points to note.

First, since the three systems have with increasing size a progressively smaller band gap, E_g, the larger systems require larger values of β and correspondingly larger Chebyshev expansion lengths, K. Furthermore, we use (for all systems) $N_s = 30$ orbitals for the sDFT calculations. Finally, note that the warm temperature calculations which require most of the numerical work have an expansion length K^w and a corresponding number of stochastic orbitals N_s^w chosen so that the total work W^{tot} is the same for each value of β/β^w (this includes sDFT at $\beta/\beta^w = 1$). This allows us to compare the efficacy of tempering in terms of the reduction of fluctuations as a function of β/β^w.

3.2.1 Prelude: Stochastic vs. Systematic Deviation

The results shown below exhibit two kinds of deviations, which are briefly reviewed; for a fuller discussion see Ref. [44]. The first deviation the usual stochastic Monte-Carlo fluctuation which scales with the number of samples as $N_s^{-1/2}$. The other kind is a systematic deviation. Such deviation scales as N_s^{-1} and appears whenever the results of the Monte-Carlo sampling are used in an iterative self-consistent process (see, e.g., [53]). Here, since the density is prepared from a finite number of stochastic orbitals and the filtered stochastic orbitals depend on the density, the

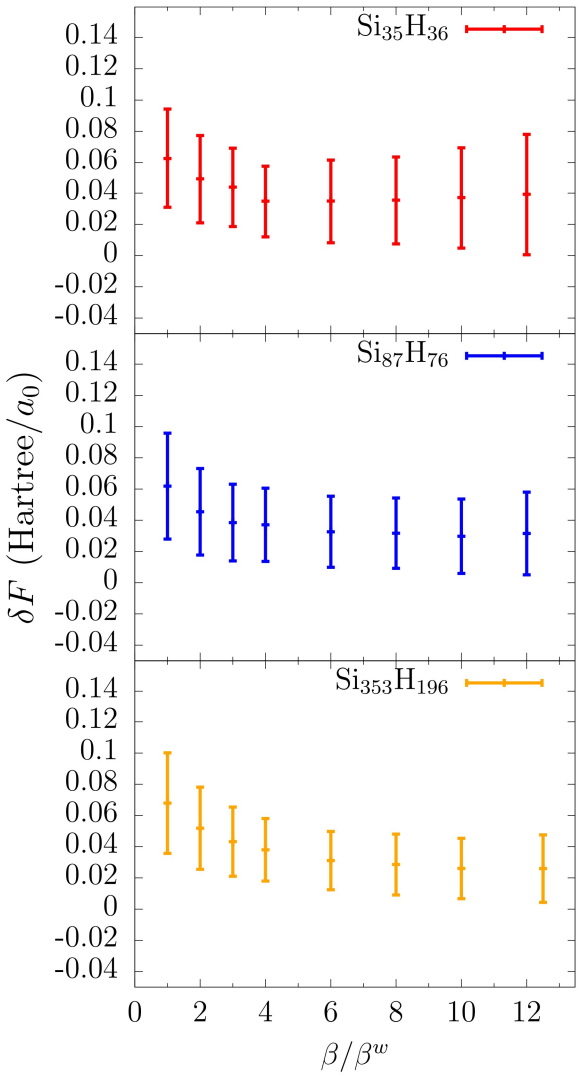

Figure 3: Analogous to Fig. 2 but shows δF,the error in the averaged force relative to the deterministic value, normalized over all silicon atoms and over the $N_{\text{ind}} = 10$ runs, and the associated standard deviation σ_F (Eqs. (49,50)). In sDFT δF is significantly larger than σ_F, indicating some amount of systematic deviation. In t-sDFT, around $\beta/\beta^w \sim 7 - 10$, both the stochastic and especially the systematic errors decrease, i.e., σ_F decreases and $\delta F \sim \sigma_F$.

self-consistent SCF procedure has a systematic deviation.

The practical effect of the systematic deviation is simple to state: when doing calculations with a finite N_s, and repeating these calculations many times, the averaged result would differ from the true $N_s \to \infty$ result. As we show below, in most of our calculations, tempering reduces and practically eliminates this systematic deviation, avoiding the need to use jackknife or bootstrap methods [54].

3.2.2 Energies

Fig. 2 shows, for the three different systems, the averaged total energies per particle based on the $N_{\rm ind} = 10$ runs and the associated error bars for different β^w values. For simplicity, the results are depicted as a function of β/β^w. In addition, we include the deterministic DFT values for comparison. Interestingly (see also the SI), for a fixed N_s, the systematic deviation decreases by a factor of 2 when the system size increases by a factor of 10, while the stochastic error decreases by a larger factor with system size, scaling as $N_e^{-1/2}$ due to self-averaging.

Consider first the starting point for each figure, $\beta/\beta^w = 1$, which is simply sDFT (i.e., with no correction terms). Since we only use $N_s = 30$ stochastic orbitals, a very small number, the sDFT calculations show a significant systematic deviation, i.e., the averaged energy-per-particle is several standard deviations away from the deterministic value.

Turning to t-sDFT (i.e., $\beta^w < \beta$), we see that both the systematic deviation and the stochastic error decrease as β/β^w increases. As evident from Fig. 2 (and verified by a second-order polynomial fit of the error in the energy as a function of β/β^w in the SI) once $\beta/\beta^w \sim 7 - 12$ there is essentially no systematic deviation while the stochastic fluctuations decrease by a factor of around 4-5.

The reduction in the statistical error and systematic deviation as β/β^w increases relative to sDFT ($\beta/\beta^w = 1$) for a fixed W^{tot} results from the fact that we can use a much larger number of stochastic orbitals for the warmer temperature density, the significant contribution to the density,

compared to the sDFT density. Despite the need to use a longer Chebyshev expansion, we can use fewer stochastic orbitals for the correction term since its contribution to the density is smaller.

Finally note in Fig. 2, with increasing system size, the optimal β/β^w values (i.e., that results in the smallest small statistical fluctuations in the energy per electron) shift to larger ratios. This is partially a result of quantum confinement, so that with increasing size the KS gap decreases and therefore β increases, and partially due to the modified density of state structure for the larger clusters which causes the optimal β^w to increase with system size. Therefore, the smallest system the optimal ratio is about $\beta/\beta^w \sim 3-4$, and for the larger clusters minimum error is obtained for $\beta/\beta^w \sim 7-12$.

3.2.3 Forces and Density

We next show how t-sDFT reduces the errors in the atomic forces compared to sDFT (we only analyze the forces on the silicon atoms for comparison). Fig. 3 is similar to the energy plot in Fig. 2, but here we plot the normalized deviation of the averaged stochastic forces from the deterministic forces, δF:

$$(\delta F)^2 \equiv \frac{1}{N_{\rm Si}} \sum_{i=1}^{N_{\rm Si}} |\bar{\boldsymbol{F}}^i - \boldsymbol{F}^{d,i}|^2, \tag{49}$$

where a bar indicates averaging over the $N_{\rm ind} = 10$ runs and "d" stands for deterministic; i is an index over the silicon atoms. The error bars in Figure 3, σ_F, indicate the standard deviation of the normalized averaged force of the silicon atoms, i.e.,

$$\sigma_F^2 = \frac{1}{(N_{\rm ind}-1)N_{\rm ind}N_{\rm Si}} \sum_{j=1}^{N_{\rm ind}} \sum_{i=1}^{N_{\rm Si}} |\boldsymbol{F}^{i,j} - \bar{\boldsymbol{F}}^i|^2, \tag{50}$$

where $\boldsymbol{F}^{i,j}$ is the force over atom i in the j'th independent run.

Note that the magnitude of the errors in the forces is large, but this could be reduced by increas-

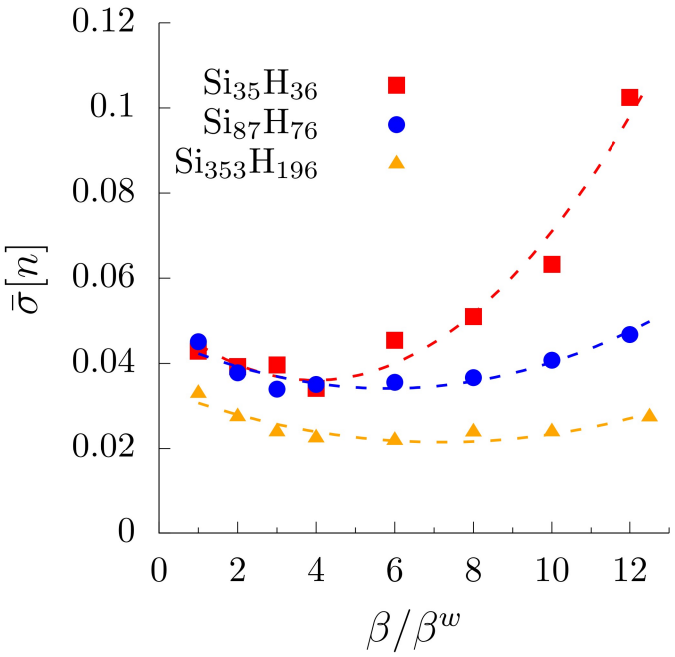

Figure 4: The normalized integral of the standard deviation of the density per electron, together with a parabolic fit. For the smaller system, the density deviation decreases between $\beta/\beta^w = 2-4$, and for the two larger systems the stochastic errors decrease around a larger range $\beta/\beta^w = 2-10$, by up to 30%-40%.

ing the number of independent runs or stochastic orbitals. However, since the purpose of the study is to uncover the behavior with respect to β^w, we use a small number of stochastic orbitals to reduce the computational effort and thus, apply the approach for many values of β/β^w and different system sizes. Further, note that since the stochastic errors are generally not systematic, especially once tempering is used, these forces can be used for Langevin molecular dynamics if we increase the number of sampling orbitals by about an order of magnitude. We have indeed applied sDFT (with an order of magnitude more orbitals than in this study) in a Langevin molecular dynamics study and have shown that that the Langevin dissipation matrix is then easily modified to include the effect of the fluctuations in the sDFT force, and the correct pair distribution is then obtained [55, 56].

As previously discussed, the sDFT forces are similar in the three systems, since the local environment is similar and therefore the errors are primarily a function of the number of stochastic orbitals, N_s. The reduction in the errors in the forces is appreciable but less significant than for the energy. Using again a 2nd order fit (see the SI), we get that the reduction in error in the forces is about 30% for the smallest system and goes up to 50% for the largest cluster.

To compare the deviation in the density using t-sDFT to sDFT, we use the integral of the standard deviation of the averaged density per electron, defined as $\bar{\sigma}[n] \equiv (N_{\text{ind}} - 1)^{-1/2} N_e^{-1} \sum \sigma(n(\boldsymbol{r}))dV$, where $\sigma(n(\boldsymbol{r}))$ is the standard deviation in the density at grid point $\boldsymbol{r}$. Fig. 4 shows that tempering again reduces the stochastic error for β/β^w values around 7-10. The reduction in the deviation of the density is similar in magnitude to that of the total atomic forces, up to 30%-40%, and is much less dramatic than the error reduction in the total energies.

Finally, note that Fig. 4 shows that when the value of β/β^w is very large, the density fluctuations start rising with the β/β^w ratio; for large ratios, the warm density deviates significantly from the low-temperature density, so the difference between the two densities is significant which causes large statistical fluctuations.

3.3 Conclusions

We presented here a tempering method for stochastic density functional theory that reduces the statistical error in the total energy. Our scheme (t-sDFT) relies on decomposing the density into a large high-temperature term with a correction density. The new method expands the density in terms of the inverse temperature, β, to take advantage of the fact that with lower β (i.e., a higher effective temperature) fewer Chebyshev polynomials are needed, thus enabling the use of more stochastic orbitals without increasing the computational cost.

A natural extension of this work is the implementation of multiple-β tempering with more than two values of β, as done earlier for deterministic renormalization-group studies; the formalism is presented in Appendix A. In this work we have not implemented embedded fragments, an approach that independently reduces the standard deviation in the energy and forces. In future work, the two methods will be combined to hopefully further reduce the stochastic error. Further work will also explore how to optimize the choice of β values and the number of stochastic orbitals used to reduce the stochastic deviations.

Our method reduces the standard deviation in the total energy by a factor of around 4-5, which corresponds to reducing the total number of required stochastic orbitals by more than an order of magnitude. This is only for the total energies, while the error in the forces and density is reduced by a smaller amount only 30%-50% and 30%-40% respectively. Interestingly, this is the opposite behavior relative to energy-window sDFT where the error in the forces is improved significantly but not the error in the total energies. Another interesting aspect is that sDFT almost eliminates systematic deviation.

The main conclusion of our work is that for the same overall cost, tempering improves the accuracy by 1.4-4 depending on the quantity studied while also shrinking the systematic deviation so that the results are closer to the deterministic value even for a small number of samples. Equivalently, for the same stochastic deviation, tempering reduces the overall effort by a factor ranging

from $\approx$2 to ≈ 20, depending on the desired quantity.

Acknowledgments

This work 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. In addition, RB gratefully acknowledges the support from the US-Israel Binational Science Foundation (BSF) under Grant No. 2018368. Computational resources were supplied through the XSEDE allocation TG-CHE170058.

Appendix A: Multiple β

The general expansion of the filter $\hat{\rho}_\beta$ for L temperatures, ordered so $\beta \equiv \beta_1 > \beta_2 > ... > \beta_L$, is

$$\hat{\rho}_\beta = \hat{\rho}_{\beta_L} - \sum_{\ell=1}^{L-1} \Delta\hat{\rho}_\ell, \tag{51}$$

where

$$\Delta\hat{\rho}_\ell = \hat{\rho}_{\beta_{\ell+1}} - \hat{\rho}_{\beta_\ell}. \tag{52}$$

This expansion leads to an expression for the density similar to Eqs. 45-46. The case we studied in this chapter is simply $L = 2$, with $\beta_1 \equiv \beta$ and $\beta_2 \equiv \beta^w$.

