Astronomical Observing Techniques 2019

Lecture 9: Silicon Eyes 1

Huub Rottgering
rottgering@strw.leidenuniv.nl

Modern Detectors

Photon detectors

Respond to individual photons and releases electrons

X-rays to IR

Examples: photoconductors, photodiodes, photoemissive detectors

Thermal detectors

Absorb photons which changes temperatures and/or resistance

IR and sub-mm detectors

Examples: bolometers

Coherent receivers

Responds directly to electrical field and preserve phase,

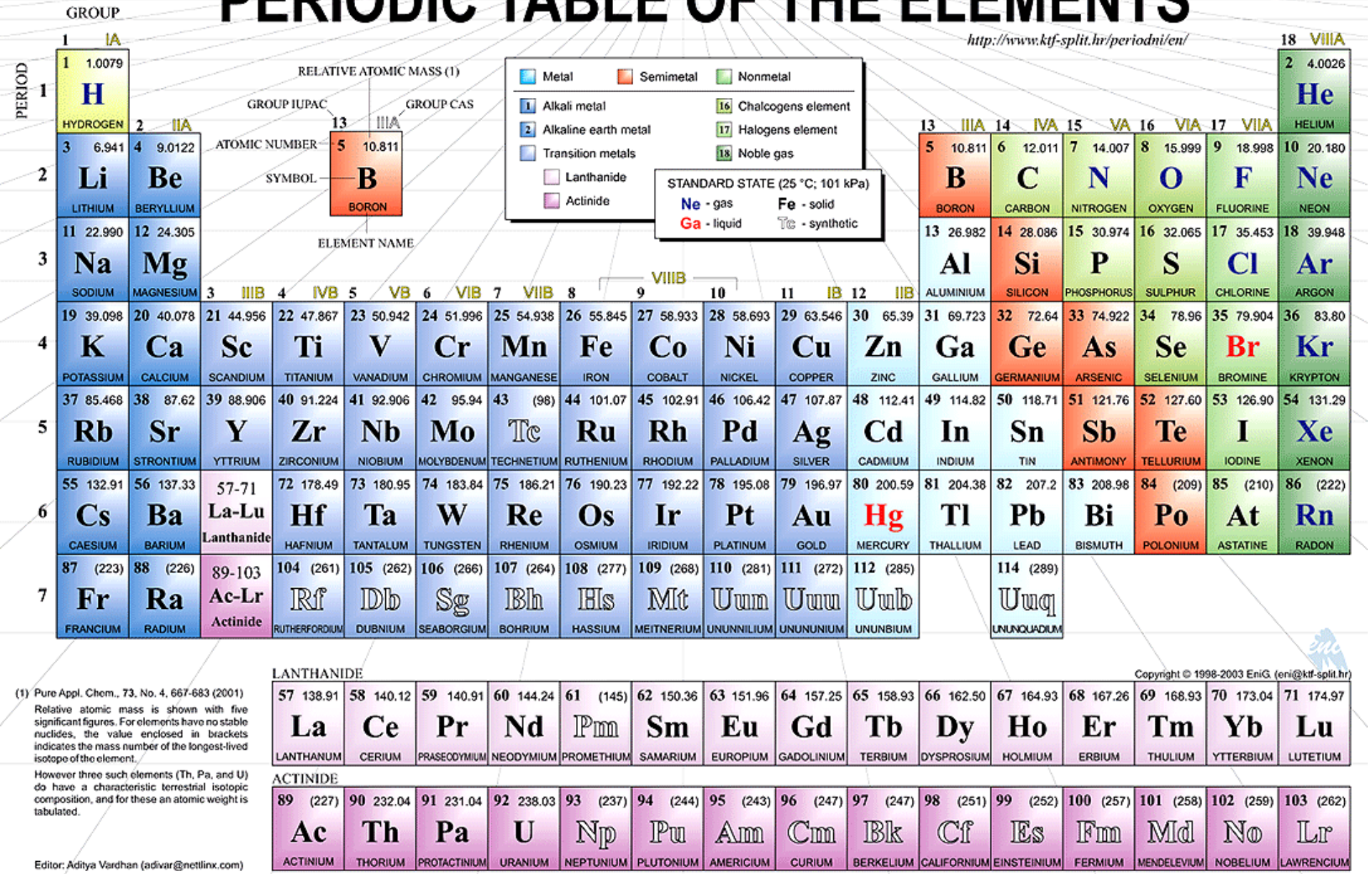
mainly used in the sub-mm and radio regime

Examples: heterodyne receivers

Content

1. Physical principles
   - Periodic table, Covalent Bond, Crystal lattices, Electronic Bands, Fermi Energy and Fermi Function, Electric Conductivity, Band Gap and Conduction Band
2. Intrinsic Photoconductors
   - Photo-Current
3. Extrinsic Photoconductors
   - Depletion Zone
4. Photodiodes
5. Charge Coupled Devices

PERIODIC TABLE OF THE ELEMENTS

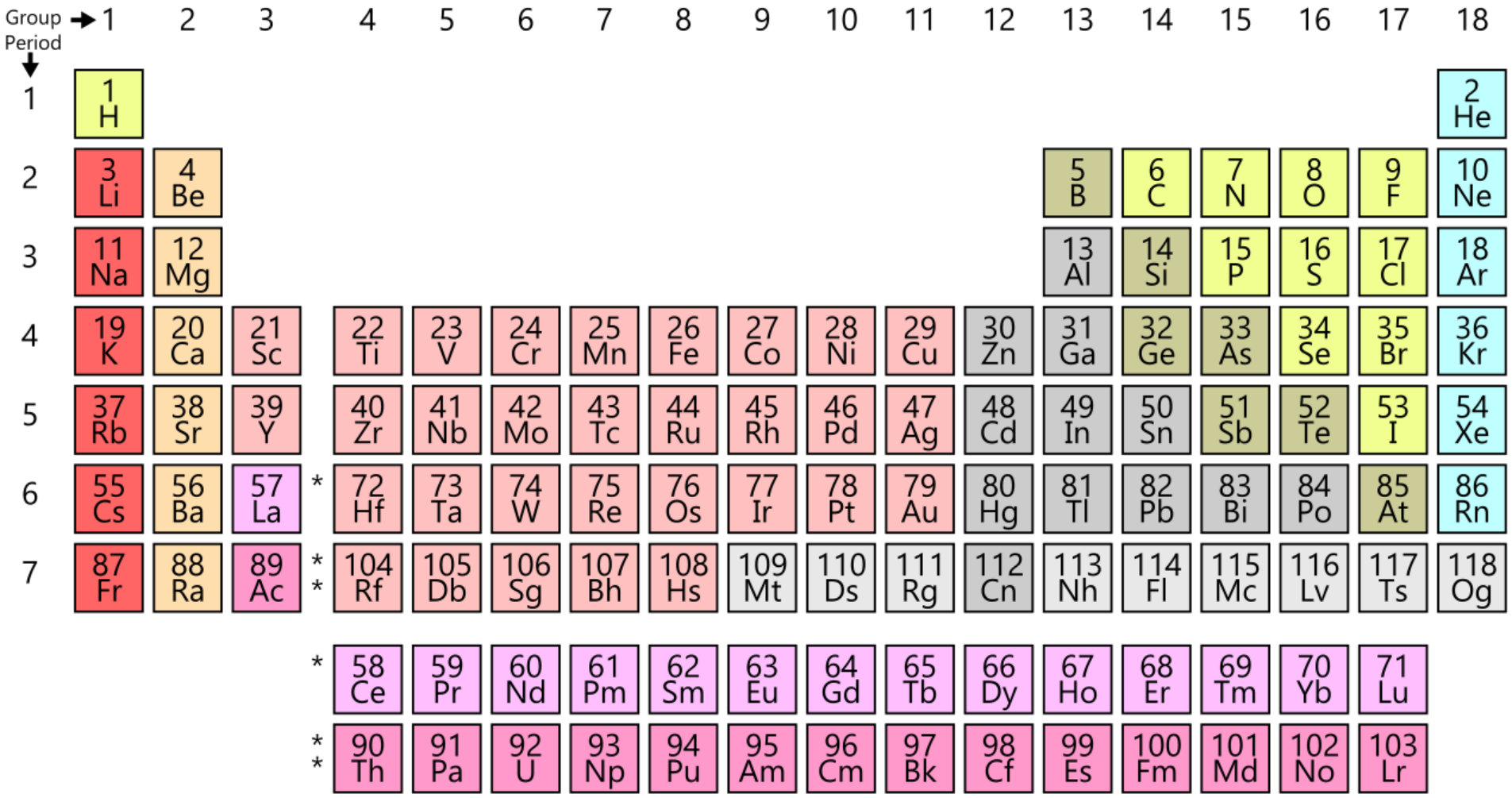


(1) Pure Appl. Chem., 73, No. 4, 667-683 (2001)

Relative atomic mass is shown with five significant figures. For elements have no stable nuclides, the value enclosed in brackets indicates the mass number of the longest-lived isotope of the element.

However three such elements (Th, Pa, and U) do have a characteristic terrestrial isotopic composition, and for these an atomic weight is tabulated.

Editor: Aditya Vardhan (adivar@netlinx.com)


Group → 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Period ↓
1 1 H 2 He
2 3 Li 4 Be 5 B 6 C 7 N 8 O 9 F 10 Ne
3 11 Na 12 Mg 13 Al 14 Si 15 P 16 S 17 Cl 18 Ar
4 19 K 20 Ca 21 Sc 22 Ti 23 V 24 Cr 25 Mn 26 Fe 27 Co 28 Ni 29 Cu 30 Zn 31 Ga 32 Ge 33 As 34 Se 35 Br 36 Kr
5 37 Rb 38 Sr 39 Y 40 Zr 41 Nb 42 Mo 43 Tc 44 Ru 45 Rh 46 Pd 47 Ag 48 Cd 49 In 50 Sn 51 Sb 52 Te 53 I 54 Xe
6 55 Cs 56 Ba 57 La * 72 Hf 73 Ta 74 W 75 Re 76 Os 77 Ir 78 Pt 79 Au 80 Hg 81 Tl 82 Pb 83 Bi 84 Po 85 At 86 Rn
7 87 Fr 88 Ra 89 Ac * 104 Rf 105 Db 106 Sg 107 Bh 108 Hs 109 Mt 110 Ds 111 Rg 112 Cn 113 Nh 114 Fl 115 Mc 116 Lv 117 Ts 118 Og
* 58 Ce 59 Pr 60 Nd 61 Pm 62 Sm 63 Eu 64 Gd 65 Tb 66 Dy 67 Ho 68 Er 69 Tm 70 Yb 71 Lu
* 90 Th 91 Pa 92 U 93 Np 94 Pu 95 Am 96 Cm 97 Bk 98 Cf 99 Es 100 Fm 101 Md 102 No 103 Lr

The **periodic table**, also known as the **periodic table of elements**, is a tabular display of the chemical elements, which are arranged by atomic number, electron configuration, and recurring chemical properties. The structure of the table shows *periodic trends*. The seven rows of the table, called periods, generally have metals on the left and non-metals on the right. The columns, called groups, contain elements with similar chemical behaviours. Six groups have accepted names as well as assigned numbers: for example, group 17 elements are the halogens; and group 18 are the noble gases.

https://en.wikipedia.org/wiki/Periodic_table

The Boron group

The **boron group** are the chemical elements in **group 13** of the periodic table, comprising boron (B), aluminium (Al), gallium (Ga), indium (In), thallium (Tl), The elements in the boron group are characterized by having three electrons in their outer energy levels (valence layers)

Z	Element	No. of electrons per shell
5	boron	2, 3
13	aluminium	2, 8, 3
31	gallium	2, 8, 18, 3
49	indium	2, 8, 18, 18, 3
81	thallium	2, 8, 18, 32, 18, 3

Carbon group

Z	Element	No. of electrons/shell
6	Carbon	2, 4
14	Silicon	2, 8, 4
32	Germanium	2, 8, 18, 4
50	Tin	2, 8, 18, 18, 4
82	Lead	2, 8, 18, 32, 18, 4

The **carbon group** or **Group 14** is a periodic table group consisting of carbon (C), silicon (Si), germanium (Ge), tin (Sn), lead (Pb), and flerovium (Fl)

In the field of semiconductor physics, it is still universally called **Group IV**.

