Surface Science 132 (1983) 543–576
North-Holland Publishing Company

HETEROSTRUCTURE DEVICES: A DEVICE PHYSICIST LOOKS AT INTERFACES

Herbert KROEMER

Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA

Received 18 October 1982; accepted for publication 30 December 1982

The band offsets occurring at abrupt hetero-interfaces in heterostructure devices serve as potential steps acting on the mobile carriers, in addition to the macroscopic electrostatic forces already present in homostructure devices. Incorporation of hetero-interfaces therefore offers a powerful device design parameter to control the distribution and flow of mobile carriers, greatly improving existing kinds of devices and making new kinds of devices possible. Unusual device requirements can often be met by band lineups occurring in suitable semiconductor combinations. Excellent theoretical rules exist for the semi-quantitative ($< \pm 0.2$ eV) prediction of band offsets, even unusual ones, but no quantitatively accurate ($< \pm 1\ kT$) purely theoretical predictive rules are currently available. Poorly-understood second-order nuisance effects, such as small interface charges and small technology-dependent offset variations, act as major limitations in device design. Suitable measurements on device-type structures can provide accurate values for interface physics parameters, but the most widely used measurements are of limited reliability, with pure $I-V$ measurement being of least use. Many of the problems at interfaces between two III/V semiconductors are hugely magnified at interfaces between a compound semiconductor and an elemental one. Large interface charges, and a strong technology dependence of band offsets are to be expected, but can be reduced by deliberate use of certain unconventional crystallographic orientations. An understanding of such polar/nonpolar interfaces is emerging; it is expected to lead to a better understanding and control of III/V-only device interfaces as well.

1. Introduction

This paper takes a look at interfaces in submicron structures, from the point of view of a device physicist who is interested in incorporating semiconductor hetero-interfaces into future *high-performance* semiconductor devices.

A significant fraction of such devices will be compound semiconductor rather than silicon devices. Before long, most compound semiconductor devices will involve heterostructures [1,2]. Homostructure devices made from a single compound semiconductor will probably be relegated to the low-performance/low-cost end of compound semiconductor technology, although silicon device technology will very likely continue to be dominated by homostructure devices. Furthermore, high performance in devices usually means

0039-6028/83/0000-0000/$03.00

minimizing the non-active part of the device volume, to the point that the device turns from a collection of semiconductor regions separated by interfaces, to a collection of interfaces with a minimum of semiconductor between them.

As this development progresses, it calls for a constant interchange of ideas between the device physicist and the more fundamentally-oriented "basic" surface/interface physicist. This interchange goes both ways: On the one hand, the device physicist (even if inclined to do so) can less and less rely on "cookbook empiricism"; instead he must closely follow the basic physicist in assimilating and utilizing the new fundamental knowledge that the latter has acquired. On the other hand, device physics constantly poses new problems to the basic physicist; and experiments on device-type structures (sometimes deliberately "misdesigned" as devices) offer themselves as powerful tools for basic research. One of the purposes of this paper is to contribute to this necessary interchange of ideas between the device physicist and the basic physicist.

Throughout the paper, the term *heterostructure device* is to be understood in the sense that the hetero-interface plays an essential role in the operation of the device, rather than just serving as a passive interface between what is basically a homostructure device and a chemically different substrate as in silicon-on-sapphire structures. In many cases, the interface *is* the actual device. The emphasis must therefore be on "good" interfaces made by "good" technology. Various kinds of interface defects, although never totally absent, can then at least be assumed to be present in only such small densities that their effect can be treated as a perturbation of a defect-free interface model, rather than as dominating the physics. These assumptions are by no means unrealistic "academic" ones, made to simplify the problem in neglect of practical realities: They spell out the conditions that a heterointerface must satisfy to be of interest for incorporation into the active portion of a high-performance device. This poses stringent demands on the concentrations of these defects, to the point that they can rarely be neglected altogether.

The main device physics problems of hetero-interfaces can be roughly divided into problems of the static energy band structure, and problems of the electron transport *within that structure.* I shall concentrate here on the band structure aspects, and ignore the transport aspects. This is not because I consider transport problems less interesting or important (heaven forbid!), but simply because the transport aspects of the device physics are well covered by others at this Symposium. Instead, I will address myself at the end to an area of electronic structure that is not yet in the mainstream of heterostructure device development: The problems of achieving device-quality polar/nonpolar interfaces, involving such pairs as GaAs-on-Ge or GaP-on-Si. This is already an area of active interest to the basic physicist, but so far only from the structural point-of-view, largely neglecting the electrical properties that are the essence of device. Currently, the device physicist is disenchanted about the consistently miserable electrical properties that have resulted whenever device-type structures of this kind have been attempted. I believe that device-quality interfaces in such systems *can* be achieved, but only if both structural and electrical considerations are pursued jointly. This raises some new kinds of problems that simply do not exist in III/V-only systems, but the understanding of which is likely to have benefits far beyond these esoteric mixed systems themselves, feeding back even on such much simpler systems as the familiar GaAs/(Al, Ga)As systems.

2. Energy band diagrams of hetero-interfaces

2.1. Band offsets: the Shockley–Anderson model

From a device physics point-of-view the most important aspect of a semiconductor hetero-interface, and the point of departure for all subsequent considerations is the energy band diagram of the interface. We assume that the transition from one semiconductor takes place over at most a few lattice constants. For such abrupt interfaces the "canonical" energy band model is the *Shockley–Anderson model* [3–6], (Fig. 1). Its characteristic feature is an abrupt change in energy gap at the interface, leading to discontinuities or *offsets* in the conduction and valence band edges. The magnitudes of these offsets are assumed to be characteristic properties of the semiconductor pair involved, essentially independent of doping levels and hence of Fermi level considerations, but possibly dependent on the crystallographic orientation and on other factors influencing the exact arrangement of the atoms near the interface. Far away from the interface, the band energies are governed by the requirement that a bulk semiconductor must be electrically neutral, which fixes the band energies relative to the Fermi level. Except for certain fortuitous doping levels, the combination of specified band offsets with specified band energies at infinity calls for band bending, accommodated by space charge layers near the interface, similar to the space charge layers at p–n homojunctions. The calculation of the exact shape of this band bending is an exercise in electrostatics and Fermi statistics, not of interest here [5].

The band diagram shown in fig. 1 is for an n–n structure (often written n–N structure, to indicate the change in energy gap). As the figure shows, the conduction band offset then leads to a shallow potential notch and a Schottky-barrier-like potential spike barrier, both of which play large roles in the electrical properties of such junctions. Fig. 2 shows two other possibilities, an N–p junction and an n–P junction.

From the device physics point-of-view, the band offsets are the dominant aspect of heterostructure interfaces, and their existence is in fact the principal

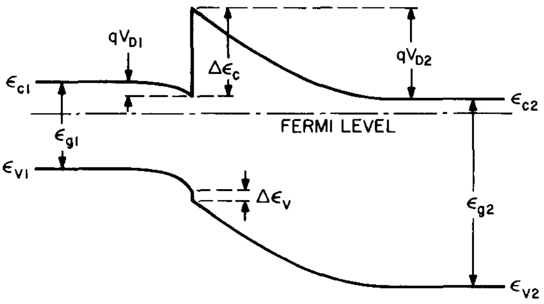
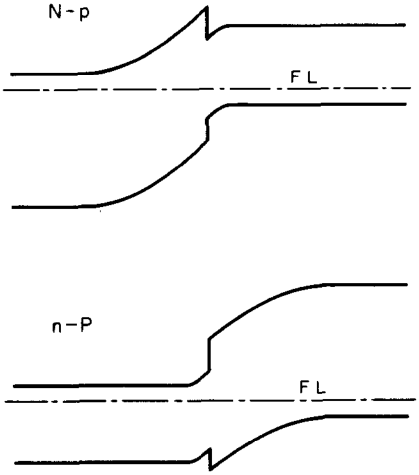


Fig. 1. Band diagram of the Shockley–Anderson model for an abrupt unbiased n–N heterojunction, showing the band edge discontinuities (or offsets) that are the characteristic feature of the model. The specific lineup shown is the "normal" lineup, for whch the narrower forbidden gap falls within the wider gap at the interface.

reason why heterostructures are incorporated into semiconductor, devices: The band offsets act as potential barriers, exerting very strong forces on electrons and holes. These quantum-mechanical "quasi-electric" forces exist in addition to those purely classical electrostatic forces that are due to space charges and applied voltages, which govern carrier flow and distribution in homostructures made from a single semiconductor. The band offset forces may be made either to assist or to counteract the classical electrostatic forces. This gives the device



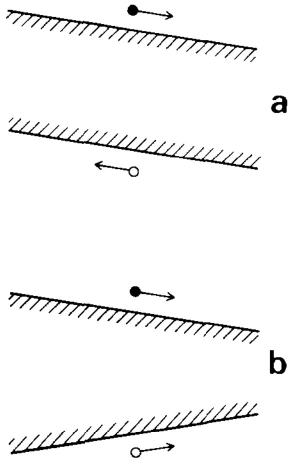


Fig. 2. Band diagrams for N–p (top) and n–P (bottom) heterojunctions.

Fig. 3. Forces on electrons and holes. In a uniform-gap semiconductor (top) the two forces are of equal magnitude but opposite direction, equal to the electrostatic forces $\pm qE$. In a graded-gap structure (bottom) the forces on electrons and holes may be in the same direction. From ref. [2].

physicist an extraordinary new degree of design freedom in controlling the distribution and flow of carriers, to improve the performance of existing devices, and to make possible new kinds of devices.

Basically, it is not the electrostatic force $\pm qE$ that acts as force on the carrier, but the slope of the band edge of the band containing the carrier, multiplied by the sign of the charge of the carrier. In a homostructure, the slopes are necessarily equal to each other and to qE (fig. 3a). But in a heterostructure, energy gap variations cause the slopes of the conduction and valence bands to differ from each other and from the electrostatic force. The case of abrupt band offsets is simply a limiting case; the underlying physics is perhaps clearer by considering the more general case of a graded energy gap, as in fig. 3b, in which only band edge slopes are visible, with no hint as to the magnitude or even the direction of the electric field.

This *general heterostructure design principle* [1,2] may be used in many different ways. A judicious combination of classical electrostatic forces and band gap variations (fig. 3b) makes it possible in a bipolar structure on control the flow of electrons and holes separately and independently. This principle is the basis of operation of the double-heterostructure laser [7,6] that serves as the heart of emerging light-wave communications technology. It also forms the basis of new kinds of improved bipolar transistors [2], and probably of other future devices.

