

Table des matières

Table des matières	2
Chapitre 1 : Semi-conducteurs hors équilibre	3
1 Introduction.	3
2 Recombinaison et génération.	3
3 Courants dans les semi-conducteurs.	4
3.1 Courant de conduction.	4
3.2 Courant de diffusion.	5
3.3 Courant total.	5
4 L'équation de continuité – longueur de diffusion.	6
Chapitre 2 : Jonction PN	8
1 Introduction.	8
2 Jonction PN à l'équilibre thermodynamique.	8
2.1 Tension de diffusion :	9
2.2 Champ et potentiel dans la ZCE	10
2.2.1 Champ électrique $E(x)$:	10
2.2.2 Potentiel électrique $V(x)$:	12
2.2.3 Largeur de la ZCE W :	12
3 Jonction PN sous polarisation.	12
3.1 Densité de porteurs injectés - approximation de Boltzmann.	13
3.2 Distributions des porteurs dans les régions neutres.	14
Chapitre 3 Chapitre 4 : Transistor Bipolaire	34
Figure 6.1.a : capacité MOS Figure 6.1.b : les différents paramètres énergétiques.....	67
Chapitre 4 Figure 6.3 : variation de la charge en fonction du potentiel de surface dans une structure MOS.	73
Chapitre 5 Figure 6.4 : bandes d'énergie et densité de charge associée	74
Chapitre 6 Figure 6.6 : Vue en coupe d'un MOS-FET (Taur)	81
Figure 6.7 : schéma simplifié du MOS (Taur)	82
Chapitre 7 88	
Figure 6.10 : variation de la mobilité (a) et de la vitesse (b) en fonction du champ électrique	89
Figure 6.9 : claquage du transistor MOS-FET (Taur et al)	90
Figure 6.10 : origines des capacités (Singh)	92
Chapitre 8 Figure 7.1 : symbole, table de vérité et caractéristique de transfert idéale d'un inverseur 8-96	
Chapitre 9 Figure 7.2 : schéma générique d'un inverseur NMOS	9-97

Chapitre 1 : Semi-conducteurs hors équilibre

1 Introduction.

Ce chapitre va nous fournir les outils qui vont nous permettre dans la suite de ce cours de comprendre le fonctionnement des composants actifs à semi-conducteurs. Dans le cours de « propriétés de la matière » de ESINSA 3, on a envisagé l'étude des semi-conducteurs sans polarisation, soit encore des semi-conducteurs à l'équilibre thermodynamique. Le composant étant « fait » pour être polarisé, on doit donc modifier notre perception du semi-conducteur qui devient polarisé, donc dans un état hors équilibre thermodynamique.

2 Recombinaison et génération.

Lorsque le semi-conducteur est à l'équilibre thermodynamique, la densité de porteurs obéit à la **loi d'action de masse** :

$$pn = n_i^2 \quad (1)$$

Cette relation n'est plus vérifiée dans le cas d'un semi-conducteur hors équilibre et sera remplacée par une relation que l'on verra en fin de chapitre. Cette modification de la **loi d'action de masse** provient des phénomènes de **génération et de recombinaison**. On caractérise la création de porteurs dans le semi-conducteur par un paramètre g' qui mesure le nombre de porteurs qui se créent par unité de temps et de volume, soit $[g'] = s^{-1}cm^{-3}$. De la même manière, le nombre de porteurs qui se recombinent, est donné par r' qui a les mêmes dimensions que g' . Bien entendu, en régime stationnaire le nombre de porteurs est constant et $g' = r'$. Il faut noter que r' est fonction des propriétés du matériau et que g' est fonction d'un terme « thermique » (saut d'un électrons de la BV vers la BC par exemple) et d'un phénomène externe (rayonnement, excitation externe). De ce fait, on écrit généralement $g' = g + g_{th}$. Le **taux net de recombinaison** s'écrit alors :

$$g' - r' = g + g_{th} - r' = g - r \quad \text{avec} \quad r = r' - g_{th} \quad (2)$$

La variation du nombre, par exemple d'électrons par unité de temps et de volume est donc donnée par :

$$\frac{dn}{dt} = g - r \quad (3)$$

Le taux g est fonction du processus d'excitation. La recombinaison des porteurs définit par r s'effectue de deux manières distinctes : une directe qui est proportionnelle à l'excédent d'électrons par rapport à l'équilibre et une par centre d'impuretés ou de recombinaisons. Dans le cas de recombinaison directe, le taux de recombinaison des électrons (trous) r_n (r_p) s'écrit :

$$r_n = \frac{\Delta n}{\tau_n} = \frac{n - n_0}{\tau_n} \quad (4a)$$

$$r_p = \frac{\Delta p}{\tau_p} = \frac{p - p_0}{\tau_p} \quad (4b)$$

avec τ la **durée de vie** des électrons ou des trous.

Dans le cas de matériau où il y a peu de porteurs libres, la recombinaison est **assistée** par **centres de recombinaisons** et s'écrit :

$$r = \frac{1}{\tau} \frac{pn - n_i^2}{2n_i + p + n} \quad (5)$$

C'est l'équation de Shockley-Read.

3 Courants dans les semi-conducteurs.

Dans la plupart des dispositifs électroniques, le courant (le déplacement de charges) a pour origine les champs électriques et les gradients de concentration de porteurs. Le courant est alors la somme d'un courant de conduction et d'un courant de diffusion respectivement.

3.1 Courant de conduction.

Dans un cristal, à température finie, du fait de l'agitation thermique, les électrons possèdent une **vitesse thermique** donnée par :

$$V_{th} = \sqrt{\frac{3kT}{m^*}} \quad (6)$$

où m^* est la **masse effective** des électrons dans un potentiel périodique. Cette diffusion (les chocs) des électrons est isotropes, c'est à dire qu'en « moyenne » ces électrons sont immobiles. En revanche, si on applique un champ électrique E , la force électrique résultante va induire un mouvement de dérive qui se superpose au mouvement désordonné de l'agitation thermique. Entre deux chocs espacés en moyenne de τ , l'électron acquiert une vitesse moyenne $\bar{v} = \frac{-e\tau}{m^*} E$, τ étant le temps moyen entre deux

chocs, c'est le **temps de relaxation**. En introduisant la mobilité qui est toujours positive, on peut réécrire la vitesse :

$$v_n = -\mu_n E \quad \text{avec} \quad \mu_n = \frac{e\tau}{m_e^*} \quad (7a)$$

$$v_p = +\mu_p E \quad \text{avec} \quad \mu_p = \frac{e\tau}{m_h^*} \quad (7b)$$

L'équation aux dimensions de la mobilité donne $[\mu] = cm^2V^{-1}s^{-1}$. Pour le silicium la mobilité des électrons est de l'ordre de $2000 cm^2V^{-1}s^{-1}$.

Si on veut calculer la densité de courant de conduction, qui est le nombre de charges qui traverse une surface unité par unité de temps, alors cette densité de courant s'écrit pour les électrons (J_n) et pour les trous (J_p) :

$$\begin{aligned} J_n &= -nev_n = ne\mu_n E \\ J_p &= nev_p = ne\mu_p E \end{aligned} \quad (8)$$

La densité de courant de conduction est bien évidemment la somme des deux composantes, ce qui donne :

$$J_C = J_n + J_p = (ne\mu_n + pe\mu_p)E \quad (9)$$

En définissant la conductivité σ du matériau, l'équation (9) se réécrit $J_C = \sigma E$, ce qui est une autre forme de la **loi d'Ohm**.

3.2 Courant de diffusion.

Lorsque les porteurs de charges ne sont pas uniformément répartis, distribués dans le matériau, ces porteurs de charges diffusent depuis la région à forte concentration vers la région à faible concentration. On dit que ces porteurs subissent le **processus général de la diffusion** décrit par la **première loi de Fick**. Si on définit par n_D^x , le nombre d'électrons qui diffuse dans la direction x par unité de temps et de surface (c'est le flux), et par D_n la constante de diffusion des électrons (qui dépend entre autre de la température), alors on peut écrire :

$$n_D^x = -D_n \frac{dn}{dx} \quad (10)$$

Le signe moins traduit le fait que les porteurs libres diffusent dans la direction de plus faible concentration. Une relation similaire est obtenue pour les trous :

$$p_D^x = -D_p \frac{dp}{dx} \quad (11)$$

La densité de courant de diffusion est la somme des deux contributions (électrons et trous), soit encore :

$$J_D = e(-n_D^x + p_D^x) = eD_n \frac{dn}{dx} - eD_p \frac{dp}{dx} \quad (12)$$

3.3 Courant total.

La densité de courant dans le semi-conducteur est la somme des deux contributions (diffusion et conduction) et s'écrit :

$$J_T = J_D + J_C \quad (13)$$

soit, en exprimant le courant d'électrons et le courant de trous :

$$J_n = ne\mu_n E + eD_n \frac{dn}{dx} \quad (14a)$$

$$J_p = pe\mu_p E - eD_p \frac{dp}{dx} \quad (14b)$$

La mobilité et la constante de diffusion indiquent l'aptitude des porteurs libres à se mouvoir dans le matériau semi-conducteur. On conçoit alors qu'ils doivent être intimement liés, la relation entre ces deux coefficients étant la **relation d'Einstein** :

$$\frac{D}{\mu} = \frac{kT}{e} \quad (15)$$

4 L'équation de continuité – longueur de diffusion.

Pour calculer les courants, on n'a pas tenu compte des processus de recombinaison et de génération qui altèrent la distribution des porteurs dans le matériau et donc le courant. Pour décrire le transport et la recombinaison des électrons et des trous injectés, on doit appliquer une équation que l'on appelle équation de continuité. Soit la figure 1 qui définit une région de l'espace de volume $A\Delta x$. Si on s'intéresse à la variation du nombre de particules dans ce volume, on doit écrire que cette variation est égale à ce qui rentre (du au courant) moins ce qui sort (le courant) plus ce qui se génère moins ce qui se recombine :

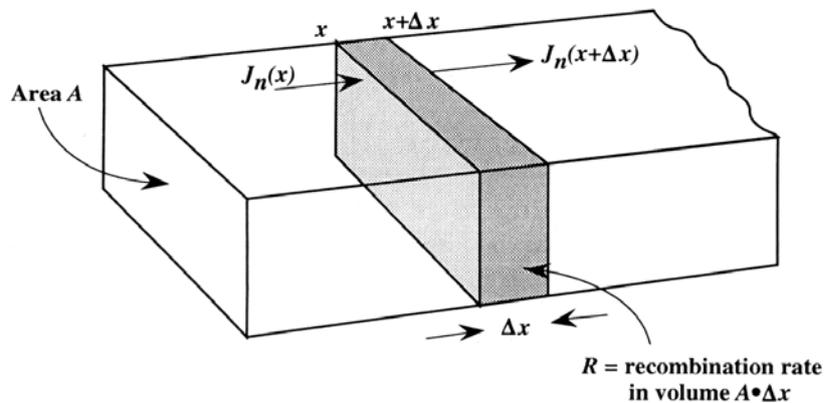


Figure 1-1 : conservation des particules appliquée à un volume $A\Delta x$ (d'après Singh)

$$A\Delta x \frac{dn(x,t)}{dt} = A \left[\frac{J_n(x + \Delta x)}{e} - \frac{J_n(x)}{e} \right] - R + G \cong A \frac{dJ_n(x)}{dx} \frac{\Delta x}{e} - R + G \quad (16)$$

Soit encore :

$$\frac{dn(x,t)}{dt} = \frac{1}{e} \frac{dJ_n}{dx} - r_n + g_n \quad (17a)$$

$$\frac{dp(x,t)}{dt} = -\frac{1}{e} \frac{dJ_p}{dx} - r_p + g_p \quad (17b)$$

Les équations 17a et 17b représentent les **équations de continuité** pour les électrons et les trous respectivement.

Appliquons, par exemple ces équations, dans le cas où le courant est un courant de diffusion exclusivement. En se servant de l'équation 12, on obtient :

$$\begin{aligned} J_n(\text{diff}) &= eD_n \frac{dn}{dx} \\ J_p(\text{diff}) &= -eD_p \frac{dp}{dx} \end{aligned} \quad (18)$$

soit, à partir des équations (17),

$$\frac{dn}{dx} = D_n \frac{d^2 n}{dx^2} - \frac{n - n_0}{\tau_n} \quad (19a)$$

$$\frac{dp}{dx} = D_p \frac{d^2 p}{dx^2} - \frac{p - p_0}{\tau_p} \quad (19b)$$

En régime stationnaire, les dérivées par rapport au temps s'annulent et les équations 19a et 19b deviennent :

$$\frac{d^2(n - n_0)}{dx^2} = \frac{n - n_0}{D_n \tau_n} = \frac{n - n_0}{L_n^2} \quad (20a)$$

$$\frac{d^2(p - p_0)}{dx^2} = \frac{p - p_0}{D_p \tau_p} = \frac{p - p_0}{L_p^2} \quad (20b)$$

On a introduit dans les équations ci-dessus, les **longueurs de diffusion** des électrons (L_n) et des trous (L_p). Ces longueurs de diffusion représentent les distances sur lesquelles les porteurs injectés dans le matériau diminuent d'un facteur $1/e$. La **longueur de diffusion** représente également la distance moyenne qu'un électron (un trou) va parcourir par diffusion avant qu'il ne se recombine avec un trou (un électron).

Chapitre 2 : Jonction PN

1 Introduction.

La nécessité de réaliser un composant électronique à réponse non linéaire (c'est une des demandes les plus fortes en électronique) a conduit à l'élaboration en premier lieu de jonctions qui font parties d'une classe importante de composants : les dispositifs «redresseurs» ou «rectifier devices». Un simple semi-conducteur ne peut avoir ce comportement mais la juxtaposition de deux matériaux semi-conducteurs de « type » opposés (c'est la **jonction PN**) conduit à ce comportement. On verra dans le chapitre suivant qu'un contact Métal - Semi-conducteur conduit au même résultat mais à un avantage en vitesse de commutation, c'est la **diode Schottky**.

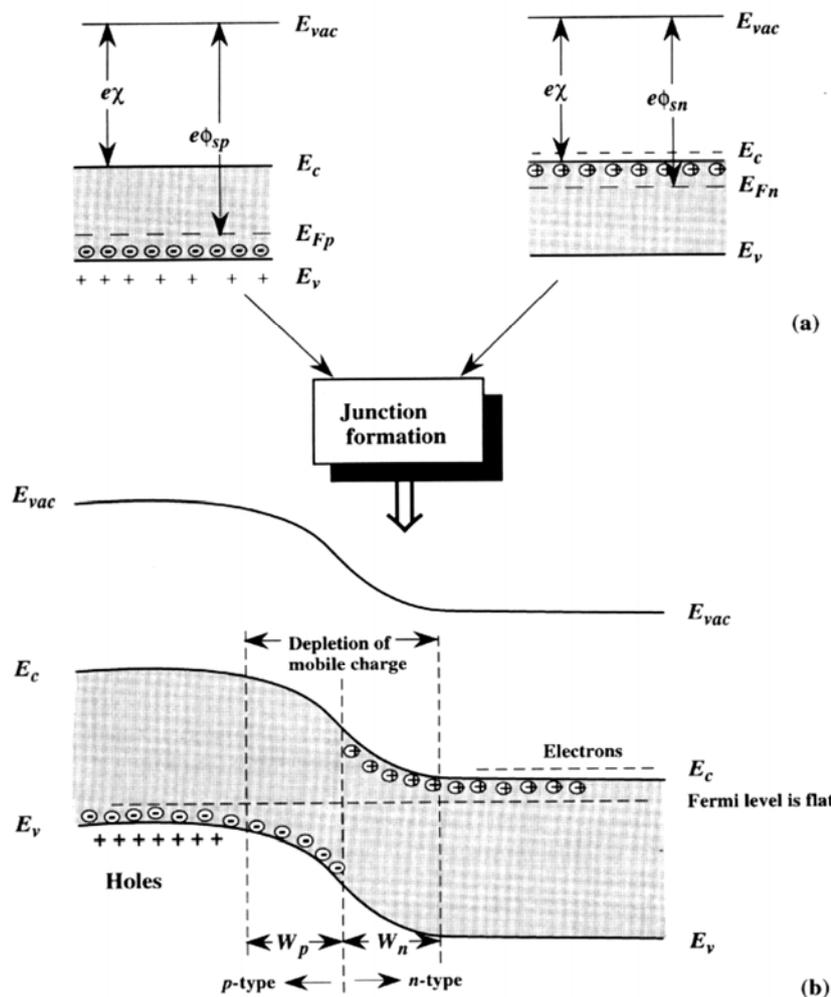


Figure 2-1 formation de la jonction PN (Singh)

2 Jonction PN à l'équilibre thermodynamique.

La jonction PN est, comme son nom l'indique la juxtaposition d'un semi-conducteur de type P dopé avec N_A accepteurs et d'un semi-conducteur de type N dopé avec N_D donneurs. La situation lorsque les deux parties ne sont pas en contact est représentée sur la figure 1a. La mise en contact conduit à un processus de mise à l'équilibre de la

jonction. Les électrons étant abondants (ce sont les porteurs majoritaires) dans la région n diffusent vers le matériau de type p, le même processus ayant lieu pour les trous mais dans le sens opposé. Ce déplacement de porteurs libres laisse la place aux atomes dopants ionisés à la frontière des deux régions et dont la charge n'est plus compensée : apparaît alors un champ électrique interne E_i à l'interface des deux matériaux qui va s'opposer au processus de diffusion des porteurs. Ce champ va croître avec le processus de diffusion jusqu'à l'obtention d'un équilibre entre les forces de diffusion et les forces électriques créées par E_i . Trois régions peuvent être alors identifiées :

- La région P loin de la jonction où le matériau est neutre et les bandes d'énergie sont plates.
- La région N loin de la jonction où là encore le matériau est neutre et les bandes d'énergie sont plates.
- La région à l'interface des deux matériaux qui est vide de porteur, on parle de zone dépeuplée (ou désertée, de déplétion, déplétée) qui est elle chargée : c'est la **zone de charge d'espace** ou **ZCE**.

2.1 Tension de diffusion :

Cette tension que l'on désigne par V_D ou V_{bi} (pour « built-in potential »), est par définition la différence de potentiel entre la région neutre N et la région neutre P :

$$V_D = V_{bi} = V_N - V_P \quad (1)$$

Pour la calculer, il suffit d'écrire que le système est à l'équilibre thermodynamique, c'est à dire que la diode (nom courant d'une jonction PN) est non polarisée. Une des façons de l'écrire est de dire que le courant de trous (ou des électrons) est nul. Si on s'intéresse aux trous, la densité de courant $J_p(x)$ est la somme de deux courants : un courant de diffusion et un courant de conduction, et cette somme est nulle :

$$J_p(x) = e \left[\mu_p p(x) E(x) - D_p \frac{dp(x)}{dx} \right] = 0 \quad (2)$$

Le premier terme de l'équation (2) est la composante conduction et le deuxième la composante diffusion. Ce qui se réécrit :

$$\frac{\mu_p}{D_p} E(x) = \frac{1}{p(x)} \frac{dp(x)}{dx} \quad (3)$$

En faisant apparaître le potentiel ($E(x) = -dV(x)/dx$) et à partir de la relation d'Einstein, on obtient :

$$\frac{-e}{kT} \frac{dV(x)}{dx} = \frac{1}{p(x)} \frac{dp(x)}{dx} \quad (4)$$

En intégrant de la région P à la région N, on obtient :

$$\frac{-e}{kT} \int_{V_p}^{V_n} dV(x) = \int_{p_p}^{p_n} \frac{dp(x)}{p(x)} \quad (5)$$

où p_p et p_n sont les densités de trous dans la région P et N respectivement. On obtient alors l'expression du potentiel de diffusion :

$$V_D = \frac{kT}{e} \ln\left(\frac{p_p}{p_n}\right) \quad (6a)$$

$$V_D = \frac{kT}{e} \ln\left(\frac{n_n}{n_p}\right) \quad (6b)$$

L'équation (6b) étant obtenue à partir de l'expression du courant d'électrons. En exprimant p_p et p_n en fonction des taux de dopage et de n_i , les équations (6a) et (6b) s'écrivent :

$$V_D = \frac{kT}{e} \ln\left(\frac{N_A N_D}{n_i^2}\right) \quad (7)$$

Nota : Lorsque la diode sera polarisée par une tension externe V , les expressions que nous allons écrire par la suite seront valables en remplaçant V_D par $V_D + V$.

2.2 Champ et potentiel dans la ZCE

Pour calculer le profil de potentiel $V(x)$ et le profil du champ électrique $E(x)$ dans la zone de charge, il suffit d'intégrer l'équation de Poisson entre $-W_p$ et W_n , qui sont les frontières de la ZCE :

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_{sc}} \quad (8)$$

avec $\rho(x)$ la densité de charges dans la structure. Elle est bien évidemment nulle dans les régions neutres. Si on considère que la ZCE est vide de porteurs, alors :

$$\frac{d^2V(x)}{dx^2} = -\frac{e}{\epsilon_{sc}} N_D \quad \text{pour } 0 < x < W_N \quad (9a)$$

$$\frac{d^2V(x)}{dx^2} = +\frac{e}{\epsilon_{sc}} N_A \quad \text{pour } -W_p < x < 0 \quad (9b)$$

2.2.1 Champ électrique $E(x)$:

Pour obtenir le champ ($E(x) = -\frac{dV(x)}{dx}$), il suffit d'intégrer une fois l'équation de Poisson, ce qui donne :

$$E(x) = +\frac{eN_D}{\epsilon_{sc}}(x - W_N) \quad \text{pour } 0 < x < W_N \quad (10a)$$

$$E(x) = -\frac{eN_A}{\epsilon_{sc}}(x + W_p) \quad \text{pour } -W_p < x < 0 \quad (10b)$$

La continuité du vecteur déplacement (en $x = 0$), qui se réduit ici à la continuité du champ électrique conduit à une relation importante entre les dopages et les extensions des ZCE dans les deux régions. A partir de 10a et 10b, et en faisant $x = 0$, on obtient :

$$N_D W_N = N_A W_P \quad (11)$$

On voit également que le champ est maximum en $x = 0$ et vaut :

$$E_M = -\frac{eN_D W_N}{\epsilon_{sc}} = -\frac{eN_A W_P}{\epsilon_{sc}} \quad (12).$$

Le fait qu'il soit négatif indique tout simplement que dans notre étude, il est orienté vers les x négatifs.

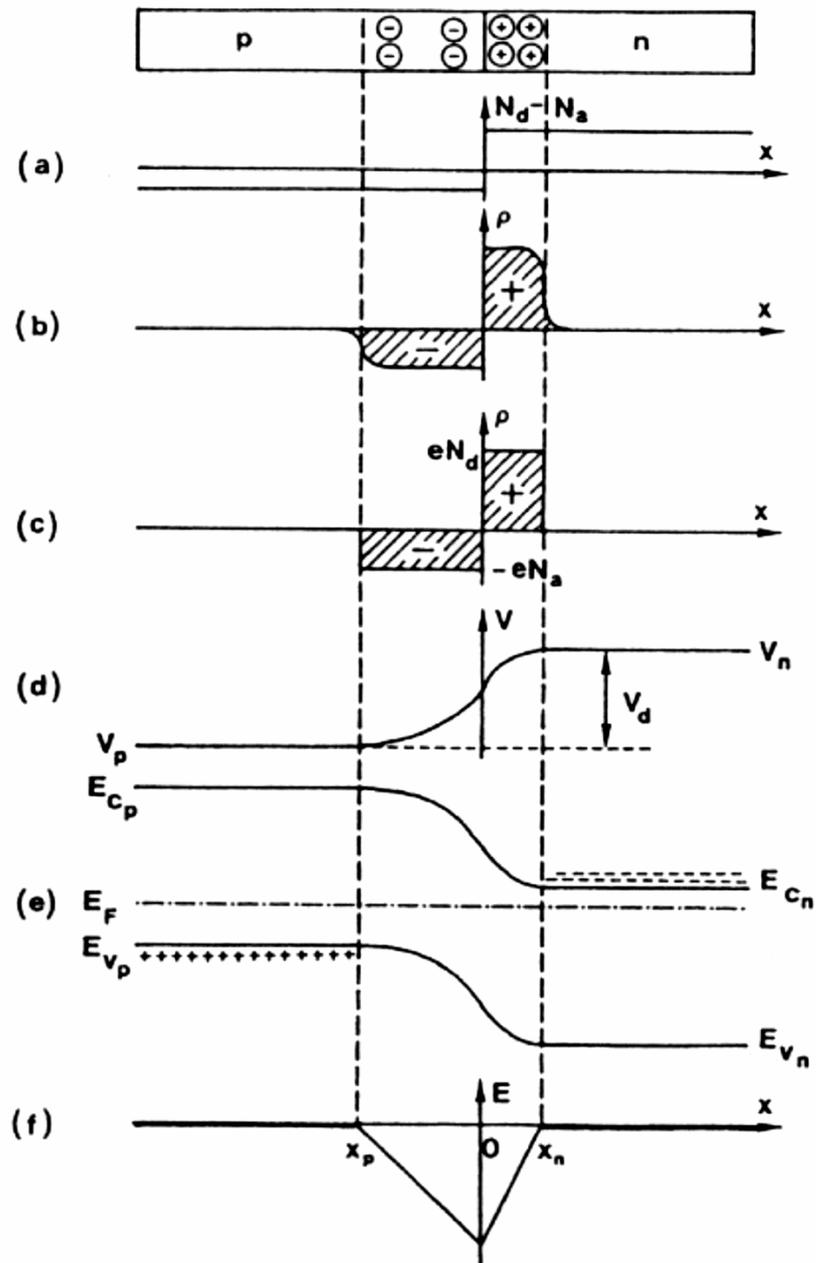


Figure 2-2 : répartition a) du dopage, b) densité de charges réelle, c) dans le modèle abrupt, d) répartition du potentiel, e) des bandes d'énergie, f) du champ électrique (Mathieu).

2.2.2 Potentiel électrique $V(x)$:

En intégrant le champ électrique, on obtient tout naturellement :

$$V(x) = -\frac{eN_D}{\epsilon_{sc}}(x - W_N)^2 + V_n \quad \text{pour } 0 < x < W_N \quad (13a)$$

$$V(x) = \frac{eN_A}{\epsilon_{sc}}(x + W_p)^2 + V_p \quad \text{pour } -W_p < x < 0 \quad (13b)$$

V_n et V_p étant les potentiels dans les zones neutres des régions n et p respectivement.

2.2.3 Largeur de la ZCE W :

À partir des relations (13a) et (13b), on peut écrire le potentiel de diffusion sous la forme :

$$V(W_n) - V(-W_p) = V_d = \frac{eN_D W_n^2}{2\epsilon_{sc}} + \frac{eN_A W_p^2}{2\epsilon_{sc}} \quad (14)$$

soit, en utilisant (11),

$$W_p(V_d) = \sqrt{\frac{2\epsilon_{sc}}{e} \frac{N_D}{N_A(N_A + N_D)} V_d} \quad (15a) \quad W_n(V_d) = \sqrt{\frac{2\epsilon_{sc}}{e} \frac{N_A}{N_D(N_A + N_D)} V_d} \quad (15b)$$

soit enfin,

$$W(V_d) = \sqrt{\frac{2\epsilon_{sc}}{e} \frac{N_D + N_A}{N_A N_D} V_d} \quad (15c)$$

3 Jonction PN sous polarisation.

Supposons maintenant que l'on polarise la jonction par une tension externe V_A appliquée sur le contact p, le contact n restant au potentiel zéro (la masse). Cette tension va rompre l'équilibre entre les forces de diffusion et de conduction et un courant va pouvoir circuler dans la structure. Pour calculer ce courant on doit faire quelques approximations :

- ZCE vide de porteurs (jonction abrupte)
- Approximation de faible injection *i.e.* que les porteurs injectés sont en densité faible devant les porteurs majoritaires
- Approximation de Boltzmann
- Toute la tension V_A est appliquée aux bornes de la ZCE
- Pas de phénomènes de génération et de recombinaison dans la ZCE

Pour calculer le courant circulant dans la diode, nous allons dans un premier temps, calculer la densité de porteurs injectés aux frontières de la ZCE, puis calculer le profil de la répartition des ces porteurs du à la diffusion dans les régions neutres. On calculera alors le courant résultant.

3.1 Densité de porteurs injectés - approximation de Boltzmann.

Dans la ZCE, on sait que deux composantes (de diffusion et de conduction) de courant s'opposent parfaitement si la tension externe est nulle. Si on calcule l'ordre de grandeur de ces composantes, on trouve autour de 10^4 A/cm², ce qui pour une surface moyenne de 10^4 μm² correspond à un courant de 1A. En appliquant une tension en direct, on sait que le courant sera de quelques dizaines de mA.

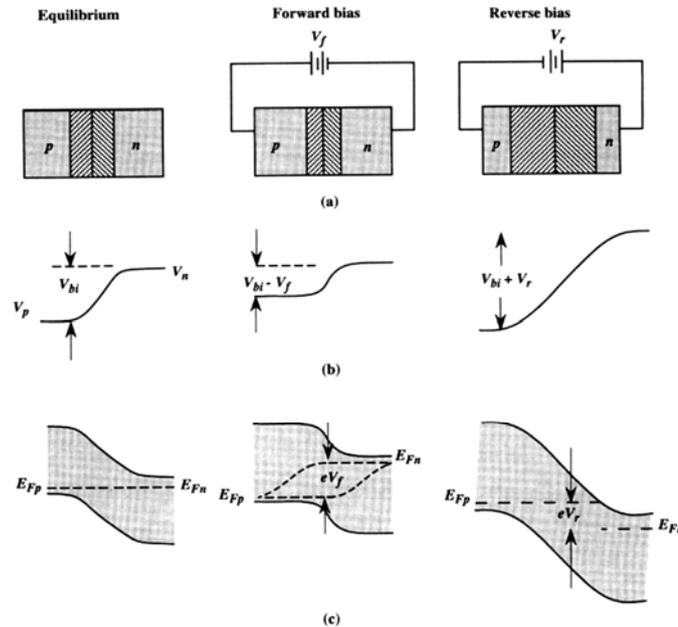


Figure 2-3: différents régimes de polarisation (Singh)

L'approximation de Boltzmann consiste à dire que la résultante des courants étant faible devant les composantes de ce courant, on considère que l'on est encore en quasi-équilibre et donc que l'équation (4) est encore valide en remplaçant V_d par $V_d - V_a$.

On écrit alors :

$$\text{Si } V_a=0 \quad \frac{p(W_N)}{p_p} = \exp\left(-\frac{eV_d}{kT}\right) \quad (16)$$

$$\text{Si } V_a \neq 0 \quad \frac{p'(W_N)}{p_p} = \exp\left(-\frac{e(V_d - V_a)}{kT}\right) \quad (17)$$

Ce qui se réécrit :

$$p'_n = p_n \exp\left(\frac{eV_A}{kT}\right) = \frac{n_i^2}{N_D} \exp\left(\frac{eV_A}{kT}\right) \quad (18a)$$

$$n'_p = n_p \exp\left(\frac{eV_A}{kT}\right) = \frac{n_i^2}{N_A} \exp\left(\frac{eV_A}{kT}\right) \quad (18b)$$

On remarque que le produit de la densité de trous (d'électrons) dans n(p) par la densité d'électrons (de trous) dans p(n) donne :

$$n'_p * p_p = p'_n * n_n = n_i^2 \exp\left(\frac{eV_a}{kT}\right) \quad (19)$$

3.2 Distributions des porteurs dans les régions neutres.

Une fois les trous injectés en W_N , ils vont diffuser et se recombiner avec les porteurs majoritaires (les électrons). En écrivant l'équation de continuité pour les trous et en se plaçant en régime continu ($V_a = cte$), on trouve :

$$p'(x) - p_n = \frac{p_n}{sh\left(\frac{d_n}{L_p}\right)} \left(e^{\frac{eV_a}{kT}} - 1\right) sh\left[\frac{(x_c - x)}{L_p}\right] \quad (20a)$$

$$n'(x) - n_p = \frac{n_p}{sh\left(\frac{d_p}{L_n}\right)} \left(e^{\frac{eV_a}{kT}} - 1\right) sh\left[\frac{(x + x'_c)}{L_n}\right] \quad (20b)$$

avec d_n et d_p les largeurs des régions neutres n et p respectivement.

Dans le cas de **régions « longues »** ($d_{n,p} \gg L_{p,n}$), les expressions (20) peuvent se simplifier :

$$p'(x) - p_n = p_n \left(e^{\frac{eV_a}{kT}} - 1\right) e^{(W_N - x)/L_p} \quad (21a)$$

$$n'(x) - n_p = n_p \left(e^{\frac{eV_a}{kT}} - 1\right) e^{(x + W_p)/L_n} \quad (21b)$$

Dans le cas de **régions « courtes »** ($d_{n,p} \ll L_{p,n}$), les expressions (20) peuvent se simplifier :

$$p'(x) - p_n = \frac{p_n}{d_n} \left(e^{\frac{eV_a}{kT}} - 1\right) (x_c - x) \quad (22a)$$

$$n'(x) - n_p = \frac{n_p}{d_p} \left(e^{\frac{eV_a}{kT}} - 1\right) (x'_c + x) \quad (22b)$$

3.3 Courants de porteurs minoritaires dans les régions neutres :

Le processus d'entraînement des porteurs dans les régions neutres est un phénomène de diffusion. En effet, en faible injection on suppose que le champ électrique dans la ZCE est beaucoup plus grand que dans les zones adjacentes que nous avons supposées neutres ($E=0$). Ce processus ne doit pas être confondu avec le processus de transfert de porteurs à travers la ZCE qui, suivant le type de polarisation, peut être un processus de diffusion ou de conduction. Le courant J_p (J_n) de diffusion des trous (électrons) dans la région n (p) est :

$$J_p(x) = -eD_p \frac{dp(x)}{dx} \quad (23)$$

$$J_n(x) = eD_n \frac{dn(x)}{dx} \quad (24)$$

En supposant que tous les dopants sont ionisés et en injectant les équations (20 à 22), on obtient l'expression des courants de trous et d'électrons injectés respectivement en $x = W_N$ et en $x = -W_P$. Dans le cas d'une diode idéale, aucune recombinaison d'électrons et de trous n'est présente à la traversée de la ZCE, de sorte que l'on peut écrire :

$$J(V) = J_p(-W_p) + J_n(W_p) = J_p(W_n) + J_n(-W_p)$$

ce qui se réécrit :

$$J(V) = J_s(e^{eV/T} - 1) \quad (25)$$

avec J_s , courant de saturation de la diode, qui s'écrit :

$$\text{Région quelconque :} \quad J_s = \frac{en_i^2 D_p}{N_D L_p \text{th}\left(\frac{d_n}{L_p}\right)} + \frac{en_i^2 D_n}{N_A L_n \text{th}\left(\frac{d_p}{L_n}\right)} \quad (26a)$$

$$\text{Région longue :} \quad J_s = \frac{en_i^2 D_p}{N_D L_p} + \frac{en_i^2 D_n}{N_A L_n} \quad (26b)$$

$$\text{Région courte :} \quad J_s = \frac{en_i^2 D_p}{N_D d_n} + \frac{en_i^2 D_n}{N_A d_p} \quad (26c)$$

Si on s'intéresse à l'expression de $J_p(x)$ dans le cas d'une région longue par exemple, cette expression s'écrit :

$$J_p(x) = e \frac{D_p}{L_p} p_n e^{-(x-W_N)/L_p} (e^{eV/kT} - 1) \quad \text{pour } x > W_N \quad (27)$$

Cette expression nous montre que le courant de minoritaires décroît très vite avec x . C'est la recombinaison de ces porteurs minoritaires avec les porteurs majoritaires (ici les électrons dans la région n) qui est responsable de cette décroissance. Pour maintenir la neutralité électrique (ici de la région n), un nombre égal d'électrons est fourni par le contact électrique de la région n et ceci en un temps de l'ordre de 10^{-12} seconde, qui est le **temps de relaxation diélectrique** τ_d . On peut alors considérer que la région est toujours neutre, ceci en comparant ce temps avec la durée de vie des porteurs minoritaires (ou le temps de transit suivant la géométrie de la jonction) qui est toujours très supérieur à τ_d . Le régime stationnaire est le résultat de la polarisation qui maintient l'injection des minoritaires. Le même processus est équivalent dans la région p. Si on veut calculer la partie électronique du courant dans la région n (courant de porteurs majoritaires), il suffit à partir de (26b) et (27) d'écrire :

$$J_n(x) = J - J_p(x) \quad \text{pour } x > W_N \quad (28)$$

soit encore :

$$J_n(x) = e \left[\frac{D_p}{L_p} (1 - e^{-(x-W_N)/L_p}) p_n + \frac{D_n}{L_n} n_p \right] (e^{eV/kT} - 1) \quad \text{pour } x > W_N \quad (29)$$

Attention !! Cette expression n'est valable que dans la région n, c'est un courant de majoritaires, d'électrons dans la région n. Elle est naturellement différente de l'équation (24).

Les expressions pour des régions courtes ou quelconques sont obtenues de la même manière. La figure xxx montre la situation dans les 2 régions, en fonction de la géométrie de la jonction.

A partir de ces différentes expressions, on voit que le courant qui circule dans une diode est loin de suivre la loi d'Ohm. Le comportement est **non linéaire** et **redresseur**. Le courant sature à la valeur J_s lorsqu'une tension négative est appliquée sur le côté P (polarisation inverse), et comme J_s est petit, la diode est dite bloquée ou non-conductrice. En revanche, si la polarisation est directe, le courant varie exponentiellement avec la tension et la diode est très conductrice.

4 La diode réelle : phénomènes de génération et recombinaison dans la ZCE.

Dans l'étude qui précède, on a considéré que la diode, le semi-conducteur était parfait et sans défauts (impuretés, défauts cristallins...). En d'autres termes, on a supposé dans la diode idéale, que les trous et les électrons injectés ne pouvaient se recombiner dans la ZCE, et par conséquent, le courant était constant à travers cette zone. En fait cette approximation est loin d'être justifiée dans une diode réelle, et l'on doit tenir compte des effets de génération thermique et de recombinaison à la traversée de la ZCE. En fait lorsque les électrons et les trous sont injectés dans la ZCE, une façon de franchir la région sans franchir la barrière de potentiel $V_d - V_a$ est de se recombiner entre eux. Ceci induit un flux de courant supplémentaire (courant de génération recombinaison) que l'on doit ajouter au courant théorique de la diode idéale.

