The electronic structure of surfaces plays a key role in the properties of quantum devices. However, surfaces are also the most challenging to simulate and engineer. Here, the electronic structure of InAs(001), InAs(111), and InSb(110) surfaces is studied using a combination of density functional theory (DFT) and angle-resolved photoemission spectroscopy (ARPES). Large-scale first principles simulations are enabled by using DFT calculations with a machine-learned Hubbard U correction [npj Comput. Mater. 6, 180 (2020)]. To facilitate direct comparison with ARPES results, a “bulk unfolding” scheme is implemented by projecting the calculated band structure of a supercell surface slab model onto the bulk primitive cell. For all three surfaces, a good agreement is found between DFT calculations and ARPES. For InAs(001), the simulations clarify the effect of the surface reconstruction. Different reconstructions are found to produce distinctive surface states, which may be detected by ARPES with low photon energies. For InAs(111) and InSb(110), the simulations help elucidate the effect of oxidation. Owing to larger charge transfer from As to O than from Sb to O, oxidation of InAs(111) leads to significant band bending and produces an electron pocket, whereas oxidation of InSb(110) does not. The combined theoretical and experimental results may inform the design of quantum devices based on InAs and InSb semiconductors, for example, topological qubits utilizing the Majorana zero modes.
a conventional semiconductor and an s-wave superconductor. Exceptional control of interface properties is necessary to realize topological superconducting phases and manipulate Majorana zero modes, key requirements in topological quantum computation proposals.\textsuperscript{19–22} The semiconductor in such a hybrid semiconductor-superconductor Majorana device is required to have a large g-factor, strong Rashba spin-orbit coupling and significant proximity-induced superconducting gap. Proximity-induced superconductivity has been studied recently in nanowires, including NbTiN/InSb\textsuperscript{23–30}, Al/InAs\textsuperscript{31–37}, Al/InSb,\textsuperscript{38–42} Pb/InAs,\textsuperscript{43} and Sn/InSb.\textsuperscript{44} High-quality, uniform and transparent superconductor–semiconductor interfaces are required to optimize topological gaps in semiconductor–superconductor heterostructures.\textsuperscript{45–47}

The functionality and performance of InX hybrid interfaces in the aforementioned applications depend critically on the interface quality. Changes in the structure of the interface at the atomic scale affect the electronic properties. The interface configuration may give rise to (desirable or undesirable) interface states, alter the band bending and band alignment, or affect the magnitude of the proximity-induced gap and spin-orbit coupling.\textsuperscript{48–50} A deeper understanding of the relation between the structure and electronic properties of InX hybrid interfaces could advance the synthesis of precisely controlled interfaces with tunable properties.

InX hybrid interfaces are often grown by molecular beam epitaxy (MBE) with InX serving as the substrate.\textsuperscript{28,31,38,41,51} The nanowire-based Majorana devices typically have multiple interfaces, including InX-superconductor and InX-oxide interfaces. Therefore, elucidating the structure and electronic properties of InAs and InSb surfaces is an important first step toward advancing the understanding of epitaxially grown InX hybrid interfaces. InAs and InSb exhibit a number of surface reconstructions, which have been observed with atomic resolution using scanning tunneling microscopy and low-energy electron diffraction (LEED).\textsuperscript{52–56} Understanding the resulting surface states and Fermi-level pinning is important for engineering appropriate interface Hamiltonian and realizing topological superconductivity hosting Majoranas. Surface reconstructions with less pinning and no problematic surface states close to the band edges are preferred.\textsuperscript{47} It may be challenging to identify the InX surface reconstructions present in experiments because some samples may contain highly reconstructed surfaces with large unit cells, disorder, or domains with related but different reconstructions.\textsuperscript{57,58} Band bending and surface states have been observed by angle-resolved photoemission spectroscopy (ARPES).\textsuperscript{59,60} However, it is a nontrivial task to assign spectroscopic signatures to specific structural features. For instance, without comparing to calculations, it is difficult to associate surface state dispersion with the corresponding surface reconstruction.\textsuperscript{61} Moreover, a number of experiments have investigated the effect of the adsorption of gases on the surfaces of InAs and observed the formation of an accumulation layer, for instance for N\textsubscript{2} on InAs\textsuperscript{62} or O\textsubscript{2} on InAs.\textsuperscript{63,64} However, the microscopic origin of the charge accumulation has not been understood. Furthermore, there has been no systematic comparison of oxygen adsorption between oxidized InAs and InSb surfaces, which is particularly important for the understanding of nanowire devices, as they can be fabricated in situ avoiding oxidation or ex situ with oxidized surfaces. The amount of surface charge accumulation has significant influence on the electronic properties of such devices.\textsuperscript{47,65}

