Breaking the theoretical scaling limit for predicting quasi-particle energies: The stochastic GW approach

Daniel Neuhauser, Yi Gao, Christopher Arntsen, and Cyrus Karshenas
Department of Chemistry, University of California at Los Angeles, CA-90095 USA

Eran Rabani
School of Chemistry, The Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Roi Baer
Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

We develop a formalism to calculate the quasi-particle energy within the GW many-body perturbation correction to the density functional theory (DFT). The occupied and virtual orbitals of the Kohn-Sham (KS) Hamiltonian are replaced by stochastic orbitals used to evaluate the Green function, the polarization potential, and thereby the GW self-energy. The stochastic GW (sGW) relies on novel theoretical concepts such as stochastic time-dependent Hartree propagation, stochastic matrix compression and spatial/temporal stochastic decoupling techniques. Beyond the theoretical interest, the formalism enables linear scaling GW calculations breaking the theoretical scaling limit for GW as well as circumventing the need for energy cutoff approximations. We illustrate the method for silicon nanocrystals of varying sizes with $N \sim 3000$ electrons.

The GW approximation [1, 2] to many-body perturbation theory (MBPT) [3] offers a reliable and accessible theory for quasi-particles (QPs) and their energies [4, 5], enabling estimation of electronic excitations [19, 25] quantum conductance [26, 30] and level alignment in hybrid systems [31, 32]. Practical use of GW for large systems is severely limited because of the steep CPU and memory requirements as system size increases. The most computationally intensive element in the GW method, the calculation of the polarization potential (screen Coulomb interaction), involves an algorithmic complexity that scales as the fourth power of the system size [33, 34]. Various approaches have been developed to reduce the computational bottlenecks of the GW approach [8, 33–36]. Despite these advances, GW calculations are still quite expensive for many of the intended applications in the fields of materials science, surface science and nanoscience.

In this letter we develop a stochastic, orbital-less, formalism for the GW theory, unique in that it does not refer to quasi-particles and orbital energies of the KS Hamiltonian. While the approach is inspired by recent developments in electronic structure theory using stochastic orbitals [37–41] it introduces three powerful and basic notions: Stochastic decoupling, stochastic matrix compression and stochastic time-dependent Hartree (sTDH) propagation. The result is a stochastic formulation of GW, where the QP energies become random variables sampled from a distribution with a mean equal to the exact GW energies and a statistical error proportional to the inverse number of stochastic orbitals (iterations, $I_{sGW}$).

We illustrate the sGW formalism for silicon nanocrystals (NCs) with varying sizes and band gaps [42, 43] and demonstrate that the CPU time and memory required by sGW scales nearly linearly with system size, thereby providing means to study QPs excitations in large systems of experimental and technological interest.

In the reformulation of the GW approach, we treat the QP energy ($\varepsilon_{QP} = \hbar \omega_{QP}$) as a variable rather than a solution to an eigenvalue problem, and rewrite the QP perturbative correction [2, 5] to the KS energy as:

$$\varepsilon_{QP}(\varepsilon) = \varepsilon + \tilde{\Sigma}^P(\omega_{QP}; \varepsilon) + \Sigma^X(\varepsilon) - \Sigma^{XC}(\varepsilon),$$

where $\varepsilon$ replaces the KS orbital energy (also treated as a variable) and can be determined from the desired value (highest occupied or lower virtual) of the density of states of the KS Hamiltonian available from the sDFT [10]. The frequency domain polarization self-energy $\tilde{\Sigma}^P(\omega, \varepsilon)$ is given in terms of the Fourier transform of the time domain counterpart $\Sigma^P(t, \varepsilon)$ and:

$$\Sigma^P(t; \varepsilon) = \frac{1}{q(\varepsilon)} \text{tr} \left[ f_{\sigma} \left( \hat{h}_{KS} - \varepsilon \right)^2 \tilde{\Sigma}^P(t; \varepsilon) \right],$$