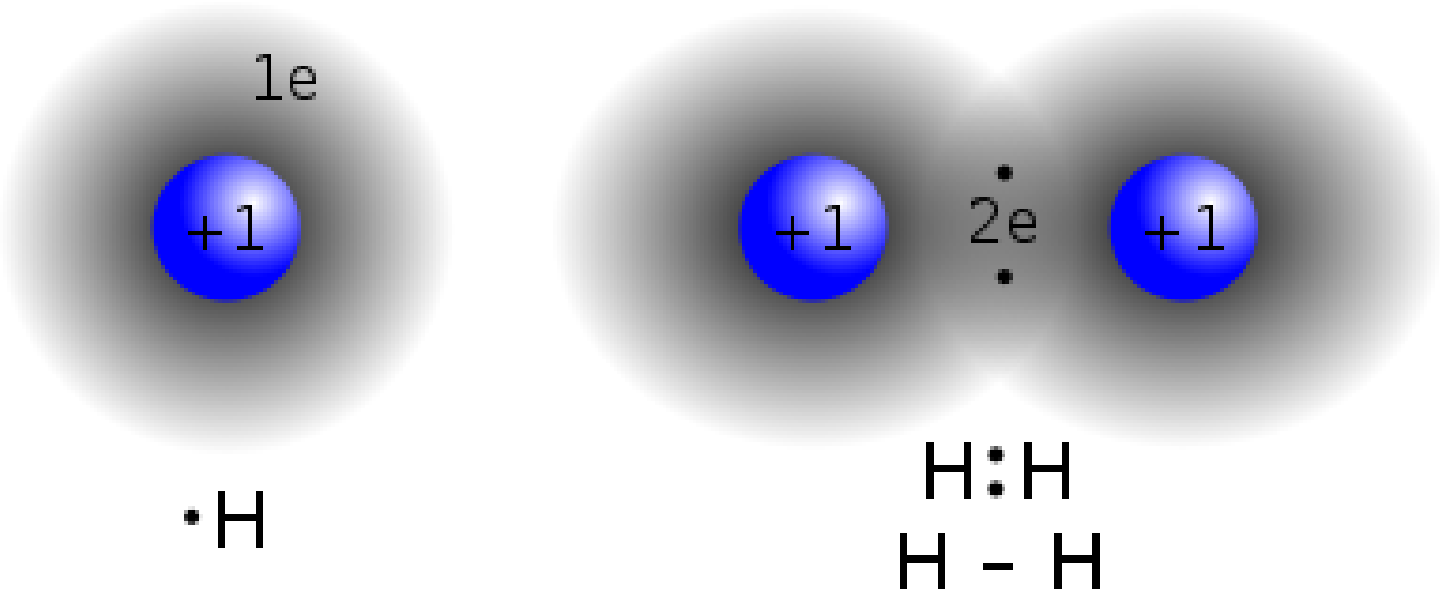
Like other groups, the members of this family show patterns in electron configuration, especially in the outermost shells, resulting in trends in chemical behavior.

The Nitrogen group

or Group 15 of the periodic table consists of the elements nitrogen (N), phosphorus (P), arsenic (As), antimony (Sb), bismuth (Bi). This group has the defining characteristic that all the component elements have 5 electrons in their outermost shell.

Z	Element	Electrons per shell
7	nitrogen	2, 5
15	phosphorus	2, 8, 5
33	arsenic	2, 8, 18, 5

Covelent bond



A covalent bond forming H_2 (right) where two hydrogen atoms share the two electrons

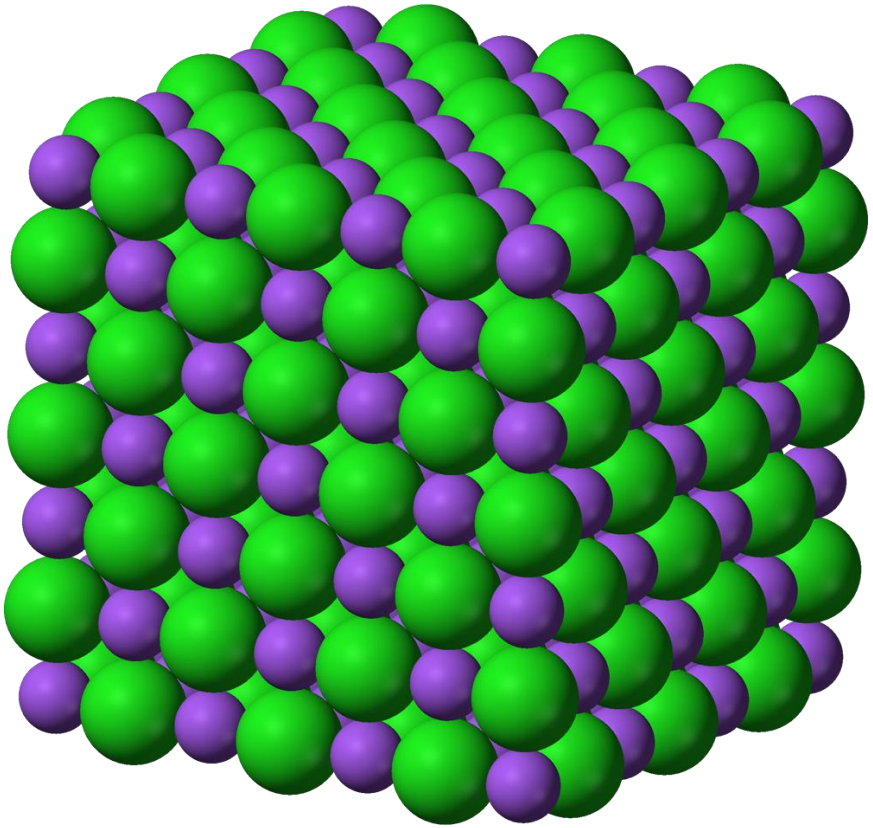
Electrons are fermions with spin ½, therefore they like to be in pairs

A **covalent bond**, also called a **molecular bond**, is a chemical bond that involves the sharing of electron pairs between atoms. These electron pairs are known as **shared pairs** or **bonding pairs**, and the stable balance of attractive and repulsive forces between atoms, when they share electrons, is known as covalent bonding. For many molecules, the sharing of electrons allows each atom to attain the equivalent of a full outer shell, corresponding to a stable electronic configuration.

https://en.wikipedia.org/wiki/Covalent_bond

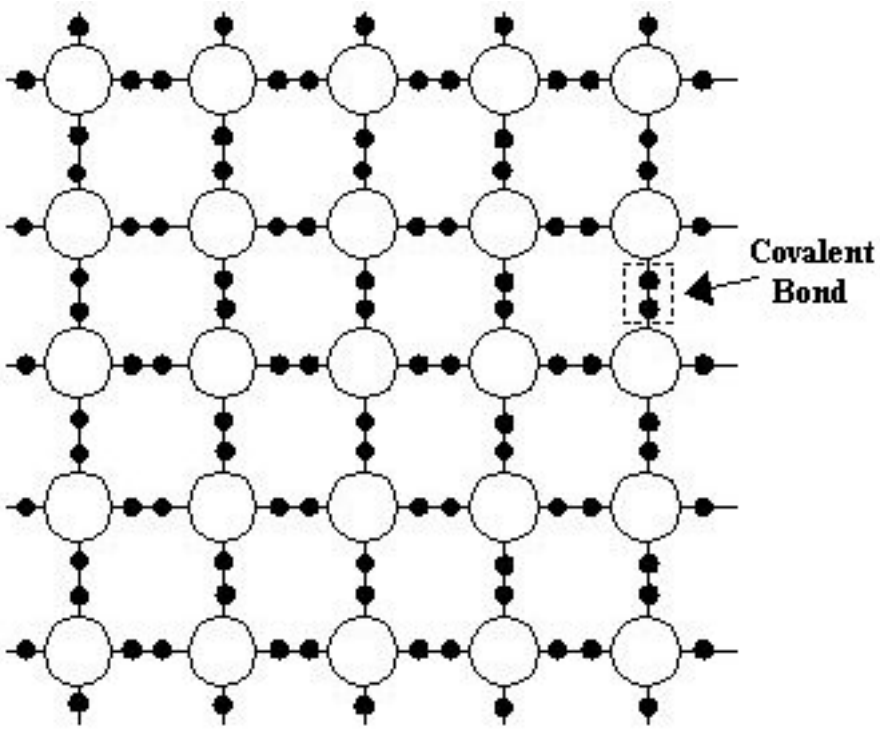
Crystal Lattice

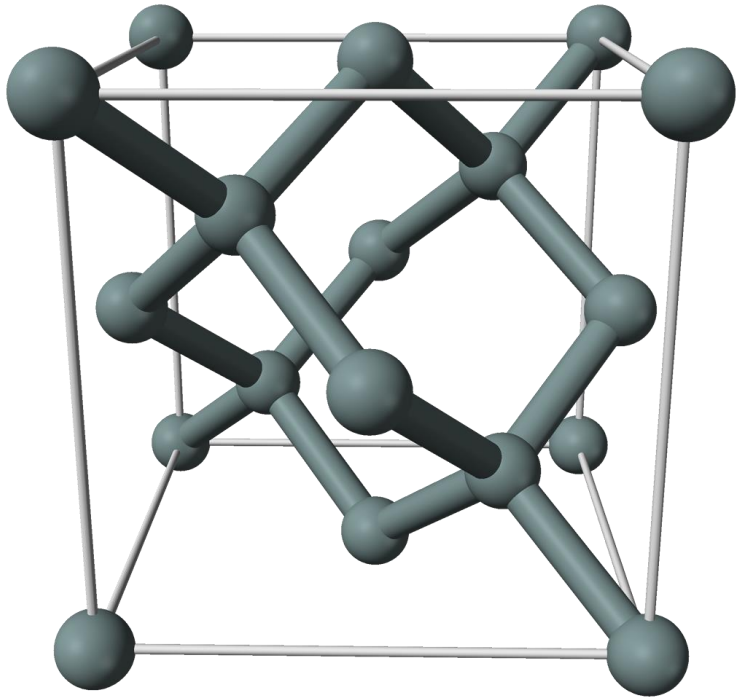
- crystals: periodic arrangement of atoms, ions or molecules
- smallest group of atoms that repeats is unit cell
- unit cells repeat at lattice points
- crystal structure and symmetry determine many physical properties



purple: Na^+, green: Cl^-

Crystalline silicon





https://en.wikipedia.org/wiki/Monocrystalline_silicon

- Elements with 4 e^- in valence shell form crystals with diamond lattice structure (each atom bonds to four neighbors).
- Covalent bond between neighbours due to “shared” electrons
- The crystalline structure of silicon forms a diamond cubic

Bands and band gaps

The **electronic band structure** (or simply **band structure**) of a solid describes the range of energies that an electron within the solid may have (called *energy bands*, *allowed bands*, or simply *bands*) and ranges of energy that it may not have (called *band gaps* or *forbidden bands*).

