Supplementary material

I. CALCULATION OF E_{isolated}

A. Isolated Gaussian charge in bulk

We define a Gaussian charge distribution as:

$$\rho(r) = \frac{q}{\sigma^3(2\pi)^{3/2}} e^{-r^2/(2\sigma^2)}, \tag{1}$$

where σ is the standard deviation. The distribution is normalized and integrates to a total charge q.

The electrostatic energy of a Gaussian distribution, embedded in a medium with a homogeneous dielectric constant ε and interacting with its own potential, is given by:

$$E = \frac{1}{2}\int \rho\phi d\mathbf{r} = \frac{q^2}{\varepsilon}\frac{1}{2\sigma\sqrt{\pi}}. \tag{2}$$

The energy diverges as the charge distribution approaches the point-charge limit ($\sigma \to 0$).

B. Point charge near a single interface

We next consider a point charge in the proximity of an abrupt interface. The analytic form of the potential can be found in elementary text books [1]. Let us denote with ε_1 the dielectric constant on the side of the interface where the charge is located, and ε_2 on the opposite side. The potential on the side of the charge is given by

$$V = \frac{1}{\varepsilon_1}\left(\frac{q}{R_1} + \frac{q'}{R_2}\right), \tag{3}$$

where R_1 is the distance to the original charge and R_2 to the image charge. The image charge corresponds to

$$q' = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2} q = k_1 q. \tag{4}$$

The potential and, consequently, the electrostatic energy of a point charge diverge at $R_1 = 0$, but the interaction with the image charge results from the second term of V:

$$E = \frac{qq'}{4\varepsilon_1 d}, \tag{5}$$

where d is the distance from the interface, i.e., $R_2 = 2d$.

C. Point charge near two interfaces

The case of two interfaces can be approached either by the image-charge method [2, 3] or by fitting potentials at the boundaries [4]. Since there are errors in the printed formulas in Refs. [3, 4], we here provide the correct expressions for a three-material system based on the former approach.

Let the interfaces be located at $-L/2$ and $L/2$. The dielectric constants of the three materials are ε_1, ε_2, and ε_3. We consider the case in which the charge is within the middle layer at a location z_0, and look for the potential at a point z.

The reflections with respect to the left and right interfaces yield image charges $k_1 q$ and $k_2 q$ at positions

$$-\frac{L}{2} - \left[z_0 - \left(-\frac{L}{2}\right)\right] = -L - z_0, \tag{6}$$

$$\frac{L}{2} + \left(\frac{L}{2} - z_0\right) = L - z_0, \tag{7}$$

respectively. k_1 and k_2 are calculated from Eq. (4). A reflection with respect to one interface aims at satisfying the boundary condition at that interface, but in order to satisfy it at the other interface, the reflections have to be continued. The image at $L - z_0$ is thus reflected with respect to the left interface, giving a new image of charge $k_2 k_1 q$ and located at

$$-\frac{L}{2} - \left(L - z_0 - \left(-\frac{L}{2}\right)\right) = -2L + z_0. \tag{8}$$

Similarly, the image at $-L - z_0$ is reflected with respect to the right interface, resulting in a new image of charge $k_1 k_2 q$ located at

$$\frac{L}{2} + \left(\frac{L}{2} - (-L - z_0)\right) = 2L - z_0. \tag{9}$$

The reflections may be continued in this manner resulting in a potential and an electrostatic energy given by an infinite sum over images:

$$V(z) = \frac{q}{\varepsilon_2}\sum_{n=0}^{\infty}(k_1k_2)^n\left(\frac{k_1}{|-(2n+1)L - z_0 - z|} + \frac{k_2}{|(2n+1)L - z_0 - z|} + \frac{k_1k_2}{|-2(n+1)L + z_0 - z|} + \frac{k_1k_2}{|2(n+1)L + z_0 - z|}\right) \tag{10}$$

$$E = \frac{1}{2}\int V\rho = \frac{q^2}{2\varepsilon_2}\sum_{n=0}^{\infty}(k_1k_2)^n\left(\frac{k_1}{|-(2n+1)L - 2z_0|} + \frac{k_2}{|(2n+1)L - 2z_0|} + \frac{2k_1k_2}{|2(n+1)L|}\right). \tag{11}$$

The potential from the point charge itself is not included in these expressions. The interaction with the image charges is thus obtained by evaluating the potential $V(z)$ at $z = z_0$.