$$\Sigma^X(\varepsilon) = \frac{1}{q(\varepsilon)} \text{tr} \left[ f_{\sigma} \left( \hat{h}_{KS} - \varepsilon \right)^2 \tilde{\Sigma}^X \right],$$

$$\Sigma^{XC}(\varepsilon) = \frac{1}{q(\varepsilon)} \text{tr} \left[ f_{\sigma} \left( \hat{h}_{KS} - \varepsilon \right)^2 \tilde{\Sigma}^{XC} \right],$$

$$Q(\varepsilon) = \text{tr} \left[ f_{\sigma} \left( \hat{h}_{KS} - \varepsilon \right)^2 \right].$$

Here, $\Sigma^X(\varepsilon)$ and $\Sigma^{XC}(\varepsilon)$ are the exchange and exchange-correlation self-energies, respectively, and $Q(\varepsilon)$ is a normalization factor. All three quantities can be calculated using a linear-scaling stochastic approach, as detailed in the supplementary information. In the above, $\tilde{\Sigma}^{XC}(\varepsilon)$ is the exchange-correlation potential of the KS-DFT Hamiltonian $\hat{h}_{KS}$ and $f_{\sigma}(\varepsilon) = e^{-\varepsilon^2/2\sigma^2}$ is an energy filter func-
tion of width $\sigma$ \cite{44}. In the GW approximation, the matrix elements of the polarization self-energy

$$\Sigma^P (r_1, r_2, t; \varepsilon) = \left\langle r_1 \left| \hat{\Sigma}^P (t; \varepsilon) \right| r_2 \right\rangle = i \hbar \tilde{G}_0 (r_1, r_2, t) \ W^P (r_1, r_2, t; \varepsilon), \quad (3)$$

are expressed in terms of the Green function, $G_0 (r_1, r_2, t)$

$$i \hbar G_0 (r_1, r_2, t) \equiv \left\langle r_1 \left| e^{-i \hbar K r_1 / \hbar} \hat{P}_\mu (t) \right| r_2 \right\rangle, \quad (4)$$

and the polarization potential

$$W^P (r_1, r_2, t; \varepsilon) \equiv \left\langle r_1 \left| u_C \otimes \chi (t; \varepsilon) \otimes u_C \right| r_2 \right\rangle. \quad (5)$$

In the above equations, $\hat{P}_\mu (t) \equiv \left( \theta (t) - \theta_\beta (\mu - \hat{h}_{KS}) \right)$, $\theta (E)$ and $\theta_\beta (E) = \frac{1}{2} (1 + \text{erf} (\beta E))$ are the Heaviside and a smoothed-Heaviside functions, respectively, $\mu$ is the chemical potential, $u_C (|r_1 - r_2|) = e^{2t}/4\pi \varepsilon_0 |r_1 - r_2|$ is the bare Coulomb potential, and $\chi (r_1, r_2, t; \varepsilon)$ is the time-ordered density-density correlation function \cite{34}. The symbol $\otimes$ represents a space convolution.

For performing the trace operations in Eqs. (1)-(3), we use a set of real stochastic orbitals $\phi (r)$ \cite{45-47} for which $1 = \left\langle \phi \left| \phi \right\rangle \right\rangle_\phi$ where $\left\langle \cdots \right\rangle_\phi$ denotes a statistical average over $\phi$. The choice of $\phi (r)$ satisfying these requirements is not unique. The form used here assigns a value of $\pm \hbar^{-3/2}$ at each grid point with equal probability, where $\hbar$ is the grid spacing. This allows us to rewrite the self-energy in Eq. (2) as:

$$\Sigma^P (t; \varepsilon) = \left\langle \int \phi_\varepsilon (r_1) \Sigma^P (r_1, r_2, t; \varepsilon) \phi (r_2) \ d^3 r_1 d^3 r_2 \right\rangle_\phi \quad (6)$$

where $|\phi_\varepsilon \rangle = f_\varepsilon (\hat{h}_{KS} - \varepsilon) |\phi \rangle$ is the corresponding filtered state at energy $\varepsilon$, which can be obtained by a Chebyshev expansion of the Gaussian function with $\sigma$ chosen as a small parameter \cite{48,49}. We note in passing that it is possible to obtain simultaneously $\Sigma^P (t; \varepsilon)$ for several values of $\varepsilon$ (more details are given in the supplementary information).