Band theory derives these bands and band gaps by examining the allowed quantum mechanical wave functions for an electron in a large, periodic lattice of atoms or molecules. Band theory has been successfully used to explain many physical properties of solids, such as electrical resistivity and optical absorption, and forms the foundation of the understanding of all solid-state devices(transistors, solar cells, etc.).

https://en.wikipedia.org/wiki/Electronic_band_structure

Why bands and band gaps occur

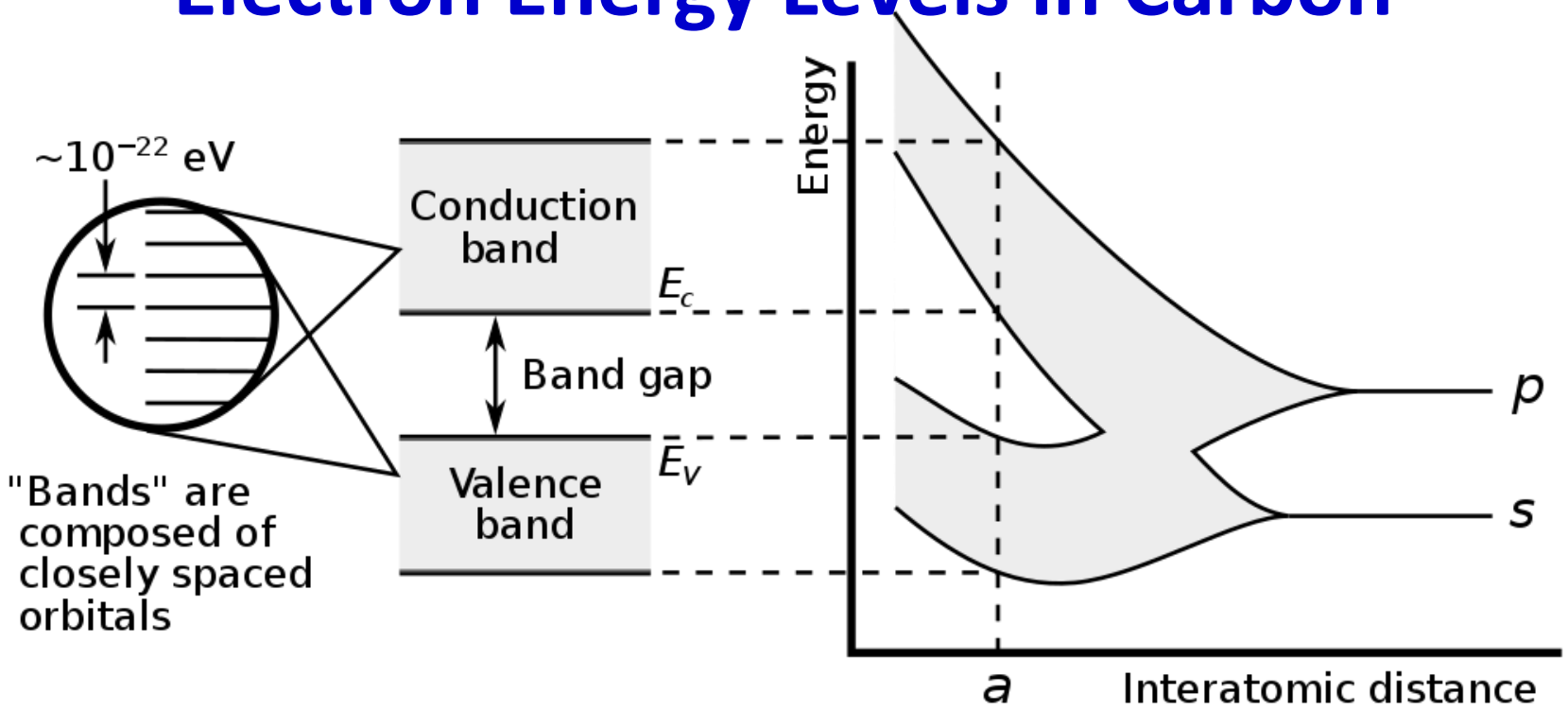
The electrons of a single, isolated atom occupy atomic orbitals each of which has a discrete energy level. When two or more atoms join together to form into a molecule, their atomic orbitals overlap. The Pauli exclusion principle dictates that no two electrons can have the same quantum numbers in a molecule. So if two identical atoms combine to form a diatomic molecule, each atomic orbital splits into two molecular orbitals of different energy, allowing the electrons in the former atomic orbitals to occupy the new orbital structure without any having the same energy.

Similarly if a large number *N* of identical atoms come together to form a solid, such as a crystal lattice, the atoms' atomic orbitals overlap. Since the Pauli exclusion principle dictates that no two electrons in the solid have the same quantum numbers, each atomic orbital splits into *N* discrete molecular orbitals, each with a different energy. Since the number of atoms in a macroscopic piece of solid is a very large number ($N \sim 10^{22}$) the number of orbitals is very large and thus they are very closely spaced in energy (of the order of 10^{-22} eV). The energy of adjacent levels is so close together that they can be considered as a continuum, an energy band.

This formation of bands is mostly a feature of the outermost electrons (valence electrons) in the atom, which are the ones involved in chemical bonding and electrical conductivity. The inner electron orbitals do not overlap to a significant degree, so their bands are very narrow.

Band gaps are essentially leftover ranges of energy not covered by any band, a result of the finite widths of the energy bands. The bands have different widths, with the widths depending upon the degree of overlap in the atomic orbitals from which they arise. Two adjacent bands may simply not be wide enough to fully cover the range of energy.

Electron Energy Levels in Carbon



Showing how electronic band structure comes about by the hypothetical example of a large number of carbon atoms being brought together to form a diamond crystal. The graph *(right)* shows the energy levels as a function of the spacing between atoms. When the atoms are far apart *(right side of graph)* each atom has valence atomic orbitals p and s which have the same energy. However when the atoms come closer together their orbitals begin to overlap. Due to the Pauli Exclusion Principle each atomic orbital splits into $N_{\text{molecular}}$ orbitals each with a different energy, where N is the number of atoms in the crystal. Since N is such a large number, adjacent orbitals are extremely close together in energy so the orbitals can be considered a continuous energy band. a is the atomic spacing in an actual crystal of diamond. At that spacing the orbitals form two bands, called the valence and conduction bands, with a band gap between them.

Fermi Energy

In quantum mechanics, a group of particles known as fermions (for example, electrons, protons and neutrons) obey the Pauli exclusion principle. This states that two fermions cannot occupy the same quantum state. Since an idealized non-interacting Fermi gas can be analyzed in terms of single-particle stationary states, we can thus say that two fermions cannot occupy the same stationary state. These stationary states will typically be distinct in energy. To find the ground state of the whole system, we start with an empty system, and add particles one at a time, consecutively filling up the unoccupied stationary states with the lowest energy. When all the particles have been put in, the **Fermi energy** is the kinetic energy of the highest occupied state.

As a consequence, even if we have extracted all possible energy from a Fermi gas by cooling it to near absolute zero temperature, the fermions are still moving around at a high speed. The fastest ones are moving at a velocity corresponding to a kinetic energy equal to the Fermi energy. This speed is known as the **Fermi velocity**. Only when the temperature exceeds the related **Fermi temperature**, do the electrons begin to move significantly faster than at absolute zero.

The Fermi energy is an important concept in the solid state physics of metals and superconductors. It is also a very important quantity in the physics of quantum liquids like low temperature helium (both normal and superfluid ^{3}He), and it is quite important to nuclear physics and to understanding the stability of white dwarf stars against gravitational collapse.

https://en.wikipedia.org/wiki/Fermi_energy

Fermi–Dirac distribution

For a system of identical fermions with thermodynamic equilibrium, the average number of fermions in a single-particle state *i* is given by the **Fermi–Dirac (F–D) distribution:**

$$\bar{n}_i = \frac{1}{e^{(\epsilon_i - \mu)/k_\mathrm{B}T} + 1}$$

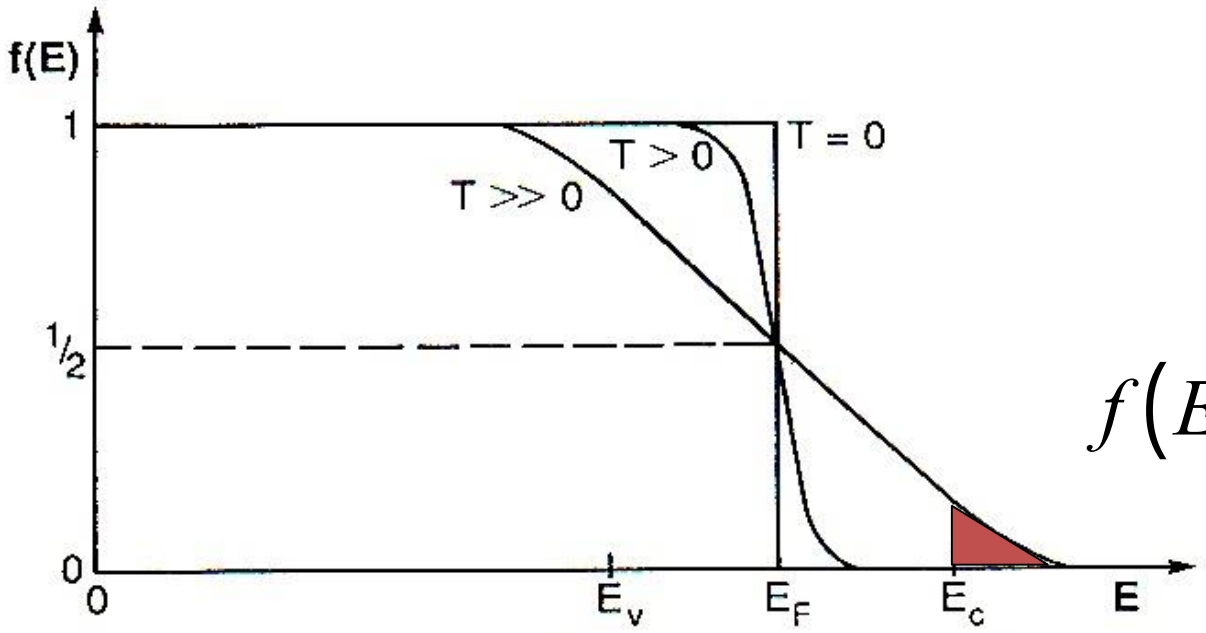
where k_B is Boltzmann's constant, T is the absolute temperature, ε_i is the energy of the single-particle state i, and μ is the total chemical potential.

At zero temperature, μ is equal to the Fermi energy plus the potential energy per electron.