4 Gapped Filtering Stochastic GW

A fundamental ingredient in many electronic structure and dynamics methodologies is the polynomial expansion of the density matrix operator. Such methods include deterministic approaches to DFT that avoid the direct diagonalization of the Hamiltonian matrix, [57, 58] or allow for fast diagonalization of the Kohn-Sham Hamiltonian [59] when the Hamiltonian is a sparse matrix. These polynomial expansions are also the primary ingredient in stochastic quantum chemistry methods, [48] including stochastic DFT [60–62] and beyond-DFT approaches. The latter include the linear-scaling stochastic-GW method (sGW) [32] which calculates the quasiparticle (QP) energy as a perturbative correction to the DFT eigenvalue [33, 63–65], as well as, e.g., stochastic-MP2 [66] stochastic Bethe Salpeter Equation [6] and stochastic Time-Dependent DFT (TDDFT) [67].

In the usual Chebyshev approach [68–70] the coefficients in the polynomial expansion of the density matrix $\hat{\rho}$ are obtained analytically by requiring that the scalar function $\Theta(\mu - E)$ converges uniformly as a function of E [69, 70]. In practice, to converge the expansion one usually replaces the sharp Heaviside function with a smoothed one. However, for most systems the gap is small relative to the full energy range of the Hamiltonian operator, so even a smoothed Heaviside function would require typically thousands of terms for convergence.

Here we suggest an alternative to the usual determination of the Chebyshev coefficients. For any desired number of coefficients, we only require that the Chebyshev expansion be correct for energies outside the gap. The logic is that it is immaterial what the values of the polynomial expansion are inside the gap since there are no eigenstates there. Within the gap the filter could have any form, including Gibbs oscillations, but they would be irrelevant to the final density matrix.

In Section II we develop the idea in detail and give a simple prescription, which we label gapped-filtering, for obtaining the Chebyshev coefficients. Numerically, gapped-filtering carries negligible overhead. One just specifies the desired length of the Chebyshev expansion, labeled $N_{\rm chb}$, and then just inverts a single matrix of rank $N_{\rm chb} + 1$.

In Section III we numerically study gapped-filtering and show that the method converges rapidly with N_{chb}, which is the only parameter in the approach. The method is more efficient than traditional smoothed filters, where one needs to converge with both a smoothness parameter and N_{chb}. When high accuracy is required, gapped filtering is 2-3 times faster than the traditional approaches.

Section IV presents an application of the method to stochastic-GW, verifying this significant reduction in a large scale calculation. Conclusions are discussed in Section V.

4.1 Methodology

4.1.1 Gapped-filtering

Ignoring spin, the density matrix, i.e., the projection operator to the occupied manifold, is formally:

$$\hat{\rho} = \sum_{i=1}^{N_{\mathrm{occ}}} |\phi_i\rangle \langle\phi_i| \,, \tag{53}$$

where $\{\phi_i\}$ are the eigenvectors of the 1-body Hamiltonian operator $\hat{H}$ and N_{occ} is the number of occupied states. We only consider here systems with a gap where the physical temperature is tiny compared to the gap, i.e., a zero-temperature description.

Even if all eigenstates and eigenvalues are known, the memory required to store the eigenstates could be enormous for giant systems with tens of thousands of electrons. In those cases, it is better to use a polynomial expansion. Thus, the density matrix is equivalent to a Heaviside function of energy centered within the band gap,

$$\hat{\rho} = \Theta(\mu - \hat{H}), \tag{54}$$

where μ is the chemical potential. To converge the expansion, a smoothed complementary-error

function is typically used:

$$\Theta(\mu - \hat{H}) \simeq \hat{\Theta}_\beta \equiv \sqrt{\frac{1}{2}\mathrm{erfc}\left(\beta\left(\mu - \hat{H}\right)\right)}, \tag{55}$$

where β is an inverse-temperature-like parameter that determines the sharpness of the function [57–59].

Note that the square root stems from the fact that in most cases, especially for stochastic quantum chemistry applications, the results of applying the filter are then squared to give the density or Green's function. Throughout this chapter, the analytical filters always refer to this square-root erfc form (abbreviated as sqrt-erfc). Further, we verified that gapped-filtering does analogously well when compared with an erfc filter without a square-root.

To apply the filter, the Chebyshev polynomial expansion is traditionally used:

$$\Theta(\mu - \hat{H}) \simeq \sum_{n=0}^{N_{\mathrm{chb}}} a_n T_n(\tilde{H}), \tag{56}$$

where the expansion coefficients, a_n, depend on μ and β and T_n which is the n'th order Chebyshev polynomial of the first kind. The argument of the Chebyshev function is a scaled Hamiltonian, $\tilde{H} \equiv \frac{\hat{H} - H_{\mathrm{avg}}}{\Delta H}$, [69, 71] where the H_{avg} and ΔH parameters are the center and half-width of the spectrum of $\hat{H}$, so that the eigenvalues of $\tilde{H}$ are between -1 and 1. The length of the expansion, N_{chb}, is approximately proportional to $\beta\Delta H$, times a factor which depends on the relative position of μ relative to the spectrum bottom, $H_{\mathrm{avg}} - \Delta H$. (Qualitatively that factor accounts for the "stretching" of the angle $\theta = \arccos(x)$ near the band bottom, where $x \to -1$ and $dx/d\theta$ becomes large. For a fuller discussion see [72].)

Usually the coefficients are determined by requiring that the expansion in Eq. (56) is valid uniformly over all E between the minimum and maximum eigenvalues of $\hat{H}$. Due to the orthogonality

of the Chebyshev polynomials, the coefficients are then simply:

$$a_n = \frac{2 - \delta_{n0}}{\pi} \int_{-1}^{1} \Theta(\tilde{\mu} - x) T_n(x) w(x) dx, \tag{57}$$

where $w(x) = (1 - x^2)^{-1/2}$ is the Chebyshev weighting function and we introduced the scaled chemical potential, $\tilde{\mu} = (\mu - H_{\text{avg}})/\Delta H$.

Here we exploit a simple realization: because there are no states within the bandgap, the behavior of the expansion within the gap does not impact the density matrix. What is important is that the expansion is 1 for the occupied states and 0 for the unoccupied states, regardless of its values within the bandgap. As we show below, the increased flexibility, as the expansion is free to vary within the band gap, suffices to reduce the number of polynomials for a given desired level of numerical accuracy.

Our derivation starts by defining a modified weighting function that vanishes within the scaled band gap:

$$w^*(x) = w(x)\big(\Theta(x_H - x) + \Theta(x - x_L)\big), \tag{58}$$

where the scaled energies are

$$x_H = (\varepsilon_{\text{H}} - H_{\text{avg}})/\Delta H + \delta$$

$$x_L = (\varepsilon_{\text{L}} - H_{\text{avg}})/\Delta H - \delta$$

where ε_H and ε_L are the HOMO and LUMO energies, while δ is a small padding factor, which we usually take to be about 1% of the scaled band gap. This padding ensures that the weighting function does not go to 1/0 at exactly the HOMO/LUMO energies, thereby allowing for any uncertainties in the values of the HOMO/LUMO energies.

Since the weighting function is zero within the gap, we call the resulting method gapped-filtering, and the associated Chebyshev expansion gapped-filter. As the Chebyshev polynomials are not orthogonal under this weighting function, the coefficients need to be rederived. This is

done by minimizing the norm squared of the difference between the Heaviside function and the Chebyshev expansion, weighted now by $w^*(x)$:

$$J = \int_{-1}^{1} \left| \Theta(x - x_0) - \sum_{n=0}^{N_{\rm chb}} a_n T_n(x) \right|^2 w^*(x) dx. \tag{59}$$

Minimizing J with respect to a_i^* gives:

$$\frac{\partial J}{\partial a_i^*} = \sum_k M_{ij} a_j - b_j = 0, \tag{60}$$

i.e., $a = M^{-1} b$, where

$$b_i = \int_{-1}^{x_H} w^*(x) T_i(x) dx,$$

and

$$M_{ij} = \int_{-1}^{1} w^*(x) T_i(x) T_j(x) dx. \tag{61}$$

The matrix elements are easily calculated. Defining $x = \cos\theta$, we get:

$$M_{ij} = F_{ij}(-1, x_H) + F_{ij}(x_L, 1), \tag{62}$$

where

$$\begin{aligned} F_{ij}(y,z) &= \int_y^z w(x) T_i(x) T_j(x) \\ &= \int_{\theta_z}^{\theta_y} \cos(i\theta)\cos(j\theta) d\theta \\ &= \frac{1}{2} \int_{\theta_z}^{\theta_y} \big(\cos((i-j)\theta) + \cos((i+j)\theta) \big) d\theta \\ &= G_{ij}(\theta_y) - G_{ij}(\theta_z) \end{aligned} , \tag{63}$$

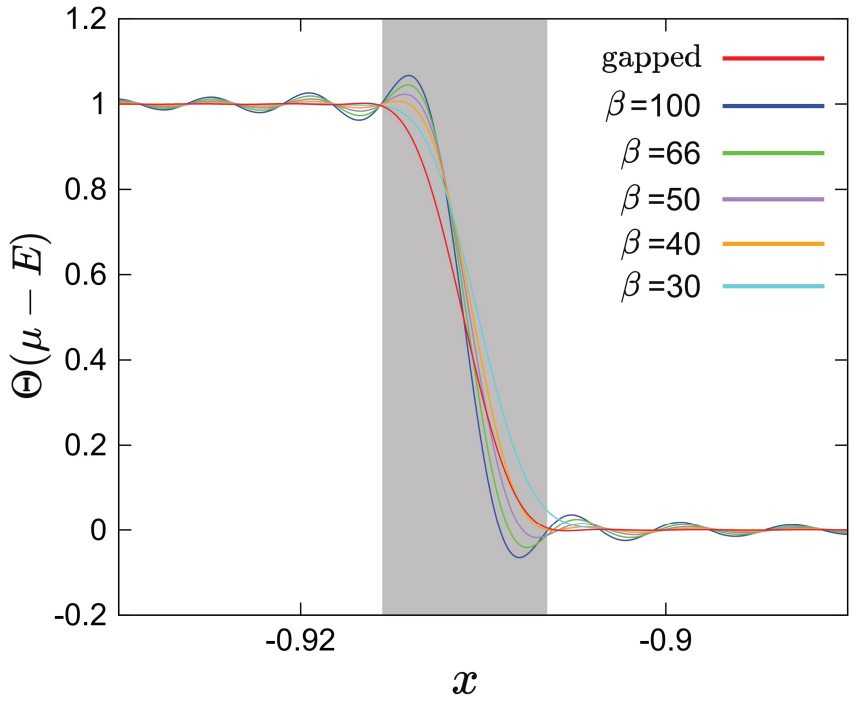

Figure 5: Gapped-filtering vs. various sqrt-erfc filters, using a Chebyshev expansion with $N_{\text{chb}} = 450$ polynomials, as a function of the scaled energy, $x = (E - H_{\text{avg}})/\Delta H$. Naphthalene parameters are used ($\Delta H = 13.5$ Hartree, $\varepsilon_{\text{gap}} = 0.123$ Hartree). The highlighted grey area denotes the gap region where the expansion need not be optimized. Note that β is always reported in inverse Hartree, so, for example, for $\beta = 30$ Hartree^{-1} the product of β with the gap is quite small, $\beta(\varepsilon_L - \varepsilon_H) = 3.7$.

where $\cos\theta_y = y$, and

$$G_{ij}(\theta) = \begin{cases} \theta & i = j = 0 \\ \frac{1}{2}\theta + \frac{1}{4i}\sin(2i\theta) & i = j \neq 0 \\ \frac{\sin((i-j)\theta)}{2(i-j)} + \frac{\sin((i+j)\theta)}{2(i+j)} & i \neq j \end{cases} . \tag{64}$$

Similarly,

$$b_i = \int_{\theta_H}^{\pi} \cos(i\theta) d\theta = \begin{cases} \pi - \theta_H & i = 0 \\ -\frac{1}{i}\sin(i\theta_H) & i > 0 \end{cases} , \tag{65}$$

where $\cos(\theta_H) = x_H$.

Additional considerations

We now discuss several issues associated with the method.

HOMO/LUMO determination: Our method requires knowledge of the HOMO/LUMO energies associated with the one-body Hamiltonian $\hat{H}$. In situations where these are not accurately given, it will often be worthwhile to do a single longer Chebyshev propagation to determine accurately, by filtering, the location of the HOMO and LUMO. For example, in stochastic GW the gapped filter will be applied thousands of times, so the initial overhead for finding the HOMO/LUMO energies by a narrow Chebyshev filter will be often small compared to the savings incurred by the subsequent use of gapped filtering in sGW.

The same considerations would also be true in DFT applications that rely on Chebyshev filtering, as long as the overall number of vectors on which one needs to apply Chebyshev filtering is much larger than 1, so that in each SCF step the extra overhead in a single long Chebyshev to determine the HOMO and LUMO would be negligible compared to the overall cost needed to filter all the required states.

HOMO/LUMO accuracy: Note that formally the HOMO and LUMO energies that are used

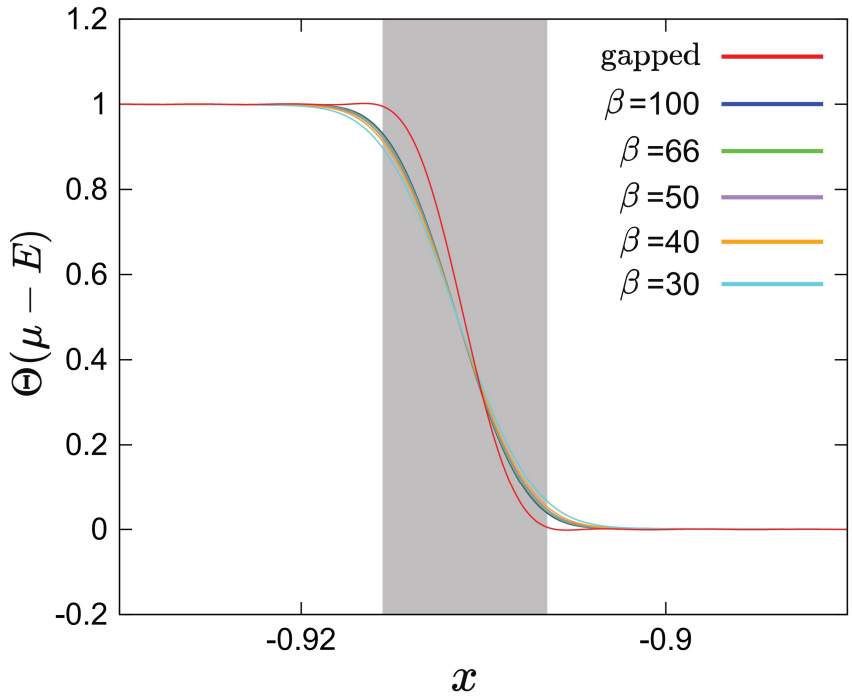

Figure 6: Analogous to Fig. 1 but now the analytical coefficients are damped using the commonly used Jackson kernel coefficients to avoid a sharp cut at N_{chb}. While the Gibbs oscillations are damped the step function is even further widened, i.e., gets further from the analytical step function demonstrating that just damping the analytical coefficients is not sufficient.

need to be lower and upper bounds to the true HOMO and LUMO, but since the method is continuous even a small (circa 5% or less) deviation in the wrong direction (e.g., an applied HOMO energy which is higher than the true HOMO by about 2-3$\%$ of the gap, and analogously for the LUMO) still gives excellent results, as we verified.

