ARTICLE   OPEN

Check for updates

Accelerating GW calculations through machine-learned dielectric matrices

Mario G. Zauchner [ID][1], Andrew Horsfield [ID][1] and Johannes Lischner [ID][1][✉]

The GW approach produces highly accurate quasiparticle energies, but its application to large systems is computationally challenging due to the difficulty in computing the inverse dielectric matrix. To address this challenge, we develop a machine learning approach to efficiently predict density–density response functions (DDRF) in materials. An atomic decomposition of the DDRF is introduced, as well as the neighborhood density–matrix descriptor, both of which transform in the same way under rotations. The resulting DDRFs are then used to evaluate quasiparticle energies via the GW approach. To assess the accuracy of this method, we apply it to hydrogenated silicon clusters and find that it reliably reproduces HOMO–LUMO gaps and quasiparticle energy levels. The accuracy of the predictions deteriorates when the approach is applied to larger clusters than those in the training set. These advances pave the way for GW calculations of complex systems, such as disordered materials, liquids, interfaces, and nanoparticles.

npj Computational Materials (2023)9:184 ; https://doi.org/10.1038/s41524-023-01136-y

INTRODUCTION

Density functional theory (DFT)[1,2] has shown tremendous success in the calculation of electronic ground-state properties. However, it is well known that band gaps of solids and HOMO–LUMO gaps of molecules are often significantly underestimated when computed using Kohn–Sham (KS) eigenvalues[3,4]. In order to remedy this issue, the GW method[5–7] is often employed in which a self-energy correction to the DFT KS energies is computed. The resulting quasiparticle energies are in agreement with experimental measurements for a wide range of materials. However, the large numerical effort required for GW calculations and the method's unfavorable scaling with system size have traditionally restricted applications to relatively small systems[8,9]. The most expensive step is the computation of the interacting density–density response function (DDRF), which is closely related to the inverse dielectric matrix. In particular, the non-interacting DDRF is typically computed by carrying out a slowly converging summation over all unoccupied states[8,10,11]. Afterward, the non-interacting DDRF must be inverted to calculate the interacting DDRF.

To overcome these limitations of the GW approach, significant efforts have been made in recent years to develop scalable implementations[12–16]. Alternatively, model DDRFs (or model dielectric functions) have been developed to accelerate GW calculations. For example, Hybertsen and Louie constructed a model dielectric matrix based on the assumption that the local screening response of the material is similar to that of a homogeneous medium with the same local density[17]. A similar model was also proposed by Cappellini et al.[18,19]. However, it has proven difficult to generalize these model dielectric functions to highly non-uniform systems, such as isolated molecules or nanoclusters whose screening properties differ substantially from uniform systems. To overcome this limitation, Rohlfing[9] proposed to express the dielectric matrix as a sum of atomic contributions attributing a density response resulting from a Gaussian-shaped charge density to each atom. This model dielectric matrix contains a number of parameters that need to be determined, for example, by comparison to calculated RPA dielectric functions.

In recent years, machine learning (ML) techniques have been widely adopted to predict scalar properties of materials, such as the total energy. A key ingredient in ML approaches is the descriptor which parametrizes the atomic and chemical structure of the material. Many descriptors used in computational chemistry are explicitly constructed to be invariant under rotations and translations: for example, ACE[20], SOAP[21], the Coulomb matrix[22,23], bag-of-bonds[24] or fingerprint-based descriptors have been shown to be reliable descriptors for the prediction of scalar quantities. When predicting tensors or functions, however, it is no longer sufficient to employ a rotationally invariant descriptor. To alleviate this problem, Grisafi et al.[25] developed a symmetry-adapted version of the SOAP kernel which is equivariant under rotations and was successfully used in the prediction of polarizability tensors and first hyperpolarizabilities[25,26], dipole moments[27] and electronic densities[28]. Several other groups also explored ML approaches for the electronic density, including Brockherde et al.[29], Alred et al.[30], and Chandrasekaran and co-workers[31]. Moreover, the construction of group-equivariant neural networks (NNs), such as Clebsch–Gordan networks[32–34], tensor-field networks[35], and spherical convolutional NNs (CNNs)[36,37] have seen significant developments in recent years, and the implementation of these methods has been significantly simplified by frameworks such as e3NN[38] developed by Geiger et al.[39], thus providing promising alternatives to the symmetry-adapted SOAP for the learning of functions.

In this work, we address the problem of predicting non-local response functions, such as the DDRF. Predicting such quantities is a formidable challenge: for example, the DDRF of a small silicon cluster can be tens of gigabytes in size when represented on a plane-wave basis, even when a modest plane-wave cutoff is used. To address this problem, we introduce a decomposition of the DDRF into atomic contributions, which can be predicted using ML techniques. To ensure that the ML model appropriately incorporates the transformation properties of the DDRF, we also develop a descriptor called neighborhood density–matrix (NDM), which transforms in the same way as the DDRF under rotations and is

---

[1]Department of Materials, Thomas Young Centre, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. ✉email: j.lischner@imperial.ac.uk

npj

used in conjunction with a dense NN to predict the atomic contributions to the DDRF. We then use the ML DDRFs to carry out GW calculations of hydrogenated silicon clusters. This approach which we refer to as the ML–GW method, produces accurate GW quasiparticle energies at a significantly reduced computational cost compared to standard implementations. We note that recently several attempts were made to use ML to directly predict quasiparticle energies in materials[40–42]. In contrast, the ML–GW approach still solves a physical model (the quasiparticle equation) but uses ML DDRFs to accelerate calculations.

RESULTS

Theoretical results

The GW method yields accurate quasiparticle energies by applying a self-energy correction to the mean-field KS energy levels. The GW self-energy $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$ is calculated from the one-electron Green's function $G(\mathbf{r}, \mathbf{r}', \omega)$ and the screened Coulomb interaction $W(\mathbf{r}, \mathbf{r}', \omega)$ according to [7,8,43]

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int e^{-i\delta\omega'} G(\mathbf{r}, \mathbf{r}', \omega + \omega') W(\mathbf{r}, \mathbf{r}', \omega') d\omega' \quad (1)$$

with δ denoting a positive infinitesimal. The screened Coulomb interaction is, in turn, computed from the bare Coulomb interaction $v(\mathbf{r}, \mathbf{r}')$ and the inverse dielectric matrix $\epsilon^{-1}(\mathbf{r}, \mathbf{r}', \omega)$ via

$$W(\mathbf{r}, \mathbf{r}', \omega) = \int \epsilon^{-1}(\mathbf{r}, \mathbf{r}_2, \omega) v(\mathbf{r}_2, \mathbf{r}', \omega) d\mathbf{r}_2, \quad (2)$$

which demonstrates that the dielectric matrix constitutes a key ingredient in GW calculations. It can be obtained from the interacting DDRF $\chi(\mathbf{r}, \mathbf{r}', \omega)$ according to

$$\epsilon^{-1}(\mathbf{r}, \mathbf{r}', \omega) = \delta(\mathbf{r}, \mathbf{r}') + \int v(\mathbf{r}, \mathbf{r}_2) \chi(\mathbf{r}_2, \mathbf{r}', \omega) d\mathbf{r}_2. \quad (3)$$

In the remainder of this paper, we will assume that the frequency dependence of the dielectric matrix can be approximated by the generalized plamon-pole approximation (GPP)[7,44,45]. As a consequence, only the static DDRF $\chi(\mathbf{r}, \mathbf{r}') \equiv \chi(\mathbf{r}, \mathbf{r}', \omega = 0)$ needs to be determined.

Within the random-phase approximation (RPA), the interacting static DDRF is given by

$$\chi(\mathbf{r}, \mathbf{r}') = \chi_0(\mathbf{r}, \mathbf{r}') \\ + \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_0(\mathbf{r}, \mathbf{r}_1) v(\mathbf{r}_1, \mathbf{r}_2) \chi(\mathbf{r}_2, \mathbf{r}') \quad (4)$$

with $\chi_0(\mathbf{r}, \mathbf{r}')$ denoting the static non-interacting DDRF, which is typically computed as a sum over empty and occupied states[10,11] according to

$$\chi_0(\mathbf{r}, \mathbf{r}') = \sum_{ij} \frac{f_i(1-f_j)}{\epsilon_i - \epsilon_j} \\ \times \left[\phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) \phi_j^*(\mathbf{r}') \phi_i(\mathbf{r}') + \text{c.c.} \right]. \quad (5)$$

Here, ϵ_i, f_i, and $\phi_i(\mathbf{r})$ denote the orbital energy, occupancy, and wavefunctions of the KS state i.

