Semiconductors

It is often said that semiconductors started the computer revolution. And that the transistor started the electronic era. And those statements are sorta true, but what exactly are semiconductors and transistors.

Semiconductors don't sit somewhere between an insulator and a conductor in terms of their ability to conduct electricity. Instead they can be subtly modified to take on either characteristic in particular ways.

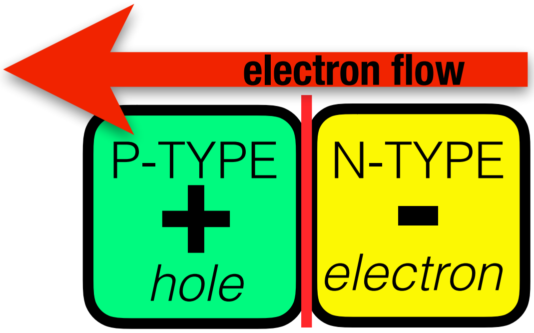
You've probably also heard the word silicon used almost synonymously with electronics, and that's because silicon is the most commonly used material for creating semiconductors, though other materials can be used.

Silicon can easily be formed into a crystal which is normally non-conductive. Its atomic structure is such that electrons cannot move freely through it. However, if you mix in a small amount of phosphorus or arsenic when forming the crystal, a process called 'doping', the crystalline structure changes. This new structure results in excess electrons and thus a propensity to have them flow out. This is called an N-Type semiconductor, since electrons have a negative charge.

If you use boron or gallium instead, the structure will have gaps, or 'holes', where electrons should be. Any free electron can readily fill these holes; a positive charge. So this is called a P-Type semiconductor.

So doping allows for the conversion of a good insulator into a reasonably good conductor. Nothing really special about that per-se, as there are much better conductors and insulators available — but when you combine the two types of semiconductors, interesting things happen.

We've already used a semiconductor in our experiments — the LED. Diodes are one of the simplest forms of useful semiconductor devices (though LEDs are a bit more complicated, in terms of the light generation aspect). They simply contain one N-Type piece and one P-Type piece, joined together in the middle. It is at this junction that the interesting things happen.

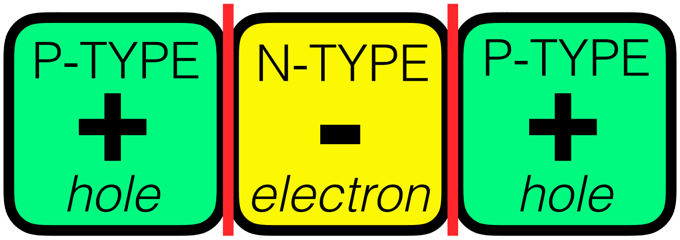


The interaction there is somewhat complex, but a simplified description is that electrons from the N-Type fill the holes in the P-Type and an equilibrium is reached at the junction. If a battery were connected such that its negative side linked to the N-Type and the positive to the P-type, electrons will be removed from the P-Type by the battery's positive terminal, causing more holes. Those holes can be filled from the N-Type as it has a supply incoming from the battery's negative terminal. Hence current flows.

If you swap the connections, the battery's negative side pulls electrons from the N-Type, so it has none to fill the holes in the P-Type. No current flows.

That's great, and useful, but the real breakthrough in computers was using semiconductors to make a transistor.

A transistor consists of three semiconductor sections. Each with a wire leg connected to it, and all three sandwiched together so that there are two junctions. Since there are two types, there can be two type of transistor, NPN or PNP.



Since this sorta looks like two diodes jammed together, it might seem like no current could ever flow through it. But because of the interaction at the junction, applying a current to the center section will cause current to be able to flow across both junctions from one end to the other. (From N to N in the NPN type or P to P in the PNP type)

This means the transistor can act like a switch. Computers work with 0s and 1s, ons and offs, as you'll learn more about soon. Early computers used physical switches, then later vacuum tube 'valves'. These were all still big and slow. But the transistor is small and fast.

When coupled with advances in miniaturization and manufacturing, electronics got smaller and cheaper and computers went from a few 1,000 switches in a room, to 10,000 in less than 1" square in just a few years time. Today's top-end Xeon CPU packs over 7.2 Billion transistors, and Altera offers a FPGA chip with over 30 Billion!

— — This page intentionally blank