Semiconductors, diodes, transistors
(Horst Wahl, QuarkNet presentation, June 2001)

- Electrical conductivity
  - Energy bands in solids
  - Band structure and conductivity
- Semiconductors
  - Intrinsic semiconductors
  - Doped semiconductors
    - n-type materials
    - p-type materials
- Diodes and transistors
  - p-n junction
  - depletion region
  - forward biased p-n junction
  - reverse biased p-n junction
  - diode
  - bipolar transistor
  - operation of bipolar pnp transistor
  - FET

ENERGY BANDS IN SOLIDS:

- In solid materials, electron energy levels form bands of allowed energies, separated by forbidden bands
- Valence band = outermost (highest) band filled with electrons ("filled" = all states occupied)
- Conduction band = next highest band to valence band (empty or partly filled)
- "gap" = energy difference between valence and conduction bands, = width of the forbidden band
- Note:
  - Electrons in a completely filled band cannot move, since all states occupied (Pauli principle); only way to move would be to "jump" into next higher band - needs energy;
  - Electrons in partly filled band can move, since there are free states to move to.
- Classification of solids into three types, according to their band structure:
  - Insulators: gap = forbidden region between highest filled band (valence band) and lowest empty or partly filled band (conduction band) is very wide, about 3 to 6 eV;
  - Semiconductors: gap is small - about 0.1 to 1 eV;
  - Conductors: valence band only partially filled, or (if it is filled), the next allowed empty band overlaps with it

ELECTRICAL CONDUCTIVITY

- In order of conductivity: superconductors, conductors, semiconductors, insulators
- Conductors: material capable of carrying electric current, i.e. material which has "mobile charge carriers" (e.g. electrons, ions,...), e.g. metals, liquids with ions (water, molten ionic compounds), plasma
- Insulators: materials with no or very few free charge carriers; e.g. quartz, most covalent and ionic solids, plastics
- Semiconductors: materials with conductivity between that of conductors and insulators; e.g. germanium Ge, silicon Si, GaAs, GaP, InP
- Superconductors: certain materials have zero resistivity at very low temperature.

Some representative resistivities ($\rho$):

- $R = \rho L/A$, $R =$ resistance, $L =$ length, $A =$ cross section area; resistivity at $20^\circ$C

<table>
<thead>
<tr>
<th>Material</th>
<th>$\rho$ (ohm.cm)</th>
<th>$\rho$ (ohm. m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum</td>
<td>$2.8 \times 10^{-8}$</td>
<td>$3.6 \times 10^{-2}$</td>
</tr>
<tr>
<td>Brass</td>
<td>$8 \times 10^{-8}$</td>
<td>$10 \times 10^{-2}$</td>
</tr>
<tr>
<td>Copper</td>
<td>$1.7 \times 10^{-8}$</td>
<td>$2.2 \times 10^{-2}$</td>
</tr>
<tr>
<td>Platinum</td>
<td>$10 \times 10^{-8}$</td>
<td>$12 \times 10^{-2}$</td>
</tr>
<tr>
<td>Silver</td>
<td>$1 \times 10^{-8}$</td>
<td>$2.1 \times 10^{-2}$</td>
</tr>
<tr>
<td>Carbon</td>
<td>$3 \times 10^{-8}$</td>
<td>$44.5$</td>
</tr>
<tr>
<td>Germanium</td>
<td>$0.45$</td>
<td>$5.7 \times 10^{-9}$</td>
</tr>
<tr>
<td>Silicon</td>
<td>$6 \times 10^{-4}$</td>
<td>$6 \times 10^{10}$</td>
</tr>
<tr>
<td>Porcelain</td>
<td>$10^{-5} - 10^{-7}$</td>
<td>$10^{6} - 10^{8}$</td>
</tr>
<tr>
<td>Teflon</td>
<td>$10^{-5}$</td>
<td>$10^{5}$</td>
</tr>
<tr>
<td>Blood</td>
<td>$1.5 \times 10^{5}$</td>
<td>$1.9 \times 10^{5}$</td>
</tr>
<tr>
<td>Fat</td>
<td>$2.4 \times 10^{7}$</td>
<td>$3 \times 10^{7}$</td>
</tr>
</tbody>
</table>
**INTRINSIC SEMICONDUCTORS**

- semiconductor = material for which gap between valence band and conduction band is small:
  (gap width in Si is 1.1 eV, in Ge 0.7 eV).
- at \( T = 0 \), there are no electrons in the conduction band, and the semiconductor does not conduct (lack of free charge carriers);
- at \( T > 0 \), some fraction of electrons have sufficient thermal kinetic energy to overcome the gap and jump to the conduction band; fraction rises with temperature;
  e.g. at 20° C (293 K), Si has 0.9x10^10 conduction electrons per cubic centimeter; at 50° C (323 K) there are 7.4x10^10;
- electrons moving to conduction band leave "hole" (covalent bond with missing electron) behind;
  under influence of applied electric field, neighboring electrons can jump into the hole, thus creating a new hole, etc. ⇒ holes can move under the influence of an applied electric field, just like electrons;
  both contribute to conduction.
- in pure Si and Ge, there are equally many holes ("p-type charge carriers") as there are conduction electrons ("n-type charge carriers");
- pure semiconductors also called "intrinsic semiconductors".