Equations (4) and (5) highlight the two main challenges in computing the DDRF: (1) the calculation of the non-interacting DDRF requires a summation of all empty states, which is slowly converging, and (2) the calculation of the interacting DDRF requires a matrix inversion which scales unfavorably with system size.

In order to bypass the expensive computation of the DDRF and pave the way toward an ML approach, we propose to express $\chi(\mathbf{r}, \mathbf{r}')$ as a sum of atomic contributions $\chi_i(\mathbf{r}, \mathbf{r}')$ according to

$$\chi(\mathbf{r}, \mathbf{r}') = \sum_{i=1}^{N} \chi_i(\mathbf{r}, \mathbf{r}'), \quad (6)$$

where i labels atoms and N is the total number of atoms.

How this partitioning is achieved is not immediately obvious. However, the atomic contributions to the DDRF should have the following properties: (1) the atomic contributions should be localized in the vicinity of the corresponding atom, (2) they should retain the global symmetry of χ, i.e., $\chi_i(\mathbf{r}, \mathbf{r}') = \chi_i(\mathbf{r}', \mathbf{r})$, and (3) they should integrate to zero, i.e., $\int \chi_i(\mathbf{r}, \mathbf{r}') d\mathbf{r} = \int \chi_i(\mathbf{r}, \mathbf{r}') d\mathbf{r}' = 0$, to ensure that the change in the charge density induced by a perturbing potential is overall charge neutral[8].

We start by expressing the DDRF in a localized basis set of real orbitals $\{\phi_{\alpha_a}^a(\mathbf{r})\}$, where a labels the atom on which the basis function is centered and α_a indexes the orbital on site a[46]. In this basis the DDRF is given by

$$\chi(\mathbf{r}, \mathbf{r}') = \sum_{a, \alpha_a} \sum_{b, \alpha_b} \chi_{\alpha_a \alpha_b}^{ab} \phi_{\alpha_a}^a(\mathbf{r}) \phi_{\alpha_b}^b(\mathbf{r}'), \quad (7)$$

where $\chi_{\alpha_a \alpha_b}^{ab}$ is a symmetric matrix. This expression suggests the following decomposition of the DDRF into atomic contributions

$$\chi_i(\mathbf{r}, \mathbf{r}') = \frac{1}{2} \sum_{\alpha_i} \sum_{b, \alpha_b} \left(\chi_{\alpha_i \alpha_b}^{ib} \phi_{\alpha_i}^i(\mathbf{r}) \phi_{\alpha_b}^b(\mathbf{r}') + \chi_{\alpha_b \alpha_i}^{bi} \phi_{\alpha_b}^b(\mathbf{r}) \phi_{\alpha_i}^i(\mathbf{r}') \right). \quad (8)$$

We refer to the representation of the DDRF in the basis $\{\phi_{\alpha_a}^a(\mathbf{r})\}$ as 2-center DDRF (2C-DDRF) because it contains pairs of basis functions which are centered on different atoms.

Using the symmetry of $\chi_{\alpha_i \alpha_w}^{iw}$ and the fact that the basis functions are real, it can be easily verified that $\chi_i(\mathbf{r}, \mathbf{r}') = \chi_i(\mathbf{r}', \mathbf{r})$. We can also ensure that $\int \chi_i(\mathbf{r}, \mathbf{r}') d\mathbf{r} = 0$ by removing all s-orbitals from the basis: see the computational methods section for details. The locality of $\chi_i(\mathbf{r}, \mathbf{r}')$ is directly inherited from the corresponding properties of the full DDRF. In particular, we have found that the expansion coefficients $\chi_{\alpha_i \alpha_w}^{iw}$ decay rapidly as the distance between atom i and atom w increases[47].

We stress that this atomic representation of the DDRF is exact, i.e., $\sum_i \chi_i(\mathbf{r}, \mathbf{r}')$ reproduces the full interacting DDRF when the local basis sets are complete. However, the atomic contributions to the DDRF contain contributions from pairs of basis functions that are centered on different atoms, see Eq. (8). These contributions are difficult to learn using atom-centered descriptors.

To make progress, we exploit the localization of $\chi_i(\mathbf{r}, \mathbf{r}')$ and expand it in terms of a set of basis functions $\psi_{nlm}^i(\mathbf{r}) = Y_{lm}(\hat{\mathbf{r}}) R_n(|\mathbf{r}|)$ (with Y_{lm} denoting the spherical harmonics and R_n a set of radial functions), which are all centered on atom i according to

$$\chi_i(\mathbf{r}, \mathbf{r}') = \sum_{nlm} \sum_{n'l'm'} \chi_{nlmn'l'm'}^{(i)} Y_{lm}(\hat{\mathbf{r}}) Y_{l'm'}^*(\hat{\mathbf{r}}') R_n(|\mathbf{r}|) R_{n'}^*(|\mathbf{r}'|) \quad (9)$$

with $\chi_{nlmn'l'm'}^{(i)}$ denoting the expansion coefficients given by

$$\chi_{nlmn'l'm'}^{(i)} = \int \int d\mathbf{r} d\mathbf{r}' \chi_i(\mathbf{r}, \mathbf{r}') R_n^*(|\mathbf{r}|) R_{n'}(|\mathbf{r}'|) Y_{lm}^*(\hat{\mathbf{r}}) Y_{l'm'}(\hat{\mathbf{r}}'). \quad (10)$$

These coefficients can be learned using a NN based on atom-centered descriptors. We refer to the representation of the DDRF in the basis $\{\psi_{nlm}^i(\mathbf{r})\}$ as 1-center DDRF (1C-DDRF) because it only contains pairs of basis functions centered on the same atom.

As discussed in the introduction, it is not appropriate to use a scalar descriptor (such as the standard SOAP descriptor[48]) that is invariant under rotations to develop an ML model for the DDRF: the behavior of the atomic DDRFs under rotations is determined by their analytical form: see Eq. (9). In particular, we show in the Supplementary Discussion that the coefficients of the atomic DDRF transform according to

$$\tilde{\chi}_{nlm_1 n'l'm_2}^{(i)} = \sum_{m, m'} D_{m_1 m}^l(\hat{R}) D_{m_2 m'}^{l'*}(\hat{R}) \chi_{nlmn'l'm'}^{(i)}, \quad (11)$$

where $\tilde{\chi}_{nlmn'l'm'}^{(i)}$ denote the coefficients of the transformed DDRF, $\hat{R}$ is a rotation and $D_{mm'}^l(\hat{R})$ is a Wigner D-matrix[49].

Next, we construct the NDM descriptor, which transforms under rotations in the same way as the atomic DDRF. The starting point for such a descriptor is a non-local extension of the smooth

neighborhood density of atom i of species η employed in the SOAP descriptor[21], defined as

$$\rho_i^\eta(\mathbf{r}, \mathbf{r}') = \sum_{k \in \eta} \sum_{l \in \eta} e^{-a(\mathbf{r} - \mathbf{r}_k)^2} e^{-a(\mathbf{r}' - \mathbf{r}_l)^2}, \tag{12}$$

where k and l run over atoms in the neighborhood of atom i within a cut-off radius R_{cut} and a is a hyperparameter that describes the size of an atom. The NDM is then expanded in a basis of spherical harmonics and radial basis functions $R_n(|\mathbf{r}|)$ according to

$$\rho_i^\eta(\mathbf{r}, \mathbf{r}') = \sum_{nlm} \sum_{n'l'm'} \rho_{nlmn'l'm'}^{(i,\eta)} Y_{lm}(\hat{\mathbf{r}}) Y_{l'm'}^*(\hat{\mathbf{r}}') R_n(|\mathbf{r}|) R_{n'}^*(|\mathbf{r}'|), \tag{13}$$

with $\rho_{nlmn'l'm'}^{(i,\eta)}$ being expansion coefficients. The above equation shows that the NDM transforms in the same way as the atomic DDRF: see Supplementary information for additional details. Therefore, we use the expansion coefficients as a descriptor for learning the DDRF.

