## Many-particle Systems, 10

#### Intrinsic semiconductors

**Intrinsic** semiconductors have negligible concentrations of impurity atoms. Their electrical conductivity arises primarily from electrons excited into the otherwise empty conduction band from the otherwise filled valence band-usually by absorbing sufficient energy from phonons at finite temperature. Exciting an electron into the conduction band leaves a vacant state in the valence band. An electron at lower energy in the valence band can fill this vacant state. That, in turn, makes available a possible state for vet another valence band electron to fill. In other words, the excitation of the electron provides a mobile charge in the conduction band as well as a mobile unoccupied state in the valence band. When an external electric field is turned on, the conduction band electron accelerates in the opposite direction while the valence band state accelerates in the same direction as the field. For this reason, the state is considered to be a mobile. effective positive charge. Such a positive, effective charge is called a "hole."

#### Doped semiconductors

In practical applications semiconducting material essentially always contains small amounts of intentionally placed impurity atoms. Such materials are **doped**. Si and Ge are the workhorses of solid-state electronics. Each has two s and two p electrons per atom, and at T = 0 K each has a full sp hybrid valence band and an empty sp hybrid conduction band (if sufficiently pure). Historically these elements are characterized as being in "Group IV" of the periodic table. When atoms from Group V with two s and three p electrons each-such as, P, As, and Sb-are introduced into Si or Ge atomic

impurity states appear in the band gap just below (a few meV-"m" is milli- not mega-) the lowest energy state in the conduction band ( $\tilde{CB}$ ). At T = 0 K (left in the figure) these impurity states are occupied. At higher temperatures (right), electrons are excited out of these states into the conduction band and the doped host material is a much better conductor than when pure. These doped states therefore "donate"



negative electrons and this kind of doping is called "n-type" (n for "negative").

On the other hand, if Group III atoms—B, AI, Ga, In—are employed as the dopant (each with two s and one p electrons), empty states (at T = 0 K (left)) close to the top of the valence band (VB) "accept" electrons from the valence band at finite temperature (right), leaving (effectively positive) holes in the valence band that can conduct. This kind of doping is called "p-type." Note that compounds of Groups III and V atoms-most notably GaAs, with a 1.4 eV band gap, and GaN, with a 3.4 eV band gap-can replace Si and Ge and can also be n- or p-type doped.



#### The pn junction diode

An n-doped semiconductor has a supply of mobile electrons in its conduction band while a p-doped semiconductor has a supply of mobile holes in its valence band. When a slab of n-type material is placed in good contact with a slab of p-type, electrons diffuse from the n-side conduction band (where their concentration is higher) to the p-

Mn10

side conduction band (where their concentration is lower), making the n-side positive and the p-side negative. Similarly, electrons from filled states in the valence band on the n-side diffuse into the empty states on the p-side which can be visualized as holes diffusing in the opposite direction; hole diffusion leads to the same kind of charge separation. When the electric forces on electrons and holes become sufficiently large the diffusive redistribution of charges stops (i.e., the concentration-difference driving force for diffusion becomes

balanced by the electric driving force for charge transport). This results in an energy band picture like that shown to the right. In the figure, the open circles depict holes and the "–"s depict electrons. The electronic band gap is raised on the p-side relative to the n-side because the negative charge accumulating on the p-side repels electrons (i.e., creates higher electronic potential energy) while the positive n-side attracts them (creates lower electronic potential energy). The transition region between the p- and nsides is partly in the p-material and partly in the n-material and is called the "junction."

This structure can serve an important electrical function when connected to an external source of emf. If the p-side is connected to a positive voltage and the n-side to a negative one, the potential energy of electrons on the p-side is lowered and raised on the n-side, thus "flattening" the potential energy hill electrons have to climb in going from n to p. This reduces the electric force suppression of electronic diffusion and results in an electrical current directed from p to n. This situation is called "forward bias." If the external voltage difference is reversed ("reverse bias") and isn't too large, the electric force is increased and electrons actually tend to migrate from p to n (against diffusion), producing a current from n to p. Because the concentration of electrons in the conduction band on the p-side is so much less than on the n-side the

magnitude of the reverse bias current is typically much smaller than the forward bias current (for the same magnitude of applied voltage difference). In a circuit such a structure is called a "pn junction diode" and its circuit diagram symbol is shown to the right (along with forward and reverse biased currents).

The diode passes current easily in one direction but not in the other. If an ac voltage is applied to the structure, the resulting current will (mostly) only flow in the forward direction, as depicted to the right. The "full-wave" input signal is "half-wave rectified" by the diode.

An interesting thing can happen if the reverse bias is too



large. This is shown to the right. At the top is an unbiased bandgap in equilibrium. Below shows what happens when the reverse bias becomes large. When the filled VB states on the p-side line up with unfilled CB states on the n-side, and when the junction distance is small enough, electrons can tunnel from p to n producing a large reverse current (from n to p). This













breakdown can produce sudden, very large currents—and therefore sudden, high heating rates—large reverse biases are generally avoided for most diode applications.

