PHYSIQUE DES SEMICONDUCTEURS III-V

S. LAVAL

*Institut d'Electronique Fondamentale, CNRS (URA 022), Bat. 220,
Université Paris Sud, F-91405 Orsay Cedex, France*

INTRODUCTION

Les matériaux semiconducteurs III-V sont des corps composés formés à partir d'un élément de la 3^e colonne et d'un élément de la 5^e colonne de la classification périodique. Parmi tous les composés binaires possibles, tous n'ont pas le même intérêt potentiel. L'étude de leurs propriétés, et en particulier de la structure de bandes, montre que les éléments les plus légers donnent des composés à large bande interdite, dont les propriétés se rapprochent de celles des isolants, et à structure de bande indirecte. Les composés incluant du bore, de l'aluminium, ou de l'azote, et le phosphore de gallium GaP entrent dans cette catégorie ; ils ont en général peu d'intérêt pour l'électronique rapide /1/, qui demande des semiconducteurs à forte mobilité de porteurs, ou pour l'optoélectronique où une structure de bande directe est nécessaire pour que les transitions optiques soient efficaces /2/. A l'autre extrémité, les éléments lourds comme le thallium ou le bismuth donnent des composés à caractère métallique. On considérera donc essentiellement les composés à base de gallium (GaAs, GaSb), ou d'indium (InP, InAs, InSb), dont les propriétés sont les plus intéressantes. Le tableau I résume quelques paramètres pour différents matériaux de la famille III-V.

L'intérêt pratique des semiconducteurs III-V est encore considérablement renforcé par la possibilité de réaliser des alliages par substitution partielle de l'un des éléments par un autre élément de la même colonne. On sait par exemple obtenir des alliages ternaires du type $\text{Ga}_x\text{Al}_{1-x}\text{As}$, $\text{Ga}_x\text{In}_{1-x}\text{As}$, ou quaternaires comme $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$. La figure 1 représente les variations de la largeur de bande interdite de l'alliage en fonction du paramètre cristallin, qui varie lui même avec la composition. Ce diagramme montre qu'il est possible d'obtenir des matériaux dont la largeur de bande interdite, et donc les propriétés optiques, varient dans une large gamme. Il existe cependant une contrainte importante pour la fabrication de ces matériaux, qui sont réalisés en couches minces par croissance épitaxiale sur un substrat binaire : le paramètre cristallin doit être très proche de celui du substrat. Le diagramme de la figure 1 permet de connaître la composition de tout alliage ternaire susceptible d'être épitaxié en couche mince sur les substrats binaires GaAs ou InP.

TABLEAU I - Paramètres caractéristiques pour les principaux composés III-V.

 E_g énergie de bande interdite ; λ_g longueur d'onde du seuil d'absorption ; m^* masse effective des électrons ; a_0 paramètre de maille du cristal.

Composé III-V	E_g (eV)	λ_g (μm)	m^*/m_0	a_0 ($\text{\AA}$)	Structure de la bande interdite
B N	$\approx 7,5$	0,17		3,615	indirecte
Al P	2,45	0,51		5,463	
Al As	2,16	0,57		5,661	
Al Sb	1,58	0,75	0,12	6,138	
Ga P	2,26	0,55	0,82	5,449	
Ga As	1,42	0,87	0,063	5,653	directe
Ga Sb	0,72	1,85		6,095	
In P	1,35	0,92	0,08	5,868	
In As	0,36	3,44	0,023	6,058	
In Sb	0,17	7,30	0,014	6,479

Fig.1 - Largeur de bande interdite en fonction du paramètre cristallin pour les alliages III-V.

Par ailleurs, l'interface entre deux composés différents (hétérojonction), présente des propriétés spécifiques, par exemple la présence d'un gaz d'électrons bidimensionnel, que l'on peut mettre à profit dans la réalisation de composants performants. L'empilement de couches alternées très minces de deux composés différents peut également faire apparaître des effets quantiques (puits quantiques multiples), ou même une périodicité supplémentaire du réseau cristallin (super-réseaux).

Les matériaux III-V offrent donc de multiples possibilités, aussi bien pour la microélectronique rapide que pour l'optoélectronique, par la grande variété des composés réalisables et l'ouverture qu'ils donnent sur la conception de dispositifs totalement nouveaux par l'ajustement des structures de bandes ("band gap engineering").

Nous décrivons les propriétés physiques, électroniques et optiques essentielles de ces matériaux avant de présenter un certain nombre de leurs applications à des composants électroniques, les dispositifs optoélectroniques étant traités par ailleurs.

I - STRUCTURE CRISTALLINE

Les semiconducteurs III-V formés à partir de Al, Ga, ou In d'une part, P, As, ou Sb d'autre part ont tous une structure de type zinc-blende [3]. Leur réseau cristallin peut se décomposer en deux sous-réseaux cubiques à faces centrées interpénétrés, l'un étant constitué des atomes de l'élément III, l'autre des atomes de l'élément V (fig. 2). Ces deux sous-réseaux sont décalés l'un par rapport à l'autre le long de la diagonale du cube, d'une quantité $(a_0/4, a_0/4, a_0/4)$, a_0 étant le paramètre cristallin, c'est-à-dire la longueur de l'arête du cube élémentaire. Chaque atome se trouve donc au centre d'un tétraèdre régulier dont les sommets sont occupés par un atome de l'autre espèce. La maille élémentaire, à partir de laquelle on peut reconstituer le cristal entier par un ensemble de translations, est formée par un atome de chaque type.

Fig.2 - Réseau cristallin dans la structure zinc-blende.

A partir du réseau cristallin, on définit le réseau réciproque, qui est le système de coordonnées [énergie - vecteur d'onde] dans lequel on représente les variations des fréquences de vibrations du réseau cristallin ou de l'énergie des états électroniques en fonction du vecteur d'onde $\vec{k}$ caractérisant la propagation de l'onde considérée (de nature vibrationnelle ou électronique) /4/.

Le réseau réciproque associé à la structure de type zinc-blende est cubique centré. Sa maille élémentaire, qui correspond à la première zone de Brillouin est représentée sur la figure 3 /5/. Elle présente un centre de symétrie Γ à l'origine ($\vec{k}=0$), et un certain nombre d'axes de symétrie : il existe trois axes équivalents $\langle 100 \rangle$ coupant le bord de la zone de Brillouin aux points X, et quatre axes équivalents $\langle 111 \rangle$, avec les points correspondants L.

Fig. 3 - Première zone de Brillouin d'un réseau cubique à faces centrées.

II - STRUCTURE DES BANDES D'ENERGIE

Les bandes d'énergie donnent les états d'énergie possibles pour les électrons en fonction de leur vecteur d'onde. On les représente donc dans l'espace réciproque, et, pour simplifier, dans les directions de plus hautes symétries de la première zone de Brillouin décrites ci-dessus. Elles se décomposent en bandes de valence et bandes de conduction (Fig.4). Ce sont la bande de valence la plus haute, la bande de conduction la plus basse, et la bande interdite qui les sépare qui déterminent principalement les propriétés de transport du semiconducteur, et ce sont celles-ci que nous considérerons par la suite.

L'allure générale des bandes est la même pour tous les composés III-V considérés /6/, dont la structure de bande est directe, c'est-à-dire pour lesquels le maximum de la bande de valence et le minimum de la bande de conduction sont situés au centre de la zone de Brillouin, en $k=0$. Le minimum central de la bande de conduction correspond à des électrons de faible masse effective, donc très mobiles. Il existe par

Fig. 4 - Structure de bandes de GaAs

ailleurs des minima secondaires en bordure de la zone de Brillouin : quatre vallées L équivalentes, et trois vallées de type X. Ces minima sont beaucoup plus plats, les électrons y ont une masse effective beaucoup plus grande, et donc une faible mobilité. Le tableau II donne les valeurs des masses effectives dans les différentes vallées et les écarts en énergie entre le minimum principal et les minima secondaires L et X pour

TABLEAU II - Paramètres de la bande de conduction dans GaAs et InP.

Température	GaAs		InP	
	300 K	77 K	300 K	77 K
ΔE_G (eV)	1,43	1,51	1,35	1,41
λ (μm)	0,87	0,82	0,92	0,88
$\Delta E_{\Gamma L}$ (eV)	0,33	0,33	0,54	0,61
$\Delta E_{\Gamma X}$ (eV)	0,48	0,46	0,76	0,755
$m_{e\Gamma}^*/m_0$	0,063	0,067	0,078	0,082
m_{eL}^*/m_0	0,22	0,22	0,22	0,22
m_{eX}^*/m_0	0,58	0,58	0,384	0,384

quelques semiconducteurs, à la température ambiante et à basse température. Notons que la largeur de bande interdite augmente lorsqu'on abaisse la température.

