We identify the deviation from the straight line error (DSLE) – i.e., the spurious non-linearity of the total energy as a function of fractional particle number – as the main source for the discrepancy between experimental vertical ionization energies and theoretical quasiparticle energies, as obtained from the GW and GW+SOSEX approximations to many-body perturbation theory (MBPT). For self-consistent calculations, we show that GW suffers from a small DSLE. Conversely, for perturbative $G_0W_0$ and $G_0W_0+$SOSEX calculations the DSLE depends on the starting point. We exploit this starting-point dependence to reduce (or completely eliminate) the DSLE. We find that the agreement with experiment increases as the DSLE reduces. DSLE-minimized schemes, thus, emerge as promising avenues for future developments in MBPT.

Electronic structure theory has developed into an essential tool in material science, because it offers a parameter free, quantum mechanical description of solids, molecules, and nano-structures. This success is due to the continuous development of electronic structure methods such as density-functional theory (DFT) \cite{Hohenberg1964} and many-body perturbation theory (MBPT) in the GW approximation \cite{Hedin1965}. This development is guided in part by exact constraints, imposed by the theoretical framework itself. We will demonstrate in this work that invaluable insight into the GW approach can be gained from such an exact constraint. The GW method has long been heralded as the method of choice for band gaps and band structures of solids and quasiparticle spectra of molecules and nano-structures \cite{Zhang1991, Caruso2010, Dauth2015}. Yet, its accuracy is not always satisfying and the starting-point dependence in the perturbative $G_0W_0$ variant can be very pronounced \cite{Zhang1991, Caruso2010, Dauth2015}. Here we will show that the accuracy of GW is closely related to the exact constraint of piecewise linearity of the total energy \cite{Zhang1991}. Its violation gives rise to the deviation from the straight line error (DSLE), which has been extensively studied in DFT \cite{Hedin1965}. The starting-point dependence in $G_0W_0$ can then be exploited to minimize the DSLE, which uniquely defines the optimal starting point.

In 1982 Perdew et al. showed that the total energy of a quantum mechanical system has to change linearly with respect to the fractional removal (or addition) of an electron \cite{Perdew1982}

$$E(f) = (1-f)E(N_0 - 1) + fE(N_0). \quad (1)$$

Here, $N_0$ is the number of electrons in the neutral system and $E(N_0)$ the associated total energy. $E(N_0 - 1)$ is the total energy of the singly ionized system and $f$ varies in the interval $[0, 1]$. This piecewise linearity condition was initially derived in the context of DFT, but applies to any total energy method.

In the following, we introduce a formal definition of the deviation from linearity at fractional occupation numbers that employs only quantities directly accessible by quasiparticle energy calculations, that is, the ionization potential (IP) and the electron affinity (EA). The linearity condition in Eq. (1) implies that the first derivative of the total energy with respect to the fractional occupation number $f$ (i.e., $\partial E/\partial f$) should be constant and that it exhibits discontinuities at integer occupations ($f = 0$ and $f = 1$). Additionally, $\partial E/\partial f$ equals the electron removal energy $E(N_0) - E(N_0 - 1)$ or, equivalently, the energy for adding an electron to the positively charged system. Therefore, the vertical IP of the neutral system with $N_0$ electrons can be expressed as:

$$\left.\frac{\partial E(f)}{\partial f}\right|_{1} = E(N_0) - E(N_0 - 1) = -\text{IP}(N_0). \quad (2)$$

A similar relation holds for the electron affinity of the
cations (EA$_c$):

$$\frac{\partial E(f)}{\partial f} \bigg|_0 = E(N_0) - E(N_0 - 1) = -\text{EA}_c(N_0 - 1). \quad (3)$$

Equations 2 and 3 illustrate that IP = EA$_c$ in an exact theory. We thus define the difference between IP and EA$_c$ as the DSLE

$$\Delta_{\text{DSLE}} = \text{EA}_c(N_0 - 1) - \text{IP}(N_0). \quad (4)$$

$\Delta_{\text{DSLE}} = 0$ is a necessary condition for piecewise linearity. However, in an approximate treatment of electronic exchange and correlation, as e.g. in GW, IP and EA$_c$ may differ. A non vanishing $\Delta_{\text{DSLE}}$ indicates a curvature in the total energy versus fractional electron number curve, as illustrated in Fig. 1, causing an erroneous deviation from the straight line.

