Nobel Lecture: Quasielectric fields and band offsets: teaching electrons new tricks

Herbert Kroemer

*Electrical and Computer Engineering Department, University of California, Santa Barbara, California 93106-9560

(Published 22 October 2001)

I. INTRODUCTION

Heterostructures, as I use the word here, may be defined as heterogeneous semiconductor structures built from two or more different semiconductors, in such a way that the transition region or interface between the different materials plays an essential role in any device action. Often, it may be said that the interface is the device.

Herbert Kroemer
Nobel Prize in Physics (2000)
“for developing semiconductor heterostructures used in high-speed- and optoelectronics”
Si: the **diamond** structure

Si crystal structure

Conventional unit cell

Primitive unit cell

http://www.webelements.com/silicon/crystal_structure.html

http://onlineheavytheory.net/silicon.html

The underlying Bravais lattice of GaAs and many III-V semiconductors are also fcc, but the two types of positions are occupied by different elements. The crystal structure is called **zincblende**. (Wide band gap nitrides, e.g. GaN, AlN, are of a different structure, to be discussed later.)

![Si: diamond](http://onlineheavytheory.net/silicon.html)

![GaAs: zincblende](https://en.wikipedia.org/wiki/Cubic_crystal_system#Zincblende_structure)
Crystallographic direction index

The related notation \([hkl]\) denotes the direction \(ha_1 + ka_2 + la_3\).

Indices defined in conventional unit cell.

Indices in angle brackets such as \(\langle100\rangle\) denote a family of directions that are equivalent due to symmetry operations, such as \([100]\), \([010]\), \([001]\).

Cubic systems (including diamond structure, e.g. Si, zincblende structure, e.g. GaAs)

Indices defined in conventional unit cell.
Indices in angle brackets such as \(\langle100\rangle\) denote a family of directions that are equivalent due to symmetry operations, such as \([100]\), \([010]\), \([001]\).
Crystallographic directions of Si

For Si, view the yellow and blue spheres as identical.

The distance between neighboring atoms in the $\langle 100 \rangle$ direction of Si is the conventional lattice parameter $a$. The unit cell of the 1D Bravais lattice is a line segment of length $a$, containing one atom. All $\langle 100 \rangle$-direction atomic lines are equivalent.

The distance between neighboring atoms in the $\langle 110 \rangle$ direction of Si is $\sqrt{2}a/2$. The unit cell of the 1D lattice is a line segment of length $\sqrt{2}a/2$, containing one atom. All $\langle 110 \rangle$-direction atomic lines are equivalent.

Distance between neighboring atoms in the $\langle 111 \rangle$ direction of Si are $\sqrt{3}a/4$ and $3\sqrt{3}a/4$. The unit cell of the 1D lattice is a line segment of length $\sqrt{3}a$, containing 2 atoms.

Each bond length is $\frac{\sqrt{3}a}{4}$

All $\langle 111 \rangle$-direction atomic lines are equivalent.
Crystallographic directions of GaAs

For GaAs, view the yellow and blue spheres as different.

The distance between neighboring atoms in the ⟨100⟩ direction of GaAs is the conventional lattice parameter $a$. The unit cell of the 1D lattice is a line segment of length $a$, containing one atom. There are 2 inequivalent types of ⟨100⟩-direction atomic lines: one of Ga and one of As.

The distance between neighboring atoms in the ⟨110⟩ direction of GaAs is $\sqrt{2}a/2$. The unit cell of the 1D lattice is a line segment of length $\sqrt{2}a/2$, containing one atom. There are 2 inequivalent types of ⟨110⟩-direction atomic lines: one of Ga and one of As.

Distance between neighboring atoms in the ⟨111⟩ direction of GaAs are $\sqrt{3}a/4$ and $3\sqrt{3}a/4$. The unit cell of the 1D lattice is a line segment of length $\sqrt{3}a$, containing one Ga and one As atom. All ⟨111⟩-direction atomic lines are equivalent.