A l'équilibre, à la température du zéro absolu, et pour un matériau pur (intrinsèque), la bande de valence est remplie d'électrons, et la bande de conduction est vide. Lorsque la température s'élève, les électrons acquièrent de l'énergie thermique et un certain nombre d'entre eux passe dans la bande de conduction, laissant un "trou" dans la bande de valence; le matériau devient conducteur. Le nombre de places pour les électrons dans les bandes est limité : entre l'énergie ϵ et $\epsilon+d\epsilon$, le nombre d'états dans un volume unité est :

$$N(\epsilon) d\epsilon = \frac{1}{2\pi^2 h^3} (2m^*)^{3/2} \epsilon^{1/2} d\epsilon \quad (1)$$

et la densité totale d'états dans la bande de conduction jusqu'à l'énergie ϵ est donnée par :

$$N = \frac{1}{3\pi^2 h^3} (2m^*)^{3/2} (\epsilon - \epsilon_c)^{3/2} \quad (2)$$

si ϵ_c est l'énergie correspondant au bas de la bande de conduction.

La répartition des électrons entre la bande de valence et la bande de conduction dépend de la température (statistique de Fermi-Dirac), mais aussi de la pureté du matériau : l'introduction d'impuretés bien choisies permet de "doper" le matériau, en augmentant soit le nombre d'électrons dans la bande de conduction (dopage N), soit le nombre de trous dans la bande de valence (dopage P). Mais il existe aussi, inévitablement, des impuretés résiduelles qui introduisent des niveaux profonds dans la bande interdite, niveaux qui peuvent avoir des répercussions importantes sur le fonctionnement des dispositifs.

III - VIBRATIONS DU RESEAU CRISTALLIN - PHONONS

Pour décrire les propriétés des modes de vibration du réseau cristallin, auxquels sont associés les phonons, nous commençons par rappeler le modèle classique d'une chaîne unidimensionnelle d'atomes, formée de deux types d'atomes différents, de masses respectives M_1 et M_2 , distants de $d/2$, et dont les déplacements par rapport à leur position d'équilibre sont, dans la maille i , u_i et v_i , /4, 7/. En supposant des forces de rappel élastiques de constante K , les équations du mouvement s'écrivent :

$$\begin{aligned} M_1 \ddot{u}_i &= K (v_i + v_{i-1} - 2u_i) \\ M_2 \ddot{v}_i &= K (u_i + u_{i-1} - 2v_i) \end{aligned} \quad (3)$$

En cherchant des solutions périodiques de la forme :

$$\begin{aligned} u_i &= u \exp[j(\omega t - qid)] \\ v_i &= v \exp[j(\omega t - qid)] \end{aligned}$$

où q est le vecteur d'onde de la vibration, avec des conditions aux limites cycliques (cristal infini), on montre qu'il existe, pour chaque valeur de q , deux fréquences propres distinctes ω_- et ω_+ telles que :

$$\omega_{\pm}^2 = \frac{K}{M_1 M_2} [(M_1 + M_2) \pm \sqrt{M_1^2 + M_2^2 + 2M_1 M_2 \cos qd}] \quad (4)$$

Fig. 5 - Courbes de dispersion des modes de vibration d'une chaîne diatomique

Les courbes correspondantes sont représentées sur la figure 5.

Au centre de la zone de Brillouin, la fréquence ω_- s'annule, et la pente de la branche inférieure correspond à la vitesse de propagation du son le long de la chaîne : il s'agit du mode acoustique. La fréquence ω_+ prend la valeur :

$$\omega_+ = \sqrt{\frac{2K(M_1 + M_2)}{M_1 M_2}}$$

dont l'ordre de grandeur est celui des fréquences optiques infra-rouge, d'où la dénomination de mode optique.

En $q = 0$, on peut remarquer que pour la branche acoustique, $u/v = 1$, les deux atomes de la maille vibrent en phase, et on a un déplacement macroscopique du réseau. Par contre, pour le mode optique, on a $u/v = -M_2/M_1$, les atomes vibrent en opposition de phase à l'intérieur de la maille.

Au bord de la zone de Brillouin ($q = \pi/d$), les branches acoustique et optique se rapprochent, avec $\omega_- = \sqrt{2K/M_1}$ et $\omega_+ = \sqrt{2K/M_2}$, en supposant que $M_1 > M_2$. On peut remarquer que, pour la branche acoustique, on a $v=0$, les atomes les plus légers restent immobiles, alors que pour la branche optique, $u=0$, et ce sont les atomes les plus lourds qui sont fixes.

Dans les cristaux réels, il y a trois degrés de liberté par atomes, donc trois branches acoustiques et trois branches optiques, dont deux correspondent à des vibrations transversales, et une à des vibrations longitudinales pour chaque type de mode, les fréquences restant dégénérées si le cristal est isotrope et covalent.

Dans certains cristaux, comme les composés III-V que nous considérons, la liaison entre atomes est partiellement ionique, et le cristal est alors qualifié de "polaire". Les vibrations peuvent en effet induire un déplacement relatif de charges, équivalent à des dipôles oscillants qui se couplent aux ondes électromagnétiques de même fréquence ω : c'est ainsi qu'un rayonnement infra-rouge incident sur le cristal est absorbé, son énergie servant à exciter les vibrations du réseau. Ceci est particulièrement important pour le mode optique au centre de la zone de Brillouin, puisque les atomes de la maille vibrent en opposition de phase. Pour les vibrations transversales, le champ électrique induit par les vibrations est faible; il est

beaucoup plus grand pour les vibrations longitudinales, et il en résulte une polarisation macroscopique qui conduit à une levée de dégénérescence entre les fréquences transversale et longitudinale, et à des interactions de type coulombien avec les électrons libres du cristal. Ce caractère polaire des phonons optiques au centre de zone de Brillouin a donc des répercussions importantes sur les propriétés de transport électronique comme nous le verrons par la suite.

La figure 6 reproduit les courbes de dispersion des phonons dans le cas de GaAs [9], dans les différentes directions de plus fortes symétries de la zone de Brillouin telles que nous les avons définies plus haut.

Fig. 6 - Courbes de dispersion des phonons dans GaAs.

IV - TRANSPORT ELECTRONIQUE EN REGIME STATIONNAIRE

Considérons un matériau semiconducteur de type N, ayant donc un certain nombre d'électrons dans la bande de conduction. A l'équilibre à la température T, ces électrons sont animés d'un mouvement brownien; la vitesse moyenne qui en résulte est nulle. Lorsqu'on applique un champ électrique, ils sont accélérés et voient donc leur énergie augmenter, c'est-à-dire qu'ils montent dans la bande de conduction. Cependant les "collisions" qu'ils subissent peuvent à la fois leur faire perdre de l'énergie (collisions inélastiques), et modifier la direction de leur vitesse, ce qui globalement ralentit leur mouvement parallèlement au champ électrique. Le régime de transport stationnaire est atteint quand il y a équilibre entre l'accélération communiquée par le champ électrique et le "freinage" sous l'effet des collisions. La vitesse moyenne a alors une composante non nulle dans la direction du champ électrique : c'est la vitesse de dérive. L'énergie cinétique globale des porteurs dans un champ électrique provient donc de la superposition d'une composante ordonnée du mouvement, liée au champ, et d'une composante désordonnée correspondant à l'énergie thermique. Les valeurs relatives de ces deux composantes dépendent de la nature des collisions et du champ électrique présent, ces deux facteurs n'étant pas indépendants puisque, comme nous allons le voir, les caractéristiques des collisions dépendent de l'énergie des électrons.

Les collisions correspondent à des interactions que subissent les électrons soit avec les impuretés présentes dans le réseau cristallin (dopage), soit avec les différents types de phonons, optiques ou acoustiques.

Les interactions avec les impuretés sont élastiques, et ont pour seule conséquence de dévier la trajectoire de l'électron, ceci d'autant moins que son énergie est élevée ; elles auront généralement peu d'influence sur le transport, sauf pour des matériaux très dopés.

Les interactions prépondérantes pour les électrons dans la vallée centrale Γ sont celles avec les phonons optiques polaires, de faible vecteur d'onde $\vec{q}$. Elles s'accompagnent de l'absorption ou de l'émission d'un phonon, et l'énergie de l'électron varie d'une quantité égale à l'énergie du phonon. L'émission d'un phonon est un processus beaucoup plus probable que l'absorption dès que l'énergie de l'électron est suffisante. Ces interactions étant de type coulombien, leur probabilité est inversement proportionnelle à q^2 , et les variations les plus faibles du vecteur d'onde de l'électron sont privilégiées, et celui-ci sera peu dévié. Les interactions avec les phonons optiques polaires sont donc inélastiques et anisotropes.

Dès que les électrons ont atteint une énergie suffisante pour pouvoir être transférés vers les vallées latérales de la bande de conduction, ce sont les interactions avec les phonons de grand vecteur d'onde, donc du bord de la zone de Brillouin, qui deviennent prépondérantes. Elles correspondent à un couplage par potentiel de déformation, sont inélastiques, et isotropes. Elles ont donc des conséquences importantes sur le transport puisqu'elles redistribuent les vitesses électroniques dans toutes les directions.

Les interactions avec les phonons acoustiques du centre de la zone de Brillouin, qui ont une très faible énergie, sont pratiquement élastiques et sont isotropes.

