

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Microscopic Origin of Polarization Charges at GaN/(Al,Ga)N Interfaces

Permalink

<https://escholarship.org/uc/item/18c3j80s>

Journal

Physical Review Applied, 19(6)

ISSN

2331-7043

Authors

Yoo, Su-Hyun

Todorova, Mira

Neugebauer, Jörg

et al.

Publication Date

2023-06-01

DOI

10.1103/physrevapplied.19.064037

Supplemental Material

<https://escholarship.org/uc/item/18c3j80s#supplemental>

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Microscopic origin of polarization charges at GaN/AlGaIn interfaces

Su-Hyun Yoo*

*Department of Materials, Imperial College London, SW7 2AZ London, United Kingdom and
Department of Computational Materials Design, Max-Planck-Institut für Eisenforschung GmbH,
Max-Planck-Strasse 1, D-40237 Düsseldorf, Germany*

Mira Todorova and Jörg Neugebauer

*Department of Computational Materials Design, Max-Planck-Institut für Eisenforschung GmbH,
Max-Planck-Strasse 1, D-40237 Düsseldorf, Germany*

Chris G. Van de Walle†

Materials Department, University of California, Santa Barbara, CA 93106-5050, USA

(Dated: May 13, 2023)

GaN/AlGaIn heterojunctions are at the heart of high-electron-mobility transistors that are being adopted for high-power and high-frequency applications. The strong polarization fields present at this interface significantly enhance the density of the two-dimensional electron gas (2DEG) that is confined on the GaN side of the junction. The microscopic origin of these electrons has been debated over the years: after excluding that they would be contributed by bulk donors, a model that identifies surface states on the AlGaIn surface as the source of electrons has become widely adopted. Recently it has become clear, however, that the measured density of surface states is insufficient to account for the high electron density in the 2DEG. Here we demonstrate, based on state-of-the-art first-principles calculations, that the source of electrons is intrinsic to the overall structure and that the negative charge in the 2DEG is balanced by fixed charge on the surface. We perform a rigorous study of polarization, using our recently developed methodology for quantifying polarization fields within the finite-sized systems that can be addressed with density functional calculations. The results show that the electrons that appear in the 2DEG originate locally at the interface, and that the net charge at the interface is predominantly compensated by fixed charge on the surface, rather than surface states. We elucidate the source of this fixed charge and associate it with surface reconstructions or the presence of heterovalent impurities (such as oxygen). Our results force a reassessment of the impact of surface states on the density of the 2DEG: rather than serving as a supply of electrons, the surface states mainly act to pin the Fermi level. Our conclusions allow a fresh interpretation of experimental observations and allow devising guidelines for optimizing carrier densities in the 2DEG.

I. INTRODUCTION

GaN/AlGaIn high electron mobility transistors (HEMTs) are a subject of intense research, because their high electron density and large breakdown field allow applications in high-power and high-frequency devices [1]. In spite of the technological importance, the origin of the electrons in the two-dimensional electron gas (2DEG) has remained a subject of debate. In conventional HEMTs, intentional doping generates the sheet of highly mobile electrons at the interface. In the GaN/AlGaIn system, however, no extrinsic doping is required [2, 3]. Over the years, it has become widely accepted that the 2DEG originates from donor states on the AlGaIn surface [4–12]. In the schematic band diagram shown in Fig. 1, electrons would be transferred from surface states on the AlN surface to the 2DEG, and the Fermi level would be pinned within the surface-state distribution.

Since the electron density in the 2DEG can exceed $2 \times 10^{13} \text{ cm}^{-2}$, the “surface donor model” requires that the density of surface states be at least as large. However, measurements of surface-state densities have systematically produced much lower values. Using photo-assisted capacitance-voltage measurements, Swenson and Mishra [13] and Yeluri *et al.* [14] reported a density of $2 \times 10^{12} \text{ cm}^{-2}$, and Hossain *et al.* [15] reported $1.2 \times 10^{13} \text{ cm}^{-2}$. More recently, Ber *et al.* [16] used UV-assisted gated van der Pauw experiments to map out the surface-state density as a function of energy in the band gap, again finding that the concentration of surface states is too low (by as much as two orders of magnitude) to explain the observed variations as a function of barrier thickness. They proposed that intrinsic fixed surface charge is needed to explain the discrepancy between the observed density of the 2DEG and the low density of surface states. Similar conclusions about low densities of surface states have been reached based on measurements of actual device structures. Ganguly *et al.* [17] performed careful experiments on GaN/AlN/oxide heterostructures and concluded that fixed charge was present at the oxide-AlN interface. Bakeroot *et al.* [18] compared metal-insulator-semiconductor (MIS) transistor characteristics

* syoo@ic.ac.uk

† vandewalle@mrl.ucsb.edu

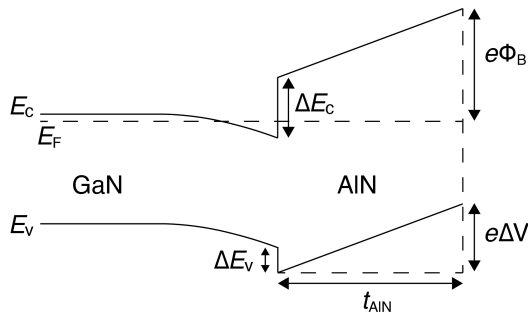


FIG. 1. Schematic band diagram illustrating the formation of a 2DEG at a GaN/AlN heterostructure. We adopt the convention that the notation A/B for a heterostructure implies that semiconductor B is grown on top of A. The slope of the bands in the AlN layer is determined by the net charge at the interface (polarization charge minus electron density). $e\Phi_B$ is the surface barrier height, and ΔV is the voltage drop over the AlN layer. The valence-band offset ΔE_v and conduction-band offset ΔE_c are indicated. t_{AlN} is the thickness of the AlN layer.

to simulations and concluded that agreement with experiment could only be obtained if they assumed a low density of surface states (containing mobile charge) in combination with fixed charge.

The experimental evidence thus strongly points to the presence of fixed surface charge. However, the microscopic origin of this surface charge has not yet been elucidated. In this paper we demonstrate, based on state-of-the-art first-principles calculations, how charge stabilization on the surface occurs. Our explicit implementation uses “pseudohydrogen” atoms with fractional valence [19] and can be mapped directly onto realistic surfaces where bound states would be stabilized via surface reconstructions or bonding with an oxide overlayer. Our methodology allows rigorously quantifying polarization fields within the finite-sized systems that can be addressed with density functional calculations; we perform slab calculations both for systems consisting of a single material and for GaN/AlN heterostructures.

Our results will allow us to explain a wide range of experimental observations, and provide pointers for optimizing 2DEG densities.

II. METHODOLOGY

Our calculations are based on density functional theory (DFT) as implemented in the Vienna *Ab initio* Simulations Package (VASP) code [20, 21] with the generalized gradient approximation (GGA) of Perdew, Burke, and Ernzerhof (PBE) [22]. These calculations use projector-augmented-wave potentials [23] with a plane-wave energy cutoff of 500 eV; the Ga *d* states are included in the valence. For the 4-atom wurtzite (WZ) primitive cells, the Brillouin zone is sampled with a $(8 \times 8 \times 6)$ Γ -centered grid; for the supercells used for slab calculations,

a $(8 \times 8 \times 1)$ grid is employed. All calculations for the electronic relaxation are performed until the total energy convergence is less than 10^{-7} eV per surface unit cell.

A vacuum region of 18 Å separates adjacent slabs, and both surfaces of a slab are passivated by pseudohydrogen atoms, as discussed in detail in Sec. III A. For the GaN slabs, all Ga and N atoms are kept fixed at their bulk positions, while the pseudohydrogen atoms are relaxed during structural optimization.

The GaN/AlN interface structure is modeled using slabs consisting of 15 bilayers (BL) of GaN, 7 BL of AlN and 18 Å of vacuum in the supercell. One bilayer consists of one atomic layer of Ga (Al) and one atomic layer of N and corresponds to a thickness of $c/2$ in the bulk WZ structure. **Tests performed in Ref. 19 showed that quantum confinement has negligible impact on the results at these layer thicknesses.** During structural relaxations, the five outermost Ga-N BLs of the N-polar $(000\bar{1})$ surface are fixed at their bulk positions, as we assume GaN to be the substrate. All remaining atoms are relaxed.

The dipole correction scheme [24] is used in all slab calculations to remove artificial interactions between two slab surfaces through the vacuum region. Our tests confirm that the application of the dipole correction scheme renders the electric field in the slab independent of the vacuum thickness.

Within this set-up, we obtain for the bulk lattice parameters $a = 3.216$ Å, $c/a = 1.629$ and $u = 0.377$ for GaN, and $a = 3.128$ Å, $c/a = 1.604$ and $u = 0.381$ for AlN, in good agreement with previous computations [25] and with experiment [26, 27]. For AlN strained to match the in-plane lattice parameter of GaN we obtain $c/a = 1.534$ and $u = 0.389$. The calculated band gaps are $E_g = 1.72$ eV for GaN, 4.06 eV for AlN and 3.50 eV for strained AlN. They are underestimated compared to the experimental band gaps [28], as expected for the PBE functional, but in good agreement with computations at the same level of theory [29]. The spontaneous polarization constants for GaN and AlN, and the polarization of strained AlN (s-AlN), are calculated by performing Berry phase calculations [30, 31] and referencing to corresponding hexagonal ($P_{\text{eff}}^{\text{H}}$) and zinc-blende (ZB) ($P_{\text{eff}}^{\text{ZB}}$) phases. The obtained values are listed in Table I. A Table comparing these values with previous calculations [25, 32, 33] is included in the Supplemental Material [34].

III. RESULTS FROM FIRST-PRINCIPLES CALCULATIONS

A. Slab calculations for GaN(0001) surfaces

To perform calculations for systems in which polarization fields are manifested, a slab geometry is typically adopted, which allows studying surfaces and interfaces while still adhering to periodic boundary conditions. Using hydrogen atoms to passivate the surface of such slabs is a well-established procedure. In a

TABLE I. Calculated spontaneous polarization constants $P_{\text{eff}}^{\text{H}}$ and $P_{\text{eff}}^{\text{ZB}}$ for GaN and AlN, and comparison with previous calculations. Polarization values for strained AlN (s-AlN) are also included, where the strain in AlN is induced by lattice matching to GaN. The superscript indicates the reference phase, either the hexagonal phase (H) or the unstrained zincblende phase (ZB). The polarization constants are expressed either in units C/m^2 or in units of electrons per area of a surface unit cell (e/uc).

