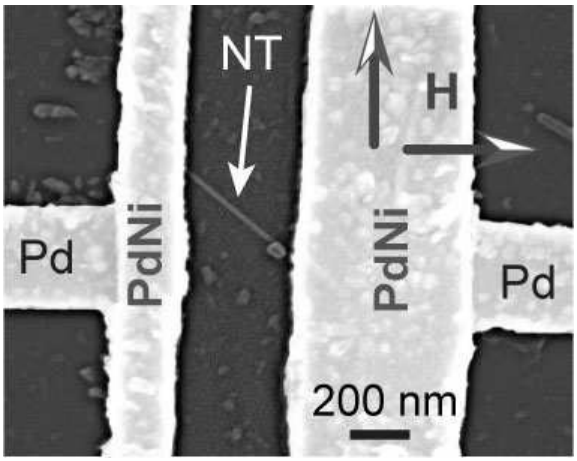
ECOLE NATIONALE SUPÉRIEURE DE L'ELECTRONIQUE ET DE SES APPLICATIONS

Le cours de physique de l'ENSEA

III. Physique des composants à semi-conducteurs



Année 2006-2007

http ://www-reynal.ensea.fr/teaching/semicond/

Auteurs, anciens auteurs et/ou enseignants : affiliations et coordonnées

Ont participé à la rédaction de ce cours, sont intervenus ou continuent à intervenir dans l'enseignement de la physique des composants à semi-conducteurs à l'ENSEA :

→ Sylvain Reynal (Laboratoire de Physique Théorique et Modélisation, UMR 8089 du CNRS) - coordination polycopié, co-auteur parties I et II, et site web du cours ; `reynal@ptm.u-cergy.fr` ;

→ Emmanuelle Bourdel (Equipe Circuit Instrumentation et Modélisation Electronique, ENSEA) ; co-auteur partie I ; `bourdel@ensea.fr`

→ Christophe Brun (Laboratoire de Photonique et de Nanostructures - UPR 20 du CNRS, Marcoussis) ; moniteur TD ;

→ Laurence Croizé (Laboratoire de Physique Moléculaire pour l'Atmosphère et l'Astrophysique, UMR 7092 du CNRS) ; moniteur TD ;

→ Bruno Darracq ; co-auteur partie II et site web du cours ; `darracq@ensea.fr` ;

→ Frédérique Giannini ; `giannini@ensea.fr` ;

→ Abdelak Kasbari ; `kasbari@ensea.fr` ;

→ Jean Gaubert (ancien co-auteur partie II - L2MP, Polytech' Marseille) ;

→ Bertrand David (ancien co-auteur partie II - département TSI de l'ENST).

Illustration de couverture : photographie SEM (scanning electron microscope) d'un spin-field effect transistor de dimension quasi-nanométrique réalisé à partir d'un nanotube (flèche "NT") et de deux contacts ferromagnétiques en alliage PdNi. (c) Institut für Physik, Universität Basel, Suisse.

Table des matières

I Semiconducteurs **7**

1 Notions de cristallographie **11**

1.1 Géométrie du réseau cristallin 11

1.1.1 Définitions 11

1.1.2 Réseaux de Bravais 13

1.1.3 Systèmes cubiques 13

1.1.4 Directions cristallographiques - indices de Miller 13

1.2 Réseau réciproque - Zones de Brillouin 16

1.2.1 Réseau à une dimension 16

1.2.2 Réseau tridimensionnel 18

1.3 Diffraction des rayons X - Loi de Bragg 19

2 Théorie des bandes et dynamique de l'électron **23**

2.1 Introduction 23

2.2 Calcul de la structure de bandes d'énergie 23

2.2.1 Approximations 25

2.2.2 Théorème de Bloch 26

2.3 Caractéristiques de la structure de bandes d'énergie et première zone de Brillouin 29

2.4 Conditions aux limites périodiques et impact sur la quantification de l'énergie 30

2.5 Bandes d'énergie : modèle des liaisons fortes et effet des impuretés 31

2.5.1 Modèle 32

2.5.2 Solutions générales pour les états stationnaires 36

2.5.3 Bandes d'énergie en l'absence d'impureté 36

2.5.4 Dynamique et vitesse de groupe 37

2.5.5 Solutions en présence de l'impureté 39

2.5.6 Piégeage par l'impureté 39

2.6 Bande de conduction et bande de valence 42

2.7 Isolant, semi-conducteur et métal 42

2.8 Electron en présence d'un champ électrique - masse effective 42

2.8.1 Action d'une force extérieure - théorème accélération 43

2.8.2 Masse effective 44

2.8.3 Étude des extréma de la fonction $E_n(\mathbf{k})$ 45

2.8.4 Courant dans une bande pleine 46

2.8.5 Notion de trou 46

3 Notions de physique statistique **49**
3.1 Fonctions d'états des systèmes ouverts 50
3.1.1 Variables d'état conjuguées 50
3.1.2 Énergie interne pour un système ouvert 51
3.1.3 Enthalpie libre pour un système ouvert 51
3.2 Caractérisation microscopique d'un système 52
3.2.1 Micro-états 52
3.2.2 Travail et chaleur à l'échelle microscopique 53
3.3 Entropie micro-canonique 55
3.4 Entropie d'information 59
3.4.1 Comment quantifier l'information manquante ? 59
3.4.2 Principe d'information minimum 60
3.4.3 Application à l'évolution d'un système isolé 61
3.5 Équilibres thermodynamiques 62
3.5.1 Équilibre thermique 62
3.5.2 Équilibre mécanique 64
3.5.3 Équilibre diffusif 65
3.6 Systèmes non-isolés 66
3.6.1 Systèmes en contact avec une source de température - statistique de Boltzmann 66
3.6.2 Systèmes en contact avec une source de température et un réservoir de particules - statistique de Gibbs 69

4 Semi-conducteur intrinsèque, dopage **73**
4.1 Semi-conducteur intrinsèque 73
4.1.1 Étude en fonction de la température 73
4.1.2 Semi-conducteur vs. métal 74
4.1.3 Populations de porteurs de charge dans les bandes 75
4.2 Semi-conducteurs dopés 84
4.2.1 Semi-conducteurs de type N 85
4.2.2 Semi-conducteurs de type P 89

5 Phénomènes de transport **93**
5.1 Modèle de Drude de la conduction et de la diffusion 94
5.1.1 Conduction électrique 94
5.1.2 Diffusion électrique 97
5.2 Équation de conservation 100

6 Recombinaisons **103**
6.1 Absorption de la lumière par les semi-conducteurs 103
6.1.1 Transitions directes 104
6.1.2 Coefficient d'absorption pour une transition directe entre extréma . 105
6.1.3 Transitions indirectes 106
6.2 Processus de génération-recombinaison 107
6.2.1 Approche qualitative 107
6.2.2 Taux de génération-recombinaison 108
6.2.3 Durée de vie des porteurs en excès 110
6.2.4 Recombinaison par centres recombinants (ou niveaux pièges) . . . 111

II Composants **115**

7 Du matériau homogène au composant **117**
7.1 Introduction . . . 117
7.2 Diagramme des bandes en présence d'un potentiel électrostatique "macroscopique" ou diagramme d'Anderson . . . 117
7.3 Les jonctions . . . 120
7.3.1 Schéma des bandes d'énergie d'une homojonction PN . . . 120
7.3.2 Étude quantitative de la jonction PN a l'équilibre . . . 121
7.4 Semi-conducteurs hors de l'équilibre thermodynamique . . . 125
7.4.1 Efficacite d'injection d'une jonction pn . . . 127
7.4.2 Relation courant-tension . . . 128
7.4.3 Approximation de boltzmann . . . 130

8 La jonction PN **131**
8.1 Régime statique . . . 131
8.1.1 Étude qualitative a partir du schema des bandes d'énergie . . . 131
8.1.2 La jonction PN sans recombinaisons . . . 133
8.1.3 Effet des recombinaisons . . . 139
8.1.4 Tenue en tension inverse . . . 142
8.2 Régime dynamique . . . 144
8.2.1 Analyse qualitative . . . 144
8.2.2 Analyse quantitative . . . 147
8.2.3 Commutation d'une jonction pn sur charge resistive . . . 153

9 Le transistor bipolaire à homojonctions **157**
9.1 Analyse qualitative . . . 157
9.2 Modèle equivalent en regime statique . . . 158
9.2.1 Cas des faibles niveaux d'injection . . . 160
9.3 Modes de fonctionnement du transistor bipolaire . . . 161
9.3.1 Mode bloqué : $V_{EB} < 0$ et $V_{CB} < 0$ . . . 161
9.3.2 Mode normal : $V_{EB} > 0$ et $V_{CB} < 0$ . . . 161
9.3.3 Mode inversé : $V_{EB} < 0$ et $V_{CB} > 0$ . . . 162
9.3.4 Mode saturé : $V_{EB} > 0$ et $V_{CB} > 0$ . . . 163
9.4 Caractéristiques statiques du transistor bipolaire . . . 163
9.5 Modèle dynamique du transistor bipolaire . . . 164
9.5.1 Charges stockées dans la région quasi-neutre de base . . . 165
9.5.2 Charges stockées dans les régions dépeuplées . . . 165
9.5.3 Modèle équivalent en régime dynamique . . . 166
9.6 Schéma équivalent dynamique petit signal en régime normal dans la configuration émetteur-commun . . . 166
9.6.1 Conductance base-émetteur . . . 166
9.6.2 Transconductance . . . 167
9.6.3 Capacité de diffusion . . . 167
9.6.4 Capacités de transition . . . 167
9.7 Le transistor en commutation . . . 168
9.7.1 Mise en conduction . . . 168
9.7.2 Bloquage . . . 170
9.8 Exercice d'application du chapitre : Analyse d'un transistor bipolaire NPN . 171

10 Les transistors à effet de champ **175**
10.1 Le transistor à effet de champ de jonction (JFET) 175
10.1.1 Principe de fonctionnement et structure 175
10.1.2 Caractéristiques statiques 176
10.1.3 Caractéristiques dynamiques 179
10.2 Le transistor mos (metal oxyde semi-conducteur) 180
10.2.1 Principe de fonctionnement et structure 180
10.2.2 Inversion des populations de porteurs en surface d'une structure MOS idéale 180
10.2.3 Calcul de la caractéristique courant tension d'un MOS idéal 181
10.2.4 Modèle équivalent simplifié 184
10.2.5 Structure MOS réelle 185
10.3 Exercices d'application du chapitre 186

A Éléments de fabrication des circuits intégrés **189**
A.1 Opérations technologiques élémentaires 189
A.2 Réalisation du substrat semi-conducteur 189
A.3 Réalisation des couches de semi-conducteur 190
A.3.1 La diffusion 190
A.3.2 L'implantation ionique 190
A.3.3 L'epitaxie 190
A.4 Réalisation de couches isolantes 191
A.5 Réalisation des interconnexions 191
A.6 Lithographie et gravure des couches 191
A.7 Étapes technologiques de la filiere cmos 192
A.8 Remarques 192

B Données numériques **199**
B.1 Caractéristiques des principaux semi-conducteurs à 300k 199

Première partie

Semiconducteurs

Introduction

La physique des semi-conducteurs est l'une des branches les plus riches et ouvertes de la physique des solides. L'intérêt particulier porté à ce type de matériau provient essentiellement de propriétés électroniques très intéressantes. En effet, le semi-conducteur est un matériau plutôt isolant intrinsèquement, mais qui sous certaines conditions peut conduire le courant électrique. On peut notamment en modifier la conductivité en y introduisant en quantité infime certains atomes, dits *dopants*, de nature différente de ceux composant le semi-conducteur.

L'industrie des circuits et composants électroniques à semi-conducteurs[1] est un des piliers fondamentaux de l'économie depuis les années 70. Elle est notamment à la base du développement d'un grand nombre de technologies modernes parmi lesquelles l'informatique, les télécommunications, le traitement de l'information et l'automatique.

Cette industrie est basée sur deux concepts : la miniaturisation et l'imprimerie[2]. La miniaturisation est rendue possible par le fait que les vecteurs d'information sont des particules élémentaires : les électrons. Les circuits et composants étant essentiellement à géométrie planaire à l'heure actuelle, la reproduction à grande échelle des circuits à été rendue possible grâce aux procédés classiques d'imprimerie transposés aux échelles micrométriques.

Les composants électroniques mettent à profit les propriétés dynamiques des électrons dans les semi-conducteurs. Il faut donc, pour tenter de comprendre le fonctionnement de ces composants, préciser ces propriétés et définir les grandeurs physiques dont les évolutions conditionnent les propriétés électriques ou optiques des composants. Les paramètres fondamentaux d'une telle étude sont d'une part l'état de la population d'électrons à l'équilibre thermodynamique (qui dépend fortement de la structure des niveaux d'énergie), et d'autre part l'évolution de cette population lorsqu'intervient une perturbation. On distingue deux types de perturbations : les perturbations douces qui donnent lieu aux phénomènes de transport, et les perturbations dures qui sont à l'origine des phénomènes de génération-recombinaison.

[1] Solid-state devices

[2] Pour cette dernière, il semble que la fin de la partie ait bientôt sonné avec l'apparition de techniques de croissance d'autoorganisée permettant de construire les composants nanométriques atome par atome...

Chapitre 1

Notions de cristallographie

Les matériaux semi-conducteurs que nous étudions dans ce cours sont exclusivement des solides cristallins. Leur propriétés électroniques, optiques, mécaniques et thermiques sont donc intimement liées à la géométrie particulière sous-tendant l'arrangement des atomes au sein du matériau : le réseau cristallin. La branche de la physique permettant de caractériser cette géométrie est la cristallographie, dont l'objectif du présent chapitre est de présenter les notions essentielles.

Pour l'essentiel, nous exploiterons ces notions d'abord lors de l'étude de la structure de bandes d'énergie des matériaux semi-conducteurs, puis dans le cadre de la modélisation des phénomènes d'interaction rayonnement-matière au sein de ces matériaux.

1.1 Géométrie du réseau cristallin

1.1.1 Définitions

L'état cristallin correspond à un arrangement ordonné d'atomes ou de molécules qui constituent le solide (fig. 1.1)[1]

La géométrie d'un solide cristallin est définie par :

[1]Dans les cristaux liquides, l'ordre est partiel : seul l'ordre résultant de l'orientation des molécule est conservé, tandis que l'agencement géométrique des molécule dans l'espace est, comme dans toute phase liquide, désordonné.

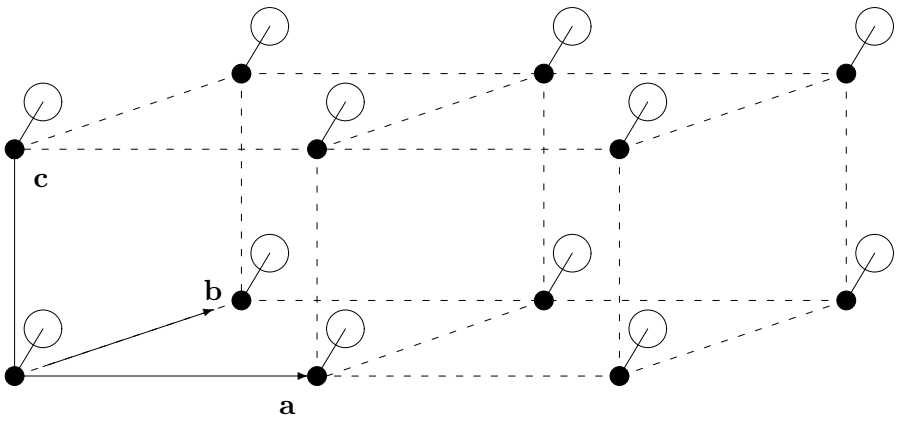


FIG. 1.1 – Un cristal est un arrangement ordonné d'atomes ou des molécules.

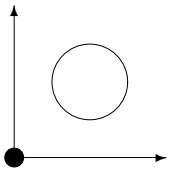


FIG. 1.2 – Un motif de base est constitué d'un groupe d'atomes ou de molécules.

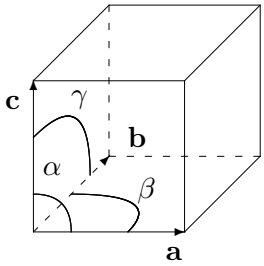


FIG. 1.3 – La maille primitive est décrite par 3 vecteurs de base ainsi que 3 angles.

- Un motif[2] de base (fig. 1.2) qui correspond à un groupe d'atomes qui se répète périodiquement pour constituer le cristal ;
- Un réseau[3] cristallin constitué de nœuds auquel est "accroché" le motif ; en dimension 3, ce réseau est caractérisé par trois vecteurs de base $\mathbf{a}, \mathbf{b}, \mathbf{c}$, qui permettent de générer l'ensemble des nœuds par des translations élémentaires à partir d'un nœud-origine. Ces vecteurs sont définis par leur longueur a,b,c ainsi que par les angles α, β, γ qu'ils font entre eux.

Les trois vecteurs précédent sous-tendent (fig. 1.3) un volume appelé **maille** du réseau. La maille **élémentaire**[4] (ou primitive en franglais) est la maille de plus petit volume parmi toutes les mailles permettant de générer les nœuds d'un réseau donné. Selon le degré de complexité géométrique du réseau, elle peut, ou non, être confondue avec la maille conventionnelle, qui est la maille utilisée par convention (et souvent par simplicité) pour décrire un réseau particulier.

La position de chaque motif du cristal est donc défini par le vecteur translation $\mathbf{T}$:

$$\mathbf{T} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$$

avec n_1, n_2, n_3 entiers, et le volume de la maille élémentaire est donné par :

$$\Omega = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$$

En outre, on appelle **rangée réticulaire** une rangée de nœuds et **plan réticulaire** un plan passant par des nœuds du réseau.

Il est important de comprendre que le réseau cristallin est un modèle idéal de cristal : en effet, les **nœuds du réseau cristallin** correspondent à la position d'équilibre des atomes du cristal (ou de son motif, dans le cas général). En toute rigueur, le réseau cristallin ne "coïncide" donc avec le cristal réel qu'à température nulle, *i.e.* lorsque l'agitation thermique est inexistante. A température non-nulle, un cristal n'est donc jamais rigoureusement périodique : ces écarts à la périodicité feront l'objet d'une attention particulière lors de l'étude des phénomènes de transport, parce qu'elle influent fortement sur la conductivité électrique du matériau.

[2] Anglais : pattern
[3] Anglais : lattice
[4] Anglais : primitive cell

1.1.2 Réseaux de Bravais

Les réseaux de Bravais sont des familles de réseaux permettant de classer aisément la plupart des réseaux cristallins[5] selon leurs symétries. Il existe 7 familles de géométrie cristalline, caractérisées par des degrés de **symétrie** plus ou moins élevés :

$cubique$	$a = b = c \quad \alpha = \beta = \gamma = 90^\circ$
$ttragonal$	$a = b \neq c \quad \alpha = \beta = \gamma = 90^\circ$
$orthorhombique$	$a \neq b \neq c \quad \alpha = \beta = \gamma = 90^\circ$
$monoclinique$	$a \neq b \neq c \quad \alpha = \beta = 90^\circ \neq \gamma$
$triclinique$	$a \neq b \neq c \quad \alpha \neq \beta \neq \gamma$
$rhombodrique\ (ou\ trigonal)$	$a = b = c \quad \alpha = \beta = \gamma \neq 90^\circ$
$hexagonal$	$a = b \neq c \quad \alpha = \beta = 90^\circ\, \gamma = 60^\circ$

Ces sept familles se répartissent en 14 modes de réseaux de Bravais, certaines familles possédant des sous-familles de symétrie différente (fig. 1.4).

1.1.3 Systèmes cubiques

Le système cubique est le système dans lequel cristallisent la plupart des semi-conducteurs (germanium, silicium, arseniure de gallium, ...). Il possède une symétrie d'ordre 4, *i.e.* il est invariant par rotation de $2\pi/4$. Il existe trois réseaux cubiques dont la maille conventionnelle est un cube d'arête a :

- le cubique simple (*simple cubic* ou SC) : un motif de base à chaque sommet du cube (figure 1.5) ;
- le cubique centré (*centered cubic* ou CC) : un motif de base à chaque sommet d'un cube et un motif au centre du cube (figure 1.6) ;
- le cubique à faces centrées (*face centered cubic* ou FCC) : un motif à chaque sommet et un au centre de chaque face d'un cube (figure 1.7).

Dans le cas du réseau SC, la maille conventionnelle et la maille élémentaire sont identiques, ce qui n'est pas le cas pour les réseaux CC et FCC. Par exemple, si l'on note l la longueur de l'arête du cube (maille conventionnelle), la maille élémentaire du réseau FCC est sous-tendue par les trois vecteurs suivants : $\mathbf{a} = (l/2, l/2, 0)$, $\mathbf{b} = (0, l/2, l/2)$ et $\mathbf{c} = l/2, 0, l/2$. Le volume généré est alors un **rhomboèdre** (trois angles non-droits, vecteurs de même longueur), dont le volume vaut $\Omega = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = l^3/4$. Ce volume est bien inférieur au volume l^3 de la maille conventionnelle. On se convaincra aisément (cf. TD "Cristallographie") que cette maille élémentaire permet de générer tous les nœuds du réseau.

1.1.4 Directions cristallographiques - indices de Miller

Du fait des symétries cristallines, certaines directions sont privilégiées. La normale à un plan réticulaire définit une direction particulière. Les **indices de Miller** permettent de préciser cette direction (et donc ce plan réticulaire) pour un réseau donné.

Pour définir les indices de Miller d'un plan réticulaire donné (cf. exemple figure 1.8), il faut :

- déterminer les coordonnés des points d'intersection du plan réticulaire avec les axes définis par les vecteurs de la maille (généralement la maille conventionnelle) : ex. $(2, 4, 3)$ ou $(1, 2, 3/2)$. Si l'intersection n'est pas un point, on prend ∞.
- prendre l'inverse de ces coordonnées : ex. $(1/2, 1/4, 1/3)$ ou $(1, 1/2, 2/3)$.

[5] A l'exclusion des quasi-cristaux

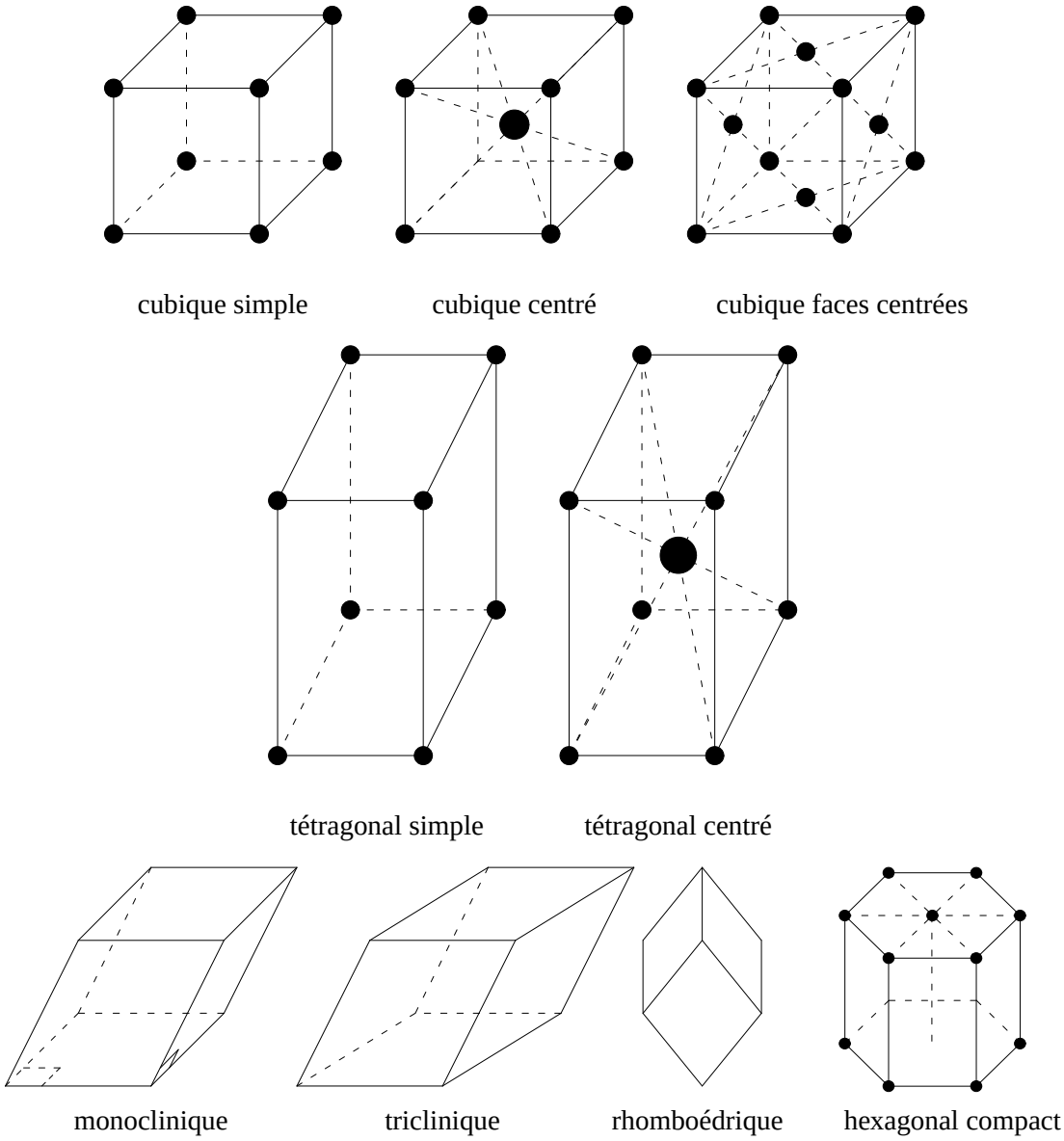


FIG. 1.4 – Réseaux de Bravais

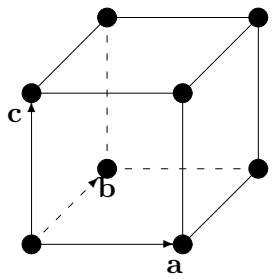


FIG. 1.5 – Maille conventionnelle (et également élémentaire) du réseau cubique simple : $a = b = c$

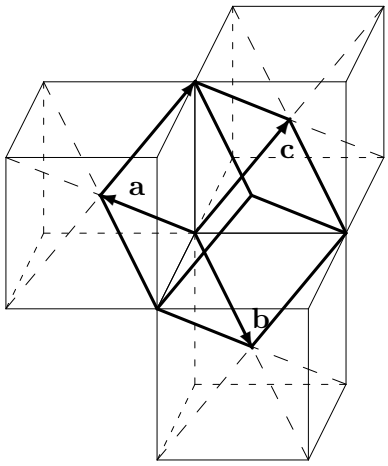


FIG. 1.6 – Maille élémentaire du réseau cubique centré (BC) : $a = b = c = l^*\sqrt{3}/2$, où l est la longueur de l'arête du cube. Cette maille n'est PAS un cube, contrairement à la maille conventionnelle...

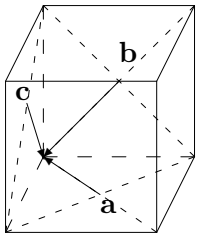


FIG. 1.7 – Maille élémentaire du réseau cubique faces-centrées (FCC) : $a = b = c = l^*/\sqrt{2}$, où l est la longueur de l'arête du cube. Cette maille élémentaire n'est PAS un cube, contrairement à la maille conventionnelle...

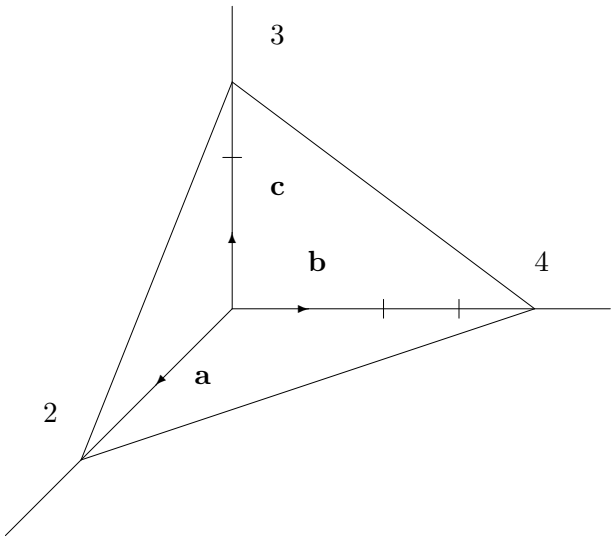


FIG. 1.8 – Indices de Miller du plan dessiné : $[6, 3, 4]$

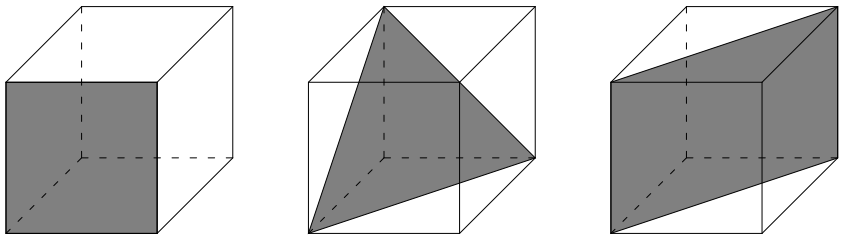


FIG. 1.9 – Les 3 plans réticulaires les plus courants pour les réseaux cubiques.

– ramener au trois plus petits entiers proportionnels en multipliant par une fraction rationnelle adhoc : ex. $[6, 3, 4]$. Lorsqu'une coordonnée est négative, on ajoute une barre au-dessus de l'indice correspondant : ex. $[\bar{1}, 0, 1]$.

Un triplet d'indices de Miller permet donc de caractériser une famille de plan réticulaires parallèles entre eux.

1.2 Réseau réciproque - Zones de Brillouin

Les symétries cristallines se retrouvent dans les propriétés des grandeurs physiques du cristal. En particulier, la **périodicité** de la structure cristalline entraîne que, sauf cas particulier, les grandeurs physiques caractérisant le cristal (densité, charge électrique, aimantation, ...) sont périodiques, et sont par conséquent décomposables en série de Fourier.

1.2.1 Réseau à une dimension

Une grandeur physique ρ (par exemple la densité de charge électrique) égale à $\rho(x)$ en un point x d'un réseau unidimensionnel est telle que :

$$\rho(x) = \rho(x + na), \forall n \in \mathbb{Z}$$

où a est la période (ou le pas) du réseau. Sa décomposition en série de Fourier s'écrit :

$$\rho(x) = \sum_{p=-\infty}^{+\infty} \rho_n e^{ip\frac{2\pi}{a}x}$$

soit encore,

$$\rho(x) = \sum_{k} \rho_k e^{ikx},\ k = p\frac{2\pi}{a},\ p \in \mathbb{Z}$$

avec

$$\rho_k = \frac{1}{a}\int_0^a \rho(x) e^{-ikx}\, dx$$

Le point d'abscisse $k = p2\pi/a$ appartient à l'espace de Fourier que l'on appelle **espace réciproque**. Plus précisément, ce point définit à l'intérieur de cet espace réciproque un nœud d'un réseau de pas $2\pi/a$, que l'on appelle **réseau réciproque**. Par exemple, le nœud d'abscisse $3 \cdot 2\pi/a$ définit la position de la troisième "harmonique" de la décomposition en série de Fourier de $\rho(x)$, et le nœud à l'origine, la "composante continue". Pour lever toute ambiguïté, nous appellerons désormais **espace direct** l'espace réel dans lequel existe le cristal, et **réseau direct** le réseau cristallin : l'espace et le réseau réciproques sont, a contrario, des objets abstraits définis pour les besoins de la décomposition en série de Fourier. Notez enfin

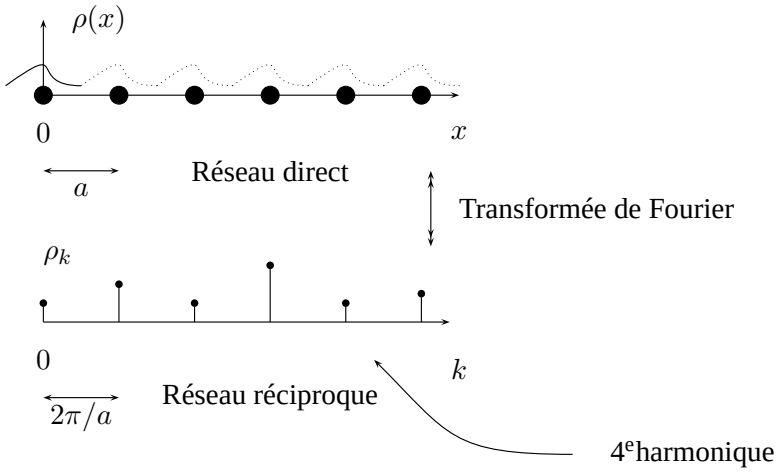


FIG. 1.10 – Décomposition en série de Fourier d'un grandeur périodique $\rho(x)$ (densité de charge électrique,...) du réseau direct. Les "raies" ρ_k sont dessinées sur le réseau réciproque de pas $2\pi/a$.

que, par ailleurs, le réseau réciproque peut être défini comme la transformée de Fourier du réseau direct. Ainsi, dans l'espace réciproque, les distances se mesurent-elles en m^{-1} (cf. en Hz en traitement du signal). Dans le cas unidimensionnel, l'analogie avec le cours de signal échantillonné est triviale, car le réseau direct est un peigne de Dirac de pas a, et le réseau réciproque un peigne de Dirac de pas $2\pi/a$. La décomposition en série de Fourier d'une grandeur périodique $\rho(x)$ revient à prendre la transformée de Fourier d'un peigne de Dirac de pas a convolué avec le motif $\rho_0(x)$ définit sur l'intervalle $[0, a)$: il s'agit donc de la transformée de Fourier du motif, $\widetilde{\rho_0}(k)$, multipliée par un peigne de Dirac de pas $2\pi/a$, c'est-à-dire "échantillonnée" en chaque nœud du réseau réciproque.

1.2.2 Réseau tridimensionnel

Dans le cas d'une grandeur physique $\rho(\mathbf{r})$ telle que :

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + \mathbf{T})$$

où $\mathbf{T}$ est un vecteur quelconque du réseau direct, la décomposition en série de Fourier donne, si et seulement si le réseau est **cubique** (cf. remarque ci-dessous) :

$$\rho(\mathbf{r}) = \sum_{\mathbf{k}} \rho_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}},\ \mathbf{k} = p_1\frac{2\pi}{a} + p_2\frac{2\pi}{b} + p_3\frac{2\pi}{c},\ \{p_1, p_2, p_3\} \in \mathbb{Z}^3$$

avec

$$\rho_{\mathbf{k}} = \frac{1}{\Omega}\int_{\Omega} \rho(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3\mathbf{r}$$

l'intégrale portant sur tout les volume de la maille élémentaire.

L'espace réciproque peut donc être subdivisé en mailles élémentaires de dimension $(2\pi/a, 2\pi/b, 2\pi/c)$. La maille élémentaire de l'espace réciproque centrée sur l'origine de l'espace réciproque est appelée **première zone de Brillouin**. Le réseau réciproque est donc totalement défini par sa période (= son pas) et par la première zone de Brillouin.

Dans le cas général (*i.e.* quelque soit le type de réseau, cubique, hexagonal, etc...), les vecteurs de base du réseau réciproque $(\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*)$ sont donnés par :

$$\mathbf{a}^* = \frac{2\pi}{\Omega}\mathbf{b}\times\mathbf{c},\ \ \mathbf{b}^* = \frac{2\pi}{\Omega}\mathbf{c}\times\mathbf{a},\ \ \mathbf{c}^* = \frac{2\pi}{\Omega}\mathbf{a}\times\mathbf{b} \tag{1.1}$$

où $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ sont les vecteurs de base du réseau direct, et Ω est le volume de la maille élémentaire du réseau direct :

$$\Omega = \mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$$

Le volume Ω^* de la maille élémentaire du réseau réciproque est donc :

$$\Omega^* = \mathbf{a}^*\cdot\mathbf{b}^*\times\mathbf{c}^* = \frac{8\pi^3}{\Omega}$$

et la décomposition d'une grandeur périodique du réseau en série de Fourier s'écrit, dans le cas général (*i.e.* quelque soit la géométrie du réseau) :

$$\boxed{\rho(\mathbf{r}) = \sum_{\mathbf{k}} \rho_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}},\ \mathbf{k} = p_1\mathbf{a}^* + p_2\mathbf{b}^* + p_3\mathbf{c}^*,\ \{p_1, p_2, p_3\} \in \mathbb{Z}^3}$$

la somme portant sur **tous** les vecteurs du réseau réciproque. En outre, on montre aisément que les formules précédentes donnant les vecteurs du réseau réciproque sont telles que, si $\mathbf{T}$ est un vecteur du réseau direct, et $\mathbf{T}^*$ un vecteur du réseau réciproque, alors

$$\boxed{\mathbf{T}\cdot\mathbf{T}^* = 2\pi}$$

Dans le cas des systèmes cubiques il est simple de montrer que :
- la maille élémentaire du réseau réciproque d'un cubique simple est cubique simple
- la maille élémentaire du réseau réciproque d'un cubique centré est cubique à faces centrées
- la maille élémentaire du réseau réciproque d'un cubique à faces centrées est de type cubique centré.

1.3 Diffraction des rayons X - Loi de Bragg

L'analyse de la diffraction de rayons X par les matériaux cristallins constitue une méthode puissante pour l'étude de la géométrie du réseau.

Une première approche qualitative permet de montrer (cf. figure 1.11) que les pics de diffraction sont obtenus pour des angles tels que :

$$2d\sin\theta = n\lambda,\ \forall n \in \mathbb{Z}$$

où λ est la longueur du rayonnement X, et d la distance entre deux plans réticulaires.

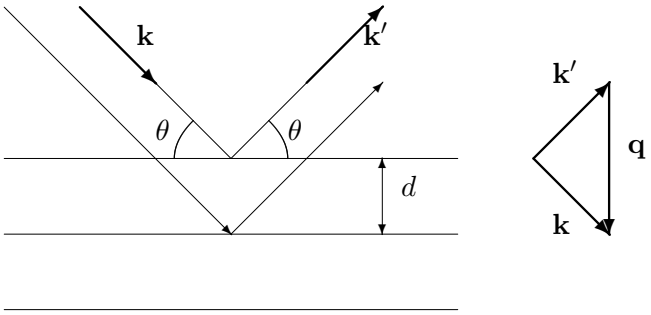


FIG. 1.11 – Loi de Bragg. $\mathbf{k}$ et $\mathbf{k}'$ sont les vecteurs d'onde des ondes incidentes et émergentes, respectivement.

Le vecteur $\mathbf{q}$ est tel que :

$$\mathbf{k} = \mathbf{k}' + \mathbf{q}$$

On sait que la quantité de mouvement des photons X vaut $\mathbf{p} = \hbar\mathbf{k}$, d'où

$$\mathbf{p} = \mathbf{p}' + \hbar\mathbf{q}$$

ce qui montre que la quantité $\hbar\mathbf{q}$ représente la **quantité de mouvement** absorbée par le cristal lors du processus de diffusion. Nous allons montrer que ce vecteur $\mathbf{q}$ doit satisfaire à une relation géométrique très générale pour que la diffusion ait lieu, relation dont la loi de Bragg constitue un cas particulier.

Une analyse plus quantitative du mécanisme de diffusion repose sur la **section efficace différentielle de diffusion**, que l'on définit par :

$$\frac{d\sigma}{d\Omega}(\mathbf{q}) \sim \frac{2\pi}{\hbar} |\left\langle \mathbf{k} \right| U(\mathbf{r}) \left| \mathbf{k}' \right\rangle|^2$$

où :

- $U(\mathbf{r})$ représente le **potentiel d'interaction** entre le photon X (*i.e.* l'onde électromagnétique qu'il "porte") et les atomes du cristal ; ce potentiel est de nature essentiellement électrostatique (dipolaire), les atomes interagissant avec le champ électrique de l'onde via leur distribution de charge électrique (il existe des termes d'interaction magnétique, mais d'intensité beaucoup plus faible) ;
- $|\mathbf{k}\rangle$ et $|\mathbf{k}'\rangle$ représentent les *kets* associés aux ondes électromagnétiques planes de vecteurs d'onde $\mathbf{k}$ et $\mathbf{k}'$;

La section efficace différentielle de diffusion indique la probabilité de diffusion d'un photon incident dans l'angle solide élémentaire $d\Omega = 2\pi \sin\theta \; d\theta$ sous-tendu par le vecteur $\mathbf{k}'$ (ou de manière équivalent le vecteur $\mathbf{q}$ une fois $\mathbf{k}$ fixé).

Remarque 1.1 *L'expression précédente n'est en toute rigueur valable que si le potentiel d'interaction $U(\mathbf{r})$ est faible devant l'Hamiltonien H_0 du cristal (cf. formule des perturbations, cours de Physique Quantique de première année).*

En explicitant le terme $\langle \mathbf{k}| \, U(\mathbf{r}) \, |\mathbf{k}'\rangle$, on obtient :

$$\frac{d\sigma}{d\Omega}(\mathbf{q}) \sim \frac{2\pi}{\hbar} \left| \int_V U(\mathbf{r}) e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}} d^3\mathbf{r} \right|^2$$

où l'intégrale de volume porte sur tout le volume du cristal, soit encore

$$\frac{d\sigma}{d\Omega}(\mathbf{q}) \sim \frac{2\pi}{\hbar} \left| \int_V U(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r} \right|^2$$

A un facteur près, la probabilité de diffusion dans la direction $\mathbf{k}'$ telle que $\mathbf{k}' = \mathbf{k} - \mathbf{q}$ est donc égale au carré de la transformé de Fourier $\widetilde{U}(\mathbf{q})$ du potentiel d'interaction entre le cristal et l'onde électromagnétique.

Le caractère périodique de ce potentiel permet d'expliciter davantage l'expression précédente. En effet, le potentiel peut s'écrire

$$U(\mathbf{r}) = \sum_{\mathbf{T}} U_a(\mathbf{r} - \mathbf{T}), \; \mathbf{T} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}$$

(la somme porte sur tous les vecteurs du réseau direct) où $U_a(\mathbf{r})$ est le potentiel d'interaction entre **un** atome et l'onde électromagnétique. La transformé de Fourier de $U(\mathbf{r})$ s'écrit alors :

$$\begin{aligned}
\widetilde{U}(\mathbf{q}) &= \sum_{\mathbf{T}} \int_V U_a(\mathbf{r} - \mathbf{T}) e^{-i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r} \\
&= \sum_{\mathbf{T}} e^{-i\mathbf{q}\cdot\mathbf{T}} \int_V U_a(\mathbf{r} - \mathbf{T}) e^{-i\mathbf{q}\cdot(\mathbf{r}-\mathbf{T})} d^3\mathbf{r} \\
&= \sum_{\mathbf{T}} e^{-i\mathbf{q}\cdot\mathbf{T}} \int_V U_a(\mathbf{r}') e^{-i\mathbf{q}\cdot\mathbf{r}'} d^3\mathbf{r}' \text{ (Jacobien=1, cristal infini)} \\
&= \left(\sum_{\mathbf{T}} e^{-i\mathbf{q}\cdot\mathbf{T}} \right) \widetilde{U}_a(\mathbf{q})
\end{aligned}$$

où

$$\widetilde{U}_a(\mathbf{q}) = \int_V U_a(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} d^3\mathbf{r}$$

est la transformée de Fourier du potentiel d'interaction entre **un** atome et l'onde électromagnétique. Pour un réseau unidimensionnel, l'expression

$$\sum_{\mathbf{T}} e^{-i\mathbf{q}\cdot\mathbf{T}}$$

s'écrit

$$\sum_{n=-\infty}^{+\infty} e^{-iqna} = \frac{2\pi}{a} \sum_{m\in\mathbb{Z}} \delta(q - m\frac{2\pi}{a})$$

(on peut aussi voir cette expression comme la transformée de Fourier d'un peigne de Dirac de période a, qui est précisément un peigne de Dirac de période $2\pi/a$ - par rapport aux tables de transformées données en traitement du signal, il faut faire $t \to x$, $f \to q/2\pi$, $\omega \to q$ et $T \to a$).

La généralisation aux réseaux tridimensionnels donne :

$$\sum_{\mathbf{T}} e^{-i\mathbf{q}\cdot\mathbf{T}} = \frac{8\pi^3}{\Omega} \sum_{\mathbf{T}^*} \delta(\mathbf{q} - \mathbf{T}^*),\ \mathbf{T}^* = p_1\mathbf{a}^* + p_2\mathbf{b}^* + p_3\mathbf{c}^*,\ \{p_1, p_2, p_3\} \in \mathbb{Z}^3$$

soit finalement

$$\widetilde{U}(\mathbf{q}) \sim \sum_{\mathbf{T}^*} \delta(\mathbf{q} - \mathbf{T}^*)\widetilde{U}_a(\mathbf{T}^*)$$

la somme portant sur tous les vecteurs $\mathbf{T}^*$ du réseau réciproque. Finalement, la section efficace de diffusion s'écrit

$$\frac{d\sigma}{d\Omega}(\mathbf{q}) \sim \sum_{\mathbf{T}^*} \delta(\mathbf{q} - \mathbf{T}^*) \left|\widetilde{U}_a(\mathbf{T}^*)\right|^2$$

Les pics de diffusion correspondront donc à des vecteurs $\mathbf{q}$ confondus avec des vecteurs du réseau réciproque, l'intensité de diffusion étant proportionnelle à "l'harmonique" donnée par le vecteur $\mathbf{q} = \mathbf{T}^*$ correspondant. Naturellement, les harmoniques de rang élevé, correspondant à des vecteurs $\mathbf{q}$ de forte amplitude, sont de plus en plus faibles : la diffusion a donc lieu essentiellement pour des vecteurs $\mathbf{q}$ d'amplitude proche de la maille du réseau réciproque.

Pour des rayons X, la diffusion est **élastique**, *i.e.* sans changement d'énergie cinétique du photon diffusé : on a donc $k \sim k'$. Or, lorsqu'il y a diffusion, $\mathbf{q} = \mathbf{T}^*$ où $\mathbf{T}^*$ est un vecteur du réseau réciproque ; ainsi

$$\mathbf{k}' = \mathbf{k} + \mathbf{T}^*$$

d'où

$$k'^2 - k^2 = T^{*2} - 2\mathbf{k}\cdot\mathbf{T}^* \sim 0$$

soit finalement

$$\left(\frac{\mathbf{T}^*}{2}\right)^2 = \mathbf{k}\cdot\frac{\mathbf{T}^*}{2}$$

Pour qu'il y ait diffusion, il faut donc :

- qu'il existe un vecteur $\mathbf{T}^*$ perpendiculaire au plan bissecteur de $\mathbf{k}$ et $\mathbf{k}'$
- que la projection de $\mathbf{k}$ sur ce vecteur T^* soit de longueur $T^*/2$.

On se convaincra à l'aide de la figure 1.12 et en raisonnant sur un cristal à deux dimensions, que ces deux conditions sont équivalente à la condition que le vecteur $\mathbf{k}$ doit se trouver sur la frontière d'une zone de Brillouin.

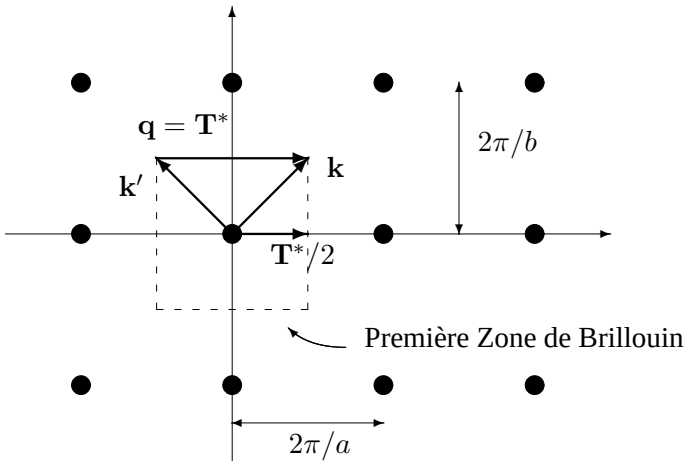


FIG. 1.12 – Condition de diffusion : le vecteur **k** doit se trouver sur la frontière d'une zone de Brillouin. Attention ! Les disques noirs représentent les points du réseau **réciproque**, pas les atomes du réseau direct !

Chapitre 2

Théorie des bandes et dynamique de l'électron

2.1 Introduction

La théorie des bandes d'énergie permet d'expliquer le comportement électrique de certains matériaux, et de justifier entre autres l'origine physique de la différence entre isolant, conducteur et semi-conducteur. Pour comprendre les propriétés électroniques des matériaux semi-conducteurs, il faut étudier le comportement des électrons dans ces matériaux. La première étape est de déterminer l'état des électrons à $T = 0\ K$ dans le cristal à l'aide de la mécanique quantique. Le nombre de particules mises en jeu (électrons et noyaux de tous les atomes présents dans un volume donné de cristal) est en fait beaucoup trop grand et conduit à un nombre d'équations tel qu'il est impossible de résoudre rigoureusement le problème. En effet, le nombre de variables est donné par $3(Z+1)N$ avec N le nombre d'atomes dans le cristal et Z le numéro atomique des atomes du cristal. Par exemple, dans le cas du Silicium, le numéro atomique est 14, c'est-à-dire qu'il possède 14 électrons autour de son noyau, chaque particule possédant trois degrés de liberté (*i.e.* , 3 coordonnées de position). Dans le Silicium il y a $5\ 10^{22}$ atomes par cm^3. Le nombre de variables du problème est donc de $3(Z+1)5\ 10^{22}$, c'est à dire de l'ordre de $2\ 10^{24}$ par cm^3 de Silicium !

2.2 Calcul de la structure de bandes d'énergie

Un cristal est un **système à** $N(Z+1)$ **particules** constitué d'électrons et de noyaux, ces derniers étant considérés comme des particules ponctuelles. D'un point de vue quantique, ce système est décrit à l'instant t par une fonction d'onde

$$\psi(t, \{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) = \psi(t, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{R}_1, \mathbf{R}_2, \ldots)$$

où $\mathbf{r}_i$ est la position de l'électron i et $\mathbf{R}_\alpha$ la position du noyau α. L'interprétation de cette fonction d'onde, $\psi(t, \{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\})$, est un peu plus compliquée que dans le cas d'un système à 1 particule (électron dans un puits de potentiel,...) : le terme $|\psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\})|^2$ représente la densité volumique de **probabilité de présence conjointe** de **chaque** électron i en $\mathbf{r}_i$ **et** de **chaque** noyau α en $\mathbf{R}_\alpha$, à l'instant t, ou encore, en termes un peu plus imagés, la probabilité que le système soit à l'instant t dans une configuration telle que chaque électron i se trouve en $\mathbf{r}_i$ et chaque noyau α en $\mathbf{R}_\alpha$.

Pour déterminer les états stationnaires de ce système[1], on résout l'équation aux valeurs propres de l'énergie :

$$H[\varphi_n(\{\mathbf{r}_i\},\{\mathbf{R}_\alpha\})] = E_n\varphi_n(\{\mathbf{r}_i\},\{\mathbf{R}_\alpha\})$$

où l'opérateur H est l'Hamiltonien du système (opérateur associé à l'énergie totale[2]), et E_n est la nième valeur propre de H, *i.e.* , le **niveau d'énergie** E_n **du système**. Ce Hamiltonien comporte les termes suivants :

- énergies cinétiques des électrons :

$$T_e = \sum_i \frac{P_e^2}{2m_e} = \sum_i \frac{-\hbar^2}{2m_e}\Delta_i$$

  où le Laplacien Δ_i est défini par rapport aux coordonnées de l'électron i (*i.e.* , $\Delta_i = \partial^2/\partial x_i^2 + \ldots$);
- énergies cinétiques des noyaux :

$$T_N = \sum_\alpha \frac{P_N^2}{2M_N} = \sum_\alpha \frac{-\hbar^2}{2M_N}\Delta_\alpha$$

  où M_N est la masse d'un noyau.
- énergies potentielles d'interaction (répulsives, donc positives) entre électrons :

$$V_e = \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}$$

- énergies potentielles d'interaction (répulsives, donc positives également) entre noyaux :

$$V_N = \frac{1}{2}\sum_{\alpha\neq\beta}\frac{Z^2e^2}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}$$

- énergies potentielles d'interaction (attractives, donc négative) entre noyaux et électrons :

$$V_{eN} = -\frac{1}{2}\sum_{i,\alpha}\frac{Ze^2}{|\mathbf{R}_\alpha - \mathbf{r}_i|}$$

- énergies potentielles de toutes les particules dans le champ électrostatique macroscopique (i.e. imposé par une f.e.m. externe : pile, GBF, ...) :

$$V = V(\mathbf{r}_1,\ldots,\mathbf{R}_1,\ldots)$$

Par la suite on prendra $V(\mathbf{r}_1,\ldots,\mathbf{R}_1,\ldots) = 0$, *i.e.* on considérera un cristal non-polarisé électriquement.

[1] Rappel : un état stationnaire est un couple formé d'un niveau d'énergie E_n et de la fonction d'onde associée φ_n

[2] Ce qu'on appelle énergie mécanique E_m en mécanique classique

2.2.1 Approximations

Le nombre de variables du problème étant beaucoup trop grand, il est nécessaire d'effectuer des approximations pour pouvoir évaluer les états du système d'électrons. La première approximation consiste à séparer la fonction d'onde du système en deux fonctions d'onde indépendantes, l'une décrivant l'état de l'ensemble des électrons et l'autre décrivant l'ensemble des noyaux. En effet les électrons et les noyaux ont des comportements fortement différents du fait d'un rapport de masse important. Les électrons ont des vitesses nettement supérieures à celles des noyaux et on peut donc utiliser l'approximation de Born-Oppenheimer[3] pour séparer la fonction d'onde des électrons de celle des noyaux.

Pour cela on suppose que (vu des électrons) les noyaux sont immobiles, situés à leur position d'équilibre aux nœuds du réseau cristallin $\mathbf{R}_\alpha^0$. On pose donc :

$$\varphi(\{\mathbf{r}_i\},\{\mathbf{R}_\alpha\}) = \varphi_e(\{\mathbf{r}_i\})\varphi_N(\{\mathbf{R}_\alpha^0\})$$

ce qui signifie que la dynamique des électrons d'une part, celle des noyaux d'autre part, sont indépendantes : les fonctions d'onde, et partant, les probabilités de présence, le sont aussi. L'équation aux valeurs propres de l'énergie s'écrit donc :

$$(T_e + T_N + V_e + V_N + V_{eN})\varphi_e(\{\mathbf{r}_i\})\varphi_N(\{\mathbf{R}_\alpha\}) = E\varphi_e(\{\mathbf{r}_i\})\varphi_N(\{\mathbf{R}_\alpha\})$$

que l'on peut réécrire, après séparation des variables, *i.e.* en tenant compte du fait que T_e et V_e n'agissent que sur les $\{\mathbf{r}_i\}$, tandis que T_N et V_N n'agissent que sur les $\{\mathbf{R}_\alpha\}$:

$$(T_e + V_e + V_{eN})\varphi_e(\{\mathbf{r}_i\}) = E_n^e\varphi_e(\{\mathbf{r}_i\}) \tag{2.1}$$

et

$$(T_N + V_N)\varphi_N(\{\mathbf{R}_\alpha\}) = E_m^N\varphi_N(\{\mathbf{R}_\alpha\})$$

Remarque 2.1 *Les électrons étant des **fermions**, ils sont soumis au principe de Pauli (on ne peut "placer" qu'un seul fermion dans un état quantique donné). Pour satisfaire cette contrainte, la fonction d'onde du système d'électrons $\varphi_e(\{\mathbf{r}_i\})$ doit nécessairement être **antisymétrique** par rapport aux variables de position $\mathbf{r}_i$, ce qui signifie que permuter deux électrons n et m,* i.e. *permuter un couple $(\mathbf{r}_n, \mathbf{r}_m)$ inverse le signe de $\varphi_e(\{\mathbf{r}_i\})$.*

*Considérons maintenant, par exemple, le cas où les deux électrons 1 et 2 sont placés dans le même état quantique : alors $\mathbf{r}_1 = \mathbf{r}_2$ (*i.e. *les probabilités de présence sont les mêmes). Par conséquent, permuter les positions des électrons 1 et 2 dans l'expression de φ_e laisse cette fonction d'onde inchangée : en effet, on a*

$$\varphi_e(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\ldots) = \varphi_e(\mathbf{r}_2,\mathbf{r}_1,\mathbf{r}_3,\ldots) = \varphi_e(\mathbf{r}_1,\mathbf{r}_1,\mathbf{r}_3,\ldots)$$

Or φ_e est antisymétrique, donc elle doit également changer de signe lors de cette permutation, ce qui implique que

$$\varphi_e(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\ldots) = -\varphi_e(\mathbf{r}_2,\mathbf{r}_1,\mathbf{r}_3,\ldots)$$

soit encore,

$$\varphi_e(\mathbf{r}_1,\mathbf{r}_1,\mathbf{r}_3,\ldots) = -\varphi_e(\mathbf{r}_1,\mathbf{r}_1,\mathbf{r}_3,\ldots)$$

ce qui impose la nullité de φ_e lorsque $\mathbf{r}_1 = \mathbf{r}_2$! Par conséquent, la probabilité de trouver les deux électrons 1 et 2 dans le même état quantique est ... nulle. Le fait d'imposer l'antisymétrie de $\varphi_e(\mathbf{r}_i)$ par rapport aux $\mathbf{r}_i$ suffit donc à assurer le respect du principe de Pauli.

[3] Egalement appellée approximation adiabatique.

Notez en outre que si l'on tenait compte du spin, alors il faudrait assurer l'antisymétrie non seulement par rapport aux positions, mais aussi par rapport aux variables de spin $\{\mathbf{S}_i\}$: conséquences, 1) deux électrons peuvent alors se trouver au "même endroit", à la condition que leurs spins soit anti-alignés ; 2) deux électrons n'ont aucune probabilité d'être à la même position si leurs spins sont alignés. Une des conséquences les plus connues de ce phénomène est la liaison covalente : les deux électrons appariés se trouvent dans la même région de l'espace, mais avec des spins anti-alignés.

On obtient donc deux relations indépendantes où E_n^e représente l'énergie des électrons qui se déplacent dans le champ créé par des noyaux fixes. Elles traduisent le fait que, pour les électrons qui sont des particules très rapides, c'est la position instantanée des noyaux qui influe sur leur état à travers le terme V_{eN}.

Pour les noyaux, leur mouvement lent n'est pas affecté par la position instantanée des électrons mais par leur mouvement d'ensemble à travers le terme E_n^e. Autrement dit, les noyaux se déplacent "lentement" dans le champ **moyen** produit par tous les électrons, et ce mouvement lent des noyaux entraîne à son tour l'ensemble de tous les électrons (en plus du mouvement propre beaucoup plus "rapide" de ces derniers).

Pour déterminer les états d'énergie possibles pour les électrons il faut résoudre l'équation aux dérivées partielles (2.1). Mais, étant donné le nombre considérable de variables, il est nécessaire de simplifier encore le problème. Pour cela on utilise l'approximation de Hartree-Fock qui consiste à ramener le problème à un seul électron en supposant que chaque électron évolue dans un potentiel créé par l'ensemble des autres électrons "fixés" à leur position moyenne (d'où également le terme de **champ moyen** appliqué à cette approximation). Dans le cadre de cette approximation, on ramène l'étude à un seul électron de fonction d'onde $\varphi(\mathbf{r})$ évoluant dans un potentiel $V_c(\mathbf{r})$ appelé **potentiel cristallin** et qui comprend l'énergie d'interaction électron-noyaux et les interactions électrons-électrons.

L'équation à résoudre devient donc celle d'un problème **à une particule** :

$$-\frac{\hbar^2}{2m_e}\Delta\varphi_n(\mathbf{r}) + V_c(\mathbf{r})\varphi_n(\mathbf{r}) = E_n\varphi_n(\mathbf{r}) \tag{2.2}$$

où l'énergie potentielle V_c dépend **elle-même** des fonctions propres de l'Hamiltonien du problème (puisque celles-ci indiqueront finalement la position moyenne des électrons, donc la topologie du champ moyen créé par eux, donc V_c). Or les fonctions propres en question dépendant de l'Hamiltonien, on est donc en présence d'un bouclage [4] et plusieurs méthodes ont été proposées pour résoudre l'équation précédente (l'une d'elle consistant à calculer les solutions par approximations successives, en partie de fonctions d'onde arbitraires mais "raisonnables", par exemple les orbitales atomiques).

L'étude de ces méthodes dépasse le cadre de ce cours et ne sera donc pas abordée. Nous allons en revanche étudier les propriétés des états électroniques et des fonctions d'onde associées.

2.2.2 Théorème de Bloch

L'énergie E d'un électron dans un solide et sa fonction d'onde ϕ sont donnés par l'équation aux valeurs propres de l'énergie (équation 2.2). Dans un réseau cristallin (*i.e.* à structure périodique), l'énergie potentielle des électrons (ou potentiel cristallin) est une fonction périodique de même période que le cristal :

$$V_c(\mathbf{r}) = V_c(\mathbf{r} + \mathbf{T})$$

[4] Pour cette raison, cette méthode est également appelée **méthode du champ auto-consistant/self-consistent field**.

où **T** représente le vecteur translation du réseau cristallin (cf. chapitre précédent).

La périodicité du potentiel cristallin entraîne une propriété fondamentale de la fonction d'onde des électrons donnée par le **théorème de Bloch** :

Théorème 2.1 (Théorème de Bloch) *Les fonctions d'onde stationnaires d'un électron dans un cristal périodique sont des* ***fonctions de Bloch*** *de la forme*

$$\varphi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$$

avec

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{T})$$

et **k** *un vecteur quelconque représentant le* ***vecteur d'onde*** *de l'électron dans le cristal,* i.e. *,* $\hbar\mathbf{k}$ *représente la quantité de mouvement de l'électron dans le cristal.*

Le théorème de Bloch montre que la fonction d'onde d'un électron est une onde plane (terme en $e^{i\mathbf{k}\cdot\mathbf{r}}$) modulée en amplitude par une modulation périodique $u_{\mathbf{k}}(\mathbf{r})$.

Démonstration :

Pour simplifier le raisonnement, nous nous limitons ici au cas d'un cristal **unidimensionnel** de maille a. On a besoin pour la suite de la démonstration de l'opérateur T_u, que l'on définit ainsi :

$$T_u = e^{\frac{i}{\hbar}uP_x}$$

u est un nombre réel homogène à une longueur, et P_x représente l'observable de quantité de mouvement selon l'axe Ox (on rappelle que $\mathbf{P} = \frac{\hbar}{i}\nabla$). Nous allons montrer que T_u agit sur toute fonction d'onde électronique en la translatant de u vers l'arrière. Tout d'abord, nous avons

$$T_u = e^{u\frac{d}{dx}}$$

d'après la définition de P_x. Ensuite, nous exploitons le développement en série de Taylor de la fonction exponentielle,

$$e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$$

ce qui nous donne[5], avec $z \to u\frac{d}{dx}$,

$$T_u = \sum_n \frac{u^n}{n!}\frac{d^n}{dx^n}$$

En appliquant T_u à une fonction d'onde $\varphi(x)$, nous obtenons

$$T_u[\varphi(x)] = \sum_n \frac{u^n}{n!}\varphi^{(n)}(x) = \varphi(x+u)$$

ce qui prouve le résultat.

L'étape suivante consiste à montrer que l'hamiltonien H du cristal et l'opérateur T_u commutent lorsque u est égal à la maille cristalline a. Souvenons-nous que H est composé

[5] En admettant que le développement en série est toujours valide bien que nous ayons désormais affaire à une fonction d'un opérateur...

de deux terme : l'énergie cinétique $P_x^2/2m$, et l'énergie potentielle $V(X)$, avec $V(X) = V(X+a)$. D'où

$$[H, T_u] = \left[\frac{P_x^2}{2m} + V(X), T_u\right] = \left[\frac{P_x^2}{2m}, T_u\right] + [V(X), T_u] = [V(X), T_u]$$

Il est en effet trivial que T_u commute avec le terme d'énergie cinétique, puisque T_u est une fonction de P_x uniquement. Reste donc à calculer le commutateur de $V(X)$ et T_u. La méthode directe consiste à calculer successivement l'action de $V(X)T_u$ et $T_uV(X)$ sur une fonction d'onde $f(x)$, et à comparer les deux résultats :

$$V(X)T_uf(x) = V(X)f(x+u) = V(x)f(x+u)$$

et

$$T_uV(X)f(x) = T_u[V(x)f(x)] = V(x+u)f(x+u)$$

Les deux résultats sont identiques si $V(x) = V(x+u)$, *i.e.* , si $u = a$. Nous avons donc le résultat fondamental suivant dans un cristal unidimensionnel de maille a :

$$\boxed{[H, T_a] = 0} \tag{2.3}$$

Ce résultat se généralise aisément au cas d'un cristal tridimensionnel de maille $(\mathbf{a}, \mathbf{b}, \mathbf{c})$, sous la forme

$$\forall \mathbf{R} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \text{ tel que } (n_1, n_2, n_3) \in \mathbb{N}^3, [H, T_{\mathbf{R}}] = 0$$

La dernière étape de la démonstration consiste à montrer que les états stationnaires (qui sont les états propres de l'hamiltonien H) sont des fonctions de Bloch. Tout d'abord, le fait que deux opérateurs commutent implique qu'il existe une base d'états propres **commune** aux deux opérateurs, *i.e.* pour toute fonction d'onde $\varphi(x)$ appartenant à cette base, nous aurons

$$H[\varphi(x)] = E\varphi(x)$$

et

$$T_a[\varphi(x)] = \lambda\varphi(x)$$

où E et λ sont les valeurs propres associées à l'état propre $\varphi(x)$. La valeur propre E représente, nous le savons, le niveau d'énergie associé à $\varphi(x)$. Mais que représente λ ? Pour répondre à cette question, cherchons les vecteurs propres de T_a : ce sont les fonctions $f(x)$ telles que

$$f(x+a) = \lambda f(x)\ \forall x$$

$f(x)$ étant de carré sommable, l'égalité précédente implique que λ est de module 1. En effet,

$$\begin{aligned}\langle f|\ f\rangle &= \int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |f(y+a)|^2 d(y+a) \\ &= \int_{-\infty}^{\infty} |f(y+a)|^2 dy = |\lambda|^2 \int_{-\infty}^{\infty} |f(y)|^2 dy\end{aligned}$$

Nous pouvons donc écrire, en toute généralité,

$$\lambda = e^{i\Phi} = e^{ika}$$

où $k \in \mathbb{R}$ est un nombre homogène à un vecteur d'onde, mais sans autre signification pour l'instant[6]. Nous voyons donc que la valeur propre λ détermine le déphasage de la fonction

[6]La seconde écriture est purement arbitraire pour l'instant ; elle s'avèrera utile par la suite, k étant *vraiment* le vecteur d'onde électronique...

d'onde lorsqu'on se déplace le long de la chaîne d'atome, et que les vecteurs propres de T_a sont des fonctions d'onde dont le module est *périodique, i.e.* , $|\varphi(x)| = |\varphi(x+a)|$. Il en va donc de même de la **probabilité de présence** : un électron placé dans un état stationnaire du cristal étant état propre de T_a, il possède une probabilité de présence périodique, et se trouve donc fondamentalement *délocalisé* sur tout le cristal. Nous avons déjà, à ce stade, un des arguments fondamentaux pour la compréhension du phénomène de conduction du courant électrique dans les cristaux : si les états stationnaires sont délocalisés, il existe des paquets d'onde constitués de ces états capables de se propager d'un bout à l'autre du cristal.

