



**HAL**  
open science

## Semiconductor physics and devices

Charbel Tannous

► **To cite this version:**

Charbel Tannous. Semiconductor physics and devices. Master. Milieux Diélectriques et Magnétiques, UBO Brest, France. 2016, pp.40. <hal-04260375>

**HAL Id: hal-04260375**

**<https://hal.science/hal-04260375v1>**

Submitted on 26 Oct 2023


**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Semiconductor physics and devices

C. Tannous 

Université de Brest, Lab-STICC, CNRS-UMR 6285, F-29200 Brest, FRANCE

(Dated: [October 26, 2023](#))

Semiconductors are the workhorse of electronic technology that is constantly evolving with smaller devices, yet more powerful and cheaper. While metals were the workhorse of the 19th century, semiconductors are considered as the workhorse of the 20th and 21st centuries with an overlap of Photonics being felt stronger in the 21st century with optical fiber technology, Quantum Communication... Semiconductor physics is described in this work with a brief introduction to standard devices such as the p-n junction, transistor and Light-Emitting Diodes as well as novel devices belonging to Spintronics and Valleytronics.

PACS numbers: 72.80.Cw, 72.10.Bg, 68.35.bg

Keywords: Elemental semiconductors, General formulation of transport theory, Semiconductors

## Contents

<b>I. Introduction</b>	2
A. Band Structure	5
B. Gap evaluation	7
<b>II. Carrier Density in Semiconductors</b>	8
A. Intrinsic Semiconductors	8
B. Doped semiconductors	10
<b>III. Band bending and contacts between different materials</b>	12
A. Metal-Semiconductor Contacts	12
B. Metal-Insulator-Semiconductor Contacts	13
<b>IV. Carrier Transport</b>	14
A. Drift Current	15
B. Diffusion Current	16
C. Tunnel Current	16
D. Spin Current	17
<b>V. Generation and recombination processes</b>	19
A. Intrinsic Band-to-Band Generation-Recombination Processes	19
B. Extrinsic Shockley-Read-Hall Generation-Recombination	20
<b>VI. Electronic devices</b>	21
A. p-n junction	21
1. Space-charge region	22
2. Varicap Diode	23
3. Debye Length	23
4. Numerical treatment	24
B. Field-Effect Transistor	26
1. MOSFET with high- $\kappa$ dielectrics	26
2. Negative capacitance ferroelectric FET	27
3. Multigate devices	28
<b>VII. Magnetic, Optical, Spintronic, Valleytronic devices</b>	28
A. Hall effect devices	28
B. Light Emitting Diode	29
C. Laser Diode	32
D. Photo-detectors	35
E. Spintronic devices	36
F. Valleytronic devices	38
<b>A. Summary of Ge, Si and GaAs properties</b>	38
<b>References</b>	39

## I. INTRODUCTION

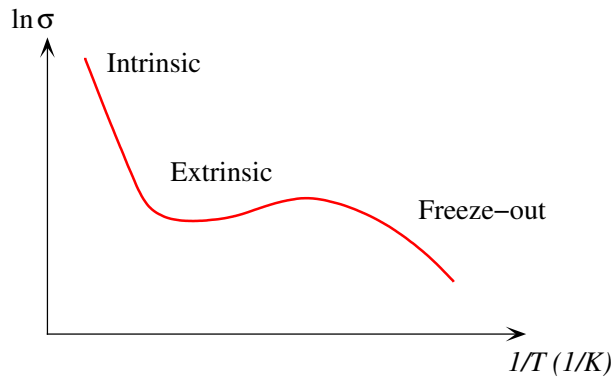
The term "semiconducting" was used by A. Volta for the first time in 1782. The first known observation of a semiconductor effect is that of M. Faraday (1833), who noticed the resistance decrease of silver sulfide with temperature that contrasted with the increase observed in metals.

In comparison with other solids, insulators have their conductivity  $\sigma_c \in [10^{-18}, 10^{-8}]$  S/cm, whereas semiconductors have  $\sigma_c \in [10^{-8}, 10^{+3}]$  S/cm, semi-metals like graphite, As, Sb and Bi or metals possess  $\sigma_c \in [10^{+3}, 10^{+8}]$  S/cm (1). S is Siemens or inverse Ohm ( $\Omega^{-1}$ ) conductance unit.

A typical metal at  $T = 300$  K has an electron density  $n_{metal} \sim 10^{23}/(\text{cm})^3$  whereas a typical semiconductor has  $n_{semi} \sim 10^{10}/(\text{cm})^3$ .

Thus the conductivity of an intrinsic (non-doped) semiconductor is very small and a semiconductor is insulating at low temperatures whereas its conductance becomes larger at higher temperature (doped and undoped) as in Fig. 1 in contrast to metals.

This provides another meaning to the prefix "semi": the material needs temperature or dopants in order to behave like a conductor.



**Fig.1:** Logarithm of conductivity versus temperature for a typical doped semiconductor. Intrinsic regime indicates behavior following carrier contribution as if free of dopants. At intermediate  $T$  we have extrinsic behavior with dopants contributing to variation of  $\sigma$ . Finally "frozen-out" regime indicates low temperature insulating behavior with no contribution of carriers to  $\sigma$ .

A semiconductor (SC) can be considered as a quantum two-level system with a lower energy set of states called the valence band (akin to the Quantum Chemistry LUMO or Lowest Unoccupied Molecular Orbital) and a higher energy set of states called the conduction band (akin to the Quantum Chemistry HOMO or Highest Occupied Molecular Orbital). The energy separating the valence from the conduction bands is the gap  $E_G$ , a measure of the relative number of excited electrons giving the ratio as  $\exp(-E_G/2k_B T)$  with  $k_B$  Boltzmann constant and  $T$  the temperature.

Electrons populate the conduction band (CB) and their absence called holes populate the valence band (VB). Holes are created in the VB when an electron is excited to travel across the gap from the VB to the CB or to some level inside the gap.

Semiconductors belong to column IV of the Periodic Table and combinations such as III-V and II-VI columns make special semiconductors with interesting optical properties (cf. Table.1).

Bonding in a semiconductor is based on strong covalent bonds involving two electrons per bond with a bonding energy of several electron-volts. Additionally the electrons are localized in the bond with a well-defined orientational behavior.

The crystal structure of Si and Ge is the diamond structure that is an FCC with origin at (000) and another one displaced by  $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ .

The crystal structure of binary semiconductors, such as GaAs or InP that are interesting for their optical properties, is the zinc-blende structure, which is the same structure for binary compounds with two different elements (cf Fig. 3) occupying alternating positions.

	Column				
Period	II	III	IV	V	VI
2		B	C	N	O
3	Mg	Al	Si	P	S
4	Zn	Ga	Ge	As	Se
5	Cd	In	Sn	Sb	Te
6	Hg		Pb		

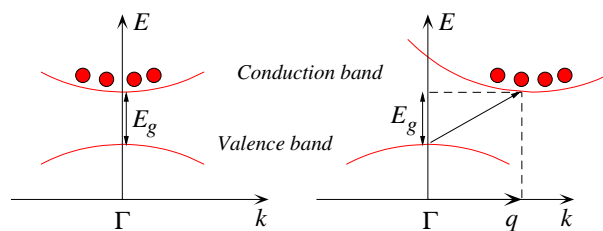
**Table 1:** Elements of the Periodic Table related to semiconductors. Adapted from Sze (1)

$E_G$	Excited electron fraction	Solid type
0 eV	1	Metal
0.25 eV	$10^{-2}$	SC
2 eV	$10^{-17}$	Largest $E_G$ for a SC
4 eV	$10^{-35}$	Insulator

**Table 2:** Effect of gap magnitude on the fraction of excited electrons. A solid is considered as a semiconductor (SC) when  $E_G \leq 2\text{eV}$ .

Crystal type	Gap type	$E_G$ (eV) [0K - 300 K]	Crystal type	Gap type	$E_G$ (eV) [0K- 300 K]
Diamond	i	5.4 5.4	AlSb	i	1.65 1.6
Si	i	1.17 1.11	PbS	d	0.286 0.34-0.37
Ge	i	0.744 0.66	PbSe	i	0.165 0.27
$\alpha$ -Sn	d	0.00 0.00	PbTe	i	0.190 0.29
InSb	d	0.23 0.17	CdS	d	2.582 2.42
InAs	d	0.43 0.36	CdSe	d	1.840 1.74
InP	d	1.42 1.27	CdTe	d	1.607 1.44
GaP	i	2.32 2.25	GaSb	d	0.81 0.68
GaAs	d	1.52 1.43	SnTe	d	0.03 0.18

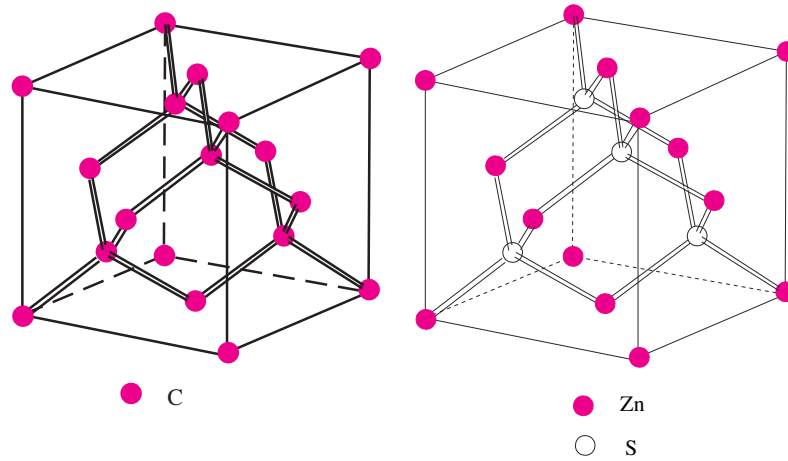
**Table 3:** Gap type (d for direct, i for indirect as depicted in Fig. 2) and its values in eV at 0K and 300 K for several semiconductors compared to insulating Diamond.



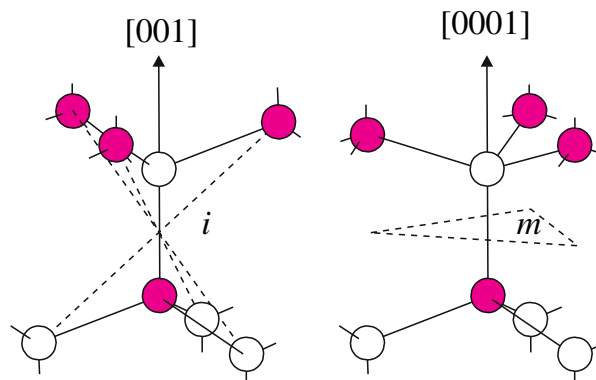
**Fig.2:** Direct (at left) with vertical electron transition and Indirect gap (at right) with diagonal electron transition requiring an excitation (such a phonon  $\mathbf{q}$ ) to compensate for momentum difference. An electron valley in the conduction band is indicated by red spots: it is at zone center ( $\Gamma$  point) in the direct gap case and away from it in the indirect gap case.

The Wurtzite structure belongs to the Hexagonal Bravais structure which is very important specially to optical properties and differs from zinc-blende as displayed in Fig.4. Miller indexing for the Hexagonal class is based on four numbers instead of three for all other Bravais classes. Index  $[hkilm]$  corresponds to  $[1/h, 1/k, 1/l, 1/m]$  intersections with the primitive  $a, b, c$  and  $z$  axes where  $z$  axis is perpendicular to the hexagonal cell plane.

From the technology side, semiconducting electronic devices follow Moore law that specifies doubling of performance



**Fig.3:** Diamond structure (at left) that describes Ge and Si and zinc-blende (at right) that is appropriate for GaAs. After Cardona *et al.* (2)



**Fig.4:** Comparison of the tetragonal bonds in the zinc-blende structure along  $[001]$  direction (at left) and wurtzite structure along  $[0001]$  direction (at right). A four index Miller notation is proper to the Hexagonal Bravais class to which Wurtzite belongs.  $i$  denotes an inversion center, whereas  $m$  a symmetry plane. After Grundman *et al.* (3)

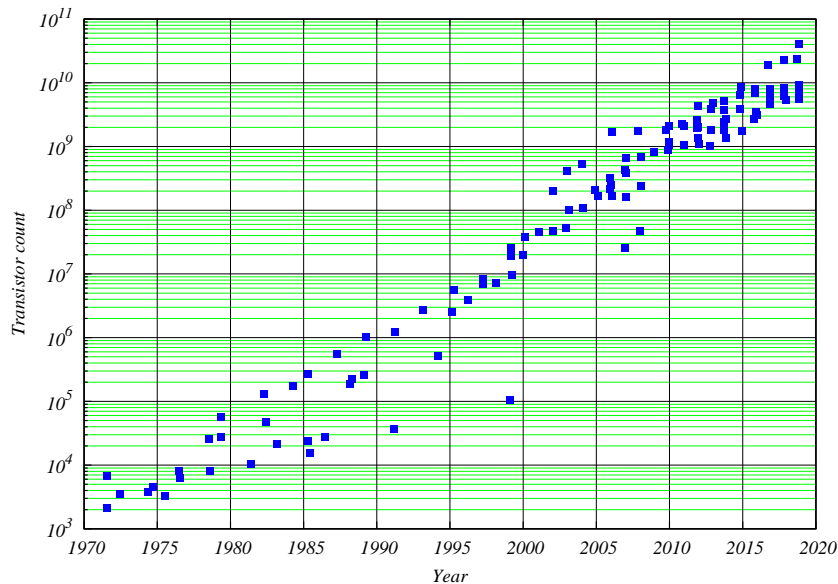
and cost decrease every eighteen months (corrected afterwards to two years) relies heavily on miniaturization since device mean size dictates its speed, energy consumption, response time...

A practical measure of miniaturization is the minimum feature (also called process node). It is the metric scale used by the chip foundry to control all sizes such as length, width and depth of various properties (gate, gate-oxide, contacts, carrier transport channels...) related to an individual device (transistor) fabrication or in making contacts between different devices (metal wire width and thickness) ...

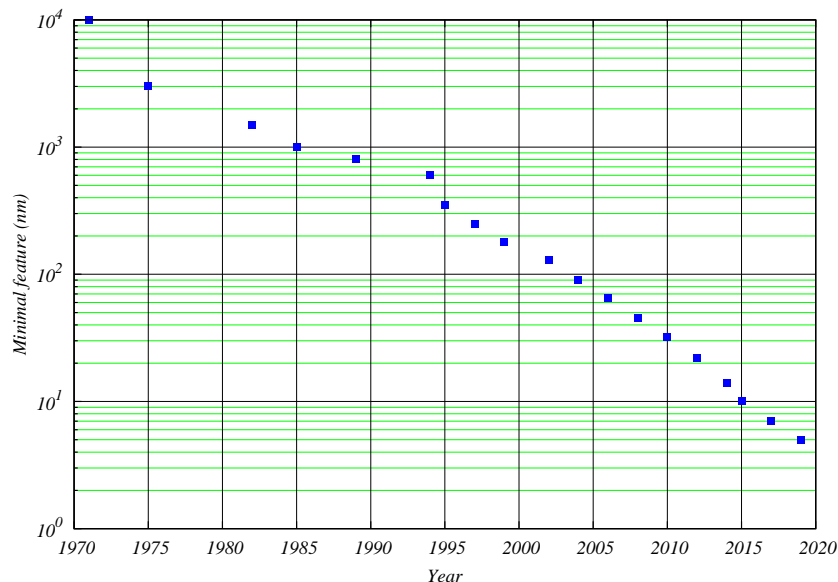
Its progress with time is depicted in Fig. 6.

Semiconductor physics is constantly evolving according to Moore law with devices getting smaller, yet more powerful while being more affordable price-wise. According to miniaturization or size scaling laws governing semiconductor devices, we are heading steadily toward the nm in terms of minimum feature that will be 2 nm in 2024. This means that the semi-classical Boltzmann approach adopted by semiconductor Physicists and Engineers is going to be no longer valid. Accordingly we must rather use Quantum Mechanics (QM) (4; 5) to adequately describe physical phenomena in these novel quantum devices.

Spin is another physical variable that ought to be considered leading to replacing traditional micro-electronics by nano-electronics and spintronics. We describe below basic semiconductor physics and pave the way leading to the introduction of spin current that will ultimately control spin-junction diodes, spin-Transistors, spin coupled light-emitting devices...



**Fig.5:** Moore law illustrated with transistor count in a CPU chip from the beginning of circuit integration circa 1970 to present times. The chips originate from Intel, Apple, AMD, Motorola, ARM and SUN Microsystems. Data gathered by Hannah Ritchie and Max Roser under CC-BY Licence.



**Fig.6:** (Color on line) Variation of process node (minimum feature) as an approximate exponential decrease with time. In 2022, 3 nm was attained and 2 nm is projected in 2024. Adapted from several chip foundries (Intel, AMD, IBM and Motorola).

## A. Band Structure

A semiconductor embodies an interacting electron ensemble where Heisenberg exchange interactions originating from wave-functions overlap along with Coulomb interactions between electrons. In contrast to a metal where electron concentration is large leading to screening, semiconductors with a smaller concentration of carriers do not benefit, in general, from this property.

Semiconductors possess a particular screening and are very different from their atomic constituents because of special effects such as orbital hybridization (6) and Pauli exclusion principle originating from electron spin-1/2

statistics and Coulomb interactions leading to many-body correlations (7).

Band structure calculations for semiconductors should account for exchange-correlation effects in solving the Schrödinger equation after using the Born-Oppenheimer approximation to obtain either the wave-function  $\Psi(\mathbf{r})$  or the Landau type electron density  $n(\mathbf{r})$  governing the many-electron system.

There are different treatments of exchange and correlation effects proper to different classes of semiconductors that are discussed further below.

Slater introduced the  $X - \alpha$  method where the exchange potential is modelled by a local potential of the form  $V_{ex} = \alpha[n(\mathbf{r})]^{1/3}$ , derived from the electron gas case and scaled by a constant  $\alpha$  to include correlations.  $n(\mathbf{r})$  is the local electron density (8).

This method, in spite of being semi-empirical, has been very successful in calculating ground-state properties and excitation states of many systems.

The  $X - \alpha$  theory may be considered as a precursor of density functional theory (DFT (9)) which has become a standard method for calculating ground-state properties of molecules and solids. Recent reviews of DFT may be found in Feiguin (9) or Kasper *et al.* (10) works.