Correlation function: Also note that when one is just looking for a correlation function, involving a matrix element of the density of state (rather than the full filtered function, Eq. (55)), then the total number of residues is halved by using simple trigonometric identities, e.g., $2\left\langle\zeta_0\right|T_{2n}\left|\zeta_0\right\rangle = \left\langle\zeta_n|\zeta_n\right\rangle - \left\langle\zeta_0|\zeta_0\right\rangle$ where $\left|\zeta_n\right\rangle \equiv T_n\left|\zeta_0\right\rangle$ (see Eq. (3.10) in [73]). This reduction by a half is valid for both the regular Chebyshev approaches and for gapped filtering, so it does not change the relative performance of the approaches.

Filter design: To conclude this section, note that the problem we specifically try to solve, an optimum Chebyshev filter where we do not care what happens inside the gap, could be viewed as a specific sub-problem of digital filter design [74, 75].

4.2 Gapped-filtering convergence study

We now turn to a numerical study of the new method. We use the band gap and energy width from an LDA calculation of Naphthalene. The calculations are detailed in the next section, but the relevant part here is that the gap is $\varepsilon_{\rm gap} \equiv \varepsilon_L - \varepsilon_H = 0.123$ Hartree, the spectrum width is 13.5 Hartree, and the scaled HOMO/LUMO energies are $(\varepsilon_H - H_{\rm avg})/\Delta H = -0.9156$ and $(\varepsilon_L - H_{\rm avg})/\Delta H = -0.9065$ i.e., a scaled gap of $(\varepsilon_L - \varepsilon_H)/\Delta H = 0.0091$.

In Fig. 5 we show, for an expansion length of $N_{\rm chb} = 450$, the improved performance of gapped filter over the traditional approaches. We specifically plot as a function of the scaled energy ($x = (E - H_{\rm avg})/\Delta H$), the numerical representation of $\Theta(\mu - E)$ (i.e., the RHS of Eq. (56)) at several β values. As expected, convergence is faster for a lower β, but if β is too low then the analytic sqrt-erfc function deviates too much from the desired step function.

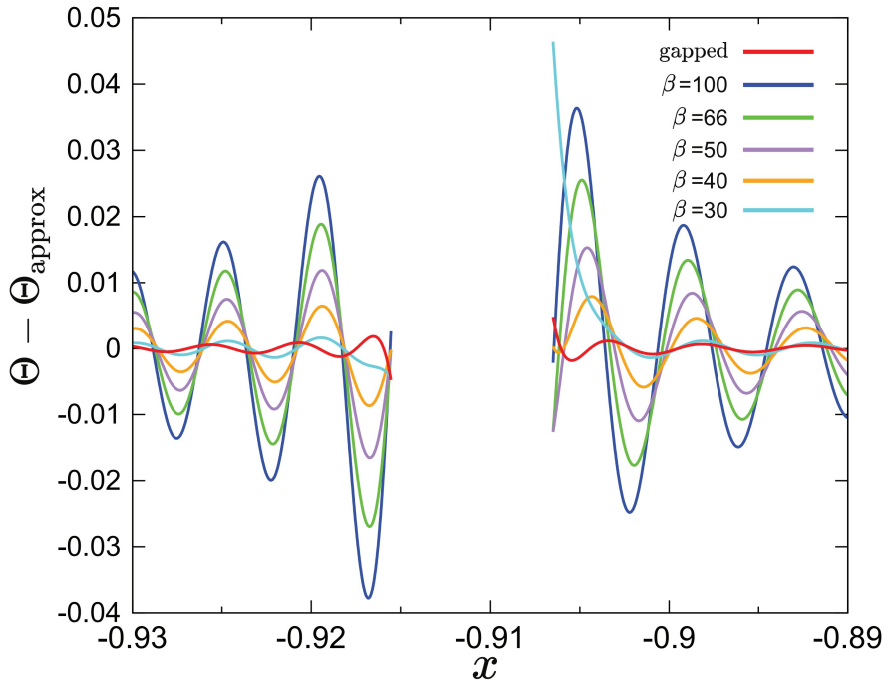

Figure 7: Analogous to Fig. 5, but showing, away from the gap, the differences between the Chebyshev expansions and an exact step function.

As a side note, a common approach [76] to avoid Gibbs oscillations is to multiply the Chebyshev coefficients by a filter, $c_n \to g_n c_n$, where g_n falls smoothly to almost 0 at $n = N$. This is shown in Fig. 6 (constructed for $N_{\text{chb}} = 450$, like Fig. 5), using the popular Jackson filter [76]. But while the smoothing of the analytic coefficients does remove the Gibbs oscillations, this comes however at the expense of making the filter wider, so it is even further from the true step function filter.

The difference in Fig. 5 at $N_{\text{chb}} = 450$ between the numerical representations and the desired step function is detailed in Fig. 7. Interestingly, at this N_{chb} the gapped filter is also smooth inside the gap (gray region in Fig. 5) even though the gapped-filtering method does not try to explicitly optimize the behaviour of the filter there. (Note: for very high N_{chb} which are irrelevant, $N_{\text{chb}} > 1300$, gapped-filtering does show Gibbs oscillations inside the gap.)

In Fig. 8 we show, again at $N_{\text{chb}} = 450$, the expansion coefficients, a_n, as a function of n.

The magnitude of the gapped-filtering coefficients decreases at a much faster rate than that of the traditional coefficients. The faster drop-off leads to a smoother expansion and therefore to less oscillations.

We now turn to an analysis of the performance of the gapped-filtering approach as a function of the number of Chebyshev polynomials. Fig. 9 shows, as a function of $N_{\rm chb}$, the maximum (over all energies outside the gap) of the absolute difference between Eq. (56) and Eq. (54) at several β values. The relative advantage of gapped filtering is more pronounced when higher accuracy is sought, i.e., at larger $N_{\rm chb}$. Thus small β values (in our case $\beta = 30$ or 40 Hartree^{-1}) require a small number of Chebyshev polynomials for convergence, that rivals the number in gapped-filtering. However, for such small β values the analytical sqrt-error function deviates significantly from the desired step function, so that the error would be high at any value of $N_{\rm chb}$. In contrast, a more accurate representation of the step function with small error requires a much larger number of polynomials for sqrt-erfc filters (see here the graphs for $\beta = 50, 66, 100$ Hartree^{-1}) than in gapped filtering.

A technical point is that for very large $N_{\rm chb}$, some of eigenvalues of M will be close to zero; this would lead to round-off errors in the matrix inversion, which in our case would deteriorate the accuracy of the gapped filter when the maximum error is on the order of 10^{-6}, i.e., beyond $N_{\rm chb} > 10\Delta H/\varepsilon_{\rm gap} = 1100$. We avoid such errors using, for such large $N_{\rm chb}$, quadruple precision in calculating the coefficients (i.e., for yielding M and b, and then for inverting $M^{-1}b$ using a quadruple precision algorithm [77, 78]). Because the dimensions of M_{ij} are $(N_{\rm chb}+1)\times(N_{\rm chb}+1)$ and $N_{\rm chb}$ is typically on the order of $100-2000$, this inversion has negligible cost. (To clarify, the coefficients are then well-behaved and are used in the usual double-precision Chebyshev algorithm, i.e., quadruple precision is only used to determine the a_n coefficients.)

An advantage of gapped-filtering is that it avoids the use of an artificial temperature parameter, β. We only need to worry about a single convergence parameter, $N_{\rm chb}$, unlike the usual procedure of converging with respect to two parameters, β and $N_{\rm chb}$.

4.3 Application: stochastic-GW

4.3.1 Filtering in stochastic-GW

The polynomial expansion of the Chebyshev operator is crucial for our stochastic quantum chemistry approaches. Here we overview the relevant parts for one such method, stochastic-GW. Only the parts of the method where filtering is relevant are discussed, see Refs. [32, 33] for further details. The stochastic-GW method, in its simplest version, calculates the HOMO (or LUMO) quasiparticle energy in the diagonal approximation ε_{QP}, i.e., based on the HOMO/LUMO DFT eigenvector whose associated eigenvalue is modified.

Stochastic-GW has two filtering stages. First, one does a stochastic realization of G. For each such realization, one takes an initially random white noise orbital $|\zeta_0\rangle$, i.e., $\zeta_0(r) \propto \pm 1$ in a grid representation (or an analogous expression in a basis-set description). Then, the white noise orbital is filtered with the Heaviside operator $\hat{\Theta} \equiv \Theta(\mu - \hat{H})$ to generate a function $|\zeta\rangle$ which contains only occupied states (with random coefficients):

$$\zeta(r) = \langle r| \hat{\Theta} |\zeta_0\rangle . \tag{66}$$

We interchangeably label such a function as a stochastic occupied orbital, or more simply a filtered orbital.

Each such filtered ζ function is then propagated in time under $\hat{H}$; the correlation function of the result with ζ_0 equals the negative-time Green's function, $\{\zeta(r,t)\zeta_0(r')\} = G_0(r,r',-t)$. The positive-time Green's function is obtained analogously. The curly brackets indicate a statistical average over the stochastic orbitals. Typically we use $N_\zeta = 200$ to 2000 different random white noise $\zeta_0(r)$ functions. Fewer samples are needed for larger systems due to self-averaging.

Then, for each such stochastic realization of G (i.e., for each ζ_0), one calculates the action of the time-dependent effective interaction $\hat{W}(t)$ on a vector related to ζ_0. This is done by choosing,

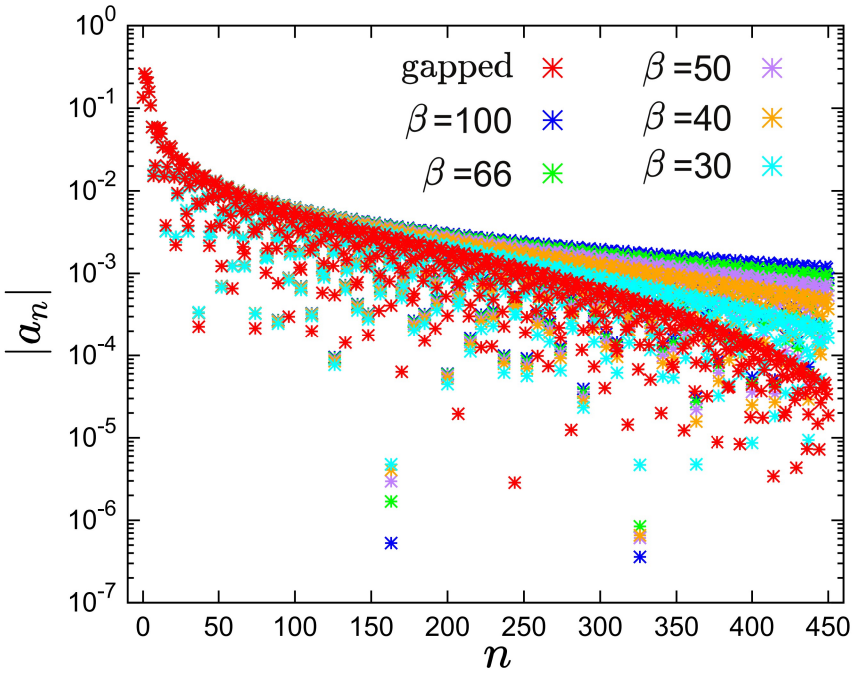

Figure 8: The absolute value of the expansion coefficients for $N_{\rm chb} = 450$ for the gapped and traditional (sqrt-erfc) filters.

for each ζ_0, several ($N_{\rm orb}$) white noise vectors, labeled $\eta_{0\ell}(r),\ \ \ell = 1, ..., N_{\rm orb}$. These vectors are also filtered, $|\eta_\ell\rangle = \Theta|\eta_{0\ell}\rangle$. This set of filtered orbitals are then excited and propagated in time, under a time-dependent Hartree Hamiltonian $H(t)$. The crucial part of the evolution Hamiltonian, which also gives the action of $\hat{W}(t)$, comes from the Hartree potential due to the time-dependent density

$$n(r,t) \simeq \frac{2}{N_{\rm orb}} \sum_{\ell=1}^{N_{\rm orb}} |\eta_\ell(r,t)|^2, \tag{67}$$

where $\eta_\ell(r, t = 0)$ is obtained by slightly "kicking" the filtered orbital $\eta_\ell(r)$ (for further details see Refs. [66, 67]). Note that we consider closed shell systems so the factor of 2 due to spin was restored here.