On sait que l'expression du taux de recombinaison, donnée par l'équation de Shockley-Read (voir chapitre précédent), s'écrit $r = \frac{1}{\tau} \frac{pn - n_i^2}{2n_i + p + n}$. On sait également que le produit de la densité de trous par la densité d'électrons par exemple en $x = W_N$ s'écrit sous polarisation V_a :

$$p(W_N)n(W_N) = p(W_P)n(W_P) = n_i^2 \exp\left(\frac{eV_a}{kT}\right) \quad (30)$$

Dans la mesure (on suppose) où le produit np est constant dans la ZCE et que ce produit est très supérieur à n_i^2 , le taux de recombinaisons étant maximum pour $n = p$ se réécrit :

$$r_{\max} = \frac{n_i}{\tau} \exp\left(\frac{eV_a}{2kT}\right) \quad (31)$$

On peut montrer que ce taux est maximum en centre de ZCE. En première approximation et en polarisation directe on peut supposer que ce taux est constant dans toute la ZCE de largeur W_T . Par définition le courant de génération recombinaison s'écrit :

$$J_{GR} = e \int_{W_p}^{W_N} r dx \quad (32)$$

En polarisation inverse, le taux r s'écrit ($pn \ll n_i^2$) $r = -\frac{n_i}{2\tau} < 0$ et le courant est un courant de génération J_{Ge} qui s'écrit à partir de (32) :

$$J_{Ge} = -\frac{en_i}{2\tau} W_T \quad (33)$$

En polarisation directe $r \approx cte = r_{max} > 0$, et le courant est un courant de recombinaison qui s'écrit :

$$J_{Re} = \frac{en_i}{2\tau} W_T \exp\left(\frac{eV_a}{2kT}\right) \quad (34)$$

Ce qui se résume en un courant de génération recombinaison J_{GR} qui s'écrit :

$$J_{GR} = J_{GR}^0 \left(\exp\left(\frac{eV_a}{2kT}\right) - 1 \right) \quad (35) \quad \text{avec} \quad J_{GR}^0 = \frac{en_i}{2\tau} W_T \quad (36)$$

Cette expression revient à dire que sous polarisation nulle, le courant de génération (effet thermique) compense exactement le courant de recombinaison et $J_{GR} = J_{Ge} + J_{Re} = 0$. Le courant total du dispositif s'écrit maintenant :

$$J(V_a) = J_S \left(\exp\left(\frac{eV_a}{kT}\right) - 1 \right) + J_{GR}^0 \left(\exp\left(\frac{eV_a}{2kT}\right) - 1 \right) \quad (37)$$

En général $J_{GR}^0 \gg J_S$ (sauf pour les semi-conducteurs à petites bande interdite à cause de la valeur de n_i), de telle sorte que sous faible polarisation c'est le terme en $\exp\left(\frac{eV_a}{2kT}\right)$ qui est dominant, alors que pour les fortes polarisations, l'exponentielle prend le dessus et le comportement théorique de la diode est prépondérant. Si l'on trace la caractéristique $J(V)$ en coordonnées semi-logarithmique on observera 2 régimes de pente $\frac{1}{2}$ et 1. On peut alors introduire le **facteur d'idéalité** de la diode n et écrire (37) sous la forme :

$$J(V) = J_0 \left(\exp\left(\frac{eV_a}{nkT}\right) - 1 \right) \quad (38)$$

Dans le cas d'une diode de très bonne qualité $n=1$, sinon n prendra une valeur intermédiaire entre 1 et 2.

5 Diodes sous forts niveaux de polarisations.

On va dans cette partie étudier l'effet des niveaux de polarisation élevés, aussi bien en direct qu'en inverse. Cette partie distincte est nécessaire du fait que précédemment on se plaçait en régime de faible injection et les hypothèses consécutives ne peuvent plus s'appliquer maintenant.

5.1 Régime de forts niveaux de polarisation en direct. La différence essentielle est au niveau de la tension appliquée aux bornes de la ZCE V_j . Compte tenu que la polarisation est forte, l'approximation selon laquelle toute la tension appliquée V_a se retrouve aux bornes de la ZCE ne tient plus. En forte injection la densité de porteurs minoritaires devient égale voire supérieure (en très forte injection) à la densité de porteurs majoritaires, la largeur de la ZCE se réduit pour tendre vers zéro et les chutes ohmiques dans les régions neutres augmentent. Le processus de déplacement des porteurs minoritaires n'est plus exclusivement de diffusion mais également de conduction du fait de la présence d'un champ électrique dans les régions non déplétées de la jonction. On peut montrer (Mathieu p 172) que le courant d'une P⁺N se met sous la forme :

$$J(V) = J_s^* \left(\exp\left(\frac{eV_a}{2kT}\right) - 1 \right) \quad (39)$$

$$\text{avec } J_s^* = \frac{en_i D_p^*}{L_p^* \operatorname{th}\left(\frac{d_n}{L_p^*}\right)}, \quad D_p^* = D_p (1 + p/n)_{W_N} \quad \text{et } L_p^* = \sqrt{D_p \tau_p^*}$$

τ_p^* est la durée de vie des trous en régime de forte injection.

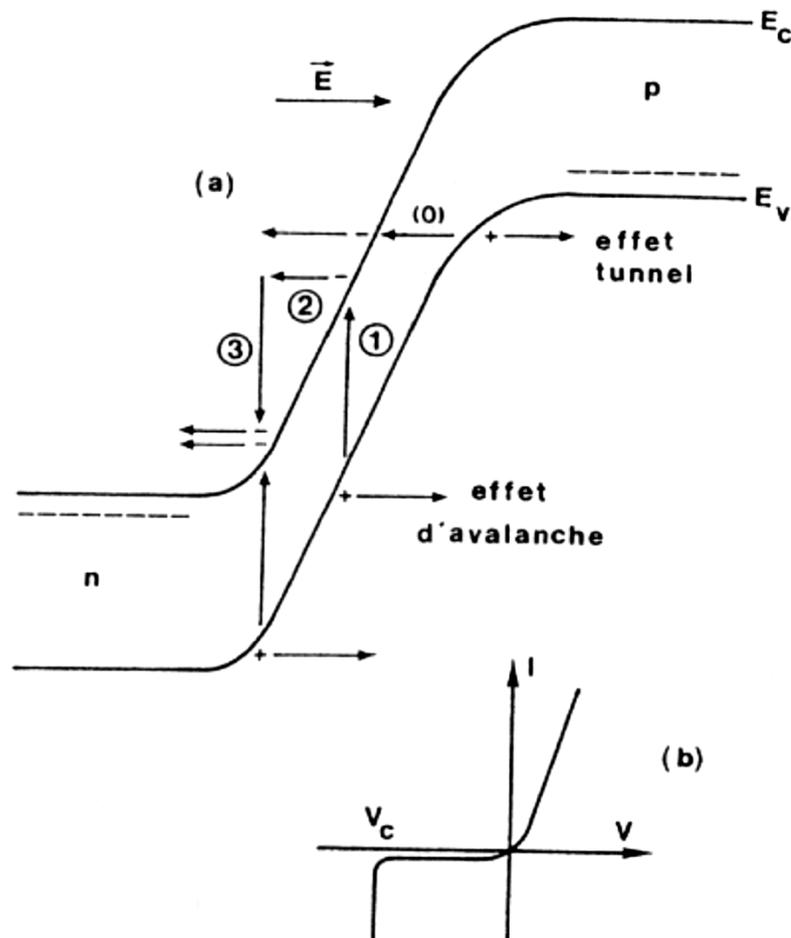


Figure 2-4:claquage de la jonction par avalanche ou tunnel (Mathieu)

Si la tension augmente encore la zone de charge d'espace s'annule et le courant suit la loi d'Ohm, et il est contrôlé par la résistance série de la diode R_s qui est la contribution des régions N et P et des résistances de contacts.

5.2 Polarisation inverse : claquage de la jonction. Lorsque la tension inverse augmente, on peut observer une brusque augmentation du courant du à un ou des phénomènes de claquage de la jonction («break-down»). Quatre phénomènes sont responsables de ce claquage, un thermique, deux par effet d'avalanche et un par perçage de la jonction.

5.2.1 Effet thermique. Ce mode est surtout rencontré dans les semi-conducteurs à petit gap. Si on se rappelle que le courant J_s est fonction de n_i , ce courant sera donc sensible à la température. Le courant inverse traversant la ZCE peut par effet Joule augmenter la température de la jonction et par la même augmenter n_i , qui lui-même augmente J_s qui par cascade augmente T et ainsi de suite jusqu'à obtenir le claquage thermique de la jonction.

5.2.2 Effet Zener. En inverse, le champ électrique dans la ZCE, augmente comme $\sqrt{(V_d - V_a)}$ ($V_a < 0$). Lorsque ce champ électrique est supérieur au champ électrique critique E_C , la force électrique que subit l'électron dans la bande de valence est suffisante pour l'arracher de sa valence et le rendre « libre » et l'amener dans la bande de conduction. Ce processus consiste à un passage par effet tunnel de l'électron de la bande de valence à la bande de conduction, laissant dans la bande de valence un trou. On a donc création d'une paire électron trou qui est évacuée par le champ de la ZCE. Cet effet sera d'autant plus présent que la largeur de la ZCE est faible (fort dopage) et la tension inverse grande (fort champ dans la ZCE). Une fois l'effet Zener présent pour une tension critique V_Z , la tension aux bornes de la ZCE n'augmente plus et est fixée à V_Z .

5.2.3 Effet d'avalanche : ionisation par impact. Lorsque la largeur de la ZCE est tel que le champ critique ne peut être atteint, un autre phénomène est présent. Sur la largeur de la ZCE, les porteurs peuvent être accélérés et acquérir une énergie suffisante pour, par impact exciter un électron de la bande de valence vers la bande de conduction. On est « parti » d'un électron qui transportait le courant, et après le choc nous avons 2 électrons et un trou. Ceci peut se reproduire plusieurs fois pendant la traversée de la ZCE. Le courant initial (transporté par l'électron) I_0 devient alors $I_0' = M(V_a)I_0$ (40) où M est le facteur de multiplication. On montre (Singh p226) que la tension de claquage V_{BD} s'écrit dans le cas d'une jonction dissymétrique :

$$V_{BD} = \frac{\epsilon E_C^2}{2eN_B} \quad (41)$$

5.2.4 Claquage par perçage. Dans le cas de diode courte, la tension inverse peut amener la ZCE à dépeupler totalement la région neutre. Prenons le cas d'une diode N⁺P ; la ZCE s'étend essentiellement coté P. Si la tension est telle que toute la région P est dépeuplée, le contact électrique va subir l'influence du champ électrique et fournir en

électrons la diode. La diode devient un court-circuit et le courant est limité par la résistance de protection R_p et la résistance série de la diode R_s .

5 Jonction en régime dynamique : capacités de la jonction.

La notion de capacité est toujours liée à la présence de régions chargées, donc de charges. Dans le cas de la jonction PN, on sait que les charges peuvent être de deux types : des charges fixes (les dopants ionisés) et des charges mobiles (les électrons et les trous). On va voir que la présence de ces 2 types de charges induit la présence de 2 types de capacités : une de **jonction** ou de **transition** liée aux charges fixes dans la ZCE et une capacité de **diffusion** ou de **stockage** liée à l'injection de porteurs minoritaires dans les régions neutres.

- **Capacité de transition ou de la jonction.** Par définition, une capacité est définie par la variation de la charge en fonction de la tension appliquée :

$$C_T = \left| \frac{dQ}{dV} \right| \quad (42)$$

La charge dans la zone de déplétion est donnée par :

$$|Q| = eAN_A W_P = eAN_D W_N \quad (43)$$

En injectant la relation (15) dans (43) et en différenciant, on obtient :

$$C_T = \frac{A}{2} \sqrt{\frac{2e\varepsilon}{(V_D - V_A)} \frac{N_A N_D}{(N_A + N_D)}} = \frac{A\varepsilon}{W_T} \quad (44)$$

L'équation (44) montre que la capacité de transition est identique à la capacité d'un condensateur plan dont l'épaisseur de l'isolant est W_T et la permittivité ε . Une des applications de la variation de la capacité en fonction de la tension appliquée est la réalisation de varicap ou varactor, très utilisés dans des circuits résonnants ou de réglage de fréquences.

- **Capacité de diffusion ou de stockage.** On vient de voir que cette capacité était due à l'injection de porteurs, elle sera donc dominante dans le cas de jonction polarisée en direct. En fait cette capacité va traduire le comportement capacitif de la jonction, c'est à dire par exemple le déphasage entre la tension et le courant en régime alternatif. Si on veut calculer la charge stockée par exemple dans la région neutre N, cette charge est due à des trous venant de P. La charge Q_{Sp} est proportionnelle au courant de trous J_p par le facteur de proportionnalité qui doit être un temps que l'on appelle τ . Par définition cette charge stockée peut s'écrire :

$$Q_{Sp} = A \int_{W_N}^{x_c} e(p'(x) - p_n) dx \quad (45)$$

où $p'(x) - p_n$ représente la densité de trous en excès dans la région neutre donnée par l'expression (20a). L'intégration de (45) conduit à :

$$Q_{Sp} = e(p'(0) - p_n)L_p \left[\coth\left(\frac{d_n}{L_p}\right) - \frac{1}{\operatorname{sh}\left(\frac{d_n}{L_p}\right)} \right] \quad (46)$$

Ce qui peut se mettre sous la forme $Q_{Sp} = \tau J_p(W_N)$ (47) en posant :

$$\tau = \tau_p \left(1 - \frac{1}{\operatorname{ch}\left(\frac{d_n}{L_p}\right)} \right) \quad (48)$$

NOTA : en fonction de la géométrie de la jonction τ est la durée de vie des trous si la région est « longue », est le temps de transit

$\tau_i = \frac{d_n^2}{2D_p}$ si la région est « courte » et un temps intermédiaire donné par (48) si la région est « quelconque ».

Cette étude dans la région N est valable dans la région P en changeant les indices. De ce fait la charge stockée dans les 2 régions N et P est donnée par :

$$Q_S = Q_{Sn} + Q_{Sp} = \tau_{(n)} J_n (-W_P) + \tau_{(p)} J_p (W_N) \quad (49)$$

En injectant cette relation dans l'expression $C_S = \frac{dQ_S}{dV}$, on obtient :

$$C_S = C_{Sn} + C_{Sp} = \frac{e}{kT} (\tau_{(n)} J_n + \tau_{(p)} J_p) \quad (50)$$

6 Jonction PN en commutation. Grands signaux.

Le fait de passer des petits signaux aux grands signaux va faire apparaître des phénomènes nouveaux sur le comportement en commutation de la jonction PN. Pour cela on va étudier le comportement de celle ci lorsque l'on va alternativement la faire passer de l'état passant à l'état bloqué. Pour simplifier l'étude on va se placer dans le cas d'une diode P⁺N, de sorte que le courant est un courant essentiellement de trous et donc les charges stockées des trous.

Supposons la jonction dans l'état passant, donc polarisée par une tension directe égale à V_A largement supérieure à la tension de mise en conduction de la diode. Dans ce cas, le courant est limité par la résistance du circuit R_p et vaut donc $I_1 = V_A/R_p$. Il peut donc être important. On doit noter également que la charge stockée dans la région N est grande. La tension aux bornes de la jonction (et non de la diode) est faible et de l'ordre de la tension de diffusion V_D (en fait

$V_D - kT/e$). Si maintenant on inverse brutalement la tension, le courant avant de se stabiliser à sa valeur I_S , commence par prendre une valeur importante $I_2 = \frac{V_2}{R_p}$ négative, cette valeur restant constante pendant un temps appelé temps de stockage, avant de commencer à diminuer (en valeur absolue) pour atteindre sa valeur à l'équilibre I_S .

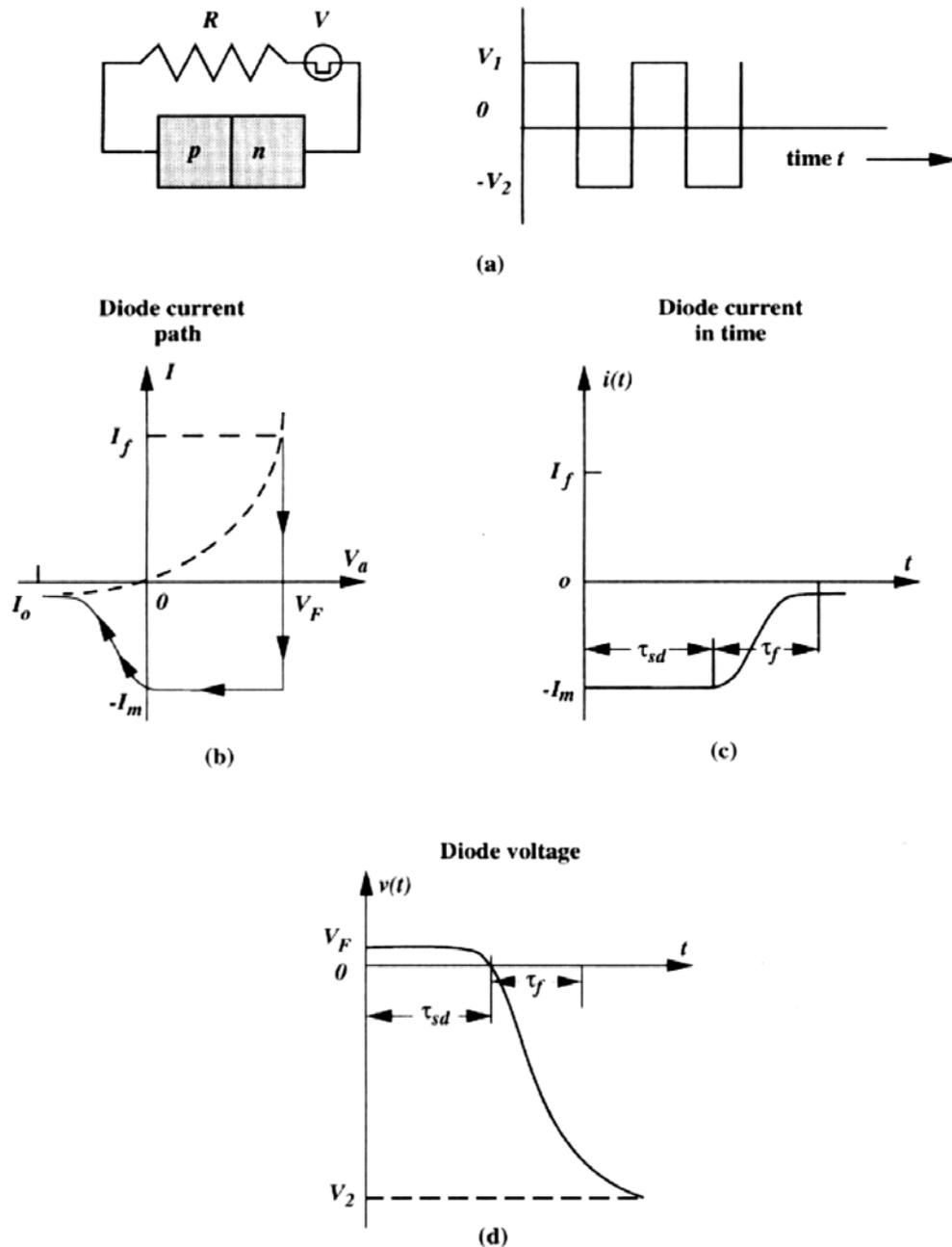


Figure 2.5 : principe physique de la diode en commutation (Singh et Ng)

Si on regarde l'expression de l'excédent de densité de trous injectés dans la région n, $p'(x_n) - p_n$, elle est donnée par (supposons une diode longue) (21a). On se

rappelle que dans cette expression, V_a représente la tension aux bornes de la jonction (faible injection). Tant que cette expression est positive, on peut considérer que la jonction est polarisée en direct et donc que la tension à ses bornes est $V_j \cong V_D$. Or il est évident que cette charge stockée ne peut s'évacuer instantanément. Pourquoi ? En fait l'évacuation est limitée par la recombinaison dans la région N et la diffusion, phénomènes qui ne sont pas immédiats. Cette tension $V_j(t)$ va donc évoluer de V_D à 0 pendant un temps τ_{sd} qui correspond au temps nécessaire pour évacuer les charge c'est à dire « faire » $p'(x_n) = p_n$. A partir de cet instant, la ZCE va commencer à s'étendre, la tension à ses bornes augmenter, sa résistance augmenter et donc le courant I diminuer pour atteindre I_S . On montre que le temps de recouvrement τ_r qui est le temps $\tau_{sd} + \tau_f$ peut se mettre sous la forme :

$$\tau_r = \frac{I_1}{I_2} \tau = \frac{J_d}{J_i} \tau$$

avec τ soit le temps de transit (diode courte), soit la durée de vie (diode longue) soit un temps intermédiaire (diode qcq). L'expression du temps de stockage est donnée par Tyagi:

$$\tau_{sd} = \tau_p \left[\ln\left(1 + \frac{J_d}{J_i}\right) - \ln\left(1 + \frac{J_d}{J_d + J_i}\right) \right] \quad (51)$$

$$\tau_f \cong 2.3 \left[\frac{\tau_F \delta + R_p C_T}{1 + \delta} \right] \quad (52)$$

avec $\delta = \frac{J_d}{J_d + J_i}$ et C_T la capacité moyenne entre $V=0$ et $V=-V_2$, et τ_F la durée de vie de la charge (\approx la durée de vie des porteurs).

On peut réduire ce temps en utilisant une diode courte (τ_p est remplacé par le temps de transit), mais le problème de **perçage** intervient. Pour des commutations rapides on préfère «tuer» la durée de vie en incorporant des centres de recombinaisons (Au par exemple dans le Si), ou en utilisant des semi-conducteurs à «gap» indirect.

7 Diode tunnel – Diode Backward.

Une diode tunnel est similaire à une diode PN. La seule différence est le niveau de dopage dans les deux régions qui est très élevé de façon à « dégénérer » le semi-conducteur. Une influence directe est la faible extension de la ZCE qui est de l'ordre de quelques nanomètres. Cette faible largeur de ZCE induit une grande capacité de transition de la diode tunnel. Si on trace la caractéristique $I(V)$ de la diode on obtient la courbe de la figure 6.6 qui est une courbe en forme de N.

Différence avec une jonction PN classique :

- En inverse la diode conduit.
- En direct la diode présente une partie à résistance dynamique négative.

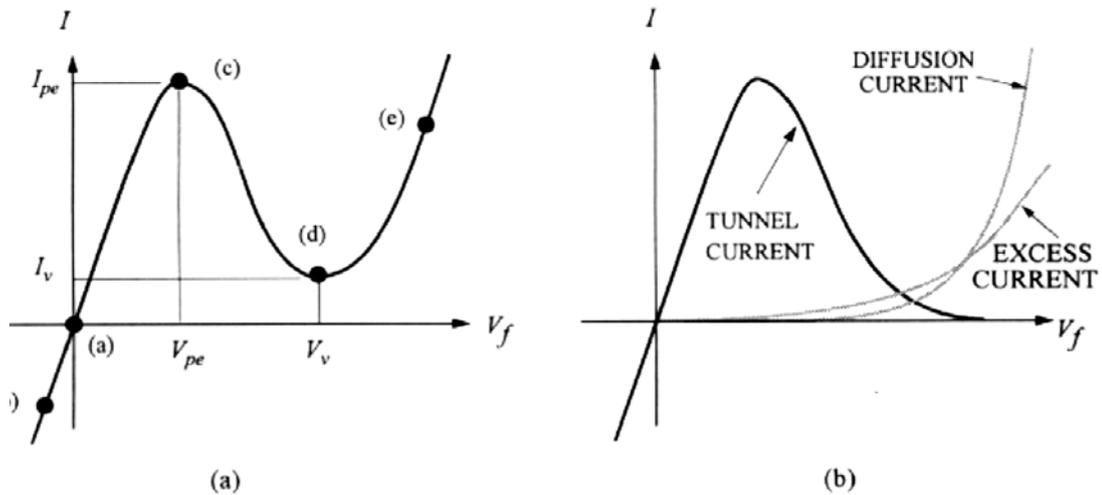


Figure 2.6 : caractéristique $I(V)$ d'une diode tunnel (Ng)

Cette caractéristique (résistance négative) est due à la présence d'une composante nouvelle de courant qui est le courant tunnel. Ce courant implique un transfert d'électrons entre la BV et la BC. La probabilité d'effet tunnel à travers une barrière de potentiel triangulaire est donnée par :

$$T_t \approx \exp\left(-\frac{4a\sqrt{2m^*e\phi_b}}{3\hbar}\right) \quad (53)$$

Dans notre cas a est la largeur de la ZCE et $e\phi_b$ la largeur de la bande interdite (voir figure). La figure 2.8 représente les différentes conditions sur les bandes énergétiques pour expliquer la forme de la caractéristique $I(V)$ de la diode tunnel. Le courant tunnel peut être évalué empiriquement par Demassa:

$$I_t = I_{pe} \left(\frac{V_a}{V_{pe}}\right) \exp\left(1 - \frac{V_a}{V_{pe}}\right) \quad (54)$$

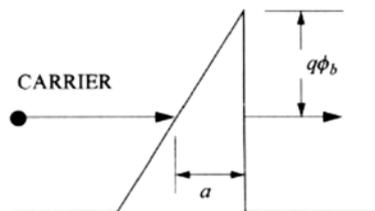


Figure 2.7 : principe de l'effet tunnel

En pratique le courant de vallée n'est pas nul, même à $T=0K$, à cause d'un courant en excès dû à la présence d'états dans le « gap » (voir figure 2.7). Un facteur de qualité d'une diode tunnel est le rapport I_{pe}/I_v .

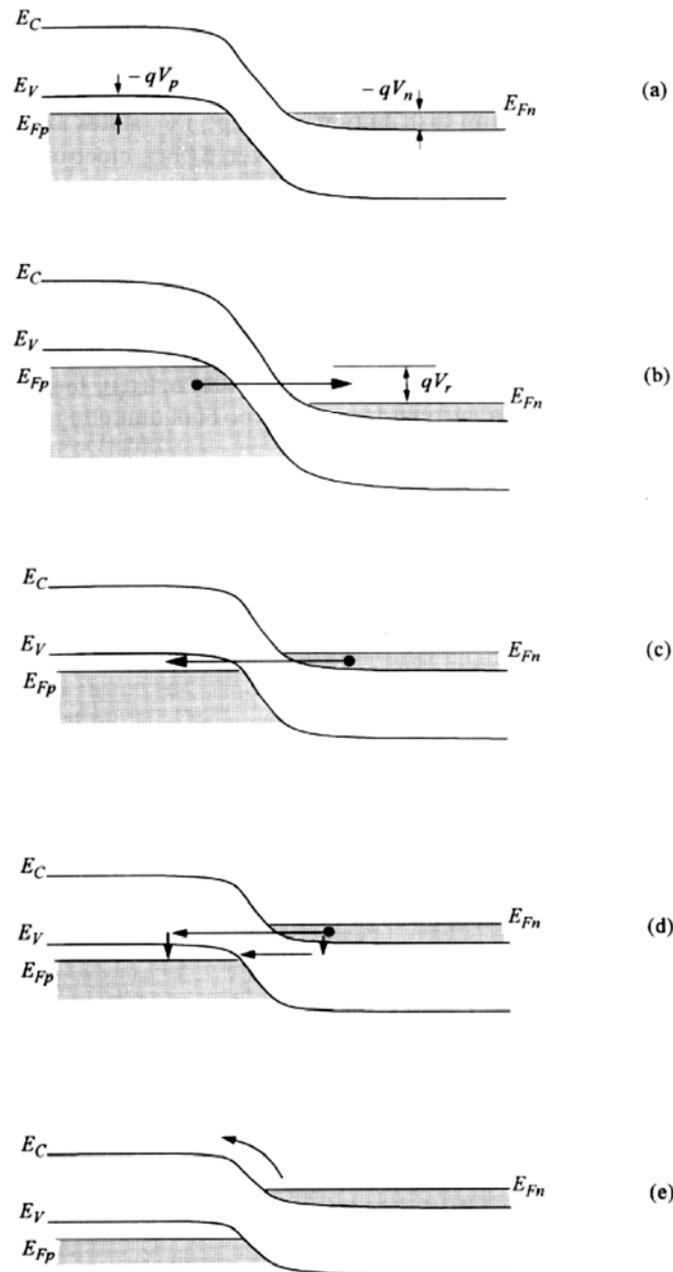


Figure 2.8 :diagramme d'énergie dans les différents régime d'une diode tunnel

Les applications sont nombreuses en modulation micro-onde, détection, oscillateur...

La diode **Backward** (pour diode en inverse) est une diode dont le matériau est juste dégénéré de façon à réduire le courant tunnel (le rendre nul) en direct et le

maintenir en inverse. La conduction est donc assurée par des porteurs majoritaires, cette diode est donc plus rapide en vitesse de commutation.

8 Références.

1. **H. Mathieu**, « physique des semi-conducteurs et composants électroniques », 4^e édition , Masson (1998)
2. **KK. Ng**, « complete guide to semiconductor devices », McGraw-Hill editors, (1995)
3. **J.Singh**, « semiconductors devices, an introduction », McGraw-Hill editors, (1994)
4. **M.S. Tyagi**, « introduction to semiconductor material and devices », John Wiley and Sons, (1991)
5. **T.A. Demassa**, « the prediction of tunnel diode voltage-current characteristics », solid-state electron.,**13**, p131, (1970)
6. **D. J. Roulston**, « Bipolar semiconductor devices », McGraw-Hill, 1990.

Chapitre 3 : Contact Métal – Semi-conducteur Diode Schottky

1. Introduction. L'utilisation du métal dans les dispositifs électroniques répond à trois nécessités ; l'interconnexions entre les circuits et composants, la réalisation de contacts ohmiques pour injecter le courant dans les composants et enfin la réalisation de composants redresseurs ou diodes Schottky.

2. La diode à barrière Schottky. Dans certaines conditions, le contact métal – semi-conducteur peut avoir une réponse non linéaire, ce qui permet de réaliser des diodes Schottky. Avant d'aller plus loin, on doit introduire un certain nombre de paramètres qui nous permettront de comprendre le fonctionnement de ce composant.

- **Travail de sortie – Affinité électronique.** Le **travail de sortie** est l'énergie qu'il faut fournir à un électron dans le métal pour l'extraire du métal. On l'appellera $e\Phi_M$ et son unité sera l'électronvolt. Il est défini comme la différence entre le niveau de vide et le niveau de Fermi dans le métal. Ce travail de sortie est constant pour un métal donné. Dans le cas du semi-conducteur, on peut définir également un travail de sortie $e\Phi_{SC}$ qui est la différence d'énergie entre le niveau de vide et le niveau de Fermi du semi-conducteur. Mais dans ce cas, $e\Phi_{SC}$ est fonction du dopage et donc varie. On préfère définir alors l'affinité électronique $e\chi_{SC}$ qui est la différence d'énergie entre le niveau de vide et la bande de conduction BC. Ce terme est constant pour un semi-conducteur donné, quel que soit le dopage de celui-ci.

Physiquement, plus le travail de sortie est grand et plus l'électron est « lié » à son milieu. En d'autres termes, si on met en contact deux matériaux de travail de sortie différents, les électrons à l'interface auront tendance à être transférés du matériau à faible travail de sortie vers le matériau à grand travail de sortie.

- **Mise en contact Métal - semi-conducteur.** Prenons par exemple un métal et un semi-conducteur de type n et tel que $e\Phi_M > e\Phi_{SC}$ et mettons-les en contact. Le système n'étant pas polarisé, les niveaux de Fermi s'alignent dans la structure. A la jonction les niveaux de vide doivent être les mêmes, ce qui conduit à un courbure de bande comme montré sur la **figure 1**. Cette courbure correspond à un transfert d'électrons du semi-conducteur vers le métal, ce qui est normal compte tenu de $e\Phi_M > e\Phi_{SC}$. Compte tenu de la très forte densité d'électrons dans le métal ($\approx 10^{23} \text{ cm}^{-3}$), les électrons transférés ne modifient pas le profil de bande et le niveau de Fermi dans le métal, et ceux ci restent plats. Le départ des électrons du semi-conducteur crée une zone de charge d'espace à l'interface dans le semi-conducteur. La hauteur de barrière de Schottky (dans le cas idéal) est donnée par :

$$e\Phi_b = e\Phi_M - e\chi_{SC} \quad (1)$$

C'est la barrière que doivent franchir les électrons pour passer du métal vers le semi-conducteur. Les électrons du semi-conducteur « voient » une barrière pour passer dans le métal égale à :

$$eV_{bi} = eV_d = e\Phi_M - e\Phi_{SC} = e\Phi_{MS} \quad (2)$$

Dans ce qui précède, on a étudié le cas idéal d'un contact M/SC. D'après ce que l'on vient de voir, la hauteur de barrière Schottky dépend des propriétés du métal et du semi-conducteur. En fait, expérimentalement, on s'est aperçu que cette hauteur était pratiquement indépendante du métal employé. L'explication provient de la présence de défauts à l'interface. Deux types principaux de défauts sont présents à l'interface : des

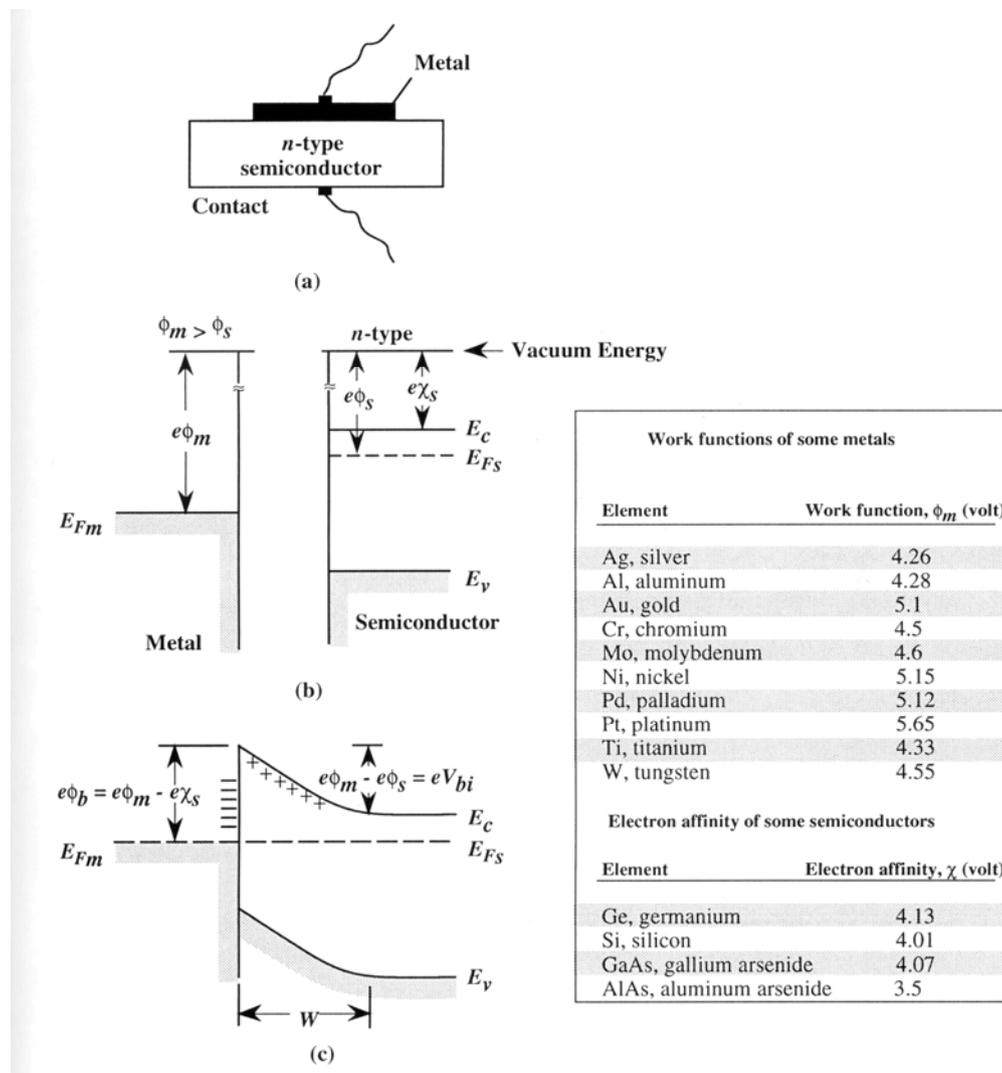


Figure 1 : diagramme énergétique d'un contact Métal / SC (Singh)

états de surfaces rapides (intrinsèques) et des états de surfaces lents (extrinsèques). Ces défauts conduisent à une distribution dans la bande interdite des niveaux énergétiques

à l'interface (figure 2). On peut caractériser cette distribution par un niveau neutre Φ_0 ou E_0 tel que les états sous ce niveau sont neutres si remplies et les niveaux au-dessus neutres si vides. Si la densité de niveaux est élevée, on peut montrer que le niveau de Fermi se bloque (« pinning of Fermi level ») sur le niveau Φ_0 ou E_0 et la barrière devient :

$$e\Phi_b = E_g - e\Phi_0 \quad (3)$$

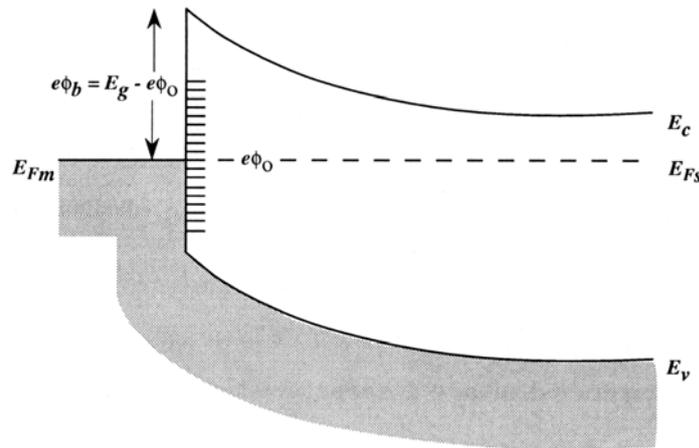


Figure 2 : Niveau neutre E_0 ou ϕ_0 (Singh)

3. Caractéristiques capacité – Tension $C(V)$. La hauteur de barrière connue, il est

facile de déterminer le profil du champ électrique, la largeur de la zone de déplétion, la capacité de transition ... Ceci se calcule de la même façon que la jonction PN, ou plus précisément d'une jonction très dissymétrique (P^+N ou N^+P). on part de l'équation de Poisson :

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho(x)}{\epsilon_{SC}} \quad (4)$$

On en déduit le champ électrique, le potentiel et la largeur de la ZCE :

$$E(x) = -\frac{eN_d}{\epsilon_{SC}}(x - W) \quad (5a)$$

$$V(x) = -\frac{eN_d}{\epsilon_{SC}}\left(\frac{x^2}{2} - W.x\right) \quad (5b)$$

$$W = \sqrt{\frac{2\epsilon_{SC}(V_{bi} - V)}{eN_d}} \quad (5c)$$

avec V_{bi} donnée par (2). L'origine des potentiels est pris à l'interface, soit encore $V(0) = 0$. La charge dans la région déplétée est donnée par :

$$Q = eN_d W = \sqrt{2e\epsilon_{sc}N_d(V_{bi} - V)} \quad (6)$$

La capacité correspondante pour un dispositif de surface A est donnée par :

$$C = A \left| \frac{dQ}{dV} \right| = A \left[\frac{e\epsilon_{sc}N_d}{2(V_{bi} - V)} \right]^{\frac{1}{2}} = \frac{\epsilon_{sc}A}{W} \quad (7)$$

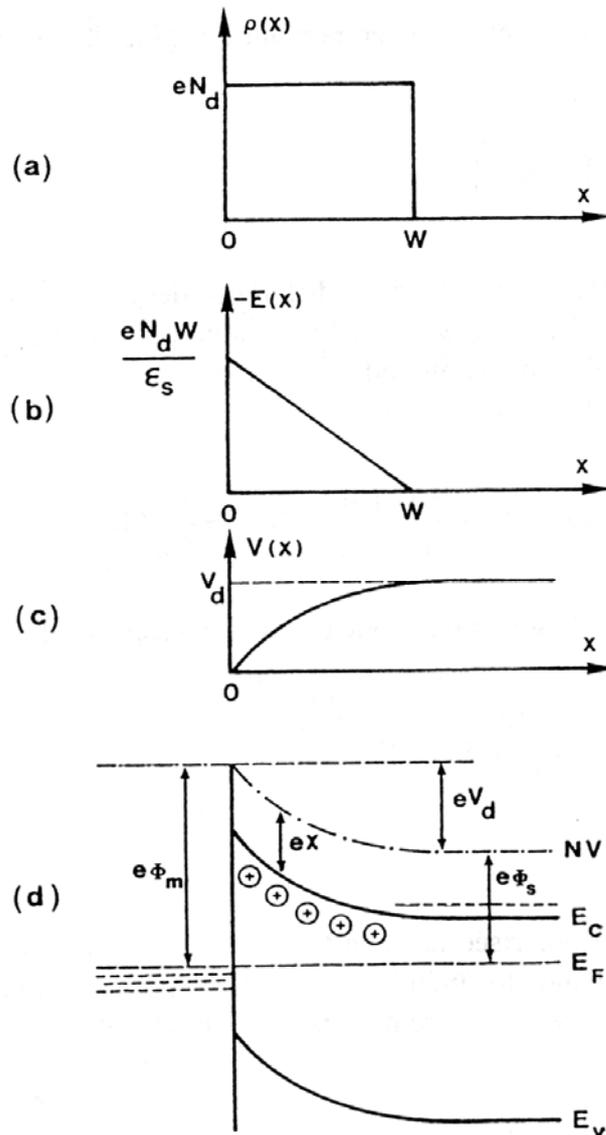


Figure 3 : densité, champ et potentiel d'une diode Schottky (Mathieu)

et donc :

$$\frac{1}{C^2} = \frac{2}{A^2 e \epsilon_{sc} N_d} (V_{bi} - V) \quad (8)$$

On retrouve les résultats obtenus sur une jonction P⁺N dans ce cas.