In DFT, the ground-state energy is a functional of the ground-state density and satisfies the variational principle with respect to density that leads to a set of single-particle equations, the Kohn-Sham (KS) (11) equations, that should be solved self-consistently:

$$\left[-\frac{\hbar^2}{2m^*}\Delta + V_H + V_{xc}\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \quad n(\mathbf{r}) = \sum_i^{\text{occ}} |\psi_i(\mathbf{r})|^2 \quad (1)$$

where  $V_H$  and  $V_{xc}$  are respectively the Hartree (9; 11) and exchange-correlation potential replacing the simple Slater  $X - \alpha$  exchange potential. The above sum is over occupied (occ) states.

In practical applications, the functional containing the effects of exchange and correlations is approximated by the local density approximation (LDA) where the density in the exchange-correlation potential of the electron gas is replaced by the local density of the real system (11). The KS eigenvalues  $\varepsilon_i$  have no clear physical meaning except for the highest occupied which corresponds to the ionization energy. Although there is no theoretical justification, they are often interpreted as single-particle excitation energies corresponding to excitation spectra of the system upon a removal or addition of an electron.

In fact two distinct routes (9) for dealing with electronic structure calculations exist:

1. Wave-function  $\Psi(\mathbf{r})$  route:

After making the orbital approximation, one may select the LCAO method, Ab-initio Self-consistent field or other methods before treatment of correlations.

2. Density  $n(\mathbf{r})$  route:

After accounting for the Hohenberg-Kohn (9) theorems, one moves on to Density Functional Theory (DFT) which is based on Landau (12) Fermi Liquid Theory. The most representative method is Kohn-Sham work (11) leading to LDA (Local Density Approximation) methods inspired from Landau Fermi Liquid Theory.

Bands possess curvature to be interpreted as follows. In the free electron dispersion relation  $E(\mathbf{k}) = \frac{\hbar^2 k^2}{2m_e}$  it is possible to define the mass  $m_e$  from the energy  $E(\mathbf{k})$  by taking the second derivative with respect to momentum  $k$  such that:  $\frac{1}{m_e} = \frac{1}{\hbar^2} \frac{\partial^2 E(\mathbf{k})}{\partial k^2}$ .

Generalizing this approach to anisotropic semi-conducting crystals, we define an anisotropic effective mass depending on band indexed by  $\lambda$  as:

$$\frac{1}{m_{ij}^*(\lambda, \mathbf{k})} = \frac{1}{\hbar^2} \frac{\partial^2 E_\lambda(\mathbf{k})}{\partial k_i \partial k_j} \quad (2)$$

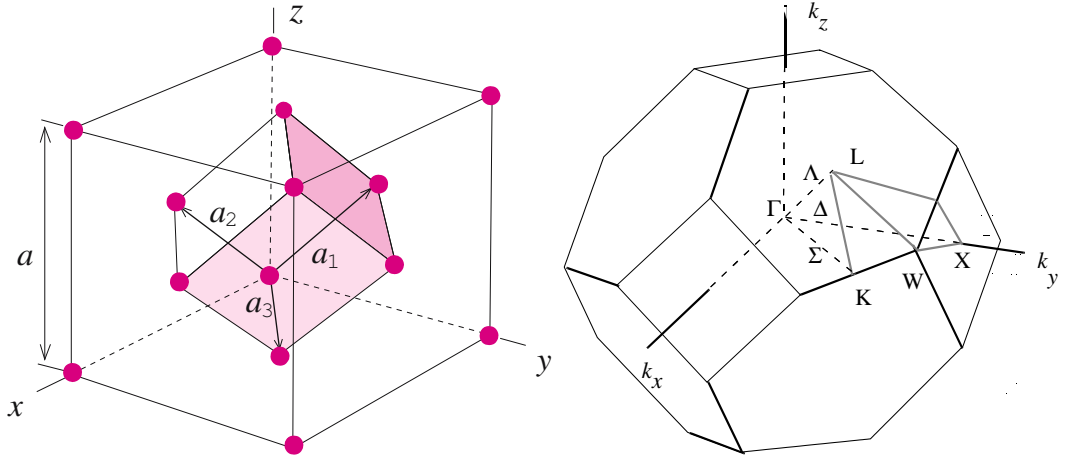


Fig.7: FCC cell (at left) and Brillouin zone (at right). After Cardona *et al.* (2)

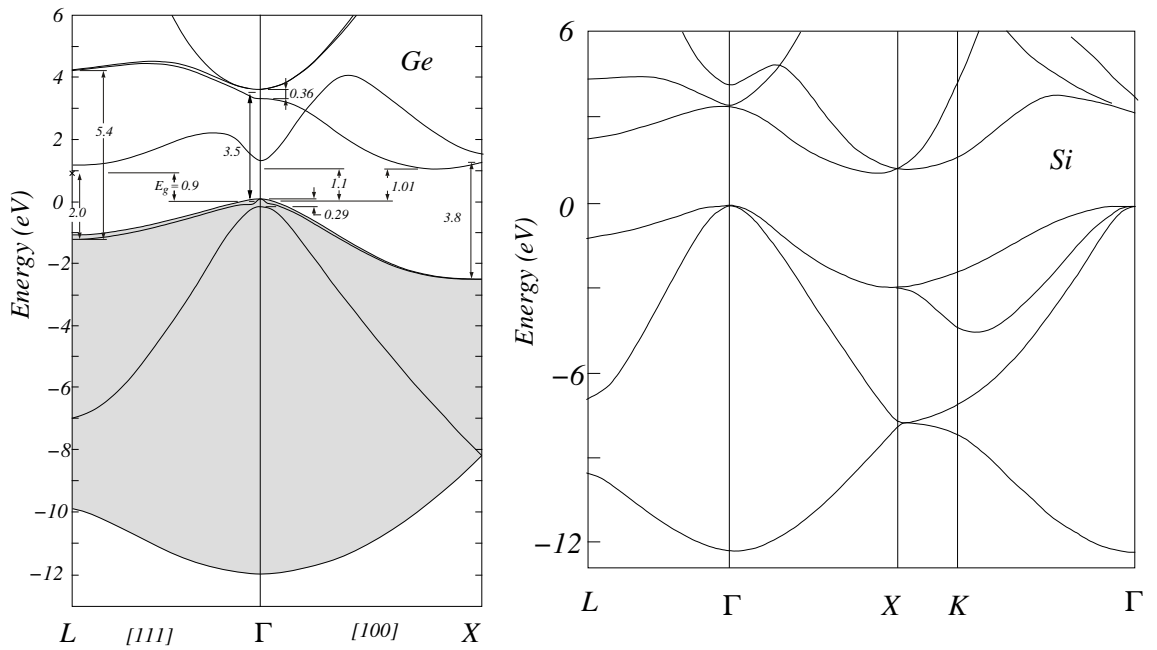


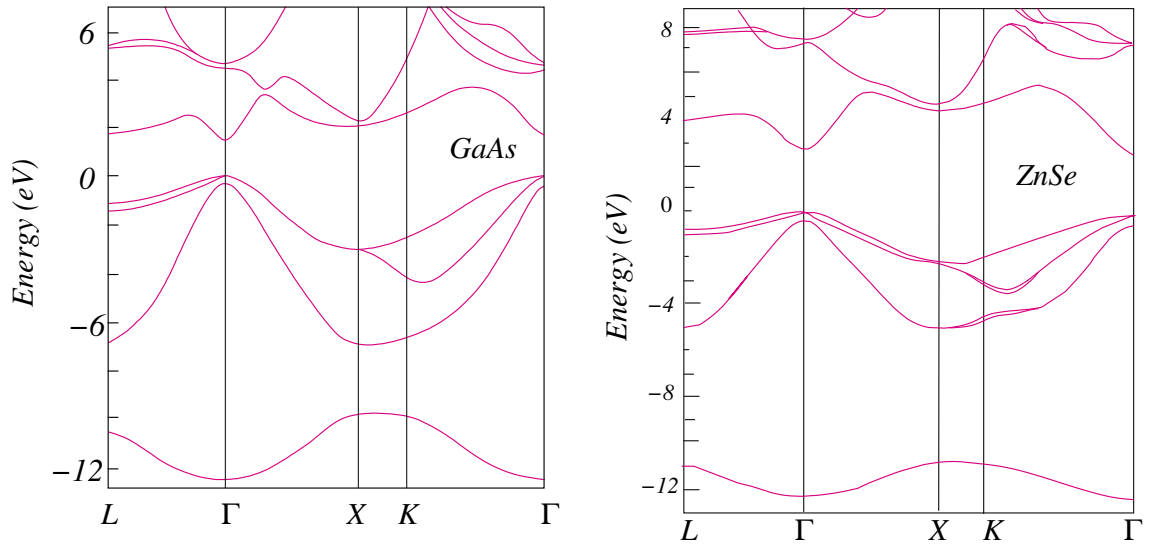
Fig.8: Indirect gap Germanium (at left) and Silicon bands (at right). All energies indicated for Ge are in eV. After Cardona *et al.* (2) and Kittel (6)

$E_\lambda(\mathbf{k})$  is a band specified with  $\lambda = n, \sigma$ , indicating band index and spin respectively.

This generalization of Newtonian mass to account for anisotropy would be extremely interesting since  $m_{ij}^*$  is not a scalar but a tensor depending on spatial coordinates, moreover it could be positive, negative and even zero (at Brillouin zone borders or in-band) leading to a potentially novel gravity theory endowed with gravitational screening akin to electromagnetic screening due to the existence of positive and negative mass (13).

## B. Gap evaluation

Despite the success of DFT and LDA, gap evaluation differed from experimental values. The GW method is an alternative to DFT and LDA for calculating ab-initio semiconductor band structure and estimating thereof the band gap value. LDA methods consistently under-estimate the gap value even obtaining a negative value for Ge and InAs



**Fig.9:** Direct gap GaAs (at left) and ZnSe bands (at right). After Cardona *et al.* (2) and Kittel (6)

Crystal	$\epsilon_r$	Crystal	$\epsilon_r$
Diamond	5.5	GaSb	15.69
Si	11.7	GaAs	13.13
Ge	15.8	AlAs	10.1
InSb	17.88	AlSb	10.3
InAs	14.55	SiC	10.2
InP	12.37	Cu <sub>2</sub> O	7.1

**Table 4:** Relative dielectric constants and gap values.

as displayed in Table. 5.

## II. CARRIER DENSITY IN SEMICONDUCTORS

### A. Intrinsic Semiconductors

Description of carrier densities in semiconductors is based on physical processes occurring on both sides of the gap: the valence band considered to be populated by  $p$  hole density or absent electrons and the conduction band considered to be populated by  $n$  electron density.

In the semi-classical two-fluid model, the semiconductor is viewed as containing two oppositely charged  $n, p$  fluids since electrons and holes coexist separately across the gap preventing them from colliding and mutually annihilating each other. Moreover both  $n, p$  contribute to total conductivity with the same sign despite their charges are opposite to each other.

In contrast to metals, carrier concentrations in semiconductor are highly  $T$ -dependent since all of the carriers in an intrinsic (undoped) material are thermally induced (i.e.  $n = p = 0$  at  $T=0$  K).

Electrons and holes are considered as classical following Boltzmann statistics (7; 14; 15) and this is true when  $n$  and  $p$  are small, meaning that Pauli principle may be ignored (non-degenerate case):

$$n = \int_{E_c}^{\infty} g_C(E) f(E) dE, \quad p = \int_{-\infty}^{E_v} g_V(E) \{1 - f(E)\} dE \quad (3)$$

where  $f(E)$  is Fermi-Dirac statistics distribution given by:  $f(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1}$ .

	LDA	GW	Exp.
Diamond	3.90	5.6, 5.33, 5.67	5.48
GaAs	0.67	1.58, 1.32, 1.22	1.52, 1.63
GaN (W)	2.3	3.5	3.5
GaN (ZB)	2.1	3.1	3.2, 3.3
GaP	1.82	2.55	2.39
Ge	<0	0.75, 0.65	0.744
InAs	-0.39	0.40	0.41
InP	0.57	1.44	1.42
Si	0.52	1.29, 1.24, 1.25	1.17
CdS (ZB)	1.37, 0.83	2.83, 2.45	2.55
CdS (W)	1.36	2.79	2.59
CdSe (ZB)	0.76	2.01	1.90
CdSe (W)	0.75	1.91	1.97
CdTe (ZB)	0.80	1.76	1.92
CdTe (W)	0.85	1.80	1.60
ZnS (ZB)	2.37	3.98	3.80
ZnS (W)	2.45	4.03	3.92
ZnSe (ZB)	1.45	2.84	2.96
ZnSe (W)	1.43	2.75	2.87

**Table 5:** Band gaps of semiconductors and insulators which have been calculated within the GW method and compared to LDA and experiment. The energy is in eV and when several values appear under LDA or GW, it means that exchange-correlation or other effects are handled differently within either method. W is for Wurtzite structure where ZB is for Zinc-Blende. Note that LDA gives a negative gap value for Ge and InAs.

Expressions for density of states  $g_C$  and  $g_V$  (conduction  $C$  and valence  $V$  bands) are needed in the following. Compared to the free electron case  $E_k = \frac{\hbar^2 \mathbf{k}^2}{2m^*}$  we have the density of states  $g(E) = \frac{(2m^*)^{\frac{3}{2}}}{2\pi^2 \hbar^3} \sqrt{E}$ . Thus,

$$g_C(E) = \frac{(2m_n^*)^{\frac{3}{2}}}{2\pi^2 \hbar^3} \sqrt{E - E_C}, \quad g_V(E) = \frac{(2m_p^*)^{\frac{3}{2}}}{2\pi^2 \hbar^3} \sqrt{E_V - E} \quad (4)$$

for  $E > E_C$  and  $E < E_V$  respectively, and zero otherwise when  $E_V < E < E_C$ .

In an intrinsic semiconductor, neutrality gives  $n = p$ , and the Fermi level  $E_F$  must lie in the band gap.  $m_n, m_p^*$  are electron and hole effective masses respectively.

If  $m_n^* \neq m_p^*$  (ie.  $g_C \neq g_V$ ), then  $E_F$ , must be adjusted with respect to the gap center such that  $n = p$ .

Furthermore, the carriers which are induced across the gap are somehow highly energetic since typically  $E_G = E_C - E_V \gg k_B T$  and  $\frac{1eV}{k_B} \sim 10000$  K.

From Boltzmann statistics,  $E - E_F \gtrsim \frac{E_G}{2} \gg k_B T$  and  $\frac{1}{e^{(E-E_F)/k_B T} + 1} \sim e^{-(E-E_F)/k_B T}$ .

A similar relationship exists for holes:  $1 - \frac{1}{e^{(E-E_F)/k_B T} + 1} \sim e^{-(E-E_F)/k_B T}$  since  $(1 - f(E)) = f(-E)$  and  $e^{(E-E_F)/k_B T}$  is small.

Consequently, the concentration of electrons  $n$

$$n \approx \frac{(2m_n^*)^{\frac{3}{2}}}{2\pi^2 \hbar^3} e^{E_F/k_B T} \int_{E_C}^{\infty} \sqrt{E - E_C} e^{-E/k_B T} dE \quad (5)$$

Performing a change of variable  $x = E/k_B T = \beta E$  we get:

$$n = \frac{(2m_n^*)^{\frac{3}{2}}}{2\pi^2 \hbar^3} (k_B T)^{\frac{3}{2}} e^{-\beta(E_C - E_F)} \int_0^{\infty} x^{\frac{1}{2}} e^{-x} dx \quad (6)$$

Thus

$$n = 2 \left( \frac{2\pi m_n^* k_B T}{h^2} \right)^{\frac{3}{2}} e^{-\beta(E_C - E_F)} = N_f^C e^{-\beta(E_C - E_F)} \quad (7)$$

Similarly (16; 17)

$$p = 2 \left( \frac{2\pi m_p^* k_B T}{h^2} \right)^{\frac{3}{2}} e^{-\beta(E_V - E_F)} = N_f^V e^{-\beta(E_V - E_F)} \quad (8)$$

where  $N_f^C$  and  $N_f^V$  are temperature-dependent effective densities of states for a 3D classical gas.

In general, in the non-degenerate limit (18),

$$np = 4 \left( \frac{k_B T}{2\pi\hbar^2} \right)^3 (m_n^* m_p^*)^{\frac{3}{2}} e^{-\beta E_G} \quad (9)$$

the mass action law is valid for both doped and intrinsic semiconductor as long as they remain in the non-degenerate limit. However, for an intrinsic semiconductor, where  $n = p$ , we get:

$$n_i = p_i = 2 \left( \frac{k_B T}{2\pi\hbar^2} \right)^{\frac{3}{2}} (m_n^* m_p^*)^{\frac{3}{4}} e^{-\beta E_G/2} \quad (10)$$

We give below some expressions relating  $n$  and  $p$  that involve  $E_C$  and  $E_V$

$$n = p = N_f^C e^{-\beta(E_C - E_F)} = N_f^V e^{\beta(E_V - E_F)} \quad (11)$$

Thus:

$$E_F = \frac{1}{2}(E_V + E_C) + \frac{1}{2}k_B T \ln \left( \frac{N_f^V}{N_f^C} \right) = \frac{1}{2}(E_V + E_C) + \frac{3}{4}k_B T \ln \left( \frac{m_p^*}{m_n^*} \right) \quad (12)$$

Thus if  $m_p^* \neq m_n^*$ ,  $E_F$  is temperature dependent.

However, when  $p$  and  $n$  are controlled in some device by external conditions, the system might be pushed far from thermal equilibrium. The variation of the intrinsic carrier concentration with temperature might be caused by: A variation of the carrier effective mass, the  $T^{3/2}$  pre-exponential term, a variation of the band gap  $E_G$  or the  $k_B T$  factor in the exponential function denominator.

Variation of the carrier effective mass for small temperature change can be neglected whereas the band gap variation for small temperature change can be described by a linear function (9).