We note that the NDM can be written as the product of two neighborhood densities $\rho_i^\eta(\mathbf{r}) = \sum_{k \in \eta} \exp\{-a(\mathbf{r} - \mathbf{r}_k)^2\}$ according to

$$\rho_i^\eta(\mathbf{r}, \mathbf{r}') = \rho_i^\eta(\mathbf{r}) \rho_i^\eta(\mathbf{r}'). \tag{14}$$

Similar to the NDM, $\rho_i^\eta(\mathbf{r})$ can be expanded in a basis of spherical harmonics and radial basis functions $R_n(|\mathbf{r}|)$ with coefficients $\rho_{nlm}^{(i,\eta)}$. It follows that

$$\rho_{nlmn'l'm'}^{(i,\eta)} = \rho_{nlm}^{(i,\eta)} \rho_{n'l'm'}^{(i,\eta)}, \tag{15}$$

which demonstrates that the coefficients of the neighborhood density contain the same information as the coefficients of the neighborhood density matrix. Indeed, we have found in our calculations that both types of coefficients perform equally when used as descriptors to predict the atomic DDRFs. We further note that the coefficients of the 3-body version of the SOAP descriptor $d_{nn'l}^{(\eta)}$ can be obtained from the NDM using

$$d_{nn'l}^{(\eta)} = \sum_{l'mm'} \sqrt{\frac{8\pi^2}{2l+1}} \rho_{nlm}^{(i,\eta)} \rho_{n'l'm'}^{(i,\eta)} \delta_{ll'} \delta_{mm'}, \tag{16}$$

in the case where there is no coupling between different atomic species η.

Machine learning

We apply our ML approach for predicting DDRFs to hydrogenated silicon clusters and then use the DDRFs to calculate GW quasiparticle energies for these systems. We refer to this technique as the ML–GW approach. The atomic positions of the clusters were constructed as described in the methods section and then relaxed using DFT.

To establish the accuracy of this approach, we first investigate the error in the GW quasiparticle energies resulting from the expansion of the DDRF in terms of the intermediate local basis $\{\phi_{a_q}^q(\mathbf{r})\}$: see Eq. (7). Figure 1 compares the HOMO–LUMO gaps obtained from mean-field DFT–PBE calculations, a standard plane-wave G_0W_0 calculation using a generalized plasmon-pole approximation (GPP)[7,45] and a G_0W_0 calculation using the 2C-DDRF, where the DDRF is expanded in terms of a modified version of the admm-2 basis set[50]: see "Methods" section. The DFT–PBE results show that the HOMO–LUMO gap decreases with increasing cluster size from $E_g \approx 4.8$ eV for the smallest cluster containing 10 Si atoms to $E_g \approx 3$ eV for the biggest cluster with almost 60 Si atoms. This decrease is a consequence of quantum confinement effects, which are less pronounced for bigger clusters. The plane-wave GW HOMO-LUMO gaps show a similar trend as a function of cluster size, but the gaps are larger than the DFT–PBE gaps by several electron volts. Interestingly, the GW corrections are larger for smaller clusters than for larger clusters. As a consequence, the reduction in the GW
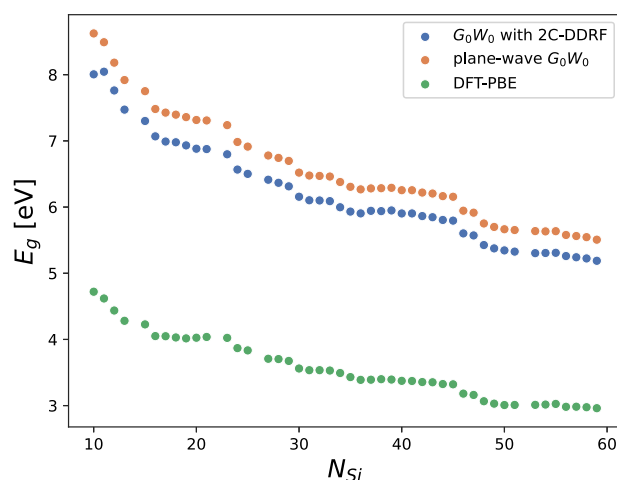

Fig. 1   HOMO–LUMO gaps of silicon clusters. HOMO–LUMO gaps of hydrogenated silicon clusters from DFT–PBE Kohn–Sham eigenvalues, plane-wave G_0W_0 and G_0W_0 calculations using the 2C-DDRF.

HOMO–LUMO gaps as a function of cluster size is larger compared to the DFT–PBE result: in particular, the gap is as large as 8.6 eV for the smallest clusters and shrinks to 5.5 eV for the largest clusters corresponding to a decrease of 3.1 eV (compared to a decrease of 1.8 eV in the DFT–PBE HOMO–LUMO gap energies). Similar results were obtained by Chelikowsky et al.[51], who also carried out GW calculations on hydrogenated Si clusters. In particular, they found that the HOMO–LUMO gap shrinks from ~9 eV for a 10 Si atom cluster to ~6.5 eV for a 47 Si atom cluster. The GW results obtained with the 2C-DDRF are qualitatively similar to the plane-wave GW results. However, the HOMO–LUMO gaps that are obtained with this approach are consistently ~0.3–0.4 eV smaller than the plane-wave results. This is a consequence of the incompleteness of the local basis set. Interestingly, the calculated HOMO–LUMO gaps exhibit step-like features at clusters with 16, 24, and 46 silicon atoms. Inspection of the atomic structure of these clusters reveals that they exhibit one or more SiH₃ units on their surface, see Fig. 2, suggesting an interesting interplay between the chemical bonding and the HOMO–LUMO gaps in these systems.

Next, we determine the 1C-DDRF. For the basis set, we use solid harmonic Gaussians with optimized decay coefficients: see the "Methods" section. Figure 3a compares the HOMO–LUMO gaps from G_0W_0 calculations with the 1C-DDRF to those obtained with the 2C-DDRF and also to plane-wave G_0W_0 results. For small clusters, the HOMO–LUMO gaps obtained with the 1C-DDRF are smaller than those obtained with the 2C-DDRF, while the opposite behavior is observed for larger clusters. The largest difference between the two methods is obtained for clusters containing ~40 Si atoms. The root-mean-square error (RMSE) of the 1C-basis results relative to the 2C-basis results is 0.22 eV, and the RMSE relative to the plane-wave results is 0.45 eV for all clusters. Figure 3b shows the HOMO and LUMO quasiparticle energies. It can be seen that better agreement with the plane-wave result is obtained for the LUMO than for the HOMO.

Figure 4a shows the quasiparticle energy corrections of the ten lowest conduction orbitals and the ten highest valence orbitals from plane-wave G_0W_0 and G_0W_0 with the 1C-DDRF. The corrections obtained with the 1C-DDRF follow a similar trend as those obtained from the plane-wave calculation. For the unoccupied states, the quantitative agreement is better than for the occupied states, but the 1C-DDRF results for the unoccupied states are scattered over a larger energy range than the plane-wave results. To analyze the errors that arise from the use of the 1C-DDRF in more detail, Fig. 4b shows a two-dimensional histogram of the difference in QP corrections between plane-
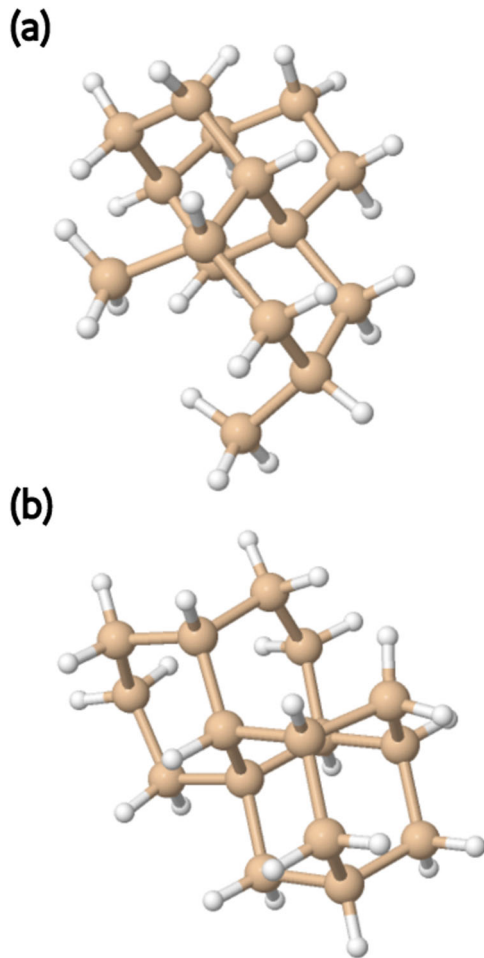
Fig. 2 Atomic structure of silicon clusters. Clusters with **a** 15 Si atoms and **b** 16 Si atoms. Notice the presence of two SiH_3 units on the surface of the cluster in (**a**). Hydrogen atoms are white; silicon atoms are brown.

wave G_0W_0 and G_0W_0 with the 1C-DDRF. For the occupied states, the differences are mostly smaller than 0.4 eV, while they are somewhat smaller for the unoccupied states. The RMSE over all energy levels is 0.32 eV.