4. Courant dans une diode Schottky. Le courant dans une diode Schottky peut impliquer de nombreux mécanismes différents. Le plus important est l'émission thermoionique dans lequel les électrons avec une énergie plus grande que la barrière $e(V_{bi} - V)$, la franchissent pour se retrouver dans le métal. Un autre phénomène est l'émission tunnel dans la barrière. Ce processus de conduction est d'autant plus présent que le semi-conducteur est fortement dopé et donc que la ZCE est étroite. **Une différence fondamentale entre la diode PN et la diode Schottky est qu'ici, le courant en direct est un courant de porteurs majoritaires.**

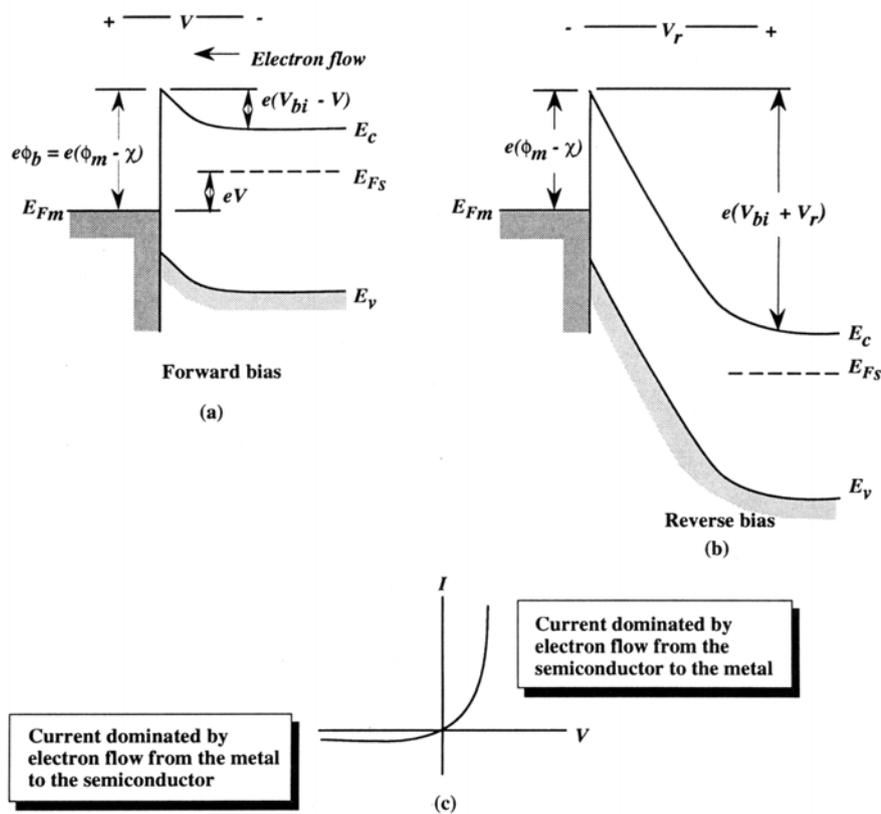


Figure 4 : Polarisation directe (a) et inverse (b) (Singh)

Les électrons qui franchissent la barrière $e(V_{bi} - V)$ sont ceux qui ont une énergie supérieure à la barrière et leur densité s'écrit :

$$n_b = n_0 \exp\left[-\frac{e(V_{bi} - V)}{kT}\right] \quad (9)$$

avec n_0 , la densité d'électrons dans la région neutre soit :

$$n_0 = N_c \exp\left[-\frac{E_C - E_F}{kT}\right] \quad (10)$$

L'équation (9) se réécrit alors :

$$n_b = N_c \exp\left[-\frac{E_C - E_F + V_{bi} - V}{kT}\right] = N_c \exp\left[-\frac{e(\Phi_b - V)}{kT}\right] \quad (11)$$

On peut montrer (Singh) que le flux d'électrons franchissant la barrière de potentiel est $\frac{\langle v \rangle n_b}{4}$ où $\langle v \rangle$ est la vitesse moyenne des électrons. Le courant d'électrons du semi-conducteur vers le métal est alors simplement donné par :

$$I_{SM}(V) = \frac{e \langle v \rangle A}{4} N_c \exp\left[-\frac{e(\Phi_b - V)}{kT}\right] \quad (12)$$

Si la tension de polarisation est nulle, le courant est nul, ce qui revient à dire que le courant résultant d'un déplacement d'électrons du métal vers le semi-conducteur compense exactement $I_{SM}(0)$, soit encore :

$$I_{MS} = -I_{SM}(0) = -\frac{e \langle v \rangle A}{4} N_c \exp\left[-\frac{e(\Phi_b)}{kT}\right] \quad (13)$$

Si on polarise le système, la barrière vue par les électrons dans le métal est la même et donc $I_{MS} = cte = I_S$. Le courant est alors donné par :

$$I = I_{SM} - I_{MS} = I_S \left(\exp\left[\frac{eV}{kT}\right] - 1 \right) \quad (14)$$

Si on suppose que les électrons suivent la statistique de Maxwell-Boltzmann, la vitesse moyenne s'écrit :

$$\langle v \rangle = \left(\frac{8kT}{\pi m^*} \right)^{\frac{1}{2}} \quad (15)$$

En remplaçant N_c par son expression, on obtient enfin l'expression du courant thermoionique dans une diode Schottky :

$$I = A \left(\frac{m^* e k^2}{2\pi^2 \hbar^3} \right) T^2 \exp\left(-\frac{e\Phi_b}{kT}\right) \left[\exp\left[\frac{eV}{kT}\right] - 1 \right] \quad (16)$$

Le premier terme, dans la parenthèse, est appelé constante de Richardson \mathfrak{R} .

A ce courant s'ajoute si la jonction (le semi-conducteur) est très dopée, un courant par effet tunnel :

$$I_{tunnel} = AJ_0^t \exp\left(\frac{eV}{E_0}\right) \quad \text{avec } E_0 = f(N_d, m^*, \dots) \quad (17)$$

5. Circuit équivalent en petits signaux.

- Résistance dynamique $R_d = \frac{dV}{dI}$
- Capacité différentielle $C_d = A \left(\frac{eN_d \epsilon_{SC}}{2(V_{bi} - V)} \right)^{\frac{1}{2}}$
- Résistance série de la diode $R_s = R_{contacts} + R_{RN}$
- Inductance parasite L_s
- Capacité « géométrique » de la diode $C_{geom} = \frac{\epsilon_{SC} A}{L}$

Tous ces éléments sont reportés sur la **figure 5** (Singh) ci dessous.

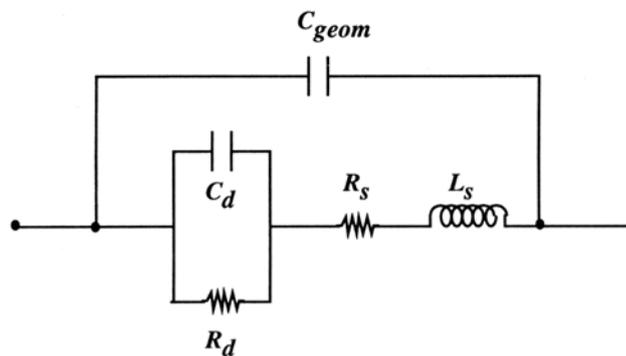


Figure 5 : Schéma équivalent en petits signaux (Singh)

Chapitre 3 Chapitre 4 : Transistor Bipolaire

1. Introduction : c'est le 23 décembre 1947, que le premier transistor est utilisé pour amplifier la voix. C'était un transistor à « contacts points », réalisé par Bardeen et brattain (Physical Review **74**, p230, 1948). Le 30 juin 1948, la nouvelle est annoncée au public. W. Shockley décrit la théorie de fonctionnement en 1949 (Bell System Tech. J. **28**, p435 1949). Le premier transistor à jonction est développé en 1950 (Physical Review **83**, p151, 1951). Le nom de bipolaire provient des charges qui contribuent au courant dans le dispositif, à la fois des électrons et des trous.

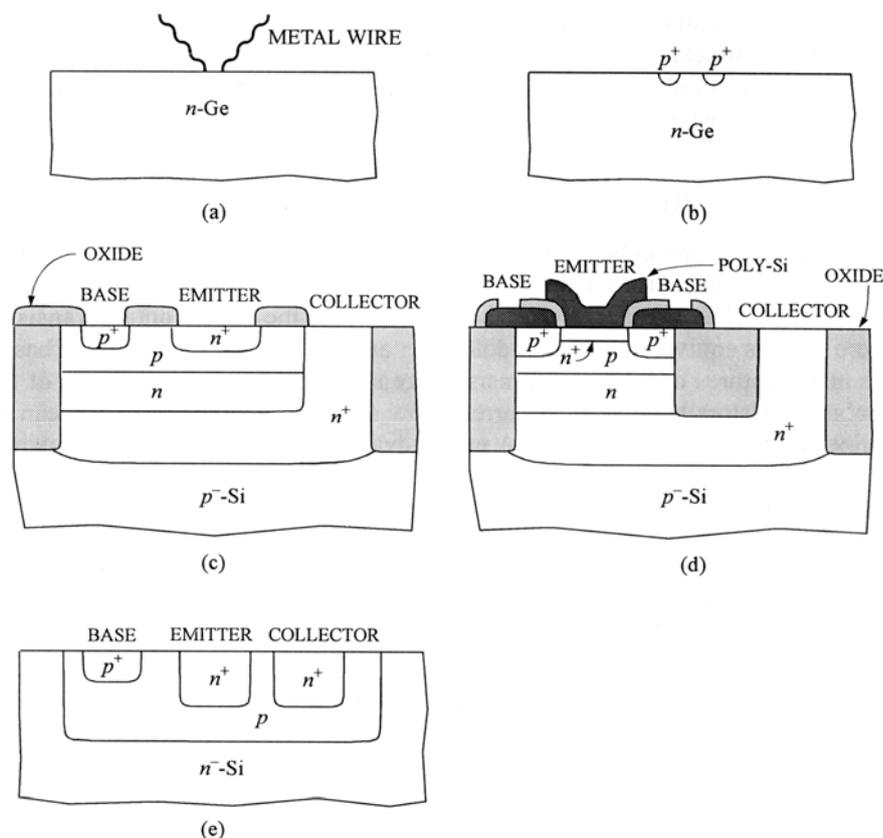


Figure 4.1 : les différents types de transistors bipolaires. (Ng)

2. Transistor bipolaire (BJT) – Principe de fonctionnement . Un des besoins très important pour les dispositifs électroniques est qu'une faible variation du signal d'entrée induise un fort changement du signal de sortie, d'où la notion de gain. Une deuxième nécessité est la notion d'isolation entre la sortie et l'entrée.

Regardons rapidement comment le profil de potentiel dans une structure semi-conductrice peut être contrôlé par un courant de base.

Nous savons que le courant inverse d'une jonction PN est proportionnel à la densité de porteurs minoritaires et indépendant de la tension inverse. Le BJT exploite ce simple fait. Le défi était de trouver un moyen simple pour injecter les porteurs de charges

minoritaires. Il suffit en fait d'utiliser le principe d'une jonction PN polarisée en direct. Dans une jonction PN, par exemple, le courant dépend des électrons injectés vers la région P et des trous injectés dans la région N. Si on utilise une diode N⁺P, on peut négliger l'injection des trous et le courant est essentiellement dû à l'injection d'électrons dans la région P. De plus, ce courant peut être facilement modifié par une faible variation de la tension de polarisation directe ($I \propto \exp(\frac{eV_A}{kT})$). Il suffit donc pour réaliser un transistor

bipolaire, par exemple un NPN, de juxtaposer une jonction N⁺P polarisée en direct qui va nous servir à injecter le courant et une jonction PN en inverse pour le collecter. La région N⁺ est l'*émetteur*, la région centrale P, la *base* et la région N le *collecteur*.

Les électrons injectés dans la base par la polarisation directe, pour peu que la base soit « fine » ou « étroite », vont pour la majorité d'entre eux se retrouver propulsés dans le collecteur par le champ électrique présent dans le ZCE de la jonction base – collecteur polarisée en inverse. Une minorité d'entre eux se seront recombinaés avec les trous majoritaires dans la base, créant un «appel» trous dans la base pour maintenir celle-ci neutre et par là même créant une composante (nous verrons négligeable) du courant de base.

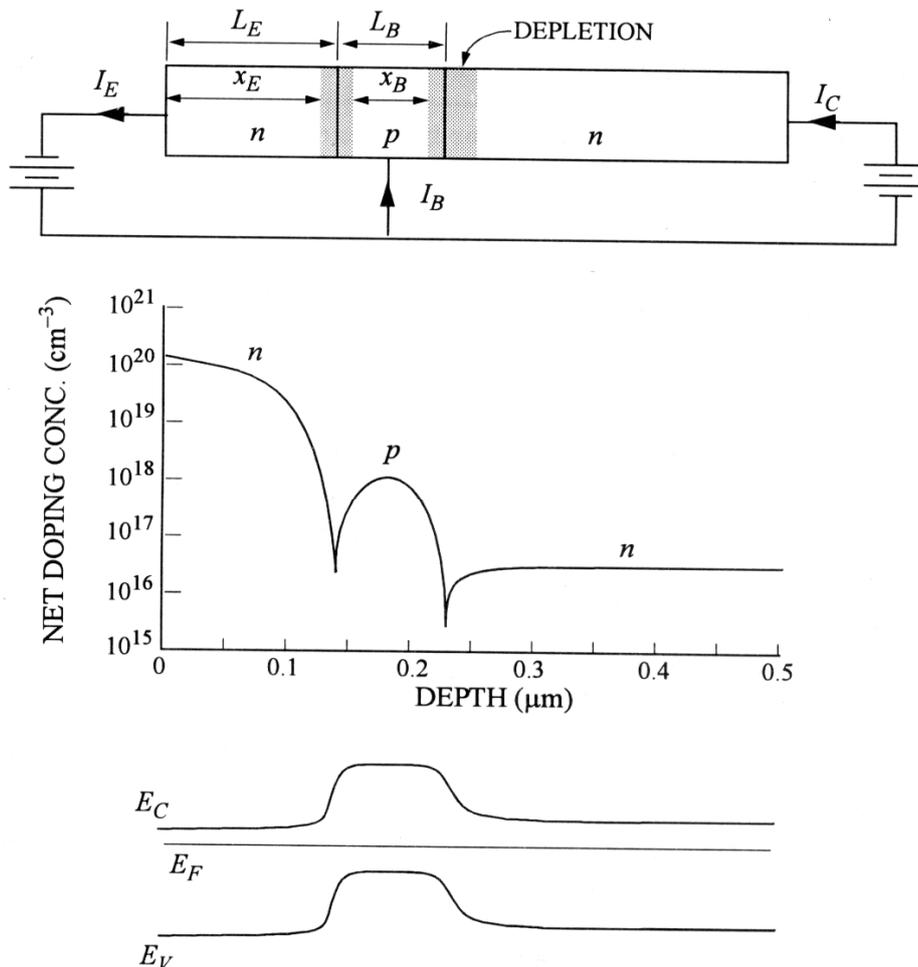


Figure 4.2 : dopage et bandes d'énergies dans un npn (Ng)

3. **Caractéristiques statiques du BJT.** Le BJT possède une orientation physique complexe du flux de courant. Les porteurs de l'émetteur sont injectés «verticalement» à travers la base dans le collecteur alors que les charges dans la base sont injectées depuis le côté du dispositif. On devrait donc faire une étude «2D». En première approximation, on traitera le problème à «1D» et on ajoutera d'autres hypothèses simplificatrices :

- Analyse 1D
- Dopage homogène de la base ($E_i = 0$)
- Faible injection
- Tensions appliquées continues

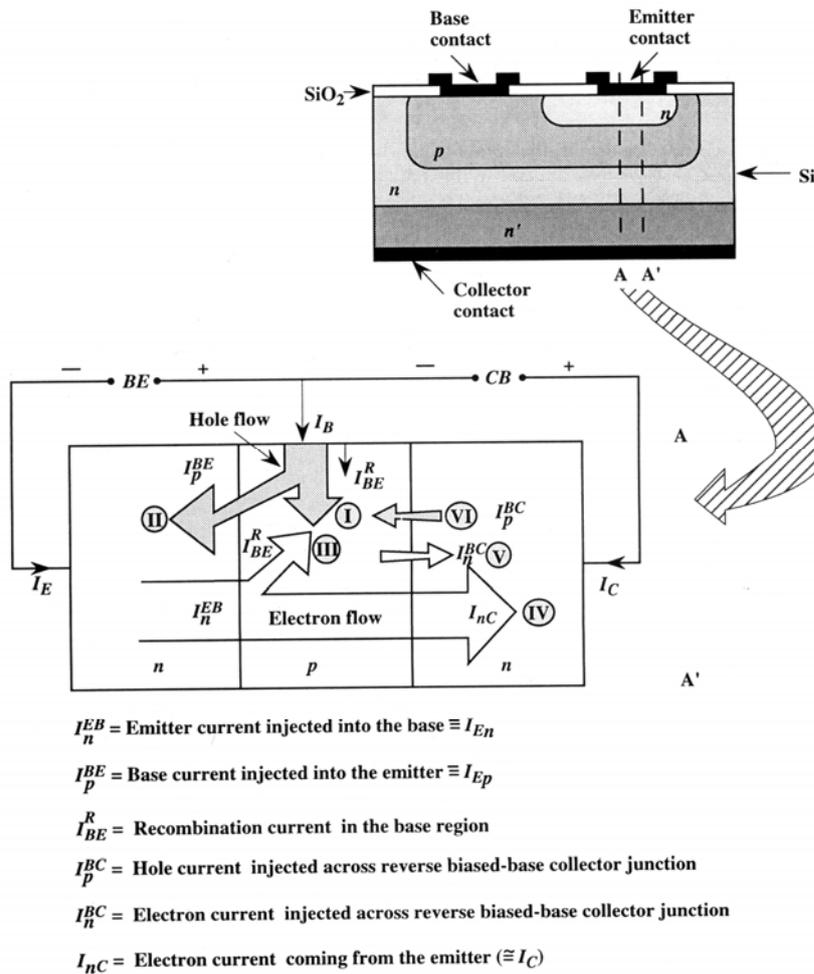


Figure 4.3 :visualisation des courant dans un transistor npn (Singh)

Un grand nombre de courants peut être identifié dans un BJT :

Dans la Base : le courant est formé de la recombinaison de trous avec les électrons injectés de l'émetteur vers la base et de trous injectés de la base vers l'émetteur. On doit également rajouter un courant de recombinaison dans la ZCE de la jonction E – B.

Dans l'émetteur : un courant d'électrons qui se recombinent avec les trous dans la base et un courant d'électrons qui est injecté dans le collecteur.

Dans le collecteur : un courant de trous et d'électrons injectés par la polarisation inverse de la jonction. En général, on néglige ces composantes devant les autres.

La figure 4.4 montre sur un diagramme énergétique ces mêmes composantes. On va maintenant, calculer toutes les composantes des courants présents dans le transistor en se servant des résultats de la jonction PN. Dans un premier temps on négligera le courant de recombinaison dans les ZCE, à savoir I_{rD} et le courant de recombinaison dans la base I_{rB} .

- **courant I_{En}** : c'est le courant des électrons injectés de l'émetteur vers le collecteur et qui circule dans la base. L'origine de ce courant dans la base est la diffusion des électrons injectés au niveau de la jonction E – B vers la jonction C – B. A partir des expressions des courants de trous et d'électrons (voir Chap. 1, eq.14) et en éliminant le champ électrique, on obtient :

$$p \frac{J_n}{eD_n} - n \frac{J_p}{eD_p} = \frac{d(p.n)}{dx} \quad (1)$$

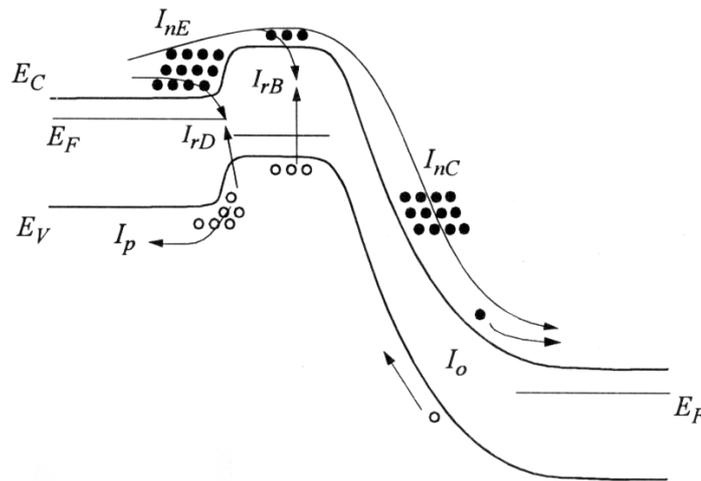


Figure 4.4 : visualisation des différents flux de porteurs dans un BJT en régime normal de fonctionnement (Ng)

Cette équation se simplifie en remarquant que dans la base $J_n \gg J_p$ et $n \ll p$ soit :

$$p \frac{J_n}{eD_n} = \frac{d(p.n)}{dx} \quad (2)$$

En intégrant cette équation de la jonction E – B à la jonction C – B, on obtient :

$$J_n = -eD_{nb}n_i^2 \frac{\left[\exp\left(\frac{eV_{BE}}{kT}\right) - \exp\left(\frac{eV_{BC}}{kT}\right) \right]}{\int_E^C p(x)dx} \quad (3)$$

ce qui se réécrit :

$$J_n = -eD_{nb}n_i^2 \frac{\left[\left(e^{\frac{eV_{BE}}{kT}} - 1 \right) - \left(e^{\frac{eV_{BC}}{kT}} - 1 \right) \right]}{\int_E^C p(x)dx} \quad (4)$$

En régime normal de fonctionnement, J_n est négatif, car les électrons se déplacent vers les x positifs, ce qui conduit à un courant négatif.

- **Courant dans l'émetteur et le collecteur** : les expressions de ces courants sont celles obtenues dans le chapitre « jonction PN » et sont donc :

$$J_{pE} = J_{spE} \left(\exp\left(\frac{eV_{BE}}{kT}\right) - 1 \right) \quad (5)$$

$$J_{pC} = J_{spC} \left(\exp\left(\frac{eV_{BC}}{kT}\right) - 1 \right) \quad (6)$$

A partir des conventions de signes, on obtient :

$$I_E = I_n - I_{pE} = -\left[\frac{Aen_i^2 D_{nb}}{C} + I_{spE} \right] \left(\exp\frac{eV_{BE}}{kT} - 1 \right) + \frac{Aen_i^2 D_{nb}}{C} \left(\exp\frac{eV_{BC}}{kT} - 1 \right) \quad (7)$$

$$I_C = I_n + I_{pC} = -\frac{Aen_i^2 D_{nb}}{C} \left(\exp\frac{eV_{BE}}{kT} - 1 \right) + \left[\frac{Aen_i^2 D_{nb}}{C} + I_{spC} \right] \left(\exp\frac{eV_{BC}}{kT} - 1 \right) \quad (8)$$

$$I_B = -I_E + I_C = I_{pE} - I_{pC} = I_{spE} \left(\exp\left(\frac{eV_{BE}}{kT}\right) - 1 \right) - I_{spC} \left(\exp\left(\frac{eV_{BC}}{kT}\right) - 1 \right) \quad (9)$$

3. **Equations d'Ebers – Moll.** Dans ce modèle, le transistor est idéalisé par deux jonctions PN tête bêche, et qui peuvent être traitées séparément. Dans le cas où la jonction E – B est polarisée en direct et la jonction C – B non polarisée, les courants I_E et I_C s'écrivent :

$$I_E = -I_{S_1} \left(\exp\frac{eV_{BE}}{kT} - 1 \right) \quad (10)$$

$$I_C = -\alpha_N I_E = +\alpha_N I_{S_1} \left(\exp\frac{eV_{BE}}{kT} - 1 \right) \quad (11)$$

I_{S_1} est le courant de saturation de la diode E – B

α_N est le gain en courant en fonctionnement normal.

On peut faire la même démarche pour la diode C – B, ce qui s'écrit :

$$I_C = -I_{S_2} \left(\exp\frac{eV_{BC}}{kT} - 1 \right) \quad (12)$$

$$I_E = -\alpha_I I_C = +\alpha_I I_{S_2} \left(\exp\frac{eV_{BC}}{kT} - 1 \right) \quad (13)$$

I_{S_2} est le courant de saturation de la diode C – B

α_I est le gain en courant en fonctionnement inverse.

Enfin, on doit envisager le cas où les deux jonctions sont polarisées en direct (régime saturé) ou en inverse (régime bloqué). Le cas général est obtenu en combinant les deux écritures (équations (10) à (13)) :

$$I_E = -I_{S_1} \left(\exp \frac{eV_{BE}}{kT} - 1 \right) + \alpha_I I_{S_2} \left(\exp \frac{eV_{BC}}{kT} - 1 \right) \quad (14)$$

$$I_C = +\alpha_N I_{S_1} \left(\exp \frac{eV_{BE}}{kT} - 1 \right) - I_{S_2} \left(\exp \frac{eV_{BC}}{kT} - 1 \right) \quad (15)$$

avec :

$$I_{S_1} = \frac{Ae^2 n_i^2 D_{nb}}{\int_E^C ep(x) dx} + \frac{Aen_i^2 D_{pe}}{\int_E^C N_D(x)} \quad (16) \quad \text{et} \quad I_{S_2} = \frac{Ae^2 n_i^2 D_{nb}}{\int_C^E ep(x) dx} + \frac{Aen_i^2 D_{pc}}{\int_C^E N_D(x)} \quad (17)$$

L'intégrale sous le premier terme de I_{S_1} (I_{S_2}) représente la charge de porteurs libres contenue dans la base. Cette charge est composée des trous (la base est dopée P) Q_B et des électrons (les porteurs minoritaires) Q_S qui sont injectés de l'émetteur et/ou du collecteur. En toute généralité, cette intégrale peut donc se réécrire $Q_B + Q_S$ où Q_S est appelée également *charge stockée*. Cette charge jouera un rôle primordial dans l'étude du transistor en commutation. L'expression littérale de ces intégrales a déjà été démontrée dans le chapitre sur la jonction PN et dépend de la géométrie de celle ci (voir équations 26a à 26c du chapitre correspondant).

Dans le cas des faibles niveaux de polarisation, la charge dans la base (région neutre exclusivement) est essentiellement Q_B et on définit le **nombre de Gummel** comme

étant $G_N = \left| \frac{Q_B}{e} \right|$ soit:

$$G_N = \int N_A dx \text{ (nnp)} \text{ ou } G_N = \int N_D dx \text{ (pnp)}. \quad (18)$$

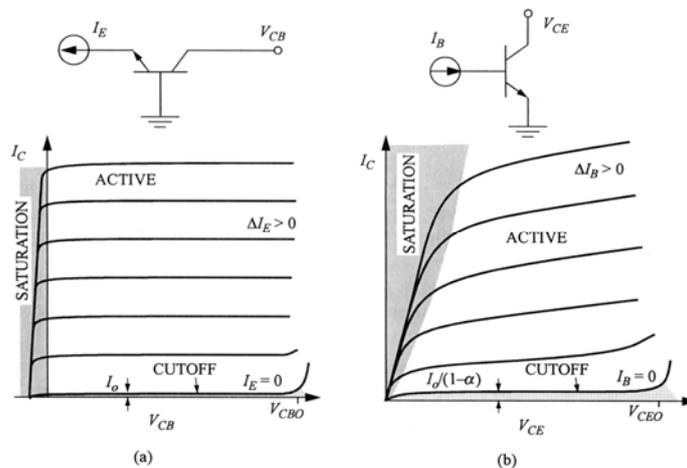


Figure 4.6 : a) caractéristiques BC et b) caractéristiques EC (Ng)

4. Paramètres statiques du transistor bipolaire. Trois montages correspondants à trois types de polarisations sont utilisés : émetteur commun, base commune et collecteur commun. Les caractéristiques sont représentées pour les deux premiers montages sur la figure 4.5. D'après ces figures, on peut observer trois modes de fonctionnement, à savoir les modes saturé (les deux jonctions en direct), normal (E – B en direct et C – B en inverse) et bloqué (les deux jonctions en inverse). Dans la suite du paragraphe on va se placer dans le régime normal de fonctionnement à savoir $eV_{BE} \gg kT$ et $eV_{CB} \gg kT$ (polarisation inverse), ce qui permet de simplifier les expressions des différents courants comme suit :

$$I_E = -\left(\frac{Ae^2 n_i^2 D_{nB}}{Q_B + Q_S} + I_{SpE}\right) \exp\frac{eV_{BE}}{kT} \quad (19a)$$

$$I_C = -\left(\frac{Ae^2 n_i^2 D_{nB}}{Q_B + Q_S}\right) \exp\frac{eV_{BE}}{kT} \quad (19b)$$

$$I_B^* = I_E - I_C = I_{SpE} \exp\frac{eV_{BE}}{kT} \quad (19c)$$

L'expression (19b) est souvent écrite sous la forme

$$|I_C| = A_E J_{C0} \exp\left(\frac{eV_{BE}}{kT}\right) \quad (19d)$$

où J_{C0} est la densité de courant collecteur saturé qui s'écrit

$$J_{C0} = \frac{e}{\int_{E'}^{C'} (p_p / D_{nB} n_i^2) dx} \quad (19e)$$

- **Efficacité d'injection d'émetteur.** L'efficacité d'émetteur est le rapport du courant d'électrons (dans un npn) du à l'injection d'électrons depuis l'émetteur au courant de trous. Donc :

$$\gamma_E = \frac{I_n}{I_{Ep}} \quad (20)$$

nota : dans la littérature anglo – saxonne l'efficacité d'injection est donnée par

$$\frac{I_n}{I_n + I_{Ep}} = \frac{I_n}{I_E} = \frac{\gamma_E}{1 - \gamma_E}.$$

- **Gain en courant en base commune.** Il est défini comme le rapport du courant collecteur au courant d'émetteur, soit :

$$\alpha = \frac{I_C}{I_E} = \frac{1}{1 + \frac{J_{Sp} \cdot (Q_B + Q_S)}{e^2 n_i^2 D_{nB}}} \quad (21a)$$

Dans le cas d'un PNP, un signe moins apparaît dans le rapport (conventions de signes !).

- **Gain en courant en émetteur commun.** il est donné par le rapport du courant collecteur au courant de base, soit :

$$\beta = \frac{I_C}{I_B} = \frac{\alpha}{1-\alpha} \quad (21b)$$

Remarquons, qu'ici compte tenu du fait que l'on **néglige les effets de recombinaisons dans la base, le gain β se confond avec l'efficacité d'injection γ_E .**

Il est bien évident que cette dernière remarque est très loin de décrire la réalité du fonctionnement du transistor. En effet deux phénomènes importants de recombinaisons dans la base sont présents et modifient considérablement les performances « théoriques » du transistor. Pour en tenir compte on va introduire le facteur de transport .

- **Facteur de transport dans la base.** Il est évident qu'à la traversée de la région de base, certains électrons (trous dans un PNP) vont se recombiner avec les trous majoritaires dans la base. Il est certain que si on veut que l'effet transistor persiste, le courant de recombinaison doit rester petit devant le courant collecteur !! Calculons le courant de recombinaison I_{rB} dans la région neutre de la base : il est donné par le rapport de la charge stockée dans la base au temps de recombinaison des électrons dans la base, ce temps est la durée de vie des porteurs τ_n . Ainsi I_{rB} est donné par :

$$I_{rB} = \frac{Q_s}{\tau_n} = \frac{AeX_{B_{eff}} n_p}{\tau_n} \quad (21)$$

Le courant collecteur peut se mettre sous la forme $I_C = Q_s / \tau_t$ avec τ_t le temps de transit des électrons dans la base.

!!!! Attention, dans le calcul de Q_s , on a considéré que la base était courte et on a « pris » le temps de transit comme temps pertinent !!!!

On peut donc écrire le facteur de transport δ comme étant égal à :

$$\delta = \frac{I_C}{I_{rB}} = \frac{\tau_n}{\tau_t} = \frac{2L_n^2}{X_{B_{eff}}^2} > 1 \quad (22)$$

nota : dans la littérature anglo – saxonne le facteur de transport (de la base) est noté B et est défini par $I_C = BI_n$ et donc dans notre supposition (**base courte**) $B = \frac{\delta - 1}{\delta} < 1$.

A partir des résultats précédents et en incluant le phénomène de recombinaisons dans la ZCE (largeur W_T) de la jonction E – B qui s'écrit (Eq 34 Chap 2)

$I_{rD} = \frac{Aen_i}{2\tau} W_T \exp\left(\frac{eV_{BE}}{2kT}\right)$, le courant total de Base s'écrit :

$$I_B = I_B^* + I_{rB} + I_{rD} \quad (23)$$

avec I_B^* le courant intrinsèque du transistor (sans recombinaisons). On peut alors écrire le gain β sous la forme :

$$\frac{1}{\beta} = \frac{I_B}{I_C} = \frac{I_B^* + I_{rB} + I_{rD}}{I_C} = \frac{1}{\gamma_E} + \frac{1}{\delta} + \frac{I_{rD}}{I_C} \quad (24)$$

- **Régime saturé** : dans ce régime, les deux jonctions sont en directes. On peut alors négliger l'unité devant l'exponentielle et les expressions (14) et (15) se résument à :

$$I_E \cong - \left(\frac{Ae^2 n_i^2 D_{nB}}{Q_S + Q_B} + I_{spE} \right) e^{\frac{eV_{BE}}{kT}} + \frac{Ae^2 n_i^2 D_{nB}}{Q_S + Q_B} e^{\frac{eV_{BC}}{kT}} \quad (25)$$

$$I_C \cong - \frac{Ae^2 n_i^2 D_{nB}}{Q_S + Q_B} e^{\frac{eV_{BE}}{kT}} + \left(\frac{Ae^2 n_i^2 D_{nB}}{Q_S + Q_B} + I_{spC} \right) e^{\frac{eV_{BC}}{kT}} \quad (26)$$

On va envisager deux cas, suivant le niveau de polarisation, faible ou fort et calculer la charge stockée dans la base Q_S qui est responsable du courant et des limitations en commutation du transistor.

- **Faible niveau de polarisation** : Ce calcul a déjà été effectué et on obtient pour le courant du aux électrons injectés dans la base une expression identique à (4), en remplaçant l'intégrale par Q_B (en faible injection $Q_S \ll Q_B$), soit $Q_B = eN_A d_{pB}$, d_{pB} étant la largeur effective de la base. Si on se place dans le cas d'une base courte, on sait (équation 46 du chapitre II) que la charge stockée Q_S (petite) qui est responsable du courant varie linéairement avec x dans la base (voir TD Jonction PN) et s'écrit $-\frac{1}{2} d_{pB} en(x = E') = -\frac{1}{2} d_{pB} \frac{n_i^2}{N_A} e^{\frac{eV_{BE}}{kT}} = \tau_t J_{sn} e^{\frac{eV_{BE}}{kT}}$. Si on fait le même raisonnement pour la charge due aux électrons injectés du collecteur vers l'émetteur, on obtient $-\frac{1}{2} d_{pB} en(x = C') = -\frac{1}{2} d_{pB} \frac{n_i^2}{N_A} e^{\frac{eV_{BC}}{kT}} = \tau_t J_{sn} e^{\frac{eV_{BC}}{kT}}$.

La charge totale stockée par un transistor saturé en faible injection est donc :

$$Q_S = \tau_t J_{sn} e^{\frac{eV_{BC}}{kT}} + \tau_t J_{sn} e^{\frac{eV_{BE}}{kT}} \quad (27)$$

On peut également redéfinir la charge stockée en saturation comme la somme d'une charge en régime normal mais sous un courant $I_{C_{sat}}$ plus la charge en excès qui sature le transistor, soit $Q_S = Q_{S_{normal}} + Q_{SAT}$ avec Q_{SAT}

qui s'écrit $Q_{SAT} = 2\tau_t J_{sn} e^{\frac{eV_{BC}}{kT}}$ (voir TD BJT). Cette charge supplémentaire introduira une dégradation des performances dynamiques du transistor (temps d'évacuation de ces charges Q_{SAT} ou temps de désaturation).