FYI: The Ga-As bond is partially ionic, therefore each ⟨111⟩ direction is “polar”. Here, we consider 2 opposite directions, e.g. [111] and [1 1 1], as “one direction”. For the zincblende structure, there are 4 such directions. Due to symmetry, this crystal structure is not polar. The wurtzite structure, with the same fourfold coordinated bonding, is polar, due to lower symmetry. GaN is wurtzite structure. We did not cover this important structure and did not discuss much about symmetry.
Miller indices for crystallographic planes

\((hkl)\) denotes planes orthogonal to the reciprocal lattice vector three Miller indices \(h, k, \ell\)

\[
\mathbf{K} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3
\]

Equivalently, \((hkl)\) denotes a plane that intercepts the three points \(a_1/h, a_2/k,\) and \(a_3/\ell,\) or some multiple thereof.

This is an easier way to get the indices:
Reciprocals of the intercepts in \textbf{real space};
No need to find the reciprocal primitive vectors.

If one of the indices is zero, it means that the planes do not intersect that axis (the intercept is "at infinity").

\textbf{Different conventions in naming the real and reciprocal primitive vectors}

Real: \(a_1, a_2, a_3;\) reciprocal \(b_1, b_2, b_3\)
Real: \(a, b, c;\) reciprocal \(a^*, b^*, c^*\)
Cubic systems

Indices in curly brackets or braces such as \{100\} denote a family of planes that are equivalent due to symmetry operations.

Easier way to get the indices: Reciprocals of the intercepts
Crystallographic planes of Si

For Si, view the yellow and blue spheres as identical.

The distance between neighboring atoms in \{100\} planes of Si is $\sqrt{2}a/2$. The unit cell of the 2D lattice is a square of side $\sqrt{2}a/2$, containing one atom. All \{100\} atomic planes are equivalent.

The unit cell of the 2D lattice is a rectangle with sides $a$ and $\sqrt{2}a/2$, containing two atoms, as shown to the left. All \{110\} atomic planes are equivalent.

In \{110\} planes of Si, the distance between nearest neighboring atoms is the bond length $\sqrt{3}a/4$. 
In \{111\} planes of Si, the distance between neighboring atoms is $\sqrt{2}a/2$.

The 2D Bravais lattice is triangular, with lattice parameter $\sqrt{2}a/2$. The unit cell contains one atom. All \{111\} atomic planes are equivalent.
Crystallographic planes of GaAs

For GaAs, view the yellow and blue spheres as different.

The distance between neighboring atoms in \{100\} planes of GaAs is $\sqrt{2}a/2$. The unit cell of the 2D lattice is a square of side $\sqrt{2}a/2$, containing one atom. There are 2 inequivalent types of \{100\} atomic planes: one of Ga and one of As.

In \{110\} planes of GaAs, the distance between nearest neighboring atoms is the bond length $\sqrt{3}a/4$.

The unit cell of the 2D lattice is a rectangle with sides $a$ and $\sqrt{2}a/2$, containing two atoms.
All \{110\} atomic planes in a family that are parallel to each other (e.g. all (110) plane) are equivalent.

Two sets of \{110\} atomic planes that are perpendicular to each other are oppositely oriented.
In $\{111\}$ planes of GaAs, the distance between neighboring atoms is $\sqrt{2}a/2$.

Each bond length is $\frac{\sqrt{3}a}{4}$

The 2D Bravais lattice is triangular, with lattice parameter $\sqrt{2}a/2$. The unit cell contains one atom. There are 2 inequivalent types of $\{111\}$ atomic planes: one of Ga and one of As atoms.
To understand the SiO2/Si(001) interface, we first look at the Si(100) surface.

Si(001) truncated, surface reconstruction not considered

Dangling bonds

The front is the gray plane (110) in the 3D picture.
4.2 Types of solid: bond picture

This is about as far as we can go easily with a band model. But let us consider the semiconductors Si and Ge again. These have the ‘diamond structure’ which is like F.C.C. (see §1.3) except that there

From Ziman, Principles of the Theory of Solids
Silicon wafers

Ga is the cation (positive ions) and N is the anion (negative ions). Shouldn’t the positive polarization charge be at the Ga face? The “classical picture” of polar materials can be confusing.

Here, “polar” means having spontaneous polarization. In other context, “polar” may mean heteroatomic. For example, the Ga-As bond in GaAs is a polar covalence bond in that sense.
Review: Polarization, Displacement field, and Electric field

\( \mathbf{P} \) is the total (net) dipole moment per unit volume:

\[
\mathbf{P} = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \sum_i \mathbf{p}_i
\]

For a dielectric (no spontaneous polarization), external (applied) field due to external field leads to \( \mathbf{P} \).