Chaque fonction d'onde étant à la fois, état propre de H de valeur propre E, et état propre de T_a de valeur propre e^{ika}, nous pourrions "indexer" les états stationnaires par les nombres E et k, *i.e.* , $\varphi_{k,E}(x)$. En réalité, nous verrons plus loin que l'énergie E dépend de k (mais non l'inverse : il n'y a pas bijection), ce qui rend superflu l'indice E : nous noterons donc $\varphi_k(x)$ les états stationnaires du cristal. Ils sont donc tels que

$$H[\varphi_k(x)] = E(k)\varphi_k(x)$$

et

$$T_a[\varphi_k(x)] = e^{ika}\varphi(x) = \varphi(x+a)$$

On remarque pour finir que cette seconde relation implique que tout état stationnaire peut aussi s'écrire

$$\varphi_k(x) = u_k(x)e^{ikx}$$

où

$$u_k(x) = u_k(x+a)$$

Nous retrouvons donc l'expression (ici restreinte au cas unidimensionnel) d'une fonction de Bloch.

Il resterait, pour parachever la démonstration du théorème de Bloch, à montrer que k est le vecteur d'onde de l'électron dans le cristal. Ceci sera fait dans le paragraphe relatif au modèle des liaisons fortes (Sec. 2.5).

2.3 Caractéristiques de la structure de bandes d'énergie et première zone de Brillouin

Il reste à déterminer le terme $u_{\mathbf{k}}(\mathbf{r})$ constituant la modulation d'amplitude de l'onde plane. Pour cela, on remplace dans l'équation aux valeurs propres de l'énergie (équation 2.2), la fonction d'onde de l'électron par sa fonction de Bloch :

$$-\frac{\hbar^2}{2m_e}\Delta\left[u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}\right] + V_c(\mathbf{r})u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} = E_n u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.4}$$

Après quelques développements algébriques, on obtient[7] :

$$-\frac{\hbar^2}{2m_e}\left[-\mathbf{k}^2 u_{\mathbf{k}}(\mathbf{r}) + 2i\mathbf{k}\cdot\nabla u_{\mathbf{k}}(\mathbf{r}) + \Delta u_{\mathbf{k}}(\mathbf{r}) + V_c(\mathbf{r})u_{\mathbf{k}}(\mathbf{r})\right] = E(\mathbf{k})u_{\mathbf{k}}(\mathbf{r}) \tag{2.5}$$

Cette relation montre que la fonction $u_{\mathbf{k}}(\mathbf{r})$ dépend a priori du **vecteur d'onde k** (d'où l'indice). De plus, comme l'énergie est une grandeur réelle, le conjugué de l'équation précédente doit redonner la même valeur $E(\mathbf{k})$. Or, prendre le conjugué de cette équation revient

[7] On utilise les deux relations suivantes : $\Delta f = div(\nabla f)$ et $div(\alpha\mathbf{A}) = \alpha div\mathbf{A} + \mathbf{A}\cdot\nabla\alpha$

à changer le signe de $\mathbf{k}$ et donc $E(\mathbf{k}) = E(-\mathbf{k})$. L'énergie d'un électron dans un cristal est donc une fonction paire de $\mathbf{k}$. On effectue l'étude de l'énergie des électrons d'un cristal dans l'espace des vecteurs $\mathbf{k}$ qui correspond à l'espace réciproque défini en cristallographie.

Le réseau réciproque d'un cristal est périodique. Qu'en est-il de l'énergie dans l'espace réciproque ? Soient $\mathbf{T}$ un vecteur du réseau cristallin défini par ses vecteurs de base $\mathbf{a}, \mathbf{b}, \mathbf{c}$ avec n_1, n_2, n_3 entiers et $\mathbf{T}^*$ un vecteur de l'espace réciproque défini par ses vecteurs de base $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$:

$$\mathbf{T}^* = m_1\mathbf{a}^* + m_2\mathbf{b}^* + m_3\mathbf{c}^*$$

avec m_1, m_2, m_3 entiers où les vecteurs $\mathbf{a}, \mathbf{b}, \mathbf{c}$ et $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ sont reliés par la relation (1.1).

Soit $\mathbf{k}'$ un vecteur d'onde tel que $\mathbf{k}' = \mathbf{k} + \mathbf{T}^*$. La fonction d'onde associée à $\mathbf{k}'$ s'écrit :

$$\varphi_{\mathbf{k}'}(\mathbf{r}) = u_{\mathbf{k}'}(\mathbf{r})e^{i\mathbf{k}'\cdot\mathbf{r}} = u_{\mathbf{k}'}(\mathbf{r})e^{i\mathbf{T}^*\cdot\mathbf{r}}e^{i\mathbf{k}\cdot\mathbf{r}} = v_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$$

Or

$$v_{\mathbf{k}}(\mathbf{r}+\mathbf{T}) = u_{\mathbf{k}'}(r+\mathbf{T})e^{i\mathbf{T}^*\cdot(\mathbf{r}+\mathbf{T})} = u_{\mathbf{k}'}(\mathbf{r})e^{i\mathbf{T}^*\cdot\mathbf{r}} = v_{\mathbf{k}}(\mathbf{r})$$

puisque $\mathbf{T}^* \cdot \mathbf{T} = 2\pi$.

$v_{\mathbf{k}}(\mathbf{r})$ est donc toujours une fonction périodique de l'espace "direct" et donc $\varphi_{\mathbf{k}'}(\mathbf{r})$ est une solution de l'équation (2.4) au même titre que $\varphi_{\mathbf{k}}(\mathbf{r})$ et possèdent donc les mêmes valeurs propres $E(\mathbf{k})$. Les états associés aux vecteurs d'onde $\mathbf{k}$ et $\mathbf{k}'$ sont donc physiquement équivalents puisqu'ils possèdent la même énergie.

Théorème 2.2 L'énergie d'un électron du cristal est une fonction périodique dans l'espace réciproque.

Cette propriété de l'énergie permet de restreindre l'étude des états stationnaires du cristal à la première zone de Brillouin qui contient l'intégralité des états quantiques du cristal (le reste de l'espace réciproque étant obtenu par "périodisation" de la première zone de Brillouin).

2.4 Conditions aux limites périodiques et impact sur la quantification de l'énergie

L'étude précédente s'applique au mouvement d'un électron dans un cristal infini. Or les électrons se déplacent dans un espace borné. Au niveau des surfaces du cristal, il y a rupture de la périodicité du potentiel cristallin et donc perturbation de la fonction d'onde. Les effets de surface n'influent que sur des distances très faibles devant les dimensions du cristal et on peut donc les négliger lors de l'étude du mouvement des électrons dans le volume cristallin. En effet, le nombre d'atomes en surface est négligeable par rapport au nombre d'atomes dans le volume. Les conditions aux limites usuelles (par exemple annulation de la fonction d'onde en dehors du cristal) peuvent être remplacées, pour des motifs purement techniques (simplicité de calcul), par des **conditions aux limites périodiques** (ou cycliques).

On considère un pavé de cristal de dimensions $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ grandes devant les dimensions de la maille primitive :

$$\mathbf{X} = N_1\mathbf{a}, \mathbf{Y} = N_2\mathbf{b}, \mathbf{Z} = N_3\mathbf{c}$$

où $\mathbf{a}, \mathbf{b}, \mathbf{c}$ sont les vecteurs de base de la maille primitive et N_1, N_2, N_3 des nombres entiers.

Les conditions aux limites périodiques s'écrivent :

$$\varphi_k(\mathbf{r}) = \varphi_k(\mathbf{r}+\mathbf{X}) = \varphi_k(\mathbf{r}+\mathbf{Y}) = \varphi_k(\mathbf{r}+\mathbf{Z})$$

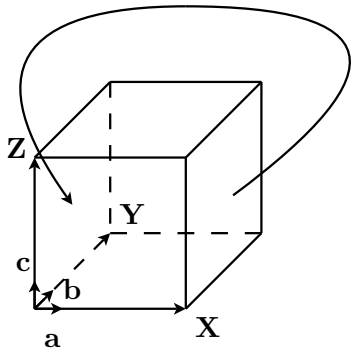


FIG. 2.1 – Conditions aux limites périodiques : on "recolle" chaque face du cristal avec la face opposée.

Ces conditions permettent de ramener l'étude du mouvement d'un électron dans un cristal fini à l'étude de l'électron dans un cristal infini. Il est donc possible d'utiliser le théorème de Bloch. En appliquant ce théorème à la fonction d'onde $\varphi_{\mathbf{k}}$, on obtient, puisque

$$\varphi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}$$

les trois relations suivantes :

$$\varphi_{\mathbf{k}}(r+X) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{X}}$$

$$\varphi_{\mathbf{k}}(r+Y) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{Y}}$$

$$\varphi_{\mathbf{k}}(r+Z) = u_{\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{Z}}$$

En appliquant les conditions aux limites périodiques, on obtient :

$$e^{i\mathbf{k}\cdot\mathbf{X}} = e^{i\mathbf{k}\cdot\mathbf{Y}} = e^{i\mathbf{k}\cdot\mathbf{Z}} = 1$$

Ces relations montrent que le vecteur d'onde $\mathbf{k}$ ne peut prendre que des valeurs discrètes de la forme :

$$\boxed{\mathbf{k} = \frac{n_1}{N_1}\mathbf{a}^* + \frac{n_2}{N_2}\mathbf{b}^* + \frac{n_3}{N_3}\mathbf{c}^*} \tag{2.6}$$

où n_1, n_2, n_3 sont des entiers. Cette relation s'obtient en appliquant les relations (1.1) sachant que $\mathbf{k}$ est un vecteur de l'espace réciproque (mais pas nécessairement du *réseau* réciproque, sauf valeurs particulières de n_1, n_2 et n_3).

2.5 Bandes d'énergie : modèle des liaisons fortes et effet des impuretés

Dans cette partie on introduit la théorie des bandes d'énergie à partir d'un modèle semi-phénoménologique, dit *modèle des liaisons fortes*., qui permet notamment de modéliser l'effet d'une impureté (dopant, défaut, ou autre) dans la chaîne d'atome. Ce modèle s'apparente à la méthode LCAO (Combinaison Linéaire d'Orbitales Atomiques) utilisée en chimie dans l'étude des orbitales hybrides. Il consiste à aborder le problème de la recherche des niveaux d'énergie du cristal selon une approche de type *dynamique temporelle* où l'effet tunnel et le recouvrement des orbitales atomiques d'atomes voisins jouent un rôle explicite.

Le modèle des liaisons fortes — en fait, ici, ce sera une version extrêmement simplifiée de ce modèle — s'appuie sur le fait que l'électron peut passer d'un atome à l'autre par effet tunnel, avec une probabilité décrite par un coefficient semi-empirique A proportionnel au recouvrement des orbitales atomiques d'atomes formant une liaison chimique. Nous verrons que cet effet est responsable de la démultiplication de chaque niveau d'énergie de l'atome **seul** en autant de niveaux qu'il y a d'atomes présents dans le cristal. Cette structure de niveaux d'énergie très nombreux fait apparaître des bandes que l'on appelle **bandes d'énergie**.

Contrairement au modèle de Krönig-Peney, ce modèle permet d'arriver avec relativement peu de calculs à la structure des bandes, et même d'obtenir les caractéristiques essentielles de la dynamique de l'électron : vitesse de groupe, accélération, réponse à un champ extérieur, effet de la présence d'une impureté dans la chaîne. Le prix à payer est celui de toute approche phénoménologique : on introduit forcément à certaines étapes de la modélisation des coefficients qui sont en réalité déterminés par l'expérience, et dont on ne peut par conséquent qu'estimer grossièrement la valeur dans une étude analytique préalable comme celle que nous proposons ici (mais qu'on se rassure : le modèle de Krönig-Peney n'est pas plus réaliste).

2.5.1 Modèle

On considère un cristal unidimensionnel, linéaire, infini, de maille a, constitué d'atomes situés aux abscisses $x_n = na$. Physiquement, un tel système correspond à un nanofil, cf. figure 2.2. On note sur la figure que les orbitales atomiques $\omega(x)$ correspondant à des atomes voisins le long de la chaîne se recouvrent : ce recouvrement, qui correspond à la réalisation effective d'une liaison chimique entre atomes, autorise également la condution électrique le long de la chaîne (modulo des conditions liées au remplissage des bandes d'énergie et à la position du niveau de Fermi).

Le potentiel cristallin dans lequel se déplace l'électron est représenté figure 2.3. Il est construit comme superposition du potentiel créé par tous les atomes de la chaîne, chaque atome créant un potentiel ayant l'allure d'un puits de potentiel, cf. figure 2.4. Sans perdre en généralité, nous avons anticipé sur la suite de cette partie et avons inclut l'effet d'un éventuel atome d'impureté placée arbitrairement à la position x_0 : il peut s'agir d'un atome dopant (phosphore ou bore pour le dopage des semiconducteurs IV) ou d'une impureté "non–désirée" due à la fabrication du cristal,.... La présence de cette impureté à la position x_0 est modélisée par un puits de potentiel de profondeur différente.

On note $\Omega_I(x)$ la fonction d'onde de l'état fondamental d'énergie E_I de l'atome constituant le cristal, et on néglige les autres niveaux d'énergie de cet atome (figure 2.4). De même, on note $\Lambda(x)$ la fonction d'onde de l'état fondamental d'énergie $E_I + F$ de l'impureté située en x_0 (l'allure de $\Lambda(x)$ étant supposée similaire à celle de $\Omega_I(x)$). L'influence de l'impureté sera d'autant plus marquée que F sera élevé en valeur absolue.

Chaque atome du cristal est alors placé sur le réseau cristallin à une position $x_n = na$, avec $n \in \mathbb{Z}^*$ (cristal infini), et on considère comme base d'états du cristal, la base hilbertienne suivante :

$$\mathcal{B} = \{\ldots, \omega_{-n}(x), \ldots, \omega_0(x), \ldots, \omega_n(x), \ldots\}, n \in \mathbb{Z}$$

avec par construction $\omega_n(x) = \Omega_I(x - na)$ si $n \neq 0$, et $\omega_0(x) = \Lambda(x)$. Il faut noter que la base $\mathcal{B}$ ne constitue qu'approximativement une base orthonormée. En effet, les orbitales atomiques d'atomes voisins se recouvrent un peu (c'est d'ailleurs nécessaire pour que la liaison chimique existe), et l'orthogonalité n'est donc pas parfaite.

Dans cette base $\mathcal{B}$, la matrice hamiltonienne H (de dimension infinie) gouvernant la

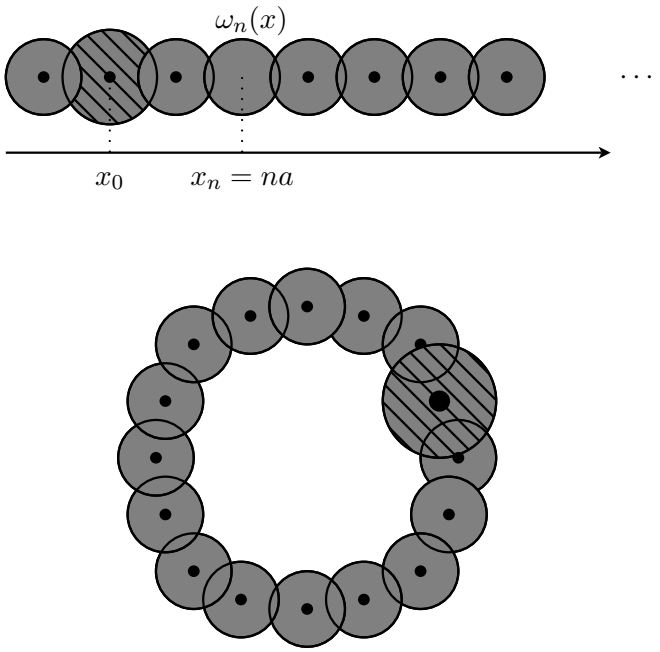


FIG. 2.2 – Orbitales atomiques de valence (*i.e.* , couches externes), notées $\omega_n(x)$, dans le cas d'un nanofil (chaîne d'atome équivalente à un cristal unidimensionnel) : le recouvrement d'orbitales voisines permet la propagation de l'électron le long de la chaîne par effet tunnel. L'impureté en x_0 est arbitrairement matérialisée par une orbitale de volume supérieur. Figure inférieure : conditions aux limites périodiques obtenues par "soudure" du fil sur lui-même. Ces conditions sont, en général, purement artificielle et ne correspondent pas à une réalité physique.

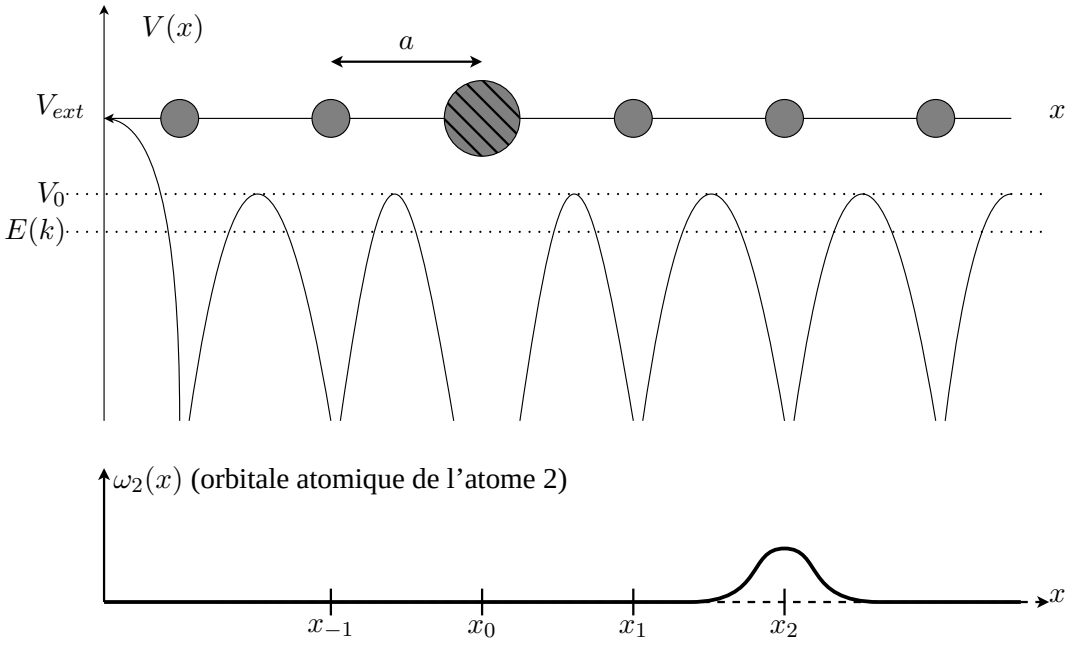


FIG. 2.3 – Le potentiel cristallin $V(x)$ représente les forces qui s'exercent sur les électrons du cristal. L'énergie V_{ext} est l'énergie que doivent posséder les électrons pour pouvoir sortir du cristal ; si leur énergie est inférieure à V_{ext} ils sont liés au cristal. Les électrons dont l'énergie potentielle est inférieure à V_0 sont liés aux noyaux et ne peuvent pas se déplacer dans le cristal, sauf si leur énergie n'est que **légèrement** inférieure à V_0, auquel cas ils sont **quasi-libres**, et peuvent alors passer d'un puits à l'autre par **effet tunnel**. Les électrons d'énergie supérieure à V_0 peuvent se déplacer librement dans le cristal, on parle d'électrons **libres**. A titre d'illustration, une **impureté** a été introduite arbitrairement dans la chaîne : sa présence se manifeste par un puit de potentiel de largeur et profondeur distinctes. **En bas** : une fonction d'onde particulière de la base des orbitales atomiques $\mathcal{B}$, en l'occurence $\omega_2(x) = \Omega_I(x-2a)$ correspondant à l'atome numéro "2" dans la chaîne.

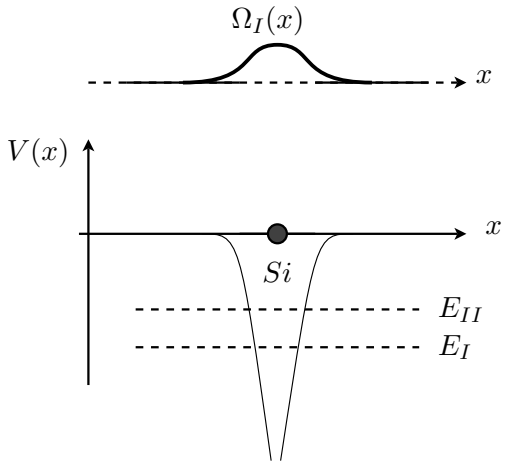


FIG. 2.4 – Potentiel $V(x)$ créé par l'atome lorsqu'il est "isolé" (ici, du silicium). On note E_I le niveau fondamental correspondant, et $\Omega_I(x)$ la fonction d'onde (i.e. l'orbitale atomique) **normée** correspondante. E_{II} représente le premier niveau excité, qui n'est pas pris en compte dans les calculs de cette partie. Il donnerait naissance à une deuxième bande d'énergie, cf. figure 2.5.

dynamique de l'électron est définie par :

$$\begin{aligned} &H_{n,n} = E_I,\ H_{n,n\pm1} = A,\ \text{pour } n \neq 0 \\ &H_{0,0} = E_I + F,\ H_{0,-1} = H_{0,1} = A \\ &H_{i,j} = 0 \text{ sinon} \end{aligned}$$

où l'on rappelle que $H_{i,j} = \langle \omega_i | \, H \, | \omega_j \rangle$ par définition. Notez que les indices de ligne/colonne peuvent être négatifs... ce qui est une convention uniquement destinée à rendre la notation plus légère dans la suite (l'impureté étant située en $n = 0$ par commodité, il peut y avoir des atomes à sa gauche). On peut schématiquement représenter H ainsi :

$$H = \begin{pmatrix} & \ddots & \ddots & \ddots & & & & & & & \\ \cdots & 0 & A & E_I & A & 0 & \cdots & & & & \\ & \cdots & 0 & A & E_I & A & 0 & \cdots & & & \\ & & \cdots & 0 & A & E_I + F & A & 0 & \cdots & & \\ & & & \cdots & 0 & A & E_I & A & 0 & \cdots & \\ & & & & \cdots & 0 & A & E_I & A & 0 & \cdots \\ & & & & & & & \ddots & \ddots & \ddots & \end{pmatrix} \tag{2.7}$$

Le terme A traduit la possibilité qu'a l'électron de sauter, par effet tunnel, d'un atome à l'atome voisin. Le terme F à la ligne "0" traduit le fait que le niveau d'énergie fondamental de l'impureté est différent[8] de E_I. Il suffit donc de faire $F = 0$ pour traiter le cas d'une chaîne d'atomes sans impuretés.

On notera que A est (en théorie !) explicitement calculable via la formule

$$A = \langle \omega_n | \, H \, | \omega_{n+1} \rangle = \int_{-\infty}^{\infty} \omega_n^*(x) H[\omega_{n+1}(x)] dx = \int_{-\infty}^{\infty} \Omega_I^*(x) H[\Omega_I(x-a)] dx$$

[8] En toute rigueur, les coefficients A de la ligne "0" devraient également avoir une valeur différente des autres A.

Ceci montre que, toutes choses égales par ailleurs, A est proportionnel au recouvrement des orbitales d'atomes voisins. Nous admettrons que sa valeur peut être déterminée expérimentalement, ou bien numériquement via une simulation sur ordinateur.

2.5.2 Solutions générales pour les états stationnaires

On note $\varphi_k(x)$ un état stationnaire de l'électron dans le cristal, et $E(k)$ l'énergie de cet état (k n'étant pour l'instant qu'un nombre permettant de repérer les différents états stationnaires du cristal et leur énergie $E(k)$). Dans la base $\mathcal{B}$, on écrit

$$\varphi_k(x) = \sum_n c_n \, \omega_n(x)$$

où les coordonnées c_n sont à déterminer... Les postulats de la mécanique quantique nous disent que les niveaux d'énergie sont les valeurs propres de H, et les états stationnaires $\varphi_k(x)$ sont les vecteurs propres correspondant. Nous devons donc chercher les fonctions d'onde telles que

$$H[\varphi_k(x)] = E(k)\varphi_k(x)$$

soit encore

$$H \left|\varphi_k(x)\right\rangle = E(k) \left|\varphi_k(x)\right\rangle$$

en notation de Dirac. En projetant[9] cette relation sur un vecteur $\omega_m(x)$ de la base $\mathcal{B}$, on obtient

$$\left\langle\omega_m(x)\right| H \left|\varphi_k(x)\right\rangle = E(k) \left\langle\omega_m(x)\right| \left.\varphi_k(x)\right\rangle$$

d'où

$$\sum_n c_n \left\langle\omega_m(x)\right| H \left|\omega_n(x)\right\rangle = E(k)c_m$$

ou encore

$$\sum_n H_{mn}c_n = E(k)c_m$$

(on reconnaîtra aisément l'équation matricielle $H \cdot \varphi = E\varphi$ dans cette formule, d'ailleurs). Compte-tenu de la définition des éléments de la matrice H (equ. 2.7), on a donc

$$\boxed{Ac_{m-1} + E_I c_m + Ac_{m+1} = E(k)c_m, \, \forall m \neq 0} \tag{2.8}$$

$$\boxed{Ac_{-1} + (E_I + F)c_0 + Ac_1 = E(k)c_0} \tag{2.9}$$

la seconde équation se ramenant au cas général en l'absence d'impureté ($F = 0$). L'ensemble de ces équations constitue donc les équations aux valeurs propres de H permettant de calculer les niveaux d'énergie $E(k)$ et les coefficients c_n des états stationnaires dans la base $\mathcal{B}$ des orbitales atomiques.

2.5.3 Bandes d'énergie en l'absence d'impureté

On fixe donc tout d'abord F à zéro. Les coefficients c_n sont donc solutions de

$$Ac_{m-1} + E_I c_m + Ac_{m+1} = E(k)c_m, \, \forall m$$

[9] En faisant le produit scalaire...

Il s'agit d'une équation aux différences finites (ou "suite récurrente"), linéaire et à coefficients constants, dont les solutions doivent être cherchées sous la forme $c_n = \alpha^n$. Ceci implique que $\varphi_k(x+a) = \alpha\varphi_k(x)$. Or, d'après le théorème de Bloch, on a

$$\varphi_k(x+a) = e^{ika}\varphi_k(x)$$

ce qui donne finalement $\alpha = e^{ika}$. Les états stationnaires s'écrivent donc

$$\boxed{\varphi_k(x) = \sum_n e^{ikna}\,\omega_n(x) = \sum_n e^{ikx_n}\,\omega_n(x)} \tag{2.10}$$

Ceci montre au passage que ces états stationnaires vérifient le théorème de Bloch avec la modulation d'amplitude $u_k(x)$ donnée par $\sum_n \omega_n(x) = \sum_n \Omega_I(x-na)$, *i.e.* , ce sont des ondes planes modulées en amplitude par les orbitales atomiques.

A partir de l'expression explicite de c_n et de l'équation 2.8, on peut désormais calculer $E(k)$:

$$E(k)e^{ikma} = Ae^{ik(m-1)a} + E_I e^{ikma} + Ae^{ik(m+1)a}$$

d'où

$$\boxed{E(k) = E_I + 2A\cos(ka)} \tag{2.11}$$

Les conditions aux limites périodiques, équation 2.6, restreintent au cas unidimensionnel, fixent la quantification de k :

$$k = \frac{2\pi l}{Na} = \frac{2\pi l}{L}$$

donc indirectement, celle de l'énergie. Les niveaux d'énergie sont donc bornés par $E_I - 2A$ et $E_I + 2A$, et forment une **bande d'énergie**. La généralisation au cas d'un atome possédant plusieurs orbitales $\{\Omega_I, \Omega_{II}, \ldots\}$ d'énergie $E_I, E_{II}, \ldots$ conduit à un ensemble de bandes d'énergie séparées par des bandes interdites, ainsi que l'illustre la figure 2.5. Il faut noter que si A est suffisamment important (en résultat à un fort recouvrement des orbitales), alors les bandes peuvent se chevaucher (cas de l'étain : on appelle alors semi-métal le matériau correspondant).

2.5.4 Dynamique et vitesse de groupe

On considère un état initial de l'électron décrit par l'état stationnaire $\psi(x, t=0) = \varphi_k(x)$ d'énergie $E(k)$. D'après l'équation de Schrödinger, l'évolution temporelle est simplement donnée par

$$\psi(x, t>0) = e^{-\frac{i}{\hbar}E(k)t}\varphi_k(x)$$

La fonction d'onde ne diffère donc de l'état initial que par un facteur de phase dépendant du temps, et l'électron demeure délocalisé sur l'ensemble du cristal (on notera en effet que, comme pour tout état stationnaire, $|\psi(x, t>0)| = |\psi(x, t=0)|$). Un état "physiquement réaliste" où la fonction d'onde électronique est partiellement localisée (disons, en gros, sur quelques distances interatomiques) est donc décrit par un paquet d'ondes stationnaires,

$$\psi(x, t=0) = \sum_k d_k\,\varphi_k(x),$$

où la somme porte sur tous les vecteurs d'onde de la première zone de Brillouin, et les coefficients d_k doivent simplement satisfaire

$$\sum_k |d_k|^2 = 1$$

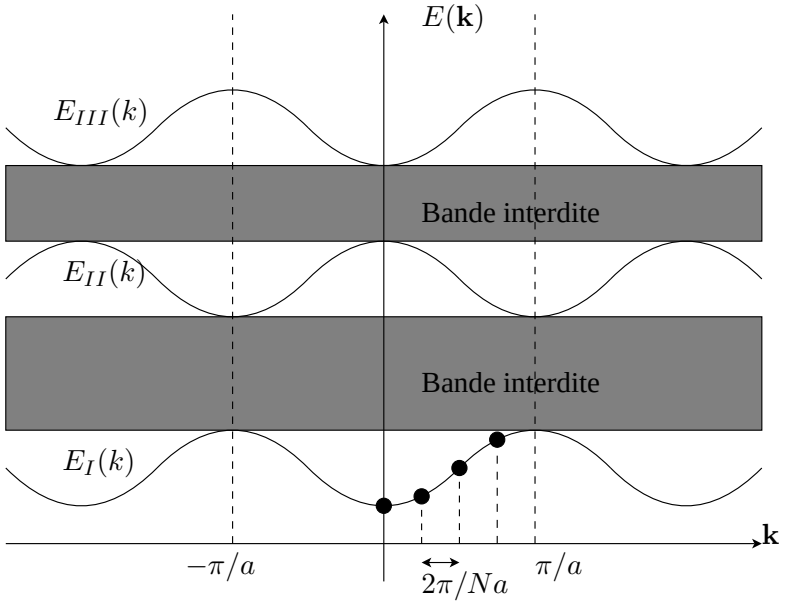


FIG. 2.5 – Bandes d'énergie dans une chaîne de N atomes : les bandes $E_I(k), E_{II}(k), \ldots$ proviennent chacune d'une orbitale de l'atome isolé. La quantification de k est due aux conditions aux limites périodiques, les points noirs représentant les différents états stationnaires.

pour que $\psi(x, t = 0)$ soit normalisée. A l'instant d'un paquet d'ondes électromagnétiques, les coefficients d_k n'ont des valeurs appréciables que pour k voisin d'un vecteur d'onde k_0.

La dynamique d'un tel paquet d'onde est donnée par l'équation de Schrödinger : ici, en vertu du fait que $\psi(x, t = 0)$ est une superposition d'états stationnaires dont on a calculé la dynamique il y a un instant, on a donc trivialement (en vertu de la linéarité de l'équation de Schrödinger)s

$$\psi(x, t > 0) = \sum_k d_k \, e^{-\frac{i}{\hbar}E(k)t} \, \varphi_k(x),$$

En remplaçant $\varphi_k(x)$ par son expression sur la base des orbitales atomiques, équation 2.10, on obtient

$$\begin{aligned}
\psi(x, t > 0) &= \sum_k d_k \, e^{-\frac{i}{\hbar}E(k)t} \sum_n e^{ikx_n} \, \omega_n(x) \\
&= \sum_n \left[\sum_k d_k \, e^{i\left(kx_n - \frac{E(k)t}{\hbar}\right)} \right] \omega_n(x) \\
&= \sum_n C_n(t) \, \omega_n(x)
\end{aligned}$$

avec

$$C_n(t) = \sum_k d_k \, e^{i\left(kx_n - \frac{E(k)t}{\hbar}\right)}$$

L'expression de $\psi(t)$ sur la base des orbitales montre que la probabilité de présence de l'électron "sur" le n*ième* atome est donnée par la quantité $|C_n(t)|^2 = |\, \langle \omega_n(x)| \, \psi(x,t) \rangle \,|^2$, qui est une onde de probabilité se propageant à la vitesse de groupe

$$v_g = \frac{1}{\hbar} \frac{dE(k)}{dk}$$

quantité qui s'interprète géométriquement comme la tangente à la courbe $E(k)$ divisée par $\hbar$. Cette expression se généralise dans le cas tridimensionnel à

$$\boxed{\mathbf{v}_g = \frac{1}{\hbar}\nabla_k E(k)} \tag{2.12}$$

2.5.5 Solutions en présence de l'impureté

Nous allons montrer que l'impureté située en x_0 provoque une diffusion de la fonction d'onde de l'électron. Ce cas correspond à $F \neq 0$: les solutions précédentes $c_n = e^{ikx_n}$ ne sont alors plus valables à cause des lignes "-1", "0" et "1" de la matrice H. On considère donc désormais des solutions c_n construites comme la somme d'une onde incidente et d'une onde réfléchie à gauche de l'impureté, et une onde transmise à droite de l'impureté :

- pour $n < 0$, $c_n = e^{ikx_n} + Re^{-ikx_n}$;
- pour $n > 0$, $c_n = Te^{ikx_n}$

où R et T sont respectivement les coefficients de réflexion et de transmission de l'onde sur l'impureté, et c_0 reste à déterminer.

On admet que la structure des bandes n'est pas modifiée par l'impureté : les nouveaux coefficients c_n étant des superpositions linéaires de solutions, ils sont donc solutions de l'équation aux valeurs propres de H pour tout $n \neq 0$, avec la même valeur propre $E(k)$ que précédemment. Reste donc à examiner les lignes $n = -1$, 0 et 1. L'équation $H[\varphi_k(x)] = E(k)\varphi_k(x)$ nous fournit dans ce cas trois équations, à savoir

$$\begin{aligned} Ac_{-2} + E_I c_{-1} + Ac_0 &= E(k)c_{-1} \\ Ac_{-1} + (E_I + F)c_0 + Ac_1 &= E(k)c_0 \\ Ac_0 + E_I c_1 + Ac_2 &= E(k)c_1 \end{aligned}$$

d'où

$$\begin{aligned} A(e^{-2ika} + Re^{2ika}) + [E_I - E(k)](e^{-ika} + Re^{ika}) + Ac_0 &= 0 \\ A(e^{-ika} + Re^{ika}) + [E_I + F - E(k)]c_0 + ATe^{ika} &= 0 \\ Ac_0 + [E_I - E(k)]Te^{ika} + ATe^{2ika} &= 0 \end{aligned}$$

La résolution fournit

$$R = \frac{-1}{1 + \frac{2iA\sin(ka)}{F}} \tag{2.13}$$

$$T = c_0 = \frac{1}{1 + \frac{F}{2iA\sin(ka)}} \tag{2.14}$$

2.5.6 Piégeage par l'impureté

Si k est imaginaire pur, on écrit $k = i/L$ où $L \in \mathbb{R}$, et $E(k)$ est alors situé "hors" de la bande d'énergie (en effet, le terme $\cos(ka)$ devient $\cosh(a/L)$, qui "sort" de l'intervalle $[-1, 1]$). On obtient également une vitesse de groupe imaginaire pure : l'électron ne se propage plus. Ceci est confirmé par la fonction d'onde, dont les coefficients c_n sont des exponentielles décroissantes, *i.e.* , $\exp(ikx_n) = \exp(-x_n/L)$. Cette situation correspond à une fonction d'onde électronique localisée : son allure est donnée par deux exponentielles décroissantes de longueur caractéristique L, cf. figure 2.7. L'électron est alors dit **piégé**.

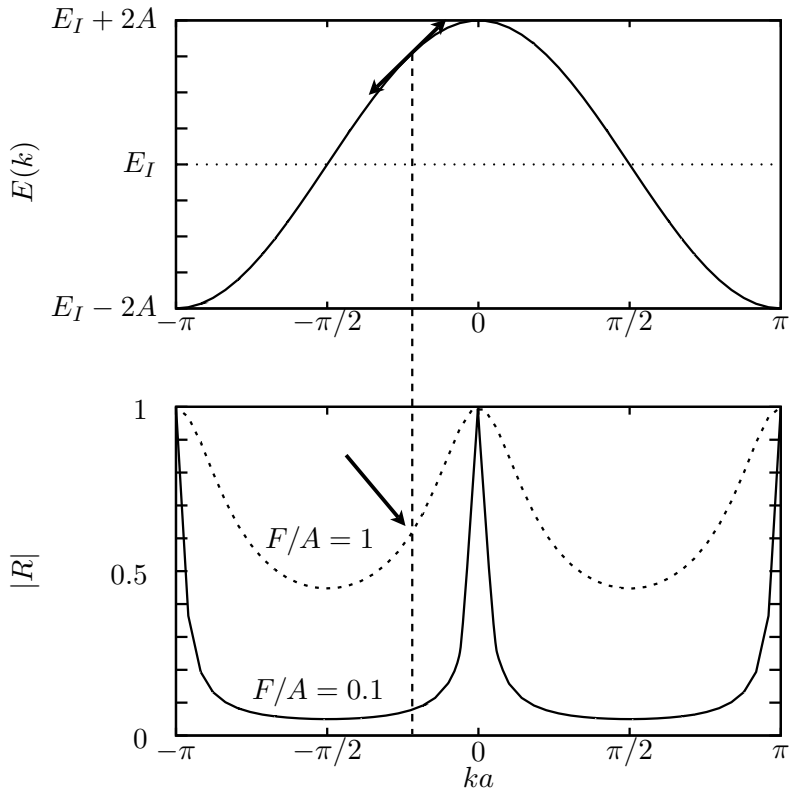


FIG. 2.6 – Haut : diagramme de bande obtenu à partir du modèle à une orbitale ; bas : coefficient de réflexion de l'onde électronique sur l'impureté en fonction du vecteur d'onde k pour $F/A = 0.1$ ("faible" impureté) et $F/A = 1$ ("forte" impureté). La tangente sur la courbe $E(k)$ illustre la vitesse de groupe pour un état proche de l'extrémum de bande ; le coefficient de réflexion correspondant sur l'impureté est indiqué par un flèche pour $F = A$. On notera que, plus v_g augmente (en valeur absolue), plus le coefficient de réflexion sur l'impureté diminue.

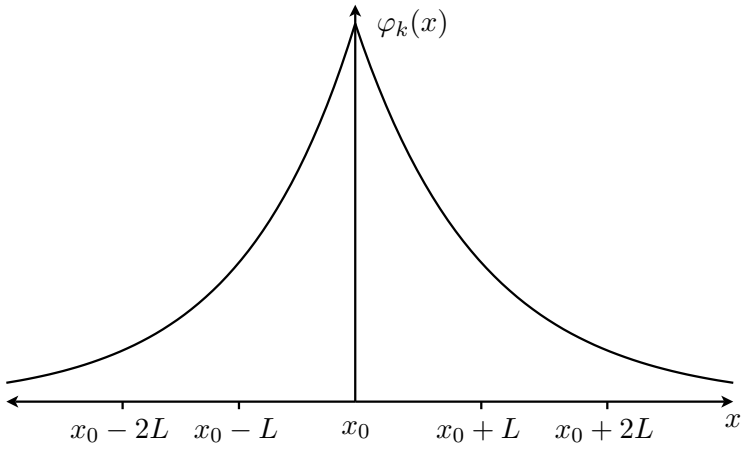


FIG. 2.7 – Fonction d'onde électronique localisée sur une impureté située en x_0 : le vecteur d'onde k est imaginaire pur et vaut i/L ; l'énergie de l'électron est par conséquent située hors de la bande permise. Un électron qui occuperait cet état quantique serait piégé par l'impureté, et ne pourrait plus participer à la conduction du courant électrique.

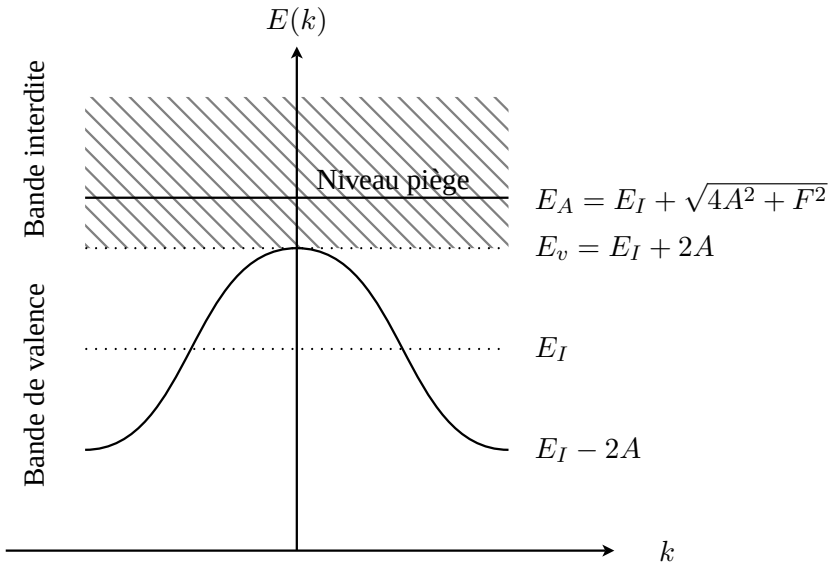


FIG. 2.8 – Les différentes niveaux d'énergie en jeu dans le piégeage d'un électron par une impureté : E_v est le maximum de la bande de valence (j'ai choisi de dessiner la bande avec $A > 0$, c'est donc une bande de valence) ; E_A est le niveau d'énergie correspondant à un électron piégé (dit "niveau piége", ici niveau accepteur par référence au dopage de type P). Remarque : il existe une autre solution en-dessous de la bande de valence, qui toutefois ne nous intéresse pas.

On peut être plus précis : l'onde pour $n < 0$ est de la forme $\exp(-x_n/L)+R\exp(-x_n/L)$. Ce n'est que si R est infini que la localisation a lieu en $n = 0$ (sinon l'onde diverge lorsque $x \to -\infty$, et sa normalisation impose qu'elle soit négligeable en x_0). Si R est infini, T l'est aussi (et $R = T$ de fait), d'où d'après l'expression de R ci-dessus (equation 2.13),

$$F = -2iA\sin(ka)$$

d'où, en explicitant k,

$$F = 2A\sinh(a/L)$$

soit finalement

$$L/a = \frac{1}{\sinh^{-1}(F/2A)}$$

pour l'extension caractéristique de la fonction d'onde de l'électron piégé. Le niveau d'énergie de l'électron piégé est quant à lui donné par

$$E(k) = E_I + 2A\cos(ka) = E_I \pm \sqrt{4A^2 + F^2}$$

Le symbole $\pm$ vient du fait que j'ai exprimé l'énergie $E(k)$ directement en fonction de A et F, il y a donc deux solutions. Dans tous les cas, on voit que $E(k)$ est bien situé *hors* de la bande permise, cf. figure 2.8. Si F augmente (impureté "forte"), on constate que :

- l'énergie de l'électron piégé s'éloigne de plus en plus de la bande permise et s'enfonce dans la bande interdite (niveau "profond")
- parallèlement, l'extension caractéristique de la fonction d'onde, L, diminue : le piégeage s'accompagne d'une localisation de plus en plus forte...

2.6 Bande de conduction et bande de valence

Dans le paragraphe précédent, on a vu que le spectre d'énergie possible des électrons d'un cristal est réparti en bandes permises séparées par des bandes interdites. Ces bandes d'énergie sont constituées d'états d'énergie discrets puisque le vecteur d'onde **k** est quantifié. Les électrons peuvent donc occuper un nombre fini de niveaux d'énergie.

Les électrons se répartissent sur ces niveaux d'énergie en occupant d'abord les niveaux de plus basse énergie. Ces niveaux correspondent aux niveaux d'énergie des atomes dans le cristal (différents de ceux de l'atome isolé). La dernière bande d'énergie accueillant des électrons correspond à la dernière couche électronique des atomes, également appelée **couche de valence**. On appelle donc **bande de valence** la dernière bande d'énergie correspondant aux électrons liés aux atomes du cristal.

Sous l'effet de l'agitation thermique, d'une force extérieure ou d'un flux lumineux, certains électrons peuvent se libérer de l'attraction de l'atome et passer sur des niveaux d'énergie supérieurs (ce sont les électrons quasi-libres). Il se trouvent alors dans une bande d'énergie que l'on appelle la **bande de conduction** (en référence aux phénomènes de conduction liés à ces électrons quasi-libres). La bande de valence et la bande de conduction sont séparées par une **bande d'énergie interdite** que l'on appelle le **gap**. La valeur de ce gap permet de faire la distinction entre matériaux isolant, semi-conducteur et conducteur. L'énergie maximum de la bande de valence est notée E_v, l'énergie minimum de la bande de conduction est notée E_c et la largeur de la bande interdite : $E_c - E_v = E_g$.

L'étude des propriétés électroniques du cristal ne fait intervenir que les électrons susceptibles de produire un courant. Or, comme il sera montré par la suite, seuls les électrons présents dans des bandes d'énergie dont tous les états ne sont pas occupés peuvent participer à des phénomènes de conduction. Seuls les bandes de valence et de conduction peuvent correspondre à ce critère puisque les électrons éventuellement présents dans la bande de conduction proviennent de la bande de valence. En effet, à la température de $0K$, tous les niveaux de plus basse énergie sont occupés, donc toute les bandes sont pleine à l'exception de la bande de conduction. Celle-ci peut donc accueillir des électrons provenant de la bande d'énergie immédiatement inférieure, la bande de valence. Ces deux bandes d'énergie ne sont donc pas totalement remplies et les électrons qui s'y trouvent peuvent participer à un courant.

2.7 Isolant, semi-conducteur et métal

La distinction entre isolant, semi-conducteur et métal provient de la valeur de leur gap E_g :

- pour les isolants $E_g > 3eV$
- pour les semi-conducteurs $E_g < 3eV$
- pour les métaux, soit il y a **recouvrement** de la bande de valence et de la bande de conduction (absence de gap), soit les états voisins du minimum d'énergie de la bande de conduction sont déjà occupés par des électrons à la température de $0K$.

2.8 Electron en présence d'un champ électrique - masse effective

Après avoir défini la structure des niveaux d'énergie électronique dans le cristal (approche "statique"), il apparaît désormais indispensable de déterminer comment les électrons réagissent lorsqu'on leur applique une force extérieure (macroscopique), par exemple une force de nature électrostatique issue d'un générateur de f.e.m. La dynamique résultant de

l'application de cette force nous permettra (mais seulement après de multiples étapes...) de déterminer (notamment) la relation courant-tension dans le cristal.

L'approche que l'on va suivre ici est dite **semi-quantique** : elle est valable si l'on applique au cristal une force **lentement variable** (en pratique, de fréquence inférieure à plusieurs centaines de GHz), et de faible intensité, de manière que l'électron ne **change pas de bande d'énergie** au cours de la dynamique. On peut alors continuer à raisonner comme si le paquet d'ondes décrivant (au niveau quantique) l'électron était équivalent à un point (donc approche "classique") situé au "sommet" de la vague.

Cette approche nous conduira à introduire le concept de masse effective, et surtout, une relation semblable à la relation fondamentale de la dynamique en mécanique classique : on pourra alors, par la suite, et moyennant quelques hypothèses, traiter le mouvement des électrons dans le cristal comme celui de particules classiques (ce qui s'avérera beaucoup plus simple qu'un traitement quantique systématique).

Le prix à payer sera la nécessité de revenir au traitement quantique, via l'équation de Schrödinger, lorsque les forces exercées seront, soit très intense (cas des champs électriques intense dans les diodes Zener), soit de très haute fréquence (interaction du cristal avec une onde électromagnétique IR ou visible, cf. chapitre 7).

2.8.1 Action d'une force extérieure - théorème accélération

L'accélération "classique" de l'électron (*i.e.* ici aussi, "classique" se réfère au sommet du paquet d'ondes) est donnée par :

$$\mathbf{a} = \frac{d\mathbf{v}}{dt}$$

soit, en utilisant la définition de la vitesse "classique" de l'électron,

$$\mathbf{a} = \frac{1}{\hbar}\frac{d}{dt}\nabla_k E_n(\mathbf{k})$$

Continuons à raisonner avec les outils de la mécanique newtonienne : si l'on applique une force extérieure $\mathbf{F}$, l'énergie de l'électron variera, d'après le théorème de l'énergie cinétique, de la quantité de travail reçu, soit

$$dE = \delta W = \mathbf{F} \cdot d\mathbf{r} = \mathbf{F} \cdot \mathbf{v} dt$$

Évidemment, cette relation n'est valable que dans le cas où les hypothèses énoncées plus haut sont vérifiées (faible intensité, faible fréquence, pas de transition inter-bande).

Or on a aussi, par définition de la vitesse de groupe,

$$dE = \nabla_k E_n(\mathbf{k}) \cdot d\mathbf{k} = \hbar \mathbf{v} \cdot d\mathbf{k}$$

En identifiant ces deux relations, on obtient

$$\boxed{\mathbf{F} = \hbar \frac{d\mathbf{k}}{dt} = \frac{d\mathbf{p}}{dt}} \tag{2.15}$$

ce qui montre que, dans l'approche semi-quantique, $\mathbf{p} = \hbar\mathbf{k}$ est équivalent à la quantité de mouvement de l'électron dans le cristal : $\mathbf{k}$ n'est pas seulement le vecteur d'onde de l'électron, il représente également ce qu'on nomme le **moment cristallin** (terminologie anglo-saxonne, "momentum" signifiant "impulsion" ou "quantité de mouvement").

Remarque 2.2 *En particulier, dans une collision, cette quantité de mouvement intervient dans le bilan ; par exemple, la diffusion d'un photon X de quantité de mouvement initiale $\hbar\mathbf{K}$, et finale $\hbar\mathbf{K}'$, s'écrira*

$$\hbar\mathbf{K} + \hbar\mathbf{k} = \hbar\mathbf{K}' + \hbar\mathbf{k}$$

où $\mathbf{k}$ et $\mathbf{k}'$ sont les moments cristallins électroniques avant et après le processus de diffusion. Or on a vu au chapitre 1, à propos de la diffusion de Bragg, que la relation suivante devait être vérifiée pour que la diffusion soit favorisée :

$$\mathbf{K} = \mathbf{K}' + \mathbf{T}^*$$

où $\mathbf{T}^*$ est un vecteur du réseau réciproque. Ainsi, dans une diffusion de type Bragg, la variation de moment cristallin de l'électron ayant diffusé l'onde est toujours un vecteur du réseau réciproque.

2.8.2 Masse effective

En reportant la relation 2.15 dans l'expression donnant l'accélération, on obtient :

$$\begin{aligned}\mathbf{a} &= \frac{1}{\hbar}\frac{d}{dt}\nabla_k E_n(\mathbf{k}) \\ &= \frac{1}{\hbar}\left(\frac{d\mathbf{k}}{dt}\cdot\nabla_k\right)\nabla_k E_n(\mathbf{k}) \\ &= \frac{1}{\hbar^2}\left(\mathbf{F}\cdot\nabla_k\right)\nabla_k E_n(\mathbf{k})\end{aligned}$$

soit, en explicitant les coordonnées de $\mathbf{a}$,

$$a_i = \frac{1}{\hbar^2}\left(\sum_j F_j\frac{\partial}{\partial k_j}\right)\frac{\partial E_n(\mathbf{k})}{\partial k_i} = \frac{1}{\hbar^2}\sum_j \frac{\partial^2 E_n(\mathbf{k})}{\partial k_j \partial k_i}F_j$$

Cette dernière relation relie **linéairement**, sous forme de relation matricielle, l'accélération de l'électron à la force appliquée, et on peut remarquer qu'elle ressemble à l'équation fondamentale de la dynamique $\mathbf{a} = m^{-1}\mathbf{F}$ à condition d'affecter à l'électron une « masse » m^* telle que :

$$\boxed{\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2}\frac{\partial^2 E_n(\mathbf{k})}{\partial k_j \partial k_i}}$$

soit, explicitement,

$$\frac{1}{m^*} = \begin{pmatrix} \frac{\partial^2 E}{\partial k_x^2} & \frac{\partial^2 E}{\partial k_x \partial k_y} & \frac{\partial^2 E}{\partial k_x \partial k_z} \\ \frac{\partial^2 E}{\partial k_y \partial k_x} & \frac{\partial^2 E}{\partial k_y^2} & \frac{\partial^2 E}{\partial k_y \partial k_z} \\ \frac{\partial^2 E}{\partial k_z \partial k_x} & \frac{\partial^2 E}{\partial k_z \partial k_y} & \frac{\partial^2 E}{\partial k_z^2} \end{pmatrix}$$

La matrice $1/m^*$ est appelée **tenseur de masse effective** de l'électron.

Dans le cas où cette matrice est proportionnelle à la matrice identité, on parle alors simplement de **masse effective**, et l'on a les relations

$$\mathbf{F} = m^*\mathbf{a}$$

$$1/m^* = \frac{1}{\hbar^2}\frac{\partial^2 E_n(\mathbf{k})}{\partial k^2}$$

Ce concept de masse effective est très utile puisqu'il permet de traiter la dynamique de l'électron quasi-libre dans un cristal en utilisant les lois de la mécanique classiques, à condition de remplacer la masse au repos de l'électron par sa masse effective. On peut interpréter cette grandeur comme représentant toutes les interactions de l'électron avec son environnement cristallin.

On notera que, alors que la vitesse de groupe était proportionnelle à la **pente** de la courbe $E(k)$, la masse effective est inversement proportionnelle à la **courbure** de cette bande. Ainsi,

- $m^* < 0$ au voisinage d'un maximum de$E(\mathbf{k})$
- $m^* > 0$ au voisinage d'un minimum de$E(\mathbf{k})$

Autrement dit, dans un cristal, l'accélération de l'électron n'est pas forcément colinéaire à la force appliquée, et peut même être de sens opposé.

2.8.3 Étude des extréma de la fonction $E_n(\mathbf{k})$

Au voisinage d'un extrémum de la courbe $E_n(\mathbf{k})$ situé en $\mathbf{k} = \mathbf{k}_0$, on peut développer la fonction $E_n(\mathbf{k})$ en série de Taylor. En se limitant au deuxième ordre, on obtient :

$$E_n(\mathbf{k}) \sim E_n(\mathbf{k}_0) + \frac{1}{2}\sum_{i,j} \left.\frac{\partial^2 E}{\partial k_i \partial k_j}\right|_{\mathbf{k}=\mathbf{k}_0} (k_i - k_{0,j})(k_i - k_{0,j})$$

Si on se place dans un système de coordonnées tel que la matrice $1/m^*$ soit diagonale, on peut exprimer l'énergie en fonction de la masse effective et du vecteur d'onde sous la forme :

$$E_n(\mathbf{k}) \sim E_n(\mathbf{k}_0) + \frac{\hbar^2}{2}\left[\frac{(k_x - k_{0,x})^2}{m_x^*} + \frac{(k_y - k_{0,y})^2}{m_y^*} + \frac{(k_z - k_{0,z})^2}{m_z^*}\right]$$

Au voisinage d'un extrémum, les surfaces de même énergie dans l'espace des $\mathbf{k}$ (ou surfaces isoénergétiques définies par$E(\mathbf{k})$=cste) sont donc des **ellipsoïdes**. Cette approximation quadratique n'est bien sûr valable qu'au voisinage des extréma de E puisqu'elle provient d'un développement limité. Heureusement, seuls les électrons d'énergie proche des extréma de $E(\mathbf{k})$ sont concernés par des phénomènes de transport comme on le verra par la suite.

Dans le cas où $m_x = m_y \neq m_z$, la surface isoénergétique est une **ellipsoïde de révolution** autour de l'axe correspondant à m_z, *i.e.* l'axe porté par k_x. On note généralement $m_z = m_l$, que l'on définit comme la masse **longitudinale**, et $m_x = m_y = m_t$ étant appelées masses **transversales**.

Enfin, dans le cas le plus simple où $m_x = m_y = m_z$, la masse effective est alors un scalaire noté m^*. Les surfaces isoénergétiques sont des sphères et l'énergie s'écrit :

$$E_n(\mathbf{k}) \sim E_n(\mathbf{k}_0) + \frac{\hbar^2}{2}\frac{(\mathbf{k} - \mathbf{k}_0)^2}{m^*}$$

Cette expression est similaire à celle de l'énergie d'un électron enfermé dans une boîte. On retrouve bien l'idée que la dynamique d'un électron cristallin est identique à celle de l'électron libre à condition de remplacer sa masse par la masse effective m^*.

Dans l'expression précédente, $E(\mathbf{k}_0)$ correspond à l'énergie **potentielle** de l'électron cristallin et le terme $E_c = \frac{\hbar^2}{2}\frac{(\mathbf{k}-\mathbf{k}_0)^2}{m^*}$ à son énergie cinétique. On a vu que la quantité de mouvement de l'électron dans le cristal valait $\mathbf{p} = \hbar\mathbf{k}$, ce qui permet de retrouver $E_c = \frac{p^2}{2m^*}$ (ici encore, la seule différence est la présence de la masse effective au lieu de la masse dans le vide).

2.8.4 Courant dans une bande pleine

Une bande pleine ne peut pas contribuer à des phénomènes de transport : en effet, la vitesse totale des électrons de cette bande pleine est donnée par :

$$\mathbf{v}_{tot} = \sum_{k \in 1ere Z.B.} \mathbf{v}_k = \sum_k \frac{1}{\hbar} \nabla_k E_n(\mathbf{k})$$

la somme portant sur tous les états électroniques de la première zone de Brillouin.

L'énergie étant une fonction paire de $\mathbf{k}$, les gradients correspondant à deux vecteurs $\mathbf{k}$ et $-\mathbf{k}$ opposés sont de signe opposés, et se compensent dans la somme, puisque dans la première zone de Brillouin, à chaque vecteur $\mathbf{k}$ correspond un vecteur $-\mathbf{k}$. On obtient donc :

$$\mathbf{v}_{tot} = 0$$

pour une bande pleine. La vitesse totale des électrons situés dans une bande pleine est donc nulle et ils ne contribuent pas à des phénomènes de transport. Pour qu'une bande puisse contribuer au courant électrique, il est donc nécessaire qu'elle ne soit que partiellement remplie. Nous verrons que c'est le cas dans deux situations :

- si le niveau de Fermi est située à l'intérieur de la bande en question (cas d'un métal),
- ou bien si un mécanisme extérieur (dopage ou température) permet de créer des places vacantes (des "trous") dans la bande d'énergie (cas d'un semi-conducteur).

2.8.5 Notion de trou

Les électrons de la bande de valence sont ceux qui participent aux liaisons de valence pour former un couche complète à 8 électrons (règle de l'octet). Par exemple, dans le cas du Silicium, les atomes de Si possèdent 4 électrons sur leur couche de valence. Dans le cristal chaque électron s'associe avec l'un des électrons de valence d'un atome voisin pour former une liaison de valence :

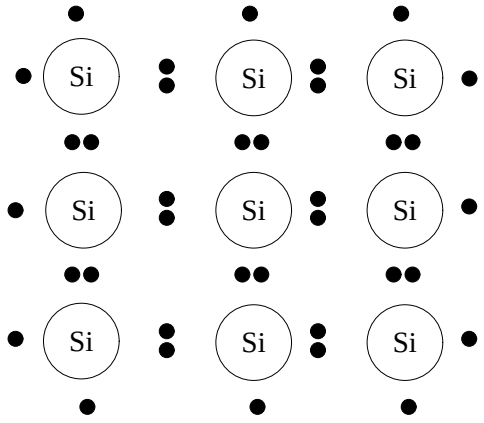


FIG. 2.9 – Représentation schématique ("à plat") des liaisons de valence entre atomes dans un cristal de Silicium.

Seuls les électrons de la bande de valence et de la bande de conduction peuvent participer à un courant, les bandes inférieures étant pleines. Le nombre d'états vacants de la bande de **valence** est très faible par rapport au nombre d'états occupés, donc par rapport au nombre d'électrons de cette bande (cf . chapitre 4, statistique de Fermi-Dirac).

La contribution de ces (nombreux) électrons au courant qui s'établit sous l'action d'un champ électrique est donnée par :

$$I = -e \sum_{\mathbf{k} \text{ occupés}} \mathbf{v}(\mathbf{k})$$

Par ailleurs, on sait que pour une bande pleine, le courant total est nul. De plus, lorsqu'on "enlève" un électron de la bande de valence, le courant total doit logiquement **diminuer** de la valeur du courant porté par cet électron avant qu'on l'enlève. Donc lorsqu'on enlève de la bande de valence initialement **pleine**, **un** électron , disons, situé dans un état de vecteur $\mathbf{k}_i$, le courant devient $-(-e\mathbf{v}(\mathbf{k}_i)$.

En répétant le processus plusieurs fois, *i.e.* en enlevant plusieurs électrons de vecteur d'onde respectifs $\mathbf{k}_1, \mathbf{k}_2, ...$, le courant électrique total (pour la bande de valence) vaut :

$$I = e \sum_{\mathbf{k} \text{ vides}} \mathbf{v}(\mathbf{k}_i)$$

Le courant de l'ensemble des électrons de la bande de valence est donc **équivalent** au courant créé par des « quasi-particules » de charge $+e$, que l'on appelle des **trous**, et qui occupent naturellement les places vacantes de la bande de valence. Dans ce cas, on peut faire comme si la bande de valence contenait, non plus des électrons, mais des trous... (et dans ce cas aussi, on ne compte plus la population d'électrons, mais celle des trous).

L'intérêt de la méthode est notamment qu'il est plus facile de compter les trous de la bande de valence que les électrons, car les premiers sont bien moins nombreux. Mais il réside aussi dans le fait que :

- l'énergie du trou est l'opposée de l'énergie de l'électron manquant ;
- le vecteur d'onde $\mathbf{k}_t$ du trou est l'opposé du vecteur d'onde de l'électron manquant ;
- la masse effective est l'opposée de celle de l'électron manquant.

De fait, alors que les électrons du haut de la bande de valence possèdent une masse effective négative (et répondent en ce sens "anormalement" à un champ électrique extérieur), les trous de la B.V. possèdent eux une masse effective positive, et répondent "normalement", *i.e.* accélèrent dans le sens de la force.

En conclusion, on peut, de manière équivalente, calculer le courant de la bande de valence, soit en comptant les contributions de tous ses électrons, soit en comptant les contributions de tous ses trous. La seconde approche consiste à considérer que la B.V. est peuplée uniquement de trous (en faible nombre), de charge positive, et surtout, de masse effective positive : en particulier, ces deux dernière caractéristique font que les trous accélèrent dans le sens du champ électrique.

Chapitre 3

Notions de physique statistique

Nous avons jusqu'alors concentré l'essentiel de notre étude sur la dynamique **d'un** électron, considéré comme "baignant" dans un potentiel (ou un champ) moyen crée par l'ensemble des autres particules du cristal, figées à leur position moyenne.

Or, le comportement réel d'un cristal de semi-conducteur, soumis de surcroît à des contraintes extérieures (différence de potentiel par exemple) est autrement plus complexe : à première vue, nous devrions prendre en compte la dynamique de **l'ensemble** des électrons (sinon des noyaux...) du cristal si nous voulions avoir accès à la grandeur qui, au final, intéresse le physicien du composant : le courant électrique. Or, cette approche est tout simplement impossible, vu le nombre d'équations de Schrödinger que nous aurions à résoudre...

En fait, lorsque nous mesurons un courant électrique à l'ampèremètre, nous n'avons pas accès au mouvement individuel de chaque électron ; au contraire, ce que nous fournit l'appareil est une grandeur macroscopique, moyennée[1] sur l'ensemble des électrons du cristal. Donc, ce qui nous intéresse finalement, ce n'est pas tant une information détaillée sur la dynamique de chaque particule du cristal, qu'une description "à gros grain" de la dynamique du gaz d'électrons : c'est précisément l'objectif de la **physique statistique**[2] que de proposer une modélisation — et des lois — qui donne du système une description *moyennée* sur l'ensemble des particules qui le composent.

On voit que cette description est assez proche dans l'esprit de celle de la thermodynamique classique[3] : on décrit un système complexe avec un petit nombre de variables macroscopique, dites *variables d'état* . En fait, la physique statistique va un peu plus loin, car elle part véritablement du modèle microscopique (ici : les bandes d'énergie, la masse effective, la répartition des électrons sur les niveaux d'énergie) pour aboutir au modèle macroscopique du système, ce que ne fait pas la thermodynamique classique (mais, en fin de compte, on aura aussi accès à l'énergie interne, l'entropie, la pression etc...).

[1] En fait, on a une petite idée du mouvement individuel des électron, car le courant est entaché d'un bruit dont l'origine microscopique est précisément la fluctuation des vitesses électroniques autour de la vitesse moyenne du gaz.