D. Gaussian charge near a single interface

When the charge distribution is spherical and localized (i.e. vanishes beyond some limit $r > R$) and the distance from the interface $d > R$, the interaction energy with the image charge is equal to that of a point charge. This is because the interface effects occur in the $1/r$ regime of the potential. The electrostatic selfenergy of an isolated charge distribution can thus easily be added to the image charge interaction of a point charge to obtain the full electrostatic interaction in the vicinity of an interface. This approximation works well for a Gaussian charge distribution provided $d \gg \sigma$.

A more detailed treatment is required when the charge distribution is found on both sides of the interface. The induced potential is solved numerically by treating the charge distribution as a superposition of point charges. The integration is complicated by the $1/r$ form of the kernel, which diverges close to the point charge and leads to a slow convergence with respect to the mesh size. To overcome this problem, we partition the Coulomb kernel $1/r$ as $\mathrm{erf}(ar)/r + \mathrm{erfc}(ar)/r$. The range separation is determined by the length a^{-1}, which is taken to be of the order of the distance between the mesh points Δx. In this way, the long-range interaction (the first term) no longer diverges near the point charge. As the second term is concerned, the interaction is extremely short-ranged. Assuming that within its range the charge density is constant, we may evaluate its contribution to the local potential analytically as $\rho(\mathbf{r}') \int [\mathrm{erfc}(ar)/r]\, 4\pi r^2 dr = \rho(\mathbf{r}')\pi/4a^2$. With this approach the integrals converge much faster with respect to the size of the numerical mesh. Generally, we use $a^{-1} = 0.4\sqrt{2}\Delta x$, which was found to give numerical errors below 1 mHa for a mesh size of $N = 20$.

Using a finite charge distribution rather than a point charge overcomes the diverging electrostatics when $d \to 0$. Figure 1 demonstrates the typical behavior of the electrostatic energy, which always assumes a finite value as the charge distribution is moved through the interface. For comparison, we also plot the approximated result which we obtained by using image-charge interactions pertaining to a point charge. This approximation coincides with the result from the proper integration for distances larger than about 2σ, but diverges at the interface.

For more complex dielectric profiles, calculating E_{isolated} through summation of image charge interactions can become difficult and even computationally tedious. In such cases, the isolated limit may be obtained more easily through a uniform extrapolation of the periodic system (E_{periodic}). As shown in Fig. 2 of the Letter, the two approaches yield equivalent results for surfaces with simple dielectric constant profile.

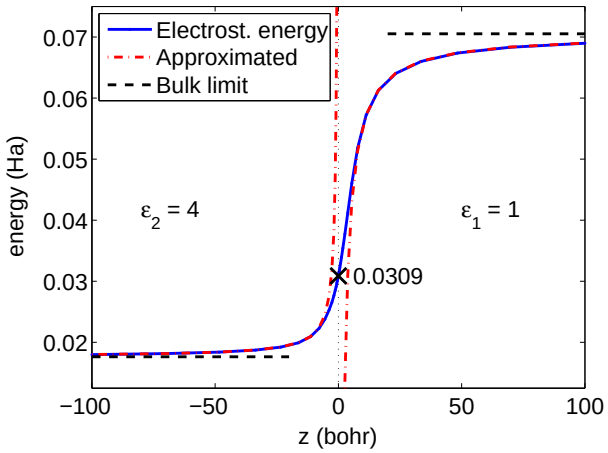

FIG. 1. Electrostatic energy (solid, blue) for a Gaussian charge distribution ($\sigma = 4$ bohr) as its center is moved across a dielectric interface. We took $\varepsilon_1 = 1$ on the right-hand side and $\varepsilon_2 = 4$ on the left-hand side of the interface located at $z = 0$. The indicated value corresponds to the energy for a distribution centered at the interface. Bulk limits are indicated with horizontal dashed lines. For comparison, we also show an approximated result obtained by using image-charge interactions pertaining to a point charge (dash-dotted, red).