The DSLE is most easily seen in the deviation from the straight line

$$\Delta E(f) = E(f) - E_{\text{lin}}(f), \quad (5)$$

where, following Eq. 1, $E_{\text{lin}}(f)$ is the straight line between $E(N_0)$ and $E(N_0 - 1)$. We will first examine the DSLE and Eq. 5 for different DFT functionals before proceeding to our GW analysis. In DFT, common (semi-)local functionals typically exhibit a convex curvature and suffer from a large DSLE, whereas Hartree-Fock (HF) is concave with a moderate DSLE.

To quantify the DSLE by means of Eq. 4 we use the eigenvalue of the highest molecular orbital (HOMO) of the neutral system ($\epsilon^H_{N_0}$) for IP($N_0$) and the eigenvalue of the lowest unoccupied molecular orbital (LUMO) of the cation ($\epsilon^{\text{LUMO}}_{N_0-1}$) for EA$_c(N_0 - 1)$.

For O$_2$, PBE gives $\Delta_{\text{DSLE}} = \epsilon^H_{N_0} - \epsilon^{\text{LUMO}}_{N_0-1} = 11.6$ eV, which agrees with the pronounced convexity observed in Fig. 2. For benzene, the DSLE in PBE reduces to half the size ($\Delta_{\text{DSLE}} = 5.9$ eV), which is also apparent from the maximal extent of $\Delta E(f)$ in Fig. 2. Conversely, HF exhibits a concave DSLE manifested in $\Delta_{\text{DSLE}} = -3.5$ eV for O$_2$ and $\Delta_{\text{DSLE}} = -2.5$ eV for benzene. All in all, the magnitude of $\Delta_{\text{DSLE}}$ can be taken as a measure for the severity of the DSLE, whereas the sign indicates the curvature. A positive value of $\Delta_{\text{DSLE}}$ corresponds to a concave and a negative sign to a convex curvature. Convexity gives rise to a delocalization of the electron density and concavity to an overlocalization.

We now move on to discuss the DSLE in the GW approximation. In GW, the total energy is only available at integer occupation numbers, because an ensemble generalization to fractional particle numbers is not straightforward. However, we can still estimate the basic shape of the total energy curve by invoking only observables at integer particle numbers. If, in addition to the quasiparticle energies, we also have access to the total energy at integer particle numbers, we can calculate the slopes of $\Delta E$

$$\partial \Delta E/\partial f = \partial E/\partial f - \partial E_{\text{lin}}/\partial f. \quad (6)$$

For the reference straight line we use $\partial E_{\text{lin}}/\partial f = E(N_0 - 1) - E(N_0)$, and $\partial E/\partial f$ gives us the GW quasiparticle energies. Equation 6 then becomes

$$\partial \Delta E/\partial f = \begin{cases} \epsilon^H_{N_0} + E(N_0 - 1) - E(N_0) \quad \text{for } N_0 \\ \epsilon^{\text{LUMO}}_{N_0-1} + E(N_0 - 1) - E(N_0) \quad \text{for } N_0 - 1, \end{cases} \quad (7a)$$

where we calculate the total energies $E(N_0 - 1)$ and $E(N_0)$ with fully self-consistent GW (scGW) [21, 22]. The resulting slopes are shown as arrows in Fig. 2.

Among the different GW flavors only scGW, in which the Dyson equation is solved iteratively, gives results that are independent of the starting point [21, 22]. Therefore, scGW provides an unbiased assessment of the DSLE. Most importantly, in scGW also the ground state is treated at the GW level and the scGW density does not inherit the (de-)localization error of the starting point, as is the case in $G_0W_0$ calculations.