## LEDs

The band gap in a light emitting diode (LED) is in the visible range, 1.8 to 3.1 eV. When the diode is forward biased, electrons preferentially travel from the n to the p side. Occasionally, such an electron in the CB will recombine with a hole in the VB, emitting a visible photon in the process. If the current is sufficiently large, the light emitted will be bright. This bright light is not produced by heating a filament (as in an incandescent light) or ionizing a gas (as in fluorescent light), so is associated with much less dissipated energy and is consequently much more efficient in converting input electrical energy into light. The record LED efficiency to date is over 10 times that of incandescent lights and twice that of fluorescent lights. Because LEDs generate much less heat than conventional lights, they tend to last longer—perhaps 50 times longer than incandescents and 5 times longer than fluorescents. Their high efficiency and long lifetime makes LEDs promising for replacing conventional home and workplace light sources. (QM =\$)

The color emitted by an LED is determined by the semiconductor band gap. Red LEDs are often made of n- and p-doped AlGaAs (where some of the Ga atoms are replaced by Al); blue LEDs are typically InGaN. The fact that LEDs emit specific colors means that it is difficult to produce white light with them—which, of course, is essential for building lighting. White LEDs are usually blue (or UV) emitters coated with fluorescent material that simultaneously emits several colors to produce white. (That same technique is used in "fluorescent" lights. The 2014 Nobel Prize was awarded to the developers of the blue LED, primarily because it is the key to making white LEDs needed for practical lighting applications.) This "down conversion" of photon energy reduces the efficiency of the LED and the quality of the light produced isn't as pleasant (yet) as that produced by conventional sources. Another difficulty for some practical applications is that the emitting area of a typical LED is only a few mm<sup>2</sup>. To illuminate a large area requires an extended array.

LEDs can be run "backward." That is, if an LED absorbs photons the result is a current or a voltage. Thus, LEDs can be used as photodetectors. The most important application of this conversion of light into electricity is solar cells. (QM =\$)

## **MOSFETs**

Mn10



Certainly one of the most important applications of pn junctions is the metal-oxide-semiconductor field-effect-transistor (MOSFET). The structure of a MOSFET is shown to the left. This particular form is an "nMOS." The black inverted "T" is a metal plate. It is separated from the semiconducting material by a thin

insulating oxide layer. In equilibrium, the energy band picture shows the gap at higher energy in the p-region than in the two n-regions. If the right-hand n-side of the MOSFET is held at a higher potential than the left-hand side the result will be as shown to the right. Remember there are lots of conduction band electrons in the two n slabs. When the "gate" (p-region) voltage is set to a sufficiently high positive value the potential energy bump in the

3

middle goes down and electrons will flow from the left side to the right producing an electrical current in the other direction. Thus, the gate voltage acts as a switch: no current when it is below some threshold value, current when it is above.

A pMOS transistor is a pnp structure that is switched on (holes flow from one side to the other) when the n-region is set to a sufficiently negative voltage. All of the stuff on a contemporary computer chip consists of combinations of nMOS and pMOS transistors. Because both types are used (usually in pairs) the technology is called cMOS, where the "c" stands for "complementary." An example is the "inverter" or "NOT gate," shown to the right. The two \_\_\_\_\_\_ widgets are transistors; the one with the little bubble attached to its gate is pMOS. When a sufficiently positive voltage ("high") is applied to the "IN" port the nMOS is turned on (i.e., its resistance becomes small), while the pMOS is off (high resistance), and the voltage at the upper end of the nMOS becomes close to –V. Thus, the "OUT" voltage goes "low." Inversely, if IN is negative, the pMOS is turned on (and the nMOS is off) and the



voltage at its lower end becomes close to +V. Then OUT is high. The combination of transistors above is often compressed to the symbol to the right. Two inverters connected as shown to the left constitute a simple ("flip-flop") **memory cell**; a short high voltage pulse at A becomes

a low voltage pulse as it passes through the first inverter, then is transformed back to high by the second inverter and can be read out as such at B after many passes though the loop.

High and low can be thought of as bits 1 and 0 (or equivalently "true" and "false"). Thus, one inverter transforms 1 into 0 (true into false) and vice versa. In logic operator language an inverter is a **NOT gate**.

cMOS transistors can be connected in various ways to execute any logic operation. For example, the circuit to the right executes the logic operator **AND**. (1 AND 1 = 1, 1 AND 0 = 0, 0 AND 1 = 0, 0 AND 0 = 0; can you see how the circuit produces this?)



OUT



The circuit to the left executes the logic operator **OR**. (1 OR 1 = 1, 1 OR 0 = 1, 0 OR 1 = 1, 0 OR 0 = 0; can you see why?)

Transistors are the basic building blocks of computer hardware (and responsible for a huge amount of economic commerce). And **they only work because of quantum mechanics**.

# As suggested throughout these notes, quantum mechanics = money!