Fig. 7 - Probabilités d'interaction des électrons en fonction de leur énergie dans GaAs.

La figure 7 représente les différentes probabilités d'interaction à la température ambiante dans GaAs /10/. On peut y remarquer les seuils correspondant à l'énergie du phonon optique polaire, et à l'écart $\Delta E_{\Gamma L}$ en énergie entre le minimum Γ de la bande de conduction et les vallées L.

Les propriétés que nous venons de décrire, jointes à l'augmentation importante de la masse effective des électrons lorsqu'ils passent dans les vallées latérales, font que la courbe représentant les variations de la vitesse électronique en fonction du champ électrique appliqué (fig.8) présente une partie de résistance différentielle négative, conduisant à des applications spécifiques (effet Gunn). Le champ correspondant au maximum de vitesse est appelé champ critique. Il est plus grand dans InP que dans GaAs dans la mesure où l'écart en énergie $\Delta E_{\Gamma L}$ est plus important et les probabilités d'interaction avec les phonons optiques polaires plus élevées. Ces courbes montrent que si l'on conserve la définition habituelle de la mobilité comme étant le rapport entre la vitesse électronique et le champ électrique appliqué :

$$\vec{v} = \mu \cdot \vec{E} \quad (5)$$

la mobilité n'est pas constante et varie fortement avec l'énergie des électrons.

Il faut remarquer qu'à 300K, la distribution de l'énergie thermique des électrons est telle que de nombreux électrons ont d'emblée une énergie supérieure à l'énergie des phonons optiques polaires. Il n'en est pas de même à 77K, et l'apparition de la possibilité d'émission de phonons optiques polaires quand les électrons ont atteint l'énergie nécessaire dans le champ électrique appliqué se traduit sur la courbe vitesse - champ par une rupture de pente, pour les matériaux peu dopés (fig.9). L'augmentation du dopage produit une diminution de la mobilité par interactions avec les impuretés.

Fig. 8 - Caractéristiques vitesse - champ électrique en régime de transport stationnaire à 300 K calculées par la méthode de Monte Carlo [11/.

Fig. 9 - Caractéristiques vitesse - champ électrique à 77 K en régime stationnaire (simulation Monte Carlo /11/).

V - TRANSPORT ELECTRONIQUE NON-STATIONNAIRE

V.1- Description théorique

Le régime de transport stationnaire n'est atteint que lorsque les électrons ont subi un nombre suffisant d'interactions, c'est-à-dire lorsque la distance parcourue est grande devant le libre parcours moyen. Dans des structures submicroniques, il devient indispensable de prendre en compte les effets transitoires qui existent lorsque les électrons sont brusquement soumis à un échelon de champ électrique, qu'il soit temporel ou spatial. En effet, aux tout premiers instants, peu d'interactions ont encore eu lieu alors que l'accélération dans le champ est effective. Il en résulte que les vitesses électroniques peuvent transitoirement être plus importantes que les vitesses en transport stationnaire : il s'agit du phénomène de survitesse. La courte période avant toute collision correspond au régime de transport balistique.

Ces phénomènes sont régis par deux temps caractéristiques : le temps de relaxation du moment τ_m qui caractérise la rapidité avec laquelle l'orientation du vecteur vitesse relaxe, et le temps de relaxation de l'énergie des électrons τ_E . Ces temps de relaxation dépendent de l'énergie des porteurs. La plupart des interactions sont isotropes et ont donc une action beaucoup plus importante sur l'orientation de la vitesse que sur l'énergie, l'énergie des phonons étant relativement faible par rapport à celle des électrons. Il en résulte que, généralement, le temps de relaxation du moment est beaucoup plus faible que celui de l'énergie.

On peut donner un modèle simplifié du phénomène de survitesse, en écrivant les équations de relaxation du moment et de l'énergie :

$$\frac{d(m^* \vec{v})}{dt} = e \vec{E} - \frac{m^* \vec{v}}{\tau_m} \quad (6)$$

$$\frac{d\mathcal{E}}{dt} = e \vec{v} \cdot \vec{E} - \frac{\mathcal{E} - \mathcal{E}_{th}}{\tau_E} \quad (7)$$

Supposons qu'au temps $t=0$ on applique un échelon de champ électrique d'amplitude supérieure au champ critique. Si $t < \tau_m$, on est dans le cas du régime balistique qui ne peut être décrit par les équations ci-dessus. Supposons donc que $t \gg \tau_m$, et que la vitesse est par conséquent en équilibre avec le champ appliqué. La première équation peut être considérée comme stationnaire ($d/dt = 0$) et donne

$$\vec{v} = (e \tau_m(\mathcal{E}) / m^*(\mathcal{E})) \vec{E} = \mu(\mathcal{E}) \vec{E} \quad (8)$$

Pour des temps suffisamment courts pour que l'énergie des porteurs ait peu augmenté ($t \ll \tau_E$), l'énergie des porteurs est voisine de l'énergie thermique $\mathcal{E}_{th}$, la mobilité est donc celle des porteurs dans le bas de la vallée Γ de la bande de conduction, $\mu(\mathcal{E}) = \mu(\mathcal{E}_{th})$, soit une mobilité élevée. Par ailleurs, le champ $\vec{E}$ étant intense, la vitesse $\vec{v} = \mu(\mathcal{E}_{th}) \vec{E}$ est donc importante. Pour des temps plus longs, l'énergie des porteurs augmente, et lorsqu'ils ont atteint l'énergie de transfert vers les vallées latérales, la mobilité décroît ; on tend vers le régime de transport stationnaire avec $\vec{v} = \mu(\mathcal{E}) \vec{E}$ (fig.10).

Explication du phénomène de survitesse dans GaAs : les électrons pénètrent un échelon de champ électrique en x_0 ($t = t_0$). Leur énergie $\mathcal{E}$ ne croît pas instantanément et leur mobilité μ , fonction de $\mathcal{E}$, conserve des valeurs élevées pendant le régime non stationnaire. Leur vitesse d'entraînement $v = \mu E$ présente des valeurs en excès, par rapport à l'état stationnaire (survitesse), sur une distance pouvant atteindre $0.5 \mu\text{m}$ dans GaAs à 300 K.

Fig. 10 - Explication schématique du phénomène de survitesse

On voit donc que, dans la mesure où la mobilité initiale en champ faible est plus élevée que la mobilité en champ fort ($\mu(\mathcal{E}_{th}) > \mu(\mathcal{E})$), un phénomène de survitesse peut être observé. Ceci est plus marqué dans les semiconducteurs III-V que dans le Silicium, du fait de la structure de bande qui induit une diminution importante de la mobilité lorsque le champ électrique croît.

Ce phénomène de survitesse a été très étudié par des simulations numériques particulières de type Monte-Carlo, dans lesquelles le mouvement des particules est calculé en fonction du champ appliqué. Dans les composés III-V, l'échelle de temps sur laquelle il se produit est la picoseconde (fig. 11). Si on applique un créneau de champ électrique spatialement, ce phénomène se produit sur des distances qui sont de l'ordre de quelques dixièmes de microns. Les vitesses maximales obtenues peuvent atteindre 6 à 8 fois la valeur de la vitesse stationnaire. Ceci doit cependant être modulé par le fait que dans un dispositif réel, c'est le temps de transit des porteurs qui est le paramètre important, et que la phase d'accélération conduit à une réduction de la vitesse moyenne sur la traversée du dispositif et peut être très pénalisante pour des composants de longueur inférieure à $0,2\ \mu\text{m}$.

Fig. 11 - Evolution temporelle de la vitesse moyenne des électrons à 300K

V.2 - Mise en évidence expérimentale

Compte tenu des échelles de temps et de distance sur lesquelles se produisent ces phénomènes, leur mise en évidence expérimentale est délicate. Il faut une résolution temporelle inférieure à $0,5\ \text{ps}$, ce qui est possible par des méthodes optiques /12/, mais l'interprétation n'en est pas immédiate. Des mesures de conductivité en ondes submillimétriques, suivies d'une transformée de Fourier, ont été tentées /13/, mais la difficulté réside alors dans la nécessité d'avoir des mesures sur une gamme de fréquences assez étendue dans ce domaine. Une méthode de spectroscopie d'électrons chauds a permis de mettre en évidence des électrons balistiques /10,14/. On peut également faire une analyse spatiale en comparant les caractéristiques d'échantillons submicroniques de différentes longueurs /15/. C'est cette méthode dont le principe est rappelé ici avec les principaux résultats obtenus.

Le tracé d'une caractéristique courant - tension dans un dispositif submicronique ne suffit pas pour en déduire les variations de la mobilité, donc de la vitesse moyenne des porteurs, en fonction du champ électrique. En effet, l'injection des porteurs aux contacts varie souvent de façon non-linéaire avec la tension, et le résultat de la mesure est alors erroné. On peut pallier cet inconvénient en utilisant une méthode de

photoconduction, méthode différentielle dans laquelle on mesure un courant de faible intensité par rapport au courant d'obscurité, évitant ainsi les problèmes de non-linéarités de l'injection.

