Supporting Information

Jing Yang,* Stefano Falletta, and Alfredo Pasquarello

Chaire de Simulation à l'Echelle Atomique (CSEA), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

E-mail: jing.yang@epfl.ch

Optimizing probe width σ

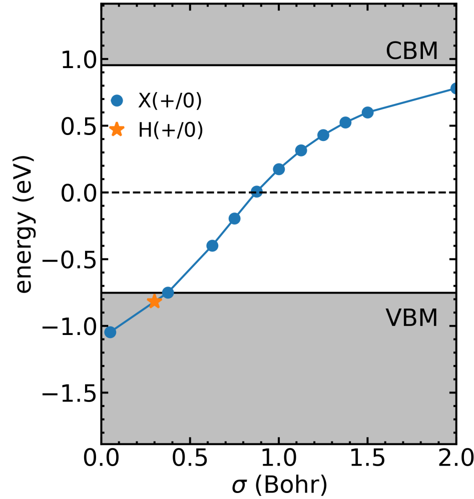

Figure S1: Transition level of the local probe as a function of its Gaussian charge distribution width σ compared to hydrogen interstitial (starred) in GaP. Note that here the single-particle energy levels are not finite-size corrected.

The probe used in this work follows a local potential

$$V_{\mathrm{loc}}(r) = -\frac{1}{r}\mathrm{erf}\left(\frac{r}{\sqrt{2}\sigma}\right), \tag{1}$$

where σ determines the width of the Gaussian charge distribution. It has been shown that increasing σ shifts the defect level up monotonously.[?] In this work, we select the optimized σ based on the level of hybridization between the defect state and the valence band. In most of the cases, this is sought by setting the not-finite-size-corrected D(0/+) transition level calculated at PBE level in the middle of the band gap. Figure S1 demonstrates this selection of σ in GaP. While the hydrogen interstitial transition level falls into the valence band, with optimized $\sigma = 0.86$ Bohr, the X(0/+) level sits right in the middle of the band gap. This optimization process allows us to achieve localized state in materials where physical defects fail to work.

In most of materials surveyed in this work, we are able to find σ that sets the transition level right in the middle of the band gap. However, in some cases, the H(0/+) transition level already sits above midgap and tuning σ fails to bring it to the middle. In these cases, a minimum σ of 0.05 Bohr is used. Another exception is that for ZnO, we found stronger interaction between the probe state and the band states with larger σ. As such, we selected the minimum σ that creates an in-gap defect state, which is 0.25 Bohr. As we have discussed in the manuscript, as far as a well-localized state is achieved, the value of σ and the specific defect type hardly affect the corresponding α_K value sought.

The case of silicon

In most of the materials we surveyed, the perturbative one-shot method produces single-particle energy levels quantitatively similar to the full self-consistently method. The rationale here is that the localized defect wave function at PBE0 level does not differ too much from the PBE one. Silicon provides an exception in the list of materials studied in this work. Figure S2 shows the occupied and unoccupied levels in the gap with varying α in silicon. The hydrogen interstitial level lies close to the band edge and hybridizes strongly with the valence band (Figure S2a). The close similarity between PBE and PBE0 defect wave function

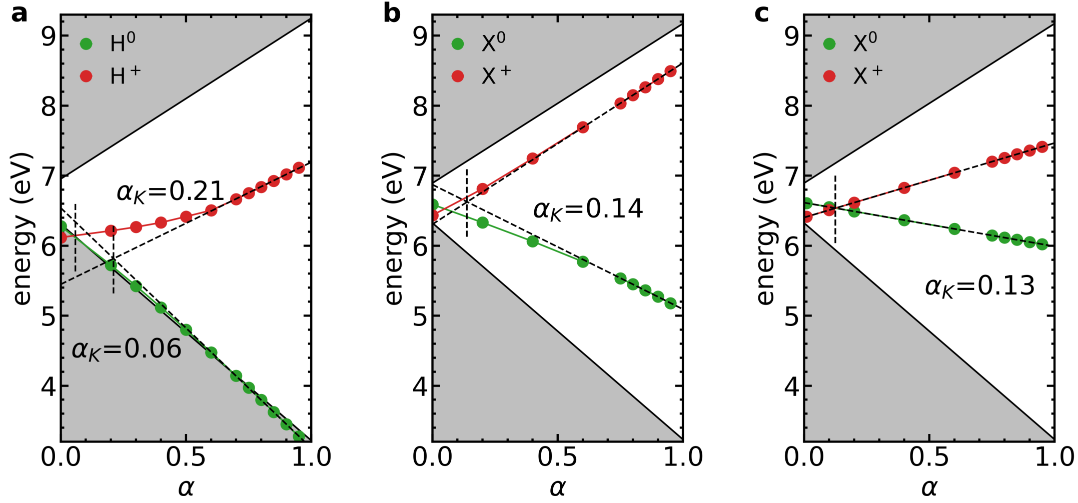

Figure S2: Occupied and unoccupied defect levels in Si with varying α for (a) hydrogen interstitial with full PBE0(α) calculation, (b) optimized probe with full PBE0(α) calculation, and (c) optimized probe with one-shot PBE0(α) calculation.

is broken in this case. As a result, we observe that the defect level versus α slope changes from the small-α regime to the large-α regime, which results in two different α_Ks. Here α_K = 0.06 corresponds to a band gap of 0.84 eV while α_K = 0.21 corresponds to a band gap of 1.54 eV, both deviating from the experimental value (1.23 eV). This is one of the cases where the hydrogen interstitial method fails.

