FROM Si TO ALLOY COMPOUND SEMICONDUCTORS: THE LIGHT EMITTING DIODE AND LAMP

Nick Holonyak, Jr.

After the discovery of the transistor by Bardeen and Brattain (December 16, 1947) and Bardeen's identification of minority carrier injection, it became essentially inevitable that the semiconductor would be studied intensely and would serve as the basis for an entirely new form of electronics, in fact, the wondrous electronics, including integrated circuits (ICs) and computers, now all around us. At first, in his modesty, Bardeen felt that early transistors would, because of their high cost, be useful only in certain special applications but might not be competitive in broad applications with the low cost vacuum tube. We are now able to smile at this honest early opinion. Bardeen also commented not long before his death (to Holonyak) that if there was a mistake made in his old Urbana semiconductor laboratory, which was founded Sept, 1952, it was not to emphasize crystal and materials studies more. This is ironic considering the fact that he knew, and wrote in his notebook at an earlier date (Feb, 1952), that Si was more important than Ge and that impurity diffusion from a single surface, a single reference plane, was the "right" way to make transistors; indeed, impurity diffusion was studied in his laboratory, as well as were surfaces and interfaces, including quantum size effects in inversion layers. It is worth mentioning that Bardeen held the basic patent (filed before the transistor patent) on the use of inversion layers in field effect devices, which gives some hint as to why he did not miss in the identification of carrier injection and later quantum size effects in surface channels.

Because of the primitive state of crystal development, it was inevitable that early transistor studies would be based on the elemental materials Ge and Si, and thus that material studies on these indirect-gap (long carrier life-time) crystals would come first. In the case of Si, these studies, as we know, continue even now. For example, we have known of the oxide on Si and its use in transistor devices for 40 years, and yet these studies continue and even expand, ever and ever toward smaller and smaller and more densely packed devices, perhaps eventually reaching down to the level of quantum size effects. We understand how important this is, and, of course, this is a source of great pride to all of us who have worked at one time or another on oxide-defined Si devices.

In the crystal and materials studies dealing with transistors, it was shown at Sony (Esaki, 1958) that at degenerate dopings an abrupt p-n junction tunnels and exhibits negative resistance. The limitations of tunnel diodes in Ge and Si immediately made the III-V semiconductors and their materials study interesting, which, indeed, led directly to III-V vapor phase epitaxy (VPE) and, for higher voltage p-n junctions, to VPE $GaAs_{1-x}P_x$ synthesis as early as 1960 -- at the time, in fact, when epitaxial Ge and Si transistors were first being devised.

The identification of injection, a fundamental idea, made it possible to understand at last why a semiconductor driven with a current (in the beginning SiC) could emit light, which was an old and poorly understood phenomenon. The turn towards III-V materials (1959-60) and the study of p-n junctions (tunnel junctions, varactors, etc.) in these systems led also in the direction of the study of III-V light emitters.

Ignoring its indirect-gap band structure and poor efficiency, one school of thought in the late 50's and early 60's focused on GaP because of its high energy gap (2.26 eV, green). Others were more concerned with GaAs in spite of its lesser energy gap (1.4 eV, infrared) because of its prospects also as a transistor material.

When a Zn-diffused GaAs varactor junction was used successfully in optical signal transmission (Rediker and co-workers, 1962), some attendees of the Device Research Conference (July, 1962) interpreted this to mean a direct-gap III-V crystal, say GaAs (or even $GaAs_{1-X}P_X$!), could be the basis for a laser. By the end of 1962 four groups succeeded in demonstrating p-n junction semiconductor lasers (Hall, et al., GaAs; Nathan, et al., GaAs; Holonyak, $GaAs_{1-X}P_X$; Quist, et al., GaAs). The semiconductor laser is, of course, important in its own right, but beyond that it made clear at once how important a direct-gap semiconductor is for light emission, including visible spectrum and not just infrared. In contrast, indirect-gap crystal (GaP) and its weaker band-to-band matrix element proved to be more or less useless as a laser material, which indicated very early its basic weakness also as a light emitter. Simply put, a direct-gap crystal is fundamentally a better light emitter material.

Whether it was appreciated or not, the demonstration of a visible-spectrum $GaAs_{1-X}P_X$ laser in 1962, as well as first practical light emitting diode (LED), signaled the future direction of laser and LED development. First of all, the visible-spectrum $GaAs_{1-X}P_X$ laser proved that III-V alloys were not inherently disturbed systems filled with defects, which was a mistaken belief of many early workers. Also, the tunability of composition(x) and energy gap of an alloy provided the necessary flexibility to make heterojunctions. Thus, as early as 1962 it could be concluded that direct-gap alloys such as $GaAs_{1-X}P_X$ were more important and indirect-gap binary materials such as GaP were less important, except perhaps as substrates.

Now we deal with a basic question: a p-n junction emits light, but how fundamental is it as a light emitter? Is it an ultimate lamp? To answer this question, we invert the problem and shine light (excitation light, $hw \gtrsim Eg$) on the center of a small rectangular slab of undoped (intrinsic) direct-gap semiconductor and drive the electron chemical potential ($E_{Fn} \equiv -q\phi_n$, the electron quasi-Fermi level) towards the conduction band edge, $E_{Fn} \rightarrow E_c$, and the hole chemical potential ($E_{Fp} \equiv -q\phi_p$, the hole quasi-Fermi level) towards the valence band edge, $E_{Fp} \rightarrow E_V$ ($E_c - E_V = E_g$). We note that charge neutrality is preserved, and the energy bands remain flat (straight). In the ideal case we can recover, with 100% quantum efficiency, all the input photons as electron-hole recombination radiation. At the moment we ignore the problem of removing the recombination-radiation from the high-index ($n \sim 3.5$) semiconductor.

