Supplementary Information

Machine Learning Dielectric Screening for the Simulation of Excited State Properties of Molecules and Materials

Sijia S. Dong[1,2], Marco Govoni[1,2,†], and Giulia Galli[2,1,*]

[1]Materials Science Division and Center for Molecular Engineering, Argonne National Laboratory, Lemont, IL 60439, USA

[2]Pritzker School of Molecular Engineering, the University of Chicago, Chicago, IL 60637, USA

[†]mgovoni@anl.gov

[*]gagalli@uchicago.edu

February 12, 2021

S1 Computational details

In the following sections we report the computational details used in this work.

S1.1 Systems

We considered the systems reported in Table S1. The electronic structure of each system was computed at the density functional theory (DFT) level of theory using plane wave basis sets and the ONCV pseudopotentials,[1] with the Perdew–Burke–Ernzerhof (PBE)[2] exchange and correlation functional. Quantum Espresso[3] (version 6.1.0) and the Qbox[4] (version 1.66.2) codes were used. For each system the macroscopic dielectric constant, ϵ_∞, was calculated by averaging the diagonal elements of the polarizability tensor computed using the Qbox code, and was labeled $\epsilon_\infty^{\mathrm{PT}}$. The electronic dipole used to compute the polarizability tensor is defined from using the center of charge of maximally localized Wannier functions (MLWF) with the refinement correction by Stengel and Spaldin.[5]

S1.2 First principles molecular dynamics simulations

First principles molecular dynamics (FPMD) simulations were carried out using the Qbox[4] code (version 1.66.2).

For liquid water, we considered unit cells with 16 or 64 water molecules. For 64-water samples, we considered snapshots from each of 10 independent FPMD trajectories (samples s0022-s0031) of the PBE400 dataset.[6] For 16-water samples, we generated 20 independent MD trajectories starting from 20 independent snapshots of 16 water molecules initiated randomly with the same atomic density as the PBE400 dataset (1.11 g/cm^3 D_2O). The initiation method is the same as the method used to generate the PBE400 dataset.[6] FPMD simulations were carried out using the Bussi-Donadio-Parrinello (BDP) thermostat[7] at 400 K with a thermostat time constant of 10000 a.u. and a time step of 10 a.u. (∼0.24 fs).

We modeled a hydrophobic Si/water interface with a slab containing 72 Si atoms, 24 H atoms terminating Si, and 108 water molecules. We modeled a hydrophilic Si/water interface with a slab containing 72 Si atoms,

Table S1: Systems considered in this work.

System	Number of atoms	Size of the cell (Å)	$\epsilon_\infty^{\mathrm{PT}}$
16-H_2O	48	$a = 7.82$	1.83±0.01
64-H_2O	192	$a = 12.41$	1.87±0.004
Si	64	$a = 5.43$	10.01
SiC	64	$a = 4.36$	6.12
C	64	$a = 3.57$	5.29
MgO	64	$a = 4.21$	3.16
LiF	64	$a = 4.03$	2.04
$Si_{10}H_{16}$	26	$a = 20.00$ to 50.00	-
$Si_{35}H_{36}$	71	$a = 25.00$	-
$Si_{87}H_{76}$	163	$a = 25.00$	-
H-Si/water	420	$a = 11.62, b = 13.42, c = 33.43$	3.34
COOH-Si/water	492	$a = 11.62, b = 13.42, c = 35.73$	3.84

24 -COOH groups terminating Si, and 108 water molecules. The geometrical configurations were taken from Pham et al.[8]

The FPMD simulations of the $Si_{35}H_{36}$ cluster were carried out using the BDP thermostat at 500 K. The thermostat time constant was 5000 a.u., and the time step was 20 a.u. (~0.48 fs). Equilibration was reached within the first 20000 time steps (~9.7 ps). The finite temperature absorption spectrum was obtained by averaging the spectra obtained for ten snapshots extracted every 5000 time steps from the FPMD trajectory, and starting after 10.16 ps.

S1.3 Calculations of absorption spectra

Absorption spectra calculations were carried out with the WEST[9]-Qbox[4] coupled codes using the FF-BSE scheme reported by Nguyen et al.[10].

For exited-state energies, a scissor operator was applied to the ground-state PBE energy levels to obtain a band gap corresponding to the value obtained at the G_0W_0@PBE level. Band gaps at the G_0W_0@PBE level were taken from the literature[8,9,11,12] except for 16-H_2O systems and the $Si_{10}H_{16}$ clusters, which we computed in this work. The values used are summarized in Table S2. For 16-H_2O snapshots, the G_0W_0@PBE band gap was computed using 640 PDEP eigenpotentials. For $Si_{10}H_{16}$, the G_0W_0@PBE band gap was computed using 1024 PDEP eigenpotentials and a cubic unit cell with a side length of 50 Å.

FF-BSE calculations were done at the Γ point. Parameters used in FF-BSE simulations are in Table S2.

The screening in FF-BSE[10] in the reciprocal space is expressed as follows:

$$\Delta\tau_{vv'}(\mathbf{G}) = \begin{cases} (\epsilon_\infty^{-1} - 1)\tau^u_{vv'}(\mathbf{G} = \mathbf{0}), & \mathbf{G} = \mathbf{0} \\ \frac{4\pi e^2}{|\mathbf{G}|^2}\frac{\rho^+_{vv'}(\mathbf{G}) - \rho^-_{vv'}(\mathbf{G})}{2}, & \mathbf{G} \neq \mathbf{0} \end{cases} \tag{S1}$$

where $\mathbf{G}$ is the reciprocal space lattice vector, and $\rho^\pm_{vv'}(\mathbf{G})$ is the Fourier component of $\rho^\pm_{vv'}(\mathbf{r})$. As described in the main text, we focus on getting a surrogate model corresponding to the $\mathbf{G} \neq \mathbf{0}$ terms; the term corresponding to $\mathbf{G} = \mathbf{0}$, the long-wavelength limit, is added separately.

For bulk Si, we used the Yambo[13,14] code (version 4.4.0) to compute BSE spectra at the Γ point and with k-point samplings. The Lanczos-Haydock solver was used, with scissor operators reported in Table S3. Details of the calculations with Yambo are reported in Table S3. For Model-BSE, the screened Coulomb potential $W = \epsilon^{-1}V_c$ was computed using ϵ defined as:

$$\epsilon^{-1}_{\mathbf{G},\mathbf{G}'} = \begin{cases} \epsilon_\infty^{-1}, & \mathbf{G} = \mathbf{G}' = \mathbf{0} \\ 1 + f, & \mathbf{G} = \mathbf{G}', \mathbf{G} \neq \mathbf{0} \\ 0, & \mathbf{G} \neq \mathbf{G}' \end{cases} \tag{S2}$$

Table S2: Parameters used to obtain the FF-BSE spectra.

System	G_0W_0@PBE band gap (eV)	Scissor operator (eV)	Bisection levels in each Cartesian direction	Bisection threshold	Kinetic energy cutoff (Ry)
16-H_2O	10.41	-	2	0.02	60
64-H_2O	8.1$^{\text{Ref. 12}}$	3.99±0.16	2	0.05	60
Si	1.37 (X_{1c}-point)$^{\text{Ref. 9}}$	0.70	3	0.02	40
SiC	2.28 (X_{1c}-point)$^{\text{Ref. 9}}$	0.95	2	0.04	70
C	5.50$^{\text{Ref. 11}}$	1.04	2	0.02	80
MgO	7.25$^{\text{Ref. 11}}$	2.48	2	0.01	60
LiF	13.27$^{\text{Ref. 11}}$	4.13	2	0.02	60
$Si_{10}H_{16}$	8.50	3.80	4	0.03	20
$Si_{35}H_{36}$	6.29$^{\text{Ref. 9}}$	2.80 (0 K); 3.58±0.17 (500 K)	4	0.03	25
$Si_{87}H_{76}$	4.77$^{\text{Ref. 9}}$	2.21	4	0.03	25
H-Si/water	1.34$^{\text{Ref. 8}}$	0.33	5	0.05	25[a]
COOH-Si/water	1.34$^{\text{Ref. 8}}$	0.27	5	0.05	25[a]

[a] Comments on this choice is discussed in Section S2.3.

where ϵ_∞ is the dielectric constant computed by Yambo, and f is the scaling factor defined in the main text. f is an input value and is specified in the caption of each Model-BSE spectrum reported in this article.

Table S3: Parameters of bulk Si for BSE calculations with the Yambo code.

System	Number of atoms	k-point	Size of dielectric matrix (in no. PWs)	Size of exchange term in the BSE kernel (in no. PWs)	Number of bands	Scissor operator (eV)
Si	64	Γ	1000	295667	320	0.70
Si	2	$12 \times 12 \times 12$	80	9185	20	0.85

S1.4 Machine learning

We carried out ML-BSE calculations by implementing an interface between WEST[9] and Tensorflow[15] (version 1.13). The convolutional model used in Eq. 10 of the main text is defined as:

$$\Delta\tau_{vv'}(x,y,z) = \sum_{i=0}^{n_x-1}\sum_{j=0}^{n_y-1}\sum_{k=0}^{n_z-1} K_{i,j,k}\tau^u_{vv'}[x+(il-m_x)\Delta x, y+(jl-m_y)\Delta y, z+(kl-m_z)\Delta z] \tag{S3}$$

where K is the convolutional filter of size (n_x, n_y, n_z), and Δx, Δy, Δz are the spacings of the uniform 3D-mesh used to represent periodic functions in real-space. $m_i, i = x, y, z$ is the size of padding in each of the x, y, z directions, and $m_i = \lfloor \frac{p_i}{2} \rfloor$ where $p_i = n_i + (n_i - 1)(l - 1) - 1$, and l is the dilation rate. We performed a hyperparameter search to determine the parameters in the training procedure. In the optimization procedure, we used the Adam optimizer[16] with a learning rate of 0.001. The loss was evaluated by mean squared error (MSE). Early stopping was used to stop training based on validation loss after 25 epochs without an improvement. Periodic boundary conditions are satisfied using padded arrays; we considered the same size for input and output arrays in Eq. 10. In order to save memory and training time, we have trained parameters for convolutional models skipping every other element of the $\tau^u_{vv'}$ and $\Delta\tau_{vv'}$

arrays in each Cartesian direction (except for 16-H_2O samples, where the size of the system allows us to consider all elements). In this way the arrays entering the training procedure have an 8 fold smaller memory footprint when the coarse grid is used. When the trained models are applied, we reconcile the fact that training was done on coarse grids, by applying a dilation rate of 2 in Eq. 10. In this way, the convolutional filter can be applied, after training, to $\tau^u_{vv'}$ arrays that are defined on the original FFT grid. Therefore, we used $l = 2$ in Eq. S3 except for the 16-H_2O system, for which we chose $l = 1$.

For each system considered here, the training and validation data come from one snapshot, and we use data from snapshots different from the training snapshot as the test set. We have trained ML models with either a global scaling factor or a convolutional model with filter size (n, n, n) with $n = 3, 5, 7, 9, 12, 15, 20$. To study the effect of the training/validation split, we considered a given snapshot (which we call s00001) of the 16-H_2O system as an example, where there are 735 pairs of τ^u and $\Delta\tau$ arrays. Three different training/validation splits of the data set were considered: (1) all pairs used for training and validation, (2) 80% pairs used for training and 20% pairs used for validation, and (3) 60% pairs used for training and 40% pairs used for validation. Note that case (1) does not give the same training and validation loss because minibatches of size 35 were used in training but all data were used in each validation. To evaluate the accuracy of the models, we trained for the models using snapshot s00001, and we used two other snapshots (i.e. s00003 and s00007) as test sets. None of these models have significant differences in the accuracy of reproducing the peak positions of the FF-BSE spectra. For all three split schemes, the average $\Delta\omega$ of the lowest-energy peak over different model architectures is -0.03 eV for s00001, -0.01 eV for s00003, and 0.00 eV for s00007. For the spectrum RMSE from ML-BSE using each split scheme, s00003 is 0.009 greater than s00001, and s00007 is 0.019 greater than s00001. These results suggest that the training/validation split within the range tested here does not significantly impact the accuracy or transferability of the ML models for predicting the absorption spectra. A subset of the dataset ($\tau^u_{vv'}$ and $\Delta\tau_{vv'}$ pairs) can be randomly selected and used as the training set in ML. The size of the subset does have an effect on the accuracy of the prediction. At least 10% of data are needed in this subset to have a converged f^{ML} value.