- **Fort niveau de polarisation** : Le fait d'être en forte injection implique que dans la base $p = \bar{p} + \Delta p = \bar{p} + \Delta n \cong \Delta n$. En d'autres termes, la densité d'électrons injectés est égale à la densité de trous dans la base $\Delta n \cong p \cong n$, ce qui assure la neutralité électrique dans la base. On a plus d'électrons et de trous qu'à l'équilibre. Compte tenu de la dissymétrie de dopage (N⁺P) de la jonction, on peut dire que $J_n > J_p$, ce qui permet d'écrire (1) sous la forme :

$$J_n = 2eD_{nB} \frac{dn}{dx} \quad (27)$$

Si on suppose que D_{nB} est constante (c'est une approximation en forte injection), on voit que la variation du nombre de porteurs (ici les électrons) est linéaire avec x. Soit :

$$J_n = \frac{2eD_{nB}}{d_{pB}} \int_{E'}^{C'} dn = \frac{2eD_{nB}}{d_{pB}} n(x=C') - \frac{2eD_{nB}}{d_{pB}} n(x=E') \quad (28)$$

On néglige la densité d'électrons en $x = C'$ (ils sont évacués vers le collecteur). Ainsi, (28) s'écrit finalement :

$$J_n = -\frac{2eD_{nB}}{d_{pB}} n_i e^{\frac{eV_{BE}}{2kT}} \quad (29)$$

Q_S s'écrit, à partir de la figure xx, $Q_S = -\frac{1}{2} en_i e^{\frac{eV_{BE}}{2kT}} d_{pB}$. A partir de (29), on peut exprimer Q_S en fonction de J_n et on obtient :

$$Q_S = \frac{d_{pB}^2}{4D_{nB}} J_n \quad (30)$$

Or ici, le transistor est saturé, donc on injecte également des électrons du collecteur vers la base et donc Q_S et J_n s'écrivent en fait :

$$Q_S = -\frac{1}{2} en_i (e^{\frac{eV_{BE}}{2kT}} + e^{\frac{eV_{BC}}{2kT}}) d_{pB} \quad (31)$$

$$J_n = -\frac{2eD_{nB}}{d_{pB}} n_i e^{\frac{eV_{BE}}{2kT}} - \frac{2eD_{nB}}{d_{pB}} n_i e^{\frac{eV_{BC}}{2kT}} \quad (32)$$

En fait des effets secondaires vont modifier ces expressions, effets que nous allons examiner maintenant.

5. Effets secondaires sur un transistor bipolaire.

- **Modulation de la largeur de la base : Effet Early.** Dans l'expression des courants intervient la charge dans la base Q_B . Or cette charge est proportionnelle entre autre à d_{pB} qui est la largeur de la région neutre de la base, soit la largeur métallurgique moins l'extension des ZCE des jonctions E – B et C – B. Cette largeur est donc fonction des polarisations des deux jonctions, si bien que le courant va avoir une dépendance supplémentaire en tension. Comme la jonction

émetteur est polarisée en directe, on peut négliger l'extension de sa ZCE et c'est donc la polarisation de la jonction côté collecteur qui va surtout influencer la courant. En fait si $|V_{BC}|$ augmente, d_{pB} va diminuer comme $V_{BC}^{1/2}$. De ce fait δ va augmenter ainsi que γ_E et donc le courant collecteur. Cet effet n'est pas souhaitable dans la mesure où l'on veut que la sortie dépende uniquement de l'entrée. on introduit le **facteur de réaction d'Early** μ_C (voir Leturcq et Rey p100) définit comme :

$$\mu_C = \frac{kT}{e} \frac{1}{Q_B + Q_S} \frac{d(Q_B + Q_S)}{dV_{BC}} \quad (33)$$

Ainsi en base commune, la pente de la caractéristique de sortie est donnée par :

$$\left. \frac{dI_C}{dV_{BC}} \right|_{I_E=cte} = (1 - \alpha) \mu_C \frac{kT}{e} I_C \quad (34a)$$

et en émetteur commun :

$$\left. \frac{dI_C}{dV_{CE}} \right|_{I_B=cte} = \mu_C \frac{kT}{e} I_C \quad (34b)$$

A partir des relations précédentes, on voit que le montage en émetteur commun sera plus sensible à l'effet Early. On définit également une tension d'Early qui est le point d'intersection obtenu par extrapolations des caractéristiques de sortie et qui est donné par :

$$\frac{dI_C}{dV_{CE}} = \frac{I_C}{V_{CE} + V_A} \quad (34c)$$

En général, la tension d'Early est beaucoup plus grande que V_{CE} . Une expression approchée de la tension d'Early est donnée par $V_A \approx \frac{eN_A X_B^2}{\epsilon_{SC}}$. **L'extrapolation graphique est plus rigoureuse. (voir figure 4.6)**

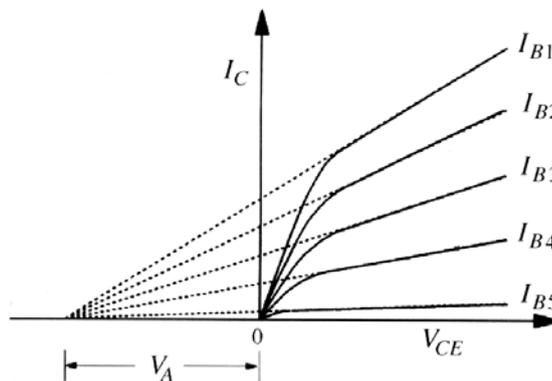


Figure 4.6: Effet Early et Tension d'Early (Taur)

Nota : si la tension augmente trop, (ou/et si la base est très fine), il peut y avoir **perçage** (« **punch through** ») de la base ($d_{pB} \Rightarrow 0$). La tension qui correspond à

cet effet est donnée, en négligeant la tension de diffusion de la jonction C – B ($\approx 0.7V$) par :

$$V_{pr} = \frac{eW_B^2 N_A (N_{Dc} + N_A)}{2\epsilon N_{Dc}} \quad (35)$$

- **Défocalisation du courant – résistance de Base («crowding effect»)** : dans tout ce qui précède, on a considéré le transistor en fonctionnement 1-D. En fait, compte tenu de la géométrie le courant qui circule de la base vers l'émetteur est perpendiculaire à la circulation du courant émetteur vers collecteur. La base étant peu dopée (résistive) et fine, il existe une chute de potentiel qui fait que la jonction E – B est plus polarisée en direct à la périphérie du contact qu'au centre du contact et donc l'injection du courant se fait de façon privilégiée à la périphérie et moins facile au centre d'où la notion de **défocalisation**. Ce phénomène est d'autant plus marqué que le niveau d'injection est élevé. Pour remédier à ce problème, on doit modifier la géométrie des contacts et en particulier tendre à augmenter le rapport périmètre/surface. On atteint cet objectif en utilisant des géométries à contacts base émetteur interdigitées.
- **Variation du gain en courant** : on constate que le gain en courant est fonction du courant collecteur, même à tension collecteur constante (ce qui élimine l'effet Early). Si on écrit le gain β en fonction des paramètres physiques du transistor, on obtient :

$$\beta \approx \left[\frac{N_A d_{pB} D_{pE}}{N_{DE} d_{nE} D_{nB}} + \frac{X_{Beff}^2}{2L_n^2} + \frac{I_{SrD} N_A X_{Beff}}{eAD_{nB} n_i^2} e^{\frac{-eV_{BE}}{2kT}} \right]^{-1} \quad (36)$$

Cette expression est valable en supposant la base **courte** et en négligeant le courant de recombinaison dans la région neutre de la base devant I_{nE} (Ng p 247). Pour des faibles tensions de polarisation V_{BE} , la composante de recombinaison I_{rD} peut être très grande et fournir la majeure partie du courant

de base I_B et le gain en courant s'écrit $\beta_N = \frac{I_C}{I_B} \approx \frac{I_C}{I_{rD}}$, qui est croissant avec

I_C . Le gain atteint alors une valeur maximum (lorsque I_C augmente) qui est de l'ordre de 50 à quelques centaines et qui est la valeur obtenue sans tenir compte des recombinaisons dans la ZCE (I_{rD} devient négligeable). Pour des polarisations fortes, on constate une décroissance de β . On obtient ce résultat

en remarquant que dans ces conditions, β s'écrit $\beta = \frac{e^2 n_i^2 D_{nB}}{J_{spE} Q_S}$ (Letrurcq p

102). Si on se rappelle du traitement de la jonction PN en forte injection, la charge stockée Q_S dans la base est proportionnelle à la densité de courant d'électrons J_n , soit à la densité de courant de collecteur J_C et la charge s'écrit

$Q_S = J_C \frac{d_{pB}^2}{4D_{nB}}$. Cette expression permet de réécrire β sous la forme

$$\beta = \frac{4(en_i D_{nB})^2}{d_{pB}^2 J_{SnE} J_C}$$
 qui est bien décroissant avec I_C . On fera attention, qu'ici, la constante de diffusion est celle donnée en forte injection !! (figure 4.7 tirée de Ng)

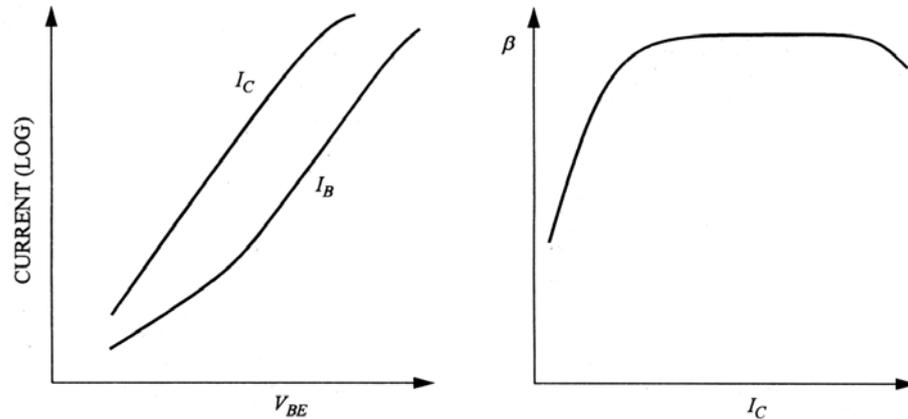


Figure 4.7 : variation du des courants en fonction de V_{BE} et variation du gain en fonction du courant collecteur (Ng)

- **Claquage par avalanche** : si on augmente trop fortement la tension inverse de la jonction C – B, le phénomène de claquage par avalanche (identique à celui de la jonction PN) se produit. Cet effet est, on le rappelle, du à la multiplication des porteurs dans la ZCE de la jonction en inverse, ici la jonction C – B. Dans le cas d'un montage BC, la tension de claquage BV_{CBO} est bien définie . En revanche, dans le cas du montage EC, le processus est plus compliqué (Leturcq p105), les courbes sont moins abruptes, et la tension BV_{CEO} une fraction de BV_{CBO} .

6. **Le Transistor Bipolaire en commutation.** Dans les applications numériques, le transistor bipolaire peut être utilisé en tant qu'interrupteur. Dans un interrupteur idéal, aucun courant ne doit circuler lorsque celui ci est ouvert, et dans l'état passant (« ON »), toute la tension doit se retrouver aux bornes de la charge. Ainsi, dans l'état « OFF », le transistor doit être bloqué, et dans l'état passant « ON », il doit être saturé. Un autre problème est de choisir le bon type de polarisation. Le choix doit être dicté par le fait que le signal de commande doit être le plus petit possible. On peut voir d'après la figure xxx, que seule le mode émetteur commun est acceptable. Ainsi, **le transistor en interrupteur est polarisé en EC.** La mise en conduction et le blocage ne se font pas instantanément, et les temps mis en jeu sont très importants dans les applications numériques. Les problèmes rencontrés sont les mêmes que dans le cas de la commutation de la jonction PN et sont dus au temps de désaturation dans la région de base. On va dans la suite étudier les temps mis en jeu.

- **Temps de mise en conduction** : le calcul du temps de mise en conduction, pour être simple, fait appel à l'équation de continuité de la charge (Roulston p412).

Si on considère l'équation de continuité dépendant du temps pour les électrons, elle s'écrit (voir chap I eq. 17):

$$\frac{dI_n}{dx} = eA\left(\frac{n-n_0}{\tau_n} + \frac{d(n-n_0)}{dt}\right) \quad (37)$$

Si on remplace le courant par un courant de diffusion, on obtient (chap I, eq. 19) :

$$D_n\left(\frac{d^2n}{dx^2}\right) = \frac{n-n_0}{\tau_n} + \frac{d(n-n_0)}{dt} \quad (38)$$

Ici le temps τ_n , est la durée de vie des porteurs dans la base, ou temps de recombinaison. Si maintenant, on intègre cette équation dans une région quelconque soit de $x = 0$ à $x = w$, on obtient

$$eAD_n\left[\left(\frac{dn'}{dx}\right)_{x=w} - \left(\frac{dn'}{dx}\right)_{x=0}\right] = \frac{d}{dt} \int_0^w eAn' dx + \frac{1}{\tau_n} \int_0^w eAn' dx \quad (39)$$

ce qui peut se réécrire :

$$I_{n(x=0)} - I_{n(x=w)} = \frac{dQ_B}{dt} + \frac{Q_B}{\tau_n} \quad (40)$$

avec Q_B , la charge instantanée stockée dans la région considérée. Si on peut négliger le courant total de minoritaires à l'abscisse w , l'équation se résume à **l'équation de continuité de la charge** qui est donnée par :

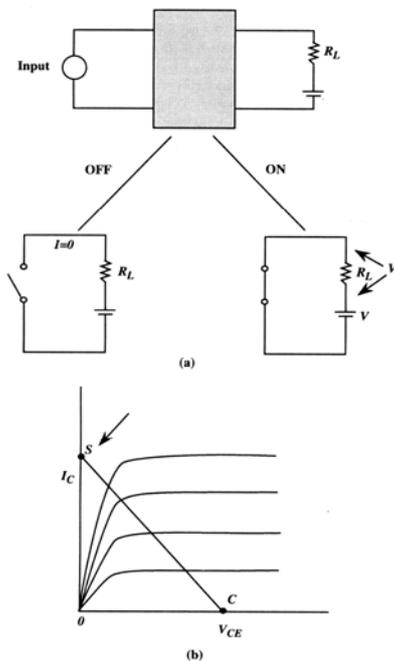


Figure 4.8 : Transistor en interrupteur (Singh)

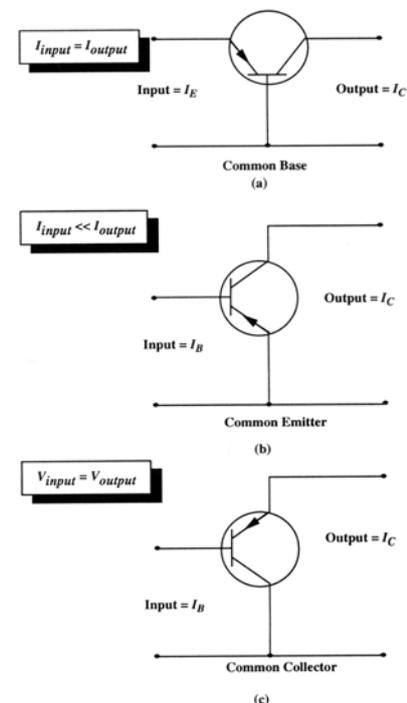


Figure 4.9 : circuits de polarisation (Singh)

$$I_n = \frac{dQ_B}{dt} + \frac{Q_B}{\tau_n} \quad (41)$$

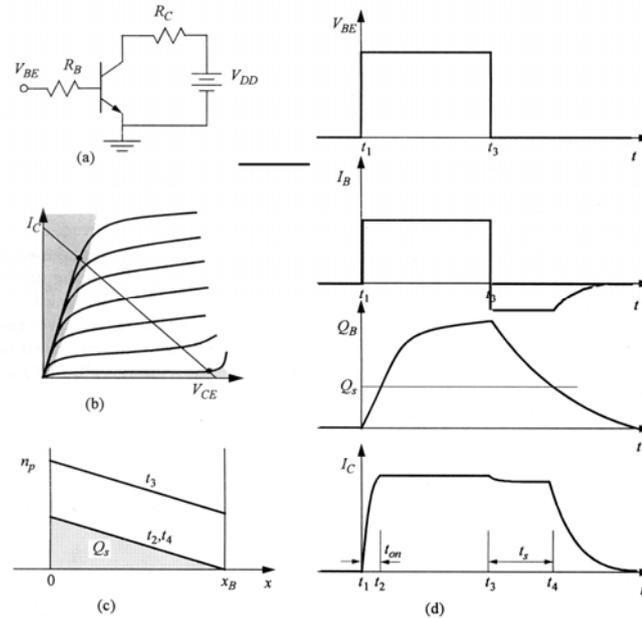


Figure 4.10 : commutation d'un transistor (Ng)

Cette équation est bien plus simple que l'équation faisant intervenir la densité de porteurs car ici, il n'est plus nécessaire de connaître la *forme* de la distribution des porteurs minoritaires. Revenons maintenant à l'étude de la commutation de notre transistor de l'état bloqué à l'état saturé (OFF et ON). A partir de (41), on peut déduire l'évolution de la charge stockée $Q_B(t)$ qui s'écrit :

$$Q_B(t) = I_B \tau_n [1 - \exp(-\frac{t}{\tau_n})] \quad (42)$$

On a remplacé le courant par le courant de base I_B car c'est lui qui va permettre à la charge $Q_B(t)$ de se créer. L'évolution du courant $I_C(t)$ est tout simplement donnée par $I_C(t) = \frac{Q_B(t)}{\tau_t}$ avec τ_t le temps de transit dans la base. Le courant collecteur

augmente jusqu'à atteindre sa valeur de saturation donnée par $I_{C_{sat}} \approx \frac{V_2}{R_L}$ en négligeant

$V_{CE_{sat}} \approx 0,1$ à $0,3$ volt. La charge limite $Q_B(t_{ON})$ pour saturer le transistor est donnée

par $Q_S = \frac{I_{C_{sat}} d_{pB}^2}{2D_{nB}}$. Le temps t_{ON} pour saturer le transistor est donc donné par :

$$t_{ON} = \tau_n \ln \left[\frac{1}{1 - (Q_S / I_B \tau_n)} \right] \quad (43)$$

Par contre la charge $Q_B(t)$ peut continuer à augmenter en fonction du niveau de polarisation de la base et donc sursaturer le transistor. D'après (42), on voit que la charge finale est donnée par $\tau_n I_B$ si t_3 est suffisamment grand.

• **Temps de blocage** : si maintenant la tension d'entrée passe à « 0 », le transistor doit se bloquer. Il est évident que le passage ne sera pas immédiat du fait du temps nécessaire pour évacuer la charge stockée $Q_B(t)$ dans la base (voir jonction PN). D'après toujours (42), on voit que la charge stockée va pendant un temps de stockage t_S décroître exponentiellement de $\tau_n I_B$ à Q_S , temps qui est donné par :

$$t_S = \tau_n \ln \left(\frac{I_B \tau_n}{Q_S} \right) \quad (44)$$

Une fois le transistor désaturé, le courant collecteur peut décroître exponentiellement de façon à bloquer le transistor. L'évolution est schématisée sur la figure xx(p250 Ng ou p328 singh).

Ce temps de stockage, comme dans la jonction PN, limite fortement la vitesse de commutation dans les circuits numériques. Deux façons permettent de réduire ce temps :

- Introduction d'impuretés dans la base pour « tuer » la durée de vie τ_n des minoritaires dans la base. La réduction de β peut être compensée par la diminution de la largeur de la base.
- Introduction d'une diode Schottky en parallèle sur la diode collecteur – base. Ceci évite au transistor de passer au-delà de la saturation et donc le temps de stockage n'a plus lieu d'être.

Rem 1 : Une conclusion supplémentaire est que pour des circuits « hautes vitesses », les transistors doivent rester dans le mode actif.

Rem 2 : Dans les transistors modernes, la charge en saturation est essentiellement dans la couche épitaxiée qui constitue le collecteur. Le temps qui intervient alors dans (44) est le temps de transit ou durée de vie (dépend de la géométrie) de cette couche. (Roulton p271)

7. **Transistor Bipolaire en HF** . Dans le modèle de fonctionnement en ac, les capacités internes parasites et les résistances doivent être incluses. En général, les résistances parasites peuvent être réduites par une géométrie *ad hoc* (grandes surfaces), des « masques » (lay-out) appropriés et des procédés de fabrication bien pensés. On peut en première approximation négliger ces résistances parasites. La figure XX (p333 Taur ou 332 Singh) représente le circuit équivalent en ac dans le cas où ces résistances sont négligées. On peut alors calculer les composants de ce circuit :

- **Transconductance** g_m : elle relie la variation du courant collecteur à la tension Base – Emetteur, soit :

$$g_m = \frac{\partial I_C}{\partial V_{BE}} = \frac{e I_C}{kT} \quad (45)$$

- **Résistance d'entrée** r_π : elle relie la variation de la tension Base – Emetteur au courant de base, soit :

$$r_\pi = \left(\frac{\partial I_B}{\partial V_{BE}} \right)^{-1} = \frac{kT}{eI_B} = \frac{\beta}{g_m} \quad (46)$$

- **Résistance de sortie** r_o :

$$r_o = \left(\frac{\partial I_C}{\partial V_{CE}} \right)^{-1} = \frac{V_A}{I_C} \quad (47)$$

où V_A est la tension d'Early définie par (34c).

- **Capacité** C_π : elle est définie comme la somme de deux capacités, la capacité de diffusion de la jonction E – B polarisée en direct et de la capacité de jonction de la zone de déplétion associée.

$$C_\pi = C_{SE} + C_{TEB} \quad (48)$$

la capacité de stockage (ou de diffusion) est donnée (voir plus haut et jonction PN) par

$$C_{SE} = \tau_F g_m \quad (48a)$$

avec τ_F le temps de transit en direct (F pour « forward »). Comme cette capacité est due à l'injection de porteurs minoritaires (diode E – B en direct), qui sont présents dans l'émetteur, la base, la ZCE E – B et al ZCE C – B, on peut écrire τ_F comme la somme de différents temps, soit :

$$\tau_F = t_E + t_{t_b} + t_{t_{BE}} + t_{t_{BC}} \quad (48b)$$

avec, t_E le temps de retard (« delay time ») de l'émetteur, t_t les temps de transit dans la base, la ZCE E – B et la ZCE C – B.

- **Capacité** C_μ : c'est la capacité de jonction de la jonction C –B polarisée en inverse, soit $C_\mu = C_{TCB}$
- **Capacité de la couche de déplétion de la diode collecteur – substrat**, C_{dCS} .

Lorsque les résistances parasites sont incorporées dans le modèle, on obtient le schéma de la figure .

- **Fréquence de coupure** f_T (« cut off frequency ») : elle est déterminée par la fréquence qui rend le gain en émetteur commun unitaire. D'après la figure , on a :

$$i_c = g_m v_{be} - j\omega C_\mu v_{be} \quad (49a)$$

$$i_b = \left(\frac{1}{r_\pi} + j\omega C_\pi + j\omega C_\mu \right) v_{be} \quad (49b)$$

le gain $\beta(\omega)$ est alors donné par $\frac{i_c}{i_b}$. En basse fréquence on obtient bien le gain statique du transistor. Pour des transistors modernes, compte tenu des paramètres physique et géométrique, le gain peut se résumer à :

$$\beta(\omega) \approx \frac{g_m r_\pi}{1 + j\omega r_\pi (C_\pi + C_\mu)} \quad (50)$$

Aux hautes fréquences, c'est la partie imaginaire qui domine et le gain s'écrit :

$$\beta(\omega) \approx \frac{g_m}{j\omega(C_\pi + C_\mu)} \quad (51)$$

L'expression de la fréquence f_T est immédiate et on obtient :

$$\frac{1}{2\pi f_T} = \tau_F + \frac{kT}{eI_C} (C_{T_{BE}} + C_{T_{CB}}) \quad (52)$$

On a, ici, négligé les résistances parasites.

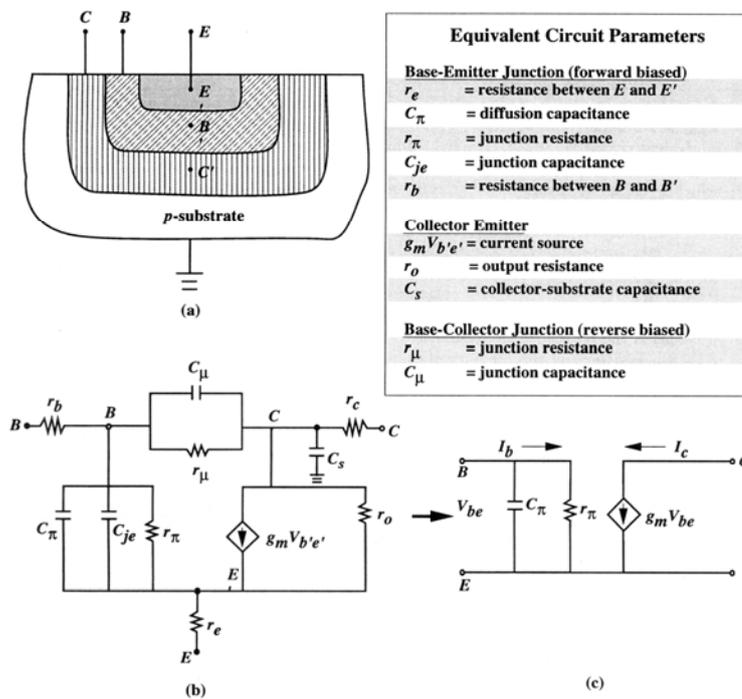


Figure 4.11 : circuit équivalent d'un TBJ en petits signaux. (Singh)

- **Fréquence d'oscillation max** f_{max} : une autre figure de mérite qui tient compte du paramètre important qui est la résistance de base (f_T n'en tient pas compte) est déterminée en considérant la fréquence pour laquelle le gain en puissance $G_{p_{max}}$ est unitaire. D'après la figure (p 177 Roulston), on obtient $G_p = \frac{g_m^2 R_L}{4r_b \omega^2 C_\pi^2}$. On obtient en final (Roulston p177) :

$$f_{\max_{osc}} = \sqrt{\frac{f_t}{8\pi C_{T_{CB}} r_b}} \quad (53)$$

Régime de fonctionnement	Jonction Emetteur – Base	Jonction Collecteur – Base
Actif, normal	Direct	Inverse
Bloqué (« cutoff »)	Inverse	Inverse
Saturation	Direct	Direct
Inverse	Inverse	Direct

Transistor NPN	Transistor PNP
$I_E = I_n - I_{pE} = -\left[\frac{Aen_i^2 D_{nb}}{C} + I_{spE}\right] \left(\exp\frac{eV_{BE}}{kT} - 1\right) + \frac{Aen_i^2 D_{nb}}{C} \left(\exp\frac{eV_{BC}}{kT} - 1\right)$	$I_E = I_p + I_{nE} = +\left[\frac{Aen_i^2 D_{pb}}{C} + I_{snE}\right] \left(\exp\frac{eV_{EB}}{kT} - 1\right) - \frac{Aen_i^2 D_{pb}}{C} \left(\exp\frac{eV_{CB}}{kT} - 1\right)$
$I_C = I_n + I_{pC} = -\frac{Aen_i^2 D_{nb}}{C} \left(\exp\frac{eV_{BE}}{kT} - 1\right) + \left[\frac{Aen_i^2 D_{nb}}{C} + I_{pC}\right] \left(\exp\frac{eV_{BC}}{kT} - 1\right)$	$I_C = -I_p + I_{nC} = -\frac{Aen_i^2 D_{pb}}{C} \left(\exp\frac{eV_{EB}}{kT} - 1\right) + \left[\frac{Aen_i^2 D_{pb}}{C} + I_{nC}\right] \left(\exp\frac{eV_{CB}}{kT} - 1\right)$
$I_B = -I_E + I_C = I_{pE} - I_{pC} = I_{spE} \left(\exp\frac{eV_{BE}}{kT} - 1\right) + I_{spC} \left(\exp\frac{eV_{BC}}{kT} - 1\right)$	$I_B = I_E + I_C = I_{pE} + I_{pC} = I_{snE} \left(\exp\frac{eV_{EB}}{kT} - 1\right) + I_{snC} \left(\exp\frac{eV_{CB}}{kT} - 1\right)$

9. Transistor Bipolaire à Hétérojonction. Dans un BJT, parmi les paramètres importants, figure les gains α et β . L'expression du gain α (en tenant compte des phénomènes de recombinaisons dans la base) s'écrit :

$$\alpha = \frac{\gamma_E}{1 - \gamma_E} \frac{\delta - 1}{\delta} = \left[1 - \frac{n_{i_e}^2}{N_{D_E}} \frac{D_{pE}}{D_{n_B}} \frac{N_{A_B}}{n_{i_B}^2} \frac{W_{Beff}}{L_{pE}} \right] \left[1 - \frac{W_{Beff}^2}{2L_{n_B}^2} \right] \quad (54)$$

Si la base est courte, le deuxième crochet se réduit à 1, et α se réduit à :

$$\alpha = \frac{\gamma_E}{1 - \gamma_E} \frac{\delta - 1}{\delta} = \left[1 - \frac{n_{i_e}^2}{N_{D_E}} \frac{D_{pE}}{D_{n_B}} \frac{N_{A_B}}{n_{i_B}^2} \frac{W_{Beff}}{L_{pE}} \right] \quad (55)$$

Si on veut un β le plus grand possible, on doit rendre α le plus proche de l'unité. Plusieurs options sont possibles :

- Diminution du dopage de la base

- Diminution de la longueur de la base, ce qui en fait diminue le nombre de Gummel $N_{A_B} \cdot W_{Beff}$.

Ces deux options augmentent la résistance de base et dégrade les performances du composant (voir équation 53).

Un autre problème qui provoque une diminution du gain β et qui domine dans les dispositifs modernes où le dopage de l'émetteur est important, est la diminution de la bande interdite du matériau en fonction du dopage (« Band gap Shrinking ») de E_g à $E_g - \Delta E_g$. Si on revient sur l'expression de α , les valeurs de n_{i_E} et de n_{i_B} ne sont donc pas égales, car l'expression de la densité intrinsèque en fonction du gap s'écrit :

$$n_i(E_g - \Delta E_g) = n_i(E_g) \exp\left(\frac{\Delta E_g}{kT}\right) > n_i(E_g) \quad (56)$$

avec E_g fonction du dopage :

$$\Delta E_g = 22.5 \left(\frac{N_B (cm^{-3})}{10^{18}} \frac{300}{T(K)} \right)^2 \text{ en meV} \quad \text{pour } N_B < 10^{19} cm^{-3} \quad (57)$$

Si le dopage est supérieur à $10^{19} cm^{-3}$ la variation du gap est plus faible que celle donnée par (57). Dans notre cas, le dopage de la base étant très inférieur à celui de l'émetteur, $E_{g(base)} > E_{g(émetteur)}$ et donc $n_{i_E} > n_{i_B}$. De ce fait, le gain α et β s'écrivent :

$$\alpha = \left[1 - \frac{n_{i_E}^2}{N_{D_E}} \frac{D_{p_E}}{D_{n_B}} \frac{N_{A_B}}{n_{i_B}^2} \frac{W_{Beff}}{L_{p_E}} \exp\left(\frac{\Delta E_g}{kT}\right) \right] \quad (58)$$

$$\beta = \frac{\alpha}{1 - \alpha} = \left[\frac{N_{D_E}}{N_{A_B}} \frac{D_{n_B}}{D_{p_E}} \frac{L_{p_E}}{W_{Beff}} \exp\left(-\frac{\Delta E_g}{kT}\right) - 1 \right] \quad (59)$$

Ici, $\Delta E_g = E_{g(base)} - E_{g(émetteur)} > 0$.

Ainsi, le gain est sensible à l'écart de bande interdite entre l'émetteur et la base. On voit donc qu'il est difficile de concilier un fort dopage d'émetteur, une base peu dopée et fine avec un gain β important. Par contre on note que si la différence de gap est négative, le gain β peut augmenter très rapidement. Une façon très élégante d'obtenir une différence de gap est d'utiliser deux matériaux différents pour l'émetteur et la base de telle sorte que la différence de gap soit bien négative. On réalise ainsi une hétérojonction entre la base et l'émetteur (voir figure 7.20 p339 du singh). Un $\Delta E_g \approx 250 meV$ à 300 K conduit à une augmentation d'un facteur 10^4 sur le gain, ce qui permet d'augmenter le dopage de la base de façon importante et réduire donc la résistance de base et améliorer f_{max} sans nuire à la valeur de β . On peut également réduire le dopage de l'émetteur sans affecter de façon importante le gain, ce qui permet d'augmenter le ZCE coté émetteur et donc de réduire la capacité d'émetteur. On peut montrer également que la tension d'Early est fortement réduite (Taur *et al* p 367).

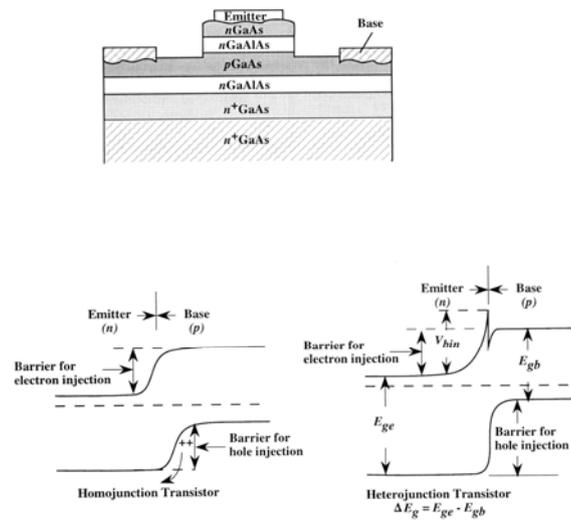


Figure 4.12 : Structure d'un HBT (Singh)

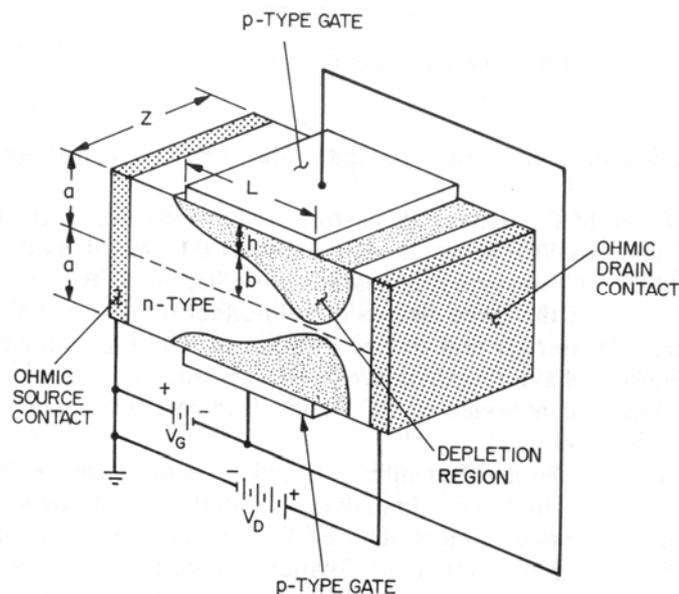
10. Références :

- H.Mathieu, « Physique des semi-conducteurs et des composants électroniques », 4^o édition, Masson 1998.
- P.Leturcq et G.Rey, « Physique des composants actifs à semi-conducteurs », Dunod Université, 1985.
- J. Singh, « semiconductors devices :an introduction », McGraw-Hill, Inc 1994.
- Y.Taur et T.H. Ning, « Fundamentals of Modern VLSI devices », Cambridge University Press, 1998.
- K.K. Ng, « complete guide to semiconductor devices », McGraw-Hill, Inc 1995.
- D.J. Roulston, « Bipolar semiconductor devices », McGraw-Hill, Inc 1990.

Chapitre 5 : Transistors à Effet de Champ J-FET, MES-FET et MOD-FET

1. Introduction. Les transistors à effet de champ seront étudiés dans ce chapitre pour ce qui concerne les J-FET, MES-FET et MOD-FET et dans un chapitre particulier pour les transistors MOS-FET, constituant une famille particulière compte tenu du domaine qu'ils couvrent.

Comme l'a défini W.Shockley, «l'effet de champ est la modulation d'un canal conducteur par un champ électrique». Dans le cas d'un FET le canal, nous le verrons, est couplé capacitivement par un champ électrique transverse.



**Figure 5.1 : coupe schématique d'un transistor MES-FET
(d'après Sze)**

2. Principe de fonctionnement. Le fonctionnement d'un FET est relativement simple comme d'ailleurs sa structure. La structure de base consiste à l'existence d'un canal conducteur pris en sandwich par deux grilles G et aux bornes duquel ont réalisé deux contacts ohmiques qui seront la source S (au potentiel «0») et le drain D polarisée par une tension V_{DS} . La présence de (des) la grille(s) permet en fait de moduler la largeur du canal. Pour se faire, nous avons deux technologies à notre disposition, qui différencieront le J-FET et le MES-FET. Dans le cas du J-FET (pour Junction Field Effect Transistor), comme son nom l'indique, on réalise une région de type opposée au canal, ce qui crée une jonction PN ou NP et on polarise cette diode en inverse. Comme nous le savons, en inverse, une ZCE large s'étend surtout dans la zone moins dopée (en général le contact Grille est très dopé) c'est à dire le canal, ce qui a pour effet de réduire la largeur du canal conducteur et donc la conductance g du canal. Dans le cas du MES-FET, on remplace la jonction PN par une diode Schottky, qui joue le même rôle. *Une remarque importante est de voir que l'entrée (le signal appliqué sur la grille) est isolée du courant qui circule dans le canal car les jonctions sont polarisées en inverses, ce qui permet d'avoir un gain*

important, qui est le rapport de la puissance de sortie (ou le courant) à la puissance d'entrée (ou le courant). On voit que les dispositifs à effet de champ sont commandés en tension et ont une résistance d'entrée très grande.

3. Le transistor J-FET. Dans le cas du J-FET sur Silicium, le substrat très dopé sert de grille arrière. La couche active (le canal) peut être obtenue soit par implantation ionique, diffusion ou par épitaxie (voir cours M. Ferry). Dans le cas de transistor sur GaAs, le substrat étant isolant, la grille inférieure est absente. Soit la figure 5.1 (SZE p 314) qui représente la structure d'un J-FET à double grille. Définissons les paramètres géométriques du dispositif :

- $2a$ est la largeur (hauteur) maximale du canal, c'est la largeur «métallurgique».
- Z est la profondeur du composant.
- h est la largeur de la ZCE sous la grille dans le canal.
- b est la largeur effective du canal.

D'après les résultats de la jonction P⁺N, la largeur de la ZCE est donnée par :

$$h = \sqrt{\frac{2\epsilon_{SC}}{eN_B}(V_{bi} - V_G)} \quad (1)$$

Ici, N_B est le dopage du canal n donc N_D , V_G la tension de grille qui polarise en inverse la jonction (elle est donc négative) et V_{bi} est la tension de diffusion de la diode. Lorsque la tension de grille est suffisante pour amener la situation $2h = 2a$, on dit que l'on a atteint le point de **pincement** du canal (« Pinch Off »). La tension de grille qui permet d'atteindre le pincement est notée V_T et appelée **tension de seuil** (« threshold voltage »). La tension de pincement, appelée encore **tension interne ou intrinsèque de pincement** est donnée par :

$$V_p = \frac{eN_D a^2}{2\epsilon_{SC}} \quad (2)$$

On en déduit alors à partir de (1) et (2), la tension de grille V_T qu'il faut appliquer pour pincer le canal (en faisant $2h = 2a$) :

$$V_T = V_{bi} - V_p \quad (3)$$

Dans le cas d'un J-FET canal n, si la tension de seuil est positive, en l'absence de toute tension de grille ($V_G = 0V$), le canal est complètement déplété. Le transistor à un mode de fonctionnement à **enrichissement**. Dans le cas contraire, le transistor conduit si la tension de grille est nulle (le canal existe) et le transistor est à **appauvrissement ou déplétion**.