II. CALCULATION OF E_{periodic}

When the dielectric-constant profile $\varepsilon(\mathbf{r})$ only depends on z, the Poisson equation reads:

$$\varepsilon(z)\nabla^2 V(\mathbf{r}) + \frac{\partial}{\partial z}\varepsilon(z)\frac{\partial}{\partial z}V(\mathbf{r}) = -\rho(\mathbf{r}). \qquad (12)$$

For a periodically repeated supercell, we found it convenient to solve this equation in Fourier space. A multiplication in real space corresponds to a convolution in Fourier space (convolution theorem) and $\partial f(x)/\partial x = -iG_x f(G)$. Moreover, since the dielectric function has no structure in the x and y directions, $\varepsilon(\mathbf{G}) = \varepsilon(0,0,G_z) = \varepsilon(G_z)$. Equation (12) thus becomes:

$$\varepsilon(G_z) * \left[G^2 V(\mathbf{G})\right] + \left[G_z \varepsilon(G_z)\right] * \left[G_z V(\mathbf{G})\right] = \rho(\mathbf{G}), \quad (13)$$

where $*$ denotes a convolution in the z direction. Hence, for any fixed G_x and G_y, we can write a linear system

$$\begin{aligned}\rho(G_x, G_y, G_z) = &\\ &\sum_{G_z'} \varepsilon(G_z - G_z') G_z'^2 V(G_x, G_y, G_z')\\ &+ \sum_{G_z'} (G_z - G_z')\varepsilon(G_z - G_z') G_z' V(G_x, G_y, G_z')\\ &+ \sum_{G_z'} \varepsilon(G_z - G_z')(G_x^2 + G_y^2) V(G_x, G_y, G_z'), \end{aligned} \qquad (14)$$

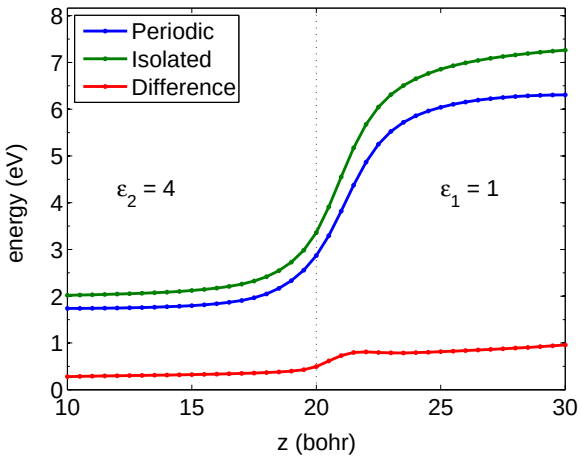

FIG. 2. Electrostatic energies of a Gaussian charge distribution in the vicinity of an interface in the isolated and periodic cases. The periodic system consists of a $20 \times 20 \times 40$ bohr^3 supercell, in which the left half has a dielectric constant $\varepsilon = 4$ and the right half a dielectric constant $\varepsilon = 1$. The Gaussian charge distribution has a width of $\sigma = 1$ bohr. The average potential is set to zero in the calculation of E_{periodic}.

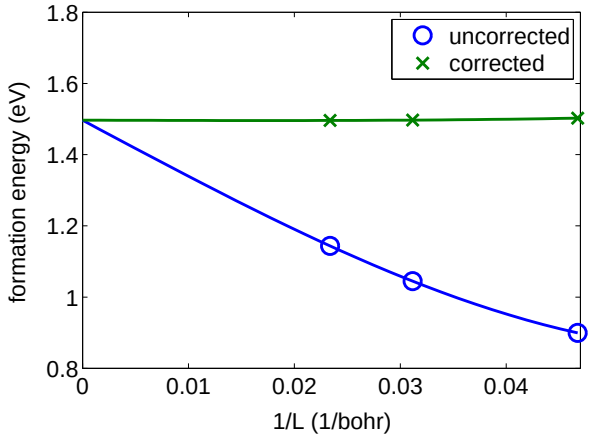

FIG. 3. Uncorrected and corrected formation energies for a Cl vacancy in the $+1$ charge state in bulk NaCl as a function of the inverse side of the supercell. The chemical potential of Cl is taken from the isolated Cl_2 molecule and the Fermi energy is set at the valence band maximum.

which can be further simplified to yield

$$\rho(G_x, G_y, G_z) = \sum_{G'_z} \varepsilon(G_z - G'_z) \times (G_x^2 + G_y^2 + G_z G'_z) V(G_x, G_y, G'_z), \quad (15)$$