In unipolar devices only one kind of carriers, usually electrons, are present. Here the band offset force has been used with great success in at least two different ways: (a) to confine electrons in quantum wells [8] that are much narrower and have much steeper walls than would be achievable by classical electrostatic forces (= doping) alone; (b) to spatially separate electrons from the donors, against their mutual Coulomb attraction [9]. The latter possibility forms the basis of a rapidly developing new class of field effect transistors [10]. Quantum well structures form the basis of new classes of lasers [11], and they will probably also be responsible for fundamentally new kinds of future device that would not exist at all without quantum wells.

In the energy band diagrams shown in figs. 1 and 2 the signs and magnitudes of the two band offsets were such that at the interface the narrower of the two gaps fell energetically within the wider gap. This "straddling" lineup is the most common case. The most extensively studied of all hetero-interfaces, GaAs/$Al_xGa_{1-x}As$, is of this kind, and its lineup is known to a higher accuracy than that of any other system: For $x < 0.45$, the range in which (Al, Ga)As is a direct-gap semiconductor, the conduction band offset is $85\% \pm 3\%$ of the total energy gap discontinuity ("Dingle's rule" [8]), which translates into a conduction band offset of 10.6 meV per percent of Al. For higher Al concentrations see Casey and Panish [6].

Although the "straddling" lineup, with varying ratios of $\Delta\epsilon_c : \Delta\epsilon_v$, appears to be the most common case, "staggered" lineups, as in fig. 4a, can also occur.

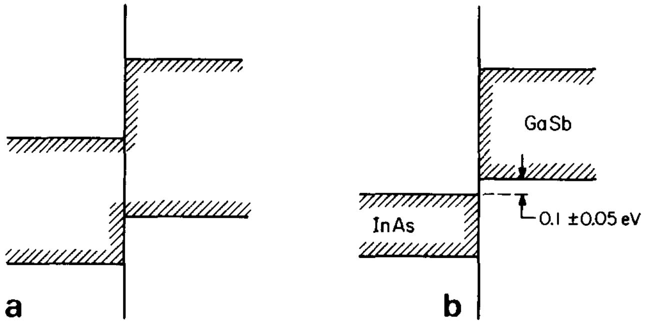


Fig. 4. (a) "Staggered" lineups are expected to occur in many semiconductor pairs. (b) The InAs/GaSb lineup has a broken gap, as shown.

One of the most extreme (and most interesting) lineups is the "broken-gap" lineup at the InAs/GaSb interfaces (fig. 4b): The conduction band edge of InAs falls below the valence band edge of GaSb, by an amount somewhere between 60 and 150 meV [12].

Such different kinds of lineups give the device physicist a powerful device design tool. One of the purposes of this paper is to give a few examples illustrating this point, another is to give some guidance about what governs the lineups in several basic heterosystems. But first we must turn to some of the nuisance effects that complicate considerably the simple Shockley–Anderson model.

2.2. Interface charges

The Shockley–Anderson model in its simplest form described above, is an oversimplification in that it neglects the possibility that there might be interface charges associated with the hetero-interface. Any such interface charge would deform the energy band diagram from that in figs. 1 and 2. Fig. 5 shows the results for an n–N heterostructure, for both signs of the charge. A negative interface charge raises the height of the spike barrier, a positive charge lowers it, and if the positive charge is large enough, the barrier is obliterated altogether, creating instead a potential well. Evidently, interface charges – if strong enough – can have a significant effect on the overall barrier heights seen by the carriers, and hence on the properties of any heterostructure device employing the offset barriers.

Interface charges may arise either from the accumulation of chemical impurities at the interface during growth, or from various kinds of structural defects at the interface. An additional mechanism discussed in detail in section 5 occurs at hetero-interfaces that combine two semiconductors from different columns of the periodic table (example: GaAs/Ge), in which case there will often exist a large net interface charge due to non-cancellation of the ion core charges at the interface.

Major modifications of the band diagram occur already for interface charge densities that are still small compared to monolayer densities. Hence, interface charges can play a non-negligible role even at hetero-interfaces which by any other criterion might be considered interfaces with a high degree of perfection.

Consider GaAs, with a lattice constant $a = 5.653$ Å and a dielectric constant $\epsilon_r = 13$. The density of atoms in a monolayer is $2/a^2 = 6.23 \times 10^{14}$ atoms per cm^2. Suppose the GaAs is doped to a level of 10^{17} cm^{-3}, and a region of $d = 10^{-5}$ cm thickness is depleted at a heterojunction, corresponding to $\sigma = 10^{12}$ charges per cm^2, a number certainly very small compared to a monolayer. The electric field supported by such a charge is $E = q\sigma/E\epsilon_0 \cong 1.4 \times 10^5$ V/cm. The accompanying band bending is $\Delta\epsilon_c = \frac{1}{2}qEd = 0.7$ eV, about twice the band bending occurring at a typical GaAs/(Al, Ga)As n–N heterojunction. Evidently, an interface charge density due to defects of, say, 10^{12} charges per cm^2, equivalent to 1.6×10^{-3} monolayer charges, will change the energy band diagram of such a heterojunction completely, and with it the electrical properties of any device containing this heterojunction. Even much smaller interface charge densities, of the order 10^{-4} monolayers, will still have a significant effect. Unfortunately, effects apparently attributable to interface charges of such small but non-negligible magnitude appear to occur frequently

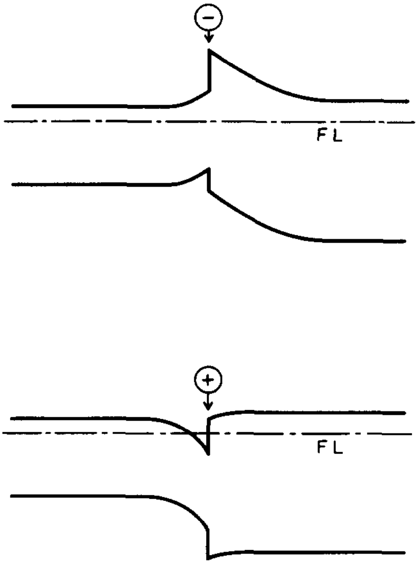


Fig. 5. Band deformation due to a negative (top) or positive (bottom) interface charge.

[13–15]. Evidently, the interface charge is an example of the high degree of sensitivity of the performance of heterojunction devices on the exact atomic structure at or near the hetero-interface, and hence an example of the interrelation between "nanostructure" and device performance.

To a basic physicist, an interface charge of, say, 10^{-3} monolayers may be all but indistinguishable from a "perfect" interface with zero interface charge. But to a device physicist such a small change is a major effect, whose neglect would be unrealistic, and which must be considered along with the band offsets. Still, the roles of the two effects are different: Whereas the band offsets are fundamental and are usually the reason for using heterostructures in devices, the interface charges are almost always a nuisance. Hence we will continue to stress the effects of offsets, raising the issue of interface charges only where necessary.

Unfortunately, interface charges are not the only nuisance: The band offsets themselves appear to be at least somewhat sensitive to exactly how the heterostructure is grown [16], on a level that is not negligible for the device properties, even though it may again be of minor concern to the basic physicist. This introduces another element of uncertainty into the device design, about which we will have to say more later.

3. Band offsets as central device design parameters

3.1. General comments

The extent to which band offsets influence device performance varies tremendously from device to device. At one extreme, the abrupt band offsets may be a nuisance. The heterojunctions in double heterostructure lasers are a good example: Although a varying energy gap is an essential ingredient of the device, a gradual variation would, for various reasons, be greatly preferable over an abrupt step [6,7]. Similar considerations apply to the p–n heterojunctions in heterostructure bipolar transistors [2]. If the semiconductors involved exhibit a continuous mutual solid solubility, the abrupt offsets are easily eliminated by gradient the transition, and this is frequently done.

Of greater interest in the context of this Symposium are devices that call for the retention of the sharp band edge discontinuities, usually with a highly specific kind of mutual band lineup. Many of the more recent heterostructure device concepts are of this kind. Such devices call for a good understanding and knowledge of the band offsets, but exactly what is needed in the way of understanding and knowledge varies greatly from case to case. It depends strongly on the nature of the device; for a given device it changes with the state of development of that device; and more often than not, the needs of the device physicist are again quite different (usually much more severe) than those of the basic physicist. Roughly, the device physicist needs three different levels of knowledge about band offsets:

(a) Semi-quantitative theoretical predictions of the band offsets for as wide a range of semiconductor pairs as possible, to assist in the selection of promising semiconductor pairs to implement new device concepts.

(b) Quantitative data about band offsets, much more accurate than ±0.1 eV, for those semiconductor pairs that are of clear interest for practical devices, to assist in the detailed development of such devices. Ideally, this should not be restricted to accurate empirical data, but would include a theoretical understanding on a level permitting theoretical predictions with this accuracy.

(c) Data about, and a theoretical understanding of, such nuisance effects as offset variations and interface charges.

In the following three sub-sections of this paper (3.2 through 3.5), these three items are taken up, one by one. Only with respect to item (a) does a satisfactory solution exist, and only with respect to this item have the needs of the device physicist been fully met by the interests of the basic physicist. One of the hopes of this writer is that this paper might stimulate the basic physicist to take up a similar interest in the other two problem areas, to contribute to a satisfactory resolution to those problems as well.

3.2. Rough device design: semi-quantitative theoretical offset rules

New heterostructure device concepts, especially the truly novel ones, usually start out as a hypothetical energy band diagram which, if it could be realized in an actual semiconductor structure, would presumably lead to the desired device properties. The solid state photomultiplier proposed by Williams, Capasso and Tsang (= WCT) [17], and discussed by Capasso earlier at this Symposium, is an excellent example. It requires a highly unsymmetric band lineup, with a conduction band offset that is larger than the gap of the narrower-gap semiconductor, and a valence band offset as small as possible. In such cases, in which the choice of semiconductors is not obvious, the first task is to determine whether the needed energy band diagram is in fact achievable by a real semiconductor combination, and whether or not any such combination is compatible with whatever other constraints may be present (lattice matching, mobilities, overall energy gap constraints, etc.). To this end, semi-quantitative predictive lineup rules are required.

The oldest and still widely used such rule is Anderson's *Electron Affinity Rule* [4–6], according to which the conduction band offset should equal the difference in electron affinities between the two semiconductors. Although the rule has been repeatedly criticized on various grounds [18–20], it is better than nothing at all. In fact, it has found vocal defenders [21,22], and it continues to be widely used despite all criticism, largely because its principal competitors, the Frensley–Kroemer theory [23] and the Harrison theory [19,24] are not so overwhelmingly superior to have caused its abandonment.