System	$P_{\text{eff}}^{\text{H}}$ (C/m^2)	$P_{\text{eff}}^{\text{H}}$ (e/uc)	$P_{\text{eff}}^{\text{ZB}}$ (C/m^2)	$P_{\text{eff}}^{\text{ZB}}$ (e/uc)
GaN	1.309	0.732	-0.033	-0.018
AlN	1.329	0.703	-0.089	-0.047
s-AlN	1.202	0.672	-0.139	-0.078

slab geometry, a number of layers of the material constitute a slab in a given orientation, surrounded by a finite thickness of vacuum (which is chosen to be large enough so that interactions between surfaces across the vacuum layer are minimized). When such calculations are performed for a monatomic semiconductor, hydrogen atoms can be used to passivate the surface (i.e., generate a surface band structure free of any surface states in the band gap) [35, 36].

For instance, in the case of Si(111), an unpassivated (1×1) surface would contain a Si atom with one dangling bond (db), filled with a single electron. Passivating that db with a H atom creates a two-electron bond, with a bonding state deep in the valence band. For a zincblende (ZB) surface, for instance of GaN, the same logic demands that a db on a Ga atom, which is filled with $3/4$ electron on the neutral surface, needs to be passivated with a “pseudohydrogen” (psH) atom with a core charge of $5/4 e$ (we will use e to denote an electron charge, with e representing a positive number) surrounded by $5/4$ electrons. This arrangement will again lead to a neutral surface with a strong two-electron bond between the Ga atom and the $\text{psH}_{1.25}$ atom. In the absence of the passivation, the $3/4$ electron in the Ga db would be unstable and, in the case of a finite (111) slab, transfer to the N-polar side (where it can fill the N db containing $5/4$ electron); in the process, a large field would be set up. What the passivation with $\text{psH}_{1.25}$ accomplishes is that the $3/4$ electron in the Ga db, which is necessary to maintain local charge neutrality, is kept on the Ga-polar surface.

This procedure has proven very useful for studying surface reconstructions in cases where slabs in a given orientation do not exhibit mirror symmetry [37]. In order to extract properties of a single surface, the back surface of the slab is passivated with appropriate psH atoms, while various surface reconstructions are studied on the top surface, allowing an analysis of differences in atomic and electronic structure and energetics.

This same procedure has been adopted for studying surface reconstructions on WZ phase semiconductors. In spite of sporadic comments about potential problems

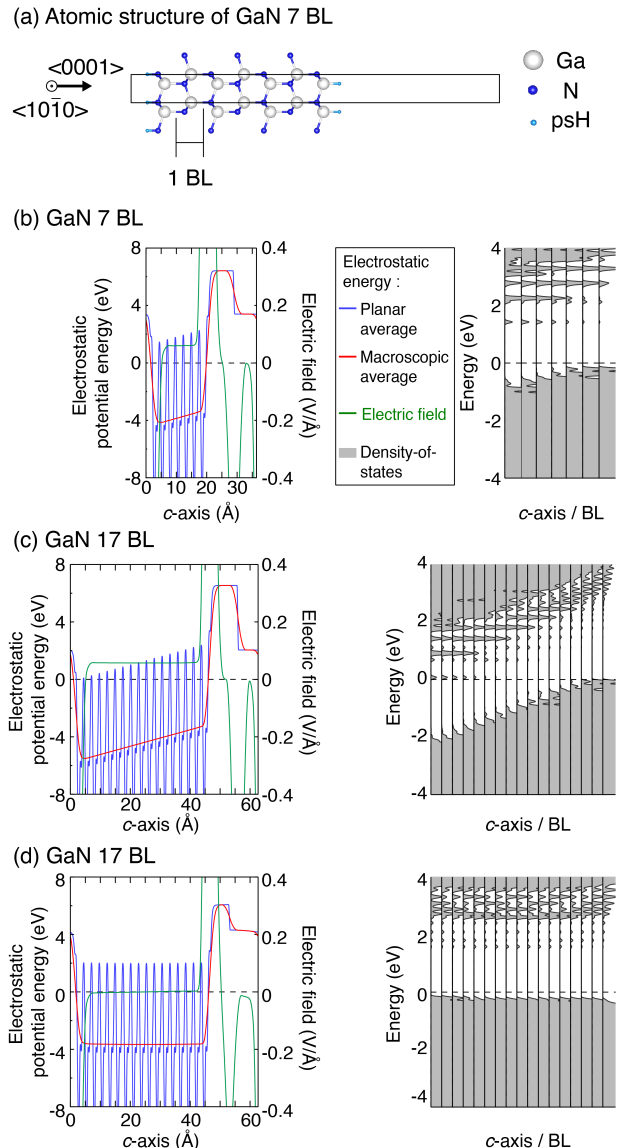


FIG. 2. (a) Atomic structure of a GaN(0001) 7 BL slab where both surfaces are passivated with pseudohydrogen atoms. (b)-(d) Electrostatic potential profile and layer-resolved density of states (LDOS) for GaN(0001) slabs with various thicknesses, terminated by a Ga-polar (0001) surface on the right and a N-polar (0001-bar) surface on the left. (b) 7 BL slab passivated with $\text{psH}_{0.75}$ and $\text{psH}_{1.25}$. (c) 17 BL slab passivated with $\text{psH}_{0.75}$ and $\text{psH}_{1.25}$. (d) 17 BL slab passivated with $\text{psH}_{0.732}$ and $\text{psH}_{1.268}$. The electric field, the planar- and the macroscopic-averaged electrostatic potential energies are shown as green, blue, and red solid lines, respectively. The behavior of the electric field in the vacuum region reflects the application of the dipole correction (see Sec. II). The LDOS along the c axis (axis perpendicular to the slab surfaces) is resolved for each Ga-NBL and shown as a grey region. The zero of energy in the DOS plots is set to the Fermi level.

arising from the presence of polarization fields [38–40], any issue relating to the presence of finite fields has been generally ignored in the literature. Figure 2 illustrates that this traditional passivation scheme has its shortcomings. Figure 2(b) depicts the results of a simulation for a slab [depicted in Fig. 2(a)] consisting of 7 BL of GaN in the (0001) direction. The 7 BL slab is terminated on the left with psH_{0.75} and on the right with psH_{1.25}. Although strong two-electron bonds are being formed [as evidenced in the layer-resolved density of states (LDOS) by the absence of any surface state, and only valence-band states being occupied], a significant electric field is present over the slab, with a magnitude $\mathcal{E}=59.83$ mV/Å. Gauss’ law tells us that this field is related to the presence of a surface charge: $\Delta\mathcal{E} = \sigma/(\epsilon_0\epsilon_r)$, where σ is the surface charge, ϵ_0 is the dielectric permittivity of vacuum, and ϵ_r is the relative dielectric constant of GaN. For consistency we use the calculated value $\epsilon_r=6.10$; this is the electronic (high-frequency) part of the dielectric constant, consistent with the absence of atomic relaxation in the GaN slab calculations. Using the observed value of $\Delta\mathcal{E}$, we obtain $\sigma = -0.018 e/\text{uc}$ (where “uc” stands for the area of a surface unit cell).

This value turns out to correspond to the predicted polarization charge on a GaN(0001) WZ surface as obtained from a Berry phase calculation for bulk WZ GaN based on the modern theory of polarization [30, 31] (see Table I). The surface bound charge on the psH_{0.75}/psH_{1.25}-passivated slab corresponds to the polarization charge when the latter is determined using the ZB phase as reference [25]; more about this in Sec. III C. We see that the passivation with psH_{0.75} and psH_{1.25} has managed to provide *insulating* surfaces, on which the polarization charge induced by the spontaneous polarization of WZ GaN can be directly observed (as described in Sec. III of Ref. 30).

Figure 2(c) depicts the results of a similar simulation on a thicker GaN slab, consisting of 17 BL of GaN, still terminated with psH_{0.75} and psH_{1.25}. As expected, the presence of the field has now led to a voltage drop over the slab that exceeds the band gap, and breakdown has occurred, as clearly observed in the layer-resolved DOS plot: electrons from the valence band at the right surface are transferred into the conduction band at the left surface. This clearly illustrates the pitfalls of using the traditional passivation scheme.

These considerations also suggest a methodology to avoid the presence of a field: apply an additional fixed positive (negative) charge on the right (left) surface to compensate the polarization-induced surface bound charge. Since (referenced to the ZB phase) the polarization charge on the GaN(0001) surface is $\sigma_{\text{GaN}} = -0.018 e/\text{uc}$ (Table I), the magnitude of this additional fixed charge should be $+0.018 e$ per unit cell. On a realistic surface, in the absence of impurities, this compensating charge can be provided by reconstructing the surface to create bound states (i.e., states associated with atomic bonding); we will return to this point in Sec. V.

Describing such a scenario directly is unfeasible in a DFT calculation; indeed, describing a charge of $0.018 e$ would require over 100 unit cells to model a system with integer electrons to an accuracy of $0.001 e/\text{uc}$, a system size beyond the capabilities of DFT. We therefore propose an approach that allows us to introduce small fractional fixed charge within a (1×1) surface unit cell.

B. Pseudohydrogen passivation for wurtzite semiconductors

Within the DFT methodology, we can introduce *neutral* entities consisting of a fractionally charged ionic core surrounded by the corresponding compensating fractional electron charge. The psH_{0.125} atom that we introduced to passivate a cation-polar ZB surface is one example. This suggests introducing a psH_{1.268} atom for passivating the Ga db on the WZ GaN(0001) surface, and a psH_{0.732} atom for passivating the N db on the (000 $\bar{1}$) surface. The net result, fully confirmed by the simulations shown in Fig. 2(d), is that we find an insulating slab with insulating surfaces (i.e., a system with a band gap common to the “bulk” and the surfaces), in which all electrons are locked in two-electron bonds (valence states), and with surfaces that are *neutral* because the polarization charge has been fully compensated by fixed charges.