The sGW formalism is efficient since very few orbitals (typically $N_{\rm orb} = 10$, and less for larger systems) are needed to describe the response of molecules and solid-state systems within the short few-fs time for which $\hat{W}(t)$ is needed. Still, a total of $N_\zeta N_{\rm orb} \simeq 2000 - 40,000$ filtering

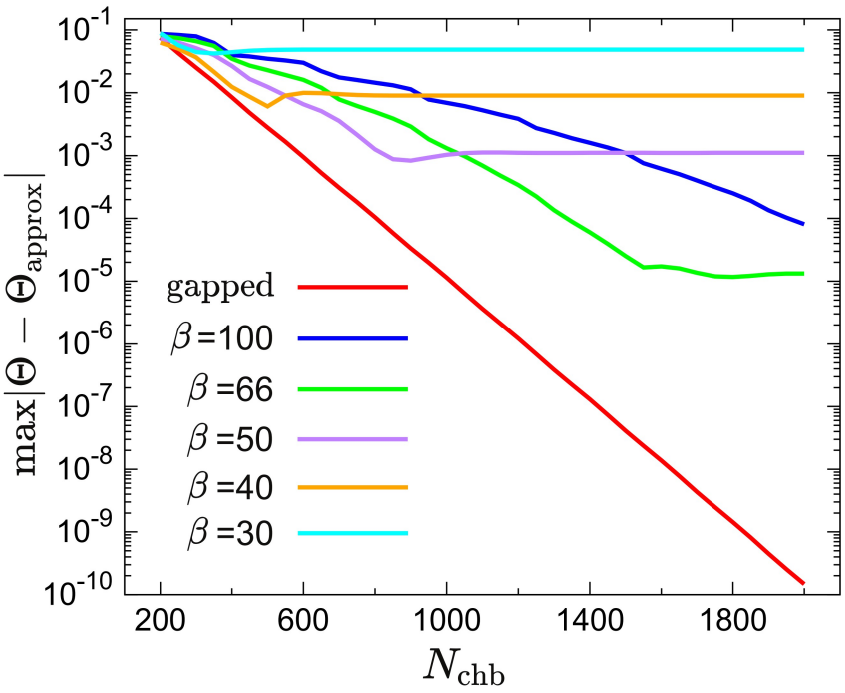

Figure 9: The maximum, taken over all energies outside the band gap, of the absolute value of the difference between the step function $\Theta(\mu - E)$ and its Chebyshev expansion (the RHS of Eq. (56)). The figure is plotted as a function of the number of Chebyshev terms, $N_{\rm chb}$. Note that smoother low-β filters converge initially faster but then their error reaches quickly a high plateau.

projections is required, therefore motivating the use of the proposed gapped-filtering approach.

4.3.2 Gapped-filtering for stochastic-GW

We now demonstrate gapped-filtering in an sGW study of naphthalene. All calculations use $N_\zeta = 1000$ stochastic samplings of G, which gives the quasi-particle energy to within a statistical error of ± 0.06 eV.

The DFT Hamiltonian $\hat{H}$ uses the LDA exchange-correlation functional with Troullier-Martins norm-conserving pseudo-potentials [79]. The grid employed has $48 \times 44 \times 24$ points with a grid spacing of 0.5 Bohr. As mentioned, the DFT bandgap for naphthalene is then $\varepsilon_{\rm gap} = 3.34$ eV, i.e., 0.123 Hartree, while the spectrum half-width ΔH is 13.5 Hartree, i.e., 110 times higher than the band gap.

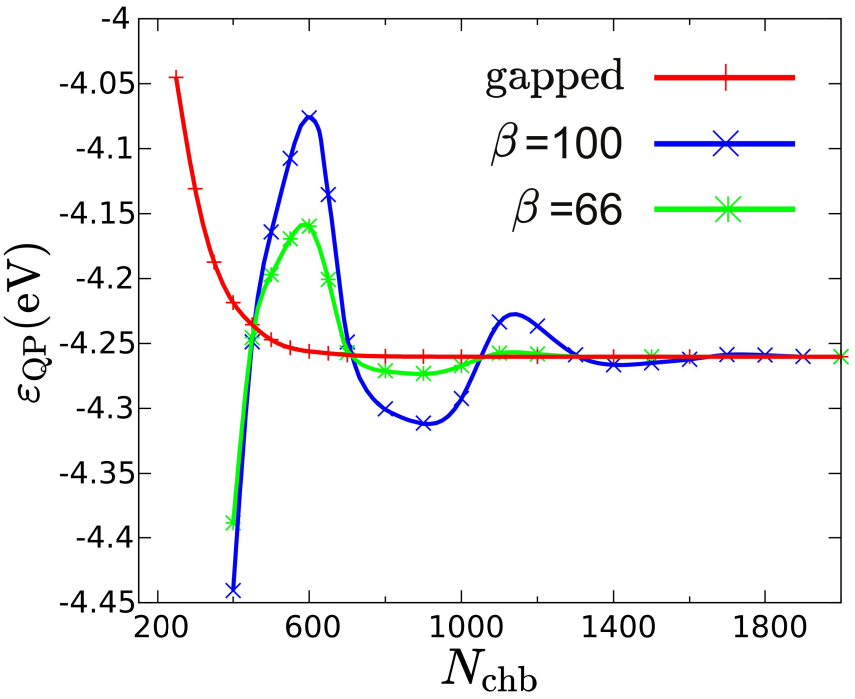

Figure 10: The stochastic-GW quasiparticle energy of naphthalene, $\varepsilon_{\rm QP}$, as a function of the number of Chebyshev polynomials, $N_{\rm chb}$, using gapped-filtering and sqrt-erfc filters with $\beta = 66$ and $\beta = 100$. Gapped-filtering converges to the asymptotic QP energy at much lower $N_{\rm chb}$ than the traditional sqrt-erfc filters.

Fig. 10 shows the resulting HOMO QP energies for gapped-filtering and for the traditional filtering method from Eq. (56) at $\beta = 66$ and $\beta = 100$. The statistical errors barely change with $N_{\rm chb}$, so we do not show them in the figure.

The figure shows that gapped-filtering requires very few Chebyshev terms. Thus, to have a systematic error of 0.015 eV or better, $N_{\rm chb} = 450$ is sufficient for gapped-filtering.

4.4 Conclusions

In conclusion, our work has shown that, for gapped systems, a step-function filter in energy is best done by optimizing the Chebyshev coefficients to fit the gapped spectrum only, rather than using a smooth filter and unnecessarily fitting the irrelevant gapped region.

The final algorithm is very easy to adopt for practical systems. Given the HOMO/LUMO energies and the spectrum overall min/max energies, one should specify an acceptable tolerance of the deviation from the step function, and then do a similar calculation the red line in Fig. 9 to extract the corresponding number of $N_{\rm chb}$ terms. With our parameters we find an acceptable tolerance at about $N_{\rm chb} \simeq 4\Delta H/(\varepsilon_L - \varepsilon_H)$, but this factor could be somewhat different depending on the application and the location of the gap relative to the spectrum minima. Our results would have immediate implication on cases such as sGW, where traditional Chebyshev filtering is expensive and the simpler alternative, explicit projection by $\hat{\rho} = \sum_{n \le N_{\rm occ}} |\phi_n\rangle\langle\phi_n|$ (see Ref. [33]), requires for large systems huge memory for storing all the occupied eigenstates.

Finally, our work indicates that when there is at least partial knowledge of the system spectrum, a polynomial expansion could be made more efficient than an expansion which is designed to be uniformly convergent as a function of energy.

Acknowledgements

We are grateful to Nadine Bradbury for useful discussions. This chapter 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 supplied through the XSEDE allocation TG-CHE170058, and the NERSC project m4022.

5 Orthogonal Projector Augmented Wave Method

5.1 Projector Augmented Wave

This section serves as an introduction to the projector augmented wave (PAW) method before orthogonal PAW (OPAW) is explained in the next section. Solving the all-electron problem, in which one considers every electron in the system, leads to two general types of wavefunctions: core and valence. Both states have very sharp oscillatory behavior near the nuclei due to the steepness of the Coulomb potential near those nuclei. The valence states, being higher in energy than the core states, are smoother and more spatially spread out. When using a plane-wave (PW) basis set, an all-electron calculation will require a large basis size corresponding to a very fine grid and a high kinetic energy cutoff to accurately describe the core and valence states near the nuclei. One method to circumvent this difficulty is the norm-conserving pseudopotential (NCPP) approach in which the more chemically inert core states are replaced with an effective potential that results in smoother pseudowavefunctions near the nuclei while reproducing correct scattering behavior after a radial cutoff. The KS equations would then only be solved for the valence states, greatly reducing the computational cost. Two important aspects when evaluating a pseudopotential are softness and transferability. Softer pseudopotentials require fewer PWs than harder ones and a more transferable pseudopotential would be able to produce reliable results for different types of systems.

A trade-off of NCPP is that large basis sets are required for first-row and transition metals as the $2p$ and $3d$ orbitals are localized in space. An alternative approach is the use of ultrasoft pseudopotentials (USPP) in which norm-conservation is dropped to allow for more parametrization and thus allowing for softer pseudopotentials [80]. Another is the augmented plane waves (APW) method where atom-centered boundary augmentation spheres are drawn. Within each sphere, the wavefunctions are expanded in terms of atom-like partial waves and envelope functions are used in the bonding regions outside the sphere [81]. Blöchl proposed the PAW method in 1994 where

he combined ideas present in the USPP and APW methods.

The central concept of PAW is to define a linear map between the smooth pseudowavefunctions, $|\tilde{\psi}_n\rangle$, and the highly oscillatory all-electron wavefunctions, $|\psi_n\rangle$:

$$|\psi_n\rangle = \hat{T}\,|\tilde{\psi}_n\rangle\,. \tag{68}$$

The transformation should only be applied to regions near the atomic nuclei as the regions away from the nuclei are already sufficiently smooth, therefore,

$$\hat{T} = \hat{I} + \sum_a \hat{T}^{(a)}, \tag{69}$$

where a indexes over all atoms in the system, and the atom-centered operator $\hat{T}^{(a)}$ is only applied within an augmentation sphere. Skipping much derivation which can be found in Refs. [82, 83], $\hat{T}^{(a)}$ can be expressed as

$$\hat{T}^{(a)} = \sum_i \left(|\phi_i^{(a)}\rangle - |\tilde{\phi}_i^{(a)}\rangle \right) \langle p_i^{(a)}|\,, \tag{70}$$

where $|\phi_i^{(a)}\rangle$ and $|\tilde{\phi}_i^{(a)}\rangle$ are the all-electron and pseudo atom-like partial waves, and $|p_i^{(a)}\rangle$ are fixed projector functions defined within an atom's augmentation sphere. The mapping between the all-electron and pseudowavefunctions is thus

$$|\psi_n\rangle = |\tilde{\psi}_n\rangle + \sum_{a,i} \left(|\phi_i^{(a)}\rangle - |\tilde{\phi}_i^{(a)}\rangle \right) \langle p_i^{(a)}|\tilde{\psi}_n\rangle\,. \tag{71}$$

This transformation leads to the pseudowavefunctions satisfying a generalized eigenvalue problem

$$\tilde{H}\tilde{\psi}_n = \epsilon_n \hat{S}\tilde{\psi}_n, \tag{72}$$

where

$$\hat{S} = \hat{T}^\dagger \hat{T} = \hat{I} + \sum_{ij,a} |p_i^{(a)}\rangle \, s_{ij}^{(a)} \, \langle p_j^{(a)}| \tag{73}$$

is an overlap operator with $s_{ij}^{(a)} \equiv \langle \phi_i^{(a)} | \phi_j^{(a)} \rangle - \langle \tilde{\phi}_i^{(a)} | \tilde{\phi}_j^{(a)} \rangle$, $\tilde{H} = \hat{T}^\dagger \hat{H} \hat{T}$, and $\langle \tilde{\psi}_i | \hat{S} | \tilde{\psi}_j \rangle = \delta_{ij}$. The PAW Hamiltonian has the form

$$\tilde{H} = -\frac{\nabla^2}{2} + V_{KS}(r) + \sum_{ij,a} |p_i^{(a)}\rangle \, D_{ij}^{(a)} \, \langle p_j^{(a)}| \tag{74}$$

where the local $V_{KS}(r)$ effective potential is composed of terms similar to those in NCPP, namely, an external, Hartree, and exchange-correlation potential. The non-local $D_{ij}^{(a)}$ terms and effective potentials depend on the on-site PAW density matrices $\rho_{ij}^{(a)} = \sum_n \langle p_j^{(a)} | \tilde{\psi}_n \rangle \, \langle \tilde{\psi}_n | p_j^{(a)} \rangle$ and the smoothed PAW density $\tilde{n}(r) = \sum_n f_n |\tilde{\psi}_n(r)|^2$. Full expressions for these terms can be found in Refs. [82, 84].

A major benefit of PAW is that lower kinetic energy cutoffs can be used than those with NCPP. In addition, the all-electron wavefunction is accessible for calculations that require core-related properties such as calculating NMR shifts [85]. Despite these advantages, the pseudowavefunctions are not orthogonal nor norm-conserving, the latter leading to the use of a compensation charge to make up for the difference in charge between the all-electron wavefunctions and pseudo wavefunctions. These factors greatly complicate the implementation of PAW as expectation values such as the total energy have to be expanded in multiple terms [82].

5.2 Orthogonal Projector Augmented Wave

Many stochastic electronic methods including sDFT and sGW were developed in the NCPP framework where the wavefunctions are orthogonal. To facilitate the task of combing PAW with these stochastic methods, orthogonal PAW (OPAW) was developed [88]. Given the generalized eigen-

problem in Eq. (72), an orthogonal transformation is be obtained:

$$\psi'_n = \hat{S}^{1/2}\tilde{\psi}_n, \tag{75}$$

resulting in

$$H'\psi'_n = \epsilon_n\psi'_n, \tag{76}$$

where $H' = \hat{S}^{-1/2}\tilde{H}\hat{S}^{-1/2}$. The task now is to develop a method that efficiently applies $\hat{S}^n$ where n is a real number.

One common assumption in PAW calculations is that the augmentation spheres do not overlap each other [86]. With this assumption, the projector functions can be independently transformed. The first step is to orthonormalize the projectors in Eq. (73). A projector overlap matrix, $L_{ij}^{(a)} = \langle p_i^{(a)}|p_j^{(a)}\rangle$ is built and then diagonalized: $L^{(a)} = U^{(a)}\lambda^{(a)}U^{(a)\dagger}$. The eigenvector matrix $U^{(a)}$ is unitary, i.e., $U^{(a)}U^{(a)\dagger} = \hat{I}$. The projectors are then rotated,

$$|\xi_i^{(a)}\rangle = \frac{1}{\sqrt{\lambda_i^{(a)}}}\sum_j U_{ji}^{(a)}\,|p_j^{(a)}\rangle\,, \tag{77}$$

so that they are orthogonal $\langle\xi_i^{(a)}|\xi_j^{(a)}\rangle = \delta_{ij}$. With this transformation, the PAW overlap operator becomes

$$\hat{S} = \hat{I} + \sum_{kl,a}|\xi_j^{(a)}\rangle\,O_{kl}^{(a)}\,\langle\xi_j^{(a)}|\,, \tag{78}$$

where $O^{(a)} = \sqrt{\lambda^{(a)}}U^{(a)}s^{(a)}U^{(a)\dagger}\sqrt{\lambda^{(a)}}$. The matrix $O^{(a)}$ is also diagonalized, $O^{(a)} = Q^{(a)}q^{(a)}Q^{(a)\dagger}$, so that

$$\hat{S} = \hat{I} + \sum_{i,a}|\zeta_i^{(a)}\rangle\,q_i^{(a)}\,\langle\zeta_i^{(a)}|\,, \tag{79}$$

where $|\zeta_i^{(a)}\rangle = \sum_l Q_{li}^{(a)}\,|\xi_l^{(a)}\rangle$.