Although none of these three rules or theories are accurate enough to base a quantitative device design on their predictions, all of them are very useful as semi-quantitative guides. In fact, in simple cases, such as the WCT device [17], even rougher guides may be useful, such as the *Equal Anion Rule* [25]. It states that, for heterojunctions in which the anion atom (the column V or VI element) is the same on both sides, most of the energy gap discontinuity occurs in the conduction band, and the valence band offset is small compared to the conduction band offset. The GaAs/(Al, Ga)As pair has a common anion, and the comparatively small valence band discontinuity in that system $\Delta\epsilon_v \sim 0.15\ \Delta\epsilon_g$ (for an Al fraction less than 0.45) demonstrates both the rule itself and its approximate nature. The rule has a theoretical foundation: For the III/V and II/VI semiconductors, the valence band wave functions are heavily concentrated around the anion atoms, with only a small part of the wave function being near the cation atom. Equal anion atoms thus naturally mean similar valence band energies [26].

Inasmuch as the WCT solid state photomultiplier calls for as small a valence band offset as possible, it naturally calls for a semiconductor pair that shares the anion species, such as a pair of phosphides, arsenides, or antimonides. Lattice matching is an additional important consideration, and because all Al and Ga compounds with the same anion tend to have very similar lattice constants [6], we can restrict the consideration further to the pairs AlP/GaP, AlAs/GaAs, and AlSb/GaSb, or related alloys. A look at the energy gaps eliminates all but the last pair, which remains as the natural candidate. With energy gaps of 1.60 eV (AlSb) and 0.72 eV (GaSb) [27], the equal anion rule predicts a conduction band offset of 0.88 eV, more than enough to exceed the gap of GaSb, and making some allowance for the approximate nature of that rule. In fact, the Harrison theory [19,24,28] predicts a valence band offset of only 0.02 eV, with the GaSb valence band edge actually the lower of the two semiconductors, that is, a very slightly staggered arrangement. Such a 20 meV prediction should not be taken seriously – the whole theory is probably not better than ±0.2 eV – but it certainly suggests that the predictions of the equal-anion rule cannot be far off, and it makes AlSb/GaSb a natural candidate for the WCT device. This is in fact one of the two systems discussed by WCT [17] for their device; the foregoing discussion was intended to illustrate by what simple considerations one arrives at this kind of selection. Because AlSb and GaSb do not lattice-match perfectly (2.66 versus 2.65 Å), the addition of a few percent of As to the AlSb is desirable and probably necessary, but this is a refinement going beyond the semi-quantitative considerations discussed here [17].

The Frensley–Kroemer theory [23] (without the doubtful dipole corrections of that theory) predicts an only slightly different band lineup: $\Delta\epsilon_v \cong 0.05$ eV, with AlSb having the lower valence band. Evidently, this changes little. The widely-uded electron affinity rule [4–6] cannot be applied to this system, because the electron affinity of AlSb is unknown, and we do not consider the use of Van Vechten's theoretical values [29] – suggested by Shay et al. [21] and by Philips [22] – as a reliable substitute: The Harrison theory tends to give more accurate values.

The equal-anion rule can be extended into a prediction of how valence band edges vary as the anion is changed: With increasing electronegativity of the anion, the valence bands tend to move to lower energy [25], essentially because the increase in electronegativity reflects a lowering of the valence electron states within the anion atomic potential. In the case of Au Schottky barriers, a quantitative correlation was found [25] between valence band energies relative to the Fermi level, and the anion electronegativity. In the case of semiconductor heterojunctions, no *quantitative* correlation exists, but the anion electronegativity rule remains a useful *qualitative* predictor – see the broken-gap lineup in InAs/GaSb [12,30] – especially if one compares semiconductors whose energy gaps are not too dissimilar. In such cases the valence bands of the phosphides should be lower than those of the arsenides, which in turn should be lower than those of the antimonides.

This kind of prediction can be of great help if – for whatever reasons – a staggered band lineup is desired. As a good example, consider a superlattice with staggered band lineup as shown in fig 6. There has recently been a strong interest in such superlattices [31], for the following reasons. In a staggered structure, any electrons would accumulate in the low-ϵ_c layers, any holes in the high-ϵ_v layers. If both kinds of layers are thin enough (≤ 100 Å), there would be significant tunneling of both electrons and holes, and the entire superlattice would behave essentially as a homogeneous substance with an overall energy gap smaller than that of either constituent compound, slightly larger than the separation between the highest valence band and the lowest conduction band. Suppose next that the low-ϵ_c layer is n-type doped, and the high-ϵ_v layer p-tape. If selective contacts are made to the n-type and p-type layers, and a bias voltage applied, the effective energy gap is varied. But a voltage-adjustable

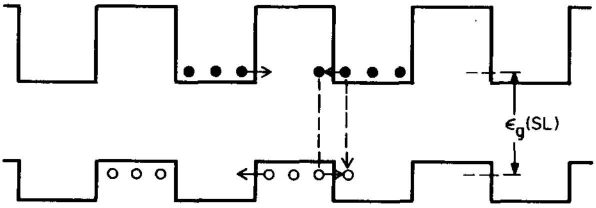


Fig. 6. Staggered-offset superlattice, in which electrons and holes (if present) accumulate in alternating layers. Because of electron tunneling, such structures can have an effective gap narrower than the gaps of both bulk semiconductors.

energy gap would of course be an extremely valueable new phenomenon.

The whole concept is simply an elaboration of the n–i–p–i superlattice concept of Döhler and Ploog [32], except that the spatial separation of the high concentrations of electron and hole from each other is now achieved very easily by the band offset forces, rather than purely electrostatically, by heavy doping.

The occurrence of a broken gap in the InAs/GaSb system suggests that less extreme cases of staggering are indeed achievable, but are they achievable in semiconductors with much larger energy gaps? The anion electronegativity rule [25] suggests that combinations of a phosphide with an antimonide form a promising point of departure. Because phosphides tend to have smaller lattice constants (and larger energy gaps) than antimonides, it is advisable to start with the phosphide that has the largest lattice constant (and the smallest gap), InP, and the combine it with the largest-gap antimonide, AlSb. For this system the Harrison theory [28] predicts indeed staggered band offsets, with a conduction band well depth $\Delta\epsilon_c = 1.20$ eV, a valence band well depth $\Delta\epsilon_v = 0.97$ eV, and a net band separation

$$\epsilon_g(\mathrm{SL}) > \epsilon_c(\mathrm{InP}) - \epsilon_v(\mathrm{AlSb}) \cong 0.4 \text{ eV}.$$

The actual superlattice gap should be somewhat larger, increasing with decreasing superlattice period.

Although the estimate was rough, the message is clear: Staggered superlattices with usefully large gaps should be achievable! Whether or not the simple InP/AlSb pair is indeed a promising pair, remains to be seen, but it is certainly a useful point of departure. If anything, the staggering is larger than needed and the effective gap ($\gtrsim 0.4$ eV) too small to be useful. Evidently, the conditions to obtain staggering may be relaxed somewhat. Now, one of the drawbacks of the InP/AlSb pair is a large lattice mismatch ($\cong 4.6\%$). Such a lattice mismatch, would almost certainly be fatal to device performance in a single-interface heterostructure device due to inevitable misfit dislocations. But it might be quite acceptable in a short-period superlattice, where the lattice misfit can be taken up by elastic strain, a point recently elaborated upon by Osbourn [31] in the context of strained-layer staggered superlattices based on the GaP/Ga(P, As) system. If necessary, the lattice misfit could be reduced by replacing AlSb with an Al(Sb, As) alloy. This would make the valence band well shallower and increase the net gap, but the Harrison theory predicts that even for perfect lattice match to InP, that is, for $\mathrm{AlAs_{0.56}Sb_{0.44}}$ [33], a valence band well of 0.46 eV and a net gap of 0.91 eV should remain. Further fine-tuning could be achieved by replacing some of the Al by Ga [6].

Two other lattice-matched pairs for which staggered lineups can be safely predicted are $\mathrm{InP/Al_{0.50}In_{0.50}As}$ ($\epsilon_g \gtrsim 1.1$ eV) and $\mathrm{Ga_{0.52}In_{0.48}P/AlAs}$ ($\epsilon_g \gtrsim 1.6$ eV).

There is some evidence [34] that the $\mathrm{GaP_xAs_{1-x}}$ system for $x > 0.5$ leads to staggered lineups with large net gaps, but for this system the theoretical predictions are not as clear-cut as for the above combination. We will return to this point later.

3.3. Quantitative device design: the absence of theoretical guidance

Although semi-quantitative lineup prediction rules are very useful in identifying promising hetero-pairs for hypothetical device applications, a detailed device design requires far more accurate values. In any device in which current flows *across* a heterostructure barrier, the current depends on the barrier height $\Delta\epsilon$ at least like a Boltzmann factor $\operatorname{Exp}(-\Delta\epsilon/kT)$, implying a factor e for every change in barrier height by 1 kT ($\cong 26$ meV at 300 K). If the current is tunneling rather than thermionic current, the dependence tends to be even steeper. There is no need to discuss here whether a prediction to some fraction of kT is necessary or whether $\pm 1\ kT$ or even $\pm 2\ kT$ would be sufficient: None of the predictive theories comes anywhere near even the less demanding limit. Those physicists (not involved in actual device design) who have expressed their satisfaction with either the electron affinity rule [21,22] or the Harrison theory [20], quote examples of "excellent agreement" between theory and experiment, in which the predicted offsets vary by 0.2 eV ($\cong 8\ kT$) or more from reliable experimental data. Presumably, then, this is roughly the level of reliability of existing predictive rules or theories. This degree of agreement may indeed be very satisfactory to the fundamental physicist, who wants a general understanding of band offsets; it is totally unsatisfactory as a quantitative basis for device design.

Nor is the need for an accurate prediction significantly less demanding in those devices in which current does not flow *across* a hetero-barrier, but *along* it, as in the new high electron mobility transistor (HEMT) [10] which represents one of the most active areas of heterostructure device research and development, also discussed (from a physics- rather than device-oriented point-of-view) by Störmer at this Symposium. One of the most important design parameters in these devices is their threshold voltage, that is, the gate voltage at which the conductance along the 2D conducting channel is effectively turned on or off (it may be either a positive or a negative voltage, depending on the desired design). To be useful in future high-performance IC's (their dominant area of interest), the threshold voltages of these devices must be predictable much more accurately than ±0.1 V, preferably to ±0.1 V, which calls for a knowledge of the band offsets to within a similar accuracy.

As the HEMT case shows, the absence of any purely theoretical predictive tools with the desired accuracy is not preventing the design of this particular device to go forward. The band offsets at the (Al, Ga)As-on-GaAs (100) interface *are* known to the required degree of accuracy [6]. But this accurate knowledge is the result of accurate *experimental* measurements [8], not of an accurate predictive theory. Once the evolution of a new heterostructure device has progressed beyond the initial speculative stage, to the point of practical device development, it is necessary that the band offsets be accurately *known*, but the knowledge need not come from a predictive theory; knowledge from accurate experimental data may actually be preferable to a theoretical prediction. This de-facto status of the band offsets is similar to that of energy gaps: Whenever available, we use accurate experimental values of energy gaps, rather than theoretical values. Only when experimental data are missing, will we use theoretical ones.