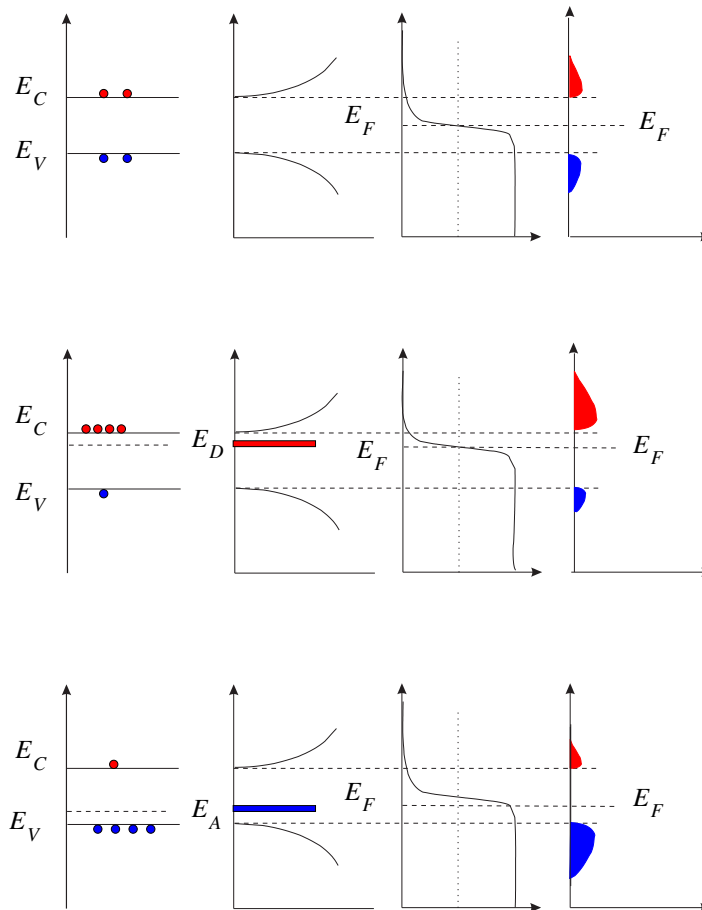
## B. Doped semiconductors

In order to increase  $n$  (or  $p$ ) to a density  $\sim 10^{18}/\text{cm}^3$  or larger, dopants are used.

A semiconductor having an electron concentration higher than the hole one is called n-type (extrinsic) semiconductor. Electron density in Si and Ge (column IV elements) can be increased by doping with column V elements, such as Phosphorus (P) and Arsenic (As) whereas p-doping is from column III elements like Boron (B) according to Table. 1.

A dopant should occupy an interstitial site normally occupied by a normal atom of the virgin crystal and should not disturb the local environment by its size or chemical interactions that should be limited to accepting or giving an electron.

In a p-type (extrinsic) semiconductor, hole density is higher than the electron one. Donor dopants introduce energy states within the energy gap, very close to the conduction band whereas acceptor dopants introduce energy states within the energy gap, very close to the valence band (cf. Fig. 10). Since ionization energies of typical donors and acceptors are rather small (10 to 50 meV), at room temperature (For  $T = 300$  K,  $k_B T = 25$  meV) almost all



**Fig.10:** (Color on-line) Upper: Intrinsic band structure with electron  $n$  (in red) and hole (in blue) density  $p$ . Middle: n-doped semiconductor with donor level  $E_D$  close to the conduction band with augmented  $n$  density. Lower: p-doped semiconductor with acceptor level  $E_A$  close to the valence band with augmented  $p$  density. In all cases we have  $np = n_i^2$ . Adapted from Sze (1).

impurities are ionized.

For example, Phosphorus or Boron will either donate or absorb additional electron (with the latter called hole creation). These additional charges will be localized around the donor or acceptor ion. The difference is that the donor (acceptor) is considered localized as if having an infinite mass. Accordingly, the binding energy is quantized and given by:

$$E_\ell = \frac{m^* e^4}{2\epsilon_r^2 \hbar^2 \ell^2}, \quad \ell \in \mathcal{N} \quad (13)$$

where  $m^*$  is electron donor or hole acceptor mass and  $\epsilon_r$  is the relative dielectric constant.

Typically  $\frac{m^*}{m} < 1$  and  $\epsilon_r \sim 10$  implying that binding energies are often much less than Hydrogen ground state value (13.6 eV). Consequently thermal excitations will often ionize dopant sites.

The law of "mass action" is valid when Boltzmann statistics is also valid i.e. if degeneracy is small. Thus, for an intrinsic or doped semiconductor we have:

$$np = N_f^C N_f^V e^{-\beta E_G} = n_i^2 = p_i^2 \quad (14)$$

The law of "mass action" does not apply when we are in the injection regime  $np > n_i^2$  or depletion regime  $np < n_i^2$ .

At thermodynamic equilibrium, the semiconductor is electrically neutral, thus:

$$n + N_A^- = p + N_D^+ \quad (15)$$

The probability that a donor (acceptor) is occupied by an electron is determined by Fermi-Dirac statistics:

$$n_D = N_D^0 = N_D \frac{1}{1 + e^{\beta(E_D - E_F)}} \quad (16)$$

$$p_A = N_A^0 = N_A(1 - f(E_A)) = N_A \frac{1}{1 + e^{\beta(E_F - E_A)}} \quad (17)$$

In order to tackle an analytically solvable case, suppose that we have an  $n$ -type semiconductor (without any  $p$ -type dopants) so that  $N_A = N_A^0 = N_A^+ = 0$ , then

$$n = N_f^C e^{-\beta(E_C - E_F)}, \quad N_D = N_D^0 + N_D^+ N_D^0 = N_D \frac{1}{e^{\beta(E_D - E_F)} + 1} \quad (18)$$

Furthermore, charge neutrality requires that

$$n = p + N_D^+ \quad (19)$$

Assume that for a doped semiconductor

$$N_D^+ \gg n_i \quad (20)$$

ie., many more carriers are provided by dopants than are thermally excited over the entire gap, then as  $np = n_i^2$ , we have  $N_D^+ \gg p$  such that

$$n \approx N_D^+ = N_D - N_D^0 \approx N_D \left( 1 - \frac{1}{e^{\beta(E_D - E_F)} + 1} \right) \quad (21)$$

Thermally induced carriers satisfy Boltzmann statistics:

$$n = N_f^C e^{\beta(E_F - E_C)} \quad (22)$$

thus we can eliminate  $E_F$  in  $n$

$$n = \frac{N_D}{1 + e^{\beta E_D} n / (N_f^C)} \quad (23)$$

This quadratic equation has only one acceptable solution

$$n = \frac{2N_D}{1 + \sqrt{1 + 4 \left( N_D / N_f^C \right) e^{\beta E_D}}} \quad (24)$$

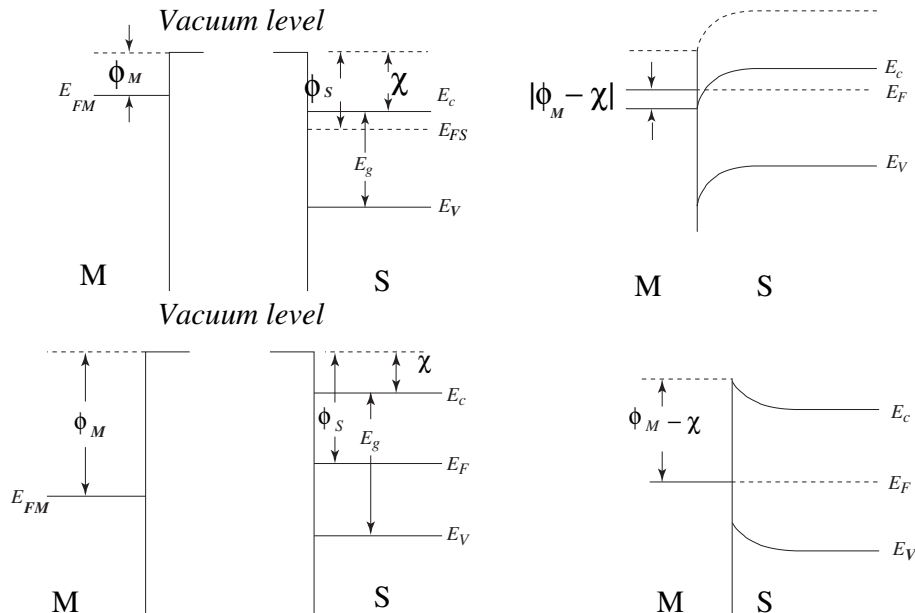
At low  $T \ll \frac{E_D}{k_B}$ ,  $\approx \sqrt{N_D N_f^C} e^{-\beta E_D}$ , at high  $T \gg \frac{E_D}{k_B}$ ,  $n = N_D$ , and at higher  $T$  the approximation breaks down so that  $N_D^+ \gg n$  since thermally excited carriers will dominate.

### III. BAND BENDING AND CONTACTS BETWEEN DIFFERENT MATERIALS

#### A. Metal-Semiconductor Contacts

When a semiconductor is to be used, metallic contacts are required in order to apply voltage, inject or collect a current... Metal-semiconductor contacts are very important to understand since they display a range of behavior from strongly blocking (rectifying) to ohmic with each regime having its own applications.

A simplified band diagram of a metal and a  $n$ -type semiconductor, before (left) and after (right) contact is displayed in Fig. 11 considering two different cases:  $\phi_M < \chi$  and  $\phi_M > \chi$  where  $\phi_M$  is the metal work-function (energy difference



**Fig.11:** Upper panel: At left, Metal (M) and semiconductor (S) before contact with work function  $\phi_M < \chi$ , the semiconductor affinity. At right, after joining M and S, we have an Ohmic contact. S bands are down bending facilitating carrier movement across S-M interface. Lower panel: Metal and n-type S before contact with work-function  $\phi_M > \chi$ . At right, after joining M and S, we have blocking contact with S bands up-bending creating a barrier across S-M interface. Adapted from Sze (1).

from Fermi to vacuum level) and  $\chi$  is the semiconductor affinity (energy difference from conduction band bottom to vacuum level).

Assuming the metal work-function  $\phi_M < \chi$  the semiconductor affinity, when M and S (cf Fig. 11) are brought into intimate contact electrons travel from the conduction band of the semiconductor into the metal until the Fermi levels equalize reaching thermal equilibrium (cf Fig. 11). M and S have an Ohmic contact. S bands are down bending facilitating carrier movement across S-M interface.

When we are in the  $\phi_M > \chi$  case, with M and S (cf Fig. 11) brought into intimate contact, equalizing Fermi levels to reach thermal equilibrium (cf Fig. 11) makes S bands bend upwards creating a barrier across S-M interface. M and S have a blocking contact with a barrier making an obstacle to carrier crossing of the interface.

In general, the barrier height created by the contact is related to the work-function of the metal, the affinity and resistivity of the semiconductor, barrier reduction due to an image force and the nature and density of semiconductor surface states (1). A high density of surface states, as encountered in Si, Ge and GaAs can effectively maintain the value of the barrier height and make it completely independent of the metal work-function. Barrier heights are also relatively insensitive to doping level provided it is below the degenerate limit  $10^{17} \text{ cm}^{-3}$ .

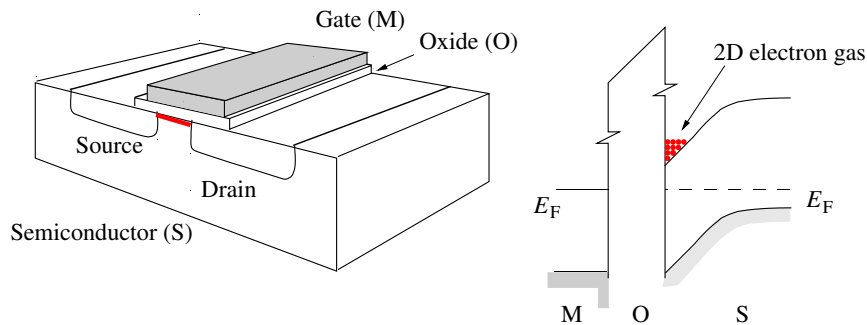
## B. Metal-Insulator-Semiconductor Contacts

The best example of a metal-insulator-semiconductor contact is the MOSFET, the basic component of any micro-electronic circuit.

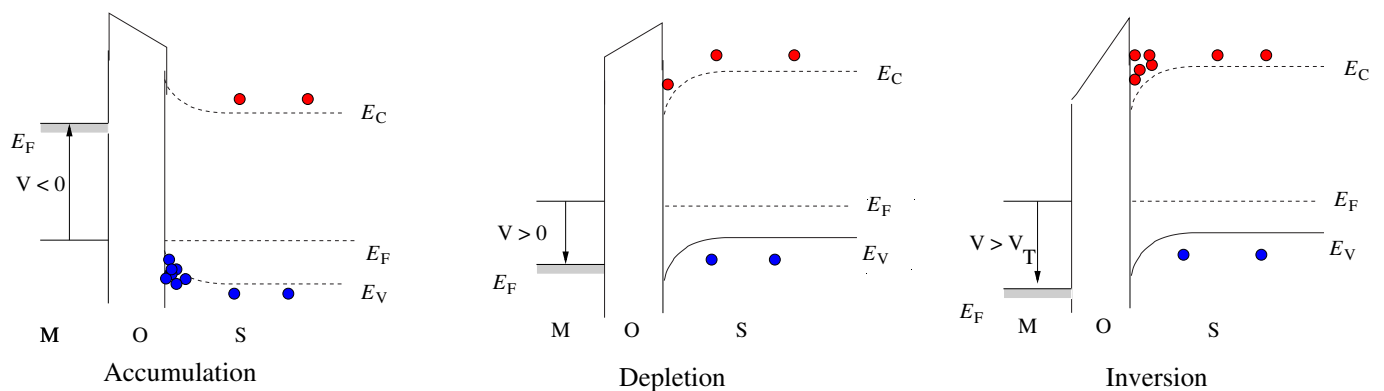
A MOSFET is a Field-Effect Transistor (FET) based on the MOS structure with the metal (M) part acting like a grid separated from the semiconductor (S) by an insulating oxide (O) controlling the source-drain current running through a channel formed by the trapped 2D electron gas at the oxide-semiconductor interface when the voltage  $V_G > 0$  is applied to the metallic gate (cf. Fig. 12). The channel is wiped out by a voltage  $V_G < 0$ .

In order to understand the formation of a 2D electron gas and band bending of the different materials involved, one has to consider a set of joining modes encountered typically in the microelectronics industry as displayed in Fig. 13.

The MOSFET, Flash transistor and other components of the microelectronics industry are explained in more detail further below.



**Fig.12:** (Color on line) At left, a MOSFET device with an electron channel (in red) running between source and drain. At right, the channel is a 2D electron gas (in red) trapped at the oxide-semiconductor interface (OSI) when the voltage  $V_G > 0$  is applied to the metallic gate. Transistor switch action is based on  $V_G$  sign: When  $V_G > 0$  electrons are attracted to the OSI and the MOSFET is ON and when  $V_G < 0$  the MOSFET is OFF since the electrons are repelled by  $V_G < 0$ .



**Fig.13:** Band bending modes for a p-type semiconductor based MOS structure. There is charge accumulation (left) for a negative applied voltage  $V$ , depletion (middle) for a small  $V > 0$  and inversion (right) for a large  $V > V_T$  the threshold voltage. Accumulation and depletion correspond to breakdown of law of mass action (respectively  $np > n_i^2$ ,  $np < n_i^2$ ). Inversion means we have electrons trapped on a p-type semiconductor surface. Note that the Fermi levels in the Metal and Semiconductor are constant with absence of band bending in the Metal and its presence in the Semiconductor. Adapted from Sze (1).

#### IV. CARRIER TRANSPORT

Current due to an electrical potential gradient is called drift current while current due to a particle density gradient is a diffusion current.

Electrons or holes produce currents when we have:

1. Electric potential gradient  $dV/dx$ ,
2. Particle density gradient  $dn/dx$ ,
3. Temperature gradient  $dT/dx$

Other currents such as tunnel, photo-induced, spin... triggered by other means (quantum effects, shining of light, magnetic field...) are discussed further below.

## A. Drift Current

P. Drude in the 1900 proposed a classical model for a free electron that predicts, thermal, electric, and optical properties of solids. Drude applied kinetic theory of gases to electrons in metals leading to description of DC and AC conductivity, Hall-effect and magneto-resistance phenomena.

However, the specific heat of metals was overestimated and the issue has been tackled, later on with the advent of Quantum Mechanics. First of all, Ehrenfest theorem justified classical single electron treatment through building quantum packets obeying classical equations of motion and Sommerfeld replaced Maxwell-Boltzmann distribution with Fermi-Dirac statistics to account for Pauli Spin-Statistics theorem.

We assume that an electron (of charge  $q = -e$  and mass  $m_e$ ) is subjected to an electric field and its Newtonian equation of motion is:

$$m_e \ddot{\mathbf{r}} = -e\mathbf{E} \quad (25)$$

Integrating once, we have:  $\mathbf{v} = \dot{\mathbf{r}} = -e\mathbf{E}t/m_e$ .

This implies that velocity increases linearly until a collision happens with other electrons impurities or metal boundaries. Thus the velocity is reset to zero and re-increases linearly for a typical time  $\tau_c$  until another collision event resets it. The average length traveled by the free electron is the mean free path  $\ell$ .

Thus we get a saw-tooth type of motion of the electron with an average velocity between two successive collisions given by:

$$\bar{v} = \frac{1}{\tau_c} \int_0^{\tau_c} dt eEt/m_e = eE\tau_c/(2m_e) \quad (26)$$

Drude suggested another mechanism which produces a smoother motion and that is a damping mechanism as a friction term proportional to velocity such that the equation of motion writes:

$$m_e \dot{\mathbf{v}} = -e\mathbf{E} - \frac{m_e \mathbf{v}}{\tau} \quad (27)$$

$\tau$  is a relaxation time responsible for damping given by  $\tau = \ell/\bar{v}$ .

Integrating we have:

$$\mathbf{v} = \frac{e\mathbf{E}\tau}{m_e} + \mathbf{v}_0 \exp(-t/\tau) \quad (28)$$

Thus, when  $t \gg \tau$ ,  $\mathbf{v} \rightarrow \mathbf{v}_\infty = \frac{e\mathbf{E}\tau}{m_e}$  resulting in a current density  $\mathbf{J} = \frac{n_e e^2 \mathbf{E} \tau}{m_e}$  for an electron density  $n_e$ . The resulting Drude conductivity  $\sigma_D$  as in  $\mathbf{J} = \sigma_D \mathbf{E}$  is  $\sigma_D = \frac{n_e e^2 \tau}{m_e}$ .

In metals, carrier densities vary somehow weakly with temperature whereas in semiconductors they vary tremendously with temperature and doping. Thus we separate the contribution of carrier densities and transport mechanisms by introducing carrier mobility.