Since the F–D distribution was derived using the Pauli exclusion principle, which allows at most one electron to occupy each possible state, a result is that $0 < \bar{n}_i < 1$

Fermi Energy

- Fermi function *f(E)* is probability that state of energy *E* is occupied at temperature *T*; $f(E_F) = 0.5$

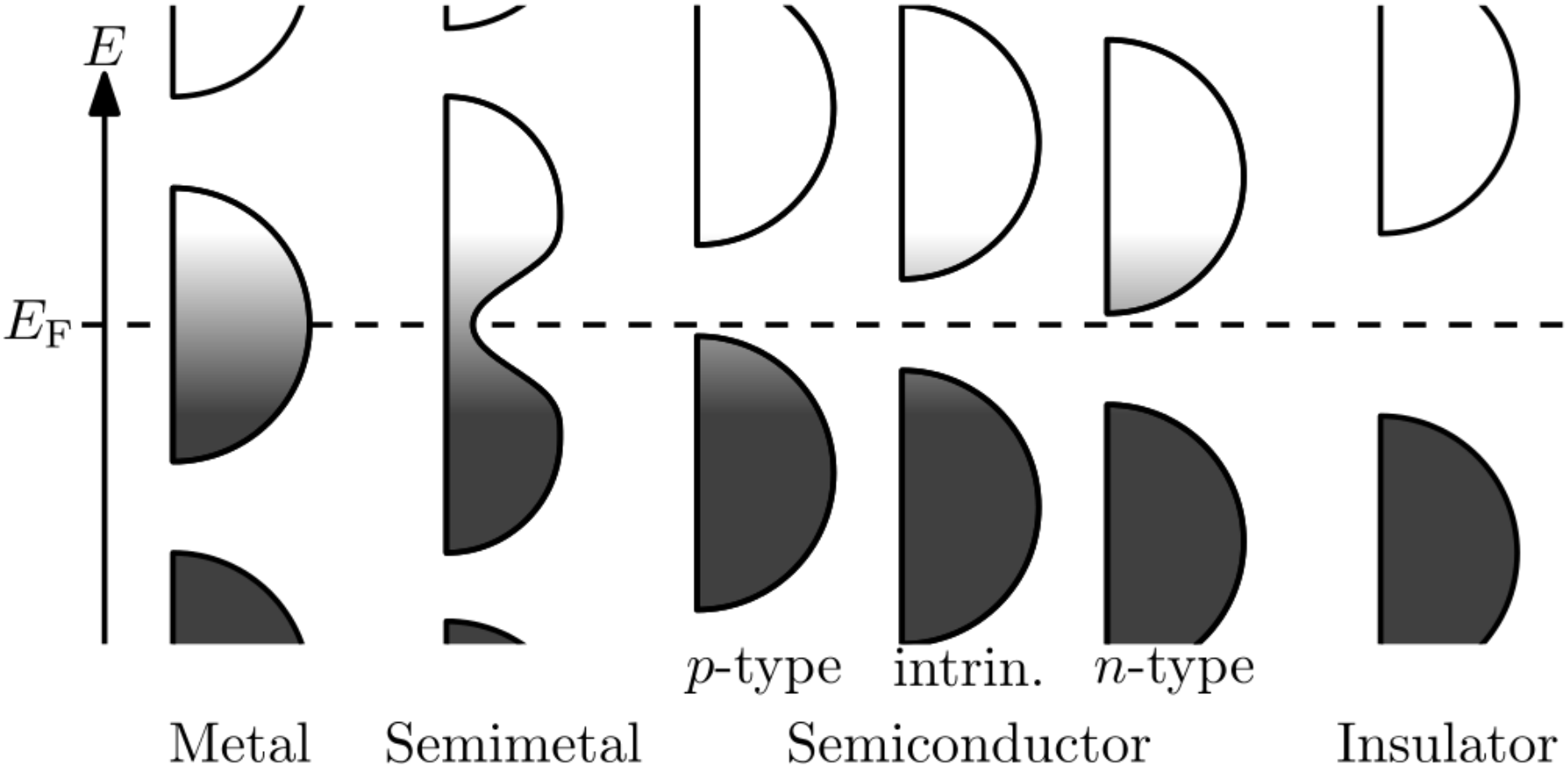


$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}}$$

Electric Conductivity

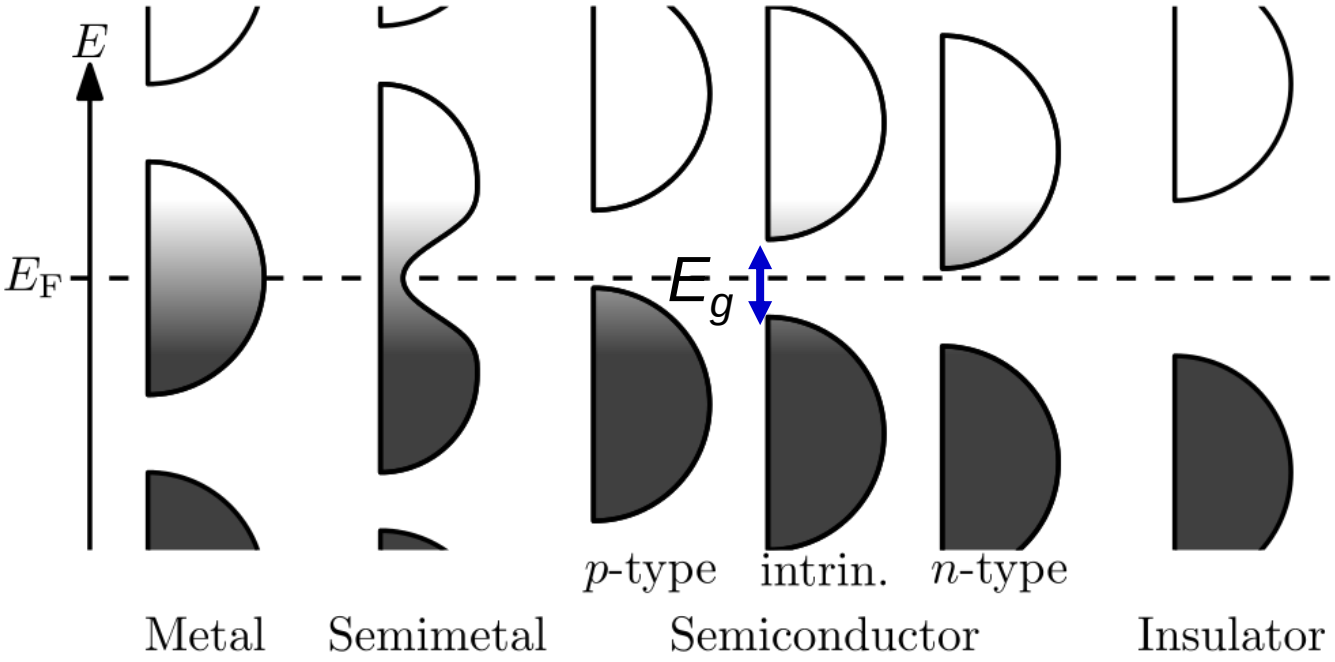
- The material's electrons seek to minimize the total energy in the material by going to low energy states; however, the Pauli exclusion principle means that only one can exist in each such state. So the electrons "fill up" the band structure starting from the bottom. The characteristic energy level up to which the electrons have filled is called the Fermi level. The position of the Fermi level with respect to the band structure is very important for electrical conduction: only electrons in energy levels near or above the Fermi level are free to move around, since the electrons can easily jump among the partially occupied states in that region. In contrast, the low energy states are rigidly filled with a fixed number of electrons at all times, and the high energy states are empty of electrons at all times.
- Electric current consists of a flow of electrons. In metals there are many electron energy levels near the Fermi level, so there are many electrons available to move. This is what causes the high electronic conductivity of metals.
- An important part of band theory is that there may be forbidden bands of energy: energy intervals that contain no energy levels. In insulators and semiconductors, the number of electrons is just the right amount to fill a certain integer number of low energy bands, exactly to the boundary. In this case, the Fermi level falls within a band gap. Since there are no available states near the Fermi level, and the electrons are not freely movable, the electronic conductivity is very low.

https://en.wikipedia.org/wiki/Electrical_resistivity_and_conductivity



Filling of the electronic states in various types of materials at equilibrium. Here, height is energy while width is the density of available states for a certain energy in the material listed. The shade follows the Fermi–Dirac distribution (**black** = all states filled, **white** = no state filled). In metals and semimetals the Fermi level E_F lies inside at least one band. In insulators and semiconductors the Fermi level is inside a band gap; however, in semiconductors the bands are near enough to the Fermi level to be thermally populated with electrons or holes.

Overcoming the Bandgap



Overcome bandgap E_g by lifting e^- into conduction band:

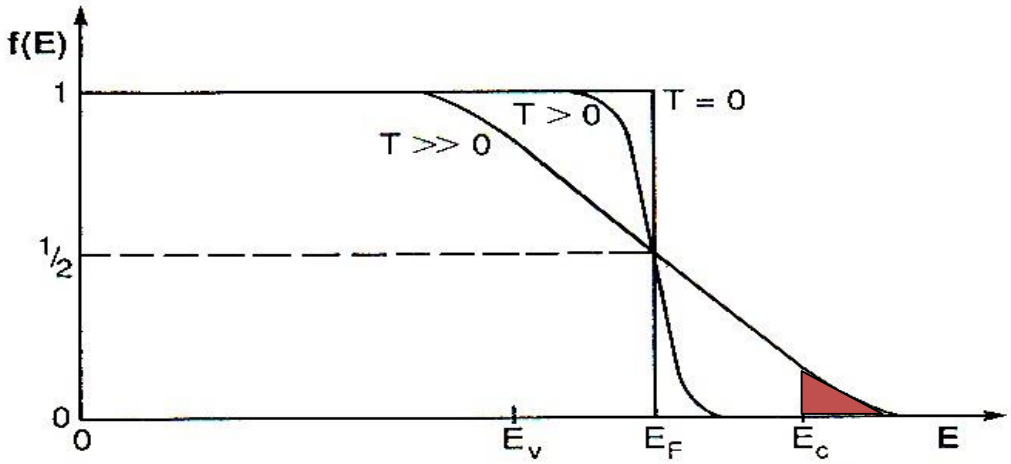
1. external excitation, e.g. via a photon ←photon detector
2. thermal excitation
3. impurities

Electrons in Conduction Band

- Number of occupied states in conduction band is given by product of number of possible states N_c in conduction band times the probability $f(E_c)$ that they are occupied
- For silicon, temperature increase of 8K doubles number of electrons in conduction band

$$n_0 = N_c f(E_c)$$

$$N_c = 2\left(\frac{2\ m_{eff} kT}{h^2}\right)^{3/2}$$



$$f(E_c) = \frac{1}{1 + e^{(E_c - E_F)/kT}} \overset{E_c - E_F >> kT}{\approx} e^{-(E_c - E_F)/kT}$$

Photoconductivity

When a photoconductive material is connected as part of a circuit, it functions as a resistor whose resistance depends on the light intensity. In this context, the material is called a photoresistor (also called *light-dependent resistor* or *photoconductor*). The most common application of photoresistors is as photodetectors, i.e. devices that measure light intensity.