Figure 2 illustrates that also scGW violates the straight line condition due to the approximate nature of the self-energy. Evaluating Eq. 4 with the scGW quasiparticle HOMO and LUMO energies, we obtain $\Delta_{\text{DSLE}} = 0.9$ eV for O$_2$. This indicates convexity, which is, however, much less pronounced than in PBE. The slopes at the two end-points of the scGW curve confirm the convex behavior, because they are pointing in different directions, i.e., they have different signs. For our second example, benzene, the behavior is markedly different. The signs of $\partial \Delta E/\partial f$ are equal at both ends of
the occupation interval. To connect both endpoints — schematically sketched by the blurred curve in Fig. 2 — \( \Delta E \) inevitably has to cross zero at some point. Hence, we expect the \( \Delta E \) curve to be divided into two regimes. Beginning from the cation, the positive slope at \( N_0 - 1 \) gives rise to a concave DSLE. When we approach the \( N_0 \) electron limit, the positive slope at \( N_0 \) requires a convex curvature. For benzene, the absolute value of the slope at \( N_0 - 1 \) is higher than at the other end of the interval. As a result, we expect the concave deviation on the \( N_0 - 1 \) side to be more pronounced than the convex part. This is also reflected by a negative \( \Delta_{\text{DSLE}} = -0.7 \text{ eV} \), which we associate with a concave DSLE.

Our benzene example illustrates that we can in principle encounter systems that are not DSLE-free, although \( \Delta_{\text{DSLE}} = 0 \). In these cases \( \partial \Delta E / \partial f \) provides additional information on the total energy curve. If \( \partial \Delta E / \partial f = 0 \) at \( N_0 - 1 \) and \( N_0 \), we can ensure that the total energy follows exactly the straight line in the vicinity of \( N_0 - 1 \) and \( N_0 \). Hence, we expect the method to be DSLE-free. These two conditions are generally applicable to any electronic-structure method that provides access to total and quasiparticle energies.

To provide a comprehensive assessment of the DSLE in the \( GW \) approximation, we have further performed calculations for a benchmark set consisting of 48 atoms and molecules selected from the quantum chemical G2 ion test set [23, 24] — referred to as G2_{ip} in the following [26]. For sc\( GW \) we find a mean DSLE of \( \Delta_{\text{DSLE}} = -0.5 \text{ eV} \). The average over the absolute \( \Delta_{\text{DSLE}} \) amounts to 0.9 eV. The sc\( GW \) DSLE is thus much smaller than that of PBE and HF (Figs. 2 and 3). Our results provide quantitative evidence that sc\( GW \) predominantly exhibits a concave DSLE. More generally, the slopes at \( N_0 \) show a small deviation from zero, \( \partial \Delta E / \partial f = 0.14 \text{ eV} \). Conversely, the slope at \( N_0 - 1 \) is higher with a mean value of \( \partial \Delta E / \partial f = 0.61 \text{ eV} \). Out of the 48 systems of the G2_{ip} set, 28 of them exhibit simultaneously concave and convex curvatures, as in the benzene case. Overall, our results suggest that fully self-consistent \( GW \) has the tendency to over-localize electron density, which is consistent with previous work [10] on the quasiparticle self-consistent \( GW \) approach [27].

Most commonly, \( GW \) is not carried out fully self-consistently, but applied in first-order perturbation theory \( (G_0W_0) \). This introduces a dependence on the reference ground state encoded in \( G_0 \). Logically, also the DSLE should depend on the chosen starting point.

For \( O_2 \), \( G_0W_0 \) calculations based on orbitals and eigenvalues from PBE \( (G_0W_0@PBE) \) yield an IP of \(-11.6 \text{ eV} \), which differs from the \( A_c \) of the cation \((-13.0 \text{ eV}) \). Thus, \( G_0W_0@PBE \) also violates the straight line condition, as quantified through Eq. (4), which yields \( \Delta_{\text{DSLE}} = 1.4 \text{ eV} \). The positive value of \( \Delta_{\text{DSLE}} \) indicates a convex total energy at fractional particle numbers as in PBE, albeit an order of magnitude smaller than in PBE. Conversely, we find \( \Delta_{\text{DSLE}} = -0.6 \text{ eV} \) with opposite sign if we use \( G_0W_0 \) based on the PBE hybrid functional PBEh(\( \alpha \)) [28, 29] with \( \alpha = 1 \).