The final step is to apply the procedure of Ono and Hirose to connect wavefunctions on differ-

ently spaced grids [87]. Typically, the projector functions are defined on a fine radial grid centered around their associated atom and interpolated onto a fine 3D cubic grid A^f. A rougher 3d cubic grid with fewer points than the fine grid is also constructed around each atom, A^r. The goal of the Ono-Hirose method is to efficiently calculate the overlap of an atom's projector functions with the pseudowavefunctions within the augmentation sphere, and this should formally be done on the fine grid. Since the pseudowavefunctions are represented on a rougher grid, an interpolation of the pseudowavefunctions from the rough grid to fine grid is needed in principal to calculate the overlap with the fine grid projectors:

$$\psi(r_f) = \sum_{r \in A^r} B(r_f, r)\psi(r), \tag{80}$$

where $B(r_f, r)$ is a linear projection matrix.

The key to the Ono-Hirose method is that the overlap of the atomic projectors and pseudowavefuntions,

$$\langle p_i^{(a)} | \psi \rangle = \sum_{r_f \in A^f} p_i^{(a)}(r_f)\psi(r_f)dv_f, \tag{81}$$

can be written in terms of an overlap on the rough grid avoiding the need to interpolate the rough-grid wavefunctions to the fine-grid:

$$\langle p_i^{(a)} | \psi \rangle = \sum_{r \in A^r} \bar{p}_i^{(a)}(r)\psi(r)dv, \tag{82}$$

where dv_f and dv are the fine and rough grid volume elements. The rough grid projectors are defined as

$$\bar{p}_i^{(a)}(r) = \frac{dv_f}{dv} \sum_{r_f \in A^f} p_i^{(a)}(r)B(r_f, r). \tag{83}$$

Ono and Hirose used cubic fitting for the smoothing matrix, $B(r_f, r)$; however, a spline fit would make the matrix separable and easier to work with [88].

With these ingredients, the fine-grid $|\zeta_i^{(a)}\rangle$ projectors in Eq. (79) are processed to be rough-grid

projectors. This processing makes these projectors no longer orthogonal on the rough grid so the orthogonalization procedure on the projectors is repeated again eventually leading to

$$\hat{S} = \hat{I} + \sum_{i,a} |\eta_i^{(a)}\rangle \, o_i^{(a)} \, \langle \eta_i^{(a)}| \,, \tag{84}$$

where $|\eta_i^{(a)}\rangle$ is orthogonal on the rough grid. With this representation, powers of $\hat{S}$ becomes

$$\hat{S}^n = \hat{I} + \sum_{i,a} |\eta_i^{(a)}\rangle \left(\left(1 + o_i^{(a)}\right)^n - 1 \right) \langle \eta_i^{(a)}| \,, \tag{85}$$

The above is derived from the identity

$$(\hat{I} + (a-1)P)^m = \hat{I} + (a^m - 1)P, \tag{86}$$

where a and m are real numbers and P is a projection operator.

The resulting OPAW pseudowavefunctions are not only orthogonal, but also norm-conserving as the transformation between the all-electron wavefunctions and the OPAW pseudowavefunctions is unitary:

$$|\psi_i\rangle = \hat{U} \, |\psi_i'\rangle = \hat{T}\hat{S}^{-1/2} \, |\psi_i'\rangle \,. \tag{87}$$

6 Time-dependent density functional theory with the orthogonal projector augmented wave method

Density functional theory (DFT) and time-dependent DFT (TDDFT) are well established ab-initio ground and excited state electronic structure methodologies. A successful tool for these methods is plane-wave basis sets where one Fourier transforms the wavefunctions between real and reciprocal space. Software that use plane-wave basis sets include Quantum Espresso [89] and VASP [90].

Closely related methods are finite element and finite difference, which use only the real-space grid, avoiding Fourier transforms. In the former, the basis set consists of local, piecewise polynomials where convergence is controlled by the number of basis functions and their polynomial orders [91]. For the latter, wavefunctions are represented by their values on the grid and the Laplacian operator is evaluated using a finite difference formula [92].

Some advantages of real-space approaches are the production of sparse matrices which make them very amenable to large-scale parallelization and the ability to use non-uniform and adaptive grids [93–96]. Successful implementations of ground state and excited state real-space code include PARSEC [96] and Octopus [97].

Both plane-wave and real-space methodologies are well suited for sufficiently smooth wavefunctions but require a large basis set for the former and denser grid points for the latter to accurately represent the highly oscillatory atomic core states. One popular method to circumvent this issue is the use of norm-conserving pseudopotentials (NCPP) in which the effects of the inert core states are replaced with an effective pseudopotential, leading to smooth pseudo valance wavefunctions that are easier to represent [50, 98]. There have been many applications of NCPP with DFT and TDDFT using plane-wave and real-space frameworks [98–101]. However, the simplicity of NCPP comes at the price of high computational costs in treating first-row elements, transition metals, and rare earth elements [102].

The projector augmented wave (PAW) method, first proposed by Blöchl, yields smoother

pseudo wavefunctions than those of NCPP by relaxing norm conservation [82]. These smoother wavefunctions have a lower kinetic energy cutoff and, therefore, can use larger grid spacing than with NCPP. The key to PAW is a linear transformation that maps the smoothed pseudo wavefunctions to the highly oscillatory all-electron wavefunctions. This transformation enables calculations of wavefunction-dependent properties that would require very high kinetic cutoffs in plane-wave NCPP [103] such as hyperfine parameters, core-level spectra, electric-field gradients, and NMR chemical shifts [85].

A trade-off of PAW is that the linear mapping results in non-orthogonality of the pseudo wavefunctions that satisfy instead a generalized eigenvalue problem complicating its use in electronic structure methods that rely on orthogonal wavefunctions such as stochastic DFT [60] or stochastic GW [32]. We recently solved this non-orthogonality problem through the combination of two techniques that were discovered earlier [88]. The first is the efficient application of powers of the PAW overlap operator to generate an orthogonal Hamiltonian and wavefunctions that are norm-preserving [88, 104]. The second is the Ono-Hirose transformation which yields accurate overlaps of coarse-grid wavefunctions with the localized dense-grid atomic projector functions [87]. With these two ingredients, our resulting orthogonal PAW (OPAW) method was demonstrated with the Chebyshev-filtered subspace iteration plane-wave DFT approach, successfully reproducing PAW band gaps from the ABINIT software [88, 105].

Many post-DFT methods, such as our stochastic GW approximation [32] and stochastic Bethe Salpeter equation methodologies [6], are easier to implement with orthogonal time-dependent wavefunctions. Here, we take the next step in developing OPAW, making it viable for TDDFT using real-time time-dependent propagation. Compared to linear-response methods, real-time propagation can calculate electronic responses to any arbitrary external stimuli, allowing for study of non-linear-response phenomena such as high-harmonic generation and exciton dynamics in photovoltaic devices [106, 107].

Our combined OPAW-TDDFT approach directly uses orthogonal wavefunctions and the Ono-

Hirose method to yield an efficient real-time propagation method which, as we will show, works well even with very large grid spacing and increased time-step. Note that other implementations of PAW have earlier been used for time-dependent propagation [108–110].

The OPAW-TDDFT method is demonstrated here with absorption spectra calculations for various organic- and chromophore-based systems. Sections 6.1 and 6.2 reviews the OPAW theory and discusses its implementation in TDDFT. In section 6.3, we present and analyze absorption spectra calculations using OPAW-TDDFT against those calculated with NCPP-TDDFT, and time-step analysis follows in 6.4. Concluding remarks follow in section 6.5.

6.1 Orthogonal Projector Augmented Wave Method

In PAW, the all-electron (AE) wavefunctions, ψ_n, are built from the pseudo (PS) wavefunctions, $\tilde{\psi}_n$, using the linear map:

$$|\psi_n\rangle = \hat{T}\,|\tilde{\psi}_n\rangle = |\tilde{\psi}_n\rangle + \sum_{a,i} \left(|\phi_i^{(a)}\rangle - |\tilde{\phi}_i^{(a)}\rangle \right) \langle p_i^{(a)}|\tilde{\psi}_n\rangle\,, \tag{88}$$

where a indexes the atoms and i stands for a combination of angular, magnetic, and principal quantum numbers to label each partial wave channel associated with an atom. The AE ($\phi_i^{(a)}$) and PS ($\tilde{\phi}_i^{(a)}$) partial wave channels equal each other outside a spherical augmentation region around each atom. The atomic projector functions, $p_i^{(a)}$, are localized within the augmentation regions and are dual to the PS partial waves there (i.e. $\sum_i |\tilde{\phi}_i^{(a)}\rangle \langle p_i^{(a)}| \simeq 1$).

Since the PS wavefunctions are not orthogonal, they fulfill a generalized eigenproblem,

$$\tilde{H}\tilde{\psi}_n = \varepsilon_n \hat{S}\tilde{\psi}_n, \tag{89}$$

where

$$\hat{S} = \hat{T}^\dagger\hat{T} = \hat{I} + \sum_{ij,a} |p_i^{(a)}\rangle\, s_{ij}^{(a)}\, \langle p_j^{(a)}| \tag{90}$$

is an overlap matrix with $s_{ij}^{(a)} \equiv \langle \phi_i^{(a)} | \phi_j^{(a)} \rangle - \langle \tilde{\phi}_i^{(a)} | \tilde{\phi}_j^{(a)} \rangle$, and

$$\tilde{H} = -\frac{\nabla^2}{2} + v_{KS}(r) + \sum_{ij,a} |p_i^{(a)}\rangle \, D_{ij}^{(a)} \, \langle p_j^{(a)}| \,. \tag{91}$$

Full details of the effective potential, $v_{KS}(r)$, and the non-local term coefficients ($D_{ij}^{(a)}$) can be found in Refs. [82, 84].

To make OPAW, we rotate the PS wavefunctions to yield orthogonal ones:

$$\psi'_n = \hat{S}^{1/2} \tilde{\psi}_n, \tag{92}$$

resulting in

$$H' \psi'_n = \varepsilon_n \psi'_n, \tag{93}$$

where $H' = \hat{S}^{-1/2} \tilde{H} \hat{S}^{-1/2}$ is the OPAW Hamiltonian. To efficiently approximate $\hat{S}^n$, where n is any real number, one first assumes that the augmentation regions of different atoms do not overlap so that the projector functions can be separately rotated around each atom [88]. With this assumption, $\hat{S}$ is readily transformed into

$$\hat{S} = \hat{I} + \sum_{i,a} |\eta_i^{(a)}\rangle \, o_i^{(a)} \, \langle \eta_i^{(a)}| \,, \tag{94}$$

where the rotated projectors, $\eta_i^{(a)}$, are orthogonal so that $\langle \eta_i^{(a)} | \eta_j^{(a')} \rangle = \delta_{ij} \delta_{a,a'}$ and $o_i^{(a)}$ are derived from transformations applied to $s_{ij}^{(a)}$. We verified earlier [88] that the projector locality assumption yields accurate band gaps, and will show below how it is also sufficient for time-dependent properties. Any power of $\hat{S}$ is then easily expressed as

$$\hat{S}^n = \hat{I} + \sum_{i,a} |\eta_i^{(a)}\rangle \, [(1 + o_i^{(a)})^n - 1] \, \langle \eta_i^{(a)}| \,. \tag{95}$$

The second key to the efficacy of OPAW is the use of the Ono-Hirose method [87] to efficiently calculate the overlap of the dense-grid projector functions with the coarse-grid wavefunctions in any application of $\hat{S}^n$ and $\hat{S}$. In the method, a smoothing matrix that connects the dense-grid and the coarse-grid is constructed using spline interpolation in the x, y, and z directions which when applied onto the dense-grid projector functions generates coarse-grid projector functions. These coarse-grid projector functions are then used in calculating coarse-grid overlaps. Further details of the transformation of the projectors and the application of the Ono-Hirose method are given in Ref. [88].

6.2 Time-Dependent Density Functional Theory with the Orthogonal Projector Augmented Wave Method

One common use of TDDFT is the calculation of absorption spectra of materials and molecules. In the linear response regime, the dipole absorption cross section tensor, $\bar{\sigma}$, is formally [111]

$$\bar{\sigma}_{ij}(\omega) = \frac{4\pi\omega}{c}\mathrm{Im}\bigg(\int dr d\bar{r}\, r_i \tilde{\chi}(r,\bar{r},\omega)\bar{r}_j\bigg), \tag{96}$$

Where $i = (x,y,z)$ and similarly for j. Eq. (96) is then obtained through the linear-response dipole-dipole correlation function $d_{ii}(t) = \int dr\, r_i \Delta n_i(r_i,t)$ where

$$\Delta n_i(r,t) = \int d\bar{r}\, \tilde{\chi}(r,\bar{r},t)\bar{r}_i = \frac{1}{\gamma}\bigg(n^{\gamma}(r,t) - n^{\gamma=0}(r,t)\bigg) \tag{97}$$

is the induced charged density and

$$n^{\gamma}(r,t) = \sum_n f_n |\Psi'^{\gamma}_n(r,t)|^2 \tag{98}$$

are the perturbed and unperturbed densities made from the time-dependent OPAW wavefunctions, labeled by γ, and detailed later.

A component of the absorption cross section tensor is then given by a Fourier transform of the dipole-dipole correlation function. For example, the xx component is

$$\bar{\sigma}_{xx}(\omega) = \frac{4\pi\omega}{c}\mathrm{Im}\left(\int_0^\infty dt d_{xx}(t)\xi(t)e^{i\omega t}\right). \tag{99}$$

The damping function $\xi(t)$ is included to avoid Gibbs oscillations. We usually use $\xi(t) = \cos^4(\eta\pi t)$ where $\eta = 1/(2t_{tot})$ so the damping factor goes to 0 at the total simulation time, t_{tot}. The absorption spectrum is the trace of the tensor, i.e. $\sigma(\omega) = \frac{1}{3}\sum_{i=(x,y,z)}\bar{\sigma}_{ii}(\omega)$.