When computing f^{Avg}, we note that numerical errors may cause the absolute value of some elements in $\Delta\tau_{vv'}(\mathbf{r})/\tau^u_{vv'}(\mathbf{r})$ to be extremely large, and these outliers need to be discarded before computing f^{Avg}. No outliers in data need to be eliminated to carry out ML to obtain f^{ML}.

S1.5 Protocol to compute absorption spectra at finite temperature

Below we present the protocol to obtain absorption spectra at finite temperature:

1. Obtain representative snapshots of the target system at a finite temperature T using FPMD.
2. For one snapshot among those selected in Step 1, perform a FF-BSE calculation to obtain $\tau^u_{vv'}$ and $\Delta\tau_{vv'}$. Note that the selected snapshot may come, in some cases, from a smaller system of similar nature as the target system (see, e.g., the example of water presented in the main text).
3. Use the $\tau^u_{vv'}$ and $\Delta\tau_{vv'}$ saved in Step 2 as the training/validation sets to machine learn the mapping between unscreened and screened Coulomb integrals.
4. For all snapshots determined in Step 1, except for the one already evaluated in Step 2-3, compute the ML-BSE spectrum using the ML model trained in the previous step as a surrogate model for the calculation of screened Coulomb integrals.
5. Obtain the optical absorption spectrum at the finite temperature T by computing the average of the absorption spectra obtained in Step 4.

S2 Comparison between FF-BSE and ML-BSE absorption spectra

S2.1 Liquid water

To quantify the accuracy of the ML-BSE spectrum, we compare it with the FF-BSE spectrum and compute the change of the energy of individual peaks ($\Delta\omega = \omega^{\text{ML-BSE}} - \omega^{\text{FF-BSE}}$, where ω is the position of the peak in energy) and the root mean square error (RMSE) of the whole spectrum in a given energy range (the range is 0.0-27.2 eV for all systems except for the interfaces, which is 0.0-13.6 eV). For a representative snapshot of the 16-H_2O cell, when using a simple scaling factor model or a convolutional model of filter size $(7, 7, 7)$, we find that for the lowest-energy peak, $\Delta\omega = -0.03$ eV in both cases, and RMSE is 0.021 or 0.018 when the scaling factor or the convolutional model is used, respectively (Figure S1).This shows that, for the 16-H_2O system, the difference introduced by using different ML models (a convolutional model versus a global scaling factor) is negligible.

Figure S2 shows the sensitivity of the position of the first peak of the spectrum to the value of the global scaling factor. We used the 16-H_2O case as an example.

Figure S3(a) shows a comparison of the experimentally measured absorption spectrum of liquid water with the one computed for the 64-H_2O system using either FF-BSE or ML-BSE. Figure S3(b) shows the sensitivity of the spectrum to different choices of the scissor operator.

Figure S4 shows the variations of individual snapshots used to calculate the averaged spectrum of liquid water (64-H_2O system) in Figure 2. The same comparison for each individual snapshot is reported in Figure S5. The model used in ML-BSE is the global scaling factor from 16-H2O.

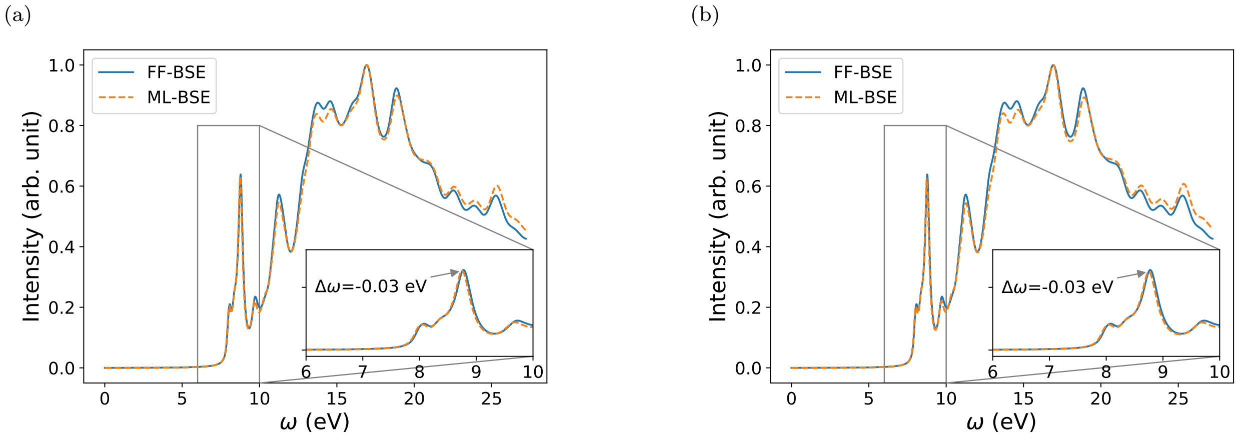


Figure S1: Accuracy of ML-BSE spectra of liquid water (16-H2O) obtained using (a) a convolutional model with filter size $(7, 7, 7)$, (b) a global scaling factor model. RMSE of the spectra is 0.018 for (a) and 0.021 for (b).

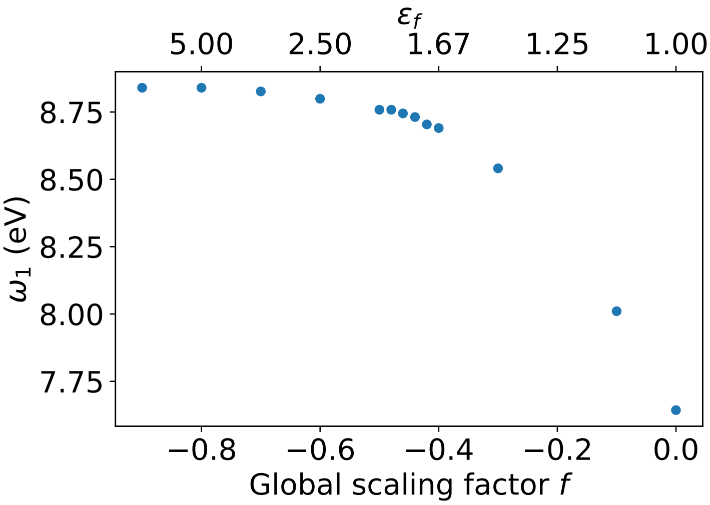


Figure S2: Sensitivity of the position of the lowest-energy peak (ω_1) of the computed absorption spectrum of water, obtained for a snapshot with 16 H_2O molecules), to the global scaling factor (f). $\epsilon_f = (1 + f)^{-1}$.

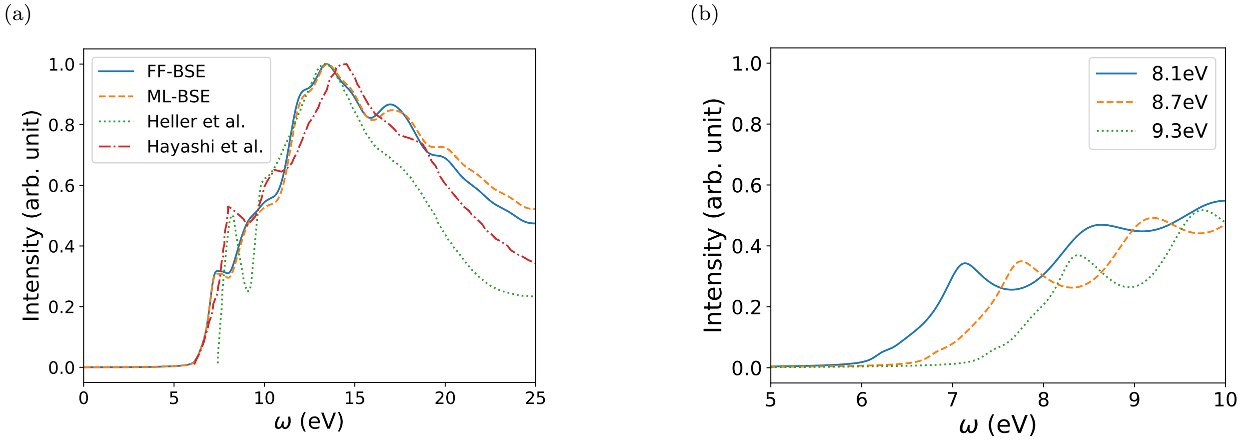


Figure S3: Absorption spectrum of liquid water (64-H_2O). (a) FF-BSE and ML-BSE are obtained computing and averaging the spectra of 10 snapshots. The position of the first peak of the experimental spectra from Heller et al.[17] and from Hayashi et al.[18] is located at 8.18 eV and 7.98 eV, respectively. The position of the first peak of the ML-BSE and FF-BSE spectra is located at 7.40 eV. (b) Sensitivity of the FF-BSE absorption spectrum of a chosen snapshot to the fundamental gap of the system. In (b), the values of the fundamental gap, obtained applying a scissor operator to the computed PBE band gap, were chosen based on the values of the experimental gap of water of 8.7±0.6 eV,[19] and the G_0W_0@PBE band gap of 8.1 eV.[12] For this particular snapshot, the energy of the first peak is 7.14 eV, 7.85 eV, and 8.47 eV, when the band gaps of the system is equal to 8.1 eV, 8.7 eV, and 9.3 eV, respectively.

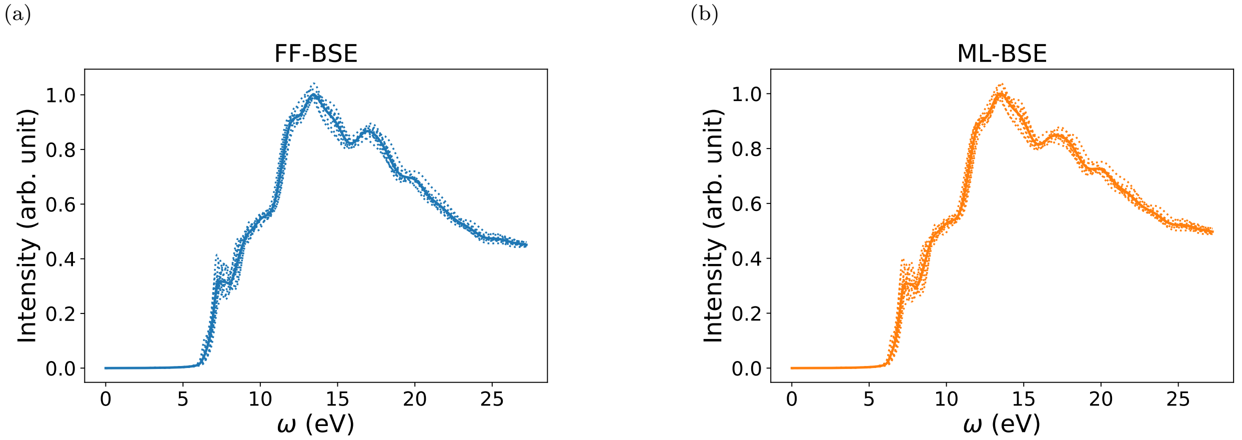


Figure S4: The absorption spectra of 10 individual snapshots of 64-H2O systems (dotted lines) and their averaged spectrum (solid line) from (a) FF-BSE and (b) ML-BSE calculations. The model used in ML-BSE is a global scaling factor obtained from simulations of a 16-water-molecule cell.