Now one could ask if our DSLE definition may prove useful for the design of novel DSLE-free approaches for quasiparticle energy calculations. Since in our two examples \( \Delta_{\text{DSLE}} \) changes sign for \( \alpha = 0 \) and \( \alpha = 1 \), it is conceivable to postulate that an intermediate, optimal \( \alpha \) exists for which the DSLE in \( G_0W_0 \) is eliminated or at least considerably reduced. To test this postulate, we evaluate the DSLE for several \( \alpha \) values in the range \([0, 1]\). As illustrated in the lower panel of Fig. 4, an increasing \( \alpha \) gradually decreases \( \Delta_{\text{DSLE}} \) for \( O_2 \) and benzene. At an optimal \( \alpha \) of \( \approx 0.4 \) for benzene and \( \approx 0.6 \) for \( O_2 \) the DSLE vanishes. Beyond this point, \( \Delta_{\text{DSLE}} \) becomes increasingly negative with a further increase in \( \alpha \), indicating an increasingly concave total energy curve.

In analogy to DFT and sc\( GW \), we can support the \( G_0W_0 \) \( \Delta_{\text{DSLE}} \) results by examining the slopes of \( \Delta E \) as defined in Eq. (5). In \( G_0W_0 \) the quasiparticle energies are equivalent to the derivative of the total energy in the random phase approximation (RPA) with respect to the particle number [30]. We show \( \partial \Delta E / \partial f \) for \( N_0 - 1 \) and \( N_0 \) as a function of \( \alpha \) for \( O_2 \) and benzene in the upper panel of Fig. 4. Both molecules show the trend expected from the \( \Delta_{\text{DSLE}} \) calculations. Beginning at small \( \alpha \), the slopes support convexity because \( \Delta E \) falls away from \( N_0 - 1 \) and rises again approaching \( N_0 \). Conversely,
Consistent with the empirical findings of Bruneval [33] and Marom et al. [17] and the work of Körzdörfer et al. [34].

Finally, we illustrate that the concept of DSLE-minimized quasiparticle calculations is generally applicable and can be transferred to other self-energy approximations. Motivated by the good results of renormalized second order perturbation theory (rPT2) for electron correlation energies [35], Ren and coworkers recently proposed a beyond-GW scheme [36] that combines GW with a second order screened exchange self-energy (G0W0+SOSEX). In the following, we apply our DSLE analysis to (G0W0+SOSEX)@PBEh(α) calculations for the G2ip test set. In the upper panel of Fig. 5 we display the corresponding DSLE and in the lower panel the IP MAE as a function of α. Compared to G0W0, the starting-point dependence is weaker and the DSLE is always negative. Also the DSLE and the MAE are minimized at smaller α values. This confirms our previous, heuristic findings, that rPT2 and G0W0+SOSEX perform best for starting points that are close to PBE. The smallest average deviation from the experimental IPs amounts to 0.21 eV, which is comparable to G0W0.

In conclusion, we have shown that the DSLE is a prominent source of discrepancy between experimental and theoretical vertical IPs. Through a formal definition of the DSLE for quasiparticle calculations, we show that the prominent GW approach has an intrinsic DSLE of 0.5 eV and a tendency towards concavity, i.e. localization of electrons. We then establish a correlation between the DSLE and the deviation from experimental ionization energies. This allowed us to propose a recipe for obtaining DSLE-minimized approximations to many-body perturbation theory. The DSLE-minimized G0W0 and G0W0+SOSEX schemes give the best agreement with experimental data, as illustrated for the 48 molecules of the G2ip test set.