We can connect our approach to generating an insulating system with the concept of “electron counting” (EC), which has been fruitfully used to explain and predict surface reconstructions [41]. The electron-counting rule (ECR) states that a surface will adopt a low-energy structure by ensuring that surface states close to the valence-band maximum (VBM) are fully occupied, and states close to the conduction-band minimum (CBM) are completely empty. The situation depicted in Fig. 2(c) thus clearly violates the ECR. Our passivation scheme in Fig. 2(d) leads to a situation where valence-band states are fully occupied and conduction-band states completely empty, thus satisfying the ECR.

We suggest that the passivation scheme proposed here, which leads to surfaces that are both insulating and neutral, is the appropriate way to perform calculations for surface properties of semiconductors that exhibit spontaneous polarization (see Ref. 19 for more details of the proposed passivation scheme). A completely analogous study for AlN slabs shows that the Al-polar (0001) surface should be passivated with psH_{1.297} and the N-polar (000 $\bar{1}$) with psH_{0.703}. The differences from the conventional psH valencies of 1.25 and 0.75 again correspond to the polarization charge obtained from a Berry phase calculation, $\sigma = -0.047 e/\text{uc}$, when the ZB phase is used as a reference (Table I).

C. The concept of fixed charge

We demonstrated that neutral, insulating surfaces of a WZ GaN slab can be obtained by passivating the surfaces with pseudohydrogen atoms with appropriate valence. We contend that actual surfaces of materials that exhibit spontaneous polarization will be passivated in a similar manner, i.e., with *fixed charge* compensating the polarization charge. Specific realizations will be discussed in Sec. V. But first we want to elucidate and define this notion of “fixed charge”, building on the results in Sec. III B.

A field-free slab with zero surface charge was obtained by passivation with psH_{1.268} on the Ga-polar (0001) surface. The 1.268 electrons provided by this psH atom form a two-electron bond, implying that before bonding with the psH atom this Ga db contained 0.732 electrons. Since the psH_{1.268} itself is neutral, the (0001) surface with a Ga db containing 0.732 electrons is therefore also neutral. The situation to strive for, if we want a field-free slab, is therefore to have a Ga bond to which the Ga atom would contribute 0.732 electrons. Of course, this situation on the ideal surface, with 0.732 electrons residing in each Ga db, is completely unstable. We therefore need to figure out a way to “fix” this charge, and the psH_{1.268} provides such a way, by locking up the 0.732 electrons in a two-electron bond. One could therefore call the charge corresponding to these 0.732 electrons, equal to $-0.732 e/\text{uc}$, the “fixed charge”. Ultimately, however, it is more convenient (and conventional) to define the fixed charge as the *deviation* of this charge from the ZB case, which would have 0.75 electrons in the db; i.e., we define the fixed charge σ_f as $-0.732 - (-0.75) = +0.018 e/\text{uc}$. With this definition, the fixed charge is the negative of the polarization charge (referenced to the ZB phase) on the WZ GaN(0001) surface.

The main reason for choosing the ZB, rather than the hexagonal phase as the reference for the polarization [25], is that ZB naturally corresponds to the situation without spontaneous polarization, where surface passivation (or charge fixing on the surface by performing surface reconstructions) is well understood and accepted. But ultimately, the use of ZB as a reference is a *choice*. This choice affects the magnitude and sign of the fixed charge, but since the fixed charge can never be observed in isolation it is acceptable to make the choice that fits best with conventional understanding of passivation. Indeed, only the *sum* of polarization charge and fixed charge is observable, and as long as the reference is chosen consistently, this sum is unaffected. Conversion from the hexagonal to the ZB reference is easy if units of e/uc are used, since the formal polarization of the ZB phase (within the modern theory of polarization) is simply $3/4 e/\text{uc}$ [25, 30]. With the ZB reference, the polarization charge on WZ GaN(0001) is $\sigma_{\text{GaN}} = -0.018 e/\text{uc}$, and the fixed charge σ_f that compensates it is $+0.018 e/\text{uc}$; the sum of the fixed charge and the polarization charge is $+0.018 - 0.018 = 0 e/\text{uc}$. With the H reference, the

polarization charge on WZ GaN(0001) is $0.732 e/\text{uc}$ (Table I), and the fixed charge σ_f^H that would compensate it is $-0.732 e/\text{uc}$ (exactly the total electron charge in the Ga db!); the sum of the fixed charge and the polarization charge is therefore $-0.732 + 0.732 = 0 e/\text{uc}$ again. Papers discussing charges that compensate polarization [17, 42] have followed the ZB-reference convention and hence reported positive charges compensating GaN or AlN(0001) surfaces.

When we analyzed Fig. 2(b) in Sec. III A, we found an electric field that corresponded to a surface charge $\sigma = -0.018 e/\text{uc}$. The nonzero surface charge arose from the fact that we passivated with a psH_{1.25} atom. Within the ZB reference, the psH_{1.25} atom corresponds to fixed charge $\sigma_f = 0$, and thus the total charge on the surface is $\sigma = -0.018 e/\text{uc}$, equal to the polarization charge on the GaN(0001) surface.

Finally, we note that our assumption that the Ga db on the GaN(0001) should contain 0.732 electrons for a field-free situation provides an illuminating and intuitive way to look at electron distributions in a material exhibiting spontaneous polarization. Gauss’ law tells us that “field-free” corresponds to “zero net charge” or “neutral”, The Ga atom has three valence electrons; they provide the negative charge $-3e$ that compensates the positive ionic charge of $+3e$ and makes the atom overall neutral. If the Ga db pointing along [0001] contains 0.732 ($=0.75-0.018$) electrons, then to maintain charge neutrality the other three Ga dbs need to contain $(0.75+0.018/3)=0.756$ electrons. We can make a similar argument for the nitrogen atom, based on charge neutrality at the (000 $\bar{1}$) surface: The N db pointing along [000 $\bar{1}$] needs to contain 1.268 electrons, and the other 3 N dbs contain $(1.25-0.018/3)=1.244$ electrons. It is easy to verify that bonds between such Ga and N atoms will all contain two electrons. And since all building blocks are locally neutral, a slab built out of such atoms will be field-free and have neutral surfaces.

A Ga atom with a db along [0001] containing 0.732 electrons, and the other three Ga dbs 0.756 electrons can be considered an atom in which the center of mass of the electron cloud, referenced to the ionic core, has shifted slightly to the left, i.e., towards (000 $\bar{1}$). The electron cloud surrounding the N atom is shifted in a similar fashion. As noted in Ref. 30, such a shift is not just a conceptual notion; it is actually a rigorous way of describing polarization if Wannier centers are used to represent electrons. This way of depicting the Ga and N atoms does not lead to any changes in the interior of a slab, where Ga–N bonds are formed; but on a surface (or at an interface), it provides an intuitive way to visualize a neutral, insulating, field-free situation. In Sec. III D we will see that this forms a sound basis for explaining the source of electrons in the 2DEG at a GaN/AlN (0001) interface.

D. Slab calculations for GaN/AlN (0001) interfaces

We now turn to interfaces. For simplicity we perform simulations of GaN/AlN (0001) interfaces, but all of our argumentation and conclusions also apply to GaN/AlGaN interfaces.

DFT calculations of interfaces are typically performed in a superlattice geometry. However, in the case of a polar orientation such as (0001), the superlattice would contain *two* interfaces, corresponding to the left and right sides of the GaN layer; as is evident from Sec. III A, these two interfaces are very different, and disentangling the properties of single interfaces would be very difficult. In order to study the properties of a *single* interface, we employ a geometry in which a slab containing a single GaN/AlN interface is surrounded by vacuum. Figure 3 depicts the results of a simulation for a finite system consisting of 15 BL of GaN and 7 BL of AlN. Since GaN and AlN have different lattice parameters, the construction of an epitaxial interface requires the presence of strain in one or both of the layers. The greater thickness of the GaN layer reflects our assumption that GaN plays the role of a substrate, i.e., remains unstrained, while the AlN layer is strained to match the in-plane lattice parameter of GaN, and relaxed in the perpendicular direction (along the c axis).

In addition to the interface, the slab has a GaN(000 $\bar{1}$) surface to the left and an AlN(0001) surface to the right; based on the information in Secs. III B and III C we know how to remove the effects of such surfaces from the calculation. The GaN(000 $\bar{1}$) surface on the left is passivated with $\text{psH}_{0.732}$, and the AlN(0001) surface on the right is passivated with $\text{psH}_{1.328}$. These choices for surface passivation were made to ensure that the surfaces would be insulating and neutral. For the GaN(000 $\bar{1}$) surface, the choice corresponds to the passivation scheme discussed in Sec. III B. Since the AlN layer is subjected to strain, piezoelectric polarization is present in addition to the spontaneous polarization. Our calculations for a strained AlN slab indicate that the resulting polarization charge is $-0.078 e/\text{uc}$ and thus the valency of the pseudohydrogen needed to passivate the AlN(0001) surface is 1.328.

Figures 3(a) and (b) show that the system has a band gap throughout the slab, is insulating in the regions near the left surface and the right surface, and that the bands are flat in the vicinity of both surfaces, indicating the absence of any charge on the surface. It is also important to note that the overall system is neutral.