Les expériences ont été faites essentiellement sur GaAs, mais aussi sur InP. Les dispositifs utilisés sont représentés schématiquement sur la figure 12. Ils sont formés d'une couche de GaAs d'épaisseur e voisine de $5\text{ }\mu\text{m}$, de dopage résiduel très faible ($\approx 10^{14}\text{ cm}^{-3}$), épitaxiée sur un substrat semi-isolant et sur laquelle sont réalisés deux contacts ohmiques, dont la distance L varie selon les échantillons de quelques microns à $0,1\text{ }\mu\text{m}$. Ils sont éclairés par des impulsions lumineuses de quelques μs de durée (éclairage quasi-continu).

Fig. 12 - Dispositif photoconducteur pour l'étude du transport non stationnaire.

Le photocourant mesuré entre les deux électrodes est proportionnel au nombre de porteurs créés par l'éclairement, et à leur vitesse moyenne sur la traversée du dispositif. Si la distance inter-électrodes est supérieure à quelques microns, la caractéristique photocourant - tension est celle du transport stationnaire. Par contre, pour les distances submicroniques, les caractéristiques reflètent les variations de la vitesse moyenne des électrons en fonction du champ électrique, et on peut ainsi mettre en évidence les effets de survitesse et de transport balistique. Un avantage supplémentaire de la méthode est que le photocourant circule dans une zone de faible épaisseur, dans laquelle le champ électrique moyen est proche du rapport V/L de la tension appliquée à la distance inter-électrodes, contrairement au courant d'obscurité qui n'est limité que par l'épaisseur de la couche. Les difficultés qui subsistent sur le plan expérimental sont la connaissance précise du nombre de photons incidents qui pénètrent dans le matériau, compte tenu de la réflexion et de la diffraction de la lumière, et la détermination de la durée de vie moyenne des porteurs $\langle\tau\rangle$, qui définit le nombre d'électrons en excès en régime d'éclairement continu.

Les résultats montrent que pour des dispositifs de quelques microns de longueur, pour différents matériaux III-V, la caractéristique est proche des courbes vitesse- champ électrique en régime stationnaire (fig. 13). Pour des distances plus faibles, les caractéristiques présentent de moins en moins de saturation (fig. 14), et ceci s'explique bien par les résultats de simulation numérique Monte-Carlo [11] qui montrent que le nombre de particules transférées dans les vallées latérales de la bande de conduction, pour un champ électrique donné, est d'autant plus faible que la distance est courte. On met ainsi clairement en évidence le phénomène de survitesse. Les effets sont encore plus marqués sur des dispositifs plus courts et à basse température (77 K). La figure 15 donne les variations du photocourant en fonction de la tension appliquée

entre les deux électrodes. On observe sur l'échantillon le plus court ($L = 0,5 \mu\text{m}$) un changement de pente à partir de 360 mV. Cette valeur correspond à l'énergie minimum nécessaire aux électrons pour être transférés dans les vallées L avec l'émission d'un phonon. Ils ont donc traversé la structure sans perte d'énergie, donc en transport balistique. Pour les distances un peu plus grandes, 0,2 et 0,4 μm , la cassure apparaît à des tensions légèrement plus élevées : les électrons ont subi quelques interactions avec des phonons optiques polaires, qui leur ont fait perdre un peu d'énergie au cours de leur trajet.

Fig. 13 - Photocourant mesuré en fonction du champ électrique moyen dans des dispositifs longs.

Fig. 14 - Photocourant en fonction de la tension appliquée pour des dispositifs submicroniques GaAs à 300 K.

Fig. 15 - Photocourant en fonction de la tension appliquée pour des dispositifs submicroniques GaAs à 77 K.

Fig. 16 - Comparaison des caractéristiques à 300 K et à 77 K pour des dispositifs GaAs de différentes longueurs.

Alors que la mobilité des électrons dépend de la température puisque les probabilités d'interaction en dépendent, le transport balistique en est indépendant. La comparaison des caractéristiques à 300 K et 77 K pour des échantillons de différentes longueurs (fig. 16) confirme ainsi l'existence du transport balistique : pour $L = 3 \mu\text{m}$, le rapport des photocourants mesurés est d'environ 2, c'est-à-dire proche du rapport des mobilités entre ces deux températures, tandis que pour $L = 0,08 \mu\text{m}$, les deux courbes sont confondues.

En estimant le nombre de photons incidents, compte tenu des effets de la diffraction, on peut évaluer les vitesses moyennes des électrons $\langle v \rangle$, qui dépassent $3 \cdot 10^7 \text{cm/s}$ dans GaAs pour des dispositifs submicroniques, soit un gain supérieur à 3 par rapport à la vitesse en régime stationnaire.

VI - PROPRIETES DE SURFACE

Les matériaux III-V sont fragiles, en particulier thermiquement dans la mesure où l'arsenic et le phosphore sont des éléments volatils qui ont tendance à quitter la surface dès que l'on dépasse des températures de l'ordre de 500°C . Il a donc fallu mettre au point des technologies dites "froides", et des procédés d'encapsulation, pour limiter les dégradations du matériau en surface lors de la réalisation des dispositifs.

Par ailleurs, la surface de GaAs est bien connue pour présenter des états de surface tels que le niveau de Fermi en surface est fixé au voisinage du milieu de la bande interdite, ce qui permet de réaliser de bons contacts Schottky par dépôt d'un métal. En conséquence, il existe sous la surface, dans un matériau de type N, une zone de charge d'espace (ZCE) vide d'électrons, qui s'étend typiquement, sous une surface infinie, sur une épaisseur de $0,3 \mu\text{m}$ si le dopage est de 10^{16}cm^{-3} . La proximité d'électrodes peut diminuer l'extension en profondeur, comme le montre la figure 17 représentant le résultat de simulations Monte-Carlo pour différentes distances inter-électrodes, sur une couche de GaAs de faible dopage (10^{14}cm^{-3}).

Fig. 17 - Répartition des électrons dans des dispositifs GaAs, montrant l'extension de la zone de charge d'espace sous la surface pour différentes distances inter-électrodes.

L'existence de cette zone de charge d'espace a des conséquences importantes, sur lesquelles nous reviendrons, pour le fonctionnement des transistors à effet de champ. Elle est également responsable de phénomènes observés en photoconduction. Lorsque le semiconducteur est éclairé, les paires électrons-trous créées sont séparées par le champ électrique perpendiculaire à la surface régnant dans la ZCE, les trous viennent s'accumuler en surface, et seuls les électrons participent au courant de photoconduction ; les porteurs étant séparés, leur durée de vie apparente est très longue. Sous l'effet de l'éclairement la charge des états de surface varie $/16/$, et simultanément la courbure de bande associée à la ZCE diminue. A fort éclairement, les deux types de porteurs circulent sous la surface, et peuvent se recombiner entre eux, retrouvant une durée de vie caractéristique du matériau massif. Ceci explique les traînes observées sur la décroissance du photocourant après une impulsion lumineuse ultra-brève, en particulier sous faible éclairement (fig. 18 a).

Pour InP, la surface est probablement en légère accumulation d'électrons plutôt qu'en désertion, ou bien la situation est proche des bandes plates en surface, selon le dopage. La décroissance du photocourant après une impulsion courte est alors beaucoup moins sensible à la puissance lumineuse (fig. 18 b).

Fig. 18 - Evolution temporelle du photocourant après une impulsion lumineuse ultra courte dans GaAs (a) et dans InP (b).

VII - PROPRIETES OPTIQUES

L'interaction d'un électron avec un photon se fait, comme toute interaction, avec conservation de l'énergie et du vecteur d'onde. Le vecteur d'onde du photon étant très inférieur à celui des électrons, les transitions optiques directes entre la bande de valence et la bande de conduction apparaissent verticales dans le diagramme de bande des électrons.

VII.1 Absorption

Lorsqu'un photon d'énergie $h\nu$ supérieure à la largeur de bande interdite E_g est envoyé sur le semiconducteur, il est absorbé en induisant la transition d'un électron de la bande de valence vers la bande de conduction, c'est-à-dire en créant une paire électron-trou [2].

Le coefficient d'absorption α est proportionnel à la densité d'électrons dans l'état initial n_1 , et à la densité de places disponibles dans l'état final n_2 :

$$\alpha = A \sum p_{12} \cdot n_1 \cdot n_2 \quad (9)$$

où p_{12} est la probabilité de transition.

La transition associe un état de la bande de valence à un état de la bande de conduction de même vecteur d'onde $\vec{k}$. En supposant les bandes paraboliques, on a donc :

$$h\nu = E_g + \frac{\hbar^2 k^2}{2m_l^*} + \frac{\hbar^2 k^2}{2m_c^*} \quad (10)$$

Cette relation définit l'ensemble des états de vecteur d'onde $\vec{k}$ qui interviennent sous l'excitation par un rayonnement de fréquence ν , à partir du seuil d'absorption donné par l'énergie E_g .