4. Caractéristiques courant – tension. La détermination de l'expression du courant de drain I_D en fonction des tensions V_{DS} et V_{GS} est assez compliquée (Mathieu p340) si on ne fait pas quelques hypothèses simplificatrices :

- *Mobilité des porteurs constante* dans le canal, c'est à dire indépendante du champ électrique. Cette approximation n'est plus valable dans les dispositifs modernes (longueur de canal faible et dans le cas de J-FET et MES-FET à base de AlGaAs).
- *Approximation du canal graduel* : (hypothèse de départ de Schokley) lorsque l'on polarise le drain - source, le champ électrique est essentiellement suivant la direction grille – substrat (V_{GS} étant bien entendu non nulle), ce qui revient à dire que le champ électrique selon Drain – source varie lentement. On en déduit que la largeur de la ZCE en un point x du canal est fonction de la différence de potentiel entre la grille et le potentiel en x dans le canal soit $V(x)$.
- *Approximation de la ZCE* : on considère que dans la ZCE, le nombre de porteurs libres est nul. La densité de porteurs libres hors de la ZCE est égale à la densité de donneurs (canal n) ou des accepteurs (canal p).

Ces hypothèses simplificatrices ne seront plus valables pour expliquer la saturation du courant, mais également dans le cas de dispositifs modernes à géométrie sub microniques et dans le cas de composants à base de AlGaAs.

La résolution du courant de drain passe par l'intégration de l'équation de Poisson à 1 dimension :

$$\frac{d^2V}{dy^2} = -\frac{\rho(y)}{\epsilon_{SC}} = -\frac{eN_D}{\epsilon_{SC}} \quad (4)$$

D'après l'équation (1), la largeur de la ZCE en un point x dans le canal est donnée par :

$$h(x) = \sqrt{\frac{2\epsilon_{SC}}{eN_D}(V_{bi} - V_G(x))} = \sqrt{\frac{2\epsilon_{SC}}{eN_D}(V_{bi} - V_G + V(x))} \quad (5)$$

$V(x)$ est le potentiel dans le canal en x , c'est donc la polarisation de drain en x . on a alors :

- $V(0) = V_S = 0V$
- $V(L) = V_D = V_{DS}$

On obtient le courant de drain (suivant x) en écrivant la loi d'Ohm dans cette direction :

$$J_x = \text{densité de charge} \times \text{mobilité} \times \text{champ électrique} \quad (6a)$$

$$J(x) = -eN_D\mu_n \frac{dV}{dx} \quad (6b)$$

Le courant dans le canal I_D (courant de drain) est compté positivement du drain vers la source. En multipliant $-J(x)$ par la section conductrice du canal $2Z(a-h(x))$ on obtient :

$$I_D = 2Z[a-h(x)]eN_D\mu_n \frac{dV}{dx} \quad (7)$$

En remplaçant $h(x)$ par son expression (eq. 5), et en intégrant sur x , on obtient :

$$I_D \int_0^L dx = e\mu_n N_D 2Z \int_0^{V_p} \left[a - \left\{ \frac{2\varepsilon_{SC} [V(x) + V_{bi} - V_G]}{eN_D} \right\}^{1/2} \right] dV \quad (8)$$

Soit encore :

$$I_D = \frac{2Z}{L} e\mu_n N_D a \left\{ V_{DS} - \frac{2[(V_{DS} + V_{bi} - V_{GS})^{3/2} - (V_{bi} - V_{GS})]}{3(eN_D a^2 / 2\varepsilon)^{1/2}} \right\} \quad (9)$$

La conductance maximum du canal est notée $G_0 = \frac{2Z}{L} e\mu_n N_D a$, ce qui permet de réécrire le courant de drain (en injectant la tension interne de pincement V_p) :

$$I_D = G_0 \left\{ V_{DS} - \frac{2[(V_{DS} + V_{bi} - V_{GS})^{3/2} - (V_{bi} - V_{GS})]}{3V_p^{1/2}} \right\} \quad (10)$$

*Cette expression est exclusivement valable dans le régime **non pincé** du transistor.*

Le régime de saturation du transistor doit faire l'objet d'un traitement particulier. Lorsque les polarisations de grille et de drain sont telles que les ZCE occupent la totalité du canal conducteur, on dit qu'il y a pincement du canal et d'après le raisonnement précédent, le courant devrait «chuter» à zéro, la conductance du canal étant égale à zéro. La réalité est plus

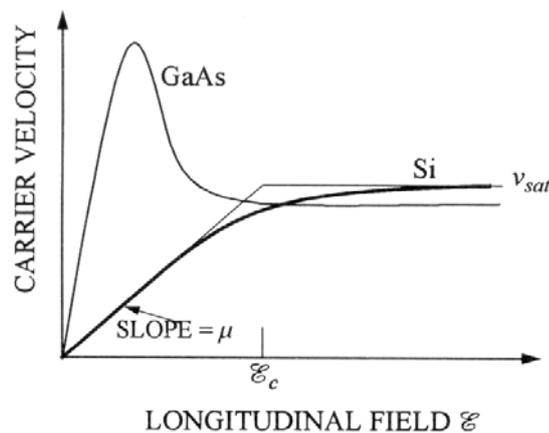


Figure 5.2 : variation de la vitesse des porteurs avec le champ électrique

complexe et l'on sait qu'au delà de la saturation, le courant atteint une limite et « sature ». Pourquoi ? Au départ, le point de saturation se localise coté drain, et ce n'est qu'en ce point que la conductance du canal est nulle et donc le courant. Par contre ailleurs elle est non nulle et le courant aussi. Si bien que, comme il doit y avoir continuité du courant, d'après (6a), la mobilité (la vitesse) doit tendre vers l'infini au point de pincement, ce qui est physiquement déraisonnable. On sait, toujours d'après (6a) que le courant en x est donné par

$I = 2Zen(x)v(x)(a - h(x))$. Donc la solution pour maintenir le courant constant lorsque $h(x)$ tend vers a est d'augmenter $v(x)$ ou $n(x)$. La vitesse $v(x)$ ne peut augmenter (figure 5.2 ou p 185 Ng) au delà de la vitesse de saturation (pour le Si) ou de la valeur pic (pour le GaAs), si bien que l'on doit créer des porteurs au niveau du point de pincement pour augmenter notablement $n(x)$. Ceci a pour effet de créer une couche d'accumulation et donc de repousser la ZCE et donc d'ouvrir le canal, le courant pouvant de nouveau passer ! Différentes

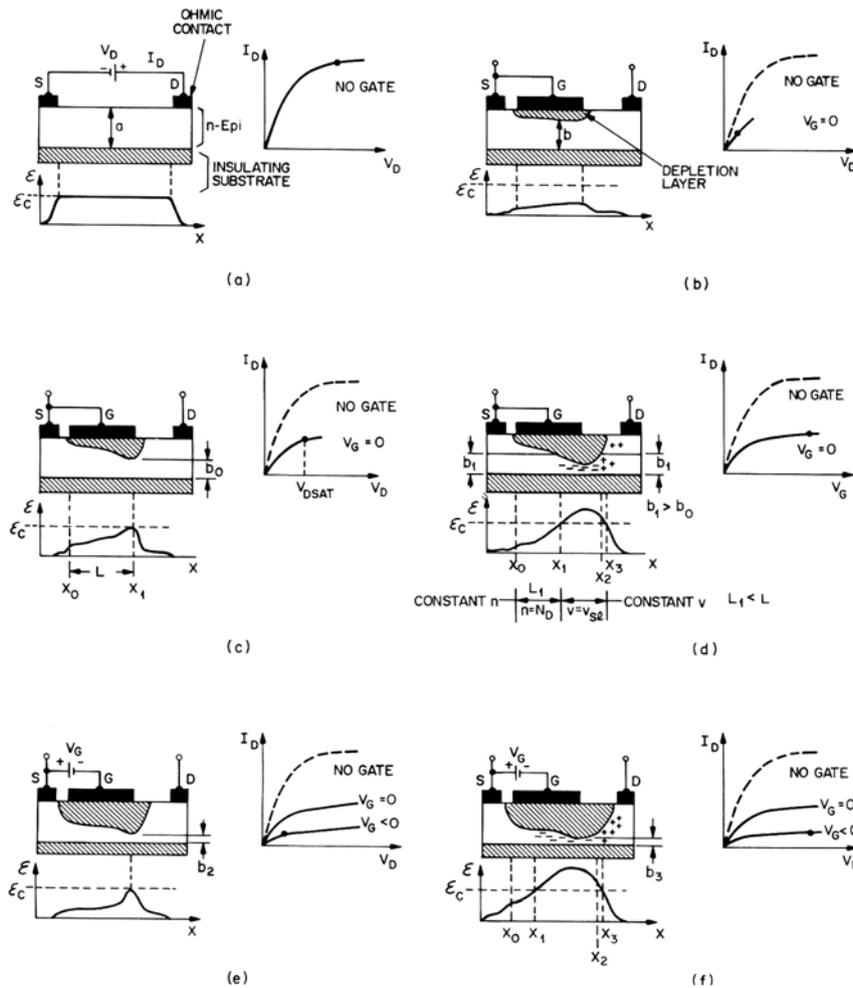


Figure 5.3: Régimes de saturation, principes physiques (Sze)

situations sont représentées sur la figure 5.3 (p 336 SZE). Une fois le phénomène physique admis, on conçoit que le courant sature donc à sa valeur I_{Dsat} qui est donnée, en considérant que la tension aux bornes du canal conducteur est égale à V_{Dsat} , par :

$$I_{Dsat} = G_0 \left[\frac{V_p}{3} - V_{bi} + V_G + \frac{2(V_{bi} - V_G)^{3/2}}{3V_p^{1/2}} \right] \quad (11)$$

avec

$$V_{Dsat} = V_p - V_{bi} + V_G = V_G - V_T \quad (12)$$

l'équation (12) vient de : $V_{GS} = V_G - V(x = \text{pincement}) + V(x = \text{pincement}) - V_S = V_T + V_{Dsat}$.

La transconductance du dispositif, qui définit le contrôle du courant de drain par la tension de grille est donnée par :

$$g_m = \left. \frac{\partial I_D}{\partial V_G} \right|_{V_D} = G_0 \frac{(V_D + V_{bi} - V_G)^{1/2} - (V_{bi} - V_G)^{1/2}}{V_p^{1/2}} \quad (13)$$

On peut donc améliorer la transconductance en utilisant des substrats à forte mobilité ou/et en diminuant la longueur du canal.

Dans le cas où la tension drain source est petite devant $V_{bi} - V_G$, on parle de fonctionnement en régime linéaire et l'expression (10) se réduit (en développant en série de Taylor) :

$$I_D = G_0 \left[1 - \frac{(V_{bi} - V_G)^{1/2}}{V_p^{1/2}} \right] V_D \quad (14)$$

et la transconductance en régime linéaire devient :

$$g_{m,lin} = \frac{G_0 V_D}{2V_p^{1/2} (V_{bi} - V_G)^{1/2}} \quad (15)$$

En saturation, à partir de l'expression (11) du courant de drain, on obtient :

$$g_{msat} = G_0 \left[1 - \frac{(V_{bi} - V_G)^{1/2}}{V_p^{1/2}} \right] \quad (16)$$

On peut remarquer que la transconductance en saturation est égale à la conductance en régime linéaire.

5. Régime de fonctionnement HF – Fréquence de coupure. L'étude du comportement en fréquence des FETs se fera sur le MES-FET pour plus de facilité mais les résultats peuvent être étendus sans restrictions aux J-FET (*en faisant attention à la différence de largeur de canal $2a$ ou a*). La course à la HF et micro-onde implique l'utilisation de transistors J-FET et MES-FET à canal n de part la plus grande mobilité des électrons par rapport à celle des trous. Le schéma en petits signaux d'un transistor MES-FET est donné sur la figure 5.4 (p385 singh). Le modèle intrinsèque intègre les capacités grille – canal (C_{GS} et C_{DG}), les résistances de canal (R_i et $R_{DS} = g_D^{-1}$) et la capacité drain canal C_{DC} . A ce modèle, on doit ajouter les paramètres «extrinsèques» (ou parasites) que sont la résistance de grille R_G , la résistance de source et de drain R_S et R_D et la capacité de couplage drain substrat C_{DS} .

Si on applique une variation de charge (par variation de V_G) ΔQ sur la grille, par principe de neutralité électrique, on va retrouver cette charge ΔQ dans le canal. Si on pose Δt , le temps que mettent les électrons pour « répondre » à cette variation de charge, la variation de courant ΔI_D résultant est donnée par :

$$\Delta I_D = \frac{\Delta Q}{\Delta t} \quad (17)$$

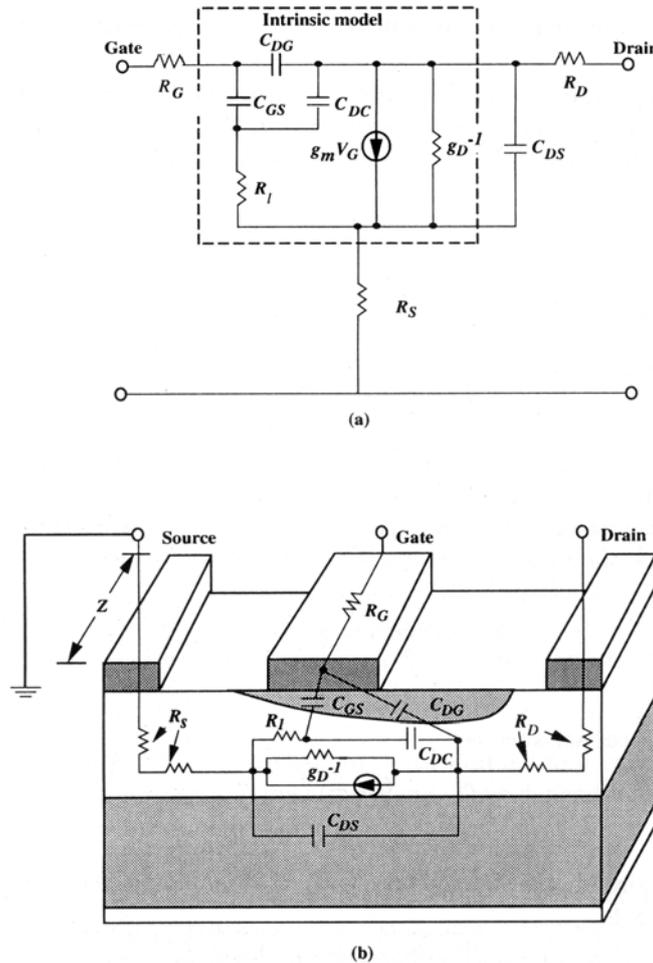


Figure 5.3 : schéma équivalent HF du FET (Singh)

Le temps Δt est bien entendu fonction de la longueur du canal et peut être vu comme le temps de transit des électrons à travers le dispositif. Suivant que le dispositif «travaille» dans le modèle « mobilité constante » ou « régime de saturation de vitesse », ce temps de

transit s'écrit $\tau_r = \frac{L^2}{\mu V_{DS}}$ ou $\tau_r = \frac{L}{v_{sat}}$ respectivement, avec L la longueur du canal, V_{DS}

la tension drain source et μ la mobilité des porteurs dans le canal. La fréquence de coupure, comme dans le cas du bipolaire correspond à un gain en courant unitaire, c'est à dire $I_{in} = I_{out}$. A partir de (17), on obtient facilement la transconductance g_m en fonction

de la capacité de grille C_{GS} et de τ_r , à savoir $g_m = \frac{C_{GS}}{\tau_r}$. La fréquence de coupure est

alors donnée par :

$$f_T = \frac{g_m}{2\pi C_{GS}} = \frac{1}{2\pi \tau_r} \quad (18)$$

Souvent, compte tenu de la réduction de la taille des composants, le modèle susceptible de « coller » à la réalité pour le calcul de la fréquence de coupure est « à saturation de vitesse » et f_T s'écrit :

$$f_T = \frac{v_{sat}}{2\pi L} \quad (19)$$

dans le cas contraire, à partir de (18), en prenant pour la transconductance, la valeur maximum (eq 16 avec $V_G = V_{bi}$) et pour C_G son expression classique $\frac{\epsilon WL}{a}$, on obtient pour $f_{T(max)}$ l'expression suivante :

$$f_{T(max)} = \frac{e\mu_n N_D a^2}{2\pi\epsilon L^2} \quad (20)$$

L'équation (20) ne tient pas compte des capacités parasites.

Un effet indésirable de la résistance série de source intervient dans le calcul de la transconductance g_m . Cette résistance série introduit une chute ohmique qui modifie la polarisation grille - source V_{GS} , si bien que l'on peut montrer que la transconductance devient $g'_m = \frac{g_m}{1 + g_m R_S}$. La valeur de la résistance de drain n'intervient pas car en saturation le courant est indépendant de la tension de drain V_{DS} .

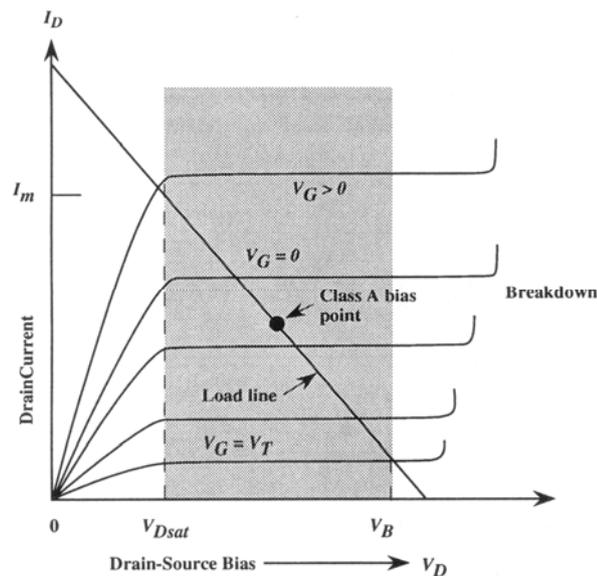


Figure 5.4 : point de fonctionnement en ampli classe A (Singh)

- 6. Applications forts signaux.** La principale application des transistors à effet de champ est le domaine des transistors de puissance. Dans cette application le transistor fonctionne dans le régime actif et saturé, soit le domaine des amplificateurs *classe A*. Si on suppose le point de fonctionnement Q centré (figure 5.4), avec un signal de commande sinusoïdal, la puissance de sortie maximum en régime dynamique est donnée par

$$P_{o\max} = V_{eff} \cdot I_{eff} = \frac{I_Q V_Q}{2}. \text{ Ce qui d'après la figure s'écrit :}$$

$$P_{\text{sortie}} = \frac{(I_{D1} - I_{D2}) \cdot (V_B - V_1)}{8} \quad (21)$$

V_B est la tension de claquage de la jonction grille – drain. Si l'on veut augmenter la puissance en sortie, on doit augmenter la tension de claquage, faire tendre I_{D2} et $V_1 \approx V_{Dsat}$ vers zéro. Le problème ici est que si on augmente V_B (en augmentant la longueur du canal L) on dégrade la réponse en fréquence (Eq. 20).

*On doit donc trouver un **compromis** entre la vitesse de fonctionnement du composant et la puissance admissible.*

7. Transistors à hétérostructure : MOD-FET, HEMT ou TEG-FET. L'intérêt croissant dans le domaine des hautes fréquences conduit à rechercher des composants avec une fréquence de coupure la plus élevée possible. Une des solutions est d'utiliser des matériaux à très forte mobilité et à dopage élevé (voit Eq. 20). La présence de diffusion par les impuretés ionisées limite la possibilité d'augmenter le dopage. L'idée (Esaki *et al*) de séparer spatialement les impuretés dopantes et les électrons provenant de ces impuretés a conduit à l'élaboration d'hétérostructures semi-conductrices comme montrées sur la figure (8.10 singh). La structure est composée de semi-conducteurs à « gaps » différents (d'où le nom d' hétérostructures) et à dopage modulés (différents) d'où le nom de MOD – FET. Lorsque la structure est réalisée, les électrons provenant des donneurs « voient » des niveaux énergétiques plus faibles dans le semi-conducteur à « petit gap » et viennent se localiser dans le puits formé. Pour améliorer la séparation spatiale on intercale une couche non dopée (« spacer ») du semi-conducteur à « grand gap ». Le confinement dans le puits de ces électrons modifie leur faculté de déplacement. On montre en fait que le déplacement suivant la direction z est interdite. Ils ne peuvent donc se déplacer que suivant les directions x et y , soit dans un plan parallèle à la structure. Ils forment un gaz bi – dimensionnel (TEG – FET) qui constituera lui-même le canal. Pour plus de détails, consulter les références en fin de chapitre.

Pour réaliser un transistor à effet de champ, on doit réaliser un contact de grille qui va comme dans le cas des J – FET et MES – FET contrôler la densité dans le canal. Cette grille constitue un contact Schottky avec une hauteur de barrière ϕ_b . Définissons les différents paramètres du dispositif (figure 7.21 p 383 mathieu) :

- d est la largeur du semi-conducteur « grand gap »
- d_s est la largeur du « spacer »
- d_d est la largeur du semi-conducteur « grand gap » dopé
- $e\phi_b$ la hauteur de la barrière Schottky
- ΔE_c la discontinuité des bandes de conduction
- $V_2(z)$ courbure de potentiel dans la zone 2 «grand gap» (la barrière)

Soit l'équation de Poisson dans la barrière :

$$\frac{\partial^2 V_2(z)}{\partial z^2} = - \frac{eN(z)}{\epsilon_b} \quad (22)$$

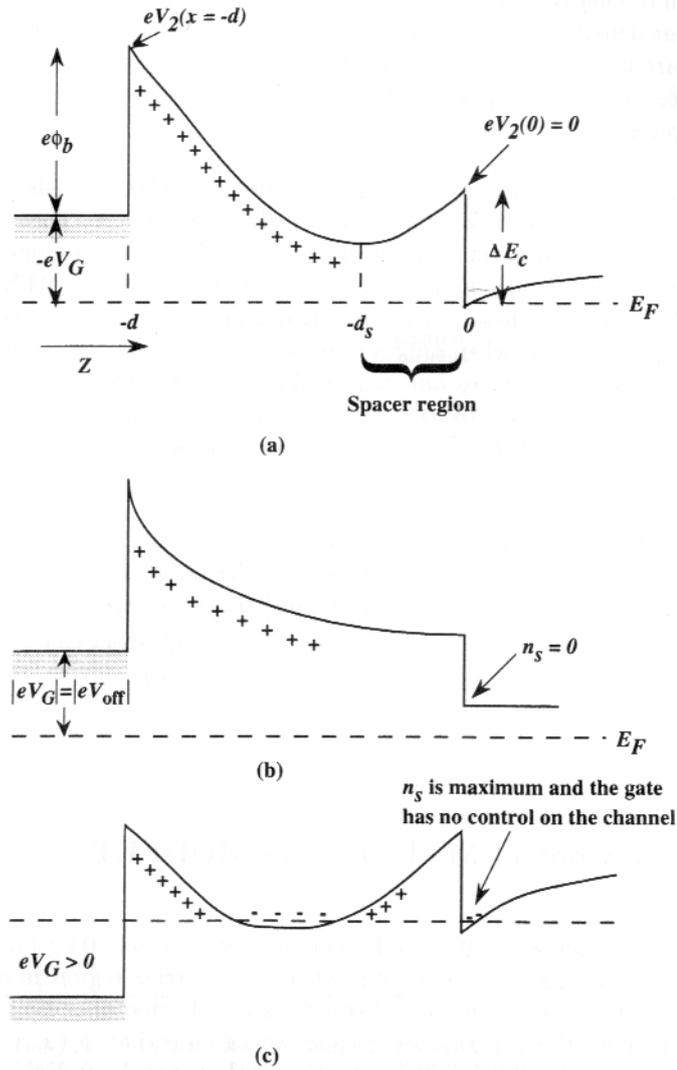


Figure 5.6 : structure MOD-FET (Singh)

avec $N(z)$ la densité de charge dans la barrière. On suppose la barrière complètement déplétée, *i.e.* $N(z < d_s) = N_D$. A partir de l'équation (22) et en intégrant dans le «spacer» on a $\frac{dV_2}{dz} = -E_2(0)$, où $E_2(0)$ est le champ électrique continu dans le «spacer». Si maintenant on intègre (22) dans la barrière dopée, on obtient (en intégrant 2 fois) :

$$V_2(z) = -\frac{eN_D}{2\epsilon_b}(z + d_s)^2 - E_2(0)z + cte \quad (23)$$

Si on calcule maintenant, la différence de potentiel dans la barrière (de $-d$ à 0), on obtient :

$$V_2(-d) - V_2(0) = -\frac{eN_D}{2\epsilon_b}(-d + d_s)^2 + E_2(0)d \quad (24)$$

A partir de la figure, on obtient facilement :

$$\Delta V = -\frac{E_F}{e} + V_g - \frac{\phi_b}{e} + \frac{\Delta E_C}{e} \quad (25)$$

Attention, ici E_F représente la dégénérescence du puits du semi-conducteur « petit gap », c'est à dire la différence entre le bas de la bande de conduction et le niveau de Fermi.

A partir de la loi de Gauss appliquée dans le puits, on obtient :

$$n_s = \frac{\varepsilon_p E_1(0)}{e} \quad (26)$$

A partir de (24) et (25), on peut écrire :

$$E_2(0) = \frac{1}{d} \left(V_g - \frac{E_F}{e} - V_t \right) \quad (27)$$

avec

$$V_t = \frac{\phi_b}{e} - \frac{\Delta E_C}{e} - \frac{eN_D}{2\varepsilon_B} d_d^2 \quad (28)$$

En écrivant la continuité du vecteur déplacement en $z=0$, et à partir de (26), on obtient finalement:

$$n_s = \frac{\varepsilon_b}{ed} \left(V_g - \frac{E_F}{e} - V_t \right) \quad (29)$$

La densité de porteurs dans le puits est nulle lorsque la tension V_g appliquée sur la grille est égale à $V_t + \frac{E_F}{e} = V_{off}$, avec $V_{off} = \phi_b - \frac{\Delta E_C}{e} + \frac{E_F}{e} - V_{P_2}$ et $V_{P_2} = \frac{eN_D}{2\varepsilon_b} d_d^2$. Les équations (27) et (29) s'écrivent respectivement :

$$E_2(0) = \frac{1}{d} (V_g - V_{off}) \quad (30)$$

$$n_s = \frac{\varepsilon_b}{ed} (V_g - V_{off}) \quad (31)$$

V_{off} est la tension de seuil du transistor. Suivant le type de métal utilisé pour la barrière Schottky et le produit dopage x épaisseur, on pourra comme dans le cas d'un MES – FET, obtenir un MOD – FET à déplétion ou à enrichissement.

Le calcul du courant en fonction de V_g et V_D est relativement compliqué du fait des effets de transports quantiques dans le puits triangulaire qui forme le canal. Un calcul (Mathieu p 390) montre que le courant de drain peut se mettre sous la forme :

$$I_D = \frac{We\mu_n\beta}{L} \left((V_g - V_t)V_D - \frac{V_D^2}{2} \right) \quad (32)$$

avec $\beta = \frac{2\varepsilon_b em_e}{2de^2 m_e + \varepsilon_b \pi \hbar^2}$.

Comme dans le cas du MES – FET, la saturation du courant est du à la saturation de la vitesse des électrons à la vitesse v_s . On montre que dans le cas d'un canal long (Mathieu p 391) le courant de saturation s'écrit :

$$I_{dsat} \approx \frac{g_0 (V_g - V_t)^2}{2V_s} \quad \text{avec } V_s = LE_C \quad (33a)$$

$$V_{dsat} \approx V_g - V_T \quad (33b)$$

V_s est la tension drain source qui établit le champ critique sur toute la longueur du canal.

En revanche dans le cas d'un canal court, V_s est petite et les expressions en saturations se réécrivent :

$$I_{dsat} \approx g_0(V_g - V_T) \quad (34a)$$

$$V_{dsat} \approx V_s \quad (34b)$$

Un effet important dans le cas où le semi-conducteur à grand gap n'est pas complètement déplété est l'apparition d'un MES – FET parasite. Le schéma équivalent est formé alors d'un HEMT en parallèle avec un MES – FET parasite.

8. Références

- S.M. Sze « Physics of semiconductors devices », 2° édition, Wiley, New York, 1981
- H.Mathieu, « Physique des semi-conducteurs et des composants électroniques », 4° édition, Masson 1998.
- J. Singh, « semiconductors devices : an introduction », McGraw-Hill, Inc 1994.
- Y.Taur et T.H. Ning, « Fundamentals of Modern VLSI devices », Cambridge University Press, 1998.
- K.K. Ng, « complete guide to semiconductor devices », McGraw-Hill, Inc
- F. Ali et A. Gupta, Eds., « HEMTs & HBTs : devices, fabrication, and circuits », Artech House, Boston, 1991.

Chapitre 6 : Structure MOS Transistor MOS-FET

Introduction. Nous étudierons dans cet important chapitre, en premier lieu la capacité MOS, qui est la base du transistor MOS – FET, et de la technologie CMOS. La difficulté d’obtenir des oxydes de bonne qualité a fait prendre du retard à la technologie MOS par rapport aux bipolaires, mais une fois cet obstacle levé, la simplicité de la structure, de la technologie et donc sa plus forte compétitivité a conduit à un développement important de cette filière.

Structure Métal – Oxyde - Semi-conducteur. Les structures MOS sont essentiellement basées sur la technologie Silicium de par la «facilité» de réaliser un oxyde de bonne qualité par simple oxydation du Silicium de façon à former une couche isolante de SiO₂. La structure typique d’une capacité MOS est représentée sur la figure 6.1.a. Le substrat est donc su Si, sur lequel on a fait pousser une fine couche d’oxyde. Un dépôt métallique vient former le contact électrique.. La figure 6.1.b donne un ordre de grandeur des différents paramètres caractéristiques de la structure. Le travail de sortie du Si est donné par :

$$\phi_{SC} = \chi + \frac{E_g}{2e} + \phi_{Fi} \quad (1)$$

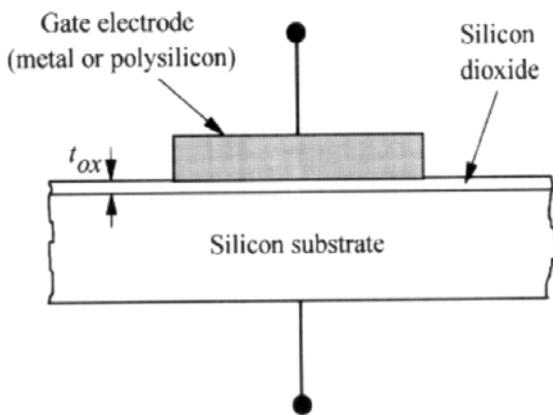


Figure 6.1.a : capacité MOS

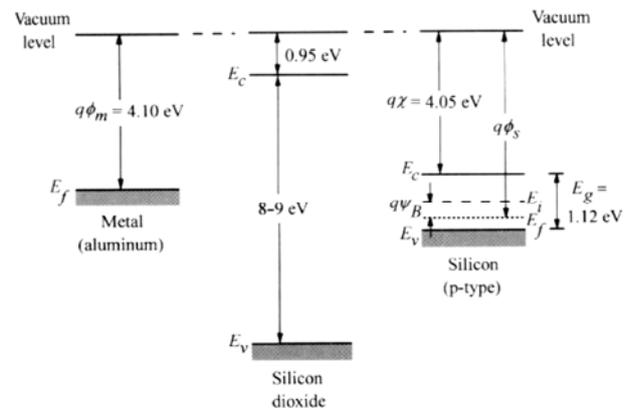
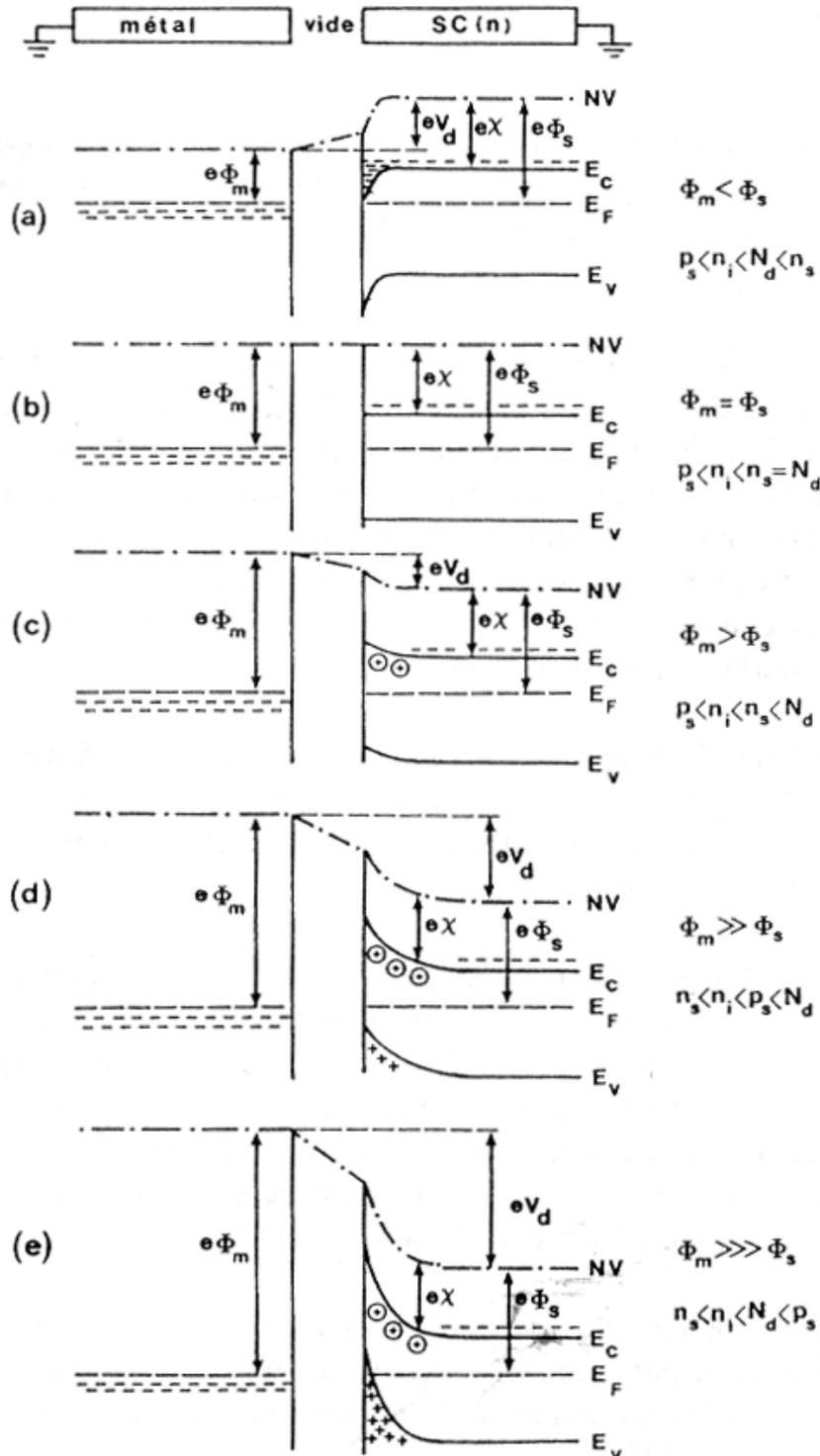


Figure 6.1.b : les différents paramètres énergétiques

On voit donc que si ϕ_{SC} est différent de ϕ_m , les niveaux de vide ne seront pas alignés, un transfert de charge aura lieu et une courbure de bande s’établira. Posons, par analogie avec la jonction PN, $NV_m - NV_{SC} = eV_{bi}$ et $V_{bi} = \phi_m - \phi_{SC}$. Relions les deux contacts au même potentiel (par exemple la masse). Lorsque la structure n’est pas polarisée, la structure est à l’équilibre thermodynamique et les niveaux de Fermi (métal et Si) sont alignés. Le champ électrique à l’interface est donné par le gradient du potentiel $E = -\frac{dV}{dx}$ et est d’autant plus élevé que l’épaisseur d’oxyde d_{ox} est faible (à V_{bi} constant). On peut donc dire que si d_{ox} est grand, le champ électrique sera faible, donc le transfert de charge également et donc la courbure de bande également.

Si maintenant, on «aminci » l'oxyde, que se passe-t-il ? La réponse est fonction du signe de V_{bi} .

Toutes les situations sont reportées sur la figure 6.2 cidessous (Mathieu).



- $\phi_m > \phi_{SC}$: la tension $V_{bi} > 0$ et le champ électrique < 0 . Ce champ provoque une accumulation de trous (porteurs majoritaires) à l'interface oxyde – SC et par influence électrostatique une accumulation d'électrons dans le métal à l'interface métal – SC. Compte tenu de la différence de densité électronique, la courbure de

bande n'affecte que le SC, la bande de valence se courbant vers le haut et se rapprochant du niveau de Fermi, rendant bien compte d'une augmentation locale de la densité de trous. **On parle de régime d'accumulation.**

- $\phi_m = \phi_{SC}$: la tension étant nulle, il n'y a pas de transfert de charges, les bandes restent plates, on dit que l'on est en situation de **bandes plates ou «flat band»**.
- $\phi_m < \phi_{SC}$: dans ce cas, des charges négatives se développent dans le semi-conducteur et en corollaire des charges positives dans le métal. Les charges positives dans le métal résultent de l'évacuation des électrons à la surface et les charges négatives dans le métal, en partie par les accepteurs ionisés qui ne sont plus compensés par les trous qui sont repoussés de l'interface (influence électrostatique) et par l'apparition d'électrons pour maintenir le produit $n.p = cte$ au voisinage de l'interface. Dans ce cas, la courbure de bandes se produit vers le bas, rapprochant la bande de conduction du niveau de Fermi. On doit ici envisager 3 cas, suivant l'importance de cette courbure. Introduisons pour cela le potentiel de surface ϕ_s qui est la courbure de potentiel à la surface du silicium. Tant que ϕ_s est inférieur à ϕ_{Fi} , on parle de régime de déplétion, la surface du Si «étant dépeuplée de trous. Dans le cas où $\phi_{Fi} < \phi_s < 2\phi_{Fi}$, à la surface le niveau de Fermi intrinsèque coupe le niveau de Fermi. La bande de conduction se retrouve alors plus proche du niveau de Fermi que la bande de valence, on a donc localement inversé le type de population. On parle de **régime de faible inversion ou «weak inversion»**. Si on poursuit le raisonnement, on atteint alors le régime de forte inversion dans le cas où $\phi_s > 2\phi_{Fi}$, on se retrouve alors en **régime de forte inversion ou «strong inversion»**.

On va s'intéresser dans ce qui suit à une capacité MOS «idéale» sur substrat de type p. Qu'entend t on par structure idéale ? Cette notion est importante et conduit à une simplification du problème. **On appelle structure MOS idéale une structure qui non polarisée est en situation de bandes plates («flat band»), avec un oxyde parfait (isolant parfait sans charge) et absence d'états d'interfaces.** On reviendra plus en détails sur ces deux paramètres plus tard. On a donc $\phi_m = \phi_{SC}$. On peut voir alors sur la figure 6.3 qu'en polarisant la structure avec une tension de grille V_g , on peut retrouver les mêmes situations que précédemment, à savoir :

- $V_g = 0$ régime de bandes plates
- $V_g < 0$ régime d'accumulation
- $V_g > 0$ régime de déplétion ou d'inversion suivant la valeur de V_g .

