FREQUENTLY ASKED QUESTIONS

February 21, 2017

Content Questions

How do you place a single arsenic atom with the ratio 1 in 100 million? Sounds difficult to get evenly spread throughout.

Yes, techniques for doping of materials with well-defined concentrations is the subject of a whole subfield of engineering... there are various techniques. A common method is to use a vapor of the dopant, which diffuses through a material. One can also use beams of atoms to implant them as dopants into a surface. Here's a Wikipedia article.

Can any element with four valence electrons be used in solid state? For example, carbon.

The key property is the small band gap. (There are also practical considerations, such as physical properties determining how easy it is to actually make devices out of the substance, cost, etc.). Carbon does work, in diamond form (not cheap!) and tin also works. Useful semiconductors can also be made out of compounds, too. Here's a table of semiconductor materials.

Qualitatively, what causes the difference in band gap in conductors vs insulators vs semiconductors?

Quantitively the band gap width is determined from quantum mechanics, i.e., solving for the wavefunctions describing the system of multiple-electron atoms in the crystal. Qualitatively, you can think about the atoms in the crystal as having electrons residing in "shells" or "orbitals" (these come from solving the Schrödinger equation – I hope this picture is familiar to you from previous courses). Each shell with principal quantum number n can hold a maximum of $2n^2$ electrons in subshells with different l values ($l = 0, ...n-1$). According to the Pauli principle, you can have up to two electrons, with opposite spins, in each state. When a shell is fully occupied by electrons, it's "closed"; further electrons at higher energies added to the atom are called "valence" electrons. You can think of a solid as a bunch of positive centers (nuclei) with closed shells of electrons, plus valence electrons. Accord-

ing to quantum mechanics, states that the valence electrons can occupy are determined from the solution to the Schrödinger equation, for which the Hamiltonian includes potentials due to interactions between atoms. These interactions will tend to modify the allowed energy levels such that valence electrons can occupy a band of energies, the so-called "valence band". There will also exist allowed, unoccupied-in-the-ground-state energy levels. Electrons in excited states occupying these could be delocalized i.e., could be found quite far from their "home" nuclei. Such states make up the "conduction band". Where the valence and conduction bands lie in energy depends on the specific nature of the atoms involved. The difference in energy between the conduction and valence bands is known as the "band gap".

Depending on the specific structure of the atoms in the crystal (note you can have compounds as well as elements), the conduction band can overlap with the valence band. In this case, an electron can easily be promoted from the valence to the conduction band with only a bit of extra energy (even just thermal energy) and go wandering around, easily influenced by external electric fields. Such materials are good conductors. If there's a big difference between the valence and conduction bands, then it takes a lot of energy to kick an electron up to the conduction band, and that material will be an insulator. A semiconductor corresponds to the case where there's a non-zero band gap, but a small one.

Why do different materials have different band gaps? It depends on their specific atomic structure. For example, if a material has a partially-filled shell, the next-state-up might be not very far away (in the same shell), and so the conduction band will be very close to the valence band– perhaps even so close that inter-atomic interactions will create energy levels easily accessible in the ground state– and you'll get a conductor. As another example, if there are *no* valence electrons (such as for noble elements like He, Ne, Ar, etc.), and just closed shells, the next highest energy state an electron can occupy could be quite far up (all the way to the next shell), so you will have large band gap and insulating properties. And another example: Si and Ge have subshells filled, but not full shells; the conduction band is nearby, but not right on top of the valence band– and you have a semiconductor. (BTW, note that you can have compounds as well as elements, with complicated electronic band structure.)

What was the logic behind the number of electrons drawn in the lattice diagram?

Each atom in the 2D lattice representation has 4 valence electrons. Each atom has also 4 perpendicular neighbors. If you draw 4 electrons around each atom's nucleus, you'll get 2 shared in the space between each atom and its perpendicular neighbor.

So you add atoms to a lattice to improve performance, but why does filling that electron spot improve performance? Does it fill the gap?

Well, here by "performance" I meant "conductivity"... when you add dopant atoms, you add extra electrons or holes at some sites. Yes, doing this does provide extra (localized) states that electrons can occupy in the gap between the valence and conduction bands (although you wouldn't say the gap is filled).