We have an alternative: Since we have photoexcited the sample and generated electron-hole pairs, maybe we can extract the excess electrons and holes as a current and not recombination radiation. This is possible (excitation light still in place) if we make an electron and a hole "Fermi-level conduction-connection" to the electron-hole pairs by doping the right half of the semiconductor n-type and the left half p-type. We do this, let us say, by some form of ideal donor or acceptor implantation (neutral impurity) with the light still ON. Instead of extracting the electron-hole pairs as a current at voltage $V \sim E_g/q$, let us place a "bucking" voltage, an opposing voltage, $V \approx E_g/q$ across the external n and p conduction connections we have just made. No current flows with the excitation light in place. We turn-off the light, and thus turn-off the light-generated internal "solar-cell" voltage. Now the external "bucking" voltage, a battery, becomes a source and supplies the electron-hole pairs by current flow via the "Fermi-level conduction connections", which, indeed, changed our quasi-two-level (E_C, E_V) direct-gap quantum system into a p-n junction.

The point of this exercise is to show that to make an ideal photoexcited two-level quantum system, i.e., a simple photopumped flat-band direct-gap semiconductor light source, into an otherwise-identical current-driven light source, we have been forced quite naturally into constructing a p-n junction. And, of course, with a current input there is no further need for photopumping. A p-n junction is, in fact, a fundamental form of light source, essentially a two-state quantum system driven directly with an electron-hole injection current (not by heat as in an incandescent lamp). Of course, an ordinary p-n junction is too simple of a light source as we have described it. It should really take the form of a double heterojunction so that the n-type and p-type input current sections ("the Fermi-level conduction connections") are wider in energy gap than the narrower gap thin active region in order to reduce the absorption. Also this makes the electron-hole pair injection into the thin active region ($L_Z<1/\alpha$, α the absorption coefficient, $\alpha \sim 10^4 cm^{-1}$) more efficient, as well as confines the carriers and to some extent the recombination radiation (as in a laser).

There remains the practical problem of how to realize a modern LED (or laser). Although $GaAs_{1-X}P_X$(the prototype) served its purpose in showing the importance of direct-gap alloys, and that alloys make possible heterojunctions, the successor alloys $Al_XGa_{1-X}As$ and $In_{0.5}Ga_{0.5}P$ modified into $In_{0.5}(Al_XGa_{1-X})_{0.5}P$ proved to be more important for heterojunctions because Al-Ga substitution allows change in the energy gap without change of the lattice constant. This is important in reducing defects and thus shunt losses. These alloy systems match the common substrate material GaAs, and of further importance can be grown economically at large scale in aluminum-capable Dupuis-style metalorganic-chemical-vapor-deposition (MOCVD) reactors.

In the interest of brevity, we consider further only thin red-orange-yellow-green (ROYG) $In_{0.5}(Al_XGa_{1-X})_{0.5}P$ heterolayers grown epitaxially (MOCVD) on relatively thick GaAs substrates. Because of its lesser energy gap, however, the GaAs substrate turns out, unfortunately, to be highly absorbing of ROYG recombination radiation. The semiconductor does not just generate the recombination radiation; its high index of refraction ($n \sim 3.5$) relative to the outside world (its transmission mismatch) causes it to act as a "box", a container, that traps the photons and annihilates them with any absorbing substance in the "box", e.g., the thick GaAs substrate. The thick GaAs substrate on which the thin $In_{0.5}(Al_XGa_{1-X})_{0.5}P$

heterolayers are grown must be removed if we are to realize a truly efficient LED. Recently Kish and Craford and their co-workers (Hewlett-Packard, 1994) have shown that the GaAs substrate can be removed and be replaced, by various "tricks" of wafer bonding, with non-absorbing GaP of sufficient thickness to make LED assembly convenient. The use of non-absorbing GaP platform makes possible as much as a 200% increase in light output so that now ROY LED's exceed, in lumens per watt, incandescent lamps. Parallel with these developments, the same III-V vapor epitaxial crystal growth technology (MOCVD) has made possible the construction of blue-emitting direct-gap III-V $In(Al_xGa_{1-x})N$ heterostructure LEDs (Nakamura, et al. at. at Nichia, 1994) of at present lesser performance but nevertheless considerable potential for improvement. Thus, III-V alloys are able to serve the entire visible spectrum (maybe even receiving help from II-VI alloys).

These are remarkable developments: After over 30 years the direct-gap III-V alloy has prevailed as an LED material and, in addition, has caused the LED, because of its better performance than an incandescent, to become a lamp. The lighting industry, whether it is appreciated or not, has been put under long range threat. It will take some time for the LED to be fully developed in all its possible display uses, and at equal performance across the entire visible spectrum. Perhaps even more time will be required for its conversion into a full-fledged lamp. Nevertheless, it will happen, as, indeed, is beginning to occur already. It is interesting how long it has taken (many decades!) to establish conclusively the special importance of direct-gap III-V alloys for LEDs, and how much time in general is required to convert scientific advances into important technological advances. It is interesting also how, in a related sense, the semiconductor, with its electron (-) and hole (+) conduction capability, as well as its ability to generate, detect and handle photons, becomes more and more the universal material of electronics, and now even begins to intrude into illumination and lighting. To answer again the question posed earlier: the p-n heterojunction is an ultimate lamp.

Finally, it is clear that Bardeen's identification of carrier injection marked the beginning of a revolution in electronics that still continues, and that is bound to continue much, much longer. It is, indeed, proper to consider John Bardeen the "godfather" of modern electronics and, as long as electronics exists, to hold him in the greatest esteem. In truth, there is no substitute for electronics and what it does, and we will always be inspired by and be in John Bardeen's debt.