Inspection of Fig. 3 reveals that electrons are present in conduction-band states on the GaN side of the interface; this is evident both from the downward band bending in the electrostatic potential profile [Fig. 3(a)] and in the layer-resolved DOS [Fig. 3(b)]. The charge distribution of these electrons is depicted in Fig. 3(c). The presence of such electrons corresponds to our notion, founded in experimental observations, that a 2DEG forms at the GaN/AlN (0001) interface. But what is the source of these electrons? No intentional donor-doping was intro-

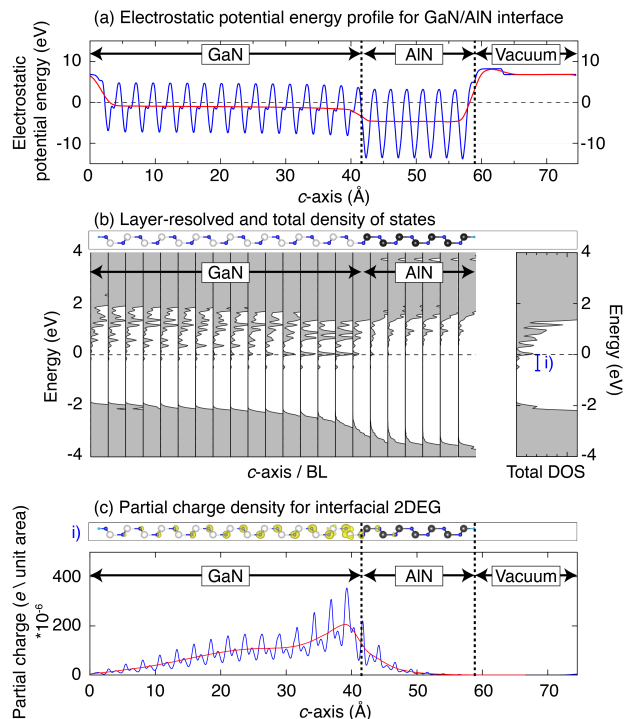


FIG. 3. (a) Electrostatic potential profile, (b) layer-resolved DOS, and (c) partial charge density for the electron concentration in the occupied conduction band of a GaN/AlN (0001) heterostructure consisting of 15 BL of GaN, 7 BL of AlN, and 18 Å of vacuum. The GaN(000 $\bar{1}$) surface is passivated with $\text{psH}_{0.732}$ and the AlN(0001) surface is passivated with $\text{psH}_{1.328}$. In (b), the zero of energy corresponds to the Fermi level, and the occupied conduction-band states are indicated with the label i) in the total DOS plot. A schematic of the structure is shown above panels (b) and (c). In (c), the schematic also shows isosurfaces of the occupied conduction-band states. Planar- and macroscopic-averaged quantities are shown in blue and red solid lines, respectively.

duced, and no surface states that can act as donors are present.

Here we explain the source of the electrons based on the description of electron distribution in a solid with spontaneous polarization outlined at the end of Sec. III C. We observed that a field-free WZ GaN slab could be constructed based on neutral Ga and N atoms in which the electron cloud has been shifted slightly to the left, along [000 $\bar{1}$]. Specifically, a Ga atom on the (0001) surface would have a db containing 0.732 electrons. We can apply similar argumentation to construct a field-free slab of strained AlN; again, the electron cloud would be shifted slightly to the left, but by a somewhat greater amount, reflecting the stronger spontaneous polarization; an Al atom on the (0001) surface would have a db containing 0.672 electrons, and conversely a N db on the (000 $\bar{1}$) surface would contain 1.328 electrons.

If we now bring the GaN and AlN slabs together and form a GaN/AlN interface, we find that the bond at the

interface contains $0.732+1.328=2.060$ electrons, i.e., an excess of 0.060 electrons over and above the two electrons needed to form the Ga-N bond. Since both the GaN and AlN layers that were brought together were field-free, these electrons will “stay local” (since they are present as part of a locally charge-neutral system), which means they need to go into the conduction band, forming the 2DEG! Of course some local band bending will occur in response to the presence of these electrons, but that does not invalidate the overall field-free nature of the system, as evidenced by the absence of electric fields in the regions of the slab away from the interface in Fig. 3(a). This view of the origin of electrons in the 2DEG is consistent with comments about charge imbalance in Ref. 16.

We explicitly verified the presence of 0.060 electrons per areal unit cell in the region of the interface by performing an integration over the electron charge density. This number, $0.060 e/uc = 0.107 C/m^2 = 6.7 \times 10^{13}$ electrons/cm², is also exactly the amount of negative charge needed to compensate the positive bound charge resulting from the polarization discontinuity at the GaN/AlN interface: $\sigma_b = P_{\text{eff}}^{\text{ZB}}(\text{GaN}) - P_{\text{eff}}^{\text{ZB}}(\text{s-AlN}) = [-0.018 - (-0.078)] e/uc = 0.060 e/uc$ (see Table I).

In the following, we will take the flat-band, neutral-interface configuration as the starting point for analyzing how the 2DEG density depends on the electronic structure of the AlN surface.

IV. PROPERTIES OF ALN BARRIERS

A. Critical layer thickness

While the presence of electrons at the interface agrees with our expectations, our simulations deviate from actual experimental observations in two ways: (1) at realistic interfaces, the 2DEG density is smaller than the magnitude of the polarization discontinuity; i.e., a net positive charge is present at the interface; and (2) in realistic AlN barriers an electric field is present (which is of course linked by Gauss’ law to the net charge at the interface). These features can be explained by noting that the system depicted in Fig. 3 can lower its energy if electrons are transferred from the interface to low-energy states on the surface. For the structure in Fig. 3, we precluded this transfer by imposing passivation with psH_{1.328} atoms on the AlN(0001) surface, but on a realistic surface the actual reconstructions would be driven by energy minimization. In Sec. V we will discuss how we can model such electron-accommodating reconstructions; for now, our model of passivation with pseudohydrogen atoms will serve to assess the impact of fixed charge on the surface.

Once the density of electrons in the 2DEG is less than the polarization discontinuity, the interface is charged and an electric field is present in the AlN layer; as a consequence, the impact of the energy-lowering driving force on the distribution of electrons in the system will depend on the thickness of the AlN barrier, as illustrated

in Fig. 4. We assume that the GaN(000 $\bar{1}$) surface has been passivated in such a way that we can assume that the thick GaN layer is field-free (except for 2DEG-related band bending near the interface).

We initially assume [Fig. 4(a)] that the AlN surface is passivated with psH_{1.268}. This provides a fixed charge $\sigma_f = 0.018 e/uc$ that is sufficient to passivate a GaN surface, but not an AlN surface. As a result, for thin AlN barriers, *all* of the electrons will be transferred from the interface to the surface. The net charge at the interface is then the bound polarization charge, given by the polarization discontinuity: $\sigma_b = 0.060 e/uc = 0.107 C/m^2$. If we assume the GaN layer is field-free, Gauss’ law tells us the electric field in the AlN barrier will be $\mathcal{E} = 142 \text{ mV/\AA}$ (we are assuming here that the dielectric constant of AlN is given by the experimental value of 8.5, Ref. 43). A rough estimate then indicates that as long as the AlN thickness is less than 30 Å, the AlN VBM at the surface will lie below the CBM of GaN, indicating that it will be energetically favorable to accommodate the electrons in bound states at the surface (i.e., states with energies below the VBM) rather than in conduction-band states at the interface (we assume, for now, that no surface state with energies in the band gap are present). Passivation with psH_{1.328} provided a positive fixed charge $\sigma_f = 0.078 e/uc$; transfer of electrons from the interface to the surface will reduce the amount of fixed positive charge on the surface; indeed, psH_{1.268} provides a fixed charge $\sigma_f = 0.018 e/uc$.

Nitride semiconductors typically exhibit unintentional *n*-type conductivity [44]; we can thus assume that the Fermi level in the system is close to the energy of the GaN CBM. In this estimate of the “critical layer thickness” t_{crit} we assume that the voltage drop over the AlN layer equals the energy difference between the GaN conduction band and the AlN valence band, i.e., $E_g(\text{AlN}) - \Delta E_c$, where ΔE_c is the GaN/AlN conduction-band offset [see Fig. 4(a)]. Experimentally it has indeed been found that observation of the 2DEG at the interface requires a minimum critical layer thickness of the AlN layer. This thickness is typically on the order of a few nm, showing a good agreement with our qualitative estimate. Comments about estimating the critical layer thickness in the case of AlGa_xN barriers, which are more common in experiment, are included in the Supplemental Material [34]. We will return to this issue below.

B. States on the surface and energy considerations

In our calculations, the bound states at the surface are simulated by adjusting the valence of the psH atom on the AlN surface. Our discussion in Sec. III D assumed a perfectly neutral and insulating AlN surface, with no states in the band gap, achieved by passivating the surface with psH_{1.328} atoms. As discussed in Sec. III C, this psH valence provides a fixed charge $\sigma_f = 0.078 e/uc$, exactly right to compensate the polarization of strained WZ