https://en.wikipedia.org/wiki/Photoconductivity

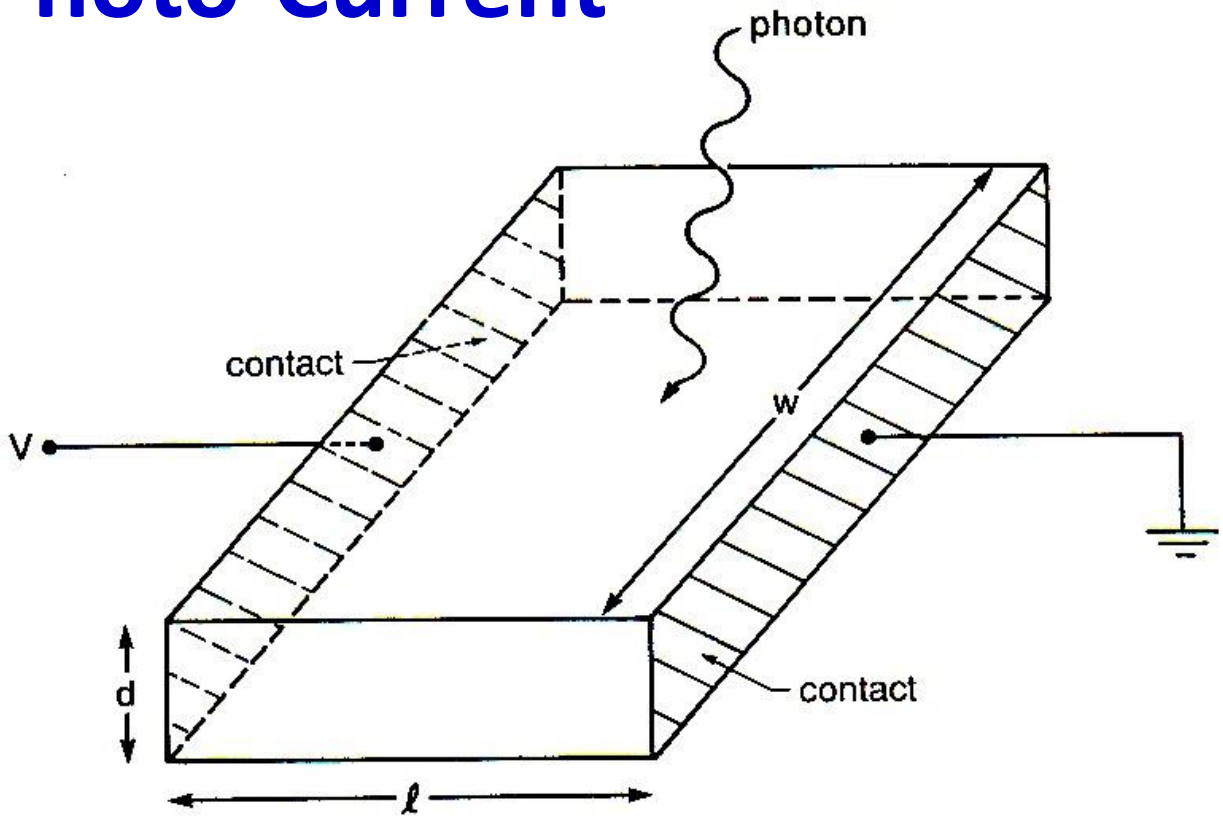
Intrinsic Photo-Conductors: Basic Principle

- photon lifts e^- into conduction band
- applied electric field drives charges to electrodes
- few charge carriers → high resistance

Photo-Current

- Conductivity: $j=\sigma E$
- Current: $I=jwd$
- $V=RI$, $E=V/l$
- $\sigma=j/E=jl/V=jl/(RI)$
  $=jl/(Rjwd)$
  $=1/R \cdot l/wd$

$$\sigma = \frac{1}{R_d}\frac{l}{wd}$$



where:

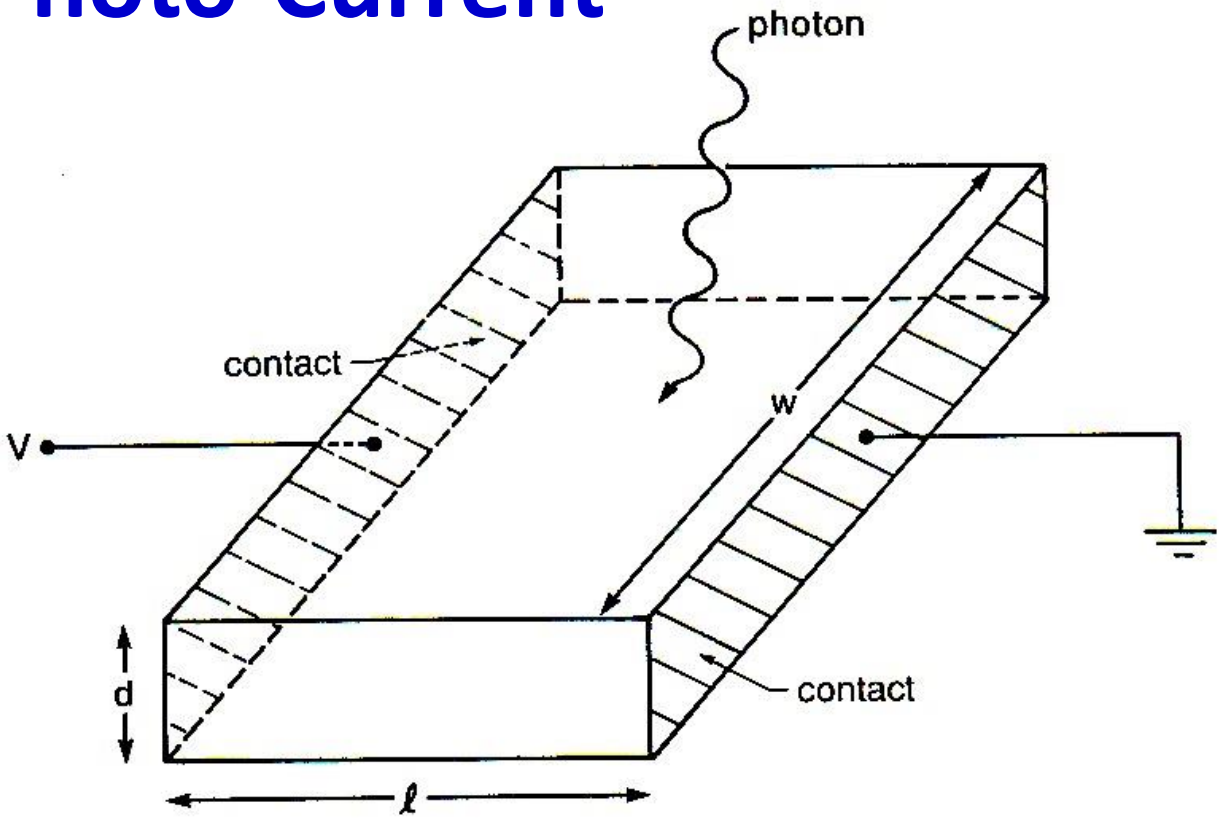
R_d = resistance

w,d,l = geometric dimensions

Photo-Current

$$\sigma = \frac{1}{R_d}\frac{l}{wd} = qn_0\mu_n$$

$$n_0 = \frac{j\,ht}{wdl}$$



where:

q = elementary charge

n_0 = number density of charge carriers

φ = photon flux

η = quantum efficiency

τ = mean lifetime before recombination

μ_n = electron mobility; drift velocity $v=\mu_n E$

Important Quantities and Definitions

Quantum efficiency η = $\dfrac{\#\text{ absorbed photons}}{\#\text{ incoming photons}}$

Responsivity S = $\dfrac{\text{electrical output signal}}{\text{input photon power}}$

Wavelength cutoff: $\lambda_c = \dfrac{hc}{E_g} = \dfrac{1.24\mu m}{E_g[eV]}$ E_g :Energy bandgap

Photo-current: $I_{ph} = q\varphi\eta G$

Photoconductive gain G: $G = \dfrac{I_{ph}}{q\varphi\eta} = \dfrac{\tau}{\tau_t} = \dfrac{\text{carrier lifetime}}{\text{transit time}}$

The product ηG describes the probability that an incoming photon will produce an electric charge that will reach an electrode

Limitations of Intrinsic Semiconductors

- long-wavelength cutoffs

$$\lambda_c = \frac{hc}{E_g}$$

Germanium:	1.85μm
Silicon:	1.12μm
GaAs:	0.87μm

- difficult to create completely pure material
- problems to make good electrical contacts to pure Si
- difficult to avoid impurities and minimize thermal noise

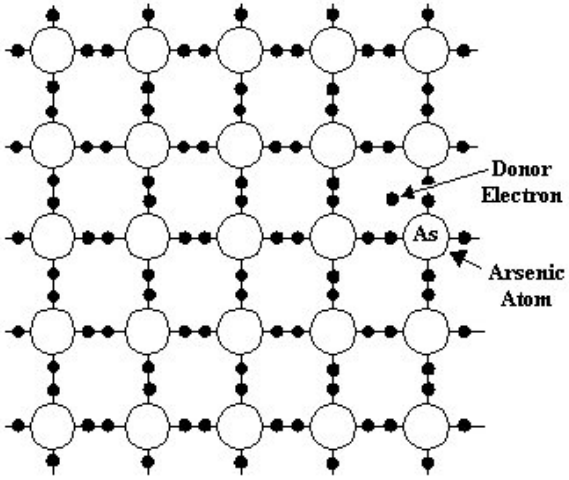
Extrinsic semiconductor

An **extrinsic** semiconductor is one that has been *doped*; during manufacture of the semiconductor crystal a trace element or chemical called a doping agent has been incorporated chemically into the crystal, for the purpose of giving it different electrical properties than the pure semiconductor crystal, which is called an *intrinsic semiconductor*. In an extrinsic semiconductor it is these foreign dopant atoms in the crystal lattice that mainly provide the charge carriers which carry electric current through the crystal. The doping agents used are of two types, resulting in two types of extrinsic semiconductor. An *electron donor* dopant is an atom which, when incorporated in the crystal, releases a mobile conduction electron into the crystal lattice. An extrinsic semiconductor which has been doped with electron donor atoms is called an **n-type semiconductor**, because the majority of charge carriers in the crystal are negative electrons. An *electron acceptor* dopant is an atom which accepts an electron from the lattice, creating a vacancy where an electron should be called a *hole* which can move through the crystal like a positively charged particle. An extrinsic semiconductor which has been doped with electron acceptor atoms is called a **p-type semiconductor**, because the majority of charge carriers in the crystal are positive holes.

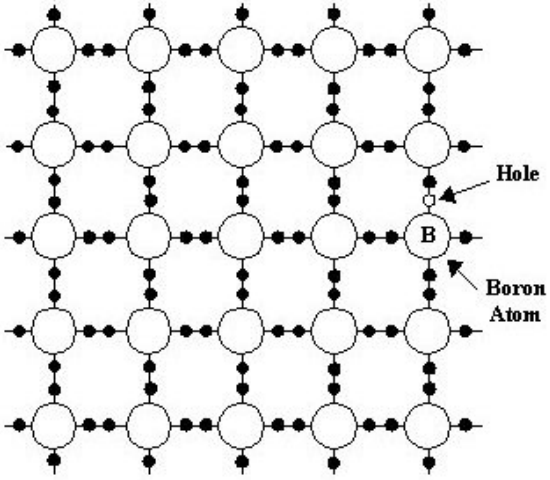
https://en.wikipedia.org/wiki/Extrinsic_semiconductor

Extrinsic Semiconductors

- extrinsic semiconductors:
  charge carriers = electrons (n-type) or holes (p-type)
- addition of impurities at low concentration to provide excess electrons or holes
- much reduced bandgap -> longer wavelength cutoff