First principles simulations based on density functional theory (DFT) can help interpret experiments and resolve the effects of surface reconstructions and defects.\textsuperscript{66–71} DFT studies of InAs and InSb surfaces and interfaces have been limited because local and semi-local exchange-correlation functionals severely underestimate the band gap to the point that it reduces to zero.\textsuperscript{72} Refs. [52, 73] focused on the InAs(001) surface reconstruction equilibrium phase diagram using the local density approximation (LDA). InAs(110) and InSb(001) surfaces have been studied using LDA with empirically modified pseudopotentials in refs. [74, 75]. Hybrid functionals, such as the the Heyd–Scuseria–Ernzerhof (HSE) functional\textsuperscript{76,77} and many-body perturbation theory within the GW approximation\textsuperscript{78,79} yield band gaps in good agreement with experimental values for bulk InAs and InSb.\textsuperscript{80,81} However, it is not feasible to use these methods for simulations of large surface and interface models due to their high computational cost. One alternative to hybrid functionals is the modified Becke–Johnson (mBJ) semilocal exchange,\textsuperscript{82} which has a local approximation to an atomic exact-exchange potential. The mBJ functional can produce band gap values similar to hybrid functionals for InAs and InSb at a lower computational cost.\textsuperscript{83} However, the mBJ functional can only be used for fully periodic systems. For surface and interface calculations with a vacuum region mBJ calculations tend to diverge because in the vacuum, where the electron density and kinetic energy density are close to zero, the functional becomes unstable.\textsuperscript{84}

Recently, we have introduced a new method of DFT with a machine-learned Hubbard U correction, which can provide a solution for accurate and efficient simulations of InAs and InSb.\textsuperscript{85} Within the Dudarev formulation of DFT+U\textsuperscript{86} the effective Hubbard U is defined as \(U_{\text{eff}} = U - J\) with \(U\) and \(J\) representing the on-site Coulomb repulsion and the exchange interaction, respectively. For a given material, the \(U_{\text{eff}}\) parameters of each element are machine-learned using Bayesian optimization (BO). The BO algorithm finds the optimal \(U_{\text{eff}}\) values that maximize an objective function formulated to reproduce as closely as possible the
band gap values and the qualitative features of the band structure obtained with a hybrid functional\cite{85}

\[ f(\bar{U}) = -\alpha_v(F_{g}^{HSE} - F_{g}^{PBE+U})^2 - \alpha_r(\Delta Band)^2 \] (1)

Here, \( \bar{U} = [U^1, U^2, \ldots, U^n] \) is the vector of \( U_{eff} \) values applied to different atomic species. \( \Delta Band \) is defined as the mean squared error of the PBE+U band structure with respect to HSE:

\[ \Delta Band = \sqrt{\frac{1}{N_e} \sum_{j=1}^{N_s} \sum_{j=1}^{N_b} (\epsilon_{HSE}^{j}\{k\} - \epsilon_{PBE+U}^{j}\{k\})^2} \] (2)

\( N_e \) represents the total number of eigenvalues, \( \epsilon \), included in the comparison, \( N_s \) is the number of \( k \)-points, and \( N_p \) is the number of bands selected for comparison.

The PBE+U(BO) method allows for negative \( U_{eff} \) values. Negative \( U_{eff} \) values are theoretically permissible when the exchange term, \( J \), is larger than the on-site Coulomb repulsion, \( U \).\cite{87–91} We have found that negative \( U_{eff} \) values are necessary to produce band gaps for narrow-gap semiconductors, such as InAs and InSb. Because the reference hybrid functional calculation is performed only once for the bulk material to determine the optimal \( U_{eff} \) values, the computational cost of PBE+U(BO) calculations for surfaces is comparable to semi-local DFT. Moreover, unlike the mBJ functional, PBE+U(BO) can be used for surface and interface models with a vacuum region.

Supercells are often used in DFT simulations to study the effect of surface reconstructions, disorder, and defects within a periodic picture. The complexity of a supercell band structure makes it difficult to compare with ARPES experiments. Band unfolding\cite{92,93} is a useful method for projecting the band structure of a supercell onto the corresponding primitive cell with appropriate spectral weights. Previously, band unfolding has been applied to surface slab band structures only in the \( xy \) plane, that is, a supercell of \( XX'YX'Z \) was unfolded onto the corresponding \( 1X'1X'Z \) slab. However, when the band structure of a slab with a large number of layers is unfolded only in the \( xy \) plane, the number of bands may still be too large to compare with ARPES. Moreover, bands corresponding to nonzero values of \( k_z \) in the primitive cell may be present in the unfolded slab band structure, creating artifacts that do not appear in ARPES. To address these issues, we have developed a “bulk unfolding” scheme, where the bands of a surface slab are unfolded onto the corresponding bulk primitive cell. This facilitates the comparison of calculated band structures to ARPES experiments.

The bulk unfolding scheme is used to compare the calculated slab band structures to ARPES measurements. The results of bulk-unfolded PBE+U(BO) simulations are in good agreement with ARPES. Moreover, the simulations help interpret features observed in ARPES. For InAs(001), the effect of surface reconstruction is investigated by comparing the band structures produced by the \( \beta^2(2 \times 4), \alpha^2(2 \times 4), \zeta(4 \times 2), \) and \( c(4 \times 4) \) reconstructions. The results of PBE+U(BO) simulations support the presence of \( \beta^2(2 \times 4) \) and \( \zeta(4 \times 2) \) coexisting domains. For InAs(111) and InSb(110), we investigate the effect of oxidation. In agreement with ARPES, PBE+U(BO) simulations show that the binding of oxygen to As on the InAs(111) increases the band bending and produces an electron pocket. In contrast, the binding of oxygen to Sb on InSb(110) does not significantly shift the valence band maximum and no electron pocket is observed. This is explained by larger charge transfer from As to O than from Sb to O.