[2] Un certain nombre d'ouvrages français utilise la dénomination "thermodynamique statistique", voire "mécanique statistique" mais j'ai voulu respecter la dénomination actuellement en vigueur dans la littérature anglo-saxonne, dans la mesure où cette branche de la physique me semble englober, à la fois l'aspect *thermodynamique* par la présence de grandeurs macroscopique, et la composante *mécanique* parce qu'on part d'un modèle microscopique pour décrire le système.

[3] Ce chapitre fait largement appel à la thermodynamique : on se reportera au chapitre "notions de thermodynamique" le cas échéant.

3.1 Fonctions d'états des systèmes ouverts

Dans le cadre de l'étude statistique des semi-conducteurs, nous devrons inévitablement prendre en compte dans la modélisation, le fait que le nombre d'électrons (ou de trous) du système peut varier dans le temps : c'est typiquement le cas, par exemple, lorsqu'on met en contact, dans une diode à jonction, deux semi-conducteurs de dopage différent, et que les électrons diffusent d'une partie de la diode vers l'autre. Les systèmes que nous considérerons seront donc **ouverts**, et nous aurons besoin d'une variable d'état supplémentaire pour les décrire, en l'occurence leur nombre de particules N. Lorsque le système comportera plusieurs types de particules (par exemple, des électrons et des trous), on notera N_1, N_2, ..., N_i le nombre de particule de chaque type.

3.1.1 Variables d'état conjuguées

Considérons un système de volume V, contenant N particules, caractérisé par une température T et une pression P. Si l'on construit un nouveau système en juxtaposant deux de ces systèmes, le volume du nouveau système vaut $2V$, et il contient $2N$ particules : ces variables d'état sont qualifiées **d'extensives**, car elles sont proportionnelles à la taille du système. En revanche, la pression et la température sont **intensives**, car le nouveau système possède la même température T et la même pression P que le système de départ (ce sera également le cas du potentiel chimique).

On définit un couple de variables d'états **conjuguées** en associant à chaque variable extensive une variable intensive obtenue par la relation suivante :

$$\text{var. intensive} = \left.\frac{\partial U}{\partial \text{var. extensive}}\right|_{\text{Toutes autres var. extens. = constantes}}$$

où U est l'énergie interne.

Ainsi, à l'entropie S est associée la grandeur intensive **température** :

$$\boxed{T = \left.\frac{\partial U}{\partial S}\right|_{V,N_1,N_2,\ldots}} \tag{3.1}$$

Cette définition est fondamentale, car c'est elle qui constituent la définition de la température d'un système : en somme, la température d'un système précise de combien varie son énergie lorsque, toutes choses égales par ailleurs, son désordre s'accroît.

De même, la pression P est associée au volume (attention au signe moins) :

$$\boxed{P = -\left.\frac{\partial U}{\partial V}\right|_{S,N_1,N_2,\ldots}} \tag{3.2}$$

et précise comment varie l'énergie d'un système lorsque, toutes choses égales par ailleurs, son volume augmente.

Cette définition nous permet d'introduire une troisième variable intensive, le *potentiel chimique* μ, associée à la variable extensive N, nombre de particules du système :

$$\boxed{\mu = \left.\frac{\partial U}{\partial N}\right|_{V,S}} \tag{3.3}$$

Lorsque le système comporte plusieurs types de particules, on définit un potentiel chimique par type, *i.e.*

$$\mu_i = \left.\frac{\partial U}{\partial N_i}\right|_{V,S,N_{j\neq i}}$$

Le potentiel chimique est également appelée *niveau de Fermi* en physique du solide, et c'est l'appellation que nous utiliserons désormais. Il précise comment varie l'énergie du système lorsqu'on lui ajoute ou qu'on lui enlève une particule : c'est en quelque sorte l'énergie portée par particule.

A quoi sert le niveau de Fermi ? Nous savons intuitivement que, lorsqu'on met en contact deux systèmes à des températures différentes, l'équilibre est atteint lorsque leur température s'égalisent : l'échange de chaleur est alors nul. De même, la position prise à l'équilibre par une membrane de haut-parleur est celle qui équilibre les pressions de chaque côté de la membrane : une fois l'équilibre atteint, les volumes (en particulier celui du baffle) ne varient plus. En somme, ces variables **intensives** sont intimement liées à la notion d'équilibre : lorsqu'on met plusieurs systèmes en contact, elles s'égalisent à l'équilibre, et les variables extensives associées ne varient plus.

Le niveau de Fermi jouera le même rôle pour les systèmes **ouverts**, et permettra de caractériser l'équilibre *diffusif* : deux systèmes ouverts mis en contact par l'intermédiaire d'une paroi perméable aux particules (une jonction PN fait l'affaire...), seront à l'équilibre lorsque leur niveau de Fermi s'égaliseront : l'équilibre atteint, les nombres de particules ne varient alors plus. Nous proposerons une démonstration de cette affirmation dans le paragraphe dédié à l'étude des équilibres.

3.1.2 Énergie interne pour un système ouvert

L'énergie interne d'un système ouvert peut être caractérisée par un triplet de variables d'état : on dit que c'est un système **trivariant**. Par exemple, nous pouvons choisir les trois variables extensives S, V et N : $U = U(V, S, N)$.

Lors d'une transformation thermodynamique élémentaire, caractérisée par une variation élémentaire des variables d'état S, V et N, la variation d'énergie interne est :

$$dU = \frac{\partial U}{\partial V}dV + \frac{\partial U}{\partial S}dS + \frac{\partial U}{\partial N}dN$$

soit, en utilisant les définitions précédentes,

$$dU = -PdV + TdS + \mu dN \tag{3.4}$$

Ce qui est intéressant, c'est que l'on retrouve dans cette expression :

- le travail reçu par le système, $\delta W = -P\,dV$;
- la chaleur reçue (si la transformation suit un chemin réversible), $\delta Q = TdS$;

Le troisième terme correspond à l'énergie reçu par diffusion, *i.e.* lorsque des particules entre ou sorte du système.

3.1.3 Enthalpie libre pour un système ouvert

Dans le cas des semi-conducteurs, nous aurons à considérer des systèmes en contact avec l'air ambiant, donc subissant des transformations thermodynamique à pression et température constante : la fonction d'état n'est pas alors la mieux adaptée à ce genre de problème, car dans une transformation isotherme et isobare, le volume et l'entropie peuvent varier. On

introduit à cet effet une nouvelle fonction d'état, notée G, et appelée *enthalpie libre*. Elle est définie à partir de U et des variables d'état :

$$\boxed{G = U + PV - TS} \tag{3.5}$$

Lors d'une transformation élémentaire, la variation de G s'écrit :

$$dG = dU + PdV + VdP - TdS - SdT$$

Compte-tenu de l'expression de dU obtenue ci-dessus, on simplifie ainsi :

$$dG = VdP - SdT + \mu dN$$

Lors d'une transformation isotherme-isobare, la variation de G se simplifie en $dG = \mu dN$, d'où l'intérêt d'utiliser cette fonction d'état lorsqu'on a à considérer ce type de transformations thermodynamiques.

3.2 Caractérisation microscopique d'un système

3.2.1 Micro-états

Considérons un gaz constitué d'atomes d'hydrogène. Comment pouvons nous caractérisé ce système au niveau microscopique ? On peut d'abord spécifier la position et la vitesse (ou la fonction d'onde, selon qu'on désire une description quantique ou classique) de chaque atome. On peut également s'intéresser aux degrés de liberté intrinsèques, et préciser les nombres quantiques n,ℓ, m et s de chaque électron (ce qui revient à dire sur quel niveau d'énergie est situé chaque électron dans "son" atome, et quels sont son moment cinétique et son spin). La données de toutes ces informations détermine un **micro-état** du système.

A partir de là, le point **fondamental** est que plusieurs micro-états (en général un très grand nombre) réalise le même macro-état, *i.e.* un état **macroscopique** spécifié par la donnée d'une pression, d'une température, etc... On voit par là comment procède la description statistique du système : elle ramène une description parfaitement détaillée du système — *i.e.* en terme de micro-états — à une description macroscopique *à grands traits*, en terme de macro-états : toute la difficulté sera désormais de déterminer la correspondance surjective entre l'ensemble — très volumineux ! — des micro-états, et l'ensemble — beaucoup plus restreint, mais accessible à l'expérience — des macro-états.

Puis, au cours de son évolution dynamique, le système va "explorer" tous les micro-états qui lui sont accessibles — en fonction des contraintes d'énergie qui lui sont imposées : pour l'observateur, qui travaille, lui, dans "l'espace" des macro-états, cela se traduira par une exploration de macro-états, donc une fluctuation des variables d'état — pression, température, etc... — autour de leur valeur moyenne.

Exemple 3.1 *Considérons un système thermodynamique minimal : un électron enfermé dans un atome modélisé par un puits de potentiel tridimensionnel d'arrête L (un cube dont il est parfaitement prisonnier, ce qui est suffisant pour notre propos) . Son énergie est donnée*[4] *par $E_{n_x,n_y,n_z} = \frac{\pi^2\hbar^2}{2mL^2}(n_x^2 + n_y^2 + n_z^2)$, la fonction d'onde électronique étant donnée, à une constante près, par : $\varphi_{n_x,n_y,n_z} = \sin(n_x\pi x/L)\sin(n_y\pi y/L)\sin(n_z\pi z/L)$. Chaque micro-état est spécifié par un triplet $\{n_x, n_y, n_z\}$ — et éventuellement le nombre quantique de spin s_z—, et spécifie l'énergie du système. Imaginons que cette énergie soit justement la variable*

[4] Voir polycopié de physique quantique, chapitre "Étude de cas".

d'état du système, accessible à l'expérience (tandis que le triplet $\{n_x, n_y, n_z\}$ ne serait pas accessible) : en quelque sorte, la donnée de E spécifierait un ***macro-état*** *du système. Si, pour simplifier la notation, on pose $\varepsilon_0 = \frac{\pi^2\hbar^2}{2mL^2}$, on a par exemple :*

- $E_{1,1,1} = 3\varepsilon_0$ *:*
- $E_{1,1,2} = E_{1,2,1} = E_{2,1,1} = 6\varepsilon_0$ *;*
- $E_{1,2,2} = E_{2,1,2} = E_{2,2,1} = 9\varepsilon_0$ *;*
- $E_{2,2,2} = 12\varepsilon_0$ *;*
- ...
- $E_{2,3,6} = E_{3,2,6} = E_{2,6,3} = \ldots = 49\varepsilon_0$ *;*

On voit bien sur cette exemple que plusieurs micro-états réalise le même macro-état : l'énergie $6\varepsilon_0$ est ***dégénérée*** *trois fois, au sens où 3 micro-états la réalise. Il n'est pas difficile d'imaginer que, pour un "vrai" système thermodynamique comportant plusieurs milliards d'atomes, une même valeur macroscopique de l'énergie est réalisée par un nombre extraordinairement grand de micro-états. Or cette énergie représente alors justement l'énergie interne U du système : on a là un premier exemple de correspondance entre spécification microscopique et grandeur thermodynamique accessible à l'expérience (encore que U ne soit pas directement mesurable, mais pour un gaz parfait elle peut l'être indirectement via la température).*

3.2.2 Travail et chaleur à l'échelle microscopique

Reprenant notre exemple précédent — exemple (3.1) —, voyons comment se concrétise, à l'échelle microscopique, la notion de transfert de chaleur et de travail. Considérons un système thermodynamique constitué d'un atome "idéalisés" (boîte cubique), possédant 4 électrons, dont la répartition sur les niveaux d'énergie est indiquée figure 3.1. Lorsque le volume du système augmente, L augmente également, et la distance entre niveaux adjacents diminue : les électrons conservent leur répartition initiale (*i.e.* leurs nombres quantiques), mais l'énergie totale diminue puisque les niveaux se ressèrent (sur la figure, elle diminue de $13.5\varepsilon_0$). Nous avons vu qu'à l'échelle macroscopique, toute variation de volume traduit un transfert de travail vers le système : $\delta W = -P.dV$. D'un point de vue quantitatif, nous pouvons écrire l'énergie totale (*i.e.* l'énergie interne) de notre système sous la forme :

$$U = \sum_i N_i E_i$$

où E_i représente chacun des niveaux d'énergie du système, et N_i le nombre d'électrons sur chaque niveau : dans notre exemple, on a par exemple $N_1 = 1, N_2 = 2, N_3 = 1$. Lorsque L varie, on peut donc écrire :

$$dU = \sum_i N_i dE_i = -\frac{2}{3V}\left(\sum_i N_i E_i\right) dV = -\frac{2U}{3V} dV$$

puisque $V = L^3$ et $\frac{dE_i}{dV} = \frac{-2E_i}{3V}$.

Remarque 3.2 *Pour un gaz parfait, on a $U = \frac{3}{2}nRT = \frac{3}{2}PV$, d'où $\frac{2U}{3V} = P$, et on retrouve naturellement $dU = -PdV$.*

Si un transfert de travail mécanique correspond à une modification de la position des niveaux d'énergie sans variation de nombres quantiques, à contrario, un transfert de chaleur se traduit par une modification de la répartition des électrons sur les différents niveaux. La

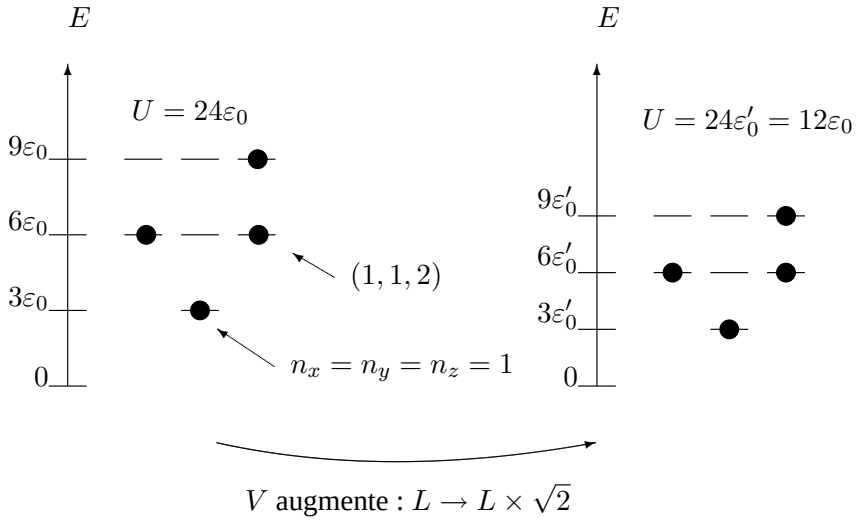


FIG. 3.1 – Toute variation lente du volume du système — ici, un atome "cubique" — se traduit par une variation d'énergie totale du système, les électrons conservant leur nombres quantiques, *i.e.* leur position initial sur les niveaux d'énergie. Dans cette exemple, l'arrête est multipliée par $\sqrt{2}$, donc $\varepsilon_0 = \frac{\pi^2\hbar^2}{2mL^2}$ est divisée par 2 : les niveaux d'énergie se ressèrent

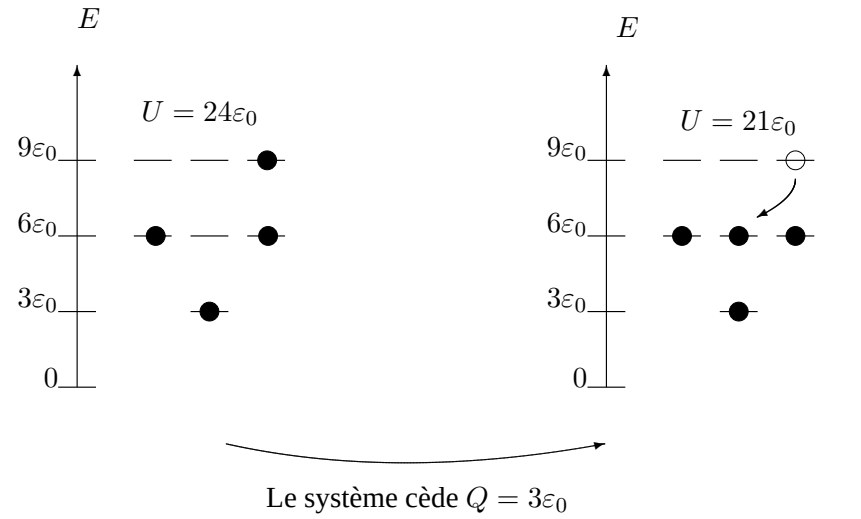


FIG. 3.2 – Le transfert de chaleur résulte de transitions énergétiques, sans modification de la position des niveaux d'énergie. Les populations électroniques N_i sur chaque niveau E_i varient, et modifient l'énergie interne du système.

figure 3.2 montre ce qui se passe lorsque le système reçoit une quantité de chaleur $Q = -3\varepsilon_0$ (*i.e.* chaleur cédée au milieu extérieur).

Là encore, on peut écrire la variation d'énergie interne du système sous la forme :

$$dU = \sum_i dN_i E_i = Q$$

Finalement, lorsque se conjuguent les deux effets, échange d'énergie et modification du volume du système, la variation d'énergie interne s'écrit :

$$dU = \sum_i dN_i E_i + N_i dE_i$$

où le premier terme s'identifie avec la chaleur reçue Q, et le second terme avec le travail échangé W.

3.3 Entropie micro-canonique

Nous allons désormais établir le lien entre la description microscopique en terme de micro-états et la mesure du désordre à l'échelle macroscopique, d'abord dans le cas d'un système isolé (hypothèse *micro-canonique*) : un tel système possède une énergie interne fixée, et ne peut échanger, ni travail, ni chaleur, avec le milieu extérieur. On sait qu'un tel système évolue toujours de telle manière que son entropie augmente (et si elle est maximum, alors c'est qu'un état d'équilibre est atteint). Le problème est donc : comment définir cette entropie ?

L'hypothèse micro-canonique affirme que :

> *A l'équilibre thermodynamique, tous les micro-états d'un système isolé sont équiprobables.*

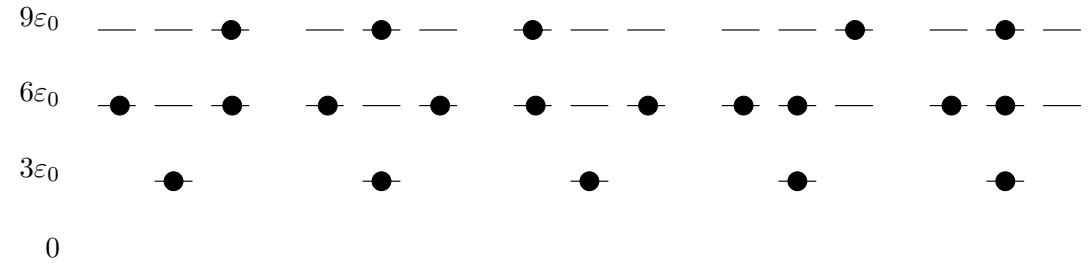
En d'autres termes, le système explore successivement tous les micro-états accessibles compte-tenu de la contrainte qui impose une énergie totale constante. Concrètement, cette exploration symbolise l'agitation thermique : les électrons passent sans arrêt d'un niveau à l'autre, définissant ainsi une succession de micro-états, mais *à énergie totale fixée*.

A partir de ce premier postulat, on peut donner une définition de l'entropie, qu'on appelle *entropie micro-canonique* :

> *L'entropie d'un système isolé, à l'équilibre thermodynamique, vaut $S = k \log \Omega$, où Ω est le nombre total de micro-états accessibles au système.*

Comme le nombre Ω de micro-états accessibles dépend de l'énergie interne U du système, on écrit souvent : $S(U) = k \log \Omega(U)$.

Exemple 3.3 *Reprenons notre exemple précédent, constitué d'un atome "cubique" possédant 4 électrons. Imaginons que son énergie interne soit fixée à la valeur $U = 24\varepsilon_0$. Compte-tenu du principe d'exclusion de Pauli, selon lequel on ne peut mettre plus d'un électron par état quantique (*i.e. *deux électrons ne peuvent pas avoir le même n-uplet de nombres quantiques), cherchons les différents micro-états accessibles au système qui respectent la contrainte $U = 24\varepsilon_0$. Ici, nous ne tiendrons exceptionnellement pas compte du spin, afin que le dénombrement des différents états reste simple : de fait, on ne peut pas mettre plus d'un électron par "case quantique" étiquetée par un triplet $\{n_x, n_y, n_z\}$ (si on tient compte du spin, on peut mettre deux électrons dans une telle case, à condition que leur spins soient antialignés). La figure suivante illustre quelques-unes des différentes configurations accessibles au système (9 en tout).*

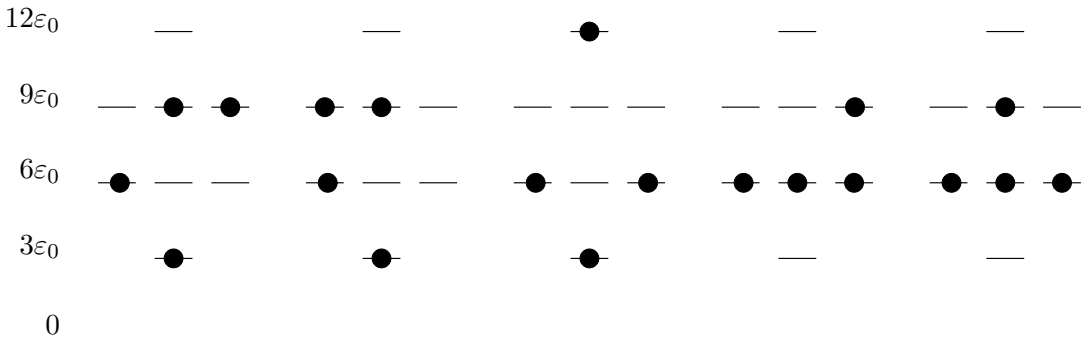


micro-état numéro 1 micro-état numéro 2 micro-état numéro 3 micro-état numéro 4 micro-état numéro 5

L'entropie du système vaut alors : $S = k \log 9 = 3.10^{-23}$ *J/ °K Quelques remarques concernant le décompte des micro-états :*

- *il faut toujours, pour des systèmes de particules* ***quantiques****, considérer les électrons comme des particules indiscernables :on ne peut pas leur mettre un coup de pinceau de couleur pour les reconnaître*[5] *! Donc, le fait d'intervertir deux électrons n'ajoute pas un micro-état à la liste...*
- *En revanche, et c'est ce qu'indique clairement la figure, si un niveau d'énergie est dégénéré (par exemple, au niveau* $6\varepsilon_0$ *correspond les trois triplets* $(1, 1, 2), (1, 2, 1), (2, 1, 1)$*), on peut déplacer un électron d'une case à l'autre, et ça fait bien un micro-état différent, car dans ce cas, on modifie le triplet* $(n_x, n_y n_z)$*, donc la fonction d'onde électronique.*

Pour bien comprendre que l'entropie dépend de l'énergie interne, on peut refaire le décompte en imposant cette fois $U = 27\varepsilon_0$*. La figure suivante montre quelques-unes des première configurations (noter que le niveau* $12\varepsilon_0$ *n'est pas dégénéré, puisqu'il lui correspond un seul triplet* $(2, 2, 2)$*) :*



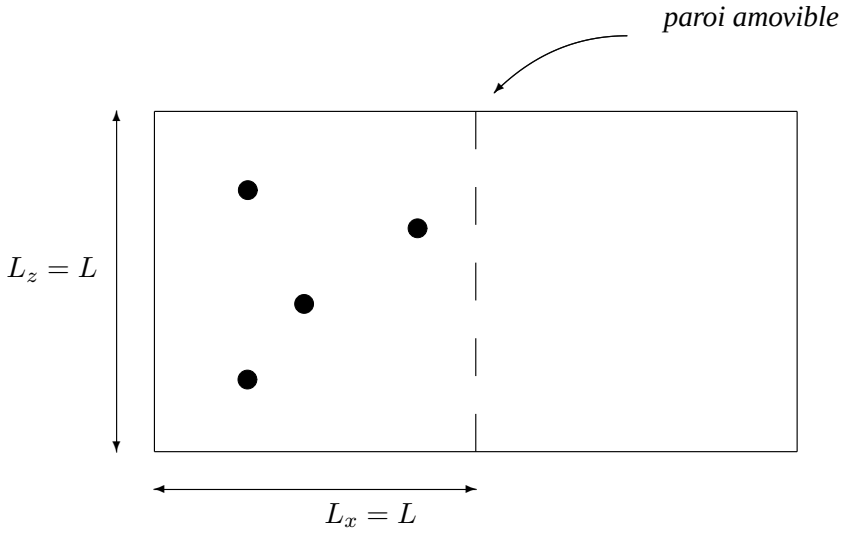
micro-état numéro 1 micro-état numéro 2 micro-état numéro 3 micro-état numéro 4 micro-état numéro 5

On dénombre en tout $9+3+3 = 15$ *micro-états, et l'entropie a bien augmenté avec l'énergie interne. Si on avait suffisamment de données à notre disposition (il faudrait calculer* S *pour suffisamment de valeurs de l'énergie interne), on pourrait même déterminer la température de ce système à l'aide de la relation* $T = \frac{\partial U}{\partial S}$*. Le point à retenir est que la plupart des*

[5] En fait, l'explication est nettement moins simpliste... mais je n'ai pas la place pour rentrer dans ce genre de détails.

systèmes sont à température positive[6] *et que par conséquent énergie interne et désordre varient dans le même sens : le nombre de micro-états accessibles croît avec l'énergie totale du système, ce qui est intuitivement compréhensible, puisqu'on "relâche" des contraintes.*

Exemple 3.4 ***Détente de Joule-Gay-Lussac***[7] ***:*** *on considère une enceinte (cf. figure ci-dessous) contenant un gaz, séparée en deux parties par une paroi amovible, et isolé du milieu extérieur. Initialement, une partie contient le gaz, et l'autre est vide. Puis on enlève brutalement la paroi, et on laisse le gaz envahir la seconde partie de manière à occuper tout l'espace disponible. Calculons l'entropie du système entre ces deux états d'équilibre, en considérant que chaque partie de l'enceinte est une boîte cubique d'arrête L.*



Le système contient initialement 4 *molécules, libres de se déplacer à l'intérieur de la demi-boîte gauche : on modélise le problème en supposant que la paroi de la boîte est équivalente à un "mur" de potentiel infranchissable, et de fait on se retrouve dans le cas d'un puits de potentiel tridimensionnel, tel que l'énergie d'une molécule est donnée par :*

$$E_{n_x,n_y,n_z} = \frac{\pi^2\hbar^2}{2m}\left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2}\right)$$

Tant que la paroi est présente, chaque demi-boîte est un cube, et par conséquent $L = L_x = L_y = L_z$*. Posons, comme précédemment,* $\varepsilon_0 = \frac{\pi^2\hbar^2}{2mL^2}$*, et fixons l'énergie initiale à* $U = 27\varepsilon_0$*. Nous avons calculé dans l'exemple précédent que l'entropie valait alors* $S = k\log 15$*. Le problème à présent, c'est de dénombrer les micro-états accessibles une fois qu'on a enlevé la paroi... en se souvenant que, puisque le système est isolé, l'énergie reste égale à* $U = 27\varepsilon_0$ *! On a désormais* $L_x = 2L$*, les autres dimensions restant inchangées, soit :*

$$E_{n_x,n_y,n_z} = \varepsilon_0\left(\frac{n_x^2}{4} + n_y^2 + n_z^2\right)$$

[6] Cette remarque semble troublante : pourtant, il existe des systèmes — les lasers et certains matériaux magnétiques — à température négative, *i.e.* tels que énergie et entropie varient en sens inverse. Un tel système possède une énergie bornée supérieurement, et possède des propriétés particulières : il présente une inversion de population (peuplement maximal des niveaux supérieurs) ; il peut céder de la chaleur à un système plus chaud ; il faut l'exciter pour faire décroître sa température... Il faut noter que cette température négative est une température thermodynamique : c'est une grandeur qui est défini à partir de U et S, pas à partir d'une échelle thermométrique ; rien ne prouve qu'elle soit mesurable (pour cela, il faudrait qu'un équilibre thermique soit possible entre le système et le thermomètre, ce qui n'est pas certain !).

[7] C'est un cas d'école en thermodynamique classique, mais ici le calcul est basé uniquement sur le modèle microscopique, contrairement à l'approche thermodynamique.

Il faut donc trouver ***tous*** *les quadruplets de triplets* (n_x, n_y, n_z) *(*i.e. *pour chacune des* 4 *molécules), tels que la somme des énergies soit égale à* $U = 27\varepsilon_0$*. Voici les premiers niveaux correspondant au (fastidieux !) décompte des différents triplets* (n_x, n_y, n_z) *(on vérifie une fois encore qu'une augmentation de volume produit un ressèrement des niveaux) :*

- $2.25\varepsilon_0 = E_{1,1,1}$;
- $3\varepsilon_0 = E_{2,1,1}$;
- $4.25\varepsilon_0 = E_{3,1,1}$;
- $5.25\varepsilon_0 = E_{1,2,1} = E_{1,1,2}$;
- $6\varepsilon_0 = E_{2,2,1} = E_{2,1,2} = E_{4,1,1}$;
- $7.25\varepsilon_0 = E_{3,1,2} = E_{3,2,1}$;
- $8.25\varepsilon_0 = E_{1,2,2} = E_{5,1,1}$;
- $9\varepsilon_0 = E_{2,2,2} = E_{4,1,2} = E_{4,2,1}$;
- $10.25\varepsilon_0 = E_{1,1,3} = E_{1,3,1} = E_{3,2,2}$;
- $11\varepsilon_0 = E_{2,1,3} = E_{2,3,1} = E_{6,1,1}$;
- $11.25\varepsilon_0 = E_{5,1,2} = E_{5,2,1}$;
- $12\varepsilon_0 = E_{4,2,2}$;
- $12.25\varepsilon_0 = E_{3,1,3} = E_{3,3,1}$;
- $14\varepsilon_0 = E_{4,3,1} = E_{4,1,3} = E_{2,2,3} = E_{2,3,2} = E_{6,1,2} = E_{6,2,1}$;
- $14.25\varepsilon_0 = E_{5,2,2} = E_{7,1,1}$;
- $15.25\varepsilon_0 = E_{3,2,3} = E_{3,3,2}$;
- $16.25\varepsilon_0 = E_{5,1,3} = E_{5,3,1}$;
- $17\varepsilon_0 = E_{4,3,2} = E_{4,2,3} = E_{6,2,2}$;
- $17.25\varepsilon_0 = E_{1,4,1} = E_{1,1,4} = E_{7,1,2} = E_{7,2,1}$;
- $18\varepsilon_0 = E_{2,1,4} = E_{2,4,1} = E_{8,1,1}$;
- ...

Quelques micro-états possibles :

- $\{E_{1,1,1}; E_{3,1,1}; E_{1,1,2}; E_{3,3,2}\}$
- $\{E_{2,1,1}; E_{2,2,1}; E_{2,2,2}; E_{4,1,2}\}$
- ...

Le dénombrement complet a été fait sous Mathematica. On commence par construire une table des micro-états (les limites sur les indices sont tels que les niveaux inférieurs à $27\varepsilon_0$ *soient tous présents) :*

```
MicroEtats =  Table[nx^2/4 + ny^2 + nz^2, {nx, 1, 9}, {ny, 1, 5}, {nz, 1,
```

On trie cette table par ordre croissant, et on ne garde que les micro-états d'énergie effectivement inférieure à $27\varepsilon_0$ *:*

```
MicroEtats = % // Flatten // Sort
MicroEtats=Take[%,78]
```

Finalement, on construit l'état du système en affectant un micro-état à chacune des 4 *molécules, puis on calcule l'énergie totale du système en prenant la somme des énergies de ces* 4 *micro-états. Si on tient compte du principe d'exclusion de Pauli (si on a affaire à un gaz d'électrons), on ne peut pas mettre plus d'un électron par micro-état, et les indices doivent toujours être 2 à 2 différents. D'autre part, le fait que les particules soient indiscernables (c'est-à-dire qu'échanger deux électrons ne changent rien au décompte) fait que les 4 boucles sont imbriquées l'une dans l'autre,* i.e. *on ne compte pas les permutations des indices.*

```
W = 1
Do[
    If[MicroEtats[[i]]+MicroEtats[j]]+MicroEtats[[k]]+MicroEtats[[l]]==27,
        {Print[W," ",i," ",j," ",k," ",l];W=W+1}],
    {i,1,78},{j,i+1,78},{k,j+1,78},{l,k+1,78}]
```

Le résultat du dénombrement est de 68 *micro-états, soit une entropie* $S = k \log 68$*. La variation d'entropie est donc* ***positive****, et vaut exactement :* $\Delta S = k \ln(68/15) = 1.51k$*. Par comparaison, le calcul de la variation d'entropie en utilisant la thermodynamique classique donne :*

$$\Delta S = nR \ln \frac{V_2}{V_1} = nR \ln 2 = Nk \ln 2 = 4k \ln 2 = 2.77k$$

(se souvenir que la constante des gaz parfaits, $R = \mathcal{N}k$*, est le produit du nombre d'Avogadro par la constante de Boltzmann). L'écart entre les deux valeurs est assez compréhensible : avec 4 molécules, on est encore assez loin de la limite* $N \to \infty$ *de la thermodynamique classique.*

Remarque 3.5 *Petit exercice intéressant : modifiez le programme pour l'adapter à des "grosses" molécules discernables et ne satisfaisant pas au principe de Pauli. Calculez le nombre de micro-états pour l'état initial et l'état final de la détente de Joule, et comparer avec les résultats de la thermodynamique classique. Au passage, vous devez obtenir beaucoup plus de micro-états (quelques dizaines de milliers...)*

On observe donc que, même pour un "petit" système composé de seulement 4 *molécules, l'entropie croît de manière vertigineuse lorsqu'on relâche un peu les contraintes (ici, le volume occupable) : il est donc évident que l'état final a une probabilité d'apparaître au cours de la vie du système beaucoup plus élevée que l'état initial, et que les chances d'observer un retour* ***spontané*** *à l'état initial sont extrêmement faibles... C'est en ce sens que cette évolution est irréversible : en théorie, rien n'empêche le système de revenir à son état initial (*i.e. *tout le gaz concentré dans la demi-boîte gauche), simplement c'est extrêmement improbable...*

3.4 Entropie d'information

3.4.1 Comment quantifier l'information manquante ?

La définition précédente de l'entropie ne s'applique qu'aux systèmes qui ont déjà atteint un équilibre thermodynamique : c'est une définition qui a cependant l'avantage immense d'être cohérente avec la définition de l'entropie proposée par Clausius en thermodynamique classique.

Une autre définition a été proposée au début du XX^{e}siècle, en liaison avec la théorie de l'information. Elle permet de généraliser aux systèmes hors d'équilibre, *i.e.* entre les deux états d'équilibre d'une transformation thermodynamique, la notion de désordre, ou plus précisément ici, **d'information manquante** : cette quantité mesure **l'information manquante** pour connaître parfaitement l'état **microscopique** du système à partir de son état **macroscopique**. Si l'état macroscopique nous permettait de déduire immédiatement l'état microscopique, c'est que le système ne pourrait explorer qu'un seul micro-état, et l'entropie d'information serait nulle. Au contraire, si le système peut explorer un grand nombre de micro-état tout en donnant l'illusion d'être dans le même macro-état (même énergie interne, même température, etc...), alors il nous manque énormément d'information : en d'autres termes, si on

essaie de deviner l'état microscopique du système à partir des variables macroscopiques, on a de très fortes chances de se tromper...

En outre, l'entropie d'information ne s'applique pas seulement aux système isolés : elle nous intéresse donc particulièrement, puisque les systèmes auxquels on s'intéressera par la suite sont généralement en contact thermique avec le milieu extérieur.

L'entropie d'information s'écrit[8], que le système soit isolé ou non :

$$\boxed{S = -k \sum_i \mathcal{P}_i \ln \mathcal{P}_i} \tag{3.6}$$

où $\mathcal{P}_i$ est la probabilité d'occurence du micro-état i, et où la somme porte sur tous les micro-états accessibles au système (de toute façon, les autres ont une probabilité d'occurence nulle...). L'idée clé ici est que, hors de l'équilibre, tous les micro-états ne sont pas équiprobables, et c'est assez logique finalement : dans la détente de Joule-Gay-Lussac par exemple, on part d'un état initial où 15 micro-états sont accessibles, et le système s'arrête dans un état final à 182 micro-états **équiprobables** ; il est à peu près évident que les 182 micro-états finals ne sont pas immédiatement accessibles avec la même probabilité, mais qu'en quelque sorte, le système commence à explorer essentiellement ceux qui sont le plus "proches" des 15 micro-états initiaux, puis élargit progressivement sont exploration jusqu'à passer le même temps dans chacun des 182 micro-états finals [9]

On observe toutefois que, à l'équilibre, la définition précédente colle avec celle de l'entropie micro-canonique à l'équilibre, puisqu'alors

$$\forall i, \mathcal{P}_i = \frac{1}{\Omega}$$

où Ω est le nombre total de micro-états accessibles **à l'équilibre**, — ils sont dans ce cas tous équiprobables — et que de ce fait,

$$S = k \sum_i \frac{1}{\Omega} \ln \Omega = k \ln \Omega$$

De même, il est intéressant de remarquer que, si on connaît **parfaitement** l'état microscopique d'un système — *i.e.* on sait **à coup sûr** qu'il est dans le micro-état numéro i – alors tous les $\mathcal{P}_i$ sauf un sont nuls, et l'entropie vaut 0.

3.4.2 Principe d'information minimum

Le principe d'information minimum est le pendant microscopique du second principe, mais il va plus loin : tandis que le second principe de s'intéresse qu'aux états d'équilibre initial et final, le principe d'information minimum permet de savoir ce qui se passe **entre** ces deux états. Il affirme en effet que

> *L'équilibre d'un système est atteint lorsque l'entropie d'information (= l'information manquante) est maximale.*

[8] Je n'ai pas la place de vous montrer comment on construit cette expression, mais vous pouvez toujours jeter un coup d'œil dans un cours de théorie de l'information : généralement, il y a ça dans tout bon cours de transmission du signal.

[9] En fait, il y a plusieurs interprétations de la transition vers l'équilibre, et ce que je vous propose ici est un choix personnel — qui me convainc particulièrement — parmi toutes les interprétations actuelles.

ce qui signifie que le système tend, lors de son évolution du passé vers le futur, à explorer le maximum de micro-états, et s'arrête lorsque ceux-ci sont tous équiprobables : l'information manquante est maximum, ou ce qui revient au même l'information sur l'état microscopique est minimale. Là aussi, il y a quelque chose d'intuitivement rassurant : on ne voit pas pourquoi le système se restreindrait spontanément à n'explorer qu'une partie des états accessibles (*i.e.* par exemple : la demi-boîte gauche dans la détente de Joule) si rien ne l'empêche de faire plus...

D'un point de vue quantitatif, son évolution est donnée, à tout instant, par l'inégalité :

$$\boxed{\frac{dS}{dt} \geq 0} \tag{3.7}$$

qui ne fait que traduire mathématiquement le principe d'information minimum. L'**équilibre** est défini, lui, par l'égalité :

$$\boxed{\frac{dS}{dt} = 0} \tag{3.8}$$

ce qui signifie, qu'au premier ordre, l'entropie est constante à l'équilibre (mais des fluctuations d'ordre supérieur ou égal à 2 existent).

3.4.3 Application à l'évolution d'un système isolé

Pour appliquer ce principe aux situations qui nous intéressent plus particulièrement, commençons par considérer le cas des systèmes isolés, *i.e.* à énergie interne constante. Écrivons que l'entropie d'information est extrémale :

$$dS = -k \sum_i d\mathcal{P}_i \ln \mathcal{P}_i - k \sum_i d\mathcal{P}_i = -k \sum_i (\ln \mathcal{P}_i + 1) d\mathcal{P}_i = 0$$

et cherchons l'expression des probabilités de chaque micro-état à l'équilibre. Outre l'équation précédente, les probabilité doivent également satisfaire à la contrainte :

$$\sum_i \mathcal{P}_i = 1$$

soit encore,

$$\sum_i d\mathcal{P}_i = 0$$

ce qui fait qu'on doit au final chercher les valeurs des $\mathcal{P}_i$ qui maximisent une fonction sous contrainte. Pour résoudre ce genre de problème, on utilise par exemple la méthode des multiplicateurs de Lagrange[10], ce qui conduit à :

$$\forall i, 1 + \ln \mathcal{P}_i + \lambda = 0$$

où λ est une constante à déterminer. L'intégration conduit à :

$$\forall i, \mathcal{P}_i = e^{-1-\lambda}$$

ce qui montre bien que tous les micro-états sont équiprobables à l'équilibre ! Comme il y a en tout Ω micro-états, chaque micro-état possède une probabilité d'occurence

$$\mathcal{P}_i = \frac{1}{\Omega}$$

[10] Voir annexe.

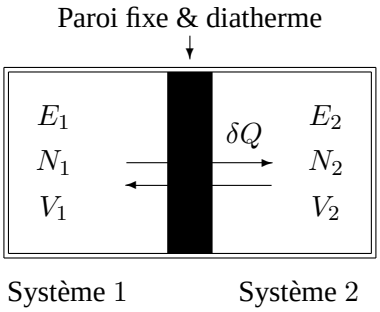


FIG. 3.3 – Mise en contact thermique de deux systèmes isolés thermiquement de l'extérieur : l'équilibre s'établit lorsque les températures s'équilibrent.

et à l'équilibre, on retrouve bien $S = k \ln \Omega$: en somme, postuler le principe d'information minimum (et accessoirement trouver une expression **satisfaisante** pour l'entropie d'information...) permet de retrouver le postulat d'équiprobabilité des micro-états à l'équilibre d'un système isolé.

3.5 Équilibres thermodynamiques

Notre objectif ultime est, pour cette partie du cours, de déterminer les conditions d'équilibre du système thermodynamique que constitue une diode PN. On peut voir ce dispositif comme étant constitué de deux sous-systèmes, en l'occurence deux cristaux de semiconducteurs "accolés" et présentant un dopage différent (l'un à forte densité d'électrons, l'autre à forte densité de trous). L'étude de la distribution de probabilité des micro-états de ce système fera l'objet de la section suivante, dans la mesure où le système en question n'étant pas isolé, il requiert un traitement particulier.

L'objet de cette partie est simplement de montrer comment les variables d'états intensives permettent, à partir de l'expression microscopique du second principe (notre principe d'information minimum), de trouver les conditions d'équilibre de deux systèmes mis en contact, soit mécanique, soit thermique, soit diffusif, soit enfin une combinaison de ces trois situations. Ces résultats seront fondamentaux pour la compréhension du rôle joué par le niveau de Fermi dans l'apparition de l'équilibre thermodynamique au sein d'une jonction PN.

3.5.1 Équilibre thermique

Considérons deux systèmes en contact thermique (fig 3.3), d'énergie respective E_1 et E_2. Ces deux systèmes sont limités par une enveloppe commune qui isole **l'ensemble** de l'extérieur : la chaleur ne peut transiter que du système 1 vers le système 2, et *vice-versa* ; l'énergie totale du "sur-système", $E_T = E_1 + E_2$, est constante. L'ensemble possède donc un seul degré de liberté, soit E_1 par exemple. L'évolution de l'ensemble des deux systèmes, considéré comme un sur-système isolé, est régie par le principe d'information minimum, équation (3.7), où S est l'entropie de l'ensemble des deux systèmes. Si on appelle S_1 et S_2 l'entropie respective de chaque système, on a bien entendu

$$S = S_1 + S_2$$

et l'évolution est régie par

$$\frac{dS}{dt} \geq 0$$

soit encore,

$$\frac{dS_1}{dt} + \frac{dS_2}{dt} \geq 0$$

Ce qui nous intéresse, c'est :

- le sens de transfert de chaleur à l'approche de l'équilibre ;
- les conditions d'équilibre.

On va donc exprimer l'inégalité précédente en fonction de E_1 (ou indifféremment de E_2 puisque les deux sont liés) :

$$\frac{dS_1}{dE_1}\frac{dE_1}{dt} + \frac{dS_2}{dE_2}\frac{dE_2}{dt} \geq 0$$

soit encore, puisque $dE_1 = -dE_2$,

$$\frac{dS_1}{dE_1}\frac{dE_1}{dt} - \frac{dS_2}{dE_2}\frac{dE_1}{dt} \geq 0$$

qui s'écrit finalement :

$$\left(\frac{dS_1}{dE_1} - \frac{dS_2}{dE_2}\right)\frac{dE_1}{dt} \geq 0$$

Si on se souvient que la température thermodynamique est donnée par (equ. (3.1)) :

$$T = \frac{\partial E}{\partial S}$$

on peut mettre l'inégalité précédente sous la forme fondamentale suivante :

$$\boxed{\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\frac{dE_1}{dt} \geq 0} \tag{3.9}$$

L'interprétation en est extrêmement simple : tant que l'équilibre n'est pas atteint, l'entropie de l'ensemble ne peut qu'augmenter, ce qui implique que le transfert de chaleur ait lieu des températures élevées vers les températures basses (en effet, l'inégalité impose que la dérivée de E_1 par rapport au temps soit du signe de $T_2 - T_1$).

L'équilibre a lieu lorsque $\frac{dS}{dt} = 0$. Reprenant le même raisonnement, mais avec une égalité cette fois, on voit que l'équilibre est atteint lorsque $T_1 = T_2$: en effet, l'égalité impose que, *quelles que soient les petites fluctuations* de l'énergie de chaque système autour de sa valeur moyenne, l'entropie reste nulle au premier ordre.

La température joue donc un rôle de **potentiel thermodynamique** vis-à-vis des échanges thermiques : à l'instar d'un potentiel électrique (penser à un condensateur se déchargeant dans une résistance), elle fixe *le sens d'écoulement de l'énergie* et la condition d'équilibre (équivalent du régime stationnaire dans le cas du condensateur).

Au-delà, ce qu'il importe de retenir est que l'équilibre thermodynamique n'implique pas "l'immobilité" : les variables d'états continuent de fluctuer autour de leur position d'équilibre, et ces fluctuations constituent la manifestation tangible de l'agitation thermique. L'équilibre signifie simplement que le désordre du système est maximum (compte-tenu des contraintes exercées), et que le système ne peut aller plus loin...

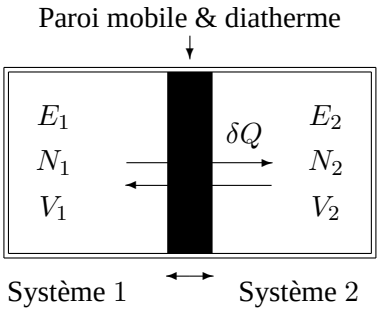


FIG. 3.4 – Mise en contact thermomécanique de deux systèmes isolés de l'extérieur : l'équilibre s'établit lorsque les températures **et les pressions** s'équilibrent.

3.5.2 Équilibre mécanique

On considère à présent un système constitué de deux sous-systèmes pouvant échanger de la chaleur, et séparés par un piston mobile (fig. 3.4) : nous allons alors mettre en évidence le rôle de potentiel thermodynamique joué par la **pression**.

On appelle respectivement V_1 et V_2 les volumes des sous-systèmes : comme le système total est isolé **mécaniquement** du milieu extérieur, le volume total $V_T = V_1 + V_2$ est constant. En plus de la variable E_1 (identique au cas précédent), nous avons donc un second degré de liberté, V_1 par exemple. Nous nous intéresserons non seulement au sens de transfert de chaleur, mais aussi à l'évolution des volumes, *i.e.* au sens de déplacement du piston.

L'équilibre est atteint lorsque l'entropie atteint son maximum :

$$\frac{dS_1}{dt} + \frac{dS_2}{dt} \geq 0$$

Cette fois, S_1 comme S_2 dépendent chacune de l'énergie **et** du volume, d'où :

$$\frac{\partial S_1}{\partial E_1}\frac{dE_1}{dt} + \frac{\partial S_1}{\partial V_1}\frac{dV_1}{dt} + \frac{\partial S_2}{\partial E_2}\frac{dE_2}{dt} + \frac{\partial S_2}{\partial V_2}\frac{dV_2}{dt} \geq 0$$

Comme, d'une part,$dE_1 = -dE_2$, et d'autre part,$dV_1 = -dV_2$, la relation précédente se simplifie en :

$$\left(\frac{\partial S_1}{\partial E_1} - \frac{\partial S_2}{\partial E_2}\right)\frac{dE_1}{dt} + \left(\frac{\partial S_1}{\partial V_1} - \frac{\partial S_2}{\partial V_2}\right)\frac{dV_1}{dt} \geq 0$$

Ici encore, on fait appel à la définition de la température (equ. (3.1)) et de la pression (equ. (3.2)) ce qui conduit à l'expression suivante pour l'approche de l'équilibre :

$$\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\frac{dE_1}{dt} + \left(\frac{P_1}{T_1} - \frac{P_2}{T_2}\right)\frac{dV_1}{dt} \geq 0$$

A l'équilibre **thermique**, le premier terme disparaît (cf. cas précédent), reste donc l'inégalité donnant l'évolution des volumes,

$$\boxed{\left(\frac{P_1}{T_0} - \frac{P_2}{T_0}\right)\frac{dV_1}{dt} \geq 0} \tag{3.10}$$

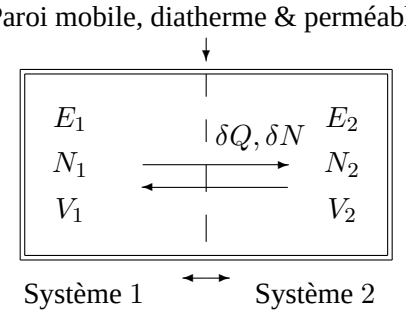


FIG. 3.5 – Mise en contact thermique, mécanique, et diffusif de deux systèmes isolés de l'extérieur : l'équilibre s'établit lorsque températures, pression, et niveau de Fermi s'équilibrent.

où T_0 est la température d'équilibre thermique. Le volume V_1 croît si $P_1 \geq P_2$, comme le dicte l'intuition : au passage, il faut ici la simple conséquence d'une définition **cohérente** de la pression, rien d'autre[11] !

L'équilibre lorsque l'inégalité précédente devient une égalité : c'est le cas si, *quelques soient les petites fluctuations de volumes*, les pressions s'égalisent (là aussi, confirmation d'une observation expérimentale bien banale, mais ça vaut là peine de le montrer à partir du second principe). Au final, on observe que la pression, comme la température, joue un rôle de potentiel thermodynamique vis-à-vis de sa **variable extensive conjuguée**, le volume. C'est là un point important : chaque variable intensive (T, P, μ) joue un rôle de potentiel vis-à-vis de sa variable extensive conjugués, et dicte les conditions d'équilibre.

3.5.3 Équilibre diffusif

Nous allons raisonner comme précédemment, mais en considérant un système constitué de deux sous-systèmes pouvant échanger des particules au travers d'une paroi perméable (fig. 3.5). Il faut voir que le raisonnement est **strictement** le même, sauf que nous ajoutons un troisième degré de liberté au système, le nombre de particule dans chaque boîte.

Ici aussi, le fait que le système soit isolé impose que le nombre totale de particule, $N_T = N_1 + N_2$ soit constant. On peut donc choisir comme degré de liberté supplémentaire, N_1 par exemple. L'approche de l'équilibre conduit à la même inégalité que précédemment,

$$\frac{\partial S_1}{\partial E_1}\frac{dE_1}{dt} + \frac{\partial S_1}{\partial V_1}\frac{dV_1}{dt} + \frac{\partial S_1}{\partial N_1}\frac{dN_1}{dt} + \frac{\partial S_2}{\partial E_2}\frac{dE_2}{dt} + \frac{\partial S_2}{\partial V_2}\frac{dV_2}{dt} + \frac{\partial S_2}{\partial N_2}\frac{dN_2}{dt} \geq 0$$

Compte-tenu de $dN_1 + dN_2 = 0$, et en utilisant la définition du *potentiel chimique - niveau de Fermi*, on obtient,

$$\left(\frac{1}{T_1} - \frac{1}{T_2}\right)\frac{dE_1}{dt} + \left(\frac{P_1}{T_1} - \frac{P_2}{T_2}\right)\frac{dV_1}{dt} + \left(-\frac{\mu_1}{T_1} + \frac{\mu_2}{T_2}\right)\frac{dN_1}{dt} \geq 0$$

A l'équilibre thermique et mécanique, les deux premiers termes disparaissent. Reste donc :

$$\boxed{\left(-\frac{\mu_1}{T_0} + \frac{\mu_2}{T_0}\right)\frac{dN_1}{dt} \geq 0} \tag{3.11}$$

où T_0 est la température d'équilibre. L'interprétation de μ est la suivante :

[11] En fait, c'est la définition des potentiels thermodynamiques qui découle de l'inégalité, et non l'inverse, mais par soucis de clarté, je n'ai pas procédé ainsi — ce qui par ailleurs eut été plus long à mettre en place.

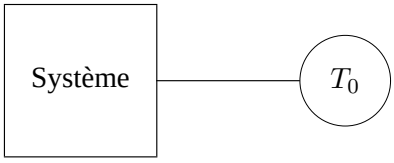


FIG. 3.6 – Le thermostat, à la température T_0, est un réservoir d'énergie de capacité calorifique idéalement infinie, en contact thermique avec le système.

A travers la paroi, les particules diffusent de la région à niveau de Fermi élevé vers la région à faible niveau de Fermi

Il est évident que cette règle est fondamentale pour la compréhension de la diffusion des électrons et des trous dans une jonction PN, où deux matériaux, de niveaux de Fermi différents (de par le dopage), sont en contact.

A l'équilibre, l'égalité impose cette fois-ci que ***les niveaux de Fermi s'égalisent****.*

3.6 Systèmes non-isolés

3.6.1 Systèmes en contact avec une source de température - statistique de Boltzmann

Intéressons-nous à présent au cas des systèmes non-isolés. Nous considérerons d'abord des systèmes fermés, *i.e.* à nombre N de particules constant, en contact avec un thermostat, *i.e.* un réservoir de chaleur de capacité calorifique suffisamment élevée (= inertie thermique) pour que l'on puisse considérer sa température constante, quelle que soient les échanges thermiques qui puissent avoir lieu entre le thermostat et le système (cf. figure 3.6).

Là encore, l'état d'équilibre est fixé par le maximum d'entropie. Mais nous avons une contrainte supplémentaire par rapport au cas des systèmes isolés : puisque le système est en contact avec l'extérieur, en l'occurence une source de température, son énergie est constante, non plus en valeur instantanée, mais seulement en valeur moyenne : la source de température permet à l'énergie du système de fluctuer autour de sa valeur moyenne, cette dernière étant imposée une fois **l'équilibre atteint**. C'est d'ailleurs, là encore, une manifestation de l'agitation thermique : il y a en permanence échange d'énergie entre le système et le thermostat, via les chocs moléculaires et les transitions énergétiques entre niveaux, mais la grande capacité calorifique du thermostat impose que ces échange s'équilibrent en moyenne[12].

Nous devons donc à présent trouver les probabilités $\mathcal{P}_i$ qui maximisent l'entropie, avec deux contraintes supplémentaires :

$$\sum_i \mathcal{P}_i = 1 \text{ et } \langle E \rangle = U = \sum_i \mathcal{P}_i E_i = \text{constante}$$

la seconde relation exprimant la contrainte imposée sur l'énergie moyenne, *i.e.* l'énergie interne[13]. En différenciant, on a pour cette dernière équation :

$$\sum_i E_i d\mathcal{P}_i = 0$$

[12] Je reconnais que je ne justifie pas **rigoureusement**, si l'on puis dire, cette hypothèse. J'essaie tout juste de vous donner une raison, qualitative certes, mais qui semble logique et cohérente avec le reste. Car en fait, la justification exacte repose sur le formalisme *canonique*, et demanderait d'aller malheureusement beaucoup plus loin dans cet exposé. Sorry !

[13] A ce propos, on se souviendra que l'énergie interne U d'un gaz parfait vaut $\frac{3}{2}nRT$, et est donc imposée dés lors que la température est fixée par un thermostat. Ici, j'ai volontairement différenciée la notation de l'énergie totale E, et de l'énergie interne, qui est la **moyenne** de l'énergie totale.

(attention, le système n'échange pas de **travail mécanique** avec l'extérieur, donc les niveaux d'énergie ne varient pas !), combinée avec les équations du paragraphe précédent, à savoir :

$$\sum_i d\mathcal{P}_i = 0$$

$$dS = -k \sum_i (\ln \mathcal{P}_i + 1) d\mathcal{P}_i = 0$$

On utilise à nouveau les multiplicateurs de Lagrange, puisque c'est si pratique et que Lagrange a pas fait tout ce boulot pour rien :

$$\ln \mathcal{P}_i + 1 + \lambda_1 + \lambda_2 E_i = 0$$

où, vous l'avez compris, les deux constantes sont à déterminer... L'intégration donne :

$$\mathcal{P}_i = e^{-1-\lambda_1-\lambda_2 E_i} = \frac{1}{Z} e^{-\beta E_i}$$

Point à noter immédiatement : *pour un système non-isolé, tous les micro-états ne sont pas équiprobables* : cette distribution de probabilité est appelée **distribution de Boltzmann**.

La constante Z est appelée **fonction de partition du système**, et, en utilisant la contrainte $\sum_i \mathcal{P}_i = 1$, on trouve qu'elle vaut :

$$Z = \sum_i e^{-\beta E_i}$$

Remarque 3.6 *Je n'insiste pas sur l'intérêt de la fonction de partition Z, car on irait trop loin, mais il faut savoir que cette fonction est très pratique parce qu'elle permet à elle seule d'exprimer toutes les fonctions d'état du système (énergie interne, pression, énergie libre, etc...). Par exemple, l'énergie libre s'écrit $F = -kT \ln Z$, l'énergie interne $U = kT^2 \frac{\partial \ln Z}{\partial T}$, $P = kT \frac{\partial \ln Z}{\partial V}$, etc...*

Trouver la constante β est un poil plus compliqué (en fait, c'est pas compliqué, simplement il faut y penser !) : on va "remonter" à la thermodynamique et voir si on peut l'identifier avec une variable d'état connue[14], tout en utilisant les équations correspondant aux contraintes (notamment, on n'a pas encore utilisé celle relative à l'énergie). Reprenons l'expression de dS en fonction des probabilités $\mathcal{P}_i$:

$$dS = -k \sum_i (\ln \mathcal{P}_i + 1) d\mathcal{P}_i$$

et remplaçons $\mathcal{P}_i$ par son expression en fonction des multiplicateurs de Lagrange :

$$dS = -k \sum_i (-\ln Z - \beta E_i + 1) d\mathcal{P}_i$$

Cette expression se simplifie très proprement, puisque $\sum_i d\mathcal{P}_i = 0$:

$$dS = k\beta \sum_i E_i d\mathcal{P}_i = k\beta d\langle E \rangle = k\beta dU$$

[14] Encore une fois, tout ceci est beaucoup plus naturel dans le formalisme de l'ensemble canonique, mais ça demanderait trop de prérequis. Voir la bibliographie en annexe.

Or (et là : premier point fondamental dans la compréhension du raisonnement), on a dit au départ que le système *n'échangeait pas de travail mécanique* avec l'extérieur, mais seulement de la chaleur : donc $dU = Q$, W étant nul.

Second point important : toutes les différentielles que vous avez au-dessus sont relatives à de petits déplacements *autour de l'équilibre* (logique, puisqu'on cherchait précisément les probabilité à l'équilibre...), et ces déplacements à la limite *infinitésimaux* correspondent à des fluctuations **réversibles**, *i.e.* **sans production d'entropie** δS^p. Du coup, toute variation d'entropie autour de l'équilibre se limite à l'entropie **reçue**, qui s'écrit :

$$dS = \frac{\delta Q}{T}$$

où, *bien entendu*, T est la température du thermostat. Voilà : à partir de là, on identifie $k\beta$ avec $1/T$, d'où :

$$\beta = \frac{1}{kT}$$

et finalement, l'expression complète de la distribution de Boltzmann est :

$$\mathcal{P}_i = \frac{1}{Z} e^{-E_i/kT}$$

Au cours de son évolution, le système explore tous les micro-états numéro i accessibles, mais avec des probabilités déterminées par cette distribution : on voit bien que, contrairement au cas du système isolé, son énergie totale peut fluctuer (autour de $U = \langle E \rangle$).

On peut d'ailleurs s'intéresser à la probabilité de le trouver avec une certaine énergie E : pour cela, il faut tenir compte de ce que plusieurs micro-états peuvent correspondre à la même énergie (cf. l'exemple de la détente de Joule dans un paragraphe précédent). Si on note $g(E)$ le **degré de dégénérescence** du niveau d'énergie E (*i.e.* le nombre de micro-états qui ont l'énergie E), la probabilité de trouver le système dans l'état d'énergie E est donnée simplement par la somme des probabilités d'apparition de tous les micro-états d'énergie E, soit :

$$\mathcal{P}(E) = \frac{1}{Z} g(E) e^{-E/kT}$$

Remarque 3.7 *$g(E)$ est également appelée **densité d'état** lorsque les niveaux d'énergie sont suffisamment ressérés pour qu'on puisse supposer que l'énergie varie continûment : c'est le cas notamment des bandes d'énergie dans les semi-conducteurs, et nous reviendrons sur ce point dans le paragraphe relatif à la distribution de Fermi-Dirac.*

Comme $g(E)$ est une fonction qui croît *en général* avec l'énergie, $\mathcal{P}(E)$ est une fonction qui commence par croître, présente un maximum, puis décroît très vite comme une exponentielle, et ce d'autant plus vite que la température est basse. Généralement, les niveaux d'énergie **inférieurs** sont donc les plus peuplés (ou ceux qui sont peuplés "le plus fréquemment" par le système) : cependant, il n'est donc pas impossible d'observer un système à basse température avec momentanément une énergie très élevée, mais ceci est simplement hautement improbable. Ceci étant, ce n'est pas impossible, et cette affirmation est fondamentale pour comprendre la dépendance thermique des réaction chimiques : lorsqu'on augmente la température, on "allonge" la distribution de Boltzmann vers la droite, et on permet aux molécules d'explorer plus fréquemment des états d'énergie élevés ; de fait, elles ont plus de chance, soit de traverser les barrières de potentiel répulsives créées par leur voisine (cas de la création d'une liaison chimique), soit de briser une liaison par choc moléculaire à vitesse élevée... Le résultat dépend de la compétition entre ces deux effets !

3.6.2 Systèmes en contact avec une source de température et un réservoir de particules - statistique de Gibbs

Les semi-conducteurs utilisés dans un dispositif électronique notamment, ne sont pas des systèmes fermés : en effet, dés lors qu'on fait circuler un courant dans le dispositif (en réalité même, dés lors qu'on fabrique une diode...), le système échange non seulement de l'énergie **mais aussi** des électrons avec le milieu extérieur. Ainsi, dans le cas d'une jonction PN constitué de deux semi-conducteurs de dopage différent (et constituant chacun un système thermodynamique), il y a échange d'électrons et de trous entre les deux. Nous allons, selon le même raisonnement que précédemment formuler une nouvelle distribution qui tiennent compte du fait que le nombre N de particules du système peut varier.

Nous allons considérer pour notre modèle que le système est en contact avec un réservoir d'énergie et de particules. Ce réservoir possède une capacité calorifique infinie, ce qui garantit que sa température est constante quels que soient les échanges d'énergie. Il possède également une taille suffisante pour que, quels que soient les échanges de particules entre le système et le réservoir, le potentiel chimique du réservoir reste constant (rappel : le potentiel chimique est la variable d'état qui détermine l'équilibre de diffusion).

A l'équilibre , un tel réservoir impose au système, à la fois une énergie moyenne constante, et un **nombre moyen de particules** constant[15].

L'état d'équilibre est là encore déterminé par le maximum d'entropie, assortie à présent de trois contraintes :

$$\sum_i \mathcal{P}_i = 1$$

et

$$\langle E \rangle = U = \sum_i \mathcal{P}_i E_i = \text{constante}$$

d'une part,

$$\langle N \rangle = \sum_i \mathcal{P}_i N_i = \text{constante}$$

d'autre part. Il y a a priori une petite difficulté conceptuelle sur laquelle il faut s'arrêter si on veut comprendre l'intérêt de cette nouvelle distribution par rapport à celle de Boltzmann. Dans cette dernière équation, N_i représente le nombre de particules du système se trouvant dans le micro-état numéro i, d'énergie E_i. Dans les deux cas précédents (système isolé et système en contact avec un thermostat), on supposait que le nombre de particules (par exemple, les 4 molécules de la détente de Joule) était fixé, et *on ne se préoccupait pas* de savoir si une ou plusieurs de ces molécules étaient dans le même état : c'est effectivement légitime dans le cas où le système est composé de (grosses) molécules, *i.e.* de particules que l'on peut décrire de manière classique (par opposition à "quantique"). Mais ça l'est beaucoup moins pour des électrons, dont on sait qu'ils ne peuvent occuper un même micro-état (*i.e.* mêmes nombres quantiques, spin compris). De fait, il est indispensable pour un système d'électrons (ou toute particule à comportement "fortement" quantique comme les nucléons du noyau ou les photons) de tenir compte du nombre de particules se trouvant dans chaque micro-état numéro i. En l'occurence, pour un système d'électrons, on verra en fin de parcours qu'il faudra imposer :

$$\forall i, N_i \leq 1$$

[15] J'insiste sur le fait que c'est seulement à l'équilibre que ces moyennes sont constantes : lors de l'approche de l'équilibre, elles varient bien entendu...

pour respecter le principe d'exclusion de Pauli[16].