Replacing the hydrogen interstitial with the optimized probe brings the defect levels into the gap and reduces the change in slope as α increases. However, there is still a non-negligible change in the resulted α_K (Figure S2b). Extrapolating from the high α regime gives α_K = 0.14, with a band gap of 1.21 eV.

In the one-shot method (Figure S2c), PBE wave function is retained throughout and the linearity is kept across the range of α. The obtained α_K is very close to that of the full calculation in the high-α regime and both predict accurate band gaps. In other words, even though the PBE wave funtion used in the one-shot method does not reproduce exactly the defect state at high α_K, the resulting α_K values are close. Out of the nineteen materials studied in this work, we only observe this behavior in Si and TiO_2. Whether this agreement

between full and one-shot calculations is coincidental or not remains an open question.

The evolution of defect wave function with α is shown in Figure S3. Here we can directly observe that, the H^0 wave function hybridizes with the valence band and is spread across the cell, whereas H^+ level moves away from the band edge as α increases and tends to localize at high α. This explains the non-linearity of H^+ level versus α in Figure S2a. In contrast, for the optimized probe, the change in the defect wave function is much less pronounced as a localized in-gap state is achieved (Figure S3c, d).

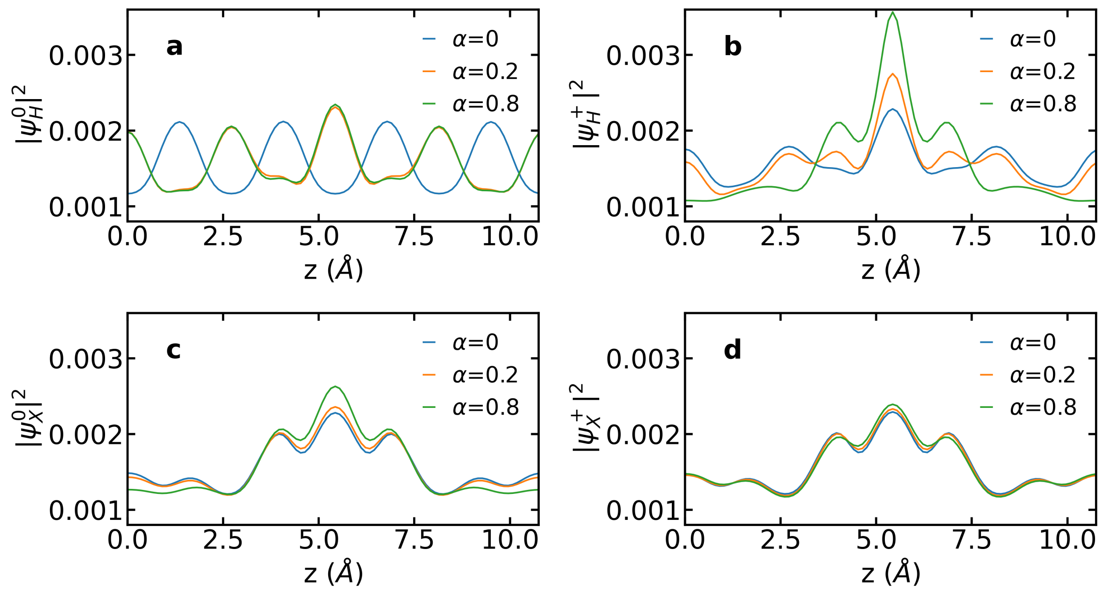

Figure S3: Wavefunction densities ($|\psi_D|^2$) averaged on the xy plane at Γ-point for (a) the highest occupied band of H^0, (b) the highest occupied band of H^+, (c) the lowest unoccupied band of X^0, and (d) the lowest unoccupied band of X^+ in Si.

Band gaps predicted by one-shot method

Table S1: Band gaps (in eV) and obtained α_K using hydrogen interstitial defect and potential probe compared with QSG$\hat{\text{W}}$ and experimental values for the one-shot method. Zero-phonon renormalizaton (ZPR) corrected experimental band gaps are taken from ref **?** and used as references for calculating the mean absolute error (MAE). QSG$\hat{\text{W}}$ are taken from ref **?** . For the validity of comparison, MAEs on the same row only include materials for which data from both methods are available. Since the hydrogen interstitial method is only valid for three $3d$ materials, it is excluded from the $3d$ MAE and total MAE comparison.