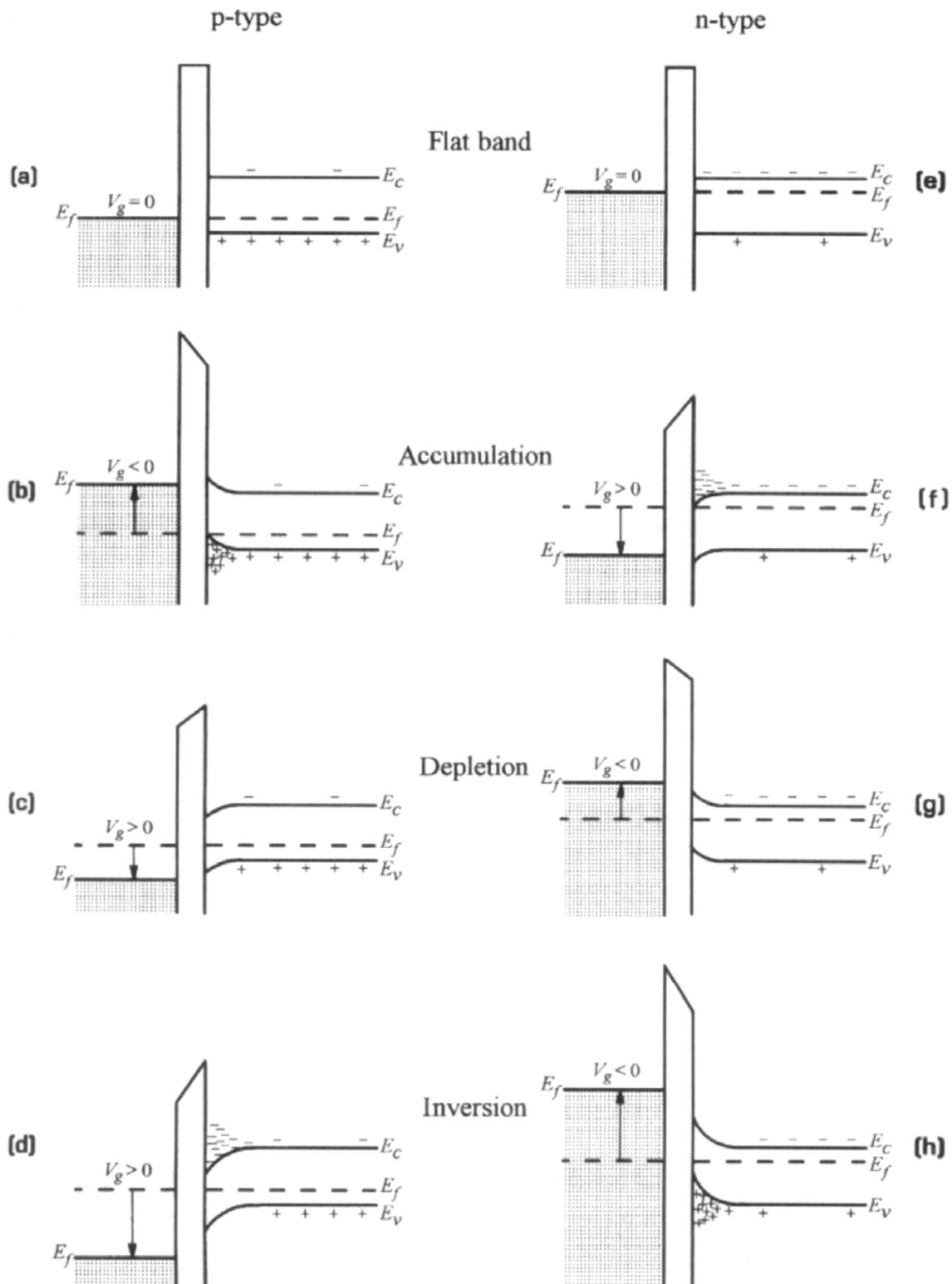


Figure 6.2 : les différents régimes d'une structure MOS

Dans le cas d'une capacité MOS sur substrat n, on retrouve les mêmes situations mais avec les polarités des tensions inversées et les différences ϕ_m, ϕ_{SC} également.

Remarquer que dans ce cas, $\phi_{SC} = \chi + \frac{E_g}{2e} - \phi_{Fi}$.

Champ, potentiel et charge dans le silicium. Il suffit, comme d'habitude maintenant, de résoudre l'équation de Poisson. Rappelons que $e\phi_{Fi} = |E_F - E_{Fi}| > 0$. On supposera le semi-conducteur au potentiel zéro (soit $V(x = \infty) = 0$) et $V(x = 0) = V_s$ et la grille polarisée par V_g .

On rappelle que l'on se place dans le cas d'un semi-conducteur de type p. La densité de charge dans le Si s'écrit :

$$\rho(x) = e[p(x) - n(x) + N_D^+(x) - N_A^-(x)] \quad (2)$$

Dans la région neutre la densité de charges est nulle, ce qui revient à écrire :

$$p_0 - n_0 = N_A^- - N_D^+ \quad (2a)$$

La densité de trous et d'électrons dans la région neutre s'écrivent :

$$n_0 = n_i \exp\left(-\frac{e\phi_{Fi}}{kT}\right) \quad \text{et} \quad p_0 = n_i \exp\left(\frac{e\phi_{Fi}}{kT}\right) \quad (3)$$

On peut alors exprimer $n(x)$ et $p(x)$ en fonction de n_i :

$$n(x) = n_0 \exp\left(\frac{eV(x)}{kT}\right) = n_i \exp\left(\frac{e(V(x) - \phi_{Fi})}{kT}\right) \quad (4a)$$

$$p(x) = p_0 \exp\left(-\frac{eV(x)}{kT}\right) = n_i \exp\left(-\frac{e(V(x) - \phi_{Fi})}{kT}\right) \quad (4b)$$

On suppose que le dopage est homogène et tous les dopants ionisés ($T=300K$), ce qui permet de réécrire (2) en utilisant (2a) :

$$\rho(x) = e \left[n_0 - p_0 + p_0 e^{-\frac{eV(x)}{kT}} - n_0 e^{+\frac{eV(x)}{kT}} \right] \quad (5)$$

L'équation de Poisson s'écrit alors :

$$\frac{d^2V(x)}{dx^2} = -\frac{e}{\epsilon_{sc}} \left[p_0 \left(e^{-\frac{eV(x)}{kT}} - 1 \right) - n_0 \left(e^{+\frac{eV(x)}{kT}} - 1 \right) \right] \quad (6)$$

En remarquant que $\frac{d^2V(x)}{dx^2} = \frac{d}{dx} \left(\frac{dV(x)}{dx} \right) = \frac{d}{dV} \left(\frac{dV(x)}{dx} \right) \frac{dV(x)}{dx}$, on peut réécrire l'équation de Poisson (6) sous la forme :

$$\frac{dV(x)}{dx} d \left(\frac{dV(x)}{dx} \right) = -\frac{e}{\epsilon_{sc}} \left[p_0 \left(e^{-\frac{eV(x)}{kT}} - 1 \right) - n_0 \left(e^{+\frac{eV(x)}{kT}} - 1 \right) \right] dV(x) \quad (7)$$

L'équation (7) est la forme classique de l'équation de Poisson dans une structure MOS. Il suffit maintenant d'intégrer cette équation depuis le volume (« bulk ») du semi-conducteur où $V(x)=0$ et $\left. \frac{dV(x)}{dx} \right|_0 = 0$ vers la surface, et l'on obtient :

$$\int_0^{dV(x)/dx} \frac{dV(x)}{dx} d\left(\frac{dV(x)}{dx}\right) = -\frac{e}{\epsilon_{SC}} \int_0^{V(x)} \left[p_0 \left(e^{-\frac{eV(x)}{kT}} - 1 \right) - n_0 \left(e^{+\frac{eV(x)}{kT}} - 1 \right) \right] dV(x) \quad (8)$$

soit encore en se rappelant que le champ électrique $E(x) = -\frac{dV(x)}{dx}$,

$$E^2(x) = \left(\frac{dV(x)}{dx} \right)^2 = \frac{2kTp_0}{\epsilon_{SC}} \left[\left(e^{-eV(x)/kT} + \frac{eV(x)}{kT} - 1 \right) + \frac{n_0}{p_0} \left(e^{eV(x)/kT} - \frac{eV(x)}{kT} - 1 \right) \right] \quad (9)$$

On introduit souvent la **longueur de Debye** qui s'écrit $L_D = \sqrt{\frac{\epsilon_{SC} kT}{e^2 p_0}}$. Physiquement, sur cette longueur, on peut considérer que le semi-conducteur n'est pas neutre électriquement, ceci lorsque il y a dans cette région une forte variation de dopage.

Une étude très détaillée sur la signification réelle de cette longueur de Debye est donnée dans **Mathieu** p 89 et p115 (attention à son expression !!) et dans une moindre mesure dans **Taur** p26. Dans ces conditions, l'équation (9) peut se mettre sous la forme :

$$E^2(x) = \left(\frac{dV(x)}{dx} \right)^2 = \left(\frac{kT}{e} \right)^2 \frac{2}{L_D^2} \left[\left(e^{-eV(x)/kT} + \frac{eV(x)}{kT} - 1 \right) + \frac{n_0}{p_0} \left(e^{eV(x)/kT} - \frac{eV(x)}{kT} - 1 \right) \right] \quad (10)$$

En $x=0$, le champ électrique est le champ de surface E_S et le potentiel est appelé potentiel de surface V_S . A partir du théorème de Gauss (on l'obtient en intégrant l'équation de Poisson) qui lie le champ électrique et la charge **totale** par unité de surface contenu dans le semi-conducteur, on obtient :

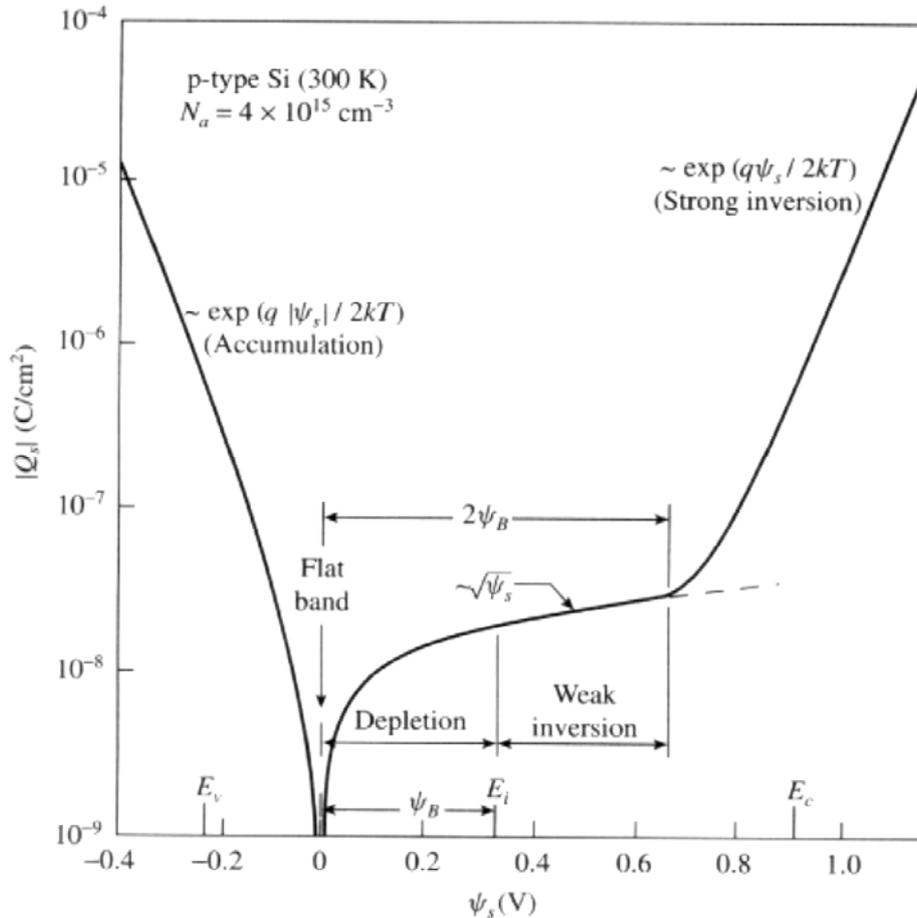
$$E_S = -\frac{Q_{SC}}{\epsilon_{SC}} \quad (11)$$

On peut alors calculer la charge dans le semi-conducteur en fonction de (10) et l'on obtient :

$$Q_{SC} = \pm \left(\frac{kT}{e} \right) \frac{\epsilon_{SC} \sqrt{2}}{L_D} \left[e^{-eV_S/kT} + \frac{eV_S}{kT} - 1 + \frac{n_0}{p_0} \left(e^{eV_S/kT} - \frac{eV_S}{kT} - 1 \right) \right]^{1/2} = -Q_{\text{métal}} \quad (12)$$

Compte tenu de l'orientation du champ électrique et de l'origine des potentiels dans le volume, le \pm dans l'expression (12) peut être remplacé par $-sign(V_S)$ qui dans une structure MOS idéale (tension de bandes plates = 0) peut être remplacé par $-sign(V_g)$. Sur la figure suivante (2.25 p65 Taur) est représentée la variation de la charge en fonction du potentiel de surface V_S .

On remarque en premier lieu que pour V_S nul, la charge est bien nulle, on est bien en situation de *bandes plates*. Ce qui est accord avec l'hypothèse de structure MOS idéale. Pour des V_S



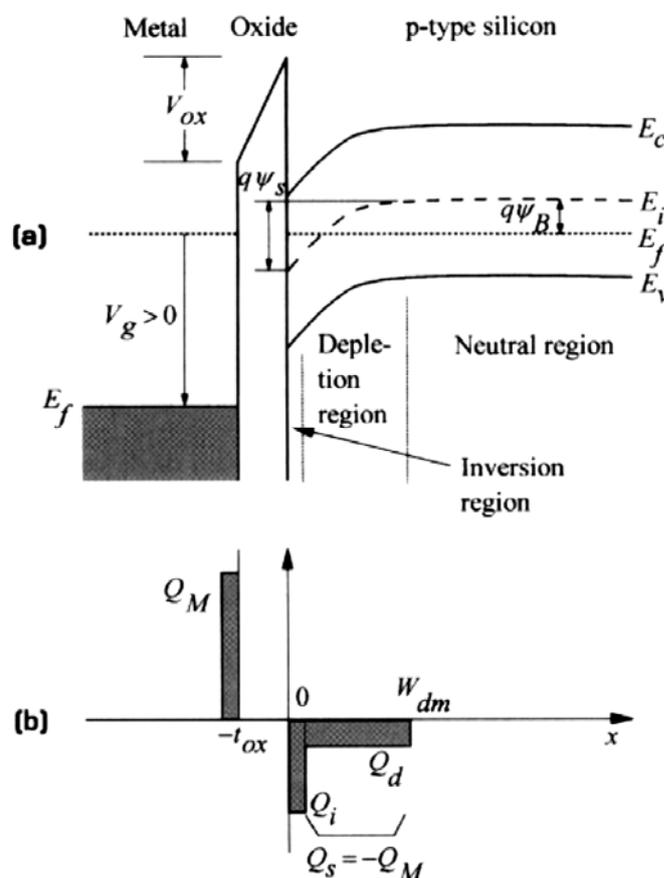
Chapitre 4 Figure 6.3 : variation de la charge en fonction du potentiel de surface dans une structure MOS.

négatifs (donc des V_g négatifs), la charge dans le métal est négative et est donc positive dans le semi-conducteur. On est en régime d'accumulation. Une fois que $eV_s/kT > 1$, c'est le terme en exponentielle négative qui devient prédominant et la variation de charges est alors exponentielle $e^{-eV_s/2kT}$. La charge est due aux trous qui viennent s'accumuler à l'interface, ce sont des porteurs libres et la largeur de la zone d'accumulation est donc faible (qq 10 \AA). Pour des tensions positives et $eV_s/kT > 1$, mais avec $e^{eV_s/kT}$ pas suffisamment important pour rendre le terme n_0/p_0 appréciable, la charge, qui est une charge essentiellement de déplétion, varie en $\sqrt{V_s}$ comme indiqué sur la courbe. Lorsque la tension augmente encore, le terme $e^{eV_s/kT}$ domine $\sqrt{V_s}$, ce qui arrive lorsque l'on atteint le régime de forte inversion où la charge est due aux électrons accumulés à l'interface Si-SiO₂. La charge d'inversion est donc proportionnelle à $e^{eV_s/2kT}$. Un critère (qui n'est pas unique) très répandu pour fixer le seuil de

forte inversion est l'égalité entre la densité de porteurs à la surface et la densité de porteurs dans le volume soit ici $n_s = p_0 = N_A$. Un calcul simple montre qu'alors :

$$V_s = 2\phi_{Fi} = \frac{2kT}{e} \ln\left(\frac{N_A}{n_i}\right) \quad (13)$$

On peut remarquer que compte tenu de la variation de la charge d'inversion en exponentielle, une fois cette couche d'inversion créée, toute augmentation de la polarisation, c'est à dire de charges côté métal sera compensée par une augmentation de la densité d'électrons dans cette couche, la couche de déplétion restant pratiquement constante (elle varie en $\sqrt{V_s}$).



Chapitre 5 Figure 6.4 : bandes d'énergie et densité de charge associée

Capacités de la structure MOS idéale. La figure ci dessous représente le principe de la mesure de la capacité de la structure MOS en fonction de la tension de grille. La rampe de pulsation ω_p sert à fixer le point de fonctionnement (de polarisation). Le signal sinusoïdal de pulsation ω_m qui se superpose au signal de polarisation sert à moduler le point de fonctionnement et à faire la mesure.

Lorsque l'on applique une tension V_g sur la grille, cette tension se répartie entre l'oxyde et le semi-conducteur qui est à la masse, soit :

$$V_g = V_{ox} + V_s = \frac{-Q_{SC}}{C_{ox}} + V_s \quad (14)$$

L'oxyde et le semi-conducteur se comporte comme des capacités, la capacité d'oxyde donnée par $C_{ox} = \frac{\epsilon_{ox}}{d_{ox}}$ et la capacité du semi-conducteur donnée par $C_{SC} = \frac{d(-Q_{SC})}{dV_s}$, la tension aux bornes du semi-conducteur étant V_s . Le signe moins tient compte du fait que lorsque la tension V_s augmente, la charge Q_s augmente en négatif ! Si maintenant, on s'intéresse à la structure complète, on a $C_{MOS} = \frac{dQ_m}{dV_g} = \frac{d(-Q_{SC})}{dV_g}$. En prenant la dérivée par rapport à $-Q_s$ de (14), on obtient finalement :

$$\frac{1}{C_{MOS}} = \frac{1}{C_{ox}} + \frac{dV_s}{d(-Q_s)} = \frac{1}{C_{ox}} + \frac{1}{C_{SC}} \quad (15)$$

La capacité du semi-conducteur qui résulte de la charge d'espace est fonction de deux types de charges, une de déplétion appelée Q_{dep} et une de porteurs libres Q_s , ce qui permet d'écrire

$$C_{SC} = -\frac{dQ_{SC}}{dV_s} = -\left(\frac{dQ_{dep}}{dV_s} + \frac{dQ_s}{dV_s}\right) = C_{dep} + C_s \quad (16)$$

C_s est la capacité dynamique associée à la charge de porteurs libres dans le semi-conducteur. On voit à partir de (15) et (16), que la capacité totale de la structure MOS est fonction à travers C_{SC} du régime de fonctionnement.

- **Régime d'accumulation** ($V_s < 0$ et $V_g < 0$). On a vu précédemment que la charge dans le semi-conducteur était due à une charge d'accumulation :

$$Q_{SC} \approx \frac{\epsilon_{SC} \sqrt{2kT}}{eL_D} e^{-\frac{eV_s}{2kT}} \quad (17)$$

La capacité du semi-conducteur s'écrit alors (Taur p 71):

$$C_{SC} = \frac{e}{2kT} Q_{SC} = \frac{e}{2kT} C_{ox} |V_g - V_s| \quad (18)$$

On obtient en final :

$$\frac{1}{C_{MOS}} = \frac{1}{C_{ox}} \left(1 + \frac{2kT/e}{|V_g - V_s|} \right) \quad (19)$$

La tension thermique étant de l'ordre de 26 mV et le potentiel de surface n'excédant pas 0.3 à 0.4 V en accumulation, dès que la tension de grille est de l'ordre de -1 à -2 V, la capacité totale est égale à la capacité d'oxyde.

- **Régime de bandes plates** ($V_s = 0$). Dans ces conditions, on montre (Singh p 420) qu'une petite variation $\delta V(x)$ du potentiel produit une variation de densité de

charges $\delta\rho(x) \approx \frac{e\delta V(x)p_0}{kT}$. En résolvant l'équation de Poisson, on obtient

$|\delta Q_{SC}| = \frac{eN_A}{bkT} \delta V_s$ avec $b = \sqrt{\frac{eN_A}{kT}}$. La capacité du semi-conducteur est alors

donnée par $C_{SC}(fb) = \frac{\epsilon_{SC}}{L_D}$, ce qui conduit à une capacité totale de la structure de :

$$C_{MOS}(fb) = \frac{\epsilon_{ox}}{d_{ox} + \frac{\epsilon_{ox}}{\epsilon_{SC}} \sqrt{\frac{kT}{e} \frac{\epsilon_{SC}}{eN_A}}} \quad (20)$$

Une application numérique montre que la capacité de bandes plates est légèrement inférieure à la capacité d'oxyde.

- **Régime de déplétion – faible inversion.** ($0 < V_s < 2\phi_{Fi}$). Le semi-conducteur à l'interface commence à devenir dépeuplé de trous (les porteurs libres majoritaires). Cette fois la variation de la tension de polarisation se manifeste par une variation de la largeur désertée. L'équation (12) peut alors se simplifier en

$Q_{SC} = -\frac{\sqrt{2}\epsilon_{SC}kT}{eL_D} \left(\frac{eV_s}{kT}\right)^{\frac{1}{2}} = -[2eN_A\epsilon_{SC}V_s]^{\frac{1}{2}} = Q_{dep}$. On obtient facilement la capacité du semi-conducteur :

$$C_{SC} = -\frac{dQ_{SC}}{dV_s} = \left(\frac{eN_A\epsilon_{SC}}{2V_s}\right)^{\frac{1}{2}} = \frac{\epsilon_{SC}}{W_{dep}} \quad (21)$$

La capacité totale de la structure MOS s'écrit alors :

$$C_{MOS}(déplétion) = \frac{\epsilon_{ox}}{d_{ox} + \frac{\epsilon_{ox}}{\epsilon_{SC}} W} \quad (22)$$

- **Régime de forte inversion.** Comme on peut le voir sur la figure, dans ce domaine, trois courbes sont présentes. Pourquoi ? Avant d'aller plus loin, on doit se demander comment et d'où viennent les électrons responsables de la couche d'inversion. Compte tenu de la polarisation et de la présence de l'isolant, ils ne peuvent provenir du métal. Le semi-conducteur étant de type p, la densité d'électrons est trop faible. En fait la seule solution est la génération thermique de paires électron – trou, les électrons restant confinés à l'interface. Les trous sont évacués vers le volume par le champ électrique en un temps caractéristique qui est le **temps de relaxation diélectrique**

$\tau = \frac{\epsilon_{SC}}{\sigma} \approx 10^{-12} s$. On a vu au chapitre 2 sur la jonction PN, que dans le cas de zone

déplétée, un courant de génération thermique était présent. Ceci à cause d'un taux net

de génération thermique donné par $g_{th} = \frac{n_i}{2\tau_m}$ (voir 33 du chap 2). Dans notre cas, au

départ la densité de porteurs est très faible, mais au fur et à mesure, la génération

thermique joue son rôle et la densité d'électrons augmente. Ainsi l'expression ci-dessus est un peu simpliste mais permet d'obtenir un bon ordre de grandeur. Un calcul plus rigoureux est donné dans **Jund et Poirier** (1966). La limite de forte inversion étant définie lorsque $n_S = N_A$; le temps τ_S pour créer N_A électrons est donné par :

$$g_{th} \tau_S = N_A \quad \Rightarrow \quad \tau_S = \frac{2N_A \tau_m}{n_i} \quad (23)$$

τ_S est défini comme le temps de stockage et c'est un paramètre très important pour les CCD. Les valeurs typiques varient de 0.1s à 1mn pour du silicium très pur. **Ainsi donc, la couche d'inversion peut mettre plusieurs secondes à se former.** En fonction de la fréquence utilisée pour faire la mesure de la capacité dynamique, cette couche pourra ou non être présente. La couche de déplétion quant à elle étant toujours existante (temps de mise en place = temps de relaxation diélectrique). On doit donc pour l'expression de la charge dans le Si, garder les termes en V_S et en $e^{\frac{eV_S}{kT}}$, qui sont respectivement la contribution de la déplétion F_{dep} et de l'inversion F_{inv} . On peut mettre ce résultat sous la forme :

$$Q_{SC} \approx -\left(\frac{kT}{e}\right) \frac{\epsilon_{SC} \sqrt{2}}{L_D} \left[\frac{eV_S}{kT} + e^{\frac{e(V_S - 2\phi_{Fi})}{kT}} \right]^{1/2} = -\left(\frac{kT}{e}\right) \frac{\epsilon_{SC} \sqrt{2}}{L_D} [F_{dep} + F_{inv}]^{1/2} \quad (24)$$

[On a exprimé le rapport $\frac{n_0}{p_0}$ en fonction de ϕ_{Fi}].

On doit maintenant envisager trois cas pour rendre compte de façon complète de la variation de la capacité MOS en forte inversion. On va donc comparer l'inverse du temps de stockage avec ω_m et ω_p .

- $\omega_p < \omega_m < \frac{1}{\tau_S}$: **capacité basse fréquence.** Dans ce cas la couche d'inversion existe et peut «suivre» la modulation de la tension de grille. Tout l'excédent de charge «demandé» par la modulation de grille est fourni par la couche d'inversion, si bien que la capacité s'écrit :

$$C_{SC} \approx C_S = \frac{\epsilon_{SC}}{\sqrt{2}L_D} e^{\frac{e(V_S - 2\phi_{Fi})}{2kT}} \quad (25)$$

En limite de forte inversion ($V_S = 2\phi_{Fi}$), (25) s'écrit tout simplement $C_{SC} = \frac{\epsilon_{SC}}{\sqrt{2}L_D}$.

En forte inversion, on voit donc que la capacité totale tend très vite vers la capacité de l'oxyde.

- $\omega_p < \frac{1}{\tau_S} < \omega_m$: **capacité haute fréquence.** Cette fois, la couche d'inversion existe toujours mais ne peut plus suivre la modulation de la tension de grille. Ainsi tout l'excédent de charge doit être fourni par la couche de déplétion. Ceci revient à poursuivre la courbe de déplétion, la capacité étant toujours donnée par $C_{SC} = \frac{\epsilon_{SC}}{W_M}$. W_M est la largeur maximum de la couche de déplétion obtenue pour $V_S = 2\phi_{Fi}$ et est donnée par l'expression suivante :

$$W_m = \left(\frac{4\epsilon_{SC}\phi_{Fi}}{eN_A} \right)^{1/2} \quad (25a)$$

On obtient alors l'expression de la capacité totale $C_{MOS}(\min) \approx \frac{\epsilon_{ox}}{d_{ox} + \frac{\epsilon_{ox}}{\epsilon_{SC}}W_m}$.

- $\frac{1}{\tau_S} < \omega_p < \omega_m$: **régime de déplétion profonde**. Dans ce cas, la variation de la polarisation de grille est trop rapide pour que la couche d'inversion ait le temps de se créer. La structure devant rester neutre, la seule solution est une augmentation de la charge de déplétion : ainsi la largeur de la ZCE augmente. Cette augmentation se poursuit jusqu'au claquage du semi-conducteur. La figure 6.5 résume l'ensemble.

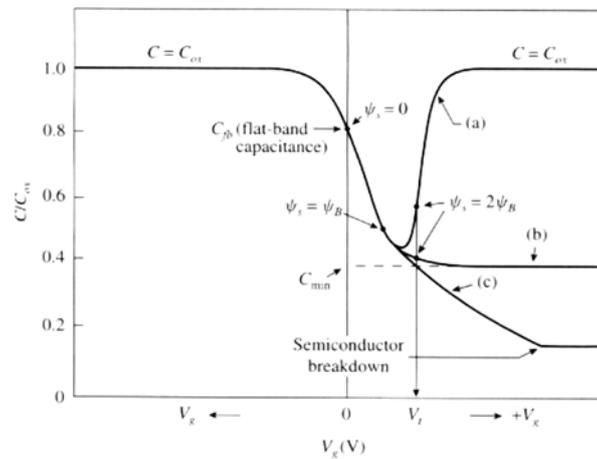


Figure 6.5 : courbe $C(V)$ d'une structure MOS

Structure MOS réelle.

La difficulté majeure pour la réalisation des transistors MOS et qui explique le retard pris par rapport à la technologie Bipolaire dans les années 60 et 70, fut la qualité médiocre de l'isolant, le SiO_2 . La présence de défauts et de charges dans l'isolant et des états d'interfaces en était les principaux responsables. Un autre paramètre qui va modifier le fonctionnement de la structure MOS est la différence des travaux de sortie entre le métal et le semi-conducteur. On va intégrer ces modifications dans les équations ci dessus et notamment sur le potentiel de surface.

- **Charges dans l'isolant et à l'interface oxyde – silicium**. En fait plusieurs types de charges sont présents dans le SiO_2 . La figure ci dessous résume les différents types. On doit remarquer que l'on parle de densité de charge par unité de surface.
 - **Charges ioniques mobiles Q_m** . En général dues aux ions sodium et potassium. Peuvent être éliminées.
 - **Charges dans le volume de l'oxyde Q_{ot}** . Elles sont créées par irradiation et dues à l'ionisation des centre pièges dans l'oxyde.
 - **Charges fixes Q_f** . Elles sont en général dans l'oxyde mais proche de l'interface semi-conducteur (30 \AA). Elles sont immobiles sous l'effet d'un champ électrique.

- **Charges piégées à l'interface** Q_{it} . elles sont dues à la rupture de la périodicité du réseau qui introduit des niveaux énergétiques dans la bande interdite. Les charges piégées à l'interface sont des électrons ou des trous sur ces niveaux. *Elles dépendent du potentiel de surface* V_S .

Pour une étude plus détaillée consulter Taur p82 et Nicollian et Brews.

- **Effet de la charge d'oxyde sur le potentiel de surface.** Il est bien évident que la présence de charges dans l'oxyde va perturber la répartition des charges dans le métal et le semi-conducteur (par influence électrostatique une charge va se créer pour partie dans le métal et pour partie dans le semi-conducteur), donc la courbure de bande et donc le potentiel de surface. Cela va finalement se répercuter sur la tension de grille qui est un paramètre mesurable. Plaçons nous dans le cas d'une structure polarisée en régime de « flat bands ». Considérons une charge pelliculaire Q_{ox} à l'abscisse x dans l'oxyde et une tension δV_G qui permet de se trouver en régime de « flat bands » et calculons δV_G . Si on applique sur la grille une tension égale à la tension de bandes plates, la charge dans le semi-conducteur est par définition nulle et donc toute la charge se retrouve dans le métal et est égale à $-Q_{ox}$. Le champ électrique créé dans l'oxyde par cette charge pelliculaire est donné par le théorème de Gauss :

$$E(x) = -\frac{Q_{ox}}{\epsilon_{ox}} \quad (26)$$

La tension induite par ce champ et *supportée par la grille* est donnée par :

$$\delta V_g = -\frac{xQ_{ox}}{\epsilon_{ox}} = -\frac{x}{d_{ox}} \frac{Q_{ox}}{C_{ox}} \quad (27)$$

Si la charge est répartie sur tout l'oxyde avec une densité de charge donnée par $\rho(x, V_S) = \rho(x) + Q_{it}(V_S)\delta(x - d_{ox})$, la tension de grille pour compenser la courbure de potentiel créée par celle ci est tout simplement donnée par :

$$\Delta V_g = -\int_0^{d_{ox}} \frac{x\rho(x, V_S)dx}{\epsilon_{ox}} = -\frac{1}{C_{ox}} \int_0^{d_{ox}} \frac{x\rho(x)dx + Q_{it}(V_S)d_{ox}}{d_{ox}} \quad (28)$$

Cette tension supplémentaire correspond (au signe près) à la courbure de bandes, même si la structure n'est pas polarisée, due à la présence de charge dans l'oxyde. Comme on l'a vu plus haut la densité de charge $\rho(x)$ est la somme des états présents dans l'isolant. Dans le cas ou cette charge est uniquement pelliculaire et présente à l'abscisse x , la tension est donnée par (27), ce qui permet de voir que l'influence d'un tel type de charge $Q_f + Q_{it}$ est important lorsque elle est positionnée

à l'interface oxyde – semi-conducteur ($x = d_{ox}$) et $\delta V_g = -\frac{Q_{ss}}{C_{ox}}$. En général, pour

simplifier l'écriture, on introduit une charge *équivalente* d'oxyde $Q_{ox}(V_S)$ par unité de surface qui permet d'écrire (28) sous la forme :

$$\Delta V_g(V_S) = -\frac{Q_{ox}(V_S)}{C_{ox}} \quad (29)$$

- **Effet de la différence des travaux de sortie.** La différence des travaux de sortie est le deuxième paramètre qui va influencer la valeur de la tension de bandes plates. Le phénomène se traduit comme pour la jonction métal – semi-conducteur par une courbure de bandes qu'il va falloir compenser par une tension appliquée sur la grille. Cette tension est tout simplement égale (voir figure) à la différence entre les deux travaux de sortie, à savoir :

$$\Delta V_g = \Phi_m - \Phi_s = \Phi_{ms} \quad (30)$$

En conclusion, la **tension de bandes plates** qui tient compte à la fois de la présence de charge dans l'oxyde et de la différence des travaux de sortie qui sont **deux effets cumulatifs** est donnée par (29) + (30) soit :

$$V_{FB} = \Phi_{MS} - \frac{Q_{ox}}{C_{ox}} \quad (31)$$

Toutes les courbes C-V sont décalées d'une tension V_{FB} , et la tension de grille s'écrit maintenant :

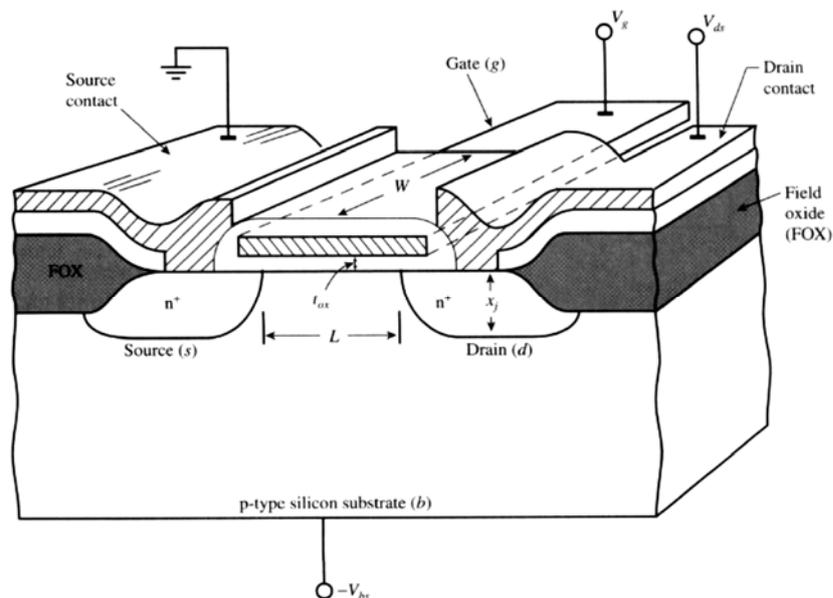
$$V_g = V_{FB} + V_{OX} + V_S = V_{FB} - \frac{Q_{SC}}{C_{ox}} + V_S \quad (32)$$

On va introduire maintenant un paramètre essentiel pour la compréhension du fonctionnement du transistor MOS – FET, la tension de seuil de la structure MOS, que l'on note V_T , pour « threshold voltage ». Cette tension est la polarisation que l'on doit appliquer sur la grille de façon à passer en limite de forte inversion. C'est donc la tension de grille V_g telle que $V_S = 2\Phi_{Fi}$ (voir 13). Sachant que $V_g = V_{OX} + V_S$ et $V_{ox} = -\frac{Q_{sc}}{C_{ox}}$, on en déduit l'expression de la tension de seuil :

$$V_T = V_g(V_S = 2\Phi_{Fi}) = \frac{\sqrt{4\epsilon_{sc}eN_A\Phi_{Fi}}}{C_{ox}} + 2\Phi_{Fi} + V_{FB} \quad (33)$$

On obtient l'expression (33) est remarquable que la charge dans le semi-conducteur est essentiellement une charge de déplétion en limite de forte inversion.

Transistor MOS – FET. Un transistor MOS-FET est réalisé à partir d'une capacité MOS et ou de part et d'autre on a ajouté deux contacts ohmiques qui vont servir de drain et de source. En général un quatrième contact est pris sur le substrat et on obtient la structure de la figure 6.6. La figure représente un MOS-FET avec un substrat de type p dans lequel on a incorporé deux régions N^+ qui servent donc de source (à la masse) et de drain. La région sous la grille entre la source et le drain est appelée *canal*. Lorsque la polarisation de grille est inférieure à la tension de seuil de la structure, le canal est en soit en accumulation soit en déplétion et on se retrouve avec un dispositif équivalent à deux jonctions NP tête-bêche, donc avec un courant inverse de jonctions. En revanche si la tension de grille est supérieure à V_T , une couche d'inversion (donc de type N) se forme pour créer un canal conducteur qui assure la continuité de type de porteurs entre la source et le drain et donc permet la circulation du courant. En modulant la tension de grille au delà de V_T , on module la densité de porteurs dans le canal donc la conductivité du canal donc le courant. On parle de *MOS-FET canal n* ou *n MOS-FET*. Les figures ci après résument les différents régimes de fonctionnement du transistor.



Chapitre 6 Figure 6.6 : Vue en coupe d'un MOS-FET (Taur)

6.1. Calcul du courant de drain. Le calcul du courant dans un transistor MOS – FET est compliqué car on doit considérer un modèle 3 – D. Dans ce cas, seule une résolution numérique permet d'aboutir à la caractéristique I – V. La figure 6.7 représente une vue en coupe du transistor avec l'orientation des axes que l'on va utiliser dans la suite. Définissons quelques paramètres :

- L longueur du canal (suivant y)
- W largeur du canal (suivant z)
- V tension dans le canal (dépend de y) $V(0) = 0, V(y = L) = V_{ds}$

On supposera que la tension de substrat est nulle et que la source est à la masse.