Passons à présent à la résolution de l'équation d'équilibre (maximum d'entropie) en tenant compte des 3 contraintes précédentes. A l'aide des désormais traditionnels multiplicateurs de Lagrange, on obtient :

$$1 + \ln \mathcal{P}_i + \lambda_1 + \lambda_2 E_i + \lambda_3 N_i = 0$$

La résolution conduit à :

$$\mathcal{P}_i = \frac{1}{Z} e^{-\lambda_2 E_i - \lambda_3 N_i}$$

Comme précédemment, Z est la **fonction de partition** du système, et s'exprime ainsi :

$$Z = \sum_i e^{-\lambda_2 E_i - \lambda_3 N_i}$$

Trouver les deux autres multiplicateurs, *i.e.* λ_1 et λ_2, procède de la même démarche qu'au paragraphe précédent. On écrit que, pour de petites évolutions réversibles autour de l'équilibre, la variation d'entropie est :

$$dS = -k \sum_i (1 + \ln \mathcal{P}_i) d\mathcal{P}_i$$

et on remplace $\mathcal{P}_i$ par son expression en fonction des multiplicateurs :

$$dS = -k \sum_i (1 - \ln Z - \lambda_2 E_i - \lambda_3 N_i) d\mathcal{P}_i$$

Comme précédemment — et comme toujours de toute façon ! — $\sum_i d\mathcal{P}_i = 0$, et cette expression se simplifie en :

$$dS = -k \sum_i (-\lambda_2 E_i - \lambda_3 N_i) d\mathcal{P}_i = k\lambda_2 d\langle E \rangle + k\lambda_3 d\langle N \rangle$$

que l'on peut indifféremment écrire,

$$dS = k\lambda_2 dU + k\lambda_3 d\bar{N}$$

à condition de se souvenir que $\bar{N}$ est la valeur moyenne du nombre de particules dans le système. En identifiant avec la relation thermodynamique donnée par l'équation (3.4),

$$dS = \frac{dU}{T} - \frac{\mu}{T} dN$$

on trouve que

$$\lambda_2 = \frac{1}{kT} \text{ et } \lambda_3 = -\frac{\mu}{kT}$$

L'expression de la probabilité d'occurence du micro-état numéro i s'écrit au final :

$$\boxed{\mathcal{P}_i = \frac{1}{Z} e^{-\frac{E_i - \mu N_i}{kT}}} \tag{3.12}$$

où

$$Z = \sum_i e^{-\frac{E_i - \mu N_i}{kT}}$$

Application à la statistique de Fermi-Dirac

La statistique de Fermi-Dirac est la distribution de probabilité de chaque micro-état pour les systèmes de *fermions*. Toutes les particules sont classables[17] en deux familles :

[16] Je ne suis vraiment pas sûr que cette explication soit vraiment convaincante, mais il est difficile de faire plus convaincant en 5 lignes...

[17] Jusqu'à nouvel ordre...

– les fermions sont des particules de spin **demi-entier** : électrons, neutrons, protons, neutrinos, etc... ; elles satisfont au principe d'exclusion de Pauli ;
– les bosons (du nom du physicien Bose) sont des particules de spin entier ou nul : photons, mésons (particules intervenant dans les réactions nucléaire), vraisemblablement le graviton (quantum du champ de gravitation), etc...

Les comportements statistiques de ces deux classes de particules sont fondamentalement différents, du fait que les premières satisfont au principe d'exclusion : deux électrons ne peuvent se trouver dans le même état quantique ; en revanche, le nombre de photons se trouvant dans le même état (*i.e.* même vecteur **k** , même énergie et même état de polarisation) n'est pas limité. Du coup, le décompte des micro-états est donne des résultats complètement différents, et de là, la valeur de l'entropie et les probabilités de chaque micro-états sont très différentes d'une famille à l'autre.

Intéressons-nous à présent à un systèmes d'électrons dans un cristal. La statistique de Gibbs nous dit que la probabilité d'un micro-état numéro i est (equ. 3.12) :

$$\mathcal{P}_i = \frac{1}{Z} e^{-\frac{E_i - \mu N_i}{kT}} \tag{3.13}$$

Première chose à faire : déterminer la fonction de partition Z. On a génériquement :

$$Z = \sum_i e^{-\frac{E_i - \mu N_i}{kT}}$$

où la somme porte sur tous les micro-états du système. Ce qui nous intéresse au final, c'est la probabilité de trouver **un** électron particulier sur tel ou tel niveau d'énergie d'une bande d'énergie donnée, et avec un vecteur k donné. On va donc considérer, à partir de maintenant, un système constitué **d'un électron**, d'énergie E et de vecteur k fixé, en "contact" avec le nuage d'électrons du cristal (et en fait tout le reste : noyaux, impuretés, etc..., le principal étant que le "reste" soit suffisamment vaste pour jouer le rôle du réservoir de chaleur et de particule à la base de la statistique de Gibbs).

Ce système possède deux micro-états :
– un micro-état pour lequel l'électron en question est présent dans le système : l'énergie de micro-état est donc $E_i = E$, et on a $N_i = 1$;
– un second micro-état pour lequel l'électron est absent, *i.e.* le système est vide : l'énergie E_i est donc nulle, et $N_i = 0$ (ça paraît bizarre de prendre ça en compte, mais il faut se souvenir qu'on travaille à nombre de particules variable) ;
– il n'y a pas de micro-états pour lesquels on trouve plus d'un électron, en vertu du principe de Pauli.

On peut donc écrire explicitement Z, en sommant sur les deux micro-états :

$$Z = 1 + e^{-\frac{E-\mu}{kT}}$$

Examinons maintenant les probabilités. La probabilité d'occurence du premier micro-état correspond à la **probabilité que notre électron se trouve sur le niveau d'énergie E, avec le vecteur** k, et qui s'écrit explicitement :

$$\mathcal{P} = \frac{1}{Z} e^{-\frac{E-\mu}{kT}} = \frac{e^{-\frac{E-\mu}{kT}}}{1 + e^{-\frac{E-\mu}{kT}}} = \frac{1}{1 + e^{\frac{E-\mu}{kT}}}$$

Cette expression de la distribution de probabilité constitue la distribution de Fermi-Dirac. En général, on écrit le potentiel chimique (appelé **niveau de Fermi**) E_F, et on utilise la dernière forme, *i.e.* :

$$\boxed{\mathcal{P}_{FD}(E) = \frac{1}{1 + e^{\frac{E - E_F}{kT}}}} \tag{3.14}$$

Remarque 3.8 *A contrario, la probabilité que notre électron soit **absent** du niveau d'énergie E correspond à la probabilité d'occurence du second micro-état :*

$$\mathcal{P}' = \frac{1}{Z} = \frac{1}{1 + e^{-\frac{E - \mu}{kT}}}$$

*On vérifie bien que **la somme de ces deux probabilités** fait* 1*...*

Chapitre 4

Semi-conducteur intrinsèque, dopage

4.1 Semi-conducteur intrinsèque

Nous nous intéressons dans cette section aux propriétés statistiques du semi-conducteur pur. La distribution de Fermi-Dirac indique la probabilité qu'un électron particulier se trouve dans un état d'énergie E, et de vecteur k, ou de manière équivalent, le nombre moyen d'électrons $\langle N(E)\rangle$ se trouvant dans cet état : en effet, puisqu'on ne peut mettre plus d'un électron par case quantique (spécifiée par un couple (k, E)), les deux points de vue sont équivalents. A partir de cette distribution, nous calculerons les populations d'électrons appartenant à chacune des bandes, et en particulier, les bandes de valence et de conduction, puis nous en déduirons notamment la conductivité électrique du cristal.

4.1.1 Étude en fonction de la température

Tout d'abord, voyons quelle est l'influence de la température sur la répartition des électrons dans les bandes d'énergie. La figure 4.1 indique l'allure de cette distribution à 0°K, et à température non-nulle. Rappelons que la distribution de Fermi-Dirac s'écrit :

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

Quelque soit la température, la probabilité de présence est de 50% au niveau de Fermi E_F : le niveau de Fermi détermine le niveau d'énergie (s'il existe *i.e.* ne se trouve pas dans une bande interdite) qui est rempli la moitié du temps. C'est en quelque sorte la **jauge de remplissage** des niveaux d'énergie.

A $T = 0^{\circ}$K, la distribution est un échelon : la répartition des électrons sur les niveaux se fait en commençant par les niveaux les plus bas, jusqu'à épuisement du stock d'électrons : le dernier niveau occupé est alors le niveau de Fermi. A cette température, il n'y a pas d'agitation thermique, et le système est *figé* dans cet état d'équilibre.

Lorsque la température augmente, la distribution "s'étale" : les niveaux supérieurs au niveau de Fermi commencent à se peupler, et ceux inférieurs à E_F se dépeuplent. Plus la température est élevée, et plus l'étalement de la distribution conduit à un peuplement de niveaux éloignés (vers le haut) du niveau de Fermi. Il est important de comprendre qu'à cette température, l'équilibre est **dynamique** : l'agitation thermique fait passez sans cesse des électrons des niveaux inférieurs aux niveaux supérieurs, et réciproquement ; la distribution n'est qu'une loi de probabilité, et sur chaque niveau d'énergie E il existe des fluctuations de populations autour de la valeur moyenne $\langle N(E)\rangle$.

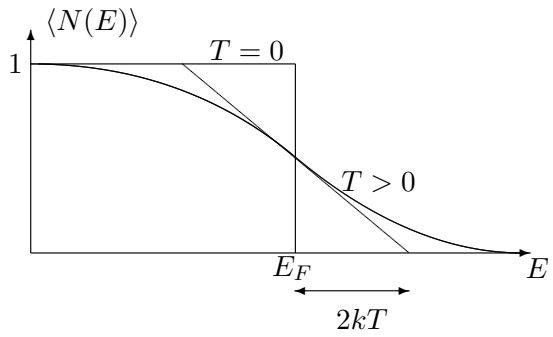


FIG. 4.1 – Distribution de Fermi-Dirac pour deux températures du cristal.

Remarque 4.1 *Ces transitions entre niveaux, qui constituent la "signature" de l'agitation thermique, se manifestent sous deux formes : elles peuvent être radiative (émission/absorption de photon), et sont à l'origine du rayonnement thermique, rayonnement d'énergie croissant avec la température ; elles peuvent aussi résulter de chocs (cas d'un transition montante), ou fournir de l'énergie au réseau cristallin sous forme de vibrations (transition descendante) ; dans tous les cas, l'énergie ne fait que se répartir différemment sur les multiples degrés de libertés du cristal et du milieu extérieur, mais conserve une "forme" désordonnée.*

Pour avoir un bonne ordre de grandeur de l'efficacité de l'agitation thermique, on peut faire un développement limité de la distribution autour de E_F :

$$f(E) \sim \frac{1}{2} - \frac{E}{4kT}$$

En confondant la distribution avec sa tangente autour du niveau de Fermi (cf. figure 4.1), on peut considérer que "l'extension" de la zone d'agitation thermique est de $4kT$: ainsi, à plus de $2kT$ du niveau de Fermi, on a très peu de chance de trouver un électron (vers le haut), ou bien toutes les chances d'en trouver un (vers le bas).

4.1.2 Semi-conducteur vs. métal

La distinction entre métal et semi-conducteur provient essentiellement de la position du niveau de Fermi dans le diagramme des bandes d'énergie. Dans le cas d'un semi-conducteur (fig. 4.2), le niveau de Fermi se trouve **dans une bande interdite**. A température nulle, la bande de valence est entièrement remplie (*i.e.* chaque niveau est occupé à 100% de chances) : elle est donc **vide** de trous ; de même, la bande de conduction est vide (d'électrons !). De fait, aucune conduction n'est possible, puisqu'il n'y a pas de porteurs de charge (que ce soit des électrons ou des trous ne change rien à l'affaire...).

Un semi-conducteur est donc isolant à température nulle. Il ne deviendrait réellement conducteur qu'à une température suffisante pour que l'agitation thermique permette à la bande de conduction de se peupler largement en électron : en gros, lorsque $4kT \sim E_g$, où E_g est le **gap** ou largeur de la bande interdite. Pour le Silicium, dont le gap vaut $1,12eV$, la température "seuil" serait :

$$T = \frac{1,12 \times 1,6.10^{-19}}{41,38.10^{-23}} \sim 3250\,K$$

A cette température, un cristal de Silicium est inutilisable... puisqu'il est liquide ! A température ambiante, l'agitation thermique est en revanche inférieure d'un ordre de grandeur à la largeur de la bande interdite : dans ces conditions, un cristal de Silicium est semi-conducteur.

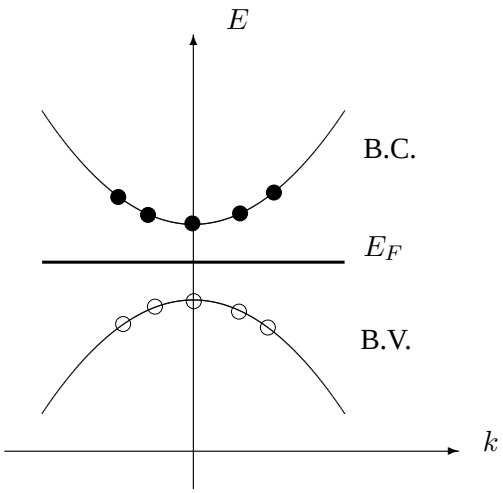


FIG. 4.2 – Dans un semi-conducteur, le niveau de Fermi se trouve dans une bande interdite. A 0˚K, la bande de valence est complètement remplie (pas de trou) et la bande de conduction est vide (pas d'électrons) : dans ce cas, le courant total est nul, qu'un champ électrique soit appliqué ou pas, et la conductivité est nulle. A température ambiante, la B.C. se remplit partiellement, la B.V. se vide partiellement (donc se peuple en trous), et un courant peut apparaître.

On voit là toute l'importance accordée aux extréma de bandes (cf. Dynamique de l'électron et la notion de masse effective) : de par la forme de la distribution de Fermi-Dirac, les électrons ou les trous ne peuvent apparaître qu'à proximité du niveau de Fermi. Or, dans un semi-conducteur, le fait que le niveau de Fermi se trouve précisément entre deux bandes permises, fait jouer un rôle particulièrement important au minimum de la bande de conduction et au maximum de la bande de valence.

Notez que le cas d'un **isolant** à température ambiante correspond simplement au cas d'un matériau à gap **très élevé** et ayant, comme un semi-conducteur, son niveau de Fermi situé dans la bande interdite[1].

Dans un **métal** , le niveau de Fermi est situé à l'intérieur d'une bande permise (fig. 4.3), qui constitue la **bande de conduction** . Ainsi, quelle que soit la température, la bande de conduction est toujours partiellement remplie, et un courant peut apparaître dés qu'un champ électrique est appliqué. Selon que ce remplissage est important ou non, *i.e.* selon la position du niveau de Fermi à l'intérieur de la bande, on a bien entendu une densité d'électrons plus ou moins élevée, et partant, une conductivité elle aussi plus ou moins élevée.

4.1.3 Populations de porteurs de charge dans les bandes

Connaître la quantité de porteurs de charges, électrons ou trous, dans un semi-conducteur, est fondamental pour la compréhension du fonctionnement des dispositifs électroniques. D'une part, de cette quantité dépend dans une large part la conductivité électrique[2]. D'autre part, nous verrons que le dopage modifie profondément l'équilibre entre populations de trous et d'électrons : dans un semi-conducteur dopé "N" par exemple, les électrons sont majoritaires. De ce déséquilibre résulte, lorsqu'on met en contact deux semi-conducteurs différem-

[1]Un problème intéressant est de savoir qu'est-ce qui détermine la position du niveau de Fermi, pour un matériau donné. Il y a beaucoup de paramètres en jeu, dont la géométrie du cristal, le recouvrement éventuel de certaines bandes, le numéro atomique et bien sûr le dopage P ou N...

[2]Les défauts cristallins et la température interviennent également.

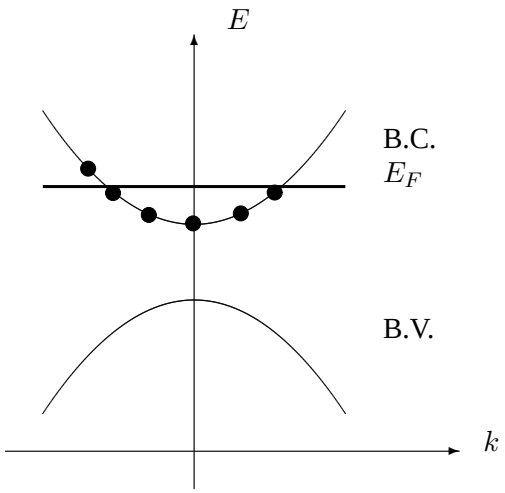


FIG. 4.3 – Dans un métal, le niveau de Fermi se trouve dans une bande permise. A 0˚K, comme à température ambiante, la B.C. reste remplie partiellement et un courant peut apparaître ; si le remplissage est faible (niveau proche du minimum), on a affaire à un semi-métal. Dans un métal, la bande de valence n'intervient pas dans la dynamique électronique : il n'y a pas de trous mobiles.

ment dopés, une diffusion des porteurs de charge d'un matériau vers l'autre, et partant des propriétés à la base du fonctionnement des jonctions PN.

En théorie, déterminer la quantité totale de porteurs dans chaque bande d'énergie, ne requiert que la connaissance du nombre moyen d'électrons dans chaque état (état caractérisé par une énergie E et un vecteur k) : il suffit ensuite de faire la somme sur tous les états de chaque bande d'énergie. En pratique, c'est plus délicat, car si l'on dispose bien du nombre moyen d'électrons dans chaque état, via la distribution de Fermi-Dirac, en revanche on ne sait pas du tout combien chaque bande contient exactement de places disponibles[3]...

Pour contourner ce problème, on exploite là encore le fait que seuls les états proches du niveaux de Fermi jouent un rôle dans la dynamique électronique. En gros, la situation est représentée sur la figure 4.4. On commence par chercher le nombre (moyen) d'électrons occupant la totalité des états de même énergie E : c'est simplement le produit de la probabilité d'occupation d'un état, par le nombre d'état qui ont la même énergie E (rappelons qu'un état est caractérisé par un point dans le diagramme des bandes : c'est donc une énergie E **et** un vecteur k). Par exemple, dans un cristal à une dimension, il y a toujours deux états (sauf pour les extréma) correspondant à la même énergie, l'un de vecteur k, l'autre de vecteur $-k$. A deux ou trois dimension, c'est plus délicat évidemment, puisqu'on a beaucoup plus de liberté sur le vecteur k (qui se promène dans un espace abstrait certes, mais de dimension identique à celle de l'espace physique...).

Le nombre d'états de même énergie est appelé **dégénérescence** ou **densité d'états**, et est noté $g(E)$. Admettons qu'on connaisse ce nombre pour toutes les énergies : il suffirait, pour connaître la population d'électrons dans la bande de conduction, de calculer :

$$n = \int_{E_{C,min}}^{E_{C,max}} g(E) f(E) dE$$

qui représente la somme, sur toutes les énergies de la bande de conduction, du nombre moyen

[3]Notre modèle simplifié de puits de potentiels couplés, qui nous a permis d'obtenir la forme des bandes d'énergie, nous permettrait bien de calculer ce nombre ; le problème est que, vu son manque flagrant de réalisme, le nombre en question risque d'être bien loin des résultats expérimentaux !

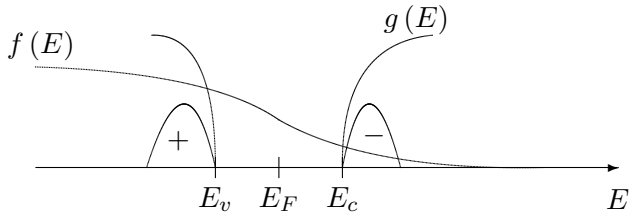


FIG. 4.4 – Pour obtenir le nombre moyen d'électrons sur un niveau, tous vecteurs k confondus, on multiplie la probabilité de présence dans un état, $f(E)$, par le nombre d'états qui ont la même énergie E, nombre appelé **densité d'état**, et noté $g(E)$.

d'électrons (habituellement, $E_{C,min}$ est notée simplement E_C, cette notation est exceptionnelle ici). Problème :

– on ne sait pas (en tout cas, pas pour un "vrai" cristal de Silicium) où s'arrête la bande conduction ;

– comme on va le voir, on ne sait calculer $g(E)$ qu'à proximité des extréma de bandes ;

En fait, l'allure de la distribution de Fermi-Dirac nous tire d'affaire : comme elle décroît très vite, on peut remplacer l'intégrale précédente par :

$$n = \int_{E_C}^{\infty} g(E) f(E) dE \tag{4.1}$$

En effet, l'erreur porte sur les quelques électrons qui osent s'aventurer dans le haut de la bande de conduction... Pour la même raison, on peut utiliser pour $g(E)$ l'expression (calculable, elle) au voisinage de l'extrémum : ce sont encore les mêmes électrons qui engendreront une petite erreur. Bien, nous y voilà : il reste à calculer $g(E)$ au voisinage d'un extrémum.

Densité d'états au voisinage d'un extrémum de bande

Pour déterminer $g(E)$, rappelons-nous que l'énergie d'un électron s'écrit, au voisinage du minimum de la bande de conduction,

$$E = \frac{\hbar^2 (\mathbf{k} - \mathbf{k}_0)^2}{2m^*} + E_C$$

où m^* est la masse effective, $\mathbf{k}_0$ le vecteur correspondant au minimum (cf. figure 4.5), E_c l'énergie du minimum de la bande de conduction (attention à ne pas confondre $E_C =$ *minimum de la bande de conduction* et $E_c =$ *énergie cinétique*).

Le problème consiste donc à compter combien on peut trouver de vecteurs $\mathbf{k}$ correspondant à l'énergie E. Par exemple, à une dimension, on a au moins deux vecteurs k pour chaque niveau d'énergie, les extréma de bande exceptés (la figure 4.5 montre même que certains niveaux admettent quatre vecteurs).

Pour un cristal bidimensionnel, la situation est illustrée figure 4.6 : on a représenté les niveaux d'énergie près d'un minimum, sous forme de courbes de niveaux, qu'on appelle également **courbes d'isoénergie**.

Remarque 4.2 *Ces courbes d'isoénergie ne sont pas nécessairement des cercles (resp. des sphères en tridimensionnel). En effet, si la maille du cristal n'est pas un cube, mais une forme géométrique moins symétrique (rhomboèdre, parallélépipède rectangle, etc.), l'anisotropie qui en résulte se "voit" aussi dans le diagramme des bandes ; en d'autres termes, l'interaction électron-cristal n'a pas la même intensité selon les deux (resp. trois) directions du*

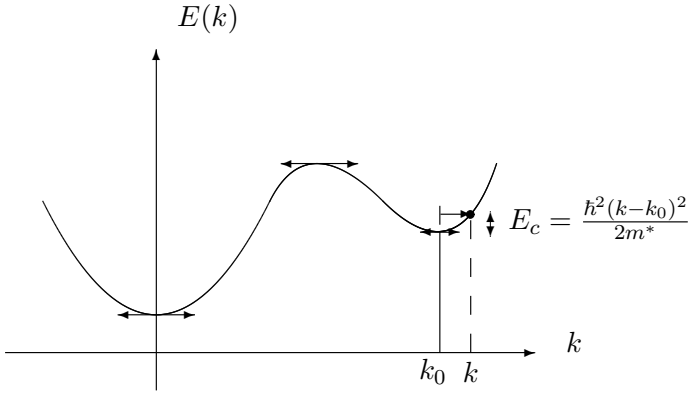


FIG. 4.5 – Énergie cinétique E_c d'un électron, à proximité d'un extrémum de bande, dans le cas d'un cristal unidimensionnel.

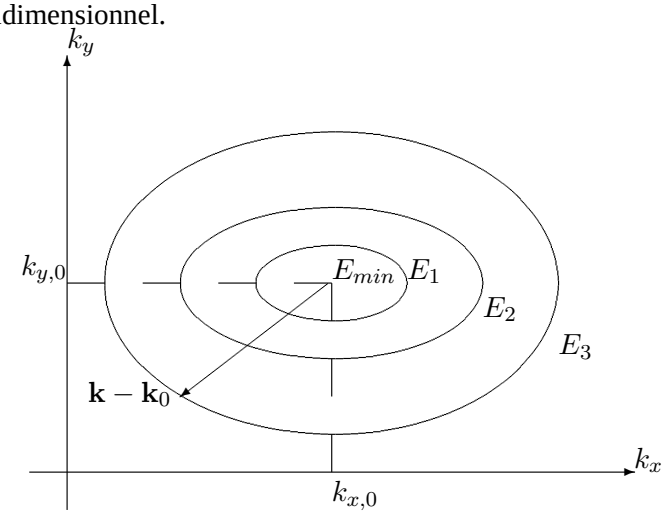


FIG. 4.6 – Représentation par courbes de niveau de l'énergie cinétique d'un électron, à proximité d'un minimum de bande, dans le cas d'un cristal **bidimensionnel** . Ces courbes sont des ellipses, car les masses effectives ne sont pas nécessairement les mêmes selon les deux directions du cristal.

cristal. Du coup, l'expression de l'énergie au voisinage d'un extrémum de bande s'écrit :

$$E = \frac{\hbar^2(k_x - k_{x,0})^2}{2m_x^*} + \frac{\hbar^2(k_y - k_{y,0})^2}{2m_y^*} + \frac{\hbar^2(k_z - k_{z,0})^2}{2m_z^*} + E_C$$

car les masses effectives ne sont pas les mêmes selon chaque direction. A E fixé, l'équation de la courbe (resp. la surface) générée par le vecteur **k** *est une ellipse (resp. un ellipsoïde).*

Le problème est de compter le nombre de vecteurs **k** associés à la même énergie (par exemple E_1 sur la figure), *i.e.* le nombre de vecteurs situé sur l'ellipse, sachant qu'en vertu des conditions aux limites périodiques (cf. Chapitre 2), chaque vecteur est séparé du précédent de $2\pi/L$, où L est la dimension du cristal selon la direction concernée (formulation équivalente : la maille de l'espace réciproque est un rectangle de côtés $2\pi/L_x$ et $2\pi/L_y$).

Manifestement, le décompte est assez délicat, mais comme les niveaux d'énergie sont très rapprochés d'une part, et que le nombre d'états par niveaux d'énergie est très élevé (sauf cas particulier) d'autre part, on va considérer que E varie **continûment**, ce qui simplifie pas mal les calculs, les sommes discrètes devenant des intégrales. Du coup, le problème se

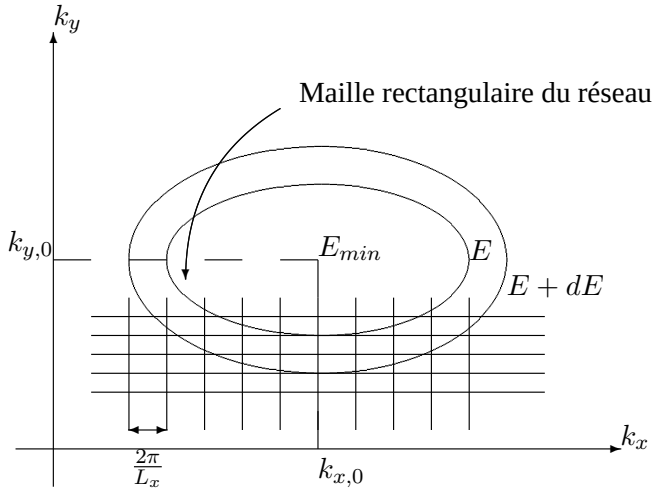


FIG. 4.7 – La densité d'états $g(E)$ représente le nombre de mailles (ici, des rectangles) inclues dans l'anneau elliptique d'épaisseur dE

formule ainsi : combien peut-on trouver de vecteurs **k** associés à une énergie comprise entre E et $E + dE$? C'est ce nombre qui représente **la densité d'états**. A deux dimensions, il correspond au nombre de rectangles (chaque rectangle étant associé à un vecteur **k**) inclus dans l'anneau elliptique d'épaisseur dE (figure 4.7).

De l'expression de l'énergie au voisinage d'un extrémum de bande (cf. remarque 4.2),

$$E = \frac{\hbar^2(k_x - k_{x,0})^2}{2m_x^*} + \frac{\hbar^2(k_y - k_{y,0})^2}{2m_y^*} + E_C$$

on déduit que la courbe d'isoénergie est une ellipse de demi-axes $\sqrt{2m_x(E - E_C)}/\hbar$ et $\sqrt{2m_y(E - E_C)}/\hbar$. L'aire de cette ellipse est

$$\mathcal{A} = \pi \times \frac{\sqrt{2m_x(E - E_C)}}{\hbar} \times \frac{\sqrt{2m_y(E - E_C)}}{\hbar} = \frac{2\pi}{\hbar^2}\sqrt{m_x m_y}(E - E_C)$$

Le nombre de rectangles (maille du réseau réciproque de surface $4\pi^2/L_xL_y$) inclus dans l'ellipse est :

$$N = \frac{\mathcal{A}}{4\pi^2/L_xL_y} = \frac{\pi\sqrt{m_x m_y}L_xL_y}{h^2}(E - E_C)$$

Finalement, le nombre de rectangles inclus dans l'anneau d'épaisseur dE est :

$$dN = \frac{\pi\sqrt{m_x m_y}L_xL_y}{h^2}dE$$

La densité d'état pour un cristal bidimensionnel s'écrit donc :

$$g(E) = \frac{\pi\sqrt{m_x m_y}L_xL_y}{h^2}$$

et représente **le nombre d'état disponibles** pour les électrons, sur le niveau d'énergie E, à dE près.

Passons à présent au cas d'un cristal tridimensionnel. Les ellipses d'isoénergie deviennent des ellipsoïdes, et la densité d'état correspond au nombre de mailles du réseau réciproques

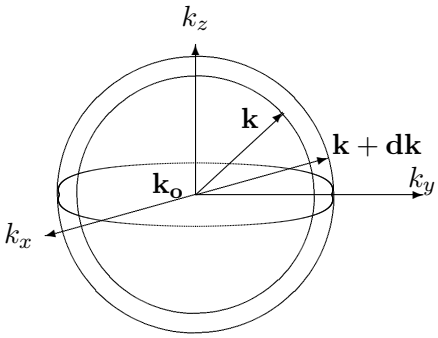


FIG. 4.8 – Combien de vecteurs k de l'espace réciproque sont associés à la même énergie E, à dE près ? Réponse : le volume de la calotte sphérique, divisé par le volume occupé par une maille du réseau réciproque.

(cubes, parallélépipèdes, ou forme plus complexe) inclus dans la calotte d'ellipsoïde d'épaisseur dE. La figure 4.8 illustre le principe du calcul pour un cristal cubique simple, pour lequel les surfaces d'isoénergie sont des sphères (les trois masses effectives étant égales).

Reprenons le calcul précédent, en généralisant à trois dimensions. L'équation donnant l'énergie,

$$E = \frac{\hbar^2(k_x - k_{x,0})^2}{2m_x^*} + \frac{\hbar^2(k_y - k_{y,0})^2}{2m_y^*} + \frac{\hbar^2(k_z - k_{z,0})^2}{2m_z^*} + E_C$$

indique que l'ellipsoïde a pour demi-axes, $\sqrt{2m_x(E-E_C)}/\hbar$, $\sqrt{2m_y(E-E_C)}/\hbar$ et $\sqrt{2m_z(E-E_C)}/\hbar$, donc pour volume,

$$\mathcal{V} = \frac{4}{3}\pi \frac{\sqrt{2m_x(E-E_C)}}{\hbar} \times \frac{\sqrt{2m_y(E-E_C)}}{\hbar} \times \frac{\sqrt{2m_z(E-E_C)}}{\hbar}$$

Le nombre de parallélépipèdes (maille du réseau réciproque de côtés $2\pi/L_x$, $2\pi/L_y$ et $2\pi/L_z$) inclus dans l'ellipsoïde est :

$$N = \frac{\mathcal{V}}{8\pi^3/L_xL_yL_z} = \frac{8\sqrt{2}\pi\sqrt{m_xm_ym_z}L_xL_yL_z}{3h^3}(E-E_C)^{3/2}$$

Finalement, le nombre de parallélépipèdes inclus dans la calotte ellipsoïdale d'épaisseur dE est :

$$dN = \frac{4\sqrt{2}\pi\sqrt{m_xm_ym_z}L_xL_yL_z}{h^3}\sqrt{E-E_C}dE$$

Pour un cristal tridimensionnel, la **densité d'état** au voisinage du minimum de la bande de conduction s'écrit donc :

$$\boxed{g_n(E) = \frac{2\pi V}{h^3}\sqrt{8m_xm_ym_z}\sqrt{E-E_C}} \tag{4.2}$$

où $V = L_xL_yL_z$ est le **volume du cristal** (le "n" de g_n est une convention pour indiquer qu'il s'agit de la bande de conduction). Naturellement, cette expression est aussi valable au voisinage du maximum de la bande de valence, à condition de changer $E - E_C$ en $E_V - E$ (où E_V est le maximum de la bande de valence), et d'utiliser pour les masses effectives, celles des trous :

$$g_p(E) = \frac{2\pi V}{h^3}\sqrt{8m_xm_ym_z}\sqrt{E_V-E}$$

On retiendra que la densité d'état croît comme la racine carré de l'énergie de l'énergie cinétique $E_c = E - E_C$, ce qui justifie l'allure de la courbe $g(E)$ sur la figure 4.4.

Populations de porteurs de charges

A présent, nous disposons de tous les résultats nécessaires au calcul des populations d'électrons dans la bande de conduction, et de trous dans la bande de valence. Rappelons-nous que la population totale d'électrons dans la bande de conduction est donnée par l'équation 4.1,

$$n = \int_{E_C}^{\infty} g_n(E) f(E) dE$$

En fait, nous avons ignoré jusqu'à présent le **spin** de l'électron : en en tenant compte, on peut placer dans chaque "case" quantique étiquetée par **k** et E, non pas **un**, mais **deux** électrons, à condition que leur spin soient **antiparallèles**. De fait, l'expression de la population d'électrons est en réalité :

$$n = \int_{E_C}^{\infty} 2 g_n(E) f(E) dE$$

En utilisant les expressions de $g_n(E)$ et de la distribution de Fermi-Dirac $f(E)$, on obtient un dinosaure incalculable[4] :

$$n = \int_{E_C}^{\infty} 2 \frac{2\pi V}{h^3} \sqrt{8 m_x m_y m_z} \sqrt{E - E_C} \frac{1}{1 + e^{(E-E_F)/kT}} dE$$

Si l'on a des velléités de calculer **ça**, il faut (au moins...) faire l'approximation suivante : dans le cas d'un semi-conducteur, le niveau de Fermi E_F est dans la bande interdite, et l'expression de la distribution de Fermi-Dirac admet, lorsque E est dans la bande de conduction, et que la température n'est pas trop élevée, l'approximation suivante :

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

appelée *approximation de Boltzmann*, parce que l'expression est identique à la distribution de Boltzmann[5]. L'expression de n, vous le noterez, en est incroyablement simplifiée :

$$n \sim \int_{E_C}^{\infty} 2 \frac{2\pi V}{h^3} \sqrt{8 m_x m_y m_z} \sqrt{E - E_C}\, e^{(E_F-E)/kT} dE$$

Moyennant quelques changements de variables (sincèrement) sans difficulté, on peut ramener cette expression à une fonction *Gamma*[6],

$$\Gamma\left(\frac{3}{2}\right) = \int_0^{\infty} x^{1/2} e^{-x} dx = \frac{\sqrt{\pi}}{2}$$

et obtenir pour n l'expression suivante :

$$n \sim \frac{2V}{h^3} \sqrt{m_x m_y m_z} (2\pi kT)^{3/2} e^{(E_F-E_C)/kT}$$

En général, on s'intéresse à la densité **volumique** d'électrons, et V disparaît. On pose également :

$$N_c = \frac{2}{h^3} \sqrt{m_{x,n} m_{y,n} m_{z,n}} (2\pi kT)^{3/2} \tag{4.3}$$

[4] En fait, si, mais c'est long : on utilise des intégrales de Sommerfeld et une palanquée de changements de variables.

[5] Ceci étant, comme des tonnes d'approximations portent le nom de Boltzmann, ça ne nous avance pas énormément...

[6] Fonction généralisant la factorielle sur $\mathbb{R}$, pour ceux qui l'ignorent...

que l'on appelle **densité équivalente d'états** dans la bande de conduction. Finalement, la densité volumique d'électrons dans la bande de conduction s'écrit :

$$n \sim N_c e^{(E_F - E_C)/kT} \tag{4.4}$$

et l'on se rappellera que cette expression n'est valable que pour des températures pas trop élevées et lorsque E_F est situé **dans** la bande interdite (ce sont les conditions de validité de l'approximation de Boltzmann et l'on voit d'ailleurs que, si $E_F > E_C$, il y a plus d'électrons que de places disponibles !).

Remarque 4.3 *On peut interpréter cette expression en disant que, en ce qui concerne le calcul des populations, l'ensemble des niveaux d'énergie de la bande de conduction peut être remplacé par un seul niveau, E_C, dégénéré N_c fois (*i.e. *comportant N_c états de même énergie E_C), et dont le remplissage en fonction de la température est donné par la distribution de Fermi-Dirac à l'approximation des faibles températures. C'est pour cette raison que N_c porte le nom de **densité équivalente d'états**.*

Le point fondamental est que la population d'électrons dans la bande de conduction dépend directement de la position du niveau de Fermi :

> La population d'électrons dans la bande de conduction augmente avec le niveau de Fermi.

Ceci confirme l'interprétation "qualitative" que nous avions donnée du niveau de Fermi : c'est un niveau "jauge" qui représente le remplissage des niveaux d'énergie par les électrons. En particulier, nous verrons que le fait de doper au Phosphore un semi-conducteur comme le Silicium ou le Germanium, fait singulièrement monter le niveau de Fermi vers le niveau E_C, grâce à l'apport (par l'atome "dopant") d'une grande quantité d'électrons.

Considérons à présent le cas de la bande de valence. La population d'intérêt est à présent la population de trous, dont l'expression est donnée par :

$$p = \int_{-\infty}^{E_V} 2g_p(E)(1 - f(E))dE$$

Ce qui est important, c'est la présence du terme $1 - f(E)$ qui donne la probabilité de **non-occupation** par un électron, donc la probabilité d'occupation par un trou. Ensuite, le calcul est identique, et donne

$$p \sim N_v e^{(E_V - E_F)/kT} \tag{4.5}$$

où

$$N_v = \frac{2}{h^3}\sqrt{m_{x,p}m_{y,p}m_{z,p}}(2\pi kT)^{3/2} \tag{4.6}$$

est la **densité équivalente d'états** dans la bande de valence (attention : les masses effectives sont celles des trous). Ici encore, le niveau de Fermi représente la jauge de remplissage : un niveau de Fermi proche du bas de la bande interdite correspond à un faible remplissage **en électrons** du haut de la bande de valence, donc à un bon remplissage **en trous** : un semi-conducteur de type P, caractérisé par une population majoritaire de trous (grâce à l'apport de trous par un atome dopant), verra son niveau de Fermi se rapprocher du bas de la bande interdite. Du coup, en mettant en contact deux semi-conducteurs de dopage différent, nous avons deux matériaux dont les niveaux de Fermi diffèrent, et phénomènes de diffusion apparaissent : cette diffusion de porteurs à travers l'interface séparant les deux matériaux est à la base du fonctionnement de la jonction PN.

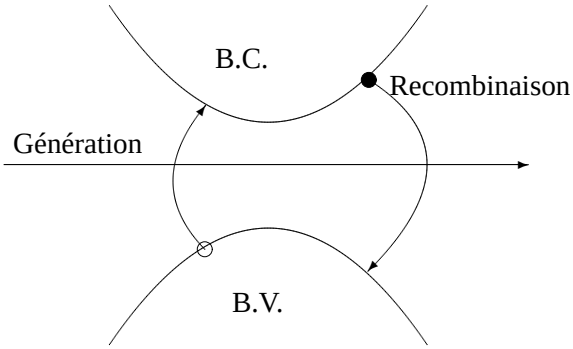


FIG. 4.9 – La loi d'action de masse traduit l'équilibre thermodynamique entre processus de génération et de recombinaison.

Loi d'action de masse

En faisant le produit de n et de p, on obtient expression suivante :

$$\boxed{np = N_c N_v e^{-E_g/kT}} \tag{4.7}$$

où $E_g = E_C - E_V$ est **le gap d'énergie** du semi-conducteur, *i.e.* l'épaisseur de la bande interdite. Cette expression ne dépend que des masses effectives et de la température : pour un cristal donné, à température fixée, le produit np est donc constant. Cette équation constitue la **loi d'action de masse** : il existe l'analogue en chimie, par exemple le produit ionique de l'eau, $[H_3O^+][OH^-] = K_e$.

Cette relation n'est que la conséquence de la forme de la distribution de Fermi-Dirac (plus exactement, de l'approximation de Boltzmann) : en toute rigueur, elle n'est valable qu'à **l'équilibre thermodynamique** du cristal, *i.e.* lorsque températures, pression, et niveau de Fermi sont stationnaires. Lorqu'on écarte le système de l'équilibre (par exemple, en l'éclairant), la loi d'action de masse n'est plus respectée, mais la tendance du système est de revenir à l'équilibre dés que la cause de déséquilibre cesse : ce mécanisme de retour à l'équilibre est par exemple le mécanisme de **génération-recombinaison** de **paires électron/trous** illustré figure 4.9.

Ainsi, lorsqu'on éclaire un cristal, on crée des **paires électron/trous** : idéalement, pour chaque photon incident absorbé, un électron est envoyé d'un niveau de la B.V. à un niveau de la B.C. et de fait, le produit np augmente. Le retour à l'équilibre se fait donc par recombinaison (retour d'un électron de la B.C. à la B.V.), qui fait simultanément disparaître un électron et un trou.

Les fluctuations de populations autour des valeurs n et p, dues à l'agitation thermique, font que la loi d'action de masse n'est jamais parfaitement respectée : le produit np fluctue lui aussi, du fait que des électrons passe sans cesse d'une bande à l'autre. Il y a sans cesse **génération-recombinaison** de paires électron-trou par l'agitation thermique.

Application au semi-conducteur intrinsèque

Dans un semi-conducteur intrinsèque, *i.e.* non-dopé, électrons de la B.C. et trous de la B.V. sont en quantité identique : chaque électron de la B.C. a donné naissance, lorsqu'il est apparu dans la B.C., à un trou. On peut donc écrire :

$$n = p$$

On note n_i cette densité de porteur, et d'après la loi d'action de masse, on a :

$$n = p = n_i = \sqrt{N_c N_v} e^{-E_g/2kT} = \frac{2}{h^3}(2\pi kT)^{3/2} \underbrace{(m_{x,n} m_{y,n} m_{z,n})^{1/4}}_{\text{électrons}} \underbrace{(m_{x,p} m_{y,p} m_{z,p})^{1/4}}_{\text{trous}} e^{-E_g/2kT}$$

L'ordre de grandeur est de 10^{10} porteurs par cm^3 pour le Silicium à température ambiante, ce qui est très faible, comparé aux métaux (10^{22} électrons par cm^3 pour le Cuivre), et provient essentiellement du positionnement du niveau de Fermi à l'intérieur d'une bande interdite.

Le point important est que la quantité d'électrons, comme de trous, est dans un semi-conducteur proportionnelle à la température (l'agitation thermique permet aux électrons de passer plus facilement de la B.V. et la B.C.), et inversement proportionnelle au gap (le gap constituant la barrière énergétique à franchir par un électron pour passer de la B.V. à la B.C.). Comment varie la conductivité lorsqu'on augmente la température d'un semi-conducteur ? C'est le résultat d'une compétition entre deux effets. D'une part, la population d'électrons augmente et favorise la conductivité ; d'autre part, l'agitation thermique des noyaux augmente aussi, créant des écarts importants à la périodicité parfaite : les chocs des ondes associées aux électrons sur les noyaux du cristal se multiplient, et diminue ce qu'on appelle la mobilité. Au final, c'est (à température ambiante du moins) le premier effet qui l'emporte : la conductivité d'un semi-conducteur augmente avec la température. Il en va différemment pour un métal : si les chocs augmentent avec la température, en revanche la densité d'électrons reste quasiment constante, puisque le niveau de Fermi est situé **dans** la bande de conduction (cf. figure 4.3), et du coup la conductivité diminue généralement lorsque la température augmente. C'est cette différence de comportement qui est d'ailleurs exploitée pour réaliser des capteurs de températures sous forme de résistances CTN ou CTP (*i.e.* à coefficient de température respectivement négatif et positif).

Comment est positionné exactement le niveau de Fermi dans un semi-conducteur intrinsèque ? De $n = p$, on déduit :

$$N_c e^{(E_F - E_C)/kT} = N_v e^{(E_V - E_F)/kT}$$

d'où le niveau de Fermi,

$$E_{F,i} = \frac{E_C + E_V}{2} + \frac{kT}{2} \ln \frac{N_v}{N_c} = \frac{E_C + E_V}{2} + \frac{kT}{4} \ln \frac{m_{x,p} m_{y,p} m_{z,p}}{m_{x,n} m_{y,n} m_{z,n}}$$

Les masses effectives des trous et des électrons étant du même ordre de grandeur en général, le niveau de Fermi est situé approximativement au milieu de la bande interdite, pour un semi-conducteur intrinsèque.

4.2 Semi-conducteurs dopés

Un semi-conducteur intrinsèque possède une faible conductivité, due principalement à la faible population de porteurs : on peut fortement accroître cette conductivité en ajoutant, en très faible proportion, un atome d'impureté appelé **dopant**. Selon le type d'atome dopant, on augmentera la conductivité par électrons (semi-conducteur dopé N) ou par trous (semi-conducteur dopé P). Le semi-conducteur ainsi dopé s'appelle **semi-conducteur extrinsèque** : à température ambiante, presque toutes ses propriétés électriques ne dépendent que de l'atome dopant et de sa concentration. Pour comprendre comment fonctionne le dopage, il faut se souvenir que le Silicium, comme le Germanium et le Carbone, forment avec leurs voisins des **liaisons covalentes**, et que leur **tétravalence** (due au fait qu'ils possèdent

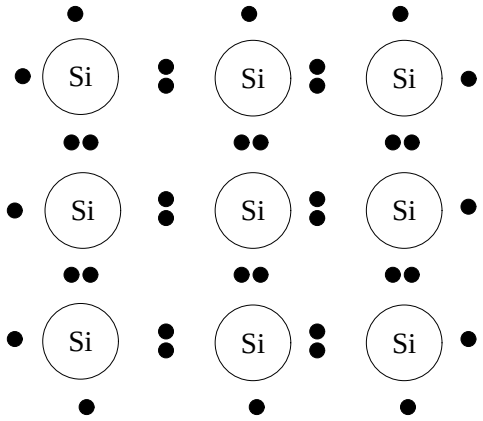


FIG. 4.10 – Dans le réseau cristallin du Silicium (ici, vu en projection sur un plan), chaque atome forme une liaison de valence avec l'un de ses quatre voisins.

4 électrons sur la dernière couche) leur permet de former 4 liaisons : on a vu dans la partie Cristallographie que le cristal ainsi formé possède une structure de type **diamant**, correspondant à 2 réseau *cubique faces-centrées* décalés l'un de l'autre d'un quart de diagonale du cube. La structure géométrique des liaisons covalentes formées par chaque atome est tétraédrique. En projection bidimensionnelle, l'allure du réseau cristallin est illustrée figure 4.10.

Le dopage consiste alors à introduire, en petite quantité, un atome d'impureté, Phosphore, Bore, Antimoine, etc., qui vient — si tout se passe bien — prendre la place occupée par un atome de Silicium (par substitution) : l'atome d'impureté "prend" alors les liaisons de l'atome de Silicium (qui est éjecté). On voit d'ores et déjà que l'atome d'impureté doit posséder une taille semblable à celle du Silicium pour que la substitution fonctionne correctement, et c'est là une des principales difficultés dans le choix d'un dopant. Malgré cela, certains atomes d'impuretés viennent parfois occuper une **lacune interstitielle** et créent une rupture brutale de périodicité, dont l'effet principal est d'accroître les probabilités de chocs, et de réduire la conductivité.

4.2.1 Semi-conducteurs de type N

Pour réaliser un dopage de type N, on ajoute au Silicium (mais il en va de même avec le Germanium et les composés III-V ou II-IV comme l'Arseniure de Gallium) un atome de la colonne V, par exemple le Phosphore, qui est **pentavalent** (figure 4.11). Le cinquième électron ne peut pas former de liaison de valence, et se trouve de fait dans un état faiblement lié à l'atome de Phosphore : sa fonction d'onde est très étalée (en générale de l'ordre de plusieurs distances interatomiques), et il faut voir le système davantage comme un couple *électron-phosphore ionisé* vu la distance moyenne entre l'électron et l'ion P^+. Une très faible énergie est nécessaire pour détacher cet électron : généralement, l'énergie d'agitation thermique est suffisante, et l'électron passe très vite dans la bande de conduction, *i.e.* devient un électron susceptible de conduire le courant électrique au même titre que les autres électrons de conduction fournis par les atomes de Silicium : pour cette raison, l'atome pentavalent est appelé **atome donneur** car il apporte des électrons au réseau. On voit qu'avec cette technique, on peut donc assez précisément contrôler la conductivité d'un semi-conducteur.

Voyons comment les électrons apportés par l'impureté pentavalente se positionnent dans le diagramme des bandes d'énergie. La figure 4.12 montre que le niveau d'énergie E_D occupé par le cinquième électron est très proche de la bande de conduction : il est situé dans

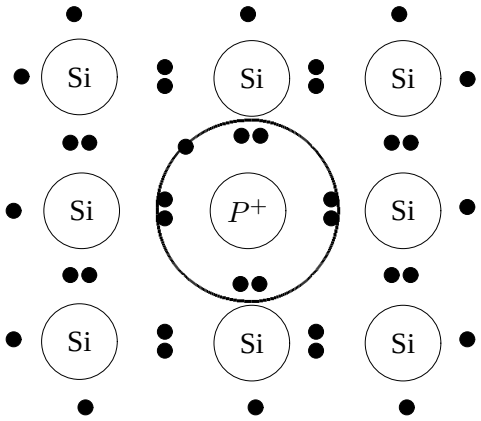


FIG. 4.11 – Principe du dopage de type N : on substitue à certains atomes de Silicium un atome d'impureté pentavalent, qui possède alors un électron faiblement lié : cette électron peut facilement, du fait de l'agitation thermique, se détacher et intégrer la bande de conduction, donnant ainsi un électron de conduction supplémentaire au réseau.

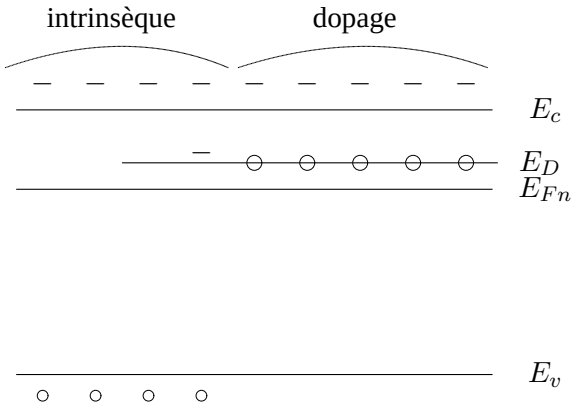


FIG. 4.12 – Avant détachement thermique, le cinquième électron du Phosphore se trouve sur le niveau **donneur** E_D, très proche de la bande de conduction : le gap pour atteindre la B.C. est faible, et peut être franchi à température ambiante. Une fois détachés, les électrons issus de ce niveau forme, avec les électrons de la B.C. déjà présent, l'ensemble des électrons de conduction du cristal.

la bande interdite (logique, puisque l'électron est malgré tout attaché au Phosphore, et ne conduit pas), mais le "gap" à franchir pour atteindre la B.C. est suffisamment faible (de l'ordre de $0,04eV$, à comparer au gap de $1,12eV$ de la bande interdite) pour que, à température ambiante (où l'énergie d'agitation thermique est de l'ordre de $kT \sim 0,025eV$), le niveau d'énergie E_D soit quasiment vide : la majorité les électrons se retrouve dans la bande de conduction (ou se trouvaient déjà quelques électrons fournis par le Silicium). Généralement, la concentration en atome donneurs est de l'ordre de $10^19\,cm^{-3}$: la quantité d'électrons fournis à la B.C. par le dopant est donc largement supérieur (de plusieurs ordres de grandeur) à la quantité d'électrons déjà présents dans la B.C. (on a vu au paragraphe précédent que $n_i \sim 10^10\,cm^{-3}$). De fait, on peut, en tout cas à température ambiante, négliger les électrons déjà présents dans la B.C. et ne prendre en compte que les électrons fournis par le dopant : voilà ce qui justifie l'appellation de semi-conducteur **extrinsèque** donnée au semi-conducteur dopé.

On vient de voir que la population en électrons de conduction d'un semi-conducteur dopé augmente singulièrement : qu'en est-il des trous ? Comme l'illustre la figure 4.12, les trous présents dans la bande de valence sont uniquement dus à la génération de porteurs par le semi-conducteur intrinsèque : en d'autres termes, le dopant n'augmente pas la concentration initiale en trous, puisque seuls des électrons sont **données** au réseau. Même, l'effet du dopant est de diminuer la concentration en trous via la loi d'action de masse : lors de l'opération de dopage, l'augmentation brutale d'électrons dans la B.C. provoque l'augmentation du produit np, et le nécessaire retour à l'équilibre se fait par **recombinaison** d'électrons, qui redescendent dans la B.V. et comblent ainsi quelques trous... Le point important à retenir est donc que, dans un semi-conducteur dopé N (N comme *négatif* pour *électrons*), la population **majoritaire** est une population d'électrons : on dit que les électrons sont les **porteurs majoritaires**.

Pour étudier les phénomènes de diffusion lors de la mise au contact de deux semi-conducteurs de dopage P et N, nous devons étudier la position du niveau de Fermi. Il nous faut avant tout une relation, analogue à la distribution de Fermi-Dirac, nous disant quelle est exactement la probabilité qu'un électron du niveau E_D passe dans la bande de conduction. Cette relation peut être démontrée à partir de l'expression de la statistique de Gibbs, comme nous l'avons fait pour la distribution de Fermi-Dirac, mais exceptionnellement je me contenterai de l'énoncer. On note :

- N_D la concentration en atomes donneurs ;
- N_D^+ la concentration en atome donneurs **ionisés** positivement par perte de l'électron faiblement lié ;

La probabilité d'ionisation d'un atome de Phosphore, *i.e.* la probabilité de non-occupation du niveau E_D, est donnée par la distribution suivante :

$$\boxed{N_D^+ = \frac{N_D}{1 + 2e^{(E_F - E_D)/kT}}} \tag{4.8}$$

En guise d'interprétation, on notera que la probabilité d'ionisation dépend des positions relatives du niveau de Fermi et du niveau donneur E_D : lorsque le niveau de Fermi monte, le remplissage du niveau E_D est de plus en plus fréquent, et les atomes de Phosphore sont donc de moins en moins souvent ionisés.

A partir de cette relation, on peut (moyennant une résolution numérique) trouver la position du niveau de Fermi. La neutralité électrique du cristal (qui, j'insiste, n'a à aucun moment été rompue puisqu'on a introduit des **atomes** de Phosphore, pas des ions...) permet d'écrire :

$$n = p + N_D^+$$

Cette relation signifie en somme que les électrons de la B.C. (en quantité totale n) ont deux provenances :

- les uns (terme p) proviennent de la B.V., et sont dus à la génération (d'origine thermique) de paires électron-trou par le semi-conducteur intrinsèque (*i.e.* ces électrons et ces trous n'ont rien à voir avec le dopage) ; leur nombre est **rigoureusement** égal au nombre de trous (chaque paire générée crée un trou **et** un électron), d'où le fait qu'ils soient associés au terme p ;
- les autres (généralement bien plus nombreux) sont dus à l'ionisation des atomes donneurs, et proviennent donc du niveau donneur E_D : ils correspondent au terme N_D^+.

Remplaçant dans l'équation de neutralité n et p par leur expression (équation 4.5 et 4.4), on obtient l'équation donnant le niveau de Fermi :

$$N_c e^{(E_F - E_C)/kT} = N_v e^{(E_V - E_F)/kT} + \frac{N_D}{1 + 2e^{(E_F - E_D)/kT}}$$

La résolution numérique est faite en T.D. (niveau de Fermi d'un semi-conducteur dopé mais on peut[7] essayer au moins qualitativement de distinguer les différents régimes de comportement du semi-conducteur dopé.

A **basse température**, l'ionisation des atomes donneurs est faible : le terme N_D^+ est proche de zéro, et le semi-conducteur se comporte comme un semi-conducteur intrinsèque, *i.e.* il est quasi-isolant. Les populations en trous et en électrons sont égales, et valent n_i (cf. loi d'action de masse).

A **température ambiante** , l'ionisation des donneurs est presque **totale** : on dit qu'il y a **saturation** de l'ionisation, et on peut écrire,

$$N_D^+ \sim N_D$$

En revanche, comme on l'a vu au début de ce chapitre, la création de paires électron-trou par le Silicium demeure assez faible (c'est bien pour cela qu'il reste quasi-isolant s'il n'est pas dopé !). En fait, la raison de cette dissymétrie est uniquement à mettre sur le compte de la différence de distance entre, d'une part E_C et E_V (paires électron-trou), et d'autre part E_C et E_D (niveau donneur) : si E_D n'était pas aussi proche de la B.C., il n'y aurait pas cette dissymétrie et le semi-conducteur resterait isolant à température ambiante. Dans cette région de température, l'apport d'électrons par l'ionisation des donneurs est donc prépondérant, et dépasse largement l'apport par génération de paires. La concentration en électron est pratiquement celle des donneurs, et on peut négliger la concentration en trous p devant celle des impuretés ionisées positivement :

$$n \sim N_D$$

Les électrons sont **majoritaires** et il y a dans le matériau :

- des électrons de conduction, majoritaires, en concentration N_D ;
- quelques trous, minoritaires ;
- des atomes donneurs ionisés, qui constituent autant de **charges positives fixes**, et sont quasiment en concentration N_D (ionisation presque totale).

Les électrons étant majoritaires, on devine que le niveau de Fermi a toutes les chances de se trouver assez près du niveau E_C : en reportant les approximations dans l'équation donnant le niveau de Fermi, on obtient :

$$N_c e^{(E_F - E_C)/kT} \sim N_D$$

[7] Comme en chimie des équilibres chimiques... le fait que la loi d'action de masse impose un **produit** de concentration, permet de faire toutes sortes d'approximations dans les **sommes** de concentrations : très souvent, l'une des deux concentrations est négligeable devant l'autre.

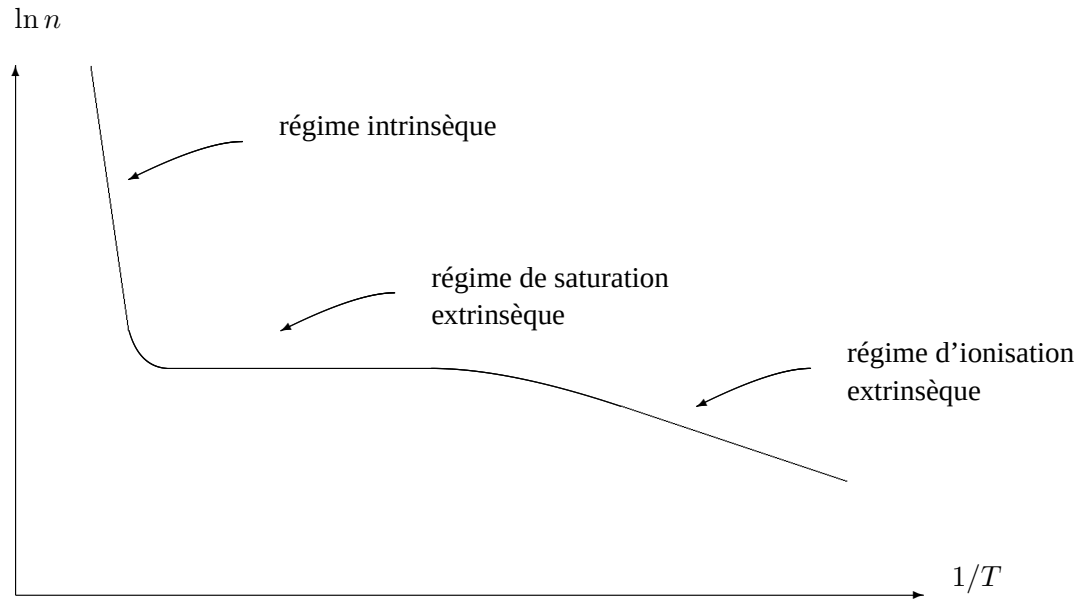


FIG. 4.13 – Évolution de la concentration en électrons de conduction en fonction de l'inverse de la température.

ce qui conduit à

$$E_F \sim E_C + kT \ln \frac{N_D}{N_C}$$

qui montre que, pour des dopages de l'ordre $10^15\,cm^{-3}$ à $10^18\,cm^{-3}$, le niveau de Fermi est légèrement en-dessous du minimum de la B.C. à température ambiante (sauf cas de *surdopage*, qui concerne les semi-conducteurs **dégénérés** et pour lesquels l'expression ne s'applique pas de toute façon).

A **haute température** , la génération de paires électron-trou par le Silicium concurrence sérieusement l'apport d'électrons par les atomes donneurs. Le comportement intrinsèque redevient prépondérant : les électrons de la bande de conduction proviennent, lorsque la température augmente, majoritairement de la génération de paires électron-trou, l'ionisation totale des atomes donneurs étant déjà totale.

La figure 4.13 montre l'évolution de la concentration en électrons de la B.C. en fonction de la température : on distingue bien les trois régimes, en particulier le régime de **saturation** qui constitue le régime utilisé effectivement dans les composants (températures entre -10˚C et 100˚C).

4.2.2 Semi-conducteurs de type P

Nous reprenons le même type de raisonnement pour les semi-conducteurs de type P. Cette fois, on introduit dans le matériau une impureté **trivalente**, comme l'illustre la figure 4.14.

Le mécanisme est en tout point semblable au cas du semi-conducteur P. L'atome d'impureté trivalent laisse apparaître une liaison pendante dans le réseau. Cette liaison est associée à une orbite vacante, qui est très rapidement (dés l'opération de dopage), remplie par un électron de valence donné par un Silicium avoisinant : l'atome de Bore s'ionise négativement, mais le trou crée par le déplacement d'un électron du Silicium vers l'orbite vacante reste, tant que l'agitation thermique est faible, **localisé** à proximité du Bore, du fait de l'attraction électrostatique. On parle alors, pour marquer l'analogie avec le cas précédent, de **trou**

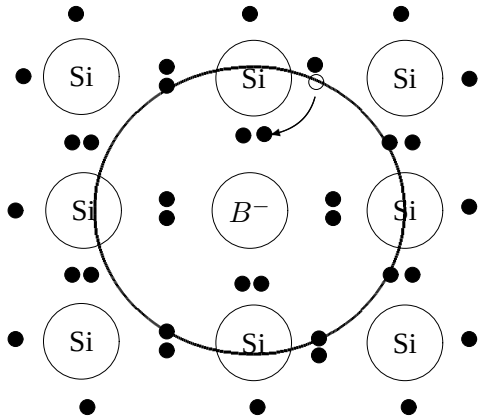


FIG. 4.14 – Principe du dopage de type P : on substitue à certains atomes de Silicium un atome d'impureté trivalent (ici, du Bore), qui fait apparaître une place vacante, située dans une liaison de valence ; cette lacune est immédiatement comblée par un électron de la bande de valence, fourni par une liaison de valence d'un atome de Silicium situé à proximité du Bore ; le trou ainsi crée est localisé dans le voisinage du Bore, du fait de l'attraction électrostatique par l'ion négatif, mais est susceptible de se détacher dés que l'agitation thermique est suffisante.

faiblement lié.

Dés que l'agitation thermique est suffisante pour vaincre l'attraction électrostatique, ce trou peut se détacher du Bore et participer, en peuplant la bande de valence, à la conduction du courant électrique. Le dopage a alors augmenté la population de trous, qui deviennent à température ambiante, **porteurs majoritaires** : la conduction du courant électrique s'effectue essentiellement grâce aux trous, et l'atome d'impureté est appelé **atome accepteur**, puisqu'il apporte des trous au matériau en "acceptant" des électrons du réseau initial.

Le niveau occupé par le trou *lié* au Bore ionisé est situé dans le bas de la bande interdite (figure 4.15) : le trou n'est pas mobile, mais la proximité de E_A et de E_V permet au trou de passer facilement à la bande de valence (ou, formulé avec les électrons : un électron de la bande de valence peut facilement passer sur le niveau E_A, *i.e.* remplir le trou lié, et ce faisant, il crée un trou libre dans la bande de valence).

La probabilité d'ionisation d'un atome accepteur est donnée par la distribution suivante :

$$\boxed{N_A^- = \frac{N_A}{1 + 2e^{(E_A - E_F)/kT}}} \tag{4.9}$$

qui est en tout point symétrique de la distribution correspondant à la probabilité d'ionisation des atomes donneurs. On peut là encore déterminer le niveau de Fermi à partir de l'équation de neutralité électrique :

$$n + N_A^- = p$$

qui signifie que les trous de la bande de valence proviennent, d'une part de la génération de paires électron-trou, d'autre part de l'ionisation des atomes accepteurs (ou plutôt, de la perte du trou lié par l'atome accepteur).

A basse température, N_A^- est proche de zéro, et la majorité des trous provient de la génération de paires : le comportement du semi-conducteur est intrinsèque, il est quasi-isolant.

A température ambiante, l'ionisation est presque totale (saturée) : la majorité des trous proviennent du dopage, presque chaque ion accepteur ayant perdu le trou qui lui était lié. On

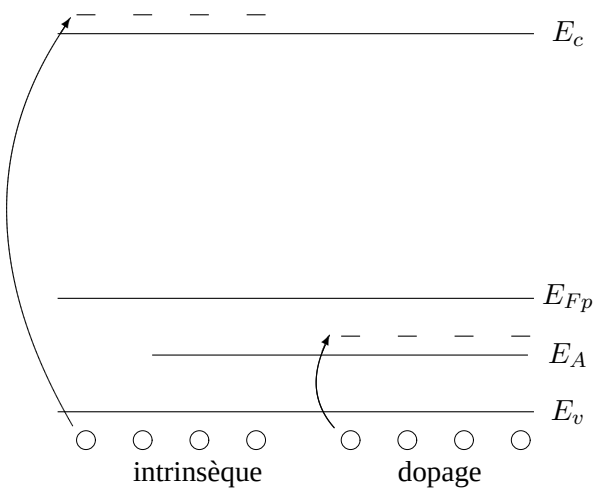


FIG. 4.15 – Avant détachement thermique, le trou lié au Bore (ionisé négativement) se trouve sur le niveau **accepteur** E_A, très proche de la bande de valence : le gap pour atteindre la B.V. est faible, et peut être franchi à température ambiante. Une fois détachés, les trous issus de ce niveau forme, avec les trous de la B.V. déjà présents, l'ensemble des trous libres (*i.e.* de conduction) du cristal.

peut négliger les trous provenant de la génération de paires (terme n) devant ceux provenant du dopage, et écrire :

$$p \sim N_A$$

On voit donc que les porteurs majoritaires dans un semi-conducteur P sont les trous, et qu'à température ambiante, ils sont quasiment en même proportion que les atomes accepteurs. On a donc, dans le matériau, à cette température :

- des trous libres, majoritaires, en concentration proche de N_A ;
- des électrons issus de la génération de paires, minoritaires ;
- des impuretés ionisées négativement, fixes (atomes de Bore ionisés).