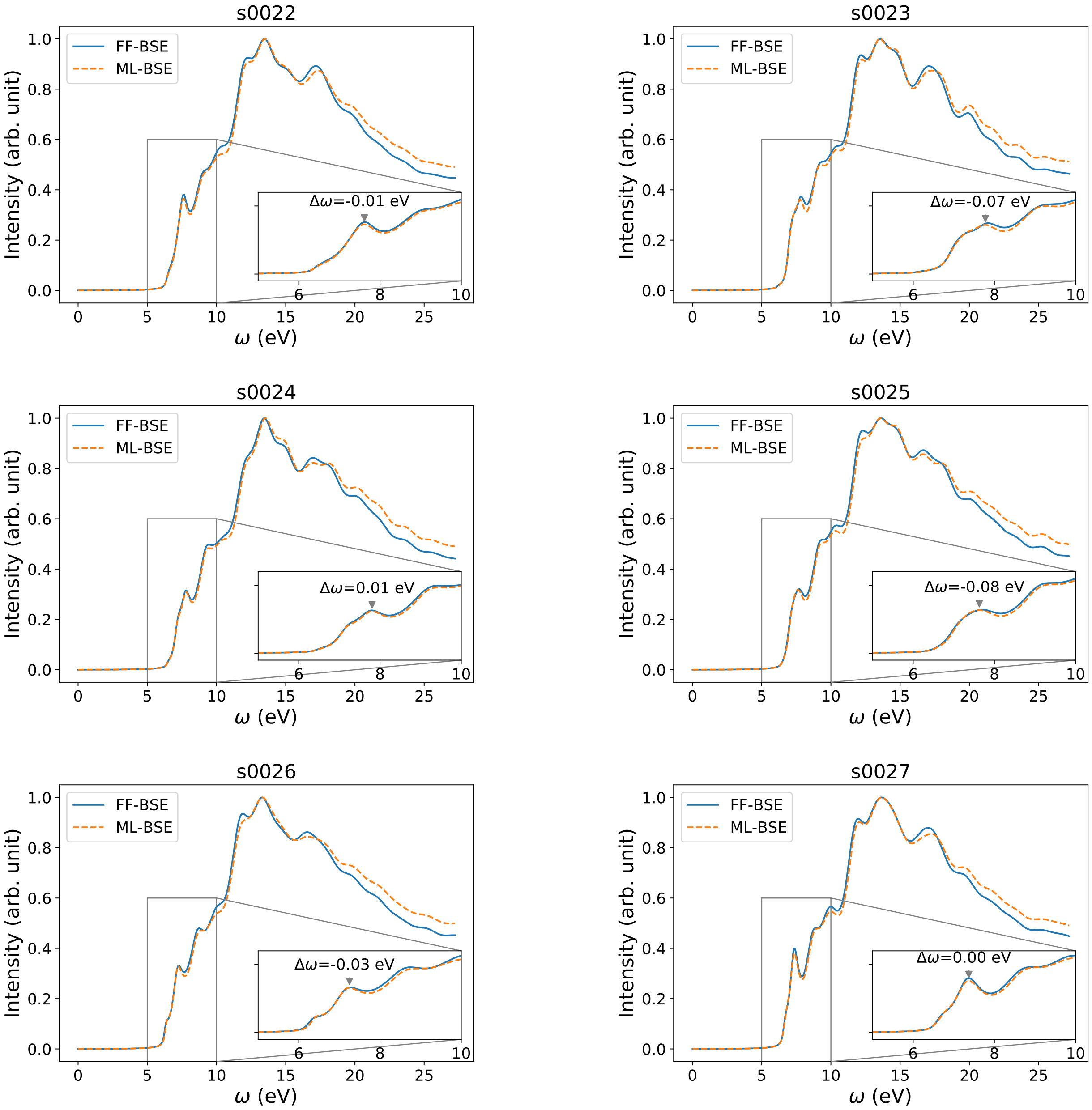


Figure S5: ML-BSE and FF-BSE spectra of 10 snapshots of the 64-H_2O system from FPMD trajectories at 400 K. The model used in ML-BSE is a global scaling factor obtained from simulations of a 16-water-molecule cell. The labels on the snapshots on top of each panel follows the labeling of snapshots of the PBE400 set (http://quantum-simulation.org/reference/h2o/pbe400/s32/index.htm).[6]

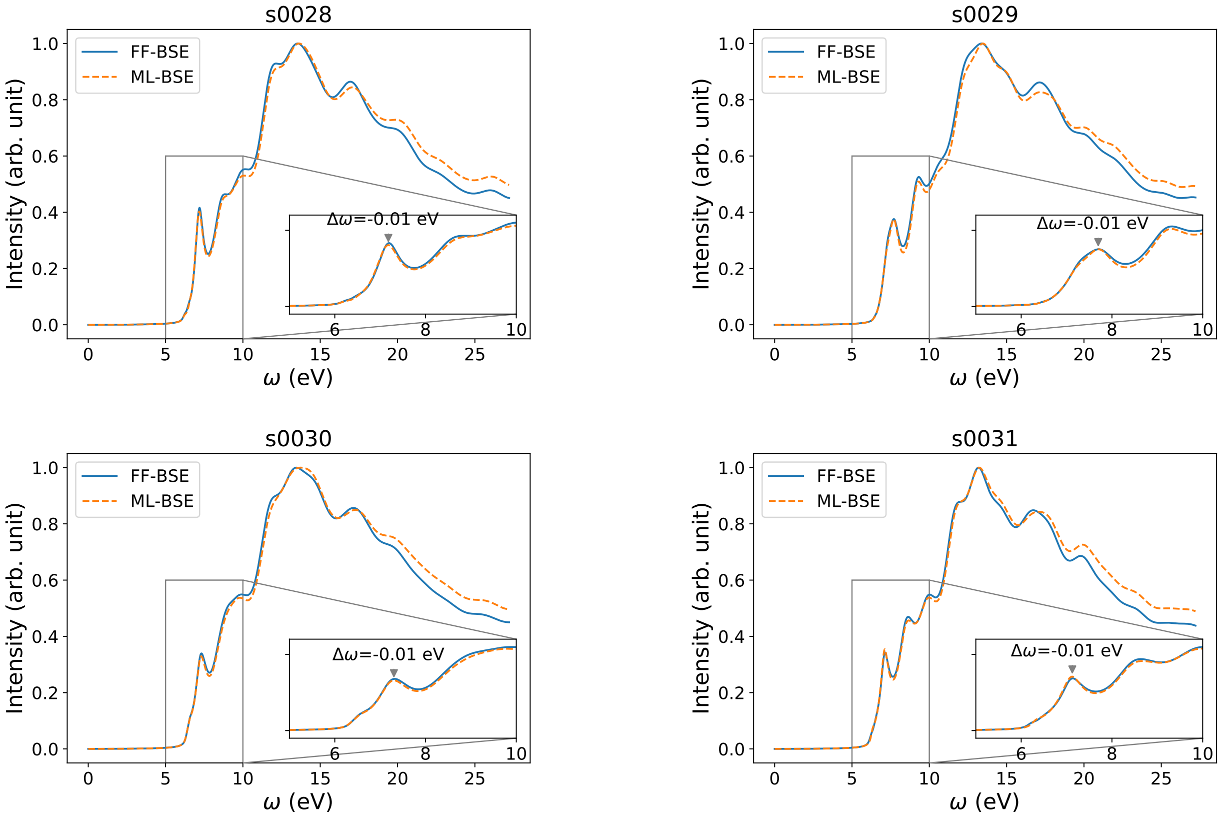


Figure S5: (Continued from the previous page) ML-BSE and FF-BSE spectra of 10 snapshots of the 64-H_2O system. The model used in ML-BSE is a global scaling factor obtained from simulations of a 16-water-molecule cell. The labels on the snapshots on top of each panel follows the labeling of snapshots of the PBE400 set (http://quantum-simulation.org/reference/h2o/pbe400/s32/index.htm).[6]

S2.2 Solids

Figures S6 and S7 show that a convolutional model has similar accuracy as a global scaling factor for Si and LiF. The comparisons between FF-BSE and ML-BSE spectrum for C (diamond), SiC, and MgO are reported in Figure S8.

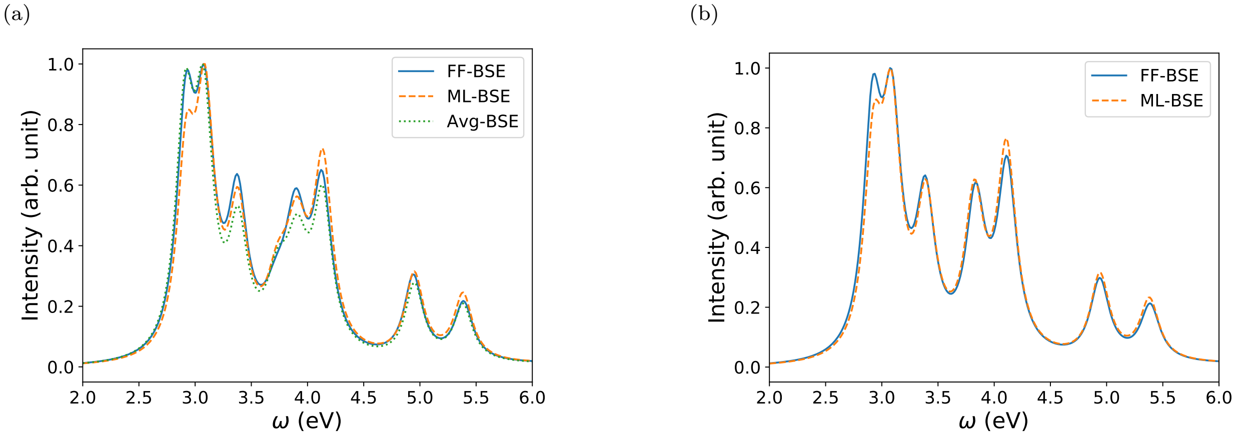


Figure S6: Accuracy of ML-BSE spectra of Si obtained using (a) a global scaling factor model and (b) a convolutional model (filter size $(7, 7, 7)$). Panel (a) also shows the BSE spectrum obtained by using a scaling factor computed from averaging $\Delta\tau/\tau^u$, labeled "Avg-BSE". The peak shifts of the ML-BSE spectrum from the FF-BSE spectrum are within 0.01 eV for all cases. The RMSE values between the ML-BSE and FF-BSE spectra are 0.019 for (a) and 0.015 for (b).