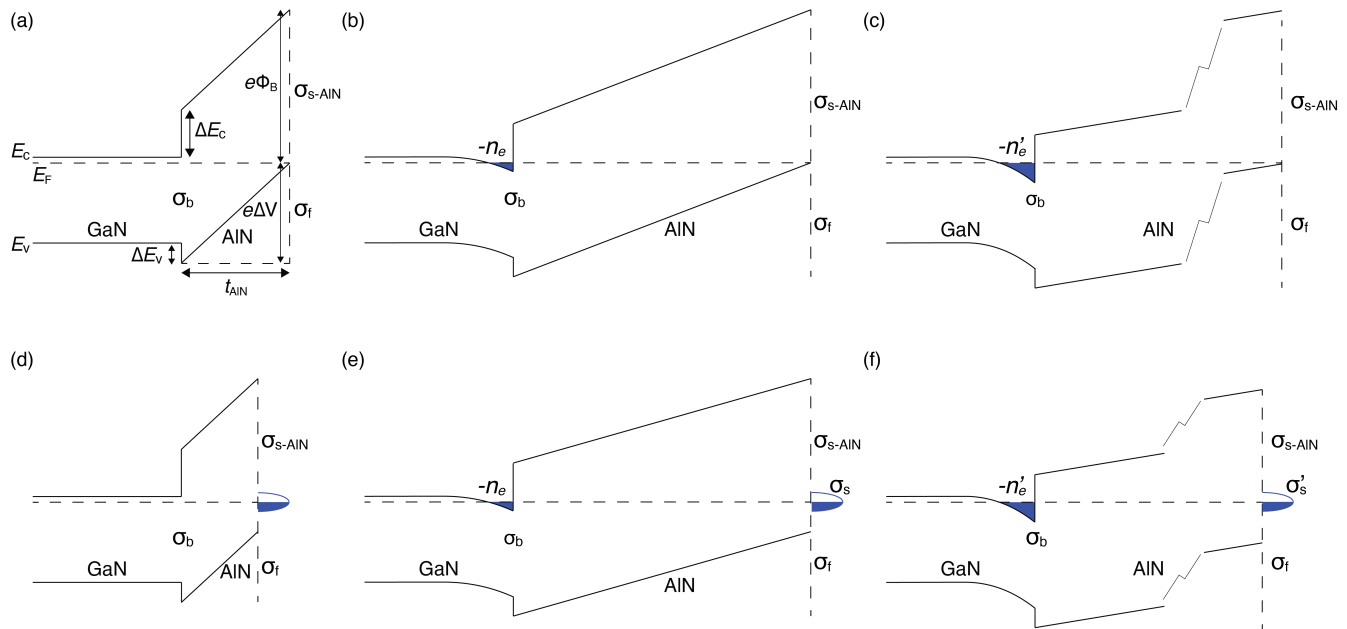


FIG. 4. Schematic band diagrams illustrating the formation of a 2DEG at a GaN/AlN heterostructure for various thicknesses of the AlN layer. Symbols were defined in Fig. 1. Panels (a)-(c) are for the situation where no surface states are present in the AlN band gap, (d)-(f) in the presence of surface states. The slope of the bands in the AlN layer is determined by the net charge at the interface (polarization charge minus electron density in the 2DEG). σ_b is the polarization charge at the interface, $\sigma_{s\text{-AlN}}$ is the polarization charge on the (0001) surface of AlN, and σ_f is the fixed charge (imposed by the formation of bound states) on the surface. n_e is the density of electrons in the 2DEG (a positive number, so that the corresponding charge is $-n_e$). (a) AlN thickness at or below the critical layer thickness. $\sigma_f = -\sigma_{\text{GaN}}$ is assumed, and no 2DEG forms. (b) AlN thickness above the critical layer thickness, and fixed charge on the surface larger than $-\sigma_{\text{GaN}}$. (c) Large AlN thickness, where the electron density in the 2DEG (n'_e) approaches the polarization charge at the interface, and the fixed charge approaches $\sigma_{s\text{-AlN}}$. (d) AlN thickness at the critical layer thickness, $\sigma_f = -\sigma_{\text{GaN}}$, no 2DEG forms; Fermi level at charge equilibration point in the surface state distribution. (e) AlN thickness above the critical layer thickness, fixed charge on the surface larger than $-\sigma_{\text{GaN}}$, and charge in surface states equal to σ_s . (f) As in (e), for large AlN thickness, where the electron density in the 2DEG (n'_e) approaches the polarization charge at the interface.

AlN. If instead of a psH_{1.328} atom we use a psH_{1.268} atom, then no mobile electrons are present in the system. Indeed, in that case $\sigma_f = 0.018 e/uc$, and combined with the polarization charge on the (strained) AlN surface we obtain $\sigma = \sigma_{s\text{-AlN}} + \sigma_f = -0.078 + 0.018 = -0.060 e/uc$, which is exactly equal and opposite to the bound polarization charge σ_b at the interface.

This reasoning indicates that as σ_f increases from $0.018 e/uc$, to $0.078 e/uc$, the density of electrons at the interface will increase and the 2DEG will build up, even if the AlN thickness is below the critical layer thickness. In the Supplemental Material [34] we demonstrate this with explicit first-principles calculations for the heterostructure. In realistic heterostructures one does of course not have such absolute control over the fixed charge at the surface, so now we explore what happens in the experimentally more relevant situation where the thickness of AlN increases.

When the AlN thickness exceeds the critical layer thickness [Fig. 4(b)], the VBM of AlN would rise above the Fermi level. This would create positive charge (holes) on the surface, which would be added to the fixed charge. As a result the magnitude of the total charge on the surface would go down, and the field in the AlN layer

would decrease, allowing the AlN VB to remain close to the Fermi level. Simultaneously, electrons appear in the 2DEG at the interface. Another way to look at this is to observe that the maximum voltage drop over the AlN layer is set at the critical layer thickness. If the barrier thickness is smaller than the critical layer thickness, the electric field stays constant. As the barrier thickness increases beyond the critical layer thickness, however, the voltage drop stays (approximately) constant, which implies that the electric field in the AlN layer decreases. Therefore, by Gauss' law, the net charge at the interface has to decrease. This happens because the density of electrons in the 2DEG increases and increasingly compensates the positive bound polarization charge. In the limit of an infinitely thick AlN barrier, the 2DEG density exactly equals the polarization charge, the net charge at the interface is zero, and there is no electric field in the AlN layer, as shown in Fig. 3.

Creating holes in the AlN VB is of course highly unlikely; instead, the requisite positive charge on the surface would be provided by adjusting the fixed charge (still under the assumption that there are no surface states in the band gap). In our model system this would happen by increasing the valency of the psH atoms above 1.268, which

corresponds to increasing the magnitude of the positive fixed charge. The magnitude of the 2DEG density grows and the electric field in the AlN layer decreases, as depicted in Fig. 4(c).

For a given fixed charge σ_f , the total charge on the surface is $\sigma = \sigma_{s\text{-AlN}} + \sigma_f$; charge neutrality dictates that $-n_e + \sigma_b + \sigma_{s\text{-AlN}} + \sigma_f = 0$, leading to $n_e = \sigma_{\text{GaN}} + \sigma_f$. For $\sigma_f = -\sigma_{\text{GaN}}$, this gives $n_e = 0$, as we found above. If we would increase the psH valency to 1.328 (corresponding to a fixed charge $\sigma_f = -\sigma_{s\text{-AlN}} = 0.078 e/\text{uc}$), then we are perfectly compensating $\sigma_{s\text{-AlN}}$ and we would have a field-free situation (even for a finite thickness of AlN, see Fig. 3) where the electrons in the interfacial 2DEG perfectly compensate the bound polarization charge: $n_e = \sigma_{\text{GaN}} + \sigma_f = \sigma_{\text{GaN}} - \sigma_{s\text{-AlN}} = \sigma_b$. In reality, the fixed charge on the surface will assume an intermediate value, allowing for a field in the AlN layer. Realistic surfaces do not feature psH atoms, of course; in Sec. V we will discuss the physical nature of such bound states and relate it to surface reconstructions.

The notion that for AlN thickness below 30 Å no electrons are present at the interface and they are all tied up in bonds (i.e., negative fixed charge) at the surface is based on several approximations. We ignore the cost of the surface reconstructions that would provide the bound charge. However, from numerous first-principles simulations we have learnt that the cost of such reconstructions is on the order of 0.1 eV/uc. In contrast, the energy that can be gained by transferring electrons from the interface to the surface, for small thicknesses of AlN, is on the order of the GaN band gap, i.e., an order of magnitude larger. One may also wonder about the energy involved in setting up the field across the AlN barrier—which is essentially the cost of charging a parallel-plate capacitor. Again, an estimate indicates that this cost would only be on the order of 0.1 eV/uc (we will return to this issue in Sec. VI). We are therefore justified in assuming that the energy lowering obtained by transferring electrons to the surface will be a dominant driving force.

C. Bound states versus surface states

In the scenario outlined above, the electrons that appear as the interfacial 2DEG when the psH valency is raised from 1.268 to 1.328 are related to fixed charge associated with specific surface structures. One may ask whether the charge could be associated with donor-like surface states, i.e., states with energies within the AlN band gap, rather than bound states. However, surface states cannot be present in the high densities corresponding to the density of the 2DEG, i.e., $6.7 \times 10^{13} \text{ cm}^{-2}$. **First, the aforementioned spectroscopic measurements of surface-state densities never observe such high values. Second, if such a high density of states were indeed present, some conductivity associated with these states should be observed, and this has never been reported.** Therefore the vast majority of the charge must be asso-

ciated with bound states, i.e., fixed charge.

However, virtually all realistic surfaces have some density of surface states, and we cannot exclude that some fraction of electrons may be associated with surface states. Unlike bound states, these surface states can easily accommodate variations in the density of electrons. For that reason, such surface states (even at relatively low densities) play an important role in pinning of the Fermi level.

We already noted that once the AlN barrier exceeds the critical thickness, the electron distribution has to change. Indeed, the electric field based on assuming full transfer of electrons from the interface to the surface would now raise the AlN VBM above the Fermi level; this can be regarded as a driving force for keeping electrons at the interface, in GaN conduction-band states. An equilibrium is established by requiring a constant Fermi level that governs the density of electrons at the interface and balances that density with charge at the surface. In our treatment above, we assumed the AlN band gap was free of surface states and the Fermi level at the surface would coincide with the AlN VBM. Any realistic surface will have some density of surface states, and it is the distribution of electrons within those surface states, rather than the AlN VBM, that will determine the Fermi-level position at the surface, resulting in a surface barrier height Φ_B as shown in Figs. 4(d)-(f).

The panels in the lower half of Fig. 4 correspond to the panels in the upper half, but allow for the presence of a distribution of surface states on the AlN surface, which can accommodate a certain density of electrons yielding a charge σ_s . Figure 4(d) corresponds to the critical layer thickness. We still assume that $\sigma_f = +0.018 e/\text{uc}$. The critical layer thickness is then determined by the point where the Fermi level is aligned with the charge equilibration point of the surface state distribution; i.e., the point where the positive charge associated with the surface states is exactly balanced by the electrons residing in the surface states, so that $\sigma_s = 0$. This pinning of the Fermi level implies that the critical thickness will be smaller than in the absence of surface states. Figures 4(e) and (f) depict situations where the AlN thickness increases. The total charge on the surface is now $\sigma = \sigma_{s\text{-AlN}} + \sigma_f + \sigma_s$, and the relation between the 2DEG density and the charges on the surface is now given by $-n_e + \sigma_b + \sigma_{s\text{-AlN}} + \sigma_f + \sigma_s = 0$, leading to $n_e = \sigma_{\text{GaN}} + \sigma_f + \sigma_s$.