Pour déterminer le niveau de Fermi dans cette approximation, on remplace p par son expression (équation 4.5) :

$$N_v e^{(E_V - E_F)/kT} \sim N_A$$

qui donne pour le niveau de Fermi l'expression suivante,

$$E_F \sim E_V + kT \ln \frac{N_v}{N_A}$$

Comme (sauf cas de surdopage) $N_v > N_A$, le niveau de Fermi est en bas de la bande interdite, à proximité de la bande de valence : ceci traduit bien une baisse de population d'électrons par rapport au Silicium pur, donc le fait que les trous deviennent majoritaires.

Enfin, à haute température, on retrouverait, pour les mêmes raisons évoquées précédemment, un comportement de semi-conducteur intrinsèque, la création de trous par génération de paires devenant très importante, et l'ionisation des atomes accepteurs arrivant à saturation : dans ce cas, le niveau de Fermi rejoint le milieu de la bande interdite, comme pour un semi-conducteur non-dopé.

Désormais, nous avons tous les résultats pour étudier le comportement d'une jonction PN, constitué par la juxtaposition d'un semi-conducteur P et d'un semi-conducteur N : nous verrons, dans la partie suivante consacrée au phénomènes de transports, que la mise

en contact s'accompagne d'une diffusion des porteurs de charge de part et d'autre de la jonction, que cette diffusion fait apparaître un champ électrique, et constitue de fait une barrière de potentiel que les porteurs ne peuvent franchir que dans un seul sens...

Chapitre 5

Phénomènes de transport

L'essentiel de l'étude a porté jusqu'à présent sur le comportement du semi-conducteur **à l'équilibre thermodynamique**. Or, dés lors qu'une jonction PN est polarisée, elle se trouve précisément **hors de l'équilibre**, puisqu'un courant (même faible dans le cas d'une polarisation inverse) y circule. L'objet de ce chapitre est donc de modéliser les différents phénomènes qui contribuent à la circulation d'un courant électrique dans le semi-conducteur : ces phénomènes sont appelés *phénomènes de transport*, et relèvent essentiellement de deux classes de phénomènes :

- la **conduction** : un courant circule sous l'action d'un champ électrique macroscopique, du par exemple à une polarisation extérieure ;
- la **diffusion** : il s'agit d'un transport de charge dû au caractère **localement inhomogène** de la répartition des charges électriques dans le matériau.

Par ailleurs, signalons pour mémoire que d'autres phénomènes de transport existent, qui portent sur des grandeurs autres que le courant électrique :

- la diffusion thermique : c'est un transport d'énergie cinétique (d'agitation thermique) créé par un gradient de température ;
- la viscosité (uniquement dans un gaz) : il s'agit d'un transport de quantité de mouvement entre couches adjacentes d'un fluide.

> **Remarque :** Ces deux derniers phénomènes ne feront l'objet d'aucune étude détaillée ici, puisqu'ils n'interviennent pas directement dans la modélisation des composants électroniques à semi-conducteurs. Retenez toutefois que la diffusion thermique, et en particulier son couplage non-linéaire à la conduction ou à la diffusion électrique — connues sous le nom d'effet Seebeck, Peltier, et Thompson — est à la base du fonctionnement des thermocouples et des modules de refroidissement Peltier, entre autres dispositifs thermo-électriques.

Ainsi, un phénomènes de transport se distingue par les propriétés suivantes :

- il correspond à un transport d'une grandeur physique dans le matériau : charges électriques, énergie cinétique...
- il résulte d'une perturbation extérieure au matériau, ou bien d'une inhomogénéité locale ;
- il traduit la **tentative de retour à l'équilibre thermodynamique** du matériau, lorsque ce dernier est soumis à une perturbation ou à une inhomogénéité.

Cette dernière propriété est sans doute la plus universelle. D'ailleurs, nous la retrouverons lorsque nous aborderons les phénomènes de génération/recombinaison : nous verrons alors que le processus de génération traduit le retour du produit np à sa valeur à l'équilibre thermodynamique, *i.e.* n_i^2 (loi d'action de masse). Et dans le cas de la diffusion, la tentative

de retour à l'équilibre se traduit par la tendance du matériau à **homogénéiser** les concentrations.

Il n'est pas évident *a priori* que l'ensemble de ces phénomènes — conduction, diffusion et recombinaison — puissent être étudiés séparément. Si c'est le cas, ce n'est qu'en raison d'une différence de plusieurs ordres de grandeur des temps caractéristiques de chacun des phénomènes précédents. En l'occurence, si l'on s'intéresse à la conduction, on constate que le temps de relaxation du gaz d'électrons — temps moyen entre deux collisions — est de l'ordre de $0,1ps$. En comparaison, la durée de vie d'une paire électron-trou — temps moyen au-delà duquel l'électron de conduction se recombine avec le trou de la bande de valence — va de $1ns$ à $1ms$, et s'avère donc de plusieurs ordres de grandeurs supérieure. La conséquence fondamentale est que l'on peut étudier la conduction et la diffusion des trous et des électrons séparément, en considérant que ce sont deux gaz qui n'interagissent pas aux échelles de temps caractéristiques de la conduction.

5.1 Modèle de Drude de la conduction et de la diffusion

Le modèle de Drude propose un traitement classique de la conduction et de la diffusion. Les électrons et les trous y sont modélisées comme des particules classiques, affublées cependant d'une masse effective traduisant leur interaction avec le réseau cristallin. Ces particules peuvent donc être considérées comme de petites "billes" enfermées dans une boîte parallélépipédique, au parois parfaitement réfléchissantes. On voit là tout l'intérêt du concept de masse effective, qui permet de s'affranchir, lors de l'étude des phénomènes de transport, de la complexité d'une modélisation quantique, tout en conservant l'ingrédient fondamental qu'est l'interaction électron-réseau.

5.1.1 Conduction électrique

La conduction électrique résulte d'une compétition entre deux phénomènes physiques :

- l'accélération du gaz d'électrons sous l'effet d'un champ électrique extérieur (polarisation, ...) ;
- le freinage dû aux chocs des électrons sur les écarts à la périodicité du réseau cristallin.

Rappelons qu'en vertu du théorème de Bloch, un électron se propage sans collision (de bout en bout) dans un réseau parfaitement périodique. Seuls des écarts à la périodicité sont donc à l'origine des diffusions (*i.e.* des chocs) des électrons/trous sur le réseau. Ces écarts sont dus pour l'essentiel :

- à l'agitation thermique des noyaux, proportionnelle à la température : des lors que la température n'est plus nulle, la position instantanée des noyaux n'est plus celle des noeuds du réseau, et le réseau n'est plus parfaitement périodique ;
- aux défauts cristallins : lacunes, interstices, dislocations sont autant de causes "mécaniques" d'écarts à la périodicité ;
- à la présence d'impuretés de dopage : dans ce dernier cas, il existe une compétition entre l'augmentation de conductivité due au dopage, et l'augmentation de résistivité due à la diffusion des électrons par les impuretés.

Retenons donc que,**toutes choses égales par ailleurs**, les écarts à la périodicité réduisent la conductivité : ces écarts ont pour causes essentielles la température et le dopage.

Dans le modèle de Drude, le traitement classique de la diffusion repose sur une modélisation probabiliste des collisions, puisque ce qui intéresse au final le physicien du composant est le courant électrique macroscopique, et non les détails du mouvement de chaque électron.

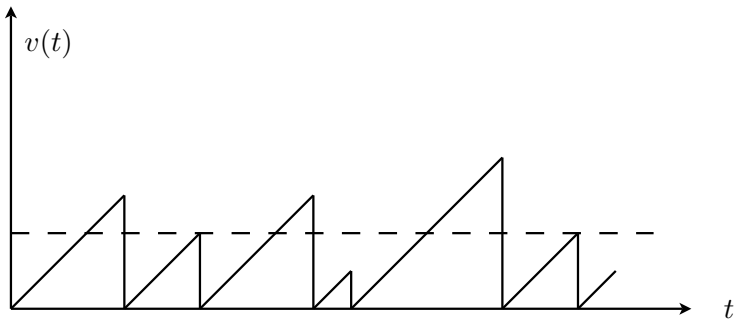


FIG. 5.1 – La vitesse d'ensemble du gaz d'électrons est le résultat d'une moyenne sur un grand nombre de chocs.

Ainsi que l'illustre la figure 5.1, chaque électron possède, si l'on se restreint à un mouvement unidimensionnel, un mouvement en dent de scie. En présence d'un champ électrique extérieur, constant dans le temps, et uniforme $\mathcal{E}$, l'électron/trou subit une force

$$\mathbf{F} = \pm q\mathcal{E} = m^*\mathbf{a}$$

et voit donc sa vitesse instantanée augmenter linéairement au cours du temps, jusqu'à ce qu'une collision avec un ion du réseau ramène sa vitesse à zéro (c'est évidemment une modélisation très simpliste, mais qui suffira amplement à notre propos) :

$$v(t) = \pm\frac{q\mathcal{E}t}{m^*}$$

L'idée clé est donc de modéliser ces collisions de manière probabiliste, puis de réaliser une moyenne sur les temps de collision pour obtenir la vitesse d'entraînement du gaz.

On postule que la probabilité de collision par unité de temps vaut $1/\tau$, τ étant un temps caractéristique que l'on suppose ici **indépendant du niveau d'énergie** $E(k)$ sur lequel se trouve l'électron.

> **Remarque :** En toute rigueur, la "sensibilité" aux collisions dépend de l'énergie cinétique de l'électron subissant le choc : en l'occurence, l'électron doit sentir d'autant plus fortement l'effet d'une collision qu'il possède une faible énergie cinétique. La valeur de τ que nous utiliserons sera donc une valeur moyennée sur l'ensemble des niveaux d'énergie, de la bande de conduction pour les électron (notée τ_n), de la bande de valence pour les trous (notée τ_p). Notez qu'en réalité, ces temps sont le plus souvent mesurés expérimentalement (cf. le TD/BE "mesures par résonance cyclotron de la structure des bandes du Silicium").

La probabilité de collision entre t et $t + dt$ vaut donc dt/τ et la probabilité de non-collision entre t et $t+dt$ vaut par conséquent $1-dt/\tau$. Considérons un électron subissant une collision à l'instant $t = 0$, et soit $p(t)$ la probabilité pour qu'il ne subisse **aucune collision** entre $t = 0$ et l'instant t. La probabilité qu'il ne subisse aucune collision entre $t = 0$ et $t+dt$ vaut donc :

$$p(t + dt) = p(t)\left(1 - \frac{dt}{\tau}\right)$$

ce qui s'écrit encore

$$p(t) + p'(t)dt = p(t)\left(1 - \frac{dt}{\tau}\right)$$

L'intégration de cette équation différentielle du premier ordre conduit au résultat

$$p(t) = e^{-t/\tau}$$

(correspondant d'ailleurs à la loi de Poisson). On en déduit que la probabilité pour qu'un électron ne subisse aucune collision entre $t = 0$ et t, et en subisse une entre t et $t + dt$ s'écrit :

$$\mathcal{P}(t)dt = e^{-t/\tau}\frac{dt}{\tau}$$

Ce résultat représente, en somme, la probabilité qu'un électron ayant subit une collision à $t = 0$, en subisse à nouveau une à l'instant t. Par conséquent, le temps moyen entre deux collisions est simplement

$$\langle t \rangle = \int_0^\infty t\, \mathcal{P}(t)dt = \tau$$

Le temps caractéristique τ représente le temps moyen entre deux collisions.

Remarque : A partir du temps moyen, on peut définir le **libre parcours moyen** l, ou distance moyenne parcourue entre deux collisions. Si v_m est la vitesse moyenne d'agitation thermique, on a $l = v_m\tau$. L'énergie cinétique d'agitation thermique vaut $e_c = \frac{3}{2}kT = \frac{1}{2}m^*v_m^2$; à $300\ K$, on obtient $v_m \sim 10^5\ m/s$ (soit $400000\ km/h$!). Avec un valeur de τ de l'ordre du dixième de picoseconde, le libre parcours moyen est de l'ordre de $10\ nm$, soit à peu près 20 distance interatomique.

En présence d'un champ électrique $\mathcal{E}$ constant et uniforme, la vitesse d'entraînement vaut donc simplement :

$$\langle v \rangle = \pm\frac{q\mathcal{E}\tau}{m^*} = \pm\mu\mathcal{E} \tag{5.1}$$

où

$$\boxed{\mu = \frac{q\tau}{m^*}} \tag{5.2}$$

est la **mobilité**. On constate bien que, comme prévu, la mobilité croît avec le temps moyen entre collision.

Finalement, le courant électrique s'écrivant

$$J_n = -nqv_n$$

pour les électrons, et

$$J_p = pqv_p$$

pour les trous (n et p étant les concentrations volumiques, respectivement en électrons et en trous), on en déduit la loi d'Ohm locale :

$$J_n = \frac{nq^2\mathcal{E}\tau_n}{m_n^*} = \sigma_n\mathcal{E}$$

pour les électrons, et

$$J_p = \frac{pq^2\mathcal{E}\tau_p}{m_p^*} = \sigma_n\mathcal{E}$$

pour les trous. Les conductivités s'écrivent donc respectivement

$$\boxed{\sigma_n = \frac{nq^2\tau_n}{m_n^*} = nq\mu_n} \tag{5.3}$$

et

$$\boxed{\sigma_p = \frac{pq^2\tau_p}{m_p^*} = pq\mu_p} \tag{5.4}$$

pour les électrons et les trous.

Le courant de conduction total est la somme des courants de trous et d'électrons,

$$J_{cond} = J_n + J_p = q(n\mu_n + p\mu_p)\mathcal{E}$$

et la conductivité totale du matériau s'écrit naturellement :

$$\boxed{\sigma = \sigma_n + \sigma_p = nq\mu_n + pq\mu_p}$$

Le tableau suivant donne quelques ordres de grandeur pour le Silicium et l'Arseniure de Gallium (AsGa) : les mobilités sont exprimées en $cm^2/V/s$.

Si	$\mu_n = 1400$	$\mu_p = 500$
Ge	$\mu_n = 4000$	$\mu_p = 1900$
$AsGa$	$\mu_n = 8500$	$\mu_p = 400$

Notons pour terminer qu'on peut modéliser l'effet des collisions sous la forme d'une force de type "frottement visqueux" s'appliquant sur la **vitesse d'ensemble** des électrons. En choisissant la force telle que

$$\mathbf{F_f} = -\frac{m^*}{\tau}\mathbf{v}$$

le principe fondamental de la dynamique s'écrit :

$$m^*\mathbf{a} = \pm q\mathcal{E} - \frac{m^*}{\tau}\mathbf{v}$$

soit encore

$$m^*\frac{d\mathbf{v}}{dt} + \frac{m^*}{\tau}\mathbf{v} = \pm q\mathcal{E}$$

qui s'intègre en

$$\mathbf{v}(t) = \pm\frac{q\mathcal{E}}{m^*\tau}(1 - e^{-t/\tau})$$

et permet d'aboutir au même résultat que précédemment (en régime permanent). L'interprétation de τ est alors celle d'un **temps de relaxation** du gaz d'électrons : c'est le temps nécessaire au gaz pour atteindre le régime permanent en réponse à un champ électrique extérieur. Ce temps définit également une fréquence de coupure haute $f_c = 1/2\pi\tau \sim 1000GHz$ au-delà de laquelle l'inertie du gaz est telle qu'il ne répond plus au champ électrique.

5.1.2 Diffusion électrique

La diffusion de particule résulte, de manière très générale, d'une inhomogénéité locale de concentrations, qui traduit une **rupture d'équilibre thermodynamique** : le retour à l'équilibre tend à homogénéiser les concentrations par transport de particules.

Comme nous l'avons vu au chapitre 4, une manière d'aborder le problème consiste à voir l'équilibre diffusif comme le résultat d'un équilibre entre **potentiels chimiques**, par exemple lorsque deux gaz sont mis en contact par l'intermédiaire d'une paroi poreuse (figure 5.2). Rappelons que, dans ce modèle, les particules diffusaient des régions à fort potentiel chimique vers les régions à faible potentiel, jusqu'à ce que les potentiels s'équilibrent.

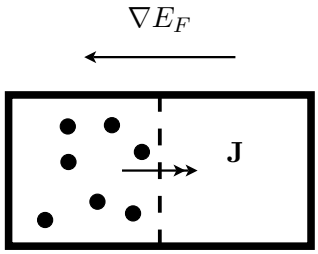


FIG. 5.2 – La diffusion à travers une paroi "poreuse" (ou bien une interface entre deux matériaux) s'établit dans le sens des niveaux de Fermi E_F (=potentiel chimique) décroissants. Une autre interprétation consiste à dire qu'elle homogénéise les concentrations de part et d'autre de la paroi. A l'équilibre thermodynamique (en l'absence de courant), le niveau de Fermi E_F est homogène.

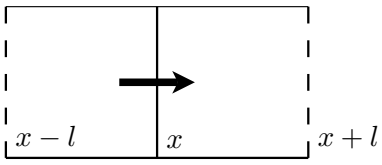


FIG. 5.3 – Une tranche de gaz d'épaisseur l petite, dans laquelle existe un gradient de concentration.

Cette approche nous fournit une première information **qualitative**, car elle nous indique dans quel sens diffusent les particules lorsque le système approche l'équilibre thermodynamique. Elle ne nous indique pas, en revanche, l'intensité de cette diffusion, *i.e.* le courant électrique résultant du transport de charge : c'est l'objet de la loi de Fick.

> **Remarque :** Il y a une autre raison à ne pas utiliser les potentiels chimiques pour traiter **quantitativement** le cas de la diffusion. La diffusion est un processus thermodynamique fondamentalement **hors d'équilibre**. Or la notion de potentiel chimique, comme de toute autre grandeur thermodynamique vu jusqu'alors, n'a de sens qu'à l'équilibre. Notez qu'il existe toutefois une méthode de résolution du problème introduisant des *pseudo-potentiels chimiques* ou *pseudo-niveaux de Fermi* ayant également un sens hors de l'équilibre thermodynamique, mais que nous n'aborderons pas ici.

Pour établir la loi de Fick, considérons (figure 5.3) une tranche de gaz[1] d'épaisseur $2l$ où l correspond au **libre parcours moyen**, dans laquelle existe un gradient de concentration ∇n dirigé selon l'axe Ox, et déterminons le courant résultant de ce gradient à travers la section s d'abscisse x.

Pour faire simple, on peut considérer que, parmi toutes les particules se trouvant entre l'abscisse $x-l$ et x, un tiers ont leur vitesse dirigée selon Ox (les deux autres tiers étant sur Oy et Oz), et parmi ces dernières, la moitié ont une vitesse positive. Comme l est le libre parcours moyen, ces dernières particules franchiront la section s sans avoir subi de collision : elle participerons donc au courant de diffusion.

Notant $n(x-l)$ la concentration volumique à l'abscisse $x-l$ (et considérant que cette concentration est constante entre $x-l$ et x, ce qui, vu l'ordre de grandeur de l, est à peu

[1] Le fait qu'il s'agisse d'un gaz de molécules, d'électrons, ... n'a aucune importance ici.

près vérifié), le nombre de particules se trouvant entre les abscisses $x - l$ et x et traversant **effectivement** la section s vaut $n(x-l)sl/6$. Cette traversée prend un temps égal à l/v_m où v_m est la **vitesse moyenne d'agitation thermique**. Par conséquent, la **densité de courant** circulant de la gauche vers la droite à travers s vaut :

$$J_1 = \frac{n(x-l)l}{6}\frac{v_m}{l} = \frac{n(x-l)v_m}{6}$$

En procédant de la même manière pour les particules se trouvant à droite de x, on obtient pour la densité de courant circulant de droite à gauche :

$$J_2 = -\frac{n(x+l)v_m}{6}$$

La densité de courant totale circulant à travers s vaut donc

$$J_D = J_1 + J_2 = \frac{(n(x-l) - n(x+l))v_m}{6} \sim -n'(x)lv_m/3$$

Vu les simplifications apportées au modèle, le facteur 3 a toutes les chances d'être peu réaliste. En revanche, la relation de proportionnalité entre gradient de concentration et courant de diffusion est tout-à-fait crédible. On écrira donc finalement

$$J_D = -D\nabla n \tag{5.5}$$

où $D = lv_m/3$ est la **constante de diffusion**. Comme D est positive, on constate que le mouvement de diffusion des particules a lieu dans le sens opposé au gradient, ce qui tend bien à **homogénéiser** les concentrations.

Le courant (la densité de courant pour être exact) que nous avons calculé est un courant de particule. Il faut donc multiplier par q ou $-q$ pour obtenir un courant *électrique*. On obtient donc

$$\boxed{J_{Dn} = qD_n\nabla n} \tag{5.6}$$

pour les électrons, et

$$\boxed{J_{Dp} = -qD_p\nabla p} \tag{5.7}$$

pour les trous.

On voit que, à l'instar de ce qui se passait pour la conduction électrique, la diffusion est d'autant plus importante que, toute chose égales par ailleurs, le libre parcours moyen est plus élevée : conduction et diffusion dépendent donc de la même façon des collisions. Cette similarité de comportement est formalisé dans la relation d'Einstein qui lie mobilité et constante de diffusion. Rappelons que la mobilité s'écrit

$$\mu = \frac{q\tau}{m^*}$$

Comme libre parcours moyen et temps moyen entre collision sont liés par la relation $l = v_m\tau$, on a

$$D = \frac{lv_m}{3} = \frac{v_m^2\tau}{3}$$

L'énergie cinétique d'agitation thermique s'écrivant

$$e_c = \frac{3}{2}kT = \frac{1}{2}m^*v_m^2$$

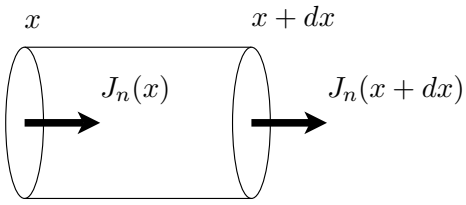


FIG. 5.4 – Équation de conservation de la charge.

on en déduit

$$\boxed{D = \frac{kT}{q}\mu = U_T\mu} \tag{5.8}$$

(*i.e.* $D_n = \frac{kT}{q}\mu_n$ et $D_p = \frac{kT}{q}\mu_p$) relation qui constitue la **relation d'Einstein** (enfin, une des nombreuses, tant cet homme-là fût prolixe, on ne le dira jamais assez...).

Enfin, précisons que, lorsqu'on est en présence, à la fois d'un champ électrique et d'un gradient de concentration, les courants de conduction et de diffusion s'ajoutent, soit :

$$\boxed{\mathbf{J}_n = nq\mu_n\mathcal{E} + qD_n\nabla n} \tag{5.9}$$

et

$$\boxed{\mathbf{J}_p = pq\mu_p\mathcal{E} - qD_p\nabla p} \tag{5.10}$$

respectivement pour les électrons et les trous, et par conséquent

$$\boxed{\mathbf{J} = q(n\mu_n + p\mu_p)\mathcal{E} + q(D_n\nabla n - D_p\nabla p)} \tag{5.11}$$

pour le courant total dans le matériau.

5.2 Équation de conservation

Terminons se chapitre par l'obtention d'une des équations les plus fondamentales pour la modélisation des composants à semi-conducteur, l'équation de conservation de la charge électrique. Considérons (figure 5.4) une tranche élémentaire cylindrique de matériau, d'épaisseur dx, et un courant électrique circulant selon l'axe Ox. La variation du nombre de charge **à l'intérieur** du cylindre est due :

- aux flux de charges entrant dans le cylindre ;
- à la création de charges (paires électron/trous) par processus de génération thermique ;
- à la disparition de charges par recombinaison.

Notant $\rho(t)$ la densité volumique de charge dans le cylindre à l'instant t, le nombre de charges à l'intérieur du cylindre vaut $\rho s dx$. Notons également U le nombre de charges disparaissant algébriquement par seconde et par unité de volume, du fait des processus de génération/recombinaison (U sera explicité au chapitre suivant : il est positif si les recombinaisons l'emportent, négatif sinon).

Le bilan s'écrit donc, puisque le courant est exclusivement dirigé selon l'axe Ox :

$$s dx\frac{\partial\rho}{\partial t} = \int_s J(x).dS - \int_{s'} J(x+dx)\ dS - U\ s\ dx = (J(x) - J(x+dx))s - U\ s\ dx$$

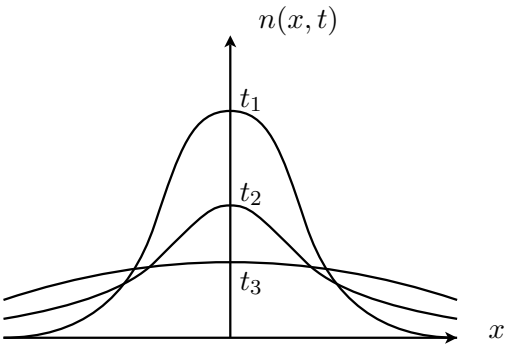


FIG. 5.5 – L'équation de diffusion traduit la tendance du système à homogénéiser les concentrations.

soit encore

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x} - U$$

La généralisation à trois dimension (ou bien le théorème d'Ostrogradsky appliqué à toute la surface du cylindre) s'écrit :

$$\frac{\partial \rho}{\partial t} = -\operatorname{div} J - U$$

Enfin, du fait de la différence d'ordre de grandeur entre le temps moyen entre collisions τ et la durée de vie d'une paire électron/trou, les courant d'électrons et de trous se conservent séparément :

$$\boxed{q\frac{\partial n}{\partial t} = \frac{\partial J_n}{\partial x} - U_n} \tag{5.12}$$

et

$$\boxed{q\frac{\partial p}{\partial t} = -\frac{\partial J_p}{\partial x} - U_p} \tag{5.13}$$

En l'absence de processus de génération/recombinaison ($U = 0$), et en considérant le champ électrique uniforme, le terme $\operatorname{div} J$ se réduit à

$$qD_n \operatorname{div} \nabla n = qD_n \frac{\partial^2 n}{\partial x^2} \text{ et } -qD_p \operatorname{div} \nabla p = -qD_p \frac{\partial^2 p}{\partial x^2}$$

pour les électrons et les trous, respectivement. Reportant dans l'équation de conservation de la charge, on obtient respectivement

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} \text{ et } \frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} \tag{5.14}$$

Ces deux dernières équations portent le nom d'**équation de la diffusion**. On voit que le second membre est un Laplacien, opérateur associé à la **concavité** d'une fonction (positif si la concavité est tournée vers le haut). S'il existe localement un maximum de concentration (cf. figure 5.5), le Laplacien prend une valeur négative, et le terme de gauche est négatif : la concentration diminue localement au cours du temps, et l'équation de conservation de la charge affirme simplement que les particules "perdues" dans la zone du maximum doivent, soit avoir éventuellement disparu du fait des recombinaisons, soit se retrouver hors de la zone du maximum (qui du coup, décroît...).

Chapitre 6

Recombinaisons

6.1 Absorption de la lumière par les semi-conducteurs

Considérons une onde électromagnétique monochromatique plane, de pulsation ω, de vecteur d'onde k, "tombant" sur un cristal semi-conducteur. Le photon associé à cette onde électromagnétique est caractérisé par :

- son énergie $E = h\nu = \hbar\omega$;
- sa quantité de mouvement $\mathbf{p} = \hbar\mathbf{k}$, avec $\omega = kc/n$, où n est l'indice optique du milieu.

D'après la *théorie des perturbations dépendant du temps* (cf. polycopié de Physique Quantique), une onde électromagnétique de pulsation ω est capable de provoquer une transition énergétique d'un électron entre deux niveaux distants d'une énergie E telle que, *en première approximation* :

$$\hbar\omega - \delta E/2 < E < \hbar\omega + \delta E/2$$

où δE la **largeur naturelle de raie d'absorption** (figure 6.1). δE vérifie la seconde relation d'incertitude de Heisenberg :

$$\delta E \times \delta t \geq \hbar$$

où δt est la durée d'action de la perturbation électromagnétique (à l'instant de l'absorption de l'onde, un photon est annihilé et la perturbation cesse).

Ce processus d'absorption est entièrement régit par les lois de la mécanique quantique. Il est donc, en toute rigueur, caractérisé par une **probabilité de transition** $\mathcal{P}_{E_i \to E_f}(t)$, dépendant entre autres :

- du temps : elle croît avec le temps pour des temps suffisamment faibles ;
- de la pulsation de l'onde et de la différence d'énergie entre les deux niveaux E_i (initial) et E_f (final) : elle est maximale lorsque $\hbar\omega = E_f - E_i$, et décroît au fur et à mesure que $\hbar\omega$ s'éloigne de $E_f - E_i$;
- de l'élément de matrice $\langle\varphi_i| H(t) |\varphi_f\rangle$ de l'Hamiltonien d'interaction *électron-champ* entre les états quantiques φ_i et φ_f : cet élément de matrice de l'Hamiltonien d'interaction dépend des caractéristiques du champ électromagnétique (direction de propagation, type de polarisation), et des orbitales initiales et finales occupées par l'électron subissant l'interaction.

Un photon d'énergie E peut donc provoquer des transitions entre des niveaux qui ne sont pas exactement séparés de E, étant entendu que, plus la durée d'action de la perturbation électromagnétique (avant absorption effective du photon) est longue, plus la probabilité de

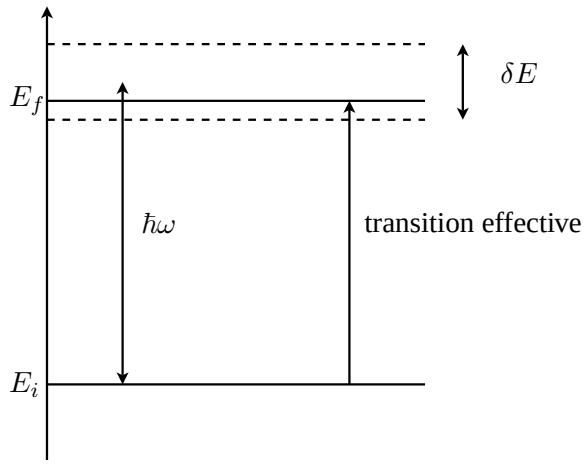


FIG. 6.1 – Une onde électromagnétique de pulsation ω peut provoquer une transition énergétique entre niveaux séparées de $E = \hbar\omega \pm \delta E/2$, δE étant d'autant plus faible que la durée d'action δt de l'onde — avant que l'absorption est effectivement lieu — est longue. Cette incertitude définit la **largeur naturelle de la raie d'absorption**.

transition est *piquée* autour de la valeur $\hbar\omega = E_f - E_i$. C'est bien ce que traduit la seconde relation d'incertitude d'Heisenberg.

6.1.1 Transitions directes

L'énergie n'est pas la seule grandeur intervenant dans un processus d'absorption. Considérons le mécanisme d'absorption comme une collision du type :

$$\text{photon} + e^- \rightarrow e^-$$

Deux grandeurs (au moins) se conservent dans une collision [1] :

- l'énergie totale : $\hbar\omega + E_i = E_f$, à δE près ;
- la **quantité de mouvement** (*i.e.* le vecteur d'onde à $\hbar$ près) :

$$\hbar\mathbf{k} + \hbar\mathbf{k}_i = \hbar\mathbf{k}_f$$

à $\delta\mathbf{k}$ près (cf. première inégalité d'Heisenberg), où $\mathbf{k}$, $\mathbf{k}_i$ et $\mathbf{k}_f$ sont respectivement les vecteurs d'onde de l'onde électromagnétique, de l'électron avant et après la collision.

Essayons d'évaluer un ordre de grandeur du vecteur d'onde de l'onde électromagnétique. Pour un semi-conducteur, le seuil d'absorption en énergie est le gap E_g, car seules les transitions **interbandes** sont radiatives, *i.e.* correspondent à une absorption/émission de lumière.

> **Remarque :** *A contrario*, on a vu au chapitre 3, que les transitions **intrabandes** étaient dues à une force lentement variable, de faible intensité, comme par exemple un champ électrique uniforme du à une polarisation extérieure par une pile ou un générateur de type GBF. Dans ce cas, la transition est régie par le *théorème accélération*, $\mathbf{F} = \hbar d\mathbf{k}/dt$.

Le fait que le gap fixe le seuil énergétique d'une absorption de lumière permet de donner un ordre de grandeur au vecteur d'onde k de l'onde "absorbable". Le gap est, pour le

[1] Qu'elle soit traitée par la mécanique newtonienne ou quantique ne change rien sur ce point-ci, hormis l'incertitude apportée par les relations d'Heisenberg, et le décompte du spin dans la conservation du moment cinétique

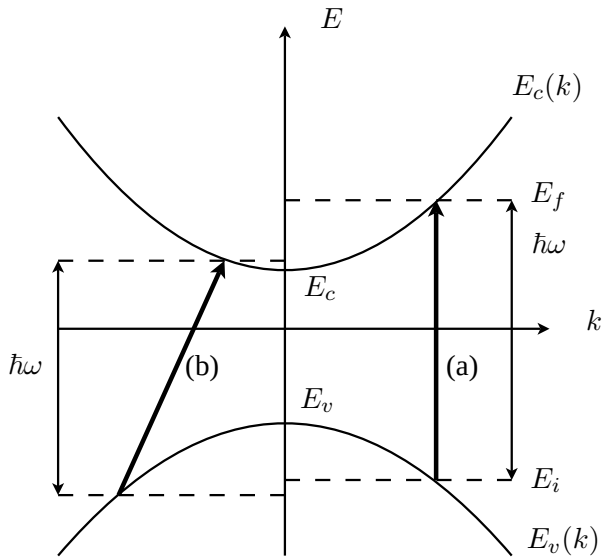


FIG. 6.2 – Absorption directe (a) ou indirecte (b) dans un semi-conducteur.

Silicium, de $1,12\ eV$, et en général, $1\ eV$ est un bon ordre de grandeur pour les autres semi-conducteurs de la colonne IV. La relation :

$$\hbar\omega > E_g$$

conduit donc, puisque $\omega = kc$, à la relation suivante pour le vecteur d'onde :

$$k > \frac{E_g}{\hbar c}$$

La longueur d'onde maximale est donc de l'ordre de $1\mu m$, soit environ 10000 fois plus que le pas habituel d'un réseau cristallin, qui définit la taille de la première zone de Brillouin. Le vecteur d'onde de l'onde électromagnétique est donc généralement négligeable devant les vecteurs d'onde électroniques k_i et k_f. On peut donc considérer que, lors de l'absorption, la quantité de mouvement de l'électron est inchangé, et cette transition porte donc le nom de **transition directe** ou **verticale** : l'électron change de bande d'énergie en conservant le même vecteur **k** (autrement dit, il reste dans la même case de l'espace réciproque, mais en changeant de bande). Cette transition est illustrée figure 6.2, et correspond au cas (a).

6.1.2 Coefficient d'absorption pour une transition directe entre extréma

L'intensité lumineuse suit une loi du type

$$I(x) = I_0 e^{-\alpha x}$$

x étant la distance parcourue par l'onde dans le matériau. Le **coefficient d'absorption** d'un semi-conducteur est donc définit comme l'inverse de la distance au-delà de laquelle 63% du flux lumineux est absorbé.

On calcule α en utilisant la théorie des perturbations dépendant du temps, α étant proportionnel à la probabilité **totale** d'absorption d'un photon, définie comme la *somme des probabilités de transitions $\mathcal{P}_{i\to f}$ entre états de la bande de valence et états de la bande de conduction.*

On montre que, pour une **transition directe entre extréma de bandes**, α s'écrit :

$$\alpha = \frac{(2m_r)^{3/2}}{2\pi\epsilon_0 cn}\frac{e^2p^2}{m^2}\frac{1}{\hbar^3\omega}(\hbar\omega - E_g)^{1/2}$$

où :

- m_r est la masse effective réduite, moyenne harmonique des masses effectives des trous et des électrons : $m_r^{-1} = m_e^{-1} + m_t^{-1}$;
- n est l'indice optique du matériau : $n = \sqrt{\epsilon_r}$:
- $p = \hbar k_i = \hbar k_f$ est la quantité de mouvement de l'électron (inchangée dans une transition directe) ;
- m est la masse de l'électron dans le vide (donc la "vraie" masse).

α est donc variable d'un semi-conducteur à l'autre via tous les paramètres précédents, en particulier les masses effectives des trous et électrons (donc au final, la courbure du diagramme des bandes à proximité des extréma). Pour un matériau donné, il est proportionnel à $(\hbar\omega - E_g)^{1/2}$: la probabilité d'absorption croît très rapidement lorsque l'énergie des photons devient très supérieure au gap.

Un ordre de grandeur moyen est :

$$\alpha = 4.10^4(\hbar\omega - E_g)^{1/2}$$

$\hbar\omega$ et E_g étant exprimés en eV. Par exemple, pour un gap de $1\ eV$ et une onde de longueur d'onde $1\mu m$ (soit $\hbar\omega \sim 1.2eV$), on trouve $\alpha \sim 10^4\ cm^{-1}$. L'onde est donc presque intégralement absorbée au-delà de quelques micromètres de la surface du matériau.

6.1.3 Transitions indirectes

Au vu des ordres de grandeur précédents, une transition indirecte, *i.e.* lors de laquelle le vecteur d'onde de l'électron est modifié, doit faire intervenir une troisième particule. C'est le rôle joué par les **phonons**, *pseudo-particules* associées aux vibrations des ions du cristal [2] : pour faire simple, un phonon est un **mode propre** de vibration du cristal, caractérisé par une **énergie** E, une pulsation ω (avec $E = \hbar\omega$), une quantité de mouvement et un vecteur d'onde $\mathbf{k}$.

Lors d'une transition indirecte, la conservation de la quantité de mouvement s'écrit :

$$\hbar\mathbf{k}_{photon} + \hbar\mathbf{k}_i = \hbar\mathbf{k}_f + \hbar\mathbf{k}_{phonon}$$

Un photon est donc absorbé lors de la transition énergétique de l'électron de la bande de valence à la bande de conduction, et, *via* le champ électrique de l'onde qui joue le rôle **d'excitateur de modes**, un phonon est crée (*i.e.* un mode propre supplémentaire s'ajoute aux vibrations du réseau cristallin). Ce phonon emporte bien entendu une partie de l'énergie de l'onde qui, pour une transition directe, aurait été disponible entièrement pour l'électron. Deux conséquences importantes :

- le vecteur d'onde d'un phonon étant du même ordre de grandeur que celui d'un électron (d'ailleurs pour les mêmes raisons liées à la taille de la première zone de Brillouin...), la quantité de mouvement de l'électron peut être modifiée lors du choc (cf. figure 6.2, cas (b)) ;

[2] En fait, ces pseudo-particules sont intimement liées à la modélisation quantique des vibrations du cristal, et traduisent la représentation de ces modes de vibration sous forme à la fois corpusculaire et ondulatoire, cf. TD 2.

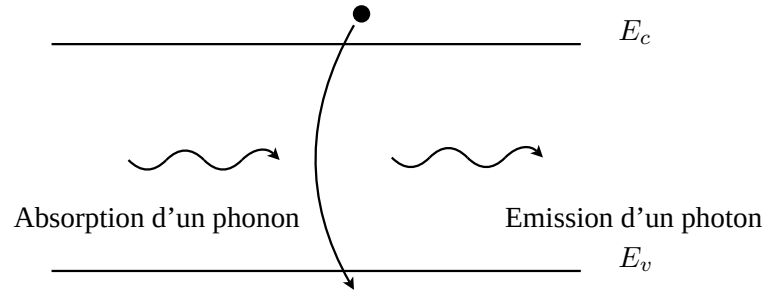


FIG. 6.3 – Transition indirecte accompagnée de l'absorption d'un phonon et d'une émission radiative. C'est un des mécanismes fondamentaux du rayonnement thermique.

– il y a augmentation des vibrations du réseau cristallin lors de la transition ; or, ces vibrations sont totalement décorrélées entre elles, et forment une part essentielle de l'agitation thermique [3] : il y a donc cession de chaleur de l'onde électromagnétique vers le réseau cristallin, c'est un des mécanismes fondamentaux du **rayonnement thermique** (le mécanisme inverse : absorption de phonon et émission de photon, existe également, cf. figure 6.3, le rayonnement thermique étant en réalité le fruit d'un équilibre entre ces deux mécanismes).

Pour mémoire, signalons également qu'il existe d'autres types de transitions interbandes :

– les transitions directes dites "interdites" : elles ont lieu entre des points du diagramme non situés sur des extréma de bandes, et sont caractérisées par un coefficient d'absorption variant en $(\hbar\omega - E_g)^{3/2}$;
– les transitions non-radiatives, du type "Auger" : un électron subit une transition descendante de la B.C. à la B.V., et transfère l'énergie ainsi libérée à un autre électron de la B.C., voire à un trou de la B.V. : ces transitions interviennent dans les processus de génération-recombinaison étudiés dans la suite de ce chapitre.

Enfin, précisons que les photons ne sont pas les seules particules capables de provoquer des transitions énergétiques interbandes. Toute particule peut provoquer une transition à condition (notamment) qu'elle soit suffisamment énergétique pour "passer le gap" : c'est le cas des rayonnements ionisants, par exemple les particules produites dans les accélérateurs de particules (protons et neutrons rapides, rayonnements radioactifs α, β et γ, particules élémentaires) ou des particules qui peuplent l'espace interstellaire et résultent des réactions très énergétiques ayant lieu dans les étoiles. Il est donc naturel que, depuis quelques décennies, le Silicium bien sûr, mais aussi d'autres semi-conducteurs comme CdTe, InP ou AsGa, soient à la base des détecteurs de particules utilisés en instrumentation nucléaire comme en astrophysique.

6.2 Processus de génération-recombinaison

6.2.1 Approche qualitative

Lorsqu'on éclaire un semi-conducteur, l'absorption de photons par le matériau provoque simultanément l'accroissement des populations d'électrons dans la B.C. et de trous dans la

[3] Le rôle des électrons est, à température ambiante, négligeable dans cette agitation, du fait de leur faible masse notamment.

B.V. : on dit qu'il y a **génération de porteurs**, chaque photon absorbé créant une transition énergétique interbande.

Un moyen de mesurer cet accroissement de population de porteurs est d'observer l'augmentation de conductivité du matériau : c'est d'ailleurs le principe des capteurs **photoconducteurs**. Or, la conductivité n'augmente pas indéfiniment : on observe assez rapidement une stabilisation à une valeur seuil pratiquement **proportionnelle** au flux incident. L'origine de cette stabilisation est le phénomène de **recombinaison**, au cours duquel un électron de la B.C. et un trou de la B.V. se recombinent (*i.e.* un électron de la B.C. "redescend" dans la B.V.) en émettant généralement un photon et/ou un phonon.

Séparons l'origine physique des deux phénomènes :

- la génération de porteurs est due à une perturbation extérieure, ici l'illumination du matériau : le taux de génération de porteurs est proportionnel au flux incident ;
- la recombinaison traduit quant à elle la **tentative de retour à l'équilibre** thermodynamique du matériau : le taux de recombinaison est d'autant plus élevé que le semiconducteur est **éloigné** de l'équilibre.

En somme, l'évolution de la valeur de la conductivité résulte d'une compétition entre ces deux effets antagonistes : lorsqu'on illumine le matériau, il y a tout d'abord génération de porteurs (électrons et trous), puis, au fur et à mesure que le système s'éloigne de l'équilibre, le mécanisme de recombinaison prend de l'importance ; la stabilisation a lieu lorsque les deux effets se compensent (*ce qui ne signifie pas qu'on se trouve à nouveau à l'équilibre thermodynamique : tant que la perturbation perdure, le système est fondamentalement hors-d'équilibre !*). Lorsque la perturbation cesse (ici, on cesse d'illuminer le semiconducteur), seule la recombinaison demeure, le temps que l'équilibre soit à nouveau atteint (cf. figure 6.4).

En fait, la tentative de retour à l'équilibre est un mécanisme très général, indépendant du type de perturbation, et même de l'existence d'une perturbation : lorsqu'un semi-conducteur est éloigné de l'équilibre thermodynamique, pour quelque raison que ce soit, un mécanisme de **génération-recombinaison** se met en place, le choix entre *générations* et *recombinaisons* dépendant du fait que les populations de porteurs sont en **défaut** ou en **excès** par rapport aux concentrations à l'équilibre. Ainsi, si dans une région d'un semi-conducteur, les populations de porteurs sont très faibles (c'est le cas dans la **Zone de Charge d'Espace** d'une diode — cf. Sec. 8.1.3), il y a génération de porteurs : cette génération est simplement due à l'agitation thermique, et est appelée **génération thermique**.

6.2.2 Taux de génération-recombinaison

A l'équilibre thermodynamique, le produit np vaut n_i^2. Lorsque on illumine le matériau, n et p augmentent, et le produit np devient supérieur à sa valeur à l'équilibre.

Notons n_0 et p_0 les concentrations de porteurs à l'équilibre, et $\hat{n}$ et $\hat{p}$ les écarts par rapport à l'équilibre :

$$n = n_0 + \hat{n}$$

$$p = p_0 + \hat{p}$$

On rappelle que, **à l'équilibre** :

- $n_0 = p_0 = n_i$ dans le cas d'un semi-conducteur pur (*i.e.* intrinsèque) ;
- $n_0 \sim N_D$ et $p_0 = n_i^2/N_D$ dans le cas d'un semi-conducteur dopé de type N ;
- $p_0 \sim N_A$ et $n_0 = n_i^2/N_A$ dans le cas d'un semi-conducteur dopé de type P.

Considérons tout d'abord le mécanisme de recombinaison directe (aussi appelée **recombinaison bande-à-bande**), au cours duquel un électron de la B.C. se recombine directement

avec un trou de la B.V. : lors de ce processus, une recombinaison fait simultanément disparaître un électron et un trou. Essayons d'évaluer le **taux de recombinaison** défini comme le taux de paires se recombinant par seconde et par unité de volume du matériau. Le taux d'électrons se recombinant par seconde est proportionnel, d'une part au nombre d'électrons présents (chaque électron ayant la même probabilité de se recombiner), d'autre part au nombre de trous de la B.V. (la probabilité qu'un électron se recombine étant proportionnelle au nombre de places vacantes dans la B.V.). On peut, en toute généralité, écrire le taux de recombinaison R sous la forme :

$$R = R_n = R_p = Anp$$

où R dépend exclusivement du matériau (les taux de recombinaison étant, pour ce mécanisme, identiques pour les trous — R_n — et les électrons — R_p, on notera les deux taux par R).

Notons $G(T)$ le taux de **génération thermique** : ce taux est proportionnel à la température, puisque c'est l'agitation thermique qui est à l'origine des générations de paires électron-trou, chaque génération de paire faisant apparaître **simultanément** un électron et un trou ; en revanche, il doit être indépendant des concentrations en porteurs de charge.

L'évolution du nombre de porteurs est alors indiquée par l'équation suivante (on suppose ici que les concentrations sont homogènes dans le matériau, afin de ne pas avoir à traiter *en sus* les problèmes de diffusion) :

$$\frac{dn}{dt} = \frac{dp}{dt} = G(T) - Anp = -U$$

où U est appelé **taux net de génération/recombinaison**. Si U est positif il y a effectivement recombinaison, tandis que si U est négatif la génération l'emporte. Avec les notations précédentes, on a donc

$$\frac{d\hat{n}}{dt} = \frac{d\hat{p}}{dt} = G(T) - A(n_0 + \hat{n})(p_0 + \hat{p})$$

En régime permanent, les concentrations de porteurs sont indépendants du temps, d'où $U = 0$ et par conséquent

$$G(T) = An_0p_0 = An_i^2$$

ce qui montre bien que $G(T)$ ne dépend que de la température (et des caractéristiques du matériau, comme le montre l'expression de n_i vue au chapitre 5). On peut donc réécrire l'équation précédente sous la forme :

$$\frac{d\hat{n}}{dt} = \frac{d\hat{p}}{dt} = An_0p_0 - A(n_0 + \hat{n})(p_0 + \hat{p}) \sim -An_0\hat{p} - Ap_0\hat{n}$$

On constate bien sur cette équation que la variation du nombre de porteur est, conformément aux signes *moins* figurant au second membre, telle que le matériau tente de rejoindre l'équilibre thermodynamique. En régime permanent, l'équilibre est atteint puisque les écarts $\hat{p}$ et $\hat{n}$ sont nuls.

En présence d'une illumination du semi-conducteur caractérisée par un taux de création de porteurs noté g (g est directement proportionnel au flux lumineux incident, mais doit tenir compte du rendement quantique, de l'état de surface du matériau, des absorptions qui ne conduisent pas à la création d'une paire mais à celle d'un phonon, etc...), l'équation précédente se réécrit :

$$\frac{d\hat{n}}{dt} = \frac{d\hat{p}}{dt} = -An_0\hat{p} - Ap_0\hat{n} + g$$

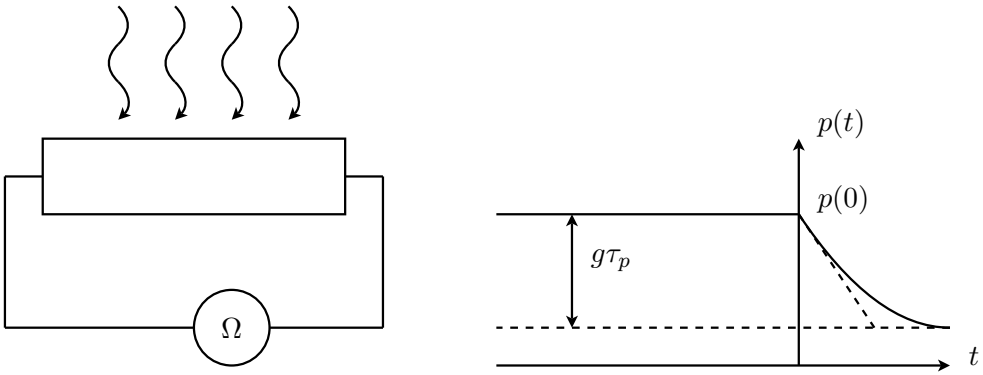


FIG. 6.4 – Évolution de la population de trous pendant et après l'illumination, mettant en évidence le rôle joué par la durée de vie des trous en excès dans le temps de retour à l'équilibre.

Cette fois, on observe qu'en régime permanent, les concentrations de porteurs sont excès :

$$An_0\hat{p} + Ap_0\hat{n} = g$$

et en particulier, sont proportionnelles au flux incident.

6.2.3 Durée de vie des porteurs en excès

Lorsque cesse l'illumination ($g = 0$), le système retourne à l'équilibre : quelle est la constante de temps de ce phénomène de **relaxation** vers l'équilibre ? Considérons un semiconducteur de type N : pour de petits écarts par rapport à l'équilibre (*i.e.* p_0, $\hat{n}$ et $\hat{p} \ll n_0 = N_D$), on peut simplifier le second membre de l'équation d'évolution des concentrations :

$$\frac{d\hat{n}}{dt} = \frac{d\hat{p}}{dt} = -An_0\hat{p} + g$$

En posant

$$\frac{1}{\tau_p} = An_0 = AN_D$$

on a :

$$\frac{d\hat{n}}{dt} = \frac{d\hat{p}}{dt} = -\frac{\hat{p}}{\tau_p} + g$$

Par conséquent, lorsque $g = 0$, les populations de trous et d'électrons retournent à l'équilibre avec **la même** constante de temps τ_p : cette constante de temps est appelée **durée de vie des porteurs en excès** et correspond à la durée de vie d'une paire *en excès* avant recombinaison. Elle est inversement proportionnelle à la population de porteurs majoritaires, donc proportionnelle à la concentration en minoritaires : un dopage élevé réduit la durée de vie des porteurs en excès.

Il faut noter que cette même durée de vie définit la **vitesse de réponse du système** en régime variable (et plus précisément, la pulsation de coupure $\omega = 1/\tau_p$), *i.e.* lorsque le flux lumineux varie dans le temps (cas d'un signal lumineux en sortie d'une fibre optique par exemple). Un dopage élevé augmente cette fréquence de coupure.

La figure 6.4 illustre l'évolution des concentrations en trous lors d'une illumination, puis lorsque la perturbation cesse. En présence d'un taux g de création de porteurs par illumination, la concentration en trou est, en régime permanent, d'après l'équation précédente, en

excès de :

$$\hat{p} = \tau_p g$$

Si la quasi-neutralité est respectée ($\hat{n} \sim \hat{p}$), la population d'électrons est aussi en excès de la même valeur : mais comme les électrons sont les porteurs majoritaires, cet excès est négligeable (sauf dans le cas d'inhomogénéités de concentrations, cf. Sec. 8.1.3, cas de la jonction longue).

Lorsque la perturbation cesse, à $t = 0$, la concentration en trou rejoint sa valeur d'équilibre p_0 avec une constante de temps τ_p : comme la conductivité suit la même loi de variation, cette expérience permet de mesurer directement la durée de vie des porteurs en excès.

6.2.4 Recombinaison par centres recombinants (ou niveaux pièges)

Dans un matériau dopé ou impur, d'autres processus de génération-recombinaison existent et entrent en concurrence (voire masquent) la recombinaison bande-à-bande. Dans de tels matériaux (cas courant en pratique, aucun matériau n'étant parfaitement pur), les recombinaisons par niveaux pièges ou **impuretés profondes** (on dit aussi **centres profonds**) sont souvent prépondérantes. Le mécanisme de **recombinaison** met en jeu deux étapes bien distinctes dans le temps (figure 6.5) :

- la capture d'un électron par un niveau d'énergie E_r situé généralement au milieu de la bande interdite et correspondant à une impureté (dans le cas d'impuretés de dopage, ces niveaux sont proches des bandes de valence ou de conduction) ;
- la capture d'un trou par la même impureté (ou, dit autrement, "l'émission" vers la B.V. de l'électron précédemment capturé par l'impureté).

Remarque : Pourquoi l'appelation "niveau piège" ? Parce qu'à un niveau d'énergie $E(k) = E_r$ situé dans une bande interdite correspond un vecteur d'onde k_r imaginaire pur : il n'y a donc pas propagation, et d'après le théorème de Bloch, la fonction d'onde de l'électron est similaire à une onde évanescente s'atténuant sur une distance $\propto 1/(ik_r)$. Un tel niveau traduit donc physiquement le "piégeage" d'un électron par une impureté ou un défaut cristallin.

Le mécanisme de génération par centres profonds est exactement l'inverse du précédent : émission d'un électron vers la B.C. par un centre profond, suivie de l'émission d'un trou vers la B.V.

Après la capture d'un électron par l'impureté, il se peut (tout dépend du type d'impureté et de la position du niveau d'énergie) que l'électron en question soit ré-émis vers la B.C. au lieu de capturer un trou : on parle dans ce cas de **piège lent**. Bien entendu, il n'y a pas recombinaison dans ce cas-là.

Ces transitions sont très généralement radiatives, les radiations émises/absorbées étant bien entendu de longueur d'onde beaucoup plus élevées que pour les recombinaisons bande-à-bande, puisque les niveaux d'énergie en jeu sont beaucoup plus proches. Le fait que ces mécanismes soient prépondérants par rapport aux processus bande-à-bande est principalement dû au fait que la probabilité de transition radiative lors d'une recombinaison bande-à-bande décroît très rapidement lorsque l'énergie du photon dépasse le seuil E_g : le fait de réaliser, grâce au centre profond, la recombinaison (ou la génération) en deux étapes de seuil énergétique beaucoup plus faible peut rendre celle-ci beaucoup plus probable.

L'étude quantitative des recombinaisons par centres profonds (durées de vie, taux de recombinaisons) est sensiblement plus complexe que l'étude des recombinaisons directes. On note dans la suite n_r la concentration en centres recombinants, et f_r la probabilité d'occupation d'un centre recombinant (*i.e.* la valeur en $E = E_r$ de la fonction de Fermi-Dirac $f(E)$).

On peut envisager quatre transitions possibles entre ce niveau E_r et les bandes de valence et de conduction, transitions qui sont représentées sur la figure 6.5 :

- la capture d'un électron de la bande de conduction par le centre recombinant (processus 1) ; ce taux de capture d'électron est proportionnel à la concentration n en électrons de la bande de conduction et au nombre $n_r(1 - f_r)$ de centres non-occupés par des électrons :
$$R_n = C_n\, n\, n_r\, (1 - f_r),$$
f_r étant la probabilité pour qu'un électron occupe un centre recombinant et C_n un coefficient dépendant du matériau ;
- l'émission d'un électron par le centre vers la bande de conduction (processus 2). Ce taux d'émission est proportionnel à la concentration en centres occupés $n_r\ f_r$, et aux places disponibles dans la bande de conduction ; cette dernière étant presque vide, on peut considérer que la concentration en places disponibles est sensiblement constante et n'intervient qu'à travers le coefficient E_n dépendant du matériau :
$$G_n = E_n\, n_r\ f_r,$$
- la capture d'un trou par le centre (processus 3), symétrique du processus 1. Le taux de capture de trous est proportionnel aux nombres de centres occupés $n_r\ f_r$ (donc susceptibles d'accueillir un trou), et au nombre de trous dans la bande de valence p :
$$R_p = C_p\, n_r\ f_r p$$
- l'émission d'un trou vers la bande de valence (processus 4), processus symétrique du processus 2. Le taux d'émission de trous est proportionnel au nombre de centres non-occupés $n_r(1 - f_r)$, puisque la concentration en trous de la bande de valence est très faible et peut-être considérée comme constante :
$$G_p = E_p n_r(1 - f_r),$$
E_p étant, comme ci-dessus, un coefficient dépendant du matériau.

Expression du taux net de génération/recombinaison. En régime stationnaire ou bien à l'équilibre thermodynamique (en l'absence de perturbation), les concentrations en trous et en électrons ne dépend pas du temps, et on en déduit :

$$R_n = G_n \Rightarrow U_n = R_n - G_n = 0$$

et

$$R_p = G_p \Rightarrow U_p = R_p - G_p = 0$$

De plus la probabilité d'occupation d'un centre est donnée par la fonction de Fermi-Dirac :

$$f_r = \frac{1}{1 + e^{\frac{E_r - E_F}{kT}}}$$

Hors de l'équilibre et en régime stationnaire, les concentrations en centres vides ou occupés sont constantes (séparément), ce qui implique :

$$R_n + G_p = G_n + R_p$$

En utilisant ces résultats on pourrait montrer, que dans le cas de centres recombinants situés au milieu de la bande interdite et jouant un rôle symétrique pour les électrons et pour les trous, le taux net de génération recombinaison est donné par la relation :

$$U = C n_r \frac{pn - n_i^2}{p + n + 2n_i} \tag{6.1}$$

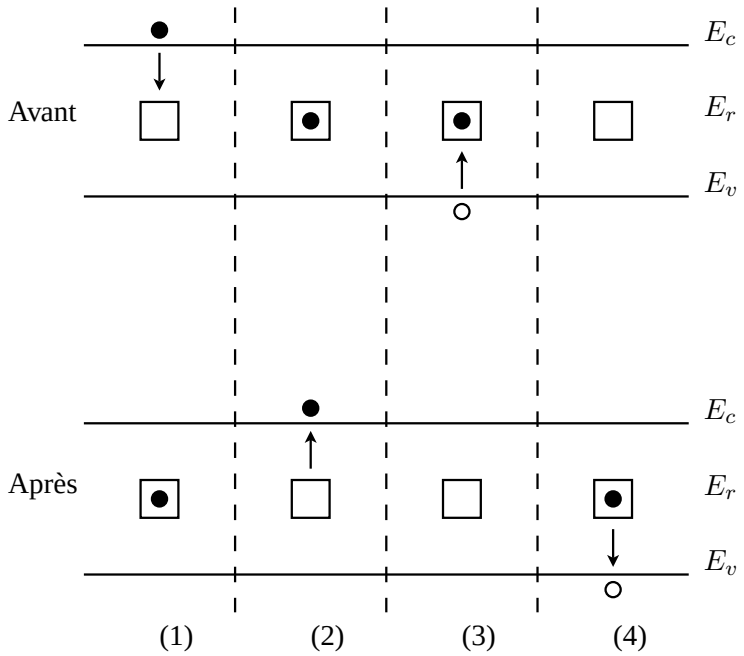


FIG. 6.5 – Différents processus de recombinaison par centres recombinants ou centres profonds : (1) capture d'un électron, (2) émission d'un électron, (3) capture d'un trou, (4) émission d'un trou. Il importe de noter que, contrairement à la recombinaison directe, aucun de ces processus ne fait simultanément disparaître un électron et un trou. E_t est le niveau piège correspondant à l'impureté.

Deuxième partie

Composants

Chapitre 7

Du matériau homogène au composant

7.1 Introduction

Le but de la physique des composants est de fournir des modèles pour les composants à semi-conducteur, et plus largement, la multitude de matériaux utilisés aujourd'hui. Ces modèles sont notamment utilisés pour :

- représenter le comportement externe du composant afin de concevoir des circuits électroniques, c'est la fonction des schémas équivalents ou des modèles internes des logiciels de Conception Assistée par Ordinateur (Spice,...),
- étudier l'influence de la structure interne du composant sur ses performances en vue d'améliorer ces dernières.

Les caractéristiques externes sont les courants et les tensions aux bornes du composant, auquels s'ajoutent les flux lumineux reçus ou émis dans le cas de composants optoélectroniques, la polarisation magnétique du courant dans le cas de composants spintroniques... Les évolutions de ces caractéristiques externes sont liées à l'évolution interne des concentration de porteurs (et de leur état de spin dans le cas de la spintronique) ; la physique des composants sert à établir ces liens.

7.2 Diagramme des bandes en présence d'un potentiel électrostatique "macroscopique" ou diagramme d'Anderson

On a vu au chapitre 2 que les niveaux d'énergie permis pour les électrons se regroupaient en un certain nombre de bandes. Seules les bandes presque vides d'électrons et les bandes presque pleines participent aux phénomènes de transport de charges, c'est ainsi que les propriétés conductrices des cristaux sont déterminées par les deux bandes d'énergie supérieures : la bande de valence et la bande de conduction. Ces deux bandes sont séparées par une bande interdite appelée gap.

En présence d'un champ électrique faible et lentement variable de nature macroscopique (polarisation externe), il existe en chaque point d'abscisse x du composant une énergie potentielle électrostatique $-qV(x)$ additionnelle (une inhomogénéité du matériau due au dopage produirait indirectement le même effet, voir l'explication plus bas). En toute rigueur, il faudrait recalculer les niveaux d'énergie en présence de ce potentiel additionnel, mais si $V(x)$ varie lentement dans l'espace et dans le temps, on peut appliquer une approximation dite **semi-classique** et supposer que cette énergie potentielle s'ajoute simplement à l'énergie cristalline que possède déjà l'électron dans le matériau : l'énergie totale est alors donnée par

$E(k) - qV(x)$, où $E(k)$ est le niveau d'énergie cristallin occupé par l'électron *en l'absence de champ électrique*.

> **Remarque :** La validité de cette approximation s'appuie sur le fait que les fonctions d'onde électronique (en fait, les paquets d'onde) sont suffisamment localisés pour que l'on puisse parler de niveaux d'énergie **à une abscisse** x particulière, et donc appliquer une approximation semi-classique. En effet, le principe d'incertitude d'Heisenberg stipule qu'on ne peut pas simultanément mesurer la position x et la quantité de mouvement $p = \hbar k$, et par conséquent, écrire quelque chose comme $E = E(k) - qV(x)$ serait en toute rigueur un non-sens... Il faut cependant noter qu'une telle localisation était déjà sous-entendu dans le chapitre 3 lorsqu'on a introduit la notion de masse effective. Il faut donc se souvenir que dans cette dernière expression, x et k sont en réalité affectés d'une incertitude dont le produit est au moins supérieur à 1.

Le diagramme d'énergie résultant (figure 7.1) est appelé **diagramme d'Anderson** : c'est un diagramme des bandes qui représente désormais l'énergie E en fonction de x et de k (cf. remarque ci-dessus pour le problème des relations d'incertitude d'Heisenberg), là où le diagramme des bandes du chapitre 2 se contentait de E en fonction du vecteur d'onde k. Comme une telle représentation est délicate, on ne trace en réalité que quelques niveaux caractéristiques.

On représente en particulier E_c, minimum de la bande de conduction, et E_v, maximum de la bande de valence. Sur le schéma de la figure 7.1 on voit apparaître en plus de ces deux niveaux :

- Le niveau "du vide" E_0, qui est le niveau d'énergie d'un électron ayant vaincu les forces d'attraction cristallines. Ce niveau est séparé du bas de la bande de conduction par **l'affinité électronique** χ, énergie qu'il faut fournir à un électron du bas de la bande de conduction pour l'extraire du cristal (cf. effet photoélectrique). Attention à ne pas confondre l'affinité avec le travail de sortie W_s, ce dernier étant une énergie moyennée sur l'ensemble des électrons de la bande de conduction.
- Le niveau de référence $E = 0$, qui correspond à l'énergie de l'électron au repos à l'infini (hors de l'influence des forces électrostatiques).

La présence d'un potentiel électrostatique macroscopique (*i.e.* non-cristallin) provoque le "déplacement" vertical de tous ces niveaux d'énergie par rapport au niveau de référence $E = 0$: on dit que les bandes sont **déformées** par le potentiel électrostatique.