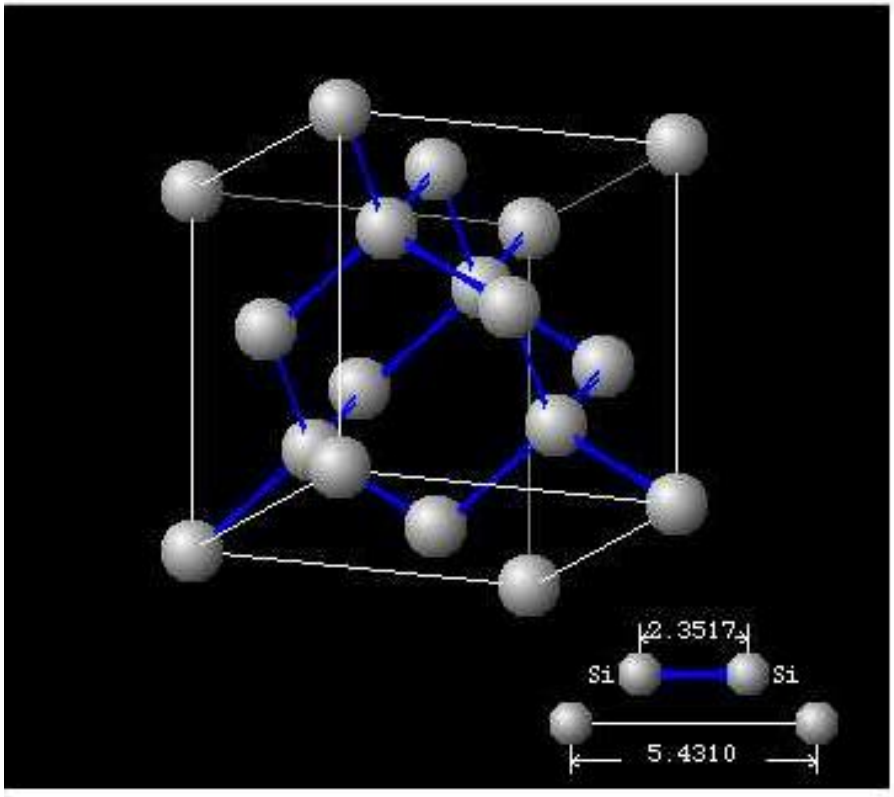
N-type silicon

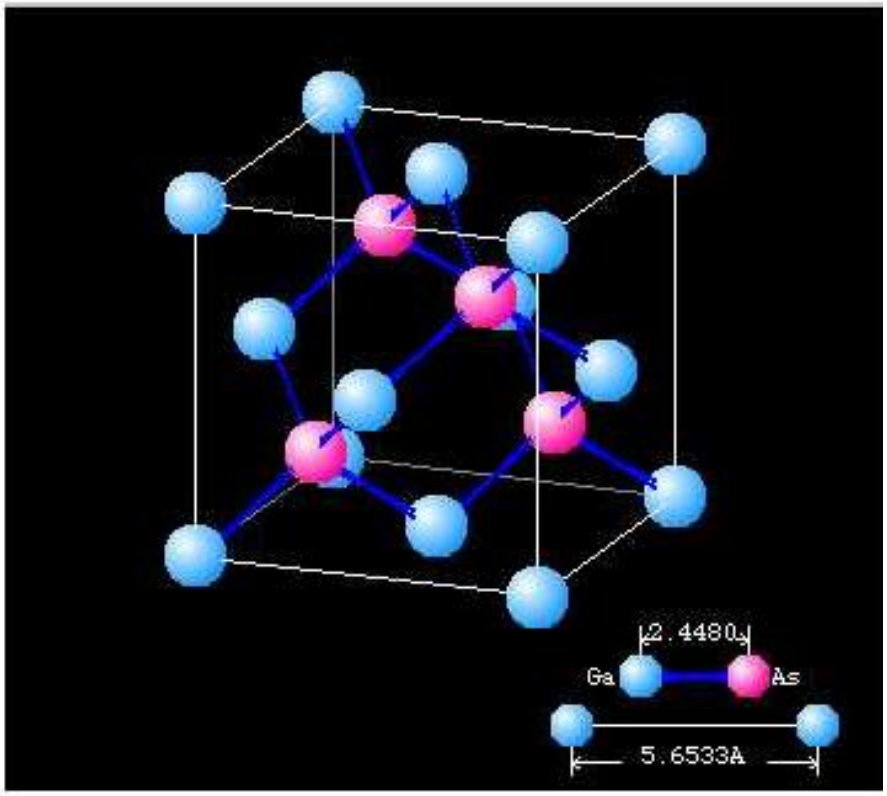


P-type silicon

Example: addition of boron to silicon in the ratio 1:100,000 increases its conductivity by a factor of 1000!

Diamond Lattices



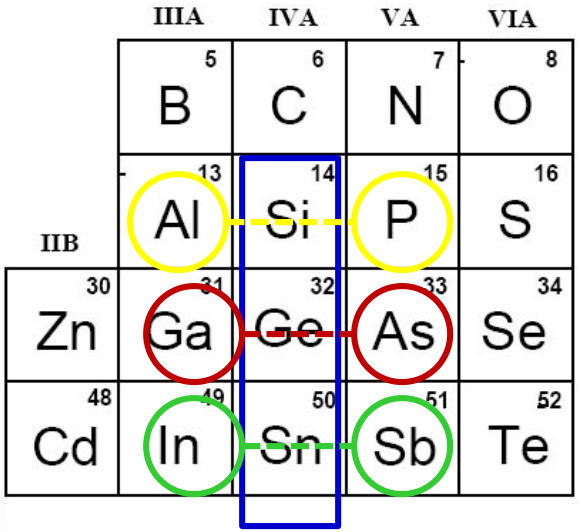


Silicon

Gallium+Arsenide

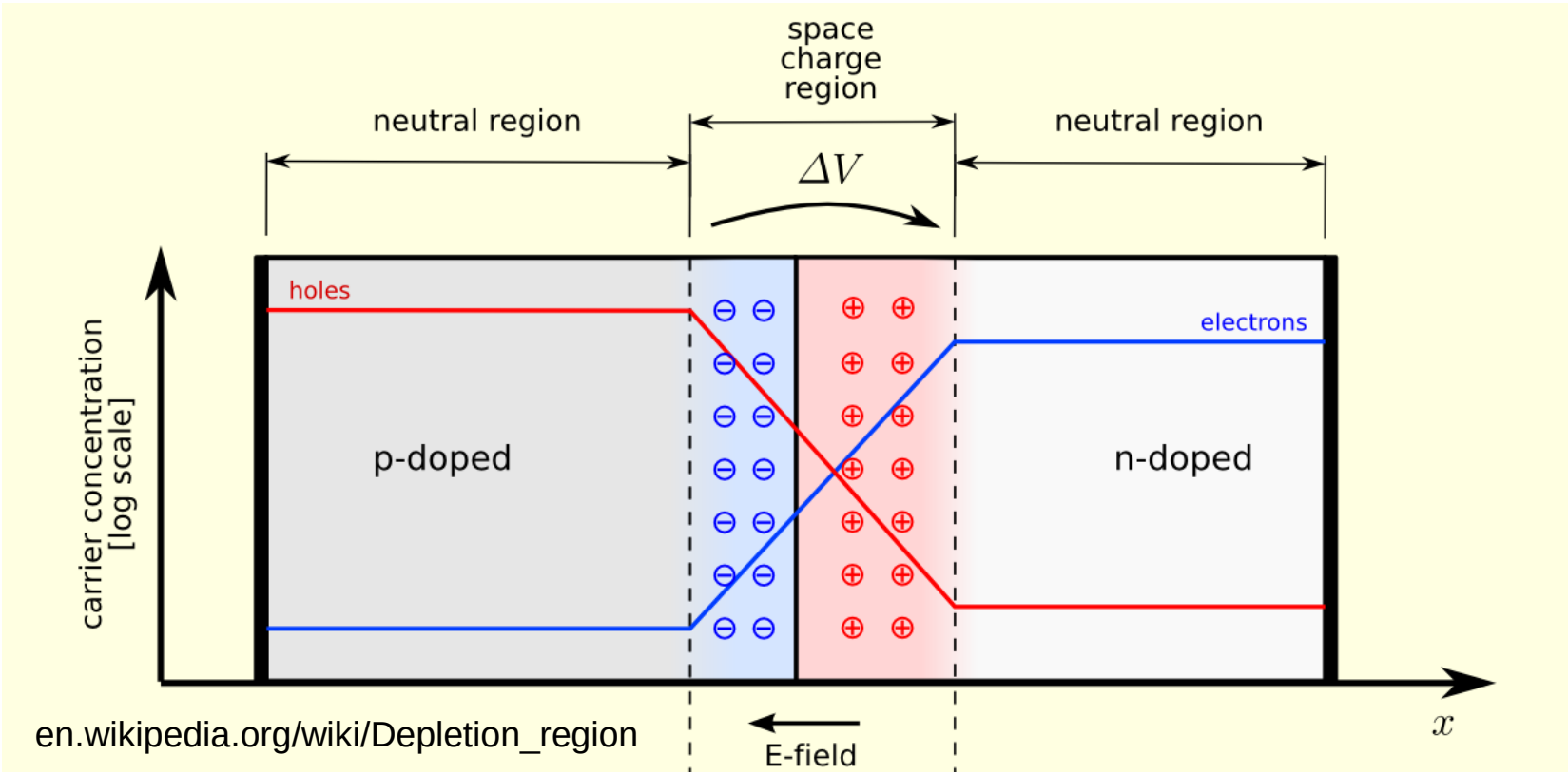
Extrinsic Semiconductor Band Gaps

		Ge	Si
Impurity	Type	Cutoff wavelength λ_c (μm)	Cutoff wavelength λ_c (μm)
Al	p		18.5[a]
B	p	119[b]	28[a]
Be	p	52[b]	8.3[a]
Ga	p	115[b]	17.2[a]
In	p	111[b]	7.9[a]
As	n	98[b]	23[a]
Cu	p	31[b]	5.2[a]
P	n	103[b]	27[a]
Sb	n	129[b]	29[a]



Problem: absorption coefficients much less than for intrinsic photoconductors → low quantum efficiency → active volumes of pixels must be large

PN Junction + Depletion Zone



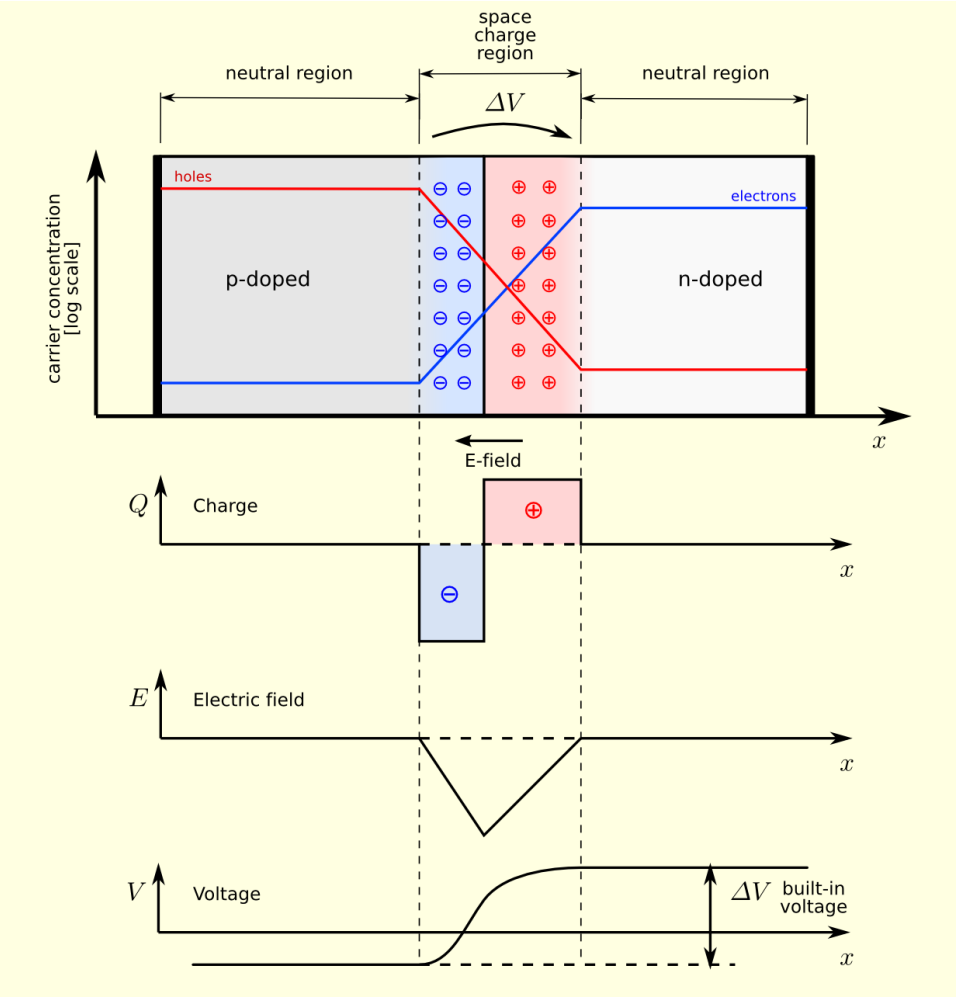
- junction between p- and n-doped Si (both are electrically neutral)
- e^- migrate to P-side, holes migrate to N-side
- e^- can only flow over large distances in n-type material, holes can only flow in p-type material