We thank Xinguo Ren for supporting us with the SOSEX calculations and Christoph Friedrich for fruitful discussions. This work was supported by the Academy of Finland through its Centres of Excellence Programme under project numbers 251748 and 284621. S.K. and M.D. acknowledge support by Deutsche Forschungsgemeinschaft Graduiertenkolleg 1640 and the Bavarian State Ministry of Science, Research, and the Arts for the Collaborative Research Network Soltech.
[1] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
[2] L. Hedin, Phys. Rev. 139, A796 (1965).
[3] W. G. Aulbur, L. Jonsson, and J. W. Wilkins, Solid State Phys. 54, 1 (1999).
[4] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601 (2002).
[5] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. 7, 126 (2005).
[6] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, Phys. Rev. B 76, 115109 (2007).
[7] N. Marom, F. Caruso, X. Ren, O. T. Hofmann, T. Körzdörfer, J. R. Chelikowsky, A. Rubio, M. Scheffler, and P. Rinke, Phys. Rev. B 86, 245127 (2012).
[8] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz Jr, Phys. Rev. Lett. 49, 1691 (1982).
[9] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Science 321, 792 (2008).
[10] A. Ruzsinszky, J. P. Perdew, G. I. Csonka, O. A. Vydrov, and G. E. Scuseria, J. Chem. Phys. 125, 194112 (2006).
[11] P. Mori-Sánchez, A. J. Cohen, and W. Yang, J. Chem. Phys. 125, 201102 (2006).
[12] Y. Zhang and W. Yang, J. Chem. Phys. 109, 2604 (1998).
[13] E. Kraisler and L. Kronik, Phys. Rev. Lett. 110, 126403 (2013).
[14] V. Vlcek, H. R. Eisenberg, G. Steinle-Neumann, L. Kronik, and R. Baer, J. Chem. Phys. 142, 034107 (2015).
[15] V. Atalla, P. Rinke, I. Zhang, O. T. Hofmann, and M. Scheffler, in preparation (2015).
[16] F. Bruneval, Phys. Rev. Lett. 103, 176403 (2009).
[17] We have performed all calculations with the Fritz-Haber-Institut ab initio molecular simulations (FHI-aims) package 24 57. We used Tier 3 numeric atom-centered orbital basis functions 24 for self-consistent GW 21 22, otherwise the Tier 4 basis sets augmented by Gaussian aug-cc-pV5Z basis functions 24 38.
[18] C.-O. Almbladh and U. von Barth, Phys. Rev. B 31, 3231 (1985).
[19] J. F. Janak, Phys. Rev. B 18, 7165 (1978).
[20] W. Yang, A. J. Cohen, and P. Mori-Sánchez, J. Chem. Phys. 136, 204111 (2012).
[21] P. Caruso, P. Rinke, X. Ren, M. Scheffler, and A. Rubio, Phys. Rev. B 86, 081102 (2012).
[22] P. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B 88, 075105 (2013).
[23] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. 106, 1063 (1997).
[24] X. Ren, P. Rinke, V. Blum, J. Wienerink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, New J. Phys. 14, 053020 (2012).
[25] We excluded the free H and He-atom from benchmark set presented in Ref. [24] as the G0W0 quasiparticle LUMO energies of the cations reside in a completely empty spin channel.
[26] See Supplemental Material at [URL will be inserted by publisher] for the systems and the corresponding numerical values of the G2ip test set.
[27] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. 96, 226402 (2006).
[28] C. Adamo and V. Barone, J. Chem. Phys. 110, 6158 (1999).
[29] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
[30] L.-W. Wang, Phys. Rev. B 82, 115111 (2010).
[31] Reference vertical ionization energies are taken from http://cccbdb.nist.gov.
[32] At the optimal starting point for each system, we observe the same MAE for the G2ip set.
[33] F. Bruneval and M. A. L. Marques, J. Chem. Theory Comput. 9, 324 (2012).
[34] T. Körzdörfer and N. Marom, Phys. Rev. B 86, 041110(R) (2012).
[35] X. Ren, P. Rinke, G. E. Scuseria, and M. Scheffler, Phys. Rev. B 88, 035120 (2013).
[36] X. Ren, N. Marom, F. Caruso, M. Scheffler, and P. Rinke, Phys. Rev. B 92, 081104(R) (2015).
[37] V. Blum, R. Gehrke, F. Hanke, P. Hauke, V. Hauke, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. 180, 2175 (2009).
[38] T. H. Dunning, J. Chem. Phys. 90, 1007 (1989).