Now that we have established the accuracy of the method used to generate the training set, we use a dense NN in conjunction with the NDM descriptor to generate the coefficients of the 1C-DDRF according to

$$\chi^{(i)}_{nlmn'l'm'} = f\left(\rho^{(i,Si)}_{nlm}, \rho^{(i,H)}_{nlm}\right),\tag{17}$$

where f is the NN function. The hydrogen and silicon environment descriptors are concatenated into a single vector before being fed into the NN. A separate network is trained for Si and H contributions to the DDRF. The exact architecture of the network as well as the practical computation of the atomic decomposition and the descriptors, are described in the "Methods" section. To generate the training data for the NN, we start from the set of relaxed hydrogenated Si clusters that were studied above. From each relaxed cluster, we generate six configurations by randomly displacing the atoms with the magnitude of the displacements being drawn from a uniform distribution with a maximum of 0.1 Å. For these clusters, we then calculate the 1C-DDRF.

Once the NN is trained on the 1C-DDRFs of the randomly displaced clusters, we use it to calculate the 1C-DDRFs of the relaxed clusters and then determine quasiparticle energies via the
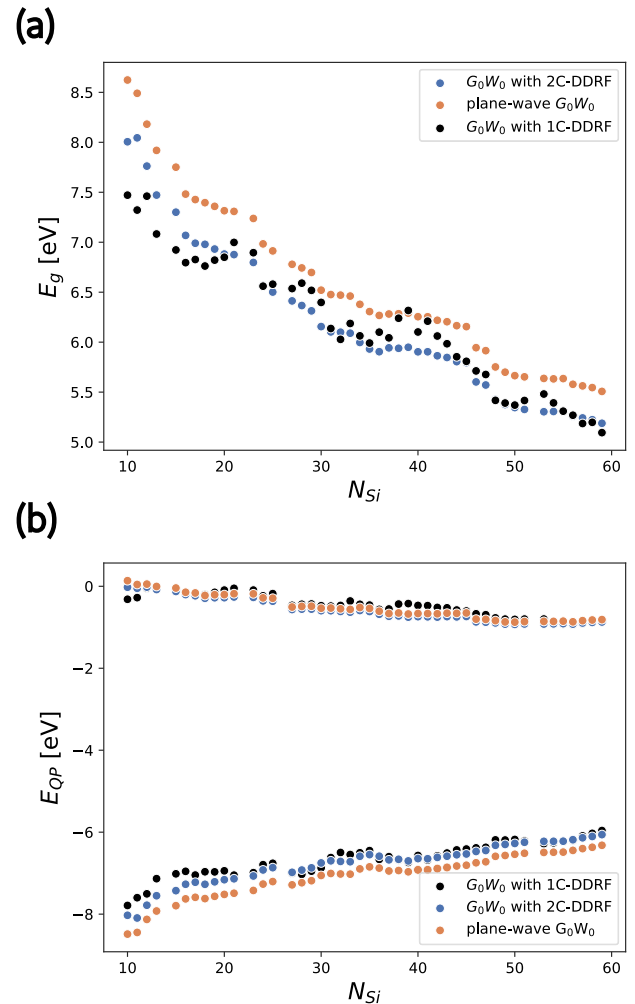

Fig. 3 HOMO–LUMO gaps, HOMO and LUMO energies of silicon clusters. **a** HOMO–LUMO gaps of hydrogenated silicon clusters from plane-wave G_0W_0 and G_0W_0 calculations using the 2C-DDRF and G_0W_0 calculations using the 1C-DDRF. **b** HOMO and LUMO energies of hydrogenated Si clusters.

ML–GW approach. Figure 5 compares the HOMO–LUMO gaps from ML–GW and GW with explicitly calculated 1C-DDRFs. Except for the smallest cluster, the ML–GW method accurately reproduces the HOMO-LUMO gaps of the explicit GW calculations. The worse performance for the smallest cluster is a consequence of the training set, which contains a large number of bigger clusters containing atomic environments that differ from those found in the smallest clusters. The overall RMSE of the ML–GW method relative to the explicit GW with the 1C-basis is only 0.15 eV but reduces to 0.06 eV when the smallest cluster is excluded.

Figure 6 shows the difference in QP corrections between ML–GW and GW with the 1C-DDRF for the 10 highest valence states and 10 lowest conduction states, with the energies of the smallest cluster excluded. ML–GW produces QP shifts for both valence and conduction states within 0.1 eV from the explicit G_0W_0 with the 1C-DDRF. The majority of valence states exhibit a positive error, while for conduction states, the error is largely negative.

Figure 7 compares the ML–G_0W_0 QP corrections to plane-wave G_0W_0 results, again with the energies of the smallest cluster excluded. As expected, the differences are very similar to those between plane-wave G_0W_0 and the explicit G_0W_0 with the 1C-basis. In particular, the RMSE is 0.34 eV for all clusters and reduces
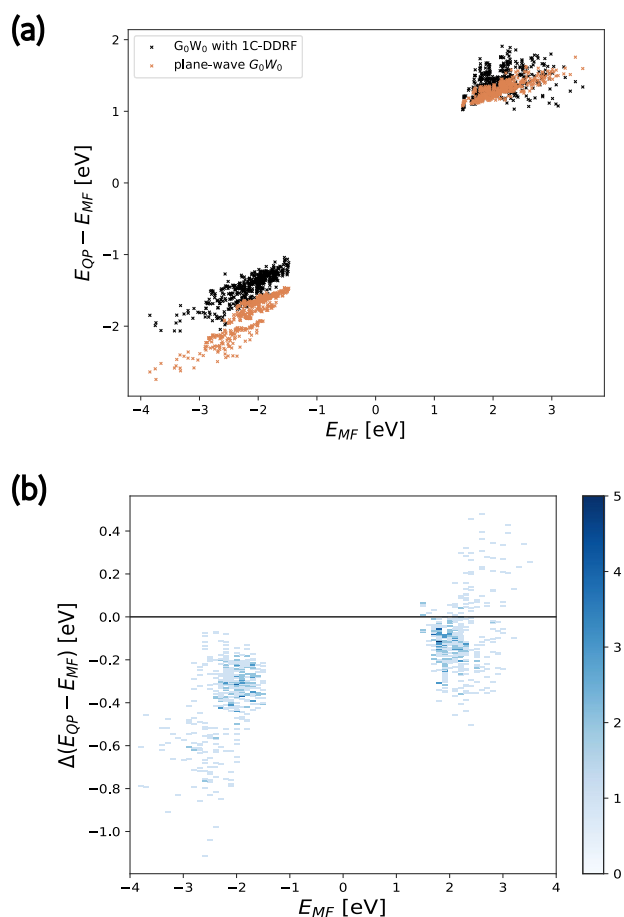
(a)

(b)