Sous fort éclaircissement à une longueur d'onde λ donnée, les états du bas de la bande de conduction se remplissent, la probabilité de transition à cette fréquence chute, et le matériau devient transparent : c'est le phénomène de saturation d'absorption, qui s'accompagne d'une augmentation de la largeur de bande interdite apparente. Plus l'énergie des photons incidents est supérieure au gap, plus il faut une densité d'énergie lumineuse élevée pour observer ce phénomène [17].

Pour des fréquences plus importantes, d'autres bandes peuvent intervenir, en particulier la bande de trous légers dite "split-off", se traduisant par des épaulements sur la courbe d'absorption en fonction de l'énergie des photons.

Fig. 19 - Spectre d'absorption de GaAs pour différentes températures.

Lorsque la bande de conduction contient des électrons, un photon incident peut être absorbé en leur communiquant de l'énergie. Cette absorption par les porteurs libres correspond à une transition intra-bande, qui nécessite donc une interaction avec un phonon pour la conservation du vecteur d'onde. Le coefficient d'absorption est proportionnel au carré de la longueur d'onde du rayonnement incident, et ce phénomène est donc plus important aux grandes longueurs d'onde (IR). Par ailleurs, dans le domaine de l'infrarouge lointain, il apparaît les bandes d'absorption associées aux modes de vibration du réseau cristallin.

La largeur de la bande interdite étant fonction de la température, le seuil d'absorption en dépend également (fig 19) /18/ On voit par ailleurs apparaître à basse température un pic au voisinage du seuil, qui correspond à l'absorption excitonique (les excitons sont des paires électron-trou non encore dissociées, l'énergie de liaison entre les deux particules introduisant un niveau légèrement inférieur au bas de la bande de conduction /19/).

VII.2 Indice de réfraction

Dans un semiconducteur, la constante diélectrique s'écrit en fonction de la fréquence sous la forme :

$$\epsilon = 1 + \frac{4\pi Ne^2}{m^* (-\omega^2 - j\Gamma\omega)} \quad \text{avec } \omega = 2\pi\nu \quad (11)$$

où Γ caractérise l'absorption du milieu. ϵ est une grandeur complexe.

De la même façon, l'indice de réfraction est complexe :

$$\tilde{n} = n - j\chi \quad (12)$$

où χ est le coefficient d'extinction, relié au coefficient d'absorption α par l'expression :

$$\alpha = \frac{4\pi\nu\chi}{c} \quad (13)$$

Les parties réelles et imaginaires de l'indice de réfraction sont liées par les relations de Kramers - Krönig :

$$n(\omega) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\omega' \chi(\omega')}{\omega'^2 - \omega^2} d\omega' \quad (14)$$

(P = partie principale)

En haute fréquence, l'indice de réfraction est lié à la constante diélectrique par la relation :

$$n = \sqrt{\epsilon}$$

Le coefficient de réflexion du matériau s'écrit :

$$R = \frac{(n - 1)^2 + \chi^2}{(n + 1)^2 + \chi^2} \quad (15)$$

Il faut remarquer que dans les alliages ternaires du type GaAlAs, l'indice de réfraction dépend de la composition, c'est-à-dire de la concentration en aluminium, et ceci est utilisé dans la réalisation des diodes laser en particulier.

VII.3 Recombinaison des porteurs, transitions radiatives

La recombinaison directe bande à bande se fait entre un électron du bas de la bande de conduction et un trou du sommet de la bande de valence, les porteurs étant thermalisés dans chaque bande. Si l'énergie correspondante est emportée par un photon ($h\nu = E_g$), la transition est radiative, et son spectre est étroit.

Le taux de recombinaison est proportionnel à la densité de porteurs dans l'état haut n_2 , et à la densité de places libres dans l'état bas n_1 :

$$R_{21} = p_{21} n_2 n_1 \quad (16)$$

p_{21} étant la probabilité de transition par porteur libre.

Le processus de recombinaison suppose que le système est initialement hors d'équilibre, ce qui peut être obtenu soit par excitation par un courant électrique conduisant à un phénomène de claquage pour

l'électroluminescence, soit par excitation optique par des photons d'énergie supérieure au gap pour la photoluminescence, soit par un faisceau d'électrons pour la cathodoluminescence.

Les recombinaisons radiatives ne sont pas les seules possibles. En effet, l'énergie libérée par la recombinaison peut être communiquée à un autre porteur, qui passe à un niveau plus excité dans la bande (effet Auger). Par ailleurs, les niveaux profonds existant dans la bande interdite peuvent jouer le rôle de centres de recombinaison : l'électron et le trou ayant des probabilités voisines de venir peupler ce niveau se recombinent alors par son intermédiaire (transition indirecte). Certains de ces niveaux peuvent accepter préférentiellement un électron ou un trou, jouant le rôle de pièges qui modifient la durée de vie effective des porteurs en fonction du coefficient d'émission du niveau.

En régime permanent de déséquilibre, par exemple sous l'effet d'un éclairage continu, le nombre de porteurs en excès résulte de l'équilibre entre la génération G de paires électron-trou et la recombinaison : $G = p_2 n_2 n_1$. Supposons un matériau de type N, on a :

$$n_2 = N + \Delta N$$

$$n_1 = \Delta P$$

où ΔN et ΔP sont les nombres de porteurs en excès, et sont égaux si les phénomènes de piégeage sont négligeables. Il vient :

$$G = p_2 (N + \Delta N) \Delta N \quad (17)$$

Si le nombre de porteurs en excès est faible par rapport au dopage ($\Delta N \ll N_D$) il est proportionnel au taux de génération :

$$\Delta N = G\tau \quad (18)$$

où $\tau = (p_2 N_D)^{-1}$ représente la durée de vie des porteurs.

Par contre, si le déséquilibre est important, pour $\Delta N \gg N_D$, on a :

$$\Delta N = \sqrt{G/p_2} \quad (19)$$

ΔN n'est plus proportionnel à G , et on ne peut pas dans ce cas-là définir une durée de vie significative puisqu'elle varie avec le taux de génération /20/. Ceci se traduit par exemple par une décroissance non exponentielle du photocourant après une impulsion lumineuse.

Dans la mesure où on peut définir des durées de vie τ et τ' respectivement pour les transitions radiatives et les transitions non radiatives, l'efficacité de recombinaison radiative est donnée par /2/ :

$$\frac{R}{R_{\text{tot}}} = \frac{1}{1 + \tau / \tau'} \quad (20)$$

L'émission dont il a été question ici est l'émission spontanée. Un photon incident peut également induire une transition radiative, avec émission d'un second photon ayant des caractéristiques identiques au premier ; c'est l'émission stimulée, qui est à la base du fonctionnement des lasers.

VII.4 Bilan des phénomènes lors de l'excitation lumineuse d'un semiconducteur

Les différents processus mis en jeu sont résumés sur la figure 20. L'absorption optique de photons d'énergie $h\nu$ crée une population d'électrons dans la bande de conduction, avec une énergie moyenne dans cette bande définie à partir de la relation (10). Les électrons se thermalisent très vite entre eux ($\approx 10^{-13}$ s), et

on peut donc définir une température électronique supérieure à celle du réseau cristallin : il s'agit d'électrons "chauds". Ces électrons relaxent vers le bas de la bande de conduction, en émettant des phonons optiques polaires, en des temps qui sont de l'ordre de quelques picosecondes. De la même façon, les trous relaxent vers le sommet de la bande de valence. Il y a ensuite recombinaison des électrons et des trous, par les processus décrits plus haut. Dans un matériau GaAs de bonne qualité, dans lequel les transitions radiatives sont prédominantes, le temps caractéristique de recombinaison est de l'ordre de quelques nanosecondes.

Fig. 20 - Transitions optiques entre la bande de valence et la bande de conduction.

Simultanément, il y a diffusion spatiale des porteurs. En l'absence de champ électrique appliqué, les deux types de porteurs diffusent simultanément, puisque le champ coulombien retient leur séparation dans l'espace. La diffusion est donc caractérisée par le coefficient ambipolaire /2/ :

$$D_a = \frac{2D_e D_t}{D_e + D_t} \quad (21)$$

où D_e et D_t sont respectivement les coefficients de diffusion pour les électrons et les trous, liés à la mobilité des porteurs par la relation d'Einstein :

$$D = \mu kT / e \quad (22)$$

VII.5 Propriétés optiques non -linéaires

La polarisation diélectrique d'un milieu peut se développer selon les puissances du champ électrique /21/ sous la forme :

$$\vec{P} = \epsilon_0 (\chi^{(1)} \vec{E} + \chi^{(2)} \vec{E} \cdot \vec{E} + \chi^{(3)} \vec{E} \cdot \vec{E} \cdot \vec{E} + \dots) \quad (23)$$

les susceptibilités non-linéaires χ étant des grandeurs tensorielles.