Mobility is introduced as a coefficient relating velocity and electric field, thus  $\mathbf{v}_\infty = \frac{e\mathbf{E}\tau_e}{m_e} = \mu_e \mathbf{E}$  where  $\mu_e = \frac{e\tau_e}{m_e}$ .

It follows that in an isotropic material, the mobility  $\mu$  is a scalar coefficient relating  $\mathbf{v}$  and  $\mathbf{E}$  whereas in the anisotropic case, the coefficient becomes a tensor  $\mu_{ij}$  such that  $\mathbf{v}_i = \mu_{ij} E_j$  according to Einstein summation rule.

$\mu$  depends on temperature and doping density especially when it is high. It also depends on the electric field and it fails at high fields. Consequently there is a limiting field to validate a mobility description.

In the Si case, the maximum drift velocity is  $10^5$  m/s for both electrons and holes. The field needed for electrons to reach this speed is about  $2 \times 10^6$  V/m, whereas holes require fields above  $10^7$  V/m.

Crystal	Electrons	Holes	Crystal	Electrons	Holes
Diamond	1800	1200	GaAs	8000	300
Si	1350	480	GaSb	5000	1000
Ge	3600	1800	PbS	550	600
InSb	800	450	PbSe	1020	930
InAs	30000	450	PbTe	2500	1000
InP	4500	100	AgCl	50	–
AlAs	280	–	AlSb	900	400

**Table 6:** Electron and hole mobilities in several semiconductors at 300 K in  $\text{cm}^2/\text{V.s}$ .

The total conductivity  $\sigma$  of an isotropic sample may be defined as the current carrying both electrons (with mobility  $\mu_e$ ) and holes (with mobility  $\mu_p$ ):

$$\sigma = ne\mu_e + pe\mu_p \quad (29)$$

with  $e$  the electronic charge.

### B. Diffusion Current

In 1D, Fick law relates a variation of carrier concentration  $n(x)$  (or temperature) to a flow of particles given by a current proportional to minus the concentration gradient:

$$J = -eDdn(x)/dx \quad (30)$$

with  $D$  is the diffusion coefficient and  $e$  the carrier charge.

When considering the diffusion of holes and electrons in the same density gradient, holes and electron fluxes are along the same direction, canceling thereby the conventional electrical current. Nevertheless,  $D$  and  $\mu$  are related at temperature  $T$  by Einstein relation:

$$D = \mu k_B T / e \quad (31)$$

that can be derived from nulling the total current  $J_T$  given by the sum of the drift and diffusion currents at equilibrium. Thus:

$$J_T = -ne\mu E - eD(dn/dx) \equiv 0 \quad (32)$$

Considering a Boltzmann distribution for  $n(x)$  at temperature  $T$ , we have:  $n(x) = n_0 \exp(-eV(x)/k_B T)$  where  $V(x)$  is an applied potential whose gradient is  $E = -dV/dx$  yielding the drift current  $-ne\mu E$ . Given that  $(dn/dx) = -(en_0/k_B T) \exp(-eV(x)/k_B T)(dV/dx) = eEn/k_B T$ , we get finally Einstein result  $D = \mu k_B T / e$ .

Both,  $D$  and  $\mu$  are impacted by the various collision processes (defects, impurities, phonons, dislocations...).

### C. Tunnel Current

Current mechanisms through insulating materials (generally oxides) are important to understand since insulators are used in many devices such as Diodes, MOSFET...

We describe below Fowler-Nordheim tunneling that has been studied extensively in MOS devices where it has been shown to be the dominant current mechanism in thick oxides and how it is applied to Flash technology where we

have both thick and thin oxides...

A Flash transistor used in USB keys and Solid State Drives (SSD) is derived from the MOSFET by including a thin metallic layer called Floating Gate (FG) very close to the semiconductor separated by a thin oxide (TO) about an Angström thick (cf. Fig. 14).

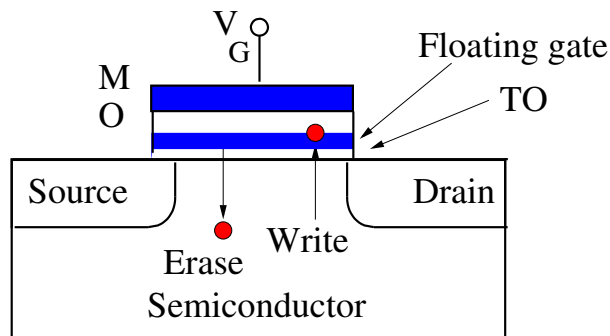
The FG is used to store electrons whose number correspond to digital bits of information and writing or erasing information is done through the application of a voltage  $V_G$  to the metallic gate. Writing consists of applying  $V_G > 0$  attracting electrons to cross the TO from the semiconductor to the FG. Erasing is when  $V_G < 0$  repelling FG electrons to cross the TO back to the semiconductor.

Quantum mechanics is needed to describe tunneling from the adjacent conductor into the thin oxide (TO) insulator limiting the current through the structure (cf. Fig. 14).

Current evaluation through the oxide based on the WKB approximation (4; 5) yields the following relation between the current density,  $J_{FN}$ , and the electric field in the TO,  $E_{TO}$

$$J_{FN} \sim E_{TO}^2 \exp\left(-\frac{4}{3} \frac{\sqrt{2m_{TO}^*} [e\phi_B]^{3/2}}{e\hbar E_{TO}}\right) \quad (33)$$

which is of the form  $J_{FN} \sim V^2 \exp\left(-\frac{A}{V}\right)$  where  $V \propto E_{TO}$ . In order to appreciate this J-V dependence, we recall the Ohmic case(1) as  $J \sim V \exp\left(-\frac{A}{k_B T}\right)$   $\phi_B$  is the barrier height at the conductor-insulator interface, for electron tunneling from n-type doped silicon into the  $\text{SiO}_2$  material.



**Fig.14:** Flash transistor derived from the MOSFET by including a thin metallic layer called Floating Gate (FG) very close to the semiconductor separated by a thin oxide (TO) about an Angström thick. The gate voltage  $V_G$  controls electron tunneling across the insulating layer between the FG and the semiconductor. Writing consists of applying  $V_G > 0$  attracting electrons (in red) to cross the TO from the semiconductor to the FG. Erasing is when  $V_G < 0$  repelling FG electrons (in red) to cross the TO back to the semiconductor.

In general Fowler-Nordheim tunneling time is somehow long, nevertheless several methods have been developed in order to shorten it (2; 3), specially in SSD (Solid State Drives) that are now competing with magnetic HDD (Hard Disk Drives).

#### D. Spin Current

Spintronics require procedures to generate, control and manipulate a spin current through its various applications (19; 20). In the ideal situation where spin (or its projection along a direction) is conserved, spin current is simply defined as the difference of electron currents between the spin  $\uparrow$  and  $\downarrow$  states.

Spin-orbit coupling inevitably makes the spin non-conserved, however the work-around is to adopt a spin relaxation time approximation in the framework of Boltzmann semi-classical transport theory (21).

Applying Boltzmann approach to spin-polarized systems consists of treating spin  $\uparrow$  and  $\downarrow$  carriers as two distinct fluids carrying two currents allowing to draw analogies with the semiconductor  $n, p$  two-fluid model (22).

Evolving from the semi-classical two-current (or two-fluid) approach (21) to a fully quantum description entails treating the spin density for a particle in a quantum spinor (5) state given by:

$$\Psi(\mathbf{r}) = \begin{bmatrix} \psi_{\uparrow}(\mathbf{r}) \\ \psi_{\downarrow}(\mathbf{r}) \end{bmatrix} \quad (34)$$

Selecting for definiteness, the spin  $z$  component with the corresponding spin operator  $\hat{s}_z$ , we have the spin density  $S_z(\mathbf{r})$  given by the expectation value in the spinor state  $\Psi(\mathbf{r})$ ,  $S_z(\mathbf{r}) = \Psi^\dagger(\mathbf{r})\hat{s}_z\Psi(\mathbf{r})$ .

The spin current density is given by the conventional definition (22)  $\mathbf{J}_s(\mathbf{r}) = \text{Re}[\Psi^\dagger(\mathbf{r})\frac{1}{2}\{\hat{\mathbf{v}}, \hat{s}_z\}\Psi(\mathbf{r})]$ , where  $\hat{\mathbf{v}}$  is the velocity operator, and  $\{, \}$  denotes the anticommutator.

With quantum mechanics (4; 5), one can derive a continuity equation relating the spin, spin-current and spin-torque densities as follows:

$$\frac{\partial S_z}{\partial t} + \nabla \cdot \mathbf{J}_s = \mathcal{T}_z. \quad (35)$$

The right hand side of the continuity equation is the torque density defined by  $\mathcal{T}_z(\mathbf{r}) = \text{Re}[\Psi^\dagger(\mathbf{r})\hat{\tau}\Psi(\mathbf{r})]$ . The spin-torque operator  $\hat{\tau}$  is defined by analogy with classical angular momentum time evolution  $\hat{\tau} = \frac{d\hat{s}_z}{dt}$ .

Applying Heisenberg time evolution (4; 5) to  $\hat{s}_z$  spin component operator, we write:

$$\hat{\tau} = \frac{d\hat{s}_z}{dt} = \frac{1}{i\hbar}[\hat{s}_z, \hat{H}] \quad (36)$$

where  $\hat{H}$  is the Hamiltonian of the system.

In some cases, due to symmetry reasons (22), the average torque may vanish for the bulk of the system, i.e.,  $(1/V)\int dV\mathcal{T}_z(\mathbf{r}) = 0$ .

Torque density may be written as a divergence of a Torque Dipole Density (TDD) (23),

$$\mathcal{T}_z(\mathbf{r}) = -\nabla \cdot \mathbf{P}_\tau(\mathbf{r}). \quad (37)$$

Moving this expression to the left hand side of (35), we have

$$\frac{\partial S_z}{\partial t} + \nabla \cdot (\mathbf{J}_s + \mathbf{P}_\tau) = 0, \quad (38)$$

This agrees with the expression of a standard source-less continuity equation (23).

Therefore, in the bulk of systems where the average torque vanishes, the transport of spin  $S_z$  is governed by the spin current density:

$$\mathcal{J}_s = \mathbf{J}_s + \mathbf{P}_\tau. \quad (39)$$

A problem with this approach is that eq. (37) does not uniquely determine the TDD  $\mathbf{P}_\tau$  from the corresponding torque density  $\mathcal{T}_z$ .

This ambiguity may be resolved by imposing physically that the TDD being a material property should vanish outside the sample.

This leads to  $\int dV \mathbf{P}_\tau = -\int dV \mathbf{r} \nabla \cdot \mathbf{P}_\tau = \int dV \mathbf{r} \mathcal{T}_z(\mathbf{r})$ .

Consequently, after performing bulk average, the effective spin current (23) density can be written as:

$$\mathcal{J}_s = \text{Re}[\Psi^\dagger(\mathbf{r})\hat{\mathcal{J}}_s\Psi(\mathbf{r})], \quad (40)$$

where

$$\hat{\mathcal{J}}_s = \frac{d(\hat{\mathbf{r}}\hat{s}_z)}{dt} \quad (41)$$

is the effective spin current operator. Compared to the conventional spin current operator, it has an additional term  $\hat{\mathbf{r}}(d\hat{s}_z/dt)$ , that accounts for spin torque contribution.

## V. GENERATION AND RECOMBINATION PROCESSES

Generally generation and recombination processes can be divided into radiative and non-radiative. The non-radiative transitions do not involve photons: they may involve the interaction of the electron with phonons or the exchange of energy or momentum with another electron or hole. Both energy and momentum must be conserved with Fermi Golden Rule (4; 5) describing the transition. Transitions may also be divided into band-to-bound-state and band-to-band types.

Fermi Golden Rule (4; 5) describes the transitions rate,  $W_{i \rightarrow f}$  for a transitions from an initial state  $\psi_i$  with energy  $E_i$  to a final state  $\psi_f$  with energy  $E_f$  when some Hamiltonian operator  $\hat{H}$  describes the transition where photons contribute with energy,  $\hbar\omega \sim E_f - E_i$ :

$$W_{i \rightarrow f} = \frac{2\pi}{\hbar} \left( \left| \langle \psi_f | \hat{H} | \psi_i \rangle \right|^2 \delta(E_f - E_i - \hbar\omega) + \left| \langle \psi_f | \hat{H} | \psi_i \rangle \right|^2 \delta(E_f - E_i + \hbar\omega) \right) \quad (42)$$

A first consequence on carrier transport is that the average distance that a carrier can diffuse before recombining is known as the diffusion length  $L_{n,p}$ . If the mean distance between collisions is  $\lambda$ , then in a time  $\tau_r$ , the recombination lifetime, a carrier can make  $\tau_r/\tau_{d_{n,p}}$  collisions and can diffuse a distance  $\lambda\sqrt{\tau_r/\tau_{d_{n,p}}}$ .

Using the average thermal velocity  $\sqrt{k_B T/m^*}$ , the diffusion length is obtained as:

$$L_{n,p} = \lambda\sqrt{\tau_r/\tau_{d_{n,p}}} = \sqrt{D_{n,p}\tau_r} \quad (43)$$

In order to determine  $\tau_r$  one has to consider several recombination scenarios involving, for instance, traps inside the gap.

When a semiconductor is doped with donor or acceptor impurities, impurity energy levels (or impurity bands) are introduced in the band gap.

The generation or recombination process can then be thought of as a non-equilibrium process when  $n_i^2 \neq np$ .

### A. Intrinsic Band-to-Band Generation-Recombination Processes

For band-to-band transitions, the rate of recombination for electrons,  $R_n$  and holes,  $R_p$  is proportional to the product of the electron,  $n$  and hole,  $p$  concentrations such that

$$R_n = R_p = \gamma np \quad (44)$$

where  $\gamma$  is the capture coefficient and the generation rate for electrons  $G_n$  or holes  $G_p$  may be written in terms of the emission rate  $\mathcal{E}$ , as  $G_n = G_p = \mathcal{E}$ .

The net recombination rate is  $R = R_n - G_n = R_p - G_p = \gamma np - \mathcal{E}$

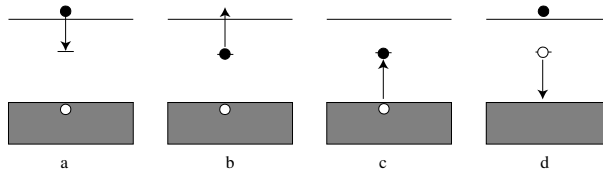
If no electric nor optical excitation of carriers are creating an external perturbation to the system, then at thermal equilibrium, the net recombination rate should be zero giving  $\gamma n_0 p_0 - \mathcal{E} = 0$  where  $n_0$  and  $p_0$  are the electron and hole

densities at thermal equilibrium i.e.  $n_0 p_0 = n_i^2$ . Thus the net recombination rate can be rewritten as  $R = \gamma(np - n_0 p_0)$ .

If the carrier concentrations for electrons and holes deviate from the equilibrium values by  $\delta n$  and  $\delta p$  respectively,  $n = n_0 + \delta n$  and  $p = p_0 + \delta p$  and therefore for low injection level (i.e. for  $\delta p \ll (n_0 + p_0)$ ), the recombination rate is  $R = \gamma(n_0 + p_0) = \delta n / \tau_R$  where  $\delta n = \delta p$  since the electrons and holes are created in pairs for interband transitions and the recombination lifetime is defined as  $\tau_R = 1/\gamma(n_0 + p_0)$ .

## B. Extrinsic Shockley-Read-Hall Generation-Recombination

The Shockley-Read-Hall processes are related to the emission or absorption of a phonon in the system.



**Fig.15:** Energy band diagram for Shockley-Read-Hall generation and recombination process (a) electron capture, (b) electron emission, (c) hole capture and (d) hole emission.

- The first process is electron capture (Fig. 15 (a)) where the recombination rate for the electron is proportional to the density of the electrons and the trap density  $N_t$  in the band gap, multiplied by the probability that the trap is empty ( $1 - f_t$ ) where  $f_t$  is the occupation function of the trap:

$$R_n = \gamma_n n N_t (1 - f_t) \quad (45)$$

where  $\gamma_n$  is the electron capture coefficient.

- The second process is electron emission (Fig. 15 (b)) where the electron generation rate is  $G_n = \epsilon_n N_t f_t$ ,  $\epsilon_n$  the emission coefficient and  $N_t f_t$  the density of traps occupied by electrons.
- The third process is hole capture (Fig. 15(c)) where the hole recombination rate is given by  $N_t f_t$  and  $R_p = \gamma_p p N_t f_t$  with  $\gamma_p$  the hole capture coefficient.
- The final process is hole emission (Fig. 15 (d)) where the generation rate is proportional to the empty density (since holes occupy)  $G_p = \epsilon_p N_t (1 - f_t)$  where the hole emission coefficient is  $\epsilon_p$ .

Since at thermal equilibrium, the net recombination and generation rates are zero, using detailed balance principle,

$$R_n - G_n = \gamma_n n_0 N_t (1 - f_{t0}) - \epsilon_n N_t f_{t0} = 0, \quad R_p - G_p = \gamma_p p_0 N_t f_{t0} - \epsilon_p N_t (1 - f_{t0}) = 0 \quad (46)$$

Note that values with a subscript 0 refer to the equilibrium values. Therefore:

$$\epsilon_n = \gamma_n n_0 (1 - f_{t0}) / f_{t0}, \quad \epsilon_p = \gamma_p p_0 (1 - f_{t0}) / f_{t0} \quad (47)$$

For degenerate semiconductors, the intrinsic concentrations may be replaced by an effective concentration,  $n_{ie}$ . The net recombination rate is

$$\begin{aligned} R_n - G_n &= R_p - G_p = \frac{(\gamma_n \gamma_p n p - \epsilon_n \epsilon_p) N_t}{\gamma_n n + \epsilon_n + \gamma_p p + \epsilon_p} \\ &= \frac{\gamma_n \gamma_p (n p - n_i^2) N_t}{[\gamma_n (n + n_0 (1 - f_{t0}) / f_{t0}) + \gamma_p (p + p_0 f_{t0} / (1 - f_{t0}))]} \end{aligned} \quad (48)$$

for e.g. an n-type semiconductor with  $n_0 \gg p$ ,  $\Delta p \gg p$ , the net recombination rate,  $R_n - G_n$  is roughly given by  $\Delta p / \tau_p$ , where  $\tau_p = 1/\gamma_p N_t$  and  $\tau_n = 1/\gamma_n N_t$  are the electron and hole lifetimes respectively.