Fig. 4 QP corrections obtained from plane-wave GW and 1C-GW. **a** Quasiparticle corrections from plane-wave G_0W_0 and G_0W_0 with the 1C-DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. **b** Histogram of difference in quasiparticle corrections from plane-wave G_0W_0 and G_0W_0 calculations with the 1C-DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO–LUMO gap.
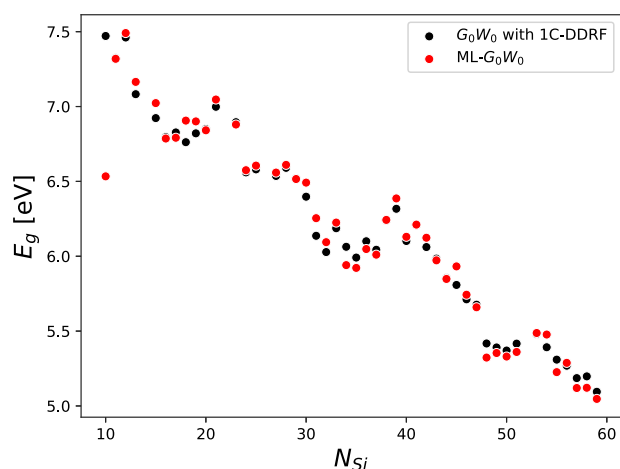
Fig. 5 1C and ML–GW HOMO–LUMO gaps. HOMO–LUMO gaps of hydrogenated silicon clusters obtained from G_0W_0 calculations using the 1C-DDRF and ML–G_0W_0.
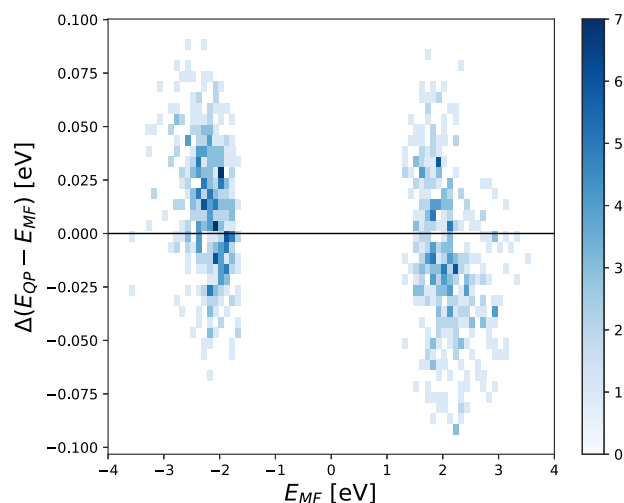
Fig. 6 ML–GW QP correction error compared to 1C-GW. Histogram of difference in quasiparticle corrections from G_0W_0 using the 1C-DDRF and ML–G_0W_0 for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO–LUMO gap. The energies of the smallest cluster were excluded.

to 0.30 eV when the smallest cluster is excluded. This result demonstrates that the key obstacle to improving the ML–GW approach is the development of a better basis set.

Finally, we test the ability of the ML–GW approach to predict the quasiparticle energies of clusters that are larger than those included in the training data. For this, we only include clusters with up to N_{max} Si atoms in the training set, with N_{max} being 60, 50, and 40. Again, the training set only includes clusters with randomly displaced atoms, and the test set consists of relaxed clusters. The predicted ML–GW for the whole set of relaxed clusters is shown in Fig. 8. From this graph, it is clear that the accuracy of the prediction for the largest clusters deteriorates as N_{max} is reduced: while for $N_{max} = 60$, the gaps and QP corrections for clusters with more than 60 Si atoms are still highly accurate, larger differences are observed for $N_{max} = 50$. For $N_{max} = 40$, errors as large as 1 eV are obtained for the gaps of clusters with around 50 Si atoms. Figure 8f shows that the large error in the gaps is a consequence of having a negative error in the QP shifts for occupied states and a positive error in the shift for unoccupied states. In other words: instead of a cancellation, we get an accumulation of errors when computing HOMO–LUMO gaps.

DISCUSSION

We have developed an ML approach to predict the interacting DDRF of materials. To achieve this, we introduce a decomposition of the DDRF into atomic contributions, which form the output of a NN. We also introduce the NDM descriptor, which is a generalization of the widely used SOAP descriptor[21]: instead of symmetrizing the descriptor using a Haar integral over a symmetry group[52], we construct the tensor product of the expansion coefficients of the neighborhood density, which transforms under rotation in the same way as the atomic contributions to the DDRF. Thus, while not fully covariant, our approach is able to distinguish between different orientations of a chemical environment, which is a key requirement for predicting functions such as the DDRF.

The ML technique for DDRFs is then combined with the GW approach. The resulting method is called the ML–GW approach. We apply this method to hydrogenated silicon clusters. The ML–GW approach reproduces HOMO–LUMO gaps and quasiparticle
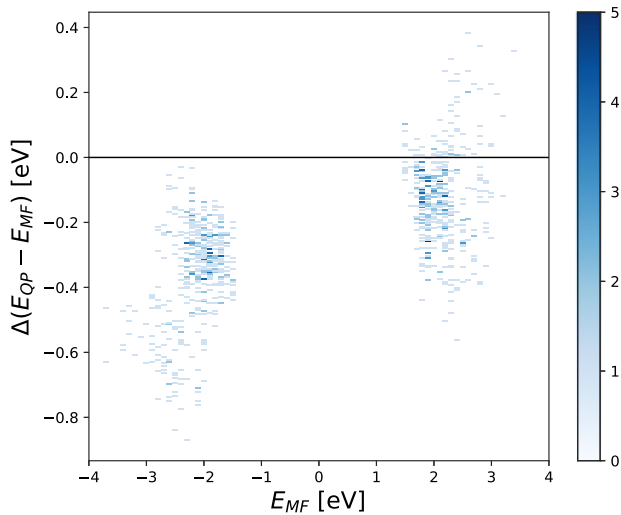
Fig. 7 QP correction error of ML–GW compared to plane-wave GW. Histogram of difference in quasiparticle corrections from plane-wave G_0W_0 and ML–G_0W_0 DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO–LUMO gap. The energies of the smallest cluster were excluded.

energies of GW calculations using the explicitly calculated 1C-DDRF, i.e., the DDRF in a pair basis where the basis functions of each pair are centered on the same atom, with an accuracy of about 0.1 eV. The accuracy of the results deteriorates when it is applied to clusters that are larger than those included in the training set.

However, the error of ML–GW is significantly larger when compared to standard plane-wave GW results: HOMO–LUMO gaps are reproduced to within 0.5 eV, but the error reduces to 0.4 eV when the smallest cluster is excluded from the test set. These errors are comparable to those obtained by Rohlfing in his GW calculations for silane using a model dielectric function[9].

These findings demonstrate that the main challenge to improving the ML–GW method is the construction of better local basis sets for the DDRF. The basis used for the 2C-DDRF can be improved straightforwardly by using larger basis sets, such as aug-admm-2, admm-3, or aug-admm-3[50]. However, it is more difficult to increase the basis used for the 1C-DDRF as this leads to linear dependencies, which deteriorate the predictive accuracy of the NN. This was also observed by Grisafi et al.[28] when predicting the expansion coefficients of the electronic density using the symmetry-adapted SOAP kernel[25]. In the future, we plan to explore the use of orthogonal radial basis sets, such as Laguerre polynomials, instead of solid harmonic Gaussians.

We expect that the ML–GW method can be applied to calculate quasiparticle energies in systems that have so far been out of reach for standard implementations. Examples include disordered materials, liquids, interfaces, or nanoparticles. It could also be combined with on-the-fly ML methods[53] to perform GW calculations on molecular-dynamics snapshots to determine finite-temperature quasiparticle energies.

METHODS
Data generation
The atomic structures of the hydrogenated silicon clusters were obtained in the same way as described by Zauchner et al.[54]: starting from the $Si_{123}H_{100}$ cluster of the silicon Quantum Dot data set[55], we remove the silicon atom furthest from the center of the cluster, terminate the dangling bonds with hydrogen atoms and relax the resulting structure using DFT. The process is repeated

until only 10 silicon atoms remain. From this set of silicon clusters, only clusters with fewer than 60 silicon atoms were used in the training set for DDRF prediction. From each cluster with fewer than 60 silicon atoms, we created six additional clusters in which random displacements were added to the atomic positions. The magnitudes of the displacements were drawn from a uniform distribution with a width of 0.1 Å. Finally, calculations were also carried out for clusters with between 60 and 70 silicon atoms. These clusters are not part of the training set but are used to test the extrapolation capacity of the ML approach. Note that all calculations were carried out for clusters in a vacuum, i.e., we did not consider the effect of a substrate or a solid matrix.