Depletion region

- In semiconductor physics, the **depletion region**, also called **depletion layer**, **depletion zone**, **junction region**, **space charge region** or **space charge layer**, is an insulating region within a conductive, doped semiconductormaterial where the mobile charge carriers have been diffused away, or have been forced away by an electric field. The only elements left in the depletion region are ionized donor or acceptor impurities.
- The depletion region is so named because it is formed from a conducting region by removal of all free charge carriers, leaving none to carry a current. Understanding the depletion region is key to explaining modern semiconductor electronics: diodes, bipolar junction transistors, field-effect transistors, and variable capacitance diodes all rely on depletion region phenomena.

Depletion Zone / PN Junction

- migrating e^- from N-side to P-side produces positive donor ion on N-side; migrating hole produces negative acceptor ion on P-side
- migrating e^- recombine with holes on P-side; migrating holes recombine with e^- on N-side
- migrating e^- and holes, mobile charge carriers are depleted
- charged ions remain adjacent to interface



en.wikipedia.org/wiki/Depletion_region

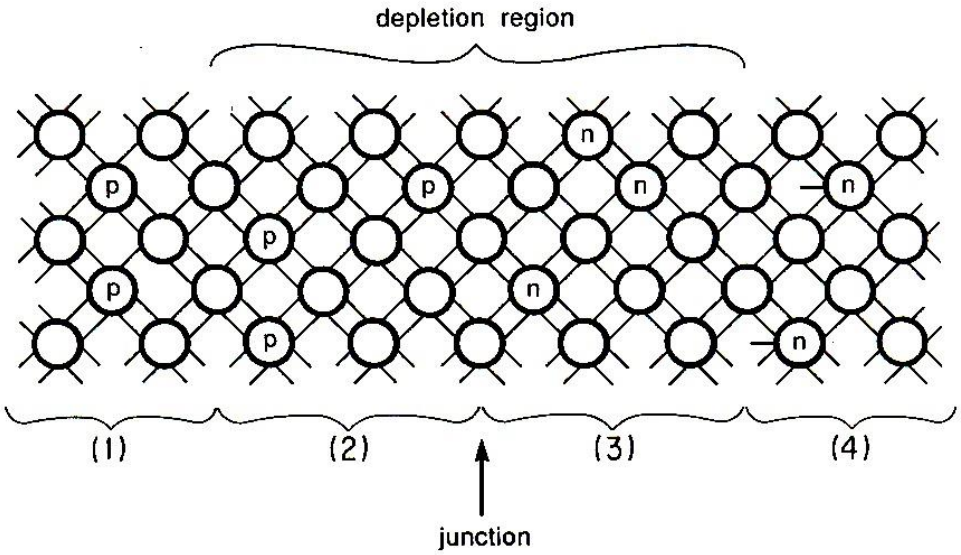
A depletion region forms instantaneously across a p–n junction. It is most easily described when the junction is in thermal equilibrium or in a steady state: in both of these cases the properties of the system do not vary in time; they have been called dynamic equilibrium.[1][2]

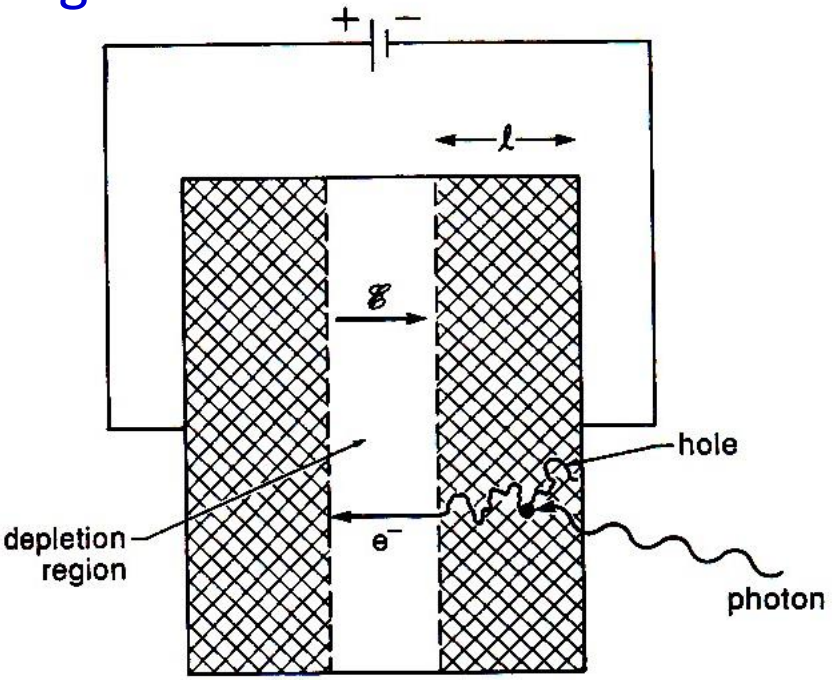
Electrons and holes diffuse into regions with lower concentrations of them, much as ink diffuses into water until it is uniformly distributed. By definition, the N-type semiconductor has an excess of free electrons (in the conduction band) compared to the P-type semiconductor, and the P-type has an excess of holes (in the valence band) compared to the N-type. Therefore, when N-doped and P-doped semiconductors are placed together to form a junction, free electrons in the N-side conduction band migrate (diffuse) into the P-side conduction band, and holes in the P-side valence band migrate into the N-side valence band.

Following transfer, the diffused electrons come into contact with holes and are eliminated by recombination in the P-side. Likewise, the diffused holes are recombined with free electrons so eliminated in the N-side. The net result is that the diffused electrons and holes are gone. In a N-side region near to the junction interface, free electrons in the conduction band are gone due to (1) the diffusion of electrons to the P-side and (2) recombination of electrons to holes that are diffused from the P-side. Holes in a P-side region near to the interface are also gone by a similar reason. As a result, majority charge carriers (free electrons for the N-type semiconductor, and holes for the P-type semiconductor) are depleted in the region around the junction interface, so this region is called the **depletion region** or **depletion zone**. Due to the majority charge carrier diffusion described above, the depletion region is charged; the N-side of it is positively charged and the P-side of it is negatively charged. This creates an electric field that provides a force opposing the charge diffusion. When the electric field is sufficiently strong to cease further diffusion of holes and electrons, the depletion region reached the equilibrium. Integrating the electric field across the depletion region determines what is called the **built-in voltage** (also called the junction voltage or barrier voltage or contact potential).

Photodiodes

- junction between *two* oppositely doped zones
- 2 adjacent zones create a depletion region



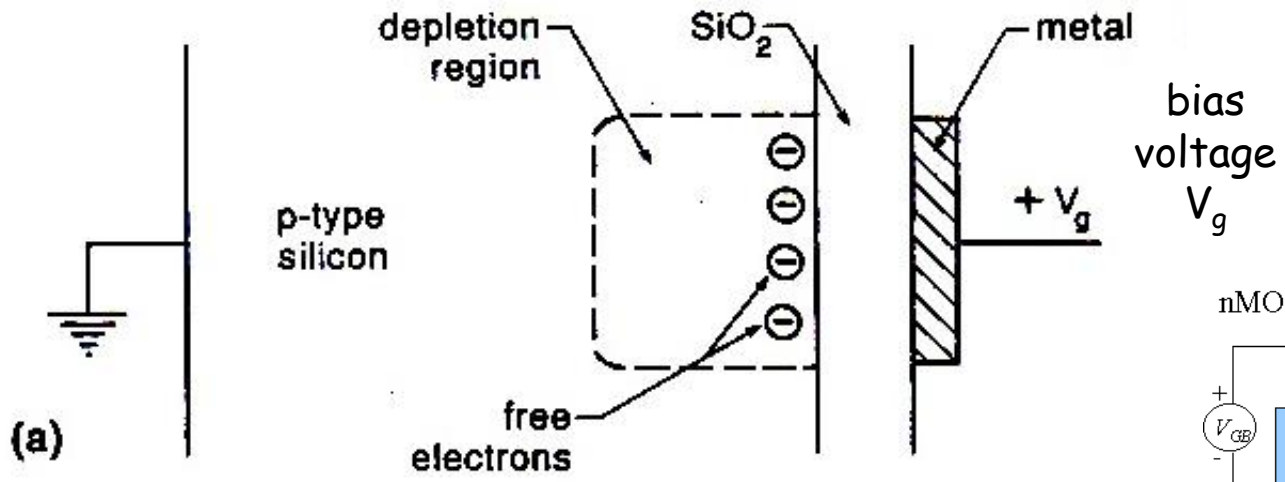


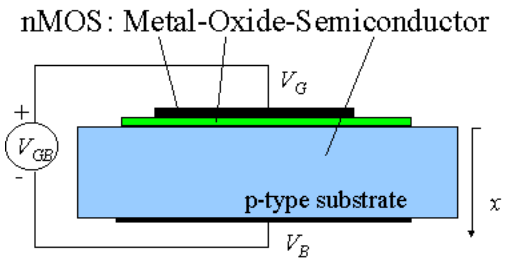
1. Photon gets absorbed e.g. in the p-type part
2. Absorption creates an e^--hole pair
3. The e^- diffuses through the material
4. Voltage drives the e^- across the depletion region → photo-current

Charge Coupled Devices (CCDs)

CCDs = array of integrating capacitors.