## VI. ELECTRONIC DEVICES

Electronic families are divided in four categories: zero-polar, unipolar, bipolar and hybrid. CCD (Charge Coupled Devices) are considered as zero-polar (CCD action does not involve any transistor), whereas MOS (Metal Oxide Semiconductor) and low power CMOS (Complementary MOS) devices are considered as unipolar since they are based on one type of semiconductors (n or p). Thus a MOSFET is a unipolar Field-Effect transistor whose operation is distinct from the BJT (bipolar junction transistor) using both types (n and p) of semiconductors.

Bipolar family requires the presence of both types of semiconductors (an example is the BJT that could be p-n-p or n-p-n...).

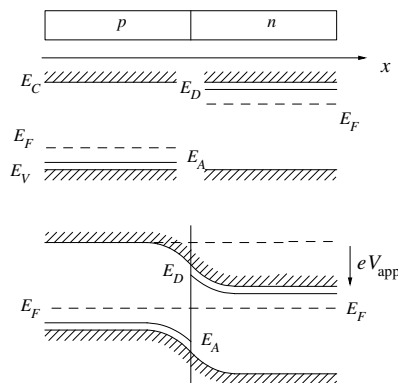
Hybrid families cover many types of devices, such as a mixture of unipolar and bipolar types with the Bi-CMOS example exploiting large signals handled with bipolar devices and low consumption with CMOS. It could also be a mixture of analog and digital devices, or a mixture of RF and low frequency devices...

### A. p-n junction

A p-n junction is formed when p-type and n-type semiconductor regions are joined to each other. The p-n junction is a semiconductor diode.

When p-type and n-type semiconductors are joined to make a p-n junction we get diffusion of one type of charge carriers into the other type of semiconductor. This leads to electron-hole recombination with the formation of a high resistance zone termed depletion layer (also called Space-Charge Region or SCR) because it contains low density localized charges.

Therefore, we have a non-equilibrium situation disobeying the law of mass action with  $np < n_i^2$ .



**Fig.16:** p-n junction at equilibrium and forward biased. Before junction is formed, Fermi levels are different in the p and n materials. However, at equilibrium after junction completion, the Fermi level is constant everywhere whereas band energy boundaries  $E_C, E_V$  and dopant levels  $E_A, E_D$  display bending.

Total current density of electrons is made of drift and diffusion is zero at equilibrium:

$$-n(x)e\mu_e \frac{dV(x)}{dx} + eD_e \frac{dn(x)}{dx} = 0 \quad (49)$$

where  $V(x)$  is the potential and  $n(x)$  the electron density at a distance  $x$  from one side of the junction.

Rearranging and integrating from  $x = x_n$ , well in the n-side of the junction, to  $x = x_p$ , well in the p-side one obtains

$$\int_{x_p}^{x_n} \mu_e \frac{dV(x)}{dx} = \int_{x_p}^{x_n} \frac{D_e}{n(x)} \frac{dn(x)}{dx} \quad (50)$$

Then

$$V_n - V_p = \frac{D_e}{\mu_e} \ln \left[ \frac{n_n}{n_p} \right] \quad (51)$$

If the junction was made by doping the n-side with  $N_D$  donors and the p-side with  $N_A$  acceptors, then using the approximation  $np \approx n_i^2/N_A$  we get  $n_n = N_D$ .

Hence

$$V_n - V_p = \frac{D_e}{\mu_e} \ln \left[ \frac{N_A N_D}{n_i^2} \right] \quad (52)$$

At equilibrium, the Fermi level is constant everywhere and the potential energy difference is equal to  $(E_c - E_F)$  on the p-side minus  $(E_c - E_F)$  on the n-side.

Combining previous results we obtain:

$$V_n - V_p = \frac{k_B T}{e} \ln \left[ \frac{N_c}{n_n} \right] - \frac{k_B T}{e} \ln \left[ \frac{N_c}{n_p} \right] \quad (53)$$

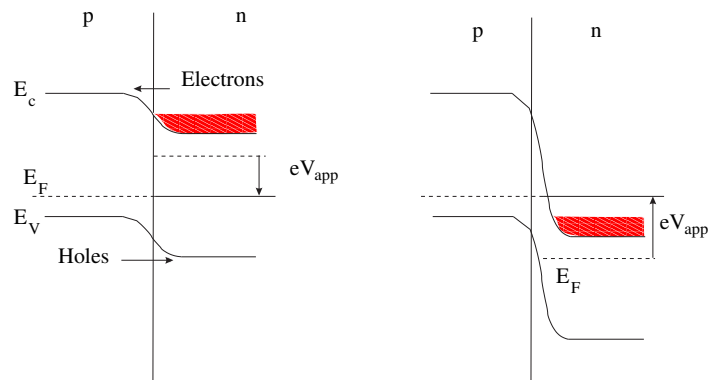
The potential difference between the two sides is termed the barrier potential  $V_b$ . Using  $n_n = N_D$  and  $n_p = n_i^2/N_A$  we obtain:

$$V_b = -\frac{k_B T}{e} \ln \left[ \frac{N_D N_A}{n_i^2} \right] \quad (54)$$

Typically in Ge and Si junctions,  $V_b$  is respectively 0.4 V and 0.8 V.

### 1. Space-charge region

In order to describe simply the physics of the SCR, we assume that doping atom density changes rapidly from one value  $N_A$  (acceptor density) on the p-side to a value  $N_D$  (donor density) on the n-side.



**Fig.17:** Energy band diagrams for a p-n junction with forward bias (left) and reverse bias (right).

Such a junction is known as a step junction or an abrupt junction. For simplicity, we assume that the total potential difference  $V_J$  across the junction (cf. Fig. 17) is such that  $V_J \gg k_B T/e$  the thermal voltage. In the depleted part of the p-region, the only charges are localized negatively charged acceptor ions.

Integrating Poisson equation:  $dE/dx = -eN_A/\epsilon_r\epsilon_0$  gives:  $E(x) = -eN_A x/\epsilon_r\epsilon_0 + constant$ . Assuming  $E = 0$  when  $x = x_p$  yields:  $E(x) = -eN_A(x - x_p)/\epsilon_r\epsilon_0$ .

Similarly, on the  $n$  side we get:  $E(x) = eN_D(x - x_n)/\epsilon_r\epsilon_0$  resulting in a potential  $V_J$  across the junction given by:

$$V_J = \frac{e}{2\epsilon_r\epsilon_0}(N_Dx_n^2 + N_Ax_p^2) \quad (55)$$

The largest value of the electric field  $E_{max}$  provides another relation between the boundary limits  $N_A|x_p| = N_D|x_n|$ .

Thus a thicker depletion layer occurs when the doping is lighter.

Consequently we suppose we can ignore the thickness and voltage in a heavily doped side. Such a junction is indicated with a plus sign ( $p^+ - n$ ), and the total thickness of the depletion layer is proportional to  $V_J$ . Accordingly the total depletion layer width  $W$  is

$$W = \sqrt{\frac{2\epsilon_r\epsilon_0(N_D + N_A)}{eN_DN_A}V_J} \quad (56)$$

## 2. Varicap Diode

By applying a reverse bias to the junction creating a depleted zone (SCR) in the metallurgical p-n junction, it is possible to make a device having a variable capacitance (cf. Fig. 17).

The bias induces a difference between the Fermi levels on both sides. Forward bias reduces the barrier potential and reverse bias increases it producing a wider SCR.

Practically the capacitance of the junction varies with the applied voltage as if the boundaries of the SCR were capacitor plates. It is, in fact, the capacitance of the depletion layer that depends on the bias voltage. The capacitance  $C(V_J)$  of the depletion layer is:

$$C(V_J) = dQ(V_J)/dV = (dQ/dW)(dW/dV) \quad (57)$$

where  $Q$  is the charge on each plate of the capacitor,  $W$  is the total thickness of the depletion layer and  $V_J$  is the total potential difference between the p and n regions. If we take a p-n<sup>+</sup> diode, then the major part of  $V_J$  and of  $W$  is in the p-region. As a result  $W \approx x_p$  and  $Q = eN_Ax_pA$  with  $A$  the junction cross-section area.

Thus

$$dQ/dV = dQ/dx_p = eN_AA \quad (58)$$

Eliminating  $x_p$  allows us to rewrite the above  $C(V_J)$  as:

$$C(V_J) = A\sqrt{\frac{eN_A\epsilon_r\epsilon_0}{2}} \frac{1}{\sqrt{V_J}} \quad (59)$$

In terms of SCR thickness, we get  $C(V_J) = A\epsilon_r\epsilon_0/W$ .

Thus  $C(V_J)$  characterizes an ordinary capacitor having the same size, shape and permittivity pertaining to the depletion layer.

## 3. Debye Length

If the doping concentration, however, changes abruptly then the bands may not follow this change as quickly as the doping profile. This is because the doping profile may be discontinuous while the intrinsic potential  $V(x)$  and the first derivatives must be continuous from any thermal diffusion effects.

There is therefore a length scale, called the Debye length,  $L_D$  which is the distance over which the bands of a semiconductor respond to a change of doping for a n-type semiconductor. In order to derive an expression for  $L_D$ , we use Poisson equation to obtain:

$$\frac{d^2V(x)}{dx^2} = -\frac{e}{\epsilon_0\epsilon_r} \left[ N_D(x) - n_i \exp\left(\frac{e(V(x) - V_F)}{k_B T}\right) \right] \quad (60)$$

$N_D(x)$  is the doping profile,  $V_F = -E_F/e$  is the Fermi potential and  $n_i$  is the intrinsic electron density.

Consider a local change  $\Delta N_D(x)$  in doping profile. The corresponding change to the intrinsic potential,  $\Delta V(x)$  can be found by expanding the above exponential term and eliminating coefficients with no space variation:

$$\left( \frac{d^2}{dx^2} - \frac{e^2 N_D}{\epsilon_0 \epsilon_r k_B T} \right) \Delta V(x) = -e \epsilon_0 \epsilon_r \Delta N_D(x) \quad (61)$$

This is a second-order ordinary differential equation (ODE) whose solution  $\Delta V(x) \sim \exp(x/L_D)$  with

$$L_D = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{e^2 N_D}} \quad (62)$$

$L_D$  is the largest distance for Coulomb interaction between charged particles present.  $L_D$  is usually much smaller than any lateral dimension of many devices (like transistors) to become a crucial parameter when miniaturization triggers quantum effects.

#### 4. Numerical treatment

Starting from the constitutive system of equations developed by Polak *et al.* (24):

$$\begin{aligned} \frac{d\psi}{dx} &= -E \\ J_n &= e \left[ D_n \frac{dn}{dx} - \mu_n n \frac{d\psi}{dx} \right] \\ J_p &= -e \left[ D_p \frac{dp}{dx} + \mu_p p \frac{d\psi}{dx} \right] \\ \frac{1}{e} \frac{dJ_n}{dx} &= R(n, p) + G(x) \\ \frac{1}{e} \frac{dJ_p}{dx} &= -R(n, p) - G(x) \end{aligned} \quad (63)$$

with the recombination term of the Shockley-Read-Hall form simplified with respect to the previous expression eq. 48:

$$R(n, p) = \frac{(np - n_i^2)}{T_1 n + T_2 p + T_3} \quad (64)$$

where  $T_1, T_2$  are time constants whereas  $T_3$  includes carrier concentration. Typically for Silicon  $T_1 = \tau_p = 10^{-5}$  sec,  $T_2 = \tau_n = 10^{-5}$  sec, and  $T_3 = (\tau_p + \tau_n)n_i$  (25).

In order to solve the boundary value problem associated with the above system (when a voltage is applied to the p-n junction) we transform it into a hybrid system of ordinary first-order (Current and carrier density equations) and a second-order Poisson ODE.

The mathematical/numerical reasons for performing this transformation reside in the fact the above system is a *singularly singular perturbed problem* (26; 27). Many algorithms (27; 28) have been developed in order to deal with this difficulty stemming from several facts:

1.  $\psi, n$  and  $p$  are fast variables in comparison with  $E, J_n$  and  $J_p$  (25; 29; 30).

2. Near the depletion layer boundaries, the values of  $n$  and  $p$  change by several orders of magnitude making the SCR a double boundary layer. This difficulty is of the same type as the one encountered in Hydro or Aero-dynamics where the fluid velocity changes by several orders of magnitude near an obstacle.

Recognizing the difficulty due to the presence of the SCR, a standard way to find a valid solution is to treat the boundary layer separately from the rest of the diode. In spite of the success of this approach (31; 32), one might feel uneasy about this methodology and rather tackle the problem with new powerful mathematical/numerical methods that will treat the layer and the rest of the device on the same equal footing.

3. When the above system is rewritten explicitly in terms of the Poisson equation as we will do below, the second spatial derivative of the electric potential is multiplied by a very small number  $\nu^2$  ( $\nu \sim 10^{-4}$  to  $10^{-3}$ ) and this is the reason this problem is called singularly perturbed: the solution with  $\nu = 0$  is entirely different from the solution with  $\nu$  finite but small (27).

The Scharfetter-Gummel algorithm (33) allow to separate fast/slow variables by integrating out the fast variables over some small interval while holding the slow variables constant over that same interval. The Scharfetter-Gummel algorithm leads to a spatial exponential discretization that will slow down somehow the swift behavior of fast variables.

We decided not to use the Gummel algorithm nor none of its flavors but rather tackle the problem head-on from the singular perturbation point of view since this approach is far more rigorous and leads to a better handling of any instability problem encountered in the semiconductor set of ODE.

Transform first the system in the following dimensionless two-point boundary value problem with no generation processes:

$$\begin{aligned}
\frac{dn}{dx} &= C_1 J_n + n \frac{d\psi}{dx} \\
\frac{dp}{dx} &= -C_2 J_p - p \frac{d\psi}{dx} \\
\frac{dJ_n}{dx} &= C_3 \frac{(np - 1)}{n + \tau_1 p + \tau_2} \\
\frac{dJ_p}{dx} &= -C_3 \frac{(np - 1)}{n + \tau_1 p + \tau_2} \\
\frac{d^2\psi}{dx^2} &= C_4 (n - p + N_D)
\end{aligned} \tag{65}$$

The constants  $C_1, C_2, C_3$  and  $C_4$  are given by:

$$C_1 = \frac{J_{SC} L_D}{e n_i D_n}, C_2 = \frac{J_{SC} L_D}{e n_i D_p}, C_3 = \frac{e L_D n_i}{J_{SC} T_1}, C_4 = \frac{e L_D^2 n_i}{\epsilon_s U_T}$$

where the Debye length  $L_D$  is given by  $L_D = \sqrt{\frac{k_B T \epsilon_s}{n_i e^2}}$  and the scaling current  $J_{SC} = \frac{n_i \mu_n k_B T}{L_D}$ .

$U_T = k_B T / e$  is the thermal voltage. The modified constants are  $\tau_1 = T_2 / T_1$  and  $\tau_2 = T_3 / (T_2 n_i)$ .

COLSYS (27; 28; 34) is a (Bézier) B-spline collocation (28) method. It is based on a meshing technique (35; 36) of the boundary layer leading to singularity damping. The tunable layer meshing being exponential encompasses the Scharfetter-Gummel (27) method and can be shown rigorously to have the form:

$$h_i = h_{i-1} \exp(h_{i-1} \delta / \nu)$$

where  $h_i$  is the  $i$ -th mesh point,  $\delta$  is a constant related to the required accuracy and nature of the B-spline whereas  $\nu$  is the singular value parameter.

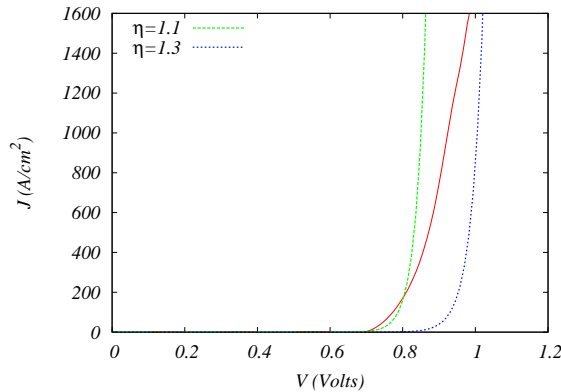
We validate our approach by the agreement quality with Shockley equation obtained after evaluating the J-V (current density-voltage) characteristics of the junction.

Since we have used the Shockley-Read-Hall recombination term, we do not expect the simple Shockley result but rather the modified form:

$$J_S = J_0[\exp(eV/\eta k_B T) - 1], \quad J_0 = \frac{eD_p n_i^2}{L_p N_D} + \frac{eD_n n_i^2}{L_n N_A}$$

The diffusion lengths are given by:  $L_n = \sqrt{D_n \tau_n}$ ,  $L_p = \sqrt{D_p \tau_p}$ .

The comparison of the obtained J-V characteristic with Shockley formula is displayed in Fig. 18. The calculated characteristic falls between the two Shockley curves  $\eta = 1.1$  and  $\eta = 1.3$ .



**Fig.18:** (Color on-line) p-n junction J-V characteristics obtained with COLSYS (in red) of the Silicon p-n diode with Shockley approximation  $J_0[\exp(eV/\eta k_B T) - 1]$  bounding curves with  $\eta = 1.1$  and  $\eta = 1.3$ . Using Silicon data (cf. Appendix) we get  $J_0 = 3.671 \times 10^{-11}$  A/cm<sup>2</sup>.

## B. Field-Effect Transistor

Silicon dioxide has been used traditionally as a gate oxide material for electric insulation. Shrinking the transistor implied reducing the SiO<sub>2</sub> gate dielectric thickness (cf. Fig. 12) in order to increase the gate capacitance and thereby drive current and device performance. When gate oxide thickness is thinner than 2 nm (predicted minimal feature length for 2024), quantum tunneling induces leakage currents across the oxide decreasing device performance. If the silicon dioxide is replaced by a high-dielectric constant material, gate capacitance increases without current leakage.