Experimentally, it is observed that the 2DEG density changes as the thickness of the AlN barrier changes. As discussed in Sec. IV B, the 2DEG density is related to the amount of charge at the AlN surface. **The major shortcoming in most discussions of the relationship between AlGaIn thickness and 2DEG density [12, 16, 45] is that they implicitly assume that the electronic structure of the surface remains unchanged when the AlGaIn thickness is varied.** I.e., a fixed distribution of surface states is assumed, which is supposed to accommodate the large variations in 2DEG density. We now know that this can-

not be correct. In order to explain the variation in 2DEG density from zero up to $6.7 \times 10^{13} \text{ cm}^{-2}$ a correspondingly high density of surface states should be present, and such a density is not observed.

Still, the charge on the AlN surface needs to change as the density of electrons in the 2DEG changes. We therefore still have to explain how this amount of *fixed* charge in *bound* states can be adjusted as the AlN thickness varies. The key here is to realize that varying the thickness of AlN can only be accomplished by changing the atomic structure, either by growth or by etching. Either way, in the course of changing this atomic structure, the system will aim to lower its energy. We have already established that the distribution of charge between interface and surface is the dominant driving force in the energetics, and that the cost of surface reconstructions is minor on the relevant energy scale. Therefore, for a given AlN thickness, the system will choose a surface reconstruction that can accommodate the right amount of fixed charge to balance the charge in the 2DEG and equilibrate the Fermi level. In Sec. V we will see examples of how this can be accomplished on realistic reconstructed surfaces.

V. REALISTIC SURFACES

In presenting our model, we assumed psH atoms with fractional valence are present to provide fixed charge. This allowed us to focus on the physics and perform simulations for 1×1 unit cells. We now discuss whether and how such charges can be accommodated at realistic surfaces, and we distinguish between conditions present during growth versus conditions after the sample has been removed from the growth chamber.

A. Stabilization of fixed charge and consequences for surface reconstructions

In Sec. III C we defined the fixed charge on a GaN slab as the amount of charge fixed in a dangling bond on the surface that would exactly compensate the polarization charge, and hence neutralize the surface. For the Ga-polar WZ GaN(0001) surface, the number of electrons needed in the bond pointing along [0001] was 0.732, and fixed charge was defined as $0.75 - 0.732 = +0.018 e/\text{uc}$. We also pointed out that even on a ZB (111) polar surface, a charge amounting to $3/4$ electron needs to be stabilized in the Ga bond pointing along [111]. It is well known that if the surfaces of a ZB (111) slab are not passivated, charge transfer out of the Ga dbs will occur, and it has been recognized that passivation with psH_{1.25} atoms solves this problem [37]. What the pseudohydrogen accomplishes is to fix the 0.75 electrons in the Ga bond; and it does this in a way that is conventionally accepted as representative of the covalent bonds that would

be formed at a heterostructure, or in a surface reconstruction.

Therefore the majority of the $-0.732 e/\text{uc}$ that is needed to neutralize the WZ GaN surface is already accounted for, through formation of covalent bonds that are taken for granted on the ZB (111) surface—hence our choice of the ZB phase as the reference for defining the polarization and the fixed charge. The challenge we face on the WZ surface is to supply additional fixed charge in the amount of $+0.018 e/\text{uc}$. The relatively small magnitude of this charge is actually a major problem for first-principles modeling, since electrons come in integer units.

Still, it is instructive to describe a conceptual example. The ideal neutral and insulating (0001) surface of GaN should have a fixed charge of 0.018 electrons per unit cell. It is well known that the energy of the system can be lowered by creating a surface reconstruction, for instance by adding one N adatom per (2×2) unit cell. The 5 electrons provided by the N adatom, combined with the 4×0.732 electrons provided by each of the 4 Ga surface atoms, allow forming three two-electron bonds that tie the N adatom to three of the Ga surface atoms, with 1.928 electrons left to fill a remaining db on the N adatom. This is clearly not a stable situation; 0.072 electrons are missing to make this a fully filled, energetically stable db.

But it can be stabilized by considering a larger-scale reconstruction. Hypothetically, let's consider a (14×12) reconstruction consisting of 42 (2×2) subunits and a total of 168 unit cells. Let's imagine 41 of the (2×2) subunits contain a N adatom and one does not. The one unreconstructed (2×2) units has $4 \times 0.732 = 2.928$ mobile electrons available. These electrons can go into the partially filled N dbs in the 41 reconstructed (2×2) units, which need $41 \times 0.072 = 2.952$ electrons. To within 0.024 electron, or less than $0.0002 e/\text{uc}$, this (14×12) reconstruction will therefore “fix” the charge that is required to make the WZ GaN surface neutral and insulating.

Note that in this example, we consider the charge that is “locked up” in a N db to be “fixed charge”, since the energy levels of N db states essentially overlap with the valence band [46]; i.e., the charge cannot be released in the type of surface science experiments that have been applied to probe charge on the surfaces of GaN-based devices. We also note, however, that our example of a reconstruction on a bare GaN surface is largely academic, since such surfaces would only be stable in ultrahigh vacuum.

The argumentation presented here should apply to the surface of *any* (thin or thick) layer of a WZ semiconductor in a (semi)polar orientation. Indeed, the need to provide a fixed charge may provide a driving force for reconstructions that can be experimentally observed, for instance using scanning tunneling microscopy. As documented in Ref. 47, a number of unusual surface reconstructions have been reported. In addition to the expected 2×2 , additional reconstructions that have been observed on the Ga-polar (0001) surface include 5×5 , 6×4 , and “ 1×1 ”

(pseudo- 1×1). On the N-polar (000 $\bar{1}$) surface, 1×1 , 3×3 , 6×6 , and $c(6\times 2)$ reconstructions have been observed. To the best of our knowledge, no consistent explanations have been advanced for the presence of these reconstructions. We suggest that the drive to accommodate fixed charge to compensate polarization charge may provide at least a partial explanation for these larger-scale reconstructions.

B. Post-growth conditions

The discussion so far has focused on “bare” nitride surfaces, with surface reconstructions that involve only Ga (Al) and N atoms. Once the sample is exposed to air, during or after cooldown, the surface invariably oxidizes [48, 49], giving rise to the presence of a few monolayers of oxide on the surface [50, 51]. In some cases the oxide is intentionally grown [17, 52]. Since the oxide does not exhibit polarization, the AlN/oxide interface behaves similarly to an AlN/vacuum interface from the point of view of polarization. High-quality bonding occurs at the AlN/Al₂O₃ interface; Al atoms are equally comfortable in tetrahedral and octahedral coordinates and can switch from bonding to nitrogen to bonding to oxygen [50, 51]. Still, given the mismatch in structure and lattice parameters, it is easy to accept that deviations from the bulk bonding arrangement can give rise to fixed charges at the densities discussed above, with deviations on the order of $0.1 e/uc$ compared to interfaces in the absence of polarization. It is particularly suggestive, as noted in Ref. 17, that oxygen acts as a donor in AlN, which renders it plausible that a fraction of the interfacial Al-O bonds play a role in providing the fixed charge that is required at the surface of the GaN/AlN (0001) heterostructure.

Indeed, we could consider the presence of oxygen donors in the context of the example we discussed in Sec. V A, for a hypothetical (14×12) surface reconstruction, in which (2×2) subunits exhibit a N-adatom reconstruction. Each of these units needs 0.072 electrons to completely fill the N-adatom db that fixes the charge. In our example, we assumed one subunit remained unreconstructed and provided 2.928 mobile electrons. Instead of this unreconstructed (2×2) cell, we could assume the presence of three oxygen atoms, which act as donors and provide 3 mobile electrons, quite close to the amount of charge (2.952 electrons) needed to fill all of the N dbs. To within 0.048 electrons, or less than $0.0003 e/uc$, this (14×12) reconstruction with three oxygen donors will therefore “fix” the charge that is required to make the WZ GaN surface neutral and insulating.

We have argued that most of the charge needed to compensate the polarization charge will be accommodated in bound states (fixed charge). However, imperfections in the bonding as well as the presence of the surface on the oxide will also give rise to a distribution of surface states within the band gap [50]. While their density is too low to explain the fixed charge needed to compensate

the polarization charge, such surface states will definitely play a role in pinning of the Fermi level, as discussed in Sec. IV C.

We note that experiments aimed at establishing surface-state densities and/or surface barrier heights on nitride surfaces are performed in air and therefore unavoidably study *oxidized* surfaces. Surface state distributions on such surfaces can vary greatly depending on growth or preparation conditions. This likely explains the differences that have been observed in different experiments, or the variations in measured surface barrier height [12].

VI. DISCUSSION AND COMPARISON WITH EXPERIMENT

A. 2DEG density and barrier height

As already discussed, our model for the origin and distribution of electrons is consistent with the existence of a critical layer thickness for the AlN barrier, below which a 2DEG is not observed. It is also consistent with the observation [12] that the density of electrons in the 2DEG increases when the barrier thickness is increased beyond the critical layer thickness, and eventually saturates.

Experimentally, a correlation between barrier thickness (or 2DEG density) and surface barrier height has also been examined. Some studies [3, 5, 6, 8] have found the surface barrier height $e\Phi_B$ (the energy difference between the Fermi level and the CBM of AlGa_xN_{1-x}) to be constant as the AlGa_xN barrier thickness increases. Other investigations, however, have reported that $e\Phi_B$ increases significantly with AlGa_xN thickness, *e.g.*, changing from 1.1 eV to 1.8 eV with increasing Al_{0.35}Ga_{0.65}N thickness [10], or even from 1.0 eV to 2.0 eV for Al_{0.19}Ga_{0.81}N or from 1.5 eV to 2.5 eV for Al_{0.29}Ga_{0.71}N [12]. As noted above, surface-state distributions can be very sensitive to preparation conditions, so the wide variety of Φ_B values that has been observed should come as no surprise.