	H(0/+)	X(0/+)			
	$E_g(\alpha_K)$	σ (bohr)	$E_g(\alpha_K)$	QSG$\hat{\text{W}}$	Expt + ZPR
sp semiconductors					
AlAs	2.03(0.12)	0.92	1.99(0.11)	2.39	2.28
AlP	2.21(0.12)	0.20	2.27(0.13)	2.56	2.54
Ar	13.70(0.51)	0.53	14.42(0.58)	14.23	14.33
BN	6.51(0.24)	0.05	6.51(0.24)	6.59	6.74
C	5.73(0.21)	0.14	5.72(0.21)	5.83	5.85
CaO	6.31(0.28)	0.15	6.35(0.28)	6.79	7.09
LiCl	9.33(0.34)	0.31	9.51(0.36)	9.87	9.57
LiF	15.25(0.48)	0.18	15.37(0.49)	15.52	15.35
MgO	8.20(0.33)	0.05	8.21(0.33)	8.37	8.36
Si	—	1.23	1.15(0.12)	1.27	1.23
SiC	2.52(0.19)	0.05	2.49(0.18)	2.56	2.52
MAE	0.28		0.20	0.12	
$3d$ materials					
GaN	—	0.29	3.35(0.19)	3.61	3.67
GaP	—	0.86	2.37(0.15)	2.41	2.43
GaAs	—	0.67	0.99(0.09)	1.58	1.57
ZnO	—	0.25	3.08(0.24)	3.41	3.60
InP	—	0.78	1.13(0.09)	1.42	1.47
ZnS	3.39(0.17)	0.30	3.52(0.19)	3.63	3.94
ZnSe	1.81(0.11)	0.07	1.91(0.13)	2.75	2.87
TiO_2	3.19(0.14)	0.05	3.21(0.14)	—	3.30
MAE			0.46	0.11	
MAE			0.29	0.11	

Band gaps predicted by the CAM functional

Table S2: Band gaps (in eV) and obtained α_K for PBE0 and CAM functionals compared to HSE06 and experimental values. Dielectric constants, HSE06-calculated band gaps, and ZPR corrected experimental band gaps are taken from ref **?** .

	PBE0(α_K)	CAM($\alpha_{s,K}$,$\alpha_l = 1/\epsilon, \mu_{\text{HSE}}$)				
	$E_g(\alpha_K)$	ϵ	α_l	$E_g(\alpha_{s,K})$	HSE06	Expt
sp semiconductors						
AlAs	1.99(0.11)	8.16	0.12	1.96(0.09)	1.93	2.28
AlP	2.31(0.14)	7.54	0.13	2.31(0.15)	2.27	2.54
Ar	14.70(0.61)	1.66	0.60	14.70(0.61)	10.36	14.33
BN	6.53(0.24)	4.50	0.22	6.79(0.26)	5.83	6.74
C	5.71(0.21)	5.70	0.18	5.75(0.23)	5.35	5.85
CaO	6.40(0.29)	3.30	0.30	6.48(0.28)	5.30	7.09
LiCl	9.65(0.37)	2.70	0.37	9.64(0.37)	7.80	9.57
LiF	15.83(0.53)	1.90	0.53	15.76(0.52)	11.50	15.35
MgO	8.32(0.34)	3.00	0.33	8.35(0.35)	6.47	8.36
Si	1.21(0.14)	12.00	0.08	1.03(0.12)	1.14	1.23
SiC	2.34(0.16)	6.52	0.15	2.66(0.26)	2.24	2.52
MAE	0.248			0.228	1.425	
3*d* semiconductors						
GaN	3.58(0.22)	5.30	0.19	3.21(0.16)	3.14	3.67
GaP	2.36(0.14)	9.10	0.11	2.59(0.25)	2.31	2.43
GaAs	1.00(0.09)	10.58	0.09	1.02(0.09)	1.26	1.57
InP	1.13(0.09)	9.60	0.10	1.17(0.08)	1.47	1.47
ZnO	3.13(0.25)	3.74	0.27	3.22(0.25)	2.43	3.60
ZnS	3.57(0.20)	5.13	0.19	3.63(0.21)	3.29	3.94
ZnSe	2.03(0.14)	5.90	0.17	2.09(0.14)	2.20	2.87
TiO_2	3.19(0.14)	5.70	0.18	2.94(0.08)	3.39	3.30
MAE	0.358			0.412		0.443
Total MAE	0.294			0.306	1.011	

The parameters of the range-separated CAM functional is determined as the following: The fraction of long-range Fock exchange α_l is set to $1/\epsilon$, where ϵ is the dielectric constant taken from experimental values.[?] The screening length μ is fixed to 0.106 Bohr^{-1}, which is the value used in HSE06 functional.[? ?] The fraction of short-range Fock exchange α_s is sought by enforcing Koopmans' condition with the optimized probe.

A complete list of the parameters for the CAM functional and comparison with PBE0(α_K)

and HSE06 band gaps is included in Table S2. We can see that for simple *sp* semiconductors, K-CAM has similar accuracy as K-PBE0, reducing the mean absolute error to 0.23 eV from the HSE06 error of 1.43 eV. However, the mean absolute error for $3d$ semiconductors is almost doubled, with K-CAM yielding an error of 0.41 eV. On the other hand, while HSE06 tends to significantly underestimate the band gaps of *sp* materials, especially those with intermediate to wide band gaps, the predictions for $3d$ materials are within reasonable range and gives a similar mean absolute error comparing to K-CAM.