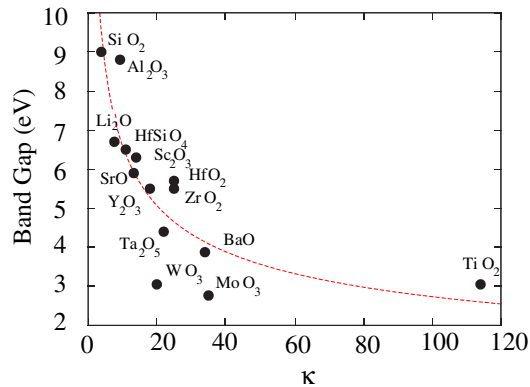
High- $\kappa$  (as Micro-electronics engineers call the relative dielectric constant  $\epsilon_r$ ) dielectrics are materials having a large band gap ( $E_G$ ) and high ( $\kappa$ ) simultaneously, at the foundations of critical components in micro-electronic devices.

High-quality interfaces between the Si substrate and SiO<sub>2</sub> gate dielectrics facilitated the steady down scaling of Si devices, which accelerated the operation speed while reducing power consumption. However, as the thickness of SiO<sub>2</sub> is decreased to less than a few nanometers, the traditional fabrication process is challenged by significant leakage currents that originate from tunneling through ultrathin SiO<sub>2</sub> dielectric.

### 1. MOSFET with high- $\kappa$ dielectrics

The solution to the above problems can be resolved by incorporating high- $\kappa$  materials such as HfO<sub>2</sub> and ZrO<sub>2</sub> instead of SiO<sub>2</sub>. These high- $\kappa$  oxides can reduce leakage currents by increasing the physical thickness of insulating layers while enhancing capacitive coupling between the channel layer and gate electrode. Unfortunately, Hafnium silicates are susceptible to trap-related leakage currents and when Hafnium concentration is increased, performance decreases.

The inverse relationship  $E_G \sim 1/\kappa$  (cf. fig 19) makes materials with large values in both properties extremely rare.



**Fig.19:** (Color on line) Band gap (eV) versus  $\kappa$  displaying inverse relationship  $E_G \sim 1/\kappa$  (in red) for many oxides.

Name	$E_G$ (eV)	$\kappa$	$E_G \times \kappa$
BiF <sub>3</sub>	6.07	28.9	175.2
BaBeF <sub>4</sub>	9.72	17.0	165.0
Tl(AlF <sub>4</sub> )	5.76	27.4	157.6
LaF <sub>3</sub>	8.84	16.5	145.9
LiSbF <sub>4</sub>	6.58	21.9	144.1
Bi <sub>7</sub> F <sub>11</sub> O <sub>5</sub>	5.15	27.5	141.5
Li(Sb <sub>2</sub> F <sub>7</sub> )	5.96	22.9	136.2
LaF <sub>3</sub>	9.90	12.8	126.6
Ge <sub>5</sub> F <sub>12</sub>	5.66	22.1	125.3
PbF <sub>2</sub>	5.77	21.7	124.9
Sn <sub>2</sub> ClF <sub>3</sub>	4.57	26.7	122.2

**Table 7:** Candidate fluorides for high- $\kappa$  dielectrics with  $E_G > 4\text{eV}$  and Figure Of Merit  $\text{FOM} = E_G \times \kappa > 120$ . Fluorides are sorted in the decreasing order of FOM.

While experimental and theoretical data on  $E_G$  and  $\kappa$  of oxides are accumulating, corresponding information for non-oxide dielectrics with anions such as C, N, F, P, S, and Cl, is still scarce.

A strong paradigm shift would be fluorides such as BiF<sub>3</sub>, LaF<sub>3</sub>, and BaBeF<sub>4</sub> serving as high- $\kappa$  dielectrics to definitely replace SiO<sub>2</sub> in electric insulation.

According to the International Roadmap for Devices and Systems (IRDS), dielectrics with  $\kappa$  of 50-100 will be required in transistors or capacitors very soon. Among the oxides, rutile TiO<sub>2</sub> or perovskite SrTiO<sub>3</sub> with  $\kappa > 100$  are attracting interests as next-generation gate dielectrics but their small band gaps cause significant leakage currents.

This indicates that a more diverse library of high- $\kappa$  materials will be beneficial in coping with the challenges encountered in next-generation semiconducting devices.

Ranking the candidate materials is based on using  $E_G \times \kappa$  as a figure of merit since  $E_G$  and  $\kappa$  are approximately proportional to the leakage current density logarithm (cf. Table. 7).

## 2. Negative capacitance ferroelectric FET

Another means to struggle against tunneling across the MOSFET oxide is to replace it (cf. Fig. 12) with a ferroelectric having a negative capacitance (37).

In a conventional FET, the gate voltage ( $V_G$ ) drops partially in the insulator ( $V_{ins}$ ) and partially in the semiconductor so that  $V_G = V_{ins} + \phi_S/e$  where  $\phi_S$  is the semiconductor work-function. The charge density  $Q$  in the channel between the drain and the source (cf. Fig. 12) is determined by  $C_S \times \phi_S/e$ , where  $C_S$  is the semiconductor capacitance. For

normal dielectrics, both  $Q$  and the current increase at best by an order of magnitude per 60 mV of increase in gate voltage at room temperature due to Boltzmann statistics.

In an NC-FET,  $V_G = V_F + \phi_S/e$ , where  $V_F$ , a voltage drop across a ferroelectric, is negative. Hence,  $\phi_S/e > V_G$ , meaning that one can achieve the same charge density in the channel with a smaller voltage at the gate. This means a current increase steeper than 60 mV per decade that could be exploited to reduce the voltage needed to turn on the transistor with respect to a conventional FET.

### 3. Multigate devices

Yet another approach to deal with miniaturization problems is to turn to multigate devices that are tailored to adapt to miniaturization and reduce leakage between source and drain with fins, nanowires or nanoribbons arranged perpendicularly to gate surface.

While fins are perpendicular to both gate and gate-oxide, nanowires and nanoribbons are perpendicular to gate but parallel to gate-oxide.

Presently the MOSFET is replaced by the finFET and in the future it will be the Ribbon FET, a version of the GAA (Gate-All-Around) transistor where nanoribbons replace the nanowires.

## VII. MAGNETIC, OPTICAL, SPINTRONIC, VALLEYTRONIC DEVICES

### A. Hall effect devices

Currently, Magnetic compasses and Magnetic field detectors in many smartphones are based on the Hall effect.

The Hall-effect was discovered by Edwin Hall who was still a student in 1879 and can be explained with the Lorentz force acting on an electric charge  $-e$  moving with velocity  $\dot{\mathbf{r}}$  in presence of  $\mathbf{E}$  and  $\mathbf{B}$  fields.

Generalizing Drude approach, we write the equation of motion in  $\mathbf{v}$  as:

$$m_e \dot{\mathbf{v}} + \frac{m_e \mathbf{v}}{\tau} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (66)$$

This is a linear system whose solution can be found from superposition of the transient response given by:

$$m_e \dot{\mathbf{v}}_T + \frac{m_e \mathbf{v}_T}{\tau} = 0, \quad \mathbf{v}_T = \mathbf{v}_0 \exp(-t/\tau) \quad (67)$$

and the stationary response given by:

$$m_e \dot{\mathbf{v}}_S = -e(\mathbf{E} + \mathbf{v}_S \times \mathbf{B}) \quad (68)$$

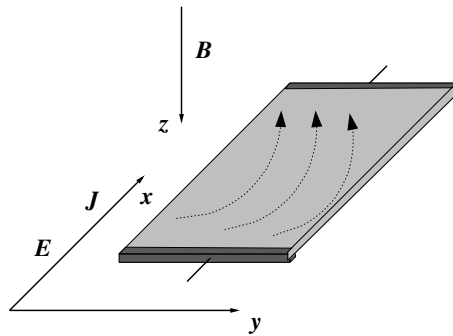
Using  $\mathbf{v}_S \sim \exp(i\omega t)$ ,  $\mathbf{E} = E_x \mathbf{x}$ ,  $\mathbf{B} = B_z \mathbf{z}$ , we get the  $x, y, z$  components as:

$$m_e i\omega v_{Sx} = -e(E_x + v_{Sy} B_z), \quad m_e i\omega v_{Sy} = e v_{Sx} B_z, \quad v_{Sz} = 0 \quad (69)$$

Solving the system we get:  $v_{Sx} = \frac{e E_x i m_e \omega}{m_e^2 \omega^2 - e^2 B_z^2}$ .

The solution reveals existence of a resonance (cyclotron) given by  $\omega_c = \frac{e B_z}{m_e}$ . The equilibrium condition  $\mathbf{v}_{Sy} = 0$  produces the electric fields:

$$E_x = v_{Sx} B_z, \quad E_y = -\frac{e \tau E_x}{m_e} B_z = -\omega_c \tau E_x \quad (70)$$



**Fig.20:** Hall effect with magnetic induction  $\mathbf{B} = B_z \mathbf{z}$  acting on electronic current  $\mathbf{J} = J \mathbf{x}$  resulting in transverse accumulation of charge due to Lorentz force leading to a transverse voltage along  $\mathbf{y}$  direction.

The transverse voltage originating from charge accumulation perpendicular to the original current direction along  $\mathbf{x}$  direction  $J_x = \frac{n_e e^2 \tau}{m_e} E_x$  is the Hall effect (cf. Fig. 20).

This leads to a Hall resistance given by  $R_H = \frac{E_y}{J_x B_z} = -\frac{1}{en_e}$  indicating sign of carriers and being inversely proportional to their density  $n_e$  giving a huge advantage to semiconductors with respect to metals in addition to the possibility of changing the sign of  $R_H$  by replacing electrons by holes.

## B. Light Emitting Diode

The LED has revolutionized communication technology like semiconductor laser diodes and photo-diodes in addition to applications in lighting, display and sensing. A LED is a p-n junction fabricated with special materials that can convert electrical energy into optical radiation. The process is named electroluminescence.

Electroluminescent light differs from thermal radiation or incandescence (radiation resulting from conductor heating in Tungsten bulbs). A LED spectral linewidth is typically 10 to 50 nm whereas a laser diode one is a fraction of to few nm.

Light is obtained by injecting minority carriers into the region of a p-n junction where radiative transitions take place. This is achieved by forward biasing the p-n junction, such that electrons and holes are pushed from the n and p regions toward the metallurgical junction to recombine releasing energy in terms of photons with energy close to the gap value (cf. Fig. 21).

In fact, there are three types of recombination processes:

- A radiative process ( $R_{rd}$ ) leading to photon emission
- A non-radiative process ( $R_{nr}$ ) with emission of phonons, i.e., heat and includes the Shockley-Read-Hall (SRH) process seen previously
- A non-radiative process ( $R_{nr}$ ) with Auger recombination process involving momentum and energy transfer from an electron-hole pair to another particle (electron or hole). The Auger process is important when carrier concentration is high.

The internal quantum efficiency (IQE) is a measure of efficiency and is related to how electrons are converted into photons within an active region (providing amplification).

All of these processes are dependent on the minority carrier concentration. The SRH process increases linearly ( $An$ ), the radiative recombination process quadratically ( $Bn^2$ ), whereas the Auger recombination process increases as the cube ( $Cn^3$ ) of  $n$ . The coefficients  $A$ ,  $B$ , and  $C$  are constants.

The IQE is defined as (38):

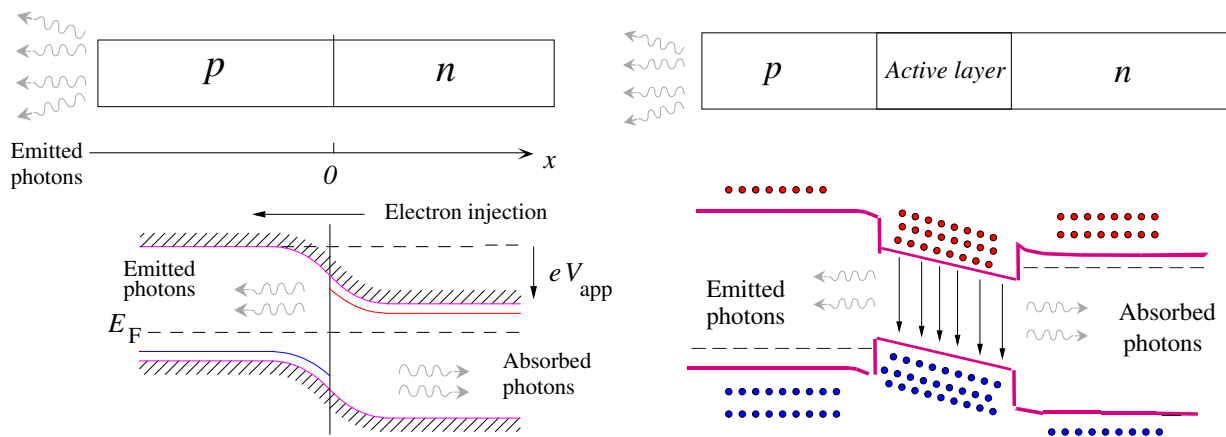
$$IQE = \frac{\text{Light produced}}{\text{Electrons injected}} = \frac{R_{rd}}{R_{rd} + R_{nr}} = \frac{Bn^2}{An + Bn^2 + Cn^3} \quad (71)$$

For low carrier concentration the SRH term ( $An$ ) dominates leading to poor IQE and significant non-radiative recombination. This is the case for a p-n homojunction LED (cf. Fig. 21). The p-n structure of the LED, is such that electrons diffuse into the p-type layer whereas holes diffuse into the n-type layer. The diffusion length for minority carriers in GaN is approximately  $1 \mu\text{m}$  (38) This causes carriers to spread out over a large region reducing their concentration (cf. Fig. 21).

The Double Heterostructure LED (cf. Fig. 21), on the other hand, confines the carriers to within an active layer allowing amplification, which is typically around 3-200 nm thick. This confinement significantly increases the carrier concentration under the same current density and enhances the probability of radiative recombination ( $Bn^2$ ) thereby increasing the the IQE.

Since the energy of the emitted photons is approximately equal to the band gap, photon wavelength can be tuned by using materials with different band gaps.

The design and construction of the LED p-n junction must maximize photon output from the diode (cf. Fig. 21). Absorption in the semiconductor and total internal reflection reduce the effective output by a factor of up to one hundred in some devices.



**Fig.21:** Left: p-n homojunction structure of a LED. Right: Double Heterostructure with an intermediary central region (active layer) rendering the LED more effective. Note that the bandgap of the active layer is smaller than the bandgaps of both n and p regions.

For a long time up to 1993, a LED could only emit green or red light. Blue and therefore White LED (required for lighting and displays) were elusive.

Lighting is based on the notion of light efficacy with the the ideal efficacy  $K_{cd} = 683 \text{ lm/Watt}$  (lumens/Watt) according to the SI units committee (39) as well as light efficiency which is the ratio of the efficacy and  $K_{cd}$  (cf. Table 8).

The luminous efficacy  $\eta_P$  of any radiation source characterized by an emitted power spectrum function  $P(\lambda)$  is given by:

$$\eta_P = K_{cd} \frac{\int_{\mathcal{D}_\lambda} P(\lambda) V(\lambda) d\lambda}{\int_{\mathcal{D}_\lambda} P(\lambda) d\lambda} \quad (72)$$

In the particular case of a source of the Black-Body type  $P(\lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{[\exp(\frac{hc}{\lambda k_B T}) - 1]}$  and the wavelength domain  $\mathcal{D}_\lambda = [0, \infty]$ .  $c$  is the speed of light in vacuum and  $V(\lambda)$  is daylight eye sensitivity given by (40)  $V(\lambda) = 1.019 \exp(-285 [(\frac{\lambda}{1000}) - 0.559]^2)$  with  $\lambda$  expressed in nm.

For instance Sun light efficacy is about 93 lm/Watt (Black-body at temperature of 6000K) and an efficiency of  $93/683=13.6\%$  whereas a Tungsten bulb is about 15 lm/Watt (temperature about 3000K) or 2% only. For a candle considered as a Black-Body at  $T=1800K$ , we get 0.6 lm/Watt which corresponds to an efficiency of  $0.6/683$  or about 0.1%.

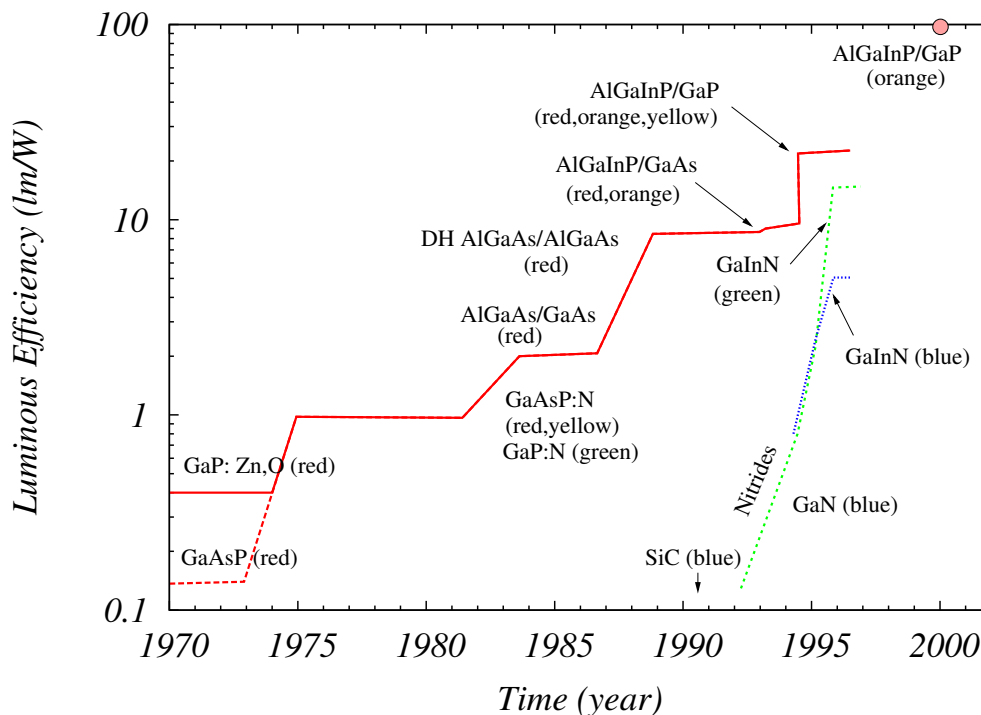
Light efficacy is a measure of perceived light power relative to the provided electrical power, of white light improved over the centuries, starting with ancient oil lamps (0.1 lm/W) and candles (0.6 lm/W), to Tungsten bulbs (15 lm/W) in the 19th century, fluorescent lamps (70 lm/W) in the 20th century, and LEDs (300 lm/W) in the 21st century.