The time-dependent Schrödinger equation in the PAW framework has the form

$$i\hat{S}\frac{\partial}{\partial t}\tilde{\Psi}_n^\gamma(r,t) = \tilde{H}\tilde{\Psi}_n^\gamma(r,t), \tag{100}$$

where $\tilde{\Psi}_n^\gamma(r,t)$ is a time-dependent PS wavefunction evolving under the PAW Hamiltonian. With the transformations in Eqs. (92) and (93), this equation in the OPAW framework becomes

$$i\frac{\partial}{\partial t}\Psi_n'^\gamma(r,t) = H'(t)\Psi_n'^\gamma(r,t), \tag{101}$$

where $\Psi_n'^\gamma(r,t)$ is the time-dependent orthogonalized PS wavefunction. The initial condition of the OPAW wavefunction in Eq. (98), is

$$\Psi_n'^\gamma(r,t=0) = e^{-i\gamma r_i'}\psi_n'(r) \tag{102}$$

where $r_i' = \hat{S}^{-1/2}r_i\hat{S}^{-1/2}$. In practice we find that we could replace r_i' by r_i without any significant change in the results.

For simplicity, we use fourth-order Runge-Kutta to approximate the time-propagation of the

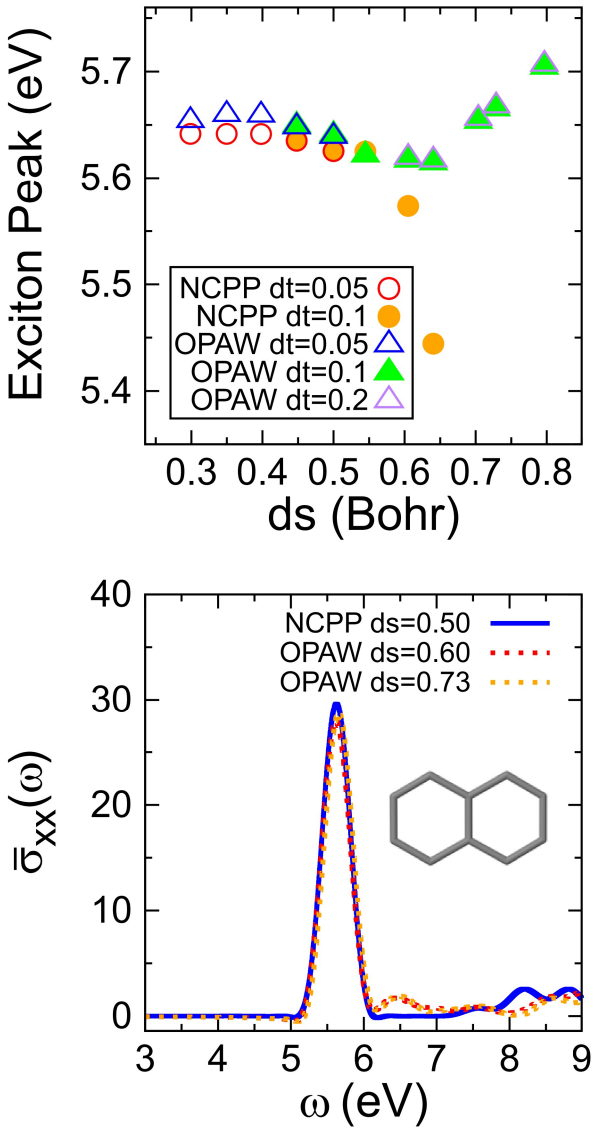

Figure 11: The top panel shows the exciton peak of the $\bar{\sigma}_{xx}(\omega)$ tensor component for naphthalene vs. grid spacing with various time-steps (in a.u.) using OPAW and NCPP. The bottom panel shows the naphthalene $\bar{\sigma}_{xx}(\omega)$ calculated using OPAW for very large grid spacings (in Bohr) and a large 0.2 a.u time-step. The results match an NCPP simulation at a lower grid spacing and a smaller 0.1 a.u. time-step.

OPAW wavefunctions in Eq. (97). Compared with other typically used approaches, e.g., the split-operator method, the Runge-Kutta approach is simple to implement as it does not require exponentiation of non-local terms in the Hamiltonian; for further details see Sec. 6.4.

6.3 Results

We demonstrate below OPAW-TDDFT and show that it allows the use of much larger grid spacing than NCPP-TDDFT. For simplicity, we abbreviate the two methods as OPAW and NCPP, omitting the TDDFT label.

For both methods, the ground state wavefunctions were calculated with the Chebyshev-filtered subspace iteration method using plane-waves[105]. The DFT Hamiltonian used the LDA functional. The simulations were non-periodic and employed the Martyna-Tuckerman approach for the Coulombic interactions [112]. For OPAW, we employed the recommended LDA-based atomic datasets from the ABINIT website [113], while for NCPP we used the Hamann form for the pseudopotentials [102]. The simulation boxes for all calculations were chosen so that there was approximately 6 Bohr of padding from the edges of the system in the x, y, and z directions.

The time-dependent propagations for both OPAW and NCPP used the Runge-Kutta method, with time-steps that were either 0.05, 0.1 or 0.2 a.u. Specifically, most results shown here use the largest time-step for which the simulations are stable for the specific system and grid spacing. Generally, OPAW enables the use of larger time-steps than what is possible with NCPP, as detailed below.

Since for each molecule the box size is unchanged between runs, the grid spacings are generally slightly different in the x, y, and z directions, and therefore we report the results against the grid spacings' geometrical averages, $ds \equiv (dx\, dy\, dz)^{1/3} = dV^{1/3}$.

We first study the convergence with grid spacing using a small molecule (naphthalene). The simulations used a box size of $28\ \mathrm{Bohr} \times 26\ \mathrm{Bohr} \times 16\ \mathrm{Bohr}$ with ds in the range of 0.3 Bohr to

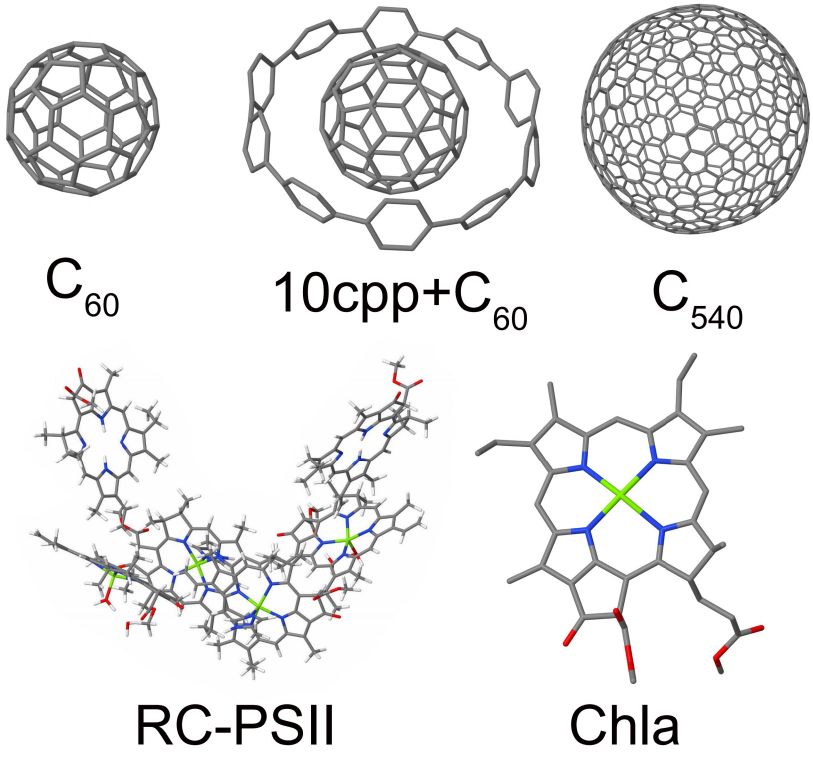

Figure 12: Structures and abbreviations for all systems used in this section.

0.8 Bohr and only the $\bar{\sigma}_{xx}(\omega)$ element of the absorption tensor is calculated.

Fig. 11 shows the exciton peaks (i.e., first significant absorption peak) for several $\bar{\sigma}_{xx}(\omega)$ spectra calculated at different grid spacings and time-steps for NCPP and OPAW. Compared to NCPP, OPAW shows excellent spectral peaks at much larger grid spacings. Specifically, note that even at $ds = 0.73$ Bohr, the OPAW exciton peak is still acceptable, i.e., different from the low-ds values by only ˜0.05 eV. We also show in Fig. 11 that the spectra for OPAW at high grid spacings are well converged compared to NCPP at lower ones.

Our convergence tests with naphthalene demonstrate that grid spacing of 0.6-0.7 Bohr are sufficient to converge OPAW at a level that requires 0.4-0.5 Bohr for NCPP. OPAW reduces the number of grid points in a single dimension by about 20%-40%, which when taken over all three dimensions, corresponds to a total number of grid points 2-4 less than that of NCPP..

Our next step is to demonstrate the power of the OPAW on a series of five larger systems, shown in Fig. 12. Included are C_{60} fullerene, C_{60} fullerene embedded inside a 10 para-substituted phenyl "nanohoop" cycloparaphenylene ring (10cpp+C_{60}) [114], C_{540} fullerene, a Chlorophyll-a chromophore with a methyl acetate ligand in place of the phytyl chain (Chla)[115], and a hexameric reaction center of photosystem II consisting of six chromophores (RC-PSII) [115]. A total simulation time of 2000 a.u. was used for the fullerenes and 1000 a.u. for the Chlorophyll based systems which is sufficient to isolate specific peaks to better than 0.05 eV for the former and 0.1 eV for the latter.

As shown in Fig. 13, OPAW generates comparable spectra to that of NCPP at much larger grid spacing for C_{60}. Similarly for the larger 10cpp+C_{60} and C_{540} systems, OPAW also yields converged spectra at large grid spacing, around 0.6-0.7 Bohr.

In addition, we compare in Fig. 14 the C_{60} spectra to the NCPP real-space time-dependent local-density approximation (TDLDA) calculations of Yabana and Bertsch [98]. An exponential damping function, $\xi(t) = \exp(-\eta t)$, was used in Eq. (99) for the NCPP and OPAW runs, with $\eta = 0.0125$ a.u., so that the energy resolution is comparable to that used in Ref. [98]. The Yabana-

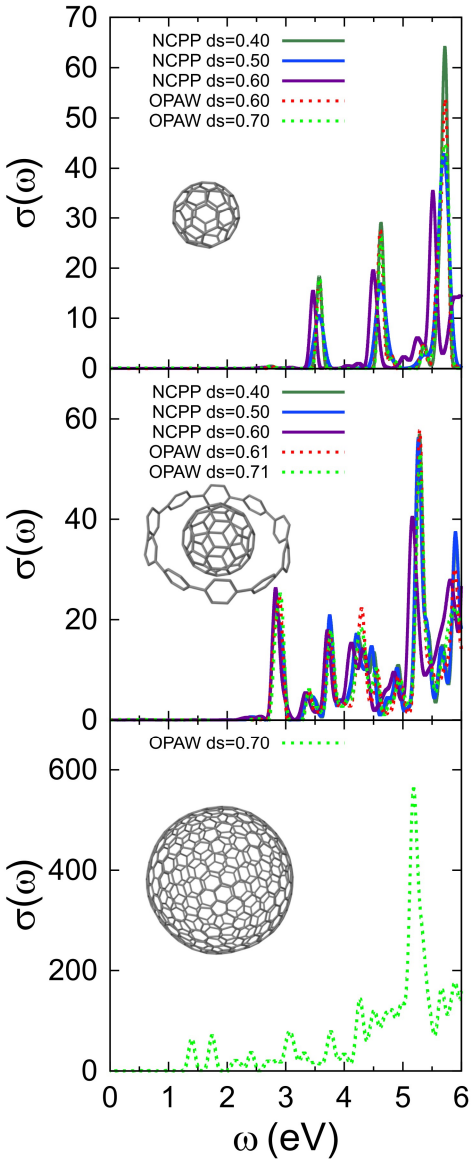

Figure 13: The top panel shows the OPAW absorption spectra of C_{60} for large grid spacing with a 0.2 a.u. time-step; these spectra match the NCPP results which required smaller grid spacings and a smaller 0.1 a.u. time-step. The middle and bottom panels show the OPAW absorption spectra for 10cpp+C_{60} and C_{540} respectively for large grid spacings with a 0.2 a.u. time-step.

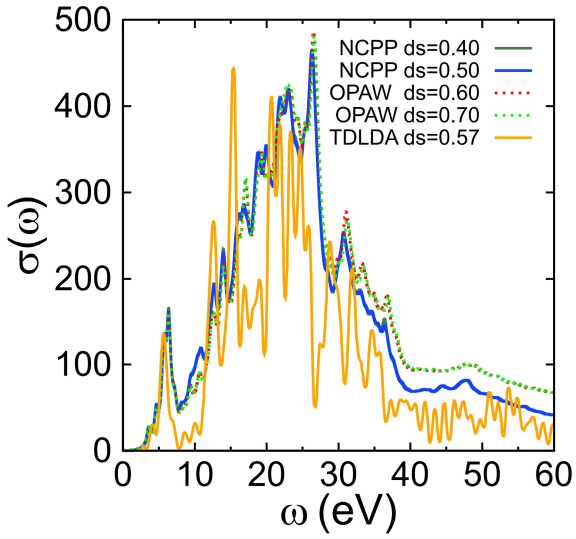

Figure 14: The OPAW and NCPP absorption spectra of C_{60} at larger energies at various grid spacings with a 0.2 a.u. time-step for the former and 0.1 a.u. time-step for the latter along with the TDLDA absorption spectra from Ref. [98].

Berstch TDLDA spectra were scaled so the first large peak height around $\sim$6 eV is similar to our NCPP and OPAW runs. The OPAW and NCPP runs are separately converged with grid spacing. Up to about 30 eV, OPAW and NCPP match well, and both match overall the trends of the Yabana-Bertsch data although there are significant differences at specific ranges of energies. Above 30 eV the overall trends remain but the spectra are quantitatively different between OPAW and NCPP, most likely due to the difference in methodology.