Pixel structure: metal "gate" evaporated onto SiO_2 (isolator) on silicon = MOS (metal-oxide-semiconductor)

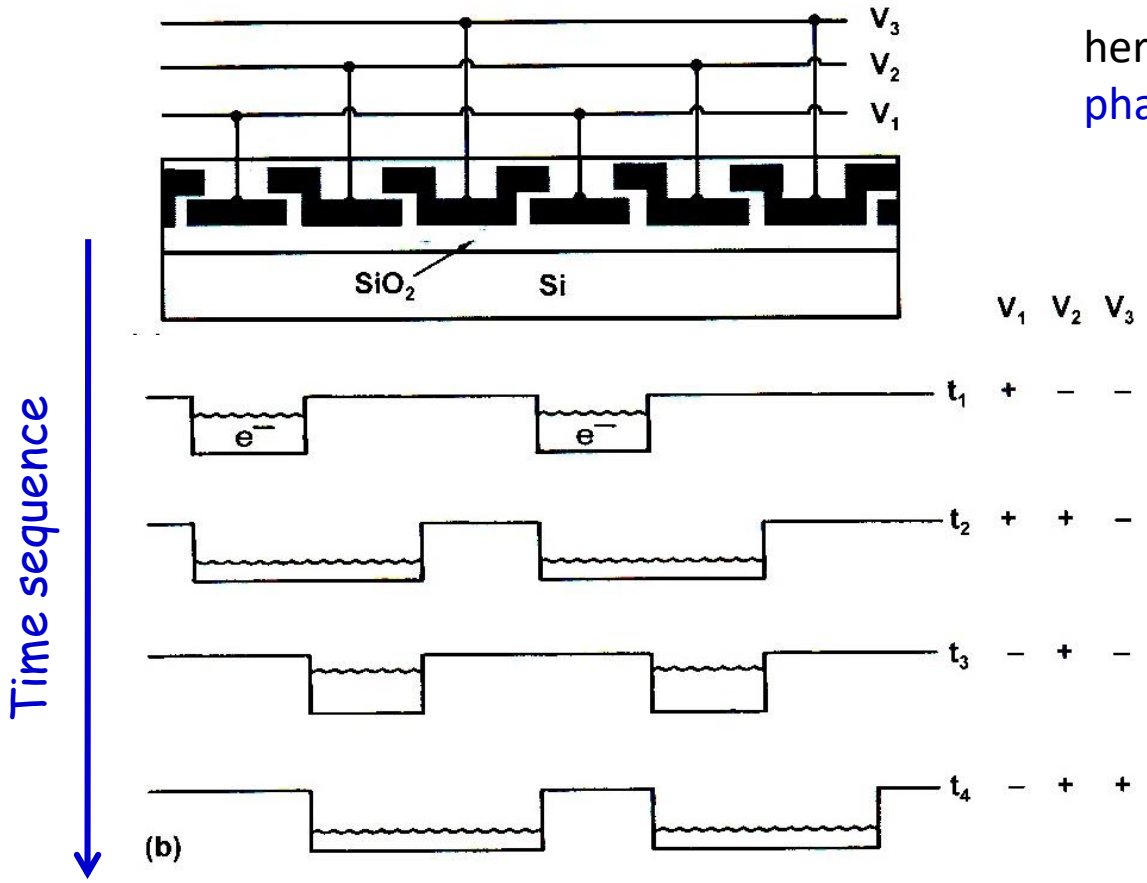




1. photons create free e^- in the photoconductor
2. e^- drift toward the electrode but cannot penetrate the SiO_2 layer
3. e^- accumulate at the Si—SiO_2 interface
4. total charge collected at interface measures number of photons during the exposure
5. → read out the number of e^-

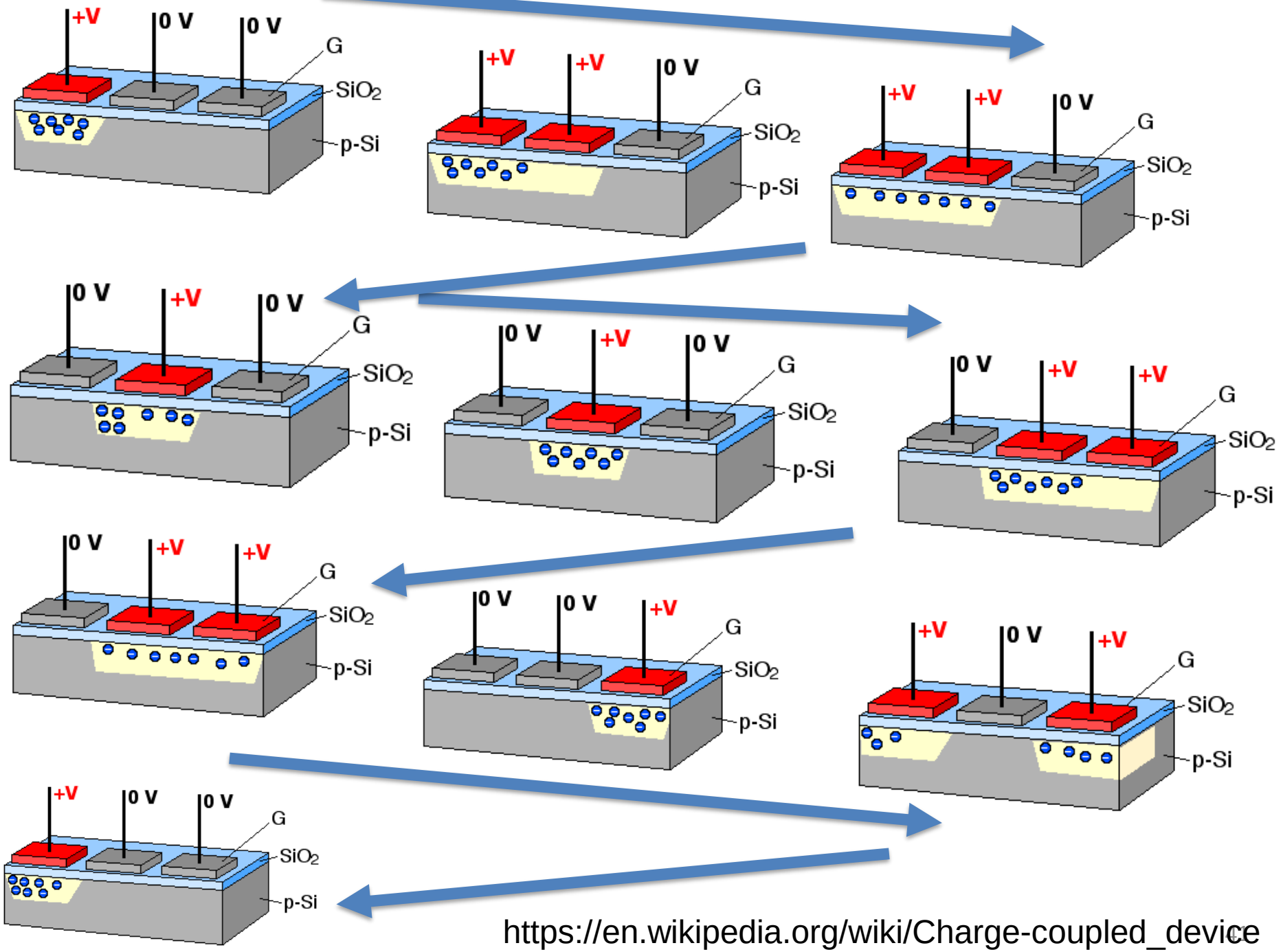
Charge Coupled Readouts

Charges are moved along columns to the edge of the array to the output amplifier

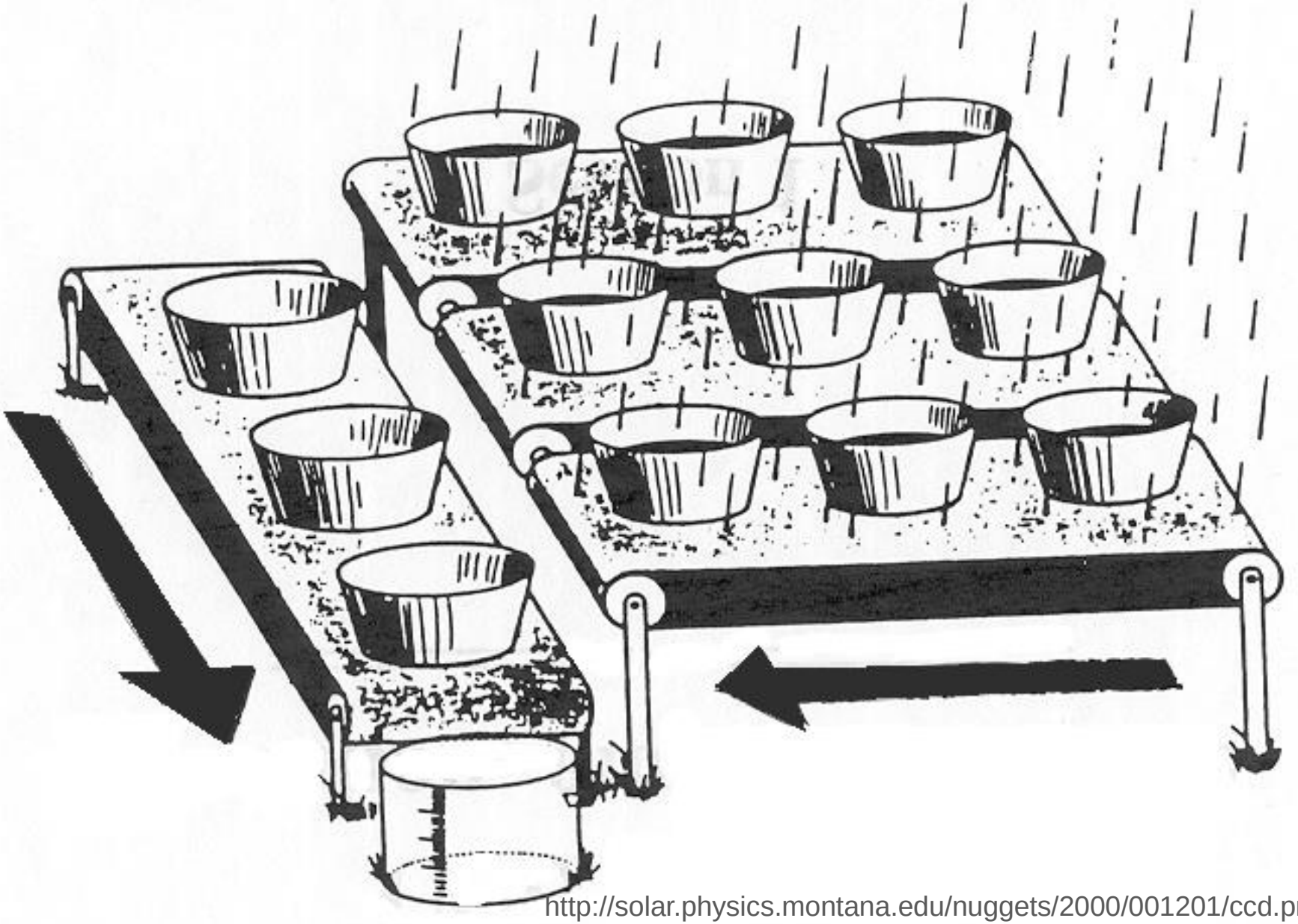


here: 3 sets of electrodes → 3-phase CCD

Charge transfer (in-)efficiencies (CTEs) due to electrostatic repulsion, thermal diffusion and fringing fields



https://en.wikipedia.org/wiki/Charge-coupled_device



http://solar.physics.montana.edu/nuggets/2000/001201/ccd.png

Charge Transfer Efficiency (CTE)

Time-dependent mechanisms that influence the CTE:

1. Electrostatic repulsion causes electrons to drift to the neighbouring electrode with time constant for charge transfer τ_{SI}.

2. Thermal diffusion drives electrons across the storage well at τ_{th}.

3. "Fringing fields" due to dependency of the well on the voltages of neighbouring electrodes (τ_{ff}).

Approximation for the CTE of a CCD with *m* phases: $$CTE = \left(1 - e^{-t/\tau}\right)^m$$

Noise from charge transfer inefficiency: ε = (1-CTE)

CCD Color Sensors

1. Three exposures through 3 filters – only works for fixed targets
2. Split input into 3 channels with separate filter and CCD
3. Bayer mask over CCD – each subset of 4 pixels has one filtered red, one blue, and two green