Defining constant	Symbol	Numerical value	Unit
Hyperfine transition frequency of Cesium	$\Delta\nu_{Cs}$	9 192 631 770	Hz
Speed of light in vacuum	$c$	299 792 458	m.s <sup>-1</sup>
Planck constant	$h$	$6.626\ 070\ 15 \times 10^{-34}$	J.s
Elementary charge	$e$	$1.602\ 176\ 634 \times 10^{-19}$	C
Boltzmann constant	$k_B$	$1.380\ 649 \times 10^{-23}$	J.K <sup>-1</sup>
Avogadro constant	$N_A$	$6.022\ 140\ 76 \times 10^{23}$	mol <sup>-1</sup>
Luminous efficacy	$K_{cd}$	683	lm . W <sup>-1</sup>

**Table 8:** The seven defining constants of the SI and the seven corresponding units they define. J is Joule, W is Watt, lm is Lumen and  $K_{cd}$  refers to the Candela efficacy which is equal to the photopic standard value  $K_m = 683$  lm/Watt (lumens/Watt) previously adopted by the SI (39) system, also called the "Mechanical equivalent of the Lumen".

In terms of lighting, for a 60 Watts Tungsten bulb, light produced is  $60\text{ W} \times 15\text{ lm/Watt} = 900$  lumens in total and that has tremendous consequences for the quality and cost of the lighting desired.

Thus a LED is expected to perform better on both lighting and energy scales as displayed in Fig. 22.



**Fig.22:** Evolution of the LED toward the Blue. DH is Double Heterostructure. Adapted from Nakamura (38).

This was extremely frustrating to Physicists, Engineers and many workers in areas of technology related to Color

which is essentially trichromatic.

Three RGB primaries are needed in order to generate all other colors and White light.

Gallium nitride (GaN) was one possible candidate, although, for a long time, no p-type or active layer could be created in a stable fashion because of the difficulty of keeping Nitrogen permanently in its interstitial position.

These challenges were ultimately overcome by Nakamura, Mukai, and Senoh (38), leading to the first efficient blue LED using GaN in 1993.

With blue LEDs, highly efficient white light sources became possible. This can be achieved by converting part of the blue light emitted from the LED to yellow (38) and recombining them as shown below.

Color algebra is based on three primary colors (Red, Green, Blue or RGB) with  $R+G+B=1$  (White) making the additive system. Hence it is a Boolean ternary algebra and not a binary one. The complementary colors (Cyan, Magenta, Yellow) yield the subtractive system with  $C=1-R$ ,  $M=1-G$ ,  $Y=1-B$ . Thus  $C+M+Y=0$  (Black)  $C+R=1$ ,  $M+G=1$  and  $B+Y=1$  (White).

The result  $B+Y=1$  might be surprising since one might think that Blue and Yellow makes Green. In fact Green is obtained from  $C+Y=G$ . Using ternary algebra, we have  $C+Y=(1-R)+(1-B)=1-R-B$  and since  $1+1=1$  and  $R+G+B=1$ , we infer that  $C+Y=G$ .

Since  $B+Y=1$ , a White LED can be created by embedding inside a blue LED, phosphors that emit Yellow light when excited by a Blue one. This process is better bandwidth-wise than adding R, G and B. Moreover, it means that Blue light is always present somehow in White light requiring filters (against Blue and UV light) to protect human eyes.

Nevertheless this required a large number of technical breakthroughs that started in 1969 and lasted over a period of 25 years in order to be fully achieved (see Table. 9).

### C. Laser Diode

Laser is Light Amplification by Stimulated Emission of Radiation acronym. A medium pumped to reach population inversion provides optical gain to amplify an optical field through Stimulated Emission (SE). This medium is called active as in the LED case, since it provides amplification. In addition to optical amplification, optical feedback is required for laser oscillation as dictated by the Barkhausen criterion.

In fact, a laser is an optical oscillator that generates a coherent signal through resonant oscillation without an input signal. No external optical field is injected into the optical cavity for laser oscillation. In the steady-state oscillation regime, the coherent laser field inside the cavity is constant versus time in both modulus and phase.

There is a nice comparison to be made with the LED case as displayed in Fig. 21: As we go from the single homojunction (with no cavity) to a Double Heterostructure containing an active layer (cavity), performance gets better with confinement. Going further entails inserting an even thinner cavity such as a Quantum Well (QW) or a series of QW (MQW: Multiple QW) structure to increase performance (see Table 9).

Light oscillation is achieved by placing the active medium between two materials acting like reflecting mirrors (cf Fig. 23). This optical cavity resonator outputs a laser light that is highly collimated spatially and temporally coherent as well.

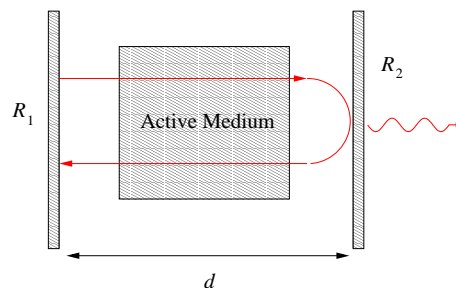
Coherence results from the fact a photon emitted by SE is coherent with the incident photon that induced the emission. The active medium emits spontaneous photons in all directions, but only the radiation that propagates along the longitudinal axis within a small divergence angle defined by the resonator provides required amplification through SE to obtain a laser.

Moreover we need a critical number of Two-Level-Systems (TLS) that are resonators with amplification, feedback and non-linearity since lasing oscillation is a collective effect involving a large number of TLS and not a single one.

An ordinary oscillator needs a (TLS) resonator with amplification, feedback and non-linearity, whereas a laser needs additionally a threshold density of TLS elements (atoms, molecules, electrons...) interacting with surrounding photons.

Material	Year	Technology breakthrough
GaN	1969	GaN epitaxial layer by HVPE
	1973	First blue Mg-doped GaN MIS LED
	1983	High quality GaN using AlN buffer by MBE
	1985	High quality GaN using AlN buffer by MOCVD
		p-type GaN using LEEBI (low hole concentration)
	1989	First p-n homojunction GaN LED
		Invention of Two-flow MOCVD
		GaN growth using GaN buffer by MBE
	1991	High quality GaN using GaN buffer by MOCVD
	1992	p-type GaN using thermal annealing (high hole concentration)
		Discovery of hydrogen passivation
InGaN	1972	InGaN growth using electron beam plasma
	1989	InGaN growth by MOCVD
	1992	InGaN layers with RT band-to-band emission
	1994	Efficient blue InGaN DH LED (1 Candela)
	1995	Efficient yellow, green, and
		blue InGaN DH QW LEDs
		First pulsed violet InGaN DH MQW LDs
		First CW violet InGaN DH MQW LDs
	1996	Commercialization of white LED using
		InGaN DH blue LED

**Table 9:** Timeline of technology breakthroughs ultimately leading to the white LED for GaN and InGaN on sapphire. HVPE: hydride vapor phase epitaxy, MIS: Metal Intrinsic Semiconductor, MOCVD: Metal Organic Chemical Vapor Deposition, LEEBI: Low-Energy Electron Beam Irradiation, DH: Double Heterostructure, QW: Quantum Well, MQW: Multiple Quantum Well, CW: Continuous Wave, RT: Room Temperature. Adapted from Nakamura (38).



**Fig.23:** (Color on-line) Laser resonator cavity filled with an active or amplifying medium and surrounded by two materials acting like reflecting mirrors with  $R_1 \sim 100\%$ , whereas  $R_2 < 100\%$  to allow escape of laser light. Since light makes round trips in the cavity and active medium, a phase condition is required (see text) to avoid destructive interference.

Townes was the first to show that a maser (41), the microwave counterpart of the laser, needs a large amount of Ammonia molecules (TLS) above a critical density.

Actually, the threshold is an indication that a phase transition (42) has taken place triggered by the many-body effect produced by a large number of interacting TLS.

To summarize, we need two conditions: threshold (gain) and phase as described below.

### 1. Gain condition

We derive the threshold condition for the laser by considering Fig. 23 where we have an active medium providing amplification to light performing a round trip between two mirrors.

The Barkhausen condition is given by:  $R_1 R_2 \exp[-2(\alpha_1 + \alpha)d] \geq 1$  where  $R_1, R_2$  are the reflection coefficients of the mirrors and  $\alpha_1$  are the ordinary loss coefficient whereas  $\alpha_1$  represent the active medium amplification.

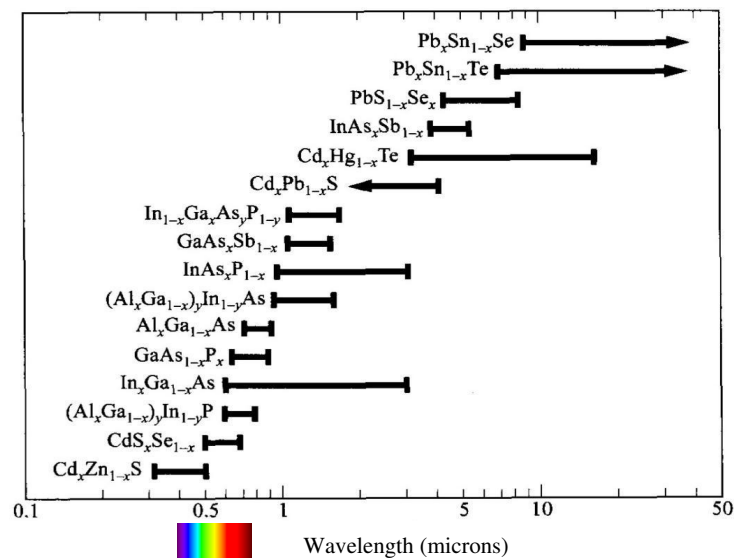
In order to have amplification by the active medium one should have  $\alpha < 0$  and therefore  $-\alpha \geq \alpha_1 - \frac{1}{2d} \ln(R_1 R_2)$ .

### 2. Phase condition

Another condition for lasing is the existence of a selected mode wavelength  $\lambda_{12}$ . This condition is related to phase, since light makes round trips in the cavity requiring a non-destructive interference constraint given by  $2\left(\frac{2\pi}{\lambda_{12}}\right)d + 2\phi(\lambda_{12})d = 2\pi p$  where the first term is the phase shift resulting from cavity sweeping whereas  $\phi(\lambda_{12})$  is the medium induced phase shift and  $p \in \mathcal{N}$ .

The first material to lase was GaAs and  $\text{Al}_x\text{Ga}_{1-x}\text{As}$  heterojunctions that have been fully studied (1). A new class of nitride-based materials (38)  $\text{Al}_x\text{Ga}_{1-x}\text{N}$  and  $\text{Al}_x\text{In}_{1-x}\text{N}$  emerged later and allowed to decrease the operating wavelength limit to about  $0.2 \mu\text{m}$  (1).

Laser diodes cover a large range from near ultra-violet to far infrared with compound semiconductor materials as displayed in Fig. 24.



**Fig.24:** (Color on-line) Laser emission wavelengths obtained with compound semiconductor materials. Visible region  $[0.380:0.780] \mu\text{m}$  in color is shown in order to compare to full wavelength coverage. Adapted from Sze (1).

Laser diodes have many applications in Fiber Communications, Guidance, Target Tracking, Metrology, Multimedia applications, Printing, Interferometry... however one interesting application in School, University, Business presentations is the unavoidable Laser pointer that requires also color diversity.

As in the LED case and for a long time Laser pointers were only red and again the issue of color diversity had to be addressed and solved with DPSSFD (Diode Pumped Solid State Frequency-Doubled) shortened to DPSS.

For example, a Green laser pointer light is generated with a high-power (typically 100-300 mW) infrared AlGaAs laser diode operating at 808 nm that pumps a Neodymium-doped Yttrium Vanadate (Nd:YVO4) crystal to lase in the infrared at 1064 nm.

Nd:YVO4 crystal is coated with a dielectric mirror that reflects at 1064 nm and transmits at 808 nm. The crystal 1064 nm output is fed into a KTP crystal (1) in the laser cavity resonator.

The Nd:YVO4 outputs polarized light and performs frequency doubling halving the wavelength to 532 nm (Green light). An infrared filter behind the mirror removes potentially harmful (to eyes) IR radiation from the output beam. The output power of most green laser pointers is on the order of 5 mW.

DPSSFD technique is used to create several colored laser pointers such as Yellow, Orange, Violet...

#### D. Photo-detectors

Photo-detectors like a photoconductor or a CCD (charge coupled device) can be considered as the converse of the LED since they transform light into charge carriers which is the opposite of the LED action produced also by other devices as displayed in Table. 10.

Photodetector	Junction Type	Gain	Response time (s)
Photo-conductor		1-10 <sup>6</sup>	10 <sup>-8</sup> - 10 <sup>-3</sup>
Photo-diodes	p-n junction	1	10 <sup>-11</sup>
	p-i-n junction	1	10 <sup>-10</sup> -10 <sup>-8</sup>
	Metal-semiconductor diode	1	10 <sup>-11</sup>
CCD		1	10 <sup>-11</sup> -10 <sup>-4</sup> (*)
Avalanche photodiode		10 <sup>2</sup> -10 <sup>4</sup>	10 <sup>-10</sup>
Photo-transistor		≈ 10 <sup>2</sup>	10 <sup>-6</sup>

**Table 10:** Typical values of Gain and Response Time for several common photodetectors. CCD is Charge-coupled device. p-i-n junction is a diode made by joining p-intrinsic-n materials. (\*) Limited by charge transfer. Large integration time is an advantage for CCD yielding high sensitivity. Adapted from Sze (1)

A CCD consists of a closely spaced array of MOS capacitors on a continuous insulating layer covering a semiconducting substrate, thus it is considered as zero polar since no transistor is involved in the CCD action.

A CCD can perform a wide range of electronic functions, including image sensing, video and signal processing. The operating principle of the CCD involves the charge storage and transfer actions controlled by gate electrodes. In fact, the CCD was initially targeted to be a Shift Register.

The CCD working principle is the following. It is first submitted to sufficiently large pulses applied to all the electrodes to produce surface depletion. Then a slightly larger bias is applied to a central electrode (CE) to produce larger depletion resulting in a potential well.

When electrons are injected from the CE left side, they will be collected by the potential well. If the potential of the CE right side electrode is increased to exceed that of the CE, electrons will be transferred from the CE to its right side one. Subsequently, carriers can be transferred successively along a linear array, which is the basic operation of a Shift Register, a Camera or a Video display.

Since a CCD does not obey scaling laws, it belongs to zero-polar family of devices, whereas MOS and CMOS are the archetypical unipolar family of the Moore type scalable device. This is the reason the CMOS device equivalent to a CCD has known tremendous development in its performance and scaling properties specially in sensors, e-tablets, smartphones, laptops, and cameras.

Sanguinetti *et al.* (43) used uniform illumination of smartphone camera image sensor by a LED to produce a number of photons generated per pixel.

Smartphone	ATIK 383L	Nokia N9
Noise, $\sigma_n$ ( $n$ )	10	3.3
Saturation ( $n$ )	$2 \times 10^4$	500
Illumination ( $n$ )	$1.5 \times 10^4$	410
Output bits per pixel	16	10

**Table 11:** Experimental parameters for two smartphone cameras employing a CCD (ATIK 383L) or a CMOS (Nokia N9). All data except last line, are expressed in number of electrons ( $n$ ). Adapted from Sanguinetti *et al.* (43).

The CCD image sensor has detection capability of 16 bits per pixel and a photon flux producing  $2 \times 10^4$  electrons per pixel, whereas the CMOS image sensor had a smaller photon flux since it has 10 bits per pixel with only 500 electrons per pixel (cf. Table 11).

## E. Spintronic devices

Basic devices belonging to this emerging technology are the diode and the transistor that we describe below by drawing analogies from the description of non-magnetic devices and extending them.

Barrier reduction occurs for both valence hole and conduction electron transport in a  $p-n$  diode under forward bias. Fig 25 shows that this leads to an increase in the conduction electron current to the left and the valence hole current to the right. Because the carriers have opposite charge, both increases result in an increased charge current to the right.

A spin-diode is the magnetic analogue of the electronic p-n junction. Replacing the p region by a ferromagnetic with  $\uparrow$  spin-polarized electrons, the n region by a ferromagnetic with  $\downarrow$  spin-polarized electrons and the SCR by a magnetic domain wall (23), requires analyzing the polarization changes across the device with an applied bias.

For the spin diode only the barrier for spin- $\uparrow$  electrons moving to the left is reduced, the barrier for spin- $\downarrow$  electrons moving to the right is increased. The charge current is thus directed to the right and the spin current to the left. Under reverse bias the barriers for carrier transport are both increased in the  $p-n$  diode (cf Fig 25), yielding rectification of the charge current. For the spin diode, one barrier is reduced and the other increased. Thus the charge current is not rectified but the spin current is.

Applying assumptions analogous to Shockley (1) pertaining to an ideal junction diode, the charge current density  $J_q$  and the spin current (44) density  $J_s$  are found to depend on applied voltage  $V$  according to (23):

$$J_q = 2eJ_0 \sinh(eV/k_B T), \quad (73)$$

$$J_s = 2\hbar J_0 \sinh^2(eV/2k_B T) = \hbar J_0 (\cosh(eV/k_B T) - 1) \quad (74)$$

where  $J_0 = Dn_m/L_m$ ,  $n_m$  is the minority carrier density, and  $L_m$  is the minority spin diffusion length. Note the different units of  $J_q, J_s$  as  $eJ_0, \hbar J_0$  respectively.