Examinons tout d'abord la situation à l'équilibre thermodynamique : il n'y a alors pas de courant électrique net circulant dans la structure, et le niveau de Fermi E_F est plat. Or, comme on le constate sur la figure 7.1, la présence d'un potentiel électrostatique $V(x)$ implique que la distance $E_c(x) - E_F$ n'est plus constante. Comme la concentration en électrons dépend de $E_c - E_F$ (équation 4.4), elle dépend donc désormais, de facto, de x :

$$n(x) = N_c e^{E_F - E_c(x)}$$

et de même pour $p(x)$, concentration en trous en fonction de x. Ceci induit donc un courant de diffusion à cause du gradient de concentration : pour une géométrie unidimensionnelle, ce courant vaut

$$J_{Dn} = qD_n \frac{dn(x)}{dx} = \frac{-qD_n}{kT} \frac{dE_c(x)}{dx} n(x)$$

soit encore, en invoquant la relation d'Einstein (Eq. 5.8),

$$J_{Dn} = n(x) q \mu_n \frac{dV(x)}{dx} = -n(x) q \mu_n \mathcal{E}$$

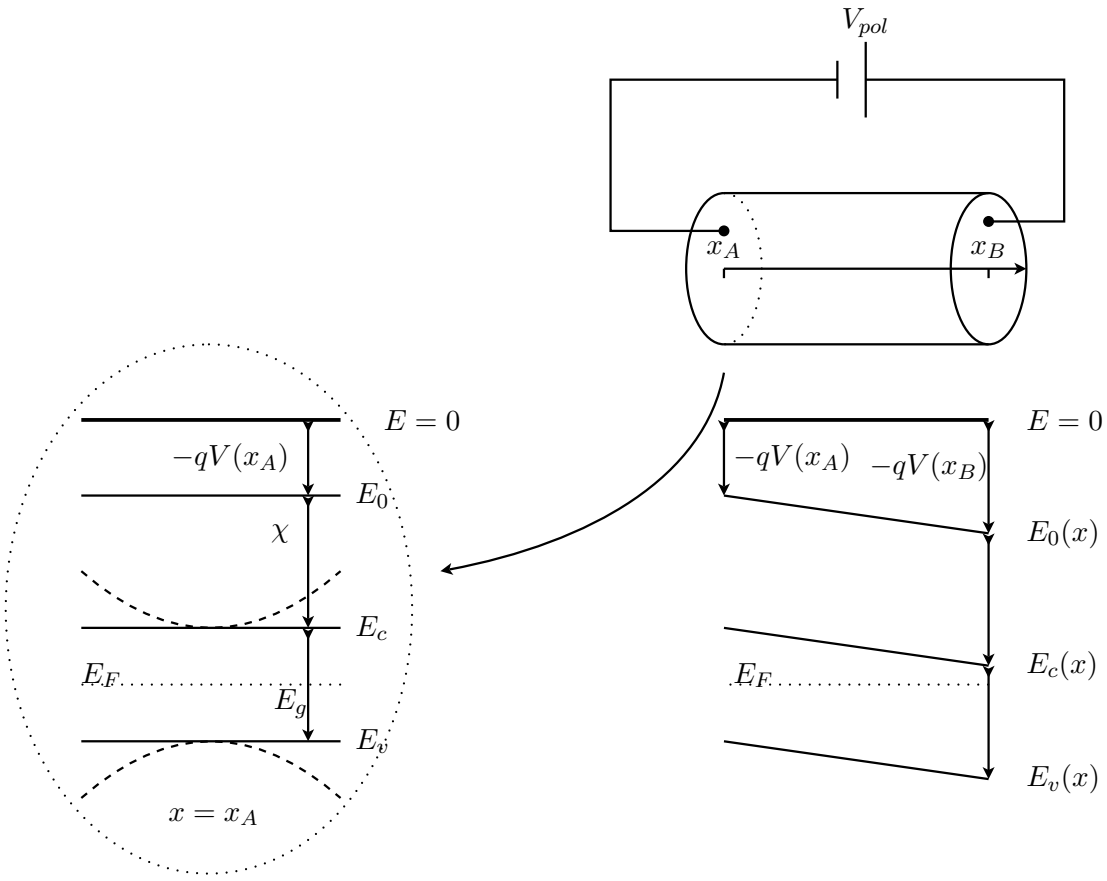


FIG. 7.1 – Diagramme dit "d'Anderson" des bandes d'énergie dans un matériau semiconducteur de géométrie unidimensionnelle soumis à une polarisation externe (ou à un dopage inhomogène, ce qui produirait le même diagramme). A gauche, diagramme des bandes à l'absisse x_0 (traits-tirets : allure des bandes de conduction et de valence à proximité des extremas) ; χ représente l'affinité électronique du matériau (proche, mais néanmoins distincte du travail de sortie W_s). A droite, les niveaux d'énergie caractéristiques en fonction de l'abscisse x avec un potentiel électrique noté $V(x)$: $E = 0$ (niveau de référence, *i.e.* , électron à l'infini), E_0 (électron hors du cristal mais encore soumis au potentiel électrostatique de polarisation électrique $V(x)$), E_c minimum de la bande de conduction, E_F niveau de Fermi (plat à l'équilibre thermodynamique, "penché" lorsqu'un courant circule), E_v maximum de la bande de valence.

Si maintenant on calcule le courant de conduction induit par le champ électrique dérivant de $V(x)$, on obtient :

$$J_{Cn} = n(x)q\mu_n\mathcal{E}$$

c'est-à-dire exactement l'opposé de J_{Dn} : le courant total est donc nul, ce qui est cohérent avec notre argument initial selon lequel le niveau de Fermi est plat à l'équilibre ! Bien entendu, une telle situation où il n'existe aucun courant net ne peut arriver en pratique que si le matériau n'est pas branché à un circut extérieur... auquel cas $V(x)$ est constant, ce qui n'est pas très intéressant... La seule exception correspond, en fait, au cas d'une inhomogénéité dans le matériau : celle-ci est en effet capable de provoquer indirectement un gradient de potentiel électrostatique,

- soit parce que le dopage est inhomogène : dans ce cas, $n(x) = N_D(x)$ par exemple (dopage de type N en concentration N_D), et $E_c(x) - E_F$ épouse la forme dictée par la formule $N_D(x) = N_c e^{E_F - E_c(x)}$; du coup, si $E_c(x)$ varie en fonction de x, $V(x)$ en fait de même... et il y a bien un potentiel électrostatique induit. A la limite (cf. section suivante), une jonction PN est un cas "abrupt" de dopage inhomogène où l'on passe brutalement d'un dopage N à un dopage P, et il y a bien, là aussi, un potentiel électrostatique induit bien qu'aucun courant ne circule tant que le composant n'est pas branché à un circuit extérieur.
- soit parce qu'il existe une perturbation extérieure inhomogène, par exemple le fait d'éclairer une seule des deux faces du composant (effet Dember) ; la conséquence est semblable au dopage, car le flux lumineux inhomogène donnera naissance à un gradient de concentration $n(x)$, et donc à un gradient de potentiel électrostatique selon le même mécanisme que précédemment.

En conclusion de cette partie, il est important pour la suite de vous entraîner à visualiser mentalement le diagramme des bandes d'Anderson comme un outil indiquant les variations de concentrations en électrons et en trous, *i.e.* , selon que les extrema de bandes E_c et E_v s'éloignent ou se rapprochent du niveau de Fermi. Il ne faudra jamais oublier que ce niveau de Fermi étant plat à l'équilibre (pas de branchement à un circuit extérieur), des variations de concentration induisent indirectement un potentiel électrostatique inhomogène, donc un champ électrostatique, ce qui permet au courant de diffusion d'être toujours équilibré par un courant de conduction. Notez qu'hors d'équilibre en revanche, un courant circule et le niveau de Fermi n'est plus plat (cf. pseudo-niveau de Fermi du chapitre "transport").

7.3 Les jonctions

Le contact ou la succession de deux matériaux est appelé une jonction, et c'est la succession de jonctions qui forme les composants électroniques[1]. On distingue les homojonctions qui sont la succession de deux matériaux identiques de dopages différents (la jonction PN en est l'exemple type), et les hétérojonctions qui sont la succession de deux matériaux différents (la jonction Schotky par exemple).

7.3.1 Schéma des bandes d'énergie d'une homojonction PN

On obtient le schéma des bandes d'une jonction PN en procédant de la manière suivante :

- on trace le schéma des bandes des deux régions P et N isolées (*cf.* figure 7.2),
- on fusionne ces deux schémas en alignant les niveaux de Fermi (celui-ci est constant à l'équilibre), et en faisant subir au niveau du vide E_0 une variation continue entre les

[1]Les composants de la spintronique sont construits à partir de jonctions entre matériaux ferromagnétiques.

deux cotés de la jonction (cette variation est continue si l'on suppose qu'il n'existe pas de distribution surfacique de charges) ; les niveaux d'énergie E_c et E_v se déduisent du niveau E_0 par des translations successives de valeurs $q\chi$ et E_g (car ces deux grandeurs dépendent seulement de la nature du matériau et non du dopage) ; on obtient alors le schéma de la figure 7.3 qui reste encore qualitatif dans la direction des abscisses.

◇ Commentaires Ce diagramme montre qu'il existe une différence de potentiel entre les régions N et P notée ϕ, et appelée tension de diffusion. L'énergie $q\phi$ qui lui correspond est appelée barrière de potentiel interne, elle se retrouve sur les variations de E_c et de E_v ; c'est la barrière de potentiel qui s'oppose au passage des porteurs majoritaires vers les régions où ils sont minoritaires. Les porteurs minoritaires ne sont pas concernés par cette barrière. Nous verrons plus loin que cette barrière est modifiée par la polarisation extérieure : si la jonction est polarisée en inverse cette barrière est renforcée et seul un courant de minoritaires (donc très faible) peut transiter au travers de la jonction. Si la jonction est polarisée en direct, la barrière vue par les majoritaires est réduite et le courant augmente rapidement. L'effet redresseur peut donc être pressenti en observant le schéma des bandes.

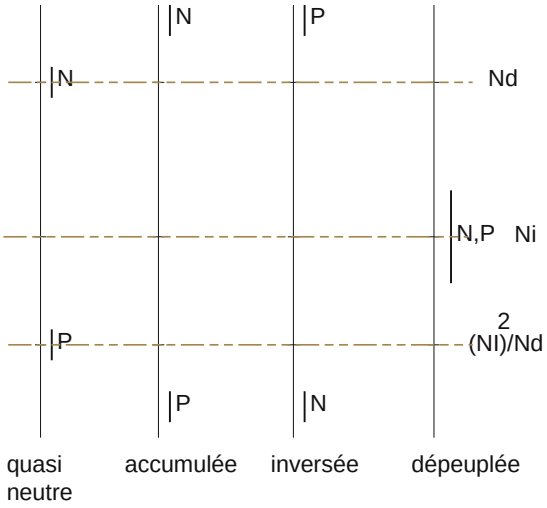


FIG. 7.2 – Schéma des bandes des régions P et N séparées

7.3.2 Étude quantitative de la jonction PN a l'équilibre

La figure 7.4 reprend le schéma des bandes d'une jonction PN et indique toutes les notations que nous utiliserons. On raisonnera sur une jonction abrupte, c'est à dire une jonction où le dopage qui est constant et égal à N_A du coté P, passe brutalement à la valeur N_D du coté N.

En examinant ce schéma des bandes, on s'aperçoit qu'une jonction PN peut être scindée

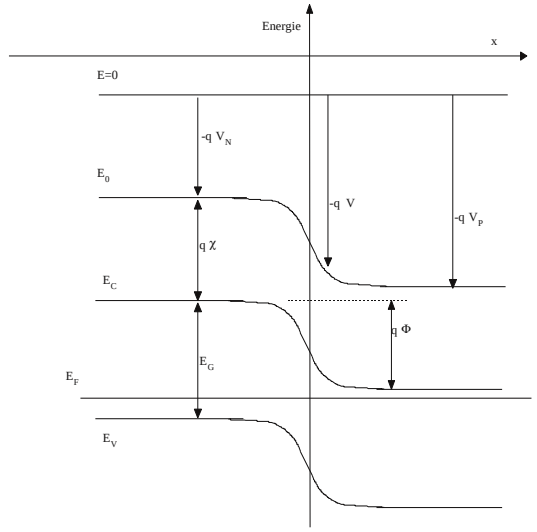


FIG. 7.3 – Schéma des bandes d'une jonction PN

en trois régions distinctes : deux régions quasi neutres et une région dépeuplée.

• La région dépeuplée se situe de part et d'autre de la jonction métallurgique (endroit où le dopage passe de N_A à N_D), en effet les expressions des concentrations de porteurs

$$n = N_c e^{-\frac{E_c - E_F}{kT}} \quad \text{et} \quad p = N_v e^{-\frac{E_F - E_v}{kT}}$$

montrent que dans cette région d'extension W_T les valeurs N et P sont toujours très inférieures au concentrations des dopants. La densité des charges mobiles $q(p-n)$ est très inférieure à la densité des charges fixes $q(N_D - N_A)$: et l'équation de Poisson se réduit à :

$$-\frac{d^2V}{dx^2} = \frac{dE}{dx} = q(N_D - N_A)$$

les valeurs du potentiel et du champ électrique varient alors d'autant plus rapidement que le dopage est élevé. Ainsi, une région dépeuplée de 1 micron dans du silicium dopé à 10^{17}cm^{-3} entraînerait une variation de champ électrique de $1.6\,10^6$ V/cm, et une variation de potentiel de 80 Volts.

• Loin de la jonction métallurgique ces mêmes expressions nous montrent que les concentrations de porteurs sont identiques à celles rencontrées dans deux semi-conducteurs distincts N et P. On peut donc délimiter une région quasi neutre P où le potentiel est constant et les concentrations de porteurs données par :

$$p \approx N_A \quad \text{et} \quad n \approx \frac{n_i^2}{N_A}$$

et une région quasi neutre N où :

$$n \approx N_D \quad \text{et} \quad p \approx \frac{n_i^2}{N_D}$$

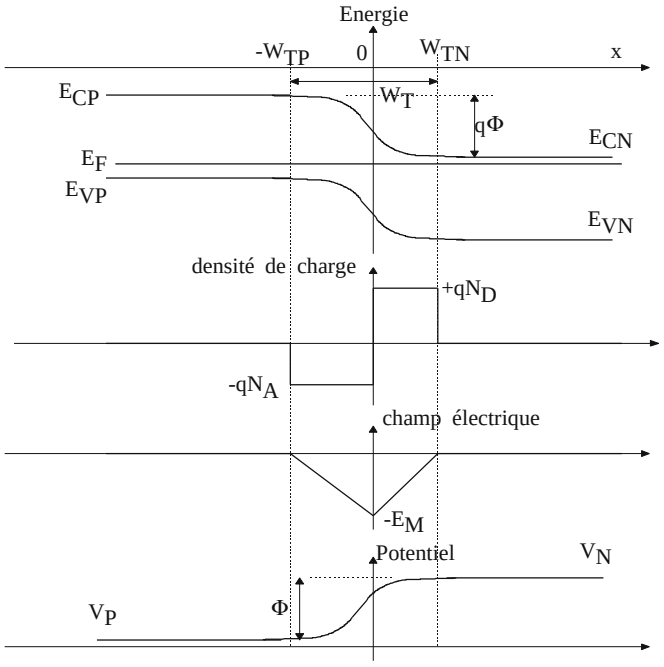


FIG. 7.4 – Schéma des bandes, densité de charges et champ électrique dans une jonction PN abrupte

Dans les régions quasi neutres la densité de charge totale $q(p - n + N_D - N_A)$ est très inférieure à la densité des charges fixes $q(N_D - N_A)$

◇ Calcul de la tension de diffusion ϕ Comme l'indique le schéma des bandes la tension de diffusion à pour expression :

$$\phi = \frac{E_{CP} - E_{CN}}{q}$$

avec, coté N, $n \approx N_D$ et par conséquent

$$E_{CN} = E_F + kT \ln \frac{N_C}{N_D}$$

et d'autre part coté P, $n \approx \frac{n_i^2}{N_A}$ soit

$$E_{CP} = E_F + kT \ln \frac{N_A N_C}{n_i^2}$$

en reportant dans l'expression de ϕ on trouve :

$$\phi = \frac{kT}{q} \ln \frac{N_A N_D}{n_i^2} \tag{7.1}$$

Dans du silicium dopé avec $N_A = 10^{18}\text{cm}^{-3}$ et $N_D = 10^{16}\text{cm}^{-3}$, valeurs typique pour une jonction PN, on trouve une tension de diffusion de l'ordre de 800mV à la température ambiante.

> **Remarque :** Cette tension de diffusion n'est pas directement mesurable, car une jonction PN réelle doit être entourée de part et d'autre par deux contacts dits ohmiques pour assurer les connections avec le circuit extérieur. Ces contacts ohmiques sont des hétérojonctions et, en traçant le schéma des bandes de la structure globale, on s'aperçoit que la différence des niveaux du vide entre les deux extrémités est nulle et donc qu'il n'existe pas de d.d.p. aux bornes de la structure complète.

◇ Calcul de la profondeur de la région dépeuplée Dans la région dépeuplée les porteurs libres très raréfiés découvrent les impuretés ionisées. la charge d'espace ainsi crée est représentée sur la figure 7.4. En prenant l'origine des abscisses à la jonction métallurgique, et en considérant le champ électrique nul dans les régions quasi neutres, on obtient en intégrant l'équation de Poisson coté P :

$$\frac{dE}{dx} = -\frac{q}{\epsilon}N_A$$

et par conséquent

$$E = c^{te} - \frac{q}{\epsilon}N_A x$$

et si $-W_{TP}$ est la frontière de la région dépeuplée coté P, et en considérant que $E(-W_{TP}) = 0$, on obtient :

$$E(x) = \frac{q}{\epsilon}N_A(-W_{TP} - x)$$

coté P. Les mêmes hypothèses conduisent à l'expression :

$$E(x) = \frac{q}{\epsilon}N_D(x - W_{TN})$$

côté N.

La figure 7.4 montre l'allure des variations du champ. La continuité du champ à la jonction se traduit par :

$$-E(0) = E_M = \frac{q}{\epsilon}N_A W_{TP} = \frac{q}{\epsilon}N_D W_{TN}$$

ce qui impose les relations suivantes entre les extensions des régions dépeuplées :

$$\frac{W_{TP}}{W_{TN}} = \frac{N_D}{N_A}$$

ainsi la région dépeuplée s'étend préférentiellement du coté *le moins dopé* :

$$W_{TP} = \frac{N_D}{Na + N_D} \quad \text{et} \quad W_{TN} = \frac{N_A}{Na + N_D}$$

La variation de potentiel étant égale à l'intégrale du champ, celle ci est donc représentée par l'aire sous la courbe du champ sur la figure 7.4. Cette valeur correspond bien sur à la différence de potentiel ϕ existant entre la région quasi neutre N et la région quasi neutre P :

$$\phi = \frac{1}{2}E_M W_T$$

et en remplaçant E_M par sa valeur en fonction de W_T

$$E_M = \frac{q}{\epsilon}\frac{N_A N_D}{N_A + N_D}W_T$$

on obtient :

$$\phi = \frac{q}{2\epsilon}\frac{N_A N_D}{N_A + N_D} W_T^2$$

ou encore

$$\boxed{W_T = \sqrt{\frac{2\epsilon}{q}\frac{N_A + N_D}{N_A N_D}\phi}} \tag{7.2}$$

et

$$\boxed{E_M = \sqrt{\frac{2q}{\epsilon}\frac{N_A N_D}{N_A + N_D}\phi}} \tag{7.3}$$

◇ Commentaires Dans la région dépeuplée les concentrations de porteurs présentent des gradients importants qui devraient entraîner des courants de diffusion importants, en effet la concentration des trous passe de la valeur N_A à n_i^2/N_D alors que la concentration des électrons passe de N_i^2/N_A à la valeur N_D. C'est seulement l'existence du champ interne que nous venons de calculer qui s'oppose à ces courants de diffusion et qui impose un bilan nul.

7.4 Semi-conducteurs hors de l'équilibre thermodynamique

L'étude de l'équilibre n'est pas suffisante pour expliquer le comportement des composants semi-conducteurs. En effet l'application d'une polarisation extérieure peut modifier de manière considérable les caractéristiques qui prévalent à l'équilibre et ainsi, par exemple, les expressions donnant la concentration des porteurs à l'équilibre (Eq. 4.4 par exemple) ne sont plus vérifiées ; il en est de même de la constance du niveau de Fermi sur toute la structure. Seules demeurent vraies l'expression de la densité de courant :

$$J = J_n + J_p = q(D_n\frac{dn}{dx} + n\mu_n E) + q(-D_p\frac{dp}{dx} + p\mu_p E)$$

et bien entendu l'équation de Poisson. C'est donc à partir de ces équations que l'on doit bâtir tout modèle susceptible d'expliquer le comportement des composants hors équilibre.

Les concentrations des porteurs n et p sont susceptibles de varier au cours du temps ou lorsque l'on se déplace dans la structure. Les mécanismes responsables de ces variations sont les mécanismes de génération recombinaison, ainsi que la divergence des densités de flux de porteurs, mécanismes qui sont pris en compte dans les deux équations de continuité du chapitre "transport", Equ. 5.12

$$q\frac{\partial n}{\partial t} = \frac{\partial J_n}{\partial x} - U_n$$

et 5.13

$$q\frac{\partial p}{\partial t} = -\frac{\partial J_p}{\partial x} - U_p$$

Ces équations traduisent le fait qu'à un endroit donné du cristal et à un instant donné, les concentrations de porteurs diminuent comme le taux net de recombinaison et comme le gradient de la densité de flux. Un taux de recombinaison U positif signifie qu'il disparaît U porteurs par unité de volume et de temps, la variation de la concentration de porteurs par unité de temps qui en découle est donc égale à $-U$, cela correspond à une diminution de concentration. Une divergence de densité de flux positive indique qu'à un endroit donné il part plus de porteurs qu'il n'en arrive, ce qui tend à faire diminuer la concentration.

Cas des régions quasi neutres. Dans les régions quasi neutres on a l'habitude de scinder les concentrations de porteurs hors équilibre n ou p en deux contributions. La première, n_0 ou p_0, correspond à la concentration à l'équilibre. La deuxième $\hat{n}$ ou $\hat{p}$ correspond à l'excès de porteurs par rapport à l'équilibre, et seule cette contribution présente des variations temporelles.

$$p = p_0 + \hat{p} \quad \text{et} \quad n = n_0 + \hat{n}$$

L'équation de neutralité électrique entraîne avec ces notations :

$$\hat{n} = \hat{p}$$

Deux situations intéressantes se présentent en pratique :

- la faible injection où $\hat{n} = \hat{p} \ll N_D - N_A$
- la faible injection où $\hat{n} = \hat{p} \gg N_D - N_A$

Remarque : En faible injection les concentrations des porteurs majoritaires sont proches de celles de l'équilibre. Pour un type N on a :

$$n \approx n_0 = N_D \quad \text{et} \quad p \ll n$$

En haute injection les concentrations des deux type de porteurs sont très supérieures à celles des majoritaires à l'équilibre. Pour un type N on a :

$$p \approx \hat{p} \approx \hat{n} \approx n \gg n_0 = N_D$$

ces deux situations seront analysées en détail ultérieurement lorsque nous étudierons la diode.

Dans le cas où il n'existe pas de taux de génération externe (génération thermique uniquement), l'expression du taux de recombinaison dans les régions quasi neutres peut se simplifier. Dans le cas d'un mécanisme bande à bande

$$U = C^{te}(pn - n_i^2) = C^{te}[(p_0 + \hat{p})(n_0 + \hat{n}) - n_i^2]$$

et compte tenu de $p_0 n_0 = n_i^2$; pour un type N en faible injection où $n_0 = N_D$, $p_0 \ll n_0$, et $\hat{p} = \hat{n}$ à cause de la quasi neutralité :

$$U = C^{te} N_D \hat{p} = \frac{\hat{p}}{\tau_p}$$

où la constante τ_p, homogène à un temps, est appelée durée de vie des porteurs minoritaires. Dans le cas d'un SC du type P on aurait trouvé

$$U = \frac{\hat{n}}{\tau_n}$$

Dans le cas d'un mécanisme par centre recombinant :

$$U = C N_r \frac{pn - n_i^2}{p + n + 2n_i}$$

les mêmes hypothèses conduisent à :

$$U = \frac{\hat{p}}{\tau_p} \quad \text{avec} \quad \tau_p = \frac{1}{C N_r} \tag{7.4}$$

pour un type N, et à

$$U = \frac{\hat{n}}{\tau_n} \quad \text{avec} \quad \tau_n = \frac{1}{C N_r} \tag{7.5}$$

pour un type P.

En haute injection le taux de recombinaisons par centres s'écrit, compte tenu de ces hypothèses ($p = n \gg n_0$ et p_0) :

$$U = \frac{CN_r}{2\tau_{ph}}$$

où τ_{ph} est la durée de vie des porteurs en haute injection. Les équations de continuité dans une région quasi neutre peuvent donc s'écrire :

$$\frac{\partial p}{\partial t} = -\frac{\hat{p}}{\tau_p} - \frac{1}{q}\frac{\partial J_p}{\partial x} \tag{7.6}$$

pour les trous et

$$\frac{\partial n}{\partial t} = -\frac{\hat{n}}{\tau_n} - \frac{1}{q}\frac{\partial J_n}{\partial x} \tag{7.7}$$

7.4.1 Efficacite d'injection d'une jonction pn

Considérons une jonction PN abrupte délimitée par deux contacts ohmiques (*cf.* figure 7.5). Un contact ohmique est un contact métal semi-conducteur dont le métal a été choisi pour que le passage d'un courant ne modifie pas la valeur des concentrations de porteurs dans son voisinage par rapport à celles d'équilibre, et pour que la d.d.p supportée par le contact soit indépendante de la polarisation. En conséquence $pn = n_i^2$ au contact ohmique, et nous verrons plus loin que toute la d.d.p. appliquée à la structure se reporte sur le semi-conducteur.

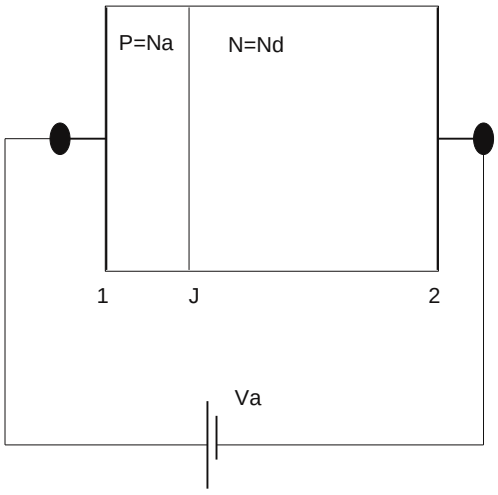


FIG. 7.5 – Jonction PN et contacts ohmiques

L'existence d'une polarisation extérieure entraîne le passage d'un courant dans la structure. En utilisant la relation d'Einstein ($D = U_T\mu$), on peut réécrire les expressions des densités de courant sans faire apparaître les mobilités :

$$J_n = qD_n(\frac{\partial n}{\partial x} + \frac{n}{U_T}E) \tag{7.8}$$

$$J_p = qD_p(-\frac{\partial p}{\partial x} + \frac{p}{U_T}E) \tag{7.9}$$

et en éliminant le champ électrique entre ces deux équations on obtient la relation :

$$\boxed{\frac{pJ_n}{qD_n} - \frac{nJ_p}{qD_p} = \frac{\partial(pn)}{\partial x}} \tag{7.10}$$

Si l'on intègre cette expression entre les deux contacts ohmiques 1 et 2 où $pn = n_i^2$ on obtient :

$$\int_1^2 \frac{pJ_n}{D_n}dx = \int_1^2 \frac{nJ_p}{D_p}dx$$

En régime statique et dans une jonction où les recombinaisons jouent un rôle négligeable l'équation de continuité montre que les densités de courant sont constantes le long de la structure, on peut donc les sortir des intégrales et calculer leur rapport appelé efficacité d'injection :

$$\frac{J_p}{J_n} = \frac{\int_1^2 \frac{p}{D_n}dx}{\int_1^2 \frac{n}{D_p}dx} = \gamma$$

En faible injection où les concentrations des porteurs majoritaires restent proches de celles de l'équilibre, on a :

$$\int_1^J pdx \approx N_A W_p \gg \int_J^2 pdx$$

$$\int_J^2 ndx \approx N_D W_n \gg \int_1^J ndx$$

où W_n et W_p sont les profondeurs des régions N et P de dopages respectifs N_D et N_A. L'expression de l'efficacité d'injection se réécrit alors :

$$\boxed{\gamma = \frac{D_p N_A W_p}{D_n N_D W_n}} \tag{7.11}$$

Dans une jonction P^+N où le dopage N_A de la région P est très supérieur au dopage N_D de la région N la densité de courant de trous est très grande devant celle des électrons et la valeur de γ est grande devant l'unité. De telles jonctions sont très utilisées dans les composants semi-conducteurs ; dans les transistors bipolaires c'est le caractère très dissymétrique de la jonction émetteur base qui permet d'obtenir un gain en courant important, car le courant de trous est fourni par le courant de base alors que le courant d'électrons est fourni par le courant de collecteur.

7.4.2 Relation courant-tension

Considérons une jonction PN abrupte entourée de deux contacts ohmiques, polarisée par une tension extérieure V_a, et calculons la différence de potentiel $V_1 - V_2$ existant entre

les points extrêmes du semi-conducteur (*cf.* figure 7.5). En partant de la relation donnant la densité de courant d'électrons :

$$J_n = q(D_n \frac{\partial n}{\partial x} + n\mu_n E)$$

et en utilisant la relation d'Einstein $D = U_T \mu$, on peut exprimer le champ électrique par la relation :

$$E = \frac{J_n}{qn\mu_n} - \frac{U_T}{n}\frac{\partial n}{\partial x} \tag{7.12}$$

en intégrant cette relation entre les points 1 et 2 où les contacts ohmiques imposent $N(1) = n_i^2/N_A$, et $N(2) = N_D$, on obtient :

$$V_1 - V_2 = \int_1^2 \frac{J_n}{q\mu_n n} dx - U_T \ln \frac{N_A N_D}{n_i^2} \tag{7.13}$$

Le premier terme de cette expression est la tension de diffusion ϕ calculée précédemment, le deuxième représente la tension appliquée à la jonction V_a. En effet si l'on considère que les contacts ohmiques sont des dispositifs qui imposent $pn = n_i^2$ dans le semi-conducteur dans leur voisinage immédiat, et qui ne subissent pas de variation de potentiel quel que soit le courant qui les traverse, on peut les considérer d'un point de vue électrique comme des sources de tension de valeur constante V_{oh1} et V_{oh2}, et donner d'une jonction PN entourée de deux contacts ohmiques la représentation suivante : On a donc

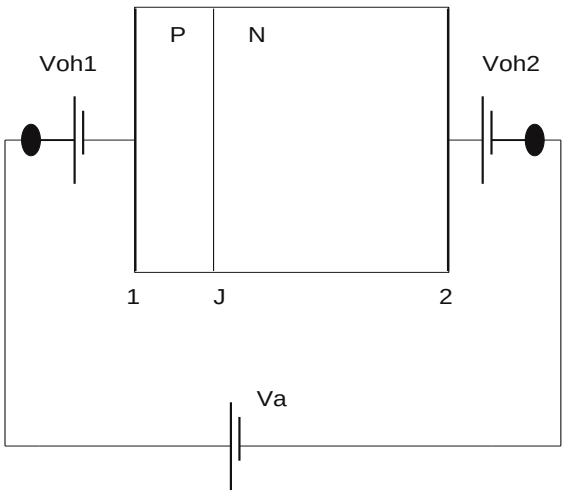


FIG. 7.6 – Jonction PN et ddp de contact ohmique

$$V_a = (V_{oh1} + V_{oh2}) - (V_2 - V_1)$$

Or à l'équilibre $V_a = 0$ et $V_2 - V_1 = \phi$, par conséquent $V_{oh1} + V_{oh2} = \phi$, et dans tous les cas la différence de potentiel entre le coté N et le coté P vaut :

$$V_n - V_p = V_2 - V_1 = \phi - V_a \tag{7.14}$$

ce qui donne, compte tenu des expressions 7.13 et 7.14 :

$$\boxed{V_a = \int_1^2 \frac{J_n}{qn\mu_n} dx} \tag{7.15}$$

Remarque : L'expression 7.14 montre que la polarisation extérieure V_a modifie la d.d.p. $V_2 - V_1$ supportée par la jonction :

$$V_2 - V_1 = \phi - V_a$$

une polarisation directe ($V_a > 0$) diminue donc la hauteur de barrière de la jonction.

7.4.3 Approximation de boltzmann

En divisant les deux membres de la relation 7.10 par pn , on obtient :

$$\frac{J_n}{qD_n n} \approx \frac{J_p}{qD_p p} = \frac{\partial \ln pn}{\partial x}$$

et en intégrant entre le point 1 où $pn = n_i^2$, et le point J correspondant à la jonction métallurgique (*cf.* figure 7.6), on trouve :

$$\int_1^J \frac{J_n}{qD_n n} dx - \int_1^J \frac{J_p}{qD_p p} dx = \ln \frac{pn}{n_i^2}$$

Cette expression peut être développée de la manière suivante :

$$\int_1^2 \frac{J_n}{qD_n n} dx - \int_J^2 \frac{J_n}{qD_n n} dx - \int_1^J \frac{J_p}{qD_p p} dx = \ln \frac{pn}{n_i^2}$$

d'après l'expression 7.15 la première intégrale correspond à V_a/U_T, tandis que les deux autres intégrales peuvent être négligées dans la plupart des cas pratiques où les chutes ohmiques dans les régions quasi-neutres sont petites devant U_T. Dans ces conditions, la valeur du produit pn au niveau de la jonction est donné par l'approximation de Boltzmann :

$$pn(J) = n_i^2 e^{\frac{V_a}{U_T}} \tag{7.16}$$

Chapitre 8

La jonction PN

8.1 Régime statique

8.1.1 Étude qualitative a partir du schema des bandes d'énergie

Les figures 8.1, 8.2 et 8.3 montrent l'effet d'une polarisation externe V_a sur le schéma des bandes d'énergie d'une jonction PN. Les hypothèses sont les suivantes :

- il existe encore une région dépeuplée de porteurs encadrée par deux régions quasi-neutres,
- les chutes de potentiels dans les régions quasi-neutres sont faibles devant l'amplitude de la tension appliquée et par conséquent la zone de transition supporte intégralement les variations de tension appliquée V_a

◇ $\boxed{V_a=0}$ Cette situation correspond à l'équilibre thermodynamique présentée dans le chapitre précédent. On rappelle l'existence d'une barrière de potentiel $q\phi$ qui s'oppose au passage des porteurs majoritaires vers les régions où ils sont minoritaires. Le champ électrique résultant de cette variation de potentiel crée dans la zone de transition un courant de conduction qui équilibre parfaitement le courant de diffusion engendré par les variations de concentration de porteurs.

◇ $\boxed{V_a<0}$ Une tension V_a négative renforce la différence de potentiel supportée par la zone de transition, et la barrière de potentiel qui s'oppose au passage des porteurs majoritaires prends la valeur $\phi - V_a$ supérieure à la valeur d'équilibre. Les porteurs minoritaires, qui n'ont pas de barrière de potentiel à franchir pour traverser la jonction, sont entraînés par le champ électrique et donnent naissance à un courant très faible appelé courant de saturation inverse. La différence de potentiel supportée par la région dépeuplée étant augmentée, cette région voit son extension W_T croître, il s'ensuit une diminution du gradient de porteurs, et par conséquent l'équilibre entre courant de conduction et courant de diffusion, qui régnait à l'équilibre thermodynamique, est rompu au profit de la composante de conduction.

◇ $\boxed{V_a>0}$ Une tension V_a positive s'oppose à la tension de diffusion, et par conséquent la barrière de potentiel vue par les majoritaires est affaiblie, un grand nombre de porteurs est désormais capable de traverser la jonction, le courant extérieur croit très vite avec Va. La différence de potentiel supportée par la région dépeuplée étant réduite, cette région voit son extension Wdep diminuer, et il s'ensuit une augmentation du gradient de porteurs, et par conséquent l'équilibre entre courant de conduction et courant de diffusion, qui régnait à l'équilibre thermodynamique, est rompu au profit de la composante de diffusion.

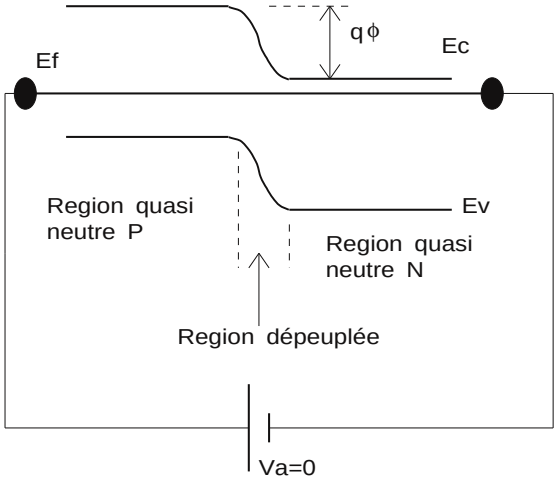


FIG. 8.1 – Équilibre

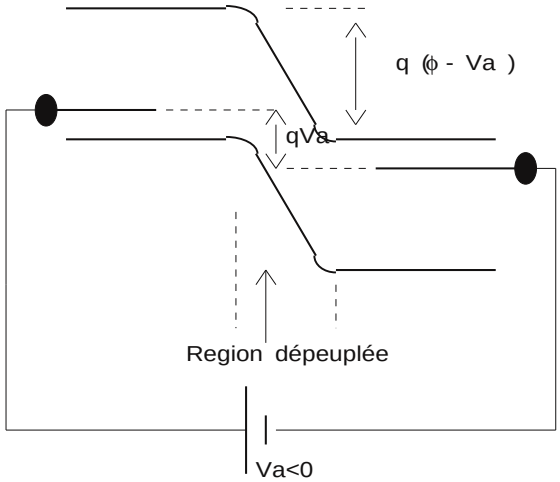


FIG. 8.2 – Polarisation inverse

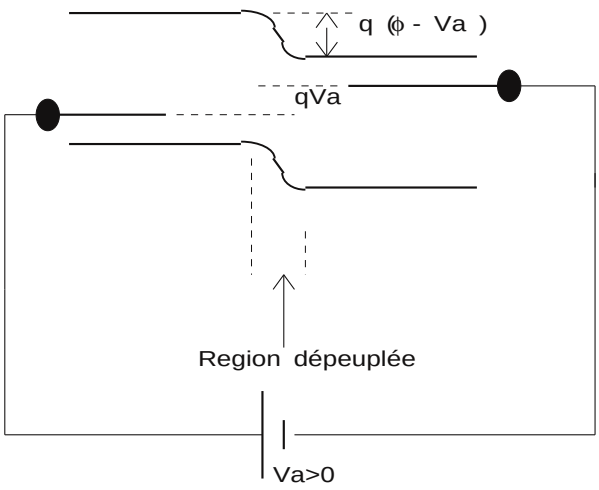


FIG. 8.3 – Polarisation directe

8.1.2 La jonction PN sans recombinaisons

Nous allons calculer la caractéristique courant tension d'une jonction PN dans le cas simple d'une jonction P^+N abrupte (dopage constant dans chaque région P ou N), et lorsque l'effet des recombinaisons peut être négligé. Ce cas n'est pas un cas d'école n'ayant aucune réalité pratique, mais au contraire représente bien la plupart des jonctions réelles. La figure 8.4 donne les notations utilisées dans ce paragraphe.

◇ Calcul de la caracteristique courant tension En éliminant le champ électrique entre les expressions des densités de courant de trous et d'électrons on obtient l'expression 7.10 reproduite ci-dessous :

$$\frac{pJ_n}{qD_n} - \frac{nJ_p}{qD_p} = \frac{\partial pn}{\partial x}$$

En intégrant cette expression entre la jonction métallurgique J et le contact ohmique 2, on trouve :

$$\int_J^2 \frac{pJ_n}{qD_n} dx - \int_J^2 \frac{nJ_p}{qD_p} = pn(2) - pn(J)$$

Lors du chapitre 2 nous avons vu que lorsque les recombinaisons étaient négligeables, l'équation de continuité montrait que les densités de courant étaient constantes, et que le rapport entre la densité de courant de trous et d'électrons était donné par la valeur de l'efficacité d'injection γ.

Dans le cas d'une jonction P^+N, l'efficacité d'injection est grande devant l'unité et par conséquent :

$$J_p \gg J_n$$

De plus dans la région N où sont effectuées les intégrales, et dans le cas des faibles niveaux d'injection où l'excès de porteurs est faible devant le niveau de dopage, la concentration en

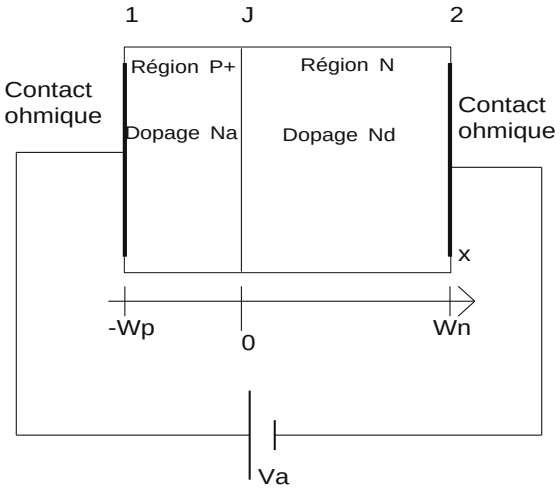


FIG. 8.4 – Description de la jonction

électrons n est très supérieure à la concentration en trous p. On a donc :

$$\int_J^2 \frac{pJ_n}{qD_n}dx \ll \int_J^2 \frac{nJ_p}{qD_p}dx$$

De plus le contact ohmique impose :

$$pn(2) = n_i^2$$

et l'approximation de Boltzmann 7.16 impose :

$$pn(J) = n_i^2 e^{\frac{V_a}{U_T}}$$

On obtient donc :

$$J_p \int_J^2 \frac{n}{qD_p}dx = n_i^2(e^{\frac{V_a}{U_T}} - 1)$$

En posant :

$$Q_b = \int_J^2 qndx = \int_J^2 qN_d dx + \int_J^2 q\hat{n}dx = Q_{b0} + Q_S \tag{8.1}$$

où Qb représente la charge des électrons dans la région N *par unité de surface*, qui en faible injection se réduit à $Q_{b0} = qN_dW_n$, on obtient la relation classique :

$$J_p = \frac{q^2 n_i^2 D_p}{Q_{b0}}(e^{\frac{V_a}{U_T}} - 1) \tag{8.2}$$

Remarque : en toute rigueur, l'intégrale donnant la valeur de Qb0 doit être effectuée sur la région quasi-neutre N et non sur tout le coté N.

La densité de courant totale J s'en déduit aisément ; elle est donné par la relation bien connue :

$$J = J_S(e^{\frac{V_a}{U_T}} - 1) \tag{8.3}$$

où la densité de courant de saturation J_S est donnée par l'expression :

$$J_S = \frac{q^2 n_i^2 D_p}{Q_{b0}} \left(1 + \frac{1}{\gamma}\right) \tag{8.4}$$

La figure 8.5 montre l'allure de la caractéristique courant tension prévue par cette relation.

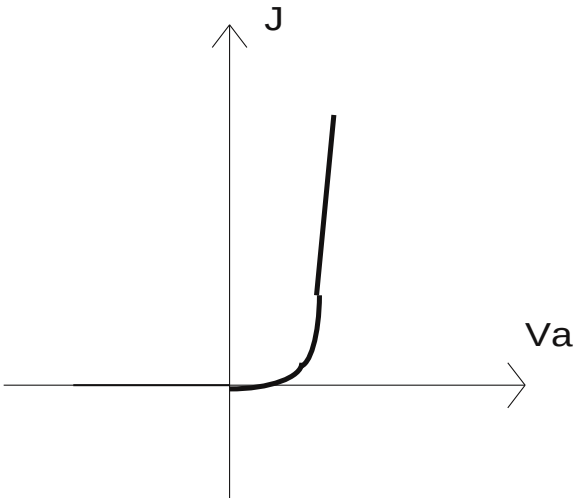


FIG. 8.5 – Caractéristique courant tension d'une jonction PN idéale

La valeur de la charge stockée Q_S par les porteurs en excès revêt une grande importance en régime dynamique, et il est intéressant de l'évaluer. La relation 7.10 réécrite dans la région quasi-neutre N, aux faibles niveaux d'injections devient :

$$-\frac{nJ_p}{qD_p} \approx \frac{dpn}{dx}$$

soit encore puisque n est sensiblement égale à N_d

$$J_p = -qD_p \frac{d\hat{p}}{dx}$$

Cette relation montre que le courant d'une jonction PN en faible injection est un courant de diffusion de minoritaires (trous dans la RQN N), et par ailleurs, que dans le cas où les recombinaisons sont négligeables, la répartition de l'excès de porteurs est linéaire comme le montre la figure 8.6 puisque les densités de courant sont constantes.

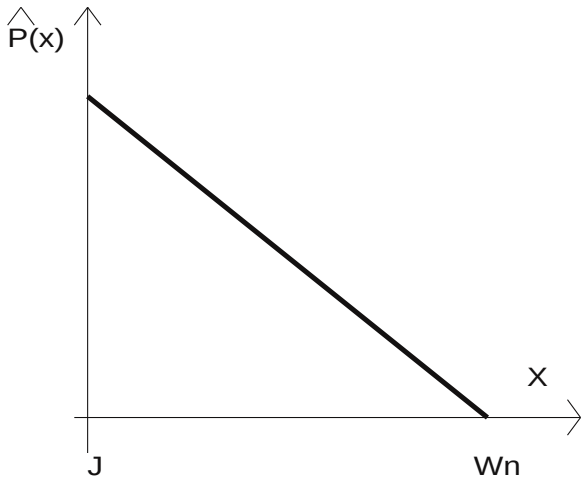


FIG. 8.6 – Répartition de l'excès de trous dans la région quasi neutre N

La charge stockée Q_s, intégrale de l'excès de porteurs, est donc proportionnelle à l'aire sous la courbe de la figure 8.6 (se rappeler que dans une région quasi-neutre l'excès de trous est égal à l'excès d'électrons), soit :

$$Q_s = \frac{1}{2} q[\hat{p}(J) - \hat{p}(W_n)]W_n$$

au niveau de la jonction l'approximation de Boltzmann donne :

$$pn(J) = pN_d = n_i^2 e^{\frac{V_a}{U_T}}$$

et donc

$$\hat{p}(J) = p - \frac{n_i^2}{N_d} = \frac{n_i^2}{N_d}(e^{\frac{V_a}{U_T}} - 1)$$

tandis qu'au niveau du contact, le contact ohmique impose un excès de porteurs nul. On obtient donc :

$$Q_s = \frac{1}{2} q \frac{n_i^2}{N_d}(e^{\frac{V_a}{U_T}} - 1)W_n \tag{8.5}$$

Le rapport entre la charge stockée Q_s et le courant de trous J_p est homogène à un temps appelé temps de transit des porteurs, dont la valeur est constante dans toute la plage des faibles niveaux d'injection :

$$\tau_B = \frac{W_n^2}{2D_p} \tag{8.6}$$

Le temps de transit représente le temps moyen que mettent les trous injectés dans la région N pour la traverser et arriver jusqu'au contact, nous verrons plus loin que le temps de transit est un paramètre très important en régime dynamique.

◇ Cas des forts niveaux d'injections Lorsque le niveau d'injection dans la région quasi-neutre N est élevé, la concentration des porteurs dépasse le niveau de dopage, et la valeur de Q_b devient égale à la charge stockée par les porteurs dans la région quasi-neutre N Q_s, on a donc :

$$\hat{n} = \hat{p} \gg N_d$$

et

$$Q_{b0} \ll Q_s \approx Q_b$$

Dans la région quasi-neutre N, la relation 7.10 devient dans ces conditions, et dans le cas d'une efficacité d'injection γ grande devant l'unité :

$$-\frac{\hat{p}J_p}{qD_p} = \frac{d\hat{p}^2}{dx}$$

ou encore :

$$\frac{d\hat{p}}{dx} = -\frac{J_p}{2qD_p}$$

ce qui montre, puisque J_p est constant, que la concentration des trous varie de manière linéaire, comme en faible injection, entre la jonction et le contact dans la région quasi neutre N, la figure 8.6 représente encore parfaitement la variation de l'excès des porteurs.

La charge stockée Q_s, intégrale de l'excès de trous, est donc proportionnelle à l'aire sous la courbe de la figure 8.6, soit :

$$Q_s = \frac{1}{2}q[\hat{p}(J) - \hat{p}(W_n)]W_n$$

au niveau de la jonction, l'approximation de Boltzmann donne :

$$pn(J) = \hat{p}^2 = n_i^2 e^{\frac{V_a}{U_T}}$$

tandis qu'au niveau du contact, le contact ohmique impose un excès de porteurs nul. On obtient donc :

$$Q_s = \frac{1}{2}qn_i e^{\frac{V_a}{2U_T}}W_n \tag{8.7}$$

et en reportant dans l'expression 8.2 de la densité de courant J_p, on trouve :

$$\boxed{J \approx J_p = \frac{2qn_iD_p}{W_n}e^{\frac{V_a}{2U_T}}} \tag{8.8}$$

Cette relation montre qu'aux forts niveaux de polarisation directe le courant croit moins vite avec la tension appliquée ; lorsque l'on trace le courant en fonction de la tension en coordonnées semi logarithmiques on obtient une caractéristique linéaire par morceaux dont la pente vaut $1/U_T$ en faible injection et $1/(2U_T)$ en forte injection comme le montre la figure 8.7. Les coordonnées du point d'intersection des asymptotes Jh et Vh, sont données par les relations :

$$J_H = \frac{4D_pQ_{b0}}{W_n^2} \quad \text{et} \quad V_H = 2U_T \ln\frac{2Q_{b0}}{qn_iW_n}$$

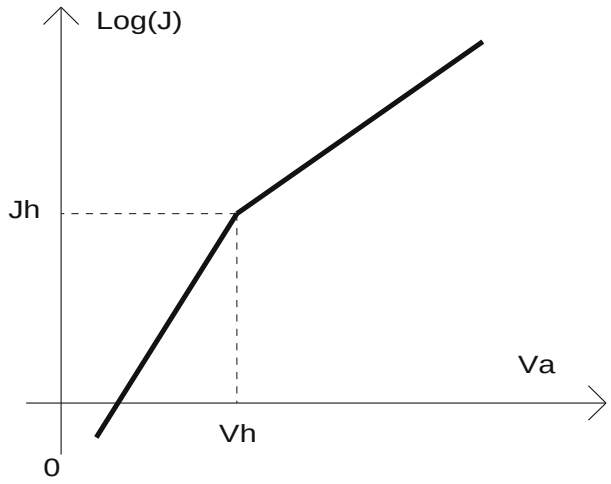


FIG. 8.7 – Infléchissement de la caractéristique dans le cas des fortes injections

Remarque : On peut aussi en forte injection calculer le temps de transit des trous dans la région quasi-neutre N, on trouve :

$$\boxed{\tau_{bH} = \frac{W_n^2}{4D_p}} \tag{8.9}$$

Remarque : On aurait pu calculer le courant d'électrons en adoptant un raisonnement identique mais en travaillant dans la région quasi-neutre P. On aurait trouvé :

$$J_n = \frac{q^2 n_i^2 D_n}{Q_E}(e^{\frac{V_a}{U_T}} - 1) \tag{8.10}$$

avec

$$Q_E = \int_1^J qpdx = qN_aW_p$$

dans le cas d'une jonction abrupte, car étant donné la concentration des dopants dans la région P+, les forts niveaux d'injection ne sont jamais atteints dans cette région.

8.1.3 Effet des recombinaisons

caractéristique courant tension

Nous allons maintenant reprendre le calcul de la caractéristique courant-tension d'une jonction PN sans négliger l'effet des recombinaisons. La relation 7.10

$$\frac{pJ_n}{qD_n} - \frac{nJ_p}{qD_p} = \frac{\partial pn}{\partial x}$$

se réécrit, en faible injection, et dans la région quasi-neutre N,

$$J_p = -qD_p\frac{\partial p}{\partial x} = -qD_p\frac{\partial \hat{p}}{\partial x} \tag{8.11}$$

Par ailleurs, l'équation de continuité des trous dans la région quasi-neutre N s'écrit en régime statique :

$$\frac{\partial p}{\partial t} = 0 = -\frac{\hat{p}}{\tau_p} - \frac{1}{q}\frac{\partial J_p}{\partial x} \tag{8.12}$$

En combinant les relations 8.11 et 8.12 on obtient l'équation différentielle suivante :

$$\frac{\partial^2 \hat{p}}{\partial x^2} - \frac{\hat{p}}{D_p\tau_p} = 0 \tag{8.13}$$

dont la solution est :

$$\hat{p}(x) = \hat{p}_0\frac{\mathrm{sh}\,[\frac{W_n - x}{L_p}]}{\mathrm{sh}\,[\frac{W_n}{L_p}]} \tag{8.14}$$

Dans cette relation la grandeur L_p est homogène à une longueur, appelée longueur de diffusion des trous, elle s'exprime en fonction de la durée de vie et de la constante de diffusion, par la relation :

$$\boxed{L_p = \sqrt{D_p\tau_p}} \tag{8.15}$$

L'excès de trous dans la région quasi-neutre au voisinage de la jonction se calcule à partir de l'approximation de Boltzmann, et vaut :

$$\hat{p}_0 = \frac{n_i^2}{N_d}(e^{\frac{V_a}{U_T}} - 1) \tag{8.16}$$

En reportant ces expressions dans la relation 8.11, on obtient l'expression de la densité de courant de trous dans la région quasi-neutre N :

$$J_p(x) = qD_p\frac{\hat{p}_0}{L_p}\frac{\mathrm{ch}\,[\frac{W_n - x}{L_p}]}{\mathrm{sh}\,[\frac{W_n}{L_p}]} \tag{8.17}$$

En adoptant le même raisonnement dans la région quasi-neutre P, on aurait obtenu des expressions similaires concernant l'excès d'électrons et la densité de courant d'électrons.

La figure 8.8 montre l'allure des variations de la concentration des porteurs minoritaires dans les régions quasi-neutres que nous venons de calculer. La densité de courant totale traversant le dispositif est constante, pour la calculer il suffit donc de connaître en un point de la structure la densité de courant de trous et la densité de courant d'électrons. En polarisation directe l'extension de la zone dépeuplée est très faible devant les profondeurs des régions

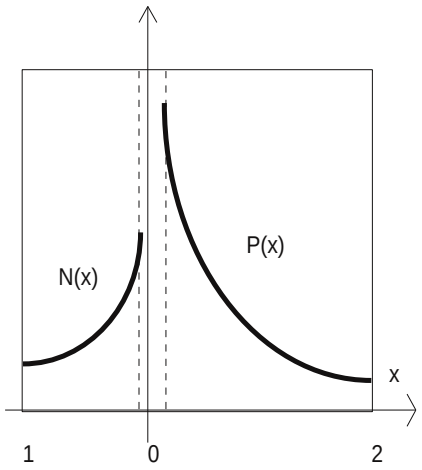


FIG. 8.8 – Densités de porteurs libres avec recombinaisons

quasi neutres, si bien que les mécanismes de recombinaisons n'entraînent qu'une très faible variation des densités de courant pendant la traversée de la région de transition, on a donc :

$$J = J_p(0) + J_n(0) \approx= qn_i^2[\frac{D_p}{L_pN_d}\coth\frac{W_n}{L_p} + \frac{D_n}{L_nN_a}\coth\frac{W_p}{L_n}](e^{\frac{V_a}{U_T}} - 1) \qquad (8.18)$$

et on retrouve encore une caractéristique courant tension de la forme

$$J = J_S(e^{\frac{V_a}{U_T}} - 1)$$

La figure 8.9 montre l'allure des variations des densités de courant le long de la structure.

Remarque : Pour des valeurs de durée de vie élevées, les longueurs de diffusionL_p et L_n deviennent grandes devant les profondeurs des régions quasi-neutres W_n et W_p, et les expressions que nous venons d'établir tendent alors vers les expressions obtenues en ignorant les mécanismes de recombinaisons. On appellera donc *jonctions courtes* des jonctions dont les profondeurs des régions N et P sont très inférieures à la longueur de diffusion des porteurs minoritaires, et *jonctions longues* celles dont les dimensions sont grandes devant la longueur de diffusion des porteurs.

◇ Cas des jonctions longues : Dans le cas des jonctions longues, la simplification des résultats précédents pour une jonction P^+N par exemple, entraîne les résultats suivants :

$$J \approx \frac{qn_i^2D_p}{L_pN_D}(e^{\frac{V_a}{U_T}} - 1) \qquad (8.19)$$

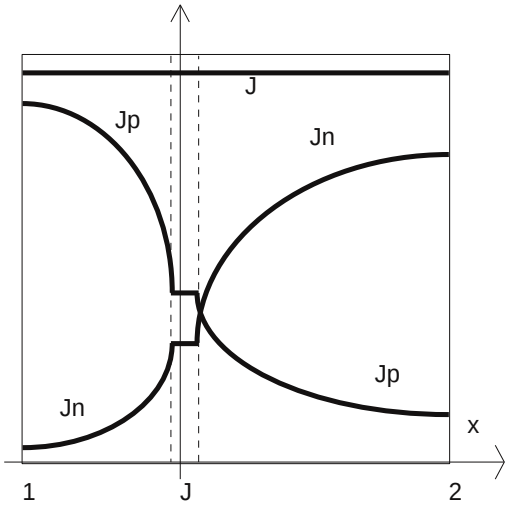


FIG. 8.9 – Variations des densités de courant dans une jonction PN avec recombinaisons

$$Q_S = \tau_p J \tag{8.20}$$

La longueur de diffusion des trous L_P a remplacé la profondeur de la région quasi neutre W_N, et la durée de vie le temps de transit des trous dans cette même région.

Courant de saturation inverse

Le courant qui traverse une jonction PN sous polarisation inverse présente deux contributions : le courant de saturation inverse de diffusion J_S que nous avons déjà calculé, et le courant de génération thermique dans la région de transition que nous noterons J_G.

Si l'on calcule l'expression du taux de génération recombinaison dans la région de charge d'espace où $pn \ll n_i^2$, on s'aperçoit qu'il est négatif et qu'il y a donc génération de porteurs dans cette région :

$$U = CN_R \frac{pn - n_i^2}{P + N + 2n_i} \approx -CN_R \frac{n_i}{2} = -\frac{n_i}{2\tau} \tag{8.21}$$

L'intégration des équations de continuité dans la région dépeuplée montre alors qu'à la traversée de la région dépeuplée le courant incident J_S subit un accroissement ΔJ du à la génération thermique :

$$\frac{\partial P}{\partial t} = 0 = -U - \frac{1}{q}\frac{\partial J_p}{\partial x}$$

d'où

$$\boxed{\Delta J_p = \Delta J = J_G = \frac{q n_i}{2\tau} W_T} \tag{8.22}$$

Ce courant, qui est prépondérant aux températures ambiantes jusqu'aux environs de 200°C, est proportionnel à la concentration intrinsèque Ni, et croit donc avec la température avec approximativement la même loi. Par ailleurs ce courant étant proportionnel à l'extension de la région dépeuplée le courant inverse croit sensiblement avec la tension appliquée.

8.1.4 Tenue en tension inverse

Dans une jonction PN polarisée en inverse la différence de potentiel supportée par la région de transition est augmentée de la valeur de la tension appliquée V_a (*cf.* figure 8.2). Les relations établies au chapitre précédent qui permettent de calculer le champ maximum et l'extension de la région dépeuplée sont encore valables si l'on remplace ϕ par $\phi - V_a$ dans ces expressions.

Pour une jonction P^+N, par exemple, on obtient :

$$\boxed{E_M = \sqrt{\frac{2qN_d}{\epsilon}(\phi - V_a)}} \tag{8.23}$$

et

$$W_T \approx \sqrt{\frac{2\epsilon}{qN_d}(\phi - V_a)} \tag{8.24}$$

On remarque que l'augmentation de la tension inverse en valeur absolue se traduit par une augmentation du champ maximum et de l'extension de la région dépeuplée.

◇ Phénomènes limitant la tension blocable par une jonction Si la tension inverse appliquée à une jonction est augmentée au delà d'une certaine valeur V_b appelée tension de claquage (breakdown voltage), le courant inverse augmente brutalement et plus rien ne s'oppose au passage du courant au niveau de la jonction. Deux mécanismes physiques déterminent cette tension de claquage ; le mécanisme d'avalanche, ou l'effet Zener pour les jonctions fortement dopées.

Claquage par avalanche

Si le champ électrique qui règne dans la région dépeuplée est suffisamment fort, les porteurs qui y transitent pourront acquérir entre deux collisions l'énergie nécessaire pour ioniser un atome du réseau cristallin, et donc générer une paire électron trou. On conçoit aisément que ce taux de génération par avalanche doit être proportionnel au champ électrique local, et au flux de porteurs qui traverse la région dépeuplée :

$$G_A(x) = A(E)\frac{J}{q}$$

Les équations de continuité dans la zone désertée s'écrivent donc (en négligeant les autres taux de génération recombinaison) :

$$\frac{\partial P}{\partial t} = 0 = G_A(x) - \frac{1}{q}\frac{\partial J_p}{\partial x}$$

$$\frac{\partial N}{\partial t} = 0 = G_A(x) + \frac{1}{q}\frac{\partial J_p}{\partial x}$$

ce qui se traduit pour les courants d'électrons et de trous, par une variation de courant ΔJ à la traversée de la zone désertée comme le montre la figure 8.10 on obtient :

$$\Delta J = \Delta J_p = -\Delta J_n = J \int_{W_T} A(E)dx$$

Sans le mécanisme d'avalanche le courant serait égal au courant de saturation de la jonc-

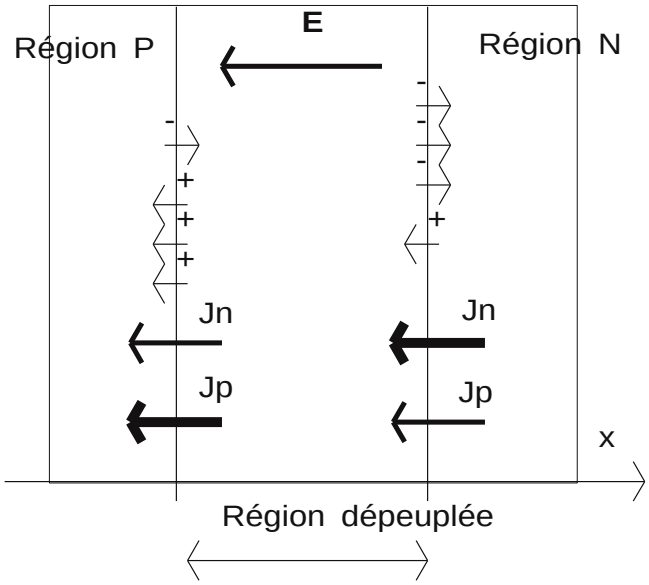


FIG. 8.10 – Densités de courant dans la jonction

tion J_S somme des courants incidents d'électrons et de trous J_{Ni} et J_{Pi}. En raison de ce mécanisme il vaut :

$$J = J_{Ni} + J_{Pi} + \Delta J = J_S + \Delta J = \frac{J_S}{1 - \int\limits_{W_T} A(E)dx} \tag{8.25}$$

Le courant devient très grand lorsque l'intégrale s'approche de l'unité, ce qui correspond au claquage de la jonction. Pour minimiser la valeur de cette intégrale d'ionisation il convient de réduire au maximum la valeur du champ électrique, ce qui s'obtient en choisissant le dopage d'une des deux régions P ou N de la jonction suffisamment faible. C'est pourquoi la tension de claquage des jonctions diminue lorsque le dopage de la région la moins dopée augmente, comme l'indique la relation empirique suivante :

$$V_b = 60[\frac{10^{16}}{N_d \text{ ou } N_a}] \tag{8.26}$$

V_b est exprimé en volts et N_d ou N_a en cm^{-3}. On obtient par exemple, $V_b = 60$V pour $N = 10^{16}$cm^{-3}, et $V_b = 1900$V pour $N = 10^{14}$cm^{-3}.

Claquage par effet zener

Pour des jonctions fortement dopées $N > 10^{18}$cm^{-3}), l'extension de la zone désertée est très réduite, si bien que le mécanisme d'avalanche n'apparaîtrait que pour des valeurs de champ électrique considérables. Or au delà de valeur de l'ordre de 10^6V/cm dans le silicium, le champ électrique génère des paires électron trou par effet tunnel à travers la bande interdite, et c'est donc ce mécanisme appelé effet zener qui limite la tension blocable des jonctions fortement dopées.

8.2 Régime dynamique

8.2.1 Analyse qualitative

Cas d'un échelon de tension

On considère une jonction P^+N sans recombinaisons et en régime de faible injection, polarisée en direct par la tension V_0, elle conduit donc le courant I_0 qui provient essentiellement du courant de diffusion des trous dans la région quasi neutre N. On a donc :

$$I_0 = SJ_p(J) = -qSD_p\frac{\partial P}{\partial x}$$

ce qui montre que le courant en régime statique est proportionnel au gradient des trous dans la région quasi neutre N. La figure 8.11 montre l'évolution de la concentration des trous dans la région quasi neutre N, lorsque l'on applique un échelon de tension à cette jonction P^+N. Comme nous venons de le voir, la concentration des trous évolue au cours du temps entre les courbes $t = t_{initial}$ où le gradient de porteurs est l'image du courant initial I_0 et la courbe $t = t_{infini}$ où le gradient de porteurs est l'image du courant final.