Does any of this mean, however, that the attempts to predict band offsets theoretically have no value beyond the crude semi-quantitative value discussed earlier? Far from it! First of all, the purpose of theories of band offsets (e.g. electron affinity rule, Harrison's theory, etc.) is only partially to provide the device physicist with quantitative design data. A more important role is to test the assumptions that go into each theory, and thereby to test our fundamental understanding of what determines the band offsets. This is similar to the way band structure calculations test our understanding of band structures more than providing accurate theoretical gap values when accurate experimental values are already available. All these are *retrodictive* theories more than *predictive* ones! By that standard, neither the electron affinity rule not the Harrison theory, with their ±0.2–0.3 eV accuracy, are doing badly (nor does the Frensley–Kroemer theory, which is of similar accuracy). Inasmuch as the present paper is to describe a device physicist's view of hetero-interfaces, it does not provide a suitable forum to discuss exactly how well these theories meet the needs of the basic physicist, and much less to discuss critically the enthusiastic support that Shay et al. [21] and Philips [22] have expressed for the electron affinity rule, and Margaritondo and his co-workers [20] for the Harrison theory.

A second reason why more accurate theoretical predictions could be useful as quantitative rather than merely semi-quantitative predictive tools occurs whenever the accuracy of the existing theories is insufficient to yield a clear-cut yes–no decision about a speculative device, but in which experimental data would require the prior development of an elaborate technology. A theoretical guidance on whether or not the development of this technology is worthwhile would be highly useful in such cases [18].

A good example is once again at hand. There has been considerable speculation [31] that a $GaP/GaP_{0.6}As_{0.4}$ superlattice would be of the interesting staggered variety shown in fig. 6. This speculation is partially based on the electron affinity rule, using the electron affinity value of 4.3 eV quoted by Milnes and Feucht [5] without giving any source. Partially it is based on a highly indirect claim by Davis et al. [35] (contradicting other data) that the conduction band offset in the GaP/GaAs system should be near zero. A very careful measurement of the electron affinity has recently been performed by Guichar et al. [36], yielding 3.70 ± 0.05 eV. Using this value, and the known electron affinity for GaAs, 4.07 eV, and making due allowance for the change from direct gap to indirect gap in going from GaAs to GaP, one predicts a conduction band offset for the superlattice of only 0.02 eV, just very slightly staggered. The Harrison theory predicts the same value [28]. With the reliability of both the electron affinity and the Harrison theory rule being no better than ±0.2 eV, this prediction is simply a draw. Inasmuch as the development of an entire superlattice technology hinges on this prediction, it is an excellent example of why more accurate predictions would indeed be desirable.

Recent experiments suggest [34] that the superlattice is indeed staggered, by about 0.2 eV. If future measurements confirm this result, this would show that both theoretical predictions are indeed incorrect by about 0.2 eV.

3.4. The nuisance effects: offset variations and interface charges

In the preceding discussion we pointed out the device physicist's need for knowing band offsets to an accuracy much better than ±0.1 eV. But this request implicitly assumed that the band offsets are in fact constants that characterize a given semiconductor pair, rather than being variables themselves. As was pointed out by Bauer [37] and by Margaritondo [20] at this Symposium, evidence is accumulating [16,38] that the offsets are process-dependent, changeable over a finite range outside of the tolerance limits of the device designer. A dependence on crystallographic orientation is almost to be expected, and while it might be a nuisance, it does not introduce any problems into device design. Nor do we need to be surprised about large offset variations in systems in which a compound semiconductor (GaAs, GaP) is grown on an elemental semiconductor (Ge, Si), or vice versa, the cases of particular interest to Bauer [37] and Margaritondo [20]. We shall argue in section 5 that in such systems technology-dependent offset variations and interface charges *are to be expected*. What *is* disturbing are offset variations and interface charges in such supposedly well-behaved lattice-matched systems as GaAs/(Al, Ga)As. It was found by Waldrop et al. [16] that for ⟨110⟩-oriented MBE growth at a substrate temperature of 580°C the band lineup depends noticeably on whether AlAs is grown on GaAs ($\Delta\epsilon_v \cong 0.15$ eV), or GaAs on AlAs ($\Delta\epsilon_v \cong 0.40$ eV). By comparison, the ⟨100⟩-lineup data of Dingle [8] for GaAs/(Al, Ga)As, extrapolated to Gas/AlAs, corresponds to an in-between value of $\Delta\epsilon_v \cong 0.20$ eV.

Although differences between ⟨100⟩ and ⟨110⟩ might have been expected, the strong growth sequence dependence for the ⟨110⟩ orientation comes as a rude shock. For a given orientation, band offsets can depend on growth sequence only through differences in the exact atomic arrangement near the interface. Evidently the atomic arrangements for ⟨110⟩ interfaces depend strongly on growth sequence. Put bluntly: At least for this orientation the offsets depend quite strongly on technology [39] rather than being a fundamental materials parameter! The question naturally arises whether or not this might quite generally be the case. Might there be a similar growth sequence dependence for {100} growth? I find it hard to believe that any significant growth sequence asymmetry of {100} band offsets would leave intact the superb fit of Dingle's superlattice data (which automatically involve both growth sequences) to a single-offset model, especially considering Dingle's wide range of layer thicknesses. Yet there exists strong evidence that, if not the band offsets, at least the transport properties in the 2D electron gas along GaAs/(Al, Ga)As {100} heterojunctions, depend quite strongly on the growth sequence [38], with higher mobilities occurring for (Al, Ga)As-on-GaAs than for GaAs-on-(Al, Ga)As. In fact, it appears that in structures containing multiple interfaces, the properties of the interfaces grown first differ from those grown later [40]!

One frequently hears the argument that effects such as these are somehow artifacts of the growth process, reflecting "bad" interfaces. While in a practical sense this might be true, it avoids the fundamental issue: Even a "bad" interface must have some atomic configuration that causes these effects, and which configuration constitutes "badness"? And can this "badness" in fact be avoided under the numerous constraints imposed upon the growth of an actual device?

We clearly need an understanding of these effects, and this may indeed by one of the most urgent research topics in which the device physicist would like to see the basic physicist take an active interest. To the basic interface physicist, offset variations of $\equiv 50$ meV might be a minor nuisance, negligible to the basic understanding of the interface physics. But the degree to which these offset variations can be controlled, may be decisive for the role heterostructure FET's will play in future high-speed VLSI technology.

A return to the earlier example of HEMT threshold voltages will illustrate the urgency. As we stated, these threshold voltages depend on several structural parameters, one of which is the conduction band offset. Now the most important envisaged applications of this transistor is in future very fast large-scale digital integrated circuits which may contain anywhere from 10^3 to 10^6 identical FET's per chip. For a variety of reasons, it is necessary that the threshold voltages of all transistors on the same chip have essentially the same value, *and* that this design value can be technologically maintained from chip to chip and even from wafer to wafer. Threshold voltage variations far below ± 0.1 V on a single chip are essential, or else the IC will simply not work, and variations below 10 mV are desirable. Worse, the variations from chip to chip should not be much larger. Evidently this calls for tight tolerances on the band offsets and on residual interface charges.

To achieve these tolerances requires an understanding of what causes offset variations and interface charges, not just purely empirical tight process control. In fact, it is probably more important to develop a physical understanding of offset *variations* on the ± 5 meV level than to be able to *predict* the exact magnitude of these offsets to better than ± 0.1 eV.

4. On measuring band offsets experimentally

4.1. Introductory comments

There does not exist any experimental technique to determine band offsets that is simultaneously simple, reliable, and universally applicable.

The most careful and presumably most accurate determination of any band lineup is Dingle's well-known work [8] on the infrared absorption spectra of superlattices of weakly-coupled multiple GaAs/(Al, Ga)As quantum wells. Dingle was able to fit large numbers of data, for wells of various widths, to a single model in which the conduction band offset is 85% ± 3% of the energy gap difference.

For sufficiently narrow wells, the method is fairly insensitive to errors by small interface charges. Major distortions in the well shape would quickly destroy the excellent fit of the experimental data to the simple square-well model. Dingle's data prominently include transitions involving the higher energy levels in the wells, which would be especially sensitive to any distortions of the well shape. It is hard to believe that the large number of observed transitions, over a wide range of well widths, could be fitted just as well to a significantly different well shape. This same quality-of-fit argument also speaks against various kinds of modifications in the band offsets, such as growth sequence asymmetries, etc. Certainly, the burden of the proof for any such modifications lies with those who would propose such modifications. Note, however, that Dingle's data, being strictly {001} data, in no way rule out any offset dependence on crystallographic orientation.

A second widely used technique to determine band offsets is based on photoelectron spectroscopy [20,41], executed with various levels of sophistication. It is even less sensitive to interface charges, and is in principle capable of giving quite accurate offsets, perhaps more directly than Dingle's technique. Especially the Rockwell group of Kraut, Grant, Waldrop and Kowalczyk [41] has cultivated this technique to a high level of perfection, to the point that in favorable cases offsets with (believable) uncertainties of ±0.03 eV were obtained. Inasmuch as Margaritondo, another practitioner of this technique, has discussed it at this Symposium, we refer to his paper [20] for more information and references.

Both the superlattice absorption technique and the photoelectron spectroscopy technique are "physicist's techniques", rather than device-type techniques. Now we argued earlier in this paper that the properties of heterostructure devices depend sensitively on band offsets. It should therefore be possible to extract accurate band offsets from measurements on devices. Because of the simplicity of purely electrical measurements, such attempts have indeed often been made [5], and many band offsets found in the literature were in fact obtained from purely electrical measurements, usually on simple p–n or n–n heterojunctions. Unfortunately, such measurements are sensitive not only to band offsets; they are just as sensitive to other phenomena that deform the band diagram, especially interface charges. Most of the electrical measurements have difficulty separating these effects. More often than not the data are merely fitted to the simple Shockley–Anderson model ignoring such complications, which can lead to grossly inaccurate band offsets.

Inasmuch as this paper represents a review of hetero-interfaces from the device physicist's point-of-view, a critical review of the main techniques is in order.