For each G_x and G_y, this one-dimensional system is written in matrix form $\rho = AV$ and solved separately to obtain $V(\mathbf{G})$. It should be noted, that the resulting matrix A is singular when $G_x = G_y = 0$. This problem is here overcome by setting the head component ($G_z = 0, G'_z = 0$) to any nonzero value. This choice only affects the average potential, which is later set to zero. When ε does not depend on z, the only nonvanishing component of ε is at $G_z - G'_z = 0$. This results in a diagonal matrix with $A_{G_z,G_z} = \varepsilon \cdot (G_x^2 + G_y^2 + G_z^2)$, thereby recovering the standard form of the Poisson equation.

The present formulation can trivially be extended to the case in which the dielectric constant is not isotropic, with the dielectric constant in the direction perpendicular to the interface being different than in parallel directions [$\varepsilon_\perp(z) \neq \varepsilon_\parallel(z)$]. In this case, the Poisson equation becomes:

$$\varepsilon_\parallel(z)\frac{\partial^2}{\partial x^2}V(r) + \varepsilon_\parallel(z)\frac{\partial^2}{\partial y^2}V(r) + \varepsilon_\perp(z)\frac{\partial^2}{\partial z^2}V(r) + \frac{\partial}{\partial z}\varepsilon_\perp(z)\frac{\partial}{\partial z}V(r) = -\rho(r). \quad (16)$$

III. CORRECTION AT A SINGLE INTERFACE

We here focus on an isolated interface and display in Fig. 2 the correction $\delta E = E_{\text{isolated}} - E_{\text{periodic}}$, which corresponds to E_{corr} in Eq. (1) of the main manuscript apart from the missing alignment term $q\Delta V$. The electrostatic energies E_{isolated} and E_{periodic} show similar z-dependent profiles, as both energies are dominated by the physical image charge due to the interface. Their difference ΔE is much smaller and results from the effect of the additional image charges induced by the supercell periodicity. At distances beyond about 2σ from the interface, the correction ΔE assumes nearly constant values on either side of the interface.

IV. Cl VACANCY IN NaCl

A. Formation energy in the bulk

Figure 3 shows uncorrected and corrected formation energies of a Cl vacancy in bulk NaCl as the supercell size is scaled. The corrected formation energies are obtained with the scheme proposed by Freysoldt, Neugebauer, and Van De Walle [5]. The correction is found to perform very well, yielding a value of 1.50 eV for all considered supercells and agreeing well with the extrapolation of the uncorrected results [6].

The applied correction scheme is expected to work well only when the charge is well localized within the supercell [6]. This is indeed the case for the present defect. Following Ref. 6, we not only inspected the wave functions, but also checked that the Kohn-Sham levels locate well within the band gap and that the delocalized screening charge far from defect approaches the value $(1 - 1/\varepsilon)q/\Omega$.

B. Dielectric constant profile and potential alignment

For the dielectric-constant profile across an interface, we adopt a model expression based on an error-function form:

$$\varepsilon(z) = \tfrac{1}{2}\,(\varepsilon_2 - \varepsilon_1)\cdot \mathrm{erf}[(z - z_0)/\beta] + \tfrac{1}{2}\,(\varepsilon_1 + \varepsilon_2)\,, \quad (17)$$

where z_0 represents the interface position, ε_1 and ε_2 the dielectric constants on either side of the interface, and β the smoothness of the dielectric-constant profile.

The adopted model expression describes well the dielectric-constant profile extracted from the DFT calculation (cf. main paper). From the fit of the DFT profile [Fig. 4(a)], we obtain z_0 located at 2.3 bohr ($0.22\ a_{\rm lat}$) from the outermost layer of atoms and a smoothness parameter β of 2 bohr. For this model $\varepsilon(z)$ [model 1 in Fig. 4(a)], we find that a shift ΔV of -20 meV accounts well for the differences between the model and the DFT potentials on both sides of the surface (differences of about -10 and -30 meV) [Fig. 4(b)]. The same ΔV also ensures that the potentials match along the lateral direction [Fig. 4(c)].