The paper is organized as follows. In Section 2, we describe our methods, including the bulk unfolding scheme, DFT settings, and slab construction. In Section 3 experimental and theoretical results are presented for InAs(001), InAs(111), and InSb(110).

2. Methods

2.1. Band Unfolding

In the following, the bulk unfolding scheme is described. As an example, the InAs(001) surface with a \( \beta^2(2 \times 4) \) reconstruction is used, illustrated in Figure 1a. The slab model comprises a \( 2 \times 4 \) supercell in the \( xy \) plane with 20 layers and a vacuum region of 40 Å in the \( z \) direction. Band unfolding is based on the relation between the basis vectors of the supercell and the primitive cell\cite{92}

\[ \mathbf{A} = \mathbf{M} \cdot \mathbf{a} \] (3)

where \( \mathbf{A} \) and \( \mathbf{a} \) are matrices with the supercell and primitive cell basis vectors as rows, respectively, and \( \mathbf{M} \) is the integer transformation matrix between the two. A similar relation holds in reciprocal space

\[ \mathbf{B} = \mathbf{M}^{-1} \cdot \mathbf{b} \] (4)

where \( \mathbf{B} \) and \( \mathbf{b} \) are basis vector matrices in reciprocal space corresponding to the supercell and primitive cell, respectively. Given a wave vector \( \mathbf{k} \) in the primitive cell Brillouin zone, there is exactly one wave vector \( \mathbf{K} \) in the smaller supercell Brillouin zone such that the two vectors are related by a reciprocal lattice vector \( \mathbf{G} \) of the supercell

\[ \mathbf{k} = \mathbf{K} + \mathbf{G} \] (5)

The unfolded band structure is formed by weighted dots with the weights given by the spectral function of the primitive cell,

\[ A(\mathbf{k}, \epsilon) = \sum_m P_{\mathbf{k}m}(\mathbf{k}) \delta(\epsilon_{\mathbf{k}m} - \epsilon) \] (6)

where \( \epsilon \) is the energy, \( m \) is the band index of the supercell band structure, \( P_{\mathbf{k}m}(\mathbf{k}) \) is the projection of the supercell state \( |\mathbf{k}m\rangle \) on all the primitive cell states \( |\mathbf{n}\rangle \) with a given \( \mathbf{k} \) and \( \mathbf{n} \) is the primitive cell band index

\[ P_{\mathbf{k}m}(\mathbf{k}) = \left\langle \mathbf{k}m | \langle \mathbf{k}m | \mathbf{n}\rangle \right\rangle^2 \] (7)

Because the supercell comprises several primitive cells, the supercell states may contain contributions from several primitive cell states. \( P_{\mathbf{k}m} \) provides information about how much of the character of \( |\mathbf{n}\rangle \) is preserved in \( |\mathbf{k}m\rangle \).
Figure 1. a) Top view and side view of the $\beta(2 \times 4)$ reconstruction of InAs(001). InAs(001) $\beta(2 \times 4)$ band structures with b) $xy$-unfolding, c) $z$-unfolding, and d) bulk unfolding. e) The surface Brillouin zone for InAs(001) $\beta(2 \times 4)$. Red lines correspond to $k_z = v|\Gamma X|$ and blue lines correspond to $k_z = 0$.

Figure 1b shows the result of the standard $xy$ unfolding scheme, obtained the VaspBandUnfolding code, available at: https://github.com/QijingZheng/VaspBandUnfolding. Within this scheme, the band structure is unfolded onto a $1 \times 1$ slab in the $xy$ plane with the original dimension in the $z$ direction, illustrated under the band structure. The $xy$-unfolded band structure is still quite complex with a large number of bands due to the large supercell size in the $z$ direction. The supercell band structure may be further unfolded onto a periodic bulk unit cell with the same orientation as the surface supercell. Most band unfolding codes can, in principle, perform such “$z$-unfolding.” However, the slab must be constructed such that its length in the $z$-direction is an integer multiple of the $z$-dimension of the periodic unit cell with the same orientation in order to preserve the relation between the supercell and primitive cell as described by Equations (1)– (3). The $z$-unfolded band structure, shown in Figure 1c, has fewer bands and appears more similar to the bulk band structure shown in Figure 2. However, some extra bands are present around $-3$ eV. The extra bands appear because when the primitive Brillouin zone of InAs bulk is projected onto the (001) periodic slab, the bands along $XW$ at $k_z = |\Gamma X|$ overlap with $X\Gamma$ at $k_z = 0$, as shown in Figure 1e. Thus, the $z$-unfolding scheme eliminates the effect of the slab thickness but cannot decompose and eliminate the bands corresponding to nonzero values of $k_z$.