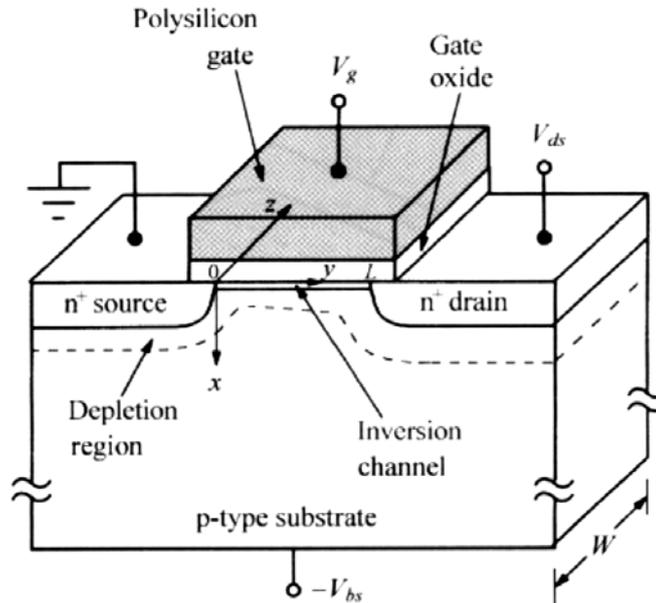


Figure 6.7 : schéma simplifié du MOS (Taur)

6.1.1. approximation de la charge surfacique. Pour résoudre analytiquement le problème, on fait l'hypothèse que la charge d'inversion dans le canal n'a pas d'épaisseur et qu'il n'existe pas de potentiel et de courbure de potentiel à travers cette couche d'inversion. La première étape pour calculer le courant est d'établir la relation entre la charge Q_{inv} dans le canal à l'interface oxyde – semi-conducteur et la tension de grille V_g . Les différentes densités de charge sont données par :

$$Q_{dep} = -eN_A W_M = -\sqrt{2eN_A \epsilon_{SC} V_S(y)} = -\sqrt{2eN_A \epsilon_{SC} (2\Phi_{Fi} + V(y))} \quad (34)$$

$$Q_{sc}(y) = -Q_{métal}(y) = -C_{ox}(V_g - V_{FB} - V_S(y)) = -C_{ox}(V_g - V_{FB} - 2\Phi_{Fi} - V(y)) \quad (35)$$

$$Q_{inv} = Q_{sc} - Q_{dep} = -C_{ox}(V_g - V_{FB} - 2\Phi_{Fi} - V(y)) + \sqrt{2eN_A \epsilon_{SC} (2\Phi_{Fi} + V(y))} \quad (36)$$

A partir de (36), on peut exprimer la densité d'électron dans le canal :

$$n_s(y) = \frac{|Q_{inv}|}{e} = \frac{|Q_{sc}|}{e} - \frac{|Q_{dep}|}{e} = \frac{C_{ox}(V_g - V_{FB} - 2\Phi_{Fi} - V(y))}{e} - \left[\frac{2N_A \epsilon_{SC} (2\Phi_{Fi} + V(y))}{e} \right]^{\frac{1}{2}} \quad (37)$$

Le courant dans le canal étant donné par :

$I_{DS} = \text{densité de charge} \times \text{mobilité} \times \text{champ} \times \text{largeur de grille}$

On obtient :

$$I_{DS} = Q_{inv} \mu_n \frac{-dV}{dy} W, \text{ soit encore } I_{DS} dy = -Q_{inv} \mu_n dV W$$

Il suffit alors d'intégrer de $y = 0$ à $y = L$, soit :

$$I_{DS} \int_0^L dy = \mu_n W \int_0^{V_{DS}} -Q_{inv} dV$$

et on obtient finalement :

$$I_{DS} = \mu_n \frac{W}{L} C_{ox} \left[(V_g - V_{FB} - 2\Phi_{Fi} - \frac{V_{DS}}{2}) V_{DS} - \frac{2}{3} \frac{\sqrt{2\varepsilon_{sc} e N_A}}{C_{ox}} \left[(2\Phi_{Fi} + V_{DS})^{\frac{3}{2}} - (2\Phi_{Fi})^{\frac{3}{2}} \right] \right]$$

(38)

Cette équation (38) représente la caractéristique I – V du transistor MOS – FET dans l'approximation de la charge surfacique. A partir de cette équation, on peut distinguer deux régimes en fonction de la tension de polarisation de grille V_g : le régime « *linéaire* » ou « *triode* » qui correspond à une augmentation linéaire du courant avec V_g puis un ralentissement jusqu'à atteindre une valeur de saturation d'où le nom de régime « *saturé* ».

6.1.2 Régime linéaire ou triode. En pratique, on utilise une expression simplifiée de (38) dans le cas où le transistor n'est pas saturé, expression qui devient :

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T - \frac{V_{DS}}{2}) V_{DS} \right] \quad (39)$$

Dans le cas où la tension drain – source est petite, on peut faire un développement limité au premier ordre de (38) en V_{DS} , ce qui permet d'obtenir :

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_{fb} - 2\Phi_{Fi} - \frac{\sqrt{4\varepsilon_{sc} e N_A \Phi_{Fi}}}{C_{ox}}) V_{DS} \right] = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T) V_{DS} \quad (40)$$

On rappelle que V_T est la tension de seuil du transistor ou tension de mise en conduction. Lorsque la tension de grille V_g est inférieure à V_T , un faible courant circule dans le canal et on parle de « *subthreshold current* » ou « *subthreshold region* ».

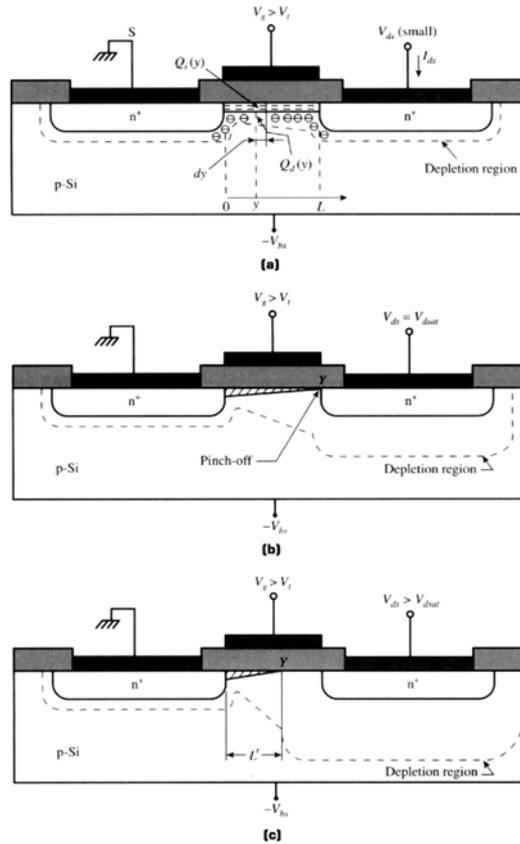


Figure 6.8 : les différents modes de fonctionnement vu en coupe (Taur)

6.1.3 Régime saturé. Pour des tensions drain – source élevées, le développement en série doit conserver le terme d’ordre 2 qui n’est plus négligeable et l’on obtient :

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left[(V_{gs} - V_T) V_{DS} - \frac{m}{2} V_{DS}^2 \right] \quad (41)$$

avec

$$m = 1 + \frac{\sqrt{\varepsilon_{sc} e N_A / 4 \Phi_{Fi}}}{C_{ox}} = 1 + \frac{C_{dm}}{C_{ox}} = 1 + \frac{3d_{ox}}{W_m} \quad (42)$$

C_{dm} est la capacité de déplétion dans le Si en limite de forte inversion.

Attention, l’expression (41) n’est valide qu’en limite de saturation ($V_{DS} < V_{Dsat}$).

D’après (41), le courant suit une parabole jusqu’à ce qu’il atteigne le sommet ou la valeur de saturation. Cette valeur est atteinte pour la tension de saturation

$$V_D = V_{Dsat} = \frac{(V_{gs} - V_T)}{m}, \text{ pour laquelle :}$$

$$I_{DS} = I_{Dsat} = \mu_n C_{ox} \frac{W}{L} \frac{(V_{gs} - V_T)^2}{2m}$$

Cette expression peut se simplifier en se rapprochant de la réalité, à savoir un niveau de dopage faible du Si et une épaisseur d'oxyde mince, ce qui conduit à $m=1$ et on obtient ainsi l'expression familière du courant de saturation d'un transistor MOS – FET, à savoir :

$$I_{DS} = I_{Dsat} = \mu_n C_{ox} \frac{W}{2L} (V_{gs} - V_T)^2 \quad (43)$$

Lorsque l'on atteint la saturation, en premier lieu coté Drain, la couche d'inversion disparaît c'est à dire que $Q_{inv}(L) = 0$ ou encore $Q_{inv}(V = V_{Dsat}) = 0$. Lorsque V_{Dsat} augmente au delà de la tension de saturation, le point de pincement se déplace vers la source (voir figure 6.8) mais le courant reste constant car la tension aux bornes du canal est constante et égale à V_{Dsat} . En fait le courant qui est donné par

$\int_0^{L'} I_{DS} dy = \mu_n W \int_0^{L'} (-Q_{inv}(V)) dV$ augmente légèrement du fait de la diminution de la longueur effective du canal, passant de L à L' . Ce phénomène (« modulation de la longueur du canal ») est surtout sensible pour des transistors à canal court. Le calcul de V_{Dsat} { il faut faire $Q_{inv}(V = V_{Dsat}) = 0$ dans (36), et en déduire V_{Dsat} } donne :

$$V_{Dsat} = V_{gs} - V_{FB} - 2\Phi_{Fi} + \frac{\epsilon_{sc} e N_A}{C_{ox}^2} - \sqrt{\frac{2\epsilon_{sc} e N_A}{C_{ox}^2} (V_{gs} - V_{FB} + \frac{\epsilon_{sc} e N_A}{2C_{ox}^2})} \quad (44)$$

En injectant cette expression dans (38), on obtient l'expression du courant de saturation

$$I_{Dsat} = \mu_n \frac{W}{6L} C_{ox} \left[(V_{Dsat} + 2\Phi_{Fi})(V_{Dsat} + 2\Phi_{Fi} + 2V_{gs} + 2V_{FB}) - 12\Phi_{Fi}(V_{gs} + V_{FB} - \Phi_{Fi} - \frac{4}{3} \frac{(eN_A \epsilon_{sc} \Phi_{Fi})^{\frac{1}{2}}}{C_{ox}}) \right] \quad (45)$$

Ces deux expressions ne sont que très peu utilisées, et en supposant encore une fois que le dopage est faible et l'oxyde mince (donc C_{ox} grand), l'expression (33) de la tension de seuil se réduit à $V_T \approx 2\Phi_{Fi} + V_{FB}$ ou encore $V_{gs} - V_{FB} \approx 2\Phi_{Fi}$, ce qui permet de réécrire (44) et (45) sous la forme usuelle bien connue :

$$\begin{aligned} V_{Dsat} &\approx V_{gs} - V_T \\ I_{Dsat} &\approx \mu_n \frac{W}{2L} C_{ox} (V_{gs} - V_T)^2 = \mu_n \frac{W}{2L} C_{ox} V_{Dsat}^2 \end{aligned} \quad (45)$$

6.1.4. Caractéristiques sous le seuil. Trois régimes de fonctionnement sont possibles dans un MOS – FET, les régimes linéaire, saturé et « bloqué ». Les deux premiers régimes ont été décrit précédemment, le troisième le sera dans cette partie. En général, on considère que lorsque la tension de grille est inférieure à la tension de seuil V_T , le transistor est en mode « bloqué », et aucun courant ne circule. La raison en est simple : si

$V_g < V_T$, le régime de forte inversion n'est pas atteint et le canal ne peut se former. En fait on sait qu'avant ce régime, une couche dite de « faible inversion » existe, et donc le canal aussi, bien qu'il soit moins conducteur (donc plus résistif). Le passage « bloqué » « conducteur » n'est donc pas si brutal et on parle d'un régime « sous le seuil » ou « subthreshold regime » qui correspond au régime de faible inversion dans le cas de la capacité MOS. On a donc un courant même pour $V_g < V_T$. Ce régime perdure tant que nous sommes dans la situation d'un potentiel de surface tel que $\Phi_{Fi} < V_S < 2\Phi_{Fi}$. On peut montrer que le processus de conduction dans ce régime est un phénomène de diffusion. Un calcul montre, dans le cas où $|V_S - 2\Phi_{Fi}| \ll \Phi_{Fi}$, que le courant s'écrit :

$$I_{DS} = \mu C_{ox} \frac{W}{L} (m-1) \left(\frac{kT}{e} \right)^2 e^{e(V_g - V_T)/mkT} (1 - e^{-eV_{DS}/kT}) \quad (46)$$

Dès que la tension drain – source est supérieure à quelques kT , le courant est indépendant de celle – ci. Si on s'intéresse à la dépendance du courant en fonction de la tension de grille, on trouve :

$$S = \left(\frac{d(\log_{10} I_{DS})}{dV_g} \right)^{-1} = 2.3 \frac{mkT}{e} = 2.3 \frac{kT}{e} \left(1 + \frac{C_{dm}}{C_{ox}} \right) \quad (47)$$

On voit que la dépendance du courant est indépendante des paramètres du dispositif (à part au travers de C_{dm}). Par contre, *il est très sensible à la température*.

6.1.4. « effet substrat » et effet de la température sur la tension de seuil. La connexion du substrat est largement utilisée notamment dans les applications mémoires et numériques. En fait, cette polarisation du substrat V_{SB} entraîne, comme on va le voir, un décalage de la tension de seuil. Sans polarisation de substrat, la forte inversion est obtenue pour un potentiel de surface égal à $2\Phi_{Fi}$. D'après la figure 6.9, on voit que cette tension est augmentée de $+V_{SB}$ si le substrat est polarisé. La charge de déplétion maximum (en limite de forte inversion) n'est plus W_{max} mais elle est augmentée pour absorber la tension supplémentaire V_{SB} . Cette largeur devient (on remplace dans 25a, $2\Phi_{Fi}$ par $2\Phi_{Fi} + V_{SB}$) :

$$W = \sqrt{\frac{2\epsilon_{sc}(2\Phi_{Fi} + V_{SB})}{eN_a}} \quad (48)$$

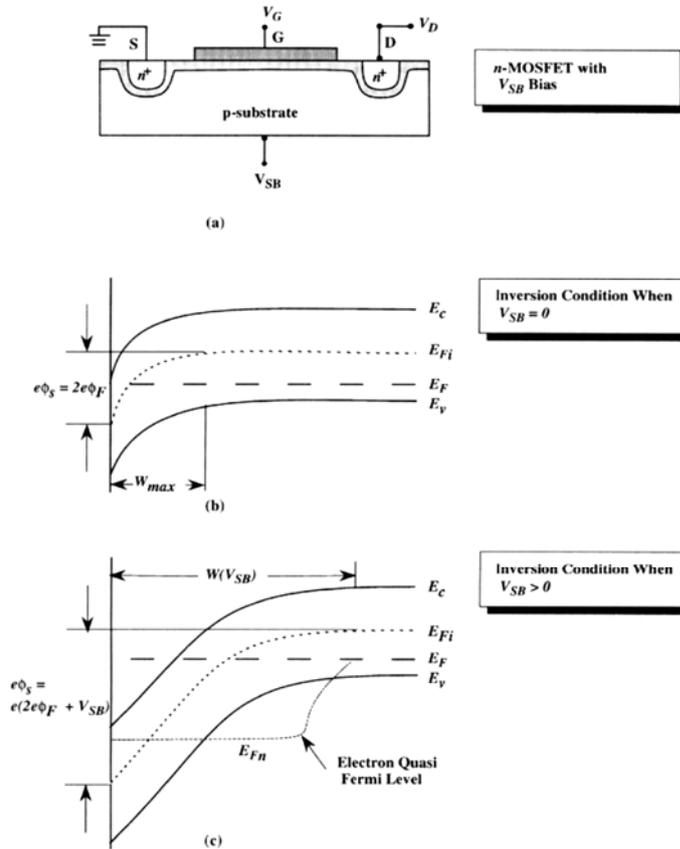
la charge supplémentaire est donc donnée par :

$$\Delta Q = -eN_a W (V_{V_{gs} + V_{SB}}) + eN_a W_m = -\sqrt{2eN_a} [\sqrt{2\Phi_{Fi} + V_{SB}} - \sqrt{2\Phi_{Fi}}] \quad (49)$$

On obtient alors la sensibilité de la tension de seuil en fonction de la tension de substrat :

$$\frac{dV_T}{dV_{SB}} = -\frac{1}{C_{ox}} \frac{dQ}{dV_{SB}} = \frac{\sqrt{\epsilon_{sc} e N_a / 2(2\Phi_{Fi} + V_{SB})}}{C_{ox}} \quad (50)$$

$$V_T = V_{T0} + \gamma \left(\sqrt{|2\Phi_{Fi}| + V_{SB}} - \sqrt{|2\Phi_{Fi}|} \right) \quad \text{avec } \gamma = \frac{\sqrt{2eN_A \epsilon_{SC}}}{C_{ox}} \quad (50a)$$



Pour assurer un décalage positif de la tension de seuil, V_{SB} doit être positif.

Figure 6.9

Un autre effet important sur la tension de seuil, est la température. La température modifie à la fois le « gap » et Φ_{Fi} . Un calcul montre que la variation de la tension de seuil avec la température se met sous la forme :

$$\frac{dV_T}{dT} = -(2m-1) \frac{k}{e} \left[\ln \left(\frac{\sqrt{N_C N_V}}{N_a} \right) + \frac{3}{2} \right] + \frac{m-1}{e} \frac{dE_g}{dT} \quad (51)$$

Des valeurs typiques avec du silicium donnent $\frac{dV_T}{dT} \approx -1mV/K$. Ce décalage de la tension de seuil induit une variation du courant sous le seuil. Cette variation peut entraîner une augmentation du courant à l'état bloqué d'un facteur 30 à 50 fois entre 25°C et 100°C.

6.2. Caractéristiques dynamiques. Deux figures de mérites des transistors MOS – FET sont la conductance de drain ou conductance de sortie g_D et la transconductance g_m . La conductance définit le contrôle du courant de drain par la tension de polarisation du canal et se met sous la forme :

$$g_D = \left. \frac{\partial I_D}{\partial V_D} \right|_{V_g = \text{cte}} = \mu_n \frac{W}{L} C_{ox} \left[V_{gs} - V_{FB} - 2\Phi_{Fi} - V_{DS} - \frac{\sqrt{2\varepsilon_{sc} N_a}}{C_{ox}} \sqrt{V_{DS} - 2\Phi_{Fi}} \right] \quad (52)$$

En général, on distingue deux cas, en fonction du régime de fonctionnement du transistor. Lorsque celui-ci est en régime purement linéaire (faible tension de drain), l'expression (52) se simplifie notablement pour donner :

$$g_{D_{lin}} = \frac{\mu_n W}{L} C_{ox} (V_{gs} - V_T) \quad (53)$$

En revanche, en régime saturé, dans le modèle simple, on doit trouver $g_{D_{sat}} = 0$, puisque le courant « sature » avec la tension. En fait, dans le cas de transistors à canal court, ce n'est plus le cas et le courant augmente légèrement avec la tension de drain. Cet effet peut être gênant lorsque la charge aux bornes du transistor varie.

La transconductance est liée à la vitesse du dispositif. En régime linéaire (faible tension de drain), elle est donnée par :

$$g_{m_{lin}} = \frac{\mu_n W}{L} C_{ox} V_{DS} \quad (54)$$

Dans le cas d'un fonctionnement en régime de saturation, la transconductance se met sous la forme :

$$g_{m_{sat}} = \frac{\mu_n W}{L} C_{ox} (V_{gs} - V_T) \quad (55)$$

7. Effets parasites sur les transistors réels. Nous allons dans ce paragraphe, examiner les différents effets qui modifient les expressions et le fonctionnement du transistor.

Chapitre 7

7.1. Variation de la mobilité. Deux facteurs sont à l'origine de la variation de la mobilité, la polarisation de grille et la polarisation de drain (en fait le champ électrique). Dans le premier cas, lorsque V_{gs} augmente, le puits le potentiel se creuse, augmentant par la même la densité de porteurs dans la couche d'inversion du fait de l'augmentation du champ électrique à l'interface oxyde – semi-conducteur. Ce champ « colle » les électrons plus proche de l'interface, là où la densité de défauts est importante et la qualité de la surface médiocre, ce qui a pour effet de diminuer la mobilité. Le deuxième facteur, l'augmentation de la tension de drain, va engendrer une augmentation du champ électrique le long du canal, donc la vitesse des porteurs ($\vec{v} = \mu \vec{E}$). En fait cette vitesse ne

peut dépasser une certaine valeur et va saturer à sa valeur maximum $v_{sat} \approx 10^7 \text{ cm/s}$ pour le silicium (voir figure 6.10). Cet effet à lieu pour des champs supérieurs au champ critique E_c ($\approx 1-2 \cdot 10^4 \text{ V/cm}$). Dans ce cas, même si la tension est inférieure à la tension de saturation V_{Dsat} mais égale à $V_D \approx E_c \cdot L$, le courant va saturer prématurément à la valeur $I_{Dsat} = WC_{ox}(V_g - V_T)v_{sat}$. Cet effet est surtout sensible pour les MOS – FET canal court.

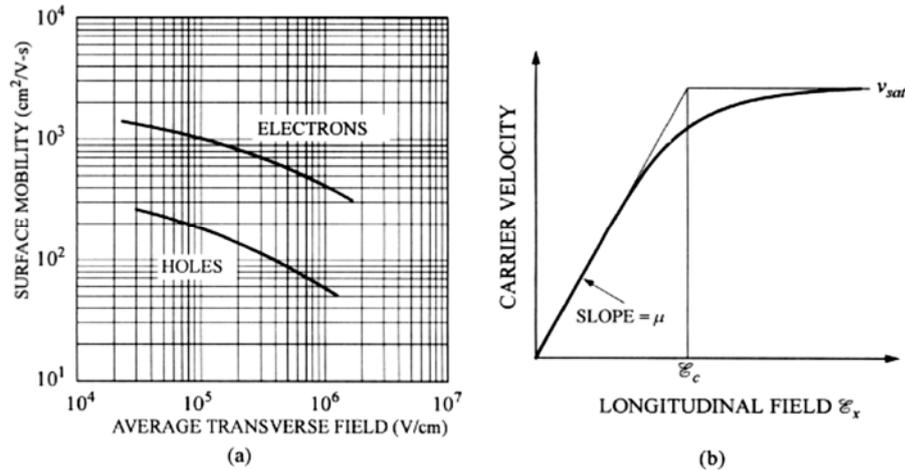


Figure 6.10 :variation de la mobilité (a) et de la vitesse (b) en fonction du champ électrique

7.2. modulation de la longueur du canal. Comme on l’a vu précédemment, lorsque la saturation est atteinte, le point de pincement se déplace vers la source, réduisant ainsi la longueur effective du canal. Ainsi, le courant qui est proportionnel à L_{eff}^{-1} va augmenter :

$$I_D = \frac{L}{L - \Delta L(V_{DS})} I_{Dsat} \quad (56a)$$

$$\Delta L(V_{DS}) = \sqrt{2 \frac{\epsilon_{Si}}{eN_A} (V_{DS} - (V_{GS} - V_T))} \quad (56b)$$

On introduit alors le paramètre λ (« LAMBDA » sous *Spice* level 1) qui décrit cette modulation de la longueur du canal. La relation empirique qui lie la tension V_{DS} et ΔL est donnée par :

$$1 - \frac{\Delta L}{L} \approx 1 - \lambda V_{DS} \quad (56c)$$

Si on suppose que $\lambda V_{DS} \ll 1$, le courant de saturation peut s’écrire :

$$I_D(sat) = \frac{\mu_n C_{ox} W}{2} \frac{V_{GS} - V_T}{L} (1 + \lambda V_{DS}) \quad (56d)$$

On constate alors une augmentation de la conductance de sortie.

7.3. Claquage du transistor. Ici encore, deux effets sont responsables du claquage du dispositif. Lorsque la tension de polarisation V_{ds} est importante, le champ électrique peut dépasser des valeurs de l'ordre de 10^5 V/cm, ce qui accélère les électrons du canal. Leur énergie est suffisante pour, par impact, ioniser et donc créer des paires électrons – trous. Les électrons sont propulsés dans le drain, augmentant le courant et les trous collectés via le substrat, ce qui crée un courant à travers le substrat. Ce courant dans le substrat crée une chute de potentiel à travers le silicium p qui polarise en direct la jonction source – substrat. Ceci diminue la tension de seuil du transistor et provoque donc une augmentation du courant. On parle de **claquage par ionisation**. En général, les dommages sont irrémédiables car les électrons à hautes énergies cinétiques « pénètrent » dans l'oxyde au niveau de la zone de recouvrement Grille – Drain, créant des défauts et des charges permanentes. Un deuxième effet qui est souvent rencontré dans les structures à canal court, est le **claquage par perçage** (« punchthrough breakdown »). Les ZCE source – substrat et drain – substrat peuvent occuper une large partie du canal et une polarisation drain source peut provoquer la fusion des deux ZCE. Dans ce cas les électrons sont directement injectés de la source vers le drain, le courant ne saturant pas dans ce cas.

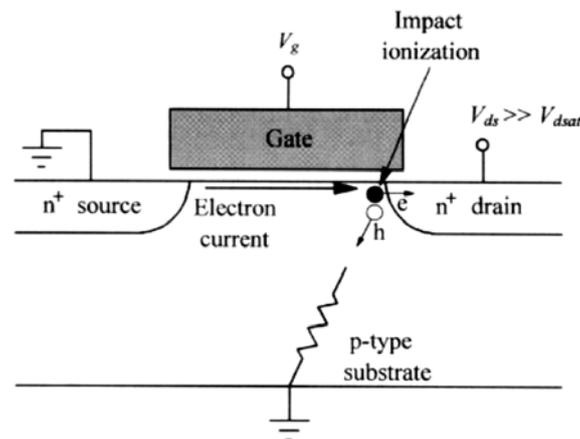


Figure 6.9 : claquage du transistor MOS-FET (Taur et al)

7.4. Transistors à canal court. La course à la miniaturisation des dispositifs avec maintenant des longueurs de canal qui peuvent descendre en dessous de $0.1\mu\text{m}$ conduit à un autre fonctionnement des transistors MOS – FET. Plusieurs effets se produisent que nous allons présenter dans ce qui suit.

7.4.1. Réduction de la tension de seuil. La réduction de la tension de seuil ou « short channel effect » se produit lorsque l'on diminue la longueur du canal. Dans le calcul de la tension de seuil, on a jusqu'à présent, considéré que les charges dans le canal étaient dues uniquement à l'application de la tension de grille. En fait des zones de déplétion existent du fait de la présence des régions n^+ (la source et le drain) qui entraînent des ZCE. En effet deux diodes n^+p entre ces régions et le substrat existent qui sont à l'origine de ces charges (plus développées coté drain du fait de la polarisation) qui participent en partie aux charges dans le canal. De ce fait une « partie du chemin » est déjà faite pour amener la zone du canal en forte inversion et la tension de seuil s'entrouve réduite. Cette réduction apparaît surtout pour des

tensions de polarisation de drain forte (\approx la tension d'alimentation). Compte tenu des tolérances dans la réalisation des transistors, un même lot de transistors réalisé sur la même tranche de silicium peut avoir des V_T différents. On doit tenir compte de cet effet lors du « design » des transistors.

Ainsi la tension de seuil résultante s'écrit (**Kang et Leblebici**):

$$V_T = V_{T(CL)} - \Delta V_T$$

avec

$$\Delta V_T = \frac{1}{C_{ox}} \cdot \sqrt{2e\epsilon_{Si} N_A |2\phi_{Fi}|} \cdot \frac{x_j}{2L} \left[\left(\sqrt{1 + \frac{2x_{ds}}{x_j}} - 1 \right) + \left(\sqrt{1 + \frac{2x_{dd}}{x_j}} - 1 \right) \right] \quad (57)$$

x_j : profondeur de la zone diffusée N^+ des régions Source et Drain

x_{ds} et x_{dd} sont les largeurs de ZCE côté Source et Drain respectivement. Noter que x_{dd} est fonction de la tension V_{DS} .

7.4.2. Transport 3D. On ne peut plus raisonner en considérant que les champs électriques sont orientés préférentiellement suivant la direction source – drain et utiliser ainsi l'approximation du canal graduel. Le courant devient ainsi 2D et si en plus la grille W n'est pas trop large, c'est un calcul numérique à 3D que l'on doit considérer. Une conséquence importante est la réduction de la tension de seuil du transistor.

7.4.3. Effets des électrons chauds. Pour une même polarisation, la longueur du canal étant réduite, les électrons acquièrent une énergie cinétique très importante (on parle de porteurs de charges « chauds ») qui leur permet par effet tunnel de traverser la grille et de créer un courant de grille. Cet effet dégrade le transistor par détérioration de l'oxyde (création de défauts et de charges).

8. Caractéristiques Hautes Fréquences. Deux facteurs limitent la fréquence de coupure du transistor, le temps de charge des capacités et le temps de transit des porteurs à travers le canal. Les figures ci dessous représentent le schéma équivalent d'un transistor MOS – FET en petits signaux :

- R_S, R_D : résistances associées à la source et au drain
- C_{GS} : capacité grille – source (inclus les capa de recouvrement)
- C_{GD} : capacité grille – drain (inclus les effets parasites)
- C_{DS} : représente la capacité substrat – drain de la jonction pn
- V_g' : tension interne grille – source ($= V_g - R_S I$)

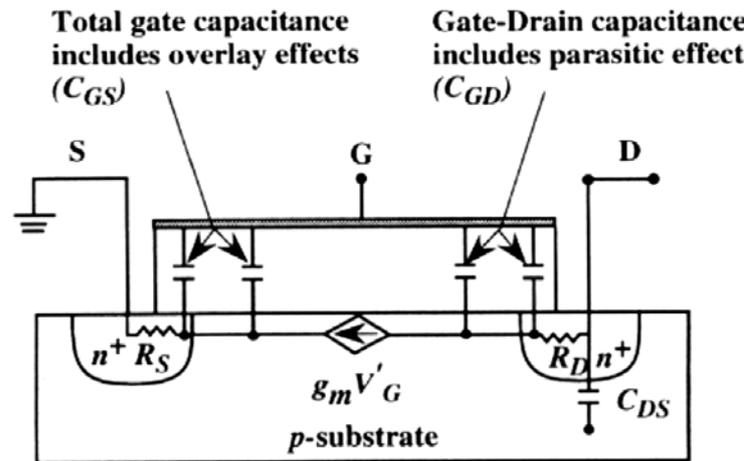


Figure 6.10 : origines des capacités (Singh)

On doit noter que les capacités d'oxyde sont fonction du régime de fonctionnement du MOS-FET. A partir de ces schémas, on peut calculer la fréquence de coupure du transistor qui est donnée pour un gain en courant unitaire. Pour simplifier le problème, on néglige les résistances et C_{DS} . Le courant d'entrée (la grille) est donné par :

$$I_{in} = j\omega C_{GS} V_g + j\omega C_{GD} (V_G - V_D) \quad (58)$$

on a également :

$$\frac{V_D}{R_L} + g_m V_g + j\omega (V_D - V_G) = 0 \quad (59)$$

A partir de (58) et (59) on en tire (en négligeant $j\omega R_L C_{GD}$ qui est petit) l'expression

$$I_{in} = f(V_g) : \quad I_{in} = j\omega (C_{GS} + C_{GD} (1 + g_m R_L)) V_G \quad (60)$$

Cette expression s'écrit souvent en faisant apparaître la **capacité de Miller** $C_M = C_{GD} (1 + g_m R_L)$.

Cette capacité de Miller est essentiellement due aux capacités de recouvrement Grille – Drain, capacités qui peuvent être fortement réduite en utilisant la technologie d'**autoalignement** en se servant de la grille comme masque.

En injectant la capacité Miller, on peut réécrire (60) sous la forme :

$$I_{in} = j\omega (C_{GS} + C_M) V_G \quad (61)$$

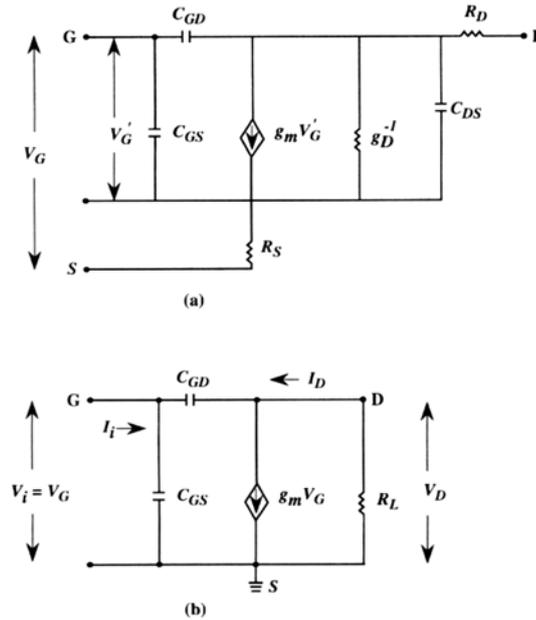


Figure 6.11 : schéma équivalent extrinsèque (a) et intrinsèque (b) (Singh)

Le courant de sortie (I_D) est simplement donné par $g_m V_g$, d'où l'expression de la fréquence de coupure :

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_M)} \quad (62)$$

On note qu'il est important d'avoir une capacité de Miller la plus faible possible. Dans le cas idéal où elle est nulle et en prenant $C_{GS} = C_{ox} WL = \frac{\epsilon_{ox} WL}{d_{ox}}$, on obtient l'expression de la fréquence de coupure maximum :

$$f_{T_{max}} = \frac{\mu_n(V_G - V_T)}{2\pi L^2} \quad (63)$$

Cette expression est très approximative dans le cas des transistors à canal court, en ce sens que la mobilité n'est plus constante, la vitesse sature et le canal se réduit. Dans ce cas, on préfère l'expression tenant compte du temps de transit τ_{tr} des porteurs dans le canal qui est donné par L/v_s , ce qui permet d'écrire :

$$f_T = \frac{v_s}{2\pi L} \quad (64)$$

Appendice 1

Capacités d'oxyde : comme on l'a vu plus haut, les capacités d'oxyde sont fonction du régime de fonctionnement du transistor MOS-FET. En fait la capacité est la somme d'une capacité dépendant du régime de fonctionnement et d'une capacité dépendant du recouvrement (« overlay » ou « overlap ») de la grille sur la source et le drain et égale à $C_{GS}(overlap) = C_{GD}(overlap) = C_{ox}.W.L_D$, si L_D est la largeur de recouvrement entre la grille et le drain.

On peut montrer que les capacités se résument alors à ce tableau :

Capacité	Bloqué	Triode	Saturé
$C_{gb}(total)$	$C_{ox}WL$	0	0
$C_{gd}(total)$	$C_{ox}WL_D$	$\frac{1}{2}C_{ox}WL + C_{ox}WL_D$	$C_{ox}WL_D$
$C_{gs}(total)$	$C_{ox}WL_D$	$\frac{1}{2}C_{ox}WL + C_{ox}WL_D$	$\frac{2}{3}C_{ox}WL + C_{ox}WL_D$

Equations Tension – Courant d'un MOS – FET canal n

$I_D = 0$	pour $V_{GS} < V_T$	(Bloqué)
$I_D(lin) = \frac{\mu_n C_{ox}}{2} \frac{W}{L} [2(V_{GS} - V_T)V_{DS} - V_{DS}^2]$		pour $V_{GS} \geq V_T$ et $V_{DS} < V_{GS} - V_T$
$I_D(sat) = \frac{\mu_n C_{ox}}{2} \frac{W}{L} [V_{GS} - V_T]^2 (1 + \lambda V_{DS})$		pour $V_{GS} \geq V_T$ et $V_{DS} \geq V_{GS} - V_T$ (saturé)

Equations Tension – Courant d'un MOS – FET canal p

$I_D = 0$	pour $V_{GS} > V_T$	(Bloqué)
$I_D(lin) = \frac{\mu_n C_{ox}}{2} \frac{W}{L} [2(V_{GS} - V_T)V_{DS} - V_{DS}^2]$		pour $V_{GS} \leq V_T$ et $V_{DS} > V_{GS} - V_T$
$I_D(sat) = \frac{\mu_n C_{ox}}{2} \frac{W}{L} [V_{GS} - V_T]^2 (1 + \lambda V_{DS})$		pour $V_{GS} \leq V_T$ et $V_{DS} \leq V_{GS} - V_T$ (saturé)

9. References

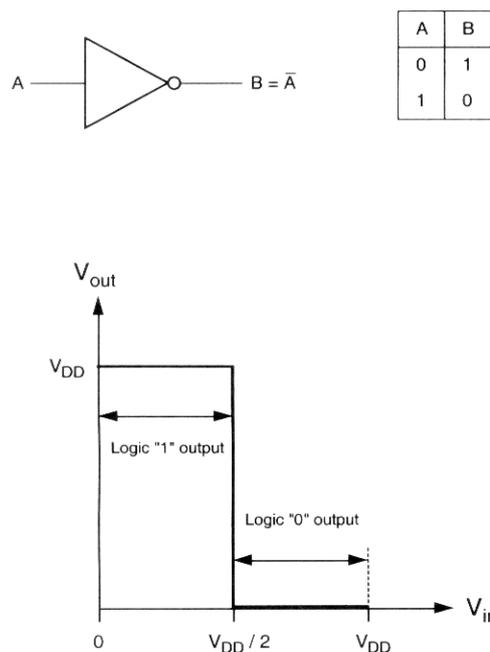
- S.M. Sze « Physics of semiconductors devices », 2° édition, Wiley, New York, 1981
- H.Mathieu, « Physique des semi-conducteurs et des composants électroniques », 4° édition, Masson 1998.
- J. Singh, « semiconductors devices : an introduction », McGraw-Hill, Inc 1994.
- Y.Taur et T.H. Ning, « Fundamentals of Modern VLSI devices », Cambridge University Press, 1998.
- K.K. Ng, « complete guide to semiconductor devices », McGraw-Hill, Inc, 1995
- E. H. Nicollian et J. R. Brews, « MOS Physics and Technology », John Wilet and Sons, 1982.

Les figures sont tirées en grande partie de ces différents ouvrages

Chapitre 7 : Inverseurs à transistors MOS

L'utilisation d'inverseurs MOS est fondamentale pour la réalisation de fonction logique. Divers types d'inverseurs peuvent être utilisés, mais on verra dans la suite de ce chapitre que deux montages voire un seul dominant dans les circuits numériques pour des raisons que nous allons détailler.

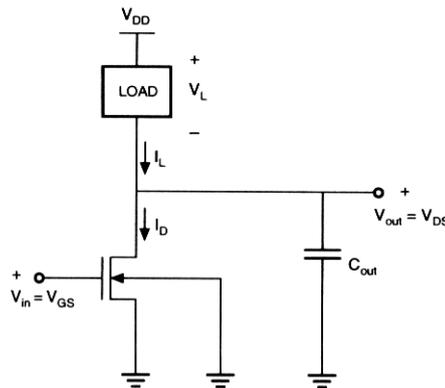
- 1. Introduction :** la définition de l'inverseur est très simple : définissons une tension d'entrée V_{in} et une tension de sortie V_{out} . Lorsque la tension d'entrée est au niveau « 1 » (représenté par la tension V_{DD}), la sortie est au niveau logique « 0 » et vice versa, le passage du « 1 » au « 0 » devant se faire de manière « brutale ». La tension de seuil d'inversion V_{th} idéale est donnée par $V_{DD}/2$. Toute tension comprise entre 0 et V_{th} représentera un « 0 » logique, et toute tension comprise entre V_{th} et V_{DD} indiquera un niveau logique haut « 1 ». (voir figure 7.1).