In Fig. 4, we also demonstrate the effect of using other $\varepsilon(z)$ profiles. In particular, when z_0 is taken to coincide with the outermost layer of atoms [model 2 in Fig. 4(a)], the deviations between the model and DFT potentials cannot be described by a rigid shift ΔV, leaving at best residual errors of ± 0.15 eV depending on the side of the interface considered. Consequently, this leads to an ambiguity of 0.3 eV in the determination of ΔV value and highlights the importance of a proper model for $\varepsilon(z)$.

It is of interest to evaluate the possibility of modeling $\varepsilon(z)$ without requiring a DFT calculation. For this purpose, we estimate the interface position z_0 on the basis of the average ionic radius of Na and Cl (1.415 Å), which also corresponds to half the distance between the atomic planes, i.e., $a_{\rm lat}/4$. The dielectric-constant profile is taken to be nearly abrupt [model 3 in Fig. 4(a)] to avoid the need of extra parameters. As can be seen in Fig. 4, this model is for NaCl nearly as effective as the one fitted to the DFT profile.

In contrast with its behavior along the z direction, the potential in the lateral directions x and y remains unaffected by the adopted model for the dielectric-constant profile [Fig. 4(c)]. Due to this robustness, we calculated ΔV by aligning the model and DFT potentials along a lateral direction for all results given in the paper. The potential difference in the z direction then serves as an indicator of the quality of the model adopted for the dielectric-constant profile.

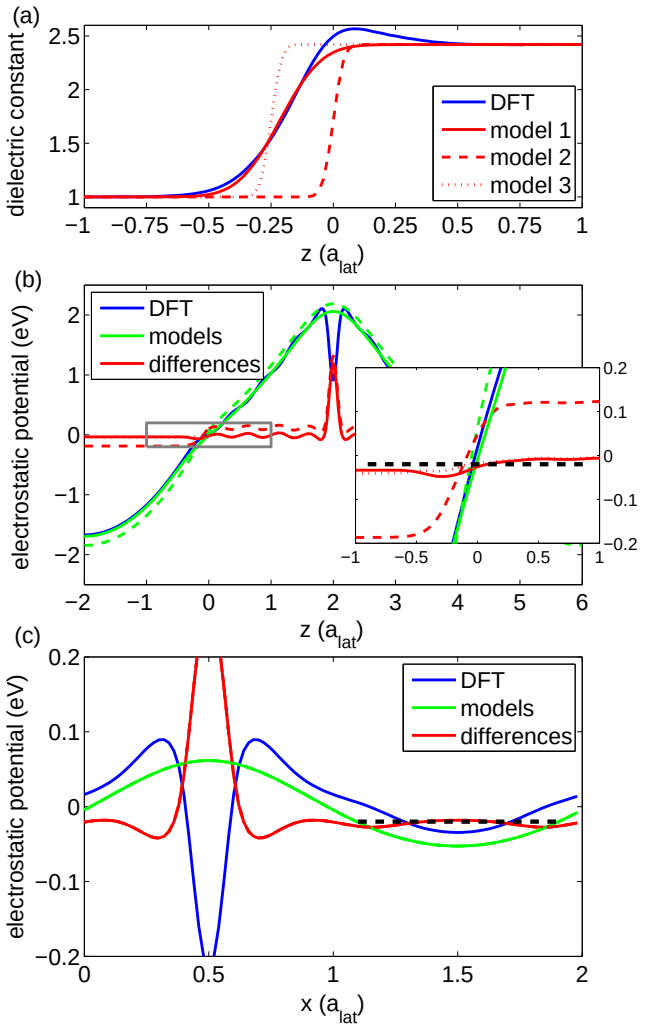

FIG. 4. (a) Dielectric-constant profile of the NaCl surface as obtained from a DFT calculation and three model profiles based on different error-function forms. Model 1 is obtained from a fit of the DFT result. Model 2 places the surface in correspondence of the outermost layer of atoms and defines an abrupt profile. Model 3 corresponds to an abrupt profile in which the interface occurs at a distance of $a_{\rm lat}/4$ (half the distance between atomic layers) from the outermost layer of atoms. (b) Corresponding plane-averaged electrostatic potentials from the models and the DFT calculation, and their differences. In the inset, the potential difference in the region denoted by a gray square is smoothened by a moving average over half the lattice constant. The adopted value of ΔV (-20 meV) is shown by a horizontal line (dashed, black). (c) Same as in (b), but along the lateral direction x. The potentials obtained with the three models overlap.