**DOPED SEMICONDUCTORS:**

- "doped semiconductor" (also "impure", "extrinsic") = semiconductor with small admixture of trivalent or pentavalent atoms:
  - **n-type material**
    - donor (n-type) impurities:
      - dopant with 5 valence electrons (e.g. P, As, Sb)
      - 4 electrons used for covalent bonds with surrounding Si atoms, one electron "left over";
      - left over electron is only loosely bound⇒ only small amount of energy needed to lift it into conduction band (0.05 eV in Si)
      - ⇒ "n-type semiconductor", has conduction electrons, no holes (apart from the few intrinsic holes)
      - example: doping fraction of 10^-8 Sb in Si yields about 5x10^16 conduction electrons per cubic centimeter at room temperature, i.e. gain of 5x10^6 over intrinsic Si.
  - **p-type material**
    - acceptor (p-type) impurities:
      - dopant with 3 valence electrons (e.g. B, Al, Ga, In) ⇒ only 3 of the 4 covalent bonds filled ⇒ vacancy in the fourth covalent bond ⇒ hole
      - "p-type semiconductor", has mobile holes, very few mobile electrons (only the intrinsic ones).
      - advantages of doped semiconductors:
        - can"tune" conductivity by choice of doping fraction
        - can choose "majority carrier" (electron or hole)
        - can vary doping fraction and/or majority carrier within piece of semiconductor
        - can make "p-n junctions" (diodes) and "transistors"
**DIODES AND TRANSISTORS**

- **p-n JUNCTION:**
  - p-n junction = semiconductor in which impurity changed abruptly from p-type to n-type;
  - “diffusion” = movement due to difference in concentration, from higher to lower concentration;
  - in absence of electric field across the junction, holes "diffuse" towards and across boundary into n-type and capture electrons;
  - electrons diffuse across boundary, fall into holes ("recombination of majority carriers");
  - ⇒ formation of a “depletion region” (n region without free charge carriers) around the boundary;
  - charged ions are left behind (cann’t move);
    - negative ions left on p-side ⇒ net negative charge on p-side of the junction;
    - positive ions left on n-side ⇒ net positive charge on n-side of the junction;
  - ⇒ electric field across junction which prevents further diffusion

- **Formation of depletion region in pn-junction:**

- **DIODE**
  - diode = "biased p-n junction", i.e. p-n junction with voltage applied across it
  - “forward biased”: p-side more positive than n-side;
  - "reverse biased": n-side more positive than p-side;
  - forward biased diode:
    - the direction of the electric field is from p-side towards n-side
    - ⇒ p-type charge carriers (positive holes) in p-side are pushed towards and across the p-n boundary
    - n-type carriers (negative electrons) in n-side are pushed towards and across n-p boundary
  - ⇒ current flows across p-n boundary

- **Forward biased pn-junction**
  - Depletion region and potential barrier reduced
Reverse biased diode

- Reverse biased diode: applied voltage makes n-side more positive than p-side
  - Electric field direction is from n-side towards p-side
  - Pushes charge carriers away from the p-n boundary
  - Depletion region widens, and no current flows

- Diode only conducts when positive voltage applied to p-side and negative voltage to n-side
- Diodes used in "rectifiers", to convert ac voltage to dc.

Depletion region becomes wider, barrier potential higher

Transistors

- Bipolar transistor = combination of two diodes that share middle portion, called "base" of transistor; other two sections: "emitter" and "collector"
- Usually, base is very thin and lightly doped.
- Two kinds of bipolar transistors: pnp and npn transistors
- "Pnp" means emitter is p-type, base is n-type, and collector is p-type material.
- In normal operation of pnp transistor, apply positive voltage to emitter, negative voltage to collector.

Operation of pnp transistor:

- If emitter-base junction is forward biased, "holes flow" from battery into emitter, move into base.
- Some holes annihilate with electrons in n-type base, but base thin and lightly doped ⇒ most holes make it through base into collector.
- Holes move through collector into negative terminal of battery; i.e. "collector current" flows whose size depends on how many holes have been captured by electrons in the base.
- This depends on the number of n-type carriers in the base which can be controlled by the size of the current (the "base current") that is allowed to flow from the base to the emitter; the base current is usually very small; small changes in the base current can cause a big difference in the collector current.
Transistor operation

- Transistor acts as amplifier of base current, since small changes in base current cause big changes in collector current.

- Transistor as switch: if voltage applied to base is such that emitter-base junction is reverse-biased, no current flows through transistor -- transistor is "off"

- Therefore, a transistor can be used as a voltage-controlled switch; computers use transistors in this way.

- "Field-effect transistor" (FET)
  - In a pnp FET, current flowing through a thin channel of n-type material is controlled by the voltage (electric field) applied to two pieces of p-type material on either side of the channel (current depends on electric field).

- Many different kinds of FETs; FETs are the kind of transistor most commonly used in computers.