For studies that have been systematically performed as a function of thickness, we suggest that the observed increase in Φ_B may be attributable to the energy related to the voltage drop ΔV over the barrier layer (see Fig. 1). This energy is essentially that of a parallel-plate capacitor, *i.e.*, $\epsilon_0\epsilon_r(\Delta V)^2/t$ where t is the thickness of the barrier layer. For small values of t , this energy is sizable enough for the system to have an incentive to keep ΔV small. The surface barrier height $e\Phi_B$ is roughly equal to $e\Delta V + \Delta E_c$ (neglecting the energy difference between the Fermi level and the CBM at the interface), and hence Φ_B will tend to be small for small thicknesses t , as is indeed observed experimentally. When t is large, the energy $\epsilon_0\epsilon_r(\Delta V)^2/t$ will be small and not play an important role in the overall energy minimization of the system. Energy minimization will then be dominated by other factors, which apparently favor larger Φ_B values. A larger Φ_B value is energetically favorable because (if we

assume a broad distribution of surface states in the band gap) it leads to less occupation of surface states. Placing electrons in surface states is energetically less favorable than accommodating them in bound states (with energy levels within the valence band). This may explain why (at least in some studies) higher Φ_B values have been found for larger barrier thicknesses.

We note that the increase in barrier height as a function of barrier thickness reported in Ref. 12 was observed in samples in which the thickness was varied by etching, and only when samples were unannealed. For *annealed* samples, the barrier height was observed to be independent of thickness. This suggests that annealing allows for an equilibration (energy minimization) process in which compensation of polarization fields is enabled by fixed charge, and depletion of surface states plays only a minor role.

B. Experimental verification

A verification of the model outlined here has actually already been experimentally implemented, in careful experiments by Ganguly *et al.* [17] on GaN/AlN/oxide heterostructures. While our discussion focused on barrier layers terminated at a surface, much of it can be carried over to structures where an oxide layer is present on top of the Al(Ga)N. While the presence of an oxide may change some of the electrostatics of the problem, from the point of view of polarization the oxide behaves the same as vacuum, since it does not have any polarization fields. In Sec. VB we already discussed the presence of thin spontaneously formed oxides on the surface. Ganguly *et al.* [17] reported the presence of positive fixed charge at the AlN/Al₂O₃ interface; they proposed that this fixed charge effectively compensates the polarization charge at the AlN surface. That was also the conclusion of a study of GaN/AlGaN metal-insulator-semiconductor heterostructures by Matys *et al.* [52]. Ganguly *et al.* [17] suggested that oxygen atoms at the AlN/Al₂O₃ interface could act as donor dopants; we commented on this issue in Sec. VB.

The numbers provided by Ganguly *et al.*'s analysis [17] are perfectly in line with our model. For instance, they state that if the fixed charge at the AlN/oxide interface (which they refer to as Q_{it} , and which corresponds to σ_f in our model) is equal in magnitude to the GaN polarization charge, then the 2DEG density is zero and the field outside the AlN layer is also zero. This is consistent with the discussion in Sec. IV C and Fig. 4(d), for $\sigma_f = -\sigma_{\text{GaN}} = +0.018 \text{ e/uc}$.

Ganguly *et al.*'s analysis [17] revealed that in their actual devices Q_{it} was equal to $6 \times 10^{13} \text{ e/cm}^2$. They pointed out that this was close to the value of the polarization discontinuity ($6.1 \times 10^{13} \text{ e/cm}^2$). We argue that this is just a coincidence; instead, this charge at the surface of the AlN layer is better explained in terms of the model we presented in Fig. 4(e): it will be set by the

barrier height and the AlN thickness at the point where the growth of the AlN layer stops.

While growing the AlN layer, the Fermi level will be pinned by surface states; let's assume these surface states lead to a surface barrier height $e\Phi_B$ of about 3.5 eV (Ganguly *et al.* did not provide an explicit value, because their experiments did not examine the bare AlN surface; however, this value is in line with other experiments on AlN and AlGaIn barriers). Based on the arguments in Sec. VIA this value of $e\Phi_B$ is roughly equal to $e\Delta V + \Delta E_c$, so we have an estimate for the voltage drop over the AlN layer. Combined with the thickness $t_{\text{AlN}} = 4 \text{ nm}$, we can then estimate the charge at the GaN/AlN interface by setting the voltage drop obtained from Gauss' law, $t_{\text{AlN}} \times (\sigma_b - n_e)/(\epsilon_0 \epsilon_r)$, equal to ΔV , leading to $n_e = \sigma_b - \Delta V(\epsilon_0 \epsilon_r)/t_{\text{AlN}} = 4.2 \times 10^{13} \text{ e/cm}^2$. The charge at the interface, $\sigma_b - n_e$ needs to be equal and opposite to the charge on the AlN surface, $\sigma_f + \sigma_{\text{AlN}}$. This leads to $\sigma_f = 7.9 \times 10^{13} + 4.2 \times 10^{13} - 6.1 \times 10^{13} = 6.0 \times 10^{13} \text{ e/cm}^2$, exactly the Q_{it} that Ganguly *et al.* derived.

We argue that while the fixed charge on the surface can vary as the AlN layer is being grown, once an AlN/Al₂O₃ heterostructure is grown this charge is now fixed, leading to the Q_{it} value observed at that interface. Different thicknesses of the oxide layer will lead to different values of the 2DEG density n , but this is now accommodated by fixed charge (and some contribution from surface states) on the *oxide* surface. This model could be tested by performing experiments similar to those of Ganguly but with a different AlN thickness t_{AlN} .

C. Two-dimensional hole gases

Our discussion was for GaN/AlN (0001) heterointerfaces, which give rise to a 2DEG. However, the entire discussion can also be applied to AlN/GaN (0001) interfaces, which should give rise to a two-dimensional hole gas (2DHG). Such hole gases were recently experimentally observed [53] in structures containing no acceptor dopants; previous observations of hole gases were in modulation Mg-doped heterostructures or structures in which both mobile electrons and holes were present.

As we discussed above, formation of a 2DEG requires positive fixed charge to be present on the surface. Interestingly, the formation of a 2DHG on AlN/GaN (0001) still requires positive fixed charge on the surface, in order to compensate the negative polarization charge ($\sigma_{\text{GaN}} = -0.018 \text{ e/cm}^2$) on the GaN surface. The difficulties in achieving a 2DHG may be in part related to the fact that nitride (0001) surfaces have a tendency to form donor-like surface states in the upper part of the band gap; these states can act as a sink for holes.

VII. CONCLUSIONS

We have provided an intuitive model for the origin of the 2DEG at GaN/AlGaN interfaces. We demonstrated, within the modern theory of polarization, that the electrons in the 2DEG are *intrinsic* to the interface, and are expected to be found there as part of a locally charge-neutral structure. The challenge in achieving high 2DEG densities is therefore not to provide a *source* of electrons (as has often been implicitly or explicitly assumed, e.g., in models based on donor-like surface states); rather, the challenge is to prevent the interfacial electrons from escaping to lower-energy states on the surface.

We have demonstrated that this is accomplished by providing *fixed charge* on the surface; i.e., charge that is tied up in bound states, and (largely) compensates the surface polarization charge. We provided specific examples of how such fixed charge can form on realistic surfaces. The role of surface states with energies in the band gap is largely to pin the Fermi level. Our approach provides a consistent, rigorous, but ultimately simple methodology to understand and analyze the problem of 2DEG formation. **We hope that it will stimulate new experiments aimed at quantifying the role and na-**

ture of fixed surface charge, and aid in guiding the optimization of carrier densities in the 2DEG, and potential also in 2DHGs.

VIII. ACKNOWLEDGMENTS

We gratefully acknowledge D. Ritter, S. Keller, C. E. Dreyer, D. Wickramaratne, and L. Weston for input and helpful discussions. C.V.d.W. was supported by the US Department of Energy, Office of Science, Basic Energy Sciences (BES), under Award No. DE-SC0010689. S.-H.Y. acknowledges funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska - Curie grant agreement No. 101034297. M. T. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy-EXC 2033-Projektnummer 390677874. Computational resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), supported by the National Science Foundation (ACI-1548562).

-
- [1] S. Chowdhury and U. K. Mishra, “Lateral and vertical transistors using the AlGa_N/Ga_N heterostructure,” *IEEE Trans. Electron Dev.* **60**, 3060 (2013).
 - [2] E. T. Yu, G. J. Sullivan, P. M. Asbeck, C. D. Wang, D. Qiao, and S. S. Lau, “Measurement of piezoelectrically induced charge in Ga_N/AlGa_N heterostructure field-effect transistors,” *Appl. Phys. Lett.* **71**, 2794 (1997).
 - [3] O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, W. J. Schaff, L. F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck, “Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGa_N/Ga_N heterostructures,” *J. Appl. Phys.* **85**, 3222 (1999).
 - [4] I. P. Smorchkova, C. R. Elsass, J. P. Ibbetson, R. Vetry, B. Heying, P. Fini, E. Haus, S. P. DenBaars, J. S. Speck, and U. K. Mishra, “Polarization-induced charge and electron mobility in AlGa_N/Ga_N heterostructures grown by plasma-assisted molecular-beam epitaxy,” *J. Appl. Phys.* **86**, 4520 (1999).
 - [5] J. P. Ibbetson, P. T. Fini, K. D. Ness, S. P. DenBaars, J. S. Speck, and U. K. Mishra, “Polarization effects, surface states, and the source of electrons in AlGa_N/Ga_N heterostructure field effect transistors,” *Appl. Phys. Lett.* **77**, 250 (2000).
 - [6] I. P. Smorchkova, L. Chen, T. Mates, L. Shen, S. Heikman, B. Moran, S. Keller, S. P. DenBaars, J. S. Speck, and U. K. Mishra, “AlN/GaN and (Al,Ga)N/AlN/GaN two-dimensional electron gas structures grown by plasma-assisted molecular-beam epitaxy,” *J. Appl. Phys.* **90**, 5196 (2001).
 - [7] S. Keller, S. Heikman, L. Shen, I. P. Smorchkova, S. P. DenBaars, and U. K. Mishra, “GaN–GaN junctions with ultrathin AlN interlayers: Expanding heterojunction design,” *Appl. Phys. Lett.* **80**, 4387 (2002).
 - [8] S. Heikman, S. Keller, Y. Wu, J. S. Speck, S. P. DenBaars, and U. K. Mishra, “Polarization effects in AlGa_N/Ga_N and Ga_N/AlGa_N/Ga_N heterostructures,” *J. Appl. Phys.* **93**, 10114 (2003).
 - [9] M. Higashiwaki, N. Hirose, and T. Matsui, “Cat-cvd Si_N-passivated AlGa_N-Ga_N hfets with thin and high Al composition barrier layers,” *IEEE Electron Device Lett.* **26**, 139 (2005).
 - [10] G. Koley and M. G. Spencer, “On the origin of the two-dimensional electron gas at the AlGa_N/Ga_N heterostructure interface,” *Appl. Phys. Lett.* **86**, 042107 (2005).
 - [11] N. Onojima, M. Higashiwaki, J. Suda, T. Kimoto, T. Mimura, and T. Matsui, “Reduction in potential barrier height of AlGa_N/Ga_N heterostructures by Si_N passivation,” *J. Appl. Phys.* **101**, 043703 (2007).
 - [12] M. Higashiwaki, S. Chowdhury, M. S. Miao, B. L. Swenson, C. G. Van de Walle, and U. K. Mishra, “Distribution of donor states on etched surface of AlGa_N/Ga_N heterostructures,” *J. Appl. Phys.* **108**, 063719 (2010).
 - [13] B. L. Swenson and U. K. Mishra, “Photoassisted high-frequency capacitance-voltage characterization of the Si₃N₄/Ga_N interface,” *J. Appl. Phys.* **106**, 064902 (2009).
 - [14] R. Yeluri, B. L. Swenson, and U. K. Mishra, “Interface states at the Si_N/AlGa_N interface on Ga_N heterojunctions for Ga and N-polar material,” *J. Appl. Phys.* **111**, 043718 (2012).
 - [15] T. Hossain, D. Wei, J. H. Edgar, N. Y. Garces, N. Nepal, J. K. Hite, M. A. Mastro, C. R. Eddy Jr., and