Using the hyperbolic identity:  $\tanh(z/2) = (\cosh z - 1)/\sinh z$  where  $z = eV/k_B T$ , the resulting current spin polarization (23) becomes:

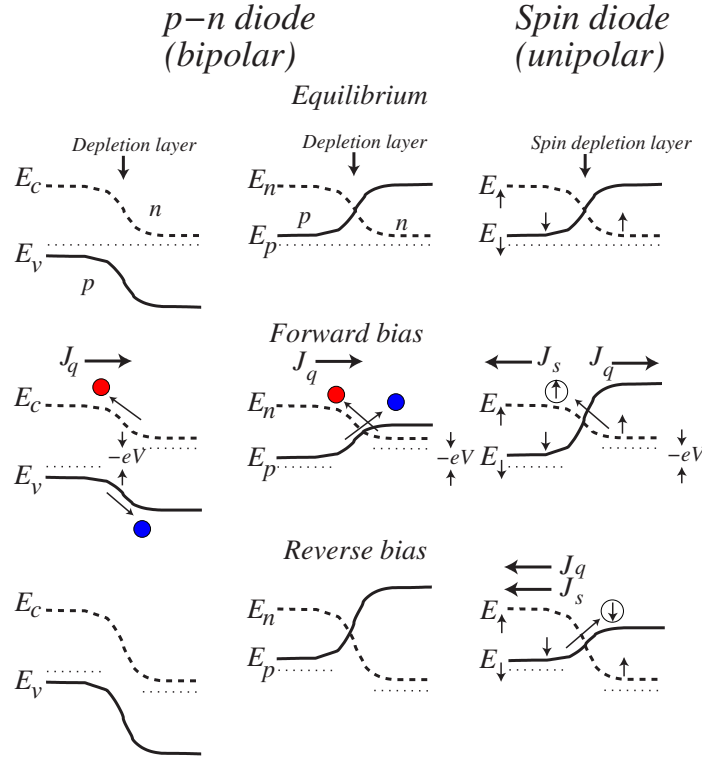
$$P = \frac{2e J_s}{\hbar J_q} = \frac{2e}{\hbar} \tanh(eV/2k_B T). \quad (75)$$

Thus the spin polarization approaches unity as  $V \rightarrow \infty$ , and approaches 0 when  $V \rightarrow 0$ . The relative directions of charge and spin currents, in the forward and reverse bias cases, are shown in Fig 25.

A spintronic device such as a spin transistor should have charge and spin current gain tunable with a magnetic field or preferably with an electric field using spin-orbit effect (23).

By simple analogy with a non-magnetic BJT (1), transistor analysis (23) gives the collector current density as:

$$I_C = -\frac{eJ_0}{\sinh(W/L_m)} [(\exp(-eV_{EB}/k_B T) - 1) - (\exp(-eV_{CB}/k_B T) - 1) \cosh(W/L_m)] - eJ_0 [\exp(eV_{CB}/k_B T) - 1] \quad (76)$$



**Fig.25:** Spin-diode diagram compared to an ordinary p-n diode. Electrons (in red) and holes (in blue) are likened to spin  $\uparrow$  and spin  $\downarrow$  carriers. The current densities  $J_q$  are for charge carriers whereas  $J_s$  are for spin carriers.

and the emitter current as:

$$I_E = -\frac{eJ_0}{\sinh(W/L_m)} [(\exp(-eV_{EB}/k_B T) - 1) \cosh(W/L_m) - (\exp(-eV_{CB}/k_B T) - 1)] + eJ_0 [\exp(eV_{EB}/k_B T) - 1] \quad (77)$$

Following standard notation (1), the base width is  $W$ , the voltage between emitter and base is  $V_{BE}$ . The voltage between collector and base is  $V_{CB}$ . The base current is  $I_B = I_E - I_C$ . When  $W/L_m$  is small,  $I_B \ll I_C$ , which is the situation required for transistor operation (current gain  $I_C/I_B \gg 1$ ).

Given acceptable values (23) of  $V_{EB}$  and  $V_{CB}$  ( $V_{EB} < 0$  and  $V_{CB} > 0$ )

$$I_C = -\frac{eJ_0}{\sinh(W/L_m)} [(\exp(-eV_{EB}/k_B T) - 1) + \cosh(W/L_m)] - eJ_0 [\exp(eV_{CB}/k_B T) - 1] \quad (78)$$

$$I_E = -\frac{eJ_0}{\sinh(W/L_m)} [(\exp(-eV_{EB}/k_B T) - 1) \cosh(W/L_m) + 1] - eJ_0 \quad (79)$$

The emitter efficiency, defined (23) as the ratio of the majority spin-direction charge current  $I_{E\downarrow}$  to the total emitter current  $I_E$  (1), is  $1 - \exp(eV_{EB}/k_B T)$  and thus very close to one.

However, in contrast to non-magnetic BJT, the collector multiplication factor  $M$ , defined as the ratio of the full collector current  $I_C$  to the majority spin-direction charge current  $I_{C\downarrow}$  (1), is given by:

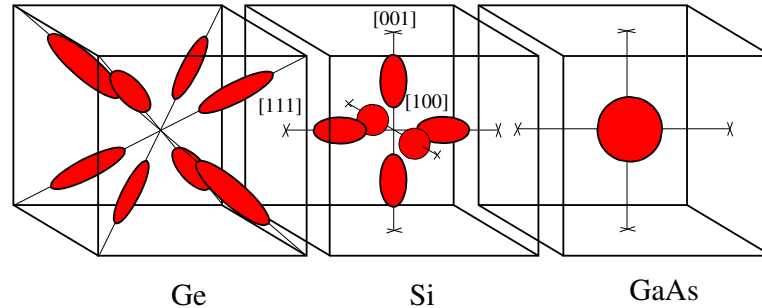
$$M = 1 + \sinh(W/L_m) \exp(e[V_{CB} + V_{EB}]/k_B T) \quad (80)$$

This implies  $M \approx 1$  only if  $W/L_m$  is small.

## F. Valleytronic devices

The conduction band of semiconductors contains valleys in the energy spectrum  $E(\mathbf{k})$  that are minima of constant energy surfaces close to the conduction band edge (cf. Fig. 2). Electrons can be trapped in the valleys since they can lower their energies, however they are not localized spatially but trapped into a preferred momentum. This implies the valleys channel charge flow in a particular way.

Typically valleys are ellipsoidal in shape for Ge, Si and spherical for GaAs as depicted in Fig. 26.



**Fig.26:** Band structure electronic valleys (in red) of Germanium, Silicon and GaAs. Valley shapes are ellipsoidal for indirect gap Ge, Si and spherical for direct gap GaAs. Their positions with respect to zone center are off for Ge, Si and on for GaAs (cf. Fig. 2). After Cardona *et al.* (2) and Kittel (6)

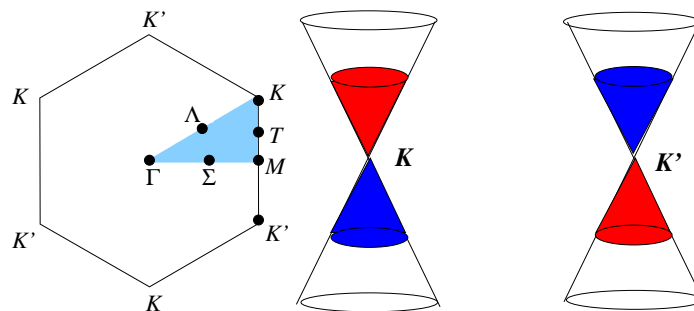
Hole valleys are the opposite of electron valleys in the sense they are near a maximum of the valence band.

If there are two or several conduction (or valence) band valleys in momentum space, then confining charge carriers in one of these valleys allows the fabrication of valleytronic devices possessing a novel electronic control parameter.

In addition to manipulating the charge or spin of electrons, another way to control electric current is by using the valley degree-of-freedom of electrons.

The first demonstration of the generation, transport and detection of valley-polarized electrons in bulk diamond was made by Isberg *et al.* (45) opening up novel ways to control quantum electronic devices.

In fact, 2D materials are more promising in this respect and Graphene is particularly interesting with valleys at  $K$  and  $K'$  points of the Brillouin zone (depicted in Fig. 27) where the shape of the valence-conduction bands look like a double cone each having a node at  $K$  or  $K'$  called Dirac points because of the relativistic speed of the carriers.



**Fig.27:** Graphene Brillouin zone showing alternating  $K$  and  $K'$  Dirac points where valence-conduction bands look like double cones. Upper-lower color ordering indicates valley index that is +1 for  $K$  and -1 for  $K'$ .

## APPENDIX A: Summary of Ge, Si and GaAs properties

Property	Ge	Si	GaAs
Atom/molecule density (/cm <sup>3</sup> )	4.42×10 <sup>22</sup>	5.0×10 <sup>22</sup>	4.42×10 <sup>22</sup>
Atomic/molecular weight	72.60	28.09	144.63
Crystal structure	Diamond	Diamond	Zinc-blende
Lattice constant, <i>a</i> (Angströms)	5.64613	5.43095	5.6533
Density (g/cm <sup>3</sup> )	5.3267	2.32S	5.32
Relative Dielectric constant	16.0	11.8	13.1
Intrinsic Diffusion constant (cm <sup>2</sup> /s)			
electron, <i>D<sub>n</sub></i>	100	39	220
hole, <i>D<sub>p</sub></i>	49	13	10
Effective DOS (/cm <sup>3</sup> )			
in conduction band, <i>N<sub>c</sub></i>	1.04×10 <sup>19</sup>	2.8×10 <sup>19</sup>	4.7×10 <sup>17</sup>
in valence band, <i>N<sub>v</sub></i>	6.0×10 <sup>18</sup>	1.04×10 <sup>19</sup>	7.0×10 <sup>18</sup>
DOS Effective mass ( <i>m<sub>0</sub></i> )			
electron, <i>m<sub>de</sub><sup>*</sup></i>	0.22	1.18	0.067
hole, <i>m<sub>dh</sub><sup>*</sup></i>	0.18	0.81	0.53
Electron affinity, <i>χ</i> (V)	4.0	4.05	4.07
Band gap, <i>E<sub>G</sub></i> (eV)	0.67	1.12	1.42
Refraction index	4.0	3.4	3.3
Intrinsic carrier concentration, <i>n<sub>i</sub></i> (/cm <sup>3</sup> )	2.4×10 <sup>13</sup>	1.08×10 <sup>10</sup>	2.1×10 <sup>6</sup>
Linear thermal expansion (/° C)	5.8×10 <sup>-6</sup>	2.6×10 <sup>-6</sup>	6.86×10 <sup>-6</sup>
Melting point (° C)	937	1415	1238
Mobility (cm <sup>2</sup> /V-s)			
electron <i>μ<sub>n</sub></i>	3,900	1,500	8,500
hole <i>μ<sub>p</sub></i>	1,900	500	400
Optical-phonon energy (eV)	0.037	0.063	0.035
Specific heat (J/g-° C)	0.31	0.7	0.35
Thermal conductivity (W/cm-° C)	0.6	1.5	0.46
Thermal diffusivity (cm <sup>2</sup> /s)	0.36	0.9	0.44
Breakdown field (V/cm)	≈ 10 <sup>5</sup>	≈ 3×10 <sup>5</sup>	≈ 4×10 <sup>5</sup>

**Table 12:** Physical properties of Ge, Si, GaAs at room temperature (300K). DOS is Density of States. Adapted from Sze (1) and M A Green, "Intrinsic concentration, effective densities of states, and effective mass in silicon," J Appl Phys , **67**, 2944 (1990).

## References

- [1] S.M. Sze, *Physics of Semiconductor Devices*, 2nd edition, Wiley, New York (1981)
- [2] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors*, Springer-Verlag (2005).
- [3] M. Grundman, *The Physics of Semiconductors*, Springer, Berlin-Heidelberg, (2006).
- [4] Undergraduate Quantum Mechanics textbooks: E. Merzbacher, *Quantum Mechanics*, Wiley New-York (1970); G. Baym, *Lectures on Quantum Mechanics*, Westview Press, Perseus, New-York (1990); R. Liboff, *Introductory Quantum Mechanics*, Addison Wesley, 4th edition (2002); D. J. Griffiths, *Introduction to Quantum Mechanics* Benjamin-Cummings, 2nd edition (2004); B.H. Bransden and C.J. Joachain *Quantum Mechanics*, Benjamin Cummings, 2nd edition (2000).
- [5] Graduate Quantum Mechanics textbooks: J. J. Sakurai *Modern Quantum Mechanics*, Revised edition by S. F. Tuan, Addison-Wesley, New-York (1994); K. Gottfried, T. M. Yan *Quantum Mechanics: Fundamentals*, 2nd edition, Springer, New-York (2003).
- [6] C. Kittel, *Introduction to Solid State Physics*, 8th edition, Wiley (2008).
- [7] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems*, McGraw-Hill, New York (1971).
- [8] J C Slater 1930 *Phys. Rev.* **35** 210, J C Slater (1951a) *Phys. Rev.* **81** 385, J C Slater (1951b) *Phys. Rev.* **82** 538, J C Slater (1953) *Phys. Rev.* **91** 528, J C Slater (1974) *Quantum Theory of Molecules and Solids*, **IV**, *The Self-consistent field for Molecules and Solids* (McGraw-Hill, New York)
- [9] A. Feiguin, *Modern Computational Methods in Solids*, Physics 5870, University of Wyoming (2011).
- [10] E. Kasper and D. J. Paul, *Resumé of Semiconductor Physics*, in *Silicon Quantum Integrated Circuits: Silicon-Germanium Heterostructure Devices: Basics and Realisations*, pp 49-116, Springer Berlin Heidelberg (2005).

- [11] W Kohn and L J Sham, Phys. Rev. **140** A1133 (1965), L J Sham and W Kohn, Phys. Rev. **145** 561-7 (1966)
- [12] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part1*, Third Edition, **5**, Pergamon, Oxford (1976). E. M. Lifshitz and L.P. Pitaevskii *Statistical Physics, Part2*, Second Edition, **9**, Pergamon, Oxford (1980).
- [13] B Das, S Datta and R Reifenberger, Phys. Rev. B **41**, 8278 (1990).
- [14] F. Reif, *Statistical and Thermal Physics*, McGraw-Hill, New-York (1985).
- [15] D. A. MacQuarrie, The Chemical Educator, **1**, 1, Springer-Verlag, New-York (1996).
- [16] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, C. W. Clark, *NIST Handbook of Mathematical Functions* Cambridge University Press, New York, (2010).
- [17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, New-York (1980).
- [18] M. Abramowitz and I.S Stegun, *Handbook of Mathematical Tables*, Dover, New-York (1960).
- [19] G.A. Prinz, Science **282**, 1660 (1998); S.A. Wolf et al, Science **294**, 1488 (2001).
- [20] I. vZutić, J. Fabian, and S.D. Sarma, Rev. Mod. Phys. **76**, 323 (2004).
- [21] R. Coehoorn, Chapter 1 in Handbook of Magnetic Materials, edited by K.H.J. Buschow (2003)
- [22] D. Culcer, J. Sinova, N. A. Sinitsyn, T. Jungwirth, A. H. MacDonald and Q. Niu Phys. Rev. Lett. **93**, 046602 (2004).
- [23] M. E. Flatté and G. Vignale, Appl. Phys. Lett. **78**, 1273 (2001), G. Vignale and M. E. Flatté, Phys. Rev. Lett. **89**, 098302 (2002), M. Deutsch, G. Vignale and M. E. Flatté, J. App. Phys. **96**, 7424 (2004).
- [24] S.J Polak, C. Den Heijer, W.H.A. Schilders and P. Markowich, *Semiconductor Device Modeling from the numerical point of view*, Int. J. for Num. Methods in Engineering, **24**, 763-838 (1987).
- [25] C. A. Ringhofer and C. Schmeiser, *A modified Gummel Method for the basic Semiconductor Equations*, IEEE Transactions CAD **7**, 251-253 (1988).
- [26] P. Markowich and C. A. Ringhofer, *A Singularly Perturbed boundary value problem modelling a semiconductor device*, SIAM J. Appl. Math. **44**, 231-256 (1984).
- [27] U.M. Ascher, R.M. Mattheij and R. D. Russel, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Prentice-Hall (Englewood Cliffs).
- [28] U.M. Ascher and R. Weiss, *Collocation for singular perturbation problems I: First-order systems with constant coefficients*, SIAM J. Num. Analysis **20**, 537-557 (1983).
- [29] M. S. Mock, *On the convergence of Gummel numerical algorithm*, Solid-State Electronics, **15**, 1-4 (1972).
- [30] C. P. Please, *An analysis of Semiconductor PN junctions*, IMA Journal of Applied Mathematics, **28**, 301-318 (1982).
- [31] A. De Mari, *An accurate numerical steady-state one dimensional solution of the PN junction*, Solid State Electronics, **11**, 33-58 (1968).
- [32] V. Arandjelovic, *Accurate Numerical Steady-State solutions for a diffused one dimensional junction diode*, Solid State Electronics, **13**, 865-871 (1970).
- [33] D. L. Scharfetter and H.K. Gummel, *Large-Signal Analysis of a Silicon Read Diode Oscillator*, IEEE Transactions ED-16, 64-77 (1969).
- [34] P. Markowitch and C. Schmeiser, *Uniform asymptotic representation of solutions of the basic semiconductor-device equations*, IMA Journal of Applied Mathematics, **36**, 43-57 (1986).
- [35] P. Markowich and C. A. Ringhofer, *Collocation methods for boundary value problems on long intervals*, Math. Comp. **40**, 123-150 (1983).
- [36] Y. He and G. Cao, *A generalized Scharfetter-Gummel method to eliminate Cross-Wind effects*, IEEE Transactions CAD-10, 1579-1582 (1991).
- [37] G. Catalan, D. Jiménez and A. Gruverman, Nature Materials, **14**, 137 (2015).
- [38] S. Nakamura, *Nobel Lecture: Background story of the invention of efficient blue InGaN light emitting diodes*, Rev. Mod. Phys. **87**, 1139 (2015).
- [39] Report of the 21st meeting (23-24 February 2012) of the Consultative Committee for Photometry and Radiometry (CCPR), Bureau International des Poids et Mesures (2012).
- [40] J. M. Palmer, Chapter 36: *Radiometry and Photometry: Units and Conversions*, p. 36.10 in *Handbook of Optics*, Third edition, edited by M. Bass, SPIE and McGraw-Hill (New-York, 2010).
- [41] J. P. Gordon, H. J. Zeiger and C. H. Townes, *The Maser-New Type of Microwave Amplifier, Frequency Standard, and Spectrometer*, Phys. Rev. **99**, 1264 (1955).
- [42] H. Haken, Light, **2**, *Waves, Photons, Atoms*, North-Holland (1981).
- [43] B. Sanguinetti, A. Martin, H. Zbinden and N. Gisin, Phys. Rev. X **4**, 031056 (2014).
- [44] G. Tatara, H. Kohno and J. Shibata, Physics Reports, **468**, 213 (2008).
- [45] H. Löfas, A. Grigoriev, J. Isberg R. and Ahuja, AIP Advances **1**, 032139 (2011) and Nature Mater. **12**, 760-764 (2013)