DFT and GW calculations
The DDRF and QP corrections were calculated using the BerkeleyGW software package[7,56]. This code uses a plane-wave basis to represent the DDRF which makes it possible to systematically converge results by increasing the plane-wave cutoff. In contrast, it is often more difficult to achieve convergence when GW implementations based on local orbitals are used. Mean-field DFT calculations were performed using the Quantum Espresso code[57,58]. Norm-conserving pseudopotentials from the Quantum Espresso Pseudopotential Library were used. The parameters of the DFT calculations were the same as those used by Zauchner et al.[54]: a plane-wave cut-off of 65 Ry and a supercell with sufficient vacuum to avoid interactions between periodic images. For the calculation of the DDRF, a total of 1000 Kohn–Sham states were used in the summation. Also, a plane-wave cut-off of 6 Ry and a truncated Coulomb interaction was used. The QP corrections were calculated using the GPP[7], an explicit sum over 1000 Kohn–Sham states, and also a static remainder correction[59]. To calculate the HOMO and LUMO energies, the vacuum level was determined by averaging the electrostatic potential over the faces of the supercell.

Projection onto the intermediate basis
We first use BerkeleyGW to calculate the inverse dielectric matrix $\epsilon_{\mathbf{GG'}}^{-1}$ in a plane-wave basis[56]. From this, we determine the interacting DDRF via

$$\chi_{\mathbf{GG'}} = (\epsilon_{\mathbf{GG'}}^{-1} - \delta_{\mathbf{GG'}})/v_{\mathbf{G}} \tag{18}$$

with $v_{\mathbf{G}}$ being the Fourier transform of the truncated Coulomb interaction.

Next, the DDRF in real space is obtained as

$$\chi(\mathbf{r}, \mathbf{r'}) = \frac{1}{V} \sum_{\mathbf{G}, \mathbf{G'}} e^{i\mathbf{G} \cdot \mathbf{r}} \chi_{\mathbf{GG'}} e^{-i\mathbf{G'} \cdot \mathbf{r'}}, \tag{19}$$

where V is the volume of the supercell.

Starting from a set of real atom-centered basis functions $\phi_{a_i}^i(\mathbf{r})$, where a_i labels the basis function on atom i, we construct an orthogonal basis set $\tilde{\phi}_{a_i}^i(\mathbf{r})$

$$\tilde{\phi}_{a_i}^i(\mathbf{r}) = \sum_k \sum_{a_k} A_{ik}^{a_i a_k} \phi_{a_k}^k(\mathbf{r}), \tag{20}$$

where $A_{ik}^{a_i a_k}$ is the matrix of eigenvectors of the overlap matrix. The coefficients of the DDRF, when expanded on the orthogonalized basis, are

$$\tilde{\chi}_{a_i a_j}^{ij} = \frac{1}{V} \sum_{\mathbf{G}, \mathbf{G'}} \chi_{\mathbf{G}, \mathbf{G'}}$$
$$\times \int_{-\infty}^{\infty} \tilde{\phi}_{a_i}^i(\mathbf{r}) e^{i\mathbf{G} \cdot \mathbf{r}} d\mathbf{r} \int_{-\infty}^{\infty} e^{-i\mathbf{G'} \cdot \mathbf{r'}} \tilde{\phi}_{a_j}^j(\mathbf{r'}) d\mathbf{r'}, \tag{21}$$

where, due to the localized nature of the basis functions, we extended the integral from an integral over the supercell to an integral over all space. These integrals are proportional to the Fourier transforms of the basis functions (or their complex
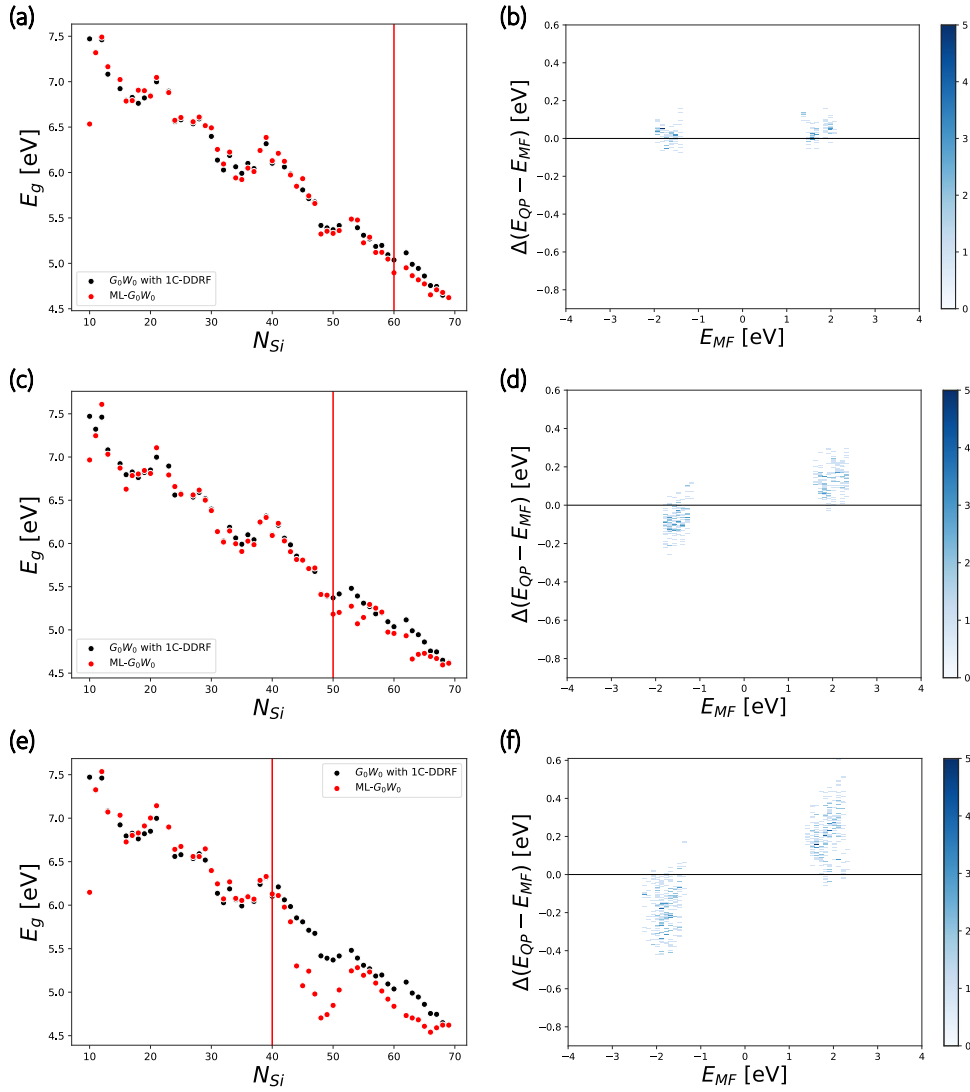
Fig. 8 Performance of ML–GW when extrapolating to larger clusters. HOMO-LUMO gaps (left panels - **a**, **c**, **e**) and errors in quasiparticle shifts (right panels - **b**, **d**, **f**) from explicit G_0W_0 calculations with the 1C-DDRF and from ML–G_0W_0 trained on clusters containing up to $N_{max} = 60$ Si atoms (upper panels - **a**, **b**), $N_{max} = 50$ Si atoms (middle panels - **c**, **d**) and $N_{max} = 40$ Si atoms (lower panels - **e**, **f**). The red vertical line indicates N_{max}. The panels on the right-hand side (**b**, **d**, **f**) only contain results for clusters with more Si atoms than N_{max}. The mean-field energies are referenced to the middle of the mean-field HOMO–LUMO gap.

conjugates). We note that it is possible to skip this step if a GW implementation based on local orbitals is used[16].