To obtain $\Sigma^P (r_1, r_2, t; \varepsilon)$ in Eq. (6) we need to calculate the non-interacting Green function $i \hbar G_0 (r_1, r_2, t)$ in Eq. (4) and the polarization potential $W^P (r_1, r_2, t; \varepsilon)$ in Eq. (5). For the former, we introduce an additional set of real stochastic orbitals, $\zeta (r)$, and describe $a$:

$$i \hbar G_0 (r_1, r_2, t) = \left\langle \zeta_\mu (r_1, t) \zeta (r_2) \right\rangle_\zeta, \quad (7)$$

where $\zeta_\mu (r, t) = \left\langle r \left| e^{-i \hbar K r / \hbar} \hat{P}_\mu (t) \right| \zeta \right\rangle$ is a “propagated-projected” stochastic orbital which can be obtained by a Chebyshev expansion of the function $e^{-i \varepsilon t/\hbar} (\theta (t) - \theta_\beta (\varepsilon - \mu))$ \cite{48,49}. One appealing advantage of the stochastic form of Eq. (7) is that it provides a compact representation for $G_0 (r_1, r_2, t)$, equivalent to matrix compression where $r_1$ and $r_2$ are decoupled. This allows a drastic simplification of the representation of the polarization self-energy obtained by combining Eqs. (6) and (7):

$$\Sigma^P (t; \varepsilon) = \left\langle \left\langle \phi_\varepsilon \zeta_\mu (t) \right| u_C \otimes \chi (t) \otimes u_C | \zeta \rangle \right\rangle_\phi \quad (8)$$

Next, we employ a temporal decoupling scheme achieved by introducing an additional set of real stochastic orbitals $\psi (r)$:

$$\Sigma^P (t; \varepsilon) = \left\langle \left\langle \phi_\varepsilon \zeta_\mu (t) \right| \psi \right\rangle \left\langle \psi \left| u_C \otimes \chi (t) \otimes u_C | \zeta \rangle \right\rangle_{\phi \psi} \quad (9)$$

which allows us to treat the term $\left\langle \phi_\varepsilon \zeta_\mu (t) \right| \psi \rangle$ separately from the term $\left\langle \psi \left| u_C \otimes \chi (t) \otimes u_C | \zeta \rangle \right\rangle_\phi$. Note that the average $\left\langle \cdots \right\rangle_{\phi \psi}$ in Eq. (9) is performed over $L_{\text{GW}}$ pairs of $\phi$ and $\zeta$ stochastic orbitals, and for each such pair we use a different set of $N_{\phi \psi}$ stochastic $\psi$'s. The term $\left\langle \phi_\varepsilon \zeta_\mu (t) \right| \psi \rangle$ is straightforward to obtain while $\left\langle \psi \left| u_C \otimes \chi (t) \otimes u_C | \zeta \rangle \right\rangle_\phi$ is determined from the time-retarded polarization potential, $\left\langle \psi \right| u_C \otimes \chi^* (t) \otimes u_C | \zeta \rangle \phi$, calculated from the linear response relation:

$$\left\langle \psi \right| u_C \otimes \chi (t) \otimes u_C | \zeta \rangle \phi = \left\langle \psi \right| u_C \delta n (t) \rangle,$$
The sDFT method was used to generate the Kohn–Sham Hamiltonian within the local density approximation (LDA). The calculations employed a real-space grid of spacings \( h = 0.6 \alpha_0 \), the Troullier-Martins norm-conserving pseudopotentials and fast Fourier transforms for implementing the kinetic and Hartree energies. The CPU time needed to converge the sDFT to a statistical error in the total energy per electron of about 10 meV was \( \approx 5000 \) hrs for the entire range of systems studied.