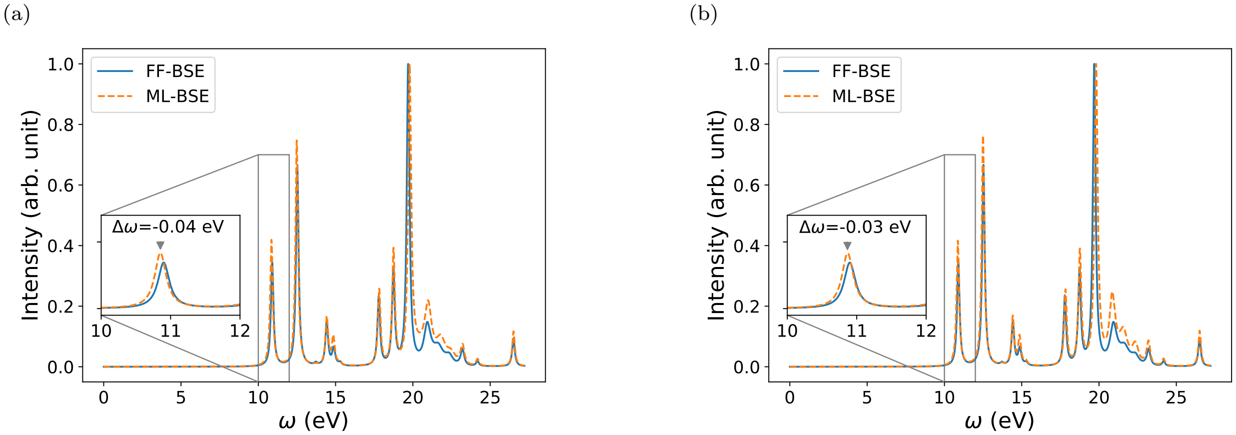
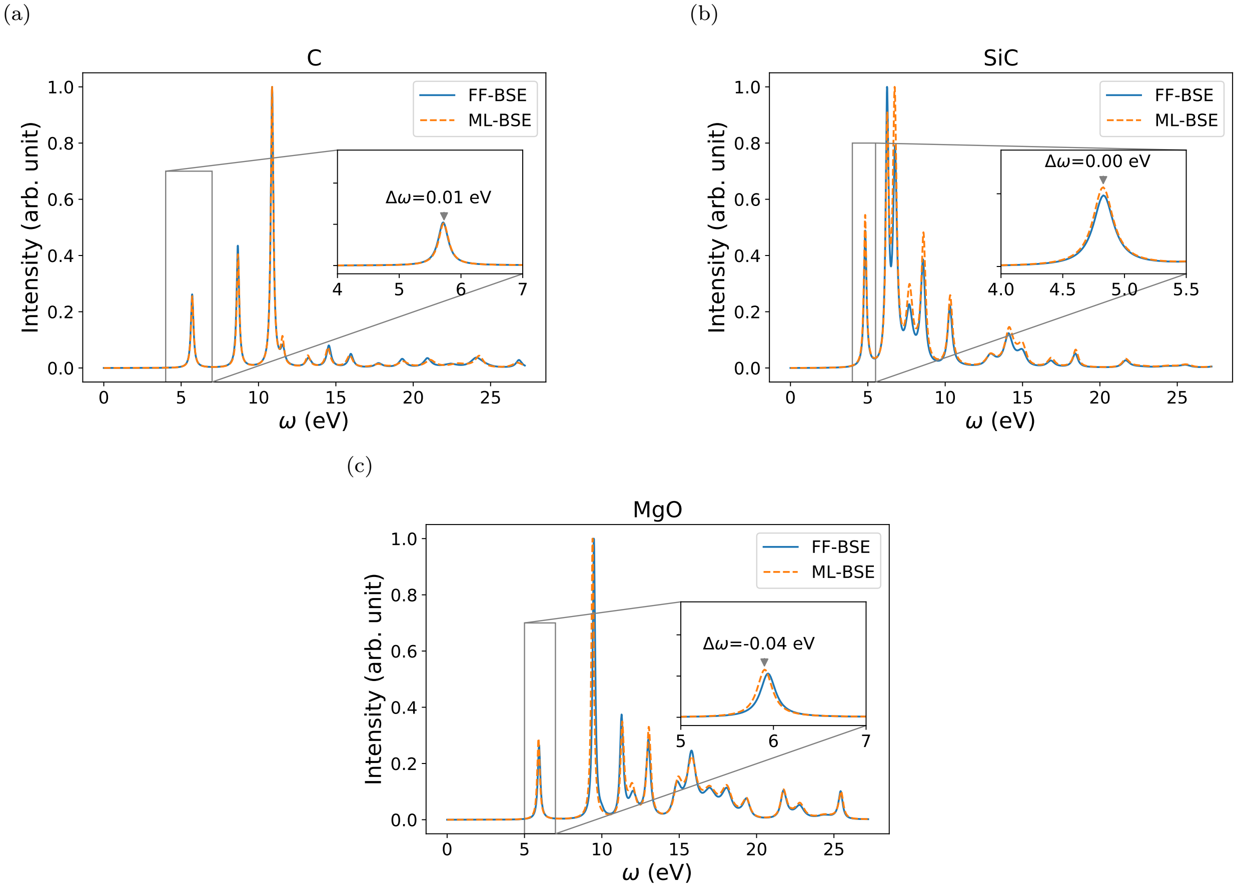


Figure S7: Accuracy of ML-BSE spectra of LiF obtained using (a) a global scaling factor model and (b) a convolutional model (filter size $(7, 7, 7)$). The RMSE values between the ML-BSE and FF-BSE spectra are 0.052 for (a) and 0.058 for (b).


(a)

(b)

(c)

Figure S8: Comparison between FF-BSE and ML-BSE spectrum of (a) diamond, (b) SiC, and (c)MgO. ML-BSE results are obtained from a global scaling factor of the respective system. The RMSE values between the ML-BSE and FF-BSE spectra are 0.005 for (a), 0.027 for (b), and 0.044 for (c).

Figure S9 shows the comparison between the absorption spectrum of Si computed with Yambo and WEST at the Γ point.

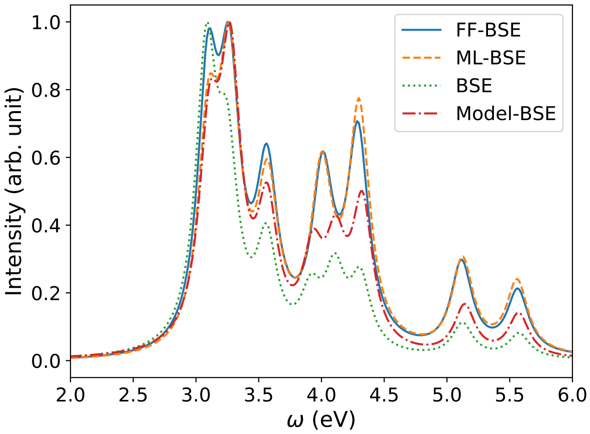


Figure S9: FF-BSE (WEST), ML-BSE (WEST) with a global scaling factor $f^{\mathrm{ML}} = -0.81$, BSE (Yambo), and Model-BSE (Yambo) with $f = -0.81$ for Si with a 64-atom supercell at the Γ point. The head of the dielectric matrix used in this figure is $\epsilon_\infty = 22.11$, and was obtained for the same cell at the Γ point in Yambo.

S2.3 H-Si/water interface (hydrophobic interface)

Comments on the kinetic energy cutoff

For the H-Si/water interface, we tested two different kinetic energy cutoff values for wavefunctions: 25 Ry and 60 Ry. From Figure S10(a), we conclude that 25 Ry is sufficient to obtain the low-energy peaks of the Si/water interface FF-BSE spectrum. Comparing Figure 5(a) and Figure S10(b), we conclude that using 25 Ry or 60 Ry will not change our conclusions for the Si/water interface. The $f_{p2}^{\mathrm{ML}}(z)$ models for the 25 Ry case and the 60 Ry case are very similar, with $\epsilon_f^{\mathrm{ML}}(z)$ being 2.28 and 2.95 for z in the respective regions of Si and water for the 25 Ry case, and 2.25 and 2.99 for z in the respective regions of Si and water for the 60 Ry case. Therefore, we used 25 Ry for all Si/water interfaces in this work.

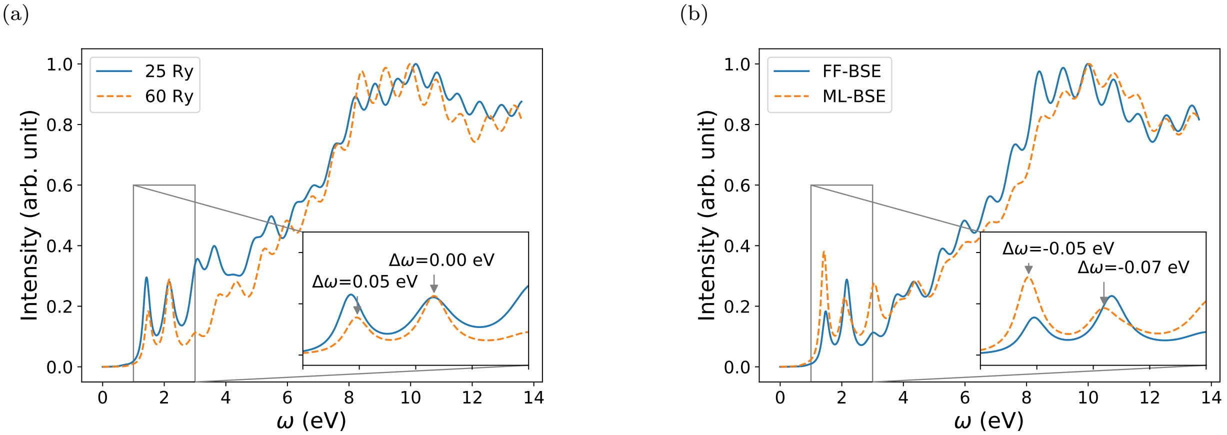


Figure S10: (a) FF-BSE spectrum from using a kinetic energy cutoff of 25 Ry or 60 Ry for wavefunctions. (b) Accuracy of ML-BSE using a $f_{p2}^{\mathrm{ML}}(z)$ model and a kinetic energy cutoff of 60 Ry for wavefunctions.

A position-dependent model $f(\mathbf{r})$ is necessary

From Figure S11, we can see that when using a model which treats Si and water on the same footing, using average scaling factors, results for the absorption spectrum are not accurate.

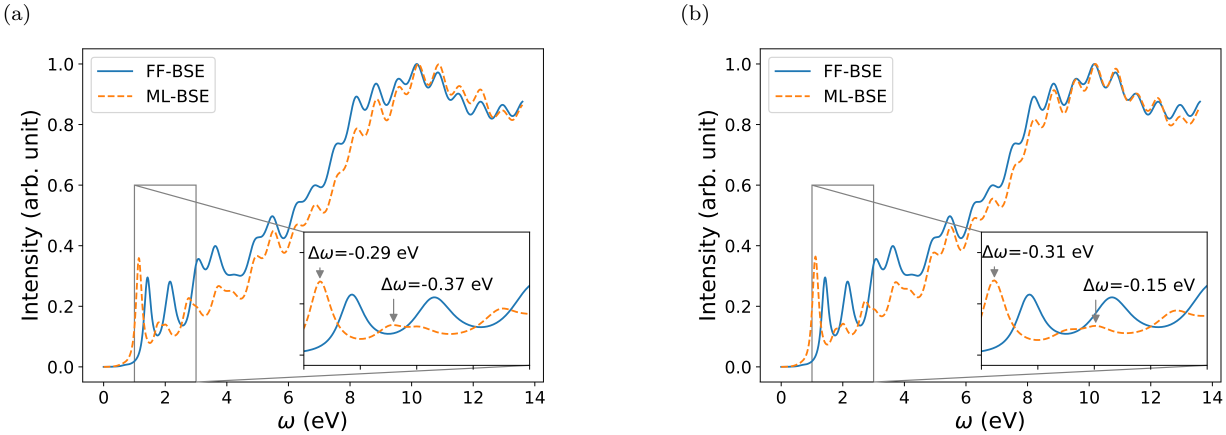


Figure S11: Accuracy of ML-BSE using (a) a global scaling factor model, and (b) a convolutional model (filter size $(7, 7, 7)$) for a snapshot representing a H-Si/water interface (see text). The RMSE values between the ML-BSE and FF-BSE spectra are 0.078 for (a) and 0.072 for (b).

Comparing Figure S11 and Figure S12, we find that a position-dependent model is necessary to obtain sufficiently accurate absorption spectra for the H-Si/water interface.