The bulk unfolding scheme unfolds the supercell band structure onto the bulk primitive cell, as shown in Figure 1d. It is noted that the band structure path for bulk unfolding is $X$–$\Gamma$–$X$, which is mapped to $M$–$\Gamma$–$M$ in the (001) projection, as illustrated in Figure 1e. Bulk unfolding is not straightforward because the matrix $M$ in Equations (1) and (2) must be an integer matrix. However, the transformation between a “standard slab,” and the primitive bulk basis vectors may produce a noninteger matrix. The “standard slab” means that the slab’s unit cell structure is standardized according to the crystal symmetry and lengths of basis vectors [95]. For instance, the basis vector matrix of a zinc blende primitive cell is given by

$$a = \begin{bmatrix} 0 & x & x \\ x & 0 & x \\ x & x & 0 \end{bmatrix}$$

A (111) slab with one unit cell in the $xy$ plane and $N$ unit cells in the $z$ direction may be generated by applying the integer transformation matrix $M_1$.

$$A_1 = \begin{bmatrix} 0 & -x & x \\ -x & x & 0 \\ -2Nx & -2Nx & -2Nx \end{bmatrix} = M_1 \cdot a$$
Because the basis vectors comprising the matrix $A_1$ are not standard, commonly used software packages, such as Python Materials Genomics (pymatgen)\cite{96} and Atomic Simulation Environment,\cite{97} will automatically apply a transformation, $M_1$, into the standard basis vectors

\[
A_2 = \begin{bmatrix}
\frac{x}{\sqrt{2}} & \frac{\sqrt{3}x}{\sqrt{2}} & 0 \\
\frac{x}{\sqrt{2}} & \frac{\sqrt{3}x}{\sqrt{2}} & 0 \\
0 & 0 & 2\sqrt{3}N_x
\end{bmatrix} = M_1 \cdot A_1 \tag{10}
\]

The final transformation matrix, $M_1M_2$, used to obtain the standard basis vectors comprising the matrix $A_2$ is not an integer matrix due to rotation, which prevents the usage of band unfolding. Hence, in order to perform bulk unfolding, the standard slab represented by the matrix $A_2$ should be converted into the nonstandard slab, represented by the matrix $A_1$, which can be obtained from the primitive cell via an integer transformation matrix. An open-source script, compatible with VaspBandUnfolding, was developed that takes the bulk structure and surface Miller indices as input and outputs the correct nonstandard slab structure for band unfolding and the integer transformation matrix $M_1$. The code is available for download at: https://github.com/Shuyangzero/Unfolding as well as in the VaspVis visualization package, available for download at https://github.com/DerekDardzinski/vaspvis. As seen in Figure 1d, bulk unfolding onto the primitive cell results in a band structure, whose prominent features closely resemble that of the bulk material. The bands corresponding to $k_z = [\Gamma X]$ are eliminated, producing a band structure that is directly comparable to ARPES measurements. Surface states arising from the reconstruction appear as weak signatures in the background. The photon energy used in an ARPES experiment determines whether or not such surface states can be detected. From here on, only bulk-unfolded band structures are compared to ARPES experiments. The $x_p$-unfolded and $z$-unfolded band structures are provided in the Supporting Information.

\section*{2.2. DFT Settings}

DFT calculations were performed using the Vienna ab initio simulation package\cite{98} with the projector augmented wave (PAW) method.\cite{99,100} The generalized gradient approximation (GGA) of Perdew, Burke, and Ernzerhof (PBE) was employed for the description of exchange-correlation interactions among electrons.\cite{101,102} A machine-learned Hubbard U correction was applied to the p-orbitals of In, As, and Sb. The optimal values of $U_{\text{eff}}^{\text{In}}, U_{\text{eff}}^{\text{As}}, U_{\text{eff}}^{\text{Sb}}$ were found by Bayesian optimization.\cite{85} For InAs, $U_{\text{eff}}^{\text{In}} = -0.5$ eV and $U_{\text{eff}}^{\text{As}} = -7.5$ eV. For InSb, $U_{\text{eff}}^{\text{In}} = -0.2$ eV, $U_{\text{eff}}^{\text{Sb}} = -6.1$ eV. With these parameters, PBE+U(BO) yields a band gap of 0.31 eV for InAs and 0.14 eV for InSb, in good agreement with the experimental values of 0.41 and 0.24 eV, respectively.\cite{103} Spin-orbit coupling was included throughout and the energy cutoff was set to 400 eV.

For verification, the band structures obtained with PBE+U(BO) were compared to those obtained with HSE for bulk unit cells of InAs and InSb with the same orientations as the surface slabs studied here. A 8x8x8 k-point grid was used for these calculations. A representative example for InAs(001) is shown in Figure 2. PBE+U(BO) is in close agreement with HSE for the band gap and the band structure features in the vicinity of the Fermi level. However, farther from the Fermi level the agreement deteriorates and the band width is underestimated. The band width is similarly underestimated by PBE+U(BO) away from the Fermi level for all surfaces studied here.

\section*{2.3. Slab Construction}

Slab models were constructed using the experimental lattice constants of 6.058 Å for InAs and 6.479 Å for InSb.\cite{103} A vacuum region of about 40 Å was added in the z direction to avoid spurious interactions between periodic replicas (for the purpose of band unfolding the closest integer number of primitive cells to 40 Å was used). To avoid surface states due to dangling bonds as well as charge transfer and spurious interactions between the two surfaces of the slab,\cite{105} the As/Sb atoms on the surface were passivated by pseudo-hydrogen atoms with 0.75 fractional electrons and the In atoms on the surface were passivated by pseudo hydrogen atoms with 1.25 fractional electrons. Specifically, for InAs(100) slab models, the unreconstructed bottom surfaces were passivated; for bare InAs(111) and InSb(110), the top and bottom surfaces were passivated; for oxidized InAs(111) and InSb(110), the bottom surfaces (without adsorbed oxygen) were passivated. Structural relaxation was performed for the surface In, As, and Sb atoms, the passivating pseudo-hydrogen atoms, and the adsorbed oxygen on InAs(111) and InSb(110) until the change of the total energy was below $10^{-5}$ eV. A 8x8x1 k-point grid was used to sample the Brillouin zone of surface slab models and dipole corrections were applied. The Fermi level position was not constrained in any of the calculations. Illustrations of the high-symmetry paths in the Brillouin zone used for surface band structure calculations are provided in the Supporting Information.