Next we performed a set of sGW calculations for a series of hydrogen passivated silicon NCs as detailed in Table I. The number of electrons \( (N_e) \), size of grid \( (N_g) \), number of sDFT iterations \( (I_{sDFT}) \), number of stochastic orbitals in sTD-DFT \( (N_e) \), the value of \( \beta \) in the sGW calculation, and the resulting QP energy gap \( (E_{gap}) \). The upper panel of Fig. 2 shows the scaling of the entire sGW approach for the combined calculation of \( \Sigma^X (\varepsilon) \), \( \Sigma^NC (\varepsilon) \), and \( \Sigma^P (t; \varepsilon) \). The scaling of the approach is nearly linear with the number of electrons, breaking the quadratic theoretical limit. It is important to note that for almost the entire range of NC sizes the sGW calculations were cheaper than the sDFT.

In conclusion, we have reformulated the GW approach for the combined calculation of \( \Sigma^X (\varepsilon) \), \( \Sigma^NC (\varepsilon) \), and \( \Sigma^P (t; \varepsilon) \). The scaling of the approach is nearly linear with the number of electrons, breaking the quadratic theoretical limit. It is important to note that for almost the entire range of NC sizes the sGW calculations were cheaper than the sDFT.

In conclusion, we have reformulated the GW approach for the combined calculation of \( \Sigma^X (\varepsilon) \), \( \Sigma^NC (\varepsilon) \), and \( \Sigma^P (t; \varepsilon) \). The scaling of the approach is nearly linear with the number of electrons, breaking the quadratic theoretical limit. It is important to note that for almost the entire range of NC sizes the sGW calculations were cheaper than the sDFT.
mation to many-body perturbation theory for calculating QP energies as a stochastic process without referring to the single-particle orbitals (or equivalently to the single-particle density matrix) and the corresponding orbital energies. The paradigm shift presented here by the fully quantum sGW formalism also removes the barrier for addressing larger systems, close to the mesoscopic limit. Indeed, the application to silicon NCs with system sizes far exceeding the current state-of-the-art indicates that the formalism scales nearly linearly with system size, breaking the theoretical limit. Some of the concepts presented here may likely be applicable to other forms of many-body perturbation theory approximations, such as propagator [55] and Green’s function theories [56].

The sGW developed here has several appealing advantages:

- **Representation:** It is especially suitable for real-space-grid/Plane-waves pseudopotential representations for which the Hamiltonian operation on a stochastic orbital scales linearly. These representations are the natural for large-scale electronic structure computations. The approach will also be useful for periodic systems with very large super-cells.

- **Matrix compression:** The introduction of the stochastic orbitals circumvents the need to store huge matrices of the Green function and the polarization potential (or the dielectric matrix $\epsilon$) and thus achieves huge savings in memory. The scaling of storage is $O(N_g)$, which makes the sGW approach applicable to large system without referring to cutoff approximations in the unoccupied space [13] [57] [58].

- **Computational scaling:** The sTDH developed here to obtain the polarization potential scales linearly with system size rendering the entire sGW a linear scaling approach. We find that the scaling of our approach is insensitive to the sparsity of the density matrix and thus represents a significant improvement over existing GW approaches.

- **Parallelization:** The stochastic character of the approach allows for straightforward parallelization of the sGW calculation: The self-energies are averaged over different stochastic orbitals so each processor performs its own independent contribution to this average.

These features make the sGW a method of choice for studying QP excitations in large complex materials not accessible by other approaches.