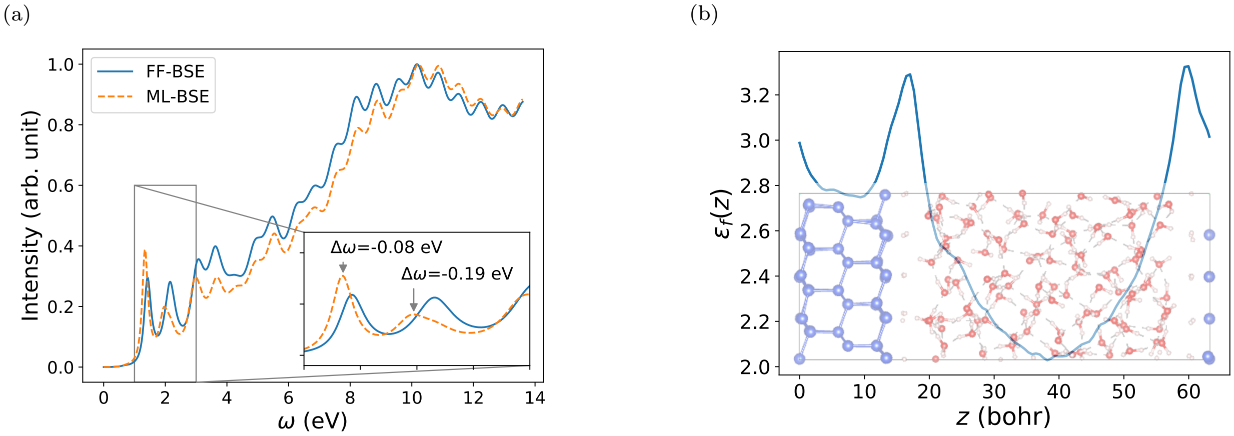


Figure S12: Accuracy of ML-BSE using (a) a position-dependent model with 108 parameters ($f^{\mathrm{ML}}_{p108}(z)$), and (b) $\epsilon^{\mathrm{ML}}_f(z)$ corresponding to the $f^{\mathrm{ML}}(z)$ profile used to compute (a), for a snapshot representing a H-Si/water interface (see text). The RMSE value between the ML-BSE and FF-BSE spectra is 0.059.

Interpretation of the $f(\mathbf{r})$ profile

In Figure S13 we show the charge density (a), $f^{\mathrm{Avg}}(\mathbf{r})$ (defined in the main text) (b), PDEP eigenpotential corresponding to the most negative eigenvalue (c), and the component of $f^{\mathrm{Avg}}(\mathbf{r})$ that corresponds to the most negative PDEP eigenvalue ($f^{\mathrm{Avg}}_1(\mathbf{r})$, defined in Eq. 11 of the main text) (d) for the H-Si/water interface. This shows that the maxima of $\epsilon^{\mathrm{ML}}_f(z)$ (Figure S12(b)) at the interfaces stem from the contribution of the PDEP eigenpotential with the most negative eigenvalue.

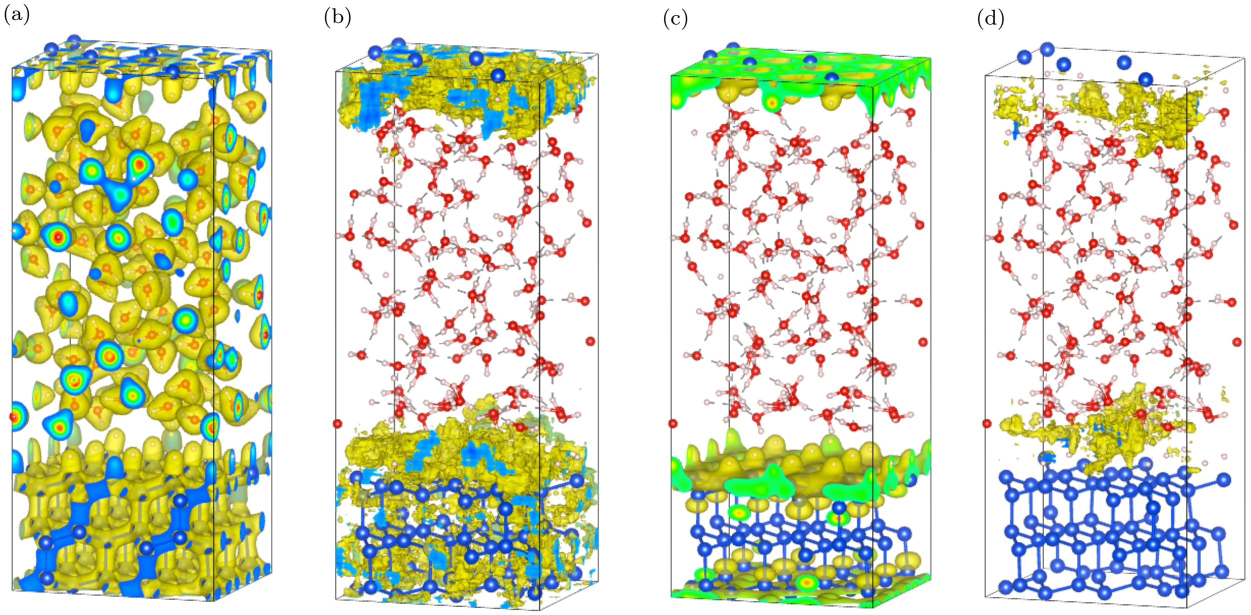


Figure S13: (a) Charge density, isovalue 0.050; (b) $f^{\mathrm{Avg}}(\mathbf{r})$ from averaging using a moving window of $8\times8\times8$, isovalue -0.644; (c) PDEP eigenpotential corresponding to the most negative eigenvalue, isovalue 3.64; (d) $f_1^{\mathrm{Avg}}(\mathbf{r})$ from averaging using a moving window of $8 \times 8 \times 8$, isovalue -0.370. The isosurface at the isovalue (i.e. value that has the greatest absolute value) is in yellow. The 3D visualization was rendered in VESTA (version 3.4.0).[20]

3D grid model from ML

In Figure S14 we show that a 3D grid model $f^{\mathrm{ML}}(\mathbf{r})$ yields an accurate ML-BSE spectrum for the H-Si/water interface. The 3D grid model has the same RMSE as the z-dependent model, and for the lowest-energy peak $\Delta\omega = -0.11$ eV, close to the -0.08 eV for the z-dependent model. The peak position from 3D grid model is slightly less accurate than the z-dependent model largely because the 3D grid model is more susceptible to the local variation in the training data.

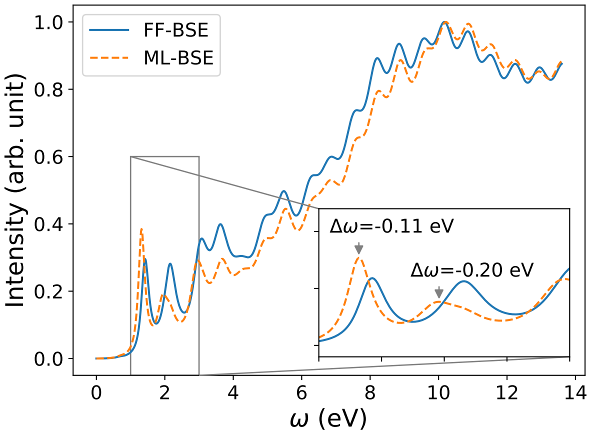


Figure S14: Accuracy of ML-BSE using the 3D grid model of the H-Si/water interface. Each sub-domain is a cube, and the side length of each sub-domain is 2.6 Å. The RMSE value between the ML-BSE and FF-BSE spectra is 0.059.

Transferability of position-dependent models

To test the transferrability of the $f^{\mathrm{ML}}(z)$ model across different snapshots, we consider a $f^{\mathrm{ML}}(z)$ model, $f^{\mathrm{ML}}_{p216}(z)$, more fine-grained than $f^{\mathrm{ML}}_{p2}(z)$ and $f^{\mathrm{ML}}_{p108}(z)$. The use of this more fine-grained model is to test whether overfitted $f^{\mathrm{ML}}(z)$ models can still be transferrable across different snapshots. As shown in Figure S15, the z-dependent model for the Si/water interface is transferable between different snapshots of the Si/water interface.

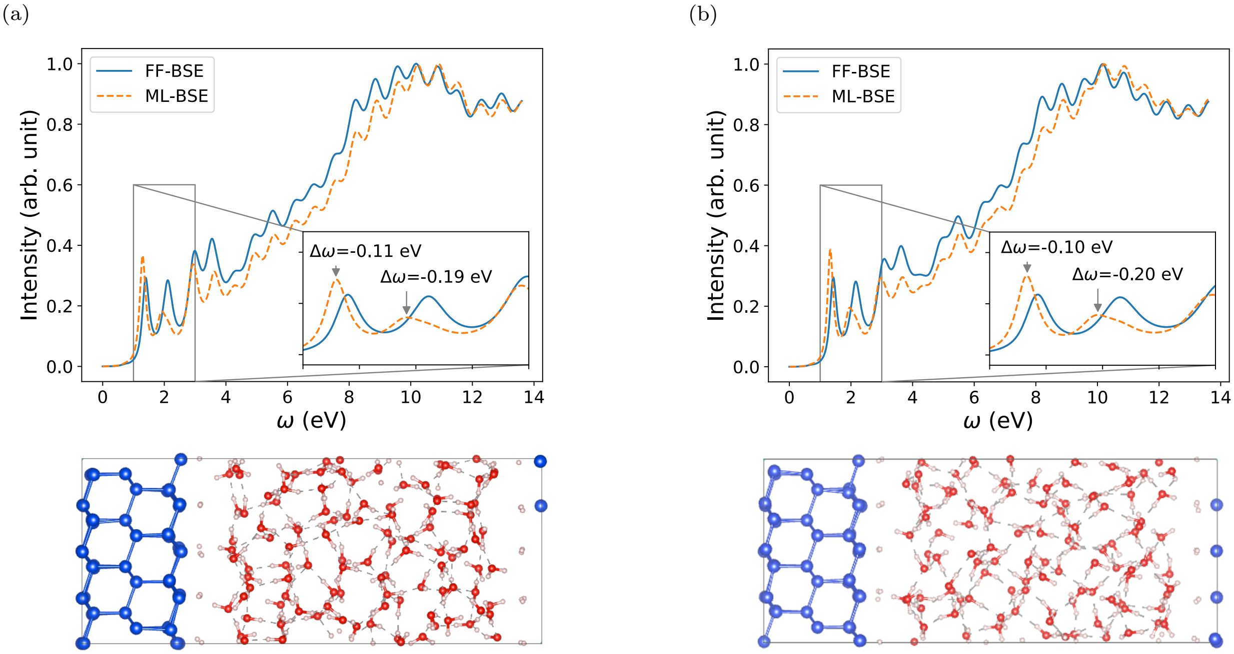


Figure S15: Accuracy of ML-BSE using the z-dependent model $f^{\mathrm{ML}}_{p216}(z)$ for the H-Si/water interface trained from a given snapshot (bottom of panel (b)) applied on a different snapshot (bottom of panel (a)). The spectrum corresponding to each snapshot is above its structural representation. The RMSE values is 0.064 for (a) and 0.059 for (b).

S2.4 COOH-Si/water interface (hydrophilic interface)

For the hydrophilic interface COOH-Si/water, we have found that the performance of the $f^{\mathrm{ML}}_{p108}(z)$ model is also similar to $f^{\mathrm{ML}}_{p2}(z)$ (Figure 5(b) of the main text), consistent with our results for the H-Si/water interface.