Compared to the previous systems, Chla and RC-PSII require slightly smaller grid spacing for both NCPP and OPAW, but OPAW still yields converged results at much larger spacing of 0.65 Bohr compared to 0.45 Bohr with NCPP. Taken over the 3D grid this yields a factor of 3 reduction in the total number of grid points. Although some of the peak shapes of the OPAW spectra of Chla in Fig. 15 start to differ from lower NCPP at energies greater than 3 eV, the lower-energy peak positions match well.

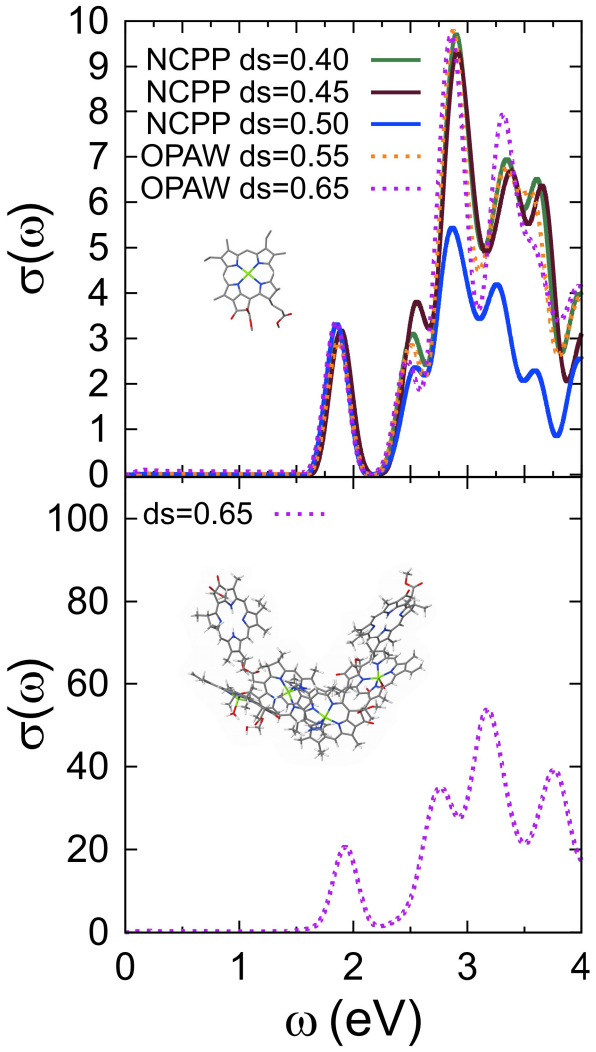

Figure 15: The top and bottom panels show the OPAW absorption spectra of Chla and RC-PSII respectively for large grid spacings with a 0.1 a.u. time-step

6.4 Computation Time

We now turn to a discussion of the time requirements and computational savings in OPAW. We verified that the time to construct the Hamiltonian and perform Runge-Kutta propagation at every time-step is similar between OPAW and NCPP given the same time-step and grid and that both scale linearly with number of orbitals and the number of grid points (i.e., quadratically overall with system size). Thus, for our largest system (RC-PSII) with 676 occupied orbitals and a grid with 987,840 points, the total wall time to perform a single OPAW-TDDFT calculation on RC-PSII, using a 6 modern node with 676 AMD-ROME cores, was close to a wall time of 44 hours, i.e., about 30,000 CPU core hours; the times for the smaller systems are correspondingly smaller.

When comparing OPAW and NCPP, one should note that NCPP could have used a non-Runge-Kutta approach, i.e., a symplectic (split-operator) propagation (a symplectic representation of OPAW is more complicated and was not pursued here). But while symplectic propagation is fast per time-step, it necessitates a shorter time-step (for the split operator, we verified that for most systems the time-step can be at most 0.05 a.u. with NCPP). The costs per time-step are different in the two methods, and we find that a symplectic propagation typically requires half the cost of Runge-Kutta, due to the balance of several terms:

- In Runge-Kutta, the cost per time-step includes four Hamiltonian operations and a single calculation and "broadcasting" of the time-dependent potential across all cores;
- in symplectic methods, the costs include calculating the kinetic energy propagator (the most expensive part), two local and non-local potentials, and the calculation and "broadcasting" of the time-dependent potential.

Thus, overall, the cost of NCPP with symplectic propagation with a 0.05 a.u. time-step is comparable to that of using Runge-Kutta with a 0.1 a.u. time-step. Given that the Runge-Kutta time-step is 0.1 or 0.2 a.u. in our calculations, the lack of symplectic propagation is not a problem for OPAW.

6.5 Conclusions

Our OPAW method has been implemented here in the TDDFT framework using fourth-order Runge-Kutta for the time propagation. We demonstrated that the convergence of OPAW with respect to the grid spacing and time-step is achieved at much larger values compared to that of NCPP, resulting in significant computational savings.

We also showed, using a series of hydrocarbon and chromophore based systems, that OPAW successfully produces converged spectra at large grid spacing and time-step. Future work will focus on implementing OPAW into our other post-DFT stochastic methods that require orthogonal time-dependent wavefunctions, namely stochastic GW [32] and Bethe-Salpeter equation [6, 116] methods.

Acknowledgments

We are grateful to Nadine Bradbury and Tucker Allen for useful discussion. This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program, under Award No. DE-SC0022198. This work used the Expanse cluster at San Diego Supercomputer Center through allocation PHY220143 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program [117], which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

References

1. Nguyen, M., Duong, T. & Neuhauser, D. Time-dependent density functional theory with the orthogonal projector augmented wave method. *The Journal of Chemical Physics* **160** (2024).

2. Bradbury, N. C., Allen, T., Nguyen, M. & Neuhauser, D. Deterministic/Fragmented-Stochastic Exchange for Large-Scale Hybrid DFT Calculations. *Journal of Chemical Theory and Computation* **19,** 9239–9247 (2023).

3. Bradbury, N. C., Allen, T., Nguyen, M., Ibrahim, K. Z. & Neuhauser, D. Optimized attenuated interaction: Enabling stochastic Bethe–Salpeter spectra for large systems. *The Journal of Chemical Physics* **158** (2023).

4. Nguyen, M. & Neuhauser, D. Gapped-filtering for efficient Chebyshev expansion of the density projection operator. *Chemical Physics Letters* **806,** 140036 (2022).

5. Yeo, C., Nguyen, M. & Wang, L.-P. Benchmarking Density Functionals, Basis Sets, and Solvent Models in Predicting Thermodynamic Hydricities of Organic Hydrides. *The Journal of Physical Chemistry A* **126,** 7566–7577 (2022).

6. Bradbury, N. C., Nguyen, M., Caram, J. R. & Neuhauser, D. Bethe–Salpeter equation spectra for very large systems. *The Journal of Chemical Physics* **157** (2022).

7. Nguyen, M., Li, W., Li, Y., Rabani, E., Baer, R. & Neuhauser, D. Tempering stochastic density functional theory. *The Journal of Chemical Physics* **155** (2021).

8. Jang, H., Qiu, Y., Hutchings, M. E., Nguyen, M., Berben, L. A. & Wang, L.-P. Quantum chemical studies of redox properties and conformational changes of a four-center iron CO 2 reduction electrocatalyst. *Chemical Science* **9,** 2645–2654 (2018).

9. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136,** B864–B871. `https://link.aps.org/doi/10.1103/PhysRev.136.B864` (3B Nov. 1964).

10. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **140,** A1133 (1965).

11. Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87,** 897–923. https://link.aps.org/doi/10.1103/RevModPhys.87.897 (3 Aug. 2015).

12. Van Mourik, T., Buhl, M. & Gaigeot, M.-P. *Density functional theory across chemistry, physics and biology* 2014.

13. Kohn, W. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.* **71,** 1253–1266. https://link.aps.org/doi/10.1103/RevModPhys.71.1253 (5 Oct. 1999).

14. Perdew, J. P. Density functional theory and the band gap problem. *International Journal of Quantum Chemistry* **28,** 497–523 (1985).

15. Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Physical Review* **139,** A796 (1965).

16. Reining, L. The GW approximation: content, successes and limitations. *WIREs Computational Molecular Science* **8,** e1344. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1344. https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1344 (2018).

17. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B* **34,** 5390–5413. https://link.aps.org/doi/10.1103/PhysRevB.34.5390 (8 Aug. 1986).

18. Vlcek, V., Rabani, E., Neuhauser, D. & Baer, R. Stochastic GW calculations for molecules. *Journal of chemical theory and computation* **13,** 4997–5003 (2017).

19. Aryasetiawan, F. & Gunnarsson, O. The GW method. *Reports on Progress in Physics* **61,** 237 (1998).

20. Reining, L. The GW approximation: content, successes and limitations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8,** e1344 (2018).

21. Faber, C., Boulanger, P., Attaccalite, C., Duchemin, I. & Blase, X. Excited states properties of organic molecules: From density functional theory to the GW and Bethe–Salpeter Green's function formalisms. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **372,** 20130271 (2014).

22. Deslippe, J., Samsonidze, G., Strubbe, D. A., Jain, M., Cohen, M. L. & Louie, S. G. BerkeleyGW: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. *Computer Physics Communications* **183,** 1269–1289 (2012).

23. Shishkin, M. & Kresse, G. Self-consistent G W calculations for semiconductors and insulators. *Physical Review B* **75,** 235102 (2007).

24. Pham, T. A., Nguyen, H.-V., Rocca, D. & Galli, G. G W calculations using the spectral decomposition of the dielectric matrix: Verification, validation, and comparison of methods. *Physical Review B* **87,** 155148 (2013).

25. Nguyen, H.-V., Pham, T. A., Rocca, D. & Galli, G. Improving accuracy and efficiency of calculations of photoemission spectra within the many-body perturbation theory. *Physical Review B* **85,** 081101 (2012).

26. Hutchinson, M. F. A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun Stat Simul Comput.* **19,** 433–450 (1990).

27. Baer, R., Neuhauser, D. & Rabani, E. Stochastic Vector Techniques in Ground-State Electronic Structure. *Annual Review of Physical Chemistry* **73,** 255–272. ISSN: 1545-1593. `https://www.annualreviews.org/content/journals/10.1146/annurev-physchem-090519-045916` (2022).

28. Baer, R., Neuhauser, D. & Rabani, E. Self-Averaging Stochastic Kohn-Sham Density-Functional Theory. en. *Phys. Rev. Lett.* **111,** 106402. ISSN: 0031-9007, 1079-7114. https://link.aps.org/doi/10.1103/PhysRevLett.111.106402 (2019) (Sept. 2013).

29. Chen, M., Baer, R., Neuhauser, D. & Rabani, E. Energy window stochastic density functional theory. en. *J. Chem. Phys.* **151,** 114116. ISSN: 0021-9606, 1089-7690. http://aip.scitation.org/doi/10.1063/1.5114984 (2021) (Sept. 2019).

30. Li, W., Chen, M., Rabani, E., Baer, R. & Neuhauser, D. Stochastic embedding DFT: Theory and application to p-nitroaniline in water. *The Journal of Chemical Physics* **151,** 174115. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5110226/13806620/174115_1_online.pdf. https://doi.org/10.1063/1.5110226 (Nov. 2019).

31. Chen, M., Baer, R., Neuhauser, D. & Rabani, E. Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials. *The Journal of Chemical Physics* **150,** 034106. ISSN: 0021-9606. eprint: https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.5064472/15553348/034106_1_online.pdf. https://doi.org/10.1063/1.5064472 (Jan. 2019).

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. *Physical Review Letters* **113,** 076402 (2014).

33. Vlček, V., Li, W., Baer, R., Rabani, E. & Neuhauser, D. Swift G W beyond 10,000 electrons using sparse stochastic compression. *Physical Review B* **98,** 075107 (2018).

34. Jones, R. O. Density functional theory: Its origins, rise to prominence, and future. en. *Reviews of Modern Physics* **87,** 897–923. ISSN: 0034-6861, 1539-0756. https://link.aps.org/doi/10.1103/RevModPhys.87.897 (2019) (Aug. 2015).

35. Kolb, B. & Thonhauser, T. Molecular Biology at the Quantum Level: Can Modern Density Functional Theory Forge the Path? *Nano LIFE* **2.** http://www.worldscientific.com/doi/pdf/10.1142/S1793984412300063 (2012).

36. Selli, D., Fazio, G. & Di Valentin, C. Modelling realistic TiO_2 nanospheres: A benchmark study of SCC-DFTB against hybrid DFT. *The Journal of Chemical Physics* **147.** Publisher: AIP Publishing LLC, 164701 (2017).

37. Freysoldt, C., Grabowski, B., Hickel, T., Neugebauer, J., Kresse, G., Janotti, A. & Van de Walle, C. G. First-principles calculations for point defects in solids. *Rev. Mod. Phys.* **86.** Publisher: American Physical Society, 253–305. https://link.aps.org/doi/10.1103/RevModPhys.86.253 (Mar. 2014).

38. Baer, R. & Head-Gordon, M. Sparsity of the Density Matrix in Kohn-Sham Density Functional Theory and an Assessment of Linear System-Size Scaling Methods. *Phys. Rev. Lett.* **79,** 3962–3965. https://link.aps.org/doi/10.1103/PhysRevLett.79.3962 (2018) (Nov. 1997).

39. Scuseria, G. E. Linear scaling density functional calculations with Gaussian orbitals. *J. Phys. Chem. A* **103,** 4782–4790 (1999).

40. Ozaki, T. Efficient low-order scaling method for large-scale electronic structure calculations with localized basis functions. *Phys. Rev. B* **82,** 075131 (2010).

41. Romero-Muniz, C., Nakata, A., Pou, P., Bowler, D. R., Miyazaki, T. & Pérez, R. High-accuracy large-scale DFT calculations using localized orbitals in complex electronic systems: The case of graphene–metal interfaces. *Journal of Physics: Condensed Matter* **30.** Publisher: IOP Publishing, 505901 (2018).

42. Neuhauser, D., Baer, R. & Rabani, E. Communication: Embedded fragment stochastic density functional theory. *J. Chem. Phys.* **141,** 041102 (2014).