We then transform back to the non-orthogonal localized basis set using Eq. (20) to find

$$\chi(\mathbf{r}, \mathbf{r}') = \sum_{a_i a_j} \sum_{ij} \tilde{\chi}^{ij}_{a_i a_j} \tilde{\phi}^i_{a_i}(\mathbf{r}) \tilde{\phi}^j_{a_j}(\mathbf{r}')$$

$$= \sum_{a_k a_l} \sum_{kl} \sum_{a_i a_j} \sum_{ij} A^{a_i a_k}_{ik} A^{a_i a_k}_{jl} \tilde{\chi}^{ij}_{a_i a_j} \phi^k_{a_k}(\mathbf{r}) \phi^l_{a_l}(\mathbf{r}')$$

$$= \sum_{a_k a_l} \sum_{kl} \chi^{kl}_{a_k a_l} \phi^k_{a_k}(\mathbf{r}) \phi^l_{a_l}(\mathbf{r}'), \tag{22}$$

where we defined

$$\chi^{kl}_{a_k a_l} = \sum_{a_i a_j} \sum_{ij} \tilde{\chi}^{ij}_{a_i a_j} A^{a_i a_k}_{ik} A^{a_j a_k}_{jl}. \tag{23}$$

The basis functions we employed are the real solid harmonic Gaussians as defined in LibInt[60]

$$\phi_{lm}(r, \theta, \phi) = N_l(\beta) r^l e^{-\beta r^2} R_{lm}(\theta, \phi), \tag{24}$$

where β is a decay parameter, $N_l(\beta)$ is a normalization factor, and R_{lm} are the real spherical harmonics given by[61]

$$R_{lm}(\theta, \phi)$$
$$= \begin{cases} \frac{i}{\sqrt{2}}\left(Y_{l-|m|}(\theta,\phi) - (-1)^m Y_{l|m|}(\theta,\phi)\right) & \text{if } m < 0 \\ Y_{lm}(\theta,\phi) & \text{if } m = 0 \\ \frac{1}{\sqrt{2}}\left(Y_{l-|m|}(\theta,\phi) + (-1)^m Y_{l|m|}(\theta,\phi)\right) & \text{if } m > 0, \end{cases} \tag{25}$$

where $Y_{lm}(\theta, \phi)$ are the complex spherical harmonics with the Condon–Shortley phase convention. Kuang and Lin showed that the Fourier transform of the complex solid harmonic Gaussians is again a solid harmonic Gaussian[62]

$$\frac{1}{(2\pi)^{3/2}} \int d\mathbf{r} e^{-i\mathbf{G}\cdot\mathbf{r}} N_l(\beta) r^l e^{-\beta r^2} Y_{lm}(\hat{\mathbf{r}})$$
$$= (-i)^l \tilde{N}_l(\beta) G^l e^{-G^2/(4\beta)} Y_{lm}(\hat{\mathbf{G}}), \tag{26}$$

with $\tilde{N}_l(\beta) = N_l(\beta)/(2\beta)^{3/2}$. The Fourier transform of the real solid harmonic Gaussians can then be easily computed using Eq. (25).

The basis set used in this work is a modified version of the admm-2 basis set[50] (see Supplementary Methods for details), in which the s-orbitals were removed and contracted Gaussians were uncontracted into individual basis functions. Removing the s-orbitals ensures that $\int d\mathbf{r}\chi(\mathbf{r},\mathbf{r}') = 0$ since only the Fourier transform of s-orbitals has a $\mathbf{G}=0$ contribution.

Projection onto the atomic basis

The fully atom-centered basis set also consists of solid harmonic Gaussians. The basis set was constructed following the same procedure as in the DScribe library[63], where individual basis functions are given by

$$\psi_{nlm}(r,\theta,\phi) = N_l(\beta_{nl})r^l e^{-\beta_{nl}r^2} R_{lm}(\theta,\phi), \qquad (27)$$

where the basis set is truncated at a maximum angular momentum l_{max} and a maximum principal quantum number n_{max}. For silicon atoms we use $l_{max} = n_{max} = 4$. For hydrogen atoms we use $l_{max} = n_{max} = 3$.

The exponents β_{nl} are constructed such that the corresponding basis functions decay to zero at a cutoff radius R_n, i.e., $\beta_{nl} = -\ln(\frac{T}{R_n^l})/R_n^2$ with $T = 10^{-3}$ Å^l being a threshold parameter. The cutoff radius $R_n = R_i + (R_o - R_i)/n$ lies between an inner radius R_i and an outer radius R_o. For hydrogen atoms, we used $R_i = 0.1$ Å and $R_o = 3.0$ Å and for silicon atoms, we used $R_i = 1.0$ Å and $R_o = 8.0$ Å. Additionally, for silicon atoms, we also included the basis functions of the modified admm-2 basis. Both R_i and R_o were optimized to minimize linear dependencies in the basis set, as such dependencies significantly deteriorate the accuracy of the NN predictions. A similar observation was made by Grisafi et al.[28] when learning electron densities, although a different approach was taken to remedy this issue in their work.

In order to compute the coefficients of the atomic contributions to the DDRF in the fully atom-centered basis, the same procedure as in the intermediate basis was used: the basis was first orthogonalized by computing the eigenvectors of the overlap matrix. Then the atomic DDRFs in the intermediate basis were projected onto the orthogonalized fully-atom centered basis with overlaps between the different basis functions being computed using LibInt[60]. Then the atomic DDRFs were transformed back to the non-orthogonal basis producing the desired coefficients $\chi^{(i)}_{nlmn'l'm'}$.

Descriptors

The basis set for neighborhood densities was generated using the same procedure as for the fully atom-centered basis for the DDRF. However, s-orbitals were not removed and the basis functions of the admm-2 basis set were not included. We used $R_i = 1.0$ Å for both hydrogen and silicon atoms $R_o = 4.0$ Å for hydrogen atoms and $R_o = 9.0$ Å for silicon atoms. The exponents of the Gaussians in Eq. (12) were set such that the standard deviation of the Gaussians is 0.5 Å. LibInt[60] was again used to compute the required integrals for the projection.

Neural network

A dense NN with four hidden layers with 2000, 1500, 1000, and 2000 nodes, respectively, was constructed for both silicon and hydrogen atoms. Each layer uses a Leaky–ReLu activation function with a leak parameter of 0.1. The output layer was further symmetrized by adding its transpose. The loss used was the mean-squared error between the predicted and true expansion coefficients $\chi^{(i)}_{nlmn'l'm'}$. The NN was trained on the perturbed clusters for 20,000 epochs. We found that adding dropout to the layers does not significantly improve the quasiparticle energies resulting from the predictions, which is likely due to the similarity between the atomic environments in the training and test set.

REFERENCES

1. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).
2. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
3. Sham, L. J. & Schlüter, M. Density-functional theory of the energy gap. *Phys. Rev. Lett.* **51**, 1888–1891 (1983).
4. Schultz, P. A. Theory of defect levels and the "band gap problem" in silicon. *Phys. Rev. Lett.* **96**, 246401 (2006).
5. Hedin, L. New method for calculating the one-particle green's function with application to the electron-gas problem. *Phys. Rev.* **139**, A796–A823 (1965).
6. Strinati, G., Mattausch, H. J. & Hanke, W. Dynamical aspects of correlation corrections in a covalent crystal. *Phys. Rev. B* **25**, 2867–2888 (1982).
7. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. *Phys. Rev. B* **34**, 5390–5413 (1986).
8. Onida, G., Reining, L. & Rubio, A. Electronic excitations: density-functional versus many-body green's-function approaches. *Rev. Mod. Phys.* **74**, 601–659 (2002).
9. Rohlfing, M. Electronic excitations from a perturbative LDA + GdW approach. *Phys. Rev. B* **82**, 205127 (2010).
10. Adler, S. L. Quantum theory of the dielectric constant in real solids. *Phys. Rev.* **126**, 413–420 (1962).
11. Wiser, N. Dielectric constant with local field effects included. *Phys. Rev.* **129**, 62–69 (1963).
12. Del Ben, M. et al. Large-scale GW calculations on pre-exascale HPC systems. *Comput. Phys. Commun.* **235**, 187–195 (2019).
13. Govoni, M. & Galli, G. Large scale GW calculations. *J. Chem. Theory Comput* **11**, 2680–2696 (2015).
14. Wilhelm, J., Golze, D., Talirz, L., Hutter, J. & Pignedoli, C. A. Toward GW calculations on thousands of atoms. *J. Phys. Chem. Lett.* **9**, 306–312 (2018).
15. Förster, A. & Visscher, L. Low-order scaling G0W0 by pair atomic density fitting. *J. Chem. Theory Comput.* **16**, 7381–7399 (2020).
16. Duchemin, I. & Blase, X. Cubic-scaling all-electron GW calculations with a separable density-fitting space–time approach. *J. Chem. Theory Comput.* **17**, 2383–2393 (2021).
17. Hybertsen, M. S. & Louie, S. G. Model dielectric matrices for quasiparticle self-energy calculations. *Phys. Rev. B* **37**, 2733–2736 (1988).
18. Cappellini, G., Del Sole, R., Reining, L. & Bechstedt, F. Model dielectric function for semiconductors. *Phys. Rev. B* **47**, 9892–9895 (1993).
19. Bechstedt, F., Sole, R. D., Cappellini, G. & Reining, L. An efficient method for calculating quasiparticle energies in semiconductors. *Solid State Commun.* **84**, 765 – 770 (1992).
20. Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
21. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
22. Hansen, K. et al. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **9**, 3404–3419 (2013).
23. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
24. Hansen, K. et al. Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.* **6**, 2326–2331 (2015).
25. Grisafi, A., Wilkins, D. M., Csányi, G. & Ceriotti, M. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.* **120**, 036002 (2018).