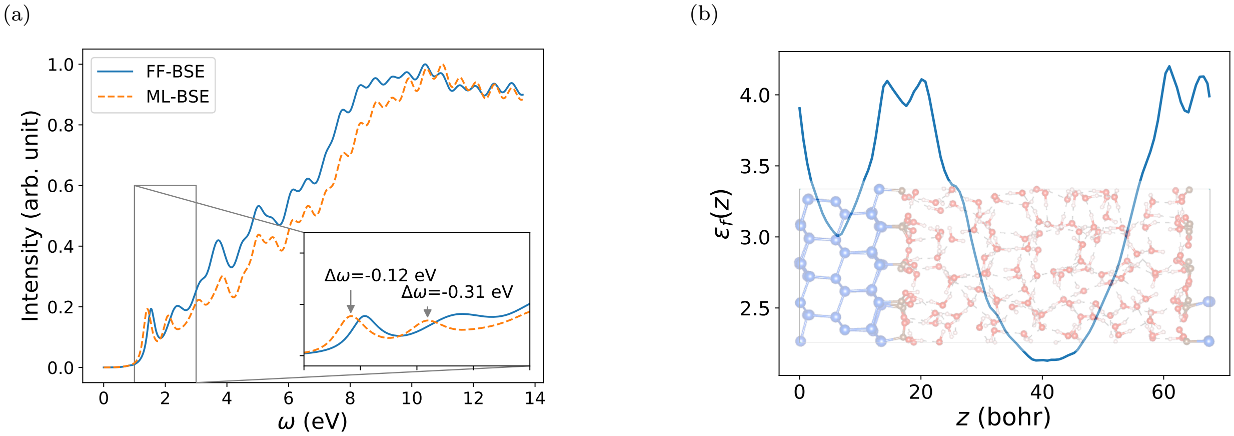


Figure S16: Accuracy of ML-BSE using (a) a position-dependent model with 108 parameters ($f^{\mathrm{ML}}_{p108}(z)$), and (b) $\epsilon^{\mathrm{ML}}_{f}(z)$ corresponding to the $f^{\mathrm{ML}}(z)$ profile used to compute (a), for a slab representing a COOH-Si terminated surface interfaced with water. The RMSE value between the ML-BSE and FF-BSE spectra is 0.078.

S2.5 Si clusters

For Si clusters, we were able to obtain a linear regression model (that results in a global scaling factor $f^{\mathrm{ML}} = -0.28$) and convolutional models in the same way as for homogeneous systems. However, the global scaling factor we obtained from linear regression does not yield accurate spectra, as shown in Figure 6 of the main article. In addition, a global scaling factor derived from averaging $\Delta\tau/\tau^{u}$, $f^{\mathrm{Avg}} = -0.54$, is different from from f^{ML} and yields an even worse spectrum , as shown in Figure S17. The value of the scaling factor may be different for simulation cells of different sizes , even if the absorption spectra have converged with respect to the size of the simulation cell. For a cubic simulation cell 30 (40) Å in length, $f^{\mathrm{ML}} = -0.28$ (-0.22).

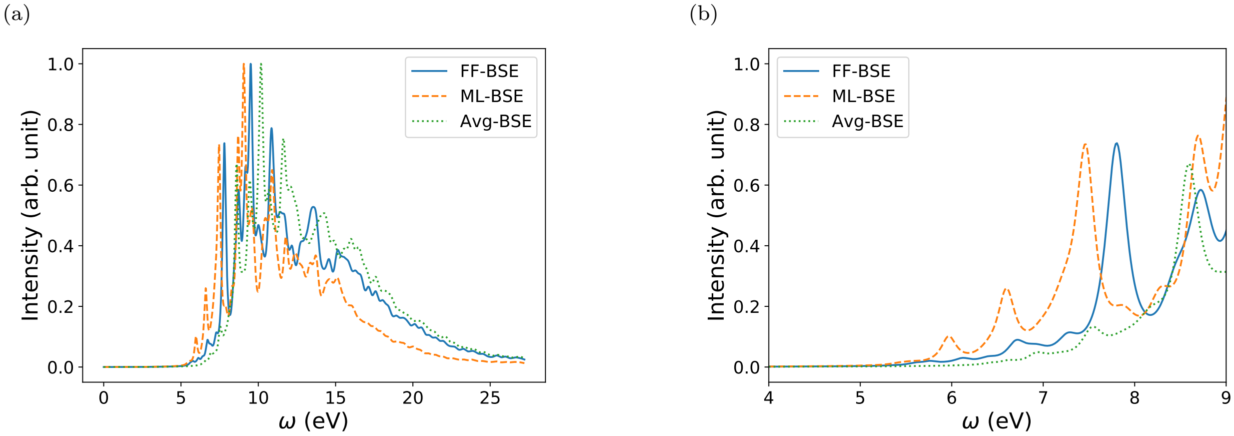


Figure S17: Comparison of ML-BSE and Avg-BSE to FF-BSE for $\mathrm{Si_{10}H_{16}}$ (30 Å cell). ML-BSE spectra are obtained using $f^{\mathrm{ML}} = -0.28$, and Avg-BSE is from using $f^{\mathrm{Avg}} = -0.54$. In (b) we show the same data as in (a) on a smaller energy interval (4 to 9 eV).

The reason behind this behavior is that, although the charge density of the Si cluster decays rapidly as a function of the distance from the cluster, the ratio $\Delta\tau/\tau^{u}$ does not. To have an ML model robust to the size of the simulation cell, we impose a threshold to the Si cluster data so that small elements (e.g. elements smaller than a chosen threshold) in $\tau^{u}_{vv'}$ are set to 0. Then we use this data to obtain our ML models.

The spectra obtained from various models are shown in Figures S18. Comparing Figures S18 and Figure 6 of the main article, it is clear that imposing a threshold on the data improves the accuracy of the spectra. However, a global scaling factor model still performs poorly compared to convolutional models. The convolutional models give highly accurate absorption spectra regardless of whether the data are treated with a threshold. This suggests that convolutional models are robust to sparse data.

In the previous two paragraphs, we applied the threshold on the training data, and applied the resulting ML models without imposing a threshold. Another way to consider the effect of the vacuum is to use the original data in training, but use a threshold on $\tau^u_{vv'}$ when the model is applied. Figure S19(a) and Figure 6(a) show that, for convolutional models obtained using the original, un-thresholded data, imposing a threshold when applying the model can improve the accuracy of the spectrum. Figure S19(b)(c) show that this will not improve much for the model based on scaling factors, even if a position-dependent model is used. This confirms that convolutional models are superior to models based on scaling factors.

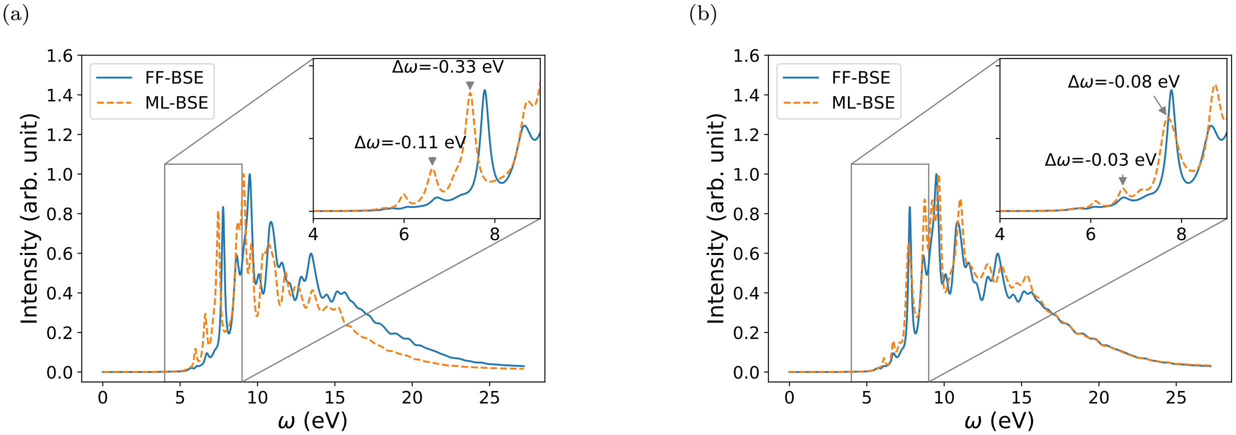


Figure S18: Accuracy of ML-BSE spectra of $Si_{10}H_{16}$ (40 Å cell) obtained using (a) a global scaling factor, and (b) a convolutional model (filter size $(7, 7, 7)$) from a smaller cell (30 Å cell). The models are obtained by considering only regions of $\tau^u_{vv'}$ above a charge density threshold (10^{-5} times the largest charge density). The RMSE value between the FF-BSE and ML-BSE spectra are 0.119 for (a) and 0.061 for (b).

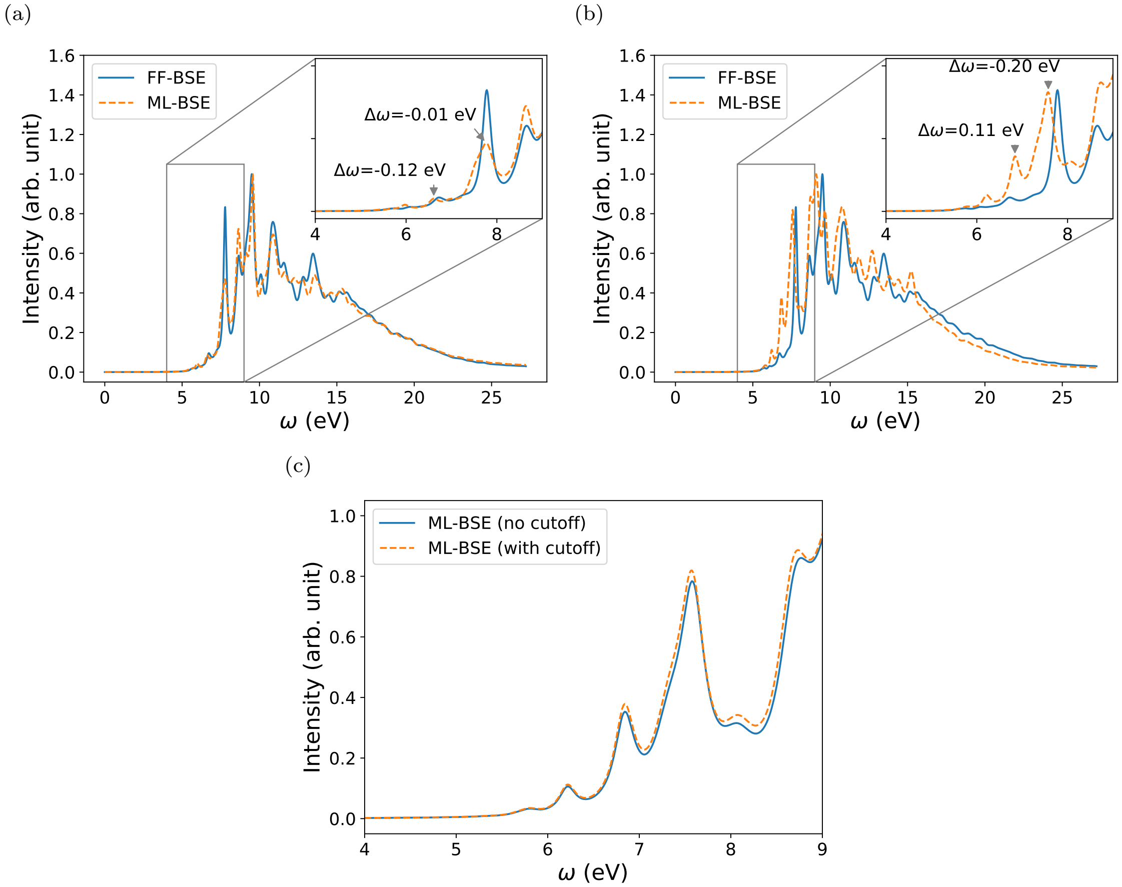


Figure S19: Accuracy of ML-BSE spectra of Si$_{10}$H$_{16}$ (40 Å cell) obtained from (a) a convolutional model (filter size $(7, 7, 7)$), and (b) a 3D grid model $f^{\mathrm{ML}}(\mathbf{r})$. When the models are applied to obtain ML-BSE spectra in (a) and (b), only regions of $\tau^{u}_{vv'}$ above a charge density threshold (10^{-5} times the largest charge density) are considered. The RMSE value between the FF-BSE and ML-BSE spectra are 0.107 for (a) and 0.044 for (b). In (c) we compare ML-BSE spectra from the 3D grid model when the charge density threshold is applied (with cutoff) or not applied (no cutoff).