V. Si DANGLING BOND AT THE H-PASSIVATED Si(001)2×1 SURFACE

To demonstrate the general applicability of our correction scheme, we address in this section a Si dangling bond at the H-passivated Si(001)2×1 surface. In our calculation, the plane-wave cutoff is fixed at 30 Ry. The k-point sampling in the lateral directions corresponds to a 12×12 mesh in the unit surface cell, while the sole Γ-point is used in the perpendicular direction.

The 2×1 reconstruction of the Si(001) surface features rows of Si-Si dimers and dangling bonds, which can be passivated by hydrogen when available in a sufficiently abundant way. Full hydrogen passivation is adopted here. In our calculation, we consider a finite Si slab and use the experimental lattice constant $a_{\rm lat} = 5.43$ Å. Only the surface Si and H atoms are allowed to relax. The side of the surface unit cell is $a_{\rm lat}/\sqrt{2}$. On opposite sides of the slab, the dimer rows are oriented

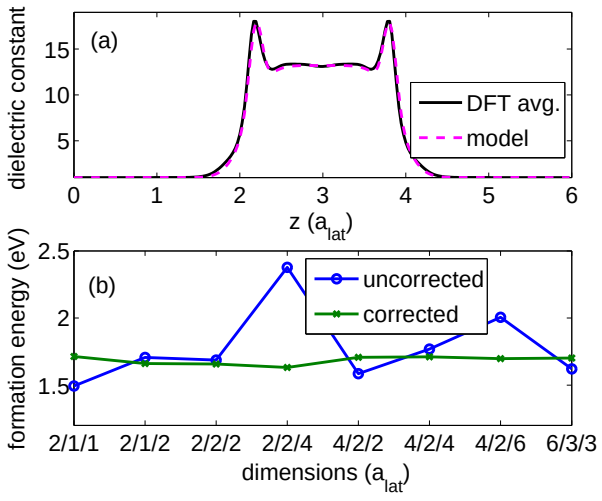

FIG. 5. (a) Unit-cell averaged dielectric constant profile obtained from a DFT calculation and the adopted model profile. (b) Uncorrected and corrected formation energies for the Si dangling bond at the Si surface in the -1 charge state, for various supercell shapes. The three figures denote the lateral thickness, the slab thickness, and the vacuum thickness, in units of $a_{\rm lat}/\sqrt{2}$, $a_{\rm lat}$, and $a_{\rm lat}$, respectively.

perpendicularly. Therefore, the minimum lateral size is 2×2 ($2a_{\rm lat}/\sqrt{2}\times2a_{\rm lat}/\sqrt{2}$). Due to the surface reconstruction and to the passivating H layer, the dielectric constant profile shows a fairly strong peak at the surface, as shown in Fig. 5(a). To describe this behavior we add an extra Gaussian in the model used for the profile of the dielectric constant, Eq. (17), and the energy $E_{\rm isolated}$ is determined from the extrapolation of $E_{\rm periodic}$ as discussed in Sec. I D. This approach was found to improve the corrected formation energies.

The dangling bond defect is created by removing one of the passivating H atoms. The negative charge state is obtained through the addition of one electron. Only the Si atom containing the dangling bond is allowed to relax. The performance of the correction scheme is illustrated in Fig. 4(b) of the main paper upon uniform scaling of the cell parameters and in Fig. 5(b) for supercells of other shapes.

We note, that in the case of Si slabs, the highest occupied state is very sensitive to the slab thickness due to the confinement effect. We thus obtained the valence band maximum by aligning the Si band structure from a separate bulk calculation through the electrostatic potential in the central part of the Si slab.

[1] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1998) ISBN 047130932X.

[2] A. Pasquarello, L. C. Andreani, and R. Buczko, Phys. Rev. B **40**, 5602 (1989).

[3] A. Pasquarello, M. S. Hybertsen, and R. Car, Phys. Rev. B **53**, 10942 (1996).

[4] M. Kleefstra and G. C. Herman, J. Appl. Phys. **51**, 4923 (1980).

[5] C. Freysoldt, J. Neugebauer, and C. G. Van de Walle, Phys. Rev. Lett. **102**, 016402 (2009).

[6] H.-P. Komsa, T. T. Rantala, and A. Pasquarello, Phys. Rev. B **86**, 045112 (2012).