Interior of dielectric is neutral, no electrical effect.

Obvious to see polarization charges at the two surfaces:

Define surface normal \( \mathbf{\hat{n}} \) pointing to interior, surface polarization charge density \( \sigma_P = -\mathbf{P} \cdot \mathbf{\hat{n}} \) on either surface.

Polarization charge results in a polarization electric field:

\[
\mathbf{E}_P = \sigma_P \mathbf{\hat{n}} / \varepsilon_0 = - (\mathbf{P} \cdot \mathbf{\hat{n}}) \mathbf{\hat{n}} / \varepsilon_0 = -\mathbf{P} / \varepsilon_0 ,
\]

which is against the external field \( \mathbf{E}_{\text{ext}} = \sigma \mathbf{\hat{n}} / \varepsilon_0 \), where \( \sigma \) is external charge surface density (on either plate, + or −).

Therefore the name “dielectric”.

\( \mathbf{P} \) is a linear response to the total field \( \mathbf{E} = \mathbf{E}_{\text{ext}} + \mathbf{E}_P = \sigma \mathbf{\hat{n}} / \varepsilon_0 - \mathbf{P} / \varepsilon_0 \), rather than just \( \mathbf{E}_{\text{ext}} \).

Define the proportional constant of the linear response as \( \chi \varepsilon_0 \), i.e., \( \mathbf{P} = \chi \varepsilon_0 \mathbf{E} \).
\[
E = E_{\text{ext}} + E_P = \sigma \hat{n}/\varepsilon_0 - P/\varepsilon_0
\]

\[
\Rightarrow \varepsilon_0 E = \sigma \hat{n} - P \Rightarrow \varepsilon_0 E + P = \sigma \hat{n}
\]

\[
\Rightarrow \varepsilon_0 (1 + \chi_e) E = \sigma \hat{n}
\]

Define \(1 + \chi_e = \varepsilon_r\) and \(\varepsilon = \varepsilon_r \varepsilon_0\), then we have Gauss’s law considering external charge only: \(\varepsilon E = \sigma \hat{n}\)

\[
\text{Define } D = \varepsilon E = \varepsilon_0 E + P
\]

\[
\Rightarrow D = \sigma \hat{n} \quad \text{(Gauss’s law in special form for parallel capacitor model)}
\]

This is the special form of Gauss’s law in the parallel-plate capacitor; the general form is:

\[
\nabla \cdot D = \rho \quad \text{(Gauss’s law in general form)}
\]

Similarly, \(-P \cdot \hat{n} = \sigma_P \Rightarrow \nabla \cdot P = -\rho_P\)

Essentially, total field is contributed by external and polarization charges:

\[
\varepsilon_0 \nabla \cdot E = \varepsilon_0 \nabla \cdot E_{\text{ext}} + \varepsilon_0 \nabla \cdot E_P = \rho + \rho_P
\]
Internal electric field due to spontaneous polarization

If a material possesses spontaneous polarization $\mathbf{P}_0$ and there is no external field, then

Parallel-plate capacitor model

\[ \mathbf{D} = \sigma \mathbf{n} = 0 \]

General

\[ \nabla \cdot \mathbf{D} = \rho = 0 \text{ everywhere} \]

\[ \mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}_0 = 0 \]

\[ \Rightarrow \quad \mathbf{E} = \mathbf{E}_p = -\mathbf{P}_0 / \varepsilon_0 \]

Thus we see that $\mathbf{P}_0$ points from Ga-face to N-face in wurtzite GaN:

Screening charge to be explained later
In this picture, \( P_0 \) is a bulk property. When the crystal is truncated, the surface density of polarization charge is the projection of \(-P_0\):

\[
\sigma_P = -P_0 \cdot \hat{n}
\]

\[
E = E_P = -P_0 / \varepsilon_0
\]

The internal field simply results from polarization charge:

Thus we see that \( P \) points from Ga-face to N-face in wurtzite GaN:

\[
Q_{\pi} = |\sigma_P|
\]
The wurtzite structure

Zincblende conventional unit cell

Compare:
zincblende [111] direction & wurtzite [0001] direction

These indices will be explained later.