Le premier terme correspond à la polarisation linéaire classique, l'induction électrique s'écrivant :

$$\vec{D}_L = \epsilon_0 \vec{E} + \vec{P}_L = \epsilon_0 (1 + \chi^{(1)}) \vec{E} = \epsilon_0 \epsilon \vec{E} \quad (24)$$

Le terme en $\chi^{(2)}$ conduit à la génération de nouvelles fréquences par somme ou différence des fréquences de deux champs incidents. Un cas particulier est celui de l'effet électrooptique, où l'un des champs est un champ continu ou basse fréquence. Le terme en $\chi^{(3)}$ peut s'interpréter dans certains cas comme une modification des propriétés optiques, en particulier de l'indice de réfraction (effet Kerr).

a) Effet électrooptique ou effet Pockles

En présence d'un champ électrique continu $E^{\rightarrow}$, l'induction électrique à la fréquence ω_{op} s'écrit :

$$\vec{D}(\omega_{op}) = \epsilon_0 (1 + \chi^{(1)} + \chi^{(2)} E^{\rightarrow}) E^{\rightarrow}_{op} \quad (25)$$

et la polarisation non -linéaire est donc équivalente à une variation $\Delta\epsilon = \chi^{(2)}E$ de la constante diélectrique, donc à une variation de l'indice de réfraction du matériau.

On définit le coefficient électrooptique $r_{ij,k}$ tel que :

$$\Delta\left(\frac{1}{n^2}\right)_{ij} = \sum r_{ij,k} E_k \quad (26)$$

L'application d'un champ continu modifie donc l'ellipsoïde des indices /22/.

Dans le cas de GaAs, qui a une symétrie cubique, si on applique par exemple un champ parallèle à l'axe $\langle 111 \rangle$, on observe une diminution de l'indice de réfraction pour une induction associée au champ optique parallèle au champ appliqué, et une augmentation dans la direction perpendiculaire. Le semiconducteur devient alors biréfringent.

L'application d'un champ basse fréquence (par rapport aux fréquences optiques) permet ainsi de réaliser sur des substrats de matériaux III-V des modulateurs en phase ou en amplitude, ce qui est particulièrement intéressant en optique intégrée. Par ailleurs, on peut mettre à profit cet effet pour caractériser le champ de fuite au voisinage d'une ligne de propagation dans un circuit intégré, ou le champ électrique en certains points du circuit par des mesures optiques, donc sans contacts, basées sur la biréfringence induite par l'effet électrooptique /23/.

b) Non-linéarités optiques d'ordre 3

On peut distinguer deux types de non-linéarités :

- les non-linéarités paramétriques, lorsque les fréquences des champs optiques sont dans une zone de transparence, et qu'il n'existe donc pas d'échange d'énergie avec le milieu (génération de nouvelles fréquences, effet Kerr)
- les non-linéarités dynamiques pour lesquelles des transitions réelles sont mises en jeu, la fréquence de l'un au moins des champs optiques étant proche d'une résonance. Il y a échange d'énergie entre le faisceau lumineux et le semiconducteur, par exemple par création de porteurs.

Supposons qu'il n'existe pas de terme non-linéaire du second ordre, ce qui est le cas dans un matériau centrosymétrique. Dans le cas d'une interaction dégénérée, dans laquelle toutes les fréquences mises en jeu sont identiques, la polarisation s'écrit :

$$\vec{P} = \epsilon_0 (\chi^{(1)} + \chi^{(3)} E^2) \cdot \vec{E} \quad (27)$$

et la non-linéarité se traduit par une variation $\Delta\epsilon = \chi^{(3)}E^2$ de la constante diélectrique, proportionnelle à l'intensité du faisceau incident.

La relation (11) montre que si le nombre d'électrons varie, il y a modification des propriétés optiques du semiconducteur, indice de réfraction et absorption, en fonction de la puissance lumineuse incidente. On peut écrire :

$$n = n_0 + n_2 I_{op} \quad (28)$$

Cet effet est largement utilisé pour la réalisation de dispositifs bistables optiques, ou d'éléments non-linéaires pour des systèmes logiques entièrement optiques /24/.

Lorsque la puissance lumineuse augmente, plusieurs phénomènes interviennent en plus des variations d'indice et d'absorption, comme la saturation d'absorption qui a déjà été mentionnée et qui entraîne un décalage du spectre d'absorption vers les hautes énergies, et l'effet de "renormalisation du gap", lié à un écrantage du potentiel coulombien pour les fortes densités de porteurs, qui diminue l'énergie nécessaire pour créer une paire électron-trou. Par ailleurs, il faut noter que l'indice de réfraction varie fortement avec la température, et les effets thermiques peuvent se superposer aux phénomènes non-linéaires, diminuant leur efficacité.

Lorsque plusieurs faisceaux cohérents et de même fréquence sont simultanément présents dans le semiconducteur, ils peuvent interférer. Les franges d'intensité lumineuse induisent des variations périodiques du nombre de porteurs photocréés, et donc un réseau d'indice de réfraction de pas $\Lambda = (\lambda/2)\sin\theta$, si l'angle entre les deux faisceaux incidents est 2θ [25]. Le réseau transitoire induit disparaît soit par recombinaison, soit par diffusion des porteurs si le temps de recombinaison est assez long, ce qui permet alors de mesurer le coefficient de diffusion ambipolaire (eq.21).

Si l'un des faisceaux (pompe) est d'intensité très supérieure à l'autre (signal), on peut avoir amplification du faisceau signal par diffraction de Bragg du faisceau pompe sur le réseau induit.

VIII - LES HETEROJONCTIONS

Une hétérojonction est réalisée par épitaxie, sur un substrat semiconducteur, d'un autre semiconducteur de composition différente, donc de largeur de bande interdite différente. Pour que le matériau épitaxié ne présente pas trop de défauts, il faut que les mailles cristallines des deux composés soient aussi proches l'une de l'autre que possible, et la figure 1 montre les compositions d'alliages possibles en fonction des substrats utilisés.

Fig. 21 - Diagramme des bandes dans une hétérojonction GaAs/GaAlAs.

Considérons deux semiconducteurs de gaps respectifs E_{g1} et E_{g2} , d'affinité électronique X_1 et X_2 (énergie entre le bas de la bande de conduction et le niveau de libération du vide). Lorsqu'on accole les deux matériaux (fig. 21), à l'équilibre thermodynamique les niveaux de Fermi s'alignent, et dans le modèle

d'Anderson, les affinités électroniques sont conservées. Il en résulte que la différence des gaps se répartit entre les discontinuités des bandes de valence et de conduction, avec une courbure des bandes au voisinage de l'hétérojonction, qui dépend de la nature des matériaux et de leur dopage. Il peut apparaître à l'hétérojonction un puits de potentiel dans lequel les électrons s'accumulent en un gaz bidimensionnel /1, 26, 27/.

L'avantage principal d'une telle structure est qu'elle permet de séparer spatialement les électrons libres des impuretés de dopage qui leur ont donné naissance, et d'éviter ainsi de limiter la mobilité par interaction sur les impuretés pour des matériaux très dopés, en particulier à basse température. Les variations de mobilité en fonction de la température sont représentées sur la figure 22. Pour GaAs massif, la mobilité à haute température est limitée par les interactions avec les phonons optiques polaires, elle décroît ensuite lorsque la température diminue au-dessous de 50K, et est globalement d'autant plus faible que le dopage est important. Par contre, dans une structure GaAs / GaAlAs, dans laquelle les électrons du gaz bidimensionnel sont dans GaAs non dopé, au voisinage de l'hétérojonction avec GaAlAs qui contient les impuretés de dopage, la mobilité à basse température ne décroît pas et peut atteindre des valeurs supérieures à $10^6 \text{ cm}^2/\text{V.s}$, en particulier lorsqu'une fine couche de GaAlAs non dopé est intercalée entre les deux matériaux /28/.

Fig. 22 - Mobilité des électrons en fonction de la température.

De nombreuses structures sont possibles en associant plusieurs hétérojonctions /29,30/. Par exemple, en faisant croître successivement des couches GaAlAs / GaAs / GaAlAs avec une épaisseur de la couche centrale de faible gap de quelques dizaines d'Angström, on réalise un puits quantique : il apparaît une quantification des niveaux d'énergie des électrons dans le puits (fig. 23)

Fig. 23 - Quantification des niveaux d'énergie électroniques dans un puits quantique.

On peut associer plusieurs de ces puits quantiques pour former des structures plus complexes :

- si l'épaisseur de la barrière est assez grande (qq 100 Å), on obtient une structure à puits quantiques multiples (Fig. 24a)
- si au contraire l'épaisseur de la barrière est suffisamment faible pour que les fonctions d'onde des électrons dans les différents puits soient couplées entre elles (fig. 24b), la périodicité de l'empilement des couches qui se rajoute à celle du réseau cristallin est vue par les électrons. Il s'agit alors de super-réseaux, dont la structure de bande est modifiée par l'existence de cette périodicité supplémentaire, faisant apparaître des "mini-bandes" dans la zone de Brillouin.

Fig. 24 - Schéma de bande pour une structure à puits quantique multiple (a) et un super-réseau (b).

Comme dans les simples hétérojonctions, les électrons circulent dans les puits, alors que ce sont les barrières qui contiennent les impuretés de dopage, et ce sont donc des structures à haute mobilité.