43. Chen, M., Baer, R., Neuhauser, D. & Rabani, E. Overlapped embedded fragment stochastic density functional theory for covalently-bonded materials. *J. Chem. Phys.* **150,** 034106. ISSN: 0021-9606. https://aip.scitation.org/doi/10.1063/1.5064472 (2019) (Jan. 2019).

44. Fabian, M. D., Shpiro, B., Rabani, E., Neuhauser, D. & Baer, R. Stochastic density functional theory. en. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **10.1002/wcms.1412,** e1412. ISSN: 1759-0884. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1412 (2019) (2019).

45. Chen, M., Baer, R., Neuhauser, D. & Rabani, E. Stochastic density functional theory: Real- and energy-space fragmentation for noise reduction. en. *J. Chem. Phys.* **154,** 204108. ISSN: 0021-9606, 1089-7690. https://aip.scitation.org/doi/10.1063/5.0044163 (2021) (May 2021).

46. Baer, R. & Head-Gordon, M. Energy renormalization-group method for electronic structure of large systems. *Physical Review B-Condensed Matter* **58,** 15296–15299 (1998).

47. Kosloff, R. Time-Dependent Quantum-Mechanical Methods for Molecular- Dynamics. *J. Phys. Chem.* **92,** 2087–2100 (1988).

48. Sankey, O. F., Drabold, D. A. & Gibson, A. Projected random vectors and the recursion method in the electronic-structure problem. *Phys. Rev. B* **50,** 1376 (1994).

49. Perdew, J. & Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation-Energy. *Phys. Rev. B* **45,** 13244–13249 (1992).

50. Troullier, N. & Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. *Phys. Rev. B* **43,** 1993–2006 (1991).

51. Kleinman, L. & Bylander, D. M. Efficacious Form for Model Pseudopotentials. en. *Phys. Rev. Lett.* **48,** 1425–1428. ISSN: 0031-9007. https://link.aps.org/doi/10.1103/PhysRevLett.48.1425 (2020) (May 1982).

52. 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.* **110,** 2810–2821 (1999).

53. Neuhauser, D., Baer, R. & Zgid, D. Stochastic self-consistent second-order Green's function method for correlation energies of large electronic systems. *J. Chem. Theory Comput.* **13,** 5396–5403 (2017).

54. Shao, J. & Wu, C. F. J. A General Theory for Jackknife Variance Estimation. *The Annals of Statistics* **17,** 1176–1197. ISSN: 00905364. http://www.jstor.org/stable/2241717 (1989).

55. Arnon, E., Rabani, E., Neuhauser, D. & Baer, R. Equilibrium configurations of large nanostructures using the embedded saturated-fragments stochastic density functional theory. *J. Chem. Phys.* **146,** 224111 (June 2017).

56. Arnon, E., Rabani, E., Neuhauser, D. & Baer, R. Efficient Langevin dynamics for "noisy" forces. *The Journal of Chemical Physics* **152,** 161103. https://doi.org/10.1063/5.0004954 (Apr. 2020).

57. Baer, R. & Head-Gordon, M. Chebyshev expansion methods for electronic structure calculations on large molecular systems. *The Journal of Chemical Physics* **107,** 10003–10013 (1997).

58. Jay, L. O., Kim, H., Saad, Y. & Chelikowsky, J. R. Electronic structure calculations for plane-wave codes without diagonalization. *Computer Physics Communications* **118,** 21–30 (1999).

59. Liang, W., Saravanan, C., Shao, Y., Baer, R., Bell, A. T. & Head-Gordon, M. Improved Fermi operator expansion methods for fast electronic structure calculations. *The Journal of Chemical Physics* **119,** 4117–4125 (2003).

60. Baer, R., Neuhauser, D. & Rabani, E. Self-averaging stochastic Kohn-Sham density-functional theory. *Physical Review Letters* **111,** 106402 (2013).

61. Neuhauser, D., Baer, R. & Rabani, E. Embedded fragment stochastic density functional theory. *Journal of Chemical Physics* **141,** 041102 (2014).

62. Baer, R., Neuhauser, D. & Rabani, E. Stochastic Vector Techniques in Ground-State Electronic Structure. *Annual Review of Physical Chemistry* **73** (2022).

63. Friedrich, C. & Schindlmayr, A. Many-body perturbation theory: the GW approximation. *NIC Series* **31,** 335 (2006).

64. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B* **34,** 5390 (1986).

65. Vlcek, V. Stochastic many-body methods for quasiparticle excitations in realistic nanoscale systems. *Bulletin of the American Physical Society* **66** (2021).

66. Neuhauser, D., Rabani, E. & Baer, R. Expeditious stochastic approach for MP2 energies in large electronic systems. *Journal of Chemical theory and Computation* **9,** 24–27 (2013).

67. Gao, Y., Neuhauser, D., Baer, R. & Rabani, E. Sublinear scaling for time-dependent stochastic density functional theory. *The Journal of Chemical Physics* **142,** 034106 (2015).

68. Tal-Ezer, H. & Kosloff, R. An accurate and efficient scheme for propagating the time dependent Schrödinger equation. *The Journal of Chemical Physics* **81,** 3967–3971. `https://doi.org/10.1063/1.448136` (Nov. 1984).

69. Kosloff, R. Time-dependent quantum-mechanical methods for molecular dynamics. *The Journal of Physical Chemistry* **92,** 2087–2100 (1988).

70. Weiße, A., Wellein, G., Alvermann, A. & Fehske, H. The kernel polynomial method. *Reviews of Modern Physics* **78,** 275–306. https://link.aps.org/doi/10.1103/RevModPhys.78.275 (1 Mar. 2006).

71. Mandelshtam, V., Ravuri, T. & Taylor, H. Calculation of the density of resonance states using the stabilization method. *Physical Review Letters* **70,** 1932 (1993).

72. Suryanarayana, P. On spectral quadrature for linear-scaling Density Functional Theory. *Chemical Physics Letters* **584,** 182–187 (2013).

73. 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. *The Journal of Chemical Physics* **102,** 8011–8022 (1995).

74. McClellan, J., Parks, T. & Rabiner, L. A computer program for designing optimum FIR linear phase digital filters. *IEEE Transactions on Audio and Electroacoustics* **21,** 506–526 (1973).

75. Parks, T. W. & Burrus, C. S. *Digital Filter Design* (J. Wiley, 1987).

76. Weiße, A., Wellein, G., Alvermann, A. & Fehske, H. The kernel polynomial method. *Reviews of Modern Physics* **78,** 275–306 (2006).

77. Dongarra, J. J., Moler, C. B., Bunch, J. R. & Stewart, G. W. *LINPACK users' guide* (SIAM, 1979).

78. Lawson, C. L., Hanson, R. J., Kincaid, D. R. & Krogh, F. T. Basic linear algebra subprograms for Fortran usage. *ACM Transactions on Mathematical Software (TOMS)* **5,** 308–323 (1979).

79. Hamann, D. Optimized norm-conserving Vanderbilt pseudopotentials. *Physical Review B* **88,** 085117 (2013).

80. Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41,** 7892–7895. https://link.aps.org/doi/10.1103/PhysRevB.41.7892 (11 Apr. 1990).

81. Madsen, G. K., Blaha, P., Schwarz, K., Sjöstedt, E. & Nordström, L. Efficient linearization of the augmented plane-wave method. *Physical Review B* **64,** 195134 (2001).

82. Blöchl, P. E. Projector augmented-wave method. *Physical Review B* **50,** 17953 (1994).

83. Rostgaard, C. The projector augmented-wave method. *arXiv preprint arXiv:0910.1921* (2009).

84. Torrent, M., Jollet, F., Bottin, F., Zérah, G. & Gonze, X. Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure. *Computational Materials Science* **42,** 337–351 (2008).

85. Pickard, C. J. & Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Physical Review B* **63,** 245101 (2001).

86. Zwanziger, J. Computation of NMR observables: Consequences of projector-augmented wave sphere overlap. *Solid State Nuclear Magnetic Resonance* **80,** 14–18 (2016).

87. Ono, T. & Hirose, K. Timesaving double-grid method for real-space electronic-structure calculations. *Physical Review Letters* **82,** 5016 (1999).

88. Li, W. & Neuhauser, D. Real-space orthogonal projector-augmented-wave method. *Physical Review B* **102,** 195118 (2020).

89. Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter* **21,** 395502 (2009).

90. Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of computational chemistry* **29,** 2044–2078 (2008).

91. Pask, J. E., Klein, B. M., Sterne, P. A. & Fong, C. Y. Finite-element methods in electronic-structure theory. *Computer Physics Communications* **135,** 1–34 (2001).

92. Chelikowsky, J. R., Troullier, N & Saad, Y. Finite-difference-pseudopotential method: Electronic structure calculations without a basis. *Physical Review Letters* **72,** 1240 (1994).

93. Wang, L.-W. *A brief comparison between grid based real space algorithms and spectrum algorithms for electronic structure calculations.* tech. rep. (Lawrence Berkeley National Laboratory., Dec. 2006). `https://escholarship.org/uc/item/9bg6t4tg`.

94. Modine, N., Zumbach, G. & Kaxiras, E. Adaptive-coordinate real-space electronic-structure calculations for atoms, molecules, and solids. *Physical Review B* **55,** 10289 (1997).

95. Gygi, F. & Galli, G. Real-space adaptive-coordinate electronic-structure calculations. *Physical Review B* **52,** R2229 (1995).

96. Kronik, L., Makmal, A., Tiago, M. L., Alemany, M., Jain, M., Huang, X., Saad, Y. & Chelikowsky, J. R. PARSEC–the pseudopotential algorithm for real-space electronic structure calculations: recent advances and novel applications to nano-structures. *Physica Status Solidi (B)* **243,** 1063–1079 (2006).

97. Andrade, X., Strubbe, D., De Giovannini, U., Larsen, A. H., Oliveira, M. J., Alberdi-Rodriguez, J., Varas, A., Theophilou, I., Helbig, N., Verstraete, M. J., *et al.* Real-space grids and the Octopus code as tools for the development of new simulation approaches for electronic systems. *Physical Chemistry Chemical Physics* **17,** 31371–31396 (2015).

98. Yabana, K. & Bertsch, G. Time-dependent local-density approximation in real time. *Physical Review B* **54,** 4484 (1996).

99. Shepard, C., Zhou, R., Yost, D. C., Yao, Y. & Kanai, Y. Simulating electronic excitation and dynamics with real-time propagation approach to TDDFT within plane-wave pseudopotential formulation. *The Journal of Chemical Physics* **155** (2021).

100. Bertsch, G. F., Iwata, J.-I., Rubio, A. & Yabana, K. Real-space, real-time method for the dielectric function. *Physical Review B* **62,** 7998 (2000).

101. Yabana, K, Nakatsukasa, T, Iwata, J.-I. & Bertsch, G. Real-time, real-space implementation of the linear response time-dependent density-functional theory. *Physica Status Solidi (B)* **243,** 1121–1138 (2006).

102. Hamann, D., Schlüter, M & Chiang, C. Norm-conserving pseudopotentials. *Physical Review Letters* **43,** 1494 (1979).

103. Sadoc, A., Body, M., Legein, C., Biswal, M., Fayon, F., Rocquefelte, X. & Boucher, F. NMR parameters in alkali, alkaline earth and rare earth fluorides from first principle calculations. *Physical Chemistry Chemical Physics* **13,** 18539–18550 (2011).

104. Ljungberg, M., Mortensen, J. J. & Pettersson, L. An implementation of core level spectroscopies in a real space projector augmented wave density functional theory code. *Journal of Electron Spectroscopy and Related Phenomena* **184,** 427–439 (2011).

105. Zhou, Y., Chelikowsky, J. R. & Saad, Y. Chebyshev-filtered subspace iteration method free of sparse diagonalization for solving the Kohn–Sham equation. *Journal of Computational Physics* **274,** 770–782 (2014).

106. Pauletti, C. F., Coccia, E. & Luppi, E. Role of exchange and correlation in high-harmonic generation spectra of H2, N2, and CO2: Real-time time-dependent electronic-structure approaches. *The Journal of Chemical Physics* **154** (2021).

107. Petrone, A., Lingerfelt, D. B., Rega, N. & Li, X. From charge-transfer to a charge-separated state: a perspective from the real-time TDDFT excitonic dynamics. *Physical Chemistry Chemical Physics* **16,** 24457–24465 (2014).

108. Walter, M., Häkkinen, H., Lehtovaara, L., Puska, M., Enkovaara, J., Rostgaard, C. & Mortensen, J. J. Time-dependent density-functional theory in the projector augmented-wave method. *The Journal of Chemical Physics* **128** (2008).

109. Enkovaara, J., Rostgaard, C., Mortensen, J. J., Chen, J., Dułak, M, Ferrighi, L., Gavnholt, J., Glinsvad, C., Haikola, V, Hansen, H., *et al.* Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of Physics: Condensed Matter* **22,** 253202 (2010).

110. Akimov, A. V. & Prezhdo, O. V. The PYXAID program for non-adiabatic molecular dynamics in condensed matter systems. *Journal of chemical theory and computation* **9,** 4959–4972 (2013).

111. Onida, G., Reining, L. & Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Reviews of Modern Physics* **74,** 601 (2002).

112. 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. *The Journal of chemical physics* **110,** 2810–2821 (1999).

113. Jollet, F., Torrent, M. & Holzwarth, N. Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format. *Computer Physics Communications* **185,** 1246–1254 (2014).

114. Xu, Y., Kaur, R., Wang, B., Minameyer, M. B., Gsanger, S., Meyer, B., Drewello, T., Guldi, D. M. & von Delius, M. Concave–convex π–π template approach enables the synthesis of

[10] cycloparaphenylene–fullerene [2] rotaxanes. *Journal of the American Chemical Society* **140,** 13413–13420 (2018).

115. Forster, A. & Visscher, L. Quasiparticle self-consistent GW-Bethe-Salpeter equation calculations for large chromophoric systems. *Journal of Chemical Theory and Computation* **18,** 6779–6793 (2022).

116. Rabani, E., Baer, R. & Neuhauser, D. Time-dependent stochastic Bethe-Salpeter approach. *Physical Review B* **91,** 235302 (2015).

117. Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L. & Towns, J. *ACCESS: Advancing Innovation: NSF's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support* in *Practice and Experience in Advanced Research Computing* (Association for Computing Machinery, Portland, OR, USA, 2023), 173–176. ISBN: 9781450399852. https://doi.org/10.1145/3569951.3597559.