Zincblende: 4 equivalent body diagonals.

Wurtzite: one unique c axis i.e. [0001] direction.
If $\bigcirc$ stands for Ga, then the bottom face is the Ga face (0001)!

Surfaces terminate in ways that minimize density of dangling bonds.
Recall this. Zincblende is similar, but with two types of atoms.
Zincblende is similar to diamond, but with two types of atoms.

Well will look at the $(\bar{1}10)$ plane with $[111]$ pointing up.
The difference is in stacking sequence.
The difference is in stacking sequence.

As a result, different symmetries:

In zincblende, the 4 bonds of each atom are along the 4 equivalent body diagonals.

In wurtzite, the [0001] is a unique direction. The bond along this axis may have a length different from the other three in the thermodynamically stable structure.

This deviation from the “ideal” structure gives rise to spontaneous polarization.

In the “ideal wurtzite structure”, the pyramids are regular tetrahedrons, thus the length of each bond is $\sqrt{3/8} \ a$, which is $3/4$ of the height of the pyramid. Therefore,

$$c = \left(2 + \frac{2}{3}\right) \sqrt{\frac{3}{8}} a = \frac{8}{3} \sqrt{\frac{3}{8}} a = \frac{8}{3} a$$
But, the idea that polarization results from non-overlapping positive and negative charge centroids does not make sense for bulk crystals with delocalized valence electrons.

- Textbook picture (Claussius-Mossotti)
- But does not correspond to reality!

\[
\mathbf{P} = \lim_{\Delta V \to 0} \frac{\sum p_i}{\Delta V}
\]

\[
\mathbf{P} = \frac{\mathbf{p}_{\text{unit cell}}}{V_{\text{unit cell}}}
\]

We stopped here on Thu 10/21/2021. Slides have updated and added after the class.
What is the origin of this difficulty?

Let’s first consider ultrathin films (few unit cells in thickness direction). No problem with the picture of non-overlapping positive and negative charge centroids.

Here, z and w polymorphs of a compound may both be polar.

The ultrathin zincblende film deviates from bulk structure.

In most practical devices, however, layers are many, many unit cell thick. The deviation at interfaces from bulk structure is insignificant, and $P_0$ aught to be regarded as a bulk property, whose projection at an interface $P_0 \cdot \hat{n} = -\sigma_P$.

But, bulk polarization $P_0$ (with the concomitant $E_P$) contradicts periodicity of crystal.

Furthermore, $E_P$ leads to a voltage $E_P t$ across thickness $t$ of bulk crystal. For a macroscopic this voltage is huge, but was never observed of a w-GaN(0001) wafer or its likes. This is due to the “screening charge” mentioned earlier. The unsustainable high voltage that would be present, were it not been for the screening charge, is referred to as “polar catastrophe.”
A simple classical picture illustrates the “polar catastrophe.”

The “catastrophe” is avoided by surface charge.

Notice that $\sigma' \neq \sigma$.

The screen charge can be due to structural deviation from bulk or simply contamination.
At the interface between two materials (1 & 2),

\[ \sigma_{P1} = -P_{01} \cdot \hat{n}_1 \]
\[ \sigma_{P2} = -P_{02} \cdot \hat{n}_2 = P_{02} \cdot \hat{n}_1 \]

Therefore, the total (net) interface charge

\[ \sigma_{P1} + \sigma_{P2} = -P_{01} \cdot \hat{n}_1 + P_{02} \cdot \hat{n}_1 = -(P_{01} - P_{02}) \cdot \hat{n}_1 \]

The interface charge is simply the polarization difference.

If \( P_{02} = 0 \), we simply have \( \sigma_{P1} + \sigma_{P2} = -P_{01} \cdot \hat{n}_1 \).

This simple observation allows for calculation of spontaneous polarization as bulk property from first principles.
In 1990s, “modern theory of polarization” calculated polarization as a bulk property from first principles.

Here, we illustrate it with an example: wurtzite BeO.