26. Wilkins, D. M. et al. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl. Acad. Sci. USA* **116**, 3401–3406 (2019).

27. Veit, M., Wilkins, D. M., Yang, Y., DiStasio, R. A. & Ceriotti, M. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J. Chem. Phys.* **153**, 024113 (2020).

28. Grisafi, A. et al. Transferable machine-learning model of the electron density. *ACS Cent. Sci.* **5**, 57–64 (2019).

29. Brockherde, F. et al. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **8**, 872 (2017).

30. Alred, J. M., Bets, K. V., Xie, Y. & Yakobson, B. I. Machine learning electron density in sulfur crosslinked carbon nanotubes. *Compos. Sci. Technol.* **166**, 3–9 (2018).

31. Chandrasekaran, A. et al. Solving the electronic structure problem with machine learning. *npj Comput. Mater.* **5**, 22 (2019).

32. Kondor, R., Lin, Z. & Trivedi, S. Clebsch–Gordan nets: a fully Fourier space spherical convolutional neural network. In *Advances in neural information processing*, 10117–10126 (vol. 31, Curran Associates, Inc., 2018).

33. Kondor, R. & Trivedi, S. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *Proceedings of the 35th International Conference on Machine Learning*, 2747–2755 (Proceedings of Machine Learning Research vol. 80, PMLR, 2018).

34. Anderson, B., Hy, T. S. & Kondor, R. Cormorant: Covariant molecular neural networks. In *Advances in Neural Information Processing*, 14537—14546 (vol. 32, Curran Associates, Inc., 2019).

35. Thomas, N. et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. Preprint at *arXiv* http://arxiv.org/abs/1802.08219 (2018).

36. Cohen, T. S., Geiger, M., Köhler, J. & Welling, M. Spherical CNNs. Preprint at *arXiv* http://arxiv.org/abs/1801.10130 (2018).

37. Cohen, T. & Welling, M. Group equivariant convolutional networks. In *Proceedings of The 33rd International Conference on Machine Learning*, 2990–2999 (Proceedings of Machine Learning Research vol. 48, PMLR, 2016).

38. Lapchevskyi, K. et al. Euclidean neural networks (e3nn) v1.0, version v1.0. Available at https://www.osti.gov/servlets/purl/1770279 (2020).

39. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks. Preprint at *arXiv* https://arxiv.org/abs/2207.09453 (2022).

40. Westermayr, J. & Maurer, R. J. Physically inspired deep learning of molecular excitations and photoemission spectra. *Chem. Sci.* **12**, 10755–10764 (2021).

41. Knøsgaard, N. R. & Thygesen, K. S. Representing individual electronic states for machine learning GW band structures of 2D materials. *Nat. Commun.* **13**, 468 (2022).

42. Golze, D. et al. Accurate computational prediction of core-electron binding energies in carbon-based materials: a machine-learning model combining density-functional theory and gw. *Chem. Mater.* **34**, 6240–6254 (2022).

43. Hybertsen, M. S. & Louie, S. G. First-principles theory of quasiparticles: calculation of band gaps in semiconductors and insulators. *Phys. Rev. Lett.* **55**, 1418–1421 (1985).

44. Lischner, J., Sharifzadeh, S., Deslippe, J., Neaton, J. B. & Louie, S. G. Effects of self-consistency and plasmon-pole models on GW calculations for closed-shell molecules. *Phys. Rev. B* **90**, 115130 (2014).

45. Sharifzadeh, S., Tamblyn, I., Doak, P., Darancet, P. T. & Neaton, J. B. Quantitative molecular orbital energies within a G0W0 approximation. *Eur. Phys. J. B* **85**, 323 (2012).

46. We note that locality was already exploited through a local orbital representation in one of the first applications of the GW method to study a real material by Strinati et al. [6].

47. Mussard, B. & Ángyán, J. G. Relationships between charge density response functions, exchange holes and localized orbitals. *Comput. Theor. Chem.* **1053**, 44–52 (2015).

48. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).

49. Rose, M. *Elementary Theory of Angular Momentum 1st edn.* Structure of matter series (Wiley, 1957).

50. Kumar, C. et al. Accelerating Kohn-Sham response theory using density fitting and the auxiliary-density-matrix method. *Int. J. Quantum Chem.* **118**, e25639 (2018).

51. Tiago, M. L. & Chelikowsky, J. R. Optical excitations in organic molecules, clusters, and defects studied by first-principles green's function methods. *Phys. Rev. B* **73**, 205334 (2006).

52. Langer, M. F., Goeßmann, A. & Rupp, M. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. *npj Comput. Mater.* **8**, 41 (2022).

53. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).

54. Zauchner, M. G., Forno, S. D., Csányi, G., Horsfield, A. & Lischner, J. Predicting polarizabilities of silicon clusters using local chemical environments. *Mach. Learn.* **2**, 045029 (2021).

55. Barnard, A. W. & Hugh. Silicon quantum dot data set. *CSIRO***v2**. Dataset at https://doi.org/10.4225/08/5721BB609EDB0 (2015).

56. Deslippe, J. et al. Berkeleygw: a massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. *Comput. Phys. Commun.* **183**, 1269–1289 (2012).

57. Giannozzi, P. et al. Advanced capabilities for materials modelling with quantum espresso. *J. Phys. Condens. Matter* **29**, 465901 (2017).

58. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).

59. Deslippe, J., Samsonidze, G., Jain, M., Cohen, M. L. & Louie, S. G. Coulomb-hole summations and energies for g w calculations with limited number of empty orbitals: a modified static remainder approach. *Phys. Rev. B* **87**, 165124 (2013).

60. Valeev, E. F. Libint: A library for the evaluation of molecular integrals of many-body operators over Gaussian functions. http://libint.valeyev.net/ (2022). Version 2.8.0.

61. Schlegel, H. B. & Frisch, M. J. Transformation between cartesian and pure spherical harmonic Gaussians. *Int. J. Quantum Chem.* **54**, 83–87 (1995).

62. Kuang, J. & Lin, C. D. Molecular integrals over spherical gaussian-type orbitals: I. *J. Phys. B* **30**, 2529–2548 (1997).

63. Himanen, L. et al. DScribe: library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).

64. Zauchner, M., Lischner, J. & Horsfield, A. Accelerating GW calculations through machine learned dielectric matrices. Dataset at https://archive.materialscloud.org/record/2023.119 (2023).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

M.Z. developed the methodology, implemented the code, wrote the first draft, and contributed to the presentation of results and revisions of the paper. J.L. and A.H. supervised the project and contributed to the presentation of results and to revisions of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-023-01136-y.

Correspondence and requests for materials should be addressed to Johannes Lischner.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.