When constructing a surface or interface slab model, it is important to converge the number of layers.\cite{107} An indicator that the number of layers is converged is that the band gap of the surface slab approaches the bulk limit. Figure 3 shows the band gap as a function of the number of atomic layers for InAs(111) and InSb(110), as representative surfaces. It is noted that the convergence test was only conducted for models without surface...
reconstructions and the surfaces were passivated, such that there were no surface states in the band gap. In each iteration the number of layers was increased by 4. If the band gap difference between the current iteration and the previous iteration was within $5 \times 10^{-2}$ eV, the current number of layers was regarded as converged. For InAs(111), 50 layers are required to converge the band gap. For InSb(110), 42 layers are required. For both surfaces the converged band gap value is close to the bulk value. For InAs(001), modeling surface reconstructions required constructing supercells in the $xy$ plane. Therefore, the maximal number of layers was limited by the computational cost. The largest models that were able to be calculated were a 20-layer slab for the $\beta 2(2 \times 4)$, $\alpha 2(2 \times 4)$, and $\zeta(4 \times 2)$ surface reconstructions and a 21-layer slab for the $\epsilon(4 \times 4)$ surface reconstruction. These slabs had a reconstructed surface on one side and the other side was passivated with pseudo hydrogen atoms. All of these supercell models contained more than 300 atoms. For the $2 \times 2$ supercell of oxidized InSb(110), the largest model that was possible to calculate was a 22-layer slab. A comparison between the band structures of 42-layer and 22-layer slabs of bare InSb(110) is provided in Supporting Information. Although the band gap values are different, no significant difference is found in position of the valence band maximum.

2.4. Experimental Details

The ARPES measurements were performed at the soft-X-ray ARPES facility of the ADRESS beamline of the Swiss Light Source, PSI, Switzerland. To maximize the coherent spectral fraction, impaired by the thermal atomic motion, the measurements were performed at 12 K. The combined (beamline and analyzer) energy resolution varied from 50 meV at photon energies $h \nu$ around 400 eV to 90 meV around 700 eV, with an angular resolution of the ARPES analyzer of 0.1°. Other relevant details of the soft-X-ray ARPES experiment, including the experimental geometry, can be found in ref. [108]. An intensity transfer function provided in Supporting Information has been applied to the ARPES data shown here. The background intensity was removed by subtracting a fraction of the angle-integrated spectrum. The low energy electron diffraction (LEED) measurements were performed at $T = 12$ K with a SPECS ErLEED 150 system in the same vacuum system where the ARPES measurements were performed.

For InAs(001), the samples were epi-ready wafers supplied by “Wafer Technology Ltd” (UK). The wafers were undoped n-type with a nominal carrier density of $n = (1-3) \times 10^{16}$ cm$^{-3}$. They were cleaned in ultra-high vacuum for 20 min at $T = 250$ °C under a stream of atomic hydrogen that was produced by an MBE-Komponenten hydrogen atomic beam source (HABS) at a hydrogen pressure of $p \approx 1 \times 10^{-5}$ mbar. For InSb(110), the samples were n-type nominally undoped epi-ready wafers supplied by “Wafer Technology Ltd” (UK). The samples were cleaved in UHV to expose a fresh, mirror-like surface. For InAs(111), the samples were n-type nominally undoped epi-ready wafers supplied by “Wafer Technology Ltd” (UK). They were subsequently hydrogen cleaned and an epilayer of InAs was grown in a molecular-beam epitaxy chamber. The samples were then shipped in air to the ARPES facility and cleaned in ultra-high vacuum for 20 min at $T = 250$ °C under a stream of atomic hydrogen at a hydrogen pressure of $p = 1 \times 10^{-5}$ mbar.

3. Results and Discussion

3.1. InAs(001)

Figure 4 shows the LEED pattern obtained for the InAs(001) sample. Comparison to simulated patterns suggests coexisting domains of $2 \times 4$ and $4 \times 2$ reconstructions. This is consistent with earlier reports in the literature that such domains can coexist. The $x2$ orders appear smeared for both reconstructions, which indicates the presence of disorder at the surface. Whilst the LEED pattern reveals the size and orientation of the surface reconstruction, the precise atomic structure cannot easily be inferred from it, because LEED is only sensitive to the surface unit cell geometry and not the atomic positions inside the unit cell.