Chapitre 8 *Figure 7.1 : symbole, table de vérité et caractéristique de transfert idéale d'un inverseur*

Le circuit générique d'un inverseur n – MOS est représenté sur la figure 7.2. La tension d'entrée est appliquée sur la grille du transistor ($V_{in} = V_{GS}$), la tension de sortie est « récupérée » sur le drain du transistor qui est lui même connecté à V_{DD} à travers une charge. La capacité de charge C_{out} représente la capacité équivalente de l'étage suivant, un inverseur étant généralement connecté à un autre pour réaliser des fonctions logiques. Comme le

courant de grille en statique est négligeable dans un MOS, il n'y a pas de courant qui circule



aussi bien en entrée qu'en sortie.

Chapitre 9 Figure 7.2 : schéma générique d'un inverseur NMOS

D'après la figure ci dessus, on voit que le courant qui circule dans le transistor n – MOS est égal au courant qui circule dans la charge, soit $I_{DS}(V_{in}, V_{out}) = I_L(V_L)$. La caractéristique de transfert en tension est obtenue en résolvant numériquement cette équation, et on obtient alors une caractéristique qui ressemble à celle de la figure 7.3. On peut à partir de cette caractéristique définir quelques notions importantes.

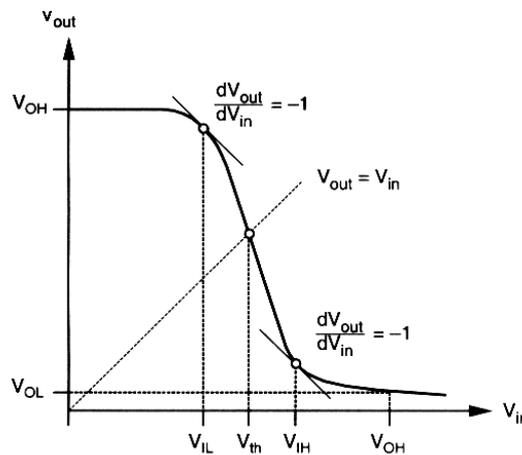


Figure 7.3 : caractéristique de transfert d'un inverseur

On remarque que pour des tensions d'entrée faible, la tension de sortie est égale à V_{OH} . Le transistor n – MOS est dans ce cas bloqué, le courant qui circule dans le composant qui sert de charge est nul et donc la tension de sortie au niveau haut (« 1 »). Lorsque la tension d'entrée (soit encore la tension de grille) augmente, le transistor pilote passe de bloqué à passant, le courant qui circule augmente, et la tension de sortie peut commencer à diminuer. La variation de la tension de sortie de l'état haut à l'état bas n'est pas brutale et on peut définir deux tensions d'entrée critique qui correspondent aux points pour lesquels la caractéristique à une

pente égale à -1 , c'est à dire encore $\frac{dV_{out}}{dV_{in}} = -1$. La tension d'entrée la plus petite est appelée V_{IL} et la tension la plus élevée V_{IH} . Lorsque la tension d'entrée augmente encore, la tension de sortie atteint une tension minimale appelée V_{OL} . Une autre tension importante qui est définie sur la caractéristique est la tension de seuil d'inversion V_{th} qui correspond au point $V_{in} = V_{out}$; on parle également de tension de transition de l'inverseur. Les définitions de ces tensions peuvent être résumées par :

- V_{OH} : tension maximum de sortie au niveau haut
- V_{OL} : tension minimale de sortie au niveau bas
- V_{IL} : tension d'entrée maximale qui peut être interprétée comme un « 0 »
- V_{IH} : tension d'entrée minimale qui peut être interprétée comme un « 1 »

Si on reprend la définition de V_{IH} , c'est la tension minimale qu'il faut appliquer à l'entrée pour être « sûr » que la sortie sera au niveau logique « 0 ». En d'autres termes, toute tension comprise entre 0 volt et V_{IL} sera interprétée par un « 0 » logique et la sortie sera à « 1 », toute tension comprise entre V_{IH} et généralement V_{DD} (la tension d'alimentation du circuit) sera interprétée comme un « 1 » logique et la sortie sera à « 0 », et **toute tension entre V_{IL} et V_{IH} ne pourra pas de façon certaine être interprétée, on sera dans le bruit**. L'immunité au bruit d'un inverseur va dépendre des tensions définies ci dessus. Le bruit provient en général de l'interconnexion entre différents inverseurs, ce bruit générant des tensions parasites qui viennent se superposer sur les tensions d'entrée et de sortie, générant des « fautes » logiques (voir figure 7.4).

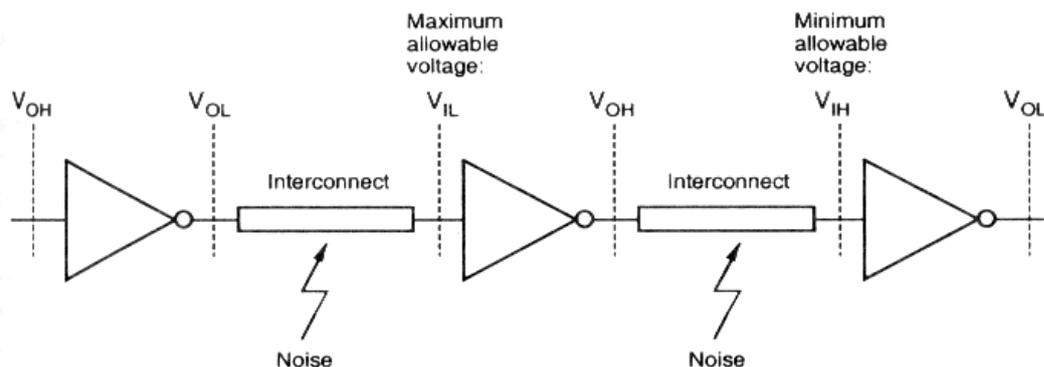


Figure 7.4 :

On introduit pour caractériser l'immunité au bruit d'un inverseur la notion de marges de bruit (« noise margins »), l'immunité étant d'autant plus grande que les marges de bruit le sont. Les marges de bruit pour le niveau haut et le niveau bas sont données respectivement par :

$$\begin{aligned}
 NM_H &= V_{OH} - V_{IH} \\
 NM_L &= V_{IL} - V_{OL}
 \end{aligned}$$

Ces valeurs sont représentées graphiquement sur la figure 7.5. On note sur cette figure la région de transition qui correspond à une région pour laquelle l'inverseur peut avoir un comportement aléatoire. C'est une région à éviter, c'est à dire une région à réduire (en faisant $V_{IL} = V_{IH}$?).

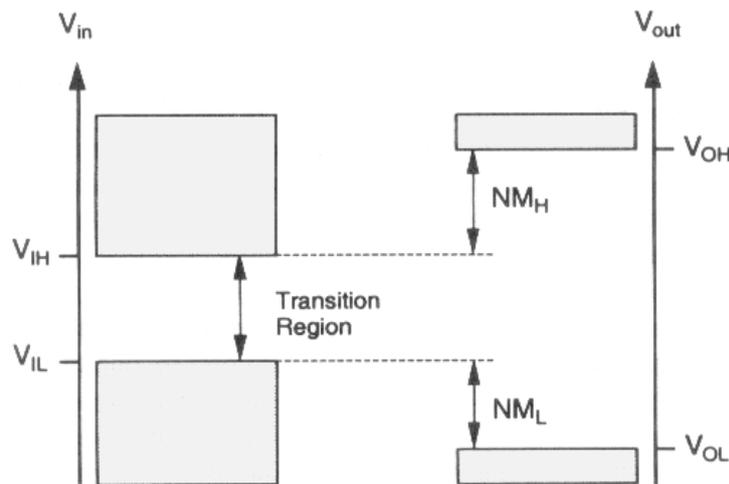


Figure 7.5 :

2. **Inverseur à charge résistive** : la structure d'un tel inverseur est représentée sur la figure ci – dessous.

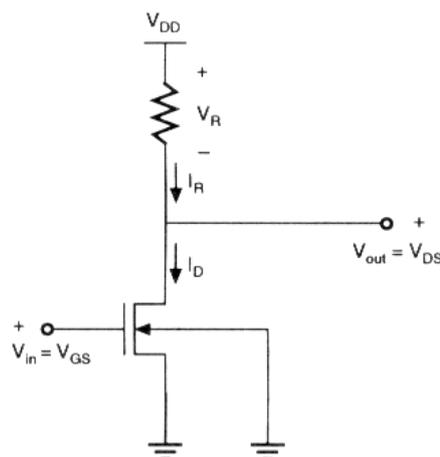


Figure 7.6 : Inverseur à charge passive ou résistive (Kang et Leblebici)

Comme l'on s'intéresse ici aux caractéristiques statiques, la capacité de sortie n'est pas représentée. Le transistor de commande (pilote ou « driver ») est un n – MOS et le substrat est connecté à la tension la plus négative (ici zéro, la source). Dans le cas de l'utilisation d'un p – MOS le substrat serait connecté à la tension la plus positive. Lorsque la tension d'entrée est au niveau « 1 », le n – MOS est conducteur et la tension de sortie se retrouve au niveau « 0 »,

dont la tension est fonction de la résistance R_{ON} dans l'état ON du n – MOS et bien évidemment de la valeur de la résistance de charge R_L . Le tracé de la caractéristique de transfert s'effectue en superposant la droite de charge et le réseau de caractéristiques du transistor n – MOS. Le courant traversant la résistance étant le même que celui qui traverse le transistor MOS, les points d'intersection donneront les points $V_{out}(V_{in})$. On peut également calculer les tensions critiques de l'inverseur, tensions qui détermineront les marges de bruit. D'abord déterminons les régimes de fonctionnement du transistor MOS. Par exemple, lorsque la tension d'entrée est inférieure à la tension de seuil, le transistor est bloqué. Le tableau résume les 3 régimes de fonctionnement :

Tension d'entrée	Régime de fonctionnement
$V_{in} < V_{Tn}$	Bloqué
$V_{Tn} \leq V_{in} < V_{out} + V_{Tn}$	Saturé
$V_{in} > V_{out} + V_{Tn}$	linéaire

On remarque donc qu'en fonction de la tension d'entrée, le transistor passe successivement de l'état bloqué, puis saturé et enfin linéaire.

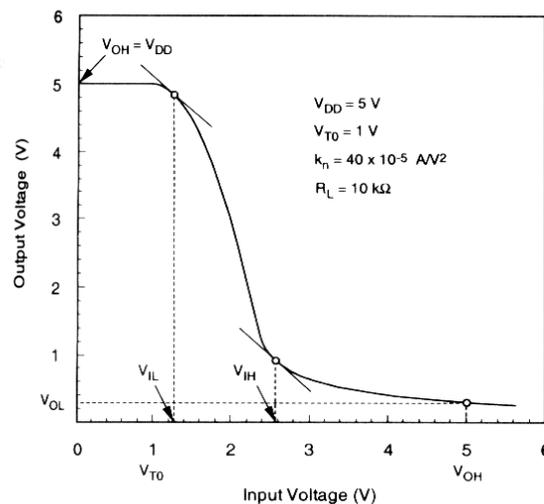


Figure 7.7 : caractéristique de transfert de l'inverseur à charge passive (Kang, Leblebici)

$$\begin{aligned}
 V_{OH} &= V_{DD} \\
 V_{OL} &= V_{DD} - V_{Tn} + \frac{1}{k_n R_L} - \sqrt{\left(V_{DD} - V_{Tn} + \frac{1}{k_n R_L}\right)^2 - \frac{2V_{DD}}{k_n R_L}} \\
 V_{IL} &= V_{Tn} + \frac{1}{k_n R_L} \\
 V_{IH} &= V_{Tn} + \sqrt{\frac{8}{3} \frac{V_{DD}}{k_n R_L}} - \frac{1}{k_n R_L}
 \end{aligned} \tag{1}$$

On peut remarquer, que si on augmente le produit $k_n R_L$, la caractéristique de transfert à un passage « 1 » => « 0 » plus brutal.

3. Inverseur n – MOS à charge active.

Dans ce type de montage, la résistance de charge est remplacée par un transistor à enrichissement dont sa résistance sera fonction de son point de fonctionnement. Il existe deux montages, un dit à charge saturée et l'autre à charge ... non saturée.

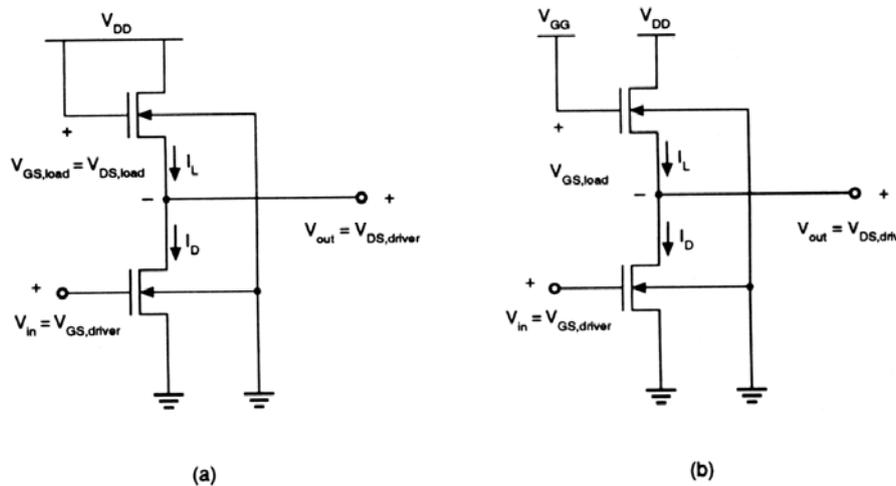


Figure 7.8 : a) Inverseur à charge saturée à transistor n – MOS à enrichissement. b) Inverseur à charge linéaire . (d'après Kang et Lebleci)

- **Inverseur à charge saturée.** (Fig 7.8.a) Dans ce montage, la grille du transistor de charge est au même potentiel que le drain, si bien que $V_{GS} = V_{DS}$. De ce fait, V_{DS} est toujours supérieur à $V_{DSat} = V_{GS} - V_{Tn}$, et le transistor est saturé

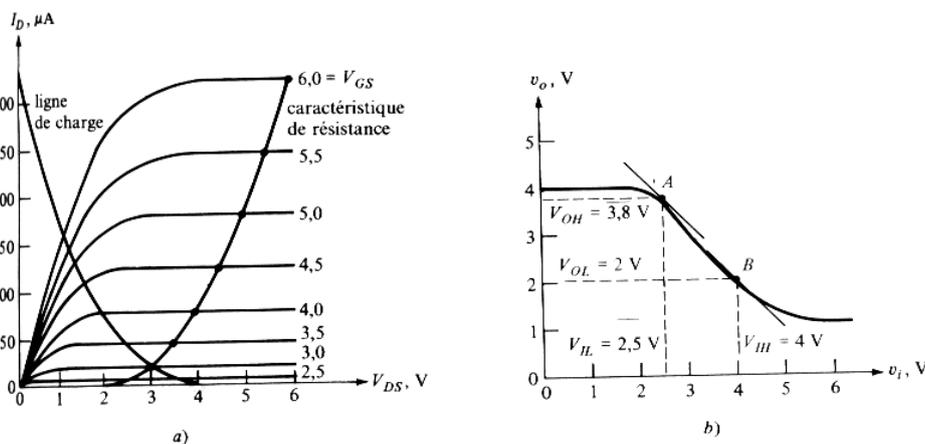


Figure 7.9 : a) caractéristique de résistance non linéaire et ligne de charge d'un inverseur à charge saturée. b) caractéristique de transfert. (d'après Grabel)

On remarque que l'allure de la caractéristique de transfert est loin d'être comparable avec celle d'un inverseur idéal. En pratique, on voit que V_{OH} est limitée à $V_{DD} - V_{Tn}$ et surtout que dans le cas ci dessus la marge de bruit au niveau haut est *négative*. On peut montrer qu'en « jouant » sur le rapport géométrique W/L, la caractéristique est nettement améliorée.

- **Inverseur à charge non saturée. (Fig 7.8.b)** Une autre façon d'améliorer la caractéristique de l'inverseur ci dessus (à part de modifier W/L) est de travailler avec une charge non saturée. Dans ce montage, le transistor de charge est toujours polarisée dans la région linéaire. Il suffit pour cela que $V_{GG} - V_{DD} > V_T$. L'inconvénient majeur de ce montage est la nécessité d'une deuxième source de polarisation par rapport au montage à charge saturée. Enfin un point dramatique en intégration VLSI est la consommation importante au repos en DC, aussi bien pour ce montage que le précédent.

Un moyen de réduire les désavantages est d'utiliser un transistor de charge à déplétion (ou appauvrissement). L'inconvénient apporté par la complication du processus de fabrication est largement compensé par les améliorations sur les caractéristiques de l'inverseur comme nous allons le voir.

- **Inverseur NMOS à appauvrissement de charge.(Fig 7.9.a)** Le transistor pilote (driver) est un N-MOS à enrichissement ($V_{Tn,driver} > 0$) alors que la charge est assurée par un N-MOS à appauvrissement ($V_{Tn,load} < 0$). La tension grille – source du transistor de charge est nulle. Comme la tension de seuil du N-MOS de charge est négative, la condition $V_{GS,load} > V_{Tn,load}$ est toujours assurée et le transistor n'est jamais bloqué. Une remarque importante doit être faite à ce niveau. D'un point de vue technologique, les deux transistors sont réalisés sur le même substrat P qui est relié à la masse. La tension source – substrat n'est donc pas nulle, mais au contraire égale à la tension de sortie, $V_{SB,load} = V_{DS,pilote} = V_{out}$. Aussi, la tension de seuil du transistor de charge est sujet à l'*effet substrat* vu dans le chapitre précédent (eq 50.a Chap 6) et on doit tenir compte de cet effet pour le calcul de la caractéristique de transfert de cet inverseur. Ici, la nouvelle tension de seuil s'écrit :

$$V_T(V_{out}) = V_{T0} + \gamma \left(\sqrt{|2\Phi_{Fi}| + V_{out}} - \sqrt{|2\Phi_{Fi}|} \right) \quad (2)$$

Le calcul des quatre tensions qui définiront les marges de bruits de cet inverseur se déroule de la même façon que précédemment et on obtient les expressions :

$$\begin{aligned}
 V_{OH} &= V_{DD} \\
 V_{OL} &= V_{OL} - V_{TO,driver} - \sqrt{(V_{OH} - V_{TO,driver})^2 - \frac{k_{load}}{k_{driver}} |V_{T,load}(V_{OL})|^2} \\
 V_{IL} &= V_{T,driver} + \frac{k_{load}}{k_{driver}} [V_{out} - V_{DD} + |V_{T,load}(V_{out})|] \\
 V_{IH} &= V_{T,driver} + +2V_{out} + \frac{k_{load}}{k_{driver}} [-|V_{T,load}(V_{out})|] \frac{dV_{T,load}}{dV_{out}}
 \end{aligned} \tag{3}$$

On voit, à partir de ces expressions que trois paramètres sont importants, les tensions de seuils des 2 transistors, en général figées par la technologie utilisée et le rapport géométrique des 2 transistors $\frac{k_{load}}{k_{driver}}$. C'est surtout ce paramètre qui est ajusté pour le dessin des inverseurs N-MOS, et notamment le rapport géométrique W/L des transistors de charge et pilote.

Un point sensible au niveau de la forte intégration VLSI, est la consommation de ces circuits. Dans le cas où le niveau de sortie est à «1» ou V_{OH} , le transistor pilote ne supporte pratiquement aucun courant, le transistor de charge par conséquent non plus. En revanche, en appliquant un niveau haut à l'entrée, les deux transistors sont passant et un courant non négligeable circule donné par $I_{DC}(V_{in} = V_{DD}) = I_{load}(sat) = I_{driver}(lin)$. En supposant que l'inverseur est « haut » pendant 50% du temps et à zéro le reste du temps, on peut calculer la puissance dissipée par l'inverseur en DC et on obtient :

$$P_{DC} = \frac{V_{DD}}{2} \frac{k_{load}}{2} [-V_{T,load}(V_{OL})]^2 \tag{4}$$

Cette consommation peut être rédhibitoire dans des circuits comportant des millions de transistors, ce qui a amené à la conception d'un autre type de circuit inverseur autre que ceux présentés ci-dessus dans une autre technologie.

4. **Inverseur CMOS**. Jusqu'à présent, la structure « inverseur » était constituée d'un transistor N-MOS à enrichissement dit pilote et d'une charge pouvant être soit résistive, soit un N-MOS à déplétion soit enfin un N-MOS à enrichissement. Un CMOS a une structure complètement différente, composée d'un transistor N-MOS et d'un transistor P-MOS jouant alternativement le rôle de pilote et de charge, on parle parfois de topologie « PUSH-PULL ». lorsque l'entrée est au niveau « haut », le N-MOS est le pilote, le P-MOS jouant le rôle de charge, la situation étant inversée avec l'entrée au niveau « bas ». Le CMOS a principalement deux avantages sur les inverseurs N-MOS : une consommation négligeable voire nulle au repos (à « 1 » ou à « 0 »), ie une dissipation nulle. De plus la caractéristique de transfert montre une plage en tension de sortie comprise entre 0 et V_{DD} et une caractéristique très proche de celle idéale (transition abrupte). La plus grande complexité de la technologie et donc son prix plus élevé est compensé par d'autres avantages déterminants que nous allons voir par la suite.

- **Principe de fonctionnement.** La figure d'un inverseur CMOS nous montre que la tension d'entrée est appliquée simultanément sur les deux grilles du N-MOS et du

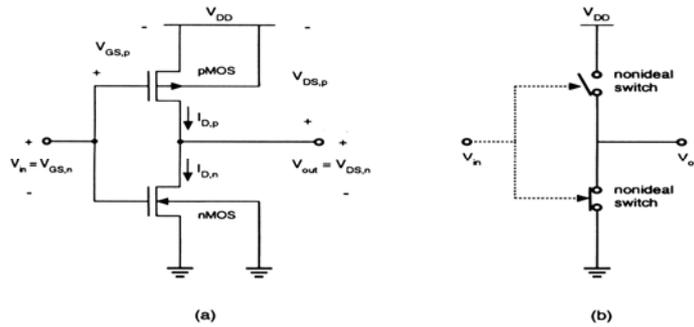


Figure 7.10 : inverseur C-MOS (Kang et lebleci)

P-MOS. Les deux transistors sont donc commandés par la même tension d'entrée V_{in} . Le substrat du N-MOS est connecté à la masse et celui de P-MOS à la tension d'alimentation, V_{DD} . Ainsi, pour les deux transistors, la tension source – substrat est nulle et aucun effet substrat ne sera à prendre en compte dans le fonctionnement du circuit. Si on résume dans un tableau les tensions pratiques appliquées sur le circuit, on obtient :

$V_{GS,p}$	$-(V_{DD} - V_{in})$
$V_{DS,p}$	$-(V_{DD} - V_{out})$
$V_{GS,n}$	V_{in}
$V_{DS,n}$	V_{out}

Considérons deux cas pour expliquer simplement le fonctionnement de l'inverseur C-MOS. Supposons que la tension d'entrée est inférieure à la tension de seuil du NMOS. Dans ce cas le NMOS est bloqué ($V_{GS,n} < V_{Tn}$), par contre le PMOS conduit puisque $V_{GS,p} < V_{Tp}$. Ainsi le courant dans le circuit est nul et la tension de sortie se retrouve à $V_{out} = V_{DD} = V_{OH}$. Dans l'autre cas, lorsque la tension de grille est supérieure à $V_{DD} + V_{Tp}$, le PMOS est bloqué, le NMOS est en régime linéaire mais le courant de drain est nul ; de ce fait, la tension de sortie est donnée par $V_{out} = V_{OL} = 0$.

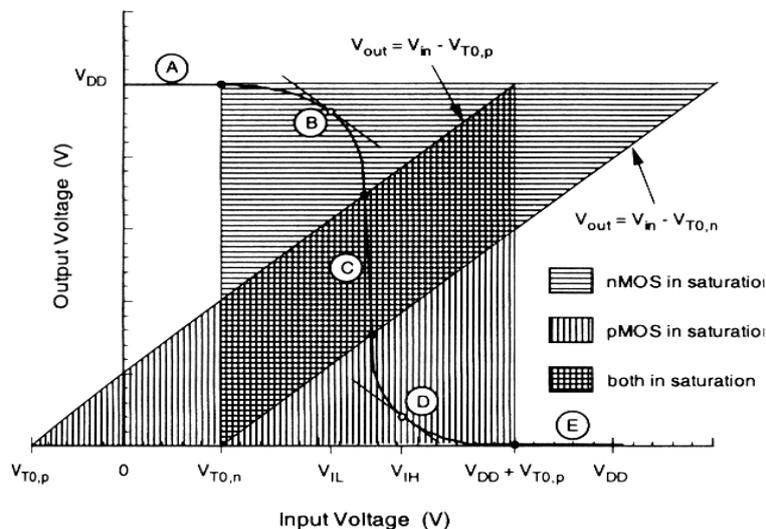


Figure 7.11 : caractéristique de transfert et régimes de fonctionnement du NMOS et du PMOS d'une porte CMOS (d'après Kang et Lebleci)

En analysant plus finement le fonctionnement du CMOS, c'est à dire en traçant la caractéristique de transfert (cf figure 7.11) , on peu remplir le tableau suivant qui résume le fonctionnement.

Région	V_{in}	V_{out}	NMOS	PMOS
A	$< V_{Tn}$	V_{OH}	Bloqué	Linéaire
B	V_{IL}	$\ll 1 \gg \approx V_{OH}$	Saturé	Linéaire
C	V_{th}	V_{th}	Saturé	Saturé
D	V_{IH}	$\ll 0 \gg \approx V_{OL}$	Linéaire	Saturé
E	$> (V_{DD} + V_{Tp})$	V_{OL}	Linéaire	Bloqué

5. **Inverseur CMOS en commutation. Effets des interconnexions et caractéristiques transitoires.** Pour étudier les caractéristiques en commutation d'une structure CMOS, supposons que deux « portes » CMOS sont connectées entre elles comme le montre la figure 7.12. Comme l'on s'intéresse aux phénomènes transitoires, on doit tenir compte des capacités parasites (voir chapitre précédent). Parmi toutes ces capacités parasites, la capacité notée C_{int} représente les effets des connexions métal ou polysilicium entre les deux portes.

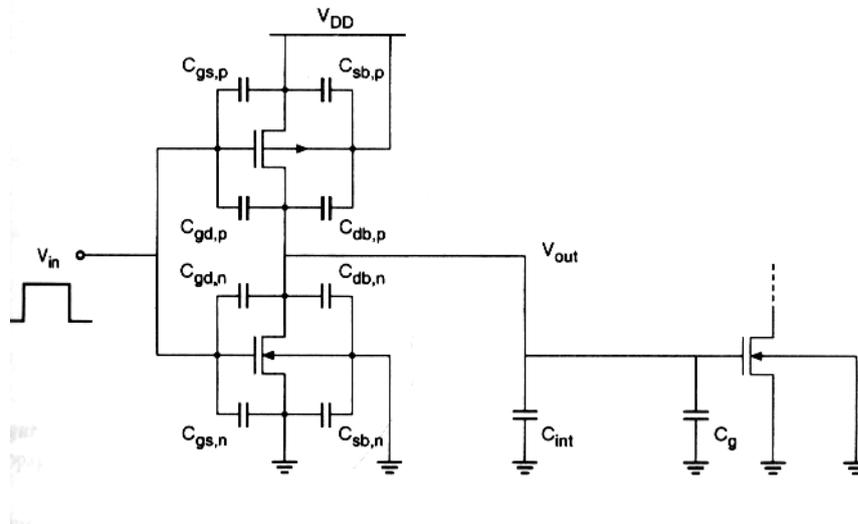


Figure 7.12 : étages d'inverseurs CMOS « cascades » équivalente (d'après Kang et Lebleci)

Dans le but de simplifier l'analyse comportementale du circuit, on se ramène en général à une capacité équivalente qui est connectée entre le nœud de sortie de la porte et la masse. Cette capacité est usuellement appelée C_{load} et est définie par :

$$C_{load} = C_{gd,n} + C_{gd,p} + C_{db,n} + C_{db,p} + C_{int} + C_g \quad (5)$$

Une fois cette simplification d'écriture réalisée, l'étude du circuit est largement simplifiée. En effet l'étude des transitoires se résume à la charge et la décharge d'une seule capacité de valeur C_{load} , comme le montre la figure 7.13.

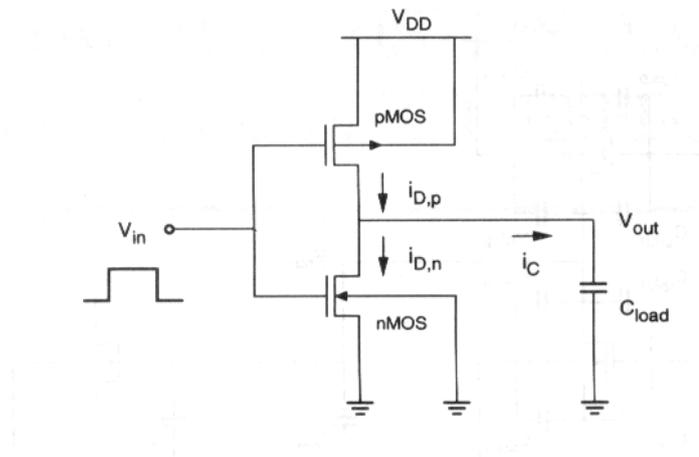


Figure 7.13 : circuit équivalent avec une capacité de sortie équivalente (d'après Kang et Leblebici)

- **Temps de retard – définitions** : on définit en général deux temps de retard (« delay time »), que l'on nomme τ_{PHL} et τ_{PLH} qui sont les temps mis pour que la tension de sortie passe de V_{OH} à $V_{50\%}$ ou de V_{OL} à $V_{50\%}$ respectivement. La tension $V_{50\%}$ est définie par $V_{50\%} = \frac{1}{2}(V_{OL} + V_{OH})$. Le temps de propagation moyen est donné par $\tau_p = \frac{\tau_{PHL} + \tau_{PLH}}{2}$. La figure ci dessous visualise ces différents temps.

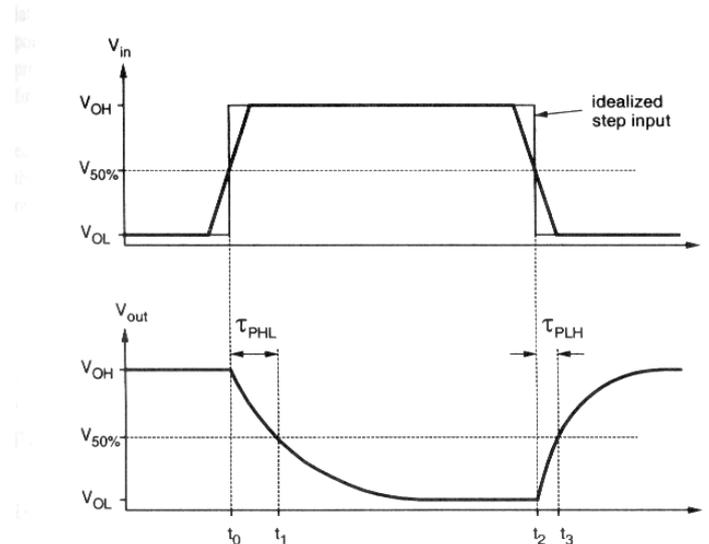


Figure 7.14 : temps de propagation d'un inverseur CMOS (d'après Kang et Leblebici)

- **Temps de montée et de descente** : on utilise les définitions courantes pour ces temps. On définit deux tensions, $V_{10\%}$ et $V_{90\%}$. Les temps de montée (« rise ») et de descente (« fall ») sont donnés sur la figure 7.15.

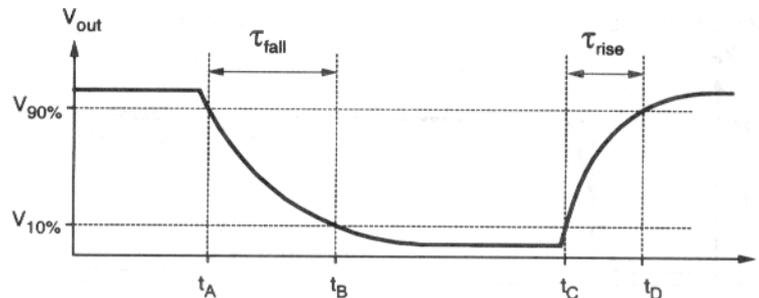


Figure 7.15 : temps de montée et de descente

Le calcul des différents temps est effectué à partir de l'équation différentielle associée au nœud de sortie, qui est donnée par :

$$C_{load} \frac{dV_{out}}{dt} = i_C = i_{D,p} - i_{D,n} \quad (6)$$

L'expression du temps de retard τ_{PHL} s'écrit alors :

$$\tau_{PHL} = \frac{C_{load}}{k_n(V_{OH} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{OH} - V_{T,n}} + \ln \left(\frac{4(V_{OH} - V_{T,n})}{V_{OH} + V_{OL}} - 1 \right) \right] \quad (7)$$

- **Dissipation de puissance en commutation d'une porte CMOS.** Comme on l'a vu précédemment, la consommation en statique d'un CMOS est pratiquement nulle. En revanche, durant la commutation, les charges et décharges de la capacité C_{load} vont entraîner une consommation et donc une

dissipation de puissance. Lorsque la tension d'entrée passe de « 1 » à « 0 », le PMOS conduit alors que le NMOS est bloqué, le courant de drain du PMOS sert à charger C_{load} . Dans le cas où la tension d'entrée passe de « 0 » à « 1 », cette fois-ci c'est le NMOS qui est passant, le PMOS bloqué et la capacité C_{load} se décharge à travers le NMOS. Si on calcule la puissance dissipée sur une période (en supposant des tensions de sortie et d'entrée périodiques de période T), cette puissance moyenne s'écrit :

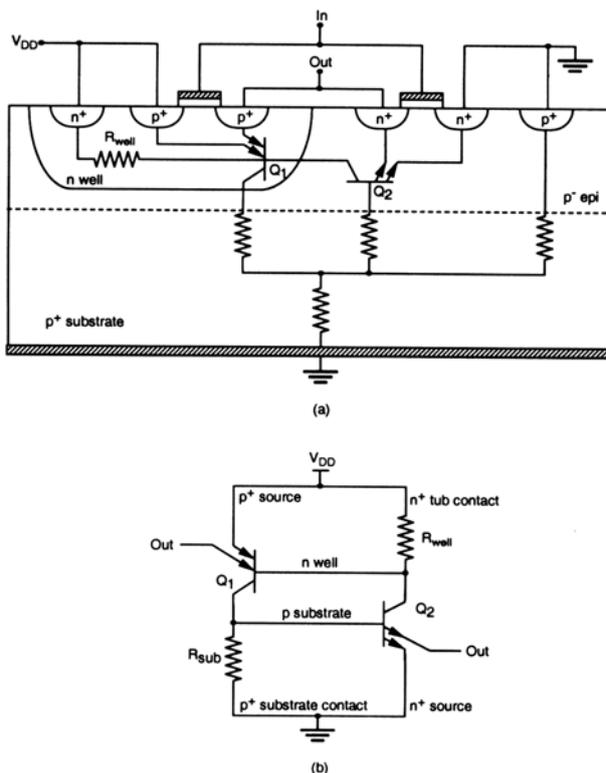
$$P_{moy} = \frac{1}{T} \int_0^T v(t) \cdot i(t) dt \quad (8)$$

en décomposant par demi période (PMOS ou NMOS conduit !), on obtient la relation classique de la consommation d'un CMOS :

$$P_{moy} = C_{load} V_{DD}^2 f \quad (9)$$

On voit donc qu'à fréquence de commutation élevée, l'avantage du CMOS est moins évident. Une autre source de consommation (donc de dissipation) est la dissipation de court circuit. En effet, durant la transition « 0 » à « 1 » ou « 1 » à « 0 », les deux transistors conduisent en même temps (d'où la notion de court circuit) et un chemin pour le courant existe entre la tension d'alimentation et la masse. La figure (4.37 du Weste) montre bien ce phénomène. En fait, lorsque le CMOS se trouve dans les régimes B, C et D de la figure 7.11, ce courant existe. La forme de ce courant est généralement « triangulaire ». **Il est à noter que dans le cas d'inverseurs chargés par une capacité, ce courant est marginal dans le courant total responsable de la consommation du circuit.**

Appendice 1 : le « latch up ».



Le phénomène de latch-up est identifié comme la mise en conduction d'une structure parasite PNPN équivalente à un thyristor ou plutôt à une diode Schokley, ou encore à deux transistors bipolaires imbriqués. Ce phénomène qui se traduit par une forte élévation du courant prive le circuit de sa fonctionnalité et s'il se prolonge peut provoquer des dégâts irréversibles. En général ce phénomène apparaît à la mise sous tension du dispositif, mais un système mal protégé des parasites (de quelques natures qu'ils soient) peut être sensible au « latch-up ». La structure d'un inverseur CMOS faisant apparaître les transistors bipolaires parasites est présentée ci contre.

Rappelons rapidement le fonctionnement d'une diode Schokley (« silicon-controlled rectifier ou SCR »). La structure PNPN peut se résumer à trois diodes PN, deux polarisées en direct et une, prise en sandwich entre les deux, polarisée en inverse. Tant que la structure Schockey n'est pas déclenchée, le courant qui circule est uniquement du au courant inverse de saturation de la diode centrale. Les courants collecteurs des deux transistors sont les courants de fuite des jonctions base-collecteur I_{CB0} (les transistors sont bloqués). Il faut noter également que leurs gains en courant sont faibles. Lorsque le dispositif est déclenché, les transistors saturent et on se retrouve avec le schéma équivalent ci-dessous. Cet état sera maintenu même si le facteur déclenchant disparaît pourvu que le courant se maintienne à une valeur supérieure à I_H appelé courant de maintien (« Holding current ») et une valeur de tension $V_H = V_{BE1,sat} + V_{CE2,sat} = V_{BE2,sat} + V_{CE1,sat}$, appelée tension de maintien.

schéma, il est aisé de montrer que le courant qui I qui circule dans le SCR est donné par :

$$I = \frac{I_{CB01} + I_{CB02} - (I_{RS}\alpha_1 + I_{RW}\alpha_2)}{1 - (\alpha_1 + \alpha_2)} = \frac{I_{CB0} - (I_{RS}\alpha_1 + I_{RW}\alpha_2)}{1 - (\alpha_1 + \alpha_2)}$$

On voit donc que si la somme des gain en courant est proche de 1, on se trouve dans le cas de figure de contre réaction positive qui va faire diverger le courant vers des valeurs pouvant être très élevée, c'est le « latch-up ». Pour que le déclenchement du latch-up se produise, il faut donc que $I \geq I_H = (V_{DD} - V_H) / R_T$ avec R_T résistance équivalente parasite du dispositif à l'état ON. On peut alors montrer que la condition sur les gains en courant s'écrit :

$$\alpha_1 + \alpha_2 \geq 1 + \left(\frac{\frac{R_T}{R_W}\alpha_1 + \frac{R_T}{R_{sub}}\alpha_2}{\left(\frac{V_{DD}}{V_{BE}} - 2 \right)} \right)$$

Pour éviter le déclenchement du latch-up, on a donc intérêt à augmenter au maximum la valeur du terme entre crochet. Pour ce faire on voit que l'on doit diminuer au maximum les résistances du puits n et du substrat