- H. M. Meyer III, "Effect of GaN surface treatment on $\text{Al}_2\text{O}_3/\text{n-GaN}$ mos capacitors," *J. Vac. Sci. Technol. B* **33**, 061201 (2015).
- [16] E. Ber, B. Osman, and D. Ritter, "Measurement of the variable surface charge concentration in gallium nitride and implications on device modeling and physics," *IEEE Trans. Electron Dev.* **66**, 2100 (2019).
- [17] S. Ganguly, J. Verma, G. Li, T. Zimmermann, H. Xing, and D. Jena, "Presence and origin of interface charges at atomic-layer deposited $\text{Al}_2\text{O}_3/\text{III-nitride}$ heterojunctions," *Appl. Phys. Lett.* **99**, 193504 (2011).
- [18] B. Bakeroot, S. You, T.-L. Wu, J. Hu, M. Van Hove, B. De Jaeger, K. Geens, S. Stoffels, and S. Decoutere, "On the origin of the two-dimensional electron gas at Al-GaN/GaN heterojunctions and its influence on recessed-gate metal-insulator-semiconductor high electron mobility transistors," *J. Appl. Phys.* **116**, 134506 (2014).
- [19] S.-H. Yoo, M. Todorova, D. Wickramaratne, L. Weston, C. G. Van de Walle, and J. Neugebauer, "Finite-size correction for slab supercell calculations of materials with spontaneous polarization," *npj Comput. Mater.* **7**, 58 (2021).
- [20] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Phys. Rev. B* **54**, 11169 (1996).
- [21] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Comput. Mater. Sci.* **6**, 15 (1996).
- [22] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.* **77**, 3865 (1996).
- [23] P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B* **50**, 17953 (1994).
- [24] J. Neugebauer and M. Scheffler, "Adsorbate-substrate and adsorbate-adsorbate interactions of Na and K adlayers on Al(111)," *Phys. Rev. B* **46**, 16067 (1992).
- [25] C. E. Dreyer, A. Janotti, C. G. Van de Walle, and D. Vanderbilt, "Correct implementation of polarization constants in wurtzite materials and impact on III-nitrides," *Phys. Rev. X* **6**, 021038 (2016).
- [26] H. Schulz and K. H. Thiemann, "Crystal structure refinement of AlN and GaN," *Solid State Commun.* **23**, 815 (1977).
- [27] I. Vurgaftman and J. R. Meyer, "Band parameters for nitrogen-containing semiconductors," *J. Appl. Phys.* **94**, 3675 (2003).
- [28] O. Madelung, *Semiconductors — Basic Data* (Springer Berlin, Heidelberg, 1996).
- [29] J. L. Lyons, A. Janotti, and C. G. Van de Walle, "Effects of carbon on the electrical and optical properties of InN, GaN, and AlN," *Phys. Rev. B* **89**, 035204 (2014).
- [30] D. Vanderbilt and R. D. King-Smith, "Electric polarization as a bulk quantity and its relation to surface charge," *Phys. Rev. B* **48**, 4442 (1993).
- [31] R. Resta, "Macroscopic polarization in crystalline dielectrics: the geometric phase approach," *Rev. Mod. Phys.* **66**, 899–915 (1994).
- [32] F. Bernardini, V. Fiorentini, and D. Vanderbilt, "Spontaneous polarization and piezoelectric constants of III-V nitrides," *Phys. Rev. B* **56**, R10024 (1997).
- [33] F. Bernardini, V. Fiorentini, and D. Vanderbilt, "Accurate calculation of polarization-related quantities in semiconductors," *Phys. Rev. B* **63**, 193201 (2001).
- [34] See the Supplemental Material at [<https://DOI:XXX>] for a table with calculated values of polarization constants, an estimation of the critical layer thickness for GaN/AlGaIn systems, and calculations elucidating the impact of psH valency on the formation of the 2DEG.
- [35] S. Ciraci, I. P. Batra, and W. A. Tiller, "Electronic structure of the (111) surface of semiconductors," *Phys. Rev. B* **12**, 5811–5823 (1975).
- [36] S. Ciraci and Inder P. Batra, "Surface metallization of silicon by potassium adsorption on Si(001)-(2×1)," *Phys. Rev. B* **37**, 2955–2967 (1988).
- [37] K. Shiraishi, "A new slab model approach for electronic structure calculation of polar semiconductor surface," *J. Phys. Soc. Jpn* **59**, 3455 (1990).
- [38] B. Meyer and D. Marx, "Density-functional study of the structure and stability of ZnO surfaces," *Phys. Rev. B* **67**, 035403 (2003).
- [39] M. Himmerlich, L. Lymperakis, R. Gutt, P. Lorenz, J. Neugebauer, and S. Krischok, "GaN(0001) surface states: Experimental and theoretical fingerprints to identify surface reconstructions," *Phys. Rev. B* **88**, 125304 (2013).
- [40] P. Strak, P. Kempisty, K. Sakowski, A. Kaminska, D. Jankowski, K. P. Korona, K. Sobczak, J. Borysiuk, M. Beeler, E. Grzanka, E. Monroy, and S. Krukowski, "Ab initio and experimental studies of polarization and polarization related fields in nitrides and nitride structures," *AIP Adv.* **7**, 015027 (2017).
- [41] M. D. Pashley, "Electron counting model and its application to island structures on molecular-beam epitaxy grown GaAs(001) and ZnSe(001)," *Phys. Rev. B* **40**, 10481–10487 (1989).
- [42] M. Esposito, S. Krishnamoorthy, D. N. Nath, S. Bajaj, T.-H. Hung, and S. Rajan, "Electrical properties of atomic layer deposited aluminum oxide on gallium nitride," *Appl. Phys. Lett.* **99**, 133503 (2011).
- [43] Michael E. Levinstein, Sergey L. Rumyantsev, and Michael S. Shur, *Properties of Advanced Semiconductor Materials: GaN, AlN, InN, BN, SiC, SiGe* (John Wiley and Sons, Inc., New York, 2001).
- [44] Pietro Pampili and Peter J. Parbrook, "Doping of III-nitride materials," *Mater. Sci. Semicond. Process.* **62**, 180–191 (2017).
- [45] L. Gordon, M. S. Miao, S. Chowdhury, M. Higashiwaki, U. K. Mishra, and C. G. Van de Walle, "Distributed surface donor states and the two-dimensional electron gas at AlGaIn/GaN heterojunctions," *J. Phys. D* **43**, 505501 (2010).
- [46] Chris G. Van de Walle and David Segev, "Microscopic origins of surface states on nitride surfaces," *J. Appl. Phys.* **101**, 081704 (2007).
- [47] R. M. Feenstra, J. E. Northrup, and J. Neugebauer, "Review of structure of bare and adsorbate-covered GaN(0001) surfaces," *MRS Internet J. Nitride Semicond. Res.* **7**, 3 (2002).
- [48] M. Kočan, A. Rizzi, H. Lüth, S. Keller, and U. K. Mishra, "Surface potential at as-grown GaN(0001) mbe layers," *Phys. Status Solidi B* **234**, 773 (2002).
- [49] Y. Dong, R. M. Feenstra, and J. E. Northrup, "Electronic states of oxidized GaN(0001) surfaces," *Appl. Phys. Lett.* **89**, 171920 (2006).
- [50] M. S. Miao, J. R. Weber, and C. G. Van de Walle, "Oxidation and the origin of the two-dimensional electron gas in AlGaIn/GaN heterostructures," *J. Appl. Phys.* **107**,

- 123713 (2010).
- [51] J. H. Dycus, K. J. Mirrielees, E. D. Grimley, R. Kirste, S. Mita, Z. Sitar, R. Collazo, D. L. Irving, and J. M. LeBeau, “Structure of ultrathin native oxides on III-nitride surfaces,” *ACS Appl. Mater. Interfaces* **10**, 10607 (2018).
- [52] M. Matys, B. Adamowicz, A. Domanowska, A. Michalewicz, R. Stoklas, M. Akazawa, Z. Yatabe, and T. Hashizume, “On the origin of interface states at oxide/III-nitride heterojunction interfaces,” *J. Appl. Phys.* **120**, 225305 (2016).
- [53] R. Chaudhuri, S. J. Bader, Z. Chen, D. A. Muller, H. G. Xing, and D. Jena, “A polarization-induced 2d hole gas in undoped gallium nitride quantum wells,” *Science* **365**, 1454 (2019).