DFT simulations were conducted to investigate the influence of surface reconstruction on the band structure. We have focused on the $\beta 2(2 \times 4)$ reconstruction, illustrated in Figure 1a, which is considered as a stable $2 \times 4$ surface reconstruction in As-rich conditions, the $\alpha 2(2 \times 4)$ reconstruction, shown in Figure 5a, which is considered as a stable $2 \times 4$ surface reconstruction in As-poor conditions, the $\zeta(4 \times 2)$ reconstruction, shown in Figure 5b, which was predicted to be the most stable $4 \times 2$ surface reconstruction in In-rich conditions by DFT calculations, and the $\epsilon(4 \times 4)$ reconstruction, shown in Figure 5c, which was predicted to be stable in As-rich conditions by DFT calculations and observed experimentally. The results are compared to the ARPES experiment in Figure 6. The ARPES spectra were measured at a photon energy of $h \nu = 405$ eV, which is known to correspond to the $k_z = 0$ plane in the 3D
Brillouin zone of InAs(001),\textsuperscript{[117]} which means that both the quantum well subbands derived from the conduction band, as well as the dispersion at the top of the valence band can be observed. The bulk-unfolded band structures corresponding to different surface reconstructions show distinctive signatures of surface states. The $\beta2(2 \times 4)$ band structure shown in Figure 1d, the $\alpha2(2 \times 4)$ band structure shown in Figure 6a and the $\zeta(4 \times 2)$ band structure, shown in Figure 6b, exhibit surface states concentrated mainly around the valley of the heavy hole band between $\Gamma$ and $X$. Both of these reconstructions do not produce surface states near the valence band maximum (VBM). In contrast, the $\zeta(4 \times 2)$ band structure, shown in Figure 6c, exhibits surface states not only in the valley between $\Gamma$ and $X$ but also at the top of the valence band at the $\Gamma$ point. None of the reconstructions studied here produces surface states in the band gap. Hence, the surface reconstruction does not affect the position of the Fermi level.

Overall, the main features of the PBE+U(BO) band structures are in good agreement with ARPES, although the bandwidth of the valence band is underestimated in our calculations. The $\beta2(2 \times 4)$, $\alpha2(2 \times 4)$, and $\zeta(4 \times 2)$ band structures are in somewhat better agreement with experiment than the $\zeta(4 \times 4)$ band structure because no additional features are experimentally observed near the VBM. We note, however, that the relatively high photon energies used in our experiment are insensitive to the signal from surface states because the extended mean free path of the photoelectrons results in small overlap with states that are sharply localized at the surface.\textsuperscript{[118]} Therefore, the comparison of the simulated band structures with different reconstructions to ARPES is inconclusive. It would be possible to better distinguish the type of surface reconstruction from the type of surface states with lower photon energies. In fact, the surface states of InAs(001) $c(8 \times 2)/2 \times 4$ have been observed with VUV-ARPES at a lower photon energy of 21 eV.\textsuperscript{[59]} PBE+U(BO) underestimates the bandwidth of the heavy hole band, consistent with the trend shown in Figure 2 and in the Supporting Information. Although a hybrid functional would produce a band width in closer agreement with experiment, it is currently unfeasible to conduct such calculations due to the prohibitive computational cost. The top of the VBM in the ARPES is about 0.2 eV lower than in the simulated band structures and a small electron pocket is seen in the ARPES, which is not present in the simulated band structures. This may be caused by the presence of disorder, as indicated by the LEED results, which could lead to Fermi-level pinning.\textsuperscript{[119]}

### 3.2. InAs(111)

Figure 7 shows the LEED pattern obtained for the InAs(111) sample. Comparison to a simulated pattern indicates that the surface is not reconstructed. Because this sample was grown on the B-side substrate, we assume that the surface is As-terminated. Here
Figure 7. a) LEED results for InAs(111). b) A simulated LEED pattern corresponding to a 1×1 surface.

Figure 8. Electronic structure of InAs(111): ARPES results obtained for a) fresh hydrogen cleaned surface and b) oxidized surface. The ARPES data is compared with bulk-unfolded PBE+U(BO) band structures of c) bare InAs(111) and d) InAs(111)+O. A is the point along \( K - \Gamma \) with the coordinates \( (\frac{1}{3}, \frac{1}{3}, 2) \). The band path \( A - \Gamma - A \) for bulk unfolding is mapped to \( \bar{K} - \bar{\Gamma} - \bar{K} \) in the (111) projection. The corresponding surface Brillouin zone is provided in Supporting Information.

Figure 9. Top and side view of the structure of a) bare InAs(111) and b) oxidized InAs(111).

we show the results from two ARPES experiments: Figure 8a shows a measurement of the bare surface after hydrogen cleaning, whilst Figure 8b shows a measurement along the same direction after exposing the surface to 1800 L of pure oxygen in the UHV chamber. One can clearly see that after oxygen exposure, the valence band shifts to higher binding energies and the size of the electron pocket increases due to charge accumulation.