Model: w-BeO/z-BeO superlattice.
Use bulk-truncate structure for w-BeO layer; lattice match z-BeO to w-BeO while keeping cubit symmetry (all bonds equal length), such that $P_0 = 0$ in z-BeO.

Therefore, interface charge is all contributed by w-BeO. $\sigma_P = P_0$. Notice the direction of $P_0$.

For a review, see R. Resta, Rev. Mod. Phys. 66, 899 (1994).
In 1997, polarizations of technologically important III-N semiconductors were calculated from first principles.


First, crystal structures were determined.

<table>
<thead>
<tr>
<th></th>
<th>$\frac{c}{a}$</th>
<th>$u$</th>
<th>$\frac{P_0}{C/m^2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlN</td>
<td>1.6190</td>
<td>0.380</td>
<td>-0.081</td>
</tr>
<tr>
<td>GaN</td>
<td>1.6336</td>
<td>0.376</td>
<td>-0.029</td>
</tr>
<tr>
<td>InN</td>
<td>1.6270</td>
<td>0.377</td>
<td>-0.032</td>
</tr>
<tr>
<td>ideal wurtzite</td>
<td>1.6330</td>
<td>0.375</td>
<td></td>
</tr>
</tbody>
</table>

They all deviate from “ideal wurtzite”, AlN the farthest.

For GaN:

$$E_P = \frac{P_0}{\varepsilon_0} \approx \frac{0.03 \, C/m^2}{9 \times 10^{-12} \, \text{F/m}}$$

$$\approx 3 \times 10^9 \, \text{V/m} = 30 \, \text{MV/cm}$$

Compare this to the breakdown field $\sim 2 \, \text{MV/cm}$. Without the screening charge, GaN will be broke down by its own polarization field!
|σ_P| = P_0 = \frac{3 \times 10^{-2} \text{ C}}{10^4 \text{ cm}^2} \times \frac{e}{1.6 \times 10^{-19} \text{ C}} \approx 2 \times 10^{13} \text{ e/cm}^2

This is a very high charge density!

\textbf{P}_0 is due to overall charge distribution (including valance electrons and ion cores).

When AlN is forced to take on ideal wurtzite structure, it still exhibits \( \textbf{P}_0 = -0.033 \text{ C/m}^2 \), already comparable with GaN and InN.

\textbf{Negative sign} means \( \textbf{P}_0 \) points from the cation-face to the N-face!

This counters the intuition of positively charged cation-face and negative anion-face.

If \( \bigcirc \) stands for Ga, then the bottom face is the Ga face (0001)!
We have omitted lots of details. Additional reading FYI:


Something interesting from Vanderbilt’s 2006 presentation:

– Theory of polarization (King-Smith & Vanderbilt)

• A parable about referee reports...

Referee’s Report: Manuscript #LJ4237

Title: Soft self-consistent pseudopotentials in ...

Author: D. Vanderbilt

This paper is not suitable for publication in Phys. Rev. Letters. The basic idea the author presents is 30 years old, it is not clear that his reformulation is of any use, and technical refinements of pseudopotential theory are by now primarily of interest to specialists. A somewhat revised version should be suitable for publication as a brief report or regular paper in Phys. Rev. B. I will expand on these summary comments below.

While the formal connection between the present work and PK is obscure, the qualitative features clearly indicate that this is a big step backwards. To be acceptable for publication in Phys. Rev. B, a revised version...
Consider two slabs of dissimilar polar insulators. One way to arrange the screening charge:

\[
\frac{|\sigma_2| + |\sigma_1|}{2}
\]

\[
-|\sigma_1|
\]

\[
+|\sigma_1|
\]

\[
-|\sigma_2|
\]

\[
+|\sigma_2|
\]

\[
- \frac{|\sigma_2| + |\sigma_1|}{2}
\]

Charge

\[
|\sigma_1| - |\sigma_2|
\]

\[
\frac{|\sigma_2| - |\sigma_1|}{2}
\]

\[
\frac{|\sigma_2| - |\sigma_1|}{2}
\]

Electrostatic potential energy for electron
Polar HFET

Screening charge at substrate interface provided by defect states in substrate, usually very defective.

Polarization charge areal density $Q_\pi$ is our $|\sigma_P|$.

Polarization charge at AlGa-face is partly screened by ionized surface states.