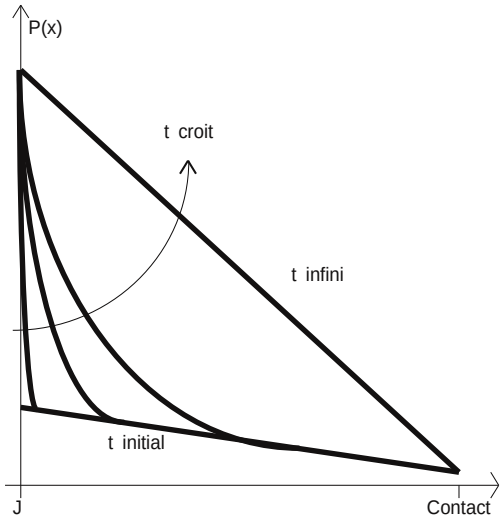


FIG. 8.11 – Évolution de la densité de porteur pour une jonction soumise à un échelon de tension

Lorsque la tension V_0 augmente brutalement jusqu'à la valeur $V_0 + \Delta V$, après un bref retard correspondant à la propagation des ondes électromagnétiques le long des régions quasi neutres, l'incrément de tension ΔV vient modifier la hauteur de barrière vue par les porteurs, et au bout d'un intervalle de temps de l'ordre de grandeur du temps de relaxation la concentration des porteurs au niveau de la jonction est augmentée. Cette augmentation se calcule aisément en utilisant l'approximation de Boltzmann :

$$p(J) = \frac{n_i^2}{N_d} e^{\frac{V_a}{U_T}}$$

Avant l'échelon et jusqu'à l'instant initial on a donc :

$$p(J, t = 0) = \frac{n_i^2}{N_d} e^{\frac{V_0}{U_T}}$$

et un bref instant après l'échelon :

$$p(J, t = 0 + \epsilon) = \frac{n_i^2}{N_d} e^{\frac{V_0 + \Delta V}{U_T}}$$

Si la concentration des trous au niveau de la jonction suit de manière quasi instantanée les variations de tension appliquée, il n'en est pas de même dans toute la région quasi neutre, les porteurs se déplaçant par diffusion on n'obtient la répartition des porteurs finale, qu'au bout d'un intervalle de temps comparable au temps de transit des porteurs. Le courant de trous au niveau du contact étant lui aussi proportionnel au gradient, est par conséquent inférieur au courant de trous à la jonction durant toute la phase transitoire. Ceci montre qu'il entre dans la région quasi neutre N plus de trous qu'il n'en sort, et donc il y a augmentation de la charge stockée pendant tout le transitoire. Cette région étant neutre il y a un stockage identique d'électrons, en effet, la continuité du courant impose que le courant d'électrons soit plus important au niveau du contact qu'au niveau de la jonction pour que la somme des courants des deux types de porteurs soit constante dans la structure. Les flux d'électrons étant de sens contraires au courants, il rentre donc dans la région quasi neutre plus d'électrons qu'il n'en sort, et la charge stockée en électrons croît de la même manière que celle des trous. Cette augmentation de charge stockée cesse lorsque le régime permanent est atteint, ce qui se produit lorsque les trous ayant diffusés jusqu'au contact on a $J_p(J) = J_p(C)$. La figure 8.12 représente les variations temporelles de ces densités de courant.

Cas d'un échelon de courant

Dans le cas d'un échelon de courant, le gradient des trous à la jonction qui est proportionnel au courant subit une brusque variation à l'instant initial et sa valeur ne change plus par la suite, c'est la diffusion des trous qui va établir progressivement ce gradient dans toute la structure comme le montre la figure 8.13. Ici encore, comme le montre la figure 8.14 le courant de trous au contact ne sera égal à celui à la jonction qu'en régime permanent, et une variation de charge stockée dans la région quasi neutre en découlera, la constante de temps sera ici aussi de l'ordre du temps de transit des porteurs dans la région quasi neutre

> **Remarque :** Nous venons de voir que dans les jonctions sans recombinaisons le temps de transit des porteurs dans les régions quasi neutres et la charge stockée par ces porteurs dans ces régions gouvernaient le régime dynamique. Dans le cas des jonctions longues où les mécanismes de recombinaisons sont prépondérants les mécanismes transitoires sont similaires, mais le régime permanent sera atteint après un intervalle de temps proche de la durée de vie des porteurs. La rapidité d'une jonction PN est donc étroitement liée au temps de transit des porteurs dans le cas des jonctions courtes, et à la durée de vie dans le cas des jonctions longues.

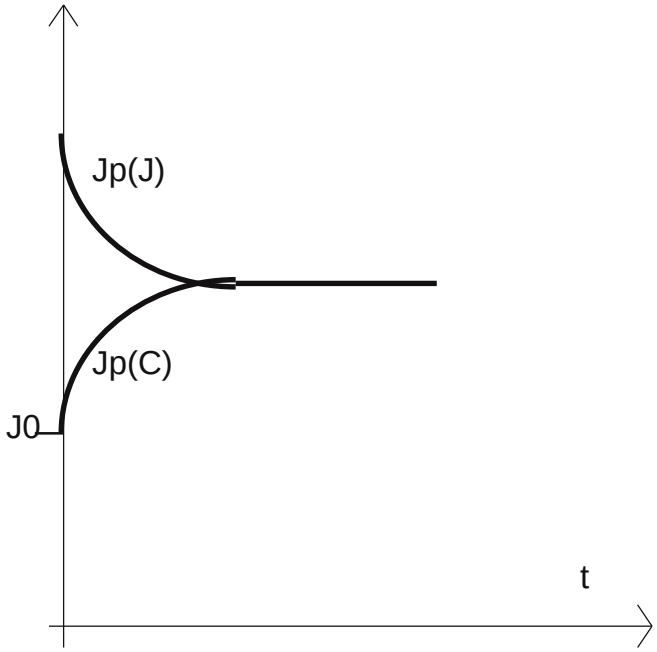


FIG. 8.12 – Évolution des courants dans le temps

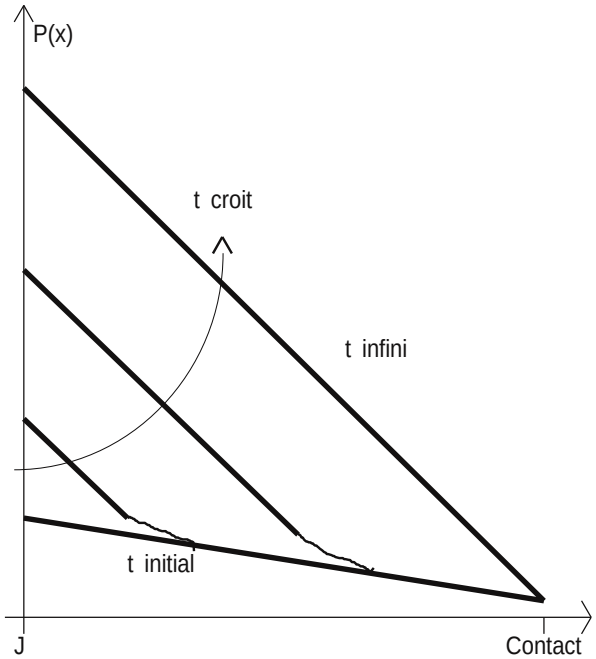


FIG. 8.13 – Évolution de la densité de porteur pour un échelon de courant

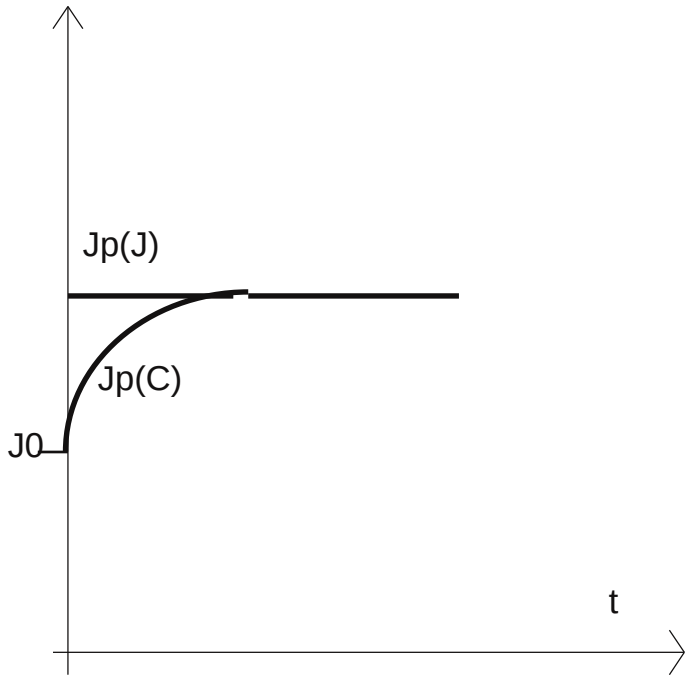


FIG. 8.14 – Évolution des courants dans le temps

8.2.2 Analyse quantitative

On se limitera, pour simplifier les calculs et pour obtenir des solutions analytiques, au cas des faibles niveaux d'injection et aux excitations lentement variables (dans le domaine fréquentiel leur spectre sera limité aux fréquences faibles par rapport à l'inverse du temps de transit des porteurs).

Calcul des courants en régime dynamique dans les jonctions sans recombinaisons

On considère une jonction PN polarisée par la tension V_a (*cf.* figure 8.15). On se place dans la région quasi neutre N où les hypothèses énoncées précédemment (faible injection et sans recombinaisons) nous permettent d'écrire :

$$\begin{cases} \dfrac{\partial N}{\partial t} \approx \dfrac{1}{q}\dfrac{\partial J_n}{\partial x} \\ J_n \approx qD_n \dfrac{\partial N}{\partial x} \end{cases}$$

On a donc l'équation différentielle

$$\frac{\partial N}{\partial t} = D_n \frac{\partial^2 N}{\partial x^2}$$

qui se traduit dans le domaine fréquentiel par l'équation :

$$pN(x,p) = D_n \frac{\partial^2 N(x,p)}{\partial x^2}$$

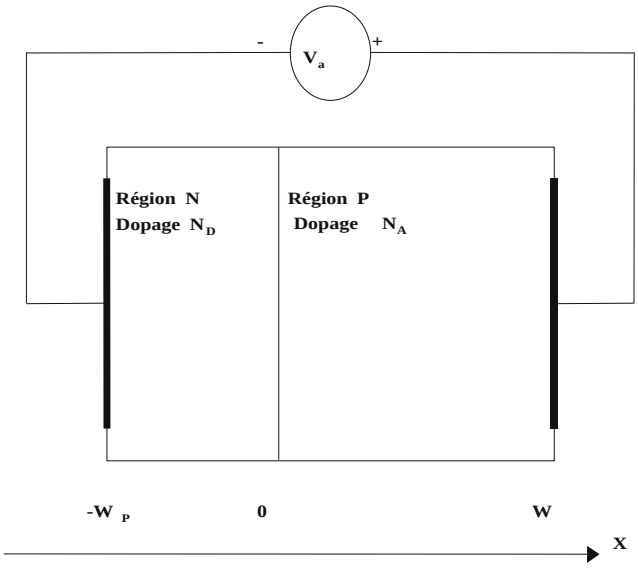


FIG. 8.15 – Jonction PN polarisée

où p est la variable de Laplace. La résolution de cette équation donne :

$$N(x,p) = \frac{N_W - N_0 e^{-Z}}{e^Z - e^{-Z}} e^{\sqrt{p/D_n}x} + \frac{N_0 e^Z - N_W}{e^Z - e^{-Z}} e^{-\sqrt{p/D_n}x}$$

avec

$$Z = \sqrt{p/D_n} W$$

N_0 et N_W sont les transformées de Laplace des concentrations d'électrons au voisinage de la jonction et du contact ohmique :

$$\begin{cases} N_0 = N(0,p) = \mathcal{L}(\frac{n_i^2}{N_a} e^{\frac{V_a}{U_T}}) \\ N_W = N(W,p) = \mathcal{L}(\frac{n_i^2}{N_a}) \end{cases}$$

Les courants d'électrons aux abscisses correspondantes ont donc pour expression :

$$\begin{cases} J_n(0,p) = \frac{qD_n}{W}(\frac{Z}{\operatorname{sh} Z} N_W - Z \coth(Z) N_0) \\ J_n(W,p) = \frac{qD_n}{W}(-\frac{Z}{\operatorname{sh} Z} N_0 - Z \coth(Z) N_W) \end{cases}$$

Pour des variations lentes de l'excitation une solution approchée peut être obtenue en faisant un développement limité des fonctions hyperboliques :

$$\begin{cases} \dfrac{Z}{\operatorname{sh} Z} \approx 1 \\ Z \coth Z \approx 1 + Z^2/2 \\ Z \ll 1 \;\Rightarrow\; \omega \ll \dfrac{1}{\tau_b} \end{cases}$$

On obtient alors dans le domaine fréquentiel :

$$\begin{cases} J_n(0,p) = \dfrac{qD_n}{W}(N_W - N_0 - Z^2/2) \\ J_n(W,p) = \dfrac{qD_n}{W}(N_W - N_0 + Z^2/2) \end{cases}$$

ce qui se traduit dans le domaine temporel par les relations :

$$\begin{cases} J_n(0,t) = -\dfrac{qn_i^2 D_n}{N_a W}(e^{\dfrac{V_a}{U_T}} - 1) - \dfrac{d}{dt}(\dfrac{W^2 q n_i^2 D_n}{2 D_n N_a W} e^{\dfrac{V_a}{U_T}}) \\ J_n(W,t) = -\dfrac{qn_i^2 D_n}{N_a W}(e^{\dfrac{V_a}{U_T}} - 1) \end{cases}$$

qui se réécrivent en utilisant les expressions du temps de transit des porteurs et de la densité de courant de saturation :

$$\begin{cases} J_n(0,t) = -J_{SN}(e^{\dfrac{V_a}{U_T}} - 1) - \dfrac{dQ_{SN}}{dt} \\ J_n(W,t) = -J_{SN}(e^{\dfrac{V_a}{U_T}} - 1) \end{cases} \tag{8.27}$$

Les expressions précédentes montrent qu'en régime dynamique lentement variable le courant d'électrons dans la région quasi neutre P au voisinage de la jonction résulte de la superposition du courant de diffusion et d'un courant réactif résultant des variations de la charge stockée par les électrons. Par ailleurs on doit noter que les expressions du courant de diffusion et de la charge stockée restent identiques à celles du régime statique. La solution précédente, obtenue par un développement limité élémentaire des fonctions hyperboliques, est identique à celle qui résulterait d'une analyse quasi statique de ce composant.

Un calcul identique réalisé dans la région quasi neutre N aurait montré que le courant de trous au voisinage de la jonction résulte de la superposition du courant de diffusion des trous dans cette région et du courant résultant des variations de charge stockée par ces trous. La continuité des courants au niveau des contacts ohmiques est assurée par les courants de majoritaires.

Pour obtenir le courant dynamique total il convient d'ajouter aux courants précédents le courant de déplacement qui n'a pas été pris en compte dans cette analyse.

Une analyse similaire des jonctions longues (recombinaisons dominantes) aboutirait au même résultat, la durée de vie des porteurs remplaçant seulement le temps de transit.

Modèle dynamique de la diode

Compte tenu du paragraphe précédent le modèle dynamique de la diode sans recombinaisons et en faible injection est le donné en figure 8.16 Les équations correspondantes

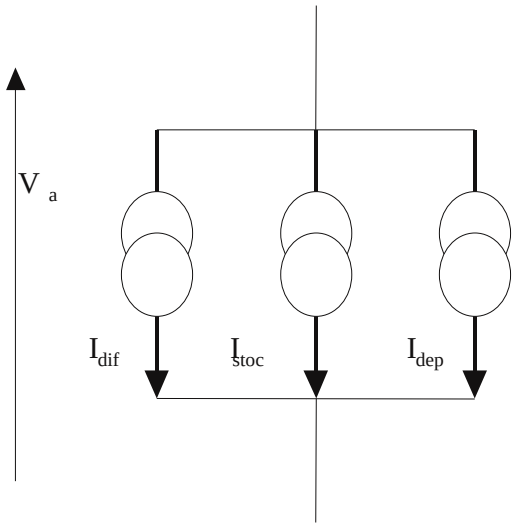


FIG. 8.16 – Modèle dynamique de diode

s'écrivent :

$$\begin{cases} I_{dif} = I_S(e^{\dfrac{V_a}{U_T}} - 1) \\ I_{stoc} = \tau \dfrac{\partial I_{dif}}{\partial t} \\ I_{dep} = \epsilon \dfrac{dE_M}{dt} \end{cases} \tag{8.28}$$

où τ est la durée de vie ou le temps de transit des porteurs.

Modèle dynamique petit signaux

En régime de petits signaux la linéarisation des trois générateurs de courant autour d'un point de fonctionnement statique (V_0, I_0) permet de donner un schéma équivalent électrique de la diode autour de ce point de fonctionnement. Ce schéma fait apparaître une conductance dynamique dont l'origine est liée aux variations du courant de diffusion, une capacité appelée capacité de stockage liée aux variations de charge stockée, et une capacité de transition modélisant le courant de déplacement résultant des variations dynamiques du champ électrique qui règne dans cette région.

◇ Conductance dynamique Si on superpose à la tension continue V_0 de polarisation la tension alternative v, et si celle ci est lentement variable par rapport au temps de transit des porteurs et à la durée de vie, le courant dans cette jonction sera donné par :

$$I_{dif} = I_S e^{\dfrac{V_0 + v}{U_T}}$$

et lorsque les variations de v sont petites devant U_T, un développement limité au premier ordre montre que le courant présente une composante variable i_{dif} dont l'amplitude est proportionnelle à v :

$$I_{dif} = I_0 + \frac{I_0}{U_T} v$$

On a donc la relation :

$$i_{diff} = \frac{I_0}{U_T} v = gv$$

où g est la conductance dynamique de la jonction. Cette conductance, qui a pour valeur

$$\boxed{g = \frac{I_0}{U_T}} \tag{8.29}$$

ne dépend pas de la jonction considérée mais uniquement des conditions de polarisation. Une jonction présente donc une conductance très faible en inverse et fortement dépendante de la polarisation en direct.

◇ Capacité de stockage Comme nous l'avons vu au paragraphe 8.2.1, les variations de tension entraînent une variation de la charge stockée dans les régions quasi neutres et par conséquent un effet réactif. Lorsque les variations de tensions sont lentes par rapport à la constante de temps de la dynamique des charges (temps de transit ou durée de vie), la charge stockée suit les variations de tension au travers des variations de courant par la relation établie au paragraphe 8.1.2 :

$$Q_S = \tau J_{dif} = \tau J_s e^{\frac{Va}{U_T}}$$

la valeur de τ étant donnée par le temps de transit dans le cas des jonctions courtes, ou par la durée de vie dans le cas des jonctions longues. Ces variations de charge stockée donnent naissance au courant variable :

$$i_{stoc} = S\frac{dQ_S}{dt} = S\frac{Q_S}{dV_a}\frac{dV_a}{dt} = C_S\frac{dV_a}{dt}$$

On obtient donc :

$$\boxed{C_S = \tau g} \tag{8.30}$$

◇ Capacité de transition La région de transition de la jonction supporte la quasi totalité de la tension extérieure, lorsque celle-ci varie, ces variations se répercutent sur la région de transition qui voit son extension varier et avec elle sa charge induite par le découvrement des impuretés fixes. Ces variations de charges induisent des variations du champ électrique et donc un courant de déplacement qui peut être modélisé par une capacité de transition en régime de petit signaux. Le champ électrique à la jonction est donné par la relation :

$$E_M = \sqrt{\frac{2qN_aN_d}{\epsilon(N_a+N_d)}(\phi - V_a)}$$

Les variations de tensions entraînent donc l'apparition du courant de déplacement :

$$i_{dep} = S\epsilon\frac{dE_M}{dt} = S\epsilon\frac{dE_M}{dV_a}\frac{dV_a}{dt}$$

soit

$$i_{dep} = S\sqrt{\frac{\epsilon q N_a N_d}{2(N_a+N_d)(\phi - V_a)}}\frac{dV_a}{dt} = C_T\frac{dV_a}{dt}$$

On en déduit la valeur de la capacité de transition

$$C_T = S\sqrt{\frac{\epsilon q N_a N_d}{2(N_a+N_d)(\phi - V_a)}} \tag{8.31}$$

Remarque : Cette expression aurait pu être obtenue en appliquant la formule du condensateur plan avec comme épaisseur l'extension de la région dépeuplée.

Schéma équivalent petit signal d'une jonction PN

Le schéma équivalent petit signal d'une jonction PN fait donc apparaître une conductance g qui ne dépend que de la polarisation de la jonction, une capacité de stockage C_S résultant du stockage des charges dans les régions quasi neutres, et d'une capacité de transition C_T provenant de la région dépeuplée qui se comporte comme un condensateur plan à épaisseur variable. La capacité de transition est prépondérante en inverse, alors que la capacité de stockage domine très vite en polarisation directe.

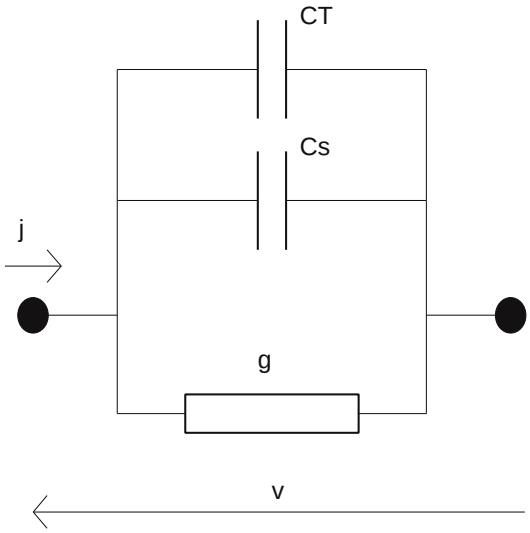


FIG. 8.17 – Modèle petit signal de la jonction PN

Remarque : *modélisation grand signal dans les logiciels de CAO*
Dans les logiciels de CAO la réponse grand signal est généralement obtenue par discrétisation dans le domaine temporel :
à chaque instant t le schéma équivalent est calculé et la réponse à l'instant t+dt est obtenue en remplaçant la diode par son circuit linéaire équivalent. La validité de cette méthode suppose de limiter les incréments de temps dt de manière à obtenir des incréments de tension appliquée à la diode compatibles avec la validité du modèle petit signal.

8.2.3 Commutation d'une jonction pn sur charge resistive

Mise en conduction

On attaque une jonction P^+N en série avec une résistance R par un échelon de tension comme le montre la figure 8.18. A l'instant initial, la concentration des porteurs dans la

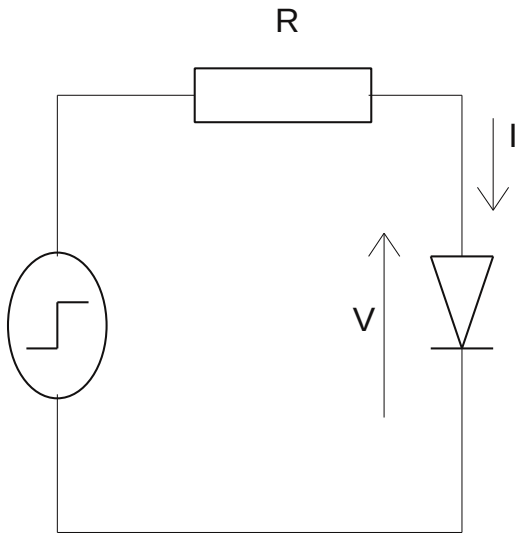


FIG. 8.18 – Commutation d'une diode sur charge R

structure est celle de l'équilibre. Après un bref délai nécessaire à la charge de la capacité de transition vers la valeur de la tension de diffusion, le courant s'établit à la valeur E/R dans la mesure où la DDP supportée par la jonction en direct reste faible devant la chute ohmique dans la résistance R.
Le courant dans la structure étant sensiblement égal au courant de trous au niveau de la jonction, et celui-ci étant un courant de diffusion, il s'établit donc de manière quasi instantanée, un gradient de trous à la jonction qui se propage ensuite par diffusion dans toute la région quasi neutre N. Au début du transitoire, le gradient des trous étant nul au niveau du contact,

un courant d'électrons provenant du contact assure la continuité du courant, et la neutralité de la région quasi neutre. La différence entre le courant de trous à la jonction et au contact pendant le régime transitoire entraîne une augmentation de la charge stockée. Les figures 8.19 et 8.20 illustrent l'analyse précédente.

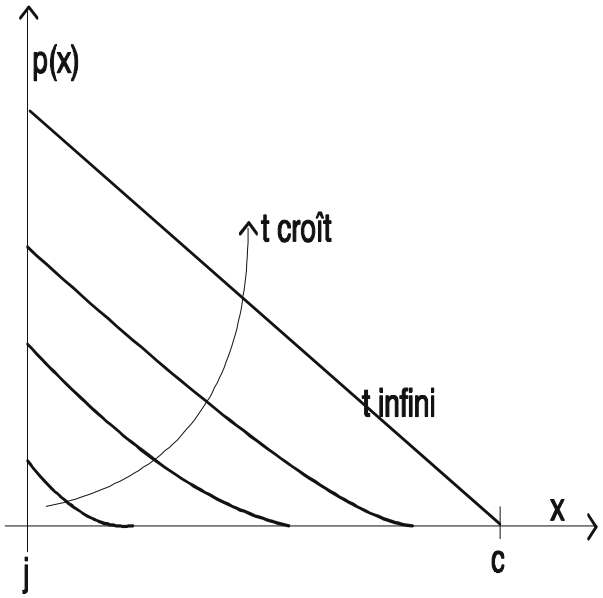


FIG. 8.19 – Évolution de la densité de porteurs

La chute de tension dans la diode se déduit des concentrations de porteurs par l'approximation de Boltzmann, elle doit donc évoluer entre une valeur nulle et la valeur correspondant au régime permanent. Toutefois, cette approximation négligeant les chutes de tension dans les régions quasi neutres n'est pas toujours vérifiée au début du régime transitoire, tout particulièrement dans les jonctions possédant une région quasi neutre profonde et faiblement dopée, comme les diodes de puissance. On observe souvent dans ce type de jonctions une surtension due à cet effet résistif.

Blocage

La jonction conduisant le courant direct en régime permanent, on retourne brutalement la tension E.

La capacité d'une jonction PN à bloquer une tension inverse repose sur l'existence d'une région dépeuplée de porteurs, or la charge stockée par les porteurs en excès doit d'abord être évacuée (au moins autour de la jonction) pour que celle-ci se développe, la jonction ne peut donc encore supporter une tension inverse. La DDP supportée par la jonction peut être évaluée par l'approximation de Boltzmann, elle est positive au début du transitoire (tant que le produit pn au niveau de la jonction est supérieur à n_i^2), et est très faible par rapport à E. Le courant s'établit donc instantanément en sens inverse à la valeur $-E/R$. Ce courant étant proportionnel au gradient de porteurs à la jonction, ce gradient s'inverse donc au début du

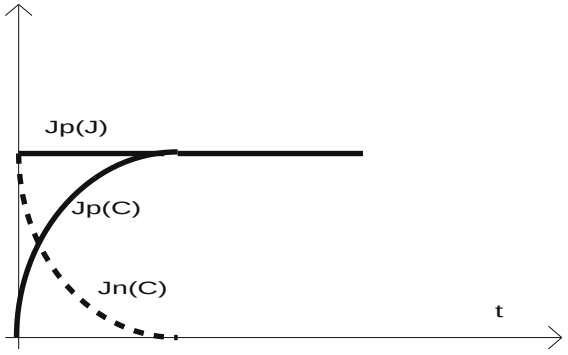


FIG. 8.20 – Évolution des densités de courant

transitoire, et reste constant tant que la charge stockée est présente dans toute la région quasi neutre. Le courant inverse circulant dans la jonction permet d'évacuer les porteurs en excès dans la région quasi neutre, et de réduire la valeur de la charge stockée et des concentrations de porteurs. Dans le cas des jonctions longues la charge stockée est réduite aussi par le jeu des mécanismes de recombinaisons.

Après un certain retard appelé temps de désaturation, le produit pn à la jonction devient inférieur à n_i^2 et une région dépeuplée peut alors apparaître, ainsi qu'une tension aux bornes de la jonction en inverse ce qui permet une réduction du courant inverse. Lorsque toute la charge stockée a été évacuée, le courant s'annule et la phase de blocage est terminée.

Chapitre 9

Le transistor bipolaire à homojonctions

9.1 Analyse qualitative

La figure 9.1 représente la structure d'un transistor bipolaire PNP idéalisé, et définit les notations que nous utiliserons par la suite. Nous raisonnerons en considérant le cas d'un transistor PNP, mais la transposition au cas d'un transistor NPN s'effectue facilement en inversant le sens des tensions et courants et en permutant les indices 'n' et 'p' relatifs aux électrons et aux trous.

Un transistor bipolaire comporte trois couches semiconductrices délimitées par deux jonctions PN. La couche la plus dopée constitue l'émetteur. La base est suffisamment mince pour que les recombinaisons n'y jouent qu'un rôle mineur ; elle est également moins dopée que l'émetteur afin que l'efficacité d'injection de la jonction émetteur base soit importante. Enfin la région du collecteur doit être profonde et peu dopée, car la jonction collecteur-base étant normalement polarisée en inverse, c'est là que se développe la majeure partie de la région dépeuplée.

En régime normal de fonctionnement la jonction émetteur-base est polarisée en direct et la jonction collecteur-base en inverse. L'émetteur injecte donc dans la base un fort courant de trous dont la majeure partie atteint la jonction de collecteur par diffusion (la base est courte). Les trous rencontrent au niveau de la jonction collecteur-base le fort champ électrique de la

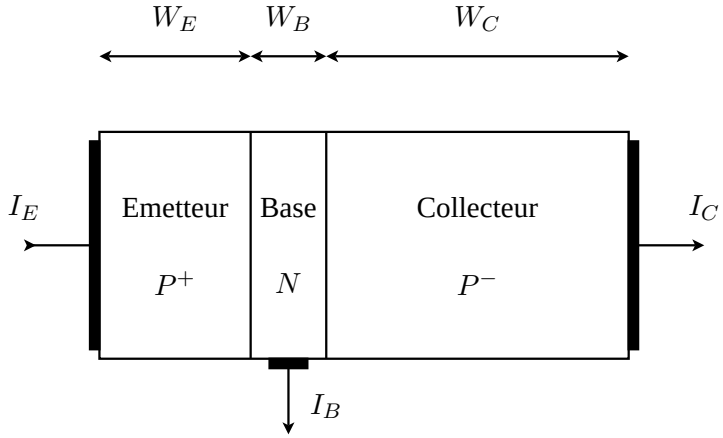


FIG. 9.1 – Transistor PNP idéalisé

zone dépeuplée, ce champ favorise leur passage et les entraîne dans le collecteur. La jonction collecteur-base étant polarisée en inverse, le courant qui lui est propre est négligeable devant le courant de trous injecté depuis l'émetteur, c'est donc ce courant de trous qui constitue la quasi totalité du courant de collecteur.

Le courant issu de la jonction base-émetteur, qui constitue le courant d'émetteur, se divise donc en deux (*cf.* figure 9.2), le courant de trous qui comme nous venons de le voir constitue le courant de collecteur, et le courant d'électrons qui constitue le courant de base. Ce courant qui provient de l'injection des électrons dans la région P^+ est toujours très petit par rapport au courant de collecteur (dans une jonction P^+N, on a $J_p \gg J_n$), on a donc les relations suivantes :

$$\begin{cases} I_E = I_p + I_n \\ I_C = I_p \approx I_E \\ I_B = I_n \ll I_C \end{cases}$$

Par ailleurs, la tension collecteur-base qui est une tension inverse, peut être très supérieure à la tension base-émetteur qui avoisine $0,6V$, on a donc :

$$|I_B|\ |V_{BE}| \ll |I_C|\ |V_{CE}|$$

ce qui montre, qu'en utilisant la tension bas-émetteur comme grandeur de commande et la tension collecteur-émetteur comme grandeur de sortie, on dispose d'une amplification de puissance conséquente.

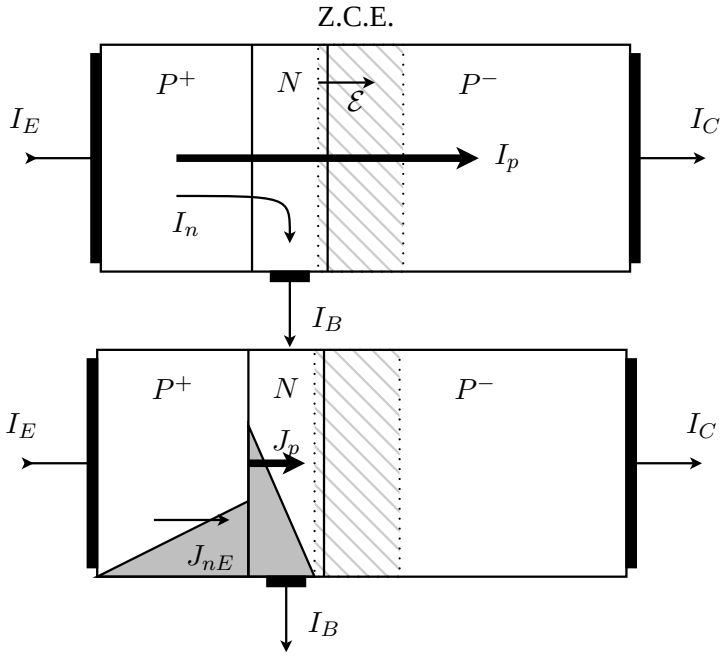


FIG. 9.2 – Courants de diffusion et concentrations de minoritaires dans un transistor PNP.

9.2 Modèle equivalent en regime statique

Pour établir ce modèle nous allons calculer le courant de trous qui circule dans la région quasi-neutre N. Nous nous placerons dans l'hypothèse d'une région courte, ce qui est toujours le cas pour un transistor bien construit. Ce calcul est semblable au calcul du courant

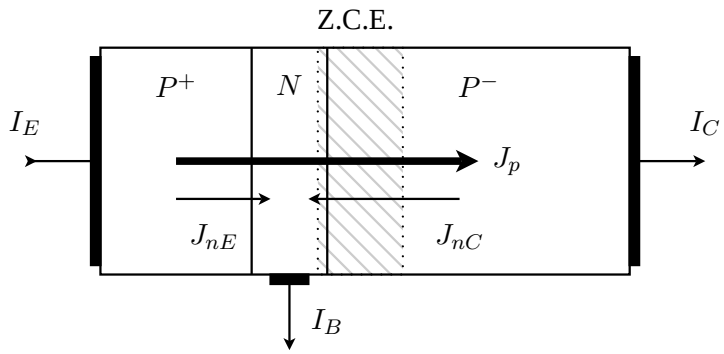


FIG. 9.3 – Définition des notations utilisées pour les densités de courants

d'électrons dans la région quasi-neutre N d'une jonction P^+N, la seule différence réside dans le fait que le contact ohmique est remplacé par la jonction collecteur-base. La figure 9.3 définit les notations utilisées.

Dans la région quasi-neutre de base, la relation 7.10 se simplifie, et devient :

$$-\frac{nJ_p}{qD_p} = \frac{d\,pn}{dx}$$

en intégrant cette relation entre les frontières E et C de la région quasi-neutre on trouve :

$$J_p = \frac{q^2 D_p[pn(J_{BE}) - pn(J_{BC})]}{\int_E^C qn(x)dx}$$

L'utilisation de l'approximation de Boltzmann fournit les valeurs du produit pn au voisinage des jonctions :

$$pn(J_{BE}) = n_i^2 e^{V_{EB}/U_T}$$
$$pn(J_{BC}) = n_i^2 e^{V_{CB}/U_T}$$

Par ailleurs, la charge des électrons dans la région quasi-neutre de base peut se scinder en deux contributions, la charge des dopants, et la charge des porteurs en excès :

$$\int_E^C qn(x)dx = \int_E^C qN_D dx + \int_E^C q\hat{p}(x)dx$$

soit,

$$Q_B = Q_{B0} + Q_{SB}$$

Le courant de trous a donc pour valeur :

$$\boxed{J_p = \frac{q^2 n_i^2 D_{pB}}{Q_B}\left[e^{V_{EB}/U_T} - e^{V_{CB}/U_T}\right]}$$

Le même raisonnement appliqué aux régions quasi-neutres d'émetteur et de collecteur permettrait de calculer le courant d'électrons qui y transite, on trouverait :

$$\boxed{J_{nE} = \frac{q^2 n_i^2 D_{nE}}{Q_E}\left[e^{V_{EB}/U_T} - 1\right]}$$

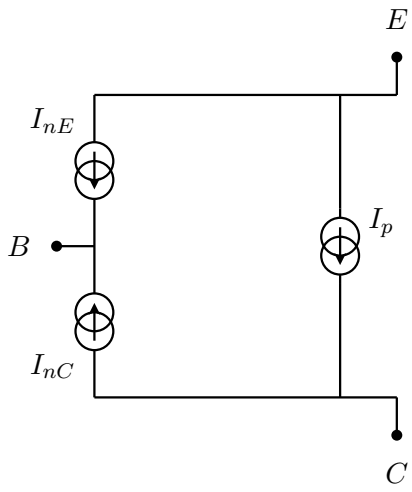


FIG. 9.4 – Modèle équivalent en régime statique

$$\boxed{J_{nC} = \frac{q^2 n_i^2 D_{nC}}{Q_C}\left[e^{V_{CB}/U_T} - 1\right]}$$

Le modèle équivalent du transistor en régime statique est indiqué figure 9.4.

Remarque : Ce schéma est valable quelque soit le signe des tensions V_{EB} et V_{CB}. L'effet des fortes injections est inclus dans les termes Q_B, Q_E, et Q_C.

9.2.1 Cas des faibles niveaux d'injection

Lorsque le niveau d'injection reste faible, les valeurs de Q_B, Q_E, Q_C, peuvent être considérés comme constantes en première approximation, et on peut écrire dans ces conditions :

$$\begin{aligned} I_p &= \frac{Sq^2 n_i^2 D_{pB}}{Q_{B0}}\left[(e^{V_{EB}/U_T} - 1) - (e^{V_{CB}/U_T} - 1)\right] \\ &= I_{SP}(e^{V_{EB}/U_T} - 1) - I_{SP}(e^{V_{CB}/U_T} - 1) \\ &= I_{CC} - I_{EC} \end{aligned}$$

$$I_{nE} = \frac{Sq^2 n_i^2 D_{nE}}{Q_{E0}}\left[e^{V_{EB}/U_T} - 1\right] = I_{SnE}[e^{V_{EB}/U_T} - 1]$$

$$I_{nC} = \frac{Sq^2 n_i^2 D_{nC}}{Q_{C0}}\left[e^{V_{CB}/U_T} - 1\right] = I_{SnC}\left[e^{V_{CB}/U_T} - 1\right]$$

Par ailleurs, les rapports entre les courants de saturation étant constants dans ce cadre d'hypothèse, on peut poser :

$$BF = \frac{I_{CC}}{I_{nE}} = \frac{I_{SP}}{I_{SnE}} = \frac{Q_{E0} D_{pB}}{Q_{B0} D_{nE}} = \gamma_{EB}$$

$$BR = \frac{I_{EC}}{I_{nC}} = \frac{I_{SP}}{I_{SnC}} = \frac{Q_{C0} D_{pB}}{Q_{B0} D_{nC}} = \gamma_{CB}$$

où γ_{EB} et γ_{CB} sont les efficacités d'injection des jonctions EB et CB. On obtient alors le modèle équivalent statique de Gummel-Poon simplifié, qui est représenté sur la figure 9.5.

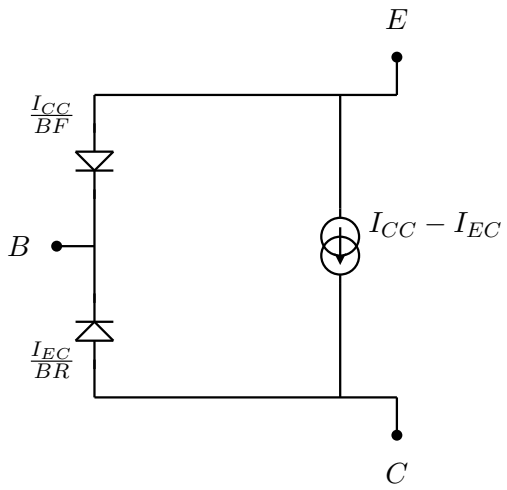


FIG. 9.5 – Modèle équivalent statique de Gummel-Poon simplifié

Remarque : Dans certains transistors l'effet des recombinaisons n'est pas totalement négligeable, il s'ensuit un courant de recombinaison qui vaut Q_{SB}/τ_p. Cette composante est négligeable devant le courant de trous si bien que les valeurs des courants de collecteur et d'émetteur ne sont pas affectées, en revanche elle peut être supérieure au courant d'électrons et modifier considérablement le courant de base. Même dans ces conditions on peut encore définir un gain en courant BF constant qui vaut :

$$BF = \frac{I_{CC}}{(Q_{SB}/\tau_p)} = \frac{\tau_p}{\tau} = \frac{\text{durée de vie}}{\text{temps de transit}}$$

9.3 Modes de fonctionnement du transistor bipolaire

Selon le signe des tensions V_{EB} et V_{CB} appliquées aux deux jonctions du transistor bipolaire on peut définir quatre régimes de fonctionnement.

9.3.1 Mode bloqué : $V_{EB} < 0$ et $V_{CB} < 0$

Lorsque les deux jonctions sont polarisées en inverse les courants qui traversent le composant sont de l'ordre de grandeur des courants de saturation des jonctions qui le constituent. De plus les jonctions étant en inverse elles peuvent supporter des tensions importantes, le transistor est bloqué, et il se comporte entre émetteur et collecteur comme un interrupteur ouvert avec un faible courant de fuite.

9.3.2 Mode normal : $V_{EB} > 0$ et $V_{CB} < 0$

Comme son nom l'indique le mode normal est le régime de fonctionnement le plus souvent utilisé et pour lequel sa structure a été optimisée. C'est dans cet état de polarisation que nous avons décrit le transistor au paragraphe 9.1. Rappelons qu'un courant important traverse le dispositif de l'émetteur vers le collecteur, et qu'un courant faible est évacué par l'électrode de base. Les répartitions des porteurs minoritaires dans les régions quasi neutres,

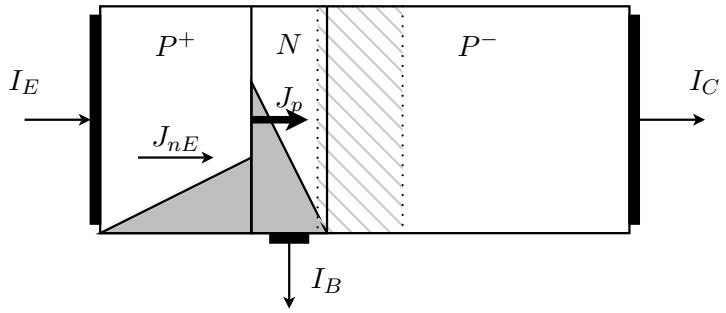


FIG. 9.6 – Concentration des minoritaires et sens des courants de diffusion pour le mode normal.

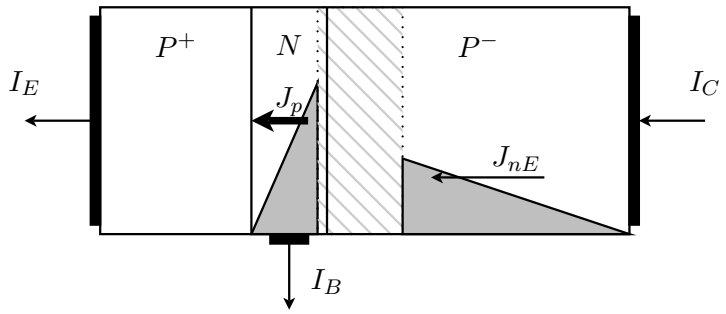


FIG. 9.7 – Concentration des minoritaires et sens des courants de diffusion pour le mode inversé.

responsables des courants de diffusions observés en régime normal, sont représentées sur la figure 9.6.

L'analyse du modèle de Gummel-Poon nous indique que :

$$\begin{cases} I_C = I_{CC} \\ I_B = \dfrac{I_{CC}}{BF} \\ I_E = I_{CC} + \dfrac{I_{CC}}{BF} = (1 + \dfrac{1}{BF}) I_C \end{cases}$$

Le paramètre BF du modèle peut donc être défini comme le gain en courant en émetteur commun du dispositif. Sa valeur est égale à l'efficacité d'injection de la jonction EB, ou au facteur de transport. Une valeur importante est toujours recherchée, et est atteinte en surdopant l'émetteur par rapport à la base, la centaine est une valeur typique pour un transistor de type signal.

9.3.3 Mode inversé : $V_{EB} < 0$ et $V_{CB} > 0$

Dans cette configuration la symétrie de la structure permet de transposer les résultats du régime normal en intervertissant les rôles du collecteur et de l'émetteur. Le sens des courants et les répartitions de porteurs sont représentés sur la figure 9.7.

L'analyse du modèle statique de Gummel-Poon donne :

$$\begin{cases} I_E = I_{EC} \\ I_B = \dfrac{I_{EC}}{BR} \\ I_C = I_{EC} + \dfrac{I_{EC}}{BR} = (1 + \dfrac{1}{BR}) I_E \end{cases}$$

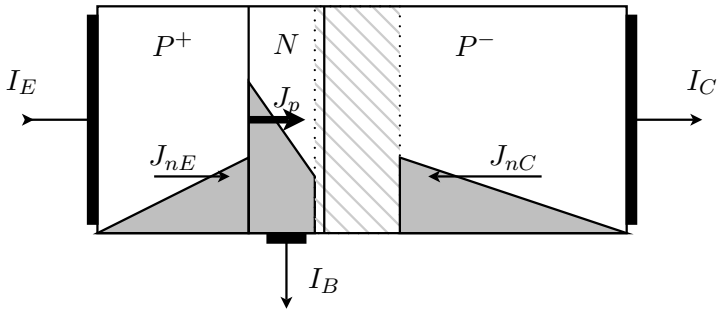


FIG. 9.8 – Concentration des minoritaires et sens des courants de diffusion pour le mode saturé.

On pourrait penser à première vue que le transistor a les mêmes performances en régime inverse qu'en régime normal. Malheureusement la nécessité d'avoir un collecteur moins dopé que la base entraîne une faible valeur pour BR, et le rendement du dispositif dans ce régime est très médiocre, si bien que le transistor n'est jamais utilisé dans cette configuration.

9.3.4 Mode saturé : $V_{EB} > 0$ et $V_{CB} > 0$

En régime saturé les deux jonctions du dispositif sont polarisées en direct, ce qui entraîne l'existence d'une charge stockée dans chaque région quasi-neutre du transistor. La figure 9.8 montre alors les répartitions de porteurs minoritaires le long de la structure. Le collecteur étant profond et peu dopé c'est souvent dans cette région que se concentre la majeure part de la charge stockée, celle-ci a un effet bénéfique en ce qui concerne les chutes ohmiques dans la région de collecteur lorsque l'on utilise le transistor en interrupteur, mais ralentit considérablement la transition vers l'état bloqué lors de la commutation d'ouverture.

Les deux jonctions supportant des tensions directes, l'approximation de Boltzmann montre que la d.d.p. aux bornes de chaque jonction avoisine leur tension de diffusion. La d.d.p. entre émetteur et collecteur $V_{EC} = V_{EB} - V_{CB}$ est dans ces conditions faible ($< 0,6\ V$), et très peu dépendante du courant collecteur. Le transistor est donc équivalent à un interrupteur fermé. En régime saturé, le courant de base résulte de l'injection d'électrons dans l'émetteur et dans le collecteur (*cf.* figure 9.8),

$$I_B = S(J_{nE} + J_{nC}) = \frac{I_{CC}}{BF} + \frac{I_{EC}}{BR},$$

tandis que le courant de collecteur provient de la différence entre le courant de trous et le courant d'électrons dans le collecteur,

$$I_C = S(J_p - J_{nC}) = (I_{CC} - I_{EC}) - \frac{I_{EC}}{BR}.$$

Le gain en courant I_C/I_B est toujours plus faible qu'en régime normal, et dépend de l'état de saturation du transistor.

9.4 Caractéristiques statiques du transistor bipolaire

Les caractéristiques statiques dans la configuration émetteur commun sont représentées sur la figure 9.9.

Ces caractéristiques découlent directement du modèle statique de Gummel-Poon simplifié que nous avons défini précédemment. La frontière entre le régime saturé et le régime

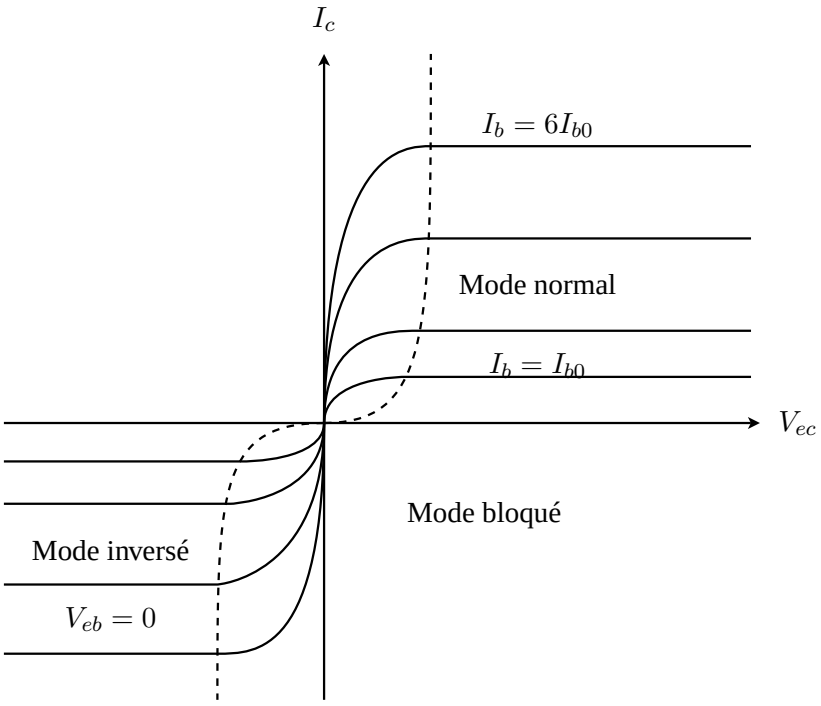


FIG. 9.9 – Caractéristiques du transistor bipolaire PNP

normal se produit lorsque $V_{BC} = 0$, on obtient alors une caractéristique du type diode qui est tracée sur la figure. Dans ce modèle simplifié on doit garder à l'esprit que :

- Le rapport I_C/I_B est constant quel que soit la d.d.p. appliquée à la jonction en inverse (la jonction BC en régime normal), si bien qu'aucune pente n'est relevée sur ces courbes contrairement aux observations expérimentales.
- La pente observée sur les caractéristiques réelles provient de la modulation de la charge Q_B par la tension inverse appliquée à la jonction collecteur-base (en régime normal) qui vient modifier la profondeur de la région quasi neutre. Dans le modèle de Gummel-Poon complet, cet effet qui porte le nom d'effet Early, est pris en compte.
- L' effet des fortes injections n'étant pas pris en compte, le rapport I_C/I_B ne dépend pas du niveau de courant. Dans la réalité on observe, pour les très forts niveaux de courant I_C un tassement du gain en courant, ce phénomène est dû au passage en forte injection de la région quasi-neutre de base, ce qui entraîne une dépendance différente par rapport à la tension de commande des courant de trous et d'électrons. Cet effet est pris en compte dans le modèle de Gummel-Poon complet.

9.5 Modèle dynamique du transistor bipolaire

Pour obtenir le modèle dynamique du transistor bipolaire il convient d'ajouter au modèle de Gummel-Poon statique les effets liés au stockage des charges dans les différentes régions du dispositif. Une approche simplifiée consiste à ne considérer que les charges stockées dans la région quasi-neutre de base, et les charges liées aux impuretés dans les régions dépeuplées. Nous nous limiterons encore dans cette analyse au cadre de l'hypothèse quasi-statique.

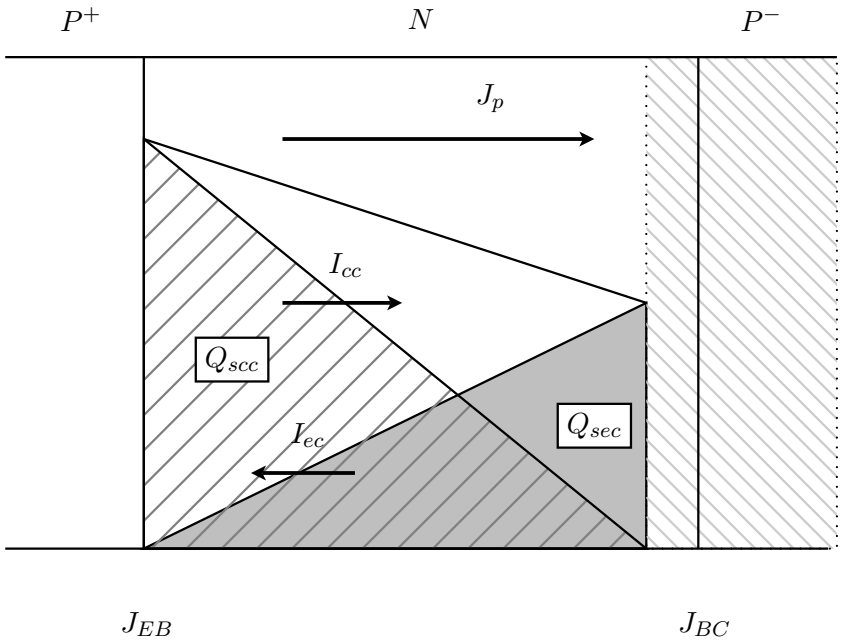


FIG. 9.10 – Contributions à la charge stockée dans la base en régime saturé

9.5.1 Charges stockées dans la région quasi-neutre de base

En régime normal, La charge stockée dans la base est due à l'injection de porteurs par l'émetteur, elle est donc proportionnelle au courant I_{CC} (pour un transistor PNP, le courant I_{CC} est le courant de trous, courant qui sera collecté au collecteur, d'où sa dénomination) et au temps de transit dans la base :

$$Q_s = \tau I_{CC}$$

En régime inversé, La charge stockée dans la base est due à l'injection de porteurs par le collecteur, elle est donc proportionnelle au courant Iec (pour un transistor PNP, le courant I_{EC} est le courant de trous, courant qui sera collecté à l'émetteur, d'où sa dénomination) et au temps de transit dans la base :

$$Q_s = \tau I_{EC}$$

Dans le cas général, la charge stockée dans la base résulte de la superposition des deux contributions précédentes :

$$Q_s = \tau I_{CC} + \tau I_{EC} = Q_{sCC} + Q_{sEC}$$

La figure 9.10 montre ces deux contributions en régime saturé.

La charge Q_{sCC} étant controlée par la tension émetteur base il faudra donc ajouter une capacité non linéaire (appelée C_{DE}, capacité de diffusion émetteur, dans le modèle de Gummel-Poon) entre les plots d'émetteur et de base. La charge Q_{sEC} étant contrôlée par la tension collecteur-base on ajoutera une capacité non linéaire (appelée C_{DC}, capacité de diffusion collecteur, dans le modèle de Gummel-Poon) entre les plots de collecteur et de base.

9.5.2 Charges stockées dans les régions dépeuplées

Ces charges sont encore contrôlées par les tensions appliquées aux deux jonctions, on ajoutera donc une capacité de transition C_{TE} entre les plots d'émetteur et de base, et une capacité de transition C_{TC} entre les plots de collecteur et de base.

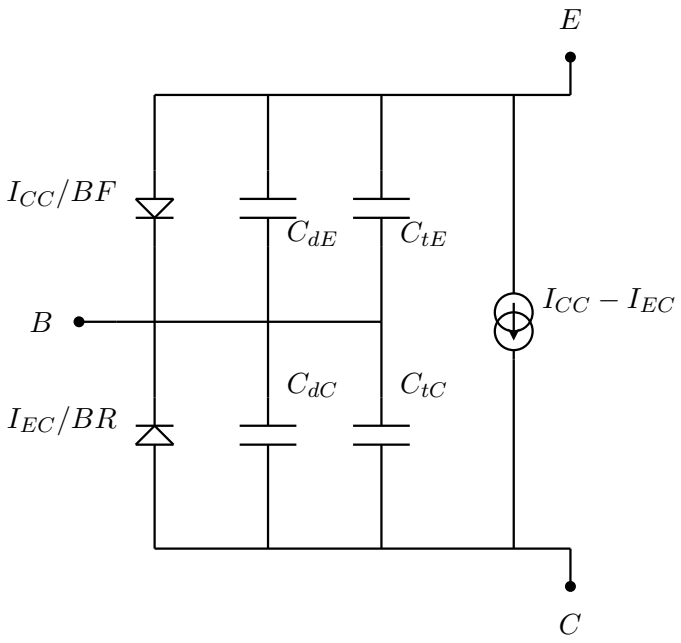


FIG. 9.11 – Schéma équivalent petit signal

9.5.3 Modèle équivalent en régime dynamique

Compte tenu des modifications apportées au modèle statique que nous venons de présenter, on obtient le modèle équivalent représenté sur la figure 9.11, valable pour des fréquences inférieures au temps de transit des porteurs dans la base. Ce modèle non linéaire peut être utilisé pour des signaux de forte amplitude, la réponse est obtenue de proche en proche dans le domaine temporel, en utilisant une discrétisation suffisamment fine pour que les éléments du modèle puissent être considérés constants dans un intervalle de temps.

9.6 Schéma équivalent dynamique petit signal en régime normal dans la configuration émetteur-commun

Le schéma équivalent dynamique petit signal résulte de la linéarisation du modèle grand signal autour d'un point de polarisation. En régime normal la tension V_{CB} étant négative, le courant I_{EC} est négligeable devant les autres composantes du courant, et il en est de même du courant réactif qui lui est associé, tout les éléments résultant de la linéarisation de I_{EC} seront donc négligés (conductance base-collecteur, capacité de diffusion base-collecteur, transconductance collecteur-émetteur).

9.6.1 Conductance base-émetteur

La conductance base-émetteur relie les variations du courant de base aux variations de la tension base-émetteur, on a donc la relation :

$$g_{BE} = \frac{dI_B}{dV_{EB}} = \frac{d}{dV_{EB}}\left[\frac{I_{SP}}{BF}e^{V_{EB}/U_T}\right] = \frac{I_{CC}}{BF\,U_T} = \frac{I_B}{U_T}$$

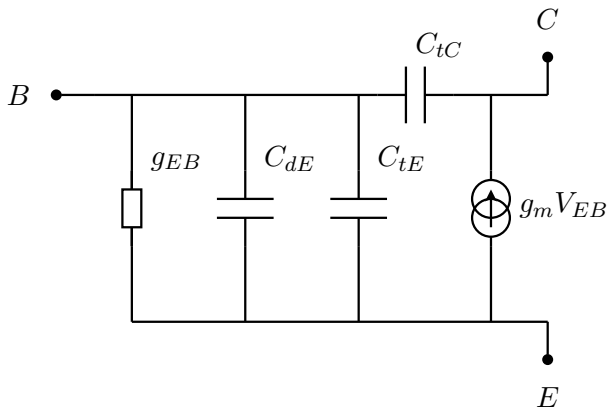


FIG. 9.12 – Schéma équivalent petit signal d'un transistor PNP

9.6.2 Transconductance

La transconductance relie les variations du courant de collecteur aux variations de la tension base-émetteur, on a donc la relation :

$$g_m = \frac{dI_C}{dV_{EB}} = \frac{d}{dV_{EB}}\left[I_{SP}e^{V_{EB}/U_T}\right] = \frac{I_{CC}}{U_T} = \frac{I_C}{U_T}$$

9.6.3 Capacité de diffusion

La capacité de diffusion base-émetteur résulte des variations de charge stockée liées aux variations de la tension base-émetteur :

$$C_{DE} = \frac{dQ_S}{dV_{EB}} = \frac{d\tau I_{CC}}{dV_{EB}} = \tau\frac{I_C}{U_T}$$

9.6.4 Capacités de transition

Les capacités de transition C_{TE} et C_{TC} ont des expressions identiques à celles obtenues pour la jonction PN :

$$\boxed{C_{TE} = S\sqrt{\frac{\epsilon q}{2}\frac{N_E N_B}{N_E + N_B}\frac{1}{(\phi - V_{EB})}}} \tag{9.1}$$

$$\boxed{C_{TC} = S\sqrt{\frac{\epsilon q}{2}\frac{N_C N_B}{N_C + N_B}\frac{1}{(\phi - V_{CB})}}} \tag{9.2}$$

N_E, N_C et N_B sont les dopages des régions d'émetteur, de base, et de collecteur, et S est la surface de la région active du dispositif. On obtient le schéma équivalent représenté sur la figure 9.12.

Sur la figure 9.13 est donné à titre de complément le schéma équivalent pour un transistor NPN.

Les schémas que nous venons de présenter modélisent un transistor idéal. Pour des calculs plus précis, on utilise souvent le schéma équivalent représenté sur la figure 9.14 (cas d'un NPN), qui prend en compte :

– l'effet des résistances d'accès R_E, R_B et R_C,
– l'effet de modulation de l'épaisseur de la base par la tension collecteur-base (effet Early) par l'adjonction de la conductance G_a,

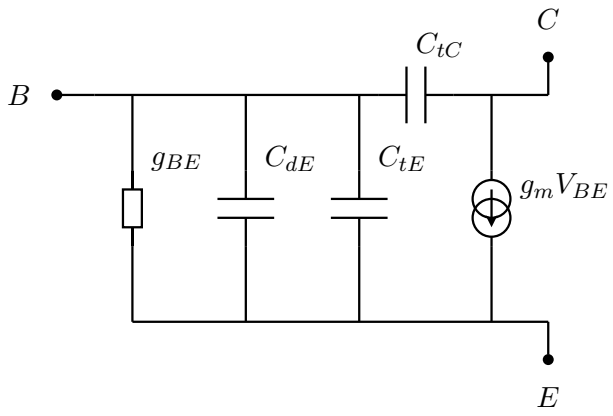


FIG. 9.13 – Schéma équivalent petit signal d'un transistor NPN

– la distribution de la capacité collecteur-base entre la région active et la région passive qui offre souvent la plus grande contribution.

Remarque : Ces mêmes améliorations apportées au modèle non linéaire, et l'inclusion des phénomènes de forte injection permettraient d'obtenir le modèle de Gummel-Poon complet du transistor bipolaire qui est utilisé par la plupart des logiciels de CAO des circuits.

9.7 Le transistor en commutation

On étudie ici le cas d'un transistor NPN.

9.7.1 Mise en conduction

Le transistor initialement bloqué, on applique à l'instant $t = 0$, un échelon de courant de base. Le circuit électrique correspondant est représenté sur la figure 9.15.

Si l'amplitude de E_b est grande devant la tension de diffusion de la jonction émetteur base, l'échelon de tension est transformé en échelon de courant de valeur E_b/R_b. La jonction BE n'injectant des porteurs que lorsque la tension base-émetteur avoisine la tension de diffusion, le courant de collecteur reste faible durant la première phase du transitoire. Cette phase cesse au bout de l'intervalle de temps t_d, qui correspond à la charge de la capacité de transition émetteur de 0 jusqu'à la valeur de la tension de diffusion, et à la charge de la capacité de transition collecteur de $-E_c$ jusqu'à la valeur $-E_c + \phi$. On a donc la relation :

$$t_d \frac{E_b}{R_b} = \int_0^{\phi} C_{TE} dV_{BE} + \int_{-E_C}^{-E_C+\phi} C_{TC} dV_{BC}$$

A partir de l'instant t_d le courant de collecteur commence à croître. Le courant entrant par la base peut se scinder en deux composantes :

– une composante permettant l'augmentation de la charge stockée.
– une composante entretenant le passage du courant de collecteur comme en régime permanent, avec le gain BF,

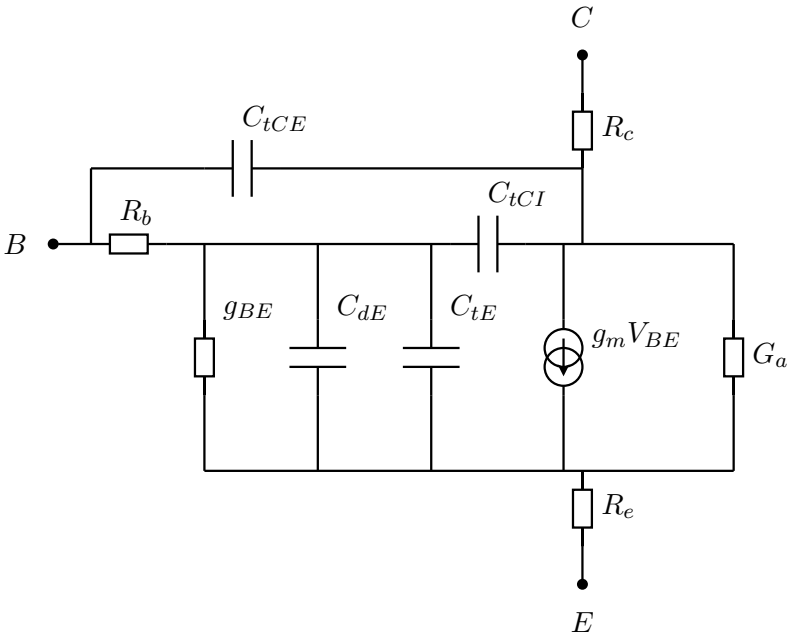


FIG. 9.14 – Schéma petit signal d'un transistor NPN non idéal

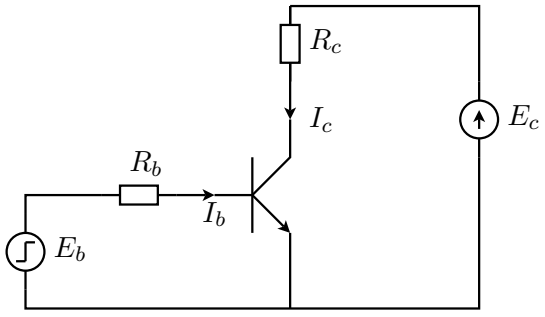


FIG. 9.15 – Transistor NPN en commutation

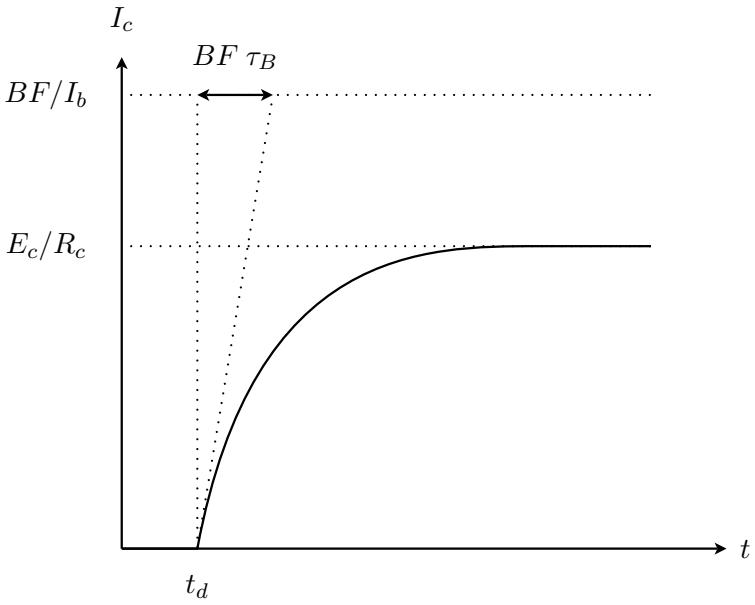


FIG. 9.16 – Évolution du courant de collecteur à la commutation

– la charge stockée étant reliée au courant de collecteur par le temps de transit dans l'hypothèse quasi statique, on a donc la relation :

$$\tau_B \frac{dI_c}{dt} + \frac{I_C}{BF} = \frac{E_b}{R_b}$$

Pendant la deuxième phase du transitoire, le courant de collecteur croit donc exponentiellement avec une constante de temps $BF\tau_B$, vers la valeur$BF\ I_B$.