To model oxidized InAs(111), three oxygen adsorption sites have been considered: O bound to a surface As atom, O bound to a surface In atom, and O bound to one As and two In atoms. The model from ref. [120] is not considered here because it is of an In-terminated surface. The first model, shown in Figure 9, is predicted to be the most stable by DFT calculations. The other two models are shown in the Supporting Information. To explain the difference between the two ARPES experiments, we compare the band structures of a 50-layer As-terminated InAs(111) slab passivated with pseudo hydrogen atoms on both sides and a slab with an oxygen atom bound to As on one side and the other side passivated. The resulting band structures show in Figure 8c,d, respectively. Overall, the PBE+U(BO) results are in good agreement with ARPES. The bandwidth is somewhat underestimated by PBE+U(BO). For bare InAs(111) in Figure 8c, the VBM is located at \(-0.24\) eV, whereas for oxidized InAs(111) in Figure 8d, the VBM is located at \(-0.47\) eV. This indicates that oxidation would lead to band bending, in agreement with the corresponding ARPES results. Oxidation introduces surface states in the band gap, which are not present in the band structure of the “bare” surface slab. A significant electron pocket is seen in the simulated band structure of the oxidized surface, in agreement with ARPES. A small electron pocket is also seen in the ARPES data in Figure 8a, which is not present in the simulated band structure of the bare InAs(111) surface. In this case, there is no evidence of disorder based on the LEED experiment. However, it is possible that the hydrogen cleaning may create some defects that are not taken into account in our simulations.

To explain why surface oxidation causes band bending, we examined the Bader charge difference between InAs(111)+O and InAs(111). The charge difference for a specific atom A, \( C_{\text{diff}}(A) \), is given by

\[
C_{\text{diff}}(A) = C_{\text{oxidized}}(A) - C_{\text{bare}}(A)
\]

where C is the atom’s Bader charge. The charge differences of the surface As atom bonded to O and the In atom in the second layer from the surface are \(-1.08e\) and \(+0.11e\), respectively. For
the oxidized surface, there is significant charge transfer from the surface As atom to the O atom, accompanied by redistribution of the charge on the In atom below the surface. This leads to increased band bending and produces an electron pocket.

### 3.3. InSb(110)

Figure 10 shows the LEED pattern obtained for the InSb(110) sample. Comparison to a simulated pattern indicates that the surface is not reconstructed. Similar to InAs(111), here we show the results from two ARPES experiments: Figure 12a shows a measurement of the clean surface after hydrogen cleaning, whilst Figure 12b shows a measurement along the same direction after exposing the surface to 1350 L of pure oxygen in the UHV chamber. In contrast to InAs(111), no significant change of the valence band maximum position and no electron pocket are observed for the oxidized InSb(110) surface.

To model oxidized InSb(110), two oxygen adsorption sites have been considered: O bound to a surface Sb atom and O bound to a surface In atom. The first model, shown in Figure 11, is predicted to be the most stable by DFT calculations. The other model is shown in the Supporting Information. To elucidate the effect of surface oxidation, we compare the band structures of a 42-layer InSb(110) slab, whose surfaces were passivated with pseudo-hydrogen atoms and a (2×2) supercell slab with an oxygen atom bound to Sb on one side and the other side passivated. The resulting band structures are shown in Figure 12c,d, respectively. Overall, the PBE+U(BO) results are in good agreement with ARPES for the main features of the band structure. Good agreement with ARPES is obtained for the effect of spin-orbit coupling and the position of the split-off band. PBE+U(BO) underestimates the bandwidth and band curvature, consistent with the trend shown in Figure 2 and in the Supporting Information.

In agreement with ARPES, the VBM position does not differ significantly between the bare InSb(110) and oxidized InSb(110) simulated band structures. In contrast to InAs(111), no states appear in the band gap of InSb(110) upon oxidation. In addition, an electron pocket does not appear in the InSb(110)+O band structure, indicating that oxidation does not lead to increased band bending. A surface state that appears at the top of the valence band in the simulated InSb(110)+O band structure is not observed in ARPES, possibly because the photon energy used in the experiment is insensitive to surface states. To explain

---

**Figure 10.** a) LEED results for InSb(110). b) A simulated LEED pattern corresponding to a 1×1 surface.

**Figure 11.** Top and side view of the structure of a) bare InSb(110) and b) InSb(110)+O (2×2).

**Figure 12.** Electronic structure of InSb(110): ARPES results obtained for a) freshly cleaved surface and b) oxidized cleaved surface. The ARPES coherent spectral component was enhanced by subtracting the incoherent k-integrated one approximated by the angle-integrated spectrum, compared with bulk-unfolded PBE+U(BO) band structures of c) bare InSb(110) and d) InSb(110)+O. B is the point along X–Γ with the coordinates (0.25, 0.25, 0). The band path B–Γ–B for bulk unfolding is mapped to Y–Γ–Y in the (110) projection. The corresponding surface Brillouin zone is provided in Supporting Information.
the difference in the effect of oxidation between InAs(111) and InSb(110), we examined the Bader charge difference using Equation (11). The charge difference of the surface Sb atom bonded to O is $-0.79e$ and no significant charge difference is found in the second layer below the surface. This indicates that oxidation does not induce as much charge redistribution in InSb(110) as in InAs(111). Interestingly, the Fermi-level pinning position we find for the oxidized samples is in good qualitative agreement with the branching point theory popularized by Tersoff and others, which predicts Fermi-level pinning in the conduction band for InAs$^{[121,122]}$ and pinning just above the valence band for InSb.$^{[122,123]}$

4. Conclusion

In summary, we have studied the electronic structure of the InAs(001), InAs(111), and InSb(110) surfaces using DFT simulations and ARPES experiments. The DFT simulations employed a recently developed method of adding a Hubbard $U$ correction determined by machine learning to a semi-local functional. The optimal values of $U_{\text{eff}}$ for InAs and InSb were found by a Bayesian optimization algorithm designed to reproduce as closely as possible the band gap and band structure features obtained with the HSE hybrid functional for the bulk materials. The efficiency of PBE+$U(BO)$ enabled us to perform simulations of surface slab models with several hundred atoms, which would not be feasible with HSE or GW due to their high computational cost.