Spectrum of Si$_{35}$H$_{36}$ at zero temperature

The 0 K spectrum of Si$_{35}$H$_{36}$ differs from the 500 K spectra, as shown in Figure S20 and Figure 7.

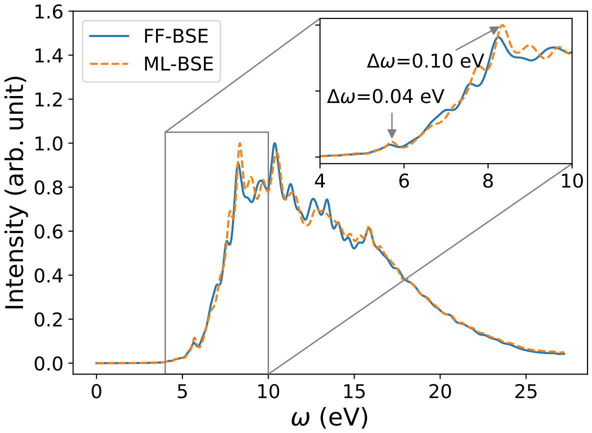


Figure S20: Accuracy of ML-BSE spectra obtained from a convolutional model (filter size (7, 7, 7)) for the 0 K geometrical configuration of $Si_{35}H_{36}$. The RMSE value between the ML-BSE and FF-BSE spectra is 0.029.

Finite temperature spectra of $Si_{35}H_{36}$

The variations of individual spectra used to calculate the averaged spectrum of $Si_{35}H_{36}$ are shown in Figure S21. The comparison between FF-BSE and ML-BSE spectra of each snapshot is in Figure S22. The model used in ML-BSE uses a convolutional model (7, 7, 7) trained on data obtained for the 0 K snapshot, the same model used in Figure S20.

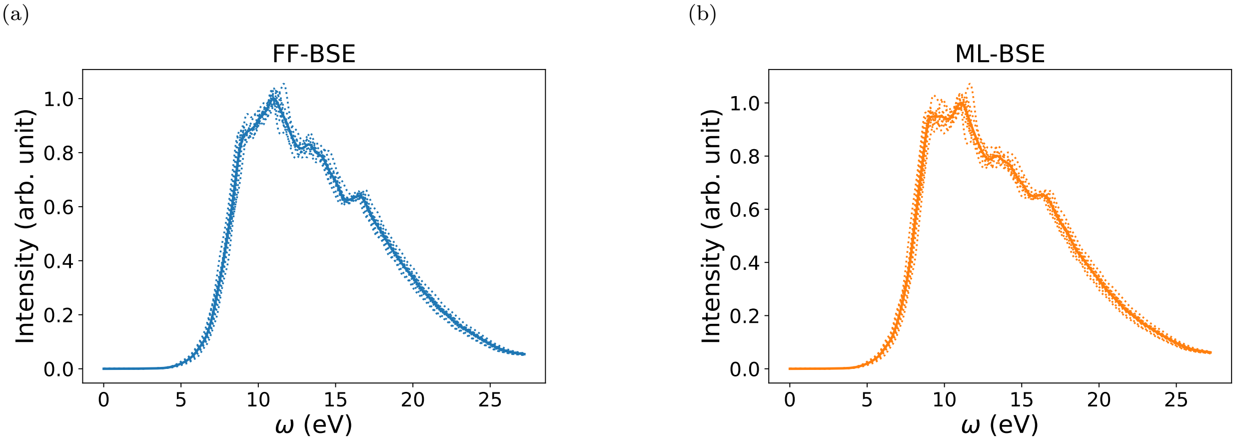


Figure S21: The absorption spectra of 10 individual snapshots of $Si_{35}H_{36}$ (dotted lines) and their averaged spectrum (solid line) from (a) FF-BSE and (b) ML-BSE calculations.

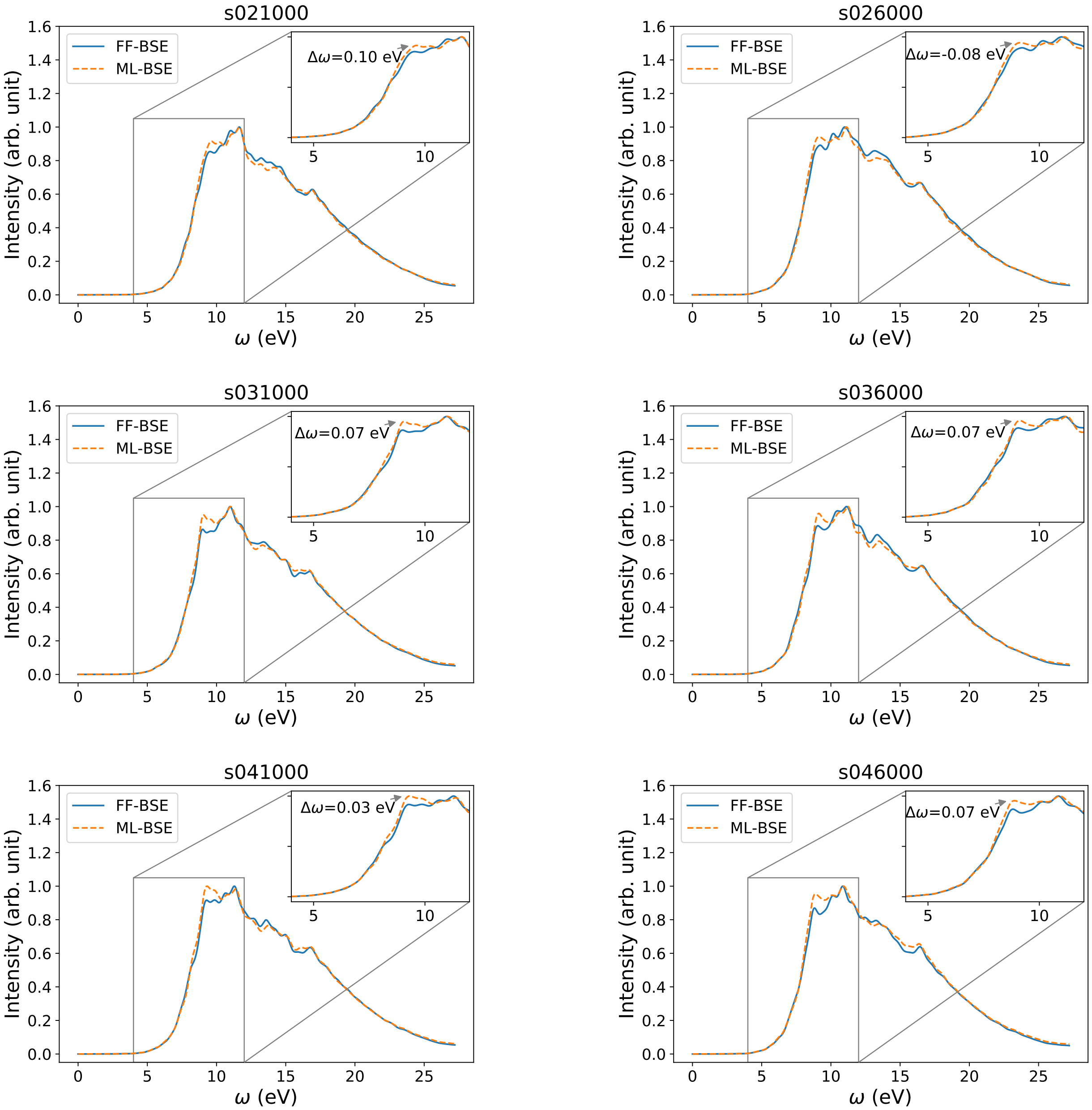


Figure S22: ML-BSE and FF-BSE spectra of 10 snapshots of $Si_{35}H_{36}$ snapshots extracted from FPMD trajectories at 500 K. The model used in ML-BSE obtained using a convolutional model $(7, 7, 7)$ trained with data obtained for the 0 K snapshot, the same model used in Figure S20. The 10 snapshots were extracted every 5000 time steps from the FPMD trajectory, starting from the 21000th time step (10.16 ps onwards). We label these snapshots from s021000 to s066000.

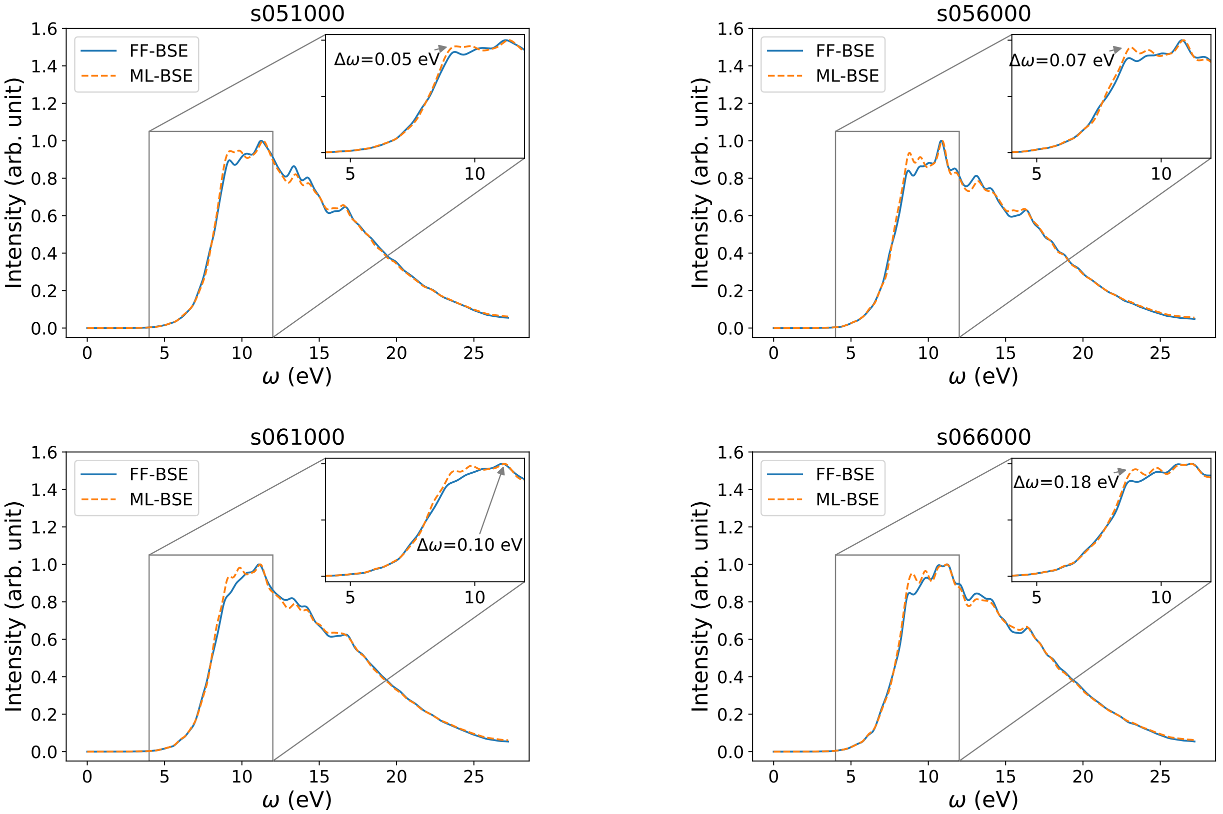


Figure S22: (Continued from the previous page) ML-BSE and FF-BSE spectra of 10 snapshots of $Si_{35}H_{36}$ snapshots extracted from FPMD trajectories at 500 K. The model used in ML-BSE obtained using a convolutional model $(7, 7, 7)$ trained with data obtained for the 0 K snapshot, the same model used in Figure S20. The 10 snapshots were extracted every 5000 time steps from the FPMD trajectory, starting from the 21000th time step (10.16 ps onwards). We label these snapshots from s021000 to s066000.