2DEG results near the dip of the electrostatic energy ($E_{\text{vac}}$) profile.

Gate is deposited on the AlGa-face. 
Negative $V_G$ depletes 2DEG, as in conventional (AlGaAs/GaAs) HFET.

GaN may be n-doped
Compare: conventional AlGaAs delta-doped HFET
Now we clarify the indices we have used to refer to directions and planes in wurtzite (or any hexagonal) crystal structure.

A plane perpendicular to \( \hat{c} \) is a basal plane. The 2D Bravais lattice of the basal plane is triangular. The two ways of defining a set of \( \mathbf{a}_1 \) and \( \mathbf{a}_2 \) are equally good to define the same unit cell, thus ambiguity when we use \( u \mathbf{a}_1 + v \mathbf{a}_2 \) to define a direction.

We thus define 3 basis vectors, 120° from each other. \([ U \ V \ T ] = U \mathbf{a}_1 + V \mathbf{a}_2 + T \mathbf{a}_3\) defines a direction in the basal plane, with the constraint \( U + V + T = 0 \).

For example, \([1 \ 1 \ 0 ] = \mathbf{a}_1 - \mathbf{a}_2\) and \([1 \ 0 \ 1 ] = \mathbf{a}_1 - \mathbf{a}_3\) are two equivalent directions.

\([1 \ 1 \ 2 ] = \mathbf{a}_1 + \mathbf{a}_2 - 2\mathbf{a}_3\) and \([1 \ 2 \ 1 ] = \mathbf{a}_1 - 2\mathbf{a}_2 + \mathbf{a}_3\) are two equivalent directions.
Adding a 4th number \( W \), \[
[U \ V \ T \ W] = Ua_1 + Va_2 + Ta_3 + Wc
\] is a direction in 3D.

\[
[U \ V \ T \ 0] = Ua_1 + Va_2 + Ta_3
\] is a direction in the basal plane.

\([0001]\) is the \( \hat{c} \) direction.

\([000\bar{1}]\) is the \(-\hat{c}\) direction.

The \((hki0)\) plane is the plane whose normal \(// [hki0]\).

The relation \([hkl] \perp (hkl)\) is lost in the 3-number indices.
Miller-Bravais indexing (hkil) of planes

(01\bar{1}0) plane

(10\bar{1}0) plane

(2110) plane

(1120) plane

(0110) plane

(1210) plane

(1100) plane

Ga\textsubscript{c}uc

Ga\textsubscript{c}uc

(0001) is the Ga-face.

(000\bar{1}) is the N-face.
SiC and SiC-based devices

SiC has >170 polymorphs (polytypes; polytypism is a narrower concept than polymorphism).
Zincblende and wurtzite are just 2 polymorphs of an sp³-coordinated octet (II-VI, III-V, IV-IV) compound semiconductor.

3 bilayers per period in stacking direction

Zincblende
(\(\bar{1}10\)) plane with [111] pointing up

Wurtzite
(1\(\bar{2}10\)) plane with [0001] pointing up

2 bilayers per period in stacking direction
3C-SiC is zincblende

C means cubic.

3 bilayers per period in stacking direction

Figure adapted from Morkoç et al, Large-band-gap SiC, III-V nitride, and II-VI ZnSe-based semiconductor device technologies, *Journal of Applied Physics* **76**, 1363 (1994); https://doi.org/10.1063/1.358463

Similarly, 2H is wurtzite.
Besides 3C, 4H and 6H are the common polytypes.

The indexing of crystalline directions and planes discussed earlier for Wurtzite applies to all sp$_3$-coordinated hexagonal crystals, including hexagonal SiC (e.g. 4H, 6H), with unit cells larger in [0001] direction.
Different band structures and phonon dispersions give rise to different electronic properties of polytypes of SiC.

![Graph showing velocity-field relationships for Si, GaAs, InP, 3C, βSiC, and 6H SiC](image)

**FIG. 1.** Velocity-field relationships of Si, GaAs, InP, 3C and 6H SiC illustrating the superior high-field electron drift velocities and breakdown field of SiC. 3C SiC has a higher saturation velocity than 6H SiC due to reduced phonon scattering.

As for any MOS(MIS)FETs, the semiconductor-insulator interface is the key.