Remarque : Nous avons négligé dans ce calcul les variations de charge dans les régions dépeuplées.

Si le courant de base est suffisamment important pour assurer la saturation du transistor, la fin du transitoire consistera en une phase de stockage où la valeur du courant collecteur variera peu et lentement vers une valeur très voisine de E_c/R_c. La figure 9.16 montre l'allure des variations du courant collecteur.

9.7.2 Bloquage

Le transistor initialement saturé on applique, à l'instant $t = 0$, un échelon de tension sur la base de valeur $-E_b$. Le circuit électrique reste celui de la figure 9.15. La première phase du transitoire s'appelle la phase de désaturation. pendant cette phase les deux jonctions sont saturées de porteurs et ne peuvent donc pas bloquer de tension inverse. La tension collecteur-émetteur étant très faible en régime saturé, le courant collecteur reste sensiblement inchangé durant toute cette période ($I_c = -E_c/R_c$). La jonction base-émetteur étant elle aussi saturée, le courant de base n'est limité que par le circuit extérieur et vaut $I_b = -E_b/R_b$. Le courant de collecteur étant constant le gradient des porteurs minoritaires l'est aussi dans la région quasi neutre de base, et la charge ΔQ_S qui doit être extraite par le courant de base pour désaturer la jonction de collecteur apparaît clairement sur la figure 9.17. On a donc la relation :

$$t_s I_B = \Delta Q_S$$

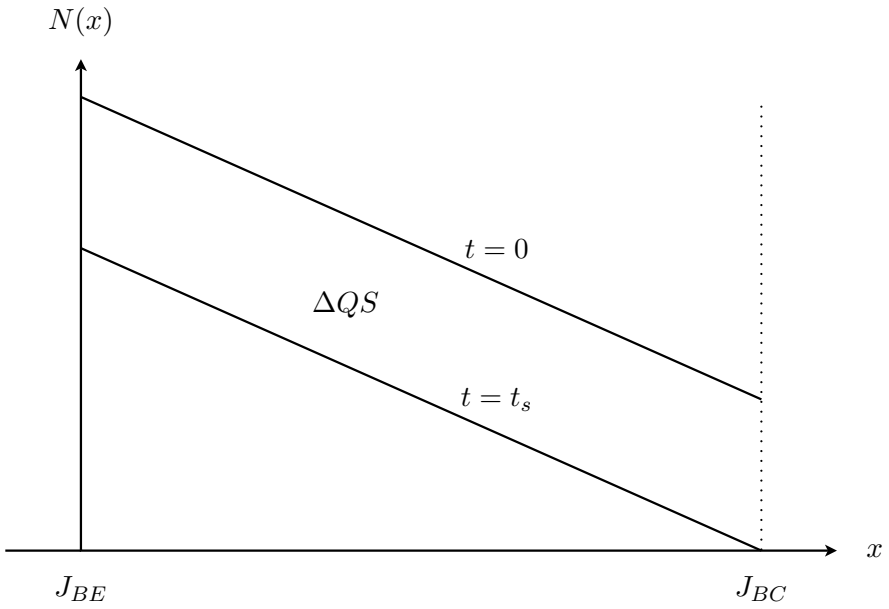


FIG. 9.17 – Variation de la densité de porteur électrons

Une fois la jonction collecteur-base désaturée le transistor se trouve en régime normal et la situation est identique à celle rencontrée au cours de la mise en conduction. La décroissance du courant est donc exponentielle avec une constante de temps $BF\tau_B$. Les formes d'ondes du courant de base et de collecteur son représentées sur la figure 9.18.

9.8 Exercice d'application du chapitre : Analyse d'un transistor bipolaire NPN

Le matériau est du Silicium dont la température est maintenue à 300K, et on prendra :

$$n_i = 10^{10} cm^{-3}$$

$$\epsilon = 10^{-12} F/cm$$

$$k = 1.3810^{-23} J/K$$

Les caractéristiques du transistor sont résumées ci-dessous :

- **Emetteur N^+ :** dopage $N_E = 5\ 10^{18}\ cm^{-3}$; épaisseur $W_E = 10^{-4}\ cm$; constante de diffusion des trous $D_{pE} = 5\ cm^2/s$.
- **Base P :** dopage $N_B = 2\ 10^{17}\ cm^{-3}$; épaisseur $W_B = 0,8\ 10^{-4} cm$; constante de diffusion des électrons $D_{nB} = 20\ cm^2/s$, durée de vie des électrons $\tau_n = 500\ ns$.
- **Collecteur N^-** : dopage $N_C = 10^{16}\ cm^{-3}$; épaisseur $W_C = 10\ 10^{-4}\ cm$; constante de diffusion des trous $D_{pE} = 10\ cm^2/s$.

Surface active : $10^{-3}\ cm^2$; densité de courant nominale $250 A/cm^2$.

On polarise le transistor en régime normal à son courant nominal avec une tension collecteur-base égale à 5 Volts.

1) Calculer les courants de collecteur, de base, et d'émetteur, ainsi que la d.d.p. base-émetteur (on négligera l'extension des régions dépeuplées dans les régions quasi-neutres).

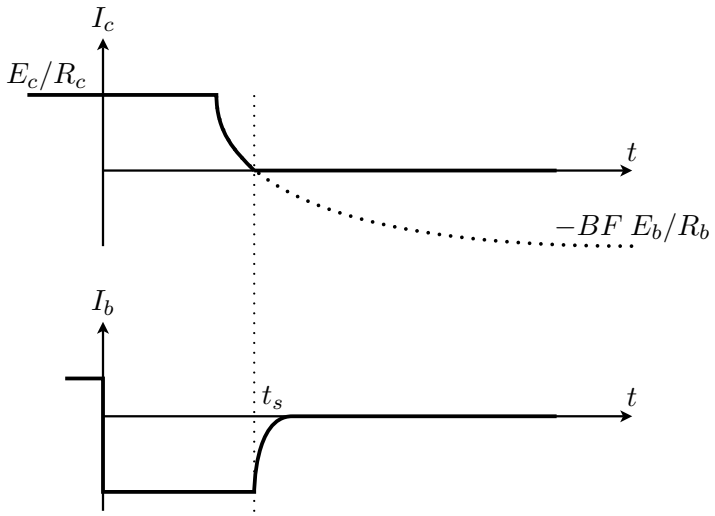


FIG. 9.18 – Évolution de I_C et I_B au blocage

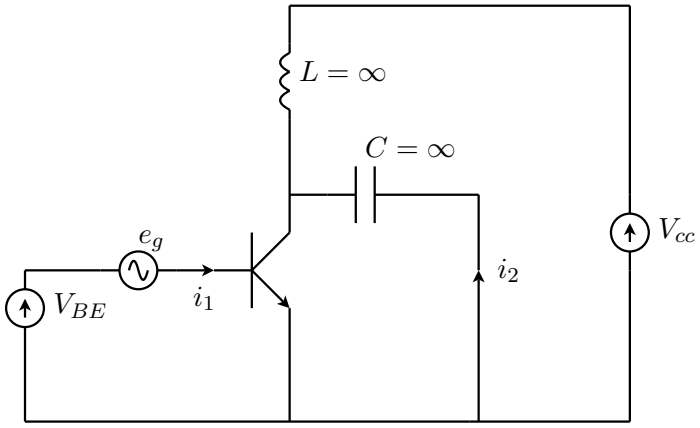


FIG. 9.19 – Montage petit signal

2) Calculer les éléments du schéma équivalent petit signal du modèle de Gummel-Poon simplifié autour de ce point de polarisation.
3) On place un court-circuit entre collecteur et émetteur pour le régime variable, on appelle i_2 le courant circulant dans ce court circuit et i_1 le courant fourni par un générateur alternatif à la base du transistor (cf. figure 9.19 ci-dessous). Donner l'expression du gain en courant i_2/i_1 en fonction des éléments du schéma équivalent. Calculer la fréquence de cassure du gain en courant et la fréquence de transition.
4) Calculer les formes d'ondes (courant de collecteur et de base, tension collecteur-émetteur) correspondant au circuit de la figure 9.20 , la source de tension E_b délivre un échelon de tension de $10,7\ V$ à partir de l'instant $t = 0$. On supposera que :
   – Dans la phase de déblocage du transistor les effets réactifs prépondérants son liés aux variations de charges dans la région dépeuplée de la jonction base-émetteur.
   – Dans la phase de montée du courant les effets réactifs prépondérants sont liés aux variations de charges dans la région quasi-neutre de base.
   – Le passage entre ces deux régimes se produit brutalement lorsque la tension base émetteur atteint $0,7\ V$.
   – Lorsque le transistor conduit sa ddp base-émetteur est égale à $0,7\ V$.

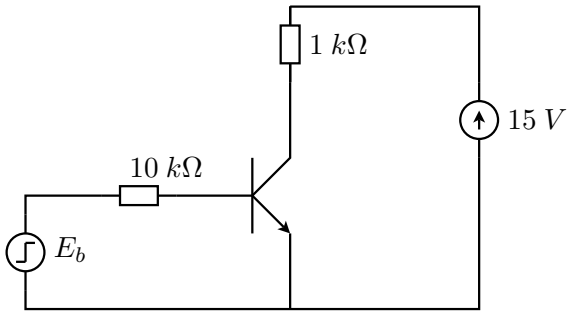


FIG. 9.20 – Commutation

Chapitre 10

Les transistors à effet de champ

10.1 Le transistor à effet de champ de jonction (JFET)

10.1.1 Principe de fonctionnement et structure

La structure idéalisée d'un transistor à effet de champ de jonction (JFET) est représentée sur la figure 10.1. Le principe de ce composant est la modulation de l'épaisseur d'un canal conducteur par la zone dépeuplée d'une jonction PN, la commande s'effectue en tension par l'intermédiaire de l'électrode de grille.. Sur la figure, le canal de type N est constitué par la région N, et la grille par la région P+. Lorsque l'on polarise en inverse la jonction $P^{+}N$ Grille Source, une région dépeuplée envahit une partie du canal qui voit son épaisseur diminuer et donc sa résistance augmenter au fur et à mesure que la DDP inverse grille canal augmente. Si l'autre extrémité du canal appelé Drain est polarisé par rapport à la source, le courant Drain Source sera modulé en intensité par la DDP inverse Grille Source.

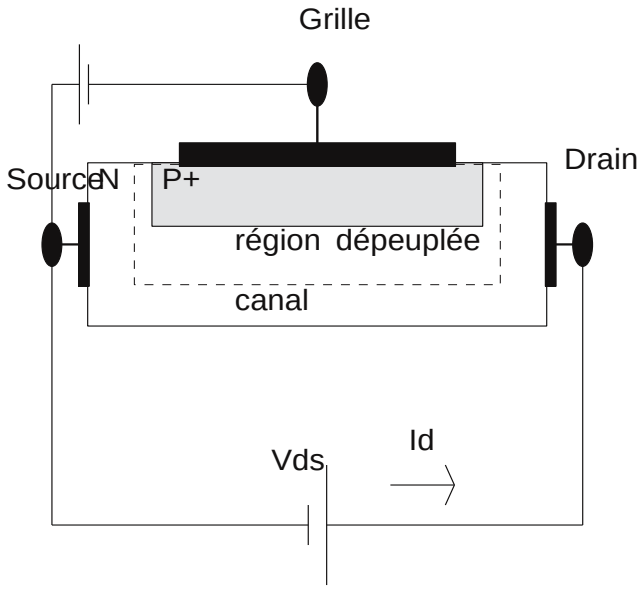


FIG. 10.1 – Structure d'un transistor à effet de champ

10.1.2 Caractéristiques statiques

La figure 10.2 indique les notations utilisées pour le calcul de la caractéristique courant tension du JFET.

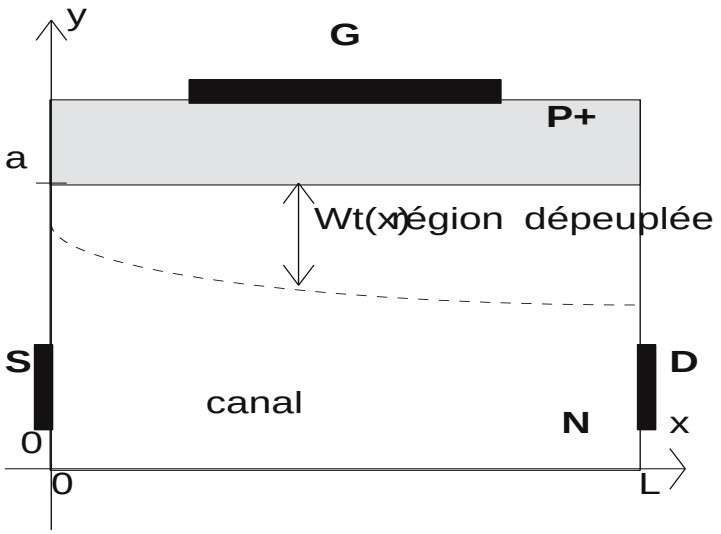


FIG. 10.2 – Extension de la zone dépeuplée

Cas des faibles tensions V_{ds} : régime ohmique

Lorsque la valeur de la tension Drain Source V_{ds} est faible devant la valeur de la tension Grille Source, l'épaisseur W_t de la région dépeuplée est constante tout le long du canal et n'est fonction que de la valeur de V_{gs} :

$$W_t(x) = C^{te} = \sqrt{\frac{2\epsilon}{qN_D}(\phi - V_{gs})} \tag{10.1}$$

ϕ est la tension de diffusion de la jonction Grille canal, Nd est le dopage du canal. La conductance du canal a pour expression :

$$G = \frac{qN_D\mu_n b}{L}(a - W_t) = \frac{qN_D\mu_n ab}{L}[1 - \sqrt{\frac{2\epsilon}{qN_D a^2}(\phi - V_{gs})}] \tag{10.2}$$

On peut réécrire l'expression précédente en faisant apparaître la conductance du canal sans région de charge d'espace G_0 :

$$G_0 = \frac{qN_D\mu_n b}{L} \tag{10.3}$$

et la tension V_p au delà de laquelle tout le canal est déplété et aucun courant I_d ne circule :

$$V_p = \phi - \frac{qN_D a^2}{2\epsilon} \tag{10.4}$$

on obtient alors :

$$\boxed{G = G_0[1 - \sqrt{\frac{\phi - V_{gs}}{\phi - V_p}}]} \tag{10.5}$$

Dans ces conditions ($V_{ds} \ll -V_{gs}$), le courant de Drain I_d croît linéairement avec la tension V_{ds} :

$$I_D = GV_{ds}$$

Tensions V_{ds} intermédiaires

Si la tension V_{ds} n'est pas trop élevée, il existe partout entre source et drain une portion de canal d'épaisseur $W_t(x)$ donnée par la relation :

$$W_t(x) = \sqrt{\frac{2\epsilon}{qN_D}[\phi - (V(G) - V(x))]} = \sqrt{\frac{2\epsilon}{qN_D}[\phi - V_{gs} + V(x)]} \tag{10.6}$$

et la conductance d'une tranche de canal dx vaut :

$$\begin{aligned} G(x) &= \frac{qN_D\mu_n b}{dx}[a - \sqrt{\frac{2\epsilon}{qN_D}[\phi - V_{gs} + V(x)]}] \\ &= G_0[1 - \sqrt{\frac{2\epsilon}{qN_D a^2}[\phi - V_{gs} + V(x)]}]\frac{L}{dx} \\ &= G_0\frac{L}{dx}[1 - \frac{W_t(x)}{a}] \\ &= G_0\frac{L}{dx}[1 - \frac{\sqrt{\phi - V_{gs} + V(x)}}{\sqrt{\phi - V_p}}] \end{aligned}$$

Si dV est la DDP supportée par la tranche dx, on a :

$$I_D = G(V)dV$$

et en intégrant entre la source où $V(x) = 0$, et le drain où $V(x) = V_{ds}$, on obtient :

$$\int_0^L I_D dx = \int_0^{V_{ds}} G(V)dV$$

Le courant de drain I_D étant constant, on obtient :

$$\boxed{I_D = G_0[V_{ds} - \frac{2}{3\sqrt{\phi - V_p}}([V_{ds} + \phi - V_{gs}]^{\frac{2}{3}} - [\phi - V_{gs}]^{\frac{2}{3}})]} \tag{10.7}$$

Lorsque V_{ds} atteint la valeur V_{dsat}, le canal se pince au niveau du drain, on a alors les relations :

$$a = \sqrt{\frac{2\epsilon}{qN_D}(\phi - V_{gs} + V_{dsat})} \tag{10.8}$$

$$\boxed{V_{dsat} = \frac{qN_D a^2}{2\epsilon} - \phi + V_{gs} = V_{gs} - V_p} \tag{10.9}$$

et

$$\boxed{I_{dsat} = G_0(\phi - V_p)[\frac{V_{gs} - V_p}{\phi - V_p} - \frac{2}{3} + \frac{2}{3}(\frac{\phi - V_{gs}}{\phi - V_p})^{\frac{3}{2}}]} \tag{10.10}$$

$V_{ds} > V_{dsat}$: régime saturé

Lorsque la tension V_{ds} est supérieure à la valeur V_{dsat} le canal est pincé à partir d'une certaine abscisse $x = L_{ef}$, la différence de potentiel entre ce point et la source vaut alors V_{dsat}. La disparition du canal n'entraîne pas pour autant la disparition du courant de drain, en effet la différence de potentiel existant dans la zone pincée ($V_{ds} - V_{dsat}$) implique la présence d'un champ électrique dans la direction des x favorable au transfert des électrons vers le drain. La figure 10.3 montre l'allure de la région dépeuplée en régime saturé. L'équation 10.8

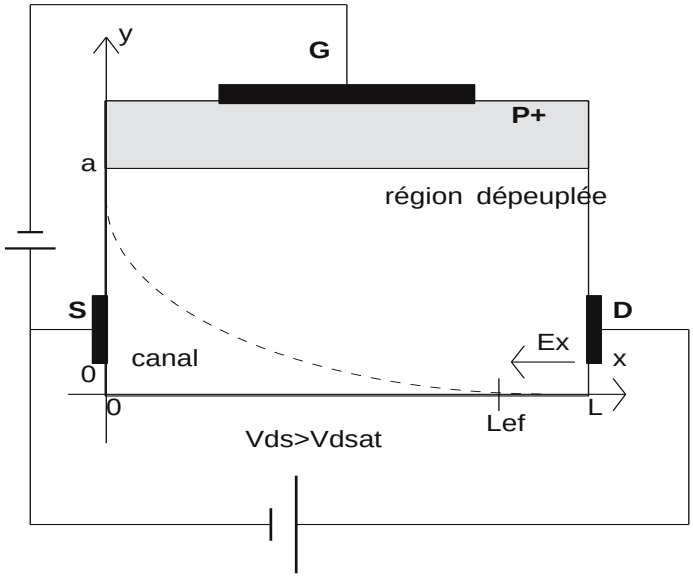


FIG. 10.3 – Transistor FET en régime saturé

qui nous a servi à calculer le courant de drain pour $V_{ds} = V_{dsat}$ reste valable en régime saturé en remplaçant la longueur L du canal par sa longueur effective L_{ef}. Il s'ensuit donc une augmentation progressive de G_0 et donc du courant I_D lorsque V_{ds} croit. En effet, l'abscisse de pincement se rapproche de la source au fur et à mesure que la tension drain croit. Cet effet entraînant une valeur non nulle de la conductance de sortie sera d'autant plus marqué que la longueur initiale du canal est court. Au premier ordre cet effet peut être négligé et on peut approximer la valeur du courant de drain Id en régime saturé par la valeur I_{dsat} qui ne dépend que de V_{gs} et des caractéristiques physiques du composant. La frontière marquant la limite du régime saturé sur les caractéristiques statiques s'obtient en portant la valeur V_{dsat} dans la relation donnant I_D :

$$I_D = G_0[V_{dsat} - \frac{2}{3\sqrt{\phi - V_p}}([V_{dsat} + \phi - V_{gs}]^{\frac{2}{3}} - [\phi - V_{gs}]^{\frac{2}{3}})] \tag{10.11}$$

La figure 10.4 montre les caractéristiques statiques calculées à partir des relations que nous venons d'établir.

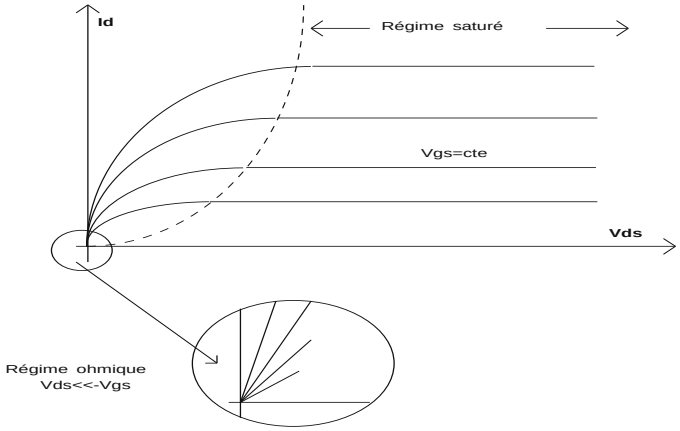


FIG. 10.4 – Caractéristiques statiques

10.1.3 Caractéristiques dynamiques

Nous nous limiterons au modèle équivalent petit signal en régime saturé qui découle de la linéarisation des équations précédemment établies pour ce régime, et de l'effet capacitif dû aux variations de la région dépeuplée avec les tensions de commande V_{gs} et V_{ds}. Le schéma équivalent du transistor intrinsèque (sans éléments parasites) est représenté sur la figure 10.5 dans la configuration source commune.

La transconductance G_m se calcule à partir de l'expression du courant de drain en régime saturé 10.9 :

$$G_m = \frac{\partial I_D}{\partial V_{gs}} = G_0(1 - \sqrt{\frac{\phi - V_{gs}}{\phi - V_p}}) \tag{10.12}$$

La conductance de sortie est nulle si l'on ne prends pas en compte les variations de la longueur effective du canal comme c'est le cas dans l'expression 10.9. Les capacités C_{gs} et C_{gd} résultent des variations de charge de la région dépeuplée Q, qui sont commandées par les tensions V_{gs} et V_{ds}. On a les relations :

$$C_{gs} + C_{gd} = \frac{\partial Q}{\partial V_{gs}}|_{V_{ds}=C^{te}} \quad \text{et} \quad C_{gd} = \frac{Q}{V_{ds}}|_{V_{gs}=C^{te}} \tag{10.13}$$

L'expression de la charge Q en fonction des tensions de commande étant difficile à établir, nous ne développerons pas le calcul et nous nous bornerons à des résultats qualitatifs : en régime saturé C_{gs} est généralement bien plus grande que C_{gd}, et ces capacités sont en première analyse proportionnelles à la surface bL de grille et inversement proportionnelles à l'épaisseur a du canal.

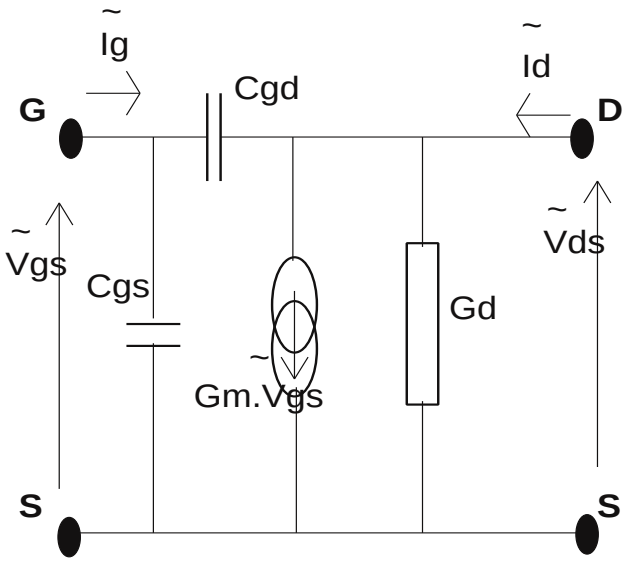


FIG. 10.5 – Schéma équivalent petit signal

10.2 Le transistor mos (metal oxyde semi-conducteur)

Nous limiterons notre étude au transistor MOS à enrichissement, le fonctionnement des MOS à déplétion étant proche de celui du transistor JFET.

10.2.1 Principe de fonctionnement et structure

La structure d'un transistor MOS est représentée sur la figure 10.6.

En l'absence d'une polarisation V_{gs} adéquate, il existe toujours une jonction polarisée en inverse entre la source et la drain, par conséquent aucun courant significatif ne peut circuler entre ces deux électrodes et le transistor est bloqué. Lorsqu'on applique une tension grille source positive la capacité métal oxyde semi-conducteur se charge et une charge négative apparaît dans le semi-conducteur, cette charge est fournie pour les faibles valeurs de V_{gs} par les charges des accepteurs ionisés d'une région dépeuplée qui apparaît sous l'oxyde. Lorsque la tension V_{gs} dépasse une certaine valeur appelée tension de seuil V_t, il apparaît sous la grille une charge de porteurs mobiles concentrée à la surface du semi-conducteur. Cette charge d'électrons permet alors le passage d'un courant entre source et drain.

10.2.2 Inversion des populations de porteurs en surface d'une structure MOS idéale

La figure 10.7(a) représente le schéma des bandes d'énergie d'une structure MOS à l'équilibre en l'absence de charges piégées dans l'oxyde et aux interfaces, dans une structure idéale où le travail de sortie du métal ϕ_m serait tel que le niveau du vide E_0 serait constant dans toute la structure. On est alors dans la situation de "bandes plates".

Si l'on applique une d.d.p. V_G positive, il y a décalage des niveaux de Fermi et apparition d'une charge positive sur l'électrode métallique et négative dans le semi-conducteur. Malgré

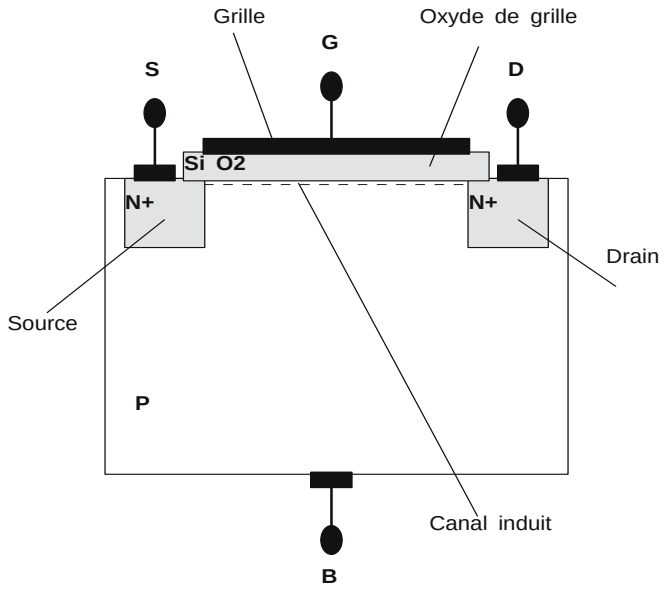


FIG. 10.6 – Structure d'un transistor MOS

la polarisation, la présence de l'isolant permet de considérer le niveau de Fermi constant dans le semi-conducteur ainsi que d'utiliser les expressions des densités de porteurs établies pour l'équilibre thermodynamique :

$$N = N_C e^{-\frac{E_c - E_F}{kT}}$$

$$P = N_V e^{-\frac{E_F - E_v}{kT}}$$

On constate alors, en utilisant ces expressions, la présence d'une région dépeuplée dans le semi-conducteur sur la figure 10.7(b). Si l'on augmente V_G au delà d'un certain seuil, la différence énergétique $E_c - E_F$ à la surface du semi-conducteur devient comparable à la différence $E_F - E_v$ en volume, et la concentration en minoritaires à la surface devient égale à la concentration en majoritaires dans le volume ; on dit que l'on a atteint le seuil d'inversion. La d.d.p. supportée par le semi-conducteur vaut alors :

$$\frac{1}{q}(E_{cvolume} - E_{csurface}) = kT \ln \frac{N_a^2}{N_i^2}$$

Au delà de ce seuil, la dépendance exponentielle des distributions de porteurs avec l'énergie, fait que l'on peut approximer ces régions inversées par des distributions surfaciques de charges et négliger leur chute de potentiel. La tension supportée par le semi-conducteur est alors constante et vaut Φ.

10.2.3 Calcul de la caractéristique courant tension d'un MOS idéal

La figure 10.8 indique les notations utilisées pour le calcul de la caractéristique courant tension. On notera $W_t(x)$ l'épaisseur de la région dépeuplée à l'abscisse x, ϕ la tension de

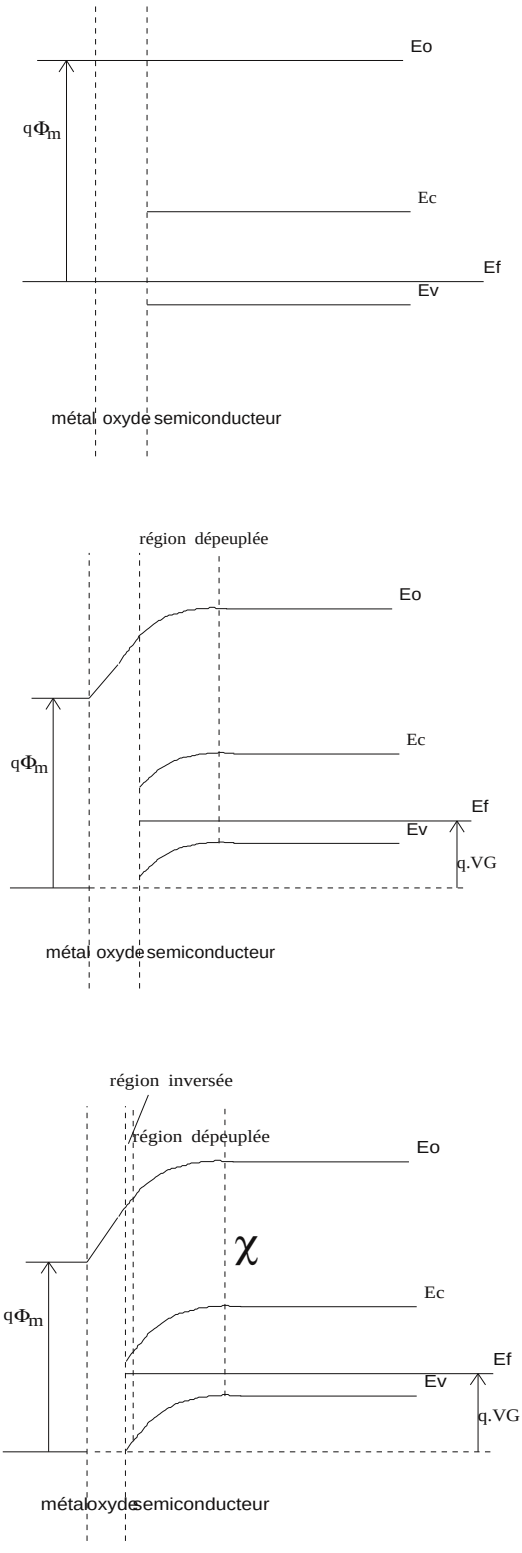


FIG. 10.7 – Schéma des bandes

diffusion de la jonction N^+P, $V_{sc}(x)$ la ddp supportée par le semi-conducteur à l'abscisse x, et V_{sb} la polarisation source substrat.

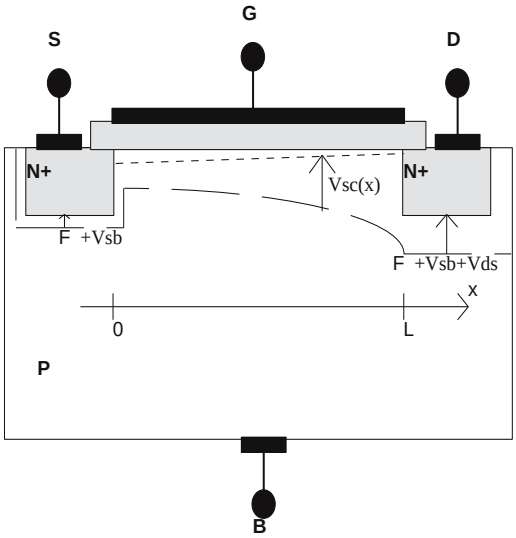


FIG. 10.8 – Structure du transistor MOS

La différence de potentiel entre la grille G et le substrat B est supportée pour partie par l'oxyde et pour partie par le semi-conducteur :

$$V_G - V_B = V_{ox}(x) + V_{sc}(x)$$

La densité de charge Q portée par le semi-conducteur à l'abscisse x vaut :

$$Q(x) = -C_{ox}V_{ox}(x) = -C_{ox}[V_G - V_B - V_{sc}(x)] \tag{10.14}$$

$C_{ox} = \epsilon_{ox}/W_{ox}$ est la capacité d'oxyde par unité de surface, ϵ_{ox} la permittivité de l'oxyde, et W_{ox} son épaisseur. La région dépeuplée supportant la quasi totalité de la ddp dans le semi-conducteur on peut écrire :

$$W_t(x) = \sqrt{\frac{2\epsilon V_{sc}(x)}{qN_a}} \tag{10.15}$$

La charge dans le semi-conducteur se décompose en une charge fixe Q_f due aux impuretés dans la région dépeuplée :

$$Q_f(x) = -qN_aW_t(x)$$

et en une charge mobile d'électrons Q_{mob} :

$$Q_{mob} = Q - Q_f = -C_{ox}[V_G - V_B - V_{sc}(x) - \frac{\sqrt{2\epsilon qN_aV_{sc}(x)}}{C_{ox}}] \tag{10.16}$$

Cette charge mobile existe au niveau de la source, où la ddp supportée par le semi-conducteur vaut $\phi - V_{sb}$, si la tension grille source V_{gs} est supérieure à la valeur seuil V_t :

$$V_{gs} = V_T = \phi + \frac{\sqrt{2\epsilon qN_a(\phi + V_{sb})}}{C_{ox}} \tag{10.17}$$

On constate sur cette expression que la ddp source substrat V_{sb} permet de moduler la tension de seuil d'un transistor MOS. La résistance élémentaire du canal à l'abscisse x a donc pour expression :

$$dR = \frac{dx}{\mu_n W Q_{mob}(x)} \tag{10.18}$$

si W est la largeur du canal. Un calcul similaire à celui effectué pour le JFET conduit à l'expression du courant de drain. Dans le cas particulier où la source et le substrat sont reliés, on obtient :

$$\boxed{I_D = \frac{\mu_n W C_{ox}}{L}[(V_{gs} - \phi - \frac{V_{ds}}{2}) - \frac{2\sqrt{\epsilon q N_a}}{3C_{ox}}((\phi + V_{ds})^{\frac{3}{2}} - \phi^{\frac{3}{2}})]} \tag{10.19}$$

cette expression étant valable en régime non pincé, c'est à dire si le canal existe partout entre grille et drain. Lorsque V_{ds} est égale à la valeur V_{dsat} :

$$\boxed{V_{Dsat} = V_{gs} - \phi + \frac{\epsilon q N_a}{C_{ox}^2}[1 - \sqrt{1 + \frac{2C_{ox}^2}{\epsilon q N_a} V_{gs}}]} \tag{10.20}$$

le canal se pince en $x = L$, et pour les valeurs de V_{ds} supérieures à V_{dsat} le courant sature à la valeur $I_{dsat} = Id(V_{dsat})$. L'allure des caractéristiques de sortie $Id(V_{ds}, V_{gs})$ est représentée sur la figure 10.9.

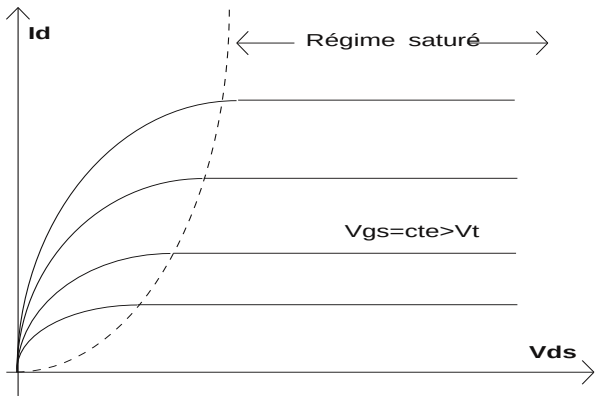


FIG. 10.9 – Caractéristiques statiques

10.2.4 Modèle équivalent simplifié

Regime statique

On peut simplifier les équations établies précédemment avec un bonne approximation en utilisant les expressions suivantes :

◇ Pour les très faibles valeurs de V_{ds} ($V_{ds} \ll \phi$ et $V_{vds} \ll V_{gs} - V_T$

$$I_D \approx \frac{\mu_n W C_{ox}}{L}(V_{gs} - V_T)V_{ds} \tag{10.21}$$

c'est le régime linéaire

◇ Pour les faibles valeurs de V_{ds} ($V_{ds} \ll \phi$)

$$I_D \approx \frac{\mu_n W C_{ox}}{L}(V_{gs} - V_T - \frac{V_{ds}}{2})V_{ds} \tag{10.22}$$

◇ En régime saturé ($V_{ds} > V_{dsat} = V_{gs} - V_T$

$$I_D \approx \frac{\mu_n W C_{ox}}{2L}(V_{gs} - V_T)^2 \tag{10.23}$$

Schéma equivalent dynamique

Le schéma équivalent du MOST est représenté sur la figure 10.10. La topologie est semblable à celle obtenue pour le JFET, avec en plus une capacité drain source C_{ds} dont l'origine est en grande partie liée à la région dépeuplée de la jonction drain substrat. La transconductance G_m s'obtient facilement à partir des expressions précédentes, tandis que les valeurs des capacités sont difficiles à obtenir analytiquement en raison du caractère bidimentionnel de phénomènes physiques mis en jeu dans ce composant. Toutefois on peut obtenir l'ordre de grandeur des capacités C_{gs} et C_{gd} en négligeant la capacité due au semi-conducteur qui se trouve en série avec la capacité d'oxyde :

$$C_{ox}WL = C_{gs} + C_{gd}$$

10.2.5 Structure MOS réelle

Dans une structure MOS réelle, il faut tenir compte des charges piégées dans l'oxyde et aux interfaces, ainsi que de la différence des travaux de sortie entre le métal et le semiconducteur. Lorsque la tension V_{gs} est nulle on n'est plus dans la situation des "bandes plates" comme c'était le cas pour une structure idéale. Pour retrouver cette situation, on doit appliquer entre grille et source la ddp V_{FB}. On peut tenir compte de ces défauts en modifiant l'expression de la tension de seuil :

$$V_T = V_{FB} + \phi + \frac{\sqrt{2\epsilon N_a(\phi + V_{sb})}}{C_{ox}} \tag{10.24}$$

Remarque : Dans le procédé de fabrication d'un TMOS, on ajuste toujours la tension de seuil à la valeur désirée par une implantation précise de charges.

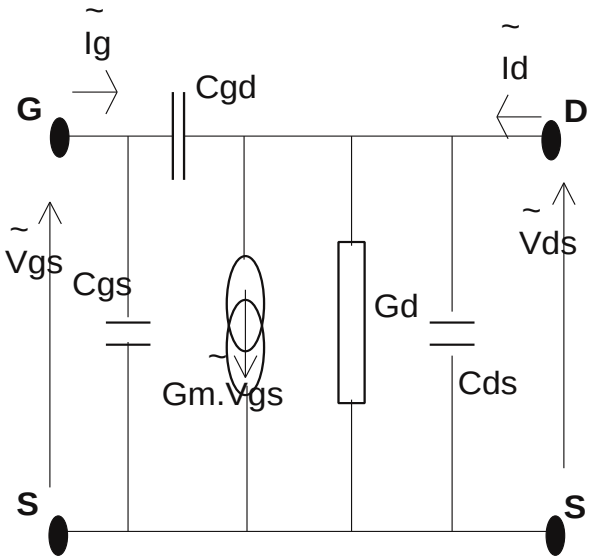


FIG. 10.10 – Schéma équivalent petit signal du MOST

10.3 Exercices d'application du chapitre

◇ I

Un transistor JFET à les caractéristiques suivantes : épaisseur du canal : a=2.5μm, longueur du canal :L=10μm, largeur du canal :b=100μm région de grille : Na=10^{18}cm^{-3} ; région de canal : Nd=10^{15}cm^{-3}, μn=1000cm^2/V.s. N$_i$=10^{10}cm^{-3}, ϵ= 10^{-12}F/cm

1 Calculer la valeur de la tension de pincement Vp, et de la conductance du canal sans charge d'espace Go.

2 Tracer la caractéristique Id=f(V_{ds}) pour les valeurs de V_{gs} suivantes : 0,-1, -2, -3, -4V, (0<V_{ds}<5V avec un pas de 1Volt).

3 Pour V_{ds}=5V, tracer la caractéristique de transfert Id = f(V_{gs}). Quelle valeur de V_{gs} donne une transconductance maximale, calculer sa valeur.

4 Calculer la somme des capacités Cgs et Cgd en régime linéaire pour V_{gs}=0. **5** On se place à V_{ds}=5V et V_{gs}=0V, et on suppose : Cgd/Cgs = 1/10 ; Cgd + Cgs = 0.5 x (valeur de Cgs+Cgd en régime linéaire pour le même V_{gs})

- Calculer les éléments du schéma équivalent petit signal.

- Calculer la fréquence de transition Ft (fréquence où le gain en courant est égal à l'unité).

- Exprimer la valeur de Ft en fonction des paramètres technologiques et discuter les solutions pour améliorer les performances fréquentielles du composant.

◇ II

Etude de l'inverseur CMOS

Le schéma électrique de l'inverseur et les notations utilisées sont donnés sur la figure ci-dessous.

Les caractéristiques des transistors sont les suivantes :

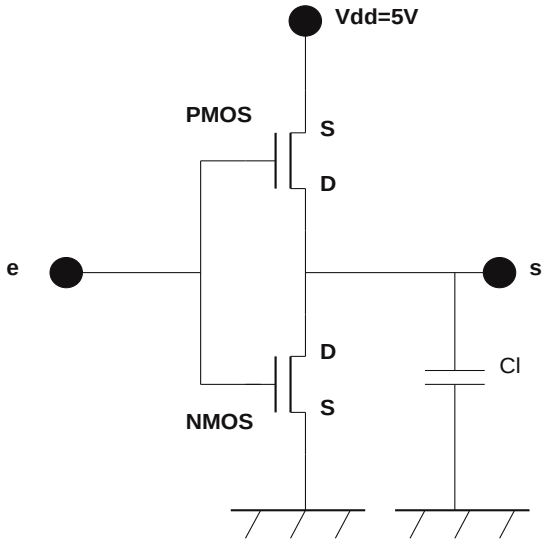


FIG. 10.11 – Schéma électrique

Cox=50nF/Cm^2. **NMOS** : V_{TN}=1V ; μ_N=500Cm^2/Vs ; W_N/L_N=1 **PMOS** : V_{TP}=-1V ; μ_P=250Cm^2/Vs ; W_P/L_P=2

1 Tracer les caractéristiques statiques du NMOS $I_{DN}(V_S)$ pour V_E=0 ; 1 ; 2 ; 3 ; 4 ; 5.

2 Tracer sur le même graphe les caractéristiques statiques du PMOS $I_{DP}(V_S)$ pour V_E=0 ; 1 ; 2 ; 3 ; 4 ; 5.

3 Déduire du graphe précédent la caractéristique de transfert de l'inverseur V_S=f(V_E)

4 Calculer la tension de commutation de l'inverseur V_C définie V_S=V_E. Que vaudrait elle si W_P/L_P=1 ?.

5 L'entrée de l'inverseur étant attaqué par un signal carré 0 5V, de fréquence 10Mhz, calculer les formes d'ondes en sortie de l'inverseur pour une capacité Cl de 0,1pF.

Annexe A

Éléments de fabrication des circuits intégrés

L'objectif de ce chapitre est de donner quelques notions sur les problèmes relatifs à la technologie des semi-conducteurs, la fabrication des circuits intégrés, la filière technologique CMOS.

A.1 Opérations technologiques élémentaires

Pour fabriquer un circuit intégré on doit être capable de réaliser à partir d'une plaquette de semi-conducteur monocristallin,

- des régions de semi-conducteur de dopages différents en type et en concentration ;
- des couches d'interconnexions métalliques ou semiconductrices permettant de relier les différents composants élémentaires à l'intérieur du circuit intégré ;
- des couches isolantes.

Dans ce paragraphe, nous décrirons succinctement les procédés de réalisation de ces différents constitutifs élémentaires du circuit intégré.

A.2 Réalisation du substrat semi-conducteur

Il s'agit d'obtenir, à partir de quartz, un cristal de silicium très pur avec très peu d'atomes d'impuretés électriquement actives, ainsi qu'une structure cristalline quasi exempte de défauts d'empilement des atomes. On réduit d'abord du quartz dans un four à arc par du carbone :

$$SiO_2 + 2C \rightarrow Si + 2\,CO$$

Le silicium obtenu, pur à 99%, est ensuite purifié par différents procédés chimiques jusqu'à l'obtention d'un silicium polycristallin très pur chimiquement (quelques 10^{-9}, résistivité>200Ω.cm). Le monocristal de silicium est obtenu par solidification du matériau précédent aux environs de 1400°C sur un germe de monocristal. On obtient alors un lingot dont le diamètre peut atteindre 20cm, et qui peut être dopé uniformément durant cette phase. Le lingot est ensuite découpé en tranches, puis poli. On obtient alors des plaquettes de semi-conducteur dont le diamètre maximal atteint 20cm environ, et ayant une épaisseur de quelques centaines de microns, aptes à la réalisation de circuits intégrés.

A.3 Réalisation des couches de semi-conducteur

Pour réaliser des régions de semi-conducteur de dopages, de types, et d'épaisseurs souhaitées on peut procéder par :

- diffusion de dopants ;
- implantation ionique de dopants ;
- croissance cristalline d'une couche de semi-conducteur à partir du substrat.

A.3.1 La diffusion

Le processus de diffusion est le procédé le plus courant pour obtenir une jonction. Les atomes d'impuretés dopantes du silicium (B, As, Ph,..) ont une constante de diffusion très faible aux températures ambiantes, pour les faire pénétrer dans le cristal de silicium à partir de la surface, on doit procéder dans un four à température élevée (1200°C typique), avec des durées de diffusion d'autant plus grandes que la profondeur souhaitée est importante (plusieurs dizaines de minutes pour des profondeurs de quelques microns). La concentration des dopants diminue de la surface vers le volume selon une loi en erfc ou Gaussienne selon que la diffusion s'effectue à concentration de surface constante ou quantité de dopants constante. Le procédé de diffusion s'effectuant à haute température, toutes les impuretés présentes dans le cristal diffuseront simultanément, et les jonctions précédemment réalisées verront leurs caractéristiques changer.

A.3.2 L'implantation ionique

L'implantation ionique consiste à accélérer des ions dopants et à les projeter sur la surface du substrat. La profondeur de pénétration des ions qui dépend de leur énergie, et le nombre total d'ions implantés qui dépend de la durée d'irradiation sont parfaitement maîtrisés. On peut, par cette technique, réaliser aisément des jonctions peu profondes (<0,5 micron) et très dopées qu'il est difficile de réaliser par diffusion. l'arrêt des ions dans le cristal s'effectuant par chocs sur les atomes du cristal, le réseau cristallin est très perturbé et il est nécessaire d'effectuer un recuit rapide (quelques secondes vers 1000°C) pour réarranger les atomes du cristal et les atomes dopants. De plus, ces temps très courts et ces températures plus faibles sont un avantage en ce qui concerne la rapidité de réalisation et la non modification des couches existantes.

A.3.3 L'epitaxie

Le processus d'épitaxie consiste à faire croître à la surface du substrat une couche mince de semi-conducteur. On distingue deux grands type d'épitaxie, les dépôts par réaction chimique en phase vapeur (CVD : Chemical Vapor Deposition), et les dépôts par évaporation ou projection moléculaire (PVD : Physical Vapor Deposition).

Les CVD sont les plus courants et les plus faciles à mettre en oeuvre, des composés gazeux de silicium (SiH4, SiH2Cl2,..) se décomposent, et une couche de silicium monocristallin se dépose à la surface du substrat avec une vitesse de l'ordre du micron par minute. On peut doper ces couches (uniformément ou non) en ajoutant un dopant dans les gaz. Les températures nécessaires dépassent 1000°C à la pression atmosphérique, mais peuvent être considérablement réduites à des pressions plus basses ou avec l'assistance de plasmas(PECVD).

Les PVD, très difficiles à mettre en oeuvre sont utilisés pour les composants microondes et optoélectroniques avec des semi-conducteurs composés, ou pour déposer des couches minces magnétiques ou piézo-électriques.

A.4 Réalisation de couches isolantes

Les couches isolantes sont réalisées le plus souvent par oxydation du silicium, ou parfois par dépôt de matériau isolant.

La plupart des couches isolantes sont réalisées directement avec l'oxyde de silicium SiO2 par oxydation "sèche" ou "humide". L'oxyde sec s'obtient par oxydation thermique($900^{\circ}C < T < 1100^{\circ}C$) du silicium sous atmosphère d'O2. On obtient de cette manière un isolant de très bonne qualité électrique (utilisé comme diélectrique de grille dans les TMOS) mais de croissance lente réservé aux couches minces. Sous atmosphère de vapeur d'eau on obtient thermiquement l'oxyde "humide" qui croit plus vite, mais de moindre qualité électrique il est réservé à la réalisation de couches plus épaisses isolant les régions actives du semiconducteur des interconnexions.

Pour réaliser une couche isolante qui n'est pas en contact avec le substrat on peut réaliser une couche d'oxyde par dépôt CVD, ou déposer du nitrure de silicium Si_3N_4 par CVD ou PECVD à basse température si des réalisations antérieures ne supportent pas des températures élevées(connexions en aluminium)

A.5 Réalisation des interconnexions

Les interconnexions sont réalisées par dépôt PVD (évaporation ou pulvérisation cathodique) de matériaux conducteurs (Al pour les longues distances, W pour les distances moyennes,..) ou par croissance épitaxiale de silicium polycristallin pour les interconnexions à courte distance. Le Si polycristallin sert aussi pour la réalisation des grilles des TMOS, de résistances, d'armatures de condensateurs.

A.6 Lithographie et gravure des couches

Pour fabriquer un circuit intégré il faut être capable de définir de manière précise la géométrie des différentes couches que nous venons de décrire, c'est le rôle de la lithographie et de la gravure. Par exemple, pour obtenir une jonction PN de géométrie bien définie, on peut à partir d'un substrat N (Cf. Figure A.1 à figure A.3) :

- faire croître une couche d'oxyde épais sur toute la surface du substrat (a) ;
- déposer sur cet oxyde une couche de laque photosensible appelée résist (b) ;
- cacher par un masque le motif désiré et éclairer par une lumière de longueur d'onde appropriée. La région insolée devient alors sensible à un solvant chimique et peut être facilement dissoute (c,d) ;
- graver la couche d'oxyde non protégée par le résist, par attaque chimique de la couche d'oxyde (e) ;
- faire diffuser les dopants P, la couche d'oxyde servant de masque à la diffusion (f).

On peut de la même manière délimiter géométriquement les couches d'interconnexions et les couches isolantes. Pour des questions de résolution, la lithographie optique qui est limité autour du micron en raison de la diffraction de la lumière, est souvent remplacée par une ``insolation`` du résist à l'aide d'un faisceau d'électrons piloté par ordinateur ou par une

irradiation avec des rayons X. On atteint avec ces techniques des résolutions submicroniques nécessaires pour la fabrication des circuits intégrés actuels.

A.7 Étapes technologiques de la filiere cmos

Nous allons décrire dans ce paragraphe les principales étapes technologiques permettant la réalisation des circuits intégrés CMOS. Les figures A.4 à A.8 résume schématiquement ces étapes :

- Le substrat de base peut être de type P avec un dopage de l'ordre de 10^{16}cm^{-3} (a).
- On procède à la réalisation de caissons de type N qui serviront de substrat locaux pour la réalisation des PMOS. On procède par implantation ionique suivi d'un recuit à haute température ; on obtient alors une concentration de quelques 10^{16}cm^{-3} sur une épaisseur de quelques microns (5μm typique) (b).
- On effectue ensuite une oxydation localisée d'isolement ($0{,}85\mu$m typique) (c).
- On réalise ensuite par oxydation sèche l'oxyde de grille des MOS (quelques dizaines de nanomètres typique). Cette étape doit être particulièrement soignée pour éviter une densité d'états d'interface trop importante (d).
- On procède alors à la réalisation des grilles des transistors en polysilicium (quelques centaines de nanomètres). la définition géométrique submicronique est atteinte par gravure sèche (assistée par plasma), et ce type de grille supporte des étapes ultérieures à haute température (e).
- On effectue successivement les implantations de source et de drain des NMOS et des PMOS (f, g).
- On effectue ensuite une oxydation à basse température permettant d'isoler la couche de polysilicium ainsi que de planariser la surface du circuit intégré. On retire alors l'oxyde pour l'ouverture des contacts de drain et de source (h).
- On réalise ensuite le premier niveau d'interconnexion métallique (i). On procède alors au dépôt d'une couche isolante, et ensuite à la réalisation d'une deuxième couche d'interconnexion généralement en aluminium (non représentés sur la figure).
- On procède enfin au dépôt d'une couche passivante (nitrure typiquement) (j). -Il reste enfin à graver les plots de contact qui serviront pour les connexions du circuit avec le boîtier, puis, après découpe de la plaquette, on obtient chaque circuit intégré qui est testé et monté individuellement.

A.8 Remarques

- De nombreuses étapes technologiques ont été passées sous silence, en particulier des implantations d'ions pour ajuster les tensions de seuil de chaque type d'interconnexions.
- Il y souvent plusieurs niveaux de polysilicium et d'interconnexions ce qui permet de réaliser des capacités et des résistances performantes et la filière technologique CMOS peut être utilisée pour la fabrication de circuits analogiques ou mixtes.
- A partir de la filière CMOS à été développée la filière BiCMOS qui combine sur une même plaquette de semi-conducteur les avantages de la filière CMOS (faible consommation, haute densité d'intégration,..) aux avantages des C.I. bipolaires (rapidité, aptitude à la réalisation de circuits analogiques,..).
- La mise au point d'une filière technologique performante demande des années de recherche et de développement (une dizaine d'années), et des équipements technolo-

giques extrêmement lourds. La durée de vie d'une filière avoisine 5 ans, et une nouvelle génération apparaît chaque 2 ans environ. Les améliorations obtenues entre deux filières successives résultent généralement d'une réduction des dimensions physiques des composants (actuellement la longueur du canal des MOS est inférieure à 0,5μm), ce qui entraîne une augmentation des densités d'intégration (plusieurs dizaines de millions de composants par puces actuellement) et des performances des circuits.

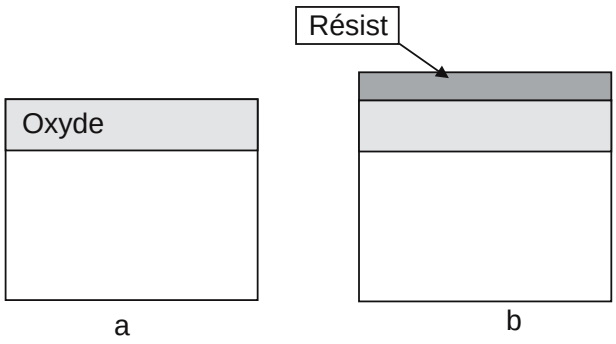


FIG. A.1 – Lithographie et gravure des couches (1)

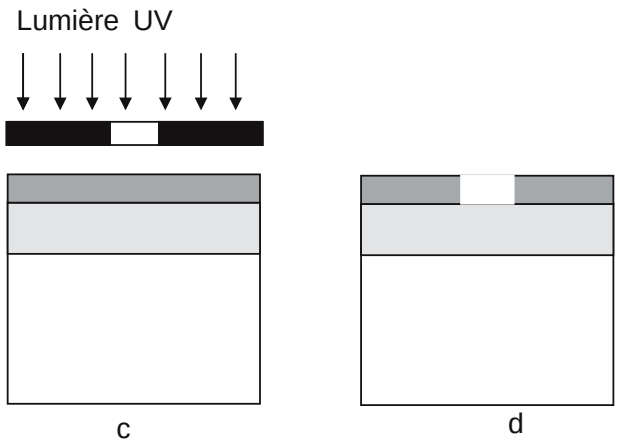


FIG. A.2 – Lithographie et gravure des couches (2)

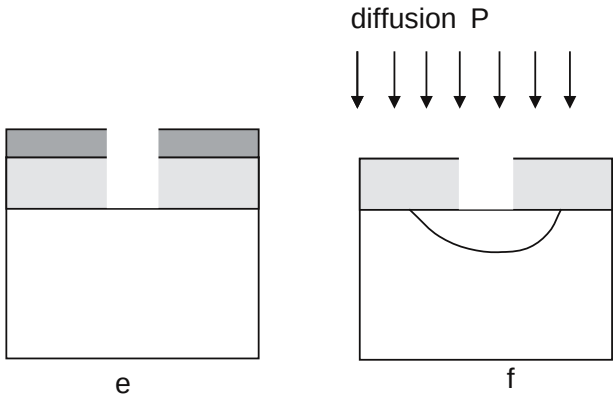


FIG. A.3 – Lithographie et gravure des couches (3)

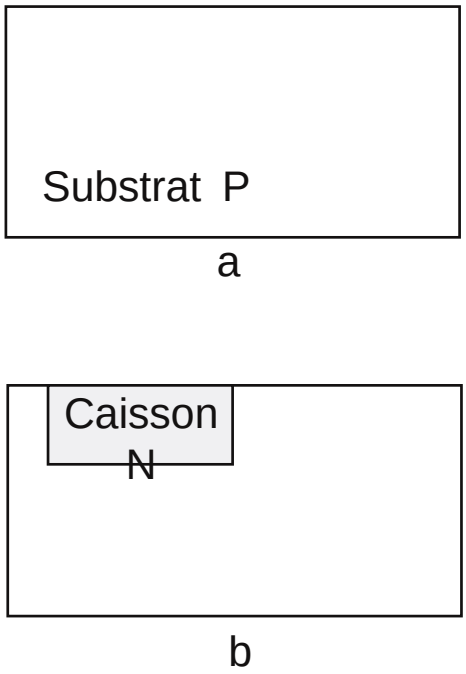


FIG. A.4 – Réalisation d'un CMOS (1)

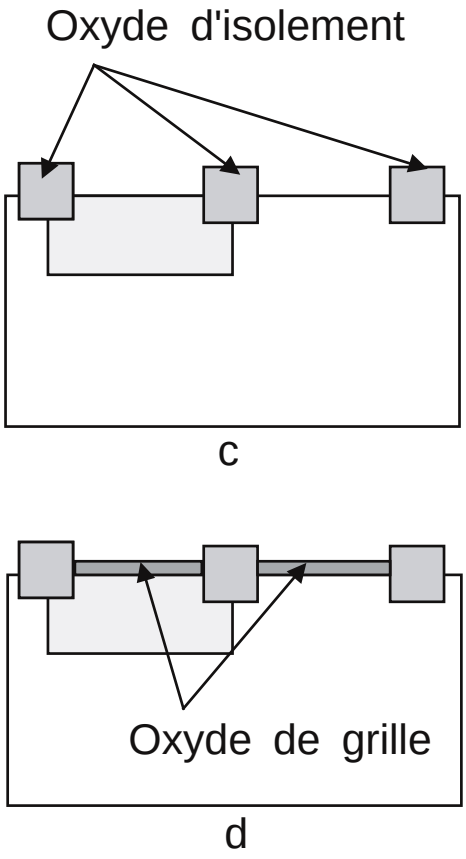


FIG. A.5 – Réalisation d'un CMOS (2)

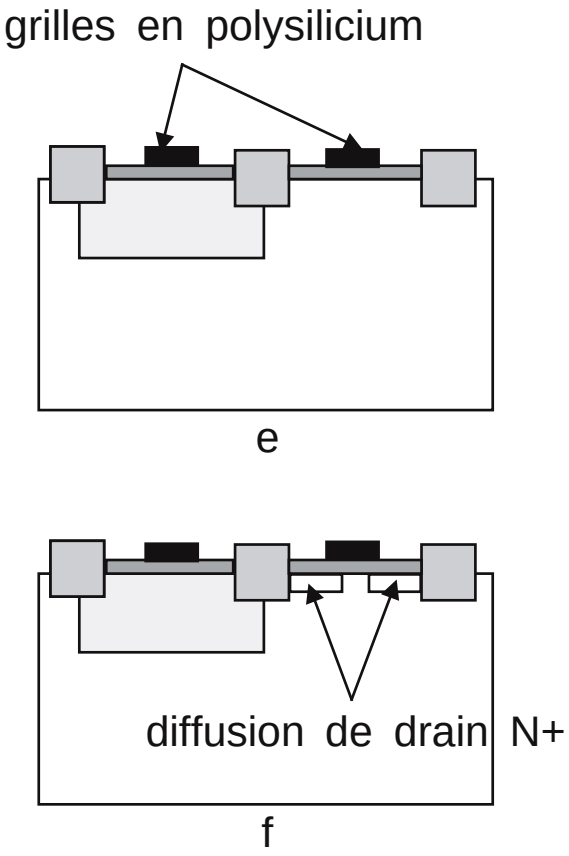


FIG. A.6 – Réalisation d'un CMOS (3)

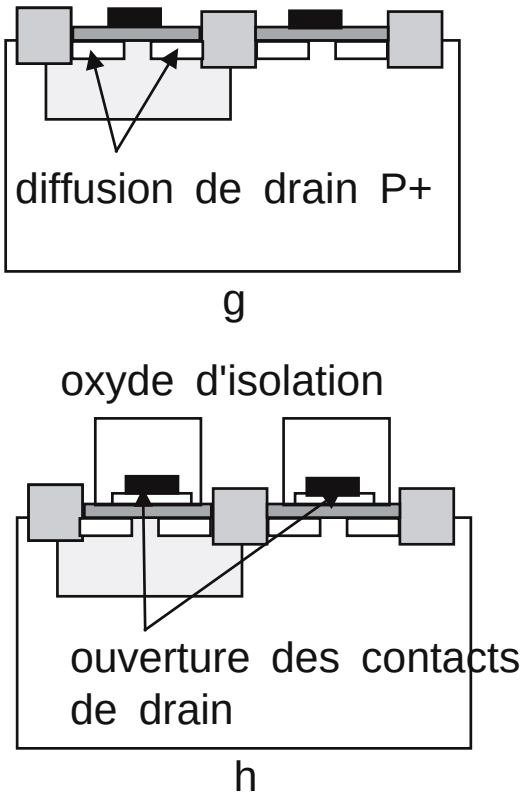


FIG. A.7 – Réalisation d'un CMOS (4)

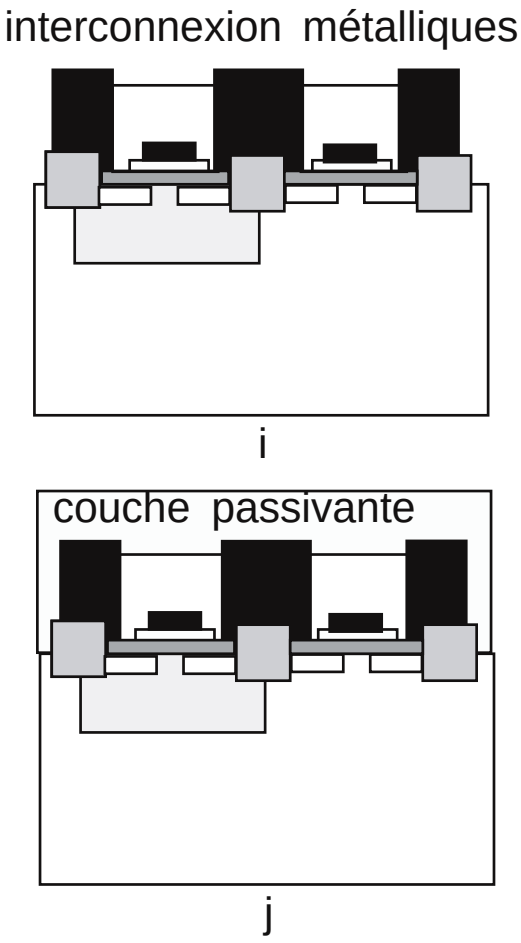


FIG. A.8 – Réalisation d'un CMOS (5)

j

Annexe B

Données numériques

B.1 Caractéristiques des principaux semi-conducteurs à 300k

Le tableau ci-dessous résume les caractéristiques physiques de trois semi-conducteurs ; le Germanium pour son caractère historique[1], le silicium car c'est le plus utilisé et qu'il est à l'origine du développement de la microélectronique, et l'arséniure de galium pour son utilisation en haute fréquence.

	E_G **(eV)**	N_i **(cm^{-3})**	μ**N cm^2/Vs**	μ**P cm^2/Vs**	ϵ **F/cm**	**claquage V/cm**
Ge	0,66	2,4.10^{13}	4000	1900	$1,44\ 10^{-12}$	10^5
Si	1,12	1,5.10^{10}	1400	500	$1,04\ 10^{-12}$	$3\ 10^5$
AsGa	1,43	10^7	8500	400	$0,96\ 10^{-12}$	$4\ 10^5$

	atomes/cm3	**cond. therm. w/cm°C**
Ge	4,42.10^{22}	0,64
Si	5,0 10^{22}	1,45
AsGa	2,21.10^{22}	0,46

[1] Noter toutefois le retour en force du Germanium dans des alliages comme SiGe utilisé en hyperfréquences.