La variété des matériaux et les possibilités quasi-infinies d'association entre eux font que les applications potentielles sont extrêmement nombreuses, et que des dispositifs complètement nouveaux peuvent être imaginés à l'aide du "band-gap engineering" qui est de mieux en mieux maîtrisé grâce aux diverses techniques d'épitaxie des matériaux.

IX - TRANSISTORS A EFFET DE CHAMP SUR SEMICONDUCTEURS III-V

Les transistors à effet de champ sont des composants unipolaires, ne mettant en jeu qu'un seul type de porteurs, et dans lesquels le courant circulant entre les électrodes de source et de drain est contrôlé par variation de la conductivité d'un canal conducteur, soit par modification de sa section (MESFET, JFET), soit par accumulation plus ou moins importante de porteurs le long d'une interface (MISFET, TEGFET), cette variation étant commandée par une troisième électrode, la grille [26].

Alors qu'en technologie silicium, l'excellente qualité de l'oxyde permet de réaliser des transistors à effet de champ à grille isolée (MOSFET) très performants, ceci n'existe pas pour GaAs sur lequel on ne peut pas obtenir d'oxyde ou d'isolant de bonne qualité. Par contre, le blocage du niveau de Fermi en surface en milieu de bande interdite permet de réaliser d'excellentes diodes Schottky par dépôt d'un métal sur le semiconducteur, et c'est ce qui est utilisé dans la technologie MESFET. Dans le cas d'InP, la réalisation de transistors avec un isolant de grille (MISFET), ou de transistors à jonction (JFET) est possible.

IX 1 - MESFET GaAs

Le principe de fonctionnement est rappelé sur la figure 25. A faible tension entre le drain et la source V_{DS} , pour une tension grille V_{GS} donnée, le courant I_{DS} est proportionnel à V_{DS} . L'épaisseur de la

Fig. 25 - Principe de fonctionnement du MESFET.

couche active conductrice étant contrôlée par la tension grille qui repousse les électrons, le courant I_{DS} est d'autant plus faible que celle-ci est négative. Lorsque la tension drain-source est augmentée, la zone de charge d'espace créée par la grille se déforme vers le drain, la différence de potentiel y étant plus importante. A partir d'une certaine valeur de V_{DS} , la zone de charge d'espace s'étend jusqu'au fond de la couche active,

et le canal est alors "pincé". Les électrons qui sont accélérés par la tension appliquée au drain traversent cette zone sous l'effet du champ électrique longitudinal qui y règne. Une augmentation de V_{DS} au delà de la tension de pincement fait légèrement reculer le point de pincement dans le canal, mais en modifie peu la résistance globale ; on observe alors une saturation du courant drain.

Ce principe de fonctionnement est vérifié si la longueur de la grille du transistor est supérieure au micron. Pour des grilles submicroniques, les phénomènes de transport non-stationnaire que nous avons décrits interviennent. Le canal sous la grille étant très court, et le champ électrique élevé, les électrons y sont en régime de survitesse jusqu'à ce que leur énergie atteigne l'énergie de transfert vers les vallées latérales de la bande de conduction ; la vitesse se sature alors brusquement, et c'est ce phénomène qui produit la saturation du courant drain, sans que le canal soit pincé, comme le montrent les simulations particulières Monte-Carlo [1].

Si le blocage du niveau de Fermi en surface permet la réalisation de la grille par contact Schottky, il crée également comme nous l'avons vu une zone de charge d'espace sous la surface du semiconducteur. Ceci réduit la section de la couche active dans les zones d'accès entre la source et la grille et entre la grille et le drain, et empêche toute commande par une polarisation positive de la grille, le courant étant limité par ces zones d'accès quelle que soit la tension grille. Une façon de pallier cet inconvénient est de creuser la couche active à l'emplacement de la grille avant de déposer celle-ci (fig. 26). De nouvelles technologies se développent également utilisant des métaux réfractaires pour la grille afin de pouvoir aligner les contacts de source et de drain de part et d'autre de celle-ci.

Fig. 26 - Structure d'un MESFET à grille enterrée.

IX 2 - Transistor à effet de champ à hétérojonction (TEGFET ou HEMT)

Le temps de réponse des dispositifs électroniques est directement lié au temps de transit des porteurs dans le composant. C'est pour cette raison que la plupart des transistors mettent en jeu un transport électronique plutôt qu'un transport par trous, dont la mobilité serait plus faible, et que l'on a cherché à réaliser des dispositifs sur semiconducteurs III-V, où la mobilité des porteurs est plus élevée que dans le silicium. Il est donc naturel de chercher à mettre à profit les fortes mobilités observées dans les gaz bidimensionnels d'électrons au voisinage d'une hétérojonction. Par ailleurs, les fortes densités électroniques qu'ils procurent sans altération de la mobilité, contrairement à ce qui se produit lorsqu'on augmente le nombre de porteurs libres par augmentation du dopage dans un matériau massif, assurent une charge électronique suffisante pour la commande d'autres composants par le courant de sortie du premier.

Le schéma de principe d'un HEMT est donné sur la figure 27, avec la structure de bande associée. La couche de GaAlAs est généralement assez mince pour être totalement déserte d'électrons, donc non conductrice. La grille commande la densité d'électrons dans le gaz 2D, et permet, pour des tensions suffisamment négatives, de pincer le canal.

Fig. 27 - Schéma de principe d'un HEMT.

Comme dans le cas du MESFET, pour les longueurs de grille submicroniques, les effets de transport non-stationnaire entrent en jeu. Un phénomène supplémentaire apparaît cependant dans le cas des transistors GaAs / GaAlAs : la discontinuité de la bande de conduction à l'hétérojonction étant plus petite que l'écart en énergie entre les vallées Γ et L de GaAs, les électrons acquièrent assez d'énergie pour passer dans la couche de GaAlAs avant de pouvoir être transférés dans les vallées L de GaAs. Il y a transfert dans l'espace réel, avec chute de mobilité puisque les électrons arrivent dans un matériau très dopé. Les vallées latérales étant plus proches en énergie de la vallée Γ dans GaAlAs que dans GaAs, les électrons y sont facilement transférés, et leur mobilité est considérablement réduite. Ce transfert spatial a en outre l'inconvénient d'autoriser un courant de fuite qui peut être important sur la grille du transistor.

D'autres structures dérivées de la structure ci-dessus ont été proposées [1], comme le SISFET ou le MIS-like FET, dans lesquelles on cherche à éviter tout courant grille en intercalant un isolant, ou un semiconducteur de forte résistivité.

X - TRANSISTORS BIPOLAIRES

Un transistor bipolaire est caractérisé par une faible résistance d'entrée, une résistance de sortie élevée, et un gain en courant voisin de 1, ce qui en fait un amplificateur de tension. Le principe de fonctionnement est basé sur le fait que dans une jonction P-N polarisée en inverse, le courant est très sensible au flux de porteurs minoritaires arrivant en limite de la zone de charge d'espace (électrons du côté P, trous du côté N), ceux-ci étant accélérés par le champ électrique régnant dans cette région. Le contrôle du flux de porteurs minoritaires est équivalent à celui du courant de fuite de la jonction. L'injection des porteurs minoritaires est assurée par une autre jonction polarisée en direct. Un transistor bipolaire est donc formé par une structure N-P-N (ou P-N-P) comme celle schématisée sur la figure 28, comportant un émetteur de type N, une base de type P et un collecteur de type N, la dénomination des trois régions spécifiant leur rôle.

Fig. 28 - Schéma d'un transistor bipolaire.

Pour que le gain en courant soit aussi voisin de 1 que possible, il faut que le courant d'électrons injecté de l'émetteur vers la base soit très supérieur au courant de trous injecté de la base vers l'émetteur (efficacité d'injection de l'émetteur), et que la majorité des électrons injectés par l'émetteur soit recueillie au collecteur (facteur de transport dans la base) [26].

En ce qui concerne le facteur de transport dans la base, pour qu'il soit proche de 1 il faut que le temps de recombinaison des électrons avec les trous de la base soit très supérieur à leur temps de transit, donc que l'épaisseur de la base soit faible, ce qui va également dans le sens d'une réduction du temps de réponse du transistor.

Le courant d'électrons traversant la jonction émetteur-base polarisée en direct s'écrit [26] :

$$J_e = e \frac{n_{iB}^2}{N_A} \frac{D_n}{w_B} \left(\exp \frac{eV_e}{kT} - 1 \right) \quad (29)$$

où n_{iB} , N_A et w_B sont respectivement la concentration intrinsèque, le dopage et l'épaisseur de la base, D_n le coefficient de diffusion des électrons, V_e la barrière de potentiel vue par les électrons.

De la même façon, le courant de trous de la base vers l'émetteur est :

$$J_p = e \frac{n_{iE}^2}{N_D} \frac{D_p}{w_E} \left(\exp \frac{eV_p}{kT} - 1 \right) \quad (30)$$

n_{iE} , N_D et w_E étant respectivement la concentration intrinsèque, le dopage et l'épaisseur de l'émetteur, D_p le coefficient de diffusion des trous, V_p la barrière de potentiel correspondante.