For a donor impurity (making an n-type semiconductor), electrons in the new states can easily jump to the conduction band. For an acceptor impurity (making a p-type semiconductor), electrons in the valence band can jump to the new state, leaving a hole that's free to travel.

See the diagrams in Eggleston 3.1.3.

What's the difference between majority and minority carriers?

For doped semiconductors, in particular in the p-n junction case, "majority charge carriers" are the relatively mobile electrons and holes associated with the dopants. These provide most of the conductivity of the doped semiconductor. The "minority charge carriers" are the less-mobile electrons and holes associated with the pure semiconductor: once in a while, due to thermal fluctuations, an electron can leave its valence band and inhabit a state in the conduction band. There are typically many fewer of these in a doped semiconductor, so minority current flow tends to be small.

Why does the voltage decrease across the depletion zone?

As thermal fluctuations cause the dopant-associated majority charge carriers to diffuse across a p-n junction, one side (the p-type side) becomes electrically negative and the other side (the n-type side) becomes electrically positive:

two sheets of opposite charge create an electric field. Associated with an electric field is a voltage drop, where the higher potential is on the positive charge side (which is the n-type side into which positive holes from the p-type side have burrowed). The electron energy is higher on the p-type side.

Is the ΔE due to the potential difference in the depletion zone? In an unbiased condition, why is equilibrium not established before a depletion zone can form?

Yes, for an unbiased p-n junction, the ΔE is due to the potential difference in the depletion zone. When p-type and n-type are stuck next to each other there is a thermally-driven diffusion of electrons from n to p type, and of holes from p type to n-type near the boundary, which establishes a separation of charge (excess negative charge on the p side and excess positive charge on the n side). The diffusion happens until enough charge has built up to create an electric field that prevents *further* charge buildup – that's when equilibrium is established, and flow is balanced in each direction. At equilibrium, there is non-zero charge separation and a depletion zone in the region near the boundary, from which the charge carriers have drifted. This configuration, with a potential difference and electron energy increase on the p-side, has required energy input to the system. This energy comes from thermal energy in the environment for an unbiased junction— in this configuration a non-zero diffusion results in a ΔE. (A battery provides energy for a biased junction.)

Why are large depletion regions poor conductors, and vice versa?

In a depletion region after negative (majority) charge has diffused from n to p, there are not many charge carriers left (electrons have left and holes have been filled), so the material in the depletion region is a poor conductor.

(Another way of thinking about it: remember from E&M that a perfect conductor must have zero electric field inside it, as charges will rearrange themselves to cancel any field. After diffusion, the n-type side has an excess of positive charge, and the p-type side has an excess of negative charge. There is then an electric field across the region from positive to negative, pointing from the n-type to the p-type; only a poor conductor can support an electric field.)

Outside of the depletion region, there are still charge carriers around to conduct current.

Why does the depletion zone get larger in reverse bias and vice versa?

When the external voltage is applied, the depletion region gets thinner or fatter depending on the external voltage's magnitude and polarity.

When the p-n junction is unbiased, the p-type side is negative and the n-type side is positive, from thermal diffusion of carriers.

In the reverse-bias case, the n-type side is made more positive by the battery. The battery is doing work, shoving charges around to force the n-type side to be more positive and the p-type side to be more negative. The applied voltage forces negative charge carriers to the negative side and positive charge carriers to the positive side, so there's even more depletion in the charge-separated region, and conduction gets even harder in that region. The bigger the externally applied voltage, the wider the region with an electric field across it.

In the forward-bias case (negative terminal of battery attached to n-type side, positive side of battery attached to p-type side), now the external voltage is working *against* the diffusion-created p-n voltage. The n-type side, which was positive from diffusion, becomes now less positive; the p-type side, which was negative from diffusion, becomes less negative. The electric field in the depletion region becomes smaller, and the potential difference decreases (over a smaller depletion region). If the forward-bias potential difference increases yet more, eventually the depletion region disappears altogether, so there's *no* poorly-conducting region. The p-n junction effectively becomes a resistor, with a potential drop across it determined by the properties of the material. The majority (dopant) carriers can flow through the junction.

How are p-n junctions used?

Various devices– notably diodes and transistors – have p-n junctions inside of them. Transistors, in particular, are cool, because they allow you to control energy flow in a circuit using a voltage. Circuits built of many transistors can perform very complicated (and useful) operations. We'll see more in the next lectures!