4.2. Capacitance–voltage profiling

Probably the best of the purely electrical measurement techniques is based on a powerful adaptation of conventional C–V impurity profiling, recently developed by Kroemer et al. [14,42]. It can, under favorable circumstances, give reliable *separate* values for both the band offsets and any interface charges. The method requires an n–n heterojunction whose doping profile is known, a condition often satisfied for junctions grown by highly developed technologies such as MBE. A Schottky barrier is placed on the outer surface of the heterostructure, parallel to the hetero-interface, and the C–V relation of the Schottky barrier rather than of the heterojunctions itself is measured. The method works best with heterojunctions exhibiting poor rectification, which are particularly hard to evaluate by other means. An *apparent electron concentration* $\hat{n}$ is determined by the conventional interpretation of C–V profiling theory [42],

$$\frac{\mathrm{d}}{\mathrm{d}V}\frac{1}{C^2}=\frac{2}{q\epsilon}\frac{1}{\hat{n}(x)}, \tag{1}$$

where C is the capacitance per unit area, and $x=\epsilon/C$. The $\hat{n}(x)$ profile will differ both from the doping profile $n_\mathrm{d}(x)$ and from the true electron concentration $n(x)$. But if the doping distribution $n_\mathrm{d}(x)$ is known, the interface charge is easily obtained by integrating the apparent difference distribution $\hat{n}(x)-n_\mathrm{d}(x)$, and the conduction band offset is obtained from the first moment of this difference distribution. The true electron distribution is not needed! The method is simple and powerful, and readily applicable to any technology that permits the growth of heterostructures in which the doping level can be kept accurately constant on both sides of the interface, with an abrupt switch at the interface. The two constant doping levels need not even be predetermined; they may be extracted from the C–V profile itself.

The method may be made self-checking, by using the two doping values, the interface charge, and the band offset, to simulate on a computer the C–V profile that *should* have been seen experimentally, and by comparing this reconstructed profile with the profile actually observed.

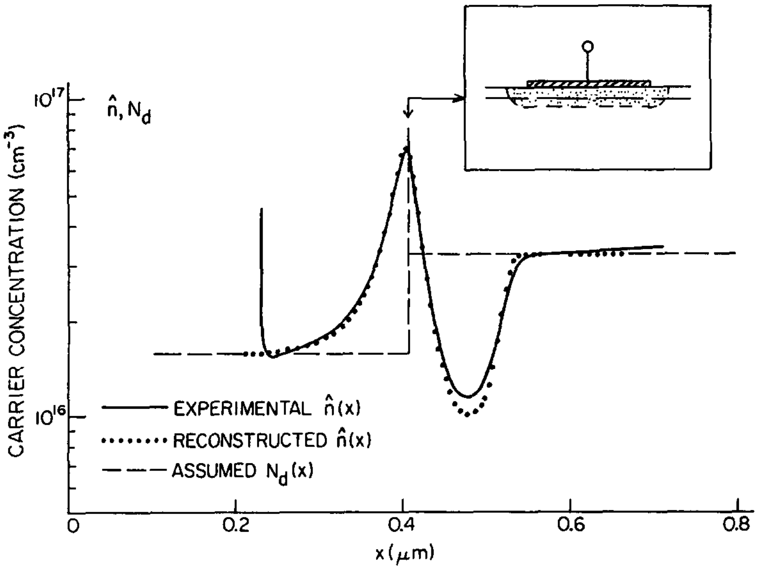


Fig. 7. C–V profiling through an LPE-grown GaAs/(Al, Ga)As n–N junction, after ref. [14]. From the measured apparent electron concentration $\hat{n}(x)$ (solid curve) and the assumed donor distribution $n_D(x)$ (broken curve) one can calculate a conduction band offset $\Delta\epsilon_c = 0.248$ eV and an interface charge density $\sigma_I = +2.7 \times 10^{10}$ cm^{-3}. The inset shows the basic test arrangement.

Fig. 7 shows an example, from ref. [14], for an LPE-grown n–N heterostructure, not ideally suited for the purpose, but so far the only published result in which the method has been used for a quantitative determination of both a band offset and an interface charge, including the self-consistency check. The technique should be even better suited to MBE- or MOCVD-grown interfaces, in which an abrupt transition with flat adjacent doping levels is more easily achieved, and this writer does in fact expect that it will be widely used in the future.

4.3. The C–V intercept method

When the doping level n_d and hence the electron concentration n in an n-type semiconductor is position-independent, the C–V profiling theorem (1) yields a linear C^{-2}-versus-V plot. This remains true for the capacitance of a p–n junction, including a p–n heterojunction, if the carrier concentrations on both sides are constant. This has led to the *C–V intercept method*, which claims that the intercept voltage V_{int} in such a linear C^{-2}-versus-V plot is exactly equal to the total built-in voltage of the heterojunction (fig. 8), sometimes called the *diffusion voltage*,

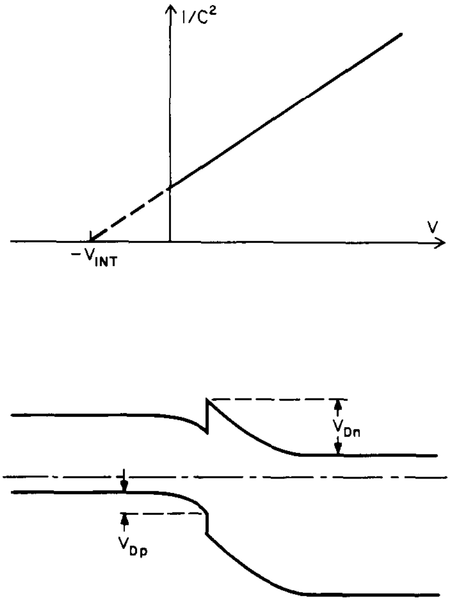


Fig. 8. The C–V intercept method of determining the band offsets at p–N heterojunctions. *If* the heterojunction is abrupt, with constant doping levels right to the interface (no grading), and without any interface charges, then the intercept voltage V_{int} in a C^{-2}-versus-V plot is related to the two diffusion voltages V_{Dn} and V_{Dp} via eq. (3). If both doping levels (and hence both Fermi energies) are known, this permits a determination of the band offsets. The method is sensitive to errors caused by grading or interface charge effects.

$$V_{int} \overset{?}{=} V_{Dn} - V_{Dp}. \tag{2}$$

For known doping levels, the energy separations between the bulk band edges and the Fermi level are known, and hence the band offsets are known if $V_{Dn} + V_{Dp}$ is known. Unfortunately, the accuracy of eq. (2) is largely a (persistent) muth. First of all, (2) neglects the so-called Gummel–Scharfetter correction [43]; it should really read

$$V_{int} = V_{Dn} + V_{Dp} + 2kT/q, \tag{3}$$

a small correction, but not a negligible one. More important: Even in the form (3), the intercept rule is strictly valid only if both doping levels are constant right to the hetero-interface, forming an abrupt transition there, and if no interface charges are present [42,44]. Interface charges tend to lower the intercept voltage, whereas impurity grading effects raise it. A small region right at the interface always remains inaccessible, even if C–V profiling is extended to forward bias values. Any space charge re-adjustments entirely inside this region will not affect the linearity of the C^{-2}-versus-V plot unless the charge inside the depletion region somehow depends on the applied voltage (which may be the case for deep levels, but not otherwise). Although these facts have been established for some 25 years [44], they remain strangely ignored except by a small fraternity of semiconductor device physicists intimately familiar with $C-V$ profiling theory. Even as astute a researcher as Phillips [22] writes in a recent paper: "The great merit of this technique is that it is self-checking, i.e., when chargeable traps are present at the interface, C^{-2} is not a linear function of V_a. The deviations from linearity automatically provide estimates of the accuracy of the determination of V_D and from it the accuracy of ΔE_C and ΔE_V." Well, they don't. To get experimental access to the charges located right near the interface, one must profile *through* the heterojunction from the outside, as described earlier, not from the interface outward.

Considering this inherent weakness of the intercept method, it is not surprising that the offset values determined by it have fluctuated widely whenever data from more than one investigator have been available, and often even for the data from the same group. Two examples are provided by the chaos in the offset data reported for GaP/GaAs and Ge/GaAs. In most of these measurements, $C-V$ intercept data were not used alone, but in conjunction with current–voltage ($I-V$) data. However, this hardly excuses the failure of the intercept method to "catch" the ever greater inadequacies of the $I-V$ techniques.

In the case of GaP/GaAs, the reported conduction band offsets vary by at least 0.65 eV: Weinstein et al. [45] claim $\Delta\epsilon_c \cong 0.22$ eV, Alferov et al. [46], $\Delta\epsilon_c \cong 0.65$ eV, and Davis et al. [35], $\Delta\epsilon_c \cong 0$. It is anybody's guess which of these values is least far away from the truth – if there is in fact a single "true" value.

The situation for Ge/GaAs is, if anything, even worse. Conduction band offsets varying from 0.09 to 0.54 eV can be found in the literature, a range corresponding to 68% of the energy gap of the narrower-gap semiconductor, Ge. The reason is probably only partially due to erratic measurements. As we shall see later, for polar/nonpolar systems such as GaAs/Ge, an erratic technology-dependence of the offsets should be expected.

Despite this history of unreliable results, the intercept method should be capable of yielding accurate offsets *if* the uncertainties inherent in it are treated with due respect, and are eliminated by suitable complementary data, especially for interfaces grown by one of the better and more tractable technologies, such as MBE or MOCVD. There is something inherently satisfactory about $C-V$ profiling measurements: They are essentially purely electrostatic measurements of equilibrium charge distributions versus position, almost completely unencumbered by transport effects.

4.4. *Current–voltage measurements*

Whatever criticisms one might have of band offsets based primarily on $C-V$ intercepts, most of those based on current–voltage ($I-V$) data on p–n or n–n heterojunctions are even less well-founded. Exceptions tend to occur for systems with unusual band lineups, in which the $I-V$ data on heterojunctions differ already qualitatively in drastic ways from those of ordinary p–n homojunctions. The outstanding (but not the only) example is the striking broken-gap lineup at the InAs/GaSb interface (fig. 4b), for which the first experimental evidence was obtained [12] from systematic rectification experiments with lattice-matched Ga(As, Sb)/(Ga, In)As p–n heterojunctions of varying (lattice-matched) alloy compositions. As the GaSb/InAs end was approached, all rectification effects suddenly disappeared, due to the "uncrossing" of the forbidden gaps.

But $I-V$ data on p–n heterojunctions without special lineup feature tend not to contain enough qualitatively different detail to be useful for quantitative offset determination, although they may be useful to supplement other data.