Le rapport des courants J_p / J_e est donc proportionnel à la quantité :

$$\frac{n_{iE}^2}{n_{iB}^2} \frac{N_A}{N_D} \frac{w_B}{w_E} \exp \left(\frac{e(V_p - V_e)}{kT} \right)$$

Pour une homojonction, les matériaux étant identiques de part et d'autre, les seuls facteurs qui diffèrent sont les rapports N_A/N_D et w_B/w_E . Une bonne efficacité d'injection demande donc que l'émetteur du transistor soit beaucoup plus dopé que la base. Or, si la base n'est pas elle-même suffisamment dopée, la résistance d'accès entre la partie active de la base et la source de polarisation est élevée, et par ailleurs la pénétration de la zone de charge d'espace base-collecteur dans la base quand la polarisation inverse de cette jonction augmente induit une variation de l'épaisseur de base néfaste au fonctionnement du transistor. C'est pourquoi l'utilisation d'une hétérojonction apporte de nets avantages aux performances des transistors bipolaires.

En effet, pour une hétérojonction avec un émetteur formé d'un matériau à grand gap et d'une base à petit gap, on a $n_{iE}^2 \ll n_{iB}^2$ et $\exp(e(V_p - V_e)/kT) > 0$ puisque la barrière de potentiel est plus élevée pour les trous du fait de la discontinuité de la bande de valence; ceci permet d'avoir un dopage plus élevé dans la base que dans l'émetteur sans que l'efficacité d'injection en souffre.

Fig. 29 - Structure de bande d'une hétérojonction abrupte (a), graduelle (b), décalée (c).

Pour une hétérojonction abrupte, la discontinuité de la bande de conduction fait souvent apparaître un pic qui introduit une barrière à l'injection des électrons (fig. 29a). Cet inconvénient disparaît si l'on réalise une hétérojonction graduelle, dans laquelle la composition du matériau varie progressivement (fig. 29b). Ce pic est également un handicap si l'on veut réaliser une structure symétrique, avec un collecteur de même matériau que l'émetteur : les électrons s'accumulent contre l'hétérojonction et ne peuvent être

collectés par le collecteur. La plupart des transistors bipolaires à hétérojonction ont donc un collecteur de même nature que la base. Une coupe typique est donnée sur la figure 30. Dans les cas où la structure symétrique est nécessaire, la séparation de la jonction électrique P-N, restant une homojonction, et de l'hétérojonction réduit cet effet (fig. 29c).

Fig. 30 - Structure d'un transistor bipolaire à hétérojonction.

L'épaisseur de la base étant faible, le transport non-stationnaire peut là encore être mis à profit pour augmenter la rapidité des dispositifs. Pour des épaisseurs de base de l'ordre de 100nm, les électrons peuvent la traverser en régime de transport balistique. Le problème est d'éviter autant que possible la zone d'accélération des électrons qui réduit leur vitesse moyenne sur la traversée de la base, et donc le temps de transit. On utilise alors la discontinuité de la bande de conduction pour que les électrons soient injectés dans la base avec une énergie de l'ordre de 300 meV. Le champ électrique dans la base étant très faible, ils la traversent sans atteindre l'énergie de transfert vers les vallées latérales, et une majorité d'entre eux peuvent atteindre le collecteur sans subir d'interactions avec les phonons optiques polaires, donc en vol balistique.

Fig. 31 - Pourcentage d'électrons balistiques en fonction de la distance parcourue dans la base.

La figure 31 montre le pourcentage d'électrons balistiques, et d'électrons ayant subi 1, 2, ou 3 interactions en fonction de l'épaisseur de base traversée /10/.

Des transistors bipolaires à hétérojonction à très haute fréquence de coupure ont été réalisés, avec des progrès très rapides au cours des deux dernières années, les fréquences maximales étant proches de 100 GHz.

CONCLUSION

Les matériaux III-V présentent des propriétés très intéressantes pour l'électronique rapide. Les mobilités électroniques élevées, en particulier dans les gaz d'électrons bidimensionnels ou les structures à puits quantiques à basse température, la possibilité de mettre à profit les phénomènes de transport non-stationnaire qui diminuent encore le temps de transit des porteurs dans les composants, et d'associer des matériaux de différentes compositions font que des transistors à très haute fréquence sont réalisés. Par ailleurs, les propriétés optiques largement utilisées en optoélectronique font que ce sont les semiconducteurs choisis actuellement pour l'intégration micro-optoélectronique. Le développement des diodes laser à 1,3 et 1,5 μm favorise les recherches dans la filière sur substrat InP, avec les composés du type GaInAsP bien que la technologie des composants électroniques soit moins avancée que dans la filière GaAs / GaAlAs.

Cependant, les progrès des hétéroépitaxies de matériaux III-V sur Silicium, et la recherche de diodes laser à émission à plus grande longueur d'onde (2 à 4 μm) risquent de modifier le développement actuel de la microoptoélectronique III-V.

REMERCIEMENTS

Je tiens à remercier vivement C. BRU, R. CASTAGNE, et P.HESTO pour les documents et résultats qu'ils ont bien voulu me communiquer pour la préparation de ce cours, ainsi que P. DANSAS, M. MOUIS, N. PARAIRE, D.PASCAL et J.F. PONE pour de nombreuses et fructueuses discussions dans ce domaine.

REFERENCES

- /1/ CASTAGNE, R., DUCHEMIN, J.P., GLOANNEC, M., RUMELHARD, Ch., Circuits Intégrés en Arséniure de Gallium, Collection technique et scientifique des télécommunications, Masson (1989)
- /2/ PANKOVE, J.L., Optical processes in semiconductors, Dover publications Inc. (1971)
- /3/ WILLARDSON, R.K., BEER, A.C., Semiconductors and semimetals, Academic Press.
- /4/ KITTEL, C, Introduction à la physique de l'état solide, Dunod (1972)
- /5/ CHELIKOWSKY, J.R., COHEN, M.L., Phys. Rev B 14 (1976) 556.
- /6/ BLAKEMORE, J.S., J. Appl. Phys. 53 (1982) R123.
- /7/ BRILL LOUIN, L., Wave propagation and group velocity, New-York Academic Press (1960).

- /8/ POULET, H., MATHIEU, J.P., Spectres de vibration et symétrie des cristaux, Gordon and Breach(1970).
- /9/ DOLLING, G., WAUGH, J.L.T., Lattice dynamics, WALLIS, R.F., Ed., Pergamon Press (1963) 19.
- /10/ HESTO, P., Doctorat d'Etat, Orsay (1984).
- /11/ BRU, C., Doctorat d'Etat, Orsay (1988)
- /12/ SHANK, C.V., FORK, R.L., GREENE, B.I., et al., Appl. Phys. Lett. 38 (1981) 104
- /13/ ALLEN, S.J., TSUI, D.C., DE ROSA, F., THORNBUR, K.K., WILSON, B.A., J. Phys.Paris C7 42 (1981) 369.
- /14/ HEIBLUM, M., NATHAN, M.I., THOMAS, D.C., KNOEDLER, C.M., Phys. Rev. Lett. 55 (1985) 2200.
- /15/ LAVAL, S., BRU, C., CASTAGNE, R., ARNODO, C., Inst. Phys. Conf. Series 56 (1980) 171.
- /16/ DANSAS, P., BOUCHEMAT, M., BRU, C., PASCAL, D., LAVAL, S., Solid State Electronics 41 (1988) 1327.
- /17/ OUDAR, J.L., Nonlinear Optics : Materials and Devices, FLYTZANIS, C., OUDAR, J.L., Ed., Springer-Verlag (1986).
- /18/ STURGE, M.D., Phys. Rev. 127 (1962) 768.
- /19/ KITTEL, C, Théorie quantique du solide, Dunod (1967).
- /20/ BOUCHEMAT, M., Thèse de Doctorat, Orsay (1986)
- /21/ BLØMBERGEN, Nonlinear Optics, W.A. Benjamin, Inc. (1965)
- /22/ YARIV, A., Introduction to Optical Electronics, Holt, Rinehard & Winston (1976).
- /23/ KOLNER, B.H., BLOOM, D.M., IEEE Journ. QE 22 (1986) 79.
- /24/ GIBBS, H., Optical Bistability : Controlling Light with Light, Academic Press (1985).
- /25/ HUIGNARD, J.P., ROOSEN, G., Nonlinear Optics : Materials and Devices, FLYTZANIS, C., OUDAR, J.L., Ed., Springer-Verlag (1986).
- /26/ VAPAILLE, A., CASTAGNE, R., Physique et technologie des dispositifs à semiconducteurs, Dunod, (1987).
- /27/ ANDO, T., FOWLER, A.B., STERN, F., Review of Modern Physics 54 (1982) 437.
- /28/ HIRTH, Doctorat d'Etat, Paris VI (1985).
- /29/ OUDAR, J.L., PALMIER, J.F., L'Echo des Recherches 114 (1983) 21.
- /30/ WEISBUCH, C., Device and circuit applications of III-V super lattice and modulation doping, DINGLE, R., Ed., Academic Press (1987).