To compare the resulting band structures to ARPES experiments, we have presented a “bulk unfolding” scheme, which unfolds the calculated band structure of a surface slab supercell onto the bulk primitive cell. For all three surfaces, the bulk-unfolded band structures calculated with PBE+$U(BO)$ were found to be in good agreement with bulk sensitive soft X-ray ARPES experiments. This confirms the transferability of the $U_{\text{eff}}$ values obtained for the bulk materials to surface slabs and demonstrates the accuracy of PBE+$U(BO)$. For InAs(001) the effect of surface reconstruction on the electronic structure was studied. The simulated band structures revealed the distinctive signatures of the $\beta 2(2 \times 4)$ reconstruction, the $\alpha 2(2 \times 4)$ reconstruction, the $\zeta (4 \times 2)$ reconstruction, and the $c(4 \times 4)$ reconstruction. The surface reconstructions studied here do not produce any states in the band gap and therefore they do not lead to band bending. The $c(4 \times 4)$ reconstruction is predicted to exhibit a surface state at the top of the valence band, which is not observed in ARPES experiments. Therefore, the results of our simulations support the coexistence of $2 \times 4$ and $4 \times 2$ domains. To observe the surface states that were predicted by our calculations, more surface sensitive ARPES measurements at lower photon energies are needed in future work.

For InAs(111) and InSb(110), we studied the effect of surface oxidation on the electronic structure. ARPES experiments and DFT simulations show that oxidation of InAs(111) leads to increased band bending and a larger electron pocket, whereas oxidation of InSb(110) does not significantly affect the position of the valence band maximum and does not produce an electron pocket. The simulated band structures show that surface states appear in the band gap of the oxidized InAs(111) slab, whereas no states appear in the gap of the oxidized InSb(110) slab. The band bending may be attributed to more significant charge transfer from the surface As to O in InAs(111) than from the surface Sb to O in InSb(110), as well as to the presence of gap states. Our systematic theoretical and experimental comparison of surface oxidation in InAs and InSb indicates that charge accumulation is preferred in native oxides of InAs, whilst native oxides of InSb exhibit pinning in the gap, consistent with earlier theoretical results of branching point theory. Our calculations reveal that the microscopic origin of this effect appears to be an increased surface charge redistribution for the native oxide of InAs which is less pronounced in the native oxide of InSb.

Our results have important implications for proximitized semiconductor nanowire devices. The results elucidate the underlying properties of III–V semiconductor surfaces used to create proximitized p-wave superconductors. In particular, they are important for understanding the Fermi-level pinning resulting from reconstructions which, in turn, is relevant for tuning such devices into the topological regime. Similarly, our results on effect of oxidation should underpin thinking about the tunnel junction physics, which controls coupling to Majorana modes. Both of these insights are crucial for the conceptual understanding of device performance as well as for guiding topological qubit design optimization. The results presented here demonstrate that PBE+$U(BO)$ is an efficient and reliable method for simulations of III–V semiconductor surfaces. PBE+$U(BO)$ may be used in the future to conduct simulations of additional surfaces and interfaces. This could advance our understanding of how the electronic structure is affected by atomic structural features. Thus, first principles simulations combined with spectroscopy may aid the discovery and realization of new materials systems for quantum computing.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Work at CMU was funded by the National Science Foundation (NSF) through grant OISE-1743717. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. This work is based on support by the U.S. Department of Energy, Office of Science through the Quantum Science Center (QSC), a National Quantum Information Science Research Center.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

angle-resolved photoemission spectroscopy, density functional theory, surface physics, quantum materials, III–V semiconductors
[116] P. De Padova, P. Perfetti, C. Quaresima, C. Richter, M. Zerrouki, K. Hricovini, in Proc. of the 14th Indium Phosphide and Related Materials Conf., IEEE, Piscataway, NJ 2002, pp. 597–600.

[117] S. Schuwalow, N. B. M. Schröter, J. Gukelberger, C. Thomas, V. Strocov, J. Gamble, A. Chikina, M. Caputo, J. Krieger, G. C. Gardner, M. Troyer, G. Aeppli, M. J. Manfra, P. Krogstrup, Adv. Sci. 2021, 8, 2003087.

[118] V. Strocov, J. Electron. Spectrosc. Relat. Phenom. 2018, 229, 100.

[119] R. E. Allen, O. F. Sankey, J. D. Dow, Surf. Sci. 1986, 168, 376.

[120] A. V. Bakulin, S. E. Kulkova, M. S. Aksenov, N. A. Valisheva, J. Phys. Chem. C 2016, 120, 17491.

[121] J. Tersoff, Surf. Sci. 1986, 168, 275.

[122] W. Mönch, J. Appl. Phys. 1996, 80, 5076.

[123] J. Tersoff, Ph.D. Thesis, American Vacuum Society, New York 1986.