Worst, $I-V$ data on n–N rather than p–n heterojunctions, although they could in principle be quite informative, have in the past been largely worthless. For example, the claim that the conduction band offset of GaP–Si interfaces is essentially zero, is based on nothing more than the failure to observe any rectification effects in Si-on-GaP n–n junctions even at liquid nitrogen temperature [47]. More recent data on this system show [48,49] this claim to be quite false. How erroneous such absence-of-rectification data can be, is illustrated by what is now the best understood heterostructure of all, the GaAs/(Al, Ga)As structure: Most early data on this system showed a more or less complete absence of rectification in n–N junctions [50]. The explanation in terms of zero conduction band offset flatly contradicted Dingle's lineup data. The problem seems to have gone away with subsequent improvements in technology; it was almost certainly due to donor-like defects at the interface, as first proposed by Kroemer et al. [13]. Similar donor-like defects were probably responsible for the lack of rectification in Si/GaP heterojunctions [47].

5. Polar/nonpolar heterostructures

5.1. *Motivation*

Almost all heterostructure *device* structures currently under active investigation employ heterostructures between III/V compounds only. There are strong incentives to extend heterostructure device technology to other systems, especially to combinations of a III/V semiconductor with one of the elemental semiconductors, Ge or Si. Natural pairs, because of their close lattice match, would be GaAs/Ge and GaP/Si. The latter is particularly interesting. If device-quality interfaces between GaP and Si could be achieved, this would be a major advance towards bridging the wide gap between highly-developed Si technology and the rapidly developing technology of III/V compounds, with potentially far-reaching device applications.

A number of attempts to grow such polar/nonpolar heterostructures have led to disappointing results: These systems are clearly far more difficult than III/V-only heterosystems. However, a physical understanding of these systems is beginning to emerge that explains why many of the earlier purely empirical "cookbook" approaches *should* have failed, and which suggests that a better understanding of both the growth mechanism and the electronic structure of these interfaces might make possible substantial progress towards the elusive goal of device-quality polar/nonpolar heterostructures.

In fact, the incentives to achieve such a better understanding go far beyond the device utilization of polar/nonpolar interfaces themselves: It would also advance the understanding of more "ordinary" III/V-only interfaces. Many of the problems that occur at polar/nonpolar interfaces are simply hugely magnified versions of problems that occur already at the GaAs/(Al, Ga)As interface. Examples: Residual interface charges, offset variations, crystallographic orientation dependence, and technology dependence. The difference is purely quantitative: In the III/V-only cases these problems are second-order nuisances, in the polar/nonpolar cases they dominate. I believe this dominance is the reason why polar/nonpolar interfaces have so far proven so intractable. It is reasonable to expect that a better understanding of these effects, leading to control in the polar/nonpolar case, will also greatly benefit the III/V-only case.

5.2. Interface neutrality and crystallographic orientation

In 1978, Harrison, Kraut, Waldrop and Grant (HKWG) published a classical paper [51] that forms the point of departure for any rational understanding of the problems of polar/nonpolar interfaces. The authors studied the electrostatics of the simplest possible atomic configurations for the three lowest-index orientations of an ideal GaAs/Ge hetero-interface. They showed that for both the {100} and {111} orientations these atomic configurations correspond to a huge net electrostatic interface charge, of the order of one-half of a monolayer charge. The argument is brought out in fig. 9 for the (001) interface, viewed in the $[\bar{1}10]$ direction. The black circles represent Ga atoms, the white circles As atoms, and the shaded ones, Ge. An alternate possibility has Ga and As interchanged. An important point in the HKWG argument is a point emphasized earlier by Harrison [52]: The tetrahedral bond configuration *guarantees* that each of the bonds connecting each atom to its four nearest neighbors contains exactly two electrons, just as in Ge, and regardless of whether the bonds are Ge–Ge, Ga–As, or mixed Ga–Ge or Ge–As bonds. Only the electron distribution along each bond depends on these details, not the overall bond charge. This means that the net electrical charge associated with the overall interface region can be determined by simply counting each column-V atom as having one extra proton charge relative to a neutral column-IV atom, and each column-III atom as missing one such charge. The overall interface charge is easily obtained by a fictitious process, whimsically called "theoretical alchemy", in which one pretends that the GaAs portion of the heterostructure has been obtained from a Ge single crystal by moving a proton lattice from one-half of the Ge atoms to the other half of the Ge atoms, creating Ga and As in the process. Depending on whether the fictitious proton motion is away from the interface or towards it, a negative or positive charge imbalance is thereby created at the interface. The bottom half of fig. 9 shows the electrostatic potential resulting from a proton transfer away from the interface, with the electron distribution along the bonds initially kept fixed. The potential staircase on the GaAs side is evident. The average slope of this staircase represents a net electric field, which is easily shown to be that of a charge of $-q/2$ per interface atom. With an interface atom density of $2/a^2$, this is a charge density $-q/a^2$. The important point is now that the bond charge relaxation following the proton transfer does not change the net

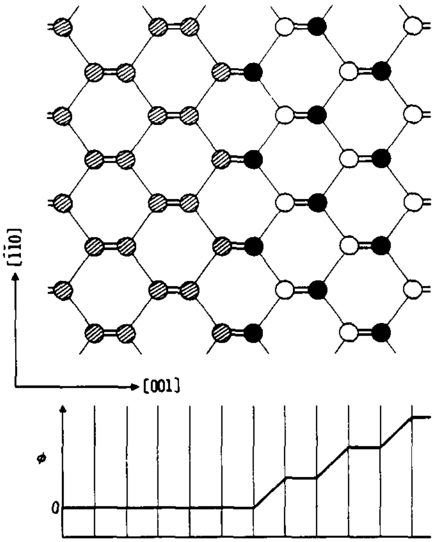


Fig. 9. Atomic arrangement and electrostatic potential at an idealized unreconstructued Ge/GaAs(001) interface, from ref. [51]. The idealized atomic arrangement exhibits a large charge imbalance at the interface, leading to a staircase potential with a large net electric field on the GaAs side. The full circles represent Ga atoms, the open circles As atoms.

interface charge, even though it is strong enough to actually reverse the sign of the net charge on the Ga and As atoms inside the GaAs side. But the total charge per bond always remains at exactly two electrons; no net charge crosses the Ga and As atomic planes inside the GaAs side, implying conservation of net interface charge during the relaxation. In terms of the potential diagram in fig. 9, the shape of the individual steps in the staircase changes, but the net *average* slope remains unchanged.

As HKWG point out, the field supported by the net interface charge is huge ($E = q/a^2\epsilon \cong 4 \times 10^7$ V/cm, assuming the dielectric constant of GaAs), sufficient to guarantee an atomic re-arrangement during the crystal growth itself, to minimize those interface charges. The authors give two specific atomic configurations which lead to zero interface charge, shown in figs. 10 and 11. The first of these contains one mixed-composition layer, but it retains a finite interface dipole. In the second configuration, containing two mixed-composition layers, the interface dipole has also been obliterated. The authors speculate that the second configuration might actually arise during epitaxial growth.

It is at this point that we must differ from HKWG. Although there can be no doubt that a drastic atomic re-arrangement will take place, and almost certainly in the general direction postulated by HKWG, it appears inconceiva-

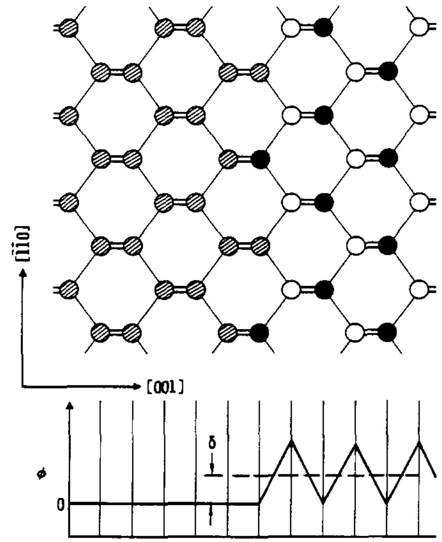


Fig. 10. Modified atomic arrangement and electrostatic potential at a Ge/GaAs(001) interfa containing one atomic plane of mixed composition, with zero net interface charge, but retaining finite interface dipole. From ref. [51].

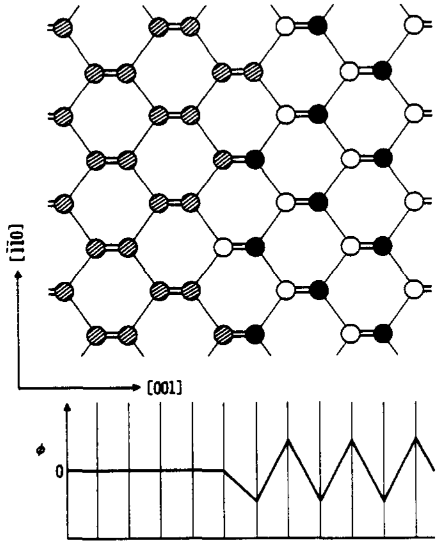


Fig. 11. Further modification of the atomic arrangement at a Ge/GaAs (001) interface, containing two atomic planes of mixed composition yielding both zero interface charge and a zero interface dipole. From ref. [51].

ble that any such re-arrangement goes sufficiently far towards completion that the remaining interface charge becomes negligible for device purposes. We recall that even a charge of only 10^{-3} monolayers is still a large interface charge for device purposes; even if the interface atomic re-arrangement goes 99% towards completion, this would still leave an intolerably large charge five times as large.

We therefore conclude that, at least for the {100} orientation, large residual interface charges must be expected at GaAs/Ge and similar polar/nonpolar interfaces. Worse, the exact amount of interface charge left must be expected to depend on the growth process. Hence the interface charges will not only be large, but technology-dependent. Finally, because even for zero interface charge the residual interface dipoles still depend on exactly which atomic re-arrangement was created, the band offsets must also be expected to be technology-dependent and hence poorly reproducible.

There are mitigating circumstances present if the growth sequence is nonpolar-on-polar. Harrison has pointed out [53] that the electrostatic arguments of HKWG also apply, with some modification, to the free surface of a compound semiconductor. A GaAs {001} surface terminating in complete Ga or As planes is electrostatically just as unfavorable as an ideal GaAs/Ge interface. The actual atomic configuration present at a free GaAs {100} surface will already be such that the net surface charge is minimized. If all dangling surface bonds dimerize, apparently a good first-order approximation, an atomic arrangement leading to a neutral surface will also lead to a neutral Ge/GaAs interface, if the vacuum is subsequently replaced by Ge.

But this argument does not apply if GaAs is grown on Ge. Thus we are led to a second prediction: Polar/nonpolar interfaces must be expected to exhibit drastic growth sequence dependences, much stronger than those observed in the GaAs/(Al, Ga)As system. Unfortunately, the more difficult polar-on-nonpolar growth sequence is demanded in the majority of device applications. In my opinion, attempts to grow GaAs/Ge or similar polar-on-nonpolar {100} heterojunctions or – worse – polar/nonpolar superlattices with this orientation, in the hope that device-quality interfaces will somehow result, are likely to be little more than a waste of time. The fact that this orientation is so successful for III/V-only growth is quite irrelevant. The likely answer – if any – to the quest for successful polar-on-nonpolar growth lies in the use of one of the nonpolar orientations to be discussed presently.