R. B. and E. R. gratefully thank the Israel Science Foundation, Grants No. 1020/10 and No. 611/11, respectively. R. B. and D. N. acknowledge the support of the US-Israel Bi-National Science Foundation. D. N., Y. G., C. A. and C. K. are part of the Molecularly Engineered Energy Materials (MEEM), an Energy Frontier Research Center funded by the DOE, Office of Science, Office of Basic Energy Sciences under Award No. de-sc0001342. Some of the simulations used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

---

[1] L. Hedin, Phys. Rev. 139, A796 (1965)
[2] C. Friedrich and A. Schindlmayr, NIC Series 31, 335 (2006)
[3] A. L. Fetter and J. D. Walecka, Quantum Thoery of Many Particle Systems (McGraw-Hill, New York, 1971) p. 299
[4] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. 55, 1418 (1985)
[5] M. S. Hybertsen and S. G. Louie, Phys. Rev. B 34, 5390 (1986)
[6] W. G. Aulbur, M. Stadale, and A. Gorling, Phys. Rev. B 62, 7121 (2000)
[7] A. Stan, N. E. Dahlen, and R. van Leeuwen, EPL (Europhysics Letters) 76, 298 (2006), [http://iopscience.iop.org/0295-5075/76/2/298](http://iopscience.iop.org/0295-5075/76/2/298)
[8] M. Shishkin and G. Kresse, Phys. Rev. B 75, 235102 (2007), [http://prb.aps.org/abstract/PRB/v75/i23/e235102](http://prb.aps.org/abstract/PRB/v75/i23/e235102)
[9] P. Huang and E. A. Carter, Annu. Rev. Phys. Chem. 59, 261 (2008), [http://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.59.032607.093528](http://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.59.032607.093528)
[10] P. E. Trevisanutto, C. Giorgetti, L. Reining, M. Ladisa, and V. Olevano, Phys. Rev. Lett. 101, 226405 (2008), [http://prl.aps.org/abstract/PRL/v101/i22/e226405](http://prl.aps.org/abstract/PRL/v101/i22/e226405)
[11] C. Rostaig, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B 81, 085103 (2010), [http://prb.aps.org/abstract/PRB/v81/i8/e085103](http://prb.aps.org/abstract/PRB/v81/i8/e085103)
[12] P. Liao and E. A. Carter, Phys. Chem. Chem. Phys. 13, 15189 (2011), [http://pubs.rsc.org/en/content/articlehtml/2011/cp/c1cp20829b](http://pubs.rsc.org/en/content/articlehtml/2011/cp/c1cp20829b)
[13] J. Berger, L. Reining, and F. Sottile, Phys. Rev. B 85, 085126 (2012), [http://prb.aps.org/abstract/PRB/v85/i8/e085126](http://prb.aps.org/abstract/PRB/v85/i8/e085126)
[14] M. van Setten, F. Weigend, and F. Evers, J. Chem. Theory Comput. 9, 232 (2012), [http://pubs.acs.org/doi/abs/10.1021/ct300648b](http://pubs.acs.org/doi/abs/10.1021/ct300648b)
[15] N. Marom, F. Caruso, X. Ren, O. T. Hofmann, T. Kürzdörfer, J. R. Chelikowsky, A. Rubio, M. Scheffler, and P. Rinke, Phys. Rev. B 86, 245127 (2012), [http://prb.aps.org/abstract/PRB/v86/085103](http://prb.aps.org/abstract/PRB/v86/085103)
[16] L. Y. Isseroff and E. A. Carter, Phys. Rev. B 85, 235142 (2012), [http://prb.aps.org/abstract/PRB/v85/i23/e235142](http://prb.aps.org/abstract/PRB/v85/i23/e235142)
[17] S. Refaely-Abramson, S. Sharifzadeh, N. Govind, J. Autschbach, J. B. Neaton, R. Baer, and L. Kronik, Phys. Rev. Lett. 109, 226405 (2012)
[18] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B 88, 075105 (Aug 2013), [http://link.aps.org/doi/10.1103/PhysRevB.88.075105](http://link.aps.org/doi/10.1103/PhysRevB.88.075105)
[19] J. C. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. 86, 472 (2001)
[20] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601 (2002)
[21] M. L. Tiago and J. R. Chelikowsky, Phys. Rev. B 73, 205334 (2006), [http://prb.aps.org/abstract/PRB/v73/i20/e205334](http://prb.aps.org/abstract/PRB/v73/i20/e205334)