Transferability of ML models for $Si_{35}H_{36}$

As shown in Figure S22 (snapshot s041000) and Figure S23, the accuracy of the convolutional model derived from data for the 0 K and 500 K snapshots is similar in predicting the absorption spectrum of the 500 K snapshot. This suggests that the convolutional model is indeed transferable from the 0 K geometry to 500 K geometries for this Si cluster.

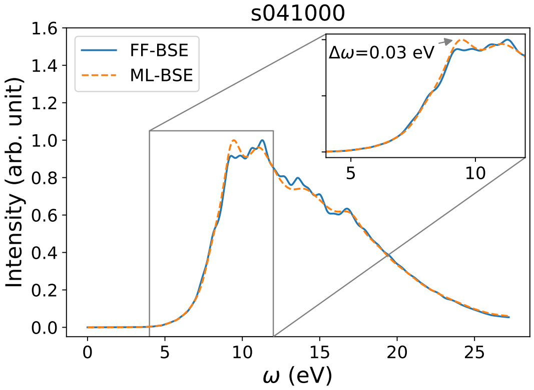


Figure S23: Accuracy of the ML-BSE spectrum of $Si_{35}H_{36}$ (a 500 K snapshot s041000) obtained by using a convolutional model $(7, 7, 7)$ derived from the data for snapshot 041000. The RMSE value between the ML-BSE and FF-BSE spectra in this figure is 0.019, to be compared with RMSE=0.020 for the same snapshot's spectrum in Figure S22.

$\mathbf{Si_{87}H_{76}}$

The ML-BSE spectrum of $Si_{87}H_{76}$ obtained by applying various models obtained from $Si_{87}H_{76}$ is similar to the one obtained by applying the convolutional model obtained from $Si_{35}H_{36}$, as shown in Figure S24 and Figure 8.

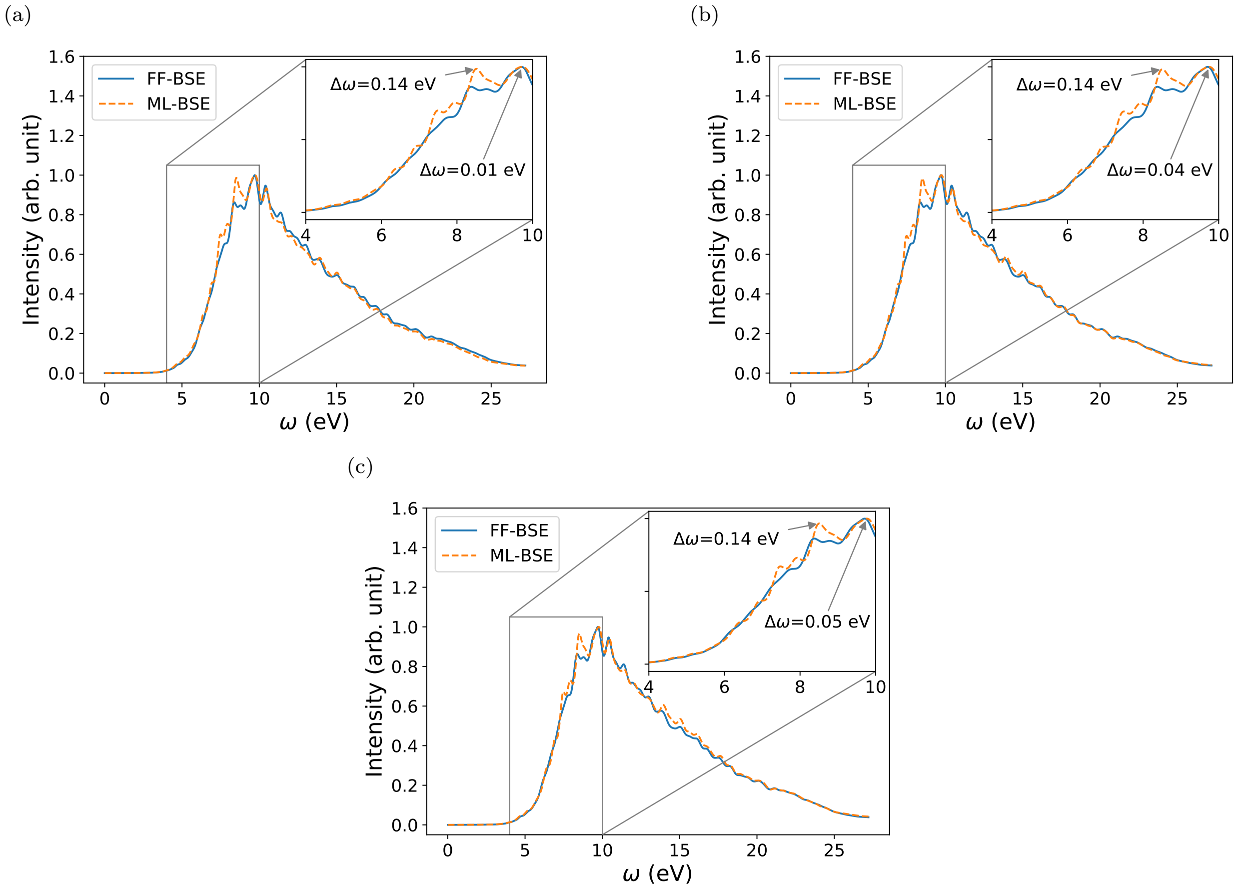


Figure S24: Accuracy of the ML-BSE spectrum of $Si_{87}H_{76}$ obtained by using (a) a global scaling factor, (b) a convolutional model $(5, 5, 5)$, and (c) a convolutional model $(7, 7, 7)$. The RMSE of the spectra are 0.025 for (a), 0.024 for (b), and 0.020 for (c).

Computational time savings

As demonstrated in Figure S25, the greater the simulation cell is (the larger the number of plane waves n_{pw} is), the more substantial are the savings in the computational time.

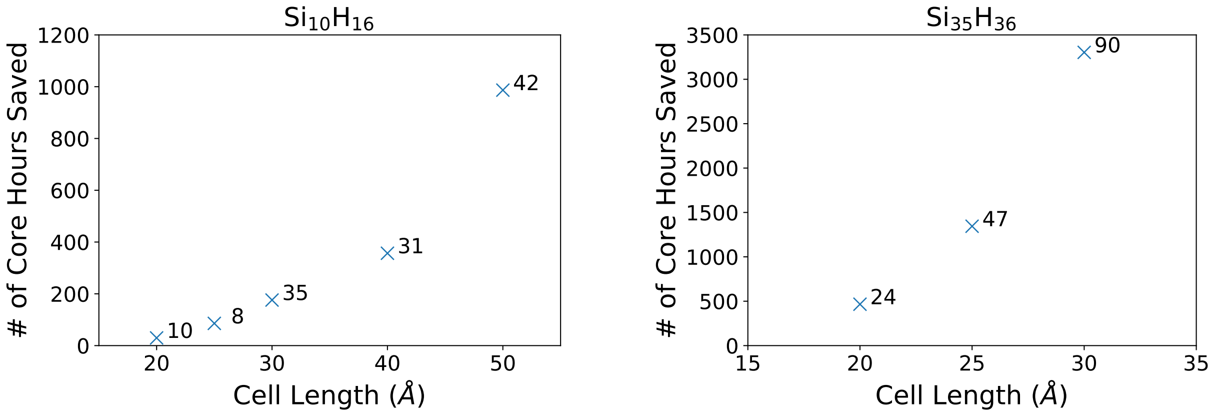


Figure S25: Number of core hours saved for the $Si_{10}H_{16}$ and $Si_{35}H_{36}$ clusters as a function of the size of the simulation cell when using ML-BSE instead of FF-BSE. The numerical values beside each data point is the corresponding α_d. Because of the relatively small computational cost of ML-BSE for $Si_{10}H_{16}$, its α_d is more sensitive to the machine status than that of $Si_{35}H_{36}$ and is not necessarily monotonic as a function of the size of the simulation cell.

S3 Timing and scaling of ML-BSE

The calculation of absorption spectra can be decomposed into the following three steps:

1. Calculation of unscreened integrals using Eq. 8;
2. Calculation of screened integrals using Eq. 9 (in FF-BSE) or Eq.10 (in ML-BSE);
3. Calculation of the frequency-dependent spectrum using Eq. 1.

Here we have replaced Step 2 with a machine learning model. For the largest system considered in this work (COOH-Si/water interface), at least ~50% of the total computational time is spent in Step 2. We define the speed-up $\alpha_d = t_d^{\text{FF-BSE}}/t_d^{\text{ML-BSE}}$ as the ratio of the core hours needed to carry out Step 2 with FF-BSE or ML-BSE. We found that α_d increases as n_{int} and the number of plane waves n_{pw} increase.

References

[1] M. Schlipf and F. Gygi, *Computer Physics Communications*, 2015, **196**, 36–44.

[2] J. P. Perdew, K. Burke and M. Ernzerhof, *Physical Review Letters*, 1996, **77**, 3865–3868.

[3] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. d. Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari and R. M. Wentzcovitch, *Journal of Physics: Condensed Matter*, 2009, **21**, 395502.

[4] F. Gygi, *IBM Journal of Research and Development*, 2008, **52**, 137–144.

[5] M. Stengel and N. A. Spaldin, *Physical Review B*, 2006, **73**, 075121.

[6] W. Dawson and F. Gygi, *The Journal of Chemical Physics*, 2018, **148**, 124501.

[7] G. Bussi, D. Donadio and M. Parrinello, *The Journal of Chemical Physics*, 2007, **126**, 014101.

[8] T. A. Pham, D. Lee, E. Schwegler and G. Galli, *Journal of the American Chemical Society*, 2014, **136**, 17071–17077.

[9] M. Govoni and G. Galli, *Journal of Chemical Theory and Computation*, 2015, **11**, 2680–2696.

[10] N. L. Nguyen, H. Ma, M. Govoni, F. Gygi and G. Galli, *Physical Review Letters*, 2019, **122**, 237402.

[11] M. Shishkin and G. Kresse, *Physical Review B*, 2007, **75**, 235102.

[12] T. A. Pham, C. Zhang, E. Schwegler and G. Galli, *Physical Review B*, 2014, **89**, 060202.

[13] A. Marini, C. Hogan, M. Grüning and D. Varsano, *Computer Physics Communications*, 2009, **180**, 1392–1403.

[14] D. Sangalli, A. Ferretti, H. Miranda, C. Attaccalite, I. Marri, E. Cannuccia, P. Melo, M. Marsili, F. Paleari, A. Marrazzo *et al.*, *Journal of Physics: Condensed Matter*, 2019, **31**, 325902.

[15] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, `https://www.tensorflow.org/`, Software available from tensorflow.org.

[16] D. P. Kingma and J. Ba, *arXiv preprint arXiv:1412.6980*, 2014.

[17] J. M. Heller, R. N. Hamm, R. D. Birkhoff and L. R. Painter, *The Journal of Chemical Physics*, 1974, **60**, 3483–3486.

[18] H. Hayashi, N. Watanabe, Y. Udagawa and C.-C. Kao, *Proceedings of the National Academy of Sciences*, 2000, **97**, 6264–6266.

[19] A. Bernas, C. Ferradini and J.-P. Jay-Gerin, *Chemical physics*, 1997, **222**, 151–160.

[20] K. Momma and F. Izumi, *Journal of Applied Crystallography*, 2011, **44**, 1272–1276.