The HKWG argument is by no means restricted to the {100} orientation. Qualitatively similar arguments with only minor quantitative modifications can be made for {111}-oriented interfaces, and in fact for all interface orientations except those in which the interface is parallel to one of the ⟨111⟩ bond direction.

The condition for this can be expressed as a mathematical condition on the Miller indices (*hkl*) of the interface [54]. Let [*hkl*] be the direction perpendicular to the interface plane. The plane is parallel to one of the ⟨111⟩ bond

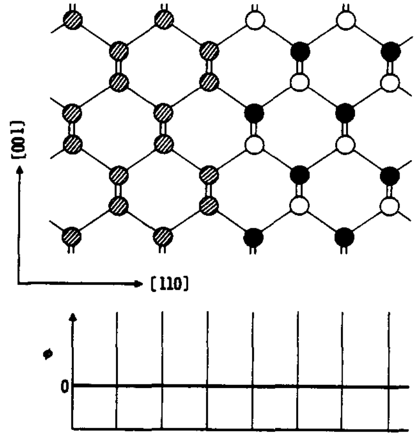


Fig. 12. Atomic arrangement and electrostatic potential at an ideal Ga/GaAs(110) interface. Each GaAs plane parallel to the interface contains an equal number of Ga and As atoms and is hence electrically neutral. From ref. [51].

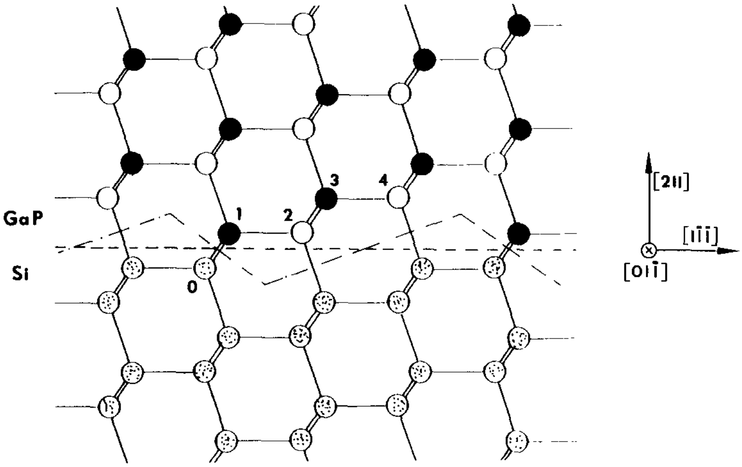


Fig. 13. Atomic arrangement at idealized GaP/Si(211) interface, from ref. [54]. As in the {110} case, each GaP plane parallel to the interface contains an equal number of Ga and P atoms and is hence electrically neutral. But in addition, the bonding of the "black" sublattice sites across the interface is much stronger (two bonds) than that of the "white" sublattice sites (one bond). When GaP is grown on Si, this bonding difference can be utilized to achieve growth free of antiphase disorder, with the "black" sublattice occupied by P atoms, the white by Ga atoms.

directions if $[hkl]$ is perpendicular to that direction. This implies

$$[hkl]\cdot\langle 111\rangle = \pm h \pm k \pm l = 0,$$

for at least two of the eight possible independent sign combinations. The simplest such orientation is the {110} orientation, already recognized as such and intensively discussed by HKWG. The next-simplest orientation is {112}, followed by {123}, etc. Figs. 12 and 13 show the atomic arrangements at a (110) and at a (112)-oriented polar/nonpolar interface, both viewed again in the [$\bar{1}$10] direction.

In the absence of specific reasons to do otherwise, it is probably advisable to use the lowest-index orientation for the epitaxial growth. If only the nonpolar-on-polar growth sequence is needed for a particular device, the {110} orientation may indeed be the preferred orientation. Inasmuch as the {110} planes are the natural cleavage planes of III/V compounds, this happily coincides with the natural interest of the surface physicist in this orientation: Most of the non-device studies of the initial growth of Ge on GaAs have indeed used these planes. However, if the polar-on-nonpolar growth sequence is demanded (which automatically induces polar/nonpolar superlattices), altogether new considerations intervene.

5.3. Polar-on-nonpolar growth: the site allocation problem

When, in a polar/nonpolar heterosystem, the polar (compound) semiconductor is to be grown on the nonpolar (elemental) one, a new problem arises [54,55]: Avoiding antiphase disorder in the growing compound semiconductor. This problem does not exist at all in element-on-compound growth, and it is at most a minor problem in compound-on-compound growth. But for compound-on-element growth it is as severe and fundamental as the interface neutrality problem at {001} polar/nonpolar interfaces, and it totally dominates the problem of polar-on-nonpolar growth for nonpolar orientations, such as {110} and {112}.

When a binary compound with two different atoms per primitive cell (e.g. GaAs, GaP) is grown on an elementary substrate (e.g. Ge, Si) in which the two atoms are identical, there exists an inherent ambiguity in the nucleation of the compound, with two different possible atomic arrangements, distinguished by an interchange of the two sublattices of the compound. If different portions of the growth exhibit different sublattice ordering, antiphase domains result, separated by antiphase domain boundaries, a defect similar to grain and twin boundaries. For high-performance devices, antiphase domain boundaries must almost certainly be avoided, which calls for a rigorous suppression of one of

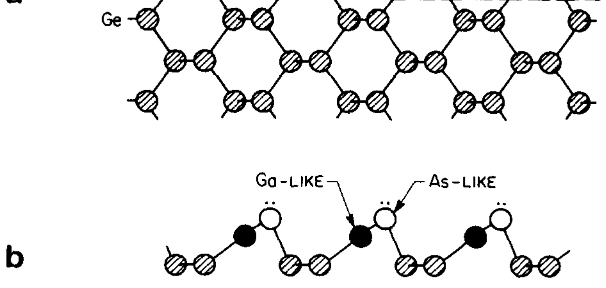


Fig. 14. (a) Occurrence of antiphase domain disorder in the growth of GaAs on an unreconstructed Ge {110} surface, due to the absence of a built-in bonding difference for the as-yet unoccupied surface sites belonging to the two sublattices. (b) Creation of Ga-like and As-like electronic configurations in the top Ge {110} atomic layer, due to reconstruction, aiding in the suppression of antiphase disorder inside the GaAs. From ref. [55].

the two nucleation modes. The problems in doing so depend very strongly on the exact atomic arrangement and on the dangling-bond configuration at the surface of the elemental semiconductor substrate. Unfortunately, they are particularly severe for the simplest nonpolar interface orientation, the {110} orientation. The situation is illustrated in fig. 14a, which shows that on an ideal and perfectly flat (= unreconstructed) Ge {110} surface the sites subsequently to be occupied by Ga and by As atoms have no built-in distinction between themselves. The relative Ga/As ordering at different nucleation sites should therefore be perfectly random, which in turn would lead to a high degree of antiphase domain disorder, with domain sizes of the order of the nucleation site separation, which is usually very small for good epitaxial growth.

The situation on the {112} surface is far more favorable. As fig. 13 shows, the unoccupied sites ahead of an ideal (112) surface are of two quite different kinds: Sites (labelled 1 in fig. 13) with two back bonds to the Si surface, and sites (Nos. 2 and 4) with only one back bond. One easily sees that the two kinds of sites belong to the two different sublattices. Now it is well known that the column-V elements P, As, and Sb, form chemical compounds with Ge and Si, whereas the column-III elements Al, Ga and In do not. One might therefore expect that the strongly-bonding column-V atoms might displace any column-III atoms from the doubly back-bonded sites (No. 1). But once site No. 1 has been occupied by a column-V atoms, site No. 2 becomes more favorable for occupancy by a column-III atom than by a column-V atom. This, in turn, favors occupancy of site No. 3 by another P atom, followed by another Ga atom on site No. 4. Apparently, this is indeed that happens: We have grown GaP on Si {112} by MBE [54], and tests show that the observed sublattice ordering is as described here, with no evidence of antiphase domains. Furthermore, although the electrical properties of these first GaP-on-Si {112} interfaces are still far from ideal, we were able to build bipolar n–p–n transistors with an n-type GaP emitter on a Si p–n base/collector structure, with emitter injection efficiencies up to 90%. This is still far below what would be desirable for practically useful devices (> 99%), but is far better than anything else ever achieved in the very difficult GaP-on-Si system. It raises the hope that device-quality polar-on-nonpolar hetero-interfaces might in fact be achievable.

Our above theoretical speculation was oversimplified in that the reconstruction of the free Ge or Si surface, which is unquestionably present, was ignored. because of the strong bonding difference present already in the unreconstructed {112} surface, any reconstruction on that surface [56] should be little more than a quantiative complication, unless the reconstruction somehow destroys the strong inherent surface site inequivalence, which is extremely unlikely. The situation on the {110} surface is entirely different. Here any reconstruction would *create* a site inequivalence (see fig. 14b), and if this inequivalence is of the right kind, it might convert a hopeless orientation into a promising one. As we have pointed out elsewhere [55], the simplest possible reconstruction, a bond rotation similar to that on GaAs {001}, and postulated by Harrison [57] to occur on Si {110}, is exactly of the most desirable kind. In fact, growth of GaAs on Ge {110} apparently free from antiphase disorder can be achieved under certain growth conditions [55], which unfortunately however do not appear to lead to device-quality electrical properties. The {112} surface, which has a built-in strong site inequivalence, is therefore preferable over the reconstructed {110} surface, which must rely on a tenuous surface reconstruction to achieve site selection. Our experimental experience [54] strongly confirms this expectation. We therefore consider our own former advocacy [55] of the reconstructed {110} surface as having been superseded by the subsequent realization of the inherently greater promise of the {112} orientation.

5.4. Small misorientations: nuisance or design parameter?

There is no such thing as a perfectly-oriented crystallographic interface. Any real interface will have deviations from perfect flatness and perfect orientation, as a result of which the $\langle 111 \rangle$ bonds are rotated out of the true hetero-interface plane by a small but non-zero angle θ. At apolar/nonpolar interface this will cause a finite built-in interface charge to appear, and even for small misorientations the resulting charge may be large by device standards. For the {112} interface, the charge density is easily shown to be

$$\sigma = \left(q\sqrt{3}/a^2\right) \sin\theta .$$

If the tilt angle is small enough, this charge is not likely to be removed by the HKWG atomic re-arrangement, but is likely to act as a permanent *tilt doping*. A wafer orientation to within $\pm 0.5°$ ($\cong$ 10 milliradian) is roughly the practical limit of current *routine* wafer orientation techniques. Assuming the lattice constant of GaAs, such a misorientation corresponds to an interface charge density of 4.7×10^{12} elementary charges per cm^2. This is a large charge, and much more accurate wafer orientation techniques than are in current use will be necessary. This is of course possible, but is a major nuisance. A highly (112)-selective etch would certainly help. However, one man's nuisance is often the next man's design parameter. *If* the orientation could be controlled to significantly better than 10^{-3} radian, a deliberate misorientation might become a practical means of introducing desirable interface charges into devices such as HEMT's. Because the interface charges would not be randomly distributed, but be located on quasiperiodic interface steps, they would scatter less, and even new superlattice effects might arise. Finally, by deliberately creating a controlled local variation in the interface tilt, one might even introduce lateral "doping" variations into device structures. It is a fitting notion on which to close a paper that addresses itself to the role of interfaces in submicron structures, more specifically, to the role of the interface nanostructure in determining the properties of devices containing those interfaces.

Acknowledgments

It is a pleasure to thank Dr. R.S. Bauer for inviting me to present this paper at this Symposium, and thereby providing the stimulus to order my thoughts on the topics discussed and to put them down on paper, something that otherwise would have been unlikely to occur. Many thanks are due to Drs. E.A. Kraut, J.R. Waldrop, R.W. Grant, D.L. Miller and S.P. Kowalczyk, for uncounted discussions. Last, but not least, I wish to acknowledge the profound influence that Professor W.A. Harrison has had on my thinking.

References

[1] See, for example, H. Kroemer, Japan. J. Appl. Phys. 20, Suppl. 20-1 (1981) 39.
[2] H. Kroemer, Proc. IEEE 70 (1982) 13.
[3] W. Shockley, US Patent 2,569,347, issued 25 Sept. 1951.
[4] R.L. Anderson, Solid-State Electron. 5 (1962) 341.
[5] For a general review, see A.G. Milnes and D.L. Feucht, Heterojunctions and Metal–Semiconductor Junctions (Academic Press, New York, 1972).
[6] An excellent recent review is contained in chs. 4 and 5 of H.C. Casey and M.B. Panish, Heterostructure Lasers (Academic Press, New York, 1978).
[7] H. Kroemer, Proc. IEEE 51 (1963) 1782.
[8] R. Dingle, in: Festkörperprobleme/Advances in Solid State Physics, Vol. 15, Ed. H.J. Queisser (Vieweg, Braunschweig, 1975) p. 21.
[9] R. Dingle, H.L. Störmer, A.C. Gossard and W. Wiegmann, Appl. Phys. Letters 33 (1978) 665.
[10] For a recent review, see T. Mimura, Surface Sci. 113 (1982) 454.
[11] For a review, see N. Holonyak, R.M. Kolbas, R.D. Dupuis, and D.D. Dapkus, IEEE J. Quantum Electron. 16 (1980) 170.
[12] H. Sakaki, L.l. Chang, R. Ludeke, C.-A. Chang, G.A. Sai-Halasz and L. Esaki, Appl. Phys. Letters 31 (1977) 211;
see also L.L. Chang and L. Esaki, Surface Sci. 98 (1980) 70.
[13] H. Kroemer, W.-Y. Chien, H.C. Casey and A.Y. Cho, Appl. Phys. Letters 33 (1978) 749.
[14] H. Kroemer, W.-Y. Chien, J.S. Harris, Jr. and D.D. Edwall, Appl. Phys. Letters 36 (1980) 295.
[15] Y.Z. Liu, R.J. Anderson, R.A. Milano and M.J. Cohen, Appl. Phys. Letters 40 (1982) 967.
[16] See, for example, J.R. Waldrop, S.P. Kowalczyk, R.W. Grant, E.A. Kraut and D.L. Miller, J. Vacuum Sci. Technol. 19 (1981) 573.
[17] G.F. Williams, F. Capasso and W.T. Tsang, IEEE Electron Devices Letters 3 (1982) 71;
see also F. Capasso, Surface Sci. 132 (1983) 527.
[18] H. Kroemer, Critical Rev. Solid State Sci. 5 (1975) 555.
[19] W.A. Harrison, J. Vacuum Sci. Technol. 14 (1977) 1016;
see also ref. [24] below.
[20] G. Margaritondo, A.D. Katnani, N.G. Stoffel, R.R. Daniel and T.-X. Zhao, Solid State Commun. 43 (1982) 163;
see also G. Margaritondo, Surface Sci. 132 (1983) 469.
[21] J.L. Shay, S. Wagner and J.C. Phillips, Appl. Phys. Letters 28 (1976) 31.
[22] J.C. Phillips, J. Vacuum Sci. Technol. 19 (1981) 545.
[23] W.R. Frensley and H. Kroemer, Phys. Rev. B16 (1977) 2642.
[24] W.A. Harrison, Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond (Freeman, San Francisco, 1980); see especially section 10F.

[25] J.O. McCaldin, T.C. McGill and C.A. Mead, Phys. Rev. Letters 36 (1976) 56. These authors expressed the correlation between valence band lineup and anion electronegativity for Schottky barriers; the approximate applicability of their result to heterojunctions appears to have been discussed first by W.R. Frensley and H. Kroemer, J. Vacuum Sci. Technol. 13 (1976) 810; see also ref. [23].

[26] For a very "physical" discussion of this theoretical foundation, see Harrison, ref. [24], especially chs. 1–3 and ch. 6.

[27] S.J. Anderson, F. Scholl and J.S. Harris, in: Proc. 6th Intern. Symp. on GaAs and Related Compounds, Edinburgh, 1976, Inst. Phys. Conf. Ser. 33b (Inst. Phys., London and Bristol, 1977) p. 346.

[28] The numerical values are based on Harrison's table 10-1 on p. 253 of ref. [24], except that we use the values from ref. [27] for the energy gaps of GaSb and AlSb.

[29] J.A. Van Vechten, Phys. Rev. 87 (1969) 1007. Van Vechten gives an extensive table of theoretical ionization energies, from which electron affinities are easily obtained by subtracting the energy gaps.

[30] This broken-gap lineup is, in fact, predicted by all three major predictive theories: The electron affinity rule, the Frensley–Kroemer theory, and the Harrison theory.

[31] G.C. Osbourn, J. Appl. Phys. Letters 53 (1982) 1536; J. Vacuum Sci. Technol. 21 (1982) 469; see also ref. [34] below.

[32] G.H. Döhler, Phys. Status Solidi (b) 52 (1972) 79,553;
G.H. Döhler, H. Künzel and K. Ploog, Phys. Rev. B25 (1982) 2365.

[33] Our calculation is to illustrate the basic idea only. The quoted composition falls into a solid solubility gap of uncertain width the existence of which has been reported. It may therefore be difficult or impossible to prepare. For a discussion and further references on this point see ch. 5 of ref. [6].

[34] P.L. Gourley and R.M. Biefeld, J. Vacuum Sci. Technol. 21 (1982) 473;
G.C. Osbourn, R.M. Biefeld and P.L. Gourley, Appl. Phys. Letters 41 (1982) 172.

[35] M.E. Davis, G. Zeidenbergs and R.L. Anderson, Phys. Status Solidi 34 (1969) 385.

[36] G.M. Guichar, C.A. Sébenne and C.D. Thuault, Surface Sci. 86 (1979) 789.

[37] R.S. Bauer and H.W. Sang, Jr., Surface Sci. 132 (1983) 479.

[38] H. Morkoc, L.C. Witkowski, T.J. Drummond, C.M. Stanchak, A.Y. Cho and J.E. Greene, Electron. Letters 17 (1981) 126;
see also H.L. Störmer, Surface Sci. 132 (1983) 519.

[39] It has been suggested by W.I. Wang (personal communication) that the ⟨110⟩ sequence dependence might be related to an as yet unexplained instability of ⟨110⟩-oriented (Al, Ga)As growth observed by him. For another report of a different kind of ⟨110⟩ growth instability see P. Petroff, A.Y. Cho, F.K. Reinhart, A.C. Gossard and W. Wiegmann, Phys. Rev. Letters 48 (1982) 190.

[40] R.C. Miller, W.T. Tsang and O. Munteanu, Appl. Phys. Letters 41 (1982) 374.

[41] E.A. Kraut, R.W. Grant, J.R. Waldrop and S.P. Kowalczyk, Phys. Rev. Letters 44 (1980) 1620.

[42] H. Kroemer and W.-Y. Chien, Solid-State Electron. 24 (1981) 655.

[43] H.K. Gummel and D.L. Scharfetter, J. Appl. Phys. 38 (1967) 2148;
see also C. Kittel and H. Kroemer, Thermal Physics, 2nd ed. (Freeman, San Francisco, 1980) ch. 13. For very unsymmetrically doped junctions, the GS correction is between 1 kT/q and 2 kT/q.

[44] H. Kroemer, RCA Rev. 17 (1956) 515.

[45] M. Weinstein, R.O. Bell and A.A. Menna, J. Electrochem. Soc. 111 (1964) 674.

[46] Zh.I. Alferov, V.I. Korolkov and M.K. Trukan, Soviet Phys.-Solid State 8 (1967) 2813.

[47] G. Zeidenbergs and R.L. Anderson, Solid-State Electron. 10 (1967) 113.

[48] N.N. Gerasimenko, L.V. Lezheiko, E.V. Lyubopytova, L.V. Sharanova, A.Ya. Shik and V. Shmartsev, Soviet Phys.-Semicond. 15 (1981) 626.

[49] S.L. Wright, PhD Thesis, University of California, Santa Barbara, CA (1982).
[50] See, for example, C.M. Garner, C.Y. Su, Y.D. Shen, C.S. Lee, G.L. Pearson, W.E. Spicer, D.D. Edwall, D. Miller and J.S. Harris, Jr., J. Appl. Phys. 50 (1979) 3383; see also the references quoted there.
[51] W.A. Harrison, E.A. Kraut, J.R. Waldrop and R.W. Grant, Phys. Rev. B18 (1978) 4402.
[52] W.A. Harrison, in: Festkörperprobleme/Advances in Solid State Physics, Vol. 17, Ed. H.J. Queisser (Vieweg, Braunschweig, 1977) p. 135.
[53] W.A. Harrison, J. Vacuum Sci. Technol. 46 (1979) 1492.
[54] S.L. Wright, M. Inada and H. Kroemer, J. Vacuum Sci. Technol. 21 (1982) 534.
[55] H. Kroemer, K.J. Polasko and S.L. Wright, Appl. Phys. Letters 36 (1980) 763.
[56] R. Kaplan, Surface Sci. 116 (1982) 104.
[57] W.A. Harrison, Surface Sci. 55 (1976) 1.